The University Of Sheffield.

Structural and Functional Studies of the AHL Tripartite Pore Forming Toxin Proteins from *Aeromonas hydrophila*

Jason Wilson

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department of Molecular Biology and Biotechnology

September 2016

Structural and Functional Studies of the AHL Tripartite Pore Forming Toxin Proteins from *Aeromonas hydrophila*

Abstract

The Cytolysin A (ClyA) family of pore forming toxins include mono-, bi-, and tripartite members. Haemolysin BL (HBL-B,L_1,L_2) and non-haemolytic enterotoxin (NHE-A,B,C) are tripartite toxins where each subunit has a specific role in pore formation, and where pore formation is more complex than that of the monopartite ClyA.

Homology searches have identified a large number of NHE homologues in a diverse group of bacterial species, including *Aeromonas hydrophila.* Each component of the *Aeromonas hydrophila* tripartite toxin system, AhlA, AhlB, and AhlC, was successfully overexpressed and purified, and haemolysis assays confirmed that these NHE homologues are haemolytic, and require all three components for maximum lysis.

AhlC was successfully crystallised, and the structure was solved of two crystal forms to ~2.5 Å. The crystal structure of AhlC shows that the protein is a structural homologue to NheA, HblB, and ClyA, and consists of a bundle of four helices, and a long helical stalk and a hydrophobic hairpin. AhlC forms a tetramer to protect the hydrophobic hairpin from the solvent, unlike other ClyA family toxins that shield their membrane-spanning regions with conformational rearrangements.

Electron microscopy studies have shown that AhlB is able to form pre-pores of ~12 nm in diameter, and that AhlB and AhlC together form pores, although more readily than AhlB in isolation. AhlBC pores also localise to certain liposomes, leading to saturation of pores in these vesicles.

AhlB was successfully crystallised in a number of conditions, and analysis showed that many crystallisation conditions contained MPD. In MPD-containing crystallisation conditions, large unit cells were present, with a conserved 115 Å cell dimension, and a 10-fold non-crystallographic symmetry axis was identified in analysis of self rotations. These crystals are predicted, therefore, to contain the pre-pore observed by EM, and attempts are on-going to solve the structure.

Acknowledgements

Many people have been instrumental in helping me complete the work carried out during my PhD. Firstly I would like to thank Prof. Pete Artymiuk, who first inspired my love of Structural Biology, and then gave me the opportunity to come and work in his group studying pore forming toxins. Secondly I'd like to thank Dr. Patrick Baker for taking me in for the second half of my PhD, and for his guidance, support, and patience throughout these years. Thank you also to my Advisors Dr. Jeremy Craven and Dr. Rosie Staniforth for guidance each year. Finally I'd like to thank Prof. David Rice, Dr. John Rafferty, Dr. Svetlana Sedelnikova, Dr. Linda Britton, Dr. Claudine Bisson and Fiona Rodgers for your vast knowledge and (especially Sveta) patience throughout the duration of my work.

I'd also like to thank those who have helped in small ways that have made all the difference to my work. Those include Dr. Svet Tzokov, Dr. Julia Pridgeon, Dr. Simon Thorpe, and Prof. Per Einar Granum.

Next I would like to thank those friends I have made during my PhD, within the crystallography group, and within the department. Those in the crystallography group, namely Dr. Claudine Bisson, Hayley Owen, Dr. George Mobbs, Dr. Sophie Bliss, Sam Dix, with whom I've shared many 24 hour Diamond trips with, I'd like to thank you for being there, and for many great times together outside of the lab. I'd also like to acknowledge all of the other friends I have made in the crystallography group, of which everyone is included. You have set the bar very high for friendly work environments. To my friends in the department, of which luckily there are too many to mention individually, thank you to you all for trips to the pub, nights out, board games nights, Christmas dinners and Cinema trips. Special thanks, however, go to Oishik Banerji for help with EM work, and Liam Aubrey for help preparing liposomes.

I'd like to thank my girlfriend Ilona, who has been instrumental in the final year of my PhD, in keeping my sanity and providing many great distractions from the ever-increasing work and writing. Finally, I would like to thank the most important people to me, my Mum, my Dad, and my Brother, for their immense support and love over the years, and I also look forward very much to spending more time with you all soon.

Dedicated to Professor Pete Artymiuk

Table of Contents

Chapter 1 – Introduction and literature review

Pore forming toxins (PFTs) are proteins used both offensively and defensively by all kingdoms of life to disrupt biological membranes. They are of great interest to fundamental science in understanding how biological membranes can be disrupted by proteins. Further, these systems can aid in antimicrobial design, and a better understanding is leading to biotechnological exploitation of PFTs. This chapter discusses the current understanding of these toxins, beginning firstly with a broad introduction. The latter part of this chapter introduces a specific toxin system that is the focus for the research that has been undertaken and described in subsequent chapters.

1.1 Biological membranes

1.1.1 Cell membranes as selectively permeable barriers

Biological membranes are a vital part of cellular life, constituting the material that makes up cellular boundaries and forms a selectively permeable barrier (Alberts et al., 2002). In both prokaryotic and eukaryotic organisms, membranes separate the internal cytosolic cell contents from the external environment in which the cells exist. This separation allows biomolecules to be concentrated inside the cell, generating favourable environments for biological processes, as well as regulating matter, information, and energy flow (Martin and Russell, 2003). Membranes also allow an electrochemical gradient to be established, which forms the basis for energy production by means of ATP synthesis using the proton motive force, as well as driving solute transport against a chemical gradient to concentrate important biomolecules inside the cell (Nicholls and Ferguson, 2013).

Membranes are also required to further compartmentalise eukaryotic cells, forming specialised organelles that are encapsulated in lipid membranes, sometimes with complicated architectures (Munro, 2004). Organelles allow separate compartments to be established which may contain compounds that would be dangerous in the cytoplasm, such as compounds found in peroxisomes and lysosomes. Organelles also allow cargo to be trafficked about the cell in vesicles, protected from the cytoplasm by the membrane (Alberts et al., 2002).

1.1.2 Anatomy of phospholipid membranes

Biological membranes are formed mostly of phospholipids, which are organic molecules containing a hydrophilic phosphate head moiety, and a hydrophobic carbon chain tail (Figure 1.1). Three types of spontaneous lipid arrangements form naturally, micelles, liposomes, and lipid bilayers. The majority of lipid membranes in life are formed of bilayers. Bilayers bury the hydrophobic carbon chain inwards, and expose only the hydrophilic head to the solvent (Figure 1.2). The hydrophobic core then forms a barrier to many classes of hydrophilic molecules (ions, drugs, peptides, proteins), creating one of the most hydrophobic microenvironments found in nature (van der Goot, 2001).

The membrane acts as a semi-fluidic material, allowing lateral diffusion of lipids and proteins in the plane of the membrane. Transfer of hydrophilic moieties contained on membrane proteins, from one side of the bilayer to the other, are energetically unfavourable and so transfer is catalysed by chaperones and flippases.

Hydrophobic chain length, and head group size and identity induce curvature and form microenvironments within the membrane that concentrate lipid types and proteins in certain areas, forming domains and 'lipid rafts' (Brown and London, 2000; Lingwood and Simons, 2010). In this way, lipids can control the shape of the membrane, the local environment, and which proteins can exist within them (Figure 1.3).

1.1.3 Membrane disruption by proteins and peptides

Some proteins and peptides are able to overcome the hydrophobic barrier to polar molecule diffusion across the membrane, disrupting the hydrophobic microenvironment by two broad mechanisms.

The first mechanism for membrane disruption involves proteins working against the membrane composition by enzymatically altering the nature of the lipids constituting the membrane. By altering the lipid, for example by removing the hydrophilic head group, the barrier is disrupted as the head group provides stability to the bilayer. In eukaryotes,

phospholipases are mediators of signal transduction, by producing signalling molecules from lipid head groups. As such, this sub-class of membrane damaging proteins includes a large number of phospholipases used by eukaryotic organisms for signalling transduction (Eyster, 2007) .Pathogenic bacteria also use phospholipases, to overcome the bilayer barrier of host

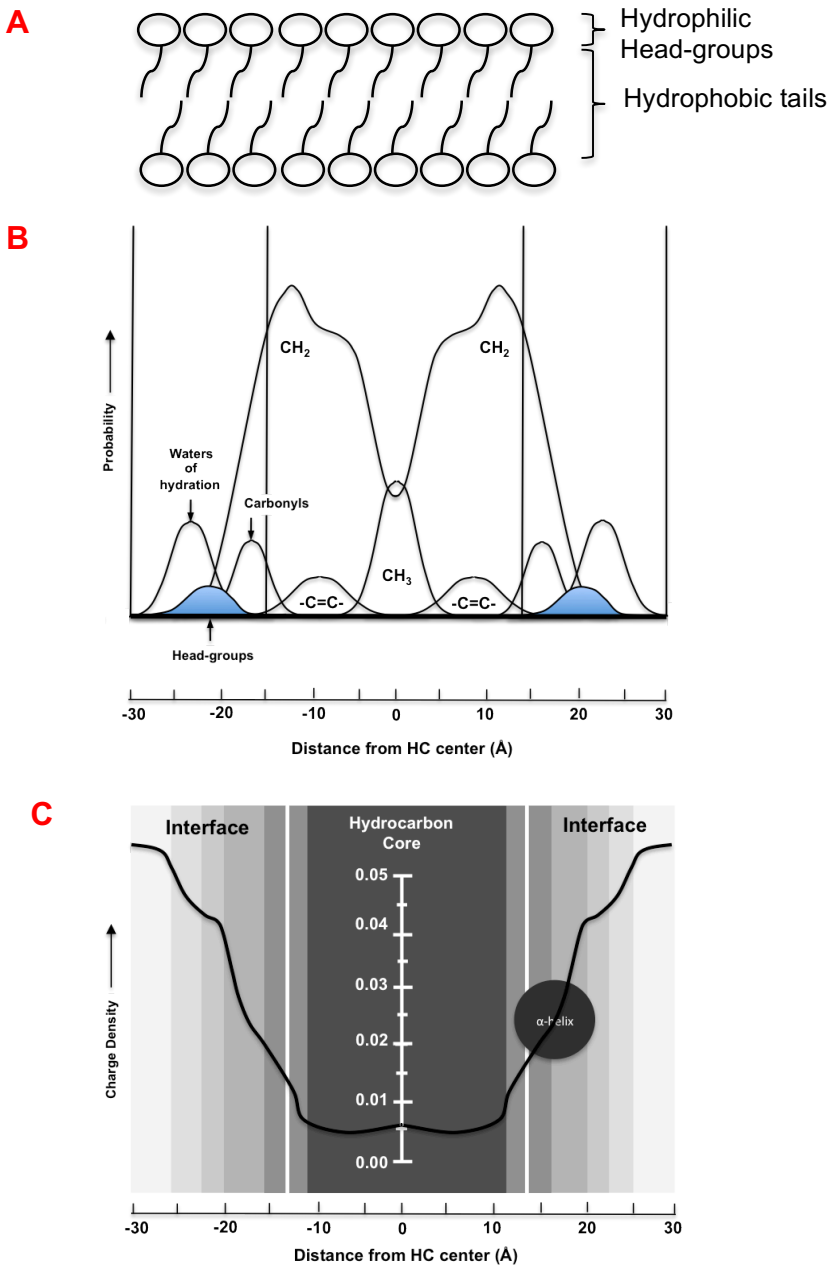


Figure 1.1. Lipid bilayer organisation and characteristics. (**A**) A simple schematic representation of a lipid bilayer, in which the hydrophilic head groups align outwardly to face solvent, and the hydrophobic core face inwards generating a hydrophobic barrier. (**B**) Schematic diagram showing the distribution of different lipid moieties across the distance of the lipid bilayer. The hydrophilic head group moieties e.g.. Phosphates and glycerol groups etc are shown in blue and are found ~ 20 Å from the hydrophobic core. The center of the bilayer contains the CH_3 group 0 Å from the hydrocarbon center. (**C**) Schematic diagram showing the charge density (unit) distributed across the lipid bilayer. The white vertical lines at ±14 Å indicate the most hydrophobic region of the bilayer, and the black circle shows how an amphipathic α-helix may straddle this hydrophobic region inserting partially into the membrane. Figures 1A and 1B were adapted from (van der Goot, 2001).

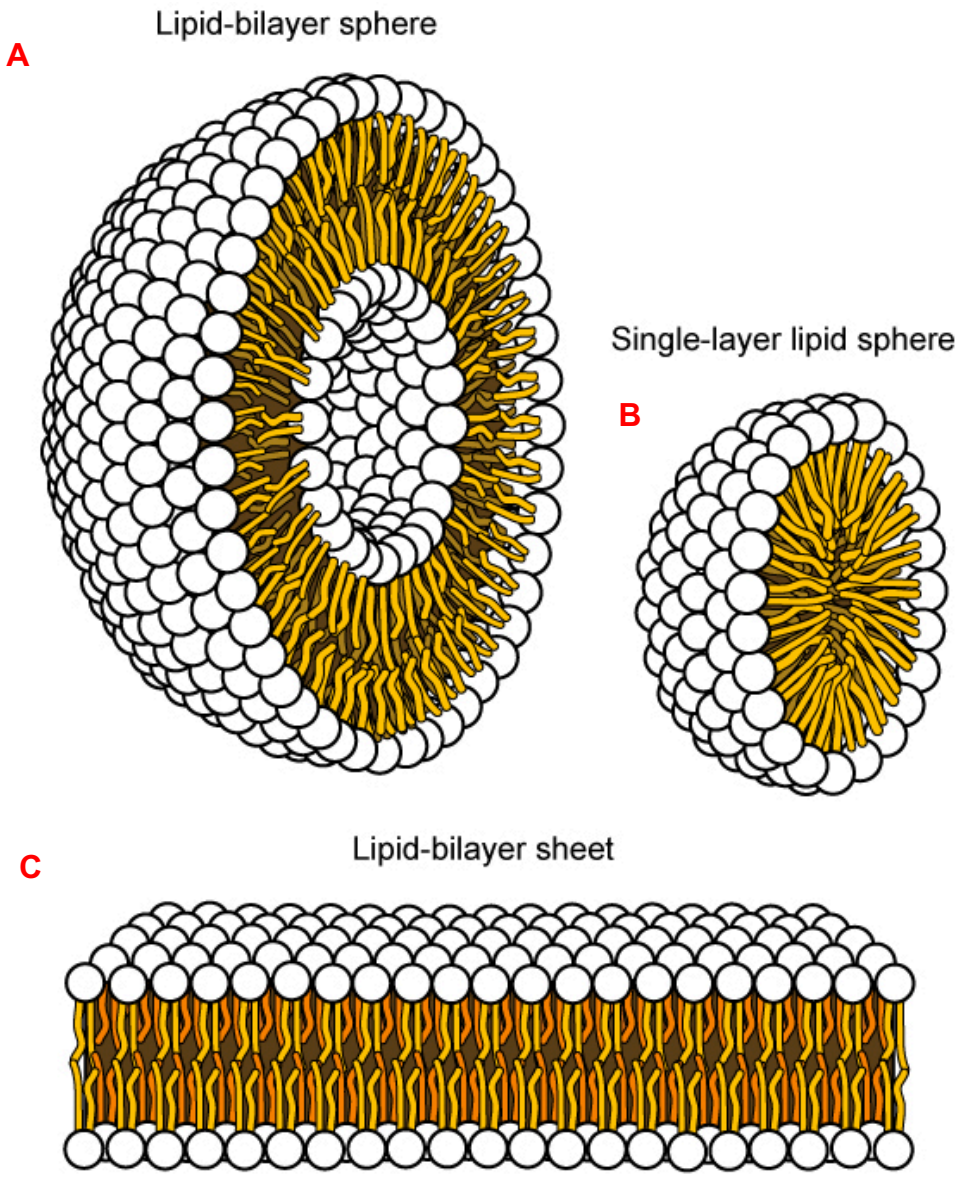


Figure 1.2. Three types of spontaneously forming lipid arrangements. Schematic representations of three spontaneously forming lipid species in solution. The hydrophilic head groups are shown as white circles, and the hydrophobic tails are shown as orange sticks. (**A**) A lipid bilayer sphere or liposome, where the outer layer protects the hydrophobic moieties from the solvent, whilst an inner layer encapsulates solvent and so forms an encapsulated sphere. (**B**) A lipid micelle, which contains the same outer sphere, but lacks the inner layer seen in liposomes, and instead all hydrophobic tails meet at the center to exclude solvent from the inside of the sphere. (**C**) A planar lipid bilayer, similar to that found making up the cell membrane, in which the structure of the hydrophobic tails induces little curvature, and so the membrane forms a plane rather than a sphere.

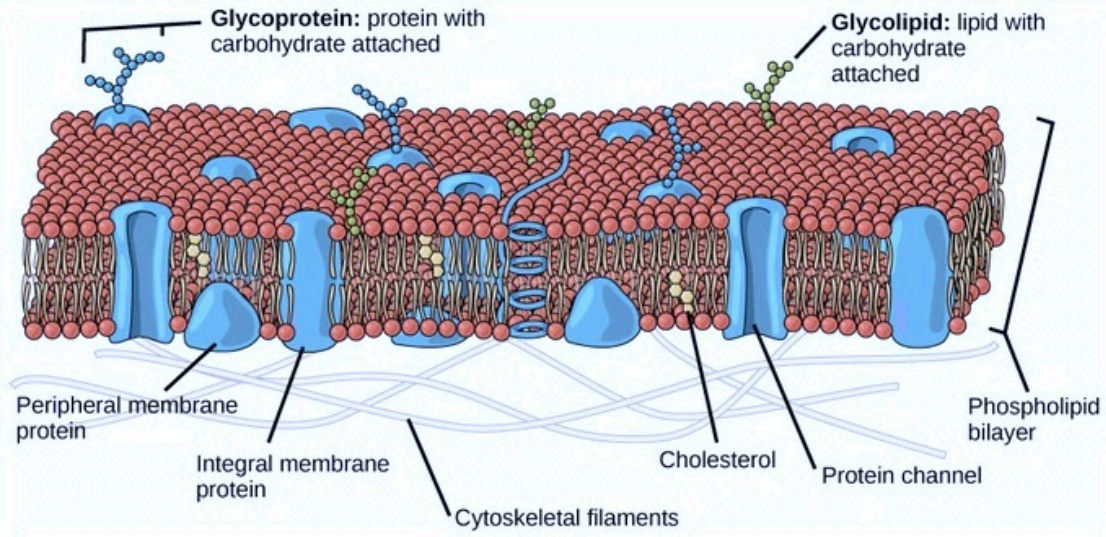


Figure 1.3. A cartoon model of a cellular membrane including common membrane features. A cartoon representation of the membrane of a eukaryotic cell, showing many of the molecules found at the membrane interface. The membrane concentrates such molecules as glycoproteins for signalling and cell adhesion, protein channels for import and export of molecules from the cell, cholesterol, to mediate rigidity in the membrane, and a large number of membrane proteins inserted either peripherally or integrally, brought to the membrane to function. Image modified from OpenStax Biology.

cells (Schmiel and Miller, 1999), and to mimic signal transduction, inducing pathological levels of signalling molecules to build up in the host cell (Sakurai et al., 2004).

The second mechanism of overcoming the membrane barrier involves proteins or peptides, which work with the membrane composition, inserting through the membrane and creating a channel in which polar molecules can pass. These channels can either allow polar molecules such as potassium, sodium, and water to pass through in a controlled manner through a channel, or in an uncontrolled manner via a pore. Many examples of non-pathogenic pores also exist, such as outer membrane beta barrel pores and multi-drug efflux pumps, which allow smaller solutes to flow freely between the periplasm and the surrounding environment. These non-pathogenic pores are carefully designed however to not disrupt the cells regular function and osmotic barrier. Pore forming toxins and peptides are also members of this second class of the membrane disrupting proteins, with their membrane-spanning domain forming a pore in the target cell membrane (Dal Peraro and van der Goot, 2016).

1.2 Pore forming toxins

Pore forming toxins and membrane disrupting peptides are employed by organisms in all kingdoms of life (Dal Peraro and van der Goot, 2016). They are thought to make up approximately 30 % of all bacterial toxins identified to date (Alouf, 2000), and are involved in the innate immune response of many animals and eukaryotes (Dunkelberger and Song, 2010). PFTs work by overcoming the membrane permeabilisation barrier, to disrupt the inner osmotic balance of the target cell, or as a means to insert an intracellular secondary toxin through the membrane, such as the anthrax binary toxin (Esser, 1994; Rojko et al., 2016; Yip and Reed, 2008). Pore forming toxins are used, therefore, as virulence factors by bacteria to gain nutrients from harsh environments such as the gut, or to allow the bacteria to colonize an environment, by mediating cellular effects by secondary toxins.

PFTs such as complement are also produced by eukaryotes, but are usually involved in immune responses, killing invading bacteria by forming pores in the bacterial membrane (Dunkelberger and Song, 2010), or forming pores in other eukaryotic cells to induce apoptosis (Qian et al., 2008).

The remarkable feature of pore forming toxins is that they lead a dual life, as both soluble proteins that can be excreted from the cell in which they are produced, and then through

conformational change and oligomerisation, as a membrane bound pore protein. These transformations can be triggered by a number of events, such as proteolysis or receptor binding (either host protein or lipid constituent) (Bann, 2012; Wilke and Bubeck Wardenburg, 2010).

Pore forming toxins are categorised in a number of ways, although often they are characterised based on the secondary structural motif used to traverse the lipid membrane (Gouaux, 1997). These categories are proteins that utilise α-helices to form the pore (α-PFTs), and proteins, which utilise β-strands to form β-barrel pores through the membrane (β-PFTs).

There are many different types of PFT, varying by family type, stoichiometry of the oligomer, and the different receptor types (Dal Peraro and van der Goot, 2016) (Table 1), The table shows that many more β-PFT structures are known in both the soluble and pore form, owing to the stability of β-barrel pores once formed (Tamm et al., 2004). α-PFTs are usually much less stable, and so indirect methods are used to study them and infer the mechanism by which they form a pore (Cosentino et al., 2016).

1.2.1 General pore forming toxin mechanisms

The large number of pore forming toxins and peptides studied to date display a great variety in pore forming mechanisms, even within pore families (Alouf, 2000). However, a number of gross steps in pore formation are conserved, owing to the fact that the same processes are required to overcome the membrane barrier (Gouaux, 1997).

These conserved steps include:

- Excretion of the protein in a soluble form
- Attachment of the protein to the membrane via a receptor
- Conformational change which inserts a part of the protein into the membrane
- Oligomerisation to form a pore unit

These steps, although required, can be achieved in a variety of ways, a few of which are shown in Figure 1.4, and described below for each stage.

Pore forming toxin	PFT family	Secondary structure features	Organism of origin	Pore stoichiometry	PFT receptor
Colicin E1	Colicins	α	*E. coli*	1–n	At IM/OM
Colicin Ia	Colicins	α	*E. coli*	1–n	At IM/OM
Colicin A	Colicins	α	*E. coli*	2	At IM/OM
Colicin N	Colicins	α	*E. coli*	2–6	IM/LPS-OM
Equinatoxin II (EqtII)	Actinoporins	α	*A. equina*	4	Sphingomyelin
Sticholysin II (StnII)	Actinoporins	α	*S. helianthus*	4	Sphingomyelin
Fragaceatoxin C (FraC)	Actinoporins	α	*A. fragacea*	8–9	Sphingomyelin
Cytolysin A (ClyA also known as HlyE) *	ClyA	α	*E. coli*, *S. enterica*, *S. flexneri*	12	Cholesterol
Non-haemolytic tripartite enterotoxin (Nhe)	ClyA	α	*B. cereus*	-	Cholesterol
Haemolysin BL (Hbl)	ClyA	α	*B. cereus*	7–8	Cholesterol
α-haemolysin (Hla) *	Haemolysins	β	*S. aureus*	7-8	PC/ADAM10/disintegrin
γ-haemolysin (Hlg)	Haemolysins	β	*S. aureus*	8	PC
Leukocidins (for example, HlgACB, LukED) *	Haemolysins	β	*S. aureus*	8	CCR5, CXCR1, CXCR2, CCR2, C5aR C5L2,
Necrotic enteritis toxin B (NetB)	Haemolysins	β	*C. perfringens*	7	Cholesterol
δ-toxin *	Haemolysins	β	*C. perfringens*	7	Monosialic ganglioside 2 (GM2)
V. cholerae cytolysin (VCC) *	Haemolysins	β	*V. cholerae*	7	Glycoconjugates
V. vulnificus haemolysin (VVH)	Haemolysins	β	*V. vulnificus*	7	Glycerol, *N*-acetyl-d-galactosamine
Aerolysin *	Aerolysin	β	*Aeromonas* spp.	7	GPI-anchored proteins (for example, CD52)
α-toxin	Aerolysin	β	*Clostridium* spp.	-	GPI-anchored proteins
Hydralysin	Aerolysin	β	*Cnidaria* spp.	-	-
ε-toxin *	Aerolysin	β	*C. perfringens*	7	HAVCR1
Enterotoxin (CPE)	Aerolysin	β	*C. perfringens*	6	Claudin
Haemolytic lectin (LSL)	Aerolysin	β	*L. sulphureus*	4–6	Carbohydrates
Kysenin	Aerolysin	β	*E. fetida*	3–6	Sphingomyelin
Perfringolysin O (PFO) *	CDCs	β	*C. perfringens*	30–50	Cholesterol
Suilysin (SLY)	CDCs	β	*S. suis*	30–50	Cholesterol
Intermedilysin (ILY) *	CDCs	β	*S. intermedius*	30–50	Cholesterol, CD59
Listeriolysin O (LLO)	CDCs	β	*L. monocytogenes*	30–50	Cholesterol
Lectinolysin (LLY)	CDCs	β	*S. mitis*	30–50	Cholesterol, CD59
Anthrolysin O (ALO)	CDCs	β	*B. anthracis*	30–50	Cholesterol
Streptolysin O (SLO)	CDCs	β	*S. pyogenes*	30–50	Cholesterol
Plu-MACPF	MACPF	β	*P. luminescens*	>30	-
Bth-MACPF (BT 3439)	MACPF	β	*B. thetaiotaomicron*	>30	-
HlyA	RTX	α?	*E. coli*	-	-
Bifunctional haemolysin–adenylyl cyclase toxin (CyaA)	RTX	α?	*B. pertussis*	-	-
MARTX	RTX	α?	*A. hydrophila*	-	-

Table 1.1. A non-exhaustive table of bacterial pore forming toxins from a number of different species. In each case, the name and abbreviation of the pore forming toxin is given in column 1, along with its PFT family in column 2. The category of each PFT is given in column 3, based on the transmembrane secondary structure feature (Gouaux, 1997). The organism for which each PFT belongs is given in column 4. The stoichiometry of the pore is given in column 5 showing that pore stoichiometry is sometimes not fixed. Finally, the putative receptor for binding of the PFT protomer to the membrane is in column 6. Finally, the PFTs are grouped by their secondary structure category, and coloured red or blue for α or β respectively, and PFTs for which a structure is available are marked with an *.

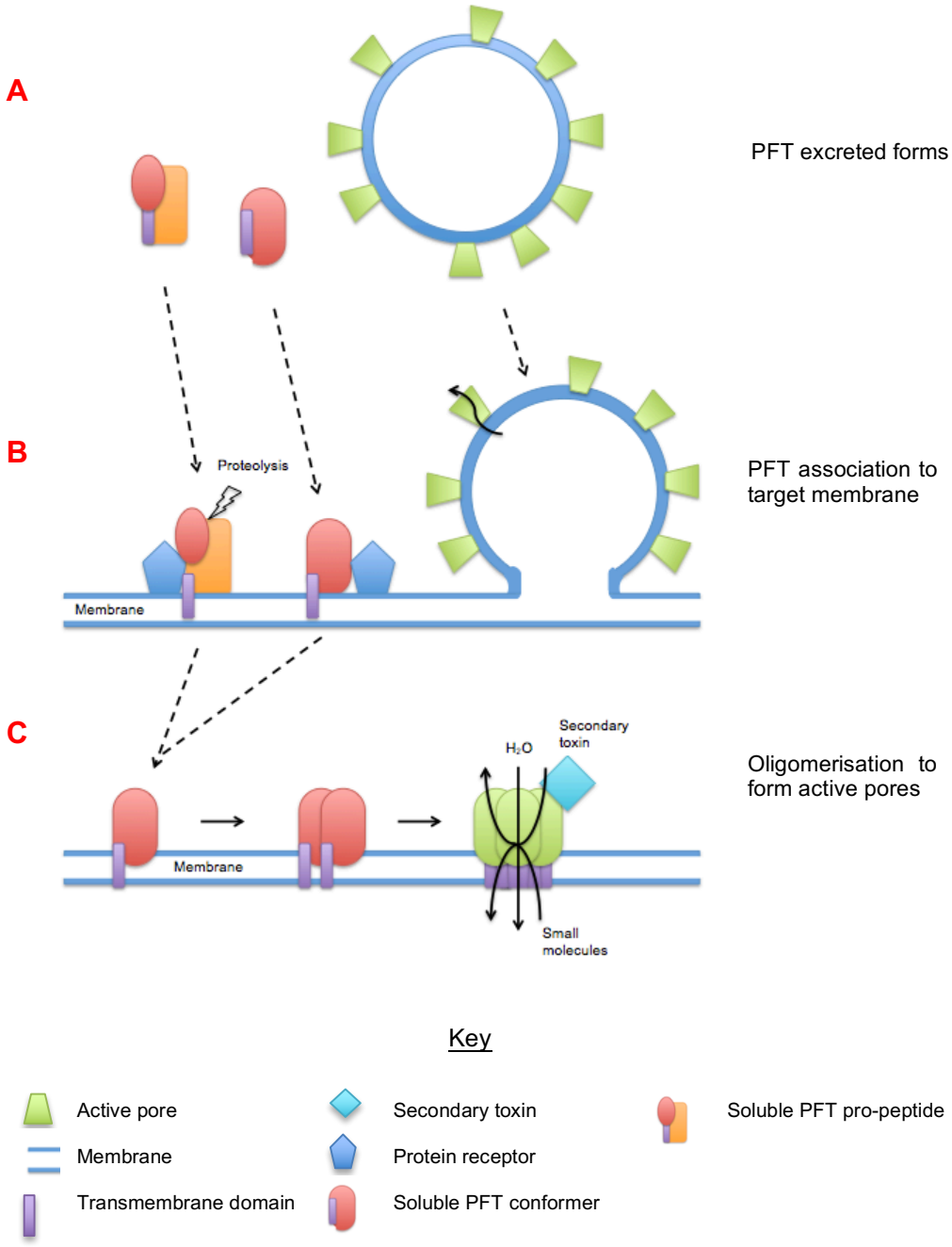


Figure 1.4. Conserved stages of pore formation by pore forming toxins. A schematic diagram showing the different stages in pore formation by pore forming toxins, and the different methods employed by PFTs for each stage. (**A**) The excretion of the PFT as a soluble protein, a requirement for the protein to reach the target membrane. PFTs can either be excreted as propeptides (left), soluble conformers (middle) or as pores in OMVs (right). (**B**) The attachment of PFTs with host membrane, leading to proteolysis of propeptides (left), conformational change (middle) or OMV membrane fusion (right). (**C**) The final stage of pore formation, through oligomerisation and opening of the pore, which goes on to cause damage either by osmotic lysis or secondary toxin insertion.

1.2.2 Excretion of the PFT in a soluble form

Three methods have been observed for bacteria exporting PFTs from host cells, two include exporting the protein as a soluble form of the protein, the last method involves inserting 'pre-pores' into outer membrane vesicles in the host bacteria, which can then associate with target membranes in a currently unknown way.

The first way that proteins can be excreted in a soluble form are as a soluble conformer, for example colicin pore forming domains, which contains a hydrophobic transmembrane helix, which is hidden in solution by the other helices of the protein (Greig et al., 2009)(Figure 1.5a).

The second method for soluble PFT excretion is as a pro-peptide which contains a protective antigen domain that must be cleaved by proteolytic activity to release the active pore-forming unit. One such example is the *Aeromonas hydrophila* proaerolysin to aerolysin conversion by removal of a C-terminal peptide (Figure 1.5b), or anthrax toxin, where the active toxin is cleaved from the pro-toxin, once the pro-toxin is associated with its host receptor.

Finally, pores can be excreted as already-formed assemblies in outer membrane vesicles, such as the excretion of Cytolysin A (ClyA) from *E. coli.* Electron tomography and microscopy have been used to observe these outer membrane vesicles containing ClyA (Wai et al., 2003)(Figure 1.5c), although it is unknown if this is a biologically relevant means of pore excretion.

1.2.3 PFT attachment to the membrane via a receptor

Many pore forming toxins have been associated with binding to host cell receptors, which allow the toxin protein to bind at the membrane interface, even when the hydrophobic region of the toxin is masked from the solvent in the soluble form of the protein (Abrami et al., 2000). Receptors are also thought to allow PFTs to recognise target vs. self by being specific for a moiety not found on the excreted cell, such as cholesterol (Tweten, 2005), or protein receptors (Giddings et al., 2004). By binding a receptor, the protein is brought to the surface of the correct cell types, where it can then undergo a conformational change in order to expose and insert the hydrophobic region into the target membrane. In the case of PFTs where proteolysis occurs, binding of the protein at the receptor also reveals and favourably positions the cleavage site for protease activity (Degiacomi et al., 2013).

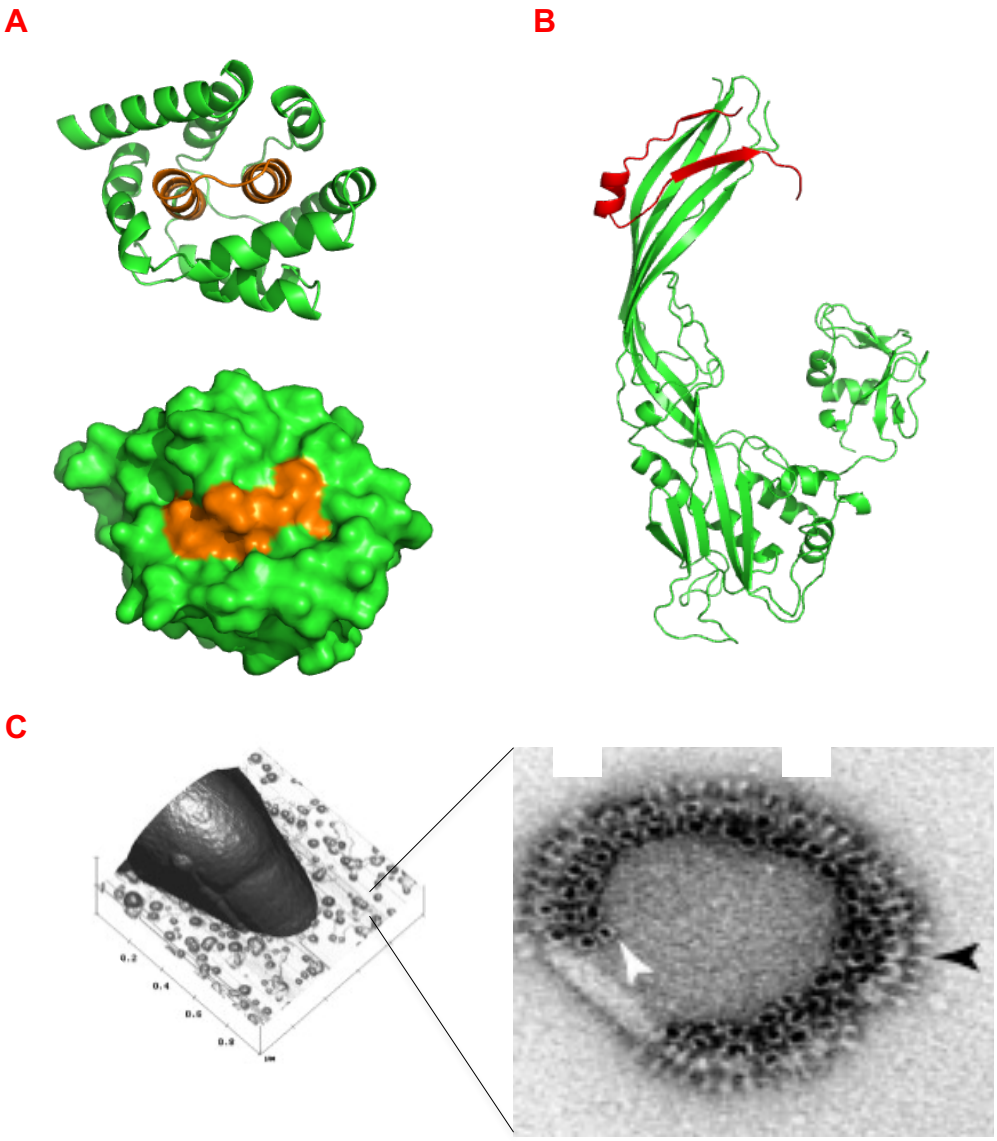


Figure 1.5.The means by which pore-forming toxins are excreted from host cells. Pore forming toxins must be able to reach the target membrane via a soluble intermediate, examples of which are highlighted here. (**A**) The ribbon and surface representation of colicin pore forming domain from colicin Ia (PDB accession 1CII) on the top and bottom respectively. The membrane inserting domain shown in orange is protected from the solvent by the hydrophilic parts of the protein (green), forming a distinct 'inside-out' fold that solubilises the protein during excretion. (**B**) The ribbon representation of the pro-aerolysin structure (PDB accession 1PRE) is shown in green, which contains a C-terminal inactivating peptide shown in red that helps to solubilise the protein and prevent early oligomer formation. (**C**) shows an electron tomography image (left), and a negative stain image of an OMV containing the pore form of ClyA (right). In the tomography image, the *E. coli* cell in black is surrounded by small OMVs that have blebbed from the main cell membrane. In the negative stain image, top views of ClyA pores are indicated by white arrows, and side views of ClyA pores are indicated by the black arrow.

Examples of known receptors are highlighted in Table 1.1, and include small molecules such as cholesterol and Sphingomyelin, as well as protein receptors such as CD59 for Intermedilysin, and the von Willebrand factor A domain for anthrax toxin (Figure 1.6), and, in some cases, a mixture of protein and lipid receptor. For example, studies of α-haemolysin show that the protein is required to first bind to the ADAM10 protein receptor prior to pore formation (Wilke and Bubeck Wardenburg, 2010). However, the crystal structure of α-haemolysin in its pore form shows that glycerophosphocholine and MPD bind at a specific site on the protein to induces pore formation, showing that the ADAM10 receptor binding is not an essential requirement for pore formation *in vitro* (Tanaka et al., 2011).

Some PFTs such as α-haemolysin also subvert the receptor, in this specific case ADAM10, which is constitutively activated by α-haemolysin. This creates a downstream effect that further increases the damage to the cell by uncoupling normal regulation of the receptor function (Inoshima et al., 2011). This added function of receptor binding allows α-haemolysin to cause damage to epithelial cells even at sub-lytic concentrations (Vandenesch et al., 2012).

1.2.4 Conformational change allowing the protein to insert fully through the membrane

Often, after associating with the membrane, further large conformational changes of the protein are associated with pore formation, even in toxins in which proteolysis rather than an initial soluble conformation exists for PFT solubilisation. One such example in which large conformational changes have been observed is in the pore formation of the α-PFT ClyA, in which both domains of the protein are involved in a conformational change, one of which that refolds from a β-strand to an α-helix (Figure 1.7)(Mueller et al., 2009). A similar level of extensive conformational changes are observed in β-PFTs, such as perfringolysin O (PFO) from *Clostridium perfringens*, where two of the four protein domains completely refold in the transition from soluble to membrane bound form, separate, and then double over to form an arch which deforms the membrane before forming a pore complex (Figure 1.8)(Rossjohn et al., 2007).

Structural analysis of the PFTs has been instrumental in deciphering these large conformational changes. The above examples are based on structures of the PFT in both the soluble and pore form, often with trapped intermediates, which show stages along the conformational change. Similar conformational changes are seen in other PFT proteins, and are expected to occur for proteins in which the structure of the pore form is still unknown. In some

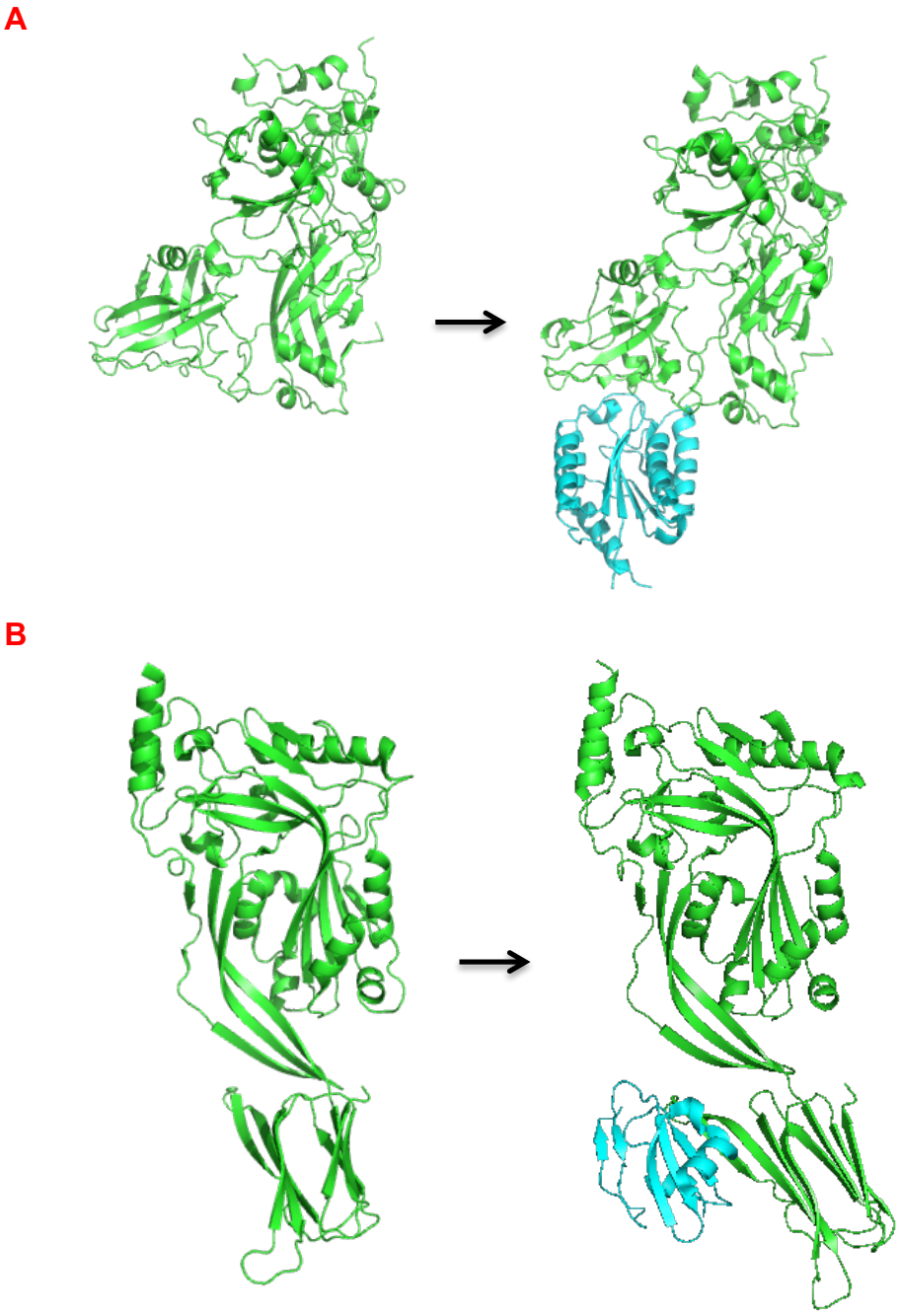


Figure 1.6. **Structures of PFTs bound to target cell protein receptors.** (**A**) Cartoon representation of the soluble domain of anthrax pore forming protective antigen (PDB accession 1TZN). Shown right is the soluble domain (green) bound to the von Willebrand factor receptor (blue). The protective antigen is able to form a heptameric assembly whilst still associated with its membrane receptor. (**B**) Cartoon representation of intermedilysin (green)(PDB accession 1S3R). Binding of the intermedilysin monomer to CD59 (blue) induces a conformational change that allows the monomer to oligomerise (PDB accession 5IMT). Intermedilysin lacks the Ig domain present on many other CDC toxins for membrane binding.

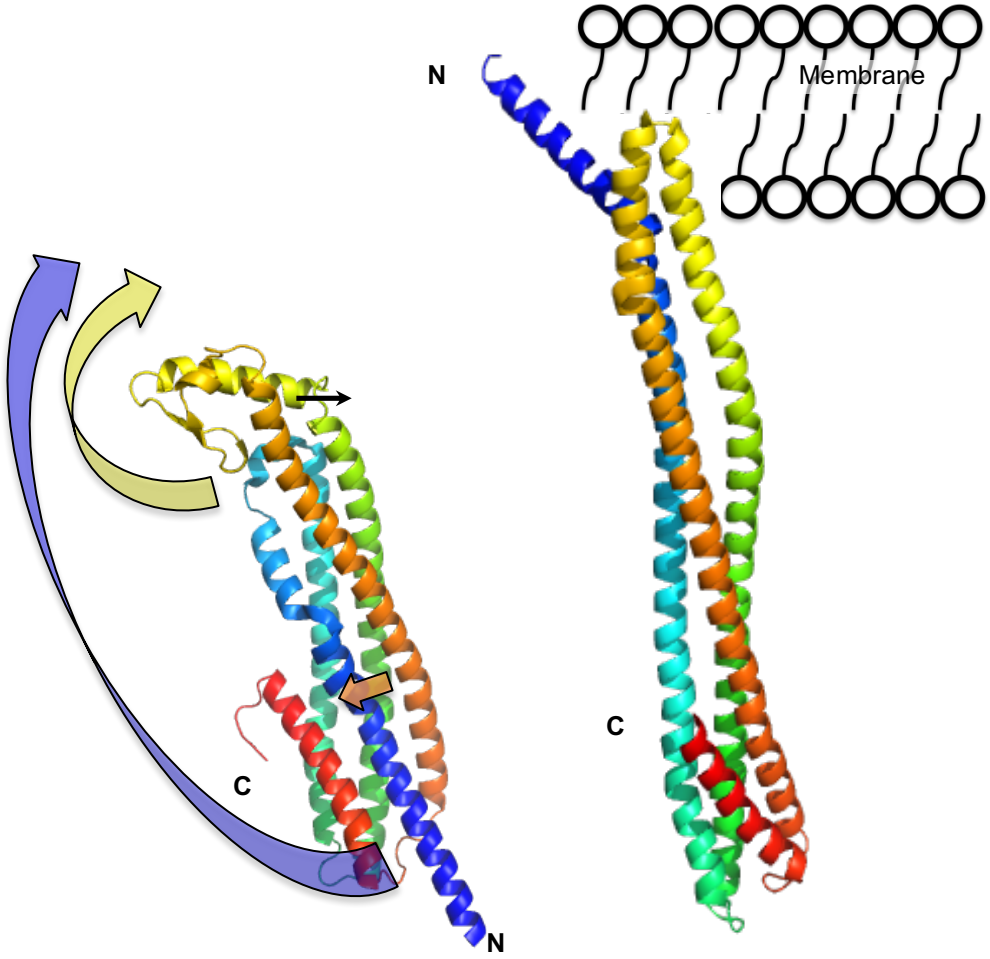


Figure 1.7. Conformational changes associated with the transition of ClyA in solution to ClyA in the pore form. Cartoon representation of ClyA PFT in both water soluble form (left)(PDB accession 1QOY), and pore form (right)(PDB accession 2WCD). Large conformational changes occur in the transmembranous N-terminus (blue), as well as transmembranous β-hairpin (yellow).

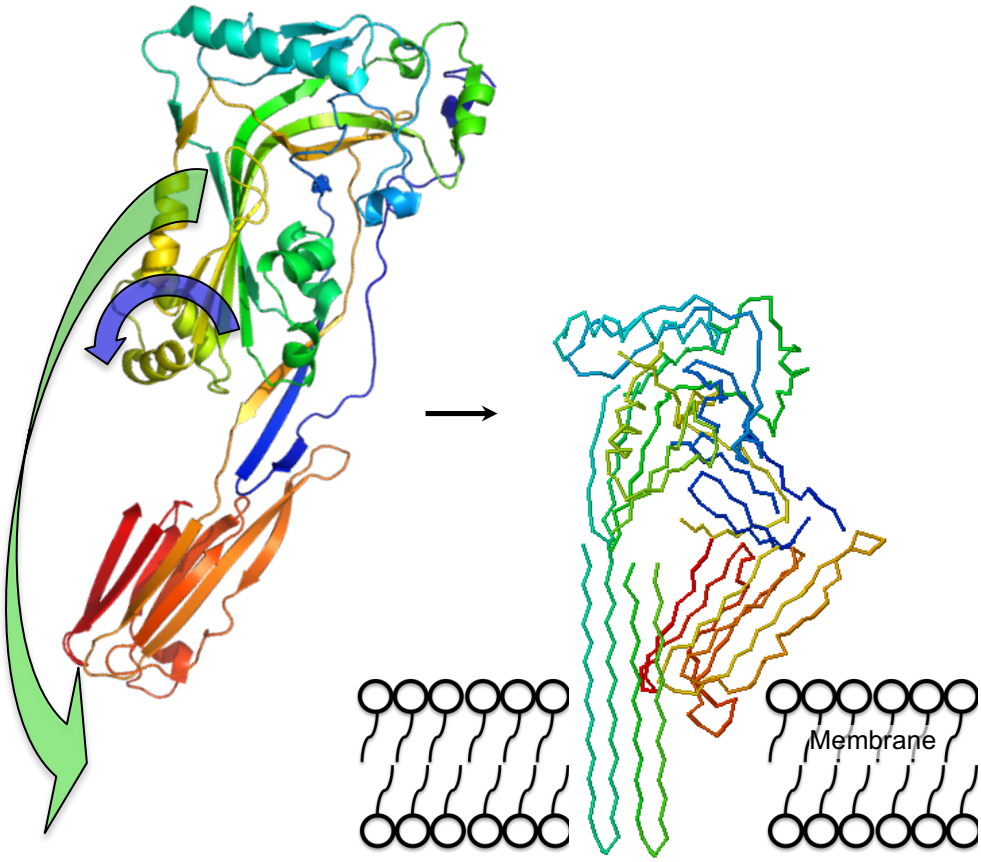


Figure 1.8. Conformational changes associated with the transition of CDC PFT Perfringolysin O in solution to the pore form. Cartoon representation of Perfringolysin O in both water-soluble form (left)(PDB accession 1M3J), and the pore form (right)(The pore form structure is of a related CDC PFT Pneumolysin (PDB accession 2BK1). Large conformational changes occur in the hinge between the N-terminus (blue), and the C-terminus (red). There is also a complete refolding of the transmembrane domain from bent β-strands to straightened strands inserted through the membrane (green).

proteins, such as ClyA, an incorrect model for pore formation had been proposed based on an initial structure of the soluble form of the protein (Wallace et al., 2000). Understandably, the large conformational changes that have been subsequently observed in this system were not predicted (Mueller et al., 2009), showing how essential a structure of the pore form is.

1.2.5 Oligomerisation to form a pore unit

Once a monomer of the PFT is inserted either fully or partially through the membrane, membrane inserted monomers then oligomerise to form either circular pores (Figure 1.9), or non-circularised assemblies that disrupt the membrane barrier. Often, stable circular pores of different stoichiometries are able to form. This phenomenon of multiple stable stoichiometries is observed for ClyA, which forms between 10 – 13 protomer pores (Eifler et al., 2006; Tzokov et al., 2006; Vaidyanathan et al., 2014), and for anthrax toxin and leukocydin, which are both able to form 7 – 8 protomer β-barrels (Kintzer et al., 2010; Miles et al., 2002; Sugawara-Tomita et al., 2002). The oligomerisation of PFT monomers at the lipid surface is rate enhanced by the collapse of the three dimensional diffusion of the proteins in solution, to the two-dimensional diffusion of proteins through the plane of the membrane. Because of this, most PFTs studied undergo the oligomerisation stage of pore formation at the membrane once the monomer has become attached (Leung et al., 2014; Thompson et al., 2011), although some toxins, such as Cry1Ab, bind at the membrane surface once oligomers have formed (Bravo et al., 2004).

Many studies have focused on the oligomerisation of PFT monomers in pore formation, as this is often a very rapid stage, which occurs spontaneously and accurately (Nguyen et al., 2003; Roderer and Benke, 2014; Thompson et al., 2011). Studies of α-haemolysin showed that protein monomers collide in the membrane to form transient dimers, which grow by addition of further protein monomers in a step-wise fashion (Thompson et al., 2011). The authors noted that a lack of pore intermediates were observed at all time-points, with only fully formed heptametrical pores or protein monomers observed, suggesting that the dimer is short-lived. This observation was rationalised by a mathematical model, in which short-lived dimers served as a means to prevent kinetically trapped pore intermediates forming that are refractory to full, open pore formation. In contrast, the model suggested that in the case where pores of larger stoichiometries form, such as *Streptococcus pneumoniae* pneumolysin, pore intermediates are more long-lived, which allows 30-40mers of PFT protomers to assemble in biologically relevant

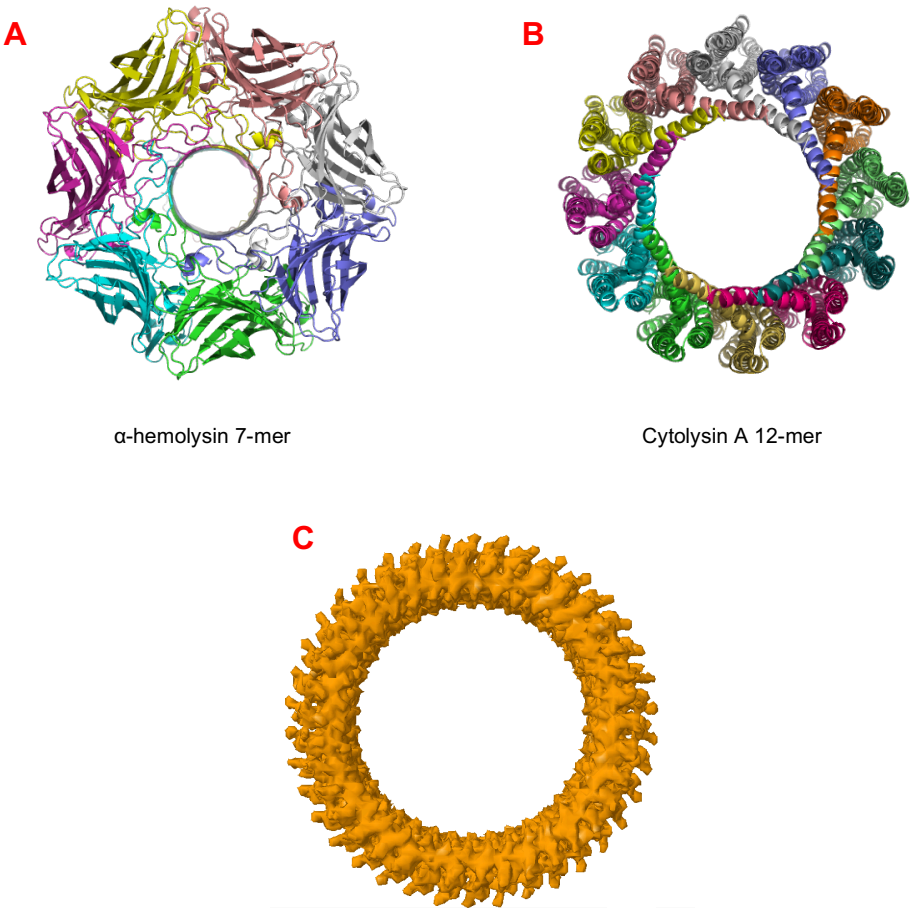


Figure 1.9. Oligomeric states of well studied pore forming toxins. (**A**) Ribbon representation of an α-hemolysin 7-mer pore where each protomer is coloured differently (PDB accession 7AHL). (**B**) Ribbon representation of a cytolysin A 12-mer pore where each protomer is coloured differently (PDB accession 2WCD). (**C**) Surface representation of Pneumolysin 37-mer pore, contoured in Jmol to show the surface accessible surface (PDB accession 2BK1).

timescales. The long-lived pore intermediates however also lead to the formation of incomplete pores and 'arcs', which have been observed experimentally (Leung et al., 2014). AFM studies show that these pore arcs are still somewhat active as the arcs form holes through the membrane, meaning that cell contents can leak and cell lysis can occur. The trade-off between accuracy and speed of pore formation for these large pore assemblies is therefore favourable in these cases where active pore intermediates exist.

1.3 Categorising pore forming toxins

1.3.1 Barrel stave vs. toroidal vs. carpet model pore formation

A number of different geometries for membrane inserting protein domains and membrane insertion mechanisms can lead to membrane disruption by PFTs (Figure 1.10). These mechanisms either work with the varying lateral polarity of the membrane, such as barrel stave pores, or by working against the varying lateral polarity of the membrane, such as toroidal pores.

The majority of large bacterial and eukaryotic pore forming proteins associate with the membrane to form ordered structures traversing the entirety of the membrane, as stable barrel stave pores. Once the protein has inserted into the membrane and formed a pore, hydrophilic residues are solvent exposed at either end of the bilayer, and in the interior pore lining. Hydrophobic residues contact the hydrophobic acyl chains of the membrane fatty acids, partitioning the hydrophobic lipid groups away from solvent, as in the undisturbed membrane. The resulting pore is completely enclosed, with extensive contacts between protomers, and no gaps in the circumference of the pore. Therefore, the stability of barrel stave pores can be remarkably high. An example of such stability is observed for the aerolysin pore, which can be incubated for 24 h in 8 M urea without disassembly or any change in tryptophan fluorescence (Lesieur et al., 1999).

The toroidal mechanism of pore formation induces a curvature in the membrane by incomplete incorporation of protein, so that the resulting pore lining is made up of a mix of protein and polar lipid head groups. Toroidal pore types can be further broken down into two groups. The first group consists of pores where the lipid is interchelated between the protein units, giving a mixed pore. The second group of toroidal pores is created by insertion of protein

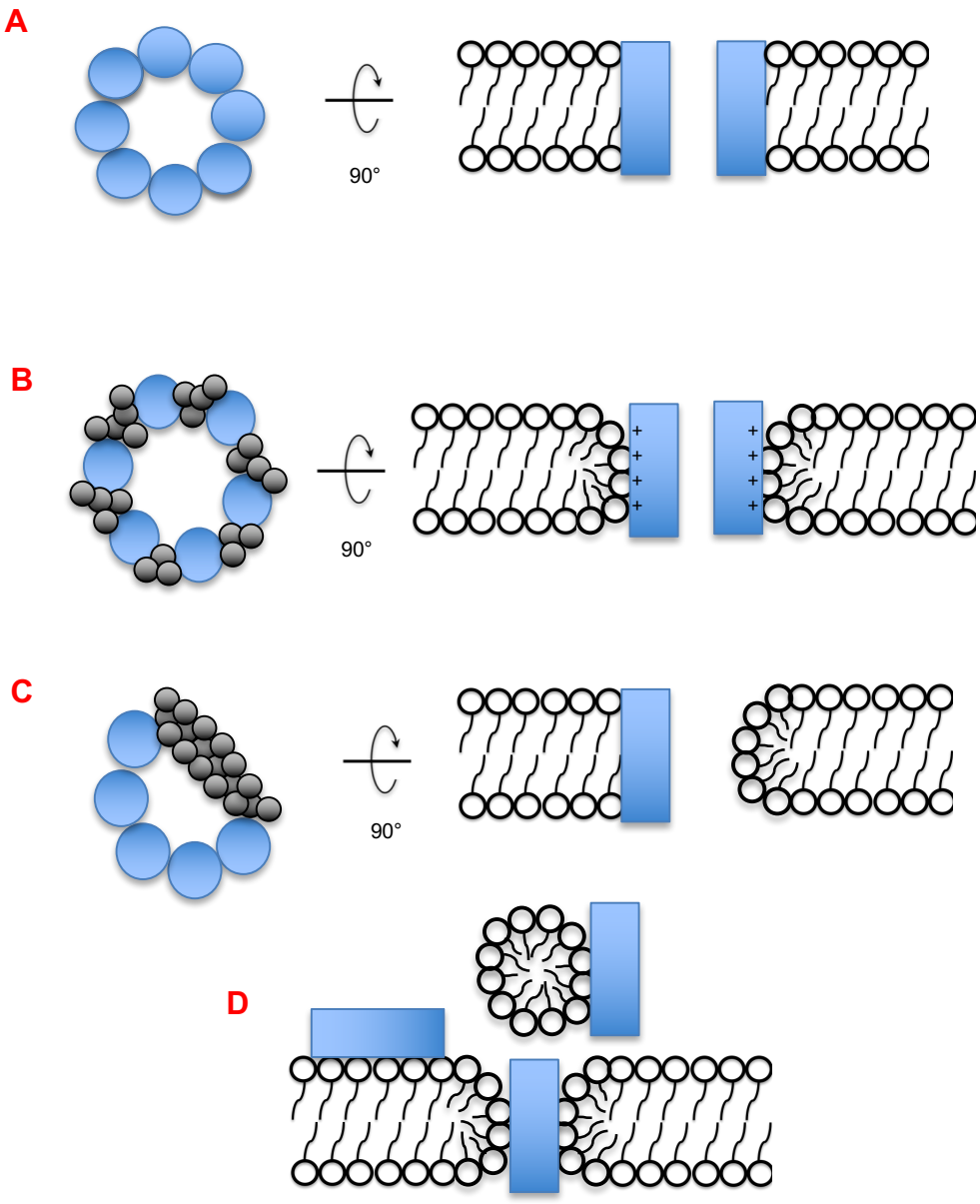


Figure 1.10. Types of pores formed by PFTs. (A) Schematic diagram of a barrel stave pore, with protein in blue and lipid in black. Top-down view (left) shows the pore contains no gaps. Side-views (right) show the protein inserts fully through the membrane partitioning the membrane. (B) Schematic diagram of a toiroidal pore. Top-down view shows that gaps in the protein are interchelated with lipid, with induced curvature. Top-down views (left) shows the induction of curvature due to the charge on the outer face of the protein. (C) Schematic diagram of an arc toroidal pore. Top-down view shows that one side of the pore is made up of protein, and the other entirely of lipid. Side-view (right) shows that the protein partitions the membrane where it inserts, and induces curvature across from itself. (D) Carpet model pore formation. Proteins bind in many orientations to the membrane and act as a surfactant depleting the lipid bilayer.

on one side of the pore, where the other side is created entirely by lipid, creating a partitioned pore.

Some interesting work has been carried out on toroidal pores, such as those created by the small alpha helical PFTs, which often form much less well ordered pores than β-PFT. Instead, these proteins induce curvature in the membrane and form holes through the membrane without forming a barrel. These PFTs form instead the toroidal pore type, in which the polar head group of the phospholipid is also included in the pore formation. There is a strain in the resulting pores, meaning they are often short-lived compared to barrel-stave pores, although the pore is more stabilised than those formed spontaneously in membranes under an applied voltage, owing to the stability of the membrane curvature by the PFT. The inclusion of lipid in toroidal pores also means the pores are difficult to study structurally by X-ray crystallography, and so historically structures of toroidal pores have been lacking. However, recently a structure of the sea anemone *Actinia fragacea* Fragaceatoxin C (FraC) has been solved by growing crystals in lipidic mesophase, and the resulting structure shows lipid bound in holes created in the interface between protomers in the oligomeric assembly (Tanaka et al., 2011). Toroidal pores have also been observed directly by glancing-angle diffraction (Qian et al., 2008) and electron microscopy, such as the higher eukaryotic bax protein from human, observed in proteolipidic discs (Xu et al., 2013). In the case of the bax protein, the pore forms an incomplete arc, where one half is protein, and the other is completely lipidic. The bax protein pore formation mechanism problem is also compounded by its possible carpet model mechanism, and so the actual method of pore formation by bax is still highly contested (Westphal et al., 2014).

Small molecular pore forming peptides, such as antimicrobial peptides, often form pores via a carpet model of pore formation (Naito et al., 2000). In this model, many peptides bind and insert partially into the membrane, at the interface between hydrophilic and hydrophobic regions of the bilayer, and disrupt the hydrophobic microenvironment of the membrane. The peptides then associate through protein-protein interactions and induce a curvature in the membrane, forming a curved carpet of proteins on the membrane surface, similar to toroidal pores. However, these peptides work in a similar way to detergents to form small to medium sized micelles of lipid, which depletes the membrane and forms transient pores through the membrane (Naito et al., 2000).

1.3.2 α-Pore forming toxins vs. β-Pore forming toxins

The most prevalent means of categorising PFTs in the literature is by the secondary structure motifs on the protein that spans the membrane in the final pore assembly, namely α-Pore forming toxins for those which utilise α-helices, and β-Pore forming toxins for those which utilise β-strands.

α-PFTs use alpha helices to generate pores through membranes, and include the large proteins such as Colicins and ClyA, as well as many of the small peptides such as melittin toxin produced in bee venom (Matsuzaki et al., 1997). The structure of an alpha helix with each mainchain amine and carbonyl group forming intra-helical hydrogen bonds means that alpha helices lack the 'inter-' hydrogen bonding potential present in β-strands that allow 'zipping' of β-strands together. Therefore interactions between α-helices are mediated exclusively between sidechains, generally via hydrophobic interactions (Lupas and Gruber, 2005), with the alpha helical pore forming toxins forming a more generally diverse and more transiently formed class of PFTs than the β-PFTs.

A number of alpha helical PFTs do however form stable barrel stave pores, such as Cytolysin A (ClyA), a 34 kDa PFT protein which forms large α-helical barrels 100 Å in length, with the majority of the protein involved in protein-protein interactions between ClyA monomers, and sitting outside the invaded membrane (Mueller et al., 2009). The transmembrane helices, which traverse the membrane fully, form an angled iris-like locking structure, similar to the α-helices in the predicted pore assembly of the colicins, suggesting a possible conserved α-helix transmembrane geometry for alpha helical PFTs.

β-PFTs are multi-domain proteins, where the transmembrane domain generally forms a beta barrel in the membrane, similar to that of porins that are found in the outer membrane of Gram-negative bacteria (Tamm et al., 2004). They resemble most closely the non-selective porins, which allow small solutes up to 1 – 4 kDa to flow freely based on a concentration gradient (Bhakdi and Tranum-Jensen, 1991). These β-PFTs differ however from selective porins, which usually contain a protein plug to select for different solutes or translocating conditions (Tamm et al., 2004). These regions are not required by β-PFTs, as β-PFTs are constitutively open. General porins are also made of a single protein chain and do not require oligomerisation to form the β-barrel structure. Some porins and channels also contain an opening in the barrel

laterally, which allow them to expand. Those porins and channels which do not open to accommodate larger substrates, such as TolC, are generally ~10 Å in internal diameter, allowing only small solutes to flow freely, whereas beta pore forming toxins form larger non-selective holes, causing damage to the membrane potential and disrupt cell permeability balance (Gilbert, 2002).

1.4 Case studies of pore forming toxins

In order to more fully explain how different PFT proteins attain their functional state, an example of a translocating pore and large conductance pore is given for both the β-PFT (Figure 1.11) and α-PFT (Figure 1.12) types, focusing on the structural information that is available, and the mechanism of pore formation understood for each.

1.4.1 Example of a translocating β-PFT – Anthrax toxin

The pathogenic Gram-positive bacteria *Bacillus anthracis* produces a number of toxins, the most well associated with pathogenicity of the organism is anthrax toxin, a tripartite virulence factor made up of the protective antigen (PA), the calcium-calmodulin (CaM) dependent adenylyl cyclase oedema factor, and the lethal factor. The protective antigen is excreted from the bacteria as an 83 kDa pro-peptide, and is able to bind to the von Willebrand domain of either the CMG2 or TEM8 receptors (Liu and Crown, 2009). Once bound, furin-like proteases cleave the 83 kDa pro-peptide to form a shorter 63 kDa active protein (Klimpel and Molloy, 1992), which is then able to form heptameric or octameric β-barrel pores in membranes. The β-barrel pore formed has a small ~ 10 Å opening, lined with a ring of phenylalanine residues at the opening, which forms a plug which is unusual for β-barrel PFTs (Krantz et al., 2005).

The phenylalanine plug is present in the anthrax toxin due to the role of the pore in virulence. The anthrax toxin is not designed to allow cell contents to leak, and instead the pore is a means to insert two associated secondary toxins into the target cell. These other two toxins, the oedema and lethal factors, can associate to the pore via their pore binding domains, and are then inserted through the pore where they gain access to the cell to carry out their toxic cellular function (Young and Collier, 2007).

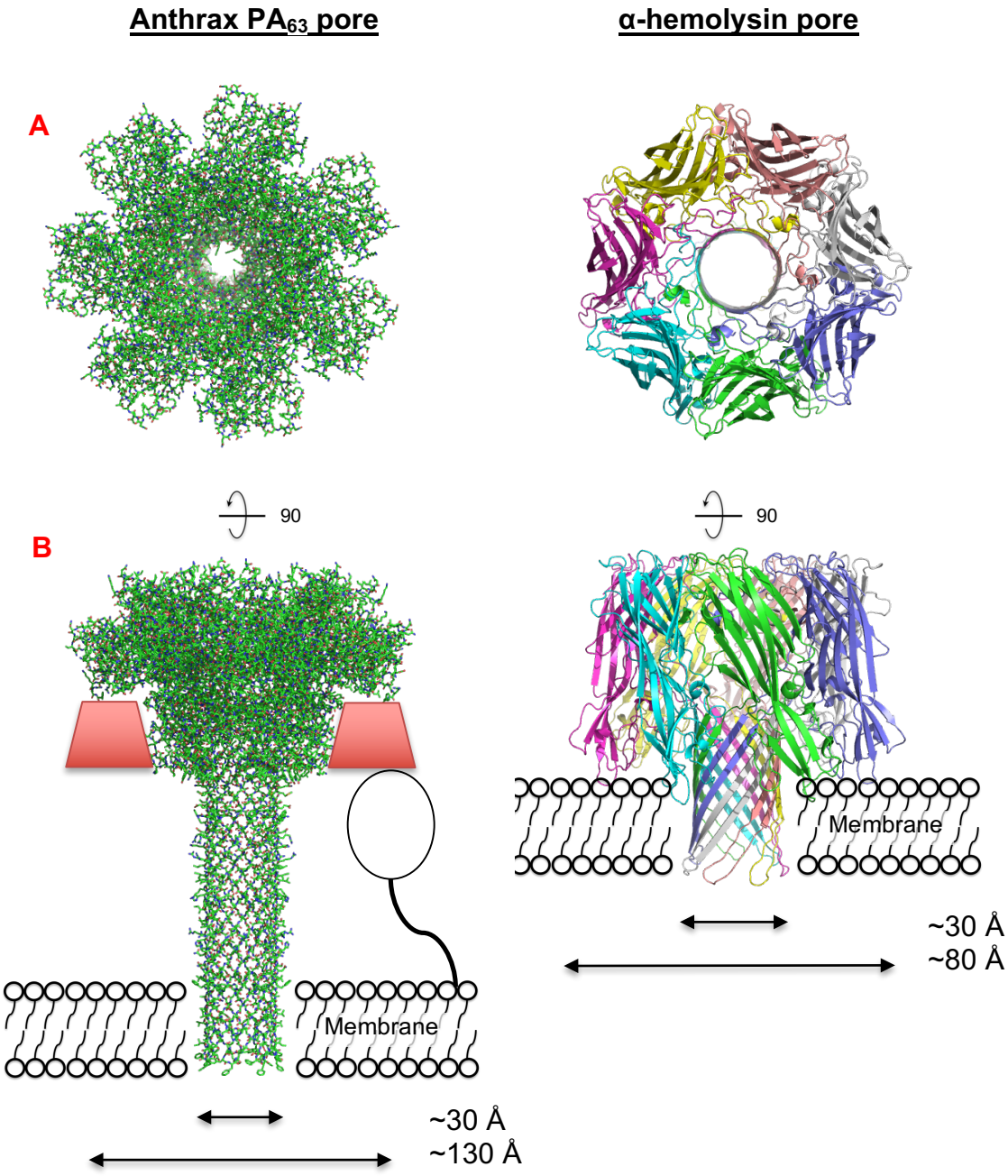


Figure 1.11. Pore structures of two β-PFTs anthrax toxin and α-hemolysin. The pore assemblies two β-PFT, the anthrax PA_{63} and α-hemolysin shown from the top down (**A**), and from the front on (B). The top view shows that both pore structures are heptameric, although other stoichiometries are also known to form. The structure of the anthrax toxin was solved by electron microscopy (Accession 3J9C) where domain 4 is disordered, but is represented in red, based on crystal structures of the cap domain (**B**). The membrane spanning regions are shown for both toxins, showing that the anthrax toxin has a large extramembranous β barrel positioning domain 4 for interaction with the receptor. α-hemolysin sits much closer to the membrane, forming more extensive interactions with the ADAM10 receptor. Both pores form with an internal diameter of ~30 Å, although the anthrax toxin has a larger overall diameter.

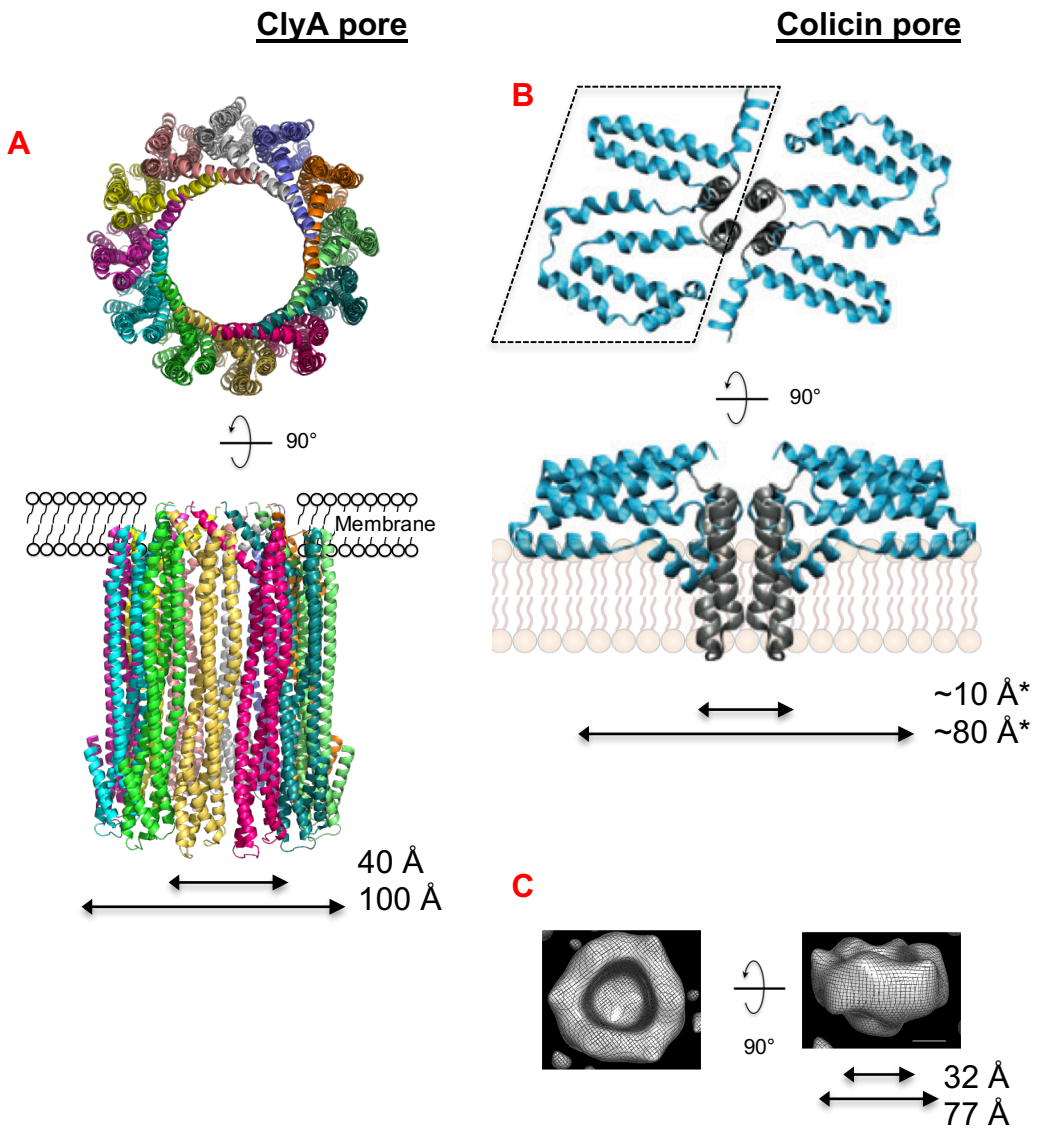


Figure 1.12. Structure representations of two αPFTs, ClyA and the inner membrane spanning C domain of Colicin Ia. (**A**) The ribbon representation of the ClyA pore, with each protomer coloured differently. The top view shows the iris-like pore opening, and the side on view shows how the iris-like opening spans the membrane. The internal and external pore diameter of ClyA is 40 Å and 100 Å respectively. (**B**) A ribbon representation model of the putative umbrella pore formed by colicins (Dunkel et al., 2015), where a single protomer is highlighted in the dashed box. The transmembrane hydrophobic helices are coloured dark grey. Based on this model the internal and external diameters of the pore are ~10 Å and ~80 Å respectively. (**C**) 3D reconstruction of the pore formed by colicin Ia in a lipid bilayer by single particle analysis, at 29 Å. The low-resolution structure has a height of ~43 Å, and an internal and external diameter of ~32 Å and ~77 Å respectively.

1.4.2 Structure of Anthrax toxin PA

The structure of the soluble anthrax PA toxin has been solved to atomic resolution by X-ray crystallography (Lacy et al., 2002). The structure of the pre-pore octameric assembly has also been solved to atomic resolution (Kintzer et al., 2010). More recently, a heptameric 'full pore' structure, including the transmembrane beta barrel domain, which was missing in the pre-pore structure, has been solved by electron microscopy to a resolution of 2.9 Å (Jiang et al., 2015)(Figure 1.11). These two assembly structures show the ability of the anthrax toxin to form pore assemblies of different stoichiometries.

The structure of anthrax toxin PA63 contains a mostly β-stranded secondary structure that forms two lobes. The stem lobe contains the hydrophobic residues that penetrate the membrane, which are protected from solvent in the soluble form by the second lobe, containing the inactivating domain. The second lobe also contains the binding domain for the secondary toxins that are used to dock the secondary toxins to the pore opening once the pore has formed (Lacy et al., 2002).

The structures of the pre-pore and pore states of the toxin show the conformational changes that occur on transition from one to the other. This approach of trapping or capturing the protein in each form during pore formation has been very successful in aiding characterisation of PFTs (Kawate and Gouaux, 2003; Roderer and Benke, 2014; Tilley and Saibil, 2006).

1.4.3 Example of a large conductance β-PFT – α Haemolysin

Staphylococcus aureus is a Gram-positive coccal bacterium, commonly associated with a number of diseases including skin infections and bacteraemia. *S. aureus* secrete a number of exotoxins which mediate these disease states by causing cell damage. These exotoxins include the β-PFTs α-haemolysin and Panton-Valentine Lukocydin, the latter of which is a bi-component toxin. α-Haemolysin was first named for its ability to lyse red blood cells, but it has since been shown that α-Haemolysin is able to cause a complex response in nucleated cells, such as deregulation of cellular processes and damage to the tissue barrier. In fact, α-Haemolysin has been shown to intoxicate a wide range of cell types including immune system T-cells, neutrophils and macrophages, as well as endothelial and epithelial cells, causing damage at sub-lytic concentrations (Hildebrandlg et al., 1991). The formation of the α-Haemolysin pore in

erythrocytes and nucleated cells allows Ca^{2+} influx and K^{+} efflux, which triggers necrotic cell death at sub-lytic concentrations by disrupting homeostasis (Vandenesch et al., 2012).

1.4.4 Structure of α Haemolysin

The structure of the α-Haemolysin pore was solved in 1996, and as such became the prototypical β-PFT (Song et al., 1996). The structure of α-Haemolysin revealed many conserved features of β-PFTs, which have been shown more recently by structure elucidation of many more β-PFTs. Although the pore form of α-Haemolysin was solved very early, the structure of soluble α-Haemolysin was determined much more recently (Sugawara et al., 2015). The α-Haemolysin monomer structure is mainly formed of beta strands with some loop regions. The protein is formed of the rim, the cap, the amino latch and the stem/pre-stem (Figure 1.13). The pre-stem contains the hydrophobic transmembrane residues, which traverse the membrane in the pore form. In the pore form of α-Haemolysin, the pre-stem extends into a curved stem, which is constructed exclusively from beta strands, and forms a β-barrel during oligomerisation (Figure 1.13). In the monomer form of α-Haemolysin the stem hides many of the hydrophobic residues of the transmembrane domain by packing against the cap region, with the amino latch loop forming specific hydrogen bonds between the two domains, such as D45 in the loop to Y118 in the pre-stem, which prevents the pre-stem unfurling in solution (Sugawara et al., 2015).

1.4.5 Example of a translocating α-PFT - Colicins

The colicin family of PFTs is a large family of toxins produced by *E. coli* to kill competing *E. coli* species (Cascales et al., 2007). Colicins are multi-domain proteins that contain a number of conserved domains, and a secondary toxin domain, which differ between colicin types (Figure 1.14). As colicins function against Gram-positive *E. coli*, they must first traverse the outer membrane and periplasm, mediated by receptor binding, to gain access to the inner membrane. Many colicins contain toxic domains that work intracellularly to affect cellular functions, such as ribosome inactivation and non-specific nuclease activity. Some colicins, however, lack the secondary domain and contain instead only the membrane traversing domains. Bioinformatical studies have indicated that many colicins contain hydrophobic regions (Cascales et al., 2007), and were predicted to form pores to deliver the catalytic domain to the cytosol of the target

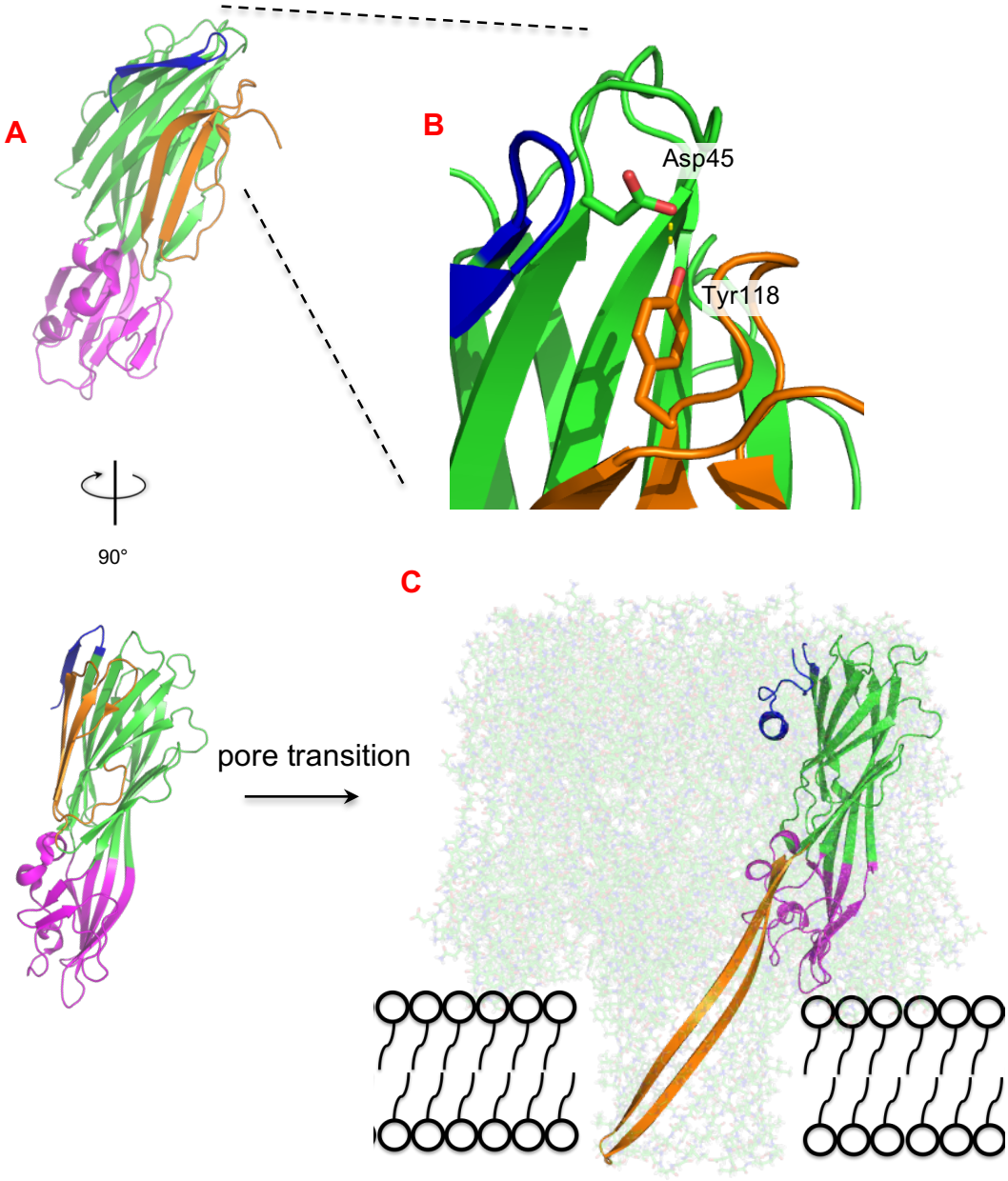


Figure 1.13. Comparison of the soluble and pore structures of α-hemolysin The crystal structure of α-hemolysin (PDB accession 4YHD) with the different domains coloured differently (**A**). The rim domain (magenta) comes into contact with the lipid and forms specific interactions with the membrane surface. In **B**, the transmembrane domain (orange) is shown in the pre-stem conformation, and packs against the cap domain (green), forming 'latching' hydrogen bonds between Tyr118 of the pre-stem and Asp45 of the cap in blue. In (**C**), the amino latch (blue) is released, and the protein undergoes a large conformational change as the pre-stem extends to form the membrane inserted pore conformer. Finally, a single protomer of the pore assembly is highlighted in a ribbon representation, with the other six protomers shown faintly as sticks, showing the orientation of the protomer in relation to the membrane.

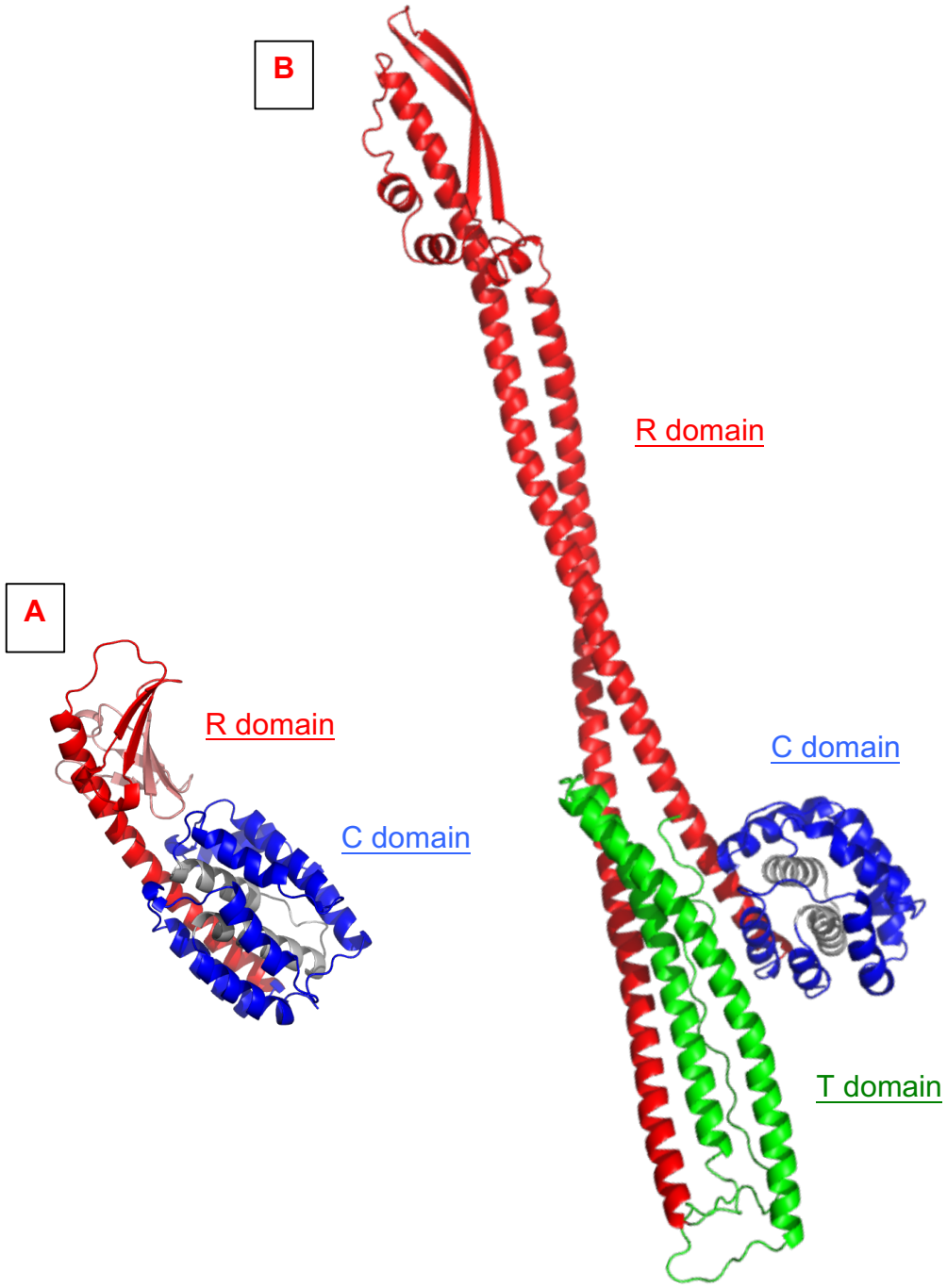


Figure 1.14. Structures of different colicins in the soluble form, from the colicin family. Two colicin family proteins Colicin N (**A**), and Colicin Ia (**B**) are shown in ribbon representations. Each is coloured based on structural and function features, as described in (Zakharov and Cramer, 2002). The R domain which is suggested to aid in receptor binding and spanning of the periplasm is shown in red for each protein. Colicin Ia also contains an N-terminal T domain (green), which is implicated also in aiding the protein to span the periplasm. The channel forming C domain is coloured blue, with the hydrophobic hairpin coloured grey. The overall architecture of many colicin proteins are similar to these, containing the same functional features, although some also contain a secondary toxin eg. nuclease

bacteria. In the case of colicins lacking this extra domain, the pore forming domain would then be sufficient to cause cell damage.

Colicins have no means to insert through the outer membrane independently, and instead bind to membrane bound porins as PFT receptors, to localise the soluble colicin to the membrane and insert the protein into the periplasm. One such example of a porin receptor has been identified for Colicin Ia which has been shown to bind TolC and BtuB before translocating the pore forming and catalytic domains (Zakharov et al., 2012).

1.4.6 Structure of colicins

The first structure solved of a colicin was of colicin a, which showed that the hydrophobic region adopted an alpha helical hairpin, surrounded by a bundle of aliphatic helices, forming an 'inside out membrane protein' fold (Knibiehler and Lazdunski, 1987)(Figure 1.12). This means of packing helices together is a common feature of α-PFTs, due to the predisposition of alpha helices to pack together mediated by hydrophobic interactions. The soluble structure shows homology to domains of the type III secretion system, a large multi-component export system found in *E. coli* and also involved in virulence to humans and other competing *E. coli* (Hueck, 1998).

The structure of colicin is a good example of structural features found in large alpha helical PFTs, although the structure of the pore form of the protein is not currently known due to the more transient nature of the pore compared with other PFTs. It is thought that the pore forming domain of colicins form an 'umbrella' dimer, which then form dimers of dimers that constitute the active pore (Figure 1.12). This tetramer would have an expected internal diameter of the pore around 10 Å, and so would have to expand to insert the secondary toxin through the membrane. 2D crystals of the membrane spanning domain of colicin Ia show a pore species exists for colicins in which a much larger internal pore diameter is observed (Greig et al., 2009).

1.4.7 Example of a large conductance α-PFT – Cytolysin A (ClyA)

Cytolysin A is a α-PFT produced by many pathogenic species of *E. coli* and *Shigella*. Although the gene for ClyA is also present in the non-pathogenic *E. coli* strain K12, detectable expression is not observed under regular growth conditions. However, when transcriptional regulators that regulate ClyA expression in pathogenic species are introduced to the *E. coli* K12

lab strain, detectable haemolytic activity as a result of ClyA expression can be observed (Oscarsson and Mizunoe, 1996). Furthermore, expression of ClyA in strains of *E. coli* lacking other known PFTs is sufficient confer a haemolytic phenotype, highlighting the importance of ClyA in *E. coli* pathogenicity. ClyA is also expressed in some pathogenic strains of *Salmonella* and *Shigella*, and has been shown to be important in survival of *Salmonella* in macrophages (Libby et al., 1994), as well as for invasion of *S. typhi* in human HEp-2 epithelial cells *in vitro* (Fuentes et al., 2008).

1.4.8 Structure of ClyA

ClyA is the most well studied and characterised example of a α-PFT, due to the wealth of biochemical and biophysical data, and the crystal structure of both the soluble and pore form of the protein being known (Mueller et al., 2009; Wallace et al., 2000). The structure of ClyA in its soluble form is almost entirely α-helical, forming two domains, a tail domain made up of a five helical bundle, and a head domain containing two short β-strands forming a 'β-tongue' (Figure 1.15). The head domain is connected to the tail domain by a large 'backbone' created by two helices, α3 and α4, that run the length of the protein. At the top of these backbone helices the β-tongue folds back downwards, and packs against the body of the protein. The β-tongue contains exclusively hydrophobic residues, which were predicted to insert into the membrane during pore formation, even prior to the ClyA pore structure (Wallace et al., 2000). The rearrangement of the β-tongue domain to α-helices during pore formation was however not predicted until the ClyA pore structure was known.

The pore structure of ClyA revealed that the pore is a dodecameric assembly, with 25 hydrogen bonds and 13 salt bridges forming between each protomer pair, locking the proteins together. Analysis of the structures of ClyA in the soluble and pore form led to a proposed mechanism for pore formation, which also agrees with biochemical data. The pore formation mechanism for ClyA is shown diagrammatically in Mueller et al., 2009, and consists of the following stages:

- The β-tongue of soluble ClyA first contacts the membrane, either as the first stage or after ClyA has attached to an unknown receptor on the target membrane. Once the ClyA is attached to the membrane, the β-tongue undergoes a conformational change from β-strands to a α-helical hairpin. The α-helical hairpin forms a 14 Å extension of α3 and α4,

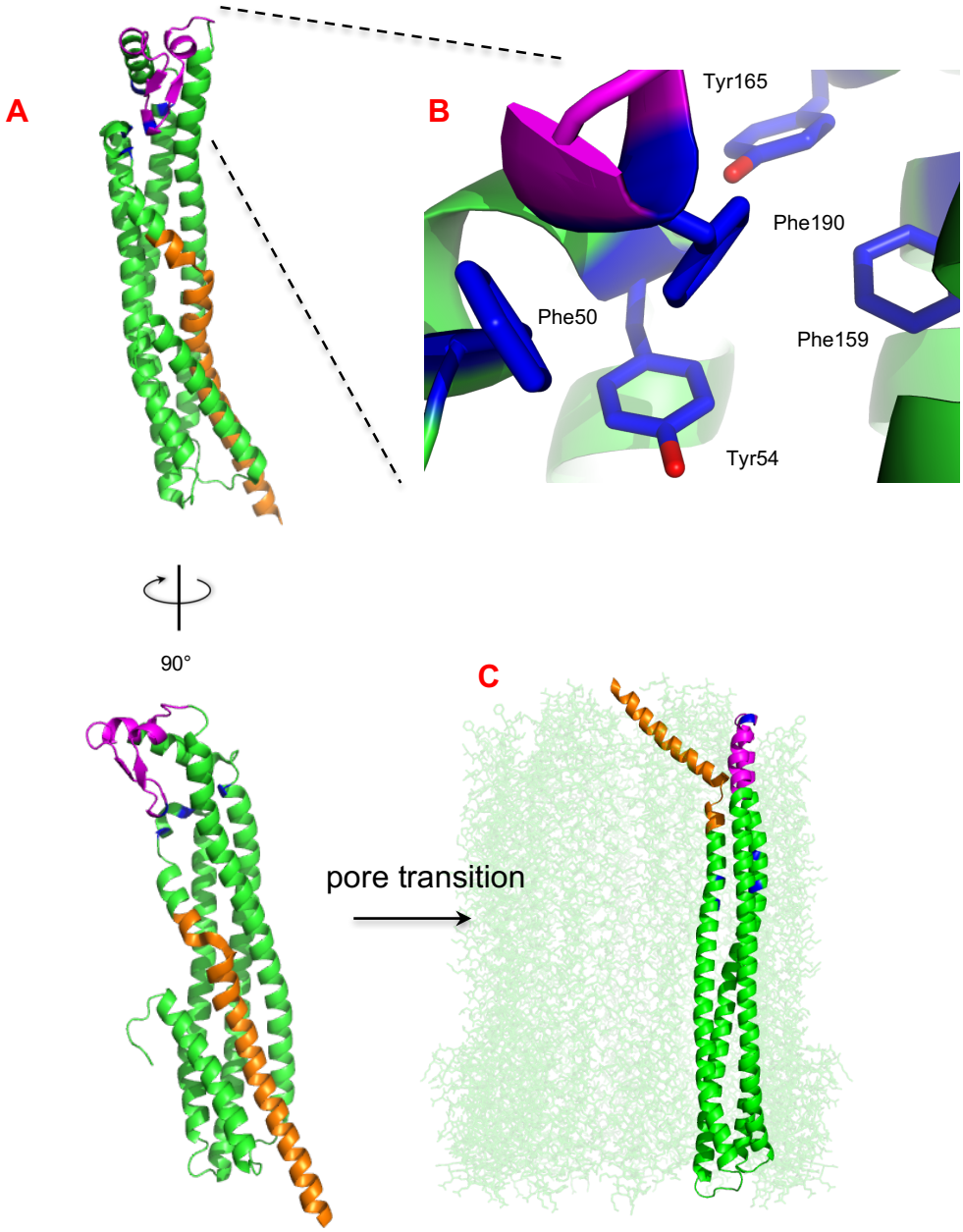


Figure 1.15. Comparison of the soluble and pore structures of cytolysin A. The crystal structure of Cytolysin A (PDB accession 1QOY) with the different domains coloured differently (**A**). The β-tongue (magenta) comes into contact with the lipid and undergoes a conformational change to α-helical hairpin. The aromatic residues shown in blue (**B**) stabilise the β-tongue in the soluble form, prior to conformational change. The transmembrane domain (orange) packs against the main body of the protein (green) in the soluble protomer form, and forms the iris structure in the oligomeric pore form (**C**).

the height of the hydrophobic extension region being 14 Å, which is long enough to insert through one leaflet of the lipid bilayer. At the interface between hydrophobic and hydrophilic residues in α3 and α4 is a tyrosine residue, which is thought to anchor membrane proteins at the membrane solution interface.

- The conformational change in the beta tongue to an α-helix allows the N-terminal helix which packs against the body of ClyA to undergo a 180° flip, into a position normally blocked by the β-tongue in the soluble form of ClyA. The N-terminus then forms an extension of α2, with a kink at the membrane solution interface. One face of the N-terminal helix faces towards the membrane, and buries its hydrophobic residues into the membrane. The other face contains hydrophilic residues and lines the interior of the pore.
- ClyA must oligomerise to form a pore, from protein monomer to 10-13mer, which forms a sealed pore via extensive surface contacts between neighbouring protomers. Unusually, this oligomerisation of the ClyA monomers is not the rate-limiting step, with the previous conformational change being the rate-limiting step (Eifler et al., 2006). Each kinked N-terminal α-helix then comes together in the pore assembly to form an iris-like opening at the top of the assembly.
- The iris-like structure of the interlocked N-terminal α-helices is then predicted to induce a lateral strain on the lipid bilayer as the N-terminal alpha helices insert through the outer leaflet of the membrane. Full insertion of the N-terminal α-helices through the membrane is the last step in pore formation, forming the open pore, which is the form seen in the crystal structure and EM images of pores in detergent and lipid (Mueller et al., 2009; Tzokov et al., 2006; Wallace et al., 2000).

The discrepancy between oligomeric states of ClyA pores is well documented, as the crystal structure of ClyA contains 12 units per pore, and a number of EM data show between 10 and 13 units forming pores of different diameters (Eifler et al., 2006; Tzokov et al., 2006). The electron micrographs show that the sample contained a majority of pores with internal diameter of 42-55 Å and external diameter of 70 – 90 Å, but also a minor species (~30 %) for which the internal diameter was larger at 55 – 60 Å and external diameter of 90 – 105 Å. Differences in oligomeric

states of PFTs has also been shown for different pore forming toxins such as α-haemolysin, and show that pore populations can be heterologous and form different oligomers stochastically.

The crystal structures of ClyA have greatly increased our understanding of αPFTs, but lacks information about assembly dynamics of the soluble protein becoming the pore. Many studies have been carried out to determine the assembly dynamics of ClyA using biophysical methods (Lee and Senior, 2016; Roderer and Benke, 2014; Roderer et al., 2016; Vaidyanathan et al., 2014). The conformational change of ClyA from soluble form to membrane-bound pore form is suggested to proceed via a molten globule pathway (Benke et al., 2015). Mutations have been made in the ClyA protein to engineer disulphide bridges, in order to trap the pore assembly along this pathway and study the different stages in pore formation (Roderer and Benke, 2014). One such artificial disulphide bridge introduced at the tail of the protein blocks the assembly of the ClyA pore and traps the protein in a monomeric intermediate. The monomer has increased hydrophobicity suggesting a conformational change has occurred to expose the membrane spanning regions. When these trapped ClyA intermediates are reduced, an increased amount of haemolysis is observed compared to wildtype ClyA. It is possible this occurs as the disulphide bridge stabilises the protein in the molten globule state, and promotes correct rearrangement of the protein during conformational changes (Roderer and Benke, 2014).

The assembly dynamics of oligomerisation of ClyA protomers have also been studied. Modelling of the pore assembly based on membrane lysis dynamics have shown insights into how ClyA pores form (Vaidyanathan et al., 2014), which may be similar for many other unrelated PFTs (Lee and Senior, 2016). The study showed that ClyA pores form via addition of compatible oligomers, beginning with addition of ClyA monomers to form dimers, and then proceeding via monomer addition as well as oligomer addition to form mature pores. In the case of ClyA, this may explain why 10 - 13 units form the pore, as many oligomeric intermediates can be added to give 12 units, leading to a large number of compatible oligomer additions available to form pores.

A number of ClyA homologues have been identified in which multiple specialised protomers are required for pore formation, compared to ClyA, which is homo-oligomeric. Examples for other PFTs for which multiple components are required are known for other PFT, such as Leukacydin, a binary homologue of α-haemolysin (Wardenburg et al., 2007), and the human

complement octapartite homologue of pneumolysin (Rosado et al., 2008). In multicomponent PFTs, the specific steps in pore formation can be split between domains on each protein, so that each protein does not in isolation contain the means for form pores. This is the case for ClyA family tripartite PFTs, in which each protein seems to have a specific role in pore formation.

1.4.9 ClyA-related crystal protein Cry6Aa

Bacillus thuringiensis produces a number of protein inclusions that naturally form crystals within the bacteria, called Crystal (Cry) toxins. Recently, the crystal structure of residues 1 – 396 of a Cry family protein from nematicidal *Bacillus thuringiensis* has been solved, and been shown to be related in structure to ClyA (Huang et al., 2016). The structure contains many of the same features of ClyA, including the same core domain and a stretch of hydrophobic residues. However, the hydrophobic region does not form the same β-tongue as seen in ClyA, and instead forms a loop at the end of two downward facing helices (Figure 1.16). The N-terminal helix is also discontinuous and made up of three smaller perpendicular helices, forming a more open conformation, which are suggested to reassemble in a similar manner to in the ClyA pore formation. This structure on one hand adds to the family of ClyA proteins containing a single protomer unit, but differences in its soluble structure raise questions as to its mode of action in pore formation.

1.5 Tripartite homologues to ClyA identified in *Bacillus cereus*

Bacillus cereus is a Gram positive, motile, spore-forming rod bacterium, and a major foodborne pathogen that is able to cause a number of pathologies in humans, including necrotic enteritis, vomiting, and diarrhoea (Turnbull, 1976). Four major toxins have been identified in pathogenic strains of *B. cereus*, haemolysin BL (HBL), non-haemolytic enterotoxin (NHE), CytK, and enterotoxin FM (entFM). HBL and NHE are two different tripartite pore forming toxins with structural homology to ClyA.

1.5.1 HBL tripartite pore forming toxin

The first tripartite pore forming toxin identified in *B. cereus*, HBL, has been extensively characterised (Beecher and Macmillan, 1991). It has been shown that HBL is able to cause

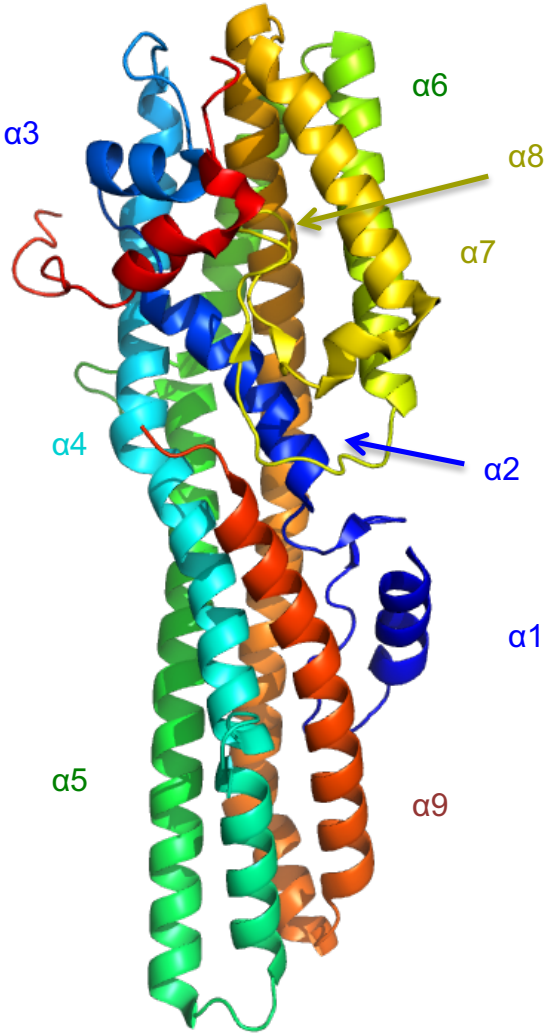


Figure 1.16. Ribbon respresentation of trypsin activated Cry6Aa toxin from *Bacillus thuringiensis.* The ribbon representation of trypsin activated Cry6Aa (PDB accession 5KUC) coloured from blue to red for N-terminus to C-terminus. The hydrophobic stretch of residues forming the putative membrane insertion domain are shown in yellow, and form a loop rather than an ordered β-tongue as seen for ClyA. The orientation of the loop is also altered, due to two downward facing helices α6 and α7.

vascular permeability and dermonecrotic pathologies when expressed in *B. cereus*, and also causes accumulation of fluid in rabbit Ileal loops, which suggests HBL is involved in causing the diarrheal symptoms in *B. cereus* infection (Beecher et al., 1995).

HBL PFT is made up of three proteins found in an operon in the *B. cereus* genome, namely Hbl-L1 (38 kDa), Hbl-L2 (43 kDa), and Hbl-B (41 kDa). The L in L1 and L2 stands for lytic, due to the lytic activity of these proteins (Beecher et al., 1995), and the B in Hbl-B stands for binding, as Hbl-B is predicted to have a membrane binding role. All three HBL protein are required for lysis, and it is suggested that the proteins form a membrane attack complex of sorts (Beecher and Wong, 1997).

An unusual lysis pattern is observed when *B. cereus* secreting HBL are grown on blood agar (Beecher and Wong, 1997). Lysis of the blood in the agar medium is lysed around the bacteria, but a halo of no lysis occurs in very close proximity to the *B. cereus* colony. This observation is rationalised by the ability of Hbl-B to inhibit the activity of Hbl-L1 when at equimolar concentrations. As Hbl-B and Hbl-L1 diffuse from the colony of excretion the concentration of Hbl-B drops when releases inhibition and causes Hbl-mediated haemolytic activity.

1.5.2 Hbl-B structure

The structure of the Hbl-B protein has been solved in the soluble form by X-ray crystallography (Madegowda et al., 2008), and consists of a similar fold to ClyA (Figure 1.17), even though the sequence identity between the two proteins is low (32 % identity over 38 % of the protein). The head domain of HBL contains an extra fold compared to the head domain of ClyA, which places the beta tongue region facing upwards, compared to the downward beta tongue seen in the soluble structure of ClyA (Figure 1.17). Preliminary gel filtration of HblB in the presence of detergent shows that HblB is able to oligomerise to form a heptamer or octamer (Madegowda et al., 2008), suggesting the head domain is able to interact with detergent despite the change in positioning of the head domain compared to ClyA.

The structural similarity of Hbl-B to ClyA, and the amino acid homology between each component in the HBL and NHE tripartite proteins gave the first insight that both NHE and HBL may be functional homologues of ClyA. These three PFT members together form a novel α-PFT super-family, with ClyA as the prototypical member.

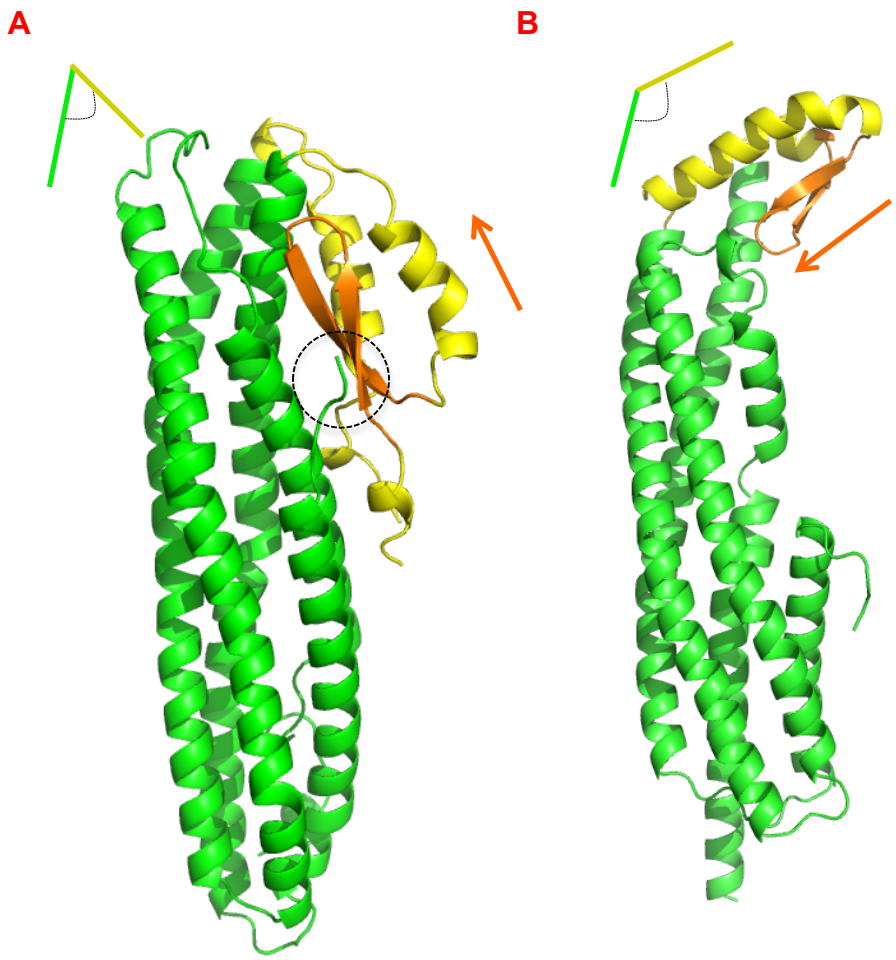


Figure 1.17. Comparison of the pitch and orientation of the head domain in the crystal structure of Hbl-B and soluble ClyA. (**A**) Cartoon representation of Hbl-B crystal structure (PDB accession 2NRJ) and the soluble ClyA structure (**B**)(PDB accession 1QOY), coloured such that the main body is in green, the α-helical part of the head is in yellow, and the β-tongue is in orange. An arrow in orange denotes the direction of the β-tongue in each structure, and an angular diagram shows the direction of the helical part of the head domain with relation to the main body of the protein. The increased length of the helical head portion in Hbl-B positions the β-tongue such that it faces upwards, and forms a latch with the C-terminus (circled).

HBL is absent in some pathogenic strains of *B. cereus*, which has made HBL less attractive for study, especially as the related tripartite system, NHE, from the same organism, has been shown to be much more widely distributed among pathogenic strains of *B. cereus* and *B. thuringiensis* (Ngamwongsatit et al., 2008). This assertion is, however, disputed for the *B. cereus* strain ATCC 10876, for which an antibody depletion of Hbl, or gene deletion of Hbl, led to decreased virulence in *B. cereus* (Sastalla et al., 2013). This strain however contains two copies of the *hbl* genes, which may account for their increased role in virulence.

1.5.3 NHE tripartite pore forming toxin

The second *B. cereus* tripartite PFT system, NHE, was identified during a large food poisoning outbreak in 1995, when a strain of *B. cereus* (NVH 0075/95) was isolated in which the previously identified enterotoxins were not present (Lund and Granum, 1996). Microbiological studies into this strain identified an operon of three proteins, NheA, NheB, and NheC as the major cause of virulence (Lund and Granum, 1996).

Studies into the prevalence of HBL and NHE expression in *B. cereus* have shown that at least one of the *nhe* or *hbl* operons are found in all pathogenic strains of *B. cereus*. The *nhe* genes are found more often than *hbl*, and are always present in gastroenteritic strains. NHE is not enough to convey a pathogenic phenotype to all B. cereus infections however, as was highlighted by LDH release from Human small-airway epithelial cells when incubated with different strains of *B. cereus* under microaerobic conditions. These studies highlight that other toxins contribute more to virulence in these conditions than does NHE (Kilcullen et al., 2016). Despite this, expression of NHE has been shown to be a great indicator of *B. cereus* cytotoxic activity and virulence in food poisoning cases (Moravek et al., 2006), and new technologies are being developed for detecting dangerous levels of *B. cereus* in food and drug products by identification of NHE transcription (Ceuppens et al., 2013; Hansen and Hendriksen, 2001).

Cellular lysis assays have shown that all three proteins are required to cause cell damage, and that the maximum lysis of Vero cells (epithelial cells) by NHE is observed when NheA, NheB, and NheC are incubated at a ratio of 10:10:1 respectively (Lindbäck et al., 2004). These same lysis assays carried out on different cell lines, monitoring PI uptake and LDH release from

cells when incubated with NHE showed that different NHE components are minimally required for cytotoxic activity of the different cell types. In GH4 pituitary cells, only NheA and NheB are required for cell lysis (Haug et al., 2010), whereas in Vero epithelial cells, similar to the cells faced by *Bacillus in situ* in the gut, all three proteins (NheA, NheB, and NheC) were required for cell lysis. It has been suggested that in the case where NheC is not required, a host protein on the surface of the pituitary cell acts as a receptor for PFT attachment, as seen for many other pore forming toxins, as discussed in 1.2.3. This degeneracy in receptor requirement is the first such example, and suggests an interesting novel cell binding function of the NheC protein in pore formation.

NheC has been shown to associate with NheB in solution, with low NheC concentrations relative to NheB in solution aiding in Vero cell LDH release (Lindbäck et al., 2010), whereas high NheC concentrations relative to NheB in solution becomes inhibitory (Lindbäck et al., 2004). Studies have also shown that NheB and NheC form a small conductance pore (Zhu et al., 2016), with NheA conveying the most lytic activity to the pore. When NheB is preincubated with detergent (Phung et al., 2012), pore formation can also be abolished, showing that a hydrophobic interaction is involved in the pathway of NheB forming a pore. These results, taken together with the recent observation that NheB and NheC form a small pore (Zhu et al., 2016), suggest that NheB undergoes a conformational change in a hydrophobic environment, which is inhibited when NheC concentration becomes too high. The NheB and NheC pore is further activated by NheA, and this final pore produces maximum lysis.

NHE pore formation also shows a dependence on the binding order of each component to the target membrane, with the optimum order being NheC followed by NheB and NheA (Lindbäck et al., 2010). When NheB is pre-incubated with cells, followed by NheC and NheA, lysis is abolished, showing that incubating first with NheB leads to intermediates that cannot form pores. *In vitro* assays to date have been carried out with a combination of recombinant protein, refolded protein, and protein purified or used natively from cell isolates (Lindbäck et al., 2004). These studies have resulted in a number of conflicting results, complicating analysis of the protein system.

1.5.4 Structure of NHE proteins

Kyte and Doolittle hydropathy plots calculated from the amino acid sequence of each protein in the NHE system, and indeed the HBL system, agree with biochemical observations that each protein has a specific role in pore formation. This is evident in that each protein has a different, and conserved pattern in the predicted beta tongue head domain (Figure 1.18). The amino acid sequence of NheA lacks any extended hydrophobic region, unlike ClyA, and so must contain an amphipathic head or no head at all. NheB contains a 53 residue hydrophobic stretch of amino acids in the putative head domain, compared to the 25 hydrophobic residues in the head domain of ClyA, making the NheB hydrophobic stretch long enough to traverse the membrane fully as an α-helical hairpin (opposed to the single leaflet the ClyA head inserts into). The amino acid sequence of NheC contains a similar span of hydrophobic residues to ClyA, both in length and amino acid identity, suggesting the head domains of ClyA and NheC may server a similar role in pore formation. It is clear that structures of each component of NHE in the soluble form would aid in determining the role of each protein and in elucidating the NHE pore forming mechanism.

1.5.5 NheA structure

The structure of NheA has been solved in the soluble form to a resolution of 2.1 Å by X-ray crystallography, and shows that NheA has a similar topology to ClyA in its soluble form (Ganash et al., 2013)(Figure 1.19). NheA is made up of 5 alpha helices, which form a tail domain made up of a 4/5 helical bundle. The N-terminus of NheA is disordered, compared to the N-terminal helix of ClyA, meaning that NheA lacks a prominent 5 helical bundle, and instead has a partial 4/5 helical bundle (Figure 1.19). The head domain of NheA differs from the head domain in ClyA on a primary and tertiary structural level. The difference in primary sequence between the two is due to the amphipathic nature of the head region compared to the strictly hydrophobic nature of the ClyA head. The ClyA head inserts through the membrane as a hydrophobic helical hairpin as discussed previously in 1.4.8. The amphipathic nature of the NheA head has led to the proposal that NheA may insert its head domain as a beta strand, similar to a β-PFT (Figure 1.20). This proposal is also based on the tertiary structure of the NheA head, in which the beta tongue is much longer than in ClyA, and in fact is long enough to insert through the membrane without rearrangement.

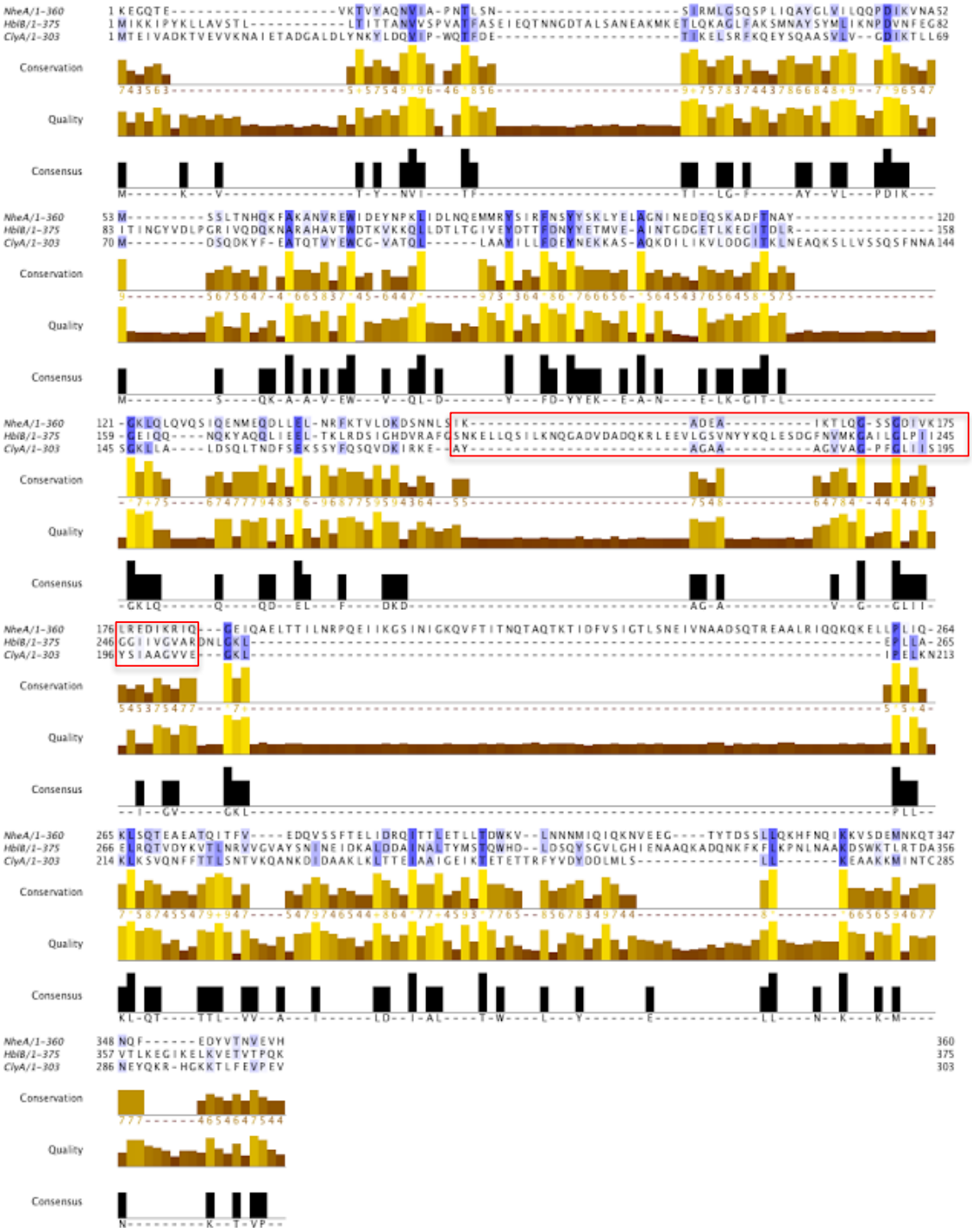


Figure 1.18. Amino acid sequence alignment between ClyA, Hbl B, and NheA. Amino acid sequence alignment between NheA, HblB, and ClyA, with residues highlighted based on the sequence identity between all proteins. Very few residues are 100 % conserved between all three proteins (between 6.1 % and 7.2 %), showing that although structurally similar, the three proteins are distant homologues. Below each sequence alignment line is also a bar graph of conservation, quality, and the consensus sequence based on all three alignments. Also highlighted in the red box is the area that constitutes the hydrophobic head region in ClyA, highlighting the poor sequence identity in this region between proteins. Image was produced using Jalview software and using the T-Coffee alignment algorithm with default settings.

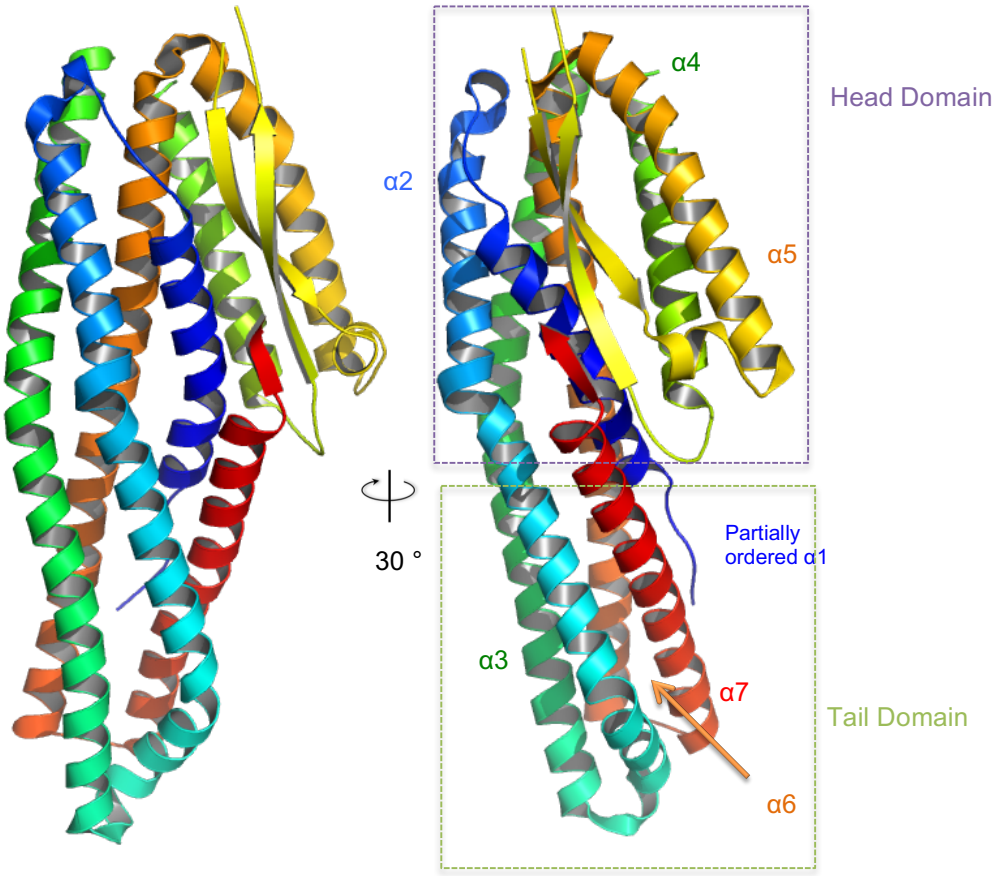


Figure 1.19. The crystal structure of NheA. A chain of the NheA present in the crystal structure (PDB accession 4K1P), coloured from blue to red for the N-terminus to the C-terminus of the protein. Each α-helix and β-strand is labelled, and the two domains, the head domain and the tail domain are highlighted.

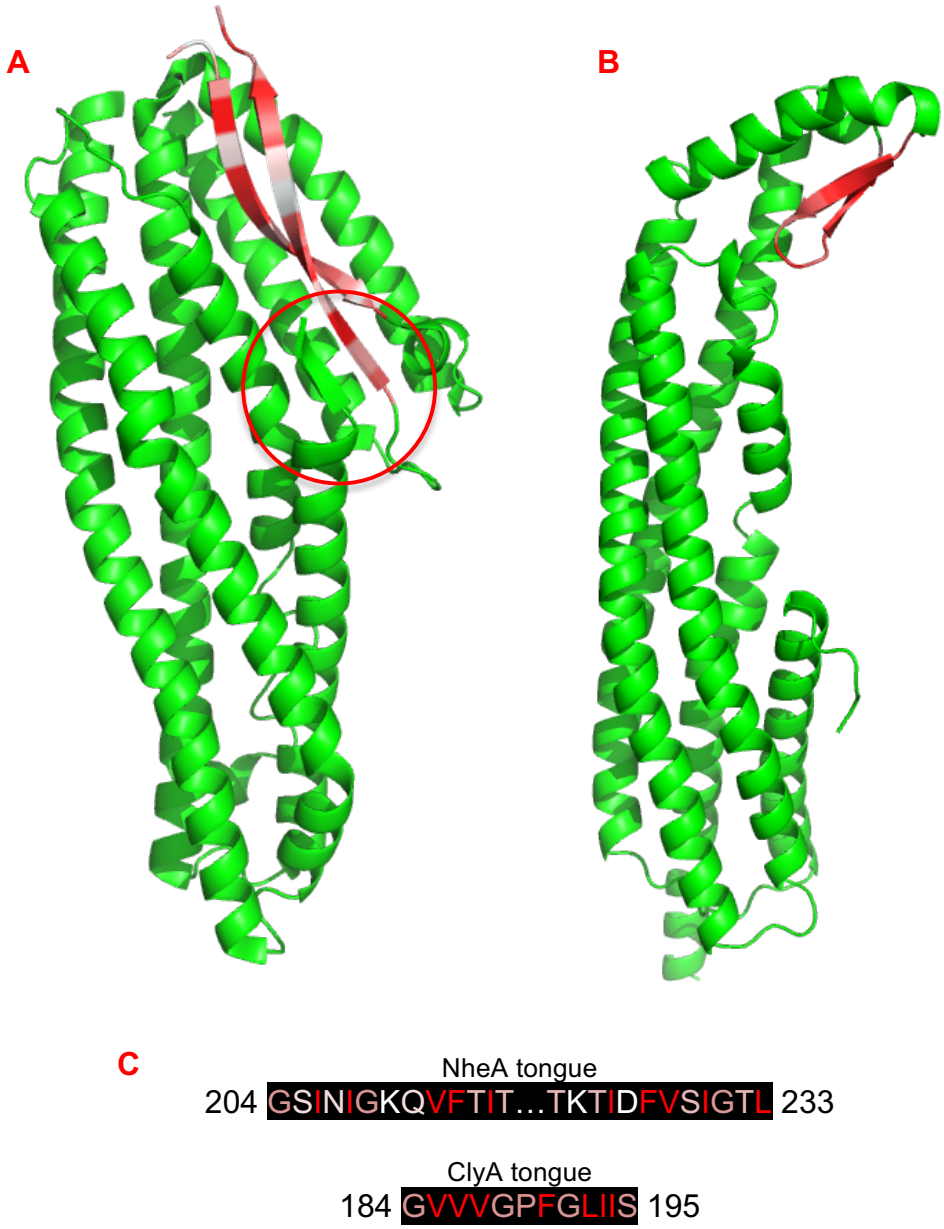


Figure 1.20. **Comparison of the β-tongue motif in NheA and ClyA.** (**A**) Ribbon representation of Hbl-B crystal structure (PDB accession 2NRJ) and (**B**) the soluble ClyA structure (PDB accession 1QOY), coloured such that the main body is in green and the hydrophobic head is coloured based on the hydrophobicity of each residue. Hydrophobic residues are coloured redder, and residues with charge are coloured white, with more hydrophilic residues forming a spectrum of colours between the two based on their hydrophobicity. (**C**) The amino acid sequence for each head domain is also shown highlighting specific amino acids

The structure of the Hbl-B, and of NheA, as well as the amino acid sequence similarity between all 6 components of HBL and NHE suggest structural homology with ClyA. Differences between ClyA and Hbl-B and NheA in the head domain, and the fact that the ClyA head domain is important in the pore formation of ClyA, make a similar mechanism of pore formation for NHE of HBL difficult to rationalise without the structure of the other NHE components, or a pore structure of the whole system.

1.6 Scientific interest in Pore forming toxins

Interest in research into pore forming toxins has been driven by the understanding that PFT are an ancient, diverse group of virulence factors, and by the knowledge that an understanding of their pore forming mechanisms is sure to aid in development of new classes of antimicrobials. These new antimicrobials, in contrast to those targeting pathogen survival, would target pathogen virulence, and could thus potentially decrease pathogenic species acquiring resistance (Clatworthy et al., 2007). More recently PFT have also been explored for technological uses as biological nanopores, for use in biotechnology (Panchal et al., 2002). Current developments have exploited PFTs as both nanosensors and mediaters to target chemotherapies to cancer cells, as well as to promote controlled protein export from cells. Both avenues of interest are discussed briefly below, in order highlight in context the importance, and current state of research in the PFT therapeutics and biotechnology fields.

1.6.1 Pore forming toxins as therapeutic targets

A number of therapies for diseases caused by *S. aureus* are being developed, some of which specifically target α-haemolysin. It was found that anti-α-haemolysin antibodies are able to protect against lethal staphylococcal pneumonia, reduce absess formation, and could be used synergistically with other therapies such as vancomycin. Development of therapies based on these results include naturally ocurring compounds such as oroxylin A, DNA aptamers, molecules targeting of the pore receptor ADAM10, and nanotoxoids in which liposomes are loaded with α-haemolysin to induce an immune response and antigen presentation (Dal Peraro and van der Goot, 2016; Dong et al., 2013; Vivekananda et al., 2014).

Small molecules such as GPI and GPI analogues have also been tipped to be useful against infections where the bacteria produce PFTs in which GPIs are required for pore formation, such as aerolysin and α-haemolysin (Wu and Guo, 2010).

Finally, attenuated toxins in which key residues are mutated or removed have been shown to make effective vaccine targets, as is shown specifically with Δ6 pneumolysin, which lacks toxic activity, but retains immunogenicity (Cockeran et al., 2011).

1.6.2 Biotechnological applications of pore forming toxins

The abundance of pore forming toxins found in nature with different specificities, different conductances and different modes of pore formation offers a large repository from which to design PFTs for biotechnological applications.

It has been shown that the pore formed by α-haemolysin is of a suitable size to thread DNA, and by engineering the internal face of the pore (Astier et al., 2006), DNA can be threaded at rates which allow the nucleotide identity to be determined as it is threaded. This has led to development of a new generation of DNA sequencing in which single molecules of DNA can now be sequenced with 99.8 % accuracy (Clarke et al., 2009).

Pores also offer an intriguing capability to design recombinant proteins, which can selectively bind and kill cancer cells. A recent example of this is the use of streptolysin O, which was expressed in tumour cells by insertion of the gene via an adenovirus (Yang et al., 2006). Expression of the gene led to pore formation from within the cell, ATP depletion, and cell necrosis. By targeting the plasma membrane directly, these treatments may work in cancer cells not sensitive to other treatments, and to slowly proliferating cancers such as prostate cancer. As the virus confers the specificity of the PFT to cancer cells, the pore forming toxin used can be in principle of broad specificity.

The final area of interest for PFTs in biotechnology centers on single molecule kinetic studies and diagnostic applications using ClyA pores as enzyme chambers (Soskine et al., 2015). Reports show that an enzyme of compatible size can be inserted into the lumen of a ClyA pore, and have lifetimes inside the lumen up to minute timescales. Enzymes that have shorter lifetimes can be made to insert for longer lifetimes by mutagenesis of ClyA, to insert the protein for biologically relevant timescales. Then, in a similar manner to the DNA sequencing using α-haemolysin, a voltage is applied across the pore, and substrate is added to one side of

a chamber. As the substrate is turned over, the voltage across the pore is changed, and so the kinetics of the reaction can be followed, on a single molecule scale, resolving each step in a multi-step reaction.

1.7 The aims of this thesis

Understanding the mechanism of ClyA pore formation has led to developments in biotechnological applications for the pore (Franceschini et al., 2013) and a greater understanding of α-PFTs in general (Mueller et al., 2009). The existence of a tripartite system with structural similarity to ClyA, but a novel mode of action, has warranted further investigation by a number of labs including our own. Structural and functional studies on the tripartite system NHE have led to some interesting insights into the mechanism of these toxins, including the elucidation of structure of NheA, which on the surface is difficult to rationalise as a α-PFT (Ganash et al., 2013). Developments in understanding of the NHE pore forming mechanism could lead also to novel anti-virulence drugs that are able to attenuate or completely abolish *B. cereus'* ability to cause diarrhoeal diseases.

The main aims of this thesis are to continue work into the NHE system, whilst also working to identify homologues to the NHE tripartite system, in which all three components are present, and to characterise these proteins biochemically and structurally. By gaining structures of each soluble component of the tripartite system by X-ray crystallography, and of assembled pores and pore intermediates by X-ray crystallography and electron microscopy, the biochemical results previously obtained for NHE can maybe be rationalised and combined to generate a detailed mechanism of pore formation and cell lysis for this family of tripartite pore forming toxins.

Important questions to ask about this tripartite system within the scope of this thesis include:

- Do homologous tripartite systems have the same minimal and maximal requirements for lysis?
- What is the specific role of each of the components of the system?
- How is the presence of three different components rationalised in the 10:10:1 model proposed by Granum et al.?

- How do the tripartite pores form?

Structural studies into these proteins may offer insight into these outstanding questions, and into pore forming toxins in general.

Chapter 2: X-ray theory

X-ray crystallography in the context of this thesis is the study of structures determined from the diffraction of X-rays from protein molecules. The X-ray diffraction technique is effective in determining protein structures as the distances that must be resolved between atoms are of the same order as X-ray wavelengths. Proteins, on the scale of 10 – 100 nm, cannot be studied by light microscopy, which employs visible light in the range of 380 to 750 nm. However, using X-rays for diffraction experiments in protein structure determination has two major drawbacks. The first drawback of the technique is that proteins diffract X-rays very weakly. This problem is circumnavigated by the formation of protein crystals; in which ordered lattices of protein molecules diffract X-rays constructively, and produce measurable diffracted rays. The second drawback of X-ray diffraction is that the phases of each diffracted wave are lost, as X-rays cannot currently be focused preserving phase information, and so the pre-focused diffraction image is recorded in the experiment.

The theory of X-ray crystallography are outlined in this chapter, based on 'Macromolecular Crystallography' and 'Crystallography Made Crystal Clear', beginning with producing a suitable protein sample, and working through the experimental setup to solve the phase problem and determine an electron density map describing the protein structure, building a protein model that describes the electron density, and validating said model for deposition in a protein structure database.

2.1 Protein crystallisation

2.1.1 Protein sample preparation

Crystallisation of a protein to be studied requires the protein be highly pure (upwards of 95 % pure), and highly concentrated (usually 10 to 20 mg/ml) in order to allow crystal formation. Impurities in protein samples can prevent crystal growth, and in some unfortunate cases lead to crystallisation of a contaminating protein. Protein concentration is important in being able to reach the limits of the proteins solubility, which is a requirement for crystal growth.

2.1.2 Protein crystallisation

The most common methodology for protein crystallisation involves supersaturating the protein in a buffered solution containing some salt and precipitant components. A drop of protein is mixed with the supersaturating crystallisation solution, and either suspended on a coverslip, or placed in a small well (Figure 2.1). A second solution, containing no protein, is placed into a second well and both are sealed, so vapour diffusion can occur across the two drops.

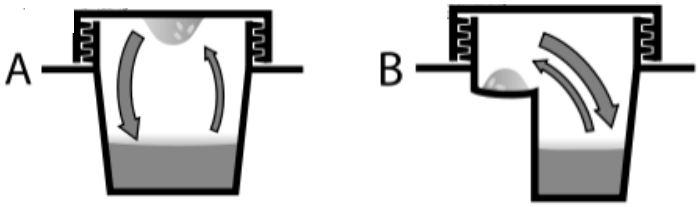


Figure 2.1 Methods of growing protein crystals. Two methods for producing crystals by vapour diffusion, namely hanging drop (left) and sitting drop (right). In both cases a small drop containing both protein and crystallisation are mixed and suspended in the same sealed volume, but separated from, a second solution containing only crystallisation solution. The two drops equilibrate via vapour diffusion due to differences in condition, leading in positive cases to crystal formation.

Vapour diffusion increases the effective concentration of protein, until one of two scenarios occur, granted the protein concentration is high enough. The protein either forms an amorphous precipitation, or begins the process of crystal formation, crystal nucleation (Figure 2.2). The formation of crystal nuclei reduces the concentration of protein in solution, and so no further nucleation occurs, and instead crystals begin a growth phase, and can grow in size to up to a millimetre in size.

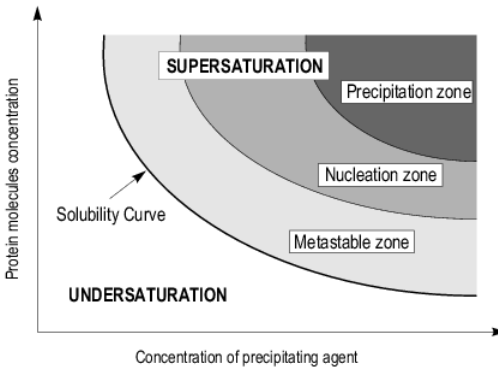


Figure 2.2 The solubility curve of protein a crystallisation experiment. The solubility curve of protein a crystallisation experiment, which begins the experiment as undersaturated, and then by vapour diffusion and concentration of both precipitant and protein passes the solubility curve to either the nucleation zone where crystals can form, the precipitation zone where the protein precipitates, or the metastable zone where crystals cannot form, but can grow if already nucleated.

Often crystals growing in initial conditions are of insufficient size or quality for diffraction experiments. These crystals are used to test that the crystals are protein in nature, and then optimisations are carried out to improve the quality of crystal growth. These optimisations generally use larger quantities of protein, to produce larger crystals, and a varied set of buffer pH and precipitant concentrations are tested around the original conditions. For crystals in which multiple crystals grow from a single nucleation site, such as needles and stacks of plates, additives such as salts can be used to try and produce single separated crystals. Once single crystals of adequate size and quality are formed, X-ray diffraction experiments can be carried out in order to determine the protein structure.

2.2 Radiation damage and Cryogenic temperature data collection

Although historically X-ray diffraction data was collected from crystals at room temperature, the radiation damage to the crystal due to X-ray exposure the succession of room temperature data collection by data collection at cryogenic temperatures (~100 K). Radiation damage is generated in crystals by inelastic scattering of X-rays and the generation of free electrons by the photoelectric effect in proportions of 8 % and 84 %, respectively for 12.4 KeV X-rays (Ravelli

and Garman, 2006). Each photoelectron generated by an incident X-ray can generate up to 500 secondary electrons, which propagate through the crystal by diffusion and lead to free radical formation and radiation damage symptoms. These symptoms are characterised by a decrease in the diffraction intensity and resolution, an increase in the overall temperature factors of atoms within the crystal, an increase in the cell dimensions of the crystal as the cell expands, and site specific damage. These site-specific damages include breakages of disulphide bonds, decarboxylation and dehydroxylation of tyrosine hydroxyl groups, in this order, as the X-ray dose increases (Ravelli and Garman, 2006). The global effects of radiation damage such as non-isomorphism between different stages in data collection also affect the ability to determine quality phases during experimental phasing, as will be discussed later.

Due to the deleterious effect of radiation damage, from difficulty in experimental phasing to errors in biological interpretation of the resulting protein structures, a means to reduce these effects has been determined. By cooling the crystal sample to cryogenic temperatures, the exposure lifetime of a general crystal can be prolonged by a factor of 70 (Nave and Garman, 2005), and the dose required to reduce the diffraction quality of a crystal by a half is around 2 x 10^7 Gy. Therefore, by careful experimental design, radiation damage can be minimised.

Radiation damage of protein crystals also increases as a function of wavelength, where wavelengths close to absorption edges of any heavy atoms present in the protein e.g. iron, leads to increased absorption, inelastic scattering and site-specific radiation damage to these specific atoms. Therefore, it is even more important to carefully design experiments that exploit these atoms, to prevent the earlier onset of radiation damage symptoms, which can include shorter data collections, and also back-soaking of any metal soaks, which otherwise will fill the solvent channels of the crystal with this absorbing metals, leading to huge radiation damage after very short X-ray exposures.

2.3 X-ray diffraction equipment

X-rays are generally generated for crystallography experiments by the acceleration of electrons, followed by either collision with metal ions to generate X-rays, or by radial acceleration of the electrons in a magnetic field leading to X-ray photon emission, termed X-ray tubes radiation and synchrotron radiation, respectively. The recent development of powerful synchrotron facilities

has diminished the requirement of X-ray tubes, although their mechanism of X-ray generation and advantages are discussed for historical insight.

2.3.1 Rotating anode tube X-ray generation

X-ray tubes, in recent time generated by rotating anodes, generate X-rays by bombarding a rotating metal disc anode with electrons from a fixed cathode. The incoming electron displaces a low-lying orbital electron from a metal atom in the anode by collision, which is replaced by an electron from a higher orbital. The excess energy from the higher orbital electron is emitted as a photon at the energies of an X-ray, and as such is emitted at a characteristic energy dependent on the energy spacing between the two orbitals. Home source X-ray tubes generally contain copper, in which an electron drops from the L or M shell to replace a displaced K electron, known as a Kα or Kβ transition respectively. These transitions emit characteristic X-rays of 1.54 Å and 1.39 Å wavelength respectively. The Kβ emissions are then absorbed by filters or removed using monochromators to ensure monochromatic light is produced from X-ray tubes. Monochromatic light is a requirement for X-ray diffraction, as two distinct wavelengths of X-rays would generate two spheres of reflection, and overlapping diffraction images.

The upgrade of a fixed X-ray tube to rotating anode confers the advantage of heat dissipation, as the surface of the metal is greatly enlarged and rotates to spread out the electron bombardment. This heat dissipation allows for a higher flux of X-rays, yielding more intense X-rays.

2.3.2 Synchrotron radiation

Synchrotron radiation has many advantages over X-ray tubes, namely that they produce X-rays of high brightness and intensity, and so exposure times and experiment lengths can be reduced greatly (minutes to milliseconds per image). Synchrotron radiation is also highly focused through collimators and optics, which decreases noise from solvent scattering, allows smaller crystals to be used, and multiple datasets to be collected from larger crystals. Synchrotron radiation is also tuneable, as it is not dependent on characteristic orbital energy gaps. Tuneable wavelengths allow access to multi-wavelength anomalous dispersion (MAD) experiments that require data collected at multiple different wavelengths, as well as single isomorphous replacement with anomalous scattering (SIRAS) experiments using heavy atoms

with absorption edges away from energies available on home sources, maximising phasing power.

Synchrotron radiation is produced by acceleration of electrons to GeV energy levels around a closed circuit by strong magnetic fields, generated by bending magnets. These electrons are stored in a storage ring, where they travel around the ring losing energy as X-rays, and are boosted externally by a radio-frequency cavity to compensate for the energy lost by the electron as synchrotron radiation on each rotation around the circuit, and to keep the electron energy high. As the electrons are accelerated, they emit X-ray radiation at a tangent to the direction of travel (Figure 2.3). On these tangents beamlines are built where users can place crystals for diffraction experiments.

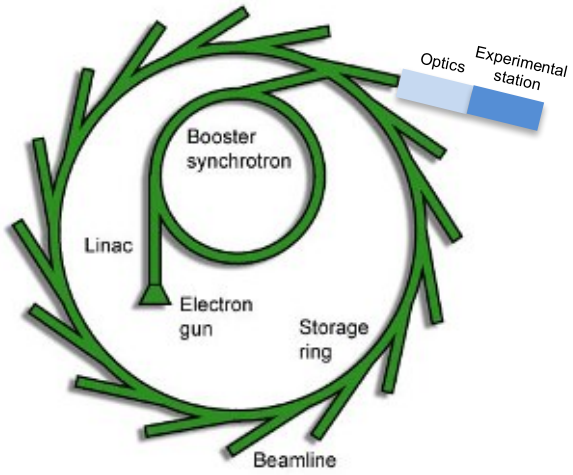


Figure 2.3 Schematic representation of a synchrotron. Schematic representation of a synchrotron showing the origin of the high velocity electrons in the center of the ring from the electron gun. The electrons are accelerated linearly in the linac, and then boosted up to GeV energies in the booster synchrotron. Finally the electrons are released to the storage ring where they are accelerated or bent by large electromagnets around the ring, which produce X-rays at tangents to the acceleration where beamlines are placed to utilise the photons produced.

In third generation synchrotrons, more intense X-rays can be emitted by electrons by manipulating their path within the closed circuit. So-called insertion devices manipulate electrons by a tightly spaced succession of magnets with alternating poles, which cause the electron to bend many hundreds of times within the insertion device (Figure 2.4). Such devices are called 'wigglers' and undulators, and the resulting electrons emit much more intense X-rays again at a tangent to the direction of travel.

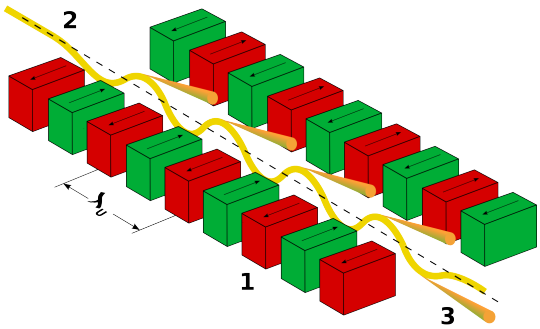


Figure 2.4 A schematic representation of an undulator or wiggler insertion device. The green and red bricks (1) represent closely spaced alternating pole electromagnets spaced to quickly change the direction and acceleration of the incoming electron beam (2). The result is the release of high-energy photons that can be focused by collimators to produce a brilliant, high-energy, monochromatic X-ray beam for X-ray diffraction experiments.

2.4 X-ray scattering and diffraction principles

X-ray diffraction occurs when single photons or wave packets interact with and excite all electrons within the photons coherence length, leading to those excited electrons emanating virtual waves, which interfere constructively and destructively in certain directions. The diffraction pattern is then generated by the sum of all scattered photons in the experiment, identically to the phenomenon observed in the double slit experiment. Each incident photon is not required to be coherent to one another for X-ray scattering to occur, and only depends on the coherence length of each photon. The constructive and destructive interference generating

the distinct diffraction pattern is contributed by a number of means discussed below, making up the geometric principles of X-ray diffraction used in X-ray diffraction techniques.

2.4.1 X-ray scattering by electrons in a single atom

In order to work towards determining protein structures by X-ray diffraction, it is first useful to determine how X-rays are diffracted by single atoms. Although the electrons present in atoms are confined to defined orbitals, the overall atomic electron density ρ(r) can be approximated as spherical. The atomic scattering function, which is also the Fourier transform of the atomic electron density ρ(r) can be described as

$$f_s = \int_r^{V(atoms)} \rho\,(r) exp(2\pi i S r)\, dr$$

Where $\rho(r)$ is the electron density of the volume unit with radius r, $2\pi iSr$ is the general expression for the phase difference of the scattered wave in reference to a fixed origin, and **S** is the scattering factor $\mathbf{S}_1$ - $\mathbf{S}_0$ where $\mathbf{S}_1$ is the scattered wave and S_0 is the incident wave. When S is set to 0, the exponent becomes 1 and the atomic scattering factor f_s becomes equal to the number of electrons in the atom. As **S** increases, the scattering becomes increasingly weak, leading to a Gaussian distribution centred on **S**=0. This phenomenon leads to two conclusions, the first being that scattering in the backwards direction is greatly decreased, which means that scattering drops off at higher resolutions i.e. at larger scattering angles, and secondly that the majority of scattering is at the same angle as the incident X-ray, although this is masked by the unscattered X-rays.

2.4.2 Diffraction from adjacent atoms, and from a molecule

Scattering of X-rays by two identical adjacent atoms leads to a maximum scattering twice that seen by a single atom, as now there is twice as many scattering electrons. The scattering is also further modulated by constructive and destructive interference between the partial waves emanating from the two atoms, related by the scattering angle and the interatomic distance d, giving rise to the diffraction pattern seen in X-ray experiments, but in one dimension. As the phase difference is defined as $\Delta\varphi = 2\pi Sr$, where r is the distance between atoms, maximum

constructive interference will occur at interger values of Sr, that is where $\Delta\varphi = 2\pi$. It therefore follows from this one-dimensional crystal form where the distance between atoms is the cell basis vector **a**, that in the three-dimensional molecular case, maximum constructive interference will occur when the following equations are integer values for each direction, i.e. $S \cdot a = n1,\ S \cdot b = n2,$ and $S \cdot c = n3$. These three equations are known as the Laue equations.

Scattering from a molecule is therefore further modulated by the irregular distances between atoms within the molecule, and so scattering of X-rays by a single molecule leads to a continuous decaying complex function, given by a summation of all of the partial waves emanated from each atom. This is given by:

$$F_s = \sum_{j=1}^{atoms} f_{s,j}^{0} \cdot exp(2\pi i S r)$$

Where the contribution of each atom to the measured intensity is given by the atomic scattering factor $f_{s,j}^{0}$. In reality, scattering from a single molecule is unmeasurable for atomic resolution X-ray diffraction, and so the scattering is amplified by producing an ordered array of proteins all scattering in phase, to produce measurable reflections. Scattering from a crystal is the summation of aligned unit cells, where a perfect alignment between cells leads to constructive interference, and an increase in scattering proportional to the number of cells in the crystal. Therefore, the overall structure factor for a molecule and a crystal is identical, however with the addition of Ncell to sum the number of identical cells present in the crystal.

It is usual to substitute the S and r terms in the general structure factor equation, to remove the wavelength and cell dimensions from the exponent, such that the final general structure factor equation is

$$F_{hkl} = \sum_{j=1}^{atoms} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)}$$

2.4.3 X-ray wave theory and production of electron density maps

Bragg's law elegantly reduces the **S**r dependence of diffraction to a simple equation as follows

$$2d\ sin\theta = n\lambda$$

where d is the lattice spacing, θ is the angle of incidence and scattering, n is a positive integer, and λ is the wavelength of the X-rays. Bragg's law can also be shown geometrically (Figure 2.5), showing how when the equation is satisfied, the partial waves emanating from the X-ray scattering event are in phase and interfere constructively.

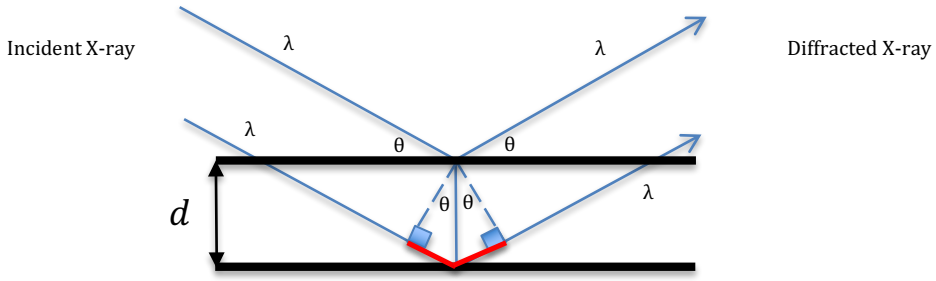


Figure 2.5. Diagrammatic representation of Bragg's law. When the distance 2d•sinθ (red lines) is equal to a whole number of wavelengths nλ, the incident X-rays will diffract from the set of planes, with distance d, in phase.

Braggs insight that led to this equation was to rearrange the scattering vector to face upwards, such that X-ray scattering occurs on a series of planes, and so that the magnitude of S was the reciprocal of the interplanar distance d. These d spacing's can then be given indices values, called miller indices, based on how they divide the cell, such that h00 describes planes parallel to the a cell unit, 0k0 describes planes parallel to the b cell dimension, and 00l describes planes parallel to the c cell dimension, and combination of hkl give rise to different scattering planes dividing the unit cell in reciprocal space (Figure 2.6).

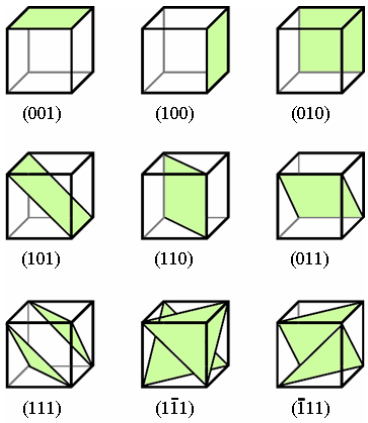


Figure 2.6. Planes shown representing those generated by different Miller indices. Miller indices describe the reflection planes within the unit cell, shown as a cube, bringing different atoms onto the different planes, satisfying the Laue equations and leading to diffraction in phase and a recorded reflection. Combinations of different indices leads to planes in different directions, angles, and number within a cell i.e. 001 is parallel to a and splits the cell 0 times, whereas 110 splits the cell in two and is relatively at a 45° angle.

These miller indices then are integer values that satisfy the Laue equations at certain scattering angles, and can be substituted into the structure factor equation in place of the scattering vector S, with the vector r also being replaced by fractional co-ordinates x,y, and z. Diffraction then occurs at scattering angles or interplanar distances that satisfy these Laue equations, and can be predicted effectively using a geometric construct called an Ewald sphere or construct.

The Ewald construction can be generated to show geometrically where a scattered wave occurs, and therefore where diffraction will occur on the imaging plate, and when in the rotation of the crystal certain reflection planes will lead to diffraction. The Ewald construction makes use of the reciprocal space, and the conditions where Braggs law are satisfied in reciprocal space are given by the modified equation

$$\mathrm{d}^*_{hkl} = \frac{1}{d_{hkl}} = \frac{2\,sin\theta}{n\lambda}$$

and it can be shown that d^*_{hkl} = |S|, where |S| is the scattering vector. By constructing a sphere of radius $1/\lambda$, and rotating the reciprocal lattice around an arbitrary lattice point placed on the surface of the sphere at the point opposite the incoming X-ray, shown as (000) in Figure 2.7, any other lattice point which lies on the surface of the sphere satisfies the Bragg equation, resulting in a reflection.

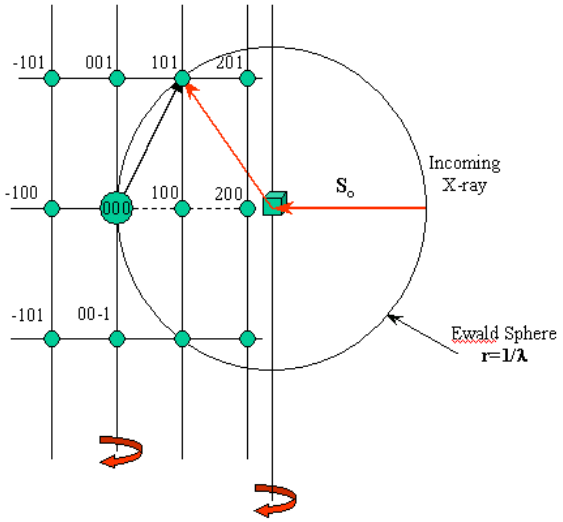


Figure 2.7. A schematic representation of the Ewald construction. Ewald construction above shows the incident X-ray vector $\mathbf{S}_0$ incoming from the right, and being diffracted by the crystal shown as a cube. The sphere shown here in two dimensions has radius $1/\lambda$, and as miller indices are rotated into contact with the surface of the sphere, a reflection is recorded, in this case reflection 101 in relation to the origin 000. Rotation of the crystal rotates the reciprocal lattice so new reflections come into contact with the sphere and satisfy Bragg's law.

X-ray diffraction experiments are designed such that the crystal is rotated in the beam, so that every lattice point is brought to the surface of the Ewald sphere, or made to satisfy the Bragg equation, and so every reflection is recorded in a single experiment. In order to achieve this, the crystal is mounted on a goniostat, which is able to rotate the crystal through 360° around phi (Figure 2.8). Goniostats can also rotate the crystal through kappa in order to sample a different

orientation of the crystal, especially useful for low symmetry spacegroups where the number of symmetry equivalent reflections are low. As the position of the reflections on the detector depends on the satisfaction of the Bragg equation by the alignment of certain planes designated by Miller indices, the distance between reflections is fixed by the unit cell dimensions of the crystal (and the distance between two adjacent reflections is the reciprocal of the unit cell dimension for that direction). The intensity of each reflection is determined by the atomic positions for each plane, as well as the symmetry of the spacegroup, which leads to systematic absences of certain reflections. As such, the spacegroup and unit cell dimensions can be determined directly from the diffraction pattern.

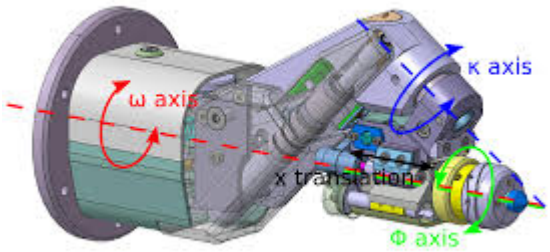


Figure 2.8. A schematic representation of a modern mini-kappa goniostat. A modern electronic goniostat with rotation capabilities in ω and φ, which are coincident at a κ angle of 0°, and allow for the full reciprocal space of reflections to be sampled from a single crystal. When the κ angle is altered, new orientations of the crystal are aligned for rotation about Φ. The goniostat also allows translation to align the crystal within the X-ray beam.

2.4.4 Describing electron density as a complicated wave function

Calculating an electron density map that describes the electron cloud surrounding a protein is the goal of a crystallographer during X-ray crystallography experiments. The electron density map for a protein is a complicated three-dimensional wave function, for which a triple sum equation, including amplitude, frequency, and phase, is able to describe the volumetric electron density distribution $\rho(x, y, z)$.

$\rho(x, y, z)$ can be described by:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} . e^{-2\pi i(hx+ky+lz)}$$

where V is the volume of the cell, F_{hkl} is a structure factor described below, and $e^{-2\pi i(hx+ky+lz)}$ is an exponent and is equal to $cos2\pi i(hx+ky+lz) + isin2\pi i(hx+ky+lz)$. The structure factor F_{hkl} contains the amplitude of the waves at hkl, as well as the phase α at each hkl (α_{hkl}), as described below

$$F_{hkl} = |F_{hkl}| \cdot e^{i\alpha_{hkl}}$$

where $|F_{hkl}|$ is the measurable amplitude of the reflection at hkl, and α_{hkl} is the phase of the reflection, and is lost during the experiment. This representation of F_{hkl} can then be substituted into the $\rho(x,y,z)$ equation to explicitly describe $\rho(x,y,z)$ as a sum of amplitudes, frequencies, and phases:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F_{hkl} . e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})}$$

The phase α'_{hkl}, now explicit in each Fourier series, must be calculated for each reflection as it is not recorded during the experiment. The Fourier transform of $\rho(x,y,z)$ produces the structure factor equation, which describes the wavefunction of each reflection measured in a X-ray diffraction experiment

$$\boldsymbol{F}_{hkl} = \int_{V} \rho(x,y,z) . e^{2\pi i(hx+ky+lz)}\ dV$$

where the integral $\int_V \ dV$ is the sum of infintesmally small units of volume, and can be performed due to the volume being continuous, unlike the discreet reflections which take integer values. As each structure factor is the sum over the entire volume, each reflection hkl has contributions from each atom within the unit cell. The structure factor equation can also be described as a sum of the atomic contributions, rather than the volume electron density, such that each reflection is described by X-ray scattering from each atom, such that:

$$\boldsymbol{F}_{hkl} = \sum_{j}^{n} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)}$$

The atomic structure factor equation here relates the above electron density equation in terms of volume to the equation in terms of scattering atoms. In this final form, the relationship between the electron density as a cloud of electrons over a volume, and the scattering of X-rays by individual atoms is made clear.

2.4.5 Structure factors as complex vectors

Structure factors for waveforms of $cos\theta + isin\theta$ can be plotted as complex vectors (Figure 2.9), where the length of the vector is the structure factor amplitude, and the angle of the vector is the phase α.

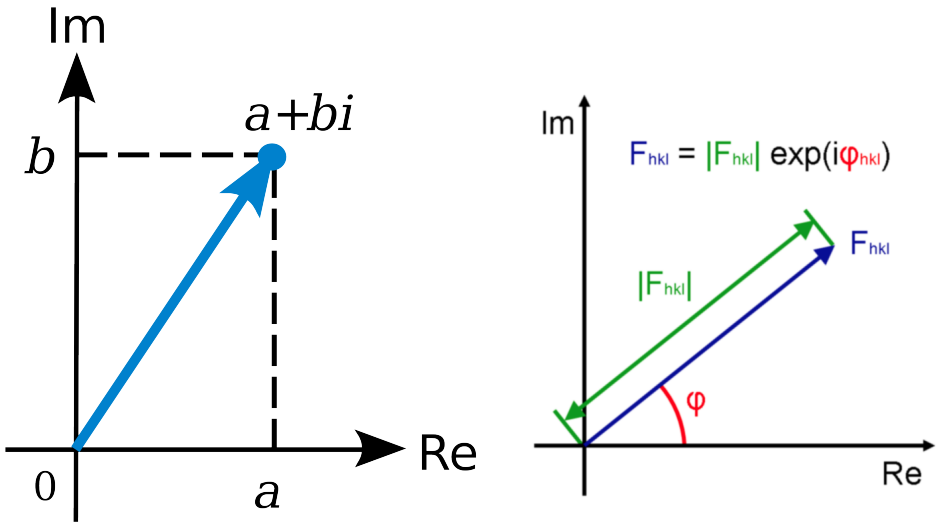


Figure 2.9. Vector representation of complex wave functions. Representing a vector as the sum of a in the real direction, and b in the imaginary direction (left) can be useful in describing structure factors of the form $cos\theta + isin\theta$. In this form the magnitude of the vector represents the amplitude of the wavefunction $|F_{hkl}|$, and the direction of the vector represents the phase of the wave function φ_{hkl} (right).

Representing structure factors as vectors has a useful application, in that the sum of the Fourier terms which describes the electron density or scattering from a group of atoms can be determined by vector addition (Figure 2.10), where the amplitude and phase of each complex structure factor is determined by summing the atomic structure factors. An elaboration of these Argand diagrams is particularly useful in determining phases in experimental phasing, which is discussed later.

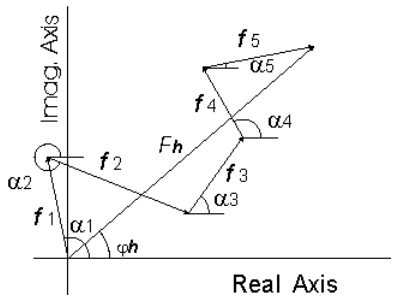


Figure 2.10. An Argand diagram representing vector addition of atomic structure factors. The vector addition of each atomic structure factor represented as f1-f5 (where each represents a different atomic structure factor $\mathbf{F}_{jhkl}$) can be shown diagrammatically as the head-to-tail alignment of each successive vector starting from the origin. The resulting vector can then be generated by drawing a vector from the origin to the head of the final vector with magnitude Fh and phase φh for the structure factor F_{hkl}.

2.5 Data collection

The X-ray experiment undertaken to determine the structure of a protein requires the intensities of all of the diffracted X-rays from the crystal to be measured. To achieve this, the crystal is mounted first in a loop, and then the loop is mounted in the X-ray beam, and cooled to ~100 K to reduce radiation damage. The crystal is rotated around an axis by the goniostat, and a series of images of the diffracted X-rays are measured using a detector. Software is then used in order to process the images for structure determination.

2.6 Data processing

2.6.1 Indexing reflections

Once a full dataset or test diffraction dataset has been collected, the lattice symmetry can be determined and the spacegroup estimated. It is possible to assign multiple lattice symmetries with different cell dimensions to a given diffraction pattern, and so each symmetry is tried as an indexing solution, and then a penalty calculated based on how well the diffraction agrees with the rules for that given spacegroup. Cell dimensions are also estimated for each solution based on Fourier methods, to generate an orientation matrix and a set of basis vectors and cell dimensions. It is then convention to choose the highest lattice symmetry for which the penalty is lowest, although in difficult cases in the test stage, it is always best to index into a lower symmetry and collect more data as if the symmetry is higher it is possible to merge these reflections and increase redundancy of each reflection, but it is not possible to regenerate uncollected reflections if a higher symmetry is not assumed that is present in reality. Once a lattice symmetry and point group is estimated, equivalent reflections are grouped and indexed i.e. the Miller indices of the reflections **h** or hkl are assigned, and the raw intensity values for each reflection are determined.

2.6.2 Integration and initial scaling of indexed reflections

Once the lattice type and cell dimensions are estimated, the unit cell dimensions are first refined by inclusion of all of the images, to give a new slightly altered unit cell before integration of all reflections. In modern data collections where fine slicing is employed, it is desirable to collect a single reflection over a number of images, as partial reflections. In order to determine the intensity for the reflection, as well as an error estimation based on the counts outside of the

reflection profile, the many partial reflections over a number of frames are integrated to give a single intensity profile. Each intensity must be converted to a structure factor amplitude, and it can be shown that the relationship between measured intensities and structure factor amplitudes is:

$$I_{hkl} = I_0 k \frac{N}{U} \lambda^3 LPA F_{\boldsymbol{h}}^2 = \frac{F_{\boldsymbol{h}}^2}{k \cdot k'_{\boldsymbol{h}}}$$

where I_0 is the intensity of the incoming beam, N is the number of unit cells and U the unit cell volume, L is the Lorentz factor that corrects for relativistic phenomenon, A is the absorbance, and F is the structure factor amplitude, proportional to the square root of the intensity. In the second part of the equation, the angle independent factors are grouped into k, and the angle dependant scaling factors are groups into $k'_{\boldsymbol{h}}$. Once the intensities are converted to structure factors, a few temporal scaling factors can be applied. The merging R-factors for all reflections on a frame should be the same for each frame, and so should produce a linear reduced chi-squared plot when plotted against all frames in the dataset. Scaling can be applied in cases where this is not the case, such as in areas of the crystal that are thinner than others e.g. for a plate morphology crystal. A second scaling factor can be applied to each reflection, where a number of identical reflections are measured separated temporally. In a later frame, the intensity for a measured reflection, also measured earlier in the data collection may be lower than the previously measured intensity, due to radiation damage. For non-extreme cases, these reflections can be scaled to correct for radiation damage, and in extreme cases, smaller wedges may be scaled individually, or the radiation-damaged frames can be removed from the data collection altogether to improve scaling results.

2.6.3 Merging scaled reflections

Once indexed and scaled, equivalent reflections are merged to give an overall averaged intensity value for each reflection in the crystal lattice. Merging is performed based on the rules of the crystal symmetry, and then a number of metrics are calculated in order to determine the quality of the merging process. These metrics are calculated for the high-resolution bin, useful

for determining a suitable high-resolution cut-off of the data, the low-resolution bin, which should be higher quality than the high-resolution data, and for all of the data, which can be useful for determining if a too high symmetry spacegroup has been assigned. The most common quality indicator for merging is the linear merging R-value:

$$R_{merge} = \frac{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} |I_{(\boldsymbol{h})\boldsymbol{i}} - \bar{\mathrm{I}}_{(\boldsymbol{h})}|}{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} I_{(\boldsymbol{h})\boldsymbol{i}}}$$

Where **h** is each reflection hkl, N is the number of reflections merged, I is the intensity of the specific reflection and $\bar{\mathrm{I}}$ is the averaged intensity of each reflection hkl. This linear R-value can also be used for merging different kinds of reflections i.e. symmetry-related reflection R_{sym} and merging Bijvoet pairs for anomalous data collection R_{anom}, given by:

$$R_{anom} = \frac{\sum_{\boldsymbol{h}} |I_{(-\boldsymbol{h})} - I_{(+\boldsymbol{h})}|}{\sum_{\boldsymbol{h}} \bar{\mathrm{I}}_{(\boldsymbol{h})}} = 2\frac{\sum_{\boldsymbol{h}} |I_{(-\boldsymbol{h})} - I_{(+\boldsymbol{h})}|}{\sum_{\boldsymbol{h}} (I_{(-\boldsymbol{h})} + I_{(+\boldsymbol{h})})}$$

For highly redundant data collections, such as those collected in anomalous datasets, the linear R-value will increase as the multiplicity increases, as the redundancy of the data is not taken into account. A second precision merging R-value R_{pim} is favoured in these cases, given by:

$$R_{pim} = \frac{\sum_{\boldsymbol{h}} (\frac{1}{N-1})^{1/2} \sum_{i=1}^{N} |I_{(\boldsymbol{h})\boldsymbol{i}} - \bar{\mathrm{I}}_{(\boldsymbol{h})}|}{\sum_{\boldsymbol{h}} \sum_{i=1}^{N} I_{(\boldsymbol{h})\boldsymbol{i}}}$$

Where the $1/(N-1)^{½}$ term accounts for many cases of the same reflection being measured. For that reason, R_{pim} is much lower usually than R_{merge}, and can often be lower than 0.1 for the low resolution bin. R_{pim} can also be useful in determining the amount of anomalous signal in a dataset by comparing the value with R_{anom}, where R_{anom} is analogous to signal, and R_{pim} is analogous to noise, and a value of R_{anom}/R_{pim} = 1.5 indicates a substructure solution is likely achievable (Rupp, 2010).

2.7 Solving the phase problem

2.7.1 The phase problem

As discussed in previously, the diffraction pattern collected during an X-ray diffraction experiment contains frequency information in the position of each reflection in the diffraction when the unit cell dimensions are known, and the amplitude measured as reflection intensity. The phase is however lost in the experiment, and so the crystallographers job is to determine this phase of each reflection, the so-called phase problem.

There are two main methods for determining phases in crystallography, experimental phasing methods, and molecular replacement. Experimental phasing methods take advantage of the Patterson function, a variation of the Fourier sum, which requires no phase information. Molecular replacement also exploits the Patterson function, and exploits the fact that a Patterson map of a homologous protein structure can be rotated and translated into the unknown unit cell, minimising the difference between the model Patterson map and the experimental Patterson map, to generate a starting set of phases for improvement by model rebuilding and refinement. Both techniques are explained briefly in the context of this thesis.

2.7.2 Experimental phasing

Experimental phasing begins with simplifying the phase problem, by incorporating and exploiting heavy atoms into the crystal and determining the positions and hence amplitudes phases of these atoms. Once determined, they can be used in estimating a starting set of phases for the experimental structure factors. Experimental phasing can either exploit heavy atom contributions to diffraction, in the heavy atom method, or the breakdown of Friedels law in anomalous scattering when the X-ray wavelength is close to an absorption edge of a heavy atom.

2.7.3 Heavy atom methods for calculating phases

This approach takes advantage of the fact that the structure factors of a heavy atom-containing crystal ($\mathbf{F}_{PH}$) are equal to the sum of the protein structure factors ($\mathbf{F}_{P}$) and the heavy atom structure factors ($\mathbf{F}_{H}$) (Figure 2.11):

$$\boldsymbol{F}_{PH} = \boldsymbol{F}_{H} + \boldsymbol{F}_{P}$$

By measuring the structure factor amplitudes of each $\boldsymbol{F}_P$ and $\boldsymbol{F}_{PH}$ experimentally, and determining the amplitude and phase of $\boldsymbol{F}_H$, the phases of $\boldsymbol{F}_{PH}$ can be estimated. This relationship can again be shown graphically using vectors (Figure 2.11).

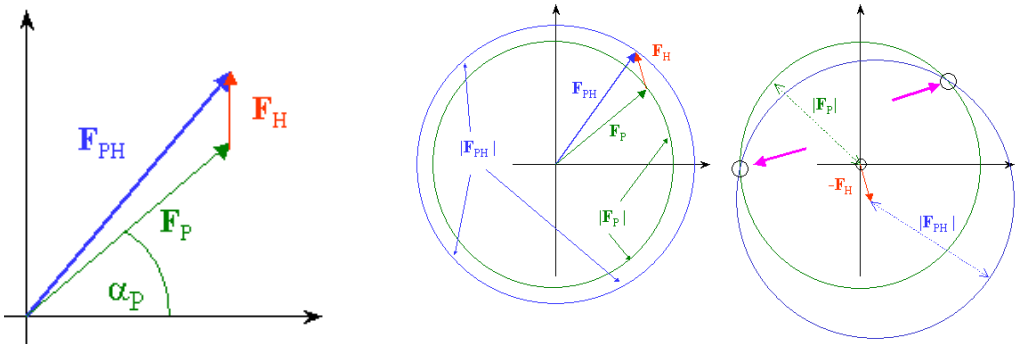


Figure 2.11. Solving the protein phase F_p by Harker construction. The vector representation of $\mathbf{F}_{PH}$, $\mathbf{F}_P$ and $\mathbf{F}_H$ show how the vector addition of $\mathbf{F}_P$ and $\mathbf{F}_H$ equals FPH. When the phase of $\mathbf{F}_{PH}$ and $\mathbf{F}_P$ are unknown but the amplitudes are, the vectors can be drawn as circles to represent all phases. If the $\mathbf{F}_{PH}$ circle is offset form the origin by the calculated $\mathbf{F}_H$ vector (where both the amplitude and phase are known), the phase is reduced to two possibilities symmetric around $\mathbf{F}_H$.

From this vector representation, we are able to plot the experimental situation, the one in which the length (amplitude) of $|F_H|$ and $|F_{PH}|$ are known, but the phases are unknown, and so drawn as a circle where the circle represents all phases. With known amplitudes and phases for $\mathbf{F}_H$, we can draw the circle for $|F_{HP}|$ offset by vector $\mathbf{F}_H$, to gain two possible phases for $|F_P|$. In reality the calculated amplitude is smaller than shown, and phase for $\mathbf{F}_H$ will have an error, and so the error in the phase must be small enough to generate an interpretable electron density map for phase improvement by means discussed later.

2.7.4 Anomalous scattering methods for calculating phases

With no anomalous scattering atoms present, the reflections h,k,l and –h,-k,-l have the same intensity, but with phases differing by 2π. These related pairs are called Friedel pairs. Phasing by anomalous scattering methods takes advantage of the symmetry of F(h,k,l) and F(–h,-k,-l)

reflections (Friedel's law), and uses the absorption of X-rays with wavelengths close to the absorption edge of certain atoms to break this symmetry (Figure 2.12). At these wavelengths, the heavy atom absorbs some of the photons, which excite an inner electron to a higher energy level. Upon returning of the excited electron to the lower energy, a new photon is emitted, with an altered phase. In order to incorporate these changes in the anomalously scattered wave, a real and imaginary component are included in the altered structure factor identity from 2.7.3

$$F_{PH}^{\lambda 2} = F_{PH}^{\lambda 1} + \Delta F_r + \Delta F_i$$

The $F_{PH}^{\lambda 1}$ and ΔF_r are combined to give the dispersive component f', and ΔF_i is known as the anomalous component f''. A Harker construction of these differences shows how $F_{PH}^{\lambda 1}$ is related to $F_{PH}^{\lambda 2}$ geometrically (Figure 2.12)

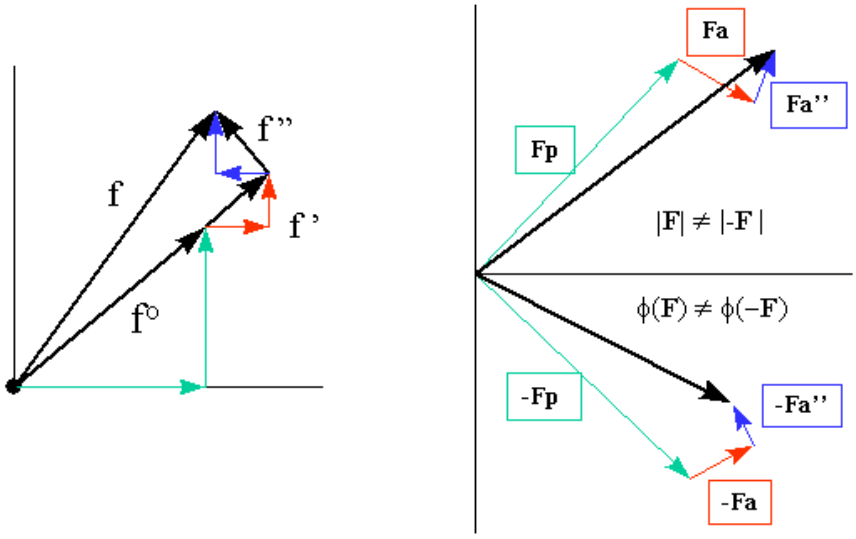


Figure 2.12. An Argand diagram showing how the anomalous f' and dispersive f'' contributions to F_{PH} break Friedels law. The anomalous protein structure factor f in an anomalous experiment is the sum of the protein structure factor f° and the real and imaginary anomalous scattering terms. As the imaginary term f'' is not symmetrical about the real axis, its contribution to hkl and –h-k-l are different, leading to an amplitude and phase differences for these otherwise identical reflections.

The contribution of the anomalous scattering factors to $\mathbf{F}_P$ have different effects on $\mathbf{F}_P$ and $-\mathbf{F}_P$, or Friedel pairs. As can be seen in Figure 2.12, the real part of the anomalous scatterer, the **F**a term between $\mathbf{F}_P$ and $-\mathbf{F}_P$, is inverted through the real axis, whereas the **F**a'' is inverted through this same real axis, but then also changes sign, so that it runs the opposite way. This difference in the imaginary contribution to anomalous scattering leads to a different amplitude **and** phase for $\mathbf{F}_{PH}$ and $-\mathbf{F}_{PH}$, breaking the symmetry of Friedels law, such that:

$$|F_{PH}^{(+)}| \neq |F_{PH}^{(-)}|$$

Determining phases by anomalous scattering is similar to SIR in the sense of the above vector diagrams, but is complicated by the structure factor of the protein and anomalous scatterer being shifted relative to the native protein structure factor by a real and imaginary vector (Figure 2.12). The 'native' $\mathbf{F}_P$ structure factors in an anomalous experiment are usually collected from the same derivitised protein, with data collected away from the wavelength of anomalous scattering of the heavy atom, where Friedels law is restored.

In order to determine suitable wavelengths for SAD and MAD experiments to maximise the f' and f'' contributions, a fluorescence scan is performed around the known absorption edge for the anomalously scattering atom. The real f'' contribution can be determined directly from the results of the fluorescence scan, and plotted graphically (Figure 2.13).

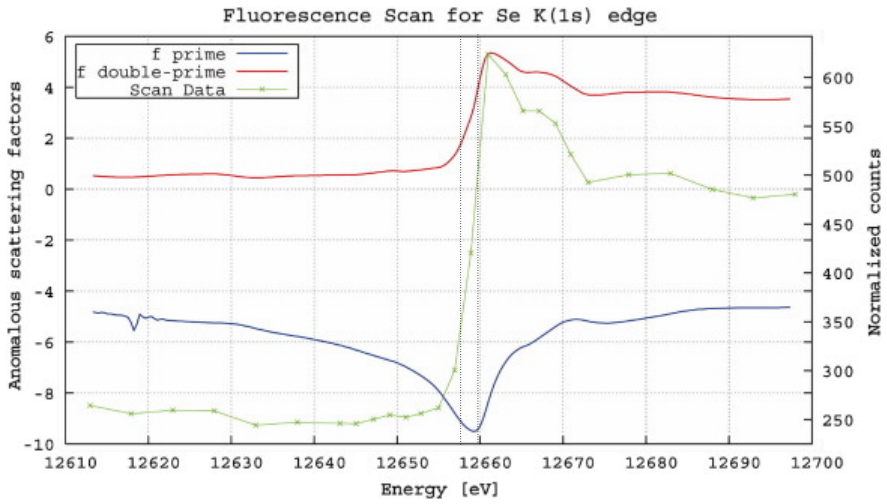


Figure 2.13. The output fluorescence scan from Chooch with calculated f' and f'' plots. A fluorescence scan for the K edge of a Se atom (green) shows that the atom absorbs X-rays around 12661 eV, and the absorption steadily decreases above this energy. From this scan, the real and imaginary scattering contributions can be calculated (red and blue lines respectively) in order to design a SAD or MAD experiment. Dotted lines show the inflection (left), which maximises the anomalous scattering contributions, as seen by the negative peak in the calculated anomalous scattering. The middle dotted line shows the peak of fluorescence and the real anomalous scattering factor f'', and finally the right most dotted line shows the high energy remote wavelength, usually collected a few 100 eV above the peak energy.

The imaginary f' can be determined indirectly from the derivative of the f'' plot via the Kramer-Kronig relation, from the same fluorescence scan. Via the Kramer Kronig transformation, the electron contributions to each f' and f'' can be determined. To maximise the f'' contribution, the wavelength is chosen to be at the region of peak absorption, that is at the peak of the f'' plot. The maximum f' contribution is reached when the gradient of the f'' peak is at its highest, called the inflection point. A third wavelength is usually also collected, at a wavelength remote from the peak and inflection point, called the high energy or low energy remote, dependent on which direction the wavelength is moved.

The two anomalous contributions f' and f'' are wavelength dependent, and the anomalous difference, i.e. the difference between amplitudes of the structure factors of |F+| and |F-| using the following equation:

$$\Delta F_{anom} = [|F^+| - |F^-|] \cdot \frac{f_0 + f'}{2f''}$$

2.7.5 Determining phases by single isomorphous replacement with anomalous scattering

The two above methods of heavy atom isomorphous replacement and anomalous scattering can often be combined, if the absorption edge of the heavy atom to be incorporated into the protein has an absorption edge wavelength available at synchrotrons. This method leads to a breaking of the phase ambiguity faced in SIR and SAD due to the centrosymmetry of the phase calculation in both, because in combining the SIR phase ambiguity where the symmetry lies around the heavy atom phase αH with the ambiguity of SAD which is centred around αH-90°, only one intersection point remains, that being the true phase of the protein (Figure 2.14).

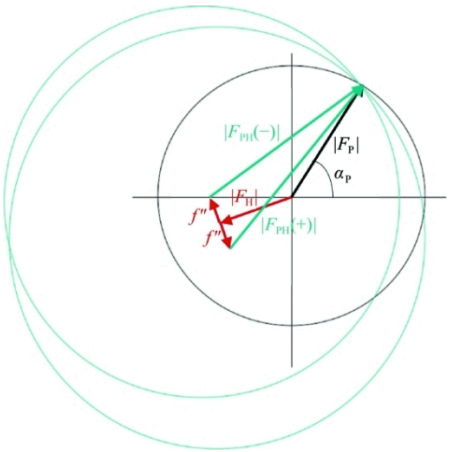


Figure 2.14. Harker construction of SIRAS methods. A Harker construction of a SIRAS experiment, showing that the phase ambiguity is broken by SIRAS methods due to only one point of intersection in common between heavy atom isomorphous

differences in $\mathbf{F}_{PH,}$ and the anomalous scattering vectors for $\mathbf{F}_{PH}^{(+)}$ and $\mathbf{F}_{PH}^{(-)}$, due to the anomalous scattering components f' and f".

2.7.6 Errors in phase determination

In reality, the above geometric Harker constructions for each phasing method are shown for ideal cases where no experimental errors exist such as no error in the intensity measurement for each reflection, or no error in the determination of the heavy atom phases. These errors, which can be minimised by careful experimental design, but that still exist in real experiments can be taken into account in programs used for phasing, and shown in the geometric Harker construction by the following means:

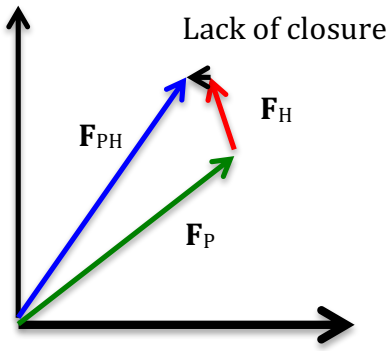


Figure 2.15. Argand diagram showing the lack of closure due to experimental error. In real cases, the phase and amplitudes of $\mathbf{F}_{H}$ **and** $\mathbf{F}_{PH}$ are not measured perfectly, and so there is a lack of closure in the three vectors $\mathbf{F}_{PH}$, $\mathbf{F}_{H}$, and $\mathbf{F}_{P}$. The phase of FP is therefore chosen to minimise the lack of closure between the three structure factors.

The above diagram shows that instead of a single phase value generated by infinitesimally thin circles, a centroid phase probability is generated (Figure 2.16), and the phase is chosen to give the smallest distance of closure of $\mathbf{F}_{H}$ and $\mathbf{F}_{PH.}$ Each phase is then also given a figure of merit based on this error, which then weights the structure factor in calculating the electron density map.

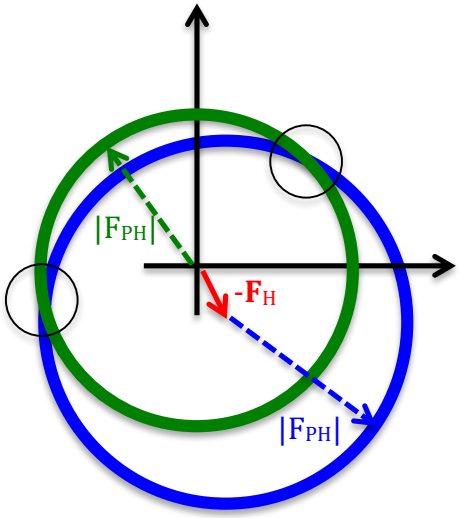


Figure 2.16. Harker construct showing the phase errors arising from errors in intensity measurements. As |FPH| is not measured infinitely accurately, an error is propagated, which can be shown as a widening of the $\mathbf{F}_{PH}$ circle. Therefore, the point at which the two circles cross, indicating the phase, has a distribution of values rather than a single value.

2.7.7 Locating heavy atoms in the unit cell

The above methods can be used to determine the phases of $\mathbf{F}_{PH}$, but each requires the location of the heavy atoms to be known in order to calculate a $\mathbf{F}_{H}$ vector to offset the $\mathbf{F}_{PH}$ vector in the $\mathbf{F}_{P}$ construction. The most powerful and commonly used method for locating heavy atoms is the Patterson method, which utilised a variation of the Fourier sum that requires no phases, called the Patterson function, given as:

$$\rho(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} |F_{hkl}^{2}| \cdot e^{-2\pi i(hu+kv+lw)}$$

where u, v, and w are Cartesian co-ordinates in Patterson space, and $|F^2{}_{hkl}|$ is the square of one structure factor, which can be determined as proportional to a measured reflection intensity.

The Patterson function generates a contour map containing peaks at locations corresponding to vectors between atoms. The number of peaks is equal to n(n-1) where n is the number of atoms in the cell, and so for a protein is a huge number. A more useful Patterson function for phasing purposes is the difference Patterson function where the amplitudes come from $F_{PH} - F_P$ for heavy atom experiments and Δf_{anom} for anomalous experiments, which plots only the distances between heavy atoms in the cell. Based on the spacegroup of the crystal, there are then special regions in the difference Patterson map where one dimension cancels out, such that the peaks for the heavy atoms are found in a plane. By looking at these regions, called Harker sections, the calculation for determining the position of the heavy atoms is made simpler, and once determined can be used to generate heavy atom structure factors and therefore the vector F_H

2.7.8 Molecular replacement methods for calculating phases

Molecular replacement is a powerful method in determining phases for isomorphous and non-isomorphous but homologous protein structures from known structures. For isomorphous protein models, where a small change like the soaking of a ligand into a crystal, the new protein structure factors can be calculated using the new measured amplitudes $|F_{hkl}^{new}|$ and the previously calculated phases α'^{calc} from the phasing model. For non-isomorphous proteins, a homologous protein for which a structure is known can be used. In this case, the orientation and position of the protein is unknown in the cell, although the unit cell dimensions and symmetry are known from the diffraction, and can limit the position of the protein in the cell. In order to place the known protein model correctly in order to calculate some starting phases, first the orientation of the protein is calculated. This part is carried out first exploiting the self-vectors in the Patterson map of both the calculated and observed data, as the self-vectors in the Patterson function are rotation dependent but translation independent. The trial rotations can then be scored, based on how well the calculated Patterson at a certain orientation matches the observed data Patterson map. This scoring is called the rotation function Z-score. Once a series of well scoring orientations are found, the translation of these models are tested, in a translation function, namely the T2 translation function (Harada and Lifchitz, 1981). This translation function exploits the cross-vectors in the Patterson map for the model placed in the desired cell, and the observed cross-vectors, subtracting the self-vectors already calculated for the rotation

function. Translations are scored by the Translation Z-score, based on how well the Patterson maps agree at certain translations.

2.7.9 *ab initio* phasing

Historically, *ab initio* methods for solving the phase problem, specifically direct methods, have been used to solve the structure of molecules with relatively few atoms (<200), and for those molecules that diffract to high resolutions. The invention of the Shake and Bake algorithm (Miller et al., 1994) expanded the number of atoms for which this method works to ~ 1000, increasing the potential for atomicity based methods.

Current *ab initio* phasing methods within the scope of this thesis make use of small secondary structure fragments, especially α-helices, rather than a selection of randomly placed atoms, in order to reduce the resolution requirement, and also increase the number of atoms per molecule for which *ab initio* methods can be successful. Arcimboldo (Sammito et al., 2014), tries to place a helix fragment or known substructure by molecular replacement, and then performs density modification and main chain tracing in order to identify potential solutions. These solutions are then cycled and improved until a sufficient set of starting phases can be calculated, as determined by the correlation coefficient.

2.8 Calculating, manipulating and interpretation of electron density

2.8.1 Calculating an electron density map

Once starting phases have been determined experimentally or through molecular replacement or *ab initio* methods, an electron density map can be calculated using the familiar equation using the observed amplitudes and the calculated phases

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{obs}| \cdot e^{-2\pi i(hx + ky + lz - \alpha'_{calc})}$$

Once generated, the map can be manipulated by prior knowledge of the known solvent content of the crystal and the fact that proteins form globular shapes and so generally there is an area of high density, and a region of low density, and also that the density of related region in the electron density i.e. those related by non-crystallographic symmetry can be averaged. Both of

these methods aim to improve the interpretability of the initial electron density map where phases may not be accurate enough from initial calculations (due to errors in experimental phasing or non-isomorphism of model and crystal protein for molecular replacement). These methods, collectively called density modification, are used in many programs used in crystallography now, and have greatly improved the ability to solve otherwise uninterpretable electron density maps.

2.8.2 Solvent flattening

Solvent flattening looks to reduce the electron density in solvent cavities to a value that reflects the disordered solvent contents present in these regions. Initially calculated phases may yield noisy solvent regions, however in reality these regions must contain low solvent content, and so can be artificially flattened in order to improve the starting phases.

Solvent flattening is carried out by placing a 3D grid over the electron density, and determining whether the density within this region is that of protein or solvent. As such, the requirement of solvent flattening is at least a semi defined protein-solvent boundary. Spheres of varying radii are placed throughout the density and deemed either protein and left unchanged, or as solvent, and flattened to a constant 0.33 $e^-/Å^3$ for water, or slightly higher for salt containing crystallisation solutions. Once flattened, the electron density map can be Fourier transformed to generate structure factors with altered phases due to the modification of the electron density. These new phases are then combined with the old unmodified phases, and a new electron density map is calculated. Multiple cycles of solvent flattening then improves the overall phase of the structure factors, and produces an electron density map of sufficient quality for manual or automatic model building.

2.8.3 NCS averaging

A powerful method for improving starting phases prior to model building is non-crystallographic symmetry averaging, which increases in power based on the number of molecules in the asymmetric unit related by NCS. The technique is so powerful in fact that in cases of virus capsids, containing 60 NCS related particles (Chapman et al., 1998), phases can be calculated by randomly generated starting phases. The process of NCS averaging is more complex than

necessary for the scope of this thesis, and so the relevant literature is cited instead of discussed here (Rupp, 2009).

2.8.4 Model building and refinement

Once initial phases are determined and manipulated to generate an interpretable map, a model can be built, in order to describe the arrangement of atoms within the electron density, and to give the structure of the protein. Two maps are often used in this model building stage to minimise the biased generated from the calculated phases dominating the Fourier series (especially useful in molecular replacement). The first of these has coefficients |n|Fo|-d|Fc||, where n and d are calculated based on the quality of the current model, by the program SigmaA (Read, 1986). These coefficients are often close to n=2 and d=1, giving a map showing density for both modelled and unmodelled areas. The second map, with coefficients ||Fo|-|Fc|| is called the difference map in which, areas for which not enough scattering matter is placed, the amplitude will be positive and contain a positive peak. For areas of this map where too much scattering matter is placed, the amplitude will be negative and show as a negative peak in the map. The crystallographer can then use this to build new regions into the protein model. Amino acids are then fitted to these maps, sometimes via automatic building software such as ARP/wARP (Langer et al., 2008) and Bucanneer (Cowtan, 2006), and then refined in real space, such that the atoms fit well within the electron density maps.

Once a round of building has been performed, a reciprocal space refinement is carried out, and then the process of real space and reciprocal space refinement is repeated iteratively until the model is correctly built and the difference map no longer contains significant features. Reciprocal space refinement aims to minimise the difference between the observed amplitudes Fobs, and the calculated amplitudes F_{calc}, generated from the atomic positions in the currently built model. This refinement tries to minimise these differences for each reflection by changing atomic positions, atomic B-factors, atom occupancies, and a number of other properties depending on the data to parameter ratio i.e. more of these parameters can be refined at higher resolutions. At lower resolutions, some of these parameters are restrained, based on known values for bond lengths, bond angles, and grouping B-factors such that each residue has a grouped B-factor. B-factors can also be described spherically rather than by an ellipsoid, in

order to decrease the number of parameters required to describe the movement. The reciprocal space refinement is monitored by a residual index or R factor, given by

$$R = \frac{\sum_{\boldsymbol{h} \in free} ||F_{obs}| - |kF_{calc}||}{\sum_{\boldsymbol{h} \in free} |F_{obs}|}$$

which ranges from 0 for complete agreement between Fobs and F_{calc}, to 1 for complete disagreement, with proteins generally ranging from 0.2 for high resolution structures at the end of model building, to 0.4 for molecular replacement solutions of homologous proteins.

A second R factor is also used, in order to reduce bias in the refinement, where ~5 % of the reflections are not used in the refinement, giving the equation for R_{free} as:

$$R = \frac{\sum_{\boldsymbol{h} \notin free} ||F_{obs}| - |kF_{calc}||}{\sum_{\boldsymbol{h} \notin free} |F_{obs}|}$$

This refinement free R factor is called the R_{free}, and only when both the R factor and the Rfree decrease together during iterative model building is bias avoided. In cases of overfitting, the Rfactor may artificially become low, but the Rfree will remain higher, and so a divergence of R factor and Rfree will signify the overfitting. The Rfactor can also help in showing twinning or other crystal properties, in that a model that describes the electron density map fully, but leads to high R factors, can signify incorrect spacegroup assignment or twinning, or in most cases the former caused by the latter.

2.8.5 Model validation

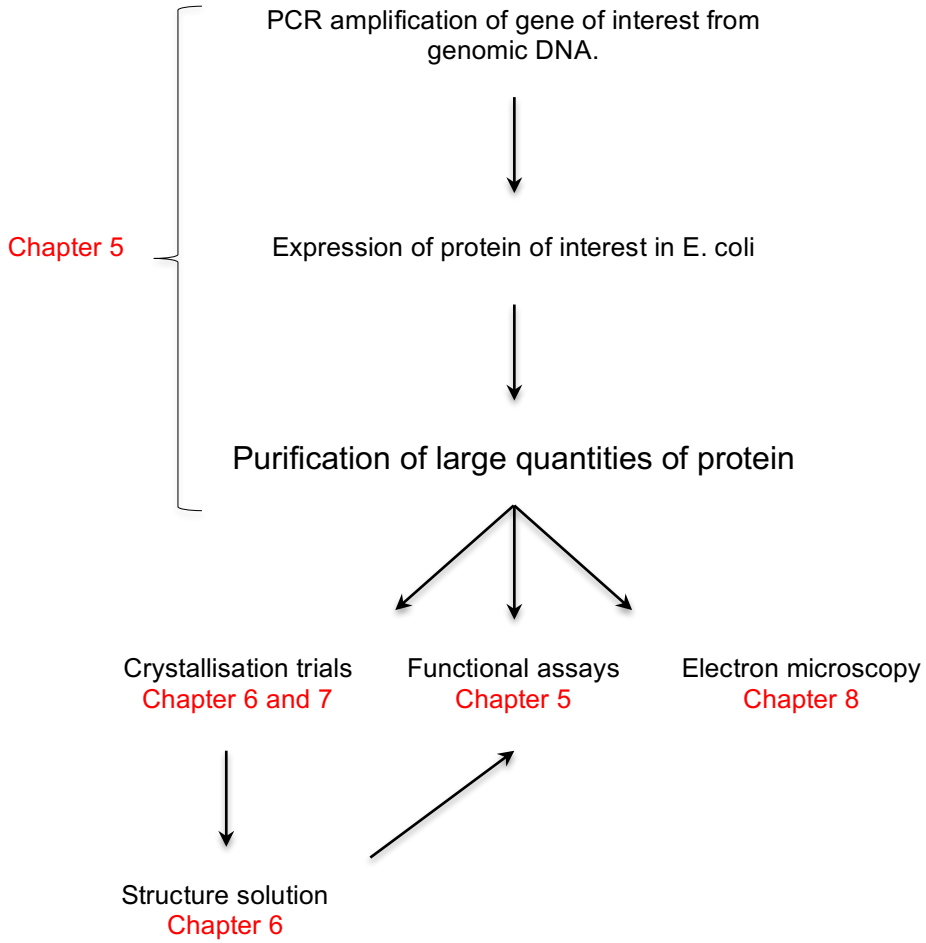
Finally, once all density features in the $||F_{obs}|-|F_{calc}||$ difference electron density map have been accounted for and the R factor and R_{free} are suitably low for the given resolution, the model can be validated in order to check that the model makes chemical sense i.e. main chain torsion angles lie with Ramachandran limits, bond lengths and bond angles are correct, and amino acid sidechain rotamers make biological sense. Programs such as Molprobity and PROCHECK carry out these analyses during the refinement and rebuilding process, and also score the protein structure on clashes between atoms, and quality of fit of the model to the electron density.

Chapter 3: Materials and Methods

This chapter explains and summarises the methodology used in the experiments described throughout this thesis. Specifically, this includes cloning, overexpression, and purification of a gene of interest, to generate sufficient quantities of protein for crystallisation and structure determination. In addition, experiments carried out to characterise each protein function are also summarised. Figure 3.1 shows the stages and workflow of these experiments, and this chapter is generally laid out in the order that the experiments appear in this thesis.

All equipment used was standard Molecular Biology lab equipment, and specific materials are identified and suppliers listed, as necessary.

Figure 3.1. schematic layout of experiments



3.1 General methods

3.1.1 Protein concentration determination

Protein concentration was determined in two ways, by Bradford assay (Bradford, 1976), and by absorbance at OD_{280} with an estimated molar extinction co-efficient determined using the Expasy ProtParam program (Gasteiger et al., 2005). Bradford assays consist of aliquoting 1 – 20 µl protein into a plastic cuvette, and adding 0.8 ml milliQ H_2O, followed by 0.2 ml BioRad Bradford reagent and mixing by inversion, before taking an absorbance reading at OD_{595}. The Bradford reagent binds to protein, and changes colour from red to blue, increasing absorbance at this wavelength. The protein concentration was determined using the following formula:

$$Protein\ concententration\ (mg/mL)\ =\ \frac{OD_{595\,nm}\quad X\quad 15}{volume\ protein\ (\mu L)}$$

Protein sample concentration was also calculated from the absorbance at OD_{280}, using extinction coefficients for each protein shown in table 3.1, calculated from the amino acid sequence using the online ProtParam software. Absorbance measurements at 280 nm were carried out on a Implen P300 spectrophotometer, using a Nanodrop to use minimal sample. The Nanodrop included a lid attachment, which effectively changed the path length by a lid factor or dilution amount i.e. 50x lid factor. The Beer-Lambert law was used to determine concentration in mg/ml, where multiplying by the lid factor gives a path length of 1.

$$Protein\ concententration\ \left(\frac{mg}{mL}\right) = \frac{OD_{280\,nm}\quad X\quad lid\ factor}{Abs\ 0.1\ \%\ \left(=1\frac{g}{L}\right)} = \frac{A\ X\ lid\ factor}{\varepsilon T}$$

where the absorbance A multiplied by the lid factor gives the absorbance when T = 1.

Table 3.1. Protein extinction coefficients

Protein	Molar extinction coefficient (M^{-1} cm^{-1})	Abs 0.1% (=1 g/l)
AhlA link C-terminal His	32890	0.80
AhlB C-terminal His	25440	0.66
AhlC C-terminal His	1490	0.05
AhlC HM C-terminal His	1490	0.05

3.1.2 SDS PAGE and Native PAGE electrophoresis

SDS PAGE gel electrophoresis was used to determine the levels of protein expression from overexpression trials, and to estimate purity during protein purification. SDS PAGE electrophoresis separates and resolves proteins based on molecular weight (Laemmli, 1970), using a protein standard marker as a reference. SDS PAGE gels were prepared manually between two glass plates, as detailed in Table 3.2, using stacking and resolving gels of 6 % and 12 % respectively. A plastic comb was used to create 15 wells at the top edge of the gel in which to load protein samples of ≤20 µl.

Table 3.2. SDS PAGE gel recipe

12 % Resolve gel	6 % Stacking gel
2.5 ml 30 % Acrylamide	0.75 ml 30 % Acrylamide
2.35 ml Tris pH 8.8	0.47 ml Tris pH 6.8
1.28 ml MilliQ water	2.46 ml MilliQ water
62.5 µl 10 % (w/v) SDS	37.5 µl 10 % (w/v) SDS
10 µl TEMED	5 µl TEMED
62.5 µl 10 % (w/v) APS	37.5 µl 10 % (w/v) APS

3.1.3 SDS PAGE sample preparation

Protein samples containing 15 – 20 µg protein (as determined by Bradford assay) were added to LDS buffer and reducing agent (Novex NuPage reducing agent), to give 1x of each. Samples were then incubated at 95 °C for 10 minutes. Samples greater than 20 µl were heated in Eppendorf tubes with the lid open to evaporate excess water from the sample to reduce the volume. Samples were centrifuged at 13000 rpm in a micro centrifuge for 5 s and loaded on to the SDS PAGE gel. A 60 V stacking stage was run for 10 minutes, which brings the samples to the interface between the stack and the resolve gel and ensures good resolution of protein bands on the gel. The voltage was then increased to 200 V for 45 minutes, or until the leading Coomassie front reached the bottom of the gel.

In order to observe the bands of protein on the SDS PAGE gel, the gel was stained in either Coomassie stain (0.1 % Coomassie blue, 1 volume methanol, 1 volume glacial acetic acid, 1 volume deionised water) for 20 minutes followed by a destaining solution (10 % (v/v) glacial acetic acid, 10 % (v/v) methanol) overnight, or in Instant Blue Stain (Expedeon) for 1 hour. Destained gels were stored in deionised water and dried between two acetate sheets for long-term storage.

3.1.4 Hemolysis assays

Haemolysis assays were carried out using Horse and Sheep blood, using methods described by Rowe and Welch (Rowe and Welch, 1994). Defibrinated Horse and Sheep blood was purchased from Thermo Scientific and stored at 4 °C for up to 4 weeks. Blood was washed and prepared for assays by diluting to 10 % w/v in 10 mM PBS pH 7.4 and centrifuged at 1000x g for 2 minutes. The supernatant was removed and the washing step was repeated three times. The final loose pellet of blood cells was diluted to 1 % w/v in 10 mM PBS pH 7.4. Each AHL component was buffer exchanged into 10 mM PBS pH 7.4 and the final concentration quantified by absorbance at 280 nm.

Haemolysis assays were set up in 1.5 ml Eppendorf tubes using 1 ml 1 % w/v blood and 2 – 10 µg each component, in different combinations described in Chapter 5. The protein was aliquoted to the wall of the epindorf tube and mixed with the blood by inversion once each protein was added. Final reactions were incubated for 1 – 2 hours at 37 °C on a blood wheel to mix throughout the experiment.

After incubation each experiment was centrifuged at 1500x g for 2 minutes at 4 °C to pellet the remaining erythrocytes. The supernatant was removed into a cuvette. Haemolysis was measured by determining the absorbance of the supernatant at 542 nm. A positive control consisted of 1 ml 1 % w/v erythrocytes pelleted at 1500x g for 2 minutes, and resuspended in 1 ml milliQ water followed by pipetting up and down to induce lysis by hyper-osmotic pressure. The positive control was incubated at 37 °C along with the experimental tubes. A negative control consisted of 1 ml 1 % w/v erythrocytes incubated at 37 °C without any protein added.

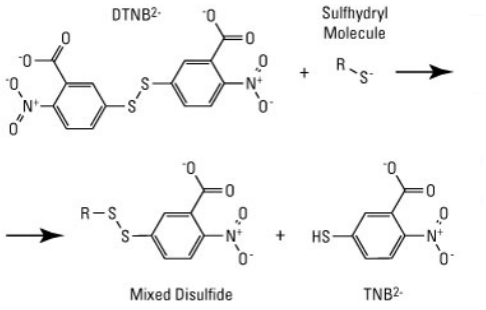
Final lysis was measured as a percentage of the total lysis in the positive control, and normalised against the lysis present in the negative control, using the following formula:

$$percentage\ lysis = \frac{reaction\ (A_{542nm}) - negative\ control\ (A_{542nm})}{positive\ control\ (A_{542nm})}\ x\ 100$$

3.1.5 Sulfhydryl DTNB assay

Sulfhydryl DTNB assays were carried out in order to determine the amount of reduced, available cysteine sulfhydryl groups present on a protein, using Ellmans reagent (Figure 3.2)(Ellman, 1959). Ellmans reagent binds to free cysteines to form a mixed disulphide bond between DTNB and the cysteine. The cysteine linked DTNB then absorbs at 412 nm, which can be followed by spectrophotometry.

Figure 3.2. DTNB sulhydryl reaction



The assay was performed by pipetting 0.9 ml of either native buffer A, containing 0.1 M sodium phosphate pH 7.3 and 1 mM EDTA, or protein denaturing buffer containing 0.1 M sodium phosphate pH 7.3, 1 mM EDTA, and 6 M GuHCl into a plastic cuvette. 0.1 ml of 10 mM DTNB in buffer A was added to the cuvette, and mixed. The sample was then used as a blank, taking an absorbance reading at 412 nm. 10 – 50 µl of 0.5 mM protein was added to the blanked sample, mixed by inversion, and then incubated for 2 minutes at room temperature. A second absorbance measurement was taken at 412 nm, to assess how much of the reagent had reacted with the protein. The extinction coefficient of Ellmans reagent is 0.0137, which was used to calculate the amount of reacted Ellmans reagent. The amount of free sulfhydryl groups in the denaturing buffer B showed the percentage of cysteines not forming disulphides bridges, whilst the amount of free sulfhydryl groups in the native buffer A showed how many cysteines were present on the surface of the protein. This assay was used in the context of this thesis to determine surface cysteines for heavy metal derivitisation.

3.1.6 Lipid stocks and preparation

Unilamellar lipid vesicles were generated, for use in circular dichroism and electron microscopy, from a number of different lipids. Total lipid extracts including *E. coli* and liver total lipid extracts (Avanti Polar Lipids) were used in electron microscopy, and the total lipid components of each extract are shown in Table 3.3. DHPC and DHPG lipids were used to generate bicelles with the detergent component CHAPSO. 100 mg of each total lipid extract was purchased as lyophilised solids, which were resuspended in 1 ml 2:1 chloroform: methanol solution and frozen at -20 °C until needed. DHPC and DHPG were stored as solids at -20 °C until required.

Table 3.3. Total lipid extract contents

Component (%)	*E. coli* total lipid extract	Liver total lipid extract	Brain total lipid extract
PE	57.5	22	33.1
PG	15.1	-	-
PC	-	42	12.6
PI	-	8	-
Lyso PI	-	1	4.1
PS	-	-	18.5
Cholesterol	-	7	-
PA	-	-	0.8
CA	9.8	-	-
Others	-	21	-
Unknown	17.6	-	30.9

3.1.7 Liposome preparation

Liposomes were made from frozen stocks of lipid extract by first removing the organic solvent under a nitrogen stream. A 100 µl 100 mg/ml aliquot of lipid suspension was placed in a round bottomed flask and the solvent was evaporated first under a gentle nitrogen stream, increasing the flow of nitrogen as the film formed. The flask was swirled to generate a thin film on the bottom of the flask, and evaporation was completed when the flask stopped feeling cold to the touch under nitrogen.

The bottom of the flask was submerged in liquid nitrogen to flash freeze the lipid film, and then the flask was placed on a vacuum pump at 1.5 mbar for 2 hours to remove all residual solvent.

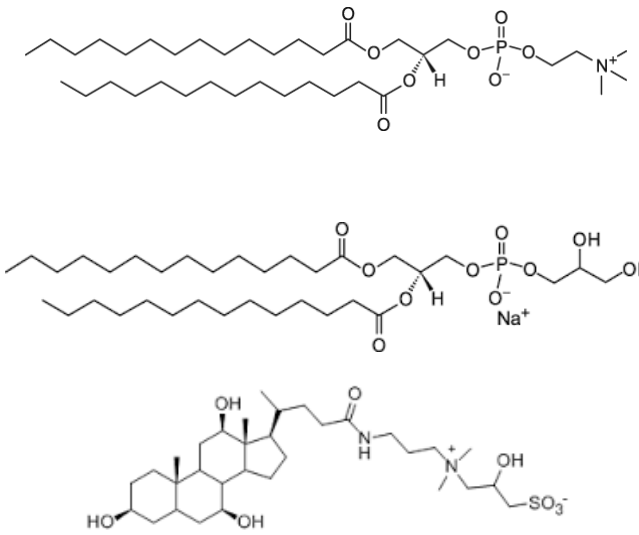
1 ml 10 mM PBS buffer pH 7.4 was added to the lipid film and the suspension vortexed for 10 minutes to resuspend the lipids fully. Once resuspended the lipids were extruded to form liposomes using a mini extruder (Avanti Polar Lipids). To extrude the lipids, the resuspended lipids were placed in a 1 ml Hamilton syringe, and attached to the mini extruder device, containing a 0.1um filter membrane. Lipids were syringed through the membrane to the second syringe attached at the opposite end, and then syringed back and forth >10 times to form a uniform solution of unilamellar liposomes. The final solution was always taken from the second syringe, as this syringe should not contain any residual pre-extruded lipid.

Extruded lipids were used in experiments immediately, or stored at 4 °C for up to one week. Liposome quality was ensured by electron microscopy, as old liposomes form aggregates and rupture easily, visible under the microscope.

3.1.8 Bicelle preparation

Lipid bicelles were used as a membrane model for crystallisation and electron microscopy studies, and were prepared from a combination of lipid and detergent (Figure 3.3). Zwitterionic phospholipid DHPC was used as the main lipid component, and supplemented with negatively charged DHPG to generate bicelles of different charges. The detergent component was CHAPSO.

Figure 3.3. Bicelle components



To produce bicelles, lipid and detergent was mixed at a 3:1 mass ratio lipid to detergent and suspended in milliQ water to produce a 20 % wt. mixture. The solution was heated to 40 °C for 1 minute and vortexed for 1 minute. The solution was then placed on ice until cold and vortexed for 2 minutes. This heating and cooling was repeated for between 40 – 70 minutes until the solution was a clear gel at 40 °C, and a clear liquid when cooled on ice.

The bicelle mixture was then frozen at -20 °C until required. Freeze thaw cycles do not affect the bicelle properties of the mixture, and so bicelles were not aliquoted into smaller volumes before freezing.

3.1.9 Circular dichroism

Circular dichroism experiments were carried out on a Jasco J810 CD spectrometer, scanning from 190 nm to 250 nm, with a scanning interval of 50 nm min^{-1}. In each case, the

sample was placed in a quartz cuvette with a path length of 0.1 mm, and 200 µl of the sample was placed in the cuvette for each reading. For each wavelength, an ellipticity and a high-tension measurement were taken, in units of millidegrees and Volts respectively. The number of readings taken is given where appropriate, and where more than a single measurement was taken; an average was used plotting the ellipticity. In each specific case the concentration of the protein used for CD measurements is also given, as well as any ligand added to the sample and incubation conditions.

The ellipticity measurements in millidegrees were converted to the more usual mean residue ellipticity ($[\theta]_{MR}$), making each measurement concentration and pathlength independent. The following formula was used to convert from ellipticity to mean residue ellipticity:

$$[\theta]_{MR} = \frac{100\ x\ \theta}{(C_{MR}\ x\ I)}$$

Where θ is the ellipticity in degrees, C_{MR} is the mean residue concentration calculated as the concentration in molar multiplied by the number of residues, and I is the path length in centimetres. A formula was also used to predict the mean residue ellipticity measurement at 222 nm for a completely α-helical protein ($^{max}[\theta]_{222}$), as follows:

$$^{max}[\theta]_{222} = -40{,}000\ x\ (1 - \frac{2.5}{n})$$

where n is the number of amino acid residues. The $^{max}[\theta]_{222}$ values for AhlB and AhlC were 39727 deg.cm^2.dmol^{-1} and 39636 deg.cm^2.dmol^{-1} respectively.

3.2 Cloning methods

3.2.1 PCR amplification of genes from gDNA

PCR was used to clone genes of interest from genomic DNA and insert restriction sites for cloning of the gene into *E. coli* expression vectors (Bartlett and Sterling, 2003). Biomix red 2 x master mix (Bioline) was used as a reaction mix from which amplification was achieved by PCR. PCR reactions were set up as follows for the first time PCR of new genes, and optimised changing the PCR temperatures and amount of genomic DNA based on initial results:

25 µl Biomix red master mix

200 nM Forward primer

200 nM Reverse primer

100 ng genomic DNA

The reaction volume was made up to 50 µl with autoclaved deionised water. PCR was run using a Techne 3prime thermo cycler using the following general program:

Table 3.4. Cloning PCR parameters

Stage	Temperature	Time	Number of cycles
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	10 sec	10
Annealing	X_1 °C	30 sec	
Extension	72 °C	1 min/ kb	
Denaturation	95 °C	10 sec	20
Annealing	X_2 °C	30 sec	
Extension	72 °C	1 min/ kb	
Final extension	72 °C	2 min/ kb	1
Final hold	10 °C	∞	-

The reaction was run in two distinct stages, denoted by the red and green shading in the above table, for 10 cycles and 20 cycles respectively. The difference between the two stages was the annealing temperature, which was calculated as the lowest melting temperature of the primers minus 5 °C. In the red stage, each primer melting temperature was calculated from only the complementary DNA sequence to that of the gene i.e. excluding the restriction sites etc. In the green stage the melting temperatures of each primer was calculated using the entire primer sequence. Using this method, potential large differences in melting temperatures are taken into account, which otherwise would not allow the primer to anneal.

3.2.2 Agarose gel electrophoresis

Agarose gels can be used to separate DNA based on its size by electrophoresis. The technique was used in this research to confirm successful cloning reactions, successful restriction digests of plasmids, and also after colony PCR to ensure than the correct insert was present in the cloned plasmid.

In all experiments 1 % (w/v) agarose gels were cast by suspending 0.4 g agarose in 40 ml TAE buffer (89 mM Tris pH 7.6, 89 mM boric acid, 2 mM EDTA). The agarose was solubilised in the TAE buffer by microwaving at medium power for 1 minute. Once fully dissolved, the TAE-agarose was allowed to cool briefly until warm to the touch. Once cooled, 4 µl 10,000x GelRed (Bioscience) was pipetted and mixed into the solution by swirling, and the gel was poured into a cast, using a ladder insert until the gel had solidified. Once solidified, the ladder insert was removed and the gel submerged in TAE buffer in a BioRad agarose gel electrophoresis container (BioRad).

Samples for electrophoretic analysis were prepared by mixing DNA with 4X DNA loading dye in a 3:1 ratio, mixing by gentle pipetting. Samples were loaded into each well of the gel, and 2 µl Hyperladder DNA ladder (Bioline) placed in lane one. Electrophoresis was carried out at 100 V for 38 minutes or until the leading edge of the loading dye reached the end of the gel. The gel was then removed and visualised on a UV trans-illuminator. Photos were taken of the gel, and then the gel was discarded or bands were cut from the gel for use in gel extraction.

3.2.3 Purifying PCR product DNA from the PCR reaction mix

Once PCR products were determined to be of the correct size, and to contain no contaminating bands, by electrophoresis, PCR mixtures were prepared for restriction digest by

purification using the Qiagen PCR clean-up kit (Qiagen). In the first stage of PCR clean-up, 5 volumes of PB buffer are added to the PCR reaction mixture. The resulting solution is passaged through a centrifugation filter column at 13,000 rpm for 2 minutes. 700 µl PE buffer are then washed through the column by centrifugation at 13,000 xg for 1 minute, and then 30 µl autoclaved milliQ water is incubated on the filter for 5 minutes at room temperature. The DNA is then eluted from the column in the water by centrifugation, and stored at -20 °C until needed.

3.2.4 Plasmid DNA purification

In order to generate microgram amounts of plasmid DNA for molecular cloning and protein overexpression, plasmid was amplified in high copy cloning strains of E. coli DH5α. A single colony containing the plasmid to be purified was selected and grown in 5 ml LB media at 37 °C overnight, and the resulting culture was pelleted by centrifugation for 15 minutes at 4500 xg. The pellet is then ready to purify the plasmid DNA using the Qiagen Miniprep kit (Qiagen). First the pellet is resuspended in 250 µl P1 buffer, and then cells are lysed by addition of 250 µl P2 buffer. 350 µl N3 buffer is added, quickly mixed by inversion, and then centrifuged at 13000 rpm in a microcentrofuge for 10 minutes. The result is a soluble suspension containing protein and plasmid DNA, with the insoluble cell debris and genomic DNA forming a loose white pellet. The soluble fraction is applied to a Miniprep centrifuge filter, which binds the DNA, and then the solution is passaged through the filter by centrifugation at 13,000 xg for 1 minute. 500 µl PB buffer is pipetted into the centrifuge filter, and then passaged through the filter by centrifugation at 13,000 xg for 1 minute. The process is repeated with 500 µl PE buffer, and then the filter is placed in an empty Epindorf, and centrifuged for 30 s at 13,000 xg to remove any remaining PE buffer, which is inhibitory to the next stage of DNA elution. Finally, the filter is placed in another Epindorf, 30 µl of autoclaved milliQ water is added, and then incubated for 5 minutes at room temperature. The DNA is eluted by centrifugation for 2 minutes at 13,000 xg, and the DNA content of the eluted solution is analysed by absorbance at 260 nm, and then stored at -20 °C until required.

3.2.5 Purification of PCR product from agarose gel

In some cases, PCR product DNA contains contaminants, which cannot be purified by column methods, and so instead the DNA is separated by agarose gel electrophoresis, and then purified by extraction of the correct band from the gel. To begin, all remaining PCR

reaction after analysis was loaded and run on an agarose gel . Once run, the band was cut from the gel using a scalpel, minimising illumination of the band under a trans-illuminator. The agarose gel was suspended in five times volume of QC buffer available in the Qiagen gel extraction kit, and then incubated at 50 °C for 10 minutes in order to dissolve the gel. Once dissolved, the solution was passaged through a centrifuge filter, identical to that used in 3.2.4, by centrifugation at 13,000 xg for 1 minute. The column was washed with 500 µl PE buffer, repeating the second centrifugation step to remove residual PE, and then 30 µl autoclaved milliQ water was applied to the column and incubated for 5 minutes at room temperature. Finally the DNA was eluted from the column by centrifugation at 13,000 xg for 2 minutes, and then stored at -20 °C or used immediately for downstream applications.

3.2.6 Restriction digestion of DNA fragments

In order to insert PCR gene products into commercially available plasmid DNA, sites are engineered into a polyclonal site that can be cut by restriction enzymes producing compatible ends that can be ligated together with PCR products. All restriction enzymes and buffers used in experiments were purchased from NEB. Restriction digests were carried out either as double digest reactions, or as sequential single restriction digests. Double digests were used where the enzymes were compatible, but in the case of XhoI, the restriction efficiency is greatly reduced on supercoiled DNA, and so NdeI restriction was carried out first followed by XhoI. Double digests were carried out with the following reaction mixes and then incubated for 2 hours at 37 °C.

1 µl of each restriction enzyme
30 µl plasmid DNA or PCR product
5 µl CutSmart 10x buffer
13 µl autoclaved milliQ water

Single digests were also carried out, using the same above mixture, minus one of the restriction enzymes. The reaction mix was incubated for 1 hour at 37 °C, and then heat inactivated for 10 minutes at 80 °C. Once inactivated, 1 µl of the second restriction enzyme was

added, and incubated for a further 1 hour. The resulting mix of either single or double digest were purified by PCR cleanup.

3.2.7 Phosphatase treatment of plasmid DNA

Antarctic Phosphatase enzyme and buffer were purchased from NEB. Phosphatase treatment of plasmid DNA was used in order to remove 5' terminal phosphates, required in the ligation reaction. As such, the resulting plasmid is unable to ligate to the polyclonal site present in the plasmid mixture, and only with the PCR product DNA from gene PCR. This was used to increase the cloning efficiency in NHE experiments.

Phosphatase treatment of plasmid DNA was carried out by incubating the following mixture at 37 °C for 30 minutes, followed by heat inactivation at 80 °C for 10 minutes.

17 µl restriction digested and purified plasmid DNA

2 µl Antarctic Phosphatase reaction buffer 10X

1 µl (5 U) Antarctic Phosphatase enzyme

3.2.8 Ligation of DNA fragments to plasmid DNA

Plasmid DNA and DNA inserts treated with complimentary restriction enzymes can be joined together to yield a continuous circularised plasmid, containing the insert DNA in the expression region of the plasmid. Ligations were carried out using a 4:1 molar ratio of insert to plasmid DNA calculated roughly as the amount of DNA in ng μl^{-1} divided by the length of the DNA in base pairs. In general DNA fragments used in these experiments were four times smaller in length than plasmid DNA (1200 bp vs. 5-6 kbp), and so DNA amount in ng μl^{-1} could be used directly to calculate ligation mix components. Ligation mixes contained the following

4 µl 4Xng insert DNA

4 µl 1Xng plasmid DNA

2 µl 10X T4 ligase

2 µl autoclaved deionised water

And were ligated at room temperature for 2 hours. Ligation mixes were then used directly in transformation of chemically competent *E. coli.*

3.2.9 Colony PCR

Colony PCR was carried out either from a colony grown on an LB agar plate, or from plasmid miniprepped from a colony. In the case where a colony was used, the colony was dissolved in 10 µl autoclaved milliQ water and boiled for 10 minutes at 95 °C, and then centrifuged for 2 minutes at 13,000 rpm to pellet the insoluble matter. PCR was then carried out as described in 3.2.1, where the genomic DNA was substituted for the boiled soluble colony DNA. In the case where miniprepped DNA was used, a single colony was grown in 5 ml LB media overnight at 37 °C, and the resulting colony miniprepped as described in 3.2.4.

In both cases the primers used for colony PCR were T7 promoter and T7 terminator primers, available from Invitrogen.

3.2.10 Site directed mutagenesis

Site directed mutagenesis was carried out using the Q5 mutagenesis kit purchased from NEB. Primers were designed using the NEBaseChanger software, with the insertion method.

SDM reaction mixes contained the following components

12.5 µl Q5 Hot start High-Fidelity 2X master mix

1.25 µl 10 µM Forward primer

1.25 µl 10 µM Reverse primer

1 µl 25 ng/µl template DNA

9 µl autoclaved MilliQ water

and were run in the PCR thermocycler under the following conditions

Table 3.5. Q5 Mutagenesis PCR parameters

Stage	Temperature	Time	Number of cycles
Initial denaturation	98 °C	30 sec	1
Denaturation	98 °C	10 sec	20
Annealing	61 °C	30 sec	
Extension	72 °C	30 sec/ kb	
Final extension	72 °C	2 min	1
Final hold	10 °C	∞	-

Once the PCR reaction was complete, the template DNA was digested in a KLD reaction, with the following components

1 µl PCR product

5 µl 2X KLD reaction buffer

1 µl 10X KLD enzyme mix

3 µl autoclaved MilliQ

which were incubated for 5 minutes at room temperature, and then transformed into 50 µl chemically competent DH5α as described in 3.3.2.

3.3 Microbiological methods

3.3.1 Growth media and conditions – LB media and rich 4xYT media

Luria-Bertani media was prepared by weighing out 10 g/l tryptone, 10 g/l NaCl, and 5 g/l yeast extract and diluting in 800 ml/l deionised water. Once dissolved, the solution was made up to 1 l/l and transferred into appropriate containers. LB media was then immediately autoclaved and then stored at room temperature until needed.

LB agar was produced by adding 15 g/l bacterial agar to LB media in a Duran bottle. Agar was partially dissolved by vortexing and the mixture autoclaved immediately, whereupon the rest of the agar was dissolved. The LB-agar was stored at room temperature until needed. For use, LB agar was melted by microwaving on a low heat and 25 ml LB-agar was poured into petri dishes to produce a single plate. The plates were supplemented with 100 µg/ml carbenecillin or

50 µg/ml kanamycin once cool enough so that the antibiotic was not denatured. Plates were used immediately once set, or sealed with Parafilm and stored at 4 °C for a maximum of 2 weeks, or until needed.

4xYT media was used for growth of 50 ml *E. coli* cultures for test overexpression, as well as for the production overexpression runs (500 ml). The media was prepared by adding 32 g tryptone, 20 g yeast extract, and 5 g NaCl to 800 ml deionised water. Once dissolved, the media was made up to 1 l and immediately autoclaved, then stored at room temperature until needed.

3.3.2 Transformation of *E. coli* with plasmid vector

In order to amplify DNA and overexpress proteins, vectors were inserted into either cloning strains or overexpression strains of *E. coli*. These strains of *E. coli* were purchased as chemically competent cells, stored at -80 °C in glycerol. Cells were thawed on ice before addition of 50 – 100 ng plasmid, mixed gently and incubated for 30 minutes on ice. Cells were then heat shocked by incubating at 42 °C for 25 s and then incubating on ice for 2 minutes. Transformants were incubated at 37 °C with shaking at 220 rpm for 1 hour to recover. Finally the recovered cells were plated onto an LB-agar plate supplemented with a selection antibiotic depending on the transformed plasmid and incubated at 37 °C for 16 hour.

3.3.3 Glycerol stocks

For short-term storage, *E. coli* transformants containing useful plasmids were stored on an LB-agar plates for a few weeks, before becoming non-viable. For longer storage of plasmid in bacteria, cells were stored in glycerol at -80 °C. These glycerol stocks were generated by centrifuging a 5 ml overnight culture at 4000 xg for 10 minutes at 4 °C. The supernatant was removed and the pellet resuspended in 0.6 ml fresh LB media and 0.4 ml glycerol, yielding a 40 % (v/v) glycerol stock, which was stored at -80 °C until required.

3.4 Protein expression and purification methods

3.4.1 Protein overexpression in *E. coli*

E. coli can be exploited to produce large quantities of useful protein by incorporation of a commercially available overexpression vector. To find conditions in which large amounts of soluble protein can be expressed, small-scale expression trials were carried out in which 50 ml cultures were used to overexpress protein by induction. Once conditions for maximum expression had been determined, a scaled version using 500 ml cultures was used to produce

large amounts of protein for structural studies. The specific conditions for each construct are given in the appropriate chapters, but an overview of the method is given here.

3.4.2 pET vector overexpression

A 2 % (v/v) inoculation of 50 ml overnight culture was used to inoculate 500 ml liquid media supplemented with a working concentration of the selection antibiotic (50 – 100 µg/ml). The 500 ml culture was grown at 37 °C with shaking until the absorbance at 600 nm measured 0.6 – 0.8. The culture was then inoculated with 1 mM IPTG and grown for either 4 hours at 37 °C or overnight at 16 °C or 25 °C.

Cells were then harvested by centrifugation at 20,000x g for 20 minutes at 4 °C to pellet the cells. The cells were resuspended in a small volume of LB and centrifuged in a 50 ml Falcon tube, resulting pellet was frozen at -80 °C until required.

3.4.3 Minimal media SeMet overexpression

Minimal media was used in overexpression cultures to incorporate selenomethionine into proteins of interest. Minimal media was made by adding 10.5 g/l K_2HPO_4, 1 g/l $(NH_4)SO_4$, 4.5 g/l KH_2PO_4, 0.5 g/l Tri-sodium citrate, 5 g/l glycerol, and 0.5 g/l of each nucleobase Adenine, Guanosine, Thymine, and Uracil to deionised water, and then autoclaving. The nucleobases did not dissolve until autoclaved, and were checked after autoclaving to ensure that they had dissolved. A separate mixture of amino acids was made by adding 1g/l $MgSO_4 \cdot 7H_2O$, 4 mg/l Thiamine, 100 mg/ml of each of L-lysine, L-phenylalanine, L-threonine, and 50 mg/ml of each of L-isoleucine, L-leucine, and L-valine to deionised water, and then stored at -20 °C until required. The mixture was dissolved in 20 ml at 50x concentration, so that each 1 l overexpression flask received 20 ml of amino acid mixture. The amino acid mixture was added to the autoclaved minimal media prior to overexpression, and 40mg/l of seleno-L-methionine was added and dissolved, to make the completed minimal media for overexpressions.

Overexpression cultures were started and grown identically to regular overexpressions in LB media until an optical density at 600 nm reached 0.8. The culture was then centrifuged at 5000 xg at 20 °C for 15 minutes to pellet the E. coli cells by a minimally harsh procedure. The cells were washed twice in a small amount of minimal media (~50 ml per 500 ml culture) and centrifuged to re-pellet the cells. Finally the cells were resuspended in 500 ml minimal media and grown at 37 °C for one hour. The optical density at 600 nm was measured before and after

the growth period, to monitor the recovery of the cells, and so long as the OD had not dropped considerably i.e. to zero, the cells were induced with 1 mM IPTG, and grown at the usual overexpression temperature determined by overexpression trials. Minimal media overexpressions were grown twice as long as LB media overexpressions, and so 4 hour cultures were grown overnight, and overnight cultures were grown for two days. Cells were then harvested by centrifugation at 8500 xg for 15 minutes at 4 °C, and the pellets were stored at -80 °C until required.

3.4.4 Cell free extract preparation

To recover the expressed protein of interest, the cells were broken using sonication, and then separated into cell debris and cell free extract by centrifugation.

The cell pellet, stored at -80 °C, which aids in the breaking step, was resuspended in 10 ml/g lysis buffer (50 mM tris pH 8, 0.2 M NaCl). The suspension was separated into smaller containers and kept on ice for sonication. Each sample was sonicated at 15 μm for 2 x 20 s for large samples and 2 x 5 s for smaller volumes, using the largest probe possible for the container. The process was carried out in a rotation, to allow cooling between each pulse. The resulting lysate was centrifuged for 10 minutes at 40000 xg to pellet the cell debris and any insoluble proteins. The cell free extract was then recovered for purification.

3.5 Protein purification

Two main methods were used during experiments to purify each protein, as each was expressed with a poly 6x Histag. The two purification techniques include immobilised metal ion affinity chromatography (IMAC) exploiting the Histag present on each protein, and size exclusion based gel filtration which was used to purify the protein further and also to analyse the size properties of the samples.

3.5.1 Nickel affinity chromatography

Nickel affinity chromatography was carried out as the initial stage in purification of the AHL proteins. The cell free extract containing AHL proteins, in 50 mM tris pH 8 and 0.2 M NaCl, were applied to a 5 ml Histrap column and then the column was washed in buffer A (50 mM tris pH 8 and 0.2 NaCl) until the UV trace returned to the baseline. A 0 – 100 % gradient of buffer B (50

mM tris pH 8, 0.2 M NaCl, and 0.5 M imidazole) was applied over 50 ml to elute the protein from the column. 2.5 ml fractions were collected during the chromatography, and were analysed by SDS PAGE electrophoresis. Fractions determined to contain the protein to be purified were pooled and concentrated for gel filtration or crystallisation preparation.

3.5.2 Gel filtration chromatography

Gel filtration is a method of size exclusion, which utilises highly cross-linked dextran to create a matrix able to separate proteins based on size. Proteins smaller than the pore size can enter the pores, and so elute later in the chromatography. Larger proteins travel through a smaller column volume, as they are excluded from the pores, and so elute from the column sooner. Each column is calibrated using known protein standards to give a spectrum of elution volumes based on size and shape of the protein molecule.

Gel filtration can be used to purify proteins, separating the protein of interest from larger and smaller protein contaminants, or analytically, to determine the oligomeric state of proteins in a number of buffer conditions.

3.5.3 Preparation of proteins for crystallisation

Protein from nickel affinity chromatography was generally desalted into a buffer containing no imidazole (50 mM tris pH 8 and 0.2 M NaCl) using a Zeba spin column (Thermo Fisher) prior to crystallisation. The Zeba column was washed three times in 300 µl of the desired buffer by centrifugation at 1500 xg for 1 minute. Up to 130 µl protein was then applied to the Zeba column, and centrifuged at 1500 xg for 2 minutes, to recover the protein in the new buffer.

All protein concentrations were calculated prior to crystallisation, and either concentrated, by centrifugation at 4500 xg in a concentrator until the desired concentration was reached, or diluted in buffer if the concentration was too high. Initial crystallisations were carried out at 15 mg/ml, and then the concentration was varied depending on the result.

Proteins were stored in high salt buffer at 4 °C for up to 4 weeks, and small amounts were desalted as required, in order to preserve the protein for longer durations.

3.6 Crystallographic methods

3.6.1 Initial robot screening

Initial screening for crystallisation conditions by vapour diffusion were carried out on the Matrix Hydra II PlusOne crystallization robot in a 96 well format. Each plate contained 96 isolated experiments consisting of a large well for mother liquor and two small wells, one of which was used in each experiment to contain a 1:1 ratio mix of mother liquor and protein. 200 µl of each crystallisation solution was aspirated into the large well, and then 200 nl of this dispensed in the small well. 200 nl protein was added to the small well also, and then the plate was sealed with transparent crystal clear tape. Plates were centrifuged at 2000 rpm for 2 minutes to ensure proper mixing of the small well drops, and stored at 17 °C or 7 °C. During incubation, vapour-water equilibration between the wells could occur, leading to formation of protein crystals. Crystal growth was checked by observing each drop under a microscope. Crystallisation trials were carried out using the commercially available crystallisation screens PACT, JCSG+, MPD, pH clear, PEGs, AmSO4, Proplex, and Memgold 1 and 2 screens (Molecular Dimensions).

3.6.2 Crystal optimisation

Crystals grown in crystallisation screening were further optimised in order to increase crystal quality and size. Optimisations were carried out in hanging drop format using 24-well plates and siliconised coverslips.

3.6.3 Siliconising coverslips

Coverslips used in hanging drop optimisations were first siliconised to produce a hydrophobic surface for the drops to sit on. 5 ml siliconising solution (Dimethyldichlorosilane solution) was placed in a small beaker and ~100 round coverslips were placed on a large glass petri dish above the beaker. This setup was contained in a large bell jar, of which the lid was placed on and the air intake tap opened. A tube connected to a water pump was attached to the air intake valve and switched on. A vacuum was formed over 5 minutes, which was then maintained by closing the air intake valve before turning off the water pump. The vacuum was maintained for 30 minutes to allow the siliconising solution to vaporise and adhere to the glass coverslips. Once complete, the air intake tap was slowly opened, and the coverslips removed and stored in a petri dish.

Before use, siliconised coverslips were polished using lens tissue to buff the surface of the coverslip, and using an air stream to remove any debris from the surface of the glass. The coverslips were generally polished just before use, but sometimes were also polished in large quantities and stored in a petri dish for up to 4 weeks.

3.6.4 Hanging drop optimisation

Conditions that produce protein crystals were optimised to yield crystals of better diffraction quality, by varying buffer pH, salt concentration and precipitant concentration. A 24 well hanging drop plate was used in each experiment, to also scale up the crystallisation conditions, and to yield larger crystals. For each optimisation, 1 ml of crystallisation solution was made up in the large well. A polished siliconised coverslip was then set up, with 1 µl of the large well solution, and 1 µl of the protein, and mixed by gently pipetting up and down and stirring. The coverslip was inverted and sealed above the large well using high vacuum grease. Hanging drop plates were then incubated at 17 °C or 7°C to allow water-vapour equilibration and crystallisation to occur. Trays were checked daily for crystal growth for one week then weekly thereafter.

3.6.5 Cryoprotection solutions

Crystallisation conditions contain high volumes of water, and generally protein crystals are constituted of ~50 % solvent content. In cooling to 100 K to reduce radiation damage of data collection, water in the crystal freezes, and expansion upon entering the solid ice phase damages the crystal. To negate these effects, a cryoprotection solution in which an amount of the water was replaced with a cryosolution was used as an intermediate to remove some of the water from the crystal. The crystallisation condition was made up in a 1 ml Eppendorf tube, but some of the water was replaced with 10 – 30 % cryosolution. Generally, ethylene glycol, the monomeric form of PEG, found in many crystallisation conditions was used. In conditions not containing PEGs, 10 – 30 % glycerol was used instead.

Cryosolutions were screened for their ability to negate water ice formation by exposure to X-rays, as effective solutions reduce/remove the water ice diffraction present at ~3.5 Å. As removal of water from a crystal can also affect its diffraction quality, the lowest percentage cryosolution that worked was used (usually 15 % cryoprotectant).

3.6.6 Mounting crystals

For data collection, crystals were removed from the drop in which they formed, and mounted into loops. Loops are available in a variety of sizes, and it is preferred that the size of the loop matches the size of the crystal as closely as possible. Under a microscope, the drop was exposed by removing the clear tape covering the well, and a small drop of Cryoprotection solution was placed in the small well below the well containing crystals. The loop, attached to a wand by the magnetic base, was used to loop a single crystal, extracting the crystal from the drop in a side-on manner to reduce surface tension. The crystal was then placed briefly into the cryoprotectant solution before extracting again and placing into liquid nitrogen to cool to 100 K. Crystals grown in conditions that act as cryoprotectants were looped straight from the drop they were grown in into liquid nitrogen. The cryoprotection drop was then blotted to remove as much of it as possible, and the condition resealed with more clear crystal clear tape, and placed back at incubation temperature.

The looped crystal was then mounted on a goniostat using the base magnet, and aligned in the beam for suitable data collection.

3.6.7 Data collection

Data was collected both in-house on the CuKα source and at Diamond using synchrotron radiation. The in-house MAR345 setup was used to screen for crystals of good diffraction quality where size permitted, and to determine spacegroup and cell dimensions where possible.

Crystals of good diffraction quality were then shipped to Diamond in liquid nitrogen for full datasets, where the X-ray intensity is much higher. When larger quantities of crystals were being produced, and suitable cryosolutions known, crystals were also sent straight to Diamond without in-house screening.

3.6.8 Data processing

Data used in this thesis were processed automatically at the Diamond synchrotron using the three Xia2 pipelines and one FastDP pipeline, and each program used by each pipeline for indexing, integrating, scaling, and merging the data are highlighted in Table 3.6. In each case in the thesis, the data processing pipeline used is given at the appropriate time.

Table 3.7. Xia2 pipeline programs used

Pipeline	Indexing and integration	Scaling and merging
fastdp	XDS (Kabsch, 2010)	XDS, Pointless (Evans and Murshudov, 2013), SCALA
2d	Mosflm (Battye et al., 2011)	SCALA
3d	XDS	XSCALE (Kabsch, 2010)
3dii	XDS (indexing with peaks from all images)	XSCALE

3.6.9 Model rebuilding and model refinement

Once a starting map and model were determined by experimental means or molecular replacement, iterative model building and refinement was carried out using COOT and REFMAC respectively (Emsley et al., 2010; Murshudov et al., 1997). Any other programs used in the model building process are given for each specific example where appropriate throughout this thesis.

3.6.10 Structure validation

Once all interpretable features were accounted for in the electron density map and the difference Fourier map, and once the refinement had converged, the model building was deemed to be finished in terms of rebuilding. Once finished, X-ray structures were validated to ensure correct bonding angles, bond lengths, Ramachandran angles, and valid rotamers. The Molprobity server (Chen et al., 2010) was used to validate structures, uploading the model co-ordinates to the server, to generate a validation report. Hydrogen atoms were added to the model for this process, and where there is clear evidence, sidechains were flipped to satisfy correct hydrogen bonding. Models were then refined in COOT and REFMAC in order to improve the Molprobity score, and improve the model compared with other comparable models in the protein database.

3.7 Electron microscopy methods

3.7.1 Protein liposome preparation

Liposomes were prepared as described in 3.1.7. Proteoliposomes were generated by incubating an amount of protein described in each specific case with 20 ug prepared liposomes in 100 µl, at 37 °C for 1 hour. The sample was then incubated on ice and used in preparing carbon grids immediately.

3.7.2 Glow discharging and using carbon grids

Svet Tsokov prepared carbon grids by floating a thin carbon film onto copper grids. Grids were stored at room temperature in a petri dish for up to 4 weeks. Each grid was glow discharged before sample application to promote adsorption of the sample to the grid.
Up to 5 grids were placed on a parafilmed microscope slide and placed into glow discharge chamber of a Cressington 208 Carbon Coater. A vacuum was established within the chamber, and a small amount of air was allowed back into the chamber to promote plasma formation during discharge. Once the vacuum was stable, the timer for glow discharging was set based on the age of the carbon grid. Once finished grids were used immediately (within 30 minutes) and kept out of direct sunlight.

On a piece of parafilm a column of 50 µl drops was dispensed, 2 drops distilled water, and 2 drops 1 % (w/v) uranyl formate. Grids were clamped in a pair of tweezers and 1 – 5 µl sample was applied directly to the carbon side of the grid. Sample was allowed to adsorb for 1 minute. Excess liquid was blotted away immediately and the grid placed in the first drop of distilled water. The grid was again blotted and placed in the second drop of distilled water. After blotting again, the grid was placed in the first drop of uranyl formate and blotted. Finally, the grid was placed in the second uranyl formate for 20 seconds, and then dried using a air vacuum until all liquid was removed. Dried carbon films were placed in a grid box and stored for up to 5 weeks.

3.7.3 Data collection on a CM100 transmission electron microscope

Electron microscopy data was collected on a Phillips CM100 with a Tungsten filament set to HT 100 kV, with an attached Gatan MultiScan 794, 1K x 1K CCD camera, and using Gatan Digital Software. A gain reference image was taken at the beginning of data collection once the CCD cooling had finished to determine and minimise background noise. The objective aperture was set to 1, the spot size of the electron beam was set to 2.

The samples were first observed with a ‘search’ setting, in which images were collected with a 0.1s exposure, image binning of 2, and using the full CCD. A second focus setting was used to adjust the focus of the area once an area of interest had been found on the grid in search mode. The focus setting was set to a 0.4 second exposure, an image binning of 1, and using the central half of the CCD. Finally, images were captured in a ‘record’ mode, with a 1 second exposure, image binning of 1, and using the full CCD.

In focus mode, images were focused such that the first ring present in the Fourier transform of the image was to the edge of the Fourier transformed image. The electron beam itself was set to under focus, to increase the contrast of the image, and then secondary focusing was achieved as described.

Images were processed in ImageJ to determine profile plots for use in estimating the pore sizes, with the interactive 3D surface plot plugin for generating temperature plots (program created by Kai Uwe Barthel).

Chapter 4: NHE experiments and identification of new tripartite pore forming toxin family members

At the start of this project, the number of members of the tripartite family of pore forming toxins with homology to ClyA was very limited. The only known members were NHE and HBL, both from pathogenic strains of Bacillus species (Lund and Granum, 1997). It is unlikely that these tripartite toxins exist in isolation, due to their distant homology with ClyA. Having a limited number of examples of this toxin family available restricts the ability to study these toxins, from both a structural and biological standpoint. Although the structures of NheA and HblB have been solved by X-ray crystallography (Ganash et al., 2013), in isolation they cannot provide deep insights into the mechanism of the tripartite toxin system. In both cases, the NheA and Hbl B structures have been compared instead to the better-understood ClyA structures.

This chapter outlines the experiments undertaken on NheB and NheC to attempt to improve protein expression for structural studies. Difficulties in expression of these other components of the NHE toxin, NheB and NheC continued from previous work carried out in the Sheffield Laboratory (Ganash, 2012). The second part of this chapter describes Bioinformatic studies on this protein family, and the identification of novel tripartite toxin systems that share homology to NHE and may aid in interpreting the current toxin structures. Mapping of the conserved residues onto known tripartite toxins are explored, and finally a new tripartite toxin target is chosen for structural studies.

Firstly, the domain features of ClyA and the NHE proteins are shown diagrammatically, to show the features present in each protein, which will then be used in discussion throughout the thesis (Figure 4.1)

4.1 Preliminary experimental results for studies carried out on NHE

4.1.1 NheB cloning results

Attempts were made to overexpress NheB, in order to repeat the successful crystallisation trials that had been previously documented (Ganash, 2012). In these studies, NheB was expressed in *B. subtilis* strain JH642, although repeated overexpression trials yielded low expression levels. Indeed, only once was NheB successfully expressed for crystallisation.

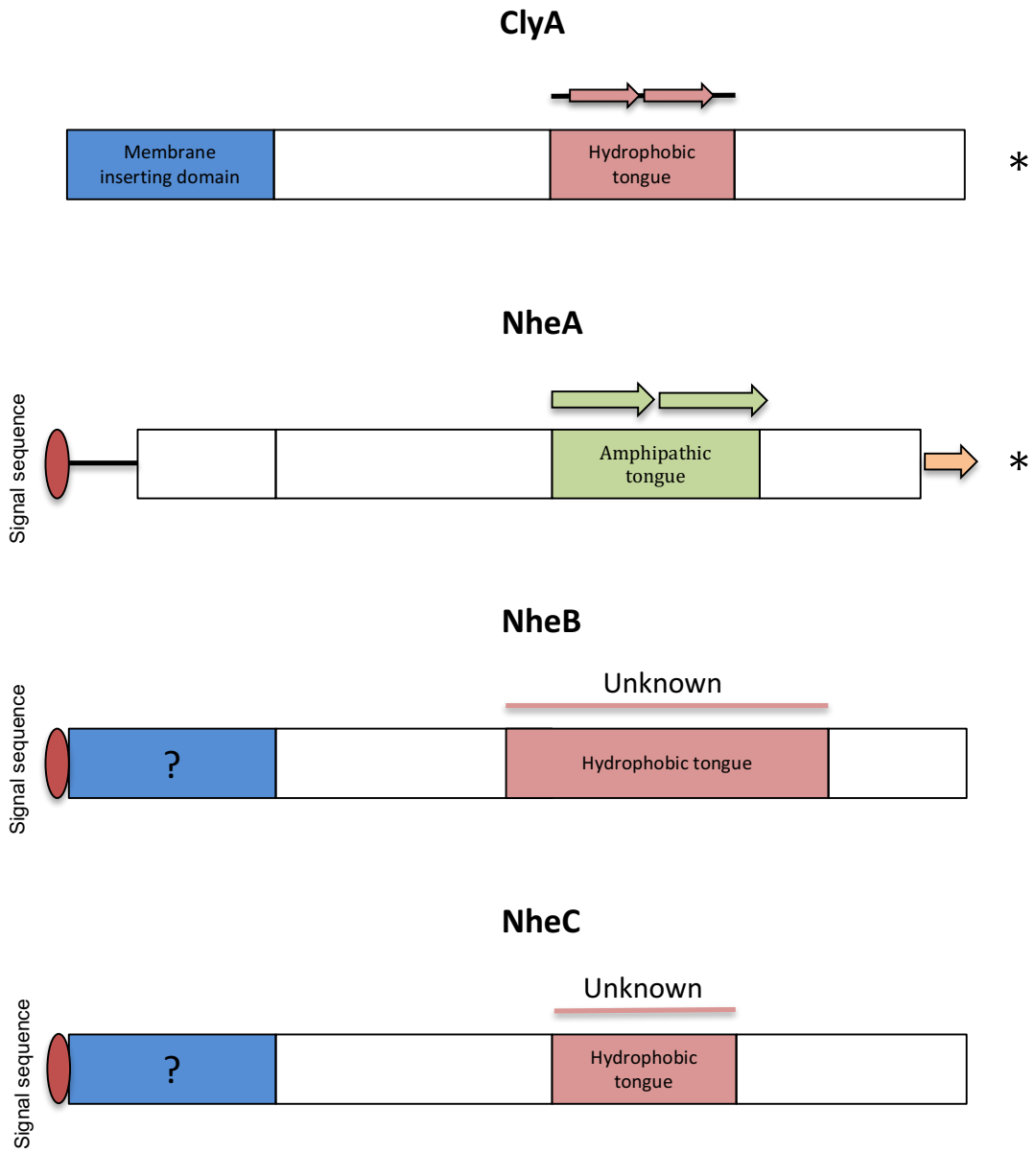


Figure 4.1. The predicted or observed pore forming domains found in ClyA and the Nhe proteins. The structure of ClyA in the soluble and pore form revealed that the protein contained an N-terminal membrane inserting helix (blue), and a hydrophobic β-tongue that inserts into a single leaflet of the membrane. The structure of NheA shows that the N-terminus is partially disordered, and that the amphipathic tongue region (green) does form a β-tongue as in ClyA. The longer tongue is held in place by a C-terminal β-strand latch (orange). NheB contains a hydrophobic stretch twice as long as ClyA (pink), and it is unknown whether this forms a β-tongue. NheC contains a hydrophobic stretch the same length as ClyA, and again the secondary structure is unknown. In both NheB and NheC, the presence of a potential N-terminal membrane inserting helix has not been confirmed (blue). The proteins for which structures are known are denoted by *****.

Expression of NheB in *E. coli* was of importance, as methodologies for growing *B. subtilis* in minimal media for selenomethionine incorporation are not developed, and so attempts were made to express NheB in *E. coli*.

When full length NheB was expressed in *B. subtilis*, the protein was exported to the media (Ganash, 2012), and the signal sequence was removed, but no mass spectrometry was carried out on the protein to determine the correct start point of NheB. A number of NheB constructs were generated, as the predicted signal sequence of NheB was vague, and was predicted to lie between $_{AES}35$, $_{HAV}43$, and $_{VAK}45$ in the N-terminus of the protein (Figure 4.2). NheB beginning with AES was used in these experiments, and was cloned into pET28a for expression in *E. coli* (Figure 4.2).

PCR cloning was carried out using pDG148-NheB *B. subtilus* expression vector provided by Magdah Ganash, in order to produce constructs in which the NheB gene began at $_{AES}35$, $_{HAV}43$, and $_{VAK}45$. PCR was carried out as described in 3.2.1 with 5 cycles at the 'pre-' temperature of 52 °C and 25 cycles at the 'post-' temperature of 58 °C. The presence of the correct DNA bands was confirmed by agarose gel electrophoresis (Figure 4.2), and then each PCR reaction was purified by PCR cleanup. Both pET28a and the purified PCR products were restriction digested for 2 hours with NdeI and XhoI, as described in 3.2.6. The digested pET28a was purified by gel extraction, and the digested PCR products were purified by PCR cleanup. The digested pET28a vector was further phosphatase treated, as described in 3.2.7, using 30 µl pET28a plasmid from the gel extraction. Once treated, the phosphatase treated plasmid was purified by PCR cleanup. The $_{AES}35$, $_{HAV}43$, and $_{VAK}45$ digested inserts were each ligated into the digested pET28a, and then used directly in transformation into DH5α Gold efficiency as described in 3.3.2.

Colony PCR was performed on colonies from each NheB construct, using the primers used in PCR i.e. the cloning primers. Colony PCR was carried out as described in 3.2.9, with annealing temperatures of 58 °C. Only one colony from an AES construct contained an insert of correct size. DNA sequencing confirmed that this AES construct contained the correct sequence (Figure 4.3), with only one base pair substituted that would result in the substitution of Thr78 to an alanine (Ala78). This construct was taken forward for overexpression.

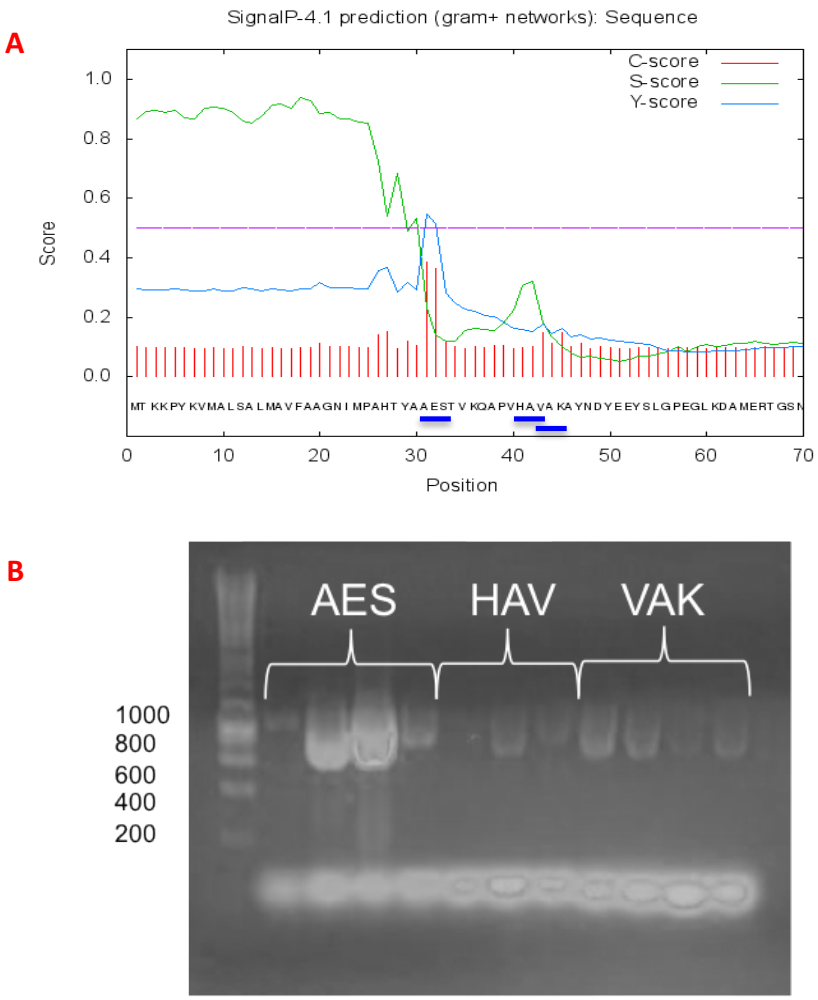


Figure 4.2. Cloning of NheB AES, HAV, and VAK constructus, based on SignalP results. (**A**) The SignalP signal peptide prediction graph for the N-terminal region of NheB, showing the C-score (red lines), the S-score (green line), and the Y-score (blue line). The starting residues for the mature NheB protein are ambiguous, and the mature NheB protein is predicted to start at $_{31}$AES, $_{41}$HAV, or $_{43}$VAK by the SignalP 4.1 server. (**B**) A photograph of the agarose gel run after PCR amplification of the three NheB constructs beginning AES (lanes 2 to 5), HAV (lanes 6 to 8), or VAK (9 to 12). Each NheB construct starting at the AES, HAV, or VAK residues were designed and cloned from genomic DNA, and ligated into a pET28a vector for overexpression trials.

A

```
79    AESTVKQAPVHAVAKAYNDYEEYSLGPEGLKDAMERTGSNALVMDLYALTIIKQGNVNFG  258
      AESTVKQAPVHAVAKAYNDYEEYSLGPEGLKDAMERTGSNALVMDLYALTIIKQGNVNFG
31    AESTVKQAPVHAVAKAYNDYEEYSLGPEGLKDAMERTGSNALVMDLYALTIIKQGNVNFG  90

259   NVSSVDAALKGKVIQHQDAARGNAKQWLDVLKPQLISTNQNIINYNTKFQNYYDTLVAAV  438
      NVSSVDAALKGKVIQHQD ARGNAKQWLDVLKPQLISTNQNIINYNTKFQNYYDTLVAAV
91    NVSSVDAALKGKVIQHQDTARGNAKQWLDVLKPQLISTNQNIINYNTKFQNYYDTLVAAV  150

439   DAKDKATLTKGLTRLSSSINENKAQVDQLVEDLKKFRNKMTSDTQNFKGDANQITSILAS  618
      DAKDKATLTKGLTRLSSSINENKAQVDQLVEDLKKFRNKMTSDTQNFKGDANQITSILAS
151   DAKDKATLTKGLTRLSSSINENKAQVDQLVEDLKKFRNKMTSDTQNFKGDANQITSILAS  210

619   QDAGIPLLQNQITTYNEAISKYNAIIIGSSVATALGPIAIIGGAVVIATGAGTPLGVALI  798
      QDAGIPLLQNQITTYNEAISKYNAIIIGSSVATALGPIAIIGGAVVIATGAGTPLGVALI
211   QDAGIPLLQNQITTYNEAISKYNAIIIGSSVATALGPIAIIGGAVVIATGAGTPLGVALI  270

799   aggaaavgggtagivlaKKELDNAQAEIQKITGQVTTAQLEVAGLTNIKTQTEYLTNTID  978
      AGGAAAVGGGTAGIVLAKKELDNAQAEIQKITGQVTTAQLEVAGLTNIKTQTEYLTNTID
271   AGGAAAVGGGTAGIVLAKKELDNAQAEIQKITGQVTTAQLEVAGLTNIKTQTEYLTNTID  330

979   TAITALQNISNQWYTMGSKYNSLLQNVDSISPNDLVFIKEDLNIAKDSWKNIKDYAEKIY  1158
      TAITALQNISNQWYTMGSKYNSLLQNVDSISPNDLVFIKEDLNIAKDSWKNIKDYAEKIY
331   TAITALQNISNQWYTMGSKYNSLLQNVDSISPNDLVFIKEDLNIAKDSWKNIKDYAEKIY  390

1159  AEDIKVVDTKKA  1194
      AEDIKVVDTKKA
391   AEDIKVVDTKKA  402
```

B

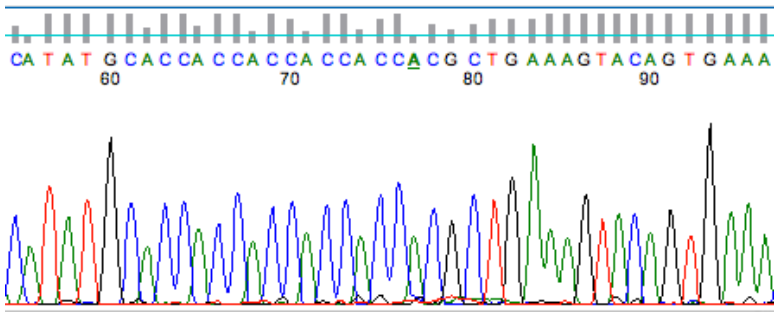


Figure 4.3. Successful cloning of the NheB AES construct. The sequenced gene contained within the pET21a vector was converted to its corresponding amino acid sequence, and compared with the amino acid sequence of the translated genomic DNA template. (**A**) Sequence alignment of the cloned NheB AES construct shown in the top line, with the translated deposited gene sequence on the bottom line shows that the amino acid sequence is almost completely identical throughout the sequence, except for a mutation at residue 78 of a threonine to an alanine highlighted by the red box. (**B**) The sequencing chromatogram also shows that the gene construct contains an N-terminal 6x Histag, prior to the AES starting residues.

4.1.2 NheB overexpression and purification results

Analysis of the codon usage of the NheB gene showed that the gene encoding NheB AES contained nine rare codons, some of which were in close proximity to one another in the gene sequence (Figure 4.4). The presence of such rare codons has been shown to affect protein production, and supplementation with co-expressing tRNAs for these codons can increase yield of protein (Wakagi and Oshima, 1998). *E. coli* lab strain BL21 codon plus RILP was used in order to improve expression levels of NheB. Expression of NheB in BL21 codon plus RILP trials showed that soluble protein was expressed at 25 °C and 37 °C incubations overnight after induction with 1 mM IPTG (Figure 4.5). Soluble protein was separated from cell debris as described in 3.4.4 and purified by NiHP followed by gel filtration (Figure 4.6). NheB eluted from gel filtration column in the void volume, suggesting the protein was aggregating or misfolded (Figure 4.6).

4.2 Bioinformatics introduction

Bioinformatical studies allow us to determine and annotate certain properties of proteins such as conserved amino acids and domains, *in silico*, and to use those properties to look for related proteins with potentially similar properties. One of the strongest tools available in bioinformatics is the ability to compare known gene and protein sequences against unknown sequences. As protein sequence and structure, and therefore function, are related, families of related proteins in sequenced genomes could be identified, to expand the number of members of a protein family.

Finding protein homologues in a range of organisms can have biological implications, if the newly found protein is homologous to a protein with a known function, suggesting the organism in which the new gene is found also has this function. In addition, this process is particularly useful to Structural Biologists, as any new family members identified increase the chances of obtaining a representative structure for a given protein (Gibrat et al., 1996). Often, the sequence variation between the proteins in a family resides on the surface, affecting how the molecules may pack together in a crystal (Rupp, 2009). Studying many examples of members of a protein family helps understand general functions and mechanisms of action of protein classes.

```
  1 gct gaa agt aca gtg aaa caa gct cca gtt cat gcg gta gca aaa gct tat aat gac tat
 21 gag gaa tac tca tta gga cca gaa ggc tta aaa gat gca atg gaa aga aca ggt tca aac
 41 gct tta gta atg gat ctg tac gct tta aca att att aaa caa ggt aat gtt aac ttt gga
 61 aat gta tca tct gtt gat gcg gct tta aaa ggg aaa gtg att cag cac cag gat aca gct
 81 aga gga aat gcg aag caa tgg tta gat gta ttg aag cca cag ctt att tca acg aat caa
101 aat atc att aac tac aat acg aaa ttc caa aac tat tat gat act tta gtt gct gcg gta
121 gat gca aaa gat aaa gcg acg ctt acg aaa ggc CTA act aga tta tca agt agt att aat
141 gaa aat aaa gca caa gtg gat cag tta gta gaa gac ttg aaa aaa ttc cga aat aaa atg
161 acg tcg gat acg caa aac ttc aag ggt gat gca aat caa att aca tct ATA tta gct agt
181 caa gat gca gga att ccg ctt ctg caa aat caa att aca acg tac aat gag gca att agt
201 aaa tat aat gca att att atc ggt tca tca gtt gcg aca gct CTA gga cca att gca att
221 att ggt ggt gca gta gtt att gct acg ggc gca gga acg ccg CTA gga gtc gca tta att
241 gca ggt ggt gca gca gca gct gta ggc ggt ggt aca gcc ggt atc gta tta gcg aag aaa gaa
261 ctt gat aat gca caa gct gaa att caa aaa ATA aca gga caa gtt aca act gct caa tta
281 gaa gta gct ggg tta acg aac att aaa aca caa acg gag tat tta aca aat acg att gat
301 act gca att aca gcg ttg caa aac att tca aac caa tgg tac aca atg ggg tca aaa tac
321 aat tct tta ctt caa aat gta gat tca att agt ccg aat gac ctt gtt ttt att aaa gaa
341 gat tta aac att gcg aaa gat agc tgg aaa aac att aaa gac tat gca gaa aag att tat
361 gct gaa gac att aaa gta gta gac aca aaa aaa gca taa
```

Figure 4.4. Rare codons present in the gene sequence for NheB. The NheB gene sequence contains 9 out of 373 rare codons (2.4 %), which are highlighted in the gene sequence by colouring based on amino acid translation, and underlined. The NheB gene contains 4 rare arginine codons (red), 3 rare leucine codons (green), 2 rare isoleucine codons (blue), and no rare proline codons. The rare codons CTA and aga highlighted in the red box are within close proximity, which can have a compounded effect on protein translation (Wakagi and Oshima, 1998). Rare codon analysis was performed using the rare codon calculator (RaCC)(nihserver.mbi.ucla.edu/RACC/).

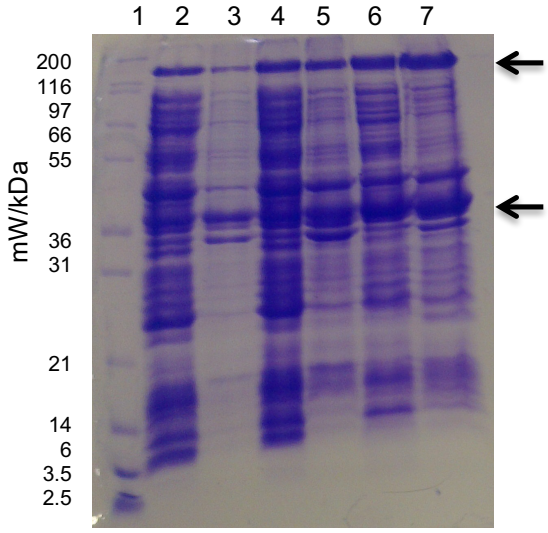
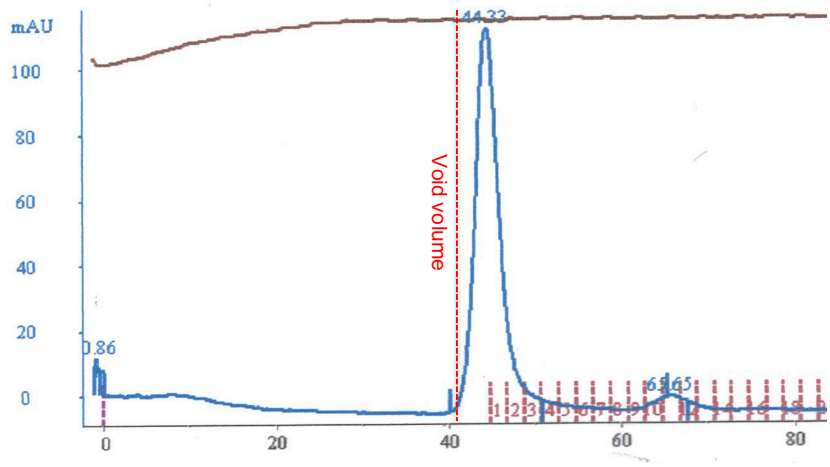


Figure 4.5. Overexpression trials of the NheB AES construct in *E. coli* DE3 Codon Plus RILP. An SDS PAGE gel containing soluble and insoluble fractions of sonicated *E. coli* cells before and after overexpression. Lane 1 contains Mark12 ladder, and molecular weights of the ladder are shown. Lanes 2 and 3 contain the pre-induction soluble and insoluble fractions, prior to NheB AES overexpression. Lanes 4 and 5 show the soluble and insoluble fractions of NheB AES expressed at 25 °C overnight respectively. Lanes 6 and 7 show the soluble and insoluble fractions of expressions carried out at 37 °C for 4 hours respectively. Black arrows indicate the molecular weight of NheB at 40 kDa, and the expected SDS-resistant high molecular weight band present for boiled samples.

A



B

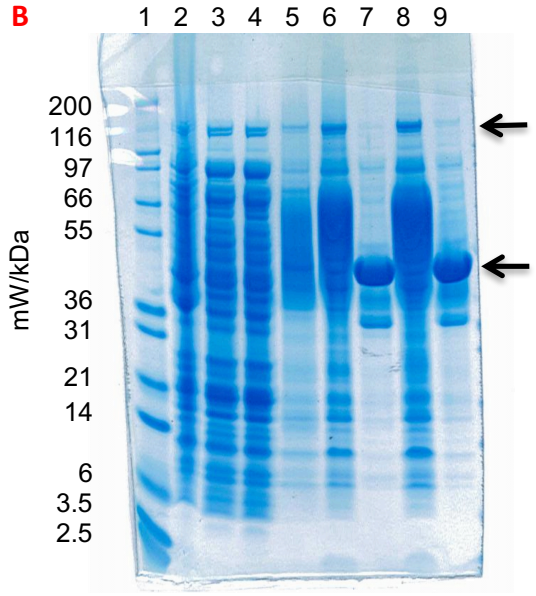


Figure 4.6. Gel filtration purification of NheB AES expressed in *E. coli* DE3 Codon Plus RILP. (**A**) A chromatogram showing the absorbance in mAU for the elution of NheB AES from a Superdex 200 pg column. The absorbance in mAU is shown as the blue trace, and the red dashed line shows the void volume. There were two peaks in the gel filtration elution, at 44.33 ml and 65.65 ml. These peaks were analysed by SDS PAGE gel electrophoresis. (**B**) An SDS PAGE gel of the above gel filtration. Lane 1 contains Mark12 ladder, and the molecular weights of each mark are shown. Lane 2 contains the cell debris and lane 3 the cell free extract. Lane 4 contains the NiHP unbound. Lane 5 contains the gel filtration load. Lane 6 and 7 contain fractions 1 and 2, and 10 – 13 respectively, with samples boiled. Lanes 8 and 9 contain the same as lanes 6 and 7, but without boiling.

Finally, conserved amino acids in a family can be mapped onto known protein structures, which can highlight amino acids that are otherwise separated in sequence but brought together by the fold of the protein. These patches of conserved amino acids can then be used to identify important areas on the protein such as active sites, binding pockets, domain interfaces etc.

All of these techniques were used in identifying and predicting important regions in new homologues to the *Bacillus cereus* NHE proteins, and are discussed below.

Homology searches were carried out in this thesis using the Basic Local Alignment Search Tool (BLAST) online server. The BLAST server 'finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.' (Altschup et al., 1990; Camacho et al., 2009).

4.3 Bioinformatical identification and selection of Nhe homologues

In order to expand the known members of the tripartite PFT family, the amino acid sequences of each protein from the Nhe tripartite system (NheA, NheB, and NheC) were used in Blastp searches to identify homologues, excluding hits from other *Bacillus* species. Blastp results were analysed in order to select suitable candidates for further study. These studies identified a number of proteins that were potentially members of the PFT superfamily, and are shown in Table 4.1, and discussed below.

4.3.1 Rickettsiella grylli Nhe homologues

The gene RICGR_0158 was identified in *Rickettsiella grylli* (GenBank AAQJ02000001.1), an intracellular pathogenic bacterium of *Gryllus bimaculutus,* and related species of crickets (Cordaux et al., 2007). This putative gene product of RICGR_0158, EDP46979.1 shared 29 % sequence identity to NheA over 23 % of the sequence residues, with an E-value of 0.01. The flanking genes surrounding RICGR_0158 were analysed for possible protein partners, in order to identify putative members of a tripartite system. The 5' upstream gene was 12 bp away from the hit, and was predicted to code for a viral A-type inclusion protein (protein ID EDP46038.1). The 3' flanking gene at the time of the homology search (2014) was predicted to be an amidase, with 99 % similarity to glutamyl-tRNA (Gln-tRNA) amidotransferase subunit A from a different

Organism	Pathogenicity	A component homologue accession and seq. identity	B component homologue accession and seq. identity	C component homologue accession and seq. identity
Rickettsiella grylli	Arthropod pathogen	EDP46979.1 (29 % identity over 23 %)	N/A	N/A
Xenorhabdus nematophila	Insect pathogen	CBJ91797.1 (57 % identity over 33 %)	CBJ91798.1 (43 % identity over 26 %)	N/A
Yersinia enterocolitica	Human pathogen	-	-	N/A
Aeromonas hydrophila	Fish pathogen	AHX33179.1 (30 % identity over 21 %)	AHX33180.1 (26 % identity over 92 %)	AHX33181.1 (16 % identity over 54 %)

Table 4.1. Initial hits from blast searches with NHE components. Initially NheA and NheB were used in blast searches, excluding hits from other *Bacilli* species. Three major hits, *Rickettsiella grylli, Xenorhabdus nematophila,* and *Aeromonas hydrophila* were identified. *Yersinia enterocolitica* was identified as containing homologues by literature search of the XaxAB proteins and no gene sequence is available. Each organism is a known pathogen, and references are given for each. The accession code for each protein is also given, along with the sequence identity to each NHE homologous counterpart.

strain of *Rickettsiella grylli* (protein ID EDP45780.1). The lack of flanking genes with homology to any PFTs suggested the gene RICGR_0158 might at most be a single component enterotoxin similar to ClyA. However, there is no sequence similarity between EDP46979.1 and ClyA, and so EDP46979.1 may represent a single component NheA homologue.

Recently (2015), a new *Ricketsiella grylli* genome sequence has been published (NZ_AAQJ02000001.1). In this genome, the NheA homologue WP_006035943.1 lies upstream of the gene WP_006035025.1, which shares 20 % sequence identity to NheC over 70 % of residues. 48 bp downstream of WP_006035943.1 lies WP_006035665.1, with 23 % sequence identity over 81 % of residues to NheB, restoring the tripartite nature of these homologues, and highlighting the presence of these toxins in another pathogenic bacterium.

4.3.2 Xenorhabdus nematophila Nhe homologues

A separate hit was identified in the genome of the motile bacteria *Xenorhabdus nematophila,* namely XaxA and XaxB*. Xenorhabdus nematophila* is a Gram negative bacteria found in symbiotic relationships with insect-parasitizing nematodes of *Steinernema carpocapsae. Xenorhabdus nematophila* is known to excrete a number of toxic proteins in outer membrane vesicles including PFTs (Khandelwal and Banerjee-bhatnagar, 2003). XaxA shared 57 % sequence identity with NheA (over 33 % of residues). XaxA lies upstream of XaxB, and these two genes have previously been identified as a binary toxin system, studied for its hemolytic activity (Vigneux et al., 2007). This binary toxin system may well give insights into the relationships between ClyA (the single PFT protein), and the tripartite system NHE found in *Bacillus cereus.* Interestingly, a homologue to XaxAB has also been identified in *Yersinia enterocolitica* (YaxAB), and YaxAB has been implicated in mouse spleen infection, and cell death by osmotic lysis (Wagner et al., 2013).

4.3.3 Operon of three genes in *Aeromonas hydrophila* with homology to the Nhe operon

The final candidate, and the one on which this thesis is based, was AHX33180.1 from a pathogenic species of *Aeromonas hydrophila*, a major fish and amphibian pathogen (Janda and Abbott, 2010a). The initial hit was identified by its sequence homology to NheB, with a 26 % sequence identity over 92 % of residues (E value 1e-13). Although no direct hits with BLASTP were discovered for NheA and NheC, analysis of flanking genes identified genes either side of

AHX33180.1, in close proximity (25 bp and 46 bp upstream and downstream respectively)(Figure 4.7). These flanking genes consisted of 30 % sequence identity over 21 % of residues (E value 3.9), and 16 % sequence identity over 54 % of residues (E value 0.18) respectively to NheA and NheC respectively, above the default E value cutoff in BLASTP. These analyses suggested that the three AHL proteins lie in an operon, and constitute a tripartite toxin similar to the NHE system from *Bacillus cereus*, and have been named as AhlA, AhlB, and AhlC (Figure 4.7).

The N-terminal 150 residues of both HblB and NheB have been characterised as the Hbl domain (Phelps and McKillip, 2002), and AhlB also shares sequence similarity in this region. However, the flanking genes AhlA and AhlC do not share sequence similarity in this region, and are not recognised as containing a Hbl domain. They do however contain a coiled coil domain, and a transmembrane domain respectively. The presence of these domains were encouraging, as this family of proteins are known to be formed of coiled coils, and contain regions which insert through membranes.

Kyte and Doolittle hydropathy plots were generated for each *Aeromonas hydrophila* protein, and compared with the corresponding NHE protein from Bacillus cereus (Figure 4.8). Each of the AHL proteins had a very similar hydropathy to the corresponding NHE protein, such that AhlA and NheA contain no regions of high hydrophobicity, AhlB and NheB, and AhlC and NheC have segments of between 54 and 63 residues and 22 and 23 residues respectively with high hydrophobicity.

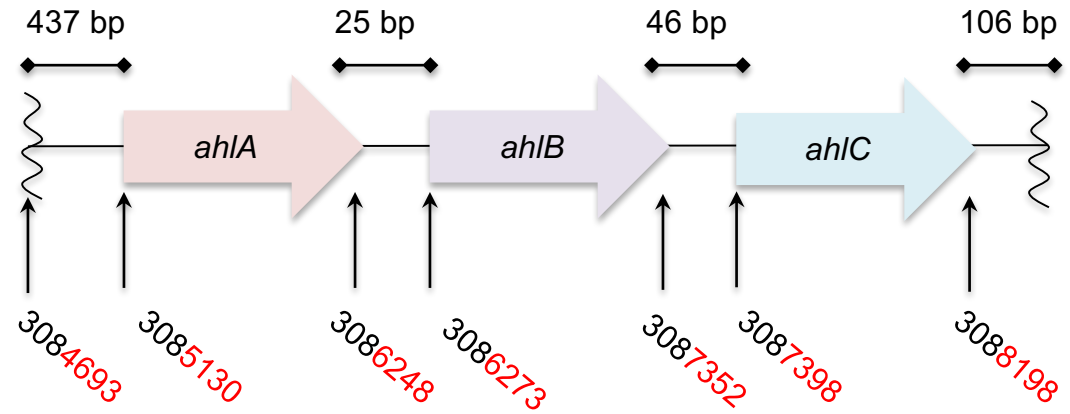
The AHL components were also used in BlastP searches limited only to the Aeromonas Taxid (Figure 4.9), and were found in 27 other species of Aeromonas, including AL09-71 and ML09-119, which are categorised as epidemic vAH strains of Aeromonas and known pathogens of fish species.

These three *Aeromonas hydrophila* proteins AhlA, AhlB, and AhlC were thus selected as suitable targets for structural studies, which are described in the following chapters.

4.4 Using AHL as a search criterion to expand the tripartite family further

As it appeared likely that the AhlA, AhlB, and AhlC proteins were a member of the tripartite PFT family, attempts were made to identify further members using BLASTP. Each of the AHL

A



B

Protein	Number of amino acids	Molecular weight	Theoretical pI
AhlA	372	39676.6	4.73
AhlB	359	37642.6	4.82
AhlC	266	28346.1	4.68

Figure 4.7. The organisation of the putative *ahl* operon, and the characteristics of the resulting protein products. (**A**) The organisation of the three *ahl* genes on the Aeromonas genome show that each lies in close proximity to one another, with a larger gap found at either end. Based on this analysis, the ahl genes are predicted to lie in an operon. The resulting putative protein products from each gene are shown (**B**), with the number of amino acids, the molecular weights, and the theoretical isoelectric point for each (based on Protparam analysis)

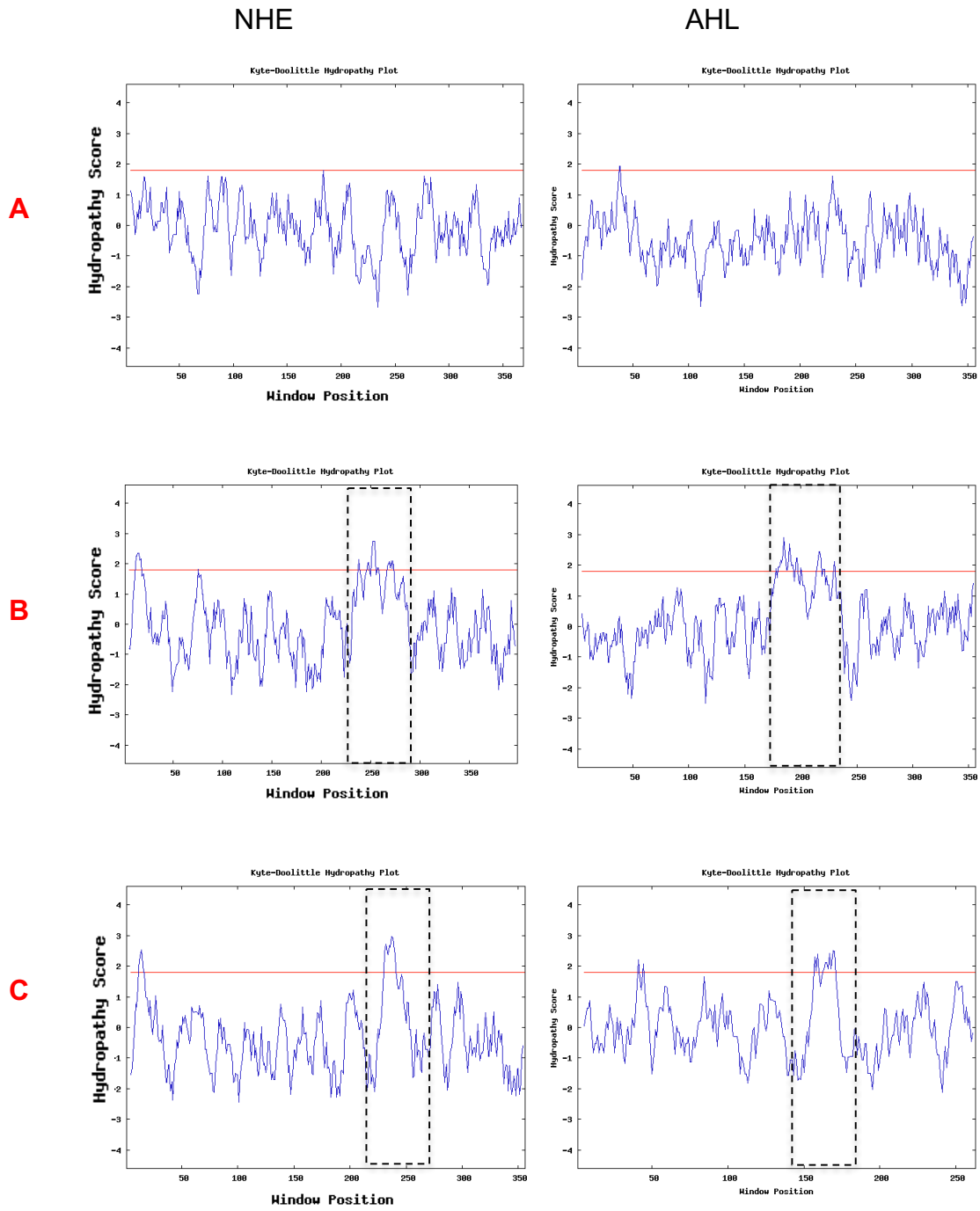


Figure 4.8. Kyte and Doolittle hydropathy plots generated for each NHE and AHL components. The A components of both NHE (**A** left) and AHL (**A** right) contain no stretch of hydrophobic residues for the head region (residues 197 to 242 for NheA based on structure). Both of the B components for NHE (**B** left) and AHL (**B** right) contain a 54 and 63 residue span of hydrophobic residues, which are predicted to form two transmembrane helices for both NHE and AHL (based on TMHMM plots). The C components of each NHE (**C** left) and AHL (**C** right) contain a shorter stretch of hydrophobic residues, and are predicted to form a single transmembrane helix (based on TMHMM plots). Stretches of hydrophobic residues are defined from peaks in the hydropathy plot where the hydropathy score exceeds 1.8 (Kyte and Doolittle, 1982). Kyte and Doolittle plots were generated at (http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm). Transmembrane helix prediction was carried out using the TMHMM Server v. 2.0, and are not shown.

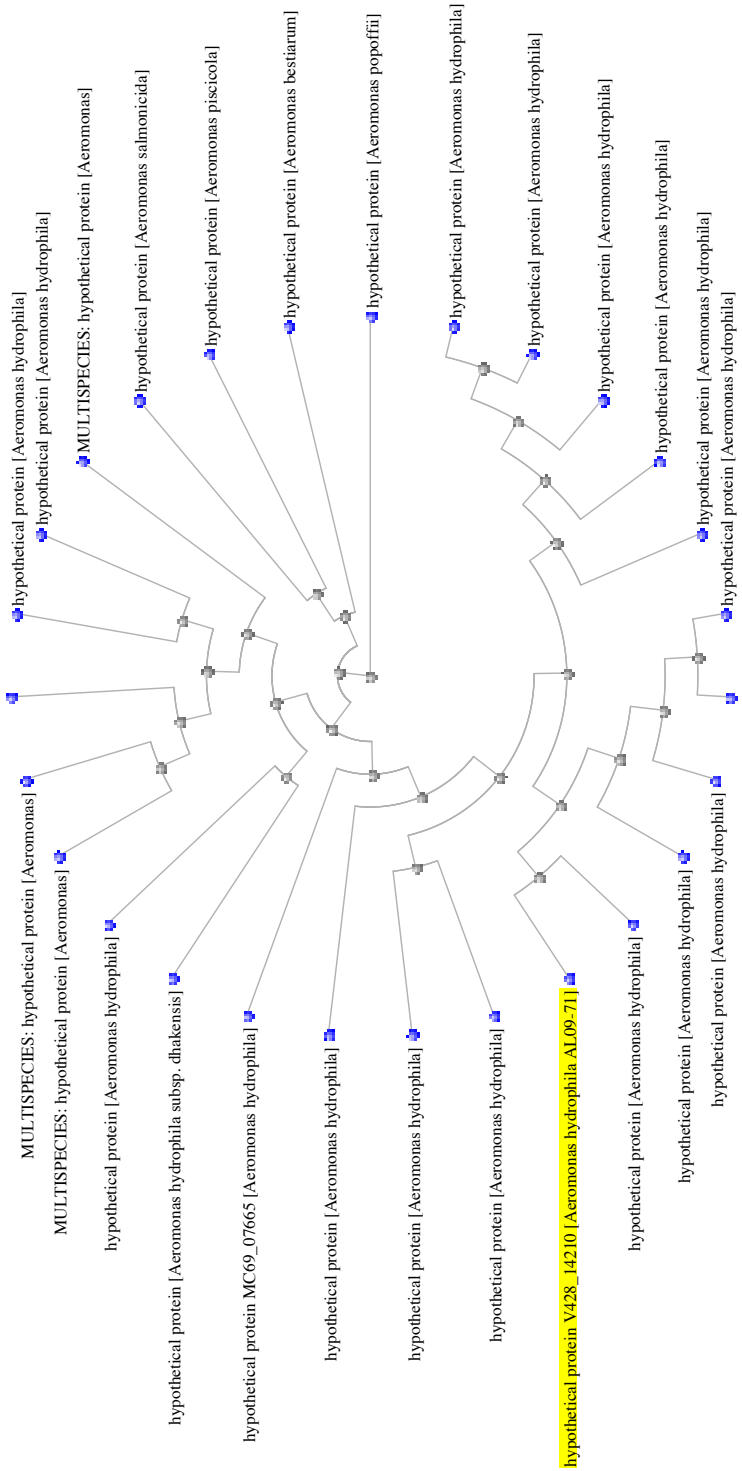


Figure 4.9. The Blast sequence identity tree of Aeromonas species and strains containing AhlC genes. AhlC was used in a blastp search, allowing only hits from Aeromonas groups (taxid = 644). Searches carried out with AhlA and AhlB show that the species are all identical; so only one tree is shown. The search sequence from *Aeromonas hydrophila* AL09-79 is highlighted in yellow, and each node (blue) indicates a different strain, group in sub-trees at grey nodes. Distances between nodes are determined by the Grishin distance method (Grishin et al., 2000). 25 strains of *Aeromonas hydrophila* carry the gene for Ahl, including AL09-71 and ML09-119, which are categorised as epidemic vAH strains (Pridgeon and Klesius, 2011b; Tekedar and Waldbieser, 2013) .

protein sequences were used in further BLASTP searches, excluding hits from *Aeromonas* and *Bacillus* species. These searches showed that the Aeromonas proteins were homologous to five other putative groups of three proteins, listed in Table 4.2, where all three AHL components have a corresponding homologue in the identified bacterial genomes. All of the AHL homologues that were identified also existed as three genes in close proximity to one another.

To further confirm that these three proteins are related to the NHE system, hydropathy plots were generated for each protein, and compared to the relevant NHE protein. Kyte and Doolittle hydropathy plots (Kyte and Doolittle, 1982) were generated for each homologue protein sequence, and all plots were found to contain the signature hydropathy observed for Nhe and Hbl, further confirming homology of these new proteins with AHL.

Four of the seven species from which these homologues were identified are known pathogenic bacteria, some with an associated haemolytic infection phenotype (Table 4.2)(Amin et al., 2010; Mahlen, 2011; Orata and Hedreyda, 2011). Phylogenetic analysis of the bacterial species containing homologues to AHL show that the bacteria are greatly removed from one another evolutionarily, suggesting an early evolution and divergence of each component over time (Figure 4.10). Interestingly, all of the results found from the AHL homologue searches were from Gram-negative bacteria, maintaining the fact that NHE and HBL are, to date, the only known members of this tripartite toxin family found in Gram-positive bacteria.

4.5 Analysis of conserved amino acids and mapping onto known structures for each tripartite component

4.5.1 Sequence alignment of the A components, and mapping of conserved residues on the structure of NheA

As many new members in the tripartite PFT family had been discovered, their amino acid sequences were analysed, to determine conserved residues and regions. As the structures of NheA (Ganash et al., 2013) and Hbl B (Madegowda et al., 2008) are known, for these components sequence patterns could be mapped onto the structure, to identify possible functional significance.

The amino acid sequences alignment for the homologues of the A component, aligning NheA and AhlA with 7 other homologues, were carried using the T-Coffee program (Notredame et al., 2000), and the results are shown in Figure 4.11. Initially we can see that NheA is at least

Organism	Pathogenicity	AhlA homologue accession and seq. identity	AhlB homologue accession and seq. identity	AhlC homologue accession and seq. identity
Erwinia mallotivora	Papaya Dieback Disease (amin et al, 2010)	WP_034933552.1 47 %	WP_034933555.1 79 %	WP_034933556.1 52 %
Chromobacterium piscinae	Not reported	WP_052247043.1 51 %	WP_043629747.1 78 %	WP_043629750.1 47 %
Rhizobium sp. Leaf262	Not reported	WP_062452822.1 27 %	WP_062452821.1 47 %	WP_062452820.1 31 %
Serratia liquefaciens	Very rare cases of human disease (Mahlen, 2011)	WP_044553510.1 45 %	WP_044553512.1 60 %	WP_044553514.1 46 %
Serratia plymuthica	Very rare cases of human disease (Mahlen, 2011)	WP_006320606.1 43 %	WP_043912873.1 62 %	WP_006320604.1 38 %
Tolypothrix sp. PCC 7601	Not reported	WP045872305.1 27 %	WP_045872306.1 48 %	WP_045872307.1 30 %
Vibrio campbellii	Emerging fish pathogen (Orata and Hedreyda, 2011)	WP_005532943.1 36 %	WP_005532945.1 63 %	WP_005532948.1 48 %
Average sequence identity		39.4 %	62.4 %	41.7 %
Number of homologues with E values below 0.01 not included here		4	>43	8

Table 4.2. A comparison of seven AHL homologues. Seven other tripartite systems were identified in Blastp searches using each AHL component, with the organism shown in column 1. All three components are present in each case, and the sequence identity to the corresponding AHL component are shown, as well as the accession codes for each protein. Any known pathogenicity of the organism, based on literature searches, is also given in the table of hits. Finally, the number of AHL component homologues from species that do not contain all three components are shown, with the average sequence identity between these proteins.

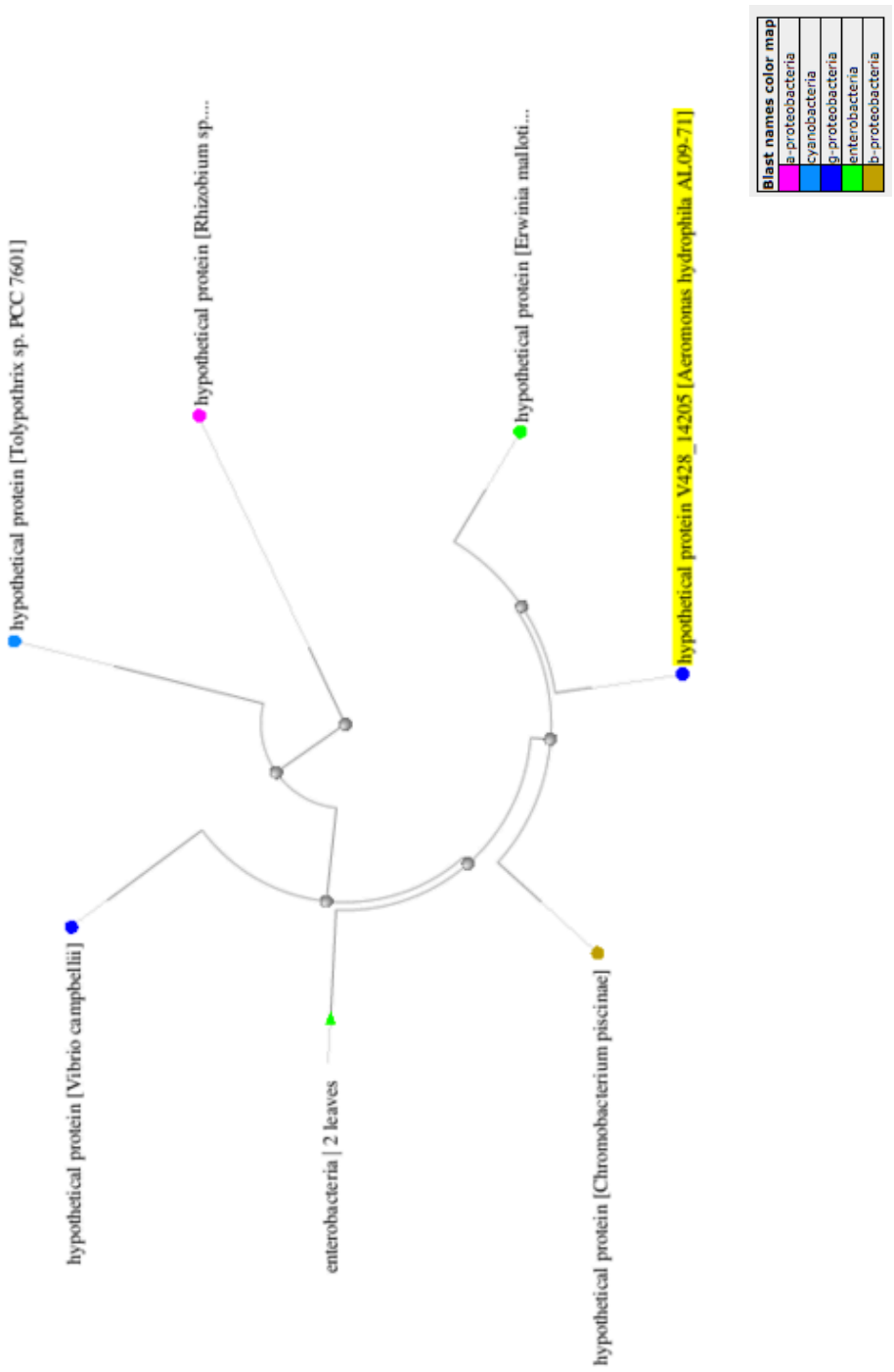
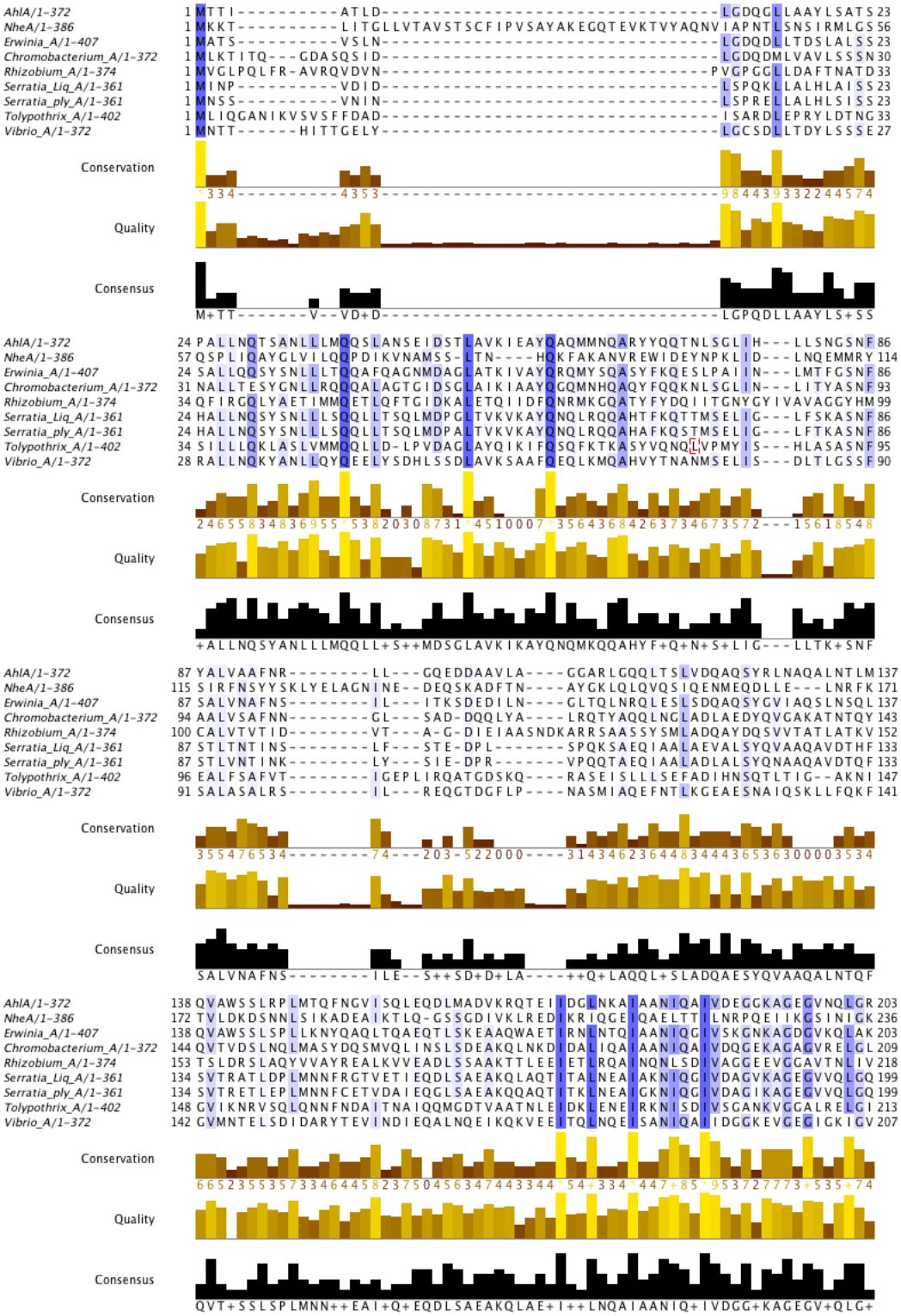


Figure 4.10. The sequence identity tree of AHL homologues where all three members are present, compared specifically between the B components. The organisms for which there was a homologue present for all three AHL components are shown on a phylogenetic tree, generated by Blast using the Grishin distance method (Grishin et al., 2000). The distance between each branch is based on the sequence identity between each protein, rather than the phylogeny, but gives an intuitive perspective on the distance between each protein in sequence similarity to AhlB.

Figure 4.11


AhlA/1-372
NheA/1-386
Erwinia_A/1-407
Chromobacterium_A/1-372
Rhizobium_A/1-374
Serratia_Liq_A/1-361
Serratia_ply_A/1-361
Tolypothrix_A/1-402
Vibrio_A/1-372
Conservation
Quality
Consensus

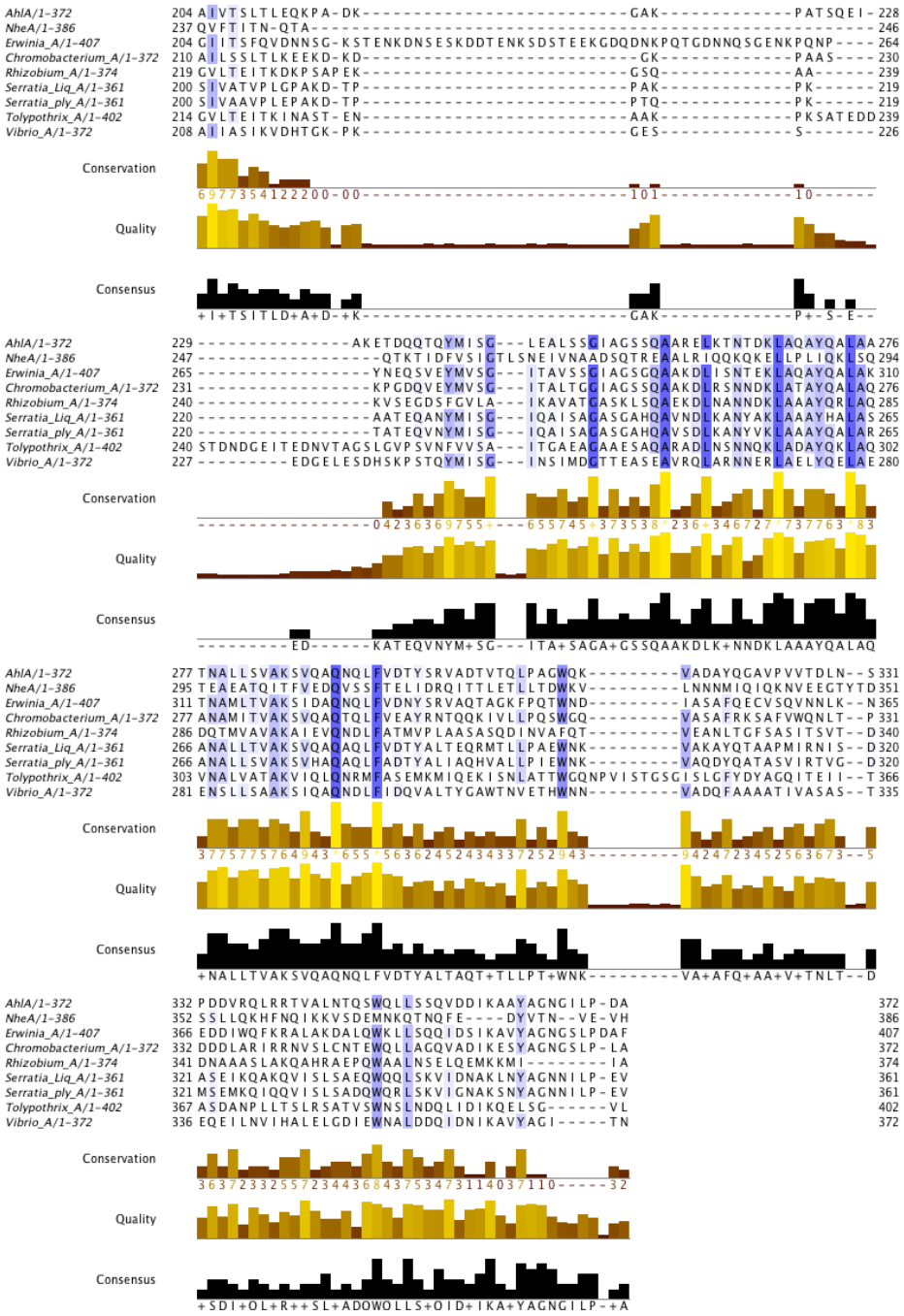


Figure 4.11. The sequence alignment between each A component of the tripartite pore forming toxin systems. The amino acid sequences of the A components of *Aeromonas hydrophila* (AhlA), *Bacillus cereus* (NheA), *Erwinia mallotivora* (Erwinia_A), *Chromobacterium piscinae* (Chromobacterium_A), *Rhizobium sp. Leaf262* (Rhizobium_A), *Serratia liquefaciens* (SerratiaL_A), *Serratia plymuthica* (SerratiaP_A), *Tolypothrix sp. CC 7601* (Tolypothrix_A), and *Vibrio campbellii* (Vibrio_A) were aligned using the JalView software. Each sequence is aligned with the strongly conserved residues highlighted in purple. Beneath the sequence alignment is a bar graph showing the number of sequences for which a residue is conserved, with more highly conserved residues shown a brighter yellow. Finally a consensus sequence for the aligned amino acid sequences is given, with a + symbolising no consensus for the given residue number.

26 residues longer than all of the other homologues, at the N-terminus, but otherwise is similar in length to all the other homologues (within 5 %). The conserved amino acids were then mapped onto the structure of NheA (Figure 4.12). The levels of conservation in the alignment are placed into two categories, residues that are 100 % conserved in identity, and residues that are conserved in residue type i.e. charge type. The alignment highlights three interesting regions of conservation in context with the NheA structure mapping.

4.5.2 Amino acid conservation in the head domain of NheA homologues

The first area of conservation in the residues of the A components map onto the head region of NheA. In NheA, the two helices α3 and α4 are bent downwards from the main body of the protein, and lead to the beta tongue region, contain a number of highly conserved leucine and isoleucine residues. These residues form a leucine zipper between the helices, and keep the two helices closely associated. The leucine and isoleucine residues are highly conserved in the other A component sequences (Figure 4.11). In addition, NheA has three leucine and isoleucine residues in the beta tongue itself, and these are loosely conserved in the A component of the family. The beta tongue of NheA also contains a conserved phenylalanine, which in all the family member sequences is either a phenylalanine or tyrosine, and in NheA faces toward the interface between the downward helices and beta tongue. In NheA this phenylalanine is surrounded by two other phenylalanine residues, forming pi-stacking interactions, however these residues cannot be obviously identified in the surrounding sequence of the other family members. Taken together, these conserved residues show an extensive network of hydrophobic interactions, which form a leucine zipper in the structure of NheA, as well as an interface between the downward helices of the head and the beta tongue region.

4.5.3 Family conserved Glutamine triad at the head domain interface in NheA

The second area of amino acid conservation in the A component is at the interface between the body of the protein and the head domain, highlighted in Figure 4.12 In the core of this region are three 100 % conserved glutamine residues, one found on each of helices α1, α2, and α6. In NheA, these residues are Gln45 on α1 and Gln282 on α6, which form polar interactions with the beta tongue main chain atoms, with Gln60 of NheA forming polar contacts with the loop between α1 and α2. Interestingly, the triad of Gln residues also occurs in Hbl B, a C family

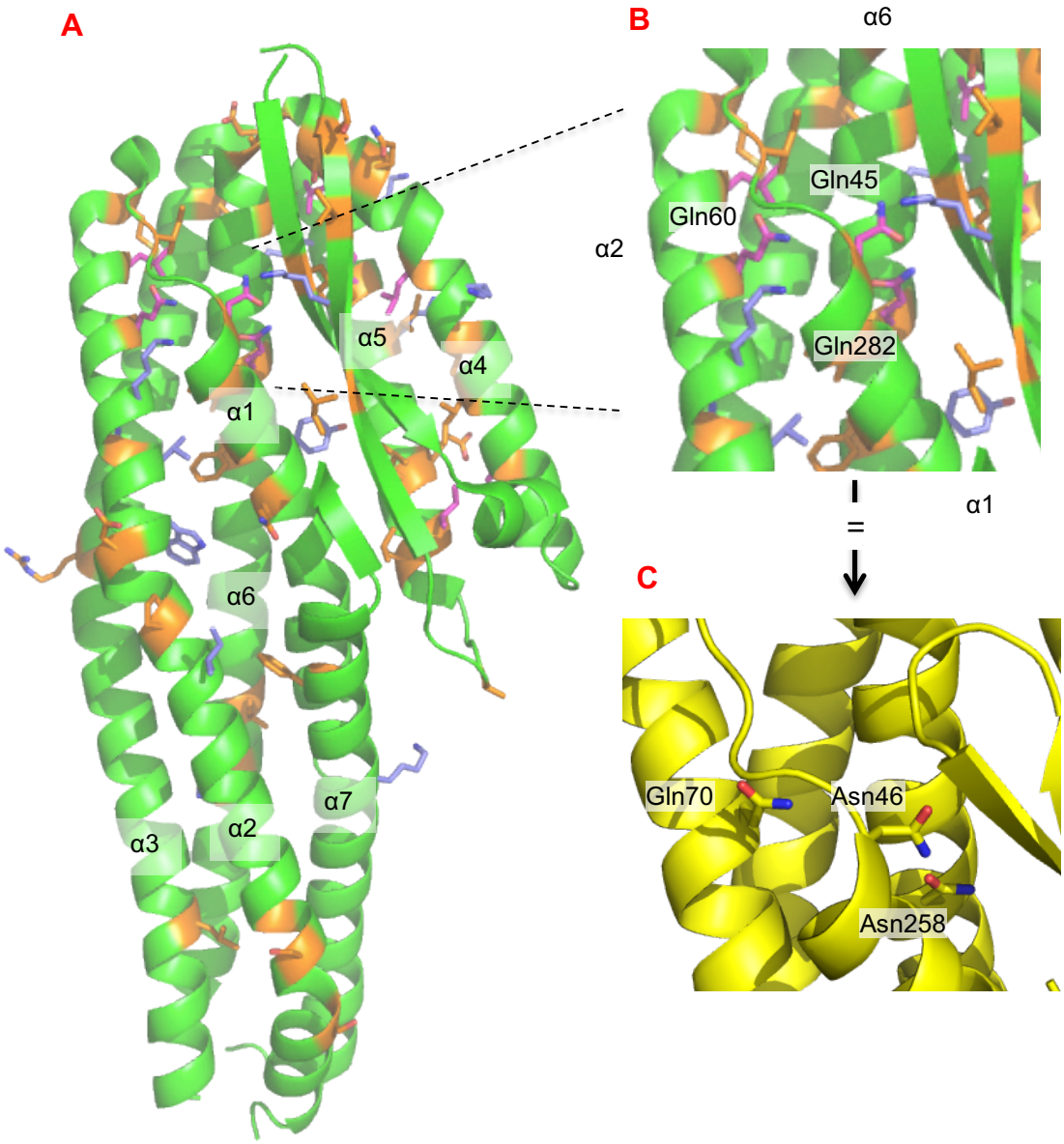


Figure 4.12. The crystal structure of NheA with family conserved residues highlighted. (**A**) The family conserved residues determined by amino acid sequence alignment of all tripartite PFT A components are mapped onto the crystal structure of NheA (Accession 4KIP). Magenta residues are 100 % conserved in identity between all A component homologues, orange residues are conserved in the characteristics of the residue. (**B**) A magnified view of the head domain of NheA shows that three family conserved glutamines are positioned at the interface between α1, the β-tongue, and the main body of the protein. (**C**) These residues are also conserved as glutamine or asparagine residues in the equivalent positions in HblB (PDB accession 1NRJ), a C component of the tripartite Hbl from *B. cereus.*

member that shows a similar structure to NheA, although in this case, one of the Gln residues is replaced by an asparagine. α1 also contains a highly conserved (six out of ten sequences) tyrosine (Tyr38) that again sits at the interface between helical body and beta tongue. Although the function of these residues is not certain, they may be involved in co-ordinating a conformational change, from the soluble form of NheA to the pore form, similar to that of the conformational change seen in ClyA (Mueller *et al.*, 2013).

4.5.4 Lack of amino acid conservation in the tail domain of NheA

In contrast, the main body of the tail domain in NheA, that makes up the five helical bundle, has a distinct lack of sequence identity. The conservation of amino acids in the tail domain is much less than in the head domain. Only a few hydrophobic and aromatic residues, that in NheA face inwards forming kinks and bulges in the five helical bundle, are conserved in the other species.

4.5.5 Sequence alignment of the B components, and analysis based purely on conserved residues

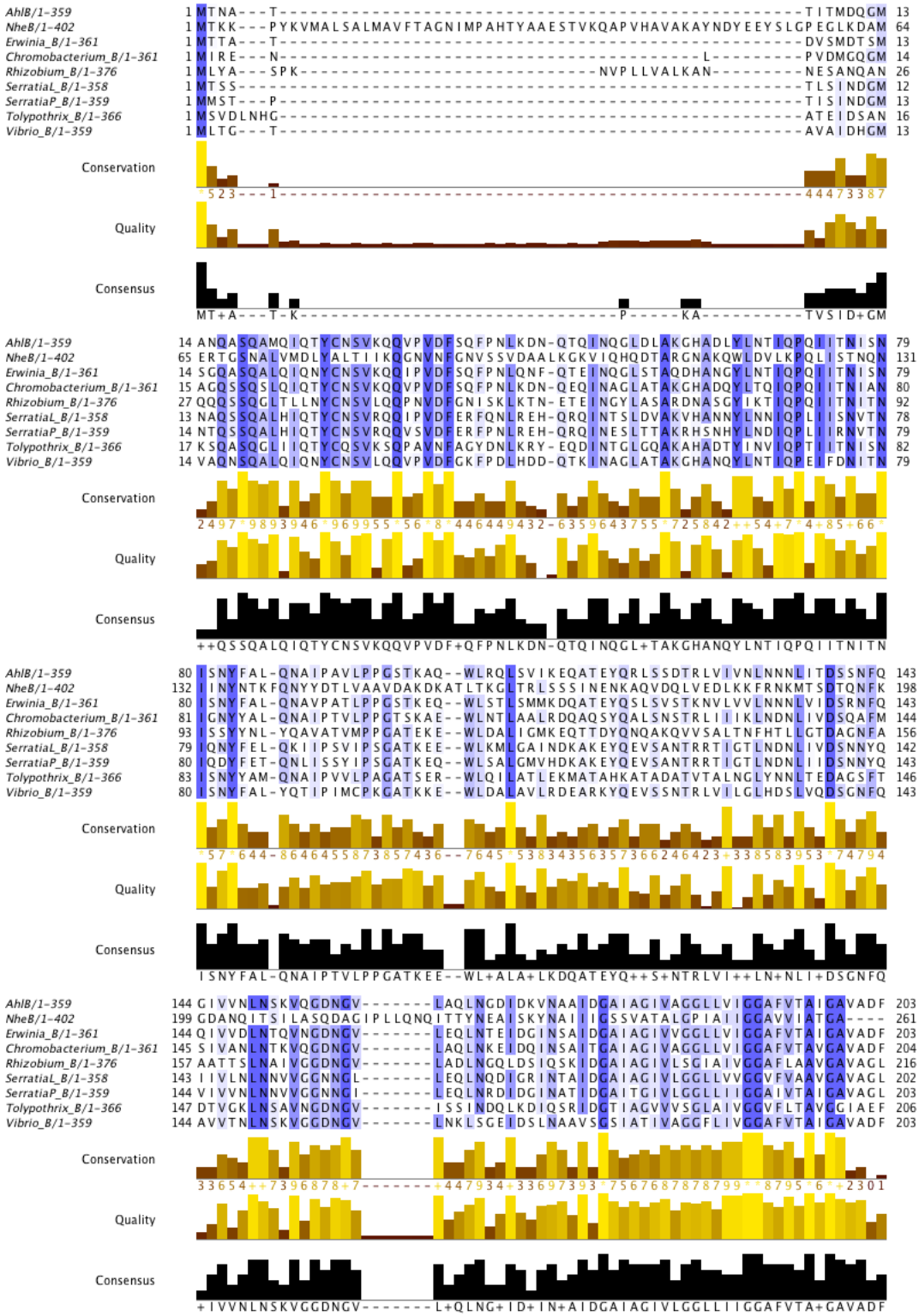
The B component of the tripartite PFT family is the only one lacking a known structure as NheA and Hbl B represent the A and C components respectively. Within the family, the B components have the highest sequence identity between homologues identified from blast searches (average 62 % identity). All proteins used in the alignment were identified by blast searches using AhlB as a search sequence, with the lowest sequence identity being 41 % identity over 92 % of the residues. The sequence alignment for the B components is shown in Figure 4.13.

The information available based on sequence conservation is limited with the B component alignment due to no currently available structures. However, a few noteworthy observations can be made, and kept in mind for analysis when structures are made available in the future.

4.5.6 Family conserved tyrosine residue also observed in ClyA

Present in ClyA is a tyrosine Tyr27 at the C-terminal end of α1, which in the pore structure sits at the proposed membrane interface of the inserted pore (Figure 4.14). A tyrosine residue is also conserved 100 % in the sequences of the B components, at residue 21 of AhlB. Tyrosines are often found at membrane interfaces of membrane proteins forming an ‘aromatic belt’

Figure 4.13


AhlB/1-359
NheB/1-402
Erwinia_B/1-361
Chromobacterium_B/1-361
Rhizobium_B/1-376
SerratiaL_B/1-358
SerratiaP_B/1-359
Tolypothrix_B/1-366
Vibrio_B/1-359
Conservation
Quality
Consensus

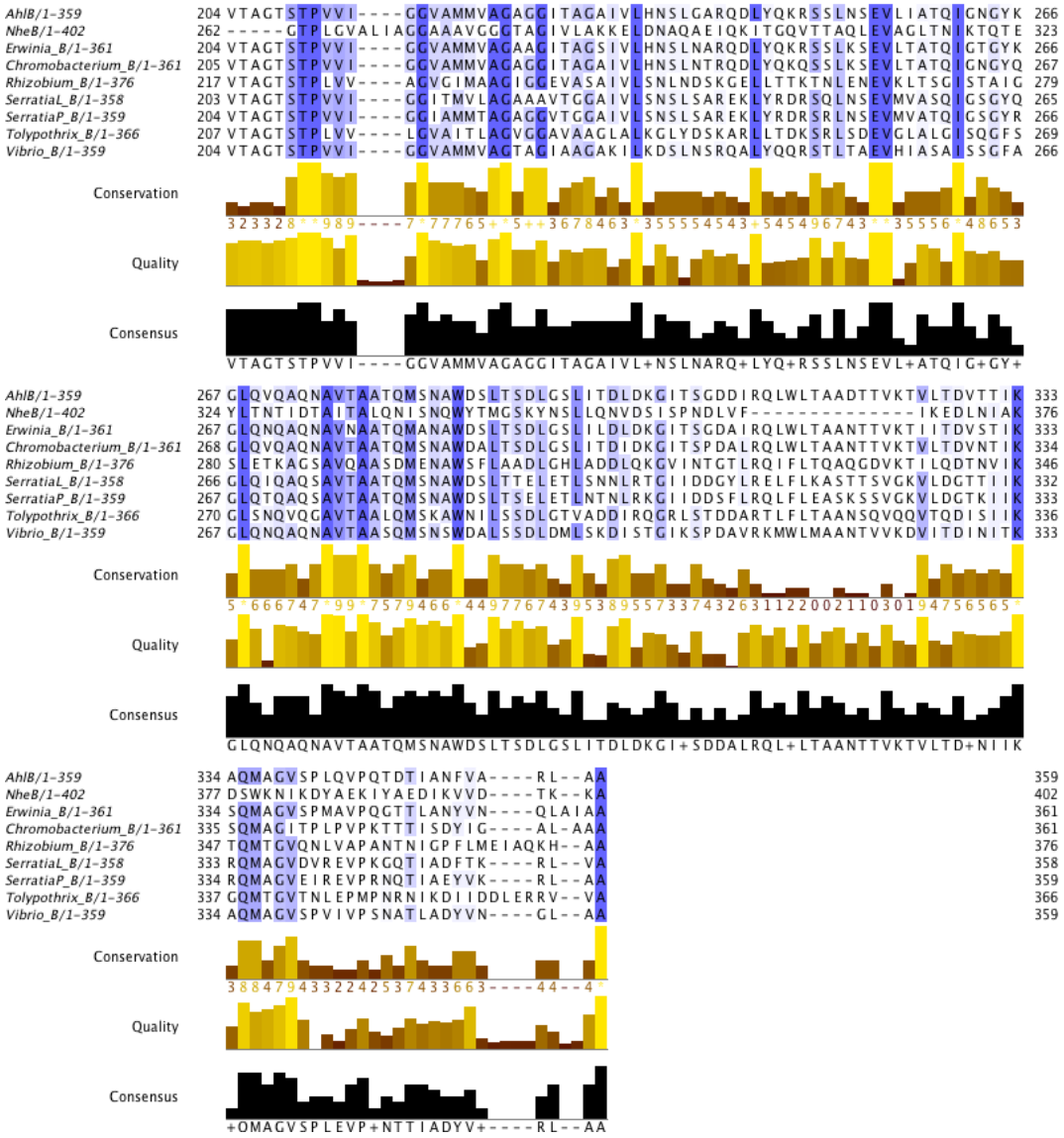


Figure 4.13. The sequence alignment between each B component of the tripartite pore forming toxin systems. The amino acid sequences of the B components of *Aeromonas hydrophila* (AhlB), *Bacillus cereus* (NheB), *Erwinia mallotivora* (Erwinia_B), *Chromobacterium piscinae* (Chromobacterium_B), *Rhizobium sp. Leaf262* (Rhizobium_B), *Serratia liquefaciens* (SerratiaL_B), *Serratia plymuthica* (SerratiaP_B), *Tolypothrix sp. CC 7601* (Tolypothrix_B), and *Vibrio campbellii* (Vibrio_B) were aligned using the JalView. Each sequence is aligned with the strongly conserved residues highlighted in purple. Beneath the sequence alignment is a bar graph showing the number of sequences for which a residue is conserved, with more highly conserved residues shown a brighter yellow. Finally a consensus sequence for the aligned amino acid sequences is given, with a + symbolising no consensus for the given residue number.

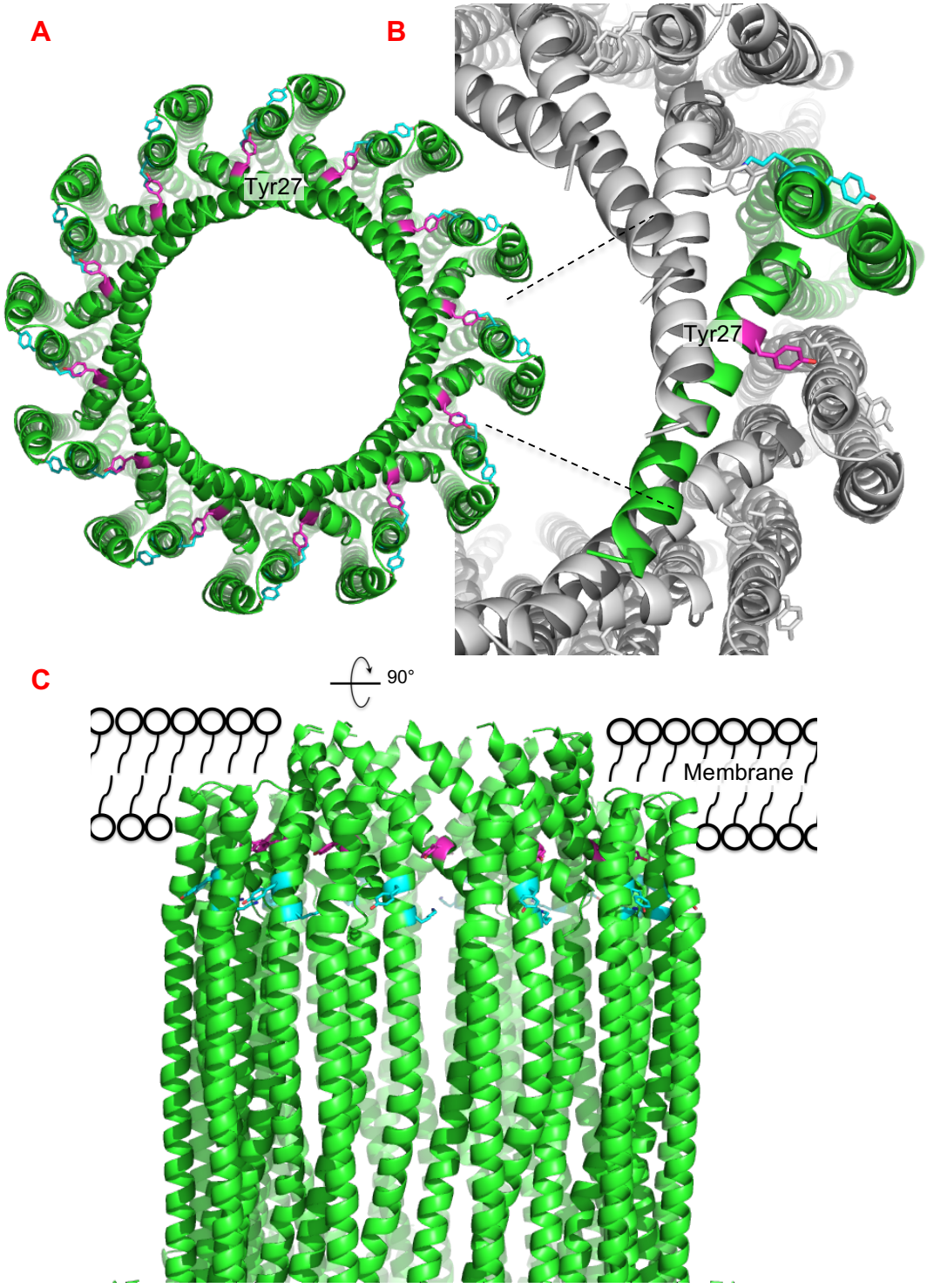


Figure 4.14. A family conserved tyrosine on the B component homologues maps onto an important region on the structure of ClyA. Residue 21 of AhlB is a family conserved tyrosine, which when aligned with the sequence of ClyA aligns to Tyr27. This residue Tyr27 in ClyA is part of the N-terminal helix that forms the N-terminal iris, which is proposed to penetrate the membrane in the pore form of ClyA (**A**). The conserved Tyrosine is highlighted in magenta (**B**), with the tyrosine and charged lysine flanking the hydrophobic hairpin in cyan. (**C**) Ribbon representation of ClyA pore showing the part of the protein that inserts into the membrane. The hydrophilic tyrosine and lysine (cyan) can be used to predict where in the membrane the pore will sit, which places the conserved Tyr27 at the membrane interface.

(Heijne, 1996; Pilpel et al., 1999), where the large aromatic benzene ring is able to insert into the hydrophobic core of the membrane, and the hydroxyl group prevents full insertion, and instead interacts with the polar head groups of the membrane. The high level of conservation and position of tyrosine in the newly discovered B components may therefore serve a similar role, implicating the N-terminus in membrane insertion.

4.5.7 Conserved stretch of hydrophobic residues for the B component homologues

The hydrophobic residues present in the B and C components are suggested to form a β-tongue, as seen in all tripartite toxin structures to date. The sequence of ClyA contains a stretch of 29 residues that form a β-tongue in the soluble structure, and a α-hairpin in the pore structure. The length of this hairpin is important, as it is long enough to insert into a single leaflet of a membrane. The stretch of hydrophobic residues in all B components is 26-32 residues, suggesting that a similar α-hairpin conformation would allow this region to insert into a single leaflet of the membrane in a similar way to ClyA.

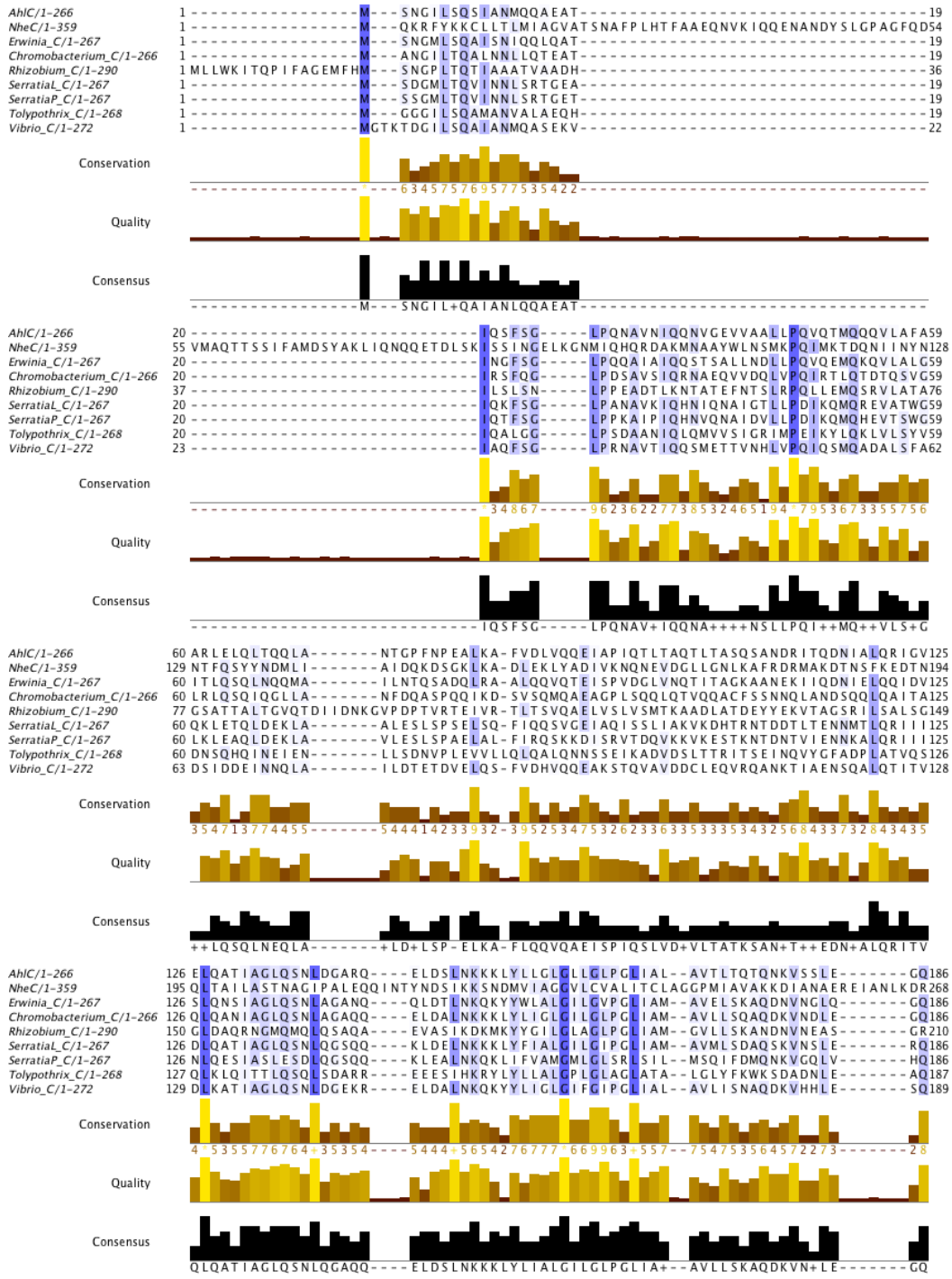
4.5.8 Sequence alignment of the C components, and mapping of the conserved residues onto the structure of Hbl B

The amino acid sequence alignment of the C components was carried out using AhlC, NheC, and the related homologues, together with HblB, for which there is a known structure (Madegowda et al., 2008)(Figure 4.15). The C components exhibited the lowest sequence similarity between the A, B and C family members, with only 4 identical residues conserved between the sequences, and with 35 residues having similar characteristics. The conserved residues were analysed and mapped onto the structure of HblB to try and draw conclusions on the features of the HblB structure in light of these conserved residues (Figure 3.16).

Initially we can see that HblB is much longer in the N- terminal region than the other proteins, as is NheC, although NheC is predicted to contain an N-terminal signal sequence. Alignment of these sequences suggests an end point for the signal sequence of NheC in a similar manner to NheB alignments. There are also three other regions in which HblB is longer than the other proteins, from residues 94 to 111, 262 to 275, and at the C-terminus from residue 328 to the end of HblB (Figure 3.16).

The first region of amino acid conservation is in the head domain of Hbl B (Figure 3.16), between α1 and α2.

Figure 4.15


AhlC/1-266
NheC/1-359
Erwinia_C/1-267
Chromobacterium_C/1-266
Rhizobium_C/1-290
SerratiaL_C/1-267
SerratiaP_C/1-267
Tolypothrix_C/1-268
Vibrio_C/1-272
Conservation
Quality
Consensus

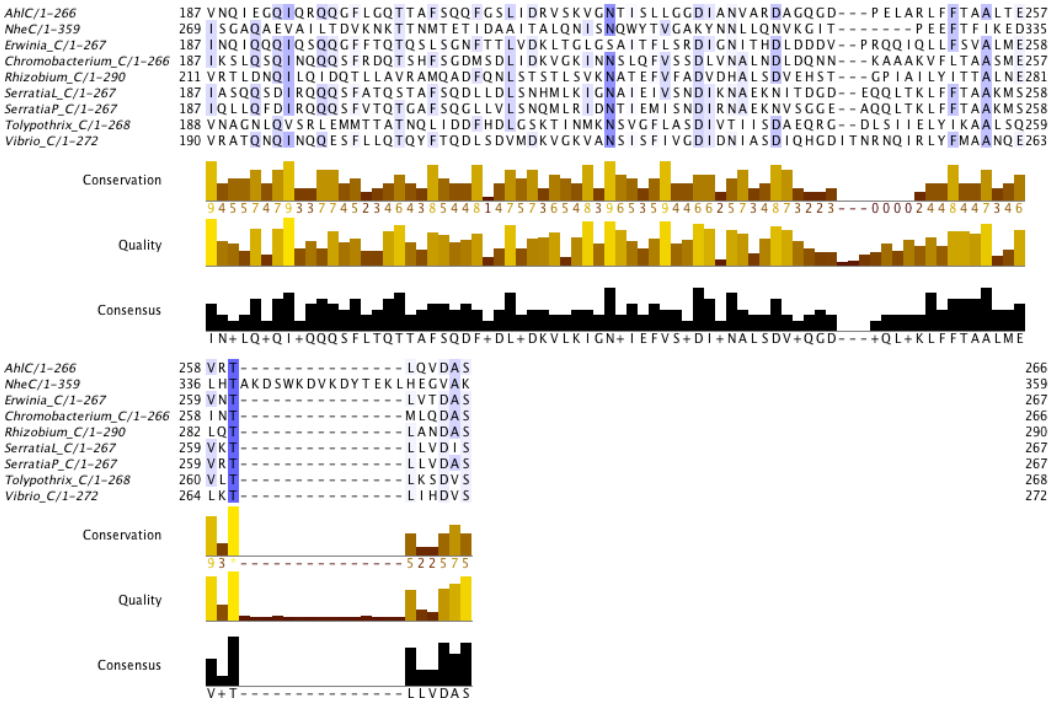


Figure 4.15. The sequence alignment between each C component of the tripartite pore forming toxin systems. The amino acid sequences of the C components of *Aeromonas hydrophila* (AhlC), *Bacillus cereus* (NheC), *Erwinia mallotivora* (Erwinia_C), *Chromobacterium piscinae* (Chromobacterium_C), *Rhizobium sp. Leaf262* (Rhizobium_C), *Serratia liquefaciens* (SerratiaL_C), *Serratia plymuthica* (SerratiaP_C), *Tolypothrix sp. CC 7601* (Tolypothrix_C), and *Vibrio campbellii* (Vibrio_C) were aligned using the JalView. Each sequence is aligned with the strongly conserved residues highlighted in purple. Beneath the sequence alignment is a bar graph showing the number of sequences for which a residue is conserved, with more highly conserved residues shown a brighter yellow. Finally a consensus sequence for the aligned amino acid sequences is given, with a + symbolising no consensus for the given residue number.

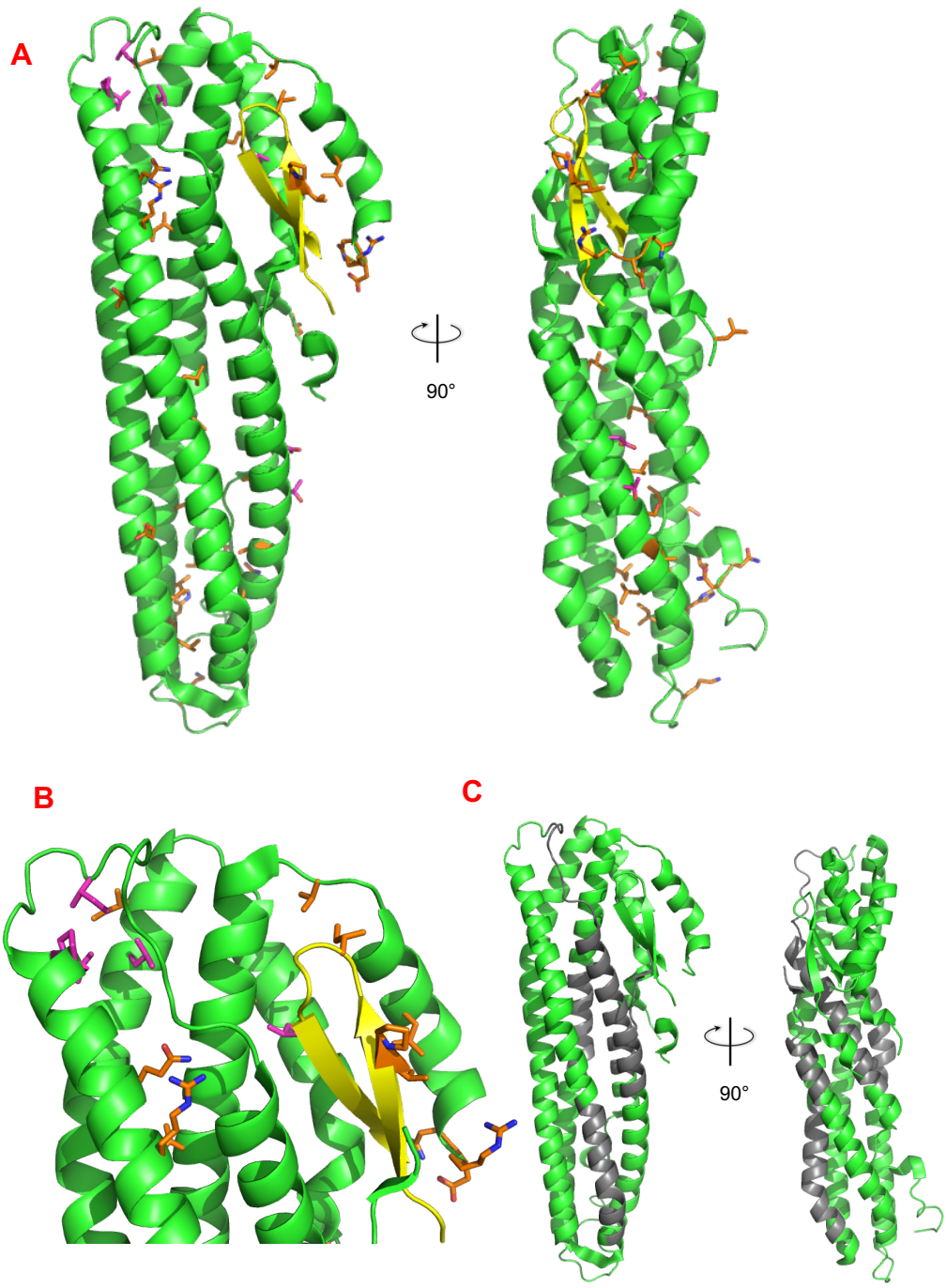


Figure 4.16. The structure of Hbl B, with the family conserved residues highlighted. The family conserved residues of the sequence alignment between all C components of the tripartite PFTs are mapped onto the crystal structure of Hbl B (PDB accession 2NRJ) (**A**). Magenta residues are 100 % conserved in identity between all A component homologues, orange residues are conserved in the characteristics of the residue i.e. charge etc, and the hydrophobic region in the head domain is highlighted in yellow. Many of the conserved residues in the C components are located at the head region of the protein (**B**) Hbl B is obviously much longer in a number of regions than all of the other C component homologues, and these extra regions are highlighted in grey in (**C**).

The glutamine asparagine triad highlighted in Hbl B based on its conservation in the A component is not conserved within other C components, i.e. the C components, other than Hbl B, do not contain a conserved triad of glutamines/asparagines. A number of hydrophobic leucine and isoleucine residues are wholly conserved in the head domain of the C components, with no obvious conclusion to their function.

Within the head domain of Hbl B, the β-tongue is constituted from a stretch of ~20 hydrophobic residues. This pattern is conserved in all the C component sequences, with the exception of *S. plymuthica* where an arginine lies in the middle of its hydrophobic region. A cluster of up to three conserved lysine residues can be identified at the N-terminal side of the hydrophobic residues, with many of the C components also containing a tyrosine in this region.

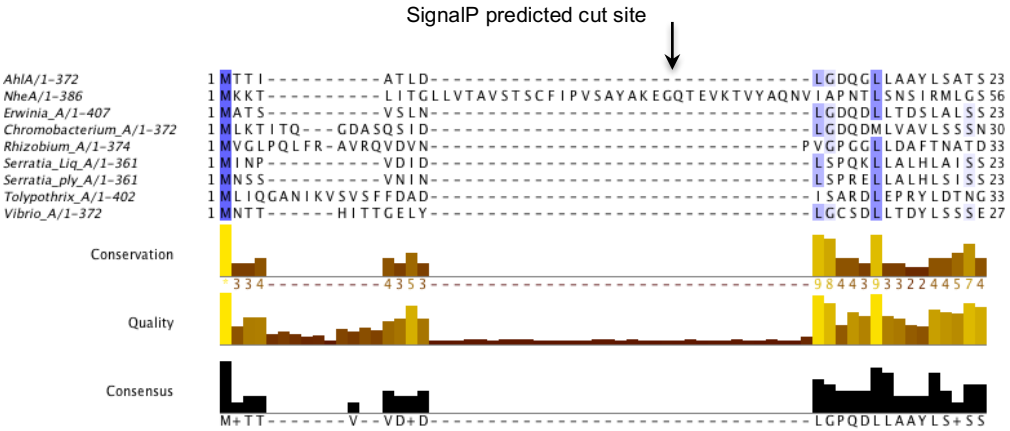
4.5.9 Analysis of the signal sequences present in the Gram-positive tripartite toxin family members

The Gram-positive tripartite toxins have been known to contain signal sequences that are involved in export of the protein from the cell via the Sec secretion pathway (Fagerlund, 2010). These signal sequences have hindered expression of the NHE proteins in *E. coli* expression systems, and prediction of the correct termination site has proven difficult. By analysing the family member alignments, a more confident prediction can be made on the correct N-terminus starting point for each protein.

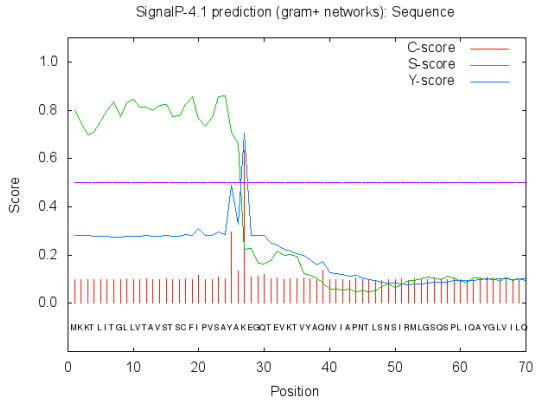
4.5.10 Signal sequence prediction for NheA, based on the structure of NheA and the family members sequence alignment

The first protein to be analysed to predict the start of the mature protein based on the amino acid sequence alignment to the newly discovered proteins was NheA (Figure 4.17). The sequence alignment shows that ~8 residues from each new Gram-negative NheA homologue are aligned with the extreme N-terminus of NheA. Downstream of this, the proteins come back into alignment at Leu9 for AhlA. The predicted cleavage site for NheA based on SignalP is after residues $_{KEG}$29 (Figure 4.17. Lund and Granum observed that the mature NheA protein was 11 residues shorter than the predicted size for cleavage at $_{KEG}$29 (Lund and Granum, 1996). The newly predicted cleavage site would therefore be within the proximity of the $_{KEG}$29 site, and up to 11 residues downstream. The crystal structure of NheA is known, and so can also be analysed for the start point in the electron density. As the NheA protein used to determine the structure was expressed in *E. coli* with a start site of $_{KEG}$29, the protein in the crystal structure

A



B



C

Chain	Start Site (black)
A	IAPNTL
B	VKTVYAQNVIAPNTL
C	IAPNTL
D	EVKTVYAQNVIAPNTL
E	IAPNTL
F	IAPNTL
G	IAPNTL
H	VIAPNTL

Figure 4.17. The N-termini of the newly identified A component AHL homologues aligned with NheA. (**A**) The alignment of all of the AhlA homologues with NheA shows that the alignment between all proteins begins at Ile42. (**B**) The SignalP server predicts that the signal sequence in NheA ends at Gly29, with a weaker signal present ending at Gln39, much below the threshold value of 0.5. (**C**) The starting amino acid of each chain in the NheA protein are tabled, and show that most chains begin near to or past Pro43. Chain B and D contain an extra N-terminal span of residues, which form a small helical region in the protein involved in crystal packing. Residues shown in red are missing from the chains electron density.

contains the protein with no further signal sequence processing. As such, the structure can be compared to the new predicted start of the mature protein based on the amino acid sequence alignment, to see if the structure correlates with the prediction. In the eight chains present in the NheA structure (PDB accession 4K1P), only two chains contain more residues present at the N-terminus of Ile42, and the other six begin either at the Ile42 residues, or one or two amino acids downstream of Ile42. These results combined suggest at least that the mature NheA protein may begin at residue Ile42, 12 residues downstream of the previously predicted Gly29 of the $_{KEG}$29 motif, and in accordance with the mature protein observed in NheA purifications performed by Lund and Granum.

4.5.11 Updated signal sequence prediction for NheB

Next, the start of the mature NheB protein was predicted by aligning all of the amino acid sequences from the newly discovered Gram-negative tripartite toxin B components with that of NheB (Figure 4.18). Previously the start of the mature NheB protein predicted by the SignalP server was ambiguous, between $_{AES}$33 and $_{VAK}$45 (Fagerlund, 2010). However, based on the amino acid sequence alignment, the start of the protein may be a further thirteen residues downstream of the $_{AES}$33 predicted site. This prediction may explain why attempts to express NheB starting at $_{AES}$33 led to soluble expression but misfolded protein, as discussed earlier. As such it may be of interest to reclone this gene to take into account the newly predicted start site for the mature NheB protein, to attempt to produce soluble correctly folded protein for crystallisation experiments.

4.5.12 Difficulty in assigning a signal sequence length for NheC

Based on the alignment of the N-termini of the C components (Figure 4.19), it is clear that the predicted signal sequence cleavage site predicted for NheC by SignalP is much to long, if compared with the start of the other proteins. However, the sequence alignment alone gives no clear indication of a start site, other than perhaps the family conserved isoleucine, present in NheC at residue 84 (Figure 4.19).

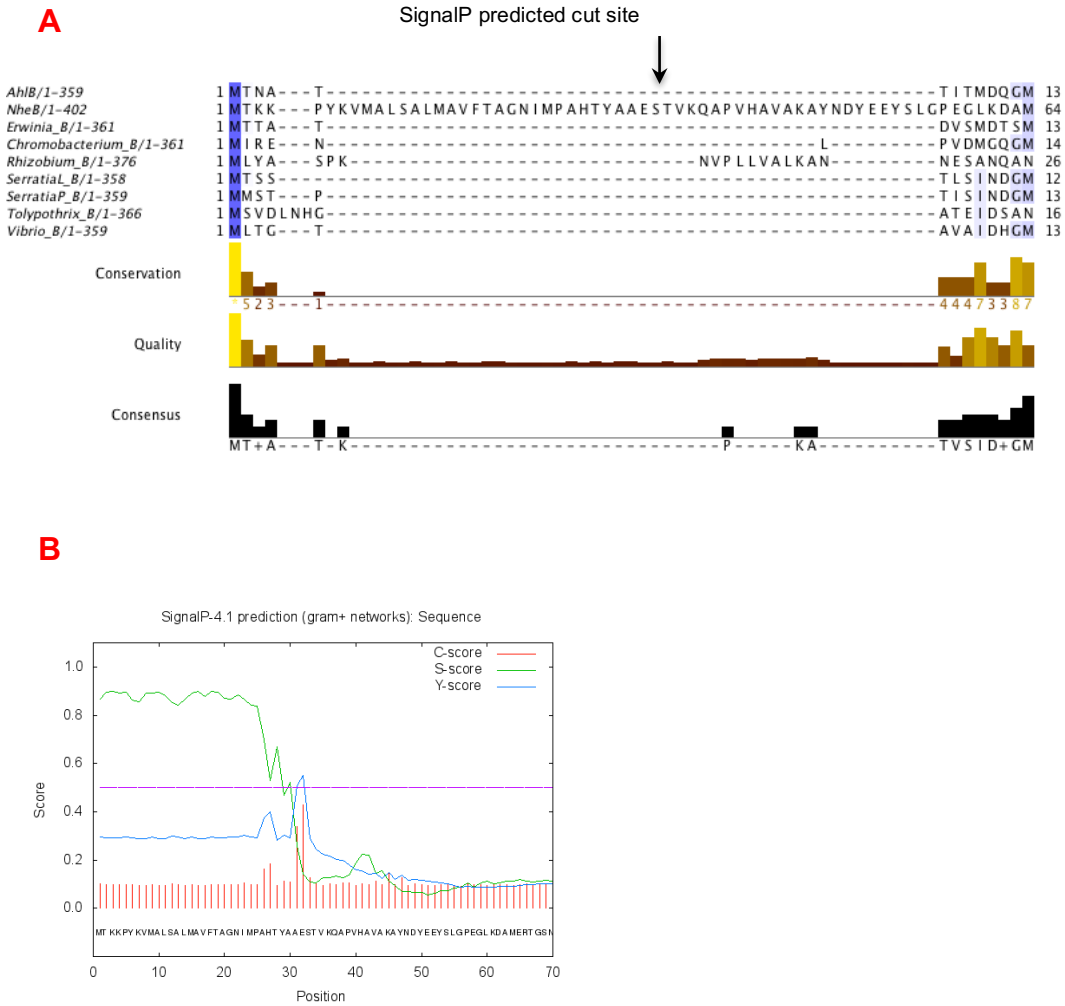


Figure 4.18. The N-termini of the newly identified B component AHL homologues aligned with NheB. (**A**) The alignment of all of the AhlB homologues with NheB shows that the alignment between all proteins begins at Met64 of NheB (conserved in 7 sequences). (**B**) The SignalP server predicts that the signal sequence in NheB ends at $_{AES}33$, $_{HAV}43$, or $_{VAK}45$, with only $_{AES}33$ rising above the 0.5 cut-off threshold for the Y-score.

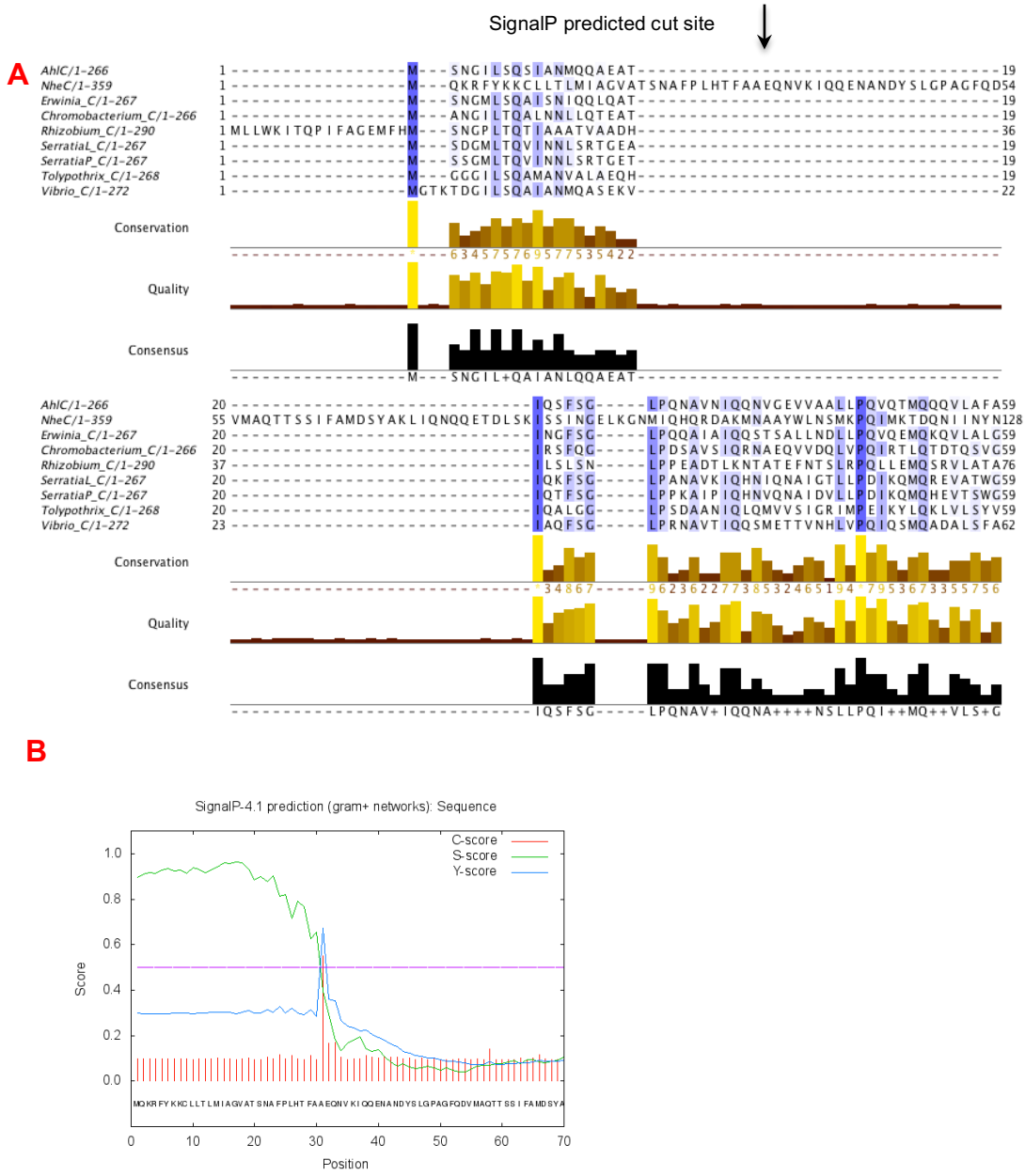


Figure 4.19. The N-termini of the newly identified C component AHL homologues aligned with NheC. (**A**) The alignment of all of the AhlC homologues with NheC shows that the alignment between all proteins begins at Ile84. (**B**) The SignalP server predicts that the signal sequence in NheC ends at $_{FAA}32$.

4.6 Conclusions of Bioinformatical analysis and experimental results

4.6.1 Bioinformatical conclusions

Six new homologues of the Nhe tripartite PFT family have been identified by Bioinformatical studies. Studies *in silico* have been carried out to highlight putative conserved regions within each component of the tripartite family, and these conserved residues were mapped onto currently known structures, such as NheA. The amino acid sequence alignments provide more contexts to the previously solved structures NheA and Hbl B, and provide a starting point for designing experiments to probe the function of each component in the tripartite PFT system. These homologues also offer a large pool of targets to select from for structural studies, to better understand the pore assembly, and to probe role of each component in the pore mechanism.

More generally, the presence of many tripartite homologues to NHE raises an interesting question as to the pathogenicity of organisms that contain the tripartite PFT genes. It has been shown that NHE expression in Bacillus is sufficient to cause diarrheal disease symptoms in humans (Lund and Granum, 1996), and so the presence of homologues to NHE in other pathogenic bacteria may suggest an important role of these proteins for the pathogenicity of these bacteria.

The putative tripartite PFTs from *Aeromonas hydrophila* were chosen for study in this thesis, based on the pathogenicity of the organism, its emergence since 2009 as a major fish pathogen, and the availability of genome DNA for the pathogenic strain AL09-71 (Pridgeon et al., 2014). As the Bacillus protein systems did not prove amenable to further study, attention shifted to the tripartite system identified in Aeromonas hydrophila (AhlA, AhlB, and AhlC). An introduction to the *Aeromonas hydrophila* organism follows, and studies on these three proteins are described in the rest of this thesis.

4.7 Aeromonas hydrophila

The *Aeromonas* genus is a large group of Gram negative, ubiquitous, water-borne bacteria, containing the *Aeromonas hydrophila* species. *Aeromonas* was originally part of the *Vibrionaceae* family, before being reclassified into its own family *Aeromonadaceae* (Colwell et al., 1986; Janda and Abbott, 2010b). Many species of Aeromonas are pathogenic, which were

historically broadly grouped into two groups, the non-motile and psychrophilic species that includes *A. salmonicida,* and the motile and mesophilic species that include *A. hydrophila, A. caviae, and A. sobria.* More recently, DNA hybridisation groups (HG) have expanded the number of different groups and associated phenospecies, diversifying the categorisation of the organism. HG1 is associated with *Aeromonas hydrophila,* a motile mesophilic species that is a known pathogen of Carp, Catfish, and Salmon, as well as being one of the two species found in 85 % of clinical isolates of *Aeromonas* from diseased humans (Ko et al., 2000). The disease caused by A. *hydrophila* is hugely variant and a result of a large number of extracellular toxins, including the PFT aerolysin discussed briefly throughout Chapter one. The hybridisation group classification system of typing Aeromonas has now also been superseded, but is relevant for the level of discussion in this chapter.

4.7.1 Diseases caused by *A. hydrophila*

Aeromonas hydrophila are predominantly pathogenic to aquatic poikilothermic animals such as fish and amphibians, causing fatal haemorrhagic septicaemia in a number of fish species (Thune et al., 1993)(Figure 4.20). Since 2009, motile *aeromonas* septicaemia (MAS) has cost West Alabama and East Mississippi U.S. catfish aquaculture an estimated $60-70 million, and epidemics are also now emerging in other areas in the U.S., and in Asia (Hossain et al., 2014; Zheng et al., 2012). *Aeromonas* has also been associated with opportunistic human disease, causing extraintestinal septicaemia, gastroenteritis, and more lethal wound infections, pneumonia, and meningitis in immunocompromised patients (Gold and Salit, 1993; Kelly, 1993; Ko et al., 2000).

A number of studies (Pang et al., 2015; Pridgeon and Klesius, 2011a; Pridgeon et al., 2013) have identified a subset of *A. hydrophila* strains which are much more virulent than other strains, and have probed the genetic elements conserved in these strains , as well as the toxins present in the secretome of these bacteria. The most virulent strain of *A. hydrophila* isolated from the infected channel Catfish was identified as AL09-71.

4.7.2 Aeromonas hydrophila AL09-71

Aeromonas hydrophila AL09-71 was first isolated from an outbreak of haemorrhagic septicaemia in channel catfish in 2009, which caused an estimated $3 million damage in 48

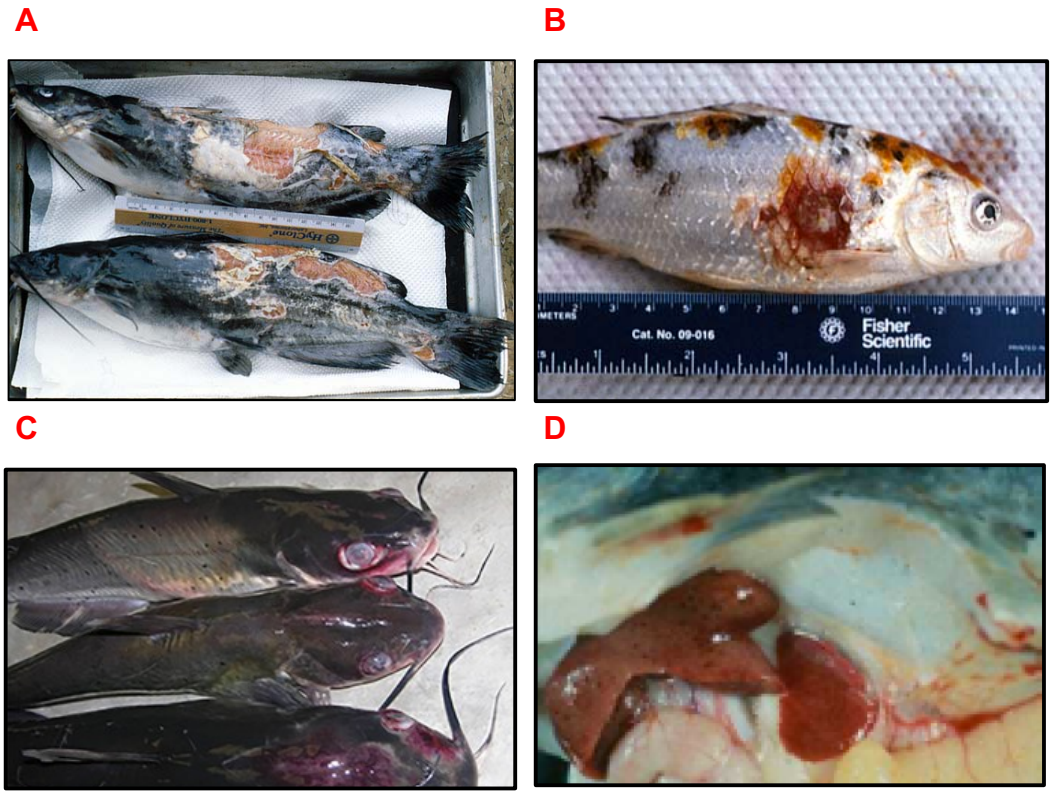


Figure 4.20. The pathology of Motile Aeromonas Septicaemia (MAS) in fish. *Aeromonas hydrophila* and related species are able to cause tissue damage and haemorrhaging in channel catfish (**A**), and rainbow trout (**B**). (**C**) and (**D**) The pathology of the disease also involves damage and haemorrhaging to soft tissues such as the eyes and liver.

catfish farms in Alabama. It was one of three highly virulent strains isolated and characterised, namely AL09-71, -72, and -73, and AL09-71 was found to be slightly more virulent than the other two strains (Pridgeon and Klesius, 2011a). Molecular characterisation of AL09-71 strain of *Aeromonas* has identified that 23 extracellular proteins are found in the toxic fractions of the extracellular products of *A. hydrophila* AL09-71. However, it was unknown how many other virulence related genes existed in the genome of the highly virulent strains that may have been contributing to the phenotype of the MAS disease. Eventually the complete genome was sequenced and made available in 2014 (Pridgeon et al., 2014), which has allowed genomic studies of the virulence factors not identified in the molecular characterisation of the extracellular fraction.

Although often thought of as a secondary pathogen, these new highly virulent strains have been shown to become primary pathogens capable of causing huge agricultural damages, and so study into potential virulence factors of these more virulent strains are of great interest.

Chapter 5: characterising the tripartite pore forming toxin properties of AHL

5.1 Introduction

In order to determine whether the three AHL proteins identified by Bioinformatical studies were indeed a member of the tripartite ClyA-like PFT family, experiments were designed to test their haemolytic potential. Each protein was cloned into an expression vector, overexpressed, purified, and assayed for its ability to cause cell lysis. This chapter outlines the experiments involved in producing large quantities of pure protein for crystallisation, and the assays carried out to show that AHL was indeed a tripartite pore-forming toxin.

5.2 Cloning results

5.2.1 Cloning of *ahlA*, *ahlB*, and *ahlC* and cloning into pET21a

Genomic DNA from pathogenic *Aeromonas hydrophila* strain AL09-71 was kindly provided by Dr. Julia Pridgeon (United States Department of Agriculture), and primers were designed to clone each tripartite toxin gene, each with a C-terminal 6x His-tag (Table 5.1). Each gene was cloned from the genomic template DNA, with the addition of sticky end restriction sites to clone the resulting gene into the pET21a expression vector. Restriction enzymes were chosen so that the sites were lacking in the gene sequence itself, and such that the 6x his-tag included at the 3' end of the vector would be included in the final gene product with minimum linker. Therefore, the restriction enzymes NdeI and XhoI were used in cloning *ahlB* and *ahlC* as they are positioned at the extremes of the polyclonal region in pET21a. However, as the *ahlA* gene contained a XhoI site, the XhoI enzyme was substituted with the next available restriction site in the polyclonal region in pET21a, HindIII, for *ahlA* cloning.

PCR cloning of the *ahl* genes from genomic DNA was performed using the Biomix red 2x master mix as described in 3.2.1. A two stage PCR protocol was followed, as described in 3.2.1 to account for pre- and post-restriction site incorporation melting temperatures. *ahlA* and *ahlC* were cloned with pre- temperatures of 57 °C and post- temperatures of 65 °C, and *ahlB* was cloned using pre- temperature of 58 °C and post- temperature of 63 °C. The *ahlB* reaction was also supplemented with 3 % (v/v) DMSO in order to increase PCR product yield.

Primer	Primer sequence	Melting temperature (ºC)
AhlA his forward	**GGC GCT AGG TAC TA**C ATA TGA CCA CGA TCG CCA CCC TGG	69.2
AhlA his reverse	**ATC TA**A AGC TTT AG TGG TGG TGG TGG TGG TGA GCG TCT GGC AGG ATG CC	71.3
AhlA link his forward	GGC GGA CAC CAC CAC CAC CAC CAC	68.3
AhlA link his reverse	TCC GCC AGC GTC TGG CAG GAT GCC	69.6
AhlB his forward	**CAG CC**C ATA TGA CCA ACG CAA CAA CCA TCA CCA TGG ACC AG	68.8
AhlB his reverse	**AT**C TCG AGG GCG GCC AGG CGC G	70.8
AhlC his forward	**GGC GCT AGG TAC TA**C ATA TGA GCA ACG GCA TTC TTT CC	65.6
AhlC his reverse	**TAA TC**C TCG AGT AG TGG TGG TGG TGG TGG GAA GCG TCC ACC TGC	71.4

Table 5.1. DNA primers designed to clone each gene from genomic DNA, and to insert a linker region in AhlA. Each primer used in cloning experiments is shown above, with the identity of the primer, the DNA sequence, and the calculated melted temperature for stage 1 and stage 2 of the PCR reaction. The SDM primers to insert a linker into AhlA are highlighted in purple. Each DNA sequence is shown 5' to 3' left to right, with the underlined region highlighting the restriction site. The bold DNA is any nonsense DNA incorporated into the primer to adjust melting temperatures, to ensure primer partners (highlighted in matching colours) were within 5 ºC of each other in both stage 1 and stage 2.

The reaction products from each reaction were run on a 1 % (w/v) agarose gel to determine the quality of the resulting PCR amplification and purity of the correct PCR products (Figure 5.1). As Biomix Red contains its own dye, no loading buffer was required for electrophoresis. The PCR for *ahlB* and *ahlC* resulted in a single band of correct size, and so *ahlB* and *ahlC* PCR products were purified by PCR clean up (3.2.3). However, *ahlA* amplification produced a number of different sized bands, including one band of the correct size (Figure 5.1). The *ahlA* PCR products were thus purified by gel extraction, such that only the band of correct size was carried through to the restriction digestion and ligation stages.

Restriction digest of *ahlB* was carried out as two single restriction digest reactions. *ahlB* was digested first with NdeI for 2 hours, as described in 3.2.6, and then heat inactivated at 80 °C for 20 minutes. The partially restricted *ahlB* gene was then digested with XhoI for 2 hours followed by a PCR cleanup. The *ahlA* and *ahlC* restriction digests were carried out using a double digest protocol, as described in 3.2.6, with NdeI and HindIII, or NdeI and XhoI respectively. The pET21a plasmid was also restricted with complementary enzymes to produce two unique sticky ends.

Ligations were carried out for 2 hours at room temperature, containing the ligation mix described in 3.2.8, and an 4:1 insert:plasmid molar ratio for each reaction.

Once ligated, the ligation mix was used directly in transformation of chemically competent DH5α silver efficiency cells. Colonies were checked for the presence of an insert of the correct size by colony PCR, as described in 3.2.9. Those colonies containing an insert of the correct size, as determined by agarose gel electrophoresis, were sequenced using T7 promoter and terminator primers (GATC genomics). Using this protocol, the *ahlA* and *ahlC* genes were successfully cloned without any mutations, whereas for the *ahlB* gene, a single mutation was present which substituted a guanosine for an adenine, and resulted in a methionine change to isoleucine at residue 336 (Figure 5.2). As this mutation was present in all sequenced *ahlB* gene inserts, it is possible that it is present in the genome itself, and arose from a sequencing error of the genome. This construct of AhlB was taken forward with the *ahlA* and *ahlC* clones for overexpression trials.

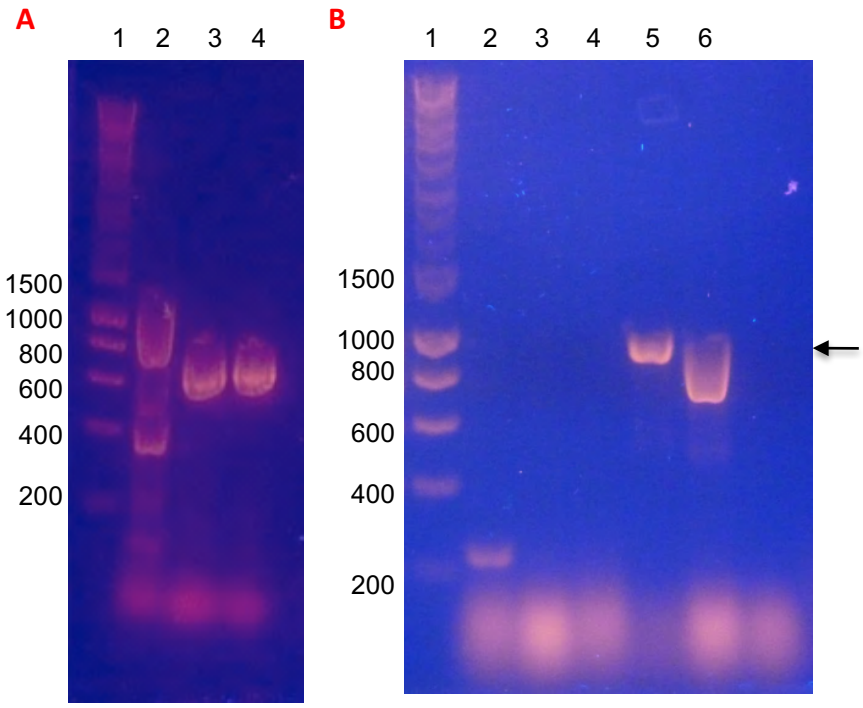
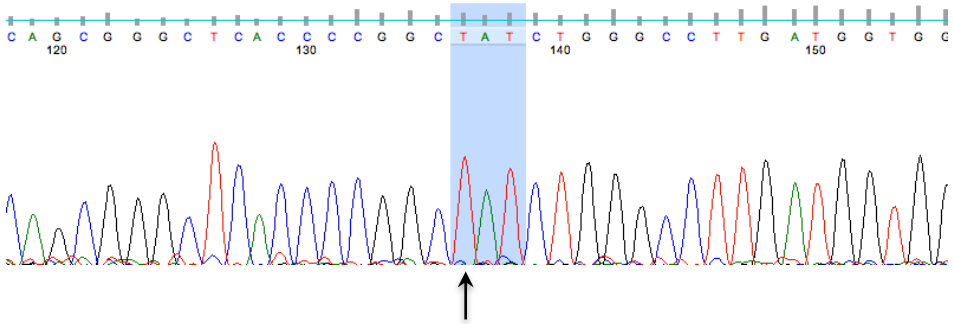


Figure 5.1. PCR products from cloning *ahlA, ahlB, and ahlC* run on an agarose gel. (**A**) 1 % Agarose gel run at 100 V for 40 minutes containing the PCR products generated during PCR of AhlA c-term link his and AhlC c-term his. Lane 1 contains 2 µL Hyperladder 1kb ladder. Lane 2 contains 2 µL PCR product obtained from 50 µL Biomix PCR of AhlA c-term link his from Al09-71 gDNA template. Lane 3 and 4 contain 2 µL of PCR product obtained from 50 µL Biomix PCR of AhlC and AhlC c-term his respectively. (**B**) 1 % Agarose gel run at 100 V for 40 minutes, containing PCR products generated after PCR clean-up or gel extraction. Lanes 1 contains 2 µL Hyperladder 1kb ladder. Lane 5 contains 2 µL of the 30 µL AhlA c-term link his generated after gel extraction of the correct band from gel A. Lane 6 contains 2 µL from the 30 µL PCR clean-up of AhlC c-term his.

A



B

Range 1: 1 to 359 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
709 bits(1831)	0.0	Compositional matrix adjust.	358/359(99%)	359/359(100%)	0/359(0%)

```
Query  1    MTNATTITMDQGMANQASQAMQIQTYCNSVKQQVPVDFSQFPNLKDNQTQINQGLDLAKG  60
            MTNATTITMDQGMANQASQAMQIQTYCNSVKQQVPVDFSQFPNLKDNQTQINQGLDLAKG
Sbjct  1    MTNATTITMDQGMANQASQAMQIQTYCNSVKQQVPVDFSQFPNLKDNQTQINQGLDLAKG  60

Query  61   HADLYLNTIQPQIITNISNISNYFALQNAIPAVLPPGSTKAQWLRQLSVIKEQATEYQRL  120
            HADLYLNTIQPQIITNISNISNYFALQNAIPAVLPPGSTKAQWLRQLSVIKEQATEYQRL
Sbjct  61   HADLYLNTIQPQIITNISNISNYFALQNAIPAVLPPGSTKAQWLRQLSVIKEQATEYQRL  120

Query  121  SSDTRLVIVNLNNNLITDSSNFQGIVVNLNSKVQGDNGVLAQLNGDIDKVNAAIDGAIAG  180
            SSDTRLVIVNLNNNLITDSSNFQGIVVNLNSKVQGDNGVLAQLNGDIDKVNAAIDGAIAG
Sbjct  121  SSDTRLVIVNLNNNLITDSSNFQGIVVNLNSKVQGDNGVLAQLNGDIDKVNAAIDGAIAG  180

Query  181  IVAGGLLVIGGAFVTAIGAVADFVTAGTSTPVVIGGVAMMVAGAGGITAGAIVLHNSLGA  240
            IVAGGLLVIGGAFVTAIGAVADFVTAGTSTPVVIGGVAMMVAGAGGITAGAIVLHNSLGA
Sbjct  181  IVAGGLLVIGGAFVTAIGAVADFVTAGTSTPVVIGGVAMMVAGAGGITAGAIVLHNSLGA  240

Query  241  RQDLYQKRSSLNSEVLIATQIGNGYKGLQVQAQNAVTAATQMSNAWDSLTSDLGSLITDL  300
            RQDLYQKRSSLNSEVLIATQIGNGYKGLQVQAQNAVTAATQMSNAWDSLTSDLGSLITDL
Sbjct  241  RQDLYQKRSSLNSEVLIATQIGNGYKGLQVQAQNAVTAATQMSNAWDSLTSDLGSLITDL  300

Query  301  DKGITSGDDIRQLWLTAADTTVKTVLTDVTTIKAQIAGVSPLQVPQTDTIANFVARLAA  359
            DKGITSGDDIRQLWLTAADTTVKTVLTDVTTIKAQ+AGVSPLQVPQTDTIANFVARLAA
Sbjct  301  DKGITSGDDIRQLWLTAADTTVKTVLTDVTTIKAQMAGVSPLQVPQTDTIANFVARLAA  359
```

Figure 5.2. Single point mutation in the sequence of AhlB leading to mutation of Met336 to isoleucine. (**A**) DNA sequencing chromatogram of the reverse compliment strand of the gene encoding AhlB, cloned into pET21a. The chromatogram shows that a mutation of guanosine to adenine is present in the gene sequence. (**B**) A single point mutation in the DNA sequence of AhlB from G to A leads to the mutation of Met336 to isoleucine. DNA chromatogram generated in FinchTV, and amino acid sequence blasted against the protein accession AHX33180.1 (AhlB).

Only his-tagged variants of each gene were produced, namely *ahlA* with linker and C-terminal his-tag, *ahlB* with C-terminal his-tag, and *ahlC* with C-terminal his-tag. Therefore, each protein will for brevity be referred to as the protein name only, i.e. AhlA, AhlB and AhlC.

5.3 Protein overexpression results

5.3.1 Expression trials

Each AHL containing plasmid was transformed into separate aliquots of the BL21 expression strain of *E. coli,* and used in expression trials to ascertain the amount of soluble protein expressed. Expression trials were carried out for each AhlA, AhlB, and AhlC as described in 3.4.1.

Each starter culture was grown to an OD_{600} of 0.6 prior to induction using IPTG, and then grown for a further 4 hours for 37 °C trials, or overnight for 16 °C and 25 °C overexpression trials. After this expression period, the cells were lysed, and the cell debris and cell free extract were analysed by SDS gel electrophoresis. The temperature was varied from 16 °C to 37 °C to optimise soluble protein expression levels for each protein, and the soluble and insoluble fractions were run on an SDS PAGE gel to ascertain the best conditions to use in large scale overexpression cultures. For AhlC, the largest amount of soluble protein was produced at 25 °C, which was of sufficient yields to use in large-scale over-expressions (Figure 5.3).

In the case of AhlB, apparently only insoluble AhlB was present on the gel (Figure 5.4), and so a small scale Histrap purification, described in 3.5.1, was carried out to enrich the soluble fraction of any soluble AhlB, so it could be seen on an SDS PAGE gel. SDS PAGE gels of the crude HisTrap purification of AhlB showed that a high molecular weight band was present for AhlB when the sample was boiled at 95 °C for 10 minutes, and that by omitting the boiling step in SDS PAGE sample preparation, the high molecular weight band disappeared, and the band of the correct molecular weight increases in amount of protein (Figure 5.5). Membrane proteins have been reported to form large oligomeric complexes in SDS when boiled, due to an association of the SDS detergent with hydrophobic regions of the protein, and so non-complete unfolding of the protein occurs (Rath and Glibowicka, 2009). In the light of observation during crude purification, the overexpression SDS page gel was rerun with the samples both boiled and unboiled which showed that indeed AhlB was expressed in a soluble form during

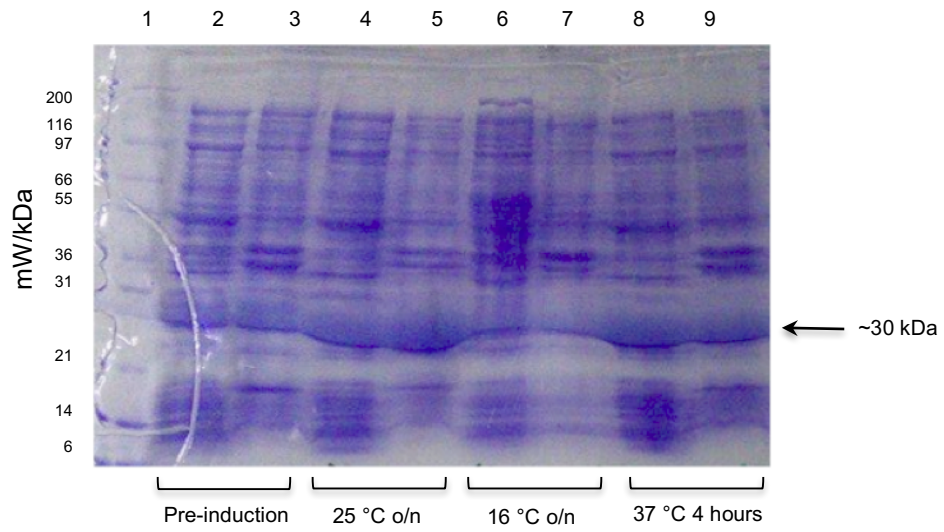


Figure 5.3. SDS PAGE of the AhlC overexpression trial. An SDS PAGE gel showing the small-scale overexpression trials undertaken to find conditions yielding soluble AhlC expression. Lane 1 contains Mark12 protein ladder, and molecular weights for each band are given. Lanes 2 and 3 contain the pre-induction soluble and insoluble fractions respectively. Lanes 4 and 5 contain the soluble and insoluble fractions respectively for the overexpression of AhlC at 25 ºC overnight. Lanes 6 and 7 contain the soluble and insoluble fractions respectively for the overexpression of AhlC at 16 ºC overnight. Finally, lanes 8 and 9 contain the soluble and insoluble fractions respectively for the overexpression of AhlC at 37 ºC for 4 hours. A black arrow marks the approximate correct molecular weight for AhlC. AhlC can be seen slightly in the pre-induction sample, as well as the smeary bands in the 25 ºC and 37 ºC overexpression conditions.

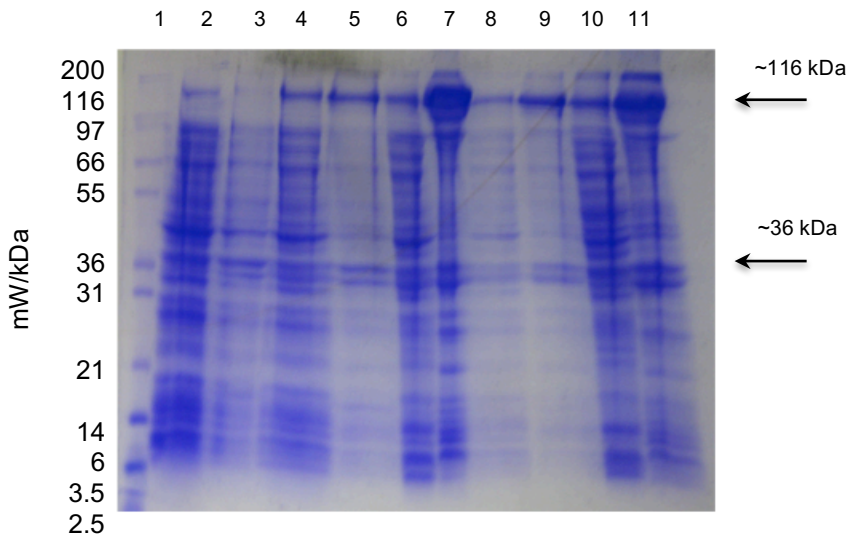


Figure 5.4. SDS PAGE of the AhlB overexpression trial. The initial SDS PAGE of the AhlB overexpression trial stained with Coomassie blue, and with Mark12 ladder in lane 1. Lanes 2 and 3 show the pre-induction soluble and insoluble fractions respectively. Lanes 4, 5, 6, and 7 show alternating soluble and insoluble fractions of overexpression at 16 °C for 4 hours and overnight respectively. Lanes 8, 9, 10, and 11 show the same layout as for 16 °C but for expressions carried out at 25 °C. Two black arrows show the apparent molecular weight of AhlB at ~116 kDa, as well as the predicted molecular weight of ~36 kDa.

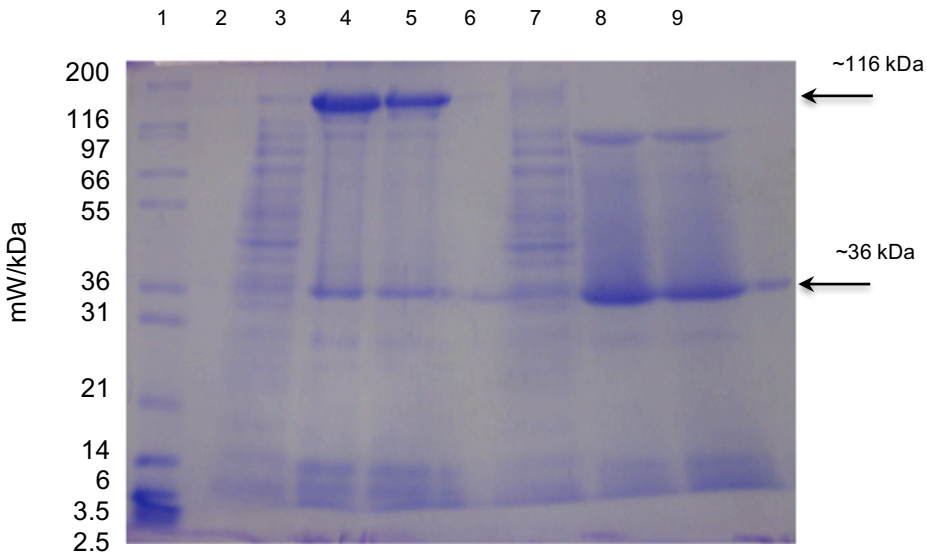


Figure 5.5. SDS PAGE following the Crude Ni-HP purification of AhlB soluble fraction. An SDS PAGE stained with Coomassie Blue, with Mark12 ladder in lane 1. Lane 3 shows the flowthrough from the AhlB 25 °C expression soluble fraction applied to a 1 ml Ni-HP column. Lanes 4 and 5 show the eluted fractions when 1 M imidazole was applied to the Ni-HP column. The samples in lane 3, 4, and 5 have been boiled for 10 minutes at 95 °C. Lanes 7, 8, and 9 show the identical samples to lanes 3, 4, and 5, but in this case, the boiling stage was omitted. Two arrows indicate the presence of AhlB at ~116 kDa in the boiled samples, and at ~36 kDa for the unboiled samples.

overexpression (Figure 5.6). The greatest yield of AhlB was expressed at 25 °C, and that it was present on the gel at ~37 kDa when the sample was not boiled.

For AhlA overexpression trials, all temperatures initially produced completely insoluble protein. In the structure of NheA (Ganash et al., 2013), the C-terminus of NheA packs against the head domain to form an antiparallel sheet. Sequence analysis shows that this situation may also occur in AhlA, and that such packing may be hindered by the presence of the C-terminal his-tag. Therefore, a 4x Gly linker was engineered between the C-terminus residues of AhlA and the 6x Histag using the SDM protocol described in 3.2.10, to minimise any unfavourable interactions. SDM was carried out using the protocol, and successful clones were confirmed by sequencing. Overexpression of this modified construct following the same overexpression trial protocol as for AhlA yielded soluble expression of AhlA with the glycine linker at 16 °C (Figure 5.7).

5.3.2 Large scale overexpression of each AHL protein

As each protein had been successfully expressed on a small 50 ml scale, yielding soluble protein, these same conditions were used in scaling the protocol to large (>1l) expressions. Colonies from freshly transformed (the day before expression) BL21, containing the relevant plasmid were grown for 8 hours in 5 ml LB containing ampicillin, and then used to inoculate overnight 50 ml cultures, which were grown at 37 °C 220 rpm overnight.

Each overexpression was carried out in 2l batches. 2 l LB media in 4 x 500 ml flasks were supplemented with ampicillin and inoculated with 10 ml overnight culture, and grown at 37 °C 220 rpm to an optical density at 600 nm of 0.8. Each culture was then induced with 1 mM IPTG and grown overnight at the optimal expression temperature determined from the small-scale trials. Cells were harvested by centrifugation at 18,000 xg in a benchtop centrifuge for 15 minutes. Pellets were resuspended in a minimum volume of LB media, and aliquoted into 50 ml falcon tubes for storage at -80 °C until required.

5.4 Purification results

5.4.1 Purification of AhlA link His

2 g of cell paste was prepared as in 4.5. Cell free extract containing AhlA was then purified by HisTrap Nickel affinity chromatography followed by gel filtration. The cell free extract was loaded through a 5 ml HisTrap column in buffer A (50 mM Tris pH 8 and 0.5 M NaCl), and then

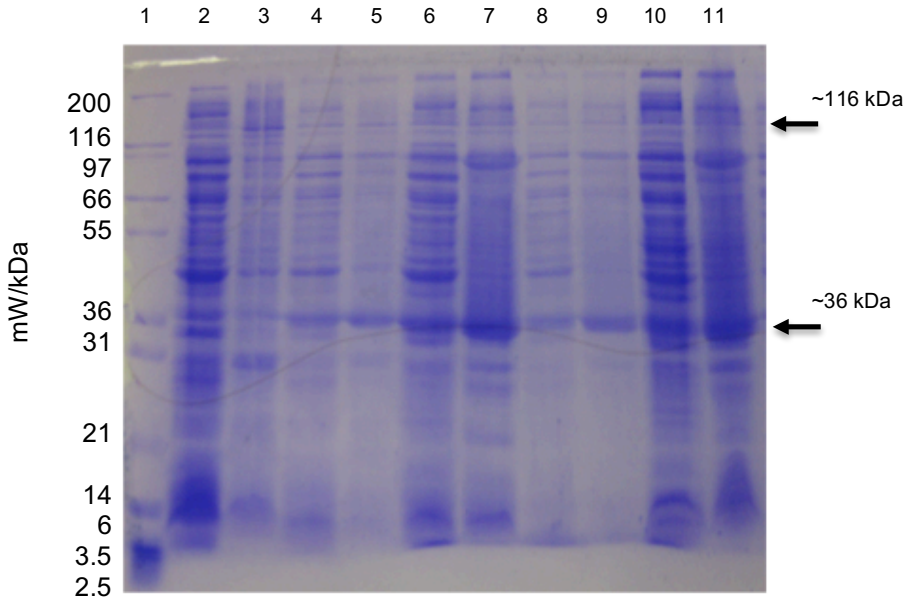


Figure 5.6. SDS PAGE of AhlB overexpression where each sample was unboiled. The SDS PAGE gel is laid out identically to the SDS PAGE gel in Figure 5.4, the difference between the two being that samples in the above gel were not boiled prior to loading. Two arrows show the apparent molecular weight of AhlB when samples are boiled at ~200 kDa, and the new position of AhlB on the gel when samples are unboiled, at ~36 kDa. The overexpression gel shows that AhlB is expressed well at 25 °C with an overnight induction period (lane 7). Two arrows show that the bands present at ~116 kDa has now disappeared, and a band is present at ~36 kDa.

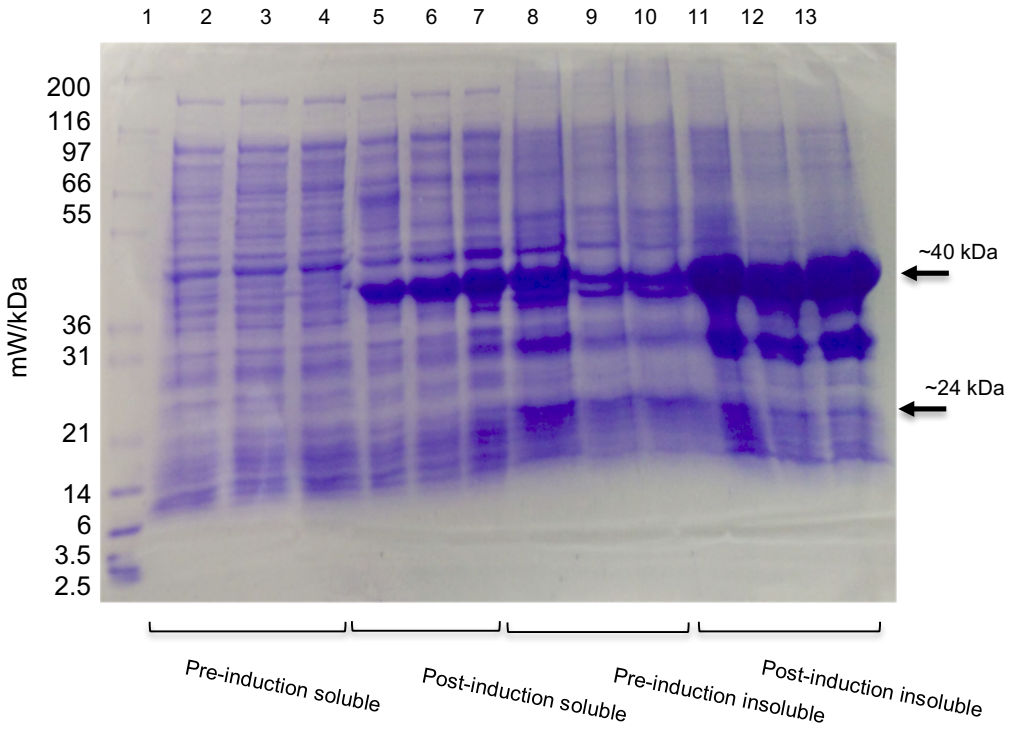


Figure 5.7. AhlA link His overexpression. SDS PAGE gel showing the overexpression trials carried out to determine conditions that yielded soluble AhlA. Lane 1 contains Mark12 ladder and molecular weights of markers are given. Lanes 2, 3, and 4 contain soluble pre-induction samples for 16 °C, 25 °C, and 37 °C overexpressions respectively. Lanes 5, 6, and 7 contain soluble post-expression samples for 16 °C, 25 °C, and 37 °C overexpressions respectively. Lanes 8, 9, and 10 contain the insoluble pre-induction samples for the 16 °C, 25 °C, and 37 °C overexpressions respectively. Lanes 11, 12, and 13 contain the insoluble post-induction samples for the 16 °C, 25 °C, and 37 °C overexpressions respectively. Finally, two arrows indicate the molecular weight of AhlA at ~40 kDa, and the lack of a prominent band at ~24 kDa which then is present during purification.

washed in the same buffer to elute unbound protein. The bound protein was then eluted with a gradient of 0 – 100 % buffer B (50 mM Tris pH 8, 0.5 M NaCl, and 0.5 M imidazole), checking each fraction for protein content by UV absorption.

Fractions from the peaks in the UV trace from the nickel affinity chromatography (Figure 5.8) were run on an SDS PAGE gel, and fractions 20 and 21 were determined to contain AhlA. Fractions 20 and 21 were combined and concentrated by centrifugation to 2 ml in a Vivaspin concentrator with a 10,000 mW cut-off at 13,000 xg. The 2ml sample was then subjected to gel filtration using a Superdex200 pg gel filtration column and run as described in 3.5.2 AhlA eluted from the gel filtration column at 75 ml in a single peak, corresponding to a molecular weight of 50 kDa (Figure 5.8). There was also a peak at 87.5 ml (~23 kDa), however on SDS PAGE gel, this peak did not agree with AhlA subunit molecular weight. The SDS PAGE gel shows that the peak at 87.5 ml contained a band of molecular weight ~20 kDa. The AhlA peak which eluted at 75 ml also contains the smaller molecular weight band of ~20 kDa, and so the origin of this band may be partial proteolysis of AhlA (Figure 5.8).

In order to determine the identity of the SDS PAGE gel bands, each was sent for mass spectrometry analysis, carried out by Simon Thorpe in the University of Sheffield Chemistry Department. The results showed that two major species existed in the 75 ml fraction, with molecular weights of 40734.06 Da and 23551.57 Da (Figure 5.9). These masses were analysed using the Expasy FindPept software (Gattiker et al., 2002) to determine amino acid sequences that could constitute these molecular weights. The 40734.06 Da mass corresponded to the sequence of AhlA with the C-terminus HisTag, but excluding the first 4 residues (MTTI) with a Δmass of 3.6 Da between the expected and the measured molecular weight (Figure 5.9). The 23551.57 Da fragment corresponded to residues 155-377 with a Δmass of -0.210 assuming an average isotopic resolution. This fragment lacked the C-terminus his-tag, and so must have been generated after the Ni^{2+} affinity chromatography step, and thus may have been due to a protease contamination present during purification. This observation is also supported by the lack of a 23 kDa band in the cell free extract. As the corresponding residues 1 – 155 were not present on the gel or in the mass spectrometry analysis, this suggested that the protein had been digested to many smaller fragments.

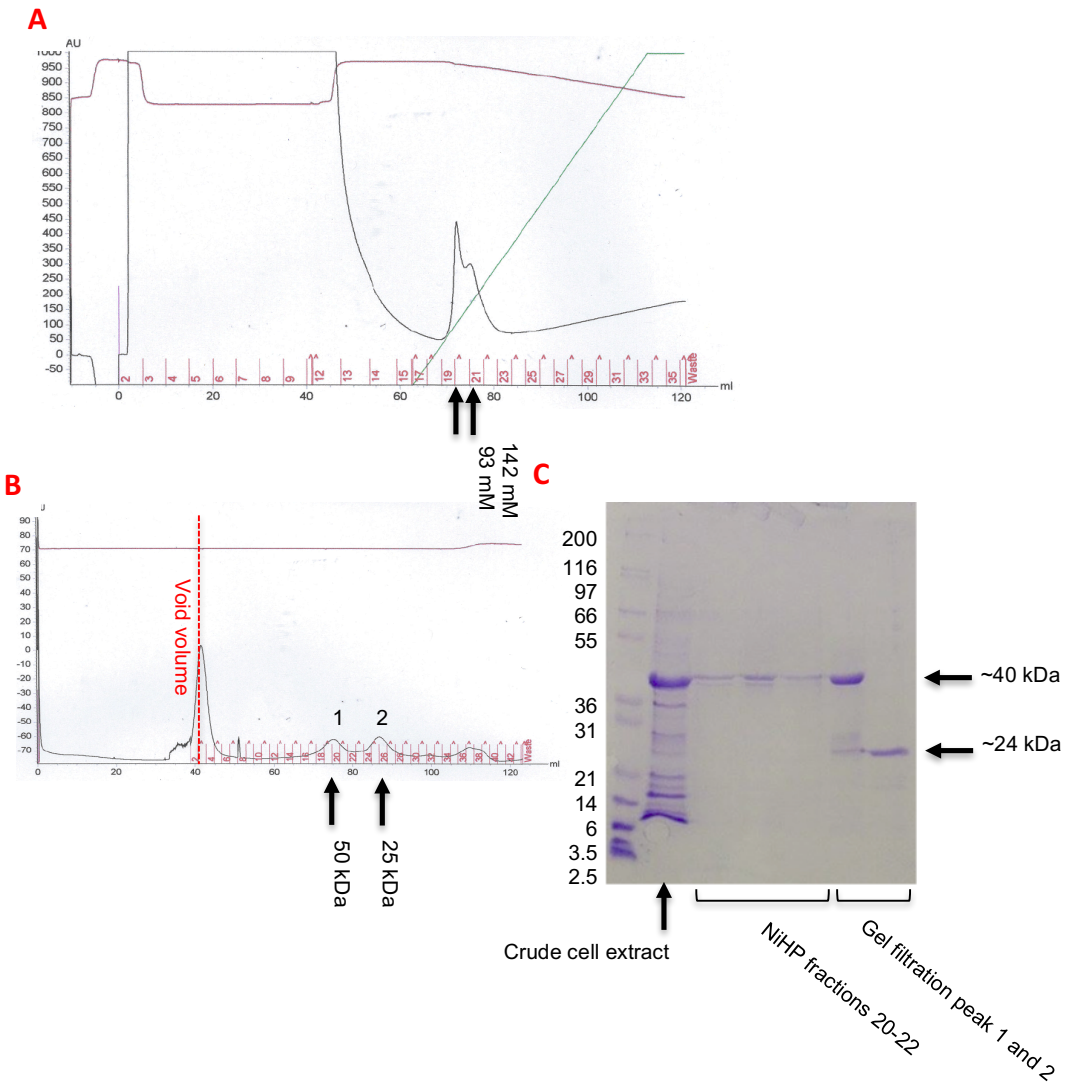
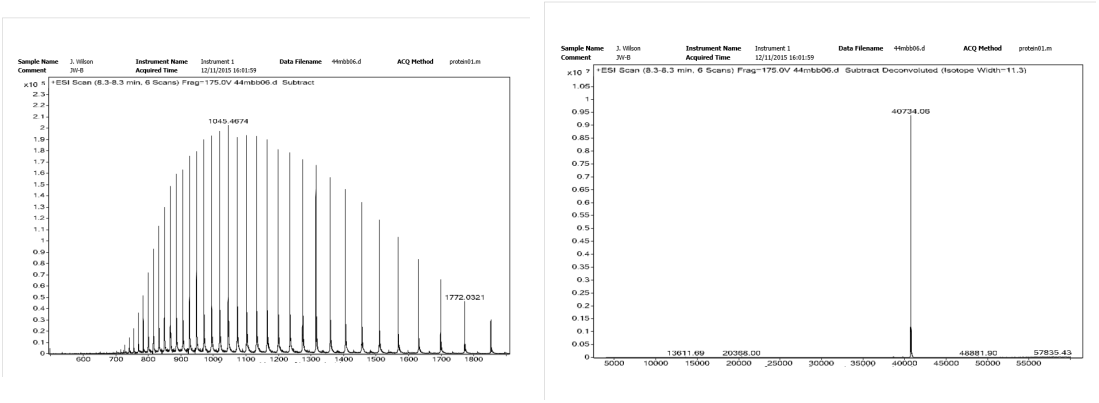
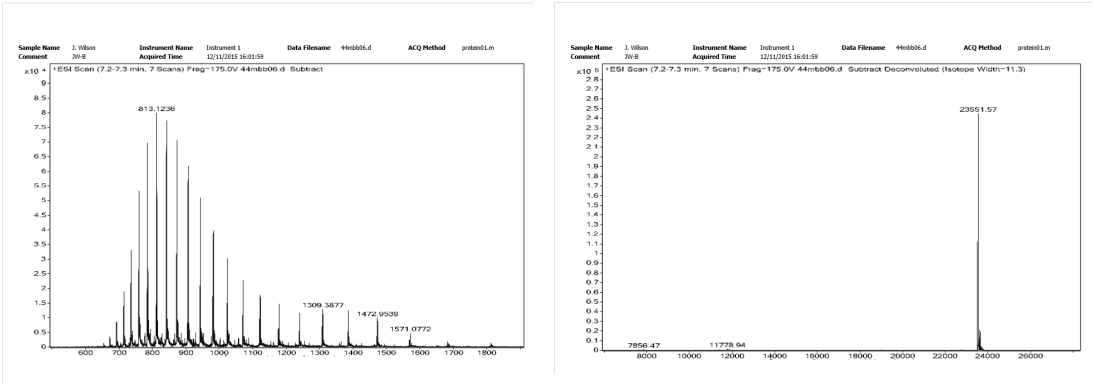


Figure 5.8. Purification of AhlA link His. (**A**) The UV trace chromatogram for the initial purification of AhlA by NiHP. Once the sample was applied, a gradient of 0 % to 100 % 0.5 M imidazole was applied, and two overlapping peaks are present in the UV trace at 93 mM and 142 mM imidazole. (**B**) The UV trace chromatogram for the gel filtration purification of AhlA NiHP fractions 20 to 22. A large peak is present at the void volume (red dashed line), as well as two other major peaks at 75 ml and 87.5 ml, marked as peak 1 and peak 2 respectively. (**C**) SDS PAGE analysis of the purification of AhlA. Lane 1 contains Mark12 ladder and molecular weights of protein bands are given. Lane 2 contains the crude cell extract with AhlA present at ~40 kDa. Fractions 20 to 22 of the NiHP imidazole elution are shown in lanes 3 to 5, corresponding to the elution of AhlA at 93 to 142 mM imidazole. An extra smaller band is present at ~39 kDa in lane 4. Finally, lanes 6 and 7 contain gel filtration peaks 1 and 2 respectively, with a band present at ~40 kDa, and a band present at ~20 kDa.

A



B



C

Mass spec mW	Sequence mW	Δmass (Daltons)
40734.080	40738.727	4.647
23551.570	23553.263	1.692

40734 Da proteolysis site

```
      10         20
MTTIATLDLG DQGLLAAYLS
```

23551.57 Da predicted proteolysis site

```
160         170              370         380
 EQ DLMADVKRQT EIIDGLN  ..... LP DALEGGGGHH HHHH
```

Figure 5.9. Mass spectrometry analysis of peak 1 and peak 2 from gel filtration of AhlA. (**A**) The mass/charge peaks for the species in mass spectrometry of peak 1 (left), and the deconvoluted mass peak (right). (**B**) The mass-to-charge peaks for the species in the mass spectrometry of peak 2 (left), and the deconvoluted mass peak (right). (**C**) The mass spectrometry molecular weights and the results from analyses by the FindPept server. The difference between the measured mass and the closest resulting fragment are shown, with the start and end amino acids in the sequence shown in red.

To attempt to prevent proteolysis of AhlA, purification was repeated, but with the addition of protease inhibitor cOmplete EDTA-free (Sigma) in the lysis buffer (50 mM Tris pH 8 and 0.5 M NaCl). Purification using these protease inhibitors yielded a similar sample of AhlA, in which the 23 kDa contaminant eluted from gel filtration at 87.5 ml. cOmplete EDTA-free is unable to inhibit metalloproteases due to the lack of EDTA, but was selected as to not interfere with the Ni^{2+} affinity chromatography column.

As there was no improvement of the AhlA purification using cOmplete EDTA, the protease inhibitor was omitted from further purifications. Therefore, degradation of AhlA was taken into account in subsequent assays and studies, and thus such assay results are qualitative in nature.

5.4.2 Purification of AhlB His

30 ml AhlB CFE was prepared from 2 g cell paste as described in 4.5.1. CFE containing AhlB was initially purified identically to the protocol used to purify AhlA, using nickel affinity chromatography followed by gel filtration, checking each fraction for protein content by UV absorption (Figure 5.10)

Fractions from peaks in the UV trace were analysed by SDS PAGE electrophoresis, which showed that fractions 19 to 21 contained AhlB (Figure 5.11). Fractions 19 to 21 were pooled and concentrated to 2 ml in a Vivaspin 10,000 MWCO column. The protein was applied to a 2 ml loop, and gel filtration was carried out as described in 3.5.2.

AhlB eluted as a sharp peak, with a slight skew towards earlier elution, at 79 ml, corresponding to a molecular weights of approximately 35 kDa (Figure 5.10). The elongated shape of proteins in this family could affect the elution profile on gel filtration, and so the 35 kDa was interpreted as either a monomer or a dimer.

As the purity of AhlB after nickel affinity chromatography was determined to be upwards of 90 % pure, gel filtration was omitted from subsequent purifications. The gel filtration step was substituted with a buffer exchange step, exchanging the protein into 50 mM tris pH 8 + 0.5 M NaCl, using a Zebaspin desalting column.

5.4.3 Purification of AhlC His

20 ml AhlC CFE was generated from 1 g cell paste as described in 4.5.1. The yield of AhlC during purification was an order of magnitude higher than that of AhlA and AhlB, and so less cell

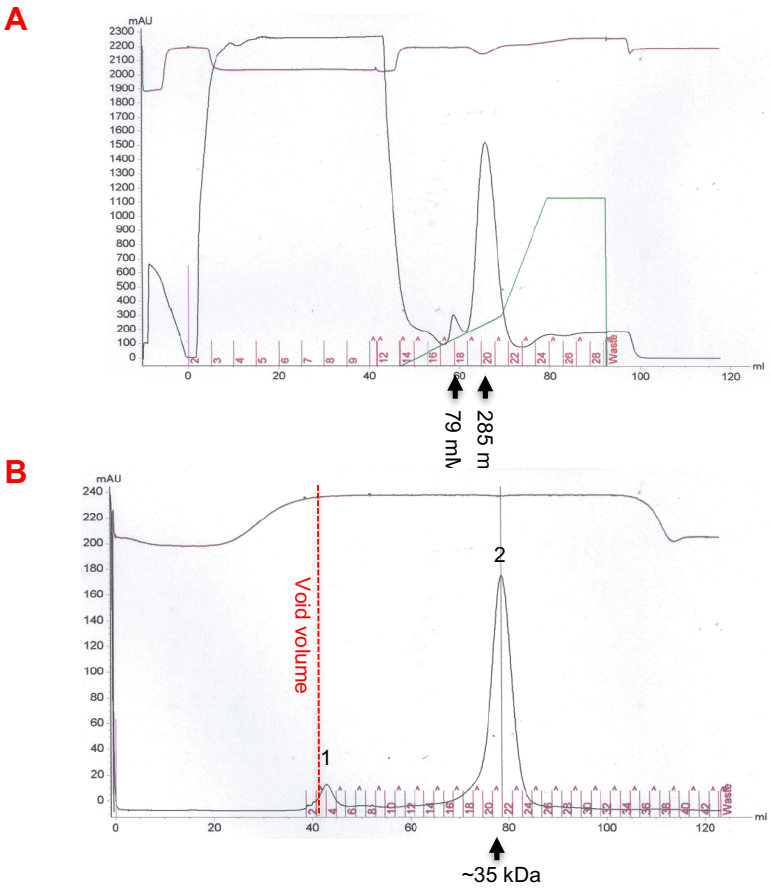


Figure 5.10. Chromatograms following the NiHP and gel filtration stages of AhlB purification. (**A**) A chromatogram for the NiHP chromatography step in the purification of AhlB. The black line shows the UV trace for OD_{280}, and the green line shows the percentage of imidazole in the gradient elution. Two peaks are highlighted in the imidazole gradient elution of AhlB, one at 79 mM imidazole, and one at 285 mM. The first peak contains the contaminants, whereas peak 2 contains AhlB. (**B**) A chromatogram for the gel filtration purification and analysis of AhlB. Fractions from peak 2 were applied to a Superdex 200 pg, and the black line shows the UV trace as the applied protein eluted. A small peak (peak 1) is observed at the void volume (red dashed line). A sharp second peak (peak 2) is seen at 75 ml, corresponding to a molecular weight of ~39 kDa.

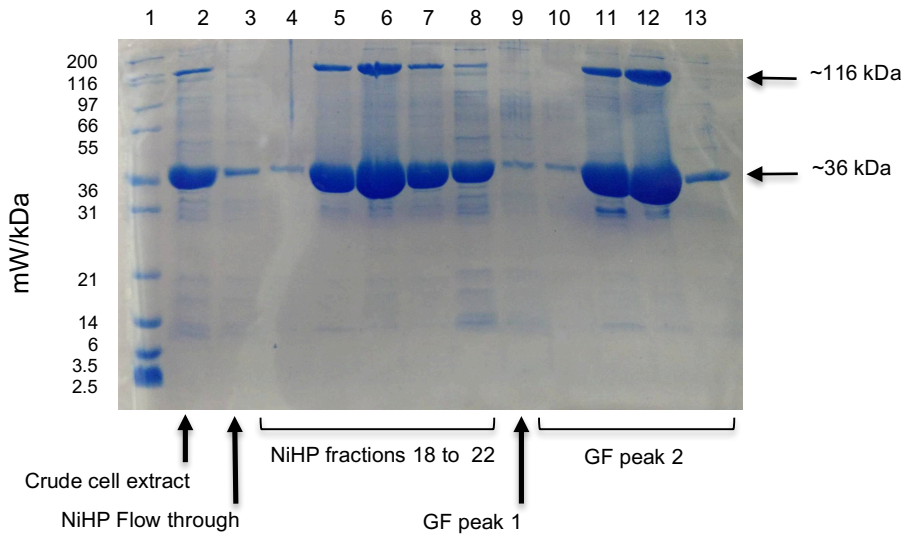


Figure 5.11. SDS PAGE analysis of the purification of AhlB. An SDS PAGE gel stained with instant blue, analysing the NiHP and gel filtration steps in the purification of AhlB. Lane 1 contains Mark12 ladder and molecular weights of each band are given. Lane 2 contains the soluble crude cell extract after sonication of the cell paste. Lane 3 shows the 0 % imidazole flowthough from the NiHP chromatography. Lanes 4 to 8 show the eluted protein from the UV peak corresponding to fractions 18 to 22. Lane 9 shows the peak present in the void volume elution from the gel filtration column. Finally, lanes 10 to 13 contain the fractions from gel filtration peak 2, present at 75 ml. Two arrows mark the high molecular weight SDS resistant oligomers of AhlB, and the monomeric protein present at ~36 kDa. Samples were not boiled in preparation for this gel.

paste was used in each experiment. The CFE containing AhlC was initially purified identically to AhlA and AhlB (Figure 5.12).

SDS PAGE analysis of the peaks in HisTrap UV trace (Figure 5.13) showed that fractions 17 and 18 contained AhlC. Fractions 17 and 18 were pooled and concentrated in a 10,000 MWCO Vivaspin concentrator, to a final volume of 2 ml. The sample was subjected to gel filtration on a Superdex200 pg column and run as described in 3.5.2.

AhlC eluted in a single peak at 69 ml corresponding to an approximate molecular weight of 80 kDa (Figure 5.12). This molecular weight corresponds to a tetramer, although as with AhlB, the elongated shape of this family of proteins may alter the elution volume and thus the molecular weight estimation. The peak was analysed by SDS PAGE and fractions 16 to 18 containing AhlC were combined (Figure 5.13). Due to the low extinction coefficient of AhlC, protein concentration was carried out on the combined fractions, concentrating the protein repeatedly in a Vivaspin 10,000 kDa MWCO concentrator until an absorbance could be taken. The protein was concentrated to 40 mg/ml, in 500 µL, with a total yield per gram cell paste of 20 mg.

As with AhlB, analytical gel filtration was used to try to ascertain oligomeric state of AhlC, but as the protein was pure enough after the first step of affinity chromatography, gel filtration was subsequently omitted from further purifications. AhlC was buffer exchanged in the case of omitting gel filtration, as described in 4.5.3.

5.5 Haemolytic activity results

In order to determine whether the pure samples of AhlA, AhlB, and AhlC were able to act as pore forming toxins, haemolytic assays were carried out on each AHL protein in isolation, and in combination, against Sheep and Horse erythrocytes, using a variation of the method of Rowe and Welch (Rowe and Welch, 1994). In these analyses, individual AHL components, or mixtures of AHL components were incubated with an isotonic solution of 1 % (w/v) erythrocytes for one hour at 37 °C. The cell debris was pelleted by centrifugation and the soluble haem was measured by absorbance at 542 nm. A positive control consisted of erythrocytes suspended in distilled water, such that they ruptured by osmotic pressure. Centrifugation of the resulting suspension yielded no pellet, and so no erythrocytes remained intact.

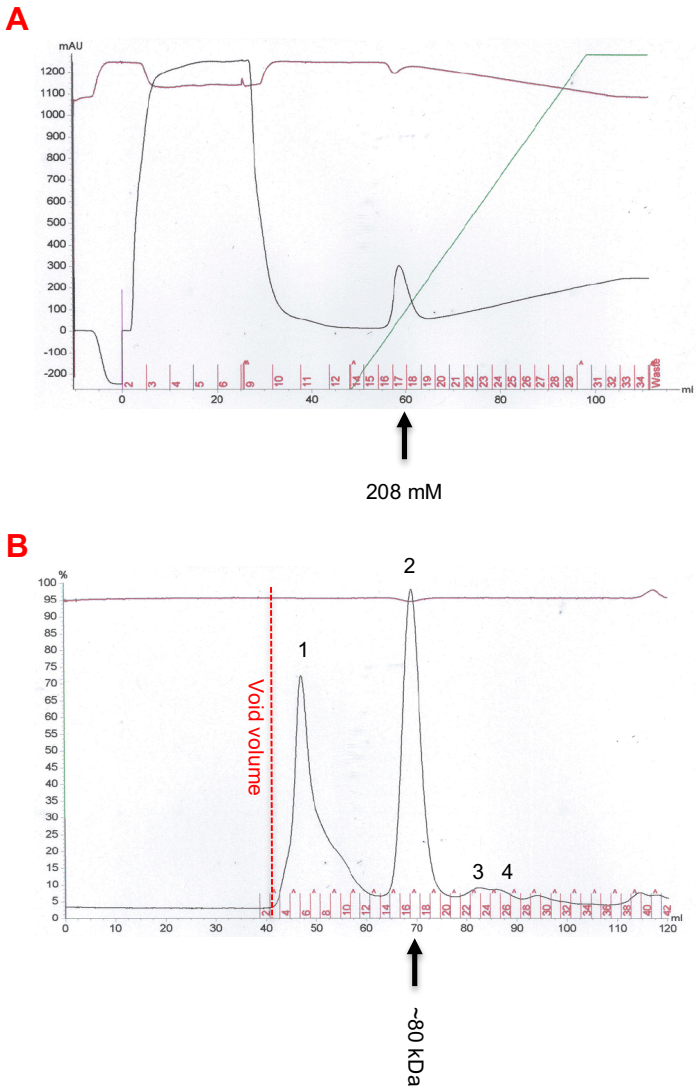


Figure 5.12. Chromatograms following the NiHP and gel filtration stages of AhlC purification. (**A**) NiHP chromatogram with the UV trace (black line) and percentage imidazole (green line) shown. AhlC elutes from the NiHP column at 208 mM imidazole (arrow). (**B**) Gel filtration chromatogram showing the UV trace (black line). The void volume is shown by the red dashed line. A sharp peak present close to the void volume indicates a large aggregate. There are two other peaks of interest present at 69 ml and 84 ml (arrows).

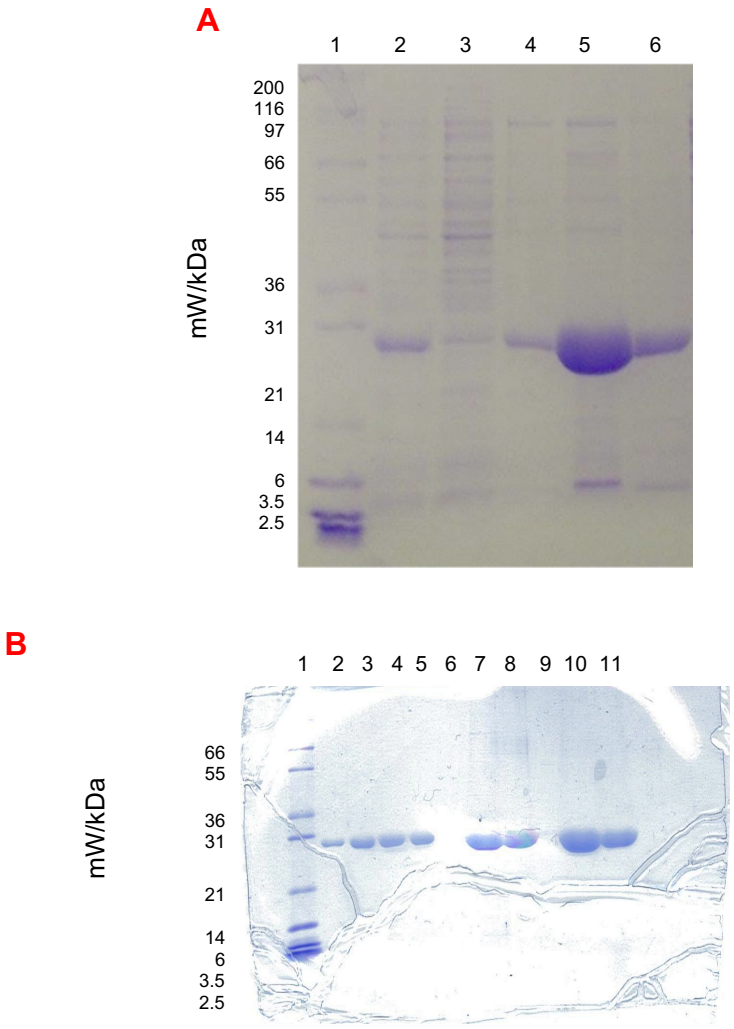


Figure 5.13. SDS PAGE analysis of the purification of AhlC. (**A**) SDS PAGE gel showing the crude cell extract of AhlC, and fractions collected during NiHP purification. Lane 1 contains Mark12 ladder, and molecular weights of each band are given. Lane 2 contains the cell free extract, and lane 3 contains the flowthrough from the NiHP column in low imidazole. Lanes 4 to 6 contain the NiHP elution fractions 16 to 18. (**B**) SDS PAGE analysis of peaks eluted from gel filtration purification. Lane 1 again contains Mark12 ladder, and molecular weights of each band are given. Lane 2-4 contains the fractions 4, 6, 8, and 10 from peak 1. Lane 7 and 8 contain fractions from peak 2, and lanes 10 and 11 contain fractions from peak 3 and 4.

All results are subsequently reported as a percentage of erythrocyte lysis relative to the positive control. Experiments were repeated at least twice, and percentage lysis is shown in Figure 5.14.

5.5.1 Minimum requirement for Ahl haemolytic activity and requirements for maximum lysis

Each of the AHL proteins in isolation exhibited no haemolytic activity with either Horse or Sheep erythrocytes using 10 µg of each protein in the assay (Figure 5.14). Lysis was also not observed when AhlA and AhlC were incubated together, and only a negligible amount of lysis was seen when AhlA and AhlB were incubated together (Figure 5.14).

When equal amounts of AhlB and AhlC were incubated together with either horse or sheep erythrocytes, lysis did occur, at 10 % of the positive control. In this experiment, flocculation of the remaining blood cells occurred, which could be seen as cells rapidly falling out of solution to the bottom of the Epindorf over a few seconds (Figure 5.14). However, maximal lysis occurred when all three AHL components were incubated in equal amounts with erythrocytes (68.7 % and 57.3 % lysis for horse and sheep erythrocytes respectively).

In previous studies with NHE, maximal lysis was observed when NheA, NheB, and NheC were incubated at a 10:10:1 molar ratio respectively, with inhibition of lysis observed as the NheC concentration was increased towards that of NheA/NheB (Lindbäck et al., 2010). To determine if a similar phenomenon is observed with AHL, the relative amount of AhlC to AhlA and AhlB was varied. In these experiments, maximal lysis was observed when AhlA, AhlB, and AhlC were incubated at a 1:1:1 ratio, respectively. It should be noted, however, that as the purification of AhlA showed that some degradation was occurring, these results should perhaps not be regarded as quantitative. However, they do show at least that there is not any inhibition when AhlC is increased in concentration relative to AhlA and AhlB.

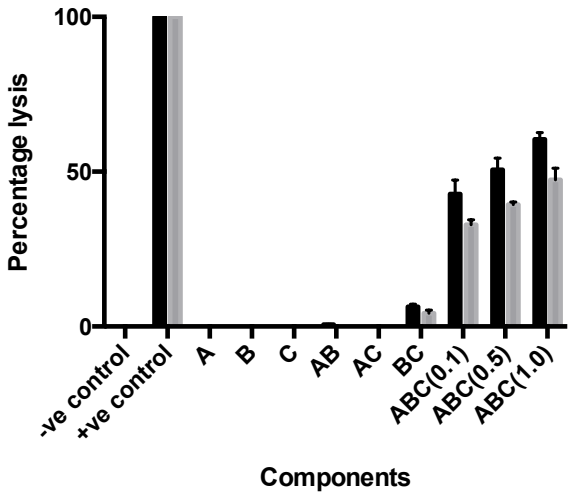
5.5.2 Determining the binding order requirement for AHL haemolytic activity

In the NHE system, haemolytic activity has been shown to be dependent on the order of incubation of each protein. In these experiments, NheB incubated with cells prior to adding NheC decreased lysis compared with that observed when both NheB and NheC were incubated simultaneously (Lindbäck et al., 2010). Similar experiments were carried out with AHL to determine if the incubation order of the AHL components affected haemolysis. For these experiments only AhlB and AhlC were used in order to minimise any experimental differences arising from degraded AhlA. 10 µg of AhlB or AhlC was first incubated with erythrocytes for one

A

	Percentage lysis of erythrocytes			
	Horse		Sheep	
-ve control	0.00	0.00	0.00	0.00
+ve control	100.00	100.00	100.00	100.00
A	0.00	0.00	0.00	0.00
B	0.00	0.00	0.54	0.00
C	0.00	0.00	0.00	0.00
AB	0.00	5.84	0.54	3.53
AC	0.00	0.00	0.00	0.00
BC	12.8	11.80	9.42	12.72
ABC(0.1)*	30.00	39.61	48.35	31.76
ABC(0.5)*	57.50	58.44	63.74	38.82
ABC(1.0)*	68.12	69.20	69.80	44.71

B



C

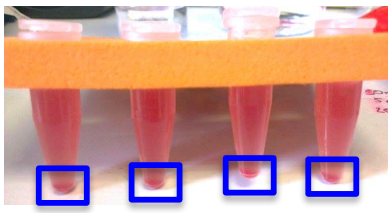


Figure 5.14. Hemolysis assay results from AhlA, AhlB, and AhlC incubated in different combinations. (**A**) A table showing the percentage lysis of Horse (column 2 and 3) and Sheep (column 4 and 5) erythrocytes during experiments. Each component was incubated at 10 µg with 1 % (w/v) erythrocytes. Each percentage is calculated by normalising the absorbance at OD_{542} to the negative control. (**B**) A graph of table A, with mean bars and SD error bars. The lysis of Horse and Sheep erythrocytes are shown as black and grey bars respectively. (**C**) Sedimentation of cell debris observed when Horse erythrocytes were lysed by combination of AhlBC (left) or all AHL components with AhlC in ratios of 0.1, 0.5, and 1.0 respectively. Bright red sedimentation fell to the bottom of the tube over second timescales.

hour, before adding 10 µg of the other component and incubating at 37 °C for a further hour. Haemolytic activity was assayed as previously described, and quantified as percentage lysis (Figure 5.15). The results show that the order of incubation of AhlB or AhlC does not affect the haemolytic activity of the proteins by an appreciable amount, as in each case lysis is ~ 10 %.

5.5.3 AhlB and AhlC do not form a stable complex in solution

The binding order inhibition of cell lysis, seen for NHE, has been suggested to be associated with the formation of a complex between NheB and NheC in solution at a 1:1 ratio, with this complex inhibitory to cell lysis (Lindbäck et al., 2010). To determine if AhlB and AhlC also formed such a complex in solution, gel filtration assays were undertaken. 8 mg of purified AhlB was incubated with 12 mg of purified AhlC, in 0.1 M HEPES pH 7.5 and 0.1 M NaCl at room temperature for 1 hour, and then run on a superdex200 column pre-equilibrated with 0.1 M NaCl and 50 mM HEPES pH 7.5. The UV trace for the chromatography is shown in Figure 5.16. There are two major peaks, present at ~69 ml and ~76 ml, in the UV trace of the gel filtration, as well as a small peak at the void volume. SDS PAGE analysis of each peak in the UV absorbance trace shows that AhlB does not form a complex with AhlC in solution, and that each protein elutes from the gel filtration column in generally the same position as when run on gel filtration separately (Figure 5.16). This lack of complex formation between AhlB and AhlC in solution under the explored conditions may explain why the binding order of AhlB and AhlC is not important in erythrocyte lysis, and why at ratios of AhlA to AhlB to AhlC of 1:1:1, no inhibition of erythrocyte lysis is observed. Taken together, this would suggest that a different pore formation pathway exists for AHL, compared to that for NHE.

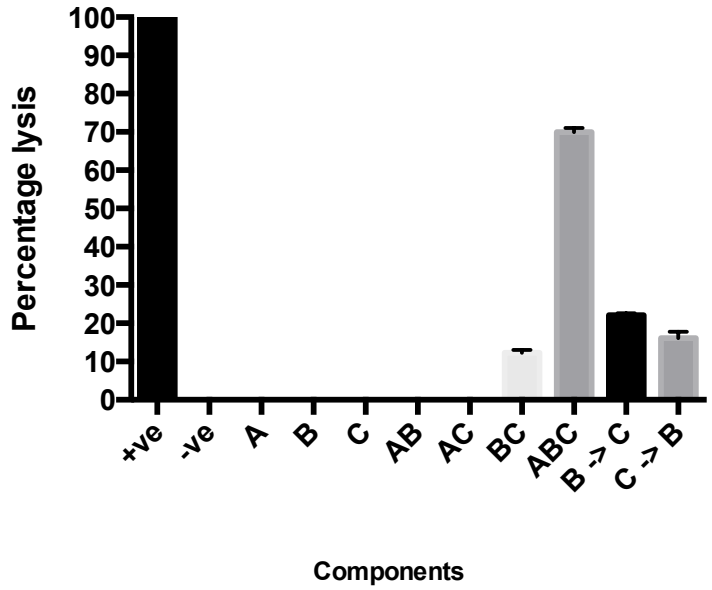
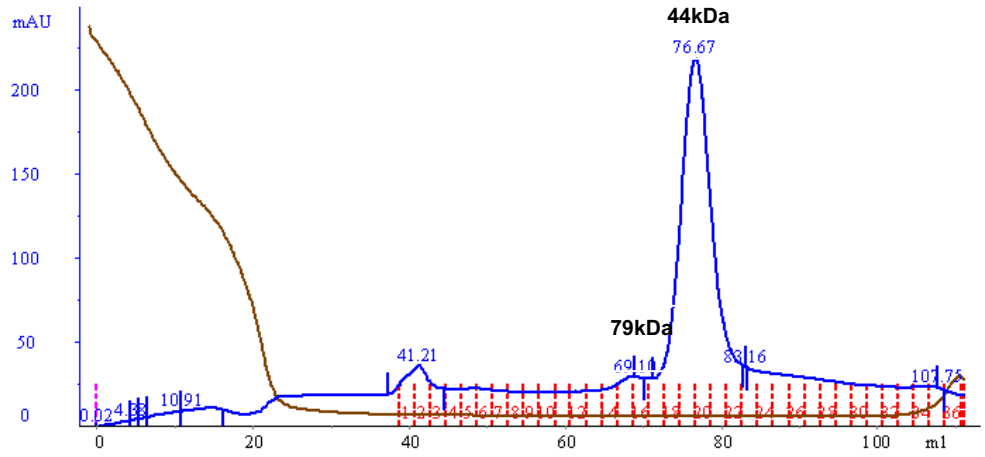
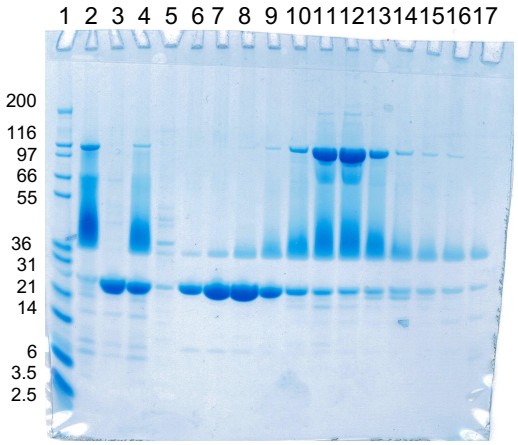


Figure 5.15. Hemolysis assay results from AhlB and AhlC incubated in different orders. A graph of percentage lysis of a suspension of 1 % (w/v) Horse erythrocytes when incubated with 1 µg of each AHL component as shown on the horizontal axis. Each percentage lysis is normalised with the positive control as 100 % lysis, and the negative control as 0 % lysis. The experiments labelled as B -> C and C -> B are the order dependent experiments in which the -> means followed by e.g. B followed by C.

A



B



1. Mark12
2. AhlB
3. AhlC
4. AhlB + AhlC GF loading
5. Fraction 2
6. Fraction 14
7. Fraction 15
8. Fraction 16
9. Fraction 17
10. Fraction 18
11. Fraction 19
12. Fraction 20
13. Fraction 21
14. Fraction 22
15. Fraction 24
16. Fraction 26
17. Fraction 28

Figure 5.16. Gel filtration of AhlB incubated with AhlC in order to determine if both associate with one another in solution. 8 mg of AhlB was incubated with 12 mg of AhlC and then run on gel filtration, looking for a larger molecular weight oligomer to form. (**A**) Gel filtration chromatogram with UV trace (blue line). The main elution peak was found at 76 ml corresponding to a molecular weight of ~44 kDa. A second small peak is present at 69 ml, corresponding to ~79 kDa. (**B**) SDS PAGE gel with lane contents shown (right). Lane 1 contains Mark12 ladder, and molecular weights of each marker are given.

5.6 Summary

Successful cloning, over-expression and purification of each AHL component allowed the haemolytic properties of the proteins to be assayed. During purification, it was observed that AhlA was possibly being degraded, and so could not be used in quantitive assays, as the amount of full length AhlA could not be determined accurately. However, AhlA was confirmed to be active in haemolytic assays, as maximum lysis of erythrocytes was observed only when AhlA, AhlB, and AhlC were incubated together. Lysis of erythrocytes increased from 10 % when AhlB and AhlC were incubated with erythrocytes, to 80 % when all three components were incubated together, showing that a synergistic effect occurs when all three components are incubated together with erythrocytes. The properties of AHL lysis differed greatly from NHE, with different minimum requirements for lysis, no independent binding order requirement and no association of the B and C components seen in solution.

The results presented here confirm that the AHL protein has haemolytic activity, and that the expanded family of proteins identified from AHL constitute new additions to the tripartite pore forming toxin system. Further experiments are, however, required to determine the role of each protein in lysis, and to confirm that indeed the mode of lysis is through pore formation. In addition, changes to the soluble expression of AhlA that would eliminate the proteolytic degradation of the protein would be very valuable. This would enable quantitive and kinetic studies of haemolysis, which would help in understanding the mechanism of pore formation, along with the other complimentary techniques that are discussed in the remainder of the thesis.

Chapter 6: Structural and functional studies on AhlC

The third component of the AHL tripartite pore system is AhlC. Like AhlB, AhlC has no haemolytic activity by itself, but is able to lyse erythrocytes when combined with AhlB, or in combination with AhlA and AhlB, as discussed in Chapter 5. AhlC is the smallest of the three AHL proteins (266 residues and 28.3 kDa), and has a similar hydropathy plot to NheC, in that the hydrophobic stretch is present, but shorter than that found in the B components. This suggests that AhlC may have a similar membrane binding function as NheC, as has been determined by previous studies (Lindbäck et al., 2004). Structural studies were carried out to determine the structure of AhlC, and to characterise its role in pore formation and cell lysis.

6.1 Crystallisation and structural determination of AhlC

6.1.1 Protein preparation for crystallisation

AhlC was purified as described in 5.4.3 and prepared for crystallisation by buffer exchange into a low salt buffer (50 mM tris pH 8 + 50 mM NaCl). The protein was concentrated to a final concentration of 40 mg/ml by centrifugation in a Vivaspin 10,000 MWCO concentrator. Protein concentration was determined by absorbance at 280 nm using a spectrophotometer, and a theoretical absorption for AhlC of 0.053 for a 1 mg/ml solution, as determined by ProtParam (Gasteiger et al., 2005). Protein was prepared just prior to crystallisation, and leftover protein was stored at 4 °C until required (<2 weeks) for optimisation experiments.

6.1.2 Selenomethionine derivitised AhlC production

In order to prepare seleno-methionine labelled protein, AhlC was expressed in 500 ml minimal media after growing to an OD_{600} of 0.7 in LB and washing the cells to remove residual LB media. Cells were grown to an OD_{600} of 0.7 at 37°C, and once induced with 1 mM IPTG, were incubated overnight at 20 °C. Cells were harvested at 18,000 xg for 15 minutes and stored at -80 °C until required. SeMet AhlC was purified identically to native AhlC, and SDS PAGE analysis showed that the amount of the protein expression was comparable to native AhlC (Figure 6.1). SeMet AhlC was prepared for crystallisation identically to the native protein as described in 5.1.

6.1.3 Initial screening of AhIC crystallisation conditions

Initial crystallisation trials were carried out using commercial 96 well screens from Molecular Dimensions and Nextal, namely the PACT, JCSG+, Classics, Proplex, and PEGs suites. 96 well sitting drop vapour diffusion screens were generated using a Matrix Hydra II PlusOne crystallization robot. Trays was sealed and centrifuged, and then stored at 17°C.

6.1.4 Initial crystal hits from screening

Crystal screens were checked manually by microscope daily for 3 days and then every week to observe crystal growth and drop precipitation. Roughly half of the drops had precipitated after 3 days suggesting the protein concentration was suitable for optimal crystal growth.

Crystals grew in two related conditions, the first (JCSG+ B8) contained 0.2 M $MgCl_2$, 0.1 M Tris pH 7.0, and 10 % (w/v) PEG 10,000, whereas the second (PACT D7) contained 0.2 M NaCl, 0.1 M Tris pH 8.0, and 20 % (w/v) PEG 6000. Figure 6.2 shows crystals in each of the conditions. Crystals from these initial hits were of sufficient size and quality to be screened for diffraction quality and to collect data, prior to optimisation of the crystal growth.

6.1.5 Native data collection from initial crystal hits

Crystals from initial screen hits PACT D7 and JCSG+ B8 were mounted into litholoops of a suitable size (Molecular Dimensions) under a microscope and transferred to liquid nitrogen. Loops were stored in pucks and sent to the Diamond synchrotron in Oxford for diffraction tests and data collection.

Test data was collected on each crystal to determine crystal quality (diffraction resolution, Laue group, and mosaicity). For test data collection, the crystal was exposed at 5 positions 45° apart with a 0.2° oscillation at each position. Mosflm (Diamond FastDP auto processing)(Battye et al., 2011) was used to index reflections on each image, and an initial Laue group and data collection strategy was estimated. Data from the JCSG+ B8 sample showed that the crystals were predicted to belong to Laue group P3, and to diffract to 2.5 Å.

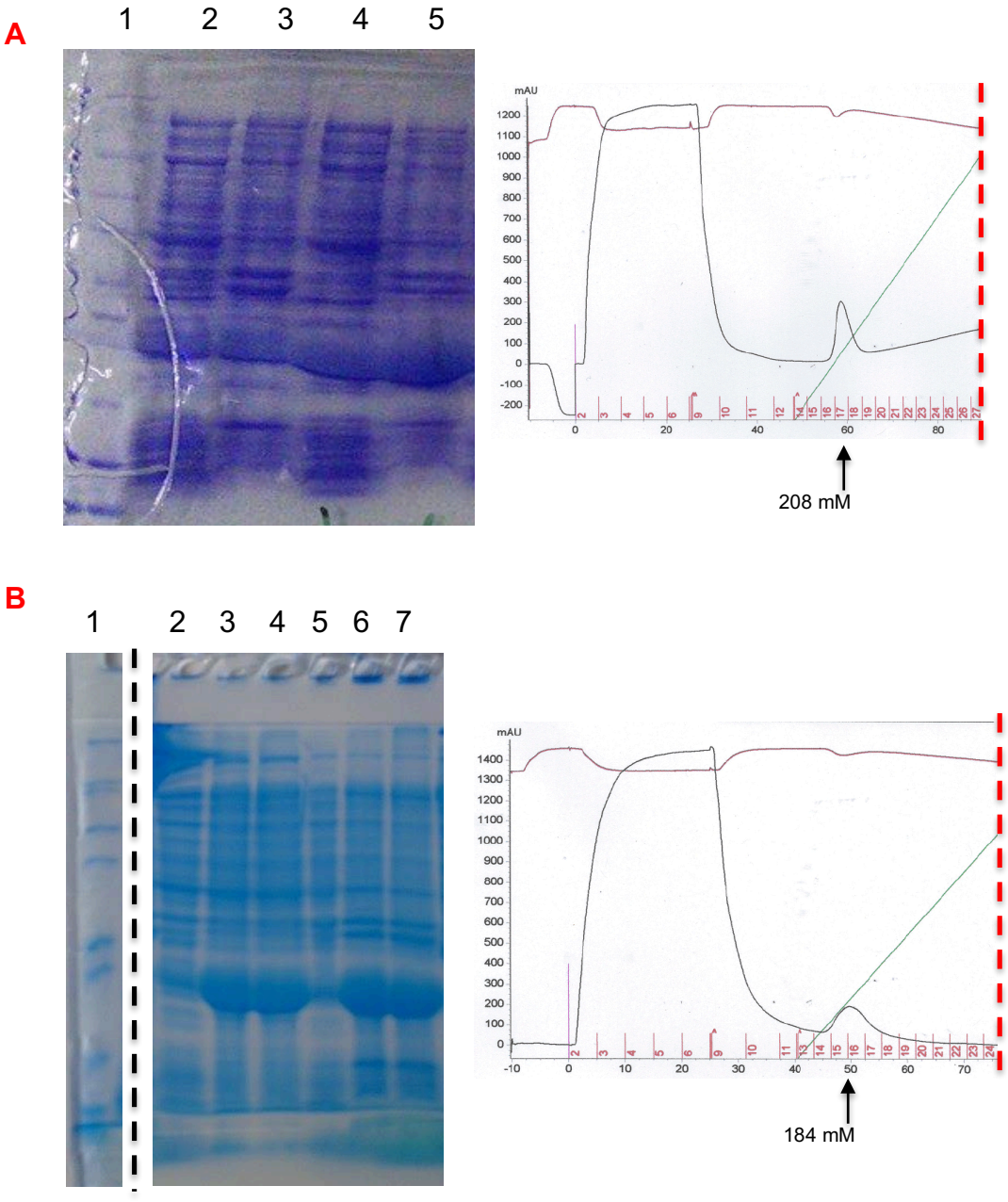
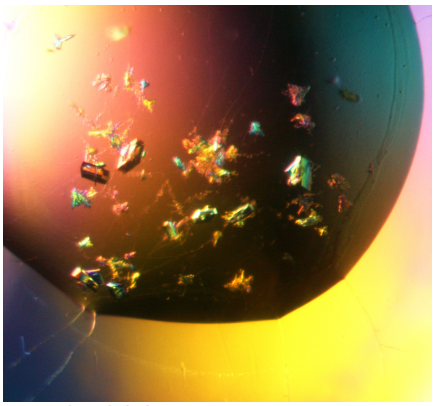
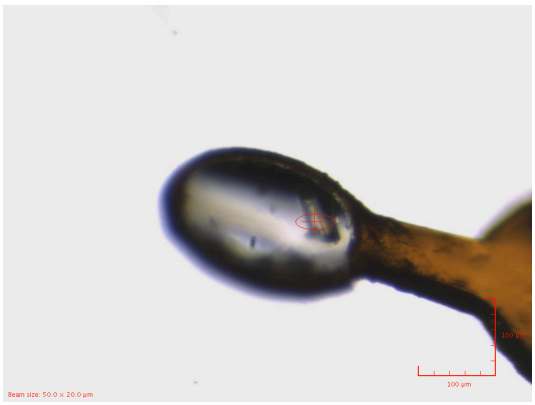


Figure 6.1. Comparison of AhlC overpression for native and SeMet labelled AhlC protein. (**A**) SDS PAGE gel showing the overexpression of native AhlC (left), and NiHP chromatography UV trace, showing AhlC elutes at 208 mM Imidazole (right). Lane 1 contains (**B**) SDS PAGE gel showing the overexpression of SeMet derivitised AhlC (left), and NiHP chromatography UV trace (right), showing AhlC SeMet elutes at 184 mM Imidazole.

A

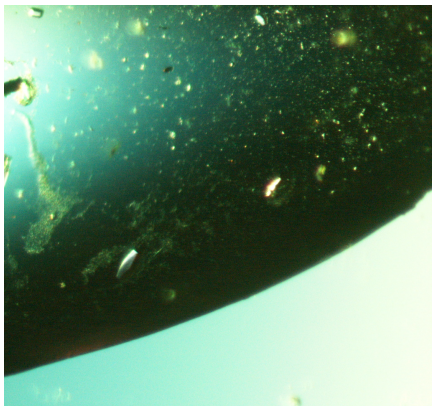


AhlC JCSG+ B8 crystals in drop

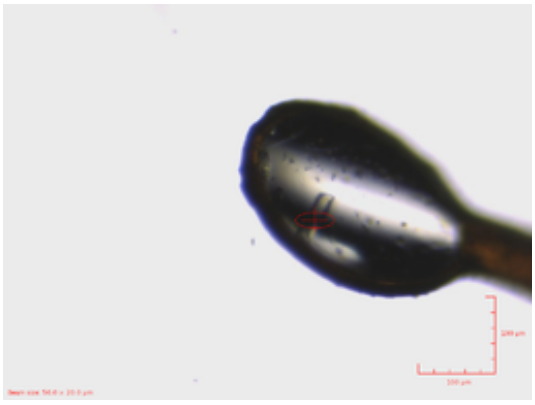


Mounted AhlC JCSG+ B8 crystal

B

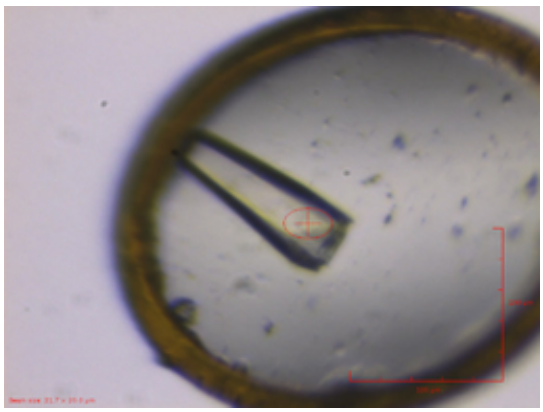


AhlC PACT D7 crystals in drop



Mounted AhlC PACT D7 crystal

C



Mounted SeMet AhlC JCSG+ B8 crystal

Figure 6.2. **AhlC crystals used for data collection.** (**A**) Picture of initial JCSG+ B8 crystals in crystallisation drop (left), and mounted in a litholoop (right). (**B**) Picture of PACT D7 crystal growing in crystallisation drop (right), and a PACT D7 crystal mounted in a litholoop (right). (**C**) Picture of an optimised JCSG+ B8 crystal grown from AhlC SeMet, mounted in a litholoop.

Crystals grown in PACT D7 diffracted to 2.8 Å and were predicted to be in Laue group P2 from the test images, collected identically to above.

6.1.6 Full data collection and processing for crystal form 1 (JCSG+ B8)

As the B8 crystals appeared to diffract to the highest resolution, a full dataset was collected, using the strategy for data collection generated by EDNA (Incardona et al., 2009). A start phi angle of 0° was used, and 200° of data was collected with an exposure of 0.2 seconds and a rotation of 0.2° per image, at an energy of 12699 eV. An example of the diffraction quality is given in Figure 6.3, and shows that the crystals visibly diffracted to an approximate resolution of 2.9 Å. The data was processed automatically at the Diamond synchrotron using the XIA2 pipeline, using the 3d, 3dii, and 2d options (Winter, 2010). The best processing based on resolution cut-off and R_{merge} was the 3dii processing with the full data statistics shown in Table 6.1. The final data processed in the spacegroup $P6_122$ with cell dimensions of a = b = 134.4 Å, c = 145.8 Å. The resolution cut-off determined by $CC_{1/2}$ was 2.65 Å, with an R_{pim} in the outer shell of 0.23, and an overall R_{pim} of 0.028.

The Matthews coefficient of all datasets were calculated in CCP4, and in each case either 2 molecules with a solvent content of 62.1 %, or 3 molecules with a solvent content of 43.2 % were expected (Figure 6.4).

6.1.7 Full data collection and processing for crystal form 2 (PACT D7)

A full dataset was also collected from an AhlC crystal grown in PACT D7. A start phi angle of 32° was used, and 180° of data was collected with an exposure of 0.05 seconds and a rotation of 0.2° per image, at an energy of 12699 eV. The shorter exposure time was offset by increasing the relative transmission by a factor of 4, in order to collect data quicker. An example of the diffraction quality is given in Figure 6.5, and shows that the crystals visibly diffracted to an approximate resolution of 3 Å. The data were processed automatically using the Xia2 3dii pipeline, in spacegroup $P2_1$ with cell dimensions of a = 65.1 Å, b = 61.7 Å, c = 129.8 Å, and a β angle of 92.0°. The resolution cut off of the processed data determined by the $CC_{1/2}$ was 2.62 Å with an R_{pim} in the outer shell of 0.51 and an overall Rpim of 0.058. The full data processing statistics are shown in Table 6.2. The Matthews coefficient analysis of crystal form 2 suggested that 4 molecules per unit cell was most likely, with a solvent content of 44.1 %, or 3 molecules with a solvent content of 59.6 % solvent content (Figure 6.6).

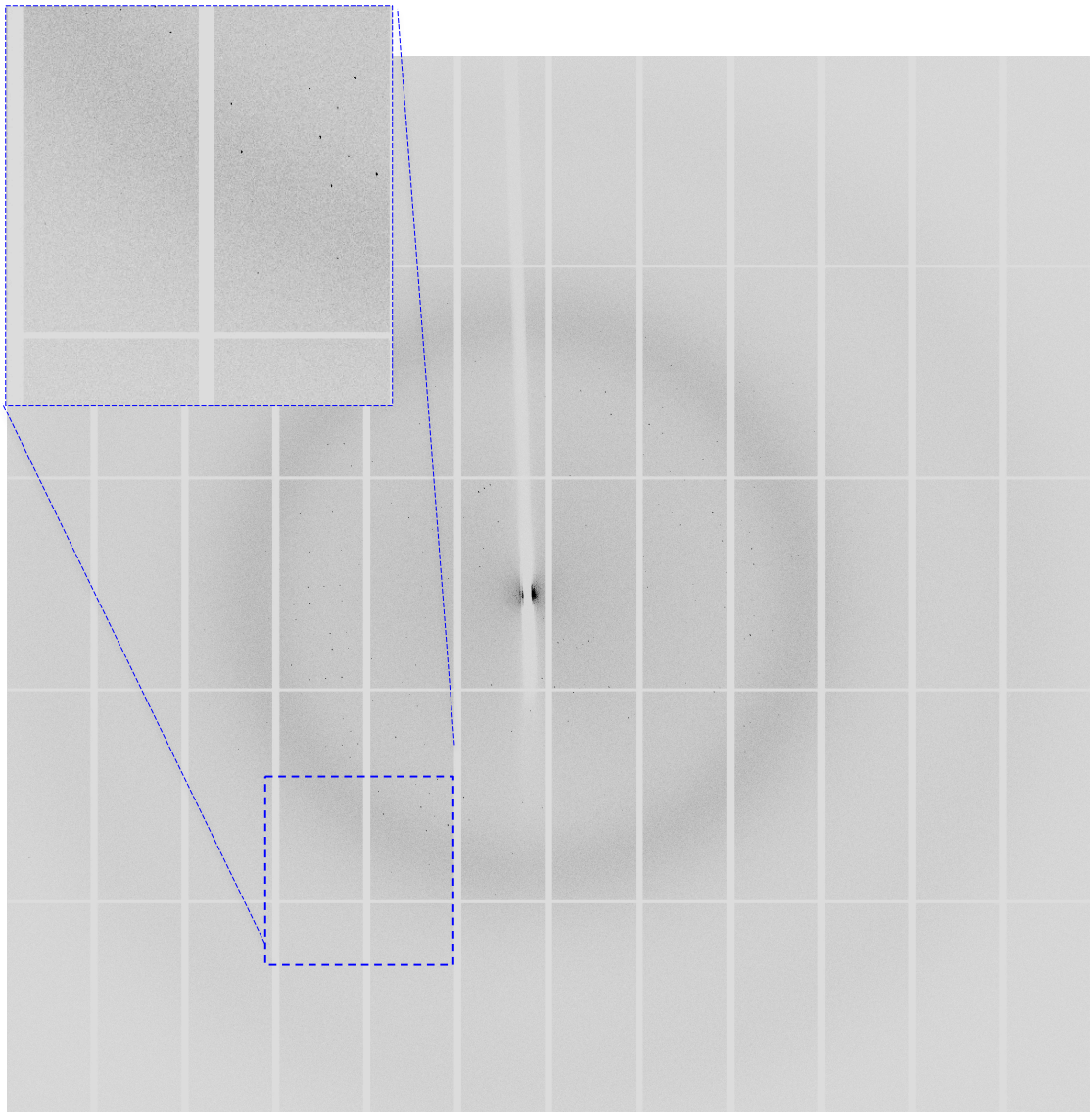
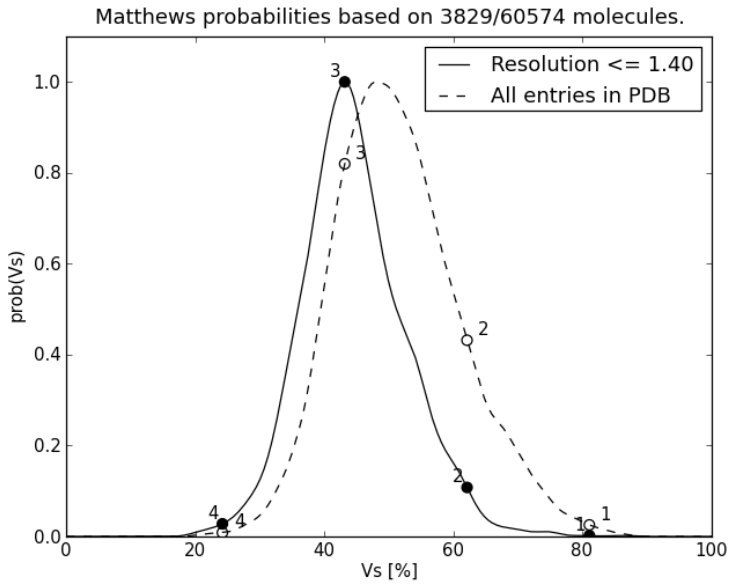


Figure 6.3. **Diffraction image from data collected on AhlC JCSG+ B8 crystal.** A diffraction image collected at phi = 0° with an oscillation of 0.2° from the JCSG+ B8 AhlC crystal. The inset square highlights a selection of the diffraction and a 2x magnification is shown to show more clearly reflections, and also to show the visible limit of diffraction. The image plate is set such that the edge of the image is at 2Å resolution.

	Overall	Low	High
Spacegroup	$P6_122$		
a (Å)	134.4		
b (Å)	134.4		
c (Å)	145.8		
α (°)	90		
β (°)	90		
γ (°)	120		
High resolution limit	2.65 Å	11.85 Å	2.65 Å
Low resolution limit	116.40 Å	116.40 Å	2.72 Å
Completeness	99.9 %	99.7 %	100 %
Multiplicity	20.7	15.5	21.6
I/sigma	19.9	51.3	4.0
R_{merge}	0.123	0.049	1.036
$R_{meas}(I)$	0.129	0.053	1.036
$R_{meas}(I+/-)$	0.129	0.052	1.084
$R_{pim}(I)$	0.028	0.015	0.231
$R_{pim}(I+/-)$	0.038	0.017	0.319
Wilson B factor	35.87		
Total observations	479843	5111	36014
Total unique	23189	329	1670

Table 6.1. Xia2 3dii processing statistics for native AhlC JCSG+ B8 data collection. The final spacegroup solution, resolution limit and merging statistics are shown for the best diffracting native AhlC crystal grown in JCSG+ B8. The R values for merging etc are defined in Chapter 2. The final table is reproduced from the merging table output by Aimless.

A



B

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0019	0.0198	6.49	81.05	29266.00
2	0.0949	0.3359	3.25	62.10	58532.00
3	0.8787	0.6377	2.16	43.16	87798.00
4	0.0245	0.0066	1.62	24.21	117064.00

Figure 6.4. Matthews coefficient graph and table for AhlC JCSG+ B8 data. A graph of Vm ($Å^3$/Da) and percentage solvent vs. probability plotted using all entries at in the PDB (dashed) and all entries at 2.7 Å (solid line). Highlighted on the graph is each incremental number of monomers in the AU, with 3 being the most probable (highest in Y). Below is a table of Vm's and numbers of molecules with probability. The Graph and table were generated using the online software at http://www.ruppweb.org/mattprob/.

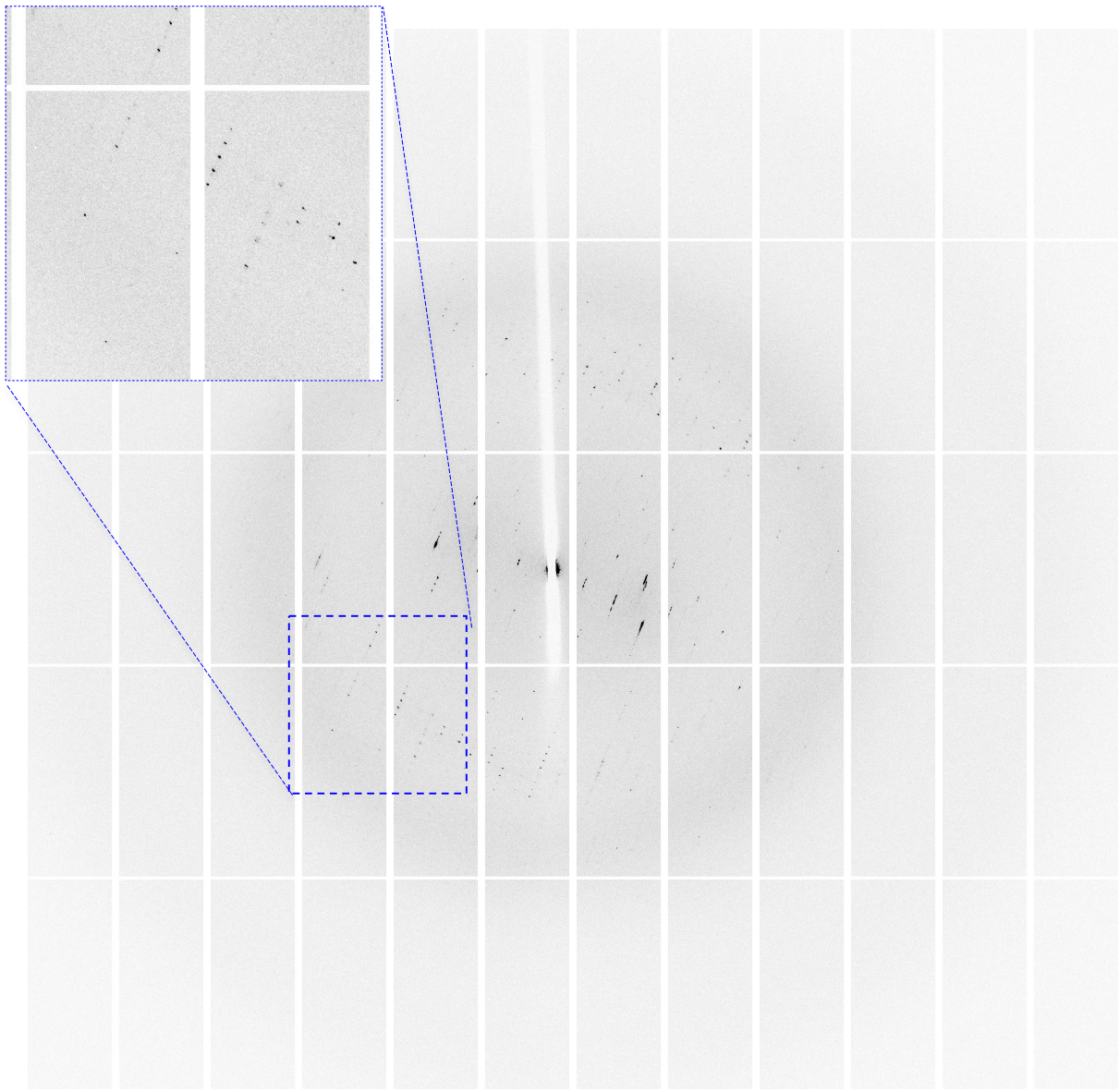
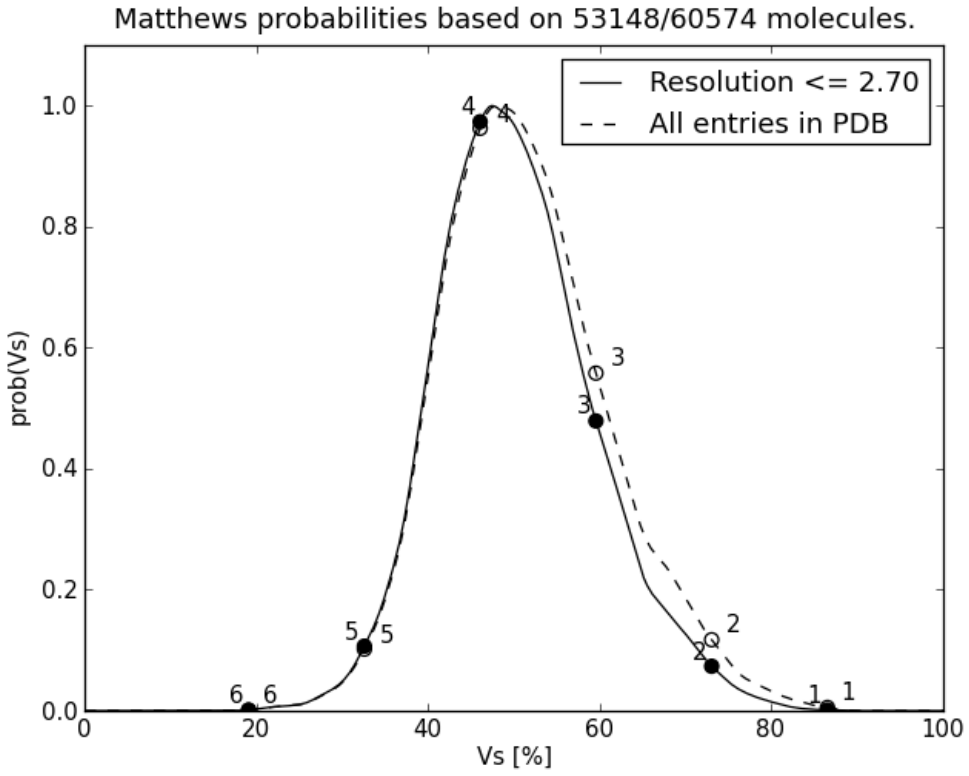


Figure 6.5. Diffraction image from data collected on AhIC PACT D7 crystal. A diffraction image collected at phi = 0° with an oscillation of 0.2° from the PACT D7 crystal. The inset square highlights a selection of the diffraction and a 2x magnification is shown to show more clearly reflections, and also to show the visible limit of diffraction. The image plate is set such that the edge of the image is at 2Å resolution.

Table 6.2. Xia2 3dii processing statistics for native AhlC PACT D7 data collection.

	Overall	Low	High
Spacegroup	$P2_1$		
a (Å)	65.1		
b (Å)	61.7		
c (Å)	130.0		
α (°)	90		
β (°)	92.0		
γ (°)	90		
High resolution limit	2.62 Å	11.72 Å	2.62 Å
Low resolution limit	46.74 Å	46.74 Å	2.69 Å
Completeness	99.7 %	93.1 %	100 %
Multiplicity	3.4	2.9	3.4
I/sigma	11.0	36.3	1.8
R_{merge}	0.074	0.022	0.710
$R_{meas}(I)$	0.108	0.039	0.954
$R_{meas}(I+/-)$	0.103	0.031	0.977
$R_{pim}(I)$	0.058	0.023	0.509
$R_{pim}(I+/-)$	0.071	0.022	0.669
Wilson B factor	37.858		
Total observations	105013	1039	7858
Total unique	31150	362	2309

A



B

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0008	0.0031	9.12	86.52	28500.00
2	0.0450	0.0672	4.56	73.04	57000.00
3	0.2929	0.3195	3.04	59.56	85500.00
4	0.5953	0.5512	2.28	46.08	114000.00
5	0.0653	0.0583	1.82	32.59	142500.00
6	0.0006	0.0006	1.52	19.11	171000.00

Figure 6.6. Matthews coefficient graph and table for AhlC PACT D7 data. A graph of % solvent vs. probability plotted using all entries at in the PDB (dashed) and all entries at 2.7 Å (solid line). Highlighted on the graph is each incremental number of monomers in the AU, with 4 being the most probable (highest in Y). Below is a table of Vm's and numbers of molecules with probability. Graph and table were generated using the online software at http://www.ruppweb.org/mattprob/.

6.1.8 Optimisation of initial hits using SeMet derivitised AhlC

As both the JCSG+ B8 and PACT D7 sulphur-methionine crystals diffracted well, attempts were made to grow similar crystals using the SeMet protein, to use for anomalous scattering phase determination. Initial hits from the native AhlC grown in JCSG+ B8 and PACT D7 were optimised in hanging drop format using SeMet AhlC, changing both the pH and the PEG concentration. 1 µl of protein was mixed with 1 µl crystallisation solution on a coverslip above well solution and stored at 17°C. Optimisations were checked daily for 3 days, before storing and checking weekly for crystal growth.

6.1.9 Seleno-methionine anomalous data collection for crystal form 1

Crystals of SeMet AhlC identified during JCSG+ B8 optimisations, grown in 0.2 M $MgCl_2$, 0.1 M tris pH 7.0, and 10 % PEG8000, were cryo-cooled in liquid nitrogen as described in 3.6.5, and sent for data collection at the Diamond synchrotron. Diffraction tests were carried out by collecting diffraction data at 5 positions 45° apart, with an oscillation per image of 0.2° and an exposure of 0.05 seconds per image (with four times relative transmission). An example of the diffraction quality is given in Figure 6.7, and shows that the crystals visibly diffracted to an improved quality compared to initial JCSG+ B8 hit crystals (Figure 6.7), and to an approximate resolution of 2.6 Å FastDP predicted the crystal to be of the Laue group P3, with unit cell dimensions of a = b = 134.7 Å, and c = 145 Å, similar to the S-Met data.

A fluorescence scan on a Se-Met crystal was carried out on the I04 beamline, optimised for the Se K edge, to determine the anomalous signal presence and to calculate f' and f'' values. The fluorescence scan (Figure 6.8) confirmed the presence of selenium, with the peak and inflection positions occurring at 12659.0 eV and 12657 eV respectively. The program Chooch was used to calculate f' and f'' values for both the peak wavelength of -7.45e and 5.79e, and -10.1e and 3.58e respectively. Datasets were collected from the same crystal at the Se peak and inflection energies, as well as at 12699.39eV, chosen as a high-energy remote wavelength.

The peak dataset was collected at a wavelength of 0.9790 Å with 3600 images collected with an oscillation per image of 0.1° (360° total). FastDP auto processing was run automatically and determined the data was in Laue group P622 (Xia2 3dii subsequently determined the spacegroup as $P6_122$), with a high-resolution cutoff of 2.8 Å. The cell

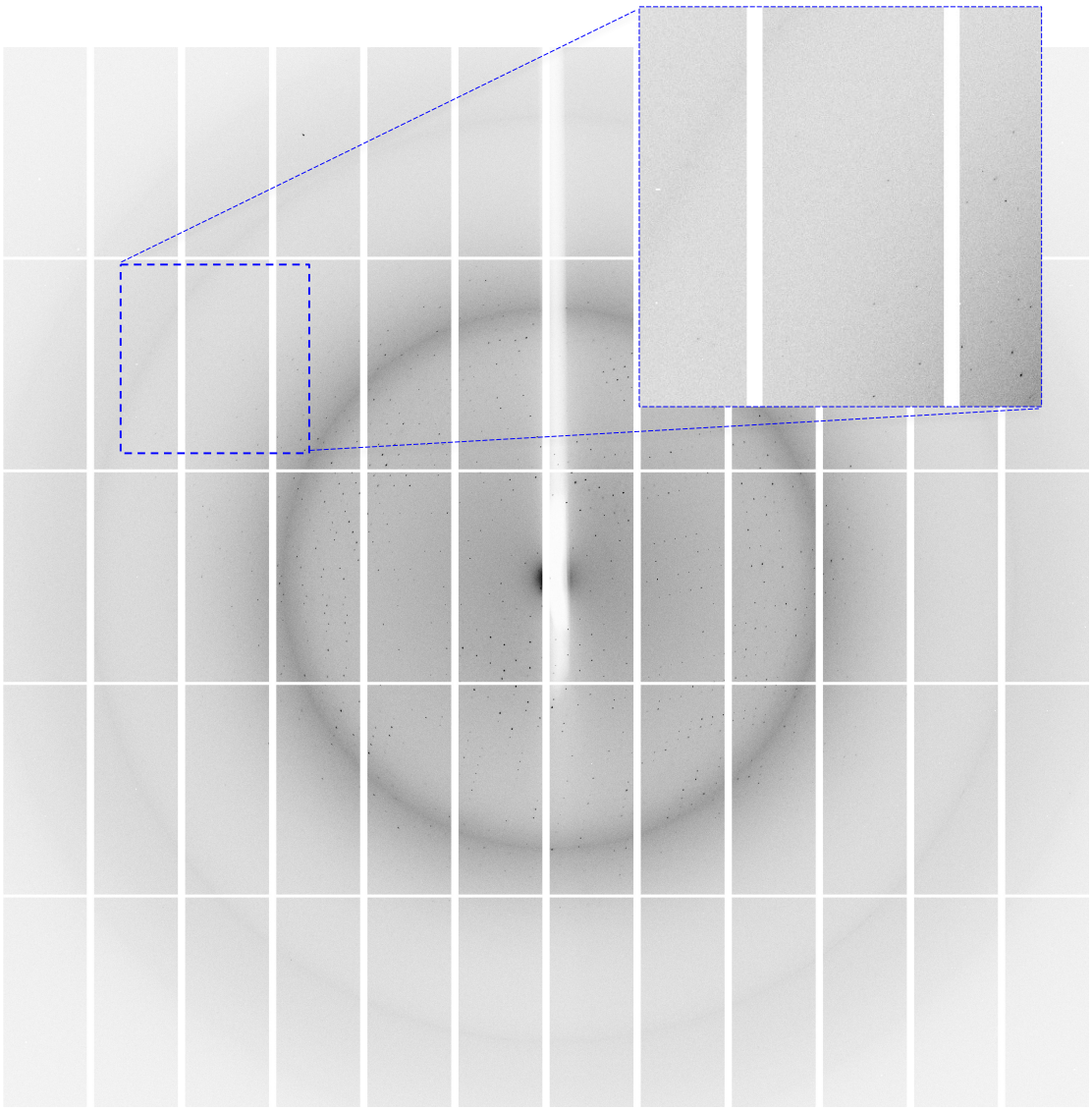
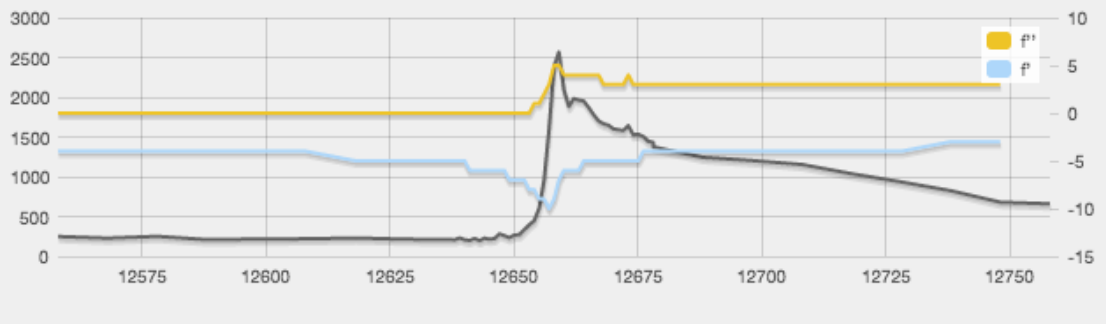


Figure 6.7. Diffraction image from data collected on AhlC SeMet JCSG+ B8 crystal. A diffraction image collected at phi = 0° with an oscillation of 0.2° from the SeMet JCSG+ B8 crystal. The inset square highlights a selection of the diffraction and a 2x magnification is shown to show more clearly reflections, and also to show the visible limit of diffraction. The image plate is set such that the edge of the image is at 2.3Å resolution.

A



B

	Wavelength	f''	f'
Peak	12658.99 eV (0.9794 Å)	5.79 e^-	-7.45 e^-
Inflection	12657 eV (0.9796 Å)	3.58 e^-	-10.07 e^-

Figure 6.8. Fluorescence scan of AhlC SeMet crystal at Se K edge. (A) A fluorescence scan of the selenium K absorption edge carried out on with 1 second exposure and 6.4 % transmission. The black line shows the fluorescence at each wavelength. The yellow line plots f'' and the blue plots f'. (B) A table showing the determined peak and inflection wavelengths and energies, and the f'' and f' values at the peak and inflection wavelengths, calculated by CHOOCH.

dimensions were a = b = 134.6 Å, c = 145.3 Å. The FastDP processing (Table 6.3) met the minimum requirement for fastEP to be run, which was used to calculate initial phases (Fisher et al., 2015).

An inflection dataset was also collected, translating the crystal slightly to expose a separate part of the crystal and minimise radiation damage. The inflection data were collected at 12656.6 eV, and otherwise identically as the peak dataset. Lastly, a high-energy remote dataset was collected, again translating the crystal to minimise radiation damage. The data were collected at a wavelength of 12699.4, and otherwise identical to the peak and inflection datasets. A final comparison of the Xia2 processing statistics for each peak, inflection and high-energy remote data are given in Table 6.4.

Anomalous signal was present in all three datasets, and radiation damage as determined by the resolution cut-off of each dataset, the merging Rvalues per batch, and the unit cell expansion were not large. The anomalous signal present in the peak wavelength dataset was however sufficiently high for the fastEP processing to be run automatically at the Diamond Synchrotron (Fisher et al., 2015).

6.1.10 Cell content analysis of crystal form 1 and number of expected seleniums

AhlC contains two methionines per chain (excluding the N-terminal Met), and so based on the Matthews coefficient , either 4 or 6 selenium atoms were expected based on packing.

6.1.11 FastEP phasing of AhlC SeMet crystal form 1

The peak dataset contained a midslope of 1.36 for the anomalous normal probability, and an R_{pim} in the inner shell of 0.032. FastEP is the phasing pipeline available at the Diamond synchrotron, which runs multiple SHELX jobs to calculate initial phases with minimal user input, by changing number of heavy atom sites, solvent content and spacegroup enantiomorphs (Figure 6.9).

Analysis of the data using SHELXC showed that anomalous signal was present (>0.8 d"/sig) to 2.8 Å. SHELXD was run using each enantiomorph of the P622 spacegroup, and the best scores were obtained for 4 Se sites found in $P6_122$, with a CC_{all} of 51.2 and CC_{weak} of 29.8. The residual map for the selenium sites were examined manually and showed that four of the 8 sites were present in the electron density when set to 4σ (Figure 6.10), which is expected based on the sequence of AhlC and the Matthews coefficient. Finally, phases were

Table 6.3. FastDP processing statistics for SeMet AhlC JCSG+ B8 data collection.

	Overall	Low	High
Laue group	P622		
a (Å)	134.6		
b (Å)	134.6		
c (Å)	145.3		
α (°)	90		
β (°)	90		
γ (°)	120		
High resolution limit	2.81 Å	12.57 Å	2.81 Å
Low resolution limit	29.54 Å	29.54 Å	2.88 Å
Completeness	99.8 %	92.5 %	98.1 %
Multiplicity	38.6	28.7	37.7
I/sigma	28.1	90.8	4.8
R_{merge}	0.129	0.032	1.013
R_{meas}(I)			
R_{meas}(I+/-)			
R_{pim}(I)			
R_{pim}(I+/-)			
Wilson B factor			
Mid-slope	1.36		
dF/F	0.059		
dI/sig(dI)	1.37		
Total observations	751732	7346	51617
Total unique	19468	256	1369

Table 6.4 Aimless statistics for Xia2 3dii processing.

	Peak	Inflection	High energy remote
Spacegroup	$P6_122$	$P6_122$	$P6_122$
a (Å)	134.6	134.7	134.8
b (Å)	134.6	134.7	134.8
c (Å)	145.3	145.3	145.3
α (°)	90	90	90
β (°)	90	90	90
γ (°)	120	120	120
High resolution limit	2.45 Å	2.35 Å	2.66 Å
Low resolution limit	58.3 Å	58.3 Å	61.7 Å
Completeness	99.9 %	99.8 %	99.9 %
Multiplicity	39.1	39.9	19.3
I/sigma	18.7	19.8	14.4
R_{merge}	0.169	0.138	0.157
$R_{meas}(I)$	0.177	0.144	0.167
$R_{meas}(I+/-)$	0.174	0.141	0.165
$R_{pim}(I)$	0.028	0.023	0.038
$R_{pim}(I+/-)$	0.038	0.031	0.051
Wilson B factor	42.5	48.9	42.9
Anomalous completeness	99.9 %	99.7 %	99.9
Anomalous multiplicity	21	21	10.1
Anomalous correlation	0.589	0.529	0.240
Anomalous slope	1.22	1.20	1.093
dF/F	0.072	0.062	0.081
dI/s(dI)	1.24	1.20	0.935
Total observations	1137906	1293084	444893
Total unique	29135	32942	3842

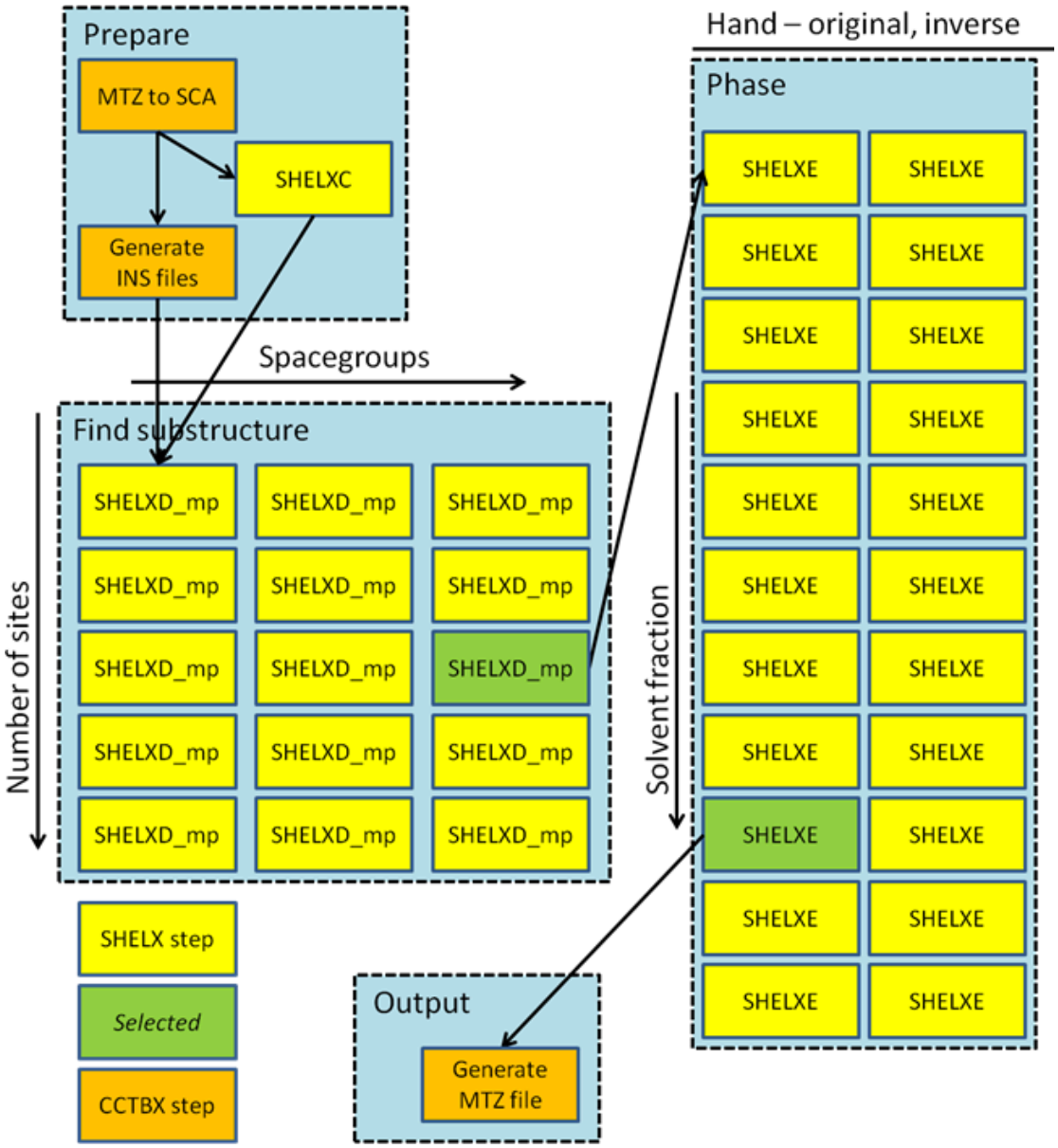
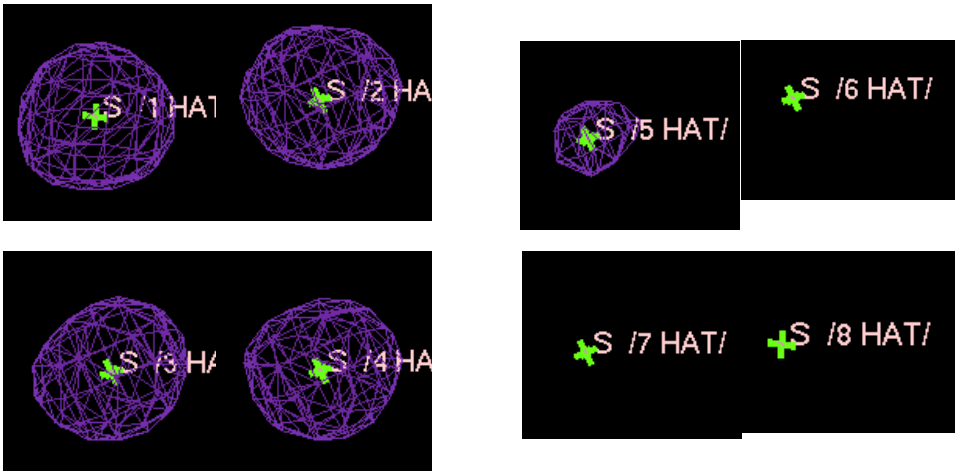


Figure 6.9. Workflow diagram of FastEP. Schematic flow diagram of programs used in the brute force method of phasing employed by FastEP auto processing at Diamond light source. The substructure solution is run multiple times in parallel on the cluster with a number of different sites to search for and all enantiomorphs of the spacegroup chosen in FastDP. Shelxe is then run in both original and inverse hand with a number of different solvent contents, and finally the result with the highest score is output and an mtz file containing phased structure factors is generated.

A



B

```
         10          20          30          40          50          60
MSNGILSQSI ANMQQAEATI QSFSGLPQNA VNIQQNVGEV VAALLPQVQT MQQQVLAFAA

         70          80          90         100         110         120
RLELQLTQQL ANTGPFNPEA LKAFVDLVQQ EIAPIQTLTA QTLTASQSAN DRITQDNIAL

        130         140         150         160         170         180
QRIGVELQAT IAGLQSNLDG ARQELDSLNK KKLYLLGLGL LGLPGLIALA VTLTQTQNKV

        190         200         210         220         230         240
SSLEGQVNQI EGQIQRQQGF LGQTTAFSQQ FGSLIDRVSK VGNTISLLGG DIANVARDAG

        250         260
QGDPELARLF FTAALTEVRT LQVDAS
```

Figure 6.10. Heavy atoms found by FastEP SHELXD. (A) Heavy atoms identified in SHELXD, and the residual density set at 4σ rmsd. The first four sites have strong density for each site, with the last four containing weak or no density. (**B**) Amino acid sequence of AhlC with the 2 non-terminal methionine's highlighted.

calculated and density modification performed using SHELXE, in a number of different runs, where the fractional solvent content of the cell was varied between 0.25 and 0.75. The best electron density map was obtained for the inverted hand of the Se sites, a solvent content of 0.5, and in the enantiomorph $P6_522$. Manual examination of the original and inverted hand map showed that the inverted hand was indeed correct, and the initial phases were of sufficient quality to see a continuous chain of density for the protein (Figure 6.11).

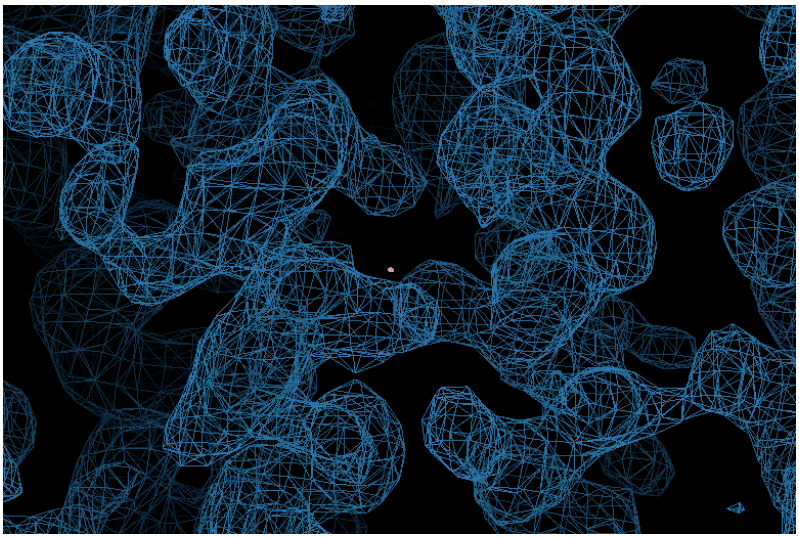
At this stage, the automatic phasing from fastEP was decided to be of sufficient quality to proceed to model building, rather than incorporating the inflection and high-energy remote wavelength datasets to the phase calculation.

6.1.12 Structure building using Buccaneer and rebuilding in Coot and Refmac5

The map generated in fastEP was used in Buccaneer to build a starting model for rebuilding and refinement (Cowtan, 2006). The map containing the initial phases and the protein sequence was supplied, along with the model produced in SHELXE, and the program was run with 5 cycles of building and refinement. The final cycle of building in Buccaneer had placed 548 residues in the map, built in 6 fragments, the longest of which contained 269 residues. 502 of the built residues had been assigned an amino acid type, and were allocated into 3 chains. The final cycle of refinement yielded an R factor of 0.30 and R_{free} of 0.36.

Using this model as a starting point, the protein was further built manually in Coot (Emsley et al., 2010) and refined in Refmac5 (Murshudov et al., 1997). Chains B and D were determined to be from the same molecule of AhlC and so were combined. Missing residues were added where there were features in both the |2mFo-DFc| electron density and the |Fo-Fc| difference electron density maps. This process resulted in a model containing two AhlC molecules in the asymmetric unit. Between rounds of building the model was refined against the original data in spacegroup $P6_522$. The structure was rebuilt and refined for 39 cycles until no new features could be interpreted in the difference map. The final model fitted the density well, with an R_{work} and R_{free} of 0.22 and 0.28, respectively. Figure 6.12 shows an example of the starting and final electron density map in a particularly difficult region to build, namely the hairpin formed between α3 and α4.

Original hand



Inverted hand

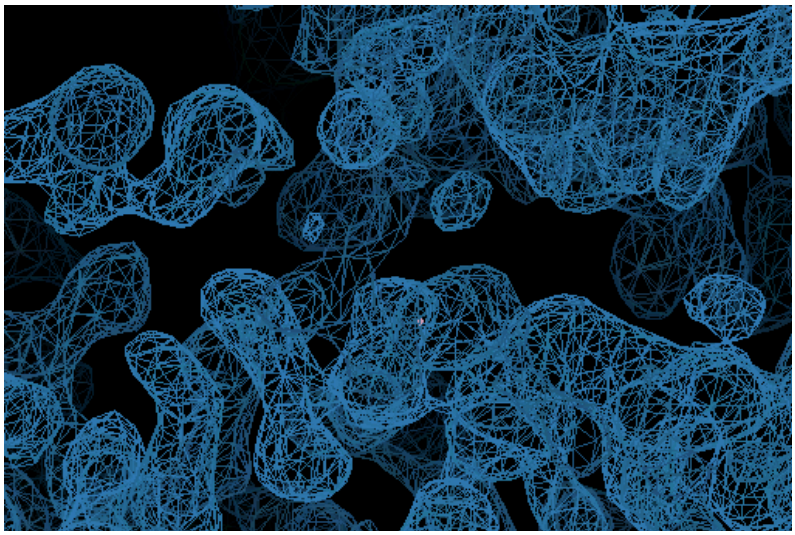


Figure 6.11. Electron density maps generated in the original and inverted hand by shelxe. Shelxe output after density modification and chain building, in the inverted and correct hand, and in the original hand. The inverted hand shows fully connected density for two alpha helices, each with obvious sidechains present. The original hand shows good solvent boundary, but disconnected density and more noise present.

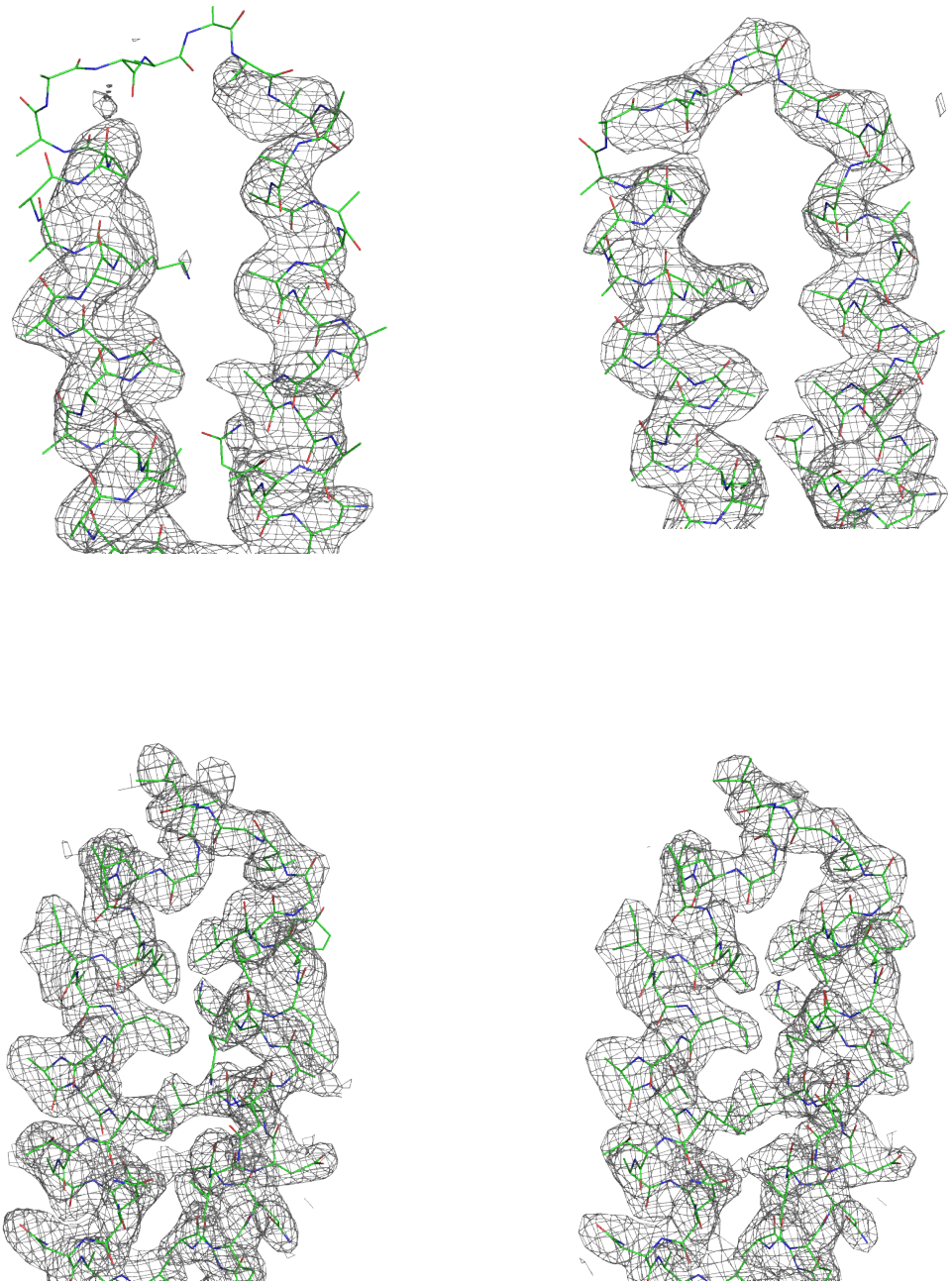


Figure 6.12. Initial and final 2Fo-Fc electron density maps of the AhlC head region after rebuilding and refinement. The initial electron density maps for chain A (top) and chain B (bottom) showing the final AhlC model with the electron density as it looked after buccaneer (left). The final electron density (right) for AhlC chain A head shows great improvement with the full carbon backbone present. The initial electron density map of chain B already contained the majority of the AhlC head, although density improved during rebuilding in COOT, allowing more sidechains to be built into the final electron density map.

6.1.13 Structure validation of AhlC crystal form 1 structure

The final model of AhlC contained 97.4 % of residues (533 residues) with favoured Ramachandran angles, and no residues were Ramachandran outliers Figure 6.13. Of the 548 residues, 2.7 % or seven residues had poor rotamers. In areas of poor density rotamers were assigned based on likelihood of particular rotamers, whilst also choosing the rotamer of best fit. Those that were left as poor rotamers included Phe211, which was in the core of the four-helical bundle, and had good density showing that an unusual rotamer existed, found in both chains. Other poor rotamers included Thr172, and Ser226, which refined to unfavoured rotamer positions to avoid clashing with atoms from surrounding residues. Nevertheless, in Molprobity, the overall clash score was 2.21, and the structure was in the 100th percentile for both clash score and overall Molprobity score, when compared with other structures of a similar resolution (N=6237, 2.60Å ± 0.25Å)(Table 6.5)

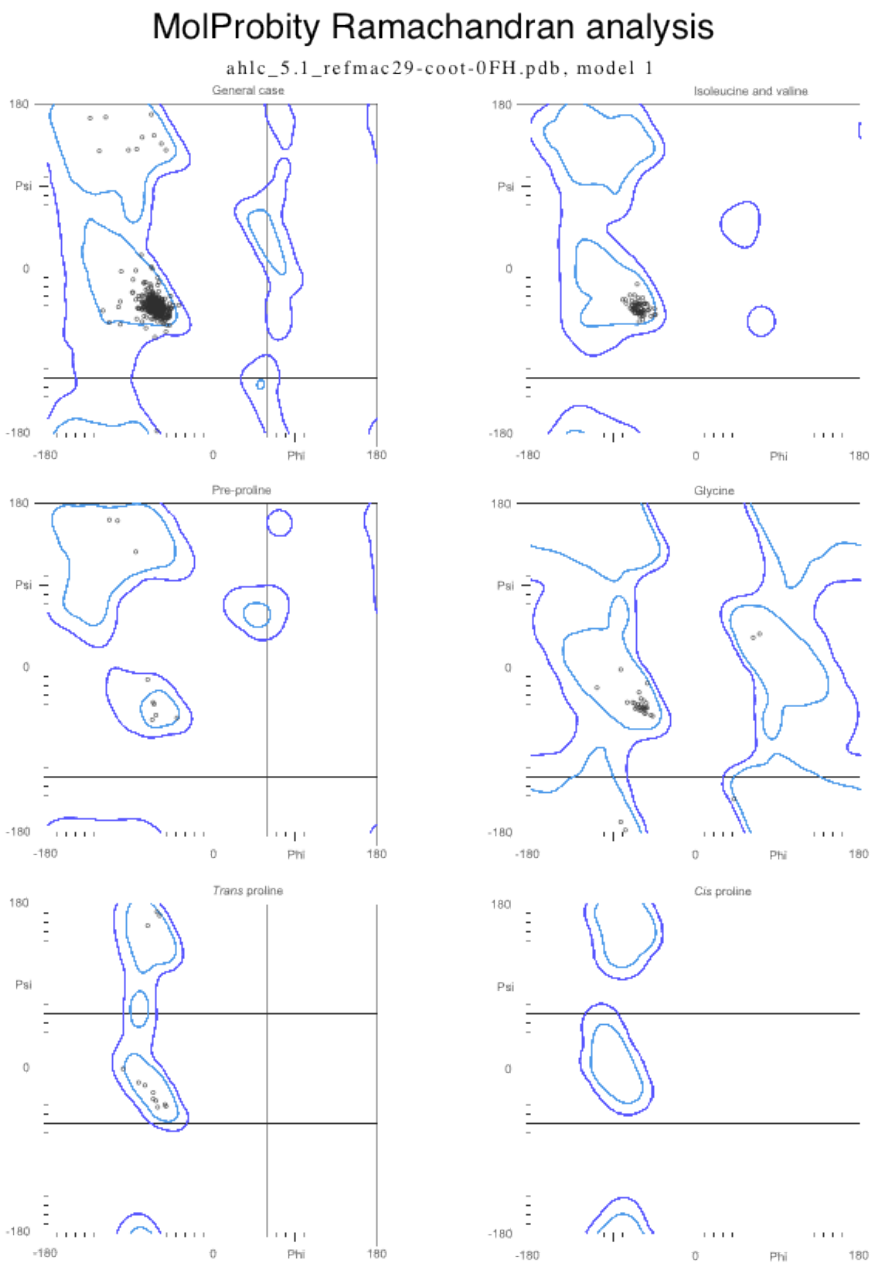
6.1.14 Solving the AhlC crystal form 2 (PACT D7) structure by molecular replacement

As no crystals were grown in PACT D7 optimisations using seleno-methionine AhlC, the seleno-methionine AhlC JCSG+ B8 structure was used to generate a model for molecular replacement, to calculate starting phases for the AhlC sulphur-methionine PACT D7 data. Chain B from the JCSG+ B8 structure was the most complete molecule of the two that existed in the B8 asymmetric unit, and so was chosen to use in molecular replacement. The isolated chain B was copied into a separate file and used as a search model in PhaserMR (McCoy et al., 2007), with a sequence identity set to 30 %, to increase the allowed rmsd and clashing cut offs. The Matthews coefficient suggested there were four molecules of AhlC in the asymmetric unit, and so the program was set to find four copies of the search model.

A single solution was output by PhaserMR, with TFZ scores per molecule of 7.5, 15.6, 14.4, and 11.7 and an overall LLG of 2585, indicating a correct solution. The final refinement in PhaserMR after placing all 4 monomers gave an R factor of 0.48, and an R_{free} of 0.48. 10 cycles of restrained refinement were run in Refmac5 before manually rebuilding in Coot. The final R factor and R_{free} after restrained refinement were 0.31 and 0.38 respectively.

6.1.15 Building and refinement of the crystal form 2 AhlC structure

The structure of the PACT D7 AhlC was rebuilt in Coot, removing residues that did not fit the electron density, and building in residues and sidechains where new density was present in



97.4% (481/494) of all residues were in favored (98%) regions.
100.0% (494/494) of all residues were in allowed (>99.8%) regions.

There were no outliers.

Figure 6.13. Molprobity Ramachandran analysis of AhlC crystal form 1. Molprobity Ramachandran analysis of Phi and Psi torsion angles. Each plot is for a specific type of amino acid, including the general case, isoleucines and valines, pre-proline residues, glycines, trans-prolines, and cis-prolines. Phi angles are plotted against Psi angles for each residue, with islands of preferred Ramachandran angles in light blue, and allowed Ramachandran angles in dark blue. Outside of this region are disallowed Ramachandran angles. For AhlC crystal form 1 structure, 97.4 % of residues were in favoured regions, and 100 % of all residues were in allowed regions.

All-Atom Contacts	Clashscore, all atoms:	2.21		100th percentile* (N=227, 2.60Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	10	2.72%	Goal: <0.3%
	Favored rotamers	332	90.22%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	481	97.37%	Goal: >98%
	MolProbity score^	1.45		100th percentile* (N=6237, 2.60Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.21%	Goal: 0
	Bad bonds:	0 / 3690	0.00%	Goal: 0%
	Bad angles:	3 / 5022	0.06%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	0	0.00%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%

Table 6.5. Molprobity analysis of the final refined structure of AhIC crystal form 1 (PACT B8). The final refined structure was validated using the Molprobity online server. The output shows that the final structure is in the 100th percentile for both the clash score, and the overall Molprobity score (N=6237, 2.60Å ± 0.25Å).

the |2mFo-DFc| electron density and the |Fo-Fc| difference electron density map. In between each round of rebuilding, refinement was carried out in Refmac5 using restrained refinement and isotropic Bfactor refinement. The head domain of each AhlC molecule varied compared with its position in the B8 structure, and so was completely rebuilt. In some chains, the head domain density was of insufficient quality to confidently build in the complete protein chain (Figure 6.14). Overall, the quality of the data and therefore the quality of the model built from the D7 data were lower than that of the B8 data.

6.1.16 Structure validation of the AhlC PACT D7 structure

The final model of AhlC crystal form 2 contained 4 unique chains and 807 residues built, out of 1104 possible (73 % of all residues). This low number is due to the difficulty in building the head domain of AhlC crystal form 2. The final R factors were 0.29 and 0.37 for R_{work} and R_{free} respectively. When the map could no longer be interpreted, building was suspended and the structure was validated for what could be built using Molprobity (Table 6.6). The final model contained 96.9 % favoured Ramachandran angles, with only one Ramachandran outlier (Figure 6.15). Only 85.7 % of rotamers were favoured, and 30 rotamers were poor. Despite this, the overall Molprobity score was good for the resolution of the data, with the model in the 100th percentile for clash score, and in the 98th percentile for the overall Molprobity score, when compared with 6055 structures of a similar resolution.

6.2 Structural analysis of AhlC

6.2.1 Overall structure of AhlC

As the AhlC model for the B8 crystals was the most complete, it is used here for the majority of the time to describe the structure. However, all built AhlC chains are shown in Figure 6.16 to highlight the different levels of map interpretation achieved for each chain. The structure of AhlC is made up of five α-helices, connected by short coils and loop regions (Figure 6.17). The helices form two major domains, the head and tail domain, connected by a stalk running the length of the protein (Figure 6.18), similar to other proteins in the PFT family. The head domain is made up of an elongated helical hairpin, which is generated by α3 and α4, that run the entire length of the protein, contributing to the tail, the stalk, and the head domains. The tail is also similar to other family proteins, forming a bundle of four

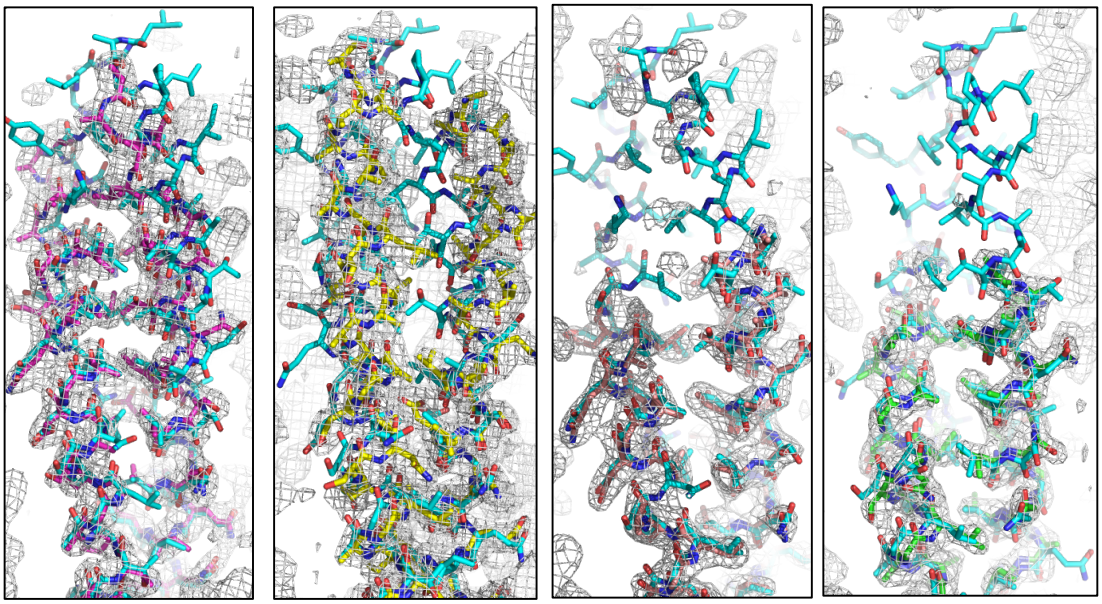


Figure 6.14. Final electron density present for the head domain of each chain present in crystal form 2. For each chain, the final electron density is shown for the head domain, set to 1σ rmsd. For each, the head domain of crystal form 1 chain B is superposed in cyan, and the crystal form 2 head domain is magenta for chain A, yellow for chain B, orange for chain C, and green for chain D. Both chain A and B contain the most electron density for this region, and the head follows a similar path to that of crystal form 1 chain B. In chain C and D, the head domain is much more flexible and 15 and 17 residues are missing in each respectively.

All-Atom Contacts	Clashscore, all atoms:	2.83		100th percentile* (N=226, 2.62Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	30	5.78%	Goal: <0.3%
	Favored rotamers	445	85.74%	Goal: >98%
	Ramachandran outliers	1	0.13%	Goal: <0.05%
	Ramachandran favored	750	96.90%	Goal: >98%
	MolProbity score^	1.84		98th percentile* (N=6055, 2.62Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.13%	Goal: 0
	Bad bonds:	0 / 5616	0.00%	Goal: 0%
	Bad angles:	1 / 7650	0.01%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 11	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	1	0.14%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%

Table 6.6. Molprobity analysis of the final refined structure of AhlC crystal form 2 (PACT D7). The final refined structure was validated using the Molprobity online server. The output shows that the final structure is in the 100th percentile for Clashscore, and the 98th percentile for the overall Molprobity score (N=6055, 2.62Å ± 0.25Å).

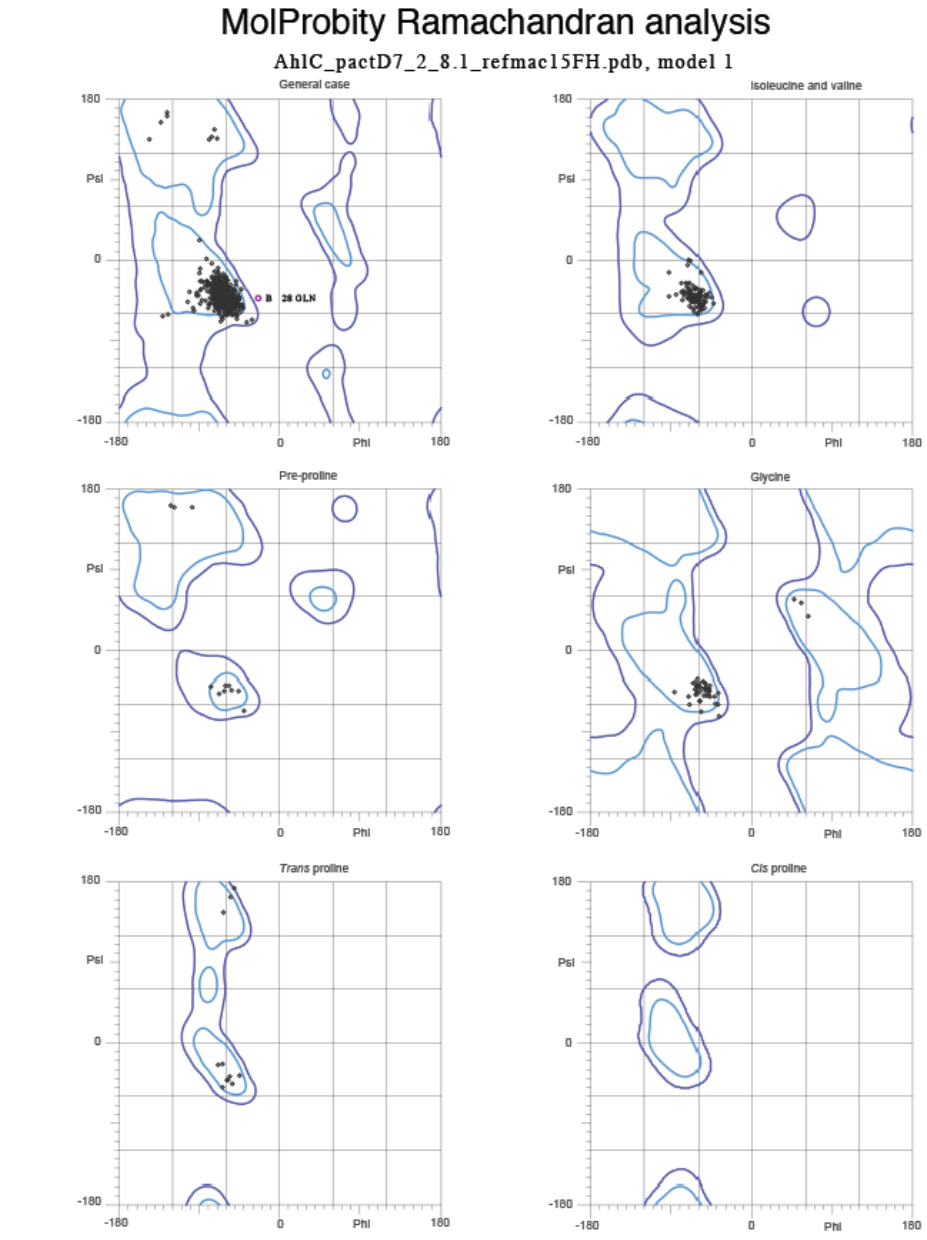


Figure 6.15. Molprobity Ramachandran analysis of AhlC crystal form 2 (PACT D7). Molprobity Ramachandran analysis of Phi and Psi torsion angles. Each plot is for a specific type of amino acid, including the general case, isoleucines and valines, pre-proline residues, glycines, trans-prolines, and cis-prolines. Phi angles are plotted against Psi angles for each residue, with islands of preferred Ramachandran angles in light blue, and allowed Ramachandran angles in dark blue. Outside of this region are disallowed Ramachandran angles. For AhlC crystal form 2 structure, 96.9 % of all residues were in favoured regions, and 99.9 % of all residues were in allowed regions. The Ramachandran outlier was Gln28 of chain B.

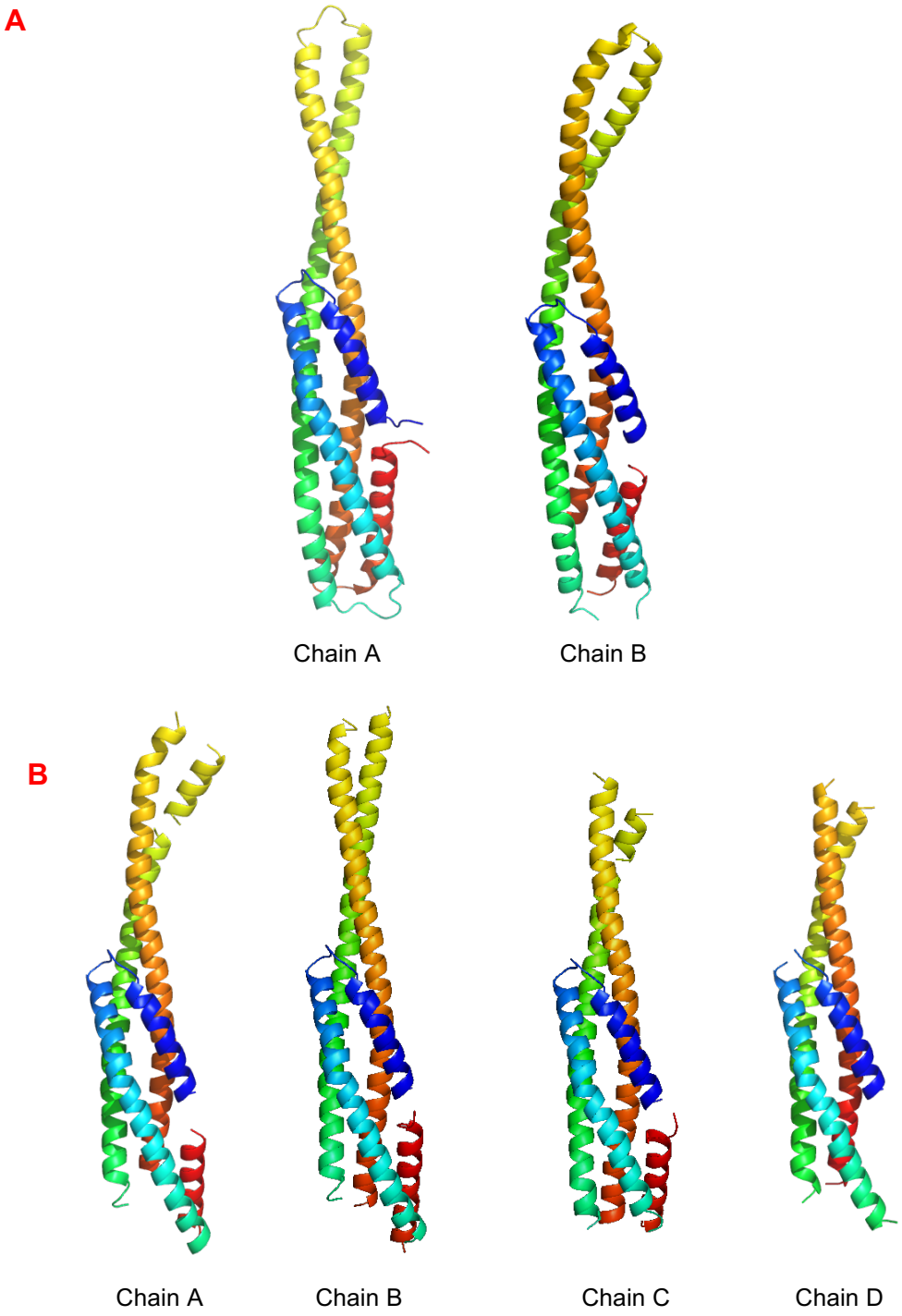


Figure 6.16. Each AhlC molecule present in both crystal form 1 and crystal form 2. Each chain present in crystal form 1 (**A**) and crystal form 2 (**B**) are shown as cartoon representations. Overall each chain contains the same secondary and tertiary structural features, although the degree to which each chain could be rebuilt varied quite considerably. In each case, the disorder of the head domain correlated with disorder in the tail domain for which the head packs i.e. neighbouring heads and tails in the tetramer assembly are either both ordered or both disordered.

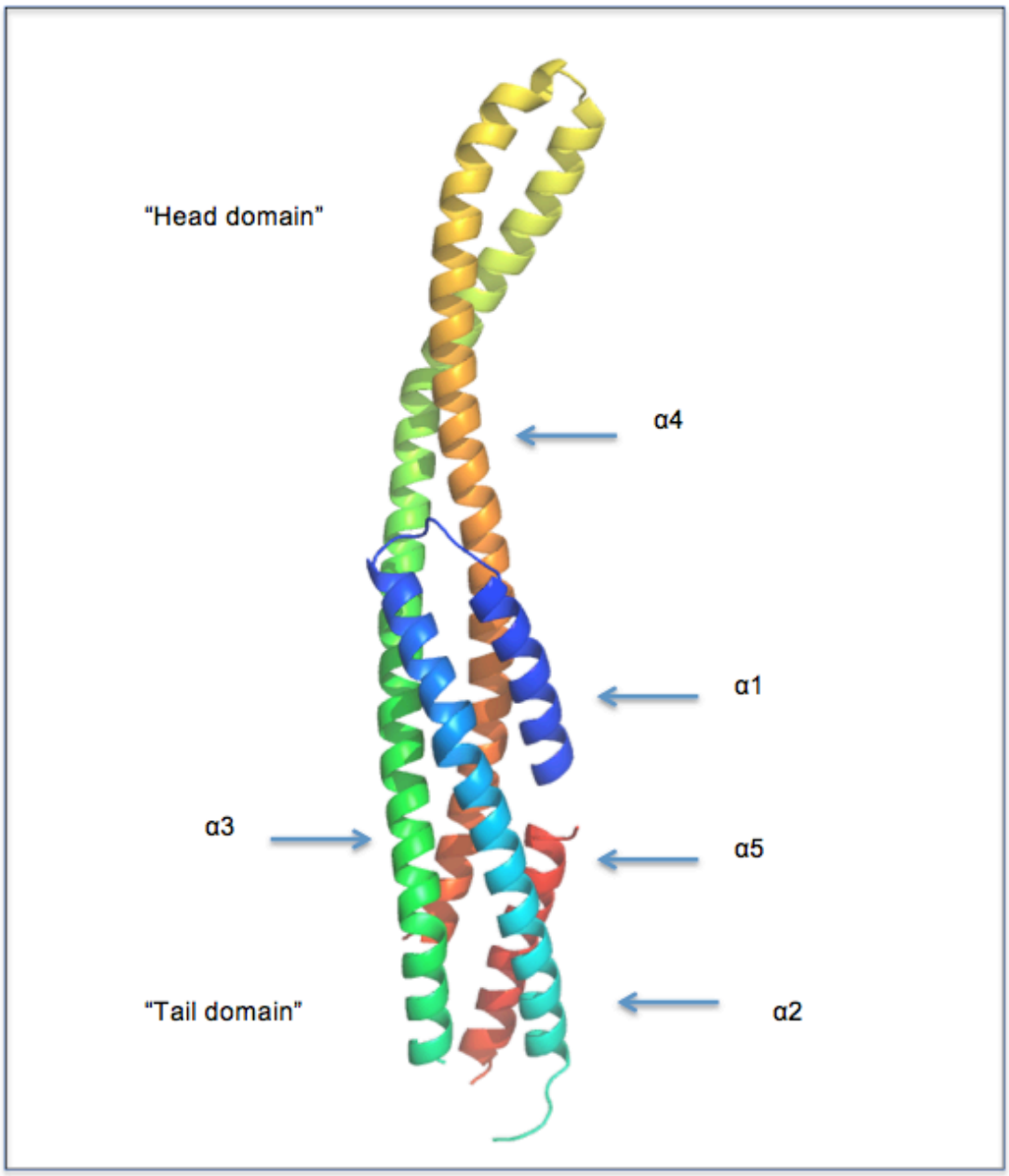


Figure 6.17. Cartoon representation of chain A and chain B in the JCSG+ B8 model after building and refinement. Cartoon representation of AhlC coloured from blue to red indicating N-terminus to C-terminus. The chain shown is chain B from the $P6_522$ structure. The helices are labelled 1 – 5 from the N-terminus to the C-terminus. Also highlighted are the head and tail subdomain, observed in a number of related toxins.

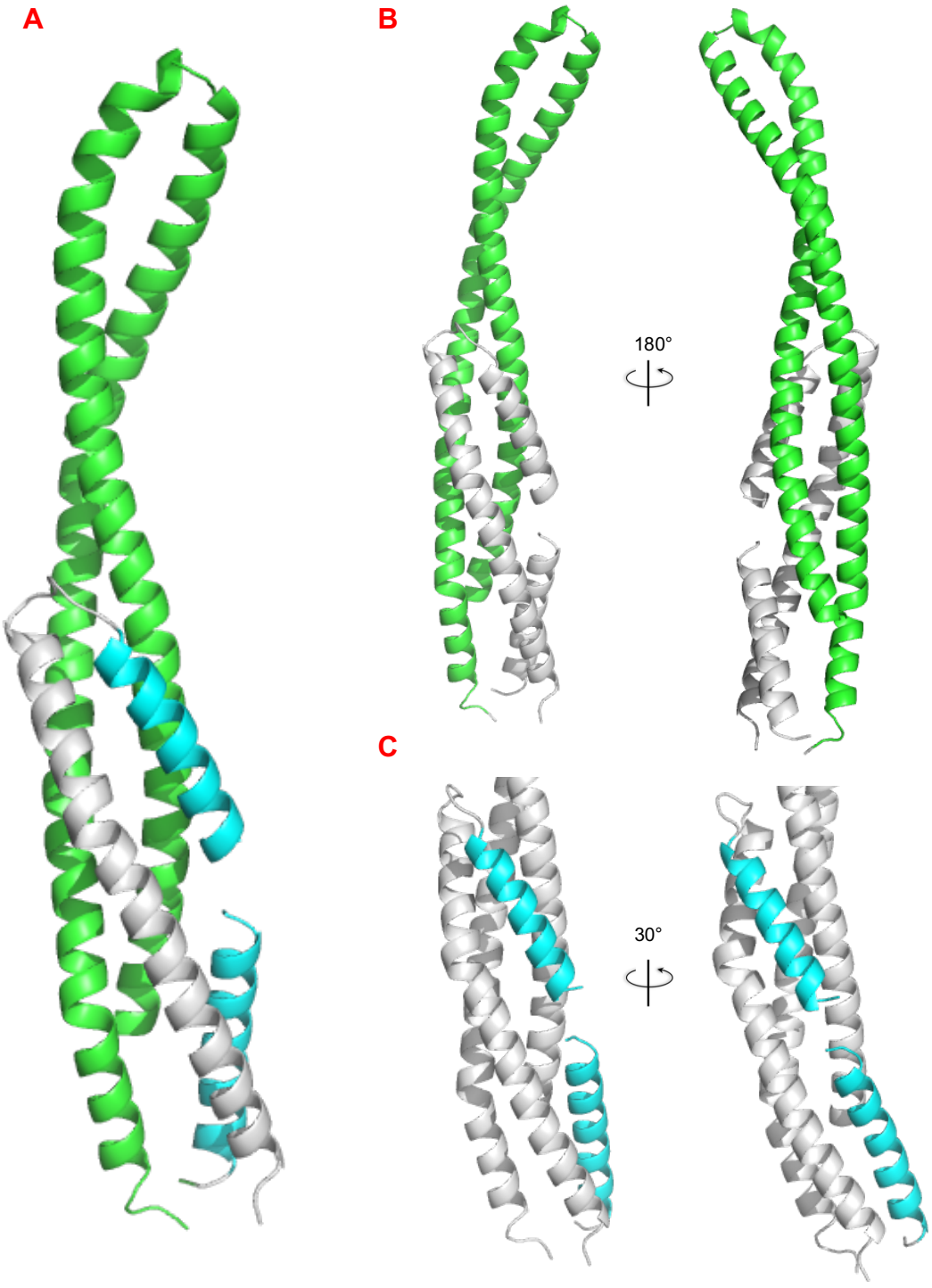


Figure 6.18. Structural features of AhlC. (**A**) Ribbon representation of AhlC coloured grey, with two structural features of the protein, the stalk and the N and C termini coloured green and cyan respectively. (B) Ribbon representation of the stalk of AhlC. Two elongated α helices run the entire length of the protein and form a hairpin at the top of the protein. (C) Ribbon representation of the tail domain of AhlC, with the N and C termini coloured cyan. As each terminus is shorter than seen for other ClyA family PFTs, the termini do not cross, and so the tail consists only of a 4 helical bundle.

helices. However, the five-helical bundle seen in the other family members is not present due to shortened N and C-termini in AhlC, which do not overlap to form the five helical bundle, but rather meet end to end forming a four helix bundle (Figure 6.18).

The helical hairpin of AhlC contains 19 exclusively hydrophobic residues (LYLLGLGLLGLPGLIALAA) that form a hairpin between L158 to L160. As such, the length of the hairpin on either side is ~15 Å (Figure 6.19). This length is similar to the ClyA hydrophobic hairpin (~16 Å) in the ClyA pore protomer. The head domain of Chain B is completely present in the electron density, whereas the electron density of chain A is somewhat ambiguous, and so a poly-ALA chain was built. The stalk that leads to the head domain contains 11 leucine and isoleucine residues, that all face inwards to the interface between the two helices α3 and α4, forming a leucine zipper. The pattern of leucines and isoleucines is likely to lead to the twisting of the helices (Figure 6.20). This formation of a leucine zipper is common in coiled coils of α-helices, and offers stability to the stalk domain (Barta et al., 2012a).

6.2.2 Biological assembly of AhlC

The structures of AhlC in the two different crystal forms were analysed using the PISA server (Krissinel and Henrick, 2007) and the CCP4 program Areaimol (Lee and Richards, 1971), to determine if the oligomeric states present in the structures could be biologically relevant. Analysis of the JCSG+ B8 structure in $P6_522$ with two molecules in the asymmetric unit suggested three different oligomeric states that would be stable in solution. These states are shown in Table 6.7. The largest oligomeric state is an A_2B_2 tetramer, made up of four chains from two adjacent asymmetric units (Figure 6.21). The delta free energy of this complex formation is predicted to be largely negative (-154.8 kJ mol^{-1}). The second largest assembly is an AB dimer, which has a predicted free energy on formation of -45.6 kJ mol^{-1}. The final assembly predicted for the B8 structure is a B_2 dimer. The buried surface area for this assembly is 10 % (2500 $Å^2$ buried surface area of 24600 $Å^2$ total area). The A_2 dimer is not determined by PISA to be sufficient to be stable in solution, which may be due to the movement of the hydrophobic head region. The total area of contact for each chain was determined using the Areaimol program to be 3295 Å of 11130 Å for chain A (30 %), and 3101 Å of 10480 Å for chain B (29.6 %).

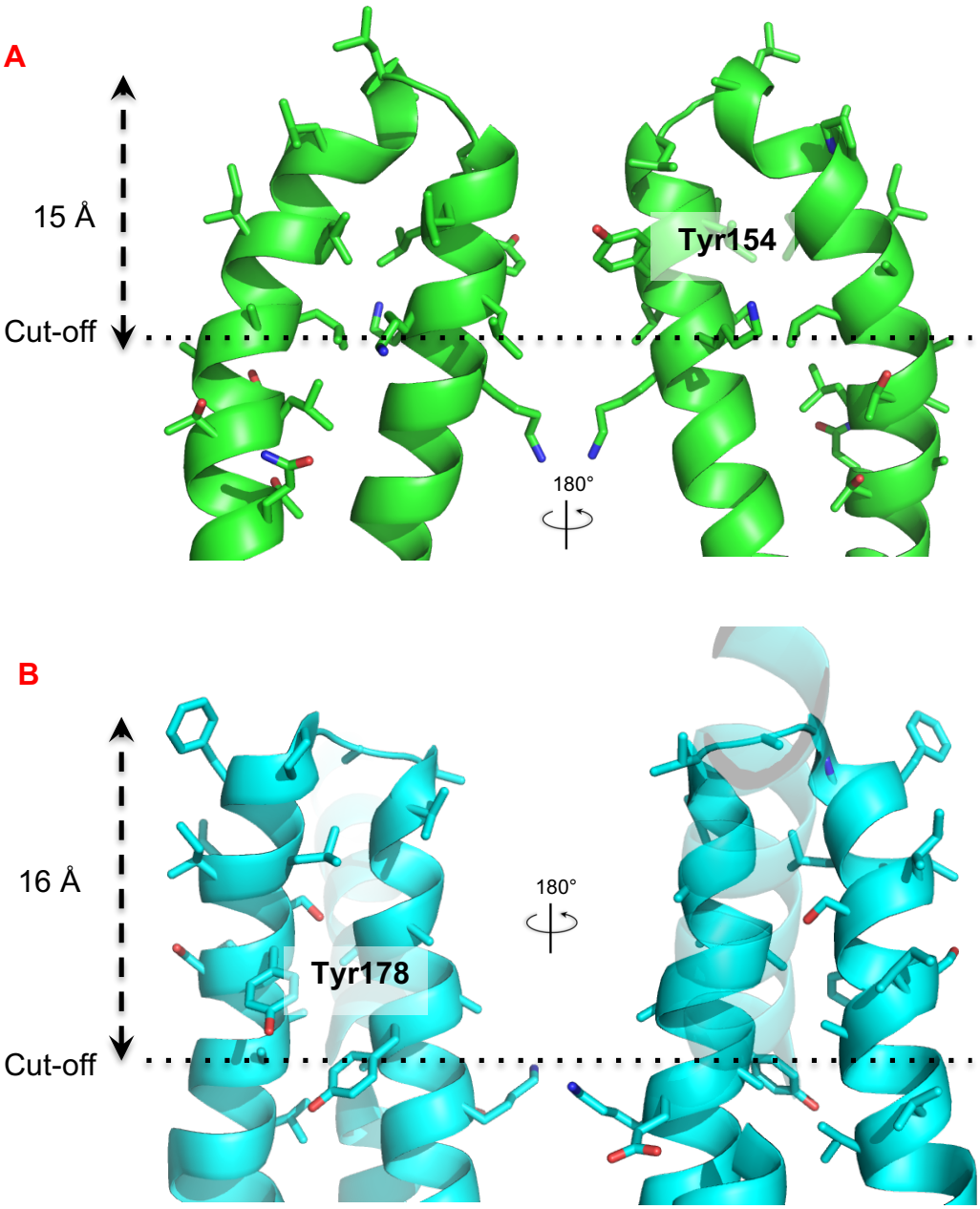


Figure 6.19. Comparison of the head domains of AhlC and ClyA pore form (PDB accession 2WCD). The head domain of AhlC is shown as a cartoon representation with sidechains present (**A**), which are used to predict a maximum amount of the protein able to insert into a membrane. From the tip of the head to the base, where charged residues are present, there is a vertical height of 15 Å. This 15 Å is similar to the head domain hairpin present in the pore form structure of ClyA (**B**). Again, from the tip of the hairpin to the base where charged residues are present is ~16 Å, similar to that of AhlC. In both structures the cutoff has been placed one helix turn from a tyrosine residue.

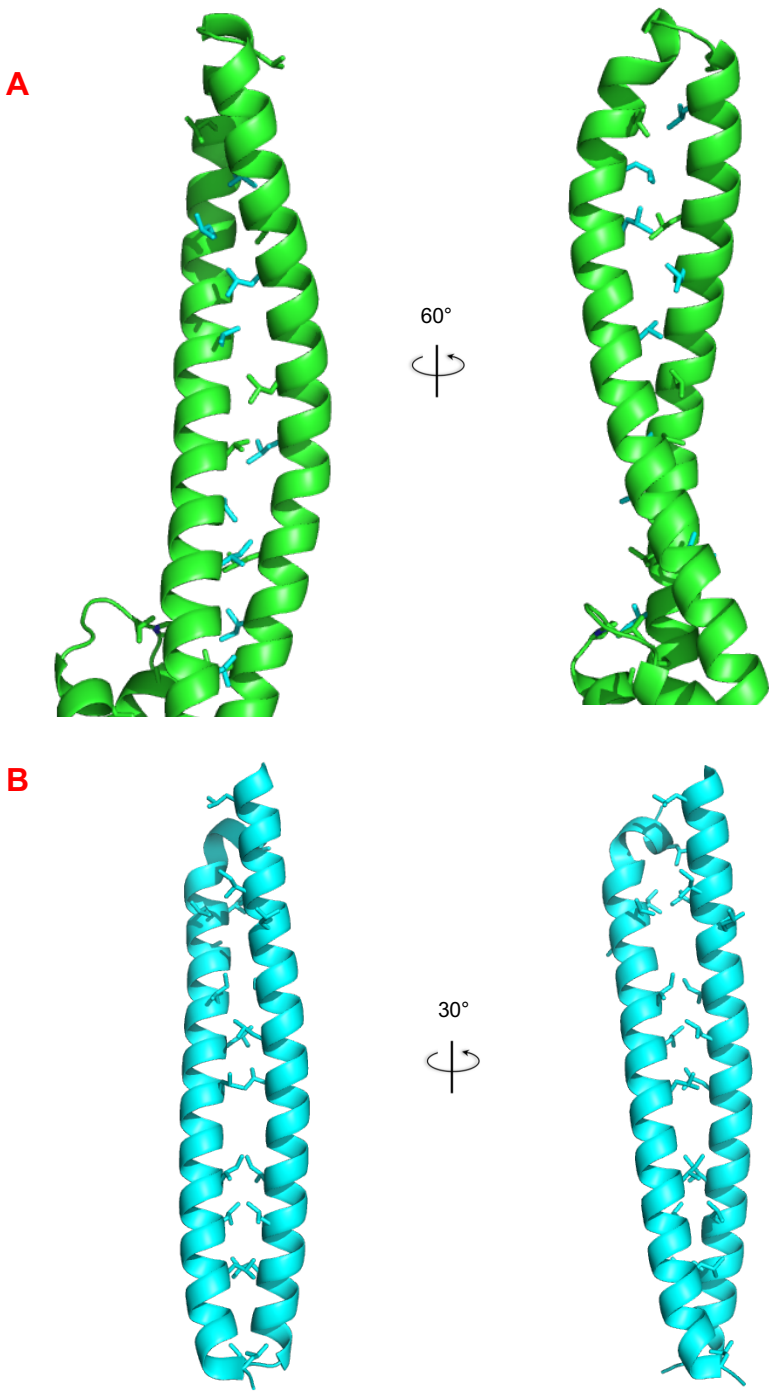


Figure 6.20. Leucine zipper contained within the stalk and head domain of AhlC, and an example of another Leucine zipper containing protein cGMP-Dependent Protein Kinase (PKG) I alpha. The stalk and head domain of AhlC contain a Leucine zipper motif made up of Leucine and isoleucine residues (**A**). Leu and Ile residues conserved in all C component proteins are shown in green, whereas those not conserved are highlighted in green. For comparison, the Leu and Ile residues contained in PGKI are shown (**B**), which form a similar twisted Leucine zipper coiled coil domain.

A

Assembly	Surface area (Å)	Buried area (Å)	ΔG^{int}	ΔG^{diss}				
A_2B_2	44344.5	9627.9	-37.0	10.1	N_{HB}	N_{SB}	N_{DS}	CSS
B + A		1512.3	-10.9		6	5	0	1
B + B		1239.2	-11.9		10	8	0	1
A + A		550.1	-3.3		4	6	0	0.114

B

Interface	Hydrogen bonds	Salt bridges
A<>B	## Structure 1 Dist. [Å] Structure 2 1 B:GLN 8[NE2] 3.18 A:GLY 185[O] 2 B:SER 22[OG] 2.80 A:THR 19[OG1] 3 B:LYS 152[NZ] 3.08 A:THR 256[OG1] 4 B:LYS 152[NZ] 2.60 A:GLU 257[OE2] 5 B:GLU 184[OE1] 2.95 A:ARG 217[NH2] 6 B:GLY 185[O] 2.92 A:GLN 8[NE2]	## Structure 1 Dist. [Å] Structure 2 1 B:LYS 152[NZ] 3.73 A:GLU 257[OE1] 2 B:LYS 152[NZ] 2.60 A:GLU 257[OE2] 3 B:GLU 184[OE1] 3.77 A:ARG 217[NH1] 4 B:GLU 184[OE1] 2.95 A:ARG 217[NH2] 5 B:GLU 184[OE2] 3.60 A:ARG 217[NH2]
A<>A	## Structure 1 Dist. [Å] Structure 2 1 A:ARG 122[NH1] 2.95 A:GLU 126[OE1] 2 A:ARG 122[NH2] 3.25 A:GLU 126[OE1] 3 A:GLU 126[OE1] 2.95 A:ARG 122[NH1] 4 A:GLU 126[OE1] 3.25 A:ARG 122[NH2]	## Structure 1 Dist. [Å] Structure 2 1 A:ARG 122[NH1] 2.95 A:GLU 126[OE1] 2 A:ARG 122[NH2] 3.25 A:GLU 126[OE1] 3 A:ARG 122[NH2] 3.79 A:GLU 126[OE2] 4 A:GLU 126[OE1] 2.95 A:ARG 122[NH1] 5 A:GLU 126[OE1] 3.25 A:ARG 122[NH2] 6 A:GLU 126[OE2] 3.79 A:ARG 122[NH2]
B<>B	## Structure 1 Dist. [Å] Structure 2 1 B:ARG 122[NH1] 2.66 B:GLU 126[OE1] 2 B:ARG 122[NH2] 3.44 B:GLU 126[OE2] 3 B:GLN 128[NE2] 2.67 B:GLN 121[OE1] 4 B:ASN 223[ND2] 3.31 B:GLN 143[O] 5 B:ASN 223[ND2] 2.85 B:ASP 146[OD2] 6 B:GLU 126[OE1] 2.66 B:ARG 122[NH1] 7 B:GLU 126[OE2] 3.44 B:ARG 122[NH2] 8 B:GLN 121[OE1] 2.67 B:GLN 128[NE2] 9 B:GLN 143[O] 3.31 B:ASN 223[ND2] 10 B:ASP 146[OD2] 2.85 B:ASN 223[ND2]	## Structure 1 Dist. [Å] Structure 2 1 B:ARG 122[NH1] 2.66 B:GLU 126[OE1] 2 B:ARG 122[NH1] 3.67 B:GLU 126[OE2] 3 B:ARG 122[NH2] 3.85 B:GLU 126[OE1] 4 B:ARG 122[NH2] 3.44 B:GLU 126[OE2] 5 B:GLU 126[OE1] 2.66 B:ARG 122[NH1] 6 B:GLU 126[OE2] 3.67 B:ARG 122[NH1] 7 B:GLU 126[OE1] 3.85 B:ARG 122[NH2] 8 B:GLU 126[OE2] 3.44 B:ARG 122[NH2]

Table 6.7. PDBePISA analysis of AhlC JCSG+ B8 A_2B_2 assembly. (**A**) The assemblies predicted by PDBePISA, and the combined surface area, buried surface area, the predicted free energy of association and dissociation, and the number of interactions formed between each monomer in the assembly. (**B**) shows each assembly interface and highlights the residues and types of interaction involved in creating the A_2B_2 assembly, including hydrogen bonding and salt bridges.

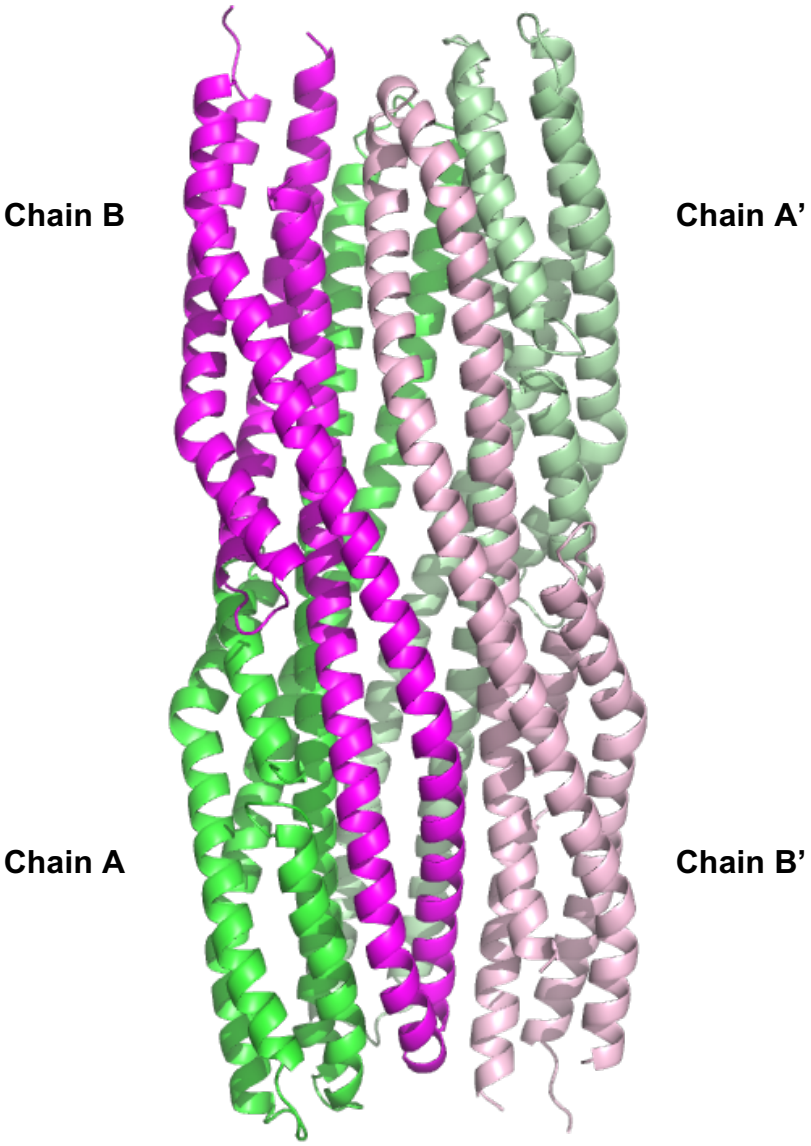


Figure 6.21. The A_2B_2 tetramer assembly generated by PDBePISA. The output from PDBePISA containg 2 chains of B (Magenta and pink), and two chains of A (Green and mint). Each interface consists of a head to tail packing, generating an overall pseudo-222 symmetry. The hydrophobic head in each B chain and B' chain packs between the tail of the two flanking monomers B' and A', and B and A respectively. The hydrophobic head on each A chain does not pack between the tails of each flanking chain, and instead is solvent exposed.

Analysis of the PACT D7 structure of AhlC in $P2_1$ with four molecules in the asymmetric unit fails to highlight the same assemblies by PISA analysis, although AreaImol highlights the same level of contact area for each chain in the D7 structure (30 % contact area for each chain). Regardless of these analyses, it is obvious that the tetrameric assembly present in crystal form 2 is identical to that found and highlighted in crystal form 1. These identical assemblies present in both crystal forms with different packing and crystal contacts, together with the gel filtration analysis that suggest a dimer or tetramer is present, show that this tetramer most likely to be the biologically relevant multimer.

6.2.3 The nature of the biological tetramer of AhlC

The nature and the identity of the residues in the oligomeric interface between each molecule in the B8 tetramer can be analysed, and show that many of the interactions between molecules are electrostatic (Table 6.7). There is a pseudo two-fold rotational symmetry that lies in the center of the two dimer interfaces, creating a pseudo-222 symmetry. The tetramer can be described as being made up of two A and two B units, creating two AB interfaces, an AA interface, and a BB interface, with the AA and BB being identical in crystal form 1, or slightly different in crystal form 2.

Each interface is analysed in Table 6.7, highlighting the residues involved in forming the oligomeric states. Considering the high percentage of buried surface formed upon tetramer assembly, only relatively few residues are involved in hydrogen bonding or salt bridge interactions. There is an asymmetry to the interactions, as well as a few residues that form symmetrical interaction, in that two neighbouring chains share the same interactions. This includestwo specific Arg – Glu salt bridge interactions, those between Arg217 and Glu184, and between Arg122 and Glu126. The Arg217 and Glu184 interaction is seen in the AB interface, whereas the Arge122 and Glu126 salt bridge is formed between the AA and BB interfaces. The assymetric Arg217 and Glu184 interaction is only seen in between the molecule for which the hydrophobic head is inserted into the tetramer, and is partially conserved between homologues suggesting this may have a role on tetramer assembly or dynamics.

6.2.4 The interface between the tetramer inserted head with the two adjacent sandwiching tail domains

There are also interactions formed between the hydrophobic head domain of chain B and the two sandwiching molecules either side (Figure 6.22). The hydrophobic head domain of AhlC contains a majority of leucine residues, namely Leu156, Leu158, Leu160, Leu163, and an isoleucine residue Ile167, which are predicted to form the hairpin, which inserts into one leaflet of the membrane. The leucine and isoleucine residues present in the head are arranged as shown in Figure 6.22, with Leu158 facing to the tail domain of the symmetry related chain B, and Leu 156, Leu160, Leu163, and Ile 166 facing the tail domain of the symmetry related chain A. These residues form hydrophobic leucine zipper interactions, and stabilise the head domain of chain B against the two tails sandwiching the head. Tyr154 in the head domain of chain B also forms pi-stacking interactions with Phe250 of the symmetry related B chain, and Lys152 of chain B forms ionic interaction with Glu257 of chain A, which is highlighted in the PISA analysis. Overall, these interactions create a stable conformation, as also indicated by the B-factors in this region.

In contrast, the sidechains present in the tail domains are positioned such that they are not able to form the same interactions with the non-inserted chain A head domain. The asymmetry in the packing of the heads means that in both crystal forms, one head is sequestered by the tail domains, and the other is either completely disordered, as in crystal form 2, or is present in the electron density, as in crystal form 1, due to crystal contacts between asymmetric units (Figure 6.23). This forced asymmetry may therefore have a function in making one head domain available for membrane binding. Binding at the membrane of this exposed head may then allow the tetramer to disassemble, localising the other three molecules to the solvent membrane interface.

6.2.5 Analysis of conserved residues

The C components of the tripartite toxin system discussed in 4.5.8 were aligned, highlighting conserved residues across all family members (Figure 6.24), and then residues with a conservation score (Livingstone and Barton, 1993) above 7 were mapped onto a ribbon representation of AhlC (Figure 6.25). The majority of the conserved residues are concentrated at the junction between the N-terminal helix, facing inwards to the core of the helical bundle, and in the coiled coiled helices of the stalk and head domain of the protein

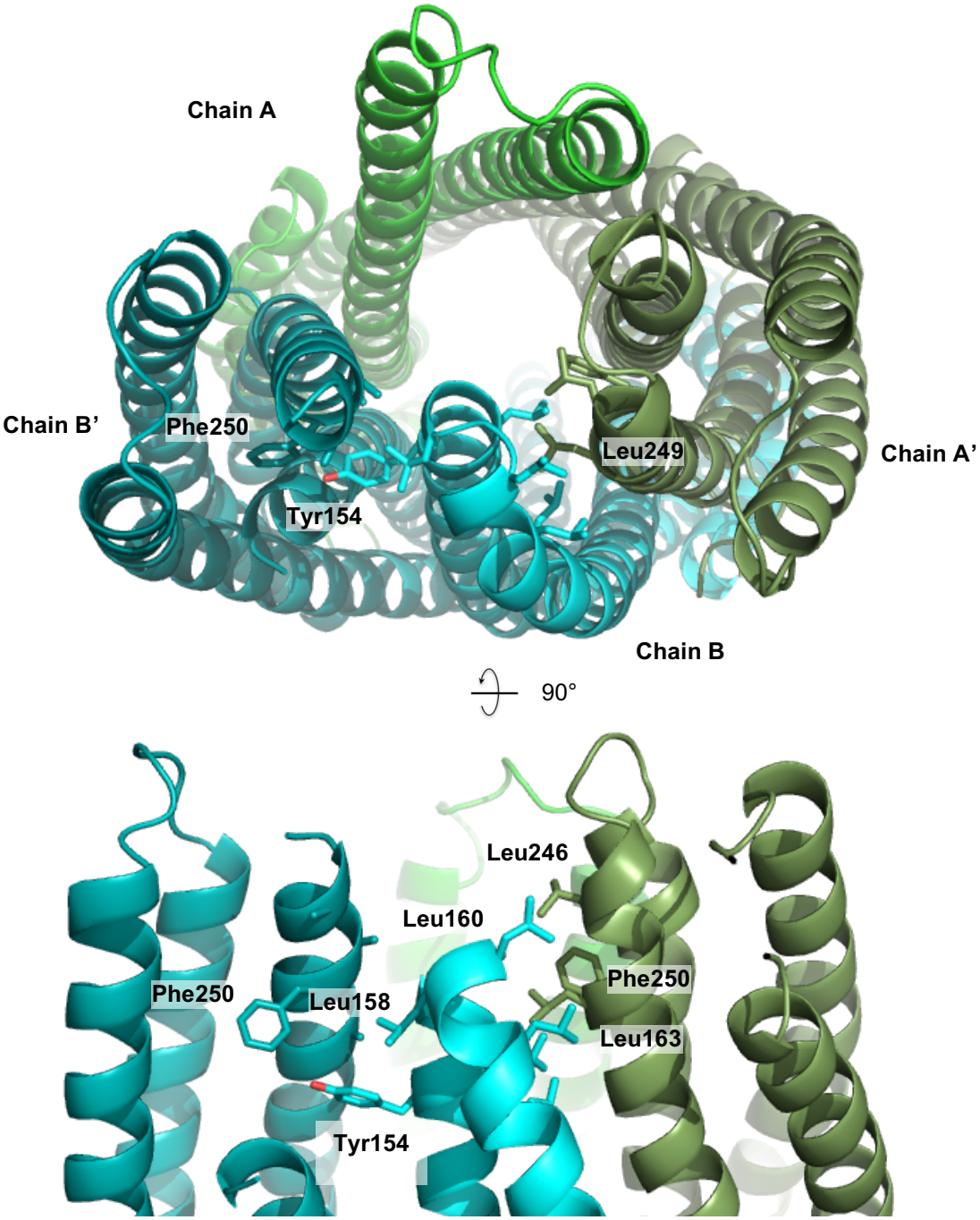


Figure 6.22. Interactions of the Chain B head with the sandwiching tail domains of the tetramer. Ribbon representation of the tetramer of AhlC. The head domain of chain B (cyan) packs between the two symmetry related adjacent tails of chain B' and A'. Leu160 and Leu163 of chain B form interactions with Leu246 and Leu249 of chain A'. However, due to the asymmetry of the interactions, the equivalent residues do not pack in A and B'. Chain B head also forms interactions with the tail of chain B', to form a hydrophobic environment for the hydrophobic hairpin.

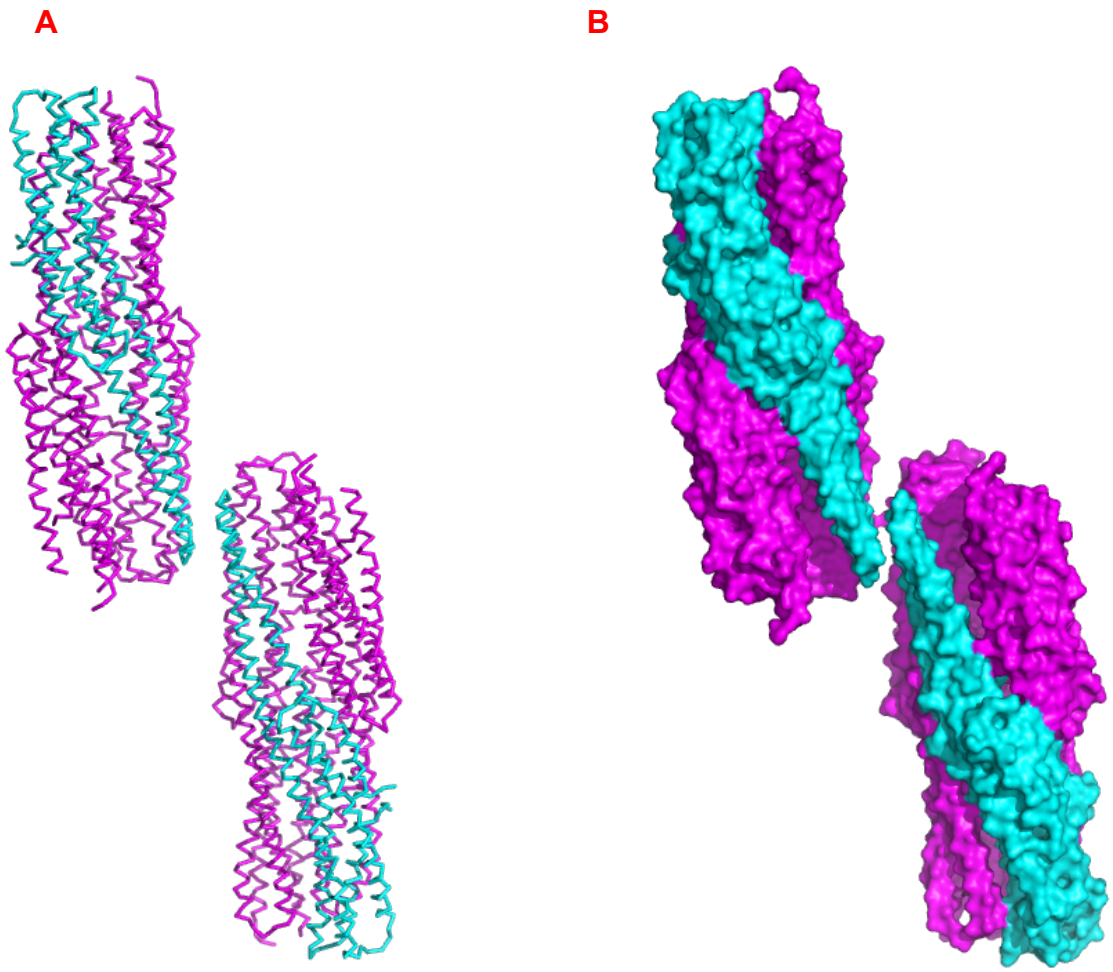
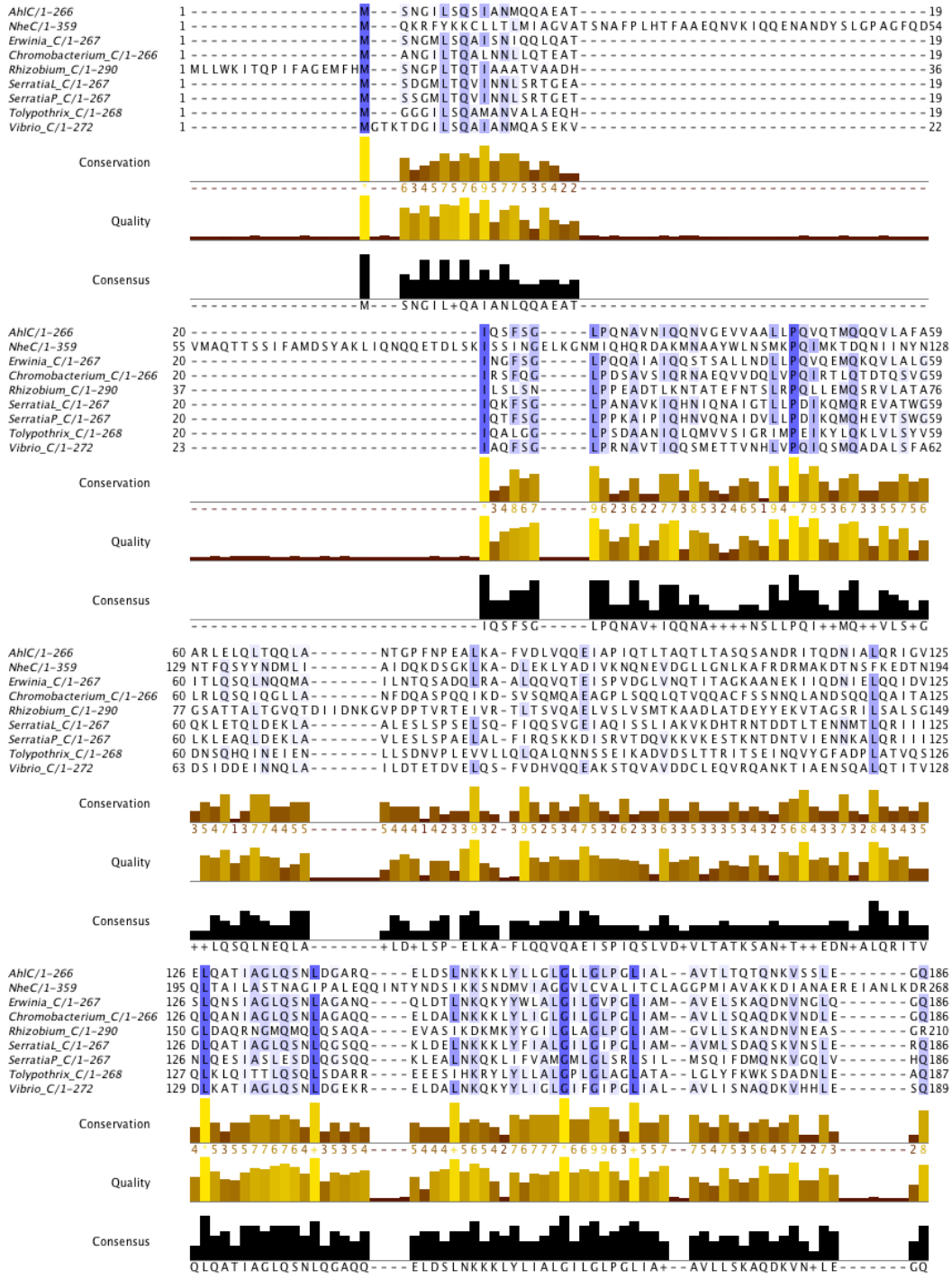


Figure 6.23. The crystal packing of head domains from neighbouring tetramers. (**A**) A ribbon representation of two neighbouring tetramers generated by crystal symmetry. Highlighted in cyan are two chains of AhlC in which the hydrophobic heads twist away from the assembly and instead interact with the other symmetry related molecule in a neighbouring tetramer. **(B)** Van der waal surface representation of the same assembly, showing the proximity of each head in neighbouring tetramers.

Figure 6.24. The sequence alignment between each C component of the tripartite pore forming toxin systems.



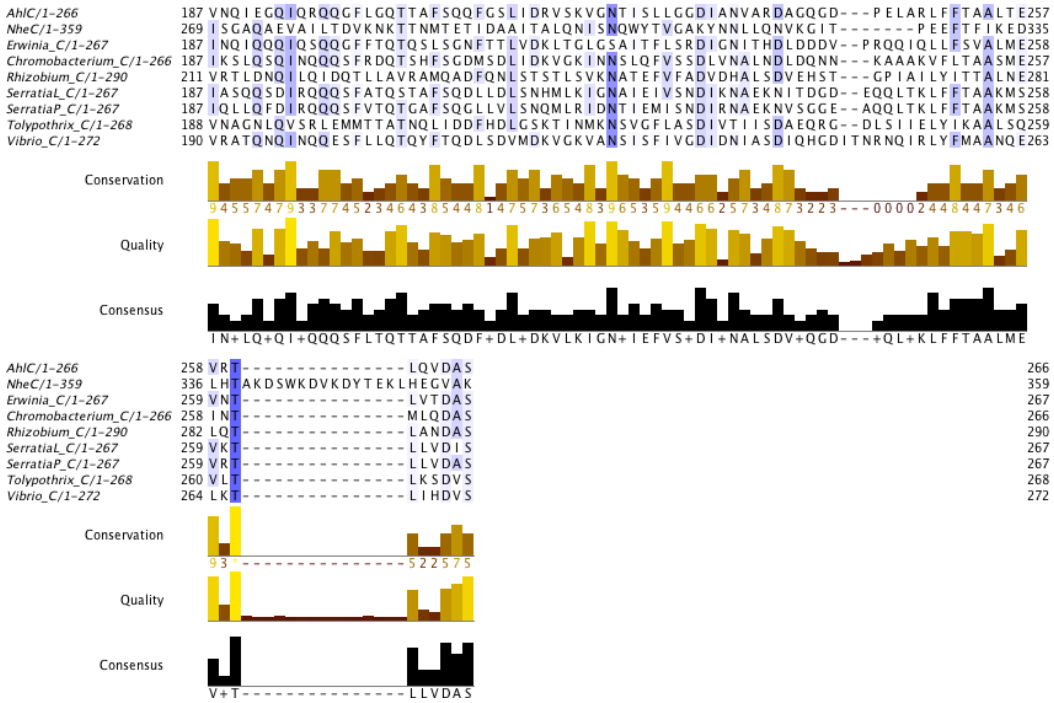


Figure 6.24. The sequence alignment between each C component of the tripartite pore forming toxin systems. The amino acid sequences of the C components of *Aeromonas hydrophila* (AhlC), *Bacillus cereus* (NheC), *Erwinia mallotivora* (Erwinia_C), *Chromobacterium piscinae* (Chromobacterium_C), *Rhizobium sp. Leaf262* (Rhizobium_C), *Serratia liquefaciens* (SerratiaL_C), *Serratia plymuthica* (SerratiaP_C), *Tolypothrix sp. CC 7601* (Tolypothrix_C), and *Vibrio campbellii* (Vibrio_C) were aligned using the Jalview software (Waterhouse et al., 2009). Each sequence is aligned with the strongly conserved residues highlighted in purple. Beneath the sequence alignment is a bar graph showing the number of sequences for which a residue is conserved, with more highly conserved residues shown a brighter yellow. Finally a consensus sequence for the aligned amino acid sequences is given, with a + symbolising no consensus for the given residue number.

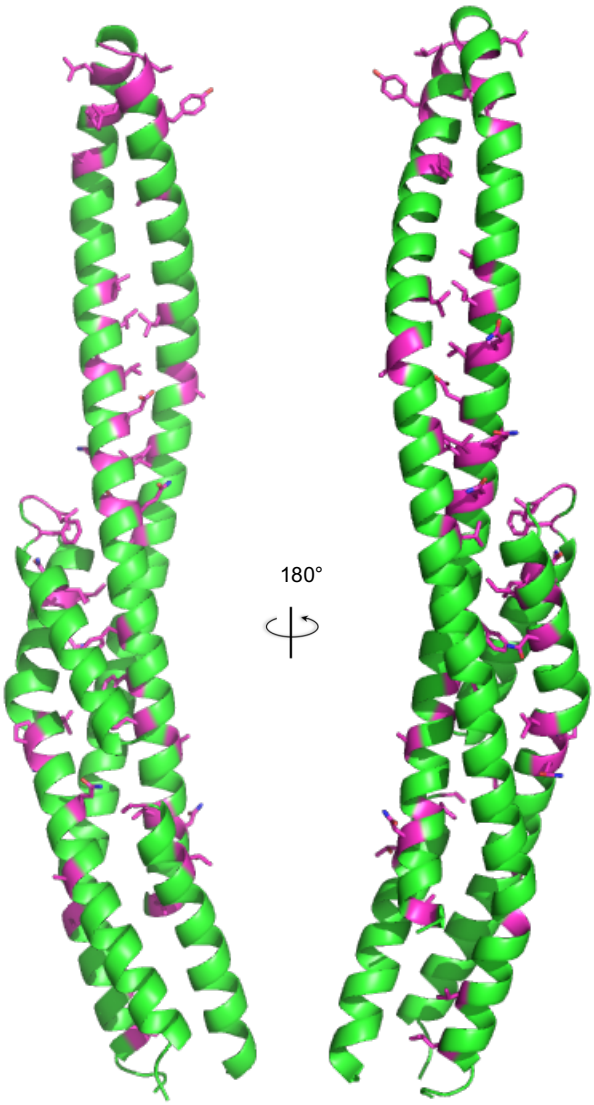


Figure 2.25. Mapping the conserved residues of the C components onto the structure of AhlC. Ribbon representation of AhlC in green, with conserved residues (above 7 for the conservation score in T-Coffee) highlighted in magenta. Areas of conservation include the head domain helices, and the intersection between the head and tail domain.

(Figure 6.26). At the junction between the N-terminal helix and α2, a network of hydrophobic residues including Ile20, Phe23, Leu23, and Ile33 are conserved, as well as two carboxamide amino acids Gln34 and Asn36. The core helical bundle contains a number of bulky hydrophobic acids that aid in forming the bulged helical bundle, namely Phe207, Phe211, Leu44, Leu214. The kink in α2 is generated by the conserved Pro46. Finally the stalk and head contain a number of Leucine and isoleucine residues including Leu120, Leu138, Leu148, Leu156, Leu158, Leu163, Leu166, Val180, Leu183, Val 187, and Ile194 (Figure 6.27). These residues form a Leucine zipper with twists modulated by the irregularity of such residues. Also in this region are some outward facing polar residues with no obvious function in the current structure. These residues are Gln186, Glu191, Gln197 and Gln198, and may have a function in pore formation, and so would not be particularly insightful in the soluble tetramer structures.

6.3 Comparison of AhlC with other ClyA-like PFT components

As the structure of AhlC showed a number of similarities to structures of other ClyA-like pore forming toxins, AhlC was compared with each known ClyA-like PFT component to try to determine a function for the protein. Chain B was superposed with each other ClyA family protein with a known structure, using Lsqkab, in order to determine similarities in the overall structures.

6.3.1 Structure Comparison of AhlC and ClyA in the soluble and pore form

The structure of AhlC and the structure of the soluble form of ClyA (PDB accession 1QOY) were superposed using Lsqkab in CCP4 in order to compare the two structures (Figure 6.28). The average rmsd over 138 residues was 3.8 Å, with distances of aligned α carbon atoms ranging from 0.8 Å to 7.7 Å. The overall alignment shows that the structures are similar in the central part of the protein, where both AhlC and ClyA contain a four-helical bundle that align well. However, the N-terminal helix in ClyA is much longer, and goes on to form a five-helical bundle at the bottom of the tail, not present in AhlC. This N-terminal helix in ClyA packs between α4 and α5 separating them, and the alignment to AhlC is poorer in this region. The lack of a long N-terminal region in AhlC is associated with a twisting of α4 relative to the corresponding stalk helix in ClyA, so that the tail end of α4 aligns with the N-

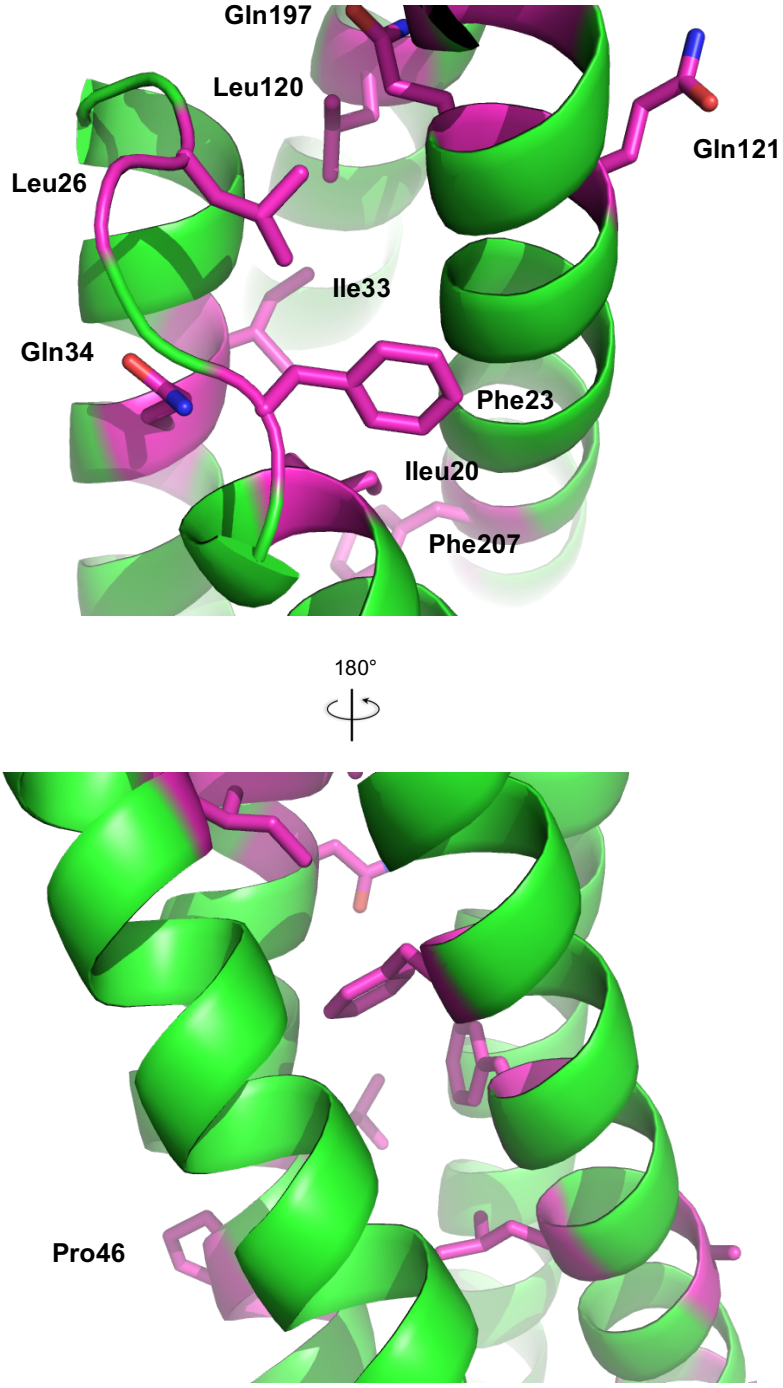


Figure 6.26. Conserved residues in the tail domain 4-helical bundle. Ribbon representation of AhlC, highlighting the conserved residues in the core of the protein at the intersection between the head and tail domain. Many of the conserved residues are hydrophobic and face inwards to the core of the 4-helical bundle. However, a few residues including Gln34, Gln121, and Gln197 are conserved within the family but have no obvious function.

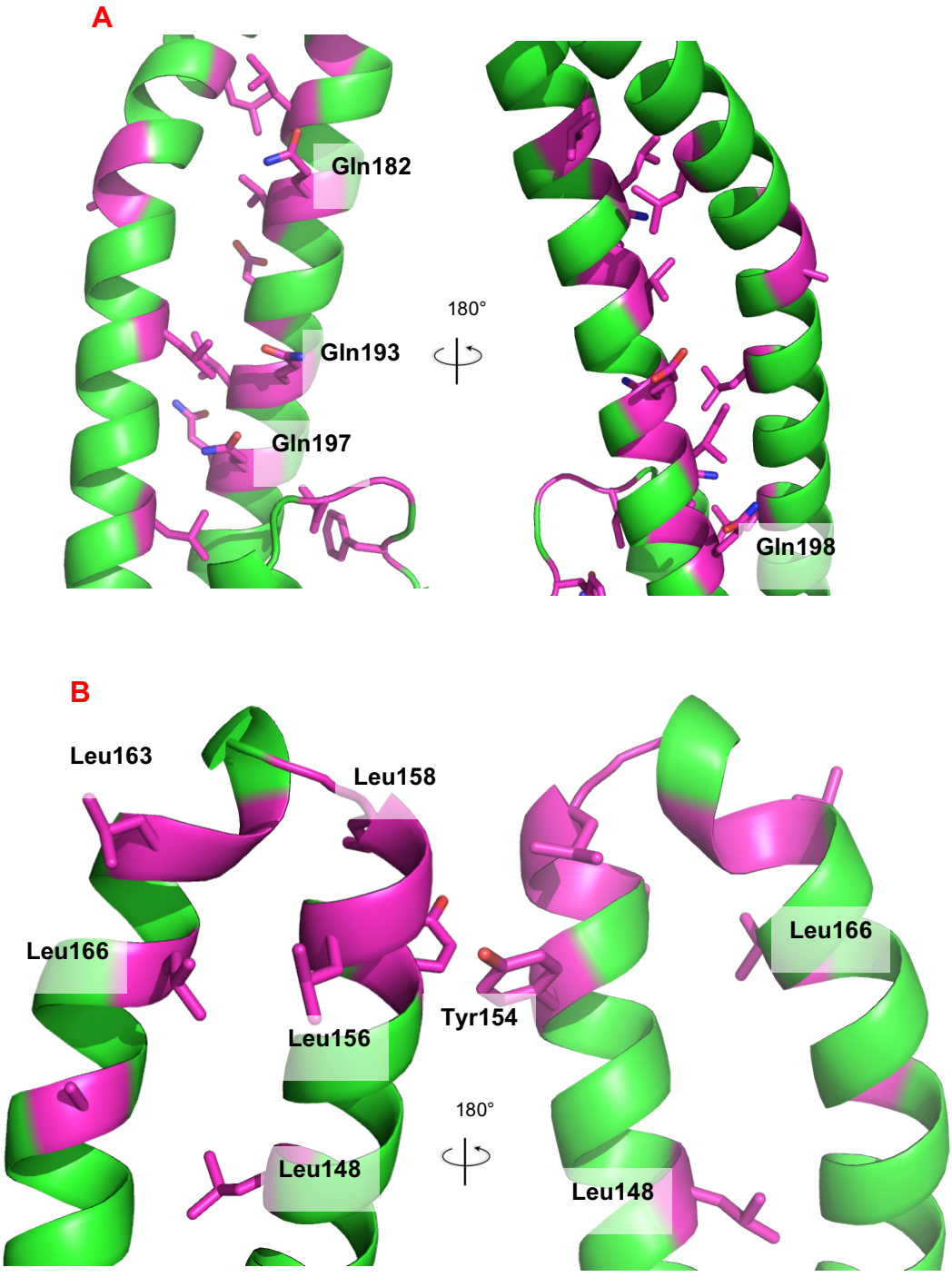
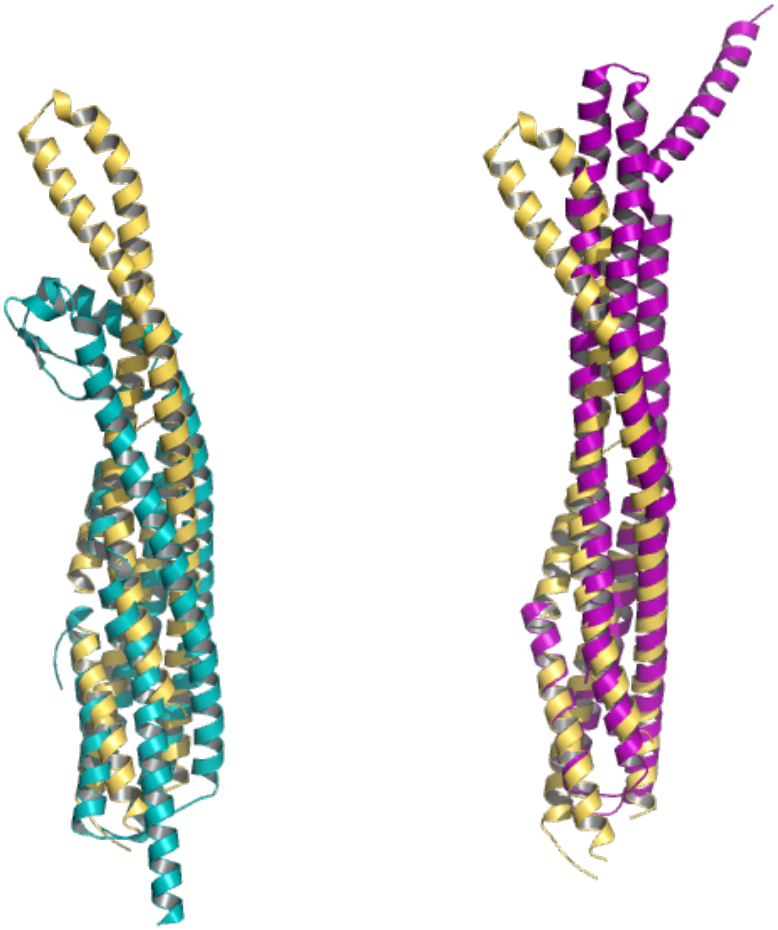


Figure 6.27. Conserved residues in the stalk and head domain helices of AhlC. Ribbon representation of the stalk ad head domain helices of AhlC, with family conserved residues coloured magenta. (**A**) A face of Gln residues is formed on α4 by Gln186, Gln193, Gln197, and Gln198. The other conserved residues are leucines involved in Leucine zipper interactions. (**B**) The head domain contains a number of conserved Leucine residues, and a family conserved tyrosine154.



ClyA Soluble form (1Q0Y) ClyA pore form (2WCD)

Figure 6.28. Superposition of AhlC with ClyA in the soluble (Left) and pore (Right) form. Cartoon representation of AhlC superposed with ClyA using Lsqkab with secondary structure matching. Both ClyA structures (Left soluble form and right pore form) superpose well with the soluble form rmsd 3.8 Å over 138 residues, and the pore form rmsd 3.2 Å over 182 residues.

terminus of ClyA, and the N-terminus of AhlC aligns with the top of the N-terminus of ClyA (Figure 6.29).

AhlC was also superposed with the pore form of ClyA (PDB accession 2WCD)(Figure 6.28). Lsqkab was run with secondary structure matching, and showed that the rmsd between Chain B of AhlC and the pore form of ClyA was 3.2 Å over 182 residues. The increased number of equivalent residues in the structural alignment between AhlC and the pore form of ClyA is due to two conformational changes in ClyA making it more similar to AhlC. The first conformational change takes place in the β tongue region of soluble ClyA from a β tongue to an alpha helical hairpin in the pore form. The hydrophobic hairpin of ClyA adopts a conformation similar to the hydrophobic hairpin in AhlC. However, in chain B of AhlC crystal form 1, the tip of the helical hairpin in ClyA is twisted related to that of the AhlC hairpin, preventing a close overlap between the two proteins in this area. Chain A of AhlC is more untwisted, and so aligns better with the tip of the hydrophobic head of ClyA, suggesting that the solvent exposed AhlC hydrophobic head in the tetramer more closely resembles the pore form of ClyA than the B chain does. The second conformational change in ClyA that leads to a better alignment of ClyA pore form and AhlC is the N-terminal helix transition in ClyA, which changes from forming part of the five helix bundle to forming the membrane inserted kinked helix. This movement of the N-terminal helix allows α4 and α5 to come closer together to form a four-helix bundle, which aligns well with the four-helical bundle present in AhlC (Figure 6.29).

Therefore, AhlC can be thought of as a hybrid between the soluble and pore form of ClyA, in that the hydrophobic hairpin of AhlC is in an alpha helical conformation, similar to that of the pore structure of ClyA. However, the N-terminal helix of AhlC is still part of the tail of the protein, similar to the same region in the soluble form structure of ClyA, although the shorter N-terminal helix in AhlC means that the four-helical bundle aligns better with that of the pore form helical bundle.

6.3.2 Structure Comparison of AhlC with NheA

The structure of AhlC and NheA had the lowest rmsd between all of the structures superposed with AhlC, with an rmsd of 2.8 Å over 195 residues. The increased alignment of NheA and AhlC compared with ClyA and AhlC is due to the improved fit of the main body of

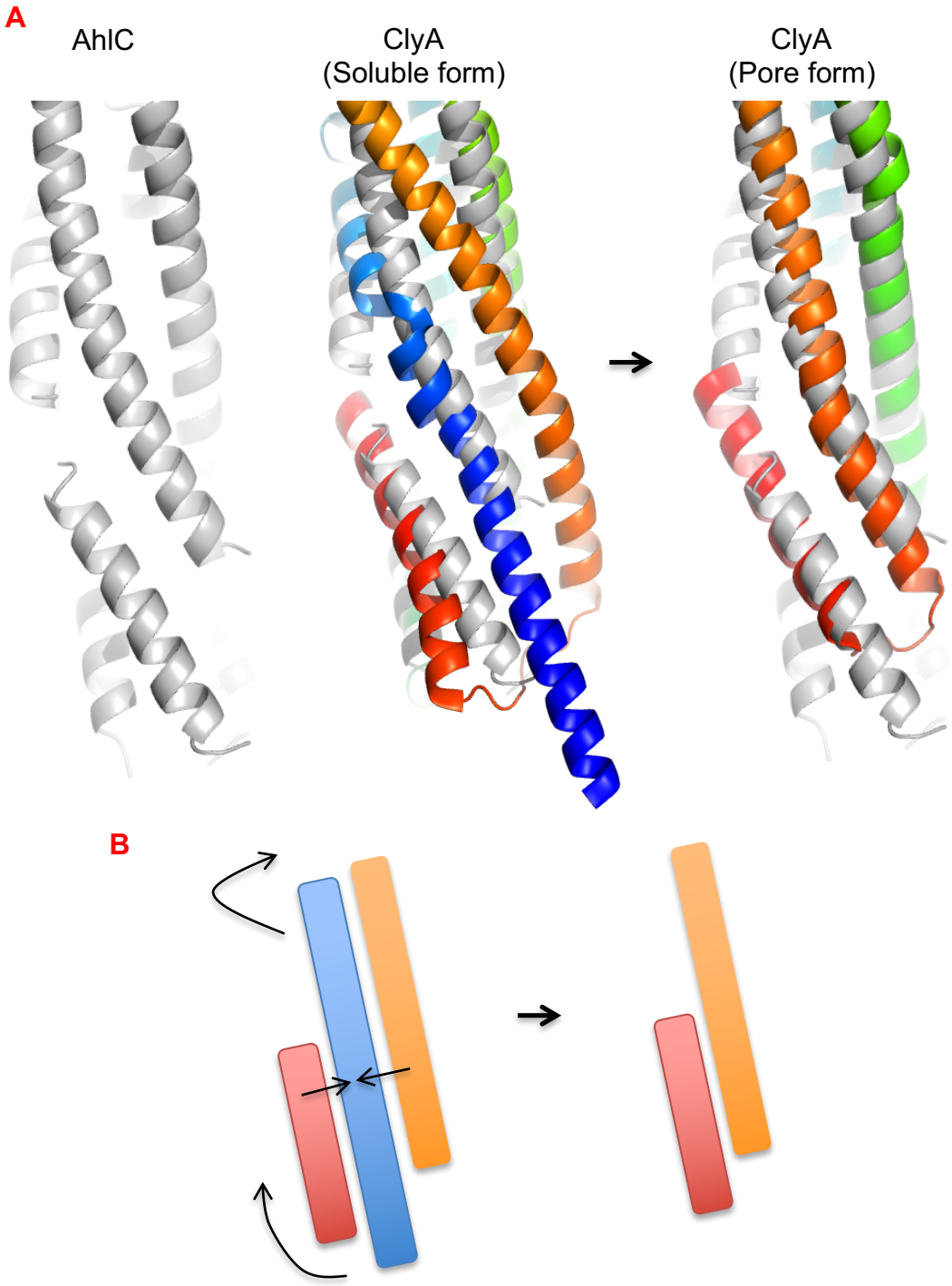


Figure 6.29. Movement of ClyA tail from soluble to pore form and increase in the alignment between the AhlC and ClyA tails. AhlC contains a four-helical bundle (**A** left) that does not align well with the equivalent helices in the ClyA tail due to the interchelation of the N-terminal helix (blue)(**A** middle, AhlC in white and ClyA coloured). The conformational change of the N-terminal helix leads to a better alignment as α4 and α5 in ClyA come closer together (**A** right). (**B**) shows the conformational change of ClyA in a schematic representation, where the colour of helices corresponds to the structure of ClyA in A.

each protein, in which each of the alpha helices align very well (Figure 6.30). In NheA the N-terminal helix is more disordered than the same region in ClyA, and so forms a minimal region of five helical bundle in the tail domain. As the N-terminal helix does not run the length of the tail domain, α6 and α7 are closer, and so align well with AhlC. The C-terminus of NheA is also slightly longer, and forms part of the latch for the head domain, although the majority of the NheA N-terminal helix aligns well with AhlC α5.

The only major deviation between both structures is in the head region, in which NheA forms two folded alpha helices followed by the beta hairpin. Sequence alignment between the two proteins show that all 100 % conserved residues are found in the main body of the protein, with some sequence identity between the head of AhlC with the two head helices of NheA. This suggests that the two helices of NheA may undergo a conformational change to extend as in AhlC, to form part of an extended helix system in NheA. A conformational change of this kind would extend NheA to be the same length as AhlC and ClyA in the pore form.

6.3.3 Structure Comparison of AhlC with HblB

In the arrangement of each of the tripartite proteins in an operon, AhlC and HblB are both 'C' proteins. Therefore it would suggest that both of these proteins would have a similar function, as well as structure. To probe this assumption, the structures of the two were compared (Figure 6.31).

The most striking difference between these two proteins is the structure of the head domain in both. The head domain of HblB adopts the β tongue conformation, similar to other soluble forms of proteins in the tripartite family. This conformer means that the protein is solubilised without the need to form oligomers, as is the case for AhlC, and, based on the crystal structure and gel filtration data (Madegowda et al., 2008), HblB appears to instead form a dimer, again more similar to other proteins in the family (Ganash et al., 2013; Wallace et al., 2000). As HblB is supposed to be the closest homologue to AhlC, based on position in the operon, sequence similarity, hydropathy plot, and function in haemolytic activity, a large difference in the structures seen raises the question as to the structures of the other C components. Based on the sequence of the AhlC homologues found in Gram-negative bacteria, it would seem that the structure of these proteins would form similarly to AhlC, and

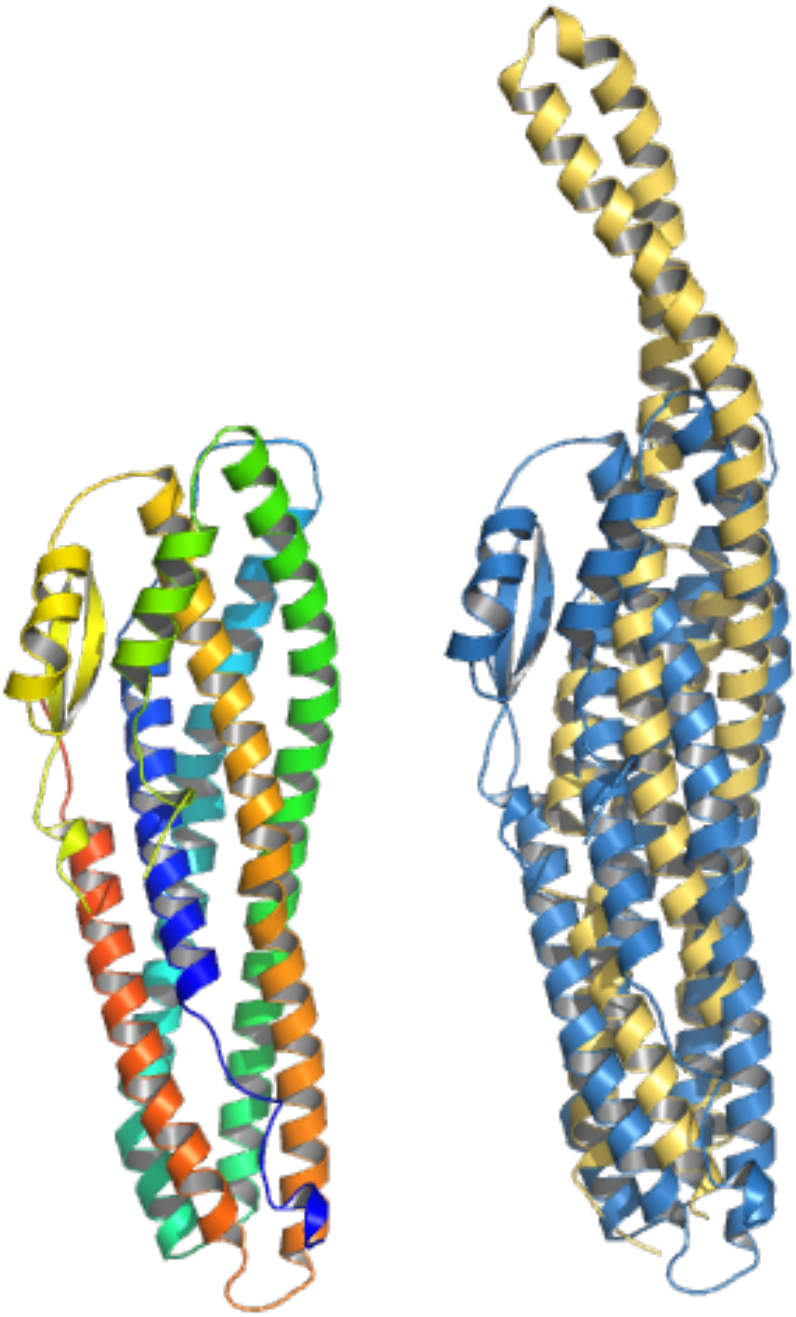


Figure 6.30. Superposition of AhlC with NheA. Ribbon representation of the soluble form of NheA (PDB accession 4K1P) is shown coloured from N to C termini as blue to red (left). Also shown is the superposition of the NheA structure (blue) with chain B of AhlC (yellow). The overall rmsd over 195 residues is 2.8 Å, and the general form of each protein overlaps in the tail domain of the proteins.

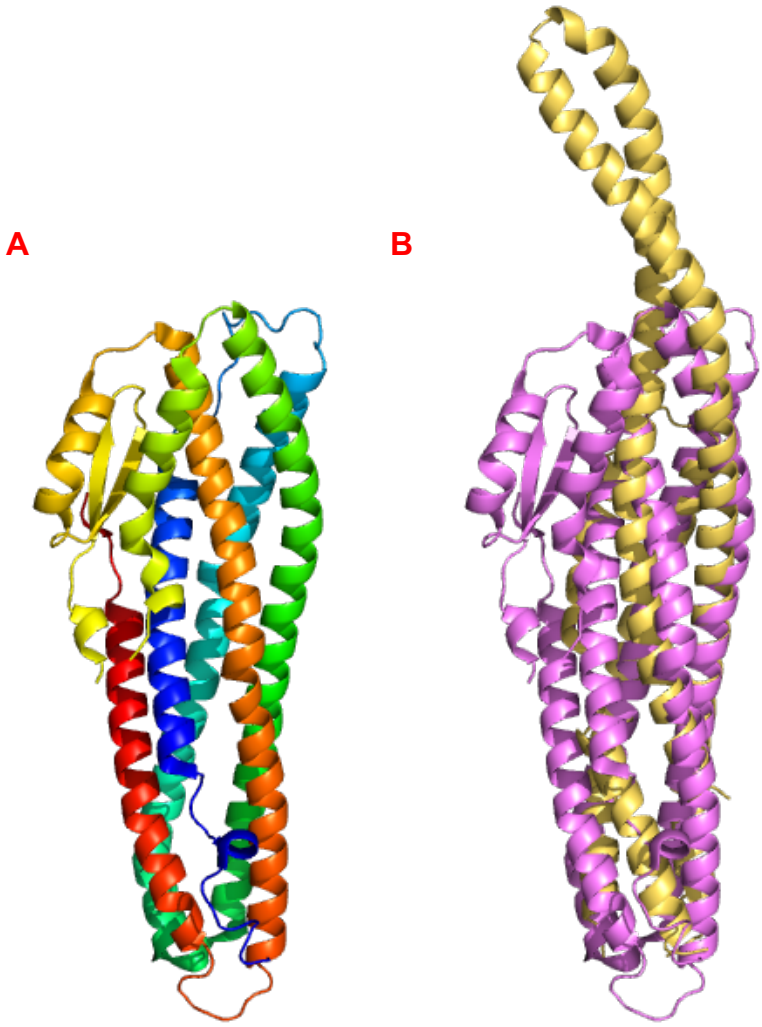


Figure 6.31. Superposition of AhlC with HblB. (**A**) Ribbon representation of the soluble form of NheA (PDB accession 2NRJ) is shown coloured from N to C termini as blue to red (left). (**B**) Also shown is the superposition of the NheA structure (pink) with chain B of AhlC (yellow). The overall rmsd over 166 residues is 2.9 Å. As both proteins are C component proteins, the overall structure is quite different, with HblB forming a β-tongue, and AhlC forming an α-helical hairpin.

that NheC would be similar in structure to HblB, dividing the C component into two sub-families.

6.4 Dali server analysis of AhlC

As the structure of AhlC was a potentially novel structure, the Dali server was used to look for similar proteins based on structural elements. Chain B of AhlC in the PDB format was uploaded to the Dali server (Holm and Rosenström, 2010) for structural analysis. Dali searches for similarity in structure between the query protein and the database of known protein structures. Each match is scored by the number of aligned residues, the rmsd between Cα's when imposed on one another, and the sequence identity between proteins. The match is also given an overall Z-score, which can be used to determine if a match is significant or not, with a cut-off of 8 used to indicate a reasonable similarity for this analysis.

The top results from the Dali server were those of other haemolysins in the family, namely NheA (PDB 4KLP) with a Z-score between 14.3 and 14.6, ClyA in the pore form (PDB 2WCD) with a Z-score between 10.6 and 13.5, Hbl-B with a Z-score of 11.8, and soluble ClyA with a Z-score of 10.2. A large number of other structures with Z-scores greater than 8 were identified, which can be grouped into two main categories.

The first category of Dali matches are protein involved in translocation and host cell binding of secretion systems. Two protein were highlighted by the Dali search, namely the soluble N-terminus of T3SS invasin IpaB (PDB accession 3U0C) with a Z-score of 9.5, and the T4SS RGD protein CagI (PDB accession 3ZCJ) with a Z-score of 9.5 (Barden et al., 2013; Barta et al., 2012b). IpaB has been previously associated with pore-forming toxins, due to its role in membrane invasion, as well as the structural similarity of IpaB and SipB, a *Salmonella enterica* homologue, to colicin translocator domains. However, the similarity to the ClyA family toxins has as of yet not been described. The superposition of AhlC with IpaB (Figure 6.32) aligned the stalk of AhlC with the two elongated helices in IpaB. Interestly both aligned helical hairpins contained three lysine residues (Figure 6.32), although IpaB did not also contain a stretch of hydrophobic residues as seen in AhlC. The second secretion system Dali hit was the T4SS CagL from *Helicobacter pylori*, a protein that decorates the pilus, and is reported to may function as an adhesin. The protein is formed of a four-helical bundle, containing a strictly required RGD tripeptide involved in binding intergrins.

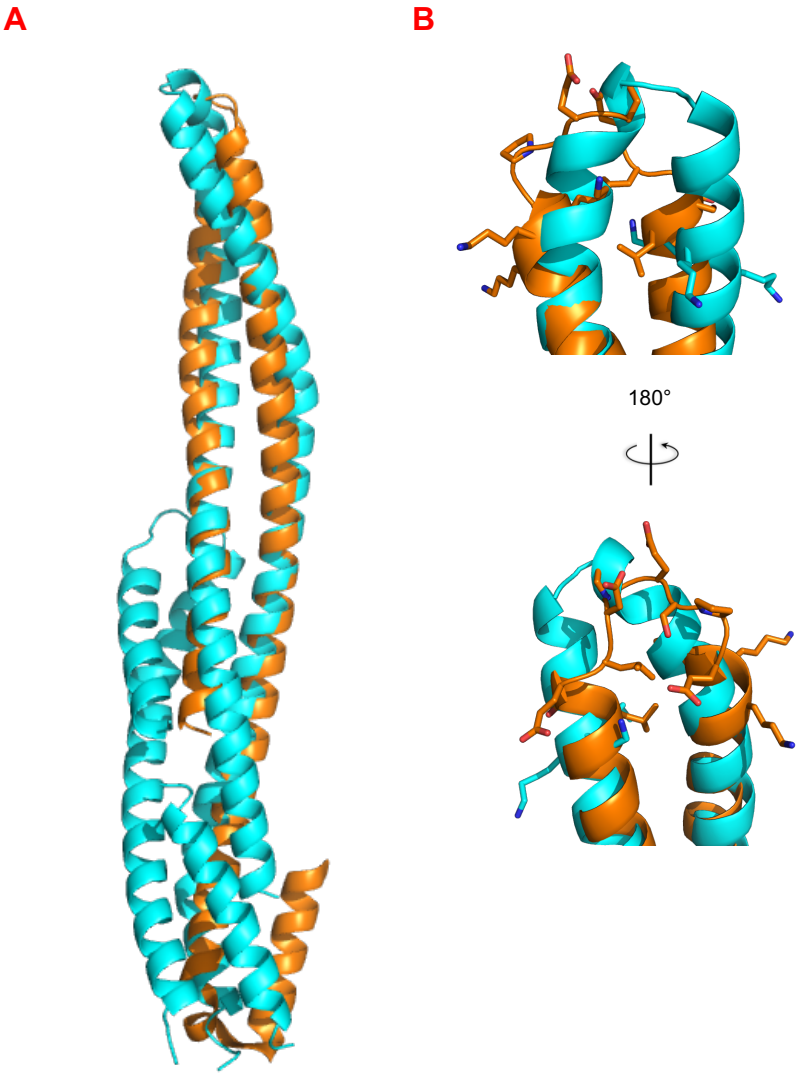


Figure 6.32. AhlC superposed with the soluble N-terminus of T3SS translocator IpaB. Ribbon representation of AhlC (cyan) superposed with the N-terminus of the T3SS translocator protein IpaB (orange)(PDB accession 3U0C). The two long elongated helices of IpaB align well with α3 and α4 of AhlC. (**B**) Ribbon representation of the head domain of AhlC aligned with the helical hairpin of IpAB. Both proteins contain a stretch of 3 lysine residues as the beginning of the helical hairpin.

The second category of Dali matches are those containing BAR domains, specifically the BAR domain from Human Bin1/Amphiphysin II (Casal et al., 2006)(Figure 6.33). This BAR domain protein is involved in associating with and/or remodelling membranes. The bridging integrator highlighted in the Dali search, with a Z-score of 10.4, forms banana shaped anti-parallel dimers and induces curvature in membranes by electrostatic interactions, and may bind other proteins in order to induce curvature, or may induce curvature through direct interaction with the membrane, although no such interaction has been observed to date.

Although both homologues proteins contain different methods for attaching to the membrane, both are involved in membrane integration, along with AhlC. It is interesting that all three contain a similar overall tertiary structure, and this structure may have some advantage in membrane insertion and membrane binding function that is currently not understood.

6.5 Structure function relationship of AhlC

The structure of AhlC shows a large similarity to previously solved ClyA like proteins, such as NheA, HblB, and ClyA itself. Strikingly, the structure most closely resembles the pore form of ClyA, as in both these proteins the head region is elongated and forms an alpha helical extension of α3 and α4. In AhlC this region is made up of a stretch of highly hydrophobic residues ($_{153}$LYLLGLGLLGLPGLIALAA$_{172}$), which forms a hairpin structure. The length of this region is ~14 Å, similar to that of the ClyA hydrophobic hairpin present in the pore form. The ClyA head region is suggested to insert into one leaflet of a lipid bilayer based on the dimensions of this region, and so it would be plausible that AhlC functions in a similar way.

The haemolytic activity of ClyA is generated by the formation of the pore, which is dependent on a conformational change associated with the insertion of this head into one leaflet of the lipid bilayer, followed by insertion of the N-terminal helix through the bilayer, forming an iris-like opening in the final pore form. Analysis of the AhlC sequence and structure, and comparison to ClyA, shows that as the protein is shorter than ClyA at both the N and C termini, a similar conformational change would not allow the N-terminal helix of AhlC to insert into the membrane in a similar fashion to ClyA. The apparent lack of a

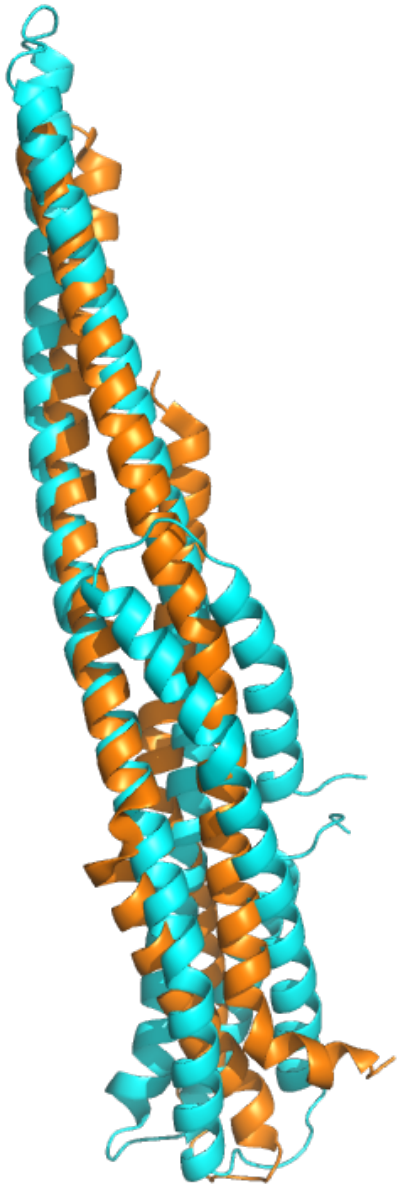


Figure 6.33. AhlC superposed with the BAR domain of human Bin1/Amphiphysin II. Ribbon representation of both AhlC (cyan), and the Dali sever hit, the BAR domain of human Bin1/Amphiphysin II, superposed using Lsqkab. The majority of the alignment is between α3 and α4 of AhlC with α2 and α3 of the BAR domain protein.

mechanism to produce a lipid bilayer-spanning pore would suggest a reason as to why AhlC alone has no haemolytic activity.

It was shown, however, in 5.5.2, that AhlC is absolutely required for any haemolytic activity, either with AhlB alone, or with both AhlA and AhlB together. The structure of AhlC suggests that it is able to insert into at most one leaflet of a lipid bilayer only. In order for AhlC to insert into the membrane, the protein must dissociate, as the head is sterically blocked by the flanking tail domains in the tetramer assembly, and some of the inserting residues are involved in forming the tetramer assembly.

The binding of AhlC to a lipid bilayer may then form a receptor for the other components of the toxin, i.e. AhlA and AhlB to bind to. Acting as a receptor may then increase the rate of pore formation by allowing the other components to bind at the membrane, or AhlC could play another as yet undetermined role in activating a pore once formed.

6.6 Mutational analysis of the AhlC hydrophobic head

6.6.1 Cloning a triple mutant variant of AhlC

Based on the structural analysis of AhlC, a series of mutations were designed to decrease the hydrophobicity of the head region (Leu156 -> Thr, Leu160 -> Thr, and Leu161 -> Thr). These mutations were chosen as they were sufficient to reduce hydrophobicity below 1.5 based on a Kyte-Doolittle hydrophobicity plot (Figure 6.34), 1.5 being the cut-off for a predicted transmembrane domain. Forward and reverse primers were designed using the NEBaseChanger online software for a substitution reaction using the Q5 site directed mutagenesis kit, and the sequences of each were GGAACTACTGGGCTGCCGGGCCTCATC for the forward primer, and CAGCCCAGTCAGGTAGAGTTTTTTCTTGTTGAGGGAGTCG for the reverse primer. Site directed mutagenesis was carried out using the following reaction mixtures, given in 3.2.10 and with the PCR protocol described in 3.2.10 with an annealing temperature of 61 °C. Correct mutations in the gene were confirmed by DNA sequencing, and once confirmed the AhlC head mutant ($AhlC^{HM}$) construct was transformed into BL21 DE3 and optimal overexpression conditions were determined by a series of overexpression trials. The best conditions for $AhlC^{HM}$ overexpression were found to be identical to those for the wild-type AhlC, and the mutant protein was purified using the same protocol as that for the wild-type

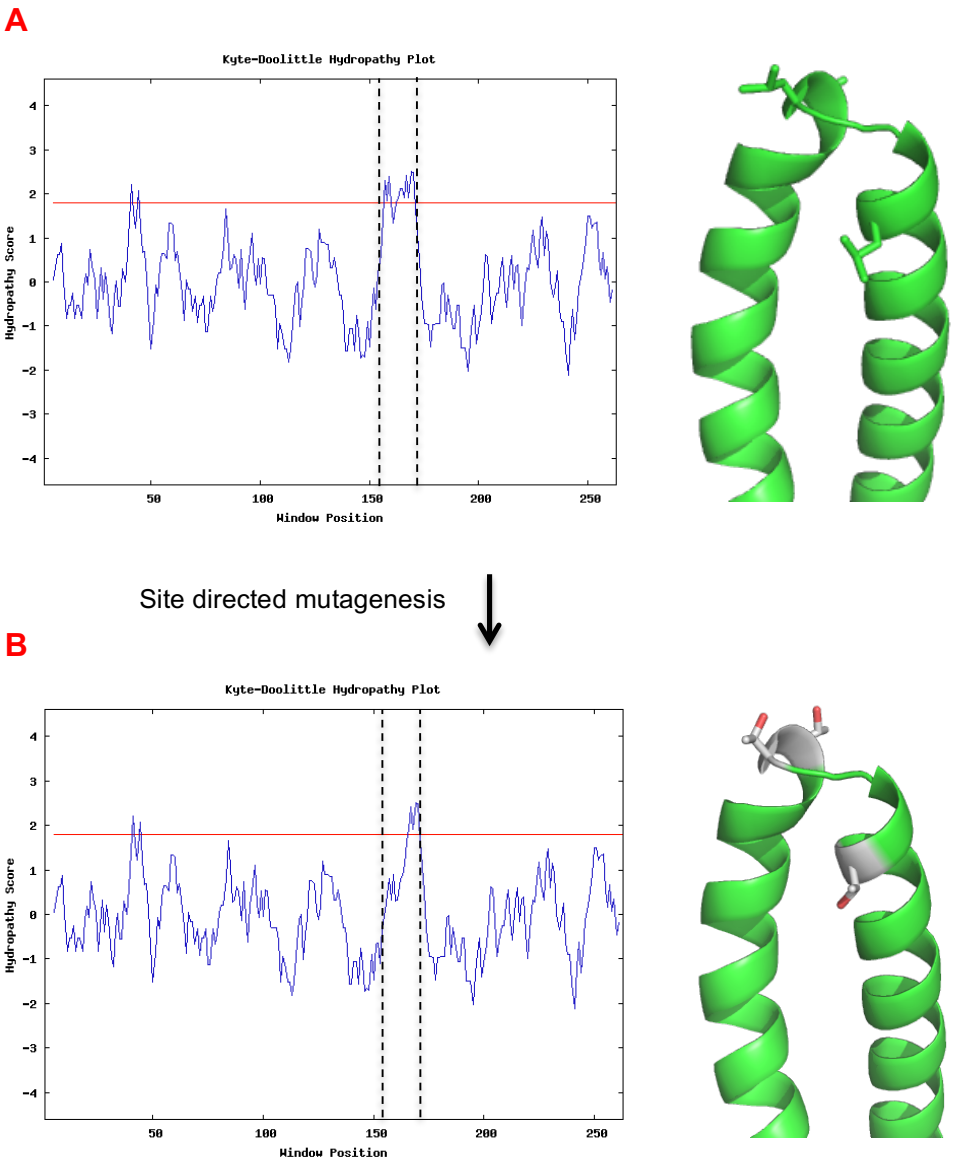


Figure 6.34. The effect of mutating Leu156, Leu160 and Leu161 to threonine residues on the hydropathy plot and the predicted effect on the structure of the protein. (A) Kyte and Doolittle plot for wild type AhlC (left), with the hydrophobic hairpin highlighted by dashed lines. Ribbon representation of AhlC with the residues to be mutated shown. (B) The Kyte and Doolittle hydrophobicity plot for AhlC in which the three Leucine residues are mutated to threonines (left). A ribbon representation of AhlC in which the three Leucine residues have been mutated to threonines.

AhlC. However, the yield of the mutant protein was reduced compared to that of the wild type protein (Figure 6.35).

6.6.2 AhlC head mutant haemolysis assay

The haemolysis assays as described in 5.5.2 were repeated using AhlA, AhlB, and $AhlC^{HM}$, to determine the effect of the AhlC head mutation on the haemolytic activity. These assays showed that lysis activity was severely attenuated when AhlC was substituted for $AhlC^{HM}$ in the AhlABC lysis system (Figure 6.36). These assays suggest that the hydrophobic nature of the AhlC head domain is essential for correct pore formation, giving further evidence that AhlC inserts into one leaflet of the membrane, during pore formation.

Further experiments would look to disrupt any binding of AhlA/AhlB to AhlC at the membrane by mutation of residues in AhlC that are conserved between AhlC and other tripartite toxins, which in the structure of AhlC face the solvent, and so may be involved in protein-protein interactions with the other toxin components.

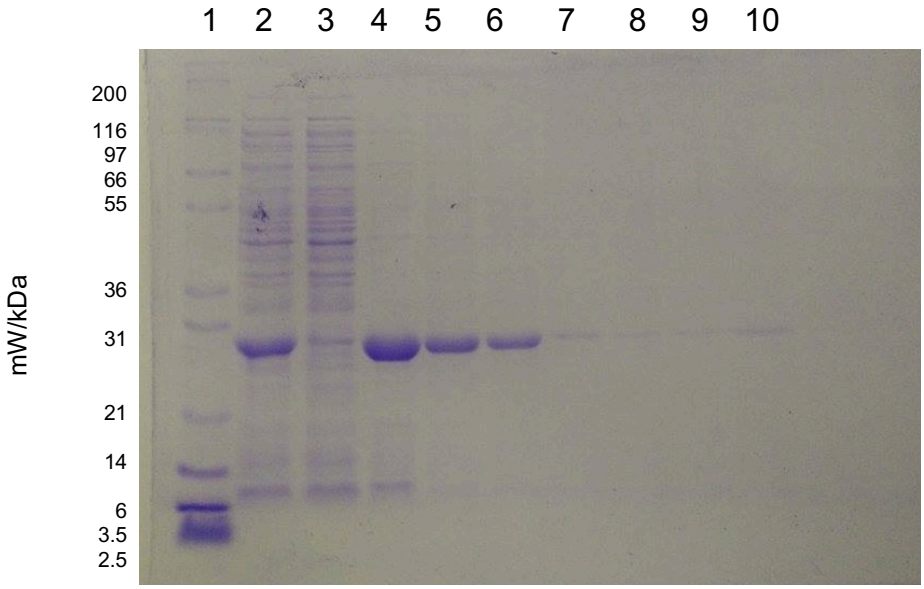


Figure 6.36. The purification of AhlC^{HM} shown by the UV trace chromatograms for NiHP and gel filtration chromatography and the resulting SDS PAGE. The SDS PAGE analysis of the purification of AhlC^{HM}. Lane 1 contains Mark12 ladder, and the molecular weights of each band are given. Lane 2 contains the cell free extract, and lane 3 contains the NiHP flowthrough. Lanes 4 to 10 contain the elution peak from the NiHP chromatography, with lanes 4 and 5 contains the most AhlC.

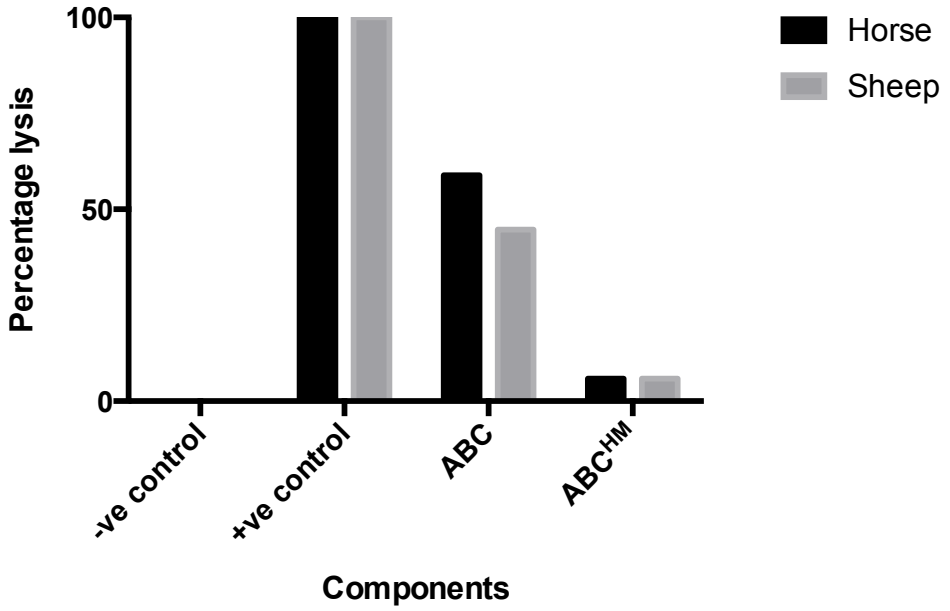


Figure 6.37. Percentage haemolysis of erythrocytes by AHL with AhlC^{HM}.A graph showing the haemolytic activity of AhlABC using both AhlC WT and AhlC containing the triple head mutation. 2 µg of each Ahl component was incubated for 1 hr. at 37 °C with 1 % (w/v) horse (black) and sheep (grey) erythrocytes in 10 mM PBS pH 7.4. Lysis was determined spectrophotometrically at 542nm on the supernatant once erythrocytes were centrifuged 2 minutes at 1500 xg. A negative control was measured as 1 % (w/v) erythrocytes without toxin, and a positive control measured as erythrocytes lysed in distilled water.

Chapter 7: X-ray crystallography studies of AhlB

7.1 Introduction

The structure of AhlC described in Chapter 6, the first of any component of these Aeromonas tripartite toxin family proteins, shows a distinct difference to the other greater family members, namely NheA, HblB, and ClyA. This difference is based in the head domain, which does not fold to form a β-tongue, and thus represents a departure from the dogma of these proteins, based on the structures of NheA, HblB, and ClyA. This raises questions as to the structure of the other two proteins AhlA and AhlB, and whether they adopt the beta tongue conformer, or are solubilised by oligomerisation, as with AhlC. As the yield of AhlA was too low to carry out crystallographic studies, subsequent crystallisation trials instead focused on AhlB, in order to determine its mode of solubilisation and role in pore formation, as well as to represent the first B component tripartite PFT structure.

This chapter describes the structural studies that were carried out in an attempt to determine the structure of AhlB by X-ray crystallography, and to characterise its function in cell lysis. Although the structure of AhlB was not determined, the experimental approaches in phasing attempts are outlined, and analyses of the symmetry present in certain crystal forms are detailed, which leads to a conjecture as to the function of AhlB.

7.2 AhlB crystallisation results

7.2.1 Initial Crystallisation trials of AhlB

AhlB was purified as described in 5.4.6, concentrated to 15 mg/ml, and buffer exchanged into various buffers for crystallisation trials. The buffer condition for initial crystallisation trials was 10 mM Tris pH 8 with no salt. Screening was carried out in sitting drop format in 96 well plates, and JCSG+, PACT, Classics, PEGs, MPD, and Morpheus screens were set up using a Matrix Hydra II PlusOne crystallization robot. Plates were incubated at 17 °C and checked routinely for crystal growth. Hit conditions were then optimised in hanging drop vapour diffusion format as described in 3.6.4, with a protein to precipitant ratio of 1µl:1µl, and incubated at 17 °C.

A number of crystallisation conditions led to crystals, which are shown in Table 7.1. These conditions can be broadly categorised into conditions containing MPD, and those not containing

Crystallisation condition	Crystal form	Salt	Buffer	Precipitant	Diffraction quality
PACT F4	1	0.2 M Potassium thiocyanate	0.1 M Bis-tris propane pH 6.5	20 % (w/v) PEG 3350	2.0 Å
JCSG+ A11	2	0.2 M Ammonium phosphate	0.1 M Tris pH 8.5	50 % (v/v) MPD	2.55 Å
JCSG+ H12	3	0.2 M Ammonium acetate	0.1 M Hepes pH 7.5	45 % (v/v) MPD	~3.3 Å
Proplex H9	3	0.1 M Imidazole	None	50 % (v/v) MPD	3.5 Å
MPD H6	4	None	None	30 % (v/v) MPD, 20 % (v/v) Ethanol	4.3 Å
Non-diffracting crystals					
MPD H3	-	0.2 M tri-sodium citrate	0.1 M NaHEPES pH 7.5	30 % (w/v) MPD	-
JCSG+ H5	-	0.2 M ammonium acetate	0.1 M bis-tris pH 5.5	45 % (w/v) MPD	-
pH clear H4	-	None	0.1 M HEPES pH 7.0	40 % (w/v) MPD	-
MPD D5	-	0.2 M sodium bromide	None	40 % (w/v) MPD	-
MPD B9	-	0.2 M sodium formate	None	40 % (w/v) MPD	-
Classics B9	2 (identical to JCSG+ A11)	0.2 M ammonium phosphate	0.1 M Tris pH 8.5	50 % (w/v) MPD	-

Table 7.1. Crystallisation conditions in which AhlB crystals were grown. Table 1 shows the conditions for which AhlB crystals grew, starting with diffracting crystals. For each condition, the crystal form, based on unit cell dimensions are given, along with the crystallisation buffer components, and the best diffraction observed for each crystal. The second half of the table shows conditions for which crystals were grown, but that diffracted very poorly.

MPD. Three crystal forms are discussed in this chapter, one grown in PACT F4, not containing MPD, and two grown in MPD containing conditions JCSG+ A11 and JCSG+ H12 (Figure 7.1).

7.3 AhlB crystal form 1 (grown in PACT F4)

A single large crystal of approximately 100 µm x 40 µm was grown from AhlB in PACT F4, containing 0.2 M potassium thiocyanate, 0.1 M bis-Tris propane pH 6.5, and 20 % (w/v) PEG 3350. These crystals were cryo-protected by transferring the crystals briefly into a solution containing 0.2 M potassium thiocyanate, 0.1 M bis-tris propane pH 6.5, 20 % (w/v) PEG 3350, and 20 % (v/v) ethylene glycol. The crystals were then cooled in liquid nitrogen and sent to the Diamond Synchrotron for data collection.

7.3.1 Data collection and processing

Test diffraction was carried out at the Diamond synchrotron, consisting of 5 images collected 45° apart, with a 0.2 second exposure and a 0.2° oscillation per image. The data was indexed by Mosflm as part of the fastDP pipeline, which predicted that the crystal belonged to space group P3, with cell dimensions a = b = 77.9 Å, c = 222.7 Å.

A full dataset was collected consisting of 360° oscillation collected over 3600 images with an oscillation per image of 0.1° and an exposure per image of 0.1 seconds. The full dataset was processed automatically by XDS using the 3dii pipeline, which determined the spacegroup of the data to be $P3_121$, with unit cell dimensions of a = b = 77.7 Å, c = 228.8 Å, and α = β = 90°, γ = 120°. The high resolution cut-off of the data, determined in aimless by the $CC_{1/2}$ < 0.5 value, was 2.02 Å. The full processing statistics are given in Table 7.2.

The Matthews coefficient was calculated to be 2.47 $Å^3\,Da^{-1}$ for two molecules of AhlB in the asymmetric unit, using the Matthews_coef software in CCP4 (Figure 7.2). The solvent content for such a unit cell was 51 %. Other less likely unit contents based on the Matthews coefficient were 1 molecule and 3 molecules in the AU, with Matthews coefficients of 4.93 $Å^3\,Da^{-1}$ and 75.1 % solvent content, and 1.64 $Å^3\,Da^{-1}$ and 25.1 % solvent content respectively.

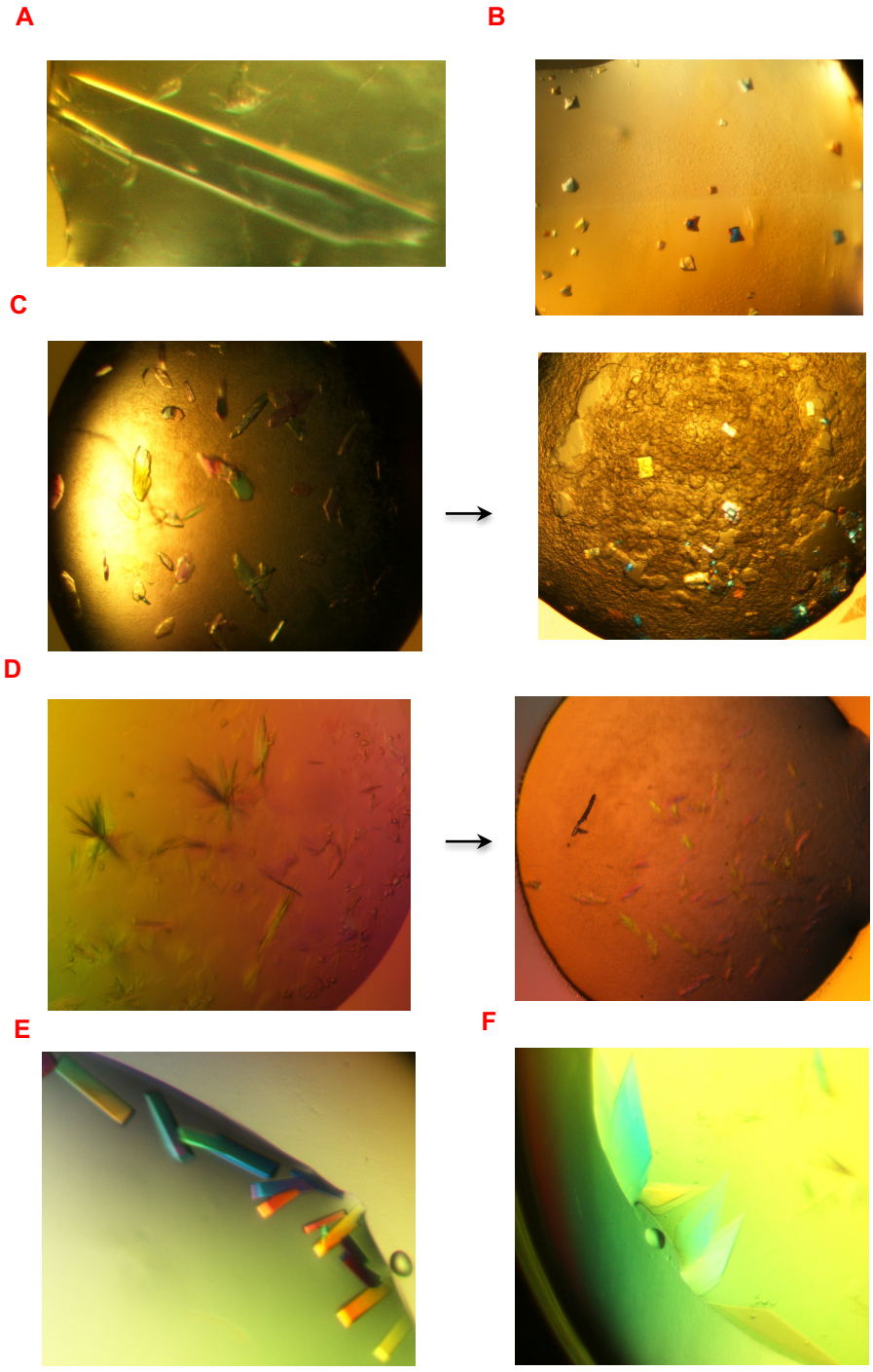


Figure 7.1. Crystals grown from AhlB protein. The only non-MPD condition PACT F4 crystal (**A**). Initial hits from JCSG+ A11 (**B**). Initial hits from JCSG+ H12 (**C** left) and optimised crystals grown in hanging drops (**C** right). Initial hits of crystals grown in MPD H6 (**D** left) and optimised crystals (**D** right). Finally, initial hits of crystals grown in Proplex H9 (**E**) and MPD H6 (**F**)

Table 7.2. Processing statistics for the PACT F4 crystal form 1 data.

	Overall	Low	High
High resolution limit	2.02 Å	9.03 Å	2.02 Å
Low resolution limit	64.55 Å	64.55 Å	2.07 Å
Completeness	100 %	99.7 %	99.9 %
Multiplicity	17.5	17.6	15.7
I/sigma	14.8	29.2	4.2
R_{merge}	0.147	0.118	0.577
$R_{meas}(I)$	0.155	0.121	0.611
$R_{meas}(I+/-)$	0.156	0.123	0.616
$R_{pim}(I)$	0.037	0.028	0.153
$R_{pim}(I+/-)$	0.05	0.036	0.214
Wilson B factor	21.97		
Anomalous completeness	100	100	99.9
Anomalous multiplicity	9.2	11.4	8.1
Anomalous correlation	-0.22	-0.509	-0.155
Anomalous slope	0.691		
dF/F	0.06		
dI/s(dI)	0.636		
Total observations	921531	12446	60429
Total unique	52750	706	3842

A

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0166	0.0749	4.93	75.06	38558.00
2	0.9662	0.9153	2.47	50.12	77116.00
3	0.0172	0.0098	1.64	25.17	115674.00

B

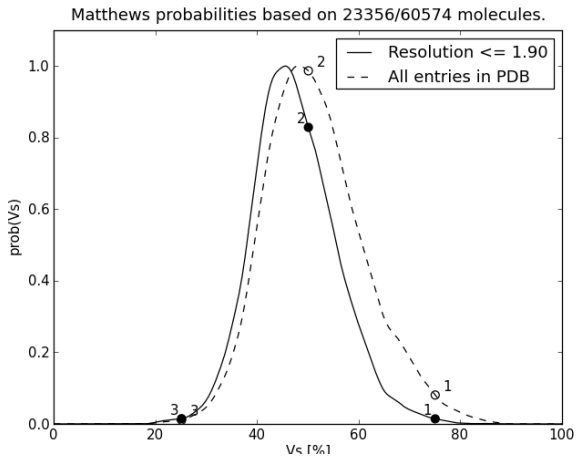


Figure 7.2. Matthews coefficient analysis of crystal form 1 grown in PACT F4. (**A**) The Matthews coefficient analysis of crystal form 1 predicts that 2 molecules are most likely to be present in the asymmetric unit, with a Vm of ~2.5, and solvent content of 50.1 %. Other possibilities for the contents of the asymmetric unit are one molecule and three molecules, with a Vm of 4.9 and 1.6 respectively, with solvent contents of 75 % and 25 % respectively. Both of these alternative AU contents have probabilities of < 2 %, based on the resolution of the data. (**B**) A plot of the probability of a given solvent content, prob(Vs), against solvent content, Vs, for all structures in the pdb, showing that the most likely cell contents is two AhlB per A.U.

7.3.2 Crystal form one molecular replacement attempts using a modified NheA chain as a search model

As no structures of proteins with >30 % sequence similarity to AhlB were known, it was envisaged that molecular replacement would be difficult. Nevertheless, the overall similarity in structure of the known family members, and the pattern of secondary structure prediction for AhlB, indicated that AhlB might contain a similar fold similar to other ClyA family toxins. Phyre2 predicted that the structure of AhlB would be most similar to NheA, and as such, molecular replacement was initially attempted using a modified NheA model, in which the sidechains were removed to produce a poly-alanine chain, and the head domain and loop regions were removed to generate a single model containing three helices (Figure 7.3). Molecular replacement was run using this model as a search ensemble, searching for 2 molecules, with a sequence identity value set to 20 %. There were 81 top solutions in the initial PhaserMR run, as determined by the LLG and TFZ scores, but PhaserMR discarded all solutions as they failed in the subsequent packing test. All of the solutions had TZ scores above 12, and the highest TFZ was 17.8. A TFZ score above 12 indicates a definite solution, so scores of 17.8 indicates that failing of the packing test had an underlying cause. The PhaserMR program was rerun fixing the position of the first molecule, and turning off the packing test, so that all solutions were allowed. The top three solutions were then written to co-ordinates in a .pdb files. Manual observation of the top three solutions showed that each of the three chains in the unit had been placed directly on top of one another, with a slight translation along the length of the helices.

As all molecular replacement solutions were superposed on top of one position, the first chain placed in molecular replacement was explored as the complete unit cell contents, with a large solvent content of 76 %. A complete molecule of NheA was superposed onto the position of the search model placed by molecular replacement, and all symmetry related molecules generated (Figure 7.4). With symmetry related molecules shown, it could be seen that the symmetry related chains clashed, suggesting either the solution was incorrect, or parts of the protein needed to be removed and rebuilt in order to fit. Also there were gaps in the packing of the molecules such that a complete lattice was not formed. Next, the electron density map was analysed for how well the density fit the molecule, any returning sidechain features, and peaks in the $|F_o-F_c|$ difference electron density map.

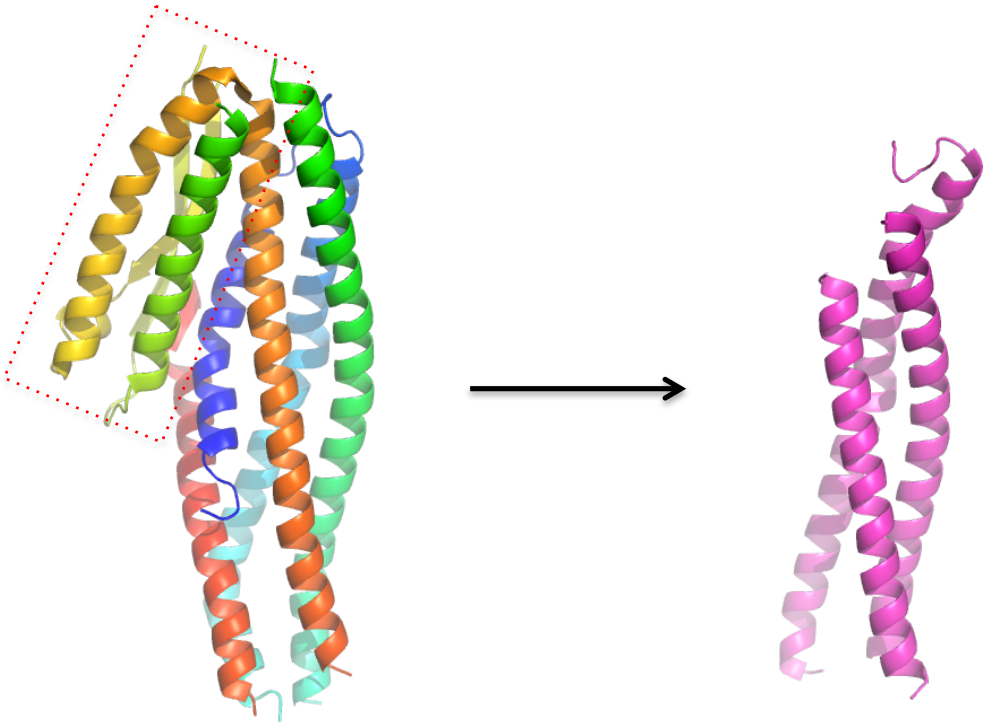


Figure 7.3. The molecular replacement search model for AhlB. Chain A of NheA (PDB accession 4K1P) was converted to a poly-alanine chain using the chainsaw software in CCP4, and then regions of high variance between ClyA-like toxins were removed, including the head domain (red dashed box), and the N and C termini (Blue and red helices respectively).

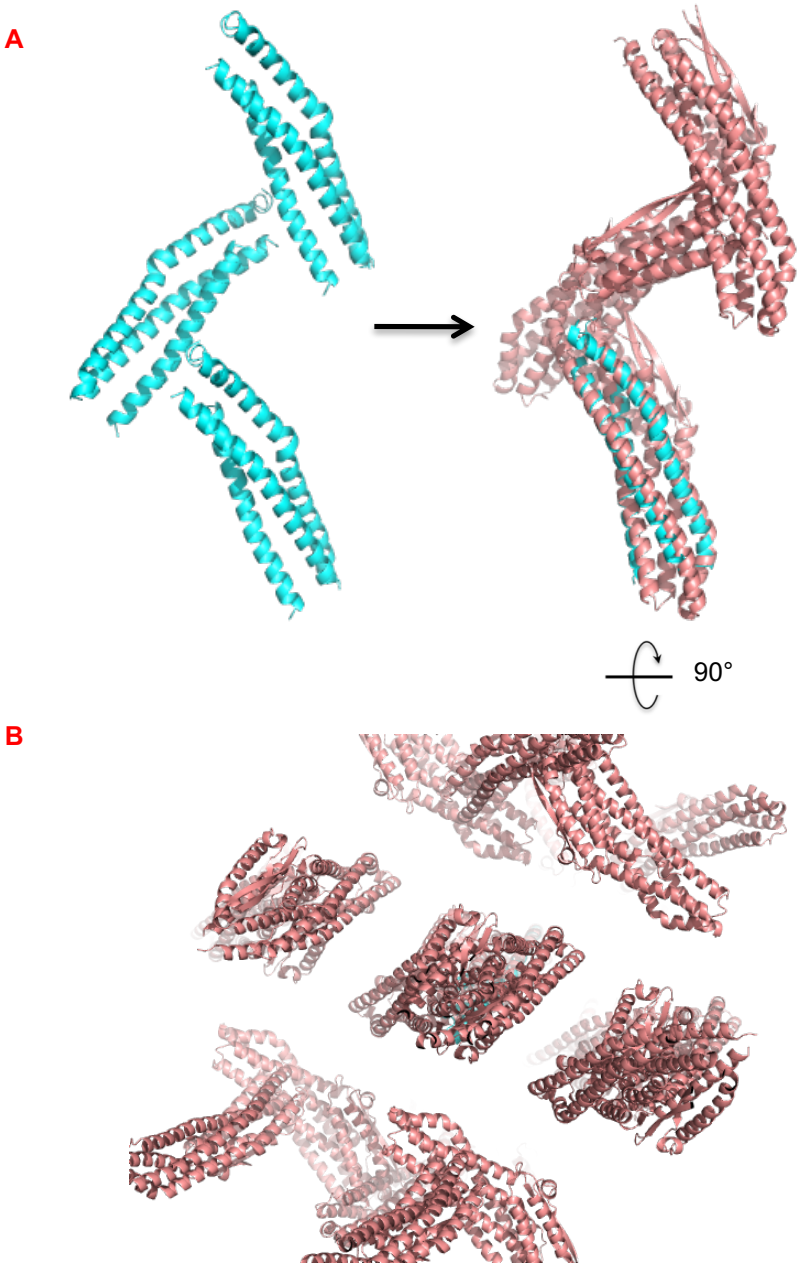


Figure 7.4. Examining the packing of the molecular modified NheA search model molecular replacement. (**A**) The molecular replacement solution (cyan) contained one chain in the AU, with no clashing between molecules (left). However, when a full chain of NheA (pink) was superposed on the solution, and symmetry generated, each NheA molecule did clash (right). (**B**) The superposed NheA molecules (pink) with symmetry generated molecules present did not form a lattice in certain direction, indicating an incorrect or incomplete solution.

Overall, the density fitted the molecule well (Figure 7.5), although no extra density was present in either the |Fo-Fc| or the |2Fo-Fc| electron density maps at the ends of each α-helix. Interestingly a number of sidechains seemed present in the electron density, despite the search model containing only alanines. One such feature was electron density for a tryptophan residue present at the relevant position in the amino acid sequence, at residues 286 (Figure 7.5). As examination of the electron density map was ambiguous, in that some sidechain features were present, but secondary structure features were not, the MR solution was tested for bias. A 7 amino acid long stretch of the protein model was deleted from the pdb file, centred on the putative tryptophan density (Figure 7.5), and 10 cycles of restrained refinement were performed using Refmac5 (Murshudov et al., 1997). The R factor during refinement decreased from 0.522 to 0.508 suggesting that one molecule did not describe the data correctly. In addition, the electron density for the deleted part of the protein model did not return in either of the |Fo-Fc| or the |2Fo-Fc| electron density maps, further suggesting the molecular replacement solution was incorrect.

7.3.3 Molecular replacement attempts using a modified search ensemble made up of known ClyA homologue structures

In an attempt to improve the molecular replacement solutions, the procedure was repeated using a search ensemble generated from the other ClyA family members with known structure. The structures of ClyA, NheA, Hbl B, and AhlC, were aligned using Lsqkab (Kabsch, 1976), and pruned to form a poly-alanine chain suitable for molecular replacement. In addition, the areas of large deviation between structures i.e. the whole head domain and part of the tail domain, were also removed from each structure, to leave five helices for each protein in the ensemble (Figure 7.6). The overall rmsd between the final helices was 3-4 Å even though the sequence similarity between each full protein was only <20 %. Molecular replacement was run using this truncated model, searching for two monomers in the asymmetric unit. It was however understood that molecular replacement would still be difficult due to the ensemble only representing 15 % of the scattering matter in the crystal.

In the ensemble molecular replacement, the result indicated that the problem of the superposing chains had been resolved, however the TFZ scores for the top solutions were only <6, indicating no solution had been found. PhaserMR failed to find a complete solution in which

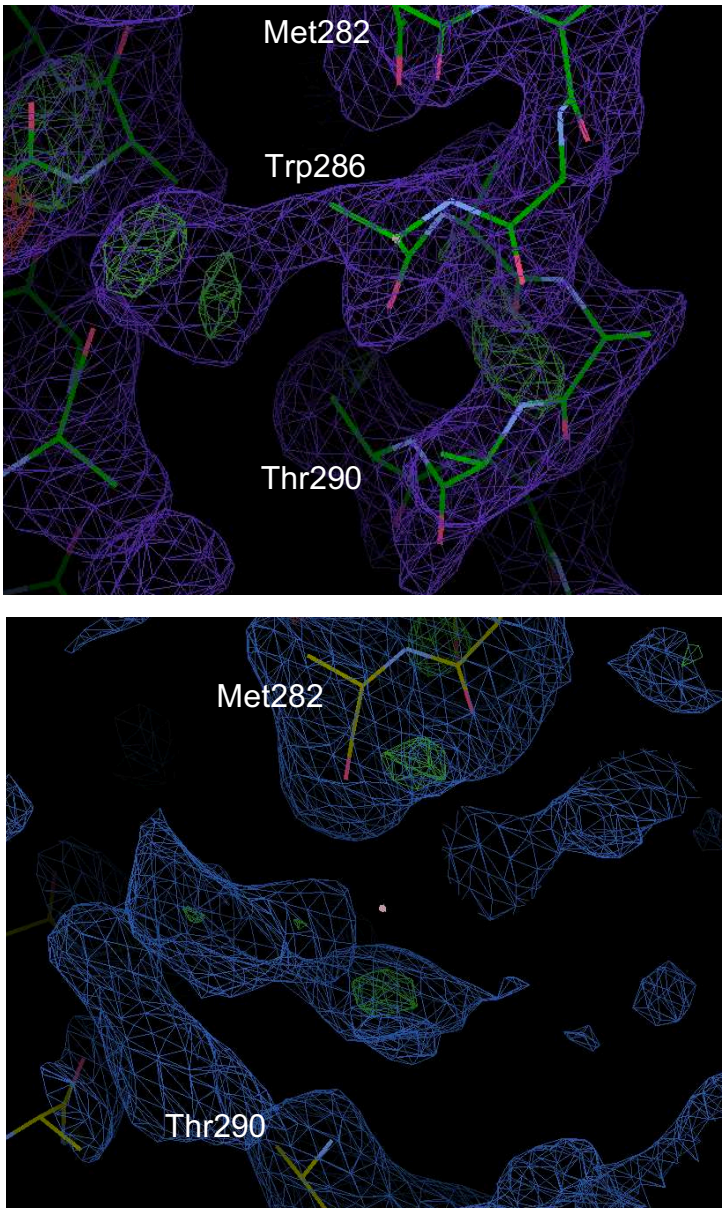


Figure 7.5. Bias check on Trp286 in the AhlB PACT F4 molecular replacement. (**A**) Electron density map of the molecular replacement solution from 7.3.2. Electron density is present in both the electron density map and difference map for a tryptophan residue that was not present in the search model. As a bias check residues 283 to 289 were removed and 10 cycles of restrained refinement were performed in Refmac. (**B**) The electron density map of the molecular replacement solution after residues 283 to 289 were removed and refinement was performed. The density for the mainchain atoms is no longer present, along with the electron density for Trp286.

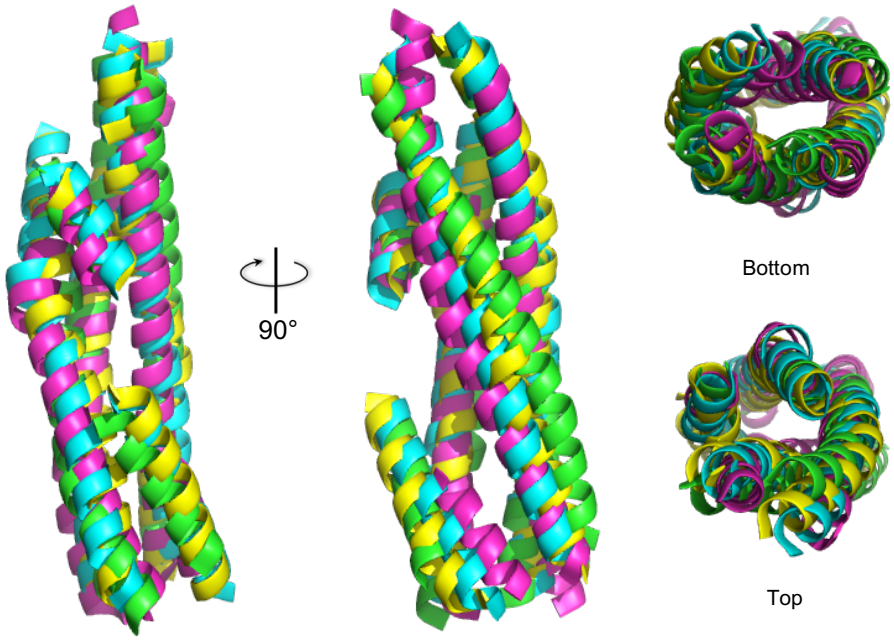


Figure 7.6. The final aligned search ensemble for molecular replacement of the PACT F4 crystal form 1 data. Each of the ClyA family toxin proteins ClyA (Green), NheA (Magenta), HblB (Yellow), and AhlC (Cyan) were aligned using the Lsqkab software in CCP4, and then visualised in Pymol for overlapping areas. The head domain of each protein was removed completely, as well as regions of the tail domain not aligned well between each of the proteins. The final selection for each protein was saved into a new co-ordinates file, and reduced to a poly-ALA chain using the chainsaw software in CCP4. The final aligned models were used in ensemble molecular replacement attempts for the PACT F4 data in PhaserMR software.

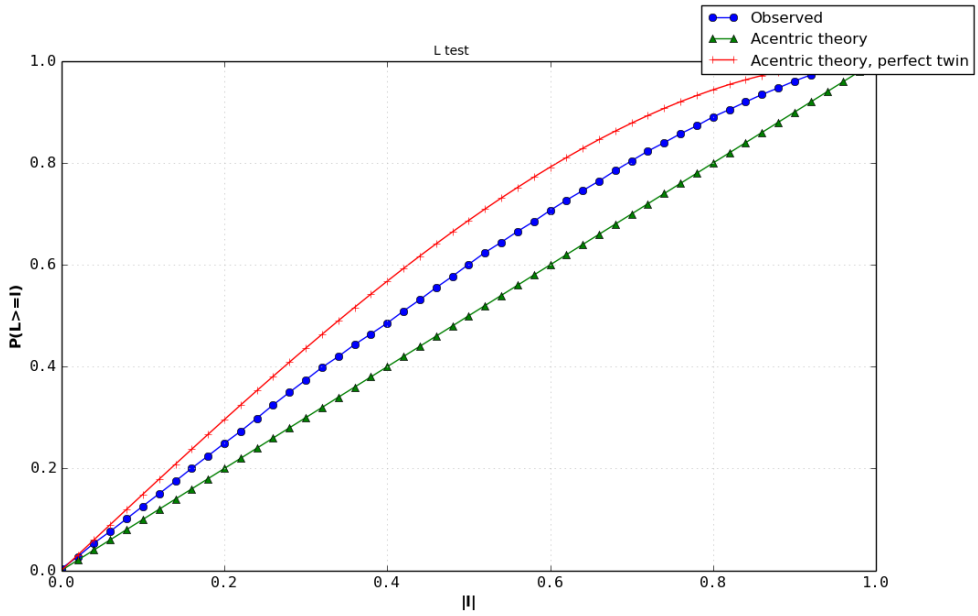
all chains had been placed, and so terminated early. As such, it is believed that the sequence identity between each molecule and AhlB is not sufficient to find a molecular replacement solution. Also, the superposed regions of the search ensemble only represent half of the AhlB molecule, and so each chain represents ~9 % of the total scattering matter, compounding the problem of identifying a molecular replacement solution.

7.3.4 Twinning analysis on the PACT F4 data

The cumulative intensity distribution for the PACT F4 data suggested significant twinning, also highlighted in the L test for twinning and the Wilson ratio and moments test carried out in Phenix (Zwart et al., 2005)(Figure 7.7 and 7.8). The L-test is particularly useful as it can differentiate between twinning and anisotropy and pseudosymmetry based departures from regular intensity distributions (Padilla and Yeates, 2003). In order to determine if any translational NCS was present in the crystal, a self-Patterson function was performed and peaks in the self-Patterson were analysed. The self-Patterson showed that the largest peak present was 3.8 % that of the origin over a resolution range of 10 Å and 2.8 Å, at fractional co-ordinates of 0.075, 0.15, and 0.49, with a p-value of 0.99. Therefore, the self-Patterson search indicated no translational NCS was present. Taken together these results suggest that the data are truly twinned, and as such, the spacegroup assignment was checked by reindexing the data into a number of different spacegroups and analysing the processing statistics.

7.3.5 Attempts to determine the correct spacegroup for the PACT F4 data by reindexing the data and analysing the merging statistics

Firstly, the lower symmetry spacegroup P3 was checked as a possible spacegroup for the twinned data. The data were indexed and integrated in Mosflm, and scaled and merged in pointless and aimless, enforcing the P3 spacegroup on the data. The final merging statistics output by aimless are shown in Table 7.3, and compared with the P321 data. Overall, the statistics show that merging in P3 produces good data, with a low R_{pim} in the low-resolution bin of 0.042, and an overall R_{pim} of 0.054. The same twinning tests performed on the P321 data were carried out on the P3 data using Xtriage in Phenix, and showed that more substantial twinning was present, and an estimated twin fraction of 0.46 based on the H-test of acentric data and Britton analyses using the twin law h,-h-k,-l (Figure 7.9). It is expected however that if the crystal really is in the spacegroup P321 then reflections h, k, l, and h, -h-k,-l will be



Acentric reflections	
$\langle I^2\rangle/\langle I\rangle^2$	1.755
$\langle F\rangle^2/\langle F^2\rangle$	0.835
$\langle \lvert E^2-1\rvert\rangle$	0.648
Centric reflections	
$\langle I^2\rangle/\langle I\rangle^2$	2.449
$\langle F\rangle^2/\langle F^2\rangle$	0.726
$\langle \lvert E^2-1\rvert\rangle$	0.854

Figure 7.7. The twinning analysis results for the crystal form 1 data in spacegroup $P3_121$. (**A**) The L test graphs plots |l| against P(L=>l). The green line shows the theoretical values for untwinned data and the red line shows theoretical values for perfectly twinned data. The blue line plots the measured intensity values for the PACT F4 data. (**B**) Twinning was also suggested by the Wilson ratio and moments, where values for untwinned and perfectly twinned data are 2.0 and 1.5 respectively for $\langle I^2\rangle/\langle I\rangle^2$, 0.785 and 0.885 respectively for <F>2/<F2>, and 0.736 and 0.541 respectively for <|E2-1|>. The measured Wilson ratio and moments show partial twinning for the data in P321.

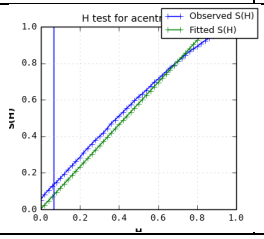
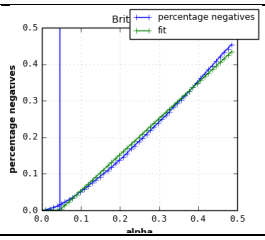
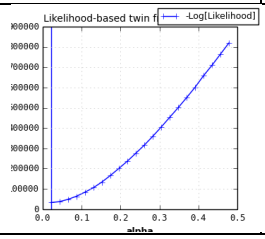
Twin Law	H test	Britton analysis	Likelihood based twin fraction	Estimated twin fraction
-h, -k, l				~0.1

Figure 7.8. The twinning analysis results for the crystal form 1 data in spacegroup $P3_121$ based on the twin law –h, -k, l. Twinning analysis based on the twin law –h, -k, l were carried out in Xtriage in Phenix. For each test, the blue line on the graph represents the estimated twin fraction, and each test shows that the estimated twin law is low at ~0.1.

Table 7.3. Processing statistics for the reindexing of crystal form 1 data into P3, P2, and P1, with the P321 statistics for comparison.

	P321	**P3**	**P2**	**P1**
a (Å)	77.7	77.7	77.8	77.6
b (Å)	77.7	77.7	228.5	77.7
c (Å)	228.8	228.8	77.7	227.9
α (°)	90	90	90	90.2
β (°)	90	90	120	90.3
γ (°)	120	120	90	120
Resolution limit	2.02 Å	1.92 Å	1.99 Å	2.03 Å
I/sigma (overall)	14.8	9.0	7.1	4.6
I/sigma (High resolution)	4.2	1.1	1.3	0.9
Completeness (overall)	100 %	99.3 %	98.0 %	89.6 %
Completeness (High resolution)	99.9 %	95.3	91.0 %	75.2 %
Multiplicity (overall)	17.5	8.1	4.1	3.9
R_{merge} (overall)	0.147	0.156	0.218	0.175
R_{merge} (high resolution)	0.577	1.47	0.999	0.740
R_{merge} (Low resolution)	0.118	0.127	0.179	0.129
R_{pim} (overall)	**0.037**	**0.054**	**0.105**	**0.114**
R_{pim} (High resolution)	0.153	0.664	0.547	0.551
R_{pim} (Low resolution)	0.028	0.042	0.087	0.083
CChalf (overall)	0.996	0.992	0.957	0.963
CChalf (High resolution)	0.951*	0.456	0.545	0.517
Total observations	**921531**	**949549**	**771865**	**850953**
Total unique observations	52750	117343	156930	266716

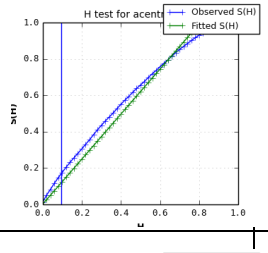
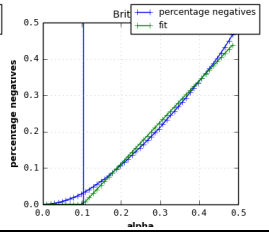
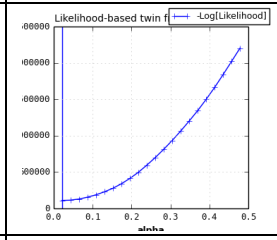
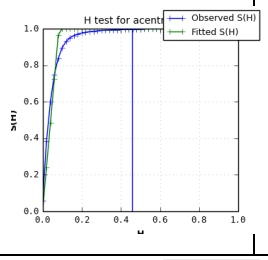
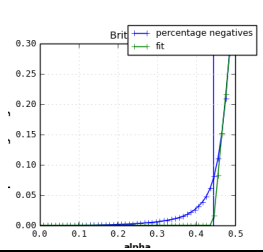
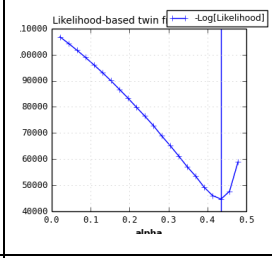
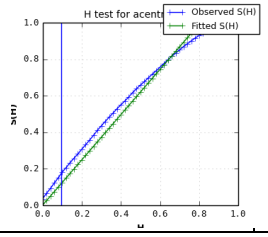
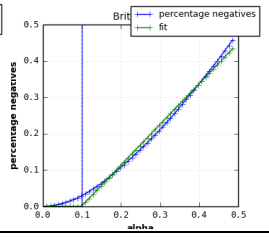
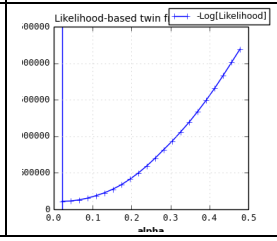
Twin Law	H test	Britton analysis	Likelihood based twin fraction	Estimated twin fraction
-h, -k, l				~0.1
h, -h-k, -l				~0.45
-k, -h, -l				~0.1

Figure 7.9. The twinning analysis results for the crystal form 1 data reindexed into lower symmetry spacegroup P3. A H-test, Britton analysis and the Log-likelihood based twin fraction analysis were carried out on the P3 data using each possible twin law, in the Xtriage software. For both the twin laws –h, -k, l, and –k, -h, -l, each test and analysis shows that the twin fraction is very low, at ~ 0.1 for each operator, and so twinning is not present using these twin laws (estimated twin fraction marked by vertical line in each graph). However, the analysis of twin law h, -h-k, -l shows a large twin fraction of ~0.45 mean twin fraction over the three tests. However it should be noted that for the higher spacegroup P321, the compared reflections are symmetry equivalents and so will look like twinning when analysed for correlation in the lower symmetry P3.

equivalent as they are symmetry equivalents, and so the twinning tests will show a high twin fraction although one does not necessarily exist. The estimated twin fraction of 0.45 is well within the limit of the quality of the data, based on the merging statistics shown in Table 7.3, and so either the data are either perfectly hemohedrally twinned and in P3 or are in P321, based on these twinning tests.

Regardless, the P3 data were used in molecular replacement to try to alleviate packing clashes due to possible incorrect symmetry. The same search ensemble was used for molecular replacement of the P321 data, and a separate molecular replacement job was run for each enantiomorph, P31 and P32. The Matthews coefficient for the P3 data predicted four or five molecules in the asymmetric unit were most likely, with Matthews coefficients of 2.58 $Å^3Da^{-1}$ and 2.06 $Å^3Da^{-1}$ respectively, and a solvent content of either 52.3 % or 40.3 % respectively. Molecular replacement was run searching for 4 chains and 2 chains, based on the possible solution in P321. Again, no MR solution was found, and so reindexing the data to P3 was not sufficient to overcome any twinning problem, and twinning was not obscuring a correct molecular replacement solution.

To check that no other spacegroups better defined the diffraction data, the data were reindexed to all probable Mosflm indexing solutions spacegroups, P6, P3, P2, P1, and C2, and then scaled and integrated using Xia2 3dii. All merging statistics are shown in Table 7.3. Attempts to merge the data in the spacegroup C2 led to a large number of rejected reflections, 40 % overall completeness, and a high-resolution limit of 4.5 Å. The poor merging statistics ruled out the C face centred spacegroup.

The P1 data and the P2 data merged well, and processing statistics are shown in Table 7.3. Although the P2 indexing merging statistics are good, far fewer reflections were observed when indexing the data in P2 (>147,000 reflections rejected). The precision merging R-value was also much higher in P2 at 0.11, compared with 0.04 and 0.05 for P321 and P3 respectively. The Data processed in P1 had unit cell dimensions of a = 77.6 Å, b = 77.7 Å, and c = 227.9 Å, with angles of α = 90.2°, β = 90.3°, and γ = 120°, suggesting the higher symmetry trigonal P3 or P321 again better explained the data.

Based on these analyses, it would seem most likely that the correct spacegroup for the PACT F4 data is either P3 with twinning present, or P321 and the contents of the unit cell are

effecting the cumulative intensity distribution making the data look twinned, although the L-test should differentiate between these.

7.3.6 AhlB PACT F4 crystal form one summary

It is hoped that in the future, when a structure with greater sequence similarity is made available, the PACT F4 data can be solved by molecular replacement, regardless of issues with twinning. Advances in *ab initio* structure determination programs such as Arcimboldo (Sammito et al., 2014) may also lead to a structure solution, due to the high α-helical percentage present in the tripartite toxins. Finally, attempts to optimise AhlB crystallisation in crystal form 1 may lead to crystals that can be used in experimental phasing experiments.

7.4 AhlB crystal form 2 (JCSG+ A11)

Small crystals of ~ 20 x 40 µm were grown from AhlB at 15 mg/ml in 10 mM Tris pH 8, in JCSG+ condition A11, which contained 0.2 M Ammonium phosphate, 0.1 M Tris pH 8, and 50 % (v/v) MPD. These crystals were of sufficient size to be tested for X-ray diffraction at Diamond synchrotron. In parallel to this, optimisations were carried out, varying protein concentration, PEG concentrations and salt concentrations. Optimisations yielded slightly larger crystals grown from 20 mg/ml AhlB in 50 mM tris pH 8 and 0.2 M NaCl storage buffer. The crystals grew in 0.2 M Ammonium phosphate, 0.1 M Tris pH 8, and 50 % (v/v) MPD, and were also sent to the Diamond synchrotron for data collection.

Due to the presence of high percentage of MPD in the JCSG+ A11 crystallisation solution, the mother liquor from which the crystals were grown was sufficient as a cryoprotection solution and crystals were cooled in liquid nitrogen with no further additions. Cooled crystals were sent to the Diamond synchrotron for data collection.

7.4.1 Data collection and processing

Diffraction tests were carried out at the Diamond Synchrotron and showed that JCSG+ A11 crystals diffracted to varying resolutions, from 4 Å to 2.5 Å, and test diffractions were collected from three crystals. Test diffraction was carried out with a 0.2° oscillation and a 0.2 s exposure per image, and 5 images were collected 45° apart for indexing. The test images were indexed by Mosflm, as part of the fastDP pipeline, and Mosflm predicted the spacegroup of the crystal to

be $C222_1$, with unit cell dimensions of a = 116 - 117 Å, b = 175 - 177 Å, and c = 484.3 – 484.6 Å between two of the crystals, and indexing failed for the third crystal.

A full dataset was collected from the crystal which diffracted to 2.5 Å, consisting of 900 images, with a 0.2° oscillation and 0.2 s exposure per image, for a total oscillation of 180°. The diffraction data was processed automatically by XDS at Diamond, using the Xia2 pipelines 3da. The best merging statistics resulted from the Xia2 2da processing, with a high-resolution cut-off of 2.55 Å based on I/σI < 2. Full processing statistics are shown in Table 7.4. The data were processed in the spacegroup $C222_1$, with cell dimensions of a = 117.7 Å, b = 178.2 Å, and c = 485.6 Å. The Matthews coefficient was calculated and showed that the asymmetric unit contained between 7 and 18 molecules of AhlB, with the extremes containing a crystal volume per protein mass (Vm) ranging from 1.82 $Å^3Da^{-1}$ to 4.68 $Å^3Da^{-1}$(between 32 % and 74 % solvent content)(Figure 7.10).

7.5 SeMet derivitised AhlB protein

In order to try to determine phases for both AhlB crystal forms 1 and 2, AhlB was derivitised with selenomethionine. AhlB overexpression was repeated using the SeMet protocol as described in 3.4.3 to produce protein in which the methionine residues were replaced with selenium methionines. The resulting protein was purified identically to native AhlB, and SDS PAGE analysis showed that the purity of the resulting protein was similar to that of native AhlB.

To determine the success of the derivitisation, purified SeMet AhlB was sent for mass spectrometry analysis to determine the amount of selenium incorporation. Mass spectrometry and data analysis was kindly carried out by Dr. Simon Thorpe in the Chemistry Department, and the results showed that there was full incorporation of selenium in the derivitised AhlB (six selenium methionines per molecule). This fully derivitised protein is referred to as AhlB SeMet in all following discussion (Figure 7.11).

Large-scale crystallisation screening was carried out using AhlB SeMet using PACT, JCSG+, MPD, Proplex, and pegs screens, as well as optimisations of the PACT F4 and JCSG+ A11 conditions varying the buffer pH and precipitant concentration. Although optimised extensively, no successful crystals were grown from PACT F4 optimisations with SeMet AhlB. Crystals grew of AhlB SeMet from optimised JCSG+ A11 conditions containing 0.2 M ammonium phosphate, 0.1 M tris pH 8.5, and 45 % (w/v) MPD, although these crystals only

Table 7.4. Processing statistics for the JCSG+ A11 crystal form 2 data.

	Overall	Low	High
High resolution limit	2.55	7.21	2.55
Low resolution limit	39.55	39.55	2.73
Completeness	100	99.4	100
Multiplicity	6.4	6.2	6.6
I/sigma	6.6	18.0	2.7
R_{merge}	0.188	0.054	0.628
$R_{meas}(I)$	0.225	0.064	0.751
$R_{meas}(I+/-)$	0.225	0.065	0.748
$R_{pim}(I)$	0.088	0.025	0.290
$R_{pim}(I+/-)$	0.122	0.035	0.402
Wilson B factor	18.06		
Anomalous completeness	99.6	99.2	99.7
Anomalous multiplicity	3.2	3.3	3.3
Anomalous correlation	-0.01	-0.073	0.000
Anomalous slope	0.965		
dF/F	0.176		
dI/s(dI)	1.042		
Total observations	1053972	47257	195594
Total unique	165659	7673	29762

A

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0000	0.0000	32.74	96.24	38558.00
2	0.0000	0.0000	16.37	92.49	77116.00
3	0.0000	0.0001	10.91	88.73	115674.00
4	0.0003	0.0014	8.19	84.97	154232.00
5	0.0013	0.0039	6.55	81.22	192790.00
6	0.0039	0.0080	5.46	77.46	231348.00
7	0.0095	0.0170	4.68	73.71	269906.00
8	0.0191	0.0302	4.09	69.95	308464.00
9	0.0297	0.0424	3.64	66.19	347022.00
10	0.0566	0.0667	3.27	62.44	385580.00
11	0.0860	0.0967	2.98	58.68	424138.00
12	0.1300	0.1329	2.73	54.92	462696.00
13	0.1612	0.1551	2.52	51.17	501254.00
14	0.1743	0.1602	2.34	47.41	539812.00
15	0.1550	0.1366	2.18	43.66	578370.00
16	0.1014	0.0872	2.05	39.90	616928.00
17	0.0447	0.0383	1.93	36.14	655486.00
18	0.0182	0.0155	1.82	32.39	694044.00
19	0.0060	0.0052	1.72	28.63	732602.00
20	0.0018	0.0016	1.64	24.87	771160.00
21	0.0009	0.0008	1.56	21.12	809718.00
22	0.0000	0.0000	1.49	17.36	848276.00
23	0.0000	0.0000	1.42	13.60	886834.00

B

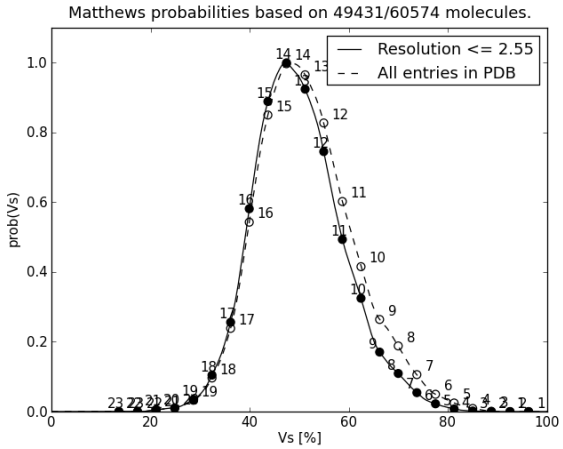
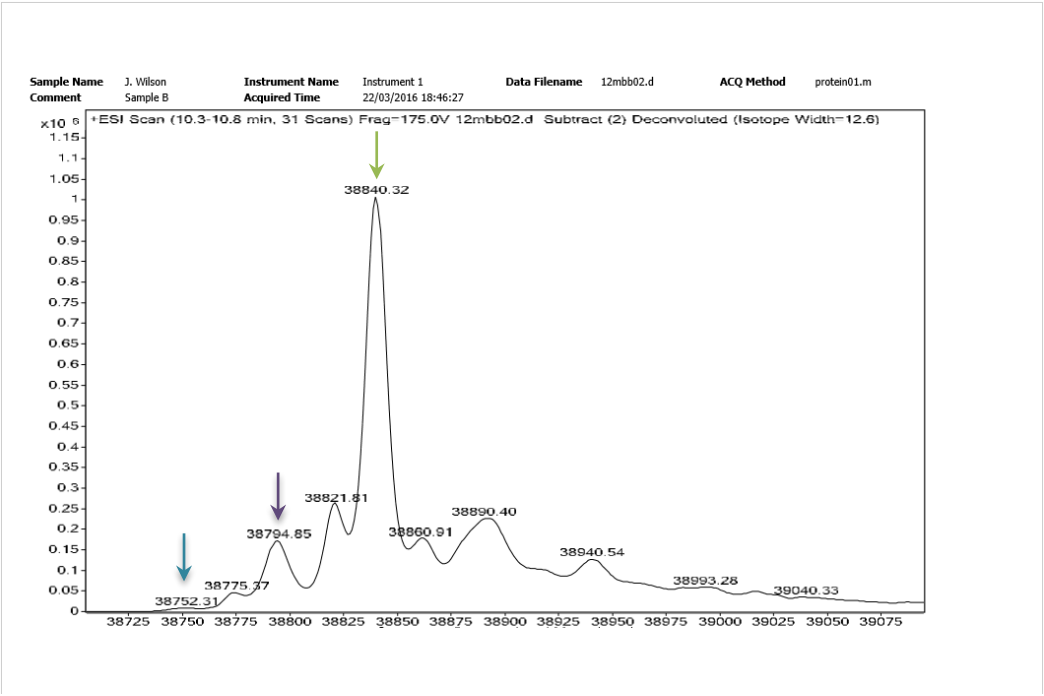


Figure 7.10. Matthews coefficient analysis of crystal form 2 grown in JCSG+ A11. (**A**) The Matthews coefficient analysis of crystal form 2 predicts that 12 and 14 molecules are most likely to be present in the asymmetric unit, with a Vm ranging from 2.73 $Å^3Da^{-1}$ and 2.18 $Å^3Da^{-1}$ respectively, and solvent content of between 54.9 % and 43.7 % respectively. (**B**) A plot of the probability of a given solvent content, prob(Vs), against solvent content, Vs, for all structures in the pdb, showing that the most likely cell contents is two AhlB per A.U.

A



B

Number of SeMets	Molecular weight of AhlB
0	38558.5
1	38605.4
2	38652.3
3	38699.2
4	38746.1
5	38793
6	38839.9

Figure 7.11. The mass spectrometry analysis of the AhlB SeMet purified protein. The mass spectrometry analysis of AhlB SeMet, with deconvolution carried out from 38700 Da to 39100 Da (**A**). Many peaks are present in the deconvoluted spectrum, and each was analysed with respect to the predicted species possible for a fully labelled, partially labelled, and unlabelled SeMet AhlB protein (**B**). The unlabelled AhlB protein has a mass of 38558.5 Da, calculated by Expasy Protparam. Each successive labelled methionine with selenium increased the molecular weight by 46.9 Da, such that the maximum molecular weight of fully labelled AhlB has a mass of 38839.9. The mass spectrum shows that the main species present is the fully labelled SeMet AhlB (highlighted by green arrow). There are also lower yields of 5-SeMet and 4-SeMet labelled protein (purple and blue arrows respectively).

diffracted to 6 Å. However, from the crystallisation screening, a largely reproducible crystal form was grown in a number of related conditions, which are highlighted in Table 7.1, and discussed below.

7.6 AhlB crystal form 3 (JCSG+ H12 and others)

The most successfully reproducible crystals of AhlB SeMet were grown in JCSG+ H12, and optimisations thereof. The crystals were generally plates (Figure 7.1) of different thicknesses that varied in diffraction quality based on growth conditions and optimisations. Test diffractions were carried out on many crystals, collecting 5 images 45° apart, with an oscillation and exposure time per image of 0.1° and 0.1s respectively. Indexing of the test data by Mosflm showed that the form 3 crystals belonged to spacegroup $C222_1$, with unit cell dimensions generally of a= ~115 Å, b= ~361 Å, and c= ~427 Å. Crystals grown in JCSG+ H12 and optimisations thereof diffracted to at most 3 Å, and more generally varied from 3 Å to 8 Å.

The Matthews coefficient analysis of crystal form 3 indicated that there were likely to be between 13 and 31 molecules in the asymmetric unit, with Vm's of 4.5 $Å^3$/Da and 1.9 $Å^3$/Da and solvent contents of 72 % and 34 % respectively. The most likely solution was 24 molecules per AU with a Vm of 2.4 $Å^3$/Da and a solvent content of 49 % (Figure 7.12).

7.6.1 SeMet MAD data collection and processing

JCSG+ H12 crystals that were grown from AhlB SeMet protein and diffracted well were used to collect largely redundant datasets at the peak, inflection, and high-energy remote wavelengths, in order to try to determine initial phases. Each anomalous wavelength was determined by a fluorescent scan at the selenium peak, and analysis of the Chooch plot showed that the peak and inflection wavelengths were at 0.9794 Å and 0.9795 Å respectively (Figure 7.13). The f' and f'' at peak and inflection wavelengths were -7.4e and 6.2e, and -10.1e and 3.2e respectively. The peak and inflection datasets were collected at the calculated wavelengths, and a high-energy remote dataset was collected at 0.9686 Å. For each, 250° of data was collected over 2500 images, with an oscillation per image of 0.1° and an exposure per image of 0.1 s, with an attenuated transmission of 10 % to minimise radiation damage. Each dataset was processed automatically by Xia2 automatic pipeline, and the best processing method was chosen to be 3dii based on the method that gave the lowest overall R_{merge} statistics. The processing statistics for each dataset are shown in Table 7.5.

A

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0000	0.0000	58.14	97.88	38558.00
2	0.0000	0.0000	29.07	95.77	77116.00
3	0.0000	0.0000	19.38	93.65	115674.00
4	0.0000	0.0000	14.54	91.54	154232.00
5	0.0000	0.0000	11.63	89.42	192790.00
6	0.0002	0.0003	9.69	87.31	231348.00
7	0.0005	0.0008	8.31	85.19	269906.00
8	0.0012	0.0015	7.27	83.08	308464.00
9	0.0020	0.0024	6.46	80.96	347022.00
10	0.0032	0.0037	5.81	78.84	385580.00
11	0.0047	0.0050	5.29	76.73	424138.00
12	0.0076	0.0081	4.85	74.61	462696.00
13	0.0110	0.0115	4.47	72.50	501254.00
14	0.0156	0.0161	4.15	70.38	539812.00
15	0.0198	0.0206	3.88	68.27	578370.00
16	0.0235	0.0240	3.63	66.15	616928.00
17	0.0303	0.0305	3.42	64.04	655486.00
18	0.0396	0.0399	3.23	61.92	694044.00
19	0.0494	0.0494	3.06	59.81	732602.00
20	0.0596	0.0594	2.91	57.69	771160.00
21	0.0716	0.0714	2.77	55.57	809718.00
22	0.0815	0.0811	2.64	53.46	848276.00
23	0.0876	0.0870	2.53	51.34	886834.00
24	0.0908	0.0902	2.42	49.23	925392.00
25	0.0906	0.0899	2.33	47.11	963950.00
26	0.0839	0.0832	2.24	45.00	1002508.00
27	0.0732	0.0726	2.15	42.88	1041066.00
28	0.0570	0.0565	2.08	40.77	1079624.00
29	0.0389	0.0386	2.00	38.65	1118182.00
30	0.0237	0.0236	1.94	36.53	1156740.00
31	0.0145	0.0144	1.88	34.42	1195298.00
32	0.0086	0.0085	1.82	32.30	1233856.00
33	0.0045	0.0044	1.76	30.19	1272414.00
34	0.0026	0.0026	1.71	28.07	1310972.00
35	0.0013	0.0012	1.66	25.96	1349530.00
36	0.0007	0.0007	1.62	23.84	1388088.00
37	0.0005	0.0005	1.57	21.73	1426646.00
38	0.0002	0.0002	1.53	19.61	1465204.00
39	0.0000	0.0000	1.49	17.50	1503762.00
40	0.0000	0.0000	1.45	15.38	1542320.00
41	0.0000	0.0000	1.42	13.26	1580878.00
42	0.0000	0.0000	1.38	11.15	1619436.00

B

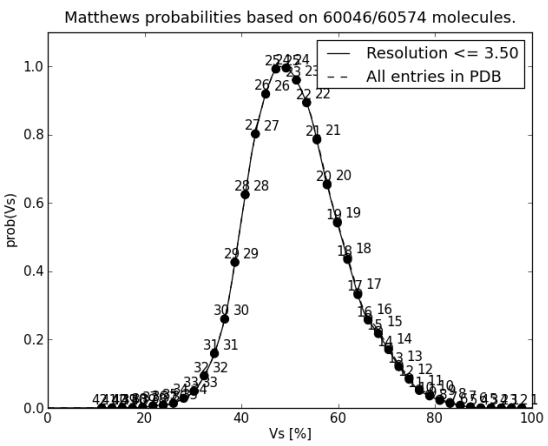
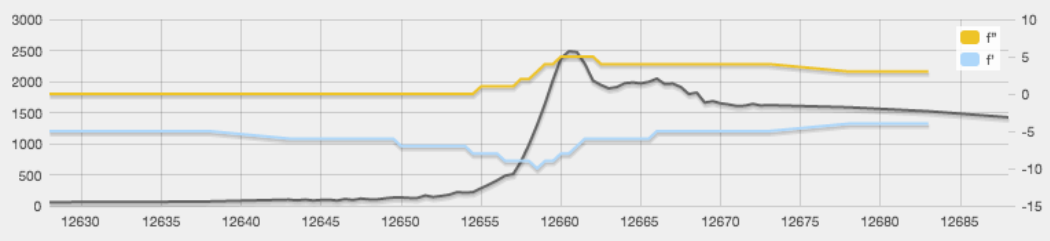


Figure 7.12. Matthews coefficient analysis of crystal form 3 grown in JCSG+ H12. (**A**) The Matthews coefficient analysis of crystal form 3 predicts that between 20 and 28 molecules are most likely to be present in the asymmetric unit, with a Vm ranging from 2.91 $Å^3Da^{-1}$ and 2.08 $Å^3Da^{-1}$ respectively, and solvent content of between 57.7 % and 40.8 % respectively. (**B**) A plot of the probability of a given solvent content, prob(Vs), against solvent content, Vs, for all structures in the pdb, showing that the most likely cell contents is 24 AhlB per A.U.

A



B

	Wavelength	f''	f'
Peak	12660.5eV (0.9793Å)	5.79	-7.45
Inflection	12658.5eV (0.9795Å)	3.58	-10.07

Figure 7.13. The selenium K edge fluorescence scan and calculated f' and f'' values. The selenium K edge fluorescence scan performed on a crystal of AhlB grown in JCSG+ H12 is shown in black (**A**), along with the calculated f' and f'' traces (blue and yellow respectively). The peak and inflection wavelengths were chosen to maximise f'' and minimise f' respectively, and their specific wavelengths and electron values are shown in **B**.

Table 7.5. Processing statistics for AhlB form 3 MAD data.

	Peak	Inflection	High energy remote
Spacegroup	$C222_1$	$C222_1$	$C222_1$
a (Å)	115.5	115.6	115.7
b (Å)	365.2	365.7	366.1
c (Å)	428.4	428.6	428.8
α (°)	90	90	90
β (°)	90	90	90
γ (°)	90	90	90
High resolution limit (Å)	3.35 (3.44)	3.56 (3.65)	3.76 (3.86)
Low resolution limit (Å)	142.8 (14.98)	168.2 (15.92)	168.36 (16.82)
Completeness (%)	100 (100)	100 (100)	100 (100)
Multiplicity	9.5 (9.7)	9.5 (9.5)	9.4 (9.3)
I/sigma	7.3 (1.0)	7.4 (1.2)	7.2 (1.4)
R_{merge}	0.275 (2.132)	0.280 (1.734)	0.294 (1.505)
$R_{meas}(I)$	0.314 (2.407)	0.318 (1.956)	0.332 (1.704)
$R_{meas}(I+/-)$	0.309 (2.386)	0.314 (1.945)	0.329 (1.694)
$R_{pim}(I)$	0.101 (0.768)	0.103 (0.629)	0.148 (0.772)
$R_{pim}(I)$ Low resolution	0.018	0.017	0.016
$R_{pim}(I+/-)$	0.139 (1.065)	0.141 (0.874)	0.148 (0.772)
CC half			0.994 (0.569)
Wilson B factor ($Å^2$)	131.49	185.75	172.59
Anomalous completeness (%)	100 (99.8)	99.9 (99.9)	99.9 (100)
Anomalous multiplicity	4.9 (4.9)	4.9 (4.9)	4.9 (4.7)
Anomalous correlation	0.121 (0.000 – 0.564)	0.085 (0.001 – 0.0473)	0.058 (0.008 – 0.026)
Anomalous slope	1.094	1.067	1.049
dF/F	0.162	0.164	0.171
dI/s(dI)	0.896	0.868	0.852
Total observations	1233738	1030786	875563
Total unique	130209	108989	92882

The high-resolution cut-off for each dataset was calculated in aimless as 3.4 Å, 3.6 Å, and 3.8 Å for the peak, inflection, and high-energy remote data respectively. The loss of diffracting power of the crystals after the peak dataset is to be expected due to radiation damage. Radiation damage was confirmed by analysing the merging R-values per batch, which increasing over the dataset when radiation damage is present.

The anomalous signal in each dataset was estimated by analysis of the midslope of anomalous normal probability, and the anomalous correlation coefficient (CCanom) per resolution bin. The cut-off for useful anomalous signal based on these two values was 6 Å for the peak dataset, 6 Å for the inflection dataset, and 7.3 Å for the high-energy remote dataset, further indicating radiation damage between datasets.

With an asymmetric unit that contained ~24 38 kDa monomers of AhlB, and each AhlB molecule containing up to 6 Se atoms, ~144 Se atoms were contained in each asymmetric unit. This made solving the substructure at low resolutions such as 5-7 Å very difficult. The SHELX and autoSHARP pipelines were used to try and solve the substructure, although no convincing substructure was ever determined.

SHELXC was run in order to determine the resolution cut-off for significant anomalous signal, based where the d”/sig approaches 0.8. Based on this, the anomalous signal at each wavelength approached 0.8 at 6 Å. The correlation coefficient was also determined between each dataset for signed anomalous differences, and fell below 30 % at 6 Å for each dataset comparison. As such, SHELXD was run with a high-resolution cut-off of 6 Å initially, searching for 120 Se heavy atoms. The SHELXD Patterson correlation coefficient for all data and for weak data was ~ 33 and ~ 6, with no large differences between runs, indicating that a correct solution was not found. The heavy atom sites from the SHELXD run were analysed for the presence of NCS using Phenix and professs (Adams et al., 2010; Winn et al., 2011), but no NCS operators could be generated.

Finally, in case the SHELXD solution was correct, SHELXE was run using both the original hand and the inverse hand of the heavy atom co-ordinates, and after 5 cycles of density modification and autotracing, the hand ambiguity was unbroken, the correlation coefficient for each hand against the native data was only 9 %, and visual observation of the output electron density map confirmed that the phases were not of sufficient quality to interpret the electron density.

7.6.2 Attempts to solve AhlB JCSG+ H12 form by collecting SeMet peak datasets from a number of crystals to increase the anomalous signal

As AhlB SeMet JCSG+ H12 crystals diffracted poorly to ~3.5 Å, with anomalous signal to ~6Å, many datasets were collected from many crystals to try to improve the merging statistics and overall anomalous signal. It has been shown that scaling and merging multiple datasets from different crystals can greatly increase anomalous signal, and lead to successful substructure solution, especially with low-resolution diffraction (Liu et al., 2011). The diffraction images for three peak data collections collected from two crystals were placed into a single folder and Xia2 3dii was run using all of the images. In this way, all of the reflections were locally scaled, and then globally scaled and merged. Results of scaling and merging multiple datasets are shown in Table 7.6. Each dataset was scaled and merged in isolation during the collection of each dataset, the statistics of which were used to determine data quality, and then these were compared to the combined data. The overall anomalous signal quality of this merged data was improved over the single datasets, with a CC_{anom} greater than 0.15 at 5 Å when all data is included, compared to 6 Å for the individual data sets. The R_{merge} per bin was used to determine any radiation damage in each dataset, and these images were removed and the scaling and merging repeated, although no increase in the resolution of useful anomalous signal was gained.

Phasing attempts were repeated with the combined data by Se-SAD. The anomalous signal was determined by SHELXC, followed by attempts to solve the substructure with SHELXD, searching for 120 Se sites, with a resolution cut-off set to 5 Å, based on the anomalous signal present determined in SHELXC. During 1000 tries to solve the substructure, no bimodal distribution of results was established, and the CC_{all} and CC_{weak} were all centred around ~44 and ~3 respectively, suggesting no correct solution had been determined. Density modification and chain tracing was attempted in SHELXE using the best solution from the SHELXD run, in both the original and inverted hand, in order to break the phase ambiguity and improve the phases in case a useful initial electron density map could be generated. As expected from the SHELXD results, no difference in CC% was established for the original hand and inverted hand, and the overall CC% against the native data was only 8 %. Finally, observation of the resulting electron density maps showed no clear solvent boundaries or continuous electron density that may be interpreted as protein electron density. As such, even combining multiple peak

Table 7.6.Merging statistics for three combined AhlB crystal form 3 peak datasets.

	Data 1	**Data 2**	**Data 3**	**Combined Data**
a (Å)	115.5	115.5	11.6.1	116.2
b (Å)	365.2	364.0	362.8	363.0
c (Å)	428.4	428.1	428.7	429.0
α (°)	90	90	90	90
β (°)	90	90	90	90
γ (°)	90	90	90	90
Resolution limit (high)	3.35 – 3.44	3.37 – 3.46	3.35 – 3.44	3.93 – 4.03
Resolution limit (low)	14.98 – 142.80	15.07 – 87.12	14.98 92.27	17.58 – 90.75
Resolution limit (overall)	3.35 – 142.80	3.37 - 87.16	3.35 – 92.27	3.93 – 90.75
I/sigma (overall)	7.3	6.6	7.5	13.9
I/sigma (High resolution)	1.0	1.1	1.2	3.6
Completeness (overall)	100	96.6	100	100
Completeness (High resolution)	100	98.4	100	99.3
R_{merge} (overall)	0.275	0.182	0.574	0.268
R_{merge} (high resolution)	2.132	1.266	3.732	0.570
R_{merge} (Low resolution)	0.037	0.031	0.076	0.084
R_{pim} (overall)	**0.139**	**0.101**	**0.101**	**0.038**
R_{pim} (High resolution)	0.768	0.660	0.677	0.204
R_{pim} (Low resolution)	0.018	0.021	0.018	0.012
CChalf (overall)	0.996	0.993	0.994	0.998
CChalf (High resolution)	0.464	0.464	0.565	0.897
Total observations	1233738	601058	4457175	3739313
Total unique observations	130209	122393	130101	81187

wavelength datasets from a number of crystals was unable to increase the phasing power and lead to a successful structure determination.

7.6.3 Heavy metal derivitisation of crystal form 3

In order to improve the phasing power at low resolutions and simplify the substructure solution, two strategies were employed. Soaking crystals with heavy atoms can be used in order to determine starting phases for model building. Mercury atoms contain 80 electrons, making it heavier than selenium by 46 electrons, giving it a larger scattering factor and a bigger anomalous signal. AhlB contains a single cysteine, and so mercurial compounds that covalently modify cysteine's can be used to produce a mercury derivitised protein.

The second strategy employed was derivitisation of AhlB with a much larger compound than Se, namely the Ta_6Br_{12} cluster, with a molecular weight of 2204.3 g/mol, and a molecular radius of ~4.3 Å. Each $Ta_6Br_{12}^{2+}$ ion adds 856 electrons to a protein, contributing hugely to scattering and phasing power of large proteins and multimeric systems (Knäblein et al., 1997). At resolutions lower than 6 Å, the cluster can be treated as a single scattering point, giving the resulting cluster a large contribution to anomalous scattering (Thygesen et al., 1996). Ta_6Br_{12} clusters target soft sites, meaning they do not covalently modify the protein, or form specific interactions with certain sidechains. Instead they target large pockets in the protein, which may aid in binding in the case where a large macromolecular assembly is formed.

Derivatives of protein crystals were generated by soaking crystals with heavy atoms and heavy atom clusters. Analysis of the protein sequence showed that AhlB contains 1 cysteine, which could be used to create a covalent bond with an ethyl-mercury compound for anomalous diffraction. The single site would simplify the substructure solution problem, but with only 1 heavy atom per 367 residues, the phasing power would be low.

Crystals from JCSG+ H12 optimisations were soaked in ethyl mercury phosphate by placing a small amount of ethyl-mercury powder in the drop containing crystals and soaking for between one hour and overnight. Crystals were then back-soaked to remove excess mercury, in a drop of the crystallisation reservoir to remove excess mercury. Back soaked crystals were then immediately placed in liquid nitrogen and stored for data collection.

AhlB crystals from JCSG+ H12 optimisations were also soaked in Ta_6Br_{12} by placing a small amount of the cluster, as a solid, into the crystallisation drop containing crystals. The soaks

were left overnight, and observed for the stages of soaking. At first the solid Ta_6Br_{12} dissolved into the drop, turning the entire drop green. After an overnight soak, the solution surrounding the crystals turned colourless again, and the green Ta_6Br_{12} compound was concentrated into the crystals, turning the crystals a deep green. Crystals were briefly backsoaked into a drop of mother liquor to remove excess Ta_6Br_{12}, and to reduce the increased radiation damage from excess Ta_6Br_{12} at peak wavelengths.

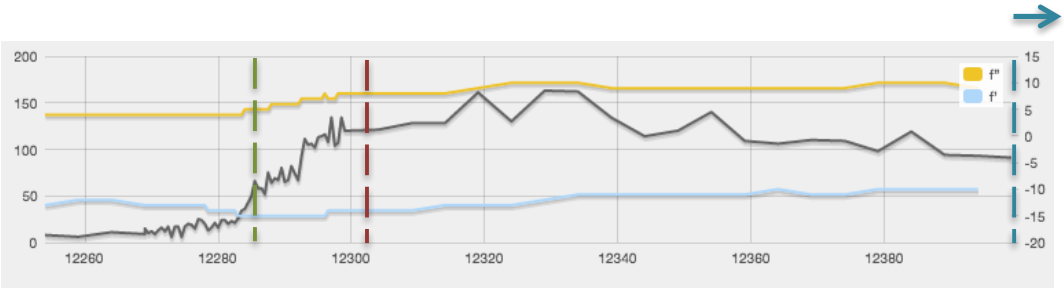
7.6.4 Mercury derivitised AhlB data collection and attempts to solve the structure

A fluorescence scan was carried out at the Diamond synchrotron, from 12240 eV to 12400 eV, scanning the Hg LIII edge (Figure 7.14). The fluorescent scan showed that crystals soaked for one hour and overnight in ethyl mercury phosphate contained very little mercury. f' and f'' values were estimated from the fluorescence scan using the program Chooch (Evans and Pettifer, 2001), and a peak wavelength for data collection was also estimated. A dataset was collected at the peak wavelength of 1.00 Å, collecting 180° of data with an oscillation and exposure per image of 0.2° and 0.1s respectively. The data were processed automatically by the 3dii Xia2 pipeline, and were processed into $C222_1$, with unit cell dimensions of a= 115.7 Å, b= 362.5 Å, and c= 428.2 Å, and a high-resolution cut-off of 3.3 Å (Table 7.7). The overall midslope of the data was 1.09, but with a high $CC_{1/2}$ value in the high-resolution bin. The absence of strong anomalous signal was confirmed by the CCanom value per resolution bin, where the CCanom value did not rise above 0.15 even at the lowest resolution. The R_{merge} per batch was consistently 0.2, and overall the R_{pim} for the low-resolution bin was 0.04, suggesting the data are good quality, and so taken together, it seems that mercury has not bound to the protein in the ethyl mercury phosphate soaking experiments. Three other crystals from the soaking experiment were also tested for presence of mercury by collecting peak datasets in the same way, at 1.0075 Å wavelength. In each case the CC_{anom} per resolution bin was below the 0.15 cut-off for all resolution bins, suggesting that none of the crystals had successfully bound mercury.

7.6.5 Sulfhydryl assay to determine the availability of the cysteine present in AhlB for derivitisation

As the mercury derivitisation experiment had not worked, the availability of the single cysteine present in AhlB to bind to mercury was tested by a sulfhydryl assay, using Ellmans

A



B

	Wavelength	f''	f'
Peak	9882.4eV (1.2546Å)	21.83	-17.25
Inflection	9879.8eV (1.2549Å)	14.27	-25.40

Figure 7.14. The mercury LIII edge fluorescence scan and calculated f' and f'' values. The mercury LIII edge fluorescence scan performed on a crystal of AhlB crystal form 3 is shown in black (**A**), along with the calculated f' and f'' traces (blue and yellow respectively). The vertical dotted lines show the wavelengths selected for the peak (red), inflection (green), and high-energy remote (blue) data collections. The peak and inflection wavelengths were chosen to maximise f'' and minimise f' respectively, and their specific wavelengths and electron values are shown in **B**.

Figure 7.7. Xia2 3dii processing statistics for AhlB Hg derivitised crystal form 3 crystals

	Overall	Low	High
High resolution limit	3.43	15.34	3.43
Low resolution limit	59.14	59.14	3.52
Completeness	99.9	97.4	100
Multiplicity	6.8	5.6	7.0
I/sigma	3.9	12.3	1.0
R_{merge}	0.401	0.104	1.669
$R_{meas}(I)$	0.480	0.126	2.033
$R_{meas}(I+/-)$	0.473	0.125	1.966
$R_{pim}(I)$	0.183	0.053	0.766
$R_{pim}(I+/-)$	0.249	0.069	1.029
Wilson B factor	105.045		
Anomalous completeness	99.5	95.8	99.8
Anomalous multiplicity	3.4	3.2	3.5
Anomalous correlation	0.023	0.057	0.022
Anomalous slope	1.006		
dF/F	0.215		
dI/s(dI)	0.819		
Total observations	823060	8274	61868
Total unique	120966	1474	8874

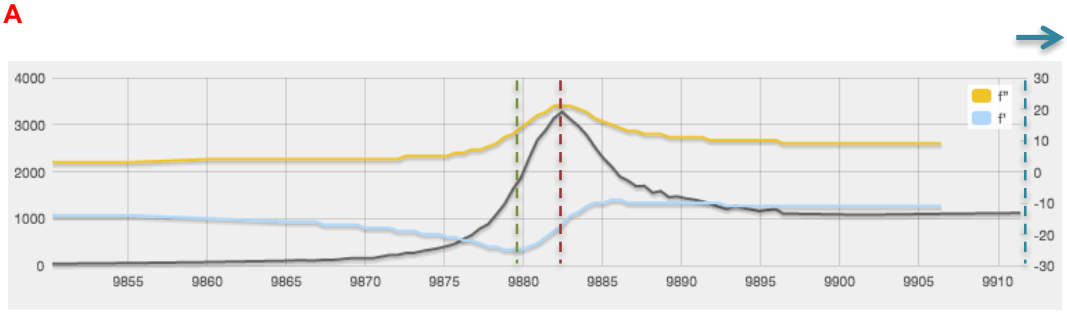
reagent, and performed under native and denaturing conditions. The assay was performed as described in 3.1.5. 10 and 20 µg of AhlB were incubated with 1 mM Ellmans reagent at room temperature for 2 minutes in either native or denaturing buffer, and then an absorbance reading was taken at 412 nm. Converting the absorbance to concentration of absorbing Ellmans reagent showed that the absorbances corresponded to no binding of the Ellmans reagent to AhlB. Therefore the Ellmans reagent was unable to bind to the cysteines present in the protein, suggesting the cysteine is not surface accessible. In the denaturing guanadinium-containing buffer, large absorbances at 412 nm were recorded. Conversion to concentration of Ellmans reagent reacted showed that all of the protein present in each experiment had reacted with Ellmans reagent, and that the cysteines were available when AhlB is unfolded.

This result suggests that mercury did not bind to AhlB in the crystal form 3 during soaking experiments, as the cysteine is unavailable. However, due to the change in the oligomeric state of AhlB from dimer in solution to the species present crystal form 3, and the possible conformational change associated with this change, it cannot be ruled out that the cysteine may become available in the crystal, although the lack of mercury signal after soaking suggests otherwise.

7.6.6 Ta_6Br_{12} derivitised AhlB data collection and processing

Crystals soaked overnight in Ta_6Br_{12} were backsoaked in cryosolution to remove excess tantalum cluster compound and then cooled in liquid nitrogen and sent to the Diamond synchrotron for data collection. A fluorescence scan was carried out first to determine the correct wavelength for each of the peak, inflection, and high-energy remote data collections. The fluorescence scan also confirmed the presence of tantalum, although due to soaking, tantalum would be present in the crystal lattice even in the absence of specific ordered binding to the protein. The fluorescence scan was collected at the Ta K edge, from 9850 eV to 9912 eV, and a peak in fluorescence at the Ta edge was present at 9982 eV (Figure 7.15). The peak and inflection wavelengths were calculated from the fluorescence scan, as well as f'' and f' for each wavelength (Figure 7.15). The peak wavelength was calculated at 9882.4 eV, with a f'' and f' of 21.8e and -17.3e respectively, and the inflection wavelength was calculated to be at 9879.8 eV with a f'' and f' of 14.27e and -25.40e respectively.

A



B

	Wavelength	f''	f'
Peak	9882.4eV (1.2546Å)	21.83	-17.25
Inflection	9879.8eV (1.2549Å)	14.27	-25.40

Figure 7.15. The tantalum edge fluorescence scan and calculated f' and f'' values. The mercury LIII edge fluorescence scan performed on a crystal of AhlB crystal form 3 is shown in black (**A**), along with the calculated f' and f'' traces (blue and yellow respectively). The vertical dotted lines show the wavelengths selected for the peak (red), inflection (green), and high-energy remote (blue) data collections. The peak and inflection wavelengths were chosen to maximise f'' and minimise f' respectively, and their specific wavelengths and electron values are shown in **B**.

A full highly redundant dataset was collected from a number of crystals based on these energies, collecting a peak, an inflection, and a high-energy remote dataset from each. The high-energy remote wavelength was chosen to be 10399 eV. At each wavelength, 250° of data were collected over 1250 images, with an oscillation per image of 0.2° and an exposure time of 0.05s. The X-ray beam transmission was attenuated to minimise radiation damage of anomalous datasets. Each dataset was processed automatically by the Xia2 automatic pipeline at the Diamond synchrotron, and the processing statistics for each of the peak, inflection, and high energy remote for the best diffracting crystal sample are shown in Table 7.8

Despite reducing the exposure for the crystal, radiation damage was clearly present in the processing statistics, from the expansion of the crystal cell dimensions by more than 1 % and a severe loss of diffraction resolution from 4 Å to 6 Å between peak and high energy remote data collections. Each of the anomalous datasets were analysed for how much anomalous signal was present in each. The metrics used to determine the level of anomalous signal were the midslope of anomalous normal probability, the CCanom per resolution bin and the R_{anom}.

7.6.7 Attempts to solve the structure of AhlB crystal form 3 by Ta_6Br_{12}

Each of the peak, inflection, and high-energy remote datasets were used in phasing attempts using either Ta-SAD or Ta-MAD using SHELX. The anomalous signal present in each dataset was determined in SHELXC, as well as a suitable resolution cut-off for substructure resolution, based on where the d"/sig fell below 1. SHELXC also output a recommended resolution cut-off for use in SHELXD based on where d"/sig fell below 0.8, a less conservative estimate.

SHELXD was run with the recommended resolution cut-off of 5 Å, with 1000 trials to solve the substructure. The average CCall and CCweak were 43 and 6 respectively, and no bimodal distribution between runs was achieved, indicating that no successful solution had be determined.

7.6.8 Other MPD containing crystallisation conditions yielding crystals of AhlB crystal form 3

Other MPD-containing conditions yielded crystals of diffraction quality with similar large unit cells, which are summarised in Table 7.1. Many of these crystals diffracted poorly, and so were not useful in solving the crystal structure of AhlB, but did provide more useful information as to

Figure 7.8. Xia2 3dii processing statistics for MAD data collections on AhlB crystal form 3 crystals derivitised with Ta_6Br_{12}.

	Peak	**Inflection**	**HE remote**
Wavelength	1.2546 Å	1.2549 Å	1.19216 Å
Spacegroup	$C222_1$	$C222_1$	$C222_1$
a (Å)	116.5	117.6	118.2
b (Å)	363.3	366.2	368.0
c (Å)	429.9	435.1	438.4
Resolution limit (Å)	4.40	5.23	6.03
I/sigma (overall)	3.2	3.4	3.3
I/sigma (High resolution)	1.4	0.9	0.7
Completeness (overall)	99.9	99.9	99.9
Completeness (High resolution)	99.9	100	100
Multiplicity (overall)	9.2	9.1	8.9
R_{merge} (overall)	0.652	0.572	0.562
R_{merge} (high resolution)	1.537	2.390	3.185
R_{merge} (Low resolution)	0.129	0.135	0.203
R_{pim} (overall)	0.253	0.22	0.223
R_{pim} (High resolution)	0.598	0.882	1.212
R_{pim} (Low resolution)	0.064	0.067	0.097
CChalf (overall)	0.978	0.976	0.969
CChalf (High resolution)	0.798	0.567	0.183
Anomalous multiplicity	4.8	4.8	4.7
Anomalous slope	**0.939**	**0.908**	**0.935**
Total observations	534760	329348	215545
Total unique observations	58425	36100	24156

the cell contents of the crystals containing MPD, and are included here as potential optimisation targets for future experiments.

7.7 Non-crystallographic symmetry analysis of AhlB crystal forms 2 and 3

AhlB crystal forms 2 and 3 both contained large, unrelated unit cell volumes, and a conserved cell dimension of a=115 Å, indicating that a related non-crystallographic symmetry may be present in both crystal forms. Analyses were carried out on the unit cell contents of crystal form two and three by analysing self-rotation functions generated for both crystal forms of AhlB. The self-rotation plots were generated with varying parameters, to determine if any NCS was present, and to posit the possible significance of any NCS as to the biological function of AhlB.

7.7.1 NCS analysis of AhlB crystal form 2

Self-rotation functions were carried out on the best dataset (resolution cut-off of 2.5 Å) at various integration radii (20 – 50 Å) and resolution cut-offs (3 and 6 Å) using Polarrfn and Molrep self-rotation function programs in CCP4. The default ncode 1 was used in the self-rotation functions, which places a, c*xa, and c* along x, y, and z. The peaks in the self-rotation function were plotted for peaks greater than 30 % of the origin in Polarrfn, and the top isoline level was set to 3 sigmas for the Molrep output. The self-rotation functions were carried out at a range of resolutions and integration radii in order to increase the significance of any peaks, and to determine where the NCS falls off.

The results of the Polarrfn show that there are significant peaks in the self-rotation at kappa = 180°. These peaks are present at omega = 90, at 36° intervals in the phi angle (Figure 7.16 and 7.17). At both 30 Å and 40 Å integration radii these peaks are present at 18° intervals. By working out the periodicity of these peaks i.e. 360/18, it was determined that these peaks represented 20 two-fold axes of NCS. The orientation of these peaks at omega = 90° and along phi, using the ncode 1, places these 20 2-fold symmetries in a plane

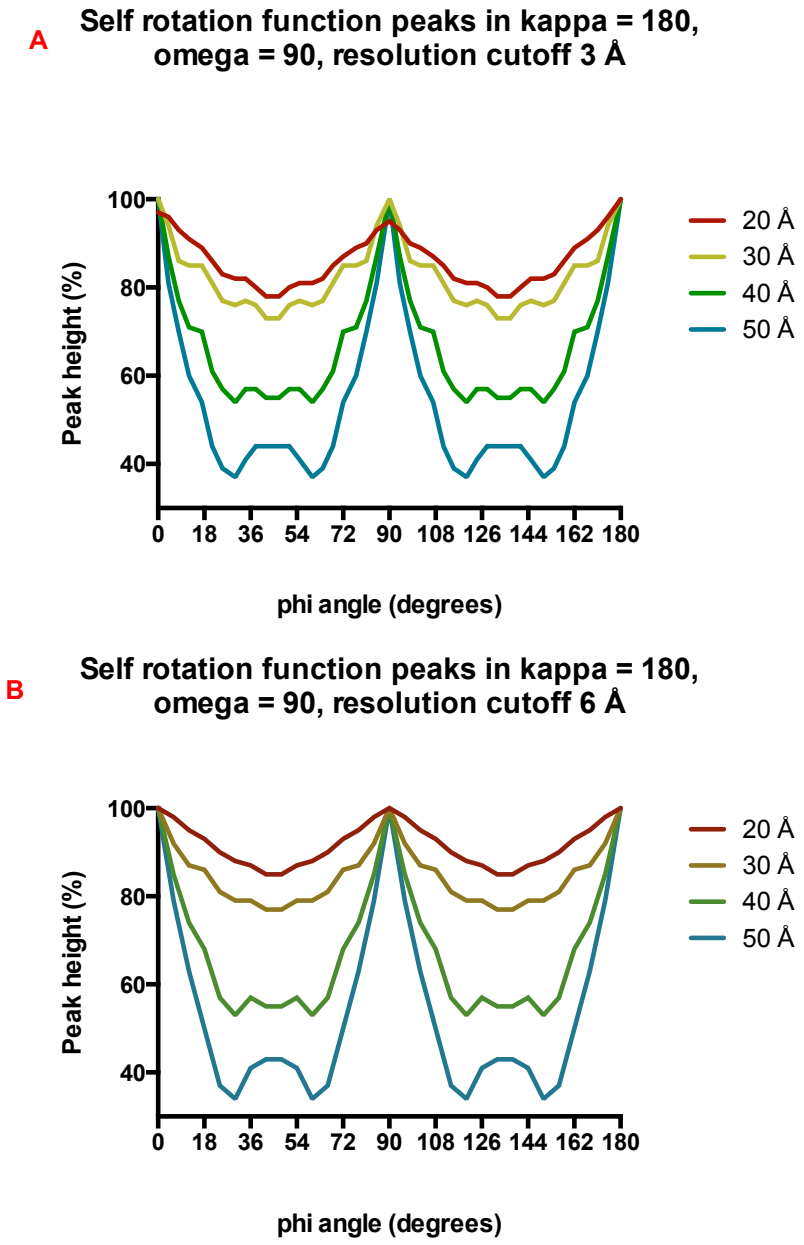
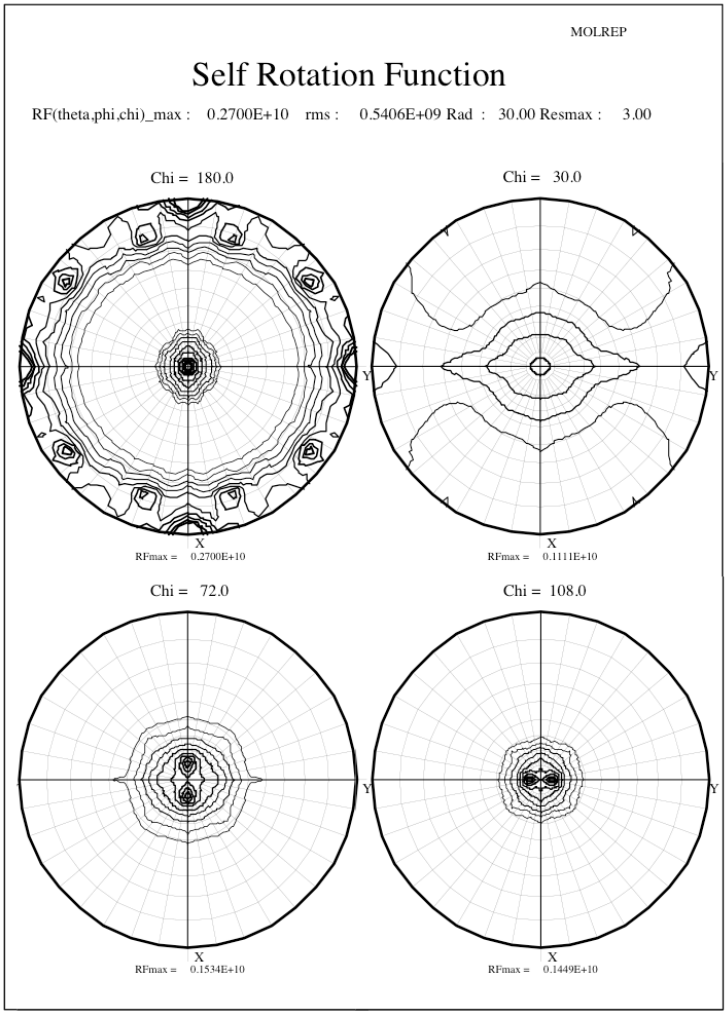


Figure 7.16. The peak height for each phi angle plotted on a graph for a number of radii of integration, and for two resolution cut-offs for crystal form 2. Self rotation functions were carried out first including all high resolution data (**A**) at 20 – 50 Å radii of integration, and the peaks at kappa 180°, omega 90°, and all phi are plotted. At 20 Å the definite peaks for ten two-fold symmetries is unclear, and becomes more clear as the radius of integration is increased. (**B**) The same general trend is observed in the SRF when the high-resolution data are omitted with a resolution cut-off of 6 Å, but more resolved for each radius of integration.

A



B

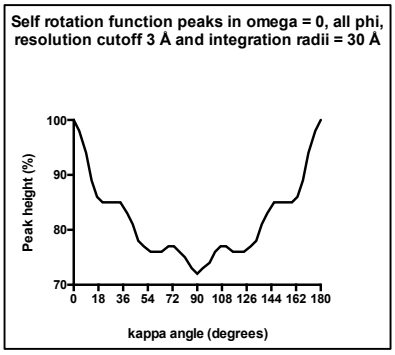


Figure 7.17. Self-rotation plots of AhlB crystal form 2. (**A**) Self-rotation plots of kappa = 180°, kappa = 36°, kappa = 72°, and kappa = 108°, generated with a radius of integration of 30 Å, and a resolution cut-off of 3 Å. In the kappa = 180°, peaks are present at omega = 0 and phi intervals of 36 °. In subsequent kappa plots excluding 36°, a peak is present at omega = 0° and phi = 0°. A slight figure of eight is present, suggesting that the NCS is not exactly parallel to C, but slightly misaligned, as indicated by the slight inward positioning of the peaks at phi intervals and kappa = 180° (**B**) The peaks in omega = 0°, phi = 0°, and intervals of kappa are plotted as the peak percentage of the origin.

perpendicular to C. As 20 two-fold symmetries were present, the kappa = 18° section was checked to determine whether a 20-fold symmetry was also present along C. The peaks in varying kappa angles showed that there was a peak at omega = 0°, and all phi angles. These peaks were plotted, and showed that although no peaks were present at 18°, peaks were present at 36°, 72°, 108°, and 144°, signifying 10-fold, 5-fold, 3.3-fold, and 2.5-fold NCS respectively (Figure 7.16 and 7.17). No peak is visible in due to the sigma cut-off and the closeness of 36° to the origin peak, however the peak at 36° can be seen when the peak heights are plotted on a graph (Figure 7.16) The alignment of these peaks at omega = 0° places the symmetry along the Z axis, which corresponds to the C cell edge in this orthorhombic crystal form.

A model was devised in order to try to explain the self-rotation function results. The combination of the 20 two-fold symmetries in the plane perpendicular to C, and the 10-fold symmetry along the C axis was rationalised as two 10-fold symmetric assemblies, aligned in the cell such that each was related to the other by a two-fold symmetry. In this model, each monomer in the ten-fold symmetric unit is related by a 36° rotation to its neighbour, and multiples of this 36° to superpose the starting molecule onto any of the other 9 molecules. The whole ten-fold symmetric assembly is then related to the second ten-fold symmetric assembly by a two-fold symmetry axis along B. The extra two-fold symmetries present, that increase the number of two-fold symmetries to 20 are present in between molecules, which also relate one assembly to another, and explain why the peaks at kappa = 180° are seen at 18° intervals, but no peaks are observed at kappa = 18°.

The 10-fold non-crystallographic symmetry parallel to C, and 20 two-fold non-crystallographic symmetries perpendicular to C also confine the diameter of the resulting molecule, as the NCS axes cannot overlap the crystallographic two-fold symmetries present in the $C222_1$ spacegroup. As such, the largest the symmetric assembly can be is 115 Å, the length of the A unit cell dimension. An exception to this rule of not overlapping exists if the 10-fold symmetry is parallel and coincident with the crystallographic two-fold symmetry axes. However, in this case, the maximum diameter of the 10-fold assembly is still 115 Å. Based on these analyses, it is predicted that AhlB crystal form 2 contains 10 molecules in the asymmetric unit, with the 10-fold axis parallel to C and with one of the 20 2-fold axes along A.

7.7.2 NCS analysis of AhlB crystal form 3

The same analyses presented above for crystal form 2 were undertaken using data collected from AhlB crystal form 3, due again to its large cell volume. Polarrfn and Molrep self-rotation functions were carried out identically to those carried out for crystal form 2, and analysed for the presence of NCS.

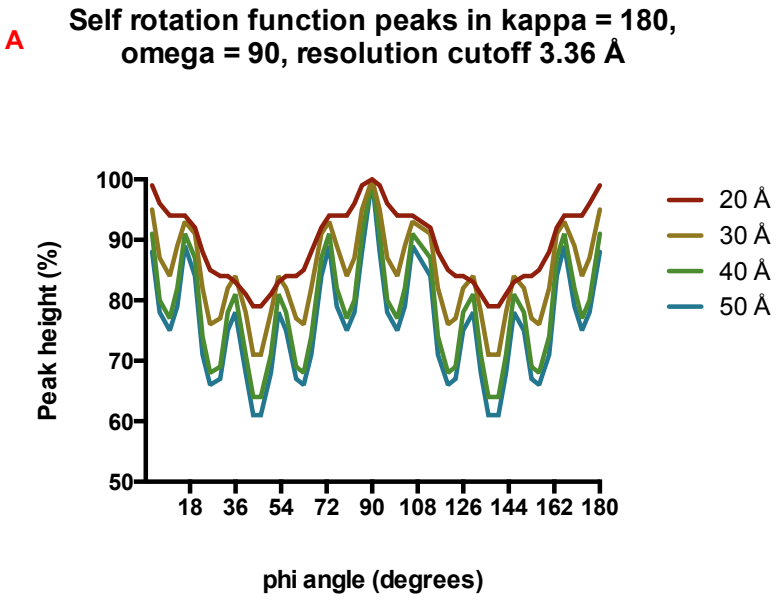
Once again, peaks were present in the self-rotation functions at kappa = 180°, omega = 90°, at phi angle multiples of 18° (Figure 7.18 and 7.19). The peaks at these angles are present for integration radii ranging from 20 Å (weakly present) to 50 Å, and at 3.4 Å and 6 Å high-resolution cut-offs (Figure 7.18). Peaks were also present at kappa angles 36°, 72°, 108°, and 144°, at omega = 0° and all phi angles, or along C. These results are identical to the self-rotation function results of crystal form 2. However, the difference in cell volume suggests that this crystal form contains 20 molecules, and so the symmetry is internal to the asymmetric unit, rather than between two adjacent asymmetric units as is predicted for crystal form 2.

7.7.3 Using the NCS present in crystal form to try and identify a correct substructure solution in the heavy atom derivitised data

In the light of the high degree of NCS present in crystal form 3, the data were revisited in order to try to identify a correct substructure solution. A single mercury site present per monomer of AhlB had an added advantage in crystal form 3, in that the 10-fold symmetry would place all mercury atoms in a plane, allowing the correct substructure to be determined more easily, and the location of mercury atoms that are not present in the first heavy atom Patterson search to be identified by determination of symmetry of the initial mercury sites (Fortelle and Bricogne, 1997).

7.7.4 The biological significance of the conserved 10-fold symmetry in crystal forms containing MPD

The presence of 10-fold symmetry is very unusual in biological crystals, and high degrees of non-crystallographic symmetry usually arise from biological assemblies. However, as AhlB elutes from gel filtration as a dimer, it is difficult to conclude that such a large assembly has formed in solution, or is at least present in the conditions of the gel filtration analysis. It was recently reported that the presence of MPD in the crystallisation of pore forming toxin α-haemolysin was able to induce pore formation, and the structure of this pore has been



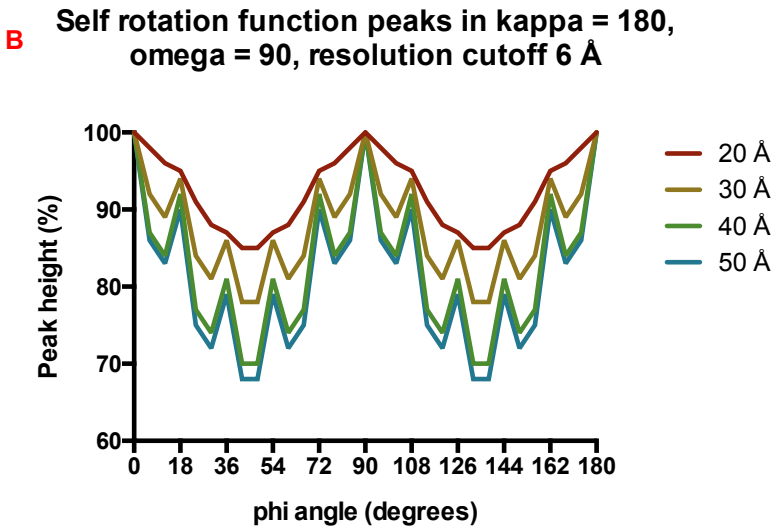
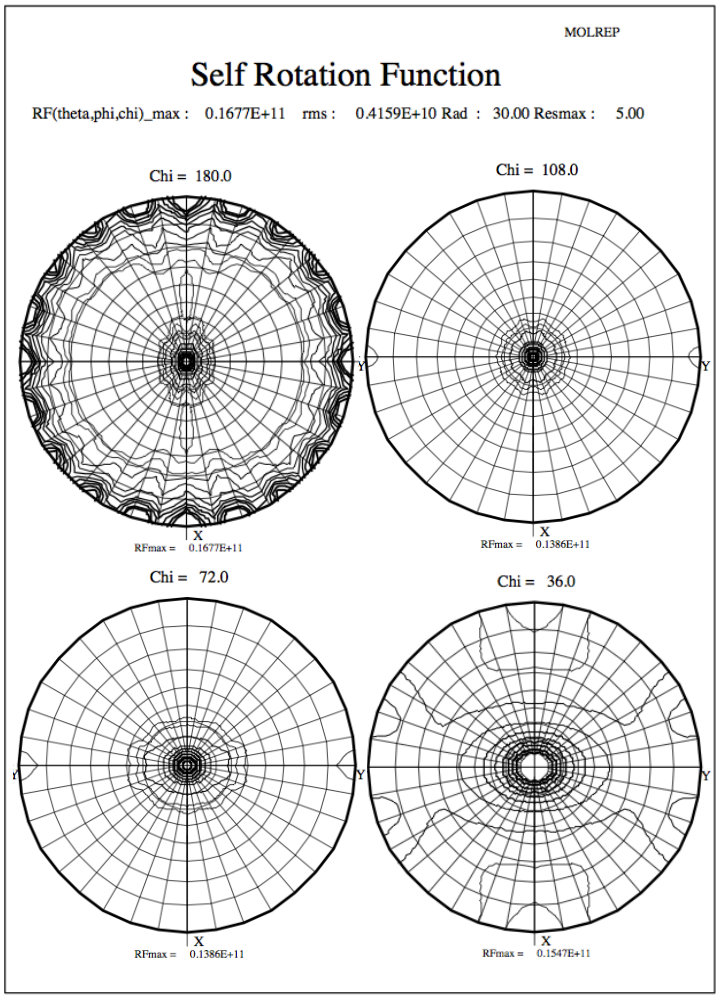


Figure 7.18. The peak height for each phi angle plotted on a graph for a number of radii of integration, and for two resolution cut-offs for crystal form 3. Self rotation functions were carried out first including all high resolution data (**A**) at 20 – 50 Å radii of integration, and the peaks at kappa 180°, omega 90°, and all phi are plotted. At 20 Å the definite peaks for ten two-fold symmetries is unclear, and becomes more clear as the radius of integration is increased. (**B**) The same self rotation plot as in A, but with a resolution cut-off for the self rotation calculation of 6 Å.

A



B

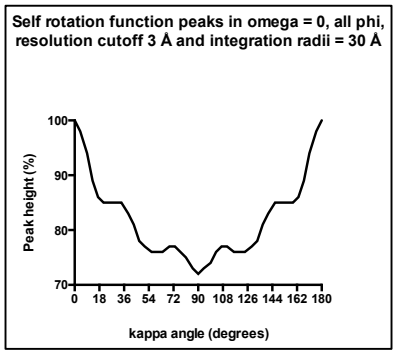


Figure 7.19. Self-rotation plots of AhlB crystal form 2. (**A**) Self-rotation plots of kappa = 180°, kappa = 36°, kappa = 72°, and kappa = 108°, generated with a radius of integration of 30 Å, and a resolution cut-off of 6 Å. In the kappa = 180°, peaks are present at omega = 0 and phi intervals of 18°. In subsequent kappa plots (108°, 72°, and 36°), a peak is present at omega = 0° and phi = 0°. (**B**) The peaks in omega = 0°, phi = 0°, and intervals of kappa are plotted as the peak percentage of the origin.

solved (Tanaka et al., 2011). The research found that the MPD bound to α-haemolysin in the same pocket as lipid, and so induced the conformational change required to form a pore. It would therefore follow that the highly symmetric assembly present in crystal form 2 and 3 may be a 'pore' formed of AhlB.

The orientation of the molecules present in crystal form 2 and 3 are reminiscent of the ClyA pore structure (PDB accession code = 2WCD), in which the asymmetric unit contains two 12-mer pores related by two-fold symmetry (Figure 7.20). The asymmetric unit of the ClyA pore structure contains two pores stacked tail to tail, generating the two-fold symmetry between each monomer in the pore to the opposite pore.

As it is has so far proved impossible to fully determine the structure of AhlB, the proposed pore form that the studies described here indicate is examined in the next chapter using electron microscopy. It was hoped that a molecular envelope generated by EM may have proved useful in either generating a starting set of phases either by MR, or by *ab initio* techniques, or may provide a NCS mask for averaging of poor experimentally phased data.

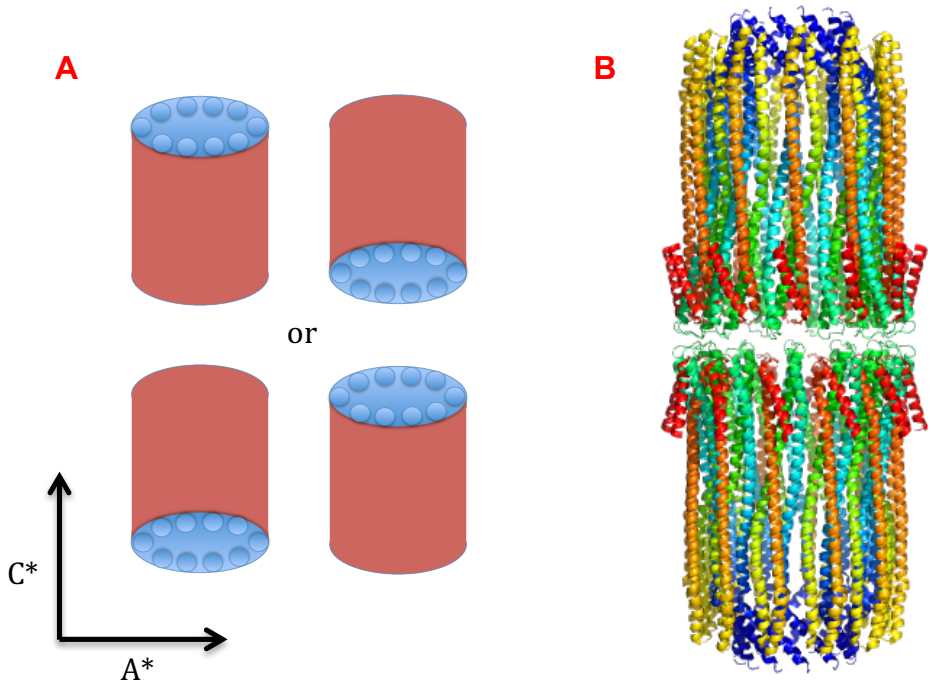


Figure 7.20. The prediction of the cell unit contents for AhlB crystal form 2 and 3 based on the self rotation function results. (**A**) In both crystal form 2 and 3, there is a ten-fold symmetry around C, with ten and twenty two-fold symmetries perpendicular to C. As such, a prediction of a ten-fold pore stacked either tail to tail or head to head, aligned in the unit cell as shown by the direction vectors. (**B**) Ribbon representation of ClyA pore form AU, containing two pores (PDB accession 2WCD). The AhlB MPD crystal forms are predicted to be similar to the crystal packing seen in the ClyA pore structure where two pores stack tail to tail in the unit cell.

Chapter 8: Characterising the pores formed by AHL

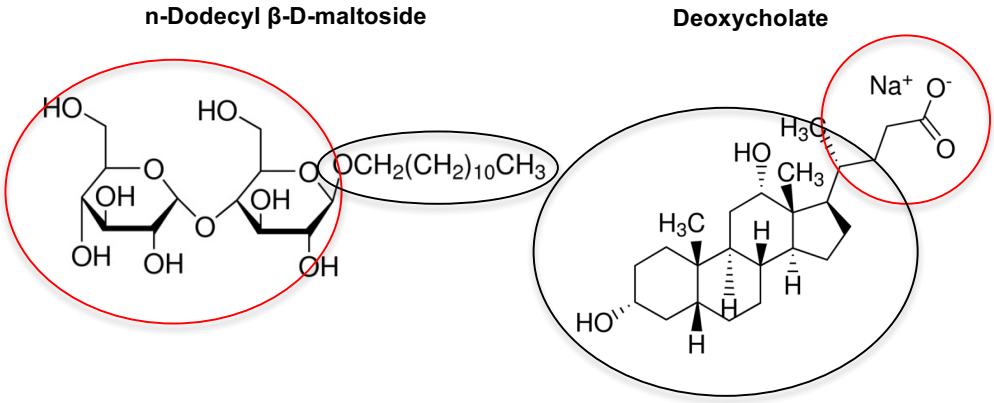
The final chapter of this thesis explores the nature of the AHL pore forming proteins in the membrane bound state i.e. the association of monomers with membranes, and formation of pores in membranes and detergents. An initial series of experiments using circular dichroism and gel filtration were carried out to determine the effect of detergents and lipids on AHL proteins. These preliminary experiments were followed by electron microscopy studies to visualise AHL proteins in membranes, and to observe and characterise the pores formed. Initial cryo-EM studies were also carried out that show that the AhlB and AhlC pores are amenable to structural analysis using cryo-EM.

8.1 Studies of AHL conformational change and oligomerisation when incubated with detergents and lipids

In order to study membrane proteins structurally, they must first be solubilised in lipids and/or detergents, and then purified in conditions that maintain the native structure. Experimentally this can be difficult, and usually extensive screening of detergents and lipids are required (Privé, 2007). PFTs are produced in the cell cytosol as soluble proteins, which eliminate the problems with initial solubilisation prior to purification that occur with many membrane proteins. Detergents and lipids are, however, required to induce pore formation in PFTs, in which conformational changes and oligomerisation occur. For detailed study, the resulting pores must then be purified in order to separate away any pore intermediates and aggregates.

Experiments were carried out with the AHL components and various detergents to attempt to generate pore assemblies that were amenable to crystallographic studies and single particle electron microscopy. A number of detergents were used in these studies, including n-Dodecyl β-D-maltoside (DDM) and deoxycholate (DOC) (Figure 8.1). DDM has been explored previously with the ClyA family PFTs to generate pores for crystallisation (Mueller et al., 2009) and electron microscopy studies (Eifler et al., 2006; Hunt et al., 2008; Tzokov et al., 2006). DOC was used as it is an anionic detergent and the negative charge was hypothesised to be complementary to the belt of lysine residues on AhlC, next to the hydrophobic hairpin. Modest success has also

A



B

Detergent name	Description	Critical micelle concentration	Micelle molecular weight
n-Dodecyl β-D-maltoside (DDM)	Non-ionic	0.15 mM at 25 °C	~50,000
Sodium deoxycholate (DOC)	Anionic	2-6 nM at 25 °C	~1200-5000

Figure 8.1. The chemical structures and properties of n-Dodecyl β-D-maltoside and Deoxycholate. (**A**) The chemical structures of n-Dodecyl β-D-maltoside and Deoxycholate show the hydrophilic head groups of each (red box), and the hydrophobic moieties (black box). (**B**) n-Dodecyl β-D-maltoside is a non-ionic detergent with a micellar molecular weight of ~50,000. Deoxycholate is negatively charged and has a much smaller micellar molecular weight of ~1200-5000.

been achieved with DOC as a pore inducing detergent for α-hemolysin (Bhakdi, 1981). Detergents and lipids were also used in attempts to understand the dynamics of pore formation, and probe more deeply the requirement of each AHL component in pore formation.

8.2 Probing AHL conformational changes by CD

Circular dichroism experiments based on those undertaken by Benke on ClyA (Benke et al., 2015), were used to determine if any changes occur in secondary structure, possibly due to conformational rearrangements, upon incubation of AHL components with detergent and lipid. In these reported experiments with ClyA, incubation of the protein with DDM caused a decrease in ellipticity at 222 nm and 208 nm, followed by an increase to greater negative ellipticity, than ClyA in solution. These results are reported to follow the conformational change of ClyA from a soluble protomer to an intermediate state, and finally to the pore form, in which more helical secondary structure is present.

8.2.1 Circular dichroism of AhlB in the presence and absence of DDM

The CD spectrum of AhlB in the soluble state was measured in order to estimate its secondary structure features in solution. Circular dichroism experiments were carried out as described in materials and methods section 3.1.9. 200 µL of 1 mg/mL AhlB was aliquoted into a 0.1 cm path length quartz cuvette and an ellipticity reading was measured for AhlB, from 190 – 250 nm. The high-tension voltage, which can be treated as a pseudo-transmission, where an increased voltage correlates to a decreased transmission of light, was used in these experiments to determine a cut-off for meaningful data. Where the high-tension voltage increased above 550 V, the ellipticity values were deemed non-meaningful (Dichroweb User Guide). The CD spectrum of AhlB shows that the protein is largely alpha helical, as there are large negative degrees of ellipticity at 222 nm and 208 nm, of -25768.4 $deg.cm^2.dmol^{-1}$ and -22794 $deg.cm^2.dmol^{-1}$ respectively, corresponding to an alpha helical content of ~ 65 % (Figure 8.2). This result is in agreement with the secondary structure of AhlB as predicted from the amino acid sequence and Phyre modelling (Figure 8.2), and the structure of NheA, which contains 72 % helical content.

The CD experiments were repeated for AhlB in the presence of 0.1 % (w/v) DDM detergent, in order to compare the spectrums of AhlB and AhlB in the presence and absence of detergent.

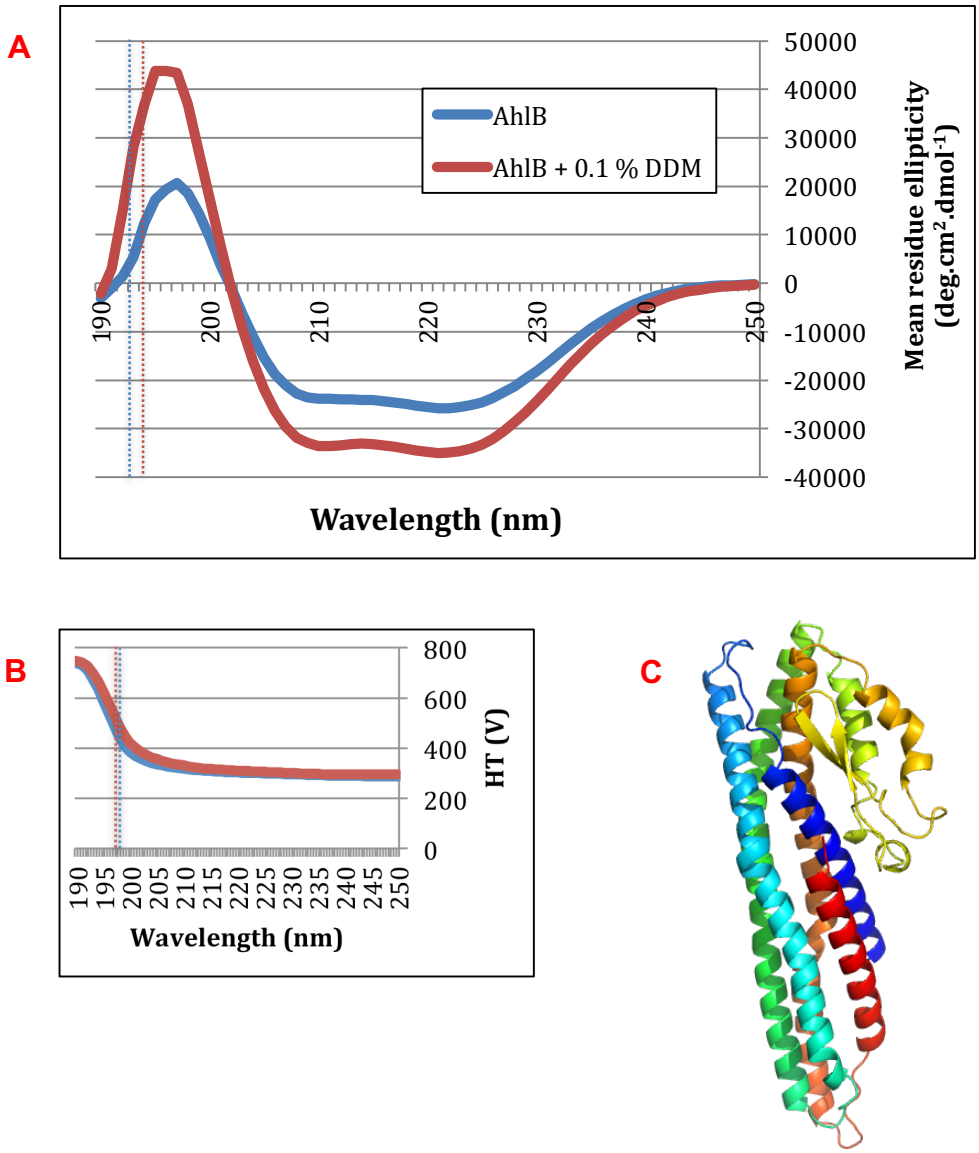


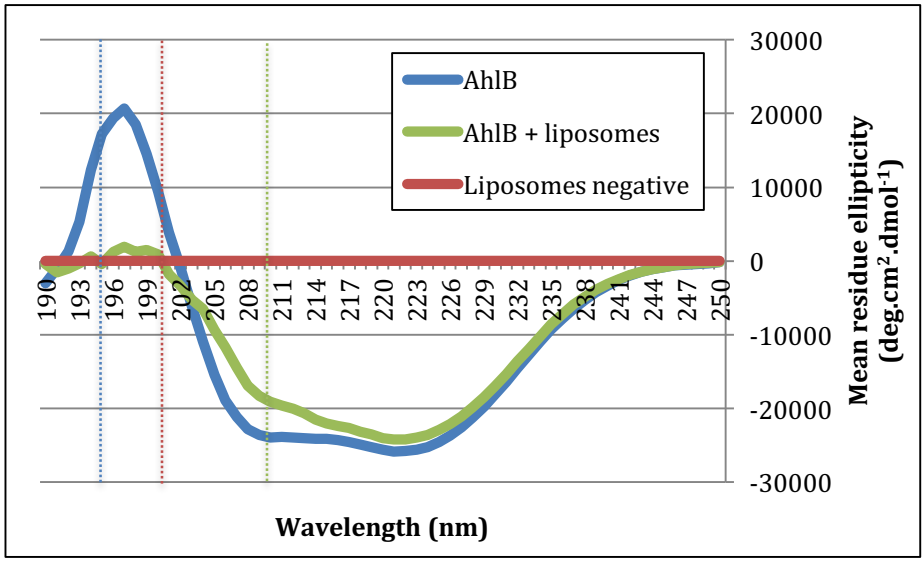
Figure 8.2. Results of the CD measurements taken for AhlB in 10 mM PBS pH 7.4 and in the same buffer containing 0.1 % (w/v) DDM. (A) The circular dichroism spectrum from 190 nm to 250 nm for 1 mg/ml AhlB (blue line) and for 200 µg AhlB incubated with 0.1 % (w/v) DDM for 1 hour at 37 °C (red line). Raw CD measurements were converted to mean residue ellipticity using the formula in 3.1.9. The dashed vertical lines indicate the cut-off for useful data as determined by the high-tension increasing above 550 V. Readings were measured three times and averaged for the final values (**B**) The high tension of the CD instrument during ellipticity measurements from 190 nm to 250 nm. The values are used here as a pseudo-transmission measurement, where a HT voltage of 550 V indicates the signal-to-noise ratio is high, and the mean residue ellipticity measurements are insignificant (Miles and Wallace, 2016). (**C**) A ribbon representation of the predicted AhlB structure, based on the structure of NheA (PDB accession 4K1P), which agrees with the MRE profile for AhlB.

A 200 µL sample was prepared with 1 mg/mL AhlB final concentration and 0.1 % (w/v) DDM, and this was incubated at 37 °C for 1 hour. A 90 – 250 nm absorbance spectrum was measured by circular dichroism. Addition of 0.1 % (w/v) DDM to AhlB increases the negative ellipticity measurement of the protein at 208 nm and 222 nm compared with AhlB without detergent (Figure 8.2). This change in ellipticity suggests that the protein has become more α-helical in nature in the presence of detergent, with an increase of ~ 65 % alpha helices to ~87 %. This result is similar to those seen for ClyA, in which the ellipticity at 208 nm and 222 nm also increased when incubated with 0.1 % (w/v) DDM. In ClyA, a conformational change of the β-tongue to α-helical hairpin occurs (Mueller et al., 2009), and the increase in α-helix content results in the change in ellipticity. The sequence analyses of AhlB described in Chapter 4 show that AhlB has a hydrophobic stretch of residues. These residues could either form a β-tongue, as seen in ClyA, HblB, and NheA, or could be α-helical as seen in AhlC. One possible explanation for the increase in ellipticity seen with AhlB and DDM could be that these residues do form a β-tongue, which changes conformation to α-helices in the presence of DDM.

8.2.2 Circular dichroism of AhlB in the presence and absence of lipids

In order to investigate the effect on secondary structure of AhlB and AhlC in the presence of lipid, liposomes were prepared from *E. coli* total lipid extract as described in 3.1.4, and incubated with each of the proteins in turn. Firstly, 0.1 mg/ml of AhlB was incubated with 10 µg *E. coli* liposomes in 200 µl PBS buffer pH 7.4 at 37 °C for 15 minutes. In this case, the degree of ellipticity at 222 nm and 208 nm does not vary greatly compared to AhlB in isolation (Figure 8.3). Interestingly, the high-tension voltage has increased in the sample containing AhlB and liposomes, even though each in isolation does not affect the high tension voltage (Figure 8.3). It has been reported that proteins bound to large fragments of membrane (>100 nm) can cause differential light scattering, which can decrease transmission, and thus increase the high-tension voltage . For example, in a study in which bacteriorhodopsin was isolated in purple membrane fragments, the non-random distribution of chromophores caused differential absorption flattening, which in turn caused a decrease in ellipticity of bacteriorhodopsin at 222 nm and 208 nm, and the wrongful assignment of secondary structure features (Figure 8.3). The chromophores in question are the peptide bonds for wavelengths between 190 nm and 240 nm. In the experiments of AhlB with liposomes, the increase in HT voltage within the range of the

A



B

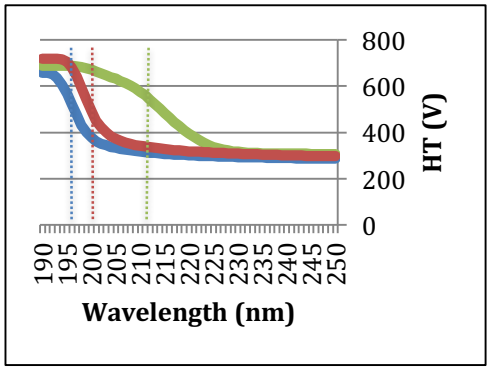


Figure 8.3. Circular Dichroism spectrum for AhlB, and for AhlB incubated with liposomes for 1 hour at 37 °C. **(A)** The circular dichroism spectrum from 190 nm to 250 nm for 1 mg/ml AhlB (blue line) and for 200 µg AhlB incubated with 10 µg *E. coli* liposomes for 1 hour at 37 °C (red line). Raw CD measurements were converted to mean residue ellipticity using the formula in 3.1.9. The dashed vertical lines indicate the cut-off for useful data as determined by the high-tension increasing above 550 V. Readings were measured three times and averaged for final values. (**B**) The high tension of the CD instrument during ellipticity measurements from 190 nm to 250 nm. The values are used here as a pseudo-transmission measurement, where a HT voltage of 550 V indicates the signal-to-noise ratio is high, and the mean residue ellipticity measurements are insignificant (Miles and Wallace, 2016).

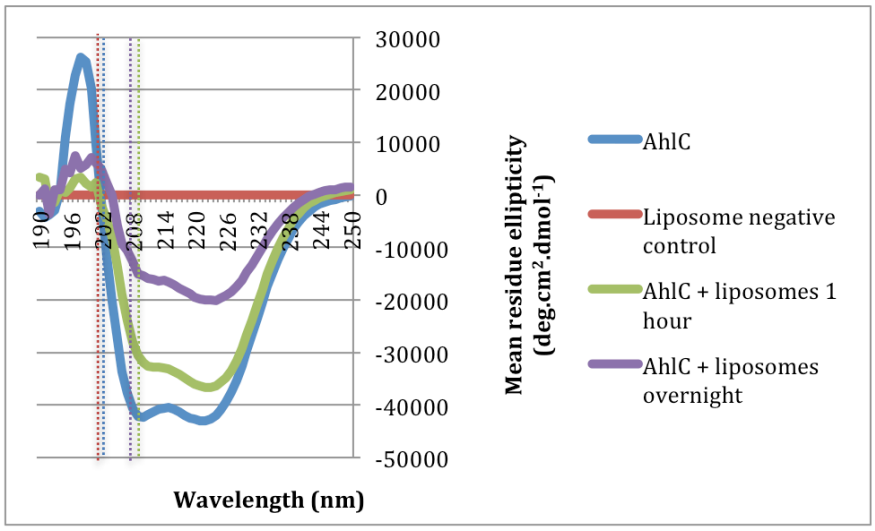
peptide bond at 190 – 240 nm could suggest that AhlB is concentrating on the liposome surface, and forming a non-random distribution of protein.

8.2.3 Circular dichroism of AhlC in the presence and absence of lipids

The same CD experiment was carried out on AhlC, first in isolation, to determine its secondary structure, and then in the presence of liposomes (Figure 8.4). The crystal structure of AhlC has been determined, and shows that the protein is ~100 % α-helical. Upon insertion of the protein into lipid, it is predicted that the usual conformational change in the head domain from β-strand to α-helix, seen for ClyA, would therefore not be required, and no change in ellipticity should be observed by CD. CD measurements were carried out on 200 µL of 0.1 mg/mL AhlC in isolation, identically to the AhlB measurements in 8.2.5. The degree of ellipticity had peaks at 208 nm and 222 nm of -42267.4 deg.cm^2.dmol^{-1} and -43016.4 deg.cm^2.dmol^{-1} respectively, in agreement with the completely alpha helical structure of AhlC observed in the crystal structure. The ellipticity at 222 nm is actually more negative than the predicted ellipticity for 100 % alpha helix content based on the equation in 3.1.9, and may be due to the increased helix packing in the tetramer assembly. The AhlC CD experiment was repeated in the presence of lipid, where 0.1 mg/ml AhlC was incubated with 10 µg *E. coli* total lipid extract liposomes at 37 °C for one hour. In the presence of lipids, the α-helical nature of AhlC was preserved, although the minima at 208 nm and 222 nm were less intense (Figure 8.4). Once again, the high-tension voltage increased to above 550 V at 205 nm, when AhlC was incubated with liposomes, suggesting possible absorbance flattening and secondary scattering.

As the high tension increases when liposomes are incubated with AhlC, the profile of the individual ellipticity plots cannot be compared. However, the overall degree of ellipticity is interesting, and the decrease in the overall degree of ellipticity at 222 nm is similar to that seen for AhlB. A model in which AhlC inserts its α-helical hairpin into a single leaflet of the bilayer, as proposed from the crystal structure of AhlC (Chapter 6) would result in an increase in differential light scattering and a consequent increase in HT. In turn, this would decrease the overall negative ellipticity, in agreement with the observed results. However, the large decrease in ellipticity of AhlC at 225 nm after being incubated with liposomes for 72 hours cannot be explained by this model, and is more likely due to protein aggregation or liposome degradation.

A



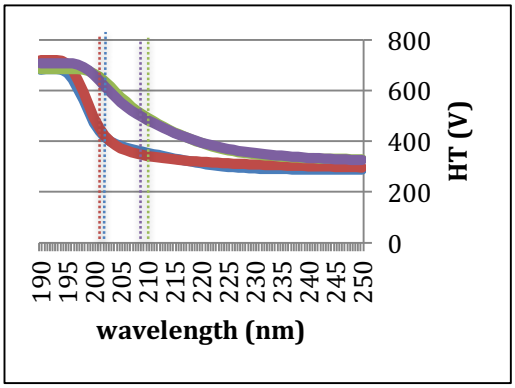


Figure 8.4. Circular Dichroism spectrum for AhlC, and for AhlC incubated with liposomes for 1 hour and overnight at 37 °C. **(A)** The circular dichroism spectrum from 190 nm to 250 nm for 1 mg/ml AhlC (blue line), 200 µg AhlC incubated with 10 µg *E. coli* liposomes for 1 hour at 37 °C (green line), and overnight (purple line). Raw CD measurements were converted to mean residue ellipticity using the formula in 3.1.9. The dashed vertical lines indicate the cut-off for useful data as determined by the high-tension increasing above 550 V. Each measurement was taken a single time, although the solution was thouroughly mixed by inversion before measurement. (**B**) The high tension of the CD instrument during ellipticity measurements from 190 nm to 250 nm. The values are used here as a pseudo-transmission measurement, where a HT voltage of 550 V indicates the signal-to-noise ratio is high, and the mean residue ellipticity measurements are insignificant (Miles and Wallace, 2016).

8.2.4 Control CD experiment with lysozyme and BSA incubated with liposomes

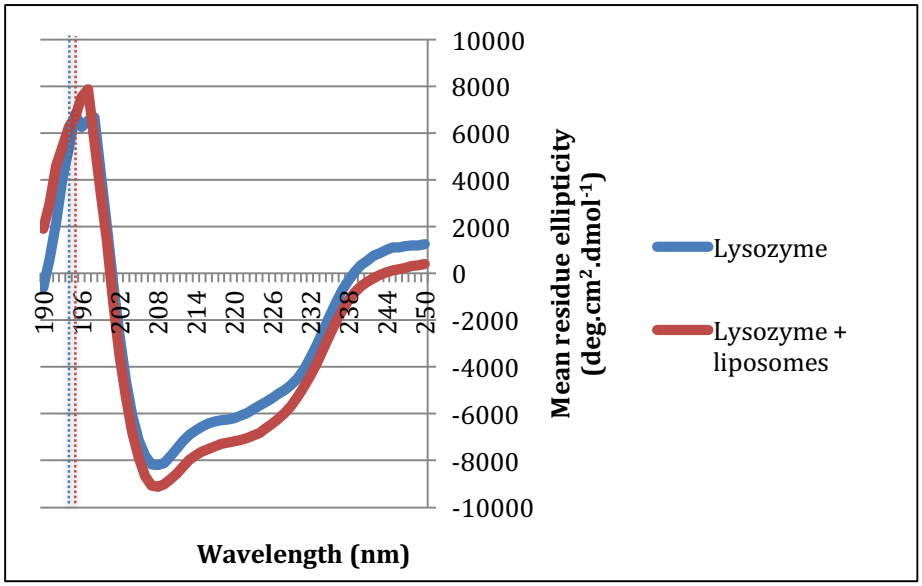
In order to test the theory that it is interaction of the AhlB and AhlC with liposomes, and the non-random distribution of protein inserted into the membrane that gives rise to the increase in high tension, control experiments were carried out with lysozyme incubated with liposomes. The above experiments were repeated, incubating 0.1 mg/ml Hen egg white lysozyme with 10 µg *E. coli* liposomes for one hour at 37 °C. The CD spectrum from 190 to 250 nm was measured for 0.1 mg/ml lysozyme in isolation, and then after one hour incubation with liposomes (Figure 8.5). The spectra show that the overall ellipticity of the protein remains unchanged, and that after incubation, the high tension does not increase (Figure 8.5). This result shows that it is not the incubation of protein with liposome at 37 °C that changes the dispersion of the solution, but a specific phenomenon occurring with AhlB or AhlC and liposomes.

Recently, the phenomenon of differential light scattering and absorption flattening has been documented thoroughly (Miles and Wallace, 2016), explaining the phenomenon and how to minimise such problems. Although the problem of differential light scattering can be reduced by using smaller liposomes (<25 nm), the problems of differential absorption flattening would persist and would possibly be compounded by the smaller liposome size. The authors do however note that data above 200 nm are still useful, and qualitative conclusions can be drawn, as have been here.

8.3 Probing AHL oligomerisation by gel filtration

ClyA-like PFTs have been shown to form oligomeric assemblies and pores when incubated with detergents (Mueller et al., 2009). Further, it has been reported that components of the NHE and HBL tripartite systems can form oligomeric complexes in the presence of detergent (Madegowda et al., 2008; Phung et al., 2012). To determine if detergents are able to induce oligomerisation and pore formation of AHL components, AhlB and AhlC were incubated with detergent, alone and in combination, and then subjected to gel filtration. Previously, it had been shown that AhlB and AhlC do not form any complex in solution when incubated together for 1 hour at room temperature (Chapter 5). Using this result as a negative control, the effect of detergents on complex formation and protein elution volume was measured. The AHL proteins AhlB or AhlC were incubated at 0.1 mg/mL, 3 mg and 5 mg total protein respectively, with 0.1 %

A



B

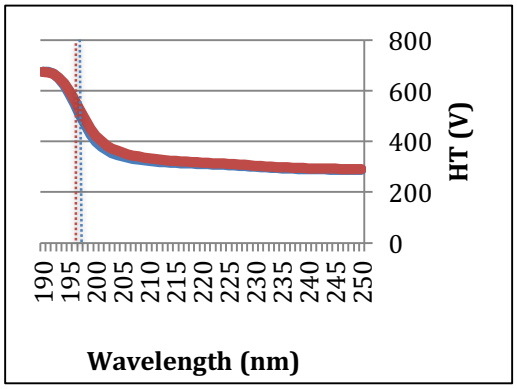


Figure 8.5. Circular Dichroism spectrum for Lysozyme and lysozyme after incubating with 10 µg *E. coli* total lipid extract liposomes for 1 hour at 37 °C. (**A**) The circular dichroism spectrum from 190 nm to 250 nm for 200 µg Lysozyme (blue line), and 200 µg Lysozyme incubated with 10 µg *E. coli* liposomes for 1 hour at 37 °C (blue line). Raw CD measurements were converted to mean residue ellipticity using the formula in 3.1.9. The dashed vertical lines indicate the cut-off for useful data as determined by the high-tension increasing above 550 V. Each measurement was taken eight times, and averaged to give each final value. (**B**) The high tension of the CD instrument during ellipticity measurements from 190 nm to 250 nm. The values are used here as a pseudo-transmission measurement, where a HT voltage of 550 V indicates the signal-to-noise ratio is high, and the mean residue ellipticity measurements are insignificant (Miles and Wallace, 2016).

(w/v) DDM. After incubation, the protein detergent sample was concentrated and gel filtration was carried out, loading 0.1 mL protein sample onto a superpose 6 10/300 GL or Superdex 200 10 300 GL gel filtration column respectively. Peak fractions were collected after the void volume, and analysed by SDS page electrophoresis and electron microscopy.

8.3.1 Gel filtration of AhlB in the presence of DDM

Gel filtration was carried out on AhlB both in the absence and presence of 0.1 % (w/v) DDM. In the absence of detergent, AhlB elutes from the gel filtration column at ~17 ml corresponding to ~ 35 kDa. There is no peak at the void volume in the absence of DDM. However, in the presence of 0.1 % (w/v) DDM, after incubation for 4 hours at 37 °C, AhlB forms oligomeric assemblies (Figure 8.6). A sharp peak at the void volume of 7.8 ml was observed in gel filtration for AhlB and DDM, which was not present when AhlB was subjected to gel filtration in the absence of detergent (Figure 8.6). A small peak was also present at 14.4 ml, corresponding to a molecular weight of ~330 kDa. Electron microscopy was used to examine the peak at 14.4 ml, in order to observe any protein assemblies formed. Electron microscopy carbon grids were generated as described in 3.7.2, and 5 µl of the gel filtration void volume fraction was adsorbed onto the grid and stained with 1 % uranyl formate. The DDM AhlB grids were observed by electron microscopy, and showed that large aggregates had indeed formed, with some areas of the images indicating that single pores may have formed, but these were few in number (Figure 8.7), and the majority of the protein on the grid was either aggregated or potentially off-pathway assemblies (Benke et al., 2015).

8.3.2 Gel filtration of AhlC in the presence of DDM

In the presence of DDM, AhlC elutes from the gel filtration column at the same volume of 15.2 ml, and same apparent molecular weight of ~80 kDa, consistent with the tetrameric form of AhlC in the absence of DDM, and so forms no higher molecular weight assemblies. This was confirmed by electron microscopy of AhlC in the presence and absence of DDM, where both samples appear identically monodisperse in the EM images. This result is in contrast to the Hbl B protein (the C component in the HBL system), which forms ~10-mer assemblies in the presence of DDM (Madegowda et al., 2008), and ClyA, which forms 12-mer pores in the presence of DDM (Mueller et al., 2009). The elution volume of AhlC at 15.2 ml in isolation

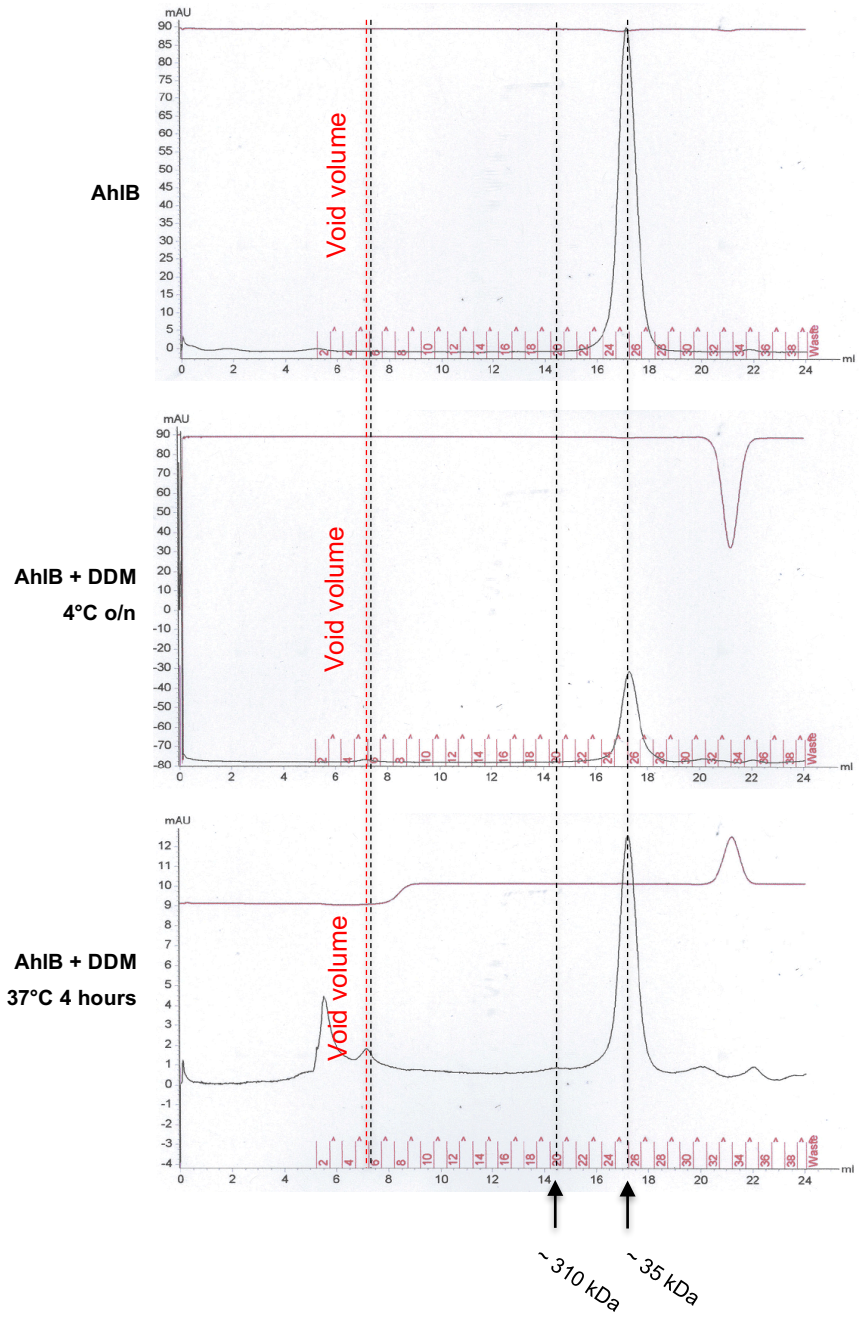


Figure 8.6. Gel filtration analysis of AhlB incubated with 0.1 (w/v) DDM overnight at 4 °C or for 4 hours at 37 °C. The gel filtration chromatograms for AhlB run on a Superose 6 10 300 GL gel filtration column are shown in the absence of detergent, after an overnight incubation with DDM, stored at 4 °C, and after 4 hours incubation at 37 °C. The black trace shows the absorbance at 280 nm of the eluate at each volume. The predicted dimer elution is highlighted in each chromatogram at 17.5 ml by a dashed line, and shows that dimeric protein is eluted from the column in each condition. The void volume is shown on each chromatogram as the red dashed line, and the black dashed line highlight the peak in UV at 280 nm at the void volume when AhlB is incubated at 37 °C. The third dashed line indicates the elution of protein at 14.4 ml when AhlB is incubated at 37 °C, corresponding to a molecular weight of ~330 kDa.

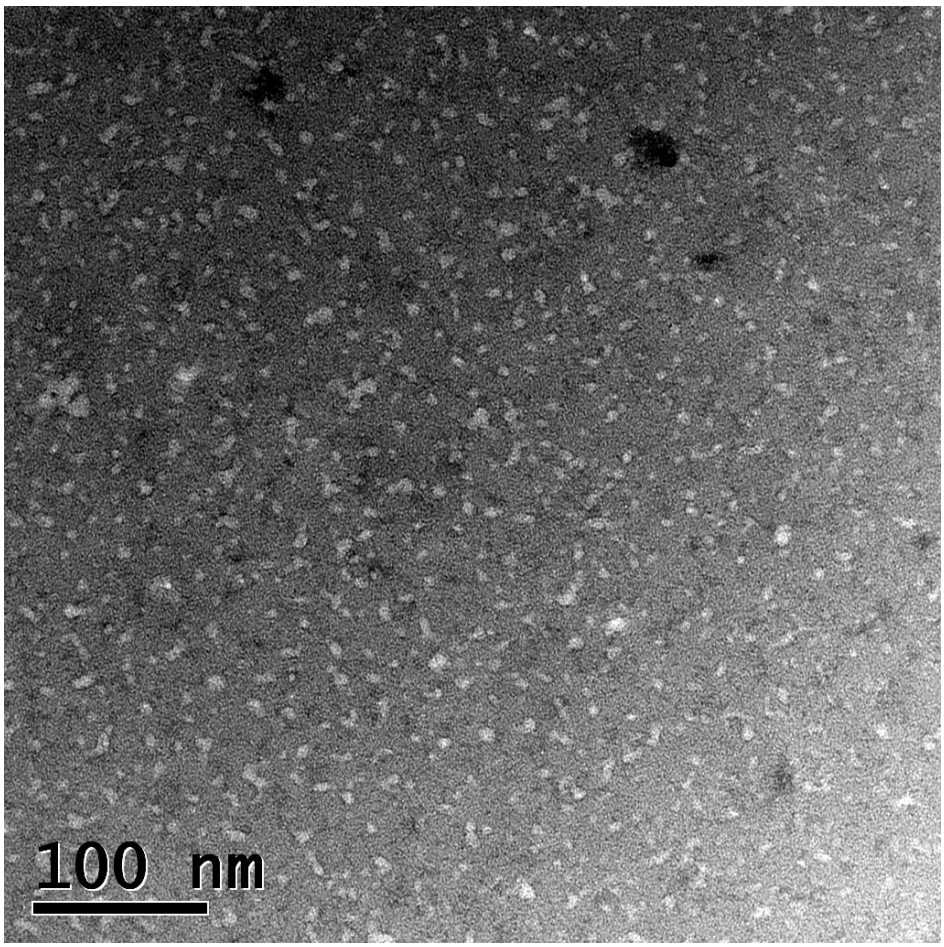


Figure 8.7. Electron micrograph of 14.4 ml elution from gel filtration of AhlB incubated with DDM. Electron micrograph of negatively stained AhlB after incubation with 0.1 % (w/v) DDM and gel filtration. Sample was formed from the elution fraction at 14 ml, corresponding to ~330 kDa. Large aggregates have formed, of ~ 10 nm, which may be AhlB pores, although no obvious pore openings are seen.

suggests the protein is a tetrameric assembly, which is also seen in the crystal structures. As the elution volume of AhlC incubated with DDM is identical to that observed for AhlC alone, it appears that DDM is unable to dissociate the tetramer of AhlC, indicating that either DDM does not interact with AhlC at all, or that more substantial interactions are required to disassemble the tetramer of AhlC. It has been reported in an increasing number of cases that PFT oligomerisation can favour a specific lipid type (Dal Peraro and van der Goot, 2016), and so it may be that in the case for AhlC, a specific lipid protein interaction facilitates the dissociation of the AhlC tetramer, prior to membrane insertion.

8.3.3 Gel filtration of AhlB and AhlC in combination, in the presence of deoxycholate

Any assembly of AHL components may require a complex formation between AhlB and AhlC. Gel filtrations showed that neither forms a complex in solution (Chapter 5), but the situation may be different in the presence of detergent. In order to determine if a complex can form between AhlB and AhlC in the presence of detergent, AhlB and AhlC were incubated at 37 °C for one hour with 10 mM deoxycholate, and then concentrated and applied to a gel filtration column equilibrated with 50 mM Tris pH 8, 0.1 M NaCl, and 5 mM deoxycholate. Incubation of AhlB and AhlC with deoxycholate resulted in the same elution profile of the proteins on gel filtration without detergent. A second experiment was carried out, in which AhlB and AhlC were incubated with 10 mM DOC overnight at room temperature, and then subjected to gel filtration on a Superdex 200 pg column. After an overnight incubation, a small peak not present in the detergent-free sample was present at 60 ml, corresponding to a molecular weight of ~500 kDa (Figure 8.8). The peak at the void volume and the peak at 60 ml were analysed by electron microscopy, adsorbing 5 µl of each to a carbon grid. Electron micrographs show that large aggregates are formed in the void volume sample, and that neither contain obvious pore-like structures (Figure 8.9)

8.3.4 Detergent experiments conclusions

No conditions were found for either AhlB or AhlC for which stable pores could be formed using the detergents used in previous studies on ClyA-like PFTs. A more substantial screen of many different detergents and incubation conditions may yield conditions that successfully form

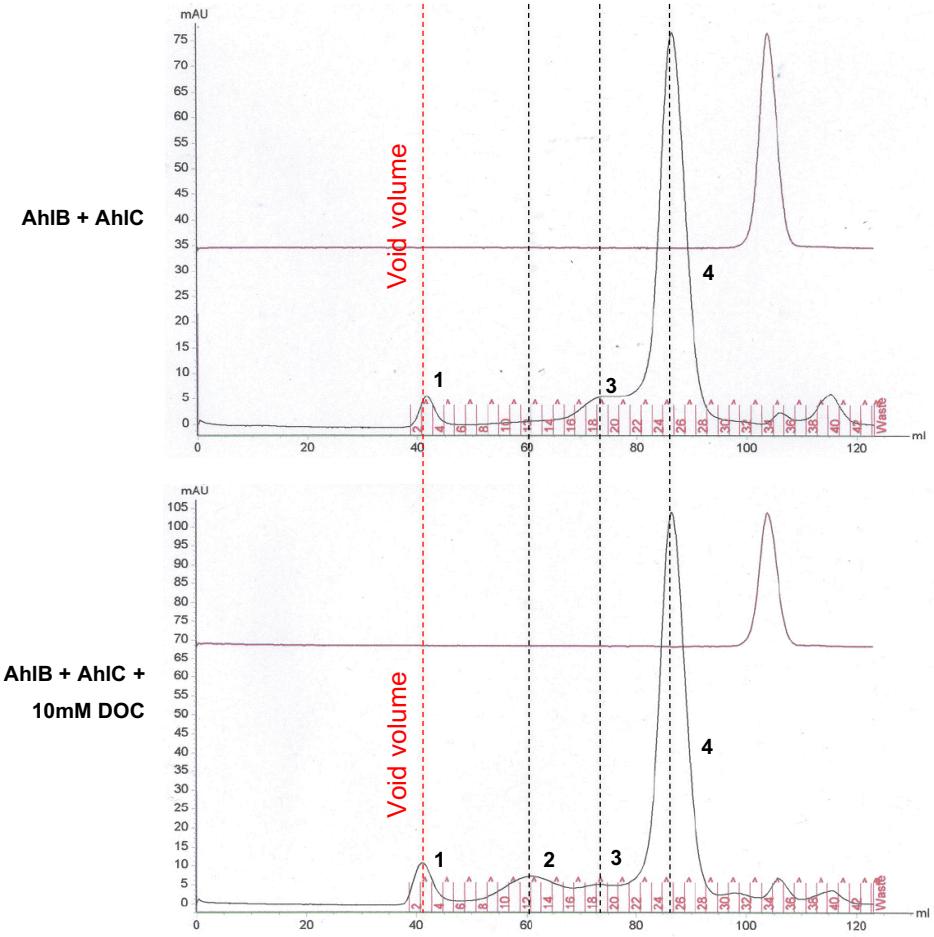


Figure 8.8. Gel filtration chromatogram of AhlB and AhlC incubated together for 1 hour at 37 °C with 0.1 M Deoxycholate. Gel filtration chromatograms for AhlB and AhlB incubated with 10 mM DOC for 1 hour at 37 °C. The elution of AhlB and AhlC are shown by dashed lines at ~87 ml (peak 4) and ~75 ml (peak 3) respectively, as is the peak that appears at 60 ml when AhlB and AhlC are incubated with DOC (peak 2). The void volume is highlighted by the red dashed line, and a small peak is present at the void volume in both the pre- and post- incubation chromatographies (peak 1).

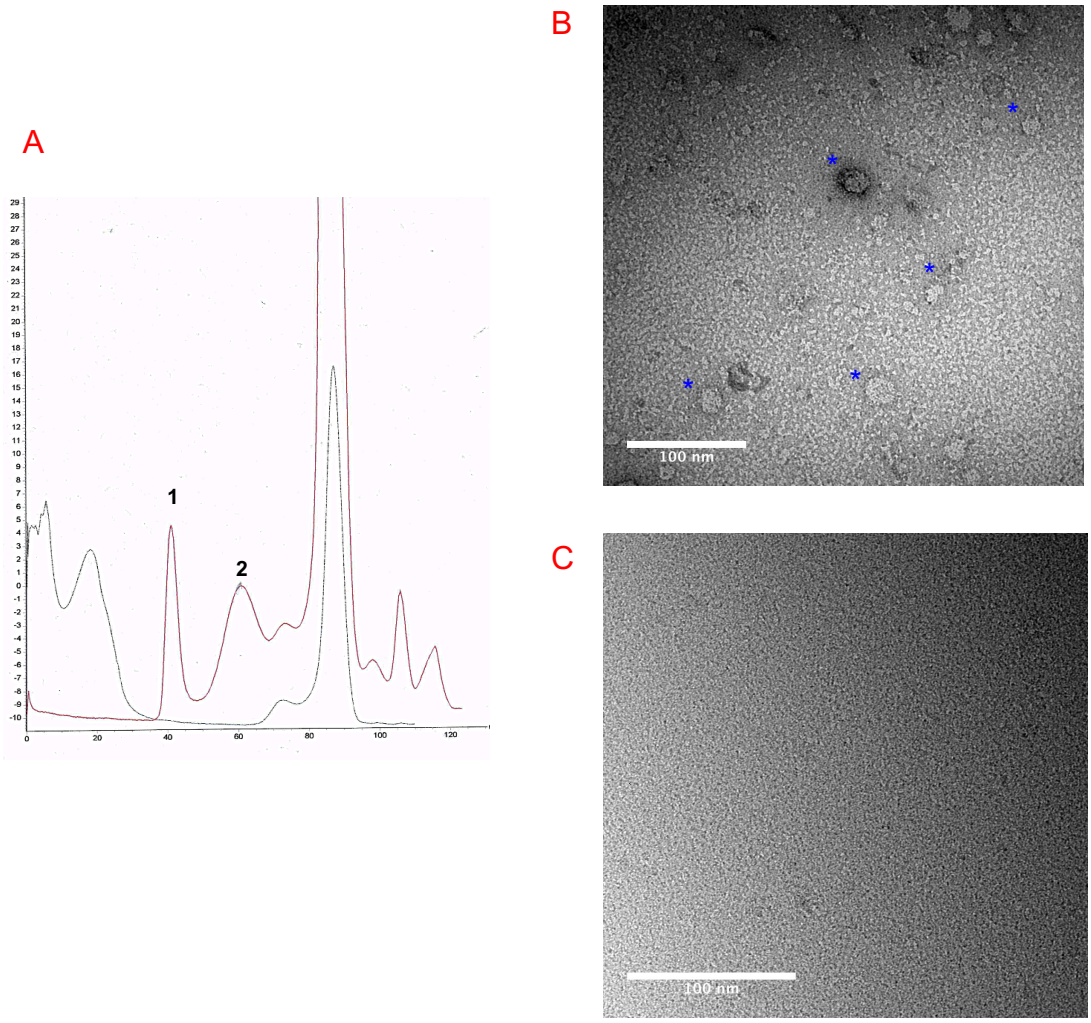


Figure 8.9. Electron micrographs of peak 1 and 2 from the gel filtration of AhlB + AhlC + 10 mM DOC. (**A**) Overlaid gel filtration chromatogram of AhlB and AhlC in isolation (black trace), and after incubation for 1 hour at 37 °C (red trace). Peak 1 is present at the void volume of 42 ml, and peak 2 is present at 60 ml. (**B**) Electron micrograph of a negatively stained sample from peak 1. A 5 µl sample of peak 1 was applied to a glow discharged grid and stained with 1 % uranyl formate. Large aggregates (20-40 nm) are highlighted by blue stars, and may contain some circular morphologies. Other small aggregates (<10 nm) are also observed across the image. (**C**) Electron micrograph of a negatively stained sample from peak 2. A 5 µl sample was applied identically to (B), and stained with 1 % uranyl formate. The grid contained a fairly uniform dispersion of protein, and possible circular morphologies present.

homogeneous pores, as has been the case for a number of other pore forming toxin systems (Saibel and Bubeck, Royal Society Membrane Pores conference 2016).

8.4 Electron microscopy experiments of AHL with liposomes

In order to probe the pore assemblies and pore assembly requirements in a more natural environment than detergent, electron microscopy experiments were carried out with the AHL components in the presence of liposomes.

8.4.1 Lipid composition of the total lipid extracts used in electron microscopy experiments

The liposomes used in all electron microscopy experiments were composed of total lipid extracts. A number of different total lipid extracts were tested, as a means of screening potential lipid compositions that would be complimentary to AHL pore formation. The composition of each total lipid extract is shown in Table 3.3. Two total lipid extracts, brain and liver, were derived from eukaryotic cell membranes, which contain more zwitterionic lipids, although no stable liposomes were produced from brain lipids. A third total lipid extract, from *E. coli*, contained more largely negatively charged lipids.

Electron micrographs of liver liposomes and *E. coli* liposomes, generated as described in 3.7.2, show that in each case, clear individual liposomes can be observed. These liposomes form an enclosed sphere of lipid, which then collapses on the carbon grid under the negative stain, to form collapsed sacs (Figure 8.10). This means that in the majority of the views of the liposomes in the micrographs, two bilayers are being observed, although around the edges of the liposomes, there is a segment where a single bilayer is presented (Figure 8.10).

8.4.2 Electron microscopy studies of AHL proteins incubated individually with liposomes

Each AHL component was incubated at 10 µg total protein in isolation with 20 µg liposomes, in order to provide control images that could be used to differentiate between the species formed each protein was incubated with liposomes, and when multiple AHL components were incubated with liposomes.

No change in the appearance of the liposomes was observed when AhlA was added in isolation. The micrographs of liposomes incubated with AhlA contain liposomes that look very similar to the liposomes observed in the preparations without protein (Figure 8.11). However, in these micrographs, the area of carbon film between the liposomes contains a high noise

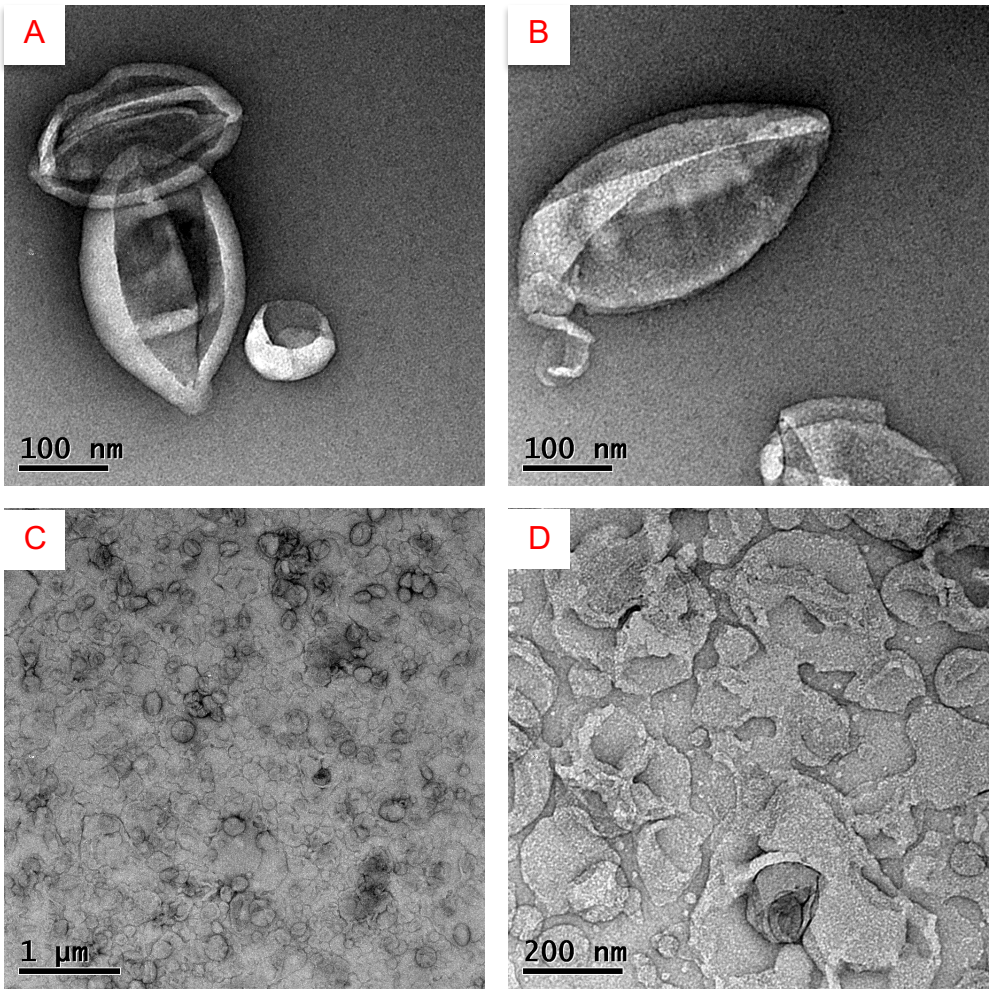


Figure 8.10. Electron micrographs of *E. coli* and liver total lipid extract liposomes. (**A and B**) Negative stain transmission electron micrographs for 10 µg *E. coli* total lipid extract liposomes. Liposomes form collapsed sacs where multiple layers of the liposome are visible stacked on top of one another. Generally the liposomes are well contrasted against the carbon film, and have smooth surfaces and edges. (C and **D**) Negative stain transmission electron micrographs for 10 µg liver total lipid extract liposomes. Liposomes form turgid liposomes without collapsing, and in some cases (**C**) seem to form extended networks of liposomes, partially fused.

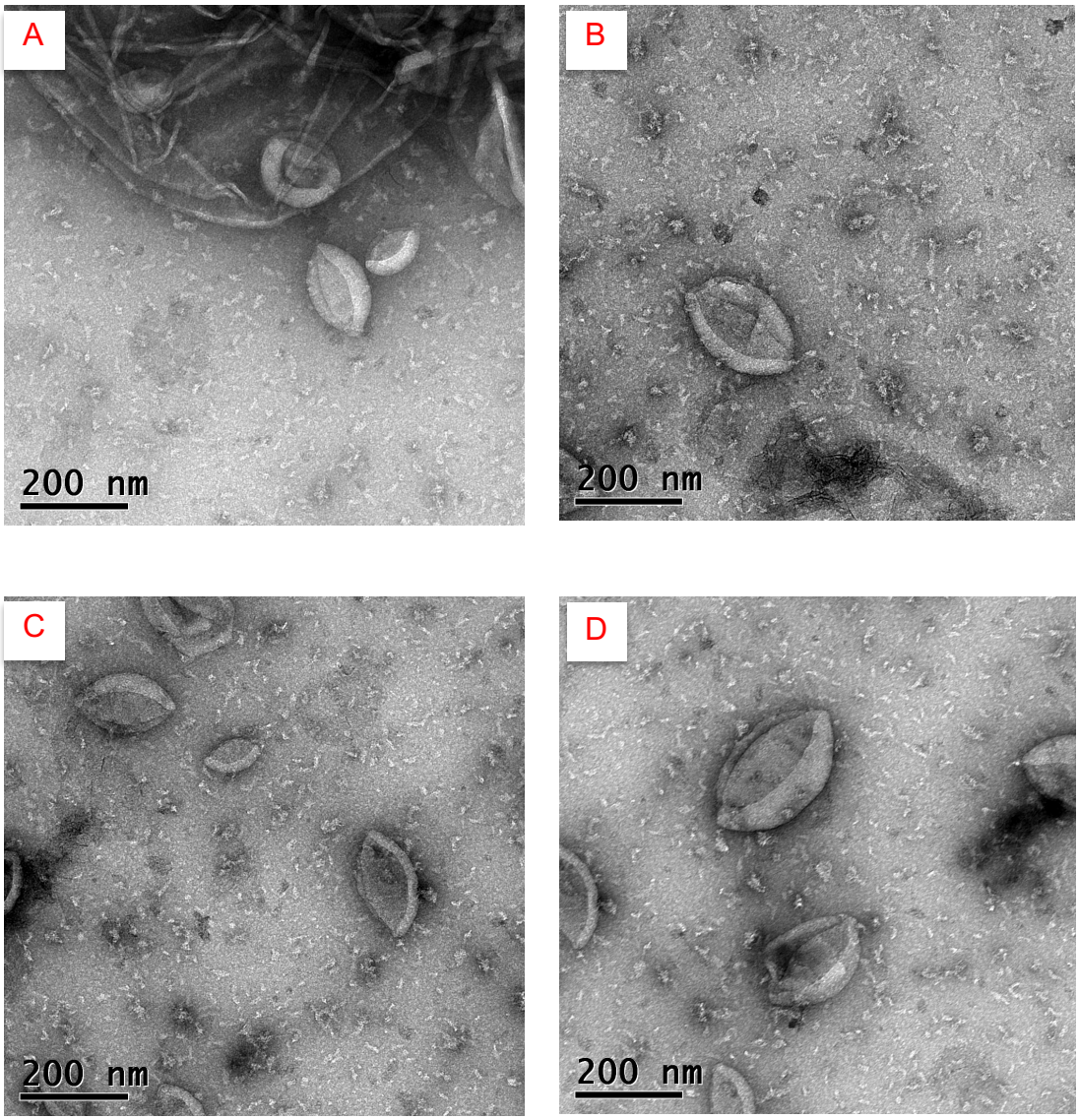


Figure 8.11. Electron micrographs of *E. coli* total lipid extract liposomes incubated with AhlA protein. (**A**-**D**) Electron micrographs showing *E. coli* liposomes incubated for with 10 µg AhlA 1 hour at 37 °C, and then applied to carbon grids. Each liposome still has a smooth surface and edges, and no protein seems to have been able to associate with the membranes of the liposomes. In each micrograph the background carbon film has many small aggregates dispersed, presumably protein aggregates.

background of small aggregates, not present in micrographs of the liposomes in isolation. These aggregates are ~40 nm in size and are presumably formed from AhlA protein. Similar results were observed in electron micrographs obtained when NheA was incubated with liposomes (Ganesh, 2012). In these micrographs the NheA protein also formed aggregates, and did not associate with the liposome membrane.

In contrast, in similar membrane preparations where 10 µg AhlB was incubated with 20 µg liposomes, the protein can clearly be seen to interact with the membrane and form assemblies. These micrographs clearly show that AhlB is able to form a small number of pores in the surface of the liposomes, each with a diameter of ~10 nm (Figure 8.12). To further investigate the AhlB assemblies, a total of 67 pores were imaged. The profile of the pores was calculated using imageJ plot profile (Figure 8.13), and these profiles were used to estimate an average pore size. The pore sizes of 67 pores are plotted in (Figure 8.13), and a Gaussian curve was fitted to determine an average pore size. Based on this, the average pore diameter was estimated at 12.0 nm with a standard deviation of 1.8 nm. In each image used, the pixel resolution was 0.72 nm per pixel. This estimation of pore size is within the limits of the pore predicted in crystal forms 3 and 4 of AhlB in crystallographic studies, with a maximum allowed diameter of ~11.5 nm.

A strange behaviour of the liposomes was observed in the presence of AhlC, with the liposomes becoming fragmented, and laying flat along the carbon grid (Figure 8.14). However, despite the morphological change, no ordered assemblies of AhlC could be seen in the membrane fragments. Similar liposome fragment morphologies were also observed when NheA was pre-incubated with 1 % (w/v) DDM, and then mixed with liposomes, to produce a protein/detergent/lipid mix. For the NheA experiment, the DDM detergent solubilises the liposomes, forming the fragments, whereas AhlC causes this phenomenon in the structure of the liposomes in the absence of detergent. It is thus clear that AhlC does indeed interact with the lipid bilayer of the liposomes, as suggested by the CD results in the presence of liposomes, but that AhlC does not form any pore structures in isolation.

The observation that AhlA does not interact with the liposomes in isolation has been predicted previously for NheA, and agrees with the results of the haemolysis assays (Chapter 5). However, the interaction of AhlB and AhlC with the membrane is unexpected, based on the

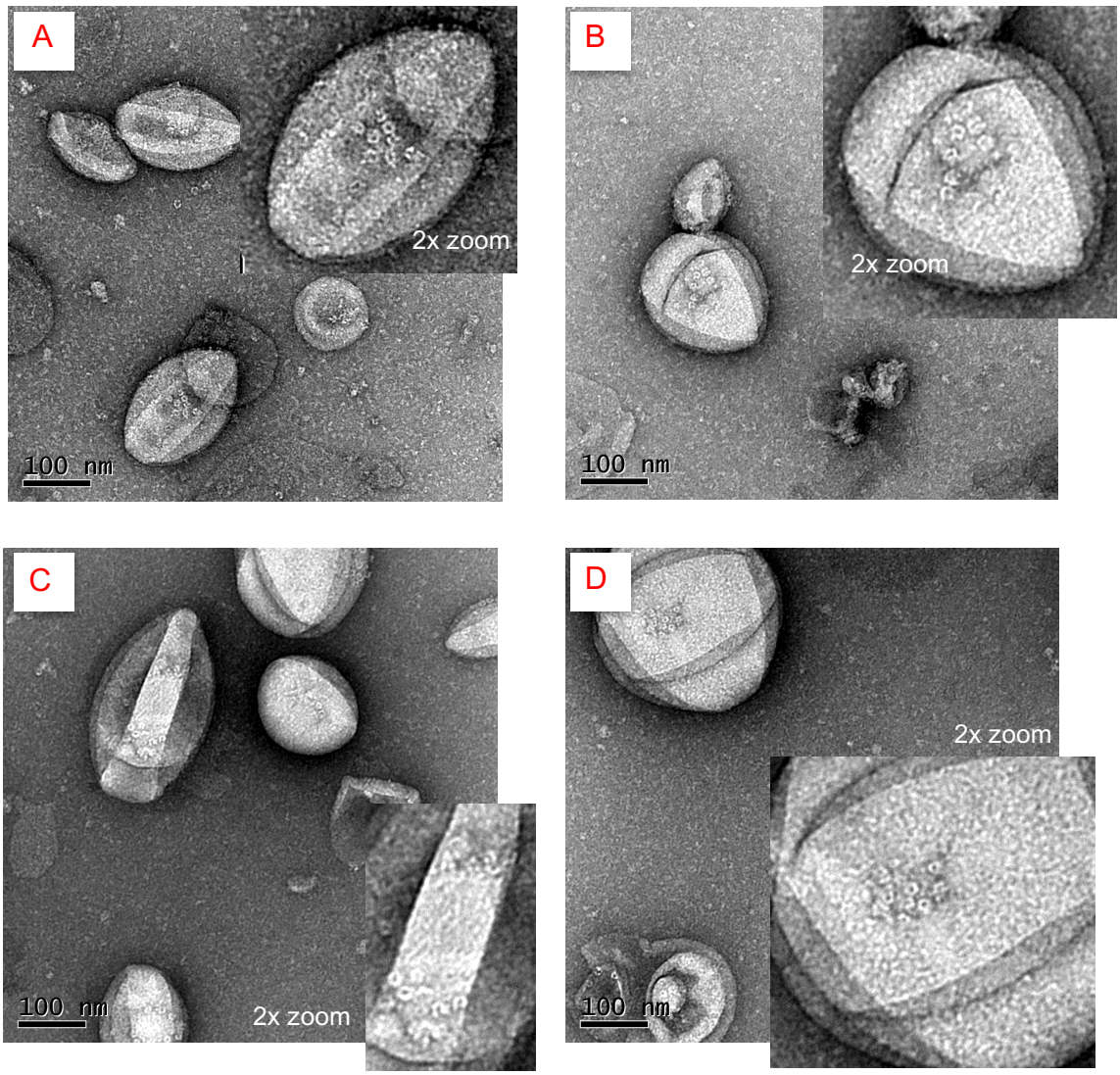


Figure 8.12. Electron micrographs of *E. coli* total lipid extract liposomes incubated with AhlB protein. (**A**-**D**) Electron micrographs showing *E. coli* liposomes incubated with 10 µg AhlB for 1 hour at 37 °C, and then applied to carbon grids. Each liposome contains a small number of circular highly contrasted assemblies on the membrane surface. Each liposome is also magnified by 2 to highlight the assemblies.

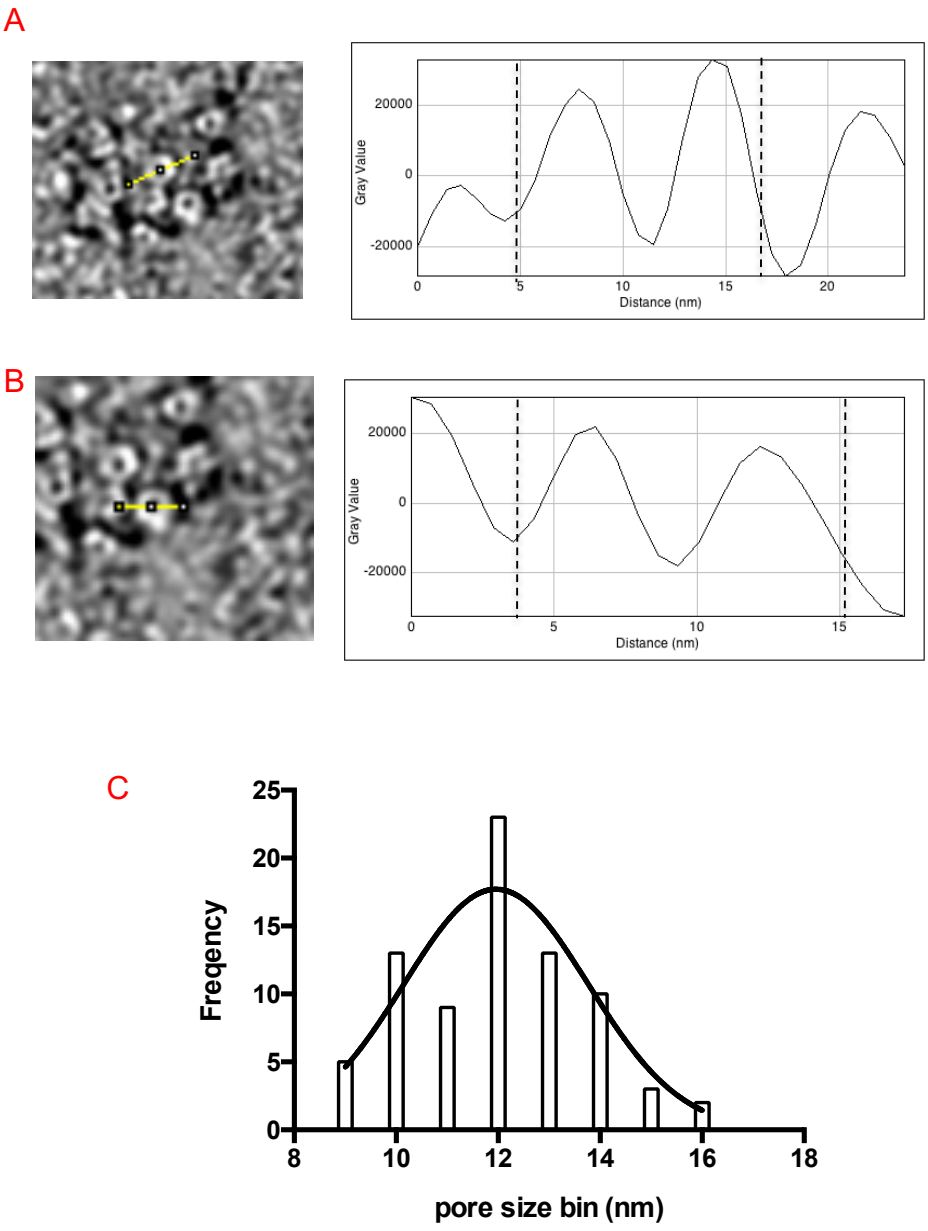


Figure 8.13. Estimation of the size of AhlB pores. (**A**) and (**B**) Electron micrograph of AhlB pores formed in *E. coli* liposomes (left), with a yellow line showing the profile plot of pixel intensity. The size of the pore was estimated in the profile plots (right) by taking the diameter as where the intensity dropped below that of the center of the pore (shown in dotted lines). (**C**) Frequency distribution of pore sizes, binned into 1 nm bins.

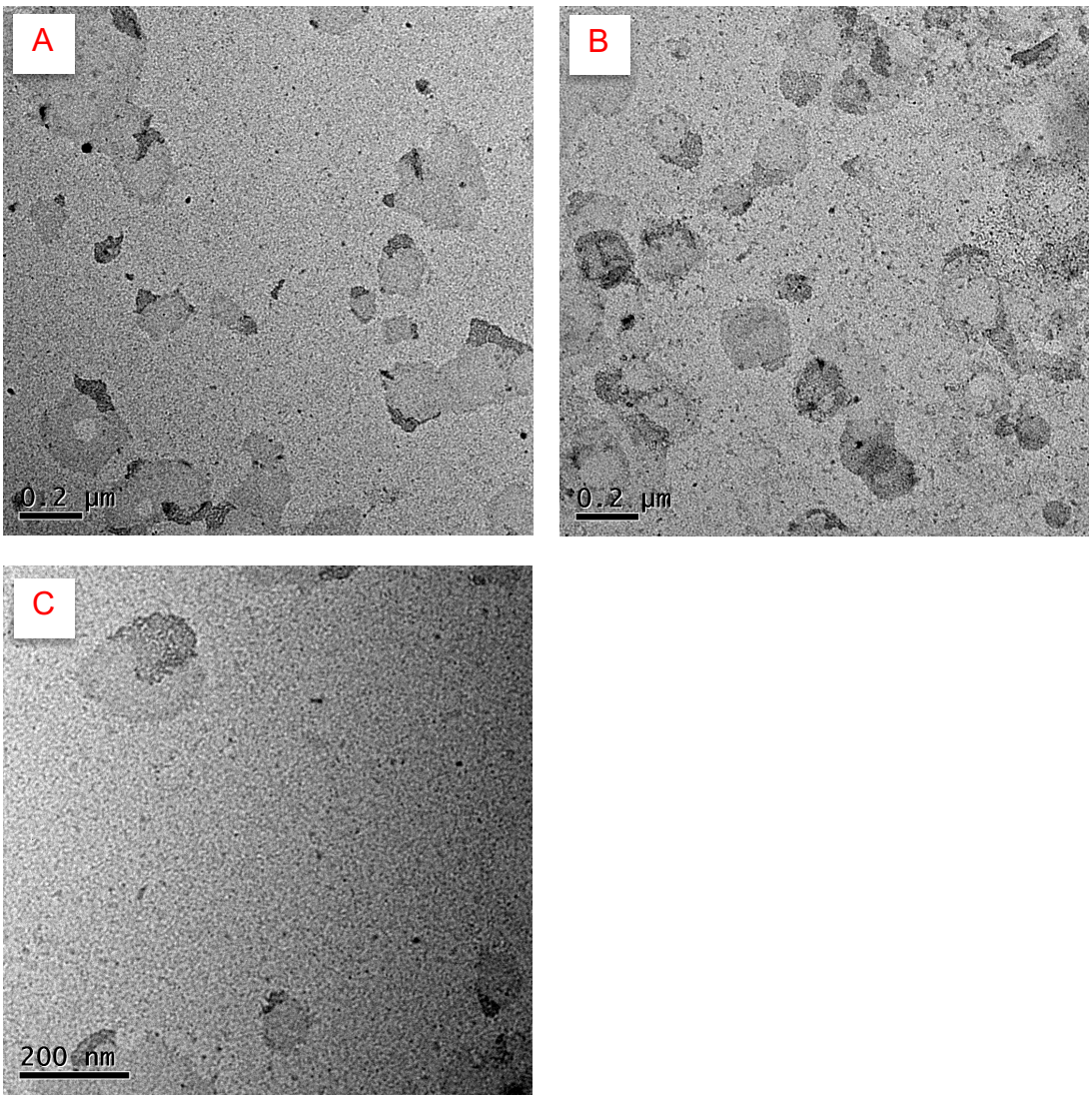


Figure 8.14. Electron micrographs of *E. coli* total lipid extract liposomes incubated with AhlC protein. (**A**-**B**) Electron micrograph showing *E. coli* liposomes incubated with 10 µg AhlC for 1 hour at 37 °C, and then applied to a carbon grid. No liposomes are present on the grid, and instead, flat sheet structures have formed. No obvious protein assemblies are present in the sheets. (**C**) Electron micrograph at higher magnification (scale bar shown), showing flat possibly lipid sheets, and no discernable ordered features present on the sheets.

lack of hemolytic activity of each protein in isolation. This does, however, agree with the NHE data (Lindbäck et al., 2010), where NheB and NheC can associate with the membrane even while lacking any lytic activity in isolation. Based on the results of the haemolysis assays, the pores formed by AhlB must therefore not be active, and may constitute a pre-pore, which is activated by some means slightly by AhlC, and fully by AhlA and AhlC. The observations that AhlC seems to interact with the membrane but does not form pores are somewhat expected, due to the structure of AhlC containing a ready-to-insert head domain, and the lack of lysis when AhlC is incubated in isolation in haemolysis assays. The observation of the deformation of the liposomes by AhlC is however unexpected, and may be a consequence of AhlC binding to the liposome and be important for the function of AhlC.

8.4.3 Electron microscopy studies on AHL components incubated in combinations with liposomes

Different AHL components were also incubated with liposomes in combination, based on those combinations that gave rise to haemolysis in the haemolytic assays described in 5.5.5, namely AhlB and AhlC together, and AhlA, AhlB and AhlC together.

As the haemolysis assays showed quite different amounts of lysis when mixtures of the AHL components were incubated with red blood cell, a series of electron microscopy samples were prepared both using different combinations of the protein components, and also using liposomes prepared from different total lipid extracts. For these experiments, the two mixtures that gave rise to lysis in the haemolysis experiments, namely AhlB and AhlC, and AhlA, AhlB and AhlC were used.

When 20 µg liposomes prepared from *E. coli* lipids were incubated with both AhlB and AhlC at 10 µg each, a different pattern of protein interaction with the bilayer was seen, compared to either component alone (Figure 8.15). Two different features were observed, in one case individual liposomes appeared to be totally saturated with protein (red circle), with individual pore-like structures clearly present, but due to the amount of protein, these structures were difficult to resolve. For other liposomes on the grid no protein appears to have bound, or rather no pore structures have formed (blue circle). This binary difference suggests that some lipid component, or mechanical property promotes, or inhibits, AhlB and AhlC binding. In addition, the saturation of the liposome with protein pores does not occur when the liposomes are

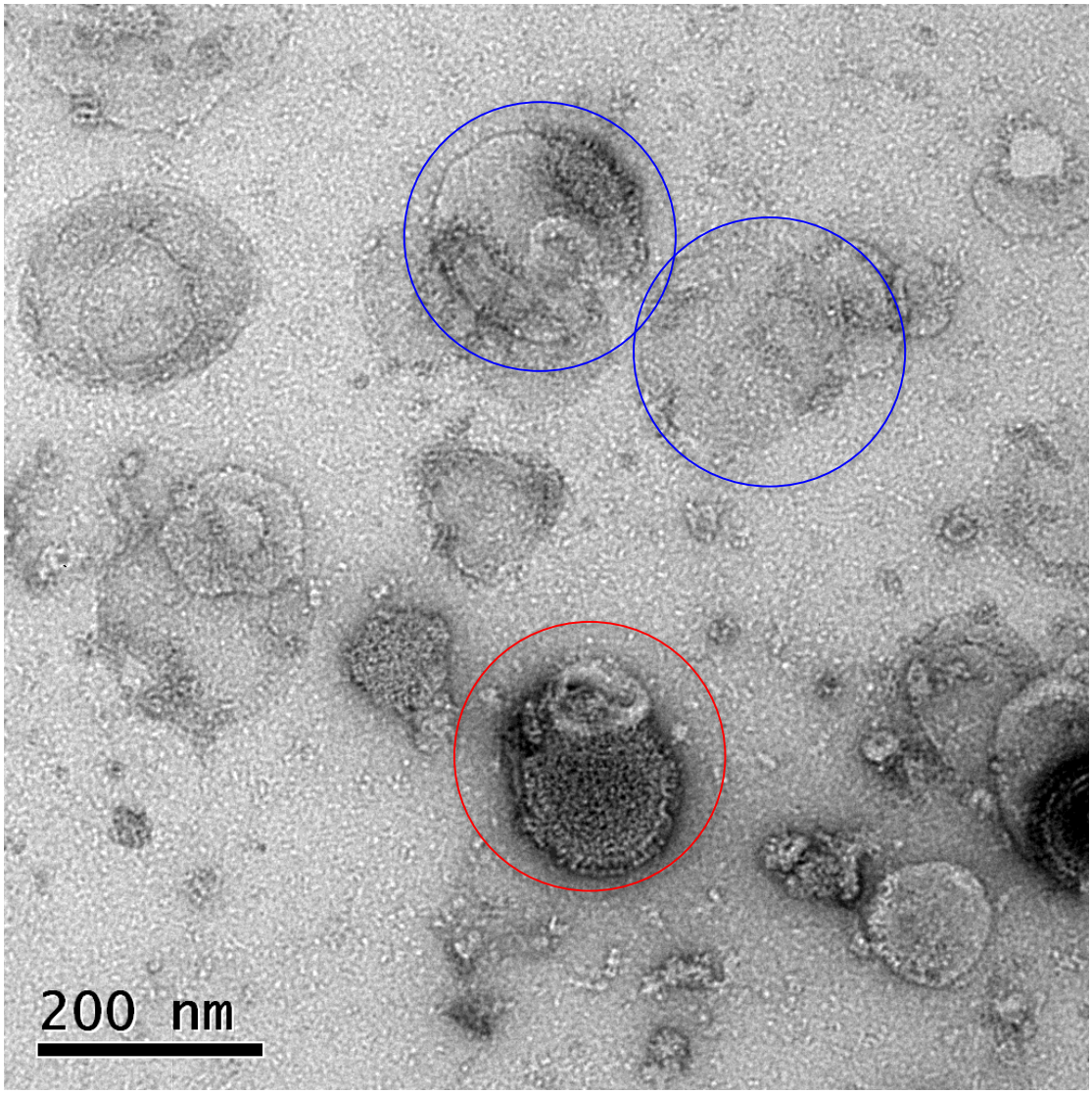


Figure 8.15. Electron micrographs of *E. coli* liposomes saturated with AhlB and AhlC. Electron micrographs of negatively stained *E. coli* liposomes incubated with AhlB and AhlC for 1 hour at 37 °C. Certain liposomes are completely saturated in protein, with some pore-like structures (red circle), with a high contrast between liposome and carbon film. Other liposomes contain no obvious pore-like structures, and are not highly contrasted against the carbon film (blue circles).

incubated with AhlB alone, suggesting that the interaction of AhlC has a concentrating effect on pore formation. In some micrographs, small fragments of lipid are liberated from the liposome, and side views of pores can be seen (red circle)(Figure 8.16). These species were also observed for the NHE protein incubated with liposomes when detergent was added (Magdah thesis, 2012). It was suggested that these 'Viking ships' were side-views of pores (Figure 8.17), which agrees with the observation of AhlBC in *E. coli* liposomes where side-views are also present on liposome surfaces, and are of the same morphology and size.

In some liposomes, single pores can be resolved in a top-down view (Figure 8.18), and are of a similar size to those of AhlB on their own. However, many pores also form arc structures in the liposomes (Figure 8.19), which have been observed for some CDC PFTs, and are suggested to be intermediates in pore assembly.

A different situation arose when 10 µg of each of AhlB and AhlC were incubated together with 20 µg liposomes prepared from liver lipids. In this case, and in contrast to that seen for *E. coli* lipid liposomes, liposomes were saturated with pores, which appeared to be in spiralling layers (Figure 8.20). Blebbing of the liposomes can also be seen (Figure 8.20), where the pores induce tight curvature in the membrane. This again suggests that a certain lipid component or mechanical property of the liposome affects the ability of pores to form in the membrane. This is a property well explored for other PFTs such as CDC pore forming toxins (Rojko and Anderluh, 2015), but has not been reported for ClyA-like PFTs.

When 20 µg liposomes prepared from *E. coli* lipids were incubated with 10 µg of each of AhlA, AhlB, and AhlC, saturation of liposomes still occurred (Figure 8.21), and the overall size of the small number of well-separated pores looked to be the same size as those formed from AhlB and AhlC (Figure 8.22). This observation suggests that AhlA does not interchelate between molecules of AhlB, extending the diameter of the pore. Instead, either AhlA may be forming another ring of molecules around the already formed AhlBC pores, or only a small number of AhlA molecules are used in pore formation. Based on the 10:10:1 stoichiometry observed for NHE, multiple rings would seem the more probable, if AHL functions similarly. However, due to the resolution of the images, these questions cannot be answered, and increased resolution by cryo-EM techniques and single particle averaging are required.

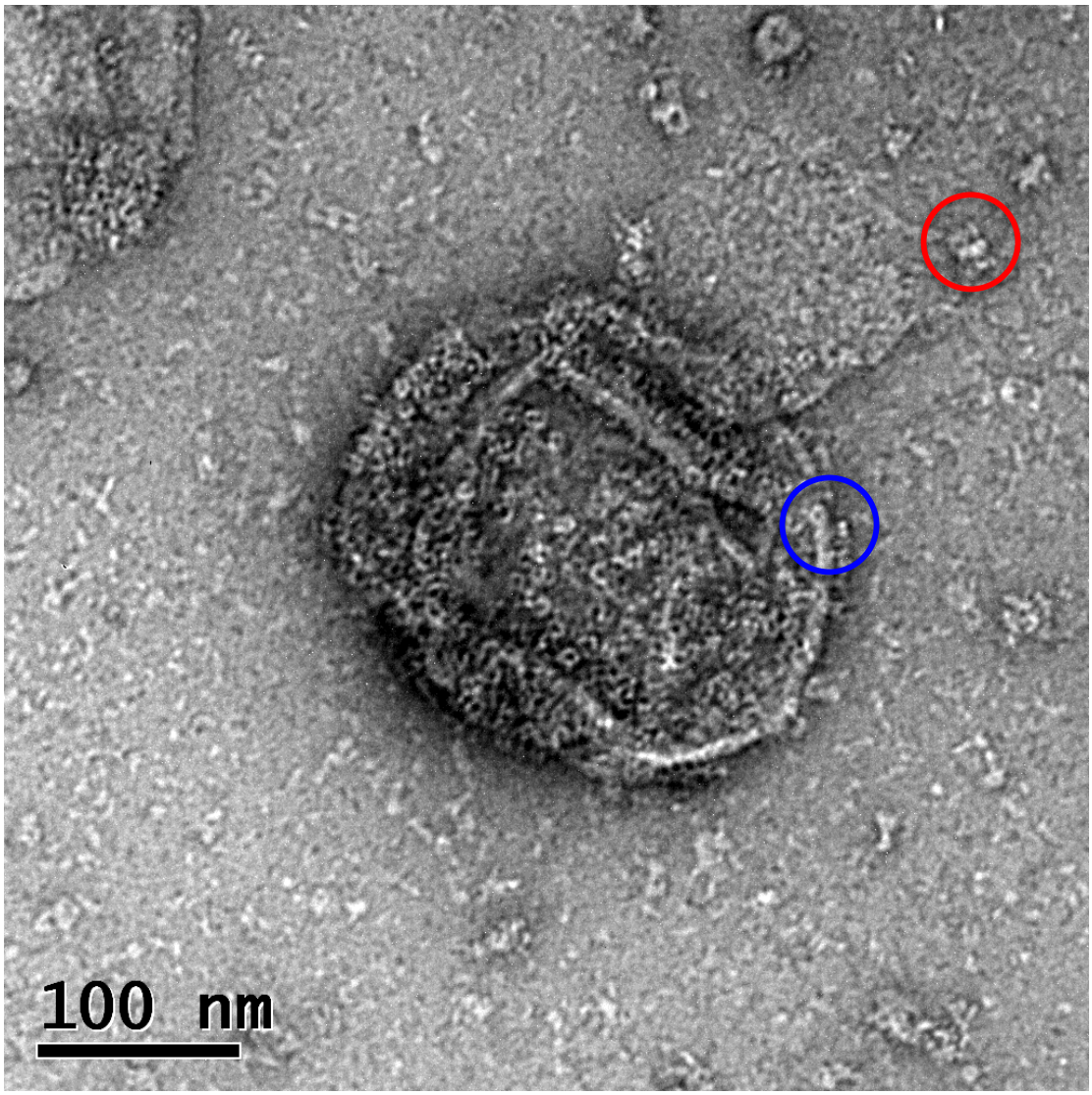


Figure 8.16. Electron micrograph of *E. coli* liposomes with AhlBC pore side-views and 'Viking ships'. Electron micrograph of negatively stained *E. coli* liposome incubated with AhlB and AhlC for 1 hour at 37 °C. Some lipid membrane fragments from the liposome to form a small disc of lipid, with pore-like structures embedded either side (red circle). These side-views of pores are similar to those seen also on the edges of the liposomes themselves (blue circle) .

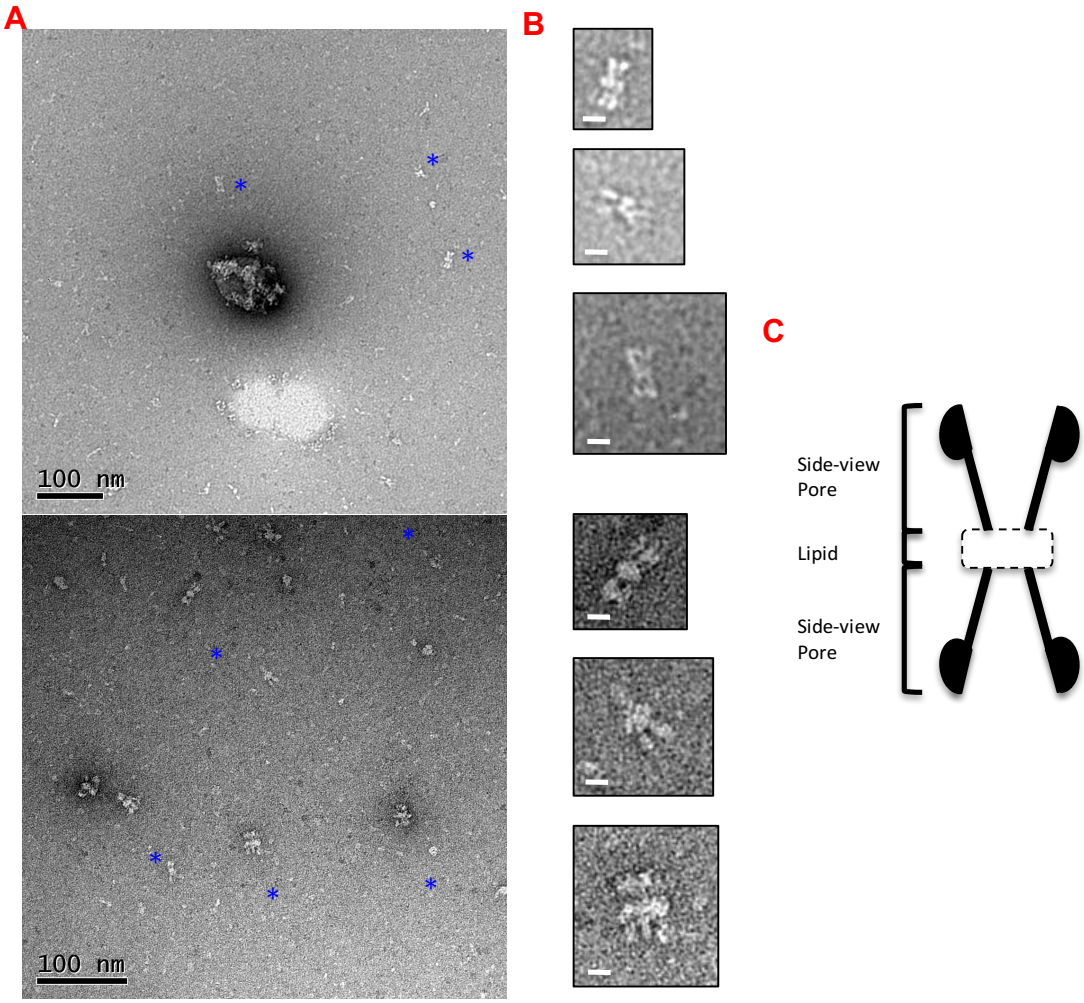


Figure 8.17. Electron micrographs and schematic diagram of the 'Viking ship' assemblies. (**A**) Electron micrographs of negatively stained liposome fragments containing AhlBC pore-like structures (blue stars). (**B**) Magnified views of a small number of these 'Viking ships', highlighting those in which only a single assembly is present on each side of the membrane. (scale bar 10 nm). (**C**) Schematic diagram showing the predicted components of the 'Viking ships', namely a central lipid bilayer with pore-like structures assembled on each face.

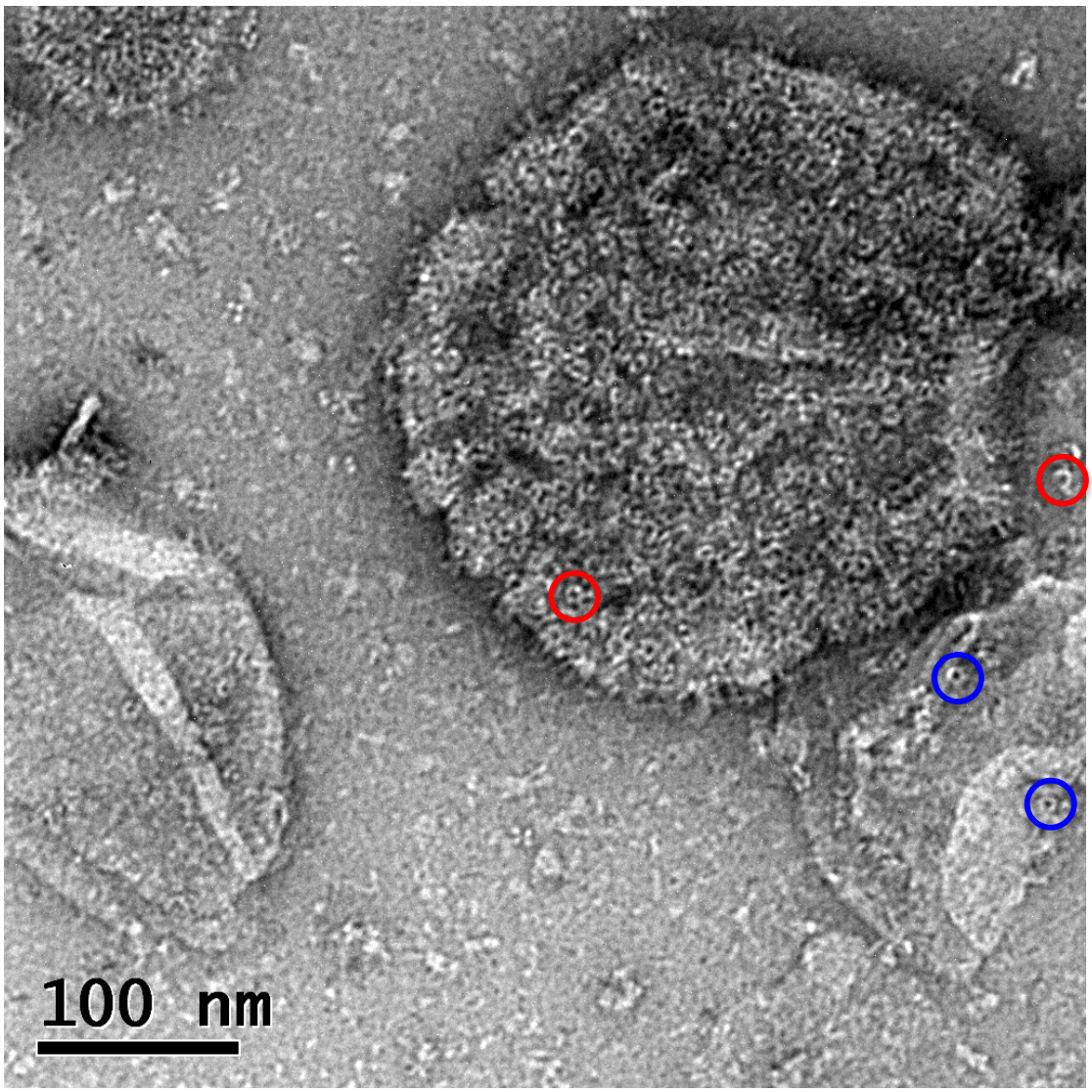


Figure 8.18. **Electron micrographs of E. coli liposomes saturated with AhlBC highlighting top-down views.** Electron micrograph of negatively stained *E. coli* liposome incubated with AhlB and AhlC for 1 hour at 37 °C. Some lipid membrane fragments from the liposome to form a small disc of lipid, with pore-like structures embedded either side (red circle). These side-views of pores are similar to those seen also on the edges of the liposomes themselves (blue circle)

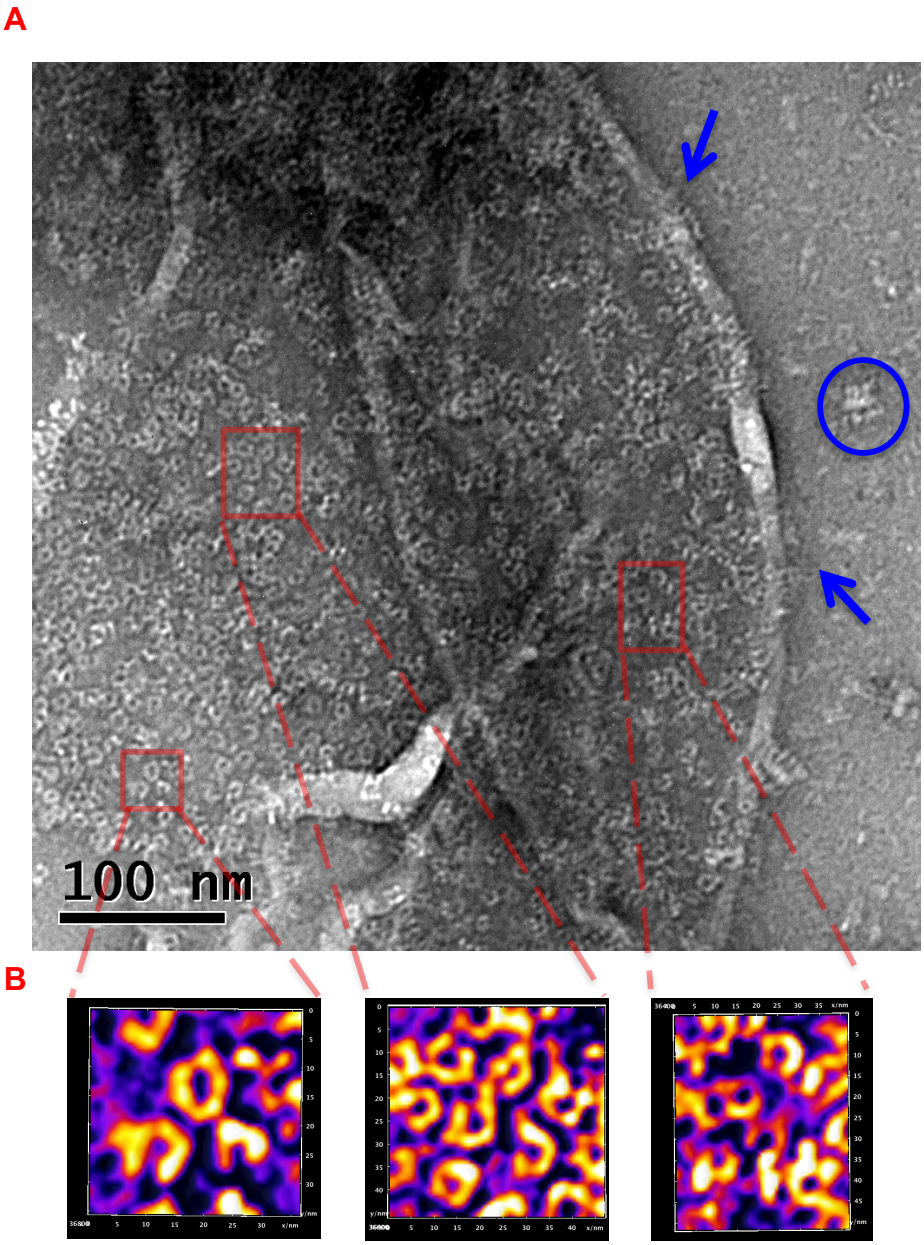


Figure 8.19. Electron micrograph of an *E. coli* liposome incubated with AhlB and AhlC, and profile plots of groups of pores. (**A**) Electron micrograph of a negative stained *E. coli* liposome with AhlBC pores formed at the membrane surface. The large size of the liposome allows single pores to be identified and isolated for diameter estimation. A few 'Viking ship' structures are also present (O), showing side views of the pores. Some side views of pores are also visible on the surface of the liposomes (←), although only in small numbers. (**B**) False colour profile plots of pixel intensities. Pores are highlighted in these images as orange. Many of the pores form either arcs or horseshoes, with thinner walls on one side.

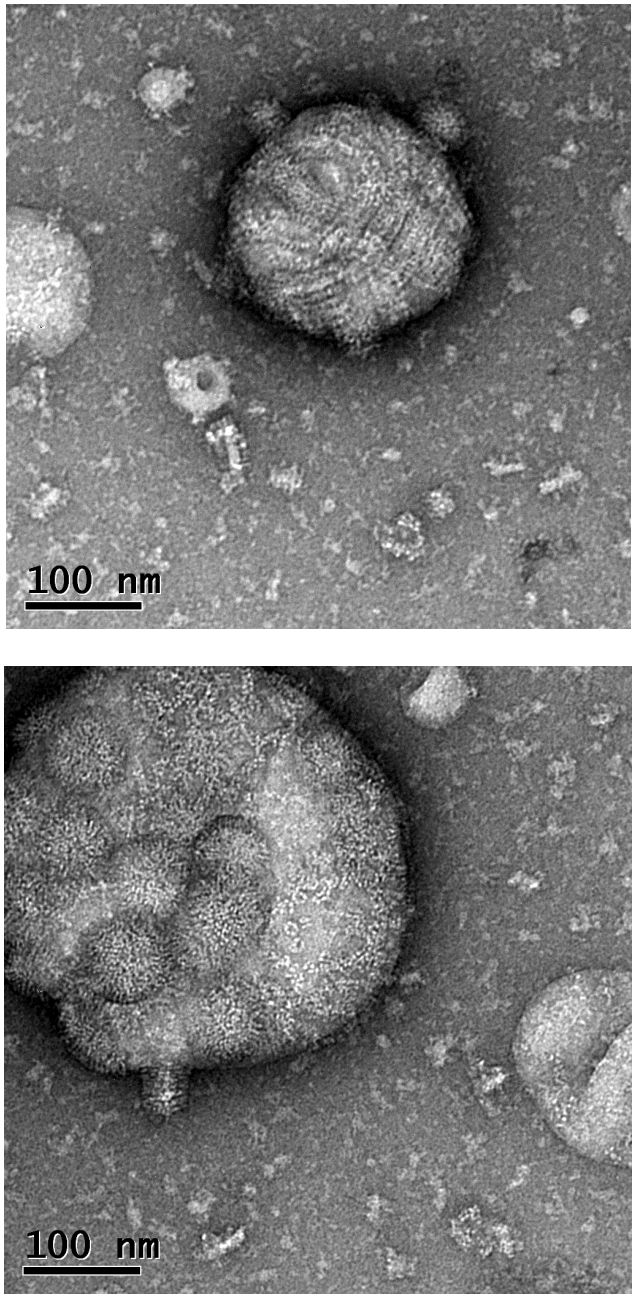


Figure 8.20. Electron micrograph of liver liposomes incubated with AhlB and AhlC. Electron micrographs of negatively stained liposomes incubated with 10 µg AhlB and AhlC for 1 hour at 37 °C. (**A**) In some liposomes spirals of protein were formed, with lines of pore-like assemblies present. (**B**) In other liposomes blebbing of the liposome was observed, where protein was saturating the liposome surface, and curvature was present in the membrane.

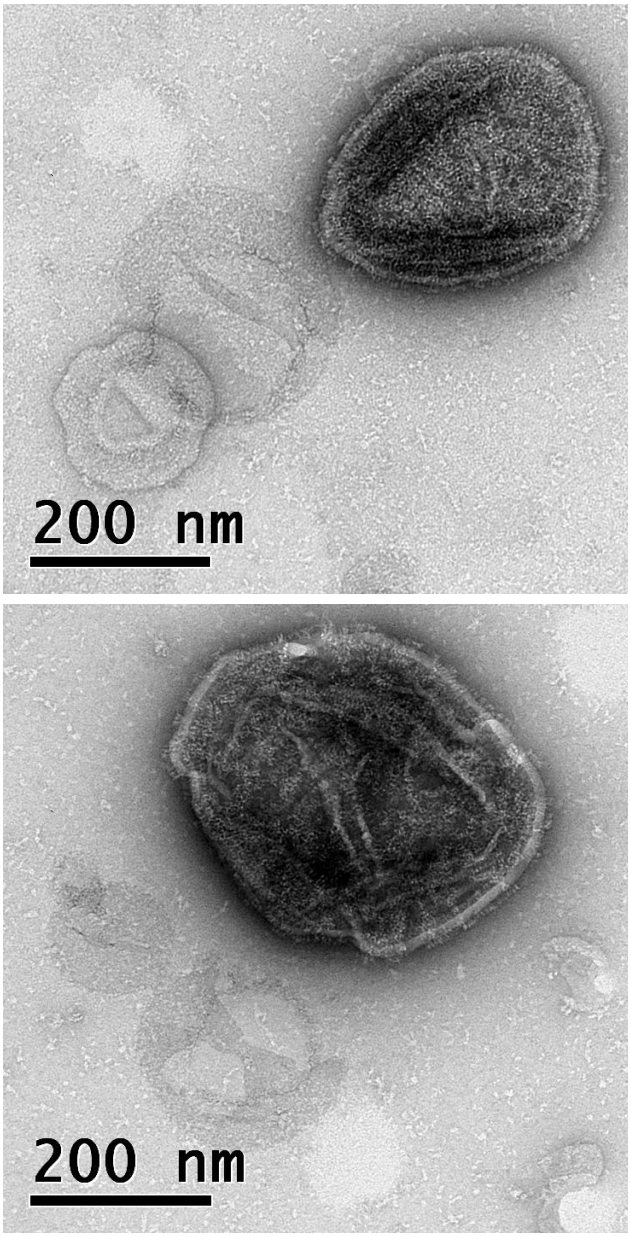


Figure 8.21. Electron micrographs of E. coli liposomes incubated with AhlA, AhlB, and AhlC. Electron micrographs of negatively stained *E. coli* liposomes incubated with AhlA, AhlB, and AhlC for 1 hour at 37 °C. In both micrographs, one liposome is saturated in protein, and is well contrasted against the carbon film, and the other liposomes present contain no pore-like assemblies, similar to that seen for AhlBC incubations.

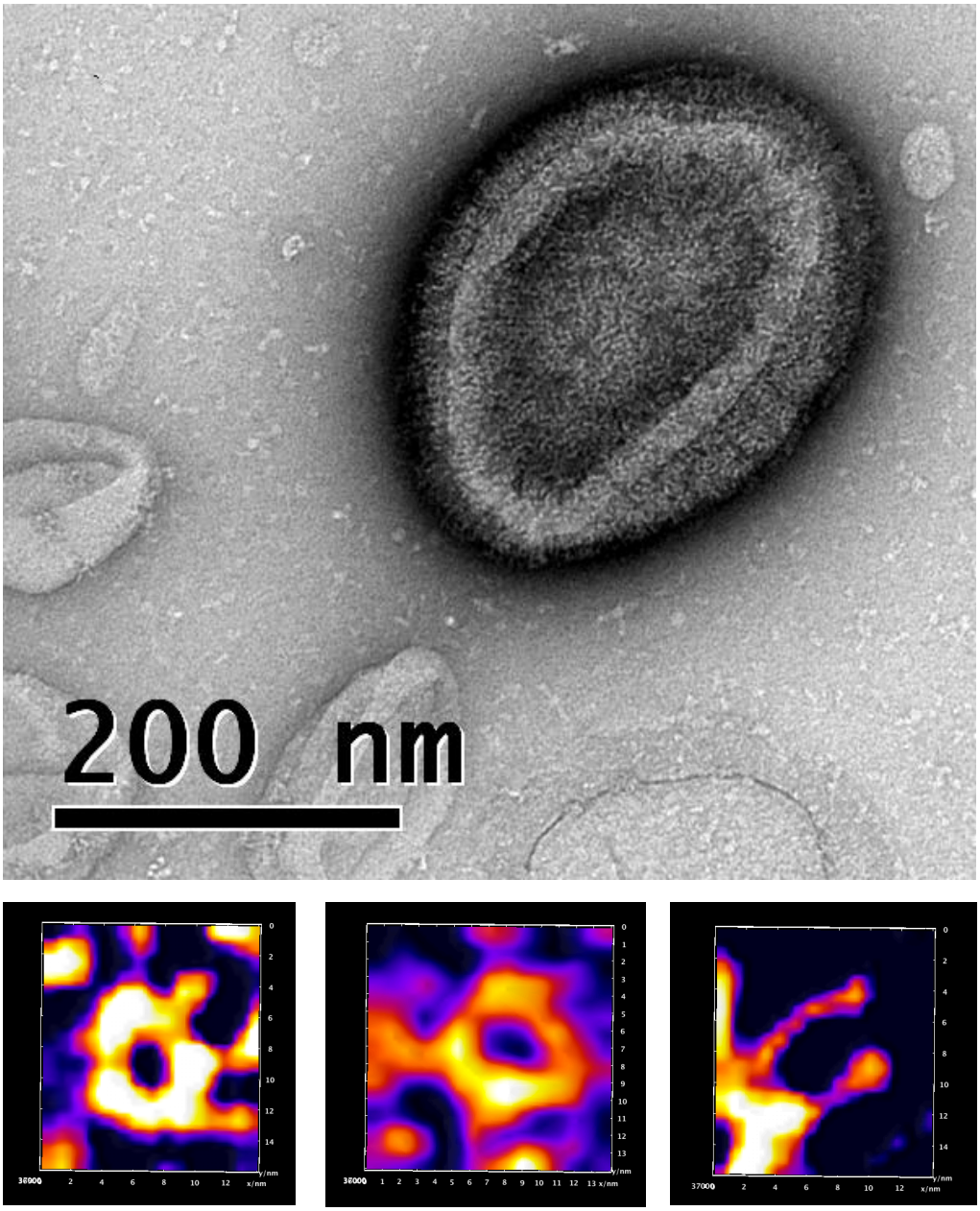


Figure 8.22. Electron micrographs of E. coli liposomes incubated with AhlA, AhlB, and AhlC. Electron micrograph of a negatively stained *E. coli* liposome incubated with AhlA, AhlB, and AhlC for 1 hour at 37 °C. A number of the pores can be resolved and the profile plot calculated. The pores are not obviously any different from AhlBC pores, as is shown by the top-down views, and the side-view (right).

8.5 Electron microscopy experiments of AHL with lipid/detergent bicelles

In order to more easily resolve individual pores, bicelles, which are small lipid discs, were explored in a series of experiments. Lipid bicelles have been used in the past 10 years to study membrane proteins in small micro-domains, and are useful in studying membrane proteins by X-ray crystallography, NMR, and electron microscopy (Dürr et al., 2012). There are two types of bicelle, which are formed of different components. The first type of bicelle is formed of long chain lipid, and either short chain lipid, or detergent. The long chain lipid forms the lipid bilayer that forms the face of the bicelle. The detergent or short chain lipid then forms a curved edge around the lipid bilayer to form a disc. The size of the lipid is determined stochastically, and is usually ~ 40 nm for bicelles formed from DHPC and CHAPSO detergent. Lipid detergent bicelles are formed by temperature cycling a mix of lipid and detergent, where the ratio of detergent to lipid and the temperature in the cooling stage determines the type of lipid structure that forms.

As the lipid requirement for AHL pore formation was not known, different ratios of negatively charged DHPG to zwitterionic DHPC were employed to alter the overall charge of the bicelles. Electron microscopy of the lipid bicelles in isolation showed that lipid bicelles formed successfully, and lay on the carbon grid both flat, and on their side (Figure 8.23). The incorporation of different ratios of charged lipids also had no effect on the size or distribution of bicelles on the carbon grid.

Next, each 10 µg of each AHL protein AhlB and AhlC were incubated with bicelles in isolation, and AhlB and AhlC in combination, were incubated with bicelles for 1 hour at 37 °C. Incubation of DHPC at 37 °C, with no charged DHPG, led to bicelles aggregation and rippling of the membrane surface. These aggregates are known to form by incubating bicelles with low charge at temperatures greater than 30°C, and so these temperatures were avoided in subsequent experiments. AhlC incubated with bicelles formed no visible assemblies in the bicelle, although bicelles were all oriented on the carbon grid such that only top-down views were visible (Figure 8.24). These views then looked similar to the liposome fragments formed when AhlC is incubated with liposomes in isolation.

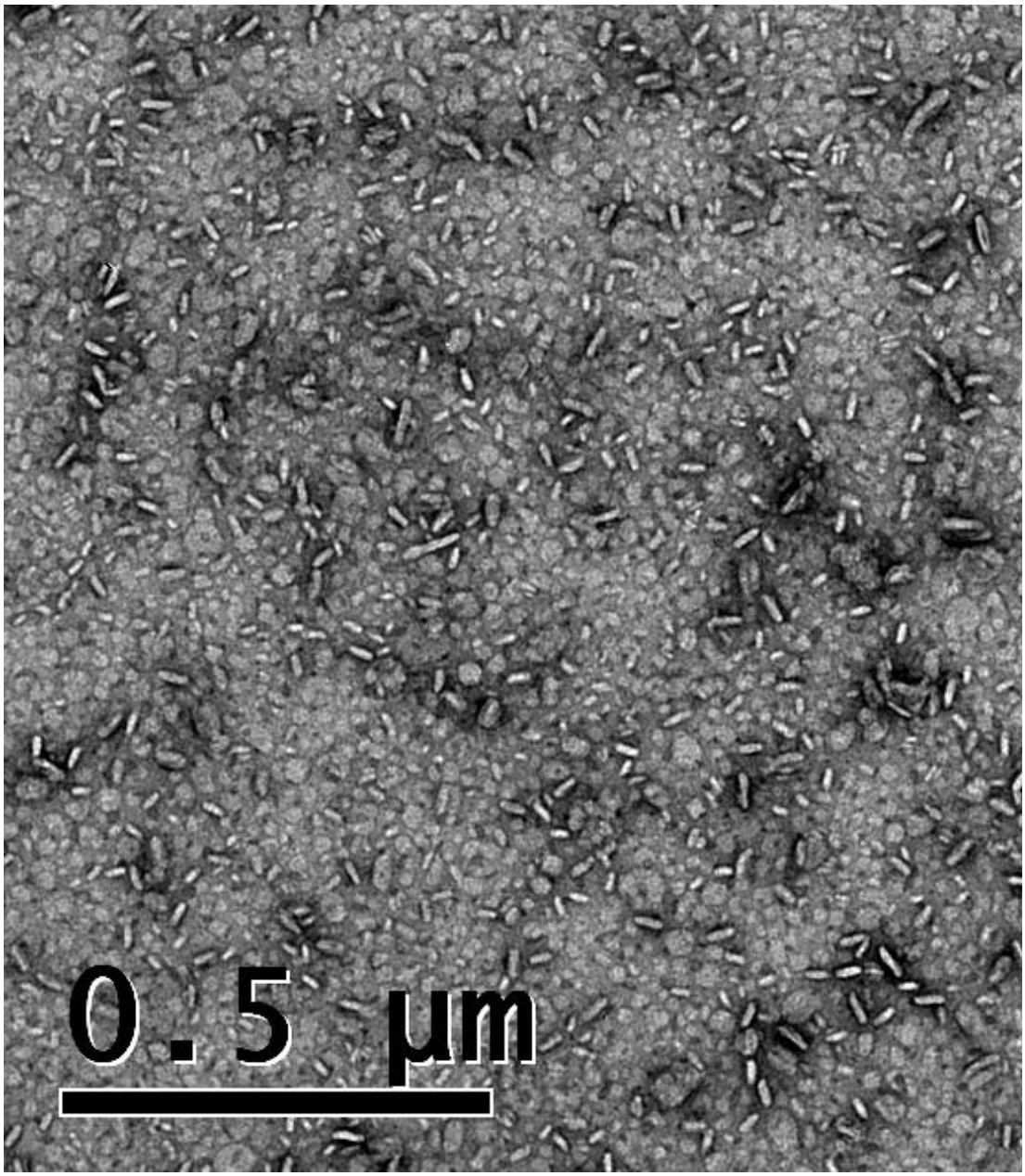


Figure 8.23. Electron micrographs of 20 % (w/v) DHPG containing lipid bicelles. Electron micrograph of negatively stained bicelles constituted of 1:4 ratio of DHPG to DHPC, with CHAPSO as the detergent. Bicelles formed as lipidic discs ~50 nm in diameter, and 10 nm in thickness. Many side-views and top-down views can be seen in the micrograph, indicating bicelles have no heavy preference on orientation on the carbon film.

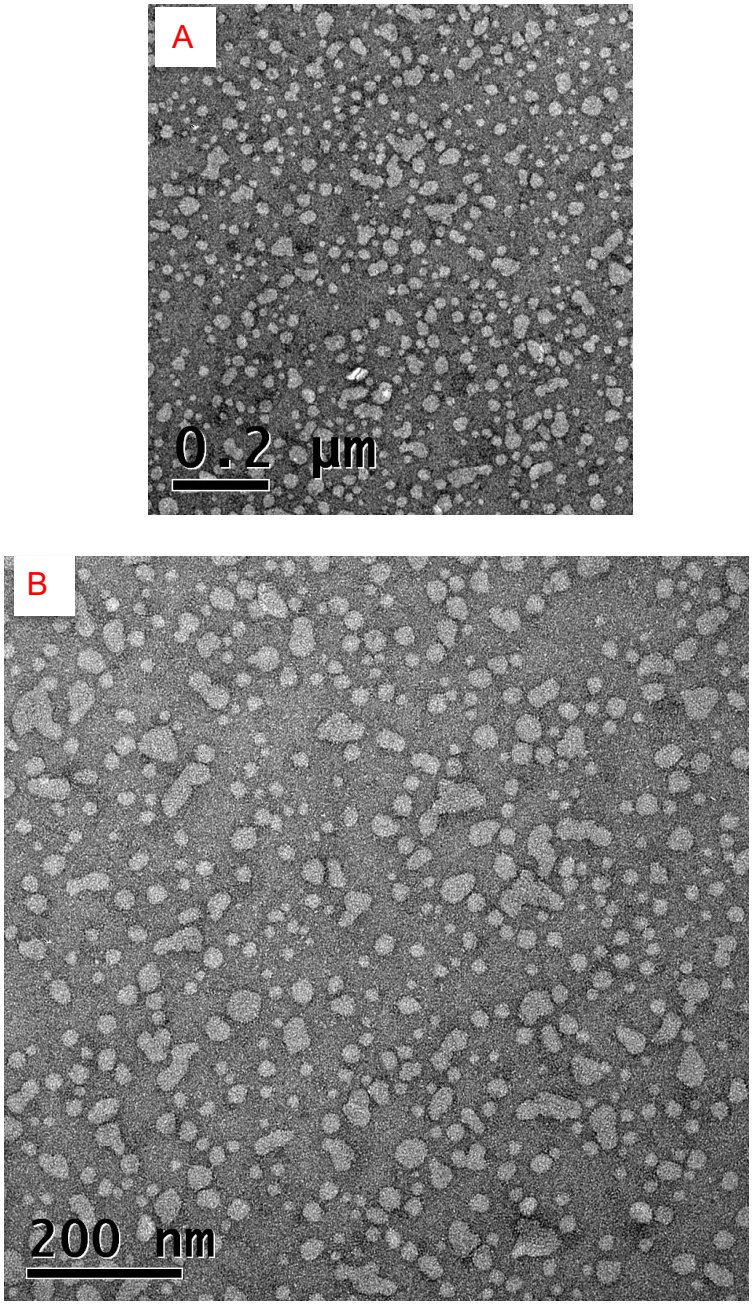


Figure 8.24. Electron micrographs of 1:4 ratio DHPG to DHPC bicelles incubated with AhlC. Electron micrographs of negatively stained 1:4 ratio DHPG to DHPC bicelles incubated with 10 µg AhlC for 1 hour at 37 °C. (A) Low magnification shows a large number of bicelles, all oriented on the carbon film such that only top-views are visible. (B) High magnification electron micrograph, showing the irregularity of the bicelles, and the variation in diameters (10 – 50 nm).

10 µg AhlB incubated with 0.125 % (w/v) bicelles containing 20 % (w/v) DHPG formed no pore-like structures (Figure 8.25), suggesting AhlB is unable to form assemblies in the bicelles, unlike in liposomes.

10 µg of AhlB and AhlC were incubated with 0.125 % (w/v) bicelles with varying concentrations of DHPG. Those bicelles containing 0 % (w/v) DHPC formed aggregates and the rippled morphologies, and so were not useful. However, bicelles containing either 20 % (w/v) or 50 % (w/v) DHPG did allow pores to form. At both concentrations of DHPG, pores were formed in bicelles leading to 'Viking ship' structures (Figure 8.26). Although not homogenous, the bicelles containing AhlB and AhlC were used in preliminary cryo-EM studies, to observe how the samples behaved in ice.

8.5.1 Preliminary cryo-EM studies on AhlB and AhlC in lipid bicelles

In order to generate a three-dimensional map of the pores formed by AHL, cryo-EM studies can be carried out to increase the effective resolution of electron micrographs collected. High-resolution structures are obtained by imaging many (>10,000) particles of the protein of interest, and then averaging these images to generate a two dimensional average. These can be classified into different orientations of the protein, and the class averages can be used to back-project the two dimensional images into a three dimensional map. In order to generate these data, a protein preparation is required in which the protein is large enough to firstly image, then pick by eye i.e. high enough contrast, and then orient based on structural features of the noisy individual images.

Preliminary cryo-EM studies were carried out in order to determine if the AhlB and AhlC pores observed in negative stain EM were amenable to cryo-EM studies, by meeting the above requirements. 10 µg of each of AhlB and AhlC were incubated with 0.125 % (w/v) bicelles, containing 20 % (w/w) DHPG: DHPC, as described above for negative stain electron microscopy studies. The samples were applied to holey film grids, blotted, and then plunge frozen, using a vitro-bot. Frozen samples were stored at liquid nitrogen temperatures and sent to the National Center for Protein Sciences microscopy facility in China for data collection. Dr. Steve Muench from Leeds University collected a series of micrographs (Figure 8.27 and Figure 8.28) at different magnifications to survey the quality of the grids.

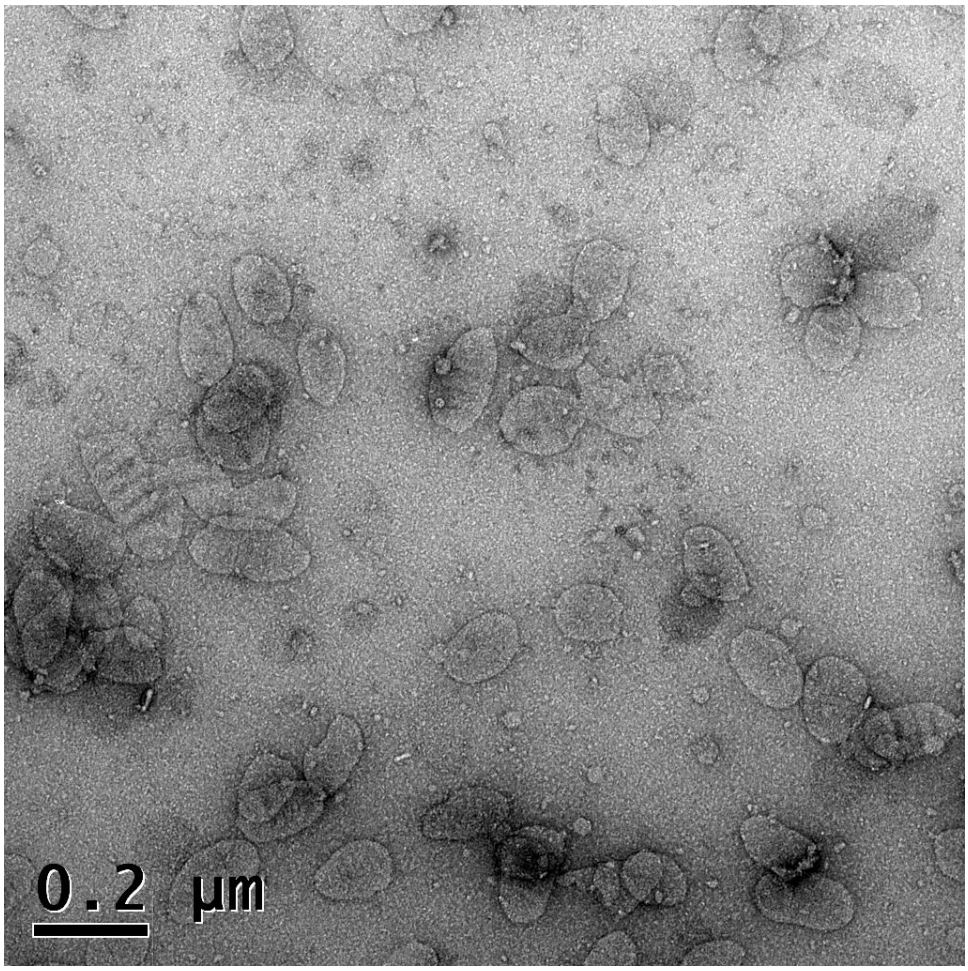


Figure 8.25. Electron micrographs of 1:4 ratio DHPG to DHPC bicelles incubated with AhlB. Electron micrographs of negatively stained bicelles containing 20 % (w/v) DHPG, incubated with 10 µg AhlB for 1 hour at 37 °C. Many bicelles are present on the grid, although no obvious pores are forming as seen with liposomes.

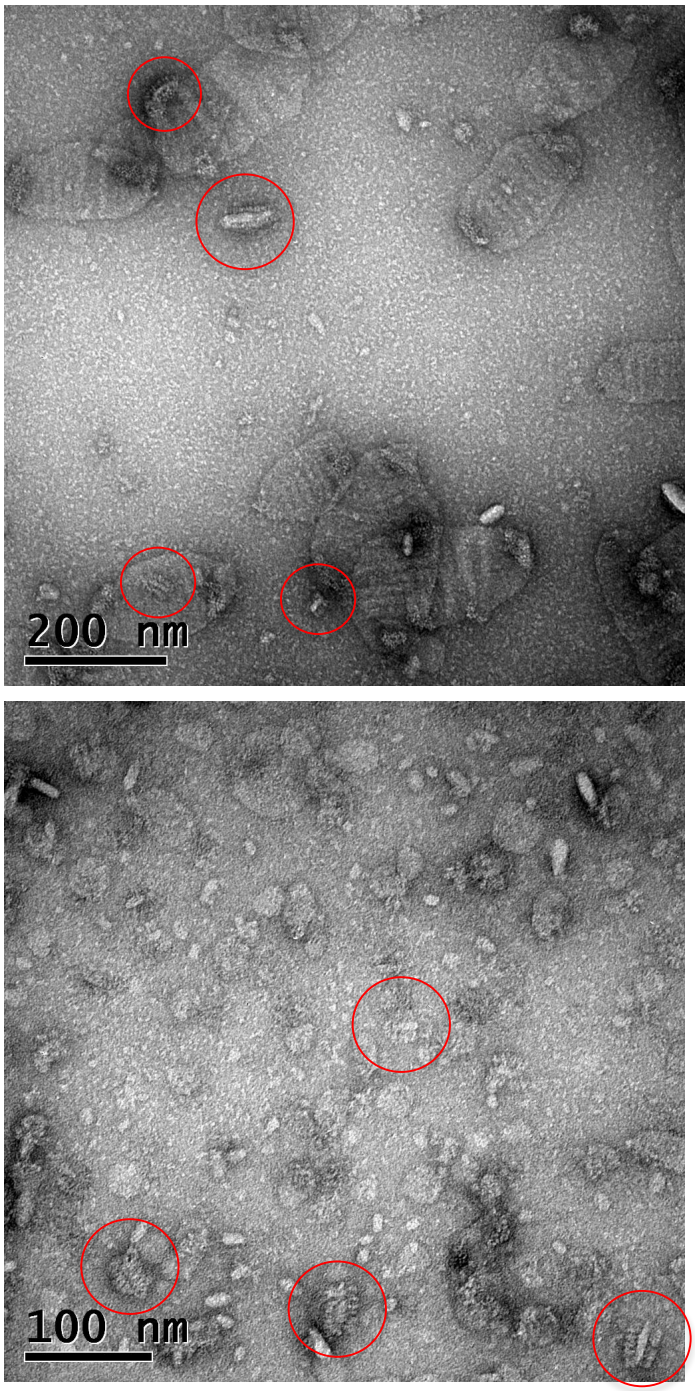


Figure 8.26. Electron micrographs of 20 % and 50 % DHPG containing bicelles incubated with AhlB and AhlC. Electron micrograph of negatively stained bicelles, at 0.25 % (w/v), incubated with 10 µg AhlB and AhlC. (**A**) Bicelles containing 20 % (w/v) DHPG, with pores forming on the edges of the bicelles (red circles). Side-views are also present, with bicelles resembling 'Viking ships'. (**B**) Bicelles containing 50 % (w/v) DHPG, again with side-views of AhlBC pores present (red circles).

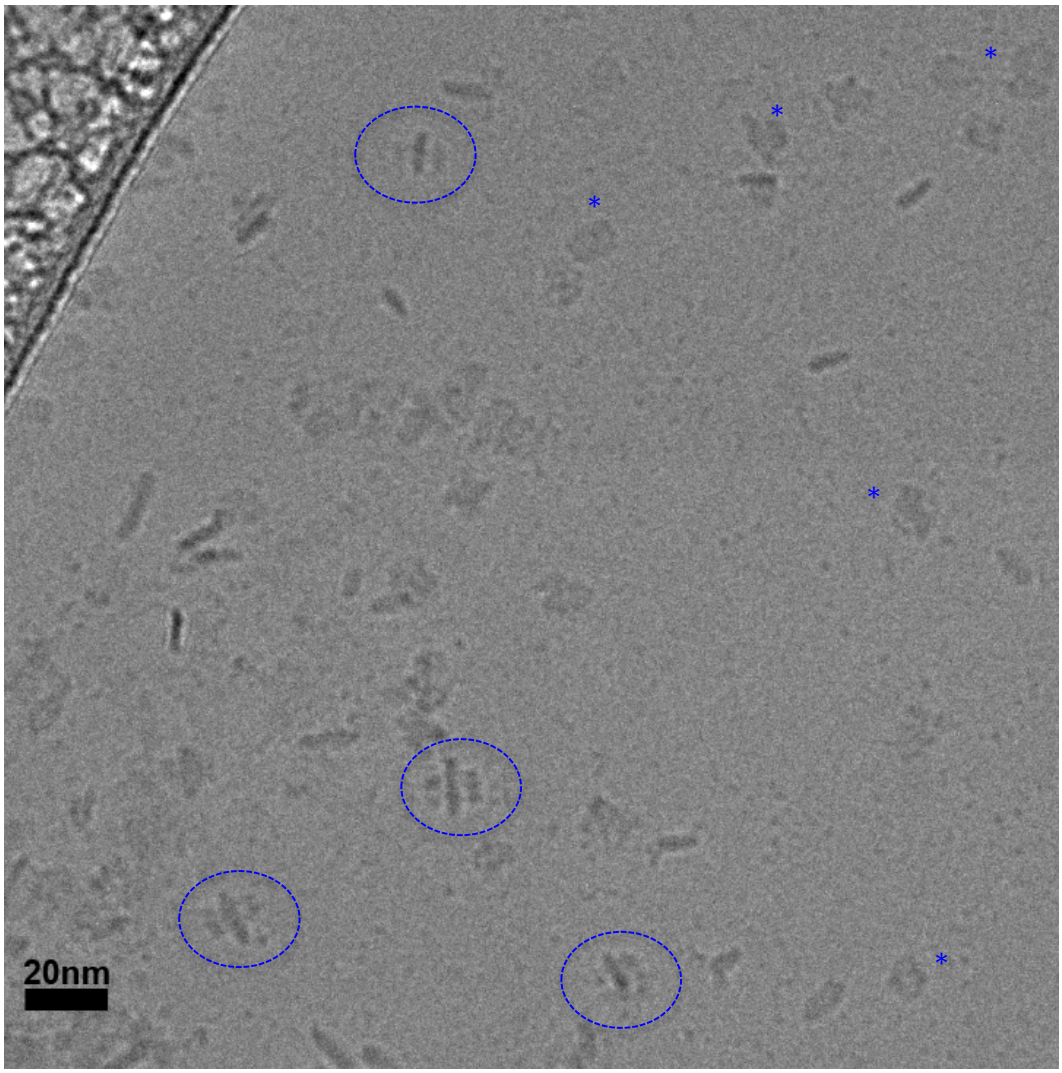


Figure 8.27. Cryo-electron micrographs of bicelles incubated with AhlB and AhlC. Cryo-electron micrographs of bicelles containing 20 % (w/v) DHPG incubated with 10 µg AhlB and AhlC for 1 hour at 37 °C. Samples were blotted and plunge frozen, and then imaged by Dr. Stephen Muench on a Titan Krios electron microscope. Top-down views of bicelles can be seen in the micrographs (*), and side-views of bicelles highlighted in blue circles show the side-views of pores embedded in the bicelles.

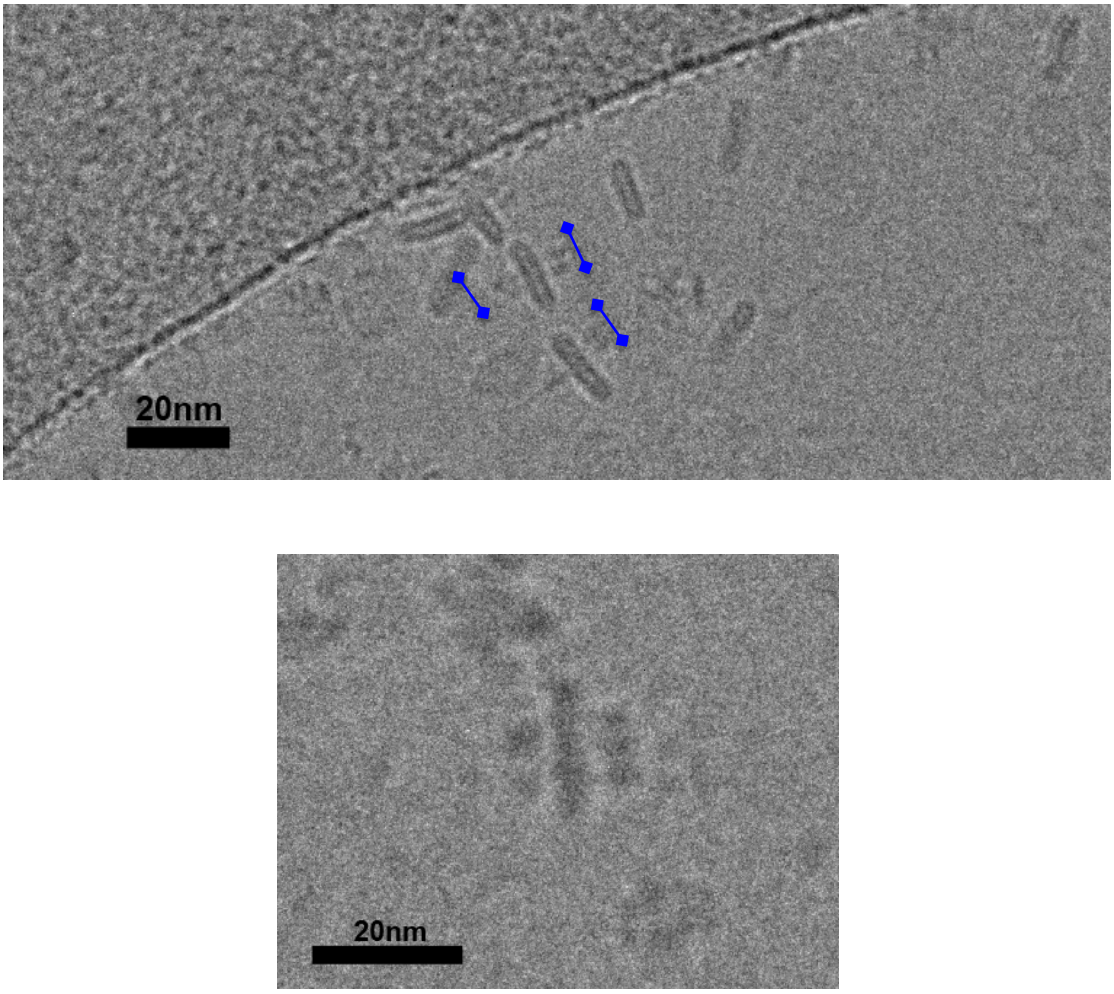


Figure 8.28. Cryo-electron micrographs of bicelles incubated with AhlB and AhlC. Cryo-electron micrographs of bicelles containing 20 % (w/v) DHPG incubated with 10 µg AhlB and AhlC for 1 hour at 37 °C. Samples were blotted and plunge frozen, and then imaged by Dr. Stephen Muench on a Titan Krios electron microscope. Some side-views of pores are highlighted showing the heterogeneous nature of the pores in the raw images. Blue scale bars are drawn across the diameter of the pore with a length of 10 nm, showing each pore is slightly less than this in diameter.

Figure 8.27 show that the bicelles are well behaved in the ice, forming a majority of side-on views, but also containing top down views of the bicelles, indicating that the bicelle orientation is not fixed in the ice. A fixed orientation of bicelles in the ice may have been problematic as protein views would also be restricted to side views. Figure 8.28 shows some higher magnification micrographs, with AhlB and AhlC protein embedded in the bicelles in a similar orientation to the structures observed by negative stain.

The homogeneity of the AhlB and AhlC structures is low, and the amount of bicelles for which these structures are embedded is relatively low, and would require optimisation in order to collect highly redundant datasets.

The high contrast of the membrane in the cryo-EM allows for localisation of viewing required to observe the AhlB and AhlC protein, improving the ability of picking these particles, and in addition, the contrast of the protein itself is also very good, with obvious 'popsicle' structures observed, similar to in negative stain electron microscopy. As the magnification and pixels per angstrom in cryo-EM are much higher than negative stain, estimation can be made of the diameter of the pores, based on the few images available from the preliminary data collection.

In cryo-EM experiments involving the insertion of CDC PFTs into lipid membranes, the formation of a pore is associated with a loss of lipid, which can be seen directly in electron micrographs (Leung et al., 2014). In the preliminary cryo-EM data for AhlB and AhlC, the pores insertion is not associated with a loss of lipid in the bicelle, suggesting a closed pore, or a pore in which the hole is too small to be resolved in raw images. Class averaging in cases where a structure cannot be obtained for the protein may still be useful in determining the open and closed forms of the pore when different AHL components are used.

8.6 Summary

Based on the structure of AhlC, and the crystallographic studies carried out on AhlB, a series of experiments were designed to probe the nature of the pores formed by AhlA, AhlB, and AhlC. For the majority of the experiments, AhlA was omitted, as a strategy for producing non-degrading protein was not found. Circular dichroism experiments showed that AhlB and AhlC are able to associate independently with liposomes, and that AhlB seems to undergo a conformational change in the presence of detergent. Gel filtration showed that DDM and DOC were not able to produce homogeneous pore preparations from AhlB and AhlC, as has been the case for other pores (Bhakdi, 1981; Madegowda et al., 2008; Mueller et al., 2009). As no homogeneous samples could be produced for crystallographic studies, electron microscopy was explored, using liposomes of different constituents. Although interesting morphologies were observed in liver total lipid extract liposomes, *E. coli* total lipid extract liposomes produced more isolated pores that may be amenable to single particle analysis. A series of experiments were performed to better characterise pores formed in *E. coli* liposomes, involving dilutions, time-courses, and different lipid manipulations. The electron micrographs confirmed that AhlB in isolation can form pores, and the pores diameter (~100 Å) was in agreement with the estimated pore diameter from crystallographic studies. The concentrating effect of AhlC seems also to be true, as a tendency for pores to form in liposomes already containing pores was observed, and some liposomes were left without pores inserted, although this may also be as a consequence of certain lipid types or mechanical properties of certain liposomes. AhlC was also able to cause liposomes to be saturated although the same concentration of AhlB was present, and AhlC forms no assemblies in isolation.

Finally, lipid bicelles were explored as a means to alleviate crowding of protein at the membrane surface, and to generate many more ‘Viking ship’ morphologies. These samples are amenable to cryo-EM, and so further study should focus on producing samples of sufficient quality for single particle experiments, for pores formed of AhlB, AhlBC, and AhlABC.

Conclusions and further work

1.8 Conclusions

The initial goals of this thesis were to continue work on the NHE toxin system whilst identifying homologue tripartite toxin family members, which were more amenable to study. During initial experiments, the difficulties experienced by Dr. Magdah Ganesh and others in working with the NHE proteins persisted, and the identification and characterisation of a group of related toxins became the main focus for the majority of this thesis. The discovery of 10 other tripartite PFTs, all in Gram-negative bacteria provided a link between ClyA and NHE/HBL, and a large repository of proteins for structural studies. The tripartite system AHL from *Aeromonas hydrophila* was chosen based on the increasing economic cost of *Aeromonas* infections with haemorrhaging and haemolytic phenotype to fish farms in the U.S. and China .

This chapter begins by summarising the work undertaken during this thesis, as well as the main conclusions. Finally, a model is proposed for AHL pore formation, taking into account the studies undertaken here, and where appropriate combining these results with observations made in studies of the NHE and HBL tripartite systems. This model is given with the aim of directing future experiments and avenues of investigation, which may lead to a greater understanding of this family of pore forming toxins.

The first series of experiments carried out on the AHL proteins, once each protein had been cloned, expressed, and purified, and were used in haemolysis assays. These experiments confirmed that AHL was indeed a member of the tripartite pore forming toxin family, as all three components were required for maximum lysis. For the AHL system, varying the order in which AhlB and AhlC were added to the lysis mixture made no difference to lysis, in contrast to the studies on NHE (Lindbäck et al., 2004), where inhibition is seen if NheB is added before NheC. However, the order of binding of all components of AHL was not determined. Similarly, changing the amount of AhlC in the lysis mixture, compared to AhlA and AhlB made no difference to the rate of lysis, again in contrast to the NHE and HBL systems, where increasing the concentration of the C component relative to the A and B components decreases the rate of lysis.

The potential reason for this lack of complex formation lies in the structure of AhlC, solved here to 2.5 Å. In this structure, a tetramer is formed, burying the hydrophobic residues proposed

to form the membrane insertion domain. This is unlike the biological dimers seen in structures of many other compounds of the family, including HblB, a C component homologue. The AhlC structure instead suggests that disassembly of the tetramer of AhlC is instigated by lipid binding, presenting the hydrophobic residues to the membrane. This is also suggested by the lack of interaction of AhlB and AhlC in solution.

The length of AhlC is much shorter than that of NheC and HblB, and comparison of the AhlC and HblB structure suggest HblB does not behave in the same way, in that a conformational change would be required to expose the HblB hydrophobic residues and for HblB to insert into the membrane. Therefore, even if the roles of these proteins are the same in pore formation, the pathway of pore formation must be different in each case. Further, gel filtration of HblB in the presence of DDM shows that HblB is able to form a 10-mer assembly (Madegowda et al., 2008), whereas such assemblies are not formed upon incubation of AhlC with detergent, suggesting that AhlC has a different mode of action in forming the lytic AHL pore.

The structure of AhlC also reveals why AhlC is unable to lyse erythrocytes by itself, unlike ClyA, which has this function. The N-terminus of AhlC is shorter than in ClyA, so much so that a similar conformational change of the N-terminus of AhlC, similar to that seen in ClyA, would not position the N-terminus of AhlC in the membrane. Therefore, in the predicted mechanism of pore formation, AhlC has no pore forming ability, and instead inserts first into the membrane to allow other AHL proteins to bind to AhlC and increase pore formation.

AhlB has been shown to form pore-like structures by electron microscopy in the presence of liposomes, where rings with central holes are formed. The crystallographic studies of AhlB in the presence of MPD support this pore structure, where 10-fold symmetric assemblies are formed in the crystal. Haemolytic assays show, however, that AhlB is not able to lyse erythrocytes on its own, and so the pore formed by AhlB in isolation is at most a pre-pore. The predicted diameter of the AhlB pre-pore is 115 Å based on the conserved cell dimension in the MPD crystallisation conditions. The electron microscopy pore diameter estimation agrees with these results, based on profile plots of 67 individual pre-pores.

The self-rotation function of the MPD-containing AhlB crystal forms 3 and 4 show that AhlB forms a 10-mer assembly, which is in accordance with the optimal stoichiometry of NHE haemolysis of 10:10:1 NheA to NheB to NheC. If AhlB indeed does form a 10-mer pore, then

the assembly of the other AHL proteins may lie around the circumference of this 10-mer pore, forming multiple layers. Alternatively, the lytic pore may be made up of interchelated units of AhlA, AhlB, and AhlC, forming a larger pore. Electron microscopy experiments and estimation of the pore sizes for the AhlBC and AhlABC does not suggest that the pore changes in size, although there are limits on resolution due to the negative stain, and the magnification of the sample at room temperature.

In the NHE and HBL systems, the inhibition of the pore formation by the C component has been suggested as a way to regulate against self-lysis (Beecher and Wong, 1997b; Lindbäck et al., 2010). In these systems the local concentration around the cell from which the A, B, and C components would be high and therefore pore formation would be inhibited by the high relative C component concentration. As the proteins diffuse away from the cell towards the target, the relative concentration of C would decrease, and lytic pores could form. However, in AHL, this regulatory system cannot occur, as there is no difference in lytic activity on varying the relative concentration of AhlC compared with AhlA and AhlB.

The structure of the AhlC tetramer offers an alternative path, in which AhlC is prevented from binding AhlB by the tetramer formation, and instead has a preference for membranes. Once bound at the membrane, the tetramer of AhlC could dissociate, to insert into the membrane without any conformational change, also leading to concentration of AhlC at the membrane surface. A lipid component or mechanical property of the membrane seems to play a role in pore formation, as in the electron microscopy studies certain liposomes are saturated with pores and arcs, whereas other liposomes are left untouched.

Taken together, these experiments have led to a preliminary model, summarised in Figure 9.1. This model takes into account the observed findings, although does not offer much insight into the role of AhlA, due to the difficulties in purifying the protein and so lack of incorporation into all of the experiments. The model for pore formation by AHL can be summarised as:

- Insertion of AhlC into the membrane by attachment of the hydrophobic head, and disassembly of the tetramer assembly.

- Attachment of AhlB to AhlC at the membrane, and conformational change. Acceleration of the association of AhlB at the membrane and insertion and oligomerisation of AhlB in the presence of AhlC.
- Formation of a pore of AhlB of ~115 Å in diameter, either with only a single AhlC present in each pore, or a 1:1 ratio of AhlB to AhlC.
- Slight activation of the resulting pore by AhlC, explaining why AhlB can form pores but has no haemolytic activity.
- Full activation of the AHL pore by AhlA, leading to full lytic capability of the pores. AhlA's role in pore formation is unknown, but its incorporation into the pore does not drastically affect the size of the resulting pores.

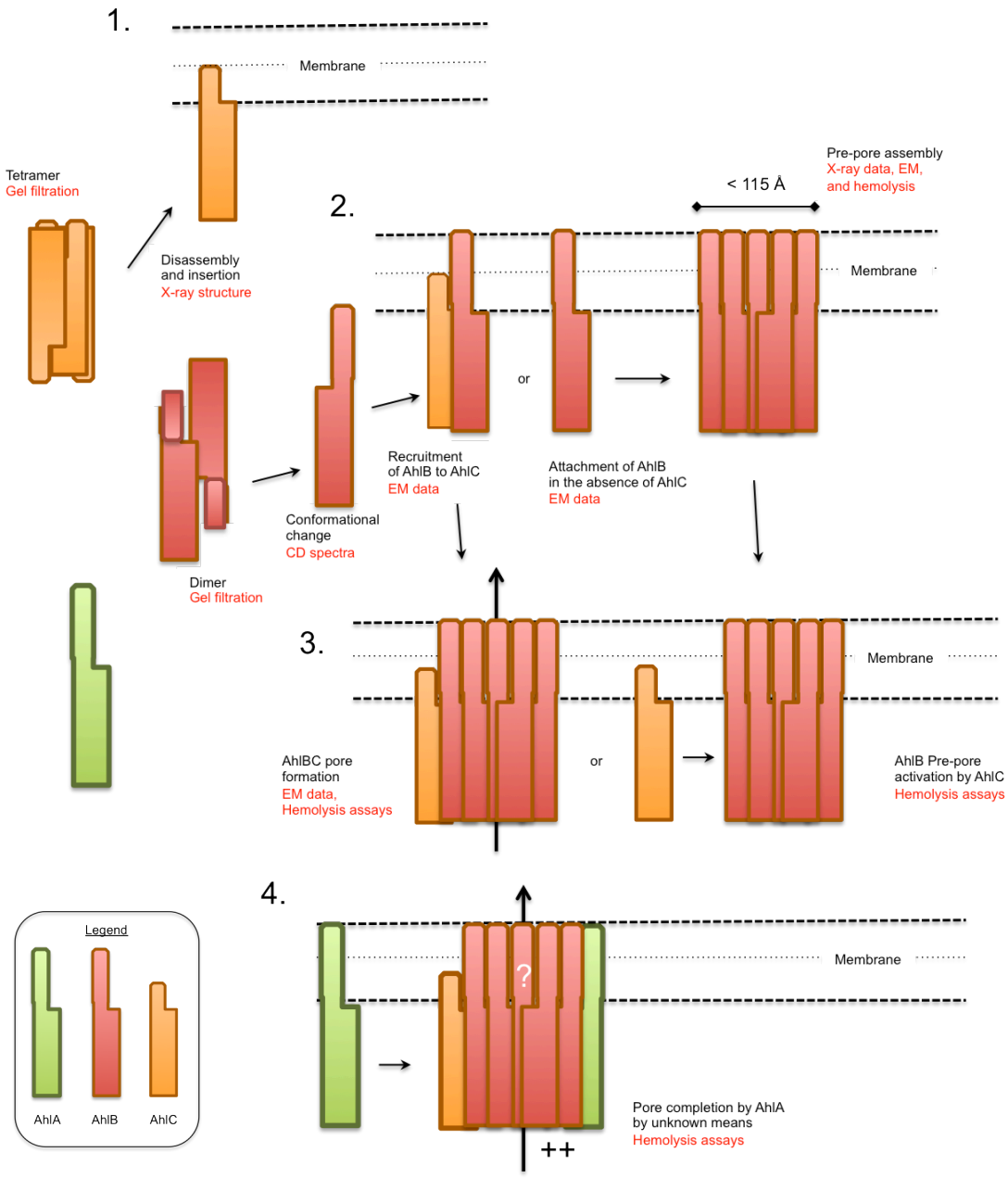


Figure 9.1. Schematic model of pore formation by Ahl, with steps in pore formation indicated with supporting experiments. Schematic diagram showing the proposed stages in pore formation by AHL. AhlC is shown in orange, AhlB is shown in red, and AhlA is shown in green. The elongated portion of each protein represents the membrane inserting domains of each protein, and can be seen situated either fully through, or partially through the lipid bilayer. Experiments supporting each stage are shown in red next to a text description of the stage.

1.9 Further Work

In order to better understand this family of tripartite toxins, a number of experiments would be of use and interest, and are listed below.

1.9.1 Cryo-EM studies of AhlB and AhlBC pores

Preliminary cryo-EM studies have shown that AhlB and AhlC embedded in lipid bicelles are amenable to structural studies. Although liposomes were not tried, the aqueous environment and vitreous ice present in cryo- sample preparation may also make liposomes suitable as a media for studying AHL pores by EM. By determining suitable conditions and collecting highly redundant datasets of single particles, single particle averaging techniques can be used to generate medium resolution (3 – 10 Å) molecular envelopes. This resolution would be sufficient to determine the stoichiometry of the pore when different components are used, and possibly to determine the mode of action of AhlA. A molecular envelope of the AhlB pore may also be of use in phasing the equivalent crystal data by providing a mask for NCS averaging, or by providing a suitable molecular replacement mask for generation of phases from the EM data.

1.9.2 Quantitative haemolysis assays and stable AhlA protein

Haemolytic assays in which AhlA were not being degraded would allow for the ratio of each component for maximum lysis to be established. As the tripartite toxins previously studied seem to have a 10:10:1 A:B:C preference, a difference from this for AHL would be of interest, in further establishing the nature of this toxin. In particular, it would be of interest to see that the maximum ratio of AhlA and AhlB were still 10, but that the ratio of AhlC is unimportant, as is suggested by the haemolytic assays shown in this thesis. Further, AhlA production would allow more quantitative experiments to be performed.

1.9.3 Solving the structure of AhlB

The structure of AhlB is of upmost importance, especially in crystallisation conditions containing MPD, as no pore structure has been solved for this tripartite pore forming system, and an inactive pore may give important information also for the assembly pathway of ClyA. As there are persistent difficulties in solving the structure of AhlB, it may be of interest to examine the other closely related species variants of AhlB. By cloning and overexpressing many of the most homologous AhlB variants and attempting to solve the structure of each, a suitable molecular replacement model may become available to solve the structure of the pore form of AhlB.

1.9.4 Characterising other tripartite and bipartite toxins

Now with a series of experiments carried out for AHL, it would be of great interest to repeat these experiments with the other proteins identified in the bioinformatics studies, namely the bipartite toxins XaxAB and YaxAB, and the other tripartite toxins such as those from *Vibrio campellii* and *Erwinia malitovora*. From the structure of AhlC, it is predicted that the AHL homologues would form a subfamily of the tripartite toxins, with different C component functions or modes of action, which all solubilise by oligomer formation rather than conformational change.

The most important experiment from a crystallographic point would be to clone, overexpress and purify all tripartite homologues identified in bioinformatical studies and attempt to crystallise them all, to be able to carry out comparative analysis on their structures, and offer new insights into the mechanism of pore formation.

1.9.5 Relating AHL studies to the fish disease caused by *Aeromonas hydrophila*

An *in vivo* look at the effect of these tripartite pore forming toxins in the disease caused by pathogenic *Aeromonas hydrophila*, and reveal the contribution of AHL to the disease phenotype. A more complete knowledge of the structure/function relationships of this tripartite toxin family may ultimately lead to the design of compounds that could alter the lytic activity, and therefore may provide a way to target the diseases caused by these organisms. Experiments designed to edit the genome of virulent strains of *A. hydrophila* could be used to knock-out the AHL genes present in AL09-71 strain, and determine if fish pathogenicity is attenuated (Jiang *et al.,* 2013)

References

Abrami, L., Fivaz, M., and Goot, F. van der (2000). Adventures of a pore-forming toxin at the target cell surface. Trends Microbiol. 168–172.

Adams, P.D., Afonine, P. V, Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D *66*, 213–221.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell.

Alouf, J.E. (2000). Cholesterol-binding cytolytic protein toxins. Int. J. Med. Microbiol. *290*, 351–356.

Altschup, S.F., Gish, W., Pennsylvania, T., and Park, U. (1990). Basic Local Alignment Search Tool Stephen. J. Mol. Biol. *215*, 403–410.

Amin, N.M., Bunawan, H., Redzuan, R.A., and Jaganath, I.B.S. (2010). Erwinia mallotivora sp., a new pathogen of papaya (Carica papaya) in Peninsular Malaysia. Int. J. Mol. Sci. *12*, 39–45.

Astier, Y., Braha, O., and Bayley, H. (2006). Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered. J. Am. Chem. … 1705–1710.

Bann, J.G. (2012). Anthrax toxin protective antigen--insights into molecular switching from prepore to pore. Protein Sci. *21*, 1–12.

Barden, S., Lange, S., Tegtmeyer, N., Conradi, J., Sewald, N., Backert, S., and Niemann, H.H. (2013). A helical RGD motif promoting cell adhesion: crystal structures of the Helicobacter pylori type IV secretion system pilus protein CagL. Structure *21*, 1931–1941.

Barta, M., Dickenson, N., and Patil, M. (2012a). The Structures Of Coiled-Coil Domains From Type Three Secretion System Translocators Reveal Homology To Pore-Forming Toxins. J. Mol. … *417*, 395–405.

Barta, M.L., Dickenson, N.E., Patil, M., Keightley, A., Wyckoff, G.J., Picking, W.D., Picking, W.L., and Geisbrecht, B. V (2012b). The structures of coiled-coil domains from type III secretion system translocators reveal homology to pore-forming toxins. J. Mol. Biol. *417*, 395–405.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. Sect. D *67*, 271–281.

Beecher, D., and Macmillan, J. (1991). Characterization of the components of hemolysin BL from Bacillus cereus. Infect. Immun. *59*, 1778–1784.

Beecher, D., and Wong, A. (1997a). Tripartite Hemolysin BL From Bacillus Cereus: Hemolytic Analysis Of Component Interactions And A Model For Its Characteristic Paradoxical Zone Phenomenon. J. Biol. Chem. *272*, 233–239.

Beecher, D., and Wong, A. (1997b). Tripartite Hemolysin BL from Bacillus cereus Hemolytic Analysis Of Component Interactions And A Model For Its Characteristic Paradoxical Zone Phenomenon. J. Biol. Chem. *272*, 233–239.

Beecher, D., Schoeni, J., and Wong, A. (1995). Enterotoxic activity of hemolysin BL from Bacillus cereus. Infect. Immun. *63*, 4423–4428.

Benke, S., Roderer, D., Wunderlich, B., Nettels, D., Glockshuber, R., and Schuler, B. (2015). The assembly dynamics of the cytolytic pore toxin ClyA. Nat. Commun. *6*, 6198.

Bhakdi, S. (1981). Staphylococcal alpha-toxin: oligomerization of hydrophilic monomers to form amphiphilic hexamers induced through contact with deoxycholate detergent micelles. Proc. … *78*, 5475–5479.

Bhakdi, S., and Tranum-Jensen, J. (1991). Alpha-toxin of Staphylococcus aureus. Microbiol. Rev. *55*, 733–751.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *254*, 248–254.

Bravo, a, Gómez, I., Conde, J., Muñoz-Garay, C., Sánchez, J., Miranda, R., Zhuang, M., Gill, S.S., and Soberón, M. (2004). Oligomerization triggers binding of a Bacillus thuringiensis Cry1Ab pore-forming toxin to aminopeptidase N receptor leading to insertion into membrane microdomains. Biochim. Biophys. Acta *1667*, 38–46.

Brown, D. a, and London, E. (2000). Structure and function of sphingolipid- and cholesterol-rich membrane rafts. J. Biol. Chem. *275*, 17221–17224.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Casal, E., Federici, L., and Zhang, W. (2006). The crystal structure of the BAR domain from human Bin1/amphiphysin II and its implications for molecular recognition. Biochemistry *45*.

Cascales, E., Buchanan, S.K., Duché, D., Kleanthous, C., Lloubès, R., Postle, K., Riley, M., Slatin, S., and Cavard, D. (2007). Colicin biology. Microbiol. Mol. Biol. Rev. *71*, 158–229.

Ceuppens, S., Boon, N., and Uyttendaele, M. (2013). Diversity of Bacillus cereus group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. FEMS Microbiol. Ecol. *84*, 433–450.

Chapman, M.S., Blanc, E., Johnson, J.E., Mckenna, R., Munshi, S., Rossmann, M.G., and Tsao, J. (1998). Use of Non-Crystallographic Symmetry for Ab Initio Phasing of Virus Structures. In Direct Methods for Solving Macromolecular Structures, S. Fortier, ed. (Dordrecht: Springer Netherlands), pp. 433–442.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D. a, Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D. Biol. Crystallogr. *66*, 12–21.

Clarke, J., Wu, H., Jayasinghe, L., and Patel, A. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nat. … *4*.

Clatworthy, A.E., Pierson, E., and Hung, D.T. (2007). Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol *3*, 541–548.

Cockeran, R., Steel, H.C., Theron, a J., Mitchell, T.J., Feldman, C., and Anderson, R. (2011). Characterization of the interactions of the pneumolysoid, Δ6 PLY, with human neutrophils in vitro. Vaccine *29*, 8780–8782.

Colwell, R., MacDonald, M., and De Ley, J. (1986). Proposal to Recognize the Family Aeromonadaceae fam. nov. R. … Syst. … *36*, 473–477.

Cordaux, R., Paces-Fessy, M., Raimond, M., Michel-Salzat, A., Zimmer, M., and Bouchon, D. (2007). Molecular characterization and evolution of arthropod-pathogenic Rickettsiella bacteria. Appl. Environ. Microbiol. *73*, 5045–5047.

Cosentino, K., Ros, U., and García-Sáez, A.J. (2016). Assembling the puzzle: Oligomerization of α-pore forming proteins in membranes. Biochim. Biophys. Acta *1858*, 457–466.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D. Biol. Crystallogr. *62*, 1002–1011.

Dal Peraro, M., and van der Goot, F.G. (2016). Pore-forming toxins: ancient, but never really out of fashion. Nat. Rev. Microbiol. *14*, 77–92.

Degiacomi, M.T., Iacovache, I., Pernot, L., Chami, M., Kudryashev, M., Stahlberg, H., van der Goot, F.G., and Dal Peraro, M. (2013). Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. Nat. Chem. Biol. *9*, 623–629.

Dong, J., Qiu, J., Zhang, Y., Lu, C., Dai, X., Wang, J., Li, H., Wang, X., Tan, W., Luo, M., et al. (2013). Oroxylin A inhibits hemolysis via hindering the self-assembly of α-hemolysin heptameric transmembrane pore. PLoS Comput. Biol. *9*, e1002869.

Dunkel, S., Pulagam, L.P., Steinhoff, H., and Klare, J.P. (2015). In vivo EPR on spin labeled colicin A reveals an oligomeric assembly of the pore-forming domain. Phys. Chem. Chem. Phys. *17*, 4875–4878.

Dunkelberger, J.R., and Song, W.-C. (2010). Complement and its role in innate and adaptive immune responses. Cell Res. *20*, 34–50.

Dürr, U., Gildenberg, M., and Ramamoorthy, A. (2012). The magic of bicelles lights up membrane protein structure. Chem. Rev.

Eifler, N., Vetsch, M., Gregorini, M., Ringler, P., Chami, M., Philippsen, A., Fritz, A., Müller, S. a, Glockshuber, R., Engel, A., et al. (2006). Cytotoxin ClyA from Escherichia coli assembles to a 13-meric pore independent of its redox-state. EMBO J. *25*, 2652–2661.

Ellman, G. (1959). Tissue Sulfhydryl Groups. Arch. Biochem. Biophys. 70–77.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. Sect. D *66*, 486–501.

Esser, A. (1994). The membrane attack complex of complement. Assembly, structure and cytotoxic activity. Toxicology.

Evans, G., and Pettifer, R.F. (2001). CHOOCH : a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. J. Appl. Crystallogr. *34*, 82–86.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. *69*, 1204–1214.

Eyster, K. (2007). The membrane and lipids as integral participants in signal transduction: lipid signal transduction for the non-lipid biochemist. Adv. Physiol. Educ. *57069*, 5–16.

Fagerlund, A. (2010). Bacillus cereus cytotoxins Hbl, Nhe and CytK are secreted via the Sec translocation pathway. BMC ….

Fisher, S.J., Levik, K.E., Williams, M. a, Ashton, a W., and McAuley, K.E. (2015). SynchWeb: a modern interface for ISPyB. J. Appl. Crystallogr. *48*, 927–932.

Fortelle, E.D. La, and Bricogne, G. (1997). SHARP: A maximum-likelihood heavy-atom parameter refinement program for the MIR and MAD methods. Methods Enzym.

Franceschini, L., Soskine, M., Biesemans, A., and Maglia, G. (2013). A nanopore machine promotes the vectorial transport of DNA across membranes. Nat. Commun. *4*, 2415.

Fuentes, J. a, Villagra, N., Castillo-Ruiz, M., and Mora, G.C. (2008). The Salmonella Typhi hlyE gene plays a role in invasion of cultured epithelial cells and its functional transfer to S. Typhimurium promotes deep organ infection in mice. Res. Microbiol. *159*, 279–287.

Ganash, M., Phung, D., Sedelnikova, S.E., Lindbäck, T., Granum, P.E., and Artymiuk, P.J. (2013). Structure of the NheA component of the Nhe toxin from Bacillus cereus: implications for function. PLoS One *8*, e74748.

Gasteiger, E., Hoogland, C., and Gattiker, A. (2005). Protein identification and analysis tools on the ExPASy server.

Gattiker, A., Bienvenut, W., Bairoch, A., and Gasteiger, E. (2002). FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification. Proteomics 1435–1444.

Gibrat, J., Madej, T., and Bryant, S. (1996). Surprising similarities in structure comparison. Curr. Opin. Struct. Biol. 377–385.

Giddings, K.S., Zhao, J., Sims, P.J., and Tweten, R.K. (2004). Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. Nat. Struct. Mol. Biol. *11*, 1173–1178.

Gilbert, R. (2002). Pore-forming toxins. Cell. Mol. Life Sci. C.

Gold, W., and Salit, I. (1993). Aeromonas hydrophila infections of skin and soft tissue: report of 11 cases and review. Clin. Infect. Dis. *16*, 69–74.

Van der Goot, F.G. (2001). Pore-Forming Toxins.

Gouaux, E. (1997). Channel-forming toxins: tales of transformation. Curr. Opin. Struct. Biol. 566–573.

Greig, S.L., Radjainia, M., and Mitra, A.K. (2009). Oligomeric structure of colicin ia channel in lipid bilayer membranes. J. Biol. Chem. *284*, 16126–16134.

Grishin, N., Wolf, Y., and Koonin, E. (2000). From complete genomes to measures of substitution rate variability within and between proteins. Genome Res. 991–1000.

Hansen, B., and Hendriksen, N. (2001). Detection of Enterotoxic Bacillus cereus and Bacillus thuringiensis Strains by PCR Analysis. Appl. Environ. … *67*, 185–189.

Harada, Y., and Lifchitz, A. (1981). A translation function combining packing and diffraction information: an application to lysozyme (high-temperature form). … Sect. A Cryst. Physics, ….

Haug, T.M., Sand, S.L., Sand, O., Phung, D., Granum, P.E., and Hardy, S.P. (2010). Formation of very large conductance channels by Bacillus cereus Nhe in Vero and GH(4) cells identifies NheA + B as the inherent pore-forming structure. J. Membr. Biol. *237*, 1–11.

Heijne, G. von (1996). Principles Of Membrane Protein Assembly And Structure Gunnar. Prog. Biophys. Mol. Biol. *66*, 113–139.

Hildebrandlg, A., Pohl, M., and Bhakdill, S. (1991). Staphylococcus aureus alpha-Toxin. J. Biol. Chem. *266*, 17195–17200.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–9.

Hossain, M., Sun, D., and McGarey, D. (2014). An Asian origin of virulent Aeromonas hydrophila responsible for disease epidemics in United States-farmed catfish. MBio *5*, 1–7.

Huang, J., Guan, Z., Wan, L., Zou, T., and Sun, M. (2016). Crystal structure of Cry6Aa: A novel nematicidal ClyA-type α-pore-forming toxin from Bacillus thuringiensis. Biochem. Biophys. Res. Commun. *478*, 307–313.

Hueck, C. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol. Mol. Biol. Rev. *62*, 379–433.

Hunt, S., Moir, A.J.G., Tzokov, S., Bullough, P, Artymiuk, P.J., and Green, J. (2008). The formation and structure of Escherichia coli K-12 haemolysin E pores. Microbiology *154*, 633–642.

Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J. Synchrotron Radiat. *16*, 872–879.

Inoshima, I., Inoshima, N., and Wilke, G. (2011). A Staphylococcus aureus Pore-Forming Toxin Subverts the Activity of ADAM10 to Cause Lethal Infection. Nat. Med. *17*, 1310–1314.

Janda, J.M., and Abbott, S.L. (2010a). The genus Aeromonas: taxonomy, pathogenicity, and infection. Clin. Microbiol. Rev. *23*, 35–73.

Janda, J.M., and Abbott, S.L. (2010b). The genus Aeromonas: taxonomy, pathogenicity, and infection. Clin. Microbiol. Rev. *23*, 35–73.

Jiang, J., Pentelute, B.L., Collier, R.J., and Zhou, Z.H. (2015). Atomic structure of anthrax protective antigen pore elucidates toxin translocation. Nature *521*, 545–549.

Jian, W., Birkard, D., Cox, D., Zhang, F., Marraffini, LA. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology. 31, 233-239.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. Acta Crystallogr. Sect. A Cryst. Physics, … 922–923.

Kabsch, W. (2010). Xds. Acta Crystallogr. D. Biol. Crystallogr. *66*, 125–132.

Kawate, T., and Gouaux, E. (2003). Arresting and releasing Staphylococcal alpha -hemolysin at intermediate stages of pore formation by engineered disulfide bonds. Protein Sci. *12*, 997–1006.

Kelly, K. (1993). Spectrum of extraintestinal disease due to Aeromonas species in tropical Queensland, Australia. Clin. Infect. … 574–579.

Khandelwal, P., and Banerjee-bhatnagar, N. (2003). Insecticidal activity associated with the outer membrane vesicles of Xenorhabdus nematophilus. Appl. Environ. … *69*, 2032–2037.

Kilcullen, K., Teunis, A., Popova, T.G., and Popov, S.G. (2016). Cytotoxic Potential of Bacillus cereus Strains ATCC 11778 and 14579 Against Human Lung Epithelial Cells Under Microaerobic Growth Conditions. Front. Microbiol. *7*, 69.

Kintzer, A.F., Sterling, H.J., Tang, I.I., Williams, E.R., and Krantz, B. a (2010). Anthrax toxin receptor drives protective antigen oligomerization and stabilizes the heptameric and octameric oligomer by a similar mechanism. PLoS One *5*, e13888.

Klimpel, K., and Molloy, S. (1992). Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. Proc. … *89*, 10277–10281.

Knäblein, J., Neuefeind, T., and Schneider, F. (1997). Ta 6 Br 12 2+, a tool for phase determination of large biological assemblies by X-ray crystallography. J. Mol. … 1–7.

Knibiehler, M., and Lazdunski, C. (1987). Conformation of colicin A: apparent difference between cytoplasmic and extracellular polypeptide chain. FEBS Lett. *216*, 183–189.

Ko, W.C., Lee, H.C., Chuang, Y.C., Liu, C.C., and Wu, J.J. (2000). Clinical features and therapeutic implications of 104 episodes of monomicrobial Aeromonas bacteraemia. J. Infect. *40*, 267–273.

Krantz, B., Melnyk, R., and Zhang, S. (2005). A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. … 777–782.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *43*.

Kyte, J., and Doolittle, R. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 105–132.

Lacy, D.B., Mourez, M., Fouassier, A., and Collier, R.J. (2002). Mapping the anthrax protective antigen binding site on the lethal and edema factors. J. Biol. Chem. *277*, 3006–3010.

Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature *227*, 680–685.

Langer, G., Cohen, S., Lamzin, V., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3*, 1171–1179.

Lee, A., and Senior, M. (2016). Dissecting the self-assembly kinetics of multimeric pore-forming toxins. J. ….

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. *55*, 379–IN4.

Lesieur, C., Frutiger, S., and Hughes, G. (1999). Increased Stability upon Heptamerization of the Pore-forming Toxin Aerolysin. J. Biol. … *274*, 36722–36728.

Leung, C., Dudkina, N. V, Lukoyanova, N., Hodel, A.W., Farabella, I., Pandurangan, A.P., Jahan, N., Pires Damaso, M., Osmanović, D., Reboul, C.F., et al. (2014). Stepwise visualization of membrane pore formation by suilysin, a bacterial cholesterol-dependent cytolysin. Elife *3*, e04247.

Libby, S., Goebel, W., Ludwig, A., Buchmeier, N., Bowe, F., Fang, F., Guiney, D., Songer, G., and Heffron, F. (1994). A cytolysin encoded by Salmonella is required for survival within macrophages. Proc. … *91*, 489–493.

Lindbäck, T., Fagerlund, A., Rødland, M.S., and Granum, P.E. (2004). Characterization of the Bacillus cereus Nhe enterotoxin. Microbiology *150*, 3959–3967.

Lindbäck, T., Hardy, S.P., Dietrich, R., Sødring, M., Didier, A., Moravek, M., Fagerlund, A., Bock, S., Nielsen, C., Casteel, M., et al. (2010). Cytotoxicity of the Bacillus cereus Nhe enterotoxin requires specific binding order of its three exoprotein components. Infect. Immun. *78*, 3813–3821.

Lingwood, D., and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. Science *327*, 46–50.

Liu, S., and Crown, D. (2009). Capillary morphogenesis protein-2 is the major receptor mediating lethality of anthrax toxin in vivo. Proc. … *106*.

Liu, Q., Zhang, Z., and Hendrickson, W. a (2011). Multi-crystal anomalous diffraction for low-resolution macromolecular phasing. Acta Crystallogr. D. Biol. Crystallogr. *67*, 45–59.

Lund, T., and Granum, P. (1996). Characterisation of a non-haemolytic enterotoxin complex from Bacillus cereus isolated after a foodborne outbreak. FEMS Microbiol. Lett.

Lund, T., and Granum, P. (1997). Comparison of biological effect of the two different enterotoxin complexes isolated from three different strains of Bacillus cerous. Microbiology 3329–3336.

Madegowda, M., Eswaramoorthy, S., Burley, S.K., and Swaminathan, S. (2008). X-ray crystal structure of the B component of Hemolysin BL from Bacillus cereus. Proteins *71*, 534–540.

Mahlen, S.D. (2011). Serratia infections: from military experiments to current practice. Clin. Microbiol. Rev. *24*, 755–791.

Martin, W., and Russell, M.J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *358*, 59–83; discussion 83–5.

Matsuzaki, K., Yoneyama, S., and Miyajima, K. (1997). Pore formation and translocation of melittin. Biophys. J. *73*, 831–838.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Miles, A., and Wallace, B. (2016). Circular dichroism spectroscopy of membrane proteins. Chem. Soc. Rev.

Miles, G., Movileanu, L., and Bayley, H. (2002). Subunit composition of a bicomponent toxin: staphylococcal leukocidin forms an octameric transmembrane pore. Protein Sci. 894–902.

Miller, R., Gallo, S.M., Khalak, H.G., and Weeks, C.M. (1994). SnB: crystal structure determination via shake-and-bake. J. Appl. Crystallogr. *27*, 613–621.

Moravek, M., Dietrich, R., Buerk, C., Broussolle, V., Guinebretière, M.-H., Granum, P.E., Nguyen-The, C., and Märtlbauer, E. (2006). Determination of the toxic potential of Bacillus cereus isolates by quantitative enterotoxin analyses. FEMS Microbiol. Lett. *257*, 293–298.

Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R., and Ban, N. (2009). The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. Nature *459*, 726–730.

Munro, S. (2004). Organelle identity and the organization of membrane traffic. Nat. Cell Biol. *6*, 469–472.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallogr. Sect. D *53*, 240–255.

Naito, a, Nagao, T., Norisada, K., Mizuno, T., Tuzi, S., and Saitô, H. (2000). Conformation and dynamics of melittin bound to magnetically oriented lipid bilayers by solid-state (31)P and (13)C NMR spectroscopy. Biophys. J. *78*, 2405–2417.

Ngamwongsatit, P., Buasri, W., Pianariyanon, P., Pulsrikarn, C., Ohba, M., Assavanig, A., and Panbangred, W. (2008). Broad distribution of enterotoxin genes (hblCDA, nheABC, cytK, and entFM) among Bacillus thuringiensis and Bacillus cereus as shown by novel primers. Int. J. Food Microbiol. *121*, 352–356.

Nguyen, V., Kamio, Y., and Higuchi, H. (2003). Single-molecule imaging of cooperative assembly of gamma-hemolysin on erythrocyte membranes. EMBO J. *22*.

Nicholls, D., and Ferguson, S. (2013). Bioenergetics.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

Orata, F., and Hedreyda, C. (2011). solation and sequence analysis of the full-length toxR gene of type strain Vibrio campbellii and use of the toxR gene sequence to evaluate variation and relatedness with other Vibrio species. Philipp Sci Lett *4*.

Oscarsson, J., and Mizunoe, Y. (1996). Induction of haemolytic activity in Escherichia coli by the s/yA gene product. Mol. … *20*, 191–199.

Padilla, J.E., and Yeates, T.O. (2003). A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. Acta Crystallogr. Sect. D *59*, 1124–1130.

Panchal, R., Smart, M., Bowser, D., Williams, D., and Petrou, S. (2002). Pore-Forming Proteins and their Application in Biotechnology. Curr. Pharm. Biotechnol. *3*, 99–115.

Pang, M., Jiang, J., Xie, X., Wu, Y., Dong, Y., Kwok, A.H.Y., Zhang, W., Yao, H., Lu, C., Leung, F.C., et al. (2015). Novel insights into the pathogenicity of epidemic Aeromonas hydrophila ST251 clones from comparative genomics. Sci. Rep. *5*, 9833.

Phelps, R., and McKillip, J. (2002). Enterotoxin production in natural isolates of Bacillaceae outside the Bacillus cereus group. Appl. Environ. Microbiol. *68*, 3147–3151.

Phung, D., Granum, P.E., Dietrich, R., Märtlbauer, E., and Hardy, S.P. (2012). Inhibition of cytotoxicity by the Nhe cytotoxin of Bacillus cereus through the interaction of dodecyl maltoside with the NheB component. FEMS Microbiol. Lett. *330*, 98–104.

Pilpel, Y., Ben-Tal, N., and Lancet, D. (1999). kPROT: a knowledge-based scale for the propensity of residue orientation in transmembrane segments. Application to membrane protein structure prediction. J. Mol. Biol.

Pridgeon, J.W., and Klesius, P.H. (2011b). Molecular identification and virulence of three Aeromonas hydrophila isolates cultured from infected channel catfish during a disease outbreak in west Alabama (USA) in 2009. Dis. Aquat. Organ. *94*, 249–253.

Pridgeon, J., Zhang, D., and Zhang, L. (2014). Complete genome sequence of the highly virulent Aeromonas hydrophila AL09-71 isolated from diseased channel catfish in west Alabama. Genome Announc. *2*, 9–10.

Pridgeon, J.W., Klesius, P.H., Song, L., Zhang, D., Kojima, K., and Mobley, J. a (2013). Identification, virulence, and mass spectrometry of toxic ECP fractions of West Alabama isolates of Aeromonas hydrophila obtained from a 2010 disease outbreak. Vet. Microbiol. *164*, 336–343.

Privé, G.G. (2007). Detergents for the stabilization and crystallization of membrane proteins. Methods *41*, 388–397.

Qian, S., Yang, L., Huang, H., and Wang, W. (2008). Structure of transmembrane pore induced by Bax-derived peptide: evidence for lipidic pores. Proc. … 1–5.

Rath, A., and Glibowicka, M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc. … *106*, 1760–1765.

Ravelli, R., and Garman, E. (2006). Radiation damage in macromolecular cryocrystallography. Curr. Opin. Struct. Biol. *16*, 624–629.

Read, R. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. Sect. A *42*, 140–149.

Roderer, D., and Benke, S. (2014). Characterization of variants of the pore-forming toxin ClyA from Escherichia coli controlled by a redox switch. Biochemistry.

Roderer, D., Benke, S., Schuler, B., and Glockshuber, R. (2016). Soluble Oligomers of the Pore-forming Toxin Cytolysin A from Escherichia coli Are Off-pathway Products of Pore Assembly. J. Biol. Chem. *291*, 5652–5663.

Rojko, N., and Anderluh, G. (2015). How Lipid Membranes Affect Pore Forming Toxin Activity. Acc. Chem. Res. *48*, 3073–3079.

Rojko, N., Dalla Serra, M., Maček, P., and Anderluh, G. (2016). Pore formation by actinoporins, cytolysins from sea anemones. Biochim. Biophys. Acta *1858*, 446–456.

Rosado, C.J., Kondos, S., Bull, T.E., Kuiper, M.J., Law, R.H.P., Buckle, A.M., Voskoboinik, I., Bird, P.I., Trapani, J. a, Whisstock, J.C., et al. (2008). The MACPF/CDC family of pore-forming toxins. Cell. Microbiol. *10*, 1765–1774.

Rossjohn, J., Polekhina, G., and Feil, S. (2007). Structures of Perfringolysin O Suggest a Pathway for Activation of Cholesterol-dependent Cytolysins. J. Mol. … *367*, 1227–1236.

Rowe, G., and Welch, R. (1994). Assays of hemolytic toxins. Methods Enzymol. *235*, 657–667.

Rupp, B. (2009). Biomolecular Crystallography.

Sakurai, J., Nagahama, M., and Oda, M. (2004). Clostridium perfringens alpha-toxin: characterization and mode of action. J. Biochem. *136*, 569–574.

Sammito, M., Meindl, K., de Ilarduya, I.M., Millán, C., Artola-Recolons, C., Hermoso, J. a, and Usón, I. (2014). Structure solution with ARCIMBOLDO using fragments derived from distant homology models. FEBS J. *281*, 4029–4045.

Sastalla, I., Fattah, R., Coppage, N., Nandy, P., Crown, D., Pomerantsev, A.P., and Leppla, S.H. (2013). The Bacillus cereus Hbl and Nhe tripartite enterotoxin components assemble sequentially on the surface of target cells and are not interchangeable. PLoS One *8*, e76955.

Schmiel, D.H., and Miller, V.L. (1999). Bacterial phospholipases and pathogenesis. Microbes Infect. *1*, 1103–1112.

Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., and Gouauxt, J.E. (1996). Structure of Staphylococcal Hemolysin , a Heptameric Transmembrane Pore. *274*.

Soskine, M., Biesemans, A., and Maglia, G. (2015). Single-molecule analyte recognition with ClyA nanopores equipped with internal protein adaptors. J. Am. … *137*, 5793–5797.

Sugawara, T., Yamashita, D., Kato, K., Peng, Z., Ueda, J., Kaneko, J., Kamio, Y., Tanaka, Y., and Yao, M. (2015). Structural basis for pore-forming mechanism of staphylococcal α-hemolysin. Toxicon *108*, 226–231.

Sugawara-Tomita, N., Tomita, T., and Kamio, Y. (2002). Stochastic Assembly of Two-Component Staphylococcal ?-Hemolysin into Heteroheptameric Transmembrane Pores with Alternate Subunit Arrangements in Ratios of 3:4 and 4:3. J. Bacteriol. *184*, 4747–4756.

Tamm, L.K., Hong, H., and Liang, B. (2004). Folding and assembly of beta-barrel membrane proteins. Biochim. Biophys. Acta *1666*, 250–263.

Tanaka, Y., Hirano, N., Kaneko, J., Kamio, Y., Yao, M., and Tanaka, I. (2011). 2-Methyl-2,4-pentanediol induces spontaneous assembly of staphylococcal α-hemolysin into heptameric pore structure. Protein Sci. *20*, 448–456.

Tekedar, H., and Waldbieser, G. (2013). Complete Genome Sequence of a Channel Catfish Epidemic Isolate, Aeromonas hydrophila Strain ML09-119 Hasan. Genome … *1*, 1–2.

Thompson, J.R., Cronin, B., Bayley, H., and Wallace, M.I. (2011). Rapid assembly of a multimeric membrane protein pore. Biophys. J. *101*, 2679–2683.

Thune, R., Stanley, L., and Cooper, R. (1993). Pathogenesis of gram-negative bacterial infections in warmwater fish. Annu. Rev. Fish Dis. 37–68.

Thygesen, J., Weinstein, S., Franceschi, F., and Yonath, A. (1996). The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies. Structure 513–518.

Tilley, S.J., and Saibil, H.R. (2006). The mechanism of pore formation by bacterial toxins. Curr. Opin. Struct. Biol. *16*, 230–236.

Turnbull, P. (1976). Studies on the production of enterotoxins by Bacillus cereus. J. Clin. Pathol. *29*, 941–948.

Tweten, R. (2005). Cholesterol-Dependent Cytolysins , a Family of Versatile Pore-Forming Toxins. Infect. Immun. *73*, 6199–6209.

Tzokov, S.B., Wyborn, N.R., Stillman, T.J., Jamieson, S., Czudnochowski, N., Artymiuk, P.J., Green, J., and Bullough, P. a (2006). Structure of the hemolysin E (HlyE, ClyA, and SheA) channel in its membrane-bound form. J. Biol. Chem. *281*, 23042–23049.

Vaidyanathan, M.S., Sathyanarayana, P., Maiti, P.K., Visweswariah, S.S., and Ayappa, K.G. (2014). Lysis dynamics and membrane oligomerization pathways for Cytolysin A (ClyA) pore-forming toxin. RSC Adv. *4*, 4930.

Vandenesch, F., Lina, G., and Henry, T. (2012). Staphylococcus aureus hemolysins, bi-component leukocidins, and cytolytic peptides: a redundant arsenal of membrane-damaging virulence factors? Front. Cell. Infect. Microbiol. *2*, 12.

Vigneux, F., Zumbihl, R., Jubelin, G., Ribeiro, C., Poncet, J., Baghdiguian, S., Givaudan, A., and Brehélin, M. (2007). The xaxAB genes encoding a new apoptotic toxin from the insect

pathogen Xenorhabdus nematophila are present in plant and human pathogens. J. Biol. Chem. *282*, 9571–9580.

Vivekananda, J., Salgado, C., and Millenbaugh, N.J. (2014). DNA aptamers as a novel approach to neutralize Staphylococcus aureus α-toxin. Biochem. Biophys. Res. Commun. *444*, 433–438.

Wagner, N.J., Lin, C.P., Borst, L.B., and Miller, V.L. (2013). YaxAB, a Yersinia enterocolitica pore-forming toxin regulated by RovA. Infect. Immun. *81*, 4208–4219.

Wai, S., Lindmark, B., and Söderblom, T. (2003). Vesicle-mediated export and assembly of pore-forming oligomers of the enterobacterial ClyA cytotoxin. Cell *115*, 25–35.

Wakagi, T., and Oshima, T. (1998). Cloning of the gene for inorganic pyrophosphatase from a thermoacidophilic archaeon, Sulfolobus sp. strain 7, and overproduction of the enzyme by coexpression of tRNA for arginine rare codon. Biosci. ….

Wallace, A., Stillman, T., and Atkins, A. (2000). E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. Cell *100*, 265–276.

Wardenburg, J., Bae, T., and Otto, M. (2007). Poring over pores: α-hemolysin and Panton-Valentine leukocidin in Staphylococcus aureus pneumonia. Nat. Med. 1405–1406.

Waterhouse, A.M., Procter, J.B., Martin, D.M. a, Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Westphal, D., Kluck, R., and Dewson, G. (2014). Building blocks of the apoptotic pore: how Bax and Bak are activated and oligomerize during apoptosis. Cell Death Differ. *21*, 196–205.

Wilke, G. a, and Bubeck Wardenburg, J. (2010). Role of a disintegrin and metalloprotease 10 in Staphylococcus aureus alpha-hemolysin-mediated cellular injury. Proc. Natl. Acad. Sci. U. S. A. *107*, 13473–13478.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D. Biol. Crystallogr. *67*, 235–242.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. *43*, 186–190.

Wu, Q., and Guo, Z. (2010). Glycosylphosphatidylinositols are potential targets for the development of novel inhibitors for aerolysin-type of pore-forming bacterial toxins. Med. Res. Rev. *30*, 258–269.

Xu, X.-P., Zhai, D., Kim, E., Swift, M., Reed, J.C., Volkmann, N., and Hanein, D. (2013). Three-dimensional structure of Bax-mediated pores in membrane bilayers. Cell Death Dis. *4*, e683.

Yang, W.S., Park, S.-O., Yoon, A.-R., Yoo, J.Y., Kim, M.K., Yun, C.-O., and Kim, C.-W. (2006). Suicide cancer gene therapy using pore-forming toxin, streptolysin O. Mol. Cancer Ther. *5*, 1610–1619.

Yip, K.W., and Reed, J.C. (2008). Bcl-2 family proteins and cancer. Oncogene *27*, 6398–6406.

Young, J. a T., and Collier, R.J. (2007). Anthrax toxin: receptor binding, internalization, pore formation, and translocation. Annu. Rev. Biochem. *76*, 243–265.

Zakharov, S., and Cramer, W. (2002). Colicin crystal structures: pathways and mechanisms for colicin insertion into membranes. Biochim. Biophys. Acta (BBA)- … *1565*, 333–346.

Zakharov, S.D., Sharma, O., Zhalnina, M., Yamashita, E., and Cramer, W. a (2012). Pathways of colicin import: utilization of BtuB, OmpF porin and the TolC drug-export protein. Biochem. Soc. Trans. *40*, 1463–1468.

Zheng, W., Cao, H., and Yang, X. (2012). Grass carp (Ctenopharyngodon idellus) infected with multiple strains of Aeromonas hydrophila. African J. Microbiol. Res. *6*, 4512–4520.

Zhu, K., Didier, A., Dietrich, R., Heilkenbrinker, U., Waltenberger, E., Jessberger, N., Märtlbauer, E., and Benz, R. (2016). Formation of small transmembrane pores: An intermediate stage on the way to Bacillus cereus non-hemolytic enterotoxin (Nhe) full pores in the absence of NheA. Biochem. Biophys. Res. Commun. *469*, 613–618.

Zwart, P., Grosse-Kunstleve, R., and Adams, P. (2005). Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. CCP4 Newsl.

Symbols and Abbreviations

ºC	Degrees Centigrade
Å	Angstrom (10^{-10} m)
a, b, c	Real space unit cell dimensions
α, ß, γ	Real space unit cell angles
Abs	Absorbance
AHL	*Aeromonas* haemolysin
amp	Ampicillin
ATP	Adenosine triphosphate
B-factor	Temperature factor
bis-tris	2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol
bis-tris propane	1,3-Bis[tris(hydroxymethyl)methylamino]propane
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Buried surface area
CA	Cardiolipin
CaM	Calmodulin
CCP4	Collaborative Computational Project Number 4
CC	Correlation coefficient
CCD	Charge coupled device
CD	Circular dichroism
CDC	Cholesterol dependent cytolysin
CHAPSO	3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate
ClyA	Cytolysin A
CMC	Critical micelle concentration
Cry	Crystal protein
CSS	Complex significance score
CytK	Cytotoxic protein K
DDM	n-Dodecyl β-D-maltoside
DHPC	1,2-diheptanoyl-*sn*-glycero-3-phosphocholine
DHPG	dihexanoyl phosphatidylglycerol
DTNB	5,5'-dithio-bis-[2-nitrobenzoic acid]
DTT	Dithiotheritol
DOC	Deoxycholate
e^-	electron
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
F or $\mathbf{F_{hkl}}$	Structure factor

|F|	Structure factor amplitude
$\mathbf{F_{obs}}$, $\mathbf{F_{calc}}$	Observed and calculated structure factor
f	Atomic scattering factor
f'	Dispersive component
f"	Anomalous component
g	Gravitational force
GeV	Gigaelectron Volt
GuHCL	Guanidinium chloride
Gy	Gray
HBL	Haemolysin BL
HblB	Haemolysin BL binding domain
$HblL_1$	Haemolysin BL lytic domain 1
$HblL_2$	Haemolysin BL lytic domain 2
HEPES	4-(2-hydrxyethyl)-1-piperazineethanesulfonic acid
hkl	Reciprocal lattice Miller indices
HT	High tension
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Kb	Kilo base pair
KDa	Kilo Dalton
KeV	Kiloelectron Volt
LB	Lysogeny broth
M	Molar concentration
MAD	Multiwavelength anomalous dispersion
MAS	Motile *Aeromonas* septicaemia
mbar	millibar
min	Minute
MES	2-(N-morpholino)ethanesulfonic acid
MCR	Multiple cloning region
MPD	(±)-2-Methyl-2,4-pentanediol
NCS	Non-crystallographic symmetry
NHE	Non-haemolytic enterotoxin
nm	Nanometer
NMR	Nuclear magnetic resonance
OD	Optical density
OMV	Outer membrane vesicle
PAGE	polyacrylamide gel electrophoresis
PA	Phosphatidic acid
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein Data Bank

PBS	Phosphate buffered saline
PEG	Polyethylene Glycol
PE	Phosphatidylethanolamine
PFT	Pore forming toxin
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
pI	Isoelectric point
PISA	Protein Interfaces, Surfaces and Assemblies
R_{free}	Free-R factor
RFZ	Rotation function score
RMS	Root mean squared
RMSD	Root mean square deviation
rpm	Revolutions per minute
R_{merge}	Merging-R factor and interations thereof e.g. R_{meas}
s	second
SAD	Single wavelength anomalous dispersion
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SE	Standard error
SIRAS	Single isomorphous replacement with anomalous scattering
TB	Terrific Broth media
TFZ	Translation function score
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
u,v,w	Patterson cell co-ordinates
UV	Ultraviolet
V	Volt
Vm	Matthews coefficient
$^{v}/_{v}$	Volume to volume
$^{w}/_{v}$	Weight to volume
wt	weight
$^{w}/_{w}$	weight to weight