CryoEM and Biochemistry-led Discovery of Protein Nanopores for Biotechnology

Peter Chong

PhD

University of York

Chemistry

September 2024

Abstract

Since nanopore sequencing was first established almost thirty years ago, it has become the fourth generation of DNA sequencing. The first nanopore was the pore toxin alpha hemolysin. Since then, other pore proteins have been repurposed as nanopores to form a collection of biological nanopores. Meanwhile, nanoscale holes drilled in silicon wafers have been used as solid-state nanopores. New pore proteins, materials, fabrication techniques, and analysis methods continue to push the field forwards. However, both biological and solid-state nanopores each have their own advantages and disadvantages. A new field of hybrid nanopores seeks to marry the two groups in order to gain the unique advantages of both types of nanopore. This research aims to use the portal proteins from the DNA packaging machinery of double stranded DNA bacteriophages as a new source of biological nanopores or as part of a hybrid nanopore. However, the conditions required for nanopore analysis can require a range of pH values and salt concentrations, depending on the analyte. Therefore, the portal proteins were sourced from thermophilic bacteriophages which live in environments of extreme temperature or pressure. Portal proteins from eight thermophilic bacteriophages were chosen as candidates. Only one, GBSV1, could be purified in a soluble form. Its structure was determined using cryo-electron microscopy. Various mutants were designed specifically for use in hybrid nanopores. However, GBSV1 could not be used as a biological nanopore because it could not insert into a lipid membrane. Meanwhile, it was able to form some hybrid nanopores, but the exact formation conditions require further optimisation.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I wish to thank my supervisors: Professor Fred Antson and Dr Sandra Greive. This project was funded by the White Rose Mechanistic Biology Doctoral Training Programme and DreamPore SAS. This project was first started by Lily Bailey. Maria Chechik mentored me when I first joined the lab during the COVID-19 pandemic and provided sage advice during protein production and cryoEM grid preparation. The YSBL technicians have provided technical support throughout. Dr Huw Jenkins provided invaluable expertise with cryoEM processing. Thank you to Dr Benjamin Cressiot and Laura Ratinho at CY Cergy Paris Université for welcoming me to France for the nanopore tests. I also wish to thank past and present members of the Antson lab, especially Dr Dorothy Hawkins, Dr Milton Ashworth, Anastasiia Herman, Dr De Sheng Ker and Jake Smith. The YSBL Coffee Club has provided moral support, caffeine and jokes. My parents, sister and family in the UK and in HK for moral support. Finally, this work is dedicated to the memory of my late grandparents.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

“加油!” (*gā yáu*)

- Literally “Add oil!”. Common expression of encouragement and support in Hong Kong.

Chapter 1 Introduction

For nearly thirty years, nanopores have been studied for their use in next generation DNA sequencing and their potential for other sequencing and detection applications (Braha et al. 1997; Kasianowicz et al. 1996). Nanopores are simply nanoscale holes in a membrane. An electric current passes through the nanopore. When an analyte, e.g. DNA base, amino acid, virus etc. passes through, it blocks the current. The current blockade is characteristic to the analyte, which can allow sequencing for DNA or identification of proteins or their conformations. The first nanopore sequencing experiment was conducted in 1996 using single stranded DNA (ssDNA) passing through α-hemolysin (αHL) (Kasianowicz et al. 1996). Since then, the range of nanopores has expanded to include other membrane pore proteins and dielectric materials. There are two broad types of nanopore in use: biological (see Biological nanopores) and solid-state nanopores (SSN) (see Solid-state nanopores (SSN)). The highest profile example of the use of nanopore sequencing has been the rise of Oxford Nanopore Technologies (ONT), which was spun out of the University of Oxford in 2005 and made an initial public offering (IPO) on the London Stock Exchange in 2021 ('London Stock Exchange Welcomes Oxford Nanopore Technologies Plc to the Main Market - Welcome Story At London Stock Exchange', n.d.). The company's MinION devices have been used for tracking COVID-19 (Nimsamer et al. 2023), analysing shark DNA (Johri et al. 2019) and on the International Space Station (ISS) (Cowing 2016). Nanopore sequencing is also called 3rd generation DNA sequencing.

1.1. Nanopores

1.1.1. Basic principles of nanopore sensing

The same basic set-up is used for both biological and solid-state nanopores. Two chambers are separated by a membrane. The chambers are filled with buffer of a chosen pH and salt concentration. The membrane can be a lipid bilayer, or SiN_x.

Several parameters are collected throughout nanopore analysis. A glossary of some common nanopore terms and parameters is included in the Appendix. Figure 1-1 shows an ideal example of a solid-state nanopore set up, but the same principles apply to biological pores. By convention, the compartment with the working electrode is the *trans* compartment, and the compartment with the ground electrode is the *cis* compartment. When positive voltage is applied (as in Figure 1-1), positively charged particles are moved by electrophoresis from the *trans* to the *cis* compartment, while negatively charged particles are moved in the opposite direction. In Figure 1-1A, no analytes are present in the compartments, so only the ions in the buffer move through the nanopore. The current measured at this stage is the current that is able to pass through the nanopore. This is the open pore

current or baseline current (I_0). The size of the current depends on the ions, their concentration and the applied voltage.

Figure 1-1B shows the current trace when an analyte (yellow rectangle) is added to the trans compartment and passes through the nanopore. The forces that move analytes into the pore are discussed in more detail later (see Electrokinetics). When the analyte enters the pore, it partially blocks the open pore current. The amount of current blocked depends on the analyte's size, and conformation. The decrease in current is the current blockade depth (DI_b). The current measured when the analyte is inside the nanopore is the blockade current (I_b). The duration that the analyte is inside the nanopore is the dwell time (T_t). The time between the end of one event and the start of the next is the inter-event time (T_i).

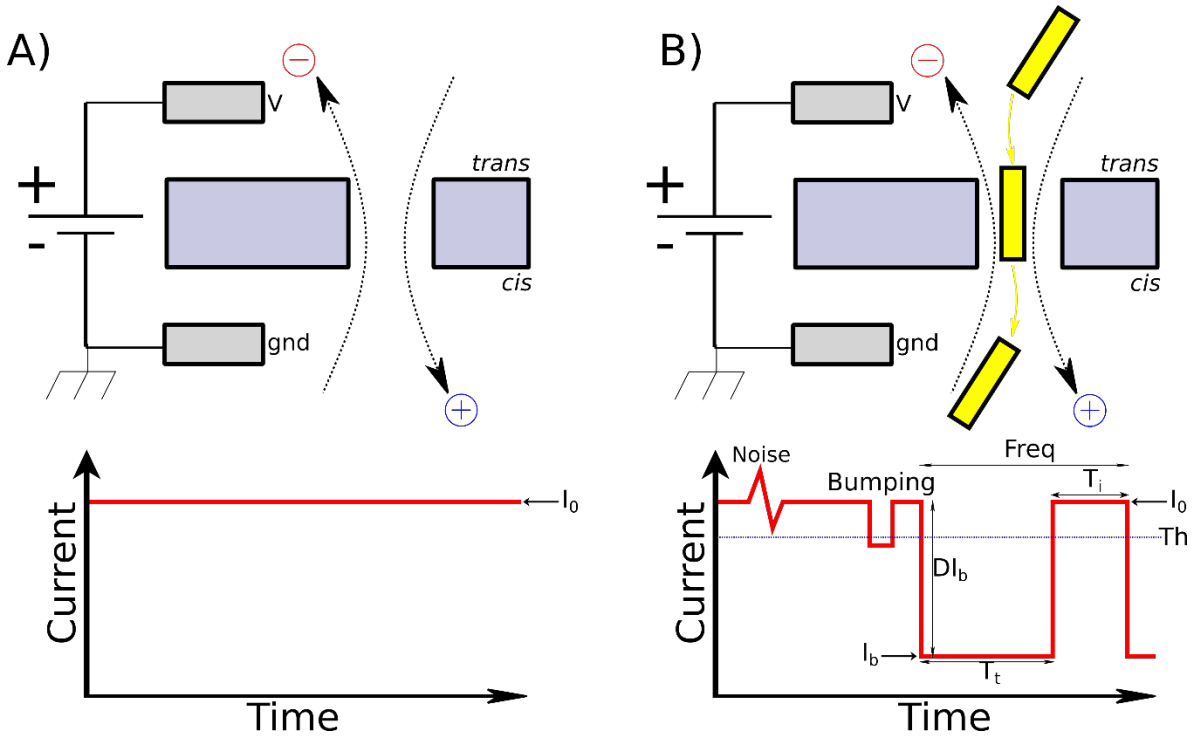


Figure 1-1: Basic principles of nanopore sensing

The basic setup for nanopore analysis. A) A bare solid-state nanopore and the corresponding current trace. Above – schematic of a nanopore with cations and ions passing through. Below – Current over time. I_0 – baseline current. B) The same nanopore with an analyte (yellow rectangle) moving through it from the trans to the cis compartment. Noise is the standard deviation of the current. Blue dashed line – threshold (Th). Bumping is a decrease in current that drops below the threshold, but not enough to be counted as an event. Ib – Blockade current; the average current of an event. DI_b – current blockade depth; the difference between I_0 and I_b. T_t - dwell time; the duration of a current blockade event. T_i – inter-event time; the time between events. Freq – frequency; time between the start of each event.

The baseline current in real nanopore experiments is not a straight line as shown in Figure 1-1, but varies (see Figure 1-2). The standard deviation of the baseline current is the current noise (σ). If every instance of the current falling below the average baseline current was taken as an event, it would include all the noise as well. Therefore, it is necessary to filter the current trace to exclude the noise

and only select those caused by an analyte of interest. A current threshold can be set a certain number of standard deviations below the baseline current. If the current drops below this threshold, it is counted as an event. The number of standard deviations can be decided based on the data, e.g. by examining a histogram of the current throughout the trace (Oukhaled et al. 2012).

However, if the threshold is low enough to exclude all the baseline noise, it can remove the data points at the start and end of an event. This can affect the measurement of event duration, which can affect later calculations (see below). Therefore, two thresholds can be used: a lower threshold (called th2 or threshold) for selecting events, e.g. I_0 - 5σ, and a higher threshold, e.g. I_0 - 3σ (called th1 or level). When the current crosses th2, it is classed as an event. The start and end points of the selected event are taken from where the current trace crosses th1 (Oukhaled et al. 2012).

Some of the selected events are caused by the analyte of interest approaching the nanopore, partially blocking the pore, then leaving again, such events are called 'bumping'. They are characterised by low current blockade, low noise (see Noise) and short dwell time. Bumping events can be filtered out using noise, e.g. noise < 1 pA and or dwell time, e.g. T_t < 200 µs (Ratinho et al. 2024; Greive et al. 2024).

The equations for binding kinetics can also be used for nanopore analysis. They are usually for analytes of interest interacting with or translocating through the pore (Oukhaled et al. 2012). The association rate (k_{on}) was found using equation (1):

$$Association\ rate\ (k_{on}) = \frac{1}{c.T_i} \quad (1)$$

where c is the concentration in M and T_i is the inter-event time in seconds. The dissociation rate (k_{off}) was found using equation (2):

$$Dissociation\ rate\ \left(k_{off}\right) = \frac{1}{T_t} \quad (2)$$

where T_t is the dwell time of an event. The dissociation constant was found using equation (3):

$$Dissociation\ constant\ (K_d) = \frac{k_{off}}{k_{on}} \quad (3)$$

The time between the start of events is the frequency. The conductance of the baseline current and blockade current can be found by dividing by the voltage (equation (4)):

$$Conductance, G\ (S) = \frac{I}{V} = \frac{1}{R} = \sigma\frac{A}{l} = \sigma\frac{\pi d^2}{4l} \quad (4)$$

where I is the current and V is the voltage. Also shown in equation (4) are some alternative methods of finding the conductance. The conductance is the inverse of the resistance (R, measured in ohms, Ω) through the pore. The conductance through the pore is proportional to the cross-sectional area of the pore (A), and inversely proportional to its length (l). The proportionality constant (σ) is the conductivity

(also called specific conductance, κ, measured in S/m). The conductivity will be set by the buffer that fills the nanopore (Bird 2010).

The volume of the analyte compared to the volume of the pore will affect how much of the open pore current is blocked. A small analyte passing through a large pore will block a small proportion of the current, so may be missed. Meanwhile, an analyte passing through a pore which is just big enough will block most of the current, so will be easily spotted. The closer the size of the pore is to the size of the analyte, the higher the signal-to-noise ratio (signal:noise, SNR) will be. The SNR is commonly defined as the ratio between the current depth (DI_b) and the baseline current root mean square (RMS) (I_{RMS}) (equation (5)):

$$signal{:}\,noise, SNR = \frac{DI_b}{I_{RMS}} \tag{5}$$

The I_{RMS} is given by equation (6):

$$I_{RMS} = \sqrt{\frac{1}{n}({x_1}^2 + {x_2}^2 + \ldots + {x_n}^2)} \tag{6}$$

The T_t is affected by the velocity of the analyte entering the pore. If the analyte translocates very fast, it may be missed, especially if it is very small. The frequency and inter-event time are affected by the concentration of the analyte in the solution and the electrokinetic forces that pull it towards/away from the pore. The analyte does not have to pass completely through the pore for analysis. Some nanopores, e.g. αHL, have a larger vestibule which an analyte may enter without being able to pass through to the other compartment. Figure 1-2 shows two additional parameters seen in actual nanopore analysis: the maximum and minimum current blockade values during an event: I_{bmax} and I_{bmin}, respectively.

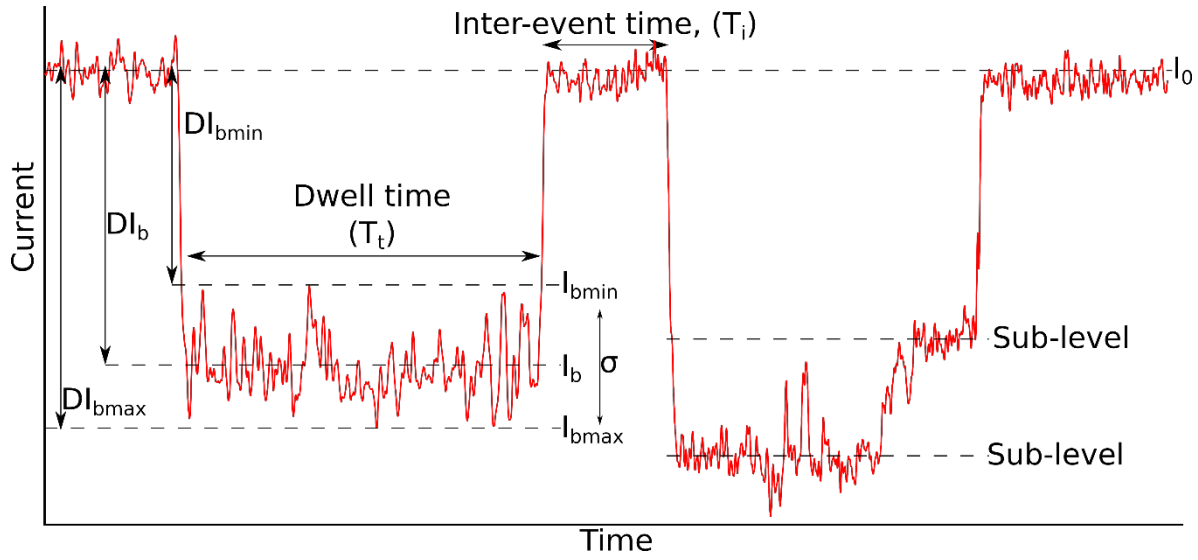


Figure 1-2: Representative ionic current recording with events labelled

The event is taken from one of the hybrid nanopore traces. I_0 - baseline current of the open nanopore. Ib – blockade current, the current when an analyte is passing through the nanopore. DI_b – blockade depth, the difference between the baseline current and the blockade current. I_{bmax} – The lowest current recorded during an event. I_{bmin} – the highest current recorded during an event. DI_{bmax} – the difference between the baseline current and the lowest current recorded during an event. DI_{bmin} – the difference between the baseline current and the highest current recorded during an event. Dwell time, T_t – the duration of an event. Inter-event time, T_i, – the time between events. σ – the noise of an event. Sub-levels are distinct levels within a single event.

1.1.2. Biological nanopores

Biological nanopores are formed by natural pore proteins, often transmembrane proteins spanning a lipid membrane. They are already being used for DNA sequencing, e.g. by Oxford Nanopore. However, they are also being tested for a variety of other biotechnology applications, e.g. protein sequencing, metabolite detection, and modification detection (Y.-L. Ying et al. 2019).

1.1.2.1. Biophysical and biochemical analysis of pores

The study of biological nanopores uses several techniques from the field of structural biology, e.g. affinity purification, enzyme linked immunosorbent assay (ELISA), cryo-electron microscopy (cryoEM) and X-ray crystallography (Iacovache et al. 2016). CryoEM and X-ray crystallography are used to build a structure of the protein which can show the topology and diameters of the tunnel. A cysteine residue can be added to allow the attachment of DNA (L. Zhang et al. 2023) or lipids to the protein (Cressiot et al. 2017a) via a thiol. Mutations to the protein, such as substituting a cysteine, or changing the tunnel diameter are made using site-directed mutagenesis (SDM) during the cloning stage. Some oligomeric biological pores have multiple oligomeric states (Soskine et al. 2013). The different oligomers may each be used as a biological nanopore with a different diameter tunnel.

Biological nanopores can also be delicate, denaturing under assay conditions, which limits the scope of experiments that can be performed. The lipid membranes required to hold biological nanopores

can vary in thickness throughout data collection. This work does not use any of the well-studied biological nanopores (Figure 1-4). However, reviewing their uses, and how they have been modified for different applications can be informative for explaining results seen with the new nanopore and proposing future work. Biological nanopores are reviewed more fully here (Crnković, Srnko, and Anderluh 2021; M. Zhang et al. 2022).

1.1.2.2. Structural features

A solid-state pore is made of the same material throughout and the thickness and internal topology of the tunnel can be determined during fabrication. On the other hand, in a given buffer, each biological pore will form a pore of the same length, diameter, charge and internal topology every time. The internal charge of biological pores is complicated by the amino acid composition of the tunnel. Depending on the buffer, the tunnel may have areas of positive, negative and neutral charge. Some biological pores are derived from toxins, e.g. αHL, ClyA, FraC and aerolysin. Figure 1-3 shows the electrostatic potential of some common biological nanopores. The surface charge of the tunnel affects how analytes enter and pass through the tunnel.

The length and narrowest point of the nanopore partly determine its sensing resolution (see above). (Cao et al. 2018; Di Muccio et al. 2019). The pore diameter at the entrance can also affect the capture of analytes into the pore. Some pores, e.g. αHL and ClyA, have sections with different diameters. The diameter of the pore can affect the entry of analytes (Wong and Muthukumar 2010). In αHL, analytes may enter the larger diameter section (also called the vestibule), but whether they pass through the narrower β-barrel section depends on the buffer conditions, e.g. salt concentration and pH, and voltage (Figure 1-3).

Most biological pores are oligomers of multiple copies of the same protein (homo-oligomers) (Figure 1-4). Some hetero-oligomer nanopores have also been used, e.g. YaxAB nanopore is made of heterodimers of YaxA and YaxB (Straathof et al. 2023). The OmpF and OmpG nanopores are comprised of a single chain that folds into an asymmetrical nanopore. Hetero-oligomeric variants of biological pores can be made in which some number of the subunits have some mutation or addition (Hall et al. 2010a). Oligomeric biological pores are held together by the interactions between their component subunits, usually H-bonds and Van der Waals forces. Biological pores may assemble in solution, e.g. phi29, or assemble on the membrane surface, e.g. aerolysin (Podobnik, Kisovec, and Anderluh 2017) before inserting into the membrane.

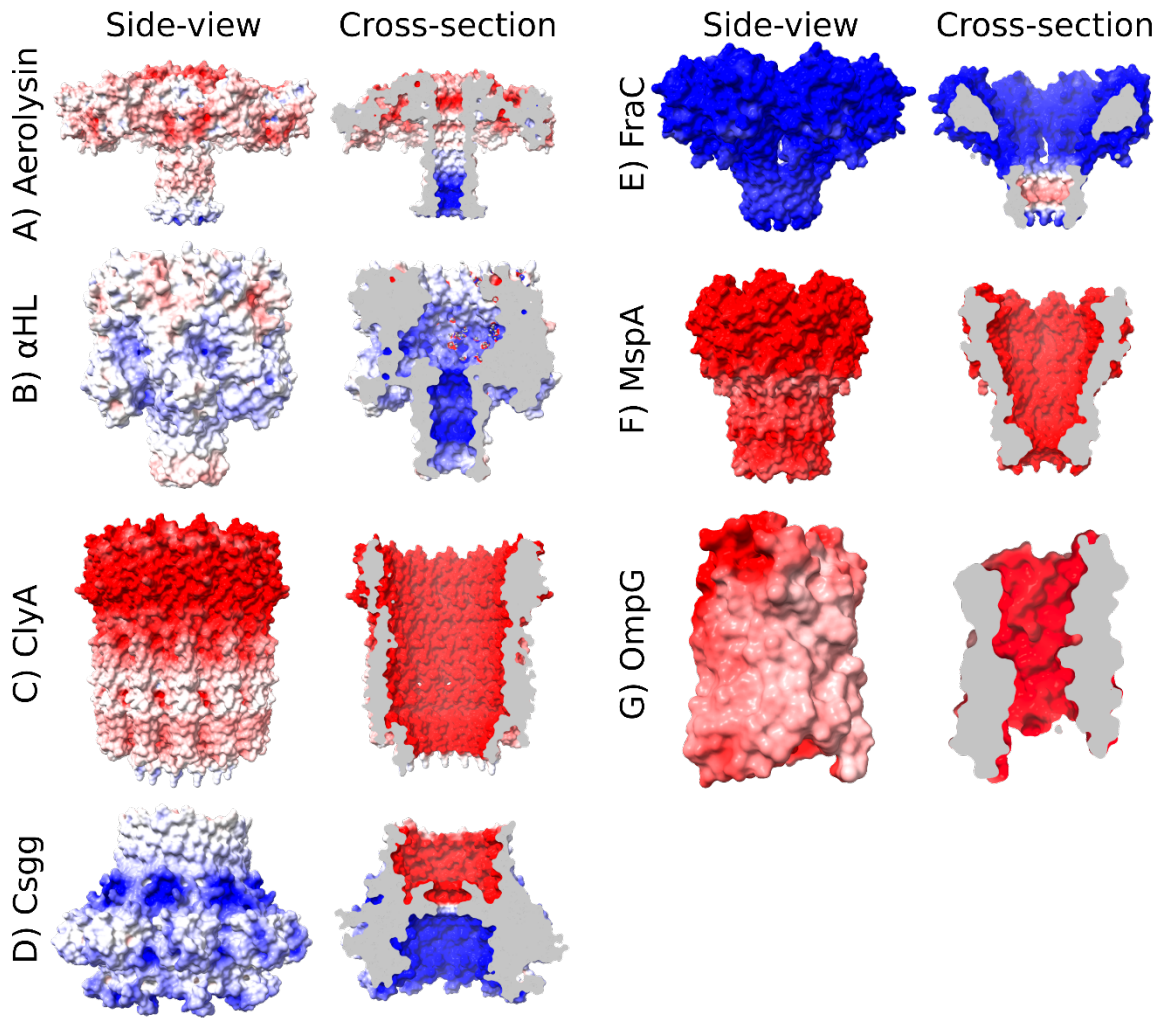


Figure 1-3: Electrostatic surface potential of common biological pores

The electrostatic potential was calculated using PDBtoPQR and APBS webserver (Jurrus et al. 2018). The conditions were pH 7 with no salt. The scale is ±10 kT/e-. Red: +10 kT/e$^-$. Blue: -10 kT/e$^-$. Figures were made in ChimeraX. The structure of the phi29 nanopore was omitted because it could not be converted into a PQR file as too many residues were missing.

Table 1-1: Intermolecular parameters of common biological nanopores

Nanopore	PDB ID	Height (Å)	Width (Å)	Tunnel width (Å)	Assembly state	Buried surface area (Å^2)	ΔG (kcal mol^{-1})	Subunit-subunit interface (Å^2)	H-bonds	Salt bridges
Aerolysin	5JZT	93	159	23	7	47,467.8	-13.2	3,138.8	44	3
αHL	3ANZ	96	91	26	7	47,206	-13.6	2,800.6	59	8
ClyA	6MRU	143	122	54	13	56,781.5	-26.5	2,183.9	24	13
CsgG	4UV3	88	120	17	7	56,730	-26	2,919.3	49	8
FraC	4TSY	75	127	18	4	40.560	-9.4	777.1	8	0
MspA	1UUN	99	99	24	8	30,250	-15.7	1,890	38	4
OmpG	2IWV	51	35	14	1	6,013.4	-195.1	n/a	n/a	n/a
Phi29	1H5W	84	137	40	12	75,950	-29.6	2725.1	26	7

Some of the nanopores have sections of different diameters. The tunnel width is the narrowest point. The buried surface area, ΔG, subunit-subunit interface area, number of H-bonds and number of salt bridges were found using PISA ('Protein Interfaces, Surfaces and Assemblies' Service PISA at the European Bioinformatics Institute', n.d.).

1.1.2.3. Common biological nanopores

(a) Aerolysin

Aerolysin is a pore-forming toxin made of seven subunits (Parker et al. 1994). The oligomer assembles on the cell membrane and inserts a long ß-barrel pore through the membrane (Iacovache et al. 2016). This introduces some complications for its use as a nanopore. An intermediate oligomer, called pro-aerolysin, can form in the membrane. This narrower diameter can give misleading readings. Unlike many other biological nanopores, the tunnel of aerolysin is a straight cylinder. Aerolysin has been used for a variety of applications. Recently, it was used to distinguish enantiomers of the vasopressin protein in a mixture of enantiomers (Ratinho et al. 2024). It has been used in protein sequencing research, though could only identify thirteen of the twenty amino acids (Ouldali et al. 2020). Aerolysin has been shown to be able to distinguish between phosphorylated and dephosphorylated proteins (Stierlen et al. 2023; Y.-L. Ying et al. 2019). This ability was used to monitor the activity of a T4 oligonucleotide kinase (PNK) in real time.

(b) Alpha hemolysin (αHL)

Alpha hemolysin (αHL) is a toxin made by *Staphylococcus aureus* (Figure 1-4B). It inserts into the cell membrane to form water-filled channels. αHL was the first protein to be used as both a biological and hybrid nanopore (see Hybrid nanopores) (Hall et al. 2010a; Kasianowicz et al. 1996). Since then it has been used as a model system for different applications of nanopores, e.g. glycomics (Bacri et al. 2011), or as a chimeric pore with T4 DNA polymerase (Olasagasti et al. 2010). One of the stranger uses has been the creation of an AND logic gate using DNA inputs (Ohara, Takinoue, and Kawano 2017).

(c) Cytolysin A (ClyA)

Cytolysin A (ClyA) is a pore-forming toxin protein. It is produced by certain members of the *Enterobacteriaceae* family. It was first tested as a biological nanopore in 2012 (Soskine et al. 2012). It can form 12-, 13- or 14mers (W. Peng et al. 2019). A full review of ClyA and its properties is outside the scope of this review. More detail can be found here (Murase 2022). Of the well-studied biological nanopores, ClyA has the widest diameter (Table 1-1). This has allowed it to be used to study folded proteins, as well as nucleic acids, like other nanopores. The large vestibule allows protein analytes to enter, but they may then be too large to translocate through the narrower end out of the nanopore. Trapping analytes in this way gives the events a long dwell time, allowing more data to be collected from a single event than if the analytes passed straight through (Soskine et al. 2012).However, the smaller diameter of the lower aperture, meant that ClyA could still be used for DNA analysis but it required some residues to be mutated to positively charged residues to allow DNA to enter (Franceschini et al. 2016).

(d) CsgG

CsgG is an amyloid secretion channel used in the secretion of curli fibres (Robinson et al. 2006). It has been studied for its role in fibre secretion (Goyal et al. 2014), but has also been used as a biological nanopore. Most famously, Oxford Nanopore Technologies (ONT) use a proprietary mutant of CsgG in their MinION, GridION and PromethION devices. They use a proprietary mutant of CsgG with a proprietary helicase attached to the mouth of the pore to feed nucleotides into it at a steady rate. It has been used for DNA (Carter and Hussain 2017) and RNA sequencing (Stephenson et al. 2022; Garalde et al. 2018). ONT have also used 2D sequencing in their MinION devices. 2D sequencing is a method of sequencing a strand of DNA then sequencing the complementary strand. ONT have previously shown 95% accuracy for 2D sequencing of a reference strand (Jain et al. 2015).

(e) Fragaceatoxin C (FraC)

Fragaceatoxin C (FraC) is a protein found in the venom of the *atinia fragacea* (aka the strawberry anemone) (Bellomio et al. 2009). It has been used in peptide sequencing (Restrepo-Pérez, Huang, et al. 2019), peptide differentiation (Huang, Voet, and Maglia 2019), detection of post-translational modifications (Restrepo-Pérez, Wong, et al. 2019).

(f) Mycobacterium smegmatis porin A (MspA)

MspA is a porin channel in the outer lipid membrane of the bacteria *Mycobacterium smegmatis*. It forms a tetrametic pore (Niederweis 2003). It has been studied for its role in solute uptake in gram negative bacteria, but has also been used as a candidate for a biological nanopore, partly because of its stability across a wide pH range (Heinz, Engelhardt, and Niederweis 2003). There are three main mutants of MspA currently being used in the literature: M1MspA, M2MspA, M3MspA (Bhatti et al. 2021). Each subsequent mutant improves its ability as a nanopore for DNA, lowering the voltage required for entry and slowing translocation speed. The mutations change the charges around the mouth of the pore to allow DNA to enter.

MspA has also been used in conjunction with enzymes for different applications. The phi29 DNA polymerase has been joined on to the MspA mouth to feed DNA into it at a set rate (Manrao et al. 2012). A helicase has been used to pull a peptide through MspA via an oligonucleotide linker to enable peptide sequencing (S. Yan et al. 2021). Recently, Brinkerhoff *et al* have tried a method of re-reading a peptide sequence of interest by having a helicase bind a DNA sequence on the end of the peptide. Once the peptide has passed through the MspA nanopore, the helicase hits the nanopore and pulls the whole construct back through, allowing the peptide to be re-read, improving sequencing precision

(Brinkerhoff et al. 2021). However, the peptide sequence can only be as long as the length of the nanopore lumen if the whole sequence is to be re-read.

(g) Phi29 portal protein (connector)

Phi29 is a portal protein from the phi29 bacteriophage (see Portal proteins). Confusingly, portal proteins are sometimes called connector proteins in the literature. It has been well-studied for both its role in dsDNA packaging in the phi29 virus, and as a biological nanopore. Unlike most other biological nanopores, the phi29 portal protein is not a membrane protein. The outside of the stem domain of the phi29 portal protein is less charged and more hydrophobic than other areas, which helps it insert into membranes (Simpson et al. 2001). It is usually inserted into lipid membranes using liposomes (Wendell et al. 2009; L. Zhang et al. 2023; S. Wang et al. 2013).

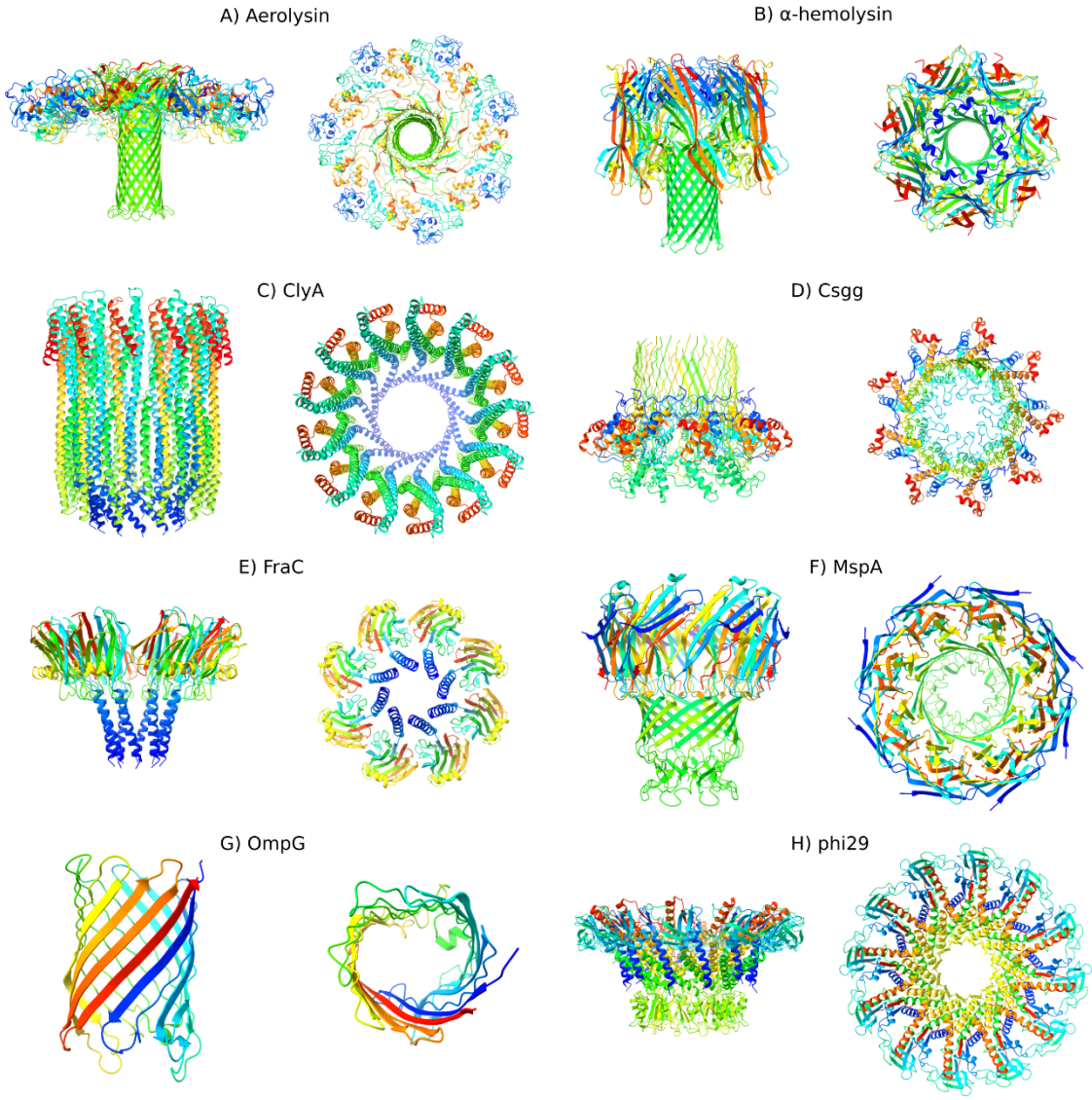


Figure 1-4: Common biological nanopores

The models are coloured by rainbow in ChimeraX. A) Aerolysin (PDB ID: 5ZJT). B) α-hemolysin (αHL) (PDB ID: 3ANZ). C) ClyA (PDB ID: 6MRU). D) CsgG (PDB ID: 4UV3). E) FraC (PDB ID: 4TSY). F) MspA (PDB ID: 1UUN). G) OmpG (PDB ID: 2IWV). H) Phi29 portal protein (PDB ID: 1H5W).

1.1.2.4. Membrane insertion strategies

While membrane proteins offer a natural source of biological nanopores, some groups have attempted to make their own by adapting existing non-membrane protein pores for lipid membrane insertion (S. Wang et al. 2013; Cressiot et al. 2017a; Ji et al. 2020). However, this is energetically unfavourable, so strategies must be employed to successfully insert them into membranes.

One method is to attach anchors to the edges of a pore, which will insert into the membrane, dragging the pore along with them. These anchors can be peptides, such as pHLIP peptides (Jia Tang and Gai 2008) or lipids, such as maleimide (Cressiot et al. 2017a). The pHLIP peptides change their structure

according to the pH of the buffer. Thus, by changing the buffer pH, membrane insertion can be triggered in a controlled manner. However, changing the pH of a solution may be impractical for nanopore experiments, as it may change the behaviour of the biological pore or the analytes. Alternatively, a lipid, such as maleimide can be attached which will naturally insert into the membrane.

Another common approach is to insert the protein into liposomes which can then fuse with the membrane. However, this requires that the protein in question is capable of insertion into liposomes in the first place. This approach has been used for αHL (Ohara, Takinoue, and Kawano 2017), phi29 (Hyun et al. 2010; S. Zhang et al. 2021; Wendell et al. 2009; S. Wang et al. 2013; Ji et al. 2020), T7 portal proteins (Ji et al. 2018), and T7 phage portal proteins (Jing, Paraiso, and Burris 2016). The common method uses sucrose for rehydration of the liposomes (Ji et al. 2018), but glycerol may be a better alternative as it prevents aggregation and leads to faster insertion times (Jing, Paraiso, and Burris 2016). The liposomes can be made to fuse with a membrane to insert the protein. A liposome containing the pore in the membrane and another containing analytes inside can be made to partially fuse with each other. This makes a bilayer between the two liposomes so that only the molecules inside the liposomes can pass through the pore (Ohara, Takinoue, and Kawano 2017).

1.1.2.5. Protein engineering

Some nanopores have been modified to include extra features, e.g. enzymes, DNA, that enhance their abilities. ClyA-aptamers have been made which act as a selectivity filter for desired analytes. The analyte binds the aptamers at the pore mouth, and is released into the pore tunnel. The mutated ClyA monomers were expressed in *E. coli* and purified using Ni-NTA affinity chromatography. The aptamer DNA had a thiol group on one end of the DNA attached to a linker. The thiolated aptamers were mixed with mutated ClyA monomers. Nonconjoined monomers were removed (Soskine et al. 2012).

Mutations have been added to both entrances of the MspA nanopore which have increased the frequency of DNA capture events and increased the dwell time of DNA inside the nanopore (Butler et al. 2008; H. Liu et al. 2017).

Usually, the analyte of interest enters the nanopore to cause a blockade. But the detection of analytes can also be indirect. The Thomsen-Friedreich (TF) binding protein was attached to the C-terminal of the phi29 portal protein. When the biomarker galectin3 (Gal3) bound to the TF binding protein, it caused a conformational change in the phi29 portal protein. The conformational change of the portal protein, rather than the entry of Gal3 caused a current blockade (L. Zhang et al. 2023).

The CsgG constriction peptide, another part of the curli fibre secretion machinery, has been attached to the CsgG nanopore to form a longer nanopore with two constriction points in the tunnel (Van der Verren et al. 2020). The dual-constriction point showed a higher base call accuracy for homopolymer nucleotides than the original CsgG nanopore. Patents have been filed for dual-constriction CsgG nanopores (Remaut et al. 2019) and for a chimeric nanopore made of two CsgG nanopores connected head-to-head (Jayasinghe et al. 2018).

More common is the attachment of an enzyme to the mouth of the pore to feed analytes into the pore. ONT use a proprietary motor attached to the mouth of their CsgG pore to feed single stranded DNA or RNA into the pore. Chimeric nanopores have been made to study DNA replication by attaching a DNA polymerase onto the mouth of an established biological nanopore (Olasagasti et al. 2010; Manrao et al. 2012). In these cases, the DNA polymerases came from T4 and phi29 bacteriophages, which are used to package dsDNA into the viral capsid during maturation. The polymerase ensures a steady rate of entry of bases into the nanopore which helps with accurate sequencing. Enzymes can also be used in conjunction with nanopores even if they are not attached to the pore itself. Recently, Brinkerhoff *et al* have tried a method of re-reading a peptide sequence of interest by having a helicase bind a DNA sequence on the end of the peptide. Once the peptide has passed through the MspA nanopore, the helicase hits the nanopore and pulls the whole construct back through, allowing the peptide to be re-read, improving sequence precision (Brinkerhoff et al. 2021).

A similar system has been used to steadily move a peptide through a nanopore (S. Yan et al. 2021). An oligonucleotide is attached to a peptide analyte. The helicase attached to the top of MspA pulls the oligonucleotide tag through the pore, dragging the peptide analyte through after it. However, other groups have used an unfoldase enzyme attached to the pore to pull the peptide of interest through (Nivala, Marks, and Akeson 2013).

A more extreme approach is to create a new biological nanopore from scratch by attaching different pore proteins together to make a large channel (S. Zhang et al. 2021). In this case, a protease enzyme was also attached to the very top of the pore. Peptides to be analysed could be fed through intact, or chopped up and fed through in pieces for peptide sequencing.

1.1.3. Solid-state nanopores (SSN)

Solid-state nanopores are simply a hole made in a thin sheet of material. The most common are silicon nitride (Si_3N_4) and silicon oxide (SiO_2). Alternatives include hafnium oxide (HfO_2), coated silicon

(Apolinário et al. 2020; Shiyu Li et al. 2020; 2022), glass (Xin Guan et al. 2023), graphene (Wasfi, Awwad, and Ayesh 2018; X. Wu et al. 2018), and molybdenum disulphide (MoS_2) (Sen, Hoi, and Gupta 2021). Since glass nanopores are always pulled into a conical shape, they always exhibit current rectification (see Current rectification). Layers of graphene have also been used as a support. Since the pores are made by the user on demand, their diameter and shape can be made to the user's specifications, unlike biological nanopores. They can also be made to be very thin, which can give them a high sensing resolution (Wanunu et al. 2010).

1.1.3.1. TEM drilling

A common way of making solid-state nanopores is by drilling a hole into a sheet using the beam of a transmission electron microscope (TEM) (Fujinami Tanimoto et al. 2021; Cressiot et al. 2018). This method allows a skilled user to determine the exact shape, size and location of the nanopore on the membrane. However, the internal topology of the nanopore depends on the beam size, the thickness of the material and height at which the beam is focused (Chen, Chang, and Wu 2019). An advantage of using TEM to drill nanopores is that they can be imaged as soon as they have been made, allowing users to see if the pore is of good quality, and is of the desired shape (Fujinami Tanimoto et al. 2021). However, this method requires optimisation and much practice and experience on the part of the user. It is also only possible to make one pore at a time and requires access to a TEM.

1.1.3.2. Controlled dielectric breakdown (CDB)

Controlled dielectric breakdown (CDB) is a relatively new method for making solid-state nanopores. Usually, dielectric breakdown in a capacitor is to be avoided. Once it occurs, the system will lose capacitance, and charge will flow across it as in any other part of the circuit, potentially causing damage to or failure of equipment. However, it can be utilised instead as a way of making SSN (Kwok, Briggs, and Tabard-Cossa 2014).

When a dielectric material such as silicon nitride separates an electrolyte, an electrical double layer will form around it (see Electrical double layer (EDL)). When a potential difference is applied across the dielectric material, the dielectric forms a capacitor with the electrolyte solution on either side of it. The area of the silicon chip exposed to the electrolyte forms the cross-sectional area of the capacitor plate. When an electric field is applied between the two electrodes, the electrons in the atoms in the silicon chip will try to move around their nuclei towards the anode. Soon, a layer of negative charge will accumulate on the side of the dielectric material facing the anode, and a layer of positive charge on the side facing the cathode. Thus, the dielectric material becomes polarised, but in the opposite

direction to the electric field between the electrodes. There is now an electric field within the dielectric material.

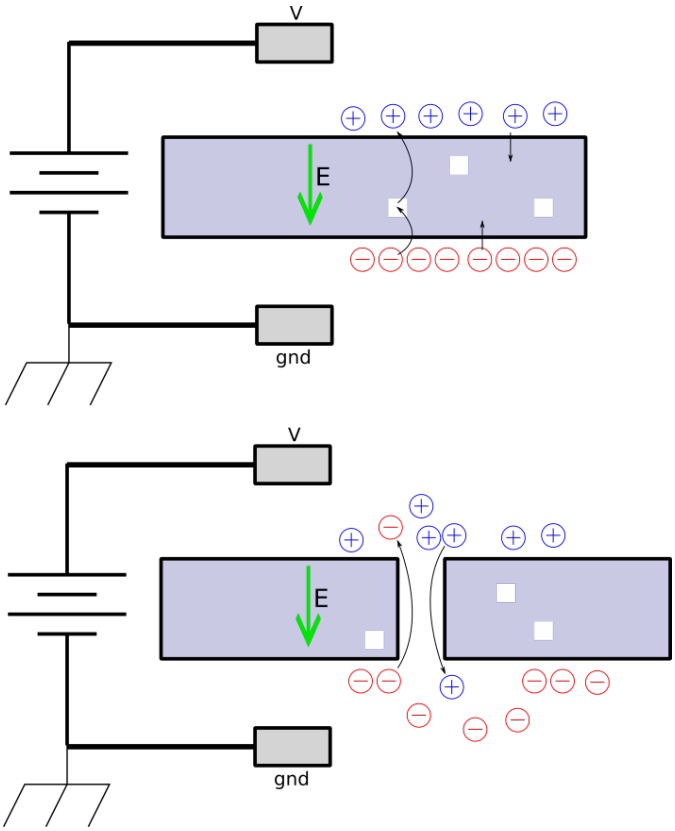


Figure 1-5: Formation of nanopore in SiN_x by CDB

A voltage is applied to a SiN_x chip. An electric field is made between the working electrode (V) and ground electrode (gnd). The concentration of cations increases on the working electrode side. The concentration of anions does the same on the ground electrode side. Top – a build-up of charge traps (white boxes) inside the SiN_x chip. Bottom – enough charge traps have accumulated to form a path through the pore, which can form a nanopore. Green arrow – the direction of the electric field.

When an electric field is applied across a dielectric material, a leakage current will flow across it. Electrons are transported across the membrane via charge traps (defects) in the dielectric material. The leakage current results in Joule heating (Fried et al. 2021). Once these defects form a path between either side of the material, a hard dielectric breakdown event occurs and a nanopore is formed (Waugh et al. 2020). The dielectric strength of a material is the maximum electric field strength that it can endure without dielectric breakdown occurring. The voltage that causes the dielectric strength is the breakdown voltage (V_m). The dielectric strength of a material is given by equation (7) where d is the distance across the material in metres. The breakdown voltage for a material depends on its thickness. The formation of the nanopore can be detected by a spike in the current through the newly formed nanopore.

$$Dielectric\ strength, E_m(V/m) = \frac{V_m}{d} \quad (7)$$

Once the pore has been made, it can be enlarged to the desired size (called Conditioning) by alternating the voltage. The voltage must be lower than that used for nanopore formation to avoid

making additional pores (Waugh et al. 2020). The mechanism for pore enlargement by conditioning is different to that of pore formation by CDB (Leung et al. 2020). Rather than Joule heating, conditioning is driven by the current passing through the pore. Lithium chloride (LiCl) and higher pH buffers lead to faster pore growth than potassium chloride (KCl).

(a) Estimating the size of the pore

Although a nanopore made on a chip by CDB can theoretically be visualised by TEM, its small size compared to the size of the chip makes this prohibitively impractical and difficult. The conductance through the pore is related to its cross-sectional area, length, and conductivity. As shown previously, the conductance through a pore depends on the dimensions of the pore (through its area and length) and the buffer (through its conductivity) (equation (4)). However, if used to find nanopore diameters, equation (4) overestimates the conductance past a certain range of diameters because it only measures the conductance through the pore itself (Kowalczyk et al. 2011). Just as the conductance through the pore can be measured, the nanopore can be thought of as a resistor with the resistance R_{ch} (equation (8)).

$$R_{ch} = \rho \frac{4l}{\pi d^2} \tag{8}$$

However, for nanopores, the total resistance is comprised of the resistance of the nanopore, and the access resistance (R_{ac}) for the entrance and exit of the pore. Access resistance is the resistance of the current when moving from the bulk medium to the constrained nanopore (Hille 1968). Thus, the nanopore can be thought of as three resistors in series (Figure 1-11B) (Vodyanoy and Bezrukov 1992). The total resistance can therefore be written as equation (9):

$$R = R_{channel} + 2R_{access} \tag{9}$$

Ohm's law can be re-written in terms of permittivity, resistivity and capacitance (equation (10)):

$$R = \frac{V}{I} = \rho \frac{A}{l} = \frac{\varepsilon \rho}{C} \tag{10}$$

To calculate the R_{ac}, the mouth of the pore is assumed to be an equipotential flat disc sitting on an insulator (James Hall 1975). The self-capacitance of an object is the capacitance between an object at the centre of a sphere with a radius of infinity and the surface of the sphere. The self-capacitance of a flat disc can be found using:

$$C\ (F) = 8\varepsilon r \tag{11}$$

where r is the radius of the disc and ε is the permittivity of the medium. If equation (11) is substituted into equation (10)), then it gives:

$$R = \frac{\rho}{8r} = \frac{\rho}{4d} \tag{12}$$

This gives the capacitance between a flat disc and a whole sphere, as if the disc were floating alone in solution. However, in the case of nanopores, the disc is on an insulating surface (the membrane);

therefore, the access resistance is the resistance between a flat disc at the centre of an infinite hemisphere, rather than a sphere, and the value should be halved.

$$\text{Access resistance}, R_{ac} = \frac{\rho}{2d} \tag{13}$$

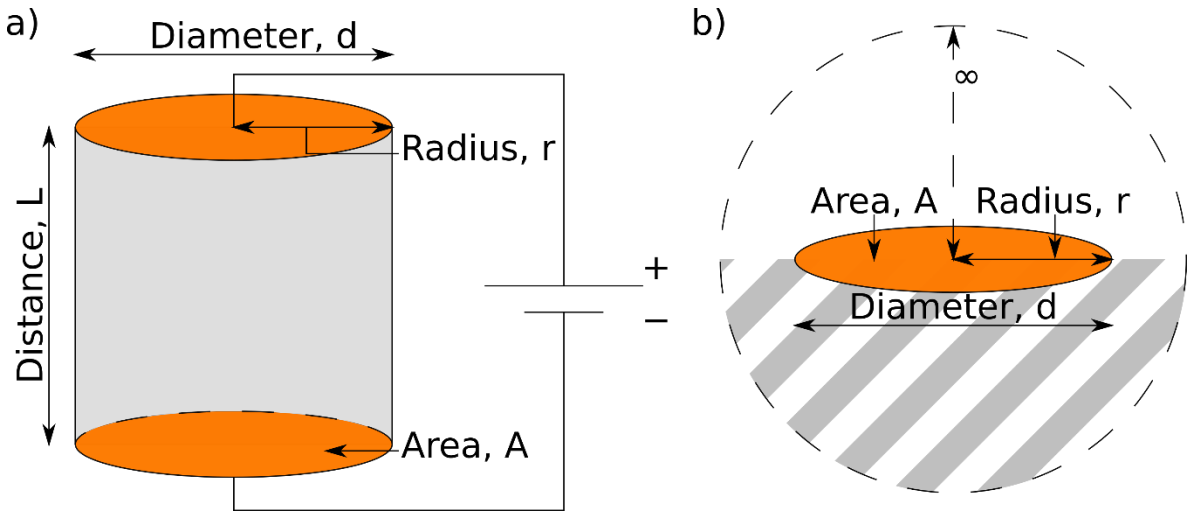


Figure 1-6: Estimations of pore size

a) The nanopore itself is assumed to be a perfect cylinder with no charge. The resistance through the pore itself (R_{ch}) is found using Ohms' law (equation (8)). b) The access resistance (R_{acc}) is found by assuming that the mouth of the pore is an equipotential flat disc. Resistance can be expressed in terms of resistivity, permittivity and capacitance. The self-capacitance of a flat disc is found by finding the capacitance between the disc and the surface of a hollow sphere of infinite radius. The disc is at the centre of the sphere. In the context of a nanopore, the disc is on an insulator (the membrane), so the access resistance is between a flat disc and an infinite hemisphere. The striped area shows the area to be ignored.

Equations (4) and (13) can be used to find the conductance through the pore itself and at the entrances of the pore. If these are substituted into equation (9), then it gives the conductance for the whole nanopore system (equation (14).

$$G = \sigma\left[\frac{4l}{\pi d^2} + \frac{1}{d}\right]^{-1} \tag{14}$$

This can be rearranged into equation (15) to give the diameter of the nanopore.

$$d = \frac{G}{2\sigma}\left(1 + \sqrt{1 + \frac{16\sigma L}{\pi G}}\right) \tag{15}$$

Buffers with high salt concentration, typically 3.6 M LiCl, are usually used for measuring conductance for pore diameter ('Northern Nanopore Instruments User Manual' 2022) because the high salt concentration masks the surface charge of the pore, which better fits the assumptions of the formula. This formula assumes that the nanopore is a perfect cylinder and ignores effects of surface charge (Kowalczyk et al. 2011).

(b) Optimisation and innovation of CDB

While a relatively new technique, CDB has already been adopted and used by different groups.

It has been performed on 12 nm and 30 nm (O'Donohue et al. 2023) thick SiN_x chips. Most groups have made their own home-made systems for performing CDB (O'Donohue et al. 2024; Kwok, Briggs, and Tabard-Cossa 2014; O'Donohue et al. 2023). A Canadian company, Northern Nanopore, has made a commercial device, called the Spark-E2, for performing CDB on 12 nm thick chips. It has already been used by several groups (Briggs et al. 2015; Waugh et al. 2020; Fujinami Tanimoto et al. 2022). As is standard for SSN, the SiN_x chips are typically cleaned with piranha or oxygen plasma (Kwok, Briggs, and Tabard-Cossa 2014) before fabrication. However, there has been one instance of using NaOCl (bleach) in the fabrication buffer without apparently cleaning with piranha or oxygen plasma beforehand (Saharia et al. 2023; Saharia, Bandara, Karawdeniya, Hammond, et al. 2021). CDB has also been performed in 95% EtOH (Briggs et al. 2015), and using LiCl as opposed to the more standard KCl (Bello and Shim 2018).

The pH of the buffer contacting the SiN_x chip can affect CDB. Originally, the same buffer was used on either side of the chip. However, if there is an acidic buffer at the working electrode and an alkali buffer at the ground electrode, then the fabrication times are much shorter (Briggs et al. 2015). The acidic buffer contains lots of H^+ ions and the alkali buffer contains lots of OH^- ions. When a positive voltage is applied, the H^+ and OH^- ions will move by electrophoresis to the SiN_x chip. This will change the pH of the EDL around each side of the chip. The abundance of H^+ and OH^- at either side allows more redox reactions to occur so that there is a greater leakage current through the chip and CDB is more likely. If the buffers are reversed, then fabrication times are much slower, taking days to fabricate and higher voltages to make the initial pore. Pores made using buffers of different pH in each chamber were found to only contact from one side, implying that they are conical in shape. The pH of the buffer has less effect on thinner membranes and they have a shorter range of fabrication times. Once formed, the conical pores can be conditioned to form cylindrical or obconical pores (C. Ying et al. 2016).

Once the initial nanopore has been made by CDB, it must be enlarged to the desired size. This process is called conditioning. Short bursts of alternative voltage are used to slowly enlarge the pore. The size is measured after a chosen number of cycles. This technique was original developed for controlling the size and reducing the low frequency noise from pores made conventionally using a TEM (Beamish et al. 2013; 2012). High electric fields can also help wet the pore (electrowetting).

The SiN_x chips used to make SSN may contain organic molecules, e.g. proteins or DNA from being handled or the environment. If left on the chip, they will be present during analysis and give their own signals. When SSN are made by TEM, these can instead cause the membrane to appear to thicken (Egerton, Li, and Malac 2004). These are usually cleaned from the surface using oxygen plasma or piranha solution. The use of high electric fields during conditioning may help remove such

contaminants (Beamish et al. 2013). Over their life time, nanopores can become clogged with analytes. SSN may be restored by washing the flow cell through with NaOCl (Saharia et al. 2023) or applying high electric fields (Beamish et al. 2012).

(c) Limitations of CDB

While CDB is easier, faster and cheaper than making SSNs using a TEM, it is not without drawbacks. Soft breakdown is the term for a false positive pore formation. A spike in current is measured, as if a pore has been formed, but in fact, it is due to conduction through the silicon chip. Such events are more common in thicker chips (Waugh et al. 2020). The dielectric strength of a material is a constant, so the greater the distance, between the capacitor plates (or the thicker the SiN_x chip), the higher the voltage required for dielectric breakdown to occur (see equation (7)).

The requirement for a high salt buffer for size analysis means that measuring the diameter as an experiment goes on may not be practical, since the test buffers use lower salt concentrations.

SSN are typically hourglass shaped (M. J. Kim et al. 2006; Min Jun Kim et al. 2007; M.-Y. Wu et al. 2009; Hout et al. 2010; Liebes et al. 2010), but equation (15) assumes that the pore is a perfect cylinder. Therefore, it gives the diameter of the cylinder that would give the same conductance as the nanopore in question. As the diameter of the pore approaches its length, measurements become less certain ('How to Measure the Size of Your Nanopore Electrically', n.d.). The diameter of the cylinder given by equation (15) may be the average diameter of the whole nanopore.

The location of the pore created by CDB on the chip is random. It will form wherever the defects are present that allow a conductive path to form between the two sides of the chip. It can be localised by thinning a small section of the chip (Waugh et al. 2020). While it is physically possible to view an SSN made by CDB using a TEM, searching the entire chip for such a small pore is very impractical. Some dielectric breakdown may occur when the two halves of the flow cell are being filled with electrolyte solution (Matsui et al. 2015). This can be a source of dielectric noise (see Noise), widen existing nanopores, or make a new nanopore. Matsui *et al* used a connecting wire between the two chambers to neutralise the charge when electrolyte solution was being poured into the flow cells. However, making SSN using CDB, necessitates that these defects in the membrane will be made. Thus, this source of noise is inherent to the technique.

1.1.4. Hybrid nanopores

A hybrid nanopore is a biological nanopore lodged inside an SSN. They are a relatively new part of the nanopore field. The first paper about hybrid nanopores was published in 2010 (Hall et al. 2010a). The idea is to combine the best of biological and solid-state nanopores. Biological nanopores give the

opportunity to mutate residues to customise the surface charge and shape. They also fold into the same shape every time, giving good reproducibility. Meanwhile, SSN are easier to make and the absence of a lipid membrane removes the possibility of multiple proteins inserting at once. This would also remove the need for the proteins to be membrane proteins, which are notoriously hard to purify and work with. Replacing the lipid membrane with SiN_x should also reduce the high frequency noise associated with biological nanopores (Fragasso, Schmid, and Dekker 2020). Most of the work so far has used alpha-hemolysin (Gornall et al. 2011; Hall et al. 2010b; Cabello-Aguilar et al. 2013; Bentin, Balme, and Picaud 2020).

1.1.4.1. Insertion strategies

There are a variety of different strategies for creating hybrid nanopores. The most common, and simplest, strategy is to plug an SSN with a biological pore (Mojtabavi et al. 2022; Hall et al. 2010b; Cressiot et al. 2018). Other strategies keep the biological pore inside a lipid membrane, then insert the whole thing into the SSN (Göpfrich et al. 2013; Sicard and Yazaydin 2022; Gornall et al. 2011). The biological pore may insert into the SSN without any external force being applied (Cabello-Aguilar *et al.*, 2013), but usually it requires an external force to move it into the SSN. This is usually in the form of electrokinetic forces which are applied through the voltage. The electrokinetic forces encompass the electrophoretic movement of the biological pore and the electro-osmotic force (EOF) (see Electro-osmotic force (EOF)) through the SSN (Sen, Hoi, and Gupta 2021; Cressiot et al. 2018). If both forces pull the biological pore in the same direction, then it will be pulled into the SSN. If they are acting in opposition, then it will only insert if the force moving towards the SSN is stronger. A more complicated approach is to use a long piece of dsDNA attached to the bottom of the pore to drag it into the SSN (Hall et al. 2010b). This requires some engineering of the protein, but ensures that the protein always enters the SSN in the desired orientation. An even more complicated method is to engineer both the protein and the SSN so that the protein is chemically held in place inside the SSN (Mojtabavi et al. 2022). This allows the pore to remain inserted in the SSN at high voltages. However, a biological pore may still enter the SSN and not make the requisite disulphide bridges to the SiN_x surface and leave again when the voltage changes.

1.1.4.2. Methods of making the SSN

The standard method of making an SSN in a SiN_x chip is to use a TEM to carefully drill a hole (Hall et al. 2010a; Cressiot et al. 2018; Mojtabavi et al. 2022). An area can be locally thinned before fabrication to make pore formation easier. The exact parameters used to make the SSN depend on the thickness of the chip and the desired shape and diameter of the SSN. The size of the TEM beam can affect the shape of the SSN (Hout et al. 2010). A thorough dissection of the fabrication of SSN using a TEM is

outside the scope of this introduction but the shape can be carefully controlled by a skilled user (Storm et al. 2003; M.-Y. Wu et al. 2005; 2009; M. J. Kim et al. 2006). Most of the SSN are made in SiN_x. A bilayer of molybdenum disulfide (MoS_2) covering a 200 nm diameter SiN_x pore has also been used to make an incredibly thin support for the biological pore (Sen, Hoi, and Gupta 2021). Other groups have used a glass nanopipette to hold a lipid bilayer containing the biological pore (Gornall et al. 2011).

1.1.4.3. Simulations of hybrid nanopores

Hybrid nanopores have also been made using molecular dynamics simulations (Sicard and Yazaydin 2022; Bentin, Balme, and Picaud 2020). The simulations can be used to simulate experiments that have already been tried, e.g. αHL in an SSN, or a new method, e.g. inserting aquaporin into a lipid nanodisc which is inserted into an SSN (Sicard and Yazaydin 2022).

1.1.5. Advantages and disadvantages of different nanopore types

The biological nanopores can require extensive mutations and have fixed diameters. They are often membrane proteins, which are difficult to work with and purify. There can also be instances of multiple proteins inserting into the lipid membrane, unless additional steps are taken to optimise the lipid membrane, which renders the nanopore useless for sequencing. Protein nanopores are also vulnerable to their environment, e.g. salt conditions of the buffer. But nanopore analysis often requires high salt for good current readings. While SSNs are easier to make than biological nanopores, it is more difficult to fabricate them at a consistent shape in terms of diameter and internal topology.

Table 1-2: Advantages and disadvantages of the different types of nanopores

Nanopore	Advantages	Disadvantages
Biological	Reproducible shape and diameter Opportunity for protein engineering	Difficult to change the size and shape Difficult to produce Membrane proteins can be difficult to work with
Solid-state (SSN)	Easy to produce on scale Diameter determined by the user Thinner nanopores can give a higher SNR	Variable diameter over time. Can be hard to estimate diameter by conductance For TEM drilling, requires training, experience and TEM access
Hybrid	Easy to make like SSN Reproducible shape of biological pores	Biological nanopore may exit the SSN Difficult optimisation Can be difficult to form the hybrids If the SSN is too large, the biological pore may translocate or move about inside, creating noise If the SSN is too small, the biological pore won't enter

Table 1-3: Summary hybrid nanopores made and their conditions

Biological pore	SSN material	SSN diameter (nm)	SSN thickness (nm)	Hybrid diameter (nm)	SSN creation	Hybrid formation method	Source
αHL	SiN_x	2.4 - 3.6	20	1.4	TEM drilled	Electrophoretically pulled by DNA	(Hall et al. 2010a)
OmpF, αHL	Glass	~230 - ~785	n/a	OmpF - αHL – 1.4	Break liposome on glass pipette	Lipid bilayer in a glass pipette	(Gornall et al. 2011)
αHL	PET	~ 6 nm	13,000 (13 µm)	1.4	Track etching to make the initial nanopore, then build a layer of zinc oxide (ZnO) and alumina (Al_2O_3) using atomic layer deposition (Apel 2001)	Insert without voltage	(Cabello-Aguilar et al. 2013)
G20c portal protein	SiN_x	5.4 - 6	30	~2.3	TEM drilled	Electrokinetic insertion	(Cressiot et al. 2018)
OmpG, αHL	BL molybdenum disulfide (MoS_2)	3.4	5	2.2	TEM drill in a sheet of MoS_2 on a base of SiN_x	Electrokinetic insertion	(Sen, Hoi, and Gupta 2021)
G20c portal protein	SiN_x	7 - 8	50	1.8	TEM drilled	Chemical fixation	(Mojtabavi et al. 2022)

1.1.6. Electrokinetics

Electrokinetics is the term for the motion of liquid relative to an adjacent charged surface, or the movement of a charged particle in liquid. It encompasses a range of effects, including electrophoresis, electro-osmosis, streaming current, streaming potential, sedimentation potential gradient, colloid vibration potential, and electrokinetic sonic amplitude (Lyklema 2001). The two effects that are most relevant for nanopores are electrophoresis and electro-osmosis. Another relevant concept which involves electrokinetics is the electrical double layer (EDL, also called double layer, DL).

1.1.6.1. Electrical double layer (EDL)

An object in a fluid will accumulate a layer of ions around it. This layer is called the electrical double layer (EDL) (Figure 1-7). Even in pure water, there will be a concentration of H^+ and OH^- ions. A layer of ions of opposing charge (counterions) will accumulate against the charged surface. This is the layer. Past this there is a layer of more loosely associated ions, called the diffuse layer. The Stern plane between the Stern layer and diffuse layer is called the Stern plane. The location of this plane may not be a set distance in the EDL. Then there is the bulk solution. The plane between the diffuse layer and bulk solution is called the slipping plane. The thickness of the EDL depends on the properties of the electrolyte solution, not on the properties of the material itself (Israelachvili 2011). Coions (ions with the matching charge to the surface wall) will be at lower concentration inside the EDL compared to the bulk solution.

The distance from the surface of the object to the bulk solution is called the Debye length (λ_D, κ^{-1}). The value of the λ_D depends on the ions and their concentration. As the concentration increases, the λ_D decreases (Kirby 2010). For symmetrical monovalent electrolytes, such as KCl, the λ_D of the ions can be estimated using equation (16).

$$Debye\ length, \lambda_D = \kappa^{-1} = \sqrt{\frac{\varepsilon_r \varepsilon_0 RT}{2e^2 I}} \quad (16)$$

ε_r is the dielectric constant (also called relative permittivity). The dielectric constant of water is 78.4 $F.m^{-1}$ (Archer and Wang 1990). ε_0 is the permittivity of free space (8.85×10^{-12} F m^{-1}). R is the gas constant (8.314462618 J K^{-1} mol^{-1}). T is the temperature in Kelvin (298.15 K for 25°C). The parameter e is the elementary charge ($1.602176634 \times 10^{-19}$ C). Since relative permittivity (ε_r) is a ratio of the permittivity of the medium to the permittivity of a vacuum (ε), some sources use the absolute permittivity of the medium (ε) instead (equation (17)).

$$\varepsilon = \varepsilon_r \varepsilon_0 \quad (17)$$

The parameter I is the ionic strength of the solution in numbers per m^3. Sometimes, the inverse of the λ_D, κ, is used (equation (18)).

$$\kappa = \frac{1}{\lambda_D} \tag{18}$$

The molar ionic strength in molar strength is given by equation (19).

$$I = \frac{1}{2}\sum_{i=1}^{n} c_i z_i^2 \tag{19}$$

To find the ionic strength in numbers per m^3, the molar ionic strength is multiplied by 1,000 and by the Avogadro constant (N_A) (6.02214076×10^{23} mol^{-1}). The surface potential (φ_0) is the difference in potential between the potential of the bulk solution and the potential of the solid wall. Confusingly, it is sometimes written as ϕ_0, Ψ_0 or ζ. When $\varphi_0 = 0$, there is no difference between the potential at the wall and in the bulk solution. The solution concentrations at which $\varphi_0 = 0$ are called the points of zero charge (pzc). The pH at which the $\varphi_0 = 0$ is written as pH_{pzc}. An EDL can still form if $\varphi_0 = 0$ (S. Lin et al. 2020). The pH_{pzc} is not necessarily the same as the isoelectric point (pI), which is the pH at which a particle has no charge at the slipping plane. If the only electrolytes are H^+ and OH^- as in pure water, then the pH_{pzc}=pI. The electrokinetic potential (ζ-potential) is a measure of flow. It applies to electrophoresis and electro-osmosis. The ζ-potential can be altered by changing the pH, electrolyte concentration, and adding surfactants that change the properties of the surface (Kirby 2010).

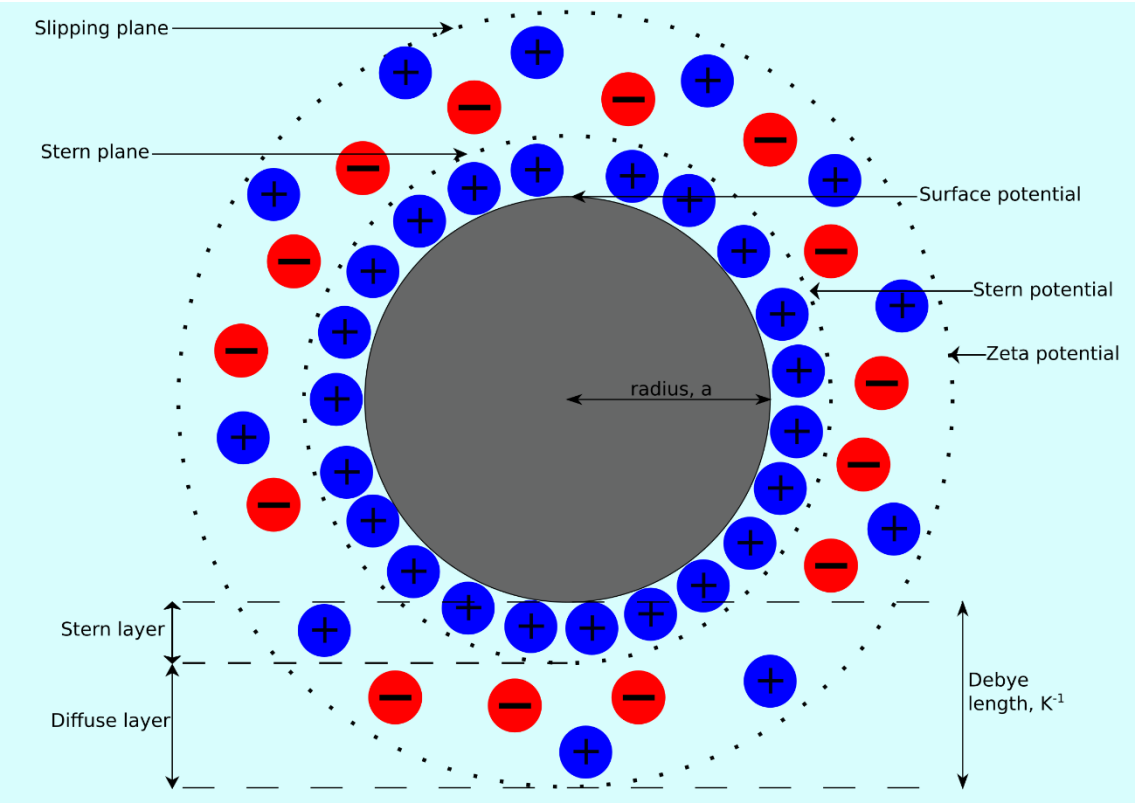


Figure 1-7: Ions around a particle

A negative charged particle (grey) in a solvent (light blue). Cations adsorb to the surface to form the Stern layer. The diffuse layer is comprised of a mixture of ions, but still with a higher proportion of cations than the bulk solution. Stern plane – boundry between the Stern and diffuse layers. Slipping plane – boundry between the diffuse layer and the bulk solution. Debye length – distance between the particle surface and the slipping plane.

1.1.6.2. Electrophoresis

Electrophoresis is the movement of charged particles in an electric field. The velocity of a charged particle is defined by the following equation (20):

$$v_d = \mu_{ep} E \tag{20}$$

where v_d is the velocity (m/s), μ_{ep} is its electrophoretic mobility, and E is the electric field (V/m). Electrophoretic mobility (μ_{ep}) is the ability of charged particles to move through a medium when in an electric field. The units are $m^2/(V \cdot s)$. If a particle's λ_D is smaller than its radius ($\kappa R > 1$) then the μ_{ep} is found using equation (21) (Paillot 1904):

$$\mu_{ep} = \frac{\varepsilon_r \varepsilon_0 \zeta}{\eta} \tag{21}$$

where ε_r is the relative permittivity (also called the dielectric constant, $\varepsilon_r(\omega)$, κ) of the medium, ε_0 is the permittivity of free space, ζ is the zeta potential, η is the dynamic viscosity of the medium. Confusingly, viscosity is sometimes represented by μ. If the λ_D is larger than the radius of the particle ($\kappa R < 1$), then it is defined by equation (22).

$$\mu_{ep} = \frac{2\varepsilon_r \varepsilon_0 \zeta}{3\eta} \tag{22}$$

Ions can be treated as charged spheres for which the λ_D is far greater than their radius. Their electrophoretic mobility is given by equation (23):

$$\mu_{ep} = \frac{Q}{6\pi a \eta} \tag{23}$$

in which *Q* is the charge, and *a* is the ion's hydrodynamic radius (Tabeling 2023).

1.1.6.3. Electro-osmotic force (EOF)

Electro-osmosis is the movement of a liquid through a porous body, in this case a nanopore. The EOF depends on the ζ-potential of the object, λ_D and the pressure through the pore. If no pressure is applied, then the velocity of an electrolyte is given by equation (24):

$$v(z) = -\frac{\varepsilon_r \varepsilon_0 E}{\eta}(\phi - \zeta) = -\frac{\varepsilon_r \varepsilon_0 \zeta E}{\eta}\left(1 - \frac{cosh(\kappa z)}{cosh\left({}^{kh}/_{2}\right)}\right) \tag{24}$$

where ϕ is the surface charge, z is the position across the pore, and h is the pore diameter. If $\kappa h < 2$, then the λ_D on opposite sides of the tunnel will overlap, and so the EDLs will overlap. If the $\kappa h > 2$, then the velocity will plateau in the middle of the pore. The velocity of electrolytes at the plateau is given by equation (25).

$$\mu_{eof} = \frac{\varepsilon_r \varepsilon_0 \zeta}{\eta} \tag{25}$$

Since the charge densities are different, when an external electric field is applied, velocity of fluid movement will be different inside and outside of the EDL. As distance from the surface wall increases, the velocity increases, until it reaches a uniform value at the bulk solution. Figure 1-8 shows the EDL and EOF against one wall of an SSN.

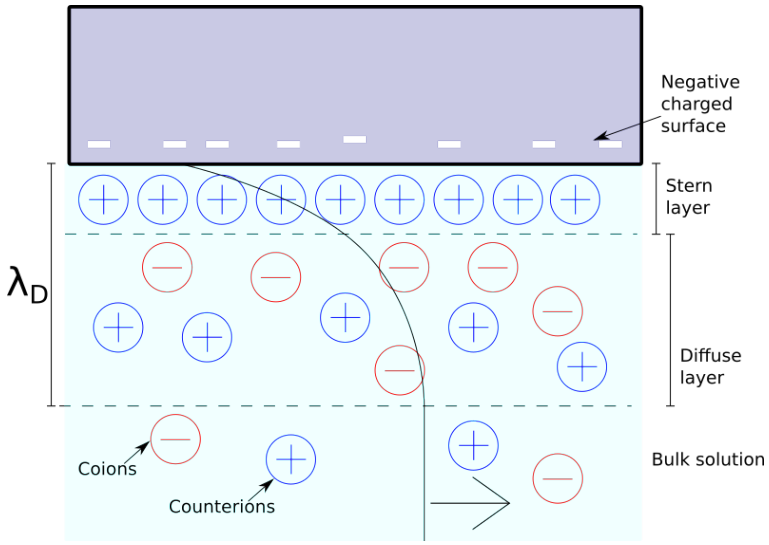


Figure 1-8: Electro-osmotic force along SiN_x surface inside an SSN.

In this example, the SiN_x surface is negatively charged. When in an electrolyte solution, an electrical double layer (EDL) will form. Cations will adsorb to the surface to form the Stern layer. The streaming plane is between the Stern and diffuse layers. The diffuse layer is loosely made up of cations and anions. The thickness of the diffuse layer is approximately the Debye length (λ_D) of the ions. The bulk solution past the diffuse layer is not charged.

The electrophoretic force and EOF acting on a particle can move it in different directions. For example, in Figure 1-8, a negatively charged particle will be moved towards the ground electrode (right to left) by EOF through the pore, but moved towards the working electrode (left to right) by electrophoresis (Figure 1-9). In such cases, the particle will move in the direction of the stronger force.

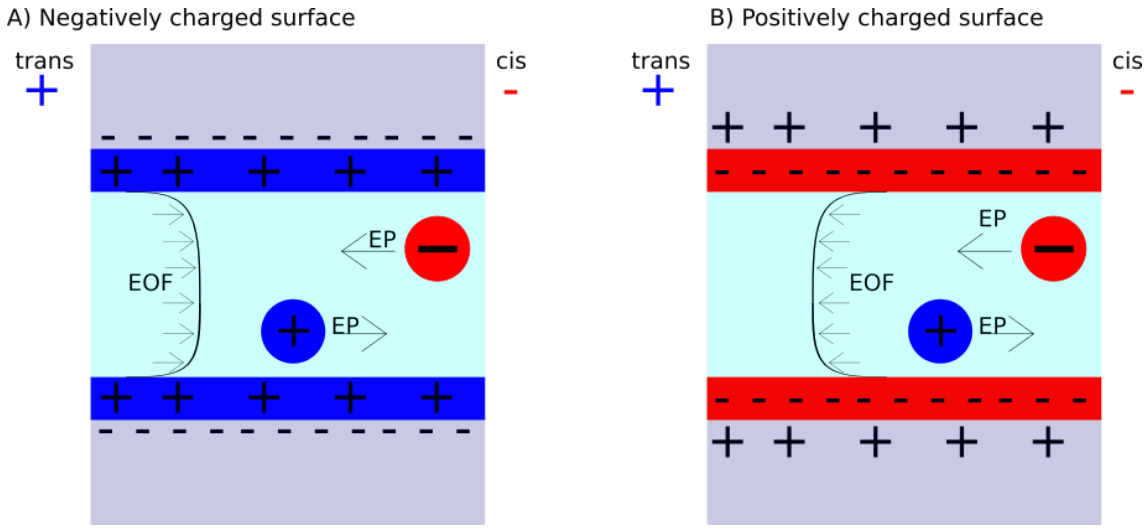


Figure 1-9: Directions of electrophoresis and electro-osmosis

A solid-state nanopore with negatively charged (A) or positively charged (B) tunnel walls. The working electrode is in the trans compartment; the ground is in the cis compartment. The direction of the electrophoretic force on the ions is the same. But the direction of the EOF changes depending on the charge of the nanopore wall.

The velocity of the EOF within the nanopore depends on its shape, diameter and length. Conical nanopores can cause EOF vortices to form within the pore. Narrower and shorter pores have a higher EOF velocity than wider and longer pores. The EOF can be a source of current rectification. Nanopores will often not have the same diameter throughout their length, so the EOF velocity will vary along the length (Rabinowitz et al. 2019; Willems et al. 2020; Aharon-Steinberg et al. 2022; Gubbiotti et al. 2022).

1.1.7. Factors affecting analyte entry into nanopores

The entry of analytes into nanopores, both biological and solid-state depends on many factors. The broad categories of factors are: buffer, pore, analyte, voltage (Table 1-4). Many of these affect each other, which can make optimisation of analysis difficult.

Table 1-4: Factors affecting analyte entrance

Buffer	Pore	Analyte	Voltage
pH	Shape	Size	Voltage
Salt concentration	Diameter(s)	Shape	
Salt identity	Surface charge	Charge	
		Polarisation	
		Concentration	

The pH, salt identity and salt concentration will affect the zeta potential of the analyte and nanopore, which will affect their EP and EOF, respectively. The size and shape of the analyte and pore will determine if the analyte can enter the pore, enter but not translocate, or translocate through. Pores

may have different sections with different diameters, which can allow an analyte to enter, but not translocate, e.g. ClyA, αHL.

The properties of the analyte can also affect its entry into the nanopore. For example, a piece of dsDNA that is shorter than its persistence length (150 bp) may not coil and flex, but behave like a rod. Meanwhile, a longer piece, or ssDNA may form a jumbled coil. As such, shorter dsDNA pieces can easily move into the large mouth of the ClyA nanopore, but ssDNA can only enter if the end of a strand is facing the entrance and must overcome an energy barrier to unwind and enter (Franceschini et al. 2016). Changing the pH of the buffer can also cause a change in the conformation of the analyte, e.g. causing DNA to unwind (Fologea et al. 2005).

If the molecule does not have a dipole in the electric field between the electrodes, then it may tumble around in solution. If the molecule can only enter the pore in certain orientations, then this can hinder its entrance. If the molecule has a dipole, then it may orientate itself in the electric field so that it always enters in the same orientation. If the analyte is charged and the mouth of the pore holds the same charge, that can act as an energetic barrier to its entry. This can be changed in biological pores by mutating chosen residues. If the mouth holds an opposing charge, then it can help attract the analyte into the pore (Fologea et al. 2005; Asandei et al. 2015).

Different electrolytes and concentrations can change whether the analytes enter by electrophoresis or electro-osmosis. At all concentrations of KCl tested, holo-human serum transferrin protein (hSTf) entered via electrophoresis. At high concentrations of LiCl, it entered by electrophoresis; but at low concentrations it entered by electro-osmosis (Saharia, Bandara, Karawdeniya, Hammond, et al. 2021).

As mentioned previously, the diameter and length of the pore will affect the velocity of the EOF through it. If the EDLs on opposite sides of the nanopore overlap, e.g. if the nanopore is very narrow, or if the λ_D is very wide (which may be due to low salt concentration in the buffer), then the EOF velocity in the centre of the tunnel will be lower than if the EDLs did not overlap (Tabeling 2023).

EOF may be the most complicated factor to optimise. Since an SSN is made of the same material throughout, the surface charge should be uniform. On the other hand, biological pores can have a variable charge through their tunnel. In both biological and SSN, as salt concentration increases, and λ_D decreases, the current blockade caused by DNA decreases (Ralph M. M. Smeets et al. 2006; Franceschini et al. 2016). In the case of ClyA, it was thought that at higher salt concentrations, the

ssDNA passes through the pore, while at lower concentrations it stays in the larger vestibule (Franceschini et al. 2016).

1.1.8. Current rectification

Current rectification (also called ion rectification) is a phenomenon when the same voltage at positive and negative polarities gives a different level of current through the pore. The shape of the tunnel of biological pores is fixed by their amino acid structure. As seen in Figure 1-4, their shape is not symmetrical relative to each end of the tunnel. Asymmetrical shape of the tunnel is a common cause of current rectification. The shape of SSNs made with a TEM is an hourglass shape but the exact topology can be controlled by the user (Kim, Lee and Kim, 2011; Muhammad Sajeer P *et al.*, 2022). SSN can also be rectified, but since their shape is chosen by the user, it can be fixed during fabrication. The asymmetric shape of a nanopore will affect the entrance of analytes and the EOF going through it. If the current is moving from the wide-end to the narrow-end of the nanopore, it will be lower than if it was passing the opposite way (Zhou, Perry, and Jacobson 2011; Min Jun Kim et al. 2007). There are a variety of models for current rectification (Siwy 2006). Electrolyte solutions with lower ionic strengths tend to show greater current rectification (Kovarik, Zhou, and Jacobson 2009; Cervera et al. 2006). This may be because the surface charge contributes proportionally more to the total ion current; whereas, at high ionic strength, the surface charge is masked. Changing the pH of the buffer can affect the ion current and rectification by changing the surface charge on the nanopore tunnel. Below the isoelectric point, the SSN material or tunnel will be protonated, so be electroneutral or positively charged (Zhou, Perry, and Jacobson 2011). Current rectification tends to decrease as the pore diameter increases (Kovarik, Zhou, and Jacobson 2009).

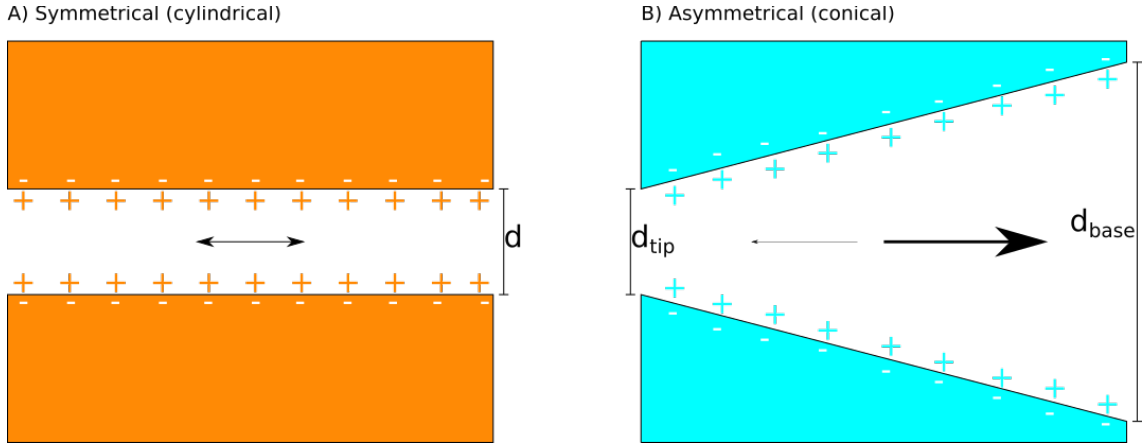


Figure 1-10: Current rectification through symmetric and asymmetric nanopores

A) Current passing through a symmetrical nanopore. The current is the same at positive and negative voltages. B) Current passing through an asymmetrical conical pore. The current is greater when travelling from the narrow end towards the wide end, and vice versa.

1.1.9. Noise

Noise is defined as unwanted electrical signals. Noise is a common problem for both biological and SSN. The power spectral density (PSD) is the power at each frequency in a signal (equation (26)).

$$S_I(f) = \frac{1}{2T} \lim_{T\to\infty} \left| \int_{-T}^{T} \Delta I(t) e^{-2\pi i\, ft} dt \right|^2 \quad (26)$$

S_I is the power, T is time, and I is current. The noise PSD can be broken into 4 components:

$$S_I(f) = S_F\left(\propto \frac{1}{f}\right) + S_W(\propto f^0) + S_D(\propto f) + S_A(\propto f^2) \quad (27)$$

where f is the frequency domain, S_F is 1/f noise (also called "flicker" or "pink" noise), S_W is white noise, S_D is dielectric noise, and S_A is amplifier noise. Each noise component tends to be the dominant source of noise at a different range of frequencies (Liang et al. 2020).

1.1.9.1. 1/f noise

1/f noise tends to dominate at frequencies < 1 kHz. The source of 1/f noise is still being investigated. Sources may include nanobubbles (Roth et al. 2008; R. M. M. Smeets et al. 2006), incomplete wetting of the pore wall (Beamish et al. 2012), and changes in the surface and bulk solution conductance. Changes in surface conductance can be from ionisation of the wall. This applies to biological and SSN.

1.1.9.2. White noise

White noise is frequency independent. The two sources of white noise are thermal (Johnson) noise and shot (Poisson) noise. Thermal noise comes from thermal changes in charge carriers. Shot noise occurs when ions cross a potential barrier, e.g. the Ag/AgCl electrodes. Some white noise will always be present, so there will always be a limit to measurements.

1.1.9.3. Dielectric noise

Dielectric noise is caused by the leakage current through the dielectric material, e.g. lipid membrane, or SiN_x. SSN tend to have higher dielectric noise, as current moves more easily through SiN_x than a lipid bilayer.

1.1.9.4. Amplifier/capacitive noise

Amplifier noise is caused by interactions of thermal noise at the amplifier input with total capacitance of the system. Capacitance is the ability of an object to hold charge. Self-capacitance is the capacitance measured between an object and the ground. Mutual capacitance is the capacitance measured between two objects. There will be capacitance between connected elements of the nanopore set up, e.g. between the membrane and electrolyte solution. If the charges on objects are +Q and –Q, then capacitance is found by equation (28):

$$Capacitance, C\ (F) = \frac{Q}{V} \quad (28)$$

where Q is the charge between the objects and V is the voltage (Harrington 2003). The unit is Farads (F). The amplifier noise is found by equation (29):

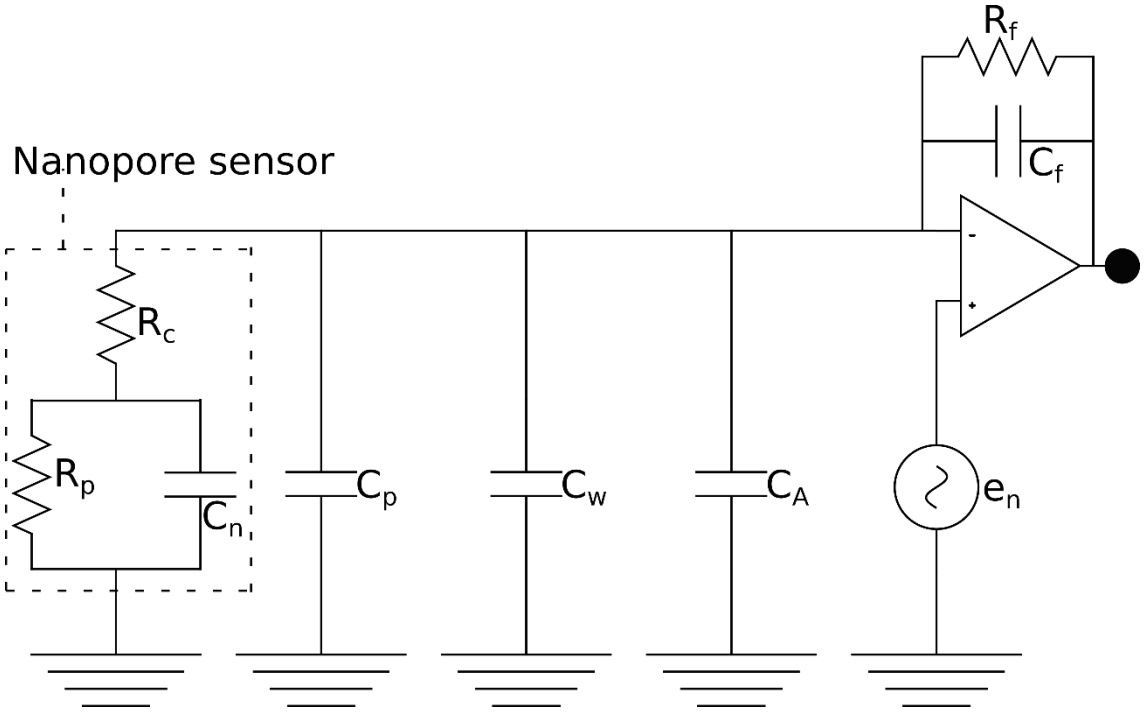
$$S_A(f) = (2\pi C_t e_n)^2 \cdot f^2 \quad (29)$$

where e_n is the voltage thermal noise (also written as v_n) of the amplifier and C_t is the total capacitance of the system. C_t is defined by equation (30):

$$C_t = C_n + C_i + C_f + C_p \quad (30)$$

where C_n is the capacitance of the nanopore itself and the membrane/chip, C_i is the capacitance of the amplifier input, C_f is the capacitance of the feedback resistor, C_p is the parasitic capacitance (Rosenstein et al. 2012; Liang et al. 2020; *Engineered Nanopores for Bioanalytical Applications* 2013).

A) Equivalent circuit diagram of the nanopore sensor and amplifier



B) Equivalent circuit diagram of the nanopore sensor itself

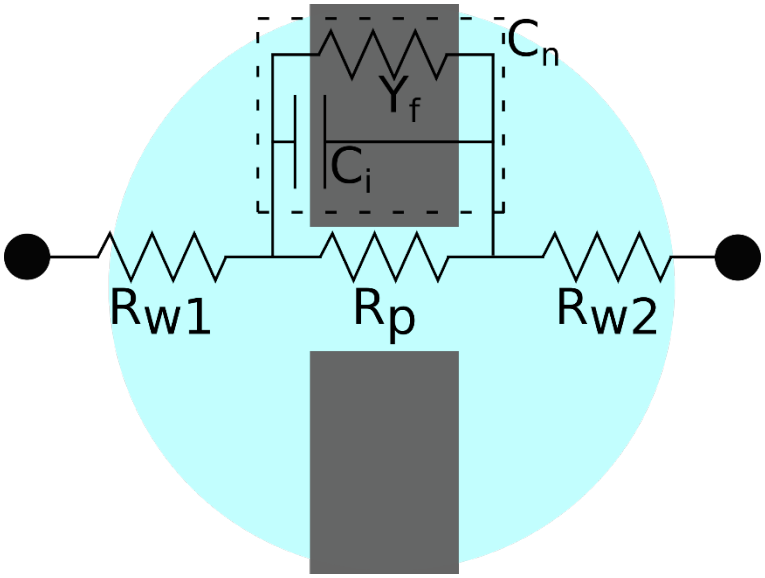


Figure 1-11: Equivalent circuit model of a nanopore

A) Equivalent circuit diagram of the whole nanopore set up. RP, RC, and Cn are the parameters associated with the nanopore sensor itself. RP is the resistance through the nanopore itself. RC is the combined resistance of the pore entrances. Cn is the capacitance of the nanopore. Cp – parasitic capacitance. Cw – capacitance of the wiring. CA – capacitance of the amplifier. en – voltage thermal noise. Rf – feedback resistance. Cf – feedback capacitance. B) Detailed circuit diagram of the nanopore sensor itself. Rw1 and Rw2 are the resistors for the entrance and exit of the pore, respectively. The capacitance of the pore can be modelled as an ideal capacitor (Ci) and an admittance (Yf).

1.1.10. Multi-level events

Events with multiple levels of current blockade can be different. The blockade current will give the average current for the whole event, which can mask the different levels present. There are some software programmes that aim to identify and analyse multi-level events, e.g. MOSAIC (Forstater et al. 2016) and Nanolyzer ('Nanolyzer Nanopore Analysis Software | Northern Nanopore Instruments', n.d.).

1.1.11. Event charge deficit (ECD)

The conventional way of analysing the blockade current of a nanopore is to subtract the average current of a blockade event from the baseline current to give the blockade depth. The amount of current blockaded by a molecule and the time it takes to move through the nanopore depend on its shape. A long thin molecule, e.g. linear dsDNA, will have a smaller current blockade, but longer dwell time, while a short wide molecule, e.g. supercoiled dsDNA, will have a higher current blockade, but shorter dwell time. Thus, the same molecule in different conformations can produce events that appear different in terms of blockade depth and dwell time. Looking at only blockade current depth to identify analytes may misattribute these to different analytes. An alternative is to integrate the area under the curve of the blockade event. This is called the event charge deficit (ECD). The same molecule should have the same ECD, even in different conformations (J. Li et al. 2003; Yupeng Wang et al. 2022). Since ECD can identify the same molecule in different conformations, the ECD can be used to find out if the molecule is interacting with something during translocation, e.g. the wall of the nanopore (Fologea et al. 2005).

ECD has been proposed as a more accurate method of measuring very short and fast events that have a dwell time that is faster than the detector can measure (Zhen Gu et al. 2015). However, this assumes that the event can be treated like a pulse train, where the event is approximated to a square wave, which may not always be accurate. The start and end of the pore event also greatly affect its analysis (Dunbar 2015).

1.1.12. Nanopore frontiers

While Oxford Nanopore Technologies have established commercial DNA sequencing using nanopores, there are other avenues that are still in development. The emerging applications of nanopore technology include ultra-long reads (Hoyt et al. 2022) (Jain et al. 2018), epigenetics (Lau et al. 2023; Hoyt et al. 2022), protein sequencing (S. Yan et al. 2021; S. Zhang et al. 2021), post-translational modifications (Stierlen et al. 2023; Y.-L. Ying et al. 2019), enzyme activity (Fennouri et al. 2013; Y.-L. Ying et al. 2019), glycomics (Bayat et al. 2022; Rivas et al. 2022; Hagan et al. 2020; Rivas et al. 2018;

Fennouri et al. 2013; Bacri et al. 2011), protein folding (S.-C. Liu et al. 2021), and biomarker detection (L. Zhang et al. 2023; Karawdeniya et al. 2018; Taniguchi et al. 2021; Stierlen et al. 2023).

1.1.12.1. Epigenetics and DNA mutations

The MspA nanopore and nanopore sequencing has been used to identify the DNA mutation O^6-carboxymethylguanine (O^6-CMG) (Xiaoyu Guan, Wang, Zhang, et al. 2022; Yu Wang et al. 2019). ONT's MinION device was used to sequence repetitive elements in the T2T-CHM13 human reference genome. The ultra-long reads allowed the construction of locations of CpG methylation, and long repeats through the genome (Hoyt et al. 2022). The methylation of cDNA from cancer and immune cells was analysed with the aim of using methylation as a marker of cancer progression (Lau et al. 2023).

1.1.12.2. Slowing DNA translocation

One method of increasing the resolution of nanopore DNA sequencing is to slow the speed of translocation of the DNA through the nanopore to give more accurate reads. For example, three aspartic acid residues were mutated to asparagine at the mouth of the MspA nanopore. DNA would enter from the opposite end of the pore compared to WT MspA and remain in the pore for longer (H. Liu et al. 2017). SSNs do not have the same engineering opportunities as biological pores, but DNA translocation can be slowed by replacing the buffer in the *trans* compartment with a polyacrylamide gel (Waugh et al. 2015).

1.1.12.3. Post-translational modifications (PTMs)

Protein sequencing is made more complicated by post-translational modifications (PTMs) on the residues. As a parallel field, nanopores are being tested to detect the presence, identity and position of PTMs on peptides. Aerolysin has been used to detect the differences caused by phosphorylation of a protein (Y.-L. Ying et al. 2019), including in different positions along the peptide chain (Stierlen et al. 2023). FraC has been shown to differentiate between unlabelled, phosphorylated and glycosylated peptides (Restrepo-Pérez, Wong, et al. 2019).

1.1.12.4. Monitoring enzyme activity

As mentioned above, enzymes can be joined to biological nanopores to form chimeric pores (Protein engineering). The enzymes provide additional functionality to the biological pores, but their activity can also be monitored from the current trace. Nivala et al use the AAA+ unfoldase enzyme ClipX to pull a protein through an αHL pore (Nivala, Marks, and Akeson 2013). As the protein is pulled through

the pore, the current trace changes to reflect the section of the protein being pulled through. An OmpG mutant with a cut side in a loop can be used to monitor activity of the nuclease-1 enzyme (Fahie et al. 2021). Since the reactants and products of an enzyme reaction give different current blockade signals, nanopores can also be used to monitor enzyme reactions in real time (Y.-L. Ying et al. 2019; Fennouri et al. 2013).

1.1.12.1. Glycomics

Monitoring oligosaccharides using nanopores is still a developing field. Both SSN (Im et al. 2019) and biological nanopores (Bacri et al. 2011) have been used to detect their presence and identities. However, the nature of carbohydrates presents some unique difficulties in detection. Polysaccharides can be made of the same component monosaccharides, but joined with different linkages, which changes their function. For example, the disaccharides maltose and trehalose are both made of two glucose monomers, but maltose has an α(1→4) linkage, while trehalose has an α(1→1)α linkage. Amylose, one of the two components of starch, is an unbranched homopolysaccharide of glucose joined by α(1→4) linkages. Cellulose is the same, except that its monosaccharides are β(1→4) linkages. The different linkages cause cellulose to fold differently to amylose. For successful polysaccharide detection, it is not enough to detect the components, the linkages must also be determined. Unlike polynucleotides or polypeptides, polysaccharides can have branches from the main chain, which would prevent them translocating through a nanopore.

Early work on the detection of oligosaccharides with nanopores compared maltose and dextran through αHL, where it was found that dwell time increased and frequency decreased with oligosaccharide size, as expected (Bacri et al. 2011). However, the normalised blockade current only increased up to ~2,000 g mol^{-1}, past which the oligosaccharides filled the αHL tunnel to the same extent regardless of size.

Glycosaminoglycans (GAG) are a class of polysaccharide made of units of disaccharides. SSNs and machine learning have been used to detect and identify heparin and chrondroitin sulphate at low concentrations (Im et al. 2019).

Another GAG which has seen a lot of investigation is hyaluronan (hyaluronic acid, HA) (Rivas et al. 2022; Fennouri et al. 2013; Rivas et al. 2018; Karawdeniya et al. 2018). HA is part of the extra-cellular matrix of most tissues and has a role in binding proteoglycans together. It is also part of the synovial fluid in joints. It is also seen in the umbilical cord (Standring 2016).

1.1.12.2. Biomarker detection

Accurate detection and quantification of biomarkers of disease is an ongoing field of research for biological (L. Zhang et al. 2023; Stierlen et al. 2023) and solid-state nanopores (Taniguchi et al. 2021; Karawdeniya et al. 2018). Biomarkers can be mutated or truncated proteins (Stierlen et al. 2023), glycans (Karawdeniya et al. 2018), or entire viruses (Taniguchi et al. 2021). The detection of the GAG HA as a biomarker has been mentioned above. Given the relative size of such analytes and the diameters of the common biological nanopores (Table 1-1), some biomarkers may not be able to enter or translocate through most of the pores in a folded state. An exception is ClyA, which has a large upper vestibule which can hold small-medium sized proteins (Soskine et al. 2012). However, translocation of the analyte through the pore, or even entry into the pore may not be necessary for the detection of biomarkers and differentiating their different forms, as long as they give identifiable current signals. Since the diameter of SSN can be set by the user, they do not suffer from this complication. A key limitation that both forms of nanopores have with regards to biomarkers is sensitivity. For these to be usable in real-world scenarios, they must be able to detect and quantify a chosen biomarker at the low concentrations found in physiological samples, e.g. from a blood sample, and from a large mixture of other molecules.

1.1.12.3. Machine learning and AI

Some groups are employing machine learning to remove user bias and increase analysis speed (Xiaoyu Guan, Wang, Shao, et al. 2022; Z.-X. Wei et al. 2019; Taniguchi et al. 2021). Machine or deep learning has been used in both biological (Z.-X. Wei et al. 2019) (Xiaoyu Guan, Wang, Shao, et al. 2022) (Xiaoyu Guan, Wang, Zhang, et al. 2022) and solid-state nanopores (Taniguchi et al. 2021). It has been used for DNA, RNA, the epigenetic mutation O^6-carboxymethylguanine (O^6-CMG), and whole coronavirus virions. It can be used to look at the shapes of the current trace within an event to identify and distinguish between similar analytes (Z.-X. Wei et al. 2019).

1.1.13. Portal proteins as nanopores

Since biological nanopores must span a lipid membrane, most of the biological pores studied thus far have been membrane pore proteins. However, non-membrane pores have also been used, such as the phi29 portal protein mentioned above. While this makes them easier to handle and purify, it means that they often need to be adapted to work as biological nanopores. Portal proteins can be inserted by making proteoliposomes and fusing them with the membrane (Guo and Wang, 2016), or by modifying them to have a skirt of porphyrin molecules around the wing domains (Cressiot *et al.*, 2017). Portal proteins are interesting as a new source of biological nanopores because they have been

extensively studied as part of the DNA packaging machine in bacteriophages. At time of writing, there are no commercial products using portal proteins as nanopores. However, there have been several patents filed for various portal proteins for their use as nanopores. There are patents for the portal proteins of SPP1 (Guo and Wang, 2017; HAQUE and Wang, 2019; Jing, 2020; Haque *et al.*, 2022), phi29 (Guo and Wang 2017; HAQUE and Wang 2019; Jing 2020; Haque et al. 2022), P22 (HAQUE and Wang 2019), RM378 (Hjörleifsdottir et al. 2002), and G20c (Wanunu et al. 2023). The use of portal proteins as nanopores may be a growing area in coming years. The patent also includes other viral proteins from many other viruses (Hjörleifsdottir et al. 2002; HAQUE and Wang 2019). The intersection of nanopores and portal proteins may be expanding in the future. At time of writing, there are only three published instances of a thermophage portal protein being used as a nanopore, all of which use the G20c portal protein (Cressiot et al. 2017a; 2018; Mojtabavi et al. 2022).

Thermus phage G20c is one of the most well studied thermophilic bacteriophages ("thermophages"). It is a member of the *oshimavirus* genus. Its small terminase is a 9-subunit oligomer (Loredo-Varela et al. 2013). There have been attempts to purify the large terminase alone, but these were not taken forwards (R. Xu 2017). The portal protein was found by X-ray crystallography to be a 12mer (L. S. Williams et al. 2013). The portal protein has been adapted for use as a nanopore (Cressiot et al. 2018; 2017a).

It has also been tested as a hybrid pore in a silicon nitride membrane. As a hybrid, it could differentiate between protein, DNA hairpins, peptide and ssDNA (Cressiot et al. 2018). The first attempt at using G20c portal protein as a hybrid nanopore inserted it into the SSN using electrokinetic force. But later work used a surface coating on the SSN to securely fasten the portal protein into the SSN, ensuring correct orientation and long lifetime (Mojtabavi et al. 2022).

1.2. Portal proteins

Portal proteins are non-membrane proteins found in the capsid of dsDNA viruses. They are the entry and exit point for DNA during DNA packaging and DNA ejection respectively. Some of the most well studied portal proteins are shown in Figure 1-12. They commonly have 12 monomers, but some, such as SPP1, have 13 (Dube et al. 1993). Portal proteins have a common fold, called the portal protein fold, made of the same domains: clip, stem, wing, crown. Some, such as P22, have an extended crown domain which forms a long barrel (Jinghua Tang et al. 2011). Despite their similar structure, portal proteins tend to have low DNA and amino acid sequence similarity.

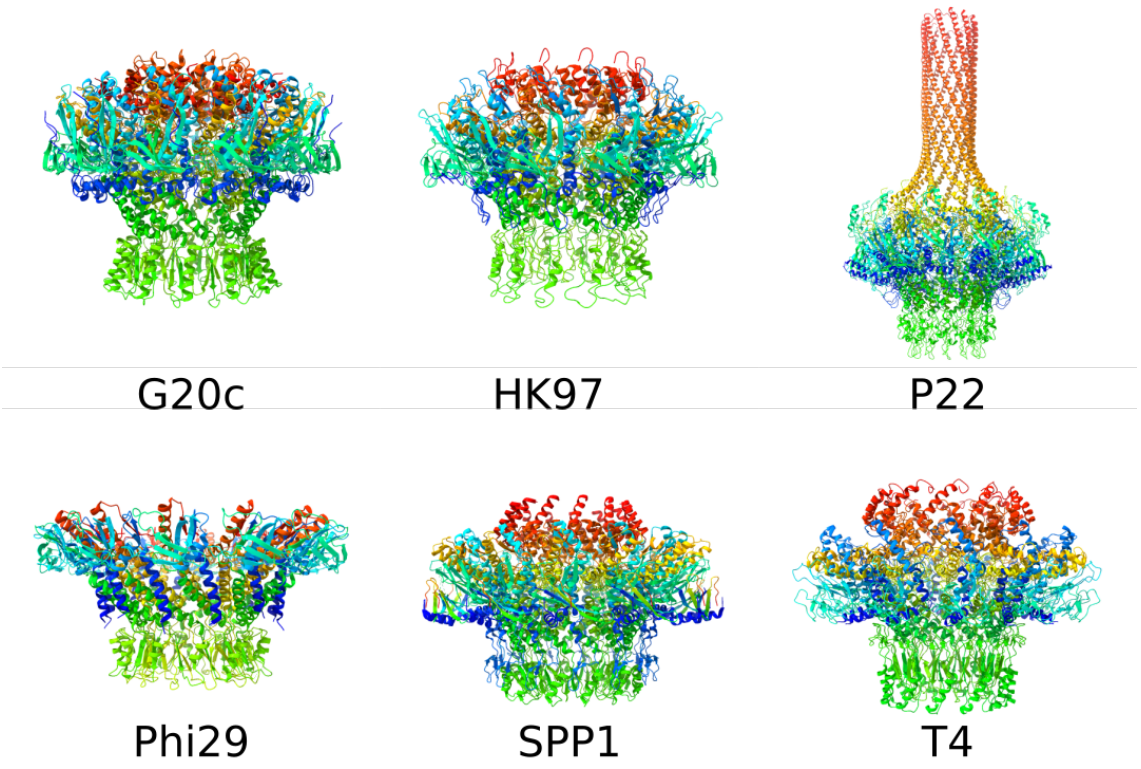


Figure 1-12: Structures of some of the most well studied portal proteins

G20c (PDB ID: 4ZJN), HK97 (PDB ID: 8CEZ), P22 (PDB ID: 3LJ5), Phi29 (PDB ID: 1H5W), SPP1 (PDB ID: 2JES), T4 (PDB ID: 3JA7).

1.2.1. Portal protein structure

1.2.1.1. Crown domain

The crown domain is the C-terminal domain of the portal protein. As seen in Figure 1-12, there are different forms. Those of GA1 and phiCPV4 phage are short and disordered (Javed et al. 2021); whereas, P22 has an enormous barrel domain on top of the crown domain (Olia et al. 2011).

1.2.1.2. Wing domain

The wing domain comprises the main body of the portal protein with the most structured components. Some key features are the tunnel loops (Tunnel loops) and the N-terminal region (N-terminal region). The tryptophan residues in the wing domains are required for the oligomerisation of the P22 portal protein (Woodbury et al. 2022).

1.2.1.3. Tunnel loops

The narrowest point in the lumen of the portal protein occurs in the middle of the lumen and is formed by the tunnel loops in the wing domains. The diameter must be large enough for B-form dsDNA to pass through (~2 nm) (Simpson et al. 2000). Some reconstructions find that the tunnel loops are very flexible and difficult to model (Javed et al. 2021; Sun et al. 2015a; Bayfield, Steven, and Antson 2020). However, the conformation of the tunnel loop may depend on the environment of the portal protein: if it is in solution, in a procapsid, or in a mature virion. The tunnel loop for the T4 portal protein was found to have a weak density when isolated alone, but well-ordered in empty virus heads (Sun et al. 2015a). This pro-head conformation made the tunnel too narrow for dsDNA to pass through. The crystal structure of the G20c portal protein showed that the crystal structure has a 14 Å diameter, with the tunnel loops as part of a structure helix; however, the cryoEM structure shows the tunnel loops have a 22 Å diameter with disordered loops. The comparison also showed that the wider tunnel is more hydrophilic (Bayfield, Steven, and Antson 2020). The variable conformations of the tunnel loops are thought to play a key role in the portal protein's function during DNA packaging (see Portal function during packaging). If the tunnel loops are deleted, from the T4 portal, more 11mers and 13mers were seen. It is possible that the tunnel loops somehow interact with neighbouring subunits to enforce 12-fold symmetry (Sun et al. 2015a). The phiCPV4 has a diameter of 36 Å at the narrowest point. The GA1 phage has a diameter of 34 Å (Javed et al. 2021).

1.2.1.4. Stem domain

The stem domain comprises two helices that connect the clip and wing domains. The two helices are at an angle to the tunnel axis to form a left-handed helix around the DNA in the channel (Simpson et al. 2000; Guasch et al. 2002; Lokareddy et al. 2017; Y.-T. Liu et al. 2019; Machón et al. 2019). When the portal protein changes to its mature conformation, the stem domain moves to elongate the portal tunnel slightly (Lokareddy et al. 2017).

1.2.1.5. Clip domain

The clip domain interacts with the other proteins required for DNA packaging, tail assembly, and DNA ejection. The clip domain is commonly involved in inter-subunit interactions. Usually, there is a three-

strand β-sheet formed of two β-strands from one subunit and one β-strand from a neighbouring subunit (Cuervo et al. 2019; Sun et al. 2015a; Javed et al. 2021; Lebedev et al. 2007). However, the clip domain of phiCPV4 was found to have a three-strand β-sheet where all the β-strands came from the same subunit. Instead, inter-subunit interactions were made between the α4 helix and β9 strands of adjacent subunits (Javed et al. 2021).

The clip domain contains two loops that interact with the DNA when it passes through the tunnel: the channel loop between the clip and stem domains, and the inner clip loop at the mouth of the pore (Padilla-Sanchez et al. 2014; Sun et al. 2015a). The loops each form a constriction point along the tunnel. In the T4 portal protein, the inner clip loop has two basic residues pointing into the tunnel which may interact with DNA in the tunnel (Sun et al. 2015a).

The clip domain is positioned close to the penton hole of the capsid, so it is possible that there are interactions between them to fix the portal in position (Bayfield, Steven, and Antson 2020). The clip domain also interacts with the components of the DNA packaging and ejection machinery, e.g. small terminase, pRNA, large terminase (Sun et al. 2015a) proteins.

However, these partners typically have five-fold symmetry, so there will be a mismatch between them and the dodecameric portal protein. However, some cryoEM studies of portal proteins have found that the clip domain is relatively flexible and could not be resolved with the rest of the portal when 12-fold symmetry was imposed (Rūmnieks, Füzik, and Tārs 2023). Rather, the clip domain was asymmetric inside the penton hole. The P22 large terminase was only able to bind the portal protein when it was inside the procapsid. This suggests that the portal protein, especially the clip domain, changes conformation when interacting with the procapsid. The clip domain was found to adopt a quasi-five-fold symmetry, which allowed it to interact with the large terminase (Woodbury et al. 2022; Lokareddy et al. 2017). The interactions between the T4 portal clip domains and its large terminase were primarily electrostatic (Sun et al. 2015a).

1.2.2. Charge

The surface of the portal protein tunnel tends to be highly negatively charged (Guasch et al. 2002; Olia et al. 2011; Sun et al. 2015a; Cuervo et al. 2019; Y.-T. Liu et al. 2019). This charge should prevent negatively charged DNA from interacting with the tunnel during packaging or ejection (Guasch et al. 2002). To support this idea, if certain negatively charged residues in the tunnel loop or clip loop of SPP1 are mutated, then DNA packaging is impaired (Lebedev et al. 2007). However, in some portal proteins, there are 1-2 rings of positive charge in the tunnel, e.g. T7 (Cuervo et al. 2019), phi29 (Guasch et al. 2002), T4 (Sun et al. 2015a). The function of these positive bands is not certain. The Epstein-Barr virus also has a portal protein, but its tunnel is positively charged, which may be a symptom of a different packaging mechanism (Machón et al. 2019). The residues at the bottom of the clip domain

of the T4 portal protein form charge interactions with the ATPase (Sun et al. 2015a). A similar interaction is seen between the T7 portal protein and its adaptor protein (Cuervo et al. 2019).

1.2.3. Conformations

The conformation of the portal protein changes throughout bacteriophage assembly and DNA packaging. Portal protein monomers require other phage proteins in order to assemble into rings. When the scaffold proteins bind the portal protein monomers, they induce a conformational change which promotes portal ring assembly (Motwani et al. 2017). When the portal protein is inside the capsid, its conformation changes from that of a portal protein alone in solution. The tunnel loops move back and the tunnel becomes more hydrophilic, which allows DNA to pass through (Bayfield, Steven, and Antson 2020).

The structures of the P22 portal protein have been solved in the procapsid and mature virus, which show that it under goes significant conformation changes when DNA has finished packaging into the capsid (Zhiwei Gu, Wu, and Wang 2024; Lokareddy et al. 2017). The P22 portal protein in the procapsid is far more asymmetric. Each monomer has a slightly different conformation, and slightly different positions along the tunnel axis. The clip domain has a quasi-5-fold symmetry. The portal protein tunnel in the procapsid is much smaller than in the mature virus after DNA packaging has finished (Zhiwei Gu, Wu, and Wang 2024; Lokareddy et al. 2017). The most obvious difference between procapsid and mature phage forms of the P22 portal proteins is the massive C-terminal barrel domain seen in the mature portal protein. The barrel forms while the capsid fills with DNA. The DNA pushes two loops at the top of the wing domains towards the crown domains, which may allow it to rotate and allow the barrel helices to straighten and form (Lokareddy et al. 2017).

The conformation of the portal protein can also change as a function of DNA packaging. The portal proteins from the bacteriophages G20c and P23-45 share 99.3% sequence identity. A comparison of the crystal structure of the G20c portal protein (PDB 6IBG) and the cryoEM structure of P23-45 portal protein (PDB 6QJT) showed a difference in the positions of the tunnel loops crown domains. The distance between the tunnel loops in the G20c structure was 1.4 Å, whereas it was 22 Å in P23-45. The tunnel was also more hydrophobic in G20c. When DNA starts to move out of the capsid during packaging, e.g. between bursts of the ATPase/large terminase, it catches on the crown domains, pushing them down. This causes a conformational change in the portal protein which pushes the tunnel loops into the tunnel to engage with DNA and stop it slipping out. Thus, the portal protein acts like a ratchet for DNA (Bayfield, Steven, and Antson 2020).

1.2.4. Portal ring assembly

Portal proteins require bacteriophage coat and scaffold proteins in order to assemble into oligomers (Bazinet et al. 1988). When the scaffold proteins bind the portal protein monomers, they induce a conformational change which promotes portal ring assembly (Motwani et al. 2017). However, portal proteins can also assemble into dodecamers in the absence of other bacteriophage proteins, e.g. if over expressed, which may compensate for the lack of phage proteins.

1.2.5. Oligomers

The most common oligomers are 12mers (Cuervo et al. 2019; Sun et al. 2015a; Rūmnieks, Füzik, and Tārs 2023; Bayfield et al. 2019). Other oligomers are also possible, e.g. 11mers (Dube et al. 1993), and 13mers (Javed et al. 2021; Dube et al. 1993; Jekow et al. 1999). Some portal proteins can form a mixture of oligomers, but only one can be used as the assembly point for viruses (Dube et al. 1993; Sun et al. 2015a; Javed et al. 2021). Portal proteins assemble into rings, but a cryoEM structure of the Pap3 portal protein found that it has a helical structure (David Hou et al. 2022).

1.2.6. Interactions with scaffold proteins

Scaffold proteins bind to the portal proteins to form the procapsid around them. They can also trigger oligomerisation of the portal protein (Sun et al. 2015a; Motwani et al. 2017). Co-expressing scaffold with portal protein was found to give better quality crystals of the portal protein (Luan 2013). Scaffold proteins bind the portal proteins and capsid together. The phiBB1 phage was found to have scaffold proteins binding to the portal protein wing domains at approximately 90° around the portal (Rūmnieks, Füzik, and Tārs 2023).

1.2.7. Symmetry mismatch between portal and capsid penton hole

The portal protein is situated inside an empty penton hole in the procapsid. There is a mismatch of the symmetries between the portal and procapsid. The portal proteins tend to have 12-fold symmetry, but the penton hole has 5-fold symmetry. This mismatch should make it difficult for the portal protein and procapsid scaffold proteins to interact with each other as the same amino acids on each portal protein monomer will not be able to interact with the same amino acids on the scaffold proteins all around the portal. The N-terminal amino acids that form a disordered region before folding into the main body of the portal can be hard to resolve by cryoEM (Bayfield et al. 2019). It is possible that they wrap around the main body of the portal protein and interact with the capsid. Since the amino acids on the very end of the N-terminal and on the edges of the wing domains are on flexible loops, it is possible that the 12-fold symmetry can allow consistent interactions between the portal protein and procapsid proteins (Bayfield, Steven, and Antson 2020).

1.2.8. N-terminal region

A common feature of portal proteins is an unstructured region at the N-terminal. This is located under the wing domain at the top of the stem domain. The nature of this domain can make it difficult to resolve in cryoEM studies of portal proteins, so it is sometimes removed during protein purification or unseen during data processing (Rūmnieks, Füzik, and Tārs 2023; Bayfield et al. 2019; Sun et al. 2015a). These residues are theorised to wrap around the outside of the portal and interact with the major capsid proteins (Sun et al. 2015a). If the N-terminal regions are deleted, head assembly is inhibited.

1.2.9. DNA interactions

The portal protein is thought to interact with DNA at various points along the tunnel. When DNA binds the portal protein, it causes a conformational change (Urbaneja et al. 1994). The T4 portal protein interacts with DNA at 3 positions: the tunnel loop, channel loop and inner clip loop. The latter two positions are in the clip domain. They are separated from each other by approximately 1 helical turn. (Sun et al. 2015a). As mentioned previously, when the capsid fills with DNA, it can trigger the assembly of the barrel domain of the P22 portal protein when the DNA acts as a kosmotropic agent (Lokareddy et al. 2017).

1.2.10. Portal function during packaging

Early models of dsDNA packaging into procapsids involved the portal rotating inside the procapsid to pump the DNA inside (Simpson et al. 2000; Hendrix 1978), or the dsDNA rotating inside the portal as it was packaged (Dube et al. 1993). However, the portal protein makes contacts with the capsid (Bayfield et al. 2019), which would make rotation within the capsid energetically unfavourable. A relatively recent theory of packaging for the phi29 phage does not include the portal protein as an active partner (Woodson et al. 2021). Instead, it is possible that the portal protein acts as a mediator between the virus capsid and the rest of the DNA packaging machinery or as a one-way valve to prevent DNA exiting the capsid during packaging (Bayfield, Steven, and Antson 2020; Lokareddy et al. 2017) (see Conformations, above).

1.2.11. Stability

In addition to other bacteriophage proteins, portal protein ring assembly is also affected by its environment. At univalent salt concentrations above 250 mM or in the presence of bivalent cations, the SPP1 portal protein formed 13mers. However, at univalent salt concentrations below 250 mM, or in the absence of bivalent cations, the ring disassembles into monomers (Jekow et al. 1999).

1.2.12. Subunit-subunit interactions

Once the oligomers have formed, a variety of inter-subunit interactions keep the portal protein dodecamer together. As mentioned previously, the clip domain plays an important role in maintaining subunit-subunit interactions through the three-stranded β-sheet that typically forms between subunits (see Clip domain) (Simpson et al. 2001; Lebedev et al. 2007). The amino acids in the wing domains of some portal proteins, e.g. P22, play a part in the portal assembly into an oligomer (Woodbury et al. 2022), though the mechanism is not clear. Subunits are held together by ionic interactions (Jekow et al. 1999; Dube et al. 1993) and charge complementarity (Simpson et al. 2001; Guasch et al. 2002; Machón et al. 2019). Hydrogen bonds and salt bridges are also required to hold them together (Bayfield et al. 2019; Sun et al. 2015a; Lebedev et al. 2007).

While mesophilic bacteriophages have been well studied, especially in relation to DNA packaging, their thermophilic counterparts have not. In order to survive the harsh environments that they inhabit, thermophage portal proteins must be resilient to disassembly by their environment, e.g. heat, pH, salt concentration. Therefore, the inter-subunit interactions should be stronger than those in mesophilic phages. Few studies have been done on this area in order to compare the stability of mesophilic and thermophilic phages. Additionally, this proposed enhanced stability would also make them ideal candidates for a new family of nanopores. Table 1-5 shows some dimensions for common portal proteins and their inter-subunit parameters. G20c is the only thermophage portal protein in the table. Despite being much smaller than P22, which has a large barrel domain (Figure 1-12) the area of contact between monomers is almost as high. It also has more H-bonds and salt bridges between monomers than the other portal proteins.

Table 1-5: Portal protein dimensions

Portal protein	PDB ID	Tunnel diameter (Å)			Width (Å)		Height (Å)	Assembly state	Monomer-monomer interface (Å²)	Buried surface area (Å²)	H-bonds	Salt bridges
		Clip	Tunnel loops	Crown	Clip	Wing						
Phi29	1H5W	40	n/a	72	72	145	74	12	2,725.10	75,950	26	7
P22	3LJ5	40	50	45	78	167	304	12	5,414.90	130,780	50	19
G20c	4ZJN	49	23	31	80	136	103	12	5,412.90	136,060	65	23
T4	3JA7	32	32	48	74	167	125	12	4,034.80	108,825.50	36	17
SPP1	2JES	33	30	45	74	162	133	13	3,231.20	107,550	29	11
HK97	8CEZ	29	28	40	72	138	77	12	4,654.30	111,811.50	24	12

The cells are colour coded. Low value – green. High value – red. Distances were measured from C_α atoms in ChimeraX. The tunnel loops in phi29 could not be modelled, so the distance cannot be measured. The P22 structure is of the portal protein in a mature virus, including the C-terminal long barrel domain. The models used were: G20c (PDB ID: 4ZJN), HK97 (PDB ID: 8CEZ), P22 (PDB ID: 3LJ5), Phi29 (PDB ID: 1H5W), SPP1 (PDB ID: 2JES), T4 (PDB ID: 3JA7). The monomer-monomer interface, buried surface area and H-bonds were found using PISA ('Protein Interfaces, Surfaces and Assemblies' Service PISA at the European Bioinformatics Institute', n.d.).

1.3. Thermophilic bacteriophages (thermophages)

Thermophages are bacteriophages which can survive in environments of extreme temperature. They have been found all over the world including Japan (Doi et al. 2013), China (L. Lin et al. 2010; B. Liu, Wu, and Xie 2010), Mexico (Yiqian Wang and Zhang 2010), Italy (Severinov et al. 2014), Russia (Yu, Slater, and Ackermann 2006), and Iceland (Hjörleifsdottir et al. 2002). They reside in environments of extreme temperature, e.g. volcanoes (Severinov et al. 2014), hot springs (Hjörleifsdottir et al. 2002; Doi et al. 2013; L. Lin et al. 2010; B. Liu, Wu, and Xie 2010), deep sea hydrothermal fields (Yiqian Wang and Zhang 2010).

Figure 1-13: Taxonomy of thermophages used in this work

Virus	Genus	Old family	Origin
Deep-sea thermophilic phage D6E	Unclassified *caudoviricetes*	*Myoviridae*	Deep sea vent, Mexico
Geobacillus virus E2	Unclassified *caudoviricetes*	*Siphoviridae*	Deep sea vent, Pacific ocean
Thermus phage phiFA	Unclassified *caudoviricetes*	*Siphoviridae*	Mount Etna, Mount Vesuvius, Italy
Thermus phage phi OH2	Unclassified *caudoviricetes*	*Siphoviridae*	Hot spring, Nagasaki, Japan
Bacillus phage 1	*Svunavirus*	*Myoviridae*	Inshore hot spring, Xiamen, China
Geobacillus phage GBSV1	*Svunavirus*	*Myoviridae*	Offshore hot springs, Xiamen, China.
Rhodothermus phage RM378	Unclassified *caudoviricetes*	*Myoviridae*	Hot spring, Iceland
Thermus phage TSP4	*Oshimavirus*	*Myoviridae*	Tengchong hot springs, China

The virus taxonomy was reorganised in March 2021, which meant that the virus families of siphoviridae, podoviridae and myoviridae were abolished (Turner et al. 2023). They are included here because such terms are still in common use.

1.4. Protein characterisation methods

1.4.1. Basic principles of microscopy

Both TEM and cryo-electron microscopes use the diffraction of electrons by a sample to image it. The small wavelength of electrons allows them to be scattered by smaller details of a sample than visible light. However, working with electrons brings some difficulties which visible light lacks. Electrons famously exhibit properties of a wave and particle (Thomas Young 1804). Both modes are used in understanding TEM and cryoEM. Many of the processes and principles that apply to TEM, also apply to cryoEM. The anatomy of a microscope is largely the same. Electrons have a wavelength small enough to allow the imaging of the positions of individual atoms, and can be produced by an electron gun, then focused onto the sample. If the sample is exposed for too long, the beam will damage it. It is not the case that multiple beams of electrons are fired through the microscope from the gun and hit the sample, then interfere with each other on the other side. Because of the speed at which the electrons are moving through the column, there is only ever one electron in the microscope at a time. This single electron hits the whole specimen, everywhere, all at once, then the next electron enters the microscope. The electrons behave as both waves and particles.

1.4.1.1. Anatomy of a microscope

An electron gun is made of three components: the filament, the Wehnelt cylinder, and the accelerator stack. The filament is the source of the electrons. The Glacios microscope at the University of York uses a field emission gun (FEG) in which an electric current is used to pull electrons from the tip. The filament is held at a very low voltage, typically -200 kV or -300 kV. The accelerator stack is a series of metal rings. Each is held at a progressively higher voltage from -300 kV to 0 V. When the electron passes through the accelerator stack, it accelerates. If it starts at -300 kV, then when it exits the stack, the electron will be travelling at 76% of the speed of light. The bottom of the microscope is grounded, so the electrons move through the electric field, down through the microscope. An electron lens is made of several coils of wire through which a current is passed. The electric fields that form around it resemble a toroid. If an electron enters the field at a divergent angle (moving away from the central axis), the fields will change its velocity so that it exits the lens at a convergent angle (moving towards the central axis). In this way, they behave much like optical lenses. The whole column is held at a high vacuum. However, there will always be some contamination in the column. The signal from electrons that hit these contaminants must be removed.

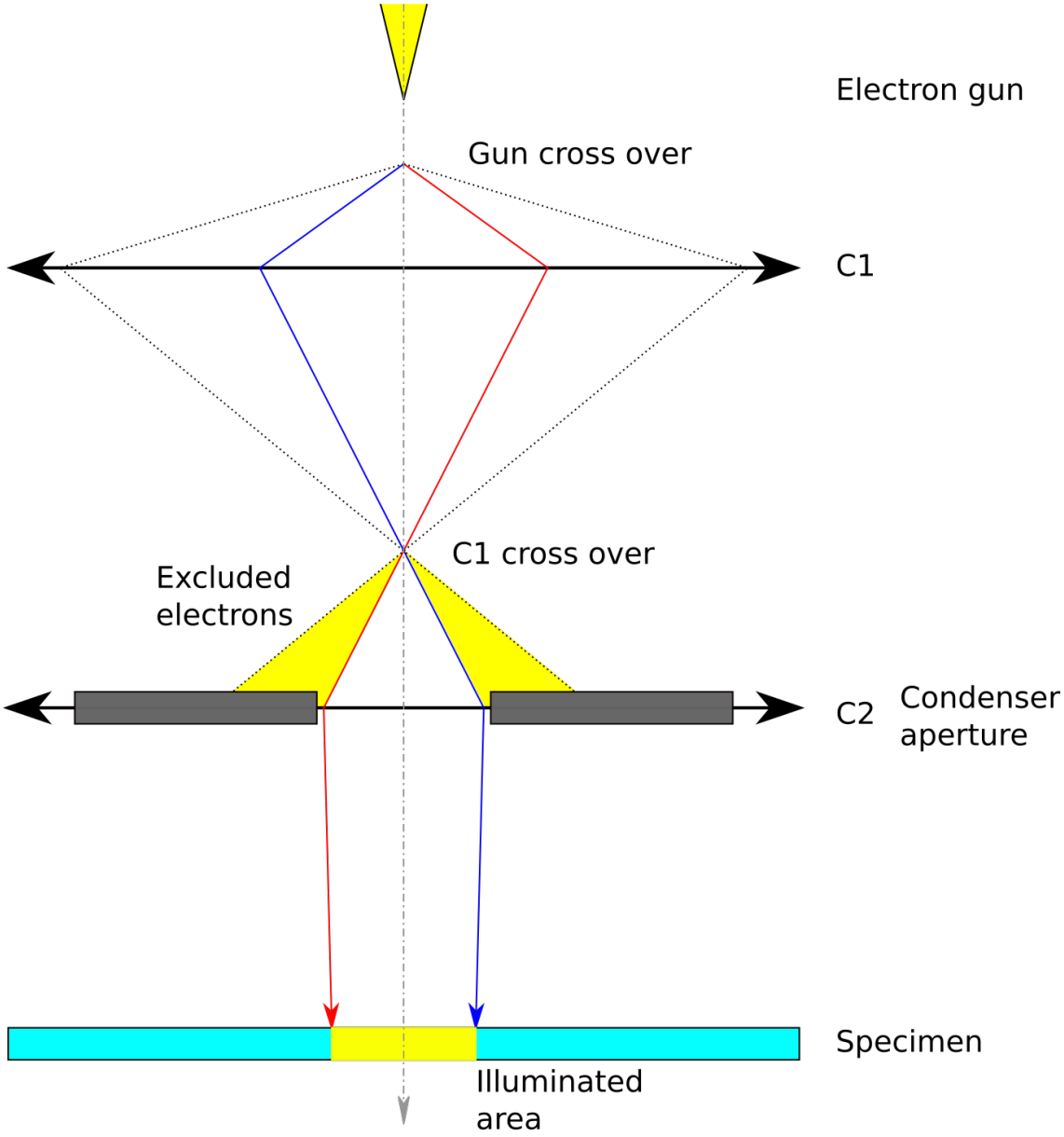


Figure 1-14: Condenser system

Electrons are pulled from the filament by an electric field. They are accelerated to typically 200 kV or 300 kV by the accelerator stack (located below the gun, but excluded for clarity). The condenser aperture is on the same plane as the C2 lens. The beam here is under focused. Red and blue lines show the most widely diffracted beams that pass through the C2 aperture. Dotted lines show electrons that are excluded by the C2 aperture.

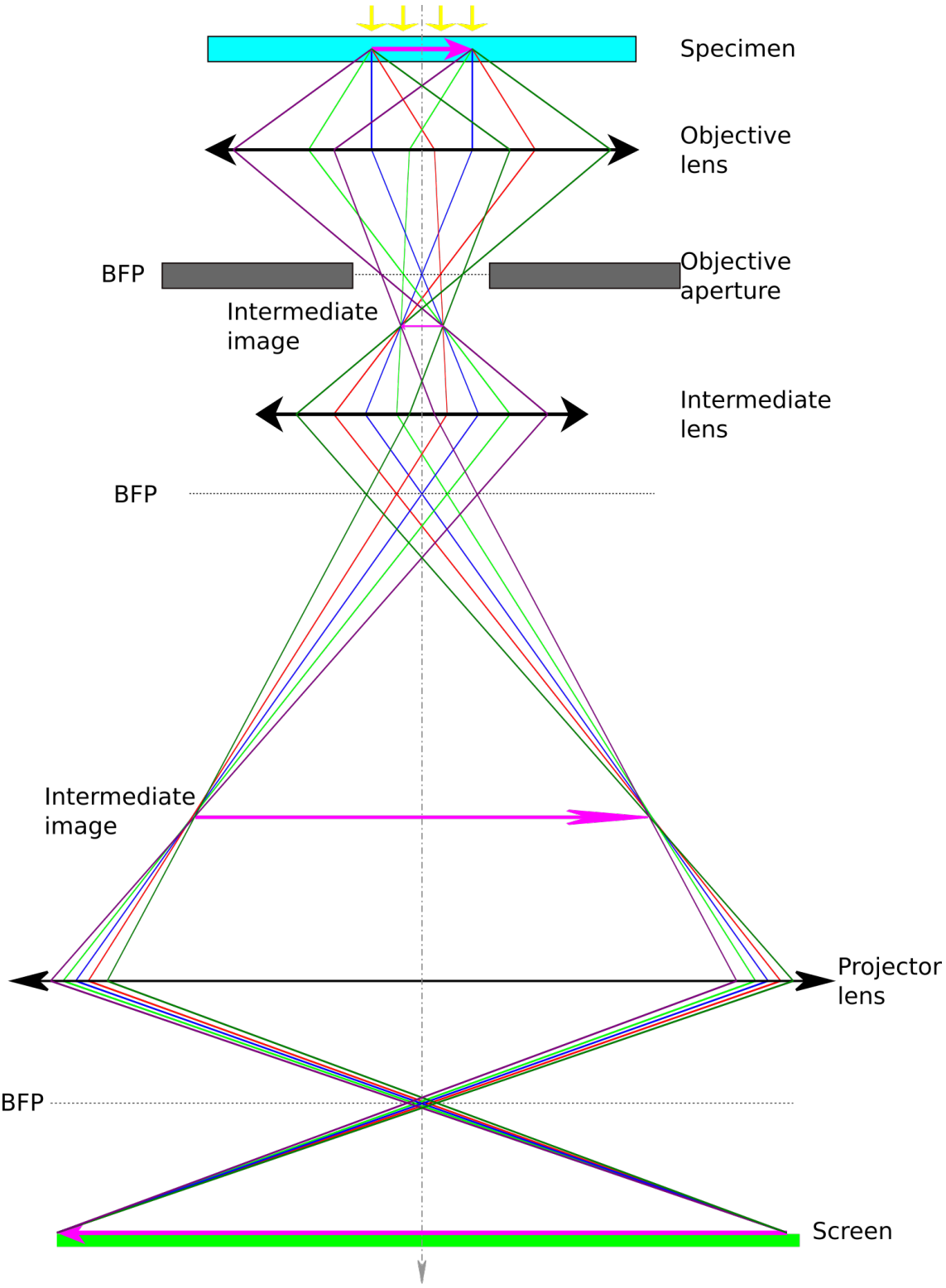


Figure 1-15: Beam diagram of a microscope in imaging mode

The objective and projector lens magnify the specimen image and remove electrons that have been highly scattered. Five rays are shown scattered from two points on the object in the specimen (purple arrow) passing through the lenses to the screen/detector. The planes in the microscope where the rays that left the object at the same angle (matching colours) intersect are the back focal planes (BFP) and form a diffraction pattern. Where the beams that left the same part of the specimen intersect, they form an image of the specimen. The detector at the bottom can be CCD camera or DED. N.B. None of this is to scale.

1.4.1.1. *Eucentric height*

The eucentric plane is a horizontal plane within the column of the microscope. It is an equal distance between the upper and lower objective lenses. It is used as a reference plane so that all calibrations will be reproducible. When the objective lens current is set to a standard reference value, the focus point (beam cross over point) will be in the eucentric plane. When the specimen is in the eucentric plane, it is said to be at the eucentric height.

The position of the grid is adjusted using three axes and two rotations (Figure 1-16). The axes are relative to the grid, not the microscope. The X-axis that runs along the arm of the sample holder intersects with the eucentric plane. If the specimen is below the eucentric plane and the sample holder is rotated around the X-axis, then the area currently being viewed will move out of the field of view. On a TEM, the α-tilt function will rotate the specimen holder around the X-axis a set distance. To move the specimen to the eucentric height, the sample holder is rotated around the X-axis (α-tilt), and then the height of the specimen is moved (in the Z-axis perpendicular to the grid, not the microscope) so that it lies in the eucentric plane. When the specimen is in the eucentric plane, then the area being viewed (that lies on the X-axis), will not move out of view when the sample holder is tilted (D. B. Williams and Carter 2009).

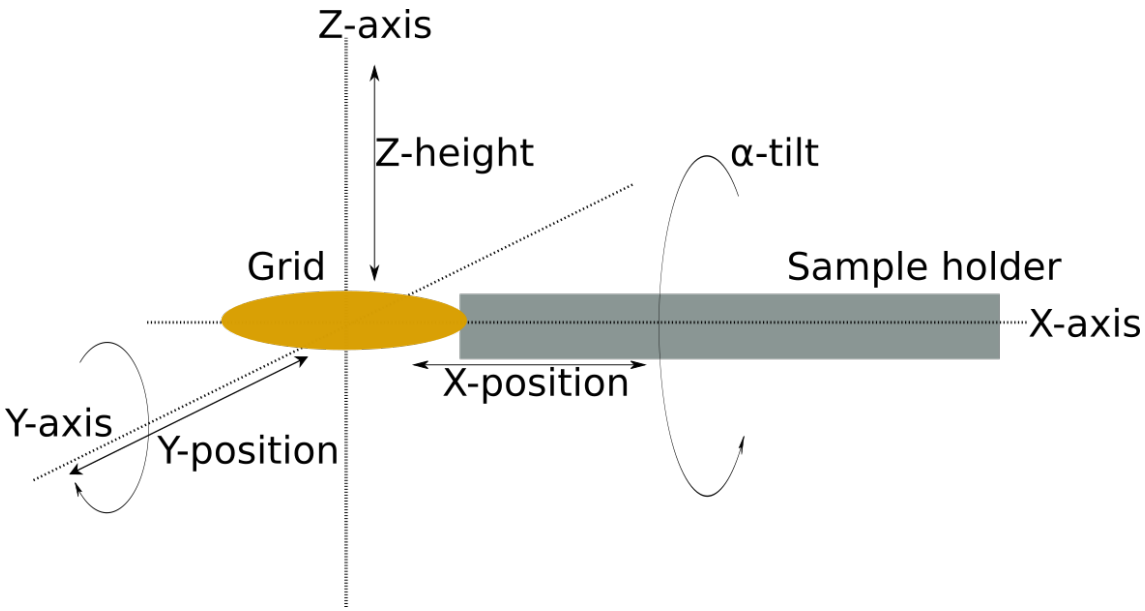


Figure 1-16: Axis for grid position

The x-axis runs along the arm of the sample holder. The y-axis runs horizontally perpendicular to the x-axis. The Z-axis is vertically perpendicular to the grid. The axes are relative to the grid and sample holder, not the microscope.

1.4.1.2. *Focus*

Like visible light, a beam of electrons can be focused using lenses. Instead of glass, lenses are electromagnets. The strength of the lens can be changed by altering the voltage passing through it.

Figure 1-17 shows the effect of changing the strength of a lens. The image plane is below the lens. The point at which the beams intersect is the beam cross over point. If the cross over point is below the image plane, the lens is weak and the image is under-focused, or defocused. If the cross over point is on the image plane, it will produce an in-focus image. If the cross over point is above the image plane, the lens is strong, and the lens is over focused. Lenses can be stacked in a series.

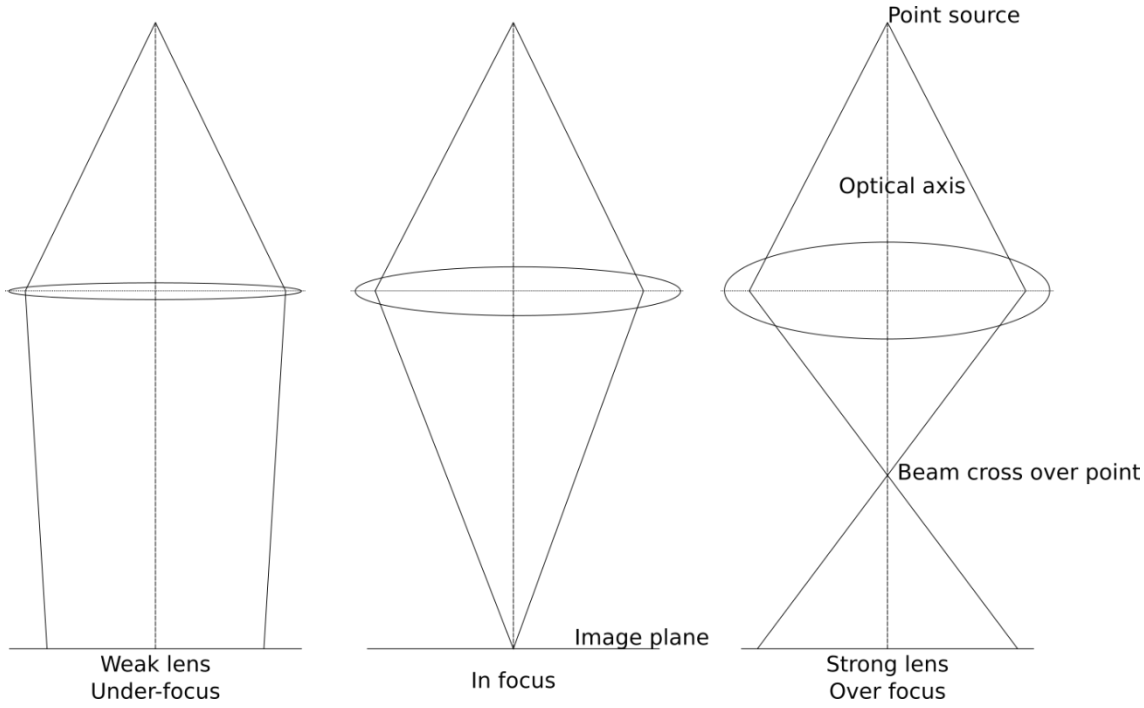
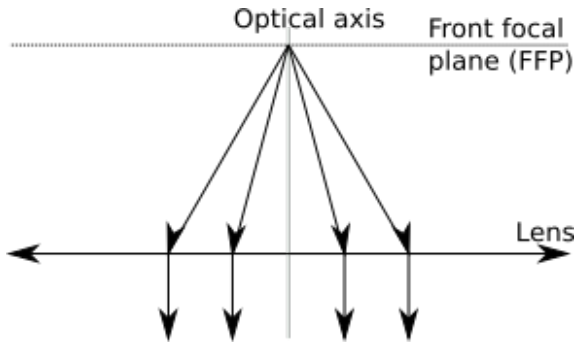


Figure 1-17: Position of beam cross over points at different foci

Lenses can be drawn as a line with arrows at each end, or an oval. Both are used in this work. In this figure, they are drawn as ovals to show differences in strength. A flatter oval represents a weaker lens; a fatter oval represents a stronger lens. When the lens is weak, the beams do not cross over at the image plane, but would cross over at a point behind it. At focus, the beams cross over at the image plane. This is the only place at which all the beams constructively interfere with each other. If the lens is too strong, the beam will cross over above the image plane. N.B. this figure is only to illustrate the positions of the cross over point. None of it is to scale.

Each lens has a front focal plane (FFP) and back focal plane (BFP). Some of the rays that hit the lens will exit the lens at the same angle. The plane before the lens at which these beams intersect is the FFP (Figure 1-18A). The BFP is the counterpart to the FFP. It is the plane after a lens at which rays that left the object and entered the lens at the same angle intersect (Figure 1-18B). The pattern of ray intersections on the BFP is called a diffraction pattern (DP). If the specimen is a crystal, then it is composed of multiple copies of the same thing (a unit cell) in the same orientation to form a lattice. The electron beam will be scattered in the same way by each unit cell. All of the beams that are scattered at a particular angle will intersect at the same point at the BFP.

A) Front focal plane (FFP)



B) Back focal plane (BFP)

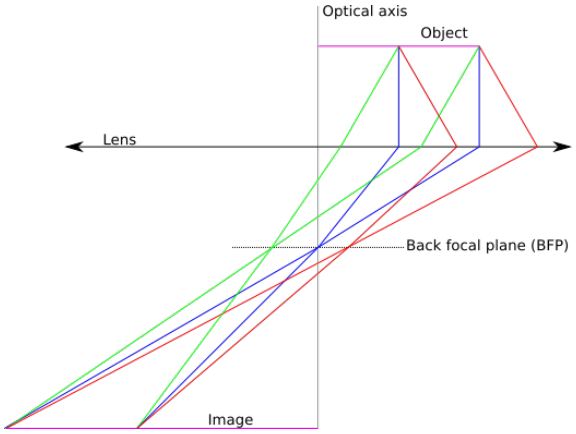


Figure 1-18: Definitions of focal planes

A) The front focal plane (FFP). The FFP is the point at which beams cross over that will exit the subsequent lens at the same angle. B) The back focal plane (BFP) is the plane at which beams cross over that entered the previous lens at the same angle. Three different coloured rays are shown from two points on the object. The green rays leave different points of the object and enter the lens at the same angle, the lens bends them to different degrees. They intersect at the BFP. The BFP shows a diffraction pattern (DP). This is the Fourier Transform of the specimen. The image plane is the plane at which all the rays from the same point in the object intersect to form an image.

1.4.1.1. Fourier transform

After the electron wave has been scattered by the sample, it will hit a detector at the bottom of the microscope. Each pixel in the detector can be on or off in order to make an image of the sample. This produces a digitised signal. Thus, there is a hard limit for the resolution of the image dictated by the pixels in the detector. The 2D image produced by the detector is processed using the Fourier Transform (FT). The FT is a way of decomposing a complicated signal into a set of composite periodic sin waves. The digitised signal is a series of numbers. The formula for the discrete Fourier Transform is given below.

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi \frac{k}{N} n} \tag{31}$$

where x_n is the function of the original signal, X_k is the Fourier Transform, N is the total number of samples, n is the sample number, and k is the coresponding signal number. This can be rewritten using Euler's formula:

$$X_k = \sum_{n=0}^{N-1} x_n . cos\left(\frac{-2\pi kn}{N}\right) + i . x_n . sin\left(\frac{-2\pi kn}{N}\right) \tag{32}$$

The term containing the cos wave represents the real components. The term containing the sin wave represents the imaginary components. The values of k, N and n are integers. The value of X_k is the sum of the formula for values of n between n=0 and n=N-1. The original signal is a series of numbers (x_0 x_1 … x_{N-1}) and the Fourier Transform is another series of numbers (X_0 X_1 X_2 … X_{N-1}). The values of k are

from k=0 to k=N-1. Some combinations of k and n will produce the same result for X_k. If it is known which combinations will produce the same result, then the calculation only needs to be done once. This is a fast Fourier Transform (FFT) and is far more efficient. As the value of k increases, the frequency of the cos and sin waves within the period of the original x_n function increases, e.g. when k=1, the cos and sin waves complete one wavelength and when k=2, they complete two wavelengths. However, once k reaches a certain value, the frequency of the cos and sin waves starts to decrease. This is the Niquist frequency and is equal to half of the sampling frequency or the number of samples taken (N). For a detector, N is the number of pixels.

The FT can also be performed on a 2D image, such as that produced by the detector in a microscope, to generate a set of 2D sin waves (Fourier components) that sum together to make the original function. The FT can be used to draw a power spectrum of that image. A power spectrum is a plot in reciprocal space. The axes are the Miller indices (the number of oscillations across the image in the y and x direction). Points close to the origin have few wavelengths across the image (low frequency), points far away have more wavelengths across the image (high frequency). Each Fourier component is plotted on the graph. The power of each wave is the amplitude squared. The power can be shown by the brightness of each dot. This graph will show a series of rings, called Thon rings (Figure 1-24 A).

1.4.1.2. Scattering

When the electron hits an atom, it will be diffracted by the electron cloud (Figure 1-19) (D. B. Williams and Carter 2009). If the electron misses the atom completely it will be unscattered and continue moving in the same direction with the same energy. If the electron beam interacts with the edge of the electron cloud, it will be scattered at a low angle (typically less than 3°). If it interacts with the cloud closer to the nucleus, then it will be scattered at a higher angle. It can also be scattered more than 90° and travel back towards the electron gun. This is a backscattered electron. In these cases, the electron retains all of its energy. These are elastically scattered electrons.

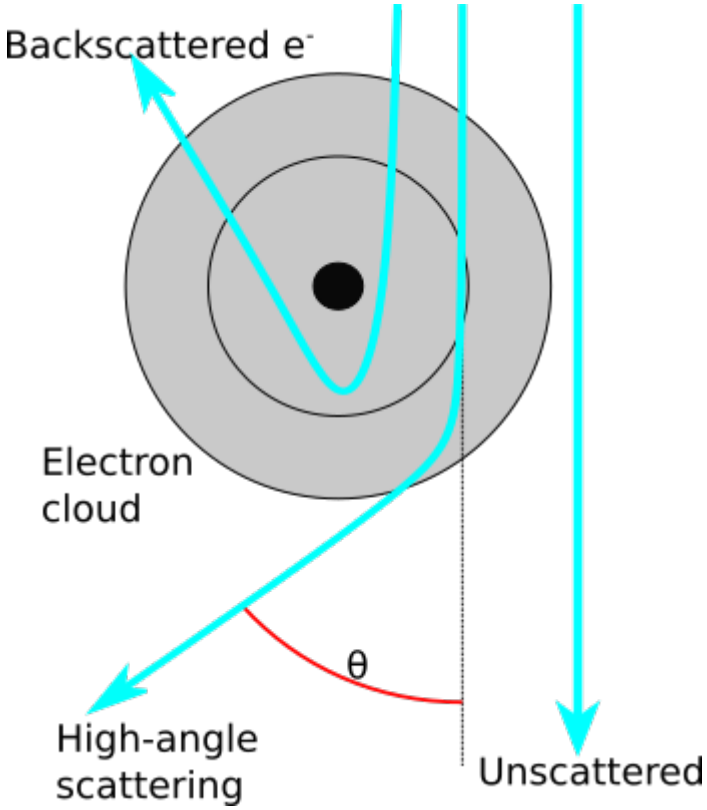


Figure 1-19: Un-scattered and elastically scattered electrons

Four types of electron scattering with an atom. If the electron misses the electron cloud of the atom, it is not scattered at all. If the electron interacts with the edge of the electron cloud, it is scattered at a low angle. Low angle scattering is defined as θ < 3°. If the electron interacts with the inner core of the electron cloud or the nucleus, it is scattered at a high angle. If θ > 90°, it is defined as a backscattered electron and can move back towards the electron source.

An electron can lose energy when it encounters an atom. This is called cathodoluminescence (CL). The incident electron hits an electron in the atom and promotes it to the next electron shell, leaving an electron hole in the original shell. The incident electron leaves with less energy. The electron will eventually drop back down to its original shell. It will lose energy as a light, e.g. an X-ray (Figure 1-20). Electrons that lose energy to the sample are called inelastically scattered electrons.

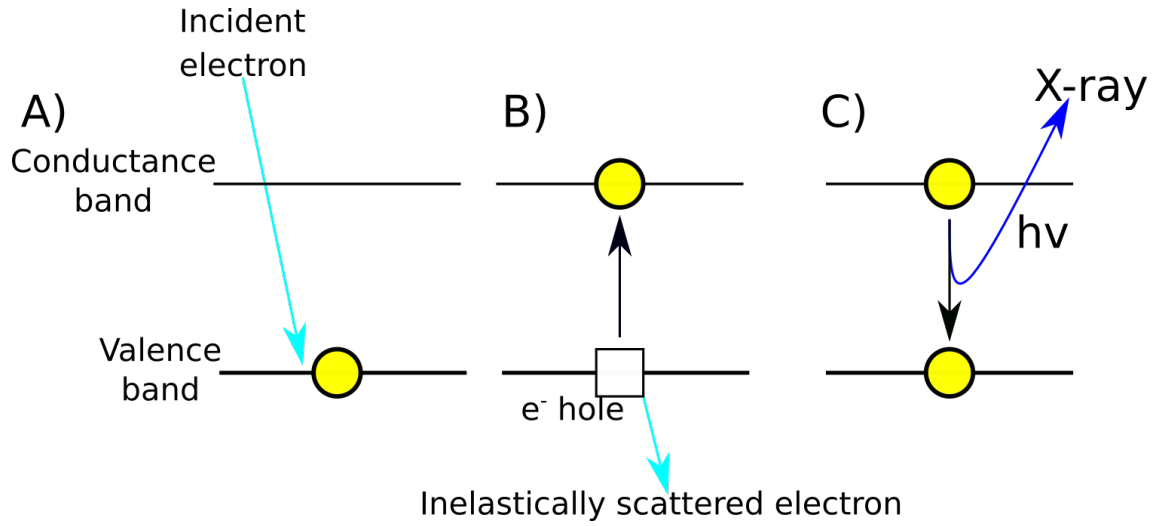


Figure 1-20: Cathodoluminescence (CL)

A) An incident electron encounters the valence band of an atom. B) An electron in the valence band gains energy and moves up to the conductance band. It leaves an electron hole in the valence band. The incident electron leaves the atom with less energy. C) Later, the electron moves back down from the conductance band to the valence band. The energy leaves as an X-ray. (D. B. Williams and Carter 2009).

The plane at which the electron beams are focused depends on the energy of the electron entering the lens. Since inelastically scattered electrons (see above) have less energy than elastically scattered electrons, they will be focused more strongly, and form an image at a plane above that of the elastically scattered electrons.

1.4.1.3. Contrast

There are two types of contrast: amplitude contrast and phase contrast. In amplitude contrast, electron beams hit a dense object in the specimen, e.g. uranyl acetate. They are scattered in many directions. Some of the electron beams are lost, e.g. they hit the microscope wall, hit the aperture etc. As a result, fewer electrons that were scattered by that point in the specimen hit the detector. Electron beams that did not hit the highly deflecting atoms are scattered less or totally unscattered, so more are focused onto the corresponding point on the detector. TEM primarily uses amplitude contrast.

Each atom in the sample is a scattering centre which will cause the electron to be scattered in all different directions. For a pair of scattering centres that are a set distance (d) apart, the electron waves that are scattered at a certain angle will be in phase with each other. In Figure 1-21, the wave from the left scattering centre had to travel an extra distance (L) before the start of the wave from the next scattering centre. If this distance is a multiple of 2π radians (or one wavelength), then the waves will be in phase with each other. Since they will hit the lens at the same angle, they will be focused onto the same point in the BFP of the lens (Figure 1-18B). The angle of scattering at which the waves will be in phase depends on the distance between the scattering centres. The closer the scattering centres are together, the higher the angle from the path of the incident electron. Thus, scattered waves from

two separate proteins (low resolution information) will be in phase at a low angle. Meanwhile, scattered waves from two atoms next to each other (high resolution) will be in phase at a very high angle. Waves that are more highly scattered will appear further away from the centre in the diffraction pattern at the BFP.

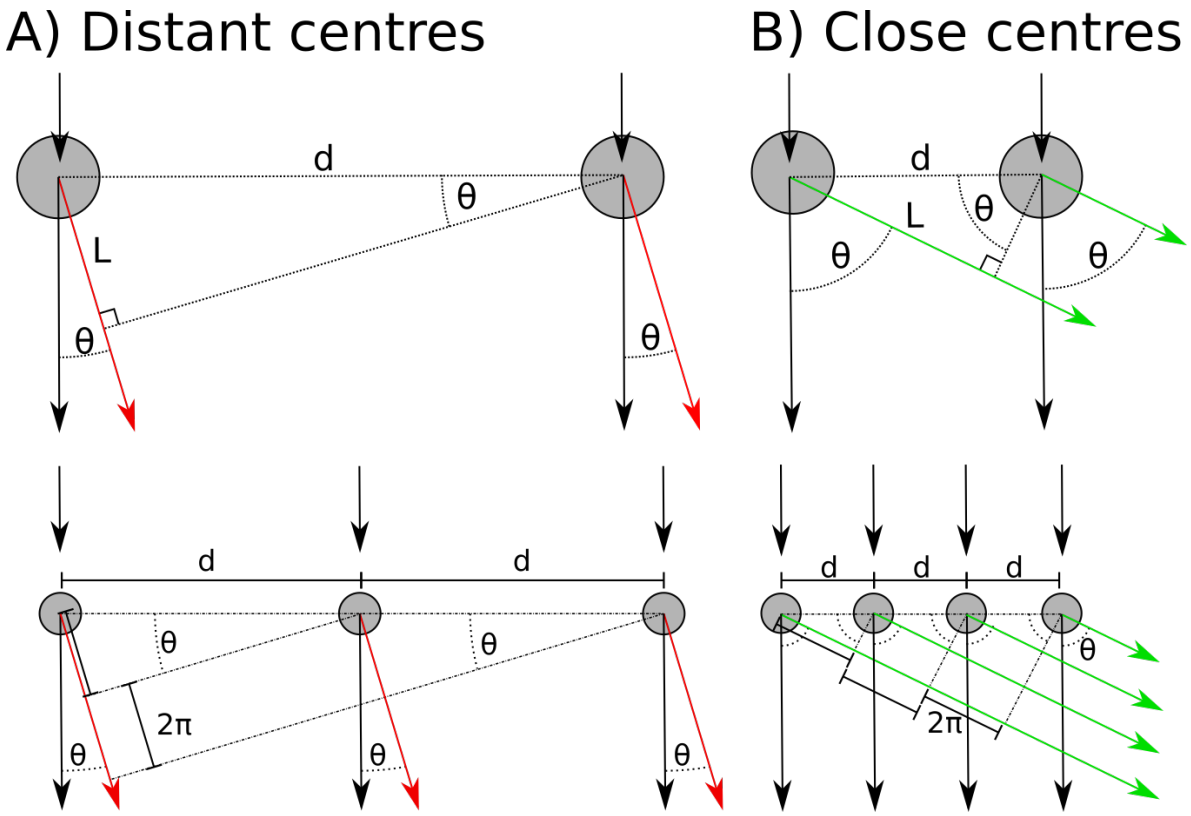


Figure 1-21: Scattering centres

The electron wave hits all the scattering centres in the specimen at once. Electron waves are scattered in all directions by each scattering centre. But only those waves that are in phase with each other will interfere constructively. A) Low-angle scattering from distant scattering centres. Top – schematic showing the distance between scattering centres (d) and the extra distance travelled by beams (L). Bottom – Example of electron waves from different scattering centres that will interfere constructively. After the wave travels 2π (one wavelength), the scattering centre from which a wave of the same angle of scattering will emerge is quite far away. B) High-angle scattering from close scattering centres. Bottom – after the wave travels 2π (one wavelength), the scattering centre from which a wave of the same angle will emerge is very close.

A wave can be drawn as a vector on an Argand diagram with real (Re) and imaginary (Im) axis (Figure 1-22). The length of the vector is the amplitude (ψ). The phase of the wave is the angle of the vector (θ). If the wave is phase shifted, the vector rotates around the origin.

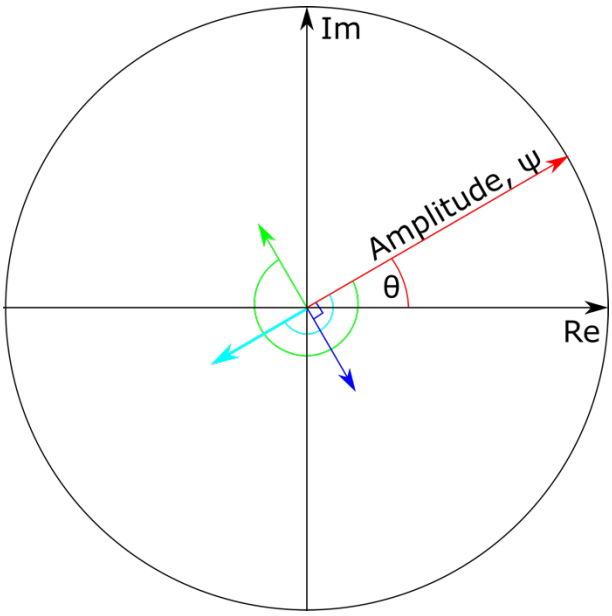


Figure 1-22: Phase circle

The amplitude (Ψ) is the length of the vector. The phase is the angle (θ). The blue vector is a wave that is phase shifted 90º (π/2 radians). The cyan vector is phase shifted by 180º (π radians) relative to the red vector. The green curve has been phase shifted 270º (3π/2 radians) relative to the red vector.

At the detector, the unscattered wave, and waves scattered by the sample will interfere with each other. Wave interference can be drawn on an Argand diagram (Figure 1-23). If the waves are in phase, the amplitude of the cumulative wave will be much larger. If the waves are 180° out of phase, the amplitude of the cumulative wave will be much smaller. If the waves are 90° out of phase, then the cumulative wave will be very similar to the original un-scattered wave, so the scattered wave will make very little difference and it will not be visible in the final image. The difference in phase between the un-scattered and scattered waves means that some will be represented more strongly in the resultant image than others. This is phase contrast. The spatial frequencies at which the contrast transfer function crosses 0, are lost information and cannot be recovered.

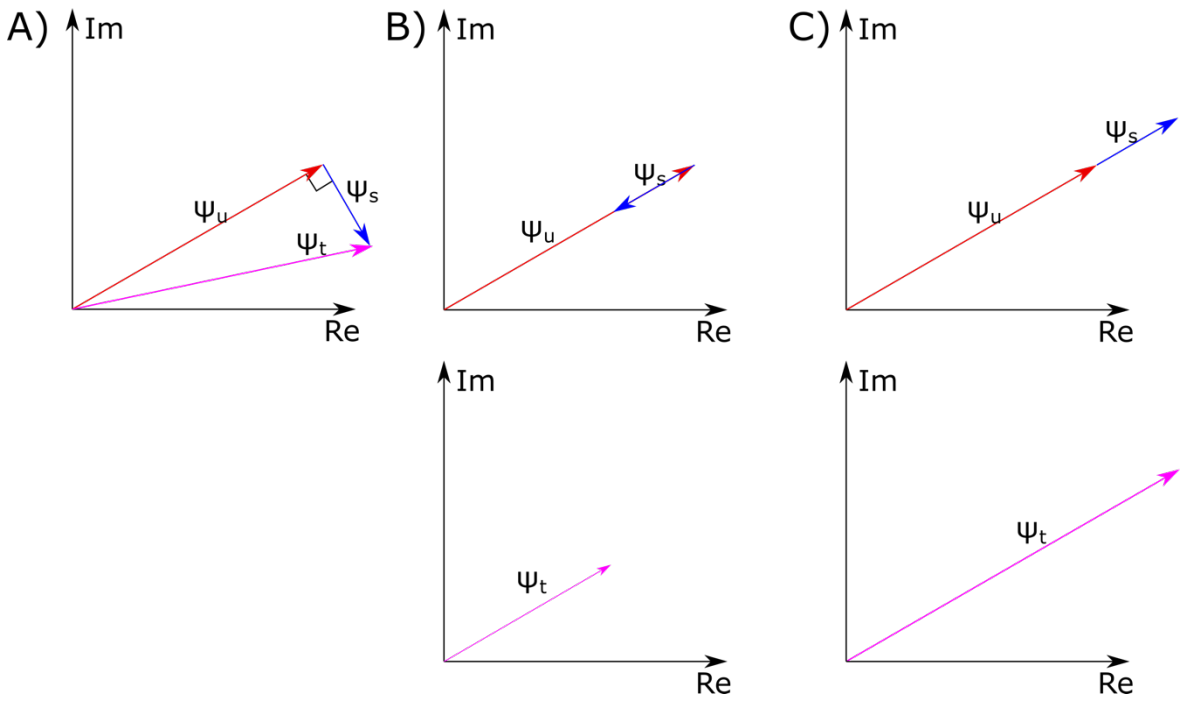


Figure 1-23: Wave interference

Ψ_U (red vector) is the amplitude of the unscattered wave. Ψ_s (blue vector) is the amplitude of the scattered wave. Ψ_t (purple vector) is the amplitude of the sum of the unscattered and scattered waves. A) The Ψ_s is phase shifted 90° relative to Ψ_U. Ψ_t is very close in amplitude to Ψ_u. If the Ψ_s has a very low amplitude, then Ψ_t will be closer to Ψ_u. Thus, waves that are phase shifted 90° relative to Ψ_U will have very little effect on Ψ_t. B) Ψ_s is phase shifted 180° relative to Ψ_u. The amplitude of Ψ_t is smaller than Ψ_u, so Ψ_s will have a large negative contribution to the image. C) Ψ_s is phase shifted 360° relative to Ψ_u. Ψ_t is much larger than Ψ_u, so Ψ_s will make a large positive contribution to the image

1.4.1.1. Contrast transfer function (CTF)

The power spectrum is a FT of the 2D image that shows the amplitude of the Fourier components as increased brightness. Components that contribute highly make up the bright spots that form the rings (Figure 1-23B-C). Components that do not contribute (Figure 1-23A) are the dark spots. The Thon rings show which spatial frequencies are represented in the image. If the strength of the objective lens is such that the focus point is on the image plane, then the image is said to be at focus. If the objective lens is weakened, then the beams will focus below the image plane. If the defocus changes, the path lengths of the scattered waves will change as well. This means that the phase shifts of the waves will change as well. Scattered waves that previously contributed nothing or negatively to the image, because they were 90° or 180° out of phase with the un-scattered wave respectively, may now contribute to the final image. Because waves of different spatial frequencies are contributing different amounts to the final image, the power spectrum (Thon rings) and CTF graph will also change. The CTF is found using the CTF equation:

$$Contrast\ transfer\ function\ (CTF) = sin\left(-\pi.\Delta z.k^2 + \frac{\pi.C_s.\lambda^3.k^4}{2}\right) \quad (33)$$

where Δz is the defocus, k is the special frequency, C_s is the spherical aberration, λ is the wavelength of the electron wave (not the wavelength of the Fourier component). C_s will depend on the microscope. λ is a constant. At low defocus, high spatial frequency waves that were highly scattered from scattering centres close together in the specimen (high resolution information) contribute more to the image, and low spatial frequencies contribute less. At higher defocus, the opposite is true. It is common practice to take images of different areas at different defocus values in order to capture low- and high-resolution information, especially in cryoEM which can reach higher resolution images than TEM. The recordings start at focus, then proceed to higher defocus, as the electron beam will damage the specimen during collection, so the high-resolution information may be lost at later stages of collection. Recording at different defocus values also captures information that is lost in other collections when the CTF crosses 0.

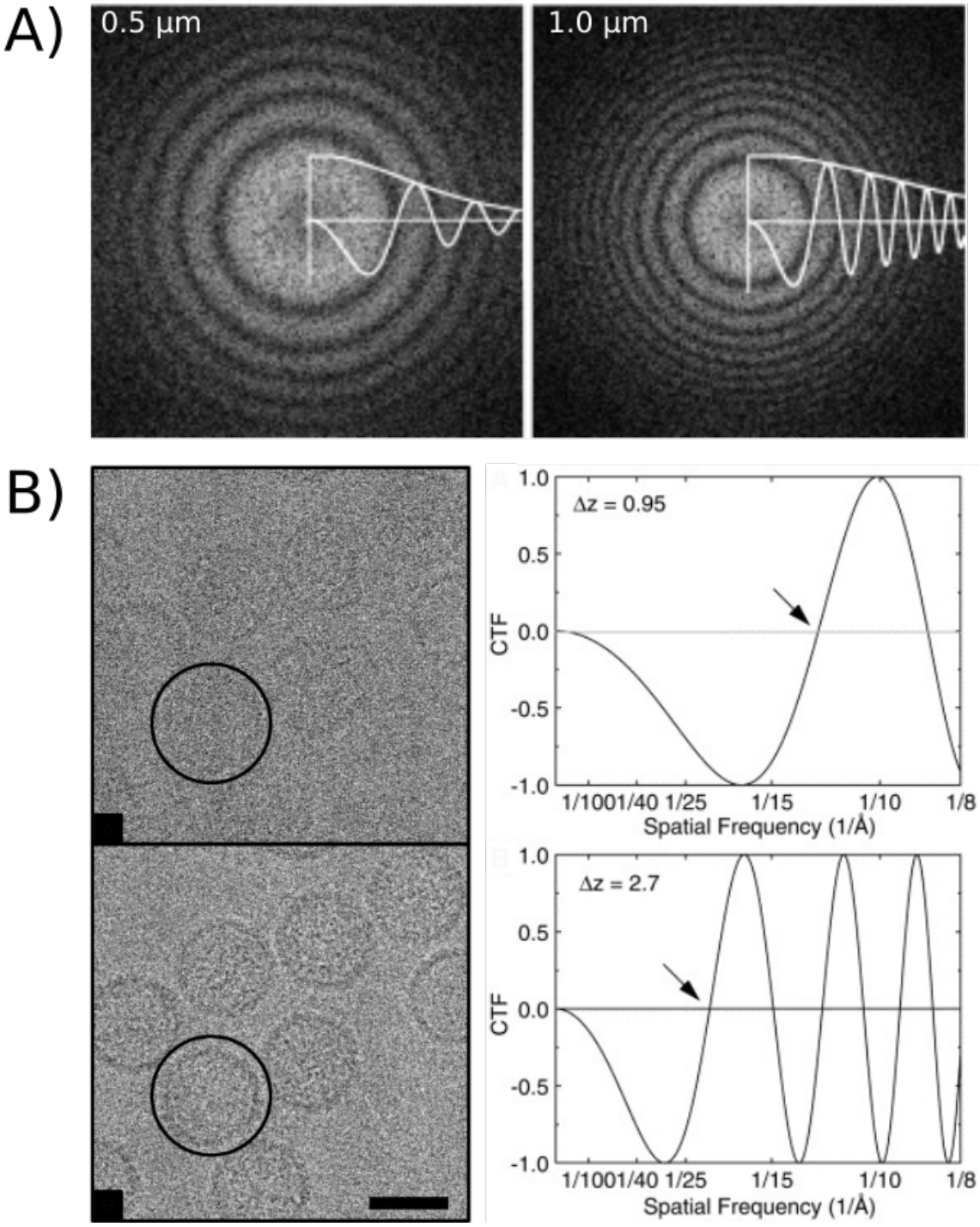


Figure 1-24: CTF, Thon rings, defocus

A) Power spectra (Thon rings) with CTF graph overlaid. The spectra are for images at different defocus values. Adapted from (Orlova and Saibil 2011). B) Micrographs of P22 procapsids and their corresponding CTF curves at different defocus values. Adapted from (Thuman-Commike and Chiu 2000).

1.4.2. Transmission electron microscopy (TEM)

TEM is a workhorse technique for structural biology to quickly check the state of samples.

1.4.2.1. Sample preparation

The sample is applied to a mesh grid. A heavy metal stain is applied. The stain surrounds the particles. If the stain is too thick, it will cover the sample and damage it. It should ideally form a thin layer around the sample (Herzik 2021). Common heavy metal stains are uranyl acetate (UA) and uranyl formate (UF). The grain of the stain will determine the resolution possible. UF has a smaller grain, so allows for a better resolution, but it has a much shorter shelf life than UA. The stain will dehydrate the sample, which damages it. When the beam passes through the stain it will be scattered at a high angle (Figure 1-19) and will lose amplitude and phase. This signal is compared to the beam that passes through the sample itself which will have lost less amplitude and phase (Dubochet et al. 1988). The sample appears white, while the background appears black, hence the name negative stain. The electron beam can cause the stain to move around the sample; using a lower electron dose may prevent this movement and give a higher resolution image (Unwin 1974). However, the stain crystals that form limit the resolution of a 3D map from a negative stain sample to 20 Å (Ohi et al. 2004).

1.4.2.2. Image formation

In negative stain, the heavy metal stain will greatly scatter electrons. Electrons that hit areas of the specimen with stain will be heavily scattered, so fewer will make it to the image plane to interfere with each other and form an image. The sample itself will scatter electrons to a lesser extent, so more of them will make it to the image plane. The amplitude of the electron waves that hit the stain has been reduced. This is amplitude contrast.

1.4.3. Cryo-electron microscopy (cryoEM)

Cryo-electron microscopy (cryoEM) is a technique that has been developed since the 1990s for imaging biomolecules in a soluble environment by rapidly freezing the surrounding buffer. In 2017, Jacques Dubochet, Joachim Frank and Richard Henderson were awarded the Nobel prize in chemistry for co-creating cryoEM.

1.4.3.1. Sample preparation

The sample is applied to a grid with nanometre scale holes. The holes allow a thin layer of liquid to form. The excess liquid is removed by blotting. This leaves only the sample inside the thin layer of liquid inside the holes (Figure 1-28). The grid is rapidly frozen in liquid ethane cooled by liquid nitrogen. This technique is called plunge freezing. The solution freezes so fast that it forms vitreous ice, which does not show any internal structure, so it appears like liquid water under a microscope (Dubochet et

al. 1988). Electrons diffract through crystalline ice, which would form if it was frozen more slowly. Crystalline ice would also damage the sample.

1.4.3.2. Image formation

The particles in the sample will be frozen in a range of orientations. The electron beam will be scattered differently by the sample depending on the particle's orientation. Unlike negative stain, in cryoEM the electrons that exit the sample at a low scattering angle are of importance for image formation. For hydrated unstained samples, such as in cryoEM, most of the signal comes from phase contrast (Dubochet et al. 1988).

However, cryoEM samples do not have something that will scatter the electrons as heavily as the heavy metal stain. The light weight atoms in biological samples do not absorb electrons. Most of the contrast in cryoEM samples comes from phase contrast. In this case, the electrons behave more like a wave. When the electron wave hits the specimen, the phase and amplitude of the wave will change. The wave will be scattered in all different directions. The scattered waves will have a longer journey from the specimen, through the lens, and onto the image plane than unscattered waves. This increased path length can also introduce phase shift. Some of the scattered waves will be in phase with each other, if they exit the specimen at the same angle.

Images of the same area are taken at a range of defocus values, starting with high defocus (Figure 1-17), then getting closer to focus. At high defocus, more high resolution will be collected. At lower defocus, low resolution information will be collected. The images must be taken in reverse order of defocus, as the beam will damage the sample during imaging, destroying high resolution information.

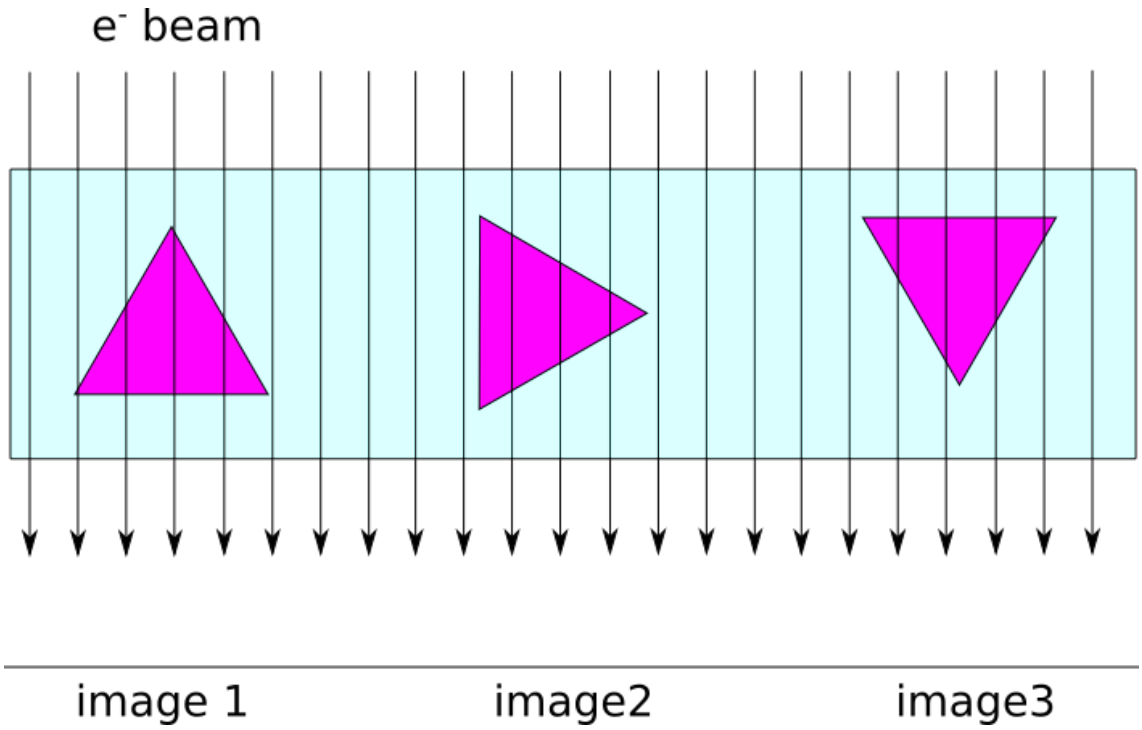


Figure 1-25: Electron beam hitting particles in different orientations

The purple triangles are individual particles in different orientations in the ice layer (blue rectangle). Each orientation will produce a different image on the image plane.

1.4.3.3. Data processing

Once the micrographs have been collected, the data must be processed to make a density map. There are a variety of different software programmes available to do this. The most common are RELION (Scheres 2012), EMAN2 (G. Tang et al. 2007), and cryoSPARC (Punjani et al. 2017). RELION is used in this work. The typical workflow for SPA is shown in Figure 1-26.

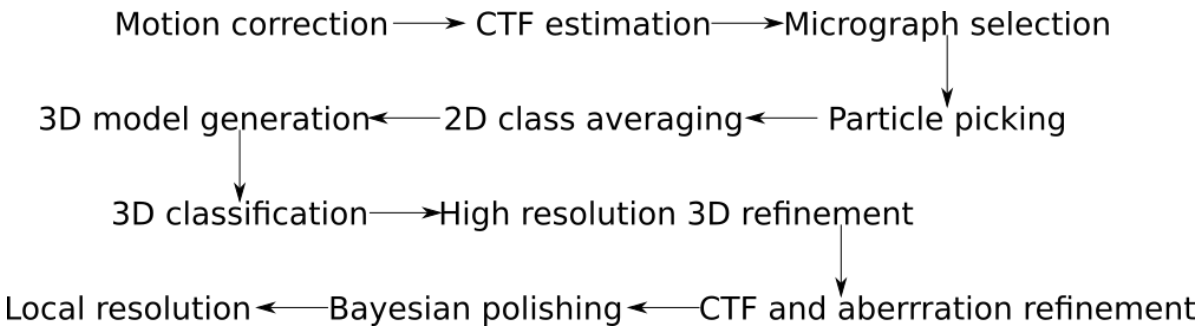


Figure 1-26: Typical workflow for single particle analysis (SPA)

The Particle Picking and 2D Class Averaging steps may be repeated several times in a cycle to remove artefacts, denatured particles or contaminants. The CTF refinement and polishing steps may also be repeated to increase the resolution of the final map, though gains will fall off.

(a) Pre-processing

Before particle images can be used to make models, the limitations of cryoEM must be considered and corrected. During imaging the electron beam will cause the ice to bulge and flow slightly within the hole (Wright et al. 2006). The beam can also cause the ice layer to bulge upwards, which causes the specimens to move. This is called beam-induced specimen movement. Including some of the carbon grid can reduce beam-induced specimen movement. A direct detector allows many sub-frames to be taken of the same area in quick succession to make "movies". By tracking the movement of a single particle through the movies, the final images can be motion-corrected. The sample will become more damaged over time from the radiation. The first frames have the least radiation damage. (Zheng et al. 2017). As mentioned above, not all of the Fourier components that make up the image are represented with equal strength in the image. This is shown by the CTF curve (Figure 1-24). The CTF for the dataset must be estimated, e.g. with a programme like CTFFIND4.1 (Rohou and Grigorieff 2015).

(b) Particle picking

The image frames will contain many particles in the ice. These particles must be picked out in a box to be further processed. Particles can be picked manually by the user, or there are different autopicking algorithms for picking particles from the images. These algorithms usually require a template of what particles to pick. The template can be from the user manually picking a small number of representative particles. This work uses the Topaz autopicker (Bepler et al. 2020; 2019).

(c) Alignment and 2D classes

RELION uses a k-means clustering and maximum likelihood method for aligning particles and making 2D classes. Other methods are used by other software packages (Orlova and Saibil 2011). A number of random particle images (k number of particles) are taken as seeds. The other particle images are aligned to them. The cross-correlation coefficient (CC) is found between each pixel in each orientation of the image and each pixel of each seed image. The angle between orientations of the test image is chosen by the user, e.g. 6° (Sigworth 1998; Scheres et al. 2005). The probability of the image being in that alignment with that seed image is calculated. Once this has been performed for each image, the images are averaged for each seed image. The probability of each image being in that class gives weight to the average. Images that have a low cross correlation with the seed have a low probability of being in that class, so they have a low weight when calculating the average. Images that have a high cross correlation, have a high probability, so have a higher weight. This is performed iteratively. Each time, the 2D class averages become clearer and more defined. The 2D class averages have a higher

signal-to-noise ratio than the individual component images. Different views of the particle should form different 2D classes. Noise or denatured particles should also form their own 'junk' classes and can be removed.

(d) Initial model

Once the 2D classes have been made, if they are of sufficient quality, they can be used to make an initial model. The FT of a 2D image of a 3D object, like the 2D classes of the particle are a central slice through the 3D FT of the object (Figure 1-27A). This is called the projection-slice theorem. The FTs of the 2D class averages are aligned to each other in Fourier space. Each FT is aligned with another by finding the common line that they share. Once all of the 2D class FTs have been fitted, it is Fourier Transformed again to make a 3D object. This serves as the initial model. If the specimen shows preferential orientation, or the classes are of poor quality, then there may not be enough views to make an initial model. In this case, a low-pass filtered map of a similar sample may be used.

(e) 3D classification

To make 3D classes, re-projections are made of the initial model. These are 2D classes of the initial model at a variety of orientations. RELION uses maximum likelihood methods for this as well. The probability of each particle aligning to a model re-projection is calculated. New 2D classes are made using the weighted probability of each image around the re-projections. These are used to make a new 3D model as above. This process can be repeated to give a better 3D model (Figure 1-27).

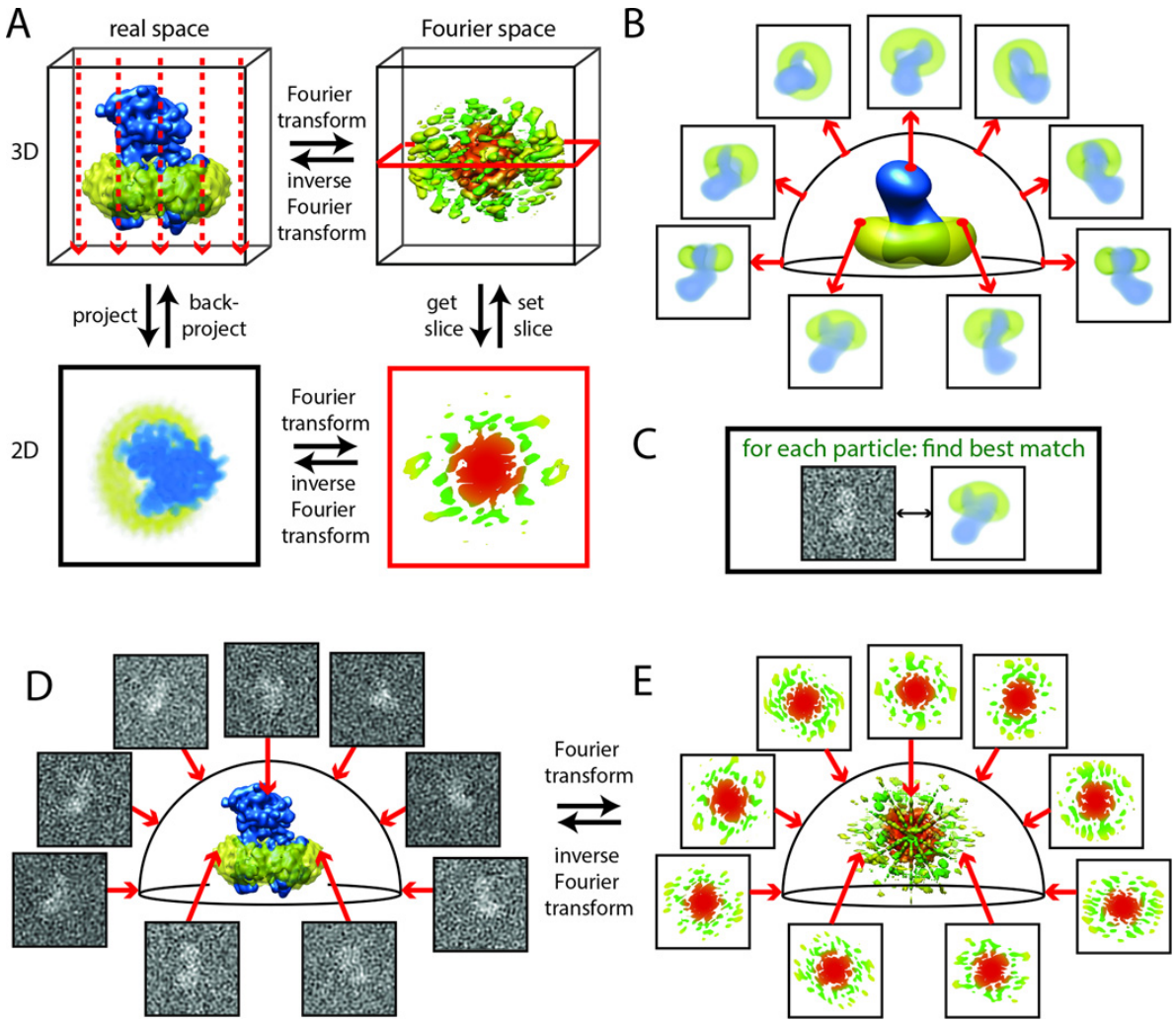


Figure 1-27: 3D reconstruction

Different stages of 3D reconstruction. A) The electron beam forms a projection image of the object in that orientation on the image plane. The projection theorem states that a Fourier Transform of the projection (red outlined image) is a slice through the Fourier Transform of the original object. B) Back projections of the initial model. These form new 2D classes. C) The particles are matched to the new 2D classes. D) The particles in the new 2D classes are used to make a new 3D model as in (A). E) FT of the 2D images to make a new 3D model. Adapted from (Nogales and Scheres 2015).

The resolution of the final map can be assessed with a Fourier Shell Correlation (FSC). The data is split into two halves. Each half is used to make a 3D model independently. The FT of each 3D model is calculated. 3D spherical regions of each FT (called shells) are compared to each other pixel-by-pixel. At low spatial frequencies (smaller shells), that correspond to lower resolution information, the maps will align well. As spatial frequency increases, the maps will start to diverge. Eventually, their alignment will not be better than random noise. The standard value for this is when the FSC is 0.143.

1.4.3.4. Common problems

A common problem in cryoEM is adsorption of the particles to the air-water interface (AWI) (Figure 1-28). In standard buffers, most particles will migrate towards and adsorb to the AWI (Kang et al. 2023;

D'Imprima et al. 2019). Upon contact with the interface, they can denature. The particles may adsorb to the AWI in a particular orientation if it contains a hydrophobic region. The thickness of the ice layer can also affect particle orientation (Noble et al. 2018; D'Imprima et al. 2019). The blotting time and time before plunge freezing can affect ice thickness. The ice layer tends to be thicker at the edges of the hole and thinner towards the centre. If most of the particles in a sample are in the same orientation, then only a single view will be seen and the data cannot be processed further. This is a common problem in cryoEM sample preparation known as preferential orientation.

Surfactants can reduce the surface tension and reduce the proportion of particles adsorbed to the AWI (Kang et al. 2023). Adding octyl-β-glucoside before freezing has been shown to help prevent preferential orientation (Sun et al. 2015a). Olek *et al* have recently released a programme to estimate and group particles by ice thickness, which can then be grouped separately (Olek et al. 2022).

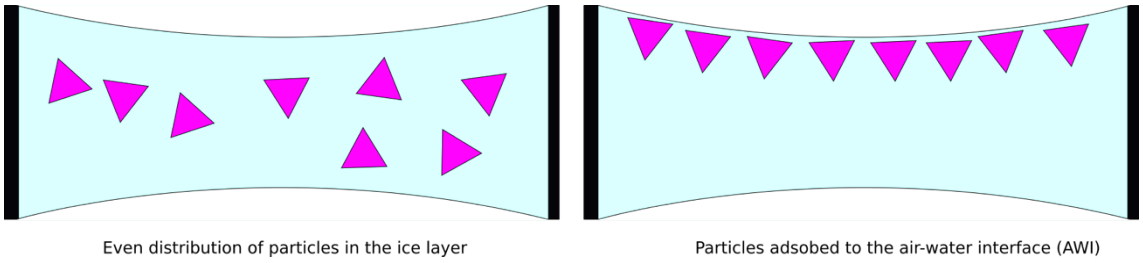


Figure 1-28: Particles and the air-water interface (AWI)

Black – carbon grid bars. Purple – particles. Blue – ice layer. Left – ideal situation. The particles are evenly distributed throughout the thin ice layer. Right – more common situation. The particles have moved to and adsorbed to the AWI. Once there, they can denature. If the particle has a particularly hydrophobic region, then that region may be what adsorbs to the AWI, which would give the sample preferential orientation as well.

1.4.4. Nano Differential Scanning Fluorimetry (nanoDSF)

In NanoDSF, a sample is heated from 20 to 95 ℃. The fluorescence at 330 nm and 350 nm is measured. Proteins contain tryptophan residues. Tryptophans and tyrosines absorb at 280 nm. In folded proteins, they emit at 330 nm. When the protein unfolds they are exposed to water and emit at 350 nm. As the protein unfolds, the ratio of 350 nm/330 nm increases. The ratio can be plotted against temperature. If the protein completely unfolds, the graph of 350 nm/330 nm vs temperature will form a sigmoid curve. The region before the gradient increases is the protein in the folded state. The region after the gradient is the protein in the unfolded state. The temperature at which the ratio starts to increase is the onset of unfolding (T_{on}). The melting temperature (T_M) is the temperature at which the gradient is steepest, and at which half of the protein is unfolded and half folded. The T_M can be found by finding the first differential of the fluorescence ratio. The backreflection scattering is a measure of protein

aggregation. When proteins unfold, hydrophobic regions may be exposed. It may be energetically favourable for these regions to aggregate together ('nanoDSF', n.d.). Performing nanoDSF on different proteins can show which are more thermally stable. It can also be used to screen different conditions for those that keep the protein folded or prevent aggregation. Equation (34) is the equation for the observed signal when moving between a native and denatured state (Tim Sharpe 2012):

$$Observed\ signal = \frac{(\alpha_N + \beta_N.x) + (\alpha_D + \beta_D.x).e^{-\left(\frac{\Delta G_{D-N}}{R.T}\right)}}{1 + e^{-\left(\frac{\Delta G_{D-N}}{R.T}\right)}} \quad (34)$$

In this case, the observed signal is fluorescence ratio. The variable x is the degree of perturbation. For thermal denaturation, x is temperature. For chemical denaturant, e.g. with urea, x is the concentration of urea. The regions of the curve corresponding to the folded (native) and unfolded (denatured) states can be approximated as straight lines. The parameters α_N and β_N are the y-intercept and gradient of the line for the folded state, respectively. The parameters α_D and β_D are the same for the unfolded (denatured) state. T is the temperature in Kelvin. R is the ideal gas constant (1.987 cal mol^{-1} K^{-1}). The ΔG_{D-N} is the free energy required to unfold the protein from the native to the denatured state. For thermal denaturation, ΔG_{D-N} is given by the modified Gibbs-Helmhotz equation (Robertson and Murphy 1997):

$$\Delta G_{D-N} = \Delta H_m \left(1 - \frac{T}{T_m}\right) + \Delta C_p \left[T - T_m - \left(T.ln\frac{T}{T_m}\right)\right] \quad (35)$$

where ΔH_M is the enthalpy of denaturation at transition midpoint, ΔC_P is the change of heat capacity of denaturation (specific heat of unfolding), T is temperature, and T_M is the midpoint of thermal denaturation (the temperature at which half of the protein is unfolded). Substituting equation (35) into equation (34) gives equation (36):

$$Observed\ signal = \frac{(\alpha_N + \beta_N.T) + (\alpha_D + \beta_D.T).e^{-\left(\frac{\Delta H_m\left(1-\frac{T}{Tm}\right)+\Delta C_p\left[T-T_m-\left(T.ln\frac{T}{Tm}\right)\right]}{R.T}\right)}}{1 + e^{-\left(\frac{\Delta H_m\left(1-\frac{T}{Tm}\right)+\Delta C_p\left[T-T_m-\left(T.ln\frac{T}{Tm}\right)\right]}{R.T}\right)}} \quad (36)$$

The x in the equation (34) has been replaced by T as the temperature is the denaturant. The curve was fitted to the points of fluorescence ratio (A350/A330) vs temperature.

For chemical denaturation, ΔG_{D-N} is given by equation (37):

$$\Delta G_{D-N} = m.(x_{50\%} - x) \quad (37)$$

The $x_{50\%}$ is the concentration of urea at the midpoint of unfolding, when half of the protein is unfolded. The parameter m is the steepest gradient of the sigmoid curve. Substituting equation (37) into equation (34) gives equation (38):

$$Observed\ signal = \frac{(\alpha_N + \beta_N.x) + (\alpha_D + \beta_D.x).e^{-\left(\frac{m.(x_{50\%} - x)}{R.T}\right)}}{1 + e^{-\left(\frac{m.(x_{50\%} - x)}{R.T}\right)}} \quad (38)$$

The variable x is the concentration of chemical denaturant, e.g. urea, guanidinium chloride. m is the steepest gradient of the sigmoid curve. Previously, nanoDSF has been used to screen enzymes and their solvent conditions for thermal stability. Binding parameters could be estimated from the T_m data (Magnusson et al. 2019). It has been used to assess thermal stability of proteins for therapeutics and the effects of pH (Lisina et al. 2023). It was found that replacing KCl with potassium glutamate (KGlu) helps stabilise protein-nucleic acid complexes (Cheng et al. 2016). It has been used in combination with AI to classify blood plasma of patients with glioblastoma (Eyraud et al. 2023). This could prove useful for portal proteins that are prone to aggregate or misfold.

1.4.5. Analytical ultracentrifugation (AUC)

Analytical ultracentrifugation (AUC) fills a niche similar to size exclusion chromatograph (SEC) (see Chapter 2) as a method of finding the size of proteins of interest. A key difference is the ability of AUC to analyse proteins under native conditions, rather than relying on the hydrodynamic radius of the protein. Oligomers may also fall apart during SEC, which would only allow it to be used to detect the presence of the protein, but not oligomers. AUC can be used to estimate the proportion of different oligomers or aggregates in a sample, e.g. in different buffer conditions, or different protein mutants. It can be used to follow the protein oligomer as it falls apart in real time under different denaturing conditions (Perez-Ramirez and Steckert 2005).

1.4.6. Circular dichronism (CD)

Circular dichronism (CD) uses the absorption of left- and right-handed polarised light by proteins of interest to detect quantities of different secondary structure elements. While it cannot assign positions within the protein to these elements, it can be used in detecting changes in secondary structure, e.g. by temperature, or denaturing agents (Nojima et al. 1977; Mehta et al. 2014). It can also be used to investigate changes in oligomerisation (Woodbury et al. 2022). It can be used to study the formation of the large C-terminal barrel domain of P22 portal protein which forms while the DNA is being packaged into the capsid (Lokareddy et al. 2017).

1.4.7. Charge-dependent mass spectrometry (CDMS)

Charge-dependent mass spectrometry (CDMS) is a variant of mass spectrometry which measures the charge of an ion as well as its mass. It can be used for larger ions, e.g. protein aggregates, whole viruses

(reviewed in (Jarrold 2022)). It has been used to determine the oligomeric state of WT and mutant portal proteins (Motwani et al. 2017; Woodbury et al. 2022).

1.5. Research Objectives

While ONT have used their CsgG mutant for DNA sequencing, other applications and types of nanopores are still being investigated. One of these areas is hybrid nanopores, which aims to combine the best features of biological and solid-state nanopores. However, since the biological component of the hybrid pore must sit in a silicon wafer, rather than a lipid membrane, most current biological nanopores may not be suitable. Thus, there is a need for new biological nanopores to fill this niche. As natural DNA translocating channels, portal proteins are good candidates for this role, especially those from thermophages. But there are few cryoEM structures available for thermophage portal proteins. In collaboration with DreamPore SAS, the aim of this work was to establish a new hybrid nanopore system based on thermophage portal proteins and SSNs made using CBD.

The objectives of this research project were:

1. To express and purify novel thermophage portal proteins – **Chapter 2**
2. To determine their cryoEM structures – **Chapter 3**
3. To characterise their thermal and chemical stability – **Chapter 4**
4. To test their applicability as biological and hybrid nanopores – **Chapter 5**

Chapter 2 Search for portal proteins

2.1. Introduction

There are many thermophages and many thermophage portal proteins (Łubkowska et al. 2021). For one of them to be suitable as a biological or hybrid nanopore, it should be easy to produce in large quantities, stable in a range of conditions and not disassemble or degrade over time. Eight genes were chosen as the first batch of thermophage portal proteins to be tested to see if they were suitable candidates: deep-sea thermophilic phage D6E, *Geobacillus* virus E2, *Thermus* phage phiFa, *Thermus* phage phiOH2, *Bacillus* phage 1, *Geobacillus* phage GBSV1, *Rhodothermus* phage RM378, and *Thermus* phage TSP4. They all belong to the *caudoviricetes* class of viruses, which are viruses with head-tail morphology (Schoch et al. 2020). All belong to unclassified genus, except for BV1, GBSV1, phiFa, and TSP4. BV1 and GBSV1 are the only members of the *svunavirus* genus (Kuhn, Adriaenssens, and Kropinski 2017). TSP4 and phiFa are members of the *oshimavirus* genus along with G20c, P23-45, P74-26 (Schoch et al. 2020). Under the previous classification system, D6E, BV1, GBSV1, RM378 and TSP4 were classed as *myoviridae*, while GVE2, phiFa, and phiOH2 were classed as *siphoviridae* (Turner et al. 2023). These terms are included as points since they are still in common usage, even though the classifications have been abolished. phiFa and TSP4 were the candidates with the most promise as their close relative G20c had previously been adapted as both a biological and hybrid nanopore (Cressiot et al. 2017a; 2018). The expression and purification of the genes served as a selection procedure for which portal proteins would be taken forwards for structure determination and nanopore testing.

2.1.1. Chosen thermophages

2.1.1.1. Deep-Sea Thermophilic Phage D6E

The bacteriophage Deep Sea thermophilic phage D6E was isolated from *Geobacillus* sp. strain E263 from a deep-sea vent off the coast of Mexico. It has an icosahedral capsid 60 nm in diameter. The tail is 16 nm wide and 60 nm long. Before the reclassification of virus families, it was classified as a member of the *myoviridae* family. The circular dsDNA genome has 49 ORF and is 49,335 bp long (Yiqian Wang and Zhang 2010). There has been some work to purify the large terminase as part of a study of thermophage large terminases, but it was not taken forwards (R. Xu 2017).

2.1.1.2. Geobacillus GVE2

Geobacillus GVE2 is one of the better studied thermophages. In later papers about GVE2, the authors refer to the paper about GVE1, so it is possible that GVE1 and GVE2 are the same phage. It was isolated from *Geobacillus sp*. E26323 from a deep-sea thermal vent in the Pacific Ocean. It was previously classified as a *siphovirus*. It has a hexagonal head that is 130 nm diameter and a tail that is 180 nm

long and 30 nm wide. It is most stable at 60C. The genome is 41 kbp long (B. Liu et al. 2006), made of linear dsDNA. It has 62 ORFs (B. Liu and Zhang 2008). Its portal protein was found to have similar predicted secondary structure to other common portals, e.g. HK97, but low sequence similarity (Yiqian Wang and Zhang 2008). GVE2 is a lytic phage. The system it uses to lyse its host cell may be different to its counterparts in mesophilic phages, as the endolysin is not exported by the sec machinery. The GVE2 endolysin accumulates in host cells and interacts with the host's ABC transporter proteins (D. Wei and Zhang 2008). There have been attempts to purify its large terminase to study dsDNA packaging in thermophages (R. Xu 2017).

2.1.1.3. Thermus phage phiFa

The *thermus* phage phiFa is a *siphovirus* which preys on the *thermus* Thermophiles HB27. It was isolated from Mount Etna and Mount Vesuvius in Italy (Lopatina et al. 2019; Severinov et al. 2014). It is somewhat related to the *thermus* P23-45 phage. It is linked to several other thermophages in the *oshimavirus* genus via their DNA polymerase-1, which share a distinctive structural motif called the SβαR motif (Ahlqvist et al. 2022). The gp15 of phiFa is similar to transcription factor gp39 of P23-45. Both inhibit transcription initiation by the host *thermus* RNAP (Severinov et al. 2014). The exact structure and nature of this protein are unknown at time of writing. The phiFa has been shown to have spacer sequences in common with other *thermus* phages across different continents, suggesting that these phages have moved great distances over time (Lopatina et al. 2019). The phiFa phage has been used to study how its host *thermus thermophilus* acquires spacer sequences to use in its CRISPR-Cas system to defend against phage infection (Artamonova et al. 2020; Aviram et al. 2022).

2.1.1.4. Thermus phage phiOH2

The *thermus* phage phiOH2 was isolated from the sediment of the Obama hot spring, Unzen, Nagasaki, Japan. It was found to infect *geobacillus kaustophilus* NBRC 102445. The host genome has 3,541,481 bp with 3,679 predicted coding sequences. Once the phiOH2 genome integrates into the host's genome, it acts like a prophage (Doi et al. 2013). There have been attempts to purify its large terminase to study dsDNA packaging (R. Xu 2017).

2.1.1.5. Bacillus BV1

Bacillus phage BV1 is a *myovirus* that infects *Bacillus* sp 6k512 from an inshore hot spring in Xiamen, China (B. Liu, Wu, and Xie 2010). The dsDNA genome is 35.055 bp in length with 54 ORF. It is one of the two members of the *svunavirus* genus along with *geobacillus* phage GBSV1.

2.1.1.6. Geobacillus GBSV1

Geobacillus GBSV1 (GenBank accession DQ340064) is a thermophilic *myovirus* from *geobacillus sp* 6k51 bacteria. It was found in offshore hot springs in Xiamen, Jujian province, China. The only research on bacteriophage GBSV1 directly has been to characterise its nuclease (Song and Zhang 2008) and characterise its component viral proteins (B. Liu et al. 2009). There has also been some work done to purify its large terminase (R. Xu 2017), but it was not successful. It is the other member of the *svunavirus* genus besides BV1.

2.1.1.7. Rhodothermus phage RM378

Rhodothermus phage RM378 is a *myovirus* that infects the eubacterium *Rhodothermus marinus*. It is the first phage to be found to infect this host. It is found in submarine hot springs in Ísafjarðardjúp, Iceland (Hjörleifsdottir et al. 2002). It was also found in Blue Lagoon and Salt factory on Reykjanes in southwest Iceland. No work has yet been done on its portal protein, only on the RNA ligase (Blondal et al. 2003) and kinase (Blondal et al. 2005). Befitting its environment, it was found to be very thermally stable. Many of the proteins in RM378 have homologues in T4 phage, e.g. their rare RNA ligase protein (Hjörleifsdottir et al. 2014).

2.1.1.8. Thermus phage TSP4

The *Thermus siphoviridae* phage TSP4 was first isolated from the Tengchong hot springs in China. (L. Lin et al. 2010). It was found to be very similar to *Thermus sephoviruses* P23-45 and P74-26 from Kamchatka peninsula hot springs in Russia. The only other work on this portal protein has been on its helicase (Q. Zhang et al. 2014).

2.1.2. Bioinformatics and structure prediction

Basic parameters were found for each protein using the ProtParam web server, e.g. pI, extinction coefficient, stability, and molecular weight (Gasteiger et al. 2005). These were used to estimate the concentration of the protein using a nanoDrop and to estimate its elution volume during purification. AlphaFold was used to give a prediction of the structure of each portal protein. The structure prediction was used to inform construct design to remove long, unstructured regions that would hinder purification. PsiPred and AIUPred were used to estimate the locations of disordered regions of the proteins.

Figure 2-1: Uniprot Information

Organism	Protein	UniProt	MW (Da)	Amino acids
Bacillus phage 1	Phage portal protein	A6XMJ5	45,503	403
Geobacillus phage GBSV1	Phage portal protein	Q0H263	45,774	406
Deep-sea thermophilic phage D6E	Portal protein	E5DV51	48,417	422
Geobacillus virus E2	Putative portal protein	A6M949	46,916	411
Thermus phage phiFa	Portal protein	A0A3Q8HYX6	52,068	467
Thermus phage phi OH2	Phage portal protein, SPP1	S6BFK8	55,587	474
Rhodothermus phage RM378	Portal protein	n/a	61,259	533
Thermus phage TSP4	Portal protein	A0A411CWA3	49,563	445

2.1.3. Cloning and construct design

The first attempt with a gene was to clone and express the full-length (FL) coding sequence of the gene. In parallel, based on the alpha fold structure prediction, some constructs were made that had the unstructured regions removed. The FL gene was cloned despite the presence of unstructured regions because a cryoEM structure of the entire protein would be more accurate. The truncations were made in parallel as an insurance choice in the likely event that the FL constructs failed to purify. The YSBL3CLIC+ vector is used throughout this work. It is a modified version of the pET28a vector. The YSBL3CLic+ vector has an N-terminal hexahistidine tag followed by a 3C cleavage site. It contains an additional alanine residue after the 3C cleavage site. This tag is commonly used in protein purification (Table 2-2). The hexahistidine tag allows for purification by affinity chromatography. The tag can then be removed using 3C protease. A cryoEM structure containing the tag is less biologically accurate or can interfere with structure studies (B. Liu et al. 2006). The tag may also interfere with the use of the portal protein as a nanopore downstream by interacting with analytes. The PhiFa_FL_340 construct uses the Champion-SUMO vector instead. The Champion-SUMO vector has a SUMO tag, then a hexahistidine tag on its N-terminal. The SUMO tag may help improve solubility of recombinantly expressed proteins. Two constructs for GVE2 and phiOH2 use a variant of the YSBL3CLic+ vector which has two hexahistidine tags at the N-terminal instead of one.

2.1.4. Recombinant protein expression

A common choice of *E. coli* cell for expression is BL21(DE3) (Table 2-1). The DE3 strains contain a copy of the T7 bacteriophage RNA polymerase downstream of a lacUV5 promotor. Expression can be induced by adding isopropyl ß-D-1-thiogalactopyranoside (IPTG) (Studier and Moffatt 1986; Studier et al. 1990). The cells used for the most part were *E. coli* BL21 Rosetta (DE3) pLysS. The Rosetta strain contains an additional plasmid which codes for rare tRNAs. Even without IPTG, there will be a low level of basal expression. The pLysS strains also have the T7 lysozyme, which suppresses expression of the T7 RNA polymerase to reduce basal expression (Moffatt and Studier 1987). These cells have an additional plasmid with a chloramphenicol resistance gene. This is particularly useful for expressing proteins that may be toxic to the host cells. While portal proteins are not natural membrane proteins, some are used in lipid membranes and they can insert under the right conditions. The pET expression system is used throughout this work (Figure 2-2). In the uninduced state, the Lac repressor (LacI) is expressed and binds the Lac operator (LacO) on the *E. coli* genome and the YSBL3CLIC+ vector. This prevents the binding of RNA polymerase. So T7 polymerase downstream cannot be made. Thus, the GOI on the YSBL3CLIC+ vector, which is under the control of a T7 promotor, is not produced. An inducer, in this case IPTG, binds the LacI protein and prevents it from binding the Lac promotor. This allows T7 RNA polymerase to be made. The T7 RNA polymerase can then attach to the T7 promotor upstream of the gene of interest (GOI) and transcribe it. A ribosome binds the ribosome binding site (RBS) upstream of the coding region.

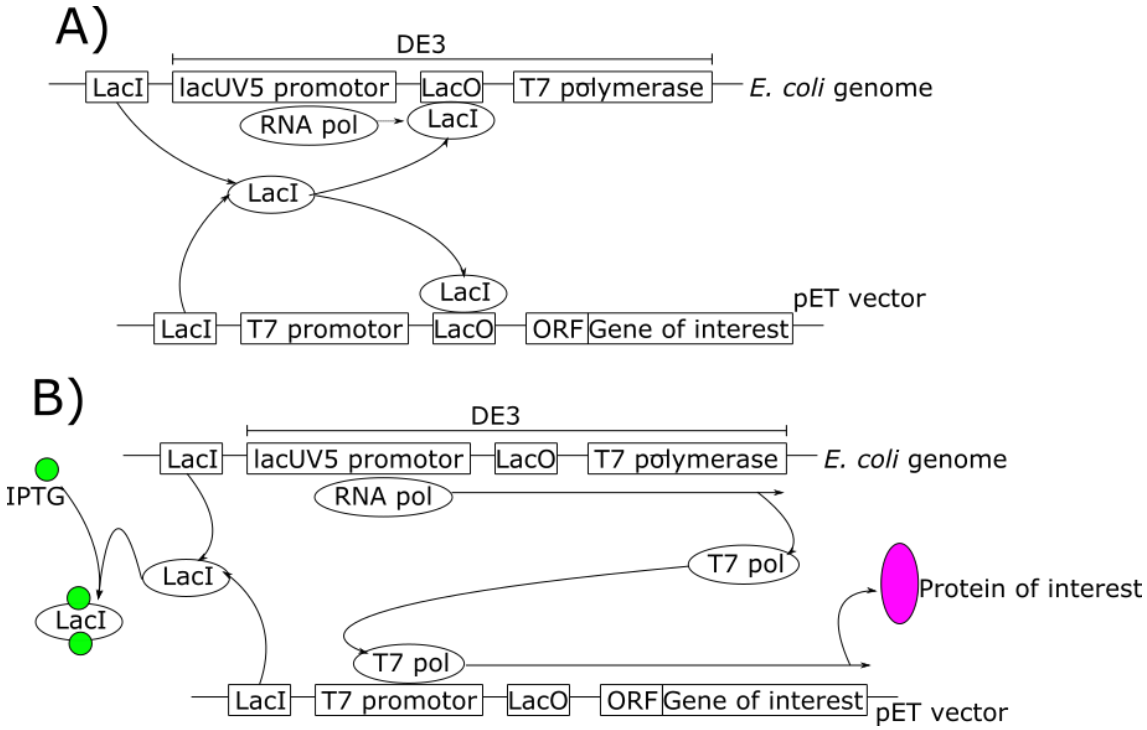


Figure 2-2: pET expression system in YSBL3CLIC+ vector

In the inactive state (A), LacI is produced. It binds to the Lac operator (LacO) which prevents the passage of RNA polymerase along the gene. B) If an activator, in this case IPTG, is added, then it binds to the LacI, preventing it from binding the LacO. RNA polymerase is free to transcribe the gene downstream, in this case the T7 polymerase. The T7 polymerase binds the T7 promotor on the pET vector and transcribes the gene of interest (GOI) downstream. A ribosome binds to the mRNA as it is being transcribed. The ribosome translates the mRNA sequence into the corresponding amino acid sequence to make the protein of interest (POI).

Table 2-1: Portal protein cloning and expression

Phage	Clone	Expression	Ref
80α	• Amplify gene from genome • Clone into pET21a with C-terminal hexahistidine tag • Use InFusion system	• *E.coli* BL21(DE3) cells, • LB medium at 37 °C, 100 μg mL^{-1} ampicillin until OD_{600} = 0.6 • Induced with 1 mM IPTG, 3 hrs, 37 °C. • Centrifuged at 10,000 g. Stored at −80 °C.	(Mukherjee et al. 2023)
PaP3	• Synthetic gene into pET-28a (+)	• *E. coli* BL21 (DE3) • Grown in LB with 50mg/L kanamycin at 37 °C until OD_{600} = ∼0.6 • Induced with 0.5 mM IPTG for 4 hours.	(David Hou et al. 2022)
phi29	• C-terminal hexahistidine tag • Clone into pET-21a(+)	• BL21(DE3) • LB medium with 50 mg mL^{-1} ampicillin • IPTG	(Jing, Burris, and Cortes 2021)
EBV	• Amplify from genome by PCR • Cut pPEU11 vector with KpnI-HindIII • Use InFusion system	• baculoCOMPLETETM ALL-IN-ONE Baculovirus Protein Expression kit (Oxford Expression Technologies) • Infect a monolayer of *Spodoptera frugiperda* Sf9 cells with baculovirus • Grown at 28 °C, 72 hr. Centrifuge cells at 3,000 g, 20 min, 4 °C	(Machón et al. 2019)
T7	• Clone T7 gp8 gene into pET28a between NcoI and NotI	• E. coli BL21(DE3) • Grown at 37 °C until OD_{600} = 0.6 • Induce with 0.4 IPTG. Grow 37°C, 3hr or 16°C ON	(Cuervo et al. 2019)
G20c	• Amplify gene by PCR	• WT G20c and G20c V325G_I328G use *E. coli* BL21(DE3)pLysS • G20c-L49C and G20c-L49C_V325G_I328G use SHuffle cells	(Cressiot et al. 2017a)

	• Clone into YSBL3CLIC+ vector with 3C-cleavable N-terminal hexahistidine tag • Use HiF DNA assembly master mix (New England Biolabs) • Mutants made using SDM	• LB media with 35 µg mL^{-1} kanamycin and 50 µg mL^{-1} chloramphenicol • 10 mL ON culture in 1L LB with antibiotics. Grow at 37°C until OD600 = 0.8 • Add 0.5 mM IPTG. Grow 16 °C ON • Centrifuge 4000 rpm, 30 min. Freeze pellets in liquid N. Store -80 °C • SHuffle cells were incubated at 30 °C before and after induction.	
GVE2	• Amplify from genomic DNA with primers with BamHI and EcoRI sites • Clone into pGEX-4T-2	• E. coli BL21 (DE3) • 200 mL of LB with 100 µg mL^{-1} of ampicillin at 37 °C until OD600 = 0.5 • 0.5 mM IPTG. Grow for 8 h at 22 °C. • Centrifuge 4000g, 5min	(Yiqian Wang and Zhang 2008)

2.1.5. Protein purification

Proteins can be purified by immobilised metal ion affinity chromatography (IMAC). Proteins with a tag of six histidines (hexahistidine tag) on one of their terminals will bind to the resin in a purification column via Ni^{2+} ions. Any molecules without the tag should pass straight through the column. The molecules bound to the resin are bound with varying levels of binding affinity. Some bound weakly, others strongly. To elute the proteins, increasing concentrations of imidazole are washed down the column. Those molecules that are bound weakly will be out competed first, then in increasing binding affinity. At a certain concentration of imidazole, the protein of interest (POI) will be out competed from the resin by imidazole and elute from the column. Since it was over expressed, it should be the most abundant molecule in the cell lysate. Its elution can be tracked by the absorbance at 280 nm.

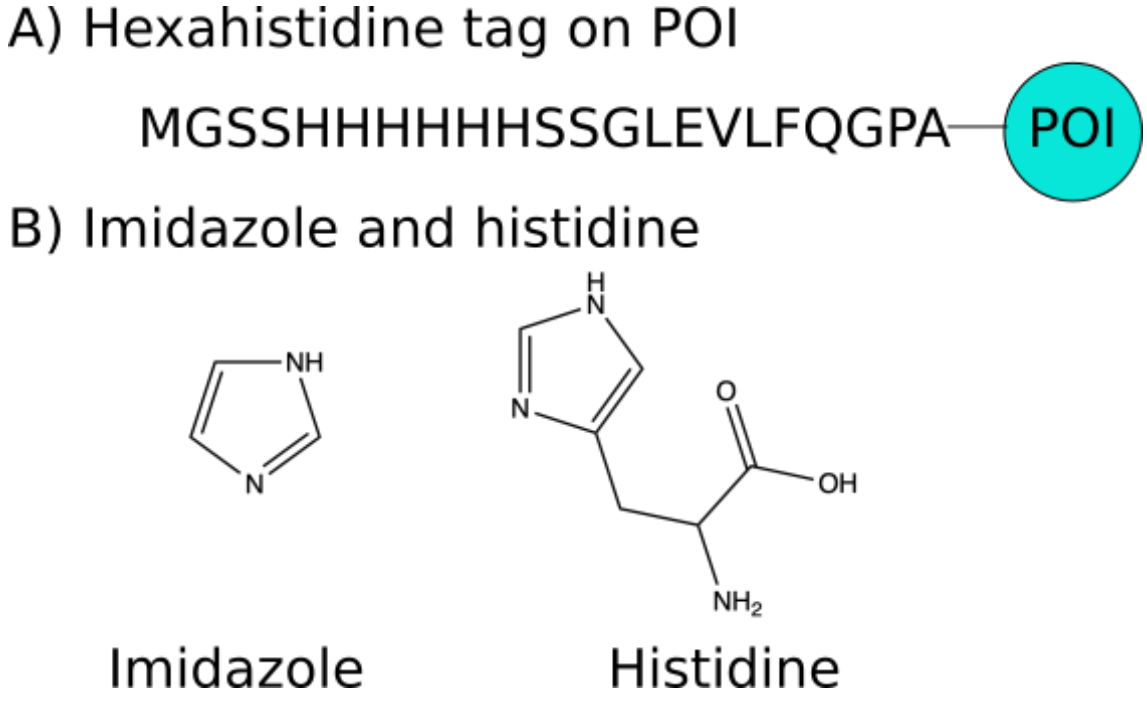


Figure 2-3: Principles of affinity chromatography

A) The hexahistidine tag attached to the N-terminal of the protein of interest (POI). B) Structures of imidazole and histidine. The histidine side chain has an imidazole group on the end. Their similarity allows imidazole to compete with histidine for binding on the Ni+ resin.

To purify DNA-binding proteins, a column packed with heparin sepharose can be used instead. The long heparin polysaccharide mimics the anionic nature of nucleic acid, which allows DNA-binding proteins to bind to it. They can be eluted from the column by increasing the ionic strength of the buffer, usually by increasing the salt concentration. Size exclusion chromatography (SEC) separates proteins based on their hydrodynamic radius, which can be used as a stand in for their size or molecular weight (MW). The column is full of porous resin. Smaller molecules will be able to fit into the pores in the resin, while larger molecules cannot. So larger molecules will move faster down the

column and elute first. The MW of the species in the fractions can be estimated based on the elution volumes of MW standards.

Table 2-2: Portal protein purification

Phage	Purification	Ref
80α	• Re-suspend in 15 mL 50 mM HEPES pH 7.4 0.5 M NaCl, 5% glycerol 20 mM imidazole (binding buffer) with presence of protease inhibitor cocktail (Sigma-Aldrich). • Lyse with Emulsiflex B15 high pressure disruptor (Avestin) • Incubate lysate with 1 µL Benzonase® nuclease (EMD Millipore) with 1 mM MgCl2 at 4 °C for 1 hr. • Centrifuge 27,000g, 30 min, 4 °C • Filter SN through a 0.22 µM syringe filter. • Purify with 1.5 mL NTA Ni-Agarose (Invitrogen) packed in Econo-Pac® Chromatography Columns (Bio-Rad). • Elute with binding buffer with 250 mM imidazole • Dialyse in 25 mM HEPES buffer pH 7.4, 1 M NaCl at 4 °C ON • Incubate at 37°C, 3-4 hr to oligomerise • Concentrate SN with Amicon Ultra 50 kDa cut-off centrifugal concentrator (Millipore) • Store in 500 µL and stored at −80 °C	(Mukherjee et al. 2023)
PaP3	• Lysed by sonication in Lysis Buffer (20 mM Tris–HCl pH 8.0, 400 mM NaCl, 1 M Urea, 5% Glycerol). • Centrifugation at 12,000 rpm 30 min • Incubated with Nickel Agarose Beads (GenScript) for 2 hr at 4 °C. • Washed beads and elute with Elution Buffer (20 mM Tris–HCl pH 8.0, 400 mM NaCl, 1 M Urea, 250 mM Imidazole, 5% Glycerol). • Superdex 200 10/30 column (Cytiva) equilibrated with Gel Filtration Buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 5 mM EDTA pH 8.0). • Use molecular weight markers to calibrate elution volumes • Concentrate ∼50 mg mL^{-1} using a 30 kDa Millipore concentrator. • Oligomerisation of FL- or ΔC-portal subunits was induced by incubating highly concentrated portal samples (∼50 mg mL^{-1}) at 22 °C overnight, • Followed by a new round of gel filtration using an analytical Superdex 200 10/30 column (Cytiva).	(David Hou et al. 2022)
phi29	• Re-suspend in 50 mM Tris, 500 mM NaCl, 10 mM imidazole (pH 7.4) with 15% glycerol	(Jing, Burris, and

	• Lyse by French Press • Centrifuge 16000 rpm, 20 min • HisPure Ni-NTA resin in a column. • Wash buffer: 50 mM Tris, 500 mM NaCl, 50 mM imidazole (pH 7.4) with 15% glycerol • Elution buffer: binding buffer with 0.5 M imidazole • Find concentration by Coomassie (Bradford) protein assay	Cortes 2021)
EBV	• Lysis buffer: 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 mM BME, 0.2% DDM (w/v) • Lyse in French Press at 25 kPsi • Centrifuge 40,000g, 20 min, 4 °C • Filter and load onto HisTrapTM HP 5mL • Binding buffer: 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM BME, 0.05% DDM (w/v), 20 mM imidazole • Elution buffer: binding buffer with 0.5 M imidazole • Purify with HiTrapTM Heparin HP 1-mL • Elute with 150 mM to 1.5 M NaCl in 50 mM Tris-HCl pH 8.0, 20 mM β-mercaptoethanol, 0.05% DDM (w/v), 1 mM EDTA • HiPrep 16/60 Sephacryl S-400 HR (GE Healthcare Life Sciences), with heparin buffer with 500 mM NaCl • Superose 6 10/300 GL (GE Healthcare Life Sciences).	(Machón et al. 2019)
T7	• Re-suspend in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 3 mM B-ME, 20 mM imidazole, 40 μg mL^{-1} DNase I, and Complete Protease Inhibitor Cocktail Tablets • Lyse by French Press • Centrifuge 30000g, 30 min. • HisTrap HP in buffer A (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 3 mM β-mercaptoethanol, 20 mM imidazole) • Elute with in buffer A with 350 mM imidazole • Equilibrate into buffer B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM dithiothreitol, 2 mM ethylenediaminetetraacetic acid) • Concentrated by 30,000 Da MW cut-off • Some cases, dialyse in TMS buffer (50 mM Tris-HCl pH 7.8, 100 mM NaCl, 10 mM MgCl2) before grid preparation	(Cuervo et al. 2019)
G20c	• Thaw cells	(Cressiot et al. 2017a)

	• Re-suspend in 5 mL/g of lysis buffer (50 mM Tris pH 8, 1 M NaCl, 10 mM imidazole, 100 mM AEBSF, 10 mg mL^{-1} lysozyme) • Lyse by sonication on ice • Centrifuge 15000 rpm, 30 min • Filter through 0.22 µm membrane • Purify with HisTrap FF 5 mL • Elute with gradient to 100% buffer B (1 M NaCl, 50 mM Tris pH 7.5, 500 mM imidazole) • Buffer exchange into 50 mM Tris pH 8, 500 mM NaCl 50 mM potassium glutamate • Cut histidine tag off using 3C protease ON • Separate cleaved and uncleaved protein with HisTrap FF 5 mL. • Purify on superpose 6 16/600 column with 20 mM Tris pH8, 1 M NaCl, 50 mM potassium glutamate • Concentrated with 100 kDa MW cut off to 4 mg mL^{-1} for biophysical studies, 10mg mL^{-1} for crystallisation. • Proteins containing cystines used buffers with 2 mM DTT	
GVE2	• Purified with Glutathione-sepharose beads (Sigma, America) according to the manufacturer.	(Yiqian Wang and Zhang 2008)
SPP1	• Concentrate cells 30x in H buffer [20 mM HEPES-NaOH, pH 7.6, 10 mM EDTA, pH 8.0, 500 mM NaCl, 10% (v/v) glycerol] • Lyse by French Press • Precipitate the protein with 75% ammonium sulphate • Re-suspend protein in L buffer [20 mM HEPES-NaOH, pH 7.6, 10 mM EDTA, pH 8.0, 40mM NaCl, 10% (v/v) glycerol] • Purify with DEAE-Sepharose Fast Flow (anion exchanger, Pharmacia) column • Purify with Mono Q column (Pharmacia). Elute with 40-500 mM NaCl step.	(Dube et al. 1993)

2.1.6. Site-directed mutagenesis (SDM)

Once GBSV1_Δ1-40_775 had been found to be soluble and easily purified, a number of mutations were made to make it more suitable for nanopores. The mutations were made by site-directed mutagenesis (SDM). Primers are designed with regions flanking the bp(s) to mutate so that it can bind the DNA template, despite the mismatching bp. PCR is run as normal to produce strands containing the mutation. Subsequent rounds of PCR will use the new mutated strand as a template, and so more and more of it will be produced.

In order to adapt GBSV1_Δ1-40_775 to be inserted into a lipid membrane, the Ser135 was mutated to cysteine to allow the attachment of the MPE PE lipid anchor (see Chapter 5). The cryoEM structure showed that the last helix of the crown domain was disordered (see Chapter 3). Since this is at the mouth of the pore, it is possible that this will flap around and cause additional noise. Therefore, a mutant without residues 383-406 was made. The cryoEM structure also showed that the tunnel lumen is strangely wide throughout, including at the tunnel loops, which are usually the most constricted part of the lumen in portal proteins. The narrowest part of the nanopore gives the best signal:noise ratio. Therefore, Pro282 was mutated to glutamine which should allow the tunnel loops to extend into the lumen, narrowing the diameter, and giving a better signal:noise. Previously, αHL was pulled into a SSN using a dsDNA tether attached to one end of its channel (Hall et al. 2010a). In order to re-create that, the Ala215 was mutated to cysteine, so that dsDNA could be attached to pull the portal protein into the SSN. The G20c portal protein (Cressiot et al. 2018) is naturally bipolar which helps it orientate in the electric field so that the clip domain enters the SSN first. However, the electrostatic surface charges of the GBSV1_Δ1-40_775 portal (see Chapter 3) show that it is negatively charged at the crown and clip domains at pH values above pH 6.5. Therefore, to help orientate it, mutations were made in the clip domain inside and outside the tunnel.

2.1.7. Chapter aims

This chapter presents the screening of different thermophage portal proteins to find suitable candidates that could be taken forwards for structure determination and nanopore testing. The proteins should be soluble, easy to purify in large quantities, not aggregate and not degrade over time.

2.2. Materials and methods

2.2.1. Bioinformatics

The PsiPred web server was used to predict areas of secondary structure of the proteins (Buchan and Jones 2019). PsiPred and AIUPred were both used to predict the disordered regions (Erdős and Dosztányi 2024)(D. T. Jones and Cozzetto 2015). Multiple sequence alignment and sequence similarity was performed using Cluster Omega (Madeira et al. 2024). The ProtParam web server was used to predict basic parameters for each construct (Gasteiger et al. 2005).

2.2.2. Alpha fold prediction of all proteins.

All of the portal proteins were submitted to ColabFold to predict their structure (Mirdita et al. 2022).

2.2.3. Cloning

Several thermophage portal protein genes were already cloned into different vectors by past members of the lab (Table 2-3).

Table 2-3: Constructs made by Lily Bailey, Dr Sandra Greive and Dr Evgeny Klimuk

Gene (accession number)	**Vector**	**Source organism**	**Made by / Gift from**
phiFa_FL (pYM340)	Champion SUMO	*Thermus* phage phiFa	Evgeny Klimuk
D6E_Δ1-34_Δ388-422 (pYM651)	YSBL3CLIC+	D6E Deep-sea thermophilic phage D6E	Sandra Greive
D6E_Δ375-422 (pYM655)	YSBL3CLIC+	D6E Deep-sea thermophilic phage D6E	Sandra Greive
D6E_Δ1-34_Δ375-422 (pYM657)	YSBL3CLIC+	D6E Deep-sea thermophilic phage D6E	Sandra Greive
GVE2_FL (pYM669)	YSBL3CLIC+	*Geobacillus* GVE2	Lily Bailey
GVE2_Δ1-18 (pYM676)	YSBL3CLIC+	*Geobacillus* GVE2	Lily Bailey
PhiOH2_Δ436-474 (pYM685)	YSBL3CLIC+_2His	*Thermus* phage phiOH2	Lily Bailey
GVE2_Δ1-18 (pYM686)	YSBL3CLIC+_2His	*Geobacillus* GVE2	Lily Bailey
PhiFa_FL (pYM714)	pMK-RQ	*Thermus* phage phiFa	Invitrogen

The accession numbers listed below the genes are those used internally by the Antson lab.

The remainder (BV1, GBSV1, RM378 and TSP4) were cloned into the YSBL3CLic+ vector (Table 2-5). The whole coding sequences for the portal proteins BV1, GBSV1, RM378 and TSP4 were ordered from Integrated DNA Technologies (IDT) to include the necessary ends for InFusion cloning on their N- and C-terminals (Table 2-4), so they could be used in InFusion cloning straight away.

Table 2-4: InFusion ends added to the ends of the new portal protein genes

InFusion end (5'-3')	Sequence
5' end	TTCCAGGGACCAGCA
3' end	CGCGCCTTCTCCTCA

Each mutant has an identifying name made of three components: the protein name, a description of the deletions, mutations etc., and a plasmid accession number used within the Antson lab (pYMxxx). For example, D6E_Δ1-34_Δ388-422_651 refers to a mutant of the D6E portal protein which has residues 1-34 and 388-422 removed, and has the accession number pYM651. The plasmids made in the Antson lab are numbered sequentially when they are made, so pYM651 is the 651st plasmid to be numbered with this system. The genes for phiFa_Δ1-19_Δ447-467_715, GBSV1_Δ1-40_775, TSP4_Δ1-43_777 constructs all had to be amplified by PCR first. Table 2-5 shows the primers used to make the different mutants.

The truncations were made using PCR. Primers were made that bound to the areas of the gene that would code for the amino acid at the start of the desired truncation. There were two sections to the primers. The 3' half bound to the gene. The 5' half contained the sequence required for InFusion cloning (Table 2-5). The PCR products were run on a 1% agarose gel. A 50 mL agarose gel was made by dissolving 0.5 g agarose (Melford) in 50 mL TAE buffer. The solution was microwaved in 30 s bursts until the agarose was completely dissolved. 0.5 µL SYBR safe DNA gel stain (Invitrogen) was added. The solution was poured into a mould with a comb and let set for 30 minutes. The gel was run at 100 mV, 400 A for 60 minutes in TAE buffer. Gels were imaged on a GelDoc.

The gene was cloned into the linearised YSBL3CLic+ vector (see Appendix) using InFusion cloning kit (Takara Bio) according to the manual instructions.

The template DNA was digested by 1 µL Dpn1 and 2 µL Tango buffer (Thermo Scientific) for 1 hr, followed by NucleoSpin Gel and PCR clean-up according to the manual (Machinery Nagel). The DNA concentration was checked using a NanoDrop UV Spectrophotometer (Thermo Scientific). The GBSV1_Δ1-40_775 gene needed gel extraction, rather than Dpn1 digestion.

Table 2-5: Constructs made during this project

Gene (Accession number)	Vector	Source organism	Source	Forward primer	Reverse primer
phiFa_Δ1-19_Δ447-467 (pYM715)	YSBL3CLIC+	*Thermus* phage phiFa	Truncation of phiFa_FL_715	pYM715_F20	pYM715_R446
BV1_FL (pYM717)	YSBL3CLIC+	*Bacillus* phage BV1	InFusion	Ordered as whole coding sequence for InFusion into YSBL3CLic+	
GBSV1_FL (pYM718)	YSBL3CLIC+	*Geobacillus* phage GBSV1	InFusion	Ordered as whole coding sequence for InFusion into YSBL3CLic+	
RM378_FL (pYM772)	YSBL3CLIC+	*Rhodothermus* RM378	InFusion	Ordered as whole coding sequence for InFusion into YSBL3CLic+	
TSP4_FL (pYM773)	YSBL3CLIC+	*Thermus* phage TSP4	InFusion	Ordered as whole coding sequence for InFusion into YSBL3CLic+	
GBSV1_Δ1-40 (pYM775)	YSBL3CLIC+	*Geobacillus* phage GBSV1	Truncation of GBSV1_FL_718	pYM775_F40	pYM775_R406
TSP4_Δ1-43 (pYM777)	YSBL3CLIC+	*Thermus* phage TSP4	Truncation of TSP4_FL_773	pYM777_F44	pYM777_R445

The accession numbers listed below the genes are those used internally by the Antson lab. The sequences of the primers are listed in the Appendix. Primers are not listed for BV1_FL_717, GBSV1_FL_718, RM378_FL_772, and TSP4_FL_773 because the whole gene was ordered with the DNA ends necessary for InFusion cloning already on the ends (see Appendix 1).

A heat-shock transformation protocol was used throughout. Stellar cells (Takara Bio) were used when transforming plasmids for cloning. Cells were thawed on ice for 20-30 minutes. 1-5 µL of 10 pg-100 ng of plasmid were added to 25 µL cells.

Table 2-6: Antibiotic requirements of cell lines used

Cell	Antibiotic resistances
E. coli Stellar	Kanamycin
E. coli BL21 Rosetta (DE3) pLysS	Kanamycin, chloramphenicol

They were mixed by gently flicking the tube. Cells were incubated on ice for 30 minutes. The cells were heat shocked in a heat block at 42 ℃ for 45 seconds, then kept on ice for 2 minutes. 300 µL of LB media was added. Cells were grown at 37 ℃, 180 rpm, for 45 min-1 hr. Cells were plated on LB agar plates with appropriate antibiotics. Kanamycin was used at 30 µg mL^{-1}. Chloramphenicol was used at 34 µg mL^{-1}. The plates were incubated at 37 ℃ overnight.

The success of cloning was checked with colony PCR using a protocol in the Antson lab. A master mix was made according to Table 2-7. 7 µL of the mastermix was aliquoted into PCR tubes. A colony was picked from the plate using a sterile pipette tip and under sterile conditions. The colony was stabbed onto a fresh plate with the same antibiotics, then placed into a PCR tube of mastermix and the tip was washed in PCR mix. The picked colonies were labelled on the original plate and the new plate. PCR was run using the PCR mixes using the settings in Table 2-7. The results were run on a 1% agarose gel.

Table 2-7: Colony PCR components

Component	Final concentration	
T7 primer, 2 µM	0.2 µM	10 µL
T7t primer, 2 µM	0.2 µM	10 µL
Clonetech PCR master mix (Takara)	1x	10 µL
dNTP, 25 mM	0.2 mM	0.8 µL
Polymerase*		0.5 µL
Water		To 100 µL

**The polymerase is DreamTaq Green DNA polymerase (Fermentas EP0711)*

The sequences of the T7 and T7t primers are shown in Table 2-8.

Table 2-8: T7 and T7t primer sequences

Primer	Sequence
T7	TAATACGACTCACTATAGGG
T7t	GCTAGTTATTGCTCAGCGG

The GeneJET Plasmid Miniprep kit (Thermo Scientific) was used to purify DNA produced by a culture of a single colony. Plasmids were sequenced by Sanger DNA sequencing (Eurofins Genomics), and later Source Bioscience.

Table 2-9: Colony PCR settings

Temperature (°C)	Time
94	1 min
94	15s
50	30s
72	1 min per kbp
72	5 min
4	HOLD

2.2.4. Expression

2.2.4.1. Small-scale test

Plasmids were transformed into *E. coli* BL21 Rosetta (DE3) pLysS cells using the same protocol as for Stellar cells above. 1 mL of cells were added to 50 mL of autoclaved LB with the appropriate antibiotics. The 50 mL of cells were grown at 37 °C, 180 rpm. The optical density at 600 nm (OD_{600}) was measured every hour until OD_{600} reached 0.6-0.8. Cells were chilled on ice for 30 minutes. A 1 mL pre-induction sample was taken. IPTG was added to a final concentration of 1 mM. The cells were grown overnight at 16 °C. The cells were harvested in 1 mL aliquots in 1.5 mL Eppendorf tubes. They were spun down at 14,300 g for 5 minutes to make cell pellets. The supernatant was discarded. The cell pellets were kept at -20 °C in a labelled bag.

For small-scale expression tests, 5 mL of autoclaved LB media was added to a 50 mL falcon tube under sterile conditions. Kanamycin and chloramphenicol were added to final concentrations of 30 μg mL^{-1} and 34 μg mL^{-1}, respectively. A single colony was picked from the agar plate and added to the media under sterile conditions. The cells were grown overnight at 37 °C, 180 rpm. For the small-scale tests of D6E_Δ1-34_Δ388-422, D6E_Δ375_422_655, D6E_Δ1-34_Δ375-422_657, the buffer was 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 30 mM imidazole, 10 mM KGlu, 10% glycerol. The method for solubility test was taken from the EMBL website. 1ml cell pellets were re-suspended in 100 μL of buffer with different concentrations of NaCl or KCl:

1. 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 5 mM EDTA, 1mg mL^{-1} lysozyme
2. 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 5 mM EDTA, 1mg mL^{-1} lysozyme
3. 50 mM Tris-HCl pH 7.5, 1 M NaCl, 5 mM EDTA, 1mg mL^{-1} lysozyme
4. 50 mM Tris-HCl pH 7.5, 0.1 M KCl, 5 mM EDTA, 1mg mL^{-1} lysozyme
5. 50 mM Tris-HCl pH 7.5, 1 M KCl, 5 mM EDTA, 1mg mL^{-1} lysozyme

The resuspension was homogenised using a vortex mixer. Cells were lysed by sonication as before. The cell lysate was spun in a microcentrifuge at 16,000 g for 5 minutes to separate out the soluble and

insoluble fractions. The supernatant was collected (the S fraction). 1 mL of acetone was added to each supernatant fraction and vortexed, then left on ice for 15 minutes. The supernatant was spun at 16,000 g for 5 minutes. The acetone was discarded without disturbing the pellet. The acetone-treated pellet was dried at 37 ℃ in a heat block, then 15 µL SDS-loading buffer was added. 25 µL SDS-loading buffer was added to the insoluble cell pellet. The samples were prepared for and run on an SDS-PAGE gel as before.

2.2.4.2. Nickel pull down assay

A nickel pull-down assay was used on RM378_FL_772 and TSP4_FL_773 to test their solubility and ability to bind Ni^{2+} resin (Generon). The buffer was 50 mM Tris-HCl pH 8, 1 M NaCl, 10 mM imidazole. The buffer for RM378_FL_772 also had 2 mM DTT, as RM378 contains cysteines. 50 µL of resin were washed with 500 µL of buffer. The cell pellet was homogenised and separated into insoluble and soluble fractions as for a small-scale expression test. The soluble fraction was added to the beads and incubated for 10 minutes. The beads were spun down. A sample of the supernatant (the fraction unbound to the beads) was taken. A sample of the beads themselves was also run on the gel.

Table 2-10: Vectors

Vector	Antibiotic resistance	Affinity tag/protein fusion
YSBL3CLIC+	Kanamycin	3C protease cleavable N-terminal 6His tag
Champion SUMO	Kanamycin	SUMO cleavable 6His tag- SUMO
YSBL3CLIC+_2His	Kanamycin	3C protease cleavable N-terminal 8His-linker-8His tag

2.2.5. Purification

2.2.5.1. Large-scale expression

For purification attempts, a large-scale expression was done. The protocol was the same as for small-scale, except that six L of LB were used as six 2.5 L flasks with 1 L in each. The starter culture was 100 mL LB. 10 mL of starter culture were added to each 1 L of media. The cells were harvested by centrifugation in an SLC6000 at 5,000 g, 4 ℃ for 45 minutes.

1 mL of cells were added to 50 mL of autoclaved LB with the appropriate antibiotics. The 50 mL of cells were grown at 37 ℃, 180 rpm. The OD_{600} was measured every hour until it reached 0.6-0.8. Cells were chilled on ice for 30 minutes. A 1 mL pre-induction sample was taken. IPTG was added to a final concentration of 1 mM. The cells were grown overnight at 16 ℃. The cells were harvested in 1 mL aliquots in 1.5 mL Eppendorf tubes. They were spun down at 14,300 g for 5 minutes to make cell pellets. The supernatant was discarded. The cell pellets were kept at -20 ℃ in a labelled bag.

2.2.5.2. Affinity chromatography

The buffers, MQ water and 20% EtOH were filtered before being put into the AKTA Pure (Cytiva). 5 mL of wash buffer was used per gram weight of cell pellet. The buffers used for each purification are shown in Table 2-13. The pellet was homogenised in a glass beaker with a magnetic stirrer on ice. To aid in cell lysis, the following were added to the buffer: 0.1 mg mL^{-1} lysozyme, 1 mM AEBSF, 0.5 µg mL^{-1} leupeptin, 0.7 µg mL^{-1} pepstatin, 1 mM $MgCl_2$, 0.025 mg mL^{-1} DNase, 0.002 mg mL^{-1} RnaseA. The cell resuspension was sonicated to lyse the cells (Bandelin Sonopuls) for one minute at 50% power, cycle5, followed by two minutes of rest. This was repeated 10 times. A sample of the cell lysate was taken before being spun down in an ultracentrifuge (Sorvall Evolution rc) using an SS34 rotor at 18,000 g, 4 ℃, 45 minutes. A sample was taken of the soluble fraction of the lysate. An AKTA Pure was prepared by running a Clean-in-place (CIP) protocol with warm water, then wash buffer. A 5 mL HisTrap FF column (Cytiva) was equilibrated according to the manual with five column volumes (CV) of MQ water, then wash buffer, then elution buffer, then wash buffer again.

The purification of BV1_FL_717 used a 1 mL HisTrap FF column. The heparin purification of GBSV1_FL_718 used a 5 mL HiTrap HP column (Cytiva).

The purification of GBSV1_Δ1-40_775 used two 5 mL HisTrap FF columns joined together. The soluble fraction was applied using a sample pump. The protein was eluted using an imidazole gradient from 0-100% elution buffer unless otherwise stated. The elution fractions were run on a 12% SDS-PAGE gel (Table 2-11 and Table 2-12). The fractions that comprised the peak were pooled together.

The GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 protein was first purified by Ni^{2+} affinity chromatography using the same buffers and methods as for other GBSV1_Δ1-40_775 mutants. When this failed, the flow-through and wash fractions were pooled together and loaded onto two 5 mL HP heparin columns joined in tandem. The wash buffer for the heparin purification was 50 mM Tris-HCl pH 8, 0.25 M NaCl. The elution buffer was the same but with 1 M NaCl. The eluted fractions were pooled and incubated with 3C protease overnight to remove the tag as above. It was run over another 5 mL FF HisTrap column using the same buffers as above to remove separate tagged and untagged proteins. The elutions were pooled and run over an Superose 6 increase 10/300 column with 50 mM Tris-HCl 0.25 M NaCl pH 8 buffer.

Table 2-11: Recipe for resolving gels for SDS-PAGE

Component	12%	15%
Water	3.2 mL	2.4 mL
1.5 M Tris-HCl pH8.8, 0.4% SDS	2.5 mL	2.5 mL
30% acrylamide	4.2 mL	5 mL
10% APS	50 µL	50 µL
TEMED	8 µL	8 µL

Table 2-12: Recipe for stacking gel for SDS-PAGE

Component	Volume
Water	3.2 mL
0.5 M Tris-HCl pH6.8, 0.4% SDS	1.3 mL
30% acrylamide	0.5 mL
10% APS	25 µL
TEMED	8 µL
1% Bromophenol Blue	10 µL

Magic dye was made by YSBL technicians as follows: 60 mg Coomassie Brilliant Blue G-250 (CBB) (Sigma B0770) was dissolved in 1 L MQ water by stirring at room temperature for 2-3 hours. 3.44 mL concentrated HCl was added to a final concentration of 35 mM HCl. The dye was stored in the dark.

Table 2-13: Buffers used for the purification of each construct

Construct	Buffer
PhiFa_FL_340	50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 10 % glycerol, (wash buffer – 30 mM imidazole/elution buffer - 0.5 M imidazole) , 10 mM KGlu.
GVE2_FL_669	50 mM Tris-HCl pH7.5, 0.5 M NaCl 1% glycerol (wash buffer – 30 mM imidazole/elution buffer – 500 mM imidazole)
GVE2_Δ1-18_676	50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 10% glycerol, (wash buffer – 30 mM imidazole/ elution buffer - 0.5 M imidazole, 10 mM KGlu). The buffer for cell lysis had no imidazole and 10 mM KGlu.
GVE2_Δ1-18_686	50 mM Tris-HCl pH 7.5, 1 M NaCl, 5% glycerol, (wash buffer – 20 mM/elution buffer - 500 mM imidazole)
BV1_FL_717	50 mM Tris-HCl pH 8, 0.5 M NaCl, 5% glycerol (elution buffer - 500 mm imidazole)
GBSV1_FL_718	Ni affinity: 50 mM Tris-HCl pH 8, 1 M NaCl, (wash buffer - 10 mM imidazole/elution buffer - 0.5 M imidazole). Heparin purification: 50 mM Tris-HCl pH 8, 10 mM imidazole, (wash - 200mM NaCl/elution – 1M NaCl).
GBSV1_Δ1-40_775	50 mM Tris-HCl pH 8, 0.25 M NaCl, 5% glycerol, 0/0.5 M imidazole.

The N-terminal affinity tags were not cut off from the GVE2_Δ1-18_676, GBSV1_FL_718, GBSV1_Δ1-40_775 and GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 purifications shown here. To remove the N-terminal tags, 3C protease was added to the pooled fractions. The PhiFa_FL_340 protein has an N-terminal hexahistidine-SUMO tag. It was removed with SUMO protease during dialysis. The solution was also dialysed in wash buffer without imidazole to remove imidazole from the pooled fractions. Dialysis used 2 L of wash buffer without any imidazole. 1 L of 2x concentrated buffer was made and filtered, then added to 1 L of filtered MQ water. The dialysis membrane had a 100 kDa cut off. It was stirred as slowly as possible in a 4 ℃ cold room overnight. The next day, the dialysed sample was passed over a 5 mL FF HisTrap using the same protocol as before to separate the cut tags, protein with tags and protein without tags. Those without should pass straight through the column in the flow-through fraction. The dialysed fractions were concentrated to 1 mL in a 100 kDa spin concentrator (Amicon Ultra-15) according to the manual instructions. A spin concentrator was used to concentrate proteins unless otherwise stated.

The GBSV1_FL_718 and the 2nd attempt with the GVE2_Δ1-18_676 protein used ammonium sulphate precipitation to concentrate the protein. 50% ammonium sulphate precipitation was used. An equal volume of saturated $(NH_4)_2SO_4$ solution was added to the pooled fractions from affinity

chromatography. It was gently mixed by inversion, then incubated on ice for 15 min. The solution was spun 4,500g 4 ℃ for 30 minutes. The supernatant was carefully discarded. The pellet was re-suspended in 0.5-1 mL of SEC buffer.

1.1.1.1. Size exclusion chromatography (SEC)

The SEC column was a superpose 6 increase 10/300 (Cytiva) unless otherwise stated. The flow rate was 0.5 mL min^{-1}. The purifications with GBSV1_Δ1-40_775 used a Hi load 16/600 superdex 200 pg column. The fractions that made up the middle 50% of the peak were collected and pooled. They were concentrated in a 100 kDa spin concentrator (Amicon) and flash frozen in liquid nitrogen in 40 µL aliquots.

Table 2-14: Buffers used in SEC purifications

Name	**Buffer**
GBSV1_FL_718	25 mM Tris-HCl pH 8, 1M NaCl.
GVE2_Δ1-18_676	50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 10% glycerol, 2 mM DTT
GBSV1_Δ1-40_775 Mutants of GBSV1_Δ1-40_775	50 mM Tris-HCl pH 8, 0.25 M NaCl

2.2.6. Negative stain

Formvar/Carbon 200 mesh Cu grids (Agar Scientific) were glow discharged using a Pelco glow discharge machine for 10s hold, 15 mA, 0.38 mBar. Two 5 µL droplets of water, two 5 µL droplets of 2% uranyl acetate (UA) kept at 4 ℃, and a 5 µL droplet of sample were pipetted onto a sheet of parafilm. Tweezers were used to gently place the grid onto the sample drop. The sample was left for 30s before wicking off with filter paper. This was repeated for the two water droplets and UA droplets. Grids were left for 1 hr under cover to dry before being stored in a grid disc until imaging. The grids were imaged at the Technology Facility at the University of York using their FEI Tecnai 12 G2 TEM. Images were taken at -500 defocus unless otherwise stated.

For the phiFa_FL_340 grids, the protein concentration in the B2 fraction was 0.064 mg mL^{-1}. The concentration in the B11 fraction was 0.144 mg mL^{-1}. The concentration in the C2 fraction should be 0.719 mg mL^{-1-}. For the GVE2_Δ1-18_676 and GBSV1_Δ1-40_775 grids, the protein concentration was 0.05 mg mL^{-1}.

2.2.7. Site directed mutagenesis (SDM)

The proposed mutants and their primers are listed in Table 2-15. The GC% and melting temperature were checked using the OligoAnalyzer tool by IDT (‘Oligo Analyzer’, n.d.). Some mutants required

several rounds of SDM to add the desired mutations (Figure 2-4). For the single mutations, the Takara CloneTech PCR mastermix (Takara Bio) was used according to the manual, except that the total reaction volume was halved, and the melting temperature was 50 ℃. The D213N primers required phusion polymerase and buffer (High GC buffer), betaine, 3% DMSO and a melting temperature of 58 ℃. Once PCR was successful, the products were treated as expressed and purified as above for GBSV1_Δ1-40_775.

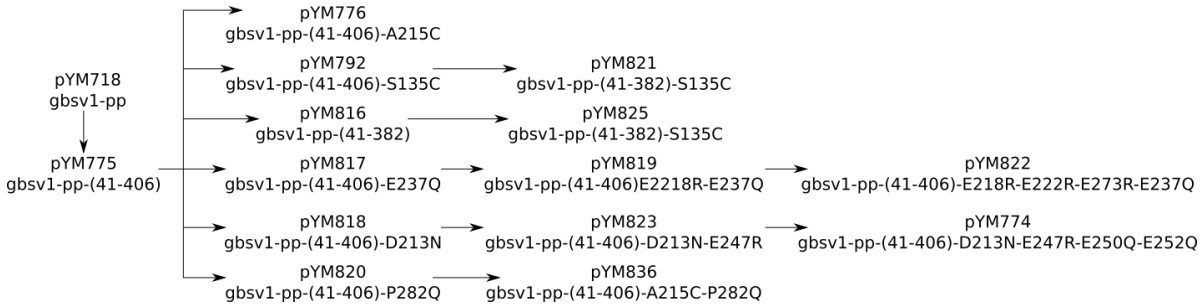


Figure 2-4: Family tree of GBSV1 portal protein mutants

The GBSV1_Δ1-40_E237Q_817, GBSV1_Δ1-40_E218R_E237Q_819, GBSV1_Δ1-40_D213N_818, GBSV1_Δ1-40_D213N_E247R_823 are intermediate constructs in order to produce GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 and GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q_774, which required several rounds of SDM.

Table 2-15: Mutations of GBSV1_Δ1-40_775

Gene (Accession number)	Vector	Desc.	Source	Forward primer (5'-3')	Reverse primer (5'-3')
GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q (pYM774)	YSBL3Clic+	Positive charge inside the clip. pt3/3	Mutagenesis of GBSV1_Δ1-40_D213N_E247R_823	pYM823_F[E250Q,E252Q]	pYM823_R[E250Q,E252Q]
GBSV1_Δ1-40_A215C (pYM776)	YSBL3Clic+	Cysteine residue at the bottom of the clip domain to attach dsDNA	Mutagenesis of GBSV1_Δ1-40_775	pym775_F[A215C]	pym775_R[A215C]
GBSV1_Δ1-40_S135C (pYM792)	YSBL3Clic+	Cysteine residue at the edge of the wing domain for putting attachment to MBP PE	Mutagenesis of GBSV1_Δ1-40_775	pYM775_F[S135C]	pYM775_R[S135C]
GBSV1_Δ1-40_ Δ383-406 (pYM816)	YSBL3Clic+	Remove the last helix of the crown domain	Mutagenesis of GBSV1_Δ1-40_775	pYM775_F(Δ383-406)	pYM775_R(Δ383-406)
GBSV1_Δ1-40_E237Q (pYM817)	YSBL3Clic+	Positive charge on outside of the clip domain. pt1/3	Mutagenesis of GBSV1_Δ1-40_775	pYM775_F[E237Q]	pYM775_R[E237Q]
GBSV1_Δ1-40_D213N (pYM818)	YSBL3Clic+	Positive charge on the inside of the clip domain. pt1/3	Mutagenesis of GBSV1_Δ1-40_775	pYM775_F[D213N]	pYM775_R[D213N]
GBSV1_Δ1-40_E218R_E237Q (pYM819)	YSBL3Clic+	Positive charge on the outside of the clip domain. pt2/3	Mutagenesis of GBSV1_Δ1-40_E237Q_817	pYM817_F[E218R]	pYM817_R[E218R]
GBSV1_Δ1-40_P282Q (pYM820)	YSBL3Clic+	Allow the tunnel loops to move further into the tunnel.	Mutagenesis of GBSV1_Δ1-40_775	pYM775_F[P282Q]	pYM775_R[P282Q]
GBSV1_Δ1-40_ Δ383-406_S135C (pYM821)	YSBL3Clic+	crownless. with cysteine for putting into membrane	Mutagenesis of GBSV1_Δ1-40_S135C_792	pYM792_F(Δ383-406)	pYM792_R(Δ383-406)
GBSV1_Δ1-40_E218R_E222R_E223R_E237Q (pYM822)	YSBL3Clic+	Positive charge on the outside of the clip domain. pt3/3	Mutagenesis ofGBSV1_Δ1-40_E218R_E237Q_819	pYM819_F[E222Q,E223R]	pYM819_R[E222Q,E223R]

GBSV1_Δ1-40_D213N_E247R (pYM823)	YSBL3Clic+	Positive charge on the inside of the clip domain. pt2/3	Mutagenesis of GBSV1_Δ1-40_D213N_818	pYM818_F[E247R]	pYM818_R[E247R]
GBSV1_Δ1-40_Δ383-406_A215C (pYM825)	YSBL3Clic+	crownless with cysteine in tail for hall mutant	Mutagenesis of GBSV1_Δ1-40_ Δ383-406_816	pym775_F[A215C]	pym775_R[A215C]
GBSV1_Δ1-40_A215C_P282Q (pYM836)	YSBL3Clic+	narrower tunnel loop with cysteine in tail for hall mutant	Mutagenesis of GBSV1_Δ1-40_P282Q_820	pym775_F[A215C]	pym775_R[A215C]

The sequences of the primers are listed in the Appendix.

2.3. Results

2.3.1. Multi-sequence alignment

The amino acid sequences of the eight portal proteins were aligned using Clustal Omega (Madeira et al. 2024) and PROMALS3D (Pei, Kim, and Grishin 2008). The sequence similarity between the portals is low (Table 2-16). Any similarities may be due to random chance. The exceptions are BV1 and GBSV1 which are the only two members of the *svunavirus* genus. Their amino acid sequences are 83.08% similar. Despite the low amino acid sequence similarity, there are regions of shared secondary structure between the eight portals (see Appendix). The N- and C-terminals tend to not share secondary structure features.

Table 2-16: Percentage identity matrix for thermophage portal proteins

	D6E	phiOH2	phiFa	TSP4	RM378	GVE2	BV1	GBSV1
D6E	100	23.75	17.16	16.44	15.36	10.11	11.49	13.03
phiOH2	23.75	100	16.83	15.1	15.36	12.82	15.65	17.18
phiFa	17.16	16.83	100	35.42	19.05	14.67	16.98	18.35
TSP4	16.44	15.1	35.42	100	17.88	15.97	13.56	15.19
RM378	15.36	15.36	19.05	17.88	100	16.62	19.1	18.64
GVE2	10.11	12.82	14.67	15.97	16.62	100	26.4	25.66
BV1	11.49	15.65	16.98	13.56	19.1	26.4	100	83.08
GBSV1	13.03	17.18	18.35	15.19	18.64	25.66	83.08	100

Found using Clustal Omega Multiple Sequence Alignment (Madeira et al. 2024). The results are coloured by a scale. Red – low values , green – high values.

2.3.2. Disorder prediction

The DISOPRED3 programme in PsiPred was run for the whole amino acid sequence of each of the eight portal proteins (see Appendix B) (D. T. Jones and Cozzetto 2015; Buchan and Jones 2019). The N- and C-terminals were often disordered (Table 2-17 and Appendix). AIUPred was also run to predict disordered regions (see Appendix B) (Erdős and Dosztányi 2024). The results predictions largely agree with each other, except that AIUPred shows more disordered regions at the far N-terminal regions.

Table 2-17: Regions of order and disorder predicted by DISOPRED3

Protein	Ordered regions (below 0.5)	Disordered regions (above 0.5)
BV1	26-385	1-25, 286-403
D6E	1-397	398-422
GBSV1	28-388	1-27, 289-406
GVE2	26-402	1-25, 403-411
phiFa	24-453	1-23, 454-467
phiOH2	1-442	443-474
RM378	39-454	1-38, 455-533
TSP4	20-426	1-19, 427-445

2.3.3. Alpha fold predictions

ColabFold was used to predict the structures of the eight portal proteins (Figure 2-5) (Mirdita et al. 2022). The portal proteins all have the portal protein fold: crown, wing, stem, clip. Most of the portals have a crown domain made of three alpha helices. D6E and phiOH2 both have long helices in their crown domains. The last helix in GVE2, BV1 and GBSV1 is less well formed. The last region of the crown domain of RM378 is very long and unstructured. Many of the portals have a long N-terminal tail under the wing domain (Rūmnieks, Füzik, and Tārs 2023; Bayfield et al. 2019; Sun et al. 2015a). The positions vary between proteins.

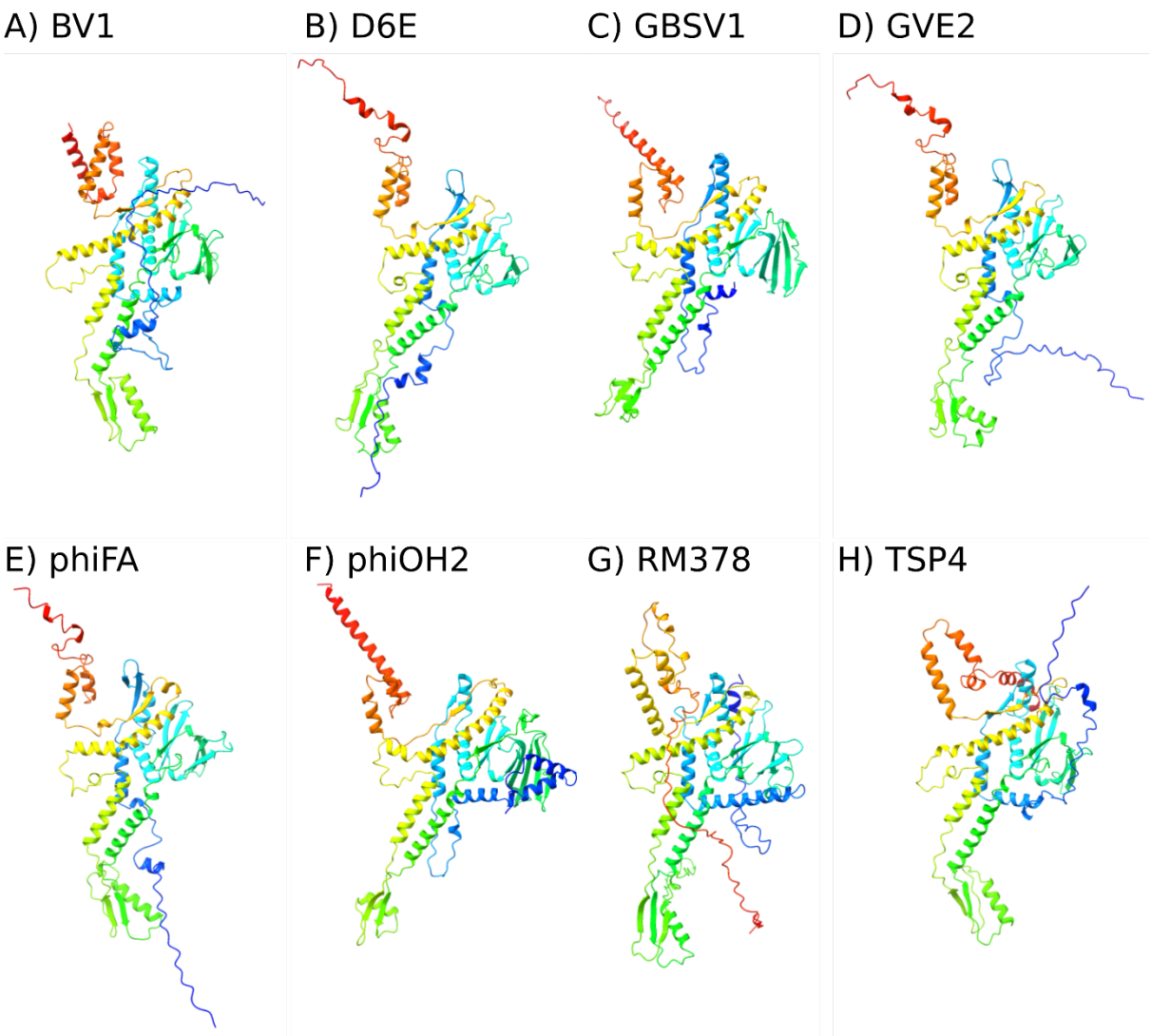


Figure 2-5: AlphaFold predictions of eight thermophage portal proteins.

Coloured by rainbow (not domain). The structures are coloured by rainbow in Chimera. A) BV1. B) D6E. C) GBSV1. D) GVE2. E) phiFa. F) phiOH2. G) RM378. H) TSP4.

2.3.4. Construct Expression Design

Long, disordered, regions can hinder protein purification by causing the proteins to aggregate. The AlphaFold predictions for the portal proteins show that several of them have long N-terminal tails, as is typical of portal proteins (Figure 2-5) (Sun et al. 2015b; Bayfield et al. 2019; Rūmnieks, Füzik, and Tārs 2023). Previously, these N-terminal regions have had to be removed for purification, or if not, they were not able to be resolved by cryoEM. The first 19 and last 20 residues were removed from the phiFa protein (phiFa_Δ1-19_Δ447-467_715) (Table 2-5). These regions were removed because they included regions that were predicted to be disordered (Appendix A, Table 2-17). The first 43 residues were removed from the TSP4 protein. This included the N-terminal region which was predicted to be disordered by AIUPred, DISOPRED3, and the AlphaFold structure predictions.

2.3.5. Expression

Three D6E constructs made by Dr Sandra Greive were tested for their expression and solubility: D6E_Δ1-34_Δ388-422_651, D6E_Δ375-422_655, D6E_Δ1-34_Δ375-422_657. A truncated form of the phiFa protein was also tested at the same time. However, none of the D6E constructs appeared to express in the conditions tested (Figure 2-6). Therefore, this protein was left while others were tested. Due to time constraints, no further expression or solubility conditions were tested. The phiFa_Δ1-19_Δ447-467_715 protein expressed but was not soluble (Figure 2-6). The PhiOH2_Δ436-474_865 mutant of phiOH2 made by Lily Bailey was expressed, but appeared to be insoluble in all conditions thus far tested. The gel resembled that of Figure 2-6B. A small-scale nickel pull down assay was done for RM378_FL_772 and TSP4_FL_773. Both were expressed, especially at 37 ℃; however, RM378_FL_772 was slightly soluble and bound to the Ni resin, while TSP4_FL_773 did not (Figure 2-6C). The small-scale expression for TSP4_Δ1-43_777 showed similar results to that for its full-length counterpart (Figure 2-6B-C).

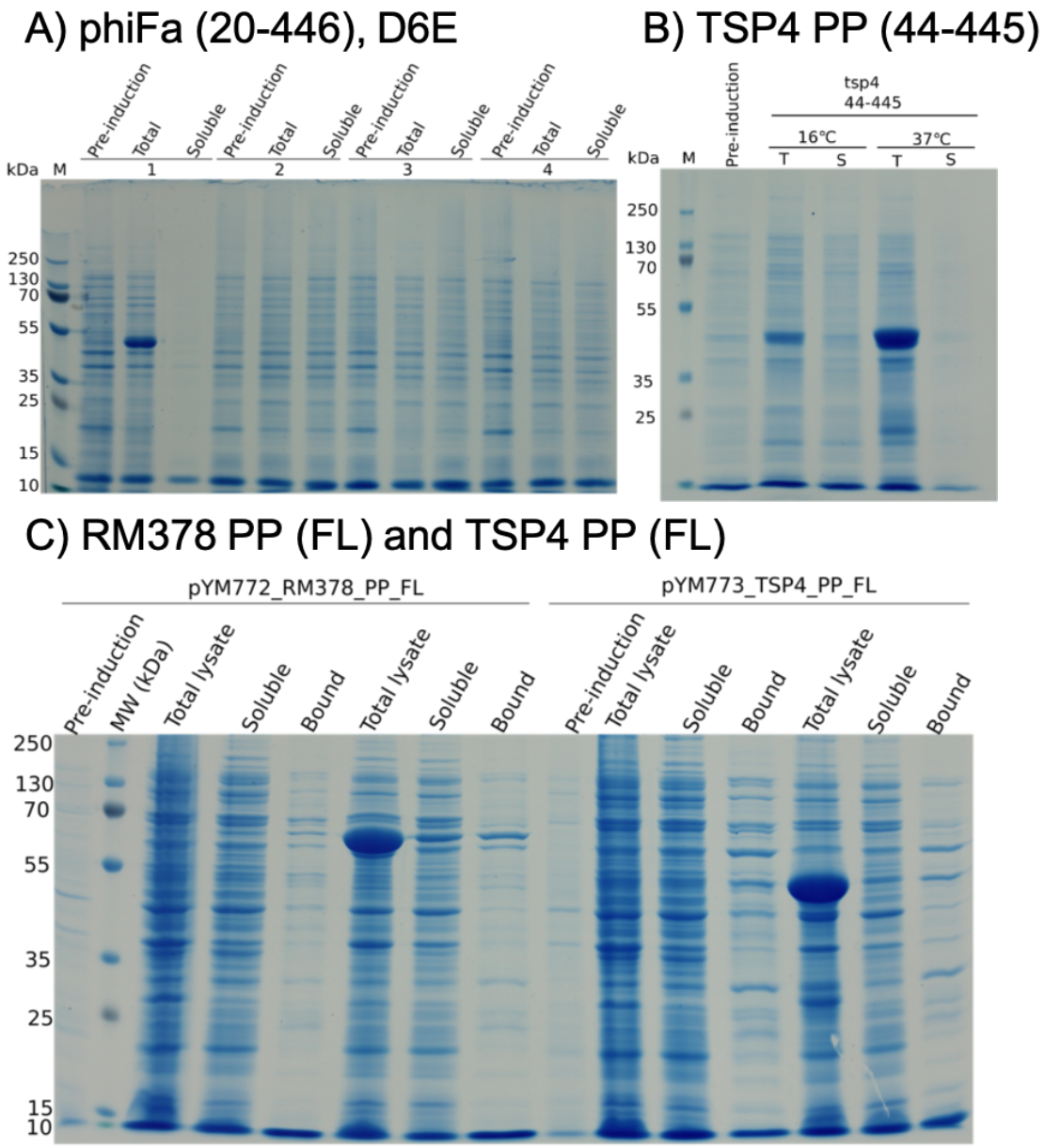


Figure 2-6: Insoluble or low expressing constructs

A) Expression test of phiFa_Δ1-19_Δ447-467_715 and different truncations of the D6E gene 1) phiFa_Δ1-19_Δ447-467_715. The D6E constructs tested were D6E_Δ1-34_Δ388-422_651 (1), D6E_Δ375-422_655 (2), and D6E_Δ1-34_Δ375-422_657 (3). B) Expression and solubility test of TSP4 44-445. C) Nickel pull down assay for RM378_FL_772 and TSP4_FL_773.

The composition of the buffer used to re-suspend the cell pellet can have a great impact on the solubility of the expressed protein, e.g. pH, salt, salt concentration, detergents, glycerol concentration. Figure 2-7 shows a representative SDS-PAGE gel of the phiFa_Δ1-19_Δ447-467_715 re-suspended with different concentrations of NaCl or KCl to see which conditions improve its solubility. In this case, none of the conditions tested improved its solubility.

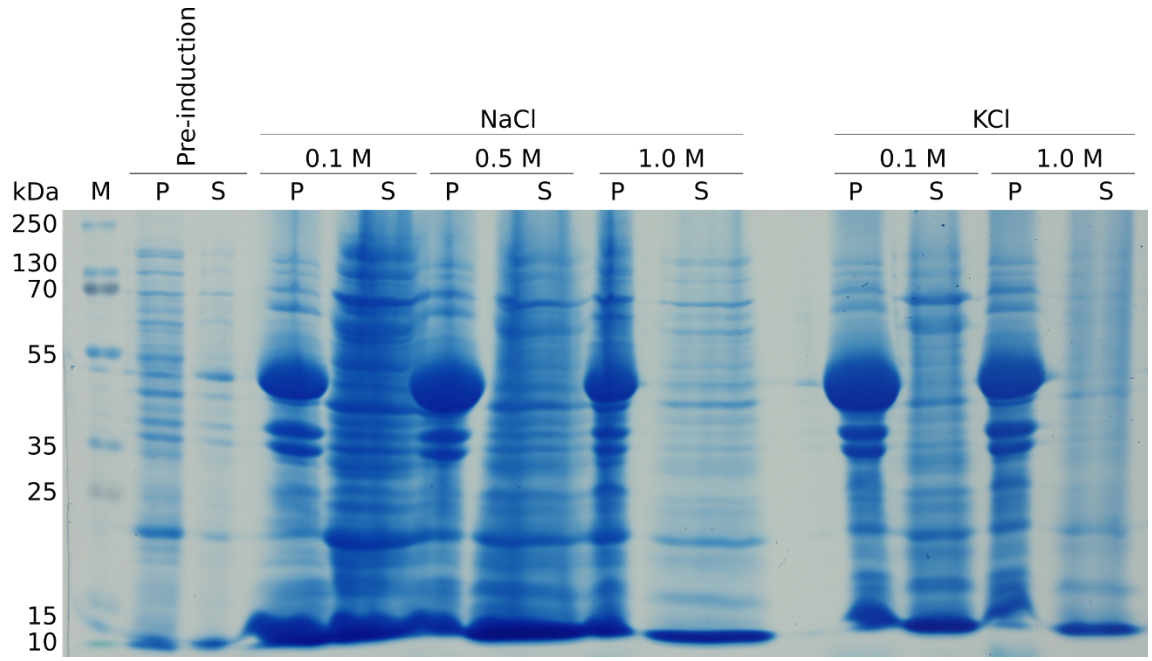


Figure 2-7: Solubility assay for phiFa_Δ1-19_Δ447-467_715 with NaCl and KCl.

P – pellet. S – supernatant. M – marker.

Table 2-18: Summary of expression of constructs

Result	Proteins
Not expressed	• D6E_Δ1-34_Δ388-420_651 • D6E PP Δ375-420_655 • D6E PP Δ1-34, Δ375-420_657
Insoluble	• phiFa_Δ1-19_Δ447-467_715 • PhiOH2_Δ436-474_685 • TSP4_FL_773 • TSP4_Δ1-43_777
Low solubility	• RM378_FL_772
Expressed	• PhiFa_FL_340 • GVE2_FL_669 • GVE2_Δ1-18_676 • GVE2_Δ1-18_686 • BV1_FL_717 • GBSV1_FL_718 • GBSV1_Δ1-40_775

2.3.6. Purification

The seven proteins that expressed well (Table 2-18) were taken forwards for purification. The PhiFa_FL_340 construct purified well, but two bands of different weights eluted from the SEC column (Figure 2-8). The N-terminal hexahistidine and SUMO tags were removed using 3C protease between affinity chromatography and SEC, so the subunit should weigh ~ 52 kDa. Based on the elution of reference molecules from the Superose 6 Increase 10/300 column, the ~18 mL could contain subunits and the low peak at ~ 11 mL could contain whole 12mers (~ 624 kDa). The sample of the 18 mL peak was taken forwards for TEM imaging (see below).

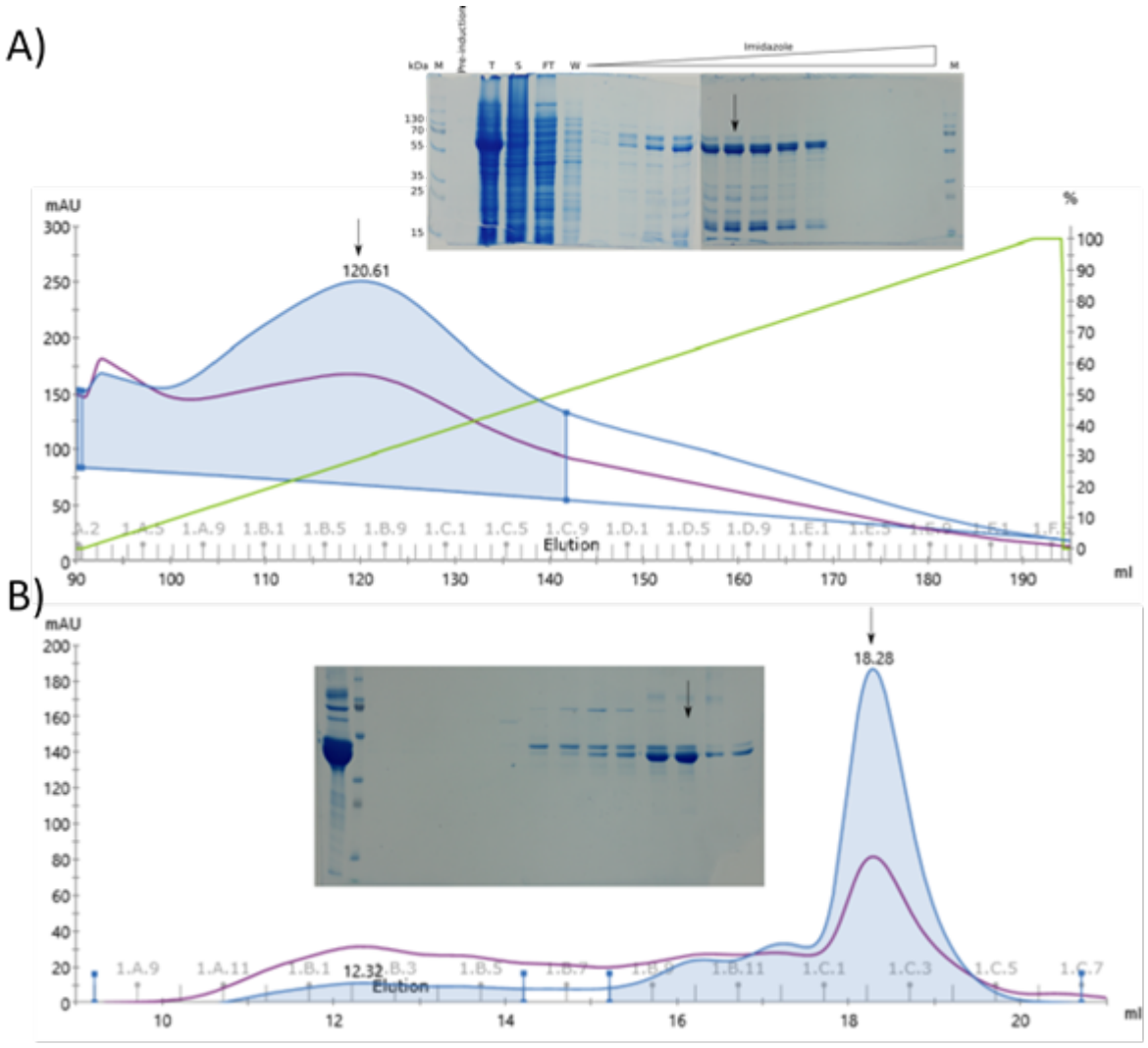


Figure 2-8: Purification of the phiFa_FL_340 protein

A standard two-step protocol was used. A) Nickel affinity chromatography with a 5 mL FF HisTrap column. The wash buffer was 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 10 % glycerol, 30 mM imidazole, 10 mM KGlu. The elution buffer was the same, but with 0.5 M imidazole instead. B) Size-exclusion chromatography with a superpose 6 increase 10/300 GL column. The buffer was 50 mM Tris-HCl, 0.25 M NaCl pH 7.5.

Very little of the GVE2_FL_669 protein binds to the Ni^{2+} resin. Most of it passes straight through into the flow-through fraction (Figure 2-9). The small quantity that does bind, also includes a lot of other molecules that bind non-specifically. The BV1_FL_717 failed to bind the Ni^{2+} resin. The results resemble (Figure 2-9).

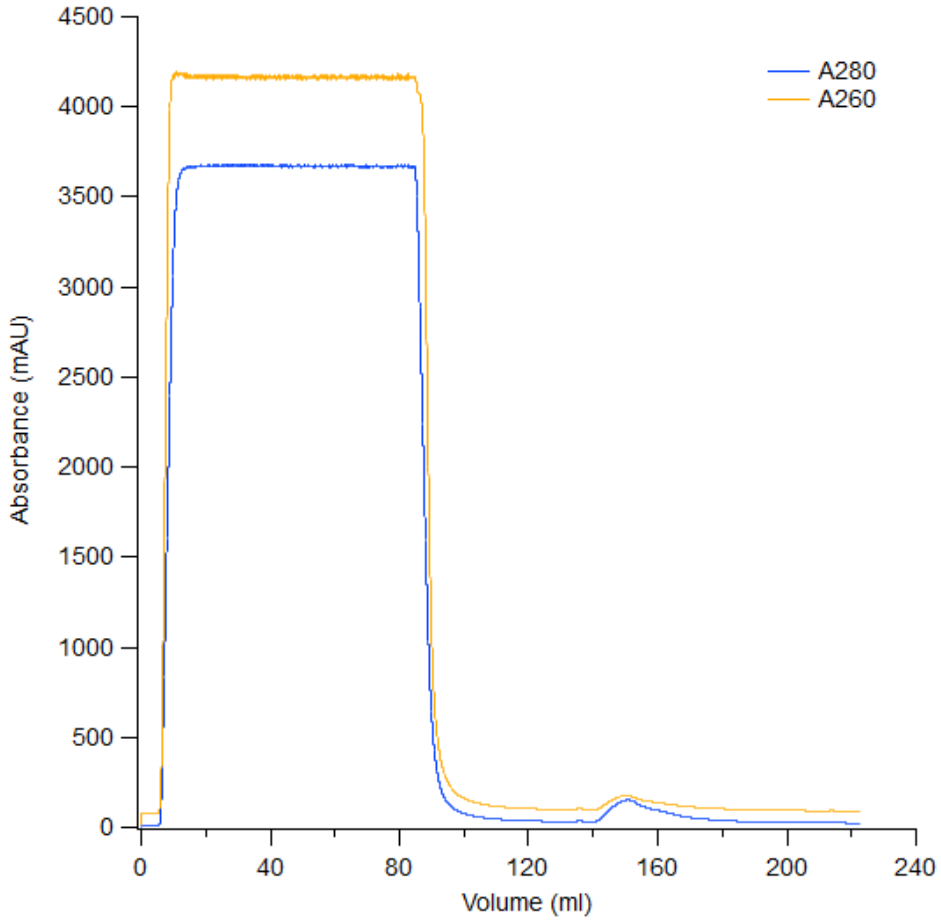


Figure 2-9: Purification of the GVE2_FL_669 protein

The GVE2_FL_669 protein does not bind to the Ni resin in the HisTrap columns. Similar results were seen for the BV1 PP (FL) protein. For the hexahistidine tags to not bind, they must be obstructed or already binding something else.

Two different methods were used to purify GVE2_Δ1-18_676 protein. The first method () used a spin concentrator as usual. The second used ammonium sulphate precipitation (Figure 2-11).

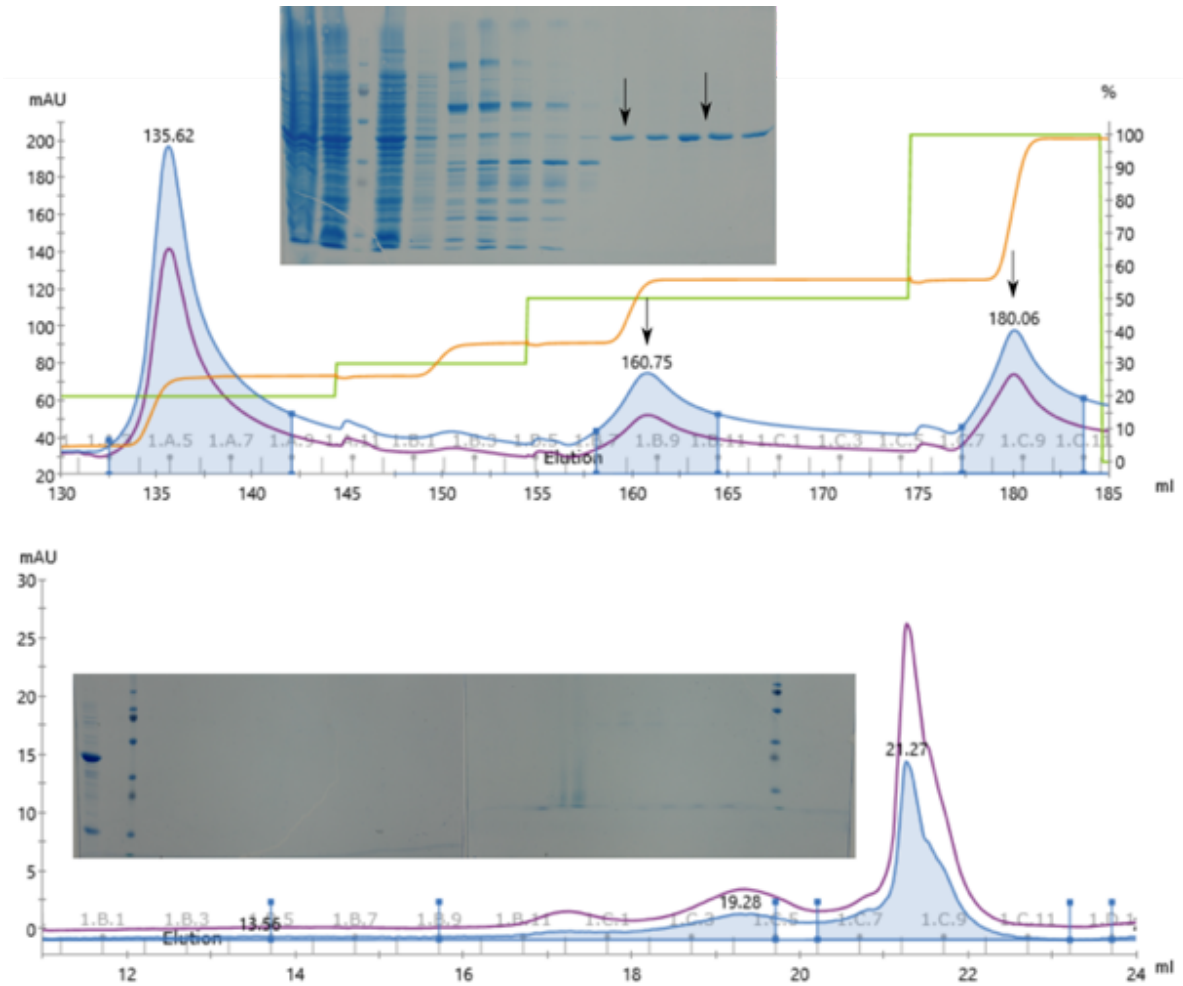
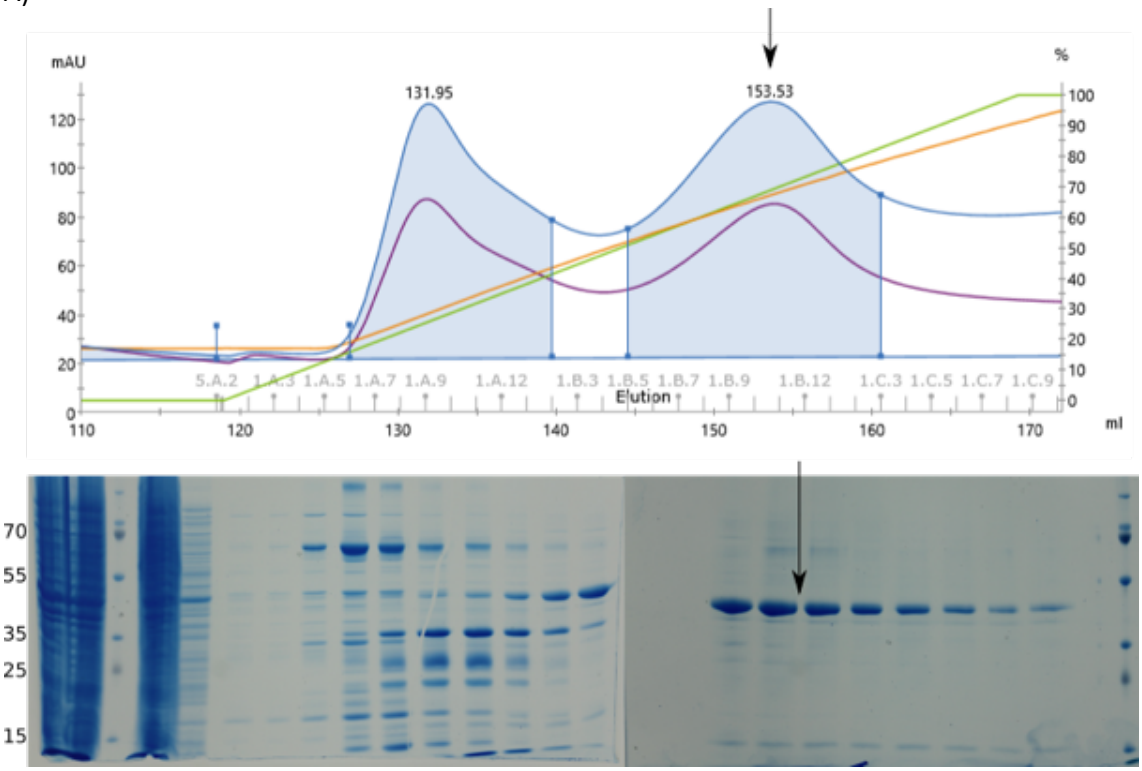


Figure 2-10: Purification of the GVE2_ Δ1-18_676 protein

Two separate purification attempts were made for the GVE2_Δ1-18_676 protein. The first attempt removed the N-terminal tag and used a spin concentrator before SEC (C). It elutes from the HisTrap column cleanly using a step elution gradient. The protein was present until it was loaded onto the SEC column. User error is possible.

The longer N-terminal histidine tag on the GVE2_Δ1-18_686 protein should help it bind more strongly to the Ni^{+} resin and give a purer product. However, the second HisTrap after dialysis with 3C protease, showed that very little of the tag had been removed (Figure 2-12).

A)



B)

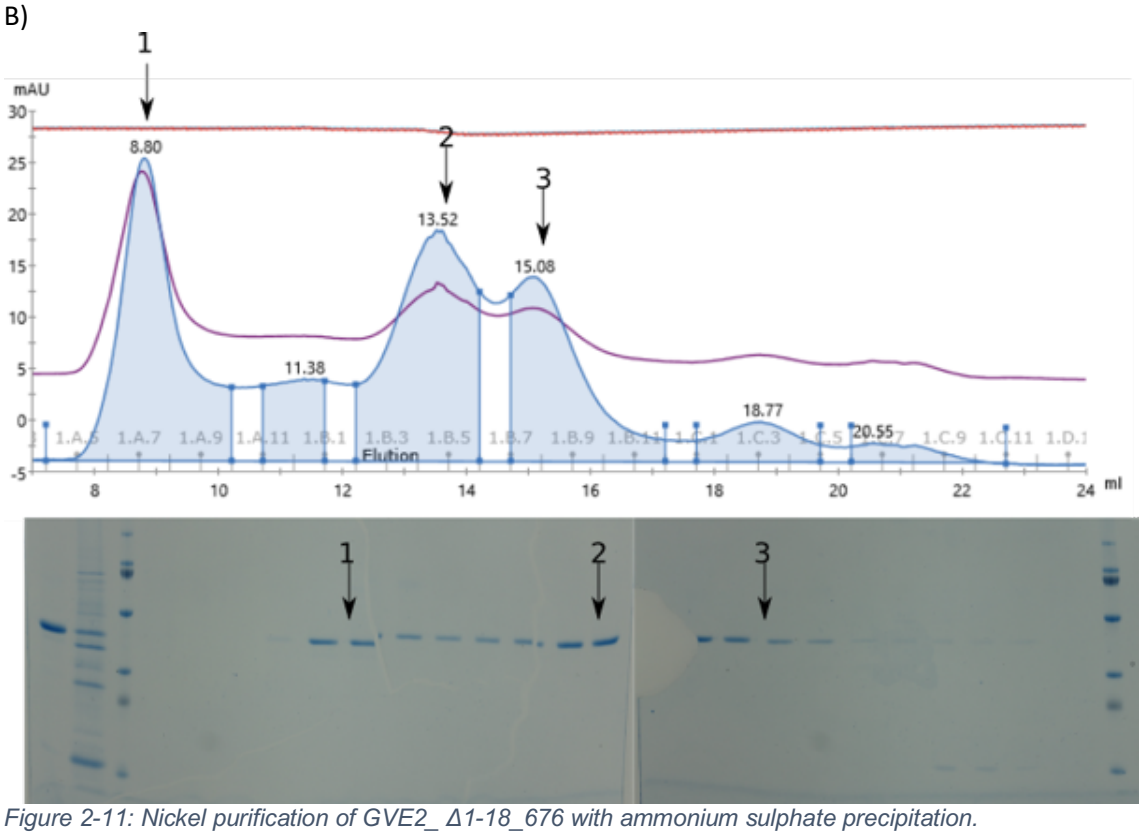


Figure 2-11: Nickel purification of GVE2_ Δ1-18_676 with ammonium sulphate precipitation.

The protein eluted cleanly from the HisTrap columns, though less so than with the step gradient. However, when it was further purified by SEC, several peaks appeared throughout. The SDS-PAGE gels show that the protein elutes across a wide range. These could be different oligomers, aggregates (especially for the first peak).

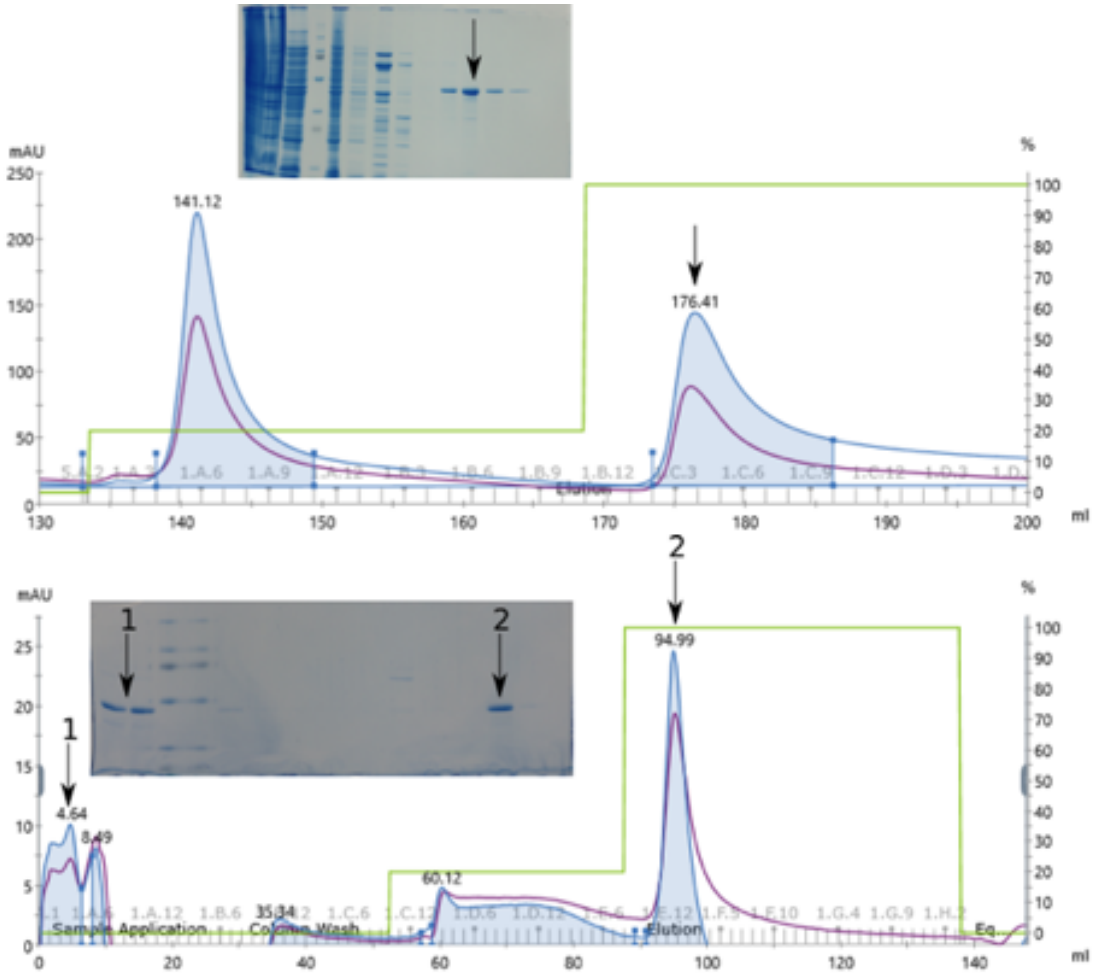


Figure 2-12: Purification of GVE2_Δ1-18_686

The purification of GVE2_Δ1-18_686 used a step gradient. It eluted cleanly from the first HisTrap purification (top), but after dialysis and tag cleavage, it appeared that very few of the tags had been removed (bottom). The cleavage site may have been obscured, or more 3C was required.

Very little of the GBSV1_FL_718 protein bound to the HisTrap column (Figure 2-13A). The small quantity that does bind, elutes with other molecules. When it is concentrated, there is a strong band for the GBSV1_FL_718 protein; however, it is lost during SEC, or the quantity was too small. By contrast, purifying it using a heparin column gives a single strong peak, though it is quite wide. When this pool was purified further by SEC, it gave one peak with significant tailing. There was also a large absorbance signal at 260 nm. This signal likely means that a large amount of nucleic acid has bound the protein and eluted with it, despite the addition of DNase and RNase to the lysis buffer (Figure 2-13B).

A) HisTrap purification

B) Heparin purification

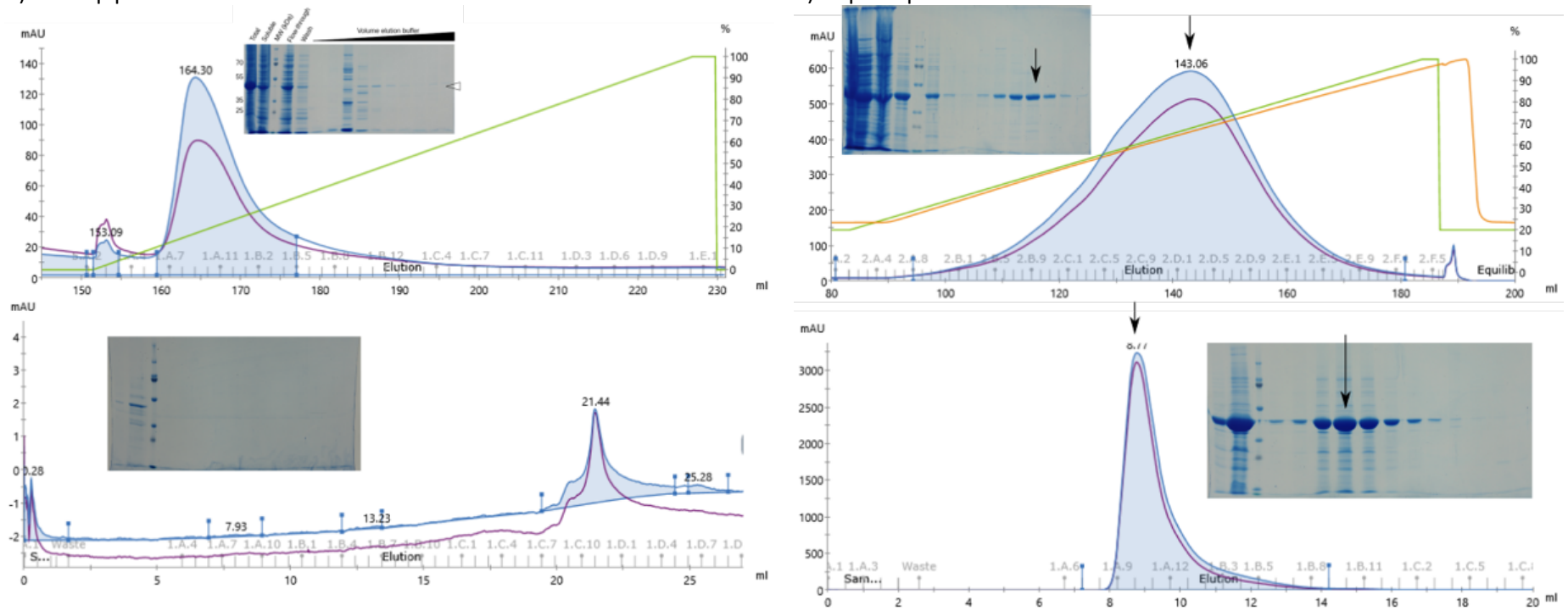


Figure 2-13: Purification attempts of GBSV1_FL_718

Two different strategies were used to purify the protein. A) Nickel affinity chromatography using a HisTrap column. B) Affinity chromatography using a heparin column.

The GBSV1_Δ1-40_775 purification used a pellet from 6 L of cell media. As such, there was a great deal more protein in this purification than the others (Figure 2-14). However, despite the increased quantity, the HisTrap separated the protein well, as seen in the chromatogram (Figure 2-14). The expected size of the GBSV1_Δ1-40_775 is ~ 45 kDa. According to the elution volumes of standards from a HiLoad superdex 16/60 200 pg column (Cytiva 2020), it should elute at ~ 80 mL. The 12mer would be ~ 540 kDa and elute at ~ 50 mL. Therefore, it appears that the GBSV1_Δ1-40_775 elutes as individual subunits. Some assemblies may be present.

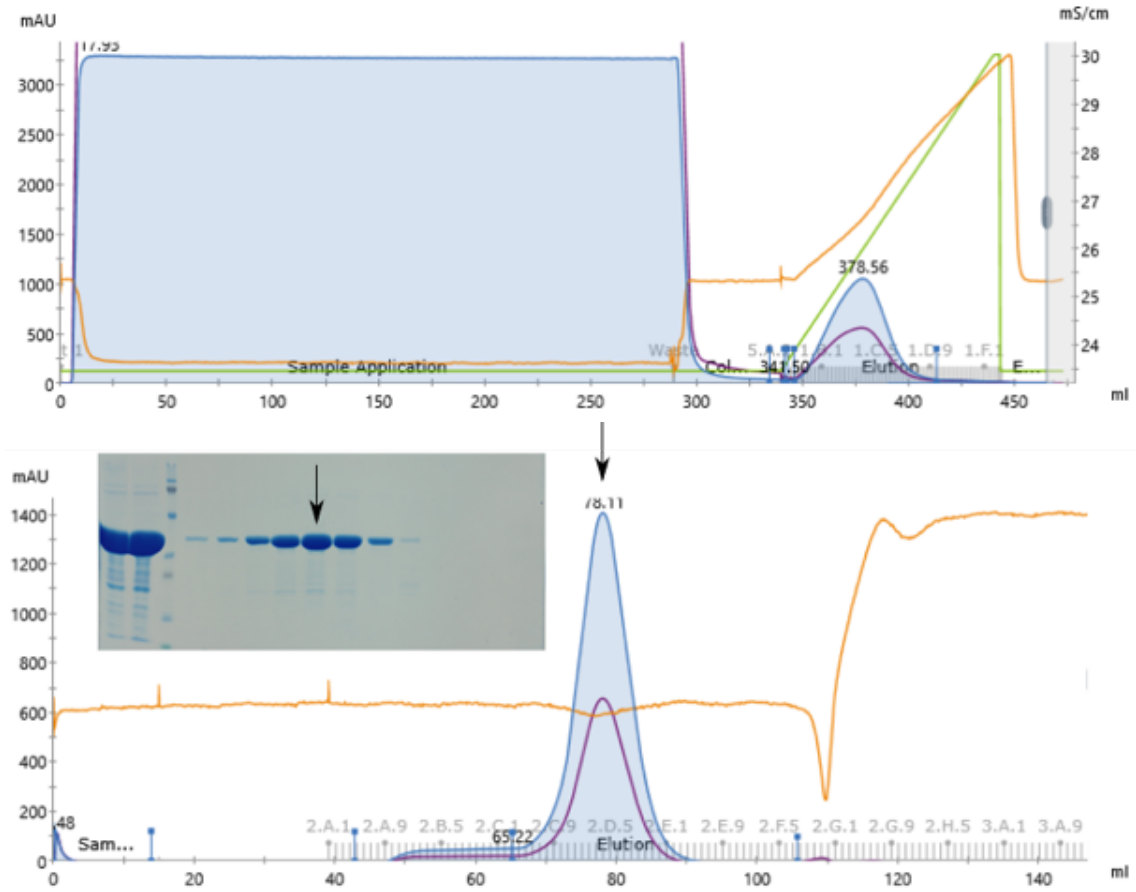


Figure 2-14: Purification of GBSV1_ Δ1-40_775

Top- affinity chromatography using two 5 mL FF HisTrap columns in tandem. Bottom – SEC using a HiLoad superdex 16/60 200 pg.

Table 2-19: Summary of the results of portal protein purifications

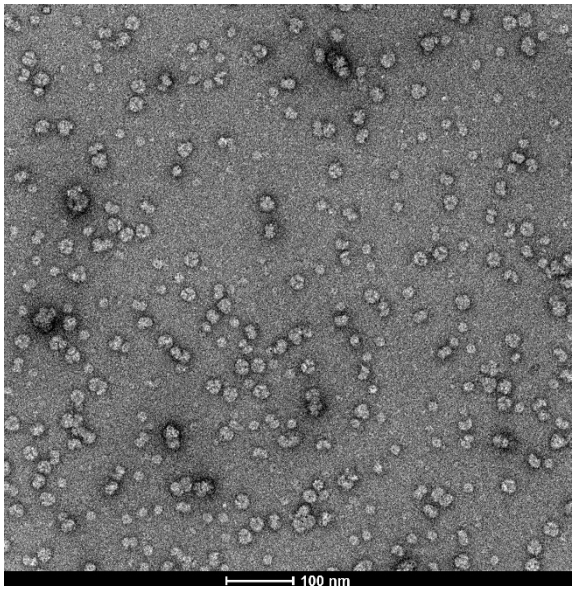
Construct	Purification result
PhiFa_FL_340	Two bands elute from SEC
GVE2_FL_669	Cannot bind Ni resin
GVE2_Δ1-18_676	Elutes from SEC across a wide range
GVE2_ Δ1-18_686	Can bind Ni resin, but cannot cut off the longer tag
BV1_FL_717	Cannot bind Ni resin
GBSV1_FL_718	Co-elutes with large quantity of nucleic acid
GBSV1_Δ1-40_775	Elutes cleanly, in large quantities

2.3.7. TEM images of the purified proteins

The purified portal proteins were imaged by TEM using negative stain to determine if they aggregated and find their oligomeric state. Grids were prepared for the phiFa_FL_340, GVE2_Δ1-18_676 and GBSV1_Δ1-40_775. The grids for the phiFa_FL_340 were made for samples at the centre of the peak, at the start of the shoulder (fraction B2), and the near shoulder (B11) (Figure 2-8A). The images were taken as 49k magnification. Proteins are not clearly visible on the TEM images (data not shown). The concentration of the protein purified may have been too low, or they did not form oligomers.

The negative stain images for GVE2 are of peak 2 (fractions B3-5) (Figure 2-15). The left image is taken at 49k magnification. The right image is taken of a different area at 68k magnification. In both, there are clearly protein aggregates. Many of the aggregates appear to be made of three oligomers bound together. They may be made of four proteins, forming a tetrahedral shape, with the fourth protein underneath. The exact reason behind this aggregation is unclear.

A) 49k magnification



B) 68k magnification

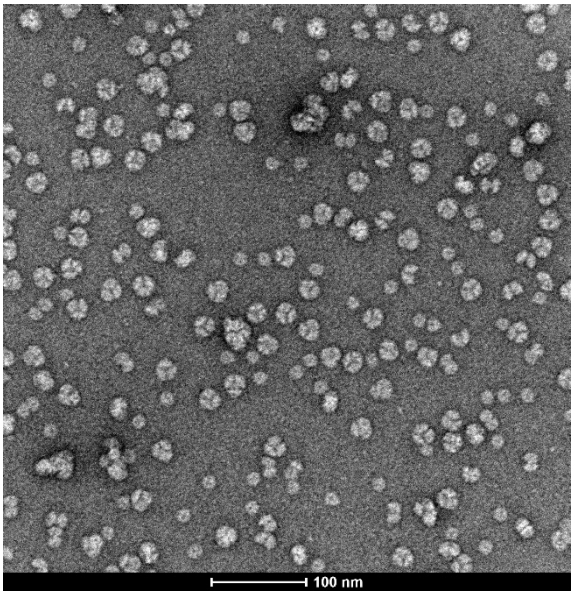
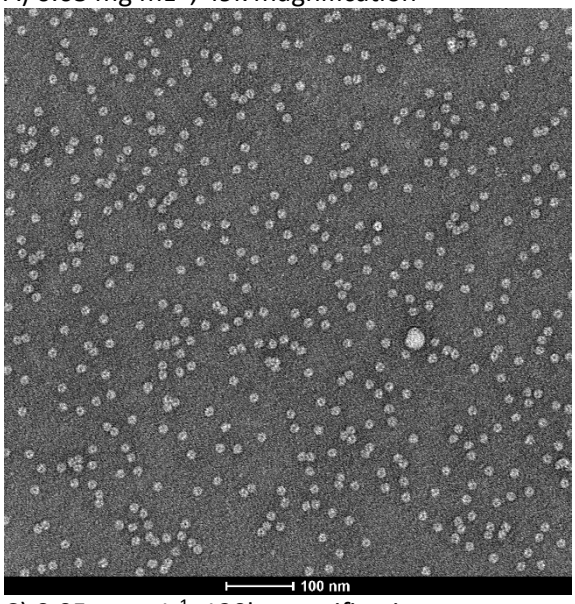


Figure 2-15: TEM images of GVE2_Δ1-18_676

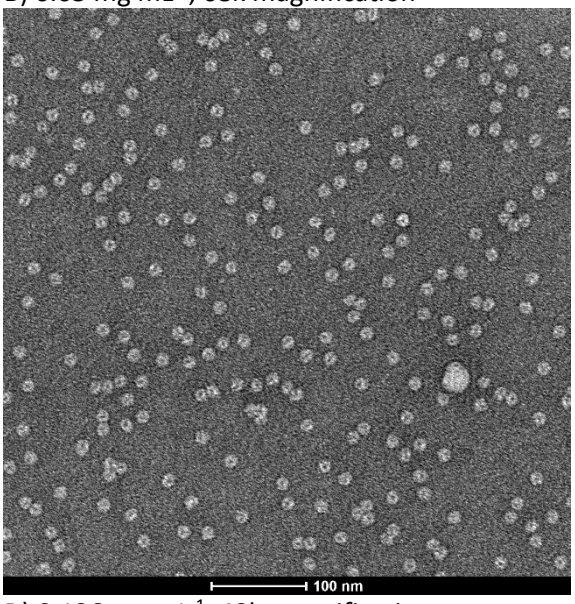
Two images of peak 2 of Figure 2-11. A) 49k magnification. This grid was made by Maria Chechik as an independent control. B) 68k magnification.

The negative stain images of GBSV1_Δ1-40_775 are from the centre of the peak (Figure 2-14). The images are of the same area at different magnifications. The protein appears to have formed oligomers, even though they eluted from the SEC column as individual subunits.

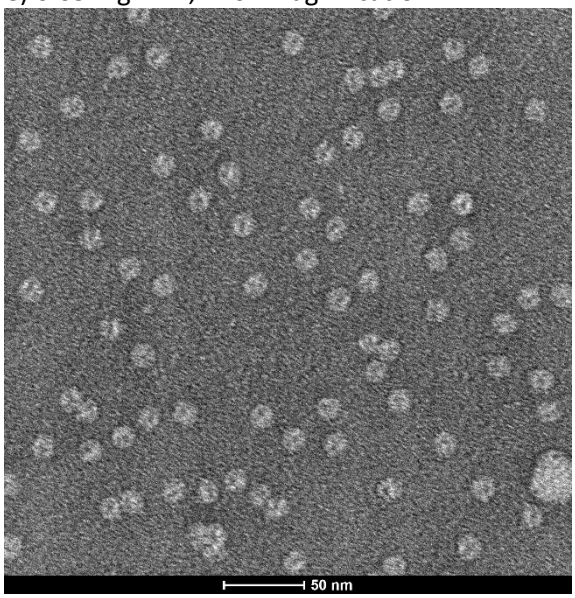
A) 0.05 mg mL^{-1}, 49k magnification



B) 0.05 mg mL^{-1}, 68k magnification



C) 0.05 mg mL^{-1}, 120k magnification



D) 0.136 mg mL^{-1}, 49k magnification

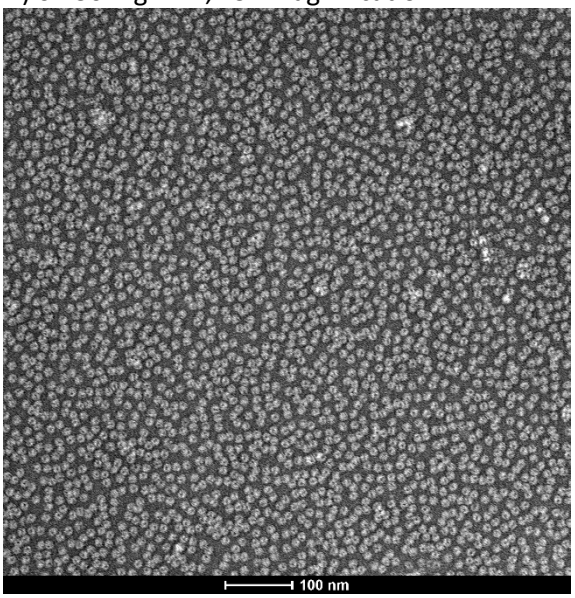


Figure 2-16: TEM images of GBSV1_Δ1-40_775.

The TEM images show successfully purified protein, but also strange aggregates. The defocus for all images shown was -500. A-C) The same area of the same grid made using the C3 fraction. The protein concentration was 0.05 mg mL-1 GBSV1_Δ1-40_775. A) 49k magnification. B) 68k magnification. C) 120k magnification. D) A TEM image of a different grid made using the same fraction from SEC. The concentration was 0.136 mg mL-1. The magnification was 49k.

Table 2-20: Summary of constructs

Gene (accession number)	Expression	Solubility	Purification	TEM
PhiFa_FL (pym340)	X	X	X	
D6E_Δ1-3_Δ388-422 (pYM649)				
D6E_Δ1-34_Δ388-422 (pYM650)				
D6E_Δ1-34_Δ388-422 (pYM651)				
D6E_Δ375-422 (pYM655)				
D6E_Δ1-34_Δ375-422 (pYM657)				
GVE2_FL (pYM669)	X	X		
GVE2_Δ1-18 (pYM676)	X	X		
PhiOH2_Δ436-474 (pYM685)	X			
GVE2_Δ1-18 (pYM686)	X	X		
phiFa_Δ1-19_Δ447-467 (pYM715)	X			
BV1_FL (pYM717)	X	X		
GBSV1_FL (pYM718)	X	X		
RM378_FL (pYM772)	X			
TSP4_FL (pYM773)	X			
GBSV1_Δ1-40 (pYM775)	X	X	X	X
TSP4_Δ1-43 (pYM777)	X	X		

Each X marks the construct's fulfilment of that stage of production. Most of the constructs could be successfully expressed and most of those were soluble. The stage which removed the most constructs was purification.

2.3.8. Mutants for nanopores

The successfully purified portal protein GBSV1_Δ1-40_775 could be taken forwards for a cryoEM structure (see Chapter 3). It could also be tested as a hybrid nanopore (see Chapter 5). For it to be used as a biological nanopore, a cysteine residue had to be added under the wing domain (see GBSV1_Δ1-40_S135C_792, Table 2-15). Different mutations were made to make different variants of the protein for testing as hybrid nanopores. A mutant was made missing the last helix of the crown domain (see GBSV1_Δ1-40_ Δ383-406_816, Table 2-15). The disordered structure seen in the

AlphaFold prediction (Figure 2-5) and confirmed in the cryoEM structure (see Chapter 3) may allow it to flap against the mouth of the pore, causing noise and blocking the entry of analytes. A cysteine was added to the clip domain to allow the attachment of dsDNA to pull the portal into the SSN as done previously with α-hemolysin (Hall et al. 2010a). Mutations were made on the inside and outside of the clip domain (see GBSV1_Δ1-40_D213N_E247R_E250Q_E252_774 and GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822, Table 2-15 and Figure 2-4) to make it more neutral or positively charged in order to give the protein a dipole. Other mutants were made that combined different combinations of previous mutants (see GBSV1_Δ1-40_ Δ383-406_S135C_821, GBSV1_Δ1-40_ Δ383-406_A215C_825, and GBSV1_Δ1-40_A215C_P282Q_836). Some mutants could not be purified at all and require further optimisation. Table 2-15 shows all of the mutants of GBSV1_Δ1-40_775 made for nanopores. No TEM images were collected for any of the mutants.

GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q_774, GBSV1_Δ1-40_A215C_776 and GBSV1_Δ1-40_ Δ383-406_S135C_821 could not be purified in time to be taken to DreamPore to be tested as nanopores. However, they both express well in the same conditions as GBSV1_Δ1-40_775 (Figure 2-17). The GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q_774 mutant was less soluble than the original GBSV1_Δ1-40_775, as expected since it has four mutations in the clip domain. But it and GBSV1_Δ1-40_A215C_776 are well expressed and soluble enough for later purification.

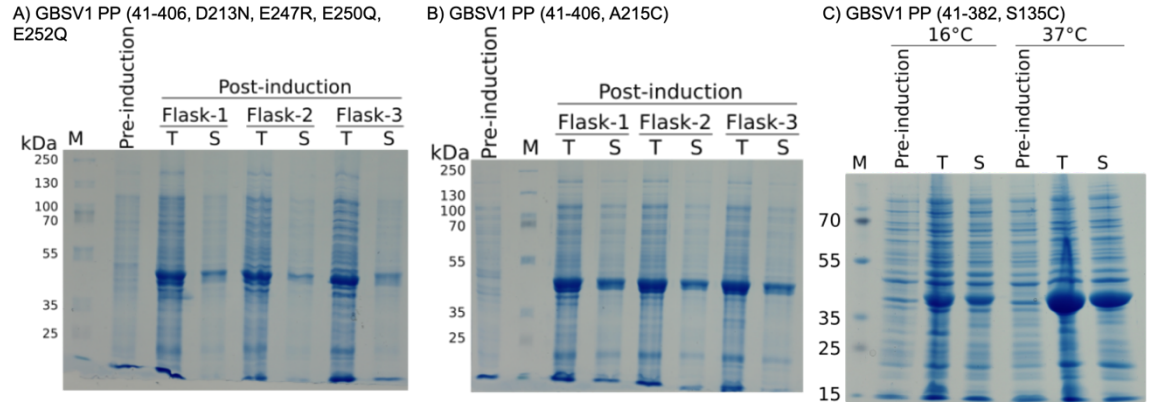


Figure 2-17: Expression of the GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q_774, GBSV1_Δ1-40_A215C_776, and GBSV1_Δ1-40_ Δ383-406_S135C_821 mutants

Both proteins express well and are soluble. The buffer was 50 mM Tris-HCl pH8, 0.25 M NaCl. The gels show expression and solubility for three flasks of a large-scale expression. A) GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q_774. B) GBSV1_Δ1-40_A215C_776. C) GBSV1_Δ1-40_ Δ383-406_S135C_821.

The GBSV1_Δ1-40_S135C_792 is the mutant made for adapting to a biological nanopore. The cysteine under the wing domain is for the attachment of the MPB PE lipid anchor. The purification protocol was the same as for GBSV1_Δ1-40_775, except that protein was eluted from the two HisTraps separately in order to maximise the yield from each column and reduce the volume of the pooled fractions. The protein purifies clearly in large quantities (Figure 2-18). Even with two columns, a lot of

protein fell through into the flow-through and wash fractions. Two peaks appeared in the SEC chromatogram (Figure 2-18B). The proteins in the two peaks appear at the same level on the SDS PAGE gel. They may be the same protein in different conformations or oligomers.

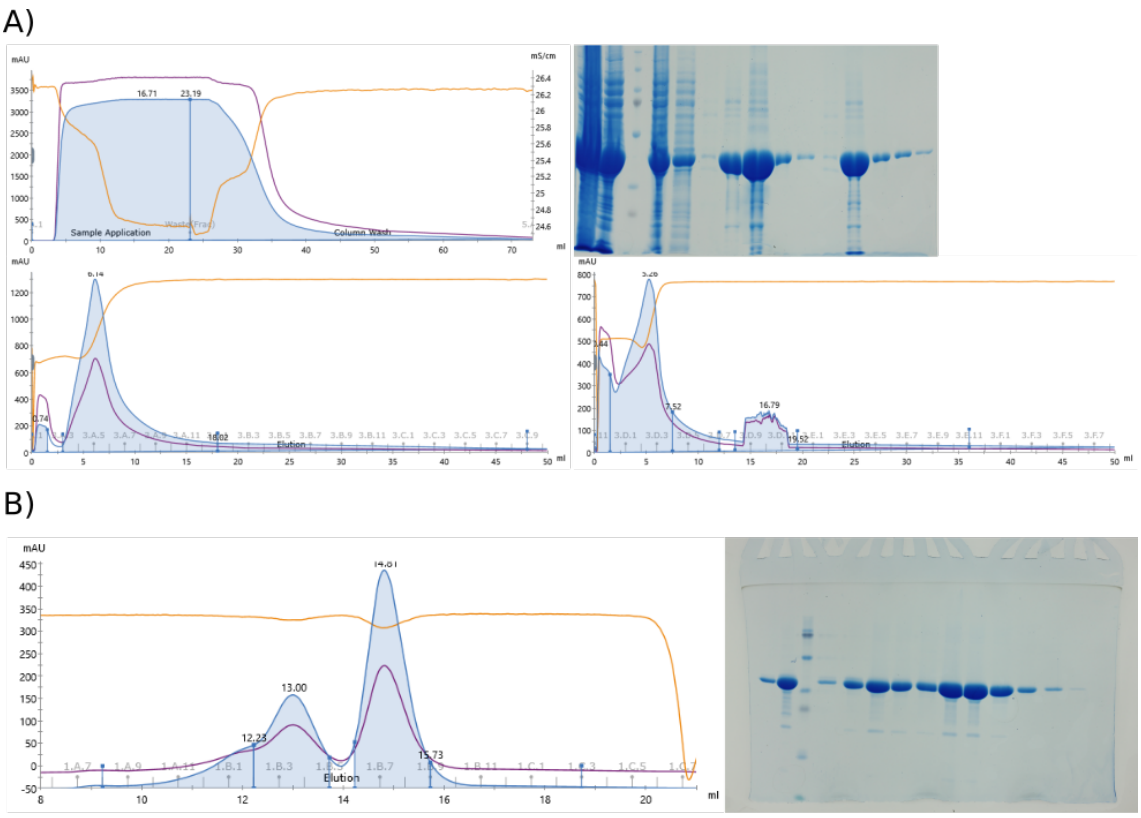


Figure 2-18: Purification of GBSV1_Δ1-40_S135C_792

The GBSV1_Δ1-40_S135C_792 protein was purified according to the methods in the Purification section. A) HisTrap. Two 5 mL FF HisTrap columns were used joined together. The protein was eluted from them separately, making two separate chromatograms (A-lower plots). A step gradient of elution buffer from 0 straight to 100% was used to elute as much of the protein as possible in a small number of fractions. B) SEC purification. Two peaks eluted from the SEC column. The second, heavier peak was taken forwards. The first peak has a slight shoulder on its left side. The bands on the gel appear to have the same weight. They could be the same protein in two different conformations or oligomers.

The GBSV1_Δ1-40_ Δ383-406_816 was purified the same way as previous mutants. The SDS-PAGE gel of the SEC purification (Figure 2-19B) shows that the protein elutes with a lot of other smaller proteins. The gel was overloaded. Multiple rounds of SEC may have given a cleaner purification. The chromatogram shows a very wide peak or rather at least 3 peaks that overlap giving the appearance of one peak with two asymmetrical shoulders. For these proteins to be of use as nanopores, they must be able to tolerate various NaCl and KCl concentrations. GBSV1_Δ1-40_ Δ383-406_816 is soluble in all NaCl concentrations tested up to 4.75 M, but the solubility decreases above 2 M. It is also soluble in all KCl concentrations tested, but both the lysate and soluble fraction samples get significantly fainter after 1 M.

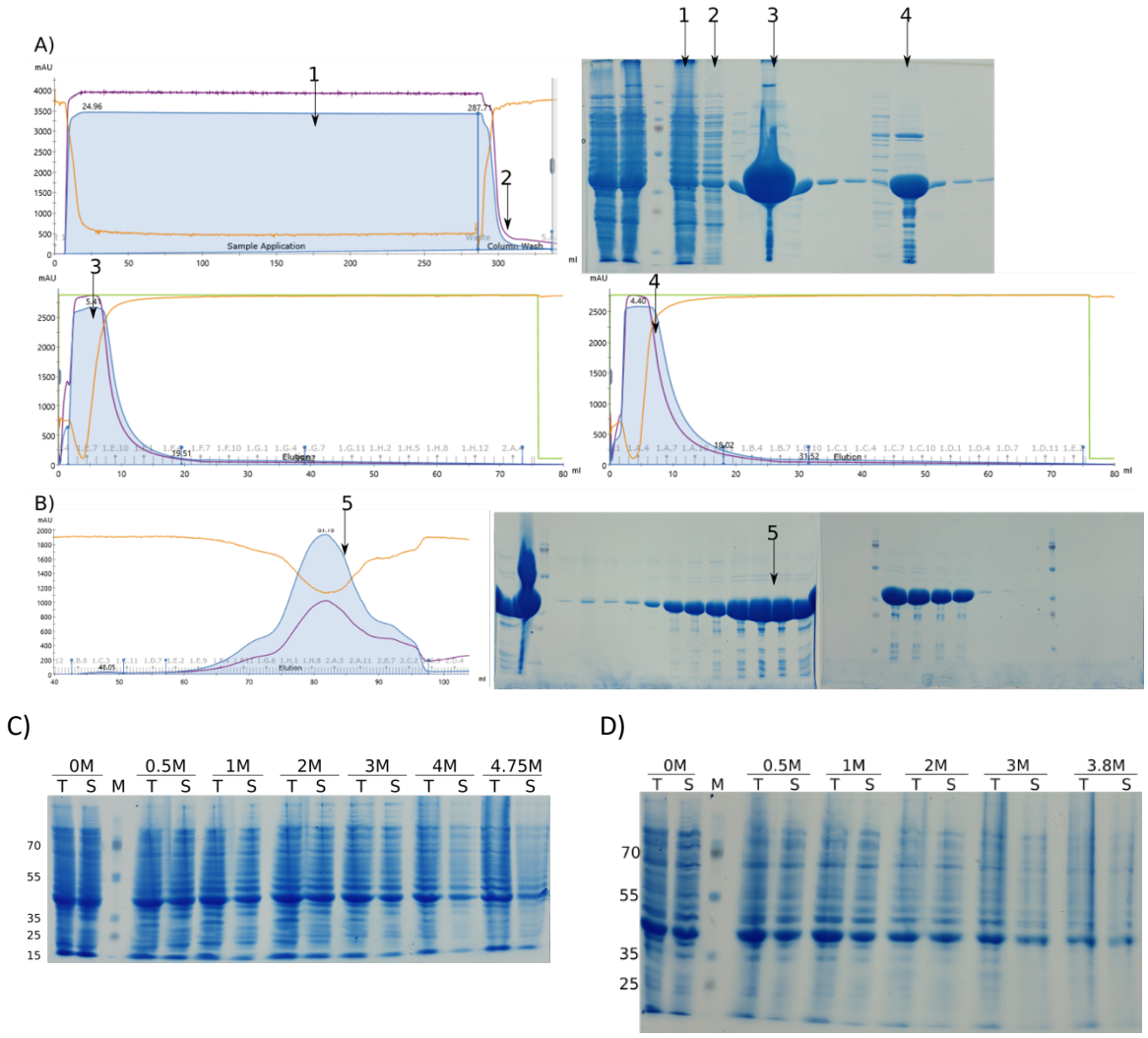


Figure 2-19: Purification of the GBSV1_Δ1-40_ Δ383-406_816 mutant

A-B) Purification. C-D) Solubility tests in different concentrations of NaCl (C), and KCl (D).

The GBSV1_Δ1-40_P282Q_820 was purified the same as previous mutants, except that the two HisTrap columns were eluted using a gradient (Figure 2-20A). The first SEC purification was over loaded with protein, giving a single very wide peak (Figure 2-20B). Based on previous purifications, the peak should be at ~15 mL for a superpose 6 increase 10/300 column. A 2nd round of SEC with a larger SEC column was used to separate the large quantity of protein. Like previous mutants, with the HiLoad column, it eluted at ~80 mL. It also eluted with other, smaller proteins.

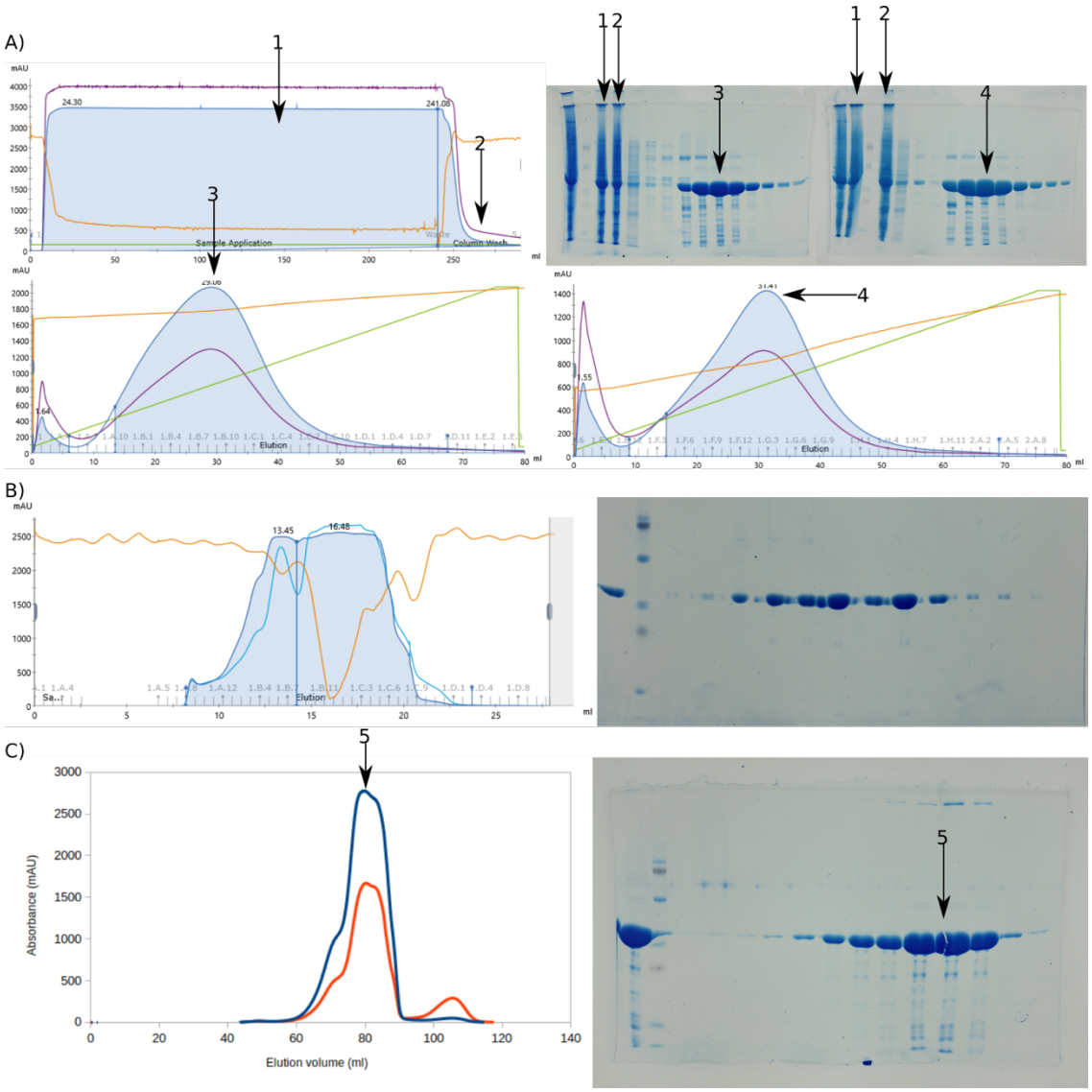


Figure 2-20: Purification of the GBSV1_Δ1-40_P282Q_820 mutant

A) HisTrap. B) SEC with superpose-6 increase 10/300 column. C) Additional round of SEC with HiLoad superdex 16/600 200pg column.

The purification of GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 used the same method as the other mutants. However, this purification failed. A HisTrap column was used for affinity chromatography (Figure 2-21A), however, most of the protein did not bind the column. The SDS-PAGE gels show that some did, and it appears to be two separate bands that are close in weight. A heparin column was tested as an alternative (Figure 2-21B), which gave much better separation. The pooled fractions were dialysed overnight with 3C protease to remove the histidine tag (Figure 2-21C). The protein was not fully cleaved (Figure 2-21D). When it was concentrated to 3 mL, it became cloudy. It was filtered before being loaded onto a HiLoad superdex 16/600 200 pg column. However, it immediately aggregated and fell out of the column (Figure 2-21E).

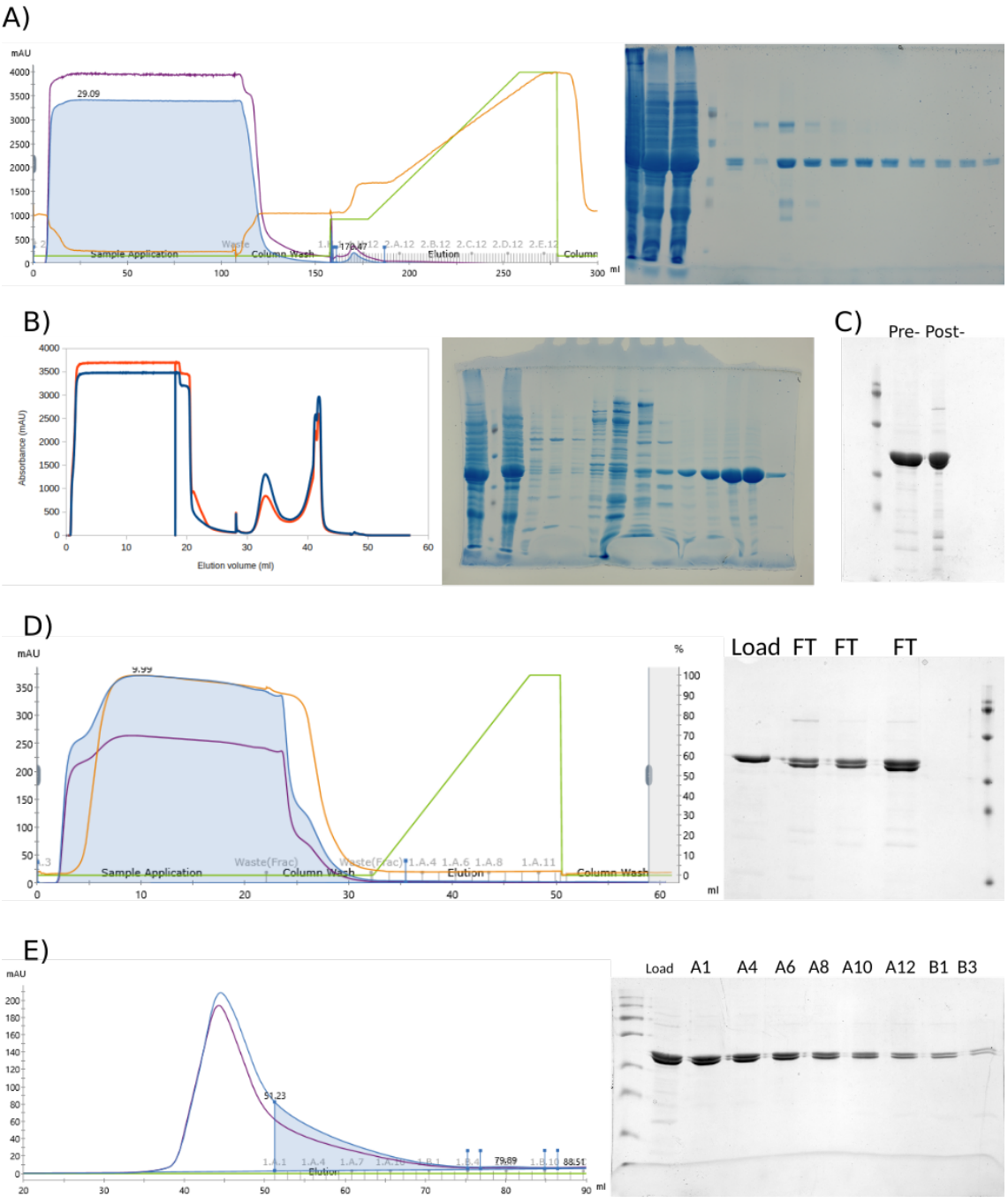
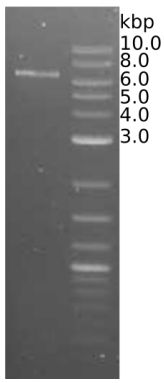


Figure 2-21: Purification of GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822

A) First attempted purification using a Ni2+ affinity chromatography as usual. Most-all of the protein did not bind the resin. B-E) 2nd attempt. B) heparin column. C) SDS-PAGE gel. D) HisTrap column. E) SEC.

A)



B)

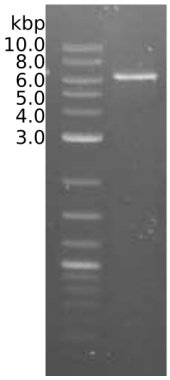


Figure 2-22: PCR of GBSV1_Δ1-40_Δ383-406_A215C_825 and GBSV1_Δ1-40_A215C_P282Q_836

These two mutants were only cloned within the time allowed. A) GBSV1_Δ1-40_Δ383-406_A215C_825. B) GBSV1_Δ1-40_A215C_P282Q_836.

Table 2-21: Nanopore mutants of the GBSV1_Δ1-40_775 protein

Gene	Accession number	Desc.	Expressed?	Purified?	Taken to DreamPore?
GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q	pYM774	Positive inside the clip. pt3/3	X	X	
GBSV1_Δ1-40_A215C	pYM776	With cysteine in clip domain for attachment of dsDNA	X	X	
GBSV1_Δ1-40_S135C	pYM792	With cysteine in the wing domain for insertion into a lipid membrane	X	X	X
GBSV1_Δ1-40_ Δ383-406	pYM816	The last helix of the crown domain was removed	X	X	X
GBSV1_Δ1-40_P282Q	pYM820	Narrower tunnel loops	X	X	X
GBSV1_Δ1-40_ Δ383-406_S135C	pYM821	The last helix of the crown domain was removed With cysteine in the wing domain for putting into membrane	X	n/a	
GBSV1_Δ1-40_E218R_E222R_E223R_E237Q	pYM822	Positive on outside the clip 3/3	X		
GBSV1_Δ1-40_ Δ383-406_A215C	pYM825	The last helix of the crown domain was removed With cysteine in clip domain for attachment of dsDNA	X	n/a	
GBSV1_Δ1-40_A215C_P282Q	pYM836	Narrower tunnel loop With cysteine in clip domain for attachment of dsDNA	X	n/a	

2.4. Discussion

Of the 16 portal protein mutants tested, containing eight separate portal proteins, only one of them was taken forwards for mutagenesis and structural determination (see Chapter 3): GBSV1_Δ1-40_775. The others could not be expressed, were not sufficiently soluble, could not be purified, or formed ordered arrays of multiple portal assemblies when imaged. The AlphaFold predictions give the portals the classical portal fold with the clip, stem, wing and crown domains. The crown domain appears ordered from the AlphaFold prediction (Figure 2-5).

2.4.1. Expression

All of the constructs tested managed to produce the desired recombinant protein, except for those using the D6E portal protein. The D6E constructs were made from a codon-optimised gene sequence for *E. coli*. A comparison of the original D6E gene and the codon optimised version has a 77% similarity. If the original gene sequence was used instead, as was done for the other portals, the results may have been different. Codon usage affects the transcription of mRNA (Gouy and Gautier 1982). However, the use of Rosetta cells should have removed the need for codon-optimisation. Thus, it is possible that the original D6E gene would also not express. An alternative method of protein expression would be to export them from the *E. coli* cells in vesicles (Eastwood et al. 2023). Exporting the protein from the cell would avoid a high local concentration which should prevent aggregation. This method also promises high protein yields. It should be tested on a reliable protein such as GBSV1 before mass adoption with other constructs. Alternatively, the entire virus could be produced and the portal protein purified from them. The whole virus could be made from *E. coli* or insect cells (Machón et al. 2019). The YSBL3CLIC+ vector used throughout this work is based on the pET28a vector from the pET series of vectors. These vectors have been used for almost thirty years (Rosenberg et al. 1987; Studier et al. 1990). Variations on these pET vectors which use the original T7 promotor and a modified translation initiation region may give improved protein expression (Shilling et al. 2020).

2.4.2. Solubility

The solubility of a protein will affect how easily it can be purified and its uses downstream, e.g. nanopores. The RM378 constructs, phiFa_Δ1-19_Δ447-467_715 protein and TSP4 constructs could be expressed, but were not very soluble, or totally insoluble (Figure 2-6). There are a variety of different possible causes. The solutions may depend on the individual protein.

This may be fixed by changing the expression conditions. Lower expression temperatures slow down cell processes, which can allow proteins more time to fold correctly. However, all of the methods used already use 16 ℃. Other portal proteins had to have their N- or C-terminal ends removed in order to make them soluble (Sun et al. 2015b). The ColabFold prediction shows these to be one long disordered

region (Figure 2-5). Other portal proteins that have been purified with their entire N-terminal lack the density to model it (Bayfield et al. 2019), indicating that the disordered N-terminal region may wrap around the outside of the wing domains of the portal and/or interact with the major capsid proteins. Solubility could be improved by adding KGlu to the buffer to help the protein fold correctly (Cheng et al. 2016). Co-expression with the relevant scaffold protein has previously helped in making protein crystals (Luan 2013). Scaffold proteins usually help portal proteins form oligomers during capsid assembly. The G20c phage is a close relative of phiFa and TSP4 as they are all part of the *oshimavirus* genus (Schoch et al. 2020). The G20c portal protein and various mutants have been expressed and purified previously (Cressiot et al. 2017a). In this case, a truncation was produced comprising residues 25-438. The last 10 residues were removed. The phiFa_Δ1-19_Δ447-467_715 construct is missing the first 19 and last 20 residues. The phiFa_FL_340 construct was shown to be soluble, but this could be due to its N-terminal SUMO tag, rather than having a complete sequence. The buffer used in purification after affinity chromatography was 50 mM Tris pH 8, 500 mM NaCl, 50 mM KGlu. The phiFa_FL_340 construct was also purified in a buffer containing KGlu. There is much less literature to draw on for RM378. It is the only thermophage tested here that infects *rhodothermus* bacteria, while the others infect *geobacillus* or *thermus* bacteria.

2.4.3. Purification

2.4.3.1. phiFa_FL_340

The PhiFa_FL_340 construct could almost be purified, except that it appeared to purify as subunits and in two forms. The two bands that appeared on the gel of the phiFa_FL_340 purification may be from subunits with or without the tag. However, the weight of the tag is too small to give such a difference. The two bands can also be seen, albeit less clearly, in the SDS-PAGE gel for the HisTrap purification (Figure 2-8), so the two bands are not caused by incomplete cleavage of the N-terminal tag. If the higher band was being degraded, there may be a smear down the gel comprised of the protein in different stages of degradation. But the bands are clearly distinct. The lighter form is more prominent. There is a slight difference in the range at which the two bands elute. The higher, heavier band elutes earlier than the lower, lighter, more well-expressed band. The different proportions of the two forms in the fractions implies that they eluted from the column separately, as is partly seen in the small peak/shoulder at 17 mL, which means that some oligomers are mostly made of the heavier form and some are mostly made of the lighter form. The asymmetric distribution could just be caused by the greater presence of the lighter form.

Since SEC and SDS-PAGE gels separate things based on hydrodynamic radius, rather than strict size, they could be two slightly different conformations of the subunit. Only one band is seen in the SDS-PAGE gels of phiFa_Δ1-19_Δ447-467_715 construct, so it may be that the N- or C- terminal residues

can cause the portal to assume one of two conformations. If this was the case, removing the N-terminal and/or C-terminal unstructured regions may force it into a single conformation.

2.4.3.2. GVE2

Since the truncated versions showed promise for purification, no further work needs to be done on the GVE2_FL_669 protein. However, the BV1 protein, which, like the GVE2_FL_669 protein, did not bind to the Ni resin, should be tested again. The lack of binding implies obstruction of the N-terminal tag. However, a heparin purification was also attempted with BV1_FL_717 and gave similar results. The protein is soluble, so the subunits are not aggregating together. Another purification with a heparin column and a salt gradient gave similar results (data not shown).

2.4.3.3. GBSV1 mutants

Since the GBSV1_Δ1-40_775 mutant purifies easily, no further work is needed on GBSV1_FL_718; however, it is worth noting that GBSV1_FL_718 could only be purified with a heparin column. When it eluted from the SEC column, the absorbance signal at 260 nm was almost as strong as that for 280 nm, which suggests that the protein was bound to a lot of nucleic acids (Figure 2-13). This was despite the inclusion of DNase and RNase in the lysis buffer. Portal proteins in mature virus capsids are known to bind DNA during packaging (Lokareddy et al. 2020).

Removing the N-terminal tail from GBSV1 portal allowed it to be purified using a HisTrap (Figure 2-14), so it may be that these residues were involved in binding whatever nucleic acids were present in the purification of the FL version.

Most of the mutants of GBSV1_Δ1-40_775 behave the same as GBSV1_Δ1-40_775 during purification. However, there are some exceptions. The SEC purification of GBSV1_Δ1-40_S135C_792 shows two peaks (Figure 2-18B): at 13 mL and 14.8 mL. The 2nd, stronger peak is at the expected elution volume for GBSV1_Δ1-40_775. There is also a small peak at 12.23 mL which appears as a shoulder to the peak at 13 mL. The fractions for the two peaks show protein of the same weight on SDS-PAGE gel.

The SEC chromatogram for the purification of GBSV1_Δ1-40_ Δ383-406_816 shows a single large peak with several smaller peaks forming shoulders on either side.

2.4.4. Protein aggregation

The purification of GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 was the only one of the mutants that failed to purify. Most of the protein did not bind the Ni^{2+} resin in the initial HisTrap (Figure 2-21A). The small amount that did bind, appears to elute as two bands, close in size. A new purification, using a heparin column gave better binding and purification.

GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 was mutated to have positive residues on the outside of the clip domain, which may account for its ability to bind the heparin resin so well. It only began to aggregate when concentrated, so its aggregation may be concentration-dependent. Performing multiple runs with low sample concentration may prevent this aggregation appearing again. The SDS-PAGE gel shows two bands, much like for the purification of phiFa_FL_340 (Figure 2-8), though these appear to be much closer in weight. Its aggregation could be studied using nanoDSF.

Many of the SEC purifications showed overlapping peaks. Many of the purifications were for large quantities of protein. When highly concentrated before being applied to a SEC column, the viscosity of the sample could increase, which may cause peak broadening and a loss of resolution. Multiple runs of less viscous samples at lower flow-rates could prevent this.

The strange aggregates seen in the negative stain images of GVE2_Δ1-18_676 appear to be made of three portals joined together. There may be a fourth underneath, forming a tetrahedral assembly. If some Ni^{+} ions had eluted with the protein, as is possible, then the tags may be mediating aggregation. But, the protein was not aggregated when they came off the SEC column, since they eluted at ~ 15 mL, which is approximately the expected volume. The aggregation must have happened between being pooled together and concentrated after SEC, and being applied to negative stain grids. The N-terminal hexahistidine tags were not removed from this sample. This purification could be repeated but with the tags removed, to rule out any role they might play in aggregation.

TEM images should be collected of the different peaks in the SEC chromatograms to start to identify them. They should be isolated separately and tested to see if they behave differently from the expected 12mers.

2.5. Summary

Expression and purification was attempted on eight thermophage portal proteins. However, only the GBSV1 portal protein missing the N-terminal tail expressed readily and was soluble. Various mutants were made of the GBSV1 portal protein in anticipation of its use as a nanopore, including narrowing the tunnel diameter, adding positive charges to the clip domain, and removing the last helix of the crown domain. The other thermophage portal proteins could not be taken forwards for different reasons, e.g. low expression, insoluble, could not be purified, or aggregated. However, they may be successfully purified with further attempts, e.g. changing the buffer, expression vector, affinity tag or truncations.

Chapter 3 CryoEM structure of the GBSV1 portal protein

3.1. Introduction

In the previous chapter, a mutant of the GBSV1 portal protein (see Chapter 2) was purified in a soluble form lacking the first 40 amino acids of the N-terminal: GBSV1_Δ1-40_775. In order to maximise its use as a nanopore, the structure of the protein should be examined. Key parameters to investigate would be the dimensions of the tunnel, stability, and internal charge. However, the structure for this portal protein has never been solved before.

In the past, X-ray crystallography was the technique of choice for structural determination (Simpson et al. 2000; Lebedev et al. 2007); however, since the 'resolution revolution', structural studies of portal proteins used cryoEM instead. As such, many recent structural studies of portal proteins have used cryoEM, e.g. T4 phage (Sun et al. 2015a), GA1 phage, phiCPV4 phage (Javed et al. 2021), phiBB1 (Rūmnieks, Füzik, and Tārs 2023), P23-45 (Bayfield, Steven, and Antson 2020), HK97 phage (D. E. D. P. Hawkins et al. 2023; D. Hawkins 2022), Pap3 (David Hou et al. 2022) (Table 3-3).

However, the conformation of the portal protein depends on whether it is in solution (Mukherjee et al. 2024) or in the capsid, and whether that capsid is part of an immature (D. Hawkins 2022; Mukherjee et al. 2024; Zhiwei Gu, Wu, and Wang 2024) or mature virus (Mukherjee et al. 2024; Y. Peng et al. 2024; Zhiwei Gu, Wu, and Wang 2024). In addition, different parts of the portal protein structure can have different resolution due to their flexibility. As mentioned previously (see Chapter 1), the tunnel loops are inherently flexible, so can be hard to resolve (Javed et al. 2021; Sun et al. 2015a; Bayfield, Steven, and Antson 2020). However, the large size of the portal protein oligomer and its symmetry should work in its favour for cryoEM.

3.1.1. Chapter Aims

This chapter aims to solve the cryoEM structure of GBSV1_Δ1-40_775. Once the structure has been solved, the inter-subunit contacts and dimensions should be found. These will be important considerations for its use as a nanopore.

3.2. Materials and methods

3.2.1. Sample preparation

The grids were Ultra Au foil grid (R1.2/1.3, #227416) (Quantifoil). The grids were glow discharged using a Pelco device (Table 3-1).

Table 3-1: Glow discharge parameters

Parameter	Value
Pressure	0.38 mBar
Current	20 mA
Hold	10s
Time	

For cryoEM grid preparation, 3 µL of 1.3 mg mL^{-1} of pYM775 in 50 mM Tris-HCl pH 8, 0.25 M NaCl was added to the gold grid using FEI vitrobot instrument (Table 3-2).

Table 3-2: Vitrobot settings

Parameter	Value
Blot time	1 s
Wait time	8 s
Force	-10
Temperature	4 ℃
Humidity	100%
Sample volume	3 µL

Table 3-3: Summary of cryoEM conditions used for portal proteins

Bacteriophage	In situ/ isolation	Grid	Grid additives	Blot time (s)	Temperature (℃)	Humidity (%)	Images	Resolution (Å)	Citation
Lambda	*In situ*	Quantifoil Cu +2 nm C R1.2/1.3 300 mesh Holey carbon	Polylysine	3.5	8	100	3310	3.2	(Zhiwei Gu, Wu, and Wang 2024)
Staphylococcus aureus bacteriophage 80α	*In situ* Isolation	Lacy carbon		3.5	8	100	*In situ*: 2796 Isolation: 2694	*In situ*: 3.2 Isolation: 3.2	(Mukherjee et al. 2024)
T5	*In situ*	Quantifoil R2/1 copper	3nm layer of continuous carbon	4	8	100	3168	3.2	(Y. Peng et al. 2024)
T4	Isolation		0.3 % octyl-β-glucoside	7		85 <	866	3.6	(Sun et al. 2015a)
P23-45	*In situ*	Quantifoil R1.2			10	90	7758	1.95	(Bayfield et al. 2019)
PaP3	Isolation	Quantifoil R 1.2/1.3 holey carbon					FL-portal: 9351 ΔC-portal: 3189	3.4	(David Hou et al. 2022)
T7	Isolation	Quantifoil Cu R2/2 holey carbon grids		3.5	10	95		5.8	(Fàbrega Ferrer 2017)
GA1 phiCPV4	Isolation	Lacey carbon		3.5	8	100	GA1:2479 phiCPV4: 2903	GA1: 3.3 phiCPV4: 2.9	(Javed et al. 2021)

in situ refers to a cryoEM reconstruction of the portal that was inside the capsid, procapsid or mature phage. Isolation refers to the portal alone in solution.

3.2.2.

3.2.2. Data collection

The grid was imaged using the 200 kV Thermofisher Scientific Glacios instrument at the University of York. Data were collected at defocus values of -2, -1.8, -1.6, -1.4, -1.2, -1, -0.8, -0.6; one shot was collected per hole, with a total of 5,489 holes imaged out of 38 grid squares. The settings are shown in Table 3-4.

Table 3-4: Data collection parameters

Parameter	Value
Voltage (kV)	200
Detector	Falcon 4
Electron exposure ($e^-/Å^2$)	50
Defocus range	-2, -1.8, -1.6, -1.4, -1.2, -1, -0.8, -0.6
Pixel size (Å)	0.574

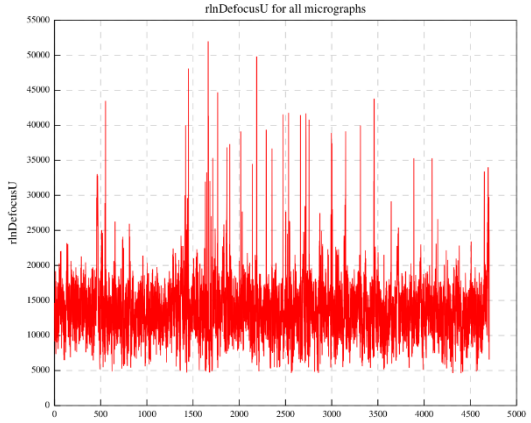
3.2.3. Data processing

RELION-4 was used for data processing unless otherwise stated. After movies were imported into RELION-4, they were sorted by time using a short script on YSBL wiki written by Dr Huw Jenkins (see Appendix). The workflow for data processing is shown in Figure 3-3. Motion correction was done using RELION's own MOTIONCOR2. CTF estimation was done using CTFFIND-4.1 (Rohou and Grigorieff 2015). The CTFFind parameters are shown in Table 3-5. 1,032 particles were picked manually from the micrographs. These were used to make 2D classes using the EM algorithm. The best 2D classes were selected and used to train Topaz auto-picker (Bepler et al. 2019). Once trained, Topaz was used to pick 973,630 particles from 4,704 micrographs. When the autopicked particles were extracted, the particle box size was set to 448 pixels and was rescaled to 64 pixels.

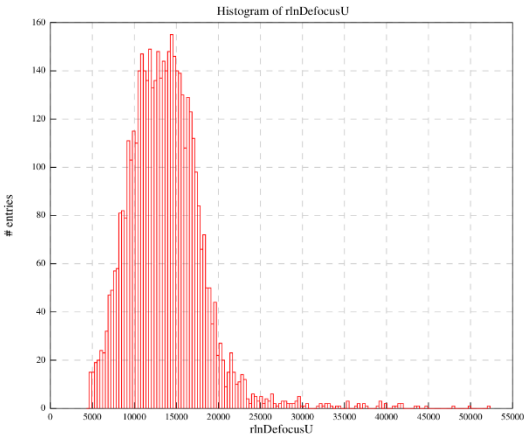
Table 3-5: Parameters for estimating the CTF

Parameter	Value
FTT box size (pix)	512
Minimum resolution (Å)	30
Maximum resolution (Å)	5
Minimum defocus (Å)	5,000
Maximum defocus (Å)	50,000
Defocus step size (Å)	500

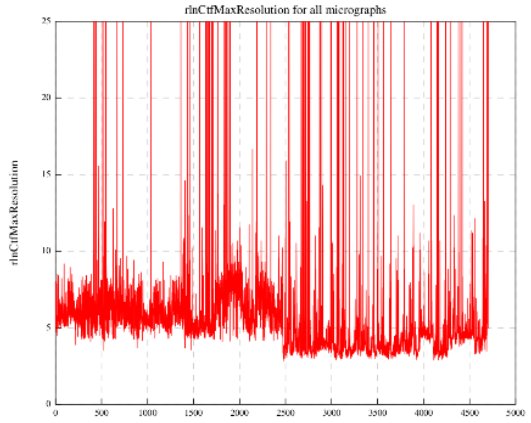
A) Defocus for all micrographs



B) Histogram of defocus



C) Maximum resolution



D) Histogram of maximum resolution

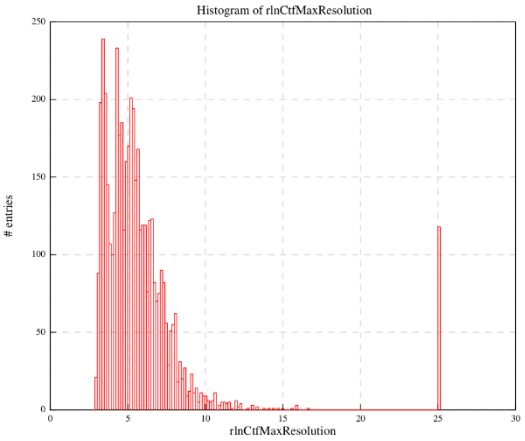


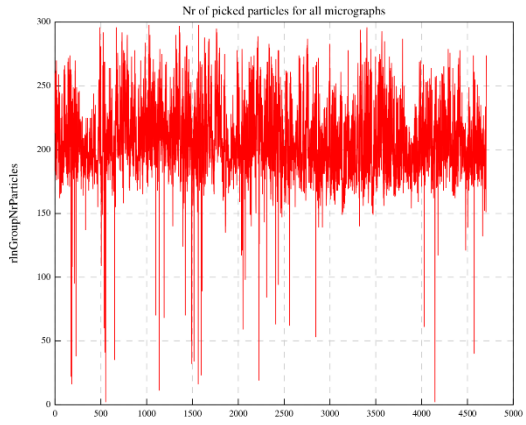
Figure 3-1: Micrograph filtering based on CTF estimation

Different parameters used to filter micrographs. A, C) Defocus and maximum resolution, respectively, derived from CTF estimation. B, D) Histogram of each parameter in A and C. A) Defocus for all micrographs. B) histogram of defocus. C) Maximum resolution for all micrographs. D) histogram of maximum resolution for all micrographs.

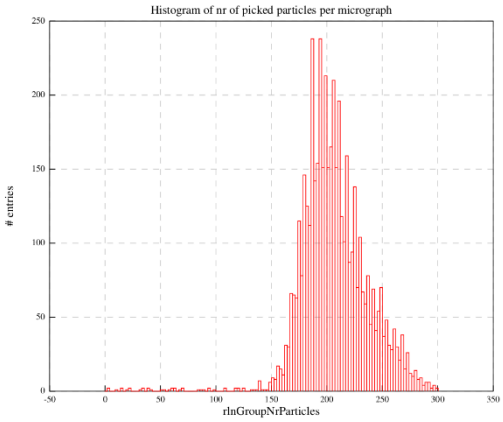
Table 3-6: Topaz autopicking parameters

Parameter	Value
Particle diameter (Å)	200
Low pass filter (Å)	20
Picking threshold	0.05
Minimum inter-particle distance (Å)	100
Maximum stddev noise (Å)	1.1
Minimum avg noise (Å)	-999

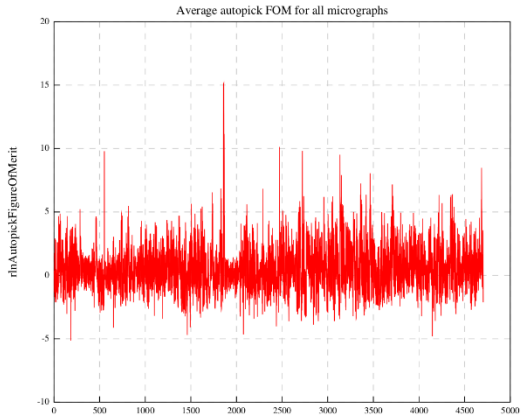
A) Number of picked particles in each micrograph



B) Histogram of picked particles per micrograph



C) Average autopick FOM for each micrograph



D) Histogram of autopick FOM

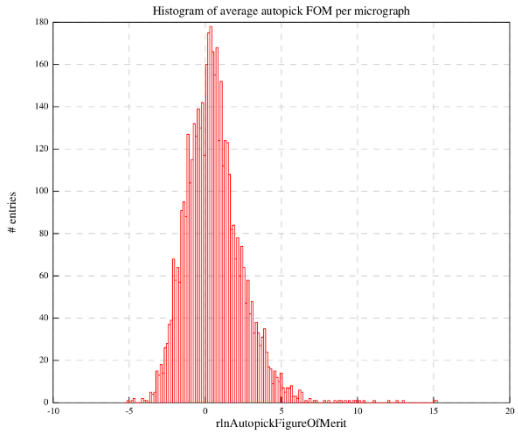


Figure 3-2: Topaz autopicking

A) number of picked particles in micrographs. B) histogram of the number of picked particles. In all micrographs. C) average autopick figure of merit (FOM) for all micrographs. D) histogram of average autopick FOM for each micrograph.

Outlying particles were removed by subset selection by only taking those particles with a maximum resolution less than 8 and those with a defocus less than 20,000. Particles were split into four batches of ~250,000 particles for 2D classification. The best 2D classes were selected from each batch using automatic selection with a minimum threshold of 0.25 and the VDAM algorithm. The classes were merged and the best classes automatically selected again.

Table 3-7: Parameters for 2D classification

Parameter	Value
Number of classes	200
Regularisation parameter	2
Algorithm	VDAM
Number of VDAM batches	200
Mask diameter (Å)	200

An initial model was made using the particles from the best 2D classes and C12 symmetry.

Table 3-8: Parameters for initial model generation

Parameter	Value
Number of classes	2
Regularisation parameter	4
Algorithm	VDAM
Number of VDAM batches	200
Mask diameter (Å)	200

The best 3D class was selected. The particles were extracted and the particle box size was reduced from 448 to 256 pixels. The best 3D class was low pass filtered to 30 Å. This was used to make a mask and low pass filtered again to 15 Å.

Table 3-9: Parameters for 3D class generation

Parameter	Value
Low pass filter (Å)	50
Symmetry	C1
Regularisation	4
Number iterations	25
Mask diameter (Å)	200

The mask was post processed. Three rounds of CTF refinement were done to estimate the 4th order aberrations, anisotropic magnification, and per-particle astigmatism. The map was refined and post-processed again. The particles were polished by Bayesian polishing. The map was refined, post-processed, and the CTF was refined as before. The map was refined and post processed a final time. The local resolution was estimated.

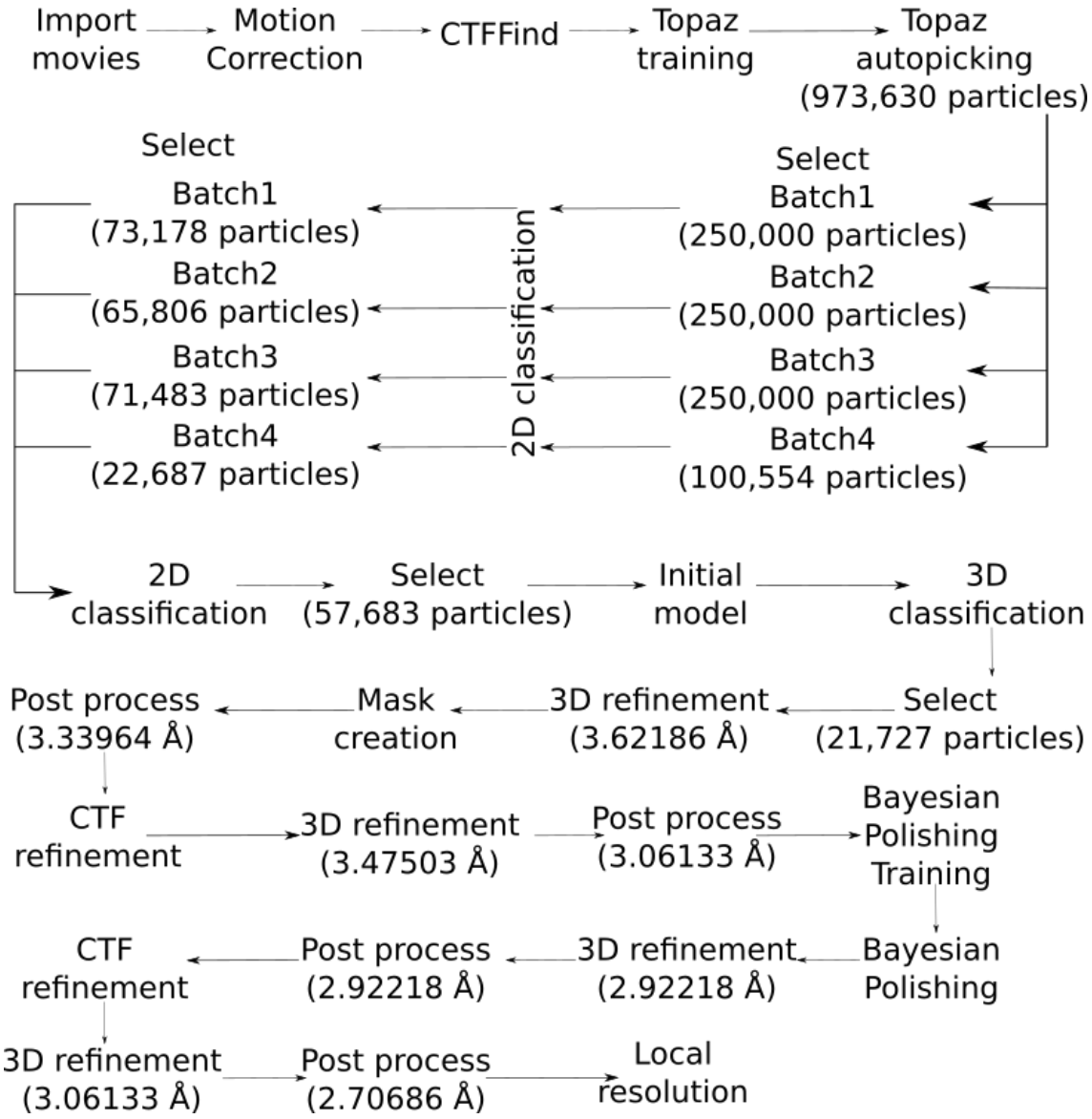


Figure 3-3: Flow diagram for RELION processing to the final mask.

Motion correction by MotionCor2. CTF estimation used CTFFIND4.1.

3.2.4. Model Building

The predicted structure of the subunit was generated from the GBSV1 amino acid sequence using ColabFold (see Chapter 2) (Jumper et al. 2021; Mirdita et al. 2022). The subunit was fitted into the density map using the Fit in Map function in Chimera. The first 40 amino acids of the sequence were deleted in accordance with the truncation of the construct. Each of the remaining residues was replaced with alanine to fit into the density then mutated back into the real amino acid. The amino acids were visually checked for good fit into the density using COOT, and a Ramachandran plot (also in COOT) was used for checking their conformation. After the subunit had been suitably placed into the density map and most amino acid side chains fitted into their respective densities, 11 copies of

the subunit were made around the central axis, to generate the 12mer, using a script written by Dr Huw Jenkins (see Appendix). The 12mer was refined using *real_space_refinement* in Phenix (Afonine et al. 2018). Changes were made to a single subunit. The rest of the previous subunits were deleted and the new corrected subunit used to populate the density map again.

3.2.5. Structure characterisation

The PDBtoPQR and APBS webservice was used to generate an electrostatic potential map of the whole 12mer at pH 8 with 0.25 M NaCl (Jurrus et al. 2018). The molecular lipophilicity potential (MLP) was calculated and imaged in ChimeraX (Laguerre et al. 1997). The inter-subunit interactions and surface area were calculated using PDBePISA (Krissinel and Henrick 2007; 'Protein Interfaces, Surfaces and Assemblies' Service PISA at the European Bioinformatics Institute', n.d.). The B-factor images were made in ChimeraX. The local resolution images were made using the results from the Local Resolution job from RELION.

3.2.6. Comparisons

The BLASTN (S. F. Altschul et al. 1997; Stephen F. Altschul et al. 2005) and BLASTP (Z. Zhang et al. 2000; Morgulis et al. 2008) software programmes were used to compare the portal protein gene and amino acid sequence respectively. The GBSV1_Δ1-40_775 structure was overlaid with the portal proteins structures of G20c, P23-45, phi29, SPP1, HK97 and T4 using the matchmaker tool in ChimeraX (Meng et al. 2006).

3.3. Results

3.3.1. CryoEM structure

The overall cryoEM processing workflow is shown in Figure 3-4. The micrographs show a high proportion of top views of the portal (views down the tunnel) (Figure 3-5). Views of the portal on its side (side views) are also present, but these are harder to spot and fewer in number. They also show fringes around the corners as a result of poor alignment of the microscope. This will affect the accuracy of CTF calculation. After pre-processing, particles were picked to train Topaz.

The 62 good 2D classes after recombining the four subsets of 250,000 particles show top views and side views. There are 218,605 particles in total. 44,593 particles make up the 15 classes containing a top view with at least one subunit missing. These make up ~20.4 % of the total particles. The preferential orientation of top views may be caused by the particles adsorbing to the air-water interface or to the grid. The missing subunit in some particles is likely to have been caused by contact with the air-water interface and starting to denature or come apart (Figure 3-5).

The 2D classes of the side views and top views containing intact rings were used to generate an initial model and, from that, 3D classes. Two of the classes were of inferior quality, with only 1-2 different views around the portal (Figure 3-4). The best 3D class had a greater range of views. This was taken forwards to use to make a mask. CTF refinement, 3D refinement, polishing and post-processing eventually led to the map at 2.7 Å resolution. The plot of the angular distribution of views shows that there is a predominance of top views and some bottom views (Figure 3-4). The quality of the map was assessed using the mtriage program in Phenix which gave a resolution of 2.7 Å (Figure 3-4). To build the protein into the density, the AlphaFold model was fitted into the density. The subunit was rebuilt as poly-alanine, then each residue was mutated to the correct sequence and fitted into the density. The density for the final helix of the crown domain (residues 379-406) was not well enough defined to be confidently modelled, and was thus omitted from the final model (Figure 3-6). A script written by Dr Huw Jenkins was used to generate a complete 12mer from the fitted structure of the subunit (see Appendix). The model was refined using *real_space_refinement* in Phenix (Afonine et al. 2018) to improve fit into density and ensure correct geometry.

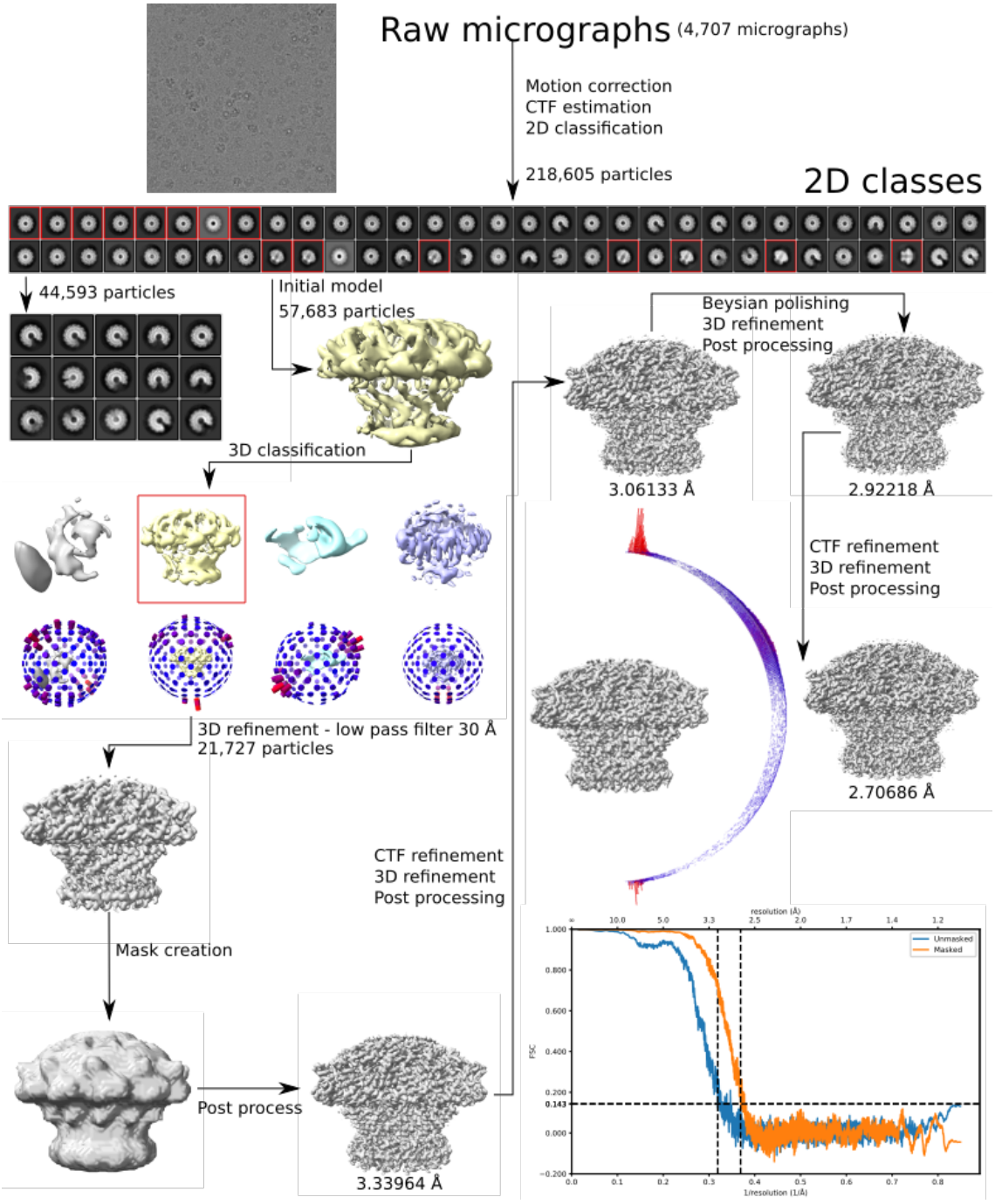
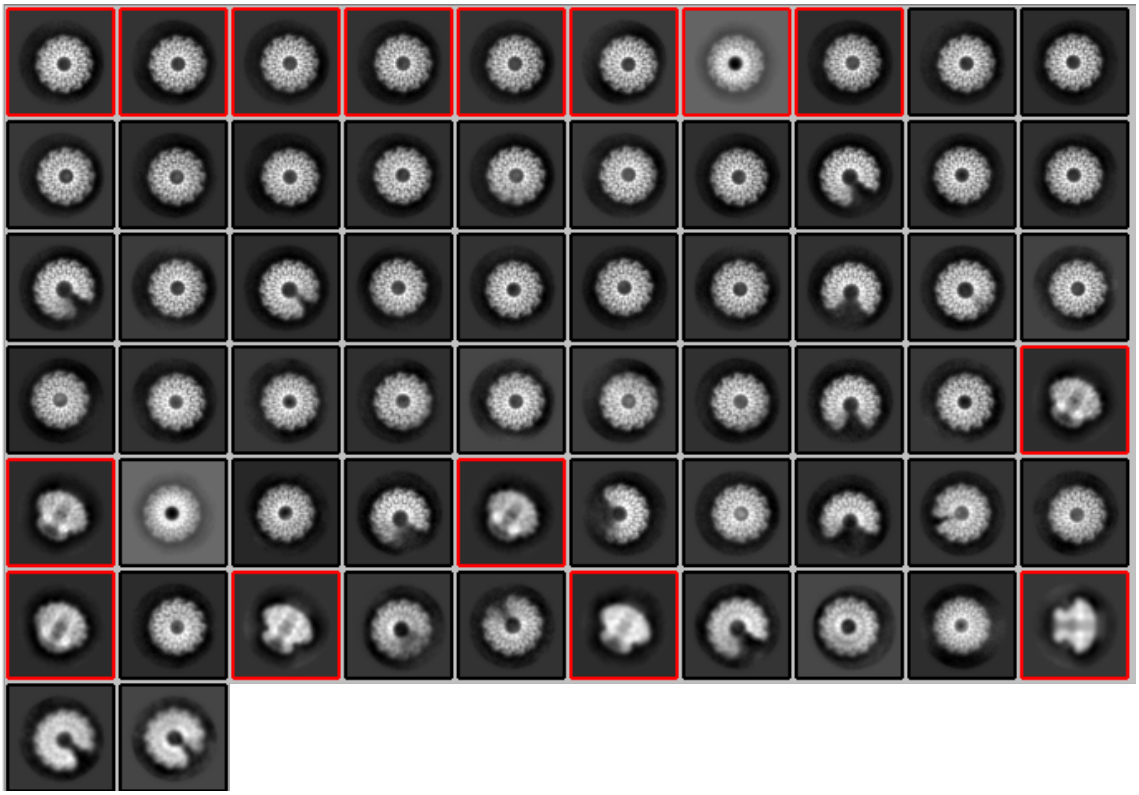


Figure 3-4: CryoEM processing workflow with maps

In total 4,707 micrographs were used for processing. After pre-processing, 218,605 particles were used to make 2D classes. The 62 best classes are shown here. Among them are 15 classes of a portal missing at least 1 subunit (44,593 particles). The best top and side view classes were selected (57,683 particles) to make an initial model. The best 3D class was used to make a mask. The mask was post-processed. CTF refinement was performed for aberrations, anisotopic magnification and per-particle refinement. The particles were polished, then CTF refined again. The angular distribution shows a preference for top views. The final resolution was 2.7 Å.

A) 2D classes



B) 2D classes with missing subunit(s)

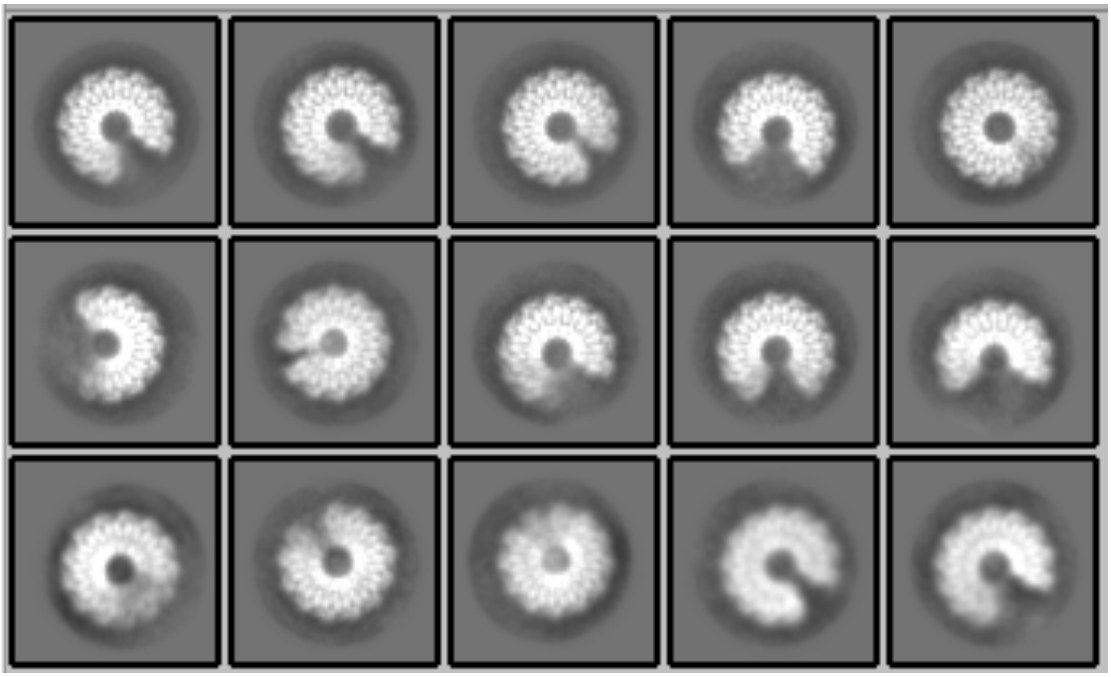


Figure 3-5: 2D classes.

A) The 62 2D classes selected by RELION's class_ranker programme based on 218,605 particles in total. The classes taken forwards to make the initial model are highlighted in red (57,683 particles, 26% particles). B) The 15 classes that contain "open rings" where the portal protein is missing one or more subunits. In total, there are 44,593 particles (20%) constituting these classes.

The predicted AlphaFold structure and the cryoEM structure are very similar, especially in the wing domain (Figure 3-6C). The main area in which they differ is the clip domain comprising the beta-sheet and alpha helix, which is significantly turned around in the predicted structure.

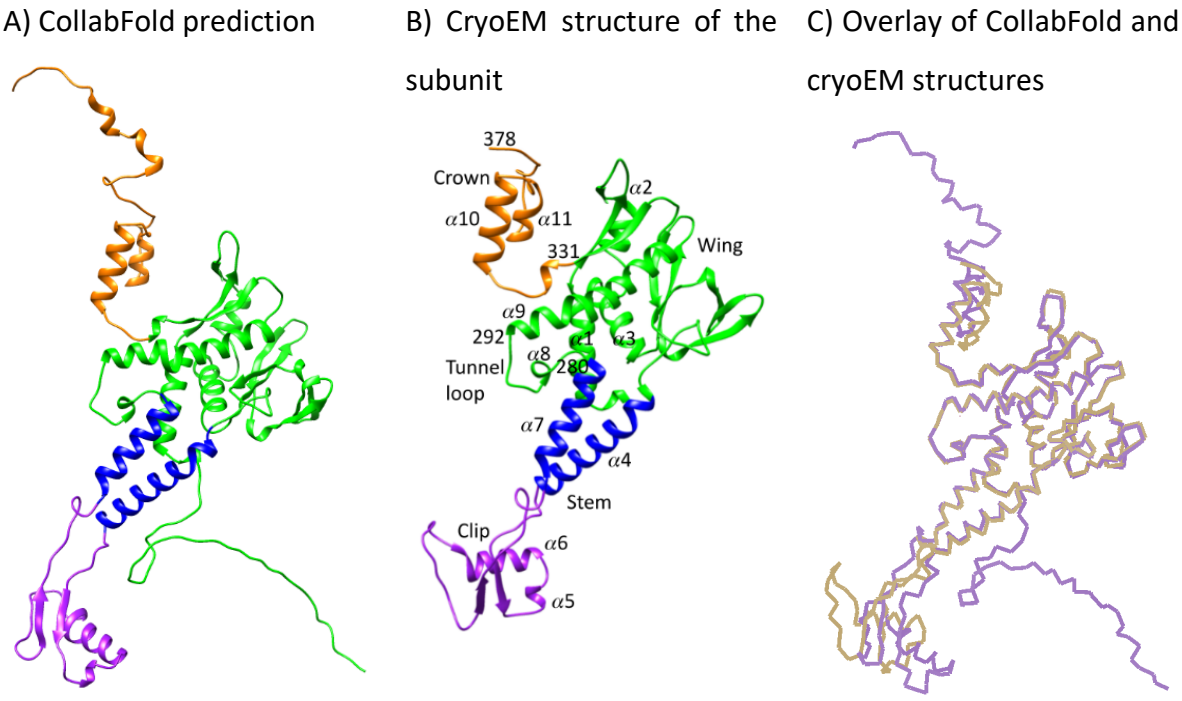


Figure 3-6: AlphaFold prediction and cryoEM structure of GBSV1_Δ1-40_775 subunit

Alpha fold prediction and final subunit from cryoEM reconstruction. A) Alphafold prediction of the whole portal protein. Orange – crown domain. Green – wing domain. Blue – stem domain. Purple – clip domain. B) Final subunit of the cryoEM reconstruction with helices, key amino acids and domains labelled. The first 40 amino acids had to be removed to make the protein soluble. The last helix of the crown domain could not be confidently modelled, so it was removed from the final model. PDB ID 9I8C. C) Overlay of the AlphaFold prediction and cryoEM reconstruction. Purple – AlphaFold prediction, brown – cryoEM subunit.

The structure is an oligomer of 12 identical subunits, as is common for portal proteins from other dsDNA viruses. Each subunit has an extended structure, comprising of clip, stem wing and crown domains (Figure 3-7). The stem domain is at an angle relative to the axis of the tunnel, resulting in a significant displacement of the clip with respect to the rest of the subunit, so that it is positioned not directly below the wing domain of the same subunit.

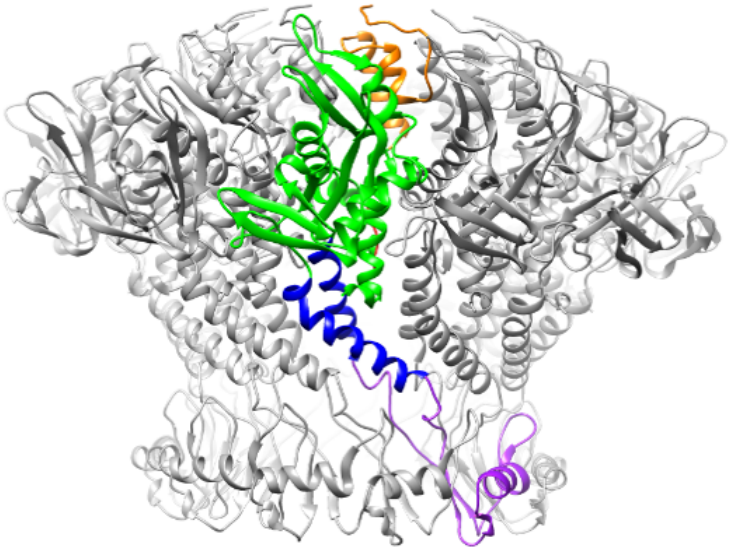


Figure 3-7: cryoEM structure of the GBSV1_Δ1-40_775

The structure is shown as ribbon diagram, with the tunnel axis vertical and with one of the 12 subunits coloured. Orange – crown domain. Green – wing domain. Blue – stem domain. Purple – clip domain. The distant subunits at the back of the 12mer have been removed for clarity. PDB ID 9I8C.

The regions that make up each domain are shown in Table 3-10.

Table 3-10: Domain boundaries

Domain	Residues
Crown	331-406
Wing	41-180, 280-331
Tunnel loop	280-292
Stem	181-200, 265-279
Clip	201-264

If the map is coloured by atomic displacement factor or local resolution, the stem and wing domains are well resolved (Figure 3-8). However, the density of the clip domains has a lower resolution than the rest of the structure, which suggests that it is flexible. The area between the stem domain helices and the clip domain is also more flexible than the wing domains. Interestingly, the tunnel loops, which in portal proteins from other bacteriophages were usually observed to be quite flexible and poorly defined in density maps, are well resolved (Figure 3-10, Figure 3-18).

Table 3-11: GBSV1_Δ1-40_775 refinement and validation statistics

Data processing	
Symmetry	C12
No. images	4,707
No. particles	21,727
Estimated resolution	2.70686
FSC threshold	0.143
Map sharpening B-factor	-62.4385
Model composition	
Chains	12
Atoms	64,752
Non-hydrogen atoms	32,292
Protein residues	4056
Nucleic acid residues	0
Model validation	
RMSD bond length (Å) (# > 4σ)	0.002 (0)
RMSD bond angle (°) (# > 4σ)	0.492 (24)
Mol probity score	1.1
Clash score	3.06
Ramachandran plot (%)	
Outliers	0
Allowed	1.51
Favoured	98.49
Rama-Z (Ramachandran plot Z-score, RMSD)	
Whole	0.59 (0.14)
Helix	1.06 (0.13)
Sheet	-1.70 (0.24)
Loop	0.49 (0.16)
Rotamer outliers (%)	0.45
Cβ outliers (%)	0
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaCLAMB outliers (%)	2.30
ADP (B-factors)	
Iso/Aniso (#)	32292/0
Protein (min/max/mean)	12.79/95.05/37.84
Nucleotide (min/max/mean)	---
Ligand (min/max/mean)	---
Water (min/max/mean)	---
Occupancy	
Mean	1
occ = 1 (%)	100.0
0 < occ < 1 (%)	0
occ > 1 (%)	0

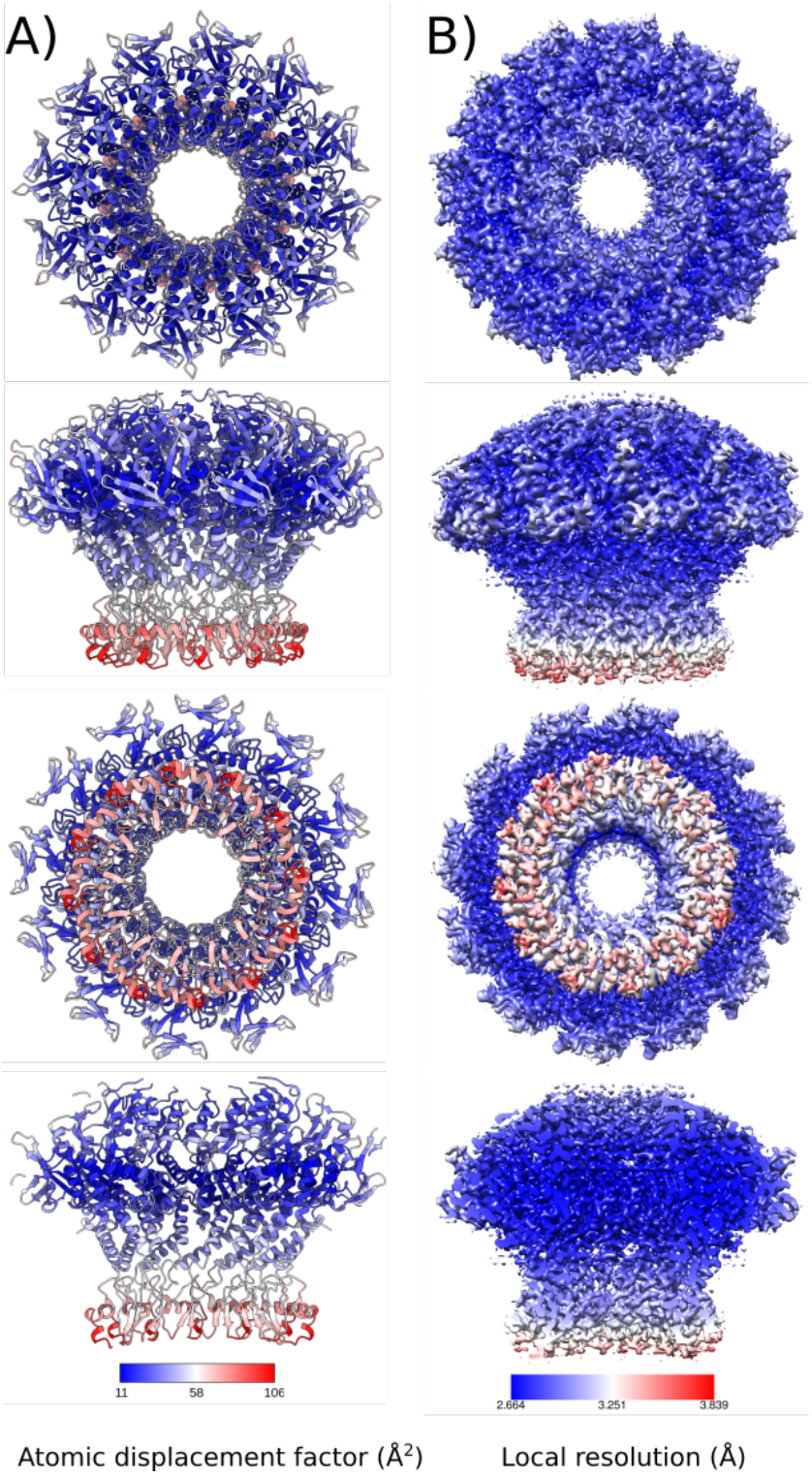


Figure 3-8: B-factors and local resolution of the final model and map

1st row - top view. 2nd row – side view. 3rd row – bottom view. 4th row – cross-section, with the portal protein viewed from inside the central tunnel. A) Ribbon diagram with B-factors, coloured by ChimeraX (Meng et al. 2023). B) The final post-processed map coloured by Local resolution using the LocalRes job from RELION.

The final helix of the crown domain could not be modelled because its corresponding density was not well-enough defined indicating flexibility. The density appeared to fold outwards over the top of the wing domains, like in a flower. Likewise, the ColabFold prediction also indicated disorder in this region. The wing domain is well resolved with the only missing part being the first 40 residues which were removed in this construct to make the protein more soluble. In portal proteins, the tunnel loops protrude into the tunnel to form a constriction point which is the narrowest part of the central tunnel. However, the measurements of the internal diameter (Figure 3-9C) show that GBSV1_Δ1-40_775 has a diameter of ~ 3.9nm between C_α atoms at this point, approximately the same as for parts of the tunnel formed by the clip and crown domains. This is because the tunnel loops are pulled back from the centre of the tunnel, as seen for the "open" state of the G20c portal protein (Bayfield, Steven, and Antson 2020). The observation that the tunnel loops are well-defined in density maps, suggests that this conformation must be stable in solution. It may be possible that under different conditions, the portal may adopt a closed conformation, similar to that of the G20c portal protein (Bayfield, Steven, and Antson 2020).

A) Width at the wing domains

B) Height and width at the clip domains

C) Diameter of the tunnel

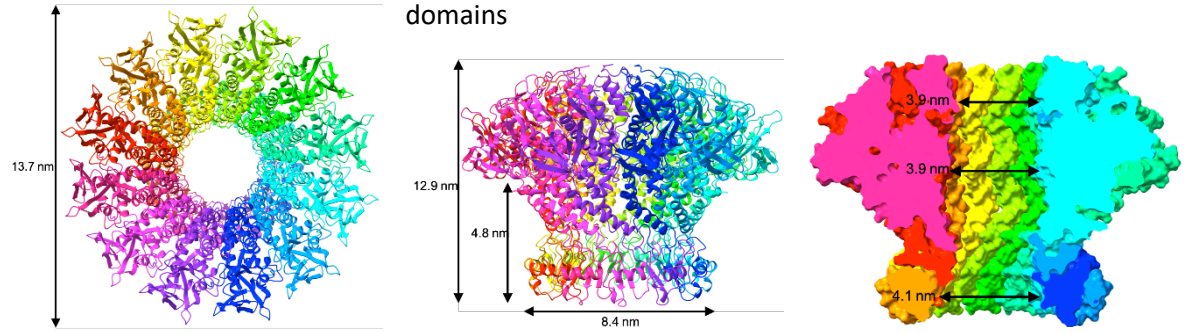


Figure 3-9: Overall dimensions of the portal protein.

The distances were measured in ChimeraX. A) top view and widest diameter across the wing domains. B) side view with the diameter at the clip domain, total height and distance from the bottom of the clip domain to the bottom of the protruding wing domain. C) side view. Internal diameters at the clip domains, tunnel loops, and at the crown domains. Strangely for a portal protein, the internal diameter appears to be ~4 nm throughout. Typically, the tunnel loops protrude into the tunnel and form a constriction.

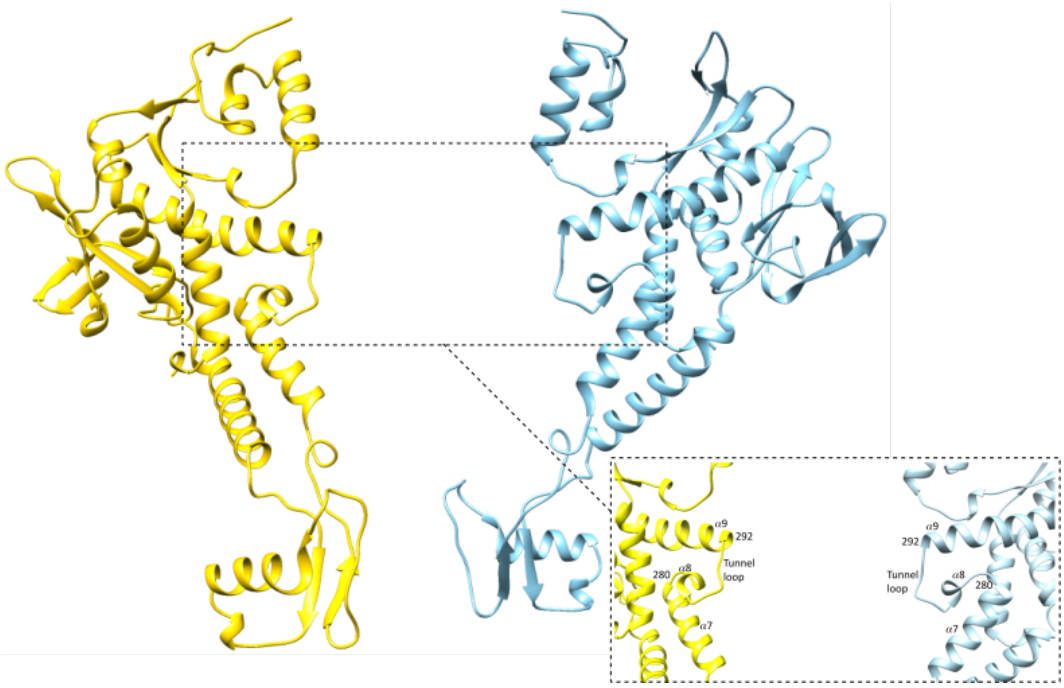


Figure 3-10: Two opposing subunits of the portal protein

The two subunits are shown as ribbon diagrams with the central tunnel axis vertical. The insert is a close-up view of the tunnel loop region.

As in most portal protein structures, the clip domain appears symmetrical. It has a ring of 3-strand β-sheets comprising two strands from one subunit and 1 strand from a neighbouring subunit as is common in portal proteins (Javed et al. 2021; Lebedev et al. 2007). Each clip domain also has an alpha helix (helix α6, Figure 3-6B) exposed at the outside, oriented perpendicular to the tunnel axis. The region after helix α6 reaches around the outside of the portal.

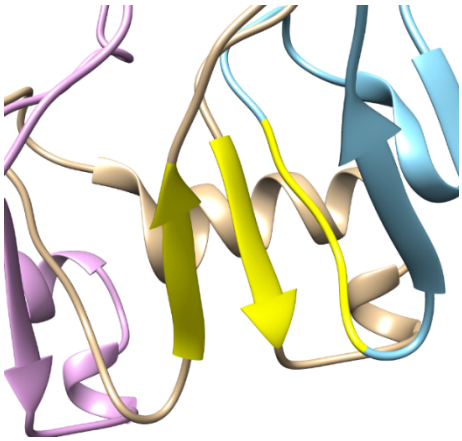


Figure 3-11: the three-strand β-sheet in the clip domain.

Three adjacent subunits are shown in pink, brown and blue. The 3-strand β-sheet between adjacent subunits is coloured yellow. Two of the strands come from one subunit, the 3rd comes from the neighbouring subunit. This also shows the alpha helix on the outside of the clip domain which is horizontal wrt the tunnel axis. Other portal proteins have an alpha helix in their clip domain, but they tend to be vertical wrt the tunnel axis.

3.3.2. Inter-subunit Interactions

The 3-stranded β-sheet in the clip domain is expected to play a major role in holding the subunits together (Figure 3-11), owing to the main-chain/main-chain inter-subunit hydrogen bonding interactions. Electrostatic potential calculations for the subunit at pH 8 (the pH of the portal protein solution used for grid preparation) reveals that the two sides of the subunit have opposing charges (Figure 3-12), in common with the bacteriophage phi29 portal protein (Simpson et al. 2000). This charge distribution will facilitate ionic attractions between adjacent subunits, facilitating the portal protein to assemble and stabilising the oligomeric protein.

Two adjacent chains of the portal were submitted to the PDB-ePISA webserver (Krissinel and Henrick 2007; 'Protein Interfaces, Surfaces and Assemblies' Service PISA at the European Bioinformatics Institute', n.d.) to analyse the interface contact. The molecular surface area between two subunits is 4,282 $Å^2$. The complex formation significance score (CCS) was 0.945, which suggests that the interface area is essential for subunit-subunit interactions. Five salt bridges were formed between neighbouring subunits (Figure 3-13). One salt bridge in the crown domain between Glu359 and Arg361. There is a salt bridge between Arg293 and Asp294 in the first turn of the long helix α9 which follows the tunnel loops. Transition from "open" to "closed" tunnel conformation is expected to involve conformational changes in the tunnel loop. To allow conformational adjustments, the N-terminal segment of the long helix α9 may have to unwind, breaking the Arg293-Asp294 salt bridge. The remaining salt bridges are found in the stem domain, between Asp187 and Asp187 in one subunit, and Lys197 in the neighbouring subunit. Subunit-subunit interactions are further stabilised by 40 H-bonds between neighbouring subunits. These are mostly found in the wing and clip domains, with some in the crown. There are no disulphide or covalent bonds between subunits.

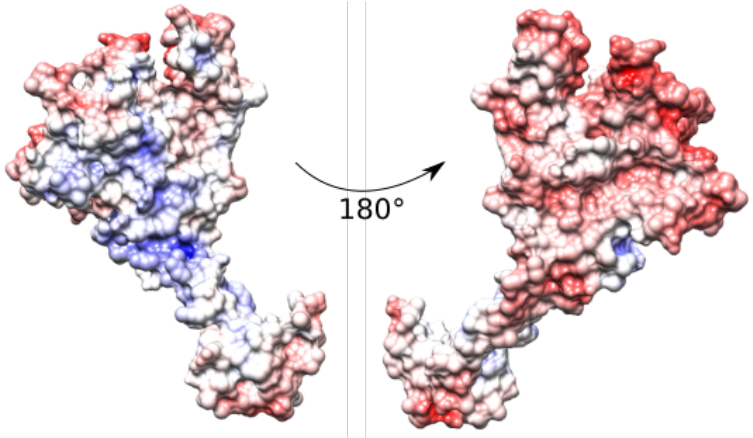


Figure 3-12: Molecular surface of two opposing subunits coloured by electrostatic potential calculated at pH 8.

One side has more positive charges, the other has more negative charges. Electrostatic potential of the portal protein subunit at pH 8 was calculated by PDB2PQR and APBS (Jurrus et al. 2018).

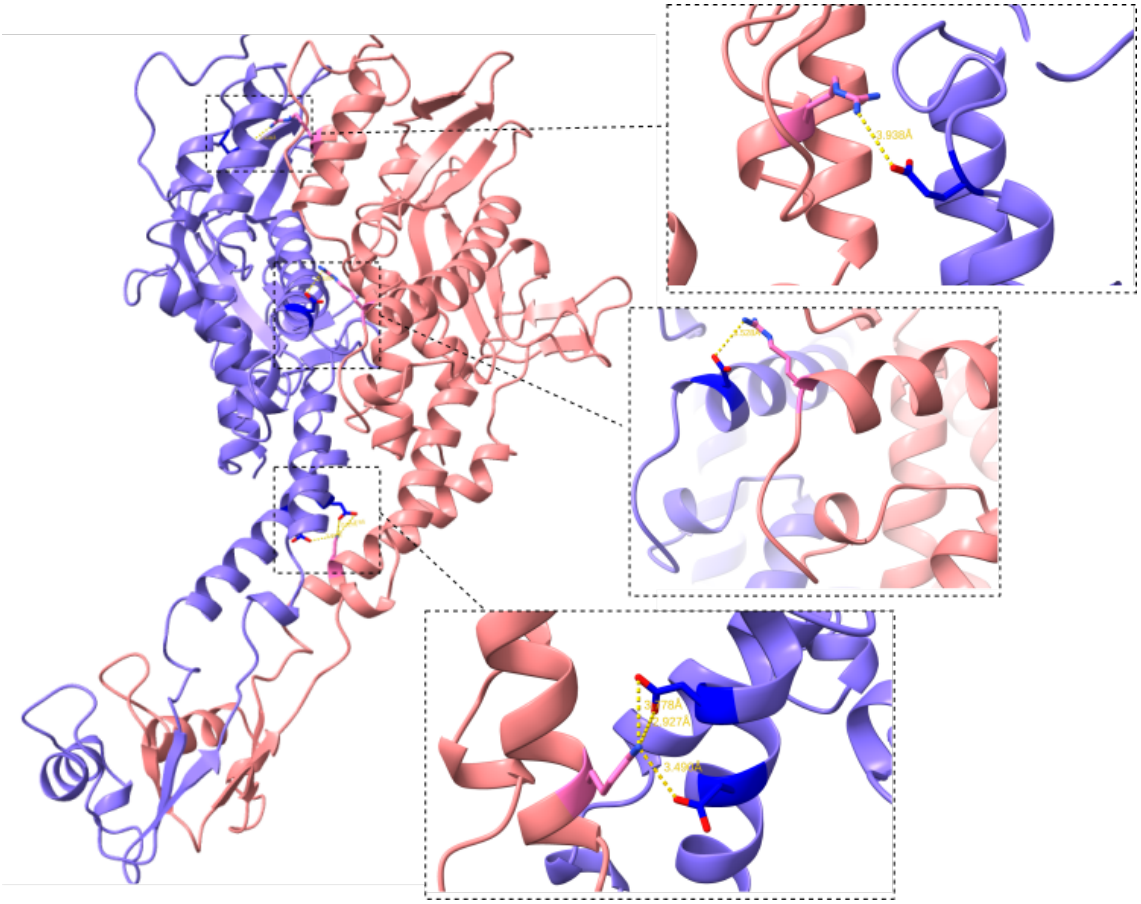


Figure 3-13: Salt bridges between neighbouring GBSV1_Δ1-40_775 subunits.

View of two subunits from inside the tunnel. The insets are close-up views of the salt bridges. The top inset shows the salt bridge between Glu359 and Arg361 in the crown domain. The middle insert shows the salt bridge between Arg293 and Asp294 at the start of helix α9, just after the tunnel loops. It is possible that this would need to break in order for the tunnel loops to expand into the tunnel. The bottom inset shows three salt bridges between Asp187 and Asp187 (blue subunit) and Lys197 (pink subunit).

There are also possible interactions between alternating subunits, e.g. subunit-A and subunit-C. These interactions are limited to four H-bonds and one salt-bridge in the clip domain. (Figure 3-14). The gain in solvation free energy of the 12mer is -533.8 kcal mol^{-1}, and the free energy of assembly dissociation is 602 kcal mol^{-1}.

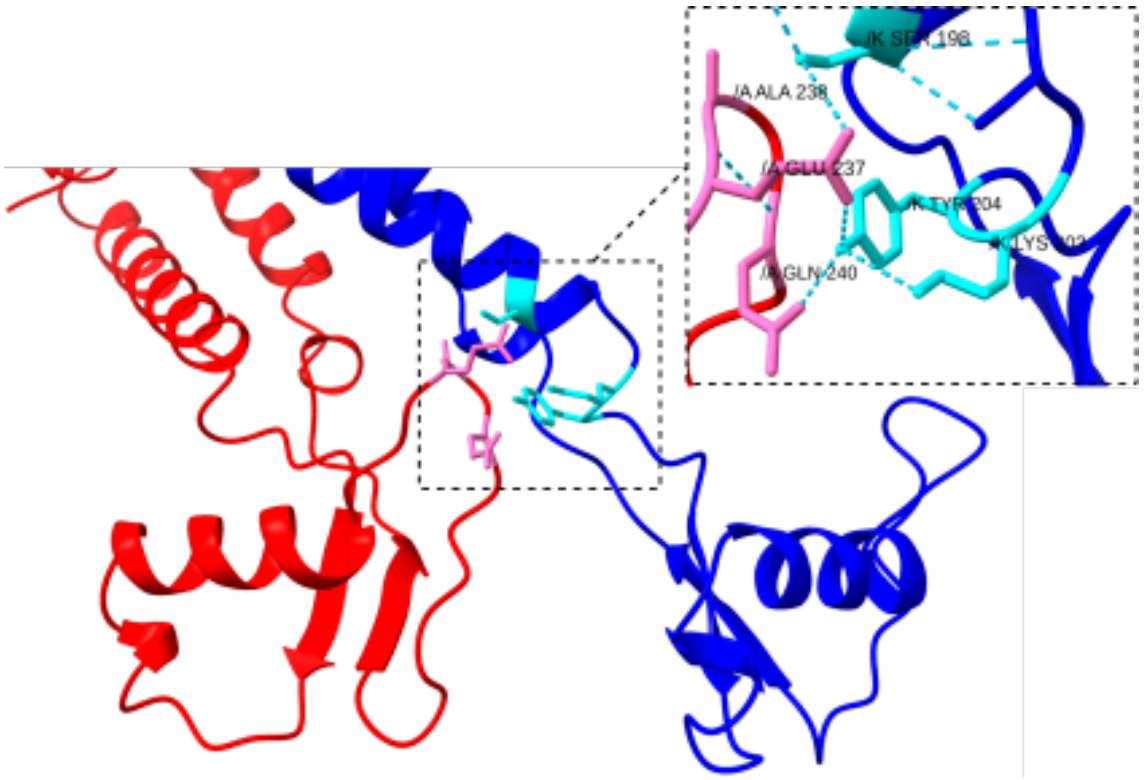


Figure 3-14: H-bonds and salt bridges between subunits A and C

The view is of the clip domains of subunits A and C from outside the portal protein. Subunit B has been removed for clarity. The inset shows the salt bridge between subunits A and C.

3.3.3. Surface properties of the portal protein

The PDBtoPQR and APBS web service was used to generate an electrostatic potential map of the whole 12mer at pH 8 with 0.25 M NaCl (Jurrus et al. 2018) (Figure 3-15). The crown domain of the portal protein is negatively charged, except for a few small pockets of positive charge. The area of the stem domain under the wing domain is positively charged. There is a band of negative charge between the stem and clip domains. The outside of the clip domain appears to have pockets of positive charge. The very bottom and inside of the clip domains are negatively charged. There are three bands of positive charge around the insert of the tunnel at the crown domain, under the tunnel loops, and between the clip and stem domains.

The molecular lipophilicity (MLP) view was used to make a map of the oligomer's MLP in ChimeraX (Laguerre et al. 1997). The prediction shows that the whole surface is hydrophilic (Figure 3-16). There are some small pockets of hydrophobicity on the surfaces of the subunit interfaces. The surface of the tunnel and the outer surface are largely hydrophilic.

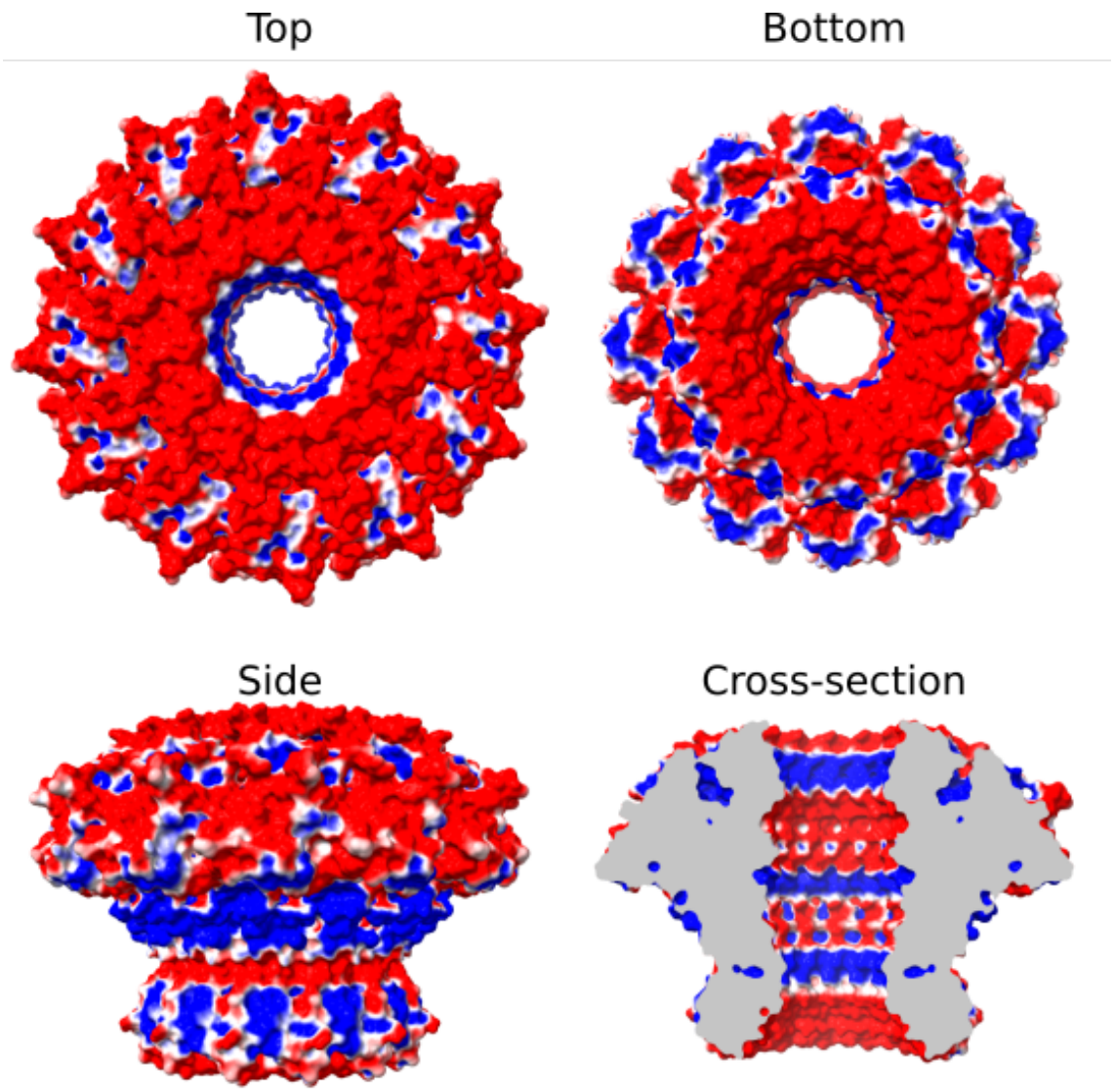


Figure 3-15: Electrostatic potential map of the GBSV1_Δ1-40_775 portal protein.

Molecular surface of the 12mer is coloured according to electrostatic potential. Red/blue colors correspond to ±1 kT/e-. The figure was generated using the PDB2PQR and APPS webservers (Jurrus et al. 2018). There are three bands of positive charge inside the tunnel: at the top of the clip domain, beneath the tunnel loops, and at the crown domains.

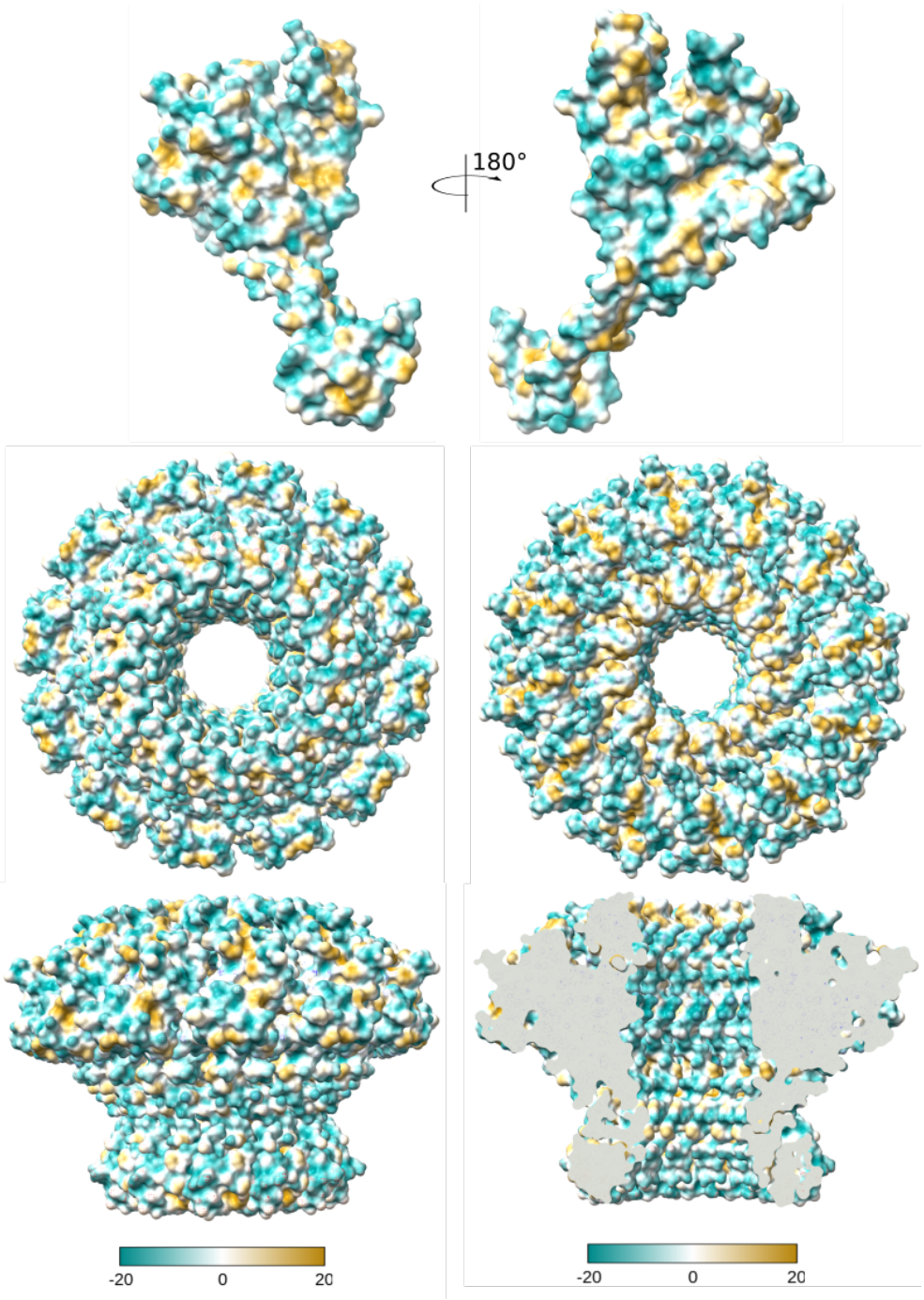


Figure 3-16: molecular lipophilicity potential (MLP) map of the GBSV1_Δ1-40_775

Gold - highly lipophilic (highly hydrophobic), green - highly lipophobic (highly hydrophilic). The portal protein appears to be very hydrophobic in the tunnel and around the outside. Nanobubbles would not be expected to form inside the tunnel. A) top view. B) bottom view. C) side view. D) cross section through the pore.

3.3.4. Homology search

In general, portal proteins with the characteristic portal protein fold have low sequence similarity. Unsurprisingly, a BLAST analysis for the GBSV1 portal protein gene (ORF20) found no homologues (S. F. Altschul et al. 1997; Stephen F. Altschul et al. 2005). Similarly, a BLASTP analysis of the GBSV1 amino acid sequence found only one result from viruses: the BV1 portal protein (see also Chapter 2) (Z. Zhang et al. 2000; Morgulis et al. 2008). The rest of the results were from firmicute bacteria (*bacillota*). BV1 and GBSV1 are the only two members of the *svunvirus* genus, so it is expected that their amino acid sequences would be similar. The purification results for the full-length BV1 and GBSV1 PP were similar (see Chapter 2). However, a purification of the BV1 protein without its N-terminal unstructured region was not attempted as it was not cloned in time. Given their similarity, it is possible that removing this domain may allow the BV1 portal protein to be purified, much like it did for GBSV1.

3.3.1. Amino acid sequence comparison with common portal proteins

The amino acid sequence of the GBSV1 portal protein was compared to that of the T7, phi29, SPP1 and G20c portal protein using BLASTP (Table 3-12) (S. F. Altschul et al. 1997; Stephen F. Altschul et al. 2005). The cover was low for all of them. The score refers to the alignment score between the GBSV1 amino acid sequence and the four subject sequences. Maximum score is the highest single score found between an alignment of the query and subject sequence. Total score is the sum of all of the alignment scores. All of the alignments had E-values greater than 1, which suggests that the alignments could be random. The E-value is the Expect value. This is the significance threshold that BLASTP uses in deciding if two sequences are similar or not. The percentage identity (Per. ident) is the percentage of the subject sequences (the four well-studied portals) that matches the query sequence (GBSV1 portal protein). The accession length (Acc. Len) is the length of the subject sequence. Accession is the accession number of the subject sequence.

Table 3-12: Portal protein amino acid sequence comparison

Description	Max Score	Total Score	Query Cover	E-value	Per. ident	Acc. Len	Accession
Chain A, Portal protein [Escherichia phage T7]	18.1	35.8	16%	2.3	35.14	547	6QWP_A
Chain A, Upper Collar Protein [Bacillus phage phi29]	16.2	16.2	7%	8.6	31.25	309	1JNB_A
Chain A, Portal Protein [Bacillus phage SPP1]	18.9	53.1	29%	1.3	30.77	503	2JES_A
portal protein [Thermus phage G20c]	18.1	36.2	23%	2.1	26.42	448	API81889.1

3.3.2. Structure comparison with other portals

The GBSV1 portal protein structure was overlaid and compared to structures of other well studied or thermophage portal proteins (Figure 3-17) using the matchmaker command in ChimeraX (Meng et al. 2006). G20c (PDB ID: 6IBG) and P23-45 (PDB ID: 6QJT) are thermophages with similar sequences and structures. Phi29 (PDB ID: 6QX7), SPP1 (PDB ID: 2JES), HK97 (PDB ID: 8CEZ) and T4 (PDB ID: 3JA7) are all used as model bacteriophages, so their portal proteins are well studied.

The GBSV1 structure appears most similar to that of G20c and P23-45. But the structure aligns well with the other portals as well, especially at the wing domains. The crown domains of G20c, P23-45, SPP1, and HK97 all have a similar composition to GBSV1 with three short alpha helices. The crown domain of GBSV1 appears to be lower relative to the others. Possibly because it is in the open conformation in which the crown domain would be expected to be lower as it is depressed by the pressure inside the capsid.

None of the other portal proteins have the same horizontal alpha helix as the GBSV1_Δ1-40_775 clip domain (helix α6 in Figure 3-6). The rest have an alpha helix in the clip domain that is more vertical relative to the tunnel axis. This may be a unique conformation of the clip domain. The amino acid sequence of the BV1 portal protein has a high similarity to the GBSV1 portal protein, so it may also have this feature.

Figure 3-18 shows the same comparisons as Figure 3-17 but focusing on the tunnel loops. The tunnel loops of the GBSV1 portal protein are all pulled further back from the tunnel compared to all other portal proteins listed. The exception is phi29, but their tunnel loops are not modelled in this PDB file. However, there are some similarities between GBSV1 and others. Both GBSV1 and HK97 have the small helix at the N-terminal end of the tunnel loop, which is labelled as α8 in GBSV1 (Figure 3-6). The α9 helix in GBSV1 appears to protrude further into the tunnel than other portal proteins. If the GBSV1 portal is in the open conformation, it may be that to form the closed conformation, the α8 helix and part of the α9 helix unwind to push the tunnel loops into the tunnel. When expressed in vitro, what determines if the protein is in open or closed conformation? In the capsid it would be osmotic pressure of the dsDNA filling the capsid. The structure of the portal protein solved from a portal protein in the procapsid or a mature virus may be different to that solved from a portal in solution. As seen in phi29, the tunnel loops can be hard to model and may be flexible in solution, but the density of the tunnel loops of GBSV1 are well resolved, so the structure seen here should be a genuine conformation.

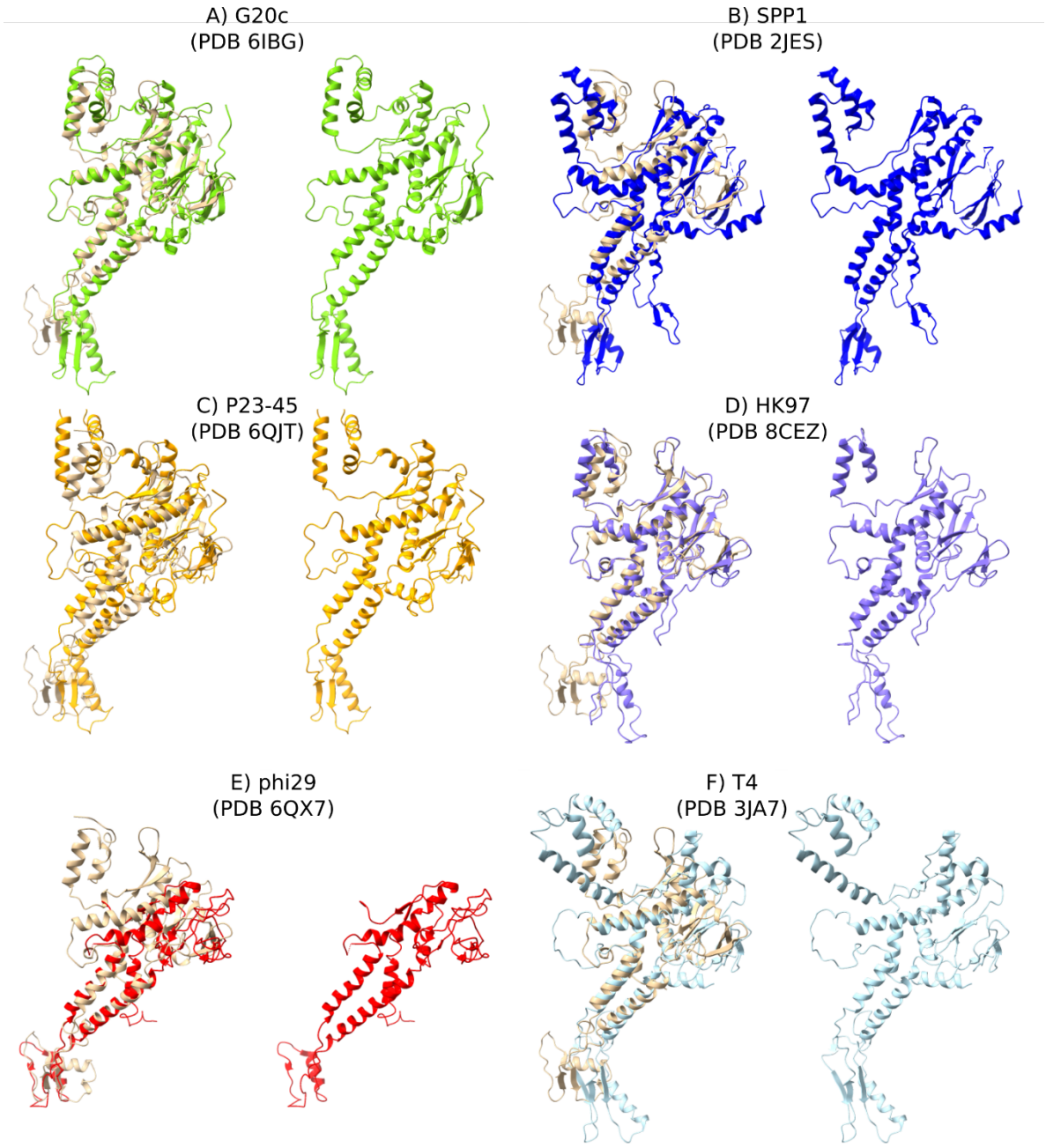


Figure 3-17: GSBV1 portal protein overlaid with common portal proteins

GBSV1 (PDB 9I8C) – beige. Left – overlay of the GBSV1 portal protein and a well studied portal protein. Right – the well studied portal protein alone. A) G20c portal protein (PDB 6IBG) – green. B) SPP1 portal protein (PDB 2JES) – blue. C) P23-45 portal protein (PDB 6QJT). D) HK97 portal protein (PDB 8CEZ) – purple. E) phi29 portal protein (PDB 6QX7) – red. F) T4 portal protein (PDB 3JA7) – light blue.

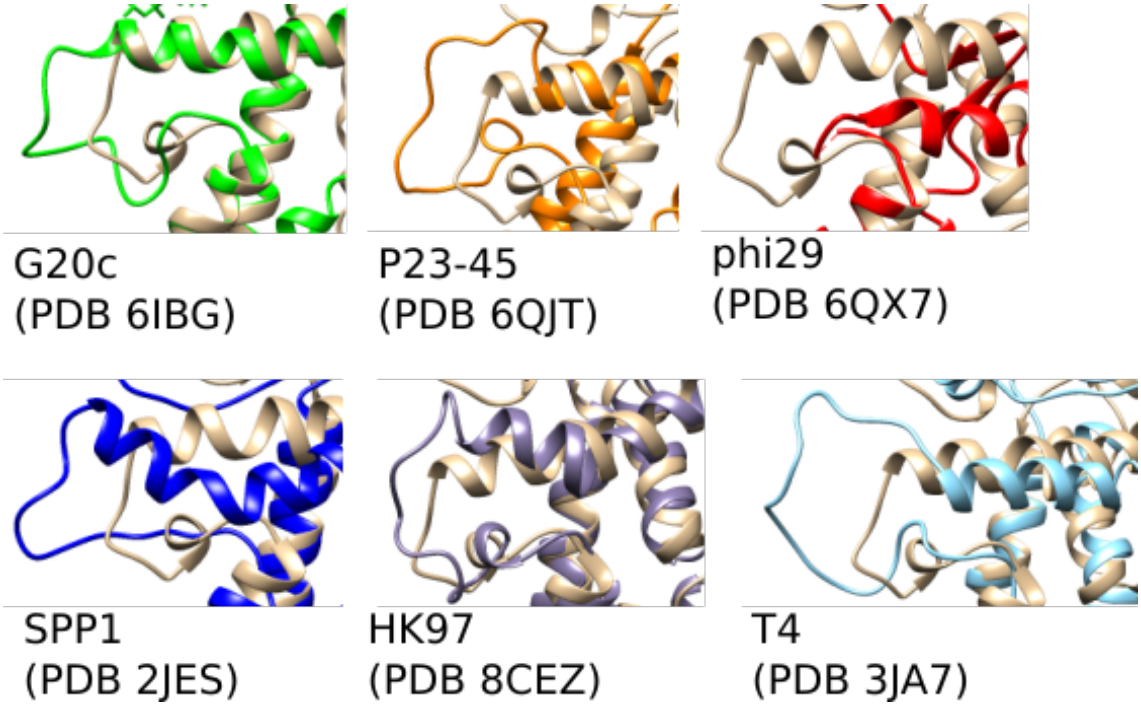


Figure 3-18: Comparison of portal protein tunnel loops

The proteins are shown as ribbon models. GBSV1 (PDB 9I8C) – beige. The tunnel loop of phi29 is disordered, so is not modelled. G20C (PDB 6IBG) – green. P23-45 (PDB 6QJT) – orange. Phi29 (PDB 6QX7) – red. SPP1 (PDB 2JES) – blue. HK97 (PDB 8CEZ) – purple. T4 (PDB 3JA7) – light blue.

3.4. Discussion

The preferential orientation showed by the GBSV1 portal protein on the grids is a common problem in cryoEM. In this case, the portal orientates itself with the tunnel axis along the electron beam, so that the view is down the barrel of the portal. This problem, usually caused by adsorption to the AWI or grid, has been resolved by replacing quantifoil carbon grids with gold grids.

3.4.1. Preferential orientation

Like other portal proteins, the cryoEM dataset of the GBSV1 portal protein suffers from preferential orientation when applied to cryoEM grids (Lokareddy et al. 2017). This limits the views that are available for reconstruction. To solve the preferential orientation problem, different grid coatings could be applied, e.g. graphene oxide or thin carbon. Additives, such as detergents, could also be tested. However, grid optimisation for the T7 portal protein in isolation found better results without a continuous carbon layer (Fàbrega Ferrer 2017).

3.4.2. Data processing

Co-expressing the portal with the GBSV1 scaffold protein may give a more stable conformation as seen with other bacteriophage portal proteins (Luan 2013). The very bottom of the clip domain and the last helix of the crown domain proved quite flexible. Some flexibility around the clip domain is expected for it to be able to interact with pRNA or small terminase (for packaging), or adaptor protein (for ejection). This has been seen in other portal proteins as well (Javed et al. 2021). The clip domain was difficult to build into, while the last helix of the crown domain had to be removed from the final model. The map could be adjusted with the programme DeepEMhancer which may help model these sections (Sanchez-Garcia et al. 2021). Focused refinement, where these areas of the map are adjusted individually could also help. The clip domain can be flexible in order to bind different parts of the packaging or ejection machinery with different symmetries, so focused refinement may show some different conformations or higher resolution (Y. Peng et al. 2024). Modelling in C1 symmetry may also show different clip domain conformations.

The open ring structures are most likely the result of the protein denaturing upon contact with the AWI. It is impossible to tell by sight if the side views are whole 12mers. An additional filtration step may be to use the Icebreaker programme to filter by ice thickness (Olek et al. 2022).

3.4.3. 2D classes of open rings

In addition to complete portal protein rings, 2D classes revealed presence of incomplete oligomers missing one or more subunits. Portal proteins with open rings have been observed in PaP3 bacteriophage (David Hou et al. 2022). However, those were found within only a small subset of

proteins. It is likely that the 'open ring' structures seen here are the result of the particles adsorbing to the AWI.

3.4.4. Comparison between AlphaFold prediction and real cryoEM structure

The similarity between the AlphaFold prediction and the real cryoEM structure is to be expected given the number of portal protein structures in the PDB and their conserved 'portal protein' fold. The main two areas of deviation are the clip and crown domains. The density for the last helix of the crown domain appeared to fold back over the wing domains. The density could be seen, but not well enough to build the model. The AlphaFold structure predicts that it is disordered. Disordered crown domains have been seen in other portal proteins (Javed et al. 2021). The clip domains could be flexible here in order to bind the small terminase equivalent or tail machinery allowing them to adopt different conformations or symmetries (Rūmnieks, Füzik, and Tārs 2023; Woodbury et al. 2022). While other portal proteins have an alpha helix in their clip domain, it tends to be in a different, more vertical orientation compared to GBSV1.

3.4.5. Open and closed conformations

It has been suggested that during DNA packaging, portal proteins adopt an open conformation, with a wider tunnel diameter, to allow the passage of dsDNA into the capsid (Bayfield, Steven, and Antson 2020). If the DNA should start to exit during packaging, the tunnel loops would be pushed into the tunnel to hold the DNA in place until packaging can resume (Bayfield, Steven, and Antson 2020). The tunnel loops of SPP1 portal were predicted to push into the major groove of the DNA (Lebedev et al. 2007). If this is the case, then the GBSV1 portal protein here may be in the open conformation. The open and closed conformations are two different energy minima in the energy landscape. The clarity of the density of the tunnel loops may imply that only the open conformation is present in this sample. So, the open conformation may be the lower energy state, and energy is required to push the portal into the closed conformation. If the portal were to move into a closed conformation, the small alpha helix (α8) at one end of the loop may have to unwind to allow it to extend into the tunnel. But this hypothetical conformational change may not be enough for the loops to extend into the tunnel to narrow the tunnel as seen in other portals. For the tunnel loops to extend into the tunnel, some of helix α9 may be required to unwind as well. The salt bridge in the helix α9 may also have to break. The diameter of B-form dsDNA is ~2 nm. The tunnel diameter is ~4 nm throughout. Other portal protein structures were collected *in situ* inside capsids (J. Xu et al. 2019).

Portal proteins assume slightly different conformations depending on their surroundings, e.g. in solution, in procapsid, during packaging, in a mature phage. Previous portal proteins have found that the tunnel loops can be hard to resolve (Javed et al. 2021; Sun et al. 2015a; Bayfield, Steven, and

Antson 2020). Future work on the structure of GBSV1 may involve finding the conditions required to trap the portal in the closed conformation, with the tunnel loops protruding into the tunnel.

3.4.6. Surface charge

Like the phi29 portal protein, each side of the GBSV1 subunits has a complementary charge (Simpson et al. 2000). These sides would have a key role in oligomer assembly. Unlike phi29 portal protein there is no band of hydrophobic residues under the wing domain. The highly lipophobic nature of the outside of the portal protein means that it may be difficult to insert it into a lipid membrane without alterations. The hydrophilic tunnel should prevent the formation of nanobubbles and make wetting easier (Paulo, Sun, Di Muccio, et al. 2023). The negative charge and hydrophilic nature of the tunnel fits its role as a dsDNA channel. The negative charge of the portal tunnel will attract a layer of cations. The negatively charged backbone of DNA will be repelled from the negatively charged tunnel surface, which should prevent it from adsorbing during packaging or ejection. The cations that surround the backbone of DNA should be substituted for those on the tunnel surface during packaging. In this way, the layer of cations may act like ball-bearings to help the passage of dsDNA into the capsid.

3.5. Summary

In summary, the cryoEM structure of the GBSV1_Δ1-40_775 protein has been solved to 2.7 Å. The portal protein is shown to be a 12-subunit oligomer with an internal tunnel, similar to structures found for other portal proteins (Sun et al. 2015b; McNulty et al. 2018; Bayfield, Steven, and Antson 2020; Javed et al. 2021; David Hou et al. 2022; Orlov et al. 2022; Mukherjee et al. 2023). There are specific structural differences which are discussed in this chapter. It shows what is likely in the open conformation of the portal, in line with the proposed mechanism of the portal protein as a one-way valve to prevent DNA slipping out during packaging (Bayfield, Steven, and Antson 2020).

Chapter 4 Thermal and Chemical Stability of GBSV1_Δ1-40_775

4.1. Introduction

For protein pores to be used as biological pores or as part of hybrid pores, they must withstand adverse conditions required by samples, e.g. low/high salt, and be stable for long periods of time. A key reason that thermophage portal proteins were chosen was for their expected thermal stability and tolerance to the harsh conditions required for nanopore analysis. Therefore, the thermal melting point of GBSV1_Δ1-40_775 was investigated using nanoDSF under different conditions.

4.1.1. Thermal stability

The Gibbs Helmholtz equation (see Chapter 1) describes the three parameters that are used in describing the thermal stability of proteins: ΔH_M is the enthalpy of denaturation at transition midpoint, ΔC_P is the change of heat capacity of denaturation (specific heat of unfolding), and T_M is the midpoint of thermal denaturation (the temperature at which half of the protein is unfolded).

$$\Delta G_{D-N} = \Delta H_m \left(1 - \frac{T}{T_m}\right) \quad (39)$$

If $\Delta G_{D\text{-}N}$ is plotted, it forms an inverted parabola (Figure 4-1, black curve). The X-intersections are the melting temperatures. There is a point of cold denaturation (T_c) and heat denaturation (T_h). The temperature of maximum stability (T_s) is the temperature at which the $\Delta G_{D\text{-}N}$ is the highest and unfolding is most energetically unfavourable. There are three ways that the thermal stability of a protein can be increased (Nojima et al. 1977). If the protein becomes more energetically stable, then the whole curve is translated upwards (black curve to blue curve). The T_s is the same, but it is thermally stable over a greater range. The curve can also translate right to a higher temperature range (black curve to yellow curve). The size of the range is the same, but the T_s also changes. Finally, the gradient of the curve can change (black curve to red curve). The gradient of the curve is dictated by the ΔC_P. The ΔC_P can be found by calorimetry, but ΔC_P=0 will give an acceptable estimate of $\Delta G_{D\text{-}N}$ for most cases (equation (39)) (Tim Sharpe 2012). In this case, the graph will be a line, rather than a parabola.

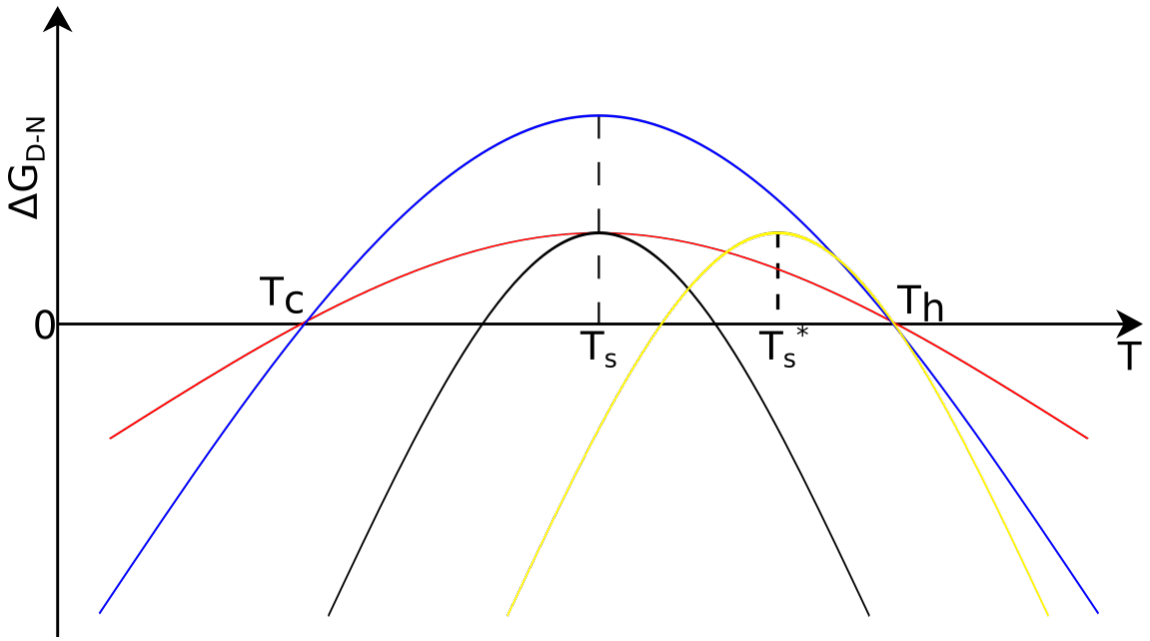


Figure 4-1: ΔG_{D-N} Vs temperature.

ΔG is the free energy of unfolding. If ΔG is negative, protein unfolding is energetically favourable. If it is positive, unfolding is energetically unfavourable. T_c – cold denaturation temperature. T_h – heat denaturation temperature. T_s - temperature of maximum stability. T_s is the modified temperature of maximum stability for the yellow curve. Black curve – original curve. Blue curve - increase thermostability by deepening the black curve. Red curve – universally flattening the black curve. Yellow curve – shift the whole curve to higher temperatures (Nojima et al. 1977).*

4.1.2. The Hofmeister series

Proteins are affected by their surrounding environment. The Hofmeister series is the series of ions based on how they interact with water to precipitate proteins (Figure 4-2) (Kunz, Henle, and Ninham 2004; Hofmeister 1888). Ions are assessed by their Jones-Dole viscosity B-coefficient which describes their interactions with the solvent (G. Jones and Dole 1929). Anions with positive B-values are usually kosmotropes, while those with a negative B-value are chaotropes, though there are exceptions (reviewed in (Zhao 2016) and (Marcus 2009)). Kosmotropes make water more structured ('structure-making') which can stabilise or precipitate proteins. Chaotropes can disrupt the structure of water ('structure-breaking'), which can destabilise proteins. Kosmotropic anions and chaotropic cations promote protein stabilisation and precipitation, while chaotropic anions and kosmotropic cations destabilise proteins. In Figure 4-2, the ions from $MeSO_4^-$ to the right are chaotropic. Ammonium sulphate can be used to precipitate proteins for storage or concentration (see Chapter 2) (also called 'salting out'). Urea and guanidinium hypochloride (GuHCl) are common chaotropic agents which are used to denature proteins. Chemical denaturation can be reversible (Nojima et al. 1977).

Kosmotropic	$PO_4^{3-} > SO_4^{2-} > EtSO_4^- > OAc^- > MeSO_4^- > Cl^- > Br^- > I^- > BF_4^- > PF_6^-$	Chaotropic
Chaotropic	$(CH_3)_4N^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Al^{3+}$	Kosmotropic

Figure 4-2: Hofmeister series

Top – anions. Bottom – cations. Ions towards the left-hand side are 'structure -making', while those towards the right-hand side are 'structure-breaking'.

4.1.3. Protein aggregation

Protein aggregation can be a problem throughout protein production and can hinder the use of proteins downstream (for examples, see Chapter 2, GVE2_Δ1-18_676 and GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822) (reviewed in (Pace, Scholtz, and Grimsley 2014; W. Wang and Roberts 2018; Housmans et al. 2023)). The exact causes of protein aggregation are still debated. It may depend on the protein structure, environment, and other experimental conditions. Aggregation can be studied using various techniques, e.g. backreflection and CD (see Chapter 1) (Nojima et al. 1977). Kosmotropic agents (see above), such as ammonium sulphate can force proteins to aggregate by giving greater structure to the surrounding water, forcing them together.

4.1.4. Portal protein stability

At time of writing, there are few studies on portal protein stability, but there have been studies on portal protein ring assembly and disassembly. Bacteriophage scaffold proteins have a role in promoting portal ring assembly (Motwani et al. 2017), but they are not essential for ring formation: P22 portal subunits will eventually come together into rings over time, especially at high concentrations and the right temperatures (Moore and Prevelige 2001). P22 portal proteins formed rings when incubated at 15-30 ℃. In addition, the buffer is also important in ring assembly. The SPP1 portal protein requires a minimum salt concentration in the buffer in order to form oligomers (Jekow et al. 1999). This may be due to the salt reducing the size of the EDL around the proteins or kosmotropic effects (see above). During purification, portal protein subunits may be present at high enough concentrations for oligomers to form and can be separated from subunits (Moore and Prevelige 2001). Urea has been used to promote DNA to exit from the capsid of a mature virus without unfolding the portal protein (Y. Peng et al. 2024). But the portal protein would have been shielded from the bulk solution, so conclusions about its stability in urea cannot be drawn. It may be reasonable to assume that the temperatures for thermophage portal protein ring assembly may be higher than for mesophilic phages.

4.1.5. Chapter Aims

The aim of this chapter was to investigate the thermal and chemical stability of the GBSV1_Δ1-40_775. Different concentrations of common salts, and different pH buffers were tested for their effect on its unfolding. The midpoint of unfolding with urea was found.

4.2. Materials and Methods

4.2.1. KCl and NaCl solubility test

The solubility test for GBSV1_Δ1-40_775 was performed using the same method as the small-scale expression test (see Chapter 1). The stock solutions were 50 mM Tris-HCl pH 8, 50 mM Tris-HCl pH 8 4 M KCl, and 50 mM Tris-HCl 4 M NaCl pH 8. The stock solutions were mixed together in different ratios to make different concentrations of each salt. The concentrations of each were 0, 0.25, 0.5, 1, 2, 3 and 4 M. The GBSV1_Δ1-40_775 sample was spun at 14,000 g for three minutes beforehand to remove any aggregates.

4.2.2. NanoDSF with different salt concentrations

NanoDSF was performed using the ThermControl programme in a Prometheus NT.48 in the Technology Facility of the University of York. A fresh aliquot of purified GBSV1_Δ1-40_775 was taken from the -80 ℃ freezer and was spun down at 14,000 g for 3 minutes, then diluted to 1 mg ml^{-1} in 50 mM Tris-HCl pH 8, 0.25 M NaCl. Stock solutions were made for baseline and different NaCl, KCl, and $MgCl_2$ concentrations (Table 4-1).

Table 4-1: Stock solutions for salt gradients

Stock buffer	Composition
A	50 mM Tris-HCl pH 8
B	50 mM Tris-HCl, 4 M NaCl pH 8
C	50 mM Tris-HCl, 3.8 M KCl pH 8
D	50 mM Tris-HCl, 20 mM $MgCl_2$ pH 8

The stock solutions B, C, or D were mixed with stock-A in different ratios to make buffers of different salt concentrations (Table 4-2).

Table 4-2: Salt concentrations for test buffers

Salt	Concentration (M)
NaCl	0.25, 0.5, 1, 2, 3, 4
KCl	1, 2, 3, 3.8
$MgCl_2$	0.5 mM, 2 mM, 5 mM, 10 mM, 20 mM

4 µL of diluted protein were added to 16 µL of chosen test buffer condition in a 500 µL Eppendorf tube for a final concentration of 0.2 mg mL^{-1}. Capillary tubes (Nanotemper Prometheus NT.40) were loaded by capillary action and tilting the tubes to fill them to the ends. The PR-ThermControl programme in the Prometheus NT48 was used. The excitation energy was 30%, and the temperature ramp was 1 ℃ min^{-1}. The graphs were drawn in Igor Pro v9.0.5.1 (64-bit).

4.2.3. Urea denaturation gradient

To test the stability of GBSV1_Δ1-40_775 under chemical denaturants, a 9 M urea stock was made from 54.05 g urea in 100 mL MQ water. A 1 ml stock of 50 mM Tris-HCl pH 8 was made from 1 M Tris-HCl pH8. Another 1 ml stock of 50 mM Tris-HCl pH8, 8 M urea was made using 50 µL 1 M Tris-HCl pH 8 and 888.89 µL 9 M urea. 18 µL of solutions of different urea concentrations from 0 to 8 M in 0.5 M steps were made in PCR tubes by mixing the 50 mM Tris-HCl pH8 and 50 mM Tris-HCl pH8, 8 M urea stock solutions in different ratios. A fresh aliquot of GBSV1_Δ1-40_775 was diluted to 5 mg ml^{-1} in the 50 mM Tris-HCl pH 8, 0.25 M NaCl. 2 µL of 5 mg ml^{-1} GBSV1_Δ1-40_775 was added to each urea solution and gently mixed by pipetting for a final concentration of 0.5 mg mL^{-1}. Capillaries were filled as above. NanoDSF was run as above but used 70% excitation energy. The graphs of the nanoDSF data were drawn using Igor Pro version 9.0.5.1. The graphs were fluorescence at 330 nm vs temperature, fluorescence at 350 nm vs temperature, fluorescence ratio (A350/A330) vs temperature, and scattering vs temperature. Prometheus also gave the first differential of each measurement, which was also plotted against temperature. The curves of fluorescence ratio (A350/A330) vs temperature were fitted using equation (34). The values for α_N, β_N, α_D, β_D, m and $X_{50\%}$ were estimated until the curve fit well. The fluorescence ratio at 25 ℃ at each concentration of urea was plotted against urea concentration. This gave a graph of change in fluorescence ratio with increasing urea concentration. These data points could be fitted using equation (35), which gave the value for the T_m. The $\Delta G_{D\text{-}N}$ for thermal denaturation was found using equation (35). The $\Delta G_{D\text{-}N}$ for chemical denaturation with urea was found using equation (34).

4.2.4. Salt gradients with 2 M urea

The method was the same as above. Because of the inclusion of urea, the same concentration ranges could not be used. The stock buffers and test buffers are shown in Table 4-3 and Table 4-4.

Table 4-3: Stock solutions for salt gradients with 2 M urea

Buffer	Components
A	50 mM Tris-HCl pH 8, 2 M urea
B	50 mM Tris-HCl pH 8, 2 M urea, 3 M NaCl
C	50 mM Tris-HCl pH 8, 2 M urea, 2 M KCl
D	50 mM Tris-HCl pH 8, 2 M urea, 20 mM $MgCl_2$.

The curves for thermal denaturation could be fitted using thermal denaturation equation (equation (35)) as above to give the thermal midpoint of denaturation for each concentration of urea.

Table 4-4: Salt concentrations tested with 2 M urea

Salt	Concentration (M)
NaCl	0.25, 0.5, 1, 2, 3
KCl	1, 2
$MgCl_2$	0.5 mM, 2 mM, 5 mM, 10 mM, 20 mM

The data points were fitted using Igor Pro. The variables α_N, β_N, α_D, β_D, ΔH_M, and T_M were estimated based on the shape of line between all points. ΔC_P was assumed to be 0.

4.2.5. pH stability Test

For the pH stability test, the buffers were 50 mM Bis-Tris pH 6 or pH 6.5 and 50 mM Tris-HCl pH 7/7.5/8. 18 µL of each buffer solution was used and added 2 µL of 5 mg ml^{-1} GBSV1_Δ1-40_775 for a final concentration of 0.5 mg ml^{-1} in a PCR tube. Prometheus NT.48 capillaries were filled by capillary action. The samples were heated using the ThermControl programme at 10% excitation energy and heated from 20 ℃ to 95 ℃ at a rate of 1 ℃/minute. Each condition was repeated three times. The traces of fluorescence ratio against temperature were fitted using the thermal fit equation above. The T_M was found for each condition. The T_M and average T_M for each condition was plotted against pH.

4.2.6. Electrostatic potential

The PDB of the refined 12mer structure of GBSV1_Δ1-40_775 (see Chapter 3) was submitted to the PDB2PQR web server at different pH values to make PQR files (Unni et al. 2011). Grid points in x, y and z were set as 500. The course sampling in x, y and z was set at 350. The fine sampling in x, y, and z was set at 250. The concentration, radii and charge of 1 M KCl, 1 M NaCl or 20 mM $MgCl_2$ were added to the input file. The values for the radii of the ions were taken from (Shannon 1976). The pH stability experiments contained no salt, so none was included in the Advanced Poisson-Boltzmann Solver (APBS) calculations. The PQR files were put into the APBS programme on the supercomputer cluster HULK at the University of York (Jurrus et al. 2018). The surface of the protein was coloured by electrostatic potential in ChimeraX (Meng et al. 2023). The scale was ± 10 kT/e^-. The electric field lines and isoelectric surfaces were drawn using the APBS tools plugin in PyMol (Baker et al. 2001; Schrödinger, LLC 2015). The scale was also ± 10 kT/e^- unless otherwise stated.

4.2.7. Thermal stability test

Two fresh aliquots of GBSV1_Δ1-40_775 were taken from the -80 ℃ freezer. Each was diluted to 5 mg/ml in 50 mM Tris-HCl pH 8, 0.25 M NaCl. One was left on the bench at room temperature. The other was left in the fridge at 4 ℃. Each day a 15 µL sample was taken and stored in a 500 µL Eppendorf

tube. 5 µL of 4x SDS dye was added. The samples were stored at -20 ℃. When the gel was to be run, the samples were thawed. They were boiled at 95 ℃ for 5 minutes as usual, then run on a 12% SDS-PAGE gel at 180 mV, 400 mA for 1 hr. The gel was made, run, stained and imaged the same as previously described (see Chapter 2).

4.3. Results

4.3.1. Solubility in NaCl and KCl

As part of its role as a potential nanopore, GBSV1_Δ1-40_775 should be able to withstand low and high concentrations of NaCl and KCl in order to adapt to the conditions of potential analytes. If it requires a specific concentration of either, then its utility is limited to analytes that are also comfortable in that range. KCl is also commonly used in nanopore experiments. Figure 4-3 shows a solubility test with 1 ml pellets of GBSV1_Δ1-40_775 at different concentrations of NaCl and KCl. The band in the soluble fraction only starts to fade at the highest concentrations of each salt. At such high concentrations, the protein does not aggregate and remains in the soluble fraction.

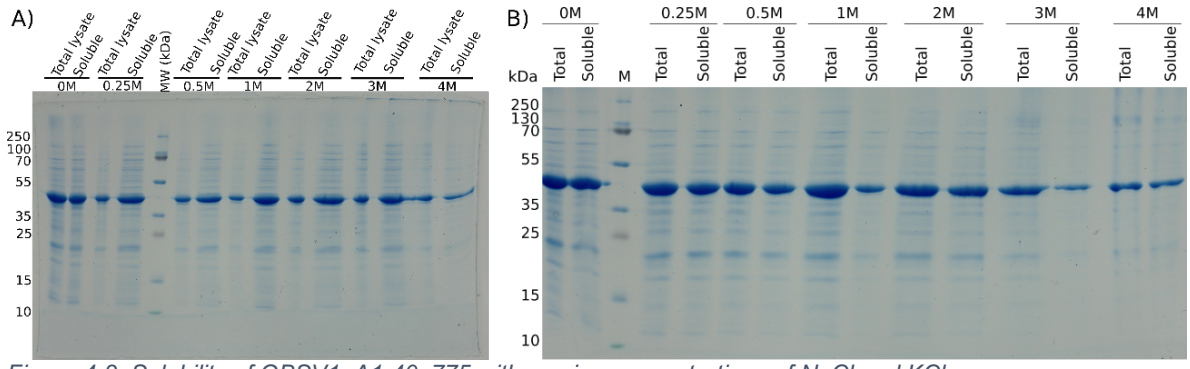


Figure 4-3: Solubility of GBSV1_Δ1-40_775 with varying concentrations of NaCl and KCl

The total and soluble fractions of cell pellets re-suspended in buffers of varying concentrations of NaCl or KCl were run on a 15% SDS-PAGE gel. The intensity of the band in the soluble fraction only starts to wane at the highest concentrations. The expected weight of the GBSV1_Δ1-40_775 subunit is 45 kDa. The buffer was 50 mM Tris-HCl pH8. A) 0, 0.25, 0.5, 1, 2, 3, 4 M NaCl or B) 0, 0.025, 0.5, 1, 2, 3, 4 M KCl

4.3.2. Melting temperature with different salts and concentrations

NanoDSF of GBSV1_Δ1-40_775 with different concentrations of NaCl, KCl or $MgCl_2$ had no effect on the melting temperature. The protein did not melt under any conditions tested (data not shown). The electrostatic potential of the subunits at pH 8 was found previously (see Chapter 3). It showed that the opposite sides of the subunit have positive and negatively charged surfaces. There are 16 H-bonds and ten salt bridges between adjacent subunits, but no covalent or disulphide bonds.

4.3.3. Urea denaturation curve

NanoDSF was repeated with different concentrations of urea in each sample (Figure 4-4). As urea concentration increased, the melting temperature decreased.

At 0 M urea, the ratio of fluorescence at 350 nm to 330 nm just starts to increase at 95 ℃. At 1-3 M urea, the A350/A330 ratio increases to a plateau, though 1 M plateaus at a lower value than the higher concentrations. There are multiple peaks within the first derivative of the 4 M urea trace, either side of the main peak at ~ 61 ℃. 5 M urea has a main peak at ~ 35 ℃, but also has a peak at ~ 50 ℃, similar

to the same peak in 4 M urea. The 6 M urea trace has a peak at ~ 29 ℃, which trails up to 40 ℃, similar to the main peak of 5 M. The ratio at 4 M is still increasing gently at 95 ℃. The fluorescence ratio at urea concentrations above 4 M, appears to plateau at similar values. They start at higher fluorescence ratio values and the first derivatives give much lower peaks. The first derivative for 5 M shows two peaks: the first larger peak at approximately 35 ℃, then a smaller peak at approximately 50 ℃. The first derivative of 6 M gives a single peak at ~28 ℃. The traces of the fluorescence ratio and first derivative of 7 and 8 M urea are flat. Starting at such high concentrations of urea, they appear to have already unfolded before the temperature started increasing.

For all of the urea concentrations tested, except for 1 M urea, as the temperature increases, the backreflection scattering decreases (Figure 4-4C). The samples with higher concentrations of urea tend to start at higher values of scattering, which would imply aggregation. However, the first differential of the backreflection (Figure 4-4D) shows that for urea concentrations above 2 M, the traces merely move around 0. If the increase in temperature was truly causing aggregated proteins to separate, the differential would be an inverted bell curve. Therefore, increasing the temperature does not cause the protein to disaggregate. The flat trace of the first differential means that there is no change of aggregation state, so it may already be aggregated. The traces for the differential of backreflection for 0 M urea show the gradient starting to increase at the highest temperatures, which suggests that the protein starts to aggregate. The differential for the trace for 1 M urea shows three peaks at ~72 ℃, ~79 ℃, and ~89 ℃. The first peak overlaps with second peak. These could represent three stages of aggregation. But the differential decreases sharply past 91 ℃. The differential trace for the backreflection for 2 M urea has a single peak, which also decreases below 0. These decreases could suggest that the protein aggregates at certain temperatures, then disaggregates at slightly higher temperatures.

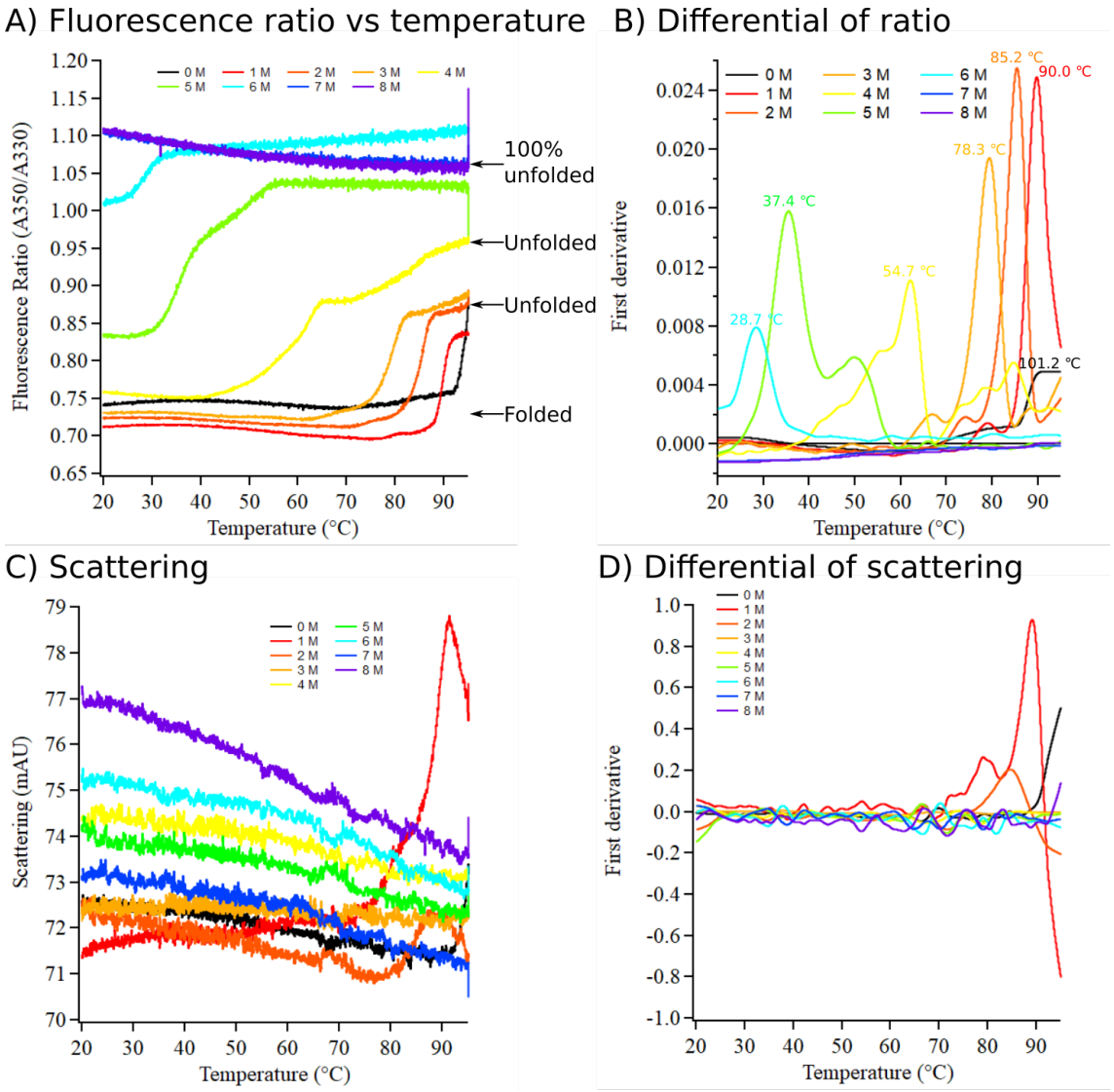
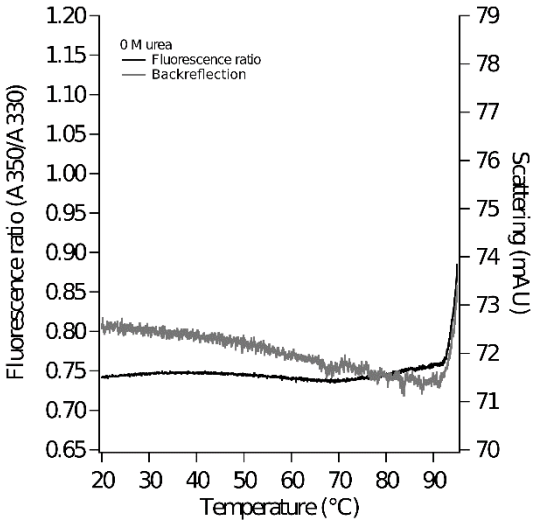


Figure 4-4: NanoDSF and backreflection of GBSV1_Δ1-40_775 at increasing urea concentrations

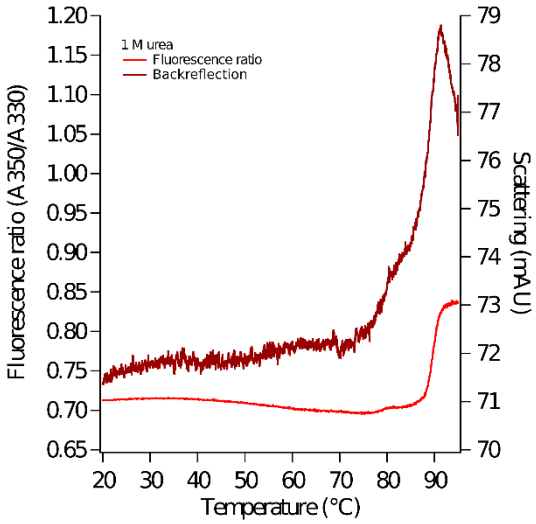
A) Fluorescence ratio (A350/A330) against temperature (℃). B) First derivative of the fluorescence ratio. C) Backreflection measurements for each urea concentration against temperature. D) First derivative of backreflection

If the traces for the fluorescence ratio and scattering are overlaid for each urea concentration tested, there appears to be a relationship between the onset of unfolding and aggregation (Figure 4-5). This is seen most clearly for the graphs for 0-2 M urea (Figure 4-5A-C). When the fluorescence ratio starts to increase, indicating the start of unfolding, scattering also increases, indicating the start of aggregation. At these low urea concentrations, the protein unfolding allows it to aggregate. At higher urea concentrations, the scattering starts at a higher value or the trace slowly decreases as temperature increases. It does not have a sigmoid shape. At 3 M urea, the trace for the scattering appears flat. At concentrations of 4 M or higher, the scattering starts to decrease as the temperature increases.

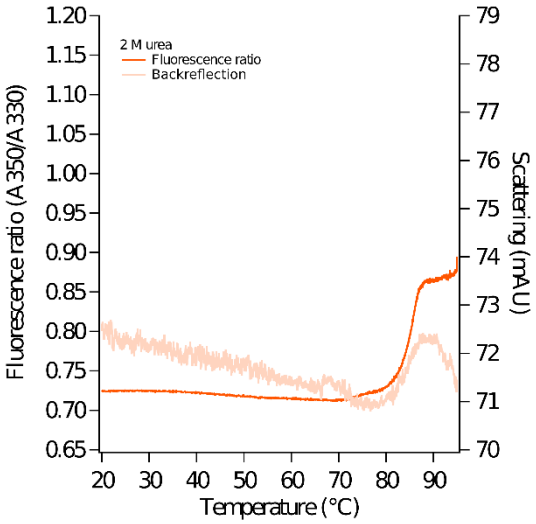
A) 0 M urea



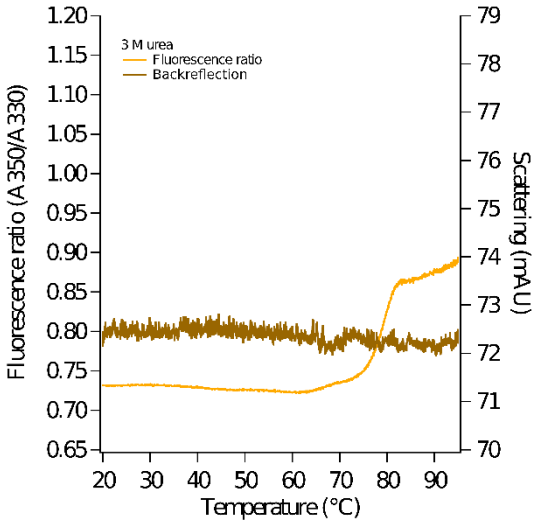
B) 1 M urea



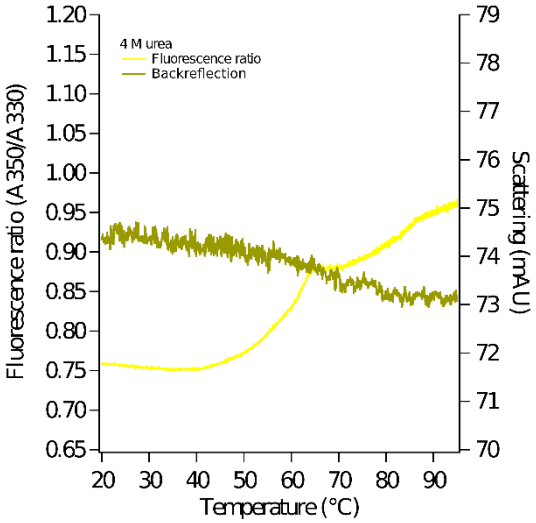
C) 2 M urea



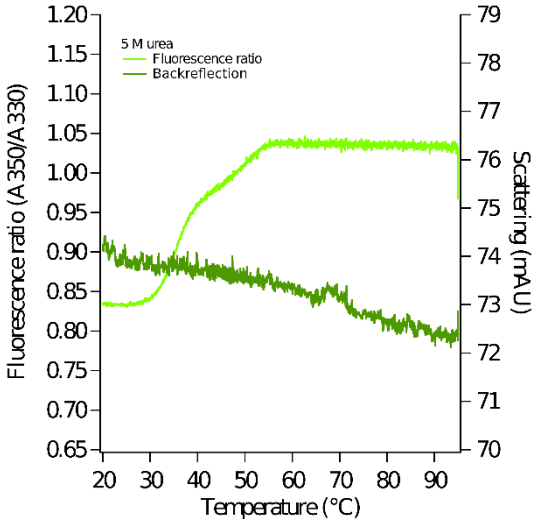
D) 3 M urea



E) 4 M urea



F) 5 M urea



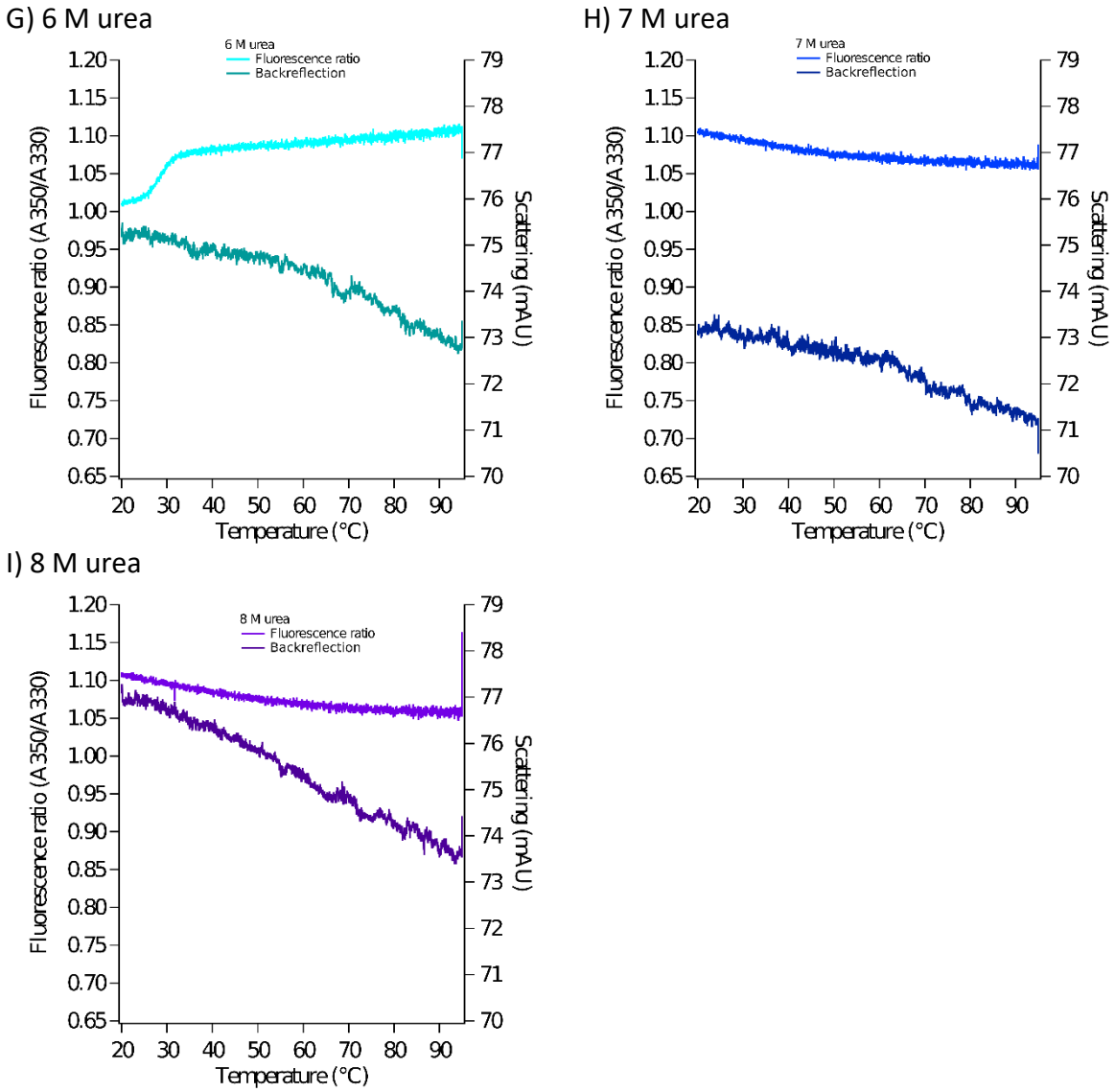
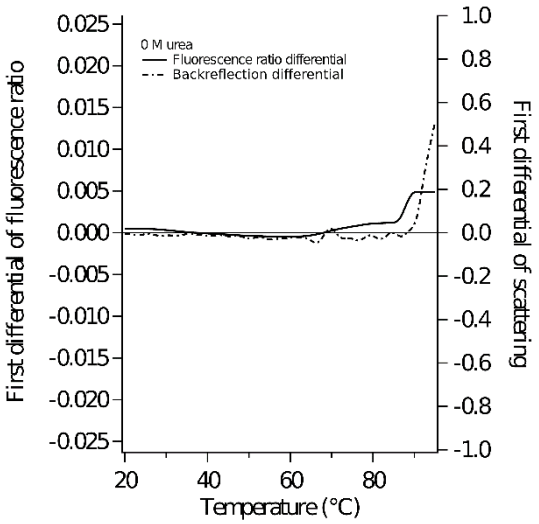


Figure 4-5: Overlay of fluorescence ratio and Dbackreflection

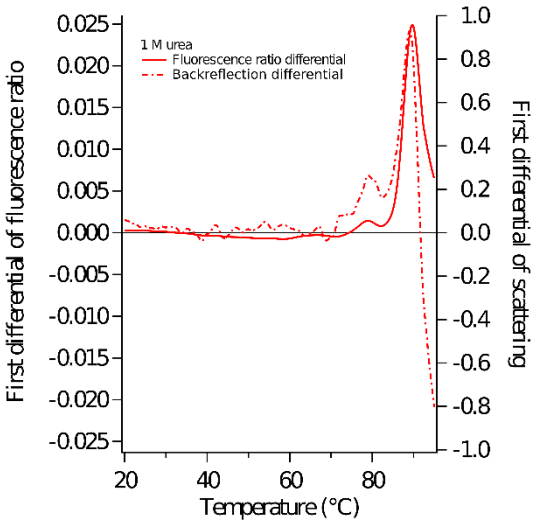
The colour of the traces for fluorescence ratio is the same as for Figure 4-4. The fluorescece ratio is plotted on the left axis. The backreflectionis plotted on the right axis. The scale of the axis is the same for all the graphs. The traces for the backreflection use a darker shade of the same colour. A) 0 M urea. B) 1 M urea. C) 2 M urea. D) 3 M urea. E) 4 M urea. F) 5 M urea. G) 6 M urea. H) 7 M urea. I) 8 M urea.

The relationship between the onset of unfolding and aggregation at lower urea concentration is also seen when the differentials of fluorescence ratio and scattering are overlaid (Figure 4-6). At 0 M urea, when the protein has unfolded, it starts to aggregate (Figure 4-6A). At 1-2 M urea, the progress of unfolding and aggregation are more in sync (Figure 4-6B-C). At higher urea concentrations, there is no apparent change in aggregation.

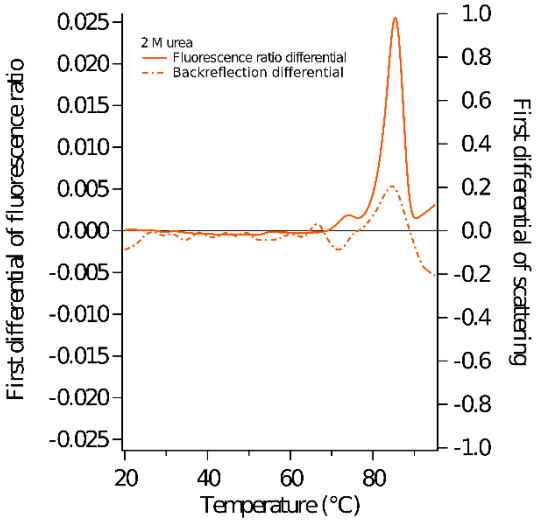
A) 0 M urea



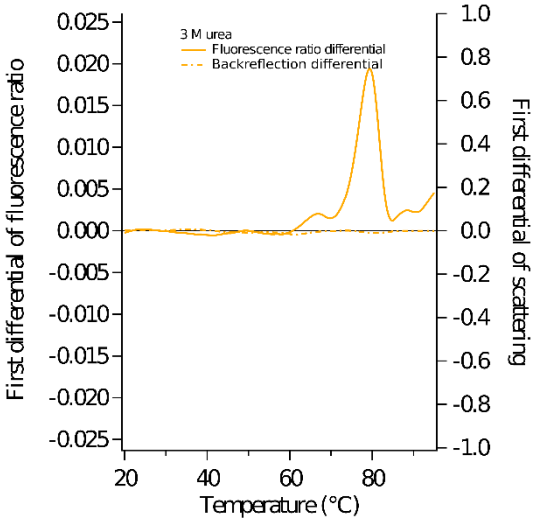
B) 1 M urea



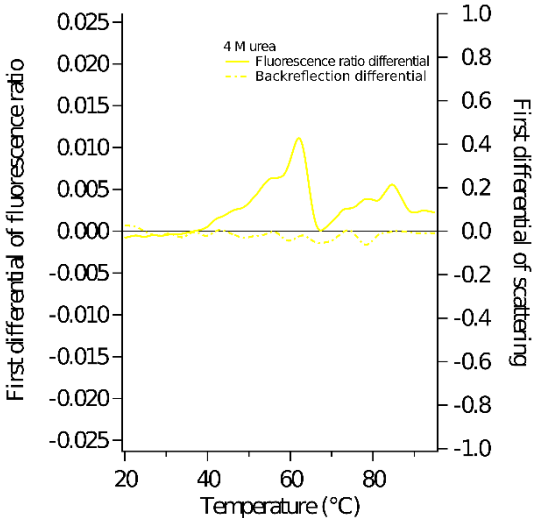
C) 2 M urea



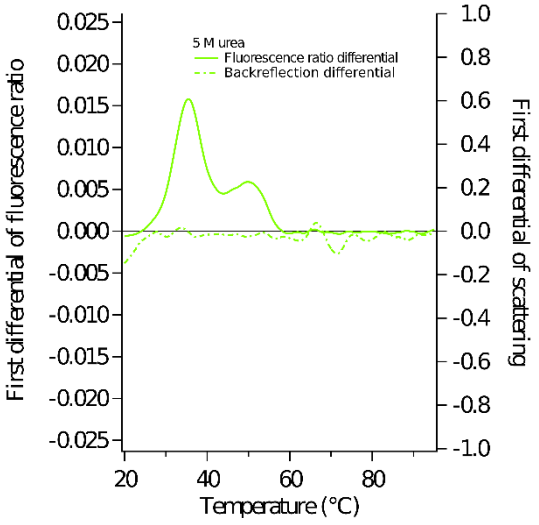
D) 3 M urea



E) 4 M urea



F) 5 M urea



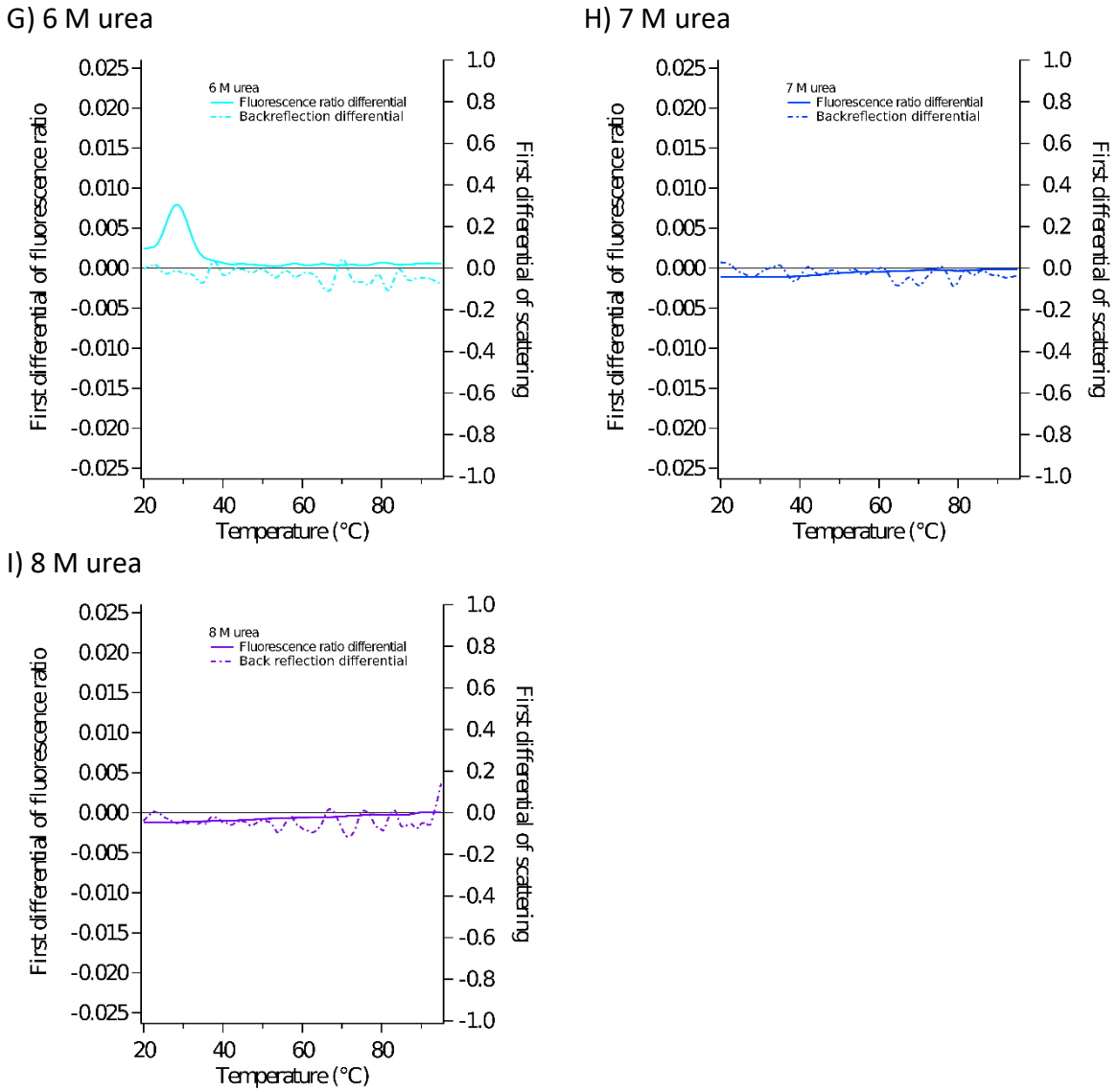


Figure 4-6: Overlays of the first derivatives for fluorescence ratio and backreflection

The traces for the fluorescence ratio are the solid line. The traces for the backreflection are dashdot lines. The scales for all of the graphs is the same. A) 0 M urea. B) 1 M urea. C) 2 M urea. D) 3 M urea. E) 4 M urea. F) 5 M urea. G) 6 M urea. H) 7 M urea. I) 8 M urea.

The fluorescence ratio traces for each urea concentration were fitted using equation (35) for thermal denaturation. This gives the thermal midpoint of unfolding (T_m) for each urea concentration (Figure 4-7A). As the urea concentration increases, the T_M decreases as the protein becomes easier to unfold. It also gives the enthalpy of denaturation for each urea concentration (Figure 4-7B). By 5 M urea, it has almost reached 0 and seems to plateau. The modified Gibbs-Helmhotz equation (equation (35)) was used to calculate the $\Delta G_{D\text{-}N}$ for each urea concentration at different temperatures (Figure 4-7C). The point at which the line crosses the X-axis is the T_m. The y-intercept is the ΔH_m. As the urea concentration increases, the $\Delta G_{D\text{-}N}$ line appears to move left and the gradient gets lower. However, if the whole Gibbs-Helmhotz equation was used including ΔC_P (see Chapter 1), then the lines would be inverted parabolas and the transition may appear as the curves translating down as the concentration

of urea increased (Figure 4-1). If this is how the thermal stability is changing, then the T_s would be the same for each urea concentration. However, since a simplified formula was used to find the ΔG_{D-N}, (equation (35)), the T_s cannot be found.

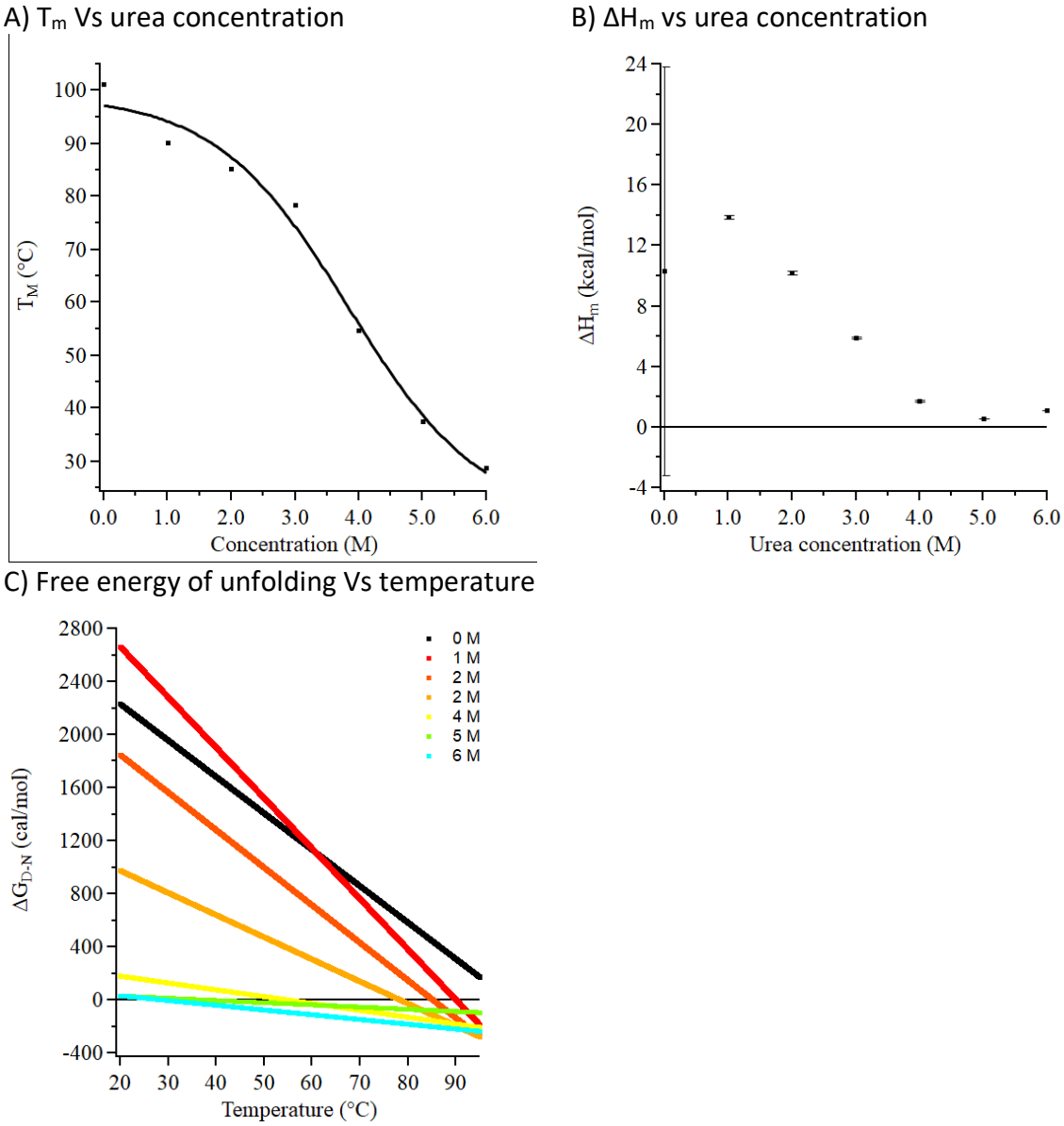


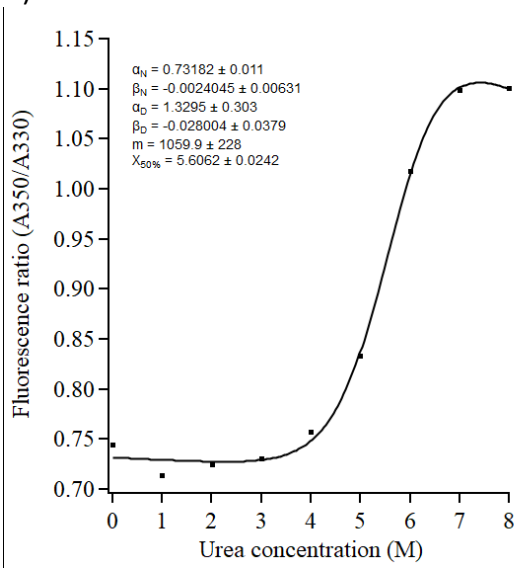
Figure 4-7: Free energy of unfolding Vs temperature for each urea concentration

Parameters for A) Melting temperature vs urea concentration. B) Enthalpy of denaturation for each urea concentration tested. C) Free energy of unfolding vs temperature for each urea concentration.

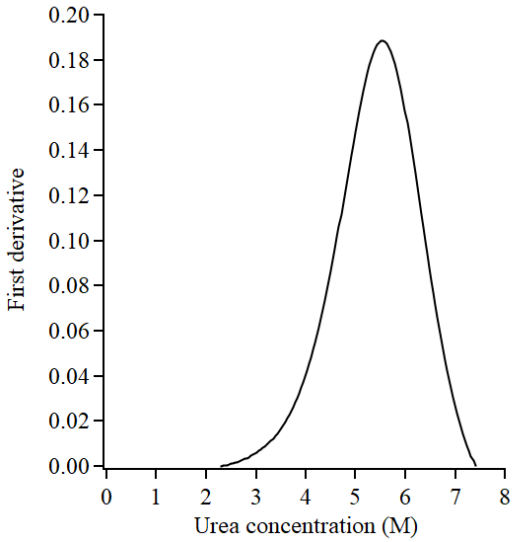
The fluorescence ratio at 25 ℃ at each urea concentration was plotted (Figure 4-8A). The data points were fitted using equation (34). This gave the midpoint of denaturation for GBSV1_Δ1-40_775 with urea as 5.6 M urea. The differential of the fitted curve shows that the fluorescence ratio starts to increase at ~ 2-3 M urea. The scattering at 25 ℃ was also measured, but does not show a clear pattern like fluorescence ratio. The ΔG_{D-N} for urea at 25 ℃ was found using equation (35). The x-intercept is the denaturation midpoint, 5.6 M. The y-intercept is 5.942 kcal mol^{-1} for 0 M urea. As the urea

concentration increases, the free energy of unfolding decreases until it passes 5.6 M, when it becomes energetically favourable.

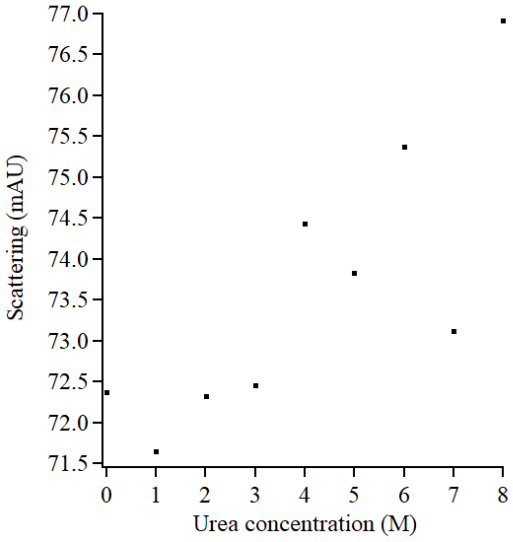
A) Fluorescence ratio at 25 °C



B) Differential of the fluorescence ratio at 25 °C



C) Backreflection at 25 °C vs urea concentration



D) ΔG_{D-N} Vs urea concentration

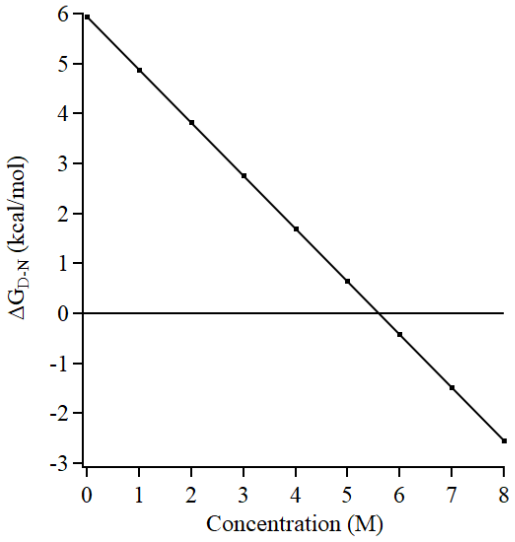


Figure 4-8: Fluorescence ratio and scattering at 25°C

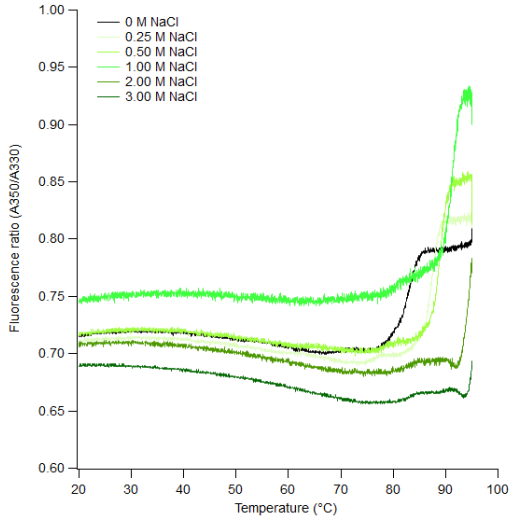
A) Fluorescence ratio vs urea concentration at 25 °C. B) First differential of the fluorescence ratio. C) Backreflection scattering at 25 °C vs urea concentration. D) ΔG_{D-N} Vs urea concentration at 25 °C. The graph is drawn using the equation for ΔG_{D-N} for chemical denaturants (see Chapter 1). The X-intercept is 5.6 M. The y-intercept is 5.942 kcal mol^{-1}.

4.3.4. Unfolding with different salt concentrations and 2 M urea

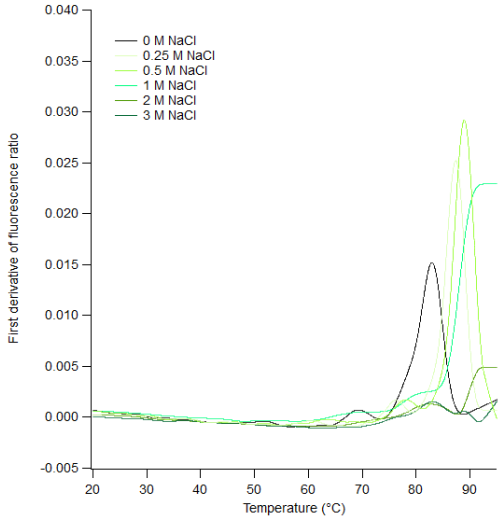
The inclusion of 2 M urea allows the protein to unfold and the effect of NaCl can be seen (Figure 4-9). With no NaCl, GBSV1_Δ1-40_775 unfolds relatively easily. As the NaCl concentration increases, the peak of the derivative shifts further to the right; the melting temperature of the protein increases.

Above 1 M NaCl, the protein does not fully unfold within the temperature range. NaCl appears to make GBSV1_Δ1-40_775 more thermally stable. However, as the NaCl concentration increases, the protein appears to aggregate more easily. The first derivative of scattering shows two peaks at ~ 82 °C and another that does not finish within the temperature range tested (Figure 4-9D).

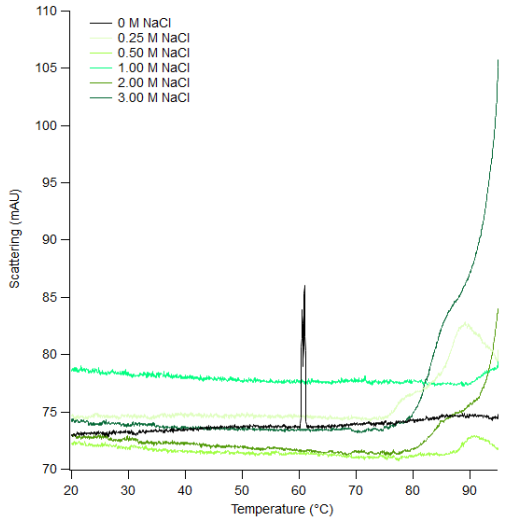
A) A350/A330 ratio vs temperature



B) First derivative of A350/A330 vs temperature



C) Scattering



D) Fist derivative of scattering

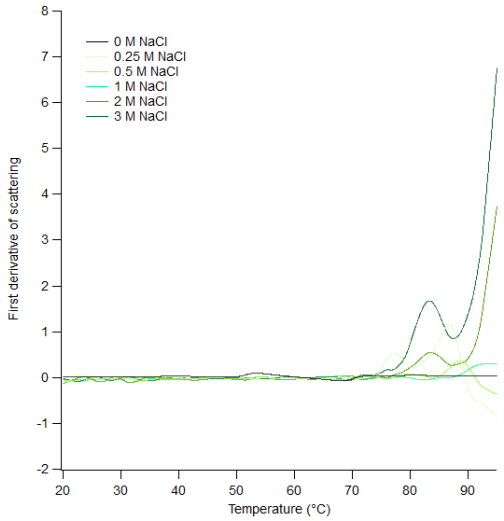


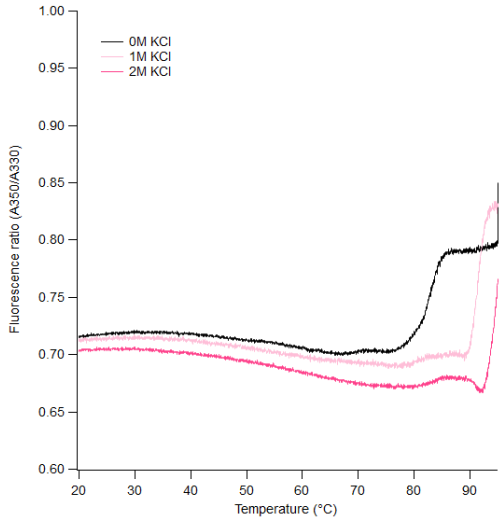
Figure 4-9: NanoDSF with 2 M urea and different concentrations of NaCl

A) fluorescence ratio (A350/A330). B) first derivative of the fluorescence ratio. C) Scattering. D) First derivative of the scattering.

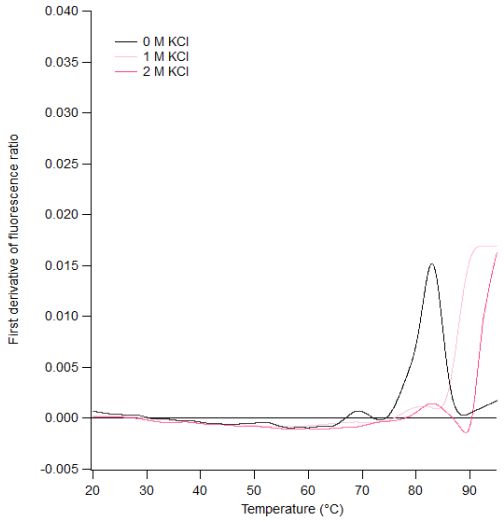
The inclusion of KCl has a similar effect to NaCl, in that it increases the melting temperature of GBSV1_Δ1-40_775 (Figure 4-10). Unfortunately, 2 M KCl was the highest reasonable concentration to test due to inclusion of 2 M urea. As the KCl concentration increases, the melting temperature increases. It was important to check the stability of GBSV1_Δ1-40_775 in KCl since it is used in buffers in nanopore experiments. Similar to the NaCl results (Figure 4-9), higher concentrations of KCl cause the appearance of two peaks in the first derivative of the scattering, one of which does not reach a

maximum within the temperature range tested. KCl appears to increase the T_M, but makes it more prone to aggregation.

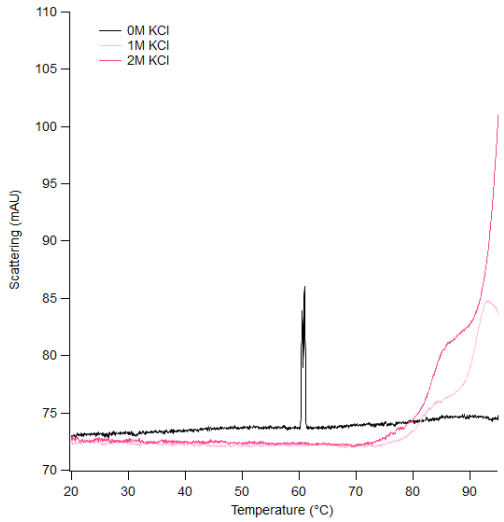
A) Fluorescence ratio vs temperature



B) First derivative of fluorescence ratio



C) Scattering



D) First derivative of scattering

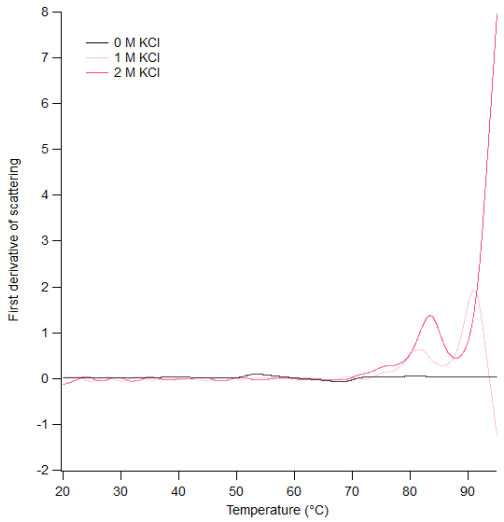


Figure 4-10: nanoDSF with 2M urea and different concentrations of KCl

A) fluorescence ratio (A350/A330) for different KCl concentrations as temperature increases. B) First derivative of the fluorescence ratio. C) Scattering. D) First derivative of the scattering.

At 0.5 mM $MgCl_2$, the T_M appears to stay the same, but the curve for the fluorescence ratio is higher relative to the 0 mM $MgCl_2$ curve, which suggests that more tryptophans are being exposed. Perhaps the protein is unfolding more than without $MgCl_2$, or unfolds in a different way. It also appears to continue increasing at a lower gradient, possibly suggesting another unfolding event. NaCl and KCl appear to increase the melting temperature as their concentration increases. However, as $MgCl_2$ increases the melting temperature appears to hit a ceiling of ~ 89 ℃ at 5 mM and above. There is no scattering for the $MgCl_2$ traces, which suggests that the protein is not aggregating as the temperature increases (Figure 4-11), unlike the traces for NaCl (Figure 4-9) or KCl (Figure 4-10).

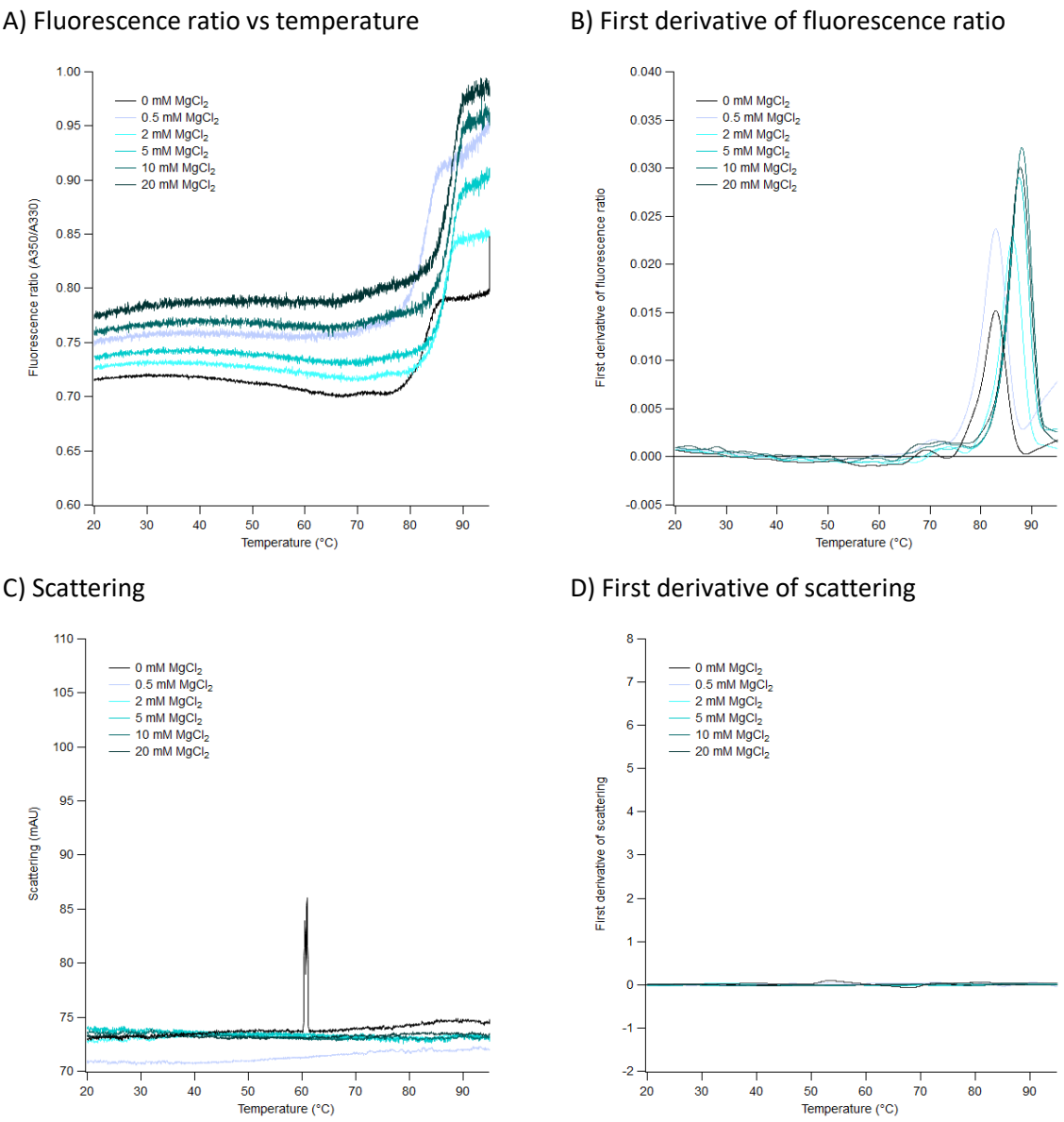
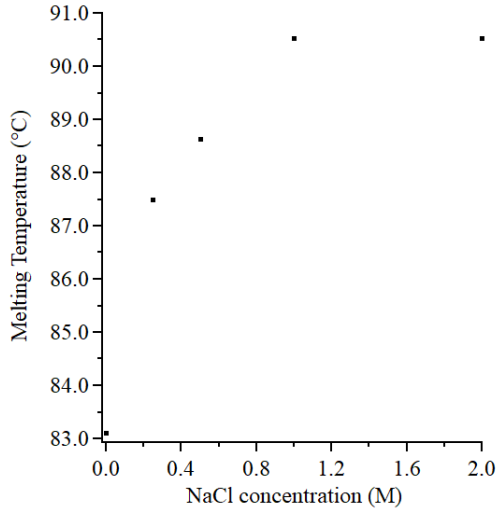


Figure 4-11: NanoDSF with 2M urea and different concentrations of $MgCl_2$.

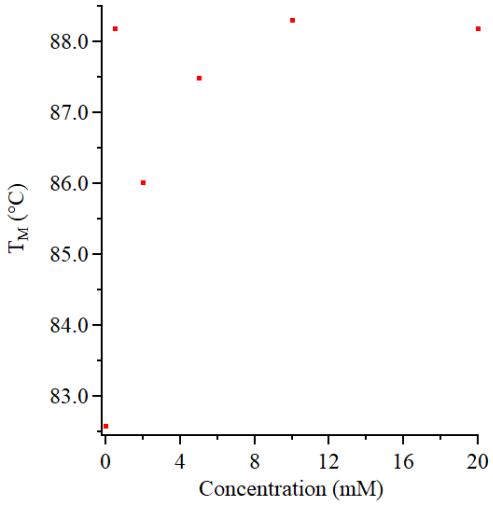
A) fluorescence ratio of A350/A330 against temperature. B) First derivative of the fluorescence ratio. C) Scattering. D) First derivative of the scattering.

The unfolding temperature, ΔH_m, and $\Delta G_{D\text{-}N}$ were found for NaCl and $MgCl_2$. The parameters could not be found for 3 M NaCl. The T_m increases with NaCl concentration. It appears to plateau at ~ 91 °C. More data points and repetitions are required for confidence and curve fitting. The ΔH_m values appear to plateau at ~160 kcal.mol^{-1}. The lines for $\Delta G_{D\text{-}N}$ appear to shift right, showing the increase in T_m. As the NaCl concentration increases, the gradient generally decreases. The traces for the fluorescence ratio and scattering at 25 °C do not show a clear sigmoid curve like those for urea, so were not analysed (data not shown). The traces for $MgCl_2$ follow a similar pattern (Figure 4-12D-F). The T_M appears to increase with $MgCl_2$ concentration to a plateau of ~ 88 °C. The ΔH_m also peaks, then decreases. Since the T_M increases then plateaus, the lines for $\Delta G_{D\text{-}N}$ appear to shift right with the inclusion of $MgCl_2$. The gradient also appears to decrease with increasing $MgCl_2$ concentration.

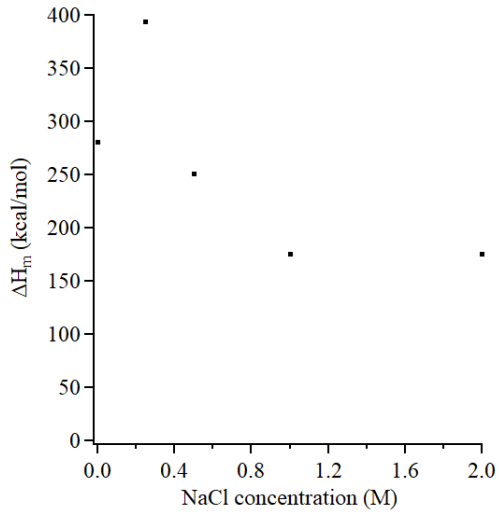
A) T_m vs NaCl concentration



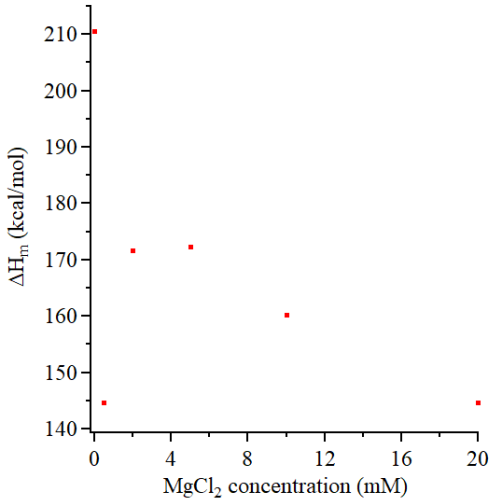
D) T_m vs $MgCl_2$ concentration



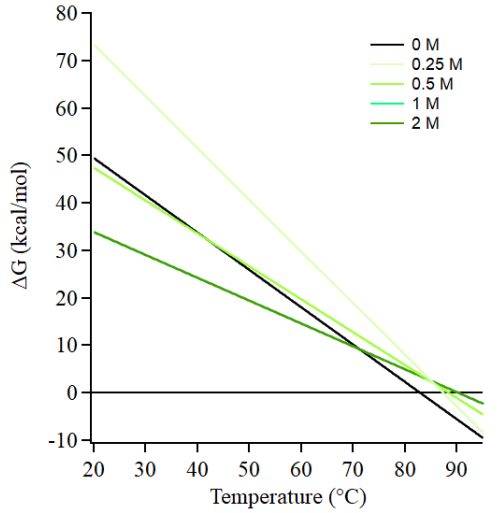
B) ΔH_m vs NaCl concentration



E) ΔH_m vs $MgCl_2$ concentration



C) ΔG_{D-N} vs temperature



F) ΔG_{D-N} vs temperature

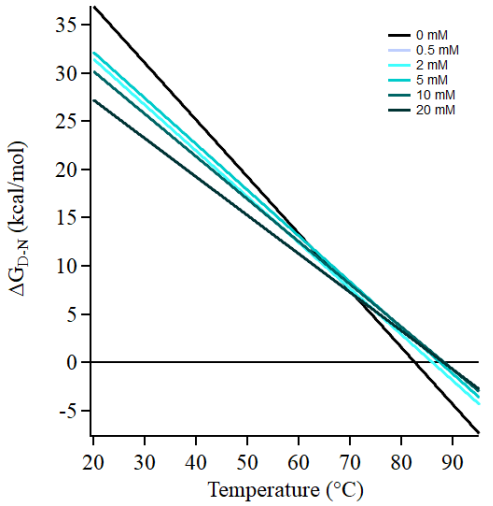


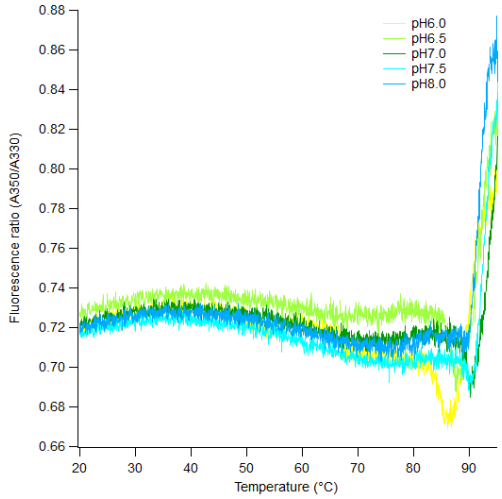
Figure 4-12: ΔG_{D-N} with different concentrations of NaCl and $MgCl_2$

A-C) NaCl graphs. D-F) $MgCl_2$ graphs. A, D) Melting temperature vs concentration. B, E) ΔH_m vs concentration. C, F) ΔG_{D-N} vs temperature.

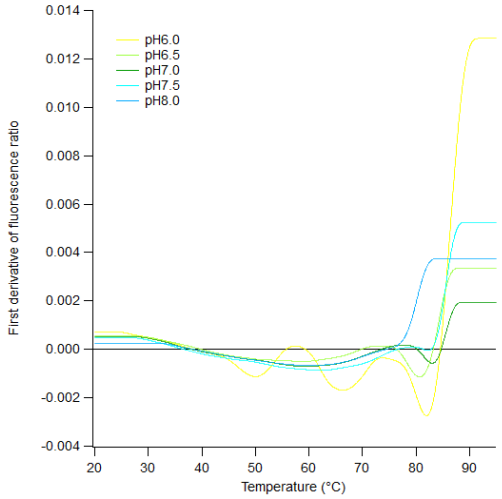
4.3.5. pH stability

To assess the melting temperature empirically, curves were fitted to the raw traces using the thermal fit equation above equation (35). The melting temperature for each pH in each repeat is shown in Figure 4-14B. The melting temperatures were plotted against pH (Figure 4-14). The average melting temperature does not significantly change as a result of pH changes. As the pH increases, the scattering tends to decrease. At pH6, the first derivative of the scattering is lower than the other pH values. This suggests that the protein aggregates at lower temperatures. Since the scattering of the pH6 trace is so high compared to the others, it may already be slightly unfolded.

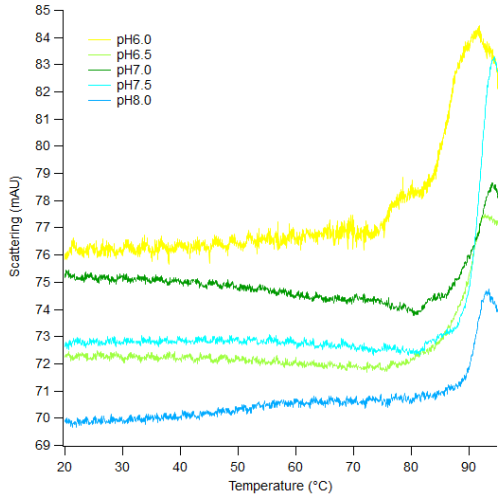
A) Ratio (A350/A330) vs temperature



B) First derivative of ratio vs temperature



C) Scattering



D) First derivative of scattering

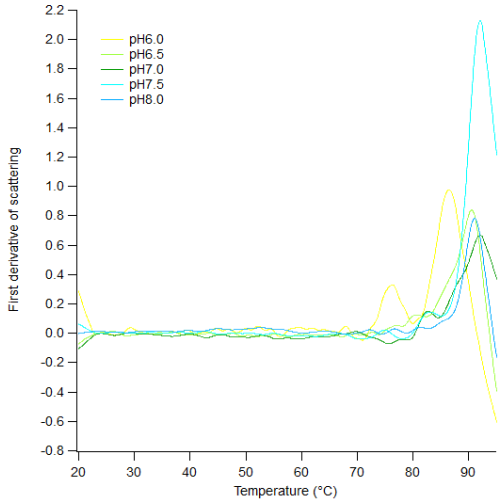
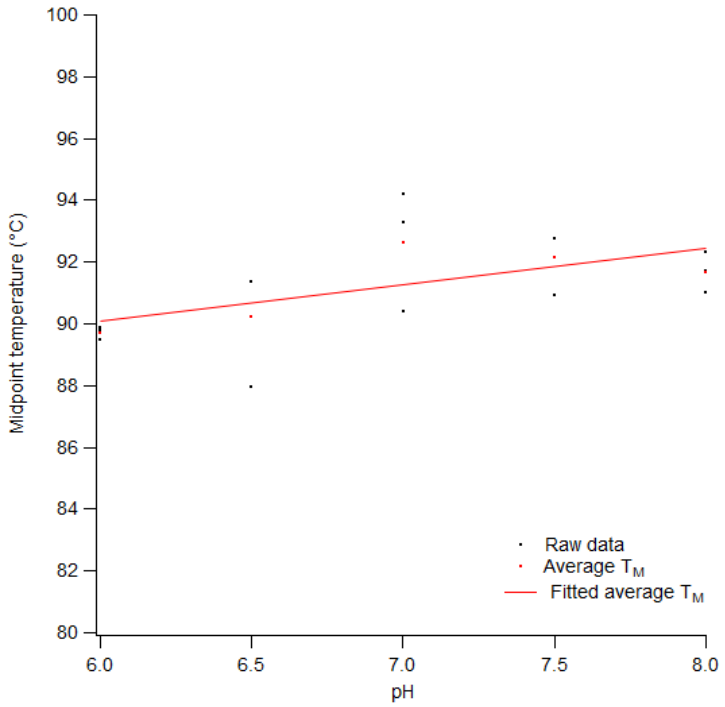


Figure 4-13: nanoDSF for GBSV1_Δ1-40_775 in different pH buffers

A) Fluorescence ratio (A350/A330). B) First derivative of the fluorescence ratio. C) Scattering. D) First derivative of scattering.

A) Melting temperature vs pH



B) Melting temperature for each repeat at each pH

pH	Melting temperature (℃)			
	Set1	Set2	Set3	Average
6.0	89.785	89.849	89.474	89.7027
6.5	87.938	91.352	91.35	90.2133
7.0	90.407	94.207	93.276	92.63
7.5	90.900	92.735	92.741	92.1253
8.0	91.000	91.716	92.308	91.6747

Figure 4-14: Average melting temperature against pH.

A) Graph of melting temperature (T_m) vs pH. B) table of values from fitting sigmoid curves to the traces.

The electrostatic potential of GBSV1_Δ1-40_775 was found at the different pH values used in the stability test. There is a sharp change between pH6 and pH6.5. The pI of the GBSV1_Δ1-40_775 subunit without the hexahistidine tag is 4.86 (see Appendices). It may be that when assembled as a 12mer, some residues are buried or neighbouring residues mask their effect on the charge of an area. At pH values 6.5 and above, the tunnel lumen becomes negatively charged throughout. At pH6, the stem helices and upper section of the clip domains are positively charged. The top of the wing domains and the crown domains are mostly neutral or negatively charged. The very bottom of the clip is still negatively charged (Figure 4-15). As a passage for dsDNA, the tunnel would be expected to be negative to avoid ionic interactions with the DNA as it passes through. If the DNA made ionic interactions with

the tunnel surface, it may bind the tunnel during DNA packaging. The negatively charged sides ensure smooth passage into the capsid. This simulation may show that the GBSV1 phage may not be viable below a certain pH. However, the portal protein alone is still stable at pH6, so it could be used as a nanopore at pH6. The altered charge of the tunnel may allow for selection of different analytes depending on the pH.

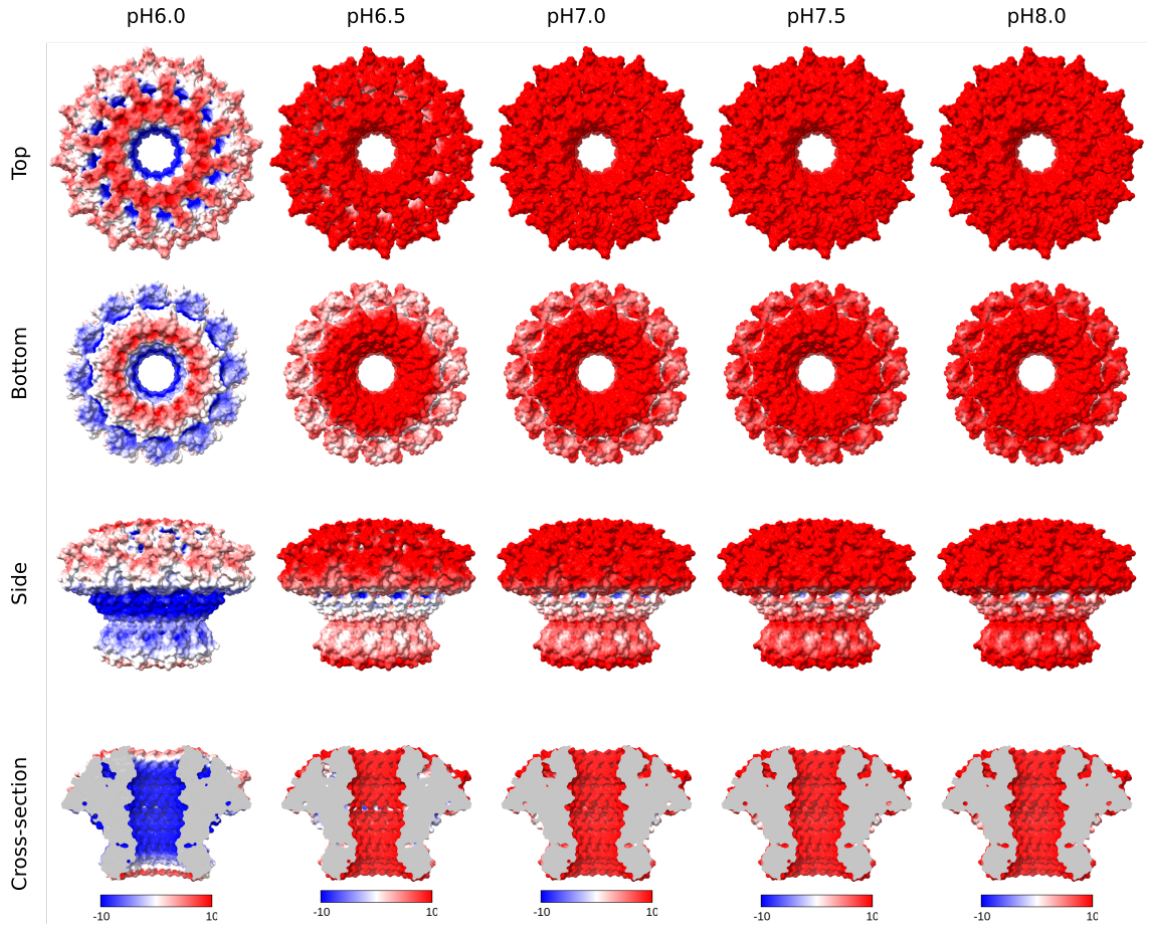


Figure 4-15: Electrostatic potential map of GBSV1_Δ1-40_775 at different pH conditions.

There is a sudden change in the surface charge between pH 6 and pH 6.5 from positive to negative. At pH6, the upper portion of the stem domains is highly positively charged. As the pH increases, this charge shifts to neutral, then more negative. The scale is ±10kT/e-.

The electric fields in GBSV1_Δ1-40_775 were visualised using APBS Tools 2.1 in PyMol (Figure 4-16). At pH6, the electric fields inside the tunnel are positive. The electric fields at the crown and clip domains are slightly negative. These negative electric fields extend to cover the mouth of the portal at both ends. The fields under the wing domains are slightly positive. At pH 6.5 and above, the fields all around the protein are negatively charged, especially around the crown domains, clip domains, and inside the tunnel. The area under the wing domains is less negative. At pH 6.5, the fields inside the tunnel, and from the crown and clip domains are far more negative than those under the wing. This trend persists as the pH increases.

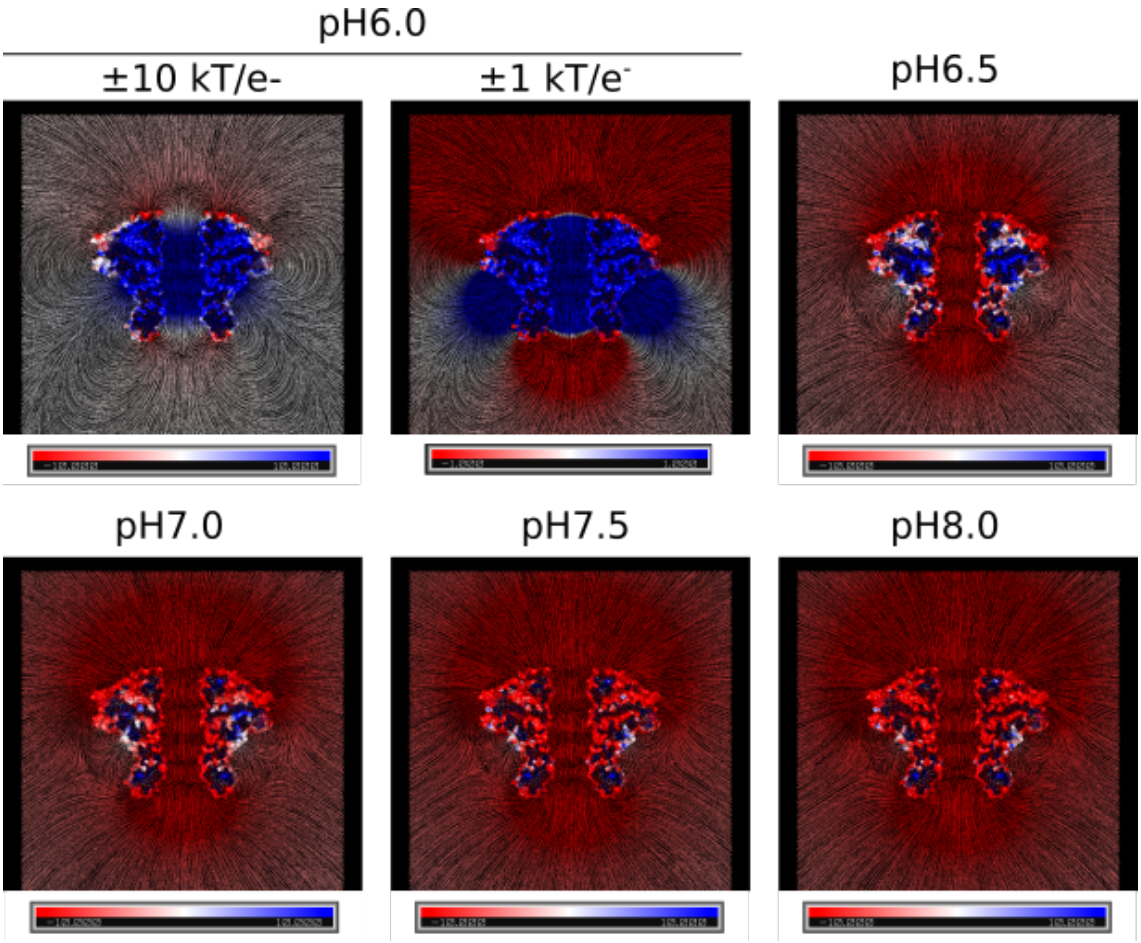


Figure 4-16: Electric fields in GBSV1_Δ1-40_775 at different pH values.

As the pH increases, the electric fields around the portal become stronger. The fields are pH 6 are shown at ±10 kT/e- and ±1 kT/e- in order to highlight the positive fields under the wing domains and negative fields at the crown and clip domains.

The isoelectric surface of the protein at different pH values was calculated (Figure 4-17). As pH increases, a similar pattern is seen to the electrostatic potential and electric field calculations. As the pH increases from pH6, the isoelectric surface away from the portal increases, especially at the crown and clip domains. At higher pH values, the isosurface at the crown domain starts to move around the sides of the portal.

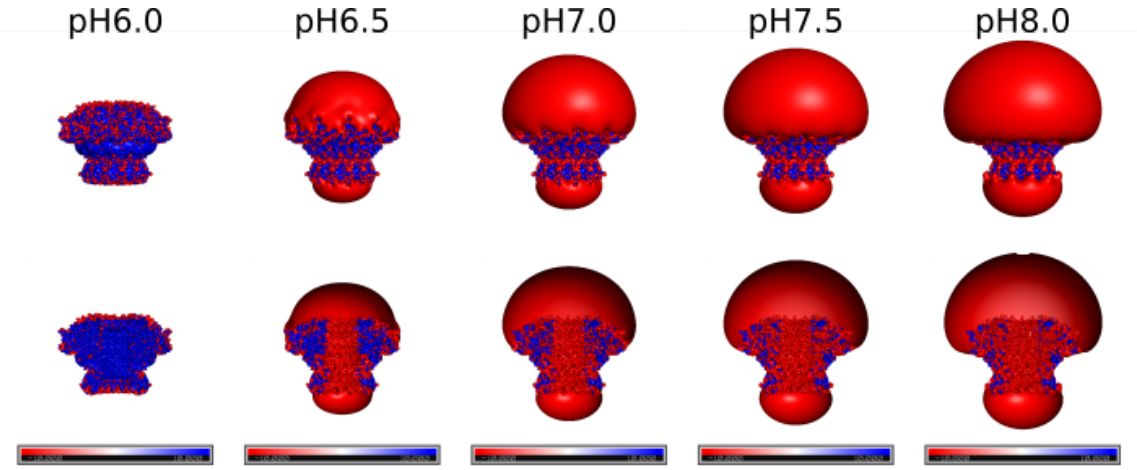


Figure 4-17: Isoelectric surface of GBSV1_Δ1-40_775 at different pH values

The top row is a side view. The bottom row is a cross-section. As pH6, there is no isosurface. As pH increases, negative isosurface extends out from the portal at the crown and clip domains. The scale is ±10 kT/e-. The diagrams were drawn in PyMol using the APBS Tools plugin.

4.3.6. Thermal stability

The GBSV1_Δ1-40_775 protein was left at room temperature or in the fridge at 4 ℃ for several days. The SDS-PAGE gels show a persistent band that does not change in intensity over time for either condition (Figure 4-18). Since the samples were heated and mixed with SDS beforehand, these bands are caused by the presence of the individual subunits, not intact 12mers or other oligomers, so this test cannot state whether the GBSV1 portal protein oligomers disassemble over time at different temperatures. However, the fact that there are still subunits present means that they don't naturally degrade over time at either temperature. Enzymes are not breaking them down over time.

A) 4 ℃ B) Room temperature

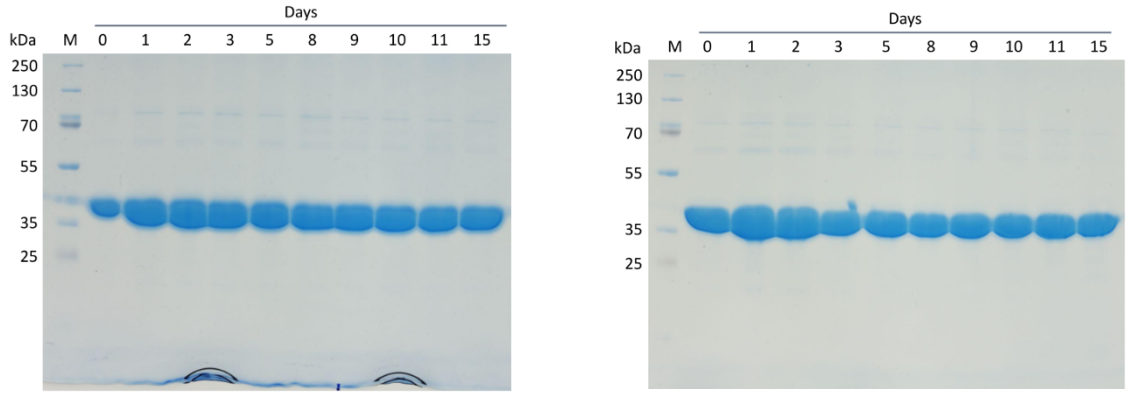


Figure 4-18: Thermal stability test of GBSV1_Δ1-40_775

15% SDS-PAGE gels of thermal stability test of the GBSV1_Δ1-40_775. A) 4℃. B) Room temperature.

4.4. Discussion

The GBSV1_Δ1-40_775 protein appears to be very thermally and chemically stable. It can stay assembled in a range of urea concentrations and pH values.

It will not unfold at high temperatures unless urea is present. The mid-point of denaturation for urea is ~ 5.6 M. The multiple peaks in the 4 M urea trace could be partial unfolding events. At certain temperatures individual domains unfold, then the main chain unfolds at ~ 61 °C. The later peaks roughly align with the peaks seen in the 3 M and 2 M urea traces. It is possible that 4 M urea allows an unfolding event at lower temperatures, and then the same unfolding events occur that are seen in 2 M and 3 M urea. 5 M urea may allow a section of the protein to unfold which would previously have unfolded at higher temperatures in lower urea concentrations, which would explain why the secondary peak overlaps with the head of 4 M urea, but not the peaks at lower concentrations.

The coincidence of the increase in fluorescence ratio and scattering for 1-3 M urea could suggest that the protein starts in a totally folded state, and then aggregates as it unfolds. Past 3 M urea the fluorescence ratio starts at a higher value at 25 °C than lower urea concentrations, so it is already slightly unfolded (Figure 4-8A-B). At this point, the scattering traces stop increasing in step with the fluorescence traces and the derivative of scattering appears relatively flat. There is a sharp jump in scattering between 3 and 4 M urea (Figure 4-8C). The plot of scattering vs urea may be expected to form a sigmoid curve much like for fluorescence ratio vs urea. The data points for 5 and 7 M urea fall below the expected values according to the trend seen in the traces for 4, 6, 8 M urea. The tests should be repeated with different protein concentrations.

The GBSV1_Δ1-40_775 protein could only unfold using temperature in the presence of 2 M urea (Unfolding with different salt concentrations and 2 M urea). 2 M was enough to lower the temperature range over which the protein is thermally stable so that unfolding events could be observed. It may be that 1 M urea is not enough to fully unfold the protein and at least 2 M is needed. When 0-2 M urea was added, protein unfolding and aggregation occurred in step (Figure 4-5 and Figure 4-6). The protein unfolding may allow the protein to aggregate. At higher urea concentrations, urea may act as a chaotropic agent to make the unfolded protein more soluble, which may prevent aggregation. However, protein concentration is a key determinant of aggregation (S. H. Kim et al. 2021). If the protein concentration was higher, the concentration of urea at which the unfolding and aggregation occur may be higher. The shapes of the traces may also be different. The urea denaturation experiments should be repeated with a range of protein concentrations.

While all salts made the melting temperature increase, they also saw the appearance of peaks in the scattering and first derivative of scattering for NaCl and KCl. The inclusion of salts increased the proteins propensity to aggregate. Higher salt concentrations caused it to aggregate more. The first derivative of scattering for both salts had a peak at ~81 °C, and another which does not finish

increasing within the temperature range tested. NaCl and KCl are acting as kosmotropic agents. The scattering and derivative for the $MgCl_2$ traces show no change. Sodium and potassium ions have the same charge, but sodium ions have a slightly smaller atomic radius, which gives them a higher charge density. (Table 4-5). The charge density of Mg^{2+} is much higher than Na^+ or K^+. There are also more Cl^- ions in the same concentration of $MgCl_2$ than in NaCl or KCl. However, the lower concentrations of $MgCl_2$ mean that the ionic strength of the solution is much lower. However, the Mg^{2+} is more of a kosmotropic cation than Na^+ or K^+, which may help keep the protein soluble and prevent it aggregating, even if it has unfolded.

Table 4-5: Atomic radii of ions used in nanoDSF experiments

Ion	Atomic radius (Å)	Charge density (C.mm^{-3})
Na^+ (VI)	1.02	24
K^+ (VI)	1.38	11
Mg^{2+} (VI)	0.72	120
Cl^- (VI)	1.81	8

The values for atomic radii were from (Shannon 1976). The values for charge density of ions were from (Rayner-Canham and Overton 2014).

There is a significant change in the scattering and charge of GBSV1_Δ1-40 induced by the buffer pH. At pH6, the protein has a highly positively charged tunnel, unfolds more quickly, and appears to aggregate more easily. The T_m does not appear to change with pH. The pH may affect the inter-subunit interactions which would allow the protein to aggregate and unfold more quickly. The interfaces of the portal protein are largely positive on one side and largely negative on the other (see Chapter 3), so electrostatic interactions may be involved in portal protein assembly, which is where the ions from NaCl, KCl or $MgCl_2$ may act to stabilise the oligomer and/or subunit. Based on the traces on the refolding experiments, the proteins do not appear to refold (data not shown). The slow decrease in fluorescence ratio could be due to aggregation. But the higher urea concentrations do not see any decrease, so the tryptophans are still exposed.

If 12mers were disassembling, some tryptophans and tyrosines would be exposed. If the subunits unfolded, further tryptophans and tyrosines would be exposed, which could be expected to give two steps in the sigmoid curve, rather than just one. Based on the increases in T_M seen when increasing NaCl, KCl or $MgCl_2$, it appears that each stabilises the protein by different amounts, possibly by different mechanisms. The difference between the use of Na^+ and K^+ ions versus Mg^{2+} ions could be about the ions that shield the GBSV1_Δ1-40_775 subunits and oligomers. An electrostatic potential map shows that the GBSV1_Δ1-40_775 subunits have opposite charges on opposite faces. The positive face would attract anions, while the negative face would attract cations. A charge map shows that GBSV1_Δ1-40_775 is negatively charged across its outer surface (see Chapter 3), even though there

are positive residues on one face. A negatively charged protein would attract a layer of cations around itself in solution. Most of the absorbance change from nanoDSF comes from the changing environments of tryptophans, however, there are only three tryptophans in GBSV1_Δ1-40_775 (Figure 4-19). W61 is buried in the wing domain. For this to be exposed, the protein would have to be totally unfolded. W205 is in the clip domain and part of the three-stranded β-sheet that holds the portal together. W326 is in the last helix of the crown domain. This tryptophan residue may be the easiest to expose to solution if it is not already.

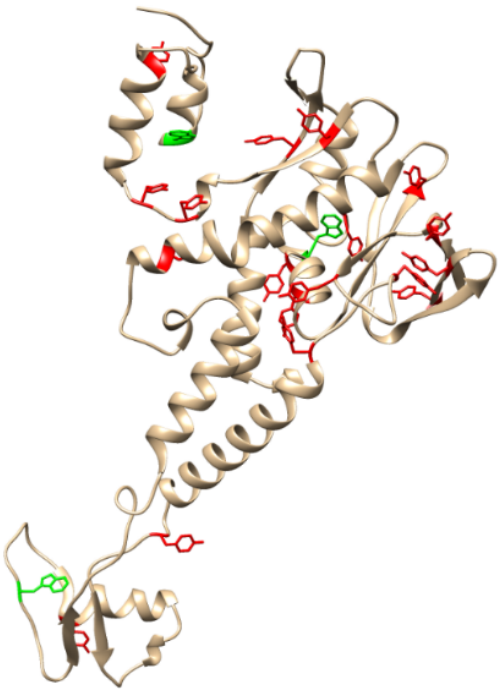


Figure 4-19: Map of the tyrosines and tryptophans on a GBSV1_Δ1-40_775 subunit

Most of the absorbance changes in nanoDSF comes when the environment surrounding tryptophans changes. There are only three tryptophans residues in a subunit: W61 in the wing domain, W205 in the clip domain, W326 in the crown domain.

SDS has previously been used to unfold and alter the charge of protein analytes passing through an SSN. If these were used with the hybrid nanopore, it may alter the charge on the inner surface of the SSN to generate a stronger EOF for inserting the portal protein. It may also generate stronger signals from peptide analytes as they would not have to stochastically unfold before entering the pore, but would already be unfolded. However, the SDS may also unfold the portal proteins themselves, which render such methods pointless. The portal protein's stability against SDS would have to be tested by monitoring the A280 reading at different SDS concentrations and by nanoDSF (Restrepo-Pérez et al. 2017; Soni et al. 2022).

Disulphide bridges have been introduced between subunits of the SPP1 portal protein in order to study the movement of subunits in portal function (Cuervo et al. 2007). Such mutations could be introduced to GBSV1 in order to further fix the shape of the tunnel in position to prevent noise during

nanopore tests. However, such mutations may also stop the shift between open and closed conformations, which is a key feature of portal proteins as nanopores.

A baseline of intact 12mer could be taken. Different treatments could be tested to see the effect on protein disassembly. Heating a sample to different temperatures then allowing it to cool would be another test of thermal stability. The masses of the molecules in the sample afterwards would show if there were still 12mers or if they had broken apart. The same could be done for samples left at different temperatures, e.g. 4 °C or room temperature, or in different concentrations of denaturant, e.g. urea. This could also be useful for testing the different peaks that elute from SEC (see Chapter 3).

4.4.1. Alternative techniques

Since the portal protein is an oligomer, it cannot be said with certainty whether the portal protein is disassembling or unfolding. NanoDSF relies on the presence of tryptophans and, to a lesser extent, tyrosines and phenylalanines in the protein. However, there are relatively few tryptophans and tyrosines in a GBSV1 subunit. Alternative techniques may validate the findings thus far and provide different information.

SYPRO orange is a commonly used protein dye which binds to hydrophobic regions of proteins. It is often used as a gel stain, and can also be used to study protein unfolding (Huynh and Partch 2015). According to electrostatic potential maps and MLP analysis of the GBSV1 portal protein (see Chapter 3), the subunit lacks large areas of hydrophobic residues between the subunits as other portal proteins have (Simpson et al. 2001). However, it may still bind the portal oligomer when it disassembles. As mentioned previously (see Chapter 1), CD can be used to detect secondary structure elements of proteins. As such it can also be used to detect the unfolding of proteins over time in different conditions.

Mass photometry could be used as a complementary technique to test whether the oligomer disassembles under different conditions. This new technique uses the interference of scattered and reflected light from a molecule to calculate its mass (Young et al. 2018).

4.5. Summary

In conclusion, the GBSV1_Δ1-40_775 is soluble in high NaCl and KCl buffers. It cannot be thermally unfolded unless urea is present in the buffer. The midpoint of unfolding for urea is 5.6 M. The protein is also not affected by pH, and does not degrade after being left at 4 °C or room temperature for several days.

Chapter 5 Optimisation of the GBSV1 portal protein as a biological and hybrid nanopore

5.1. Introduction

Previous chapters have established that the GBSV1 portal protein is soluble and stable in different salts and a range of salt concentrations (see Chapter 4). But there are many other conditions it must meet before it can be used as a biological or hybrid nanopore.

5.1.1. Insertion of portal proteins into a lipid membrane

The thermal and chemical stability studies (see Chapter 4) have established that the GBSV1 portal protein is stable in the temperature range 20-90 ℃, pH range of pH 6-8 and at urea concentrations up to 4 M. The phi29 portal protein (also referred to as "connector" in several research publications) is a well-studied biological nanopore (see Chapter 1) which has been successfully inserted into a lipid membrane (Wendell et al. 2009; Haque et al. 2012; Manrao et al. 2012; S. Wang et al. 2013; Ji et al. 2020; Jing 2020; L. Zhang et al. 2023). But the phi29 portal protein has a ring of hydrophobic residues under the wing domains (Simpson et al. 2000; Guasch et al. 2002) which the GBSV1 portal protein lacks (see Chapter 3). Therefore, if the GBSV1 portal protein is to be used as a biological nanopore, an alternative method of membrane insertion would be required. The G20c portal protein has previously been adapted as a biological nanopore following insertion into the lipid membrane (Cressiot et al. 2017b) achieved through chemical addition of porphyrin molecules to cysteine residues introduced at specific positions in the wing domain. However, this method requires time-consuming steps to conjugate the maleimide to the porphyrin and separate portal proteins with and without maleimide-porphyrin. An alternative chemical labelling approach would be to attach maleimide to the lipid which could then interact with the exposed cysteine residue of the portal proteins, but this approach would bear similar disadvantages.

5.1.2. Solid-state nanopore fabrication

Fabrication of a nanopore in solid-state materials, with high precision, is challenging. In the past, the most common method of fabricating SSN in SiN_x chips was by burning a tunnel across the material using an intensive beam of electrons available in a TEM. Using a TEM allows a user to control the size and shape of the SSN using the beam (Min Jun Kim et al. 2007; Hout et al. 2010; H.-M. Kim, Lee, and Kim 2011; Muhammad Sajeer P et al. 2022) and an image of the pore can be taken as soon as it has been fabricated. However, the precise and accurate creation of SSN using this method requires access to a TEM and a skilled user. CDB has emerged in recent years as an alternative method, which produces a pore using the dielectric breakdown phenomenon seen in SiN_x chips, when an electric current passes

through an insulating (dielectric) medium at high enough voltage, to form SSN (see Chapter 1, Controlled Dielectric breakdown).

5.1.3. Wetting

Following the assembly of a device containing an SSN, when liquid is added to the flow cell, some microfluidic principles must be considered. In order for the liquid to completely fill the SSN, the liquid must enter the pore and contact the liquid entering from the other side. This process is called wetting of the nanopore. If the liquid cannot reach or does not enter the tunnel of the pore, this would result in insufficient wetting and formation of an air bubble inside the nanopore. A bubble can also form along the side of the tunnel, effectively narrowing the tunnel diameter, or occlude the entire pore, depending on the progress of drying (Giacomello and Roth 2020). Bubble formation is prevented by treatment with Piranha cleaning solution, which makes the chip surface, including the tunnel, hydrophilic. The entry of a liquid into a pore is described by the following equation (40) (Franken et al. 1987):

$$\Delta P = -\frac{2B\gamma_L \cos\theta}{r} \tag{40}$$

where ΔP is the liquid entry pressure (LEP), B is the geometric factor (for cylindrical pores, B=1), γ_L is the surface tension, θ is the intrinsic liquid contact angle between the liquid and solid membrane, and r is the pore radius. The surface tension values of isopropanol and water are 21.7 and 71.97 mN/m, respectively (Lange and Dean 1973). Different liquids will also have different contact angles (θ). The capillary imbibition equation (equation (41)) gives the distance by which a liquid will fill a capillary:

$$l(t) = \sqrt{\frac{r\gamma_L \cos(\theta)}{2\eta}.t} \tag{41}$$

where l is the liquid-filled length, η is the solution viscosity, and t is time (Churaev et al. 1981). The viscosity values of isopropanol and water are 2.052 (Paez and Contreras 1989) and 1.0016 mPa.s (Rumble, Brunno, and Doa 2023), respectively. Due to its network of hydrogen bonds, water has a much higher surface tension. While isopropanol is more viscous, its lower surface tension allows it to more easily enter nanopores, thus promoting pore wetting.

5.1.4. Electrokinetics of portal protein insertion into an SSN

The same considerations about the entry of an analyte into a nanopore (see Chapter 1) also apply for formation of a hybrid nanopore. The surfaces of SSN and biological nanopores are charged, so both will have an EDL around them including inside their tunnels. Cations and anions are uniformly distributed in the bulk solution, so it is electroneutral. The different layers of the EDL have non-uniform concentrations of cations and anions, so they will be charged. The λ_D for K^+ and Cl^- ions in 0.4 M KCl at

room temperature is 0.48 nm. In this case, λ_D is smaller than the radius of the pores of both the SSN and portal protein. The ε_r, ε_0, η and electric field strength will be the same for EP of the portal protein and the EOF. The velocity of EP and EOF can be changed by changing the applied voltage which will change the electric field. As such, the relative velocities of EP and EOF will depend on the ζ-potentials of the SiN_x and portal protein for the EOF and EP, respectively. As the ionic concentration and pH increase, the ζ-potential decreases (Tabeling 2023). If the EOF through the pore and the EP velocity of a particle move in the same direction, then they will act in co-operation. On the other hand, if they move the particle in opposite directions, then they will be in competition. By carefully tuning the pH, salt concentration and salt identity, it can be possible to change the directions of EP and/or EOF to make them co-operative or competitive (Chinappi et al. 2020; Asandei et al. 2016; Saharia, Bandara, Karawdeniya, Hammond, et al. 2021).

The pH_{pzc} (see Chapter 1) for SiN_x in 0.1 M KCl was estimated to be pH 4.1 (K. Lin et al. 2021). At pH values higher than 4.1, the SiN_x is negatively charged, so cations will adsorb to its surface to form an electrical double layer (K. Lin et al. 2021). Under a positive voltage, the cations will move towards the ground electrode (Figure 5-1). The pI of GBSV1_Δ1-40_775 was estimated to be pH 5.12 by ProtParam (Gasteiger et al. 2005), so its ζ-potential should be negative. The ζ-potential can only be found experimentally. At pH values higher than 5.12, it will have a net negative charge. Under a positive voltage, electrophoresis will move it towards the working electrode. Therefore, at a greater range of pH values, the EOF and EP acting on the portal protein will move it in opposite directions (Figure 5-1).

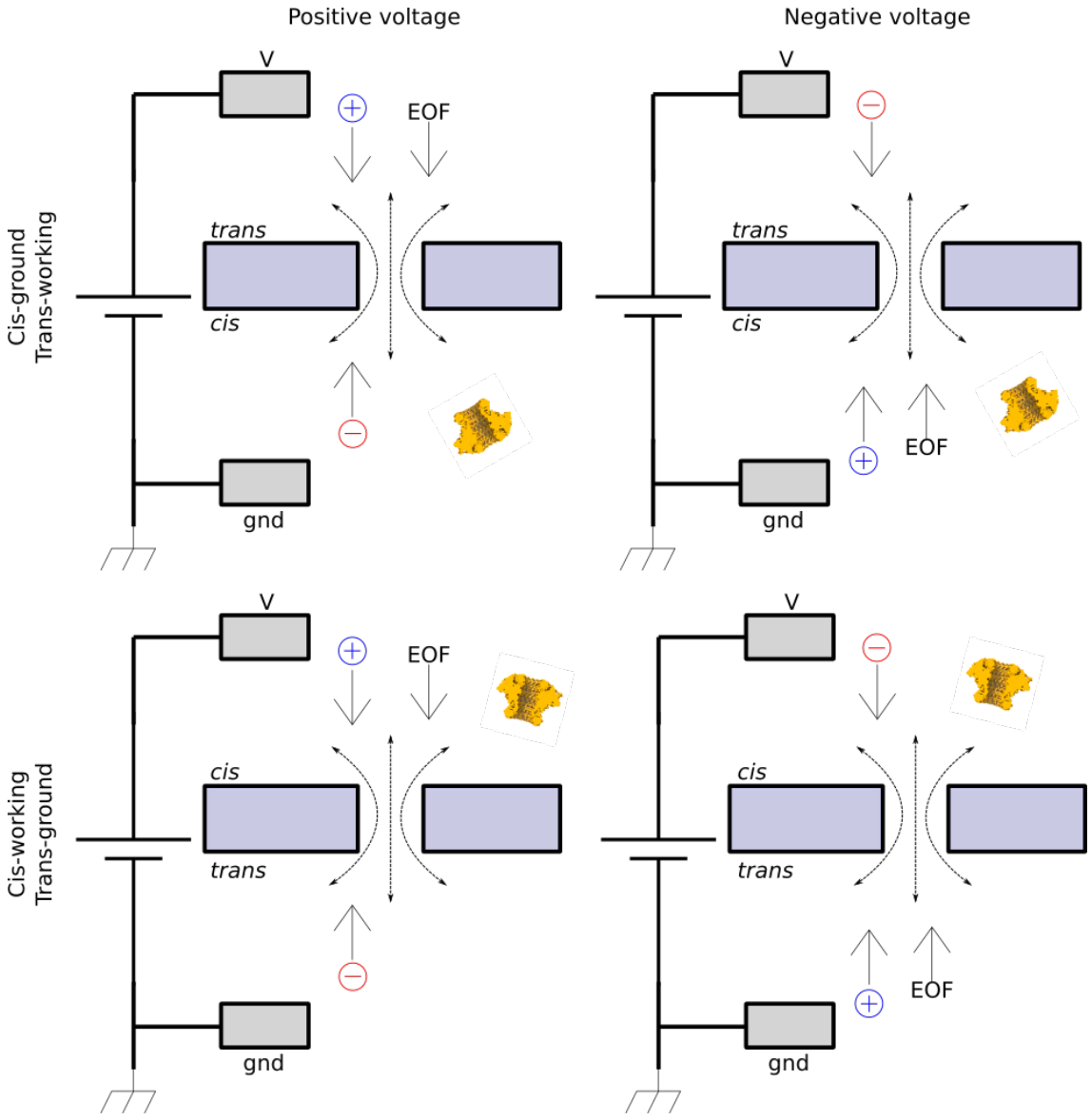


Figure 5-1: Directions of EP and EOF through an SSN

The portal protein is always added to the cis compartment unless otherwise stated. The standard set up is to have the cis compartment connected to the ground electrode (top). If the portal protein enters the SSN with the cis compartment connected to the ground electrode under positive voltage, it is because of electrophoresis acting on the portal protein. If it inserts under negative voltage, it is because of the EOF through the SSN. The opposite is true if the cis compartment is connected to the working electrode (bottom). Not to scale.

5.1.5. Rationale for mutants

The GBSV1_Δ1-40_775 protein has an internal pore and hence it could potentially be adopted as a nanopore. However, the cryoEM structure showed the diameter of the internal tunnel is ~4 nm, which may limit its uses to larger analytes. The tunnel loops, which usually form the narrowest point of the portal protein tunnel are retracted, and hence the portal protein appears to be in the open conformation which would easily allow the passage of B-form DNA, which has a van der Waals diameter of about ~2.3 nm (Bayfield, Steven, and Antson 2020). The closer the nanopore diameter is to the dimensions of the analyte, the better the signal to noise ratio (Waduge et al. 2017). Other

analytes, e.g. ssDNA or ssRNA, have smaller diameters compared to dsDNA, and so would give poorer signal to noise ratios. Previously, when the G20c portal protein was inserted into an SSN (Cressiot et al. 2018), the diameter shrank compared to when it was inserted into a lipid bilayer (Cressiot et al. 2017b). However, such tunnel reduction will depend on the size and shape of the SSN into which the GBSV1 portal protein is inserted, and also the EOF through the SSN. The size of the internal tunnel could be decreased by SDM through the introduction of bulkier amino acids into the tunnel loops. Furthermore, based on the disordered regions of the protein identified during sequence analysis (see Chapter 2) and cryoEM structure determination (see Chapter 3), it was necessary to create variants of the protein without the disordered regions which may interfere with nanopore readings (see Chapter 2).

5.1.6. Radius of gyration, R_g

The radius of gyration (gyradius, k, R_g) of an object is the distance from the axis of rotation to the point at which, if the whole mass of the object was concentrated at that point, would give the same moment of inertia as the original object. The R_g can be used as a stand-in for the approximate radius of a molecule when tumbling in solution. Thus, it can be used to estimate the fraction of the nanopore cross-sectional area blocked by an analyte, or in this case, a portal protein forming a hybrid pore. The moment of inertia for a rigid body (if treated as a series of particles) is given by equation (42) (Atkins and Escudier 2019).

$$I = \sum_i m_i r_i^2 \tag{42}$$

The parameter I is the moment of inertia. m_i and r_i are the mass of particle i and the distance between particle i and the axis of rotation, respectively. Thus, R_g is given by equation (43).

$$Radius\ of\ gyration, R_g, I = mk^2 \tag{43}$$

5.1.7. Expected current in the hybrid nanopore

Once the hybrid nanopore has been formed, the ion current between the electrodes will drop due to the reduction of the diameter of the SSN pore (Figure 5-2A). When the portal protein inserts with the clip domain towards the middle of the SiN_x chip, as seen previously with G20c (Cressiot et al. 2018), most of the current will pass through the central tunnel of the portal protein. However, some quantity of current may also pass between the outside of the portal protein and the SSN tunnel, resulting in leakage. The level of the leakage current depends on how snugly the portal protein fits into the SSN, which depends on the size and shape of the SSN and the orientation in which the portal protein inserts into the SSN. The electrokinetic forces of EP and EOF will also continue to act on the portal protein once it has formed a hybrid nanopore (Figure 5-2B). At most pH values, the EP and the EOF will pull

the portal protein in opposite directions. This may affect the dwell times of the GBSV1 portal protein inside the SSN, though it may also depend on the voltage used.

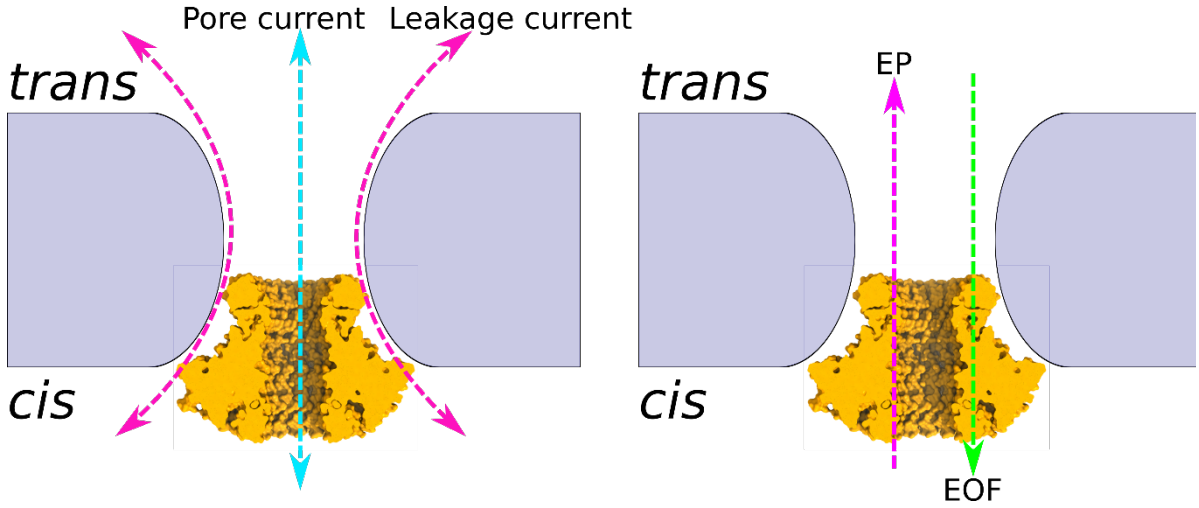


Figure 5-2: Forces acting on the hybrid pore.

A) An electric current will flow through the portal protein lumen (blue line). Depending on the size of the SSN and the orientation of the portal protein, there will also be a leakage current between the portal protein and the SSN (pink lines). Depending on the size of the SSN relative to the portal protein, there may also be an EDL between the outer surface of the portal protein and SSN surface. B) Electrokinetic forces acting on the portal protein when inside the SSN. The cis compartment is connected to the ground electrode. The trans compartment is connected to the working electrode. The portal protein is negatively charged. Not to scale.

5.1.8. Hybrid nanopore optimisation

To optimise hybrid nanopores for potential applications, several factors must be considered: the SiN_x chip, chip thickness, SSN diameter, the internal shape of the SSN, the portal protein itself, the buffer, and the voltage. The aspects of the portal protein that can be optimised for use as a hybrid nanopore are tunnel diameter, tunnel charge, overall dipole. The SiN_x chip should be clean of any contaminants that may interact with the portal protein or analytes. Contaminants could pass through the pore during analysis. They may also alter the shape of the pore and affect pore wetting. The SSN must not give excessive noise and must be hydrophilic. The SSN must also be of a suitable size and shape. The diameter of the SSN should be wide enough for the portal protein to enter easily, but not so wide that the portal protein may pass through entirely. The diameter will also affect the EOF going through the pore. The thickness of the SSN chip will affect the abilities of the hybrid pore when analytes pass through (see Chapter 1, Electrokinetics). The EOF through longer pores will have a lower velocity than those through shorter pores. The thickness of the pore will also affect the topology of the SSN tunnel and how easily it can be fabricated by CDB. The shape should be symmetrical.

5.1.9. Expected events

The formation of a hybrid pore is expected to result in a sudden decrease in current (Cressiot et al. 2018). However, in addition to hybrid pore formation, bumping events are also expected in which the

portal protein approaches the SSN or partially blocks it, then leaves again. Bumping events are characterised by their short dwell times and significant current blockades. They typically last less than 200 µs. Biological nanopores produce less noise than SSN, so successful hybrids should be characterised by a significant reduction in current noise as well (Cressiot et al. 2018; Fragasso, Schmid, and Dekker 2020).

5.1.10. Analytes for testing the hybrid nanopores

The utility of the GBSV1 portal protein as a biological or hybrid nanopore depends on its ability to distinguish between analytes of interest. To assess this, once a hybrid nanopore is formed, current measurements are performed with different analytes. The field of hybrid nanopores is still relatively new and there are challenges to the use of hybrid nanopores. Some previous hybrid pores showed significant current leakage (Engst et al. 2012; R. Wei et al. 2012), and/or were dependent on the voltage polarity to remain in their SSN support (Engst et al. 2012; Cressiot et al. 2018). Hybrid nanopores have been tested with DNA (Hall et al. 2010a; Engst et al. 2012; Cabello-Aguilar et al. 2013; Göpfrich et al. 2013; Cressiot et al. 2018; Bentin, Balme, and Picaud 2020; Sen, Hoi, and Gupta 2021; Mojtabavi et al. 2022) and peptides (Gornall et al. 2011; Cressiot et al. 2018). However, they have yet to make significant advancements in the frontier areas of the nanopore field (see Chapter 1), such as sequencing and the identification of post-translational modifications.

5.1.11. Chapter aims

This chapter presents the results of the optimisation of the conditions for the insertion of the GBSV1 portal protein for use as a nanopore.

5.2. Materials and methods

5.2.1. Biological pores

The critical micelle concentration (CMC) of 16:0 MPB PE (maleimide-DPPE) is unknown. In order to estimate it, the CMCs of similar lipids were compared. However, the CMC of lipids can vary widely. The CMC for PEG_{5000}-DPPE (also called 16:0 PEG_{5000} PE) has been calculated to be 28.1 µM (Wichit et al. 2012) and 70 µM (Sou et al. 2000) using different methods. The CMC for β-CD-DPPE is 7.17 µM (M. Yan et al. 2019). If the concentration of MPB PE exceeded its unknown CMC, it would form micelles and be unable to attach to the cysteine in GBSV1_Δ1-40_S135C_792. Therefore, the CMC of 16:0 MPB PE was taken to be 7.17 µM. MPE PB was dissolved in DMSO to 700 µM. It was further diluted in DMSO to 140 µM, then to 70 µM in 50 mM Tris-HCl pH 8, 0.25 M NaCl, 2mM DTT. Meanwhile, a fresh aliquot of GBSV1_Δ1-40_S135C_792 was taken from the -80 ℃ freezer and thawed on ice. 5 mM DTT was added and allowed to incubate for 30 minutes to reduce all of the cysteine residues. 30 µL of GBSV1_Δ1-40_S135C_792 were added to a Zebra column (ThermoFisher) to remove excess DTT. GBSV1_Δ1-40_S135C_792 was diluted in 10 mM Tris-HCl, 0.4 M KCl, pH 7.5 buffer to 7 µM. The 70 µM MPB PE and 7 µM GBSV1_Δ1-40_S135C_792 were each diluted 1/10 in 50 mM Tris-HCl pH8, 0.25 M, 2 mM DTT and left overnight to incubate at room temperature. 2-mercaptoethanol (BME) was added to a final concentration of 10 mM to remove any unbound MPB PE and incubated for 10 minutes. The mixture was filtered through a fresh Zebra column (ThermoFisher) to remove DTT, BME, and MPB PE. Cuvettes were painted with decane to make lipid membranes using methods described previously (Gutsmann et al. 2015).

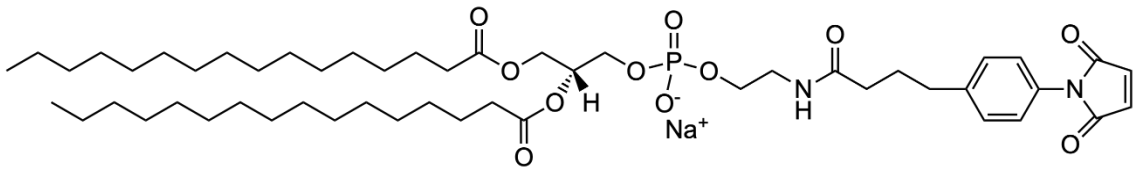


Figure 5-3: Structure of 16:0 MPB PE

5.2.2. Hybrid pores

5.2.2.1. Chip handling

Experiments used either 12 or 30 nm thick chips. The 12 nm thick chips were bought from Norcada (NXDB-50B405V122). The two wafers of 30 nm thick chips were made by Dr Benjamin Cressiot when at Northeastern University. All SiN_x chips were handled gently with tweezers holding them at the edge, being careful not to touch the middle where the nanopore would be introduced. Tweezers were wiped with 20% EtOH before use. If the chips had to be placed down, they were placed with the depression on the underside in order to not break the membrane.

5.2.2.2. Membrane check

To check if the membrane chips had burst membranes, the chips were placed into Spark-E2 flow cells and the reservoirs were filled with test buffer. The flow cells were connected to an Axopatch 200B (Molecular Devices, USA). An IV curve between ± 100 mV was collected. If the IV curve showed no current across the membrane, it was intact; if significant current was detected, then the membrane had already broken and the chip was discarded (Fujinami Tanimoto et al. 2022).

5.2.2.3. Piranha cleaning

Before using the silicon chips to make nanopores, they were first cleaned using Piranha solution to remove organic materials which might contaminate later measurements. The nanopores were first dried on absorbent paper, hole side down, for 30 minutes to remove excess water. Two pairs of nitrile gloves were worn to protect against accidental spillages and cover the wrists. Inside a fume hood, the silicon chips were each placed into a separate glass test tube, hole side down. The tubes were held in a glass beaker on a hot plate. The tubes were pointed towards the back of the fume hood so that any explosion would be away from the user. A glass Pasteur pipette was used to add three pipettes of sulphuric acid to each test tube. A fresh pipette was used to add one pipette of hydrogen peroxide to each tube. Pipettes were disposed of into a designated glass waste beaker. The bubbling piranha solution was heated to 200 °C for 20 minutes, always under observation. Afterwards, the heat was turned off, and the solution allowed to cool completely for 20 minutes. Piranha was disposed of into a bottle inside the fume hood specially designated for piranha waste. The lid was never fully tightened to prevent a build-up of pressure. The test tubes, with the chips inside were washed thoroughly and carefully with fresh MQ water six times. Freshly cleaned silicon chips were stored upright in 500 µL Eppendorf tubes in freshly degassed milli-Q water until use.

5.2.2.4. Nanopore fabrication

Silicon chips were loaded into the Spark-E2 as described in the user manual ('Northern Nanopore Instruments User Manual' 2022), except that freshly degassed milli-Q water was washed through the flow cell between the isopropanol and electrolyte solutions to prevent crystallisation of the salt inside the flow cell channels.

All solutions were freshly filtered and degassed to remove any dissolved air bubbles. When 10 nm chips were used, the two halves of the flow cell could be screwed together until flush; however, 30 nm chips were too thick for the two halves to be screwed together without breaking one of the rubber gaskets. Therefore, the flow cell was screwed together until finger tight in order to create a seal. The Nanopore Fabrication software made by Northern Nanopore was used for setting the parameters for

nanopore creation. A conductivity meter was not available, so the conductivity of the solutions was measured by running a small amount of each buffer through an AKTA Pure. The conductivity for the fabrication and conditioning solutions was set at 11.19 S/m and 16.84 S/m. The size of the SSN was calculated using equation (15) by the Spark-E2.

5.2.3. Nanopore data collection

5.2.3.1. Background readings

The portal protein stock was taken from the -80 ℃ freezer on ice. It was diluted to 100 µM working stock in test buffer. The working stock was kept in the 4 ℃ fridge when not in use. The SiN_x chip containing a solid-state nanopore was set into a Spark-E2 flow cell between two silicon gaskets with the depression facing the *trans* side. The flow cell was placed into a home-made faraday cage. The *cis* side of the flow cell was connected to the ground electrode unless otherwise stated. The current was monitored using an Axopatch 200B amplifier. The voltage was set using ClampX10 software. Once the flow cell was connected, the background readings were taken. These were IV curves at ± 100, ± 200 and ± 500 mV, and continuous recordings at 20 mV steps from 0 to 200 mV. On-the-fly current measurements at different voltages for the voltage steps during baseline measurements were taken using ClampFit. The conductance at each voltage could be estimated which gave an indication of how stable the solid-state pore was: if the conductance changed over the course of the recording, the SSN may not have been well formed. Because of the low salt content of the test buffers, the size could not be accurately estimated using the conductance formula. Later background readings included recording the current at voltages between 0 and ± 200 mV in 20 mV steps in conditioning buffer, which allowed a manual estimate of the pore size. The current and noise was measured at each voltage. The current was plotted against voltage and the noise as error bars to make an IV curve.

Unless stated otherwise, the *cis* compartment of the flow cell was attached to the ground electrode, the *trans* compartment was attached to the working electrode (Figure 5-1). At positive voltages, the current, cations and EOF move from the *trans* compartment to the *cis* compartment, while the anions move from the *cis* compartment towards the *trans* compartment.

To measure the symmetricity of the SSN (see Chapter 1), the rectification ratio (RR) was found using the current readings at each voltage. The RR is a ratio of the current at a negative voltage divided by the current at the same positive voltage (equation (44)). If the positive and negative current is the same at positive and negative voltages, then RR = 1, and the pore is symmetrical. If the positive current is higher than the negative current at the opposite voltage, then the SSN is wider at the *cis* side and

the RR < 1. If the negative current is higher than the positive current at the same voltage in the opposite polarity, then the SSN is wider at the trans side and RR > 1.

$$Rectification\ ratio\ (RR) = \frac{Current\ at\ negative\ voltage\ (nA)}{Current\ at\ positive\ voltage\ (nA)} \quad (44)$$

5.2.3.2. Hybrid nanopore formation

Unless otherwise stated, 0.2 µL of 1 mM GBSV1 portal protein was added to the *cis* compartment of the flow cell for a final concentration of 10 µM. For buffer optimisation, the buffers tested are listed in Table 5-1. Unless otherwise stated, the buffer used was 10 mM Tris-HCl, 0.4 M KCl pH 6.

Table 5-1: List of buffers tested

Buffer	Abbreviation
10 mM Tris-HCl, 0.4 M KCl pH 7.5	pH 7.5
10 mM Tris-HCl, 20 mM $MgCl_2$ pH 7.5	$MgCl_2$
10 mM Tris-HCl, 0.4 M KCl pH 6.5	pH 6.5

A range of SSNs were made. Each was given a number to identify it. When the pores are referred to in this work, the nomenclature is poreX_Ynm, where X is the pore's ID number, and Y is the diameter of the pore. Some pores were used multiple times, on different days, or after treatments, which may have changed their sizes. If this was the case, the size was measured again and a new name would be given, e.g. pore4_5nm and pore4_9nm.

5.2.3.3. Analytes

To test a positively charged analyte, 0.4 µL of 1 mM bradykinin in 25 mM Tris-HCl pH 7.5 was added to the trans compartment for a final concentration of 10 µM. Under positive voltage, it should move through the nanopore from the working electrode to the ground; thus, it should move through the hybrid pore from the clip domain to the crown domain.

In order to test a negative analyte, 0.8 µL of 250 µM dsDNA hairpin (5'-GCTGTCTGTTGCTCTCTCGCAACAGACAGC T_{50}-3') (IDT) in 10 mM Tris-HCl pH 8, was added to the *cis* compartment (ground electrode). Under a positive voltage, it should move towards the *trans* compartment (working electrode) under electrophoresis. If a hybrid nanopore has formed, it would move from the crown domain to the clip domain.

5.2.4. Data analysis

Files made by the Axopatch were exported into .abf files using ClampX. These were opened in IgorPro v9.05 32-bit using the Igor Extension ABFXOP.xop and saved as IgorPro experiment files to be opened and analysed in IgorPro v9.05 64-bit. Collections were organised into datasets based on date,

individual chip, nanopore diameter and buffer. Consecutive recordings which used the same voltage were exported as Igor binary files and joined together using the Mass Concatenation macro developed by Gabriel Gibrat (Université d'Évry). Recordings were also joined together if a potential hybrid nanopore persisted across recordings. To assess the stability of the solid-state nanopore, the conductance was measured for each pore. This also allowed comparisons across different voltages. If the solid-state nanopore was stable, the conductance should be approximately constant throughout the recording.

The average current and standard deviation (noise, σ) were measured across the recording to find the baseline current (I_0). The average conductance of the baseline was also determined to compare to baseline values. Events were identified if they crossed below five standard deviations of the baseline current. The start and end points for each peak were taken as the points when the peak crossed three standard deviations of the average (Oukhaled et al. 2012). Analysis macros written by Professor Laurent Bacri were used to identify the peaks. The baseline conductance (G_0), blockade conductance (G_b), difference in conductance (DG_b), normalised change in conductance ($DG_{b\text{-}norm}$), minimum and maximum blockade conductance were calculated using a modified CalG() macro written by Dr Sandra Greive. The normalised change in conductance was found by equation(45).

$$Normalised\ change\ in\ conductance, DG_{b-norm} = \frac{G_0 - G_b}{G_0} = \frac{DG_b}{G_0} \qquad (45)$$

The greater the blockade, the closer $DG_{b\text{-}norm}$ gets to 1.

In order to pick out current drops that likely corresponded to hybrid events, a set of rules were written (Figure 5-4). This was to remove ambiguity when picking potential hybrid formation events so that traces could be analysed consistently.

The DEEP ONES rules

1. Visual inspection of traces when imported into Igor32.
2. If no current drops are apparent, mark it as nothing and move on.
3. Make traces of the current and conductance.
4. Use Dr Bacri's macros to identify the baseline current (I_0), standard deviation (σ) and events (when the current drops below I_0 - 5σ). Measure the start and end times of events from I_0 - 3σ.
5. Use the CalG() macro to calculate the conductance for the identified events.
6. Draw a graph of normalised change in conductance ($DG_{b\text{-}norm}$) against dwell time (T_t).
7. Isolate the events with a $DG_{b\text{-}norm}$ higher or equal to 0.5.
8. Extract the parameters for each event into individual columns. Parameters: current, voltage, frequency, SSN diameter, concentration, etc.
9. Save the result waves as a deliminated text file. Use a command so that the same waves are saved in the same order every time

Figure 5-4: The DEEP ONES workflow

The chosen events were exported from IgorPro as deliminated text files. The columns in the text files were saved in the same order each time so that they could be concatenated together with others later for comparisons. Those that were collected under the same conditions (same protein mutant and same buffer) were copied into a new directory so that fair comparisons could be made. For the voltage comparisons, those of the same voltage were joined together using an IgorPro script to concatenate them together. Box plots were drawn on the data points. The box plots plotted the minimum, maximum, median, the lower quartile, upper quartile and interquartile range (IQR). The data between the minimum and the lower quartile is the lower 25% of the data. The data between the lower and upper quartile is the IQR. The data between the upper quartile and the maximum is the highest 25% of the data.

5.3. Results

5.3.1. Biological pores

One of the planned uses of the GBSV1 portal protein was to attach a lipid anchor and insert it into a lipid bilayer to form a biological nanopore, building on similar work reported earlier for the G20c portal protein (Cressiot et al. 2017b). When GBSV1_Δ1-40_S135C_792 was added to cuvettes painted with lipid to form a membrane, no change in current was observed, suggesting that it did not insert into the membrane. Several attempts were made, but no change was seen. Due to time constraints, no further tests were made.

5.3.2. Hybrid pores

5.3.2.1. Expected current blockades by hybrid nanopore formation

Before any SSNs are formed, some estimates for the proportional current blockade can be made based on the dimensions of the portal protein derived from the cryoEM structure (see Chapter 3). This is made complicated by the different orientations in which the portal protein may approach the SSN, the fact that the values for the SSN diameter are estimates based on conductance (see Chapter 1), and the unknown shape of the SSN tunnel.

The portal protein may enter the SSN with the clip domain first, side-on, or crown domain first. The portal protein dimension measured along the central axis is ~ 10 nm, i.e. distance from the clip to the crown domain (see Chapter 3). If it enters side-on, then current cannot pass through the pore unless the SSN is wide enough to let current leak between it and the portal protein. To estimate the occlusion ratio of the portal protein blocking the SSN, the programme US-SOMO was used as part of UltraScan3 to find the radius of gyration of the portal protein. The radius of gyration is 4.856 nm (Tiesinga et al. 1996; American Physical Society and American Institute of Physics 2001; Rocco and Byron 2015). Thus, the average diameter of the portal protein is 9.712 nm. Table 5-2 shows the estimated expected proportional blockade for SSN of different diameters.

However, using the R_g for calculating the expected blockades makes certain assumptions. This assumes that the portal protein is a rigid object, does not move within the SSN and that it does not have a pore. The SSN may have an hourglass shape in which the mouth is wide enough for the portal protein to enter sideways on, but the constriction point is narrow enough for it not to translocate, then the portal protein may tumble within the SSN. However, these estimates may also be used to estimate current leakage around the portal protein inside the SSN. If a hybrid pore is formed in a 10 nm SSN, then ~3% of the baseline current may be expected to pass between the outside of the portal and the SSN tunnel. The rest of the current will pass through the portal protein tunnel.

Table 5-2: Approximate proportional blockades by GBSV1_Δ1-40_775 in different sized SSN

SSN diameter (nm)	Proportional blockade	Proportion let through	SSN diameter (nm)	Proportional blockade	Proportion let through
4	> 1	0	**21**	0.46	0.54
5	> 1	0	**22**	0.44	0.56
6	> 1	0	**23**	0.42	0.58
7	> 1	0	**24**	0.40	0.60
8	> 1	0	**25**	0.39	0.61
9	> 1	0	**26**	0.37	0.63
10	0.97	0.03	**27**	0.36	0.64
11	0.88	0.12	**28**	0.35	0.65
12	0.81	0.19	**29**	0.33	0.67
13	0.75	0.25	**30**	0.32	0.68
14	0.69	0.31	**31**	0.31	0.69
15	0.65	0.35	**32**	0.30	0.70
16	0.61	0.39	**33**	0.29	0.71
17	0.57	0.43	**34**	0.29	0.71
18	0.54	0.46	**35**	0.28	0.72
19	0.51	0.49	**36**	0.27	0.73
20	0.49	0.51	**37**	0.26	0.74

These values for an object with a diameter of 9.712 nm (twice the radius of gyration of GBSV1_Δ1-40_775 as calculated by US-SUMO in UltraScanIII). The blockade values are rough guides only.

On the other hand, if the portal protein enters the SSN with the clip or crown domains first, then it will form a hybrid nanopore with a ~2-4 nm diameter, depending on if the portal protein is in the open or closed conformation. If the clip domain enters first and the SSN is wider than the clip domain (~ 8.4 nm) (see Chapter 3), then the portal protein should insert snugly into the nanopore. This is assuming that there is no leakage current between the portal protein and the SSN. However, the diameter of the G20c portal protein at the clip domain is ~ 8.5 nm and was able to insert into an SSN with a diameter of 5.4-6 nm at the narrowest point (Cressiot et al. 2018). If the SSN is narrower than this, the portal protein should not be able to insert into the SSN. Instead it would rest on top of the mouth of the pore. Based on these dimensions, it may be estimated that the best diameter for the SSN would be somewhere between 5 and 13 nm. The tunnel through the portal protein and current leakage between the portal protein and the SSN mean that these values are likely an overestimate.

5.3.2.2. Optimisation of hybrid nanopore formation

(a) Solid-state nanopore preparation

Once an SSN has been fabricated and conditioned, it must be assessed to see if it is suitable for making a hybrid pore. A good SSN should be symmetrical (see Chapter1, Current Rectification), give low noise and not show gating behaviour. The quality of an SSN can be assessed using baseline measurements of the current at different voltages, the standard deviation and the corresponding conductance values. Figure 5-5 shows a representative example of pore35. The diameter was 5.15 nm. The current is stable at different voltage values, both positive and negative. There are some background current spikes, at some voltages, but hybrid formation should produce a much deeper current drop, so these small spikes do not stop the SSN being used as part of a hybrid. The conductance was measured at positive and negative voltage. The conductance should be approximately the same across both positive and negative voltages (Figure 5-5C-D).

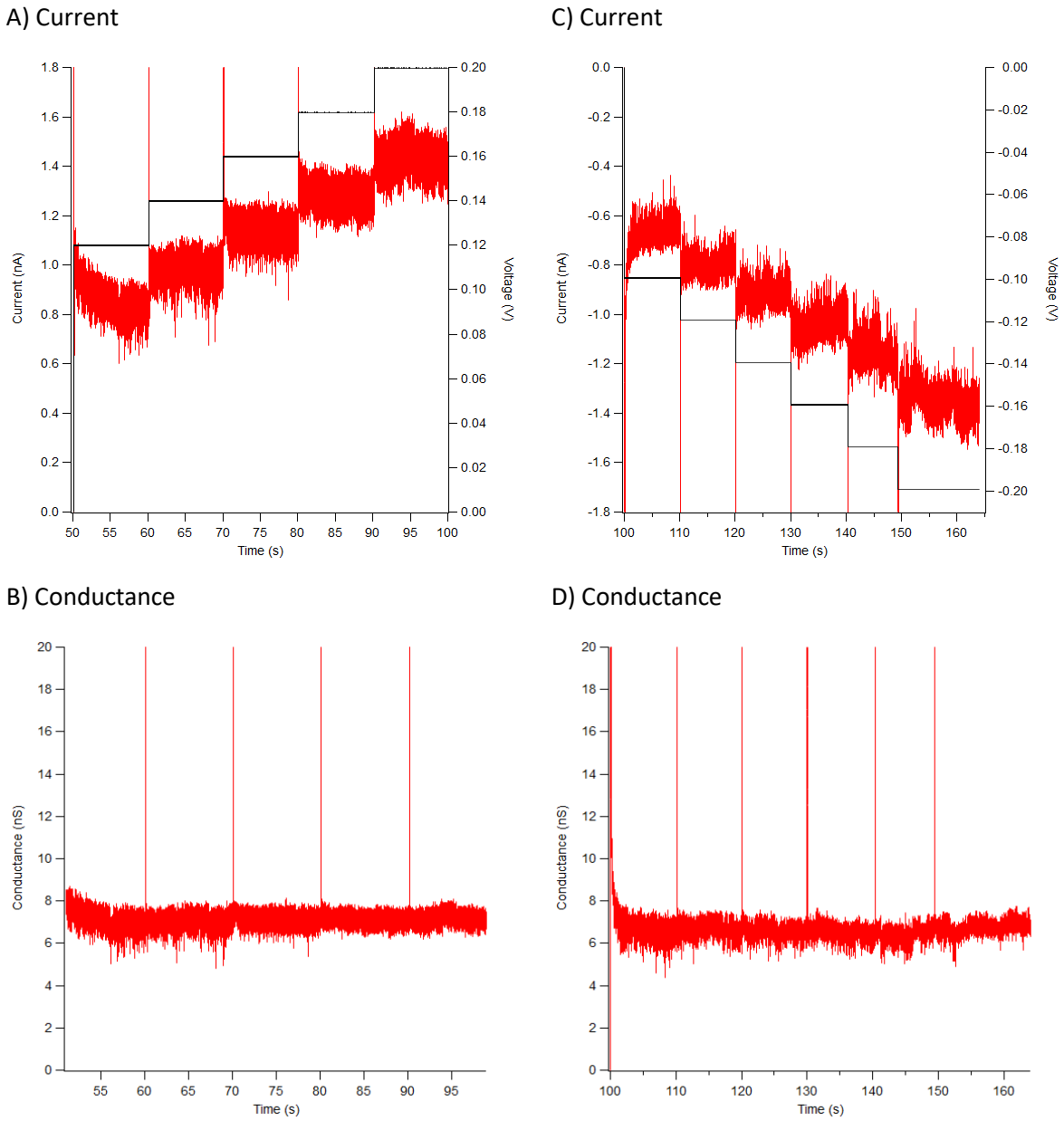


Figure 5-5: Baseline current and conductance for pore 35_5.15nm at ±100-200 mV

Baseline readings take in test buffer before the addition of portal protein. A) The baseline with positive voltage from 100 to 200 mV in 20 mV steps and the corresponding conductance trace. Red – current trace, black – voltage trace. B) The same traces as in (A) but for -100mV to -200mV. The conductance traces show if the SSN is stable at the voltages tested. Comparing the average conductance for positive and negative voltages shows if the pore is rectified. In this case, the pore is symmetrical and stable.

On the other hand, a badly formed SSN cannot be used for useful hybrid nanopores. Once identified, such an SSN should be discarded, but some attempts can be made to remedy its performance. The baseline readings for pore53_10nm are shown in Figure 5-6 and Figure 5-7 as a representative example of a bad pore. This pore was cleaned with piranha solution the day before fabrication rather than the day of its fabrication, which may well have caused its strange behaviour. When fabrication was run on the Spark-E2, a 1.1 nm nanopore was detected immediately. Conditioning was performed to enlarge the pore, but no change was observed, so this was labelled as a soft breakdown event (see

Nanopore fabrication). Fabrication was repeated and a 1.3 nm pore was detected, which was noted as another soft breakdown event, so fabrication was run a third time ('Northern Nanopore Instruments User Manual' 2022). A 7 nm pore was fabricated, far larger than a newly fabricated pore should be. This was conditioned to 8 nm. Before experiments started, the size was checked again and it was found to be 10 nm. The two instances of soft breakdown are signs that these particular chips may not have been suitable for SSN fabrication.

Figure 5-6 shows the plots of current-voltage curves (IV curves). The IV curves at ± 200 and ± 100 mV show different behaviours of the SSN. The current is not symmetric at positive and negative voltages and is very noisy. The current also displays clear hysteresis, as the reading at the same voltage is different depending on whether the voltage is increasing or decreasing. The IV curve for a perfect nanopore should be a straight diagonal line. Since current is higher when the voltage is ramping up, this is capacitive hysteresis (Bisquert 2024).

The pore also displays gating behaviour at higher voltages. (Figure 5-6A-B). The exact cause of gating is unknown.

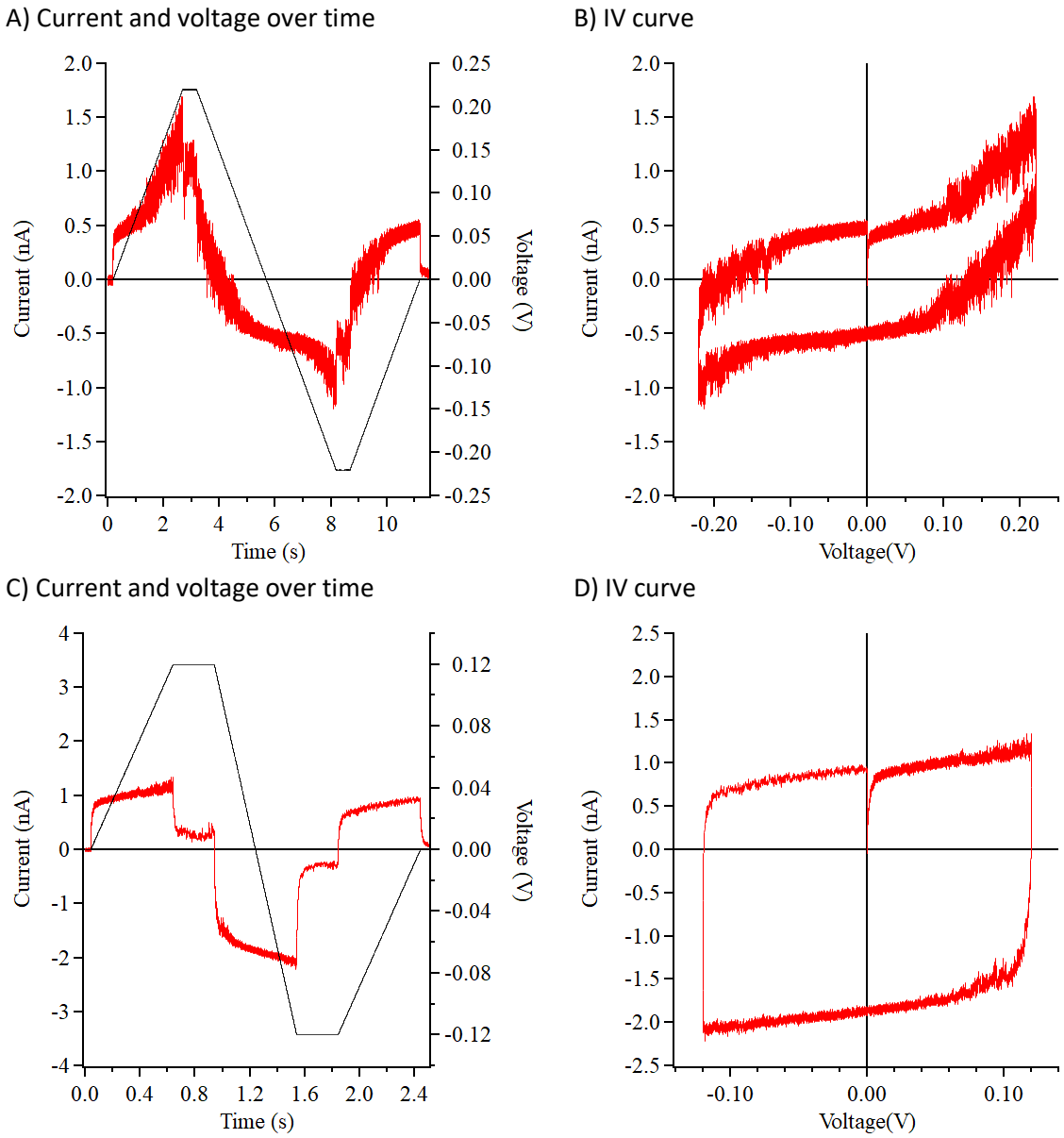


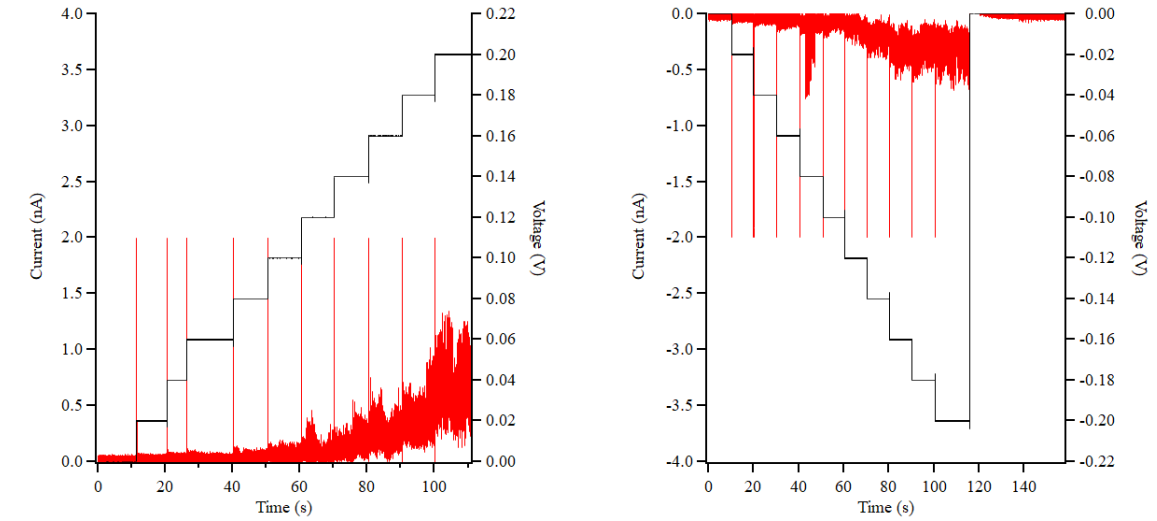
Figure 5-6: Current-voltage (IV) curves for pore53_10nm

A-B) IV curves for ± 200 mV. C-D) IV curves for ± 100 mV. A, C) Current-voltage overlays against time. Three repeats were taken for each voltage range (i-iii). Plot (iv) shows the three traces concatenated together. B, D) Current-voltage plots. The ±100 mV graphs were taken immediately after the ±200 mV traces. Each of the individual current traces (A, C i-iii) was plotted against the voltage ramp. The IV curves show clear hysteresis and rectification. The IV curves for ±200 mV (B) also show strange gating and noise at higher voltages.

The current readings at different voltages and the conductance readings are shown in Figure 5-7. The Spark-E2 found and conditioned a 10 nm SSN. An SiN_x chip with no pore would give 0 nA current readings at all voltages, but the current does increase slightly at higher voltages, so an SSN of some form is present. The small and slow increase in current suggests that the pore is much smaller than the measurement suggested. The asymmetric values of the conductance at positive and negative voltages suggest that the pore has an asymmetric shape.

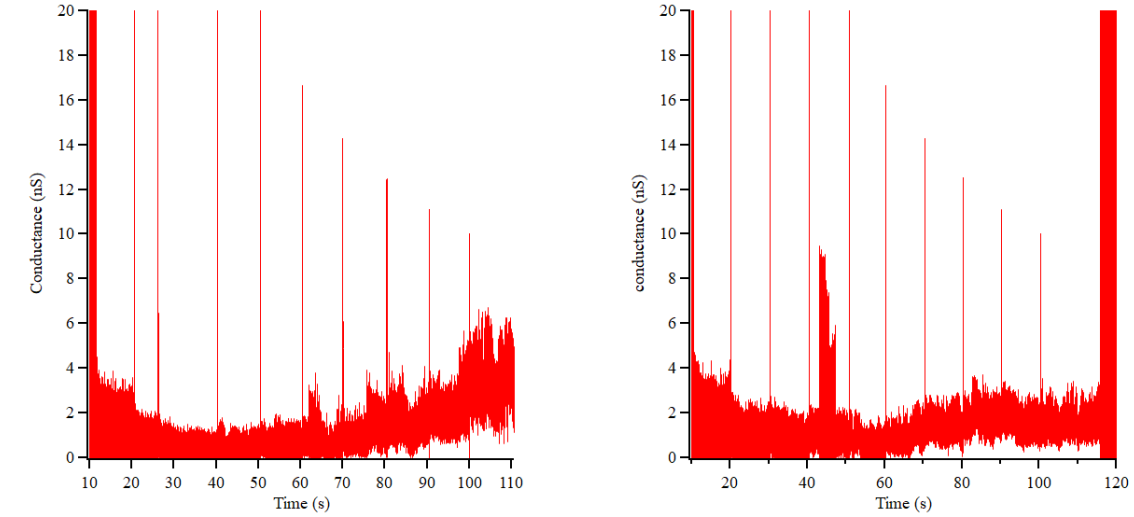
A) Current trace. 0 to 200mV

B) Current trace 0 to -200mV



C) Conductance 0 to 200mV

D) Conductance 0 to -200 mV



E) IV curve

F) Sigma vs voltage

G) Rectification ratio

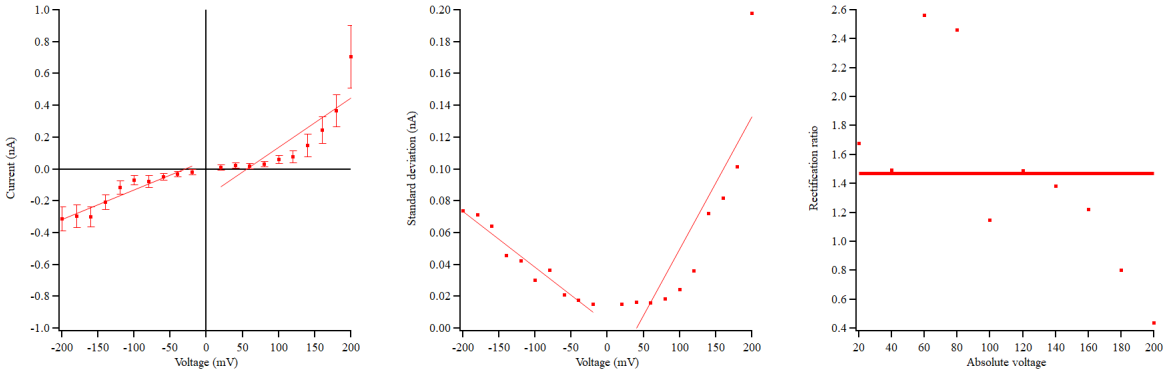


Figure 5-7: Baseline readings for pore53_10nm.

A-B) Positive voltage baseline readings. C-D) Negative voltage baseline readings. A, C) Plot of the current and voltage over time. B, D) Conductance plot corresponding to (A) and (C).

(i) Methods for fixing an unstable SSN

Not every SSN fabricated will be perfect, or even good, as soon as it has been made. As seen in Figure 5-6 and Figure 5-7, an SSN can be asymmetrically shaped, apparently blocked, or prohibitively noisy. The cause of these behaviours will inform measures taken to fix them, if they can be fixed. Figure 5-8 and Figure 5-9 show how the behaviour of a pore can be changed from asymmetric and noisy to a usable pore, even on different days. The pore is pore34_4.7nm. When it was first made, it was made at 4.7 nm diameter. It was cleaned with piranha solution before fabrication. The pore displayed asymmetric behaviour at positive and negative voltages. The first attempt to fix these was to wash test buffer through the flow cell to dislodge any nanobubbles that may have formed on the SSN surface. The baselines were taken again but still showed asymmetric behaviour (Figure 5-8A). At positive voltages, the conductance slowly increased. At negative voltages, the conductance dropped off at higher voltages. The conductance was also not the same at the same positive and negative voltages. The flow cell was washed through with MQ, then EtOH, then MQ and buffer again. This was in case the SSN surface was not sufficiently wetted and a nanobubble had formed. The baseline readings after are similar to those before but the voltage-dependent gating is less pronounced (Figure 5-8B). Re-wetting the SSN surface appears to have helped somewhat.

The nanopore's voltage-dependent gating behaviour is seen more clearly in the Figure 5-7A,E. The manual for the Spark-E2 recommends letting such chips sit in conditioning buffer or enlarging the pore slightly ('Northern Nanopore Instruments User Manual' 2022). However, this solution may be problematic if the SSN is already at the desired size. As an alternative, the pore was let sit in conditioning buffer in the flow cell (Figure 5-8C). This appears to have solved both the asymmetric current at positive and negative voltages, and the reduction in current at higher voltages. However, there are current spikes at positive voltages. They appear to reach a similar current blockade level, even at different voltages. A contaminant passing through the pore would block the same amount of current each time. Additionally, when the pore size was measured again in the Spark-E2, it was given as 6.17 nm. This is supported by the higher conductance values. Even without applying a voltage, conditioning buffer will still cause the pore to grow slowly.

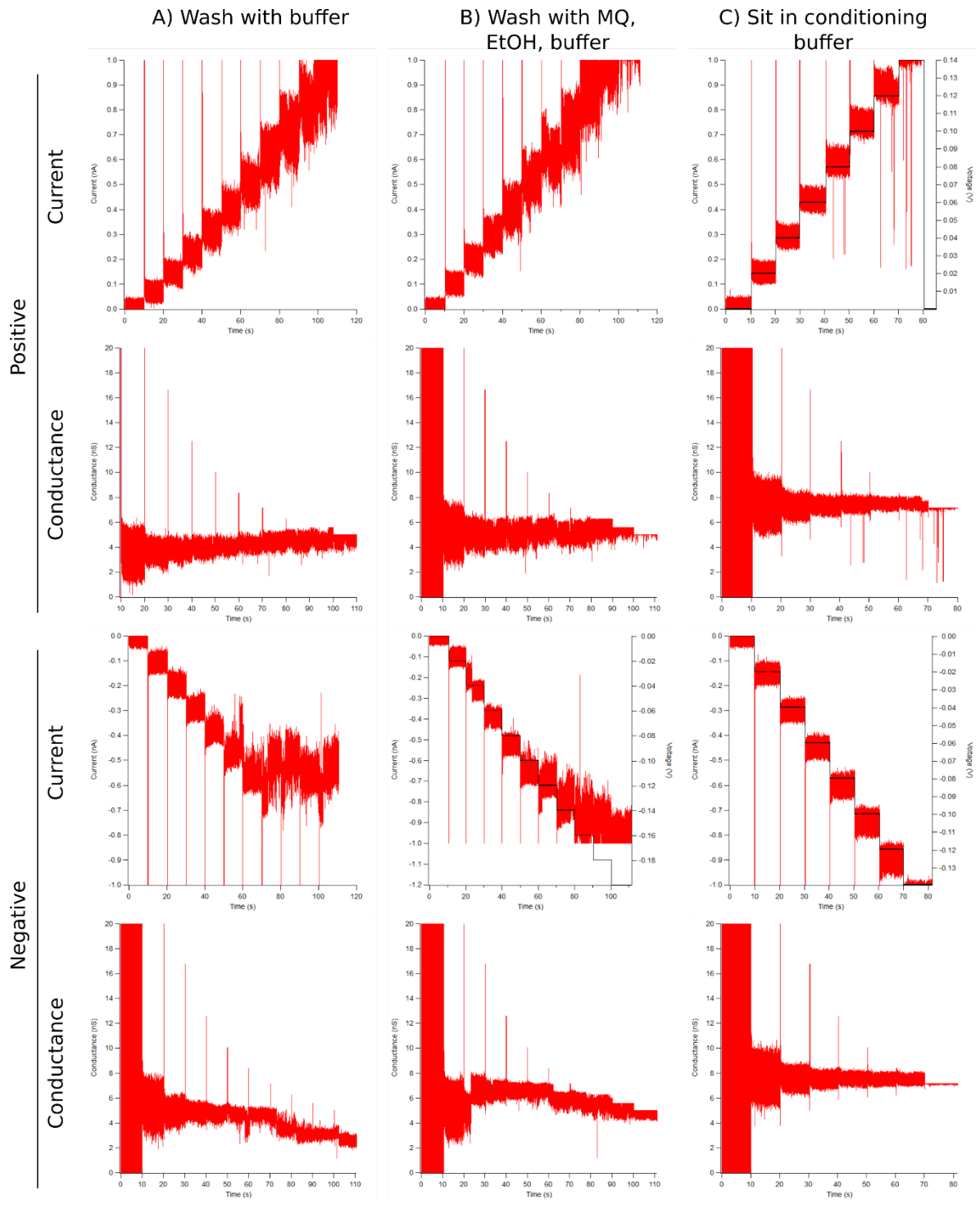


Figure 5-8: Lifetime and recovery of a stable pore

The behaviour of an SSN pore34_4.7nm after treatment with different methods to improve stability. A-C) Measurements on the same day as the pore was made. A-B) Washing test buffer (A) or MQ, 20% EtOH and test buffer (B) only made a small difference to the SSN's performance. Washing EtOH may have wetted the surface of the pore. C) The pore was let sit in conditioning buffer in the flow cell. The conductance stabilised significantly. But there are current spikes at higher positive voltages.

Figure 5-9 shows the same pore on the next day. The chip was stored in test buffer overnight in the flow cell. The size was measured with the Spark-E2 and found to be 4.02 nm. It is not possible for SSN to shrink. Instead, nanobubbles may have formed on the surface. Conditioning buffer was washed

through the flow cell (Figure 5-9A). The pore shows a significant current drop at -180 mV to a metastable baseline. The current spikes at positive and negative voltage may be the pore transiently switching to this metastable baseline. The pore was allowed to sit in conditioning buffer and the baseline readings were taken again (Figure 5-9B). The outcome is similar to Figure 5-8, in that the pore shows stable current readings and about the same conductance values at positive and negative voltages.

However, there is a limit to what conditioning buffer can fix. Figure 5-10 shows the baseline readings taken from a pore76_8nm after being left in 50% EtOH for two days. When it was made, the SSN had an 8 nm diameter. The size was measured again on day 3 and found to be 11.4 nm (pore76_11.4nm). The baseline readings also show a pore with severe asymmetry (Figure 5-10, first column). The current does not increase in proportion to positive voltage, so the conductance goes down with voltage. Meanwhile at negative voltages, the conductance jumps up at higher voltages. As before, the pore was left in conditioning buffer (Figure 5-10, 2nd column). The baseline readings afterwards show much improvement, but the conductance is still unstable and noisy. The size was measured again on the Spark-E2 at 9 nm (pore76_9nm).

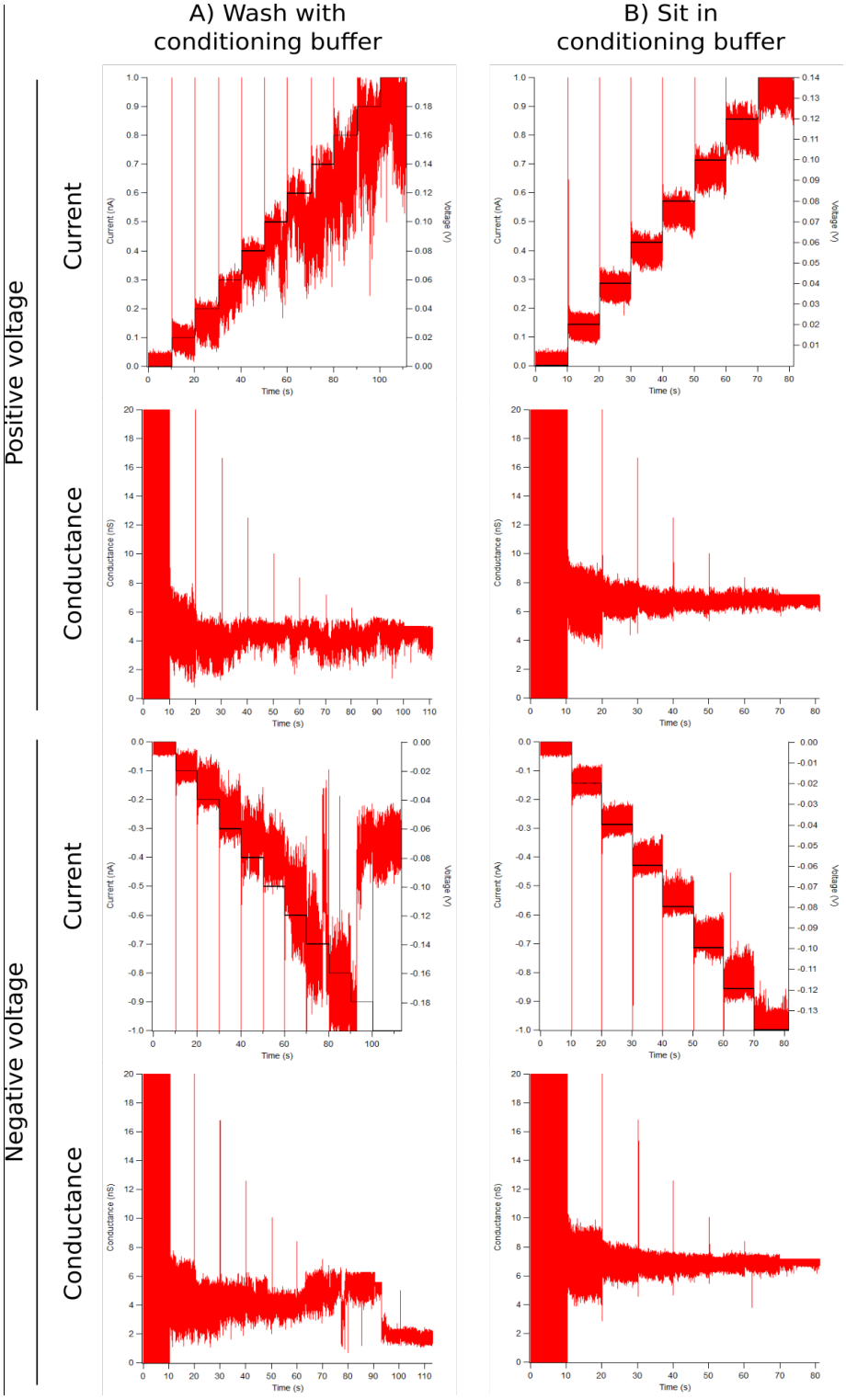


Figure 5-9: Recovery of pore34_4.7nm with conditioning buffer

The same pore as in Figure 5-8 (Pore34_4.7nm) was tested again the next day. A) The flow cell was washed through with conditioning buffer. The baseline is clearly unstable. The diameter appeared to have grown to 5.49 nm. B) The pore was let sit in conditioning buffer as before . Even after being left for a day, conditioning recovered the pore's performance. The diameter has grown again to 5.9 nm

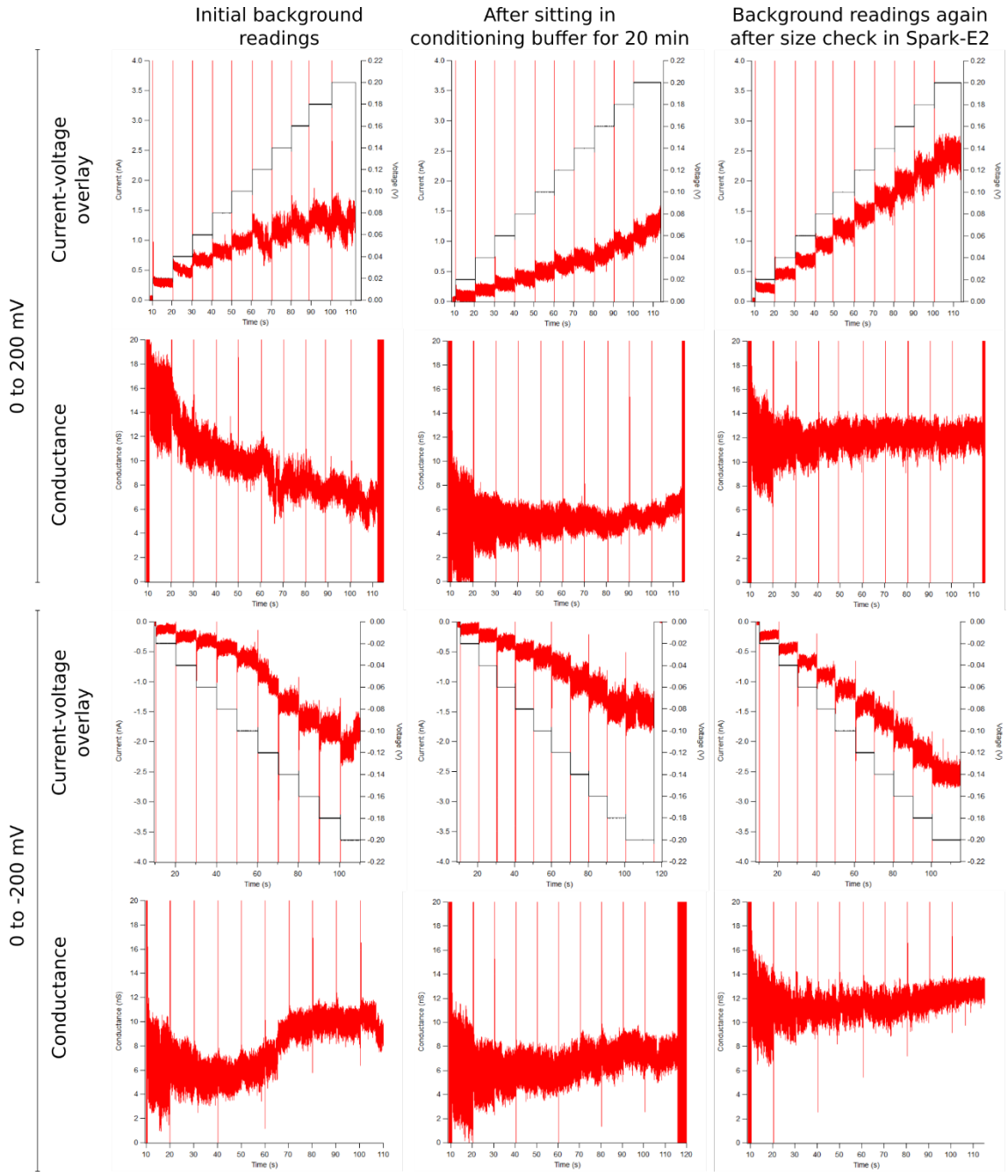


Figure 5-10: Effects of conditioning buffer on noise of pore76_8nm

Left column – pore76_11.4nm after two days in 50% EtOH/H2O. Middle column – pore76_11.4nm after sitting in conditioning buffer for 20 minutes. Right column – pore76_9nm after checking the size in the Spark-E2.

(b) Optimal solid-state nanopore length

(i) 30 nm chips

After piranha cleaning, many of the 30 nm chips were found to already be broken: the window in the chip was already at >100 nm, making them unusable for fabrication. To find out if the chips were already broken, or if the damage occurred due to the piranha cleaning, an IV curve between -100 and 100 mV was run before piranha cleaning (Fujinami Tanimoto et al. 2022) (see Materials and methods).

These showed that most of the chips in the first wafer were already broken, while the chips in the 2nd wafer were mostly intact.

However, this test could not detect chips which gained defects during their creation, which were too thick for current to pass through or for fabrication. To screen for these, a 'light test' was used (5.2. Materials and methods). The 30 nm chips were too thick for the two halves of the Spark-E2 flow cell to be screwed flush together. They were assembled until finger tight, which left a small gap between the two halves. However, even with these screening tests, SSNs could not be fabricated in 30 nm chips. The fabrication step ran for hours without change.

1. Effect of buffer pH on fabrication

Kwok *et al* found that the pH of the fabrication buffer had a significant effect on the fabrication time of 30 nm thick SSN during CDB, with extreme pH values giving shorter times (Kwok, Briggs, and Tabard-Cossa 2014). A range of buffers at different pH values were tested, but an SSN still could not be fabricated. The buffers tested were:

- 29.4 mM citrate phosphate, 41.2 mM sodium phosphate pH 4
- 1 M KCl, 10 mM sodium phosphate pH 6
- 1 M KCl, 10 mM Tris-HC, pH 9
- 50 mM glycine, 32 mM NaOH pH 10

2. Overnight fabrication

It was possible that SSN would form in the 30 nm membranes given enough time. Therefore, the fabrication protocol was run overnight in pH 4 buffer. But a pore was detected quickly, and instead the conditioning protocol was run overnight. However, when the chip was checked the next day, no pore was detected, which would suggest that the initial pore had been an instance of soft breakdown (see Introduction). Due to time limitations, no further tests were carried out on 30 nm thick membrane chips. The rest of the experiments used 12 nm thick chips (Norcada).

(c) Optimal buffer conditions

The SSN diameter of 5-6 nm was used as a starting point for SSNs as it has been used for G20c hybrid pores (Cressiot et al. 2018). As mentioned the buffer conditions will affect the SSN, the portal protein, and any potential analytes. The parameters to optimise are pH, salt type and concentration. The pH of the buffer will change the charge of the portal proteins and the surface of the SSN tunnel. At pH 6.5 and pH 7.5, the portal protein has a charge of -93.9 and -110.6, respectively (Brookes and Rocco 2018). Since the overall charge decreases between pH 7.5 to pH 6.5, it may be reasonable to assume that the ζ-potential is also lower at the lower pH. If the ζ-potential of the portal protein is lower, then its

electrophoretic velocity should also be lower (see Chapter 1). The electric field lines around the GBSV1_Δ1-40_775 protein appear similar at pH 6.5 and 7.5 (Figure 5-12).

The dipole for the GBSV1 portal protein was calculated in different buffers (Figure 5-11) using the PQR files calculated from the PDBtoPQR webserver (Unni et al. 2011) and a Chimera script (see Appendix). The dipole vectors point from the centre of the protein to the calculated negative end of the dipole. For both the pH 6.5 and pH 7.4 buffers, the dipole is parallel with the tunnel axis with the negative end at the clip domains. When an electric field is applied, the portal protein should have a propensity to orientate itself with the clip domain pointing towards the positive electrode. In the context of hybrid nanopore formation, if the portal protein is added to the *cis* compartment which has the ground electrode, then when a positive voltage is applied, it should move towards the SSN clip domain first.

A) pH 6.5 B) pH 7.5

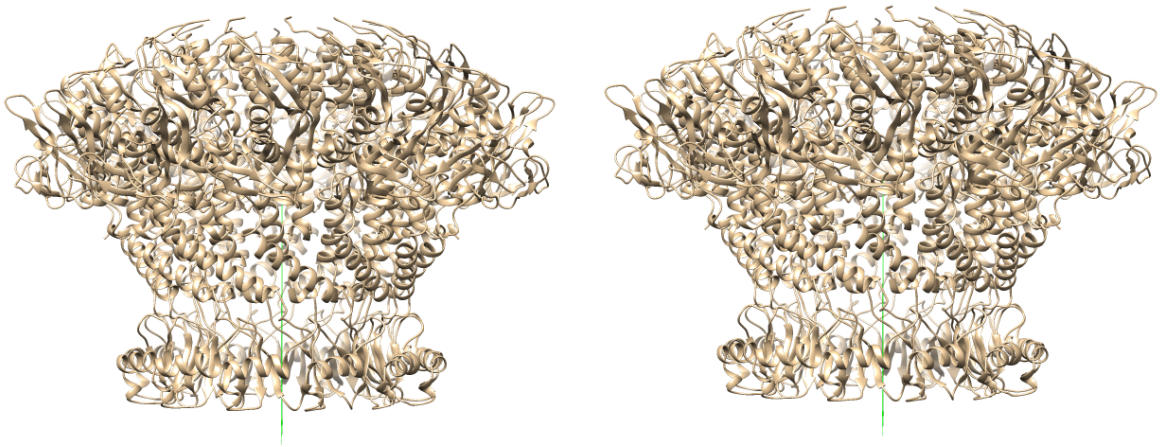


Figure 5-11: Dipole of GBSV1 portal protein

The green line shows the dipole from the middle of the portal protein. The dipole was found in different buffers. A) pH 6.5 B) pH 7.5.

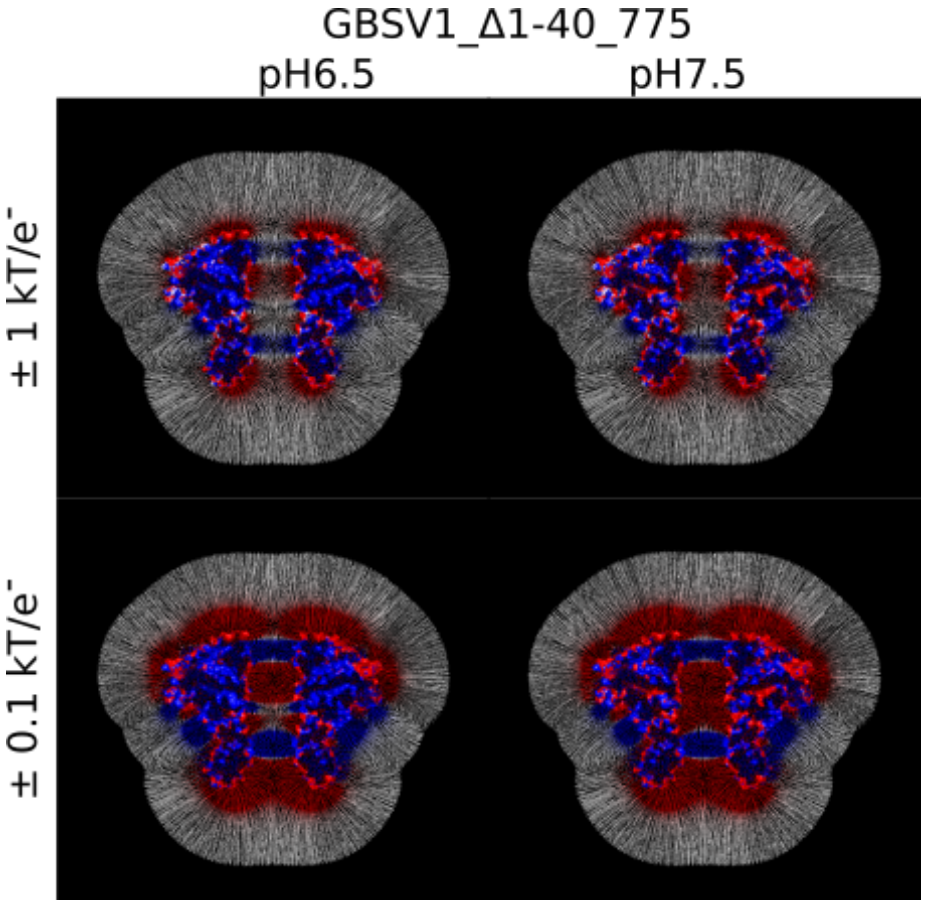


Figure 5-12: Electrostatic potential fields of GBSV1_Δ1-40_775 at different pH values.

The buffer is 10 mM Tris-HCl, 0.4 M KCl pH 6.5.

The buffers tested are listed in Table 5-1. A buffer with 20 mM $MgCl_2$ was tested because of its apparent ability to stabilise the portal protein and prevent aggregation (see Chapter 4). Divalent cations such as Mg^{2+} have a much greater effect on lowering the surface potential of objects in electrolyte solutions (Israelachvili 2011). The pore used for the $MgCl_2$ buffer was pore46_5nm (Figure 5-13A-B). The pore used for the pH 6.5 buffer was pore48_4.8nm (Figure 5-13C-D).

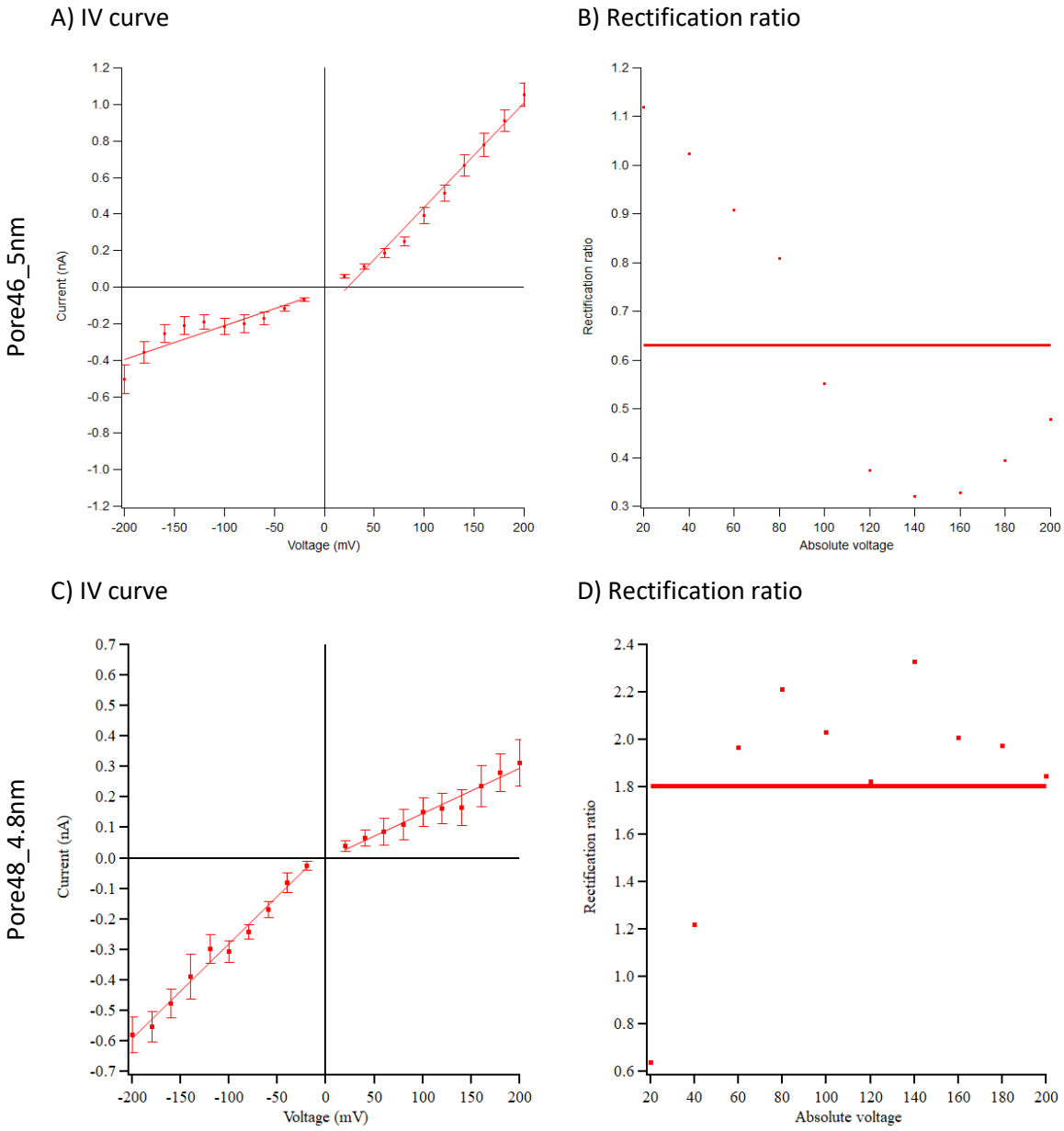


Figure 5-13: IV and RR for pore46_5nm and pore48_4.8nm

A-B) Pore46_5nm with 10 mM Tris-HCl pH 7.5, 20 mM MgCl2. C-D) Pore48_4.8nm with 10 mM Tris-HCl pH 6.5, 0.4 M KCl pH 6. A,C) IV curves made from measurements of the current between -200 and +200 mV. The error bars are the standard deviation of the current at that voltage. B, D) RR at each voltage.

The pores that used 10 mM Tris-HCl pH 6.5, 0.4 M KCl pH 7 were pore46_4.8nm, pore45_5nm, pore35_1.15nm (Figure 5-14).

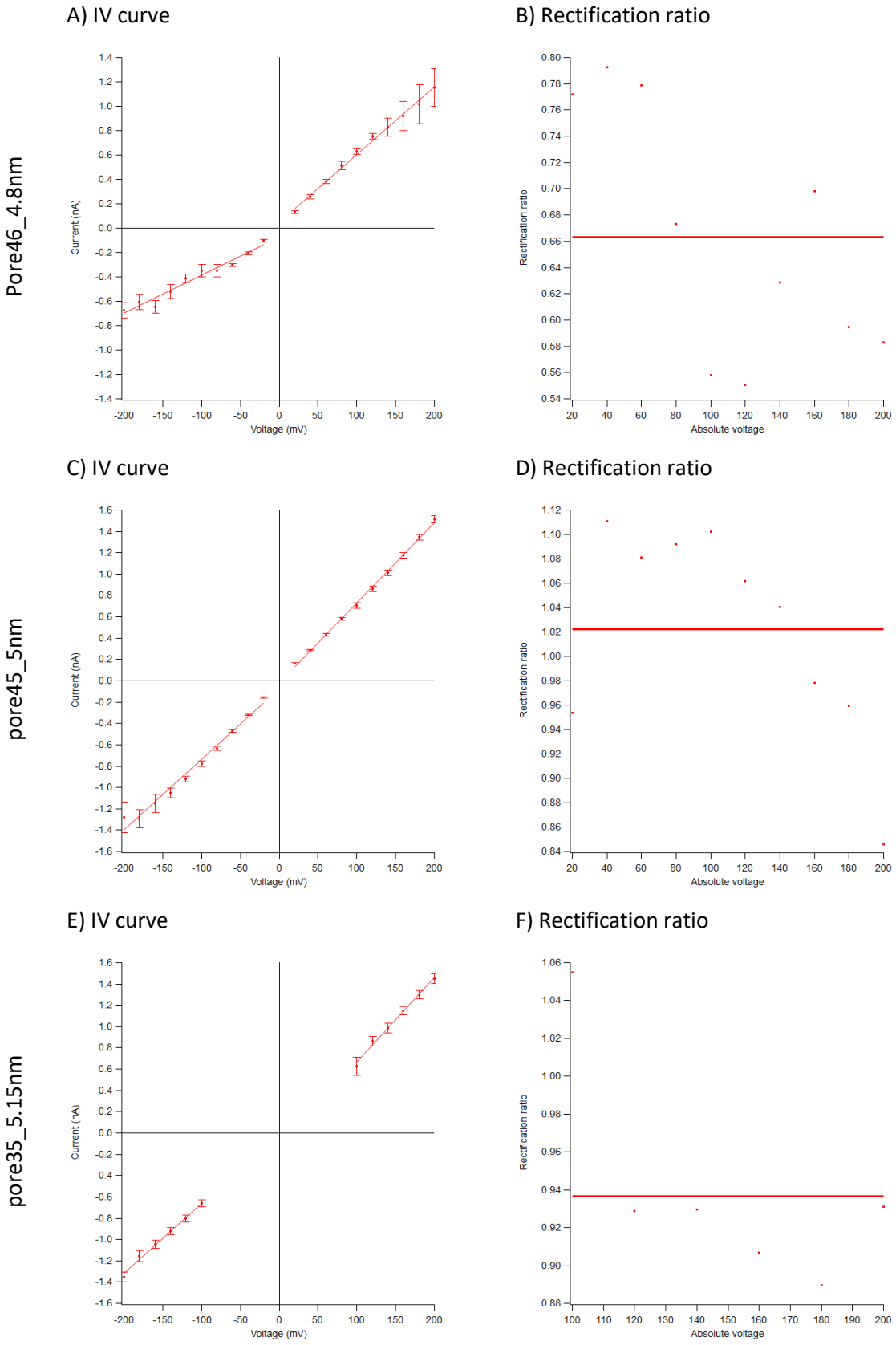


Figure 5-14: IV and RR for pore46_4.8nm, pore45_5nm, and pore35_5.15nm in pH 7.5 buffer

A-B) Pore46_4.8nm. C-D) Pore45_5nm. E-F) Pore35_5.15nm. A,C,E) IV curve. B,D,F) rectification ratio at each voltage.

Some basic parameters were compared between the different pores (Figure 5-15). The pores were used to assess hybrid pore formation. A range of different voltages were used for each pore. Therefore, conductance was used to compare them. Only background readings were collected for pore46_4.8nm in buffer at pH 7.5, so it was only used in the comparisons of baseline conductance between buffers.

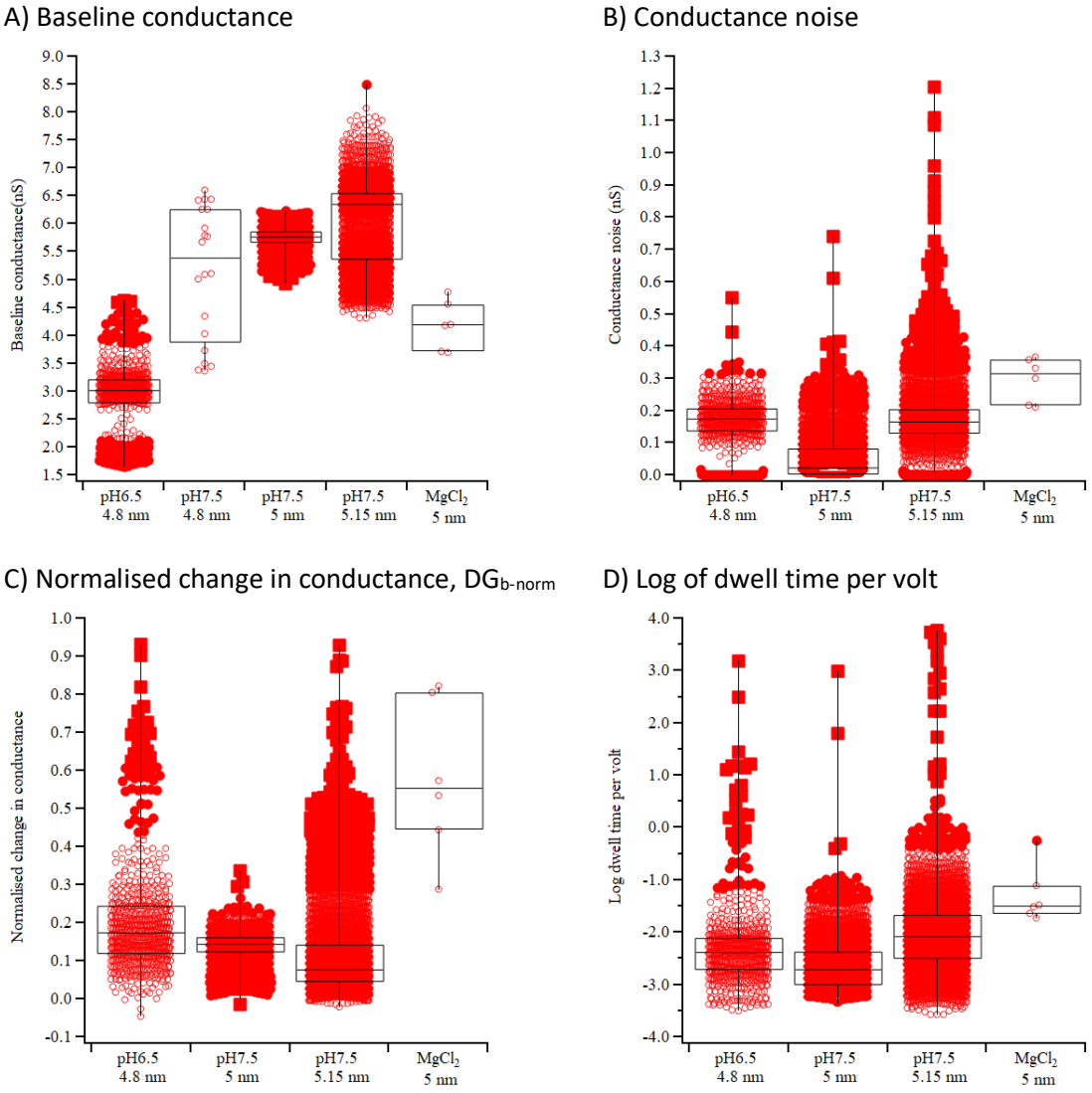
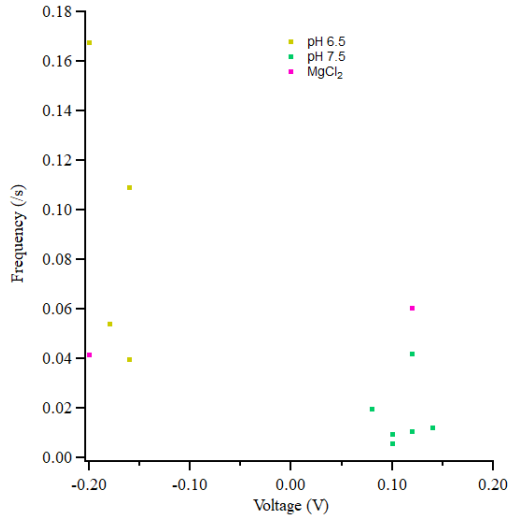


Figure 5-15: Box plots of ~ 5 nm SSN with buffers of different pH and salt

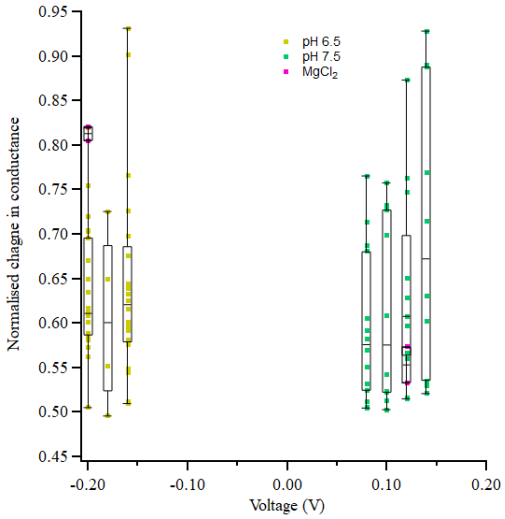
The pore used for the pH 6.5 buffer was pore48_4.8nm. The pores used for the pH 7.5 buffer were pore45_5nm, pore35_5.15nm. The pore used for the MgCl2 buffer was pore46_5nm. Basic parameters at different pores and buffers. A) Baseline conductance. B) Conductance noise. C) Normalised change in conductance. D) Log of the dwell time per volt.

However, these results included all events that crossed the threshold of 5σ of the baseline current, which included bumping events. Therefore, events were selected according to the DEEP ONES rule set (see Hybrid nanopore formation). Nanopores at a range of diameters were tested for the pH 6.5 and pH 7.5 buffers. For fair comparisons, those with a diameter of ~5 nm were compared. The results are plotted in Figure 5-16.

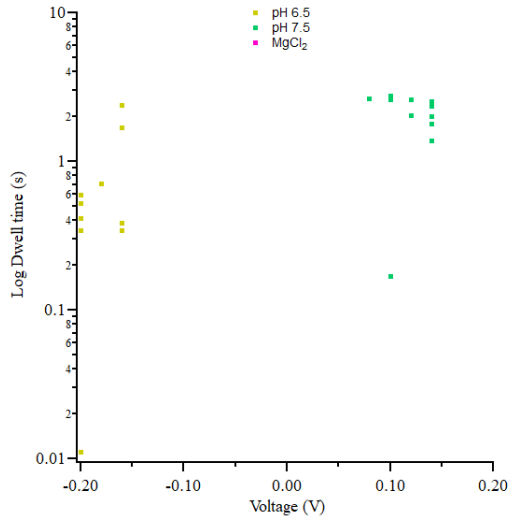
A) Frequency of hybrid events



B) Normalised conductance change



C) Log dwell time



D) G_0 Vs G_b

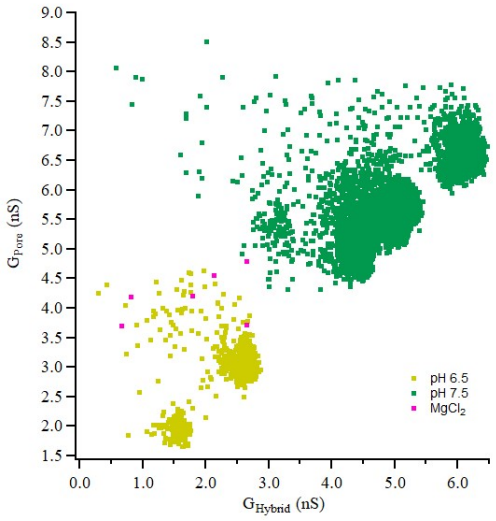


Figure 5-16: Hybrid events at different buffer conditions

The pore for the pH 6.5 buffer was pore48_4.8nm. The pores for the pH 7.5 buffer were pore45_5nm, pore35_5.15nm. The pore for the mgcl2 buffer was pore46_5nm. A) Frequency. B) Normalised change in conductance. C) Log dwell time. D) SSN conductance (Gpore) Vs hybrid conductance (Ghybrid).

There was not a clear difference between the different buffers. Due to low quality pores and the small number of data points, no firm conclusions can be drawn on the effect of the buffer on efficiency of hybrid formation from these tests.

(d) Optimal SSN diameter

The diameter of the SSN is of paramount importance to successful hybrid nanopore formation. If the SSN diameter is too small, the portal protein cannot insert and will bounce off, causing only bumping events. If the SSN diameter is too large, the portal protein may insert in different orientations, move around within the SSN, or translocate straight through. The shape of the SSN lumen should be hourglass shaped, with a constriction point halfway through. However, in CDB, the shape is influenced

by the pH of the buffer (Kwok, Briggs, and Tabard-Cossa 2014) and conditioning once the pore has been formed. Based on the shape of the portal protein (see Chapter 3), it may be expected that SSN with a diameter of ~ 9-13 nm and their hourglass topology would only allow the portal protein to stably sit in the SSN if the clip domain entered first. If the portal protein approached with the crown domain first, it should not insert as firmly and would leave. However, this requires the shape and size to be accurate.

(i) 36 nm diameter SSN

To establish an upper limit for nanopore size, some nanopores were made with diameters much greater than the portal protein, allowing it to translocate straight through. The goal was to see the GBSV1_Δ1-40_Δ383-406_816 protein translocate through the SSN.

Figure 5-17 shows baseline and raw traces for a 36 nm diameter pore. This pore was so wide that the portal protein should translocate through without touching the SSN walls. The current appears stable at each voltage tested (Figure 5-17 A), with no spikes or fluctuations. The conductance is fairly stable (Figure 5-17 B) at ~ 80 nS, though it does rise slightly at higher voltages. The IV curve was perfectly straight. The pore was tested at different voltages to look for portal protein translocations. Figure 5-17D-E show conductance traces at 60 and 80 mV respectively. Each shows frequent deep events, but also shows a drop in the baseline current from ~ 80 nS to ~ 70 nS. The baseline drops in Figure 5-17D, then returns to ~ 80 nS in Figure 5-17E. This metastable baseline is a common issue with solid state nanopores. However, the metastable baseline also shows current events. The blockade depth is similar to those seen at the original baseline (see Figure 5-17D-E). This metastable baseline is seen in other current traces for the same pore at different voltages (data not shown). The events from the two baselines are compared in Figure 5-18.

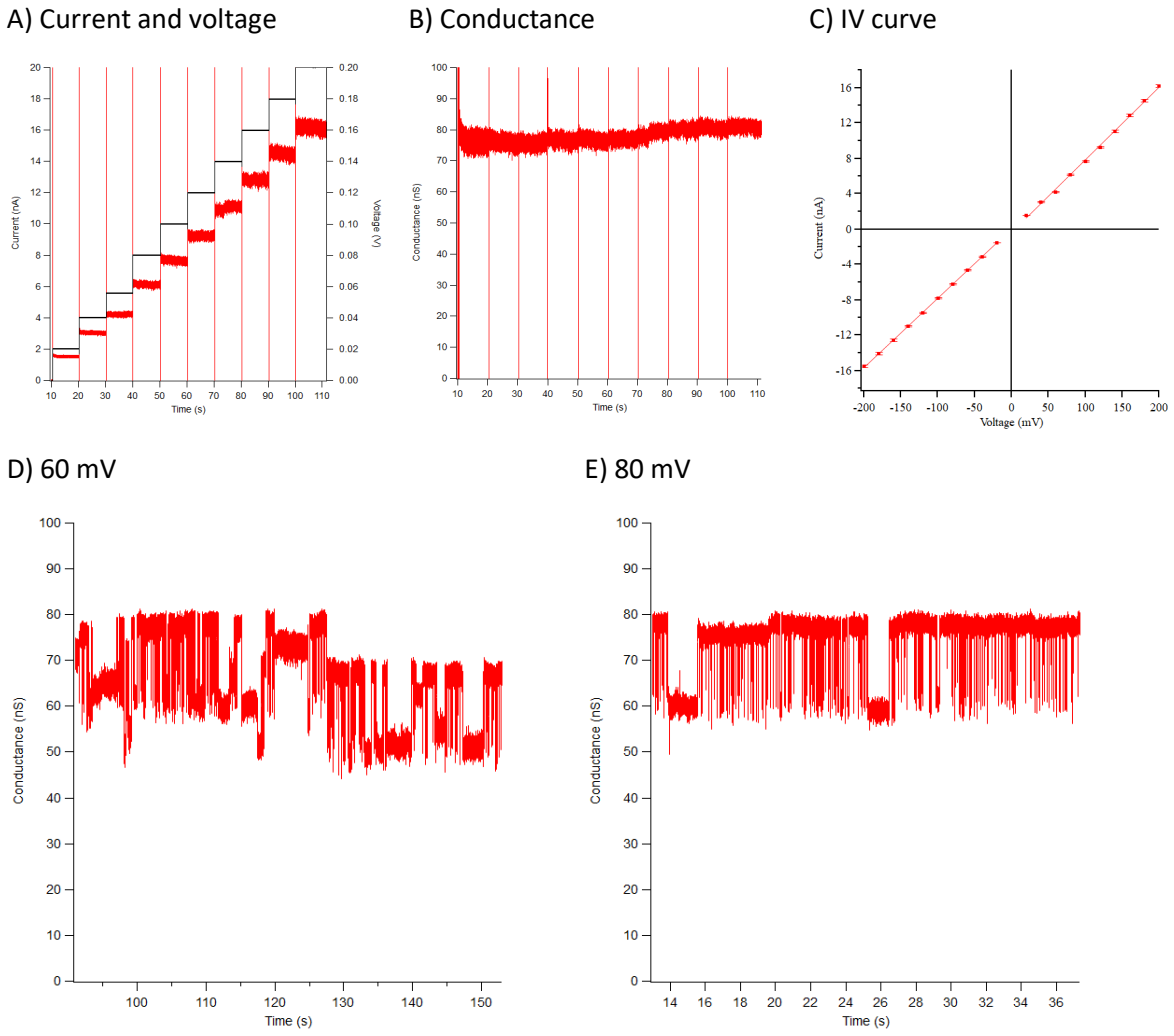
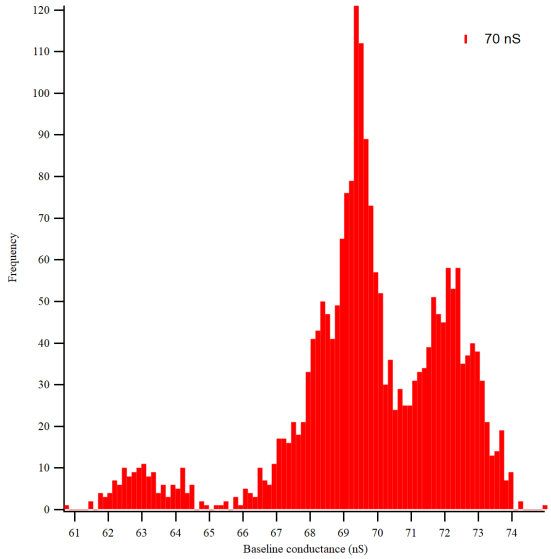


Figure 5-17: Metastable baselines in pore60_36 nm

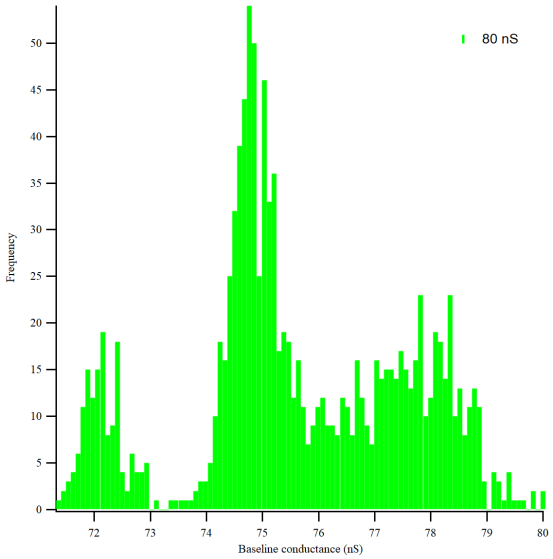
A-B) Baseline readings before the addition of GBSV1_Δ1-40_Δ383-406_816. A) Current (red) at different voltages (black) from 20 to 200 mV. B) Conductance of (A). The conductance is fairly stable throughout, but increases slightly at higher voltages. C) IV curve from the current measurements in (A). D-E) Conductance traces of events after the addition of GBSV1_Δ1-40_Δ383-406_816. Both clearly show a 2nd baseline at ~70 nS. The events at the original and 2nd baseline appear to have the same depth. The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. C) 60 mV. D) 80 mV.

The sections of the traces with the different baselines were analysed separately. Histograms for the baseline conductance for the two different baselines are shown in Figure 5-18. Both histograms have the same approximate shape with a smaller peak at lower conductance and the tallest peak in the middle. The lowest baseline conductance peak of the higher baseline (called 80 nS) overlaps with the highest baseline conductance peak of the lower baseline (Figure 5-18C). It may be possible that each of the five baseline conductance peaks is a metastable baseline. But the reason why a certain baseline becomes dominant in an area of the trace is not yet clear.

A) G_0 at 70nS



B) G_0 at 80nS



C) Overlay histograms

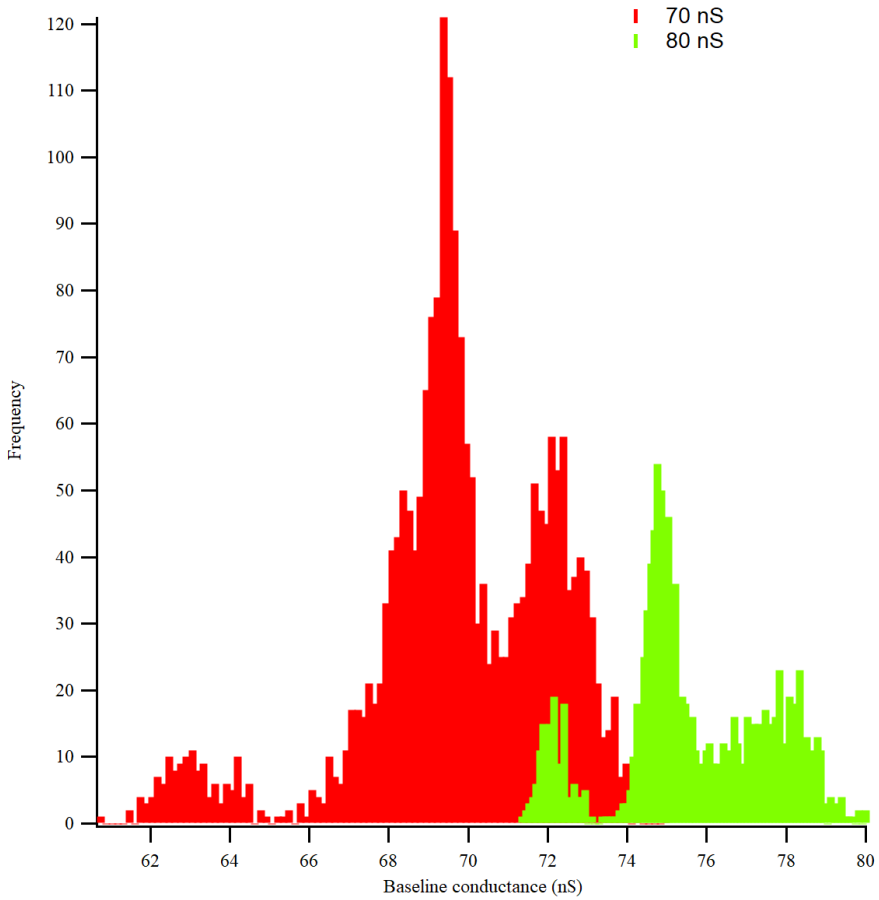


Figure 5-18: Comparison of baseline conductance for pore60_36nm

A) Baseline conductance for the trace with a lower baseline. B) baseline conductance for the area of the trace with a higher baseline. C) Overlay

Plots of the $DG_{b\text{-}norm}$ vs dwell time show approximately three clusters of events: a cluster of $DG_{b\text{-}norm}$ events, a cluster of low dwell time events, and a group of higher $DG_{b\text{-}norm}$ events (Figure 5-19). The pattern is persistent across voltages.

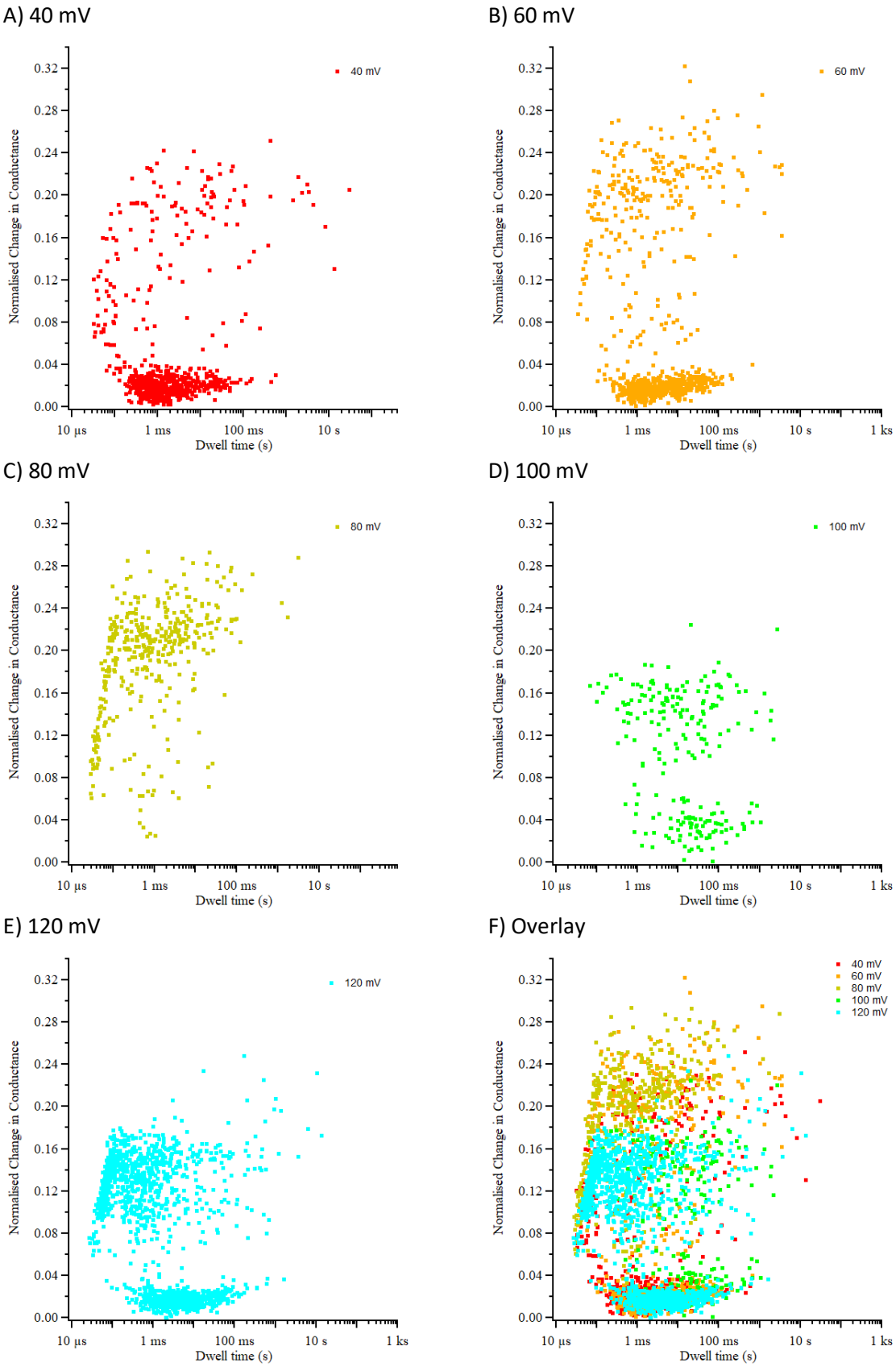


Figure 5-19: Normalised change in conductance vs dwell time for 36 nm pore

The events are taken from both the 70 and 80 nS baseline traces. A) 40 mV. B) 60 mV. C) 80 mV. D) 100 mV. E) 120 mV. F) Overlay of all voltages.

Figure 5-20 shows the $DG_{b\text{-}norm}$ and dwell time Vs voltage. The fractional blockade of the 36nm pore by the portal protein is 0.27. Since the protein has a pore, the fractional blockade will be lower. If the portal protein is translocating through the SSN, the dwell time should decrease as voltage increases.

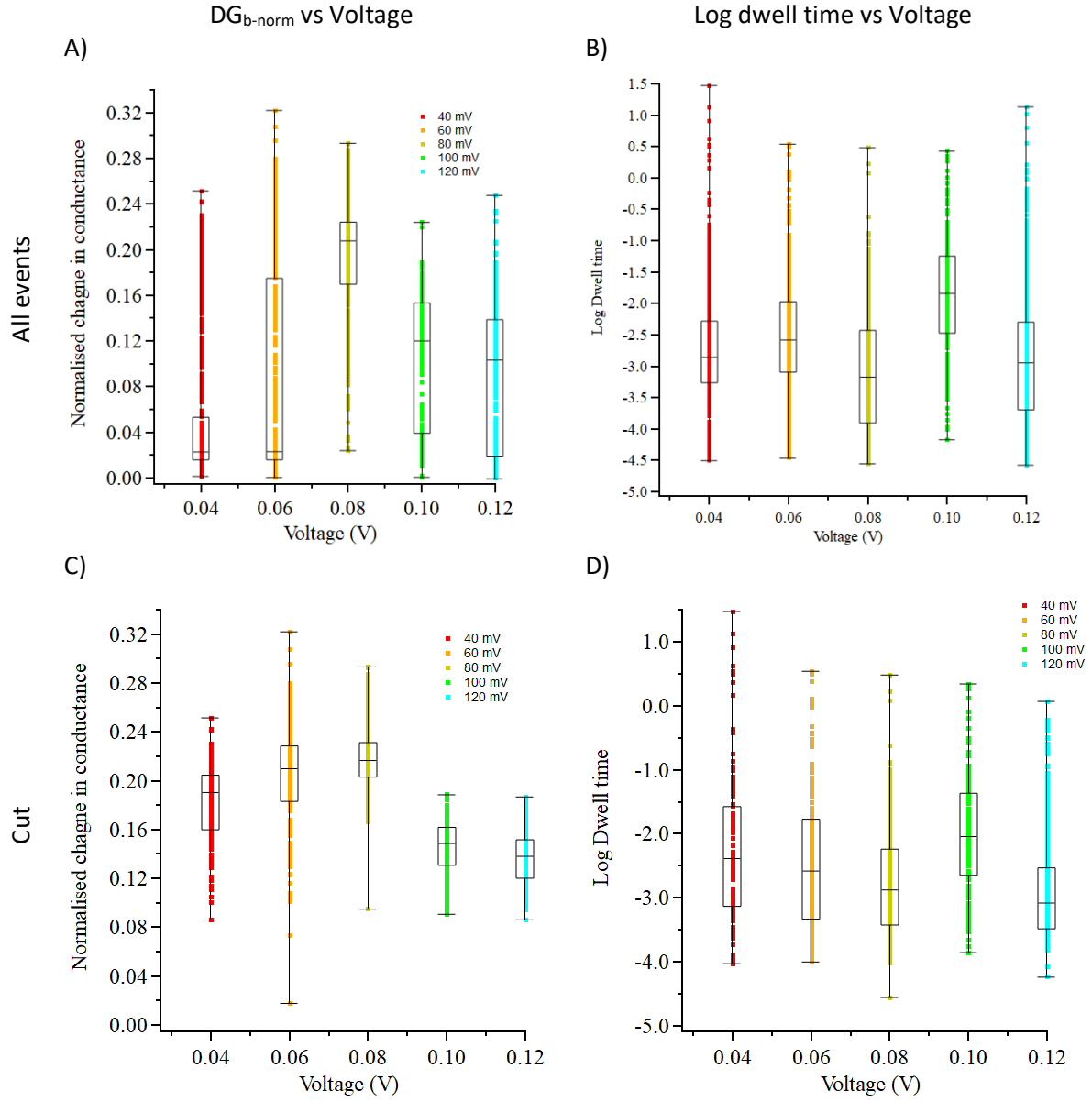
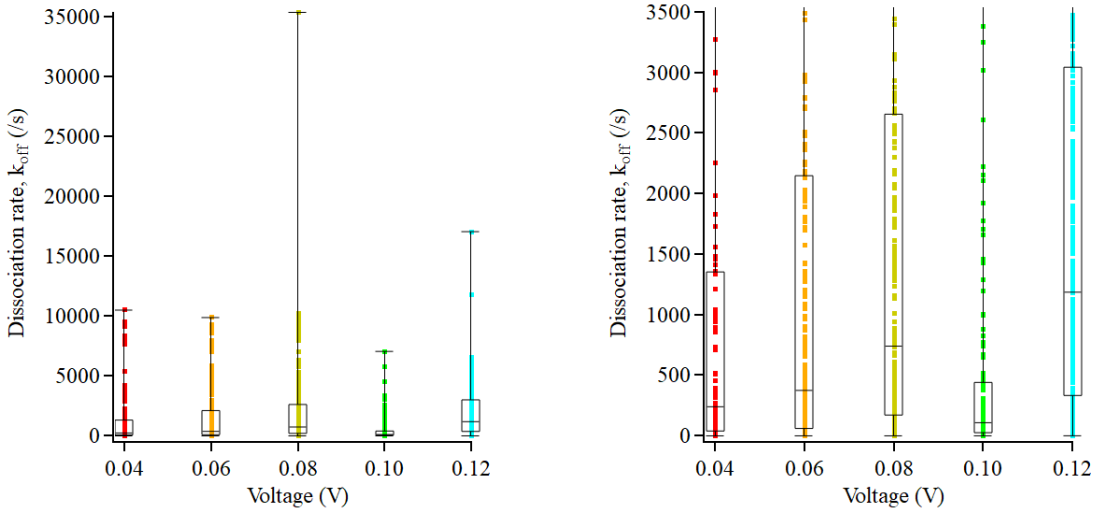


Figure 5-20: $DG_{b\text{-}norm}$ and T_t vs voltage for pore60_36nm

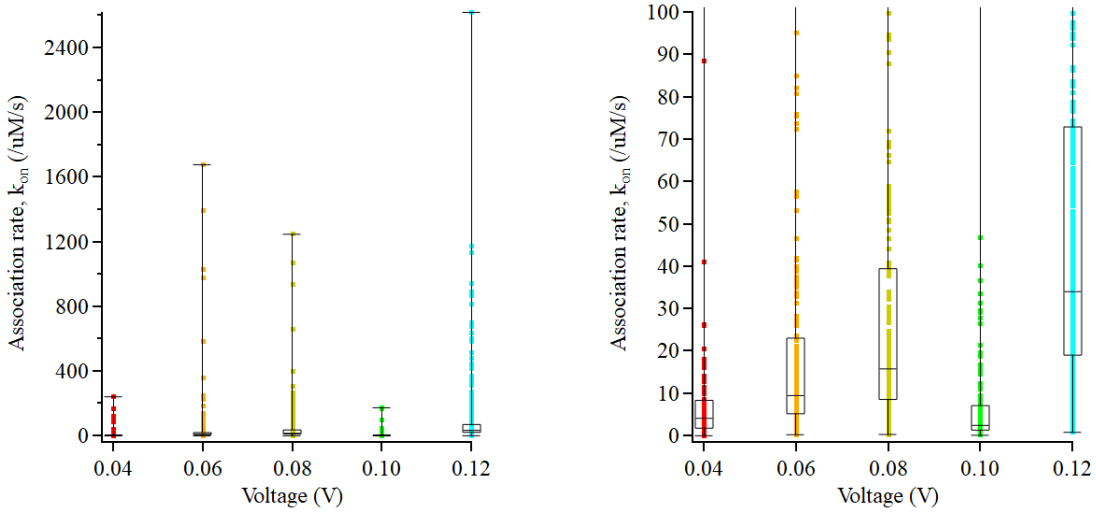
A-B) All of the events. C-D) Events with a T_t > 100 µs. A,C) Normalised change in conductance between 40 – 120 mV. B,D) Log of T_t at 40 - 120 mV. The graphs were drawn in IgorPro.

The equations for binding kinetics can be used for the entry of analytes into a nanopore (see Chapter 1). If the portal protein was confined inside the SSN, then the association rate should increase with voltage, while the dissociation rate should decrease. The voltage would keep the portal protein inside the SSN and stop it from leaving. On the other hand, if the protein was translocating through the pore, the association and dissociation rates should increase with voltage as the protein moves faster through the SSN. Figure 5-21 shows graphs for the rate constants. The association and dissociation rates increase steadily as voltage increases.

A) Dissociation rate, k_{off}



B) Association rate, k_{on}



C) Dissociation constant, K_d

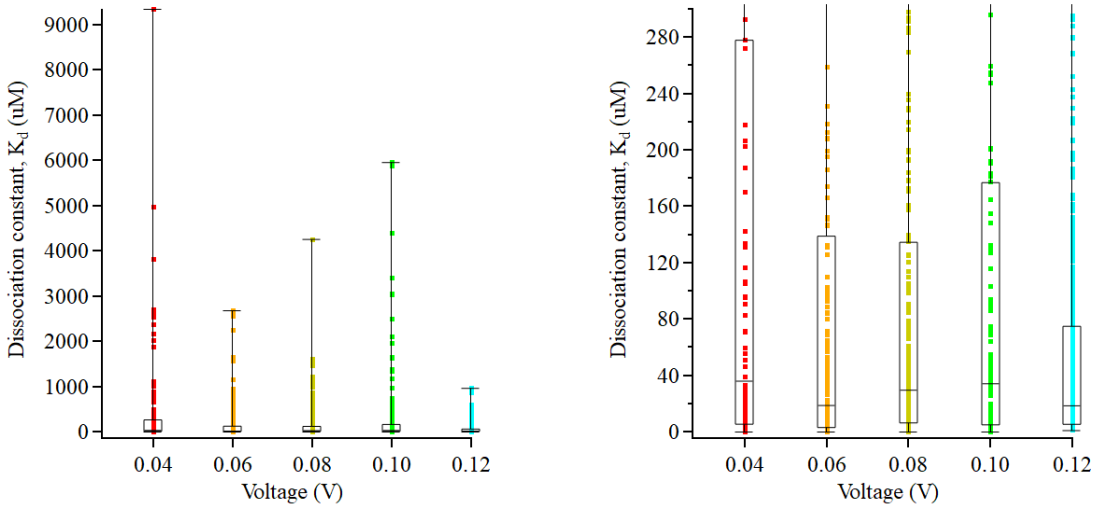


Figure 5-21: Binding kinetics graphs for pore60_36nm

Left column – whole graph. Right graph – zoom in on the box plots. A) Dissociation rate. B) Association rate. C) Dissociation constant.

(ii) Hybrid nanopore formation

In order to make a fair comparison, the data from the same buffer and pore was compared. The only difference between the GBSV1_Δ1-40_Δ383-406_816 and the GBSV1_Δ1-40_775 is the deletion of a short helix in the crown domain. This is expected to only affect the noise going through the hybrid pore, but the proteins should be compared separately just in case there are some unknown effects. The same logic applies to keeping the buffers separate, even though the dipole of the protein should be the same in each buffer.

Due to the low salt concentration of the test buffer, the nanopore diameter cannot be measured from the conductance through the pore (see Chapter 1) during experiments. Therefore, as an alternative, the baseline conductance for each event was plotted against the SSN size as found by the Spark-E2 (Figure 5-22), and against the average conductance of the event (Figure 5-23). As expected, the conductance increases with SSN diameter. Even if there are too few size repetitions for precision, the figures can be used to compare sizes and spot those that do not match the pattern. Once the diameter of the pore exceeds the pore length (in this case, 12 nm), the diameter measurements become less accurate, so this graph may be useful for estimating the approximate size of a large pore, or identifying incorrect measurements. The two instances of pores that do not fit the trend are pore49_6.2nm, which is too high, and pore79_12.54nm, which appears to be too low.

The discrepancy between measured size and apparent size in pore49_6.2nm may be due to leaving the SSN in test buffer for an extended period of time which may have allowed the pore to grow. Measurements taken before leaving it had an average baseline conductance of 2.9 nS. This group can be seen in the low tail of the box plot in Figure 5-22B and the small cluster of points in Figure 5-23B that overlaps with the points of pore48_4.8nm. The median baseline current is also higher than those of pore52_7nm, pore53_7.94nm, and pore69_8.5nm (Figure 5-22B).

A) Baseline conductance vs size

B) Between sizes 4.8 – 8.5 nm

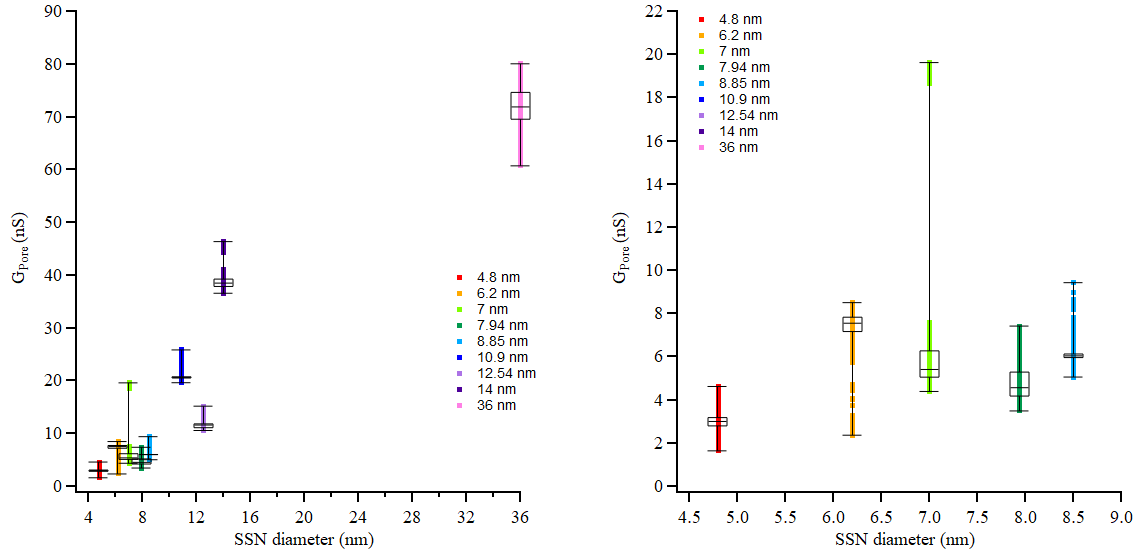
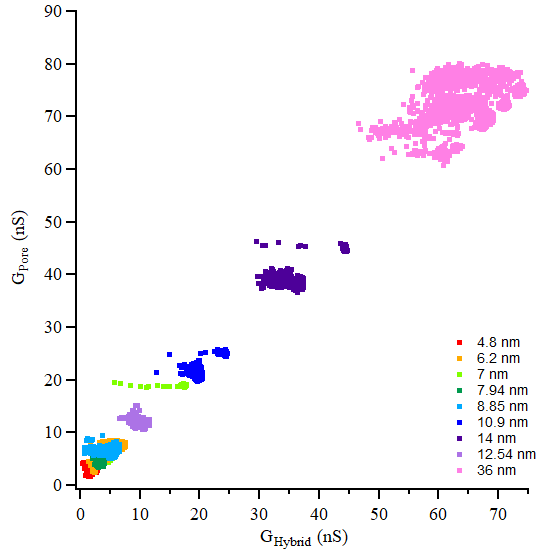


Figure 5-22: Baseline conductance vs SSN diameter

Different voltages were used for different pores, so conductance was used for comparisons, rather than current. The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. The pores used were Pore48_4.8nm at -160mV, -180mV, -200mV; Pore49_6.2nm at 200mV, 120 mV, 140 mV, 143 mV, 163 mV, 164 mV, 186 mV, 166 mV; Pore51_14nm at 163 mV, 184 mV, 204 mV, -100 mV; Pore52_7nm at 80 mV, 100 mV; Pore53_7.94nm at 407 mV, 360 mV; Pore60_36nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV; Pore69_8.5nm at -101mV, -120 mV, -140mV, -160 mV, -181 mV, -201 mV; Pore79_12.54nm at 60 mV, 80mV, 100 mV; Pore81_10.9nm at -140mV, -100 mV, 140 mV. A) G0 for SSN with a diameter between 4 and 12 nm. B) An enlarged view of the diameters between 4.5-9 nm.

As the SSN size increases, the clusters of data points become more distinct. At smaller diameters, they tend to overlap (Figure 5-23). The cluster of data points for pore49_6.2nm is much higher than expected (Figure 5-23). The actual size may be slightly larger than 8.5 nm (light blue data points).

A) Baseline vs event conductance



B) Sizes 4.8 – 8.85 nm

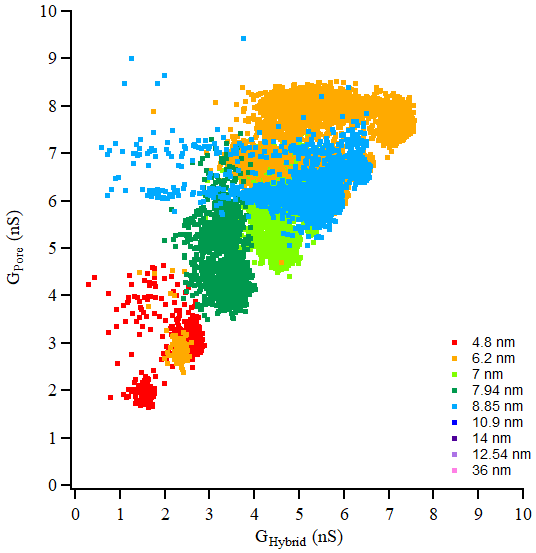


Figure 5-23: Baseline vs event conductance

The baseline conductance was plotted against the conductance of the event (A). (B) shows just the data points for sizes 4.8 – 8.85 nm. The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. The pores used were Pore48_4.8nm at -160mV, -180mV, -200mV; Pore49_6.2nm at 200mV, 120 mV, 140 mV, 143 mV, 163 mV, 164 mV, 186 mV, 166 mV; Pore51_14nm at 163 mV, 184 mV, 204 mV, -100 mV; Pore52_7nm at 80 mV, 100 mV; Pore53_7.94nm at 407 mV, 360 mV; Pore60_36nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV; Pore69_8.5nm at -101mV, -120 mV, -140mV, -160 mV, -181 mV, -201 mV; Pore79_12.54nm at 60 mV, 80mV, 100 mV; Pore81_10.9nm at -140mV, -100 mV, 140 mV. A) All events. B) An enlarged view of the smaller pores up to pore69_8.5nm.

Different voltages were used for each pore, so conductance was used for comparisons. As the SSN diameter increases, the normalised change in conductance increases. When the SSN is small, the portal protein is too large to enter it (e.g. pore48_4.8nm). But, the portal protein can still be pulled onto the mouth of the SSN and block current passing through it. The portal protein is large enough to block almost the entire mouth of the pore, which explains the high $DG_{b\text{-}norm}$ value in some of the events.

Below the diameter at which the portal protein would translocate through the SSN (~ 14 nm), the formation of a hybrid pore can be interpreted as the diameter of the nanopore shrinking to ~ 4 nm. The proportion of the original SSN current or conductance that is let through by the new 4 nm hybrid nanopore can be used as expected $DG_{b\text{-}norm}$ values for hybrid nanopore formation.

Table 5-3: Expected DGb-norm for hybrid formation

SSN diameter (nm)	Proportion blockade	Proportion let through
4	0.00	1.00
5	0.20	0.80
6	0.33	0.67
7	0.43	0.57
8	0.50	0.50
9	0.56	0.44
10	0.60	0.40
11	0.64	0.36
12	0.67	0.33
13	0.69	0.31

Figure 5-24 shows all the $DG_{b\text{-}norm}$ of all events against SSN diameter. As the SSN diameter increases, the proportion of the current blocked decreases. Most of the events have a $DG_{b\text{-}norm}$ less than the expected value for hybrid nanopore formation (Figure 5-24). Only the most extreme events for each pore appear to reach them. Pore69_8.5nm shows a spike in $DG_{b\text{-}norm}$ at its most extreme events. As the SSN diameter increases from here, $DG_{b\text{-}norm}$ decreases again. Most of the events may be bumping events. If the portal protein approaches side-on, it will block more current, as current will not be able to pass through its pore. Such events may be among those higher $DG_{b\text{-}norm}$ events.

A) Normalised change in conductance vs size B) Sizes 4.8 - 14 nm

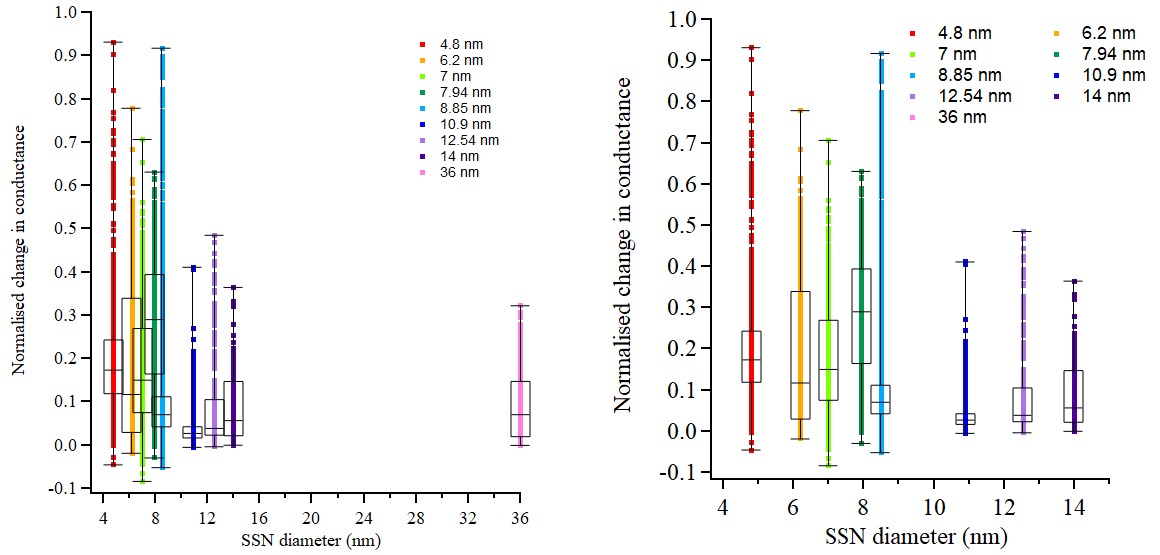


Figure 5-24: Normalised change in conductance vs size.

A) All sizes. B) Sizes 4.8 nm – 14 nm. The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. The pores used were Pore48_4.8nm at -160mV, -180mV, -200mV; Pore49_6.2nm at 200mV, 120 mV, 140 mV, 143 mV, 163 mV, 164 mV, 186 mV, 166 mV; Pore51_14nm at 163 mV, 184 mV, 204 mV, -100 mV; Pore52_7nm at 80 mV, 100 mV; Pore53_7.94nm at 407 mV, 360 mV; Pore60_36nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV; Pore69_8.5nm at -101mV, -120 mV, -140mV, -160 mV, -181 mV, -201 mV; Pore79_12.54nm at 60 mV, 80mV, 100 mV; Pore81_10.9nm at -140mV, -100 mV, 140 mV.

If the $DG_{b\text{-}norm}$ can be plotted against the dwell time, the events with higher $DG_{b\text{-}norm}$ values tend to have longer dwell times. The data points for larger pores (pore81_10.9nm onwards), tend to form a 'crab claw' shape. As the pore diameter increases, the $DG_{b\text{-}norm}$ values decrease, which can be seen in

the $DG_{b\text{-}norm}$/dwell time graphs. The graphs for pore48_4.8nm and pore69_8.5nm show the highest values of $DG_{b\text{-}norm}$ (Figure 5-24, Figure 5-25), and their $DG_{b\text{-}norm}$/dwell time graphs are a similar shape. The graph for pore49_6.2nm has a crab-claw appearance, much like the larger pores. This supports the idea that this pore grew during the experiment and that the size is an underestimate.

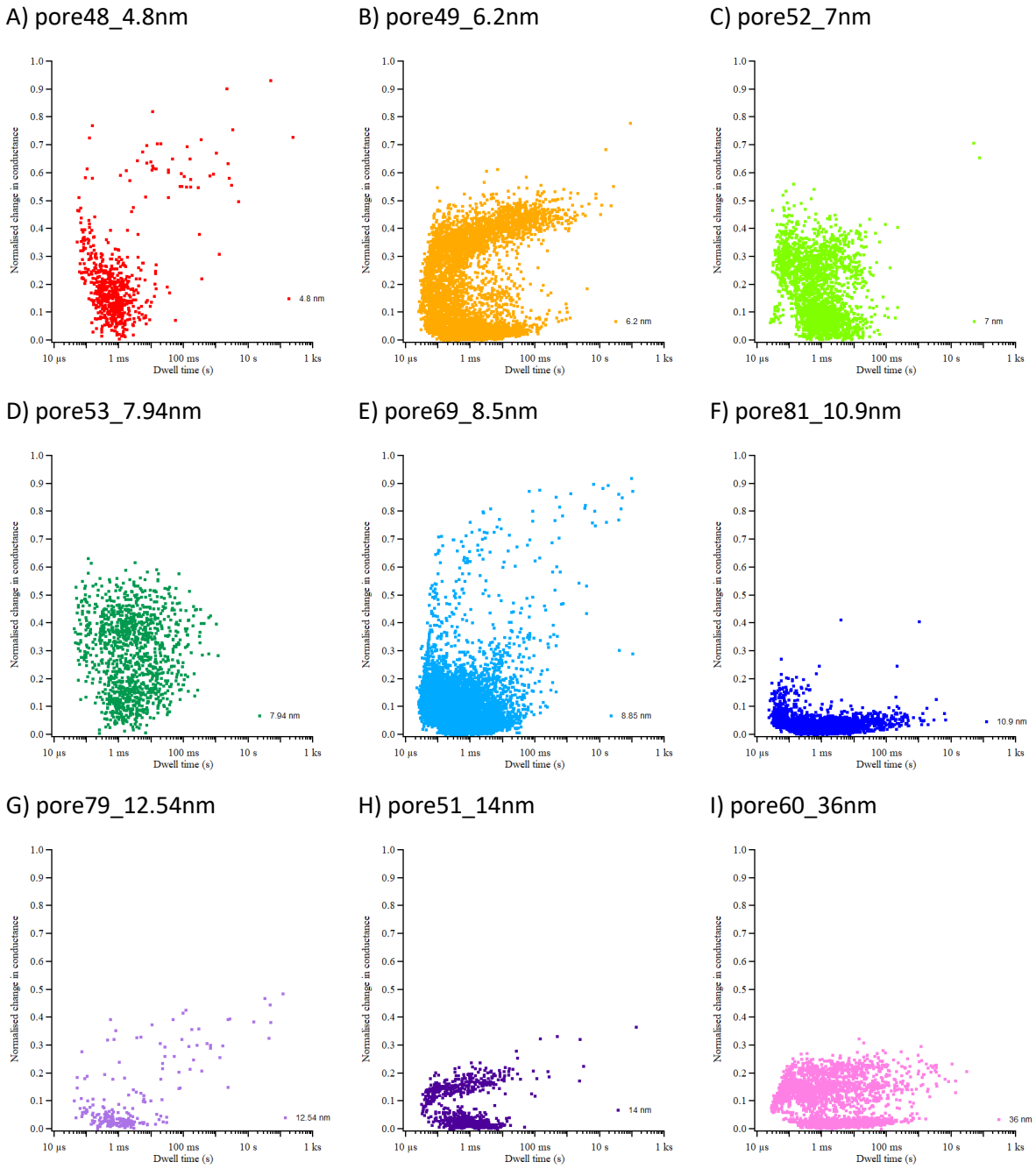


Figure 5-25: Normalised change in conductance vs dwell time

The normalised change in conductance (DGb-norm) against the event dwell time. Smaller SSN tend to have higher DGb-norm values and the data points cluster below 100 µs. The data points for larger SSN tend to form a crab-claw shape. The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. The pores used were A) Pore48_4.8nm at -160mV, -180mV, -200mV. B) Pore49_6.2nm at 200mV, 120 mV, 140 mV, 143 mV, 163 mV, 164 mV, 186 mV, 166 mV. C) Pore52_7nm at 80 mV, 100 mV. D) Pore53_7.94nm at 407 mV, 360 mV. E) Pore69_8.5nm at -101mV, -120 mV, -140mV, -160 mV, -181 mV, -201 mV. F) Pore81_10.9nm at -140mV, -100 mV, 140 mV. G) Pore79_12.54nm at 60 mV, 80mV, 100 mV.H) Pore51_14nm at 163 mV, 184 mV, 204 mV, -100 mV.I) Pore60_36nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV.

The baseline conductance vs hybrid conductance, and $DG_{b\text{-}norm}$ vs T_t graphs for the pores that used pH 7.5 buffer were drawn separately for fair comparisons (Figure 5-26 and Figure 5-27). The cluster for pore36_10.1nm is similar to that for pore81_10.9nm. Similarly, the clusters for smaller pores overlap.

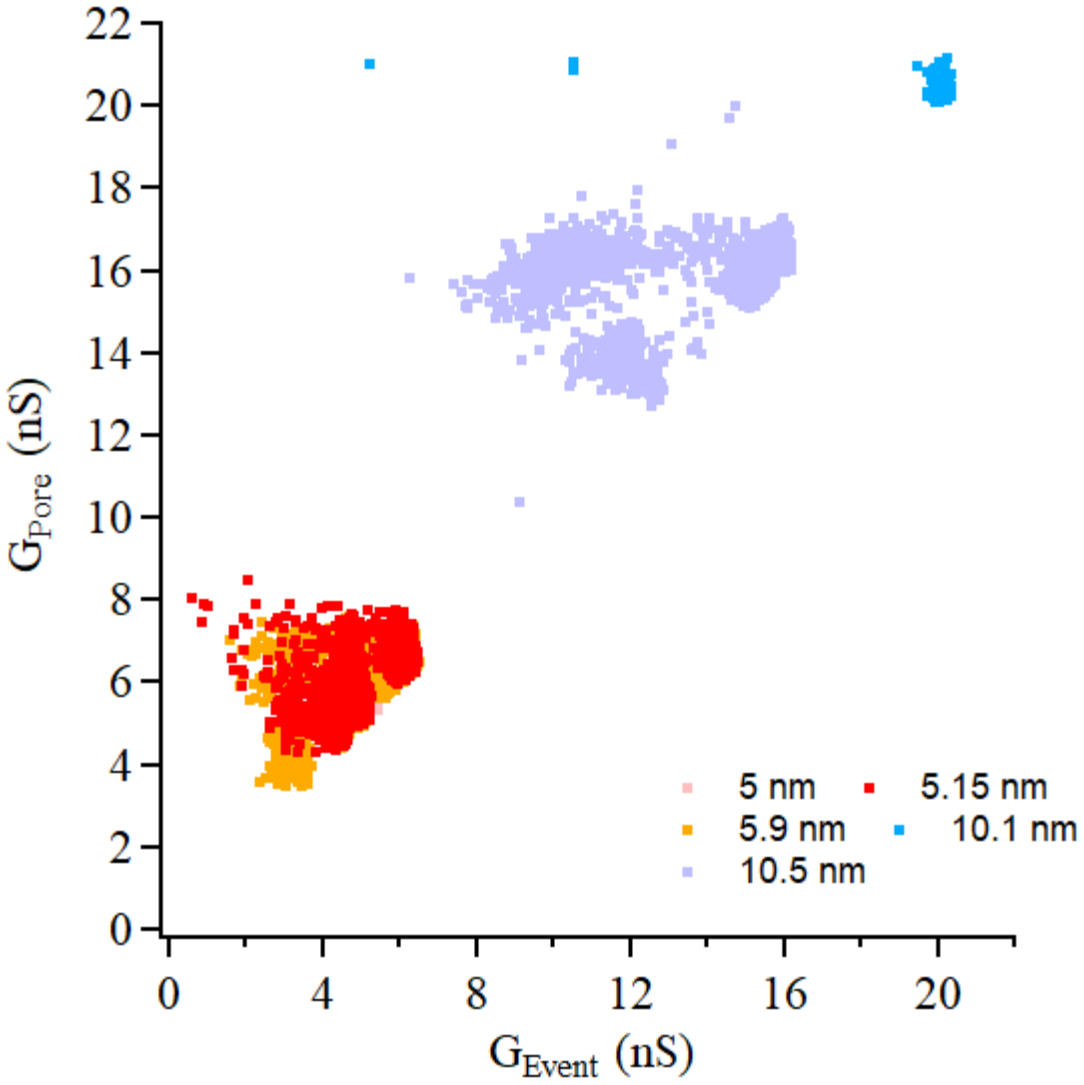


Figure 5-26: Baseline conductance for pores at pH 7.5

The buffer was 10 mM Tris-HCl, 0.4 m KCl pH 7.5. The pores are pore34_5.9nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV, 200 mV; pore35_5.15nm at 100mV, -100mV, 80 mV, 120 mV, 140mV; pore36_10.1nm at -100 mV; pore37_10.5nm at 140 mV, 150 mV, 160 mV, 180 mV, 200 mV; pore45_5nm at -101mV, -120 mV, -140 mV, -160 mV, -181 mV, -201mV.

The $DG_{b\text{-}norm}$ vs T_t graphs for the pH 7.5 pores also display a similar pattern to those in pH 6.5 buffer. Smaller pores have a large cluster at relatively short dwell time and a sparsely populated arm up to longer dwell times and higher $DG_{b\text{-}norm}$ values. The graphs for pore36_10.1nm and pore37_10.5nm appear similar to those above, even though the cluster of baseline conductance vs hybrid conductance for pore37_10.5nm is lower than pore36_10.1nm.

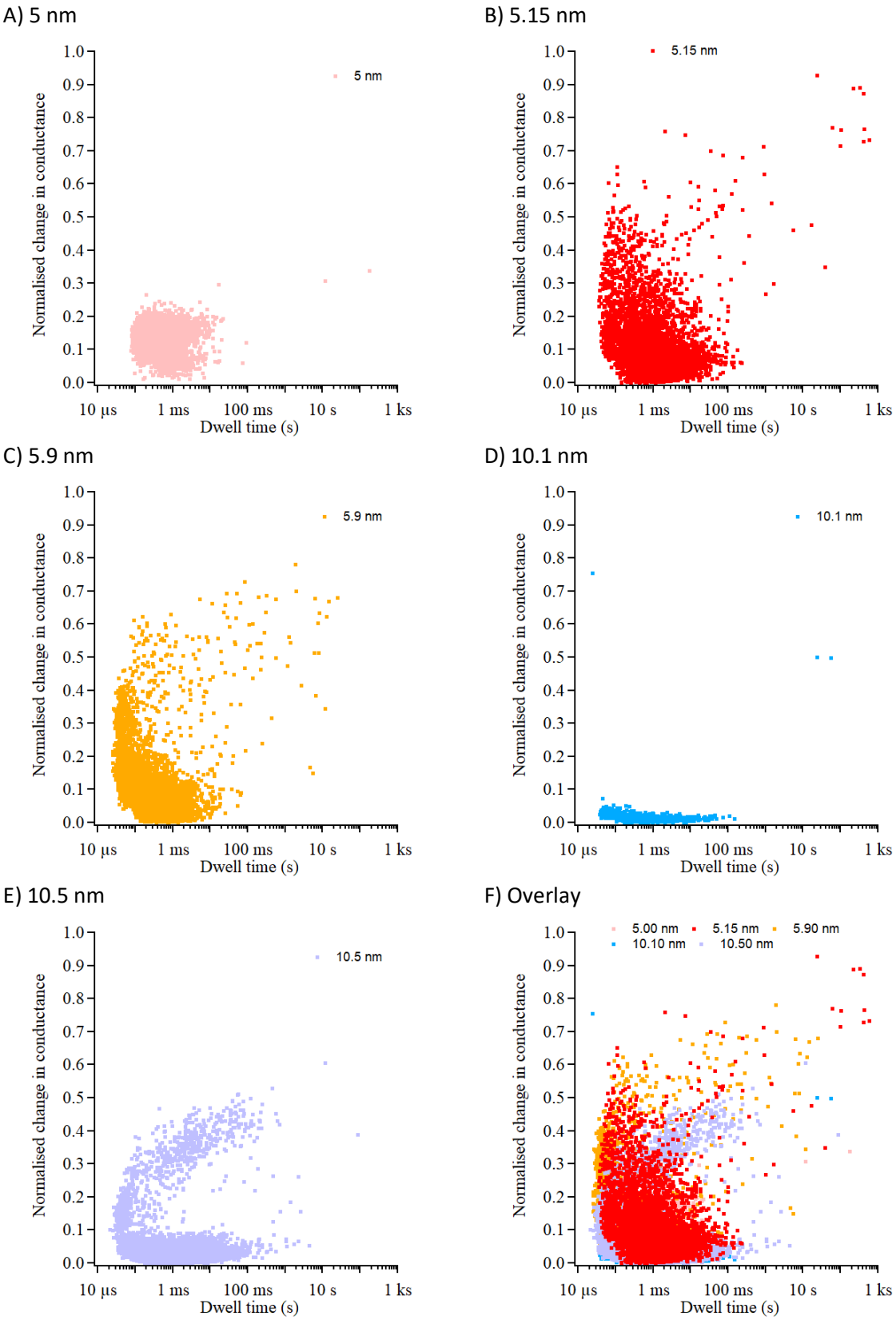


Figure 5-27: $DG_{b\text{-}norm}$ vs T_t for pores in pH 7.5 buffer

The buffer is 10 mM Tris-HCl, 0.4 M KCl pH 7.5. The pores are pore34_5.9nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV, 200 mV; pore35_5.15nm at 100mV, -100mV, 80 mV, 120 mV, 140mV; pore36_10.1nm at -100 mV; pore37_10.5nm at 140 mV, 150 mV, 160 mV, 180 mV, 200 mV; pore45_5nm at -101mV, -120 mV, -140 mV, -160 mV, -181 mV, -201mV

The $DG_{b\text{-}norm}$ vs T_t graphs for SSN with a diameter up to ~ 8 nm have a similar pattern to their events. The pore49_6.2nm should be considered to have a diameter greater than 8 nm. At such diameters, the GBSV1 portal protein cannot translocate through the SSN due to its dimensions (see Chapter 3). There is a cluster of events at low dwell time up to ~ 100 ms and low $DG_{b\text{-}norm}$. This cluster may represent bumping events, but also events in which the portal protein approaches, and lingers around the mouth of the pore for longer periods of time without blocking more of the pore. This may be from the pore adsorbing to the side of the pore mouth. There is also a cluster that resembles an arm which reaches up to longer dwell times and longer $DG_{b\text{-}norm}$ values. This may be caused by the portal protein sitting on top of the mouth of the SSN, blocking the pore.

The $DG_{b\text{-}norm}$ vs T_t graphs for SSNs with a diameter greater than ~ 8 nm have events in a 'crab claw' shape. The events can be roughly grouped into three clusters: the 'top claw', 'bottom claw', and the bumping events. The 'bottom claw' cluster contains events with a low $DG_{b\text{-}norm}$, and varying dwell times. The dwell times can be as long as those in the 'top claw' cluster. The bumping cluster is the group of events with a low dwell time and a range of $DG_{b\text{-}norm}$ values. The 'top claw' cluster is made of events with dwell time > 200 µs and a higher $DG_{b\text{-}norm}$. The bumping cluster and the 'top claw' cluster give a similar appearance to the typical graphs of blockade level vs dwell time for nanopore experiments (Stierlen et al. 2023; Greive et al. 2024). At diameters between 10-14 nm, the portal protein may be able to translocate through the SSN if it enters side-on. As the pore diameter increases, the $DG_{b\text{-}norm}$ values of the 'top claw' cluster decrease. If the portal protein completely blocked a 10 nm pore, then the expected $DG_{b\text{-}norm}$ would be 0.97. If a hybrid nanopore formed, then the diameter should be ~ 4nm, and the expected $DG_{b\text{-}norm}$ would be 0.6. Figure 5-28 shows a long-lived hybrid in a 10 nm SSN. This may be the best example of a hybrid pore formation.

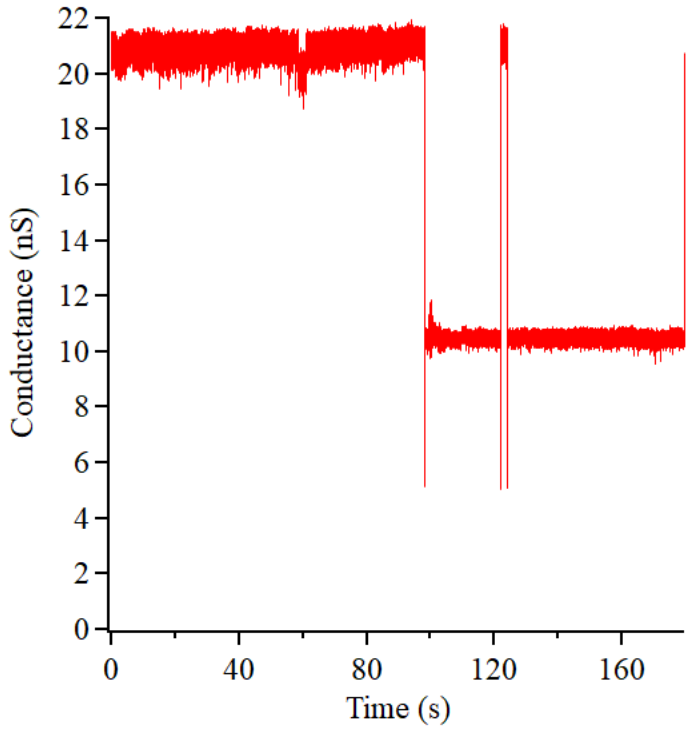


Figure 5-28: Example of probable hybrid nanopore formation in 10.1 nm SSN

The baseline conductance is ~ 21 nS. It drops suddenly at ~ 90s to ~ 11 nS. It briefly returns to the baseline at ~ 120 s, but then returns to ~ 11 nS. The pore was pore36_10.1nm. The buffer was 10 mM Tris-HCl 0.4 M KCl pH 7.5. The voltage was 100 mV.

Dwell time is another parameter that can be used to estimate the presence of hybrids and their effectiveness. Since different voltages were used for each pore, the dwell time was divided by the voltage to give the dwell time per volt (Figure 5-29). The highest median dwell time per volt was seen in pore48_4.8nm.

It is expected that dwell time inside the SSN would be affected by the voltage. If the portal protein cannot translocate through the SSN, higher voltages would give longer dwell times. On the other hand, if the SSN was wide enough for the portal protein to translocate, higher voltages should give shorter dwell times.

A) Dwell time per volt

B) Sizes 4.8 – 14 nm

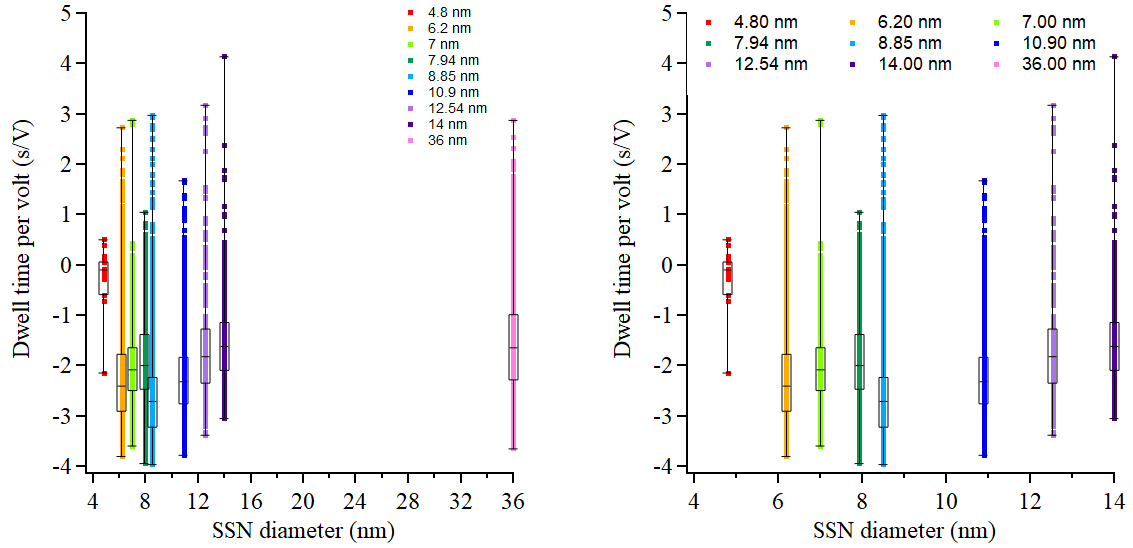


Figure 5-29: Dwell time per volt vs size

The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. The pores used were A) Pore48_4.8nm at -160mV, -180mV, -200mV. B) Pore49_6.2nm at 200mV, 120 mV, 140 mV, 143 mV, 163 mV, 164 mV, 186 mV, 166 mV. C) Pore52_7nm at 80 mV, 100 mV. D) Pore53_7.94nm at 407 mV, 360 mV. E) Pore69_8.5nm at -101mV, -120 mV, -140mV, -160 mV, -181 mV, -201 mV. F) Pore81_10.9nm at -140mV, -100 mV, 140 mV. G) Pore79_12.54nm at 60 mV, 80mV, 100 mV.H) Pore51_14nm at 163 mV, 184 mV, 204 mV, -100 mV.I) Pore60_36nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV.

The σ of the current in a hybrid nanopore is an important parameter in assessing its future utility. The events with a short duration tend to have a low level of noise, as expected. The data points for pore60_36nm show the highest median and maximum noise. At this diameter, it is possible that multiple portal proteins can enter the pore at once. However, the hypothetical normalised conductance change for the entry of a portal protein is 0.27.

As the SSN diameter increases from 4.8 nm, the conductance noise increases. However, pore69_8.5nm and pore79_12.54nm have a lower median and maximum noise level than their neighbours. As mentioned earlier, it is possible that pore79_12.54nm is actually smaller than thought. If so, the low noise maximum could be due to a snug fit between the portal protein and SSN. This could also apply to pore49_6.2nm.

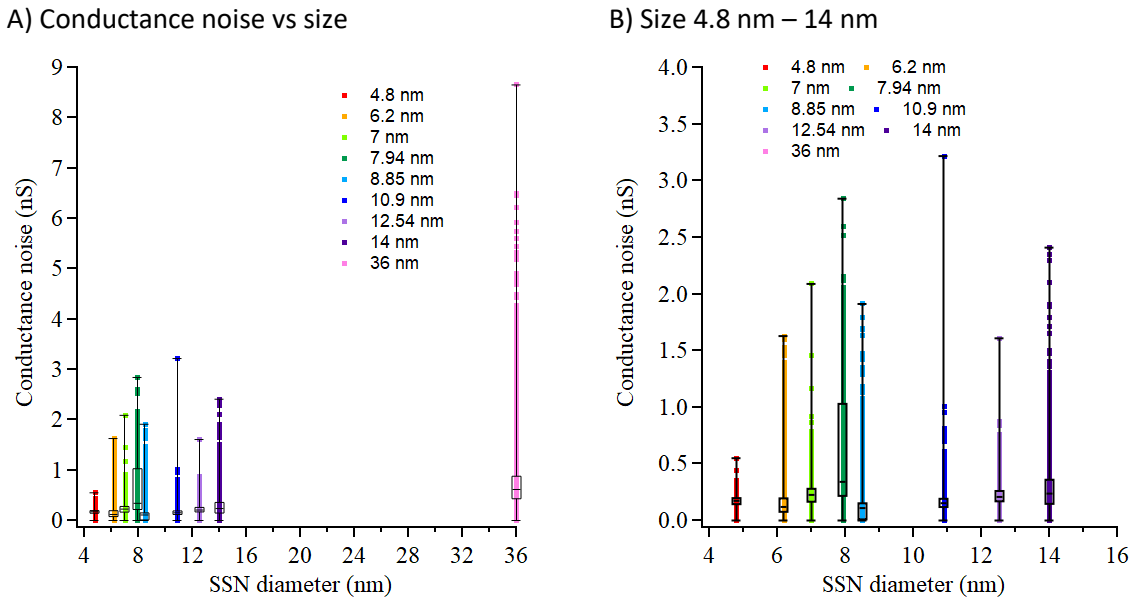


Figure 5-30: Conductance noise vs size

The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5. The pores used were A) Pore48_4.8nm at -160mV, -180mV, -200mV. B) Pore49_6.2nm at 200mV, 120 mV, 140 mV, 143 mV, 163 mV, 164 mV, 186 mV, 166 mV. C) Pore52_7nm at 80 mV, 100 mV. D) Pore53_7.94nm at 407 mV, 360 mV. E) Pore69_8.5nm at -101mV, -120 mV, -140mV, -160 mV, -181 mV, -201 mV. F) Pore81_10.9nm at -140mV, -100 mV, 140 mV. G) Pore79_12.54nm at 60 mV, 80mV, 100 mV.H) Pore51_14nm at 163 mV, 184 mV, 204 mV, -100 mV.I) Pore60_36nm at 40 mV, 60 mV, 80 mV, 100 mV, 120 mV.

(e) Voltage

The voltage dictates the electric field applied to the molecules in solution. If the voltage is too low, it may not be enough to push the portal protein into the SSN. A representative example of data is shown here for pore34_5.9nm. The hybrid events were selected using the DEEP ONES method. The $DG_{b\text{-}norm}$ vs dwell time graphs for different voltages are shown in Figure 5-31.

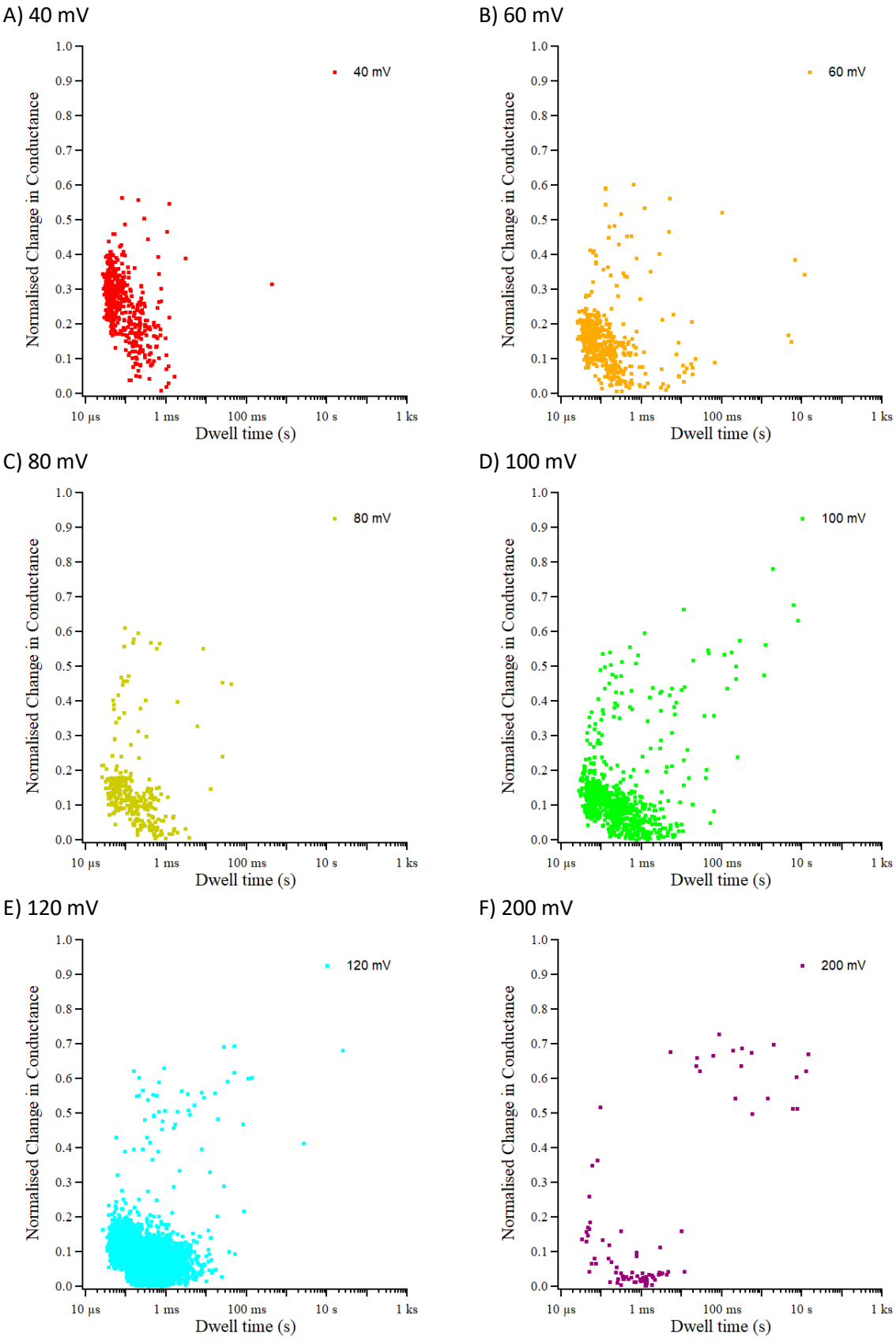


Figure 5-31: $DG_{b\text{-}norm}$ vs dwell time for pore34_5.9nm

The pore was pore34_5.9nm. The buffer was 10 mM Tris-HCl 0.4 M KCl pH 7.5

A) All selected events

B) Events selected by the DEEP ONES

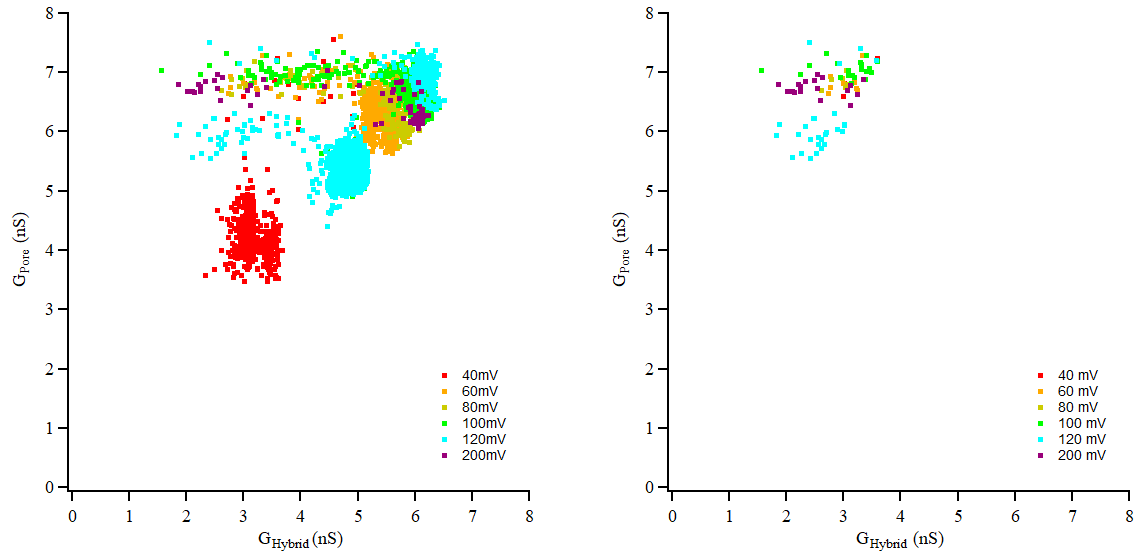


Figure 5-32: Baseline vs hybrid conductance for pore34_5.9nm

A) Data for all of the events, including hybrids and bumping. B) The data from the events chosen as hybrids according to the DEEP ONES method. The buffer was 10 mM Tris-HCl 0.4 M KCl pH 7.5.

As voltage increases, the maximum current noise level increases for both all events and hybrid events. The interquartile range for conductance noise for all the events is approximately the same across voltages. When only the hybrid events were examined, the current noise appeared to increase with voltage, but the conductance noise appeared more varied (Figure 5-33).

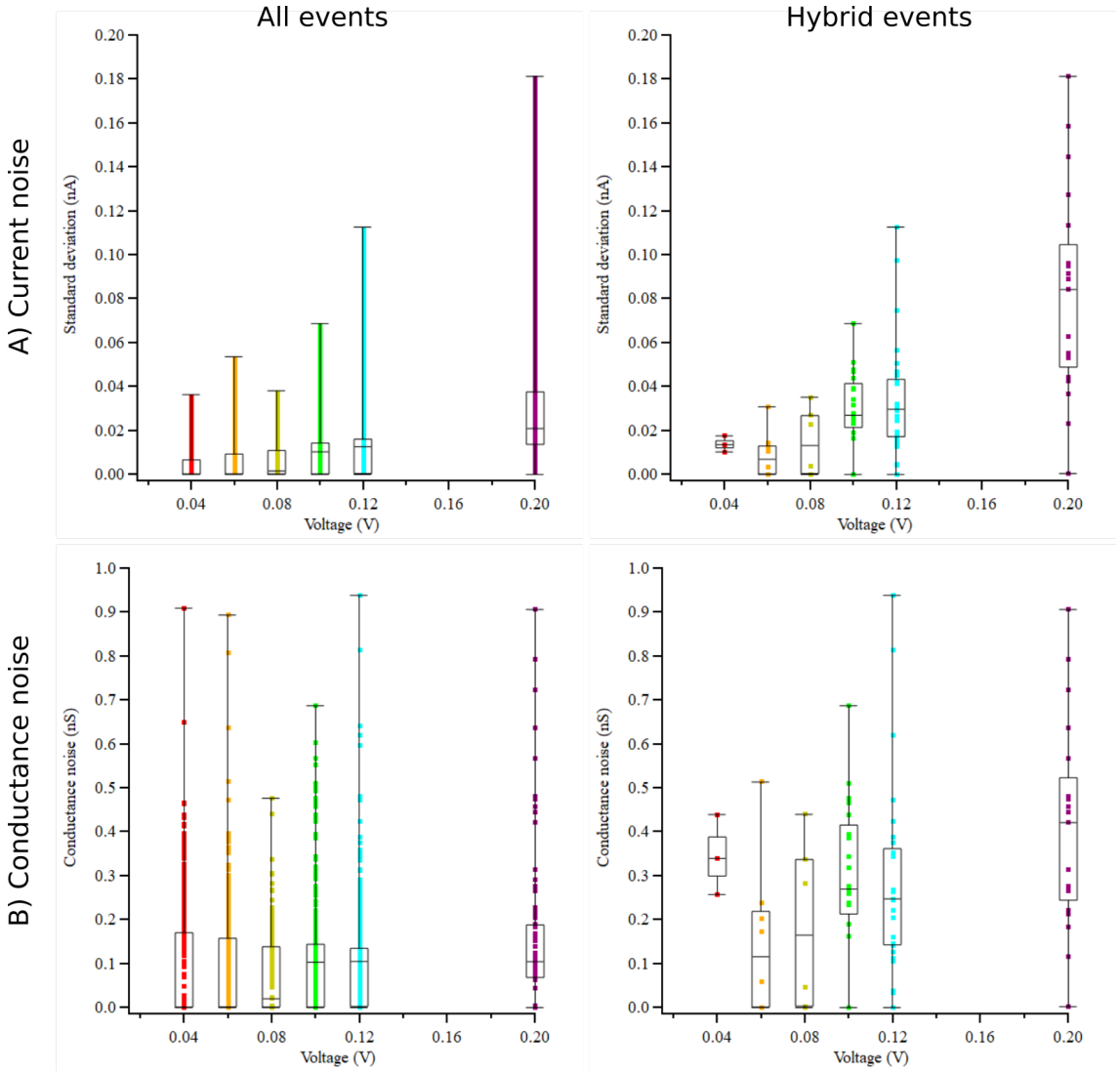
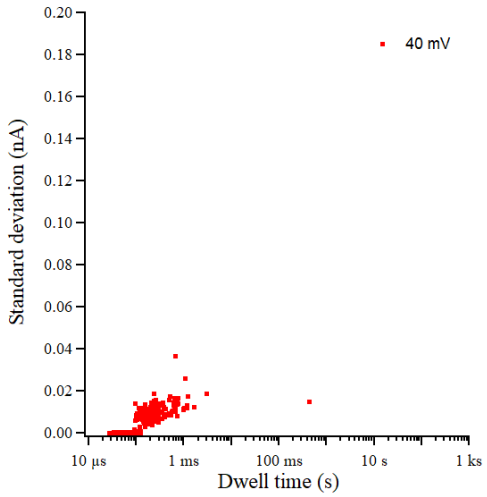


Figure 5-33: Current and conductance noise for pore34_5.9nm

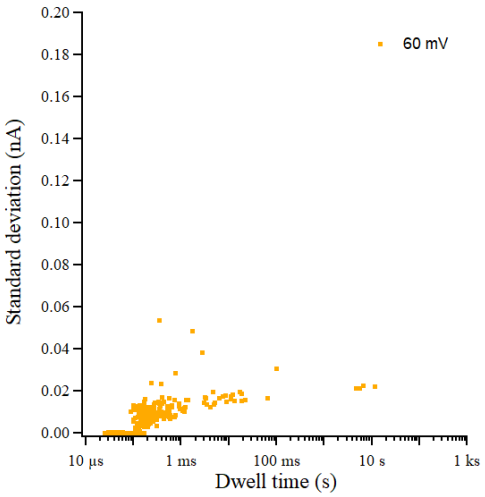
Events with noise < 1 pA can be classed as bumping. A) Current noise from all events (left) and hybrid events (right). B) Conductance noise from all events (left) and hybrid events (right).

When the standard deviation was plotted against dwell time, it showed that the shorter events below ~200 µs, which would usually be classed as bumping events, had a very low noise level (Figure 5-34) (Greive et al. 2024; Ratinho et al. 2024). Events with higher noise levels, tended to be longer. As the voltage increased, the dwell times and noise levels also increased (Figure 5-34, Figure 5-35). More data points would make the trend clearer.

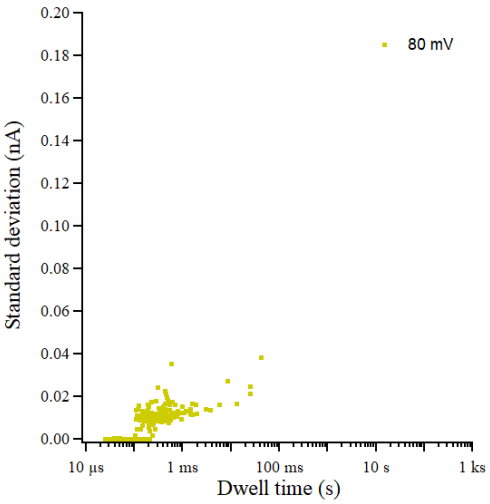
A) 40 mV



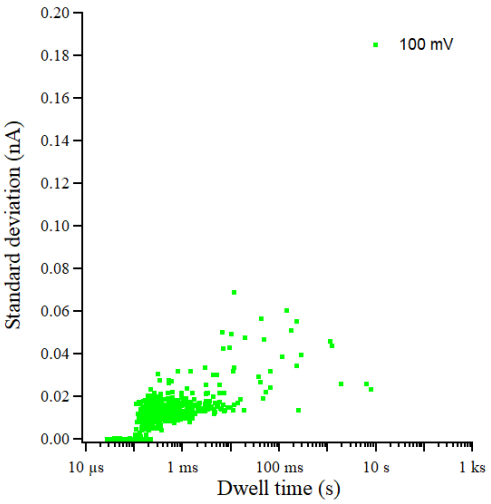
B) 60 mV



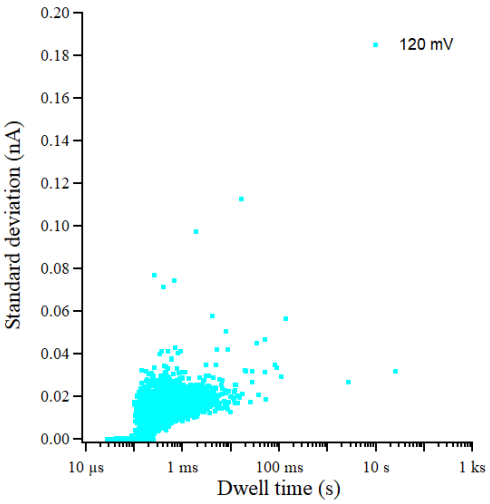
C) 80mV



D) 100 mV



E) 120 mV



F) 200 mV

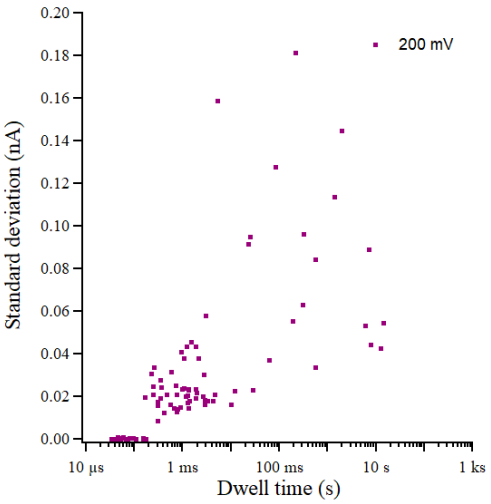


Figure 5-34: Standard deviation vs dwell time at different voltage

The pore was pore34_5.9nm. The buffer was 10 mM Tris-HCl 0.4 M KCl pH 7.5

When the hybrid events are selected using the DEEP ONES rule set, the dwell time appears to increase with voltage. However, the pattern is not very clear. Even at high voltages, there are still some events with very low dwell times.

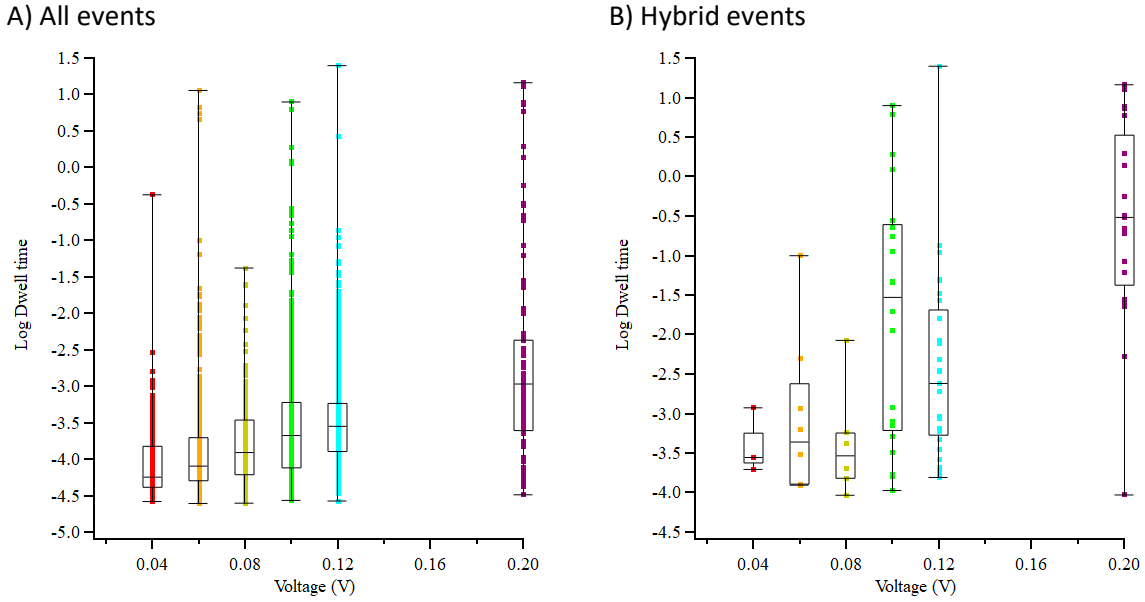


Figure 5-35: Dwell time vs voltage for pore34_5.9nm

Log dwell time for all events (A) and hybrid events (B)

The frequency of all events does not show a clear pattern with voltage (Figure 5-36). However, if only the hybrid events are selected, then the frequency shows a trend of increasing with voltage. The frequency here refers to the number of total or hybrids per second within the total recording time.

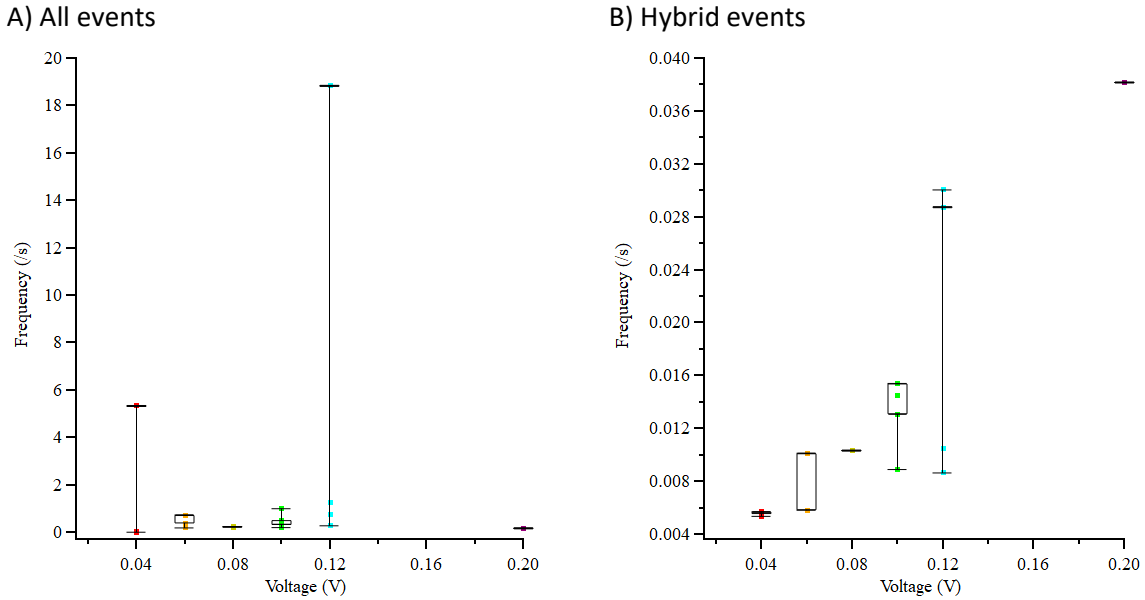


Figure 5-36: Frequency vs voltage for pore34_5.9nm

Frequency for all events (A) and hybrid events (B).

The association and dissociation rates and dissociation constant can also be found (Figure 5-37). The association rate and dissociation constants do not show a clear pattern with respect to voltage. The interquartile ranges are narrow and low, and there are a small number of extreme values. This is probably because they include hybrid events and bumping. However, as the voltage increases, the dissociation rate clearly decreases. If the protein was able to pass through the SSN, the dissociation rate should increase with voltage. It is decreasing because the voltage is keeping the portal protein inside the SSN.

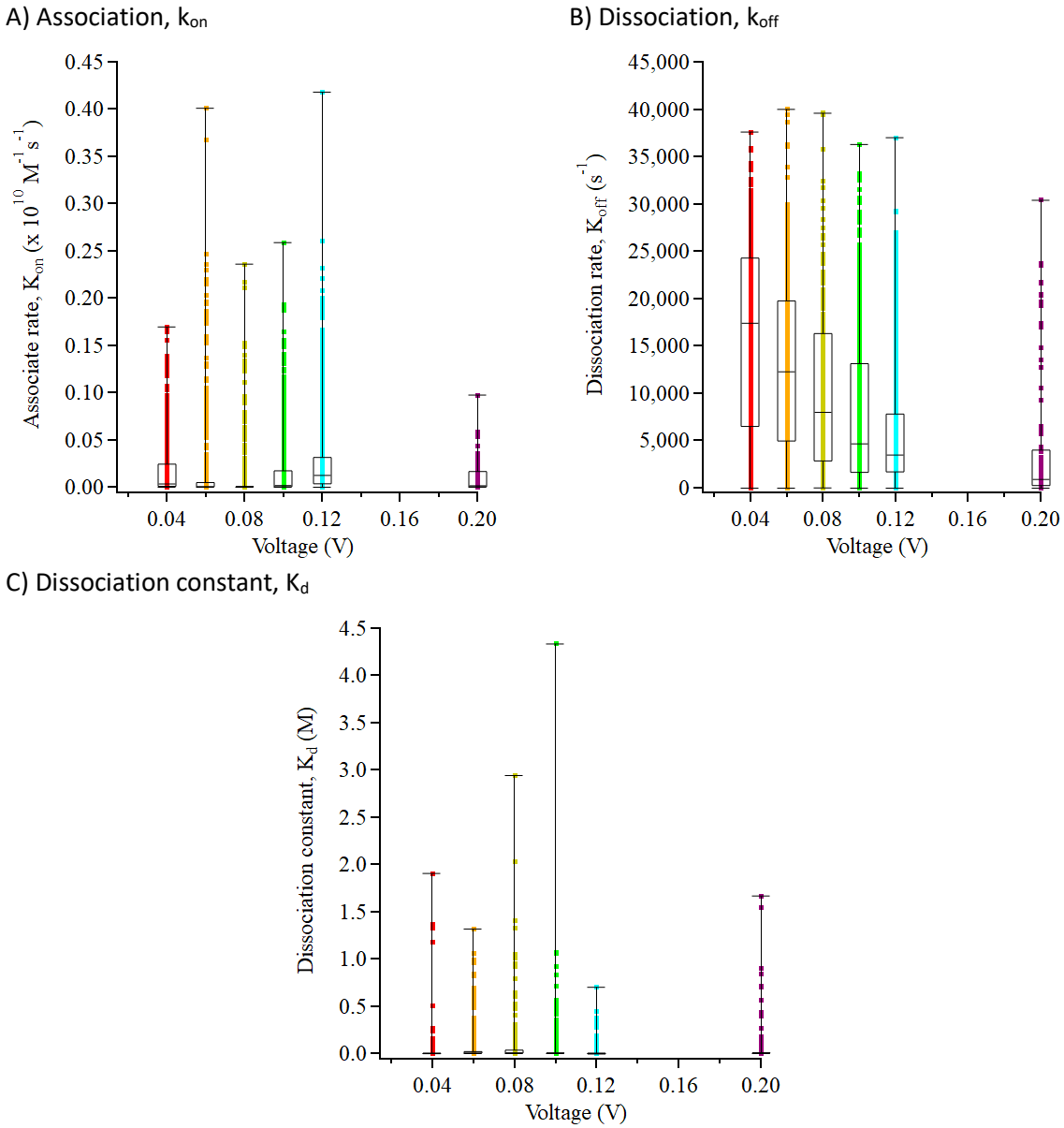


Figure 5-37: Association, dissociation rates and dissociation constant

Association rate (kon) (A), dissociation rate (koff) (B), equilibrium dissociation constant (Kd) (C).

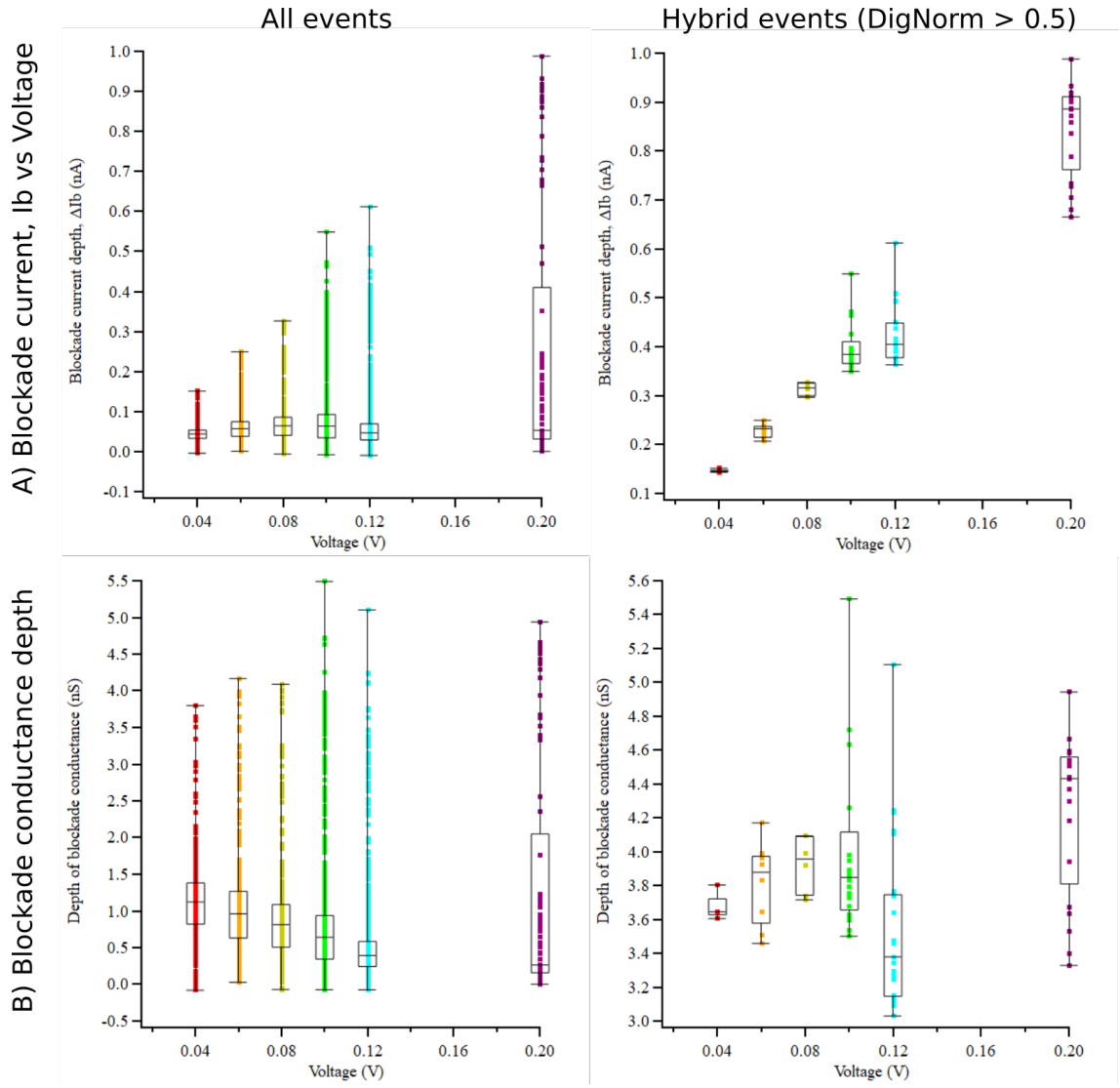


Figure 5-38: Blockade depth for pore34_5.9nm

The current depth (A) and conductance depth (B). Left – all events. Right – hybrid events.

(f) Optimal portal protein

The ideal portal protein to form hybrid nanopores would insert into the SNN easily, in the right orientation (clip domain first), give low noise, minimal current leakage, clear current changes when analytes enter or translocate, and remain stably inserted for long periods of time. Three GBSV1 mutants were tested for potential usage as nanopores: GBSV1_Δ1-40_775, GBSV1_Δ1-40_Δ383-406_816, and GBSV1_Δ1-40_P282Q_820 (see Chapter 2). GBSV1_Δ1-40_775 is a truncation of the GBSV1 portal protein which lacks the long N-terminal tail. It was the GBSV1 portal protein mutant that was successfully purified and imaged by TEM. The GBSV1_Δ1-40_ Δ383-406_816 protein is lacking the last helix of the crown domain which could not be modelled in the cryoEM structure of GBSV1_Δ1-40_775 and was therefore thought to be disordered and a potential source of noise if used in a nanopore. The GBSV1_Δ1-40_P282Q_820 protein has a mutation before the tunnel loops which is

intended to allow them to extend into the tunnel. The narrower tunnel should give the nanopore an increased SNR and allow smaller analytes.

There were far fewer tests done with the GBSV1_Δ1-40_P282Q_820 protein than for the other two proteins. No deep hybrids were observed for the tests using it. Therefore, the performance of this protein as a nanopore, and a comparison to the others cannot be made. When the GBSV1_Δ1-40_775 and GBSV1_Δ1-40_Δ383-406_816 proteins formed what could be hybrid nanopores, there was no discernible difference between their performances. The graphs for baseline vs hybrid conductance for both proteins appear similar for the sizes tested (Figure 5-23 and Figure 5-26). The same is true for the graphs of the normalised change in conductance vs dwell time (Figure 5-25 and Figure 5-27).

5.3.2.3. Adding analytes

The ultimate goal of making hybrid nanopores is to use them for probing/detecting different types of molecules in solution. After sufficient data had been collected about hybrid nanopore formation conditions, different analytes were added to the flow cell. To prevent misidentification of signals, only one analyte was added at a time. The analytes were bradykinin or hairpin DNA.

(a) Bradykinin

The NMR structure of bradykinin (PDB ID: 6f3v) was put into US-SUMO (Rai et al. 2005; Brookes, Demeler, and Rocco 2010; Brookes et al. 2010; Brookes and Rocco 2018) to find its R_g and isoelectric point. The Rg of bradykinin is 0.676886 nm, so the diameter is 1.373772 nm. The pI of bradykinin is 12, so at pH 6.5 it should be positively charged. It was added to the *trans* compartment (working electrode), so under a positive voltage, it should move through the nanopore to the *cis* compartment (ground electrode). If a hybrid nanopore was present, bradykinin would move from the clip domain towards the crown domain. The SSN used was pore81_10.9nm. The buffer was 10 mM Tris-HCl, 0.4 M KCl. pH 6.5. A probable hybrid pore was formed and 10 µM bradykinin was added to the *trans* compartment. However, after the flow cell had been left to equilibrate, the current had returned to the previous level, suggesting that the portal protein had left the SSN. Another hybrid nanopore could not be formed, so the results are inconclusive.

(b) Hairpin DNA

Hairpin DNA was added to the *cis* compartment of the flow cell. If a hybrid nanopore formed, the hairpin DNA should move from the *cis* to the *trans* compartment by EP and pass from the crown domain of the portal protein towards the clip domain. Hairpin DNA was added with the expectation that the single stranded portion would pass through the pore, but, if the portal protein was in the closed conformation, the double-stranded section would be too large to pass through (Cressiot et al.

2018). However, when the hairpin DNA was added, the current was close to 0 nA (Figure 5-39). This could be because the DNA had entered the SSN and formed a blockade inside the pore. This type of event is called a clog and is a common problem in both solid-state and biological nanopores (Kubota et al. 2019).

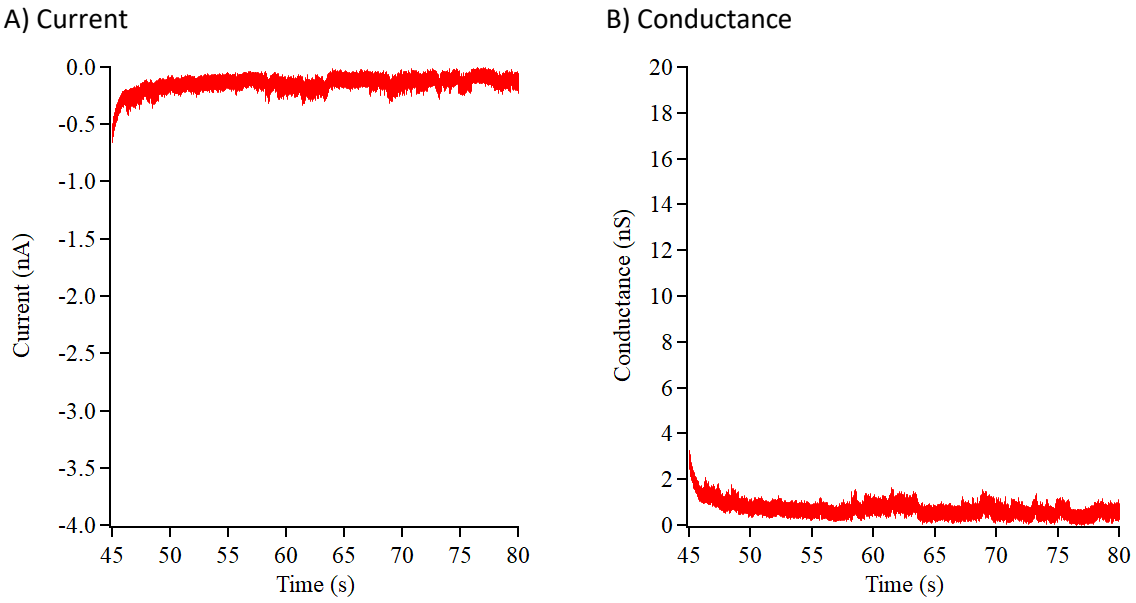


Figure 5-39: Example of hairpin DNA clogging an SSN

The pore was pore69_10.1nm. The voltage was -200 mV. The buffer was 10 mM Tris-HCl, 0.4 M KCl pH 6.5.

5.4. Discussion

5.4.1. Overall best conditions for hybrid nanopore formation

The aim of this chapter was to find the optimal conditions for forming hybrid nanopores with the GBSV1 portal protein. The ideal hybrid pore would have low noise, a long life time at a range of voltages. Due to limited time at the DreamPore lab, not all of the portal protein mutants and conditions could be tested in the time available. Not all voltages between -200 mV and 200 mV were tested for each SSN, so conclusions about the effects of pore diameter on the velocity of EP and EOF cannot be drawn.

From the analysis above, the best SSN diameter for forming hybrid nanopores is probably between ~9-10 nm. If the diameter is smaller, the portal protein may not be able to fit into the SSN and may block the mouth of the SSN. If the diameter is larger, the portal protein may translocate entirely.

The buffers tested were 10 mM Tris-HCl 0.4 M KCl pH 6.5, 10 mM Tris-HCl 0.4 M KCl pH 7.5, and 10 mM Tris-HCl 20 mM $MgCl_2$ pH 7.5. Too few tests were done for the 10 mM Tris-HCl 20 mM $MgCl_2$ pH 7.5 buffer for conclusions to be drawn. The change between pH 7.5 and pH 6.5 did not appear to make a noticeable difference. Since hybrid nanopores could not be confidently identified, analysis of formation frequency with voltage could not be done. Thus, the optimal voltage for hybrid nanopore formation is not defined. Similarly, the preference for EP or EOF in hybrid nanopore formation could not be confidently defined.

Figure 5-40: Summary of optimal conditions for hybrid nanopore formation

Condition	Optimal
Thickness (nm)	12
Buffer	n/a
SSN diameter (nm)	~9-10
Voltage	n/a
Portal protein mutant	n/a

5.4.2. Biological nanopores

5.4.2.1. Protein labelling

Due to limited time at DreamPore, only a small number of attempts were made to use GBSV1_Δ1-40_S135C_792 as a biological nanopore before efforts were refocused on its use as a hybrid nanopore. No insertions of the portal protein into the membrane were seen. It is possible that the MPB PE did not attach to the portal protein, which is why it could not enter the membrane. Since the CMC of MPB PE is unknown, it may have formed micelles in solution, which would stop it attaching to the portal protein. Finding the CMC would confirm this and inform future attempts. In order to test whether the portal protein can insert into a membrane, liposome leakage assays can be performed (Delcour 2013).

Alternatively, MPB PE could be mixed with a mixture of isopropanol/decane. The isopropanol/decane/MPB PE mixture would be used to paint lipid membranes on cuvettes. GBSV1_Δ1-40_S135C_792 would then attach to MPB PE in the membrane (Lim et al. 2016).

5.4.2.2. Alternative lipid attachment methods to portal proteins

If the MPB PE lipid does not prove a viable method, there are alternatives for inserting proteins into lipid membranes. pHLIP proteins form a-helices below neutral pH which can insert into lipid membranes (Jia Tang and Gai 2008). Lowering the pH would change the surface properties of the portal protein, and thus change its ion specificity. The volume of acid to be added to lower the pH would have to be calculated beforehand, as it would be very difficult to measure the pH of the flow cell compartment. While GBSV1_Δ1-40_775 has been shown to be stable at different pH buffers, the acidic pH may also limit the analytes that could be used and other mutants have not yet been tested. The original technique of maleimide-porphyrin could also be used as done previously for G20c portal protein (Cressiot et al. 2017b).

5.4.3. Solid-state nanopores

5.4.3.1. CDB in 30 nm thick SiN_x chips

The previous work on hybrid nanopores with the G20c protein used 30 nm thick SiN_x chips fabricated using TEM drilling (Cressiot et al. 2018). Thicker pores would result in a longer SSN and a slower EOF velocity (Melnikov, Hulings, and Gracheva 2017). Therefore, there would be less competition between the EOF and EP over the velocity of the portal protein and the net velocity of the portal protein would depend more on the EP. If the portal protein does not experience the constant push and pull of the EP and EOF while inside the SSN, it should result in longer lifetimes for the hybrid nanopores. CDB has been used successfully to fabricate nanopores in 30 nm chips before (Kwok, Briggs, and Tabard-Cossa 2014; Briggs et al. 2015), however, to the best of our knowledge, the Spark-E2 instrument specifically has not been used on 30 nm chips before. The machine was designed to fit 10 nm thick chips. The 30 nm chips were placed into the flow cells by only screwing the two halves of the flow cell together until finger tight, not until they were flush, as would be usual for 10 nm chips. When the flow cell halves were screwed together until flush, the fit was so tight that liquid could not flow through the chambers and the silicon gaskets that hold the chip broke. Screening showed that many of the 30 nm membrane chips were already broken. These included chips which had been freshly broken from the wafer, so they may have been damaged during transportation. Wafer-1 was found to contain mostly broken pores compared to wafer-2, so the problem was not the fabrication of the wafers themselves.

Kwok *et al* used a home-made CDB system to compare fabrication times for 30 nm membranes under different voltages and pH buffers (Kwok, Briggs, and Tabard-Cossa 2014). The longest fabrication time in their work was ~7 hours at 13 V using a pH 7 or pH 10 buffer. Their results for pH4 were between ~19 seconds and ~5 minutes for 16 and 13 V respectively. However, 30 nm thick chips were tested in this work in the Spark E2 with pH 4 buffer at 18 V, but no fabrication was observed. It is not possible to run the Spark-E2 at higher voltages. Unfortunately, further overnight fabrication tests could not be attempted due to time constraints. At this time, the exact reason that SSNs could not be fabricated in 30 nm thick chips is not clear. The dielectric strength (V/m) of a material is constant, so as the distance across the material increases, the break down voltage increases. However, using higher breakdown voltages comes with some complications for nanopore fabrication. When the pore is fabricated, it can grow larger than desired before the voltage is shut off ('Northern Nanopore Instruments User Manual' 2022). There is also an increased risk of fabricating multiple pores.

5.4.3.2. Alternative SSN fabrication methods

CDB in the Spark-E2 is not the only way of making an SSN. The traditional method of TEM drilling could be used as done before. Other techniques include laser etching or laser assisted controlled breakdown.

5.4.3.3. SSN tunnel topology

The CDB technique can be tailored to produce different shaped nanopores. The use of extreme pH buffers can produce conical nanopores which are severely rectified. Using very low pH on the anode side, and very high pH on the cathode side (called the 'fast conditions') could be a way of making SSN in 30 nm chips (Briggs et al. 2015). The use of NaOCl in the fabrication buffer appears to alter the surface charge of the chip, which would affect the EOF inside the SSN. Would the reduction in EOF hinder the corking of the SSN by the portal protein, or reduce any movement of the portal protein inside the SSN and reduce noise (D Y Bandara et al. 2020; Saharia et al. 2023)? If CDB cannot be made viable, SSN could be made using a TEM as done previously (Fujinami Tanimoto et al. 2021; Cressiot et al. 2018).

If the Spark-E2 or some other CDB apparatus can be used to make SSN in 30 nm thick SiN_x chips, it would allow wider diameter SSN to be made with a more accurate and precise diameter. However, thicker SSN may also give a lower signal-to-noise ratio, so even if SSNs can be made in 30 nm thick chips, they still may not be the optimal choice for hybrid nanopore formation (Waduge et al. 2017).

5.4.3.4. Persistent hysteresis

Even after piranha cleaning and using a slower ramp speed for IV curves, there was still some hysteresis in the SSN (Figure 5-6). Some of this could be attributed to capacitance noise. It may also

have been a tiny bubble forming inside the portal protein at higher voltages. Paulo *et al* created a mutant FraC protein in which they could control the presence of a nanobubble inside the tunnel by changing the voltage (Paulo, Sun, di Muccio, et al. 2023). The FraC pore begins with a nanobubble in the lumen, the "dry state". The bubble prevented the passage of ions between the *cis* and *trans* chambers, thus reducing the current. As the voltage increased, the probability of the bubble dissipating increased. Eventually, it does so, and the lumen is filled with electrolyte solution, also called the 'wet state', in which ions can move freely across, causing a jump in current. When decreasing the current again, the pore starts in the 'wet' state, so the current will be higher at the same voltages compared to when it started in the dry state when increasing the current. Thus, the pore exhibits hysteresis, or 'memory'. However, the hysteresis observed here goes the opposite way, with a higher current when increasing the voltage, and lower when decreasing. If a bubble is causing the gating in GBSV1_Δ1-40_775, it would have to be forming at higher voltages, and dissipating at lower voltages. The electrostatic charge map of GBSV1_Δ1-40_775 shows that the lumen is very negative, making it hydrophilic. The MLP map of GBSV1_Δ1-40_775 also shows that the interior should be very hydrophilic (see Chapter 3). It is not expected for a nanobubble to form naturally.

5.4.4. Hybrid nanopores

5.4.4.1. Optimal SSN diameter

The G20c portal protein was inserted into an SSN with a diameter of 5.6-6 nm (Cressiot et al. 2018). This diameter is slightly smaller than the external diameter of the G20c clip domain. However, the SSNs used were made using TEM drilling. As such, the diameter of the SSN was measured immediately after fabrication and the diameter measured was the narrowest point of the pore. For the SSNs made by CDB, equation (15) was used to calculate the diameter based on conductance. The equation assumes that the nanopore is cylindrical, but real SSNs made by CDB are hourglass shaped. The equation may give the average size of the nanopore, rather than the narrowest diameter.

An SSN which gives an average diameter of 10 nm, may have a constriction point which is much narrower. This may be why the diameters for the SSNs for hybrid nanopores for GBSV1 are higher than those for G20c. Previously, cyclodextrins have been used to approximately size the hybrid nanopore. They could also be used to find the narrowest point of an SSN made using CDB, and the diameter of an eventual hybrid nanopore.

The conical shape of an SSN made by CDB should mean that, at the right shape, if the clip domain enters first, then the whole body of the portal protein should be able to enter the SSN without translocating through. However, the range of diameters which will allow this may be small, and equation (15) which is used for estimating the SSN diameter breaks down when the diameter is equal

to or greater than the pore length (Figure 5-23). This puts an effective upper limit on the SSN diameter in a given chip thickness.

5.4.4.2. Metastable baselines

Metastable baselines (also called 'gating') are a common problem in the SSN field. The causes are currently unknown. It may depend on each SSN. One possibility is the formation of a nanobubble on the side of the tunnel or inside the whole pore. This can occur if the tunnel, or areas of the tunnel are hydrophobic, e.g. if piranha cleaning is incomplete or less effective than expected. Dewetting is the process in which liquid exits the pore to leave a bubble. Smaller diameter pores can dry more easily. Electrowetting is voltage dependent: higher voltages promote pore wetting, while lower voltages can let a pore dry or allow a bubble to form (Powell et al. 2011; Aryal, Sansom, and Tucker 2015; Trick et al. 2017; Yazdani, Jia, and Chen 2020). If the SSN is not completely fabricated, some small amount of material may be left at the mouth of the pore, which may transiently block the pore, lowering the open pore current. One of the key selling points of biological nanopores (and the hybrid pores that use them) is that biological pores assume the same shape every time and cannot suffer from incomplete fabrication. Once analytes or portal proteins are added, there are other possible causes. Proteins adsorbing to the surface of the SSN tunnel without translocating can cause a shift in the baseline (Niedzwiecki, Grazul, and Movileanu 2010). In the case of hybrids, they may partially enter the pore, lowering the baseline, before fully inserting.

Once analytes are added to a gating pore, it may be difficult to distinguish a metastable baseline from a current blockade event. The chip may move between baselines randomly, so the duration is not certain. If the chip moves to a lower baseline for an extended period of time, and events occur, then it can easily be identified (e.g. the two baselines in Figure 5-17). However, if it only assumes the lower baseline for a short time or events do not occur during the duration, then it may be mislabelled as an event.

Gating behaviour is most easily spotted when the SSN has just been made and before any analytes have been added. Under these conditions, the only cause is the SSN itself. The Manual for the Spark-E2 recommends enlarging the pore slightly to fix these ('Northern Nanopore Instruments User Manual' 2022). For good-quality data collection, if the gating cannot be fixed by the application of conditioning buffer (Figure 5-9) or enlarging the pore, it should be discarded.

5.4.4.3. Possible sources of noise

The possible GBSV1 portal protein hybrid pores seen so far with GBSV1_Δ1-40_775 are very noisy, especially compared to hybrid pores made with the G20c portal protein. As seen in Figure 5-2, the portal protein is being pulled in opposite directions by the EOF and EP. This could cause the portal

protein to shift up and down whilst in the SSN causing noise, or allowing current to leak around the outside.

The final helix of the crown domain was found to be disordered in the cryoEM structure (see Chapter 3), so it was removed from the final model. However, it is still present in GBSV1_Δ1-40_775. In addition to causing the shift between open and closed conformations, this disordered helix may flap about around the opening to the pore, causing extra noise.

If the pore is too large or unevenly shaped, it is possible that the portal protein could shift and jiggle laterally inside the SSN. This would allow leakage current between the portal protein and the edges of the SSN. The movements may not be regular, so the noise would be non-uniform.

When a chamber is filled with electrolyte solution, it generates a charge difference across the SiN_x membrane. The capacitance of the membrane is constant. The large charge difference causes a high voltage across the membrane, which can lead to small defects on the membrane. These defects can be a source of noise in SSN (Matsui et al. 2015). Matsui *et al* avoided the generation of these defects by connecting the two chambers with wire when they were being filled with electrolyte solution. However, because the Spark-E2 uses a high voltage to generate solid-state nanopores, it is possible that these defects are made during fabrication.

5.4.4.4. Hybrid selection

The aim of THE DEEP ONES rule set was to select only those events that formed hybrid nanopores. The use of a rule set (Figure 5-4, THE DEEP ONES) meant that picking out events classified as hybrid pores was consistent across datasets. The formation of a hybrid pore was thought to block most of the current passing through the bare SSN; meanwhile bumping events or partial insertions would not block as much of the current. If used on smaller pores, this method does select events that block most of the pore (Figure 5-25). However, larger SSN with a diameter of ~9-10 nm or greater, display a crab-claw pattern in the $DG_{b\text{-}norm}$ vs T_t plot (Figure 5-25). The cluster with higher $DG_{b\text{-}norm}$ values (the 'top claw') may have $DG_{b\text{-}norm}$ values < 0.5, and so would be excluded from analysis using the DEEP ONES ruleset. However, the events in this 'top claw' may be caused by the formation of hybrid nanopores. Therefore, THE DEEP ONES ruleset may exclude genuine hybrid nanopores from analysis.

The 'top claw' could be extracted instead, but it can be difficult to identify the limits. Instead, it may be better to use a negative selection method, where all events are considered and bumping events are removed based on histograms, similar to the event selection process for standard nanopore analysis.

An example of all of the picked events below I_0-5σ for pore60_36nm is shown in Figure 5-41. The histogram for conductance noise shows two strong peaks: one at very low noise, and another at ~ 0.5

nS. The lower peak can be attributed to bumping events and can be filtered out (Oukhaled et al. 2012; Greive et al. 2024; Ratinho et al. 2024). The histogram of $DG_{b\text{-}norm}$ shows a strong peak at ~ 0.02 and smaller peaks at higher values. The lower peak of shallow events is not useful for optimising hybrid nanopore formation so should be removed (Figure 5-41B). The histograms of dwell time show a very high peak at short dwell times, then some smaller peaks at higher values. The peak for short events represents bumping events and should be discarded. Based on these histograms, the events to be picked should have a $DG_{b\text{-}norm} > 0.06$, $\sigma_G > 0.05$, and $T_t > 200$ µs.

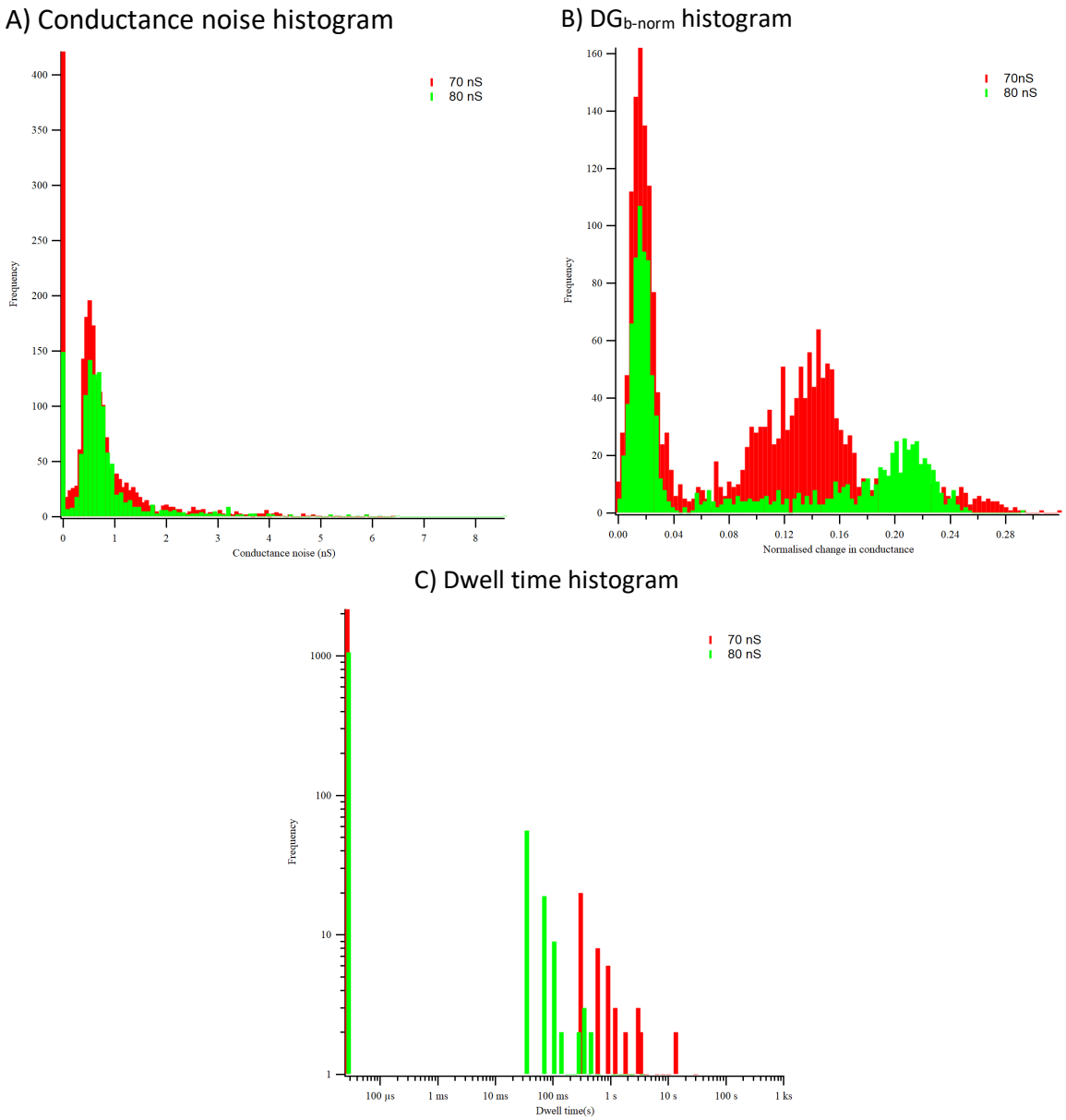


Figure 5-41: Histograms for pore60_36nm

The buffer is 10 mM Tris-HCl 0.4 M KCl pH 6.5. A) Conductance noise. B) Normalised change in conductance. C) Dwell time. Both axes are plotted on a log scale.

The clusters of data points selected by these filters are shown in Figure 5-42. The events in area 1 of Figure 5-42A are those with a dwell time < 200 µs ('bumping' events by dwell time). The events in area

3 of Figure 5-42A are those with a $DG_{b\text{-}norm}$ < 0.06. The highlighted area 2 in Figure 5-42A shows the selected data points to be taken forwards. Figure 5-42B shows conductance noise against dwell time. The dashed area shows the events selected based on conductance noise and dwell time. Due to the shape of the distribution, in this case, using either dwell time or conductance noise would have selected almost the same events.

A) $DG_{b\text{-}norm}$ vs T_t B) Conductance noise vs T_t

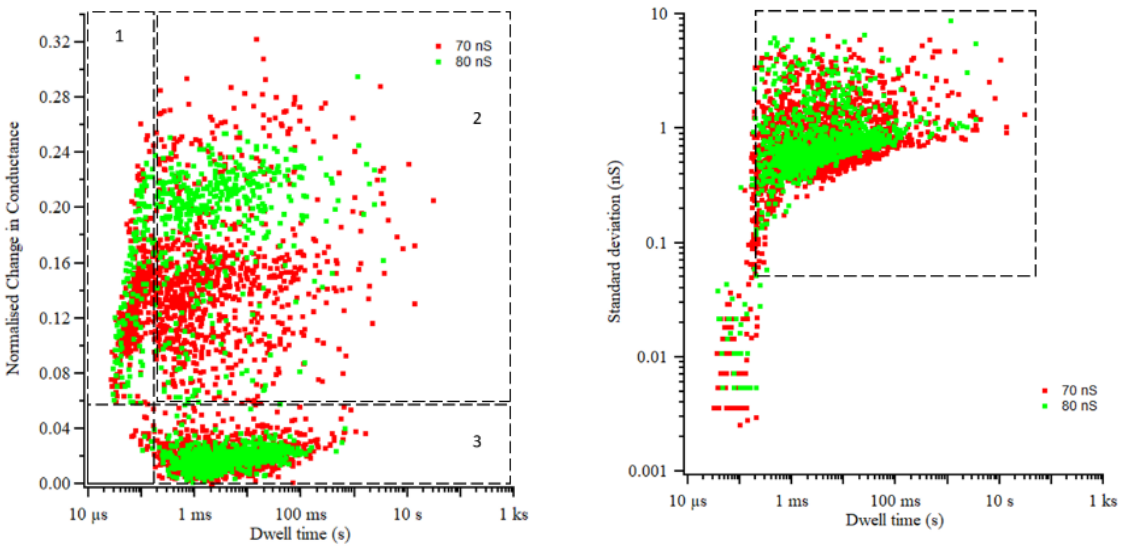


Figure 5-42: Normalised conductance change and conductance noise Vs dwell time for pore60_36nm

The dashed areas show different groupings of events. A) Normalised change in conductance Vs dwell time. Area 1 – bumping events, those with a dwell time < 200 µs. Area 2 – selected events for further analysis. Area 3 – shallow events with DGb-norm < 0.06. B) Conductance noise Vs dwell time. The dashed area shows events with dwell time > 200 µs and σG > 0.05.

5.4.4.5. Sub levels

The appearance of sub-levels in some of the traces could not be analysed due to time constraints and difficulties in identifying them. However, they may show some information about how the portal protein approaches and settles in the SSN before it forms a hybrid, or how it moves inside the SSN. The DEEP ONES identifies events in which the I_b falls below the threshold of I_0-5σ. Since the I_b is the average current of the event, it may obscure any sub-levels present. Analysis of sub-levels would require manually selecting these regions which would be time-consuming and prone to user error. An alternative would be to use software packages developed to identify these sub-levels, such as Nanolyzer (Northern Nanopore Instruments), but these are still in development. Any attempt at analysing sub-levels would require more data in optimal hybrid nanopore formation conditions.

5.4.4.6. Estimating hybrid size with CD

The diameter of the GBSV1 portal protein is estimated to be ~4 nm throughout. But when inserted into the SSN, the protein may be squashed and the lumen may be narrower. The diameter of the portal

protein in a hybrid nanopore could be tested with different cyclodextrins as for the G20c portal protein (Gidwani and Vyas 2015; Cressiot et al. 2017b). Each cyclodextrin has a different diameter (Li and Purdy, 1992). Observing CDs which pass through the pore can give a rough estimation of the pore's diameter. As previously mentioned, portal proteins can adopt different conformations depending on whether the viral DNA is moving into the capsid (clip to crown domain), or out (crown to clip domain). Thus, observing the force required to move the CD from the clip to crown domain and vice versa can also show the conformational state of the portal protein and if it is switching between them depending on the movement of the analyte. CDs can also be used to study the velocity of EOF in different conditions (L.-Q. Gu, Cheley, and Bayley 2003; Piguet et al. 2014; Bhamidimarri et al. 2016). In hybrid portal protein nanopores, they can be used to answer questions such as how does the EOF change between the open and closed conformations, or between different portal protein mutants, e.g. adding positive residues to the clip or crown domains.

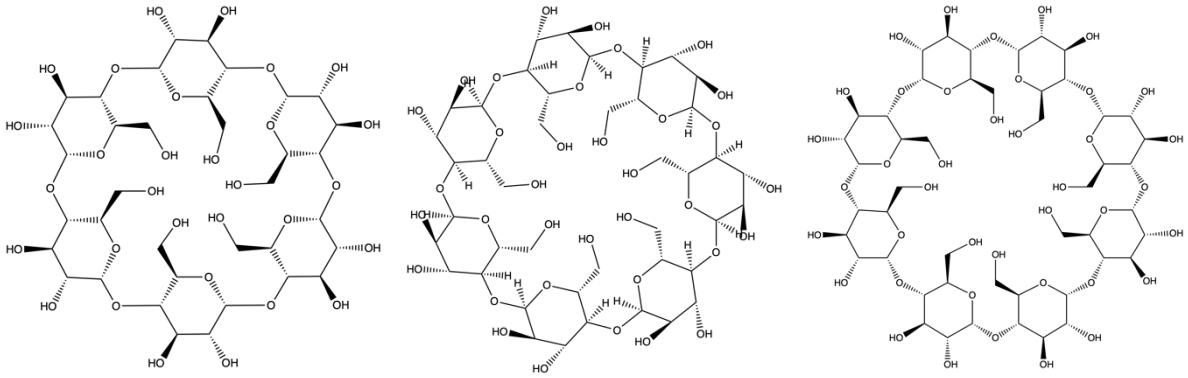


Figure 5-43: Structures of α-, β-, and γ-cyclodextrins.

α-cyclodextrins (CAS: 10016-20-3) has an outer diameter of 1.37 nm. β-cyclodextrins (CAS: 7585-39-9) has an outer diameter of 1.53 nm. γ-cyclodextrins (CAS: 17465-86-0) has an outer diameter of 1.69 nm (Song Li and Purdy 1992). They have previously been used to test the diameter of nanopores (Gidwani and Vyas 2015; Cressiot et al. 2017b).

5.4.4.7. Exit of the portal protein from the SSN

The life time of the hybrid nanopore is a very important parameter to optimise if they are to be useful in the future. A significant limitation of the analysis thus far is that it does not track if a potential hybrid left naturally or was ejected by the user. The ideal hybrid nanopore would remain in the SSN for long periods of time at different voltages. If it leaves naturally at a constant voltage, the voltage or SSN diameter may be unsuitable. The lifetime of a hybrid would be affected most by the voltage and SSN diameter. Events longer than ~1 s could be re-analysed to see if they left or were ejected. A set of conditions could have given long dwell times, but if they were truncated by the user, then that information would be lost and could give a misleading interpretation. The proportion of long events ejected could be plotted against the SSN diameter, or against voltage for the same pore.

In the present analysis, the start and end points of an event, which are used to calculate the dwell time are the points at which a current trace cross the line I_0-3σ. However, an alternative method for calculating the start and end points would be to use a point between the I_0-3σ threshold and I_0-5σ (Zhen Gu et al. 2015).

5.4.5. Portal protein components

5.4.5.1. Open and closed conformations

The tunnel of the G20c portal protein used in previous hybrid nanopores is largely positively charged. This means that the EOF flows in the opposite direction to that in the GSBV1 portal protein: from the ground electrode to the working electrode (Figure 5-44). Negatively charged molecules, e.g. DNA, were pulled by EP and EOF towards the working electrode. If the G20c portal protein inserts from the trans side with the working electrode, then negatively charged molecules will move through the portal protein from the crown domain to the clip domain (Cressiot et al. 2018).

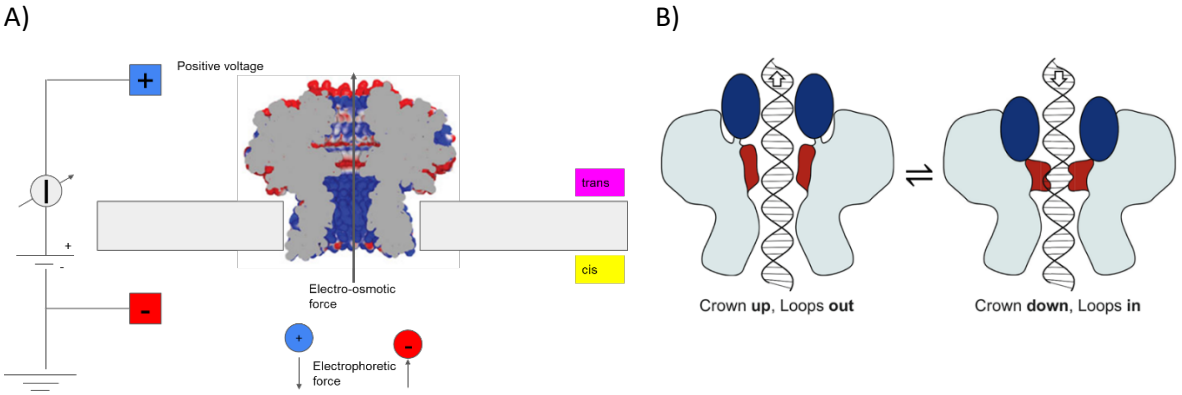


Figure 5-44: Change in tunnel diameter of the G20c portal protein depending on location of the crown domains

a) Schematic of the G20c D/N portal protein mutant as a hybrid nanopore, as made by Cressiot et al 2018. The interior of the G20c mutant is positive, so the EOF flows in the opposite direction to that in GBSV1_Δ1-40_775 portal protein. This means that negatively charged molecules, such as DNA, move with both forces from the clip to the crown domain. Adapted from (Cressiot et al. 2018). b) The left picture shows DNA moving from the clip to the crown domains, as if entering the procapsid. For the GBSV1_Δ1-40_775 hybrid nanopores, this would be like DNA moving from the trans to the cis chambers towards the cathode driven by EOF. The right picture shows the opposite, DNA moving from the crown to the clip domains. In GBSV1_Δ1-40_775 hybrid nanopores, this would be moving to the trans chamber under electrophoresis. The DNA catches the crown domains which push the tunnel loops inwards, narrowing the diameter from 2.2 nm to 1.4 nm, and changing the interior from hydrophilic to hydrophobic. Adapted from (Bayfield, Steven, and Antson 2020).

Comparison of the crystal and cryoEM structures of the G20c protein showed two different conformations with different tunnel diameters (Bayfield, Steven, and Antson 2020). In virus procapsids, the switch from the open to closed conformation, with a narrower tunnel, is thought to be triggered by DNA moving backwards through the portal protein and pushing on the crown domains. This causes a conformational change which pushes the tunnel loops inwards, constricting the pore and holding the DNA in place. Thus, DNA can freely move through the portal in the direction of clip domain to crown domain, but not vis versa.

Similarly, when the G20c portal protein was adapted as a biological nanopore, β-CD was added to either compartment so that it translocated from the clip to the crown domain or vis versa (Cressiot et al. 2017b). More force was needed to move the β-CD from the crown domain towards the clip domain than from the clip domain towards the crown domain. The direction-dependent shift between open and closed states will not only change the size of the pore, but also the EOF, as narrower pores give a higher EOF velocity (Melnikov, Hulings, and Gracheva 2017). If the pore is narrower in one direction than the other, then that will also exhibit current rectification (see Chapter 1). If the EOF and EP vectors oppose each other, and the EOF is stronger in one direction, then a greater EP velocity will be needed in that direction to move analytes into the nanopore. If the open and closed states of the portal protein could be stabilised, then they could potentially be used as separate nanopores, depending on whether the analyte enters from the crown domain (closed conformation) or the clip domain (open conformation).

5.4.5.2. Portal charge and orientation

The GBSV1 portal protein is not a perfect candidate to form a hybrid nanopore. It has several differences to the G20c portal protein (Cressiot et al. 2018). The dipole of analytes can strongly affect the types of events (Stefureac et al. 2006). Its dipole points towards the clip domain (Figure 5-11) which would, under a positive voltage, position the clip domain towards the SSN from the ground electrode (*cis* compartment). However, its electrostatic surface potential is overwhelmingly negative at the pH values used (see Chapter 3). This could prevent the portal protein orientating in the electric field so that the clip domain enters the SSN first. As mentioned previously, because the GBSV1 portal protein is negatively charged, the EP and EOF are also acting on the GBSV1 portal protein to pull it in opposite directions. The negative charge of the portal tunnel may also limit the analytes that can be used. The electric field lines of GBSV1_Δ1-40_775 show two bands of positive charge at the mouths of the pore. Negatively charged DNA may be able to enter, but it may have difficulty with the negatively charged tunnel. Meanwhile, positively charged analytes may form electrostatic interactions with the tunnel wall and clog the pore.

There are different methods of increasing the strength of the dipole or changing its direction. The pH of the buffer can affect the protein's charge and dipole, depending on the pKa values of the component residues. As shown in Chapter 4, at pH 6 the tunnel and outside of the stem domain of GBSV1_Δ1-40_775 are positively charged. In this case, the Stern layer of the EDL inside the portal protein tunnel would be comprised of cations. If GBSV1_Δ1-40_775 formed a hybrid nanopore inserting from the *cis* compartment, then under a positive voltage, the EOF would flow towards the working electrode (*trans* compartment). If negatively charged analytes were added to the *cis* compartment, the EP and EOF would both move them towards the *trans* compartment.

Positive or negative residues can be added to the poles of the protein. The G20c portal protein has natural dipole towards the crown domain. When added from the working electrode (*trans* compartment), it will orientate with the clip domain towards the ground electrode (*cis* compartment) and the SSN (Cressiot et al. 2018). The mutations to the G20c portal protein also helped it orientate itself in the electric field to be inserted clip domain first. Two of the GBSV1 portal protein mutants which could not be purified in time to be tested as nanopores, GBSV1_Δ1-40_D213N_E247R_E250Q_E252Q_774 and GBSV1_Δ1-40_E218R_E222R_E223R_E237Q_822 (see Chapter 2) have positive charges on the inside and outside of the clip domain, respectively. These may orientate the portal protein in the electric field to force it to insert with the clip domain first. This would be useful if the portal protein inserted with EOF, with the clip domain pointing towards the ground electrode (*cis* compartment). The pI of these mutants is still low, so at the pH values tested, they would be inserting with EOF rather than EP. Changing the charge of the tunnel wall may also help analytes of different charges enter the pore.

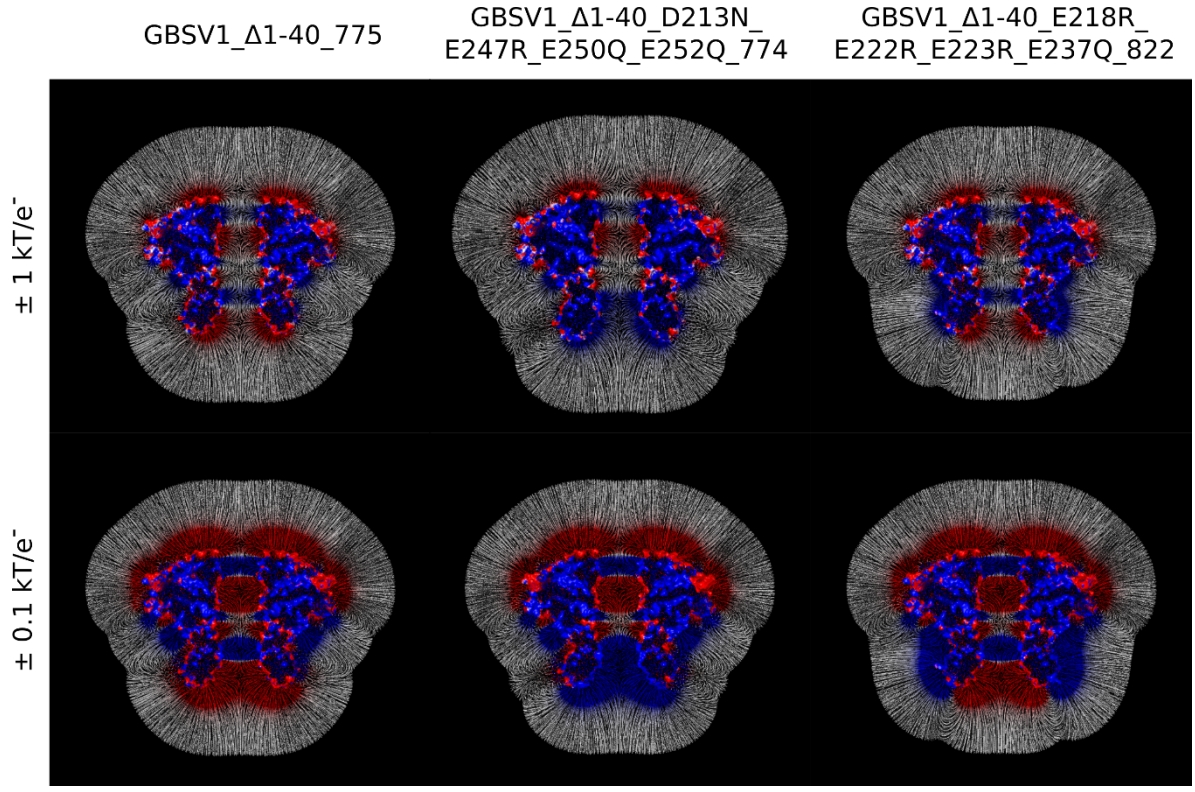


Figure 5-45: Electric fields in different mutants

The mutants were made by making the appropriate mutations in the GBSV1_Δ1-40_775 structure. The PQR files, APBS calculations and electric fields were made and visualised as before (see Chapter 3). GBSV1_Δ1-40_775 has two bands of positive electric field within the lumen: at the mouth by the crown domains, and by the constriction point in the clip domains. The GBSV1 PP Δ1-40 D213N E247R E250Q E252Q mutant had mutations targeting the interior surface of the clip domain. The positively charged mutations at the clip domain make a positive electric field which joins with that at the clip domain constriction point. This gives the protein a clear dipole along the axis of the tunnel. The GBSV1 PP Δ1-40 E218R E222R E223R E237Q mutant had mutations targeting the exterior of the clip domain. This mutant remains the negative area at the interior of the tunnel at the clip domains. However, the positive mutations at the clip domain exterior join with the small field just below the wing domain to form a positive band around the outside along the stem and clip domains.

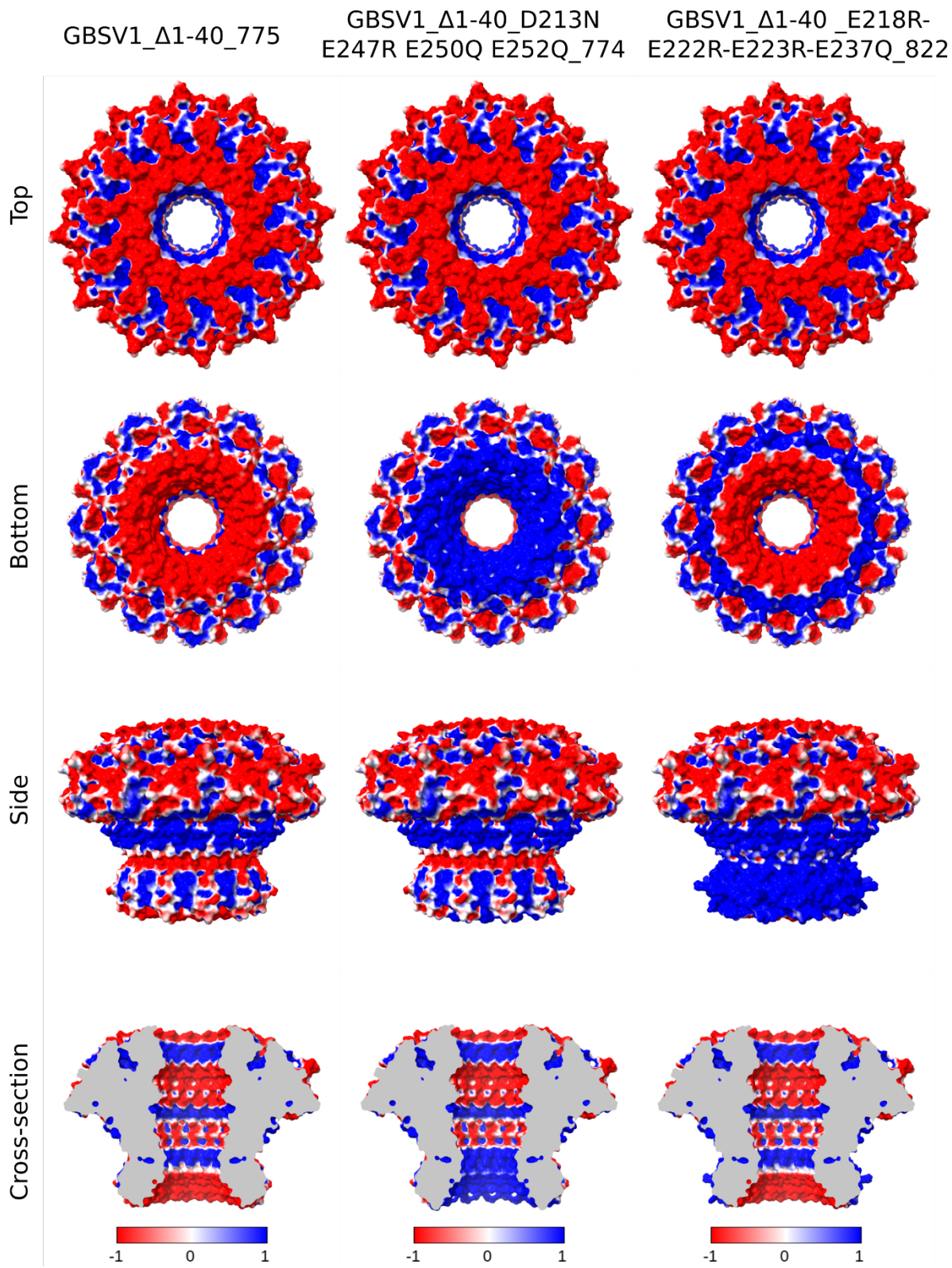


Figure 5-46: Electrostatic surface of GBSV1Δ1-40_775 and the clip domain mutants.

The APBS was calculated at 10mM Tris-HCl pH 6.5, 0.4 M KCl. There are clear differences in surface charge for the different mutants. As in the purification conditions, there are three positive bands within the tunnel. In the GBSV1 PP Δ1-40 D213N E247R E250Q E252Q mutant, the mutations at the interior of the clip domain join with the lower band to make a single large positive area. In the GBSV1_Δ1-40 E218R E222R E223R E237Q mutant, the interior is the same as the GBSV1 PP Δ1-40 mutant, but the exterior of the stem and clip domains are completely positively charged.

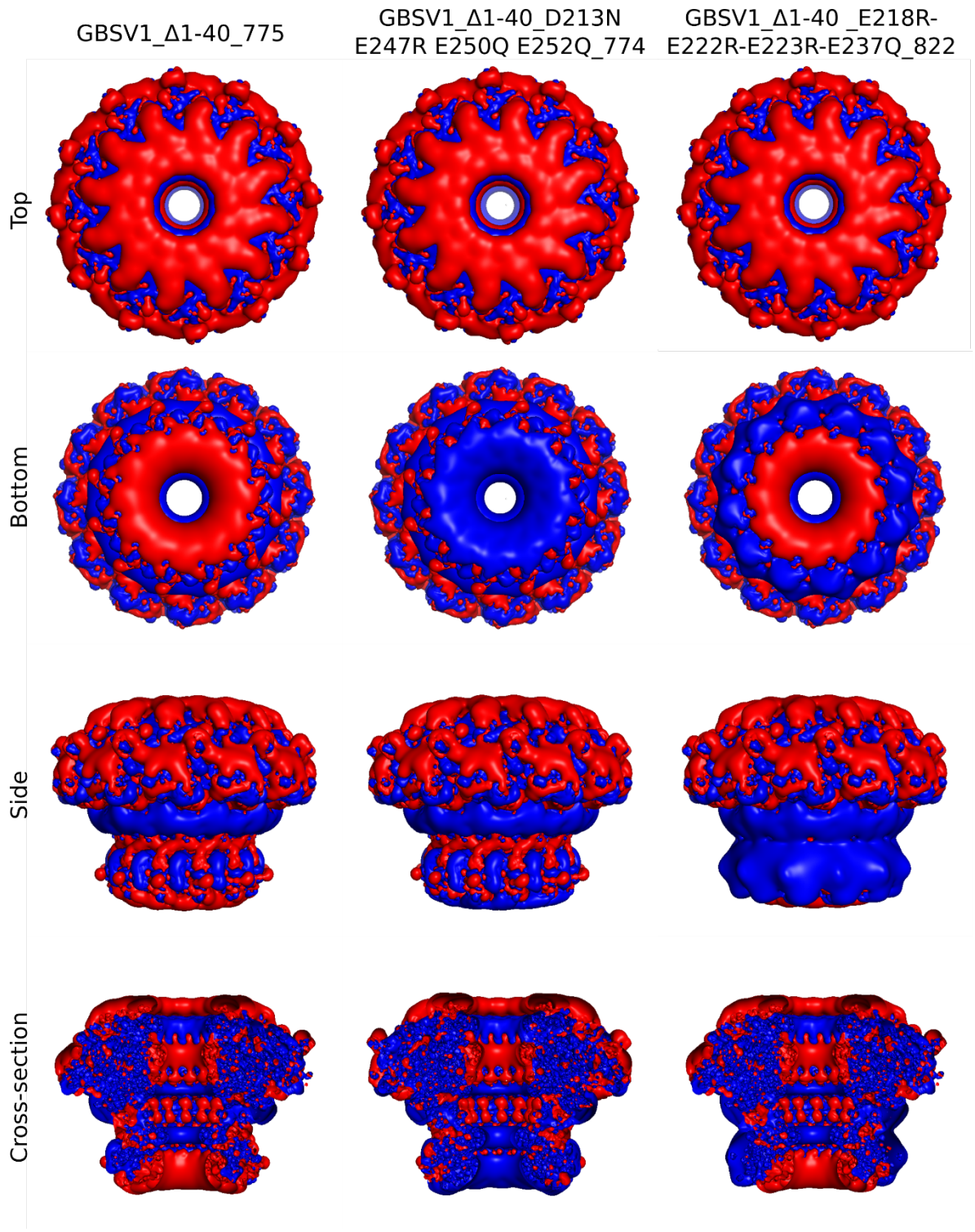


Figure 5-47: Isoelectric surface of GBSV1_Δ1-40_775 and the clip domain mutants.

The isoelectric surface for the crown and wing domains are the same between different mutants. The scale was ± 1kT/e-. Images were taken in PyMol.

One of the GBSV1 mutants (GBSV1_Δ1-40_A215C_776, see Chapter 2) which was not purified in time to be tested as a nanopore was to have a length of dsDNA attached to its clip domain in the same manner as was done with αHL (Hall et al. 2010a). The dsDNA would be pulled through the SSN first due to its negative charge, which would pull the portal into the SSN clip domain first, ensuring correct

insertion orientation and giving a clearing signal for hybrid nanopore formation. Another method would be to functionalise the surface of the SSN tunnel and immobilise the portal protein once inside. This should ensure long-lived hybrids of the right orientation (Mojtabavi et al. 2022). In this scenario, the portal protein must still enter in the correct orientation to be secured.

5.4.6. Comparisons with other hybrids.

In the previous example of hybrid nanopores using the G20c portal protein, the SSNs were drilled using a TEM (Cressiot et al. 2018). SSNs made using a TEM have an hourglass topology inside the tunnel. The narrowest diameter of the tunnel can be measured at the end of fabrication. However, this is not possible for SSNs made by CDB. Because the diameter is calculated by looking at the conductance at different voltages through the pore (see equation (15)), the narrowest diameter of the pore is unknown. The previous hybrid nanopores used SSNs with a diameter of 5.6 - 6 nm, but because of the difference in fabrication and sizing methods, these diameters may not apply to the hybrids made using the GBSV1 portal protein. SSNs made using a TEM can be made symmetrical by the user during enlargement, but SSNs made using CDB may not have symmetrical topology radially or between the top and bottom. Larger SSN tended to give less noisy IV curves. This may have been from the longer conditioning time required to enlarge the pore.

5.5. Summary

In conclusion, preliminary tests of the GBSV1 portal protein as a biological and hybrid nanopore have been attempted, but the optimal conditions for the formation of hybrid nanopores have not been found. The GSBV1 portal protein could not be used as a biological nanopore because it could not be inserted into a lipid membrane. The exact reason for this is unclear. Due to the nature of fabrication, the exact topology and diameter of the SSN cannot be certain, either just after fabrication, or over the course of an experiment. While the portal protein has a dipole in a useful direction, the near-universal negatively charged surface acts against its orientated insertion into the SSN. The Spark-E2 cannot fabricate pores in 30 nm chips, which prevents the replication of the conditions used for the G20c hybrid pores. The size measurement of the SSN breaks down when the diameter is equal to or greater than the pore length, which puts an effective limit of ~ 12 nm on SSN diameters. More tests are required at the same SSN diameter in order to give precision and confidence to the present measurements and conclusions.

Chapter 6 Conclusion

Portal proteins and nanopores have each been studied for a long time in distinct areas of research. Portal proteins have been studied primarily as an essential part of the DNA packaging machinery of dsDNA bacteriophages. Their role in this process has evolved over time, from initially being considered actively packaging the DNA (Lebedev et al. 2007), to their current understanding as molecular valves that prevent the dsDNA leaving the capsid (Bayfield, Steven, and Antson 2020). In parallel, nanopores have been studied for use in the next generation DNA sequencing and biosensing applications, leading to the development of biological nanopores, SSNs and, more recently, hybrid nanopores. This work explored the potential of thermophage portal proteins as novel nanopores. To this end, this thesis presents three-dimensional structure of a previously uncharacterised portal protein from the thermophage GBSV1, determined at 2.7 Å resolution using cryoEM. Additionally, significant insights were gained into SSN fabrication using CDB methods.

Until now, structural data on thermophage portal proteins were limited to those from the G20c and P23-45 bacteriophages, which share 99.3% sequence identity (Bayfield, Steven, and Antson 2020). The GBSV1 portal protein, characterised in this study, exhibits only 21 % sequence identity with the G20c portal protein (Madeira et al. 2024). Despite the very different sequence, it retains the canonical portal protein fold, comprising the four structurally conserved domains: clip, stem, wing and crown (see Chapter 3).

For production of a soluble protein, the N-terminal tail of the GBSV1 portal protein had to be removed . Likewise, production of soluble portal proteins of G20c (L. S. Williams et al. 2013) and T4 (Sun et al. 2015b) bacteriophages also required removal of the N-terminal segments. In other portal proteins, the portal protein was produced with an intact N-terminal, but it was too disordered to be visualised, e.g. in the case of bacteriophages P23-45 (Bayfield et al. 2019), φBB1 (Rūmnieks, Füzik, and Tārs 2023), phi29 (Guasch et al. 2002). The N-terminal tails of portal proteins have been hypothesised to interact with the scaffold proteins (Bayfield et al. 2019) and have a role in capsid assembly (Padilla-Sanchez et al. 2014). This fits with their suggested location at the outer surface of the portal protein wing domains where they are exposed to the scaffold and capsid proteins. It is likely that the N-terminal region of the GBSV1 portal protein plays a similar role in interacting with other proteins. Further work elucidating the structure of intermediate assembly states of the GBSV1 virus, or co-expressing the portal protein and scaffold proteins may answer this question with certainty.

The tunnel loops of the GBSV1 portal protein are observed in an open conformation, in which they are pulled back from the centre of the tunnel, as seen in bacteriophages T4 and T7 and G20c/P23-45. The

density for the tunnel loops was well defined, which indicates a lack of disorder, in contrast to the cyroEM structures of several other portal proteins, e.g. T4 (Sun et al. 2015b), SPP1 (Orlov et al. 2022), T7 (Cuervo et al. 2019) and the Epstein-Barr virus (Machón et al. 2019). In some cases, the tunnel loops were too disordered to be modelled at all, e.g. phi29 (Simpson et al. 2001), phiCPV4 and GA1 (Javed et al. 2021). The environment of the portal protein likely influences the conformation of its tunnel loops. For example, the tunnel loops of the T4 portal protein were disordered in isolated protein, but were well-ordered in the procapsid (Sun et al. 2015b)(Fang et al. 2020). The β-loop of HSV1 portal protein is well ordered in mature virions (Y.-T. Liu et al. 2019). Determining the structure of the GBSV1 portal protein in procapsid and mature virus forms would clarify the influence of the maturation state on tunnel loop conformation, further clarifying their open conformation observed in this study (Chapter 3).

Charge distribution analysis revealed key differences between the GBSV1 and G20c portal proteins (Cressiot et al. 2018). The GBSV1 tunnel is largely negatively charged, exhibits three bands of positive charge at the clip domain (K255, K260), just below the tunnel loops (R273), and at the crown domain (R348, R352). In contrast, a portion of the G20c portal protein tunnel formed by the stem and clip domains is negatively charged while the rest of the tunnel has a varied charge distribution, with more positive residues present towards the top of the crown domain.

Thermal and chemical stability studies of thermophage portal proteins remain limited. NanoDSF analysis of the GBSV1 portal protein revealed a single thermal unfolding peak, unlike the two peaks observed for the G20c portal protein (Cressiot et al. 2018). This suggests that the GBSV1 portal protein may dissociate without losing subunit integrity. Similar patterns were seen for aggregation as the temperature increased. The trace for G20c showed two sometimes overlapping peaks, but the GBSV1 portal protein showed only one.

Hybrid nanopores are a relatively new type of nanopore that aim~~s~~ to combine the best qualities of both biological and solid-state nanopores. Different combinations of biological and solid-state nanopore have been tested including protein pores (Hall et al. 2010a; Cabello-Aguilar et al. 2013; Göpfrich et al. 2013; Cressiot et al. 2018; Bentin, Balme, and Picaud 2020; Mojtabavi et al. 2022) and DNA pores (Engst et al. 2012; R. Wei et al. 2012), but there has been limited success so far.

The standard method of making SSNs is to drill them with a TEM (Hall et al. 2010a; Cressiot et al. 2018). But the alternative method of CDB has been used by other labs to make SSN without a TEM. Most groups use their own custom build systems to make a pore by applying the necessary voltage (Briggs, Kwok, and Tabard-Cossa 2014; Kwok, Briggs, and Tabard-Cossa 2014; Briggs et al. 2015; Matsui et al.

2015; Pud et al. 2015; Goto et al. 2016; 2019; Yanagi et al. 2017; 2018; Roshan, Tang, and Guan 2019; Yanagi, Akahori, and Takeda 2019; 2023; D Y Bandara et al. 2020; Waugh et al. 2020; Yanagi and Takeda 2020; Saharia, Bandara, Karawdeniya, Alexandrakis, et al. 2021; Kasumi Tsustumi 2023; O'Donohue et al. 2023; 2024), but some are using the Spark-E2 device, developed by Northern Nanopore Instruments (Fujinami Tanimoto et al. 2022; C. Ying et al. 2022; He et al. 2023; Alam et al. 2024; Bouhamidi et al. 2024; Elliott et al. 2024; Järlebark 2024).

This study represents the first attempt to use CDB for hybrid nanopore formation. There is a window of diameter sizes which are suitable for hybrid nanopore formation. While instances of successful hybrid nanopore formation were observed (Chapter 5), further optimisation of SSN fabrication and condition is required in order to make SSN with reliable diameters. This study showed that conditioning SSNs in buffer can help reduce current noise and stabilize the pores. Due to the low salt concentration, the SSN diameter cannot be accurately estimated by measuring its conductance. However, as shown in this thesis, plotting baseline conductance vs SSN size may be useful for estimating SSNs. Further work could refine this relationship to improve accuracy.

Appendix 1: Generation of portal protein mutants

Oligonucleotide primers

Primer name	Sequence (5'-3')
pYM715_F20	TCCAGGGACCAGCAGGCTCGACCCAGCAC
pYM715_R446	TGAGGAGAAGGCGCGTCAGTCCCGTACCGCCC
pYM775_F(Δ383-406)	AACTACATTTAGCTCGATAAAATTGGCGAT
pym775_F[A215C]	GTTGACGCGTGCACGGCAGAGCTTTCGAGT
pYM775_F[D213N]	GTGAAGGTTAACGCGGCCACGGCAGAGCTT
pYM775_F[E237Q]	CAAGCCACGCAGGCGGGGCAGCCTTGGATC
pYM775_F[P282Q]	TTTGGTGTGCAGGCTTTTTTATTGGGC
pYM775_F[S135C]	CCGTTAACGCCGTGCAAGGTAAACTTCTTG
pYM775_F40	TCCAGGGACCAGCAAGATTGAGCGATAACCCGGAAGTG
pYM775_R(Δ383-406)	TTTATCGAGCTAAATGTAGTTTTCTAGGAT
pym775_R[A215C]	CTCTGCCGTGCACGCGTCAACCTTCACGAT
pYM775_R[D213N]	CGTGGCCGCGTTAACCTTCACGATGAGCGA
pYM775_R[E237Q]	CTGCCCCGCCTGCGTGGCTTGAAGGTATTT
pYM775_R[P282Q]	TAAAAAAGCCTGCACACCAAACATGCCAG
pYM775_R[S135C]	CAAGAAGTTTACCTTGCACGGCGTTAACGG
pYM775_R406	TGAGGAGAAGGCGCGCTAGTCTGTTTGACCGTCCGC
pYM777_F44	TTCCAGGGACCAGCAGAGTTTGAGGAGATGCTGCAGG
pYM777_R445	TGAGGAGAAGGCGCGTCATCTCCTGCTCCTGTACGC
pYM792_F(Δ383-406)	AACTACATTTAGCTCGATAAAATTGGCGAT
pYM792_R(Δ383-406)	TTTATCGAGCTAAATGTAGTTTTCTAGGAT
pYM817_F[E218R]	GCCACGGCAAGGCTTTCGAGTGAAGAAGGC
pYM817_R[E218R]	ACTCGAAAGCCTTGCCGTGGCCGCGTCAAC
pYM818_F[E247R]	ATCCCGGCCAGGCTGTTAGAGGTGGAACAA
pYM818_R[E247R]	CTCTAACAGCCTGGCCGGGATGATCCAAGG
pYM819_F[E222Q,E223R]	GCCACGGCAAGGCTTTCGAGTCAAAGAGGCCGGAATGCAGTGTTC
pYM819_R[E222Q,E223R]	ATTCCGGCCTCTTTGACTCGAAAGCCTTGCCGTGGCCGCGTCAAC
pYM823_F[E250Q,E252Q]	ATCCCGGCCAGGCTGTTACAGGTGCAACAAGTCAAACCATTATCC
pYM823_R[E250Q,E252Q]	TTTGACTTGTTGCACCTGTAACAGCCTGGCCGGGATGATCCAAGG

F – forwards primer, R – reverse primer; () – deletion, truncation; [] – mutation

Gene sequence for the YSBL3CLic+ plasmid

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGA
CCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTC
GCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCA
CCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTC
AACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAA
TGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCG
CTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGAT
TATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCAT
AGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTT
CCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATG
GCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCA
CTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTT
AAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATAT
TTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGT
AACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCA
GTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACT
CTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCC
CATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCG
TTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTC
TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG
TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATA
CCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTAC
ATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGT
TGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAG
CCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCAC
GCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGA
GGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTT
ACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAG
TCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTC
ACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACT
CCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTG
ACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTC
AGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGA
AGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGT
CTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGT
AAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTT
ACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCG
GGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGG
GTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCC
AGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCA
GCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGC
CTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATG
GCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGAT
TCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGA
CCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACG
ATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGA
TCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGG

GAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCT
GAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTT
AACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAAC
GCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCG
CAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCG
CCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGAT
GCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAG
ACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAG
CGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTT
CGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATC
GCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTG
TTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTT
TTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACA
CCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTC
CGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCT
CCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGC
AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCA
CGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATA
TAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGA
GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCT
AGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATGGGCAGCAGCCATCATCATCATCATCACA
GCAGCGGCCTGGAAGTTCTGTTCCAGGGACCAGCAAGGCGCGCCTTCTCCTCACATATGGCTAGCATG
ACTGGTGGACAGCAAATGGGTCGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCG
AGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCT
GCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCT
GAAAGGAGGAACTATATCCGGAT

Translation start and N-terminal 6His tag

The coding sequence will be inserted in place of AGG

3C cleavage site

Sequences for the genes to be cloned straight into lic+

BV1 gene with ends for insertion into lic+ vector

TTCCAGGGACCAGCATTGGGATTGTTTAATTTTTTTCGGCGCAAAACACGTTCAGAACCAACGAATGC
AATAAGCTGGTTTCTGACACAAGAAGCATATGATACGCTTGCTATCCCTGGATACACTAGGTTGTCCG
ATAACCCGGAAGTAAGGATGGCCGTGCATAAAATTGCTGAACTCATATCGTCTATGACCATACACTTG
ATGCAGAATACGGATAACGGCGATATTAGGATTAAAAATGAGCTGTCGAGGAAAATAGACATCAATCC
ATATAGCCTAATGACAAGGAAAGCGTGGATGTACAATATTGTTTACACCATGCTACTCGATGGTGAGG
GAAATAGCGTCGTTTTTCCGAAATACACGACAAGCGGGTTAATCGATGAACTTATACCACTTGCTCCA
TCCAAGGTAAGTTTTGTCGATACCGATACAGGCTATCAAATTTGGTATCAAGGAAAAGCGTACAATTA
TGACGAGGTGCTGCACTTTATTGTGAACCCAGACCCAGAAAAGCCATATATGGGTAGGGGTTATAGGG
TAGTACTAAAAGACATCGTAAATAACTTGAAACAAGCAACAACAACGAAGAAAAGCTTCATGAGTGGT
AAATATATGCCTTCACTCATTGTGAAAGTCGACGCGGCTACCGCAGAGCTTTCCAGTGAAGAAGGCAG
GAACGCGGTATTTAAAAAATATCTCGAAGCAAGCGAGGCAGGACAACCATGGATTATTCCTGCTGAAT
TGCTTGACGTCGAACAGGTAAAGCCATTGTCACTTAAAGATTTAGCGATTCATGAAACAGTTGAATTA
GATAAGCGAACAGTAGCTGGCATCTTCGGAGTGCCGGCTTTTTTGTTGGGTGTCGGTAAATATGACAA
AGACGAGTATAACAACTTCATTAACTCAACTATATTGCCGATTGCCAAGGGGATTGAACAGGAACTGA
CAAGGAAACTCCTTATTAGCCCTGATTTGTATTTCAAGTTCAACCCGCGAAGCCTTTATGCATACGAC

CTGAAAGAGCTTGCGGAAGTTGGTTCCAACATGTATGTGCGTGGGCTTATGGAAGGTAACGAGGTAAG
GGATTGGCTTGGATTGTCGCCAAAAGAAGGGTTGAGTGAGCTAGTCATCTTAGAAAACTACATTCCGT
TAGACAAAATTGGCGACCAAAACAAACTGAAAGGTGGTGAGAAAGGTGGAGCGGACGGTCAAACAGAC
TAGCGCGCCTTCTCCTCA

Yellow - BV1 gp18 coding sequence

GBSV1 gp18 gene with InFusion cloning ends

TTCCAGGGACCAGCAATGGGACTATTCGATAGGTGGAGGCGAACGAAGCGTAAAAGTAAAATTAGAGC
GGACACAGGGTATGTAGGTCTATTCATGAGTGGTGAAGATGTATCCTTTCTTGTTCCCGGATACGTAA
GATTGAGCGATAACCCGGAAGTGCGCATGGCGGTGCATAAGATTGCTGATCTCATCTCGTCGATGACA
ATTTATCTCATGCAGAATACAGAGGACGGGGATATCCGCATCCGTAATGAGCTGTCGCGCAAAATTGA
TATCACTCCATATTCGCTCATGACAAGAAAGTCATGGATGTACAACATTGTGTATACGATGCTGTTAG
ATGGTGAAGGAAATAGTGTGGTGTTTCCGAAGTATACAGCGGATGGGCTGATTGATGAGCTGGTACCG
TTAACGCCGTCGAAGGTAAACTTCTTGGACACACCGGATGGATACCAGGTTTTATACGGCGGGCAGAC
ATTCAACTATGATGAGGTTTTACACTTCATCTACAATCCGGACCCGGAACGCCCGTACATCGGACGAG
GGTATCGGGTGGTATTGAAGGATATTGCGGATAACCTAAAGCAAGCAACAGCGACAAAGAAAAGTTTT
ATGAGTGGGAAATACATGCCTTCGCTCATCGTGAAGGTTGACGCGGCCACGGCAGAGCTTTCGAGTGA
AGAAGGCCGGAATGCAGTGTTCAAAAAATACCTTCAAGCCACGGAGGCGGGGCAGCCTTGGATCATCC
CGGCCGAGCTGTTAGAGGTGGAACAAGTCAAACCATTATCCCTTAAAGATATTGCGATCAACGAGGCG
GTCGAACTGGATAAGCGAACCGTAGCTGGCATGTTTGGTGTGCCGGCTTTTTTATTGGGCATCGGGGA
GTTCAACCGGGATGAGTACAACAACTTTATCAATTCGACTATCTTGCCGATTGCAAAAGGAATCGAGC
AAGAATTGACGAGAAAGCTGCTTATCAGCCCGGATCTTTATTTCAAGTTCAATCCACGAAGCCTATAC
GCTTATGACCTTAAAGAGCTGGCGGAAGTCGGTTCAAATATGTATGTCCGTGGCATTATGGAGGGAAA
TGAGGTCCGCGATTGGCTTGGGCTTTCGCCGAAAGAAGGATTGAGTGAGCTAGTTATCCTAGAAAACT
ACATTCCTCTCGATAAAATTGGCGATCAAAGCAAACTGAAAGGTGGTGATAATAGTGGAGCGGACGGT
CAAACAGACTAGCGCGCCTTCTCCTCA

RM378 gp20 gene with InFusion cloning ends

TTCCAGGGACCAGCAATGCCGTCTCTGGAAAAATATAAAAAACTGAACGAAGCTGTCAACTTCACCAA
TTTCCTTTCTCCGATGTATGGTATGGGAGCGCCGCACGGTGCGGGTGGCTCTTCCATGATTCCGATTA
ACATGTACCACCCGTTTGCGACAGCGGGGTATGCGAGCAGGTTCTACGGAGGAATAGAATTCAACCGG
TTTTTCCTCTACGATATGTACGATCGCATGGATTATACCGATCCGCTTATTTCCACGGTGCTGGATAT
CATTGCAGATGAATGCACGATTCCCAACGAAAACGGTAATATTGTGGATGTGGTAACAAAAGACATTG
AGCTTGCAAAAGCAATACTCTCCTATCTGGATTATGTTATAAATATTGAAAAGAACGCCTATCCTATC
ATTCGCAACATGATCAAATACGGAGATATGTTTCTTCATATTCTTGAAAAAGGTTCGGATGGCACGAT
CGAAAAATTTCAGGTGGTCTCTCCTTACATTTTCTCAAAGCGATATAACCCCGAAACGGATACCTGGT
ATTATGTTATTACCGACGTTTACCGTAACGTTGTAAGCGGATACTTTAACGAAGATATTCCTGAGGAA
GATGTCATTCACTTTTCTCACAAAATCGACACCAACTTCTTCCCCTATGGTAGAAGCTACCTTGAAAG
TGCTCGAGCGATATGGAACCAGCTTCGACTCATGGAAGATGCGCTCATGCTTTACCGTGTGGTTCGTA
GTGTCGATAGACGGGTGTTTTACGTAGATGTCGGAAACGTGCCGCCCGACAAGATCAACGAATATCTG
ACCAATATTGCCATGCAGTATAAGCGAGACTATTGGGTAAGAAATAATCAGAATCAATTTCTCGGAAT
CGATAACTACTTTTCTATCGAAAGCATTCTCAAAGATTACTTCATTCCCCGGCGCGGAGATCGTCGTG
CGGTGGAAATCGATATTCTTCAGGGTTCCAAAGTAGATCTTGCTGAAGATGTGGAATACATGCTCAAC
CGACTCATCTCCGCACTCAAGGTACCAAAGGCGTTCATTGGTTATGAAGGAGACGTGAATGCTAAAAA
CACACTGGCTACTCAGGATATCAAATTCAACAACACGATCAAACGCATTCAGGGGTTCTTTGTGGAAG
AACTCGAACGCATGGTGCGCATGAACAAAGAGTTTGCCGATCAGGATTTCCGCCTGGTAATGAACCGC
TCGAATTCGATTGTGGAAGGTGAGCGCTTTGCCGTGATTGAGCAGCGAATCGGTATTGCCGAACGCCT
CAAAGGCTGGGTTCGTGAAGACTGGATTTATAGTAACATTCTTCAGATTCCCTACGATCTCAAGCCCC
AGGAGGAAGTGGCGGAAGCAGCCGGTGGCGGCGGTCTTTTCGACACGGGCGGATTCGGTGAGGAAACA

ACCCCCGCCGACTTCCTTGGCGAGCGCGGCTCCCCGATCGAATCGCCAAGAGGAAGAACGGAATTCGA
TTTCGGAACGGAAGGCGGAGAAGAACTGGGCGGCGAACTTAATCTTGGCGGAGCCTTCGAAGAATTCG
AAGAAGAAACGGGTGGTGGTGAAGAAGAACTTCCATTCCCTGAAGAAGAATGACGCGCCTTCTCCTCA

TSP4 gene with InFusion ends

TTCCAGGGACCAGCAATGCCCAAAAGACGCAACAAATCCCCCACGGTTCCTGGACCTGGGACCATAGA
CCCCAACAAGACCGACCAGCTGGTGGGCCAGGACGTACCTTTCATGGAGTCCTCCTTCCTGGAGGTCG
TAGACCGAGAGTTTGAGGAGATGCTGCAGGGCAAGGACGCCCTGCTCATCTACCACAAGATGCTGGCG
GACGGCACGGTGCAGAACGCCTTGCACTACATCTTTGGCCGCATGCGCTCCGCCAAGTGGTACGTGGA
GCCTGCCAGCAGCGACCCTGAAGACGTCCAAATTGCCGCCTTTGTTCACGACCAGCTGGGCTTGGACG
ACTCCAGCGTTGGCAAGTACCCATTTGGACGACTGTTCGCCGTCTACCAGAACGCATATGTTTACGGC
ATGGCCGCCGGAGAGCTGGTGCTGTCCATGGGAGCTGACGGCAAGCTCATTCTGGACAAGGTCGTCCC
CATCCACCCCTTCAACGTGGACGAGGTTGTGTACGACAAAGATGGTGGCCCCAAGGCGCTCAAGGTAT
CCGGCAGCGTGAAGGGCAGCGCCGAACTGCGCACGGTGGAGGTGCCCATCTGGAAGACGGTGGTCTTT
GTCCACGACGACGACGGCTCCTTTACGGGCCGTTCGGCTTTGCGGGCGGCGTTCCCGCACTGGCTGGC
CAAGCGGGCCATGATTCTTCTTATCAACCACGGCCTGGAGCGCTACATGATTGGCGTTCCGGTGCTTA
AACTGCCTAAGACCGTGCGCCAGGGCACTGCACAGTGGCAGGCCGCCGAGTCCATCGTCAGGAACTTT
GTTCAAAAACCGCGCCATGGTATAATACTCCCGGACGACTGGGAATTTGACACTGTGGACCTCAAGTC
GGCCATGCCAGACGCCATCCCCTACCTCACCTACCACGATGCGGGTATCGCCAGAGCCCTTGGCATCG
ACTTTAACACTGTGCAGCTGAACTCTGGCGTGCAGGCGGTGCAGTTTGGCGAGTTCATCAGCTTGACA
CAGCAGACCATCATAGCCCTACAACGGGAGTTCGCTAGTGCCATCAACCTGTATCTCATCCCAAAGCT
GGTGCTGCCAAACTGGCCCAAGGTTACCCGCTTCCCCAAGCTGACCTTTGAAACGGAGGAGCGCAGCG
AGTTCTCCGCCGCCGCCAACCTCATGGGGATGATTATCAACGCCACCAAGGACAACCCGCAGTTCCCA
GCCGTCATGGACGACCTGATTCAAAAGCTTCCCAGCAAGATGCGCAGGGCTCTCGGTGTCGTGGACCC
GGTGAAGGAGGAGGTTCGCAAGGGAGGCGGAGCTAACTACGCGTACAGGAGCAGGAGATGACGCGCCT
TCTCCTCA

A) BV1

```
        10         20         30         40         50         60
MGLFNFFRRK TRSEPTNAIS WFLTQEAYDT LAIPGYTRLS DNPEVRMAVH KIAELISSMT

        70         80         90        100        110        120
IHLMQNTDNG DIRIKNELSR KIDINPYSLM TRKAWMYNIV YTMLLDGEGN SVVFPKYTTS

       130        140        150        160        170        180
GLIDELIPLA PSKVSFVDTD TGYQIWYQGK AYNYDEVLHF IVNPDPEKPY MGRGYRVVLK

       190        200        210        220        230        240
DIVNNLKQAT TTKKSFMSGK YMPSLIVKVD AATAELSSEE GRNAVFKKYL EASEAGQPWI

       250        260        270        280        290        300
IPAELLDVEQ VKPLSLKDLA IHETVELDKR TVAGIFGVPA FLLGVGKYDK DEYNNFINST

       310        320        330        340        350        360
ILPIAKGIEQ ELTRKLLISP DLYFKFNPRS LYAYDLKELA EVGSNMYVRG LMEGNEVRDW

       370        380        390        400
LGLSPKEGLS ELVILENYIP LDKIGDQNKL KGGEKGGADG QTD
```

B) D6E

```
        10         20         30         40         50         60
MLGLYNRYLG SGLPIQSRQL PDPKKPNNKI PNDYRGYIIN QVVGYLWGQP ISYSIDSRNY

        70         80         90        100        110        120
DEARLKEYHD RLSRFNALNS IDDLDSELGK IMSICGYAAR LLYIDKNGEE RAMNIFPWEA

       130        140        150        160        170        180
VFVEDGGEIT HAIRYYKVKD LNNNEYTKVE LYDGTNVTFF IGDGDVFIMD SEGSQPHLFD

       190        200        210        220        230        240
YVPLIRFQNN DEEQGDFEKV EALIDAYDKI ISDSVNEIET FAHAYMKFKG VEADEETINN

       250        260        270        280        290        300
AKQTGGFSVP QDGDVGFITK DINDNFVENN KKTLNENIHK FSASVDMSDE KFSGGAQTGE

       310        320        330        340        350        360
SRKWKLIALE NKAGTKARKF GKGLREQFKV LCSAWQKKGI DLDYLDIFWE FKRNVPIDLS

       370        380        390        400        410        420
YVADYASKLN GIHSKHTLLS QIPYIDDVNY ELELMRQEQE EQVNLDMFVD HGLSDQESEL

DV
```

C) GBSV1

```
        10         20         30         40         50         60
MGLFDRWRRT KRKSKIRADT GYVGLFMSGE DVSFLVPGYV RLSDNPEVRM AVHKIADLIS

        70         80         90        100        110        120
SMTIYLMQNT EDGDIRIRNE LSRKIDITPY SLMTRKSWMY NIVYTMLLDG EGNSVVFPKY

       130        140        150        160        170        180
TADGLIDELV PLTPSKVNFL DTPDGYQVLY GGQTFNYDEV LHFIYNPDPE RPYIGRGYRV

       190        200        210        220        230        240
VLKDIADNLK QATATKKSFM SGKYMPSLIV KVDAATAELS SEEGRNAVFK KYLQATEAGQ

       250        260        270        280        290        300
PWIIPAELLE VEQVKPLSLK DIAINEAVEL DKRTVAGMFG VPAFLLGIGE FNRDEYNNFI

       310        320        330        340        350        360
NSTILPIAKG IEQELTRKLL ISPDLYFKFN PRSLYAYDLK ELAEVGSNMY VRGIMEGNEV

       370        380        390        400
RDWLGLSPKE GLSELVILEN YIPLDKIGDQ SKLKGGDNSG ADGQTD
```

D) GVE2

```
        10         20         30         40         50         60
MGWWSRLTRF FRPRNETVDM TNPLLLQWLG VDPDTPRNQL SEATYFACLK ILSESLGKLP

        70         80         90        100        110        120
LKMYQKTERG IVKSDREELY NLLKLRPNPY MTSSVFWSTV EMNRNHYGNA YVWCQYSGPQ

       130        140        150        160        170        180
LQALWILPSQ YVTIVVDDRG LLGEKNAIWY RYNDPYDGKM YVFRNDEILH FKTSVTFDGI

       190        200        210        220        230        240
TGLSVRDVLK HTVDGALESQ KFMNNLYKTG LTGKAVLEYT GDLNQEARDR LVKGFEQFAN

       250        260        270        280        290        300
GSKNAGKIIP VPLGMKLVPL DIKLTDSQFF ELKKYTALQI AAAFGIKPNQ INDYEKSSYA

       310        320        330        340        350        360
SAEAQNLAFY VDTLLYVLKQ YEEEITYKIL SNDLISQGHY FKFNVNVILR ADIKTQMDSL

       370        380        390        400        410
STAVQNGIMT PNEARDYLDM PADDYGNNLM ANGNYIPLSM LGANYGKGGD S
```

E) phiFA

```
        10         20         30         40         50         60
MNIRDAPKLF LGAFGLLRNG STQHAVPREE SVPLDESDIN KLTGRQRVPI LSLDLEDEQG

        70         80         90        100        110        120
NILVQNNELT EVELSREEWL KEYHKMRLDD SVQDALNILV TAISNFRYYV FPASDEAKDV

       130        140        150        160        170        180
KIAAWLADQL GIGHPSFSKY NNFRTLLKAY ELSLIYGYSA LEIVLNEAGG LDKLVLIHPF

       190        200        210        220        230        240
SIVDIERDSK GGPKNLVVRG YVKAGTTQAS EYKEVKVPFA KVVFFAHDDE GNLDGRSILR

       250        260        270        280        290        300
AAYVPWRIKQ AMLKLVNAGY ERFLLGIPVL TVPKTVTPGT KEWKMAENTL KALAAKPRSG

       310        320        330        340        350        360
VALPEGWQLE IYTVNSQMPD ALPYIVRQDV AIKRAMGVSM AALGLDTGGN YKQSDQLSKV

       370        380        390        400        410        420
SRETALRMAK EFTDYVNLYL VNRLVYLNYP TLRRYPYLGL VSSAQADPAS VLNAFAQMLS

       430        440        450        460
AAAGAGGFNE ETYQRVAALA PRAVRDMLGF EEDRLLRLVE EARRSVG
```

F) phiOH2

```
        10         20         30         40         50         60
MLIEDLFRAP WHERALAELA KGIMTDEQLL AAIVKDWETS EKRNLMLLGD RYYRTKMDIE

        70         80         90        100        110        120
KKNQDITWRS NQKLAHDFVK KLVNQKVGYL LSKEPTIATE NETYRKIMQD MFDKRLLKVV

       130        140        150        160        170        180
KNLGKEAINK GIAFLYVYID EKGELAFKKI PSEQIIPFWK DNDHEEIVSF IRVYEEVVYT

       190        200        210        220        230        240
NTQKQLQKKV EYHHPNGINY YVWQADSLVP DVLAGVETNY HFMIDGKPYL WERMPLIAFK

       250        260        270        280        290        300
YNEEEQPLID CIKSLIDDYN LQASVNADLL ADIPNFIYKL VNYGGTNLQE FLNDLNRYRA

       310        320        330        340        350        360
VKLDENGDVD KLQADLQTDA VEKELLRIRK AIYEFGRGVD TQDENLGNAS GVALRYRYSD

       370        380        390        400        410        420
LDMDCNILET EFQSSLEHLI WFIDQYLLMT GKGDFTNEPI SIIFNRDIII NESEVIANCQ

       430        440        450        460        470
ASVGILDDQT IRENHPWYTE QVEERLKKQQ EQEQMYNGYQ GAFQQQRKDG NVNE
```

G) RM378

```
        10         20         30         40         50         60
MPSLEKYKKL NEAVNFTNFL SPMYGMGAPH GAGGSSMIPI NMYHPFATAG YASRFYGGIE

        70         80         90        100        110        120
FNRFFLYDMY DRMDYTDPLI STVLDIIADE CTIPNENGNI VDVVTKDIEL AKAILSYLDY

       130        140        150        160        170        180
VINIEKNAYP IIRNMIKYGD MFLHILEKGS DGTIEKFQVV SPYIFSKRYN PETDTWYYVI

       190        200        210        220        230        240
TDVYRNVVSG YFNEDIPEED VIHFSHKIDT NFFPYGRSYL ESARAIWNQL RLMEDALMLY

       250        260        270        280        290        300
RVVRSVDRRV FYVDVGNVPP DKINEYLTNI AMQYKRDYWV RNNQNQFLGI DNYFSIESIL

       310        320        330        340        350        360
KDYFIPRRGD RRAVEIDILQ GSKVDLAEDV EYMLNRLISA LKVPKAFIGY EGDVNAKNTL

       370        380        390        400        410        420
ATQDIKFNNT IKRIQGFFVE ELERMVRMNK EFADQDFRLV MNRSNSIVEG ERFAVIEQRI

       430        440        450        460        470        480
GIAERLKGWV REDWIYSNIL QIPYDLKPQE EVAEAAGGGG LFDTGGFGEE TTPADFLGER

       490        500        510        520        530
GSPIESPRGR TEFDFGTEGG EELGGELNLG GAFEEFEEET GGGEEELPFP EEE
```

H) TSP4

```
        10         20         30         40         50         60
MPKRRNKSPT VPGPGTIDPN KTDQLVGQDV PFMESSFLEV VDREFEEMLQ GKDALLIYHK

        70         80         90        100        110        120
MLADGTVQNA LHYIFGRMRS AKWYVEPASS DPEDVQIAAF VHDQLGLDDS SVGKYPFGRL

       130        140        150        160        170        180
FAVYQNAYVY GMAAGELVLS MGADGKLILD KVVPIHPFNV DEVVYDKDGG PKALKVSGSV

       190        200        210        220        230        240
KGSAELRTVE VPIWKTVVFV HDDDGSFTGR SALRAAFPHW LAKRAMILLI NHGLERYMIG

       250        260        270        280        290        300
VPVLKLPKTV RQGTAQWQAA ESIVRNFVQK PRHGIILPDD WEFDTVDLKS AMPDAIPYLT

       310        320        330        340        350        360
YHDAGIARAL GIDFNTVQLN SGVQAVQFGE FISLTQQTII ALQREFASAI NLYLIPKLVL

       370        380        390        400        410        420
PNWPKVTRFP KLTFETEERS EFSAAANLMG MIINATKDNP QFPAVMDDLI QKLPSKMRRA

       430        440
LGVVDPVKEE VRKGGGANYA YRSRR
```

Figure A1: Amino acid sequences of the eight portal proteins

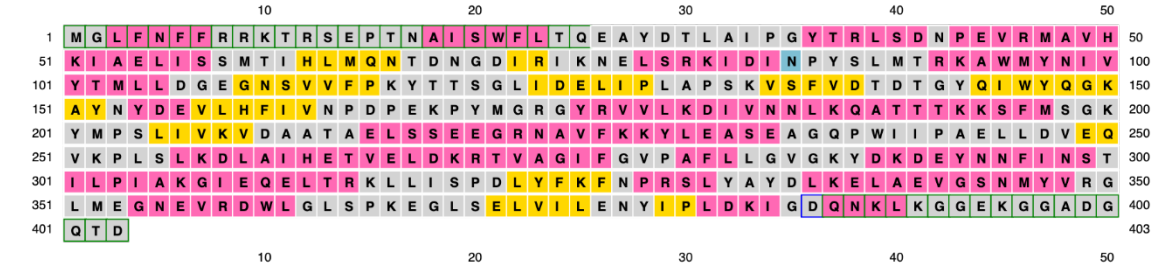
Appendix 2: Disorder predictions of the eight portal proteins

Colored PROMALS3D alignment (sequences in aligned order)

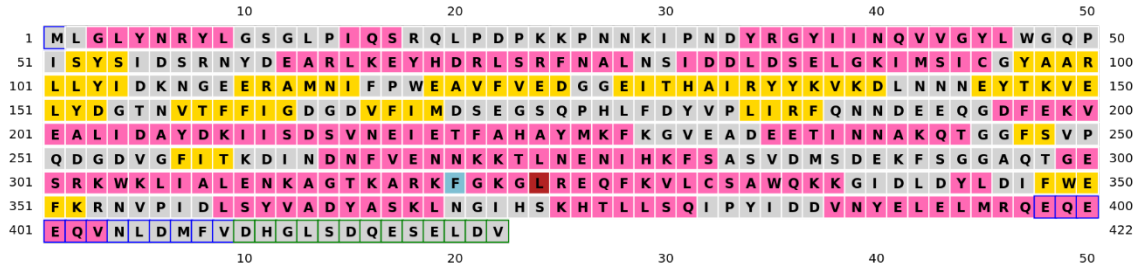
```
Conservation:                                5   7      5
TSP4          183 ---SAELRTVE---VPIWKTVVFVHDDDG----------------------
phiFA         206 TTQASEYKEVK---VPFAKVVFFAHDDEG----------------------
GVE2          157 -----DGKMYV---FRNDEILHFKTSVTFD----------------------
RMB78         191 ------YFNED---IPEEDVIHFSHKIDTNF--------------------
GBSV1         155 --------------FNYDEVLHFIYNPDPER--------------------
BV1           152 --------------YNYDEVLHFIVNPDPEK--------------------
phiOH2        182 ---TQKQLQKKVEYHHPNGINYYVWQADSLVPDVLAGVETNYHFMIDGKPYL
D6E           144 ------NEYTKVELYDGTNVTFFIGDGDVFIMDSEGS----------QPHL
Consensus_aa:                ..............h...pIh@F.hps-s.....................
Consensus_ss:                      ee    e     eeeeee

Conservation:
TSP4          213 LRAAFPHWLAKRAMILLINHGLERYMIG---VPVLKLPK-TVRQGTAQWQAA
phiFA         239 LRAAYVPWRIKQAMLKLVNAGYERFLLG---IPVLTVPK-TVTPGTKEWKMA
GVE2          186 RDVLKHTVDGALESQKFMNNLYKTGLTG---KAVLEYTG-DLNQ-EARDRLV
RMB78         219 --YLESARAIWNQLRLMEDALMLYRVVRSVDRRVFYVDVGNVPP-DKINEYL
GBSV1         179 RVVLKDIADNLKQATATKKSFMSGKYMP---SLIVKVDA-ATAE-LSSEEGR
BV1           176 RVVLKDIVNNLKQATTTKKSFMSGKYMP---SLIVKVDA-ATAE-LSSEEGR
phiOH2        249 IDCIKSLIDDYNLQASVNADLLADIPNF---IYKLVNYGG-----TNLQEFL
D6E           197 FEKVEALIDAYDKIISDSVNEIETFAHA---YMKFKGVEA-------DEETI
Consensus_aa:               ..hhbshh..bp....h.ss.h...hh.......h..s...........chh
Consensus_ss:               hhhhhhhhhhhhhhhhhhhhhhhh          eee          hhhhhhh

Conservation:                                          7
TSP4          265 ----------------RNFVQKPR---HGIILPDD-WEFDTVDLKSAMPDA-
phiFA         291 ----------------KALAAKPR---SGVALPEG-WQLEIYTVNSQMPDA-
GVE2          236 ----------------EQFANGSKNAGKIIPVPLG-MKLVPLDIKLTDSQF-
RMB78         286 QFLGIDNYFSIESILKDYFIPRRG-DRR-------AVEIDIL--QGSKVDL-
GBSV1         230 ----------------KKYLQATE-AGQPWIIPAELLEVEQV--KPLSLKDI
BV1           227 ----------------KKYLEASE-AGQPWIIPAELLDVEQV--KPLSLKDL
phiOH2        297 ----------------RY--------RAVKLDEN-GDVDKLQADLQTDAV-
D6E           243 ----------------QT--------GGFSVPQD-GDVGFITKDINDNFV-
Consensus_aa:               ..................h........pshIP....cl..l..p...s.h
Consensus_ss:                        hhh            ee       eeeeee      hhh

Conservation:       56
TSP4          311 GIDFNTVQLNSG----VQAVQ-FGE--FISLTQQTI----IALQREFASAIN
phiFA         337 GVSMAALGLDTG----GNYKQ-SDQ--LSKVSRETA----LRMAKEFTDYVN
GVE2          285 GIKPNQINDYEK----SSYASAEAQ--NLAFYVDTL----LYVLKQYEEEIT
RMB78         342 KVPKAFIGYEGDVNAKNTLA--TQDIKFNNTIKRIQ----GFFVEELERMVR
GBSV1         280 GVPAFLLGI-------GEFN--RDE--YNNFINSTI----LPIAKGIEQELT
BV1           277 GVPAFLLGV-------GKYD--KDE--YNNFINSTI----LPIAKGIEQELT
phiOH2        338 GVDTQDENLGNA----------SGV--ALRYRYSDLDMDCNILETEFQSSLE
D6E           284 SVDMSDEKFSGGAQ--------TGE--SRKWKLIALENKAGTKARKFGKGLR
Consensus_aa:               tls...b.h.............psp....ph..phh......h.cph.p.lp
Consensus_ss:                  hhh               h   hhh   hhhhhhhhh     hhhhhhhhhhhh

Conservation:                        5
TSP4          360 LPNWPKVTRFPKLTFETE--ERSEFSAAANLMGMIINATKDNPQF-PAVMDC
```

Figure A2: Multiple sequence alignment of the eight portal proteins (Pei, Kim, and Grishin 2008).

A) BV1

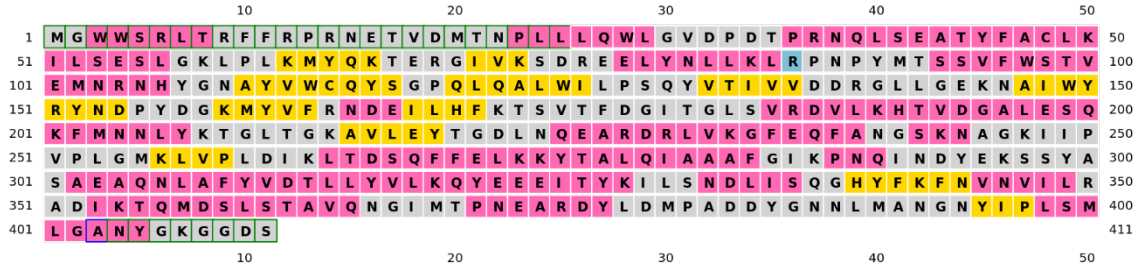


B) D6E

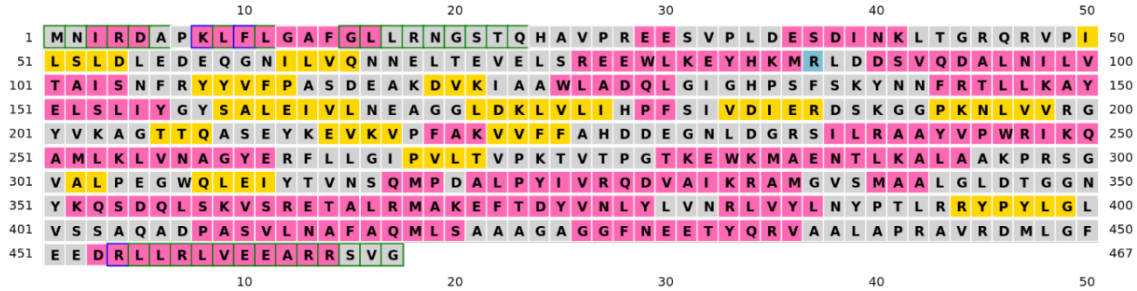


C) GBSV1

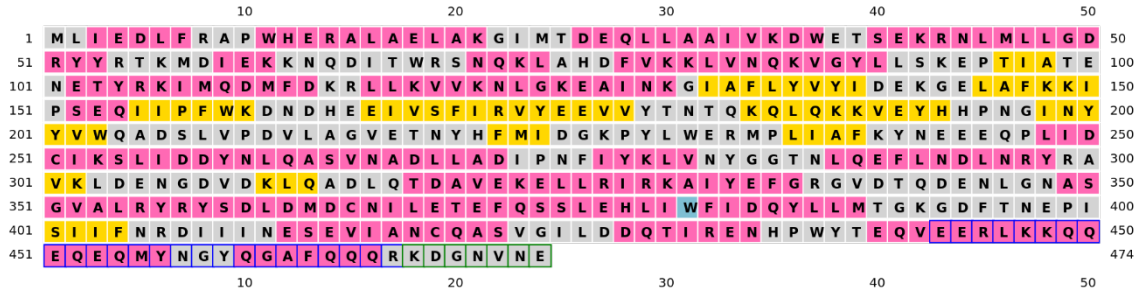
D) GVE2



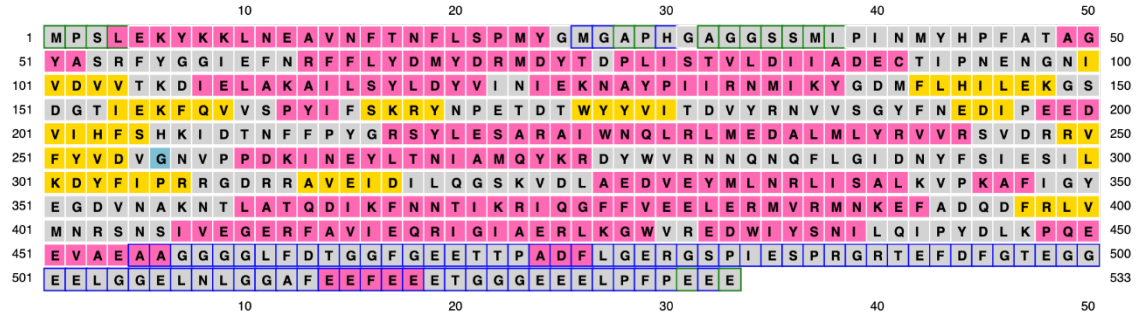
E) phiFA



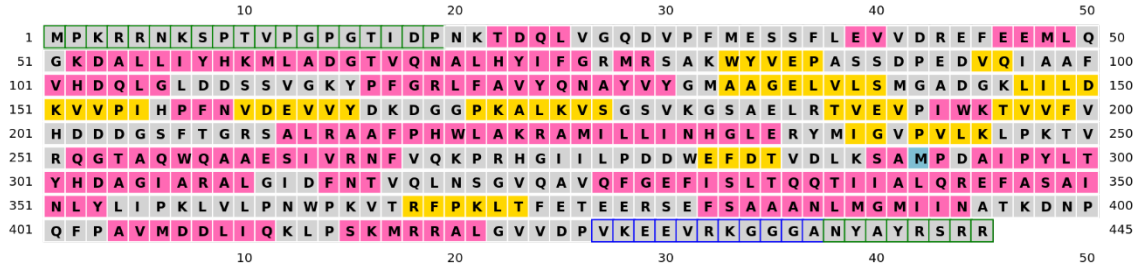
F) phiOH2



G) RM378



H) TSP4



Key

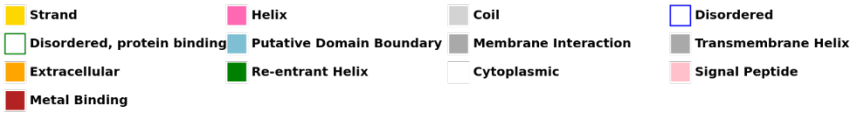
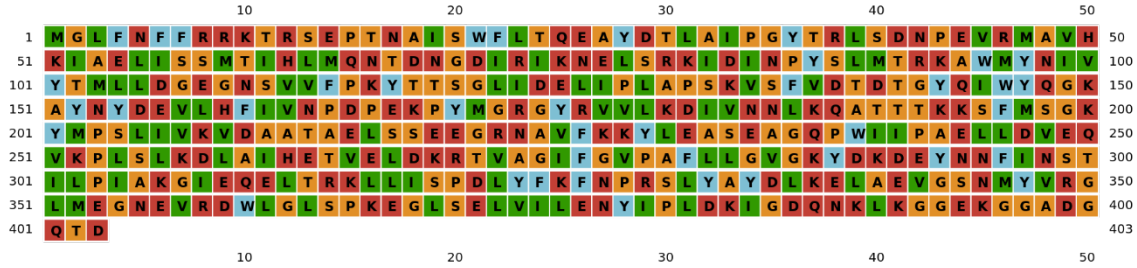


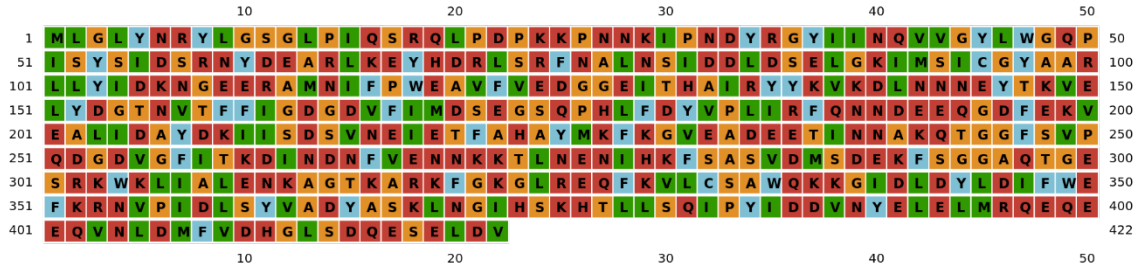
Figure A3: PsiPred analysis of secondary structure

PsiPred chart for each of the eight portal proteins (Buchan and Jones 2019)

A) BV1



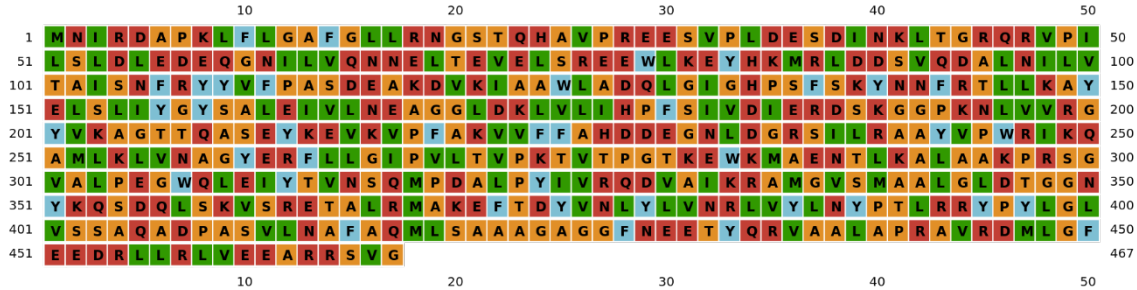
B) D6E



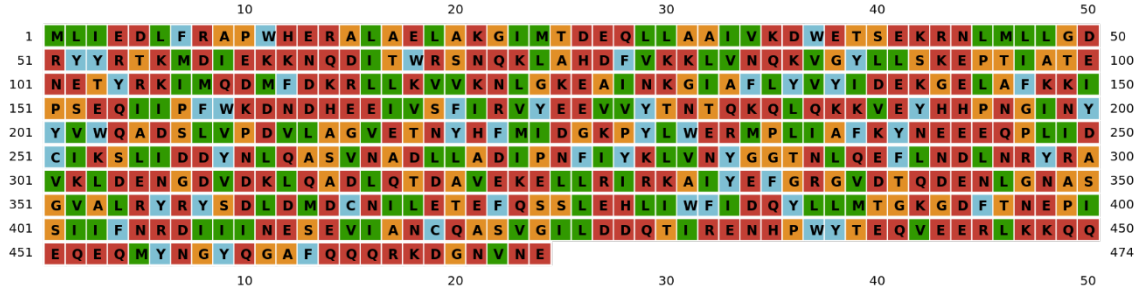
C) GBSV1

D) GVE2

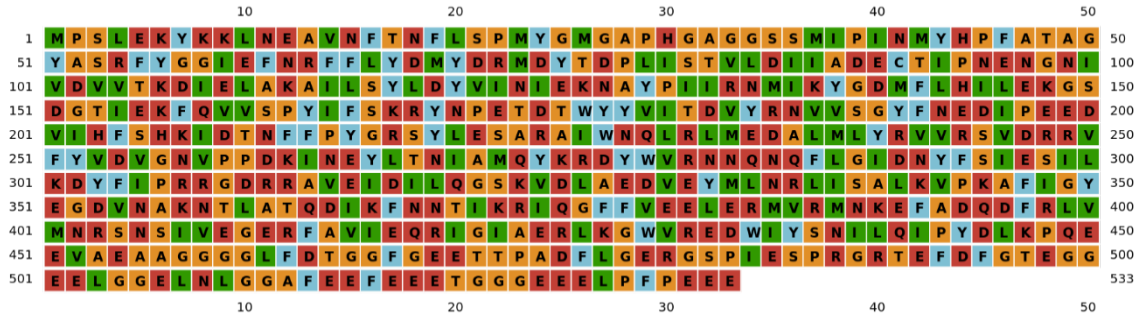
E) phiFA



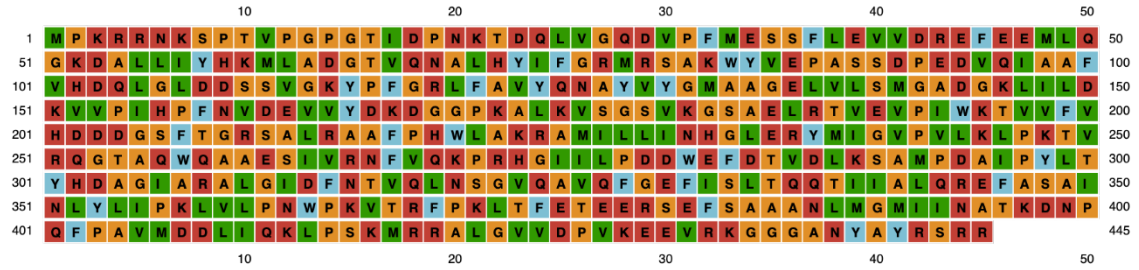
F) phiOH2



G) RM378



H) TSP4

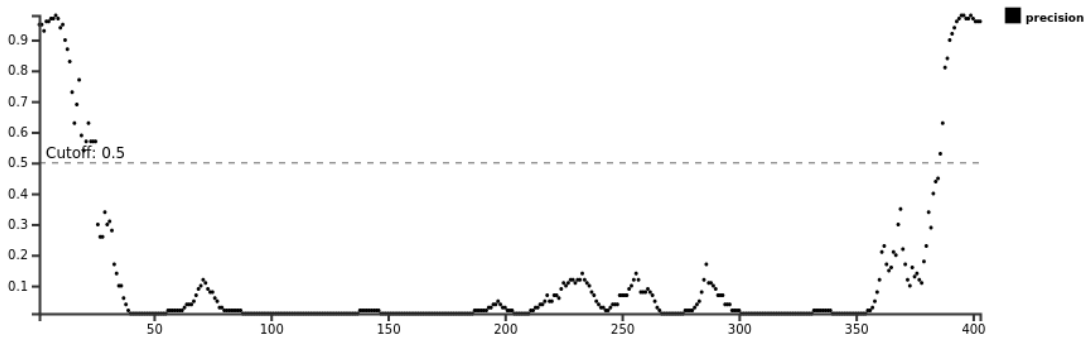


Key

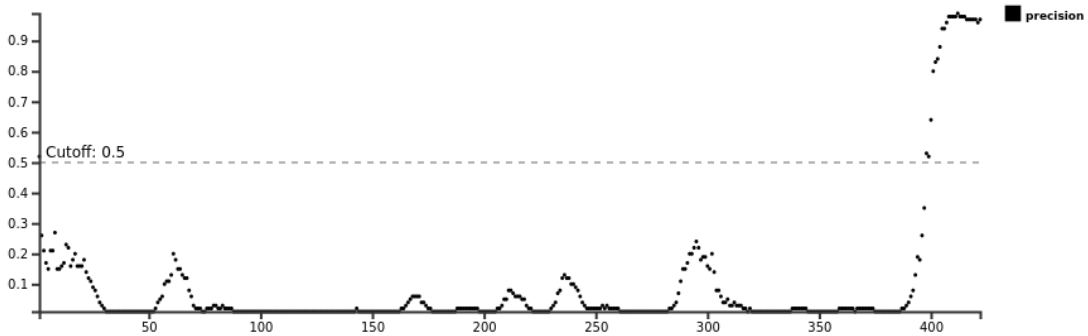
Figure A4: PsiPred of amino acid type.

Maps of amino acid types for the eight portal proteins (Buchan and Jones 2019)

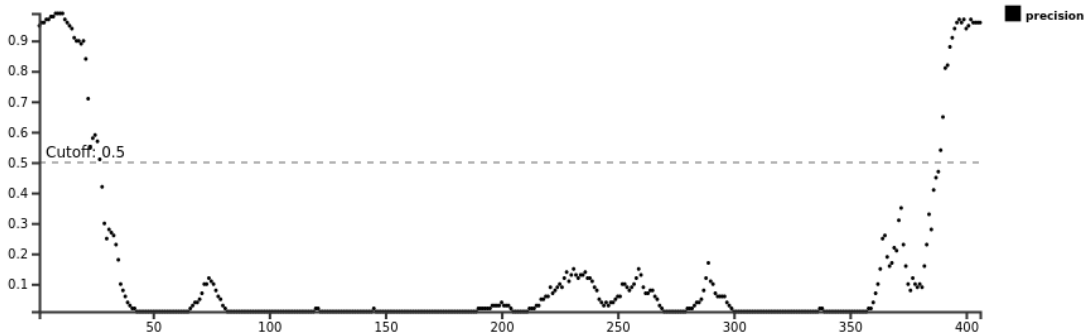
A) BV1



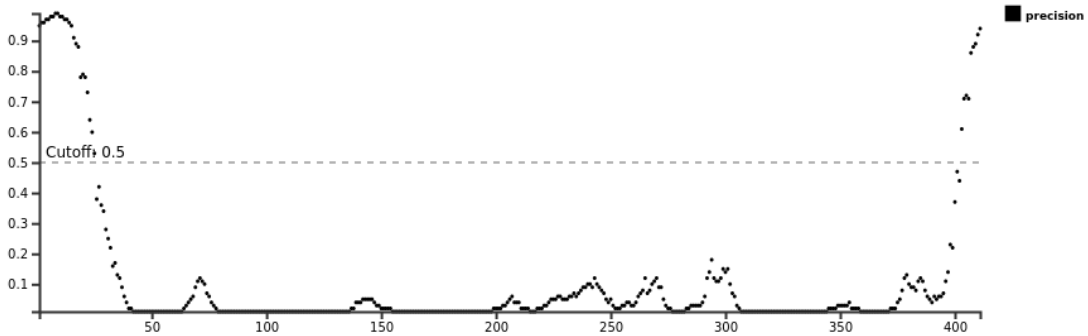
B) D6E



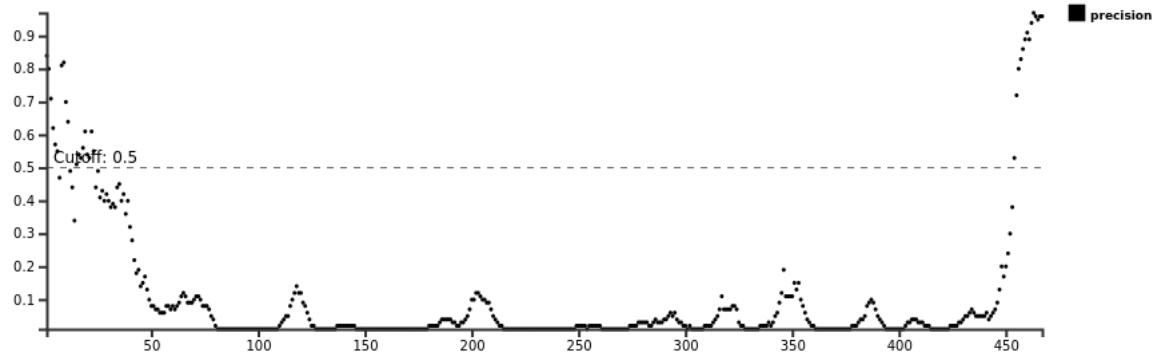
C) GBSV1



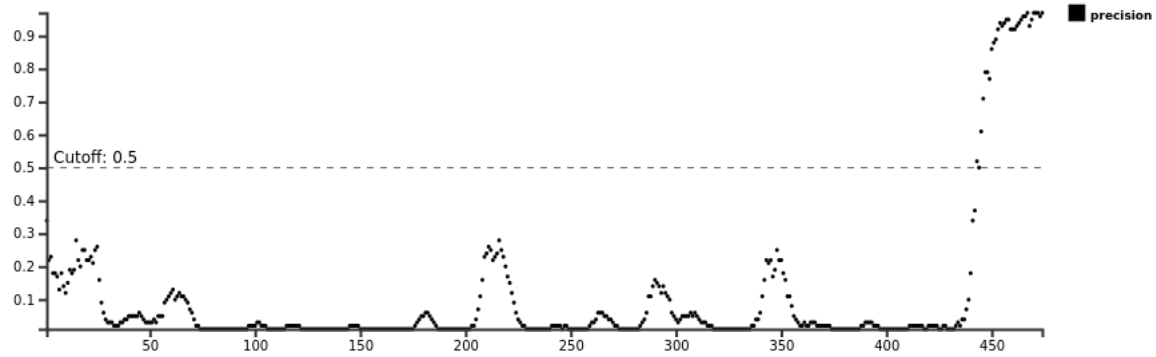
D) GVE2



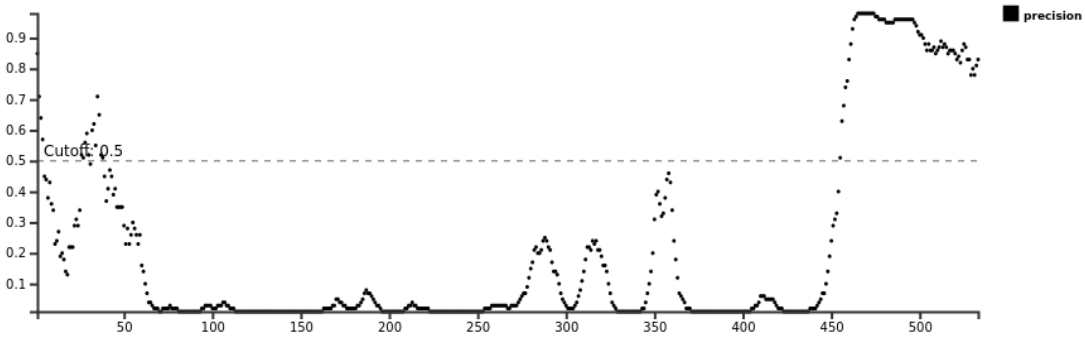
E) phiFA



F) phiOH2



G) RM378



H) TSP4

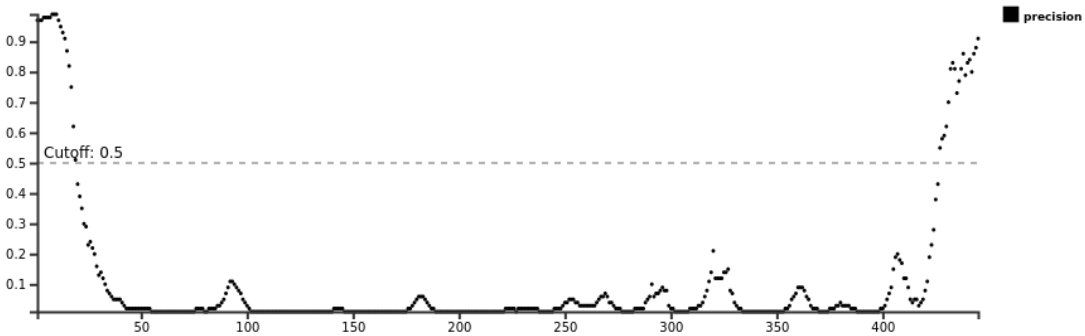
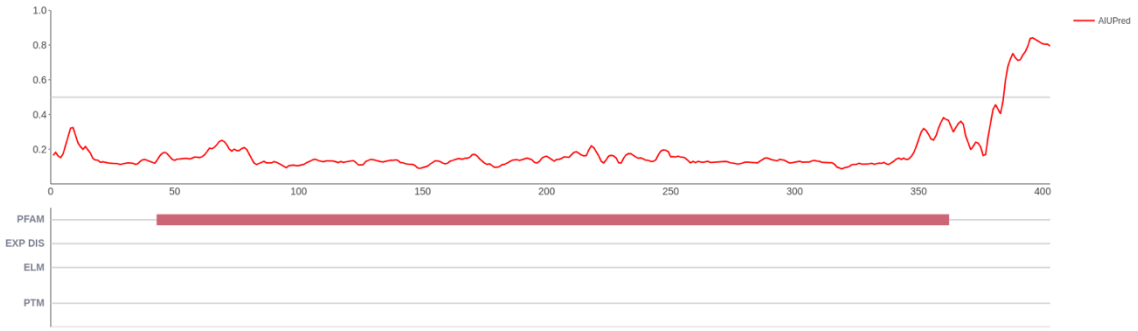


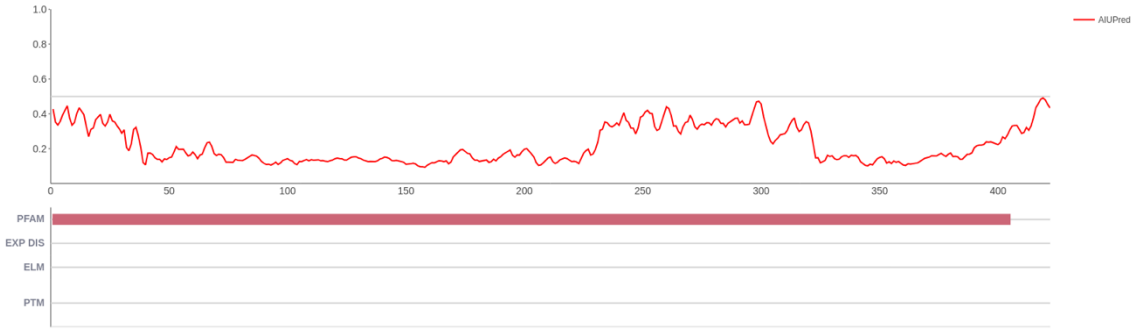
Figure A5: DISOPRED3 Plot

Plots of disorder for the eight portal proteins (D. T. Jones and Cozzetto 2015).

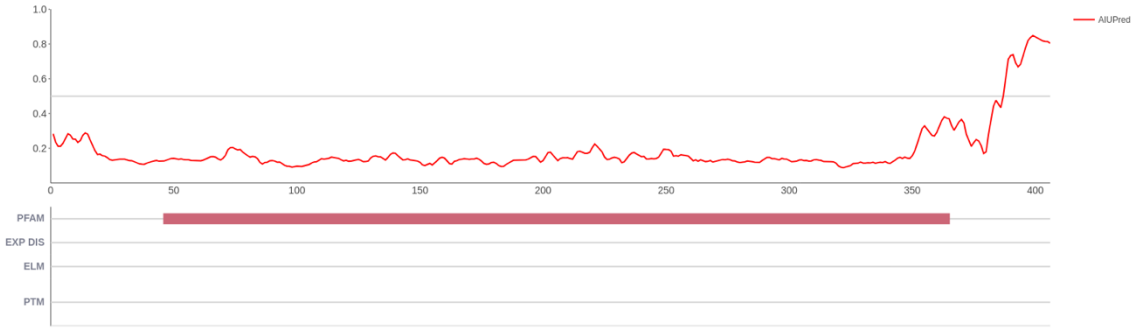
A) BV1



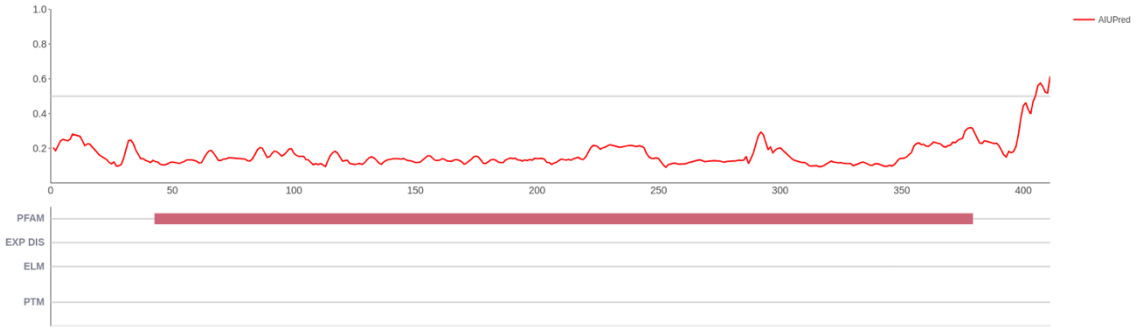
B) D6E



C) GBSV1

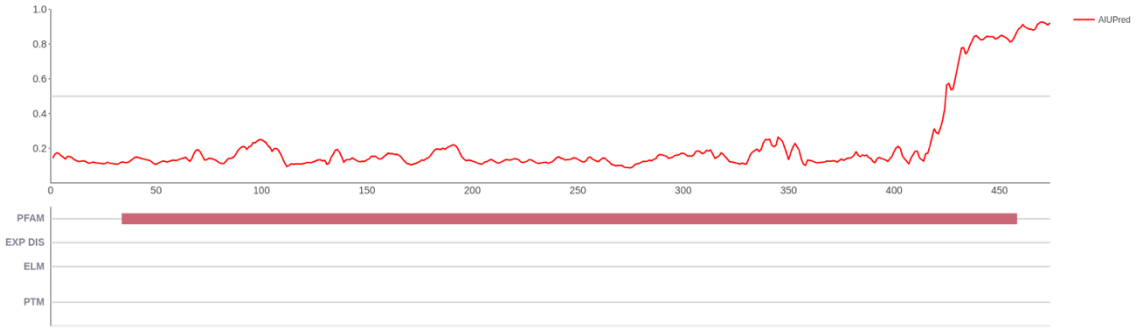


D) GVE2

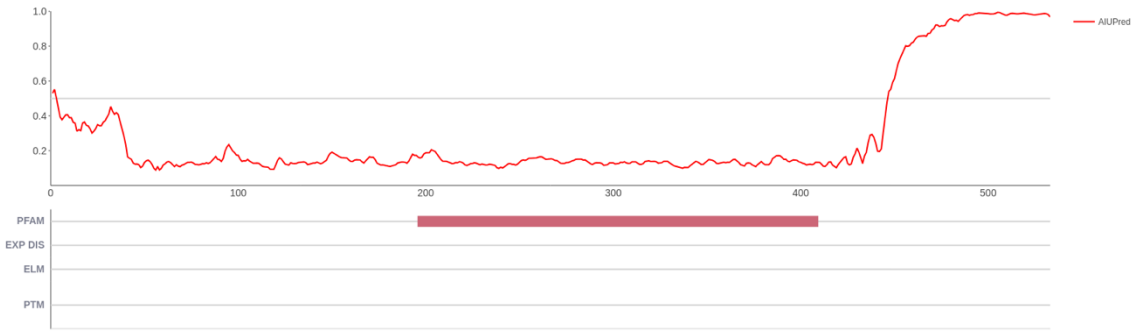


E) phiFA

F) phiOH2



G) RM378



H) TSP4

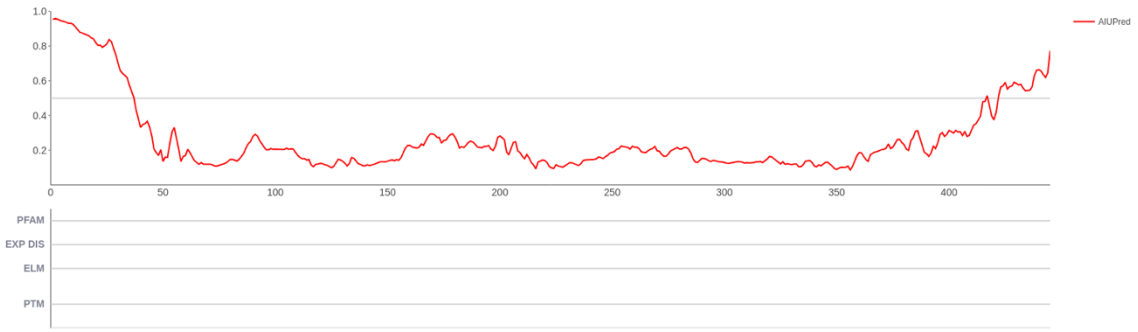


Figure A6: Protein disorder predictions by AIUPred

The PFAM measure below the graph of disorder shows regions of the protein that match certain families. In this case, all of the instances are for the portal protein family (Erdős and Dosztányi 2024).

Appendix 3: Scripts

YSBL movie sorting script

After importing the movies into RELION, they were sorted using the following short script written by Dr Huw Jenkins:

```
cp movies.star movies.star.orig
sed '/FoilHole/Q' movies.star.orig > movies.star
grep FoilHole movies.star.orig | sort -t'_' -k7 -k8 >> movies.star
```

Make 12mer script

The 11 subunits of the GBSV1 portal protein were inserted using a script written by Dr Huw Jenkins as follows:

```
# Coot script to make 12mer Portals from single chain
# Author Huw Jenkins 11.08.22

import math
# hopefully imol 0 is map or coords with correct unit cell
a, b, c, alpha, beta, gamma = cell(0)
assert a == b == c
assert alpha == beta == gamma == 90
starting_imol = first_coords_imol()
chain1 = copy_molecule(starting_imol)
set_molecule_name(chain1, '12mer')
imols = []
for ang, chain in zip(range(30,360,30),
['B','C','D','E','F','G','H','I','J','K','L']):
  ang = math.radians(ang)
  imol = copy_molecule(starting_imol)
  transform_molecule_by(imol, math.cos(ang), -1*math.sin(ang), 0,
                              math.sin(ang), math.cos(ang), 0,
                              0, 0, 1,
                              (0.5 - math.sin(math.radians(45) -
ang) * math.sqrt(0.5))*a,
                              (0.5 - math.cos(math.radians(45) -
ang) * math.sqrt(0.5))*a,
                              0
                         )
  set_molecule_name(imol, 'chain' + chain)
  imols.append(imol)
merge_molecules(imols, chain1)
```

Dipole script

Script written by Eric Pettersen from the RBVI mailing list ('Re: [Chimera-Users] Calculating Dipole Moment of a Protein in Pdb - Chimera-Users - RBVI Mailing Lists', n.d.) (Felder et al. 2007)

```
from chimera import Point, openModels, Molecule, Vector, UserError
```

```
showVector = True  # change to "False" to hide dipole vector
vectorScale = 0.05  # may want to increase/decrease for better
presentation

def centerOfMass(model):
      return Point([a.coord() for a in model.atoms],
            [a.element.mass for a in model.atoms])

# calculation method conscripted from:
#     A server and database for dipole moments of proteins
#     Nucleic Acids Res. 2007 Jul;35(Web Server issue):W512-21
#     doi:10.1093/nar/gkm307
for m in openModels.list(modelTypes=[Molecule]):
      com = centerOfMass(m)
      dipole = Vector()
      for a in m.atoms:
            try:
                  dipole += a.charge * (a.coord() - com)
            except AttributeError:
                  raise UserError("No charge assigned to %s" % a)
      # 4.803 is conversion factor to Debyes for angstrom
measurements
      print "Dipole moment for %s %s: %.3f" % (m, m.name, 4.803 *
dipole.length)
      if showVector:
            v = com + vectorScale * dipole
            bildString = ".arrow %g %g %g %g %g %g .1 .2 .9" % (
                  com[0], com[1], com[2], v[0], v[1], v[2])
            from StringIO import StringIO
            bild = StringIO(bildString)
            openModels.open(bild, type="Bild", identifyAs="%s
dipole" % m.name, sameAs=m)
```

List of abbreviations

Abbreviation	Definition
12mer	An oligomer of 12 subunits
16:0 MPB PE	1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-[4-(p-maleimidophenyl)butyramide] (sodium salt)
2D	Two-dimensional
3D	Three-dimensional
6His	Hexahistidine
Å	Angstrom
A330	Absorbance at 330 nm
A350	Absorbance at 350 nm
AC	Affinity chromatography
AWI	Air-water interface
BFP	Back focal plane
BME	2-cercaptoethanol
bp	Base pair
c	Concentration
C	Capacitance
CD	Cyclodextrin
CD	Circular dichroism
CDB	Controlled dielectric breakdown
CMC	Critical micelle concentration
CryoEM	Cryo-electron microscopy
CTF	Contrast transfer function
CV	Column volume
DG_b	Conductance blockade depth
$DG_{b\text{-}max}$	Maximum conductance blockade depth
$DG_{b\text{-}min}$	Minimum conductance blockade depth
$DG_{b\text{-}norm}$	Normalised conductance depth
DI_b	Current blockade depth
$DI_{b\text{-}norm}$	Normalised current blockade depth
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid

dsDNA	Double stranded DNA
DTT	Dithiothreitol
ECD	Event charge deficit
EDL	Electrical double layer
EDTA	Ethylenediaminetetraacetic acid
E_m	Dielectric strength
EOF	Electro-osmotic force
EP	Electrophoresis
EtOH	Ethanol
FFP	Front focal plane
FL	Full length
FraC	Fracageatoxin C
FSC	Fourier shell correlation
FT	Fourier Transform
FT	Flow-through
G	Conductance
G_0	Baseline conductance
G_b	Blockade conductance
$G_{b\text{-}max}$	Maximum blockade conductance
$G_{b\text{-}min}$	Minimum blockade conductance
GOI	Gene of interest
G_{ratio}	Conductance ratio
HCl	Hydrochloric acid
HfO_2	Hafnium oxide
I	Current
I_0	Baseline current
I_b	Blockade current
$I_{b\text{-}max}$	Maximum blockade current
$I_{b\text{-}min}$	Minimum blockade current
IDT	Integrated DNA technologies
IPTG	Isopropyl β-D-1-thiogalactropyranoside
IQR	Interquartile range
I_{rms}	Root mean square current
I_{RMS}	Root mean square current

I_{rms}	Root mean square current
IV	Current-voltage
KCl	Potassium chloride
K_d	Dissociation constant
kDa	Kilodalton; 10^3 dalsons
KGlu	Potassium glutamate
k_{off}	Dissociation rate
k_{on}	Association rate
kV	Kilovolt; 10^3 volts
mg mL^{-1}	Milligrams per millilitre
$MgCl_2$	Magnesium chloride
mL	Millilitre
MLP	Molecular lipophilicity
mM	Millimolar; 10^{-3} molar
MQ	Milli-Q water
MspA	Mycobacterium smegmatis porin A
mV	Millivolts; 10^{-3} volts
MW	Molecular weight
nA	Nanoampere; 10^{-9} amperes
NaCl	Sodium chloride
nanoDSF	Nano differential scanning fluorimetry
nm	Nanometre; 10^{-9} metre
nS	Nanosiemen; 10^{-9} siemens
NT	Nucleotide
OD_{600}	Optical density at 600 nm
ON	Overnight
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
pH_{pzc}	pH of point of zero charge
pI	Isoelectric point
POI	Protein of interest
Q	Charge

R	Resistance
R_{ac}	Access resistance
R_{ch}	Channel resistance
R_g	Radius of gyration
ROF	Open reading frame
rpm	Rotations per minute
RR	Rectification ratio
s	Second
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SiN_x	Silicon nitride
SiO_2	Silicon oxide
SNR	Signal-to-noise ratio
ssDNA	Single stranded DNA
SSN	Solid-state nanopore
T	Temperature
t	Time
TAE	Tris-acetate-EDTA
T_c	Temperature of cold denaturation
TEM	Transmission electron microscope
Th	Threshold
T_h	Temperature of heat denaturation
T_i	Inter-event time
T_m	Melting temperature
T_{on}	Onset temperature of unfolding
T_s	Temperature of maximum stability
T_t	Dwell time
UA	Uranyl acetate
V	Voltage
v_d	Velocity
V_m	Breakdown voltage
WT	Wild type
αHL	Alpha hemolysin

ΔC_P	Change of heat capacity of denaturation
ΔG_{D-N}	Free energy change from folded to denatured state
ΔH_M	Enthalpy of denaturation at the transition midpoint
ε	Absolute permittivity
ε_0	Permittivity of a vacuum
ε_r	Relative permittivity
η	Viscosity
λ_D	Debye length
μ_{eof}	Electro-osmotic mobility
μ_{ep}	Electrophoretic mobility
µL	Microlitre; 10^{-6} litres
µM	Micromolar; 10^{-6} molar
µs	Microsecond; 10^{-6} second
ρ	Resistivity
σ	Noise
σ	Standard deviation
σ	Conductivity
ϕ_0	Surface potential

Bibliography

Afonine, Pavel V., Billy K. Poon, Randy J. Read, Oleg V. Sobolev, Thomas C. Terwilliger, Alexandre Urzhumtsev, and Paul D. Adams. 2018. 'Real-Space Refinement in PHENIX for Cryo-EM and Crystallography'. *Acta Crystallographica. Section D, Structural Biology* 74 (Pt 6): 531–44. https://doi.org/10.1107/S2059798318006551.

Aharon-Steinberg, A., T. Völkl, A. Kaplan, A. K. Pariari, I. Roy, T. Holder, Y. Wolf, et al. 2022. 'Direct Observation of Vortices in an Electron Fluid'. *Nature* 607 (7917): 74–80. https://doi.org/10.1038/s41586-022-04794-y.

Ahlqvist, Josefin, Javier A. Linares-Pastén, Andrius Jasilionis, Martin Welin, Maria Håkansson, L. Anders Svensson, Lei Wang, et al. 2022. 'Crystal Structure of DNA Polymerase I from Thermus Phage G20c'. *Acta Crystallographica. Section D, Structural Biology* 78 (Pt 11): 1384–98. https://doi.org/10.1107/S2059798322009895.

Alam, Ibrar, Thitikorn Boonkoom, Harit Pitakjakpipop, Poramin Boonbanjong, Kawin Loha, Thanaya Saeyang, Jarunee Vanichtanankul, and Deanpen Japrung. 2024. 'Single-Molecule Analysis of SARS-CoV-2 Double-Stranded Polynucleotides Using Solid-State Nanopore with AI-Assisted Detection and Classification: Implications for Understanding Disease Severity'. *ACS Applied Bio Materials* 7 (2): 1017–27. https://doi.org/10.1021/acsabm.3c00998.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. 'Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs'. *Nucleic Acids Research* 25 (17): 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Altschul, Stephen F., John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu. 2005. 'Protein Database Searches Using Compositionally Adjusted Substitution Matrices'. *The FEBS Journal* 272 (20): 5101–9. https://doi.org/10.1111/j.1742-4658.2005.04945.x.

American Physical Society and American Institute of Physics, eds. 2001. *Physical Review. E: Statistical, Nonlinear, and Soft Matter Physics*. Melville, NY: Published by the American Physical Society through the American Institute of Physics.

Apel, P. 2001. 'Track Etching Technique in Membrane Technology'. *Radiation Measurements*, Proceedings of the 20th International Conference on Nuclear Track s in Solids, 34 (1): 559–66. https://doi.org/10.1016/S1350-4487(01)00228-1.

Apolinário, Arlete, Célia T. Sousa, Gonçalo N. P. Oliveira, Armandina M. L. Lopes, João Ventura, Luísa Andrade, Adélio Mendes, and João P. Araújo. 2020. 'Tailoring the Anodic Hafnium Oxide Morphology Using Different Organic Solvent Electrolytes'. *Nanomaterials (Basel, Switzerland)* 10 (2): 382. https://doi.org/10.3390/nano10020382.

Archer, Donald G., and Peiming Wang. 1990. 'The Dielectric Constant of Water and Debye-Hückel Limiting Law Slopes'. *Journal of Physical and Chemical Reference Data* 19 (2): 371–411. https://doi.org/10.1063/1.555853.

Artamonova, Daria, Karyna Karneyeva, Sofia Medvedeva, Evgeny Klimuk, Matvey Kolesnik, Anna Yasinskaya, Aleksei Samolygo, and Konstantin Severinov. 2020. 'Spacer Acquisition by Type III CRISPR-Cas System during Bacteriophage Infection of Thermus Thermophilus'. *Nucleic Acids Research* 48 (17): 9787–9803. https://doi.org/10.1093/nar/gkaa685.

Aryal, Prafulla, Mark S.P. Sansom, and Stephen J. Tucker. 2015. 'Hydrophobic Gating in Ion Channels'. *Journal of Molecular Biology* 427 (1): 121–30. https://doi.org/10.1016/j.jmb.2014.07.030.

Asandei, Alina, Mauro Chinappi, Hee-Kyoung Kang, Chang Ho Seo, Loredana Mereuta, Yoonkyung Park, and Tudor Luchian. 2015. 'Acidity-Mediated, Electrostatic Tuning of Asymmetrically Charged Peptides Interactions with Protein Nanopores'. *ACS Applied Materials & Interfaces* 7 (30): 16706–14. https://doi.org/10.1021/acsami.5b04406.

Asandei, Alina, Irina Schiopu, Mauro Chinappi, Chang Ho Seo, Yoonkyung Park, and Tudor Luchian. 2016. 'Electroosmotic Trap Against the Electrophoretic Force Near a Protein Nanopore

Reveals Peptide Dynamics During Capture and Translocation'. *ACS Applied Materials & Interfaces* 8 (20): 13166–79. https://doi.org/10.1021/acsami.6b03697.

Atkins, Anthony G., and Marcel Escudier. 2019. *A Dictionary of Mechanical Engineering*. 2nd ed. Oxford Quick Reference. Oxford: Oxford university press.

Aviram, Naama, Ashley N Thornal, David Zeevi, and Luciano A Marraffini. 2022. 'Different Modes of Spacer Acquisition by the Staphylococcus Epidermidis Type III-A CRISPR-Cas System'. *Nucleic Acids Research* 50 (3): 1661–72. https://doi.org/10.1093/nar/gkab1299.

Bacri, Laurent, Abdelghani Oukhaled, Erven Hémon, Fenseth Banzouzi Bassafoula, Loïc Auvray, and Régis Daniel. 2011. 'Discrimination of Neutral Oligosaccharides through a Nanopore'. *Biochemical and Biophysical Research Communications* 412 (4): 561–64. https://doi.org/10.1016/j.bbrc.2011.07.121.

Baker, Nathan A., David Sept, Simpson Joseph, Michael J. Holst, and J. Andrew McCammon. 2001. 'Electrostatics of Nanosystems: Application to Microtubules and the Ribosome'. *Proceedings of the National Academy of Sciences* 98 (18): 10037–41. https://doi.org/10.1073/pnas.181342398.

Bayat, Parisa, Charlotte Rambaud, Bernard Priem, Matthieu Bourderioux, Mélanie Bilong, Salomé Poyer, Manuela Pastoriza-Gallego, Abdelghani Oukhaled, Jérôme Mathé, and Régis Daniel. 2022. 'Comprehensive Structural Assignment of Glycosaminoglycan Oligo- and Polysaccharides by Protein Nanopore'. *Nature Communications* 13 (August):5113. https://doi.org/10.1038/s41467-022-32800-4.

Bayfield, Oliver W., Evgeny Klimuk, Dennis C. Winkler, Emma L. Hesketh, Maria Chechik, Naiqian Cheng, Eric C. Dykeman, et al. 2019. 'Cryo-EM Structure and in Vitro DNA Packaging of a Thermophilic Virus with Supersized T=7 Capsids'. *Proceedings of the National Academy of Sciences of the United States of America* 116 (9): 3556–61. https://doi.org/10.1073/pnas.1813204116.

Bayfield, Oliver W, Alasdair C Steven, and Alfred A Antson. 2020. 'Cryo-EM Structure in Situ Reveals a Molecular Switch That Safeguards Virus against Genome Loss'. *eLife* 9 (April):e55517. https://doi.org/10.7554/eLife.55517.

Bazinet, Christopher, Julyet Benbasat, Jonathan King, Jose Maria Carazo, and Jose L. Carrascosa. 1988. 'Purification and Organization of the Gene 1 Portal Protein Required for Phage P22 DNA Packaging'. *Biochemistry* 27 (6): 1849–56. https://doi.org/10.1021/bi00406a009.

Beamish, Eric, Harold Kwok, Vincent Tabard-Cossa, and Michel Godin. 2012. 'Precise Control of the Size and Noise of Solid-State Nanopores Using High Electric Fields'. *Nanotechnology* 23 (40): 405301. https://doi.org/10.1088/0957-4484/23/40/405301.

———. 2013. 'Fine-Tuning the Size and Minimizing the Noise of Solid-State Nanopores'. *Journal of Visualized Experiments : JoVE*, no. 80 (October), 51081. https://doi.org/10.3791/51081.

Bello, Julian, and Jiwook Shim. 2018. 'Solid-State Nanopore Fabrication in LiCl by Controlled Dielectric Breakdown'. *Biomedical Microdevices* 20 (2): 38. https://doi.org/10.1007/s10544-018-0281-9.

Bellomio, Augusto, Koldo Morante, Ariana Barlic, Ion Gutiérrez-Aguirre, Ana Rosa Viguera, and Juan Manuel González-Mañas. 2009. 'Purification, Cloning and Characterization of Fragaceatoxin C, a Novel Actinoporin from the Sea Anemone Actinia Fragacea'. *Toxicon: Official Journal of the International Society on Toxinology* 54 (6): 869–80. https://doi.org/10.1016/j.toxicon.2009.06.022.

Bentin, Jérémy, Sébastien Balme, and Fabien Picaud. 2020. 'Polynucleotide Differentiation Using Hybrid Solid-State Nanopore Functionalizing with α-Hemolysin'. *Soft Matter* 16 (4): 1002–10. https://doi.org/10.1039/C9SM01833F.

Bepler, Tristan, Kotaro Kelley, Alex J. Noble, and Bonnie Berger. 2020. 'Topaz-Denoise: General Deep Denoising Models for cryoEM and cryoET'. *Nature Communications* 11 (1): 5208. https://doi.org/10.1038/s41467-020-18952-1.

Bepler, Tristan, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J. Noble, and Bonnie Berger. 2019. 'Positive-Unlabeled Convolutional Neural Networks for Particle Picking in Cryo-Electron Micrographs'. *Nature Methods* 16 (11): 1153–60. https://doi.org/10.1038/s41592-019-0575-8.

Bhamidimarri, Satya Prathyusha, Jigneshkumar Dahyabhai Prajapati, Bert van den Berg, Mathias Winterhalter, and Ulrich Kleinekathöfer. 2016. 'Role of Electroosmosis in the Permeation of Neutral Molecules: CymA and Cyclodextrin as an Example'. *Biophysical Journal* 110 (3): 600–611. https://doi.org/10.1016/j.bpj.2015.12.027.

Bhatti, Huma, Rohil Jawed, Irshad Ali, Khurshid Iqbal, Yan Han, Zuhong Lu, and Quanjun Liu. 2021. 'Recent Advances in Biological Nanopores for Nanopore Sequencing, Sensing and Comparison of Functional Variations in MspA Mutants'. *RSC Advances* 11 (46): 28996. https://doi.org/10.1039/d1ra02364k.

Bird, John. 2010. *Electrical Circuit Theory and Technology*. 7th ed. Newnes.

Bisquert, Juan. 2024. 'Hysteresis, Rectification, and Relaxation Times of Nanofluidic Pores for Neuromorphic Circuit Applications'. *Advanced Physics Research* 3 (8): 2400029. https://doi.org/10.1002/apxr.202400029.

Blondal, Thorarinn, Sigridur Hjorleifsdottir, Arnthor Aevarsson, Olafur H. Fridjonsson, Sigurlaug Skirnisdottir, Jon Oskar Wheat, Anna Gudny Hermannsdottir, Gudmundur O. Hreggvidsson, Albert Vernon Smith, and Jakob K. Kristjansson. 2005. 'Characterization of a 5'-Polynucleotide Kinase/3'-Phosphatase from Bacteriophage RM378'. *The Journal of Biological Chemistry* 280 (7): 5188–94. https://doi.org/10.1074/jbc.M409211200.

Blondal, Thorarinn, Sigridur H. Hjorleifsdottir, Olafur F. Fridjonsson, Arnthor Aevarsson, Sigurlaug Skirnisdottir, Anna Gudny Hermannsdottir, Gudmundur O. Hreggvidsson, Albert Vernon Smith, and Jakob K. Kristjansson. 2003. 'Discovery and Characterization of a Thermostable Bacteriophage RNA Ligase Homologous to T4 RNA Ligase 1'. *Nucleic Acids Research* 31 (24): 7247–54. https://doi.org/10.1093/nar/gkg914.

Bouhamidi, Mohamed Yassine, Chunhui Dai, Michel Stephan, Joyeeta Nag, Justin Kinney, Lei Wan, Matthew Waugh, et al. 2024. 'A Method for Fabricating CMOS Back-End-of-Line-Compatible Solid-State Nanopore Devices'. arXiv. https://doi.org/10.48550/arXiv.2411.17416.

Braha, Orit, Barbara Walker, Stephen Cheley, John J. Kasianowicz, Langzhou Song, J. Eric Gouaux, and Hagan Bayley. 1997. 'Designed Protein Pores as Components for Biosensors'. *Chemistry & Biology* 4 (7): 497–505. https://doi.org/10.1016/S1074-5521(97)90321-5.

Briggs, Kyle, Martin Charron, Harold Kwok, Timothea Le, Sanmeet Chahal, José Bustamante, Matthew Waugh, and Vincent Tabard-Cossa. 2015. 'Kinetics of Nanopore Fabrication during Controlled Breakdown of Dielectric Membranes in Solution'. *Nanotechnology* 26 (8): 084004. https://doi.org/10.1088/0957-4484/26/8/084004.

Briggs, Kyle, Harold Kwok, and Vincent Tabard-Cossa. 2014. 'Automated Fabrication of 2-Nm Solid-State Nanopores for Nucleic Acid Analysis'. *Small* 10 (10): 2077–86. https://doi.org/10.1002/smll.201303602.

Brinkerhoff, Henry, Albert S. W. Kang, Jingqian Liu, Aleksei Aksimentiev, and Cees Dekker. 2021. 'Multiple Rereads of Single Proteins at Single–Amino Acid Resolution Using Nanopores'. *Science*, November. https://doi.org/10.1126/science.abl4381.

Brookes, Emre, Borries Demeler, and Mattia Rocco. 2010. 'Developments in the US-SOMO Bead Modeling Suite: New Features in the Direct Residue-to-Bead Method, Improved Grid Routines, and Influence of Accessible Surface Area Screening'. *Macromolecular Bioscience* 10 (7): 746–53. https://doi.org/10.1002/mabi.200900474.

Brookes, Emre, Borries Demeler, Camillo Rosano, and Mattia Rocco. 2010. 'The Implementation of SOMO (SOlution MOdeller) in the UltraScan Analytical Ultracentrifugation Data Analysis Suite: Enhanced Capabilities Allow the Reliable Hydrodynamic Modeling of Virtually Any Kind of Biomacromolecule'. *European Biophysics Journal* 39 (3): 423–35. https://doi.org/10.1007/s00249-009-0418-0.

Brookes, Emre, and Mattia Rocco. 2018. 'Recent Advances in the UltraScan SOlution MOdeller (US-SOMO) Hydrodynamic and Small-Angle Scattering Data Analysis and Simulation Suite'. *European Biophysics Journal* 47 (7): 855–64. https://doi.org/10.1007/s00249-018-1296-0.

Buchan, Daniel W A, and David T Jones. 2019. 'The PSIPRED Protein Analysis Workbench: 20 Years On'. *Nucleic Acids Research* 47 (W1): W402–7. https://doi.org/10.1093/nar/gkz297.

Butler, Tom Z., Mikhail Pavlenok, Ian M. Derrington, Michael Niederweis, and Jens H. Gundlach. 2008. 'Single-Molecule DNA Detection with an Engineered MspA Protein Nanopore'. *Proceedings of the National Academy of Sciences of the United States of America* 105 (52): 20647–52. https://doi.org/10.1073/pnas.0807514106.

Cabello-Aguilar, Simon, Sébastien Balme, Adib Abou Chaaya, Mikhael Bechelany, Emmanuel Balanzat, Jean-Marc Janot, Celine Pochat-Bohatier, Philippe Miele, and Philippe Dejardin. 2013. 'Slow Translocation of Polynucleotides and Their Discrimination by α-Hemolysin inside a Single Track-Etched Nanopore Designed by Atomic Layer Deposition'. *Nanoscale* 5 (20): 9582–86. https://doi.org/10.1039/C3NR03683A.

Cao, Chan, Meng-Yin Li, Nuria Cirauqui, Ya-Qian Wang, Matteo Dal Peraro, He Tian, and Yi-Tao Long. 2018. 'Mapping the Sensing Spots of Aerolysin for Single Oligonucleotides Analysis'. *Nature Communications* 9 (1): 2823. https://doi.org/10.1038/s41467-018-05108-5.

Carter, Jean-Michel, and Shobbir Hussain. 2017. 'Robust Long-Read Native DNA Sequencing Using the ONT CsgG Nanopore System'. *Wellcome Open Research* 2 (April):23. https://doi.org/10.12688/wellcomeopenres.11246.3.

Cervera, Javier, Birgitta Schiedt, Reinhard Neumann, Salvador Mafé, and Patricio Ramírez. 2006. 'Ionic Conduction, Rectification, and Selectivity in Single Conical Nanopores'. *The Journal of Chemical Physics* 124 (10): 104706. https://doi.org/10.1063/1.2179797.

Chen, Chien-Han, Xuyan Chang, and Cen-Shawn Wu. 2019. 'A Novel Shaped-Controlled Fabrication of Nanopore and Its Applications in Quantum Electronics'. *Scientific Reports* 9 (1): 18663. https://doi.org/10.1038/s41598-019-55190-y.

Cheng, Xian, Emily J. Guinn, Evan Buechel, Rachel Wong, Rituparna Sengupta, Irina A. Shkel, and M. Thomas Record. 2016. 'Basis of Protein Stabilization by K Glutamate: Unfavorable Interactions with Carbon, Oxygen Groups'. *Biophysical Journal* 111 (9): 1854–65. https://doi.org/10.1016/j.bpj.2016.08.050.

Chinappi, Mauro, Misa Yamaji, Ryuji Kawano, and Fabio Cecconi. 2020. 'Analytical Model for Particle Capture in Nanopores Elucidates Competition among Electrophoresis, Electroosmosis, and Dielectrophoresis'. *ACS Nano* 14 (11): 15816–28. https://doi.org/10.1021/acsnano.0c06981.

Churaev, N. V., G. A. Martynov, V. M. Starov, and Z. M. Zorin. 1981. 'Some Features of Capillary Imbibition of Surfactant Solutions'. *Colloid and Polymer Science* 259 (7): 747–52. https://doi.org/10.1007/BF01419320.

Cowing, Keith. 2016. 'First DNA Sequencing Conducted in Space'. *SpaceRef* (blog). 29 August 2016. https://spaceref.com/newspace-and-tech/first-dna-sequencing-conducted-in-space/.

Cressiot, Benjamin, Sandra J. Greive, Mehrnaz Mojtabavi, Alfred A. Antson, and Meni Wanunu. 2018. 'Thermostable Virus Portal Proteins as Reprogrammable Adapters for Solid-State Nanopore Sensors'. *Nature Communications* 9 (1): 4652. https://doi.org/10.1038/s41467-018-07116-x.

Cressiot, Benjamin, Sandra J. Greive, Wei Si, Tomas C. Pascoa, Mehrnaz Mojtabavi, Maria Chechik, Huw T. Jenkins, et al. 2017a. 'Porphyrin-Assisted Docking of a Thermophage Portal Protein into Lipid Bilayers: Nanopore Engineering and Characterization'. *ACS Nano* 11 (12): 11931–45. https://doi.org/10.1021/acsnano.7b06980.

———. 2017b. 'Porphyrin-Assisted Docking of a Thermophage Portal Protein into Lipid Bilayers: Nanopore Engineering and Characterization'. *ACS Nano* 11 (12): 11931–45. https://doi.org/10.1021/acsnano.7b06980.

Crnković, Ana, Marija Srnko, and Gregor Anderluh. 2021. 'Biological Nanopores: Engineering on Demand'. *Life (Basel, Switzerland)* 11 (1): 27. https://doi.org/10.3390/life11010027.

Cuervo, Ana, Montserrat Fàbrega-Ferrer, Cristina Machón, José Javier Conesa, Francisco J. Fernández, Rosa Pérez-Luque, Mar Pérez-Ruiz, et al. 2019. 'Structures of T7 Bacteriophage Portal and Tail Suggest a Viral DNA Retention and Ejection Mechanism'. *Nature Communications* 10 (August):3746. https://doi.org/10.1038/s41467-019-11705-9.

Cuervo, Ana, Marie-Christine Vaney, Alfred A. Antson, Paulo Tavares, and Leonor Oliveira. 2007. 'Structural Rearrangements between Portal Protein Subunits Are Essential for Viral DNA Translocation'. *The Journal of Biological Chemistry* 282 (26): 18907–13. https://doi.org/10.1074/jbc.M701808200.

Cytiva. 2020. 'Superdex Prep Grad and Prepakced HiLoad Columns'. Cytiva. https://cdn.cytivalifesciences.com/api/public/content/digi-11217-pdf.

D Y Bandara, Y. M. Nuwan, Jugal Saharia, Buddini I. Karawdeniya, James T. Hagan, Jason R. Dwyer, and Min Jun Kim. 2020. 'Beyond Nanopore Sizing: Improving Solid-State Single-Molecule Sensing Performance, Lifetime, and Analyte Scope for Omics by Targeting Surface Chemistry during Fabrication'. *Nanotechnology* 31 (33): 335707. https://doi.org/10.1088/1361-6528/ab8f4d.

David Hou, Chun-Feng, Nicholas A. Swanson, Fenglin Li, Ruoyu Yang, Ravi K. Lokareddy, and Gino Cingolani. 2022. 'Cryo-EM Structure of a Kinetically Trapped Dodecameric Portal Protein from the Pseudomonas-Phage PaP3'. *Journal of Molecular Biology* 434 (9): 167537. https://doi.org/10.1016/j.jmb.2022.167537.

Delcour, Anne H., ed. 2013. *Bacterial Cell Surfaces: Methods and Protocols*. Vol. 966. Methods in Molecular Biology. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-245-2.

Di Muccio, Giovanni, Aldo Eugenio Rossini, Daniele Di Marino, Giuseppe Zollo, and Mauro Chinappi. 2019. 'Insights into Protein Sequencing with an α-Hemolysin Nanopore by Atomistic Simulations'. *Scientific Reports* 9 (1): 6440. https://doi.org/10.1038/s41598-019-42867-7.

D'Imprima, Edoardo, Davide Floris, Mirko Joppe, Ricardo Sánchez, Martin Grininger, and Werner Kühlbrandt. 2019. 'Protein Denaturation at the Air-Water Interface and How to Prevent It'. *eLife* 8 (April):e42747. https://doi.org/10.7554/eLife.42747.

Doi, Katsumi, Kazuki Mori, Hindra Martono, Yuko Nagayoshi, Yasuhiro Fujino, Kosuke Tashiro, Satoru Kuhara, and Toshihisa Ohshima. 2013. 'Draft Genome Sequence of Geobacillus Kaustophilus GBlys, a Lysogenic Strain with Bacteriophage φOH2'. *Genome Announcements* 1 (4): e00634-13. https://doi.org/10.1128/genomeA.00634-13.

Dube, P., P. Tavares, R. Lurz, and M. van Heel. 1993. 'The Portal Protein of Bacteriophage SPP1: A DNA Pump with 13-Fold Symmetry'. *The EMBO Journal* 12 (4): 1303–9. https://doi.org/10.1002/j.1460-2075.1993.tb05775.x.

Dubochet, J., M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowall, and P. Schultz. 1988. 'Cryo-Electron Microscopy of Vitrified Specimens'. *Quarterly Reviews of Biophysics* 21 (2): 129–228. https://doi.org/10.1017/s0033583500004297.

Dunbar, William B. 2015. 'Comment on Accurate Data Process for Nanopore Analysis'. *Analytical Chemistry* 87 (20): 10650–52. https://doi.org/10.1021/acs.analchem.5b02281.

Eastwood, Tara A., Karen Baker, Bree R. Streather, Nyasha Allen, Lin Wang, Stanley W. Botchway, Ian R. Brown, Jennifer R. Hiscock, Christopher Lennon, and Daniel P. Mulvihill. 2023. 'High-Yield Vesicle-Packaged Recombinant Protein Production from E. Coli'. *Cell Reports Methods* 3 (2): 100396. https://doi.org/10.1016/j.crmeth.2023.100396.

Egerton, R. F., P. Li, and M. Malac. 2004. 'Radiation Damage in the TEM and SEM'. *Micron*, International Wuhan Symposium on Advanced Electron Microscopy, 35 (6): 399–409. https://doi.org/10.1016/j.micron.2004.02.003.

Elliott, Breeana, Martin Charron, John Pezacki, Erin McConnell, and Vincent Tabard-Cossa. 2024. 'Solid-State Nanopore Counting of Amplicons from Recombinase Polymerase Isothermal Amplification'. *Sensors & Diagnostics* 3 (10): 1733–42. https://doi.org/10.1039/D4SD00159A.

Engineered Nanopores for Bioanalytical Applications. 2013. Elsevier. https://doi.org/10.1016/C2010-0-66941-0.

Engst, Christian. R., Marc Ablay, Giorgio Divitini, Caterina Ducati, Tim Liedl, and Ulrich F. Keyser. 2012. 'DNA Origami Nanopores'. *Nano Letters* 12 (1): 512–17. https://doi.org/10.1021/nl204098n.

Erdős, Gábor, and Zsuzsanna Dosztányi. 2024. 'AIUPred: Combining Energy Estimation with Deep Learning for the Enhanced Prediction of Protein Disorder'. *Nucleic Acids Research* 52 (W1): W176–81. https://doi.org/10.1093/nar/gkae385.

Eyraud, Rémi, Stéphane Ayache, Philipp O. Tsvetkov, Shanmugha Sri Kalidindi, Viktoriia E. Baksheeva, Sébastien Boissonneau, Carine Jiguet-Jiglaire, et al. 2023. 'Plasma nanoDSF Denaturation Profile at Baseline Is Predictive of Glioblastoma EGFR Status'. *Cancers* 15 (3): 760. https://doi.org/10.3390/cancers15030760.

Fàbrega Ferrer, Montserrat. 2017. 'Structural Characterization of the T7 Bacteriophage Portal Protein'. Universitat de Barcelona. https://diposit.ub.edu/dspace/handle/2445/118503.

Fahie, Monifa A. V., Bach Pham, Fanjun Li, and Min Chen. 2021. 'A Selective Activity-Based Approach for Analysis of Enzymes with an OmpG Nanopore'. In *Nanopore Technology: Methods and Protocols*, edited by Monifa A.V. Fahie, 115–33. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-0806-7_9.

Fang, Qianglin, Wei-Chun Tang, Pan Tao, Marthandan Mahalingam, Andrei Fokine, Michael G. Rossmann, and Venigalla B. Rao. 2020. 'Structural Morphing in a Symmetry-Mismatched Viral Vertex'. *Nature Communications* 11 (1): 1713. https://doi.org/10.1038/s41467-020-15575-4.

Felder, Clifford E., Jaime Prilusky, Israel Silman, and Joel L. Sussman. 2007. 'A Server and Database for Dipole Moments of Proteins'. *Nucleic Acids Research* 35 (suppl_2): W512–21. https://doi.org/10.1093/nar/gkm307.

Fennouri, Aziz, Régis Daniel, Manuela Pastoriza-Gallego, Loïc Auvray, Juan Pelta, and Laurent Bacri. 2013. 'Kinetics of Enzymatic Degradation of High Molecular Weight Polysaccharides through a Nanopore: Experiments and Data-Modeling'. *Analytical Chemistry* 85 (18): 8488–92. https://doi.org/10.1021/ac4020929.

Fologea, Daniel, Marc Gershow, Bradley Ledden, David S. McNabb, Jene A. Golovchenko, and Jiali Li. 2005. 'Detecting Single Stranded DNA with a Solid State Nanopore'. *Nano Letters* 5 (10): 1905–9. https://doi.org/10.1021/nl051199m.

Forstater, Jacob H., Kyle Briggs, Joseph W. F. Robertson, Jessica Ettedgui, Olivier Marie-Rose, Canute Vaz, John J. Kasianowicz, Vincent Tabard-Cossa, and Arvind Balijepalli. 2016. 'MOSAIC: A Modular Single-Molecule Analysis Interface for Decoding Multistate Nanopore Data'. *Analytical Chemistry* 88 (23): 11900–907. https://doi.org/10.1021/acs.analchem.6b03725.

Fragasso, Alessio, Sonja Schmid, and Cees Dekker. 2020. 'Comparing Current Noise in Biological and Solid-State Nanopores'. *ACS Nano* 14 (2): 1338–49. https://doi.org/10.1021/acsnano.9b09353.

Franceschini, Lorenzo, Tine Brouns, Kherim Willems, Enrico Carlon, and Giovanni Maglia. 2016. 'DNA Translocation through ClyA Nanopores at Physiological Ionic Strengths Requires Precise Nanoscale Engineering'. *ACS Nano* 10 (9): 8394–8402. https://doi.org/10.1021/acsnano.6b03159.

Franken, A.C.M., J.A.M. Nolten, M.H.V. Mulder, D. Bargeman, and C.A. Smolders. 1987. 'Wetting Criteria for the Applicability of Membrane Distillation'. *Journal of Membrane Science* 33 (3): 315–28. https://doi.org/10.1016/S0376-7388(00)80288-4.

Fried, Jasper P., Jacob L. Swett, Binoy Paulose Nadappuram, Jan A. Mol, Joshua B. Edel, Aleksandar P. Ivanov, and James R. Yates. 2021. 'In Situ Solid-State Nanopore Fabrication'. *Chemical Society Reviews* 50 (8): 4974–92. https://doi.org/10.1039/D0CS00924E.

Fujinami Tanimoto, Izadora Mayumi, Benjamin Cressiot, Nathalie Jarroux, Jean Roman, Gilles Patriarche, Bruno Le Pioufle, Juan Pelta, and Laurent Bacri. 2021. 'Selective Target Protein

Detection Using a Decorated Nanopore into a Microfluidic Device'. *Biosensors and Bioelectronics* 183 (July):113195. https://doi.org/10.1016/j.bios.2021.113195.
Fujinami Tanimoto, Izadora Mayumi, Jiayi Zhang, Benjamin Cressiot, Bruno Le Pioufle, Laurent Bacri, and Juan Pelta. 2022. 'Dynamics of DNA Through Solid-State Nanopores Fabricated by Controlled Dielectric Breakdown'. *Chemistry – An Asian Journal* 17 (24): e202200888. https://doi.org/10.1002/asia.202200888.
Garalde, Daniel R., Elizabeth A. Snell, Daniel Jachimowicz, Botond Sipos, Joseph H. Lloyd, Mark Bruce, Nadia Pantic, et al. 2018. 'Highly Parallel Direct RNA Sequencing on an Array of Nanopores'. *Nature Methods* 15 (3): 201–6. https://doi.org/10.1038/nmeth.4577.
Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. 2005. 'Protein Identification and Analysis Tools on the ExPASy Server'. In *The Proteomics Protocols Handbook*, edited by John M. Walker, 571–607. Totowa, NJ: Humana Press. https://doi.org/10.1385/1-59259-890-0:571.
Giacomello, Alberto, and Roland Roth. 2020. 'Bubble Formation in Nanopores: A Matter of Hydrophobicity, Geometry, and Size'. *Advances in Physics: X* 5 (1): 1817780. https://doi.org/10.1080/23746149.2020.1817780.
Gidwani, Bina, and Amber Vyas. 2015. 'A Comprehensive Review on Cyclodextrin-Based Carriers for Delivery of Chemotherapeutic Cytotoxic Anticancer Drugs'. *BioMed Research International* 2015:198268. https://doi.org/10.1155/2015/198268.
Göpfrich, Kerstin, Chandrashekhar V. Kulkarni, Oliver J. Pambos, and Ulrich F. Keyser. 2013. 'Lipid Nanobilayers to Host Biological Nanopores for DNA Translocations'. *Langmuir: The ACS Journal of Surfaces and Colloids* 29 (1): 355–64. https://doi.org/10.1021/la3041506.
Gornall, Joanne L., Kozhinjampara R. Mahendran, Oliver J. Pambos, Lorenz J. Steinbock, Oliver Otto, Catalin Chimerel, Mathias Winterhalter, and Ulrich F. Keyser. 2011. 'Simple Reconstitution of Protein Pores in Nano Lipid Bilayers'. *Nano Letters* 11 (8): 3334–40. https://doi.org/10.1021/nl201707d.
Goto, Yusuke, Kazuma Matsui, Itaru Yanagi, and Ken-ichi Takeda. 2019. 'Silicon Nitride Nanopore Created by Dielectric Breakdown with a Divalent Cation: Deceleration of Translocation Speed and Identification of Single Nucleotides'. *Nanoscale* 11 (30): 14426–33. https://doi.org/10.1039/C9NR03563J.
Goto, Yusuke, Itaru Yanagi, Kazuma Matsui, Takahide Yokoi, and Ken-ichi Takeda. 2016. 'Integrated Solid-State Nanopore Platform for Nanopore Fabrication via Dielectric Breakdown, DNA-Speed Deceleration and Noise Reduction'. *Scientific Reports* 6 (1): 31324. https://doi.org/10.1038/srep31324.
Gouy, M., and C. Gautier. 1982. 'Codon Usage in Bacteria: Correlation with Gene Expressivity'. *Nucleic Acids Research* 10 (22): 7055–74. https://doi.org/10.1093/nar/10.22.7055.
Goyal, Parveen, Petya V. Krasteva, Nani Van Gerven, Francesca Gubellini, Imke Van den Broeck, Anastassia Troupiotis-Tsaïlaki, Wim Jonckheere, et al. 2014. 'Structural and Mechanistic Insights into the Bacterial Amyloid Secretion Channel CsgG'. *Nature* 516 (7530): 250–53. https://doi.org/10.1038/nature13768.
Greive, Sandra J., Laurent Bacri, Benjamin Cressiot, and Juan Pelta. 2024. 'Identification of Conformational Variants for Bradykinin Biomarker Peptides from a Biofluid Using a Nanopore and Machine Learning'. *ACS Nano* 18 (1): 539–50. https://doi.org/10.1021/acsnano.3c08433.
Gu, Li-Qun, Stephen Cheley, and Hagan Bayley. 2003. 'Electroosmotic Enhancement of the Binding of a Neutral Molecule to a Transmembrane Pore'. *Proceedings of the National Academy of Sciences of the United States of America* 100 (26): 15498–503. https://doi.org/10.1073/pnas.2531778100.
Gu, Zhen, Yi-Lun Ying, Chan Cao, Pingang He, and Yi-Tao Long. 2015. 'Accurate Data Process for Nanopore Analysis'. *Analytical Chemistry* 87 (2): 907–13. https://doi.org/10.1021/ac5028758.

Gu, Zhiwei, Kexun Wu, and Jiawei Wang. 2024. 'Structural Morphing in the Viral Portal Vertex of Bacteriophage Lambda'. *Journal of Virology* 98 (5): e00068-24. https://doi.org/10.1128/jvi.00068-24.

Guan, Xiaoyu, Yu Wang, Jinyue Zhang, Wei Shao, Shuo Huang, and Daoqiang Zhang. 2022. '[Unsupervised deep learning for identifying the O 6-carboxymethyl guanine by nanopore sequencing]'. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi = Journal of Biomedical Engineering = Shengwu Yixue Gongchengxue Zazhi* 39 (1): 139–48. https://doi.org/10.7507/1001-5515.202104068.

Guan, Xiaoyu, Yuqin Wang, Wei Shao, Zhongnian Li, Shuo Huang, and Daoqiang Zhang. 2022. 'S2Snet: Deep Learning for Low Molecular Weight RNA Identification with Nanopore'. *Briefings in Bioinformatics* 23 (3): bbac098. https://doi.org/10.1093/bib/bbac098.

Guan, Xin, Haijuan Li, Limei Chen, Guohua Qi, and Yongdong Jin. 2023. 'Glass Capillary-Based Nanopores for Single Molecule/Single Cell Detection'. *ACS Sensors* 8 (2): 427–42. https://doi.org/10.1021/acssensors.2c02102.

Guasch, Alicia, Joan Pous, Borja Ibarra, F. Xavier Gomis-Rüth, José María Valpuesta, Natalia Sousa, José L Carrascosa, and Miquel Coll. 2002. 'Detailed Architecture of a DNA Translocating Machine: The High-Resolution Structure of the Bacteriophage Φ29 Connector Particle1'. *Journal of Molecular Biology* 315 (4): 663–76. https://doi.org/10.1006/jmbi.2001.5278.

Gubbiotti, Alberto, Matteo Baldelli, Giovanni Di Muccio, Paolo Malgaretti, Sophie Marbach, and Mauro Chinappi. 2022. 'Electroosmosis in Nanopores: Computational Methods and Technological Applications'. *Advances in Physics: X* 7 (1): 2036638. https://doi.org/10.1080/23746149.2022.2036638.

Guo, Peixuan, and Shaoying Wang. 2016. Lipid bilayer-integrated spp1 connector protein nanopore and spp1 connector protein variants for use as lipid bilayer-integrated nanopore. World Intellectual Property Organization WO2016115522A2, filed 15 January 2016, and issued 21 July 2016. https://patents.google.com/patent/WO2016115522A2/en?q=(%22portal+protein%22+AND+nanopore)&oq=%22portal+protein%22+AND+nanopore.

———. 2017. Protein variants for use as lipid bilayer-integrated nanopore, and methods thereof. World Intellectual Property Organization WO2017049101A2, filed 16 September 2016, and issued 23 March 2017. https://patents.google.com/patent/WO2017049101A2/en?q=(%22portal+protein%22+AND+nanopore)&oq=%22portal+protein%22+AND+nanopore.

Gutsmann, Thomas, Thomas Heimburg, Ulrich Keyser, Kozhinjampara R Mahendran, and Mathias Winterhalter. 2015. 'Protein Reconstitution into Freestanding Planar Lipid Membranes for Electrophysiological Characterization'. *Nature Protocols* 10 (1): 188–98. https://doi.org/10.1038/nprot.2015.003.

Hagan, James T., Brian S. Sheetz, Y.M. Nuwan D.Y. Bandara, Buddini I. Karawdeniya, Melissa A. Morris, Robert B. Chevalier, and Jason R. Dwyer. 2020. 'Chemically Tailoring Nanopores for Single-Molecule Sensing and Glycomics'. *Analytical and Bioanalytical Chemistry* 412 (25): 6639–54. https://doi.org/10.1007/s00216-020-02717-2.

Hall, Adam R., Andrew Scott, Dvir Rotem, Kunal K. Mehta, Hagan Bayley, and Cees Dekker. 2010a. 'Hybrid Pore Formation by Directed Insertion of Alpha Hemolysin into Solid-State Nanopores'. *Nature Nanotechnology* 5 (12): 874–77. https://doi.org/10.1038/nnano.2010.237.

———. 2010b. 'Hybrid Pore Formation by Directed Insertion of α-Haemolysin into Solid-State Nanopores'. *Nature Nanotechnology* 5 (12): 874–77. https://doi.org/10.1038/nnano.2010.237.

Haque, Farzin, Jennifer Lunn, Huaming Fang, David Smithrud, and Peixuan Guo. 2012. 'Real-Time Sensing and Discrimination of Single Chemicals Using the Channel of Phi29 DNA Packaging Nanomotor'. *ACS Nano* 6 (4): 3251–61. https://doi.org/10.1021/nn3001615.

HAQUE, Farzin, and Shaoying Wang. 2019. Nanopore assemblies and uses thereof. Canada CA3091126A1, filed 11 February 2019, and issued 15 August 2019. https://patents.google.com/patent/CA3091126A1/en?q=(%22portal+protein%22+AND+nanopore)&oq=%22portal+protein%22+AND+nanopore.

Haque, Farzin, Shaoying Wang, Lakmal Nishantha Jayasinghe, and Michael Jordan. 2022. Pore. United States US20220162568A1, filed 9 April 2020, and issued 26 May 2022. https://patents.google.com/patent/US20220162568A1/en?oq=(oxford+nanopore)+assignee:(Oxford+Nanopore+Technologies+Limited)+17%2f602%2c339.

Harrington, Roger F. 2003. *Introduction to Electromagnetic Engineering*. Repr. Mineola: Dover Publ.

Hawkins, Dorothy. 2022. 'Cryo-EM Analysis of DNA Packaging by the HK97 Bacteriophage'. York, UK: University of York.

Hawkins, Dorothy E. D. P., Oliver W. Bayfield, Herman K. H. Fung, Daniel N. Grba, Alexis Huet, James F. Conway, and Alfred A. Antson. 2023. 'Insights into a Viral Motor: The Structure of the HK97 Packaging Termination Assembly'. *Nucleic Acids Research* 51 (13): 7025–35. https://doi.org/10.1093/nar/gkad480.

He, Liqun, Martin Charron, Philipp Mensing, Kyle Briggs, Jonathan Adams, Hendrick de Haan, and Vincent Tabard-Cossa. 2023. 'DNA Origami Characterized via a Solid-State Nanopore: Insights into Nanostructure Dimensions, Rigidity and Yield'. *Nanoscale* 15 (34): 14043–54. https://doi.org/10.1039/D3NR01873C.

Heinz, Christian, Harald Engelhardt, and Michael Niederweis. 2003. 'The Core of the Tetrameric Mycobacterial Porin MspA Is an Extremely Stable Beta-Sheet Domain'. *The Journal of Biological Chemistry* 278 (10): 8678–85. https://doi.org/10.1074/jbc.M212280200.

Hendrix, R W. 1978. 'Symmetry Mismatch and DNA Packaging in Large Bacteriophages.' *Proceedings of the National Academy of Sciences of the United States of America* 75 (10): 4779–83.

Herzik, Mark A. 2021. 'Setting Up Parallel Illumination on the Talos Arctica for High-Resolution Data Collection'. In *cryoEM*, edited by Tamir Gonen and Brent L. Nannenga, 2215:125–44. Methods in Molecular Biology. New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-0966-8_6.

Hille, Bertil. 1968. 'Pharmacological Modifications of the Sodium Channels of Frog Nerve'. *The Journal of General Physiology* 51 (2): 199–219. https://doi.org/10.1085/jgp.51.2.199.

Hjörleifsdottir, Sigridur, Arnthor Aevarsson, Gudmundur O. Hreggvidsson, Olafur H. Fridjonsson, and Jakob K. Kristjansson. 2014. 'Isolation, Growth and Genome of the Rhodothermus RM378 Thermophilic Bacteriophage'. *Extremophiles: Life Under Extreme Conditions* 18 (2): 261–70. https://doi.org/10.1007/s00792-013-0613-x.

Hjörleifsdottir, Sigridur, Gudmundur O. Hreggvidsson, Olafur H. Fridjonsson, Arnthor Aevarsson, and Jakob K. Kristjansson. 2002. Bacteriophage RM 378 of a thermophilic host organism. United States US6492161B1, filed 1 June 2000, and issued 10 December 2002. https://patents.google.com/patent/US6492161B1/en.

Hofmeister, Franz. 1888. 'Zur Lehre von der Wirkung der Salze: Zweite Mittheilung'. *Archiv für Experimentelle Pathologie und Pharmakologie* 24 (4–5): 247–60. https://doi.org/10.1007/BF01918191.

Housmans, Joëlle A. J., Guiqin Wu, Joost Schymkowitz, and Frederic Rousseau. 2023. 'A Guide to Studying Protein Aggregation'. *The FEBS Journal* 290 (3): 554–83. https://doi.org/10.1111/febs.16312.

Hout, Michiel van den, Adam R. Hall, Meng Yue Wu, Henny W. Zandbergen, Cees Dekker, and Nynke H. Dekker. 2010. 'Controlling Nanopore Size, Shape and Stability'. *Nanotechnology* 21 (11): 115304. https://doi.org/10.1088/0957-4484/21/11/115304.

'How to Measure the Size of Your Nanopore Electrically'. n.d. Northernnanopore. Accessed 5 July 2024. https://www.solidstatenanopore.com/post/how-to-measure-the-size-of-your-nanopore-electrically.

Hoyt, Savannah J., Jessica M. Storer, Gabrielle A. Hartley, Patrick G. S. Grady, Ariel Gershman, Leonardo G. de Lima, Charles Limouse, et al. 2022. 'From Telomere to Telomere: The Transcriptional and Epigenetic State of Human Repeat Elements'. *Science (New York, N.Y.)* 376 (6588): eabk3112. https://doi.org/10.1126/science.abk3112.

Huang, Gang, Arnout Voet, and Giovanni Maglia. 2019. 'FraC Nanopores with Adjustable Diameter Identify the Mass of Opposite-Charge Peptides with 44 Dalton Resolution'. *Nature Communications* 10 (February):835. https://doi.org/10.1038/s41467-019-08761-6.

Huynh, Kathy, and Carrie L. Partch. 2015. 'Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay'. *Current Protocols in Protein Science* 79 (February):28.9.1-28.9.14. https://doi.org/10.1002/0471140864.ps2809s79.

Hyun, Seok-Hee, Hee-kwon Kim, Jong-Mok Kim, and David H. Thompson. 2010. 'Oriented Insertion of Phi29 N-Hexahistidine-Tagged Gp10 Connector Protein Assemblies into C20BAS Bolalipid Membrane Vesicles'. *Journal of the American Chemical Society* 132 (48): 17053–55. https://doi.org/10.1021/ja104204z.

Iacovache, Ioan, Sacha De Carlo, Nuria Cirauqui, Matteo Dal Peraro, F. Gisou van der Goot, and Benoît Zuber. 2016. 'Cryo-EM Structure of Aerolysin Variants Reveals a Novel Protein Fold and the Pore-Formation Process'. *Nature Communications* 7 (July):12062. https://doi.org/10.1038/ncomms12062.

Im, JongOne, Stuart Lindsay, Xu Wang, and Peiming Zhang. 2019. 'Single Molecule Identification and Quantification of Glycosaminoglycans Using Solid-State Nanopores'. *ACS Nano* 13 (6): 6308–18. https://doi.org/10.1021/acsnano.9b00618.

Israelachvili, Jacob Nissim. 2011. *Intermolecular and Surface Forces*. 3rd ed. Burlington (Mass.): Academic press.

Jain, Miten, Ian Fiddes, Karen H. Miga, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2015. 'Improved Data Analysis for the MinION Nanopore Sequencer'. *Nature Methods* 12 (4): 351–56. https://doi.org/10.1038/nmeth.3290.

Jain, Miten, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani, John R Tyson, et al. 2018. 'Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads'. *Nature Biotechnology* 36 (4): 338–45. https://doi.org/10.1038/nbt.4060.

James Hall. 1975. 'Access Resistance of a Small Circular Pore'. *The Journal of General Physiology* 66 (4): 531–32. https://doi.org/10.1085/jgp.66.4.531.

Järlebark, J. 2024. 'Polymer Brush Functionalised Nanopore Sensors'. Gothenburg, Sweden: Chalmers University of Technology. https://research.chalmers.se/publication/540075/file/540075_Fulltext.pdf.

Jarrold, Martin F. 2022. 'Applications of Charge Detection Mass Spectrometry in Molecular Biology and Biotechnology'. *Chemical Reviews* 122 (8): 7415–41. https://doi.org/10.1021/acs.chemrev.1c00377.

Javed, Abid, Hugo Villanueva, Shadikejiang Shataer, Sara Vasciaveo, Renos Savva, and Elena V. Orlova. 2021. 'Cryo-EM Structures of Two Bacteriophage Portal Proteins Provide Insights for Antimicrobial Phage Engineering'. *Viruses* 13 (12). https://doi.org/10.3390/v13122532.

Jayasinghe, Lakmal Nishantha, Elizabeth Jayne Wallace, Pratik Raj Singh, Richard George Hambley, Michael Jordan, and Han Remaut. 2018. Transmembrane Pore Consisting of Two Csgg Pores, issued 22 November 2018. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2018211241.

Jekow, Petra, Joachim Behlke, Willem Tichelaar, Rudi Lurz, Manuela Regalla, Winfried Hinrichs, and Paulo Tavares. 1999. 'Effect of the Ionic Environment on the Molecular Structure of Bacteriophage SPP1 Portal Protein'. *European Journal of Biochemistry* 264 (3): 724–35. https://doi.org/10.1046/j.1432-1327.1999.00601.x.

Ji, Zhouxiang, Michael Jordan, Lakmal Jayasinghe, and Peixuan Guo. 2020. 'Insertion of Channel of Phi29 DNA Packaging Motor into Polymer Membrane for High-Throughput Sensing'.

Nanomedicine: Nanotechnology, Biology and Medicine 25 (April):102170. https://doi.org/10.1016/j.nano.2020.102170.

Ji, Zhouxiang, Xinqi Kang, Shaoying Wang, and Peixuan Guo. 2018. 'Nano-Channel of Viral DNA Packaging Motor as Single Pore to Differentiate Peptides with Single Amino Acid Difference'. *Biomaterials* 182 (November):227–33. https://doi.org/10.1016/j.biomaterials.2018.08.005.

Jing, Peng. 2020. Proteoliposome and production method thereof. United States US10859561B2, filed 20 September 2018, and issued 8 December 2020. https://patents.google.com/patent/US10859561B2/en?q=(%22portal+protein%22+AND+nanopore)&oq=%22portal+protein%22+AND+nanopore.

Jing, Peng, Benjamin Burris, and Mauricio Cortes. 2021. 'The PLB Measurement for the Connector in Phi29 Bacteriophage Reveals the Function of Its Channel Loop'. *Biophysical Journal* 120 (9): 1650–64. https://doi.org/10.1016/j.bpj.2021.02.043.

Jing, Peng, Hallel Paraiso, and Benjamin Burris. 2016. 'Highly Efficient Integration of the Viral Portal Proteins from Different Types of Phages into Planar Bilayers for the Black Lipid Membrane Analysis'. *Molecular BioSystems* 12 (2): 480–89. https://doi.org/10.1039/C5MB00573F.

Johri, Shaili, Jitesh Solanki, Vito Adrian Cantu, Sam R. Fellows, Robert A. Edwards, Isabel Moreno, Asit Vyas, and Elizabeth A. Dinsdale. 2019. '"Genome Skimming" with the MinION Hand-Held Sequencer Identifies CITES-Listed Shark Species in India's Exports Market'. *Scientific Reports* 9 (1): 4476. https://doi.org/10.1038/s41598-019-40940-9.

Jones, David T., and Domenico Cozzetto. 2015. 'DISOPRED3: Precise Disordered Region Predictions with Annotated Protein-Binding Activity'. *Bioinformatics* 31 (6): 857–63. https://doi.org/10.1093/bioinformatics/btu744.

Jones, Grinnell, and Malcolm Dole. 1929. 'THE VISCOSITY OF AQUEOUS SOLUTIONS OF STRONG ELECTROLYTES WITH SPECIAL REFERENCE TO BARIUM CHLORIDE'. *Journal of the American Chemical Society* 51 (10): 2950–64. https://doi.org/10.1021/ja01385a012.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. 'Highly Accurate Protein Structure Prediction with AlphaFold'. *Nature* 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Jurrus, Elizabeth, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E. Felberg, David H. Brookes, et al. 2018. 'Improvements to the APBS Biomolecular Solvation Software Suite'. *Protein Science* 27 (1): 112–28. https://doi.org/10.1002/pro.3280.

Kang, Joon S., Xueting Zhou, Yun-Tao Liu, Kaituo Wang, and Z. Hong Zhou. 2023. 'Theoretical Framework and Experimental Solution for the Air-Water Interface Adsorption Problem in cryoEM'. *Biophysics Reports* 9 (4): 215–29. https://doi.org/10.52601/bpr.2023.230008.

Karawdeniya, Buddini Iroshika, Y. M. Nuwan D. Y. Bandara, Jonathan W. Nichols, Robert B. Chevalier, and Jason R. Dwyer. 2018. 'Surveying Silicon Nitride Nanopores for Glycomics and Heparin Quality Assurance'. *Nature Communications* 9 (1): 3278. https://doi.org/10.1038/s41467-018-05751-y.

Kasianowicz, John J., Eric Brandin, Daniel Branton, and David W. Deamer. 1996. 'Characterization of Individual Polynucleotide Molecules Using a Membrane Channel'. *Proceedings of the National Academy of Sciences* 93 (24): 13770–73. https://doi.org/10.1073/pnas.93.24.13770.

Kasumi Tsustumi. 2023. 'Fabrication of Sub-10 Nm Solid-State Nanopores by Electrical Breakdown'. Sweden: KTH Royal Institute of Technology. https://www.diva-portal.org/smash/get/diva2:1750953/FULLTEXT01.pdf.

Kim, Hyun-Mi, Min-Hyun Lee, and Ki-Bum Kim. 2011. 'Theoretical and Experimental Study of Nanopore Drilling by a Focused Electron Beam in Transmission Electron Microscopy'. *Nanotechnology* 22 (27): 275303. https://doi.org/10.1088/0957-4484/22/27/275303.

Kim, M. J., M. Wanunu, D. C. Bell, and A. Meller. 2006. 'Rapid Fabrication of Uniformly Sized Nanopores and Nanopore Arrays for Parallel DNA Analysis'. *Advanced Materials* 18 (23): 3149–53. https://doi.org/10.1002/adma.200601191.

Kim, Min Jun, Ben McNally, Kazuyoshi Murata, and Amit Meller. 2007. 'Characteristics of Solid-State Nanometre Pores Fabricated Using a Transmission Electron Microscope'. *Nanotechnology* 18 (20): 205302. https://doi.org/10.1088/0957-4484/18/20/205302.

Kim, Soo Hyun, Han Ju Yoo, Eun Ji Park, and Dong Hee Na. 2021. 'Nano Differential Scanning Fluorimetry-Based Thermal Stability Screening and Optimal Buffer Selection for Immunoglobulin G'. *Pharmaceuticals* 15 (1): 29. https://doi.org/10.3390/ph15010029.

Kirby, Brian J. 2010. *Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices*. New York: Cambridge University Press.

Kovarik, Michelle L., Kaimeng Zhou, and Stephen C. Jacobson. 2009. 'Effect of Conical Nanopore Diameter on Ion Current Rectification'. *The Journal of Physical Chemistry. B* 113 (49): 15960–66. https://doi.org/10.1021/jp9076189.

Kowalczyk, Stefan W, Alexander Y Grosberg, Yitzhak Rabin, and Cees Dekker. 2011. 'Modeling the Conductance and DNA Blockade of Solid-State Nanopores'. *Nanotechnology* 22 (31): 315101. https://doi.org/10.1088/0957-4484/22/31/315101.

Krissinel, Evgeny, and Kim Henrick. 2007. 'Inference of Macromolecular Assemblies from Crystalline State'. *Journal of Molecular Biology* 372 (3): 774–97. https://doi.org/10.1016/j.jmb.2007.05.022.

Kubota, Tomoya, Kento Lloyd, Naoto Sakashita, Seiya Minato, Kentaro Ishida, and Toshiyuki Mitsui. 2019. 'Clog and Release, and Reverse Motions of DNA in a Nanopore'. *Polymers* 11 (1): 84. https://doi.org/10.3390/polym11010084.

Kuhn, Jens H., Evelien M. Adriaenssens, and Andrew M. Kropinski. 2017. 'To Create a New Genus, Svunavirus, within the Family Myoviridae'. https://doi.org/10.13140/RG.2.2.34207.05285.

Kunz, W., J. Henle, and B. W. Ninham. 2004. '"Zur Lehre von Der Wirkung Der Salze" (about the Science of the Effect of Salts): Franz Hofmeister's Historical Papers'. *Current Opinion in Colloid & Interface Science* 9 (1): 19–37. https://doi.org/10.1016/j.cocis.2004.05.005.

Kwok, Harold, Kyle Briggs, and Vincent Tabard-Cossa. 2014. 'Nanopore Fabrication by Controlled Dielectric Breakdown'. Edited by Adam Hall. *PLoS ONE* 9 (3): e92880. https://doi.org/10.1371/journal.pone.0092880.

Laguerre, M., M. Saux, J. P. Dubost, and A. Carpy. 1997. 'MLPP: A Program for the Calculation of Molecular Lipophilicity Potential in Proteins'. *Pharmacy and Pharmacology Communications* 3 (5–6): 217–22. https://doi.org/10.1111/j.2042-7158.1997.tb00257.x.

Lange, Norbert Adolph, and John Aurie Dean. 1973. *Lange's Handbook of Chemistry*. Edited by John A. Dean. 11. ed. New York: McGraw-Hill.

Lau, Billy T., Alison Almeda, Marie Schauer, Madeline McNamara, Xiangqi Bai, Qingxi Meng, Mira Partha, et al. 2023. 'Single-Molecule Methylation Profiles of Cell-Free DNA in Cancer with Nanopore Sequencing'. *Genome Medicine* 15 (May):33. https://doi.org/10.1186/s13073-023-01178-3.

Lebedev, Andrey A., Margret H. Krause, Anabela L. Isidro, Alexei A. Vagin, Elena V. Orlova, Joanne Turner, Eleanor J. Dodson, Paulo Tavares, and Alfred A. Antson. 2007. 'Structural Framework for DNA Translocation via the Viral Portal Protein'. *The EMBO Journal* 26 (7): 1984–94. https://doi.org/10.1038/sj.emboj.7601643.

Leung, Chelsea, Kyle Briggs, Marie-Pier Laberge, Smile Peng, Matthew Waugh, and Vincent Tabard-Cossa. 2020. 'Mechanisms of Solid-State Nanopore Enlargement under Electrical Stress'. *Nanotechnology* 31 (44): 44LT01. https://doi.org/10.1088/1361-6528/aba86e.

Li, Jiali, Marc Gershow, Derek Stein, Eric Brandin, and J. A. Golovchenko. 2003. 'DNA Molecules and Configurations in a Solid-State Nanopore Microscope'. *Nature Materials* 2 (9): 611–15. https://doi.org/10.1038/nmat965.

Li, Shiyu, Shuangshuang Zeng, Chenyu Wen, Laurent Barbe, Maria Tenje, Zhen Zhang, Klas Hjort, and Shi-Li Zhang. 2020. 'Dynamics of DNA Clogging in Hafnium Oxide Nanopores'. *The Journal of Physical Chemistry. B* 124 (51): 11573–83. https://doi.org/10.1021/acs.jpcb.0c07756.

Li, Shiyu, Shuangshuang Zeng, Chenyu Wen, Zhen Zhang, Klas Hjort, and Shi-Li Zhang. 2022. 'Docking and Activity of DNA Polymerase on Solid-State Nanopores'. *ACS Sensors* 7 (5): 1476–83. https://doi.org/10.1021/acssensors.2c00216.

Li, Song, and William C. Purdy. 1992. 'Cyclodextrins and Their Applications in Analytical Chemistry'. *Chemical Reviews* 92 (6): 1457–70. https://doi.org/10.1021/cr00014a009.

Liang, Shengfa, Feibin Xiang, Zifan Tang, Reza Nouri, Xiaodong He, Ming Dong, and Weihua Guan. 2020. 'Noise in Nanopore Sensors: Sources, Models, Reduction, and Benchmarking'. *Nanotechnology and Precision Engineering* 3 (1): 9–17. https://doi.org/10.1016/j.npe.2019.12.008.

Liebes, Yael, Maria Drozdov, Yotam Y. Avital, Yaron Kauffmann, Hanna Rapaport, Wayne D. Kaplan, and Nurit Ashkenasy. 2010. 'Reconstructing Solid State Nanopore Shape from Electrical Measurements'. *Applied Physics Letters* 97 (22): 223105. https://doi.org/10.1063/1.3521411.

Lim, Seng Koon, Camilla Sandén, Robert Selegård, Bo Liedberg, and Daniel Aili. 2016. 'Tuning Liposome Membrane Permeability by Competitive Peptide Dimerization and Partitioning-Folding Interactions Regulated by Proteolytic Activity'. *Scientific Reports* 6 (1): 21123. https://doi.org/10.1038/srep21123.

Lin, Kabin, Zhongwu Li, Yi Tao, Kun Li, Haojie Yang, Jian Ma, Tie Li, Jingjie Sha, and Yunfei Chen. 2021. 'Surface Charge Density Inside a Silicon Nitride Nanopore'. *Langmuir* 37 (35): 10521–28. https://doi.org/10.1021/acs.langmuir.1c01504.

Lin, Lianbing, Wei Hong, Xiuling Ji, Jian Han, Li Huang, and Yunlin Wei. 2010. 'Isolation and Characterization of an Extremely Long Tail Thermus Bacteriophage from Tengchong Hot Springs in China'. *Journal of Basic Microbiology* 50 (5): 452–56. https://doi.org/10.1002/jobm.201000116.

Lin, Shiquan, Liang Xu, Aurelia Chi Wang, and Zhong Lin Wang. 2020. 'Quantifying Electron-Transfer in Liquid-Solid Contact Electrification and the Formation of Electric Double-Layer'. *Nature Communications* 11 (1): 399. https://doi.org/10.1038/s41467-019-14278-9.

Lisina, Sofia, Wali Inam, Mikko Huhtala, Fadak Howaili, Hongbo Zhang, and Jessica M. Rosenholm. 2023. 'Nano Differential Scanning Fluorimetry as a Rapid Stability Assessment Tool in the Nanoformulation of Proteins'. *Pharmaceutics* 15 (5): 1473. https://doi.org/10.3390/pharmaceutics15051473.

Liu, Bin, Suijie Wu, Qing Song, Xiaobo Zhang, and Lianhui Xie. 2006. 'Two Novel Bacteriophages of Thermophilic Bacteria Isolated from Deep-Sea Hydrothermal Fields'. *Current Microbiology* 53 (2): 163–66. https://doi.org/10.1007/s00284-005-0509-9.

Liu, Bin, Suijie Wu, and Lianhui Xie. 2010. 'Complete Genome Sequence and Proteomic Analysis of a Thermophilic Bacteriophage BV1'. *Acta Oceanologica Sinica* 29 (3): 84–89. https://doi.org/10.1007/s13131-010-0039-6.

Liu, Bin, and Xiaobo Zhang. 2008. 'Deep-Sea Thermophilic Geobacillus Bacteriophage GVE2 Transcriptional Profile and Proteomic Characterization of Virions'. *Applied Microbiology and Biotechnology* 80 (4): 697–707. https://doi.org/10.1007/s00253-008-1575-2.

Liu, Bin, Fengfeng Zhou, Suijie Wu, Ying Xu, and Xiaobo Zhang. 2009. 'Genomic and Proteomic Characterization of a Thermophilic Geobacillus Bacteriophage GBSV1'. *Research in Microbiology* 160 (2): 166–71. https://doi.org/10.1016/j.resmic.2008.12.005.

Liu, Hang, Rongliang Wang, Dejian Gu, Shengwei Tan, Hongwen Wu, and Quanjun Liu. 2017. 'Expression and Purification of a Novel Mycobacterial Porin MspA Mutant in *E. Coli*'. *Journal of Nanoscience and Nanotechnology* 17 (12): 9125–29. https://doi.org/10.1166/jnn.2017.13918.

Liu, Shao-Chuang, Yi-Lun Ying, Wei-Hua Li, Yong-Jing Wan, and Yi-Tao Long. 2021. 'Snapshotting the Transient Conformations and Tracing the Multiple Pathways of Single Peptide Folding Using a Solid-State Nanopore'. *Chemical Science* 12 (9): 3282–89. https://doi.org/10.1039/D0SC06106A.

Liu, Yun-Tao, Jonathan Jih, Xinghong Dai, Guo-Qiang Bi, and Z. Hong Zhou. 2019. 'Cryo-EM Structures of Herpes Simplex Virus Type 1 Portal Vertex and Packaged Genome'. *Nature* 570 (7760): 257–61. https://doi.org/10.1038/s41586-019-1248-6.

Lokareddy, Ravi K., Ying Hui Ko, Nathaniel Hong, Steven G. Doll, Marcin Paduch, Michael Niederweis, Anthony A. Kossiakoff, and Gino Cingolani. 2020. 'Recognition of an α-Helical Hairpin in P22 Large Terminase by a Synthetic Antibody Fragment'. *Acta Crystallographica. Section D, Structural Biology* 76 (Pt 9): 876–88. https://doi.org/10.1107/S2059798320009912.

Lokareddy, Ravi K., Rajeshwer S. Sankhala, Ankoor Roy, Pavel V. Afonine, Tina Motwani, Carolyn M. Teschke, Kristin N. Parent, and Gino Cingolani. 2017. 'Portal Protein Functions Akin to a DNA-Sensor That Couples Genome-Packaging to Icosahedral Capsid Maturation'. *Nature Communications* 8 (1): 14310. https://doi.org/10.1038/ncomms14310.

'London Stock Exchange Welcomes Oxford Nanopore Technologies Plc to the Main Market - Welcome Story At London Stock Exchange'. n.d. Accessed 4 September 2024. https://www.londonstockexchange.com/discover/news-and-insights/about:blank.

Lopatina, Anna, Sofia Medvedeva, Daria Artamonova, Matvey Kolesnik, Vasily Sitnik, Yaroslav Ispolatov, and Konstantin Severinov. 2019. 'Natural Diversity of CRISPR Spacers of Thermus: Evidence of Local Spacer Acquisition and Global Spacer Exchange'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1772): 20180092. https://doi.org/10.1098/rstb.2018.0092.

Loredo-Varela, Juan, Maria Chechik, Vladimir M. Levdikov, Ahmad Abd-El-Aziz, Leonid Minakhin, Konstantin Severinov, Callum Smits, and Alfred A. Antson. 2013. 'The Putative Small Terminase from the Thermophilic dsDNA Bacteriophage G20C Is a Nine-Subunit Oligomer'. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 69 (Pt 8): 876–79. https://doi.org/10.1107/S1744309113017016.

Luan, Weisha. 2013. 'Structural Studies on Bacteriophage Portal Proteins'. Phd, University of York. https://etheses.whiterose.ac.uk/4800/.

Łubkowska, Beata, Joanna Jeżewska-Frąckowiak, Ireneusz Sobolewski, and Piotr M. Skowron. 2021. 'Bacteriophages of Thermophilic "Bacillus Group" Bacteria-A Review'. *Microorganisms* 9 (7): 1522. https://doi.org/10.3390/microorganisms9071522.

Lyklema, J. 2001. *Fundamentals of Interface and Colloid Science. 2: Solid-Liquid Interfaces*. Repr. London: Academic Press.

Machón, Cristina, Montserrat Fàbrega-Ferrer, Daming Zhou, Ana Cuervo, José L. Carrascosa, David I. Stuart, and Miquel Coll. 2019. 'Atomic Structure of the Epstein-Barr Virus Portal'. *Nature Communications* 10 (1): 3891. https://doi.org/10.1038/s41467-019-11706-8.

Madeira, Fábio, Nandana Madhusoodanan, Joonheung Lee, Alberto Eusebi, Ania Niewielska, Adrian R N Tivey, Rodrigo Lopez, and Sarah Butcher. 2024. 'The EMBL-EBI Job Dispatcher Sequence Analysis Tools Framework in 2024'. *Nucleic Acids Research* 52 (W1): W521–25. https://doi.org/10.1093/nar/gkae241.

Magnusson, Anders O., Anna Szekrenyi, Henk-Jan Joosten, James Finnigan, Simon Charnock, and Wolf-Dieter Fessner. 2019. 'nanoDSF as Screening Tool for Enzyme Libraries and Biotechnology Development'. *The FEBS Journal* 286 (1): 184–204. https://doi.org/10.1111/febs.14696.

Manrao, Elizabeth A, Ian M Derrington, Andrew H Laszlo, Kyle W Langford, Matthew K Hopper, Nathaniel Gillgren, Mikhail Pavlenok, Michael Niederweis, and Jens H Gundlach. 2012. 'Reading DNA at Single-Nucleotide Resolution with a Mutant MspA Nanopore and Phi29 DNA Polymerase'. *Nature Biotechnology* 30 (4): 349–53. https://doi.org/10.1038/nbt.2171.

Marcus, Yizhak. 2009. 'Effect of Ions on the Structure of Water: Structure Making and Breaking'. *Chemical Reviews* 109 (3): 1346–70. https://doi.org/10.1021/cr8003828.

Matsui, Kazuma, Itaru Yanagi, Yusuke Goto, and Ken-ichi Takeda. 2015. 'Prevention of Dielectric Breakdown of Nanopore Membranes by Charge Neutralization'. *Scientific Reports* 5 (December):17819. https://doi.org/10.1038/srep17819.

McNulty, Reginald, Giovanni Cardone, Eddie B. Gilcrease, Timothy S. Baker, Sherwood R. Casjens, and John E. Johnson. 2018. 'Cryo-EM Elucidation of the Structure of Bacteriophage P22 Virions after Genome Release'. *Biophysical Journal* 114 (6): 1295–1301. https://doi.org/10.1016/j.bpj.2018.01.026.

Mehta, Shyam B., Jared S. Bee, Theodore W. Randolph, and John F. Carpenter. 2014. 'Partial Unfolding of a Monoclonal Antibody: Role of a Single Domain in Driving Protein Aggregation'. *Biochemistry* 53 (20): 3367–77. https://doi.org/10.1021/bi5002163.

Melnikov, Dmitriy V., Zachery K. Hulings, and Maria E. Gracheva. 2017. 'Electro-Osmotic Flow through Nanopores in Thin and Ultrathin Membranes'. *Physical Review E* 95 (6): 063105. https://doi.org/10.1103/PhysRevE.95.063105.

Meng, Elaine C., Thomas D. Goddard, Eric F. Pettersen, Greg S. Couch, Zach J. Pearson, John H. Morris, and Thomas E. Ferrin. 2023. 'UCSF ChimeraX: Tools for Structure Building and Analysis'. *Protein Science : A Publication of the Protein Society* 32 (11): e4792. https://doi.org/10.1002/pro.4792.

Meng, Elaine C, Eric F Pettersen, Gregory S Couch, Conrad C Huang, and Thomas E Ferrin. 2006. 'Tools for Integrated Sequence-Structure Analysis with UCSF Chimera'. *BMC Bioinformatics* 7 (1): 339. https://doi.org/10.1186/1471-2105-7-339.

Mirdita, Milot, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. 2022. 'ColabFold: Making Protein Folding Accessible to All'. *Nature Methods* 19 (6): 679–82. https://doi.org/10.1038/s41592-022-01488-1.

Moffatt, Barbara A., and F. William Studier. 1987. 'T7 Lysozyme Inhibits Transcription by T7 RNA Polymerase'. *Cell* 49 (2): 221–27. https://doi.org/10.1016/0092-8674(87)90563-0.

Mojtabavi, Mehrnaz, Sandra J. Greive, Alfred A. Antson, and Meni Wanunu. 2022. 'High-Voltage Biomolecular Sensing Using a Bacteriophage Portal Protein Covalently Immobilized within a Solid-State Nanopore'. *Journal of the American Chemical Society* 144 (49): 22540–48. https://doi.org/10.1021/jacs.2c08514.

Moore, Sean D., and Peter E. Prevelige. 2001. 'Structural Transformations Accompanying the Assembly of Bacteriophage P22 Portal Protein Rings in Vitro *'. *Journal of Biological Chemistry* 276 (9): 6779–88. https://doi.org/10.1074/jbc.M007702200.

Morgulis, Aleksandr, George Coulouris, Yan Raytselis, Thomas L. Madden, Richa Agarwala, and Alejandro A. Schäffer. 2008. 'Database Indexing for Production MegaBLAST Searches'. *Bioinformatics (Oxford, England)* 24 (16): 1757–64. https://doi.org/10.1093/bioinformatics/btn322.

Motwani, Tina, Ravi K. Lokareddy, Carmen A. Dunbar, Juliana R. Cortines, Martin F. Jarrold, Gino Cingolani, and Carolyn M. Teschke. 2017. 'A Viral Scaffolding Protein Triggers Portal Ring Oligomerization and Incorporation during Procapsid Assembly'. *Science Advances* 3 (7): e1700423. https://doi.org/10.1126/sciadv.1700423.

Muhammad Sajeer P, Simran, Pavan Nukala, and Manoj M. Varma. 2022. 'TEM Based Applications in Solid State Nanopores: From Fabrication to Liquid in-Situ Bio-Imaging'. *Micron* 162 (November):103347. https://doi.org/10.1016/j.micron.2022.103347.

Mukherjee, Amarshi, James L. Kizziah, N'Toia C. Hawkins, Mohamed O. Nasef, Laura K. Parker, and Terje Dokland. 2024. 'Structure of the Portal Complex from *Staphylococcus Aureus* Pathogenicity Island 1 Transducing Particles In Situ and In Isolation'. *Journal of Molecular Biology* 436 (4): 168415. https://doi.org/10.1016/j.jmb.2023.168415.

Mukherjee, Amarshi, James Kizziah, Laura Parker, and Terje Dokland. 2023. 'High-Resolution Cryo-EM Structure of Staphylococcus Aureus Bacteriophage 80α Portal Protein and SaPI1 Capsid'. *Microscopy and Microanalysis: The Official Journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada* 29 (29 Suppl 1): 925–26. https://doi.org/10.1093/micmic/ozad067.458.

Murase, Kazunori. 2022. ‘Cytolysin A (ClyA): A Bacterial Virulence Factor with Potential Applications in Nanopore Technology, Vaccine Development, and Tumor Therapy’. *Toxins* 14 (2): 78. https://doi.org/10.3390/toxins14020078.

‘nanoDSF’. n.d. *NanoTemper Technologies* (blog). Accessed 5 July 2024. https://nanotempertech.com/nanodsf/.

‘Nanolyzer Nanopore Analysis Software | Northern Nanopore Instruments’. n.d. Northernnanopore. Accessed 17 July 2024. https://www.solidstatenanopore.com/analysis-software.

Niederweis, Michael. 2003. ‘Mycobacterial Porins--New Channel Proteins in Unique Outer Membranes’. *Molecular Microbiology* 49 (5): 1167–77. https://doi.org/10.1046/j.1365-2958.2003.03662.x.

Niedzwiecki, David J., John Grazul, and Liviu Movileanu. 2010. ‘Single-Molecule Observation of Protein Adsorption onto an Inorganic Surface’. *Journal of the American Chemical Society* 132 (31): 10816–22. https://doi.org/10.1021/ja1026858.

Nimsamer, Pattaraporn, Vorthon Sawaswong, Pavit Klomkliew, Pornchai Kaewsapsak, Jiratchaya Puenpa, Yong Poovorawan, and Sunchai Payungporn. 2023. ‘“Nano COVID-19”: Nanopore Sequencing of Spike Gene to Identify SARS-CoV-2 Variants of Concern’. *Experimental Biology and Medicine* 248 (20): 1841–49. https://doi.org/10.1177/15353702231190931.

Nivala, Jeff, Douglas B. Marks, and Mark Akeson. 2013. ‘Unfoldase-Mediated Protein Translocation through an α-Hemolysin Nanopore’. *Nature Biotechnology* 31 (3): 247–50. https://doi.org/10.1038/nbt.2503.

Noble, Alex J., Hui Wei, Venkata P. Dandey, Zhening Zhang, Yong Zi Tan, Clinton S. Potter, and Bridget Carragher. 2018. ‘Reducing Effects of Particle Adsorption to the Air-Water Interface in Cryo-EM’. *Nature Methods* 15 (10): 793–95. https://doi.org/10.1038/s41592-018-0139-3.

Nogales, Eva, and Sjors H.W. Scheres. 2015. ‘Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity’. *Molecular Cell* 58 (4): 677–89. https://doi.org/10.1016/j.molcel.2015.02.019.

Nojima, Hiroshi, Atsushi Ikai, Tairo Oshima, and Haruhiko Noda. 1977. ‘Reversible Thermal Unfolding of Thermostable Phosphoglycerate Kinase. Thermostability Associated with Mean Zero Enthalpy Change’. *Journal of Molecular Biology* 116 (3): 429–42. https://doi.org/10.1016/0022-2836(77)90078-X.

‘Northern Nanopore Instruments User Manual’. 2022.

O’Donohue, Matthew, Madhav L. Ghimire, Sangyoup Lee, and Min Jun Kim. 2024. ‘Real-Time Monitoring of Ti(IV) Metal Ion Binding of Transferrin Using a Solid-State Nanopore’. *The Journal of Chemical Physics* 160 (4): 044906. https://doi.org/10.1063/5.0185590.

O’Donohue, Matthew, Jugal Saharia, Nuwan Bandara, Georgios Alexandrakis, and Min Jun Kim. 2023. ‘Use of a Solid-State Nanopore for Profiling the Transferrin Receptor Protein and Distinguishing between Transferrin Receptor and Its Ligand Protein’. *Electrophoresis* 44 (1–2): 349–59. https://doi.org/10.1002/elps.202200147.

Ohara, Masayuki, Masahiro Takinoue, and Ryuji Kawano. 2017. ‘Nanopore Logic Operation with DNA to RNA Transcription in a Droplet System’. *ACS Synthetic Biology* 6 (7): 1427–32. https://doi.org/10.1021/acssynbio.7b00101.

Ohi, Melanie, Ying Li, Yifan Cheng, and Thomas Walz. 2004. ‘Negative Staining and Image Classification — Powerful Tools in Modern Electron Microscopy’. *Biological Procedures Online* 6 (1): 23–34. https://doi.org/10.1251/bpo70.

Olasagasti, Felix, Kate R. Lieberman, Seico Benner, Gerald M. Cherf, Joseph M. Dahl, David W. Deamer, and Mark Akeson. 2010. ‘Replication of Individual DNA Molecules under Electronic Control Using a Protein Nanopore’. *Nature Nanotechnology* 5 (11): 798–806. https://doi.org/10.1038/nnano.2010.177.

Olek, Mateusz, Kevin Cowtan, Donovan Webb, Yuriy Chaban, and Peijun Zhang. 2022. ‘IceBreaker: Software for High-Resolution Single-Particle Cryo-EM with Non-Uniform Ice’.

Structure(London, England:1993) 30 (4): 522-531.e4. https://doi.org/10.1016/j.str.2022.01.005.

Olia, Adam S., Peter E. Prevelige, John E. Johnson, and Gino Cingolani. 2011. 'Three-Dimensional Structure of a Viral Genome-Delivery Portal Vertex'. *Nature Structural & Molecular Biology* 18 (5): 597–603. https://doi.org/10.1038/nsmb.2023.

'Oligo Analyzer'. n.d. Accessed 19 July 2023. https://eu.idtdna.com/calc/analyzer.

Orlov, Igor, Stéphane Roche, Sandrine Brasilès, Natalya Lukoyanova, Marie-Christine Vaney, Paulo Tavares, and Elena V. Orlova. 2022. 'CryoEM Structure and Assembly Mechanism of a Bacterial Virus Genome Gatekeeper'. *Nature Communications* 13 (1): 7283. https://doi.org/10.1038/s41467-022-34999-8.

Orlova, E. V., and H. R. Saibil. 2011. 'Structural Analysis of Macromolecular Assemblies by Electron Microscopy'. *Chemical Reviews* 111 (12): 7710–48. https://doi.org/10.1021/cr100353t.

Oukhaled, Abdelghani, Laurent Bacri, Manuela Pastoriza-Gallego, Jean-Michel Betton, and Juan Pelta. 2012. 'Sensing Proteins through Nanopores: Fundamental to Applications'. *ACS Chemical Biology* 7 (12): 1935–49. https://doi.org/10.1021/cb300449t.

Ouldali, Hadjer, Kumar Sarthak, Tobias Ensslen, Fabien Piguet, Philippe Manivet, Juan Pelta, Jan C. Behrends, Aleksei Aksimentiev, and Abdelghani Oukhaled. 2020. 'Electrical Recognition of the Twenty Proteinogenic Amino Acids Using an Aerolysin Nanopore'. *Nature Biotechnology* 38 (2): 176–81. https://doi.org/10.1038/s41587-019-0345-2.

Pace, C. Nick, J. Martin Scholtz, and Gerald R. Grimsley. 2014. 'Forces Stabilizing Proteins'. *FEBS Letters* 588 (14): 2177–84. https://doi.org/10.1016/j.febslet.2014.05.006.

Padilla-Sanchez, Victor, Song Gao, Hyung Rae Kim, Daisuke Kihara, Lei Sun, Michael G. Rossmann, and Venigalla B. Rao. 2014. 'Structure-Function Analysis of the DNA Translocating Portal of the Bacteriophage T4 Packaging Machine'. *Journal of Molecular Biology* 426 (5): 1019–38. https://doi.org/10.1016/j.jmb.2013.10.011.

Paez, Susana, and Martin Contreras. 1989. 'Densities and Viscosities of Binary Mixtures of 1-Propanol and 2-Propanol with Acetonitrile'. *Journal of Chemical & Engineering Data* 34 (4): 455–59. https://doi.org/10.1021/je00058a025.

Paillot, R. 1904. 'M. SMOLUCHOWSKI. — Contribution à La Théorie de l'endosmose Électrique et de Quelques Phénomènes Corrélatifs (Bulletin de l'Académie Des Sciences de Cracovie; Mars 1903)'. *Journal de Physique Théorique et Appliquée* 3 (1): 912–912. https://doi.org/10.1051/jphystap:019040030091201.

Parker, Michael W., J. Thomas Buckley, Johan P. M. Postma, Alec D. Tucker, Kevin Leonard, Franc Pattus, and Demetrius Tsernoglou. 1994. 'Structure of the Aeromonas Toxin Proaerolysin in Its Water-Soluble and Membrane-Channel States'. *Nature* 367 (6460): 292–95. https://doi.org/10.1038/367292a0.

Paulo, Gonçalo, Ke Sun, Giovanni Di Muccio, Alberto Gubbiotti, Blasco Morozzo della Rocca, Jia Geng, Giovanni Maglia, Mauro Chinappi, and Alberto Giacomello. 2023. 'Hydrophobically Gated Memristive Nanopores for Neuromorphic Applications'. *Nature Communications* 14 (1): 8390. https://doi.org/10.1038/s41467-023-44019-y.

Paulo, Gonçalo, Ke Sun, Giovanni di Muccio, Alberto Gubbiotti, Blasco Morozzo della Rocca, Jia Geng, Giovanni Maglia, Mauro Chinappi, and Alberto Giacomello. 2023. 'Hydrophobically Gated Memristive Nanopores for Neuromorphic Computing'. arXiv. http://arxiv.org/abs/2306.12187.

Pei, Jimin, Bong-Hyun Kim, and Nick V. Grishin. 2008. 'PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments'. *Nucleic Acids Research* 36 (7): 2295–2300. https://doi.org/10.1093/nar/gkn072.

Peng, Wei, Marcela de Souza Santos, Yang Li, Diana R. Tomchick, and Kim Orth. 2019. 'High-Resolution Cryo-EM Structures of the E. Coli Hemolysin ClyA Oligomers'. *PLoS ONE* 14 (5): e0213423. https://doi.org/10.1371/journal.pone.0213423.

Peng, Yuning, Huanrong Tang, Hao Xiao, Wenyuan Chen, Jingdong Song, Jing Zheng, and Hongrong Liu. 2024. 'Structures of Mature and Urea-Treated Empty Bacteriophage T5: Insights into Siphophage Infection and DNA Ejection'. *International Journal of Molecular Sciences* 25 (15): 8479. https://doi.org/10.3390/ijms25158479.

Perez-Ramirez, Bernardo, and John J. Steckert. 2005. 'Probing Reversible Self-Association of Therapeutic Proteins by Sedimentation Velocity in the Analytical Ultracentrifuge'. In *Therapeutic Proteins*, by C. Mark Smales and David C. James, 308:301–18. New Jersey: Humana Press. https://doi.org/10.1385/1-59259-922-2:301.

Piguet, Fabien, Francoise Discala, Marie-France Breton, Juan Pelta, Laurent Bacri, and Abdelghani Oukhaled. 2014. 'Electroosmosis through α-Hemolysin That Depends on Alkali Cation Type'. *The Journal of Physical Chemistry Letters* 5 (24): 4362–67. https://doi.org/10.1021/jz502360c.

Podobnik, Marjetka, Matic Kisovec, and Gregor Anderluh. 2017. 'Molecular Mechanism of Pore Formation by Aerolysin-like Proteins'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1726): 20160209. https://doi.org/10.1098/rstb.2016.0209.

Powell, Matthew R., Leah Cleary, Matthew Davenport, Kenneth J. Shea, and Zuzanna S. Siwy. 2011. 'Electric-Field-Induced Wetting and Dewetting in Single Hydrophobic Nanopores'. *Nature Nanotechnology* 6 (12): 798–802. https://doi.org/10.1038/nnano.2011.189.

'Protein Interfaces, Surfaces and Assemblies' Service PISA at the European Bioinformatics Institute'. n.d. http://www.ebi.ac.uk/pdbe/prot_int/pistart.html.

Pud, Sergii, Daniel Verschueren, Nikola Vukovic, Calin Plesa, Magnus P. Jonsson, and Cees Dekker. 2015. 'Self-Aligned Plasmonic Nanopores by Optically Controlled Dielectric Breakdown'. *Nano Letters* 15 (10): 7112–17. https://doi.org/10.1021/acs.nanolett.5b03239.

Punjani, Ali, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. 2017. 'cryoSPARC: Algorithms for Rapid Unsupervised Cryo-EM Structure Determination'. *Nature Methods* 14 (3): 290–96. https://doi.org/10.1038/nmeth.4169.

Rabinowitz, Jake, Martin A. Edwards, Elizabeth Whittier, Krishna Jayant, and Kenneth L. Shepard. 2019. 'Nanoscale Fluid Vortices and Nonlinear Electroosmotic Flow Drive Ion Current Rectification in the Presence of Concentration Gradients'. *The Journal of Physical Chemistry A* 123 (38): 8285–93. https://doi.org/10.1021/acs.jpca.9b04075.

Rai, Nithin, Marcelo Nöllmann, Bruno Spotorno, Giovanni Tassara, Olwyn Byron, and Mattia Rocco. 2005. 'SOMO (SOlution MOdeler) Differences between X-Ray- and NMR-Derived Bead Models Suggest a Role for Side Chain Flexibility in Protein Hydrodynamics'. *Structure (London, England: 1993)* 13 (5): 723–34. https://doi.org/10.1016/j.str.2005.02.012.

Ratinho, Laura, Laurent Bacri, Bénédicte Thiebot, Benjamin Cressiot, and Juan Pelta. 2024. 'Identification and Detection of a Peptide Biomarker and Its Enantiomer by Nanopore'. *ACS Central Science*, May. https://doi.org/10.1021/acscentsci.4c00020.

Rayner-Canham, Geoffrey, and Tina Overton. 2014. *Descriptive Inorganic Chemistry*. Sixth edition. New York, NY: W.H. Freeman and Company.

'Re: [Chimera-Users] Calculating Dipole Moment of a Protein in Pdb - Chimera-Users - RBVI Mailing Lists'. n.d. Accessed 23 August 2024. https://mail.cgl.ucsf.edu/mailman/archives/list/chimera-users@cgl.ucsf.edu/thread/WRX23CPKFNP6AWBNB5REWUXKHF33X2EQ/.

Remaut, Han, Sander Van Der Verren, Nani Van Gerven, Lakmal Jayasinghe, Elizabeth Wallace, Pratik Singh, Richard Hambley, Michael Jordan, and John Kilgour. 2019. Novel Protein Pores, issued 3 January 2019. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2019002893.

Restrepo-Pérez, Laura, Gang Huang, Peggy R. Bohländer, Nathalie Worp, Rienk Eelkema, Giovanni Maglia, Chirlmin Joo, and Cees Dekker. 2019. 'Resolving Chemical Modifications to a Single Amino Acid within a Peptide Using a Biological Nanopore'. *ACS Nano* 13 (12): 13668–76. https://doi.org/10.1021/acsnano.9b05156.

Restrepo-Pérez, Laura, Shalini John, Aleksei Aksimentiev, Chirlmin Joo, and Cees Dekker. 2017. 'SDS-Assisted Protein Transport through Solid-State Nanopores'. *Nanoscale* 9 (32): 11685–93. https://doi.org/10.1039/C7NR02450A.

Restrepo-Pérez, Laura, Chun Heung Wong, Giovanni Maglia, Cees Dekker, and Chirlmin Joo. 2019. 'Label-Free Detection of Post-Translational Modifications with a Nanopore'. *Nano Letters* 19 (11): 7957–64. https://doi.org/10.1021/acs.nanolett.9b03134.

Rivas, Felipe, Paul L. DeAngelis, Elaheh Rahbar, and Adam R. Hall. 2022. 'Optimizing the Sensitivity and Resolution of Hyaluronan Analysis with Solid-State Nanopores'. *Scientific Reports* 12 (March):4469. https://doi.org/10.1038/s41598-022-08533-1.

Rivas, Felipe, Osama K. Zahid, Heidi L. Reesink, Bridgette T. Peal, Alan J. Nixon, Paul L. DeAngelis, Aleksander Skardal, Elaheh Rahbar, and Adam R. Hall. 2018. 'Label-Free Analysis of Physiological Hyaluronan Size Distribution with a Solid-State Nanopore Sensor'. *Nature Communications* 9 (1): 1037. https://doi.org/10.1038/s41467-018-03439-x.

Robertson, Andrew D., and Kenneth P. Murphy. 1997. 'Protein Structure and the Energetics of Protein Stability'. *Chemical Reviews* 97 (5): 1251–68. https://doi.org/10.1021/cr960383c.

Robinson, Lloyd S., Elisabeth M. Ashman, Scott J. Hultgren, and Matthew R. Chapman. 2006. 'Secretion of Curli Fibre Subunits Is Mediated by the Outer Membrane-Localized CsgG Protein'. *Molecular Microbiology* 59 (3): 870–81. https://doi.org/10.1111/j.1365-2958.2005.04997.x.

Rocco, Mattia, and Olwyn Byron. 2015. 'Computing Translational Diffusion and Sedimentation Coefficients: An Evaluation of Experimental Data and Programs'. *European Biophysics Journal: EBJ* 44 (6): 417–31. https://doi.org/10.1007/s00249-015-1042-9.

Rohou, Alexis, and Nikolaus Grigorieff. 2015. 'CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs'. *Journal of Structural Biology*, Recent Advances in Detector Technologies and Applications for Molecular TEM, 192 (2): 216–21. https://doi.org/10.1016/j.jsb.2015.08.008.

Rosenberg, Alan H., Barbara N. Lade, Chui Dao-shan, Shu-Wha Lin, John J. Dunn, and F. William Studier. 1987. 'Vectors for Selective Expression of Cloned DNAs by T7 RNA Polymerase'. *Gene* 56 (1): 125–35. https://doi.org/10.1016/0378-1119(87)90165-X.

Rosenstein, Jacob K., Meni Wanunu, Christopher A. Merchant, Marija Drndic, and Kenneth L. Shepard. 2012. 'Integrated Nanopore Sensing Platform with Sub-Microsecond Temporal Resolution'. *Nature Methods* 9 (5): 487–92. https://doi.org/10.1038/nmeth.1932.

Roshan, Kamyar Akbari, Zifan Tang, and Weihua Guan. 2019. 'High Fidelity Moving Z-Score Based Controlled Breakdown Fabrication of Solid-State Nanopore'. *Nanotechnology* 30 (9): 095502. https://doi.org/10.1088/1361-6528/aaf48e.

Roth, Roland, Dirk Gillespie, Wolfgang Nonner, and Robert E. Eisenberg. 2008. 'Bubbles, Gating, and Anesthetics in Ion Channels'. *Biophysical Journal* 94 (11): 4282–98. https://doi.org/10.1529/biophysj.107.120493.

Rumble, John R., Thomas J. Brunno, and Maria J. Doa, eds. 2023. *CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*. 104nd edition, 2023–2024. CRC Handbook of Chemistry and Physics / Chemical Rubber Company, 104th edition (2023/2024). Boca Raton London New York: CRC Press.

Rūmnieks, Jānis, Tibor Füzik, and Kaspars Tārs. 2023. 'Structure of the *Borrelia* Bacteriophage ϕBB1 Procapsid'. *Journal of Molecular Biology* 435 (24): 168323. https://doi.org/10.1016/j.jmb.2023.168323.

Saharia, Jugal, Y. M. Nuwan D. Y. Bandara, Buddini I. Karawdeniya, George Alexandrakis, and Min Jun Kim. 2021. 'Assessment of 1/f Noise Associated with Nanopores Fabricated through Chemically Tuned Controlled Dielectric Breakdown'. *ELECTROPHORESIS* 42 (7–8): 899–909. https://doi.org/10.1002/elps.202000285.

Saharia, Jugal, Y. M. Nuwan D. Y. Bandara, Buddini I. Karawdeniya, Cassandra Hammond, George Alexandrakis, and Min Jun Kim. 2021. 'Modulation of Electrophoresis, Electroosmosis and

Diffusion for Electrical Transport of Proteins through a Solid-State Nanopore'. *RSC Advances* 11 (39): 24398–409. https://doi.org/10.1039/D1RA03903B.

Saharia, Jugal, Yapa Mudiyanselage Nuwan Dhananjaya Yapa Bandara, Buddini Iroshika Karawdeniya, Jason Rodger Dwyer, and Min Jun Kim. 2023. 'Over One Million DNA and Protein Events Through Ultra-Stable Chemically-Tuned Solid-State Nanopores'. *Small* n/a (n/a): 2300198. https://doi.org/10.1002/smll.202300198.

Sanchez-Garcia, Ruben, Josue Gomez-Blanco, Ana Cuervo, Jose Maria Carazo, Carlos Oscar S. Sorzano, and Javier Vargas. 2021. 'DeepEMhancer: A Deep Learning Solution for Cryo-EM Volume Post-Processing'. *Communications Biology* 4 (1): 1–8. https://doi.org/10.1038/s42003-021-02399-1.

Scheres, Sjors H. W. 2012. 'RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination'. *Journal of Structural Biology* 180 (3): 519–30. https://doi.org/10.1016/j.jsb.2012.09.006.

Scheres, Sjors H. W., Mikel Valle, Rafael Nuñez, Carlos O. S. Sorzano, Roberto Marabini, Gabor T. Herman, and Jose-Maria Carazo. 2005. 'Maximum-Likelihood Multi-Reference Refinement for Electron Microscopy Images'. *Journal of Molecular Biology* 348 (1): 139–49. https://doi.org/10.1016/j.jmb.2005.02.031.

Schoch, Conrad L., Stacy Ciufo, Mikhail Domrachev, Carol L. Hotton, Sivakumar Kannan, Rogneda Khovanskaya, Detlef Leipe, et al. 2020. 'NCBI Taxonomy: A Comprehensive Update on Curation, Resources and Tools'. *Database: The Journal of Biological Databases and Curation* 2020 (January):baaa062. https://doi.org/10.1093/database/baaa062.

Schrödinger, LLC. 2015. 'The PyMOL Molecular Graphics System, Version 1.8'.

Sen, Payel, Hiofan Hoi, and Manisha Gupta. 2021. 'Low Noise Hybrid Nanopore with Engineered OmpG and Bilayer MoS2'. *ACS Applied Bio Materials* 4 (7): 5416–24. https://doi.org/10.1021/acsabm.1c00095.

Severinov, Konstantin, Leonid Minakhin, Shun-Ichi Sekine, Anna Lopatina, and Shigeyuki Yokoyama. 2014. 'Molecular Basis of RNA Polymerase Promoter Specificity Switch Revealed through Studies of Thermus Bacteriophage Transcription Regulator'. *Bacteriophage* 4:e29399. https://doi.org/10.4161/bact.29399.

Shannon, R. D. 1976. 'Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides'. *Acta Crystallographica Section A* 32 (5): 751–67. https://doi.org/10.1107/S0567739476001551.

Shilling, Patrick J., Kiavash Mirzadeh, Alister J. Cumming, Magnus Widesheim, Zoe Köck, and Daniel O. Daley. 2020. 'Improved Designs for pET Expression Plasmids Increase Protein Production Yield in Escherichia Coli'. *Communications Biology* 3 (1): 1–8. https://doi.org/10.1038/s42003-020-0939-8.

Sicard, François, and A. Ozgur Yazaydin. 2022. 'Biohybrid Membrane Formation by Directed Insertion of Aquaporin into a Solid-State Nanopore'. *ACS Applied Materials & Interfaces* 14 (42): 48029–36. https://doi.org/10.1021/acsami.2c14250.

Sigworth, F. J. 1998. 'A Maximum-Likelihood Approach to Single-Particle Image Refinement'. *Journal of Structural Biology* 122 (3): 328–39. https://doi.org/10.1006/jsbi.1998.4014.

Simpson, Alan A, Yizhi Tao, Mohammed O Badasso, and Paul J Jardine. 2001. 'Structure Determination of the Head±tail Connector of Bacteriophage U29'.

Simpson, Alan A., Yizhi Tao, Petr G. Leiman, Mohammed O. Badasso, Yongning He, Paul J. Jardine, Norman H. Olson, et al. 2000. 'Structure of the Bacteriophage Φ29 DNA Packaging Motor'. *Nature* 408 (6813): 745. https://doi.org/10.1038/35047129.

Siwy, Z. S. 2006. 'Ion-Current Rectification in Nanopores and Nanotubes with Broken Symmetry'. *Advanced Functional Materials* 16 (6): 735–46. https://doi.org/10.1002/adfm.200500471.

Smeets, R. M. M., U. F. Keyser, M. Y. Wu, N. H. Dekker, and C. Dekker. 2006. 'Nanobubbles in Solid-State Nanopores'. *Physical Review Letters* 97 (8): 088101. https://doi.org/10.1103/PhysRevLett.97.088101.

Smeets, Ralph M. M., Ulrich F. Keyser, Diego Krapf, Meng-Yue Wu, Nynke H. Dekker, and Cees Dekker. 2006. 'Salt Dependence of Ion Transport and DNA Translocation through Solid-State Nanopores'. *Nano Letters* 6 (1): 89–95. https://doi.org/10.1021/nl052107w.

Song, Qing, and Xiaobo Zhang. 2008. 'Characterization of a Novel Non-Specific Nuclease from Thermophilic Bacteriophage GBSV1'. *BMC Biotechnology* 8 (April):43. https://doi.org/10.1186/1472-6750-8-43.

Soni, Neeraj, Noam Freundlich, Shilo Ohayon, Diana Huttner, and Amit Meller. 2022. 'Single-File Translocation Dynamics of SDS-Denatured, Whole Proteins through Sub-5 Nm Solid-State Nanopores'. *ACS Nano* 16 (7): 11405–14. https://doi.org/10.1021/acsnano.2c05391.

Soskine, Misha, Annemie Biesemans, Marc De Maeyer, and Giovanni Maglia. 2013. 'Tuning the Size and Properties of ClyA Nanopores Assisted by Directed Evolution'. *Journal of the American Chemical Society* 135 (36): 13456–63. https://doi.org/10.1021/ja4053398.

Soskine, Misha, Annemie Biesemans, Benjamien Moeyaert, Stephen Cheley, Hagan Bayley, and Giovanni Maglia. 2012. 'An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry'. *Nano Letters* 12 (9): 4895–4900. https://doi.org/10.1021/nl3024438.

Sou, Keitaro, Taro Endo, Shinji Takeoka, and Eishun Tsuchida. 2000. 'Poly(Ethylene Glycol)-Modification of the Phospholipid Vesicles by Using the Spontaneous Incorporation of Poly(Ethylene Glycol)-Lipid into the Vesicles'. *Bioconjugate Chemistry* 11 (3): 372–79. https://doi.org/10.1021/bc990135y.

Standring, Susan, ed. 2016. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. Forty-First edition. New York: Elsevier Limited.

Stefureac, Radu, Yi-tao Long, Heinz-Bernhard Kraatz, Peter Howard, and Jeremy S. Lee. 2006. 'Transport of α-Helical Peptides through α-Hemolysin and Aerolysin Pores'. *Biochemistry* 45 (30): 9172–79. https://doi.org/10.1021/bi0604835.

Stephenson, William, Roham Razaghi, Steven Busan, Kevin M. Weeks, Winston Timp, and Peter Smibert. 2022. 'Direct Detection of RNA Modifications and Structure Using Single-Molecule Nanopore Sequencing'. *Cell Genomics* 2 (2): 100097. https://doi.org/10.1016/j.xgen.2022.100097.

Stierlen, Aïcha, Sandra J. Greive, Laurent Bacri, Philippe Manivet, Benjamin Cressiot, and Juan Pelta. 2023. 'Nanopore Discrimination of Coagulation Biomarker Derivatives and Characterization of a Post-Translational Modification'. *ACS Central Science* 9 (2): 228–38. https://doi.org/10.1021/acscentsci.2c01256.

Storm, A. J., J. H. Chen, X. S. Ling, H. W. Zandbergen, and C. Dekker. 2003. 'Fabrication of Solid-State Nanopores with Single-Nanometre Precision'. *Nature Materials* 2 (8): 537–40. https://doi.org/10.1038/nmat941.

Straathof, Sabine, Giovanni Di Muccio, Maaruthy Yelleswarapu, Melissa Alzate Banguero, Carsten Wloka, Nieck Jordy van der Heide, Mauro Chinappi, and Giovanni Maglia. 2023. 'Protein Sizing with 15 Nm Conical Biological Nanopore YaxAB'. *ACS Nano* 17 (14): 13685–99. https://doi.org/10.1021/acsnano.3c02847.

Studier, F. W., and B. A. Moffatt. 1986. 'Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes'. *Journal of Molecular Biology* 189 (1): 113–30. https://doi.org/10.1016/0022-2836(86)90385-2.

Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. 'Use of T7 RNA Polymerase to Direct Expression of Cloned Genes'. *Methods in Enzymology* 185:60–89. https://doi.org/10.1016/0076-6879(90)85008-c.

Sun, Lei, Xinzheng Zhang, Song Gao, Prashant A. Rao, Victor Padilla-Sanchez, Zhenguo Chen, Siyang Sun, et al. 2015a. 'Cryo-EM Structure of the Bacteriophage T4 Portal Protein Assembly at near-Atomic Resolution'. *Nature Communications* 6 (1): 7548. https://doi.org/10.1038/ncomms8548.

———. 2015b. ‘Cryo-EM Structure of the Bacteriophage T4 Portal Protein Assembly at near-Atomic Resolution’. *Nature Communications* 6 (1): 7548. https://doi.org/10.1038/ncomms8548.

Tabeling, Patrick. 2023. *Introduction to Microfluidics*. 2nd ed. Oxford University PressOxford. https://doi.org/10.1093/oso/9780192845306.001.0001.

Tang, Guang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. 2007. ‘EMAN2: An Extensible Image Processing Suite for Electron Microscopy’. *Journal of Structural Biology*, Software tools for macromolecular microscopy, 157 (1): 38–46. https://doi.org/10.1016/j.jsb.2006.05.009.

Tang, Jia, and Feng Gai. 2008. ‘Dissecting the Membrane Binding and Insertion Kinetics of a pHLIP Peptide’. *Biochemistry* 47 (32): 8250–52. https://doi.org/10.1021/bi801103x.

Tang, Jinghua, Gabriel C. Lander, Adam S. Olia, Rui Li, Sherwood Casjens, Peter Prevelige, Gino Cingolani, Timothy S. Baker, and John E. Johnson. 2011. ‘Peering down the Barrel of a Bacteriophage Portal: The Genome Packaging and Release Valve in P22’. *Structure (London, England: 1993)* 19 (4): 496–502. https://doi.org/10.1016/j.str.2011.02.010.

Taniguchi, Masateru, Shohei Minami, Chikako Ono, Rina Hamajima, Ayumi Morimura, Shigeto Hamaguchi, Yukihiro Akeda, et al. 2021. ‘Combining Machine Learning and Nanopore Construction Creates an Artificial Intelligence Nanopore for Coronavirus Detection’. *Nature Communications* 12 (1): 3726. https://doi.org/10.1038/s41467-021-24001-2.

Thomas Young. 1804. ‘I. The Bakerian Lecture. Experiments and Calculations Relative to Physical Optics’. *Philosophical Transactions of the Royal Society of London* 94 (December):1–16. https://doi.org/10.1098/rstl.1804.0001.

Thuman-Commike, P. A., and W. Chiu. 2000. ‘Reconstruction Principles of Icosahedral Virus Structure Determination Using Electron Cryomicroscopy’. *Micron* 31 (6): 687–711. https://doi.org/10.1016/S0968-4328(99)00077-3.

Tiesinga, Eite, Carl J. Williams, Paul S. Julienne, Kevin M. Jones, Paul D. Lett, and William D. Phillips. 1996. ‘A Spectroscopic Determination of Scattering Lengths for Sodium Atom Collisions’. *Journal of Research of the National Institute of Standards and Technology* 101 (4): 505–20. https://doi.org/10.6028/jres.101.051.

Tim Sharpe. 2012. ‘Fitting Thermal and Chemical Denaturation Data’. Biozentrum Biophysics Facility. https://www.biozentrum.unibas.ch/fileadmin/redaktion/05_Facilities/01_Technology_Platforms/BF/Protocols/Denaturation_Fit.pdf.

Trick, Jemma L., Chen Song, E. Jayne Wallace, and Mark S. P. Sansom. 2017. ‘Voltage Gating of a Biomimetic Nanopore: Electrowetting of a Hydrophobic Barrier’. *ACS Nano* 11 (2): 1840–47. https://doi.org/10.1021/acsnano.6b07865.

Turner, Dann, Andrey N. Shkoporov, Cédric Lood, Andrew D. Millard, Bas E. Dutilh, Poliane Alfenas-Zerbini, Leonardo J. van Zyl, et al. 2023. ‘Abolishment of Morphology-Based Taxa and Change to Binomial Species Names: 2022 Taxonomy Update of the ICTV Bacterial Viruses Subcommittee’. *Archives of Virology* 168 (2): 74. https://doi.org/10.1007/s00705-022-05694-2.

Unni, Samir, Yong Huang, Robert M. Hanson, Malcolm Tobias, Sriram Krishnan, Wilfred W. Li, Jens E. Nielsen, and Nathan A. Baker. 2011. ‘Web Servers and Services for Electrostatics Calculations with APBS and PDB2PQR’. *Journal of Computational Chemistry* 32 (7): 1488–91. https://doi.org/10.1002/jcc.21720.

Unwin, P. N. T. 1974. ‘Electron Microscopy of the Stacked Disk Aggregate of Tobacco Mosaic Virus Protein: II. The Influence of Electron Irradiation on the Stain Distribution’. *Journal of Molecular Biology* 87 (4): 657–70. https://doi.org/10.1016/0022-2836(74)90076-X.

Urbaneja, María Angeles, Susana Rivas, José L. Carrascosa, and José María Valpuesta. 1994. ‘An Intrinsic-Tryptophan-Fluorescence Study of Phage Φ29 Connector/Nucleic Acid Interactions’. *European Journal of Biochemistry* 225 (2): 747–53. https://doi.org/10.1111/j.1432-1033.1994.00747.x.

Van der Verren, Sander E., Nani Van Gerven, Wim Jonckheere, Richard Hambley, Pratik Singh, John Kilgour, Michael Jordan, E. Jayne Wallace, Lakmal Jayasinghe, and Han Remaut. 2020. 'A Dual-Constriction Biological Nanopore Resolves Homonucleotide Sequences with High Fidelity'. *Nature Biotechnology* 38 (12): 1415–20. https://doi.org/10.1038/s41587-020-0570-8.

Vodyanoy, I., and S.M. Bezrukov. 1992. 'Sizing of an Ion Pore by Access Resistance Measurements'. *Biophysical Journal* 62 (1): 10–11. https://doi.org/10.1016/S0006-3495(92)81762-9.

Waduge, Pradeep, Rui Hu, Prasad Bandarkar, Hirohito Yamazaki, Benjamin Cressiot, Qing Zhao, Paul C. Whitford, and Meni Wanunu. 2017. 'Nanopore-Based Measurements of Protein Size, Fluctuations, and Conformational Changes'. *ACS Nano* 11 (6): 5706–16. https://doi.org/10.1021/acsnano.7b01212.

Wang, Shaoying, Farzin Haque, Piotr G. Rychahou, B. Mark Evers, and Peixuan Guo. 2013. 'Engineered Nanopore of Phi29 DNA-Packaging Motor for Real-Time Detection of Single Colon Cancer Specific Antibody in Serum'. *ACS Nano* 7 (11): 9814–22. https://doi.org/10.1021/nn404435v.

Wang, Wei, and Christopher J. Roberts. 2018. 'Protein Aggregation – Mechanisms, Detection, and Control'. *International Journal of Pharmaceutics* 550 (1): 251–68. https://doi.org/10.1016/j.ijpharm.2018.08.043.

Wang, Yiqian, and Xiaobo Zhang. 2008. 'Characterization of a Novel Portal Protein from Deep-Sea Thermophilic Bacteriophage GVE2'. *Gene* 421 (1–2): 61–66. https://doi.org/10.1016/j.gene.2008.05.015.

———. 2010. 'Genome Analysis of Deep-Sea Thermophilic Phage D6E'. *Applied and Environmental Microbiology* 76 (23): 7861–66. https://doi.org/10.1128/AEM.01270-10.

Wang, Yu, Kiran M. Patil, Shuanghong Yan, Panke Zhang, Weiming Guo, Yuqin Wang, Hong-Yuan Chen, Dennis Gillingham, and Shuo Huang. 2019. 'Nanopore Sequencing Accurately Identifies the Mutagenic DNA Lesion O6 -Carboxymethyl Guanine and Reveals Its Behavior in Replication'. *Angewandte Chemie (International Ed. in English)* 58 (25): 8432–36. https://doi.org/10.1002/anie.201902521.

Wang, Yupeng, Jianxuan Yuan, Haofeng Deng, Ziang Zhang, Qianli D. Y. Ma, Lingzhi Wu, and Lixing Weng. 2022. 'Procedural Data Processing for Single-Molecule Identification by Nanopore Sensors'. *Biosensors* 12 (12): 1152. https://doi.org/10.3390/bios12121152.

Wanunu, Meni, Alfred A. ANTSON, Sandra GREIVE, and Benjamin CRESSIOT. 2023. Lipid-free anchoring of thermophilic bacteriophage g20c portal adapter into solid-state nanopores. European Union EP4273550A2, filed 17 May 2019, and issued 8 November 2023. https://patents.google.com/patent/EP4273550A2/en?q=(%22portal+protein%22+AND+nanopore)&oq=%22portal+protein%22+AND+nanopore.

Wanunu, Meni, Tali Dadosh, Vishva Ray, Jingmin Jin, Larry McReynolds, and Marija Drndić. 2010. 'Rapid Electronic Detection of Probe-Specific microRNAs Using Thin Nanopore Sensors'. *Nature Nanotechnology* 5 (11): 807–14. https://doi.org/10.1038/nnano.2010.202.

Wasfi, Asma, Falah Awwad, and Ahmad I. Ayesh. 2018. 'Graphene-Based Nanopore Approaches for DNA Sequencing: A Literature Review'. *Biosensors & Bioelectronics* 119 (November):191–203. https://doi.org/10.1016/j.bios.2018.07.072.

Waugh, Matthew, Kyle Briggs, Dylan Gunn, Mathieu Gibeault, Simon King, Quinn Ingram, Aura Melissa Jimenez, et al. 2020. 'Solid-State Nanopore Fabrication by Automated Controlled Breakdown'. *Nature Protocols* 15 (1): 122–43. https://doi.org/10.1038/s41596-019-0255-2.

Waugh, Matthew, Autumn Carlsen, David Sean, Gary W. Slater, Kyle Briggs, Harold Kwok, and Vincent Tabard-Cossa. 2015. 'Interfacing Solid-State Nanopores with Gel Media to Slow DNA Translocations'. *ELECTROPHORESIS* 36 (15): 1759–67. https://doi.org/10.1002/elps.201400488.

Wei, Dahai, and Xiaobo Zhang. 2008. 'Identification and Characterization of a Single-Stranded DNA-Binding Protein from Thermophilic Bacteriophage GVE2'. *Virus Genes* 36 (1): 273–78. https://doi.org/10.1007/s11262-007-0170-4.

Wei, Ruoshan, Thomas G. Martin, Ulrich Rant, and Hendrik Dietz. 2012. 'DNA Origami Gatekeepers for Solid-State Nanopores'. *Angewandte Chemie International Edition* 51 (20): 4864–67. https://doi.org/10.1002/anie.201200688.

Wei, Zi-Xuan, Yi-Lun Ying, Meng-Yin Li, Jie Yang, Jia-Le Zhou, Hui-Feng Wang, Bing-Yong Yan, and Yi-Tao Long. 2019. 'Learning Shapelets for Improving Single-Molecule Nanopore Sensing'. *Analytical Chemistry* 91 (15): 10033–39. https://doi.org/10.1021/acs.analchem.9b01896.

Wendell, David, Peng Jing, Jia Geng, Varuni Subramaniam, Tae Jin Lee, Carlo Montemagno, and Peixuan Guo. 2009. 'Translocation of Double-Stranded DNA through Membrane-Adapted Phi29 Motor Protein Nanopores'. *Nature Nanotechnology* 4 (11): 765–72. https://doi.org/10.1038/nnano.2009.259.

Wichit, Angkana, Anothai Tangsumranjit, Tasana Pitaksuteepong, and Neti Waranuch. 2012. 'Polymeric Micelles of PEG-PE as Carriers of All-Trans Retinoic Acid for Stability Improvement'. *AAPS PharmSciTech* 13 (1): 336–43. https://doi.org/10.1208/s12249-011-9749-0.

Willems, Kherim, Dino Ruić, Florian L. R. Lucas, Ujjal Barman, Niels Verellen, Johan Hofkens, Giovanni Maglia, and Pol Van Dorpe. 2020. 'Accurate Modeling of a Biological Nanopore with an Extended Continuum Framework'. *Nanoscale* 12 (32): 16775–95. https://doi.org/10.1039/D0NR03114C.

Williams, David B., and C. Barry Carter. 2009. *Transmission Electron Microscopy*. Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-76501-3.

Williams, Lowri S., Vladimir M. Levdikov, Leonid Minakhin, Konstantin Severinov, and Alfred A. Antson. 2013. '12-Fold Symmetry of the Putative Portal Protein from the Thermus Thermophilus Bacteriophage G20C Determined by X-Ray Analysis'. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 69 (Pt 11): 1239–41. https://doi.org/10.1107/S174430911302486X.

Wong, Chiu Tai Andrew, and M. Muthukumar. 2010. 'Polymer Translocation through α-Hemolysin Pore with Tunable Polymer-Pore Electrostatic Interaction'. *The Journal of Chemical Physics* 133 (4): 045101. https://doi.org/10.1063/1.3464333.

Woodbury, Brianna M., Tina Motwani, Makayla N. Leroux, Lauren F. Barnes, Nicholas A. Lyktey, Sanchari Banerjee, Corynne L. Dedeo, Martin F. Jarrold, and Carolyn M. Teschke. 2022. 'Tryptophan Residues Are Critical for Portal Protein Assembly and Incorporation in Bacteriophage P22'. *Viruses* 14 (7): 1400. https://doi.org/10.3390/v14071400.

Woodson, Michael, Joshua Pajak, Bryon P. Mahler, Wei Zhao, Wei Zhang, Gaurav Arya, Mark A. White, Paul J. Jardine, and Marc C. Morais. 2021. 'A Viral Genome Packaging Motor Transitions between Cyclic and Helical Symmetry to Translocate dsDNA'. *Science Advances* 7 (19): eabc1955. https://doi.org/10.1126/sciadv.abc1955.

Wright, Elizabeth R., Cristina V. Iancu, William F. Tivol, and Grant J. Jensen. 2006. 'Observations on the Behavior of Vitreous Ice at ∼82 And∼12 K'. *Journal of Structural Biology* 153 (3): 241–52. https://doi.org/10.1016/j.jsb.2005.12.003.

Wu, Meng-Yue, Diego Krapf, Mathijs Zandbergen, Henny Zandbergen, and Philip E. Batson. 2005. 'Formation of Nanopores in a SiN⁄SiO2 Membrane with an Electron Beam'. *Applied Physics Letters* 87 (11): 113106. https://doi.org/10.1063/1.2043247.

Wu, Meng-Yue, Ralph M. M. Smeets, Mathijs Zandbergen, Ulrike Ziese, Diego Krapf, Philip E. Batson, Nynke H. Dekker, Cees Dekker, and Henny W. Zandbergen. 2009. 'Control of Shape and Material Composition of Solid-State Nanopores'. *Nano Letters* 9 (1): 479–84. https://doi.org/10.1021/nl803613s.

Wu, Xin, Fengwen Mu, Yinghui Wang, and Haiyan Zhao. 2018. 'Graphene and Graphene-Based Nanomaterials for DNA Detection: A Review'. *Molecules (Basel, Switzerland)* 23 (8): 2050. https://doi.org/10.3390/molecules23082050.

Xu, Jingwei, Dianhong Wang, Miao Gui, and Ye Xiang. 2019. 'Structural Assembly of the Tailed Bacteriophage Φ29'. *Nature Communications* 10 (1): 2366. https://doi.org/10.1038/s41467-019-10272-3.

Xu, Ruigang. 2017. 'Structural and Biochemical Studies of the Large Terminase Protein from Thermophilic Viruses'. Phd, University of York. https://etheses.whiterose.ac.uk/18917/.

Yan, Miaomiao, Anran Cai, Jing Li, Meixiu Xin, Mingying Liu, Chunhua Wang, and Guangcheng Wei. 2019. 'Preparation of β-CD-DPPE-Dox Nanomedicine and Its' Application as the Anticancer and Antitumor Drug'. *Scientific Reports* 9 (September):13670. https://doi.org/10.1038/s41598-019-50162-8.

Yan, Shuanghong, Jinyue Zhang, Yu Wang, Weiming Guo, Shanyu Zhang, Yao Liu, Jiao Cao, et al. 2021. 'Single Molecule Ratcheting Motion of Peptides in a Mycobacterium Smegmatis Porin A (MspA) Nanopore'. *Nano Letters* 21 (15): 6703–10. https://doi.org/10.1021/acs.nanolett.1c02371.

Yanagi, Itaru, Rena Akahori, and Ken-ichi Takeda. 2019. 'Stable Fabrication of a Large Nanopore by Controlled Dielectric Breakdown in a High-pH Solution for the Detection of Various-Sized Molecules'. *Scientific Reports* 9 (September):13143. https://doi.org/10.1038/s41598-019-49622-y.

———. 2023. 'Dwell Time Prolongation and Identification of Single Nucleotides Passing through a Solid-State Nanopore by Using Ammonium Sulfate Aqueous Solution'. *ACS Omega* 8 (23): 21285–92. https://doi.org/10.1021/acsomega.3c02703.

Yanagi, Itaru, Koji Fujisaki, Hirotaka Hamamura, and Ken-ichi Takeda. 2017. 'Thickness-Dependent Dielectric Breakdown and Nanopore Creation on Sub-10-Nm-Thick SiN Membranes in Solution'. *Journal of Applied Physics* 121 (4): 045301. https://doi.org/10.1063/1.4974286.

Yanagi, Itaru, Hirotaka Hamamura, Rena Akahori, and Ken-ichi Takeda. 2018. 'Two-Step Breakdown of a SiN Membrane for Nanopore Fabrication: Formation of Thin Portion and Penetration'. *Scientific Reports* 8 (1): 10129. https://doi.org/10.1038/s41598-018-28524-5.

Yanagi, Itaru, and Ken-ichi Takeda. 2020. 'Current–Voltage Characteristics of SiN Membranes in Solution'. *ACS Applied Electronic Materials* 2 (9): 2760–71. https://doi.org/10.1021/acsaelm.0c00479.

Yazdani, Mahdieh, Zhiguang Jia, and Jianhan Chen. 2020. 'Hydrophobic Dewetting in Gating and Regulation of Transmembrane Protein Ion Channels'. *The Journal of Chemical Physics* 153 (11): 110901. https://doi.org/10.1063/5.0017537.

Ying, Cuifeng, Tianji Ma, Lei Xu, and Mohsen Rahmani. 2022. 'Localized Nanopore Fabrication via Controlled Breakdown'. *Nanomaterials* 12 (14): 2384. https://doi.org/10.3390/nano12142384.

Ying, Cuifeng, Yuechuan Zhang, Yanxiao Feng, Daming Zhou, Deqiang Wang, Yinxiao Xiang, Wenyuan Zhou, Yongsheng Chen, Chunlei Du, and Jianguo Tian. 2016. '3D Nanopore Shape Control by Current-Stimulus Dielectric Breakdown'. *Applied Physics Letters* 109 (6): 063105. https://doi.org/10.1063/1.4960636.

Ying, Yi-Lun, Jie Yang, Fu-Na Meng, Shuang Li, Meng-Ying Li, and Yi-Tao Long. 2019. 'A Nanopore Phosphorylation Sensor for Single Oligonucleotides and Peptides'. *Research* 2019 (November). https://doi.org/10.34133/2019/1050735.

Young, Gavin, Nikolas Hundt, Daniel Cole, Adam Fineberg, Joanna Andrecka, Andrew Tyler, Anna Olerinyova, et al. 2018. 'Quantitative Mass Imaging of Single Molecules'. *Science (New York, N.Y.)* 360 (6387): 423–27. https://doi.org/10.1126/science.aar5839.

Yu, M. X., M. R. Slater, and H.-W. Ackermann. 2006. 'Isolation and Characterization of Thermus Bacteriophages'. *Archives of Virology* 151 (4): 663–79. https://doi.org/10.1007/s00705-005-0667-x.

Zhang, Long, Nicolas Burns, Zhouxiang Ji, Steven Sun, Susan L. Deutscher, William E. Carson, and Peixuan Guo. 2023. 'Nipple Fluid for Breast Cancer Diagnosis Using the Nanopore of Phi29 DNA-Packaging Motor'. *Nanomedicine : Nanotechnology, Biology, and Medicine* 48 (February):102642. https://doi.org/10.1016/j.nano.2022.102642.

Zhang, Ming, Chen Chen, Yanjing Zhang, and Jia Geng. 2022. 'Biological Nanopores for Sensing Applications'. *Proteins* 90 (10): 1786–99. https://doi.org/10.1002/prot.26308.

Zhang, Qi, Qiupeng Li, Xiuling Ji, Wei Hong, Zhiyang Dong, Yunlin Wei, and Lianbing Lin. 2014. 'Identification and Characterization of a Helicase-like Protein Encoded by a Thermus Siphoviridae Phage 4 Gene'. *Pakistan Journal of Pharmaceutical Sciences* 27 (3 Suppl): 703–11.

Zhang, Shengli, Gang Huang, Roderick Corstiaan Abraham Versloot, Bart Marlon Herwig Bruininks, Paulo Cesar Telles de Souza, Siewert-Jan Marrink, and Giovanni Maglia. 2021. 'Bottom-up Fabrication of a Proteasome-Nanopore That Unravels and Processes Single Proteins'. *Nature Chemistry* 13 (12): 1192–99. https://doi.org/10.1038/s41557-021-00824-w.

Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. 2000. 'A Greedy Algorithm for Aligning DNA Sequences'. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 7 (1–2): 203–14. https://doi.org/10.1089/10665270050081478.

Zhao, Hua. 2016. 'Protein Stabilization and Enzyme Activation in Ionic Liquids: Specific Ion Effects'. *Journal of Chemical Technology and Biotechnology (Oxford, Oxfordshire : 1986)* 91 (1): 25–50. https://doi.org/10.1002/jctb.4837.

Zheng, Shawn Q., Eugene Palovcak, Jean-Paul Armache, Kliment A. Verba, Yifan Cheng, and David A. Agard. 2017. 'MotionCor2: Anisotropic Correction of Beam-Induced Motion for Improved Cryo-Electron Microscopy'. *Nature Methods* 14 (4): 331–32. https://doi.org/10.1038/nmeth.4193.

Zhou, Kaimeng, John M. Perry, and Stephen C. Jacobson. 2011. 'Transport and Sensing in Nanofluidic Devices'. *Annual Review of Analytical Chemistry* 4 (1): 321–41. https://doi.org/10.1146/annurev-anchem-061010-113938.