Label-free single-molecule nanopore analysis of protein post-translational modifications

Pablo Martin-Baniandres

St Hilda's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2023

Abstract

Despite significant advancements and refinements in single-molecule nanopore technology for DNA and RNA sequencing, its applicability to protein analysis remains limited. Proteins, consisting of 20 diverse building blocks, undergo post-translational modifications, fold into complex three-dimensional structures, and lack uniform charge distribution. These factors complicate signal interpretation and hinder their capture and movement through nanopore sensors.

In the endeavor to enhance protein sequencing methodologies, ongoing research predominantly classifies into two primary strategies. Firstly, there are initiatives aimed at emulating the ordered translocation events observed in DNA/RNA nanopore sequencing, a process guided by enzymatic mechanisms. These approaches frequently employ molecular motors, necessitating preliminary sample processing steps to introduce recognition tags and labels. Alternatively, more straightforward strategies center on the examination of short peptides. Nonetheless, a substantial challenge persists in decoding the signal and reconstructing the original protein sequence from these peptide fragments, presenting a notable bottleneck in this methodology.

Electro-osmotic flow (EOF), which occurs when an applied potential induces liquid motion in narrow tube-like conduits, is a phenomenon reported in nanopores for decades. EOF generates forces capable of increasing the capture rate and intra-pore residence time of neutral molecules and folded protein substrates. Despite promising results, limited work has explored the utilization of EOF for driving proteins across nanopores in a manner compatible with sequence-level readouts.

In this study, we capitalize on the robust and well-characterized heptameric protein nanopore, αHL. We engineer this nanopore by introducing charged residues near its constriction, creating a series of ion-selective pores with varying EOF strengths. We then use a model protein, Trx, to construct concatemeric molecules featuring 2, 4, 6, 8, and 9 repetitions of a Trx-linker unit, resulting in chain lengths ranging from approximately 300 to over 1200 amino acids. These constructs are transported across the engineered pores in a label-free, enzyme-less manner under mild denaturing conditions. This process generates distinct current signatures corresponding to the number of units within each construct, indicating a mechanism of co-translocational unfolding.

In an effort to assess the applicability of this approach, we introduce three types of post-translational modifications (PTMs)—phosphorylations, glutathionylation, and a glycosylation mimic—at five different positions within the middle Trx-linker unit of a nonameric concatemer. We engineer a universal modification site at these locations. The detection and discrimination of the three different modifications are successfully demonstrated at two of these positions.

While this approach may not provide de novo sequence information, with further optimizations, it holds the potential to become a powerful method for fingerprinting PTMs. It could enable the characterization of underivatized full-length proteoforms from cells and tissues, serving as a valuable tool for proteomic analysis.

Aknowledgements

I extend my heartfelt gratitude to my esteemed supervisor, Professor Hagan Bayley, for giving me the invaluable opportunity to embark on this journey of scientific research within one of the most enriching and stimulating environments. His profound comprehension of the intricacies of science has illuminated countless discussions during our group meetings, providing me with a deeper and more nuanced understanding of the world of science.

I am also indebted to Professor Yujia Qing and Wei-Hsuan Lan for their support, assistance, and substantial contributions to the work presented in this thesis.

Dr. Matthew Cornall and Dr. Giacomo Mazzotti deserve special recognition for their continuous support, camaraderie, life advice, and the spirited chess battles that added a touch of competitiveness to our journey.

To my parents, Jose Ramon and Josune, I owe immeasurable gratitude for their patience and unconditional support throughout this journey, which took me far from home. Your encouragement has been my rock.

I also want to express my heartfelt thanks to my grandparents, Maria Jesus and Isidro, who always believed in me even when they may not have fully understood the intricacies of my academic pursuits.

Lastly, I want to thank Carla, whom life unexpectedly gifted to me when I needed it most. You have been my “Cus D'Amato” throughout this entire journey at Oxford. I do not think this work would have come to life without you. APR

This thesis would not have been possible without the collective efforts and support of each of you. Thank you for being a part of this academic and life journey.

Ya está hecho, abuelo.

Collaborations

Part of the work contained in this thesis has been done in collaboration:

Genes encoding wild type Trx-linker concatamers were assembled by Stephanie Board, who works under the supervision of Prof. Sergi Garcia-Manyes at King's College London. Initial protein expression trials were carried out by her.

The data analysis of wt Trx-linker concatemers has been done in collaboration with Wei-Hsuan Lan and Prof. Yujia Qing.

The electrophysiology characterization and data analysis of post-translationally modified Trx-linker concatemers was undertaken in collaboration with Wei-Hsuan Lan and Prof. Yujia Qing.

It is worth noting that the research on Trx concatemers has undergone peer review and has been published in Nature Nanotechnology. Wei-Hsuan Lan and Prof. Yujia Qing are co-authors of this publication, which was prepared collaboratively. In instances where certain figures from this publication have been incorporated into this work, proper acknowledgment has been provided within the legend of each figure.

Unless explicitly acknowledged above, the research presented in this thesis represents my own work.

Contents

Abbreviations

+P -P Phosphorylated / non-phsphorylated

DPhPC 1,2-diphytanoyl-sn-glycero-3-phosphocholine

DTT Dithiothreitol

EOF Eelectoosmotic force

EPF Electrophoretic force

GSH Glutathion sulfate

I-V Current-voltage

IVTT In vitro transcription and translation

MS Mass spectroscopy

PKA Protein kinase A

POI Protein of interest

PTM Post-translational modification

SCP Single-cell proteomics

SDS Sodium dodecyl sulfate

SLN 6' sialyllactosamine

TCEP Tris(2-carboxyethyl)phosphine hydrochloride

Trx Thioredoxin

αHL α-hemolysin

1 General introduction

1.1 Stochastic sensing

The classical example of a pore-based sensing method is the Coulter counter (Beckman Coulter, Fullerton, CA, which measured resistive pulses across an aperture (ranging from ~20 um to as large as 2 mm) bathed in electrolyte solution to size and count biological cells, microorganisms, and other small particles. This approach involved generating a constant electrolyte flow driven by a pressure gradient across the aperture. When an analyte particle entered the aperture, it effectively displaced a volume element of electrolyte solution equivalent to the particle volume, generating an increased resistance during the residence time of the particle within the aperture, which was monitored by recording the ionic current under a transaperture voltage. The number and amplitude of the perturbations on the measured voltage in the form of voltage pulses, was used to determine the concentration and size of the analyte, respectively [1].

To increase the sensitivity of the technique, experimental Coulter-type devices that make use of smaller pores (~0.3 μm) were used to detect nanoscale particles and viruses [2]. In practice, depending on the size of the aperture, the analyte dimensions might range from a microbe to a molecule. As a general rule, the diameter of the analyte species must be comparable to the inside diameter of the aperture. The lower size limit

of a given aperture is determined by the electrical noise caused by the passage of ionic current through the aperture. Once the magnitude of the voltage pulse equals the noise level, detection becomes impossible. The upper size limit is determined by the ability to maintain the particle in suspension.

It was not until nanometre-sized holes, often generated by biological components such as membrane proteins, were employed, that the technique achieved enough resolution for the detection of single analytes even smaller in size, such as molecules and ions. Over the years, continuous progress in protein engineering techniques has expanded the potential of stochastic sensing with nanopores, culminating in a potent method for single-molecule analysis. It must be pointed out, that synthetic analogues of nanoscopically sized channels have also been widely employed with the purpose of molecule and ion sensing. Its fabrication and performance will be discussed in subsequent sections.

In the simplest example of single-molecule nanopore stochastic sensing, a unique nanopore is inserted into a membrane that separates two compartments, termed as the cis and trans compartments, containing a conductive electrolyte solution (Fig. 1.1). A single channel acting as a sensor is required to detect individual analyte-channel interactions. When an electric field is applied across the pore, the resulting ionic current that flows through the pore can be monitored. Molecules that are driven through the pore as a consequence of the electric field, can interact with the pore or obstruct it,

which in turn will modulate the ionic current. The transient changes in the ionic current provide detailed information about the analytes[3].

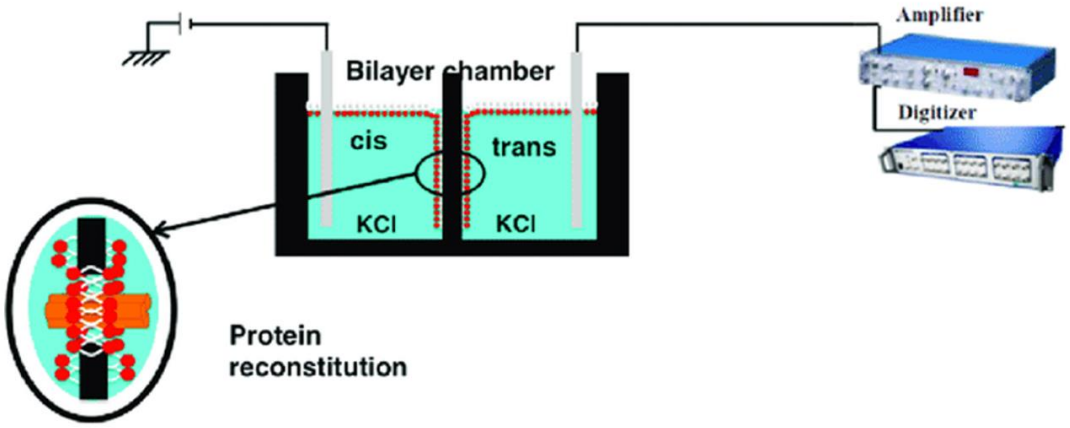


Figure. 1.1: A typical planar lipid bilayer electrophysiology experimental set up. Reproduced from [33].

The following sections of this chapter offer a comprehensive overview of various aspects within the field. These include standard electrophysiology setups, diverse categories of pores, general strategies for biomolecule sensing, the current state of proteomics and main single-cell proteomic techniques and, lastly, the distinctive intricacies associated with protein analysis using nanopores, along with the most advanced methodologies in this domain.

1.2 Experimental set-up for electrophysiology measurements

The basic structure of a nanopore can be envisioned as a nanometric aperture placed into a membrane, creating a pathway for an ionic current to pass through. If an external analyte interacts with the nanopore, by entering or coming into its close vicinity, the flow of current is perturbed. The frequency of these disruptions informs on the

concentration of the analyte, and the duration and amplitude of the interruptions can generate a distinctive signature, which can be utilized to identify the analyte.

Current methods for single-channel recording evolved from two coetaneous efforts, the patch-clamp technique, and the development of planar bilayer recording. The patch-clamp technique (Fig. 1.2) is a powerful and versatile method for studying the electrophysiological characteristics of biological membranes, for which developers, Erwin Neher and Bert Sakmann, received The Nobel prize in Physiology and Medicine in 1991. Soon after its development, it was adopted by numerous laboratories and subsequently caused a revolutionary advancement of many research areas in both cellular and molecular biology [4].

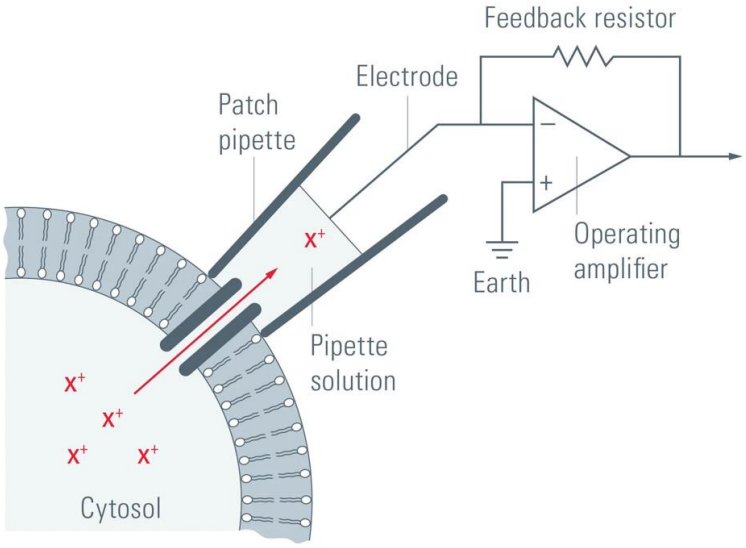


Fig. 1.2: Diagram of the patch-clamp technique. A glass pipette containing electrolyte solution is tightly sealed onto the cell membrane, electrically isolating a membrane patch. Currents flowing through the channels in this patch flow into the pipette and can be recorded by an electrode that is connected to a highly sensitive differential amplifier.

Initially, the patch-clamp approach was envisioned to study a small section of a cell membrane, which would contain multiple channels of different nature. Isolation of such section or “patch” involved using a glass pipette filled with electrolyte solution.

Two main configurations of the technique are possible with the patch-clamp technique, in what is called whole-cell measurements: In the current-clamp configuration, current is kept constant, and channels present within this patch of cell membrane pass through applied currents that flow into the pipette, which are then recorded by an electrode connected to an amplifier, informing on changes in the membrane potential. Alternatively, a voltage-clamp configuration can be utilized, in which voltage is kept constant, whereby current is injected into the cell to offset the membrane potential. By recording this compensation current, conclusions can be drawn about the conductance of the membrane and voltage-dependence of ion-channels. The latter approach permits single-channel recording directly from biological membranes, although it is limited by the number and type of pores that are found in its natural membrane environment [5].

During the same period, a very similar technique, known as single-channel recording, was first demonstrated by Hladky and Haydon in 1970, who observed the opening and closing of individual channels formed by gramacidin A, a peptide antibiotic. In this method, a synthetic bilayer was reconstituted separating two aqueous compartments, where individual channel forming proteins could be reconstituted or vectorially inserted. This method permitted similar configurations as the patch-clamp technique

(i.e., voltage-clamp and current-clamp), but unlike the latter, offered a crucial advantage from an experimental point of view, since it provided access to both sides of the membrane (cis and trans) and, therefore, both entrances of the channel, and the electrolyte solution composition on both sides [4].

Following on the latter configuration, advances in single nanopore reconstitution within synthetic bilayers, has ultimately enabled single-channel recording of protein pores outside its natural cell membrane environment. This configuration can be used to interrogate a vast number of different protein nanopores, including synthetic and chemically modified ones, which can interact with a wide spectrum of analytes in numerous ways.

In this work, for single-channel recording, a single protein nanopore was vectorially inserted into a synthetic bilayer separating two electrolyte compartments, and a fixed voltage was selected that generates an ionic current flowing through the pore that was constantly monitored. Modulation of such ionic current upon analyte interaction or passage across the pore, was used to identify and characterize the analyte.

Regarding the creation and composition of the lipid bilayer, a widely used model is the planar lipid bilayer (PLB) technique. This method involves creating a bilayer on a small aperture of about 30-100 µm diameter in a 25 µm thick PTFE film that separates two compartments filled with buffer solution (cis and trans). The volume of the compartments can be adjusted, but typically 0.5-1.0 mL is maintained to prevent

evaporation, which can become more problematic in smaller volumes, and because it facilitates the handling of pipettes and the bilayer formation process.

To generate the bilayer, the Montal-Mueller technique is frequently employed [6]. This technique involves treating the aperture with hexadecane (1-10% v/v in pentane) and letting it dry. Then, buffer solution and a few microliters of lipid dissolved in pentane are added to both sides of the bilayer. The lipid molecules form a monolayer at the air-water interface, which eventually coalesces at the aperture by repeatedly raising and lowering the buffer level in each compartment, forming a bilayer.

The lipid predominantly used in PLB studies is diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) due to its stability and lack of a phase transition at typical experimental temperatures [7], although various other lipids have been tested [8]. Asymmetric PLBs have been reported with different lipid compositions in each leaflet [9], [10], yet it is uncertain how long they remain stable and to what extent the membrane becomes homogenized once disrupted.

For electrical measurements, one Ag/AgCl electrode is placed in each chamber. In this work, the nanopore was always introduced into the cis compartment (ground), and the potential was considered positive by convention when it was greater on the trans side. Detergent-solubilized nanopores were then added on the trans solution and insertion promoted into the bilayer by applying a transmembrane potential and stirring the chamber contents, creating perturbations in the membrane structure and reducing the

energy barrier for insertion. Although not explored in this work, nanopores can also be reconstituted through the fusion of proteoliposomes [11], [12].

PLBs have a drawback in comparison to patch-clamp experiments, in that they have considerably higher capacitance, leading to amplified noise in current read-outs. Using smaller apertures is a possible solution to this issue, but it has been observed a marked reduction in nanopore insertion with a decrease in membrane area. Additionally, PLBs are metastable structures that typically last for several hours under optimal conditions, and more realistically, in the order of tens of minutes, particularly if harsh chemicals are used in the solution (e.g., denaturing agents such as urea, guanidine hydrochloride, etc). Significant advances in membrane chemistry, such as the introduction of polymerizable lipids [13], hydrogel-anchored bilayers that offer improved mechanical stability [14], and the utilization of amphiphilic polymers [15], have extended the lifespan of these bilayers.

1.3 Nanopores

Nanopores are, by definition, nanometer-sized pores. In the context of this work, nanopores refer to either pore-forming proteins, or to holes in synthetic materials, either of which are used for single-molecule sensing.

Pore-forming proteins are commonly found in nature, either as toxins or as channels for ions or molecules present in cell membranes. This type of pores, also referred to as

organic pores, are utilized as sensors of single-molecules due to its exceptional sensing capabilities, potential for high-throughput and low cost [16]. Their detection capabilities derive from the ability to allow the passage of an ion current in solution, and report on small differences in such current as analytes interact with the pore. They spontaneously insert into lipid bilayers creating nanopores with diameters ranging between ~1-4 nm, although larger pores, up to ~40 nm, can also be found in nature. In electrophysiology recordings, each pore exhibits a unique conductance that is linked to the channel characteristics, such as the internal diameter and charge.

One of the major advantages of protein nanopores in contrast with their synthetic counterparts, is that they can be subjected to Angstrom-level precise modifications using protein engineering to suit the sensitivity needs of a particular range of analytes and facilitate its detection [17]. Examples of such modifications include, i) changing the internal architecture and chemical properties by residue mutation, so that it affects the interaction and/or passage of an analyte, ii) incorporation of chemical or DNA probes, attachment of ligands/aptamers (usually near the pore entry) to bind target analytes, and iii) attachment of molecular motors such as processive enzymes to control the rate of translocation of polymeric substrates such as DNA or proteins[18]. Description of several modification strategies that enhance the sensing capabilities of biomolecules will be included in the following chapters.

However, a major limitation of biological nanopores is the noise present in the recordings, limiting the signal-to-noise ratio (SNR) and thereby the time resolution. The

electrical read-out is done by an amplifier, which detects and amplifies the electrical signal to a digitizer, which converts the analogic signal to a digital one, an operation referred as analog-to-digital conversion (ADC). Digital low-pass (LP) filtering is typically used to reduce high-frequency noise, improving the SNR, but this comes at the expense of lowering the time resolution. Because analytes may translocate the pore or undergo chemical modifications at sub-millisecond timescale, or simply require high time resolution for the purpose of identification, and detection of those analytes is simultaneously limited by the ionic current noise, there is an inherent trade-off between SNR and time resolution [17].

One approach that permits a simultaneous gain in SNR and time resolution is the reduction in translocation speed of analytes. To date, the gold standard is the use of a molecular motor that provide a ratcheting mechanism that translocate the substrate at constant discrete step sizes, a method utilized in nanopore DNA sequencing [19], [20], [21]. In this work, an alternative approach has been utilized, in which the natural tertiary structure of a protein is utilized to stall a protein within the nanopore lumen as different domains are co-translocationally unfolded. Examples of this approach have been described previously by the Bayley group, where Trx was pulled through the αHL pore by attaching a 30-mer polynucleotide tag at either the N- or C-terminus, undergoing a stepwise translocation [22], [23].

Other limitation of biological nanopores is their thermal and chemical instability, including the PLBs which are used for their reconstitution, which reduce the effective duration of their useful life during recordings [18].

For this work, mutant αHL pores were engineered through introduction of charged residues at different positions facing its lumen, with the purpose of creating ion-selective channels. αHL is an extracellular pore-forming cytotoxin secreted by *Staphylococcus aureus* bacterial strain. Structurally, it is formed by seven subunits (Fig. 1.3), which can be in vitro assembled by incubating the monomers in the presence of rabbit red blood cell membranes. Generation of the nanopore mutants was achieved after amino acid mutation within a single subunit gene, which after expression and assembly, creates homoheptameric mutants containing the mutations of choice. However, it is possible to create heteroheptamers by mixing monomers containing different mutations (or *wt* protein) at varying ratios, whereby the number of times a subunit is present within the final heteroheptamer can be partly controlled. In this instance, it is necessary a separation step via gel electrophoresis of the mixture of heptameric pores, which is enabled by the additional presence of a genetically encoded 8-residue aspartic acid tail on the C-terminus of the mutant monomer protein. The presence of this tag gradually reduces the migration of the heteroheptamers as more mutant units are present within the final assembly.

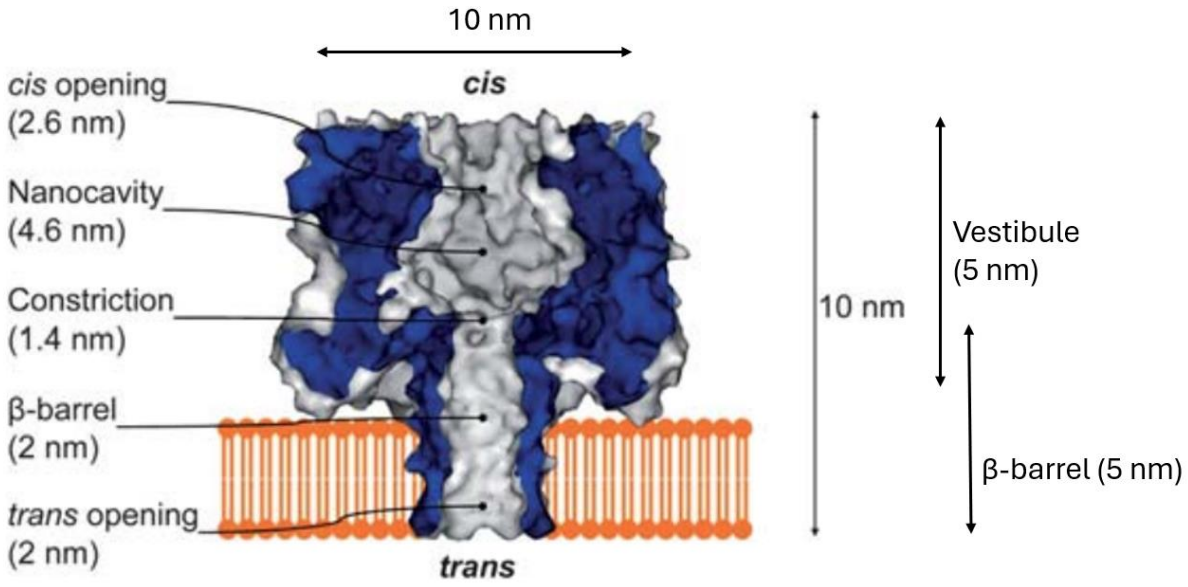


Figure. 1.3: Molecular graph of the heptameric αHL pore in the lipid bilayer. The dimensions of the various regions in the lumen of the pore are provided. Reproduced from [32].

The other main group of nanopores are synthetic ones, also known as solid-state nanopores (Fig. 1.4), including nano-scale apertures formed in silicone membranes (most commonly silicon nitride), two-dimensional (2D) materials, such as graphene, boron nitrite, molybdenum disulfide, and SiO2 (glass nanopores) [24], [25], [26], [27]. Current sensing, amplification, and recording is the same as for biological nanopores. Despite the numerous ways in which they can be drilled into membranes, e.g., laser-etching, reactive ion etching, dielectric breakdown, transmission electron microscope (TEM) focused, or by ion beam milling, they still lack atomically precise dimensions [17]. This type of pores expands the width of the material onto which they are created, ranging from ~0.3-30 nm, and feature diameters from under 1 nm to 10s of nm. Because of the ability to tune pore geometry, and the superior mechanical, chemical and thermal stability of solid-state membranes, considerable work has been performed in this field, including the development of alternative sequencing/diagnostic strategies, and new

membrane materials. However, these efforts have been only moderately successful, and to date, these types of pores can at best offer a shallow fingerprinting method for DNA and proteins, and need further development before they can match the capabilities of protein nanopores. Additionally, because of the inert nature of the materials with which they are made, they lack the chemical specificity of protein nanopores, and make the addition of chemical handles quite challenging [17], [18].

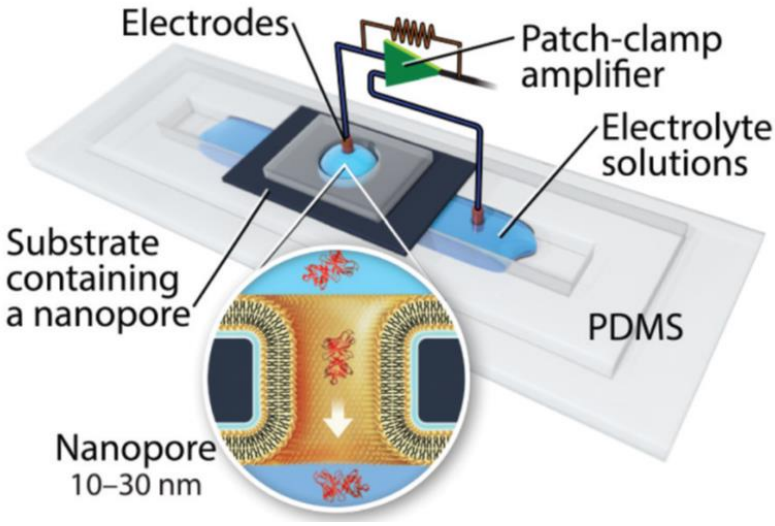


Fig. 1.4: Basic principles of single-channel recording through a solid-state nanopore. Experimental set-up for single-channel recording through an electrolyte filled solid-state nanopore; the inset shows the translocation of a single protein through a nanopore with a lipid–bilayer coating. Reproduced from [35].

However, a few examples of functionalization of solid-state nanopores can be found in the literature. One approach is to modify the surface with more reactive materials. For example, it has been shown that solid-state silicon nitride nanopores could be coated with a thin gold film to enable functionalization with a self-assembled thiol monolayer, reducing the diameter of the nanopore, and enhancing its sensitivity. Alkane-thiols

modified with nitriloacetic acid ligands were added in the monolayer, which in the presence of Ni2+, could bind His-tagged proteins and monitor its binding kinetics. Because of the strong binding, which reportedly could trap individual proteins for up to four hours, protein-protein interaction, between the anchored protein and specific antibodies, could also be characterized [28].

In another example, a non-covalent modification was reported, whereby solid-state nanopores were coated with a lipid bilayer enabling protein fingerprinting. The relatively large diameter and fluid lipid bilayer prevented non-specific protein adhesion to the nanopore walls, and protein tethers were embedded in the bilayer, to increase the time resolution by interacting with the substrates. Folded protein analytes were then trapped undergoing free rotation inside the nanopore, causing changes in the ionic current, which were used for the determination of shape and volume of proteins. Additionally, dwell time within the nanopore was correlated to the net charge [29].

Other developments include the fabrication of biomimetic pores. In one example, potassium selectivity was achieved by modifying a silicon nitride coated pore with triethoxysilypropylmaleamic acid (TESPMA) as a reactive handle for a cascade of further functionalization, ultimately enabling selectivity in pores with diameters below 3 nm [30].

Several other methods for to enable control over the translocation rate of analytes or to increase the time resolution have been described, which includes the use of ionic

liquids [31], the involvement of mechanical manipulation with a double pore system [32], optical trapping [33], [34] , and sequential DNA unzipping [35].

Finally, there is a third group of nanopores, known as hybrid pores, which combines the capabilities of biological and synthetic pores, what could be a future solution to the intrinsic limitations both types of pores offer (i.e., lack of specificity and choice for modification in solid-state nanopores and lack of stability in protein pores) and its feasibility and designs are currently being explored. The general scheme involves using the robust framework offered by solid-state nanopores to incorporate biological pores into them. For example, a thiol-terminated DNA 12-mer was clicked to an engineered cysteine residue located in an extended loop at the β-barrel opening of αHL. The DNA-modified pore was then electrophoretically captured into a solid-state nanopore, which was large enough to accommodate the β-barrel but not the cap, enabling recordings for extended periods of time [36].

The functionalization of biological and solid-state nanopores has brought about a significant transformation in nanopore approaches. Nevertheless, modifying existing pores presents certain challenges and limitations, so an emerging area which holds promising is the construction of *de novo* synthetic or semisynthetic nanopores, i.e., biological pores that do not exist in nature, in order to achieve higher modularity.

An example to this is the assembly of segments that belong to different proteins into one single nanopore construct. For example, based on the structure of Wza, an

Escherichia. coli polysaccharide transporter, a semi-synthetic self-assembled α-helical transmembrane protein was made that could successfully insert into a lipid bilayer. A consensus sequence membrane-spanning residues was identified, and a semisynthetic 35-residue peptide was synthetized that could assemble as an octamer structure [37].

Although with a much more modest results when it comes to biomolecule sensing, advances in the use of DNA origami have made it possible to assemble three-dimensional arrangements of DNA-only to form nanopores (Fig. 1.5) which can sense a variety of analytes [38], [39]. Other approaches utilize DNA as scaffolds for peptides to assemble into hybrid membrane-spanning nanopores, offering control in size and permutation around a central axis [40] . Finally, one last example includes the use of a large DNA origami sphere, permeable to ions, which was electrophoretically docked onto a lipid-coated solid-state nanopore. This set-up is known as the NEOtrap. The highly negatively charge sphere then induced a strong hydrodynamic flow caused by EOF that facilitated protein trapping and identification of several protein analytes as well their different orientation at the centre of the pore [41].

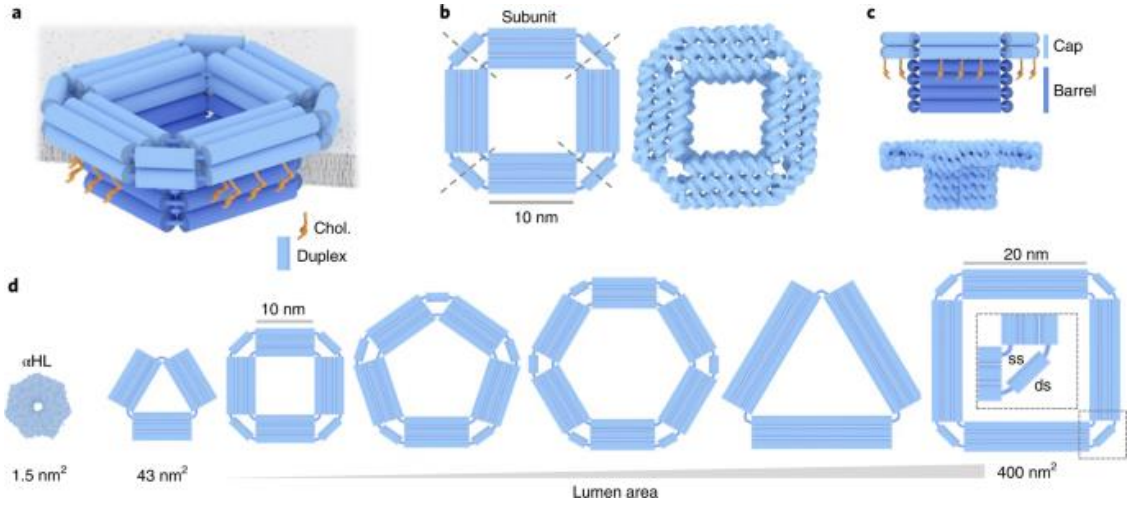


Fig. 1.5: Range of possible nanopore designs employing DNA origami. a, A DNA nanopore with square-shaped lumen composed of bundled DNA duplexes (cylinders, blue) inserted into a membrane. b, Top-down view of the DNA nanopore in schematic and molecular representations. The pore is composed of four inter-connected subunits. c, The pore in a side view illustrates its outer-membrane cap (light blue) and the membrane-spanning barrel (dark blue). Cholesterol anchors (orange) at the bottom of the cap facilitate the insertion of the pore's barrel into the membrane. d, Caps of DNA nanopores with different polygonal shapes and sizes cover a lumen area from 43 to 400 nm2. The protein pore αHL is shown for size comparison. Reproduced from [36].

1.4 Single-molecule nanopore sensing of biomolecules

Single molecule identification of biomolecules has been a long-sought goal, with essential applications at the research level, but also in medicine, i.e., diagnostics and follow-up of diseases. Nanopore technology offers a detection method capable of sensing single molecules with exquisite sensitivity. From small, simple analytes such as metal ions, small organic molecules, glucose, antibodies, etc, to more complex nucleic acids DNA and RNA, nanopore sensing has proven to be a promising technology with

potential to become integrated into the next state of the art devices for diagnostics, drug screening, food processing and environmental monitoring.

In order to improve detection potential, nanopores can be extensively modified either through mutagenesis or chemically, with unique binding components, resulting in the creation of sensors capable of identifying a diverse array of analytes (Fig. 1.6). In addition to this, monitoring of chemical and enzymatic reactions, and protein conformational changes has also been proven possible. Selected examples relevant to the present work will be described in this chapter.

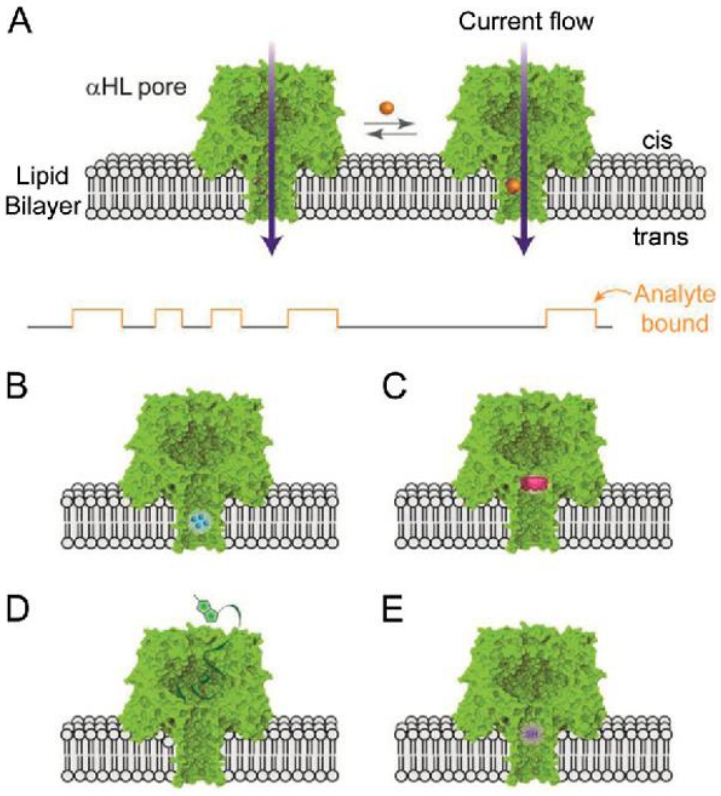


Fig. 1.6. Different approaches for single-molecule stochastic sensing with nanopores. a, Sensing with the αHL pore. The current carried by aqueous ions through an individual pore is monitored. Single analytes binding at a site or passing through the pore lumen are detected through transient changes in

the current. These stochastic current blocks function as ‘‘signatures’’ which, based on block amplitude and duration, allow for the identification of the unknown analytes. **b**, Binding sites can be formed by genetic engineering or, **c**, targeted chemical modification, e.g., with a molecule adapter (as shown). **d**, Large analytes can be detected by a ligand presented outside the pore lumen. **e**, Covalent chemical reactions occurring within the pore can also be detected. Here a single reactive thiol group is shown. Reproduced from [34].

A first group of approaches, comprise the use of nanopore sensors that can accommodate their analytes within their lumen. In one example from the Maglia group, cytolysin A (ClyA) (Fig. 1.7), a large (~5.8 nm inner diameter) nanopore was modified via substitution of a ring of glutamine residues with tryptophans [42]. This enabled prolonged capture of a set of folded ubiquitinated substrates (E2 Ubc4 enzyme) within the pore lumen, and discrimination among their different conformations. Furthermore, the ubiquitination reaction of a substrate protein could be monitored in real time.

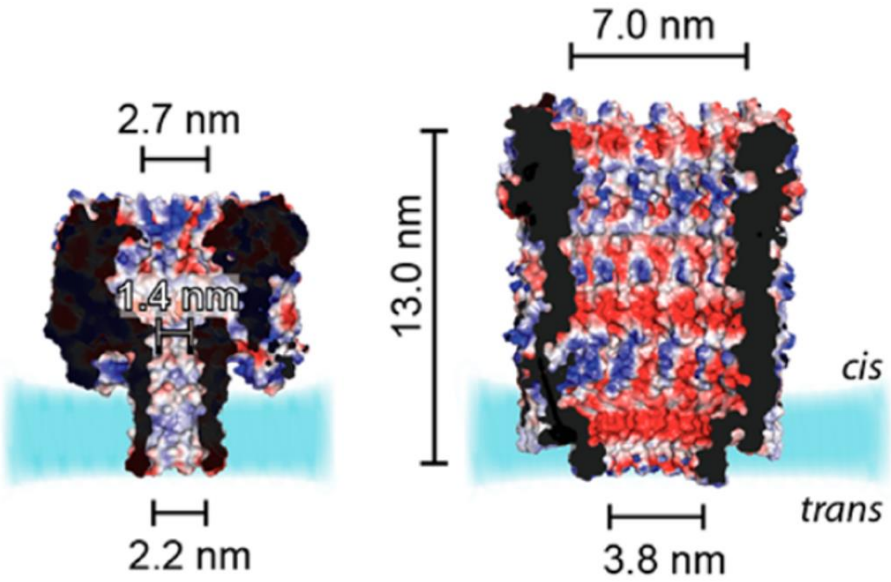


Fig. 1.7: Side-by-side comparison of sections of αHL (left) and ClyA nanopore (right). Colour coded charge density is shown. Reproduced from [37].

On a different nanopore, αHL, the protein sensor of choice in this work, was modified by introduction of positively charged residues through assembling a range of homo- and hetero- heptamers, reducing the translocation rate of DNA molecules [43]. However, the increased charge greatly reduced the passage of ions through the pore during DNA translocation, generating an almost complete blocked electrical signal, which translates in very high SNR. This example highlights the limitations of engineering nanopores via mutagenesis, as modulation of charge does also affect the ion current as well as the magnitude and directionality of driving forces.

The Maglia group has also shown other examples of engineering ion selective pores with the goal of generating an EOF that promotes analyte capture. For example, pleurotolysin A and B units (PlyAB) were oligomerized forming a large (~ 7.2 – 10.5 nm diameter) transmembrane pore which showed slight cation selectivity. By introduction of arginine residues at the constriction, this selectivity was enhanced, allowing capture of large human proteins, albumin (66.5 kDa) and transferrin (76-81 kDa) [44]. Additionally, frageatoxin C (FraC) nanopores have been extensively modified by the introduction of charged residues nearby its constriction (Fig. 1.8), which are capable of capturing and recognizing biomarkers in the form of small peptides and folded proteins with different charges, by monitoring the magnitude of current blockade and dwell times [45]. Tuning of the diameter of FraC nanopores has also been achieved by assembly of a different number of units (5, 6, and 8), with the goal of detecting a wider

range of analyte sizes [46]. Pores with smaller diameters display better resolution of smaller analytes, that would otherwise translocate the pore too quickly, and larger analytes can only fit within sufficiently large pores. Following this approach, dentification of peptide biomarkers differing by as little as one amino acid substitution (D and A) was demonstrated. A strong linear correlation between peptide size and current blockade was observed, and the application of FraC nanopores as a peptide mass identified was proposed, for which no prior knowledge on peptide identity would be required, in what would be a form of nanopore mass spectrometer [46] .

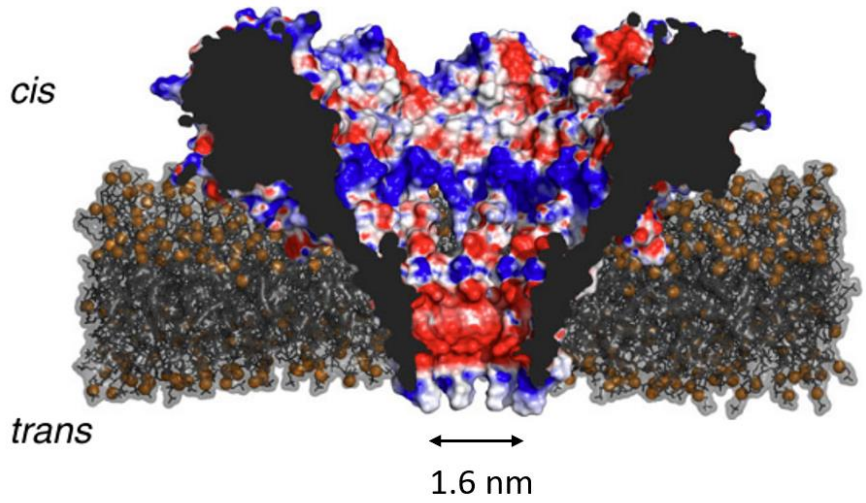


Fig. 1.8: Section of FraC nanopore. Colour coded charge density is shown. Reproduced from [38].

Another pore with potential interesting applications in DNA and protein detection is aerolysin, that has a slightly narrower diameter (1.37 nm) and displays reduced SNR compared to αHL. Modulation on capture and translocation rate of analytes has been shown by varying residues at the constriction of the channel [47]. Additionally, substitutions with alanine or tryptophan have been employed to vary the diameter of

the channel, and glutamine or glutamate to vary the charge, tailoring the pore chemistry to the specific requirements of the analyte.

Other alternatives to alter the charge of the nanopore lumen can be achieved through changes in buffer parameters such as pH or salt concentration. Examples of this have also been described by the Luchian and Maglia groups, where pH-tuning enabled balancing of EOF against EPF, enabling the capture of a range of peptides and analytes bearing different net charges [48]. In some cases, it was shown how EOF was responsible for translocation of analytes against the EPF dictated by their charge, highlighting the potential for EOF to govern directionality irrespective of charge. The combination of extreme pH values with mutagenesis for the introduction of charges residues has arisen as a promising tool for fine tuning the directionality of ion flows in pores, and ultimately determine the dynamics of the analyte in relation to the pore sensor.

An alternative to using large pores to accommodate analytes, is the use of external binding handles, such as DNA aptamers or ligands located in the vicinity of the pore entry, so that binding of analytes generates changes in the ion flow. These elements can either improve sensitivity or introduce selectivity towards a specific class of analytes. An example of this was shown by the Bayley and Maglia groups, where ClyA nanopores were engineered with DNA aptamers oriented to the cis side, evolved to bind thrombin or lysozyme [49]. These modified pores showed an increase frequency in capture by up to 15-fold, while simultaneously decreasing the capture rate of noncognate proteins

(twice as many blockades of the cognate ligand despite a ~100-fold excess of the noncognate counterpart in solution).

Protein adapters and DNA aptamers can also be incubated with the sample, rather than being covalently engineered into the nanopore architecture, and still amplify the ionic changes upon nanopore interaction. For example, glucose and asparagine, two relevant metabolites, could be directly detected from bodily fluids such as blood, urine, saliva and sweat, in a system in which analyte binding proteins behaved as sensors, and the nanopore (ClyA) as a signal transducer [50]. The analytes turned out to be too small to be directly detected, so larger SBD1 (substrate binding domain) and GBP (glucose binding domain) proteins were used, which underwent conformational changes upon asparagine and glucose binding, respectively, eliciting current changes that enabled monitoring of binding kinetics. The abundance of protein-analyte bound conformational events, directly correlated with the concentration of the analyte. A downside of this approach is that the protein sensors stochastically entered and exited the pore, in a diffusion-governed manner, which reduced the sensing time to only when the protein sensor happened to be occupying the pore. In a more sophisticated example, binding protein sensors AlkB demethylase and DHFR (dihydrofolate reductase) were trapped inside the nanopore (ClyA) cavity by engineering them with a polypeptide tag containing four additional positive charges, which favoured capture by the EPF [51]. Protein sensors remained trapped in the order of minutes inside the nanopore, greatly increasing the sensing time. This last piece of work highlights the

possibility of engineering protein adaptors by tuning their size and charge with the aim of increasing their residence time inside the nanopore.

The last class of pore modification include bioconjugation chemistry, which comprises the addition of chemical handles through modification of pre-existing reactive sites. One of the most exploited chemistries relies on the high reactivity of the cysteine residue, which sulfhydryl atom can be modified in simple, rapid click reaction steps. Cysteine residues, which are not naturally present in the wild type αHL, are introduced via mutations, which can then be modified either prior or following assembly (also known as *in situ* modification). Multiple examples can be found in the literature from the Bayley group.

In one example, using αHL as a model pore, a set of homoheptameric mutants were constructed displaying rings of seven cysteine residues at different location within the nanopore lumen. Monitoring the reaction kinetics of monomethoxypoly(ethylene glycol)-*o*-pyridyl disulfide polymers of differing lengths with the cysteine residues allowed to map the location of the pore constriction, as a trend of longer current blockades were observed when the cysteines were located closer to the constriction [52]. In a different example, photocleavable groups were used to prevent pore assembly until it was excised through light illumination [53].

Alternatively, engineered cysteines at different locations can also be further modified with chemical adapters prior to oligomerization to facilitate sensing of the analyte of

interest. Examples include cysteine functionalization with β-cyclodextrin, phenanthroline, or maleimides. In the first example, β-cyclodextrin was covalently attached in the two possible molecular orientations [54]. Two main advantages of this approach were highlighted. First, the orientation of β-cyclodextrin was controlled to allow analytes to bind through only one of the two entrances of β-cyclodextrin cavity. Second, the attachment was covalent, so the adapter could not dissociate, allowing constant monitoring without gaps. This set up was later tested with individual nucleoside 5'-monophosphates and showed continuous identification with accuracies averaging 99.8% (including methylated cytosine) based on current signature [55]. Therefore, it was proposed as an approach that could be fit for exonuclease DNA sequencing, whereby single DNA bases are enzymatically chopped at the pore entry as the DNA polymer is ratcheted towards the pore.

Other examples of the application of single-molecule nanopore sensing include using the nanopore cavity as a nanoreactor to monitor chemical reactions at the single-molecule level. For example, single cysteine heteroheptameric mutants have been used for monitoring different chemical reactions, such as the stepwise polymer growth or the kinetics of bond formation of thioarsenic bonds [56]. A nanopore pre-modified with 3,4-dimethoxy-6-nitrobenzylcarbamate was observed to undergo photoinitiated degradation via a multi-step process in which the kinetics of individual steps could be elucidated [57]. In a different example, cysteine - 2-cyanobenzothiazole (CBT) chemistry was explored [58]. CBT was observed to react reversibly with thiols, and intermediates with distinct lifetimes were detected. As a last example, maleimide has also been

explored as a reagent for modifying cysteines. In one example, αHL nanopores were pre-modified with three different maleimide linkers bearing additional sites for further conjugation: tetrazine, cyclo-octene, cyclo-octyne [59]. Cycloaddition reactions of this sites with different reagents was reported.

A last set of examples of the use of bioconjugation chemistry is the use of αHL nanopores containing one mutant subunit that contains a track of up to six cysteine footholds along the membrane-spanning β-barrel. Following this idea, a molecule capable of “walking” or “hopping” along the cystine track has been reported, using reversible thioarsenic chemistry [60] or disulfide exchange [61]. Control of directionality was achieved in the latter case, and the hopper could be functionalized to carry cargoes (e.g., a DNA molecule) demonstrated external control of directional motion (depending on the externally applied transmembrane voltage), processivity and no chemical fuel requirements. This principle was recently applied to the sequential identification of a DNA strand as it was translocated through αHL in chemical steps [62]. If sufficient throughput was achieved through further development, this approach could constitute a DNA sequencing method complementary to the enzyme-driven approach.

Another method for adding new chemical handles is the incorporation of unnatural amino acids to the nanopore, which can be achieved either by genetic engineering or by native chemical ligation. One notable example is the introduction of an amino acid with a terminal alkyne within the centre of the transmembrane β-barrel of αHL, which can later used for monitoring chemical reactions [63].

Finally, there is one last approach for functionalizing nanopores that does not require the use of genetic engineering to introduce reactive residue handles, and instead relies on the *in situ* modification of naturally present residues. Examples of such residues include lysine and methionine, which serve as potential sites for bioconjugation [64], [65]. In the particular case of methionine, its low abundance in proteins means it offers a higher degree of selectivity. Despite the fact that the chemistries required for modification are less specific than those for cysteines, several examples of successful functionalization of these residues with bioorthogonal azide and alkyne groups have been reported [66].

1.5 Current landscape of proteomics

In 1994, Marc Wilkins introduced the term proteome, as a fusion of the words protein and genome, and later founded the first dedicated proteomics laboratory in 1995. The proteome represents the complete array of proteins generated or altered within an organism or system, varying across different biological contexts. This can range from the proteome of a specific species like Homo sapiens to that of an organ such as the liver [67].

Proteins are the terminal product of central dogma, mediating almost all the functional processes of a living organism. The proteome is not static, it varies among cells and evolves over time. While the proteome to some extent mirrors the underlying

transcriptome, protein activity, often evaluated through reaction rates in associated processes, is influenced by multiple factors beyond gene expression levels. Traditionally, RNA analysis was used to assess this variability, but it was found to lack correlation with protein content. It is now understood that mRNA is not always translated into protein, and the quantity of protein produced from a given amount of mRNA depends on various factors including the gene being transcribed and the physiological state of the cell. Finally, proteins are, after all, the primary functional molecules of the cell, and its direct measurement and quantification is therefore justified. Proteomics directly confirms the presence of proteins and provides precise measurements of their abundance [68].

Proteomics offers a deeper understanding compared to genomics for several reasons:
i) Transcription levels of a gene provide only a rough estimate of its translation into a protein. Abundant mRNA may be rapidly degraded or inefficiently translated, resulting in low protein levels [69]. ii) Post-translational modifications profoundly influence protein activities; for example, phosphorylation can activate certain proteins. Techniques such as phosphoproteomics and glycoproteomics are employed to study these modifications. iii) Many transcripts give rise to multiple proteins through alternative splicing or post-translational modifications. iv) Proteins often form complexes with other molecules, such as RNA or other proteins, and their functionality is dependent on these interactions. v) The rate of protein degradation also significantly impacts protein levels within the cell, so while two protein types might be expressing at similar rates in the cell, their abundance might not correlate at all.

Proteomics comprise the expansive study of proteomes including the interactions, function, composition, and structures of proteins along with their cellular activities. It offers a deeper insight into organismal structure and function compared to genomics and its complexity surpasses that of genomics due to the dynamic alterations in protein expression influenced by temporal and environmental factors. Estimates suggest that while the human genome encodes for approximately 26,000-31,000 proteins, there are nearly one million human proteforms, as many of which undergo modifications like post-translational modifications (PTMs) and alternative splicing.

1.5.1 Single-Cell Proteomics (SCP)

Single-cell proteomics represents an emerging field which aims at deciphering the diversity of protein expression within individual cells. Unlike traditional proteomic approaches that analyse bulk samples, wherein thousands of cells are examined simultaneously, single-cell proteomics captures the unique protein expression patterns of individual cells, a capability that is crucial for unravelling the underlying mechanisms of diseases. For instance, tumours are usually composed of heterogeneous cell populations and characterizing the diverse cell types within them can improve their clinical treatment [68] . Moreover, SCP techniques enable the study of rare cell populations, such as cancer stem cells, which are inaccessible to bulk proteomics due to their limited sensitivity [70].

SCP techniques enable the identification of a vast array of proteins expressed within thousands of individual cells at a specific point in time, enabling the generation of protein maps that offer insights into disease development, progression, treatment effects, and serve as biomarkers for disease diagnosis and prognosis [68].

As mentioned earlier, solid research comparing single-cell RNA and proteome levels underscores the stability of the proteome compared to the transcriptome, highlighting the importance of exploring translation regulation at the single-cell level. Additionally, SCP provides high-resolution data at the proteome level, offering insights into post-translational modifications (PTMs) crucial for understanding early cell-signalling events

and cellular responses to environmental fluctuations and interventions. Therefore, SCP positions as a complement of single-cell RNA sequencing (scRNA-seq) by providing a deeper understanding of translational regulation and cellular heterogeneity in disease. [70].

Lastly, another key strength of SCP is that it can provide insights into protein-protein interactions and protein localization, critical for comprehending various signalling pathways, aspects not accessible by transcriptomic or genomic sequencing alone. SCP emerges as an indispensable tool for advancing the understanding of cellular biology and disease mechanisms [71].

1.5.2 SCP workflow

Single-cell proteomic technologies are still in their early stages compared to genomic and transcriptomic technologies and there are several obstacles that need to be addressed for accurately detect and quantify the numerous proteins present in a cell. One challenge is that proteins cannot be amplified like DNA or RNA, making it crucial to minimize losses during sample preparation and handling. Additionally, the dynamic range of proteins, which can vary from one to ten million copies per cell, poses a significant challenge. Current methods often only capture a fraction of this range, leading to the exclusion of low-abundance proteins, often referred to as the 'dark proteome [71], [72].

However, the most notable advancements that have allowed the progression of single-cell proteomics and its various methodologies are primarily attributed to technical innovations in sample preparation techniques that minimize loss, with the input material being a single cell [70]. Initially, the moderate success was reported with larger single cells such as blastomeres and oocytes, due to their higher protein contents, but later, more advanced sample preparation methods were developed, such as the single-pot solid-phase-enhanced sample preparation (SP3), which allowed the consistent identification of approximately 450 proteins from individual oocytes, marking a significant milestone in protein identification at that time [73].

Cell isolation

Prior to any SCP analysis, it is essential to isolate individual target cells from a heterogeneous cell population. The selection of isolation strategies depends on various factors, such as the study's objective, the cell source, the characteristics of the target cell, and the presence of potential contaminants in the source material. Common techniques for isolating individual target cells for SCP analysis include fluorescence-activated cell sorting (FACS) or flow cytometry, laser microdissection, manual cell-picking, random seeding/dilution, and microfluidics/lab-on-a-chip devices [68], but they will not be discussed further as they are beyond the scope of this work.

Sample preparation

Following the isolation of cells through suitable methods, protein extraction becomes the next crucial step. This can be achieved through various approaches, including

chemical methods utilizing lysis buffers containing combinations of ionic, non-ionic, and zwitterionic detergents, physical methods such as mechanical lysis via sonication, optical lasers, or the application of electrical fields, or a combination of both, tailored to the downstream assay platform. As already mentioned, this step is critical due to the minute size of the cells, with volumes in the femtoliter range. Considering that these cells contain approximately 60%–70% water and the remainder consists of dry matter, predominantly protein in picogram quantities, it is essential to minimize protein loss during the sample extraction process [68], [72].

Separation Techniques

When working with large cellular proteomes, a number of tools are available for separation of proteins of interest into samples of less complexity, namely, capillary electrophoresis (CE), ultrathin layer gel electrophoresis, multidimensional micro liquid chromatography.

The classical method par excellence is standard two-dimensional gel electrophoresis (2DE), which includes an initial first-dimensional separation by isoelectric focusing (IEF) based on the isoelectric point (pI) of proteins, followed by second-dimensional separation based on the molecular weight of the proteins, and is able to separate crude protein mixtures with high resolution (>3000–4000 proteins on a single gel) [68]. However, this technique requires large sample volumes, not available in SCP.

An alternative option in this case is carrying out electrophoresis in a microscale setup, a technique known as Capillary electrophoresis (CE), which comprise a family of analytical techniques that involve the separation of charged molecules within a narrow capillary tube as the separation medium, under the influence of an electric field, similar to standard electrophoresis. This microscale offers numerous advantages, including high separation efficiency, rapid analysis, minimal sample consumption (nanoliter sample volumes) and compatibility with various detection methods such as UV-absorbance, fluorescence and mass spectrometry [68].

A related technique is Ultrathin Layer Gel Electrophoresis, which achieves single-digit nanogram detection limits per spot, while also overcomes the limitations of traditional gel electrophoresis methods, offering automated, rapid separation with real-time detection, utilizing fluorescence technology. Its advantages include rapid heat dissipation and reduced sample loading compared to conventional gel electrophoresis [68].

Lastly, a second group of methods is multidimensional micro liquid chromatography (LC)**.** LC involves combining various chromatographic separation methods such as strong cation exchange (SCX) chromatography, reverse phase-liquid chromatography (RPLC), ion-exchange-reversed phase (IEC-RPLC), size exclusion-reversed phase (SEC-RPLC), and ion exchange-size exclusion (IEC-SEC), which enable the isolation of low-abundance proteins and post-translational modifications (PTMs) [68].

An alternative version of the above is high performance liquid chromatography (HPLC), wherein the solvent, which contains the sample, is delivered under high pressure e(up to 400 bar) into the packed column containing the stationary phase. The high pressure ensures a constant flow rate, which ensures reproducibility among different experiments. A yet higher resolution technique can also be found, which operates at even higher pressures (in the range of 400 – 1034 bar), which is known as ultra performance liquid chromatography (UHPLC) [68].

1.6 Analytical platforms for SCP

Analytical platforms for single-cell proteomics can be categorized in either qualitative, quantitative, or combined assay methods. And further distinguished by the technique used for analysis. Four major groups of strategies can be found, namely, antibody-based methods, which typically focus on a limited number of predetermined proteins and are commonly employed when working with a known database; bottom-up mass spectrometry (MS)-based methods, involving the digestion and analysis of the cell's proteomic content for proteome mapping; imaging-based techniques, capable of detecting and offering spatial insights into proteins; and single-molecule sequencing methods, which will be further discussed in the following chapters [68], [72].

1.6.1 Antibody-Based Assay Techniques

Fluorescence Flow Cytometry

Fluorescence flow cytometry (FFC), developed at Stanford University in the late 1960s, has become a widely used technique for single-cell protein analysis, especially in immunological studies with moderate multiplexing capabilities (up to 15 proteins). In FFC, cells or lysates are labelled with fluorescent markers specific to different proteins and then passed through a narrow flow cell. A laser beam illuminates the labelled proteins, and digital signals are generated and analysed to detect and quantify target proteins (Fig. 1.9). While FFC, exemplified by GFP-tagged workflows, has successfully identified numerous proteins from single Saccharomyces cerevisiae cells, challenges like overlapping signals and background noise persist. As a result, newer, more sensitive approaches are being explored to identify a greater number of target proteins in a single analysis [68].

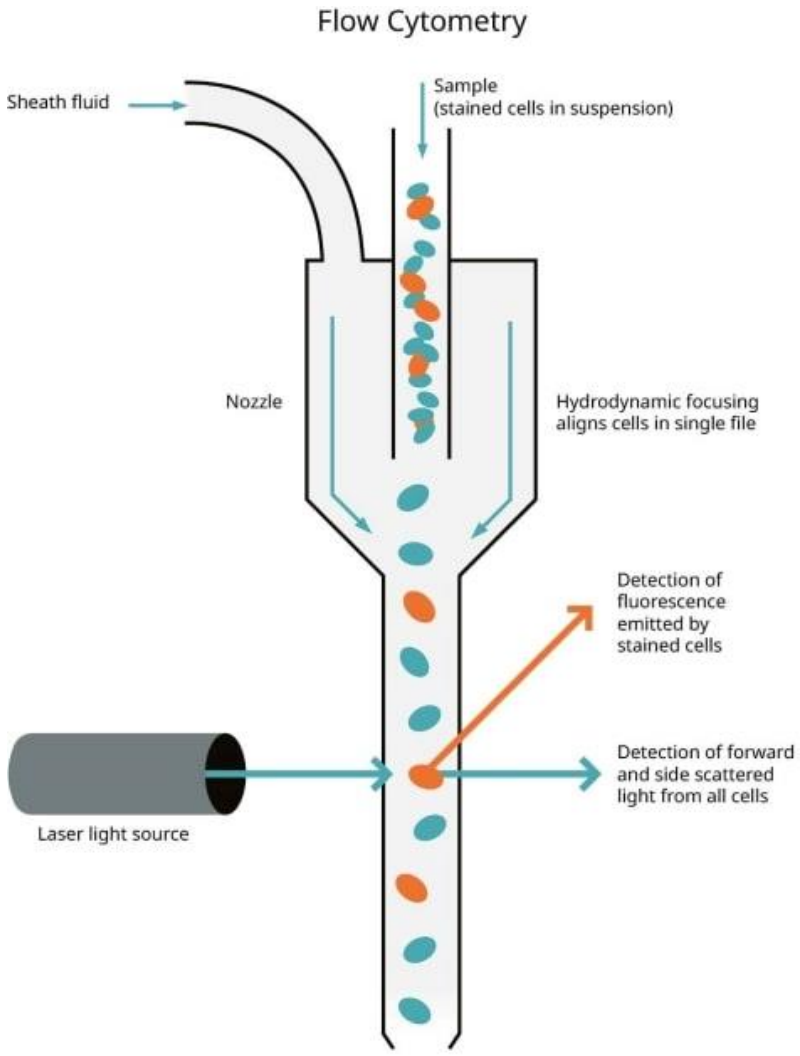


Fig. 1.9: Overview of a flow cytometer. Sheath fluid focuses the cell suspension, causing cells to pass through a laser beam one at a time. Forward and side scattered light is detected, as well as fluorescence [74].

Mass Cytometry

CyTOF represents an innovative approach to flow cytometry, which involves tagging antibodies with heavy metal ions to identify specific targets in the cells, rather than fluorochromes. Furthermore, the readout is achieved using time-of-flight mass spectrometry (TOF) (which will be further described later in this chapter), where cells labelled with these tagged antibodies are atomized and combined with heated argon

gas to desiccate the cellular components. By quantifying the heavy ions from the antibody tags, it is possible to determine the abundance of target molecules per cell, thereby inferring cellular properties (Fig. 1.10). CyTOF enables the simultaneous analysis of more than 50 markers on millions of individual cells, facilitating the characterization and identification of specific cell subtypes through proteomic profiling. Additionally, it also allows quantification of carbohydrates and lipids[68].

CyTOF offers several advantages over FFC, the main one being that data acquisition with no cell-dependent background signal interference [75].

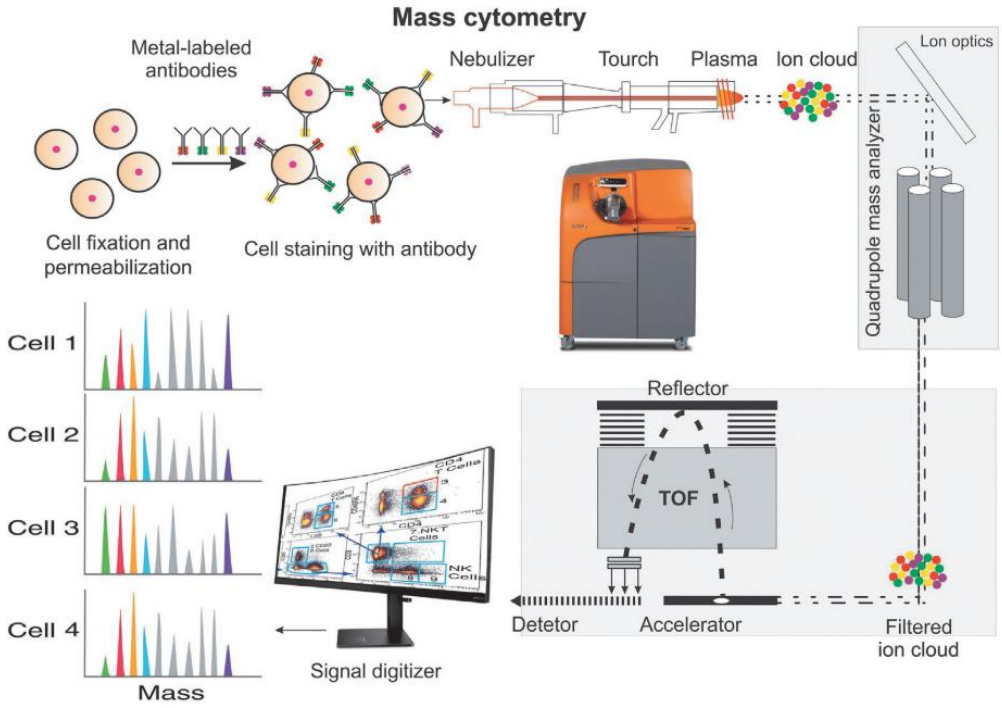


Fig. 1.10: A typical workflow in Mass Cytometry. Reproduced from [68].

Finally, a related platform worth including, chemical cytometry, involves labelling proteins with fluorophores on the lysine residues of proteins. This method is executed using microfluidics-based isolation techniques, where cells are lysed, followed by

capillary electrophoresis (CE) for separation. MS-based is then used identification. This hybrid technique combines aspects of both previous platforms, resulting in enhanced proteomics data quality [76].

Enzyme-Linked Immunospot Assay

The enzyme-linked immunospot assay (ELISpot assay) consists of a nanoscale version of the sandwich ELISA. It utilises affinity arrays of immobilized antibodies on a solid surface and reporter fluorophores which bind antigens secreted from single cells (Fig. 1.11). This method is commonly employed in immunological analyses, but it suffers from prolonged assay times and limited multiplexing capabilities [68].

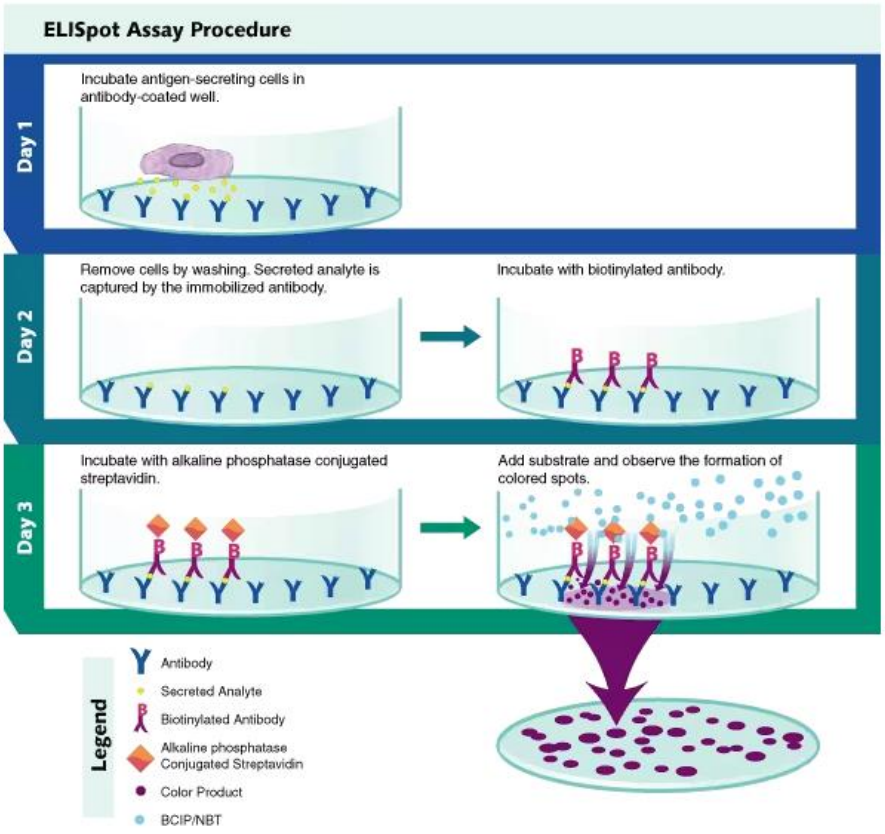


Fig. 1.11: ELISpot assay workflow. Reproduced from [77].

Single-Molecule Array (SiMoA)

SiMoA is a digital version of ELISA approach, which utilises femtoliter-sized chambers with paramagnetic beads coated with antibody capture agents. It allows the detection of thousands of molecules simultaneously in single-cell samples, with a detection range in femtoliters and the capacity to reach single-molecule resolution, theoretically enabling absolute quantification. However, limited multiplexing options, low throughput, and high costs for single-target analysis constrain its broader application [68], [75].

Microfluidic Antibody Capture Chip (MACS Chip)

This technique is a miniaturized version of ELISA, which combines fluorescent antibody labelling, a chip, and modern microscopy. It utilizes a fabricated microfluidic device with antibody patches printed on a coverslip. Optical-based methods are used for manipulations, lysis, and analysis of single cells. Upon cell lysis, which is performed by a single laser pulse, the proteins released react with the coverslip antibodies, and subsequently, the primary and/or secondary antibodies are added and binding tracked with total internal reflection microscopy (TIRF) coupled to an electron-multiplied charged coupled device (CCD), which enables single-molecule detection and reduction in background noise. Importantly, this method maintains cell viability until the moment of analysis, what may reduce proteome alterations during sample processing [68], [71].

Single-Cell Western Blotting

It is a versatile tool for single-cell resolution proteome analysis, that enables complex proteome analysis from different cell populations. Cells are settled in wells on a photoactive polyacrylamide gel, followed by in situ lysis, electrophoretic separation of cellular proteins, and subsequent antibody probing (Fig. 1.12). It can be integrated with fluorescence-activated cell sorting (FACS), allowing for target protein quantification from low starting cell numbers, with multiplexing options.

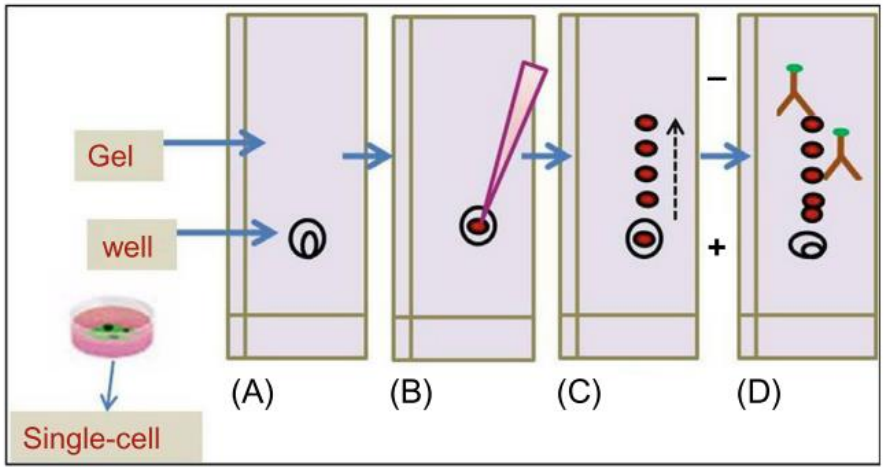


Fig. 1.12: Single-cell western blotting. a, A well on the gel is present for sample loading. b, Single-cell proteome loading from an isolated cell. c, Electrophoretic separation of proteins. d, Addition of fluorescence-tagged antibody specific for a POI and washing to confirm fluorescence signal and detection. Reproduced from [68].

1.6.2 Next-generation sequencing (NGS) combined techniques

Lastly, two more methods can be included within this group, which combine the use of Next-generation sequencing (NGS) and targeted antibody-based techniques. In contrast with single-cell RNA-seq, direct protein measurement does not contemplate molecule barcoding and amplification. However, it is possible to employ oligonucleotides for protein quantification by attaching them to antibodies that react specifically with proteins of interest. In this approach, antibodies are used as initial detectors, but protein detection and quantification is ultimately performed via the attached oligonucleotide, which can be amplified [78].

One popular method is CITE-seq, which enables measurements of both mRNA and cell surface proteins in microfluidic droplets. Here, oligonucleotide–antibody conjugates include an antibody-derived tag (ADT) sequence for antibody identification, as well as an amplification handle for PCR. Cells are co-encapsulated in microfluidic droplets along with a gel bead containing an oligonucleotide cellular barcode. This barcode is added to both the ADTs and mRNA molecules during reverse transcription (Fig. 1.13) [78].

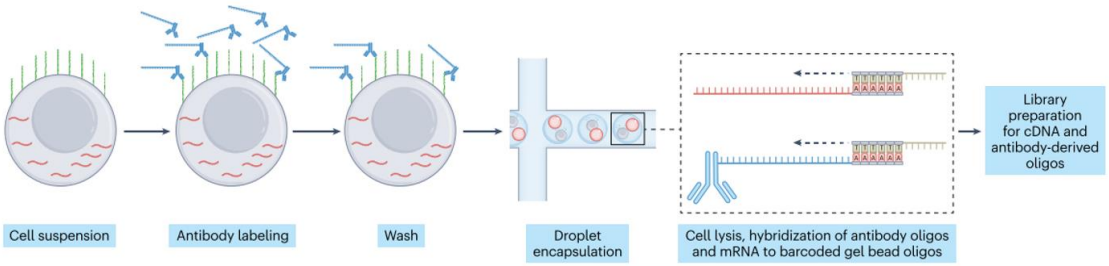


Fig. 1.13: CITE-seq. Antibody conjugates are used to label accessible epitopes on intact cells before droplet encapsulation. Once inside droplets, cells are lysed and both mRNA transcripts and antibody oligonucleotides are reverse transcribed and converted into separate NGS samples. Reproduced from [68].

One of the main limitations of CITE-seq and similar protocols is that signals can arise from any antibody conjugate in the cell suspension, what makes it crucial to remove excess conjugated through washing steps. These washes often result in cell loss, which makes it challenging working with small sampled, and despite thorough washing, background signal may persist [78].

A significant limitation of employing antibodies as reporter molecules is the potential for antibody cross-reactivity, which can pose significant challenges at higher levels of multiplexing. One way to reduce cross-reactivity is by using multiple recognition events, an approach exploited by proximity ligation assays (PEAs). In this method, signal is generated only when two antibodies bind their protein target, leading to a decrease in background compared to assays utilizing single-binder approaches. The antibody pairs are linked with reporter pairs of single-stranded DNA oligonucleotides. Upon antibody-protein binding, a DNA amplicon specific to the protein is generated, acting as a barcode. High-density real-time qPCR or an NGS platform quantifies these amplicons, enabling the identification of various proteomes (Fig. 1.14). This method, while requiring prior knowledge of expressed proteins, can be coupled with a microfluidic platform adaptation, allowing high-throughput cell assays using DNA-tagged antibodies. Additionally, it has recently been coupled to RNA sequencing, in a method

known as single-cell protein and RNA co-profiling (SPARC), whereby cDNA is synthesized from RNA by using reverse transcriptase and detected alongside coamplification and detection of cDNA [68], [78].

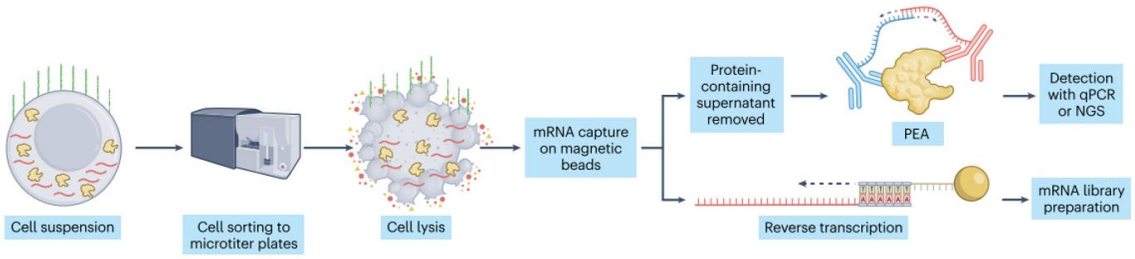


Fig. 1.14: Single-cell PEA. Individual cells are sorted in microtiter plates and lysed. mRNA molecules are captured by oligonucleotide dT magnetic beads, reverse transcribed and converted into NGS libraries. The supernatant is probed for protein levels by using a mixture of protein-specific PEA probes. Upon binding, these probes are extended and then quantified using highly parallel quantitative PCR (qPCR) or NGS. Reproduced from [68].

1.6.3 Mass Spectrometry

Single-cell proteomics using mass spectrometry (MS) has seen significant advancements in the last years and is currently gold standard for proteome analysis.

It's an analytical technique that separates ionized particles such as atoms, and molecules by using differences in the ratios of their charges to their respective masses (mass/charge; m/z) and can be used to determine the molecular weight of the particles.

The main components of a MS analyser include the following modules: an ion source, which splits the sample molecules into gas-phase ions; a mass analyser, which separates the ions according to their m/z ratio by applying electromagnetic fields; a detector,

which records the ion counts at each m/z value, providing data for calculating the abundances of each ion present; and a computer interface which control the instrument, acquire and manipulate data, and compare spectra to reference libraries [79].

There are several variations of some of these modules, mainly regarding the ion source and the mass analyser, which give way to different MS techniques.
Among the several ionization methods, soft ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are commonly employed to ionize peptides or proteins in MS.

Regarding mass analysers, which are utilized for separating ionized analytes based on their m/z ratios, there are several examples, including quadrupole (Q), ion trap (quadrupole ion trap, linear ion trap), time-of-flight (TOF), and Fourier-transform ion cyclotron resonance (FTICR) mass analysers. Additionally, it is also possible to combine different mass analysers in series, with examples such as Q-Q-Q, Q-TOF, and LTQ-Orbitrap, which enhance sensitivity and detection capabilities [68], [79].

Finally, it is also possible to combine two or more mass analysers in series, with a collision cell to promote ion fragmentation between them, what is known as tandem mass spectrometry (MS/MS). This approach allows additional ‘purification’ by selecting specific precursor molecular ions from a mixture after first MS analysis, for further analysis and characterization by the subsequent MS instrument. This is often required

when complex mixtures of tryptic peptides are separated by LC and two or more species co-elute [80].

In proteomics, two primary MS analysis methodologies can be employed: top-down and bottom-up workflows, the latter also known as peptide-based proteomics. In the bottom-up approach, proteins undergo enzymatic digestion by trypsin, followed by separation of resulting peptides through specialized columns, culminating in their MS analysis (Fig. 1.15) [79].

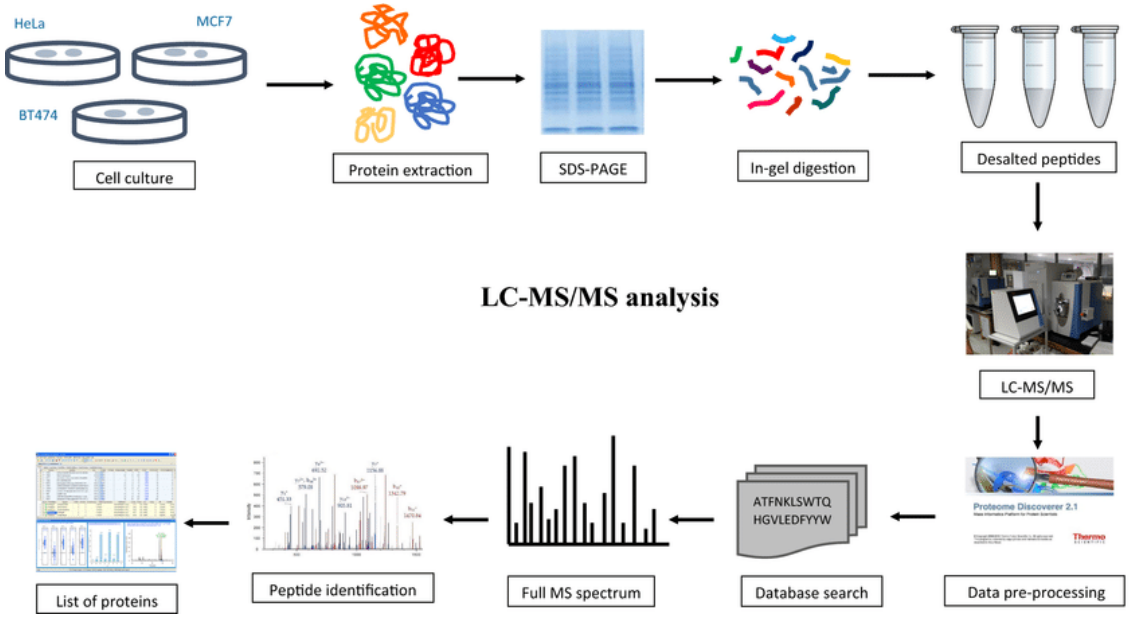


Fig. 1.15: Schematic diagram showing a typical workflow in bottom-up MS including protein preparation, identification, and data analysis. An LC–MS/MS approach is shown. Reproduced from [79].

This method can be further categorized based on fractionation techniques. One strategy involves using two-dimensional gel electrophoresis (2-DE) to isolate proteins from gels,

which are subsequently digested into peptides for MS identification. Alternatively, the "shotgun" proteomics approach involves direct protein digestion without prior fractionation, with liquid chromatography (LC) employed to separate the peptides prior to MS analysis. This method, known as LC-MS, has become a common and indispensable tool for proteomic investigation (Fig. 1.16) [81].

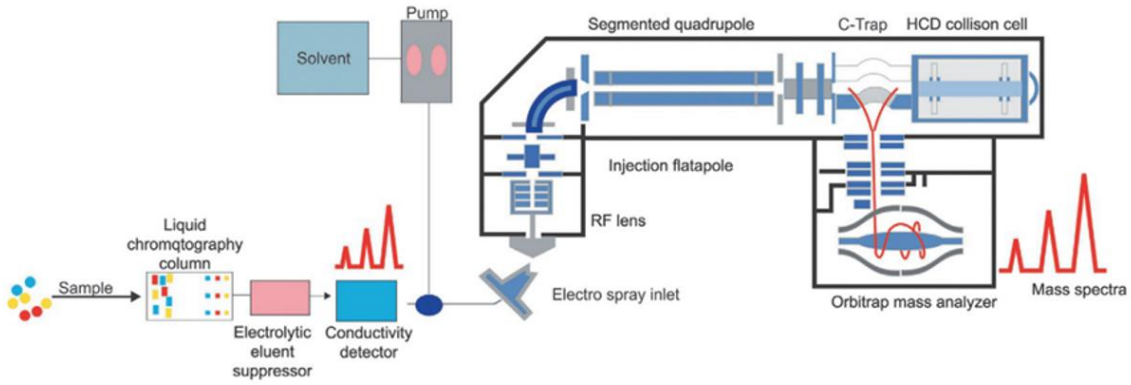


Fig. 1.16: LC-MS instrument. An electrospray ion source and a quadrupole mass analyser (Orbitrap) are shown. Reproduced from [68].

Conversely, top-down proteomics involves the direct analysis of intact proteins or polypeptides using MS. Techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) MS are commonly used to determine the molecular masses of proteins, with top-down proteomics capable of identifying proteins exceeding 200 kDa in mass [68], [79].

Each approach presents distinct advantages and limitations. While bottom-up proteomics may suffer from low sequence coverage due to the fragmented nature of recovered peptides, top-down proteomics retains the integrity of protein characteristics, allowing for the identification of various modifications and correlations,

including those resulting from the exclusion of protein digestion over time. However, top-down proteomics faces challenges regarding the solubility of proteins [70], [72].

On top of the two main categories of proteome analysis by MS, two main approaches can be utilised for protein identification: the database search method and the de novo sequencing method. The database search method is the most common one and involves comparing experimental data to theoretical data generated from a protein sequence database. Generally, the experimental spectra of a POI is compared to its theoretical spectra generated in silico using a database, finding the closest match. Several methods like SEQUEST, X!Tandem, OMSSA, and MASCOT are examples of database methods, and their differences lie in the type of scoring function utilized to rank-order the most probable peptide matches, and the type of sequence database in which the search is conducted. These database types include protein sequence, genomic, or expressed sequence tag [81]. Despite being very robust, the main limitation of this method is that the protein needs to be already known and the database used needs to contain an entry for the POI. Furthermore, erroneous matches are also common.

On the opposite end, the de novo sequencing approach identifies proteins by analysing the sequence information directly from spectrum peaks of peptide fragment ions, without relying on any pre-existing protein database, and aim at determining the longest peptide sequence that best matches the experimental spectrum [79]. This approach can be divided into two main categories. In the first category, exemplified by Sherenga and Lutefisk algorithms, the problem is approached through graph theory,

using algorithms to find the longest path in a network, thus achieving identification. In the second category, represented by PepNovo, probability models are employed to infer protein sequences from the spectrum peaks [79]. Despite being crucial for discovering new proteins and detecting amino acid mutations, it remains challenging due to inherent deficiencies in tandem mass spectra. Even if the optimal path is found, it may not always result in accurate peptide sequences because peptide fragment ions are often underrepresented, and many intense peaks in the spectra may arise from various interferences [79].

Overall, MS offers comprehensive insights into protein expression, localization, interaction, domain structure identification, modification, and activity analysis. Regarding SCP, with the aid of sophisticated automated sample preparation techniques and highly sensitive instruments capable of label-free or multiplexed data collection, it is possible to consistently identify and quantify between 1,000 to 1,500 proteins per cell. For example, techniques such as microfluidic magnetic bead-based cell isolation coupled with ultrasonication, as well as LC-MS, have enabled proteomic analysis of individual cells. One method that incorporates these techniques is Single-Cell ProtEomics by Mass Spectrometry (SCoPE-MS), which minimizes protein loss during sampling and enable peptide identification and quantification from single-cell samples through tandem mass tagging (TMT) followed by mass spectrometric analysis, enabling the identification of distinct human cancer cell types based on their proteome [71], [82].

Despite these achievements, these numbers are still relatively modest compared to the vast array of unique proteins and proteoforms present within a cell. The current status of MS-based single-cell proteomics mirrors the early stages of next-generation sequencing, highlighting the potential for further advancements and discoveries in the field [72].

1.6.4 Edman degradation

Initially described in 1967 by Pehr Edman, Edman degradation is method for sequencing amino acids within a peptide whereby individual N-terminal amino acids are sequentially cleaved from a peptide and analysed in consecutive steps [83].

Regarding the chemistry involved, the reagent phenylisothiocyanate (PITC) couples with the terminal alpha amino group of a peptide or protein to form a phenylthiocarbamyl (PTC) adduct. Under anhydrous acidic conditions the N-terminal amino acid residue is then selectively cleaved from the peptide chain as a heterocyclic derivative (an anilino-thiazolinone) through the attack of the sulfur group of the PTC adduct on the carbonyl component of the first peptide bond [83]. Subsequently, the cleaved amino acid derivative is separated from the residual peptide by extraction with an organic solvent and then converted to a more stable isomer (a phenylthiohydantoin) prior to identification by one of several possible procedures, usually chromatography or electrophoresis (Fig. 1.17).

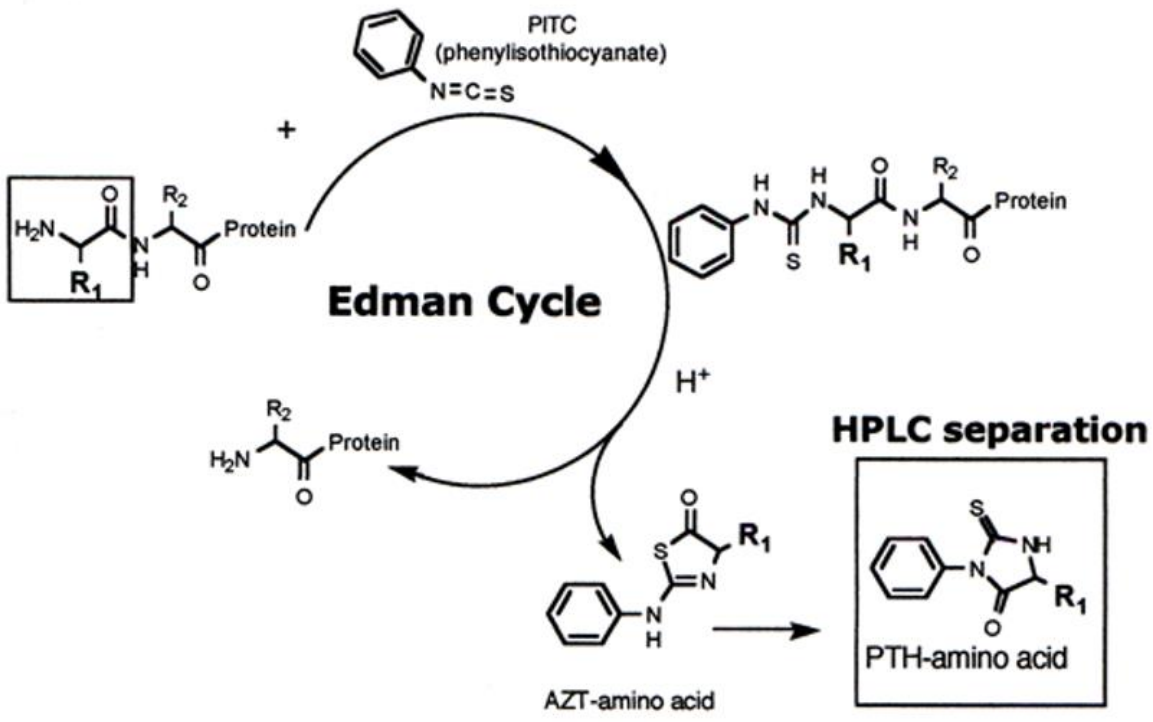


Fig. 1.17: Schematics of a single Edman degradation cycle. PITC reagent coupling and subsequent cleavage of the N-terminal amino acid is shown. Reproduced from [84].

The shortened peptide chain can then be subjected to further cycles of coupling and cleavage, with identification of each successively removed amino acid, thus establishing the amino-terminal sequence.

Although under appropriate conditions the coupling and cleavage reactions take place with close to 100% efficiency and a minimum of side reactions, realistically peptides longer than 50-60 residues cannot be sequenced (in practice around 30 amino acids). Therefore, to sequence larger proteins, these must be previously digested into shorter peptides followed by isolation by chromatography. In order to stablish the full-length protein sequence, a method of overlapping peptide digestion is used, whereby different proteases and / or chemical methods are employed in separate Edman sequencing

reactions to obtain different sets of amino acids of the same protein, which contain overlapping regions and can be used for assembly [85].

Edman degradation offers some key advantages over traditional MS in that it allows for the de novo sequencing of proteins with the lack of a protein database, and it can be used to analyse the N-terminal end of unknown proteins with high accuracy. Furthermore, it can very accurately distinguish amino acid pairs with similar molecular weights such as isoleucine / leucine and lysine / glutamic acid and can locate the position of mutation sites. This makes it still the gold standard technique for the analysis of the N-terminal sequence of proteins [85]. However, there are some important limitations of this method. As discussed above, a key disadvantage is the peptide length limitation, which makes this method not ideal for very long peptides or proteins due to decrease in cleavage efficiency and incomplete sequencing. Another limitation is that in order to achieve complete cleavage, repetitive cycles may be required for each amino acid, which makes the method very time consuming and labor-intensive (about 45 minutes per amino acid). Additionally, the N-terminus will not react with PITC if it is chemically modified, which is not trivial since most protein N-termini are found modified in cells, for example up to 80-90% of protein N-termini are found acetylated in mammalian cells. Furthermore, the cleavage reaction will not proceed if a non-α-amino acid is encountered, such as isoaspartic acid and proline, and some PTMs may also interfere with the reactivity of PITC [85].

Despite several efforts at automation, the main reason why Edman sequencing has not become a gold standard in proteome analysis is its low throughput and potential for multiplexing together its high cost and time consumption. However, it still constitutes a complementary technique to MS [85].

1.7 Approaches for single-molecule nanopore protein sequencing

Of all biomolecules, polysaccharides, nucleic acids, and proteins display the highest structural complexity, due to their polymer-like structure, made of sequential repetitions of the same set of chemical units. This makes nanopore decoding of each of these units, and hence identification, challenging if the whole molecule is sensed at once by the nanopore sensor. Nucleic acids and polysaccharide chains display no tertiary structure, and therefore it does not seem technically sound trying to study them as unique entities, with most efforts having been directed towards the analysis of their building blocks one by one. The case is different for proteins, as generally they arrange into well-defined three-dimensional arrangements. While some instances have been presented in previous sections, with successful single amino acid resolution among mixtures of different fully folded protein substrates. Attempts at identifying fully folded proteins using nanopore sensors remains an area of interest and development, comprehensive analysis and identification of polypeptides does only seem realistic if they are scanned in a sequential manner as they are driven through the nanopore.

Efforts in this direction have been fruitful regarding nucleic acids. One of the most attractive capabilities of single-molecule nanopore sequencing, and in high contrast

with other sequencing-by-synthesis-based techniques, like Illumina, is the possibility of long reads, the ability to directly sequence epigenetic marks and RNA, and the generation of real-time data from inexpensive, portable devices. Encouraged by these advances, single-cell analysis seems to be now next in the line to push the limits of nanopore technology in pursuit of protein sequencing.

Proteins are essential building blocks of life involved in nearly every biological process. The protein content of a cell provides key information for the understanding of biological processes and disease. The human genome contains an estimated 20,000 – 25,000 protein coding genes, if accounting for mRNA splicing, the number of different proteins elevates to 80,000 – 400,000, and including all possible PTMs, the number rises to >1,000,000 proteoforms. Detecting and identifying such a vast number of different protein species requires exquisite sensing capabilities [86].

As mentioned in the previous section, the current focus is on single-cell proteomics, which would be invaluable in understanding disease mechanisms, signalling pathways and diagnostics. However, it is especially challenging due to the high dynamic range in proteoform copy numbers (ranging from 1 to 10^8 copies per cell). The only methods currently available for protein sequencing are Edman degradation, mass spectrometry, or their combination [16]. The current gold standard in protein characterization, mass spectrometry (MS)-based proteomics, displays major limitations that make the technique unfit for single-cell analysis. It requires high-cost equipment and displays limited sensitivity and coverage of particularly low abundant proteoforms (detection

limit is in the picomolar range), although that could be partly overcome by enrichment using tailored antibodies for rarer proteins. Additionally, it is unable to generate full-length reads, since it requires fragmentation of protein samples into short peptides. Because of this, it struggles with identification of the location of specific modifications within individual peptides, i.e., post translational modifications (PTMs), since it only measures the mass of each peptide.

Nanopore sequencing has the potential to overcome these hurdles, since it benefits from reaching the absolute limit of detection, as theoretically single molecules can be counted, and can read full-length sequences. Nevertheless, major obstacles are still being faced that make nanopore protein sequencing still a far-off goal. There are 20 proteinic amino acids, compared to 4 nucleic acid bases, which quintuplicates the number of signals that need to be decoded. This is without taking into account post-translational modifications, that can occur in the other of hundreds of different chemical modifications, which further expands the spectrum of protein signals. However, it has been computationally estimated that by reading as little as 3 key residues, 97.5% of the whole proteome could be identified if split into 50 amino acid fragments. An additional problem is the high heterogeneity of amino acid side chains, which implies that charge is not constant throughout a protein sequence, which makes control on the direction of transport in an electric field more challenging.

As mentioned earlier, initial attempts at characterizing proteins involved driving completely folded polypeptides into large pores that could contain them in their lumen

for an extended duration of time. This enabled discrimination of individual amino acid substitutions, as well as other modifications, such as PTM mimics in selected substrates. Moreover, this technique also proved useful for the monitoring of enzymatic reactions and binding kinetics.

However sensitive this method might be, it displays two important downsides that do not make it a candidate for protein sequencing: if the target amino acid or modification happens to be buried within the hydrophobic core of the protein, sensing is not possible, and secondly, because polypeptides are not read in a sequential manner, positional information on modifications is not available. This approach might at best evolve into a future fingerprinting method, where experimental nanopore signals would be matched to an existing database of sequence / signal correlations.

To better understand the idiosyncrasies of protein sequencing, it is convenient to start by diving into the basic macromolecular mechanisms of protein translocation that would enable sequence readouts. One of the intrinsic properties of a chain in an aqueous solution is its ability to adopt a large number of conformations N. As a result, the Helmholtz free energy F of the chain is given by

$$F = E - TS = E - k_B T \ln N$$

where E is the energy of interactions among the repeat units of the chain and the solvent molecules, S is the entropy of the chain, and $k_B T$ is the Boltzmann constant

times the absolute temperature. E is determined by the intrachain interactions of the amino acid residues in the protein, which vary depending on the protein sequence and the solvent environment [87]. Additionally, changes in polymer conformations lead to reorganization of solvent molecules, contributing to the entropy.

Upon encountering a narrow restriction, such as a pore, it has been observed that the primary factor governing polymer transport is the decrease in the chain's conformational entropy [34]. As the available space is constrained, the number of conformations accessible to the chain is limited, thereby increasing the chain's free energy. This phenomenon forms the basis of the entropic barrier model for polymer transport in restricted environments.

The protein translocation phenomenon can be understood from the perspective of the Helmholtz free energy. To drive the polymer translocation process, there must be favorable energetic contributions that outweigh the entropic barrier. For example, electrostatic interactions between charged polymers and the nanopore walls, as well as applied electric fields, can provide driving forces for translocation. These interactions can lead to a decrease in Helmholtz free energy $\Delta F<0$.

Considering the unique geometry of αHL, the translocation model gains further depth by accounting for its two distinct compartments. In this model, both the pore geometry and its charge distribution play key roles in shaping the dynamics of polymer translocation. For a protein to pass through the pore from the cis-side to the trans-side,

it must first enter the relatively spacious vestibule with ~4.6 nm diameter. It then passes through the 1.5 nm constriction, which restricts the protein molecule to pass through only linearly, to enter the ~5 nm long and ~2 nm wide β-barrel. Considering the presence of both the vestibule and the β-barrel compartments, along with the chemical variations along the internal surface of the pore, three distinct contributing factors to the translocation process can be found (Fig. 1.18) [88].

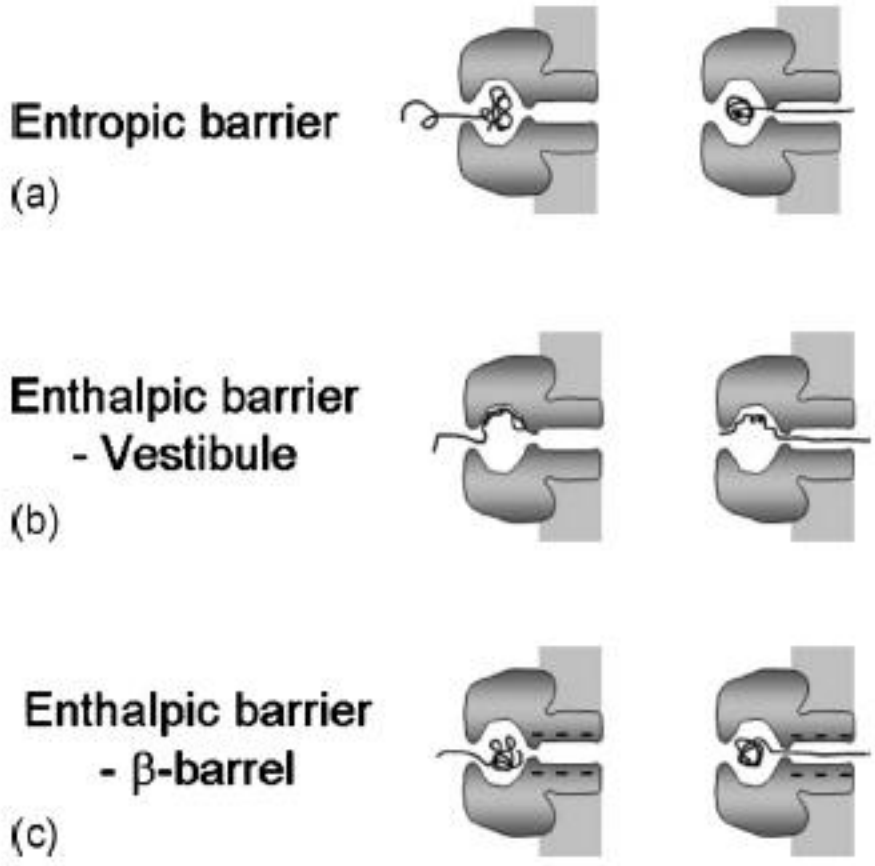


Fig. 1.18: Three components contribute to the free energy barrier for translocation through αHL. a, Chain confinement inside the vestibule leads to entropic barrier. b, Adsorption of polymer at the interior surface of the vestibule leads to enthalpic barrier. c, Electrostatic interaction between the polymer and the β-barrel results in an enthalpic barrier. Reproduced from [88].

Firstly, the protein chain tends to favorably partition into the vestibule over the β-barrel, allowing for a greater variety of conformations within the spacious vestibule. Consequently, the vestibule serves as an entropic trap for the polymer, facilitating the adoption of numerous favorable conformations. However, for translocation to occur, one end of the chain must be positioned at the entrance of the β-barrel, leading to a reduction in the chain's conformational entropy. Thus, the vestibule effectively acts as an entropic trap, resulting in an entropic contribution to the free energy barrier for translocation [88].

When molecules enter an entropic trap, they often do so because the trap exploits the tendency of them to increase their entropy by exploring different positions and orientations in space. However, once they enter the trap and become confined, their entropy decreases because their mobility and freedom of movement are restricted. So, while the process of molecules entering the trap may be driven by the desire to increase entropy, once trapped, their entropy decreases due to the confinement.

The second contributing factor is the protein-pore interaction inside the vestibule region. The attraction between the vestibule surface and the chain can hold the chain from entering the β -barrel. The chain has to overcome this attraction, and hence this results in an enthalpic contribution to the free energy barrier for translocation. Lastly, the third factor arises from the protein-pore interaction inside the β-barrel region. This also results in an enthalpic contribution to the free energy barrier. All of the above three factors are expected to contribute to the free energy barrier for translocation. To date,

the relative weights of the entropic and enthalpic contributions to the free energy barrier have not been established for the translocation of polymer molecules.

Nevertheless, proteins may not behave as other polymer-like molecules, as they typically have a folded structure with specific secondary and tertiary interactions that stabilize their conformation. When a protein threads through a nanopore, it may need to unfold or partially unfold to fit through the confined space. This unfolding process can involve breaking intra-molecular interactions, which may lead to an increase in entropy as the protein adopts a more extended, less ordered conformation. This increase in entropy could contribute favourably to the translocation process, reducing the overall Helmholtz free energy barrier.

To enable the translocation of a polymer chain through a pore, it must first be brought to the entrance of the pore, accomplished through a combination of diffusion and electrophoretic drift. At this point, a sufficient force is required in the vicinity of the pore to capture the polymer. The strong electric field generated by the externally applied voltage difference at the entrance of the pore, as well as the interaction between charges on the pore and on the translocating protein, can provide this force. Additionally, the EOF arising from internal charge within the pore can generate strong water flows which help in this mechanism. Consequently, there is a capture radius, denoted as r, which can vary over multiple orders of magnitude depending on the system. After capture, the chain must then unravel its various conformations to position one end at the mouth of the pore, facilitating single-file translocation. This final stage is

known as the threading of the polymer, which encompasses the actual translocation process.

In practice, the capture barrier has previously been solved by adding highly negatively charged tags to polypeptide C- and/or N-termini, such as short (20-30-mer) single stranded oligonucleotides [23], and charged residue-rich polypeptide regions, that lead protein translocation under the ionic current [23], [89], [90]. Such tags have also proven useful for directionality control (i.e., which end treads the pore first), but they do not allow for successful control of the translocation rate. However, threading of one protein end in the pore entry constitutes only the first bottleneck in protein sequencing.

Following this, the velocity of transport must be tightly controlled in a processive manner to permit single amino acid identification. In a free running system (i.e., non-enzymatic protein transport across the nanopore), there is an evident trade-off between capture rate and translocation speed, as increasing the forces (i.e., EFP, EOF) generated to capture an analyte of interest do also accelerate the rate of translocation. Control in the rate of transport that would in theory allow for amino acid resolution (by extrapolation from nucleic acid base callers) has been approached with the use of protein unfoldases, and DNA motors.

These current leading efforts towards protein sequencing that have the potential to overcome the aforementioned hurdles, can currently divided into two main groups: enzyme-driven and enzyme-free based approaches.

The former group comprises the use of molecular motors which substrates are either polypeptides (ClpX unfoldase) or DNA/RNA molecules (helicases, and polymerases). In one example, promising results were shown with a multidomain protein construct that was successfully captured and translocated across a nanopore via tagging with ClpX unfoldase recognition peptide sequence (ssRNA-tag) and a charged protein terminal segment that allowed threading within the pore [89], [90]. Theoretically, full-length protein substrates could be sequenced with this method, but serious efforts towards motor engineering should be conducted to compensate the lack of processivity, slipping and bias towards specific polypeptide sequences. These challenges currently limit the approach to being mostly a fingerprinting method, at best.

Other impressive examples of the use of a protein motor for controlling translocation was showcased by the Maglia group. In an early work, an αHL nanopore was engineered by adding loop sequences from GroES protein, which after membrane reconstitution, could bind a single-ring version of the GroEL chaperonin protein [91]. It was shown that this complex could assist protein folding as efficiently as native GroES. Folding was monitored by the induced changes in current signal from the protein analytes captured within the GroEL cavity.

Building on this preceding work, a heavily engineered nanopore, called proteasome-nanopore, was constructed by combination of several other proteins. The assembly was carried out in two steps. Firstly, an heptameric proteasome activator 28α (REG), which

regulates the function of the proteasome by docking on the heptameric 20S core proteasome particle, was modified by replacing its disordered region with the β-barrel transmembrane region of the heptameric anthrax protective antigen pore (*Bacillus anthracis*), using the flexible SSG motif flanking the barrel to connect to the surrounding REG. In a second step, the 20S proteasome from thermoplasma, which is made by four stacked rings of 14-α and 14- β subunits (α β β α) were fused to the previously assembled "REG-nanopore" by co-expression of the two genetically fused complexes. The internal α subunit served for docking to proteasome activator REG and contain regulatory elements. The fully assembled construct was then inserted into a lipid bilayer and an unfoldase was added along with a protein analyte. The unfoldase, VATΔN, a truncated variant of wild-type VAT Valosin-containing protein-like ATPase of *Thermoplasma acidophilum* with higher unfolding activity, was shown to bind to the outermost α-subunit of the 20S proteasome in a reversible manner, enabling ATP-driven capture, unfolding and proteolytic digestion of a substrate protein.

The previous design is of especial remark because it also allowed two approaches for single-molecule protein sequencing, a "chop-and-drop" mode, in which unfolded proteins are degraded by the proteasome and the resulting fragments delivered to the pore sensor, and a "thread-and-read" mode, where substrates are pulled through intact and its sequenced read as they translocate the pore. These two systems were initially pursued in parallel towards nucleic acid sequencing, until successful efforts on helicase and polymerase engineering and computational base calling would make the latter a better candidate for commercialization. Similarly, it highlights the possibility of

designing modular nanopore sensors by combining multiple molecular machines into a unique assembly.

In the latest proof-of-concept technique for protein sequencing, a helicase was employed to move an oligonucleotide labelled peptide across a nanopore as the motor moved along the DNA strand. The peptide was ratcheted through a *Mycobacterium smegmatis porin A* (MspA) pore at nucleic acid steps (~3 amino acids), and single amino acid resolution was demonstrated. In contrast with previous approaches where the protein analyte is pushed inside the pore, here a terminal DNA handle was used to drive translocation, and a Hel308 helicase unit pulled the peptide conjugate back to the cis side outside of the pore. The major flaw of this method is that, despite having the potential to solve the amino acid resolution issue, peptide length is restricted by 10-20s of amino acids, determined by the length of the nanopore sensing region, which would require considerable fragmentation of the original proteome, hindering recovery of original sequences. Furthermore, true amino acid resolution is far from a reality, and currently the use of constant peptide backgrounds is required for specific amino acid identification (i.e., tailored background to amplify amino acid signals), e.g., Oxford Nanopore Technologies showed double DRLY amino acids sets could be discriminated in a poly E constant background, and a sliding set of one of these amino acids along a constant background similarly generated different current fingerprints.

Another downside is that this approach will not permit de novo sequencing, since the protein variants in the sample would require fragmentation and incorporation into DNA

handles. Ultimately, for this method to be applicable to real peptide sequencing, it will most likely require the use of vast random combinatorial peptide libraries to train models, that can be later used to ID electrical fingerprints generated with real protein samples.

In practice, given the challenges in achieving full single amino acid resolution when decoding protein nanopore signals, it is expected that nanopore protein fingerprinting will be employed as a preliminary step before de novo sequencing. This approach will involve identifying a small subset of amino acid types or peptide motifs along the protein strand, which should be sufficient for unique identification based on a reference proteome.

Finally, the second main group of efforts towards nanopore proteomics relies on tuning the electroosmotic (EOF) and electrophoretic (EPF) forces, or a combination of both, to control polypeptide transport across nanopores. Some examples have been presented earlier in this chapter, where charged handles, in the form of DNA strands or charged peptide sequences needed to be fused to the POI for translocation. This strategy has proven useful for the analysis of protein folding and unfolding kinetics, characterization of protein stability, and detection of phosphorylated protein substrates, yet it remains largely unexploited.

Inspired by these encouraging results, it was decided to take on this approach for developing a novel method for protein sensing which required minimal sample

processing or enzymes to control movement across the pore, a set up that would facilitate sample handling and decrease costs if it were ever to be implemented onto a real sequencing workflow. Recent work carried out on FraC and ClyA nanopores to induce EOF proved this force to be strong enough to counteract strong EPFs induced by the charge of folded protein and peptide analytes. However, the effect of this force on the threading, unfolding and translocation of extended proteins remained unexplored, a mechanism that is necessary for reading the sequence. Hence, the main goals of this thesis where two; first was to obtain definitive evidence that long polypeptides could be captured and translocated across nanopores in the absence of significant EPFs, driven uniquely by the EOF. Secondly, it was to prove that PMTs could be detected and discriminated at specific locations along those polypeptides.

This work constitutes the first direct proof-of-concept demonstration of a long polypeptide translocating a pore in a tag-less, enzyme-less fashion, driven only by a combination of EOF and EPF. This is achieved through genetic engineering of charges in nanopores that induce EOF under an applied transmembrane voltage. The method described takes advantage of the naturally present folded state of proteins, to stall protein translocation in between spontaneously unfolded domains, causing a segment of the protein to remain within the pore sensing region for a sufficiently long duration of time to enable detection of PTMs. In the following chapters, the technical aspects of the work will be described in detail.

1.8 Motivation

As discussed in the previous sections, the traditional approach to proteomics involves analysing bulk samples, where thousands of cells are analysed at once, resulting in an average protein expression profile. This approach fails to capture the unique protein expression patterns of individual cells, which can be critical in understanding the underlying mechanisms of disease.

SCP emerges as solution to this, where MS is currently the gold standard thanks to several improvements, particularly regarding improved methods of sample extraction that minimise sample loss, and increased sensitivity. However, for MS the analysis of the proteome within a single cell remains challenging due to the high dynamic range in protein copy numbers (1 to 10^8 per cell in mammalian cells) and the impossibility of amplifying them. Furthermore, its instruments are high cost and its operation require a high level of expertise [92].

To overcome these challenges, SCP could benefit from the development of single-molecule protein sensing such as single-molecule nanopore sensing, as they offer higher sensitivity, direct quantification of individual molecules (which could gain access to yet unobserved protein species within the "dark proteome"), better proteoform characterization, increased sequence coverage, and longer reads. Furthermore, single-molecule nanopore technology comes in simpler instruments, which demand less expertise, are more cost-effective, portable, and seamlessly integrate with single-cell studies, what would ultimately make proteomics more accessible to a wider scope of researchers and clinicians [92].

Nevertheless, unlike state-of-the-art MS, single-molecule protein analysis with nanopores is still far from achieving de novo sequencing capabilities, and recent advancements in the field point towards a likely initial fingerprinting method that could potentially further develop into a full de novo sequencing technique due to its exquisite sensing capabilities. The main hurdles that need to be overcome can be responded through two main efforts. First is the development of a strategy that ensures delivery of the protein analytes across the nanopore sensor in a way that enables sequence redouts, and second is the development of nanopore sensors with increasingly improved capture, resolution and signal-to-noise ratios.

Within the first group, several key challenges arise. i) capture of one of the protein analyte ends by the nanopore sensor, which is challenging due to the lack of uniform charge, ii) unfolding of the protein into a linear form, which involve overcoming an energy barrier that can be high for highly stable proteins and may require incubation with high concentration of denaturants and other types of sample pre-treatment approaches such as chemical modification of side chains, iii) translocation of the threaded polypeptide across the nanopore at a controlled movement rate that enables adequate interrogation of the sequence, which is also challenging due to the lack of uniform charge and the lack of well characterized protein motors.

Regarding nanopore developments, most efforts have been directed towards improving resolution for limited sets of peptide analytes, which may not be generalizable to the

wider and more variable proteome, which display a high number of amino acid combinations, and while having successfully demonstrated single amino acid discrimination within simple peptides, a model which correlates the electrical signal to an individual amino acid across the full-length of a peptide analyte has not yet been generate. This would likely require training models using a vast number of combinatorial peptide libraries which are currently not accessible. More recently, pore modifications which enable the generation of an EOF have been explored for enhancing the capture and translocation of protein and peptide analytes through the pore in a charge-independent manner, but at this stage, it still remains at the proof-of-concept level and no protein detection capabilities have been demonstrated [93].

Inspired by the success of enzyme-mediated nucleic acid sequencing, several methods have been reported in the literature which exploit the use of a molecular motor, such as ClpX unfoldase, to drive full-length proteins across the pore, either in a cis-to-trans (Fig. 1.18) [89], [90] or, more recently, a trans-to-cis configuration (also known as cis-based unfoldase approach) (Fig. 1.19) [94], where the unfoldase is added in the trans or the cis, respectively. The trans-to-cis configuration, where the protein analyte is initially captured by the pore, stalled thanks to a terminal blocker domain, and subsequently pulled out by the unfoldase, has shown promising results by demonstrating single amino acid resolution, and because the unfoldase is added to the cis-side, it is compatible with MinION flow cells from Oxford Nanopore Technologies, which do not permit access to the trans side. Nevertheless, single amino acid discrimination has been achieved within heavily engineered polypeptide “cassettes”, designed to contain a

relatively “silent” poly glycine-serine background, and requires complex protein analytes with a C-terminal blocker domain bearing an unfoldase recognition tag, and an N-terminal charged leader for capture. Furthermore, it suffers from a limited throughout, in the order of low tens of events per hour [94].

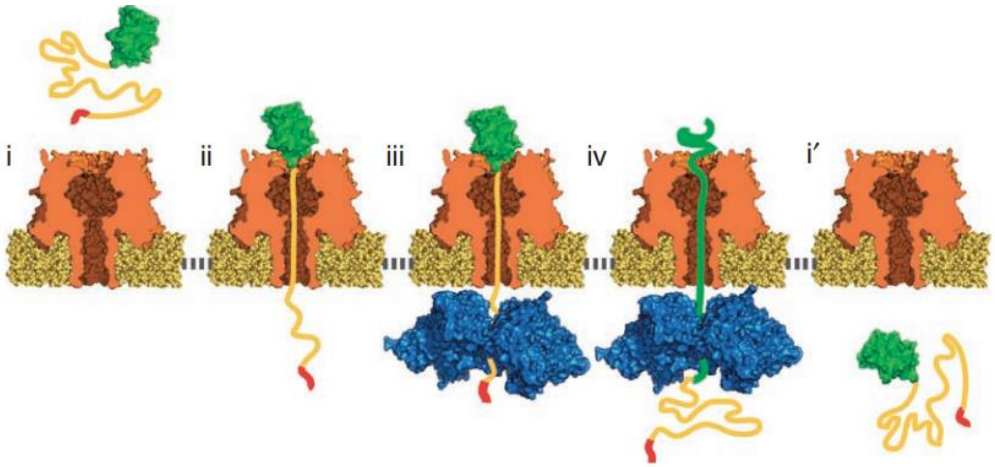


Fig. 1.18: Working model of ClpX-mediated translocation. A protein domain (green) displaying a highly charged C-terminal peptide leader (yellow) with a terminal ssrA recognition tag (red) in the cis-side (i), is first electrophoretically captured (ii). An hexameric ClpX domain recognises the ssrA tag and ratchets the protein termini (iii), and starts pulling from the protein resulting in unfolding of the folded domain (iv), and subsequent translocation (v). Reproduced from [90].

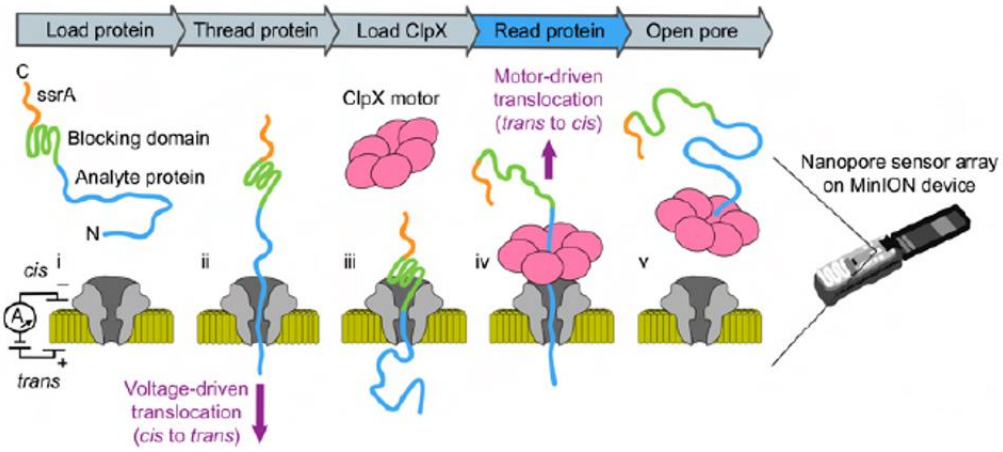


Fig. 1.19. Nanopore protein reading using an unfoldase in a trans-to-cis configuration. i-v numbers correspond the different translocation steps. Reproduced from [94].

DNA-processive motors have also been used to drive oligonucleotide-conjugated peptides through a nanopore, achieving stepwise resolution of amino acid translocation (Fig. 1.20) [95]. However, there are some major drawbacks, namely, it is highly limited by the analyte length, as it can only be used with short peptides that do not expand a longer length than the nanopore lumen and shows contribution from the DNA and the chemistry used for peptide labelling at the DNA-peptide interface, what makes identification of the amino acids closer to the DNA handles challenging.

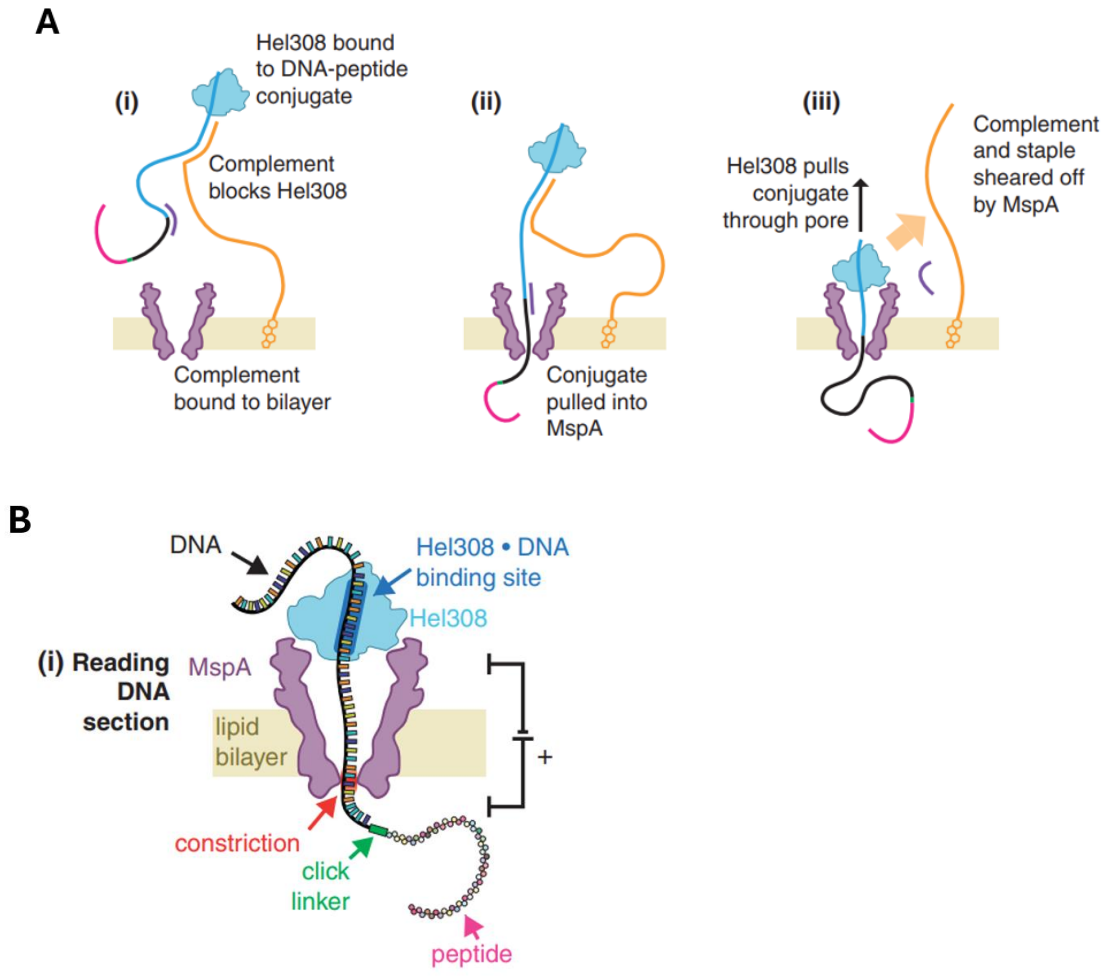


Fig. 1.20: Helicase-driven peptide reading with a nanopore. a) The DNA-peptide conjugate consists of a peptide (pink) attached via a click linker (green) to an ssDNA strand (black). This DNA-peptide conjugate is extended with a typical nanopore adaptor consisting of an extender that acts as a site for helicase loading (blue) and a complementary oligo with a 3' cholesterol modification (gold). The presence of a complementary oligo blocks the helicase, until it is pulled into the pore (ii), prompting the separation of the complementary strand (iii), thereby allowing the helicase to move along the DNA. b) As the helicase progresses through the DNA strand, it pulls it up through the pore enabling sequential readings of the DNA segment and the attached peptide. Reproduced from [95].

All the aforementioned methods require a high degree of sample processing, including chemical labelling of the protein termini, which is particularly challenging at the C-terminus due to the absence of methods that specifically target that carboxylic group. Furthermore, the N-terminus of proteins, for which more chemical labelling methods are available, are highly acetylated within mammalian cells, what requires methods for either chemical or enzymatic deacetylation. The former can be achieved through incubation with acids such as hydrochloric (HCl) and triofluoroacetic (TFA) acids, which are highly unspecific and can result in the hydrolysis of internal peptide binds within the protein. Furthermore, no enzyme has been characterized to date which can catalyse the removal of N-terminal acetyl groups.

Enzyme-free approaches that enable a controlled linear translocation of full-length proteins arises as a promising tool that could overcome all these issues. While no motor is available to control the translocation rate of the protein analyte across the nanopore, we envisioned it might be possible to harness the folded configuration of proteins to stall or "slow down" their translocation, accessing sequence information within certain

protein regions that could enable fingerprinting resolution. Furthermore, its main advantage lays within its simplicity, as no labelling is technically required.

In order to have a constant driving force for protein translocation, we decided to take advantage of EOF, and generated mutant pores with enhanced ion selectivity that enabled the generation of a directional water flux controlled by the orientation of the applied potential.

Because the feasibility of translocating full-length proteins through a nanopore, driven solely by EOF, had not yet been demonstrated at the time this work was designed, we decided to generate analytes with multiple repetitions of the same protein unit, what we decided to call concatemers, and used 109 amino acid-long thioredoxin as the repeating units, which had been thoroughly characterized previously in nanopores. We reasoned that the presence of multiple repetitive units would result into repetitive signal patterns upon complete translocation, and that number should match in both cases.

Lastly, to fully demonstrate a real applicability of our method, we decided to include PTMs (serine phosphorylation, glycosylation and glutathionylation) at different positions within the concatemers, and we demonstrated identification of all of them at some of those positions, despite being relatively close in the sequence.

2 Engineering electro-osmotic flow (EOF) in nanopores

2.1 Ion selectivity characterization

In an electrophysiology setup, when a voltage bias is applied and assuming that only a single type of salt is dissolved in both the trans and cis buffer solutions, an ideally conical-shaped pore with no exposed charges on its inner surface would allow ions and their corresponding counterions to pass through with equal ease in opposite directions. This is what we refer to as a non-ion-selective nanopore.

However, protein nanopores deviate significantly from this ideal scenario. The shape and chemical properties of the amino acid side chains that make up these pores typically introduce a bias toward the passage of either positively or negatively charged ions.

In solution, ions are surrounded by a specific number of water molecules (which depends on the type of ion). When there's a net movement of ions in one direction, it results in a directional flow of water molecules, which can create a dragging effect on analytes, regardless of their charge. This effect, induced by the applied electrical field, is known as electro-osmosis or electro-osmotic force (EOF).

The ability of a nanopore to selectively permit the passage of certain ions while blocking others is referred to as ion-selectivity, also known as permeability ratio. Conventionally,

pores that exhibit selectivity for positively charged ions are called cation-selective, whereas those favouring negatively charged ions are called anion-selective.

In order to estimate the EOF through a nanopore, first it is necessary to estimate the permeability ratio P+ /P− of the two most dominant contributors of the ionic current, K+ and Cl− ions in this case, across the nanopore. It is given by the Goldman-Hodgkin–Katz voltage equation [96], [97]:

$$\frac{P_+}{P_-} = \frac{[\mathrm{Cl}^-]_c - \exp(-eV_R/k_BT)[\mathrm{Cl}^-]_t}{\exp(-eV_R/k_BT)[\mathrm{K}^+]_c - [\mathrm{K}^+]_t},$$

where [K+] and [Cl-] are the concentrations of K+ and Cl− ions, respectively, on one side of the nanopore. The subscripts *c* and *t* denote the cis- and trans-sides, respectively. $V_R=V_t-V_c$ is the reversal voltage, which sets the ionic current passing through the nanopore zero. *e*, k_B, and *T* are the electron charge, the Boltzmann constant, and the temperature, respectively. The Goldman-Hodgkin–Katz voltage equation shows that an asymmetric salt condition, i.e., different salt concentrations on the cis- and trans-sides, is necessary for measuring the reversal voltages V_R.

To estimate the flux of the EOF, it can be assumed that each ion passing through the nanopore is coupled with N_w water molecules, which is around 10 according to an all-atom molecular dynamics simulations [98]. Therefore, the flux of the EOF, in number of water molecules per unit time, is given by the following equation:

$$J_w = N_w \frac{I}{e}\left(\frac{1 - P_+/P_-}{1 + P_+/P_-}\right),$$

where *I* is the ionic current under a given voltage, *e* is the elementary charge, and P_{K+} / P_{Cl-} the ion selectivity. This equation likely underestimates the water flux, as it does not take into account the movement of the electrical double layer.

Unless the analyte is a neutral molecule, there will also be an EFP acting on a charged particle in an electric field, which can be described using Coulomb's law. The equation for the electrophoretic force (F_{EP}) is:

$F_{EP} = q \cdot E$

Where F_{EP} is the electrophoretic force (in Newtons), q is the charge of the particle (in Coulombs), and E is the electric field strength (in volts per meter). However, for a polymer like molecule such as proteins, the bare line charge density must be calculated, which refers to the charge per unit length along an infinite chain [99]:

$\lambda_b = Q / L$

where *Q* is the total charge along the chain and *L* is the length of the chain. The concept of bare line charge density assumes that the line is infinitesimally thin and has no physical thickness. In reality, protein chains have finite dimensions and the charge distribution is not perfectly uniform along their length. However, for many theoretical calculations and simplifications, the concept of bare line charge density provides a

useful approximation. The bare electrostatic force on the protein backbone can be represented by:

$F_{bare} = \lambda_{bare} \Delta V$

which is opposed by a drag force F_{drag} caused by the electrophoretic motion of its counterions in the opposite direction. F_{drag} is therefore an intrinsic component of the translocation process that is also ultimately traceable to electrical forces. The resulting net force is the electrophoretic force driving the translocation, given by $F_{elec} = F_{bare} - F_{drag}$.

As mentioned above, one way to quantify the level of ion-selectivity in a nanopore is by experimentally determining the reversal potential (Vr), which is the voltage bias required to achieve zero current when using buffer solutions with asymmetric ionic strength concentrations on opposite sides of the bilayer (referred to as asymmetric salt conditions). In the absence of voltage bias and under such asymmetry, a non-selective pore will permit a greater net passage of ions from the side with higher ionic strength to the opposite side, driven by the concentration gradient. However, the net flow of ion charges (as measured at the trans electrode) will be zero because an equal number of positive and negative ions migrate in that direction, assuming similar migration rates for ion-counterion pairs.

This scenario is different for an ion-selective pore. For instance, in the case of a cation-selective nanopore, the passage of cations from the higher ionic strength side of the

bilayer to the lower side would be facilitated, while the movement of anions would be impeded. This would result in a net passage of cations toward the anode.

2.2. Approaches for tuning ion selectivity

The type and magnitude of ion-selectivity, and ultimately, the strength and direction of the EOF of a nanopore, depends ultimately on the surface accumulation of ionic charges, which can be modulated by several physical and chemical properties: pore geometry (size and shape), surface amino acid charge, surface chemical modifications, pH, bilayer composition, ionic strength and use of chaotropic agents.

Surface charge accumulation plays a pivotal role in ion selectivity. The presence of local surface charges within the pore induces the accumulation or depletion of ions based on their charge. This accumulation is most pronounced near the pore's surface and diminishes as you move away from the surface, following a characteristic length known as the Debye length. Consequently, this charge distribution can set the fluid in motion when subjected to an external electric field parallel to the pore walls [100].

The specific chemical properties of the amino acids oriented toward the inner pore lumen can also affect the overall electrostatic charge on the pore's surface. Recent examples in the literature highlight instances where mutations of these amino acids have significantly increased the ion-selectivity of initially non-selective nanopores, leading to a strong EOF capable of capturing and translocating protein analytes for analysis [45], [46], [93], [101], [102].

Furthermore, the surface of the nanopore can be chemically modified to enhance ion-selectivity. As explained in a previous chapter, different surface functionalizations can create specific interactions with certain ions, promoting their passage while impeding the passage of others. Typically, EOF is controlled by altering the narrowest section of the pore, known as the constriction. This region is responsible for the majority of the pore's electrical resistance and, consequently, plays a crucial role in the pore's ability to distinguish between different molecules. However, modifying this region to control EOF may potentially compromise the nanopore's sensing properties [100], [103].

Ideally, the adjustment of the EOF should focus on modifying the nanopore as far away from the constriction as possible to minimize interference with the sensing region. This strategy allows for the optimization of the constriction to enhance molecule distinguishability, while other regions can be fine-tuned to enhance EOF.

One effective method for achieving this is by altering the shape of the nanopore. Recent computational simulations have indicated that pore geometry can exert a strong influence on the migration of ions by causing them to accumulate within specific hollow structures within the pore. To implement this approach successfully, it's necessary to introduce geometric asymmetry to the pore. This asymmetry can exist within the pore lumen or extend to regions far from it, as illustrated in Fig. 2.1.

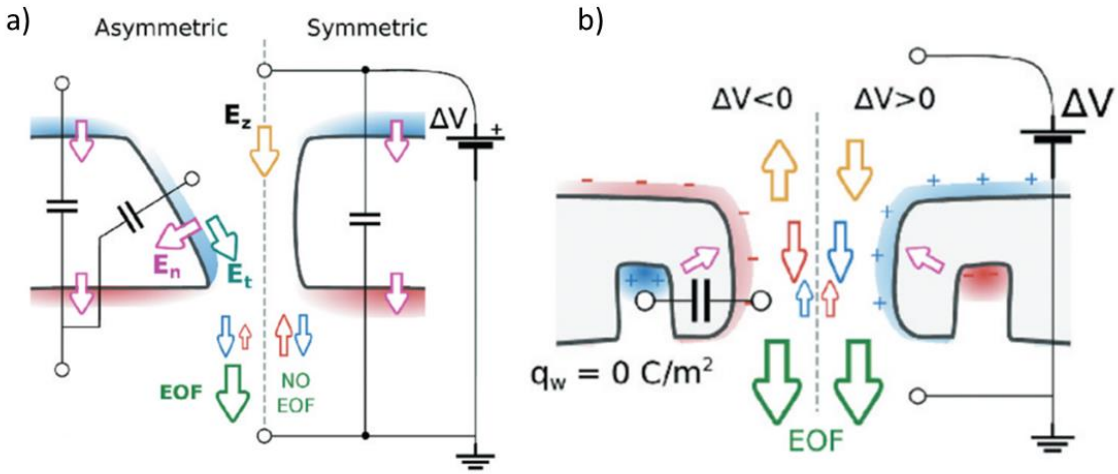


Fig. 2.1: Pore geometry can cause induced charge accumulation upon applying a voltage bias. The external electric field gives rise to induced double layers, for non-symmetric pores, this may result in charge accumulations, denoted as red or blue, for negative and positive, respectively. The applied voltage ΔV gives rise to induced Debye layers (IDLs) at the solid–liquid interfaces, the polarity of which depends on the voltage sign. The generated field acts on the net charge putting the fluid in motion. **a,**

Charge accumulation and fluxes for two different geometries, a conically shaped pore (left) and a cylindrical one (right). The normal component of the electric field at the solid–liquid interface (pink E_n), generated by an external transmembrane voltage drop ΔV, drives the formation of the Induced Debye Layer. The tangential component of the electric field at the wall, E_t, moves the accumulated charges.

b, Unidirectional EOF can be achieved on an uncharged cylindrical pore exploiting symmetry breaking due to a lateral cavity surrounding the nanopore. A unidirectional EOF is achieved regardless of the orientation of the electric field due to the simultaneous inversion of ionic selectivity. Figure reproduced from [100], [104].

Other experimental parameters can also be tuned to increase selectivity. For example, buffer pH can change the protonation state of amino acids located within the pore lumen, thereby generating surface charges. Recently, the use of chaotropes such as GdnHCl has also been shown to promote EOF in αHL [93]. In a series of computational simulations using different combinations of an electrolyte solution of GdnHCl and KCl, guanidinium cations were observed to remain bound to the nanopore surface for considerable (>10 ns) intervals of time. Bound guanidinium caused an increase in positive surface charge, which almost excluded K^+ cations from the nanopore lumen. The surface charge was compensated by Cl^- anions, which carried most of the current and ultimately caused the electroosmotic effect.

Lastly, although not initiated by an applied electrical potential and therefore not strictly classified as an electroosmotic effect, a directional movement of water molecules through a nanopore can occur due to the presence of asymmetric salt concentrations on either side of the bilayer. In a bilayer that is perfectly impermeable to water, there would be a net flow of ions from the more concentrated side to the less concentrated side, following the concentration gradient. This ion movement generates a water flux, which has the potential to exert a dragging force on analytes.

However, it's worth noting that the artificial lipid bilayers commonly used in most electrophysiology setups are composed of DPhPC, which is permeable to water molecules [105], [106]. Consequently, these bilayers allow an osmotic flow of water

towards the hypertonic solution, moving in the opposite direction compared to the ion concentration gradient.

As of the author's knowledge, these effects within the context of electrophysiology have not been harnessed in the literature for the purpose of transporting molecules, nor have they been extensively studied or simulated.

Additionally, the relationship between ionic strength and selectivity magnitude is non-linear. At low salt concentrations, selectivity tends to be low, while high salt concentrations can lead to the screening of internal amino acid charges, thereby reducing selectivity [45]. It's crucial to recognize that there is an inherent trade-off between the ionic current signal and ion-selectivity. Ionic current typically increases linearly with salt concentration, and higher currents enable a broader sampling spectrum of analyte features. However, this enhanced signal may come at the expense of pore selectivity. Therefore, it is essential to experimentally determine the optimal ionic strength that balances ion-selectivity with a current signal that suits the needs of the analyte.

2.3 Generation of ion selective αHL mutants

The focus of previous research in the field of biological nanopores has, for a long time, revolved around the electrophoretic-driven translocation of model peptides and proteins [47], [107], [108], [109]. As it was reviewed in previous chapters, even the capture of the protein analyte in the helicase-driven translocation approach, relied

initially on the capture of a highly charged peptide leader sequence [110]. However, recent studies have shed light on the relevance of electro-osmotic flow (EOF) on the transport mechanism of molecules across nanopores, which arises from the fixed charges present on the inner wall of the nanopore [45], [46], [93], [100], [101], [103]. In fact, the Bayley group had already reported enhanced capture and residence time of a neutral molecule β-Cyclodextrin within the constriction of αHL driven by EOF[96].

Within the field of proteins and peptides, an interesting example from prof. Luchian's group showed that at low pHs, the interplay between the electrophoretic driving force and the opposing EOF led to the entrapment of positively charged peptides within the nanopore [48]. This discovery highlighted the comparable strength of the EOF to the electrophoretic force, opening up possibilities for leveraging the EOF to translocate and stretch polypeptides for applications such as protein sequencing regardless of polypeptide charge.

As explained above, one straightforward way to increase the charge selectivity of nanopore is by tuning the protonation state of the amino acid side chains exposed to the inner lumen through changes in pH. Wt αHL was determined to be slightly anion-selective at neutral pH (10 mM Hepes, pH 7.2) in a 2M KCl solution ($P_{K+}/P_{Cl-} = 0.78$). Upon determining the pI of the residue segment most exposed to the lumen of the pore, this was determined to be 6.8 (Fig. 2.2). The following sequence was analysed: ARNSIDTKEYMSTLTYGFNGNVTGDDTGKIGGLIGANVSIGHTLKYVQPDFKTIL, which

contains an equal number of positively charged (Lys, Arg, in blue) and negatively charged (Asp, Glu, in red), and therefore a net charge of 0.

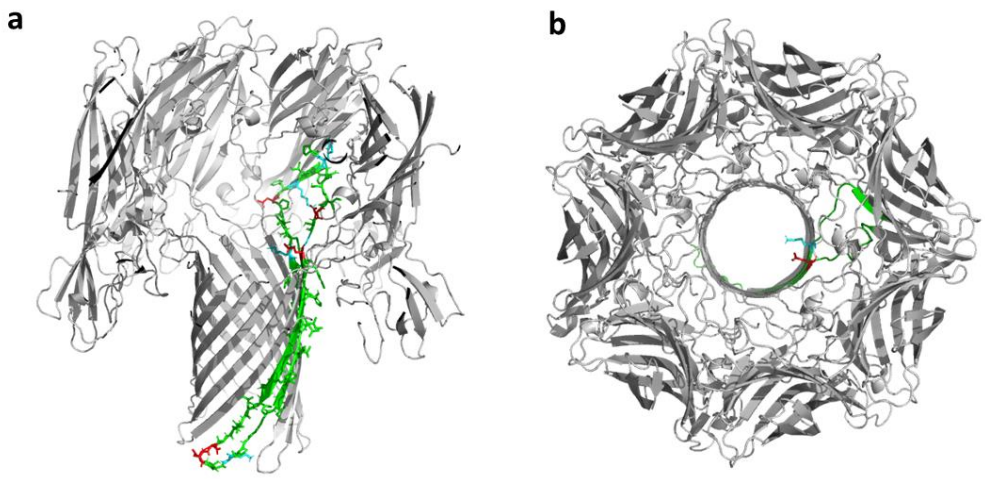


Fig. 2.2: Ribbon diagram of wt αHL, which shows almost no selectivity at neutral pHs. **a**, wt αHL front view: amino acids most exposed to the pore lumen in a single chain of the homoheptameric pore are coloured in green. Positively and negatively charged residues are coloured in blue and red, respectively. **b**, wt αHL top view: Residues E111, K147, closest to the constriction, are shown.

However, the pore ion-selectivity was reserved when increasing the pH to 11, under those conditions the pore became cation selective (P_{K^+}/P_{Cl^-} = 1.24), highlighting the tolerance of αHL to extreme pH conditions and the possibility of switching a pore selectivity by modifying the recording conditions.

Despite the ease with which protein pore selectivity can be tuned by modifying pH, adjustments in buffer pH carries an important caveat. It would inadvertently alter the

protonation state of protein analytes in a similar fashion to the nanopore's inner surface. For example, at low pH values, a biological nanopore composed of natural amino acids would tend exhibit a positive charge on its inner surface, resulting in an EOF that flows from the cis to the trans side under positive applied voltages at the trans side. However, the application of a positive potential would generate an EPF-driven pull in the trans lo cis direction for the protonated polypeptide analyte, which would counteract the EOF (Fig. 2.3). A similar trend is to be expected at high pH values in an inverted voltage orientation. In addition to this, pH tuning may compromise the stability of both the nanopore and analyte in solution, which may further complicate analysis.

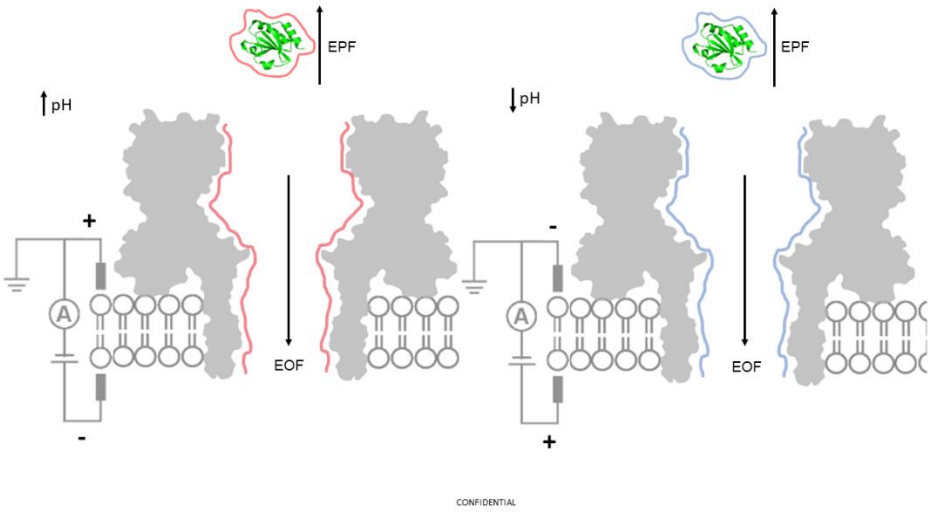


Fig. 2.3: pH changes affect simultaneously the protonation state of both the nanopore and the protein analyte, generating opposite EP and EO forces. Note that opposite pH regimes, the voltage needs to be reserved in order to maintain the directionality of both forces.

Given that the previous approach proved to be suboptimal, it was decided to induce EOF through generating mutant pores with additional charged residues. For this, acidic

or basic amino acids (for cation and anion-selectivity, respectively) were genetically substituted in the wild-type monomeric gene of αHL replacing originally neutral amino acids. Specific locations that do not compromise stability and are exposed to the pore lumen were simultaneously selected. Upon oligomerization into a heptamer, a ring of seven modified amino acids (one ring per modified amino acid in the monomer) with side chains exposed to the inner lumen are displayed. The Bayley group has performed exhaustive engineering on αHL in the past, including the fabrication of ion-selective pores. Four different mutants were selected to test selectivity and suitability for the analysis of protein analytes. The first pair of mutants *(NN-113R)*$_7$ and *(NN-113E)*$_7$, contained the following mutations over the wt pore located nearby the constriction: E111N/M113**R**/K147N, E111N/M113**E**/K147N, respectively. The asparagine mutations were previously shown to enhance the pore residence time of the neutral molecule β-Cyclodextrin, facilitating docking into the constriction, whereas introduction arginine and glutamate in position 113 were selected to confer anion and cation selectivity, respectively (Fig. 2.4).

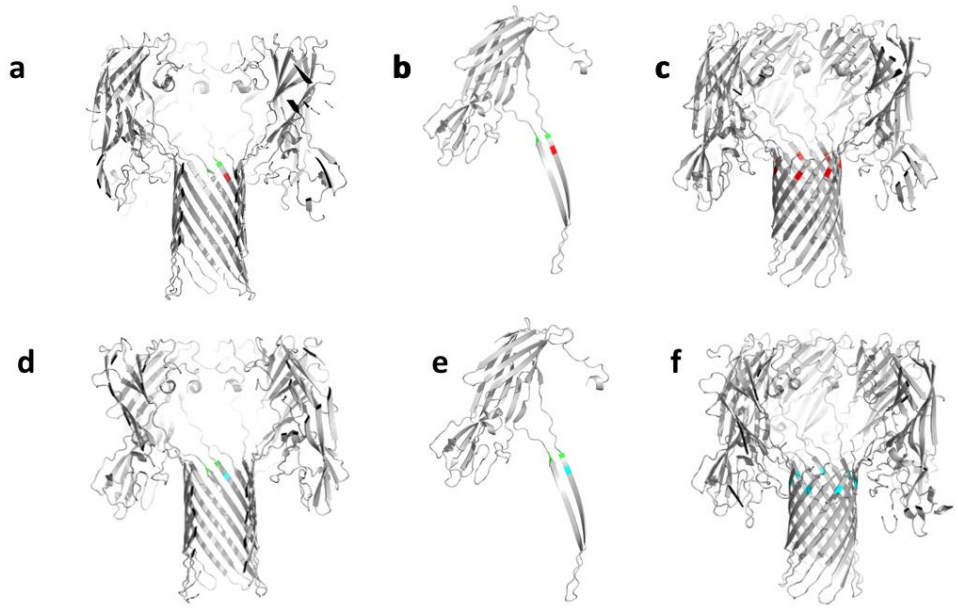


Fig. 2.4*:* Ribbon diagram of the front view of (NN-113E)$_7$ (top) and (NN-113R)$_7$ (bottom) αHL mutants. **a**, Mutated amino acids are colored in green (E111N, K147N) and red (M113E) on a single chain of the homo-heptamer. **b,** A single chain is shown **c,** All 7 Glu residues forming a ring at the constriction site are shown. **d,** Mutated amino acids are colored in green (E111N, K147N) and blue (M113R) on a single chain of the homo-heptamer. **e,** A single chain is shown. **f,** All 7 Arg residues forming a ring at the constriction site are shown.

The second pair of mutants, 4R and 4E both contained four mutations over *wt* αHL spaced out along the lumen: *(T115R/G119R/N123R/D127R-RL2)$_7$* and *(T115E/G119E/N123E/D127E-RL2)$_7$,* respectively. They were expected to show enhanced ion selectivity compared to the single single arginine / glutamate mutants, due to the presence of additional charges at the pore lumen surface (Fig. 2.5).

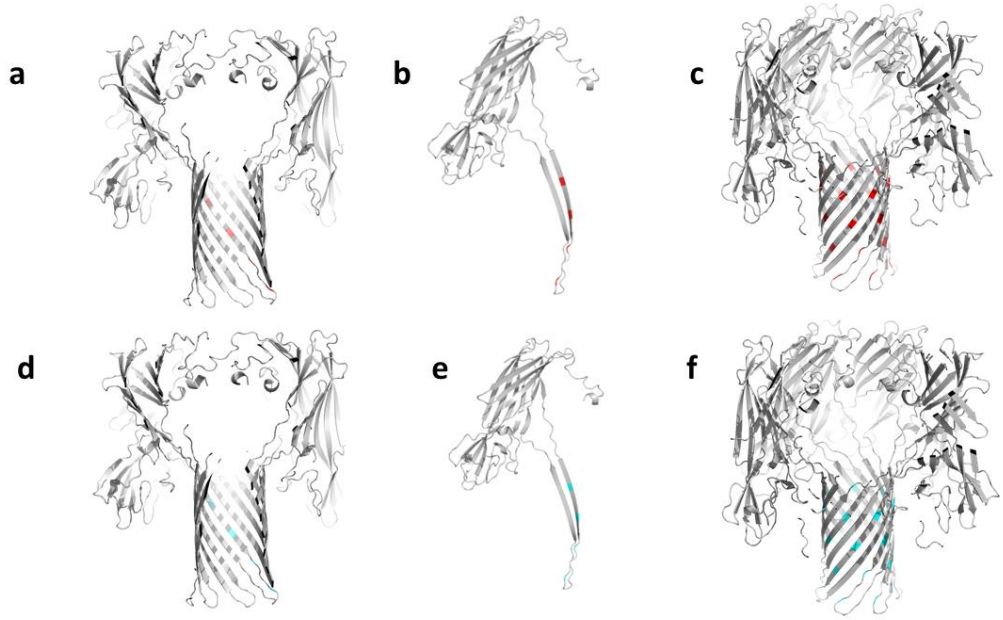


Fig. 2.5*:* Ribbon diagram of the front view of the 4E (top) and 4R (bottom) αHL mutants. **a,** Mutated amino acids are colored in red a single chain of the homo-heptamer. **b,** A single chain is shown. **c,** All 4 Glu rings across the pore β-barrel are shown. **d,** Mutated amino acids are colored in blue on a single chain of the homo-heptamer. **e,** A single chain is shown. **f,** All 4 Arg rings across the pore β-barrel are shown.

As predicted, αHL 4R mutant displayed very high anion selectivity (P_{K+}/P_{Cl-} = 0.07), while αHL 4E showed a more modest value (P_{K+}/P_{Cl-} = 3.75) (Table 2.1). Surprisingly, cation selectivity was similar for both the quadruple αHL 4E and the single *(NN-113E)*$_7$ mutant, which highlights the difficulty in predicting selectivity. Both quadruple mutant pores showed high heterogenicity in conductance, poor insertion into lipid bilayers, high levels noise and frequent gating, particularly at negative voltages, so they were discarded as nanopore sensors. On the other hand, αHL *(NN-113E)*$_7$ showed good selectivity (P_{K+}/P_{Cl-} = 3.37) and stability, but the ion carrying the current and therefore generating electroomosis was K+ (when using a KCl electrolyte). It was reasoned that

when using other electrolytes such as GdnHCl, the Gdn + ion would be carrying the charge, and, at the time, there was no information on the mobility rate of the ion available within nanopores, and therefore it was not possible to anticipate if it would be able to generate a strong enough EOF to drive protein translocation. Out of precaution, it was decided to discard this pore for further analysis, and all the work carried out onwards was done on the *(NN-113R)*$_7$ mutant.

Table 2.1: Summary of the different features displayed by the four ion-selective αHL mutants characterized.

	α-HL R	α-HL E	α-HL 4R	α-HL 4E
Conductivity (nS)	**2.07** (2M KCl)	**2.1** (2M KCl)	**1.09** (1M KCl)	**1.1** (1M KCl)
Vr (mV)	**8.57**	**-17.9**	**31**	**-19.2**
PK+/PCl-	**0.39**	**3.37**	**0.07**	**3.75**
PCl-/PK+	**2.56**	**0.3**	**14.85**	**0.27**
Noise	Moderate	Moderate	High (negative V)	High
Gating	Low (high V)	Low (high V)	High (negative V)	High

3. Translocation of monomeric proteins

3.1 Desing of Trx mutants of variable stabilities

Thioredoxin (Trx) was selected a model protein analyte to test translocation in ion-selective pores as it has been thoroughly studied in the past due to its robust expression, high stability and facility for purification. There is a high-resolution crystal structure available (PDB code 2TRX) and its stability and translocation dynamics have previously been characterized in nanopores [22], [107], [111]. Structurally, it displays a conserved fold consisting of a four-stranded β-sheet flanked by three α-helices, with catalytic cysteine pair (Cys32 and Cys35) at the N-terminal end of the second α-helix. It belongs to the oxidoreductase family of proteins (NADPH dependant), and it reduces oxidized cysteines in target proteins. Previous work at the Bayley group proved that Trx could translocate an αHL pore if derivatized with a short oligonucleotide in one termini [22], [23], [112]. The oligonucleotide strand would act as a leader, threading the pore first, electrophoretically pulling from the protein. αHL could accommodate α-helices within the pore, but unfolding was necessary for passage through the narrower ~1.3 nm constriction. A co-translocational unfolding mechanism was proposed, in which Trx unfolded in three steps (i.e., DNA strand threaded, partially unfolded intermediate, completely unfolded polypeptide), generating three different current levels associated with the time the protein would remain stalled within the pore in between those steps [111].

In this work, the wild-type version of Trx was mutated to substitute the internal cysteine pair to serines (C32S, C35S), which belong to the catalytic site. These two cysteines form a disulfide bond in the native state and it was reasoned it could hinder translocation through the pore, as it may be harder to linearize. In addition to this, C- and N- terminal cysteines were also introduced (S1C, C109) in order to have the option to functionalize the protein using the sulfhydryl handles, this mutant will hereafter be referred as Trx-V2. Its sequence is shown next, with the terminally inserted Cys coloured in green, and substituted Ser in the catalytic site in green:

CDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGT
APKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAC

N-terminal cysteine provides a key labelling site due to its specific reactivity, which allows for biorthogonal labelling to other cysteines (i.e., the C-terminal Cys), due to the availability of both the terminal amino and side-chain sulfhydryl groups. These sites can by specifically derivatized with CBT (cyanobenzothiazole) containing tags, whereas the C-terminal cysteine can be targeted with maleimide or other disulfide chemistries.

Trx-V2 was expressed in *E. coli* without a fused purification tag, and N-terminal cysteine is exposed as the starting methionine is naturally cleaved by the enzyme methionine aminopeptidase (MAP) in E coli [113]. However, the N-terminal cysteines are not reactive due to modification with cellular aldehydes such as pyruvate [114] , so they required an initial deprotection step by incubation with methoxyamine (Fig. 3.1).

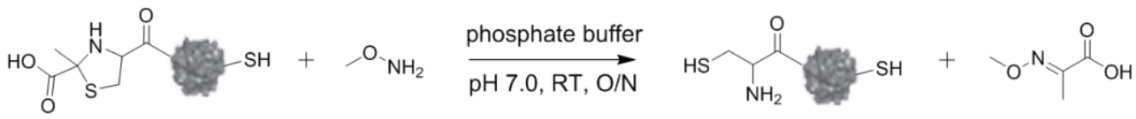


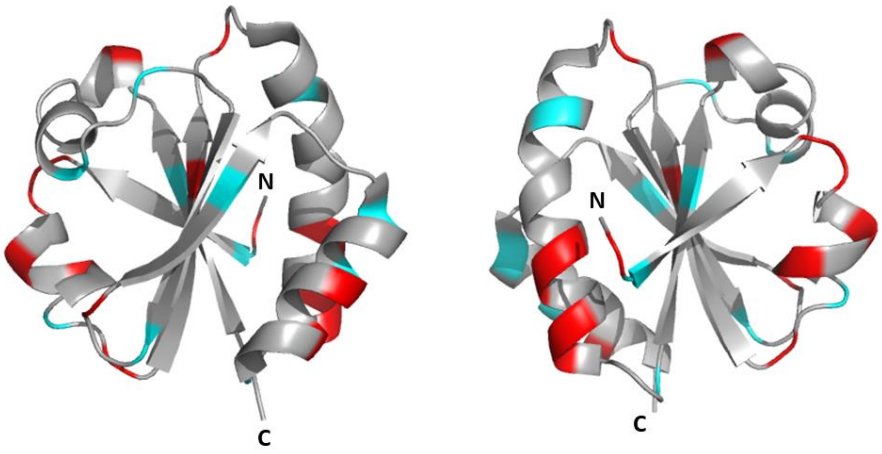
Fig 3.1: Incubation with high methoxyamine buffer is required for deprotection of the N-terminal cysteine, which is otherwise unreactive.

Translocation of Trx-V2 was initially explored in wild-type αHL, which is slightly anion-selective ($P_{K+}/P_{Cl-} = 0.79$) at neutral pHs. This was done to test whether even within weak ion-selective pores, an EOF could be generated capable of promoting capture of a protein analyte. Previous work had shown successful capture and translocation following Trx derivatization with a C-and N-terminal ssDNA tags, but results in an untagged analyte had not been reported. Pores were added to the cis side of the recording chamber for insertion into lipid bilayers. Upon single channel insertion, excess pore in solution was removed by perfusion of fresh buffer and Trx-V2 analyte was consequently added to cis, where the pore cap is oriented.

The recording buffer was 10 mM Hepes pH 7.2, and varying concentrations of salt were tested: 1M or 2M KCl, to account for charge screening within the pore (i.e., it has been reported that reducing the salt concentration may enhance the pore selectivity) [45] . Under neutral pH conditions, Trx-V2 (pI = 4.67) is expected to display a net charge of -5, which would create an EPF favoring capture and translocation.

Previously reduced (after incubation in DTT or TCEP) Trx-V2 was added on the cis side at ~ 5 uM. As a negative control, negative potentials (- 80mV to - 180 mV) were applied on trans. Under this applied voltage, an EOF and EPF in the trans-to-cis direction are expected, working against capture of the analyte. No interactions between the analyte and the pore were observed in three different pores. Next, positive potentials (+ 80mV to +180 mV) were applied, creating an EOF and EPF in the cis-to-trans direction. Synergism between these two forces still did not promote any interaction between the analyte and the pore. It was reasoned that both the EOF and EPF were too weak to cause any capture by the pore, and therefore, approaches to increase the selectivity of the pore were followed.

Next, interaction of Trx-V2 with the $(NN\text{-}113R)_7$ mutant pore was explored. Recording buffer was 10 mM Hepes, pH 7.2, 1M KCl. Under negative applied potentials, no interaction between the analyte and the pore was observed, but upon switching to positive potentials, multiple current patterns were immediately observed. These patterns showed an almost featureless partial blockade of the ionic current, highly suggestive of bumping of the analyte over the pore entry, rather than translocation events. Because the protein analyte is folded and there is no charged leader and there are not strong net charge segments across its sequence, it was reasoned that threading of the polypeptide would be highly unfavorable due to the high entropic barrier (Fig. 3.2).



CDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGT

APKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAC

Fig. 3.2: Ribbon diagram of the X-ray structure of oxidized wt Trx. Negatively charged and positively charged residues colored in red and blue, respectively. (Bottom) The sequence of wt Trx is displayed following the same color code.

Encouraged by those findings, it was decided to test high concentrations of urea added on the cis side of the chamber, to promote unfolding of the protein and facilitate end threading. Urea is a protein denaturant reported to displace the protein solvation shell via interactions with the protein backbone, hydrophilic and hydrophobic residues via hydrogen bonding, electrostatic interactions and dispersion interactions, respectively. These protein-urea interactions are more stable than the protein-protein ones, shifting the equilibrium from the native to the denatured conformation.

Before characterizing Trx-V2 in the electrophysiology setup, the thermodynamic stability of Trx-V2 in urea was experimentally measured in bulk. This type of analysis

provides an overview of the transition state of the protein from the folded state to an unfolded or semi-unfolded conformation, as a measure of denaturant concentration and allows to determine the concentration of denaturant at which the protein is completely unfolded. Technically, a series of Trx-V2 solutions with increasing concentrations of denaturant are prepared, incubated until equilibrium is reached, and then characterized with a spectroscopic technique that allows to monitor the change in protein structure.

Circular dichroism and fluorescence emission are the two main spectroscopic techniques employed to monitor protein structural changes. The first is the preferred method, as it provides with a direct measurement of the degree and type of secondary structure, but as a downside, it requires high concentration of protein due to lower sensitivity and is not compatible with denaturants that display optical activity (e.g., GdnHCl and GTC) and can interfere with the signal. On the other hand, fluorescence emission microscopy cannot report on the type of secondary structure, but displays exquisite sensitivity, allowing for very low usage of protein sample. As a fluorescent reporter, artificial dyes can be chemically attached to the analyte, but often, the naturally present fluorescent amino acids tryptophan, phenylalanine and/or tyrosine are enough. These groups display solvatochromism, a phenomenom by which the absorption and emission spectrums are modulated by the polarity of the medium. By selecting fluorophores buried within the protein core in the native state, a loss of structure will change the environment to a more polar one, resulting in changes to the fluorescence emission signal.

In this work, CD was used for the unfolding experiments with Urea, whereas fluorescence emission spectroscopy was chosen for GdnHCl and GTC (Fig. 3.3) Trx-V2 was observed to be a very stable protein and remained fully folded until 5~M urea. Up to 8M urea was needed for complete unfolding (figure). As previously reported with wt Trx, Trx-V2 followed a one-step unfolding mechanism in bulk for all chaotropes used. GdnHCl was a more potent denaturant than Urea and required a concentration of ~3 M (2.5 times less) to induce complete loss of the folded state. Finally, GTC was the most potent denaturant from all three, and showed complete unfolding at ~ 1 M, which makes it ~ 3 times more potent than GdnHCl and ~ 8 times more potent than Urea.

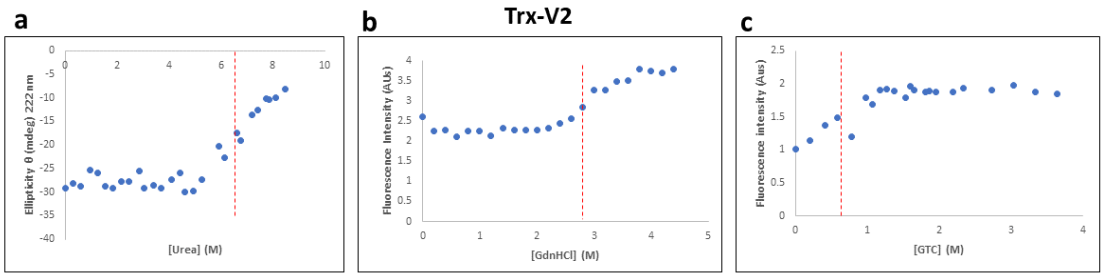


Fig. 3.3: Unfolding curves of Trx-V2 in (a) Urea, (b) GdnHCl and (c) GTC. Incubation of the protein with denaturant was done in 10 mM Hepes, 1 M KCl, pH 7.2. The dashed red line has been manually inserted to indicate the approximate concentration of denaturant at the unfolding transition state (concentration of denaturant at which 50% of the protein is unfolded).

Informed by the bulk characterization of Trx-V2 stability, it was decided to use a concentration of 9 M urea in electrophysiology experiments in order to test translocation on a completely unfolded polypeptide. At postive applied voltages, the most abundant patterns displayed no features and had duration of a few ms (current

spikes), eliciting an almost complete bockade of current. These type of patterns were highly suggestive of fast translocation events. Less often, a pattern of longer translocation time was observed, but it was not possible to identify any feature within it. This last type of pattern likely represents the interaction of a partially folded protein with the nanopore, which dwell longer time in passing through the pore.

At 9 M urea, lipid bilayer stability was greatly reduced, and recording could only be done for a few minutes. Furthemore, αHL (NN-113R)$_7$ showed reduced open pore current under those conditions, which suggests that the structure of the channel is greatly compromised.

It was reasoned that a protein of lower stability than Trx-V2 would enable working at lower concentration of denaturant and, therefore, overcome those issues, so a new Trx mutant with 4 additional mutations over Trx-V2 was designed: V2-S76 (C1S, I73R, G75A, I76S). This new mutant had the N-terminal cysteine removed, to avoid having to deprotect the protein, thus facilitating its preparation and handling. It also contained a Serine phosphorylation recognition site for PKA (RRAS), which was specifically created in one of Trx natural loops in an attempt to reduce the deleterious impact of the mutations on the protein structure. Similarly, it was reasoned that the loop would be easily accessible by PKA, facilitating its subsequent phosphorylation (Fig. 3.4)

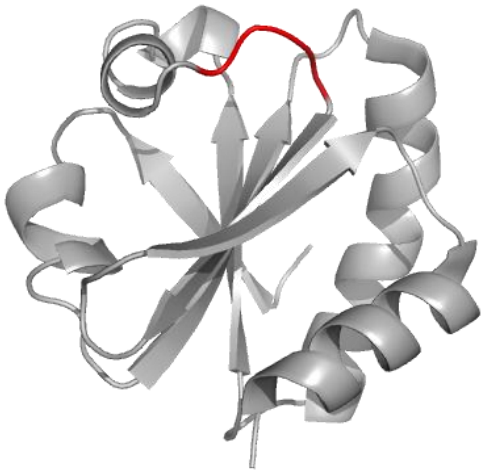


Fig. 3.4: Ribbon diagram of the X-ray structure of Trx-V2-S76 mutant. The engineered RRAS site, colored in red, is located within a flexible loop that seems accessible to the solvent.

Similarly to Trx-V2, Trx-V2-S76 was initially characterized in bulk. Tolerance to denaturant was greatly reduced for this mutant. At ~3.5 M Urea, ~2 M GdnHCl, and ~0.5 M GTC, the protein showed complete loss of structure (Fig. 3.5).

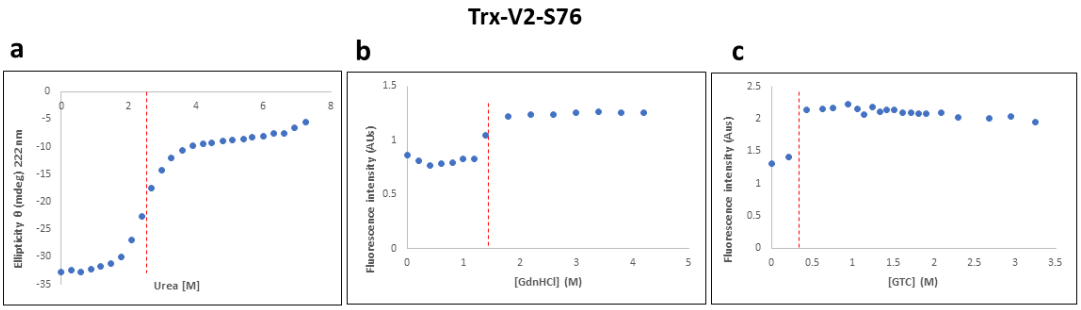


Fig. 3.5: Unfolding curves of Trx-V2-S76 in (a) Urea, (b) GdnHCl and (c) GTC. Incubation of the protein with denaturant was done in 10 mM Hepes, 1 M KCl, pH 7.2. The dashed red line has been manually inserted to indicate the approximate concentration of denaturant at the unfolding transition state (concentration of denaturant at which 50% of the protein is unfolded).

Translocation of this mutant at 9M urea elicited short patterns of almost complete current blockade (current spikes), which totally lacked any feature, similar to those observed with Trx-V2. Interestingly, no other type of pattern was observed, which suggests that the protein analyte is always found in an unfolded state when captured by the pore. The lower stability of V2-S76 mutant is a probable explanation for this observation.

3.2 Proving translocation

Encouraged by these findings, it was decided to test V2-S76 at several concentrations of more potent denaturant. Increasing concentrations of GdnHCl were titrated in the recording chamber and the current patterns characterized at 0.3 M GdnHCl (Fig. 3.6).

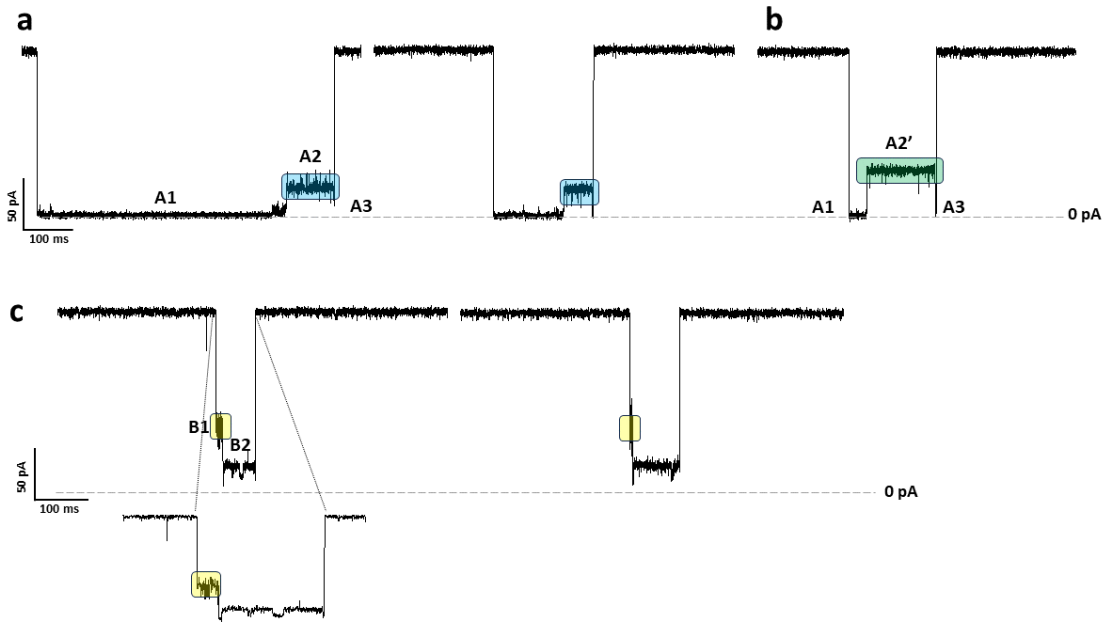


Fig. 3.6: Most characteristic patterns observed with Trx-V2-S76 analyte. **a,** pattern A shows three levels A1, A2 ($I_{RES\%}$ = 15%), A3. **b,** pattern A' is similar to pattern A, but a new level A2' ($I_{RES\%}$ = 29%) is observed.

c, A third and most abundant pattern is seen, with two main levels B1 ($I_{RES\%}$ = 33%) and B2 ($I_{RES\%}$ = 10%). Recording conditions were 10 mM Hepes, pH 7.2, 1 M KCl, 1 mM TCEP, 0.3 M GdnHCl, +140 mV, with charge-selective nanopore ((NN-113R)7).

Patterns A and A1 lasted for hundreds of ms and contained a final level of almost complete current blockade, which is suggestive of translocation events. It is also likely that each one of them corresponds to a different end of the protein threading the pore first and leading the translocation. On the other end, pattern B was considerably shorter, lacked a final level of complete current blockade, and contained a level B1 of high noise, which is usually associated with high freedom of movement of the analyte in the pore, so it most likely represents an interaction event of the analyte with the pore rather than translocation. The recording concentration of GdnHCl was of only 0.3 M, and V2-S76 mutant was shown to require ~ 2 M for complete unfolding in bulk, what likely explains why the most abundant pattern is an interaction event which does not proceed to translocation.

In order to gain more insights into the translocation mechanism of Trx-V2-S76 with GdnHCl, electrophysiology was recorded at multiple other concentrations of this denaturant (Fig. 3.7). At voltages above +100 mV, higher GdnHCl concentration correlated with higher translocation rates, indicative of lower structural stability and facilitated translocation, however, an increase in voltage seemed to have a negative effect on translocation rate at 1 M and 0.6 M GdnHCl. Generally, it is accepted that an increase in translocation dwell time as a matter of applied voltage is characteristic of analyte interaction of the pore rather than translocation, which may indicate that,

under those conditions, the protein is indeed only interacting with the pore. At the lowest denaturant regimes (i.e., 0.1 M and 0.3 M GdnHCl), voltage did not seem to affect the translocation rate, which may be indicative of a diffusion governed translocation mechanism, where once the analyte has been captured, the EOF acting on it is greatly reduced and the protein spontaneously unfolds and translocates the pore.

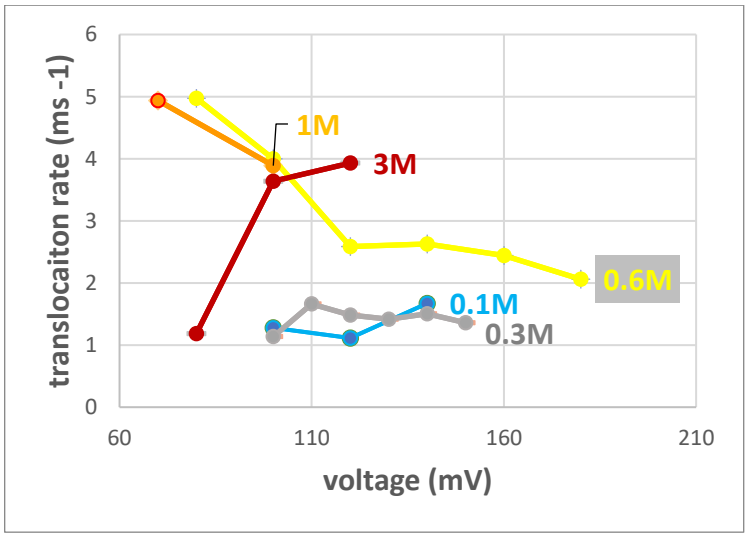


Fig. 3.7: Translocation rate of Trx-V2-S76 is dependent on GdnHCl concentration and voltage bias. Translocation rates at 5 different denaturing conditions are shown as a function of voltage (positive applied voltage on trans). In all cases, recording conditions were 10 mM Hepes, 1 M KCl, 1mM TCEP pH 7.2, with charge-selective nanopore ((NN-113R)7). At least 50 events were used to calculate the translocation rate at every GdnHCl concentration.

In order to further understand the translocation mechanism of V2-S76 and obtain more solid evidence of translocation, dimers of V2-S76 were tested on the electrophysiology setup. The dimers were formed through disulfide linkage of the C-terminal cysteines

(Fig. 3.8). It was noticed during purification that, if no reducing agent was used during cell lysis, an important proportion of the protein was found in its dimeric form, which was easily resolved from the monomer during purification by SEC.

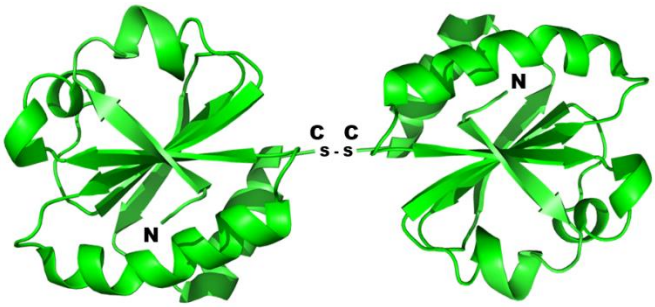


Fig. 8: Ribbon diagram of the X-ray structure of dimerized Trx-V2-S76 through its C-terminal cysteine.

In contrast with monomeric V2-S76, a single type of pattern was observed with the dimer (Fig. 3.8). This new pattern C contained three levels, C1, C2, C3, which were very similar to the ones observed in patterns A and A' with the monomer. However, level C2 showed a new current level ($I_{RES\%}$ = 20%), distinct from A2 (15%) and A2' (29%).

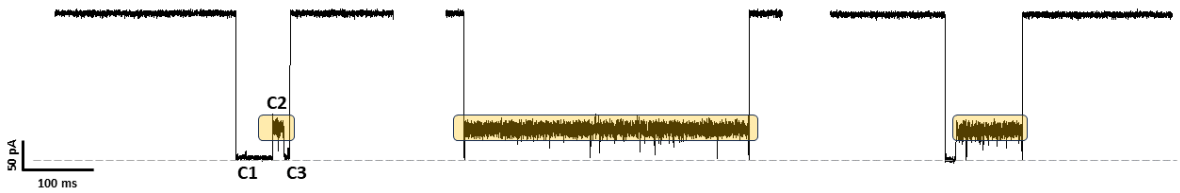


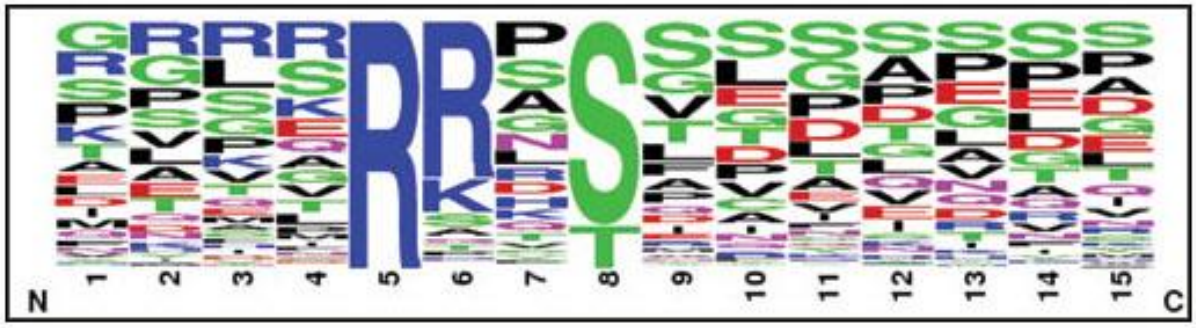
Fig. 3.8: Most representative patterns of Trx-V2-S76 translocation. A single pattern C was observed. Recording conditions were 10 mM Hepes, pH 7.2, 1 M KCl, 1 mM TCEP, 0.3 M GdnHCl, +140 mV, with charge-selective nanopore ((NN-113R)7).

Assuming the protein must thread the pore with one of its ends first, the observation of only one type of pattern is highly suggestive of a single translocation mechanism through the nanopore, which is expected due to the fact that only the N-terminal ends of the protein were free for threading the pore (since the C-terminal cysteines were forming the disulfide linkage with another unit). Therefore, the first unit entering the pore was obligated to translocate N-terminus first, whereas the second unit had to translocate C-terminus first. The presence of a new level C2 with a different current level than A2 and A2', points towards a new translocation mechanism, which likely involves the second Trx unit entering the pore before the first one has fully translocated, generating a signal overlap between the two.

3.3 Nanopore detection of phosphorylations

Encouraged by these findings, which suggested that Trx-V2-S76 might be translocating the pore, Trx-V2-S76 was phosphorylated on its serine 76 in an attempt to test whether the pore would be able to discriminate the modified from the unmodified version of the protein. Trx-V2-S76 was incubated with PKA in the presence of ATP, and the incorporation of the phosphate group confirmed by ESI-MS. Surprisingly, only ~ 70% of the protein was phosphorylated after incubation for more than 72h, which suggests that the residues flanking the RRAS recognition sequence might not be optimal. In fact, upon revisiting the consensus phosphorylation recognition site for PKA (Fig. 3.9), it was

found that a proline following the serine phosphorylation residue is a rare find. Additionally, this proline forms an invariant (Ile75-Pro76) cis prolyl peptide bond in the native state [115], which is known to be a rate limiting step in the folding of the polypeptide and may sterically limit the accessibility and / or recognition by PKA.



Trx-V2-S76: N-ter…DQNPGTAPKYGRRASPTLLLFKNG…C-ter

Fig. 3.9: PKA substrate site sequence showing the Shannon entropy content. The height of each column reflects the bias in the distribution of the amino acid residues found at that position. (Bottom) PKA recognition site engineered in Trx (red), the proline residue following the site is shown in blue.

A single pore was used to analyze unphosphorylated Trx-V2-S76 and the partially phosphorylated variant. It was not possible to observationally distinguish any new feature within the A, A1 and B1 levels that was indicative of phosphorylation sensing by the pore (Fig. 3.10). Upon representing the data in scatter plots, the 3 features analyzed tended to accumulate into three well defined populations with a wide degree of spread in noise.

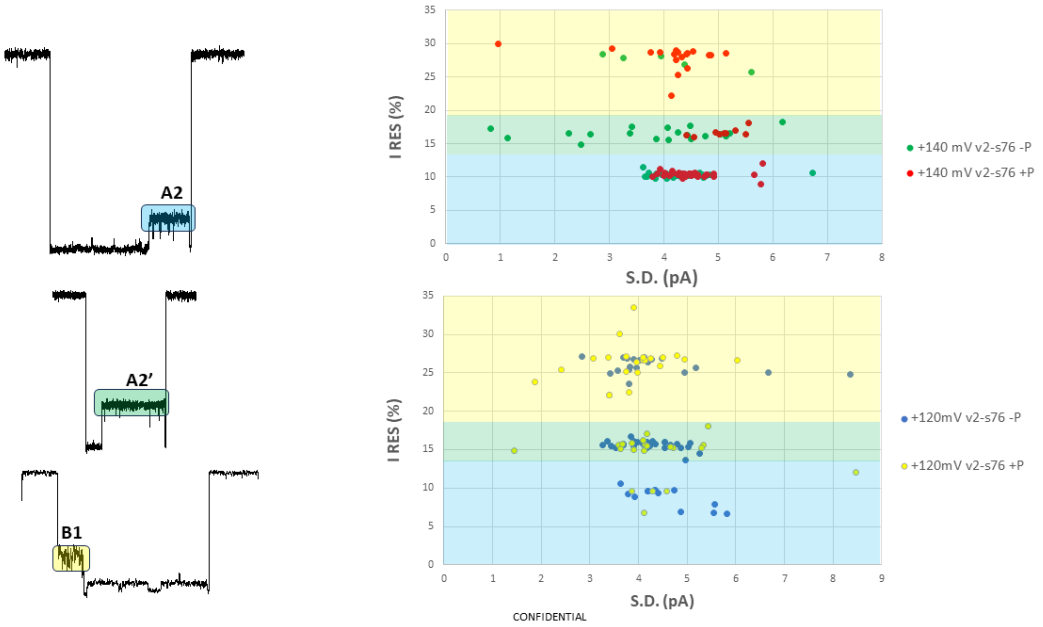


Fig. 3.10: Scatter plots of $I_{RES\%}$ and current standard deviation (noise) performed at +140 mV (top) and +120 mV (bottom) with phosphorylated and unphosphorylated Trx-V2-S76. The area where A2, A2', and B1 current levels are located within the scatter plot is shaded in blue, green, and yellow respectively. Recording conditions were 10 mM Hepes, pH 7.2, 1 M KCl, 1 mM TCEP, 0.3 M GdnHCl, at either +120 mV or +140 mV, with charge-selective nanopore ((NN-113R)7).

It is likely that the position at which the phosphorylation is located is not well sampled by the pore due to high-speed translocation of that particular segment of the polypeptide, below the detection limit of the technique. Furthermore, the lack of a strong pulling force (EOF and EPF) reduces the level of stretching of the polypeptide inside the pore, which further complicates discrimination.

In summary, label-free, EOF-driven translocation of monomeric Trx generated patterns with very variable levels, which complicates analysis (e.g., the protein analyte can

translocate with either the N- or C-terminal end first, which is expected to generate two different current patterns [22]. These levels do not resemble the ones observed during EPF-driven translocation of ssDNA derivatized Trx [107], which suggests that the protein analyte undergoes a different translocation mechanism.

Mild denaturing conditions proved useful for increasing the throughput (number of observed events) and seemed to promote translocation events over “interaction” events of the analyte with the pore, that were observed when recording at native conditions. On the other hand, strong denaturing conditions, near or over the unfolding transition state, seemed to generate short-lived patterns with a lack of features, that tended to cause an almost complete blockade of current. This lack of current makes sensing of elements at the sequence level very challenging, if not impossible, with the current methods.

It is proposed that mild denaturing conditions do not unfold, but nonetheless destabilize the tertiary structure of proteins such as Trx, which upon being captured by the pore, undergo co-translocational unfolding at a faster rate than when sampling at native conditions. Tuning the amount of denaturant based on the tolerance of a particular protein analyte to it may be key for ensuring successful translocation through the pore while maintaining structural elements that enable discrimination.

4. Translocation of Trx concatemers

4.1 Assembly and purification of concatemers with variable length

A definitive experimental strategy was conceived to prove enzyme-less, tag-less, electroosmotically-driven translocation of a protein. This approach involved the assembly of protein constructs comprising dimers, tetramers, hexamers and octamers of Trx units. The underlying rationale posited that the successful traversal of these Trx-based constructs through the pore would necessitate a series of discrete processes, including their capture, unfolding, and translocation as individual units. Consequently, it was postulated that the resultant electrical profile of the complete polypeptide would exhibit distinct discernible features corresponding to each Trx monomer present within the construct. (Fig. 4.1).

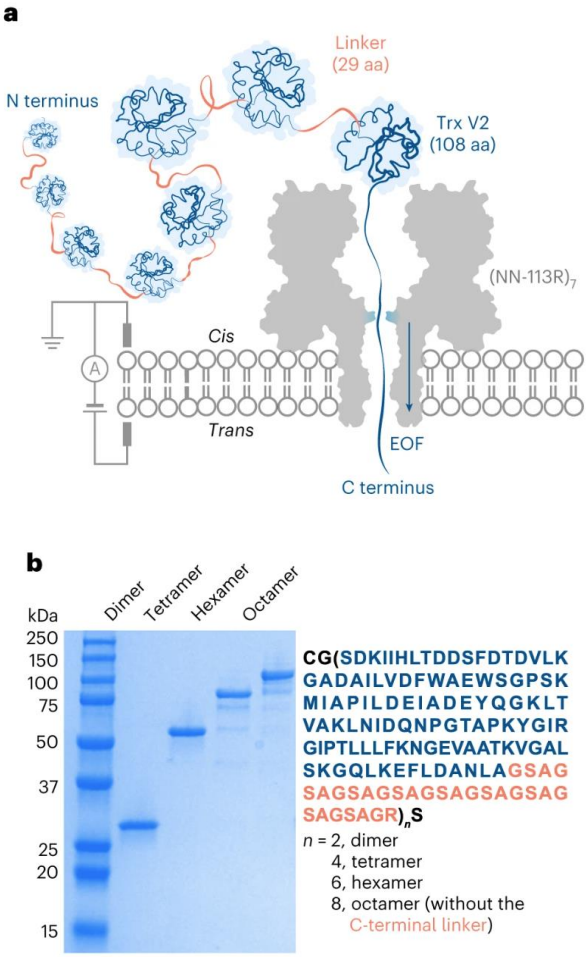


Fig. 4.1: Electro-osmosis-driven translocation of Trx-linker concatemers through a protein nanopore. a, EOF in a charge-selective αHL nanopore $(NN\text{-}113R)_7$ drives the sequential co-translocational unfolding of Trx units within a polyprotein of >1,000 aa. **b**, A sodium dodecyl sulfate–polyacrylamide gel showing the Trx-linker dimer (28 kDa), tetramer (55 kDa), hexamer (83 kDa) and octamer (110 kDa). Figure reproduced from [116].

As described on the previous chapter, Trx (108 amino acids (aa)) had the two catalytic cysteines removed (Trx:C32S/C35S) to facilitate linearization. The Trx monomers were connected by 29 aa linkers, capable of spanning the 10-nm-long lumen of the αHL nanopore when fully extended (0.35 nm per aa). This was done to prevent a Trx unit entering the pore before the previous one had completely exited and hence avoiding a

potential signal overlap. Additionally, the presence of a flexible, unstructured C-terminal linker was predicted to facilitate polypeptide threading into the pore. An exception was that the Trx-linker octamer, which had no C-terminal linker (Table 4.1).

Assembly of the concatemers was performed taking advantage of the cohesive end compatibility of DNA ends digested with BamHI and BgIII restriction enzymes. Upon hybridization of such two sites, the new sequence formed is a hybrid of the two restriction sites, which cannot be recognized and cleaved by any of the enzymes. This allowed iterative cloning of a Trx-linker monomer into another monomer, dimer into dimer and tetramer into tetramer (Fig. 4.2).

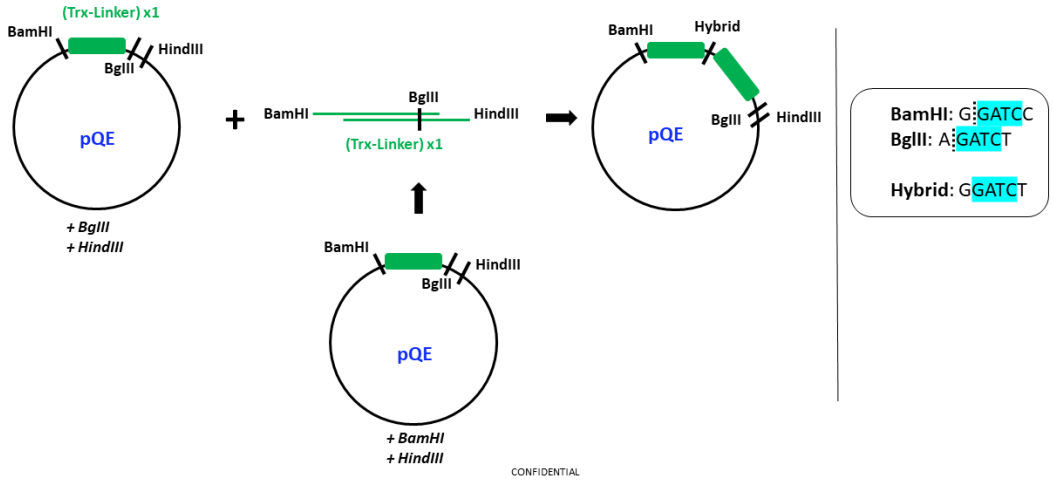


Fig. 4.2: Gene assembly of Trx-linker concatemers. A pQE plasmid containing a single Trx-linker unit is opened up upon digestion with BgIII and HindIII restriction enzymes. A second pQE plasmid containing a single Trx-linker unit is digested with BamHI and HindIII restriction enzymes and the Trx-Linker insert purified from the rest of the plasmid. The opened plasmid and purified insert are mixed and ligated together, generating a Trx-linker dimer. This process is repeated multiple times until the desired number of Trx-linker units is assembled.

Table 4.1: Sequences of the thioredoxin-linker concatemers: Blue: Trx; Yellow: linker; Red: N-terminal cysteine-glycine; White: restriction enzyme site.

Dimer (Trx-linker)2

CGSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNP
GTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAG
SAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNI
DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGS
AGSAGSAGSAGRS

Tetramer (Trx-linker)4

CGSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNP
GTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAG
SAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNI
DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGS
AGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLT
VAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSA
GSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEY
QGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAG
SAGSAGSAGSAGSAGSAGRS

Hexamer (Trx-linker)6

CGSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNP
GTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAG

SAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNI

DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGS

AGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLT

VAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSA

GSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEY

QGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAG

SAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILD

EIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGS

AGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKM

IAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKG

QLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGRS

Octamer (Trx-linker)8

CGSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNP

GTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAG

SAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNI

DQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGS

AGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLT

VAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSA

GSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEY

QGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAG

SAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILD

EIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGS

AGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKM

IAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDA

NLAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWS

GPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQL

KEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDF

WAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGA

LSKGQLKEFLDANLA

In order to ensure that each individual Trx unit within the concatemers had reached a folded state similar to monomeric Trx, a CD spectra of each of the concatemers was collected and compared to the monomeric Trx-V2 (Fig. 4.3). The spectra of all concatemers displayed a shape similar to monomeric Trx, which suggests that the Trx units were able to acquire the native folded state.

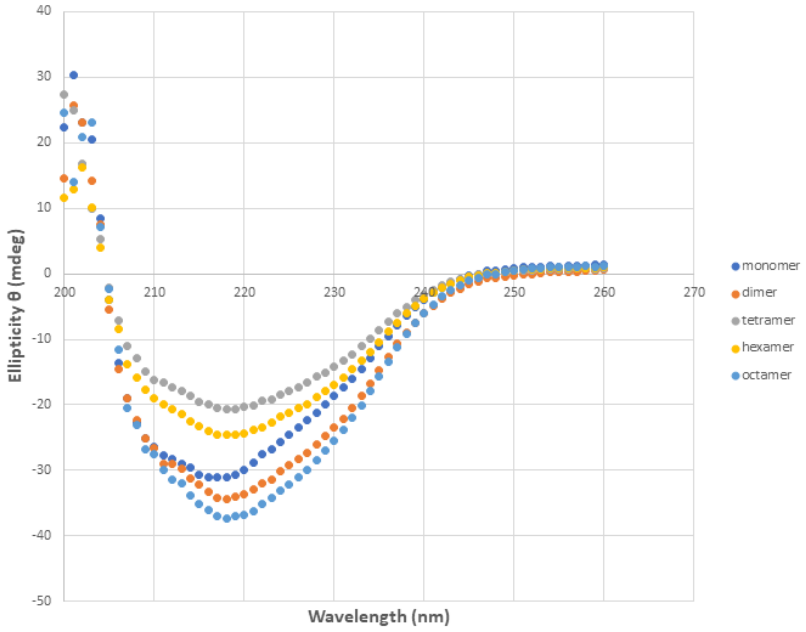


Fig. 4.3: CD spectra overlays of monomeric Trx-V2 (dark blue), Trx-Linker dimer (orange), tetramer (grey), hexamer (yellow), and octamer (light blue).

4.2 Proving translocation of Trx concatemers

4.2.1 Characterization of translocation events

Just as with Trx monomers, translocation events were not observed when testing any of the concatemers at non-denaturing conditions on WT αHL pore (P_{Na+}/P_{Cl-} = 0.78) at neither positive or negative voltage bias, what is attributed to the weak electroosmotic and electrophoretic force acting on the analytes, which is insufficient for promoting capture. A weak EOF in the cis to trans direction is expected at positive voltages applied on trans, but this proved to be insufficient to initiate translocation events.

In order to facilitate translocation, the sample buffer was supplemented with mild concentrations of denaturants to destabilize the tertiary structure of the protein and promote its unfolding. Initially, 750 mM KCl was replaced with 750 mM GdnHCl as a tentative choice for creating mild denaturing conditions, which were confirmed to be below the unfolding transition state for the Trx monomer, as discussed in the previous chapter. These altered conditions resulted in the emergence of blocking patterns characterized by repetitive features, indicative of capture and translocation events. However, it is noteworthy that these events did not progress to completion, and the arrested polypeptide required a reversal of voltage on the electrodes to revert to the open pore state. The capture rate observed under these conditions was approximately 0.11 s^{-1} μM^{-1} for all concatemers.

Given that the net force being exerted on the analytes was still not strong enough to drive translocation, it was decided to utilize the previously characterized anion-selective αHL mutant (NN-113R)7 (P_{Na+}/P_{Cl-} = 0.33) to generate enhanced electroosmosis.

An initial trial was performed at non-denaturing conditions, at which all the four Trx-linker concatemers were captured, unfolded and translocated by (NN-113R)7. However, the translocation dwelled for tens of seconds and would often not proceed to completion (voltage was reserved after 5 minutes to release the trapped polypeptide and recover the open pore state) at +140 mV. At higher voltage regimes (+160-200 mV) a few complete translocation events were observed, but the recording time was drastically reduced to only a few minutes, since the lipid bilayer is highly unstable under those conditions. The change in translocation behaviour of the concatemers was attributed to an enhanced electroosmotic force in the cis to trans direction due to the increased ion selectivity of the mutant pore compared to the WT. However, the increased net force acting on the analyte was still not optimal for single-molecule characterization of a highly stable protein such as Trx.

In a continued effort to enhance translocation throughput, mild concentrations of denaturants were added to the sample buffer to destabilize the tertiary structure and facilitate unfolding. Initially, 750 mM GdnHCl was trialled, and remarkably, the capture and translocation rate of the concatemers increased substantially. All the four Trx-linker concatemers were captured by (NN-113R)7 in the presence of 750 mM GdnHCl (Fig. 4.3) at positive applied voltages, at a capture rate ~25 times faster than that of the WT αHL pore ($k_{(octamer)} \approx 2.50\ s^{-1}\ \mu M^{-1}$ with (NN-113R)7 versus~$0.11\ s^{-1}\ \mu M^{-1}$ with wt αHL. The recording conditions were 750 mM Gdn-HCl, 10 mM HEPES at pH 7.2, +140 mV (trans), 24 ± 1 °C.

Concatemer translocation produced current patterns containing repeating features (Fig. 4.4). The most abundant feature, A, consisted of three levels (A1, A2 and A3).

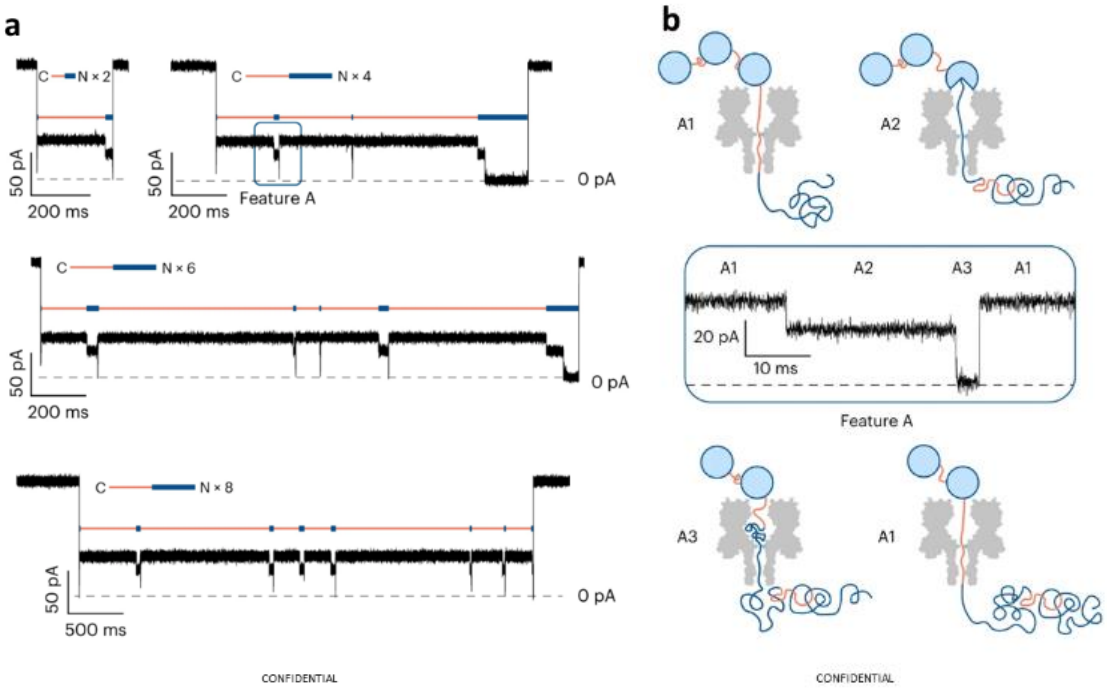


Fig. 4.4: Electro-osmosis-driven translocation of Trx-linker concatemers through a protein nanopore: a, Current recordings for the C-terminus-first translocation of a dimer, a tetramer, a hexamer and an octamer without post-acquisition filtering. The repeating features A are indicated by orange and blue bars. **b**. Zoomed-in view of the repeating feature A boxed in blue in **a** without post-acquisition filtering. Three levels are assigned as follows: A1, a linker within the pore; A2 and A3, different segments of partly unfolded Trx within the pore. Recording conditions are as follows: 750 mM GdnHCl, 10 mM HEPES, 5 mM TCEP at pH 7.2, Trx-linker concatemers (*cis*) (dimer: 2.23 µM; tetramer: 0.63 µM; hexamer: 0.25 µM; octamer: 0.81 µM), +140 mV (*trans*), 24 ± 1 °C. Figure reproduced from [116].

Interestingly, a spike to ~0 pA was seen at the beginning of almost all the translocation events and was speculated to represent the rapid unfolding and translocation of the first Trx-linker unit. The spike was followed by up to n – 1 repeats of the three-step

feature A (n is the number of Trx-linker units in the concatemer), which unambiguously demonstrated the stepwise translocation of entire polypeptide chains one unit at a time.

Trx has previously been reported to be a two-state unfolder in solution [117], [118] which is in agreement with the observations in the unfolding experiments performed (Fig. 3.5). However, the presence of A1 and A2 levels suggest a two-step unfolding mechanism, where at least one intermediate is present. In agreement with this, a previous translocation approach in which Trx was terminally labelled with a ssDNA oligo and electrophoretically driven through wt αHL also reported a two-step co-translocational unfolding mechanism [107]. The fact that Trx has been judged to be a two-state unfolder in solution is not necessarily surprising given that (i) the pore presents a physical barrier at its entrance that most likely forces the protein to take a different unfolding route and (ii) single-molecule experiments are able to discern non-rate-limiting steps that are hidden in bulk experiments [119].

Tentatively, we assigned level A1 as a threaded linker preceding the C terminus of a folded Trx unit; level A2 as the C-terminal portion of a partially unfolded Trx unit extended into the nanopore; and level A3 as the spontaneous unfolding and passage of the remaining Trx polypeptide through the nanopore (Fig. 4.4 b).

The percentage residual current ($I_{res\%}$) for each level in feature A was consistent across all such events for each polypeptide translocation and between all the individual concatemers observed with the same or different pores (Table 4.2). It was reasoned a

either a remainder of secondary structure or coiling of Trx terminal region caused the current drop observed in level A3.

Table 4.2: Percentage residual currents ($I_{res\%}$) for the three levels of repeating feature A recorded during C-terminus first concatemer translocation.

	Trx-linker dimer	Trx-linker tetramer	Trx-linker hexamer	Trx-linker octamer	Trx-linker octamer[c]	Trx-linker octamer[d]
$I_{res\%}$ (A1)[a]	34 ± 1%	35 ± 1%	34 ± 1%	35 ± 1%	26 ± 2%	35 ± 1% 32 ± 1%
$I_{res\%}$ (A2)[a]	22 ± 3%	24 ± 2%	23 ± 1%	23 ± 1%	16 ± 1%	24 ± 1%
$I_{res\%}$ (A3)[a],[b]	1.9 ± 2.4%	1.7 ± 1.7%	2.2 ± 2.3%	2.3 ± 2.7%	0.5 ± 1.7%	2.0 ± 2.0%
N[a],[b] n	105 units 2 separate pores	66 units 3 separate pores	122 units 1 pore	443 units 2 separate pores	65 units 3 separate pores	47 units 2 separate pores

[a] $I_{res\%}$ was calculated for each step in individual features A as the remaining current as a percentage of the open pore current (e.g., $I_{res\%(}A1) = I_{A1}/I_{open}\times100\%$). The standard deviations were derived for N Trx-linker units collected with 'n' separate pores. Conditions: 750 mM GdnHCl, 10 mM HEPES, pH 7.2, +140 mV (trans), 24 ± 1 °C. Concatemer concentrations (cis): dimer 2.23 µM, tetramer 0.63 µM, hexamer 0.25 µM, octamer 0.81 µM.

[b] Trx-linker units that produced a Level A3 with a dwell time 1 ms and a square shape were included in the $I_{res\%}$ analysis.

[c] Conditions: 750 mM KCl, 10 mM HEPES, 0.81 µM octamer (cis), pH 7.2, +140 mV (trans), 24 ± 1 °C. **[d]** Conditions: 750 mM GdnHCl, 10 mM HEPES, 0.81 µM octamer (trans), pH 7.2, -140 mV (trans), 24 ± 1 °C. A sub-conductance level was seen at Level A, which might be attributed to the folded thioredoxin unit adopting two conformations as a stopper under the electroosmotic force at the trans opening of the pore.

4.2.2 Distinguishing N- vs C-terminal first translocation events

During electrophysiology recordings, a less frequent type of event was also recorded, consisting of different repeating element B (Fig. 4.5 and Table 4.3). The N-terminus of the concatemeric constructs lacked a terminal flexible linker, which was predicted to

facilitate capture by the pore, therefore, it was reasoned that they might correspond to N-terminal first translocation.

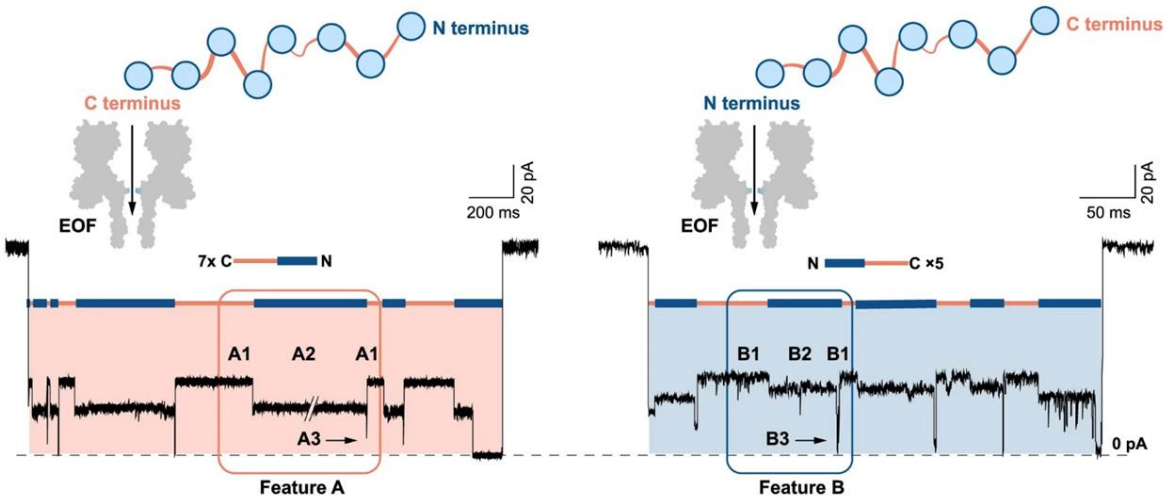


Fig. 4.5: Differential repeating current features recorded during electroosmosis-driven octamer translocation through a nanopore. Two repeating current features, A or B, were recorded with a charge-selective nanopore ((NN-113R)7) and Trx-linker octamers pre-treated with 5 mM TCEP for 10 min before their addition to the cis compartment of the recording chamber. Conditions: 750 mM GdnHCl, 10 mM HEPES, 5 mM TCEP, pH 7.2, 0.81 µM Trx-linker octamer (cis), +140 mV (trans), 24 ± 1 °C. Figure reproduced from [116].

It was hypothesized that C-terminus-first translocation occurred when features A were observed and N terminus-first translocation occurred when features B were observed. To test this, two identical concatemers were linked by a disulfide bond between the N-terminal cysteines and recorded under the same conditions, such that the new dimerized concatemers could only enter the pore C-terminus first. In this case, feature

B occurred only after feature A within each trans-location event (Fig. 4.6). Therefore, we assigned these two features as C-terminus-first (A) and N-terminus-first (B) translocation events.

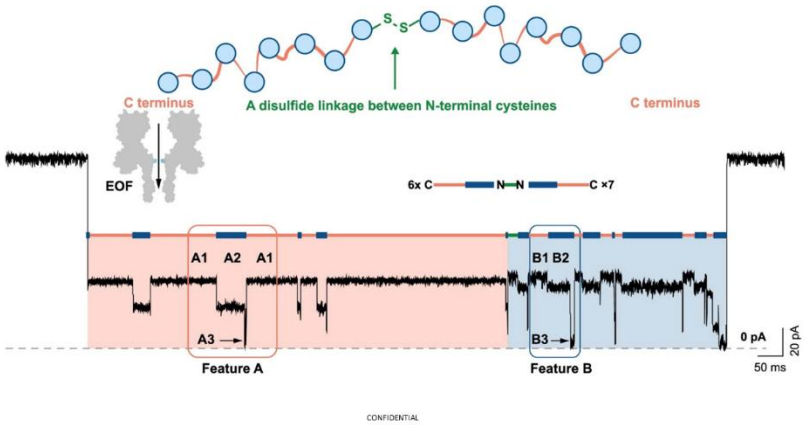


Fig. 4.6: Electroosmosis-driven translocation of dimerized octamer generates combined repeating current features. Without the TCEP pre-treatment, features A were always seen before features B when they occurred together within a single translocation event. The first two levels (B1 and B2) in features B have larger noise and higher $I_{res\%}$ compared with A1 and A2 recorded within a single polypeptide translocation event by the same pore (A1: $I_{res\%}$= 35 ± 1 %, $I_{r.m.s.}$ = 1.1 ± 0.1 pA, N = 25; A2: $I_{res\%}$ = 21 ± 1%, $I_{r.m.s.}$ = 1.5 ± 0.2 pA, N = 25; B1: $I_{res\%}$= 38 ± 1 %, $I_{r.m.s.}$ = 1.7 ± 0.4 pA, N = 39; B2: $I_{res\%}$ = 32 ± 1%, $I_{r.m.s.}$ = 2.0 ± 0.5 pA, N = 39; $I_{r.m.s.}$ values for each level were reported without subtraction of the noise of the pore; number of individual levels from multiple polypeptide translocation events recorded by the same pore are specified). The translocating molecules, which gave sequential A and B features, were assigned as dimers of octamers linked by a disulfide bond between the two N-terminal cysteines. The recorded repeating features are indicated by orange and blue bars. Conditions: 750 mM GdnHCl, 10 mM HEPES, pH 7.2, 0.81 µM Trx-linker octamer (cis), +140 mV (trans), 24 ± 1 °C. All traces were filtered at 2 kHz for clarity; transient A3 levels were truncated by filtering and therefore deviated from ~0 pA. Figure reproduced from [116].

4.2.3 Assessing the influence of the C-terminal linker on capture by the pore

To evaluate the contribution of the C-terminal linker on capture directionality, the ratio of C-terminus-first versus N-terminus-first translocation events was estimated for all concatemers. In the presence of a C-terminal linker (i.e., Trx dimer, tetramer and hexamer), the ratio of C-terminus-first versus N-terminus-first translocation events was ~2:1 at +140 mV (Table 4.3); but surprisingly in the absence of the C-terminal leader sequence (i.e., Trx octamer), the C-terminus-first pattern dominated (C versus N = ~10:1 at +140 mV) (Table 4.3). It is unclear why this trend was observed. The initial 30 amino acid segment originating from the N-terminus exhibits a net charge of -5, whereas a segment of equivalent length commencing from the C-terminus displays a net charge of +1. Assuming an electrophoretic force acting on the terminal ends of the protein to assist in capture, the N-terminus should be more easily captured by the pore, which cannot explain the higher domination of C versus N-first translocation in the octamer compared to the other constructs. This suggests that the initial capture step may be initiated by a folded Trx unit sitting at the top of the pore in an orientation which promotes C-terminal capture.

Table 4.3 Frequency of C terminus-first or N terminus-first translocation events recorded with Trx-linker concatemers[a].

	Voltage (trans)	C terminus-first translocation	N terminus-first translocation	N
Dimer	+140 mV	67%	33%	142
Tetramer	+140 mV	68%	32%	87
Hexamer	+140 mV	68%	32%	196
Octamer	+120 mV	86%	14%	87
	+140 mV	91%	9%	373
	+160 mV	94%	6%	192
	+180 mV	85%	15%	62

[a] Conditions: 750 mM GdnHCl, 10 mM HEPES, pH 7.2, the applied potential (trans) is specified in the table, 24 ± 1 °C. Concatemer concentrations (cis): dimer 2.23 µM, tetramer 0.63 µM, hexamer 0.25 µM, octamer 0.81 µM.

At 750 mM GdnHCl, ~12% of the translocated octamers produced a maximum of seven repeats of feature A following the initial spike (Table 4.4).

Table 4.4: Percentages of concatemer translocation events with N numbers of feature A repeats[a].

	N								Total events
	1	2	3	4	5	6	7	8	
Dimer[b]	44%	56%	/	/	/	/	/	/	73
Tetramer[b]	12%	31%	26%	31%	/	/	/	/	42
Hexamer[b]	14%	43%	20%	14%	3%	6%	/	/	35
Octamer[b]	10%	16%	22%	19%	3%	9%	9%	12%	86
Octamer[c]	11%	20%	29%	13%	9%	4%	7%	7%	45

[a] If the initial A3 level appeared as a spike to ~0 pA and was followed by at least one complete iteration of feature A, it was counted as one repeat.

[b] Conditions: 750 mM GdnHCl, 10 mM HEPES, pH 7.2, +140 mV (trans), 24 ± 1 °C. Concatemer concentrations (cis): dimer 2.23 µM, tetramer 0.63 µM, hexamer 0.25 µM, octamer 0.81 µM.

[c] Conditions: 750 mM KCl, 10 mM HEPES, pH 7.2, 0.81 µM octamer (cis), +140 mV (trans), 24 ± 1 °C.

A kinetic analysis revealed two populations of A3: one had a mean dwell time ~500 times longer than the other at +140 mV (<τA3> = 320 ± 60 ms versus 0.69 ± 0.04 ms) (Table 4.5). The longer-lived A3 (τA3 > 10 ms) was seen in 25% of the final features A

recorded as the translocation of an octamer was completed, but only in 3% of the preceding features A.

Table 4.5: Mean dwell times (<τ>) derived by QuB[a] for the three levels of repeating feature A (A1, A2, A3) recorded during the C-terminus first translocation of Trx-linker octamers through a single (NN_113R)7 nanopore[b].

Voltage (trans)	+140 mV	
$\langle\tau_{A1}\rangle$ / ms	270 ± 20	N = 277
$\langle\tau_{A2}\rangle$ / ms	23 ± 1	N = 277
$\langle\tau_{A3}\rangle$ / ms	320 ± 60 0.69 ± 0.04	N = 40 N = 294

[a] Dwell time analysis was performed by using the maximum interval likelihood algorithm of QuB.

[b] Conditions: 750 mM GdnHCl, 10 mM HEPES, pH 7.2, 24 ± 1 °C.

The dominant absence of a multilevel feature for the first unit and an extended duration for the last unit suggest that the unfolding kinetics of Trx units differ when the polypeptide chain is unable to fully span the lumen of the nanopore.

Alternatively, upon threading of the initial protein unit, the remainder of the concatemer might push against the pore entry driven by the local electroosmotic force, therefore accelerating its translocation. The longer dwell time for level A1 compared to level A2 (more than 10-fold) suggests that the initial unfolding of the protein C-terminus represents a higher energy barrier compared to the subsequent unfolding step of the intermediate, which is in accordance with previous data recorded on C-terminal DNA-derivatized Trx [107].

4.2.4 Chaotropic agents modulate the rate of translocation

Several other buffer conditions were tested in electrophysiology. For example, in the presence of a mild denaturing concentration of urea (recording conditions: 2 M urea, 750 mM KCl, 10 mM HEPES at pH 7.2, +140 mV (*trans*), 24 ± 1 °C) the translocation kinetics were slower than with 750 mM GdnHCl (Fig. 4.7), but faster than at non-denaturing conditions (750 mM KCl). These results were expected according to the denaturation curves obtained for Trx monomer (see previous chapter), which show that 2M urea is further away from the unfolding transition state than 750 mM GdnHCl. At 750 mM KCl, 2M urea, αHL showed a decrease in conductance compared to 750 mM KCl only, which suggests that a conformational change is triggered by this denaturant, an effect not observed at 750 GdnHCl.

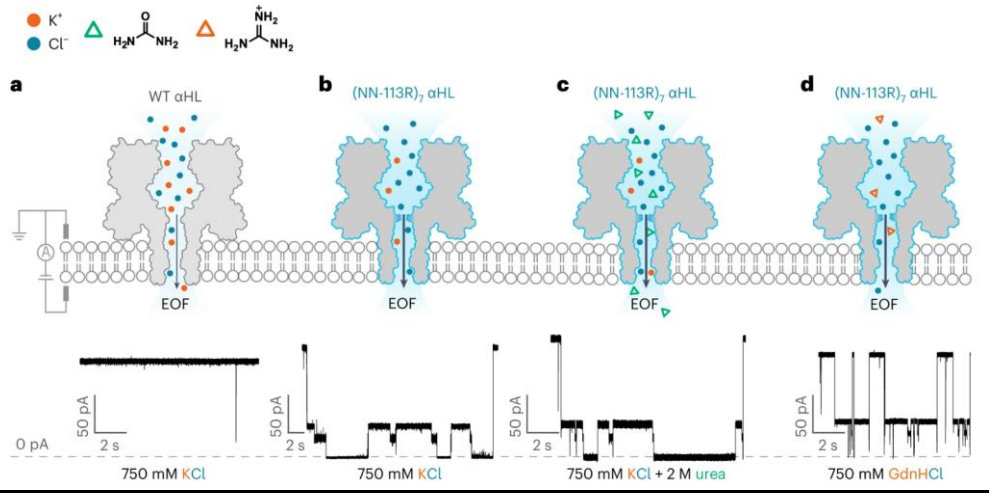


Fig 4.7: Chaotrope-facilitated electro-osmotic translocation of the Trx-linker octamers through a nanopore. **a**, Translocation of Trx-linker octamers through a weakly anion-selective WT αHL was not observed in the absence of a chaotrope. **b**–**d**, Current traces showing the translocation of Trx-linker octamers through the electro-osmotically active nanopore $(NN\text{-}113R)_7$ in the presence of 750 mM KCl (**b**), 750 mM KCl and 2 M urea (**c**) or 750 mM GdnHCl (**d**) with 2 kHz post-acquisition filtering. The use of mild

denaturing concentrations of chaotropic agents (urea and GdnHCl) accelerated the co-translocational unfolding of the Trx units. Conditions: 10 mM HEPES at pH 7.2, 0.81 µM Trx-linker octamer (*cis*), +140 mV (*trans*), 24 ± 1 °C with 750 mM KCl (**a** and **b**); 2 M urea and 750 mM KCl (**c**); 750 mM GdnHCl (**d**). Figure reproduced from [116].

It was reasoned that higher denaturant concentrations would destabilize the tertiary structure of Trx thus accelerating its unfolding and ultimately its translocation time. By doubling the GdnHCl concentration to 1.5 M, the translocation rate was accelerated, and some levels within feature A started to disappear. At 3 M GdnHCl, repeating feature A was completely lost (Fig. 4.8). Interestingly, despite both conditions not being completely denaturing for Trx monomer in solution, upon capture by the pore, the electroosmotic force is strong enough to unfold the protein at a rate below the detection limit of the amplifier (~1 ms), causing a loss in features.

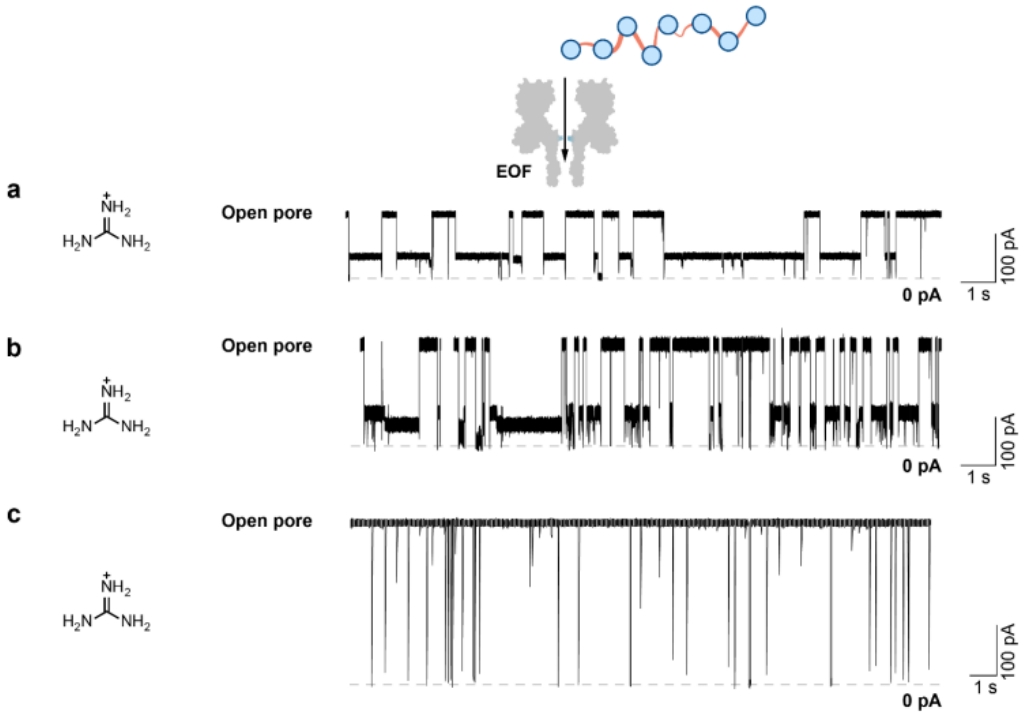


Fig. 4.8: Electroosmosis-driven translocation of Trx-linker octamers through a nanopore. Current traces for the translocation of Trx-linker octamers through a charge-selective nanopore ((NN113R)7) in the presence of 750 mM GdnHCl. **a**, 1.5 M GdnHCl. **b**, 3 M GdnHCl. **c,** without postacquisition filtering. Current features for subunit-by-subunit translocation were lost at 3 M GdnHCl (**c**). The mean number of features A recorded per concatemer is (**a**) ~4, (**b**) ~3, (**c**) 0. Conditions: 10 mM HEPES, pH 7.2, 0.81 μM Trx-linker octamer (cis), +140 mV (trans), 24 ± 1 °C, with (**a**) 750 mM GdnHCl; (**b**) 1.5 M GdnHCl; (**c**) 3 M GdnHCl. Figure reproduced from [116].

Similar repeating features were also seen in the absence of GdnHCl (recording conditions: 750 mM KCl, 10 mM HEPES at pH 7.2, +140 mV(trans), 24 ± 1 °C) or in the presence of mild denaturing concentration of urea (recording conditions: 2 M urea, 750 mM KCl, 10 mM HEPES at pH 7.2, +140 mV (trans), 24 ± 1 °C). Under these conditions, the translocation kinetics were slower (Fig. 4.9), what reduced the throughput, but the duration of each level was higher, what might translate into increased resolution. An inherent trade-off between throughput and resolution appears evident, so fine tuning of denaturing conditions in the recording buffer might be necessary depending on the analyte stability and the kind of information that is needed.

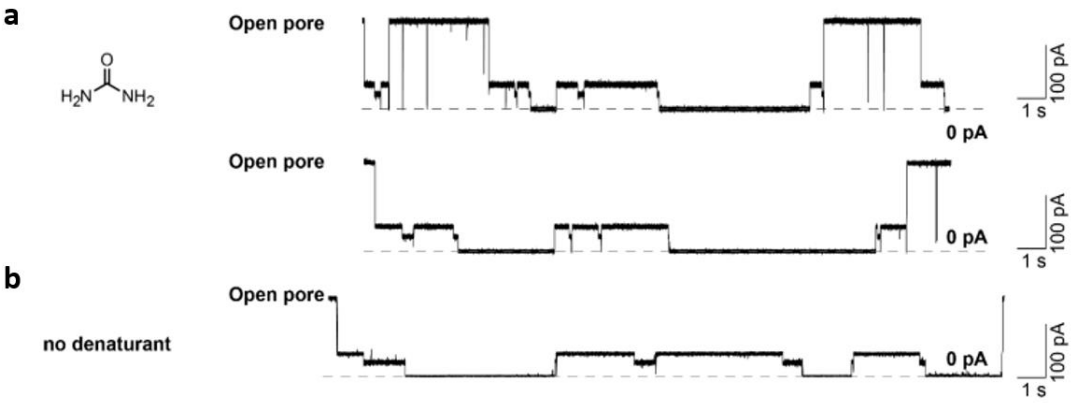


Fig. 4.9: Electroosmosis-driven translocation of Trx-linker octamers through a nanopore. Current traces for the translocation of Trx-linker octamers through a charge-selective nanopore ((NN113R)7) in the presence of 2 M urea (**a**) or no denaturant (**b**) with 2 kHz post-acquisition filtering. The mean number of features A recorded per concatemer is (**a**) ~4, (**b**) ~4. Conditions: 10 mM HEPES, pH 7.2, 0.81 µM Trx-linker octamer (cis), +140 mV (trans), 24 ± 1 °C, with (**a**) 2 M urea and 750 mM KCl; (**b**) 750 mM KCl. Figure reproduced from [116].

5. Detection of PTMs during electro-osmotic translocation

5.1 Assembly and purification of PTM-site Trx monomers

Previous work at the Bayley group had shown successful detection of phosphorylations located close to the C-terminus of Trx [112]. Using a leader Poly cytosine ssDNA oligonucleotide, individual Trx proteins were translocated through the pore following a three-step unfolding mechanism which generated current patterns consisting of three different levels. During the first step, the DNA oligonucleotide leader threaded the pore pulling from the protein C-terminus. A second step followed, during which the C-terminal end of the protein was spontaneously unfolded due to the electrophoretic force acting on it and translocated the pore; this step generated a current level which corresponded to a C-terminal section of the protein threaded inside the pore. Finally, a third and last step occurred when the remainder of the polypeptide spontaneously unfolded and translocated the pore, what caused an almost complete current block level.

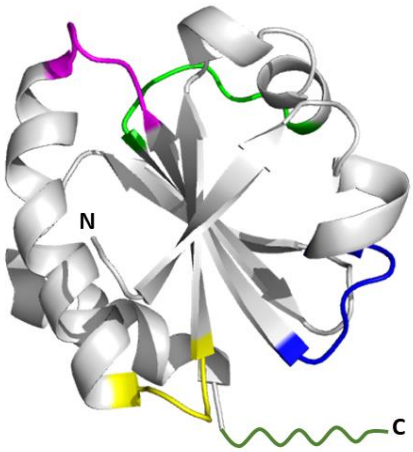
During step 2, it was estimated that the C-terminal section of the protein remained within the pore lumen for long enough that sensing of modifications located within that region would be feasible. A series of mutants were engineered which displayed the RRAS recognition motif for protein kinase A (PKA), which phosphorylates the last serine residue. A RRAS motif was added at position multiple positions nearby the C-terminal

end. Sensing of phoshphorylations within this region was exquisite with a WT αHL pore, proving site specific discrimination in mono-phosphorylated but also di-phosphorylated Trx substrates. However, it was not possible to sense phophorylations further down the protein C-terminus, due to its distance from the sensing region of the pore during this unfolding step, and therefore the sensing capabilities of the approach were limited to a very particular set of terminal amino acids.

In this work, it was sought to determine whether PTMs located anywhere within the Trx sequence could be detected during electro-osmosis-driven translocation.

To explore the sensitivity of the new approach, it was initially decided to generate single Trx-linker mutants containing a "universal" modification site (RRASAC), where a phosphoryl group could be introduced at the serine residue via incubation with PKA, and a neighbouring cystine as chemical handle which could be virtually modified with any PTM mimic. Encouraged by the previous visualization of each individual Trx-linker unit within the concatemers, it was decided to add the C-terminal linker to the monomeric Trx for two reasons: first was that the C-terminal linker was expected to facilitate capture by the pore offering at the same time a control on directionality, and second was that the presence of the linker generated level A1, opening the door for potential sensing within that region.

Six different Trx-linker mutants were engineered via site-directed mutagenesis containing the RRASAC site at six different locations across its length (Fig. 5.1).



SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA

PKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAG

SAGS

Fig. 5.1: Ribbon diagram of the 5 different sites where the RRASAC site was engineered to generate the Trx-linker library. Bottom: Trx-linker sequence with the linker regions highlighted in different colors.

Table 5.1: Sequence of the Trx-linker monomers: The color code for the different engineering sites is maintained (Fig. 5.1). The RRASAC site is highlighted in red.

Trx-linker-28S/30C

SDKIIHLTDDSFDTDVLKADGGSGRRASACGGSGGAILVDFWAEWSGPSKMIAPILDEIADEYQGK

LTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGS

AGSAGSAGSAGSAGSAGS

Trx-linker-40S/42C

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGGSGRRASACGGSGGPSKMIAPILDEIADEYQGK

LTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGS

AGSAGSAGSAGSAGSAGS

Trx-linker-61S/63C

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLGGSGRRASACGGS

GTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGS

AGSAGSAGSAGSAGSAGS

Trx-linker-84S/86C

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA

PKYGIGGSGRRASACGGSGRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGS

AGSAGSAGSAGSAGSAGS

Trx-linker-14S/16C

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA

PKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGRRASACSAGSAGSAGS

AGSAGSAGS

Trx-linker-24S/26C

SDKIIHLTDDSFDTDVLKADSGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGT

APKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSRRASA

CAGSAGSAGS

Two of those locations were located within the linker, and four more at each of the four naturally present loops within Trx (Fig. 5.1). The loops were selected over other positions within Trx sequence, as they were predicted to tolerate a higher degree of modification due to its unstructured nature. To preserve loop flexibility, the RRASAC motif was flanked by a flexible GS sequence and inserted into the loops.

Trx-linker-61S/63C failed to express and was not pursued further, but the rest of the library expressed in high yields and were successfully purified (Fig. 5.2).

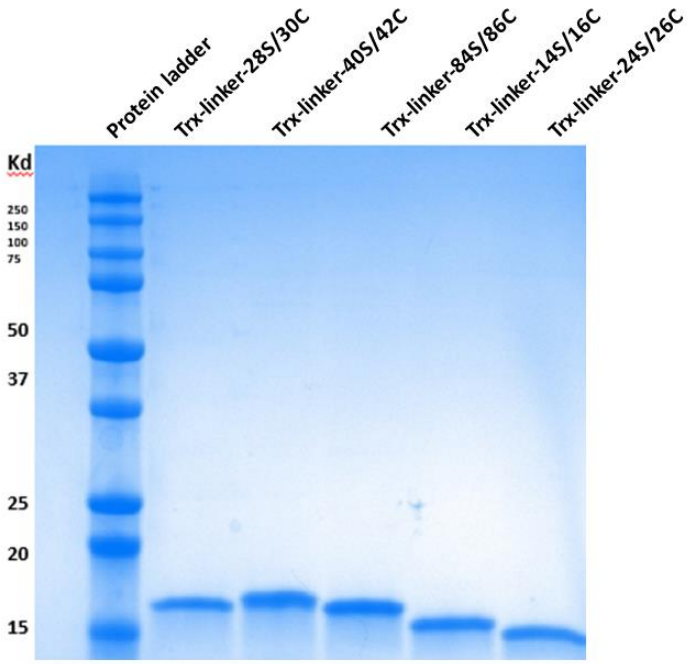


Fig. 5.2: SDS-PAGE of the five Trx-linker mutants after purification.

The mutants were subsequently modified via incubation with PKA for serine phosphorylation. Complete phosphorylation was evidenced by running each protein on a SuperSep™ Phos-tag™ Precast Gel (Fujifilm) (Fig. 5.3), an SDS polyacrylamide gel casted with Phos-tag™, a molecule that can bind phosphate groups with high efficiency while reducing the migration distance of phosphorylated proteins relative to their unphospohorylated variant, and by ESI-MS.

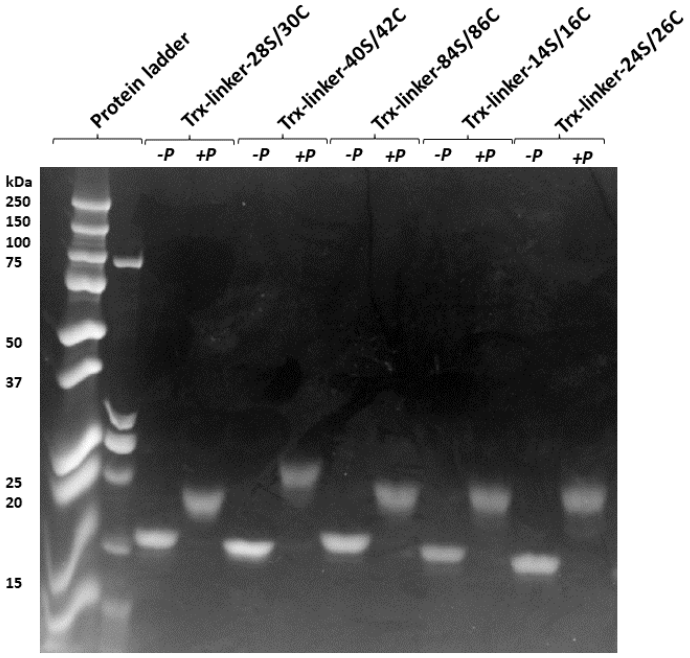


Fig. 5.3: SuperSep™ Phos-tag™ SDS-PAGE of the the five Trx-linker mutants before (-P) and after (+P) phosphorylation with PKA.

Upon electrophysiology characterization, all mutants generated translocation events, but the current profile did not resemble feature A in the concatemers, and none of the three levels (A1, A2, A3) could be identified, which further supports the observation that the unfolding kinetics of Trx units differ when the polypeptide chain is unable to fully span the lumen of the nanopore. Unfortunately, no evidence of phosphorylation sensing could be obtained in any of the five PTM modified constructs when compared to the unmodified version, and characterization of these mutants was not pursued further.

5.2 Assembly and purification of PTM-site Trx concatemers

The translocation fingerprint of each Trx-linker unit within the concatemers appeared to be much more reproducible, showing very clear features containing three levels. It was reasoned that PTM sensing would be more feasible in such construct designs, so it was decided to clone each one of these mutant PTM Trx-linker units in between two wt Trx-linker tetramers. (Fig. 5.4 and Table 5.2).

In this manner, it was sought to determine whether by adding C- and N-terminal flanking sequences, PTMs located in the middle unit of a nonamer could be detected. This new design simultaneously offered the opportunity to show detection of PTMs within a long polypeptide chain.

Table 5.2: Sequences of the thioredoxin-linker concatemers: Blue: Trx; Yellow: linker; Green: modified linker; Red: sequence for modification; Pink: flexible sequence White: restriction enzyme site.

Nonamer (Trx-linker)4(Trx-linker-24S/26C)(Trx-linker)4

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGT
APKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSA
GSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNID
QNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSA
GSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTV
AKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAG
SAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQ
GKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGS

AGSAGSAGSAGSAGSAGSAGRSGTSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPI

LDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA

GSAGSAGSAGSAGSAGSAGSRRASACAGSAGSAGRSPRRSDKIIHLTDDSFDTDVLKADGAILVDF

WAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA

ATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTD

VLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFK

NGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDD

SFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIP

TLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKII

HLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKY

GIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSA

GRS

Nonamer (Trx-linker)4(Trx-linker-14S/16C)(Trx-linker)4

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGT

APKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSA

GSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNID

QNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSA

GSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTV

AKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAG

SAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQ

GKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGS

AGSAGSAGSAGSAGSAGSAGRSGTSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIVPI

LDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA

GSAGSAGSAGRRASACSAGSAGSAGSAGSAGSAGRSPRRSDKIIHLTDDSFDTDVLKADGAILVDF

WAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGA

LSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADG

AILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAA

TKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDV

LKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKN

GEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDS

FDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPT

LLLFKNGEVAATKVGALSKGQLKEFL DANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRS

Nonamer (Trx-linker)4(Trx-linker-84S/86C)(Trx-linker)4

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA

PKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAG

SAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQN

PGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSA

GSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLN

IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSA

GSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVA

KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSA

GSAGSAGSAGSAGRSGTSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEY

QGKLTVAKLNIDQNPGTAPKYGIGGSGRRA**S**A**C**GGSGRGIPTLLLFKNGEVAATKVGALSKGQLKE

FLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSPRRSDKIIHLTDDSFDTDVLKADGAILVD

FWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGA

LSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGA

ILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATK

VGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKA

DGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEV

AATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTD

VLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKN

GEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRS

Nonamer (Trx-linker)4(Trx-linker-28S/30C)(Trx-linker)4

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA

PKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAG

SAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQN

PGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSA

GSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLN

IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSA

GSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVA

KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSA

GSAGSAGSAGSAGRSGTSDKIIHLTDDSFDTDVLKADGGSGRRASACGGSGGAILVDFWAEWSG

PSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEF

LDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSPRRSDKIIHLTDDSFDTDVLKADGAILVDF

WAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGAL

SKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAI

LVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATK

VGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKA

DGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEV

AATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTD

VLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKN

GEVAATKVGALSKGQLKEFL DANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRS

Nonamer (Trx-linker)4(Trx-linker-40S/42C)(Trx-linker)4

SDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTA

PKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAG

SAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQN

PGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSA

GSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLN

IDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSA

GSAGSAGSAGRSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVA

KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSA

GSAGSAGSAGSAGRSGTSDKIIHLTDDSFDTDVLKADGAILVDFWAEWSGGSGRRASACGGSGG

PSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEF

LDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSPRRSDKIIHLTDDSFDTDVLKADGAILVDF

WAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGAL

SKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDVLKAD

GAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVA

ATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSFDTDV

LKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKN

GEVAATKVGALSKGQLKEFLDANLAGSAGSAGSAGSAGSAGSAGSAGSAGRSDKIIHLTDDSF

DTDVLKADGAILVDFWAEWSGPSKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLL

LFKNGEVAATKVGALSKGQLKEFL DANLAGSAGSAGSAGSAGSAGSAGSAGSAGRS

Assembly of the nonamer genes was carried out following the same procedure as with the previous concatemers. A DNA cassette containing two new orthogonal restriction sites was cloned in between two tetramers of WT Trx-linker units. The two new restrictions sites allowed for subsequent cloning of the PTM site-modified Trx-linker unit of choice in between the tretramers (Fig. 5.4).

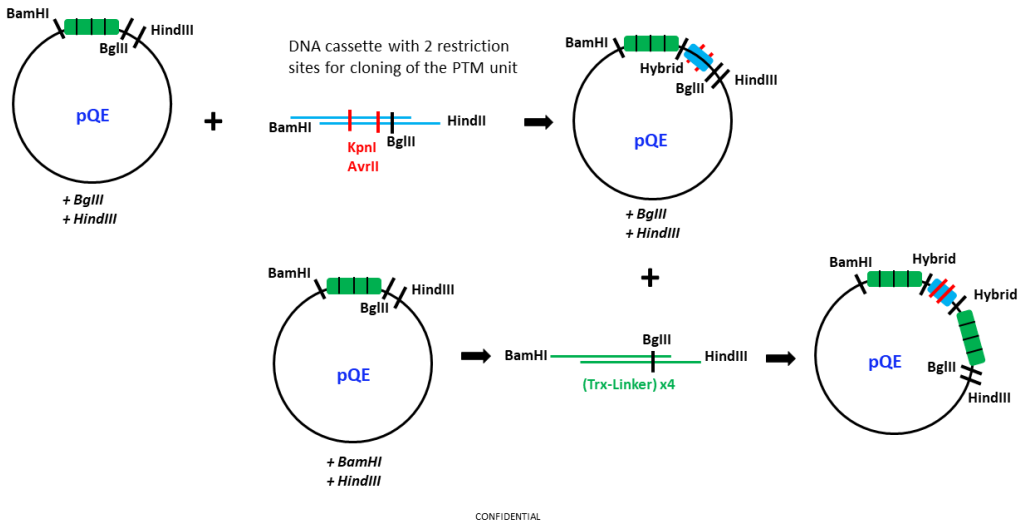


Figure 5.4: Gene assembly of Trx-linker nonamers with middle PTM unit: A pQE plasmid containing a single a tetrameric Trx-linker is opened up upon digestion with BglII and HindIII restriction enzymes. A DNA cassette is designed to contain the two internal orthogonal restriction sites KpnI and AvrII with flanking 5' BamHI and 3' HindIII. Upon digestion with BamHI and BglII, the cassette is cloned downstream the Trx-linker tetramer. Once again, the vector is opened with BglII and HindIII digestion. A second pQE plasmid containing a Trx-linker tetramer is digested with BamHI and HindIII restriction enzymes and the

Trx-Linker tetramer insert purified from the rest of the plasmid. The opened up plasmid and purified insert are mixed and ligated together, generating a Trx-linker octamer with two middle restriction sites KpnI / AvrII available for cloning of the PTM Trx-unit.

5.3 Site-selective modification and detection of PTM-site concatemers modifications

Three types of PTMs were introduced at the RRASAC PTM site of every nonamer: phosphorylation (P) at the serine via incubation with PKA and either glycosylation with a 6' sialyllactosamine (SLN) or glutathionylation (GSH) via cysteine-directed modification.

Modifications located either three of the naturally occurring Trx linkers did not generate any distinguishable signal, however, serine phosphorylation (14S-P or 24S-P) or cysteine-directed glutathionylation or glycosylation (16C-GSH, 26C-GSH, 16C-SLN or 26C-SLN) in the mutants where the PTM site was located within the linker region generated distinguishable new features containing a new level A1. (Fig. 5.5).

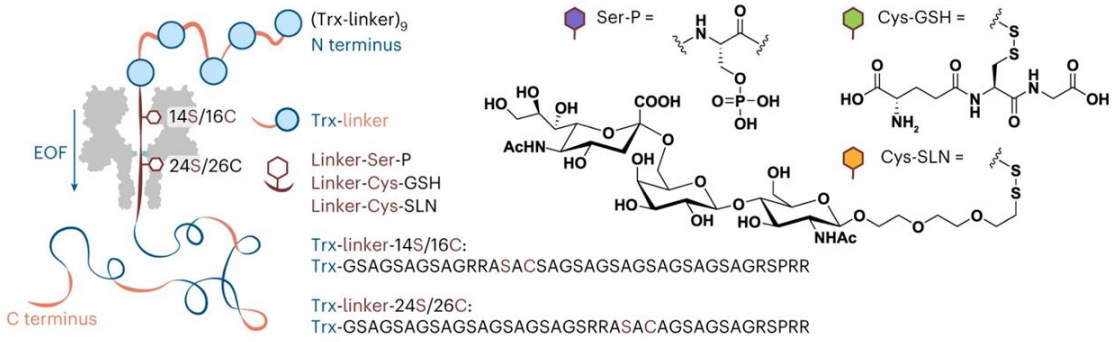


Fig. 5.5: Detection of PTMs in protein concatemers traversing a nanopore driven by EOF. Trx-linker nonamers tested with a charge-selective nanopore ((NN-113R)$_7$) containing an RRASAC sequence within

the central linker, which was post-translationally phosphorylated (purple), S-glutathionylated (green) or glycosylated (yellow). Figure reproduced from [116].

In the presence of a phosphate group (P) or glutathione (GSH) or 6'-sialyllactosamine (SLN), level A1 for the modified units exhibited a smaller $I_{res\%}$ value and higher root-mean-square noise (I_{RMS}) than that of the unmodified segments within an individual polypeptide (Fig. 5.6 and Table 5.3).

Table 5.3 Percentage residual current ($I_{res\%}$) and root-mean-square noise (I_{RMS}) characteristics of individual modifications on Trx-linker nonamers.

	$\Delta I_{res\%}$[a]	I_{RMS} / pA[c]	N
Trx-linker-14S-P	4.4 ± 0.8%	0.96 ± 0.18	19 concatemers 4 separate pores
	4.0 ± 1.2%[b]	1.6 ± 0.9[b]	19 concatemers 4 separate pores
Trx-linker-24S-P	8.3 ± 1.6%	2.0 ± 0.6	27 concatemers 3 separate pores
	9.2 ± 2.1%[b]	2.5 ± 0.9[b]	23 concatemers 3 separate pores
Trx-linker-16C-GSH	5.1 ± 0.9%	0.93 ± 0.19	46 concatemers 4 separate pores
Trx-linker-26C-GSH	8.6 ± 1.3%	1.6 ± 0.2	21 concatemers 3 separate pores
Trx-linker-16C-SLN	15 ± 1%	0.73 ± 0.28	24 concatemers 3 separate pores
Trx-linker-26C-SLN	17 ± 2%	1.7 ± 0.5	55 concatemers 5 separate pores

[a] Airs% = < $I_{res\%}$(A1, Trx-linker)> - $I_{res\%}$(A1, Trx-linker+PTM). For a C terminus-first translocation event, < $I_{res\%}$(A1, Trx-linker) > was determined as the mean $I_{res\%}$ value of the unmodified A1 levels within an individual translocation event. $I_{res\%}$ (A1, Trx-linker+PTM) was determined for the A1 level of the modified linker and appeared once per translocating concatamer. Conditions: 375 mM GdnHCl, 375 mM KCl, 10 mM HEPES, pH 7.2, +140 mV (trans), 24 ± 1 °C.

[b] Conditions: 750 mM GdnHCl, 10 mM HEPES, pH 7.2, +140 mV (trans), 24 ± 1 °C.

[c] Root-mean-square noise values (I_{RMS}) were measured from current traces after an applied post-recording filter at 2 kHz. I_{RMS} was normalised by the noise of each pore ($I_{RMS}^2 = I_{RMS}(A1, Trx\text{-}linker+PTM)^2 - I_{RMS}(open\ pore)^2$). Reproduced from [116].

Furthermore, the average increment in the current blockade was roughly proportional to the mass of the PTM, with the phosphate giving the smallest increment and the trisaccharide giving the largest (Table 5.3).

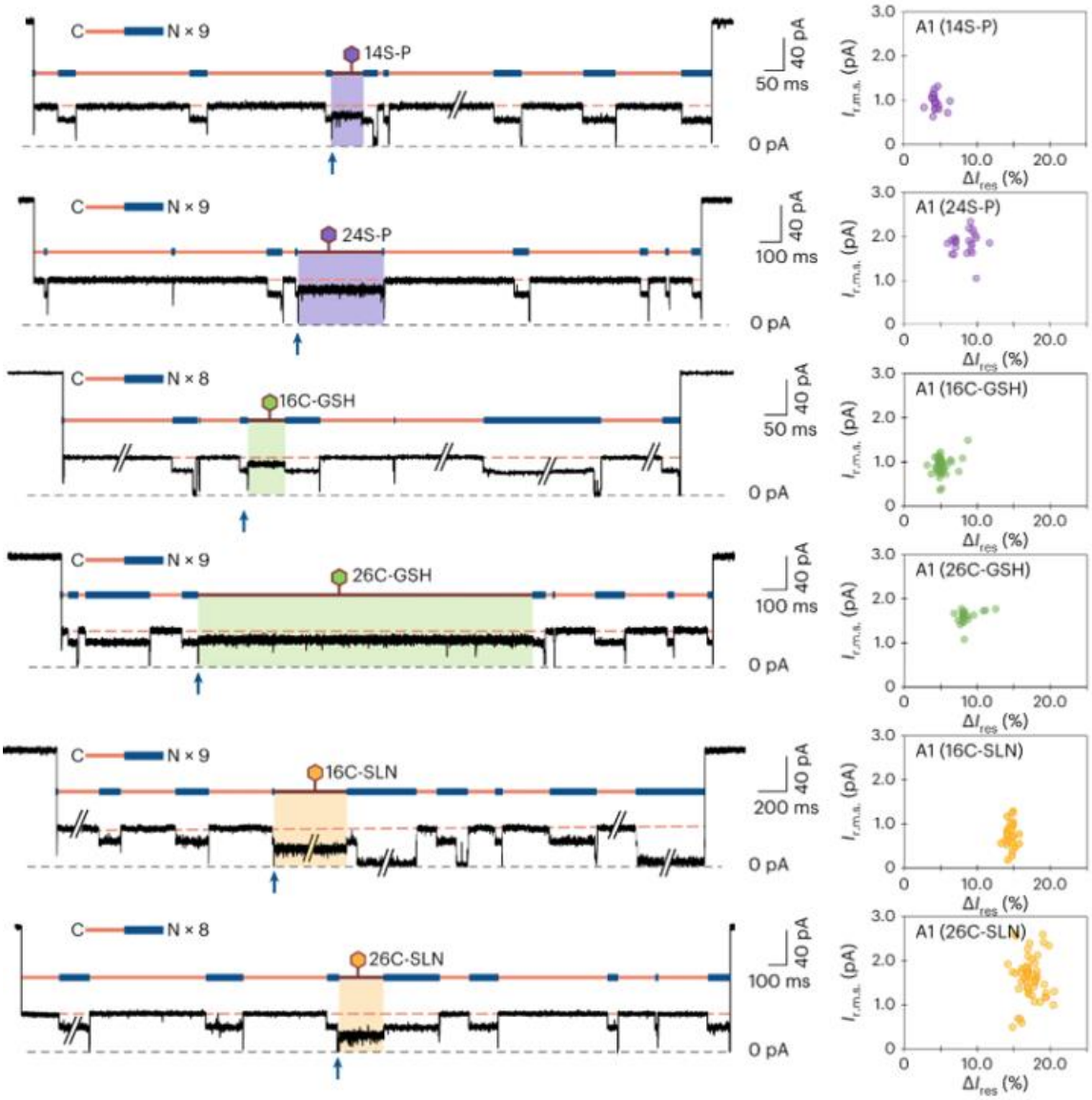


Fig. 5.6: PTM-modified protein concatemers traversing a nanopore show distinct repeating features. Recordings of C-terminus-first translocation events of Trx-linker nonamers (left), showing a distinct level A1 (boxed in purple, green or yellow) in the presence of a PTM compared with the level A1 of unmodified units (orange dash). Traces have been filtered at 2 kHz; transient A3 levels were truncated by filtering and therefore deviate from ~0 pA. The A3 level produced by the translocation of an unmodified unit before the modified linker is indicated with a blue arrow and each of the features A is indicated by orange and blue bars. The number of repeats of feature A within the polypeptide translocation event shown is specified. Scatter plots of $I_{r.m.s.}$ and $\Delta I_{res\%}$ for individual polypeptide translocation events (right), where $\Delta I_{res\%} = \langle I_{res\%}(A1, \text{Trx-linker})\rangle - I_{res\%}(A1, \text{Trx-linker + PTM})$, where $\langle I_{res\%}(A1, \text{Trx-linker})\rangle$ is the mean $I_{res\%}$ value of the remaining A1 levels for unmodified repeat units within an individual translocation event. Conditions, 375 mM GdnHCl, 375 mM KCl, 10 mM HEPES at pH 7.2, 1.2 µM Trx-linker nonamer (*cis*), +140 mV (*trans*), 24 ± 1 °C. Figure reproduced from [116].

Interestingly, despite being able to distinguish each PTM from an unmodified version of the same PTM site, there was substantial overlap between the 14S-P/24S-P and 16C-GSH/26C-GSH populations (Fig. 5.7), which would make identification of these two modifications virtually impossible in a mixture of modified protein analytes. Multiple approaches could make these two modifications more distinguishable. For example, screening of nanopore sensors for better resolution of this particular set of modifications could be done, which would usually require amino acid modifications at the reader head region. Similarly, the addition of chemical modifications or affinity tags towards one of this PTMs could be used to alter its current fingerprint. Namely, the modified concatemer could be incubated with Phos-tag™, which should bind the phosphate group effectively increasing the mass of the modification and presumably causing a higher displacement of ionic current.

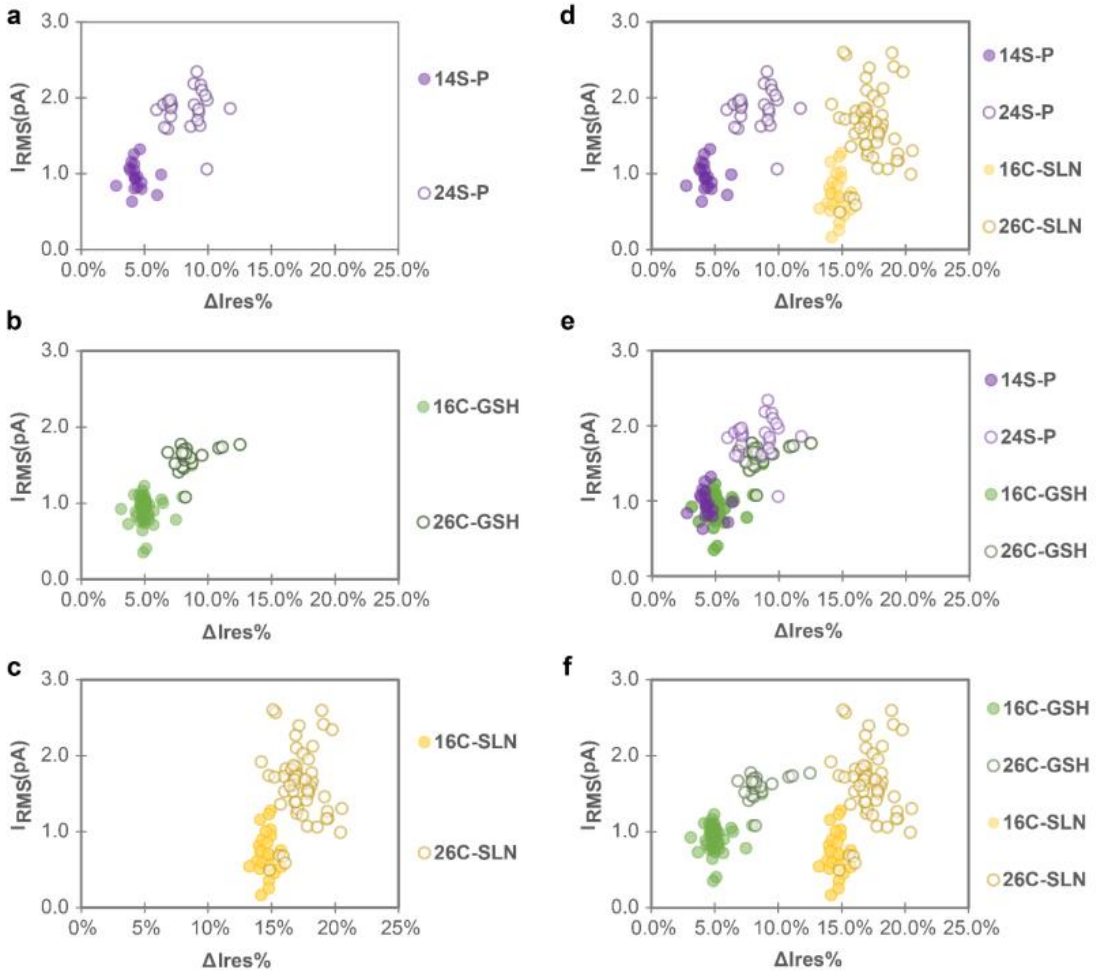


Fig. 5.7: Identification and positional discrimination of PTMs in protein concatemers translocated by electroosmotic flow through a nanopore. Protein nonamers containing a single PTM were tested. **a-c**, Scatter plots of I_{RMS} and $\Delta I_{res\%}$ showing positional discrimination of a phosphorylated serine, a glutathionylated cysteine, or a glycosylated cysteine at sites 10 aa apart ($\Delta I_{res\%}$ = <$I_{res\%}$(A1, Trx-linker)> – $I_{res\%}$(A1, Trxlinker+PTM), where <$I_{res\%}$(A1, Trx-linker)> is the mean $I_{res\%}$ value of A1 levels of an unmodified unit within a single translocation event. Conditions: 375 mM GdnHCl, 375 mM KCl, 10 mM HEPES, pH 7.2, 1.2 µM Trx-linker nonamer (cis), +140 mV (trans), 24 ± 1 °C. **d-f**, Overlaid scatter plots of I_{RMS} and $\Delta I_{res\%}$ showing discrimination between phosphorylated and glycosylated populations, glutathionylated and glycosylated populations, and overlaps between phosphorylated and glutathionylated populations. Figure reproduced from [116].

All the three PTMs tested caused a smaller current blockade at serine 14 (14S) or cysteine 16 (16C) than at serine 24 (24S) or cysteine 26 (26C) (Fig. 5.7). Given that 14S/16C must be closer to the cis opening of the αHL pore than 24S/26C in a C-terminus-first threading configuration, it is probable that the central constriction of the pore is located closer to 24S/26C, what causes a higher current displacement (Fig. 5.8). The findings also suggest that the polypeptide might not be fully extended under the EOF, which corroborates the force spectroscopy data for polypeptides under forces of <20 pN [120], [121].

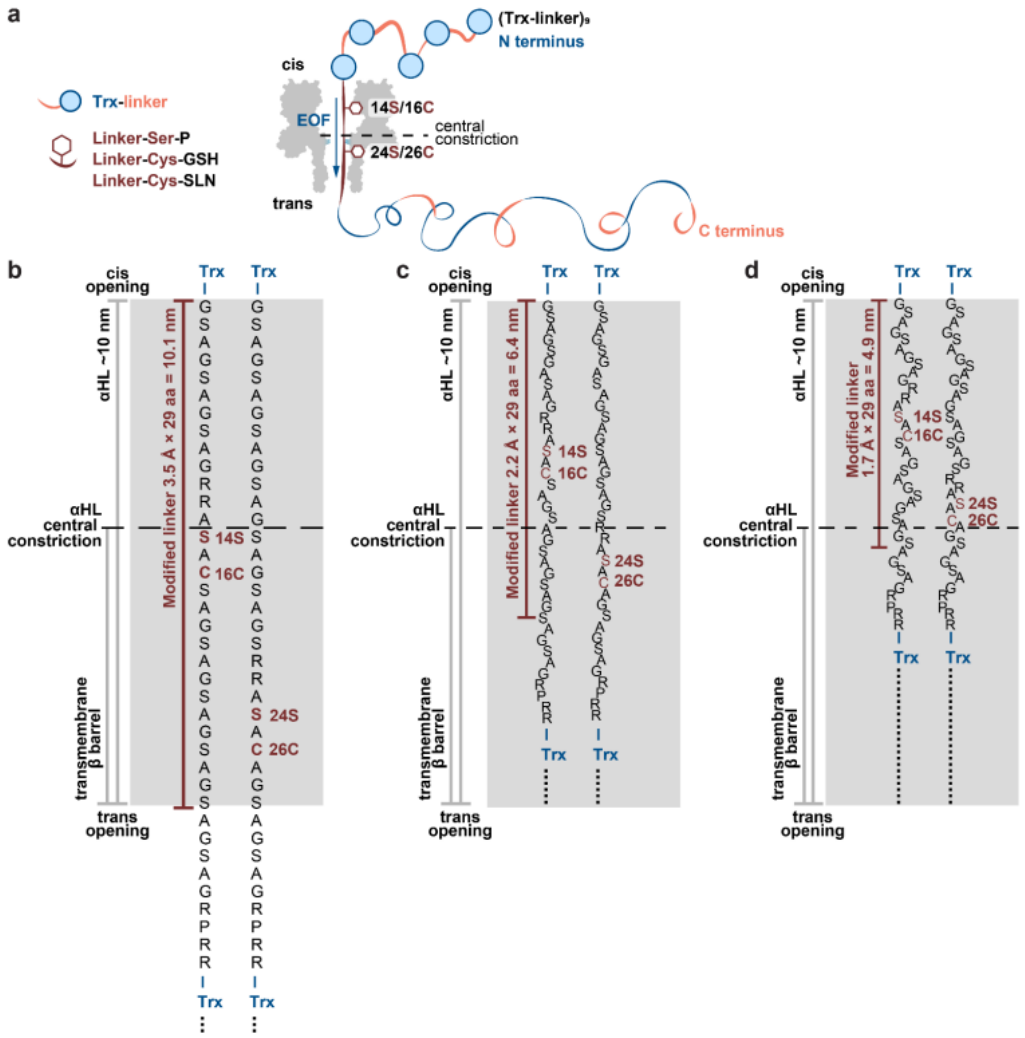


Fig. 5.8: Positions of modification sites during translocation through an αHL pore. **a**, The Trx-linker nonamers contained a RRASAC sequence within the central linker, which was post-translationally modified (hexagon). In a C-terminus-first threading configuration, as shown, the 14S/16C modification sites would be located closer to the cis opening of the αHL pore than the 24S/26C pair, when translocation is paused with a Trx unit at the cis mouth of the pore. **b-d**, depending on the degree of extension of the polypeptide chain under the EOF (3.5 Å per aa when fully extended, 1.7-2.2 Å per aa under ~5-10 pN), the 14S/16C and 24S/26C sites would be located at different positions within an αHL pore. Assuming that the N-terminal residue of the linker is at the cis opening of the pore when the translocation is arrested by a folded Trx unit, the modified linker (red) might fully span the αHL pore (**b**) or occupy only a part of the nanopore (**c,d**). When the 24S/26C sites are located nearer the central constriction of the αHL pore (**c,d**), a PTM at 24S/26C would produce a larger current blockade than that at 14S/16C (PTM = Ser-P, Cys-GSH, Cys-SLN), which is what is observed (Fig. 5.6). Given that the applied potential drops mostly across the transmembrane β -barrel [122],the current difference between 14S/16C+PTM and 24S/26C+PTM is likely to be larger in c than in d. Figure reproduced from [116].

Finally, it was sought to determine whether discrimination of two PTMs from a mixed pool of analytes was possible. Mixed nonamers containing different PTMs at the same site (26C-GSH and 26C-SLN) were tested in the same pore, and two type of events bearing features with distinct level A1 were observed, one corresponding to the previously characterized 26C-GSH and the other one to the 26C-SLN (Fig. 5.9). Although discrimination was done from a simple mix of two different samples which were anticipated to produce two well distinct signals, this experiment proves that discrimination from a more complex mixture of PTM-modified samples should be possible provided there is no overlap of signal parameters among each individual modification.

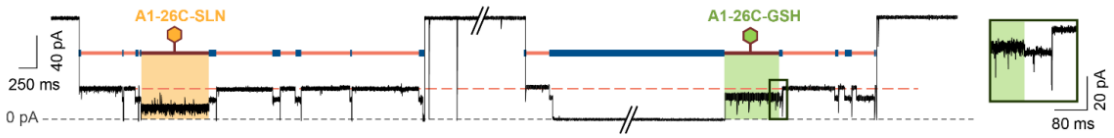


Fig. 5.9: Identification of PTMs in a mixture of two concatemers. Current traces recorded with the same nanopore for three C terminus-first translocations of a mixture of two Trx-linker nonamers containing either a GSH or SLN modification at position 26C within the central linker. The A1 level for the unmodified units (orange dash), and the A1 level for a unit modified with 26C-GSH (A1-26C-GSH, green) or 26-SLN (A1-26C-SLN, yellow) are colour-coded. The inset zooms in on the transition from A1-26C-GSH to A2. Traces have been filtered at 2 kHz. Conditions: 375 mM GdnHCl, 375 mM KCl, 10 mM HEPES, pH 7.2, 1.2 µM Trx-linker nonamer (cis), +140 mV (trans), 24 ± 1 °C. Figure reproduced from [116].

The ability to resolve a specific polymeric analyte, such as DNA or proteins, through a nanopore is contingent upon the range of electrical currents produced during its passage. It is within this current range that the various peptides or bases must be distinguished. In the case of proteins, the substantial number of constituent building blocks, coupled with the diverse post-translational modifications (PTMs), necessitates the resolution of numerous distinct signals within a relatively narrow current range.

The contribution to this signal is influenced by all the amino acids present within the nanopore at any given moment, and potentially those near its entrance, resulting in sequence segments that yield conductance values statistically indistinguishable from one another.

As previously discussed, the use of recognition tags and chemical modifications of certain PTMs may aid in identification. However, implementing this approach with PTM-

modified protein analytes in practical applications remains a challenge, given the considerable quantity and variety of modifications that can coexist within the same region. The primary obstacle to enhancing the discriminatory capability of this method resides in the limited current range that is accessible for distinguishing a large number of polypeptide segments.

In the particular case of this work, it has been estimated that αHL can accommodate sequence segments of ~30 residues, depending on the degree of stretching. The open channel current level, which reflects the total available current for resolution, was of ~ 100 pA under the conditions used (750 mM GdnHCl, 10 mM HEPES, pH 7.2, +140 mV). Current levels generated by the linker region halted in the pore was of ~35 pA (A1), followed by a second stall of a Trx unfolding intermediate of ~24 Pa (A2), and a final translocation step of ~ 2 Pa (A3). This makes the current range for this particular pore and conditions of ~33 pAs (highest current level minus the lowest current level for this particular analyte).

It was estimated that during level A1, the linker region was in a semi-extended state of ~2.2 Å per aa (3.5 Å per amino acid if fully extended). In general, a higher level of peptide stretching will produce a lower current blockade, so an increase in peptide stretching by means of an increased EOF or analyte net charge (higher EPF) could increase the current level leading to a higher total current range. An increase in EOF could be achieved, for example, by using higher concentrations of chaotropes such as GdnHCl. It has been modelled that the guanidine cation is able to bind the walls of the nanopore lumen increasing the local negative charge and therefore making the channel more

anion selective [93], which can be combined with pore engineering of ion-selective pores [101]to generate strong EOFs. Increasing polypeptide net charge could be achieved by means of chemical modification of some of their side chains.

As a rough example, assuming at least 1pA of current difference is required for accurate discrimination, only ~33 species could be uniquely identified within each ~30 aa segment if the current range is of ~33 pAs. This may seem like a modest number compared to the ~400 PTMs that can be found in proteins, but most proteins do not display that much variability throughout their sequence. For example, the first 300 residues of p53, a ubiquitous transcription factor and major tumour suppressor protein with a high level of post-translational processing, could theoretically be characterized in ~30 amino acid segments if up to 63 different signals could be decoded (Fig. 5.10). This number is as low as 3 for other segments of the protein. For most proteins, these numbers are expected to be close to the latter, since they are not as heavily involved in regulatory processes and therefore PTMs are not that abundant. In any case, pores can be engineered to display higher current range for a particular analyte.

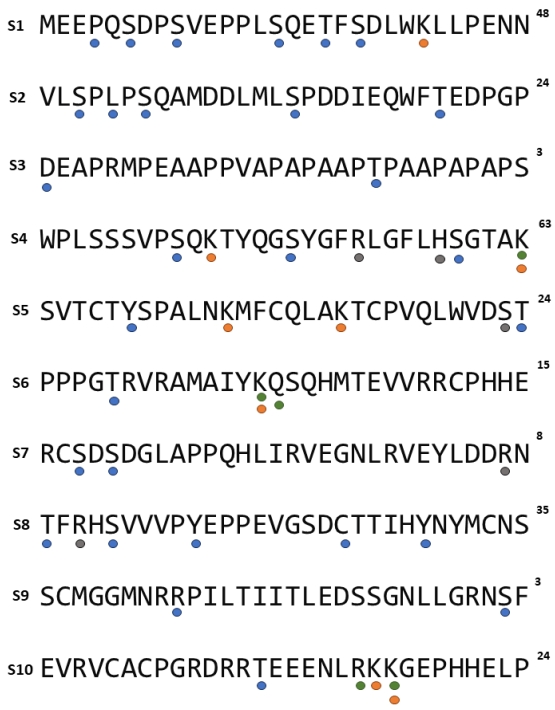


Fig 5.10: Segmented sequence of human p53 showing all reported PTMs (Phophosite.org database). 30 amino acids are shown per row. Phosphorylation is indicated in blue, ubiquitylation in orange, acetylation in green and mono-acetylation in grey. The number of possible protein species (all PTM combinations) is indicated on the right of each segment ($(n)^2$-1), where n is the number of modifications within each segment.

Another very interesting approach to overcome the limitations of a narrow current range, could be the use of time-varying cross membrane voltages to modulate the stretching level of a protein segment as it is transversing the pore [123]. As shown by prof. Gundlach and colleagues, varying the transmembrane voltage between 100 to 200 mV with a high frequency (200 Hz) can modulate the level of stretching of a translocating DNA strand in between helicase steps, what has demonstrated to increase the single-passage de novo base-calling accuracy from 62.7 to 79.3%. During each

helicase discrete step, what would resemble each one of the protein stalls (generating levels A1, A2, A3) in the system described in this work, the voltage variation can modulate the level of stretching of a given segment within the pore thus generating modulations in conductance signal that can be used to differentiate two sequences with nearly identical conductance values in a constant-voltage regime (Fig. 5.11).

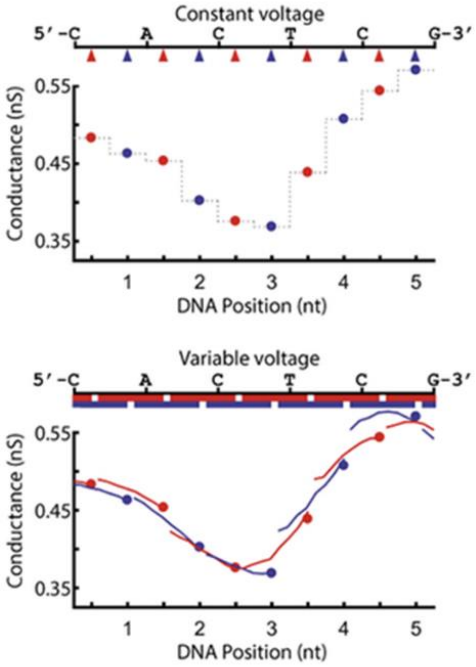


Fig. 5.11: Basic principles of constant-voltage versus variable-voltage DNA nanopore sequencing. In constant-voltage sequencing (top), the conductance is sampled at discrete positions controlled by a processive enzyme (red and blue arrows) and the resulting signal is a series of mean conductance values. In the variable-voltage regime (bottom), conductance is sampled continuously during each discrete step generating curved overlapping conductance values that contain a higher level of information. Figure reproduced from [123].

In a label-free, EOF-driven translocation approach of proteins, a transient increase in voltage could result in shorter translocation times, so fine tuning of the voltage range

and frequency would be necessary, as there will be an inherent trade-off between dwell time and level of stretching. Furthermore, discrimination of PTMs in real proteins from cells will require a method with a high dynamic range of sensitivity, due to the low abundance of some species, for what optimizing single-passage *de novo* accuracy will play a key role. Utilizing variable-voltage regimes proved to be essential in reaching this goal.

It is unlikely that a single pore will display all the ideal features to decode PTMs on any protein analyte, while it is more likely that an array of different pore variants will need to be utilized to characterize protein mixtures. Recording conditions such as temperature, concentration of denaturant, salt, etc, will also need to be fine-tuned to accommodate the requirements of a particular protein analyte, so it is likely that several different conditions will need to be utilized to expand the resolution limits. As an example, in this work phosphorylation resolution was proved to be different at the two GdnHCl concentrations tested. At 750 mM GdnHCl, $I_{res\%}$ of the modified A1 was 4.4 ± 0.8 for P-14S and 8.3 ± 1.6 for P-24S ($\Delta I_{res\%}$ = 3.9), while at 375 mM GdnHCl, 375 mM KCl, these values were 4 ± 1.2 for P-14S and 9.2 ± 2.1 for P-24S ($\Delta I_{res\%}$ = 5.2). This proves that at the higher regime of GdnHCl, the resolution between these two modifications was better.

In summary, it has been established that electro-osmotically active nanopores possess the capability to capture and sequentially unfold individual proteins, remarkably those composed of lengthy polypeptide chains exceeding 1,200 amino acids, facilitating the

identification and localization of post-translational modifications (PTMs). Fundamentally, the electro-osmotic force acting on a polypeptide remains relatively constant during its translocation, imparting a unidirectional bias preventing back-and-forth movement. In the absence of an EOF, the movement of the polypeptide across the pore would be governed by diffusion, with the overall time required for the unforced translocation of a polypeptide scaling roughly as the square of its length [124], which would render the method unfit for the sequential analysis of PTMs.

Furthermore, this approach, being label-free, eliminates the need to derivatize proteins at either the N or C terminus to facilitate electrophoretic translocation. This is advantageous, particularly for eukaryotic proteins, where challenges arise due to the widespread presence of N-acetylation and the lack of efficient N- or C-terminus-specific chemistries.

It has been shown successful location of PTMs deep within lengthy polyprotein chains by taking advantage of the stepwise unfolding. These promising results serve as a foundation for constructing comprehensive catalogs of underivatized full-length proteoforms derived from cells and tissues.

6. Final remarks

This work has demonstrated that proteins can be read on a single-molecule nanopore platform in a label-free, enzyme-free fashion. Taking advantage of the natural folded state of proteins, multiple segments within a protein can be accurately scanned during translocation, as the polypeptide gets momentarily stalled within the pore lumen, and modification within those regions finely characterized. The number of steps a protein takes to unfold upon being pulled through a nanopore is characteristic of its own structure, providing useful information that may help distinguishing it from other protein species with a different unfolding pathway. Furthermore, the two ends of a protein can lead the translocation event (i.e., C- or N-terminus), doubling the amount of information that can be extracted from a single polypeptide. This may also potentially increase the number of regions that can be accurately scanned, gaining access to more information on the sequence.

The main advantage of the method presented lays in its simplicity, as 1) no external enzyme motors are required to drive protein translocation, and 2) because EOF is able of driving the polypeptide concatemer passage through the pore irrespective of sequence charge, no prior sample processing to attach leader charged tags is required. Since it is an enzyme-less approach, no processivity or tight control in translocation was achieved, which would ultimately allow for true “amino-calling” or sequencing, but key information could be still gained due to the exquisite resolution of the nanopore sensor,

which accurately scanned the polypeptide segment occupying the pore during the unfolding stalls. Detection and discrimination among multiple post-translation modifications (phosphorylation, glutathionylation and glycosylation) at specific sites has been demonstrated.

Despite being unsuitable for de novo sequencing, this method can still be envisioned as a powerful fingerprinting approach, whereby partial information on a polypeptide sequence can be obtained, which may be enough to identify the protein by using a reference database. As a future endeavour, this method should be tested on natural protein products, to prove usefulness on “real proteins” containing naturally occurring PTMs. Ideally, these protein substrates should be extracted from individual cells, enabling an early form of single-cell proteomics. A simple enough protein product containing only a limited number of naturally occurring PTMs could be enriched from a mammalian cell lysate using specific antibodies and then tested on a nanopore platform. To deal with potential low concentration issues, a pair of droplets that form interface bilayers could be employed to encapsulate the protein mixture and then, the nanopore sensor would be inserted into the bilayer connecting both droplet contents and allowing electrophysiology recordings. The use of this system would solve excessive dilution that may happen on traditional electrophysiology chambers (0.5 – 1 mL working volume), decreasing the working volume down to the pL range. Alternatively, using multiplexed flow cells, such as Oxford Nanopore Technologies’s MinION, may also solve the low analyte concentration issue, as they compensate their a fairly large cis volume (~75 uL), with a high number of nanopores per flow-cell (2040 channels). Additionally, a tethering system is routinely utilized to increase the local concentration of analytes

nearby the pore sensors. For example, the analytes can be modified with oligonucleotide strands functionalized with cholesterol moieties that can insert into the lipid bilayer [125].

In addition to this, the protein of interest should be identifiable within complex mixtures, that is, when other protein species and even other biomolecules present in cell debris and bodily fluids, (i.e., carbohydrates, lipids, etc) are also in solution. A simple and preliminary way to test the performance of this platform under such conditions could be by mixing previously purified Trx protein concatemers with different bodily fluids (urine, blood) or simply adding the diluted E. coli cell lysate containing the over expressed concatemer to the recording chamber, prior to further processing. Because of the repetitive nature of the electrical signals generated by concatemers upon translocation, it is expected that these events will be easily distinguished from any other analyte interacting with the pore.

Importantly, the method proposed proves how useful EOF can be for controlling polypeptide movement and stretching inside the nanopore sensor. Irrespective of protein charge (concatemer substrates lack any heavily charged polypeptide regions or tags), the water flux generated by a moderate ion-selective pore was able to capture an end, unfold, and translocate proteins though the pore at a sufficiently slow pace that allows sensing. Furthermore, the approach described takes advantage of protein tertiary structure to slow down translocation and gain information on the unfolding pathway, which further enriches information on protein identity. Strong denaturing conditions and/or charged tags have been extensively employed for driving protein

passage across pores [89], [90], [93], [103], which with few exceptions [112], disregard the usefulness of tertiary structure of proteins as a mean to control translocation rate. To the best of the author knowledge, there are no previously reported works where EOF alone is employed to drive a protein through a nanopore allowing detection and identification of such protein substrates, let alone detection of PTMs.

The concatemers used in this work stalled at two different positions within the αHL lumen, generating the feature pattern described in previous chapters. The first stall, corresponding to a Trx unit sitting on top of the pore with the preceding linker threading, generated current level A1. A second stall occurred upon the C-terminal end of Trx spontaneously unfolding and the N-terminus section of the protein sitting on top of the pore with a middle segment of the protein occupying the pore, generating current level A2. Albeit not originating from a stall, a current level A3 was also generated upon translocation of the reminder N-terminus of the protein. Modifications within this protein region may also be potentially detected, but on the EOF engineered αHL sensor chosen, this level showed a very low net current (~1-5 pA), which made discrimination almost impossible.

In principle, the two polypeptide segments threading the pore during the stalls have a sufficiently long-lived dwell time (tens of ms) that should allow discrimination of modifications within that region. Unfortunately, sensing of PTMs was only successful at the linker region, but not at any of the 3 selected protein internal modification sites. One of those internal PTM sites, the closest to Trx C-terminus (positions 84S/86C), was chosen because that region was predicted to be stalled within the pore during current

level 2. However, it is possible that despite this being the case, the modifications sit far from the reader head of the pore, and therefore fall outside the detection limit.

A possible way to overcome the detection problem, could be by re-engineering the PTM site at different positions within Trx sequence, that sit closer to the pore reader head when stalled. Alternatively, the number of modifications within the same segment could be increased (e.g., adding two contiguous phosphorylation sites) or they could be made bulkier, by chemically modifying the PTM groups or by using tags that bind those modifications, or a combination of both.

For example, a tag that specifically binds phosphate groups on serine, tyrosine, threonine residues, Phos-Tag^{TM}, could be incubated with the protein of interest prior to electrophysiology measurements, in order to increase the bulkiness of phosphate groups, therefore displacing a higher level of current when occupying the pore lumen. Similar tags could be designed to bind other PTMs specifically.

As an example of a chemical modification of a PTM, one could take advantage of the formation of dehydroalanine (Dha) via elimitation of pSer in peptides and proteins. Dha is a naturally occurring amino acid formed during lanthipeptide natural product biosynthesis in prokaryotes, excretion of phosphothreonine lyases (OspF, SpvC and HopAI1) in pathogenic bacteria, and by spontaneous non-enzymatic elimination of pSer as a consequence of aging in human cells, a process which can be accelerated chemically [126], [127].

Dha has been widely employed as a robust, flexible and site-selective moiety for the in vitro insertion of labels, cargoes, PTMs, and other modification in proteins, or as an activity-based probe on its own. Dha can either be introduced either via an unnatural amino acid translation system or by chemical modification of multiple precursors such as cysteine, selenocysteine, and phosphoserine. In this case, an inverse approach is proposed, in which naturally occurring serine phosphorylations are used as precursors to introduce Dha, which can subsequently be modified with a chemical group easier to detect. A simple and "protein friendly" protocol has been described whereby pSer can initially be transformed into Dha via dephosphorylation after incubation with barium hydroxide at ambient temperature. Multiple chemical groups can then be coupled to Dha through nucleophilic addition (Thia-Michael addition, Aza-Michael addition, Selena-Michael addition) or radical addition, which can facilitate phosphorylation detection[126], [128], [129].

Finding a more suitable nanopore sensor for the detection of specific PTMs is another possible solution. In this work, the spectrum of protein pores explored has been limited to just αHL, but other structurally varied pores may offer a larger sensing region (i.e., a longer lumen that can accommodate and sense a longer peptide stretch), higher resolution (e.g., narrower reader head, optimized amino acids around it, etc), higher signal-to-noise ratio, etc. Establishing a platform for nanopore screening towards a particular analyte or modification, will likely be essential for successful discrimination of PTMs within more complex samples such as natural protein products with numerous, contiguous PTMs and varied structures. One of the main advantages of single-molecule nanopore sequencing platforms is the possibility of utilizing different protein nanopores

to interrogate samples, which can combine the resolution power of different sensors into one instrument simultaneously.

Another issue encountered with the method proposed, is the low throughput. Several hours of recording were necessary in order to observe just a few tens of squiggles corresponding to concatemer translocating events. The main limitations were low capture rate, stalling of the concatemer inside the pore, gating of the pore, and missed protein units during translocation due to spontaneous or rapid unfolding. The first three may be modestly improved by tweaking the buffer composition and finding a better suited nanopore, which exhibits enhanced EOF and less gating propensity. The latter could be improved by removing denaturant from the ionic buffer, a decision that would simultaneously necessitate of enhanced ion-selective pores to prevent stalling and slow translocation. Nevertheless, a true high-throughput method would require multiplexing an array of nanopore sensors, which is currently only possible with the flow cells commercialized by Oxford Nanopore Technologies. For example, their popular MinION flow cell, is equipped with 2040 channels, each of which can incorporate a nanopore of choice. Using such platform, the recording time would be reduced by 3 orders of magnitude. Multiple flow cells can also be used in parallel on the same instrument, which would further increase the throughput potential. For example, GridION and PromethION sequencing devices can accommodate up to 5 and 48 flow cells simultaneously.

Because of the single-molecule nature of nanopore technology, often nanopore sensors are prone to generate current squiggles that do not correspond to the analyte of

interest interacting with it, but rather artifacts such as transient gating states, or other sample contaminants that interact with the pore (i.e., any molecule which analysis is not seek). Individual polypeptides can also generate featureless events due to the high forces exerted on them upon being captured by the pore, at least under the conditions described in this work. The first polypeptide unit translocating the pore usually only generates a current spike, and single polypeptide analytes, despite being structurally similar to any of the units within a concatemer, do not generate the same current pattern.

One potential way in overcoming these issues, might be by employing a protein concatemer like the ones described in this work, or a similar construct, as a barcode attached to a protein of interest. Upon being incorporated to the protein analyte, this barcode, because of its repetitive structure, would be distinctive from any other free molecule in the sample mixture or artifact, providing at the same time a leader sequence that facilitates translocation.

Within the context of tracking biological activities, such as genetic regulation, there would be another major advantage in using a concatemeric barcode. Because of its three-dimensional structure, such reporter may encode much more information than a linear one. In fact, a reduced number of uniquely addressable reporters that can be used together while sharing a common readout is a current limitation in the field [130], [131], [132]. Most protein reporters rely on fluorescent dyes which are limited by the overlapping spectral properties. Others, contain a linear peptide segment that expand the pore reading region (~20 amino acids for most protein nanopores), but they are

theoretically limited by the number of combinations that can be encoded within that short peptide, and in practice, by the indistinguishable signals that chemically close peptide combinations generate [133].

A concatemeric protein construct would provide at least as many barcodes as existing protein repetitions constitute them, where theoretically up to ~20 position would be available for barcode encoding, or 20^{20} possible amino acid combinations. On top of that, each polypeptide within the concatemer, could potentially provide additional barcoding regions, one per unfolding step. In the case of Trx, two barcoding regions could potentially be encoded within its sequence, corresponding to the two steps in which the polypeptide unfolds upon nanopore translocation. Ultimately, barcode size and complexity would be limited by the length of the sensing region and accuracy of the nanopore sensor, and by the degree of tolerability the reporter protein displays towards mutations within its sequence without losing its folding capacity.

One last advantage can be envisioned with the use of concatemeric protein reporters, and that is the possibility of engineering directional barcodes, i.e., that exploit the differential C to N-terminal unfolding pathway. Albeit technically complex, following this idea, a barcode could be engineered in a region only available for being read when the protein was unfolded in a particular direction (e.g., C-terminus-first), and an additional "silent" reporter could be the placed in a region only accessible when the protein was unfolded in the opposite direction.

7. Materials and methods

Expression and purification of αHL pores

αHL monomers were produced by expression in an IVTT system and oligomerized into αHL heptamers by incubation with rabbit red blood cell membranes. The heptameric αHL pores were purified by SDS-polyacrylamide gel electrophoresis.

Construction of Trx-linker concatemer genes

V2-S76 mutant was generated using site-directed mutagenesis using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs).

Expression and purification of Trx monomers

Trx V2-C109 and V2-S76 genes were contained in the pET 30a+ plasmid. The plasmid was transformed into *E. coli* BL21(D3) cells (Novagen) and grown in a Luria broth medium supplemented with kanamycin (50 μg ml^{-1}) at 37 °C with continuous shaking (250 r.p.m.). Protein expression was induced in the exponential growth phase (OD_{600} = 0.6) with isopropyl-β-D-1-thiogalactopyranoside (0.4 mM final concentration). After 6 h, the cells were harvested by centrifugation (10 min, 5,000×*g*), resuspended in lysis buffer (30 mM Tris HCl, 1 mM EDTA, pH 8.3) supplemented with a protease inhibitor cocktail (cOmplete, EDTA free; Roche) and lysed by sonication. Cell debris was removed by centrifugation at 20,000×*g* for 45 min, and the supernatant was treated with a decanting 10% solution of Streptomycin Sulfate, with continuous shaking, at 4 °C. After a 2 h incubation, the solution was centrifuged at 20,000×*g* for

45 min, filtered through a hydrophilic PVDF membrane of 0.22 µm pore size, and loaded onto a Superdex® 75 10/300 GL column. Fractions containing the protein were pooled together and loaded onto HiTrap Q FF anion exchange column. Protein elution was done in a 0-1 M KCl gradient over 10 column volumes. The fractions containing the protein were pooled, aliquoted and flash frozen for storage at −80 °C.

To remove pyruvate from the N-terminal cysteine of Trx V2-S1C/C109, the protein was buffer exchanged into the deprotection buffer (100 mM sodium phosphate, 150 mM NaCl, 400 mM methoxyamine, 5 mM TCEP, pH 7.0) using an Amicon Ultra filter column (0.5 mL, 3 kDa) and incubated at room temperature overnight. The excess methoxyamine was removed, and Trx was buffer exchanged intoi 10 mM TrisHCl, 1 mM EDTA, pH 8.0 by using PD 10 desalting columns (Cytiva).

Construction of Trx-linker concatemer genes

All the reagents were purchased from New England Biolabs (NEB) and DNA oligonucleotides were obtained from Integrated DNA Technologies, unless otherwise indicated. Trx-linker concatemer genes were prepared as previously described[29]. Briefly, the Trx-linker monomer gene was amplified with a 5’ primer containing a BamHI restriction site and a 3’ primer containing a BglII restriction site, which permitted the in-frame cloning of the monomer into the vector pQE30 (QIAGEN). Synthetic genes encoding the concatemers were then constructed by the iterative cloning of monomer into monomer, dimer into dimer and tetramer into tetramer. To aid purification, an N-terminal SUMO tag was inserted between the His6 tag and the first monomer unit. In addition, N-terminal cysteine-glycine codons were included to

give the final concatemer constructs: His6-SUMO-CysGly-(Trx-linker)$_n$ (n = 2, 4, 6) and His6-SUMO-CysGly-(Trx-linker)$_7$Trx.

To produce Trx-linker nonamers (His6-SUMO-(Trx-linker)$_n$, n = 9) containing a modification site, the N-terminal cysteine-glycine codons were removed from the tetramer gene and a DNA cassette was designed to contain two terminal restriction sites (BamHI and BglII) and two internal restriction sites (KpnI and AvrII) (5′-p GATCCGGTACCGGCGGTCCTAGG AGATCTGGCGGTA-3′ and 5′-p GCCATGGCCGCCAGGATCCTCTAGACCGCCATTCGA-3′). Using the interactive cloning strategy described above, a 'cloneable' Trx-linker octamer gene was assembled with the DNA cassette as the middle unit flanked by two Trx-linker tetramer genes (that is, the final construct is His6-SUMO-(Trx-linker)$_4$-KpnI-AvrII-(Trx-linker)$_4$). A Trx-linker monomer mutant gene encoding an RRASAC peptide motif was created by site-directed insertion (forward primer, 5′- AGCGCCTGCGCGGGTTCTGCTGGTTCC-3′; reverse primer, 5′-CGCACGGCG GCTCCCTGCACTTCCGGC-3′) and subsequently cloned in between the KpnI and AvrII sites within the Trx-linker octamer to give (Trx-linker)$_4$-Trx-linker(RRASAC)-(Trx-linker)$_4$. The placement of a single correctly oriented insert was confirmed by sequencing using primers targeting the KpnI and AvrII ligation sites (forward primer, 5′-TGCGAGCGCCTGCGGTGG-3′; reverse primer, 5′-ACGCTCGCGGACGCCACC-3′).

Expression and purification of Trx-linker monomers and concatemers

Genes encoding the N-terminal His6-SUMO-tagged concatemers of Trx were cloned into the pOP3SU plasmid (kindly provided by M. Hyvönen). BLR(DE3) competent cells

(Novagen) were transformed with the plasmids and grown in a Luria broth medium supplemented with ampicillin (100 μg ml^{-1}) at 37 °C with continuous shaking (250 r.p.m.). Protein expression was induced in the exponential growth phase (OD_{600} = 0.6) with isopropyl-β-D-1-thiogalactopyranoside (0.5 mM final concentration). After 8 h, the cells were harvested by centrifugation (10 min, 5,000×*g*), resuspended in a binding buffer (30 mM Tris HCl, 250 mM NaCl, 25 mM imidazole at pH 7.2) supplemented with a protease inhibitor cocktail (cOmplete, EDTA free; Roche) and lysed by sonication. Cell debris was removed by centrifugation at 20,000×*g* for 45 min, and the supernatant loaded onto a HisTrap HP column (5 ml, Cytiva) at 0.2 ml min^{-1}. The column was washed with 50 ml of the binding buffer before single-step elution with 15 mL of 30 mM Tris HCl, 250 mM NaCl, 300 mM imidazole at pH 7.2. A single peak containing the almost pure protein was collected and dialysed (Slide-A-Lyzer G2 Dialysis Cassette, 10,000 molecular weight cutoff, 30 ml; Thermo Fisher) for 3 h against 4 l of dialysis buffer (50 mM Tris HCl, 250 mM NaCl, 2 mM 1,4-dithio-D-threitol (DTT) at pH 8.0), at 4 °C with continuous stirring, to remove excess imidazole. After injecting His6-tagged Ulp1 protease into the dialysis cassette at a molar concentration ratio of 1:200 (Ulp1:Trx-linker concatemer), the mixture was transferred into a fresh dialysis buffer overnight for SUMO-tag cleavage. The cassette was then transferred one last time into fresh dialysis buffer without DTT for 4 h. The dialysed protein was loaded onto a column packed with HisPur Ni-NTA Agarose Resin (5 ml, Thermo Fisher) equilibrated with a binding buffer (50 mM Tris HCl, 250 mM NaCl at pH 8.0) and the flow through was reapplied five more times. The final flow through containing the His6-SUMO-free protein was aliquoted and flash frozen for storage at −80 °C.

Expression and purification of SUMO protease Ulp1

The Pfget19_Ulp1 plasmid (Addgene) containing a His6-tagged Ulp1 gene was transformed into T7 Express competent cells (NEB) and grown in a Luria broth medium supplemented with kanamycin (100 μg ml^{-1}) at 37 °C with shaking (250 r.p.m.). Expression was induced at OD_{600} = 0.5 with isopropyl-β-D-1-thiogalactopyranoside (0.5 mM). Cells were harvested after 3 h by centrifugation, resuspended in lysis buffer (4 ml g^{-1}; 50 mM Tris HCl, 300 mM NaCl, 10 mM imidazole at pH 7.5) supplemented with lysozyme (1 mg ml^{-1}) and incubated on ice for 30 min before sonication. The lysate was spun at 20,000 r.p.m. for 45 min to remove the cell debris and the supernatant was applied to a column packed with HisPur Ni-NTA Agarose Resin (5 ml, Thermo Fisher) and equilibrated with a binding buffer (50 mM Tris HCl, 300 mM NaCl at pH 7.5). The column was washed with 10 column volumes of wash buffer (50 mM Tris HCl, 300 mM NaCl, 20 mM imidazole at pH 7.5) and the protein was eluted with 10 ml of elution buffer (50 mM Tris HCl, 300 mM NaCl, 300 mM imidazole at pH 7.5). The eluted protein was dialysed against a storage buffer (50 mM Tris HCl, 200 mM NaCl, 2 mM 2-mercaptoethanol) overnight, aliquoted and flash frozen as a 50% stock in glycerol.

Phosphorylation of Trx-linker concatemers

Trx-linker concatemers (1 mg ml^{-1}) were incubated with 50,000 units of the catalytic subunit of cAMP-dependent protein kinase (NEB)—which recognizes the RRAS motif within the central linker of the Trx-linker nonamer—in a protein kinase buffer (50.0 mM Tris HCl at pH 7.5,10.0 mM $MgCl_2$, 0.1 mM EDTA, 4.0 mM DTT, and 2.0 mM

ATP) (NEB) at 30 °C for 1 h. The solution was then supplemented with additional ATP at a final concentration of 2 mM and DTT at a final concentration of 2 mM before overnight incubation at 30 °C. Trx-linker concatemers were purified and concentrated using centrifugal filters (Amicon Ultra 0.5 ml, 100 K), aliquoted and flash frozen for storage at −20 °C (10 mM HEPES at pH 7.2 and 750 mM KCl). Phosphorylation of the Trx-linker concatemers at a single site was verified by liquid chromatography–mass spectrometry.

Modification of cysteines on Trx-linker concatemers

Reagents were purchased from Sigma-Aldrich, unless otherwise indicated. Trx-linker nonamer was first treated with TCEP (70 to 100 eq) at 32 °C for 2 h in a protein storage buffer (50 mM Tris HCl, 250 mM NaCl at pH 8.0). Excess TCEP was removed by a desalting column (PD MiniTrap G-25 column, Cytiva). To glutathionylate the Trx-linker nonamer, the reduced protein was reacted with oxidized glutathione (100 eq) at 32 °C overnight in a protein storage buffer (50 mM Tris HCl, 250 mM NaCl at pH 8.0) before desalting to remove the excess reagent. The modified protein was aliquoted and flash frozen for storage at −20 °C. To glycosylate the Trx-linker nonamers, the reduced protein was reacted first with 2,2’-dithiodipyridine (20 eq) at 32 °C overnight in the protein storage buffer (50 mM Tris HCl, 250 mM NaCl at pH 8.0). After the removal of excess 2,2’-dithiodipyridine with a desalting column, the activated nonamer was reacted with the 6’-sialyllactosamine derivative (NeuAcα(2-6)LacNAc-PEG3-Thiol, 5 eq; Sussex Research Laboratories) overnight at 32 °C in a protein storage buffer (50 mM Tris HCl, 250 mM NaCl at pH 8.0). Modified nonamers were desalted (PD MiniTrap G-

25 column, Cytiva), aliquoted and flash frozen for storage at −20 °C. The occurrence of glutathionylation or glycosylation at single sites was verified by liquid chromatography–mass spectrometry.

Single-channel recording

Planar lipid bilayers of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) were formed by using the Müller–Montal method on a 50-µm-diameter aperture made in a Teflon film (25 µm thick, Goodfellow) separating two 500 µl compartments (*cis* and *trans*) of the recording chamber. Each compartment was filled with a recording buffer (750 mM GdnHCl, 1.5 M GdnHCl, 3.0 M GdnHCl, 2.0 M urea/750 mM KCl or 750 mM KCl, 10 mM HEPES, 5 mM TCEP at pH 7.2 for Trx-linker dimer, tetramer, hexamer and octamer; 375 mM GdnHCl/375 mM KCl, 10 mM HEPES at pH 7.2 for Trx-linker nonamers). To record with Trx-linker dimer, tetramer, hexamer or octamer and ensure a reduced N-terminal cysteine, pretreatment of the protein samples with 5 mM TCEP was carried out for 10 min at room temperature. This pretreatment was not carried out with nonamers which lacked the N-terminal cysteine residue. Trx-linker concatemers were added to the *cis* compartment (dimer, 2.20 µM; tetramer, 0.63 µM; hexamer, 0.25 µM; octamer, 0.81 µM; nonamer, 1.20 µM). Ionic currents were measured at 24 ± 1 °C by using Ag/AgCl electrodes connected to the headstage of an Axopatch 200B amplifier. After a single (NN-113R)$_7$ pore had inserted into the bilayer, the solution was replaced with a fresh buffer by manual pipetting, to prevent further insertions. Signals were low-pass

filtered at 10 kHz and sampled at 50 kHz with a Digidata 1440A digitizer (Molecular Devices).

Data analysis

To establish the current signatures for the stepwise co-translocational unfolding of Trx concatemers, current traces were analysed using Clampfit 10.7 (Molecular Devices). The remaining current as a percentage of the open-pore current ($I_{res\%}$) was calculated for each step in individual A or B features (for example, $I_{res\%}(A1) = I_{A1}/I_{open} \times 100\%$). The standard deviations were derived from data for Trx-linker units collected using separate pores. Trx-linker units that produced a level A3 or B3 with a dwell time of <1 ms were excluded from the $I_{res\%}$ analysis due to possible undersampling. Root-mean-square noise values ($I_{r.m.s.}$) for each current level were measured from current traces after the application of a post-recording filter of 2 kHz. Unless otherwise stated, the noise of the open pore was subtracted as follows: $I_{r.m.s.}^2 = I_{r.m.s.}(A1)^2 - I_{r.m.s.}(\text{open pore})^2$. To obtain the stepwise kinetic profiles of the co-translocational unfolding of Trx concatemers, current traces were idealized using Clampfit 10.7. The dwell-time analysis was performed by using the maximum interval likelihood algorithm of QUB 2.0 software (https://qub.mandelics.com/).

CD and fluorescence unfolding experiments

A desired denaturant concentration was obtained by diluting a known volume of a freshly made 9M Urea or 6M GdnHCl solution with 10 mM Hepes pH 7.2, 1 M KCl into a series of tubes containing a constant amount of protein analyte. All protein samples

were incubated at room temperature for at least 3h for equilibration before transferring to the optical cuvette for the CD or fluorescence measurements. The final urea concentrations were checked by measuring the refractive indices of each protein solution. CD measurements were performed with a Chirascan instrument (Applied Photophysics). CD measurements were carried out at 222 nm and fluorescence emission spectra were collected at an excitation wavelength of 280 nm and an emission wavelength of 350 nm.

8 Supplementary material

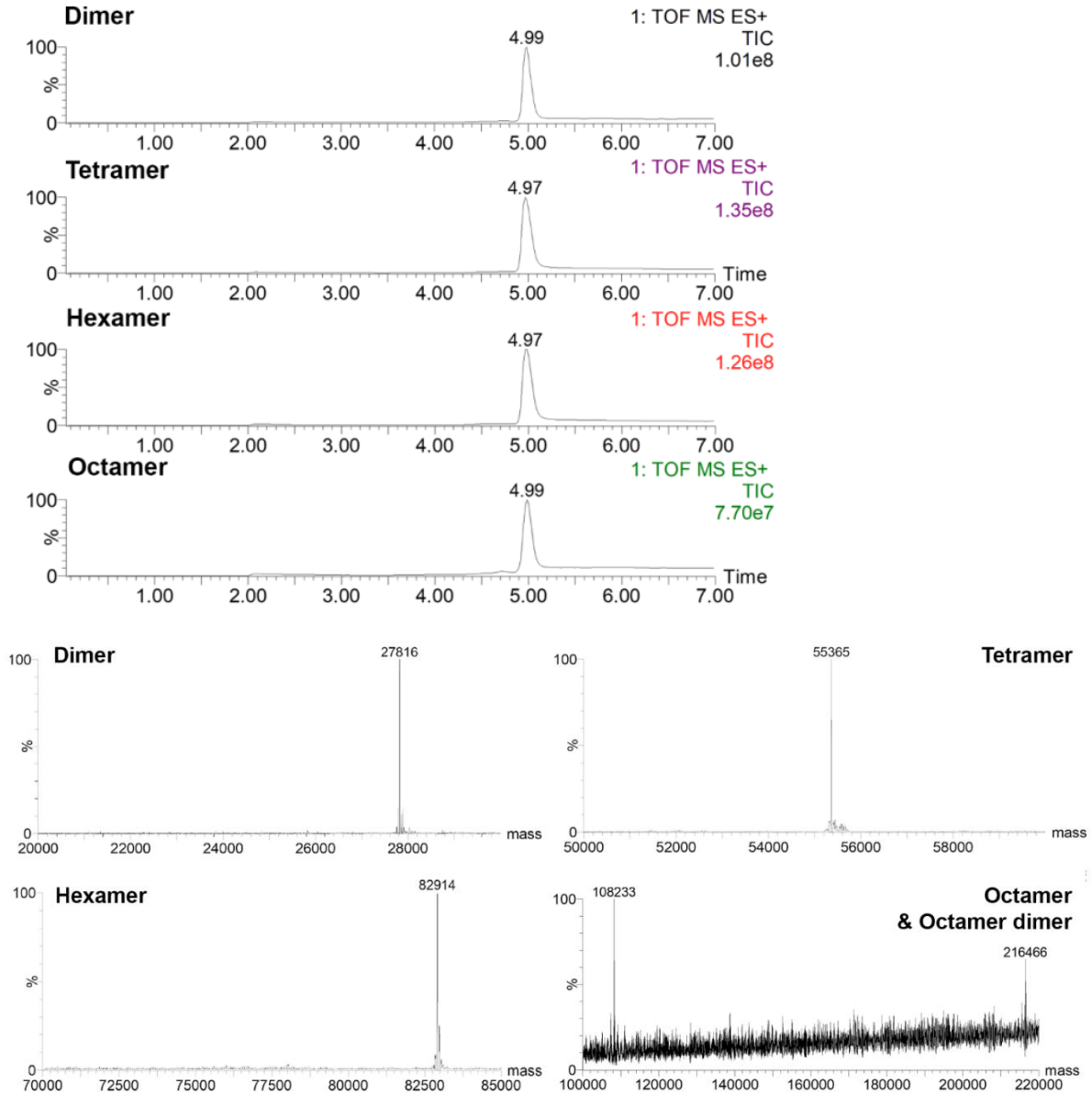


Fig. S1 LC-MS characterization of Trx-linker concatemers. LC-MS chromatograms (top) and deconvoluted ESI-MS spectra (bottom) are shown. Dimer, (Trx-linker)2: mass = 27816 Da (calc) and 27816 Da (obs); Tetramer, (Trx-linker)4: mass = 55367 Da (calc) and 55365 Da (obs); Hexamer, (Trxlinker)6: mass = 82918 Da (calc) and 82914 Da (obs); Octamer, (Trx-linker)7Trx:

mass = 108231 Da (calc) and 108233 Da (obs); Dimer of octamers: mass = 216460 (calc) and 216466 Da (obs). Reproduced from [116].

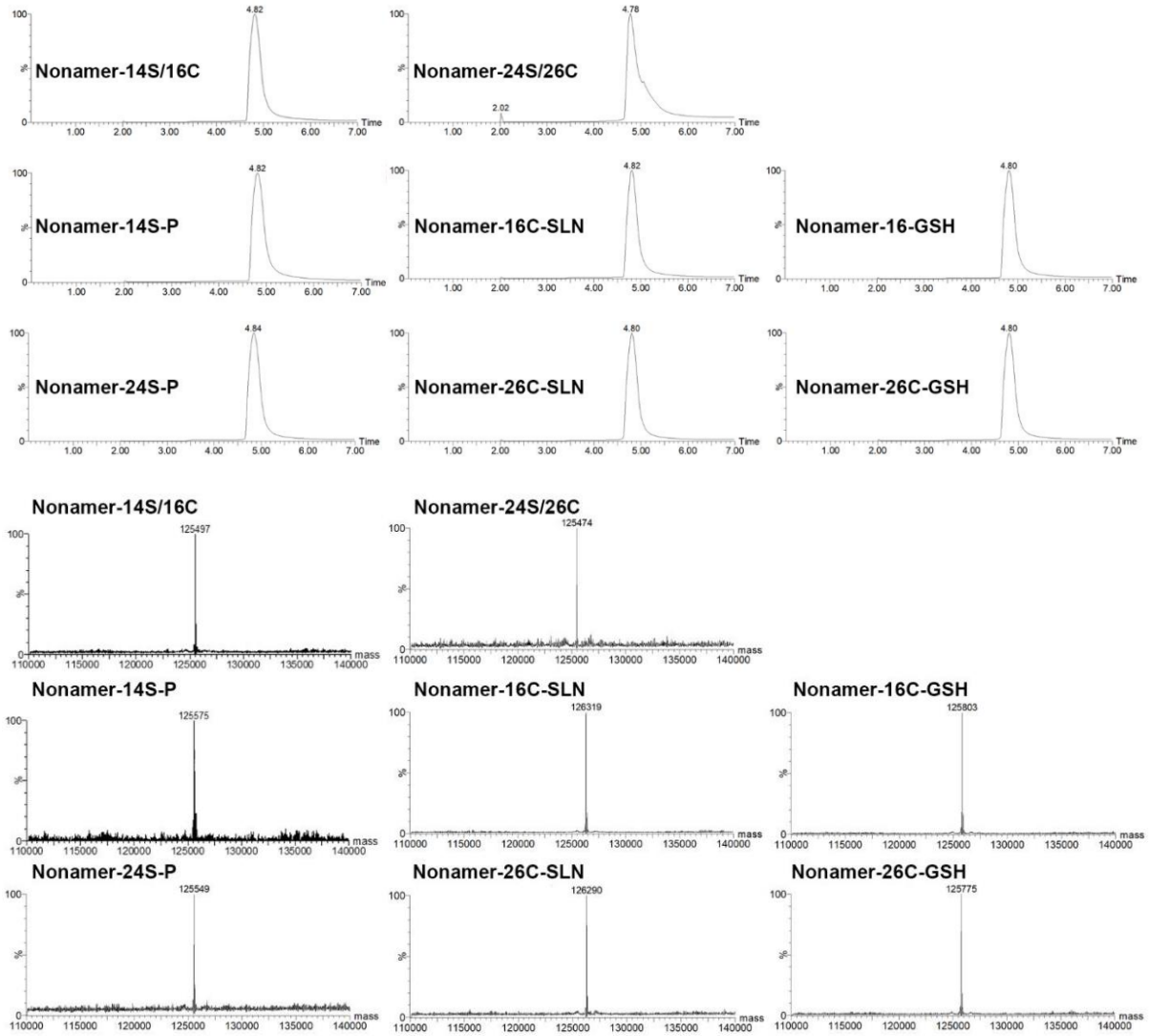


Fig. S2 LC-MS characterization of Trx-linker nonamers. LC-MS chromatograms (top) and deconvoluted ESI-MS spectra (bottom). Nonamer 14S/16C: mass = 125498 Da (calc) and 125497 Da (obs); Nonamer 24S/26C: mass = 125470 Da (calc) and 125474 Da (obs); Nonamer 14S-P: mass = 125578 Da (calc) and 125575 Da (obs); Nonamer 16C-SLN: mass = 126319 Da (calc) and 126319 Da (obs); Nonamer 16C-GSH: mass = 125803 (calc) and 125803 Da (obs); Nonamer 24S-P: mass = 125550 Da (calc) and 125549 Da (obs); Nonamer 26C-SLN: mass =

126291 Da (calc) and 126290 Da (obs); Nonamer 26C-GSH: mass = 125775 (calc) and 125775 Da (obs). Reproduced from [116].

Bibliography

[1] H. Bayley and C. R. Martin, ‘Resistive-pulse sensing - from microbes to molecules’, *Chem Rev*, vol. 100, no. 7, pp. 2575–2594, Jul. 2000, doi: 10.1021/cr980099g.

[2] R. W. Deblois, A. Charles, P. Bean, and R. K. A. Wesley, ‘Electrokinetic Measurements with Submicron Particles and Pores by the Resistive Pulse Technique’, 1977.

[3] H. Bayley and P. S. Cremer, ‘Engineering pores for a wide variety of analytes Stochastic sensors inspired by biology’, 2001. [Online]. Available: www.nature.com

[4] L. J. Mandel, ‘Modern Elect1·ochemistry: An Inter-disciplinary Study, chap’, University of California Press, 1970.

[5] G. J. S. Charlotte L. Hill, *An introduction to patch clamp recording*. 2020.

[6] M. Montal, P. MUELLERtj, and P. and, ‘Formation of Bimolecular Membranes from Lipid Monolayers and a Study of Their Electrical Properties (membrane structure/membrane reconstitution/asymmetric membranes/lipid bilayers) The following chemicals were used: glyceroldioleate (Armak Chemicals Division’, 1972.

[7] W. R. Redwood, F. R. Pfeiffer, J. A. Weisbach, and T. E. Thompson, ‘BIOCHIMICA ET BIOPHYSICA ACTA I BBA 75589 PHYSICAL PROPERTIES OF BILAYER MEMBRANES FORMED FROM A SYNTHETIC SATURATED PHOSPHOLIPID IN n-DECANE’.

[8] T. Baba, Y. Toshima, H. Minamikawa, M. Hato, K. Suzuki, and N. Kamo, ‘Formation and characterization of planar lipid bilayer membranes from synthetic phytanyl-chained glycolipids’, 1999. [Online]. Available: www.elsevier.com

[9] A. V Krylov, P. Pohl, M. L. Zeidel, and W. G. Hill, ‘Water Permeability of Asymmetric Planar Lipid Bilayers: Leaflets of Different Composition Offer Independent and Additive Resistances to Permeation’, 2001. [Online]. Available: http://www.jgp.org/cgi/content/full/118/4/333

[10] A. Wiese, J. 0 Reiners, K. Brandenburg, K. Kawahara, U. Zahringer, and U. Seydel, ‘Planar Asymmetric Lipid Bilayers of Glycosphingolipid or Lipopolysaccharide on One Side and Phospholipids on the Other: Membrane Potential, Porin Function, and Complement Activation’, 1996.

[11] A. Jo, H. Hoi, H. Zhou, M. Gupta, and C. D. Montemagno, ‘Single-molecule study of full-length NaChBac by planar lipid bilayer recording’, *PLoS One*, vol. 12, no. 11, Nov. 2017, doi: 10.1371/journal.pone.0188861.

[12] P. Jing, H. Paraiso, and B. Burris, ‘Highly efficient integration of the viral portal proteins from different types of phages into planar bilayers for the black lipid membrane analysis’, *Mol Biosyst*, vol. 12, no. 2, pp. 480–489, 2016, doi: 10.1039/c5mb00573f.

[13] B. A. Heitz *et al.*, ‘Polymerized planar suspended lipid bilayers for single ion channel recordings: Comparison of several dienoyl lipids’, *Langmuir*, vol. 27, no. 5, pp. 1882–1890, Mar. 2011, doi: 10.1021/la1025944.

[14] N. Malmstadt, T. J. Jeon, and J. J. Schmidt, ‘Long-lived planar lipid bilayer membranes anchored to an in situ polymerized hydrogel’, *Advanced Materials*, vol. 20, no. 1, pp. 84–89, Jan. 2008, doi: 10.1002/adma.200700810.

[15] W. Meier, C. Nardin, and M. Winterhalter, ‘Reconstitution of channel proteins in (Polymerized) ABA triblock copolymer membranes’, *Angewandte Chemie - International Edition*, vol. 39, no. 24, pp. 4599–4602, Dec. 2000, doi: 10.1002/1521-3773(20001215)39:24<4599::AID-ANIE4599>3.0.CO;2-Y.

[16] L. Restrepo-Pérez, C. Joo, and C. Dekker, ‘Paving the way to single-molecule protein sequencing’, *Nat Nanotechnol*, vol. 13, no. 9, pp. 786–796, 2018, doi: 10.1038/s41565-018-0236-6.

[17] A. Fragasso, S. Schmid, and C. Dekker, ‘Comparing Current Noise in Biological and Solid-State Nanopores’, *ACS Nano*, vol. 14, no. 2. American Chemical Society, pp. 1338–1349, Feb. 25, 2020. doi: 10.1021/acsnano.9b09353.

[18] D. F. Cairns-Gibson and S. L. Cockroft, ‘Functionalised nanopores: Chemical and biological modifications’, *Chem Sci*, vol. 13, no. 7, pp. 1869–1882, Feb. 2022, doi: 10.1039/d1sc05766a.

[19] J. M. Carter and S. Hussain, ‘Robust long-read native DNA sequencing using the ONT CsgG Nanopore system’, *Wellcome Open Res*, vol. 2, 2017, doi: 10.12688/wellcomeopenres.11246.1.

[20] I. M. Derrington *et al.*, ‘Nanopore DNA sequencing with MspA’, doi: 10.1073/pnas.1001831107/-/DCSupplemental.

[21] E. A. Manrao *et al.*, ‘Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase’, *Nat Biotechnol*, vol. 30, no. 4, pp. 349–353, Apr. 2012, doi: 10.1038/nbt.2171.

[22] D. Rodriguez-Larrea and H. Bayley, ‘Protein co-translocational unfolding depends on the direction of pulling’, *Nat Commun*, vol. 5, 2014, doi: 10.1038/ncomms5841.

[23] D. Rodriguez-Larrea and H. Bayley, ‘Multistep protein unfolding during nanopore translocation’, *Nat Nanotechnol*, vol. 8, no. 4, pp. 288–295, 2013, doi: 10.1038/nnano.2013.22.

[24] G. F. Schneider *et al.*, ‘DNA translocation through graphene nanopores’, *Nano Lett*, vol. 10, no. 8, pp. 3163–3167, Aug. 2010, doi: 10.1021/nl102069z.

[25] J. Kudr *et al.*, ‘Fabrication of solid-state nanopores and its perspectives’, *Electrophoresis*, vol. 36, no. 19. Wiley-VCH Verlag, pp. 2367–2379, Oct. 01, 2015. doi: 10.1002/elps.201400612.

[26] M. Van Den Hout, A. R. Hall, M. Y. Wu, H. W. Zandbergen, C. Dekker, and N. H. Dekker, ‘Controlling nanopore size, shape and stability’, *Nanotechnology*, vol. 21, no. 11, 2010, doi: 10.1088/0957-4484/21/11/115304.

[27] A. J. Storm, J. H. Chen, X. S. Ling, H. W. Zandbergen, and C. Dekker, ‘Fabrication of solid-state nanopores with single-nanometre precision’, *Nature Materials*, vol. 2, no. 8. European Association for Cardio-Thoracic Surgery, pp. 537–540, 2003. doi: 10.1038/nmat941.

[28] R. Wei, V. Gatterdam, R. Wieneke, R. Tampé, and U. Rant, ‘Stochastic sensing of proteins with receptor-modified solid-state nanopores’, *Nat Nanotechnol*, vol. 7, no. 4, pp. 257–263, 2012, doi: 10.1038/nnano.2012.24.

[29] E. C. Yusko *et al.*, ‘Real-time shape approximation and fingerprinting of single proteins using a nanopore’, *Nat Nanotechnol*, vol. 12, no. 4, pp. 360–367, May 2017, doi: 10.1038/nnano.2016.267.

[30] E. Turker Acar, S. F. Buchsbaum, C. Combs, F. Fornasiero, and Z. S. Siwy, ‘A P P L I E D S C I E N C E S A N D E N G I N E E R I N G Biomimetic potassium-selective nanopores’, 2019.

[31] S. Marion, N. Vučemilović-Alagić, M. Špadina, A. Radenović, and A. S. Smith, ‘From Water Solutions to Ionic Liquids with Solid State Nanopores as a Perspective to Study Transport and Translocation Phenomena’, *Small*, vol. 17, no. 25, Jun. 2021, doi: 10.1002/smll.202100777.

[32] S. Pud *et al.*, ‘Mechanical Trapping of DNA in a Double-Nanopore System’, *Nano Lett*, vol. 16, no. 12, pp. 8021–8028, Dec. 2016, doi: 10.1021/acs.nanolett.6b04642.

[33] U. F. Keyser *et al.*, ‘Direct force measurements on DNA in a solid-state nanopore’, *Nat Phys*, vol. 2, no. 7, pp. 473–477, Jul. 2006, doi: 10.1038/nphys344.

[34] M. J. N. Sampad, H. Zhang, T. D. Yuzvinsky, M. A. Stott, A. R. Hawkins, and H. Schmidt, ‘Optical trapping assisted label-free and amplification-free detection of SARS-CoV-2 RNAs with an optofluidic nanopore sensor’, *Biosens Bioelectron*, vol. 194, Dec. 2021, doi: 10.1016/j.bios.2021.113588.

[35] Y. Lin *et al.*, ‘Characterization of DNA duplex unzipping through a sub-2 nm solid-state nanopore’, *Chemical Communications*, vol. 53, no. 25, pp. 3539–3542, 2017, doi: 10.1039/c7cc00060j.

[36] A. R. Hall, A. Scott, D. Rotem, K. K. Mehta, H. Bayley, and C. Dekker, ‘Hybrid pore formation by directed insertion of α-haemolysin into solid-state nanopores’, *Nat Nanotechnol*, vol. 5, no. 12, pp. 874–877, 2010, doi: 10.1038/nnano.2010.237.

[37] K. R. Mahendran *et al.*, ‘A monodisperse transmembrane α-helical peptide barrel’, *Nat Chem*, vol. 9, no. 5, pp. 411–419, May 2017, doi: 10.1038/nchem.2647.

[38] N. A. W. Bell and U. F. Keyser, ‘Nanopores formed by DNA origami: A review’, *FEBS Letters*, vol. 588, no. 19. Elsevier B.V., pp. 3564–3570, Oct. 01, 2014. doi: 10.1016/j.febslet.2014.06.013.

[39] N. A. W. Bell *et al.*, ‘DNA origami nanopores’, *Nano Lett*, vol. 12, no. 1, pp. 512–517, Jan. 2012, doi: 10.1021/nl204098n.

[40] E. Spruijt, S. E. Tusk, and H. Bayley, ‘DNA scaffolds support stable and uniform peptide nanopores’, *Nat Nanotechnol*, vol. 13, no. 8, pp. 739–745, Aug. 2018, doi: 10.1038/s41565-018-0139-6.

[41] S. Schmid, P. Stömmer, H. Dietz, and C. Dekker, ‘Nanopore electro-osmotic trap for the label-free study of single proteins and their conformations’, *Nat Nanotechnol*, vol. 16, no. 11, pp. 1244–1250, Nov. 2021, doi: 10.1038/s41565-021-00958-5.

[42] C. Wloka *et al.*, ‘Label-Free and Real-Time Detection of Protein Ubiquitination with a Biological Nanopore’, *ACS Nano*, vol. 11, no. 5, pp. 4387–4394, May 2017, doi: 10.1021/acsnano.6b07760.

[43] M. Rincon-Restrepo, E. Mikhailova, H. Bayley, and G. Maglia, ‘Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes’, *Nano Lett*, vol. 11, no. 2, pp. 746–750, Feb. 2011, doi: 10.1021/nl1038874.

[44] G. Huang *et al.*, ‘PlyAB Nanopores Detect Single Amino Acid Differences in Folded Haemoglobin from Blood**’, 2022, doi: 10.26434/chemrxiv-2021-7gb4h.

[45] G. Huang, K. Willems, M. Soskine, C. Wloka, and G. Maglia, ‘Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores’, *Nat Commun*, vol. 8, no. 1, Dec. 2017, doi: 10.1038/s41467-017-01006-4.

[46] G. Huang, A. Voet, and G. Maglia, ‘FraC nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution’, *Nat Commun*, vol. 10, no. 1, Dec. 2019, doi: 10.1038/s41467-019-08761-6.

[47] C. Cao *et al.*, ‘Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores’, *Nat Commun*, vol. 10, no. 1, Dec. 2019, doi: 10.1038/s41467-019-12690-9.

[48] A. Asandei, I. Schiopu, M. Chinappi, C. H. Seo, Y. Park, and T. Luchian, ‘Electroosmotic Trap Against the Electrophoretic Force Near a Protein Nanopore Reveals Peptide Dynamics during Capture and Translocation’, *ACS Appl Mater Interfaces*, vol. 8, no. 20, pp. 13166–13179, May 2016, doi: 10.1021/acsami.6b03697.

[49] M. Soskine, A. Biesemans, B. Moeyaert, S. Cheley, H. Bayley, and G. Maglia, ‘An engineered ClyA nanopore detects folded target proteins by selective external

association and pore entry', *Nano Lett*, vol. 12, no. 9, pp. 4895–4900, Sep. 2012, doi: 10.1021/nl3024438.

[50] N. S. Galenkamp, M. Soskine, J. Hermans, C. Wloka, and G. Maglia, 'Direct electrical quantification of glucose and asparagine from bodily fluids using nanopores', *Nat Commun*, vol. 9, no. 1, Dec. 2018, doi: 10.1038/s41467-018-06534-1.

[51] M. Soskine, A. Biesemans, and G. Maglia, 'Single-molecule analyte recognition with ClyA nanopores equipped with internal protein adaptors', *J Am Chem Soc*, vol. 137, no. 17, pp. 5793–5797, May 2015, doi: 10.1021/jacs.5b01520.

[52] L. Movileanu, S. Cheley, S. Howorka, O. Braha, and H. Bayley, 'Location of a Constriction in the Lumen of a Transmembrane Pore by Targeted Covalent Attachment of Polymer Molecules', 2001. [Online]. Available: http://www.jgp.org/cgi/content/full/117/3/239

[53] C.-Y. Chang, B. Niblack, B. Walker, and H. Bayley, 'A photogenerated pore-forming protein'.

[54] H. C. Wu, Y. Astier, G. Maglia, E. Mikhailova, and H. Bayley, 'Protein nanopores with covalently attached molecular adapters', *J Am Chem Soc*, vol. 129, no. 51, pp. 16142–16148, Dec. 2007, doi: 10.1021/ja0761840.

[55] J. Clarke, H. C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley, 'Continuous base identification for single-molecule nanopore DNA sequencing', *Nat Nanotechnol*, vol. 4, no. 4, pp. 265–270, 2009, doi: 10.1038/nnano.2009.12.

[56] S. H. Shin and H. Bayley, 'Stepwise growth of a single polymer chain', *J Am Chem Soc*, vol. 127, no. 30, pp. 10462–10463, Aug. 2005, doi: 10.1021/ja052194u.

[57] T. Luchian, S. H. Shin, and H. Bayley, 'Kinetics of a three-step reaction observed at the single-molecule level', *Angewandte Chemie - International Edition*, vol. 42, no. 17, pp. 1926–1929, Apr. 2003, doi: 10.1002/anie.200250666.

[58] S. H. Shin, T. Luchian, S. Cheley, O. Braha, and H. Bayley, ‘Kinetics of a reversible covalent-bond-forming reaction observed at the single-molecule level’, *Angewandte Chemie - International Edition*, vol. 41, no. 19, pp. 3707–3709, Oct. 2002, doi: 10.1002/1521-3773(20021004)41:19<3707::AID-ANIE3707>3.0.CO;2-5.

[59] Y. Qing, G. S. Pulcu, N. A. W. Bell, and H. Bayley, ‘Bioorthogonal Cycloadditions with Sub-Millisecond Intermediates’, *Angewandte Chemie*, vol. 130, no. 5, pp. 1232–1235, Jan. 2018, doi: 10.1002/ange.201710262.

[60] G. S. Pulcu, E. Mikhailova, L. S. Choi, and H. Bayley, ‘Continuous observation of the stochastic motion of an individual small-molecule walker’, *Nat Nanotechnol*, vol. 10, no. 1, pp. 76–83, Jan. 2015, doi: 10.1038/nnano.2014.264.

[61] Y. Qing, S. A. Ionescu, S. Pulcu, and H. Bayley, ‘Directional control of a processive molecular hopper’. [Online]. Available: https://www.science.org

[62] Y. Qing and H. Bayley, ‘Enzymeless DNA Base Identification by Chemical Stepping in a Nanopore’, *J Am Chem Soc*, vol. 143, no. 43, pp. 18181–18187, Nov. 2021, doi: 10.1021/jacs.1c07497.

[63] J. Lee and H. Bayley, ‘Semisynthetic protein nanoreactor for single-molecule chemistry’, *Proc Natl Acad Sci U S A*, vol. 112, no. 45, pp. 13768–13773, Nov. 2015, doi: 10.1073/pnas.1510565112.

[64] S. Borsley and S. L. Cockroft, ‘In Situ Synthetic Functionalization of a Transmembrane Protein Nanopore’, *ACS Nano*, vol. 12, no. 1, pp. 786–794, Jan. 2018, doi: 10.1021/acsnano.7b08105.

[65] M. M. Haugland, S. Borsley, D. F. Cairns-Gibson, A. Elmi, and S. L. Cockroft, ‘Synthetically diversified protein nanopores: Resolving click reaction mechanisms’, *ACS Nano*, vol. 13, no. 4, pp. 4101–4110, Apr. 2019, doi: 10.1021/acsnano.8b08691.

[66] S. Lin *et al.*, ‘Redox-based reagents for chemoselective methionine bioconjugation’. [Online]. Available: https://www.science.org

[67] A. I. Derman *et al.*, ‘Proteomics: quantitative and physical mapping of cellular proteins’, 1993.

[68] P. Minakshi *et al.*, ‘Single-cell proteomics: Technology and applications’, in *Single-Cell Omics: Volume 1: Technological Advances and Applications*, Elsevier, 2019, pp. 283–318. doi: 10.1016/B978-0-12-814919-5.00014-2.

[69] S. P. Gygi *et al.*, ‘Correlation between Protein and mRNA Abundance in Yeast’, 1999.

[70] R. Ahmad and B. Budnik, ‘A review of the current state of single-cell proteomics and future perspective’, *Analytical and Bioanalytical Chemistry*, vol. 415, no. 28. Springer Science and Business Media Deutschland GmbH, pp. 6889–6899, Nov. 01, 2023. doi: 10.1007/s00216-023-04759-8.

[71] R. Ahmad and B. Budnik, ‘A review of the current state of single-cell proteomics and future perspective’, *Analytical and Bioanalytical Chemistry*, vol. 415, no. 28. Springer Science and Business Media Deutschland GmbH, pp. 6889–6899, Nov. 01, 2023. doi: 10.1007/s00216-023-04759-8.

[72] ‘Single-cell proteomics: challenges and prospects’, *Nature Methods*, vol. 20, no. 3. Nature Research, pp. 317–318, Mar. 01, 2023. doi: 10.1038/s41592-023-01828-9.

[73] I. Virant-Klun, S. Leicht, C. Hughes, and J. Krijgsveld, ‘Identification of maturation-specific proteins by single-cell proteomics of human oocytes’, *Molecular and Cellular Proteomics*, vol. 15, no. 8, pp. 2616–2627, Aug. 2016, doi: 10.1074/mcp.M115.056887.

[74] ‘Introduction to flow cytometry’, https://www.abcam.com/protocols/introduction-to-flow-cytometry.

[75] Y. Su, Q. Shi, and W. Wei, ‘Single cell proteomics in biomedicine: High-dimensional data acquisition, visualization, and analysis’, *Proteomics*, vol. 17, no. 3–4. Wiley-VCH Verlag, Feb. 01, 2017. doi: 10.1002/pmic.201600267.

[76] S.-E. G. J. S. &Bjørn T. G. Sonia Gavasso, ‘Single-cell proteomics: potential implications for cancer diagnostics’, Accessed: Apr. 09, 2024. [Online]. Available: https://www.tandfonline.com/doi/full/10.1586/14737159.2016.1156531

[77] ‘ELISpot Assay Principle’, https://www.rndsystems.com/products/elispot-assay-principle.

[78] H. M. Bennett, W. Stephenson, C. M. Rose, and S. Darmanis, ‘Single-cell proteomics enabled by next-generation sequencing or mass spectrometry’, *Nature Methods*, vol. 20, no. 3. Nature Research, pp. 363–374, Mar. 01, 2023. doi: 10.1038/s41592-023-01791-5.

[79] P. Wang and S. R. Wilson, ‘Mass spectrometry-based protein identification by integrating de novo sequencing with database searching.’, *BMC Bioinformatics*, vol. 14 Suppl 2, 2013, doi: 10.1186/1471-2105-14-S2-S24.

[80] P. A. DiMaggio and C. A. Floudas, ‘De novo peptide identification via tandem mass spectrometry and integer linear optimization’, *Anal Chem*, vol. 79, no. 4, pp. 1433–1446, Feb. 2007, doi: 10.1021/ac0618425.

[81] E. J. Dupree, M. Jayathirtha, H. Yorkey, M. Mihasan, B. A. Petre, and C. C. Darie, ‘A critical review of bottom-up proteomics: The good, the bad, and the future of this field’, *Proteomes*, vol. 8, no. 3. MDPI AG, pp. 1–26, Sep. 01, 2020. doi: 10.3390/proteomes8030014.

[82] M. Cui, C. Cheng, and L. Zhang, ‘High-throughput proteomics: a methodological mini-review’, *Laboratory Investigation*, vol. 102, no. 11. Springer Nature, pp. 1170–1181, Nov. 01, 2022. doi: 10.1038/s41374-022-00830-7.

[83] P. Edman and G. Begg, ‘A Protein Sequenator’, *Eur J Biochem*, vol. 1, no. 1, pp. 80–91, 1967, doi: 10.1111/j.1432-1033.1967.tb00047.x.

[84] ‘The Edman degradation cycle ’, https://www.genosphere-biotech.com/es/notas-tecnicas/protein-identification/edman-degradation/.

[85] M. M. Vecchi, Y. Xiao, and D. Wen, ‘Identification and Sequencing of N-Terminal Peptides in Proteins by LC-Fluorescence-MS/MS: An Approach to Replacement of the Edman Degradation’, *Anal Chem*, vol. 91, no. 21, pp. 13591–13600, Nov. 2019, doi: 10.1021/acs.analchem.9b02754.

[86] K. Motone, N. Cardozo, and J. Nivala, ‘Herding cats: Label-based approaches in protein translocation through nanopore sensors for single-molecule protein sequence analysis’, *iScience*, vol. 24, no. 9, p. 103032, 2021, doi: 10.1016/j.isci.2021.103032.

[87] M. Muthukumar, ‘Macromolecular Mechanisms of Protein Translocation’.

[88] C. T. A. Wong and M. Muthukumar, ‘Polymer translocation through α -hemolysin pore with tunable polymer-pore electrostatic interaction’, *Journal of Chemical Physics*, vol. 133, no. 4, Jul. 2010, doi: 10.1063/1.3464333.

[89] J. Nivala, L. Mulroney, G. Li, J. Schreiber, and M. Akeson, ‘Discrimination among protein variants using an unfoldase-coupled nanopore’, *ACS Nano*, vol. 8, no. 12, pp. 12365–12375, 2014, doi: 10.1021/nn5049987.

[90] J. Nivala, D. B. Marks, and M. Akeson, ‘Unfoldase-mediated protein translocation through an α-hemolysin nanopore’, *Nat Biotechnol*, vol. 31, no. 3, pp. 247–250, Mar. 2013, doi: 10.1038/nbt.2503.

[91] C. W. Ho, V. Van Meervelt, K. C. Tsai, P. J. De Temmerman, J. Mast, and G. Maglia, ‘Engineering a nanopore with co-chaperonin function’, *Sci Adv*, vol. 1, no. 11, Dec. 2015, doi: 10.1126/sciadv.1500905.

[92] K. Motone and J. Nivala, ‘Not if but when nanopore protein sequencing meets single-cell proteomics’, *Nature Methods*, vol. 20, no. 3. Nature Research, pp. 336–338, Mar. 01, 2023. doi: 10.1038/s41592-023-01800-7.

[93] L. Yu *et al.*, ‘Unidirectional Single-File Transport of Full-Length Proteins Through a Nanopore’, *bioRxiv*, p. 2021.09.28.462155, 2021, [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.09.28.462155v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.09.28.462155v1.abstract

[94] K. Motone *et al.*, ‘Multi-pass, single-molecule nanopore reading of long protein strands with 1 single-amino acid sensitivity 2’, doi: 10.1101/2023.10.19.563182.

[95] H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev, and C. Dekker, ‘Multiple rereads of single proteins at single-amino acid resolution using nanopores’, 2023. [Online]. Available: https://www.science.org

[96] L.-Q. Gu, S. Cheley, and H. Bayley, ‘Electroosmotic enhancement of the binding of a neutral molecule to a transmembrane pore’, 2003. [Online]. Available: https://www.pnas.org

[97] L.-Q. Gu, S. Cheley, and H. Bayley, ‘Prolonged Residence Time of a Noncovalent Molecular Adapter,-Cyclodextrin, within the Lumen of Mutant-Hemolysin Pores’, 2001. [Online]. Available: http://www.jgp.org/cgi/content/full/118/5/481

[98] A. Aksimentiev and K. Schulten, ‘Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map’, *Biophys J*, vol. 88, no. 6, pp. 3745–3761, 2005, doi: 10.1529/biophysj.104.058727.

[99] S. Van Dorp, U. F. Keyser, N. H. Dekker, C. Dekker, and S. G. Lemay, ‘Origin of the electrophoretic force on DNA in solid-state nanopores’, *Nat Phys*, vol. 5, no. 5, pp. 347–351, 2009, doi: 10.1038/nphys1230.

[100] G. Di Muccio, B. Morozzo Della Rocca, and M. Chinappi, ‘Geometrically Induced Selectivity and Unidirectional Electroosmosis in Uncharged Nanopores’, *ACS Nano*, vol. 16, no. 6, pp. 8716–8728, Jun. 2022, doi: 10.1021/acsnano.1c03017.

[101] A. Sauciuc, B. Morozzo della Rocca, M. J. Tadema, M. Chinappi, and G. Maglia, ‘Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force’, *Nat Biotechnol*, 2023, doi: 10.1038/s41587-023-01954-x.

[102] G. Huang, K. Willems, M. Bartelds, P. van Dorpe, M. Soskine, and G. Maglia, ‘Electro-Osmotic Vortices Promote the Capture of Folded Proteins by PlyAB Nanopores’, *Nano Lett*, 2020, doi: 10.1021/acs.nanolett.0c00877.

[103] A. Sauciuc, B. Morozzo, M. Tadema, M. Chinappi, and G. Maglia, ‘An engineered electroosmotic flow transports unravelled proteins across nanopores’, doi: 10.1101/2023.02.17.528930.

[104] A. Gubbiotti, M. Baldelli, G. Di Muccio, P. Malgaretti, S. Marbach, and M. Chinappi, ‘Electroosmosis in nanopores: computational methods and technological applications’, *Advances in Physics: X*, vol. 7, no. 1. Taylor and Francis Ltd., 2022. doi: 10.1080/23746149.2022.2036638.

[105] W. Shinoda, ‘Permeability across lipid membranes’, *Biochim Biophys Acta Biomembr*, vol. 1858, no. 10, pp. 2254–2265, Oct. 2016, doi: 10.1016/j.bbamem.2016.03.032.

[106] A. V Krylov, P. Pohl, M. L. Zeidel, and W. G. Hill, ‘Water Permeability of Asymmetric Planar Lipid Bilayers: Leaflets of Different Composition Offer Independent and Additive Resistances to Permeation’, 2001. [Online]. Available: http://www.jgp.org/cgi/content/full/118/4/333

[107] D. Rodriguez-Larrea and H. Bayley, ‘Multistep protein unfolding during nanopore translocation’, *Nat Nanotechnol*, vol. 8, no. 4, pp. 288–295, 2013, doi: 10.1038/nnano.2013.22.

[108] K. Willems, D. Ruić, A. Biesemans, N. S. Galenkamp, P. Van Dorpe, and G. Maglia, ‘Engineering and Modeling the Electrophoretic Trapping of a Single Protein Inside a Nanopore’, *ACS Nano*, vol. 13, no. 9, pp. 9980–9992, 2019, doi: 10.1021/acsnano.8b09137.

[109] A. Biesemans, M. Soskine, and G. Maglia, ‘A Protein Rotaxane Controls the Translocation of Proteins Across a ClyA Nanopore’, *Nano Lett*, vol. 15, no. 9, pp. 6076–6081, 2015, doi: 10.1021/acs.nanolett.5b02309.

[110] S. Zhang *et al.*, ‘Bottom-up fabrication of a proteasome–nanopore that unravels and processes single proteins’, *Nat Chem*, vol. 13, no. 12, pp. 1192–1199, 2021, doi: 10.1038/s41557-021-00824-w.

[111] D. Di Marino, E. L. Bonome, A. Tramontano, and M. Chinappi, ‘All-Atom Molecular Dynamics Simulation of Protein Translocation through an α-Hemolysin Nanopore’, *Journal of Physical Chemistry Letters*, vol. 6, no. 15, pp. 2963–2968, Aug. 2015, doi: 10.1021/acs.jpclett.5b01077.

[112] C. B. Rosen, D. Rodriguez-Larrea, and H. Bayley, ‘Single-molecule site-specific detection of protein phosphorylation with a nanopore’, *Nat Biotechnol*, vol. 32, no. 2, pp. 179–181, Feb. 2014, doi: 10.1038/nbt.2799.

[113] I. E. Gentle, D. P. De Souza, and M. Baca, ‘Direct production of proteins with N-terminal cysteine for site-specific conjugation’, *Bioconjug Chem*, vol. 15, no. 3, pp. 658–663, 2004, doi: 10.1021/bc049965o.

[114] D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir, and J. W. Chin, ‘Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal

cyanobenzothiazole condensation', *J Am Chem Soc*, vol. 133, no. 30, pp. 11418–11421, Aug. 2011, doi: 10.1021/ja203111c.

[115] D. J. A. Roderer, M. A. Schärer, M. Rubini, and R. Glockshuber, 'Acceleration of protein folding by four orders of magnitude through a single amino acid substitution', *Sci Rep*, vol. 5, Jun. 2015, doi: 10.1038/srep11840.

[116] P. Martin-Baniandres *et al.*, 'Enzyme-less nanopore detection of post-translational modifications within long polypeptides', *Nat Nanotechnol*, 2023, doi: 10.1038/s41565-023-01462-8.

[117] A. L. Pey *et al.*, 'Engineering proteins with tunable thermodynamic and kinetic stabilities', *Proteins: Structure, Function and Genetics*, vol. 71, no. 1, pp. 165–174, Apr. 2008, doi: 10.1002/prot.21670.

[118] V. Muñ and J. M. Sanchez-Ruiz, 'Sciences of the USA 17646-17651 PNAS', 2004. [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0405829101

[119] P. Tian and I. Andricioaei, 'Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mitochondrial pore', *J Mol Biol*, vol. 350, no. 5, pp. 1017–1034, Jul. 2005, doi: 10.1016/j.jmb.2005.05.035.

[120] R. S. Winardhi, Q. Tang, J. Chen, M. Yao, and J. Yan, 'Probing Small Molecule Binding to Unfolded Polyprotein Based on its Elasticity and Refolding', *Biophys J*, vol. 111, no. 11, pp. 2349–2357, Dec. 2016, doi: 10.1016/j.bpj.2016.10.031.

[121] P. Shrestha *et al.*, 'Single-molecule mechanical fingerprinting with DNA nanoswitch calipers', *Nat Nanotechnol*, vol. 16, no. 12, pp. 1362–1370, Dec. 2021, doi: 10.1038/s41565-021-00979-0.

[122] S. Howorka and H. Bayley, 'Probing Distance and Electrical Potential within a Protein Pore with Tethered DNA', 2002.

[123] M. T. Noakes *et al.*, ‘Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage’, *Nat Biotechnol*, vol. 37, no. 6, pp. 651–656, Jun. 2019, doi: 10.1038/s41587-019-0096-0.

[124] V. V. Palyulin, T. Ala-Nissila, and R. Metzler, ‘Polymer translocation: The first two decades and the recent diversification’, *Soft Matter*, vol. 10, no. 45. Royal Society of Chemistry, pp. 9016–9037, Dec. 07, 2014. doi: 10.1039/c4sm01819b.

[125] H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev, and C. Dekker, ‘Multiple rereads of single proteins at single-amino acid resolution using nanopores’, *Science (1979)*, vol. 374, no. 6574, pp. 1509–1513, 2021, doi: 10.1126/science.abl4381.

[126] J. Dadová, S. R. Galan, and B. G. Davis, ‘Synthesis of modified proteins via functionalization of dehydroalanine’, *Current Opinion in Chemical Biology*, vol. 46. Elsevier Ltd, pp. 71–81, Oct. 01, 2018. doi: 10.1016/j.cbpa.2018.05.022.

[127] A. Yang *et al.*, ‘A chemical biology route to site-specific authentic protein modifications’, *Science (1979)*, vol. 354, no. 6312, pp. 623–626, Nov. 2016, doi: 10.1126/science.aah4428.

[128] O. Harel and M. Jbara, ‘Posttranslational Chemical Mutagenesis Methods to Insert Posttranslational Modifications into Recombinant Proteins’, *Molecules*, vol. 27, no. 14, Jul. 2022, doi: 10.3390/molecules27144389.

[129] C. D. Spicer and B. G. Davis, ‘Selective chemical protein modification’, *Nature Communications*, vol. 5. Nature Publishing Group, 2014. doi: 10.1038/ncomms5740.

[130] E. A. Rodriguez *et al.*, ‘The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins’, *Trends in Biochemical Sciences*, vol. 42, no. 2. Elsevier Ltd, pp. 111–129, Feb. 01, 2017. doi: 10.1016/j.tibs.2016.09.010.

[131] D. Parrello, C. Mustin, D. Brie, S. Miron, and P. Billard, ‘Multicolor whole-cell bacterial sensing using a synchronous fluorescence spectroscopy-based approach’, *PLoS One*, vol. 10, no. 3, Mar. 2015, doi: 10.1371/journal.pone.0122848.

[132] L. Martin, A. Che, and D. Endy, ‘Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression’, *PLoS One*, vol. 4, no. 11, Nov. 2009, doi: 10.1371/journal.pone.0007569.

[133] N. Cardozo *et al.*, ‘Multiplexed direct detection of barcoded protein reporters on a nanopore array’, *Nat Biotechnol*, vol. 40, no. 1, pp. 42–46, Jan. 2022, doi: 10.1038/s41587-021-01002-6.