Engineering a biological nanopore for protein sensing



Sandra Ionescu

Exeter College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas Term 2019

“Far and away the best prize that life has to offer is the chance to work hard at work worth doing.”

— Theodore Roosevelt

Abstract

Protein nanopores have been engineered with chemical or genetic sensing elements for the stochastic detection of analytes at the single-molecule level. Upon insertion into a lipid bilayer, a nanopore allows an ionic current to flow across the bilayer under an applied transmembrane potential. Interactions between the sensing element and an analyte molecule modulate the ionic current, generating a unique electrical signature useful for analyte identification and characterisation.

This thesis describes approaches to engineer the *E. coli* porin OmpF as a novel nanopore for stochastic sensing. In contrast to previously reported monomeric nanopore sensors, OmpF is a trimer composed of three identical β-barrels. As the most abundant porin in the bacterial outer membrane, it forms a major route for antibiotic permeation and is therefore of clinical interest. First, the orientation of OmpF in synthetic lipid bilayers was determined by covalent modification of introduced cysteine residues. This is a necessary step for dissecting periplasmic from extracellular interactions in a physiological context. The targeted covalent modification approach can be generally applied to determine the orientation of other porins and toxins. Alongside this, a protocol was developed for in vitro expression of OmpF to facilitate rapid production of mutants for studying bacterial physiology or prototyping stochastic sensors. The protocol was applied to construct heterotrimers, in which one or two of the three identical OmpF barrels were differently modified. This enables the production of nanopores with only one sensing element, eliminating the possibility of multivalent interactions that would complicate analysis of the sensing output. Finally, peptide libraries were engineered into the peripheral loops of OmpF and screened by bacterial display against an antibody protein target. Although initial attempts did not yield viable nanopore sensors, a foundation was established for the screening and testing of nanopore libraries based on their interactions with a target of interest. The ultimate goal is to produce sensors for any protein target, including disease biomarkers for which no suitable ligand exists. Future improvements for successful implementation of this technology, which would enable highly selective and multiplexed protein detection, are discussed.

Acknowledgements

I have enjoyed an extraordinary network of support these past four years which made the work presented in this thesis possible. For his invaluable mentorship throughout my PhD, I would like to thank Professor Hagan Bayley, who allowed me the freedom to cultivate my scientific curiosities while never losing site of the bigger picture. I am equally grateful for the other talented members of the Bayley group, who made me a better scientist and became valued friends along the way. In particular, I'd like to thank: Yujia Qing for being an unyielding pillar of support. I aspire to her honesty, scientific rigor, and bubbly personality, which brightened some of my toughest days. The Bob collective for emergency beers, summertime BoBQs, and movie nights. Amina Ali, whose accomplishments during her Part II project rightfully earned her the title of 'Gel Girl'. And Sejeong Lee, for introducing me to the frustrating yet miraculous technology of single-channel electrophysiology.

I would also like to extend my gratitude to Exeter College, a home-away-from-home where I developed many lifelong friendships and learned lessons in leadership and comradery, and to the Amelia Jackson fund for the studentship that enabled my research. Thank you to the Super Squash Squad, with whom I ended up laughing and sharing scientific horror stories over more glasses of wine than games of squash. Especially to Owen Crowther, for his endless encouragement and dedicated proofreading—I am lucky to call you my housemate.

Finally, I would like to thank my family for their endless support. In particular, my mother and grandmother, for instilling in me the values of hard work, perseverance, and unconditional love.

Contents

1 Introduction **1**

1.1 Nanopore stochastic sensing 2
1.1.1 Planar lipid bilayers 3
1.1.2 Droplet interface bilayers 5
1.1.3 Single-molecule kinetics 7

1.2 Nanopores 8
1.2.1 Bacterial porins 10
1.2.2 Outer membrane protein F 12

1.3 Overview 14

2 Orientation of OmpF in planar lipid bilayers **17**

2.1 Introduction 17
2.1.1 OmpF reconstitution in PLBs 18
2.1.2 Methods to determine porin orientation in a PLB 19
2.1.3 Chapter overview 22

2.2 Characterisation of WT OmpF in PLBs 23

2.3 Characterisation of OmpF cysteine mutants in PLBs 25

2.4 Targeted covalent modification 28
2.4.1 Modification with MTSES 29
2.4.2 Modification with mPEG-MAL 30
2.4.3 Modification with mPEG-OPSS 36

2.5 Relative orientation of OmpF and cysteine mutants 38

2.6 Establishing orientation when the I-V curve is symmetric 41

2.7 Conclusions 43

3 In vitro expression of OmpF and heterotrimer engineering **47**

3.1 Introduction 47
3.1.1 In vitro expression systems 48
3.1.2 In vitro expression of membrane proteins 49
3.1.3 Folding and trimerisation of OmpF 51
3.1.4 In vitro expression of heteromeric nanopores 54
3.1.5 Separation of nanopores by SDS-PAGE 55
3.1.6 Chapter overview 57

3.1.7 Acknowledgement of collaboration 58
3.2 In vitro expression of WT OmpF 58
3.3 Origin of unexpected expression products 62
3.3.1 Alternative additives 66
3.4 Trimerisation of IVTT expressed monomer 68
3.5 Functional analysis of in vitro expressed OmpF 70
3.6 OmpF heterotrimers formed by IVTT 72
3.6.1 Genetic modification 73
3.7 Conclusions 76

4 Genetically engineered nanopore libraries for stochastic protein sensing 79

4.1 Introduction 79
4.1.1 Protein detection with nanopores 80
4.1.2 Chemically engineered sensors 81
4.1.3 Genetically engineered sensors 84
4.1.4 Synthetic peptide libraries 86
4.1.5 Bacterial display 88
4.1.6 Fluorescence activated cell sorting 90
4.1.7 Chapter overview 92
4.2 Library controls 93
4.3 Stochastic protein sensing with OmpF 94
4.4 Bacterial display of library controls 100
4.4.1 Surface display of peptides 104
4.4.2 Fluorescence activated cell sorting 105
4.5 OmpF bacterial display libraries 109
4.5.1 Library design 110
4.5.2 Library FACS 113
4.6 Extracting screened nanopores from bacteria 116
4.7 Electrophysiology of screened nanopores 118
4.7.1 DHB electrophysiology 119
4.7.2 PLB electrophysiology 123
4.8 Conclusions 125

5 Future developments 127

5.1 Introduction 127
5.2 The orientation of porins in planar lipid bilayers 128
5.3 Applications of OmpF heterotrimers 130

5.4 Protein sensors screened from a genetically engineered nanopore library 133

5.5 Conclusions 135

6 Materials and methods 137

6.1 General methods 137

6.2 Ch 2: Orientation of OmpF in planar lipid bilayers 137

6.2.1 OmpF expression and purification 138

6.2.2 Single-channel PLB recordings 139

6.3 Ch 3: In vitro expression of OmpF and heterotrimers engineering 141

6.3.1 Liposome preparation 143

6.3.2 IVTT 143

6.3.3 Gel extraction 144

6.3.4 Trimerisation of gel extracted monomers 144

6.4 Ch 4: Genetically engineered nanopore libraries for stochastic protein sensing 144

6.4.1 Growth media 146

6.4.2 Fluorescent cell labelling 147

6.4.3 Fluorescence activated cell sorting 147

6.4.4 Live/dead staining 148

6.4.5 Library synthesis 149

6.4.6 OmpF detergent extraction from E. coli 151

Abbreviations

αHL	α-hemolysin
βOG	*n*-octyl-β-D-glucopyranoside
BAM	β-barrel assembly machinery
CFU	Colony forming units
CMC	Critical micelle concentration
DDM	*n*-dodecyl-β-D-maltoside
DG	*n*-dodecyl-β-D-glucopyranoside
DHB	Droplet-hydrogel bilayer
DIB	Droplet interface bilayer
DPhPC	1,2-diphytanoyl-*sn*-glycero-3-phosphocholine
DTT	Dithiothreitol
DYRK1a	Dual-specificity tyrosine phosphorylation-regulated kinase 1a
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
I-V	Current-voltage
IVTT	In vitro transcription and translation
LPS	Lipopolysaccharide
mPEG-MAL	Methoxy poly(ethylene glycol)-maleimide
mPEG-OPSS	Methoxy poly(ethylene glycol)-orthopyridyl disulfide
MTSES	(2-sulfonatoethyl)-methanethiosulfonate
PLB	Planar lipid bilayer
PMMA	Poly(methyl methacrylate)
PURE	Protein synthesis using recombinant elements
OM	Outer membrane
OmpF	Outer membrane protein F
SDS	Sodium dodecyl sulfate
TEV	Tobacco etch virus
10:0 PC	1,2-didecanoyl-*sn*-glycero-3-phosphocholine
DMPC	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine

1
Introduction

Stochastic sensing with nanometre-sized pores (nanopores) has evolved into an important biotechnological tool for the detection and characterisation of biomolecules, notably sequencing DNA [1]. The first iteration of pore-based sensing took place with the invention of the Coulter counter, which sized and counted blood cells by measuring resistive pulses across an aperture (20 μm–2 mm holes) bathed in electrolyte solution with electrodes on either side [2]. The electrodes generate a constant current, which undergoes increased resistance when particles pass through the aperture and block a proportion of the electrolyte volume, registered as a voltage drop. The frequency and amplitude of the resistive pulses was used to determine to the number and size of the cells in the sample, respectively.

By using smaller (~0.3 μm) track-etched pores, the sensitivity of the technique was increased to allow detection of nanoscale particles and viruses [3]. The current state-of-the-art of resistive-pulse sensing began with the use of nanometre sized holes, usually generated by biological components—most often membrane proteins—which allow detection of even smaller analytes, such as single molecules and ions. Alongside advances in protein engineering methods, stochastic sensing with nanopores has proven to be a powerful single-molecule

approach which provides the resolution needed to discriminate analytes in a mixture [4–7], follow chemical reactions inside a nanoreactor [8–10], and sequence DNA and RNA [11–13].

This chapter provides an overview of nanopore sensing, including the most widely used electrophysiology techniques. The development of different nanopores is then reviewed, including the OmpF nanopore which is engineered for protein sensing in this thesis.

1.1 Nanopore stochastic sensing

In its simplest form, a nanopore is a hollow tube that inserts into membranes, thereby opening up a channel through which an ionic current can flow. When an external analyte interacts with a nanopore, the current is interrupted; the frequency of the interruptions yields the concentration of the analyte, whereas the duration and amplitude form a characteristic signature which can be used to identify it [14]. To increase detection specificity, unique binding elements have been engineered into nanopores to create sensors for a wide range of analytes from ions to macromolecules (Fig. 1.1).

The most powerful approach to nanopore sensing is single-channel recording, in which a single nanopore is inserted in a synthetic bilayer to monitor molecular interactions. The first definitive single-channel recording was demonstrated using gramicidin A, an antibiotic peptide [15]. Discrete increases in conductance were observed each time gramicidin A inserted into the bilayer, which became more frequent with increasing peptide concentrations. Since then, a variety of pores have been interrogated at the single-molecule level, which are further discussed in Section 1.1. After reconstitution into a bilayer, each nanopore

exhibits a characteristic conductance that is related to the channel characteristics, such as internal diameter and charge.

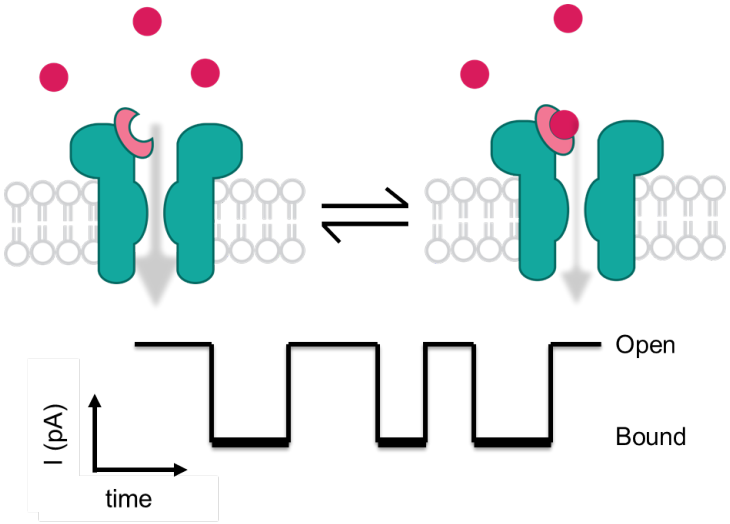


Figure 1.1 Nanopore stochastic sensing. A nanopore (teal) engineered with a binding element (pink) for a target analyte (circle) is inserted into a PLB. When a transmembrane potential is applied, an ionic current flows through the pore (open). Reversible binding of the analyte to the nanopore modulates the current, producing a characteristic signal (bound).

The two major single-channel reconstitution techniques used in this thesis are further discussed. Compared to patch clamping, which was developed around the same time and allows single-channel recording directly from biological membranes, these techniques have allowed the study of a wider range of protein channels and pores with µs resolution, including extremely toxic proteins such as the protective antigen of anthrax [16].

1.1.1 Planar lipid bilayers

Planar lipid bilayers (PLB) are routinely used as a model for understanding the structural and functional characteristics of biological membranes. In this technique, a bilayer is formed over an aperture in PTFE film (25 µm thick) that separates two buffer-filled compartments. The aperture is formed by an electric spark discharge,

with a tuneable diameter usually falling between 30-100 μm. The compartment size may also be varied but is usually 0.5-1.0 mL, as smaller volumes are sensitive to evaporation. The Montal-Mueller technique is commonly used for bilayer formation [17], in which the aperture is treated with hexadecane (1-10% v/v in pentane) and allowed to dry, after which buffer followed by a few microlitres of lipid dissolved in pentane is added to both sides of the bilayer. The lipid assembles into a monolayer at the air-water interface. By repeatedly raising and lowering the buffer level in each compartment, the monolayers eventually come together at the aperture, forming a bilayer (Fig. 1.2 A).

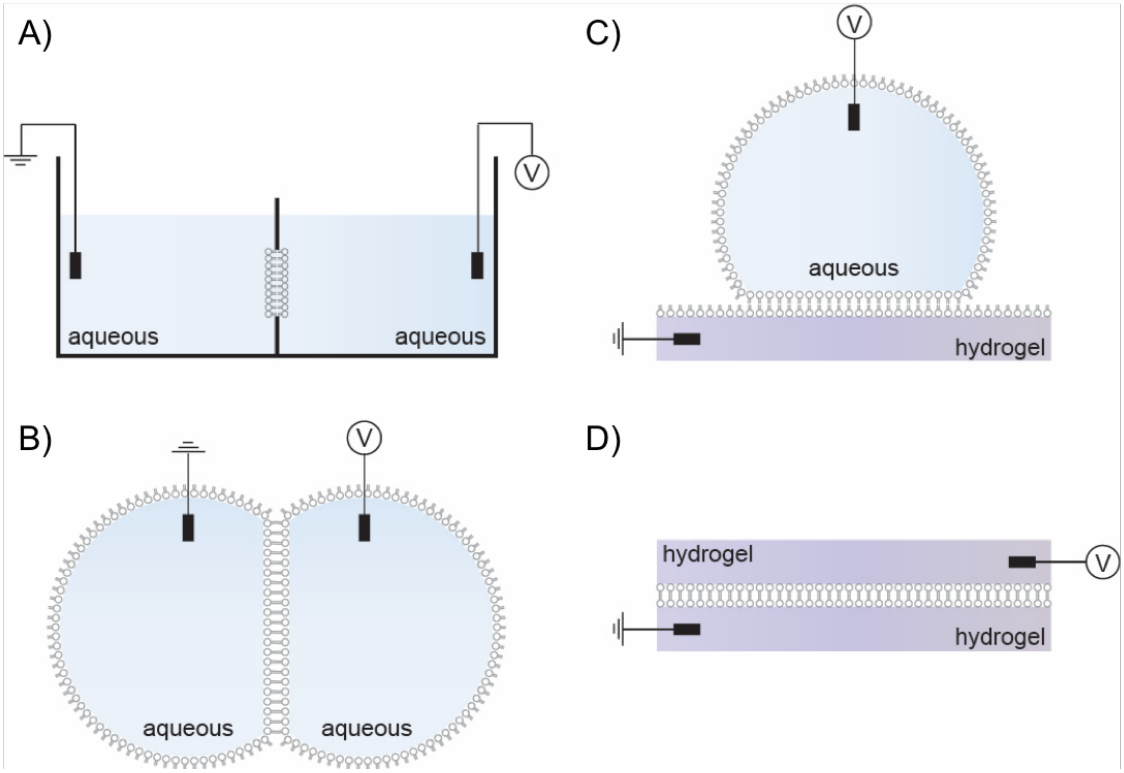


Figure 1.2 Single-channel electrophysiology techniques include **(A)** planar lipid bilayer, **(B)** droplet interface (droplet-droplet) bilayer, **(C)** droplet-hydrogel bilayer, and **(D)** hydrogel-hydrogel bilayer.

Although a variety of lipids have been trialled [18], the methylated lipid 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC) is most used in PLB studies due to its stability and lack of a phase transition over the normal experimental temperature range [19]. Asymmetric PLBs with unique lipid compositions in each leaflet have also been reported [20,21], although the stability timeframe and the extent of homogenisation once the membrane has been broken remains unclear.

To perform electrical measurements, Ag/AgCl electrodes are typically used, with one electrode placed in each chamber. In the convention used in this thesis, the nanopore is always introduced into the cis compartment (ground) and potential is defined as positive when it is greater on the trans side. Bilayer incorporation of detergent solubilised nanopores is facilitated by application of a transmembrane potential and stirring of the chamber contents, which introduces defects in the membrane structure thus lowering the energy barrier for insertion. Nanopores may also be reconstituted through the fusion of proteoliposomes [22,23].

One disadvantage of PLBs compared to patch clamp experiments is the significantly higher capacitance, which leads to increased noise in the current read-out. A simple solution is the use of smaller apertures, but this makes nanopore insertion more difficult. In addition, PLBs are metastable structures with lifetimes of several hours under ideal conditions. New advances in membrane chemistry—including polymerisable lipids [24], hydrogel-anchored bilayers with enhanced mechanical stability [25], and the use of amphiphilic polymers [26]—have increased the lifetime of these bilayers.

1.1.2 Droplet interface bilayers

A more recent technique that overcomes many limitations of PLBs is that of droplet interface bilayers (DIB). A DIB is formed between two aqueous droplets

(pico- to nano-litre volume) immersed in a lipid-containing oil (Fig. 1.2 B). The droplets are left to stabilise for a few minutes to allow monolayer formation at the water-oil interface, after which they are gently brought into contact by a micromanipulator, leading to spontaneous bilayer formation [27]. Thin Ag/AgCl electrodes (100 µm diameter) are used for electrical recordings; the tip of the electrodes may be coated with agarose to provide a hydrophilic interface that enhances the electrode-droplet interaction.

The DIB system has several advantages: The size of the bilayer can be tuned after formation by moving the droplets via a micromanipulator [27,28]. This is advantageous for single-channel recording because further nanopore insertion can be prevented by minimising the bilayer size, which also reduces noise in the electrical recording. Further, DIBs can be repeatedly ruptured and formed between the same or different droplets without cross-contamination, which is useful for high-throughput applications. In one example, an array of DIBs was utilised to screen human potassium channel inhibitors [29]. The small volume of the aqueous droplets minimises consumption of expensive reagents and produces more robust bilayers than PLBs, having lifetimes that can range from days to weeks [27].

One disadvantage is that there is no direct access to the aqueous compartments after bilayer formation. However, reagents may be introduced by fusion of a third droplet. The formation of DIBs can also be achieved with other aqueous compartments, such as hydrogels, to construct more robust droplet-hydrogel bilayers (DHB) [14,30] or hydrogel-hydrogel bilayers [31,32] (Fig. 1.2 C–D). The modular nature of the easy-to-implement DIB system, which can be built into complex bilayer networks using a 3D printer [33], has accelerated

investigation of membrane biology and has the potential to evolve into a powerful platform for synthetic biology, including the fabrication of synthetic tissues [34].

1.1.3 Single-molecule kinetics

Both techniques described above can be used to extract single-molecule kinetics of reversible interactions between analytes and nanopores. The ability to observe individual molecules reveals the heterogeneity of a population, which is masked in ensemble measurements. Various parameters can be analysed in real time, such as the open pore current, the amplitude and noise of the bound state, the binding dwell time, and the duration between binding events.

In the simplest case, a nanopore engineered to bind a particular molecule has two states: analyte bound and analyte unbound. In this pseudo-first order bimolecular reaction, in which the analyte concentration does not significantly vary over time, binding kinetics can be easily extracted from the distributions of the time spent in the bound state (τ_{off}) and the duration between binding events (τ_{on}). Each distribution is fit by an exponential decay function to yield the average dwell time. The association rate constant is then calculated as $k_{on}=1/\tau_{on}[A]$, where [A] is the concentration of the analyte, and the concentration-independent dissociation rate constant is then calculated as $k_{off}=1/\tau_{off}$. In more complex cases, such as multi-state binding, the unique current signatures can still suffice to identify and quantify the analyte even if the exact binding kinetics cannot be extracted from the current trace.

Analyte sensing with engineered nanopores has been demonstrated over a wide dynamic range covering ten orders of magnitude from millimolar [35] to near femtomolar [36] affinities. An ideal stochastic sensor should have moderate affinity for the analyte, to allow the recording of multiple binding events within a

reasonable experimental timeframe. The interaction can be modulated by changes in voltage, pH, ionic strength, temperature, protein concentration, and additives such as denaturants [37].

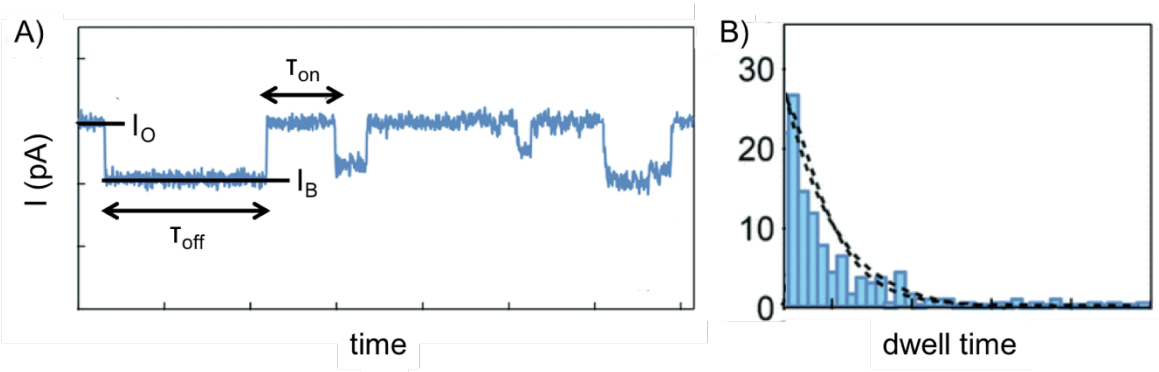


Figure 1.3 Analysis of single-molecule binding. **(A)** Representative current trace showing stochastic interactions between an analyte and a nanopore. Analyte binding blocks a proportion of the ionic current, resulting in a decrease of the open pore current (I_O) to the bound state (I_B). The duration between binding events (τ_{on}) and the binding dwell time (τ_{off}) is used to calculate rate constants for the interaction. **(B)** Example of a dwell time histogram (τ_{on} or τ_{off}) fit with an exponential decay function.

1.2 Nanopores

Stochastic sensing has been accomplished with both synthetic and biological nanopores, which have been extensively reviewed [14,38,39]. Techniques such as native chemical ligation have allowed the creation of semi-synthetic pores with unnatural amino acids alongside traditional chemical and genetic modification [8]. Hybrid pores have also been made by templating protein peptides on DNA origami [40] or by inserting biological nanopores into solid-state pores [41]. Synthetic nanopores have superior thermal and chemical stability and, importantly, their size can be tuned to accommodate various biomolecules. However, they lack the well-controlled geometry common to their protein counterparts, which can be modified

with atomic precision at specified residues. Further use of ‘nanopore’ in this thesis refers specifically to biological nanopores.

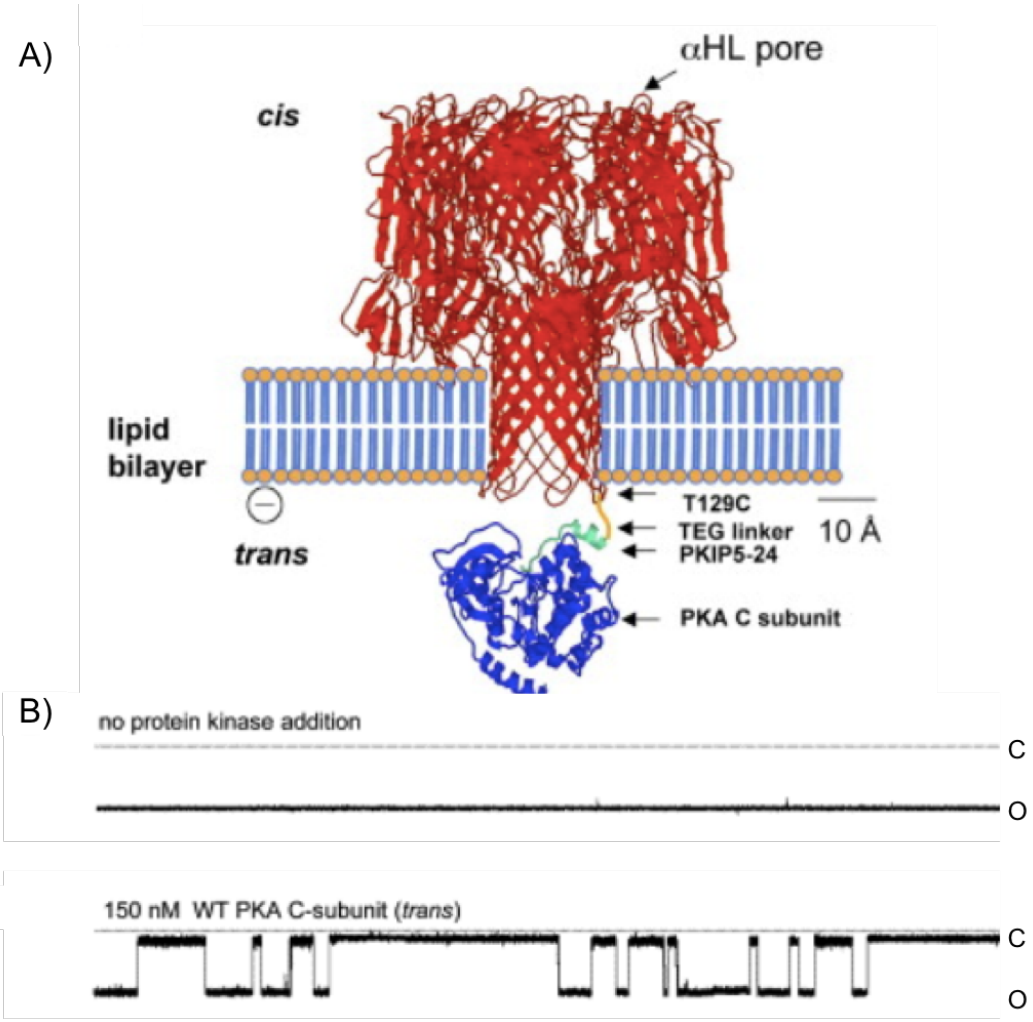


Figure 1.4 Kinase sensing with an engineered nanopore. **(A)** Structure of the αHL sensor (red) displaying a kinase inhibitor peptide (PKIP, cyan) on the trans side of the bilayer, which binds the constitutively active PKA kinase (PKA C, blue). The peptide was attached to only one of the seven subunits at an engineered cysteine residue (T129C). **(B)** Detection of individual binding events of the kinase to the αHL pore. A stable open pore current (O, open) was observed without kinase addition. Once PKA was introduced into the trans side of the bilayer, discrete binding events were observed that blocked a proportion of the ionic current through the pore (C, closed). Reproduced from [42].

Bacterial pore-forming membrane proteins are the most commonly used class of nanopores, including α-hemolysin (αHL) from *Staphylococcus aureus* which sparked the field of nanopore-based molecular analysis [43]. The shared β-barrel

structure is highly stable, showing resistance to sodium dodecyl sulfate (SDS) up to temperatures of ~65°C, and can be extensively engineered while retaining pore-forming activity. The αHL pore remained functional after the entire amino acid sequence of the barrel-forming domain was reversed, forming a 'retro' pore [44]. In a different study, it was shown that significant truncation of the pore-forming domain (up to 8 amino acids) still produced monomers which could successfully oligomerise and insert into membranes, showcasing the robustness of this class of pores/these pores [45]. The αHL pore has been extensively engineered, both by genetic and chemical means, to sense a wide range of analytes. An example of stochastic sensing with an engineered αHL pore is shown in Figure 1.4.

1.2.1 Bacterial porins

Alternative nanopore candidates comprising stable β-barrels are the bacterial porins, found in the outer membrane (OM) of Gram-negative bacteria [46]. The OM, which consists of an asymmetric bilayer with an outer leaflet rich in lipopolysaccharide (LPS), protects cells from toxic environments while still enabling efficient nutrient uptake via porins [47]. The general porin structure can be imagined as an antiparallel β-sheet, formed from a single polypeptide chain, that wraps around a central pore to form a cylinder stabilised by an inter-strand hydrogen bonding network; the barrel is closed via a salt bridge between the N- and C-terminus. This allows control over the positioning and number of modifications per barrel. Precise positioning of mutations on two of the seven monomers comprising the αHL barrel has only been demonstrated after extensive engineering in which two monomers were genetically linked [48]. Further, the use of αHL in cell-based applications is limited as overexpression compromises lipid membranes, leading to cell death. In vivo, *S. aureus* secretes αHL as a toxin to

lyse a variety of host cells, including erythrocytes [49]. In contrast, many porins are abundant in the *E. coli* OM and therefore overexpression is usually tolerated, especially in knockout cell strains [50].

The most abundant *E. coli* porins are Outer membrane protein F (OmpF) and OmpC, which form highly stable trimeric complexes present in $>10^5$ copies per cell [51]. The relative distribution of each is dictated by osmolarity, with OmpC being preferentially expressed at high osmolarity [52]. The trimers serve as conduits for the passive diffusion of sugars, ions, and small hydrophilic molecules <700 Da across the OM [47]. Other porins, such as the maltodextrin transporter LamB and the vitamin B_{12} transporter BtuB, are specific to particular solutes [47].

The monomeric *E. coli* porin Outer membrane protein G (OmpG) was previously developed for stochastic sensing [6,54,55]. The WT pore shows frequent gating behaviour, in which the current through the pore fluctuates from the open state to nearly 0 pA, the closed state, and remains mostly closed below neutral pH or when the applied potential is higher than ±100 mV [56]. This noisy background signal complicates analyte sensing, which is observed as a characteristic change in the ionic current flowing through the pore. To mitigate gating, a ‘quiet’ OmpG was genetically engineered by deleting a charged residue that destabilised the barrel and introducing a disulfide bond in an extracellular loop to restrict its conformational flexibility [53]. However, the innate gating events, although less frequent, were still present even after mutagenesis (Fig. 1.5). Further, the cysteine residues would complicate experiments that use reversible cysteine chemistry for the attachment of sensing elements or for single molecule chemistry studies, of which there have been many [10,57–59].

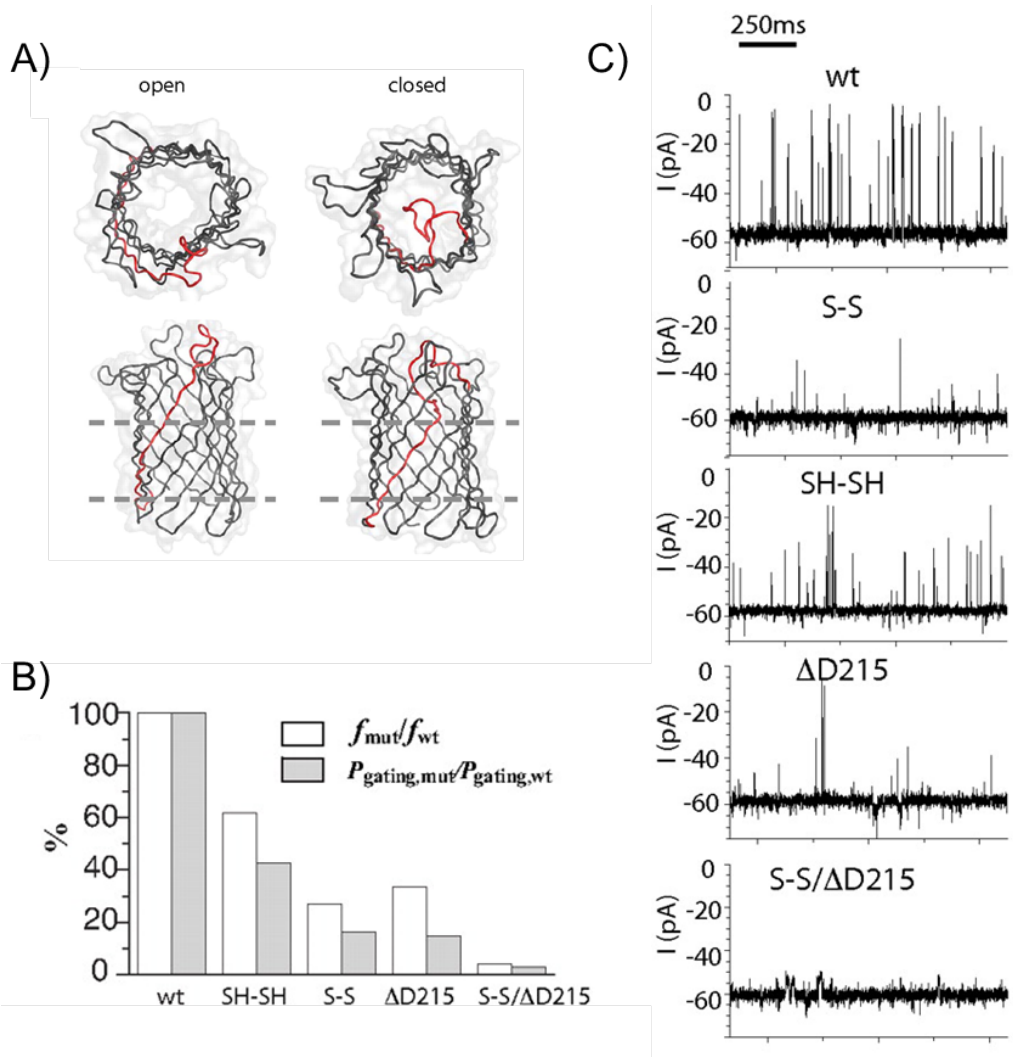


Figure 1.5 Engineering a quiet OmpG for stochastic sensing. **(A)** Top and side view of OmpG in the open (PDB: 2IWV) and closed (PDB: 2IWW) state. L6 and strand β12, thought to effect gating, are highlighted in red. **(B)** Representative current traces of WT OmpG and mutants designed in L6 (S-S) and β12 (ΔD215) to produce a quieter pore. **(C)** Comparison of gating activities of WT and mutant OmpG. The gating frequency (*f*) and probability (P_{gating}) of OmpG mutants compared to WT OmpG (defined as 100%). Reproduced from [53].

1.2.2 Outer membrane protein F

This thesis explores OmpF as a new nanopore sensing platform. The X-ray crystal structure of OmpF reveals a trimer composed of three identical 16-stranded anti-parallel β-barrels that are embedded in the membrane [60]. Eight short turns (T1-T8) connect the β-strands at the periplasmic face of the OM, whereas eight long

loops (L1-L8), which protrude over one nanometre from the extracellular surface, bridge the strands at the extracellular face (Fig. 1.6 A).

Extensive hydrophobic interactions between the barrels and latching of L2 onto the adjacent monomer stabilises the trimer [60,61] (Fig. 1.6 B). This extensive network of interactions creates a stable trimeric structure that is resistant to SDS. The conserved internal L3 loop constricts the size of the pore at about half the barrel height to an ellipse of diameter 7x11 Å and determines the cut-off size for diffusion (600-700 Da). The diameter increases abruptly on either end to roughly 15 x 22 Å, giving rise to an hourglass shape with a slightly narrower external vestibule (Fig. 1.6 B) [60]. Three arginine residues (R42, R82, R132) inside the barrel and two acidic residues (D113, E117) on L3 form the constriction zone and create a transverse electric field strong enough to orient water molecules and establish specific pathways for molecules of different charges through the porin, including antibiotics [62,63]. This property underscores the importance of electrostatic interactions even in a passive diffusion channel.

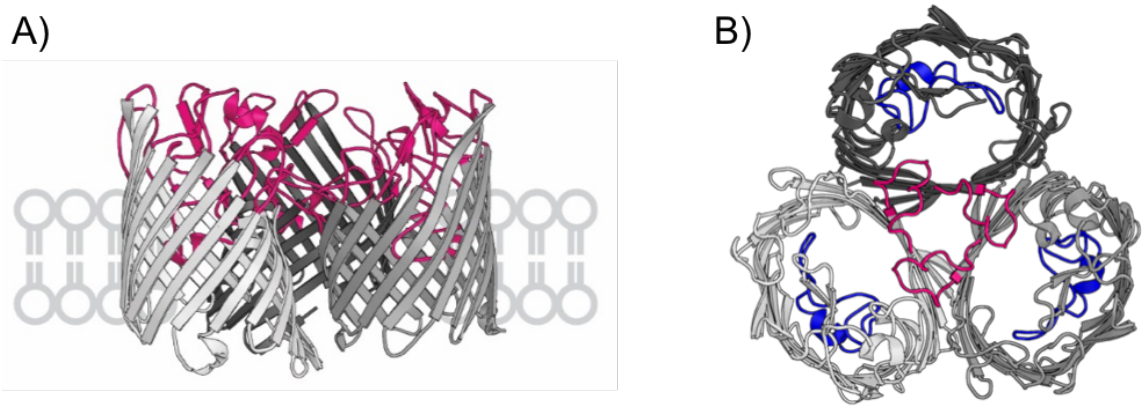


Figure 1.6 Structure of OmpF. **(A)** Side view of an OmpF trimer inserted into a bilayer, with the extracellular loops coloured in pink. **(B)** Top view showing the trimer-stabilising L2 (pink) and L3 (blue), which forms the pore constriction. The structure was visualised in PyMOL (PDF: 2OMF).

OmpF inserts easily into membranes for electrophysiological characterization. Single-channel PLB recording has been used to study the properties of the trimer, including ion transport [64,65], antibiotic permeation [66–68], and the translocation of potent antibacterial peptides known as colicins [69]. The per-trimer conductance in 100 mM KCl at neutral pH is ~1 nS [70]. Importantly, the pore exhibits minimal background noise or gating at applied transmembrane voltages of up to ~200 mV, making it suitable for nanopore sensing. However, the PLB studies carried out to date did not identify the orientation of OmpF, which can insert into the bilayer in either orientation [65,67]. This is a requisite for interrogating directional transport phenomena with molecular precision.

1.3 Overview

This thesis describes methods to engineer the OmpF nanopore for stochastic sensing. In Chapter 2, a method of determining the orientation of OmpF in a PLB by targeted covalent modification of cysteine mutants is presented. The modification reaction was carried out on single mutant trimers reconstituted in a PLB and monitored at the single-molecule level. The side of the bilayer on which the reaction occurred revealed the orientation of OmpF, which was linked to the porin's electrical asymmetry or current-voltage (I-V) curve. Orientation was then easily determined at the start of subsequent experiments by measuring an I-V curve. This approach could be applied to determine the orientation of other porins and channels [70].

In Chapter 3, a protocol for in vitro expression of OmpF using commercially available cell-free systems is described. Using this protocol, mutants were produced in a few days by following a relatively simple procedure. The protocol was further developed to generate heterotrimers, in which one or two of the three

barrels contains a unique modification. This allowed the construction of OmpF with only one sensing element compared to three in the WT homotrimer, which would make interpretation of the current read-out straightforward for single-molecule sensing applications.

Finally, a methodology for designing and screening a peptide library genetically inserted into the extracellular loops of OmpF for protein-binding variants is developed in Chapter 4. A library of nanopores would allow sensors to be developed without prior knowledge of ligands or rationally designed protein engineering steps, thereby expanding the scope of what can be detected by stochastic sensing. Future developments and applications of these three chapters are discussed in Chapter 5.

2

Orientation of OmpF in planar lipid bilayers

2.1 Introduction

Protein orientation in lipid membranes dictates biological activity and intermolecular interactions. The in vivo orientation of membrane proteins has been deduced by a variety of experimental methods including chemical modification [71], fluorescence [72], spin-labelling [73,74], X-ray crystallography [75], and neutron scattering [76], in addition to computational methods [77]. Complementary in vitro studies of channels and pore-forming proteins, such as the bacterial porins, have provided insight into membrane transport. PLB electrophysiology is one such in vitro technique which offers single-molecule resolution in real time (Section 1.1). The PLB technique has been widely used to study bacterial porins [78–83], in particular OmpF which is abundant in the OM and implicated in antibiotic resistance [35,67,84–88]. However, the orientation of porins reconstituted in these non-physiological bilayers is not obvious and often not considered.

This chapter reviews previous efforts to determine the orientation of porins in a PLB. A protocol is then presented for the unambiguous determination of

OmpF orientation in a PLB under a variety of conditions. The protocol should be generally applicable to porins and other transmembrane proteins.

2.1.1 OmpF reconstitution in PLBs

The trimeric OmpF porin is abundant in the bacterial OM and forms one of the main uptake routes of antibiotics and antibacterial peptides, known as bacteriocins [89]. Because of its clinical relevance, single-channel PLB recording has been used to better understand molecular transport across the porin and its general properties, including ion selectivity and voltage sensitivity [81,90,91]. Detergent-solubilised OmpF can be reconstituted into artificial bilayers as a functional trimer of ion conducting pores, consistent with the in vivo function [83].

The orientation of OmpF in the *E. coli* OM was established in 1993 using antibody labelling and 3-D reconstruction of crystalline porin patches: The longer loops are extracellular, and the short β-turns are periplasmic [92]. Since then, numerous PLB recordings [35,65,87,88,93,94] have been performed without a clear understanding of orientation, i.e. there are two types of observed I-V asymmetry [65], but it is not known whether the periplasmic or extracellular side penetrates the bilayer first for each type (Fig. 2.1). This information gap renders the molecular interpretation of kinetic studies thus far only speculative. The asymmetries are referred to as positive, or P-OmpF, and negative, or N-OmpF, in this chapter, depending on whether conductance—which is proportional to ion flow through the channel—is higher at positive or negative voltages, respectively. Voltages were applied to the trans side, opposite to the side of OmpF addition (cis-side).

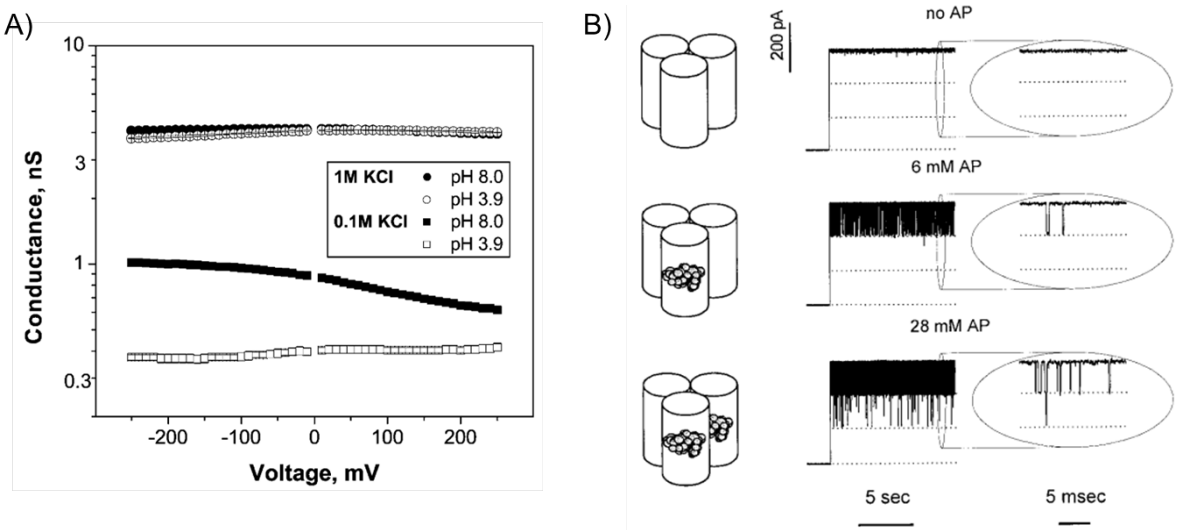


Figure 2.1 PLB recordings of OmpF. **(A)** The conductance of OmpF at different salt concentrations and pH reproduced from [65]. In this case, potential was determined as positive when it was greater at the side of OmpF addition (cis-side). **(B)** Time-resolved interactions of ampicillin (AP) molecules with a single OmpF trimer in a PLB. Without AP, OmpF exhibits a stable open current. Upon addition of increasing concentrations of AP, one or two of the three pores are transiently blocked by the antibiotic. Whether AP was added to the extracellular or periplasmic side of OmpF is not addressed. Reproduced from [87].

2.1.2 Methods to determine porin orientation in a PLB

When reconstituted into a PLB, OmpF has been found to predominantly exhibit a P-OmpF I-V curve [65,67]. The orientation that corresponds to this type of I-V asymmetry was previously inferred from antibiotic binding experiments with enrofloxacin, a fluoroquinolone antibiotic. Based on molecular dynamic simulations and mutagenesis studies, enrofloxacin is thought to have two binding sites within OmpF: one in the periplasmic vestibule and one in the extracellular vestibule [35]. The two vestibules are separated by a constriction site located halfway up the β-barrel. A crystal structure of OmpF in the presence of magnesium (1 M $MgCl_2$) showed an additional electron density around the constriction site representing a water-coordinated Mg^{2+} atom [95]. This coordination site overlaps with the

enrofloxacin binding site in the periplasmic vestibule [88], so it was assumed that OmpF binding would be altered in the presence of the ion.

In single-channel PLB recordings with buffer lacking Mg^{2+}, the interaction of OmpF with enrofloxacin was independent of the side of addition, showing discrete, step-like decreases in the open pore current in either case (K_d = 10 mM) [35]. However, when $MgCl_2$ was added to the recording buffer, OmpF interacted with enrofloxacin in a side-dependent manner. When the antibiotic was introduced to one side of the bilayer, the current traces showed fast flickering interactions, which were unresolvable as discrete events and manifested as a stable decrease in the open pore current and an increase in noise (Fig. 2.2 A); when the antibiotic was introduced to the opposite side of the bilayer, longer discrete interactions, similar to those observed in the absence of magnesium, were resolved (Fig. 2.2 B) [67].

From this behaviour, it was reasoned that the coordination of Mg^{2+} near the periplasmic entrance hinders enrofloxacin binding. Therefore, the side of addition that produces the fast flickering interactions in the presence of magnesium must represent the periplasmic side of OmpF. Based on these results, the most plausible orientation for the P-OmpF I-V asymmetry was assigned, in which the porin predominantly inserts into the bilayer with the periplasmic end first, leaving the extracellular loops in the compartment from which it was introduced. However, this indirect approach is predicated on the assumption that OmpF indeed contains only one magnesium binding site, and that the location of the ion in the crystal structure applies to the electrophysiological recordings in the study, which used 200-fold less magnesium. It also assumes that the enrofloxacin binding sites and kinetics have been correctly predicted by simulations.

The orientation of other bacterial porins such as LamB [13], OmpG [14], and mitochondrial VDAC [15] have also been indirectly inferred from side-dependant interactions with sugars, reducing agents, and by conductance changes in response to applied potentials, respectively. These approaches are not generally applicable.

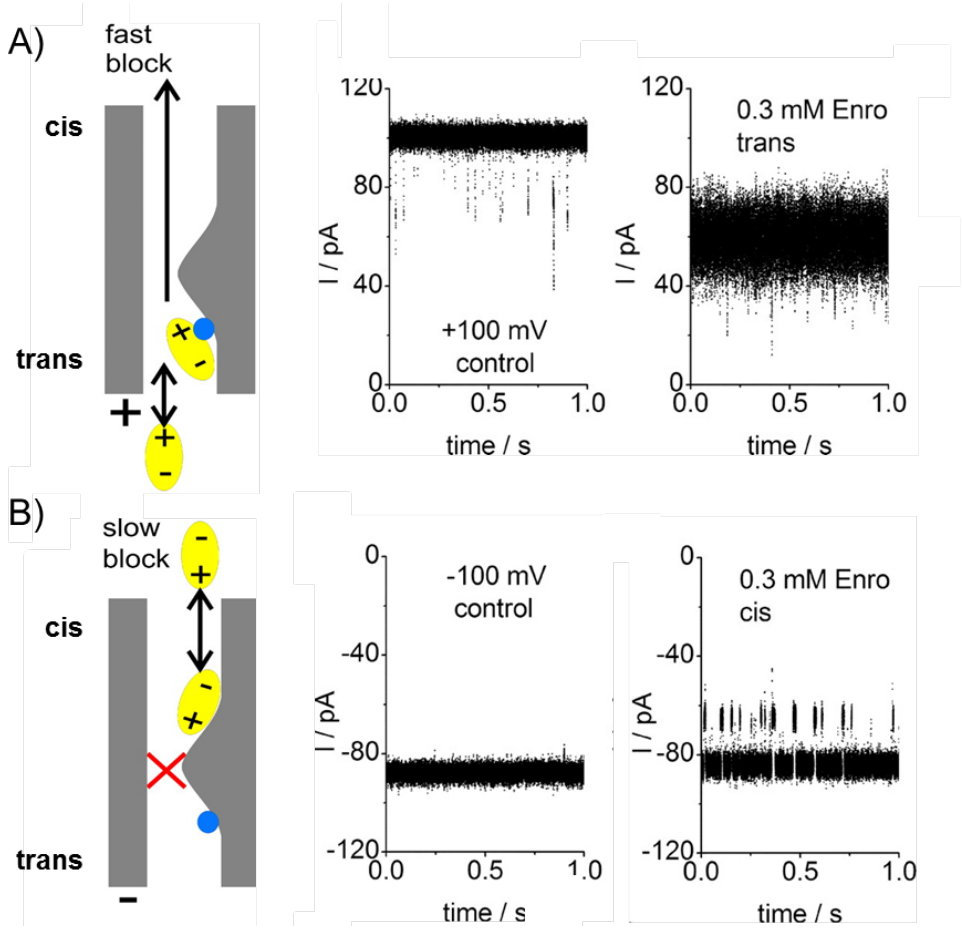


Figure 2.2 Representative current traces of OmpF interactions with enrofloxacin (yellow oval) in the presence of 5 mM Mg^{2+} (blue circle). **(A)** Addition of enrofloxacin to the trans side of the bilayer at +100 mV produces fast flickering interactions, which manifest as a decrease in current and an increase in noise. **(B)** Addition of enrofloxacin to the cis side of the bilayer at -100 mV produces discrete, step-like interactions. Reproduced from [67].

LamB, a maltodextrin transporter porin found in *E. coli* with a similar structure to OmpF, has been proposed to insert with the periplasmic side first in PLB studies, opposite to its insertion in vivo [78,80,96]. The orientation of LamB was suggested from the effects of major loop deletions and point mutations on sugar binding in

single-channel PLB recordings [79]. However, earlier studies carried out with bacteriophage-λ, known to interact exclusively with the extracellular loops, showed the opposite orientation [97]. This disagreement illustrates the challenges of finding a reliable protocol to probe the orientation of reconstituted proteins.

Overall, no direct, reproducible protocol has been established to determine the orientation of OmpF or other porins reconstituted into a PLB. In order to employ the technique to investigate directional transport through porins or to develop new sensors (Chapter 4), orientation must be explicitly assigned at the start of an experiment.

2.1.3 Chapter overview

This chapter outlines a protocol for determining the orientation of OmpF in a PLB based on targeted covalent modification of single-cysteine mutations introduced in either the extracellular or periplasmic loops. The mutants were reconstituted into a PLB, and sequential addition of bulky thiol-directed PEG reagents was used to reveal the side of the bilayer to which each cysteine was exposed. Reaction between OmpF cysteine thiols and the PEG reagent during electrical recordings caused a step-wise decrease in conductance due to the perturbation of ion flow through the porin. By correlating the results of the modification experiments with measurements of I-V curve asymmetry or the side-dependent binding of the antibiotic enrofloxacin, OmpF orientation could be quickly determined in subsequent experiments under a variety of conditions. This general and direct protocol for establishing orientation can be applied to a variety of transmembrane proteins and will allow the precise interpretation of past and future studies of antibiotic permeation and protein translocation through OmpF.

2.2 Characterisation of WT OmpF in PLBs

Single-channel recordings were used to characterise WT OmpF insertion into PLBs. All experiments were carried out in 100 mM KCl, 20 mM phosphate buffer pH 7.0 in symmetric DPhPC bilayers unless otherwise noted. OmpF stored in 1% (w/v) *n*-octyl-β-D-glucopyranoside (βOG) was diluted into the cis compartment buffer at an applied potential of ±(200–300) mV to a final βOG concentration of <0.001%. Upon insertion of an OmpF porin into the bilayer from the cis compartment, one of two types of electrical asymmetry, P-OmpF or N-OmpF, was noted in the I-V curve (Fig. 2.3 A).

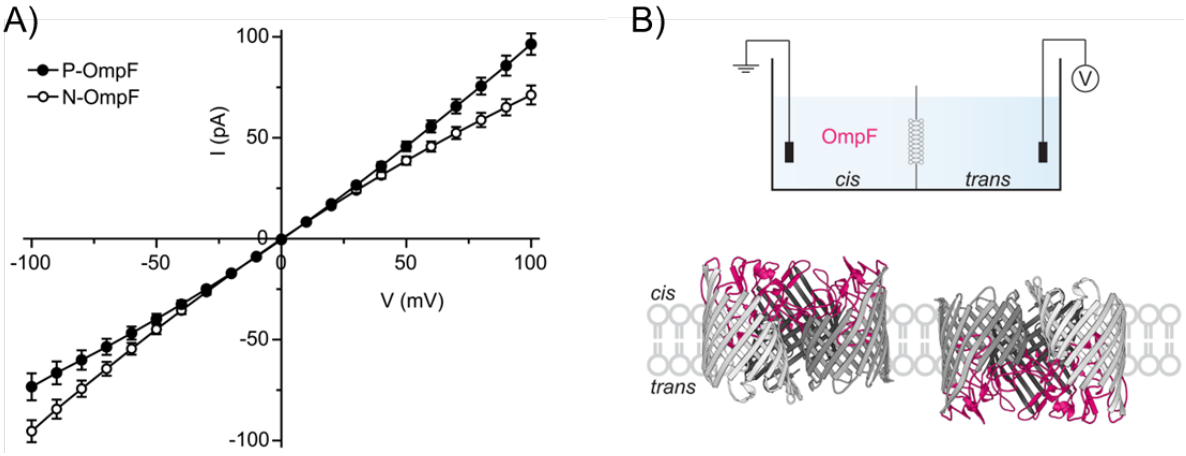


Figure 2.3 OmpF inserts into PLBs in two orientations. **(A)** WT OmpF trimers were reconstituted into planar DPhPC bilayers by addition to the cis side of the bilayer. Bidirectional insertion is evidenced by the two types of single-channel I-V curve. The conductance at positive applied potentials is larger than that at negative potentials for the P-OmpF orientation (filled circles), whereas in the N-OmpF orientation conductance is higher at negative potentials (open circles). Data are mean unitary conductance values (±SD) for OmpF trimers (P-OmpF, n=13; N-OmpF, n=12) in 100 mM KCl, 20 mM potassium phosphate pH 7.0. **(B)** OmpF added to the cis compartment inserts into the bilayer with either the long, irregular loops (pink) or the short turns first, thus producing one of the two types of observed I-V curve. In vivo, the turns are in the periplasm, and the long loops face the extracellular milieu.

This asymmetry, which has been previously reported [65], was taken to represent the two orientations of OmpF in a symmetric PLB, with either the periplasmic turns or the extracellular loops inserting first (Fig. 2.3 B). The electrical asymmetry, which is pronounced at higher potentials, must originate from the structural asymmetry of OmpF, most importantly the distribution of charged residues inside the porin lumen [65,98]. It is important to reiterate that the relationship between the orientation of a pore and its electrical asymmetry depends on the electrophysiology set-up; here, potential was applied on the trans side. If potential was applied on the cis side, then the electrical asymmetry observed for each orientation would be swapped.

Previous studies reported that the P-OmpF orientation occurs from 67-95 % of the time in single-channel PLB recordings using DPhPC bilayers [65,67,98]. In this study, only 52% of the porins (n=25) inserted in the P-OmpF orientation. This may be due to differences in reconstitution protocols, such as voltage applied and buffer composition. The I-V curves of P- and N-OmpF WT porins were linked by a 180° rotation, as expected for proteins in opposite orientations in a symmetric bilayer. The average current through a P-OmpF trimer was 98±6 pA at +100 mV (n=13), which is equivalent to a unitary (per-trimer) conductance value of 0.98 nS, in agreement with previously reported values under similar buffer conditions (0.9–1.0 nS) [65,87]. The rectification ratio, which was calculated as the current at +100 mV divided by the current at −100 mV ($I_{+100\ mV}/I_{-100\ mV}$), of P-OmpF porins was 1.3±0.1. Porins in the N-OmpF orientation (n=12) had a rectification ratio of 0.75±0.06. The presence of trimers was confirmed by voltage-induced gating, in which OmpF stochastically closes in three steps at applied potentials exceeding +200 mV (Fig. 2.4) [99].

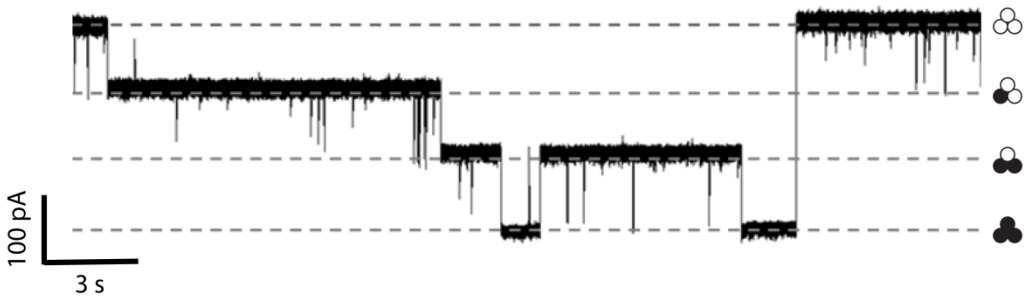


Figure 2.4 Step-wise closure (gating) of OmpF at +250 mV. A maximum of three steps was observed, indicative of a trimer. The top dashed line represents the open pore current, while the bottom dashed line represents all three monomers closed (~0 pA).

2.3 Characterisation of OmpF cysteine mutants in PLBs

Five cysteine mutants were expressed (E29C, S53C, L147C, D221C, N246C) to probe the orientation of OmpF in a PLB. The mutations were chosen based on the X-ray crystal structure of OmpF (PDB: 2OMF), such that the cysteine residues would be exposed to one of the two compartments (cis or trans) upon insertion into a bilayer. Each trimer contains three cysteine residues, as WT OmpF has none. Mutations E29C and N246C are on the extracellular loops (loops 1 and 6, respectively); S53C and L147C are on the periplasmic turns (β-turn 2 and 4, respectively); and D221C is just inside the barrel wall at the periplasmic mouth (β-turn 6).

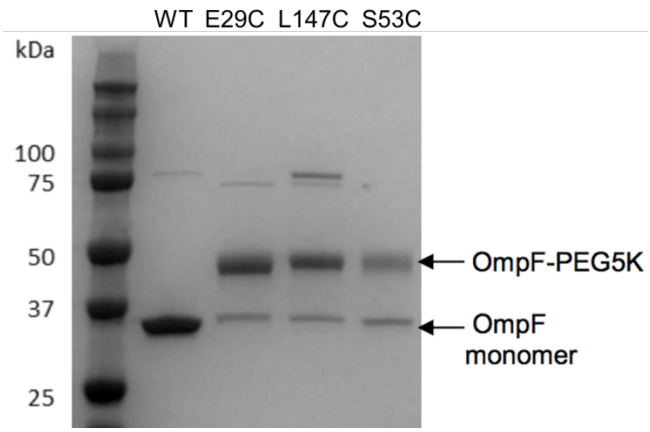


Figure 2.5 SDS-PAGE of OmpF E29C, L147C and S53C after reaction with mPEG-MAL-5K. All lanes contain OmpF (3.5 µg) reacted with mPEG-MAL-5K (1 mM) prior to denaturation by heating at 95°C for 5 min in 2% SDS to obtain monomers. The attachment of PEG-5K produces a gel shift (band marked OmpF-PEG5K). Unreacted WT OmpF monomers appear as a single band, and a faint band of folded trimers is also visible around 75 kDa.

Of the five mutants tested, two (S53C, L147C) exhibited no change in conductance when any of the thiol-directed reagents were introduced into the recording chamber (Section 2.4). This may be attributed to the peripheral location of the cysteine side chains and their orientation away from the conductive pathway. L147C and S53C did, however, react with the thiol-directed reagent methoxy poly(ethylene glycol)-maleimide (mPEG-MAL) in bulk solution, as determined by SDS-PAGE, which proved that the cysteine thiols were reactive (Fig. 2.5). The remaining three mutants (N246C, E29C, D221C) were successfully labelled during single-channel PLB recordings and will be further discussed (Fig. 2.6 A).

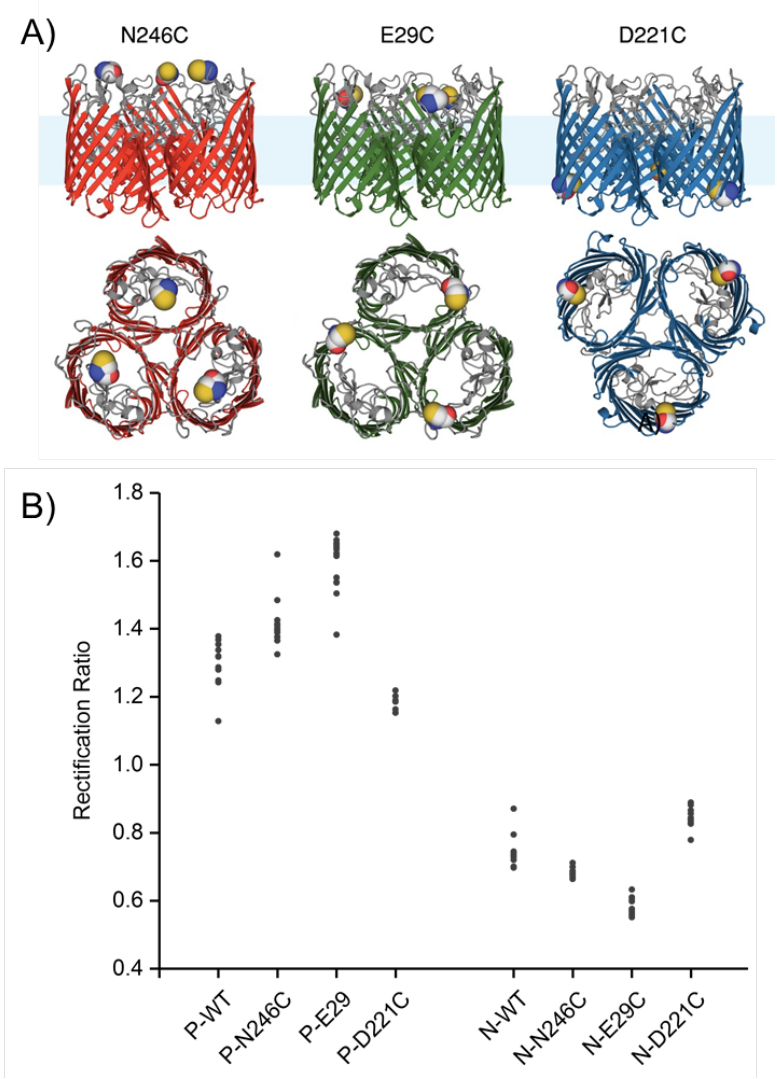


Figure 2.6 OmpF cysteine mutants used to determine OmpF orientation. **(A)** Structures of N246C, E29C, and D221C generated in PyMOL based on the WT structure (PDB: 2OMF), with the extracellular loops in grey. The extracellular cysteines of N246C and E29C and the periplasmic cysteines of D221C are depicted as spheres. **(B)** Scatter plot of rectification ratios from individual WT and OmpF mutant porins in a PLB. Each point represents the rectification ratio (I_{+100mV}/I_{-100mV}) of a single porin in the P- or N-OmpF orientation.

Single-channel recordings confirmed that N246C, E29C, and D221C exhibited bidirectional insertion (both P- and N-OmpF orientations) and that the point mutations did not eliminate the electrical asymmetry (Fig. 2.6 B). For N246C porins, the rectification ratios were 1.4±0.1 (P-OmpF; n=15) and 0.68±0.04 (N-OmpF; n=14); for E29C porins they were 1.6±0.1 (P-OmpF, n=16) and 0.57±0.04 (N-OmpF, n=18); for D221C they were 1.2±0.1 (P-OmpF, n=5) and 0.84±0.03 (N-

OmpF, n=16). The significant variation in single-channel rectification ratios observed with each point mutation underlines the previous finding that the overall charge distribution contributes to porin conductance [98]. This was especially apparent in the case of D221C, where a negatively charged residue inside the β-barrel had been neutralised. It was assumed that each type of cysteine mutant I-V curve represents the same relative orientation as the WT, i.e. a P-OmpF cysteine mutant is in the same orientation as a P-OmpF WT porin.

2.4 Targeted covalent modification

To determine the absolute orientation of P-OmpF and N-OmpF porins, single cysteine mutants were reconstituted into PLBs and labelled with thiol-targeting reagents (Fig. 2.7). First, a single porin trimer was allowed to insert into a bilayer, and an I-V curve was recorded to establish the electrical asymmetry of the porin. The reagent was then introduced on each side of the bilayer in turn; when the reagent was introduced on the side of the bilayer where the cysteine residue was exposed, generation of an adduct was expected to cause a concomitant change in the ionic current through the pore. The site of reaction (cis or trans) could then be correlated to the electrical asymmetry of the channel to establish the absolute orientation of the porin with respect to the I-V curve. All targeted covalent modification recordings were done at ±70 mV.

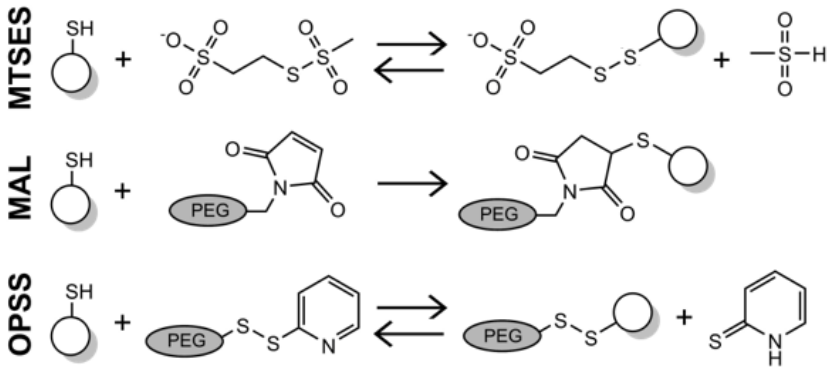


Figure 2.7 Reactions of a cysteine side chain with the thiol-directed reagents used in this study. Formation of three cysteine adducts was expected, consistent with the three cysteines in a trimer.

2.4.1 Modification with MTSES

Cysteine modification experiments were first attempted with the small molecule (2-sulfonatoethyl)methanethiosulfonate (MTSES), which had been previously used to determine the subunit composition of the Staphylococcal leukocidin pore [100]. Surprisingly, reaction between the cysteine thiol and the reagent produced an increase in conductance upon reaction with N246C, E29C, and D221C (Fig. 2.8). This may be because the negative charge imparted by the small adduct (219 Da for the anion) alters ion movement through the pore. Regardless of the side of reagent addition, rapid reaction (~1 min) was seen with all of the cysteine mutants. The reagent was not used further because of its small size, which allows it to equilibrate quickly with the internal solvent in the porin and renders it unsuitable for orientation studies.

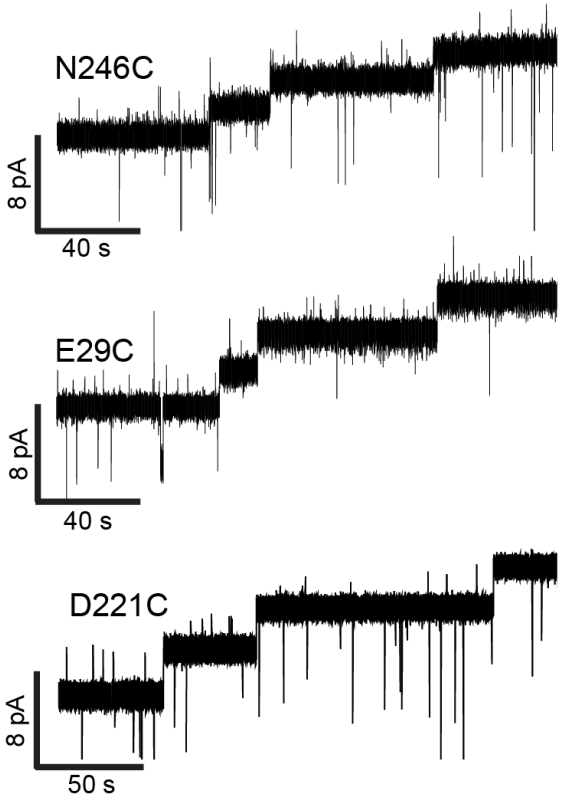


Figure 2.8 Asymmetric addition of MTSES to OmpF cysteine mutants. Reaction of P-OmpF cysteine mutant porins with MTSES (1 mM) added to either the cis or trans side of the bilayer caused step-wise conductance increases (three steps maximum). The average conductance increase per step was 1.9±0.8 pA (n=2) for N246C, 1.1±0.1 pA (n=4) for E29C, and 3.2±0.8 pA (n=2) for D221C.

2.4.2 Modification with mPEG-MAL

The bulky 5 kDa PEG-thiol, mPEG-MAL-5K, previously used to produce a gel shift on SDS-PAGE (Fig. 2.5) was then chosen for labelling to avoid reagent diffusion through OmpF to the opposite side of the bilayer, which would complicate orientation studies. The mass of the flexible PEG polymer is well above the 1.36 kDa mass estimated to divide the regimes of PEG partitioning and exclusion by OmpF [93]. Therefore, a reagent would react with a cysteine mutant only if introduced to the compartment presenting the cysteine thiol group. The maleimide

reagent reacts with the cysteine sulfhydryl group to form a chiral thioether bond, which is irreversible (Fig. 2.7).

Two sets of experiments were performed to deduce the orientation of OmpF cysteine mutants using mPEG-MAL-5K. In one set, mPEG-MAL-5K was first added to the cis compartment and electrical activity was recorded for 10–15 min before the addition of reagent to the trans compartment. In a second set, mPEG-MAL-5K was introduced first into the trans compartment and then to the cis compartment. This was done to reduce bias and ensure the reagent reacted only on one side of the bilayer. For clarity, only the cis-side-first experiment is discussed. The experiments were recorded at +70 mV for P-OmpF porins and −70 mV for N-OmpF porins to maintain current flow in the same direction with respect to the porin. The voltage was chosen such that enough current flowed through the pore to adequately resolve the decrease in conductance caused by PEG adduct formation, but low enough to avoid bilayer destabilisation.

E29C, D221C, and N246C porins in the P-OmpF or N-OmpF orientations reacted with mPEG-MAL-5K, manifesting as stable step-wise decreases in conductance in single-channel recordings (Fig. 2.9). When a decrease in conductance did not occur for >15 min, it was assumed that there was no reaction between the reagent and OmpF, and that the cysteine side chain was therefore not exposed to that side of the bilayer. When mPEG-MAL-5K (1 mM) was introduced to the cis side of the PLB, E29C and N246C porins in the P-OmpF orientation showed a three-step decrease in current; this corresponds to the formation of PEG-5K adducts on all three cysteine residues of the trimer. In contrast, P-OmpF D221C did not react until the reagent was added to the trans side (Fig. 2.9 A). For the N-OmpF orientation, the opposite pattern was observed:

D221C reacted when mPEG-MAL-5K was added to the cis side, whereas reaction with N-OmpF E29C and N246C was only observed when the reagent was added to the trans side (Fig. 2.9 B). The order of addition (trans then cis, or cis then trans) did not affect the side of the bilayer at which the reaction took place. Control experiments with WT OmpF did not produce any stable changes in conductance, regardless of which compartment contained mPEG-MAL-5K (Fig. 2.10).

The change in conductance upon PEG adduct formation varied significantly with the site of the cysteine mutation. The biggest change with mPEG-MAL-5K was for D221C where the mean per-step current drop was 4.9±0.8 pA. For N246C and E29C porins, the mean per-step current drop was 1.2±0.1 and 2.4±0.2 pA, respectively (Table 2.1). This variation correlates with the distance of the cysteine residue from the constriction zone. Residue D221 is just inside the barrel wall and is closest to the constriction zone, whereas N246 is the most peripheral of the mutated residues, at the tip of an extracellular loop. The total (three steps) decrease in conductance after PEG conjugation of all three cysteines in a D221C trimer was ~15 pA (17% of the open pore current). This is comparable to data from previous studies where attachment of a PEG-5K to a peripheral cysteine residue within the αHL barrel produced a ~20% drop in current [101].

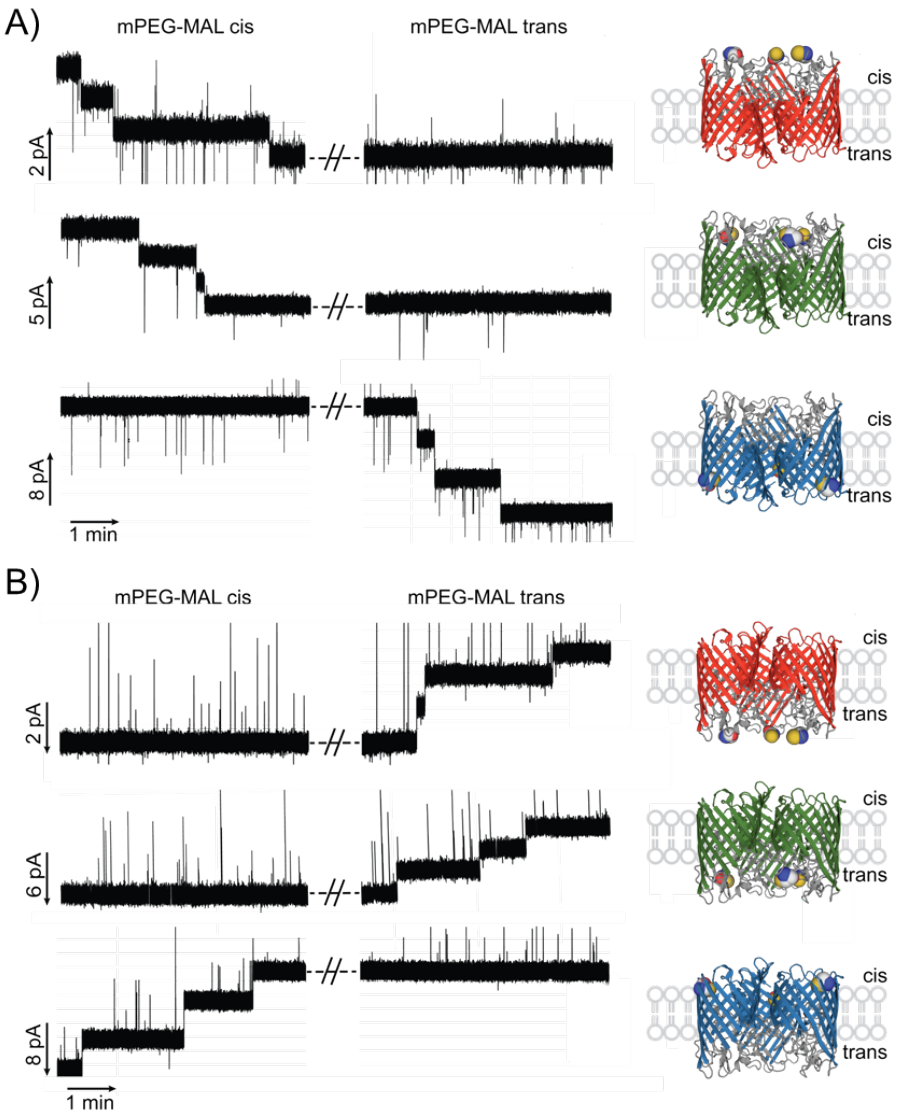


Figure 2.9 Asymmetric addition of mPEG-MAL-5K to N246C (red), E29C (green), and D221C (blue). Single-channel current traces for **(A)** P-OmpF and **(B)** N-OmpF asymmetry porins during sequential addition of mPEG-MAL-5K to the cis and then the trans side of the bilayer. Upon insertion of a single porin into a DPhPC bilayer, an I-V curve was recorded to establish P- or N-OmpF asymmetry. A three-step reaction was observed only when the reagent was introduced into the compartment to which the cysteine side chains were exposed. For the extracellular mutants N246C and E29C, the reaction was observed upon addition to the cis compartment for porins with P-OmpF asymmetry, and the trans compartment for porins with N-OmpF asymmetry. The reverse was true for the periplasmic mutant D221C. The orientation that follows from the asymmetry of the reaction is shown to the right of each trace.

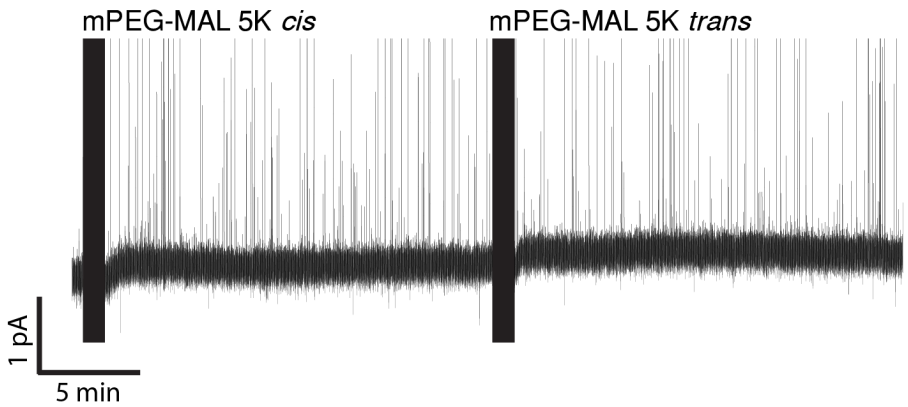


Figure 2.10 Single-channel current trace of WT OmpF in the P-OmpF orientation after asymmetric addition of mPEG-MAL-5K (1 mM) to the recording chamber. No current steps were observed for 20 minutes after addition on either side of the PLB, indicating that no reaction occurred.

Mutant	MAL-5K	MAL-2K	OPSS-5K
N246C	1.2 ± 0.1 pA	1.2 ± 0.1 pA	1.1 ± 0.2 pA
E29C	2.4 ± 0.2 pA	1.9 ± 0.3 pA	2.1 ± 0.8 pA
D221C	4.9 ± 0.8 pA	4.9 ± 0.6 pA	4.7 ± 0.5 pA

Table 2.1 The mean current step decrease (±SD) upon reaction of the PEG reagent with a cysteine thiol group on a trimer in either the P- or N-OmpF orientation at ±70 mV. The mean is derived from all three steps in a reaction observed from two or more independent OmpF trimer recordings.

In order to investigate the effect of PEG mass on the extent of current block and the reaction rate, the 5 kDa PEG was replaced with a 2 kDa PEG (mPEG-MAL-2K) and the targeted covalent modification was repeated (Fig. 2.11). Surprisingly, the average current drop for a single step with the 2 kDa polymer was not significantly smaller than that for the 5 kDa polymer for any of the cysteine mutants (Table 2.1). Presumably, the larger part of both polymers is distant from the porin entrance and has little effect on ion conductance. The smaller polymer should be less hindered at the protein surface, and therefore the rate of reaction with the maleimide of all three cysteine residues in a trimer should be faster.

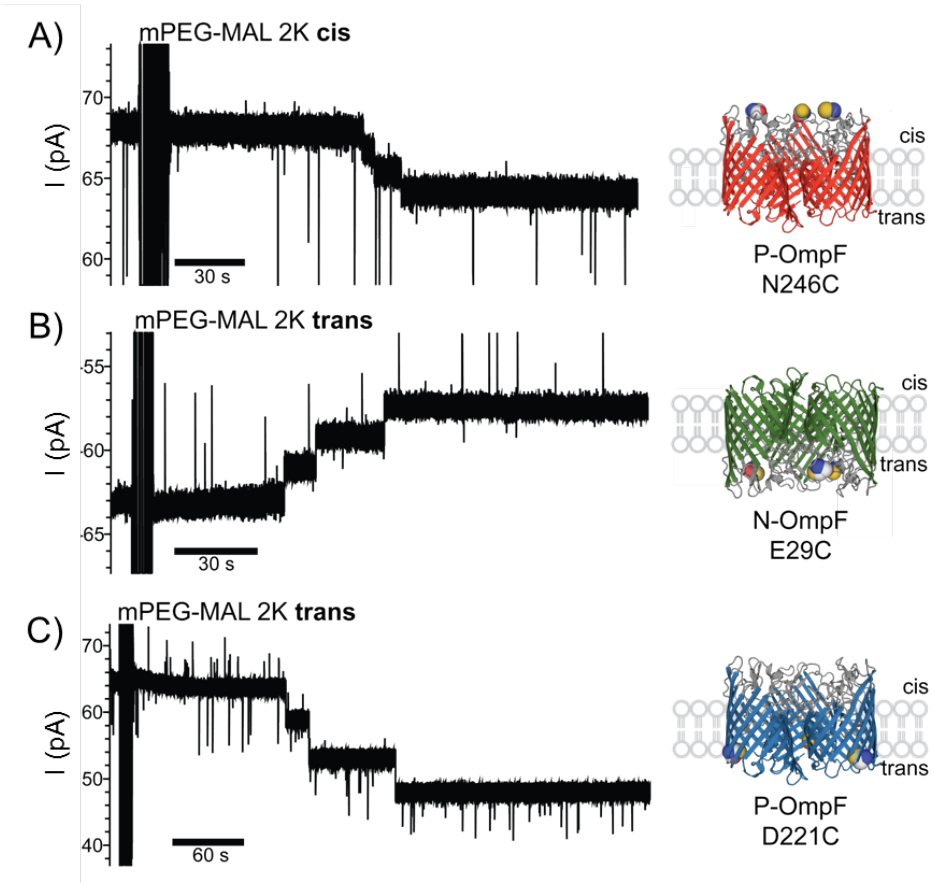


Figure 2.11 Single-channel current traces from P-OmpF and N-OmpF pores after asymmetric addition of mPEG-MAL-2K (1 mM) to the recording chamber. **(A)** N246C pores in the P-OmpF orientation reacted with mPEG-MAL-2K only when it was introduced on the cis side of the bilayer. **(B)** E29C pores in the N-OmpF orientation reacted with mPEG-MAL-2K only when it was introduced on the trans side of the bilayer. **(C)** A P-OmpF D221C pore reacted with mPEG-MAL-2K only when it was introduced on the trans side of the bilayer. Potentials of +70 mV (P-OmpF) and -70 mV (N-OmpF) were applied to keep the current flow in the same direction with respect to the pore.

A faster reaction with mPEG-MAL-2K compared to mPEG-MAL-5K was indeed observed for all cysteine mutants. For D221C, where the difference in reaction rate was most pronounced, the second-order reaction rate constant was $6.9\pm3.2\,M^{-1}s^{-1}$ (n=4) for mPEG-MAL-5K and $21\pm10\,M^{-1}s^{-1}$ for mPEG-MAL-2K (n=3), using data from both P- and N-OmpF porins. Statistical correction for the three cysteines per trimer at the start of the reaction was not made when

calculating the rate constants. These data are similar in magnitude to the rate constants obtained with PEG labelling of αHL [101].

The mPEG-MAL labelling results show that the P-OmpF orientation corresponds to a porin that has inserted with the periplasmic end first, leaving the extracellular loops facing the cis compartment, from which the protein was introduced. It follows that N-OmpF porins insert into the bilayer in the opposite orientation, thus placing the extracellular loops in the trans compartment.

2.4.3 Modification with mPEG-OPSS

The targeted covalent modification experiments were repeated with the reversible reagent methoxy poly(ethylene glycol)-orthopyridyl disulfide (mPEG-OPSS) to confirm the cysteine specificity of the reactions. The OPSS functionality reacts with a cysteine side chain to form a disulfide bond between the protein and the polymer with release of pyridine-2-thione; this bond can be cleaved by a reducing agent such as dithiothreitol (DTT) (Fig. 2.7).

The reversible reaction was observed with all three mutants (Fig. 2.12). N246C and E29C in the P-OmpF orientation reacted with mPEG-OPSS-5K (1 mM) when the reagent was added to the cis compartment (E29C, n=2; N246C, n=2). No reaction was detected after trans-side addition. In contrast, for P-OmpF D221C the reaction was observed only when the reagent was introduced on the trans side (n=2). No reaction was observed after cis-side addition. Subsequent addition of DTT (20 mM) to both sides of the bilayer cleaved the PEG adducts, which was visible in the recordings as a step-wise increase in conductance from the original open-porin level.

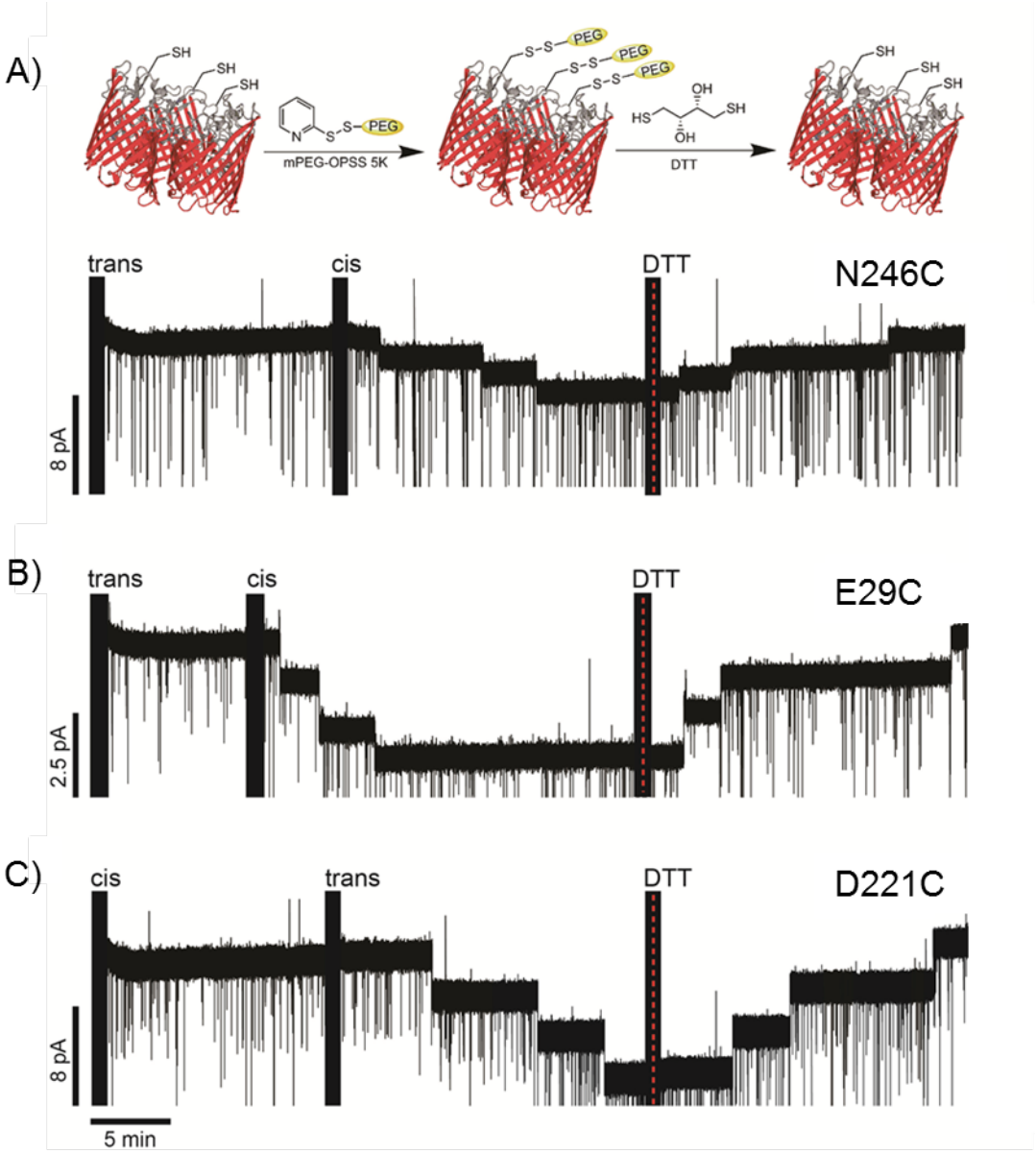


Figure 2.12 Reversible reaction of P-OmpF cysteine mutants with mPEG-OPSS-5K. **(A)** Schematic of the reversible reaction of N246C with mPEG-OPSS-5K and a single-channel trace of the reaction of N246C with mPEG-OPSS-5K added to the cis side followed by treatment with DTT. No reaction was observed when the reagent was added to the trans compartment. **(B)** Single-channel trace of the reaction of E29C with mPEG-OPSS-5K added to the cis side followed by treatment with DTT. **(C)** Single-channel trace of the reaction of D221C with mPEG-OPSS-5K added to the trans side followed by treatment with DTT.

Reaction with mPEG-OPSS-5K produced a mean single-step current drop of 1.1±0.2 pA with N246C, 2.1±0.8 pA with E29C, and 4.7±0.5 pA with D221C (averaged across porins in both orientations). These values are not significantly different from the average current drops observed for the mPEG-MAL reagent (Table 2.1). This is not surprising given that the PEG adducts differ only by the type of linkage to the protein (disulfide versus thioether). The patterns of mPEG-OPSS-5K reaction with E29C, D221C, and N246C in the P- and N-OmpF orientations matched the pattern for mPEG-MAL-5K, thus reinforcing the conclusion on the relationship between the orientation of a porin and its I-V curve.

2.5 Relative orientation of OmpF and cysteine mutants

The OmpF cysteine mutants used in this orientation study had different electrical asymmetries compared to the WT (Fig. 2.6 B). To ensure that cysteine mutants exhibiting the P-OmpF current asymmetry were in the same relative orientation as WT P-OmpF, binding studies with enrofloxacin were carried out. Enrofloxacin is a fluoroquinolone antibiotic that shows distinct interactions with OmpF in the presence of Mg^{2+} depending on whether it is introduced into the cis or the trans compartment during current recordings [67]. An X-ray crystal structure in the presence of 1 M $MgCl_2$ captured a Mg^{2+} ion in the constriction of OmpF [95]. The ion coordination site overlaps with the putative binding region of enrofloxacin in the periplasmic vestibule. All enrofloxacin experiments were therefore carried out in recording buffer supplemented with $MgCl_2$.

Once a single porin had inserted into a bilayer, 5 mM $MgCl_2$ was added to both compartments of the recording chamber. This resulted in a 10-13% drop in conductance across WT OmpF and the cysteine mutants. The Mg^{2+} that is hydrogen bonded to the L3 loop partly occludes the constriction of the porin, and

was expected to result in decreased ion flow [95]. Enrofloxacin (0.3 mM final) was then introduced into the cis compartment and the current was recorded for 10 min at ±100 mV. The solution in the cis compartment was then exchanged with fresh magnesium-containing buffer before the antibiotic was added to the trans compartment and recording was continued.

Different binding patterns were observed for WT porins in the P-OmpF versus N-OmpF orientation after addition of enrofloxacin. Because the antibiotic is zwitterionic, interactions with OmpF also depended on the applied voltage [67]. Pores in the P-OmpF orientation showed fast flickering interactions after trans-side addition at +100 mV and longer discrete interactions after cis-side addition at −100 mV. WT pores in the N-OmpF orientation showed the opposite interaction pattern (Fig. 2.13). This inverted behaviour enforces the hypothesis that the I-V asymmetries denoted as P- and N-OmpF represent pores in opposite orientations. Similar side-dependent interaction patterns were observed for E29C, D221C, and N246C, which proved that the cysteine mutants indeed maintain the same relative orientation in a PLB as the WT for both types of I-V curves (Fig. 2.13). This was expected because the peripheral cysteine mutations do not interfere with the Mg^{2+} or enrofloxacin binding site.

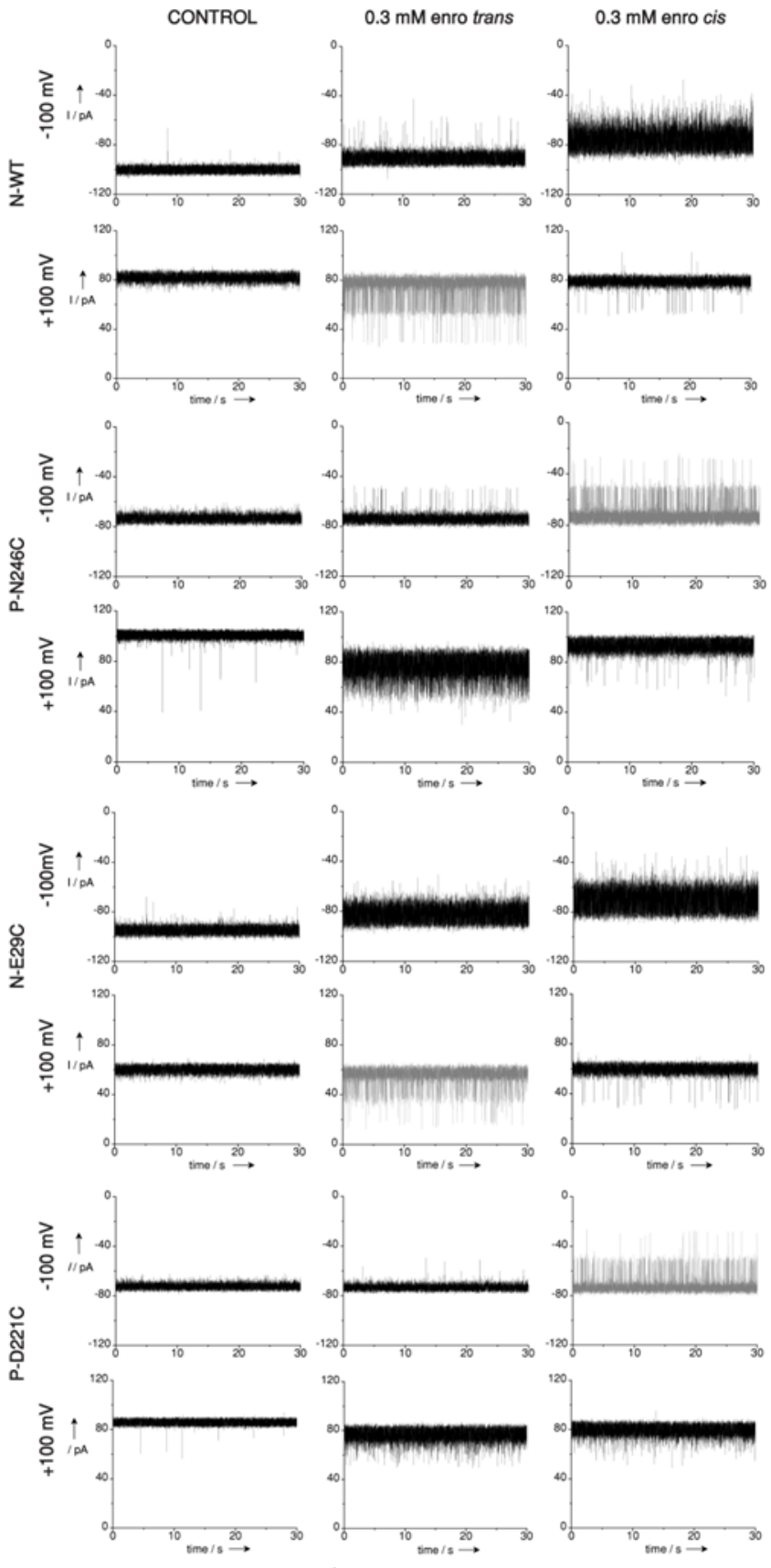


Figure 2.13 Representative current traces from a single OmpF WT or mutant pore in the presence of Mg^{2+} and enrofloxacin. Recordings were done at −100 mV and +100 mV without antibiotic (left column), with 0.3 mM enrofloxacin on the trans side (middle column), and with 0.3 mM enrofloxacin on the cis side (right column). Pores in the P-OmpF orientation showed longer discrete interactions, highlighted in grey, after cis-side addition at -100 mV, whereas pores in the N-OmpF orientation showed discrete binding interaction upon trans-side addition at +100 mV.

2.6 Establishing orientation when the I-V curve is symmetric

The method presented for determining OmpF orientation relies on its electrical asymmetry, which is affected not only by mutations, but also by the salt concentration and pH of the recording buffer [65,98,102]. The reported asymmetry in 1 M KCl (pH 8.0) at ±200 mV was only ~5 % [65]. We repeated single-channel recordings with WT OmpF in 1 M KCl, 20 mM potassium phosphate pH 7.0. For the P-OmpF orientation, the rectification ratio was 1.0±0.01 (n=7) in 1 M KCl compared to 1.3±0.1 (n=12) in 100 mM KCl. (Fig. 2.14 A). From these data, it is clear that the electrical asymmetry of the porin cannot be used as a definitive indicator of orientation in high-salt conditions. It was hypothesised that enrofloxacin binding might also be used to determine OmpF orientation when the buffer conditions diminish the electrical asymmetry of the porin.

Having established the absolute orientation of OmpF and carried out the enrofloxacin experiments in 100 mM KCl, we repeated the enrofloxacin experiments with P-OmpF porins in 1 M KCl. Addition of 5 mM $MgCl_2$ resulted in a 1–2 % drop in unitary conductance during recordings of WT OmpF. For recordings done in 100 mM KCl, fast flickering interactions were observed when enrofloxacin was on the trans side of a P-OmpF porin with a positive transmembrane potential. When the antibiotic was on the cis side, longer discrete interactions were observed, and only at negative potentials. The same pattern of interactions was observed in 1 M KCl (Fig. 2.14 B), although the enrofloxacin dwell-time for the discrete interactions observed at −100 mV with cis-side addition decreased from 5.0±0.8 ms (n=3) in 100 mM KCl to 1.0±0.5 ms (n=4) in 1 M KCl. The average

residence time of enrofloxacin was previously found to be 3 ms at −50 mV in 1 M KCl [35].

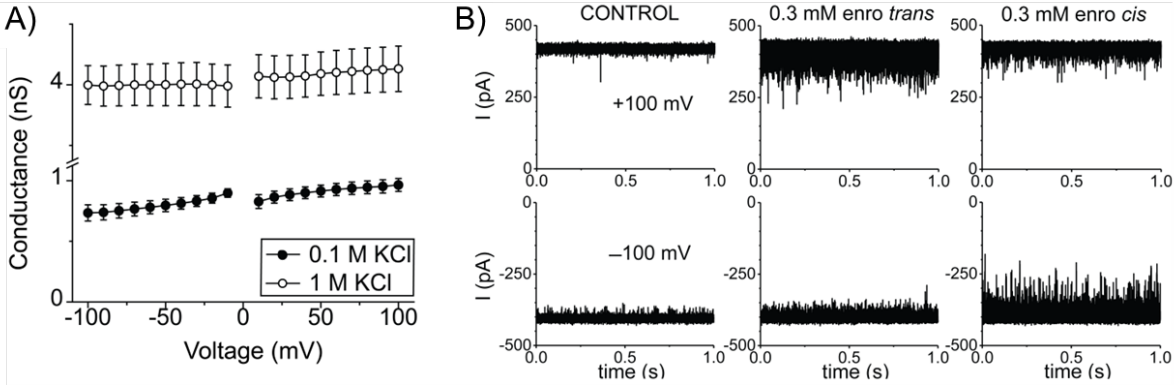


Figure 2.14 Establishing OmpF orientation by enrofloxacin binding. **(A)** Unitary conductance–voltage relationship of P-OmpF in 100 mM and 1 M KCl at pH 7.0. **(B)** Representative traces of single-channel recordings showing asymmetric binding of enrofloxacin in 1 M KCl and 5 mM $MgCl_2$. The antibiotic (0.3 mM final) was first introduced into the trans side of the bilayer, which contained a trimer in the P-OmpF orientation. The trans compartment was exchanged with fresh buffer before subsequent enrofloxacin addition to the cis side. The left panel shows WT P-OmpF in the absence of antibiotic. At +100 mV (top) enrofloxacin caused a significant increase in the current noise, caused by fast flickering interactions, which was stronger when the antibiotic was added to the trans side as opposed to the cis side. At -100 mV (bottom), the trace remained relatively unchanged compared to the control when antibiotic was added to the trans side, whereas discrete blocking events were observed when the antibiotic was added to the cis side. Data were filtered at 10 kHz and sampled at 50 kHz.

In summary, in the presence of Mg^{2+}, OmpF exhibits fast flickering interactions with enrofloxacin presented at the periplasmic side of the porin (trans for a P-OmpF porin) at positive potentials, and longer discrete events when enrofloxacin is at the extracellular side (cis for a P-OmpF porin) at negative potentials. By sequentially adding enrofloxacin to the cis and then the trans side of the bilayer, the orientation of OmpF can be determined in 1 M KCl, in which the electrical asymmetry of the porin is diminished.

2.7 Conclusions

The orientation of OmpF in symmetric DPhPC bilayers has been established. First, it was confirmed that WT OmpF and mutants E29C, D221C, and N246C exhibit two types of I-V curves in PLBs: the current through the porin was either higher at positive applied potentials (P-OmpF) or at negative potentials (N-OmpF). This electrical asymmetry was used as an indicator of the direction of insertion of OmpF, which can penetrate the bilayer from diluted detergent solution with either the periplasmic or extracellular end first.

The connection between a single-channel I-V curve and the orientation of OmpF in the bilayer was directly established by targeted covalent modification of cysteine mutants. The orientation of the extracellular mutants (N246C and E29C) and the periplasmic mutant (D221C) in a symmetric DPhPC bilayer was assigned by asymmetric addition of mPEG-MAL-5K. In the P-OmpF orientation, reaction occurred on the cis side of the bilayer for the extracellular mutants and on the trans side for the periplasmic mutant. The sites of reaction were reversed for porins in the N-OmpF orientation. From these data, it follows that a P-OmpF porin represents an OmpF trimer that has inserted from the cis compartment with the periplasmic turns first, thus leaving the extracellular loops facing the cis compartment (Fig. 2.15). N-OmpF porins are in the opposite orientation. Based on our work, the orientation of an OmpF porin can now be determined from a simple I-V curve measured at the beginning of an experiment. Our results also bolster the conclusions of previous studies, which used single-channel PLB recording of enrofloxacin transport through OmpF to determine porin orientation [35,67].

A caveat to the approach presented in this thesis is that it only works for OmpF mutants and buffer conditions where the electrical asymmetry is

pronounced. A different strategy must be considered in cases where the asymmetry has been diminished, e.g. in high-salt buffer (≥1 M KCl) [65]. One strategy is sequential addition of enrofloxacin to the cis or trans side of the bilayer in the presence of magnesium, as discussed in Section 2.6. The antibiotic shows unique interaction patterns depending on whether it is added to the periplasmic or extracellular side of OmpF. Having established the absolute orientation of OmpF in 100 mM KCl by targeted covalent modification, we were able to determine the orientation of OmpF in 1 M KCl by sequentially adding enrofloxacin first to the cis and then to the trans side of the bilayer, and then comparing the interaction patterns to those observed in 100 mM KCl.

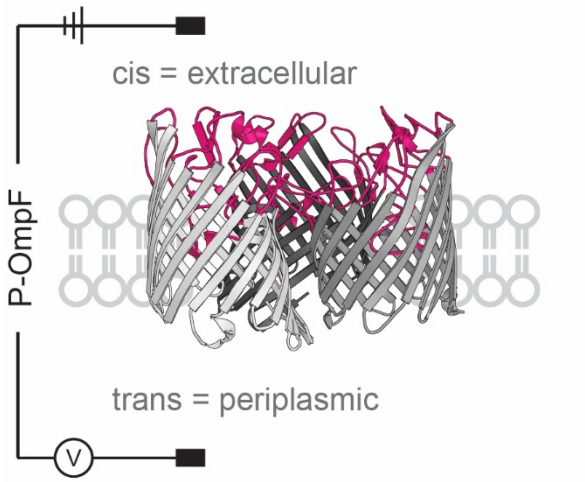


Figure 2.15 Orientation of a P-OmpF porin in a PLB. For the set-up described (OmpF introduced into the cis compartment, potential assigned as positive when it was greater on the trans side), an OmpF trimer exhibiting positive asymmetry has inserted with the periplasmic turns first, leaving the extracellular loops (pink) facing the cis compartment. The opposite orientation applies for an N-OmpF porin.

In cases where the enrofloxacin experiment does not yield clear results, mutation N246C—which does not significantly alter the electrical asymmetry compared to the WT porin—can be introduced. The orientation can then be determined by using reversible mPEG-OPSS-5K chemistry prior to the start of an experiment.

This approach will only be useful if the cysteine does not interfere with the intended outcome. If the buffer conditions do not favour OPSS chemistry, a straightforward option is to establish the orientation of a porin under the optimal conditions described in this thesis before buffer exchange to the desired conditions. It should be emphasised that the strong electrical asymmetry of OmpF persists under most physiologically relevant conditions [65,98], and a simple I-V curve will suffice to determine orientation in most cases.

3

In vitro expression of OmpF and heterotrimer engineering

3.1 Introduction

In vitro or cell-free protein expression was first developed as a controllable expression platform to decipher the link between DNA, RNA, and protein synthesis [103–105]. Since then, the technology has evolved to allow high-throughput production of functional protein for applications ranging from proteomics to artificial cells [106]. Expression in vitro is considerably faster than in vivo because gene transfection, cell culture, and extensive purification is not required. Because there is also no dependence on cell viability, expression of toxic products or proteins prone to misfolding—such as membrane proteins—is facilitated. However, amphiphiles must be supplemented during membrane protein expression to prevent aggregation and provide a hydrophobic environment for co-translational folding. Membrane protein nanopores comprising various genetic and chemical modifications for biosensing have been expressed in vitro [14,48,107].

This chapter begins by reviewing in vitro protein expression and how it has been used to develop nanopores for biotechnology. A protocol is then presented for efficient in vitro transcription and translation (IVTT) of OmpF, which folds into a

trimer with complex inter-barrel interactions. This has not been previously achieved and would allow for rapid production of mutants to study OM physiology, including antibiotic transport. The application of this protocol to construct heterotrimers, in which one of the three barrels contains a different modification, is then described. These engineering approaches might also be useful in synthetic biology, where, for example, OmpF has been used as a communication channel between artificial cells.

3.1.1 In vitro expression systems

There are three types of IVTT systems: eukaryotic cell-extracts [108–111]; prokaryotic cell-extracts [112,113]; and the Protein synthesis Using Recombinant Elements (PURE) kit, which is reconstituted from recombinant proteins [114]. The optimal system for a particular target protein depends on codon usage and the need for post translational modifications, most of which can only be achieved in eukaryotic systems. This chapter focuses on *E. coli*-based expression, which offers the highest yield and simplicity and is the obvious choice for expression of *E. coli*-derived OmpF.

Coupled *E. coli* IVTT kits, in which functional protein is transcribed from template DNA and then translated in a one-pot reaction, have been commercialised. The IVTT reaction can be completed within a few hours and requires minimal purification, depending on the application. A majority of these kits contain an *E. coli* cell extract supplemented with low-molecular weight compounds including nucleotide triphosphates, amino acids, and ATP regeneration components [115]. The reaction terminates when these compounds are depleted, although preparative-scale synthesis can be achieved if the compounds are continuously fed into the reaction [116]. Because of the defined nature of the

system, reaction conditions can be reliably modified with additives such as detergents or lipids, which are generally well-tolerated at low concentrations.

A majority of *E. coli* extract-based kits utilise exogenous T7 RNA polymerase from bacteriophage, placing gene transcription under control of the T7 promoter. The user need only supply high-purity plasmid DNA containing a T7 promoter, ribosome binding site, and the gene of interest. Because of the small expression scale (typically μL), visualisation and quantification is often accomplished by incorporating radioisotopically labelled amino acids such as ^{35}S-methionine. Although commercial cell-free systems are more expensive than traditional cell-based expression, the technology has been widely adopted because it is easy to implement and does not require expensive equipment.

3.1.2 In vitro expression of membrane proteins

Membrane proteins are estimated to account for over 50% of pharmaceutical targets [117]. However, expressing functional protein for analysis and drug screening is challenging. Over-expression in cell-based systems often leads to low yield because of aggregation, misfolding, or host toxicity. These issues stem from leaky gene expression, membrane overcrowding, or a shortage of folding pathway components such as the Sec translocon [118]. Expression in cell-free systems removes the metabolic burden on the host, focusing resources solely on production of the target protein. With the addition of amphiphiles such as detergents, liposomes, or nanodiscs—planar bilayers scaffolded by proteins—successful expression of enzymes and membrane proteins including pores, channels, transporters, and receptors has been achieved (Fig. 3.1) [119–122]. Functional multimeric complexes such as *E. coli* ATP synthase [121] and SecYEG [123] have also been produced by IVTT.

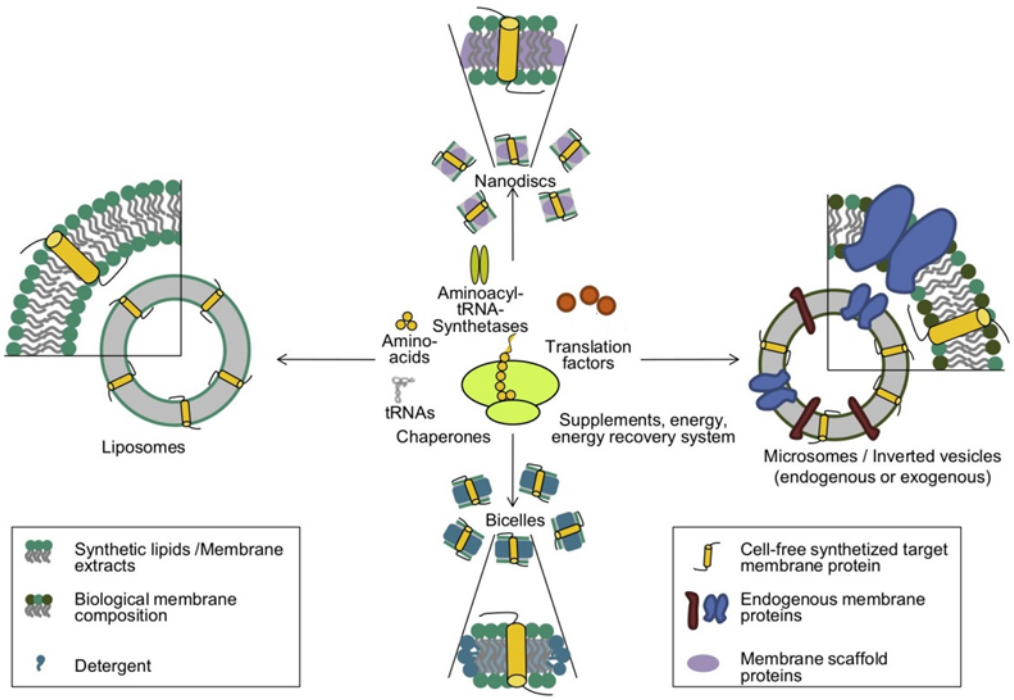


Figure 3.1 In vitro expression of membrane proteins using coupled IVTT kits, which contain all of the components required for transcription and translation. Different membrane mimics can be supplemented to provide a hydrophobic environment for membrane protein folding, including liposomes, nanodiscs or bicelles. Alternatively, native membrane structures, such as microsomes or inverted vesicles, can be used to provide endogenous proteins which assist in co-translational folding. Reproduced from [124].

Other parameters to be considered include the critical micelle concentration (CMC) of detergents used and the liposome size, if applicable, as well as the compatibility of additives with the protein synthesis reaction. For example, two 'mild' detergents with high CMCs, β-OG and CHAPS, have been shown to decrease IVTT yield [125]. Anionic lipids easily aggregate within the cell-free mixture, which contains an abundant amount of magnesium, so they are generally not tolerated [121]. For optimal yield, a range of conditions must be screened each time a new protein is expressed.

3.1.3 Folding and trimerisation of OmpF

In *E. coli,* OmpF monomers are expressed in the cytosol with an N-terminal signal sequence. This sequence is conserved among OM proteins and directs monomers to the inner-membrane SecYEG translocon for transport into the periplasm. Once there, the signal sequence is cleaved and chaperones guide the protein to the β-barrel assembly machinery (BAM) for insertion into the OM (Fig. 3.2) [126,127]. Insertion occurs without energy input, as no ATP is present in the periplasm [128]. Although the mechanism is still unclear, evidence suggests that insertion is facilitated by the membrane thinning effects of BamA—the main component of the BAM complex, a 16-stranded membrane-embedded β-barrel—and lateral opening of BamA between β-strands 1 and 16 [129]. These exposed strands are thought to function as a folding template for the nascent protein, in a mechanism analogous to amyloid nucleation [130]. Whether oligomerisation, as in the case of trimeric OmpF, is directed by the BAM machinery or occurs spontaneously in the membrane is also unknown.

Cell-free expression and trimerisation of OmpF has only been attempted using a rudimentary cell extract which achieved very low trimerisation yield (<10%) [131]. Interestingly, however, purified monomers expressed in *E. coli* have been shown to spontaneously fold into synthetic membranes devoid of chaperones and the BAM complex. After denaturation in urea, trimers have been assembled in vitro using detergents [132,133], liposomes [134], or a liposome-detergent mix [135,136]. A trimer yield of ~70% was obtained when monomers were re-folded in a 1:1 molar ratio of *n*-dodecyl-β-D-maltoside (DDM) to 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) liposomes.

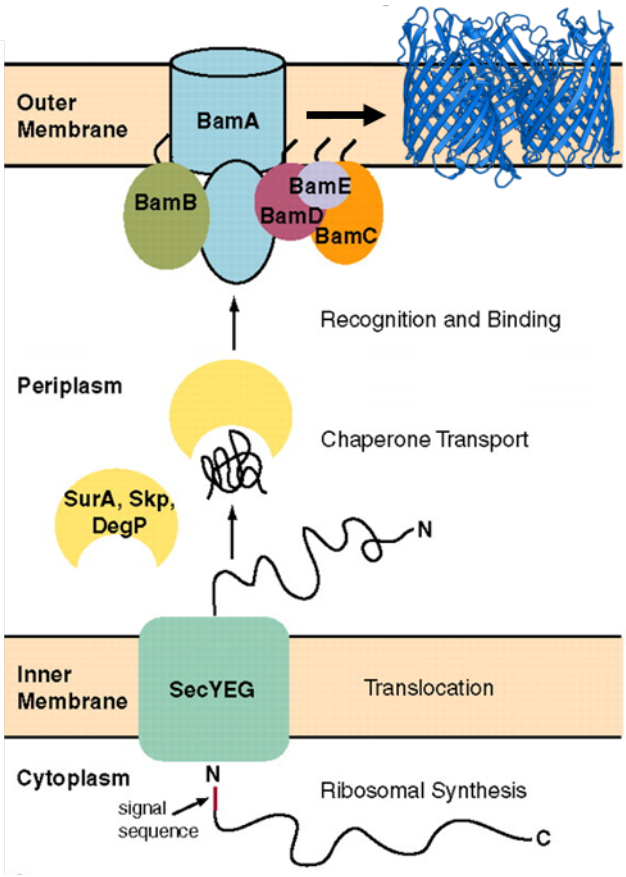


Figure 3.2 OmpF folding in vivo (blue trimer, PDB: 2OMF). Monomers expressed in the cytosol are transported to the periplasm by the Sec translocon. Chaperones then guide the polypeptide to the BAM complex for OM assembly [137].

In a systematic study on OM porin folding in the presence of liposomes only [134], efficient trimerisation of OmpF (up to 100% yield) was reported in short acyl chain lipids, such as 1,2-didecanoyl-*sn*-glycero-3-phosphocholine (10:0 PC) and 11:0 PC. A decrease in yield >50% was reported if the liposomes were composed of 12:0 PC—just one carbon longer. Interestingly, OmpF did not trimerise to any extent when PC-composed liposomes were doped with phosphoethanolamine or phosphoglycerol lipids, which are major components of the inner leaflet of the *E. coli* OM, or in vesicles formed from *E. coli* lipid extracts. A slight preference of small (30 nm) vesicles over large (100 nm) vesicles was noted [134]. The general trend that arises is that in vitro trimerisation is favourable in less stable

membranes caused by (1) shorter, saturated acyl chain lipids, (2) smaller vesicles under high curvature stress, or (3) destabilising agents such as detergents and chaotropes.

These studies demonstrate that the primary sequence of the mature polypeptide contains complete information for the re-folding and trimerisation of OmpF, which spontaneously oligomerises and inserts into membranes on the second timescale [134]. The proposed folding process of OmpF in vitro is shown in Figure 3.3.

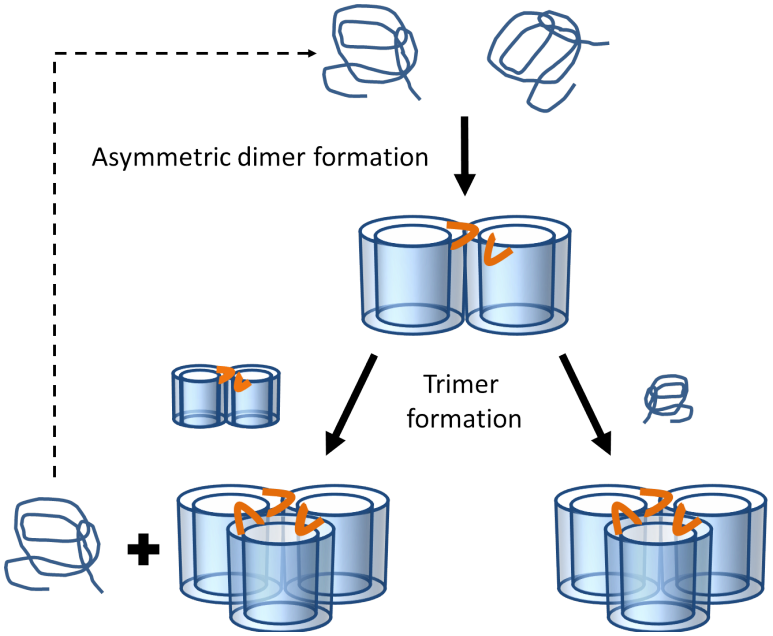


Figure 3.3 Folding pathway of OmpF in vitro. The putative trimerisation mechanism suggests that two OmpF monomers in a random coil configuration first come together to form an asymmetric dimer [132]. Extracellular loop 2, which latches onto an adjacent monomer, is drawn in orange. The asymmetric dimer, presumed to be an on-pathway intermediate, then becomes a trimer with C3 (120°) rotational symmetry by either subunit exchange with another asymmetric dimer or by the incorporation of another monomer. Although the structure of the dimer is unknown, its existence was confirmed by SDS-PAGE [132] and chemical cross-linking [138], and its asymmetric conductance was elucidated by single-channel PLB recordings [132].

3.1.4 In vitro expression of heteromeric nanopores

Nanopore sensors based on αHL have predominantly been heteroheptamers, i.e. only one of the seven monomers forming the β-barrel is modified with the sensing unit [139]. This prevents destabilisation of the pore by introducing one mutation, compared to seven in the homoheptamer. It also avoids the occurrence of parallel binding events during current recordings, which could complicate interpretation of the ionic current trace. Heteroheptamers have been made by expressing both WT and modified monomers containing a polyanionic tag by IVTT and then oligomerising different ratios of the two constructs on rabbit blood cell membranes [140]. This produced up to eight different types of pores, containing from zero to seven modified subunits, which exhibited altered electrophoretic mobilities and could be separated by SDS-PAGE (Section 3.1.6).

In a similar vein, it would be advantageous to be able to selectively change the nature of each OmpF barrel, i.e. to make heterotrimers, which could be modified to include only one binding site per trimer, or different binding sites in each of the three barrels for multiplexed sensing with one nanopore. The binding of different substrates could then be distinguished from the current signature.

OmpF has been engineered to act as a protein gate in artificial organelles for transmembrane communication [141–145]. Controlled substance release was achieved by chemically or genetically modifying OmpF to be pH sensitive [141,142] or by attaching a molecular cap that is cleaved under reducing conditions [145]. Heterotrimers would expand the bioengineering toolbox by allowing up to three different functionalities within one trimer, which could be applied to produce a dialled response based on different inputs. Unexplored methods of controlling release include covalently blocking the pore with

temperature-responsive peptides that contract to allow ion flux through the pore upon an increase in temperature [146], or attaching peptides, bulky polymers, or other blocking moieties in the barrel lumen using reversible chemistries.

3.1.5 Separation of nanopores by SDS-PAGE

Chromatographic separation methods require large amounts of protein and are therefore not suitable for IVTT reactions. Previously, αHL heteroheptamers were separated by SDS-PAGE through the addition of charged tags comprising eight aspartate residues to the C-terminus of modified monomers. The additional charge increased the electrophoretic mobility of the heptamer proportionally to the number of tags; i.e. αHL with seven charged tags migrated the fastest and αHL WT migrated the slowest. The desired heteroheptamers with only one modified subunit could then be extracted from the gel and used directly in PLB recordings (Fig. 3.4).

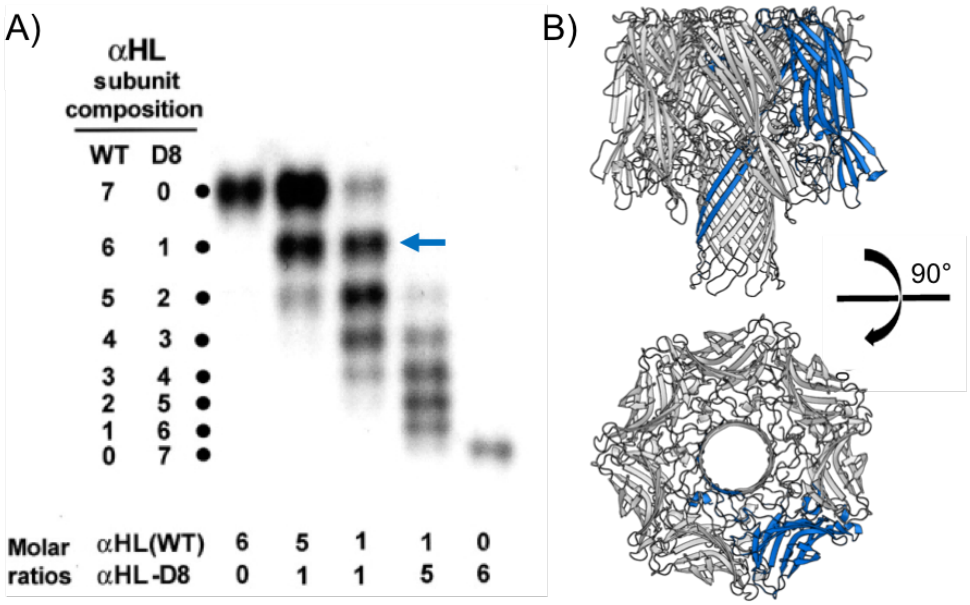


Figure 3.4 (A) Heteroheptamers generated from WT αHL and αHL-D8. The blue arrow represents a heteroheptamer with one D8 tail. Reproduced from [140]. **(B)** Structure of an αHL heteroheptamer with one modified subunit in blue, commonly used for stochastic sensing. The structure was visualised in PyMOL (PDB: 7AHL).

The faster migration of tagged αHL is surprising considering that protein migration on SDS-PAGE, in which negatively-charged SDS purportedly coats the polypeptide thus masking any intrinsic charge, should in theory be slower with increasing mass. However, membrane proteins often exhibit anomalous gel migration, especially those resistant to SDS which remain folded during electrophoresis (Fig. 3.5) [147].

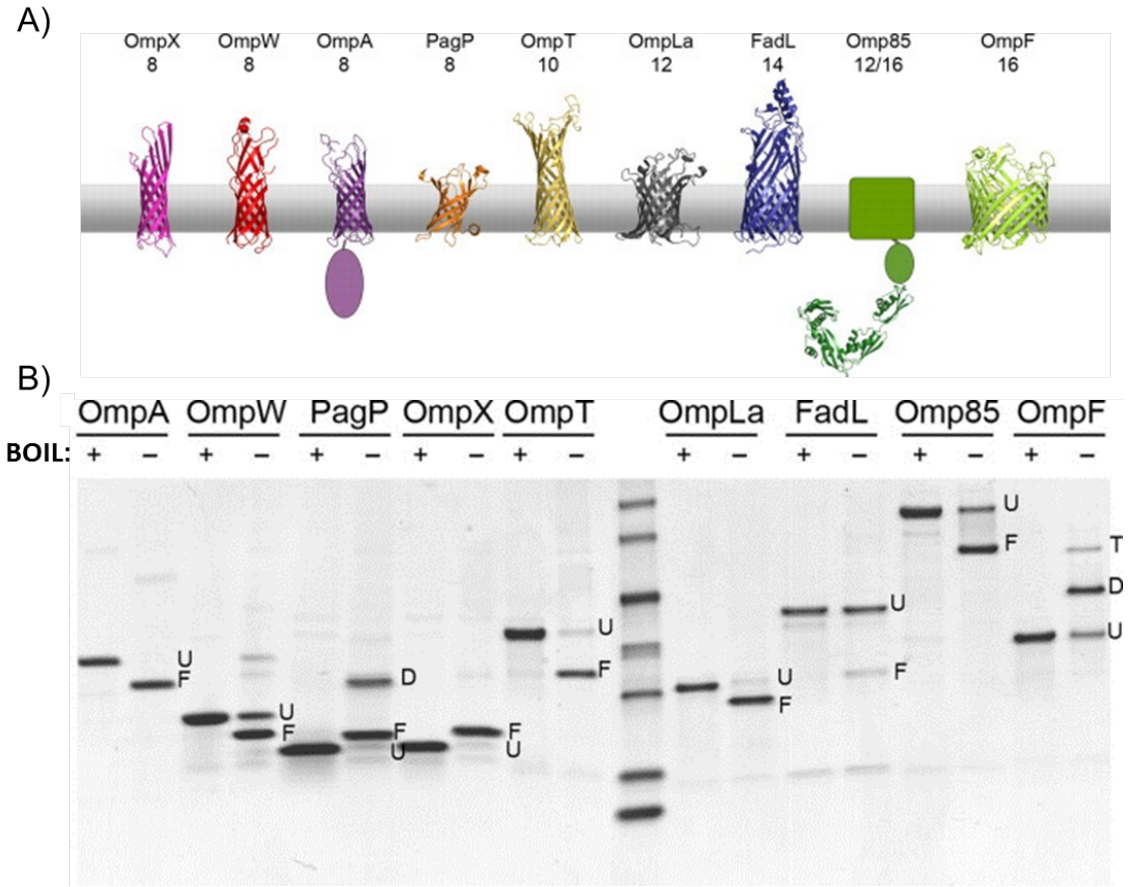


Figure 3.5 (A) Structure of nine *E. coli* OM porins listed by increasing β-strand number. **(B)** Heat modifiability of OM porins. The structure of most porins is resistant to SDS, allowing the folded (F or T:trimer) and unfolded (U) populations, along with dimer intermediates (D) in solution to be captured by SDS-PAGE. For most porins, the folded form migrates faster than the unfolded form, although in some cases (PagP, OmpX, OmpF) the folded form migrates slower. Reproduced from [134].

This migration phenomenon is likely explained by a balance between the compactness and differential SDS binding of the unfolded versus folded form. Like αHL, OM porins are resistant to SDS up to certain temperatures (~65°C), after which they will unfold—a feature known as 'heat modifiability' [148]. Therefore, the folded and unfolded populations present in solution can be captured by SDS-PAGE; the folded form migrates faster or slower depending on the porin (Fig. 3.5 B). In the case of OmpF, three species can be resolved: the unfolded monomer, the folded dimer, and the folded trimer. Following the discussion above, it is not obvious if the addition of a charged tag would cause the trimer to shift upwards (because of the added weight) or downwards (because of the added negative charge), as seen with αHL. Regardless of the direction, adequate gel separation of OmpF should reveal four trimer bands corresponding to from zero to three tagged subunits.

3.1.6 Chapter overview

This chapter presents a protocol for efficient in vitro expression of OmpF trimers using a commercial cell-free kit and subsequent efforts to make OmpF heterotrimers, in which one or two of the barrels in the trimer are different in terms of genetic or chemical modifications (Fig. 3.6). A variety of folding conditions were tested, including additives such as detergents and liposomes. The optimised conditions were then applied to trimerise a mix of WT and genetically tagged monomers in order to create heterotrimers with altered gel migration for downstream gel purification. High-throughput expression of OmpF mutants in vitro would aid fundamental biological studies of OM transport and allow rapid expression of nanopore mutants for biosensor development. The application of OmpF as a biosensor for proteins is discussed in Chapter 4.

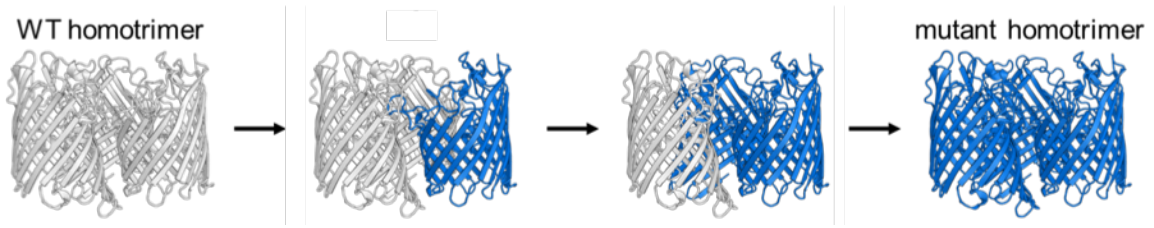


Figure 3.6. Schematic of OmpF homotrimers and heterotrimers, comprising both WT barrels (white) and modified barrels (blue). The trimers are separated by SDS-PAGE based on the altered electrophoretic mobility imparted by a genetic tag added to either the white or blue monomers.

3.1.7 Acknowledgement of collaboration

The heterotrimer project was carried out in collaboration with Amina Ali, a Part II Chemistry student under my supervision. I was responsible for experimental design throughout. Amina executed troubleshooting experiments to determine the origin of unexpected expression products (Section 3.3) and the initial trimerisation of gel-extracted monomers (Section 3.4). She also tried various chemical and genetic modifications of OmpF to enhance gel separation of heterotrimers, many of which are not discussed in this thesis but are described at length in her Master's thesis [149].

3.2 In vitro expression of WT OmpF

In *E. coli*, OmpF is expressed with an N-terminal signal sequence, which prompts translocation across the inner membrane into the periplasm before being cleaved. To express OmpF in vitro, the mature gene lacking the 22-amino-acid signal sequence was cloned into the dihydrofolate reductase (DHFR) control vector (New England Biolabs), which contains a T7 promoter, by homologous recombination to produce the IVTT plasmid pSAI (Materials and Methods, Section 6.3).

In vitro expression of WT OmpF was carried out using the PURExpress® In Vitro Protein Synthesis Kit (NEB) [114,121]. The PURE system is reconstituted

from thirty-six *E. coli* transcription and translation enzymes mixed in known quantities with ribosomes purified by size-exclusion chromatography [114,121]. This highly purified system benefits from a lack of proteases and nucleases. To distinguish the expression products from the IVTT mix protein background, ^{35}S-methionine was added to the reaction (OmpF contains 9 methionine residues per trimer). Expression was terminated with the addition of SDS-containing loading dye and then resolved by SDS-PAGE. Following electrophoresis, the gel was dried under vacuum and exposed to X-ray film overnight (~16 hours). The resultant autoradiograph showed a single band around 37 kDa, indicative of OmpF monomer expression (mass=37.2 kDa) (Fig. 3.7, Lane 1).

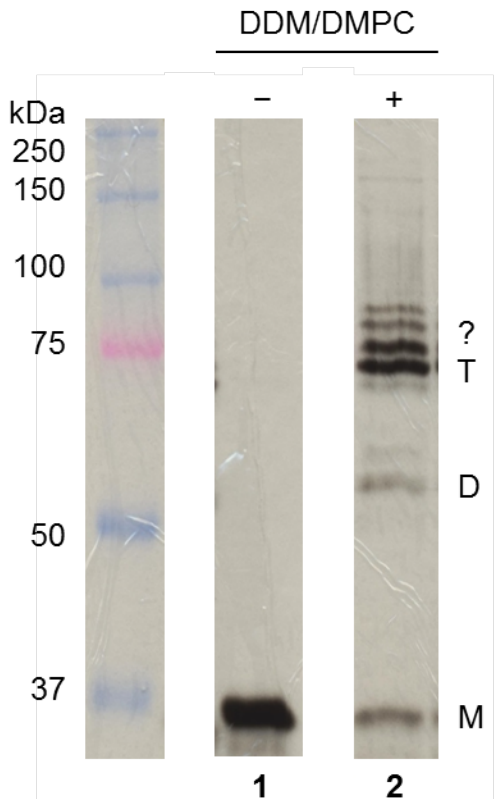


Figure 3.7 Autoradiograph of OmpF expression for 5 hours at 28ºC using the PURE system without (Lane 1) or with (Lane 2) the addition of DDM detergent and DMPC liposomes resolved by 10% SDS-PAGE. Without additives, only monomers (M) are produced. After the addition of DDM and DMPC, Dimers (D) and trimers (T) along with additional unexpected bands (?) are also observed. The initial assignment of each band was based on the gel migration of *E. coli*-expressed OmpF and later confirmed by electrophysiological analysis.

Typically, IVTT of membrane proteins is enabled by the addition of lipid vesicles (liposomes), which stabilise hydrophobic regions to prevent misfolding and aggregation [121]. In vitro trimerisation of urea-denatured monomers was previously reported in a mix of 0.2% (w/v) DDM and 0.5% (w/v) *n*-dodecyl-β-D-glucopyranoside (DG) [132]. The detergent DG is not soluble in water; addition of DG from a methanol stock increased the viscosity of the expression reaction and made it difficult to load on the gel. Therefore, this method was abandoned.

Next, expression was attempted in a mix of 0.9% (w/v) DDM and 1.3% (w/v) DMPC liposomes (7.4 mM of each), which was also previously used to produce OmpF trimers from urea-denatured monomers [135]. Trimerisation yield using this protocol was ~80% and did not vary with the size of liposomes used; therefore, 200 nm liposomes were used throughout unless otherwise stated. The IVTT reaction was repeated at 28ºC for 5 hours with the addition of DDM and DMPC, which has a phase transition temperature of 24ºC. The resultant autoradiograph of the gel showed multiple protein bands: one each for the expected monomer and dimer, along with multiple discrete high molecular weight bands above where the trimer is expected to migrate (Fig. 3.7, Lane 2). Although the trimer mass is 111 kDa, OmpF expressed in vivo migrates around 75 kDa on SDS-PAGE because it retains its native compact state in the presence of SDS [150]. The assignment of the dimer and trimer bands on the gel was later confirmed by single-channel PLB recording (Section 3.5).

The individual contribution of DDM and DMPC to trimerisation was then investigated to determine if either component was causing the multiple trimer bands. In the absence of DDM, only a single monomer band was observed (Fig. 3.8, Lane 2). In the absence of DMPC, the yield of the trimer, which still appeared

as multiple bands, was lower (Fig. 3.8, Lane 3). It was hypothesised that OmpF trimerisation is optimal in a lipid-detergent mix because the detergent destabilises the liposome membrane to create flexible regions that facilitate protein insertion [135]. Further investigation is necessary to confirm the structure of the liposome-detergent mix, such as characterisation by transmission electron microscopy [151].

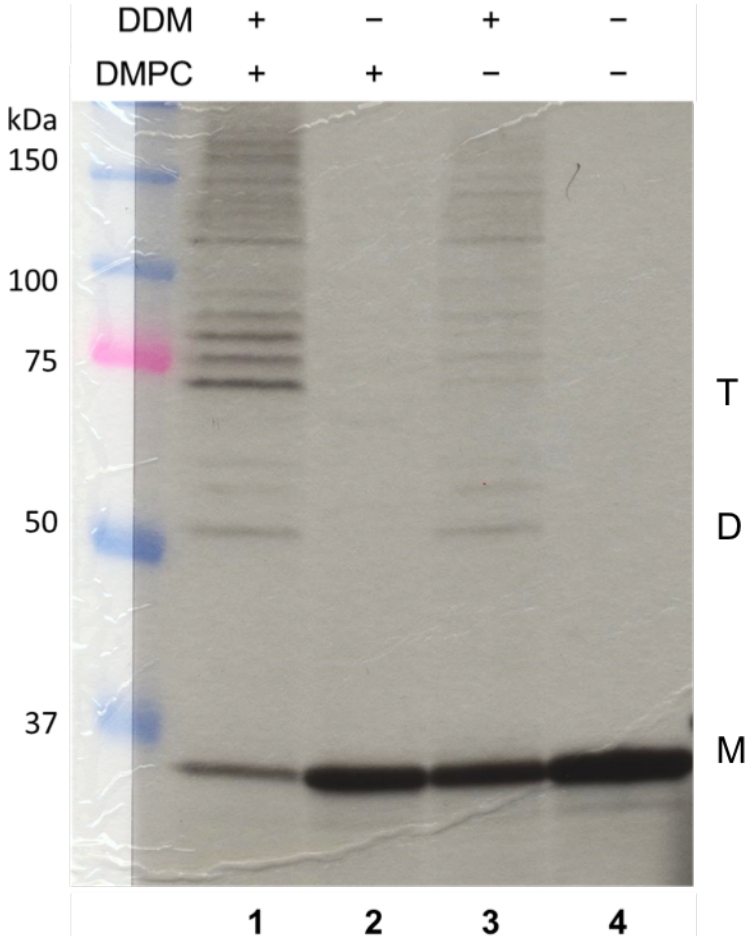


Figure 3.8 The effect of DDM and DMPC on OmpF trimerisation. In the presence of 0.9% (w/v) DDM and 1.3% (w/v) DMPC, OmpF monomers, dimers, trimers, and additional higher molecular weight bands are produced (Lane 1). Without DDM, only monomer is formed (Lane 2, Lane 4). With only DDM, the same pattern is seen as in the DDM-DMPC mix, although the yield is lower (Lane 3).

3.3 Origin of unexpected expression products

In order to resolve OmpF heterotrimers on SDS-PAGE, which was the ultimate goal of this project, adequate gel separation of four trimer bands (representing zero to four modified subunits) was required. This would not be possible with the additional bands appearing around the trimer, so the origin of these high molecular weight species was investigated.

The OmpF gene used in this study contains an additional N-terminal methionine—normally cleaved off with the signal sequence in vivo—which is necessary for initiation of translation. It is possible that this extra residue disturbs the salt bridge between the N- and C-terminus of each monomer. Previously, OmpF with an N-terminal hexahistidine-TEV tag expressed in vivo was shown to successfully trimerise by SDS-PAGE [63]. Additionally, tag removal was observed as a downward gel shift after incubation with TEV protease, which left an additional glycine residue at the N-terminus (Fig. 3.9 A). Interestingly, the first seven amino acids in the X-ray crystal structure of this construct (PDB: 4GCQ) are unresolved compared to the WT, which implies that the N-terminus does not fold correctly to form the salt bridge (Fig. 3.9 B). This creates an empty pocket between the barrels that may promote aggregation leading to the multiple trimer bands on the gel. Interaction with IVTT kit impurities is also possible, as the PURE system still contains unknown amounts of contaminating LPS (noted on the Gene Frontier website, which sells an updated version of the PURE kit).

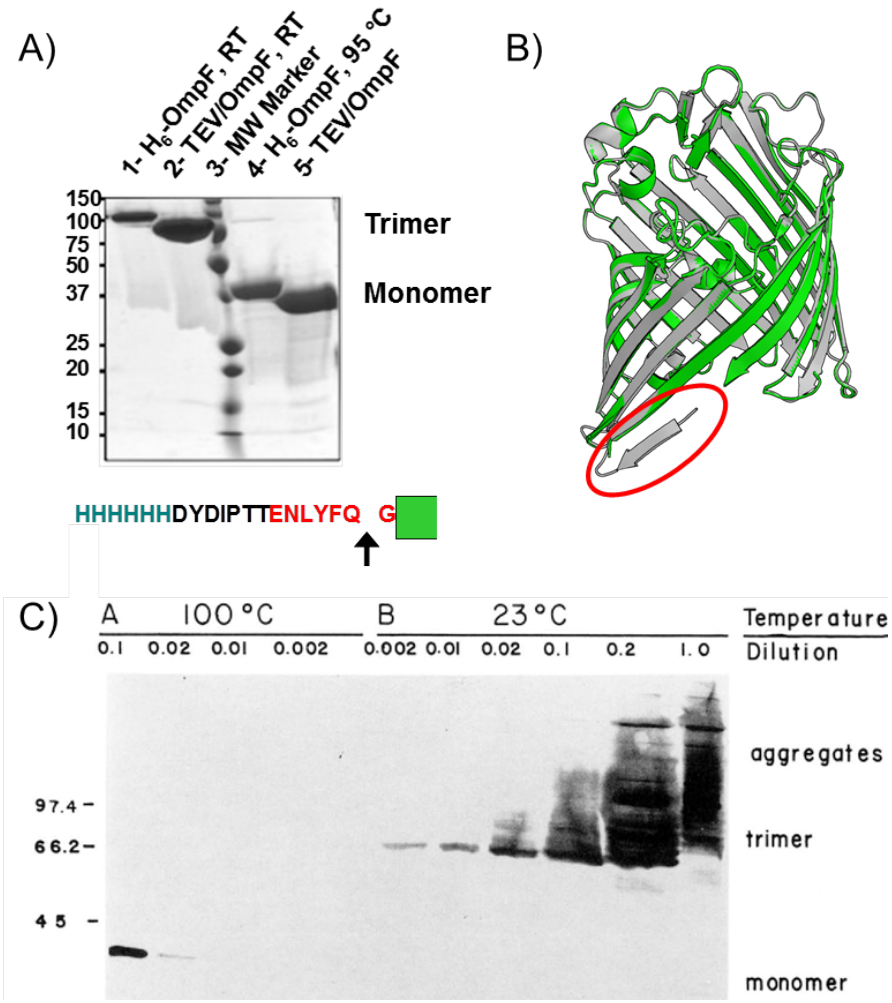


Figure 3.9 (A) SDS-PAGE of OmpF trimers (left of ladder) and monomers (right of ladder) with an N-terminal histidine-TEV tag before (Lanes 1,4) and after (Lanes 2,5) cleavage with TEV protease. Reproduced from [63]. **(B)** Superimposed structures of WT OmpF in grey (PDB: 2OMF) and the TEV-cleaved construct with the additional N-terminal glycine in green (PDB: 4GCQ). The latter is missing the seven N-terminal amino acids circled in red. Structures were visualised and aligned in PyMOL. (C) Western blot with anti-OmpF antibodies of cell envelopes solubilised in a solution containing 20 mM Tris-HCl pH 7.5, 5 mM EDTA, 500 mM NaCl, 2% SDS, and 2% DTT at 100°C (panel A) or at 23°C (panel B) for 10 min and then centrifuged. Multiple discrete bands appeared above the trimer, which were labelled as aggregates. The aggregates disappeared with increased dilution in the same buffer (dilution increases from right to left, i.e. the lane labelled '1.0' is not diluted). Reproduced from [138].

Similar high molecular weight OmpF species have been noted on Western blots, where they were labelled as aggregates (Fig. 3.9 C) [138]. The putative aggregates became less prominent with increased sample dilution in buffer containing 2% SDS. When the immunoprecipitated samples were heated, it was

found that dimers unfolded between 45 and 58ºC whereas both the trimer and aggregates unfolded above 71ºC.

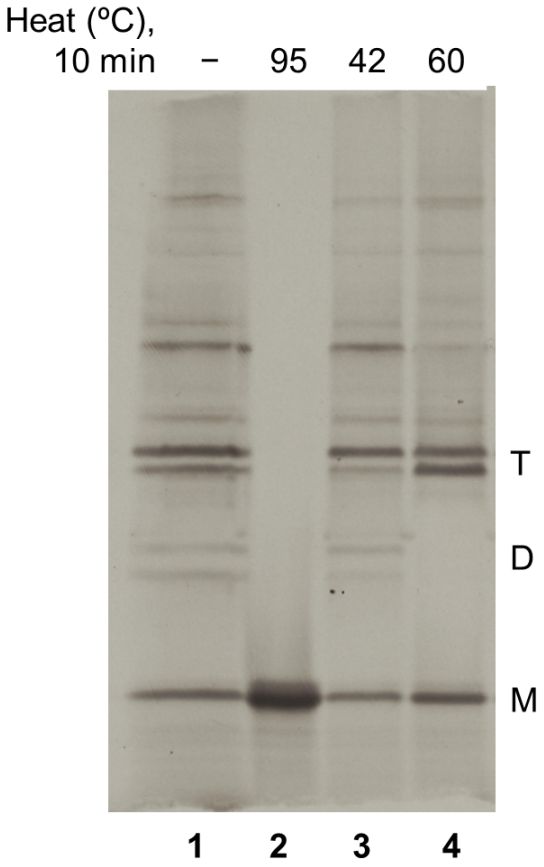


Figure 3.10 The influence of heat on OmpF expression products. After IVTT, samples were either left at room temperature (Lane 1) or heated at varying temperatures for 10 minutes (Lanes 2–4) in loading dye containing 2% SDS before SDS-PAGE.

Dilution of the IVTT reactions is not practical because of the low yield and the volume constraints of the gel wells. To test if the additional bands behaved like the previously identified aggregates, IVTT reactions were heated for 10 minutes at various temperatures (42ºC, 60ºC, and 95ºC) before SDS-PAGE (Fig. 3.10). Heating at 42ºC had little effect on the appearance of the bands, whereas heating at 95ºC dissociated everything to the monomer band. The latter was evidence that the extra bands were not caused by a different contaminating protein. Heating at

60ºC dissociated the dimer band as previously reported, with minimal effect on the trimer and additional bands. Increasing the heating time from 10 minutes to 30 minutes or increasing the heating temperature from 60ºC to 70ºC had no further effect. The additional bands appeared to have equivalent stability to the trimer, as previously reported, and are therefore likely to represent the same aggregated species previously detected on Western Blots.

It was further confirmed that only the OmpF sequence was being expressed during IVTT by carrying out the reaction (1) without the addition of DNA, which resulted in no visible bands on the autoradiograph, and (2) with a linear DNA fragment containing only the OmpF gene (Fig. 3.11). The linear fragment contained an A1M mutation, which substitutes the first alanine for a methioinine rather than adding an additional methionine codon to the mature OmpF sequence, which increases the number of amino acids in the polypeptide by one. This was done to determine if the presence of the extra amino acid was precluding formation of the salt bridge thus facilitating aggregation. The gene was isolated from the pSAI plasmid by PCR using primers with overhangs that include the T7 promoter, ribosome binding site, and T7 terminator. Expression was carried out as before, and no significant differences were observed in the appearance of the bands. These results confirmed once again that the additional bands must be a by-product of OmpF folding, and that the additional methionine was not facilitating formation of these species.

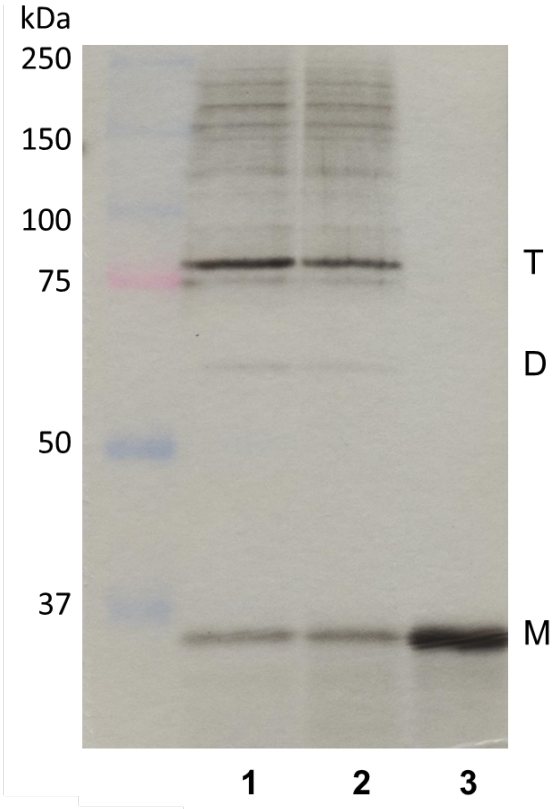


Figure 3.11 In vitro expression of OmpF A1M. The first amino acid, alanine, was mutated to methionine and expressed from either the pSAI plasmid (Lane 1) or a linear DNA template (Lane 2). Heating the DNA template IVTT reaction at 95ºC for 5 min produced a single monomer band (Lane 3), in agreement with previous results obtained from plasmid expression (Fig. 3.10, Lane 2).

3.3.1 Alternative additives

The IVTT reaction was then carried out with different additives to see if the presence of the aggregates could be altered. Previously, it was reported that OmpF trimerisation from urea-denatured monomers is 100% efficient in 100 nm liposomes composed of 10:0 PC, which is four carbons shorter than DMPC [134]. To test this, the IVTT expression was carried out in the presence of 1.3% (w/v) 10:0 PC liposomes (100 nm diameter) with or without 0.9% (w/v) DDM (Fig. 3.12, Lanes 1–2). No trimers were produced using these conditions.

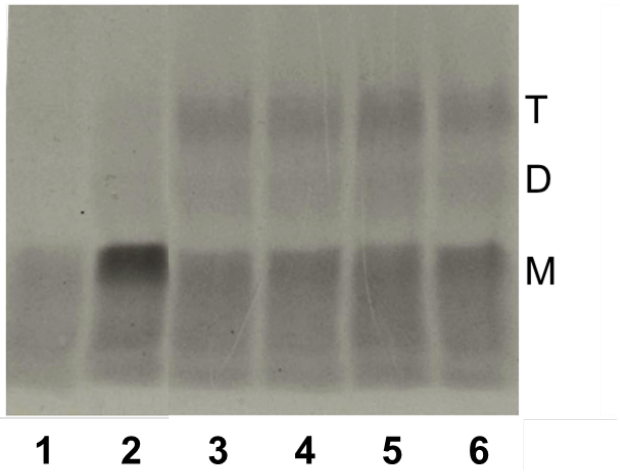


Figure 3.12 The influence of shorter acyl chain lipids and chaperones on OmpF expression and trimerisation. The IVTT additives for each lane were as follows: (1) 10:0 PC liposomes, (2) 10:0 PC liposomes and DDM, (3) DMPC liposomes and DDM along with both Skp and SurA chaperones, (4) DMPC liposomes and DDM along with SurA, (5) DMPC liposomes and DDM along with Skp, (8) standard DMPC liposomes and DDM for comparison. Samples were run on a 7.5% midi Criterion gel (Bio-Rad) rather than a preparative gel, hence the additional bands above the trimer were not resolved and appear as a smear on the gel.

The addition of chaperones to the IVTT mix has also been reported to increase protein solubility. The effects of the major *E. coli* chaperones (GroEL/ES, DnaK/DnaJ/GrpE, and trigger factor) on the expression of 788 aggregation prone proteins using the PURE system was previously investigated. From the 788 proteins tested, 526 showed a drastic improvement (>50%) in solubility by the addition of one or more chaperones [152]. Although transmembrane proteins were not tested in this study, the effect of chaperones that facilitate OmpF folding in vivo, SurA and Skp, on the additional bands produced from the IVTT reaction was tested [153]. The expression was carried out with the addition of Skp, SurA, or a 1:1 ratio of each (2 μM total) (Fig. 3.12, Lanes 3–5) and compared to standard expression in the absence of chaperones (Fig. 3.12, Lane 6). The addition of chaperones produced no discernible effect on monomer or trimer expression.

Different chaperones can be investigated in the future, although it is likely that without the other machinery involved in the OmpF folding pathway, such as the BAM, chaperones are more likely to exercise an undesirable 'holding' effect, in which they bind the protein and prevent folding.

3.4 Trimerisation of IVTT expressed monomer

The mix of DDM and DMPC used in the IVTT reaction was optimised for trimerisation of urea-denatured OmpF monomers [135]. To mimic this more closely, trimerisation of gel-purified, IVTT-expressed monomer was attempted.

The IVTT reaction was carried out as before but without the addition of detergent or lipid such that only monomers were produced. The reaction was then run on a polyacrylamide gel, which was dried under vacuum when complete and then exposed to X-ray film overnight. The developed film was used to ascertain the position of the monomer band on the gel, from which protein was extracted and concentrated using an Amicon 0.5 mL centrifugal filter with a 10K cut-off. The concentrate was diluted in folding buffer containing 100 mM KCl, 20 mM phosphate buffer pH 7.0 with DMPC and DDM in a 1:1 molar ratio (7.4 mM final) and incubated for 3 hours at 30ºC. Following this, the sample was run on a gel, dried, and visualised as before (Fig. 3.13). Two additional bands appeared at the expected dimer and trimer position, indicative of efficient trimerization in the folding buffer (>50% yield). Importantly, no additional aggregates were detected. The gel migration was similar to that observed for *E. coli*-expressed OmpF monomer and trimer (Fig. 3.13, Lanes 2–3).

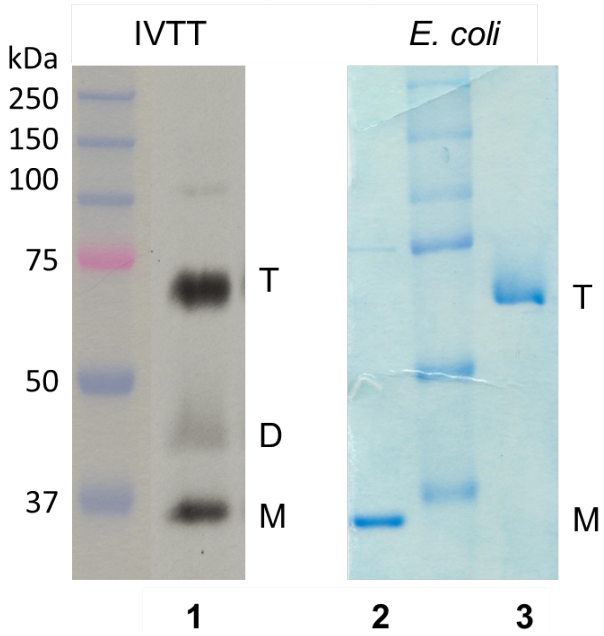


Figure 3.13 Trimerisation of gel-extracted OmpF monomers produced in vitro (Lane 1) compared to *E. coli* expressed monomer (Lane 2) and trimer (Lane 3). Monomers were obtained by heating the trimer sample at 95ºC for 5 min in 2% SDS.

To investigate if the PURE kit contains impurities, such as LPS, which interact with OmpF or promote aggregation during the IVTT reaction, the monomer folding experiment was repeated. However, this time, the monomers were incubated with the IVTT reaction mix for 10 min before dilution into folding buffer. Interestingly, multiple bands were visible on the gel above the trimer band under these conditions (Fig. 3.14). From this limited observation, it was assumed that unidentified PURE system components were causing the additional bands. Although further investigation is required to conclusively explain the phenomenon, this was considered beyond the scope of the project, and the gel-extracted monomer approach was instead used to produce OmpF heterotrimers (Section 3.6).



Figure 3.14 Trimerisation of gel-extracted OmpF monomers produced by IVTT (Lane 1). Trimerisation carried out in the presence of IVTT components (Lane 2).

3.5 Functional analysis of in vitro expressed OmpF

OmpF dimers and trimers produced by IVTT were characterised by single-channel PLB recording. Protein was extracted from the putative dimer and trimer bands, including the additional bands, produced by both methods (IVTT folding or gel-extracted monomer folding). The I-V curves of trimers produced by IVTT and trimers purified from *E. coli* did not significantly differ (Fig. 3.15 A). Additionally, the IVTT expressed trimers easily inserted into a PLB during recordings, as can be seen in the current trace showing 72 insertions over 3 minutes at +20 mV (Fig. 3.15 B). The average conductance of the insertions was 0.88±0.11 nS in 100 mM KCl, 20 mM phosphate buffer pH 7.0 (Fig. 3.15 C), which is in agreement with the previously reported per-trimer conductance of 0.93±0.18 in 150 mM KCl [99]. The large standard deviation likely arises from OmpF inserting in both orientations. Interestingly, protein reconstituted from the additional bands in the trimer region all

showed similar conductive properties to the trimer, implying that the putative aggregates or impurities on the gel disassemble to some extent in solution to allow insertion of single channels.

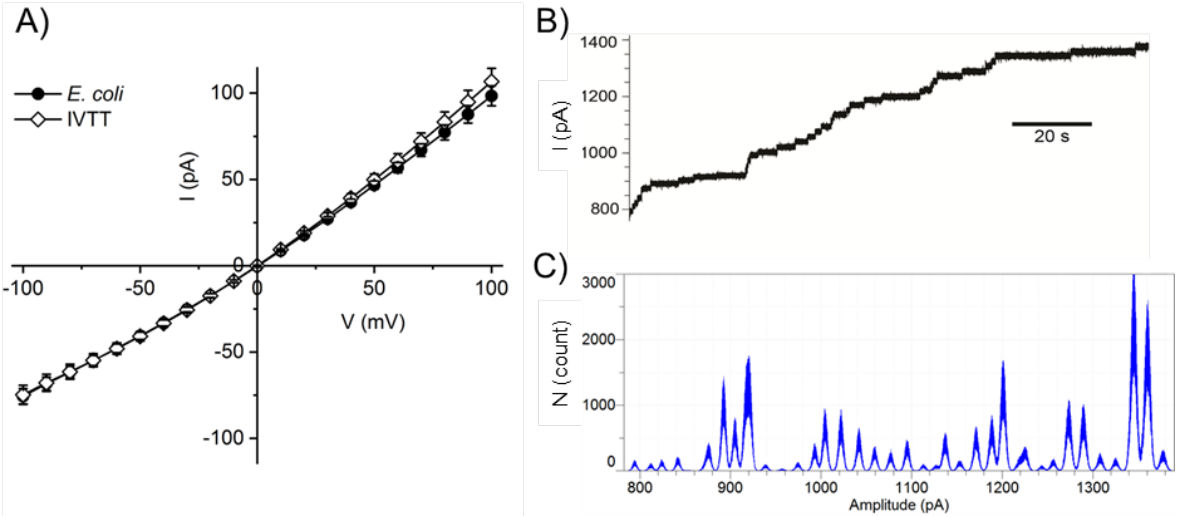


Figure 3.15 (A) Single-channel I-V curves comparing OmpF trimers purified from *E. coli* (closed circle) to those produced by IVTT (open diamond). Data are mean unitary conductance values (±SD) for OmpF trimers in the P-OmpF orientation (IVTT, n=24; *E. coli*, n=14) in 100 mM KCl, 20 mM potassium phosphate pH 7.0. **(B)** Step-wise insertion of IVTT expressed trimers into a PLB at +20 mV. **(C)** Current histograms of insertions in (B). The current distribution of each insertion was fit with a Gaussian function to obtain the mean conductance per insertion.

Aside from conductance measurements, the trimeric nature of the extracted protein was confirmed by voltage-induced gating: upon application of a high potential (>±200 mV) OmpF exhibits a step-wise decrease in conductance to nearly 0 mV. Each step corresponds to one third of the overall conductance and represents closure of a monomer subunit [99]. Single IVTT trimers were reconstituted into a PLB and a +250 mV transmembrane voltage was applied, upon which voltage gating was observed in three equal steps (Fig. 3.16 A).

Voltage gating of IVTT protein from the dimer band produced pores that gated in two steps (Fig. 3.16 B).

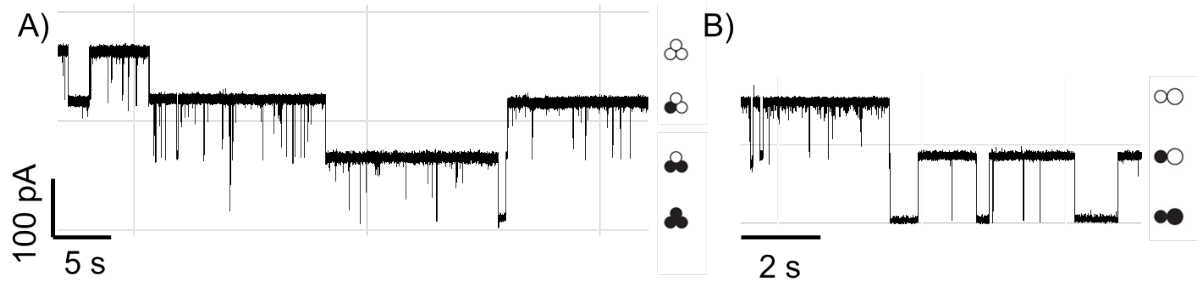


Figure 3.16. Voltage-induced gating of OmpF trimers **(A)** and dimers **(B)** in 100 mM KCl, 20 mM phosphate buffer pH 7.0 at +250 mV. The cartoon representations denote open (white) and closed (black) states.

3.6 OmpF heterotrimers formed by IVTT

Engineering OmpF such that it contains up to three types of barrels per trimer (heterotrimer) rather than one type (homotrimer) would be valuable for biosensing and investigation into OM physiology. It has previously been shown that genetic addition of charged tags to the C-terminus of αHL alters its electrophoretic mobility, leading to a gel shift on SDS-PAGE [139]. This feature was used to resolve IVTT-expressed heteroheptamers, in which only one of the seven monomers forming the barrel is modified (Section 3.1.4). Applications of heteroheptamers have included protein sensing, monitoring chemical reactions, and building molecular walkers [154–156].

Mimicking this approach, OmpF monomers were engineered to contain a tag of eight aspartate (D8) or histidine (H8) residues at the C- or N-terminus. When WT and tagged monomers are mixed together at the DNA or protein level, heterotrimers are expected to form. If the resulting heterotrimers could be

adequately resolved by SDS-PAGE, four discrete bands corresponding to OmpF with zero, one, two, or three modified subunits were expected. The relative positioning of the barrels within the C3 symmetry of OmpF was not considered as it was not expected to affect electrophoresis.

3.6.1 Genetic modification

It has been shown that OmpF with an N-terminal extension of twenty amino acids, comprising six histidines and a tobacco etch virus (TEV) protease motif separated by a linker, still trimerises in vivo (Fig. 3.9 A) [63]. The TEV motif allows removal of the tag after purification. It was therefore reasoned that a similar D8 or H8 extension would be tolerated. Only the D8 tag should be significantly charged during SDS-PAGE (resolving buffer pH=6.8), allowing testing of both negatively charged and neutral modifications.

The IVTT vector pSAI was modified to add the D8-TEV or H8-TEV tag to either the N- or C-terminus of OmpF by two rounds of PCR mutagenesis. The tagged constructs were initially expressed and trimerised in vitro with the addition of DDM and DMPC (Fig. 3.17 A). At this time, the gel-extracted monomer folding protocol presented in Section 3.4 had not been established. Compared to OmpF WT, all tagged constructs exhibited slower gel migration and the additional bands were still present above the trimer. The N-terminal tagged constructs exhibited the biggest gel shift relative to WT. Because of this, future experiments were carried out with N-terminal tags.

In a further set of experiments, D8-TEV-OmpF and H8-TEV-OmpF were expressed as before and either loaded directly onto the gel or incubated with TEV protease before SDS-PAGE (Fig. 3.17 B). Lanes containing the TEV-incubated protein showed the exact same gel shift pattern on the autoradiograph as WT

OmpF (Fig. 3.18 A, Lane 1), indicating successful tag cleavage. Trimerisation yield of the H8 construct was low in this instance, but the bands in the cleaved sample could still be resolved to ascertain tag cleavage. The additional bands observed in both the WT and tagged constructs hinder the resolution of heterotrimers on a gel, as it would not be obvious which bands comprise OmpF heteromers and which comprise homotrimer aggregates.

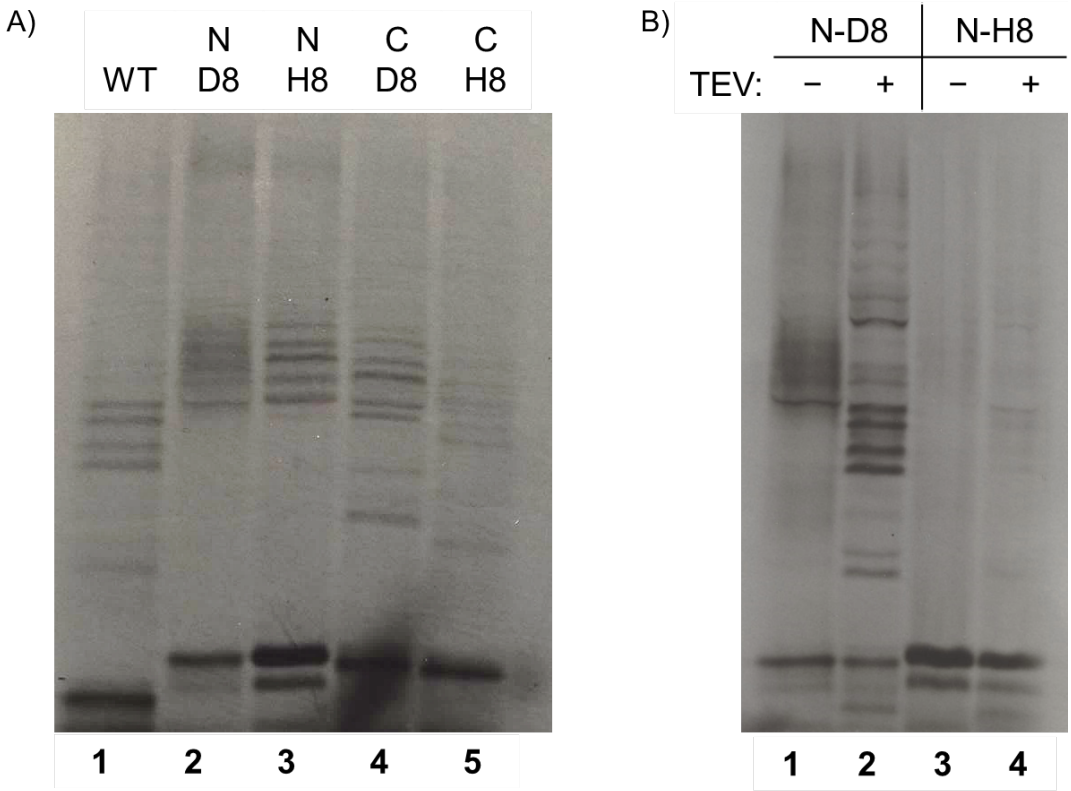


Figure 3.17 (A) Gel shift of OmpF with N- or C-terminal tags relative to the WT. **(B)** In vitro expression of D8-TEV-OmpF and H8-TEV-OmpF without (Lanes 1 and 4) or with (Lanes 2 and 4) subsequent incubation with TEV protease (50 U) at 30ºC for 1 hr.

A different strategy was then considered, in which heterotrimers were formed by folding a mixture of gel-extracted WT and tagged monomers, similar to the protocol established in Section 3.4. First, WT and D8-TEV-OmpF monomers were

expressed and resolved by SDS-PAGE. As before, the monomer band was then extracted from each lane and concentrated. To produce heterotrimers, one half of the WT and the D8-TEV-OmpF concentrate was mixed together before dilution in folding buffer containing DDM and DMPC liposomes in a 1:1 molar ratio (7.4 mM final). The remaining halves were separately diluted in folding buffer to produce WT and tagged homotrimers for comparison. The three samples were incubated at 30ºC overnight (~14 hours) and the resulting OmpF populations were resolved by SDS-PAGE. The tagged homotrimer migrated slower than the WT homotrimer, as seen in previous gels (Fig. 3.18 A, Lanes 1 and 3). In the lane containing both WT and D8-TEV-OmpF monomers, up to three bands could be resolved (Fig. 3.18 A, Lane 2). The lowest band migrates at the same location as the WT homotrimer, whereas the two bands above it migrate between the WT homotrimer and the D8-TEV-OmpF homotrimer, likely representing successful production of heterotrimers with one and two tagged subunits.

To confirm the OmpF species assigned to the different protein bands, each band was extracted from the gel and resolved in individual lanes by a second round of SDS-PAGE (Fig. 3.18 B). A clear ladder can be visualised formed by OmpF trimers comprising only WT subunits (fastest migration), followed by trimers with one D8-TEV tagged subunit, two D8-TEV tagged subunits, and trimers comprising only tagged subunits (slowest migration). The heterotrimer bands were extracted from the gel and shown to comprise functional pores by single-channel PLB recording. The tagged or WT subunit can be modified as desired for downstream applications.

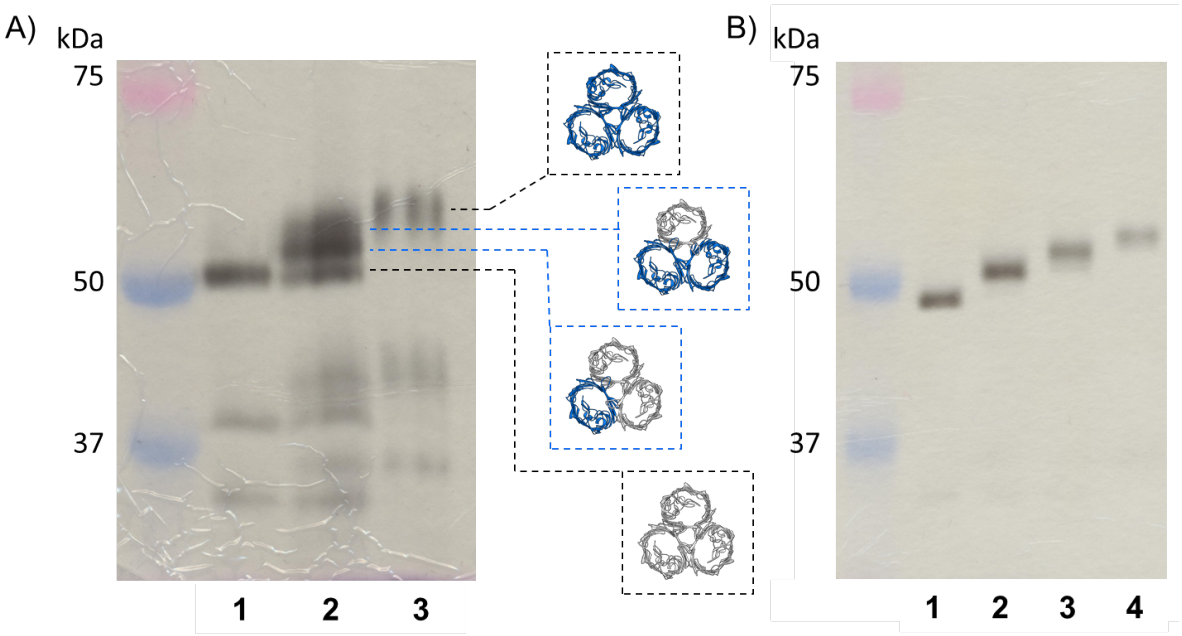


Figure 3.18 In vitro production of OmpF heterotrimers. **(A)** Trimers were formed by folding gel-extracted monomers in a mix of DDM and DMPC liposomes for ~14 hours at 30ºC. Lane 1 contains WT monomers, Lane 2 contains a mix of WT and D8-TEV-OmpF monomers, Lane 3 contains D8-TEV-OmpF monomers. The trimer composition in each band is shown in the PyMOL structures on the right (PDB: 2OMF), in which the grey subunit represents WT monomer and the blue subunit represents tagged monomer. **(B)** The four protein bands labelled with the cartoon structure in (A) were gel extracted and resolved in individual lanes on a polyacrylamide gel. Lane 1 contains WT homotrimer, Lane 2 contains a heterotrimer with one D8 tag, Lane 2 contains a heterotrimer with two D8 tags, Lane 3 contains a homotrimer with three D8 tags.

3.7 Conclusions

This chapter has described a protocol for efficient expression of OmpF homotrimers and heterotrimers in vitro, which has not been previously achieved. A variety of folding conditions were explored, including different types of detergents, liposomes, and chaperones. The optimal condition, determined from analysing expression products by SDS-PAGE, was identified to be a mix of DDM detergent and DMPC liposomes. The trimers were shown to be functional and similar to *E. coli*-expressed OmpF by single-channel PLB recordings.

Initial efforts to resolve heterotrimers containing two types of barrels on a gel were hindered by additional protein bands, later identified as probable aggregates. Appearance of the additional bands could be avoided if OmpF monomers were expressed by IVTT without additives, gel purified, and then folded in the same mixture of DDM and DMPC. Using this protocol, heterotrimers were produced from a mix of WT monomers and monomers containing a D8 N-terminal tag including a TEV cleavage site. The different heterotrimer species were resolved by SDS-PAGE.

Future work will be focused on unambiguously confirming heterotrimer formation whilst showcasing functionality in single-channel recordings. This could be achieved by both chemical modification, for example via introduced cysteine residues, and by genetic modification, for example via introduced binding motifs. Different tags may be explored to increase the gel separation of heterotrimers, as the results in this chapter indicate that migration on SDS-PAGE reflects the molecular weight and not the charge of the tag for OmpF, in contrast to observations with αHL. The application of heterotrimers for stochastic nanopore sensing and the formation of heterotrimers with three different types of barrels is further discussed in Chapter 5.

4

Genetically engineered nanopore libraries for stochastic protein sensing

4.1 Introduction

The detection and quantification of proteins in biological samples is fundamental for bioanalysis and diagnostics. Revised estimates from the Human Genome Project place the number of protein-coding genes at 20,500 [157], which result in an even higher number of proteins (>10^6) due to alternative mRNA splicing and post-translational modifications. Efforts to monitor protein levels, which are often a hallmark of disease, have been plagued by low accuracy. This is in part due to the specificity and dynamic range of existing devices, which must reliably detect the variety of protein sizes, affinities, and concentrations—which range from pg/mL to mg/mL—in the human protein repertoire [158]. Accurate protein detection for diagnosis requires not only specificity and sensitivity, but also high-throughput and multiplexing to allow the measurement of multiple analytes in parallel [159].

Nanopore stochastic sensing is a single-molecule approach to protein detection which can address these limitations (Section 1.1) [14]. Detection has previously been achieved by the addition of protein binding motifs or chemical groups at specified residues within the nanopore, which is reviewed in the

following section. However, in these examples a new sensing element had to be identified and installed onto the nanopore for each target protein. This limits the scope of analysis as not all proteins of interest, including therapeutic targets, have documented binding motifs. Furthermore, the affinity of known motifs may change in the structural context of the nanopore. For these reasons, it would be advantageous to have a library of nanopores from which stochastic sensors with optimised affinities and signal-to-noise ratios can be screened.

This chapter begins by reviewing the field of nanopore protein sensing before introducing peptide libraries and screening methods based on bacterial display. Efforts to genetically engineer peptide libraries into the OmpF nanopore towards the goal of tailored, label-free detection of any protein of interest are then detailed.

4.1.1 Protein detection with nanopores

The enzyme-linked immunosorbent assay (ELISA), which can identify and quantify proteins in the pg/mL range, is the current gold standard for protein detection and quantification [159]. Although a wide range of commercial kits are available, an antibody is required for each new target protein, and the read-out provides a static image of the antibody-antigen interaction. Other commonly used and emerging protein detection methods have been recently reviewed [159]. A drawback of a majority of these methods is that they report ensemble measurements, masking the intrinsic heterogeneity of a protein population, including changes that arise from transcription and translation errors, as well as transient events [160].

Compared to conventional protein detection methods, nanopore stochastic sensing achieves real-time sensing of individual molecules over a wide dynamic range (Section 1.1). Recent efforts in stochastic sensing have been aimed at

peptide or protein detection [37]. Because proteins are too large to bind within the lumen of most nanopores, binding elements are usually chemically or genetically introduced at the periphery of the pore. These elements must be carefully designed to ensure that the binding event is captured in the ionic current read-out. Specific examples are reviewed below.

4.1.2 Chemically engineered sensors

Protein sensing elements have been chemically attached to an engineered cysteine residue by thiol-disulfide interchange or thiol-maleimide click chemistry. In one instance, a DNA oligo was attached to the mouth of the αHL pore and subsequently hybridised to one of four DNA aptamers reported to bind thrombin [59]. Frequent background events were observed, in which the ionic current through the pore decreased by over 50% due to the movement of the oligo-aptamer duplex into the nanopore vestibule. After the addition of thrombin, binding was observed as an additional level corresponding to a ~0.01% increase (rather than decrease) in the ionic current. Despite the high background and low binding signal, it was still possible to extract kinetics for the aptamer-thrombin interaction, showcasing the flexibility of the platform.

In a different study, the αHL pore was modified with a bulky 3.4 kDa polyethylene glycol terminating in a biotin, which could rearrange in the αHL lumen to expose the biotin to either the cis or trans side of the bilayer [161]. The binding of streptavidin, a lower-affinity streptavidin mutant, and an anti-biotin antibody was compared (Fig. 4.1). The αHL pore has also been modified with sugars and an inhibitor peptide for the detection of lectins and kinases, respectively [162,163]

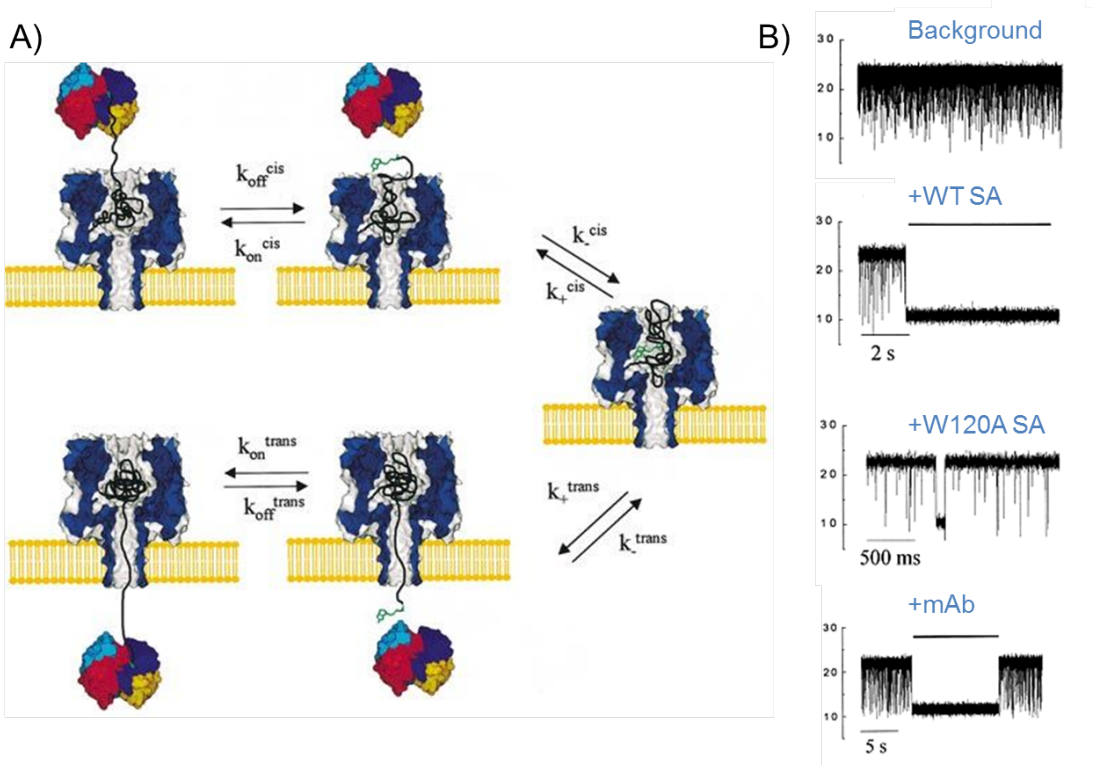


Figure 4.1 (A) Kinetic model of the interactions between an αHL pore chemically modified with a PEG-biotin and streptavidin on the cis and trans side of the bilayer. **(B)** Representative current traces of trans-side binding of biotin to WT streptavidin, a lower-affinity streptavidin mutant (W120A), and an antibody (mAb). Reproduced from [161].

The OmpG nanopore has been similarly modified with a short linker terminating in a biotin for the sensing of anti-biotin antibodies [54]. WT OmpG exhibits background gating, a phenomenon in which the current through the pore rapidly exchanges between the open and closed state [56]. By attaching the sensing element (a biotin group) to the loop which effects gating, analyte binding could be detected as a change in the gating pattern. Indeed, the identification of three different anti-biotin antibodies in a mixture was achieved by first characterising the OmpG gating signature in the presence of each antibody (Fig. 4.2) [53]. Using the same set-up, it was shown that binding of different avidin glyco-isoforms creates different gating signatures, showcasing the high sensitivity of nanopore sensing

[164]. Aside from the nature of the analyte, the current signature will depend on the salt concentration of the buffer, the applied voltage, and the linker length [55].

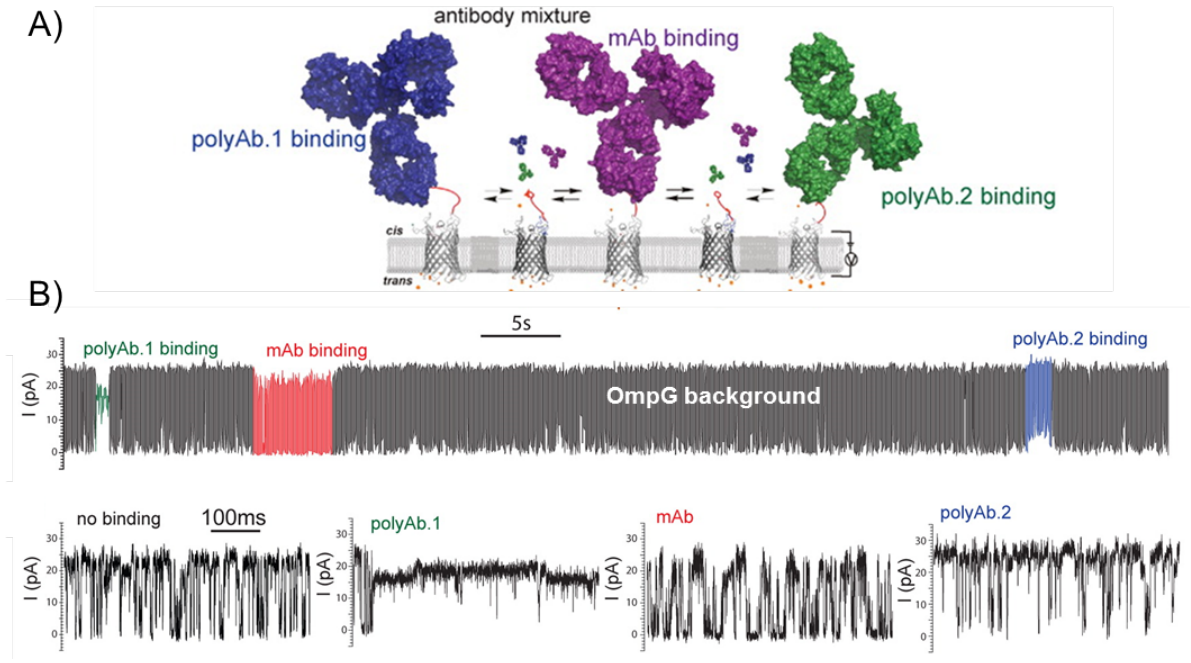


Figure 4.2 (A) Schematic of anti-biotin antibodies binding to a biotin ligand chemically attached to OmpG. **(B)** Representative current trace of a single OmpG-biotin pore showing simultaneous detection of mouse monoclonal antibody (mAb) and two polyclonal antibodies (polyAb) by changes in the gating signature. Reproduced from [53].

It is not clear if smaller protein targets would sufficiently alter gating characteristics to allow analyte discrimination, especially in heterogenous biological samples. The addition of fetal bovine serum (FBS), a complex mixture of proteins and other macromolecules, altered the gating characteristics of OmpG and even blocked the pore in some cases, thus complicating detection [6]. Although gating has been useful for molecular identification, accurate quantification of the analyte and extraction of kinetic rates might be more challenging because of the complex current read-out.

The bacterial toxin ClyA, with a 70 Å diameter (compared to 21 Å for αHL and 25 Å for OmpG), represents the largest protein pore that has been used for single-molecule stochastic protein sensing. It is large enough to accommodate proteins of at least 40 kDa, and has been modified with aptamers for the detection of thrombin [165]. However, off-target molecules that can fit in the pore will also block the ionic current, thus complicating protein detection in biological samples in which matrix proteins may bind and create high background noise. Peripheral binding sites introduced into smaller nanopores may therefore be a better alternative.

Chemical modification of introduced cysteine residues is a simple method of derivatising nanopores with sensing elements. A drawback of this technique is that modification is never 100% efficient: sensing elements were successfully attached to only 65-75% of the cysteine thiols in the studies outlined above. Ultimately, the modification efficiency will vary with the location of the cysteine and the nature of the linker, which is impractical for transferring the technology to a high-throughput platform.

4.1.3 Genetically engineered sensors

Protein detection has also been achieved by genetically engineering binding sites into the pore structure. Compared to chemical modification, genetic modification is straightforward and amenable to high-throughput screening. Kinase sensors have been developed by genetically inserting binding motifs into the peripheral loop of one of the seven αHL subunits [154,166]. In a follow-up study, kinase phosphorylation and dephosphorylation of a serine residue within the motif was monitored in real time [167]. Using a different nanopore derived from the Phi29 DNA packaging motor, detection of colon cancer-specific antibodies was

accomplished by incorporating a binding peptide into the solvent-exposed C-terminus [168].

Apart from short peptides, protein sensing has also been accomplished by the genetic addition of protein segments or whole proteins to nanopores. In one example, all seven peripheral αHL loops were extended with segments of the *E. coli* GroES co-chaperone. The modified nanopore could successfully engage the GroEL chaperone added in solution to form the functional GroEL/ES complex, which was used to monitor the dynamics of the system and chaperone-mediated protein folding [169]. In a different study, an entire protein—Barnase, a 110 amino acid bacterial RNase—was incorporated into the C- or N-terminus of the monomeric nanopore FhuA. Although the modified nanopore reliably formed pores in PLB recordings, the interaction between Barnase and its substrate, an RNase inhibitor, was only detectable once a flexible 12-residue adaptor was genetically inserted into the free N-terminus of the RNase [7]. The adaptor was long enough to reach the pore lumen and decreased the open pore current. Binding of the inhibitor changed the conformation of the adaptor, causing an increase in the nanopore current (Fig. 4.3). Although this is a useful system for monitoring protein-protein interactions, it is an indirect sensing method which would have to be re-engineered for each new protein target.

This issue with general applicability extends to all of the previously described studies, in which suitable binding partners had to be identified and optimised for each nanopore. A nanopore library would allow selection of sensors with optimised kinetics that produce easy-to-interpret binding signals for each new target protein. The screened nanopores could then be incorporated into an array for parallel sensing of multiple targets with molecular precision.

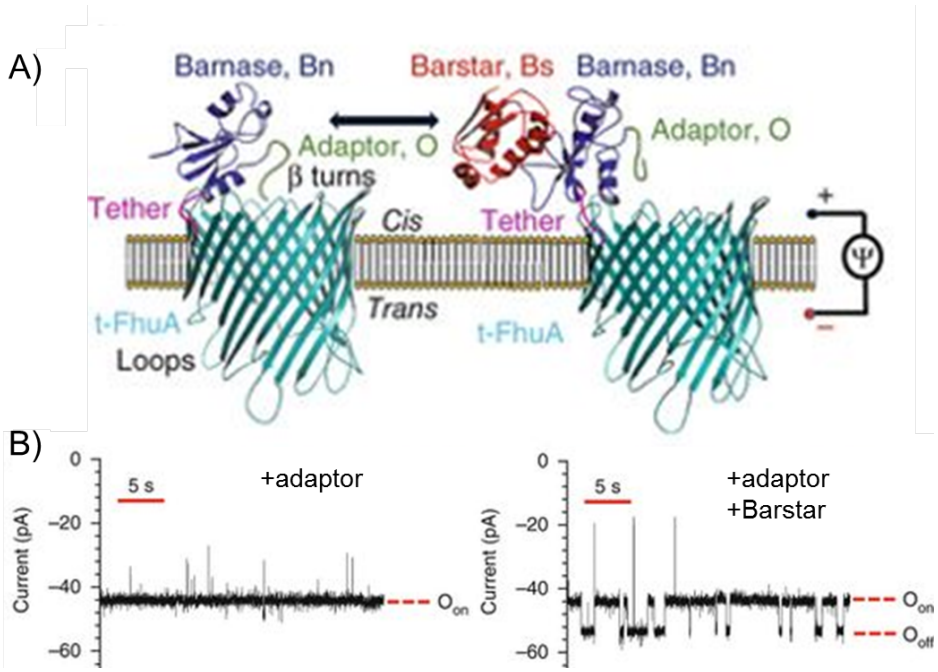


Figure 4.3 (A) Schematic of an RNAse inhibitor (Barstar) binding to an RNase (Barnase) genetically attached to the FhuA nanopore. The Barnase had to be modified with a flexible Adaptor that interacted with the nanopore for Barstar binding to be detected. **(B)** Representative current traces of the modified FhuA nanopore before and after addition of Barstar. Stochastic binding was observed through discrete current transitions to a higher conductance state, presumably due to the removal of the Adaptor from the nanopore entrance. Reproduced from [7].

4.1.4 Synthetic peptide libraries

Biomolecular interactions are largely mediated by peptide motifs: compact interaction modules that are on average six amino acids in length [170]. Of these, two to four core positions confer the majority of the binding affinity specificity [170]. Over 100,000 putative binding motifs have been identified in the human proteome [171]. A majority exist in the flexible disordered regions of proteins, which provide conformational freedom for optimal substrate engagement. Researchers have developed synthetic peptide libraries to discover new, and improve existing, motifs. The 2018 Nobel Prize in Chemistry was awarded for the directed evolution of proteins to generate new, user-defined functions or binding properties [172]. Evolution of peptide libraries through several rounds of screening and enrichment

has been applied to develop new antibodies, cell-penetrating peptides, and binding motifs for biomarkers, among other applications [173].

Libraries must be maximally diverse to ensure that rare variants with optimal characteristics can be identified during the screening process. Researchers quickly discovered issues with the degeneracy of the genetic code, which biases the amino acid distribution within randomised libraries, thereby decreasing diversity. For example, the amino acids serine, arginine, and leucine are each encoded by 6 out of 64 codons, whereas methionine and tryptophan are each encoded by only 1 codon. In addition, the 3 stop codons lead to early termination of library variants, which also reduces library quality. To address this bias, different randomisation schemes have been used. The NNK (K=T or G) and NNS (S= C or G) schemes encode all 20 amino acids using between 1 and 3 codons per amino acid, for a total of 32 codons. Additionally, 2 of the 3 stop codons are eliminated resulting in more functional variants [174]. The two schemes are mathematically identical, although codon usage suggests that NNK may be preferred in *E. coli* and NNS in human cells [175]. A summary of commonly used randomisation techniques is presented in Table 4.1.

The maximum diversity of a peptide library is dictated by the number of randomised residues and the codon scheme used. A tetrapeptide library using NNK/S codons has 6.56×10^4 unique DNA variants, whereas a decapeptide library has 1.12×10^{15} variants. In practice, the maximum size that can be achieved depends on the screening technique used. Techniques that rely on bacteria for library propagation, including phage and bacterial display, have a maximum size of ~10^9 clones because of the upper limit on the transformation efficiency of *E. coli* [173]. Although libraries containing more than seven residues are therefore likely

to be incomplete, they will include more copies of short binding motifs while allowing for longer motifs to emerge [176].

	Amino Acids	Codons	Stop Codons	$P_{stop,1aa}$	$P_{stop,10aa}$
NNN	20	64	3	0.047	0.382
NNK/NNS	20	32	1	0.031	0.270
NNC/NNT	15	16	0	0	0
NDT	12	12	0	0	0

Table 4.1 Amino acid randomisation schemes commonly used in peptide libraries. For each base in the codon, N=A,T,G,C; K=T,G; S=C,G; D=G,A,T. The per-codon probability that a stop codon will occur ($P_{stop,1aa}$) and the probability that a stop codon will occur in a 10 amino acid library ($P_{stop,10aa}$) are shown for each scheme. The proportion of a 10 amino acid library that will contain complete variants without a stop codon can then be calculated as ($1-P_{stop,10aa}$).

4.1.5 Bacterial display

Peptide library display methods physically link a phenotype, such as binding affinity, to its genotype which allows the DNA sequences of screened variants to be retrieved and characterised. Display technologies range from in vitro methods, such as mRNA [173] and ribosome display [177], to methods that depend on bacterial cells, such as phage [178] and bacterial display [179]. Although several bacteria have been explored as hosts for surface display, *E. coli* is most frequently used owing to its rapid growth rate, ease of genetic manipulation, and high transformation efficiency [180].

The most commonly used scaffolds in bacterial display are the *E. coli* porins, which comprise monomeric or trimeric β-barrels localised on the surface of the bacterium [181]. Libraries are often designed as insertional fusions in one of the extracellular loops. The most popular porin for bacterial display has been the monomeric OmpA. Libraries of 15-mer peptides engineered in an extracellular

loop of OmpA were screened to identify cell-specific peptide ligands for drug delivery by bacteria [182] or nanoparticles [183], and to identify peptides that bind with nanomolar affinity to several human protein targets, including C-reactive protein: an inflammation marker used in blood tests [184]. Screening is usually accomplished by incubating the bacterial display library with a fluorescently labelled target, followed by fluorescence activated cell sorting (FACS) to collect cells that are bound to the target and therefore emit a fluorescent signal (Section 4.1.4).

In general, amino acid insertions in the poorly conserved extracellular loops are highly tolerated [185]. Up to 162 amino acids, comprised mostly of histidine, were inserted into the trimeric porin OmpC. The resulting histidine-decorated bacteria could adsorb heavy metals, which might prove useful for bioremediation [185]. During this study, it was noticed that OmpC expression levels decreased with increased insert size. A similar trend was observed in a later study with OmpA, which showed that expression levels varied depending on not only the size, but also on the charge of the insert [186]. The expression of OmpA with negatively charged epitopes showed up to a ten-fold reduction compared to WT OmpA. It was postulated that the negative charges might experience electrostatic repulsion from the negatively charged LPS during OM insertion.

Mutations that lead to improved binding of the target protein may therefore reduce porin expression, along with host cell growth rate and viability. As a result, non-target cells which do not have the beneficial mutations become over-represented. To account for variations in display levels, a circularly permutated OmpX mutant in which both the N- and C-terminus is exposed on the cell surface, rather than in the periplasm, has been engineered [187]. The library is displayed

on the N-terminus, whereas the C-terminus contains an affinity tag that binds a control fluorescent protein [188]. For example, an N-terminal library being screened for ‘blue fluorescent protein’ has a C-terminal affinity tag that binds ‘red fluorescent protein’. The red fluorescence can therefore be used to normalise the blue fluorescence to display levels. Although OmpA and OmpX have been engineered for effective bacterial display, both form 8-stranded barrels that produce unstable pores with low conductance (50-80 pS in 1 M KCl) when reconstituted into PLBs and are therefore not suitable nanopore sensors [189,190].

4.1.6 Fluorescence activated cell sorting

Bacterial display libraries are distinct from other display technologies in that FACS provides real-time quantitative analysis of displayed peptides during screening. Bacteria pass through a laser in a single file (an event) at high speeds of up to 10,000 events per second. The light scattering characteristics of each particle provide information about cell size, which is useful for removing bacterial aggregates. In addition, detectors with different bandpass filters can be used to assess one or more fluorescence parameters in parallel, such as target binding and cell viability (Fig. 4.4). Bacteria expressing variants with desirable characteristics are sorted into one or more tubes, which represent enriched populations. The screening conditions—including target concentration, incubation times, and FACS parameters—can be modified to extract peptides with varying target binding kinetics and affinities [191]. Because of the ability to measure various outputs and tune the stringency of the screen, FACS has enabled efficient identification of peptides for a wide range of applications. This is in comparison to

panning methods such as phage display, in which screening is based only on affinity [192].

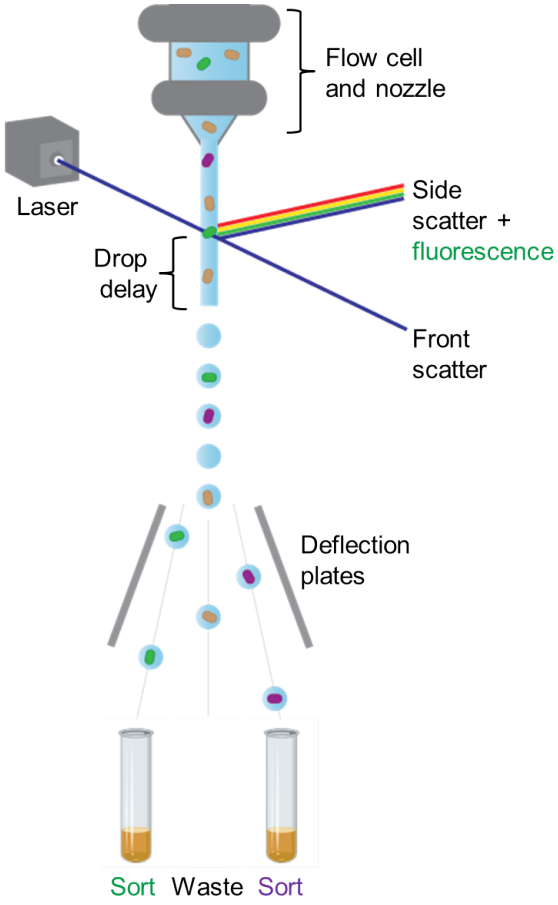


Figure 4.4 Principals of FACS. Cells in suspension are loaded into the flow cell, where they are hydrodynamically focused into a single file and interrogated by a laser. The forward and side scattered light produced by each cell is detected and can be used to discriminate cell size and morphology. In addition, any fluorophores attached to the cells will emit a fluorescent signal, which is filtered into different wavelengths and picked up by a series of detectors. Scattering and fluorescence data are analysed and used to define sort parameters to collect cells with desirable characteristics. After a critical period of time, known as the drop delay, cells are broken up into droplets by an acoustic wave. A charge is applied only to droplets that should contain target cells. Droplets subsequently pass through an electric field generated by deflection plates, which sorts the charged droplets into collection tubes, leaving uncharged droplets in the waste stream.

The limit to the number of bacteria that can be screened is determined by the event rate during sorting. A rate of 5,000 events per second or lower is recommended, as higher rates increase the number of electronic aborts and the

capture of non-target cells [191]. If a library contains 10^9 variants—the limit set by the transformation efficiency of *E. coli*—then, at 5,000 events per second, a full screen will take over one day. To maintain cell viability and consistent cell labelling conditions, the maximum recommended sort time is two hours [191]. Libraries exceeding 10^8 members are therefore enriched using another method prior to FACS, usually by magnetic activated cell sorting. This is done by incubating the bacterial display library with target-coated magnetic beads. The beads are then washed and placed in culture medium, allowing the enriched population that remains bound to divide [193].

4.1.7 Chapter overview

This chapter describes a method for generating and screening OmpF peptide libraries to identify nanopores that can stochastically sense any target protein. Label-free protein sensing with OmpF was first demonstrated by genetically engineering known binding epitopes into the peripheral loops of OmpF. Single-molecule binding of the corresponding targets—an antibody and a kinase—was observed as a change in the electrical read-out. To expand this sensing platform to any protein of interest, two high-diversity OmpF libraries were constructed and screened by bacterial display. Initial library screens against an antibody identified low-nM binding candidates by FACS. Methods to test screened sensors based on DIB and PLB electrophysiology were developed. Unfortunately, no successful antibody sensors were identified, and future optimisation of library design and screening is discussed. Once the platformed is optimised, it is envisioned that screened nanopores will be integrated into arrays for parallel and multiplexed detection of multiple analytes, such as biomarkers (Fig. 4.5).

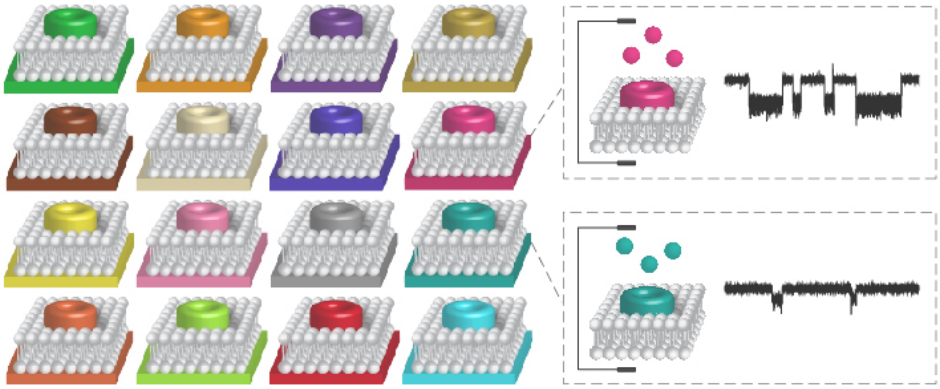


Figure 4.5 A nanopore array for protein sensing. Each differently coloured nanopore can detect a unique target protein, as shown in the electrical trace, for accurate and multiplexed protein detection. The array can be incorporated into a portable device for future applications in diagnostics and precision medicine.

4.2 Library controls

Positive and negative controls were designed to validate the approach before libraries were synthesised. As a positive control, the OmpF gene was mutated at each extracellular loop to contain the Myc epitope: a sequence of 10 amino acids, (EQKLISEEDL) derived from the *c-Myc* oncogene, which is widely used as an affinity tag [194]. This epitope was chosen because it is well characterised, and antibodies that bind the sequence with nanomolar affinity are commercially available [195–197]. It was previously shown that OmpF can tolerate deletion of all extracellular loops [90], so it was reasoned that the insertion of 10 amino acids should not affect OmpF folding.

The insertion sites of the Myc tag were designed based on the loop deletion study [90] and on the structure of OmpF (PDB: 2OMF) visualised in PyMOL (Materials and Methods, Table 6.3). Two extracellular loops, L2 and L3, were not modified because of their structural roles within the protein: L2 latches onto an adjacent monomer thus contributing to the thermodynamic stability of the trimer

[198], while L3 folds into the barrel to create the constriction zone [90]. WT OmpF was used as a negative control to establish non-specific binding in the electrophysiology and fluorescence studies discussed in later sections.

4.3 Stochastic protein sensing with OmpF

The feasibility of label-free stochastic protein sensing with OmpF was assessed using a commercially available anti-Myc tag antibody as the initial protein target. To make the positive controls, the pSAI IVTT plasmid was mutated to individually insert the Myc epitope in L1 and L4-L8. The six resulting plasmids were then expressed in vitro using the PURE system with the addition of DMPC and DDM to facilitate trimer formation (Section 3.2). The expression products were resolved by SDS-PAGE to assess trimerisation efficiency. All of the mutants formed trimers to some extent (Fig. 4.6). The trimerisation yield was lowest for the L4-Myc mutant and highest for the L1- and L8-Myc mutants.

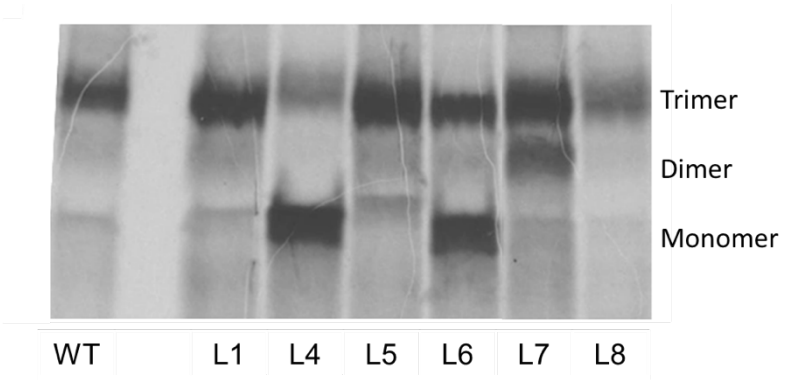


Figure 4.6 In vitro expression of ^{35}S-methionine labelled OmpF WT and Myc epitope loop mutants (L1, L4–L8) analysed by SDS-PAGE (Criterion 7.5%)

The trimer band in each lane (which includes the additional bands discussed in Chapter 3) was extracted from the gel for subsequent analysis by single-channel PLB current recording. First, a single trimer was inserted into the bilayer and the

orientation was determined from the I-V curve (Chapter 2). The open pore current was then recorded for five minutes before the addition of anti-Myc tag antibody to the extracellular side of OmpF, where the Myc epitope should be exposed. Discrete binding steps were observed following antibody addition for only L5, L6, and L7 mutants (Fig. 4.7 A); there were no changes in the current trace for L1, L4, and L8 mutants. Interestingly, the current signature of the bound state differed among the three positive control mutants: Three new current levels could be distinguished for L5-Myc, whereas only one new level with increased noise was seen for L6- and L7-Myc (Fig. 4.7 B). This difference might be useful in sensing structurally similar proteins in solution, provided that the bound state produces a unique current signature at each loop.

The multiple levels observed with L5-Myc may be due to the conformational flexibility of the antibody, which is only detectable at L5 in this case, or by bivalent antibody binding to two of the three Myc epitopes in the OmpF trimer. The latter could be tested by making heterotrimers with only one L5 Myc epitope (Chapter 3). Interaction of the antibody with neighbouring OmpF loops is yet another possibility, and has been previously reported with the OmpG nanopore [164]. A library of nanopores would allow identification of peptide inserts that generate discrete, one-step binding signatures for each protein target, removing the need to decipher complicated interaction patterns. Based on these sensing results with a large protein target (~150 kDa antibody), it was assumed that peptide inserts in L5, L6, and L7 were well-oriented for protein detection and were further considered for library synthesis.

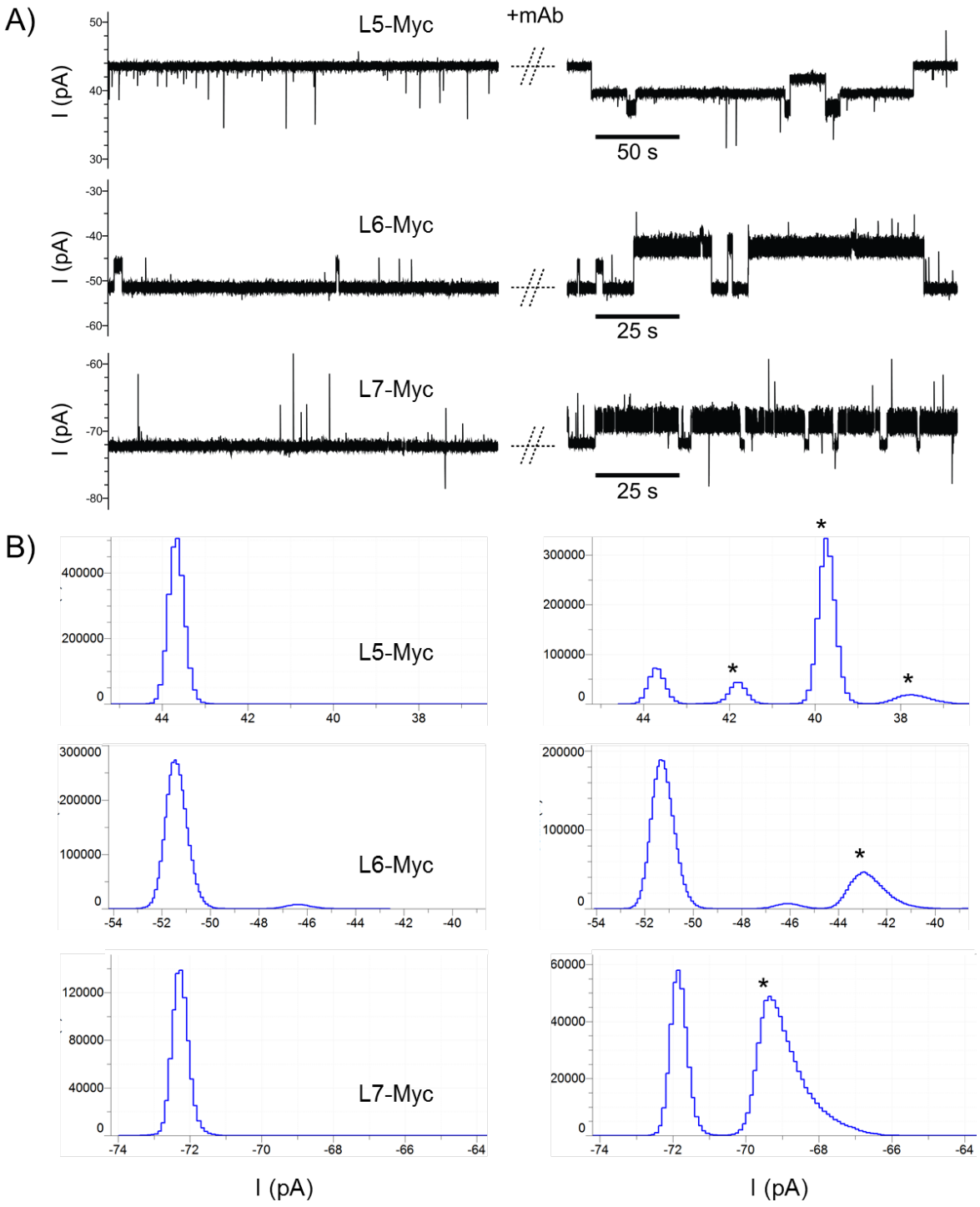


Figure 4.7 (A) Representative traces of OmpF Myc epitope loop mutants before (left) and after (right) the addition of 200 nM monoclonal anti-Myc tag antibody (mAb). The transmembrane potential was ±50 mV. **(B)** Histograms of the current trace before (left) and after (right) the addition of antibody. Additional levels arising from antibody binding are marked with an asterisk. The recording buffer was 100 mM KCl, 20 mM phosphate buffer pH 7.0 except for L7-Myc, which was recorded in 300 mM KCl.

Stochastic protein sensing was then demonstrated using a mammalian kinase, which represents a smaller and more therapeutically relevant target. The dual-specificity tyrosine phosphorylation-regulated kinase 1a (DYRK1a) was chosen, which has potential roles in Down's syndrome, cell cycle control, and tau aggregation in Alzheimer's disease [199,200]. A published binding motif of 10 amino acids (KKISGRLSPIMTEQ) was inserted at the same locations as the Myc eptiope in L5–L7 [201]. The mutants were expressed by IVTT and the trimer was extracted from the gel as before (Fig. 4.8).

The trimers were characterised by single-channel PLB recording in the absence and presence of DYRK1a. All three mutants showed discrete binding levels upon the addition of the kinase to the side of the bilayer where the peptide should be exposed (Fig. 4.9 A). The current signature again varied depending on the loop, but each binding event produced only one new current level with increased noise which reduced the open pore current by ~33%, meaning occlusion of one monomer. The large increase in noise may be caused by the flexible N-terminus of DYRK1a (42 kDa), which interacts with the ion conducting pathway of OmpF (111 kDa) once bound to a loop [202].

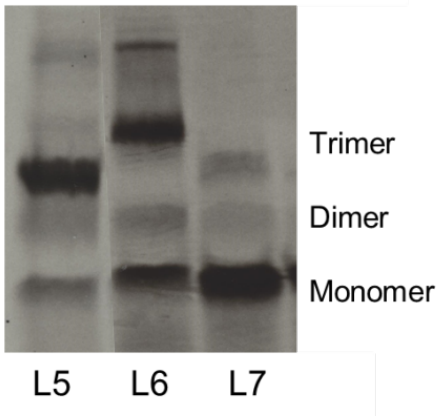


Figure 4.8 In vitro expression of ^{35}S-methionine labelled OmpF WT and DYRK1a epitope loop mutants (L5–L7) analysed by SDS-PAGE.

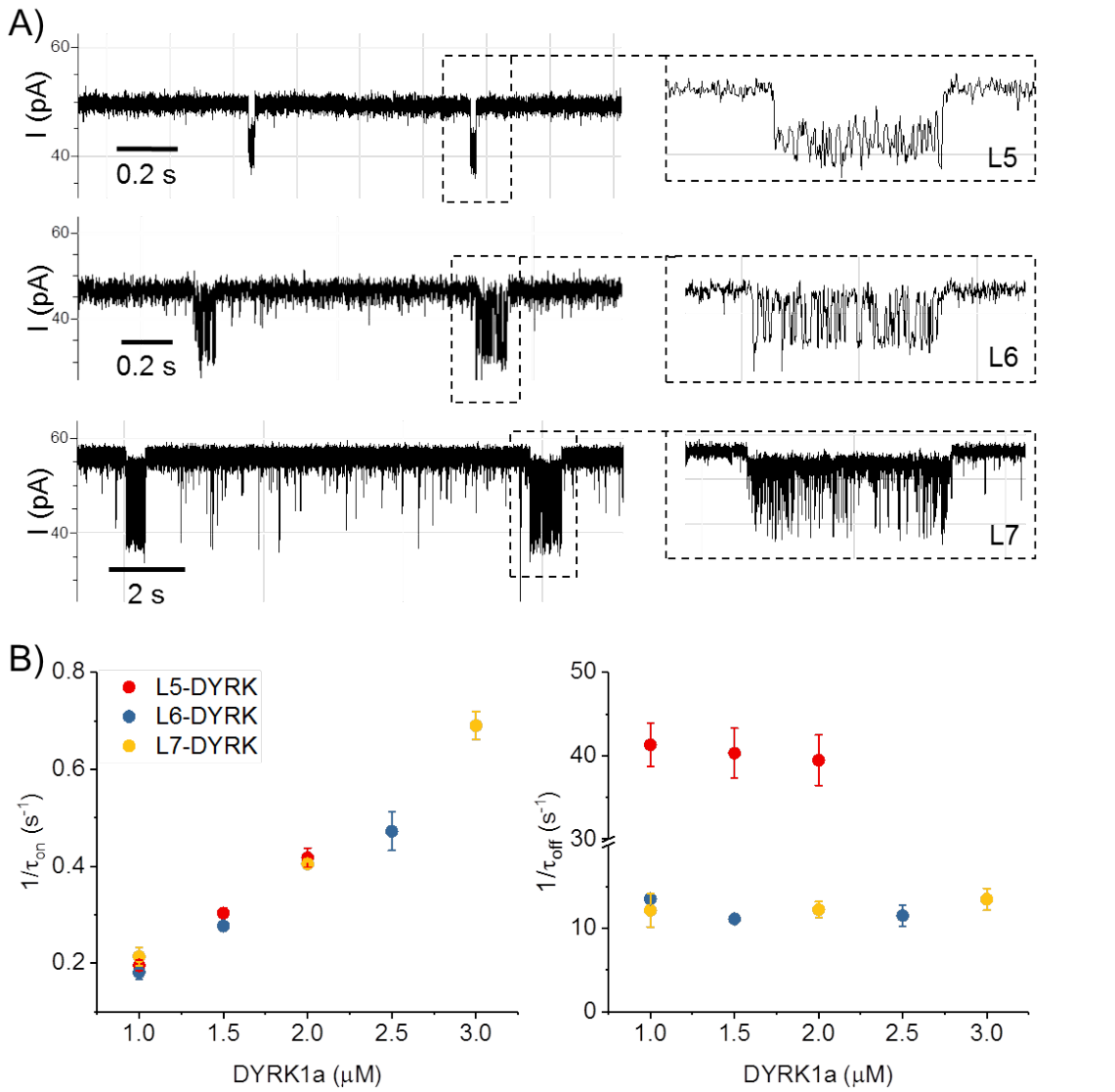


Figure 4.9 (A) Representative traces of DYRK1a binding to loop mutants (L5–L7) containing a DYRK1a binding motif. The insets show an expanded view of the binding current signature. The transmembrane potential was +50 mV. **(B)** Plots of the reciprocals of the mean dwell time between binding events (τ_{on}±SD) and the dwell times of the bound state (τ_{off}±SD) versus the DYRK1a concentration. The standard deviation is reported as the average error of the exponential fit of the dwell-time histograms.

The kinetics of DYRK1a binding were determined by measuring the time interval between binding events (τ_{on}) and the dwell time of the bound state (τ_{off}) at three different kinase concentrations. The mean on and off rate was determined by fitting dwell-time histograms for each kinase concentration to singe exponential

decay functions. Plotting the reciprocal of these values versus the kinase concentration reveals that the on rate is concentration dependent whereas the off rate is not, consistent with a unimolecular dissociation mechanism (Fig. 4.9 B). The k_{on} rate for each loop mutant was then determined from a weighted linear regression of the data points in Origin. The y-intercept was set at zero and the k_{on} value was taken as the slope. Because the dwell time of the bound state was found not to depend on kinase concentration, the k_{off} rate was determined by setting the slope to zero and k_{off} was taken as the y-intercept. The DYRK1a kinetics and the dissociation constants for each loop mutant are summarised in Table 4.2.

The results in this section demonstrate that nanopores genetically engineered to contain binding peptides can detect proteins over a wide range of sizes (150 kDa antibody, 42 kDa kinase) and binding affinities (nM antibody, μM kinase).

DYRK1a	k_{off} (s^{-1})	k_{on} ($\mu M^{-1}s^{-1}$)	K_d (μM)
L5	40.5 ± 0.5	0.203 ± 0.003	200 ± 6
L6	12.5 ± 0.8	0.185 ± 0.001	68 ± 5
L7	13.2 ± 1.1	0.208 ± 0.008	64 ± 7

Table 4.2 Kinetic rates and dissociation constants of DYRK1a binding to a peptide motif (KKISGRLSPIMTEQ) genetically inserted into L5–L7 of OmpF. Data are reported ±SD taken from the linear regression. The error of the dissociation constant, K_d, was propagated from the errors in the k_{on} and k_{off} linear regressions.

4.4 Bacterial display of library controls

The pBAD system was chosen for library expression. The vector contains the P_{BAD} promoter from the *E. coli araBAD* operon, in which expression is tightly regulated by AraC and catabolite activator protein (CAP) [203]. Induction with L-arabinose turns on expression from P_{BAD}, which exhibits low transcription in the absence of the sugar [203]. Pulse labelling experiments have shown that protein is expressed as early as 1 minute after induction, with a difference between induction versus repression expression of up to 1,200-fold [203]. Low background expression levels and rapid induction kinetics are key features for library expression, where viability is important [204]. Figure 4.10 includes a mechanistic overview of expression.

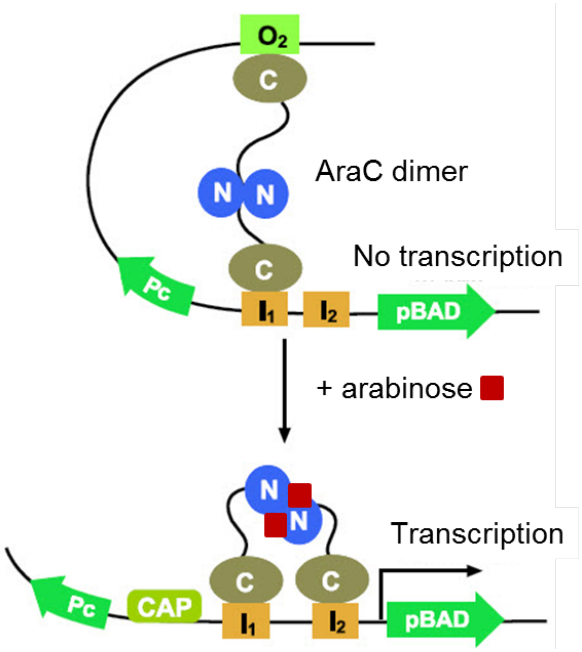
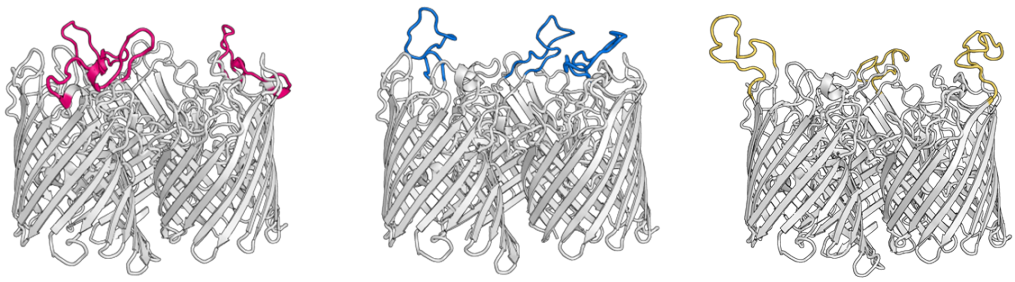


Figure 4.10 The pBAD expression system. In the absence of L-arabinose, the AraC homodimer contacts the O_2 and I_1 sites of the araBAD operon, forming a 210 bp DNA loop. For maximum transcriptional activation, two events are required. First, L-arabinose (red square) must bind to AraC which causes it to release the O2 site and bind the adjacent I_2 site thus releasing the DNA loop. Second, cAMP-binding protein (CAP) must bind between the AraC promoter (P_c) and the I_1 site to increase the AraC affinity for I_2. Transcription will proceed without CAP binding, albeit at a slower rate.

The OmpF gene containing the N-terminal signal sequence was cloned into the pBAD/*Myc*-His B vector (Invitrogen) for arabinose-induced expression in *E. coli.* Positive control plasmids containing the Myc epitope in L5–L7 were made by PCR mutagenesis, this time flanked on either side by a six-residue SG linker (total insertion=22 amino acids) (Fig. 4.11). The SG linker was added to increase the peptide distance from the cell surface for efficient labelling and also to provide conformational flexibility for optimal target binding.



L5-Myc L6-Myc L7-Myc

Figure 4.11 Models of OmpF with the Myc epitope flanked by six SG residues inserted into L5 (pink), L6 (blue), and L7 (yellow). The models were made in SWISS-MODEL, a homology-modelling server, using PDB file 2OMF as the template structure [205].

Plasmids were electroporated into *E. coli* strain MC1061, derived from the K12 lineage [206]. This cell line is routinely used for phage display and commercially available electrocompetent cells report efficiencies of ≥4×10^{10} colony forming units (CFU)/μg DNA, allowing construction of high-diversity bacterial display libraries. Further, the strain maintains a steady inducer concentration because it can intake arabinose but cannot metabolise it [207]. It has also been shown, using the OmpA nanopore, that the growth of individual library members in this strain is not biased by the charge of the inserted peptide [186], which should promote even growth of

library variants. It is important to note that the MC1061 strain is not OmpF deficient. It was reasoned that over-expression would be tolerated because OmpF is already highly abundant in the OM. Expression of additional mutant OmpF may also downregulate expression of native porins [208].

Because viable cells are a requisite for library screening, the toxicity of OmpF overexpression in MC1061 cells was assessed by performing CFU counts on agar plates. Bacteria carrying the WT plasmid were grown in LB medium to an OD=0.4–0.6 and either induced with 0.4% (w/v) arabinose for two hours or left to grow uninduced. The arabinose concentration was chosen based on a cell-labelling study which demonstrated that expression from the P_{BAD} promoter is all-or-none: at non-saturating arabinose concentrations (<0.2%), only a fraction of cells are induced. This has led to the incorrect assumption that expression levels can be tuned on a per-cell basis [209]. A fixed number of cells from both the uninduced and induced culture were plated on LB-agar before induction and at one and two hours after induction (Fig. 4.12). Compared to the uninduced sample, the induced sample showed a sharp CFU drop-off after induction.

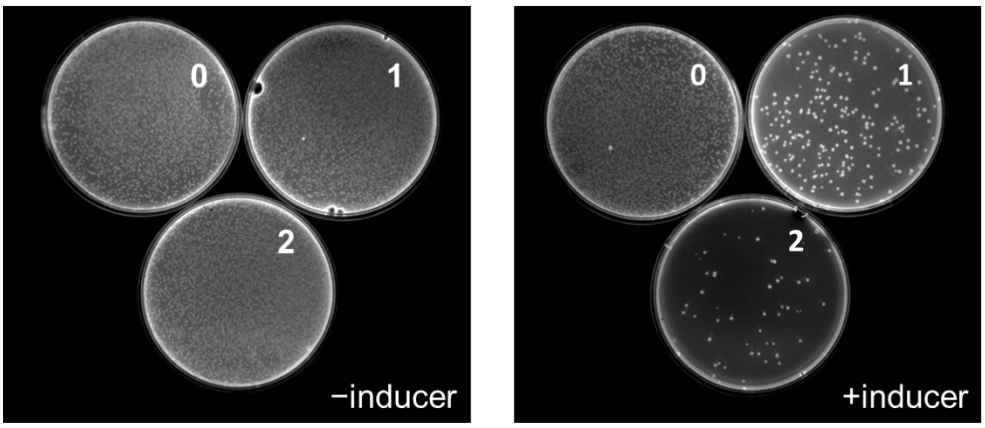


Figure 4.12 Viability of cells containing the WT OmpF plasmid that were either uninduced or induced with 0.4% (w/v) arabinose plated at the time of induction, 1 hour after induction, and 2 hours after induction. An equivalent number of cells were added to each plate.

Possible explanations for the observed cytotoxicity include (1) a high OmpF expression rate overloading the Sec translocon, which handles co-translational folding of most inner membrane proteins and transport of OM proteins into the periplasm [210], or (2) a lethal increase in OM permeability caused by the additional OmpF. To address the first hypothesis, the culture medium was switched from LB to RM medium, which is composed of M9 salts, casamino acids, $MgCl_2$, and 0.2% glucose [203]. Glucose lowers cAMP levels leading to decreased CAP binding, which in turn decreases the expression rate (Fig. 4.10). Growth curves were compared for bacteria grown in LB versus RM medium (Fig. 4.13). Uninduced samples grew more slowly in LB, and a marked decrease in OD was observed in cells induced for more than one hour compared to the uninduced control. In the RM cultures, no significant difference in growth was observed between the induced and uninduced sample until two hours after induction.

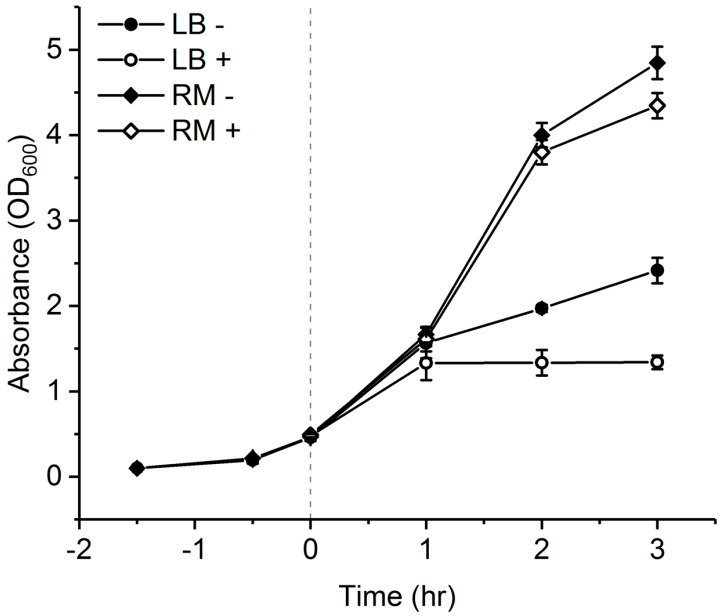


Figure 4.13 Growth curves of MC1061 cells containing the WT OmpF plasmid in LB (circle) and RM (diamond) medium. At timepoint zero, demarcated by the dashed line, the cells were either left to grow uninduced (filled symbols) or induced with arabinose (open symbols).

It is possible that the increased growth in RM medium is due to low or no expression of OmpF, however subsequent cell labelling (Section 4.4.1) and FACS (Section 4.4.2) experiments demonstrated that OmpF is expressed and folded into the OM even after induction for one hour in RM medium.

These results indicate that the presence of glucose slows the rate of protein expression, which reduces cytotoxicity. The second hypothesis—that increased OM permeability is also causing cell death—was investigated by FACS using an OmpF knockout strain (Section 4.4.2). All further experiments were done in RM medium with one hour induction to maintain cell viability.

4.4.1 Surface display of peptides

Cell-labelling experiments were subsequently carried out to ensure that peptides inserted into the extracellular loops of OmpF are efficiently displayed on the bacterial surface. MC1061 cells expressing the negative and positive control plasmids were labelled with anti-Myc tag antibody conjugated to a green fluorescent dye. A fixed number of cells were pelleted from samples induced for two hours and then incubated with 100 nM antibody for 45 min at 4°C. Following this, cells were pelleted and resuspended in phosphate buffered saline. A small volume (3–5 µL) of the labelling reaction was pipetted onto a square of LB-agar, which was then flipped over onto a glass cover slide to create a single layer of cells at the agar-glass interphase which was visualised by confocal microscopy (Fig. 4.14).

Only cells expressing L5-Myc OmpF and L6-Myc OmpF were fluorescent, which was confirmed by overlapping the bright-field and fluorescence channels. The antibody is therefore specifically binding the displayed Myc epitope, as cells expressing WT OmpF did not appear fluorescent (Fig 4.15). It is unclear why L7-

Myc OmpF was not successfully labelled. One explanation is that the orientation of the Myc epitope at the cell surface was sterically hindered or was in a suboptimal orientation leading to decreased affinity for the antibody. A new mutant, which had a shortened SG linker on either side of the Myc epitope, was successfully labelled in later experiments (Fig. 4.14, L7-Myc bottom panels). However, further studies were only carried out with L5 and L6 mutants.

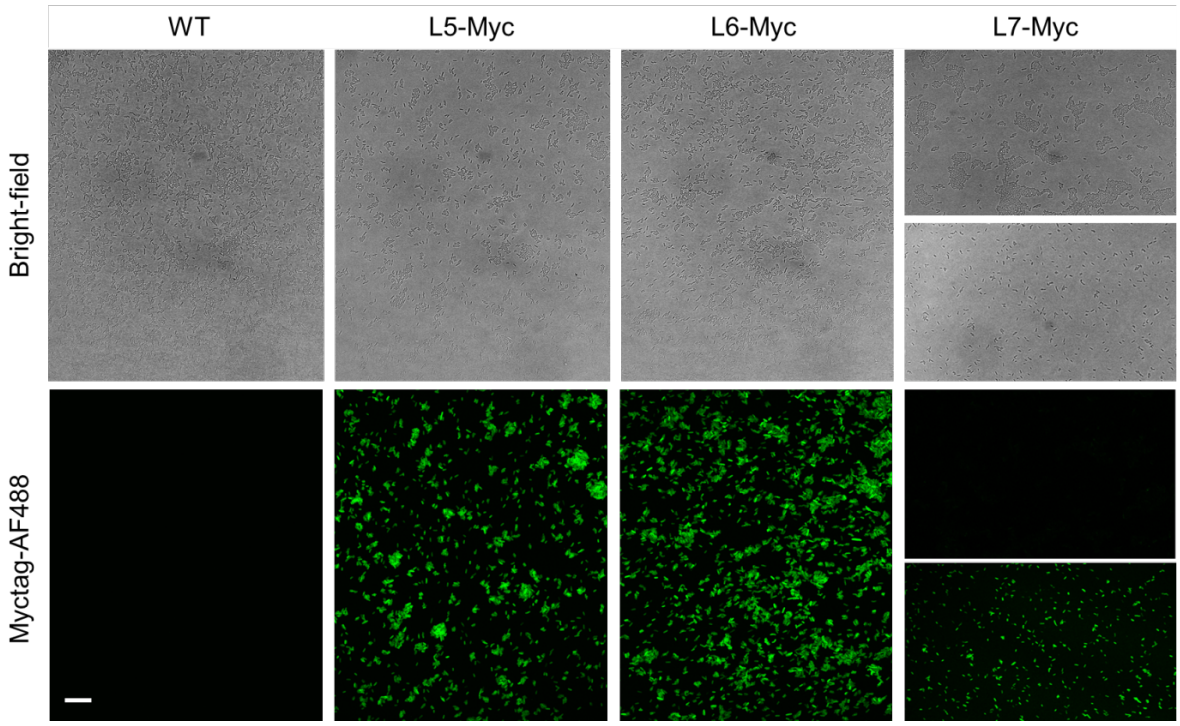


Figure 4.14 Bright-field and fluorescence images (scale bar: 20 μm) of MC1061 expressing library controls labelled with Alexa Fluor 488-conjugated anti-Myc tag antibody (100 nM). The top panels in the bright-field and fluorescent images under L7-Myc represent initial labelling attempts with the Myc epitope flanked on both sides by six SG residues. The bottom panels represent subsequent labelling attempts with the Myc epitope flanked by only two SG residues.

4.4.2 Fluorescence activated cell sorting

The established library controls were then labelled with fluorescent antibody and sorted by FACS to determine if there is sufficient signal separation between the

negative and positive control populations, as well as to test recovery and viability after cell sorting. Expression was induced for one hour before cells were labelled with 50 nM antibody. After labelling, cells were either pelleted and resuspended in RM medium (non-equilibrium sorting) or taken directly to the FACS machine (equilibrium sorting).

For library screening, the non-equilibrium mode will favour cells displaying peptides with low dissociation constants, which will remain bound to the antibody (or other target protein) even after the unbound antibody in solution is removed. It has been estimated that interactions with K_d values >10 nM have relatively fast dissociation constants, which would result in half of the bound ligand dissociating from the cell surface in a few minutes or less at 25°C [204]. Nanopores that bind the target too tightly would not be compatible with the experimental timeframe of stochastic sensing, during which hundreds of binding events must be recorded over a few hours to extract kinetics. Therefore, equilibrium mode sorting, which would collect cells displaying peptides with both high and low dissociation rates, may increase the chances of finding an optimal stochastic sensor. In general, only moderate ($K_d<10^{-6}$) to high affinity interactions are detectable by FACS because of background fluorescence [204].

To determine if both modes can be used for library screening, 100,000 events were measured for each sample. The resulting fluorescence intensity histograms show a ~50-fold increase in fluorescence signal in the positive control population compared to the negative population in the non-equilibrium mode (Fig. 4.15 A) and a ~10-fold increase in the equilibrium sort mode (Fig. 4.15 B). Although signal separation is lower in equilibrium mode, both modes provide sufficient fluorescence resolution to allow a sort gate to be applied that only

encompasses the positive population, avoiding collection of false positives during the screening process. Therefore, both strategies are viable for library screening.

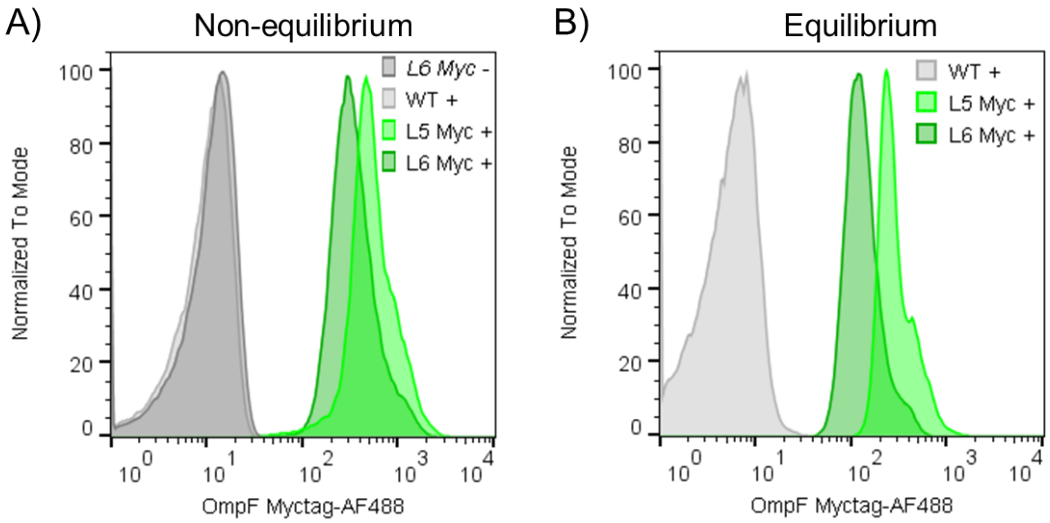


Figure 4.15 Flow cytometry fluorescence histograms of uninduced (WT and L6 Myc-) and induced (L5 Myc+ and L6 Myc+) library controls labelled with 50 nM fluorescent anti-Myc tag antibody. Each histogram comprises 100,000 events measured by the cytometer. Sorting was carried out in either non-equilibrium mode **(A)**, in which the unbound antibody was removed before the sort, or in equilibrium mode **(B)**, in which the cells were sorted in the presence of unbound antibody. The fluorescence was adjusted such that negative populations were distributed around 10 on the X-axis for comparison. Histograms were scaled to the peak height (set as 100%) of the distributions.

Cell viability must remain high after FACS to preserve library diversity. To determine if cells overexpressing OmpF survive the sorting process, 5,000 events from the equilibrium samples (Fig. 4.15 B) were sorted into 5 mL polypropylene tubes filled with 1 mL SOC medium, a rich broth which contains 0.36% (w/v) glucose. Prior to sorting, the tubes were incubated for at least 30 minutes with FBS, which improves recovery by preventing sorted cells from sticking to the tube walls and dying [211]. After the sort was complete, 100 µL was plated on RM-agar

(500 CFU expected) and the number of colonies was counted after overnight incubation (Table 4.3).

No significant difference was noted in CFU numbers between the uninduced and induced controls, indicating that overexpression of OmpF does not significantly compromise cell viability during the sorting process. However, sorting recovery—the number of CFU obtained divided by the number of CFU expected—ranged from 50-61%. Low recovery may indicate cell death during the sorting process, but is more likely to be caused by suboptimal stream alignment, cells still sticking to the tube walls (despite the FBS coating step), or the presence of debris or electronic noise in the system that is counted as an event, i.e. the number of events sorted does not represent the actual number of bacterial cells in the tube.

The threshold for an event count could not be further adjusted without excluding bacterial cells, so experiments were continued using the established parameters. Considering the high variability of plate counts [212], especially when only one plate was quantified in most cases, and the fact that the goal was not to identify the best binders (which would make poor nanopore sensors), the recovery rate was deemed acceptable for future experiments.

CFU:	Uninduced	Induced
WT	263	250
L5-Myc	305	270 ± 90
L6-Myc	270	286

Table 4.3 The CFU count on RM-agar plates from induced and uninduced samples sorted by FACS. For each sample, 5,000 events were sorted into 1 mL of SOC and 100 µL was plated, which is expected to produce 500 CFU. One plate was assessed for each sample, except for L5-Myc for which the sample was plated in triplicate to estimate the standard deviation of the plate counts.

The sorting experiment was repeated with an *ompF* knockout strain in non-equilibrium mode to see if recovery could be increased by reducing the amount of OmpF present in the OM. The resulting fluorescence histogram shows a ~100-fold increase in fluorescence in the positive control compared to the negative control, which is two times larger than the signal separation obtained with MC1061 cells (Fig. 4.16 A). However, recovery after cell sorting as determined by plate counts was only ~1% (Fig. 4.16 B), so the strain was abandoned in favour of MC1061 cells which had been used thus far.

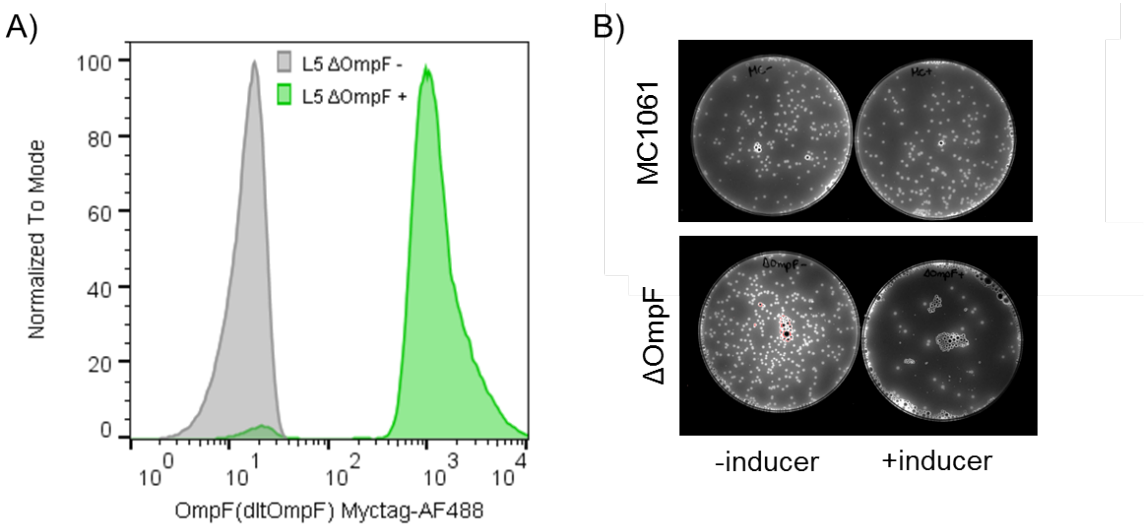


Figure 4.16 (A) Fluorescence histogram of uninduced (grey) and induced (green) OmpF knockout cells transformed with L5-Myc OmpF plasmid labelled with 50 nM fluorescent anti-Myc tag antibody. **(B)** Comparison of post-FACS viability of MC1061 and OmpF knockout cells. For each plate, 500 CFU was expected. For the uninduced samples, 225 and 250 CFU were recovered for the MC1061 and OmpF knockout cells, respectively. For the induced samples, 340 and 30 CFU were recovered for the MC1061 and OmpF knockout cells, respectively.

4.5 OmpF bacterial display libraries

Once the feasibility of the approach had been demonstrated using positive and negative controls, two high diversity peptide libraries were made in L5 and L6 of OmpF.

4.5.1 Library design

The OmpF nanopore library was made by randomising 10 amino acids in L5 and L6 at the same location as the positive Myc control (Materials and Methods, Table 6.5). A 10-residue peptide was chosen because it reflects the size of the tested Myc epitope, and because binding motifs are on average 6–7 amino acids. For the randomisation, NNK (N=A,T,G,C; K = T or G) codons were used in which all 20 amino acids are encoded for by 32 codons and 2/3 stop codons are eliminated (Table 4.1).

To construct the libraries, unique restriction sites were designed in the WT OmpF plasmid that flanked the Myc insertion sites in L5 (*SacI*/*EcoRI*) and L6 (*SacI*/*SpeI*). Further, a *StuI* site was designed in the intervening sequence between the two restriction sites to reduce library background from undigested or re-circularised singly cut plasmid. The restriction site pairs were inserted by PCR mutagenesis at each loop to yield two library plasmids (L5-lib and L6-lib). For the randomised insert, two pools of duplexed oligonucleotides containing 10 NNK residues flanked by the corresponding pair of restriction sites were commercially synthesised. The plasmids and randomised inserts were separately digested overnight to generate complementary sticky ends. The cut plasmid was then purified by gel extraction and further digested with *StuI* and Antarctic Phosphatase, which removes the 5' phosphate groups from the cut plasmid to prevent self-ligation. Both the insert and vector were purified using Zymo DNA Clean & Concentrator spin columns before ligation.

Control ligations for L5-Lib and L6-Lib were set up using 3:1, 5:1, 10:1, and 20:1 insert (~50 bp) to vector (~5000 bp) mole ratios. The highest transformation efficiency was observed using a 10:1 ratio, which was then used in the scaled-up

ligation reaction. The two libraries were separately ligated and then electroporated into MC1061 cells (Fig. 4.17). After a one-hour recovery period in rich medium at 37°C, a small aliquot of the library was plated to estimate library diversity. The remainder was used to inoculate a larger culture which was grown overnight at 16°C and then flash frozen as glycerol aliquots which contained ten times more bacterial cells than the estimated diversity of the library.

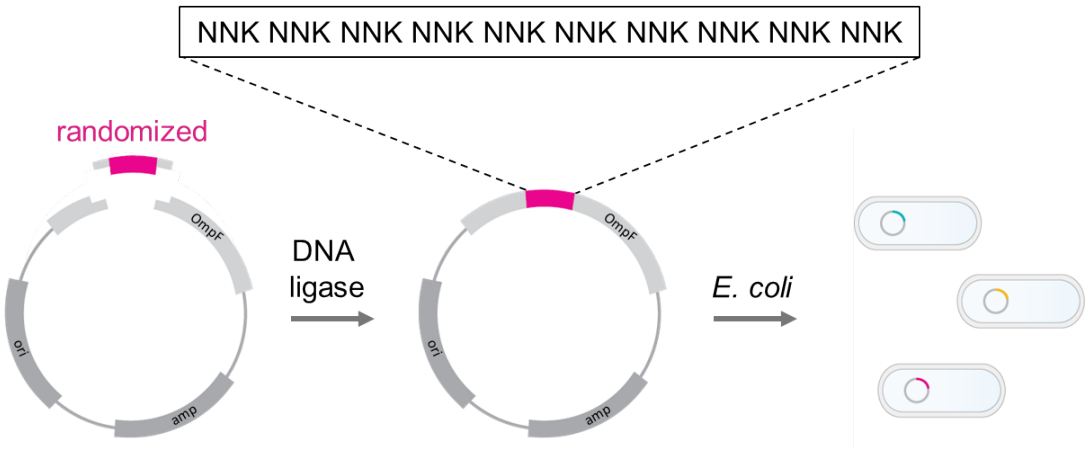


Figure 4.17 Library construction. A plasmid containing the vector backbone and an oligo pool containing 10 randomised amino acids was digested with the same set of restriction enzymes to generate complementary sticky ends. The ligated plasmids were then ligated and subsequently electroporated into competent *E. coli* to generate the bacterial display library.

Library diversity is commonly estimated at the DNA level by plating dilutions of the transformation reaction and back-calculating the total number of colonies (unique variants) that would have been formed if the entire transformation reaction was plated. This can also be written as diversity = CFU / amount of DNA plated × total amount of DNA transformed. Using this method, the L5 and L6 libraries were estimated to have 1.8×10^{8} and 8.7×10^{7} diverse peptides, respectively. This calculation yields the size of the library at the DNA level assuming that each bacterium is expressing a unique library variant. However, this is often an

overestimation of diversity because of the degeneracy of the genetic code, which means that multiple DNA sequences can code for identical peptides.

The representation of amino acids was also assessed by sequencing the DNA from at least ten colonies, all of which contained the correct insert. The observed number of each amino acid was then compared to the expected number of each amino acid in an NNK randomisation scheme [68]. Codon incidence was statistically analysed using the following equation:

$$P' = [N!/k!\,(N-k)!][f^k(t-f)^{N-k}]/t^N$$

where N is the total number of residues, k is the number of observations of a specific residue, and P' is the probability of observing k residues given that f is the number of codons encoding that residue and t is the total number of codons in the peptide set. This P' value was then corrected for the number of independent samplings by the following formula:

$$P = 1 - (1 - P')^t$$

The corrected probabilities (P) are reported in Table 4.4. In both the L5 and L6 library, glycine (GGG, GGT) was overabundant. In the L5 library, the residues tryptophan (TGG) and valine (GTG, GTT) were also overabundant, and only one stop codon was present in the ten peptide sequences analysed. Two stop codons were present in the L6 library, for which eleven peptide sequences were analysed. The prevalence of codons containing guanine or thymine may be explained by the increased coupling efficiency of the two bases during DNA oligo synthesis. From personal correspondence with Integrated DNA Technologies, it was revealed that the incorporation efficiency of guanine and thymine was up to 10% higher than adenine or cytosine in their DNA synthesis process.

		Loop 5 (1.2×10^8)					Loop 6 (8.6×10^7)				
a.a.	Codons	Count	%obs	%exp	obs/exp	P	Count	%obs	%exp	obs/exp	P
D	1	2	2.0	3.1	0.6	0.99	2	1.8	3.1	0.6	0.99
Q	1	2	2.0	3.1	0.6	0.99	2	1.8	3.1	0.6	0.99
W	1	10	10.0	3.1	3.2	0.02	8	7.3	3.1	2.3	0.26
E	1	4	4.0	3.1	1.3	0.98	3	2.7	3.1	0.9	0.99
F	1	2	2.0	3.1	0.6	0.99	0	0.0	3.1	0.0	0.46
C	1	4	4.0	3.1	1.3	0.98	5	4.5	3.1	1.5	0.94
N	1	1	1.0	3.1	0.3	0.94	1	0.9	3.1	0.3	0.90
Y	1	0	0.0	3.1	0.0	0.57	2	1.8	3.1	0.6	0.99
H	1	2	2.0	3.1	0.6	0.99	3	2.7	3.1	0.9	0.99
I	1	1	1.0	3.1	0.3	0.94	2	1.8	3.1	0.6	0.99
M	1	3	3.0	3.1	1.0	0.99	4	3.6	3.1	1.2	0.99
K	1	2	2.0	3.1	0.6	0.99	3	2.7	3.1	0.9	0.99
G	2	10	10.0	6.3	1.6	0.62	22	20.0	6.3	3.2	0.00
A	2	6	6.0	6.3	1.0	0.97	6	5.5	6.3	0.9	0.97
T	2	5	5.0	6.3	0.8	0.97	2	1.8	6.3	0.3	0.36
P	2	4	4.0	6.3	0.6	0.93	4	3.6	6.3	0.6	0.86
V	2	13	13.0	6.3	2.1	0.11	10	9.1	6.3	1.5	0.75
S	3	6	6.0	9.4	0.6	0.80	4	3.6	9.4	0.4	0.23
L	3	9	9.0	9.4	1.0	0.95	11	10.0	9.4	1.1	0.93
R	3	13	13.0	9.4	1.4	0.70	14	12.7	9.4	1.4	0.70
Stop	1	1	1.0	3.1	0.3	0.94	2	1.8	3.1	0.6	0.99
	Total	100					110				

Table 4.4 Representation of amino acids in the L5 and L6 libraries calculated by sequencing ten (L5) or eleven (L6) peptide inserts from sorted bacteria.

4.5.2 Library FACS

The anti-Myc tag antibody used in control experiments served as the initial protein target for the library screen. The screen was carried out with the L5 library in equilibrium mode (sorted in the presence of unbound antibody). A library aliquot was thawed on ice and used to inoculate a 5 mL culture which was grown until an OD=0.4-0.5 (~3.6 x 10^8 cells/mL) before induction with arabinose for one hour. Following this, a volume containing 1×10^7 cells was pelleted and resuspended in 1

mL RM medium with fluorescent antibody (10 nM). The labelling reaction was allowed to proceed for 45 min at 4°C on a rotary mixer before FACS (Fig. 4.18 A).

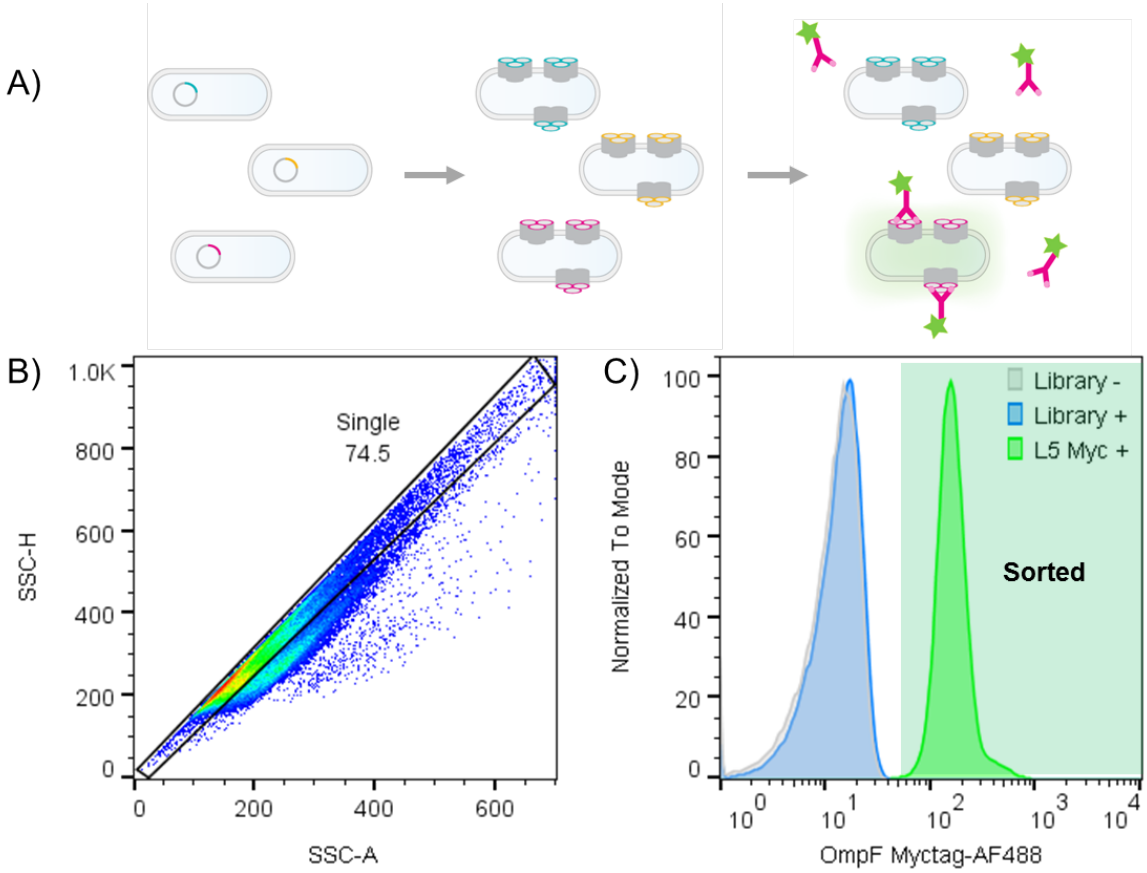


Figure 4.18 (A) Schematic of screening by bacterial display. Only cells displaying an OmpF peptide variant that binds the fluorescent antibody will fall within the sort gate and can be enriched from the rest of the population by FACS. **(B)** Density plot of side scatter area (SSC-A) versus side scatter height (SSC-H) from 100,000 L5-lib events. Only events falling within the 'Single' gate (74.5%) were considered for the sort and plotted in the fluorescence histogram. **(C)** Fluorescence histogram of cells within the 'Single' gate from the uninduced library (grey), the induced library (blue), and the positive control (green). A sort gate was set to encompass all fluorescent values above the background. Cells falling within this gate were sorted into collection tubes containing 1 mL of rich medium.

To begin the sorting process, 100,000 events were measured, and the light scattering characteristics of each cell were recorded. A plot of side scatter area versus side scatter height was used to gate for single bacterial cells (doublets or

aggregates will have a similar side scatter height, or maximum fluorescence, but a larger area, as they will take longer to pass through the interrogation point). Only events within the 'Single' gate were plotted on the fluorescence histogram and considered for the sorting process (Fig. 4.18 B). A sort gate was set to include all bacteria exhibiting fluorescence values larger than the background, determined from the initial 100,000 events (Fig. 4.18 C). During the sort, a total of 8.5×10^6 events were measured, of which 2,169 events (0.0002%) fell within the fluorescence gate and were collected in 1 mL of SOC medium. The L5-Myc positive control was sorted last to avoid contamination, and was used to ensure that expression and labelling was successful.

The sorted cells were immediately plated on RM-agar. The library sort produced 244 CFUs in total (recovery=12%), whereas the positive control recovery was 61%, in line with previous data from initial CFU counts (Table 4.3). The lower recovery for the library sort is likely due to the duration of the sort (~30 min for the library compared to a few seconds for the control), which is the only differing parameter between the two samples.

To ensure that library storage conditions and overexpression of OmpF does not compromise the viability of the library before FACS, a live/dead stain was later carried out with propidium iodide. Two library aliquots were cultured, one of which was left uninduced while the second was induced using standard conditions. As a 'dead' control, an aliquot of the uninduced cells was treated with 70% isopropanol for 30 minutes. As a 'live' control, a second uninduced aliquot was left untreated. The two controls and a sample of the induced library were then stained with propidium iodide before FACS. For each sample, 100,000 events were measured (Fig. 4.19). A small population (1.4%) of both the uninduced live control and the

induced sample was stained by the propidium iodide, indicating cell death. This viability loss may result from library preparation and storage conditions. A live/dead stain should be used in future sorts to ensure that dead bacteria, which are more likely to non-specifically bind the protein target to produce false positives, are not considered.

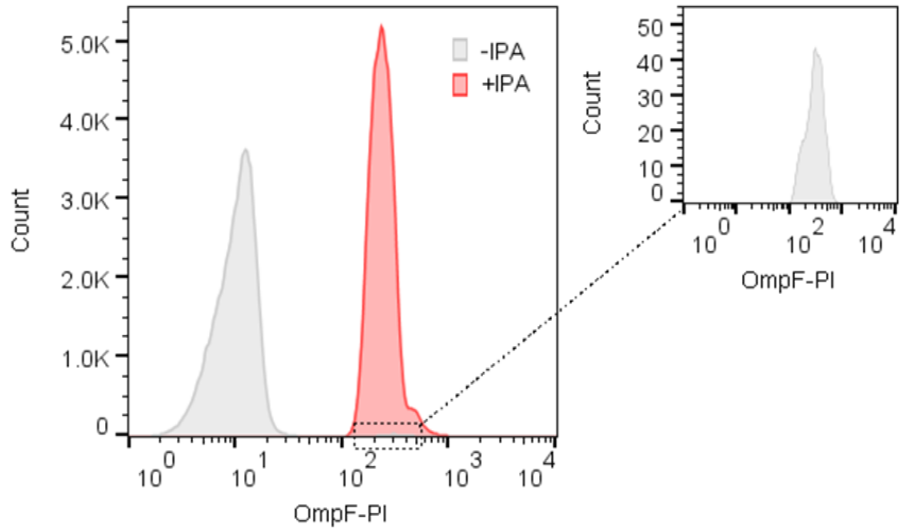


Figure 4.19 Live/dead stain of bacterial display library. Cells were stained with propidium iodide (3 μM) for 30 min before FACS. For the dead control, cells were incubated in 70% isopropanol before staining (+IPA) and compared to cells from the induced library (-IPA). Of the 100,000 events measured, only 1.4% of the cells measured fell within the dead control fluorescence values (inset).

4.6 Extracting screened nanopores from bacteria

In order to identify the optimal nanopore sensor(s), the screened candidates must be transferred from bacteria to synthetic bilayers for electrophysiological characterisation. A high-throughput protocol was established for selectively extracting OmpF from hundreds of *E. coli* colonies, which can be completed in one day.

Bacteria from a single colony were used to inoculate separate wells

containing 1 mL of RM medium in a 96-well plate. The plates were then incubated in a heating block at 37°C with shaking. At an OD=0.4-0.5, cultures were induced with arabinose for two hours and then a volume of cells corresponding to 1 mL of OD=0.3 (~80 µL) was transferred to a second 96-well plate and pelleted at 4°C. The supernatant was removed by flipping the plates upside down, and the remaining pellets were resuspended in extraction buffer: 2% βOG, 5 mM EDTA, 25 mM Tris-HCl pH 8.0. Disruption of the LPS layer by EDTA was critical for efficient OmpF solubilisation. The pellets were incubated with the extraction buffer for 15 min at 37°C with shaking and then centrifuged at 4°C for 30 min at maximum speed.

The extraction conditions were first optimised using cells overexpressing WT OmpF (Fig. 4.20 A), after which the method was tested on the two positive controls (Fig. 4.20 B), before being applied to extract screened library candidates (Fig. 4.20 C). An aliquot of the supernatant was removed from each well, mixed with loading dye, and analysed by SDS-PAGE to determine OmpF expression levels and extraction purity. A second aliquot was removed from each well for direct use in PLB current recordings (Section 4.7). All volume transfers were carried out with a multichannel pipette.

To maintain the genotype-phenotype link, the remainder of the original 96-well plate culture was also pelleted and frozen at -20°C. To prove that the DNA encoding each nanopore could be recovered, plasmids from six frozen pellets were isolated two weeks later and the loop inserts successfully sequenced. One insert contained a stop codon, which indicates sorting of false positives. Glycerol stocks can also be made for longer-term storage.

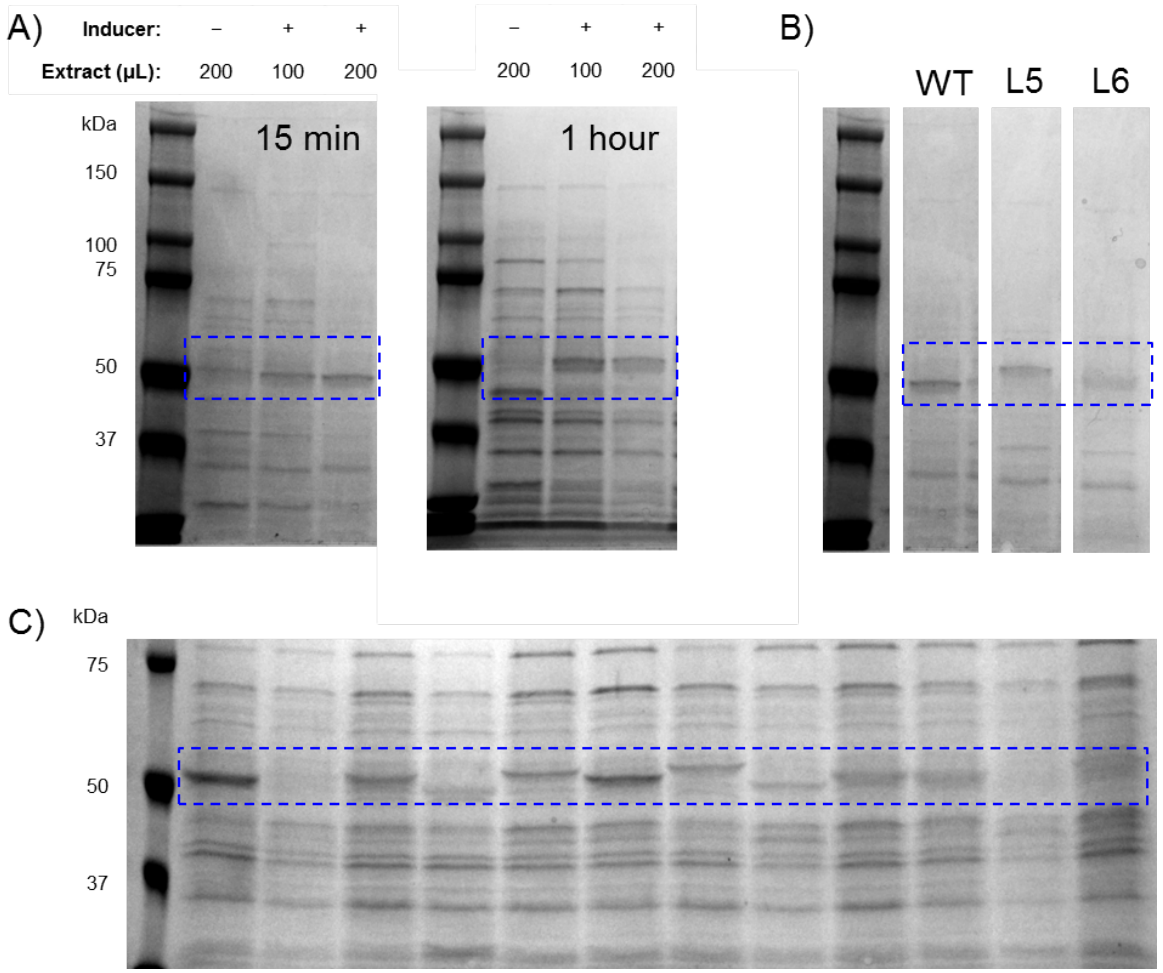


Figure 4.20 (A) SDS-PAGE of WT OmpF extracted from *E. coli* in 2% βOG, 5 mM EDTA, 25 mM Tris-HCl pH 8.0. Different extraction buffer volumes (100 vs 200 µL) and incubation times (15 min vs 1 hr) were tested and compared to the uninduced control. Incubation for 15 min in 200 µL extraction buffer was found to produce the highest purity samples. The blue dashed rectangle highlights the location of the OmpF trimer. **(B)** Extract of WT OmpF and positive controls (L5-Myc, L6-Myc) using the optimised protocol from (A). **(C)** Representative gel of FACS-screened OmpF variants extracted in a high-throughput 96-well plate format.

4.7 Electrophysiology of screened nanopores

The extracted nanopore candidates were tested for antibody binding using two electrophysiology methods, PLB and DHB recording (Fig. 4.21). This step was necessary as not all OmpF variants screened by FACS will be useful for stochastic sensing due to false positives or the orientation in which the target protein binds the nanopore, which may not always produce a detectable change in the ionic

current compared to the unbound state. Further, the binding affinity at the cell surface, which includes negatively-charged LPS and possible avidity effects from OmpF clusters [213], may not translate to the affinity observed in vitro. Single-channel PLB recording is a powerful, low-noise technique owing to the small size of the bilayer (30-100 μm) that can be formed. However, throughput is low because a fresh bilayer must be formed for each tested nanopore. Because of this, the DHB method was explored first.

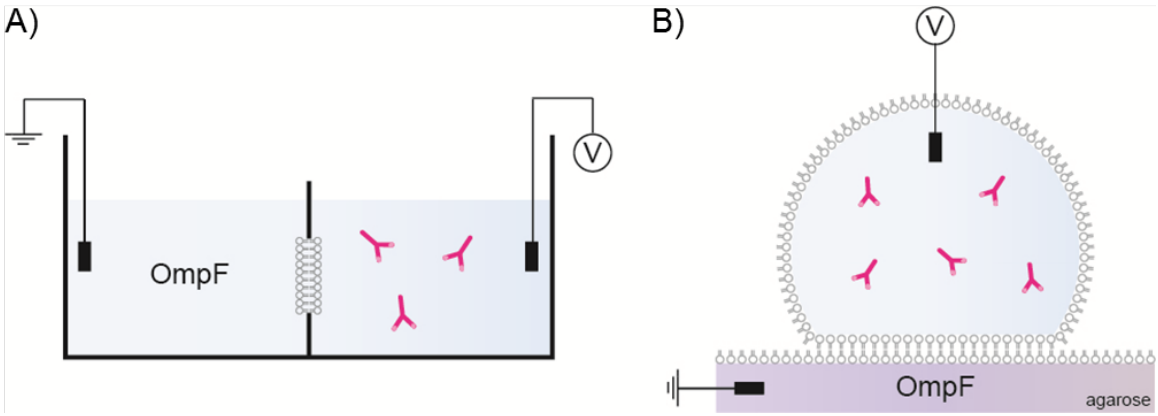


Figure 4.21 The two electrophysiology approaches used to test nanopores screened from the bacterial display library. **(A)** The PLB technique, in which a bilayer is formed between two buffer-filled compartments. OmpF is added to one side of the bilayer (ground) and the target protein (pink antibody) can be introduced on either side of the bilayer, depending on the orientation of the reconstituted nanopore. **(B)** The droplet-hydrogel bilayer technique, in which a bilayer is formed between agarose (ground) and a buffer-filled droplet. OmpF is added to the agarose and the target protein is introduced in the droplet. The contents of the droplet and hydrogel cannot be easily altered after bilayer formation.

4.7.1 DHB electrophysiology

A DHB is formed between an aqueous droplet, which can be as small as a few picolitres, and a hydrogel. The hydrogel and droplet are immersed in a lipid-containing oil to coat each surface in a lipid monolayer. When the two are brought into contact, a bilayer spontaneously forms [27]. Nanopores can be added to

either the droplet or the hydrogel, and will insert spontaneously into the bilayer. In some cases, an applied voltage is necessary for efficient insertion [27]. Interestingly, it has been shown that membranes isolated from cells can be directly pipetted onto the hydrogel before lipid is added for efficient protein incorporation into the bilayer. This method was previously used to characterise a range of proteins overexpressed in both bacteria and mammalian cells [214]. Although the reconstitution mechanism is not clear, the authors hypothesised that the membrane fragments are incorporated during formation of the DHB.

A similar approach was developed to reconstitute screened nanopores into DHBs to test antibody binding in a high-throughput manner. First, a 96-well hydrogel was formed by adding molten 1.5% (w/v) agarose in recording buffer (100 mM KCl, 20 mM phosphate buffer pH 7.0) to a custom holder machined from poly(methyl methacrylate) (PMMA). A bottomless 96-well plate was then placed on top, allowing the agarose to fill each well. After the agarose had set, a different nanopore extract was pipetted (1–3 μL) onto the agarose in each well. After 10 minutes, the wells were filled with a solution of lipid-in-oil (10 mg/mL DPhPC in hexadecane). A monolayer was allowed to form at the agarose-oil interface for at least 30 minutes (Fig. 4.22 A). In this initial version, the volume of oil in each well was 200 μL, equating to 488 mg of lipid if all 244 screened nanopores were tested. Lowering the concentration of the lipid from 10 mg/mL increased the chance of coalescence during DHB formation. Instead, a PMMA stamp was machined to create square pyramidal indentations in the agarose filling each well, which reduced the volume of lipid-in-oil needed to fill the well to a sufficient height from 200 to 50 μL (Fig. 4.22 B). The stamp was applied to the 96-well plate after it had been filled with molten agarose.

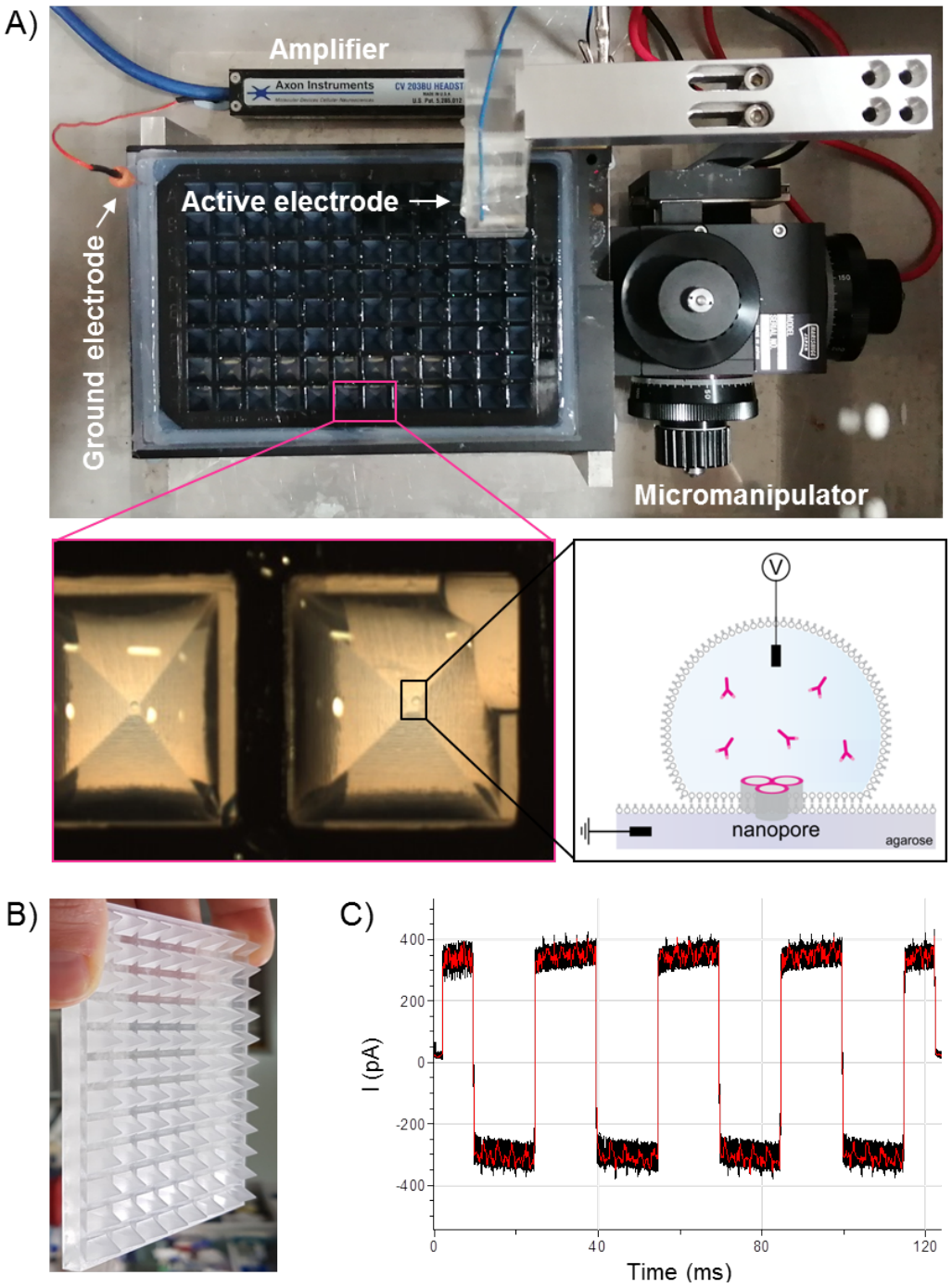


Figure 4.22 (A) The 96-well DHB set-up in a Faraday cage. The common hydrogel is grounded by one electrode, and the second active electrode can be moved in the X, Y, and Z direction using a micromanipulator to contact the droplet in each well. The stamp in (B) is placed on top of the molten agarose filling each well to create square pyramidal indentations (pink inset), which are filled with lipid-containing oil before the droplets are loaded to create a DHB in which OmpF inserts (black inset). **(B)** The PMMA stamp, containing 96 square pyramidal protrusions. **(C)** Representative current trace of the capacitance of a DHB, generated by a 1 V triangle waveform. A capacitance of 400 pF corresponds to a circular bilayer diameter of ~280 µm assuming a membrane capacitance of 0.64 µF/cm^2 for DPhPC [28].

A single droplet was introduced into each hydrogel well using one of two methods. In the first method, 200 nL (~720 μm) droplets were formed in a separate glass cuvette filled with 10 mg/mL DPhPC in hexadecane using a pipette. The droplets were then pipette transferred from the cuvette to each well. Coalescence was low (<10%) if the droplets were incubated for at least thirty minutes before transfer. In the second method, the droplets were formed by microinjection. The buffer solution, with or without the protein target (in this case, anti-Myc tag antibody), was loaded into a capillary which was then attached to the microinjector. The capillary tip was submerged into the oil and a droplet was ejected into each well. To release the droplet, the microinjector was slowly removed from the oil after a few seconds, allowing time for monolayer formation. Droplets ranging in size from 10–50 nL (~260–460 μm) were made without significant coalescence (<20%) despite the lack of incubation time, owing to their smaller size.

During current recordings, one electrode was placed in the hydrogel (cis, ground) and the other electrode was placed in a droplet (trans, voltage applied). The trans electrode was coated with a thin layer of 1.5% low-gelling agarose and lowered through the oil into the droplet using a micromanipulator. Although successful bilayer formation and protein insertion was sometimes observed by changes in the bilayer capacitance (Fig. 4.22 C), results were highly variable. This may be in part due to the roughness of the porous agarose surface or the insertion of the agarose-coated electrode (100 μm diameter) into the small droplets (~300 μm), which may distort the droplet and inhibit bilayer formation.

Because of the poor reproducibility in bilayer formation, size, and stability in the 96-well DHB set-up, the PLB method was used instead to test the screened nanopores. Further optimisation of the DHB platform, including lipid-oil

combinations, thinner electrodes, and built-in electrical circuitry could evolve the method for effective high-throughput screening. It is worth noting that DHBs are closed systems in terms of the aqueous buffers on either side of the bilayer. Because OmpF inserts into bilayers in two orientations, some pores may not display the peptide insert on the protein-containing side of the bilayer (the droplet), and will therefore not be suitable for screening.

4.7.2 PLB electrophysiology

Of the 244 CFU resulting from the L5 library cell sort, nanopores were extracted from 37 colonies and tested in single-channel PLB recording. Current recordings were carried out as described in Chapter 2. Briefly, a DPhPC bilayer was formed between two compartments filled with 100 mM KCl, 20 mM phosphate buffer pH 7.0. The detergent nanopore extract was directly added to the cis chamber and inserted into the bilayer by applying a potential. After insertion of one to a maximum of three channels (bilayers with more than three channels were re-formed), an I-V curve was measured to establish OmpF orientation. Following this, the background current was recorded for at least 5 minutes before addition of anti-Myc tag antibody (5 µL, 67 nM final). The current was then monitored for another 15–30 minutes before the trace was analysed for antibody binding, which would manifest as a step-wise change in the current.

Although the extracted nanopores inserted efficiently into a PLB (Fig. 4.23 A), of the 37 nanopore variants tested (at least 3 channels were recorded per variant) no successful stochastic sensors were identified for the anti-Myc tag antibody protein target. However, antibody detection using the positive control L5-Myc nanopore extracted from *E. coli* using the same protocol was successful (Fig. 4.23 B). Two stepwise current drops were observed after antibody addition to the

recording chamber, which were not reversible for the duration of the recording (~30 min). The antibody must therefore bind the Myc epitope more tightly with the added flexibility of the flanking SG linkers (compare to reversible binding to the Myc epitope without the added linkers in Fig. 4.7), or the antibody may bivalently bind to two Myc epitopes, explaining the two steps observed in the current trace. Regardless, this positive control experiment demonstrates that the established workflow forms a strong foundation for future identification of stochastic sensors from nanopore libraries.

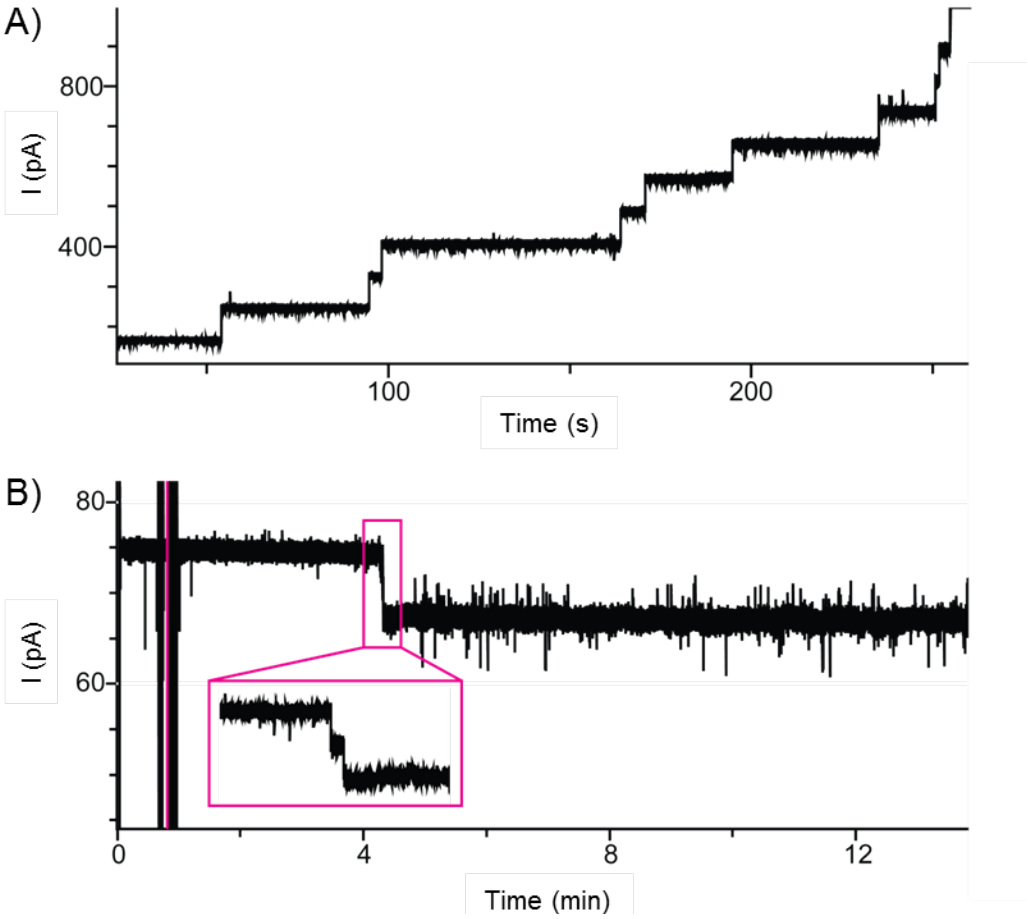


Figure 4.23 Single-channel PLB recordings of OmpF extracted from bacterial cells. **(A)** Step-wise insertion of ten OmpF trimers from the extract in ~200 seconds. **(B)** Antibody detection with the positive control L5-Myc nanopore extracted from bacteria. The antibody (67 nM) is added at the pink line and reacts within two minutes (pink inset, n=3). Antibody unbinding was not observed in the current trace over the course of 30 min. The recording buffer was 100 mM KCl, 20 mM phosphate buffer pH 7.0 and the applied potential was +70 mV.

4.8 Conclusions

This chapter has described a method for engineering and screening nanopore libraries to develop stochastic protein sensors for any protein target of interest. Compared to previous work on protein detection, in which specific epitopes were introduced into a nanopore, binding sequences will be screened within the structural context of the nanopore. This will allow the identification of sensors with optimal binding affinities and current signatures.

Label-free protein detection with OmpF was demonstrated using two targets (an antibody and a kinase) representing a range of molecular sizes and binding affinities, to showcase the dynamic range of nanopore stochastic sensing. Following this, two high-diversity libraries were engineered and screened by bacterial display, after which a high-throughput protocol was developed for extracting screened nanopores from bacteria. Progress was made in the development of a 96-well plate electrophysiology screening method for parallel testing of OmpF variants. Although no successful sensors were identified for the chosen protein target, an effective workflow was established for the design, screening, and testing of nanopore libraries. Further system improvements which might lead to successful nanopore development in the future are discussed in Chapter 5.

5

Future developments

5.1 Introduction

This thesis details work centred on the bacterial trimeric porin OmpF. Single-channel PLB current recording continues to be a popular method for characterising membrane proteins. First, the orientation of OmpF in a PLB was determined by chemical modification experiments, such that directional transport across OmpF inserted in a known orientation can now be interrogated. Next, a protocol was developed to express OmpF in vitro, allowing fast production of OmpF mutants compared to conventional in vivo expression. This protocol was applied to construct OmpF heterotrimers, in which one or two of the three barrels in the trimer can be separately modified. Finally, a peptide library was genetically engineered into the peripheral loops of OmpF, in order to develop nanopores for label-free and tailored protein sensing. Methods for screening the library and testing the screened variants were established. This chapter focuses on future developments to these three facets and their potential applications.

5.2 The orientation of porins in planar lipid bilayers

The targeted covalent modification technique used to determine the orientation of OmpF in **Chapter 2** can be generally applied to determine the orientation of other OM proteins, including the homologous trimeric porins OmpC and PhoE, as well as the siderophore transporter FhuA. Like OmpF, FhuA has been reported to exhibit bidirectional insertion into PLBs, as evidenced by its two types of electrical asymmetries [82]. Although particularly useful for membrane proteins such as porins and toxins, including the membrane attack complex perforin (MACPF) and cholesterol dependent cytolysin (CDC) superfamilies, the technique might also be used to establish the orientation of channels and transporters for in vitro studies.

The results of the OmpF orientation study are being applied in our laboratory to study colicins. Colicins are strain-specific antibacterial peptides secreted by *E. coli*, a majority of which utilise OmpF to penetrate the bacterial OM—the initial barrier to cell entry [215]. Specifically, the energy-independent threading of Colicin E9 from the extracellular side of OmpF is being examined at the molecular level by single-channel PLB current recording to construct a detailed kinetic mechanism of transport. The 61 kDa polypeptide, which contains a nuclease domain, has been reported to translocate through OmpF using an intrinsically unstructured domain by first passing through one subunit before looping around to occupy a second subunit, forming a hook-like conformation (Fig. 5.1) [216]. This hooked translocon exposes a TolB-binding epitope in the periplasm which engages the Tol system for further transport into the cytoplasm, where cytotoxicity is exerted [69,216].

Our studies are focused on showing that this translocation mechanism is kinetically feasible by monitoring the interaction of single OmpF channels of

defined orientation with Colicin E9 and dissecting the role of various segments in its translocation domain (unpublished data). Colicins and other bacteriocins kill cells with nanomolar efficacy, and have been extensively investigated because of their potential as species- and even strain-specific antibiotics [217]. However, previous PLB studies did not establish OmpF orientation and thus could not distinguish the interaction of colicins or colicin-derived peptides with the extracellular or periplasmic side of OmpF—a necessity for studying directional transport processes [95,216,218,219].

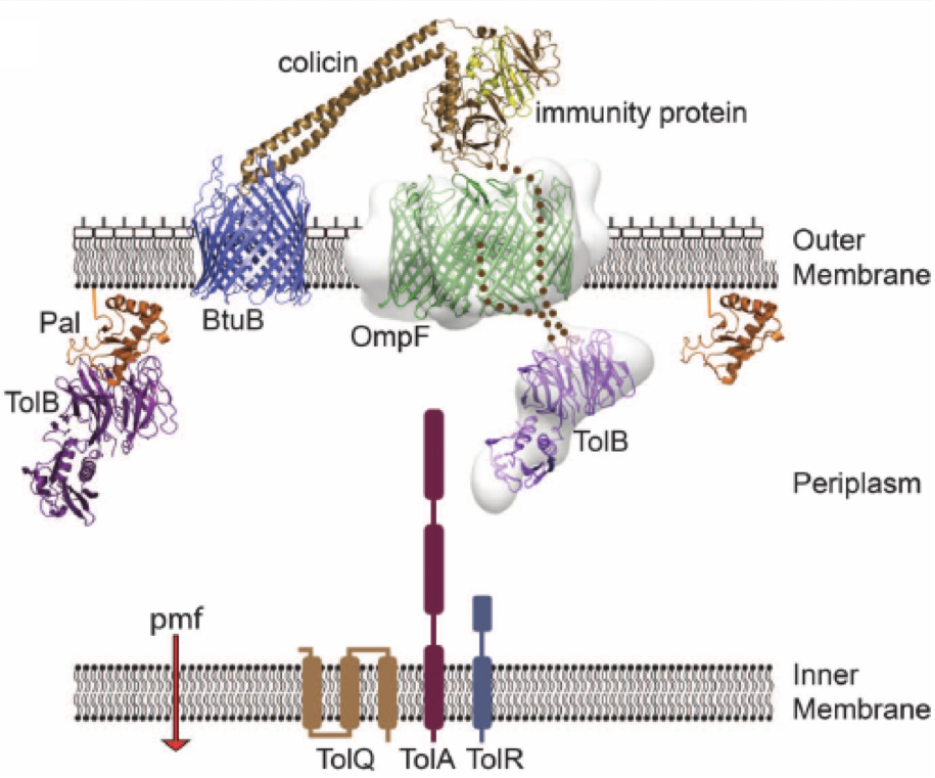


Figure 5.1 Structural representation of the intrinsically unstructured translocation domain of Colicin E9 threading through OmpF. The hooked conformation exposes a TolB binding epitope in the periplasm, which is used to engage the Tol complex for energy-dependent transport of the nuclease domain into the cytoplasm, which requires dissociation of the bound immunity protein. Reproduced from [216].

The symmetric DPhPC bilayer widely used in PLB studies is highly simplified compared to the asymmetric bacterial OM, which contains negatively charged LPS in the outer leaflet and porin microdomains [213,220]. LPS has been reported to play a role in both membrane protein insertion and the translocation of bacteriocins [221–223]. Further, bidirectional insertion of OmpF is not observed in the OM, where assembly of the porin is guided by chaperones and the β-barrel assembly machinery (Section 3.1.3) [224,225]. For future work on OM physiology, it will be important to create PLBs that more closely mimic this complex bilayer network, which is also being explored in our group. The technique described in Chapter 2 has been published as a protocol (Springer, in press), which may be applied to determine the orientation of protein components in symmetric or more native PLBs.

5.3 Applications of OmpF heterotrimers

A protocol for in vitro expression of OmpF trimers was presented in **Chapter 3**. The expression protocol was applied to engineer heterotrimers, in which one or two of the identical OmpF barrels contain a different modification. Purification of different heterotrimer species was achieved by introducing genetic tags to the N-terminus of modified monomer subunits; incorporation of each additional tagged subunit decreased the electrophoretic mobility of the trimer compared to the WT homotrimer, i.e. the tagged homotrimer migrated the slowest. By folding a combination of WT and tagged monomers expressed in vitro, different homotrimer and heterotrimer species containing from zero to three tags were produced and resolved by SDS-PAGE.

Future work will focus on producing heterotrimers with different functionalities for biosensing. In the simplest example, a single binding element

can be introduced per trimer for stochastic sensing of target analytes. The presence of three binding elements in a homotrimer can lead to parallel binding events which complicate interpretation of the current read-out. This is especially true for multivalent proteins, such as antibodies, which have the potential to simultaneously bind two of the three binding elements in a homotrimer. Heterotrimers with only one binding element would simplify analysis of the current readout for all analytes, regardless of structure or valency. Furthermore, OmpF can be engineered to contain two unique binding elements per trimer for multiplexed detection with one nanopore, which has not been demonstrated. In addition, the heterotrimer protocol might also be applied to produce chimeric trimers composed of OmpF, OmpC, or PhoE subunits, which share over 50% sequence similarity [226].

Although challenging, it would be interesting to engineer OmpF with three unique barrels (Fig. 3.6). One method of achieving this might be to use two different types of tags, which in this example will be denoted as ‘A’ and ‘B’. Mixing three different types of monomers would produce too many unique trimers for separation by SDS-PAGE. Instead, affinity or ion exchange chromatography could be used to first purify heterotrimers that contain only one A tag based on the retention time. Previously, separation of homotetrameric species with different numbers of FLAG tagged subunits (DYKDDDK) was achieved by anion exchange chromatography. The more tagged TTR subunits incorporated into the tetramer, the longer the retention time was on the anion exchange column (Fig. 5.2) [227].

The remaining three OmpF trimer species (1 A, 2 WT or 1 A, 2 B or 1 A, 1 WT, 1B) would then be resolvable by SDS-PAGE to identify the heterotrimer that only has one B tag and therefore must be made up of three unique barrels. This

approach would require more protein than can currently be produced by in vitro expression without becoming prohibitively expensive, so in vivo expression might be considered. Future work will focus on making OmpF heterotrimers in vivo by expressing WT and tagged monomers in a polycistronic vector.

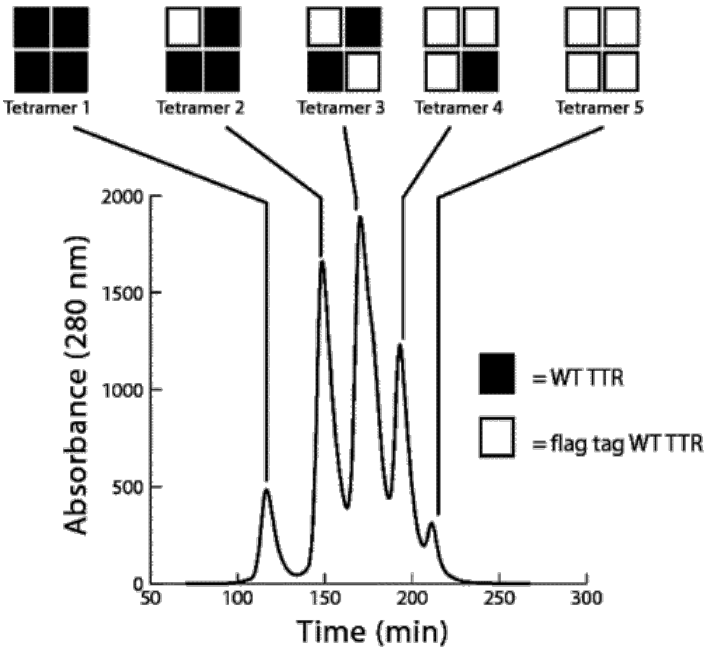


Figure 5.2 Chromatogram of transthyretin (TTR) oligomers containing zero to four FLAG tags. The different tetrameric species were separated on an anion exchange column by retention time. Reproduced from [227].

Heterotrimers with tailored non-redundant binding sites would allow for simultaneous detection of up to three analytes. The pores could also be engineered to be stimulus responsive with respect to the translocation of substances, such as antibiotics and salts, for controlled release. This was previously achieved by attaching molecular caps to OmpF that acted to block or unblock the pore upon changes in the environment, such as pH or redox potential [141,142,145]. OmpF engineered in this manner produced an all-or-none response; using heterotrimers, multiple stimuli can influence the output to produce

a dialled response. These OmpF protein gates might eventually be incorporated into artificial organelles to precisely control their function.

5.4 Protein sensors screened from a genetically engineered nanopore library

Chapter 4 details a nanopore peptide library genetically engineered into the surface-exposed loops of OmpF, which folds into the OM of *E. coli*. The bacterial display library was screened by FACS to identify a nanopore sensor for a commercially available antibody as an initial proof-of-concept. The ultimate goal of this project was to construct a nanopore library from which sensors for any protein target of interest could be screened. An optimal nanopore sensor must not only bind the target protein, but bind it in an orientation that produces a discernible change in the ionic current trace. Because of this, the screened sensors were tested in single-channel current recordings using both DHB and PLB electrophysiology. Initial efforts were geared towards making a 96-well DHB set-up for parallel current recording of screened variants, but progress was hindered because of inconsistent bilayer formation. Therefore, testing was eventually carried out using the PLB technique, which was low-throughput but provided reproducible bilayer formation and nanopore insertion.

An initial library screen against the anti-Myc tag antibody identified 244 nanopore candidates. Of these, 38 nanopores were tested by single-channel PLB recording and 0 successful sensors were identified. The Myc tag epitope (EQKLISEEDL) has a K_d of 500 nM in phosphate buffer at 25°C, similar to the binding conditions used in this thesis [197]. The contribution of each amino acid to the measured affinity was previously analysed by exchanging each position with all of the other 19 proteinogenic amino acids [195]. It was possible to truncate the

epitope to KLISEEDL with only an order of magnitude decrease in affinity (K_d = 25 μM). The core LISE and L10 positions showed no or very limited exchangeability. The affinity of the original epitope could partly be restored by certain amino acid substitutions outside the core sequence (e.g., KLISEFEL). Therefore, it was expected that a binder with sufficient affinity for nanopore sensing would be present in the library.

A more probable reason for the failed screening process was the high rate of false positives that arose from sorting <.001% of the population from only one round of FACS. In the future, bacteria which bind the target protein will first be enriched by magnetic activated cell sorting before FACS to increase the chances of identifying high-affinity nanopore sensors. If the affinity is too high for effective stochastic sensing, as was observed with the Myc epitope, point mutations can be introduced to reach the desired binding profile.

Other library designs can also be considered, including different codon randomisation schemes and linkers between the peptide insert and OmpF (currently six serine or glycine residues). Previously, a TEV cleavage site was engineered into the linker region of a kinase nanopore sensor [5]. This feature could also be incorporated into the library design if linear rather than constrained peptides are desired. Additionally, the library interface can be extended to two adjacent loops to expand the binding interface and increase specificity, which might allow detection of structurally similar proteins such as isoforms and splice variants. For example, nanopore sensors that bind only one protein isoform can be selected by labelling each isoform with fluorescent probes of different colours. During FACS, thresholds can be set to sort only cells that have high fluorescence in the desired isoform channel and low fluorescence in the remaining channels.

Future progress in high-throughput current recording will advance the development and screening of nanopore libraries. The 96-well DHB method described in Chapter 4 can be further engineered to contain built-in circuitry and switches for automated nanopore current recordings in parallel. Different lipid-in-oil mixes must also be systematically screened to identify the optimal composition for bilayer formation and stability. Once the process is optimised, the library will be screened again to see if nanopore sensors against various protein biomarkers can be identified. These sensors will then be incorporated into a chip for highly parallel and multiplexed protein detection (Fig. 4.5). The ability to measure biomarkers accurately and precisely over a wide dynamic range—which is a major challenge for existing protein detection techniques [158]—might be of particular utility for diagnosing diseases and monitoring response to therapy if issues with specificity, arising from the complexity of biological samples, are addressed [159]. However, the nanopore library technology is not limited to protein detection and may also be developed to detect small molecules, lipids, carbohydrates, and nucleic acids, or to target bacteria carrying a therapeutic payload to specific tissues, such as tumours.

5.5 Conclusions

This chapter has detailed future optimisation and applications of the OmpF nanopore engineering strategies presented in this thesis. As a well-studied and robust integral membrane protein, OmpF is a promising alternative for stochastic sensing. Future work will further expand the field of nanopore sensing for novel applications in biotechnology.

6

Materials and methods

6.1 General methods

Chemicals and oligonucleotides were purchased from Sigma Aldrich unless otherwise stated. Lipids were from Avanti Polar lipids. Enzymes and ladders were from New England Biolabs. Routine cloning was performed in XL-10 Gold cells (Stratagene) and routine plasmid purification was performed using QiaPrep Spin Miniprep or Maxiprep kits (Qiagen). Plasmid mutagenesis was confirmed by sequencing (Source BioScience) in all cases. All cell cultures were carried out in 5 mL medium in 50 mL Falcon tubes (Corning) at 37°C, 250 RPM unless otherwise noted. DNA quantitation was performed by UV-visible absorption spectroscopy using a NanoDrop 1000 (Thermo Scientific). OD measurements were taken using a Bio-Rad SmartSpec™ Plus spectrophotometer in 1 mL quartz cuvettes. Water was purified through a Milli-Q system fitted with a 0.22 μm membrane filter (Millipore).

6.2 Ch 2: Orientation of OmpF in planar lipid bilayers

The OmpF plasmids used in this study were a generous gift from Dr. Nicholas Housden and Renata Kaminska (University of Oxford). The *ompF* gene was

amplified from *E. coli* B^E3000 genomic DNA by PCR (forward primer: 5′-AAT AAA CCA TGG TGA AGC GCA ATA TTC TGG CAG TGA TCG-3′; reverse primer: 5′-AGA GGT CTC GAG TTA GAA CTG GTA AAC GAT ACC CAC-3′) and the product was sub-cloned into the *NcoI/XhoI* sites of pBAD/*Myc*-His B (Invitrogen) to produce pNGH71. A mutation (M2V) was introduced at the *NcoI* site in the second residue of the OmpF signal sequence, which was reverted to the WT sequence through whole-plasmid PCR mutagenesis to give pNGH72, which allowed arabinose-induced expression of WT OmpF. Whole-plasmid mutagenesis was used to introduce the D221C, N246C, and E29C mutations into pNGH72. Mutants L147C, S53C, and E29C were previously used for cross-linking studies and prepared in a similar manner.

6.2.1 OmpF expression and purification

Plasmids encoding WT and mutant OmpF were transformed into E. coli BZB1107 cells (*ompF-*, *ompC-*) [228]. The transformants were used to inoculate LB (5 L) containing ampicillin (100 µg/mL). Cultures were grown at 37°C, with shaking, to OD_{600}≈0.7, then expression was induced by the addition of arabinose (0.2 %, w/v). The cultures were grown for a further 4 h at 37°C before harvesting by centrifugation. OM fractionation and extraction of OmpF was carried out as previously described [69]. Pure OmpF trimer was obtained through three chromatography steps. First, detergent-extracted protein was loaded onto a 5 mL Q-Sepharose column (GE Healthcare) equilibrated in Tris-HCl (20 mM, pH 8.0) containing EDTA (5 mM) and βOG (1 %, w/v). Unbound material was washed from the column before elution of bound protein with a gradient of LiCl (0–500 mM) over ten column volumes. Eluted OmpF was pooled and further purified on a 26/60 Sephacryl 300 column (GE Healthcare) equilibrated in Tris-HCl (20 mM, pH 8.0)

containing EDTA (5 mM) and βOG (1 %, w/v). Final purification was performed on a 4.6/100 monoQ column (GE Healthcare) equilibrated in Tris-HCl (20 mM, pH 8.0) containing EDTA (5 mM) and βOG (1 %, w/v). Bound material was eluted with a gradient of LiCl (0–250 mM) over 20 column volumes and stored in elution buffer. Purity was assessed by SDS-PAGE (Fig. 6.1).

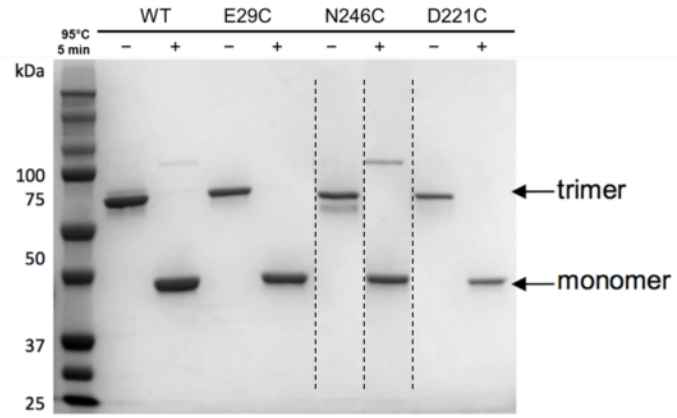


Figure 6.1 Analysis of OmpF purity by SDS-PAGE. The monomer and trimer bands of three cysteine mutants used in the orientation study are labelled. To obtain monomers, the sample was mixed with SDS-containing loading dye and heated at 95°C for 5 minutes. Lanes joined together are demarcated with a dashed line.

6.2.2 Single-channel PLB recordings

PLB recordings were carried out in 100 mM KCl, 20 mM phosphate buffer pH 7.0 with freshly added tris(2-carboxyethyl)phosphine (TCEP, 200 µM final). 1.0 L of buffer was made by mixing 15.4 mL of 0.5 M K_2HPO_4 and 9.6 mL of 0.5 M KH_2PO_4 with 7.5 g of KCl. The pH was adjusted to 7.0 with KOH. All buffers were filtered through a 0.22 µm polyethersulfone membrane. A lipid bilayer of DPhPC dissolved in pentane was formed by the Montal–Mueller solvent-free method across a 100 µm diameter aperture in a 25-µm-thick PTFE film (Goodfellow, Huntingdon, UK) [229]. The film separated two 0.5 mL Delrin compartments

designated “cis” and “trans”. The cis compartment was connected to ground, and voltage (positive or negative) was applied to the trans compartment. A positive current is defined as the flow of cations from the trans to the cis compartment. Current flow was detected with two Ag/AgCl electrodes in KCl (3 M), agarose (1.5 %) bridges and amplified by an Axopatch 200B patch-clamp amplifier (Axon Instruments, Foster City, CA) connected to a CV 203BU head stage (Axon Instruments). Data were filtered through a 2 kHz low-pass Bessel filter and digitised with a Digidata 1322A converter (Axon Instruments) at a sampling frequency of 10 kHz, unless otherwise stated. Data analysis was performed with pClamp 10.3 software (Molecular Devices). All measurements were recorded at room temperature (20.5±0.5 °C).

A single OmpF trimer was allowed to insert into the bilayer after the addition of stock OmpF (<1 µL, 13–33 µM in Tris-HCl (20 mM, pH 8.0) with LiCl (80 mM), EDTA (5 mM), and βOG (1 %, w/v)) to the cis compartment, and ±(200–300) mV was applied to induce protein insertion. Cysteine mutants were incubated with TCEP (10 mM) before addition to the cis compartment. Once insertion was detected, the buffer in the cis compartment was exchanged to prevent further insertion by replacing 20% of the chamber volume 10–15 times with stirring. An I-V curve (−100 mV to +100 mV) was recorded for each OmpF trimer to establish positive (P-OmpF) or negative (N-OmpF) asymmetry. The reported n values represent independent experiments with fresh bilayers (not multiple insertions of single porins in one experiment). Reaction with mPEG-MAL or mPEG-OPSS (Nanocs, Boston USA) was performed by adding reagent (5 mM in recording buffer) to either the cis or trans compartment (1 mM final). Reagent solutions were freshly prepared before use.

6.3 Ch 3: In vitro expression of OmpF and heterotrimers engineering

The pSAI vector was made by inserting the mature OmpF gene (no signal sequence) from the pNGH72 plasmid in the dihydrofolate reductase (DHFR) control vector provided with the PURExpress® kit (NEB). The OmpF gene and the control template backbone (without the DHFR gene) were amplified by PCR with primers containing overhangs that created 40 bp of homology between the OmpF fragment and the backbone fragment (Table 6.1). The two PCR plasmids were joined by homologous recombination [230], in which XL-10 Gold ultracompetent cells were transformed with a 3:1 molar ratio of OmpF insert to vector backbone. The in vitro expressed OmpF gene contains an additional N-terminal methionine which is normally removed during signal sequence cleavage in vivo (Fig. 6.2).

Primers to amplify the OmpF gene from pNGH72 (5'→3')	
Forward	AAGAAGGAGATATACATATGGCAGAAATCTATAACAAAG
Reverse	AATTCCCGGGATCCTCATTAGAACTGGTAAACGATAC
Primers to amplify the control vector backbone (5'→3')	
Forward	GTATCGTTACCAGTTCTAATGAGGATCCCGGG
Reverse	GTTATAGATTTCTGCCATATGTATATCTCCTTCTTAAAGTTAAAC

Table 6.1 Primers used to insert the OmpF gene into the DHFR control vector (NEB) to create pSAI, used for in vitro expression of OmpF.

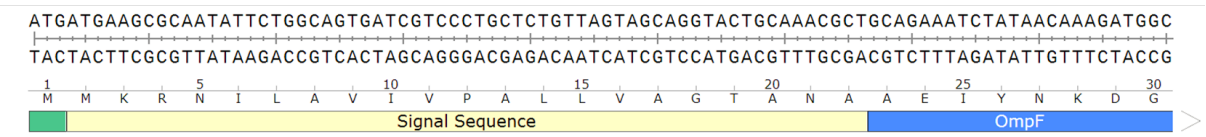


Figure 6.2 OmpF N-terminal sequence. The signal sequence is normally cleaved in vivo, yielding the mature OmpF protein beginning with an alanine. For in vitro expression, the signal sequence was removed leaving an additional N-terminal methionine, which is necessary for translation.

To prepare linear DNA template for IVTT, the OmpF gene was amplified by PCR starting with the residue A1, which was later mutated to methionine (A1M) by PCR to keep the number of amino acids in the polypeptide the same as in vivo, using the suggested primers in the PURE system protocol:

Forward Primer (49 mer): 5´ GCG AAT TAA TAC GAC TCA CTA TAG GGC TTA AGT ATA AGG AGG AAA AAA T . . . 24 bases matching gene of interest - 3´

Reverse Primer (35 mer): 5´ AAA CCC CTC CGT TTA GAG AGG GGT TAT GCT AG TTA . . . 27 bases matching gene of interest starting with ATG - 3´

N- and C-terminal H8-TEV and D8-TEV tags

The N- and C-terminus of the OmpF gene was extended by the addition of a TEV protease site (ENLYFQ-A), a linker (YDIPTT), and a charged tag (D8 or H8). This was achieved by two rounds of PCR mutagenesis, first to insert the TEV-linker and then to insert the charged tags, using the primers in Table 6.2.

TEV-linker N-term	
Forward	attccgaccaccgagaatctttattttcagGCAGAAATCTATAACAAAGATG
Reverse	atcataatcatcgtcgtcatcatcgtcatcCATATGTATATCTCCTTCTTAAAG
D8 N-term	
Forward	gacgacgatgatGATTATGATATTCCGACCAC
Reverse	atcatcgtcatcCATATGTATATCTCCTTCTTAAAG
H8 N-term	
Forward	caccatcatcacGATTATGATATTCCGACCAC
Reverse	gtggtgatgatgCATATGTATATCTCCTTCTTAAAG
TEV-linker C-term	
Forward	ctgccgaccaccgagaatctttattttcagGCAGAAATCTATAACAAAGATG
Reverse	aatataatcatcgtcgtcatcatcgtcatcCATATGTATATCTCCTTCTTAAAG
D8 C-term	
Forward	gattatgatattccgaccaccGATGACGATGATGACGAC
Reverse	gccctggaagtacaggttctcGAACTGGTAAACGATACC
H8 C-term	
Forward	caccatcatcacTGAGGATCCCGGGAATTC
Reverse	gtggtgatgatgGGTGGTCGGAATATCATAATC

Table 6.2 Primers used to make OmpF plasmids containing a TEV protease site and a charged tag, either D8 or H8, connected by a linker for in vitro expression.

6.3.1 Liposome preparation

Liposomes of different sizes were made from DMPC (200 nm) and DPPC (100 nm) lipids by extrusion. Powdered lipid was dissolved in chloroform and dried by evaporation using a nitrogen stream and then placed under vacuum for 3 hours. The lipid cake was then hydrated for >1 hour in 100 mM KCl, 20 mM phosphate buffer pH 7.0 at 30°C, which is above the transition temperature of both lipids. The hydrated lipid was then subjected to five freeze-thaw cycles and then passed through a polycarbonate membrane of appropriate size using a Mini-Extruder (Avanti Polar Lipids). The liposome suspension was stored above the transition temperature for up to two weeks.

6.3.2 IVTT

In vitro expression was carried out using the PURExpress® kit (NEB) according to the standard protocol with the addition of 1 µL RNase inhibitor (murine), 300 ng DNA (0.5-1.5 µL), and 1 µL ^{35}S-methionine (1,200 Ci mmol^{-1}, 15 mCi mL^{-1}, MP Biomedicals), and water up to 25 µL. To produce trimers, the reaction was supplemented with 3 µL DMPC liposomes (12.5 mg/mL) and 1.5 µL DDM (100 mM). The IVTT reaction was allowed to proceed for either 5 hours at 28°C or overnight (~16 hours) at 25°C. The reaction was quenched on ice with the addition of 4X Laemmli buffer (Bio-Rad) and then run on a hand-cast preparative 7.5% or 10% SDS-polyacrylamide gel (16.5 x 28 mm) for 14–16 hours at 90–120 V in a cold room (4°C). The gel was then transferred to Whatman® 3MM filter paper and dried under vacuum for 1 hour at 60°C or for 5 h at room temperature if the protein was to be extracted. For visualisation, the gel was exposed to Kodak® BioMax® MR film overnight (~16 hours) before the autoradiograph was developed using an X-ray film processor (Xograph Compact X4).

6.3.3 Gel extraction

To extract OmpF protein from gels, the polyacrylamide gel was first dried without fixation on Whatman 3MM filter paper under vacuum for 5 hours at room temperature. The dried gel was exposed to film overnight and then developed; the film was aligned with the gel to locate the desired protein band, which was excised with a scalpel. Each excised band was rehydrated in ~300 µL of 100 mM KCl, 20 mM phosphate buffer pH 7.0 containing 0.1% DDM for 10 min at RT. The filter paper was then removed, and the rehydrated gel was macerated with a plastic pestle and left for 1–2 hours at RT to allow the protein to diffuse out. The resulting slurry was filtered through a 0.2 µm PVDF membrane filter (Proteus Mini Clarification Spin Column, Generon) at 13,000 *x g* for 30 min. The filtrate was stored in 10 µL aliquots at -80°C or directly used in single-channel PLB recording.

6.3.4 Trimerisation of gel extracted monomers

Gel-extracted monomers were immediately concentrated using an Amicon 0.5 mL centrifugal filter with a 10 kDa cut-off. The concentrate was diluted by an equal volume of folding buffer containing DMPC liposomes and DDM in a 1:1 molar ratio (14.8 mM of each, 7.4 mM final) and incubated overnight at 30ºC. The resultant species were then resolved by SDS-PAGE at 4°C as described in Section 6.3.2.

6.4 Ch 4: Genetically engineered nanopore libraries for stochastic protein sensing

Electrocompetent MC1061 cells were from Lucigen and electroporation was carried out according to the suggested manufacturer protocol using an Eppendorf Electroporator 2510. Anti-Myc tag antibody clone 4A6 with or without an Alexa Fluor 488 label from Sigma was used for the fluorescent labelling and

electrophysiology studies, respectively. The kinase DYRK1a was a generous gift from Dr. Fiona Sorrell (Structural Genomics Consortium, University of Oxford). All plasmids were mutated by PCR mutagenesis with overhang primers. The PCR products were then ligated using KLD Enzyme Mix (NEB) before transformation according to standard protocols. To express OmpF with the Myc tag inserted into each extracellular loop tested, pSAI (Section 6.3) was mutated at the locations in Table 6.3 to give six plasmids.

Loop	Residues deleted	Primers (5' → 3')
L1	^{27}NGENSY^{32}	F: agcgaagaagatctgGGTGGCAATGGCGACATG R: aatcagtttctgttcACCCTTGGAGAAATAATGCAG
L4	^{165}TA^{166}	F: agcgaagaagatctgCGCCGCTCTAACGGCGAC R: aatcagtttctgttcGTCACGCTCGTTTTTACCCAGG
L5	^{199}LQEAQPLGN^{207}	F: agcgaagaagatctgGGTAAAAAAGCTGAACAGTG R: aatcagtttctgttcGTTGGTACGGTCAGCTGC
L6	^{242}NKFTNT^{247}	F: agcgaagaagatctgAGCGGCTTCGCCAACAAA R: aatcagtttctgttcAGTGATCGGCGTAGCGTTAC
L7	^{285}GIGD^{288}	F: agcgaagaagatctgGTTGATCTGGTGAACTAC R: aatcagtttctgttcTTCTACGTCTTTCGCTTTAG
L8	^{321}DNKLGVG^{327}	F: agcgaagaagatctgTCAGACGACACCGTTGCT R: aatcagtttctgttcAGAATCGATCTGGTTGATGATG

Table 6.3 Primers used to make OmpF mutants with the Myc epitope inserted into extracellular loop 1 and 4-8, replacing the residues shown.

To express OmpF in *E. coli* with the Myc tag inserted into L5–L7, pNGH72 was mutated at the same sites used for IVTT, but with an added six-residue SG linker flanking each side (SGSGSG – EQKLISEEDL – GSGSGS. Each of the primers also include a restriction site within the SG epitope or in silent mutations introduced into flanking OmpF residues (underlined) for library construction (L5: *EcoRI/SacI*, L6: *SpeI/SacI*, L7: *SacI/SalI*) (Table 6.4). The plasmids were purified by a MaxiPrep kit to increase purity for ligation.

Loop	Primers (5' → 3')
L5 (*EcoRI /SacI*)	F: gattagcgaagaagatctgggg<u>agctc</u>aggtagc GGTAAAAAAGCTGAACAGTGGG R: agtttctgttcaccagagctgcctgaa<u>gaattc</u>gt ACGGTCAGCTGCACCATA
L6 (*SpeI/ SacI*)	F: attagcgaagaagatctgggg<u>agctc</u>aggt AGCGGCTTCGCCAACAAA R: cagtttctgttcaccagagctgccgct<u>act</u> <u>AGT</u>GATCGGCGTAGCGTTAC
L7 (*SacI/ SalI*)	F: gaagaagatctgggcagcagcggcagcggc<u>gtcgac</u> CTGGTGAACTACTTTGAAG R: gctaatcagtttctgttcaccagagctgcct<u>gagctc</u> TCTACGTCTTTCGCTTTAG F2: gcgaagaagatctgagcggc<u>gtcgac</u> CTGGTGAACTACTTTGAAG R2: taatcagtttctgttcgcct<u>gagctc</u> TCTACGTCTTTCGCTTTAG

Table 6.4 Primers used to insert six SG residues on either side of the Myc epitope in L5–L7 and to introduce restriction sites (underlined) on either side of downstream ligation.

6.4.1 Growth media

LB medium was made by dissolving LB Broth (Miller) in water. RM medium (1 L) was made by autoclaving 20 g OmniPur® Casamino Acids (Millipore) in 790 mL of water. To this, 200 mL of autoclaved 5X M9 salts were added. After the solutions cooled, the following were added through a 0.22 µm PES filter (Millipore): 2.0 g glucose, 0.1 g $MgCl_2$, 0.05 g carbenicillin dissolved in water to 10 mL. The medium was stored at 4°C and replaced every month. RM medium used to resuspend samples for FACS was filtered through a 0.22 µm PES filter. RM-agar plates were made as above with the addition of agar powder. The agar was separately autoclaved in water before RM medium was added.

6.4.2 Fluorescent cell labelling

Plasmids containing OmpF with the Myc epitope inserted into L5–L7 were separately electroporated into MC1061 electrocompetent cells (Lucigen) according to standard protocol and plated on RM-agar. A single colony was used to inoculate 5 mL of RM medium, which was grown overnight and then diluted to OD = 0.1 in 5 mL RM medium . Bacteria were cultured until an OD of 0.4–0.6 before induction with 0.4% arabinose for two hours. Following this, a volume corresponding to 400 µL of OD = 0.3 cells was pelleted in a microcentrifuge tube and resuspended in 100 nM anti-Myc tag antibody in RM medium supplemented with 0.1% BSA and incubated on a rotary mixer for 45 min at 4°C. The tubes were then pelleted for 5 minutes at 5,000 *x g* and the pellet was resuspended in PBS supplemented with 0.1% BSA.

For imaging, 3–5 µL of the resuspended cells were pipetted onto a glass slide and covered with a square of agar (~ 3 mm thick), creating a single layer of cells trapped within the agar network. A confocal microscope (SP5, Leica) was used with a HC PL FLUOTAR 10X objective lens (0.3 N.A.). In all experiments the cells were first photographed in the phase contrast mode. Then a fluorescence image was taken using a blue excitation/green emission filter cube (U-MNB or EGFP, ex. 470–490 nm).

6.4.3 Fluorescence activated cell sorting

MC1061 cells were electroporated with the positive control plasmids and plated on RM-agar after 1 hour recovery in Recovery Medium (Lucigen). A single colony was used to inoculate an overnight culture which was diluted to OD 0.1 the following morning. At OD = 0.4–0.5, expression was induced with 0.4% arabinose for 1 hr. Following this, the OD was measured and a volume corresponding to 1

mL of OD = 0.1 was transferred to an Eppendorf tube and pelleted for 5 min at 5,000 *x g*. The pellet was resuspended in the desired antibody concentration in RM medium and incubated at 4°C for 45 min on a rotary mixer. Following this, cells were either pelleted again and resuspended in fresh RM medium before FACS (non-equilibrium mode) or taken directly to FACS (equilibrium mode). All cell suspensions used for FACS were diluted to 1×10^7 cells/mL. FACS was performed on an S3e™ Cell Sorter (Bio-Rad) equipped with a 100 µm nozzle. Samples were kept on ice throughout and during FACS the sample was kept at 4°C on the maximum agitation setting.

For each sample, 100,000 events were measured. Side scatter (SSC) was used to gate for single cells in order to avoid aggregates by plotting SSC-height to SSC-area. A histogram of the FL-1 channel (488 nm excitation, 530/30 nm emission) was used to determine binding of anti-Myc tag AF488 antibody. If a sort was performed, a sorting gate was set in the FL-1 histogram that started at the end of the negative control population and extended to the maximum fluorescence. Cells that fell in this gate were sorted in purity mode into 5 mL, 12 x 75 mm round-bottom polypropylene tubes (Falcon) filled with FBS and incubated for at least 30 minutes at RT. The FBS was removed and replaced with 1 mL SOC (NEB) before sorting. Cells were sorted into the SOC medium at 5,000 events per second unless otherwise stated and immediately plated on RM-agar. Recovery was calculated as $CFU_{observed}$ / $CFU_{expected}$.

6.4.4 Live/dead staining

Two library aliquots were thawed on ice and cultured as in Section 6.4.2. One sample was left uninduced, and the other was induced with 0.4% (w/v) arabinose for 1 hour. From the uninduced sample, a volume corresponding to 1×10^7 cells

was pelleted for 3 min 10,000 *x g* then resuspended in 1 mL RM medium (live control) or in 1 mL 70% isopropanol in water (dead control) and incubated in a ThermoMixer® C (Eppendorf) for 30 min at RT with shaking at 800 RPM along with a sample from the induced sample (library control) resuspended in RM medium. Following this, the three samples were pelleted and resuspended in RM medium with propidium iodide (3 µM) for 30 min at RT. FACS measurements and sorts were performed as before, however fluorescence was measured in the FL-2 channel (488 nm excitation, 585/42 nm emission).

6.4.5 Library synthesis

Randomised duplexed 10-mer DNA oligonucleotides flanked by six SG residues and the same restriction enzyme sites as the plasmids were ordered from Integrated DNA technologies (Table 6.5). Libraries were made in L5 and L6 by double digesting 6 µg of both the insert and pNGH72 vector with the appropriate set of restriction enzymes (60 U/enzyme) in a total 100 µL reaction volume for 16 hours at 37°C. Afterwards all reactions were inactivated for 20 min at 65°C. The insert was purified using a QIAquick Nucleotide Removal Kit (Qiagen) and eluted in 30 µL water. The final yield was ~75%.

Loop	Restriction enzymes	Insert sequence
L5	*EcoRI* / *SacI*	5'-GCACGCGGAATTCTTCAGGCAGCTCTGG TNNKNNKNNKNNKNNKNNKNNKNNKNNKNNK GGGAGCTCTACTAGACT-3'
L6	*SpeI / SacI*	5'-GCACGCGACTAGTAGCGGCAGCTCTGGT NNKNNKNNKNNKNNKNNKNNKNNKNNKNNK GGGAGCTCTACTAGACT-3'

Table 6.5 Sequence of the oligo pool used as to generate the randomised insert in L5 and L6. The sequence used to introduce a set of restriction enzymes for ligation is underlined in each insert.

For vector preparation, the digest was purified through a Zymo DNA Clean & Concentrator spin column and eluted in 30 µL water. The vector was mixed with 6X loading dye without SDS (NEB) and run on a 0.7% agarose gel made with SYBR Safe DNA Gel Stain (Invitrogen) for DNA visualisation. The cut plasmid band was extracted from the gel on a blue screen to prevent UV damage and purified using a Zymo gel purification kit and eluted in 40 µL water. The plasmid was then treated with *StuI* (7.5 U) and Antarctic Phosphotase (40 U) in CutSmart® buffer in a 40 µL reaction volume for 1 hr at 37°C, followed by inactivation for 20 min at 65°C. The reaction was purified using a Zymo DNA Clean & Concentrator kit and eluted in 20 µL water. The final yield was ~41%.

Test ligations were done to determine the best insert to vector ratio and establish library background (ligation with vector only). Vector to insert mole ratios of 0:1 (background), 3:1, 5:1, and 10:1 were ligated using 100 ng of vector with T4 DNA ligase (200 U) for 16 hours at 16°C. After this, reactions were inactivated for 10 min at 65°C and purified through Zymo DNA Clean & Concentrator spin columns. 30 ng of the ligation reaction was electroporated into 25 µL MC1061 cells and recovered for 1 hour at 37°C. An aliquot of the cells was diluted 100 X in RM medium, and 100 uL was plated on RM-agar. Colonies were counted the next morning and a 10:1 insert to vector ratio yielded the highest number of colonies. No colonies grew on the 0:1 plate, meaning that library background is <1%.

The ligation was then scaled up for the final library preparation. 2.4 µg of the library vector was ligated with 266 ng of insert (10:1 mole ratio) using T4 DNA ligase (200 U) in a 200 µL reaction volume, ligated as before, and purified using a Zymo DNA Clean & Concentrator spin column. Three aliquots of 25 µL MC1061 cells were thawed on ice and each electroporated with 250 ng of DNA. All three

aliquots (750 ng DNA total) were combined with 4.925 mL Recovery Medium (Lucigen) for a total 5 mL volume and recovered for 1 hour at 37°C, 250 RPM. A 10 µL aliquot was diluted and plated in triplicate to estimate library diversity from the CFU as outlined in Section 4.5.1. The final library size was 1.8×10^8 and 8.7×10^7 for L5 and L6, respectively. The remainder was used to inoculate 45 mL of RM medium in a 250 mL Erlenmeyer flask and grown for 16 hours at 18°C, 250 RPM. Aliquots containing 10 X the library diversity were mixed with 80% sterile glycerol (MP Biomedicals) to a final glycerol concentration of 16% (v/v) and flash frozen in liquid nitrogen. Library aliquots were stored at −80°C.

6.4.6 OmpF detergent extraction from E. coli

Parallel nanopore extraction from bacterial cells was carried out in 96-well V-bottom polypropylene plates, 2 mL capacity (Corning). All volume transfers were carried out using 12-channel pipettes. Each colony was used to inoculate 1 mL of RM medium in an individual well. The plate was then incubated at 37°C, 800 RPM in a ThermoMixer® C heating block until an OD=0.4–0.5 (usually 4–5 hours). Wells were then induced with 0.4% (w/v) arabinose and grown for 2 more hours. The plate was then centrifuged at 3,273 *x g* for 10 minutes at 4°C and then flipped upside down and tapped several times to remove the supernatant. The remaining pellets were resuspended in 200 µL Extraction Buffer: 2% βOG, 5 mM EDTA, 25 mM Tris-HCl pH 8.0 and incubated at 37°C, 800 RPM for 15 min. After this, the plate was centrifuged at 3,273 *x g* for at least 30 min. A 30 µL aliquot of the supernatant was carefully removed from each well for direct use in single-channel PLB current recordings or otherwise stored at 4°C. A second 30 µL aliquot was mixed with 10 µL 4X Laemmli loading dye and run on 7.5% or 10%

Criterion™ TGX™ Precast Gels (Bio-Rad) to determine expression levels and extraction purity.

References

1. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**, 1146–1153 (2008).
2. Bayley, H. & Martin, C. R. Resistive-Pulse SensingFrom Microbes to Molecules. *Chem. Rev.* **100**, 2575–2594 (2000).
3. DeBlois, R. W., Bean, C. P. & Wesley, R. K. . Electrokinetic measurements with submicron particles and pores by the resistive pulse technique. *J. Colloid Interface Sci.* **61**, 323–335 (1977).
4. Fahie, M., Chisholm, C. & Chen, M. Resolved single-molecule detection of individual species within a mixture of anti-biotin antibodies using an engineered monomeric nanopore. *ACS Nano* **9**, 1089–98 (2015).
5. Harrington, L., Cheley, S., Alexander, L. T., Knapp, S. & Bayley, H. Stochastic detection of Pim protein kinases reveals electrostatically enhanced association of a peptide substrate. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E4417-26 (2013).
6. Fahie, M. A., Yang, B., Mullis, M., Holden, M. A. & Chen, M. Selective Detection of Protein Homologues in Serum Using an OmpG Nanopore. *Anal. Chem.* **87**, 11143–11149 (2015).
7. Thakur, A. K. & Movileanu, L. Real-time measurement of protein–protein interactions at single-molecule resolution using a biological nanopore. *Nat. Biotechnol.* **37**, 96–101 (2019).
8. Lee, J. & Bayley, H. Semisynthetic protein nanoreactor for single-molecule chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13768–73 (2015).
9. Qing, Y., Pulcu, G. S., Bell, N. A. W. & Bayley, H. Bioorthogonal Cycloadditions with Sub-Millisecond Intermediates. *Angew. Chemie Int. Ed.* **57**, 1218–1221 (2018).
10. Bayley, H., Luchian, T., Shin, S.-H. & Steffensen, M. B. Single-Molecule Covalent Chemistry in a Protein Nanoreactor. in 251–277 (Springer, Berlin, Heidelberg, 2008).
11. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
12. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* **15**, 201–206 (2018).
13. Castro-Wallace, S. L. *et al.* Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci. Rep.* **7**, 18022 (2017).
14. Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* (2001).
15. HLADKY, S. B. & HAYDON, D. A. Discreteness of Conductance Change in Bimolecular Lipid Membranes in the Presence of Certain Antibiotics. *Nature* **225**, 451–453 (1970).
16. Zhang, S., Udho, E., Wu, Z., Collier, R. J. & Finkelstein, A. Protein Translocation through Anthrax Toxin Channels Formed in Planar Lipid Bilayers. *Biophys. J.* **87**, 3842–3849 (2004).

17. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 3561–6 (1972).
18. Baba, T. *et al.* Formation and characterization of planar lipid bilayer membranes from synthetic phytanyl-chained glycolipids. *Biochim. Biophys. Acta - Biomembr.* **1421**, 91–102 (1999).
19. Redwood, W. R., Pfeiffer, F. R., Weisbach, J. A. & Thompson, T. E. Physical properties of bilayer membranes formed from a synthetic saturated phospholipid in n-decane. *Biochim. Biophys. Acta - Biomembr.* **233**, 1–6 (1971).
20. Krylov, A. V, Pohl, P., Zeidel, M. L. & Hill, W. G. Water permeability of asymmetric planar lipid bilayers: leaflets of different composition offer independent and additive resistances to permeation. *J. Gen. Physiol.* **118**, 333–40 (2001).
21. Wiese, A. *et al.* Planar asymmetric lipid bilayers of glycosphingolipid or lipopolysaccharide on one side and phospholipids on the other: membrane potential, porin function, and complement activation. *Biophys. J.* **70**, 321–329 (1996).
22. Jo, A., Hoi, H., Zhou, H., Gupta, M. & Montemagno, C. D. Single-molecule study of full-length NaChBac by planar lipid bilayer recording. *PLoS One* **12**, e0188861 (2017).
23. Jing, P., Paraiso, H. & Burris, B. Highly efficient integration of the viral portal proteins from different types of phages into planar bilayers for the black lipid membrane analysis. *Mol. Biosyst.* **12**, 480–489 (2016).
24. Heitz, B. A. *et al.* Polymerized planar suspended lipid bilayers for single ion channel recordings: comparison of several dienoyl lipids. *Langmuir* **27**, 1882–90 (2011).
25. Malmstadt, N., Jeon, T.-J. & Schmidt, J. J. Long-Lived Planar Lipid Bilayer Membranes Anchored to an In Situ Polymerized Hydrogel. *Adv. Mater.* **20**, 84–89 (2008).
26. Meier, W., Nardin, C. & Winterhalter, M. Reconstitution of Channel Proteins in (Polymerized) ABA Triblock Copolymer Membranes. *Angew. Chemie Int. Ed.* **39**, 4599–4602 (2000).
27. Bayley, H. *et al.* Droplet interface bilayers. *Mol. Biosyst.* **4**, 1191 (2008).
28. Gross, L. C. M., Heron, A. J., Baca, S. C. & Wallace, M. I. Determining Membrane Capacitance by Dynamic Control of Droplet Interface Bilayer Area. *Langmuir* **27**, 14335–14342 (2011).
29. Syeda, R., Holden, M. A., Hwang, W. L. & Bayley, H. Screening Blockers Against a Potassium Channel with a Droplet Interface Bilayer Array. *J. Am. Chem. Soc.* **130**, 15543–15548 (2008).
30. Heron, A. J., Thompson, J. R., Mason, A. E. & Wallace, M. I. Direct detection of membrane channels from gels using water-in-oil droplet bilayers. *J. Am. Chem. Soc.* **129**, 16042–7 (2007).
31. Huang, S., Romero-Ruiz, M., Castell, O. K., Bayley, H. & Wallace, M. I. High-throughput optical sensing of nucleic acids in a nanopore array. *Nat. Nanotechnol.* **10**, 986–991 (2015).
32. Sapra, K. T. & Bayley, H. Lipid-coated hydrogel shapes as components of electrical circuits and mechanical devices. *Sci. Rep.* **2**, 848 (2012).
33. Villar, G., Graham, A. D. & Bayley, H. A Tissue-Like Printed Material. *Science (80-.).* **340**, 48–52 (2013).

34. Graham, A. D. *et al.* High-Resolution Patterned Cellular Constructs by Droplet-Based 3D Printing. *Sci. Rep.* **7**, 7004 (2017).
35. Mahendran, K. R. *et al.* Molecular basis of enrofloxacin translocation through OmpF, an outer membrane channel of escherichia coli - When binding does not imply translocation. *J. Phys. Chem. B* (2010).
36. Stefan Howorka, *et al.* A Protein Pore with a Single Polymer Chain Tethered within the Lumen. (2000).
37. Varongchayakul, N., Song, J., Meller, A. & Grinstaff, M. W. Single-molecule protein sensing in a nanopore: a tutorial. *Chem. Soc. Rev.* **47**, 8512–8524 (2018).
38. Shi, W., Friedman, A. K. & Baker, L. A. Nanopore Sensing. *Anal. Chem.* **89**, 157 (2017).
39. Howorka, S. & Siwy, Z. Nanopore analytics: sensing of single molecules. *Chem. Soc. Rev.* **38**, 2360 (2009).
40. Spruijt, E., Tusk, S. E. & Bayley, H. DNA scaffolds support stable and uniform peptide nanopores. *Nat. Nanotechnol.* **13**, 739–745 (2018).
41. Hall, A. R. *et al.* Hybrid pore formation by directed insertion of α-haemolysin into solid-state nanopores. *Nat. Nanotechnol.* **5**, 874–877 (2010).
42. Xie, H., Braha, O., Gu, L.-Q., Cheley, S. & Bayley, H. Single-Molecule Observation of the Catalytic Subunit of cAMP-Dependent Protein Kinase Binding to an Inhibitor Peptide. *Chem. Biol.* **12**, 109–120 (2005).
43. Song, L. *et al.* Structure of Staphylococcal alpha -Hemolysin, a Heptameric Transmembrane Pore. *Science (80-.).* **274**, 1859–1865 (1996).
44. Cheley, S., Braha, O., Lu, X., Conlan, S. & Bayley, H. A functional protein pore with a “retro” transmembrane domain. *Protein Sci.* **8**, 1257–1267 (1999).
45. Stoddart, D. *et al.* Functional truncated membrane pores. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2425–30 (2014).
46. Schulz, G. E. The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1565**, 308–317 (2002).
47. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
48. Hammerstein, A. F., Jayasinghe, L. & Bayley, H. Subunit dimers of alpha-hemolysin expand the engineering toolbox for protein nanopores. *J. Biol. Chem.* **286**, 14324–34 (2011).
49. Bhakdi, S. & Tranum-Jensen, J. Alpha-toxin of Staphylococcus aureus. *Microbiol. Rev.* **55**, 733–51 (1991).
50. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
51. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta* **1794**, 808–16 (2009).
52. Yoshida, T., Qin, L., Egger, L. A. & Inouye, M. Transcription regulation of ompF and ompC by a single transcription factor, OmpR. *J. Biol. Chem.* **281**, 17114–23 (2006).
53. Chen, M., Khalid, S., Sansom, M. S. P. & Bayley, H. Outer membrane protein G: Engineering a quiet pore for biosensing. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 6272–7 (2008).
54. Fahie, M., Chisholm, C. & Chen, M. Resolved single-molecule detection of individual species within a mixture of anti-biotin antibodies using an engineered monomeric nanopore. *ACS Nano* **9**, 1089–98 (2015).

55. Fahie, M. A., Yang, B., Pham, B. & Chen, M. Tuning the selectivity and sensitivity of an OmpG nanopore sensor by adjusting ligand tether length. *ACS sensors* **1**, 614–622 (2016).
56. Sean Conlan, ‡,§, Yong Zhang, ‡,§, Stephen Cheley, ‡ and & Hagan Bayley*, ‡,∥. Biochemical and Biophysical Characterization of OmpG: A Monomeric Porin†. (2000).
57. Xie, H., Braha, O., Gu, L. Q., Cheley, S. & Bayley, H. Single-molecule observation of the catalytic subunit of cAMP-dependent protein kinase binding to an inhibitor peptide. *Chem. Biol.* (2005).
58. Movileanu, L., Howorka, S., Braha, O. & Bayley, H. Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nat. Biotechnol.* **18**, 1091–5 (2000).
59. Rotem, D., Jayasinghe, L., Salichou, M. & Bayley, H. Protein Detection by Nanopores Equipped with Aptamers. *J. Am. Chem. Soc.* **134**, 2781–2787 (2012).
60. Cowan, S. W. *et al.* Crystal structures explain functional properties of two E. coli porins. *Nature* **358**, 727–33 (1992).
61. Prashant S. Phale, ‡ *et al.* Stability of Trimeric OmpF Porin: The Contributions of the Latching Loop L2†. (1998).
62. Tieleman, D. P. & Berendsen, H. J. A molecular dynamics study of the pores formed by Escherichia coli OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* **74**, 2786–801 (1998).
63. Ziervogel, B. K. & Roux, B. The binding of antibiotics in OmpF porin. *Structure* (2013).
64. Danelon, C., Suenaga, A., Winterhalter, M. & Yamato, I. Molecular origin of the cation selectivity in OmpF porin: single channel conductances vs. free energy calculation. *Biophys. Chem.* **104**, 591–603 (2003).
65. Nestorovich, E. M., Rostovtseva, T. K. & Bezrukov, S. M. Residue ionization and ion transport through OmpF channels. *Biophys. J.* **85**, 3718–29 (2003).
66. Kojima, S. & Nikaido, H. Permeation rates of penicillins indicate that Escherichia coli porins function principally as nonspecific channels. *Proc. Natl. Acad. Sci.* **110**, E2629–E2634 (2013).
67. Brauser, A. *et al.* Modulation of enrofloxacin binding in OmpF by Mg2+ as revealed by the analysis of fast flickering single-porin current. *The Journal of General Physiology* (2012).
68. Weichbrodt, C. *et al.* Antibiotic translocation through porins studied in planar lipid bilayers using parallel platforms.
69. Housden, N. G. *et al.* Directed epitope delivery across the Escherichia coli outer membrane through the porin OmpF. *Proc. Natl. Acad. Sci. U. S. A.* (2010).
70. Ionescu, S. A. *et al.* Orientation of the OmpF Porin in Planar Lipid Bilayers. *ChemBioChem* **18**, 554–562 (2017).
71. Frillingos, S., Sahin-tóth, M., Wu, J. & Kaback, H. R. Cys-scanning mutagenesis: a novel approach to structure–function relationships in polytopic membrane proteins. *FASEB J.* **12**, 1281–1299 (1998).
72. Raicu, V. & Singh, D. R. FRET Spectrometry: A New Tool for the Determination of Protein Quaternary Structure in Living Cells. *Biophys. J.* **105**, 1937–1945 (2013).
73. Hubbell, W. L., Altenbach, C., Hubbell, C. M. & Khorana, H. G. Rhodopsin

structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Adv. Protein Chem.* **63**, 243–290 (2003).
74. Malmberg, N. J. & Falke, J. J. Use of EPR Power Saturation to Analyze the Membrane-Docking Geometries of Peripheral Proteins: Applications to C2 Domains. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 71–90 (2005).
75. Belrhali, H. *et al.* Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* **7**, 909–917 (1999).
76. Popot, J.-L., Engelman, D. M., Gurel, O. & Zaccaï, G. Tertiary structure of bacteriorhodopsin: Positions and orientations of helices A and B in the structural map determined by neutron diffraction. *J. Mol. Biol.* **210**, 829–847 (1989).
77. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623–5 (2006).
78. Andersen, C., Schiffler, B., Charbit, A. & Benz, R. PH-induced collapse of the extracellular loops closes Escherichia coli maltoporin and allows the study of asymmetric sugar binding. *J. Biol. Chem.* **277**, 41318–25 (2002).
79. Kullman, L., Winterhalter, M. & Bezrukov, S. M. Transport of maltodextrins through maltoporin: a single-channel study. *Biophys. J.* **82**, 803–12 (2002).
80. Andersen, C. *et al.* In vivo and in vitro studies of major surface loop deletion mutants of the Escherichia coli K-12 maltoporin: contribution to maltose and maltooligosaccharide transport and binding. *Mol. Microbiol.* **32**, 851–67 (1999).
81. Alcaraz, A., Nestorovich, E. M., Aguilella-Arzo, M., Aguilella, V. M. & Bezrukov, S. M. Salting out the ionic selectivity of a wide channel: the asymmetry of OmpF. *Biophys. J.* **87**, 943–57 (2004).
82. Udho, E. *et al.* Reconstitution of bacterial outer membrane TonB-dependent transporters in planar lipid bilayer membranes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21990–5 (2009).
83. Benz, R., Janko, K., Boos, W. & Läuger, P. Formation of large, ion-permeable membrane channels by the matrix protein (porin) of Escherichia coli. *Biochim. Biophys. Acta* **511**, 305–19 (1978).
84. Kishii, R. & Takei, M. Relationship between the expression of ompF and quinolone resistance in Escherichia coli. *J. Infect. Chemother.* **15**, 361–6 (2009).
85. Cohen, S. P., McMurry, L. M., Hooper, D. C., Wolfson, J. S. & Levy, S. B. Cross-resistance to fluoroquinolones in multiple-antibiotic-resistant (Mar) Escherichia coli selected by tetracycline or chloramphenicol: decreased drug accumulation associated with membrane changes in addition to OmpF reduction. *Antimicrob. Agents Chemother.* **33**, 1318–25 (1989).
86. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta* **1794**, 808–16 (2009).
87. Nestorovich, E. M., Danelon, C., Winterhalter, M. & Bezrukov, S. M. Designed to penetrate: time-resolved interaction of single antibiotic molecules with bacterial pores. *Proc. Natl. Acad. Sci. U. S. A.* (2002).
88. Mahendran, K. R., Kreir, M., Weingart, H., Fertig, N. & Winterhalter, M. Permeation of antibiotics through Escherichia coli OmpF and OmpC porins: screening for influx on a single-molecule level. *J. Biomol. Screen.* **15**, 302–7

(2010).
89. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev.* **71**, 158–229 (2007).
90. Basle, A., Qutub, R., Mehrazin, M., Wibbenmeyer, J. & Delcour, A. H. Deletions of single extracellular loops affect pH sensitivity, but not voltage dependence, of the Escherichia coli porin OmpF. *Protein Eng. Des. Sel.* **17**, 665–672 (2004).
91. Ghai, I. *et al.* Ampicillin permeation across OmpF, the major outer-membrane channel in *Escherichia coli*. *J. Biol. Chem.* **293**, 7030–7037 (2018).
92. Hoenger, A., Pagès, J. M., Fourel, D. & Engel, A. The orientation of porin OmpF in the outer membrane of Escherichia coli. *J. Mol. Biol.* **233**, 400–13 (1993).
93. Rostovtseva, T. K., Nestorovich, E. M. & Bezrukov, S. M. Partitioning of differently sized poly(ethylene glycol)s into OmpF porin. *Biophys. J.* **82**, 160–9 (2002).
94. Kreir, M., Farre, C., Beckler, M., George, M. & Fertig, N. Rapid screening of membrane protein activity: electrophysiological analysis of OmpF reconstituted in proteoliposomes. *Lab Chip* (2008).
95. Yamashita, E., Zhalnina, M. V, Zakharov, S. D., Sharma, O. & Cramer, W. A. Crystal structures of the OmpF porin: function in a colicin translocon. *EMBO J.* **27**, 2171–80 (2008).
96. Danelon, C., Brando, T. & Winterhalter, M. Probing the orientation of reconstituted maltoporin channels at the single-protein level. *J. Biol. Chem.* **278**, 35542–51 (2003).
97. Van Gelder, P., Dumas, F., Rosenbusch, J. P. & Winterhalter, M. Oriented channels reveal asymmetric energy barriers for sugar translocation through maltoporin of Escherichia coli. *Eur. J. Biochem.* **267**, 79–84 (2000).
98. Alcaraz, A., Nestorovich, E. M., Aguilella-Arzo, M., Aguilella, V. M. & Bezrukov, S. M. Salting out the ionic selectivity of a wide channel: the asymmetry of OmpF. *Biophys. J.* **87**, 943–57 (2004).
99. Baslé, A., Iyer, R. & Delcour, A. H. Subconductance states in OmpF gating. *Biochim. Biophys. Acta* **1664**, 100–7 (2004).
100. Miles, G., Movileanu, L. & Bayley, H. Subunit composition of a bicomponent toxin: staphylococcal leukocidin forms an octameric transmembrane pore. *Protein Sci.* **11**, 894–902 (2002).
101. Movileanu, L. Location of a Constriction in the Lumen of a Transmembrane Pore by Targeted Covalent Attachment of Polymer Molecules. *J. Gen. Physiol.* **117**, 239–252 (2001).
102. Im, W. & Roux, B. Ion permeation and selectivity of OmpF porin: a theoretical study based on molecular dynamics, Brownian dynamics, and continuum electrodiffusion theory. *J. Mol. Biol.* **322**, 851–69 (2002).
103. Nirenberg, M. W., Matthaei, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1588–602 (1961).
104. Hoagland, M.B., Stephenson, M. L., *et al.* A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* **231**, 241–57 (1958).
105. Matthaei, J. H. & Nirenberg, M. W. Characteristics and stbailization of DNase-sensitive protein synthesis in E. coli extracts. *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1580 (1961).

106. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23**, 150–156 (2005).
107. Miles, G., Movileanu, L. & Bayley, H. Subunit composition of a bicomponent toxin: staphylococcal leukocidin forms an octameric transmembrane pore. *Protein Sci.* **11**, 894–902 (2002).
108. Mikami, S., Kobayashi, T., Masutani, M., Yokoyama, S. & Imataka, H. A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins. *Protein Expr. Purif.* **62**, 190–198 (2008).
109. Ezure, T. *et al.* Cell-Free Protein Synthesis System Prepared from Insect Cells by Freeze-Thawing. *Biotechnol. Prog.* **22**, 1570–1577 (2006).
110. Jackson, R. J. & Hunt, T. Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods Enzymol.* **96**, 50–74 (1983).
111. Takai, K., Sawasaki, T. & Endo, Y. Practical cell-free protein synthesis system using purified wheat embryos. *Nat. Protoc.* **5**, 227–238 (2010).
112. Swartz, J. Developing cell-free biology for industrial applications. *J. Ind. Microbiol. Biotechnol.* **33**, 476–485 (2006).
113. Kigawa, T. *et al.* Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *J. Struct. Funct. Genomics* **5**, 63–68 (2004).
114. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
115. Gregorio, N. E., Levine, M. Z. & Oza, J. P. A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc.* **2**, (2019).
116. Schwarz, D. *et al.* Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat. Protoc.* **2**, 2945–2957 (2007).
117. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
118. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
119. Niwa, T. *et al.* Comprehensive study of liposome-assisted synthesis of membrane proteins using a reconstituted cell-free translation system. *Sci. Rep.* **5**, 18025 (2015).
120. Junge, F. *et al.* Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. *N. Biotechnol.* **28**, 262–271 (2011).
121. Kuruma, Y. & Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* **10**, 1328–44 (2015).
122. Lyukmanova, E. N. *et al.* Lipid–protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: Comparison with detergent micelles, bicelles and liposomes. *Biochim. Biophys. Acta - Biomembr.* **1818**, 349–358 (2012).
123. Matsubayashi, H., Kuruma, Y. & Ueda, T. In Vitro Synthesis of the *E. coli* Sec Translocon from DNA. *Angew. Chemie Int. Ed.* **53**, 7535–7538 (2014).
124. Sachse, R., Dondapati, S. K., Fenz, S. F., Schmidt, T. & Kubick, S. Membrane protein synthesis in cell-free systems: From bio-mimetic systems to bio-membranes. *FEBS Lett.* **588**, 2774–2781 (2014).
125. Klammt, C. *et al.* Evaluation of detergents for the soluble expression of α-

helical and β-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J.* **272**, 6024–6038 (2005).
126. McMorran, L. M., Brockwell, D. J. & Radford, S. E. Mechanistic studies of the biogenesis and folding of outer membrane proteins in vitro and in vivo: what have we learned to date? *Arch. Biochem. Biophys.* **564**, 265–80 (2014).
127. Wu, T. *et al.* Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in Escherichia coli. *Cell* **121**, 235–245 (2005).
128. Wülfing, C. & Plückthun, A. Protein folding in the periplasm of Escherichia coli. *Mol. Microbiol.* **12**, 685–692 (1994).
129. Iadanza, M. G. *et al.* Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. *Nat. Commun.* **7**, 12865 (2016).
130. Noinaj, N., Kuszak, A. J., Balusek, C., Gumbart, J. C. & Buchanan, S. K. Lateral opening and exit pore formation are required for BamA function. *Structure* **22**, 1055–62 (2014).
131. Sen, K. & Nikaido, H. Trimerization of an in vitro synthesized OmpF porin of Escherichia coli outer membrane. *J. Biol. Chem.* **266**, 11295–300 (1991).
132. Visudtiphole, V., Thomas, M. B., Chalton, D. A. & Lakey, J. H. Refolding of Escherichia coli outer membrane protein F in detergent creates LPS-free trimers and asymmetric dimers. *Biochem. J.* **392**, 375–81 (2005).
133. Eisele, J. L. & Rosenbusch, J. P. In vitro folding and oligomerization of a membrane protein. Transition of bacterial porin from random coil to native conformation. *J. Biol. Chem.* **265**, 10217–20 (1990).
134. Burgess, N. K., Dao, T. P., Stanley, A. M. & Fleming, K. G. Beta-barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro. *J. Biol. Chem.* **283**, 26748–58 (2008).
135. Surrey, T., Schmid, A. & Jähnig, F. Folding and Membrane Insertion of the Trimeric β-Barrel Protein OmpF. *Biochemistry* **35**, 2283–2288 (1996).
136. Watanabe, Y. & Inoko, Y. Reassembly of an Integral Oligomeric Membrane Protein OmpF Porin in n-octyl β-d-Glucopyranoside–Lipids Mixtures. *Protein J.* **28**, 66–73 (2009).
137. Hagan, C. L., Kim, S. & Kahne, D. Reconstitution of outer membrane protein assembly from purified components. *Science* **328**, 890–2 (2010).
138. Reid, J., Fung, H., Gehring, K., Klebba, P. E. & Nikaido, H. Targeting of porin to the outer membrane of Escherichia coli. Rate of trimer assembly and identification of a dimer intermediate. *J. Biol. Chem.* **263**, 7753–9 (1988).
139. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nat. Biotechnol.* **19**, 636–639 (2001).
140. Miles, G., Bayley, H. & Cheley, S. Properties of Bacillus cereus hemolysin II: a heptameric transmembrane pore. *Protein Sci.* **11**, 1813–24 (2002).
141. Ihle, S. *et al.* Nanocompartments with a pH release system based on an engineered OmpF channel protein. *Soft Matter* **7**, 532–539 (2011).
142. Einfalt, T. *et al.* Stimuli-Triggered Activity of Nanoreactors by Biomimetic Engineering Polymer Membranes. *Nano Lett.* **15**, 7596–7603 (2015).
143. An Ranquin, Wim Versées, †, Wolfgang Meier, ‡, Jan Steyaert, † and & Gelder†, P. Van. Therapeutic Nanoreactors: Combining Chemistry and Biology in a Novel Triblock Copolymer Drug Delivery System. (2005).

144. Tanner, P., Balasubramanian, V. & Palivan, C. G. Aiding Nature's Organelles: Artificial Peroxisomes Play Their Role. *Nano Lett.* **13**, 2875–2883 (2013).
145. Einfalt, T. *et al.* Biomimetic artificial organelles with in vitro and in vivo activity triggered by reduction in microenvironment. *Nat. Commun.* **9**, 1127 (2018).
146. Yuni Jung, †,‡, Hagan Bayley, § and & Liviu Movileanu*, ⫴. Temperature-Responsive Protein Pores. (2006).
147. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci.* **106**, 1760–1765 (2009).
148. Inouye, M. & Yee, M. L. Homogeneity of envelope proteins of Escherichia coli separated by gel electrophoresis in sodium dodecyl sulfate. *J. Bacteriol.* **113**, 304–12 (1973).
149. Ali, A. Engineering Outer membrane protein F heterotrimers with applications as biosensors. (University of Oxford, 2018).
150. Rosenbusch, J. P. Characterization of the major envelope protein from Escherichia coli. Regular arrangement on the peptidoglycan and unusual dodecyl sulfate binding. *J. Biol. Chem.* **249**, 8019–29 (1974).
151. Almgren, M. Mixed micelles and other structures in the solubilization of bilayer lipid membranes by surfactants. *Biochim. Biophys. Acta - Biomembr.* **1508**, 146–163 (2000).
152. Niwa, T., Kanamori, T., Ueda, T. & Taguchi, H. Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8937–42 (2012).
153. Goemans, C., Denoncin, K. & Collet, J.-F. Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843**, 1517–1528 (2014).
154. Harrington, L., Cheley, S., Alexander, L. T., Knapp, S. & Bayley, H. Stochastic detection of Pim protein kinases reveals electrostatically enhanced association of a peptide substrate. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E4417-26 (2013).
155. Qing, Y., Ionescu, S. A., Pulcu, G. S. & Bayley, H. Directional control of a processive molecular hopper. *Science* **361**, 908–912 (2018).
156. Ramsay, W. J., Bell, N. A. W., Qing, Y. & Bayley, H. Single-Molecule Observation of the Intermediates in a Catalytic Cycle. *J. Am. Chem. Soc.* **140**, 17538–17546 (2018).
157. Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci.* **104**, 19428–19433 (2007).
158. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **13**, 786–796 (2018).
159. Cohen, L. & Walt, D. R. Highly Sensitive and Multiplexed Protein Measurements. *Chem. Rev.* **119**, 293–321 (2019).
160. Mogalisetti, P., Gorris, H. H., Rojek, M. J. & Walt, D. R. Elucidating the relationship between substrate and inhibitor binding to the active sites of tetrameric β-galactosidase. *Chem. Sci.* **5**, 4467–4473 (2014).
161. Movileanu, L., Howorka, S., Braha, O. & Bayley, H. Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nat. Biotechnol.* **18**, 1091–1095 (2000).
162. Xie, H., Braha, O., Gu, L.-Q., Cheley, S. & Bayley, H. Single-Molecule

Observation of the Catalytic Subunit of cAMP-Dependent Protein Kinase Binding to an Inhibitor Peptide. *Chem. Biol.* **12**, 109–120 (2005).
163. Howorka, S., Nam, J., Bayley, H. & Kahne, D. Stochastic Detection of Monovalent and Bivalent Protein–Ligand Interactions. *Angew. Chemie Int. Ed.* **43**, 842–846 (2004).
164. Fahie, M. A. & Chen, M. Electrostatic Interactions between OmpG Nanopore and Analyte Protein Surface Can Distinguish between Glycosylated Isoforms. *J. Phys. Chem. B* **119**, 10198–206 (2015).
165. Soskine, M. *et al.* An engineered ClyA nanopore detects folded target proteins by selective external association and pore entry. *Nano Lett.* **12**, 4895–900 (2012).
166. Cheley, S., Xie, H. & Bayley, H. A Genetically Encoded Pore for the Stochastic Detection of a Protein Kinase. *ChemBioChem* **7**, 1923–1927 (2006).
167. Harrington, L., Alexander, L. T., Knapp, S. & Bayley, H. Single-Molecule Protein Phosphorylation and Dephosphorylation by Nanopore Enzymology. *ACS Nano* acsnano.8b07697 (2019).
168. Wang, S., Haque, F., Rychahou, P. G., Evers, B. M. & Guo, P. Engineered Nanopore of Phi29 DNA-Packaging Motor for Real-Time Detection of Single Colon Cancer Specific Antibody in Serum. *ACS Nano* **7**, 9814–9822 (2013).
169. Ho, C.-W. *et al.* Engineering a nanopore with co-chaperonin function. *Sci. Adv.* **1**, e1500905 (2015).
170. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. BioSyst.* **8**, 268–281 (2012).
171. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A Million Peptide Motifs for the Molecular Biologist. *Mol. Cell* **55**, 161–169 (2014).
172. Nobel Media AB 2019. The Nobel Prize in Chemistry 2018. Available at: https://www.nobelprize.org/prizes/chemistry/2018/summary/. (Accessed: 26th August 2019)
173. Clackson, T. & Wells, J. A. In vitro selection from protein and peptide libraries. *Trends Biotechnol.* **12**, 173–184 (1994).
174. Patrick, W. M. & Firth, A. E. Strategies and computational tools for improving randomized protein libraries. *Biomol. Eng.* **22**, 105–112 (2005).
175. Sieber, T., Hare, E., Hofmann, H. & Trepel, M. Biomathematical description of synthetic peptide libraries. *PLoS One* **10**, e0129200 (2015).
176. DeGraaf, M. E., Miceli, R. M., Mott, J. E. & David Fischer, H. Biochemical diversity in a phage display library of random decapeptides. *Gene* **128**, 13–17 (1993).
177. Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci.* **91**, 9022–9026 (1994).
178. Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (80-.).* **228**, 1315–1317 (1985).
179. Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10444–8 (1993).
180. Lee, S. Y., Choi, J. H. & Xu, Z. Microbial cell-surface display. *Trends Biotechnol.* **21**, 45–52 (2003).

181. Georgiou, G. *et al.* Display of heterologous proteins on the surface of microorganisms: From the screening of combinatorial libraries to live recombinant vaccines. *Nat. Biotechnol.* **15**, 29–34 (1997).
182. Dane, K. Y., Chan, L. A., Rice, J. J. & Daugherty, P. S. Isolation of cell specific peptide ligands using fluorescent bacterial display libraries. *J. Immunol. Methods* **309**, 120–129 (2006).
183. Hall, S. S., Mitragotri, S. & Daugherty, P. S. Identification of Peptide Ligands Facilitating Nanoparticle Attachment to Erythrocytes. *Biotechnol. Prog.* **23**, 749–754 (2008).
184. Bessette, P. H., Rice, J. J. & Daugherty, P. S. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng. Des. Sel.* **17**, 731–739 (2004).
185. Xu, Z. & Lee, S. Y. Display of polyhistidine peptides on the Escherichia coli cell surface by using outer membrane protein C as an anchoring motif. *Appl. Environ. Microbiol.* **65**, 5142–7 (1999).
186. Verhoeven, G. S., Alexeeva, S., Dogterom, M. & Blaauwen, T. den. Differential Bacterial Surface Display of Peptides by the Transmembrane Domain of OmpA. *PLoS One* **4**, e6739 (2009).
187. Rice, J. J., Schohn, A., Bessette, P. H., Boulware, K. T. & Daugherty, P. S. Bacterial display using circularly permuted outer membrane protein OmpX yields high affinity peptide ligands. *Protein Sci.* **15**, 825–836 (2006).
188. Kenrick, S. A. & Daugherty, P. S. Bacterial display enables efficient and quantitative peptide affinity maturation. *Protein Eng. Des. Sel.* **23**, 9–17 (2010).
189. Arora, A., Rinehart, D., Szabo, G. & Tamm, L. K. Refolded outer membrane protein A of Escherichia coli forms ion channels with two conductance states in planar lipid bilayers. *J. Biol. Chem.* **275**, 1594–600 (2000).
190. Vogt, J. & Schulz, G. E. The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. *Structure* **7**, 1301–1309 (1999).
191. Getz, J. A., Schoep, T. D. & Daugherty, P. S. Peptide Discovery Using Bacterial Display and Flow Cytometry. *Methods Enzymol.* **503**, 75–97 (2012).
192. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).
193. Miltenyi, S., Müller, W., Weichel, W. & Radbruch, A. High gradient magnetic cell separation with MACS. *Cytometry* **11**, 231–238 (1990).
194. Kimple, M. E., Brill, A. L. & Pasker, R. L. Overview of affinity tags for protein purification. *Curr. Protoc. protein Sci.* **73**, Unit 9.9 (2013).
195. Hilpert, K. *et al.* Anti-c-myc antibody 9E10: epitope key positions and variability characterized using peptide spot synthesis on cellulose. *Protein Eng. Des. Sel.* **14**, 803–806 (2001).
196. Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5**, 3610–6 (1985).
197. Krauß, N. *et al.* The structure of the anti-c-myc antibody 9E10 Fab fragment/epitope peptide complex reveals a novel binding mode dominated by the heavy chain hypervariable loops. *Proteins Struct. Funct. Bioinforma.* **73**, 552–565 (2008).

198. Phale, P. S. *et al.* Stability of trimeric OmpF porin: the contributions of the latching loop L2. *Biochemistry* **37**, 15663–70 (1998).
199. Soppa, U. & Becker, W. DYRK protein kinases. *Curr. Biol.* **25**, R488–R489 (2015).
200. Aranda, S., Laguna, A. & de la Luna, S. DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles. *FASEB J.* **25**, 449–462 (2011).
201. Woods, Y. L. *et al.* The kinase DYRK1A phosphorylates the transcription factor FKHR at Ser329 in vitro, a novel in vivo phosphorylation site. *Biochem. J.* **355**, 597–607 (2001).
202. Ogawa, Y. *et al.* Development of a novel selective inhibitor of the Down syndrome-related kinase Dyrk1A. *Nat. Commun.* **1**, 86 (2010).
203. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–30 (1995).
204. Daugherty, P. S., Iverson, B. L. & Georgiou, G. Flow cytometric screening of cell-based libraries. *J. Immunol. Methods* **243**, 211–227 (2000).
205. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252-8 (2014).
206. Casadaban, M. J. & Cohen, S. N. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *J. Mol. Biol.* **138**, 179–207 (1980).
207. Siegele, D. A. & Hu, J. C. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci.* **94**, (1997).
208. Batchelor, E., Walthers, D., Kenney, L. J. & Goulian, M. The Escherichia coli CpxA-CpxR envelope stress response system regulates expression of the porins ompF and ompC. *J. Bacteriol.* **187**, 5723–31 (2005).
209. Schleif, R. DNA Looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
210. du Plessis, D. J. F., Nouwen, N. & Driessen, A. J. M. The Sec translocase. *Biochim. Biophys. Acta - Biomembr.* **1808**, 851–865 (2011).
211. Kenrick, S. *et al.* Flow Cytometric Sorting of Bacterial Surface-Displayed Libraries. in *Current Protocols in Cytometry* 4.6.1-4.6.27 (John Wiley & Sons, Inc., 2007).
212. Sutton, S. Accuracy of plate counts. *J. Valid. Technol.* **17**, 42–47 (2011).
213. Rassam, P. *et al.* Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature* **advance on**, (2015).
214. Leptihn, S., Thompson, J. R., Ellory, J. C., Tucker, S. J. & Wallace, M. I. In Vitro Reconstitution of Eukaryotic Ion Channels Using Droplet Interface Bilayers. *J. Am. Chem. Soc.* **133**, 9370–9375 (2011).
215. Kleanthous, C. Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat. Rev. Microbiol.* **8**, 843–848 (2010).
216. Housden, N. G. *et al.* Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science.* 340, 1570-74 (2013).
217. Gillor, O., Kirkup, B. C. & Riley, M. A. Colicins and Microcins: The Next Generation Antimicrobials. *Adv. Appl. Microbiol.* **54**, 129–146 (2004).
218. Zakharov, S. D. & Cramer, W. A. On the mechanism and pathway of colicin import across the E. Coli outer membrane. *Front. Biosci.* **9**, 1311–7 (2004).
219. Stanislav D. Zakharov, Mariya V. Zhalnina, , Onkar Sharma, and & William

A. Cramer. The Colicin E3 Outer Membrane Translocon: Immunity Protein Release Allows Interaction of the Cytotoxic Domain with OmpF Porin. (2006).
220. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
221. Sharma, O. *et al.* Genome-wide screens: novel mechanisms in colicin import and cytotoxicity. *Mol. Microbiol.* **73**, 571–585 (2009).
222. Johnson, C. L. *et al.* The antibacterial toxin colicin N binds to the inner core of lipopolysaccharide and close to its translocator protein. *Mol. Microbiol.* **92**, 440–452 (2014).
223. Ried, G., Hindennach, I. & Henning, U. Role of lipopolysaccharide in assembly of Escherichia coli outer membrane proteins OmpA, OmpC, and OmpF. *J. Bacteriol.* **172**, 6048–6053 (1990).
224. Noinaj, N., Rollauer, S. E. & Buchanan, S. K. The β-barrel membrane protein insertase machinery from Gram-negative bacteria. *Curr. Opin. Struct. Biol.* **31**, 35–42 (2015).
225. Costello, S. M., Plummer, A. M., Fleming, P. J. & Fleming, K. G. Dynamic periplasmic chaperone reservoir facilitates biogenesis of outer membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4794-800 (2016).
226. Kefala, G. *et al.* Structures of the OmpF porin crystallized in the presence of foscholine-12. *Protein Sci.* **19**, 1117–25 (2010).
227. R. Luke Wiseman, Nora S. Green, ‡ and & Kelly*, J. W. Kinetic Stabilization of an Oligomeric Protein under Physiological Conditions Demonstrated by a Lack of Subunit Exchange: Implications for Transthyretin Amyloidosis†. (2005).
228. Jeanteur, D. *et al.* Structural and functional alterations of a colicin-resistant mutant of OmpF porin from Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10675–9 (1994).
229. Thomas, G., Kozhinjampara R Mahendran4 & Mathias Winterhalter5. Protein reconstitution into freestanding planar lipid membranes for electrophysiological characterization.
230. Jacobus, A. P. *et al.* Optimal Cloning of PCR Fragments by Homologous Recombination in Escherichia coli. *PLoS One* **10**, e0119221 (2015).