Structural and Biochemical Analysis of *E. coli* ABC Transporters Implicated in Antimicrobial Peptide Resistance

Bryony Kate Ackroyd

Doctor of Philosophy

University of York

Biology

September 2018

Abstract

Cationic antimicrobial peptides (CAMPs) are a key component of the innate immune system of many organisms, including humans. They target invading pathogens in a variety of ways often integrating into, and permeabilising, bacterial cell membranes and causing cell death. In response, bacteria have developed a variety of CAMP resistance mechanisms, including those based on ATP-binding cassette (ABC) transporters such as Sap and Yej, which are the subject of studies described herein. ABC importers use an extracellular substrate binding protein (SBP) to recognise substrates and deliver them to a cognate membrane complex for uptake into the cell. A primary aim of this study was to unravel the structural basis of CAMP binding by the SBPs, SapA and YejA. CAMPs are larger than conventional peptides handled by ABC transporters and usually contain secondary structure.

SapA from diverse bacterial species proved to be insoluble. Moreover, phylogenetic analysis carried out here suggested to us that SapA is in fact a dipeptide binding protein. YejA from *E. coli* was purified and crystallised and its structure determined by X-ray diffraction methods, revealing the protein in the closed conformation with an endogenous undecapeptide, LGEPRYAFNFN, bound in a spacious cavity enclosed by two protein lobes. Remarkably this peptide is derived from sequence close to the N-terminus of YejA itself. LGEPRYAFNFN was shown via thermal shift assays and mass spectrometry to bind specifically to YejA. Other ligands of variable length derived from the N-terminus of YejA were also shown to bind to YejA; all possess the core motif EPRYAFN. We have so far been unable to demonstrate binding of CAMPs to YejA. This data leaves open the intriguing possibility that Yej function may be regulated by proteolytic processing of the extended N-terminal region of YejA and binding and uptake of the produced peptide.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank both Prof. Anthony Wilkinson and Prof. Gavin Thomas for the opportunity to complete a PhD under their supervision and for all their help, guidance and support throughout my time at York. I would also like to thank my thesis advisory panel, Dr. Martin Fascione and Dr. Marjan Van der Woude for their guidance and encouragement.

I am grateful to current and previous members of the York Structural Biology Laboratory (YSBL) and Thomas group for their company and help in keeping me motivated with many cups of tea, especially during writing. A special thanks also goes to Reyme Herman, Simon Grist and members of technical support staff, whose smooth running of both labs makes doing experiments that much easier.

Dr. Adam Dowle and Dr. Andrew Leech, thank you for all of your help with various techniques such as mass spectrometry and SEC-MALLS. My thanks also go to Dr. Jared Cartwright who kindly gifted me the pETFPP_30 vector, Prof. Eleanor Dodson who generously gave her time to aid in the initial stages of solving the structure of EcYejA and Robin Brabham for the help and guidance with peptide synthesis.

Lastly, I would like to thank my family for their continued love and support, especially during the not so good times. To all the friends that have helped me over the years, you know who you are, thanks for all the patience and kind words.

Declaration

All work carried out in this thesis was completed by the author with exception of collaborative work, which is stated within the main body of the text or outlined below. The writing of the thesis is all the authors own work. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Work carried out or overseen by others:

- Denaturing electrospray mass spectrometry experiments were performed and analysed by Dr. Andrew Leech in the York Technology Facility (TF).

- SEC-MALLS and CD experiments were overseen by Dr. Andrew Leech in the York TF.

- Native ESI and MALDI-MS/MS was carried out and analysed in collaboration with Dr. Adam Dowle in the York TF.

- A number of thermal shift assays were carried out by project student Rebecca Lees under the direct supervision of the author.

- X-ray diffraction data was collected by Dr. Johan Turkenburg and Sam Hart in YSBL.

Chapter One

Introduction

1. Introduction

1.1 Antimicrobial resistance

Antibiotics have undoubtedly saved the lives of millions of people worldwide and were seen as wonder drugs for much of the 20th century. They were considered to be magic bullets which selectively targeted disease causing microbes without harming the host. Many people thought they would stop communicable disease. The 1950s to the 1970s was considered the golden era of antibiotic discovery. However, Fleming was quick to caution about the misuse of antibiotics and the potential for resistance to emerge (Zaman *et al.*, 2017).

Antibiotics work with the immune system to clear the body of pathogens and work in a multitude of ways including inhibiting the synthesis of the bacterial cell, proteins and DNA and membrane disorganisation. Over the last 60 years, millions of metric tons of new classes of antibiotics have been manufactured. As demand for antibiotics has increased across all sectors, it has allowed for cheaper off-label drugs, helping to spread resistance (Zaman *et al.*, 2017).

Antibiotic resistance can develop very quickly (Figure 1.1) and has been found in every country (World Health Organisation, 2018). 2 million people a year in the US are infected with antibiotic resistant bacteria resulting in the death of 23,000 people, this is a huge strain on the healthcare system (CDC, 2018). In 2016 the Wellcome Trust put out a report on antimicrobial resistance that predicted that by 2050 10 million people a year would die of antimicrobial resistant infection, more than are predicted to die of cancer, costing over $1 trillion per year in healthcare fees worldwide (Wellcome Trust, 2016). However, with the improvements in information technology people are more aware than ever of antibiotic resistance, its causes and how to prevent it (Zaman *et al.*, 2017).

A whole host of mechanisms are employed by bacteria to resist antibiotic killing as well as a range of methods of acquiring those resistance traits. Acquisition of different resistance mechanisms can be split into three categories; innate, acquired or adaptive. Another major contributor to the development of antibiotic resistance is the natural mutation rate of bacteria. As mutations occur they can bring costs or benefits to the bacteria, those which are beneficial are commonly maintained and spread throughout the population via various means, sometimes these mutations are antibiotic resistance (Schroeder, Brooks and Brooks, 2017).
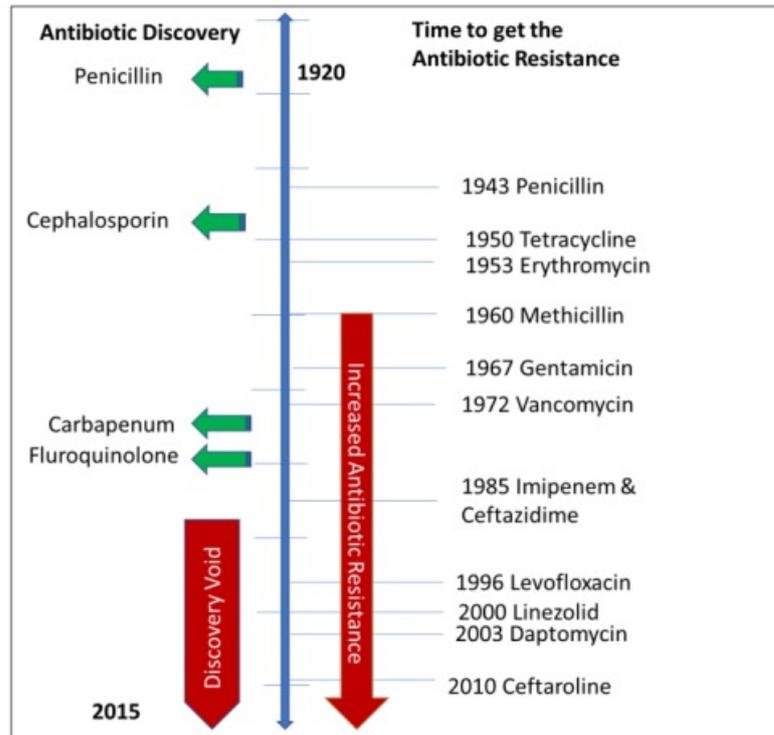
Figure 1.1. Timeline of antibiotic discovery and resistance development.

Shortly after the discovery of a new antibiotic, resistance develops. Now, due to the discovery void, there are fewer new antibiotics coming through the pipeline and so fewer are effective and available for use. Image taken from (Zaman *et al.*, 2017).

Innate resistance is the natural ability of a bacterium to resist the activity of a particular antibiotic or antimicrobial due to the inherent structure or functional characteristics of the organism (Figure 1.2). This is also sometimes known as insensitivity to particular antibiotics and antimicrobials. Innate resistance mechanisms can include a lack of affinity for the antibiotic; inaccessibility of the antibiotic to the bacterial cell; extrusion of the antibiotic from the cell or enzymes which inactivate the antibiotic. These resistance mechanisms are commonly due to the fact that the bacteria themselves produce antibiotics so these mechanisms are a form of self-immunity that can be employed against other similar antibiotics or a simple mutation which enables an efflux pump to now recognise an antibiotic (Schroeder, Brooks and Brooks, 2017).

Acquired resistance is categorised as the ability of a bacteria to not only survive antibiotic stress but to also acquire resistance under antibiotic selective pressure, for example via horizontal gene transfer (Schroeder, Brooks and Brooks, 2017). Acquired resistance is commonly spread through horizontal gene transfer (Figure 1.3), this is the passing of genetic information from one bacteria to another and contrasts with vertical gene transfer in which genetic information is passed on to the next bacterial generation via chromosome replication and cell division. Horizontal gene transfer allows the rapid passage of genetic information between multiple different bacterial cells and allows the passing of genetic information between different bacterial species, one of the reasons antibiotic resistance is spreading so quickly. There are three mechanisms of horizontal gene transfer; (i) transformation; (ii) transduction and (iii) conjugation (Von Wintersdorff *et al.*, 2016).

Transformation (Figure 1.3) is the process by which bacteria take up, integrate and express pieces of exogenous DNA from their environment (Griffith, 1928). It is therefore possible for bacteria to become resistant to antibiotics if they use this natural transformation process to pick up pieces of DNA encoding antibiotic resistance. This transformation process has been shown to transfer resistance genes between bacteria of the same and different species (Alexander, Hahn and Leidy, 1956).

Transduction (Figure 1.3) involves bacteriophages transferring genes between bacteria that are advantageous to their bacterial hosts, this in turn promotes their own survival and propagation (Modi *et al.*, 2014). As with transformation these bacteriophage transported genes are integrated and expressed within the new bacterial host. Gene transfer agents are very similar to bacteriophage, however they are host cell derived and contain only information produced by the host cell. They mediate horizontal gene transfer and package
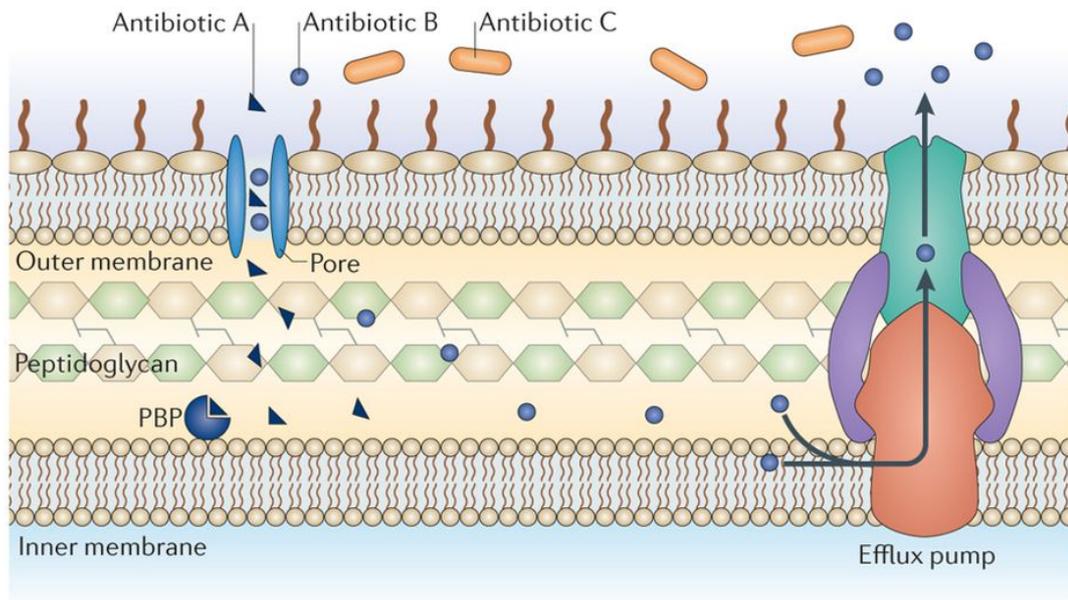
Figure 1.2. Mechanisms of innate antibiotic resistance.

Diagram depicting the different forms of innate resistance. An example of a β-lactam antibiotic targeting a penicillin binding protein (PBP), antibiotic A can enter the cell via a porin and reach its target, inhibiting peptidoglycan synthesis. Antibiotic B is removed from the periplasm via an efflux system, so is unable to reach its target and antibiotic C is not able to pass through the outer membrane so cannot target the PBP. Innate resistance is displayed against both antibiotic B and C. Image taken from (Blair *et al.*, 2015)**.**
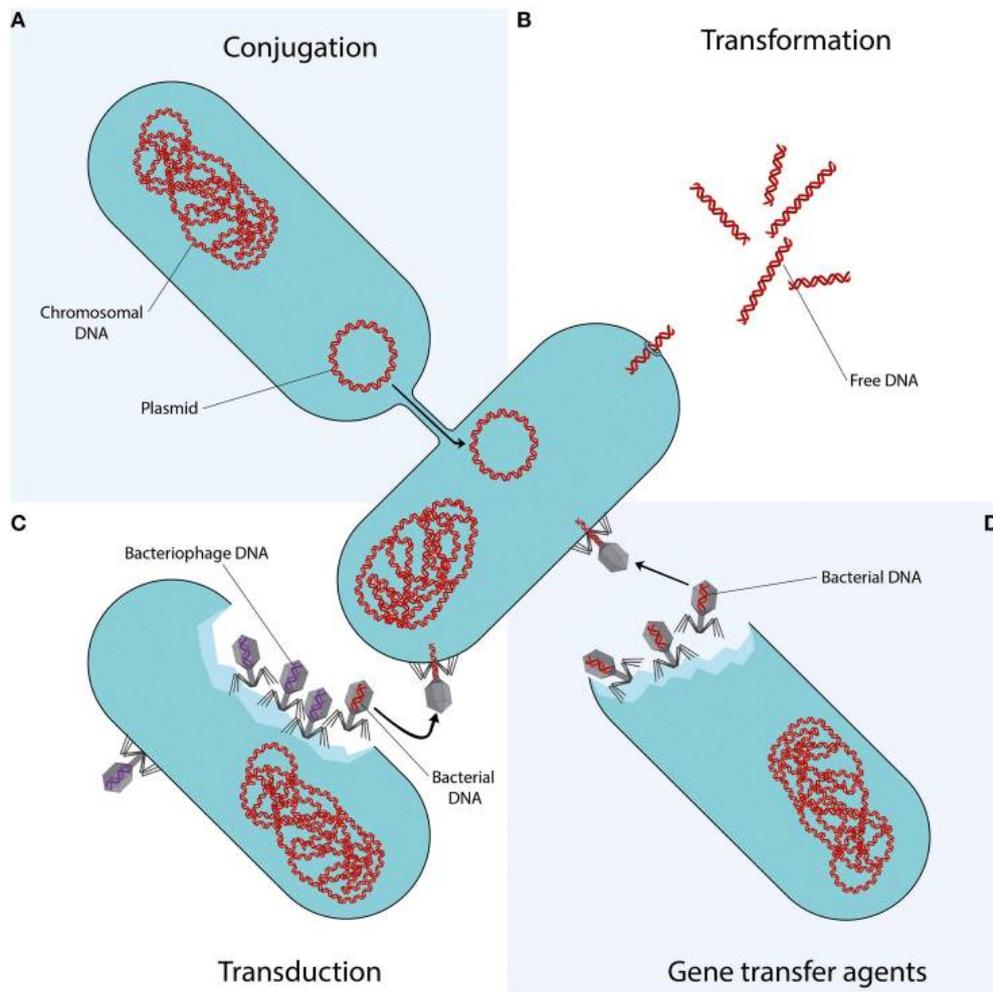
Figure 1.3. Mechanisms of horizontal gene transfer.

Methods by which bacteria acquire resistance via horizontal gene transfer. Image taken from (Von Wintersdorff *et al.,*2016).

random segments of DNA found in the host bacteria (Marrs, 1974; Von Wintersdorff *et al.*, 2016).

Finally, conjugation (Figure 1.3) is the direct transfer of DNA via cell-to-cell contact between bacteria. This DNA transfer occurs through conjugative machinery such as pili and the DNA transferred is often a plasmid or a transposon. Antibiotic resistance genes are often encoded on plasmids and transposons, making conjugation an important mechanism of spreading antibiotic resistance. Conjugation is considered the most likely mechanism of antibiotic resistance spread via horizontal gene transfer due to the fact that the DNA is more protected from the external environment than the DNA in transformation and the host range is larger than that of bacteriophage transduction (Norman, Hansen and Sorensen, 2009; Von Wintersdorff *et al.*, 2016).

Lastly, adaptive resistance is a temporary increase in the ability of a bacterium to survive a particular antibiotic assault or environmental niche resulting from alterations in gene or protein expression brought about by an environmental trigger. In the case of antimicrobial resistance the environmental trigger would be the presence of the antimicrobial itself. The intrinsic antimicrobial resistance mechanism in the bacteria that is triggered by the antimicrobial can include biofilm formation, efflux pump regulation, changes in cell morphology and permeability and antibiotic inactivation via enzymes (Schroeder, Brooks and Brooks, 2017).

1.2 Bacterial cell envelope

The bacterial cell envelope is a complex and multi-layered barrier which protect the bacterium against the external environment whilst also providing the cell with structure and rigidity. Most bacterial cell envelopes fall into one of two major categories, Gram-negative and Gram-positive, which have very different envelope architecture (Silhavy, Kahne and Walker, 2010).

1.2.1 Gram-negative cell envelopes

The Gram-negative cell envelope consists of three distinct layers; the outer membrane (OM), the peptidoglycan cell wall and the cytoplasmic or inner membrane (IM). Between the OM and the IM is the periplasm, an aqueous cellular compartment (Silhavy, Kahne and Walker, 2010).

The OM is the outermost layer of the Gram-negative cell envelope and is a lipid bilayer (Figure 1.4). The inner leaflet of the OM consists of phospholipids, whereas the outer leaflet contains glycolipids such as lipopolysaccharide (LPS). LPS is notorious for causing endotoxic shock common with septicaemia caused by Gram-negative pathogens. Therefore, the human innate immune system is highly sensitive to the presence of LPS. LPS is critical in the barrier function of the OM and consists of a glucosamine disaccharide attached to six or seven acyl chains, a polysaccharide core and an extended polysaccharide chain (O-antigen). Lipid A is a hydrophobic lipid component of LPS which anchors the LPS to the OM. The acyl chains of the LPS are predominantly saturated. This facilitates tight packing and provides an effective barrier against hydrophobic molecules. The OM also contains proteins, the majority of which form β-barrels and act as porins, allowing the passive diffusion of molecules such as polysaccharides and peptides across the OM. These OM porins limit the diffusion of hydrophilic molecules larger than 700 Da or so. This coupled with the nature of the LPS makes the OM a selective and effective permeability barrier. As a result, Gram-negative bacteria are generally more resistant to antibiotics than Gram-positive bacteria (Silhavy, Kahne and Walker, 2010).

The peptidoglycan is a rigid exoskeleton formed of repeating units of the disaccharide N-acetyl glucosamine-N-acetyl muramic acid. These repeating units are crosslinked by pentapeptide side chains. It is this peptidoglycan layer which gives bacteria their characteristic cell shape, such as rods in the case of *Escherichia coli*. In Gram-negative cells the peptidoglycan is located in the periplasm where it is anchored to the OM via the lipoprotein Lpp (Silhavy, Kahne and Walker, 2010).

The periplasm harbours the periplasmic binding proteins which function in the transport of solutes, such as sugars and amino acids, across the inner membrane (Silhavy, Kahne and Walker, 2010).

The IM of bacterial cells performs all of the membrane-associated functions of eukaryotic organelles such as energy-production, lipid biosynthesis, protein secretion and transport. The IM is formed of a phospholipid bilayer (Figure 1.4), in *E. coli* the major phospholipid components are phosphatidyl ethanolamine and phosphatidyl glycerol (Silhavy, Kahne and Walker, 2010).
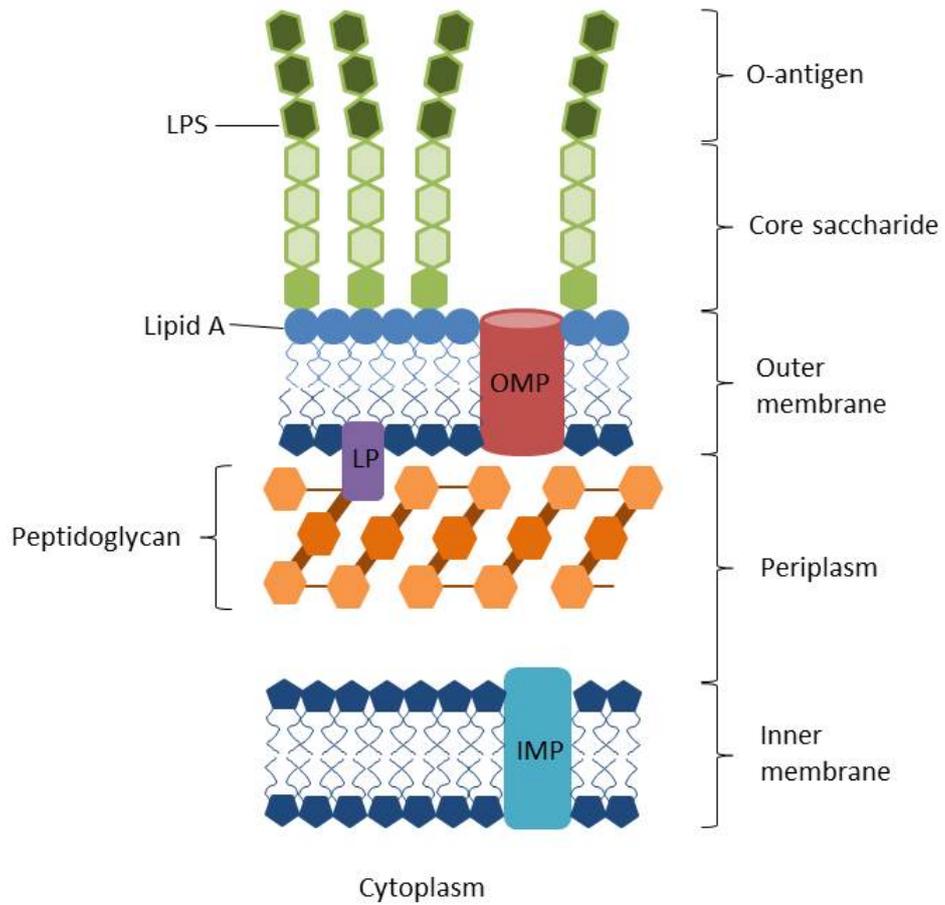
Figure 1.4. The structure of the Gram-negative bacterial cell envelope.

The Gram-negative cell envelope includes an outer membrane, a periplasm and a thin layer of peptidoglycan. LPS = lipopolysaccharide, OMP = outer membrane protein, LP = lipoprotein Lpp, IMP = inner membrane protein.

1.2.2 Gram-positive cell envelopes

The Gram-positive cell envelope structure differs in a number of key ways from the Gram-negative cell envelope structure. The most obvious difference is that Gram-positive bacteria do not contain an outer membrane. Due to the fact that the OM indirectly stabilises the IM in Gram-negative bacteria the peptidoglycan layer is relatively thin. To enable Gram-positive bacteria to withstand the turgor pressure, they are surrounded by a peptidoglycan layer that is many times thicker than that of Gram-negative bacteria (Figure 1.5). Long anionic polymers, called teichoic acids, are threaded through the peptidoglycan layers. Teichoic acids are comprised of glycerol phosphate, glucosyl phosphate, or ribitol phosphate repeats. The surfaces of the Gram-positive bacterial cell envelope can be decorated with proteins with functions analogous to those found in the periplasm of Gram-negative bacteria (Silhavy, Kahne and Walker, 2010).

Gram-positive bacterial peptidoglycan structure is similar to that of Gram-negative bacteria in that it is composed of a disaccharide repeat. The difference in the structure comes from the different crosslinks between the glycan strands, which can differ between bacterial species. Another major difference is that in Gram-negatives the peptidoglycan layer is a few nanometres thick whereas in Gram-positives the layer can be 30-100 nm thick and composed of many overlapping layers of peptidoglycan (Figure 1.5) (Silhavy, Kahne and Walker, 2010).

Teichoic acids are anionic polymers which fall into two groups, wall teichoic acids, which are coupled to peptidoglycan and lipoteichoic acids, which are anchored to the cell membrane (Figure 1.5). Due to their anionic character, they bind cations and contribute to cation homeostasis, this in turn influences the rigidity and porosity of the cell wall (Silhavy, Kahne and Walker, 2010).

1.3 Cationic antimicrobial peptides (CAMPs)

Innate immunity is the first line of defence for many organisms against invading pathogens. As many species have evolved they have maintained the innate immune system as a rapid broad-spectrum defence mechanism against pathogens. A major component of the innate immune system is cationic antimicrobial peptides (CAMPs), which in recent years have been highlighted as possible alternatives to antibiotics (Hancock and Lehrer, 1998). CAMPs are <100 residue amphipathic molecules with a variety of different antimicrobial effects and are produced by innate immune cells such as neutrophils and macrophages as well as
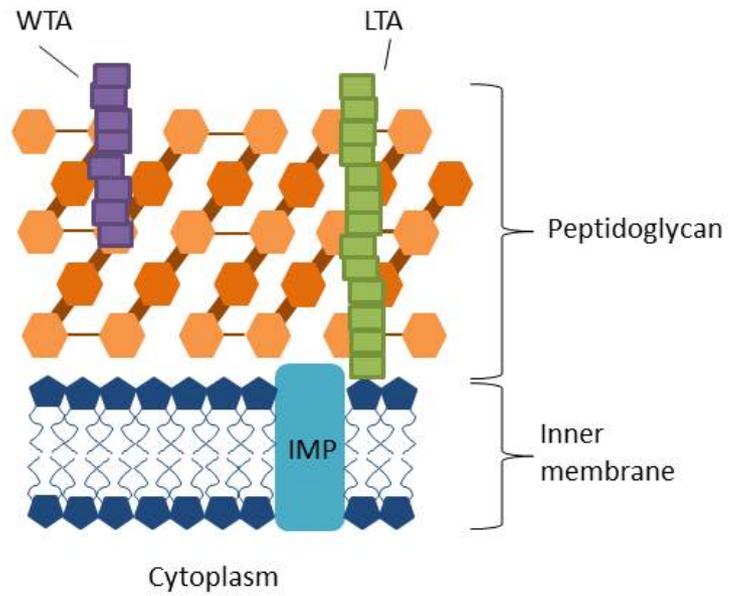
Figure 1.5. The structure of the Gram-positive bacterial cell envelope.

The Gram-positive cell envelope contains a thick layer of peptidoglycan and no outer membrane or periplasm. WTA = wall teichoic acid, LTA = Lipoteichoic acid.

being constitutively expressed by epithelial cells. Although they come in many different sizes and with a variety of secondary structures, the common feature of the CAMPs is that they are highly cationic. There are a number of different categories of CAMPs, including defensins, cathelicidins and histatins (Pasupuleti, Schmidtchen and Malmsten, 2012).

Defensins are long (18-45 residues) cationic peptides with three disulphide bonds linking six conserved cysteines. They were first discovered in human neutrophils (Ganz *et al.*, 1985a; Selsted *et al.*, 1985a) and have since been found in insects (Saito *et al.*, 1995), mammals (Ganz *et al.*, 1985b; Selsted *et al.*, 1985b) and plants (Thomma, Cammue and Thevissen, 2002). The antimicrobial activity of defensins is very broad-spectrum with activity against bacteria, fungi and enveloped viruses (Pasupuleti, Schmidtchen and Malmsten, 2012).

Cathelicidins are the second largest group of CAMPs and are identified by the far N-terminal region, the central conserved region and the variable C-terminal region (Zanetti and Gennaro, 1995). They are synthesised as pro-peptides and undergo cleavage to produce the active CAMP. The only currently known human cathelicidin is LL-37 (Zanetti and Gennaro, 1995). LL-37 (Figure 1.6A) is a 37 residue peptide which forms an α helix with a hydrophobic N-terminus which facilitates binding of LL-37 to the negatively charged hydrophobic bacterial membranes. LL-37 is able to bind both the bacterial membrane and lipopolysaccharide and has potent and widespread antimicrobial activity (Larrick *et al.*, 1995).

Histatins are histidine rich peptides ranging in size from 7 to 38 residues and are constitutively expressed in salivary glands of humans (Rijnkels *et al.*, 2003). Histatins cause bacterial membrane permeabilisation, and like cathelicidins and defensins, histatins have a broad spectrum of antimicrobial activity (Luque-Ortega *et al.*, 2008; Wiesner and Vilcinskas, 2010; Pasupuleti, Schmidtchen and Malmsten, 2012).

Melittin from bee venom is a linear 26 residue cytolytic peptide and spontaneously integrates into phospholipid bilayers. Melittin is composed of an amphiphilic α helix that has a bent rod shape (Figure 1.6B) (Terwilliger and Eisenberg, 1982). Melittin forms transmembrane pores in the phospholipid membrane of bacteria (Yang *et al.*, 2001).

The 32-residue peptide protamine is a small arginine-rich peptide with histone like properties. Protamine is synthesised in the late stages of spermatids and binds to DNA helping to condense the genome into a genetically inactive state (Balhorn, 2007).

A
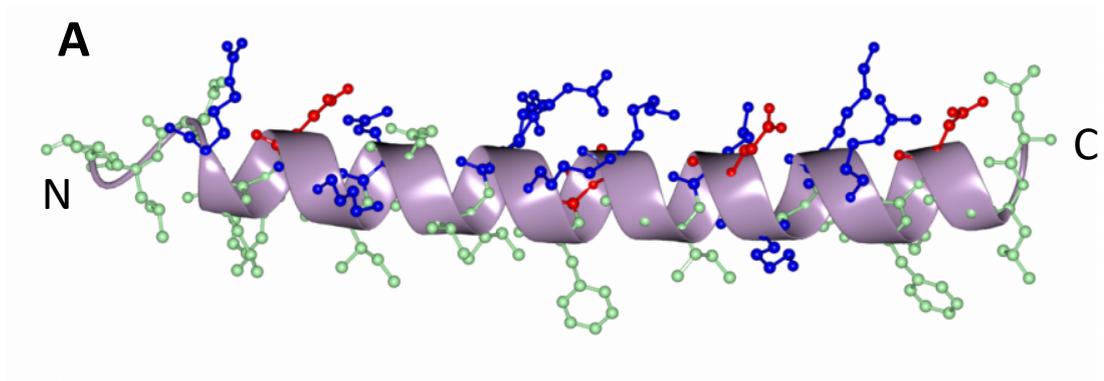
LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES

B
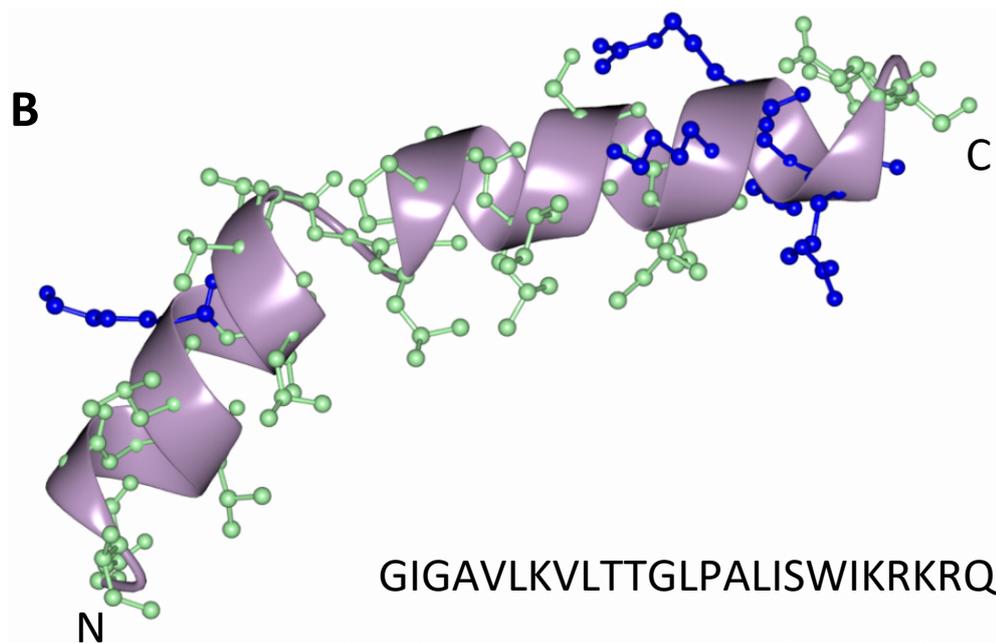
GIGAVLKVLTTGLPALISWIKRKRQQ

Figure 1.6. The structure of CAMPs LL-37 and melittin.

The structure of the CAMP LL-37 is shown in (A) and melittin is shown in (B). N and C refer to the N and C-terminus. Both are shown as a purple ribbon, with side chains displayed in ball and stick form. Green indicates an uncharged side chain, blue indicates a positively charged side chain and red indicates a negatively charged side chain. The bent helix shape of melittin can be clearly seen. Coordinates from PDB entry code 5NNM (A) and 2MLT (B).

Protamine, taken from salmon sperm, was used in early CAMP resistance studies as it showed preferential killing of *Salmonella*, unlike other compounds, and was inexpensive and easily available (Groisman *et al.*, 1992).

1.3.1 CAMP mode of action

The positive charge of CAMPs leads to their accumulation at the negatively charged membranes of both Gram-positive and Gram-negative bacteria, although the outer surfaces of these two classes of bacteria are substantially different. CAMPs appear to be able to pass easily through the porous peptidoglycan outer layer of Gram-positive bacteria (Koch, 1996; Malanovic and Lohner, 2016). Likewise, some CAMPs have the ability to cross the outer membrane of Gram-negative bacteria, however the mechanism is slightly different. To cross the outer membrane these CAMPs seem to use a charge-exchange mechanism involving Ca^{2+} and Mg^{2+} ions bound to the lipopolysaccharide, possibly aided by outer membrane protein binding according to the "self-promoted uptake hypothesis" (Anunthawan *et al.*, 2015; Malanovic and Lohner, 2016). This gives CAMPs access to the inner bacterial membrane.

Once CAMPs have crossed the outer barrier, either the outer membrane or the cell wall, they are able to exert their bactericidal effect. There are three accepted models for how CAMPs disrupt the membrane, barrel-stave, carpet or toroidal-pore. The method used by each CAMP can change depending on a variety of factors including membrane structure, aggregation, net charge and topology (Malanovic and Lohner, 2016).

A barrel-stave pore is much like a multi protein ion channel. A defined number of CAMPs are able to interact with each other in a specific way to create a pore structure (Figure 1.7). However, in the toroidal-pore system those specific CAMP interactions are not present, instead the CAMPs accumulate together and non-specifically interact with each other to disrupt the membrane curvature to create a toroid of high curvature forms (Figure 1.7). The carpeting model (Figure 1.7) is the unorganised association of CAMPs with the membrane. At high concentrations of CAMPs, they behave much like detergents forming micelles from the membrane (Wimley, 2010).

Although CAMP integration into the inner membrane of bacteria is a major cause of bacterial cell death, CAMPs are also able to stimulate the immune system which is an important feature for the hosts defence against invading pathogens. When local cells are stimulated by a microbial pathogen CAMPs are produced, this CAMP production directly
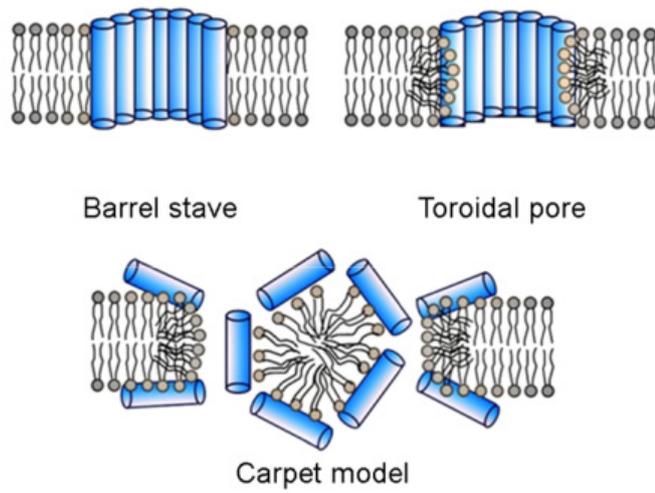
Figure 1.7. Mechanisms of CAMP action.

The lipid bilayer of the membrane is shown in grey and CAMPs are shown in blue, the different modes of CAMP action can be clearly seen in each diagram. Image adapted from (Lee and Park, 2014).

recruits leukocytes to the location and can stimulate the production of chemokines and cytokines. Production of chemokines and cytokines such as IL-8 and IFN-α recruits immune cells like dendritic cells and T cells to the site of microbial invasion. This immune cell recruitment helps to clear the infection from the host quickly. CAMPs are also able to promote wound healing, helping to prevent further infection of the affected area (Lai and Gallo, 2009).

1.3.2 Polymyxins

Although polymyxin B and polymyxin E (also known as colistin) are not thought of as traditional CAMPs, they have a similar morphology and mode of action as CAMPs. This is a result of the large number of amino groups in the molecules giving rise to a net positive charge at physiological pH. This net positive charge attracts polymyxin B and colistin to the negatively charged bacterial membrane. Initially both polymyxin B and colistin appear to target the outer membrane of Gram-negative bacteria, however at high enough concentrations (above 20 μg/ml) it has been shown that polymyxin B has the ability to depolarise the inner membrane of Gram-negative bacteria by forming pores in the membrane (Hancock, 1997; Daugelavičius, Bakienė and Bamford, 2000). This inner membrane pore forming ability of polymyxin B closely resembles the mode of action of CAMPs. Polymyxin B was isolated from *Bacillus polymyxa* and is a lipopeptide antibiotic. Its structure (Figure 1.8) consists of a cationic peptide ring and a tripeptide side chain with a fatty acid tail (Hancock, 1997). Polymyxin B and colistin differ in a single amino acid (Figure 1.8) (Li *et al.*, 2006).

1.3.3 Resistance to CAMPs

Resistance to CAMPs can be either intrinsic or acquired. Intrinsic resistance can occur via passive or inducible mechanisms. Passive intrinsic resistance commonly refers to modifications certain bacterial species have to their lipid A which makes it more positively charged. The more positively charged lipid A is, the less attractive it is to the positively charged CAMPs and therefore the less CAMP binding occurs (Manniello, Heymann and Adair, 1978; Viljanen and Vaara, 1984). The transient molecular modifications of bacteria in response to CAMPs is called inducible resistance, the reversibility of these modifications offsets the energetic burden of the changes. Again, these changes often involve components of the membrane, including incorporation of positively charged molecules to reduce the interaction of CAMPs with the bacterial cell surface (Andersson, Hughes and Kubicek-Sutherland, 2016).
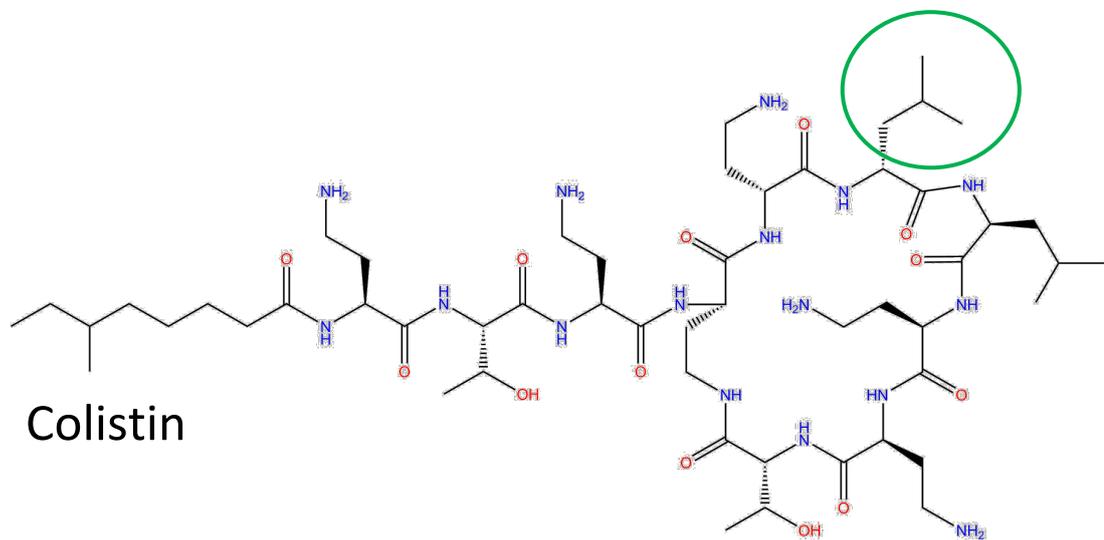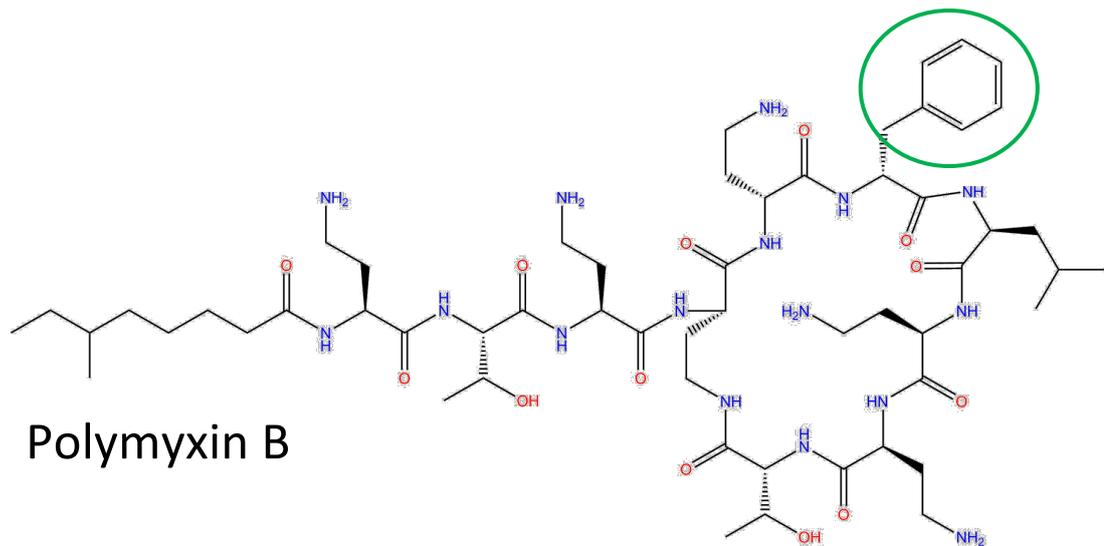
Figure 1.8. Chemical structure of polymyxin B and polymyxin E (colistin).

The amino acid residue that differs between the two structures is highlighted by the green circle, Phenylalanine for polymyxin B and Leucine for colistin. Image adapted from (Jerke, Lee and Humphries, 2016).

Although membrane modification is the most common defence mechanism against CAMPs, there are other defence strategies employed by bacteria, these include efflux of CAMPs and CAMP proteolytic degradation. The membrane protein OmpT from *E. coli* has been shown to degrade both LL-37 and the positively charged peptide protamine. Due to the fact that the active site of OmpT is extracellular these peptides would be degraded before they could integrate into the cell membrane therefore preventing membrane disruption (Stumpe *et al.*, 1998; Thomassin *et al.*, 2012). An ATP-binding cassette (ABC) transporter dependent efflux pump that acts on CAMPs is encoded by the chromosomal gene *vraFG* in *S. aureus* (Min *et al.*, 2007). Efflux pump systems that act on CAMPs have also been identified in many other bacterial species including *N. gonorrhoeae* (Shafer *et al.*, 1998) and *Yersinia* spp. (Bengoechea and Skurnik, 2000).

Interestingly, there have also been reports of ABC transporters that actively import CAMPs into the cytoplasm of bacterial cells. Although this seems counter intuitive, by importing the CAMPs into the cytoplasm, these ABC transporters prevent the CAMPs integrating into the cell membrane. These importing ABC transporters are named Sap (sensitive to antimicrobial peptides) and Yej (Groisman *et al.*, 1992; Parra-Lopez, Baer and Groisman, 1993; Eswarappa *et al.*, 2008).

1.4 ATP-binding cassette (ABC) transporters

Transport of molecules across cell membranes is critical for all organisms as it allows the cells to acquire nutrients required for survival and cell growth. In the case of active transport, transporters also allow the cells to maintain a concentration gradient across the membrane. It is possible for molecules such as oxygen, carbon dioxide and water to simply diffuse through the cellular membrane as they are small uncharged molecules. However, larger and charged molecules are actively transported across the membrane via primary or secondary transport. Primary transporters are defined as transporters that require the breakdown of the high-energy molecule ATP to catalyse the transport reaction. Secondary transport is driven by the electrochemical potential difference across the cell membrane created by pumping ions in/out of the cell. Allowing an ion to move down the electrochemical gradient increases entropy and can serve as an energy source for transporting a desired molecule. In bacteria hydrogen is the ion commonly used in secondary transport. Secondary transporters can be categorised into antiporters, where the molecules move in opposite directions across the membrane, or symporters where the

molecules move in the same direction across the membrane (Lodish H, Berk A, Zipursky SL, 2000).

ABC transporters are primary transporters and use ATP to transport their substrates across the membrane. ABC transporters have a very diverse range of substrates including; ions, sugars, amino acids, vitamins, drugs, peptides and lipid molecules (Davidson *et al.*, 2008).

1.4.1 ABC transporter structure

ABC transporters can be classified as exporters, importers or energy-coupling factors, with importers further classified as either type I or type II transporters depending on their structure and mechanism (ter Beek, Guskov and Slotboom, 2014; Wilkens, 2015a). A typical ABC transporter contains two nucleotide binding domains (NBDs) and two transmembrane domains (TMDs) (Figure 1.9). Bacterial ABC importers also have a substrate binding domain/protein (SBD/P) which delivers the substrate for transport to the transmembrane domains. In Gram-negative bacteria the SBP is free to move around the periplasm, in Gram-positive bacteria however the SBP is lipid anchored to the membrane. In bacteria it is also possible to have fused, homodimeric and heterodimeric NBDs and TMDs (Wilkens, 2015b).

The NBDs of ABC transporters are cytosolic and contain many highly conserved motifs. NBDs comprise two subdomains, one similar to the RecA protein and containing the Walker A motif and the other named the helical domain and containing the LSGGQ motif. In the transporter the NBDs are arranged so that the ATP binding interfaces are facing each other in a head to tail orientation (Figure 1.10). This arrangement allows for the binding and hydrolysis of two ATPs. Upon ATP binding the two NBDs move closer via the coupling helices, pulling the lower sections of the TMDs together and therefore changing the TMD conformation from inward facing to outward facing allowing the transport of a substrate molecule to the cytoplasm (Figure 1.11) (Locher, 2009).

All of the genomes sequenced to date encode ABC exporters. Although for several ABC exporters the physiological substrate is unknown many of them are involved in extrusion of toxic substances such as drugs. All ABC exporters share a common core architecture of 12 transmembrane helices which extend below the membrane and into the cytoplasm of the cell, often the TMDs are fused to the NBDs (Figure 1.9) (Locher, 2009).

Type I ABC importers typically take up small molecules such as ions, sugars and amino acids and usually contain 12 transmembrane helices (Hollenstein, Frei and Locher, 2007).
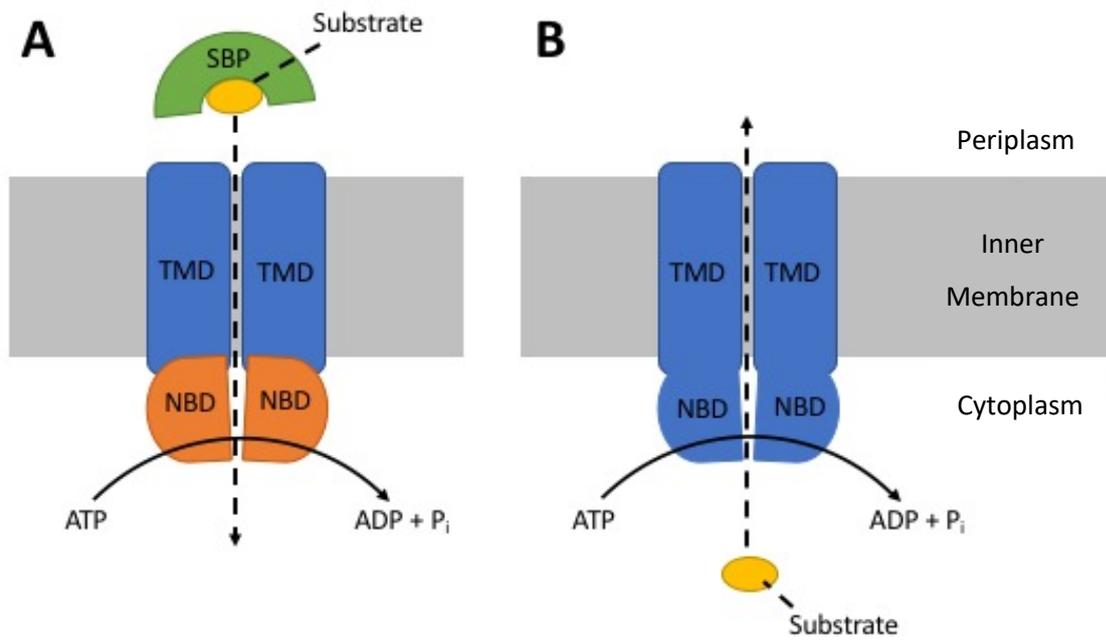
Figure 1.9. Structure of ABC Transporters.

(A) A typical ABC importer structure which includes a substrate binding protein (SBP), and (B) a typical ABC exporter which does not. TMD = transmembrane domain, NBD = nucleotide binding domain.
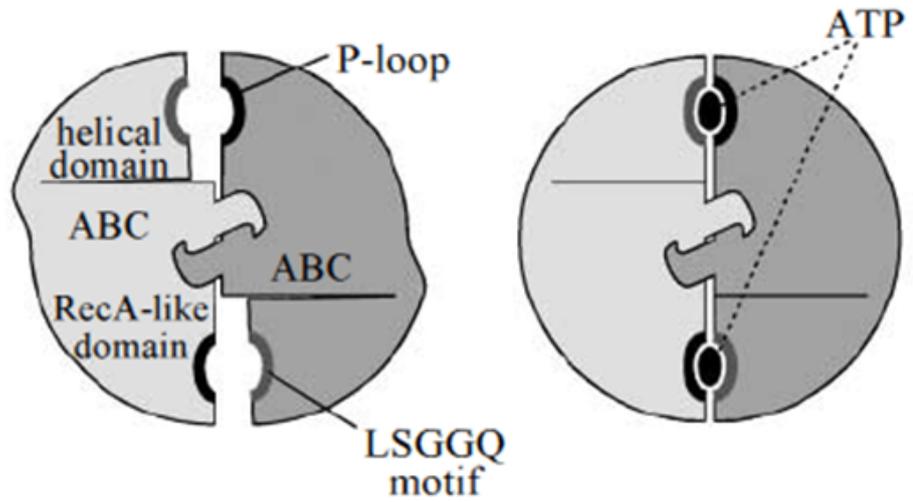
Figure 1.10. NBDs of ABC transporters have a head to tail dimer orientation.

Top down view of NBDs. The two nucleotide binding domains of ABC transporters are orientated in a head to tail orientation, as the helical domain contains the LSGGQ motif and the RecA like domain contains the P-loop, to create two identical ATP binding sites (Locher, 2009).
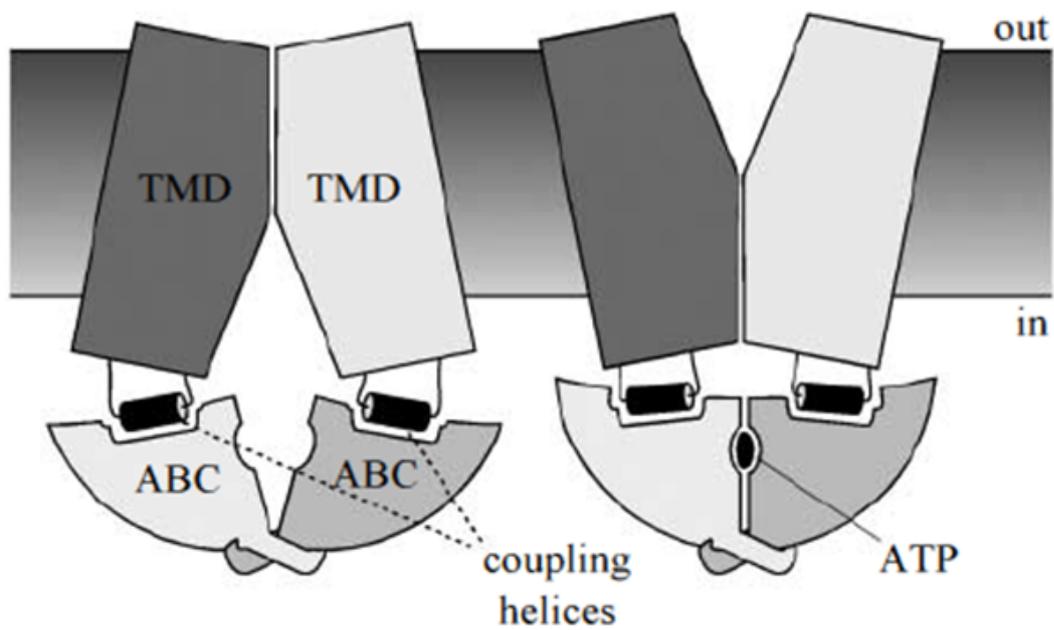
Figure 1.11. Upon ATP binding, NBDs change the orientation of TMDs of ABC transporters.

The coupling helices of the transmembrane domains (TMDs) interact with the nucleotide binding domains (NBDs), so that when the NBDs change conformation upon ATP binding the TMDs change from the inward facing to the outward facing conformation. The second ATP binding site in this image is eclipsed by the first site (Locher, 2009).

However the type II importers which transport larger molecules including haem and vitamins typically have 20 transmembrane helices (Locher, Lee and Rees, 2002). It should be noted at this point that type I importers appear to have a transient binding pocket in the TMDs. This binding pocket encourages substrate release from the SBP when it binds to the TMD and the lobes of the SBP are opened, it also helps prevent the substrate diffusing out of the TMDs before transport is affected. Upon ATP binding in the NBDs and the swivelling of the TMDs from the outward facing to the inward facing conformation the transient binding pocket is occluded and the substrate is pushed into the cytoplasm (Oldham *et al.*, 2007). In type II importers the TMDs seem to be inert with little or no affinity for their substrates (Locher, 2009).

Although type I importers have a transient binding site in the TMDs, the SBPs of both type I and type II importers have very high specificity for their substrates. Although there is considerable variation in the size of SBPs (25-70 kDa) with little shared sequence similarity, the overall fold of SBPs is well conserved. SBPs consist of two domains with the binding pocket for the substrate lying at the interface of the domains. Binding occurs in a Venus fly trap like manner accompanied by large conformational changes. Upon binding, the substrate is completely buried in a cavity between the two domains of the SBP (Locher, 2009). Berntsson *et al.*, 2010 classified SBPs into 6 structural categories (Cluster A-F) (Figure 1.12).

Cluster C SBPs bind a range of substrates including di and oligopeptides, nickel and polysaccharides such as cellobiose. One of the defining features of this group is their larger size and possession of an extra domain which is thought to extend the binding pocket to allow room for larger substrates (Berntsson *et al.*, 2010). This larger binding cavity, extra domain, and for some members of the cluster their preference for oligopeptides, may have significance for how they bind and transport CAMPs.

Energy-coupling factors (ECF) are a type of ABC transporter widely used by prokaryotes to take up micronutrients and can be either importers or exporters. Every ECF transporter contains two cytosolic NBDs, a membrane embedded substrate binding domain (EcfS) and a transmembrane energy coupling protein (EcfT) that links the NBDs and EcfS (Figure 1.13). The EcfS protein is the substrate-binding component of the transporter and determines the specificity of the transporter. These proteins are able to bind a wide range of small molecules such as vitamins, amino acids and metals. From information gathered from crystal structures, a working hypothesis has been put together to explain how substrates
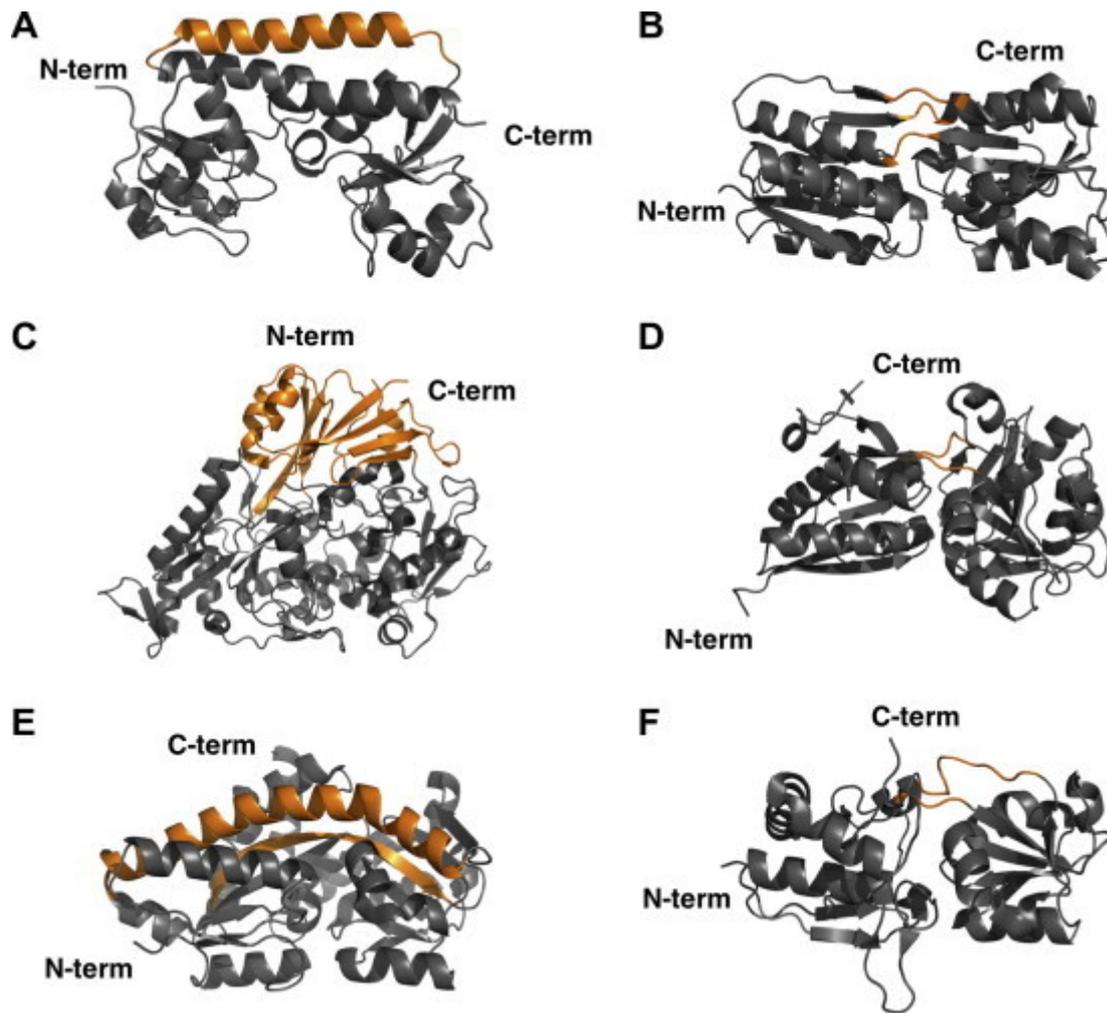
Figure 1.12. Structural classifications of SBPs.

Substrate binding proteins (SBPs) have been classified in Clusters A-F, as indicated by the lettering in the figure, based on structure. The structural differences which define a cluster are shown in orange. Cluster C SBPs have a clearly defined extra domain. Image taken from (Berntsson *et al.*, 2010).
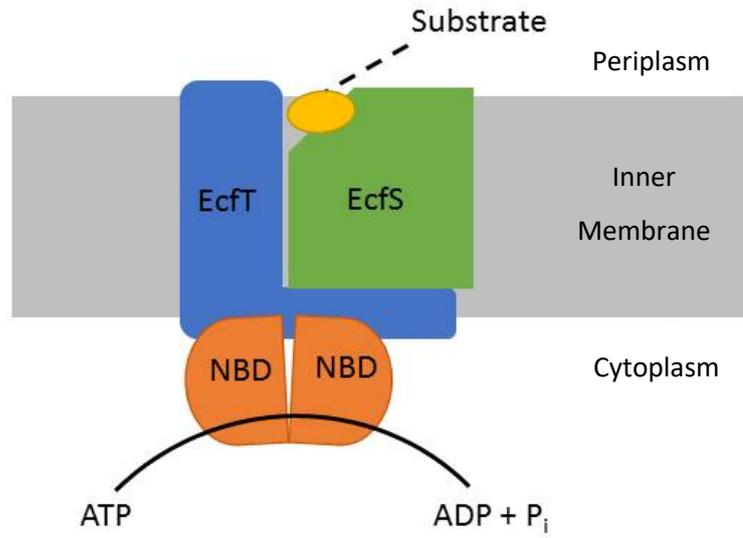
Figure 1.13. Schematic diagram of an energy-coupling factor (ECF) transporter.

ECF transporters consist of two nucleotide binding domains, a membrane-embedded substrate binding protein (EcfS) and a transmembrane energy coupling factor (EcfT).

are transported across the membrane via ECFs. In this hypothesis upon substrate binding, the EcfS protein topples over in the membrane to transport the substrate across the membrane. This process is fuelled by ATP binding and hydrolysis. EcfS is horizontally orientated in the membrane when the NBDs are in the open conformation, upon ATP binding the NBDs come together and push the EcfS into a conformation where the substrate binding site is accessible to substrates. ATP hydrolysis allows the NBDs to move apart and the EcfS to topple back over in the membrane and release the substrate to the opposite side of the membrane (Figure 1.14) (Zhang, 2013).

1.5 ABC transporters and CAMP resistance

1.5.1 ABC transporters and two-component regulatory systems in peptide antibiotic resistance

Transporters can act as co-sensors for signal transduction pathways in bacteria. These transporters can interfere with signal transduction pathways by transporting effector molecules into the cytoplasm or interacting directly with sensory components. Direct interaction with sensory components involves interaction between a membrane bound sensor domain, which binds specific substrates, and the signalling domain, which transfers the signal information into the cytoplasm of the cell. Commonly in signal transduction pathways the transporters act as inhibitors of the respective pathways, however in the case of antimicrobial peptide resistance systems the transporter is required for activation of signalling and the system remains inactive without the transporter. These antimicrobial resistance systems are widespread in Gram-positive bacteria and consist of a two-component regulatory system (TCS), where the histidine kinase lacks an obvious input domain, and an ABC transporter with an unusual 10 transmembrane helical structure and large extracellular domain. All examples of this kind of system characterised to date involve resistance to peptide antibiotics and the ABC transporter is a key component of the system (Dintner *et al.*, 2014).

An example of such a TCS and ABC transporter antimicrobial peptide resistance system is the BceRS-BceAB system in *B. subtilis* which confers resistance to the peptide antibiotic bacitracin, an inhibitor of cell wall synthesis. Bacitracin is very similar to polymyxins in that it is a nonribosomally synthesised cyclic peptide antibiotic. BceS, the histidine kinase, is unable to detect bacitracin in the absence of the ABC transporter BceAB. This led to the assumption that the BceAB transporter is the sensory domain, however the mechanism by which the transporter and the TCS (BceRS) communicate is unknown. Experimental
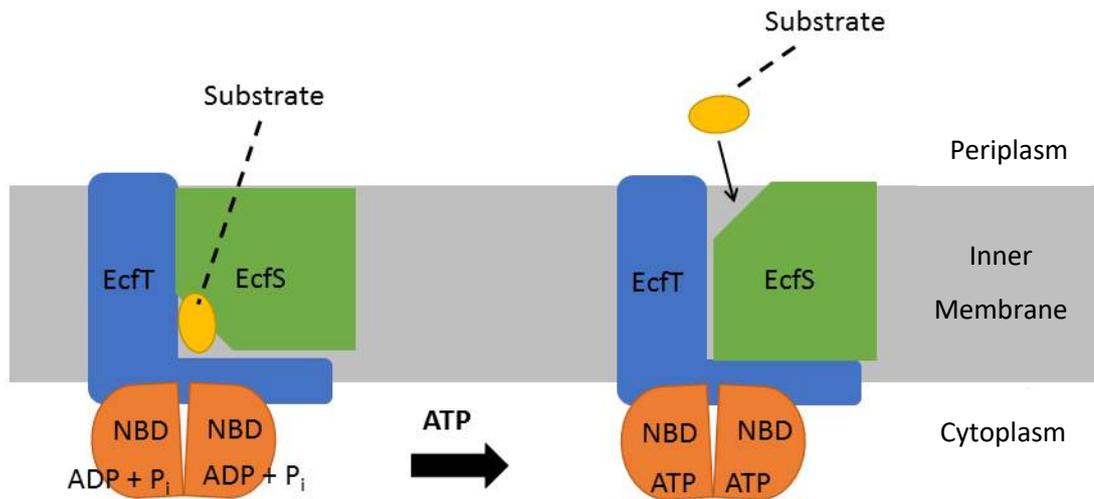
Figure 1.14. Schematic of the mechanism of transport of ECF transporters.

EcfS is horizontal in the membrane when the nucleotide binding domains (NBDs) are in the open conformation, upon ATP binding the NBDs come together and push the EcfS into a conformation where the substrate binding site is accessible to substrates. Upon ATP hydrolysis the EcfS topples back over in the membrane and releases the substrate to the opposite side of the membrane.

evidence indicates that resistance to bacitracin is via translocation of the antibiotic through the BceAB transporter to prevent contact with the cell wall synthesis machinery (Dintner *et al.*, 2014).

From the experimental evidence gathered a working hypothesis of the system has been produced in which bacitracin is bound directly by the transporter BceAB (Figure 1.15). A sensory complex is then formed in the membrane via the interaction between BceAB and BceS. When ATP is hydrolysed by the BceAB transporter, the activation of BceS is triggered, this then leads to phosphorylation of BceR. The target promoter, P_{bceA}, is activated by BceR which induces increased production of BceAB to ensure resistance. It is not known whether BceR interacts with the complex of BceAB and BceS (Dintner *et al.*, 2014).

1.5.2 Sap ABC transporter

The Sap transporter was originally discovered in *S*. Typhimurium via transposon mutagenesis along with a number of other genes which conferred resistance to protamine, however the *sap* operon gave the strongest phenotype and so was studied further (Groisman *et al.*, 1992). The *sapABCDF* operon encodes the Sap (sensitive to antimicrobial peptides) transporter which consists of an SBP (SapA), two TMDs (SapB and SapC) and two NBDs (SapD and SapF) (Figure 1.16) (Parra-Lopez, Baer and Groisman, 1993). The regulation of the *sap* operon is not currently understood.

1.5.3 The Sap transporter and CAMP resistance

Salmonella enterica serovar Typhimurium is a Gram-negative bacterium that causes a systemic disease in mice similar to typhoid in humans. *S*. Typhimurium is a key causative agent of food poisoning in humans and is a facultative intracellular pathogen that can resist CAMP attack in macrophages (Parra-Lopez, Baer and Groisman, 1993).

To identify genes involved in CAMP resistance, a MudJ transposon library was created in *S*. Typhimurium and screened for sensitivity to protamine, a cationic peptide. A number of transposon mutants were isolated that were more susceptible to protamine, including some which mapped to the *sap* operon (Groisman *et al.*, 1992).

The components of the Sap transporter were tested to determine their role in resistance to protamine. All components of the Sap transporter were required for resistance to protamine with the exception of SapA. A Δ*sapA* deletion mutant was more susceptible to protamine than the wild type strain, although it was able to survive concentrations of

46
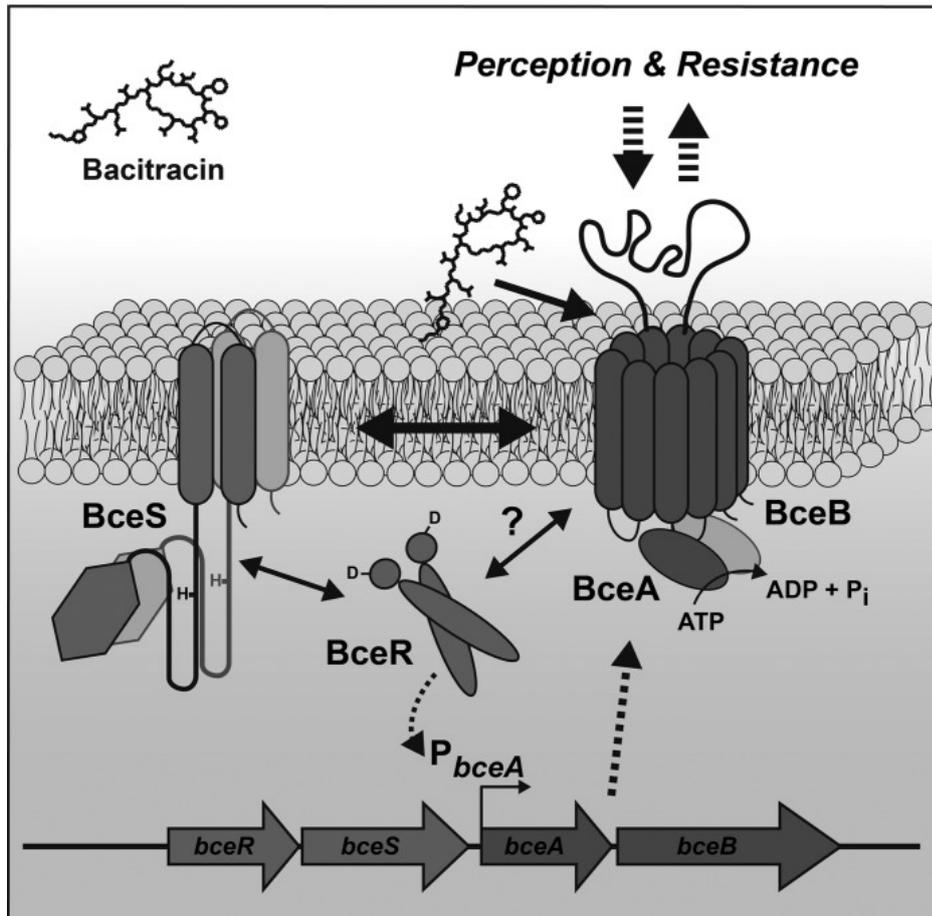
Figure 1.15. BceRS-BceAB system from *Bacillus subtilis* confers resistance against the antibiotic bacitracin.

Double headed arrows indicate interactions between proteins, dotted arrows indicate transcription events, a question mark indicates the possible interaction of BceR with the sensory complex of BceAB and BceS. Bacitracin is bound by the ABC transporter BceAB. A sensory complex is formed between BceAB and BceS. ATP hydrolysis by BceAB activates BceS which phosphorylates BceR. BceR activates the promoter P_{bceA} which increases production of BceAB. Image taken from (Dintner *et al.*, 2014).
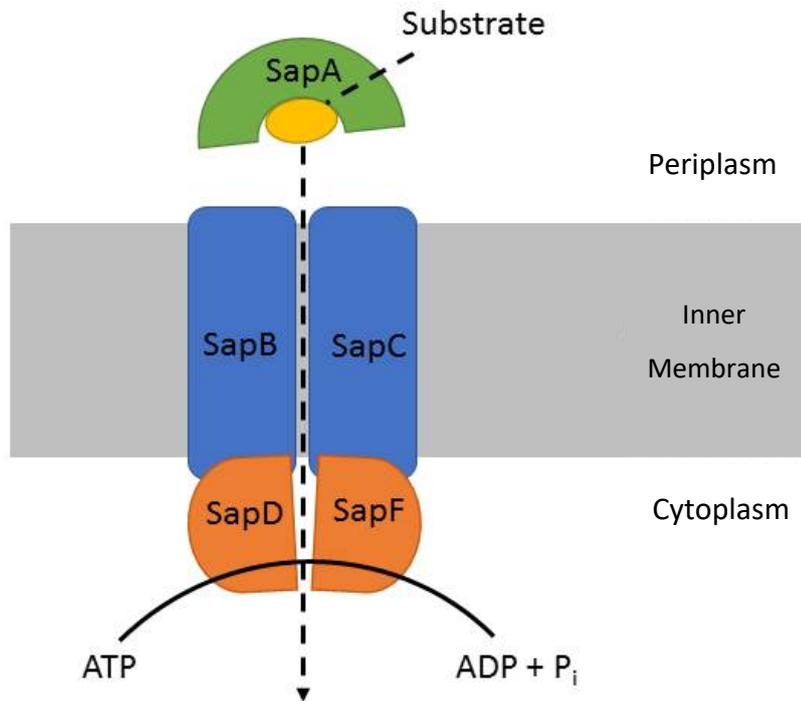
Figure 1.16. Schematic diagram of the Sap transporter.

SapA is the substrate binding protein, SapB and SapC are the transmembrane domains and SapD and SapF are the nucleotide binding domains.

protamine that prevented growth of the Δ*sapC* and Δ*sapD* mutant strains (Parra-Lopez, Baer and Groisman, 1993).

It has also been shown that the *sap* operon is upregulated in a chinchilla model of non-typeable *Haemophilus influenzae*-induced otitis media. A non-typeable *Haemophilus influenzae* (NTHI) reporter strain was created with a bioluminescent reporter that was driven by the *sap* promoter. This allowed the expression of the *sap* operon to be localised during NTHI infection of the chinchilla. The data showed that the *sap* operon was transiently expressed in the middle ear, eustachian tube, nasopharynx, and the oropharynx of the chinchilla. A non-polar Δ*sapA* mutant of NTHI was then created and tested against chinchilla β-defensin 1, a CAMP. The Δ*sapA* mutant was eight-fold more sensitive to chinchilla β-defensin 1 than the wild-type. This mutation also significantly impaired the strain's ability to colonise and survive in the middle ear and nasopharynx of the chinchilla. In competition with the wild type strain in the middle ear, the loss of the SapA protein allowed the wild type strain to outcompete the mutant (Mason, Munson and Bakaletz, 2005). Further work demonstrated that *sap* promoter activity was specifically upregulated in the presence of chinchilla β-defensin 1. This led to the hypothesis that SapA, the SBP of the Sap transporter, was involved in the recognition of, and defence against, chinchilla β-defensin 1. To test this hypothesis western immunoblot analysis and immunoprecipitation were used and showed a specific association of NTHI SapA with chinchilla β-defensin 1. This was the first demonstration of a direct interaction between a CAMP and SapA (Mason *et al.*, 2006).

A Δ*sapD* mutant was also created and inoculated into the nares and middle ears of chinchillas. Colonisation of the nasopharynx was impaired in the Δ*sapD* mutant as compared to the wild type. Complementation of the Δ*sapD* mutation restored colonisation to wild type levels. Competitive fitness was monitored for the Δ*sapD* mutant and wild type strains in the nasopharynx. The Δ*sapD* mutant strain was unable to compete with the wild type and was significantly reduced 2 days post inoculation and completely cleared 4 days after inoculation. Complementation of the Δ*sapD* mutation restored competitive colonisation to levels similar to the wild type. As with the nasopharynx, the Δ*sapD* mutant was unable to survive in the middle of ear of the chinchilla. The Δ*sapD* mutant was then tested for sensitivity to CAMPs and found to be more sensitive to chinchilla β-defensin 1, human β-defensin 3 and LL-37 than the wild type strain. These data clearly show a role for

SapD in protection against CAMPs and a requirement for SapD in the survival of NTHI in the chinchilla middle ear and nasopharynx (Mason *et al.*, 2006).

A NTHI Δ*sapBCDF* mutant caused periplasmic accumulation of CAMPs implying that SapBCDF is required for transport of CAMPs to the cytoplasm. Due to CAMPs being susceptible to cytoplasmic peptidase activity, it is hypothesised that when SapA binds CAMPs the SapBCDF transporter transports the CAMPs to the cytoplasm where peptidases degrade them to prevent their periplasmic accumulation. This leads to protection against membrane disruption caused by CAMPs (Shelton *et al.*, 2011).

Other studies have indicated that there may be a role for SapD in other cellular functions unrelated to the transport of substrates. For example studies with *Pasteurella* show that *sapD* expression is upregulated in the presence of different iron sources (Paustian *et al.*, 2002). SapD has been shown to mediate potassium uptake in *E. coli* by energising the TrkG/TrkH potassium uptake system (Harms *et al.*, 2001). The NTHI Δ*sapD* mutant was analysed for potassium uptake in different media, a 1000-fold increase in extracellular potassium was required to support minimal growth when compared with either the wild type or complemented *sapD* mutant strain (Mason *et al.*, 2006). SapD has also been implicated in polymyxin B resistance in *Proteus* species via LPS modifications. A transposon mutagenesis library was created and polymyxin B hypersensitive mutants were identified. *sapD* was one of those genes and LPS defects were found in the *sapD* mutant. It is possible that SapD indirectly affects the biosynthesis or modification of LPS in the inner membrane (McCoy *et al.*, 2001).

1.5.4 The Sap transporter and putrescine export

Polyamines, such as putrescine, contain two or more amino groups and have an important role as growth factors in animals, plants and bacteria. In the intestinal tract polyamines can impact the health of the animals either positively or negatively depending on their concentration. The polyamine concentration is regulated by the uptake and export systems of the intestinal tract bacteria. However, the mechanism of export of putrescine by these intestinal tract bacteria was unknown (Sugiyama, Nakamura and Matsumoto, 2016).

Keio collection *E. coli* strains with gene deletions in annotated transport systems were screened for lower concentrations of putrescine in the supernatant as compared to the

wild type. Strains with lower concentrations of putrescine in the supernatant than the wild type were hypothesised to contain a mutation in the genes responsible for putrescine export. Strains with the gene deletions of *sapF* and *sapD* had the lowest concentrations of putrescine in the supernatant. The other genes in the *sap* operon were tested and *sapB* and *sapC* mutants were also shown to be associated with slightly decreased putrescine supernatant concentrations, however a *sapA* mutation was shown to have no effect on the concentration of putrescine in the supernatant. This implicates the Sap transporter in putrescine export although SapA is not required (Sugiyama, Nakamura and Matsumoto, 2016).

Isotopically labelled arginine was used to demonstrate putrescine export by SapBCDF. The labelled arginine is internalised by an arginine transporter and then converted to isotope labelled putrescine. The concentration of labelled putrescine can then be measured in the supernatant from strains in which the SapBCDF transporter is present or absent. In the *sapBCDF* deleted strain there was a 62% decrease in putrescine in the supernatant as compared to the parent strain. When *sapBCDF* was complemented back into the strain the putrescine concentration in the supernatant was restored to 77% of the parental strain (Sugiyama, Nakamura and Matsumoto, 2016).

sapBCDF was also tested for resistance to LL-37. *E. coli* MG1655 and the mutant strain Δ*sapBCDF* were exposed to LL-37. Susceptibility to LL-37 was not significantly different in either strain. This result indicates that SapBCDF is not involved in resistance to LL-37 (Sugiyama, Nakamura and Matsumoto, 2016).

1.5.5 Yej ABC transporter

The Yej transporter, encoded by the *yejABEF* operon, consists of an SBP (YejA), two transmembrane domains (YejB and YejE) and a single NBD protein (YejF) (Figure 1.17). Unusually, YejF does not form a homodimer to create the NBDs of the Yej transporter, instead it appears that a single polypeptide chain contains both of the NBDs required to power the transporter (Eswarappa *et al.*, 2008). The *yejABEF* operon is regulated by a small non-coding RNA, RydC. RydC forms a complex with the RNA-binding protein Hfq and then interacts with the *yejABEF* mRNA. Impairing the expression of RydC reduces the amount of *yejABEF* mRNA, therefore endogenous expression of RydC in cells increases the amount of *yejABEF* mRNA. However, overexpression of *rydC* results in the degradation of the *yejABEF* mRNA, probably due to destabilisation (Antal *et al.*, 2005). Hfq is a global regulator of gene
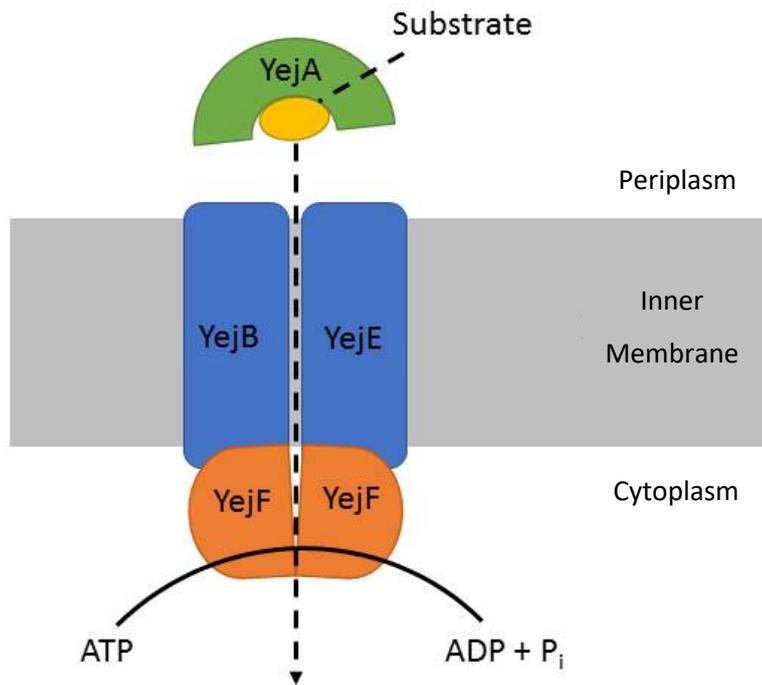
Figure 1.17. Schematic diagram of the Yej Transporter.

YejA is the substrate binding protein, YejB and YejE are the transmembrane domains and a single YejF protein forms the nucleotide binding domains.

expression in bacteria and mediates the interaction of small regulatory RNAs with mRNA (Sobrero and Valverde, 2012).

1.5.6 The Yej transporter and CAMP resistance

The role of the *yejABEF* operon in virulence was investigated in *Salmonella*. A macrophage cell line, J774A.1, and an epithelial cell line, Intestine 407, were infected with bacteria. After 12 hours the bacteria were recovered, the RNA extracted and reverse transcription carried out. The results showed that the *yejABEF* operon was upregulated seven-fold in both J774A.1 and Intestine 407 cells. This highlights the importance of the *yej* operon in the immune and epithelial cells that *Salmonella* first encounters on its course of infection (Eswarappa *et al.*, 2008).

YejF, the NBD of the Yej transporter, was deleted via the Lambda Red recombinase system. The resulting *S.* Typhimurium Δ*yejF* mutant was tested for susceptibility to different CAMPs (protamine, polymyxin B and melittin). It was found that the Δ*yejF* mutant was more susceptible to protamine, polymyxin B and melittin than the wild type strain (*S.* Typhimurium). *sap* mutants in *S.* Typhimuirum have also been shown to be sensitive to protamine (Parra-Lopez, Baer and Groisman, 1993). Therefore, the sensitivity of Δ*yejF*, Δ*sap* and Δ*yejF*Δ*sap S.* Typhimurium mutants were tested with protamine and compared. The Δ*yejF*, Δ*sap* and Δ*yejF*Δ*sap* strains were equally sensitive to 40 µg/ml protamine. Unusually the Δ*yejF*Δ*sap* double mutant was not more susceptible to protamine than the other mutants. This may be due to other genes playing a role in counteracting CAMPs when both Sap and Yej are non-functional. Δ*yejB* and Δ*yejE* strains were also tested with polymyxin B and protamine, and both showed susceptibility. However, when the Δ*yejA* mutant was tested, the strain was not more susceptible to polymyxin B, protamine or melittin, suggesting that YejA is not required for defence against CAMPs (Eswarappa *et al.*, 2008).

The morphology of the different *S.* Typhimuirum strains was investigated. It is clear from the scanning electron micrographs that the Δ*yejF* strain has several membrane irregularities. Many of the bacteria show membrane damage, extruded cytoplasms and other features associated with cell lysis (Figure 1.18) (Eswarappa *et al.*, 2008).

Like *Salmonella*, *Brucella* are intracellular pathogens. The bacteria cause severe febrile illness in humans and have evolved to exist in host macrophages in the presence of a series of stress factors including CAMPs. *Brucella melitensis* NI, the *Brucella* species used in this
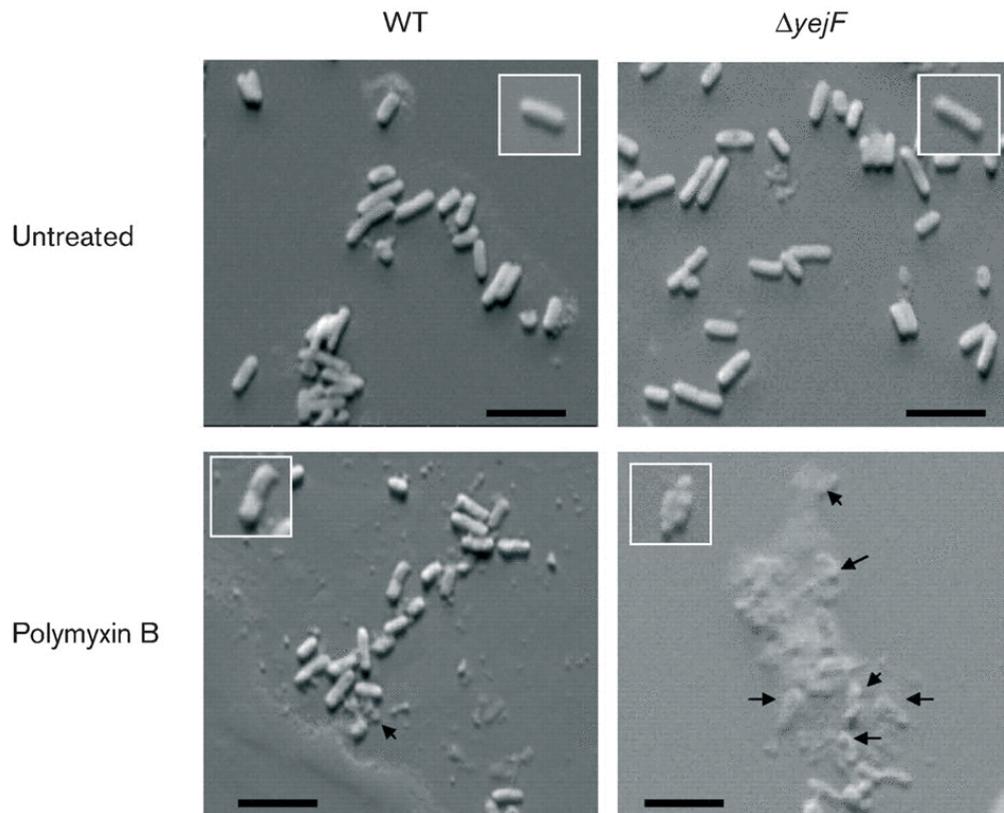
Figure 1.18. Scanning electron microscopic images of wild-type (WT) and Δ*yejF*, the Yej transporter NBD, treated with polymyxin B.

Insets are zoomed in images of single bacterial cells from the main images. When the Δ*yejF* mutant is exposed to polymyxin B membrane disruption and cell lysis occurs as shown in the above image. When cells are not treated with polymyxin B or the WT is treated with polymyxin B the same membrane disruption and lysis does not occur. Image taken from (Eswarappa *et al.*, 2008).

particular study, contains five genes within its *yej* operon. *yejA1* and *yejA2* which are the SBPs, *yejB* and *yejE* which are the TMD domains and *yejF* the NBD (Wang *et al.*, 2016).

A Δ*yejAABEF* mutant strain of *B. melitensis* NI was created, this strain was more susceptible to polymyxin B than the wild type. The Δ*yejE* strain was also more susceptible to polymyxin B than the wild type. When *yejE* was complemented back into the mutant strain resistance to polymyxin B increased to the same level as the wild type strain. However, Δ*yejA1*, Δ*yejA2*, Δ*yejB* and Δ*yejF* mutant strains showed no difference in susceptibility to polymyxin B as compared to the wild type (Wang *et al.*, 2016).

Again, scanning electron microscopy was conducted on the different strains. The Δ*yejAABEF* and Δ*yejE* mutant strains showed membrane irregularities, extruded cytoplasm and other signs of cell lysis and membrane disruption (Figure 1.19) (Wang *et al.*, 2016).

The relative gene expression level of the *yej* operon was investigated in the presence of polymyxin B. Polymyxin B was able to induce the expression of *yejA1*, *yejA2*, *yejB*, *yejE* and *yejF* in *B. melitensis* NI. The gene expression levels of *yejA1*, *yejB* and *yejE* increased by 3-4 fold in the presence of polymyxin B as compared to the untreated control (Wang *et al.*, 2016).

Together these experimental data strongly suggest a role for the Yej transporter in the defence against CAMPs. Although in different studies different components of the transporter gave the strongest/weakest CAMP resistance phenotypes, it is important to note that in most cases a transporter needs all of its components to function in transmembrane substrate translocation.

1.5.7 The Yej transporter and Microcin C

Microcin C is a bacterially produced peptide-nucleotide antibiotic and a potent inhibitor of aspartyl-tRNA synthetase. Microcin C, which is not a CAMP, is a heptapeptide with a modified AMP covalently attached to its C-terminus (Figure 1.20). Microcin C sensitive cells process Microcin C in the cytoplasm into its active form, a non-hydrolysable aspartyl-adenylate, which inhibits translation and leads to bacterial cell death (Figure 1.20). Unprocessed Microcin C has no effect on translation and processed Microcin C has no effect on Microcin C sensitive cells. This means Microcin C must be transported to the cytoplasm in its unprocessed form where it is degraded to the processed active form, a so-called Trojan horse antibiotic (Novikova *et al.*, 2007).
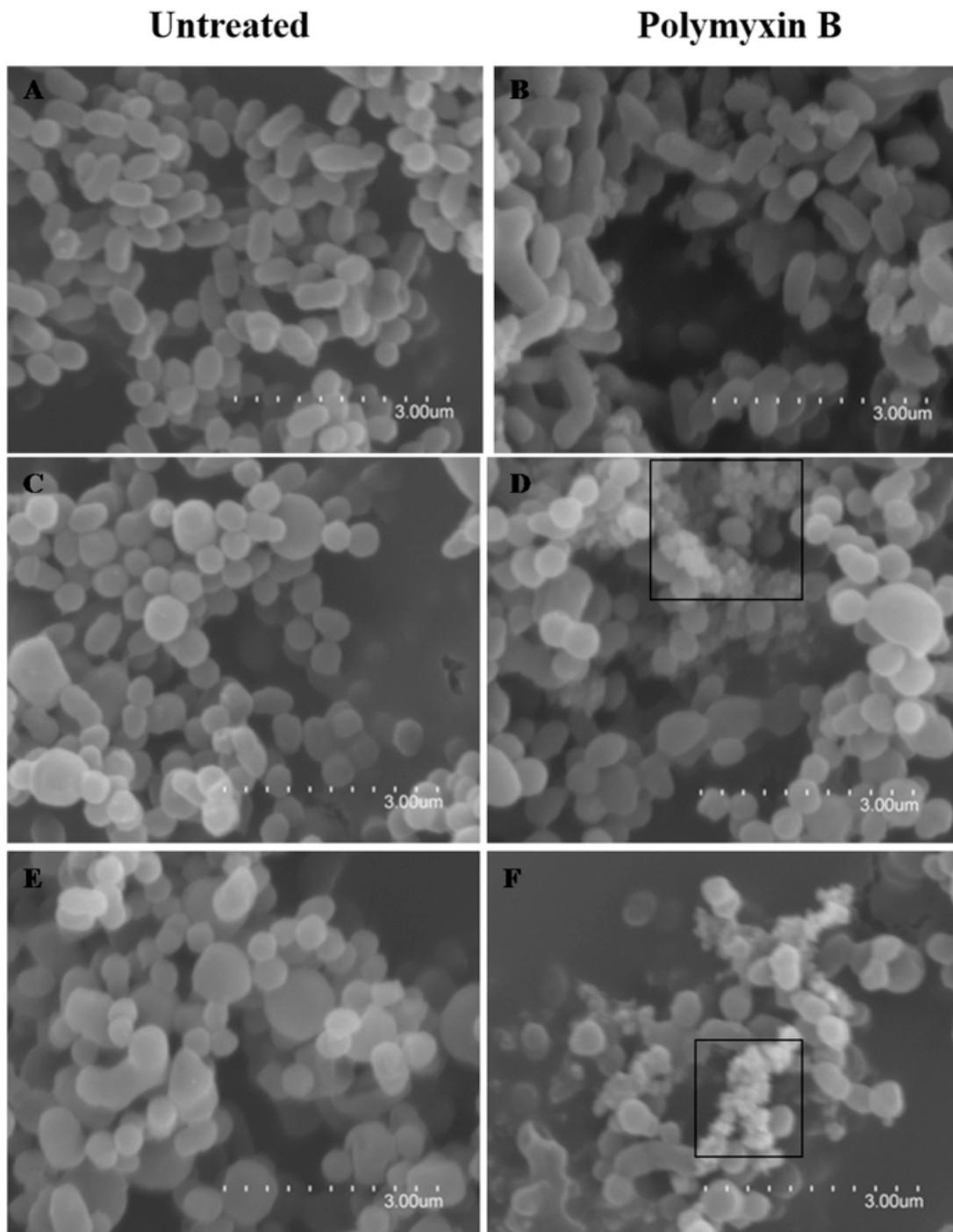
Figure 1.19. Scanning electron microscopic images of *B. melitensis* NI, Δ*yejAABEF* and Δ*yejE* mutants treated with polymyxin B.

Boxes highlight areas of membrane disruption. *B. melitensis* NI is shown in A and B, Δ*yejE* mutants in C and D and Δ*yejAABEF* mutant in E and F. All untreated samples show no cell disruption but Δ*yejE* and Δ*yejAABEF* mutants when treated with polymyxin B show signs of cell disruption and lysis whereas the wildtype has the same morphology as the untreated samples. Image taken from (Wang *et al.*, 2016).
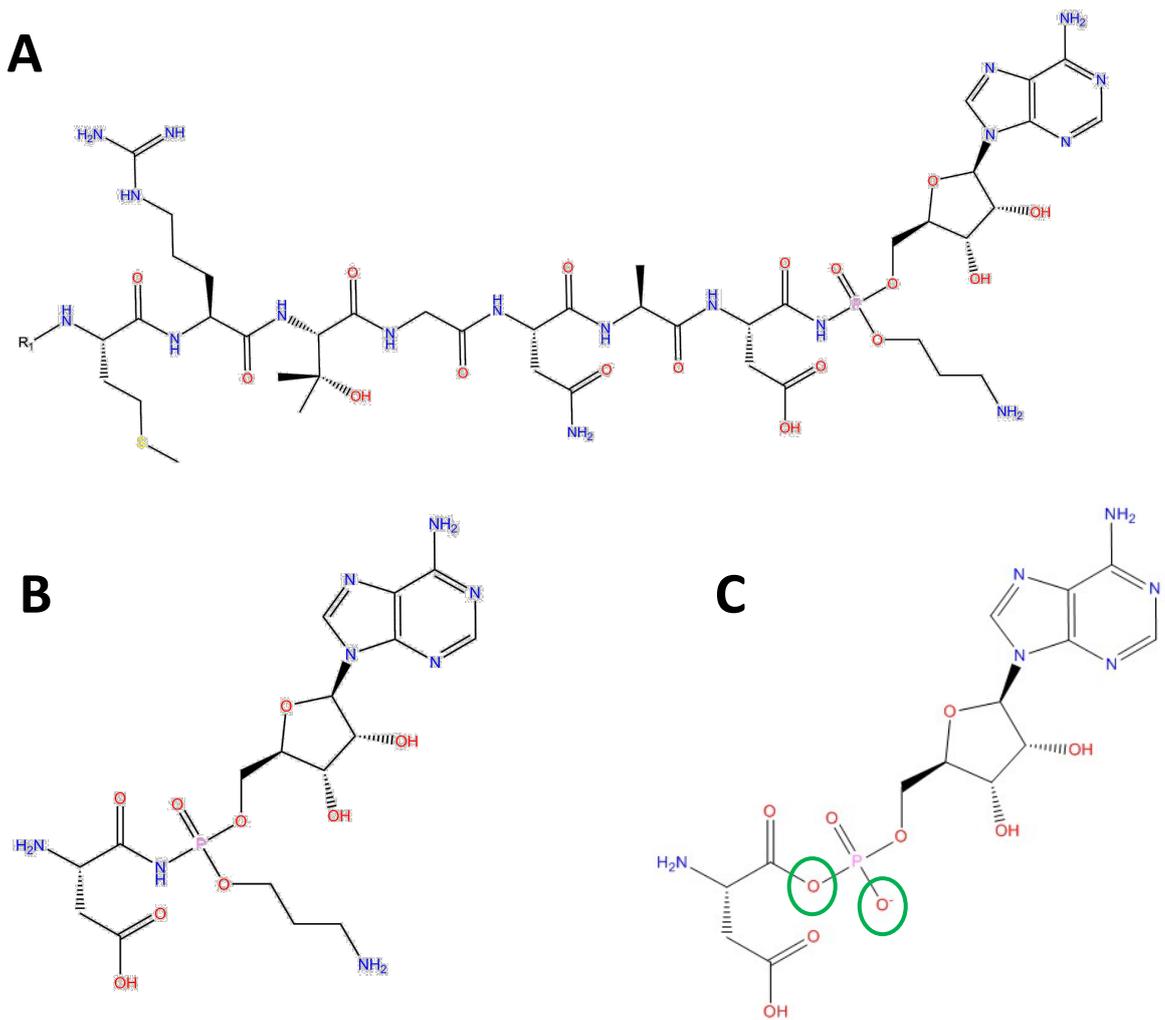
Figure 1.20. Structure of Microcin C in the unprocessed and processed forms.

(A) shows Microcin C in the unprocessed form, the form in which YejA binds Microcin C. R_1 is either CHO or H. (B) shows Microcin C in the processed form. (C) Shows aspartyl-adenylate, the molecule processed Microcin C mimics to inhibit aspartyl-tRNA synthetase. Changes between (B) and (C) are indicated by green circles on (C).

To investigate how Microcin C enters the cytoplasm, a random transposon insertion library was created with the mariner-based transposon TNSC189. Microcin C resistant colonies of *E. coli* SG289 cells were then selected and the transposon insertion sites mapped. Three different insertions were found in *yejA* and *yejB* and a single further insertion was found in *yejF*. Cells were created carrying deletions of each of the *yejABEF* genes and tested for their ability to grow in the presence of Microcin C. The results showed that deletion of any one of the *yejABEF* genes led to complete resistance to Microcin C. Therefore the Yej transporter was identified as the only transporter required for Mcirocin C uptake (Novikova *et al.*, 2007).

In later work, a series of Microcin C analogues were created to test the specificity of the Yej transporter. Aminoacyl-sulfamoyl adenosines were used to try to mimic natural Microcin C and a series of compounds were created where the heptapeptide part of Microcin C was truncated from the C-terminus, although aspartate was always maintained as the C-terminal residue. These compounds were then tested on Microcin C sensitive cells whilst the inhibition of the aspartyl-tRNA aminoacylation reaction was monitored. It was shown that Microcin C analogues required a minimum peptide chain length of 6 residues and an N-terminal formyl-methionyl-arginyl sequence to allow transport, indicating possible requirements of YejA substrates (Gaston *et al.*, 2011).

The Yej transport system therefore confers susceptibility to Microcin C, so why would a bacterium maintain a system that makes it vulnerable to Microcin C? Possibly the evolutionary advantage of maintaining the Yej transporter is to provide resistance to CAMPs.

1.6 Aims of the project

This thesis aims to biochemically characterise specific peptide SBPs which are thought to recognise CAMPs. A structural approach is used to examine how specific peptides are coordinated by the SBPs whilst a biochemical approach is used to investigate the specificity of binding. This research is conducted with a view to developing new drugs to help tackle the spread of antibiotic resistance.

SapA and YejA are the receptor components of their respective ABC-type antimicrobial peptide transporters. These proteins are therefore expected to capture the extracellular substrate and define the specificity of the transporter. Determining the structure of SapA and YejA will help to understand how they are able to transport different, and possibly

58

folded, CAMPs. Solving the structure of a SapA-CAMP or YejA-CAMP complex would give great insight into the binding mechanism of the SBPs. The technique used to determine the structure of SapA, YejA and associated CAMPs will be X-ray crystallography. For the latter studies, either Isothermal Titration Calorimetry and/or thermal shift assays will be used to study the binding of CAMPs to SapA and YejA using facilities in the Technology Facility (TF) at the University of York. This work will indicate a method of binding and highlight key residues in CAMPs, SapA and YejA which are important in binding. These residues could then be mutated or altered to enable new drug design.

Chapter Two

Materials and Methods

2. Materials and Methods

2.1 Media and Antibiotics

2.1.1 Luria-Bertani broth and agar

Luria-Bertani broth (LB) was made using 10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract and made up to 1 L with distilled water. LB agar was made up in smaller volumes of 200 ml with 4 g tryptone, 4 g NaCl, 2 g yeast extract, 3 g agar and made up to volume with distilled water. Once made up all media were sterilised via autoclave at 121 °C for 15-20 minutes.

2.1.2 Antibiotics

Where appropriate antibiotic selection was used at a working concentration of 100 µg/ml ampicillin and 30 µg/ml kanamycin, both of which were 0.22 µm filter sterilised and stored at 4 °C.

2.1.3 IPTG

Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Melford Biolaboratories) was made up as a 1 M stock in distilled water, 0.22 µm filter sterilised and stored in 1 ml aliquots at -20 °C. When added to cells the final concentration of IPTG was 1 mM or the concentration is otherwise indicated.

2.1.4 Overnight bacterial cultures

5 ml of LB with appropriate antibiotic was added to a Sterilin tube along with a single bacterial colony or a scraping from a glycerol stock. The bacterial culture was grown overnight at 37 °C with shaking at 220 rpm.

2.1.5 Glycerol stocks

800 µl of overnight culture was added to 400 µl of 50% glycerol and vortexed to create glycerol stocks. These were stored at -80 °C until further use.

2.1.6 Bacterial strains and proteins used

Table 2.1 details all of the bacterial strains, genomes and protein identifiers used throughout this work.

Protein Name	Strain Name	GenBank Genome ID	Gene Coordinates	Uniprot Protein ID	PDB ID
EcSapA	*E. coli* K-12 MG1655	U00096.3	1355467-1357110	Q47622	-
StSapA	*S.* Typhimurium LT2	AE006468.2	1783672-1785321	P36634	-
HiSapA	*H. influenzae* Rd KW20	L42023.1	1702963-1704660	P45285	-
EcYejA	*E. coli* K-12 MG1655	U00096.3	2272364-2274178	P33913	-
EcMppA	*E. coli* K-12 MG1655	U00096.3	1393227-1394840	P77348	3O9P
EcDppA	*E. coli* K-12 MG1655	U00096.3	3706098-3707705	P23847	1DPE
StOppA	*S.* Typhimurium LT2	AE006468.2	1839602-1841350	P06202	1B9J
BsAppA	*B. subtilis* 168	U20909.1	2732-4363	P42061	1XOC
LlOppA	*L. lactis* subsp. cremoris MG1363	AM406671.1	687458-689260	A2RJ53	3DRG

Table 2.1. Table of bacterial strains, genome and protein identifiers.

Information on bacterial strains used in this work along with the genome identifiers. Protein identifiers as well as PDB codes have also been added where appropriate. These proteins have been used throughout this work.

2.2 Gene Cloning

2.2.1 Agarose gel electrophoresis

A 1% (w/v) agarose gel in TBE buffer (16.2 g/L Tris, 2.75 g/L boric acid, 0.95 g/L EDTA) was used to separate DNA fragments of varying sizes. Molten agarose was mixed with either 10 µl/100 ml ethidium bromide or 1 µl/100 ml SYBR safe and poured into a gel container. A plastic comb was inserted to create a number of wells, and the gel allowed to solidify. After the gel had set the comb was removed and the gel covered with a solution of TBE buffer. 5 µl of ladder, either HyperLadder 1 kb plus (Bioline) or 2-Log ladder (NEB), was loaded into one of the wells. 5 µl of sample was mixed with 1 µl of sample loading buffer (Bioline) and 5 µl loaded into a well. The gel was run at 70 V for 50 mins and imaged using a transilluminator (Syngene Imaging System).

2.2.2 Polymerase Chain Reaction (PCR)

The high fidelity DNA polymerase Phusion was used in PCR reactions to amplify target DNA fragments. A standard PCR reaction mixture consisted of 32.5 µl water, 10 µl 5x Phusion HF buffer (Thermo Scientific), 1 µl 10 mM dNTPs (Thermo Scientific), 0.5 µl Phusion (Thermo Scientific 2U/pL), 2.5 µl genomic DNA template, 2.5 µl each of forward and reverse primers (at 10 µM). Primers were supplied by IDT and designed using the online NEBuilder Assembly Tool (Table 2.2). The mixture was stored on ice until placed in the thermocycler. An initial denaturing step at 98 °C for 2 mins was performed. After this, 35 cycles of 98 °C for 20 seconds, 60 °C for 30 seconds and 72 °C for 1.5 mins was used to amplify the DNA. A final extension at 72 °C for 5 mins was then performed. The sample was then held in the thermocycler at 10 °C until it was frozen or run on an agarose gel.

2.2.3 Restriction digest of DNA

All PCR amplified inserts were cloned into the pETFPP_30 vector for future expression. The pETFPP_30 vector is based on the pET22b+ vector and contains an optional N-terminal PelB leader sequence and a cleavable C-terminal hexahistidine tag (Figure 2.1). Both pETFPP_30 and PCR insert DNA were treated with restriction endonucleases to digest the DNA ready for assembly. Each 50 µl reaction contained 5 µl cutsmart buffer (NEB), 1 µl Nde1 (NEB), 1 µl Xho1 (NEB) and 1 µg DNA with the reaction made up to 50 µl with sterile milliQ water. The mixture was incubated at 37 °C for 1 hour and then immediately put through a PCR clean-up protocol (Qiagen PCR clean-up kit) to remove endonucleases before DNA assembly.

Primer Name	Primer Sequence
EcYejA-F	5'-CTTTAAGAAGGAGATATACATATGCAGGCTATCAAGGAAAGCTATG-3'
EcYejA-R	5'-TTGGTCCCTGGAACAGAACCTCGAGCTCTCCCTGTTTGCTGGC-3'
EcPYejA-F	5'-GCTGCCCAGCCGGCGATGGCCATGGCTCAGGCTATCAAGGAAAGC-3'
EcPYejA-R	5'-TTGGTCCCTGGAACAGAACCTCGAGCTCTCCCTGTTTGCTGGC-3'
STEcYejA-F	5'-CTTTAAGAAGGAGATATACATATGGGGCAGATAACGTTGTCAGC-3'
MTEcYejA-F	5'-CTTTAAGAAGGAGATATACATATGGCGCCAAAAGGTGGGCAG-3'
LTEcYejA-F	5'-CTTTAAGAAGGAGATATACATATGCATTTTGATTATGTGAACCCCG-3'
HiPSapA-F	5'-TCGAGCCATGGCGCCAAGTGTTCCAACATTTTTAACTGAAAATGG-3'
HiPSapA-R	5'-TCGAGCTCGAGGTATTTCTCCTGAATAAAATATAAGGTGG-3'
EcPSapA-F	5'-TCGAGCCATGGCTCCTGAATCTCCCCCGCATGC-3'
EcPSapA-R	5'-TCGAGCTCGAGTGGTTTTTTCACCTCATCC-3'
StPSapA-F	5'-TCGAGCCATGGCTACTGCGCCCGAACAAACTGC-3'
StPSapA-R	5'-TCGAGCTCGAGTGGTTTTTTCACCTCTTCG-3'
EcNSapA-F	5'-TCGAGCATATGCGCCAGGTATTATCGTCTCTTTTGGTGATTGC-3'
EcNSapA-R	5'-GCATCCTCGAGTGGTTTTTTCACCTCATCCTGTTTCTCGCG-3'
T7-F	5'-TAATACGACTCACTATAGGG-3'
T7-R	5'-GCTAGTTATTGCTCAGCGG-3'

Table 2.2. Primers used to amplify DNA.

Primers were used to either amplify genomic DNA via PCR or for sequencing of inserts once cloning had been completed.

A

pETFPP_30   pelB.ORF-3cProtease-HIS

```
  BglII                  T7 Promoter              lac operator        XbaI                        rbs
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
                                                                                Signal peptidase
                                                                      MscI                                  .
  NdeI      BspMI                   pelB leader                      NcoI                    BamHI  .
TATACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGATGGCCATGGATATCGGAATTAATTCGGATCCG
        MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAlaAlaGlnProAlaMetAlaMetAspIleGlyIleAsnSerAspPro
                                                                               -----------------------ORF

  EcoRI SacI   SalI HindIII  NotI    XhoI   HRV 3c   |                 HIS-Tag       END
AATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGGTTCTGTTCCAGGGACCAAGCAGCGGCCACCACCACCACCACCACTGAGATCCGGC
AsnSerSerSerValAspLysLeuAlaAlaAlaLeuGluValLeuPheGlnGlyProSerSerGlyHisHisHisHisHisHisEnd
 HERE------------------------↑
                                                                      T7 Terminator             .
TGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTT
```

B

pET-22b(+) sequence landmarks

T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482
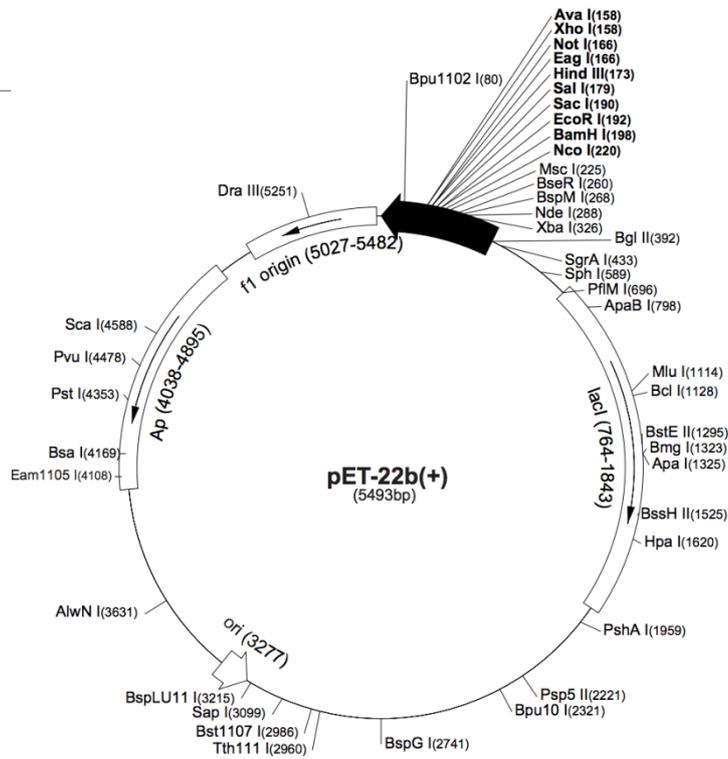
Figure 2.1. pETFPP_30 and pET22b+ vectors.

(A) Shows the open reading frame of the pETFPP_30 vector with an optional N-terminal PelB leader sequence and cleavable C-terminal hexahistidine tag. (B) shows the pET22b+ vector, the block black arrow indicates the region which changes have been made to produce the pETFPP_30 vector.

66

2.2.4 HiFi DNA Assembly

This method is used to assemble DNA fragments, in this case PCR amplified genes of interest, with overlaps matching the target vector, in this case pETFPP_30. An exonuclease creates further single-stranded 3' overhangs on both the restriction digested vector and insert which allows them to anneal. A DNA polymerase fills in any gaps between the assembled DNA and is followed by a DNA ligase which seals the nicks. The gene of interest and pETFPP_30 vector were assembled via the NEB HiFi DNA assembly protocol for 2-3 fragment assembly. The mixture was incubated at 50 °C for 15 mins and then used immediately in a transformation into Solopack Gold DH5α competent cells (Agilent) or stored at -20 °C until further use.

2.2.5 Transformation of competent cells via heat shock

An aliquot of the appropriate competent cells was thawed on ice whilst an aliquot of Super Optimal broth with Catabolite repressor (SOC) media was heated to 42 °C. SOC media is a variant of LB with added glucose which results in higher transformation efficiencies. Once the cell aliquots had defrosted they were swirled gently to mix and 0.1-50 ng of appropriate plasmid DNA or a volume of ligation mixture was added to each tube which was incubated on ice for 30 mins. The cells were then heat pulsed at 42 °C for 1 min and incubated on ice for 2 mins before 175 µl of 42 °C SOC media was added, the tubes were then incubated at 37 °C for an hour with shaking at 220 rpm. Afterwards 200 µl of the cells were plated onto LB agar plates with the appropriate antibiotics and incubated overnight at 37 °C.

2.2.6 Colony PCR

Colony PCR was used to screen colonies arising from transformation experiments for the desired clones. Template DNA for the colony PCR reaction was the DNA extracted from each colony during the initial heat step at 95 °C. Each standard colony PCR mix contained 2.5 µl of (10 mM) reverse primer, 2.5 µl of (10 mM) forwards primer, 0.5 µl of 10 mM dNTPs (Thermo Scientific), 5 µl 5x green GoTaq reaction buffer (Promega), 0.2 µl GoTaq G2 DNA polymerase (Promega) and 14.3 µl of water. The mixture was stored on ice until transferred to the thermocycler. An initial step of 95 °C for 5 mins was performed and then 35 cycles of 95 °C for 30 seconds, 60 °C for 30 seconds, 68 °C for 1 minute 45 seconds. A final extension of 72 °C for 5 mins was performed, after which the mixture was held at 12 °C until frozen or run on an agarose gel.

2.2.7 Miniprep and sequencing of DNA

Plasmids were isolated from overnight cultures of bacteria using a Qiagen miniprep kit, following the manufacturer's instructions. All plasmids were checked for accuracy and confirmed via DNA sequencing using T7-F and T7-R primers. The sequencing was carried out by GATC and analysed using Ugene software.

2.3 Expression of recombinant protein

Following DNA sequencing, correctly cloned constructs were transformed into BL21-Gold(DE3) competent cells (Agilent) for protein overexpression using the protocol detailed in 2.2.5.

2.3.1 Small scale whole cell expression trials

50 ml of LB, supplemented with appropriate antibiotic, was inoculated to an OD_{600} between 0.01-0.2 with overnight cultures. The cells were grown at a variety of different temperatures, induced with a variety of different concentrations of IPTG and grown for varying lengths of time to determine which conditions gave the highest yield of soluble protein. An initial "uninduced" 1 ml culture sample was taken immediately before addition of IPTG and additional 1 ml samples were taken every hour until the 4 hour time point. The following morning an additional 1 ml culture sample was taken. All 1 ml culture samples were centrifuged at 23897 x g for 5 mins to pellet cells immediately after harvesting. The supernatant was discarded and the pellets were stored at -20 °C.

2.3.2 Lysing cells using Bugbuster

Bugbuster disrupts the cell wall of bacterial cells to release soluble protein as an alternative to mechanical methods such as sonication. Pelleted cells from 2.3.1 were re-suspended in 50 µl Bugbuster (Novagen) and PMSF (Alpha Diagnostic International Inc.) protease inhibitor. The mixture was incubated at room temperature for 10-20 mins with shaking/rocking before centrifugation at 16000 x g for 20 mins at 4 °C. Soluble and insoluble fractions were then separated and samples were analysed via SDS-PAGE.

2.3.3 Lysing cells using sonication

Pelleted cells from 2.3.1 were re-suspended in 200 µl Buffer A (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole) with PMSF protease inhibitor and sonicated for 30 secs. The mixture was then centrifuged at 18787 x g for 20 mins. Soluble and insoluble fractions were then separated and samples were analysed via SDS-PAGE.

2.3.4 Large scale expression of EcYejA

1 L of LB, supplemented with 100 μg/ml ampicillin, was inoculated to an OD_{600} of 0.02 with overnight cultures of the EcYejA strain. They were grown at 37 °C with shaking at 150 rpm until they reached an OD_{600} between 0.4 – 0.6. At this point expression was induced with the addition of 1 mM IPTG. If EcYejA was to be treated with 2 M guanidine hydrochloride during the purification step the flasks were moved to 30 °C with shaking at 180 rpm and allowed to grow overnight. If the 2 M guanidine hydrochloride step was to be omitted during purification, cells were allowed to grow for a further 3 hours at 37 °C. Cells were harvested by centrifugation at 5000 rpm, 4 °C for 20 mins using an F10S rotor.

2.4 Preparation of bacterial extracts

2.4.1 Tris-sucrose solution supplemented with EDTA extraction of periplasm

The TSE periplasmic fraction extraction method was developed as a variation on the osmotic shock method of periplasmic fraction extraction. Osmotic shock can often contaminate the periplasmic fraction with cytoplasmic content, the TSE method has been shown to produce cleaner periplasmic fractions and so was used in this work (Quan *et al.*, 2013). Induced cells were harvested by centrifugation at 3000 x *g* for 20 mins at 4 °C. Supernatant was discarded and the pellet carefully resuspended in 1 ml/100 ml of culture TSE buffer (200 mM Tris pH 8.0, 500 mM sucrose, 1 mM EDTA) with a wire loop. Cells were incubated in TSE buffer on ice for 30 mins then centrifuged at 15000 x *g* for 30 mins at 4 °C. The supernatant and cell pellet were stored separately at -20 °C until further analysis.

2.4.2 Periplasmic fraction extraction with lysozyme

Lysozyme is a glycoside hydrolase which breaks down peptidoglycan. Induced cells were harvested by centrifugation at 3000 x *g* for 20 mins at 4 °C. The cell pellet was then weighed and resuspended in 4 ml/g of cells ICOS buffer (20 mM Tris pH 8.0, 25% (w/v) sucrose, 5 mM EDTA) and incubated on ice for 15 mins. The mixture was then centrifuged at 8500 x *g* for 20 mins, supernatant was removed and stored at -20 °C (sucrose fraction). Pellet was dissolved in 4-5 ml/g of cells of 5 mM $MgCl_2$ with protease inhibitor, 40 μl of 15 mg/ml lysozyme per gram of cells was added and mixture was incubated on ice for 30 mins. Mixture was then centrifuged at 8500 x *g* for 20 mins at 4 °C, the supernatant and cell pellet were stored separately at -20 °C until further analysis.

2.4.3 Cytoplasmic EcYejA protein recovery

Induced cells were harvested by centrifugation at 5000 rpm, 4 °C for 20 mins using an F10S rotor. If EcYejA was to be treated with 2 M guanidine hydrochloride during the purification step cell pellets were resuspended in 35 ml (for 1 L of culture) buffer A (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole) with PMSF protease inhibitor. If the 2 M guanidine hydrochloride step was to be omitted during purification, cell pellets were resuspended in 35 ml (for 1 L of culture) resuspension buffer (50 mM KPi pH 7.8, 200 mM NaCl, 20% glycerol, 10 mM imidazole) with PMSF protease inhibitor. All cells were then lysed via sonication (6 mins on time, 3 secs on, 7 secs off) and centrifuged to remove cell debris at 15,000 rpm, 4 °C for 25 mins, using an SS34 rotor. The soluble fraction containing EcYejA was then purified.

2.5 Protein purification

2.5.1 Guanidine hydrochloride denaturation to unfold-refold EcYejA followed by nickel-affinity chromatography using HisTrap HP columns

Two 5 ml HisTrap HP columns (GE Healthcare) were attached in series effectively to create a 10 ml HisTrap column and equilibrated via a peristaltic pump with buffer A (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole). The 2 M guanidine hydrochloride treated soluble fraction from 2.4.3 was flowed over the column, again using the peristaltic pump. The column was then washed with ~3 column volumes (CV) of buffer A to elute any weakly bound protein contaminants. The column was then connected to an Akta purifier and a prepared method run automatically, using the Unicorn software package, where the column was washed with 3.5 CVs of buffer A and then 2 CVs of 100% buffer B1 (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole, 2 M guanidine hydrochloride) followed by 2 CVs of 75% buffer B1, 2 CVs of 50% buffer B1, 2 CVs of 25% buffer B1 and 2 CVs 0% buffer B1 (therefore 100% buffer A). Purification was monitored by detecting the absorbance of column eluent at a wavelength of 280 nm and fractions collected using an automated fraction collector. Following this, the column was developed with a 0-100 % imidazole gradient in buffer A over 10 CV. Purification was monitored by detecting the absorbance of column eluent at a wavelength of 280 nm and fractions collected using an automated fraction collector.

2.5.2 Nickel-affinity chromatography using HisTrap HP columns of EcYejA

The soluble protein sample from 2.4.3 that was not treated with 2 M guanidine hydrochloride was purified using an Akta start and the Unicorn software package. A 5 ml

HisTrap HP column (GE Healthcare) was equilibrated with 5 CVs wash buffer (50mM KPi pH 7.8, 200mM NaCl, 20% glycerol, 40mM imidazole). The protein sample was added to a single 5 ml HisTrap HP column via the sample line and the column washed to remove any unbound protein with 10 CVs of wash buffer. Protein bound to the HisTrap column was eluted in a single step by washing the column with 5 CVs of elution buffer (50mM KPi pH 7.8, 200mM NaCl, 20% glycerol, 500mM imidazole).

2.5.3 Cleavage of Histidine tag

The hexahistidine tag of EcYejA was cleaved with human rhinovirus 3C (HRV 3C) protease in a ratio of 1:200 HRV 3C protease:EcYejA before further purification via size exclusion chromatography. Both EcYejA and HRV 3C protease were added to dialysis tubing with a molecular weight cut off of 12-14 kDa and incubated in SEC (size exclusion chromatography) buffer (50 mM Tris pH 8.0 and 150 mM NaCl) overnight at 4 °C with stirring in preparation for the size exclusion chromatography the following day.

2.5.4 Size exclusion chromatography

A superdex 200 SEC column (GE Healthcare) was equilibrated with SEC buffer overnight. Protein from the overnight cleavage of the hexahistidine tag was concentrated to less than 2 ml using a 30 kDa molecular weight Vivaspin column. Concentrated protein was then pushed straight onto the SEC column via a syringe. The column was washed with one CV of SEC buffer (120 ml) and EcYejA protein eluted. Purification was monitored by detecting the absorbance of column eluent at a wavelength of 280 nm and fractions collected using an automated fraction collector.

2.6 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

2.6.1 Buffers and gel

Polyacrylamide gels were either Mini-PROTEAN TGX stain-free 12% 15 well gels (Bio-Rad) or made using the following protocol. Resolving polyacrylamide gel was cast using 375 mM Tris-HCl pH 8.8, 12% acrylamide, 0.1% sodium dodecyl sulphate (SDS), 0.08% tetramethylethylenediamine (TEMED) and 0.05% ammonium persulfate (APS). The stacking gel was cast from 130 mM Tris-HCl pH 6.8, 4% acrylamide, 0.1% SDS, 0.1% bromophenol blue, 0.16% TEMED and 0.05% APS. Running buffer (14.4 g/L glycine, 3 g/L Tris, 0.1% SDS) was added to the gels in the gel tank.

2.6.2 Sample preparation

3 µl of 2x sample buffer (60 mM Tris pH 6.8, 10% glycerol, 5% β-mercaptoethanol, 2% SDS, 0.05% bromophenol blue) was added to 10 µl of soluble protein. Insoluble protein was resuspended in 50 x n µl 2x sample buffer (n µl is OD_{600} of cells when sample was taken). Sample mixtures were heated to 95 °C for 5 minutes before 10 µl of sample mixture was added to each well, 5 µl of broad range blue protein standard (NEB) was used as a marker. The method for naming SDS-PAGE gel samples is shown in Figure 2.2.

2.6.3 Running and staining/de-staining of SDS-PAGE gels

SDS-PAGE gels were run using a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (Bio-Rad) assembled via the manufacturer's instructions. After submerging the gel in running buffer, samples were loaded and the gel was run at 200 V for 50 mins. The gels were then rinsed with distilled water. SDS-PAGE gels were then boiled in ~200 ml deionised water in a microwave and then placed on a rocker for ~5 mins before the distilled water was drained off and ~50 ml magic dye (60 mg/L Coomassie Brilliant Blue G-250, 35mM HCl) added. The gels were then boiled again and placed back on the rocker until bands were visible and the gels imaged. To further de-stain the gels they were incubated in deionised water on the rocker.

2.7 Protein concentration determination

The Bioteck epoch 2 microplate reader with Gen5 3.00 software was used to measure the absorbance at a wavelength of 280 nm. The extinction coefficient and molecular weight of purified protein were calculated using Protparam from the Swiss Institute of Bioinformatics (https://web.expasy.org/protparam/). The Beer-lambert law was used to determine protein concentration.

2.8 Storage of protein

Protein was stored at 4 °C in SEC buffer (50 mM Tris pH 8.0 and 150 mM NaCl) and used for up to a month before being discarded and a fresh batch of protein prepared.

2.9 Peptide synthesis

Peptides were synthesised from C to N-terminus. A cartridge was prepared by adding a filter and a tap. 300 mg of both Fmoc-Gly-Rink Amide-MBHA resin (Cambridge Bioscience) and Fmoc-Asp(OtBu)-Rink Amide-MBHA resin (Cambridge Bioscience) were weighed out and added to the cartridge. The cartridge was two-thirds filled with dimethylformamide
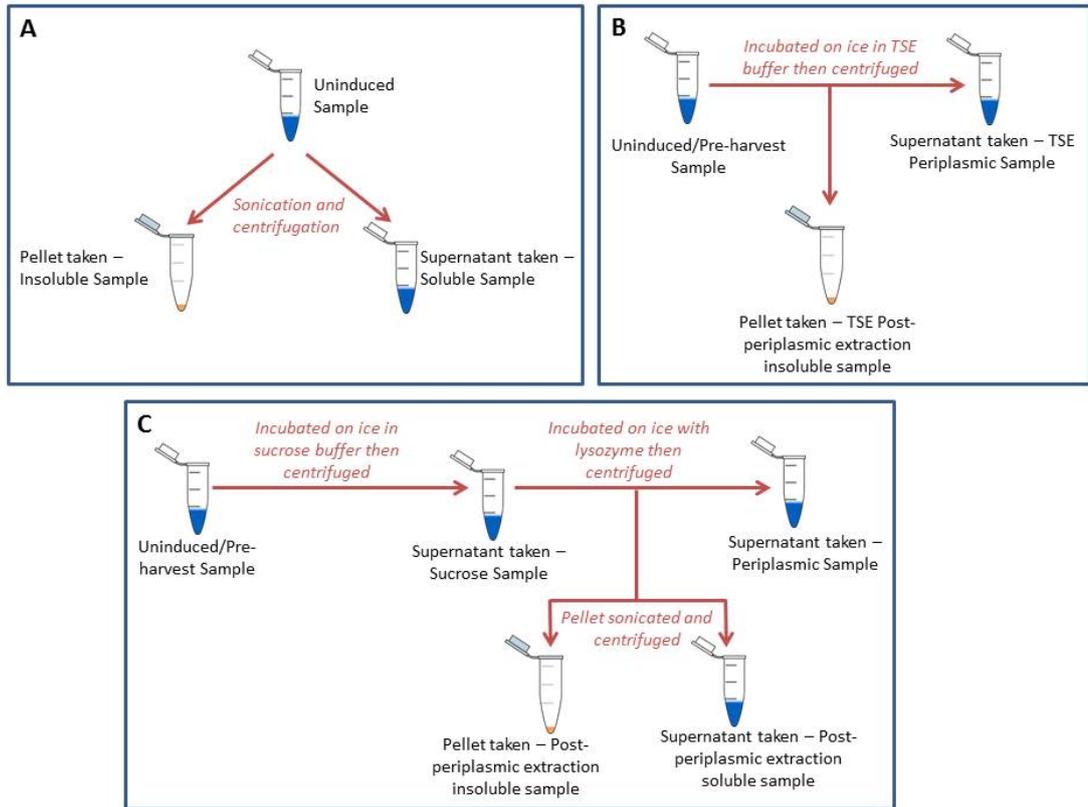
Figure 2.2. Protein sample preparation for SDS-PAGE gels.

Schematic diagram showing the three different methods of extracting, preparing and naming samples for SDS-PAGE gel analysis.

(DMF) and left to swell for 30 mins with rotation. The DMF was then drained into a manifold and the resin washed with 20% piperidine (PIP) in DMF five times followed by five washes with DMF. PIP washes consisted of half-filling the cartridge with 20% PIP, sealing and inverting the cartridge before allowing the cartridge to rotate for 2 mins and then draining in the manifold. DMF washes were carried out in identical fashion.

To couple amino acids to the peptide chain the following quantities (Table 2.3), chosen to give a yield of ~100 mg of synthesised peptide, were weighed out for the synthesis of MRTGNAD and fMRTGNAD peptides. For the synthesis of ~100 mg of fMRTGNAD(dansyl K)G, the quantities of amino acids listed in Table 2.4 were weighed out.

The residues were added to a sample vial along with O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU) at 1.02 equivalents (for MRTGNAD and fMRTGNAD 274 mg at 0.662mM and for fMRTGNAD(dansyl K)G 207 mg at 0.5 mM). The residues and HCTU were dissolved in DMF and, for MRTGNAD and fMRTGNAD 249 µl of N,N-Diisopropylethylamine (DIPEA) and for fMRTGNAD(dansyl K)G 188 µl DIPEA was added. The mixture was quickly transferred to the cartridge, which was sealed and inverted to mix and then rotated for 60 mins.

After each amino acid coupling, the resin was washed by draining the contents into a manifold and half-filling the cartridge with DMF, sealing and inverting the cartridge and leaving it to rotate for 2 mins. This was repeated three times. Fmoc cleavage was carried out by five PIP washes followed by five DMF washes. The cartridge was then either stored overnight or another amino acid coupled. At the point of addition of the final residue (fMet or Met), the MRTGNAD peptide was weighed and split in half and half added to a new cartridge. To one cartridge –Met-OH was added and to the other –fMet-OH added to create both MRTGNAD and fMRTGNAD.

Once the peptides were complete, resin shrinkage was carried out by washing the cartridge, in the same manner as the PIP and DMF washes, with dichloromethane (DCM) three times followed by three washes with methanol. The cartridge was then left overnight on a high vacuum line with the tap off and the lid on in an inverted position.

For resin cleavage 5 ml of 88:5:5:2 trifluoroacetic acid:water:dithiothreitol:triisopropylsilane was prepared for each cartridge and the

Residue	mg	mM
Fmoc-Ala-OH (Sigma)	210	0.675
Fmoc-Asn(Trt)-OH (Sigma)	403	0.675
Fmoc-Gly-OH (Sigma)	201	0.675
Fmoc-Thr(tBu)-OH (Sigma)	268	0.675
Fmoc-Arg(Pbf)-OH (Sigma)	438	0.675
--Met-OH/--fMet-OH (Sigma)	101/120	0.675

Table 2.3. Information on amino acids coupled to MRTGNAD and fMRTGNAD peptides.

Concentrations and respective weights of the amino acids coupled to synthesise ~100 mg of peptides MRTGNAD and fMRTGNAD.

Residue	mg	mM
Fmoc-L-Lys(Dansyl)-OH (Santa Cruz Biotechnology)	245	0.510
Fmoc-Asp(tBu)-OH (Sigma)	210	0.510
Fmoc-Ala-OH	159	0.510
Fmoc-Asn(Trt)-OH	304	0.510
Fmoc-Gly-OH	152	0.510
Fmoc-Thr(tBu)-OH	203	0.510
Fmoc-Arg(Pbf)-OH	331	0.510
--fMet-OH	90	0.510

Table 2.4. Information on amino acids coupled to fMRTGNAD(dansyl K)G.

Concentrations and respective weights of the amino acids coupled to synthesise ~100 mg of fMRTGNAD(dansyl K)G.

cartridges half filled with the solution. The cartridges were then sealed and inverted and left rotating for 60 mins.

30 ml of diethylether per cartridge was prepared and chilled to -80 °C and stored in a falcon tube. After resin cleavage, cartridge contents were drained into the diethylether to precipitate the peptide. The solvent and peptide were then centrifuged at 4000 x g for 5 mins at 4 °C to form a pellet, the ether was decanted off and the pellet left to evaporate off any remaining ether. The pellet was then resuspended in 30 ml of ice cold ether and the process repeated three times.

The peptide was then dissolved in 10% (v/v) aqueous glacial acetic acid and flash frozen in liquid nitrogen with rolling to maximise surface area. The sample was then lyophilised for ~24 hours to afford a fully white solid. Peptide purity was then checked using LC-MS and the yield of each peptide was ~100 mg.

2.10 General biochemical and biophysical techniques

2.10.1 Electrospray mass spectrometry

Electrospray Mass Spectrometry was conducted on a Waters LCT Premier XE system with MassLynx 4.1 software. The system was calibrated with sodium formate solution and calibration verified with horse heart myoglobin (16951.5 ± 1.5 Da). EcYejA was run at a concentration of 10 mg/ml in 2 mM Tris-HCl pH 8.0. Samples were prepared in 1:1 acetonitrile-water containing 0.1% formic acid.

2.10.2 Native electrospray ionisation (ESI) mass spectrometry and matrix-assisted laser desorption/ionisation mass spectrometry mass spectrometry (MALDI-MS/MS)

All native ESI and MALDI-MS/MS experiments were carried out with Dr. Adam Dowle in the University of York Technology Facility. For all acquisitions, protein in aqueous 1 M ammonium acetate was infused at 3 mL/min into a Bruker maXis qTOF mass spectrometer via an electrospray ionisation source. All presented spectra were summed over 1 min acquisitions at 0.1 Hz spectral acquisition rate. Ion funnel voltages were adjusted as detailed in the results to aid the preservation (100 eV) or separation (200 eV) of gas-phase complexes. Subsequently all data from section 5.2.1 have been baseline subtracted (0.8 flatness) and smoothed (0.19, Gauss, 1 cycle) before maximum entropy deconvolution to average mass. All data from section 5.2.2 have been smoothed (0.5 Da, 1 cycle, Gauss) and baseline subtracted (flatness 0.8) before maximum entropy deconvolution to average masses. Concentrations of EcYejA and GEP are noted in the results sections 5.2.1 and 5.2.2.

For MALDI-MS/MS a 100 µl aliquot of sample was acidified with the addition of 10 µl aqueous 1% trifluoroacetic acid (TFA) before extracting and desalting peptides using Promega C_{18} ZipTips. Desalted peptides were spotted out onto a MALDI target plate and overlaid with 5 mg/ml 4-hydroxy-a-cyano-cinnamic acid matrix. Peptides were analysed by MALDI-MS/MS using a Bruker ultraflex III mass spectrometer with the 30 strongest precursors, with a S/N greater than 30, selected for MS/MS fragmentation. Spectra were baseline-subtracted and smoothed (Savitsky-Golay, width 0.15 m/z, cycles 4); monoisotopic peak detection used a SNAP averaging algorithm $C_{4.9384}$, $N_{1.3577}$, $O_{1.4773}$, $S_{0.0417}$, $H_{7.7583}$) with a minimum S/N of 5. MS^2 spectra were searched against the expected sequence of EcYejA using the Mascot search program with enzyme cleavage set at wild.

2.10.3 Size Exclusion Chromatography with Multi-Angle Laser Light Scattering (SEC-MALLS)

EcYejA samples at either 1 mg/ml or 3 mg/ml in buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl) were resolved on a Superdex 200 SEC column (GE Healthcare) at 0.5 ml/min with an HPLC system (Shimadzu). Light scattering data was collected continuously on material eluting from the column using an in-line Wyatt Dawn Heleos LS detector with an inline Wyatt Optilab rEX refractive index detector and SPD-20A UV detector. Molecular weights were calculated by analysing data with the Wyatt program ASTRA version 5.3.4.14.

2.10.4 Circular Dichroism (CD) spectroscopy

CD spectra of protein were collected using a J-810 spectropolarimeter (Jasco) along with the supplied software SpectraManager version 1.53.00 (Jasco). 0.2 mg/ml (2.92 µM) protein was analysed in 20 mM Tris pH 8.0 and 50 mM NaCl buffer. Spectra were recorded at 20 °C in a 1 mm pathlength quartz cuvette (Starna) between 190 – 260 nm at 200 nm/minute with 0.5 nm pitch.

2.10.5 Thermal shift assay

To initially set the parameters for the assay, a series of dilutions of EcYejA and the assay dye, SYPRO orange (Sigma), were carried out. A protein concentration of 0.5 mg/ml (7.3 µM) EcYejA and 5 x SYPRO orange was found to give good curves without using a large quantity of protein or dye. A total volume of 25 µl per well was used throughout the thermal shift assays and each combination of EcYejA + ligand was carried out in triplicate on the same plate. For each assay 71 cycles were carried out starting at 25 °C and increasing by 1 °C every 30 secs, taking a fluorescence measurement every 1 °C.

2.10.6 Crystallisation and structure determination of protein

Purified protein was dialysed into crystallisation buffer (20 mM Tris pH 8.0, 50 mM NaCl) and concentrated to 20 mg/ml. Initial crystallisation screens were created using the HYDRA96 (Robbins Scientific) to dispense commercial screens to 96-well plates and the Mosquito Nanolitre Pipetting robot (TTP Labtech) to create the sitting drops of protein and reservoir solution using the sitting-drop vapour diffusion method where 150 nl of protein was mixed with 150 nl of reservoir solution. Two drops were set up per well. The protein concentration in the upper drop of the 96-well plate was 20 mg/ml and concentration in the lower drop was 10 mg/ml. A series of commercially available screens were used: JCSG+, PDB Min and PEG/ION. Plates were then sealed and stored at 18 °C and checked at regular intervals for crystal growth.

If crystals had grown, crystal optimisation was carried out in 24-well plates using the hanging-drop vapour diffusion method with the same or similar reservoir conditions and the same concentration of protein. A reservoir volume of 1 ml was used and a 1 µl:1 µl and 2 µl:2 µl protein:reservoir solution ratio was used. Again the plate was sealed and stored at 18 °C and checked regularly for crystal growth.

Crystals were harvested using nylon loops of varying sizes. 1 µl of a cryoprotectant solution (30% glycerol, 14 µl mother liquor to a total of 20 µl) was added to the drop containing the crystal and immediately after fishing, the crystal was cryocooled in liquid nitrogen.

The most promising crystals, EcYejA grown in 0.1 M MES pH 6.5, 12% w/v polyethylene glycol 20,000 in sitting nanodrops, were stored in liquid nitrogen and transported to the DIAMOND light source, Didcot for X-ray diffraction data collection on beamline i04-1. Data collection was performed remotely by Dr. Johan Turkenburg.

Data were processed using Xia2 followed by Aimless (Evans and Murshudov, 2013), a data reduction pipeline in CCP4i2 (Potterton *et al.*, 2018). The PDB structure 4ONY for an uncharacterised substrate binding protein from *Brucella melitensis* (Uniprot ID C0RL96) was used for molecular replacement. Searches with the complete 4ONY molecule did not give a satisfactory solution. Therefore, the 4ONY structure was separated into two domains and molecular replacement calculations were performed firstly using residues 1-261 and 545-580. This gave a convincing solution which was refined using REFMAC5 (Murshudov, Vagin and Dodson, 1997). This solution was fixed while a second molecular replacement calculation was performed using residues 262-544. This gave a solution which was later

refined in REFMAC5 (Murshudov, Vagin and Dodson, 1997) and gaps in the model filled in using BUCCANEER (Cowtan, 2006). Iterative rounds of model building using COOT (Emsley and Cowtan, 2004) and refinement in REFMAC5 (Murshudov, Vagin and Dodson, 1997) were carried out. Figures were made using the CCP4mg software program.

2.11 *In vivo* sensitivity assays

2.11.1 Disc diffusion assays

1% LB agar plates were prepared along with a bottle of 0.7% LB agar, called top agar. Small discs of filter paper were created with a hole punch and autoclaved. Just before use the poured LB agar plates were warmed to 37 °C and the top agar was melted and cooled to ~55 °C. 3.5 ml aliquots of top agar were transferred to sterile 15 ml falcon tubes and kept at 55 °C until use. Bacterial culture from overnights was added to the aliquot of top agar to a final OD_{600} of 0.05 and mixed thoroughly by pipetting. The inoculated top agar was then immediately poured onto the pre-warmed 1% agar plates. The plates were then allowed to cool and dry on the bench. Once dry, up to 4 filter paper discs were placed evenly spaced on the top agar using sterile tweezers. A total volume of 5 µl of antibiotic at various concentrations was added to the filter paper discs before the plates were incubated at 37 °C overnight and imaged the next day.

2.11.2 Determination of Minimum Inhibitory Concentrations (MICs) of CAMPs and antibiotics with *E. coli* BW25113 in liquid culture

5 ml of LB with various concentrations of antibiotics and CAMPs were inoculated to a final OD_{600} of 0.05 with overnight cultures of *E. coli* BW25113. Bacterial cultures were grown overnight at 37 °C with shaking at 220 rpm. The following day the OD_{600} of the cultures was taken and the MICs of the CAMPs and antibiotics was calculated as an approximate decrease in OD_{600} of 40-50% in the stationary phase between cultures with and without CAMPs and antibiotics.

2.11.3 Plate reader sensitivity assays

1.7 ml of sterile MilliQ water was added to the reservoirs on the outer most edge of the 96-well plate (Thermo Scientific Nunclon Delta surface) and 200 µl of sterile MilliQ water was added to outer wells (column 1, column 12, row A, row H). 195 µl of LB and CAMPs/antibiotics at different concentrations were added to the wells and 200 µl of LB only was added to wells to generate blanks. Overnight bacterial cultures were diluted in LB to create a final OD_{600} of 0.4, 5µl of this was added to the 195 µl of media and

CAMPs/antibiotics in the wells giving a final OD_{600} in the wells of 0.01. The plate was then incubated in the plate reader (Bioteck epoch 2 microplate reader, Gen5 3.00 software) at 37 °C for 48 hours with double orbital constant shaking (282 cpm, 3mm) whilst taking OD_{600} readings every 30 minutes.

2.11.4 Shake flask sensitivity assays

20 ml of LB with CAMPs/antibiotics at different concentrations was added to a 100 ml conical flask. The flasks were inoculated to an OD_{600} of 0.01 (exact OD_{600} stated) with overnight bacterial cultures and grown at 37 °C with shaking at 150 rpm for 48 hours, or the length of time is otherwise stated. OD_{600} readings of the cultures were taken at regular intervals (exact times stated).

Chapter Three

Expression of Putative CAMP Binding Proteins

3. Expression of Putative CAMP Binding Proteins

3.1 Phylogenetic analysis of Cluster C substrate binding proteins (SBPs)

SapA and YejA are SBPs of ABC transporters and each is hypothesised to bind CAMPs, however the mechanism of binding is not yet known (Parra-Lopez, Baer and Groisman, 1993; Mason *et al.*, 2006; Eswarappa *et al.*, 2008; Rinker *et al.*, 2012; Wang *et al.*, 2016). In keeping with the accepted structure based classification of SBPs, both SapA and YejA fall into the Cluster C SBPs as they contain an extra domain compared to other SBPs (Berntsson *et al.*, 2010). To better understand how the proteins relate to other Cluster C SBPs, a limited phylogenetic analysis was undertaken. This type of analysis can give insights into the possible ligands of the SBPs SapA and YejA that are both hypothesised to bind CAMPs. Depending on where SapA and YejA fall within the phylogenetic tree could either lend weight to this hypothesis or provide suggestions on a more fitting hypothesis.

To create the phylogenetic tree the protein sequences of a number of Cluster C SBPs were collected. The initial focus was on Cluster C SBPs (i) whose structures had been solved (ii) had been referred to in previous literature (iii) all the Cluster C SBPs from the organisms (*E. coli* K12, *S.* Typhimurium LT2, *L. lactis* MG1363, *H. influenzae* Rd, *B. subtilis* strain 168, *B. melitensis* biotype 1 strain 16M) where Sap or Yej had been functionally studied or the structure of the SBP had been solved. Cluster C SBPs have been shown to bind a range of different substrates e.g. DppA is known to bind dipeptides, *L. lactis* OppA has been shown to bind peptides up to 35 residues long and NikA is known to bind both Nickel and haem (Manson *et al.*, 1986; Detmers *et al.*, 2000; Heddle *et al.*, 2003; Shepherd, Heath and Poole, 2007). The collated SBPs were then aligned in Jalview using the Muscle alignment algorithm and then an unrooted Maximum Likelihood phylogenetic tree was created from the alignment using Ugene.

3.1.1 SapA is DppA-like

The phylogenetic tree (Figure 3.1) shows the SapA proteins positioned very close to the DppA proteins (36% identity between *E. coli* SapA and *E. coli* DppA) and the *H. influenzae* haem binding protein HbpA (37% identity between *E. coli* SapA and *H. inflenzae* HbpA), DppA has been shown to bind dipeptides (Manson *et al.*, 1986). A sequence alignment of some of the Cluster C SBPs from the phylogenetic tree indicated that there were four conserved cysteine residues between *E. coli* SapA and *E. coli* DppA (Figure 3.2). When a PHYRE model (Figure 3.3A) of *E. coli* SapA was created it was noticed that the four
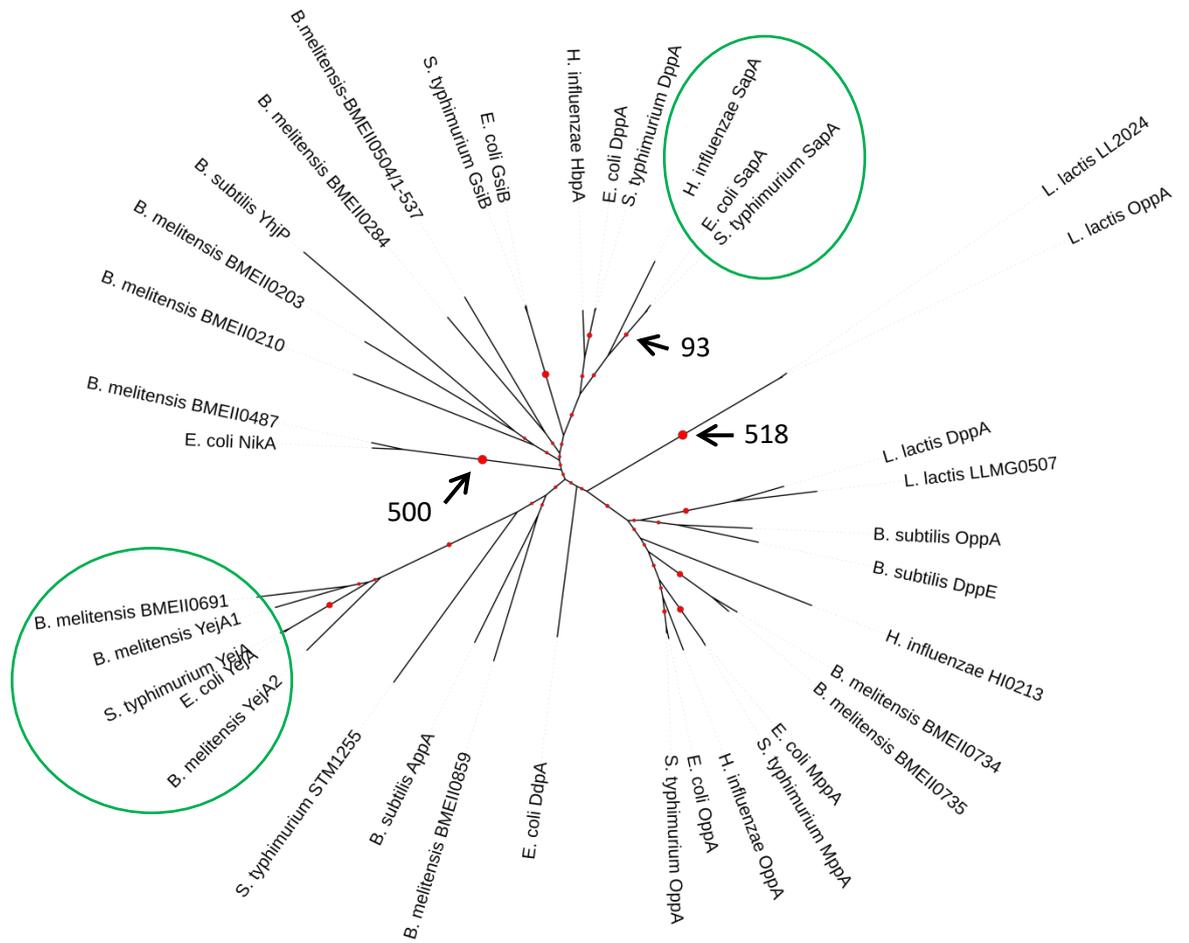
Figure 3.1. Phylogenetic analysis of Cluster C SBPs shows novel clade for YejA proteins.

Unrooted maximum likelihood phylogenetic tree populated with a series of Cluster C SBPs. E. coli = *E. coli* K12, S. typhimurium = *S.* Typhimurium LT2, L. lactis = *L. lactis* MG1363, H. influenzae = *H. influenzae* Rd, B. subtilis = *B. subtilis* strain 168, B. melitensis = *B. melitensis* biotype 1 strain 16M. Green circles indicate the SapA and YejA proteins. Red circles on the tree indicate bootstrapping. The larger the red circle the higher the bootstrapping value, some values have been given as a guide.

```
EcDppA

                  1                    10                                              20
EcDppA    MRIS......LK....KSGMLK.LGLS.........................LVAMTVAASVQ.A
EcSapA    MRQV.....LSSLLVIAGLVSGQAIA.......................APESPPHADIR.D
EcOppA    MTNI......TKRSLVAAGVLAALMAGN.V....................ALAADVPAGVTLAEK
EcYejA    ..........MIVRILLLFIALFTFGVQAQAIKESYAFAVLGEPRYAFNFNHFDYVNP...AAPKG
LlOppA    MNKLKVTLL.ASSVVLAATLLSACGSNQ.SSSTSTKKLKAG......NF.DVAYQNP.DKAIK.G
BsAppA    MKRRKTALMMLSVLMVLAIFLSACSGSK.SSNSSAKKSA...................GKPQQ.G


EcDppA          β1              η1          α1        α2              β2       β3      β4
                30         40          50          60                70       80
EcDppA    KTLVYCSEGSPEGFNPQLFTSGTTYD.ASSVPLY..........NRLVEF.KIGTTEVIPGLAEK
EcSapA    SGFVYCVSGQVNTFNPSKASSGLIVD.TLAAQFY..........DRLLDV.DPYTYRLMPELAES
EcOppA    QTLVRNNGSEVQSLDPHKIE.GVPES.NISRDLF..........EGLLVS.D.LDGHPAPGVAES
EcYejA    GQITLSALGTFDNFNRYALR.GNPG..ARTEQLY..........DTLFTTSDDEPGSYYPLIAES
LlOppA    GNLKVAYQSDSPMKAQW..LSGLSND.....ATFATMSGPGGGQDGLFFT.D.SGFKFIKGGAAD
BsAppA    GDLVVGSIGEPTLFNSL..YSTDDASTDIENMLYSFLTKTDEKLNVKLSL.A.ESIKELDG..L..


EcDppA                    β5                              α3           η2       α4
                    90        100         110         120         130         140
EcDppA    WEVSEDGKTYTFHLRKGVKWHDNKEFKPTRELNADDVVFSFDRQKNAQNPYHKVSGGSYEYFEGM
EcSapA    WEVLDNGATYRFHLRRDVPFQKTDWFTPTRKMNADDVVFTFQRIFDRNNPWHNVNGSNFPYFDSL
EcOppA    WDNK.DAKVWTFHLRKDAKWSDGT......PVTAQDFVYSWQRSVDPNTASPYASYLQYGHIAGI
EcYejA    ARYADDYSWVEVAINPRARFHDGS......PITARDVEFTFQKFMTEGVPQFRLVYK........
LlOppA    VALDKESKTATITLRKDLKWSDGS......EVTAKDYEFTYETIANPAYG..SDRWTDS..LANI
BsAppA    ......GLAYDVKIKKGVKFHDGK......LTADDVVFTYSVPLSKDYK..GERGSTYEMLKSV


EcDppA            α5    β6          β7          α6    η3      α7
                 150        160         170         180         190
EcDppA    .........GLPELISEVKKVD.DN....TVQFVLTRPEAPFLADLAMDFASILSKEYADA..MM
EcSapA    .........QFADNVKSVRKLD.NH....TVEFRLAQPDASFLWHLATHYASVMSAEYARK..LE
EcOppA    D...EILEGKKPITDLGVKAID.DH....TLEVTLSEPVPYFYKLLVHPSTSYVPK...AA..IE
EcYejA    .........G...TTVKA......IAPLTVRIELAKPGKED..MLSLFSLPVFPEKYWKDH..
LlOppA    VGLSDYHTG.KAKTISGITFPDGENGKVIKVQFKEMKPGMTQSG.NGYFLETVAPYQYLKDVAPK
BsAppA    E...................KKGDYEVLFKLKYKDG..NFYN.NALDSTAILPKHILGNVPIA


EcDppA          α8          β8        β9                          β10
                200       210         220         230         240         250
EcDppA    KAGTPEKLDLNPIGTGPFQLQQYQKDSRIRYKAFDGYWGTKPQID.......TLV.FSITPDASV
EcSapA    KEDRQEQLDRQPVGTGPYQLSEYRAGQFIRLQRHDDFWRGKPLMP.......QVV.VDLGSGGTG
EcOppA    KFGEKWTQPGNIVTNGAYTLKDWVVNERIVLERSPTYWNNAKTVI.......NQVTYLPIASEVT
EcYejA    ..KLSDPLATPPLASGPYRVTSWKMGQNIVYSRVKDYWAANLPVNRGRWNFDTIR.YDYYLDDNV
LlOppA    DLASSPKTTTKPLVTGPFKPENVVAGESIKYVPNPYYWGEKPKLN.......SIT.YEVVSTAKS
BsAppA    DLEENEFNRKKPIGSGPFKFKEWKQGQYILKLEANDDYFEGRPYLD.......VT.YKVIPDANA


EcDppA      α9              η4       α10       β11         β12
                 260        270         280         290         300
EcDppA    RYAKLQKNECQVMPYPNPAD....IARMKQDKSINLMEMPG...LNVGYLSYN.........VQ
EcSapA    RLSKLLTGECDVLAWPAASQ....LSILRDDPRLRLTLRPG....MNVAYLAFN.........TA
EcOppA    DVNRYRSGEIDMTNNSMPIEL...FQKLKKEIPDEVHVDPY....LCTYYYEIN.........NQ
EcYejA    AFEAFKAGAFDLRMENDAKNWATRYTGKHKDYKYIIKDEQKNESAQDTRWLAFN.........IQ
LlOppA    VAAL.SSSKYDIINGMVSSQ....YKQVKNLKGYKVLGGQQAMYISLMYYNLGHYDAKNSINVQDR
BsAppA    AVAQLQAGDINFF.NVPATD....YKTAEKFNNLKIVTDLA....LSYVYIGWN.E........K


EcDppA              α11             α12         β13
                   310        320         330         340
EcDppA    KKPLDDVKVRQALTYAVNKDAIIKAVYQGAGVSAKNLIPP...................TMWGY.
EcSapA    KPPLNNPAVRHALALAINNQRLMQSIYYGTAETAASILPR...................ASWAY.
EcOppA    KPPFNDVRVRTALKLGMDRDIIVNKVKAQGNMPAYGYTPP....YTDGAK..LTQPEWFGW.
EcYejA    RPVFSDRRVREAITLAFDFEWMNKALFYNAWSRTNSYFQNTEYAARNYPDAAELVLLAPMKKDLP
LlOppA    KTPLQDQNVRQAIGYARNVAEVDNKFSNGLSTPANSLIPP..................IFKQF.
BsAppA    NELFKDKKVRQALTTALDRESIVSQVLDGDGEVAYIPESP...................LSWNY.


EcDppA                  α13             β14
                   350        360         370         380         390
EcDppA    .........NDDVQ...DYTYDPEKAKALLKEAGLEKGFSIDLWAM...PVQRPYNP...NARRM
EcSapA    .........DNEAK...ITEYNPAKSREQLKSLGLE.NLTLKLWVP...TRSQAWNP...SPLKT
EcOppA    .........SQEK.......RNEEAKKLLAEAGYTADKPLT........INLLYNT...SDLHK
EcYejA    SEVFTQIYQPPVSKGDGYDRDNLLKADKLLNEAGWVLKGQQRVNATTGQPLSFELLLPASSNS..
LlOppA    .........T.SSSVKGYEKQDLDKANKLLDEDGWKLNKSTGYREKDGKELSLVYAARVGDANAE
BsAppA    .........PKDIDVPKF.EYNEKKAKQMLAEAGWKDTNGDGILDKDGKKFSFTLKTNQGNKVRE
```

```
            α14              β15            α15              β16           α16
EcDppA  ℓℓℓℓ..ℓℓℓℓ.ℓ      ━━━━━━▶    ℓℓℓℓℓℓℓℓℓ      ━━━━━━▶    ℓℓℓℓℓℓℓ
             400           410            420              430         440

EcDppA  AEMI..QADW.AKVGVQAKIVT...YEWGEYLKRAK....DGEHQTVMMGWTGDNGDPDNFFATL
EcSapA  AELI..QADM.AQVGVKVVIVP...VEGRFQEARLM....DMSHDLTLSGWATDSNDPDSFFRPL
EcOppA  KLAIAASSLWKKNIGVNVKLVN...QEWKTFLDTRH....QGTFDVARAGWCADYNEPTSFLNTM
EcYejA  QWVLPFQHSL.QRLGINMDIRK...VDNSQITNRMR....SRDYDMMPRVWRAMP.WPSSDLQIS
LlOppA  TIAQNYIQQW.KKIGVKVSLYNGKLMEFNSWVDHMTTPPGANDWDITDGSWSLAS.EPS..QQDL
BsAppA  DIAVVVQEQL.KKIGIEVKT...QIVEWSALVEQMNPP.NWDFDAMVMGWSLST.FPD..QYDI
                                                                     ▲
```

```
          α17               α18                    α19                β17
EcDppA  ℓ ℓℓℓℓℓℓ      ℓℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ         ━━━━━━━
        450       460         470          480          490        500       510

EcDppA  FSCAASEQGSNYSKWCYKPFEDLIQPARATD..DHNKRVELYKQAQVVMHDQAPALIIAHSTVFE
EcSapA  LSCAAIHSQTNLAHWCDPKFDSVLRKALSSQ..QLAARIEAYDEAQSILAQELPILPLASSLRLQ
EcOppA  LSNS....SMNTAHYKSPAFDSIMAETLKVT..DEAQRTALYTKAEQQLDKDSAIVPVYYYVNAR
EcYejA  WSSEYINSTYNAPGVQSPVIDSLINQIIAAQGNKE.KLLPLGRALDRVLTWNYYMLPMWYMAEDR
LlOppA  FSAAAP...YNFGHFNDSEITKDLNDIDSAKSENPTYRKAAFVKYQEDMNKKAYVIPTNFMLNYT
BsAppA  FHSSQIKKGLNYVWYKNAEADKLMKDAKSI.SDRKQYSKEYEQIYQKIAEDQPY.....TFLYY
          ▲            ▲
```

```
            β18            β19
EcDppA  ━....━▶  ━▶      ━▶
               520      530

EcDppA  P.....VRKEVKGYV.VDPLGKHHFENVSIE...............
EcSapA  A.....YRYDIKGLV.LSPFGNASFAGVYREKQDEVKKP........
EcOppA  L.....VKPWVGGYTGKDPLDNTYTRNMYIVKH..............
EcYejA  LAW.....WDKFSQPAVRPIYSLGIDTWWYDVNKAAKLPSASKQGE
LlOppA  PVNKRVV.....GM....TLDYGAMNTWSEIGVSSAKLATK.....
BsAppA  PNNHMAMPENLEGYKYHPKRDLYNIEKWWLAK..............
```

Figure 3.2. Alignment of a select few of the Cluster C SBPs in the phylogenetic tree.

EcDppA = *E. coli* DppA, EcSapA = *E. coli* SapA, EcOppA = *E. coli* OppA, EcYejA = *E. coli* YejA, LlOppA = *L. lactis* OppA, BsAppA = *B. subtilis* AppA, further strain information listed in Table 2.1. EcDppA secondary structure is shown along the top of the alignment. Blue arrows below alignment show conserved cysteine residues between EcDppA and EcSapA. Green arrow below alignment shows the aspartate capping residue of the binding pocket.
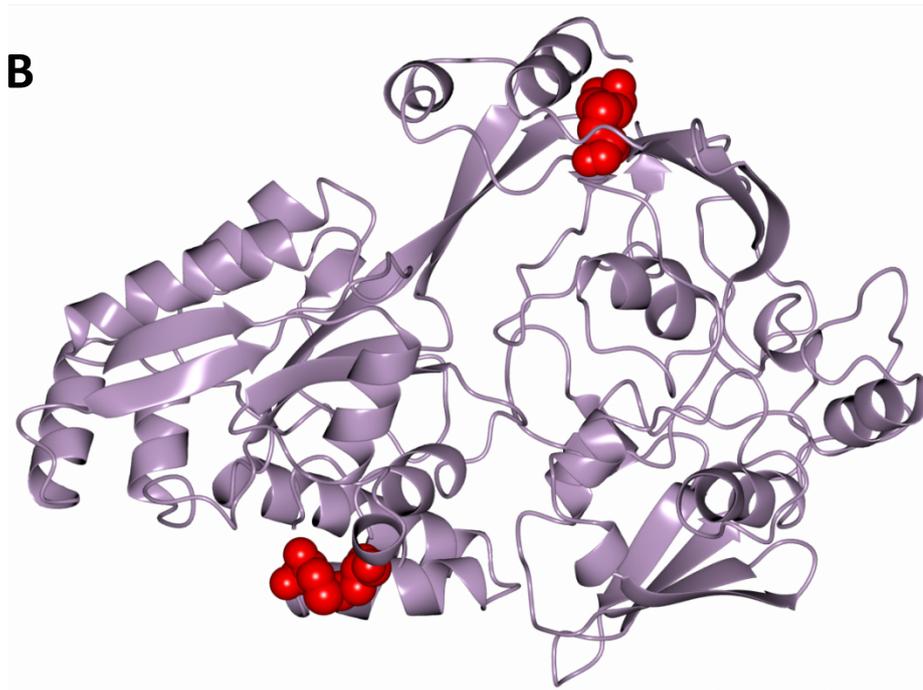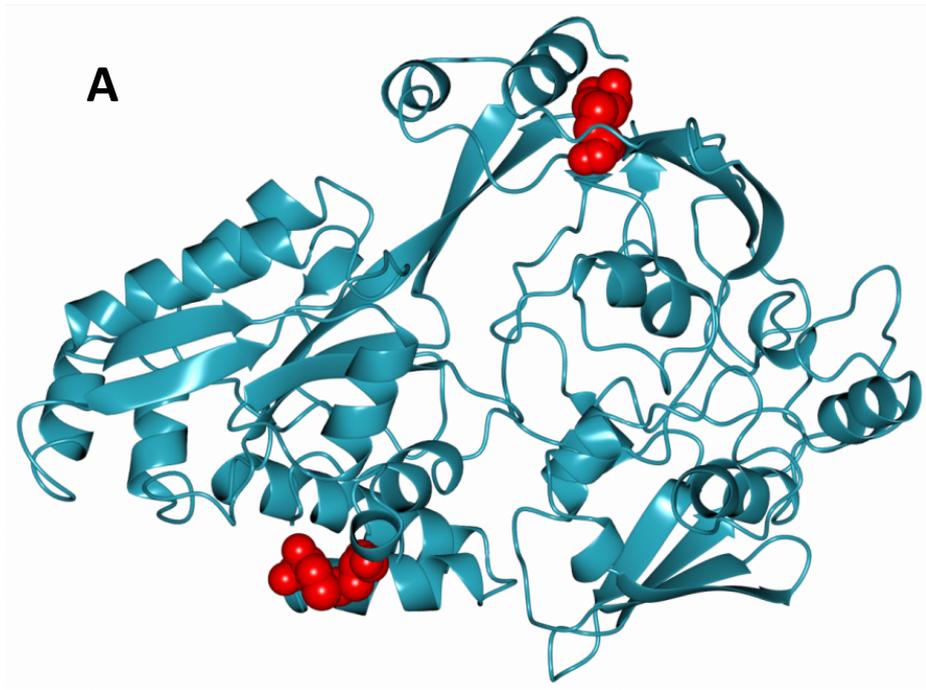
Figure 3.3. PHYRE model predicts two disulphide bonds in EcSapA and crystal structure of *E. coli* DppA shows two disulphide bonds.

PHYRE model of EcSapA (A) is shown in blue ribbon form, the four cysteine residues forming the two disulphide bonds are shown in red spheres. Crystal structure (coordinates 1DPE) of *E. coli* DppA (B) is shown in lilac ribbon, the four cysteine residues forming the two disulphide bonds are shown in red spheres.

conserved cysteines created two disulphide bonds in the predicted structure, similar to the disulphide bonds in known structures of DppA (Figure 3.3B) (Nickitenko, Trakhanov and Quiocho, 1995; Tame *et al.*, 1995; Levdikov *et al.*, 2005). The phylogenetic data and the similarities between SapA and DppA suggest that SapA is a dipeptide binding protein, however recent literature gives many examples of how this may not be the case. Via genetic and biochemical analysis SapA has been shown to be involved in resistance to CAMPs in various different bacterial species (Parra-Lopez, Baer and Groisman, 1993; Mason *et al.*, 2006; Rinker *et al.*, 2012). As CAMPs are much longer than dipeptides, commonly 20-30 residues, this literature indicates that SapA may be more than just a dipeptide binding protein.

3.1.2 YejA is part of a novel clade

Unlike the SapA proteins, the YejA proteins are not positioned close to the DppA family, nor the OppA/MppA families of oligopeptide SBPs, on the phylogenetic tree. YejA forms a novel clade on the phylogenetic tree with the clade positioned closer to, but distinct from, *B. subtilis* AppA which has been shown to bind longer peptides (Levdikov *et al.*, 2005). The evidence demonstrated here seems to correlate more closely with the hypothesis that YejA is able to bind CAMPs rather than the hypothesis that SapA is able to bind CAMPs.

Structures of some Cluster C SBPs have shown the presence of an aspartate residue in the binding pocket which appears to "cap" the end of the binding pocket and serve as an anchor for the amino terminus of peptide ligands. This aspartate forms a salt bridge with the alpha amino group of the peptide ligand and therefore prevents a longer peptide binding, "capping" the binding site. In the sequence alignment (Figure 3.2) this aspartate can be seen in EcDppA, EcSapA and EcOppA. However, in Cluster C SBPs which have been shown to bind longer peptides, such as BsAppA and LlOppA, this aspartate residue is not conserved. This presence or omission of the capping aspartate residue can therefore be used to predict whether a Cluster C SBP binds longer or shorter peptides. Using this prediction tool it is possible to hyphthesise that EcYejA has the ability to bind longer peptides as it is predicted to not contain the capping aspartate residue. Likewise, it is possible to predict that due to the predicted presence of the capping aspartate residue in EcSapA, it only binds shorter peptides, i.e. not full length CAMPs as initially hypothesised.

From the phylogenetic analysis, initially *E. coli* SapA (EcSapA) and *E. coli* YejA (EcYejA) were chosen as targets as they are represented in the phylogenetic tree and they appear to be good representatives of all the SapA and YejA proteins.

3.2 Cloning and expression of SapA

Commonly genes encoding SBPs are cloned into vectors which produce the desired protein fused to short in-frame tags to aid in the purification process, for example hexahistidine tags (PORATH *et al.*, 1975). In this study the gene *sapA* (Table 2.1, Chapter 2)*,* encoding a putative CAMP SBP, was cloned into the pETFPP_30 vector for expression and further purification.

The pETFPP_30 vector is based on the pET22b vector (Figure 2.1, Chapter 2) and contains an optional PelB leader sequence (Table 3.1), which targets the protein to the periplasm, and a cleavable C-terminal hexahistidine tag (Figure 3.4A). The hexahistidine tag is cleavable by human rhinovirus (HRV) 3C protease. *sapA* was PCR amplified from *E. coli* K12 genomic DNA and the resulting fragments were cut using restriction endonucleases then assembled into the pETFPP_30 vector via HiFi DNA assembly. Different restriction endonucleases were used to create different constructs, for constructs with a PelB leader sequence NcoI and XhoI were used, for constructs with a native leader sequence or no leader sequence NdeI and XhoI were used. All constructed plasmids were verified firstly by colony PCR to establish the presence of an insert of expected length in the plasmid (Figure 3.5). Plasmids containing the desired insert size were then sent for DNA sequencing of the open reading frame. The resulting sequence was checked to ensure the gene was in-frame and that no mutations had been introduced.

Initial small scale overexpression trials were carried out on SapA constructs to determine whether recombinant protein was being produced. Small scale overexpression trials consisted of inoculating 50 ml of LB in a conical flask with the bacterial cells harbouring the various DNA constructs. The flasks were inoculated to different starting OD_{600} between 0.01-0.2, incubated at different temperatures and for different lengths of time. The cells were induced with different concentrations of IPTG to try and find the conditions which produced the highest yield of soluble protein. Sample preparations for SDS-PAGE gels were as depicted in Figure 2.2, Chapter 2.

The cytoplasmic EcSapA protein construct was received from Dr. Tim Rasmussen (University of Aberdeen) and its solubility was assessed in small scale overexpression trials. The small scale overexpression trials were carried out at 30 °C and 37 °C and induced with 1 mM IPTG. Unfortunately, SapA was not soluble under these conditions, however an accumulation of protein at the expected molecular weight for SapA can be seen overnight

Name of leader sequence	Sequence of leader sequence	Predicted leader sequence in SignalP
PelB Leader Sequence	MKYLLPTAAAGLLLLAAQPAMA	✓
Native EcSapA Leader Sequence	MRQVLSSLLVIAGLVSGQAIA	✓
Native StSapA Leader Sequence	MRLVLSSLIVIAGLLSSQATA	✓
Native HiSapA Leader Sequence	MLRLNLRFLSFLLCIIQSVELQA	✓

Table 3.1. PelB and Native leader sequences for EcSapA, StSapA and HiSapA.

Further strain and protein information is listed in Table 2.1. The PelB leader sequence is derived from the vector, pETFPP_30, whereas the native signal sequences are unique to each protein from the different species. Both signal sequences target the proteins to the Sec translocation pathway for export to the periplasm.
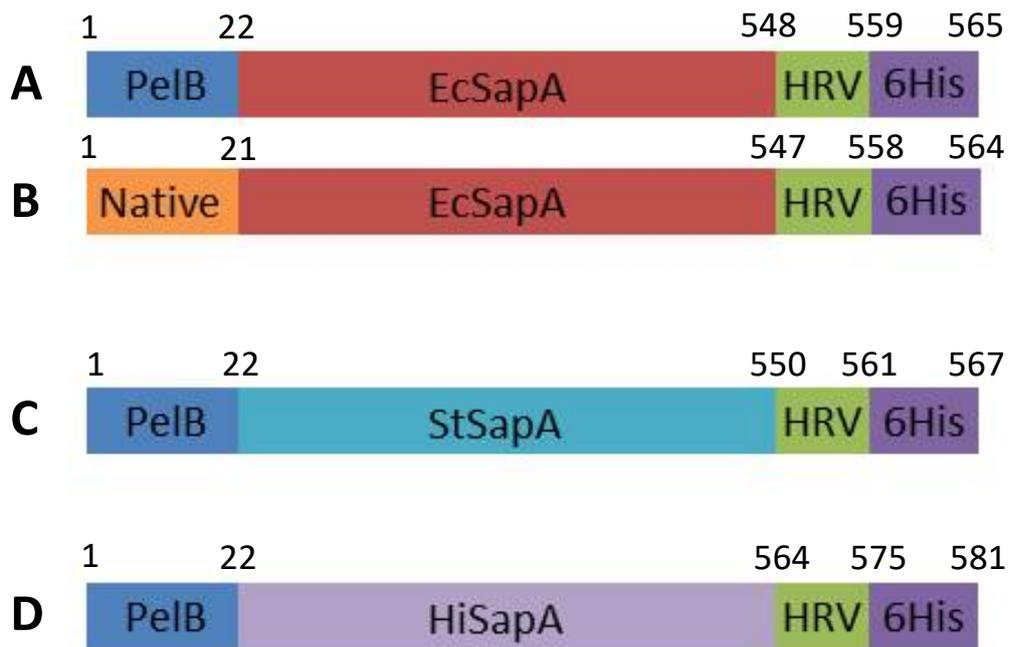
Figure 3.4. Schematic of SapA constructs used.

PelB and Native refer to the leader sequences used, HRV is the HRV 3C cleavage point and 6His is the hexahistidine tag used. Residue number of bordering amino acids are noted above constructs.
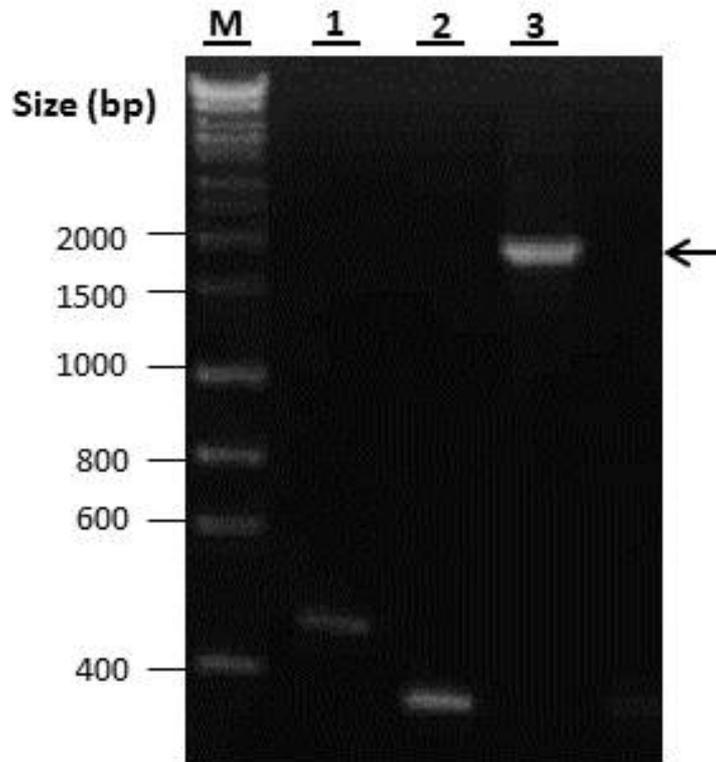
Figure 3.5. Colony PCR showing the expected size insert for cytoplasmic *ecYejA*.

M = Marker; 1, 2, 3 = Number of picked colony from cytoplasmic *ecYejA* transformation plate. A band is shown in lane 3 at the expected size (1804 bp) for cytoplasmic *ecYejA*, indicating this colony has been transformed correctly and contains plasmid DNA with the correct insert. Lanes 1 and 2 contain other colonies with DNA inserts of much lower size, indicating these colonies do not contain the correct *ecYejA* insert DNA.

in the insoluble fraction (Figure 3.6). Cytoplasmic expression of SapA yielded very low quantities of soluble protein, which may be due to the predicted disulphide bonds in SapA. Disulphide bonds, which form in oxidising conditions, are commonly important for the structure of proteins. The cytoplasm of *E. coli* is a reducing environment where disulphide bonds are not expected to form, however the periplasm has an oxidising environment allowing the formation of disulphide bonds.

Due to the insolubility of EcSapA when cytoplasmic overexpression was trialled, expression of the SapA protein in the periplasm was tested using various leader sequences (Table 3.1). The range of *sapA* genes was also broadened to include *sapA* from *S.* Typhimurium LT2 (StSapA) and *H. influenzae* (HiSapA), in attempts to increase the chances of producing a soluble version of the SapA protein. Initially *sapA* from all three species were cloned into the pETFPP_30 vector to create constructs that included both the PelB leader sequence for transport to the periplasm and the cleavable C-terminal hexahistidine tag (Figure 3.4A, 3.4C, 3.4D). The primers used to create the constructs were EcPSapA-F and EcPSapA-R for EcSapA; StPSapA-F and StPSapA-R for StSapA and HiPSapA-F and HiPSapA-R for HiSapA.

As with the cytoplasmic expression of EcSapA, SapA from all three organisms when expression cells were grown at 37 °C and induced with 1 mM IPTG produced little to no soluble protein (Figure 3.7). Although the amounts of protein loaded in the soluble lanes of Figure 3.7 were very low, as indicated by the lack of many protein bands in those lanes, when the soluble fractions were run through a Nickel affinity column a very small peak was eluted. This indicated that there was very little soluble SapA being produced and therefore it was not practical to continue with these constructs.

A variety of approaches were employed to overcome the insolubility of SapA. First different growth temperatures of the cultures were tested. Often lowering the temperature of bacterial cultures reduces the speed at which they replicate and therefore the speed and quantity of protein they produce (Schein and Noteborn, 1988). This can help with the solubility of proteins by reducing the workload of the cell and therefore avoiding protein aggregation. Temperatures were tested between 20 °C and 37 °C, none showed any improvement in solubility across all three SapA proteins (Figure 3.8). Protein bands on the SDS-PAGE gels at the expected molecular weight for SapA are not seen in any of the soluble fractions but large bands are seen at the expected molecular weight in the insoluble fractions. This implies that the majority of SapA protein that is being produced is insoluble.
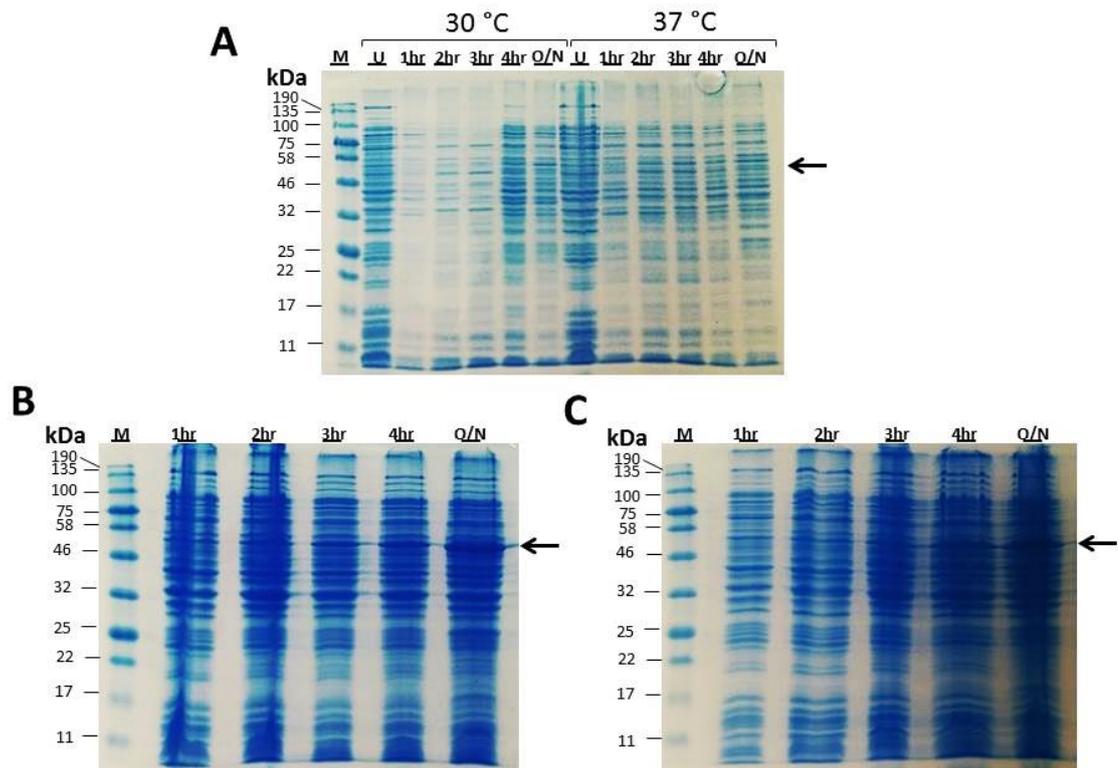
Figure 3.6. EcSapA is not soluble in the cytoplasm at 30 °C or 37 °C.

12% polyacrylamide SDS-PAGE gels stained with Coomassie blue dye. Cells were incubated at either 30 °C or 37 °C and recombinant protein production was induced with 1 mM IPTG. Gel A shows the soluble fractions at both 30 °C and 37 °C, B shows the insoluble fractions at 37 °C and C shows the insoluble fractions at 30 °C. M = Marker; U = Uninduced sample; 1, 2, 3, 4hr = Number of hours after induction that a soluble/insoluble sample was taken; O/N = Overnight soluble/insoluble sample. No accumulating band of protein at the expected molecular weight for SapA (59.4 kDa) can be seen in the soluble samples gel. There is a band at the expected molecular weight for SapA in Gel B, the band appears to increase in relative intensity through the time points indicating accumulation of insoluble protein.
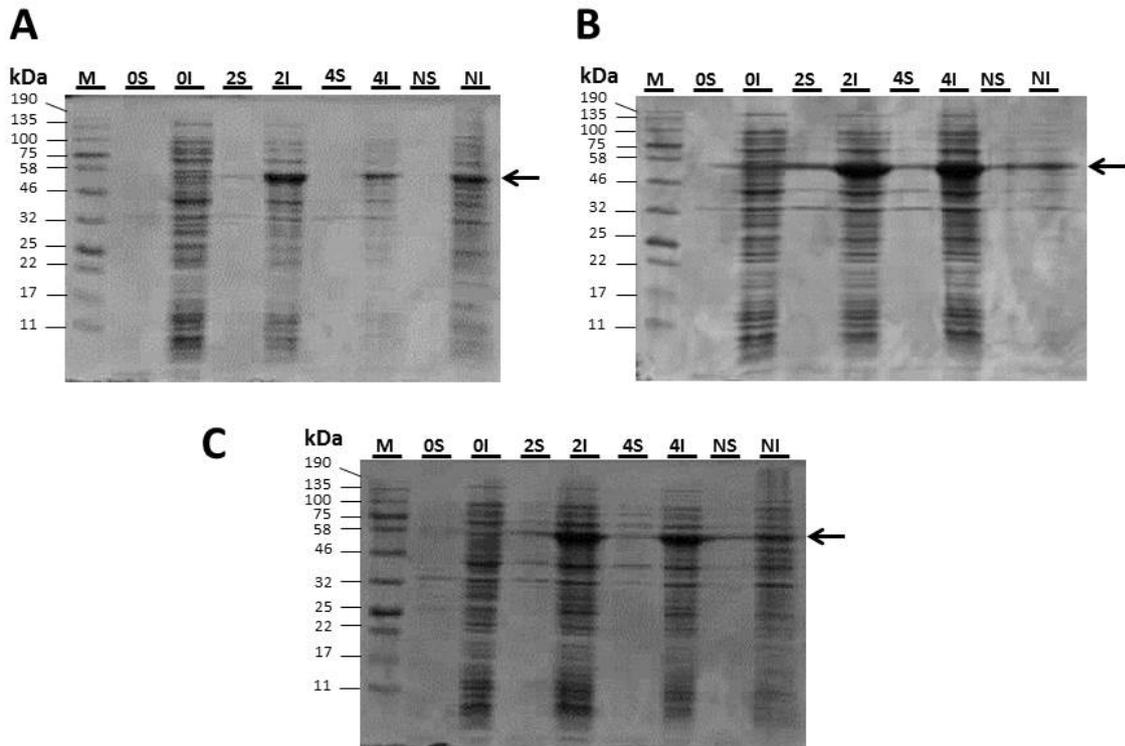
Figure 3.7. EcSapA, StSapA and HiSapA with a PelB leader sequence are insoluble at 37 °C when induced with 1 mM IPTG.

12% polyacrylamide SDS-PAGE gels stained with Coomassie blue dye. M = Marker; 0, 2, 4S = Number of hours after induction that a TSE periplasmic sample was taken; 0, 2, 4I = Number of hours after induction that a TSE post-periplasmic extraction insoluble sample was taken; NS = Overnight soluble sample; NI = Overnight insoluble sample. A = EcSapA, B = StSapA, C = HiSapA. All protein bands of the expected molecular weight for SapA (EcSapA = 59.4 kDa, StSapA = 59.5 kDa, HiSapA = 61.8 kDa) are in the insoluble fractions of the samples not the soluble fractions.

Figure 3.8. SapA with a PelB leader sequence is insoluble at different temperatures.

12% polyacrylamide SDS-PAGE gels stained with Coomassie blue dye. Images A and B are of EcSapA, C and D are of StSapA and E and F are of HiSapA. Overexpression trials show in images A, C and E were carried out at 25 °C and trials shown in images B, D and F were carried out at 37 °C. M = Marker; 1 = Uninduced sample; 2 = Pre-harvest sample; 3 = Soluble sucrose fraction; 4 = Soluble periplasmic fraction; 5 =Post-periplasmic extraction soluble sample; 6 = Blank; 7 = Post-periplasmic extraction insoluble sample. All of the gels show an increase in protein at the expected molecular weight for SapA (EcSapA = 59.4 kDa, StSapA = 59.5 kDa, HiSapA = 61.8 kDa) in the Pre-harvest whole cell sample as compared to the uninduced whole cell sample. This indicates that the process of inducing expression of SapA is functioning correctly.

Secondly a range of different IPTG concentrations were tested in small scale overexpression trials. IPTG induces the production of the target gene cloned into a pET vector by binding to the Lac repressor and causing its dissociation from the Lac operator site adjacent to the pT7 promoter, allowing transcription of the encoded gene. By titrating the concentration of IPTG it is possible to control the rate of protein production. Much like the control of temperature, lowering the rate of protein produced lowers the burden on the cell and therefore helps ensure that the protein that is produced is processed properly (Donovan, Robinson and Click, 1996). IPTG concentrations were tested in the range 0.1 mM – 1 mM, unfortunately changing the concentration of IPTG did not have an impact on the solubility of any of the three SapA proteins (Figure 3.9). In each case protein was being made, but it was insoluble. There are no bands in the soluble periplasmic fraction to suggest soluble expression, however there are bands in the pre-harvest whole cell sample at the predicted molecular weight for SapA suggesting overexpression of insoluble protein. This shows that the lower concentration of IPTG is not aiding the solubilisation of SapA in the periplasm.

Following on from the unsuccessful expression of SapA with a PelB leader sequence, the next strategy employed was to try and express EcSapA solubly in the periplasm using the protein's native signal sequence (Table 3.1, Figure 3.4B). The construct was cloned into pETFPP_30 using primers EcNSapA-F and EcNSapA-R and contained the EcSapA native signal sequence and an HRV 3C cleavable hexahistidine tag. The native signal sequence of SapA works in the same way as the PelB leader sequence, it targets the protein for secretion to the periplasm via the Sec translocase.

Again, different growth temperatures and IPTG concentrations were trialled, none of these produced soluble SapA (Figure 3.10). The uninduced samples in Figure 3.10A have a small band of protein at the predicted molecular weight for SapA, this may be due to leaky expression in the cells. However this band is not seen in the subsequent soluble samples after induction, indicating that the protein band seen in the pre-induction sample is insoluble. In Figure 3.10 gels B and C a large band of protein is shown at the predicted molecular weight for SapA and increases in size as the timepoints increase, indicating accumulation of protein. The protein that is accumulated seems to be insoluble due to it appearing in the insoluble fractions and it not being seen in the soluble fractions.

As a control, Shewana3-2073 protein was expressed. It has been previously shown that it is possible to successfully extract soluble Shewana3-2073 protein from the periplasm of *E.*
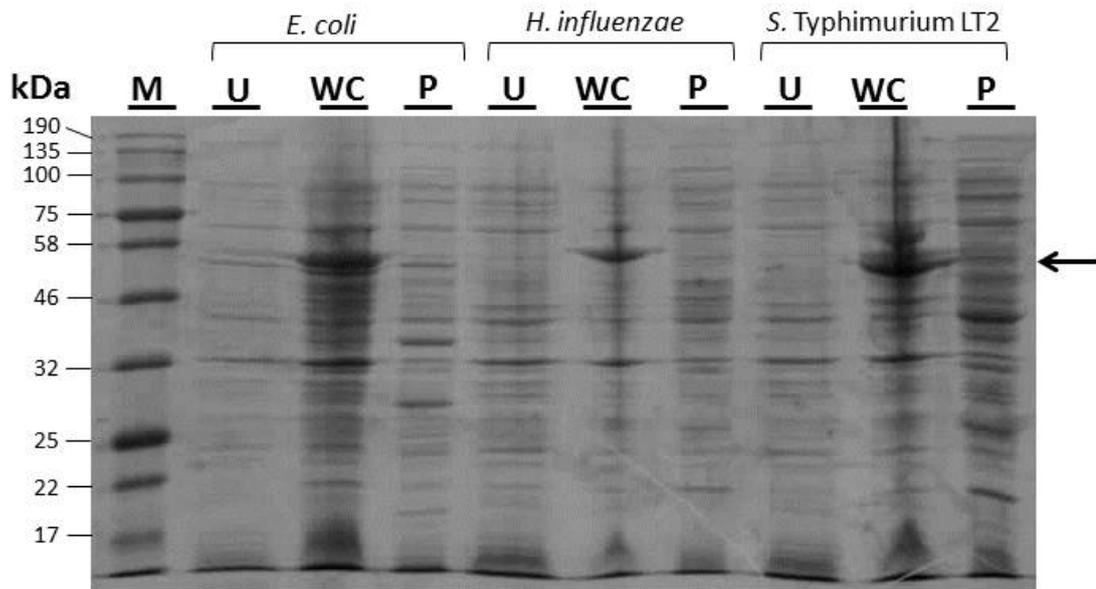
Figure 3.9. SapA with a PelB leader sequence is insoluble when induced with 0.8 mM IPTG.

12% polyacrylamide SDS-PAGE gel stained with Coomassie blue dye. M = Marker; U = Uninduced sample; WC = Pre-harvest sample; P = Post-periplasmic extraction soluble fraction. Cells were grown at 25 °C and recombinant protein production was induced overnight with 0.8 mM IPTG. In the pre-harvest samples there is an increase in protein in one of the bands, this is slightly lower than the expected molecular weight of SapA (EcSapA = 59.4 kDa, StSapA = 59.5 kDa, HiSapA = 61.8 kDa).
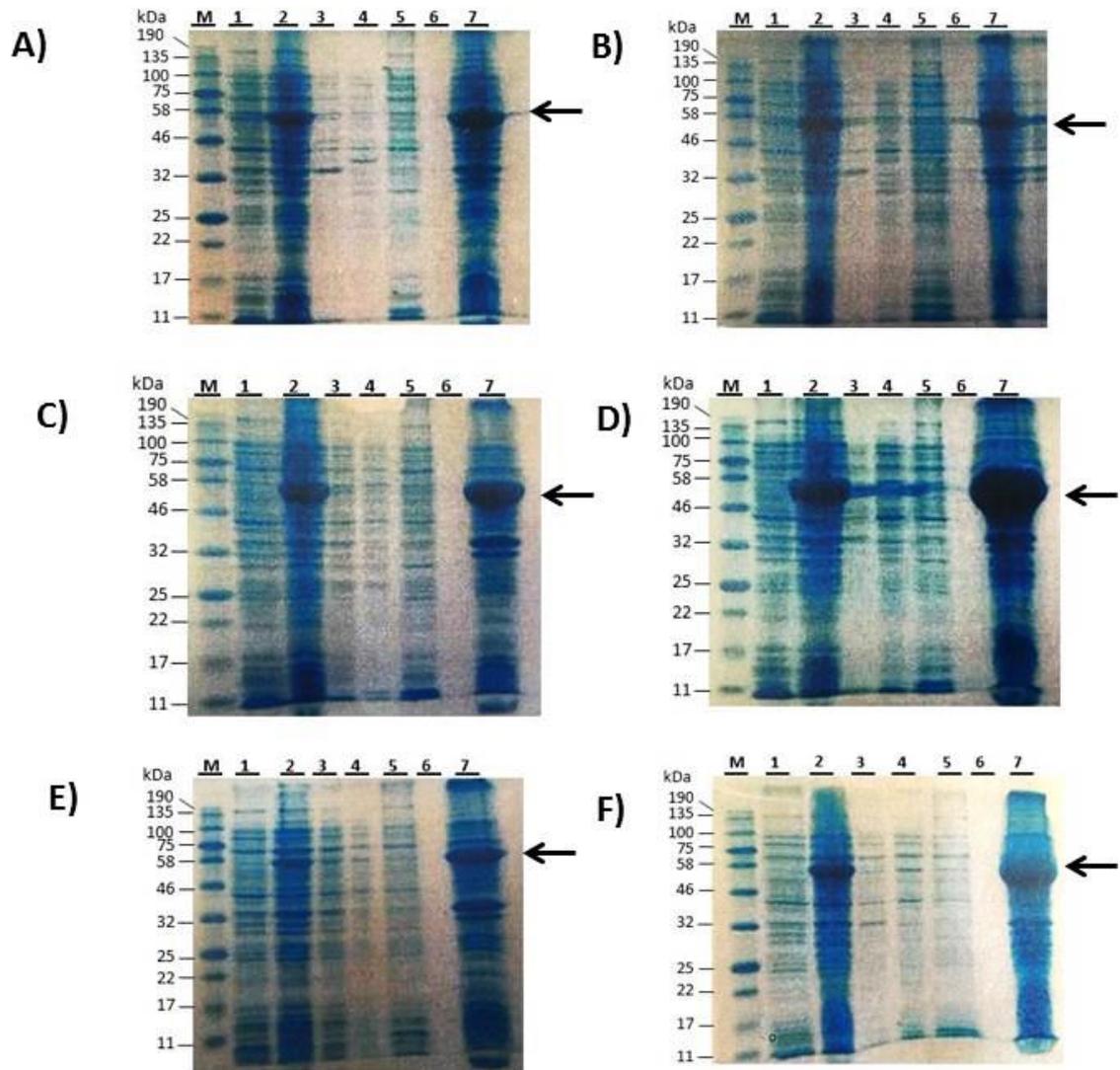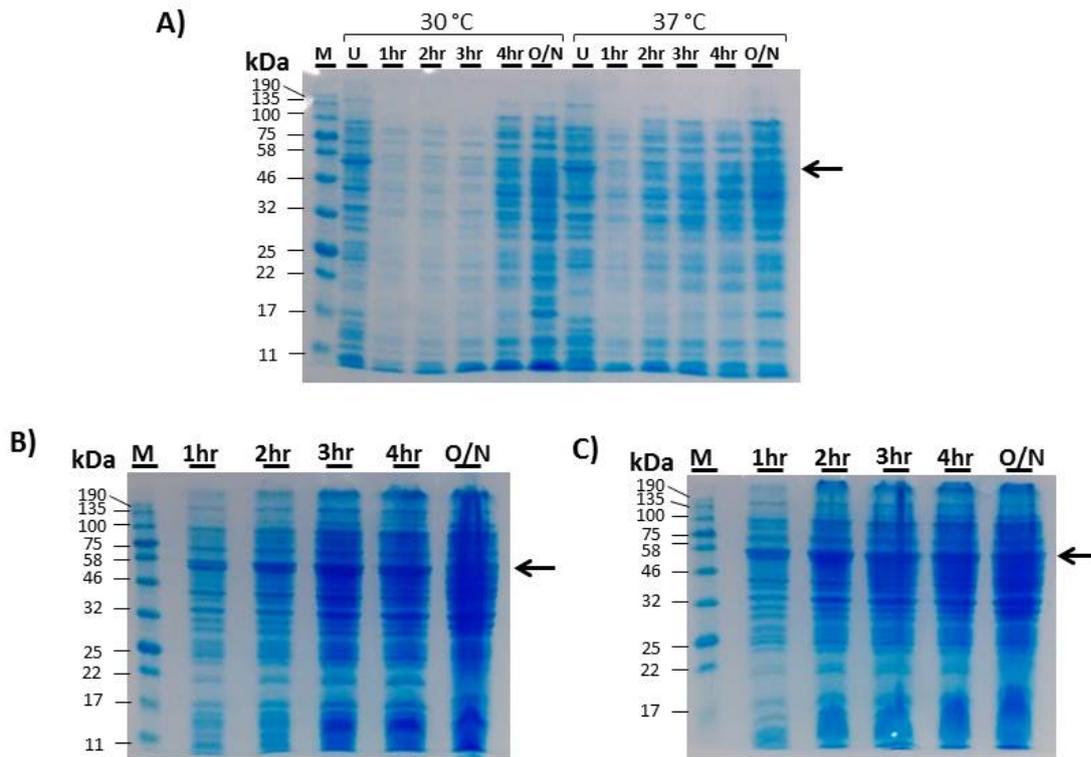
Figure 3.10. EcSapA with native leader sequence is insoluble.

12% polyacrylamide SDS-PAGE gels stained with Coomassie blue dye. A) depicts soluble protein when cells are grown at 30 °C and 37 °C, B) shows insoluble protein when cells are grown at 30 °C and C) shows insoluble protein when cells are grown at 37 °C. M = Marker; U = Uninduced sample; 1, 2, 3, 4hr = Number of hours after induction that a soluble/insoluble sample was taken; O/N = overnight soluble/insoluble sample.

coli cells using the ICOS method (Drousiotis, 2017). Shewana3-2073 is an SBP for an ABC transporter from *Shewanella ANA-3,* Shewana3-2073 has a molecular weight of ~32.5 kDa and a single disulphide bond. Shewana3-2073 was used as a positive control for periplasmic expression and extraction of protein alongside SapA (Figure 3.11). The band of protein shown slightly higher than that of Shewana3-2073 in lanes 1, 5 and 7 could possibly be Shewana3-2073 with an uncleaved signal sequence as these samples are taken from the whole cell and the signal sequence is only cleaved upon transport to the periplasm. This gel displays that it is possible to extract soluble Shewana3-2073 from the periplasm via the ICOS method, as indicated by the large bands of protein in lanes 3 and 4. However it was not possible to do the same with SapA, leading to the conclusion that the problem lies with the solubility of SapA, not the methodology employed to extract the protein.

This work shows SapA from different bacterial species is insoluble when expressed in the cytoplasm, with a PelB leader sequence or with a native leader sequence at different temperatures and IPTG concentrations.

Work was carried out on the SapA protein for almost a year when it was learned (8/8/2016) that the research group of Dr. Martin Walsh had experienced similar problems with the solubility and expression of SapA over a longer time period. They had overcome their problems and moreover had solved the structure of SapA (personal communication). Due to this new information and in the interests of time it was decided to continue on solely with EcYejA.

3.3 Cloning and expression of EcYejA

As with the SapA constructs EcYejA was cloned into the pETFPP_30 vector (Table 2.1 Chapter 2). Two different EcYejA constructs were created with the pETFPP_30 vector. One with an N-terminal PelB leader sequence and a C-terminal histidine tag, targeting EcYejA for expression in the periplasm, using primers EcPYejA-F and EcPYejA-R. The second construct, created using primers EcYejA-F and EcYejA-R, directed the synthesis of the protein without a leader sequence but with a C-terminal histidine tag and so targeted EcYejA for expression in the cytoplasm.

After DNA cloning and verification, small scale overexpression trials were carried out. Expression of the EcYejA construct lacking a leader sequence produced a large quantity of soluble protein, particularly after growth overnight at 30 °C (Figure 3.12). As this construct of EcYejA did not contain a signal sequence it is assumed that the EcYejA protein produced
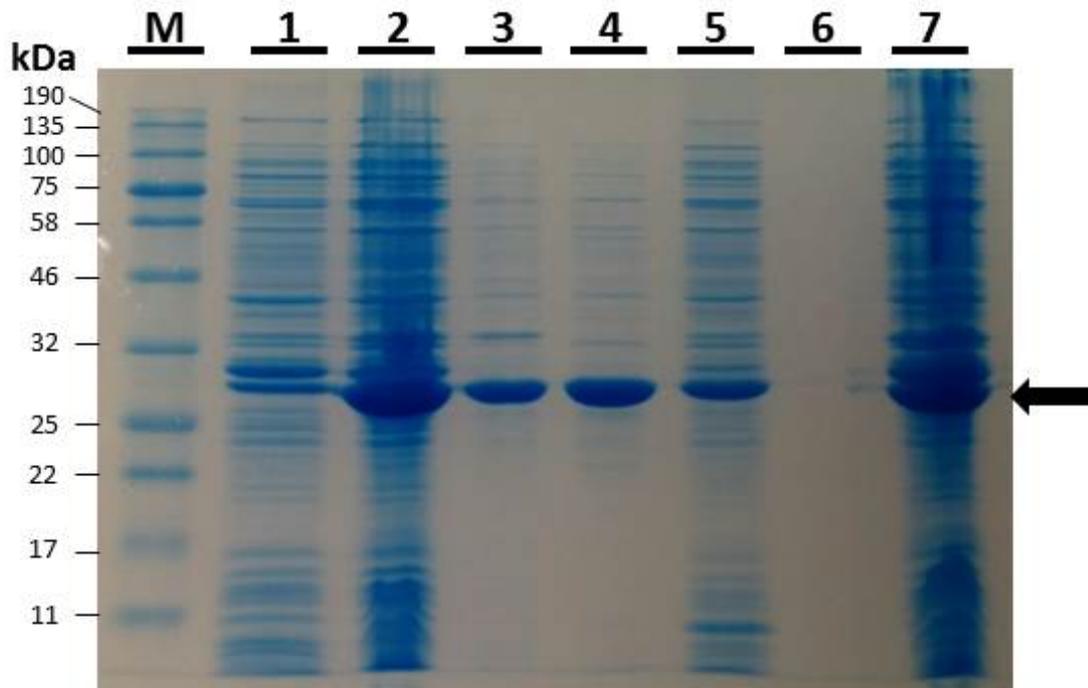
Figure 3.11. Overexpression and retrieval of soluble Shewana3-2073 from the periplasm of *E. coli*.

12% polyacrylamide SDS-PAGE gel stained with Coomassie blue dye. M = Marker; 1 = Uninduced sample; 2 = Pre-harvest sample; 3 = soluble sucrose fraction; 4 = soluble periplasmic fraction; 5 = Post-periplasmic extraction soluble sample; 6 = blank lane; 7 = Post-periplasmic extraction insoluble sample. Shewana3-2073 was expressed in BL21 (DE3) star cells at 37 °C and recombinant protein production was induced with 1 mM IPTG overnight. Shewana3-2073 protein (32.5 kDa) is indicated by the black arrow.
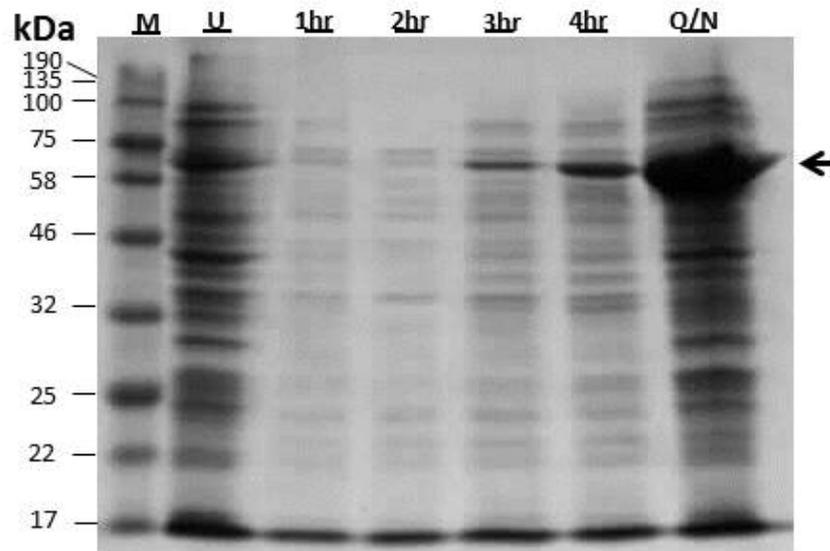
Figure 3.12. EcYejA can be produced as a soluble protein.

12% polyacrylamide SDS-PAGE gel stained with Coomassie blue dye. Cells were grown at 30 °C and recombinant protein production was induced with 1 mM IPTG, all samples were boiled at 95 °C in SDS for 5 mins before loading on the gel. M = Marker; U = Uninduced whole cell sample; 1, 2, 3, 4hr = Number of hours after induction that a soluble sample was taken; O/N = Overnight soluble sample. A band showing an increasing amount of protein at each sample time point can be seen, detailed by the arrow, at the expected molecular weight for EcYejA (68.4 kDa).

is accumulating in the cytoplasm. As with the cytoplasmic expression of EcYejA the expression of EcYejA with the PelB leader sequence was soluble and a large enough quantity was produced to carry out further data analysis (Figure 3.13). Due to the ease of purifying EcYejA from the cytoplasm, as opposed to the additional steps needed to purify from the periplasm, it was decided to remain with cytoplasmic expression and purification of EcYejA for future work. The verified plasmid encoding cytoplasmic EcYejA encodes a protein consisting of residues 2-586 of mature *E. coli* YejA fused N-terminally to a methionine residue and C-terminally to an HRV 3C cleavable hexahistidine tag.

Having established the success of the small scale overexpression trials with EcYejA, large scale overexpression and purification was carried out using the EcYejA construct lacking the leader sequence (cytoplasmic accumulation of EcYejA). SBPs can be subjected to a 2 M guanidinium-HCl treatment whilst bound to a Nickel affinity column (Lanfermeijer *et al.*, 1999). This treatment allows the release of any pre-bound ligand that might have purified with the protein from the cell extract.

3.4 Purification of EcYejA

EcYejA protein was eluted from the Nickel affinity column (Figure 3.14) and dialysed overnight into SEC buffer with HRV 3C protease (1:200 HRV 3C:protein ratio) to remove the cleavable C-terminal hexahistidine tag. Removal of the tag can aid in crystallisation as the hexahistidine tag is expected to be flexible and can therefore hinder the crystallisation process. The following day the dialysed EcYejA was subjected to size exclusion chromatography to further purify EcYejA (Figure 3.15). Another very faint band can be seen on the SDS-PAGE gel (Figure 3.15) running at approximately 46 kDa, this could be EcYejA degradation or other impurity. Due to the faintness of the band, therefore indicating low concentration compared to EcYejA, it was decided that this probably would not interfere with downstream experimentation. The resulting protein (~90 mg/L of LB liquid culture) was used for biochemical assays and crystallisation.
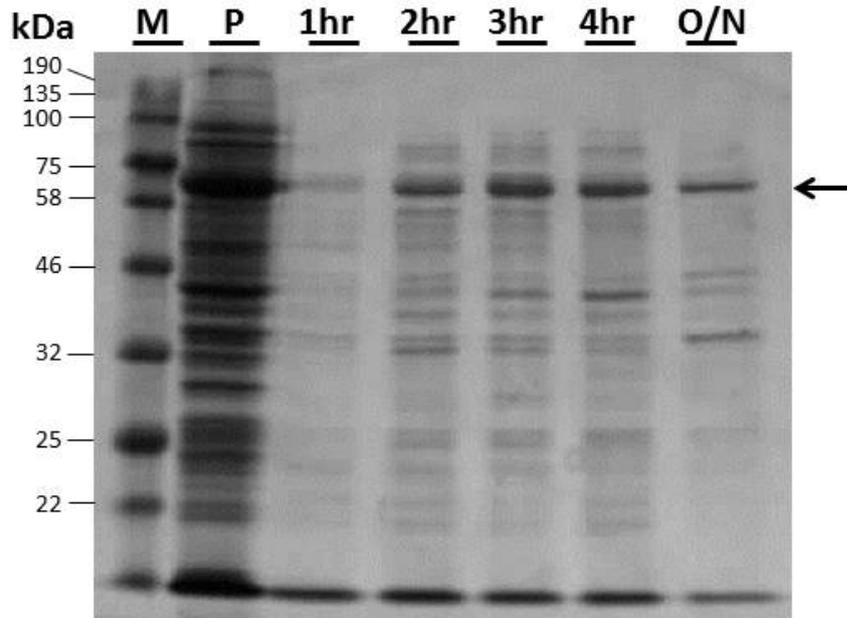
Figure 3.13. EcYejA fused to a PelB leader sequence is soluble.

12% polyacrylamide SDS-PAGE gel stained with Coomassie blue dye. M = Marker; P = Pre-harvest whole cell sample; 1, 2, 3, 4hr = Number of hours after induction that a soluble sample was taken; O/N = Overnight soluble sample. There is a band at the same molecular weight across the time points indicating that soluble EcYejA with a PelB leader sequence is being produced. The pre-harvest sample also shows a protein band at the expected molecular weight for EcYejA (68.4 kDa) with a PelB leader sequence, indicating leaky expression.
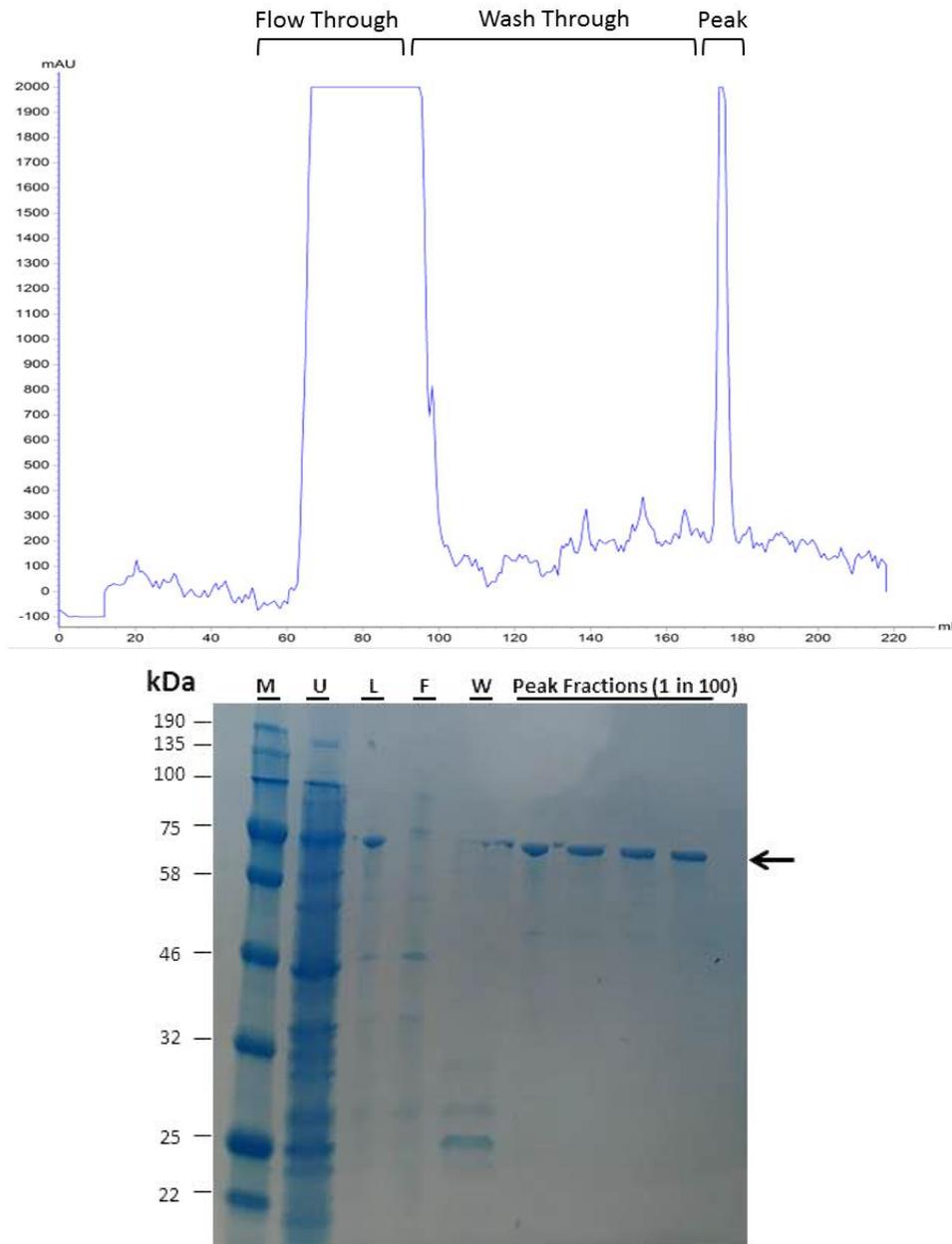
Figure 3.14. EcYejA can be purified by Nickel affinity chromatography.

12% polyacrylamide SDS-PAGE gels stained with Coomassie blue dye. Top shows the chromatogram of Nickel affinity purification of EcYejA, bottom shows the SDS-PAGE from the same purification. M = Marker; U = Uninduced sample; L = Load (1 in 100 dilution); F = Flow through (1 in 100 dilution); W = Wash through; Peak Fractions indicates fractions from the peak in the above chromatogram. There is no band on the gel in the flow through at the expected molecular weight for EcYejA (68.4 kDa) but there is a band at the expected molecular weight for EcYejA in the peak fractions, indicating EcYejA is binding to the Nickel column and eluting when imidazole is added to the column. The fraction samples are much purer than the load sample and are at the expected molecular weight for EcYejA.
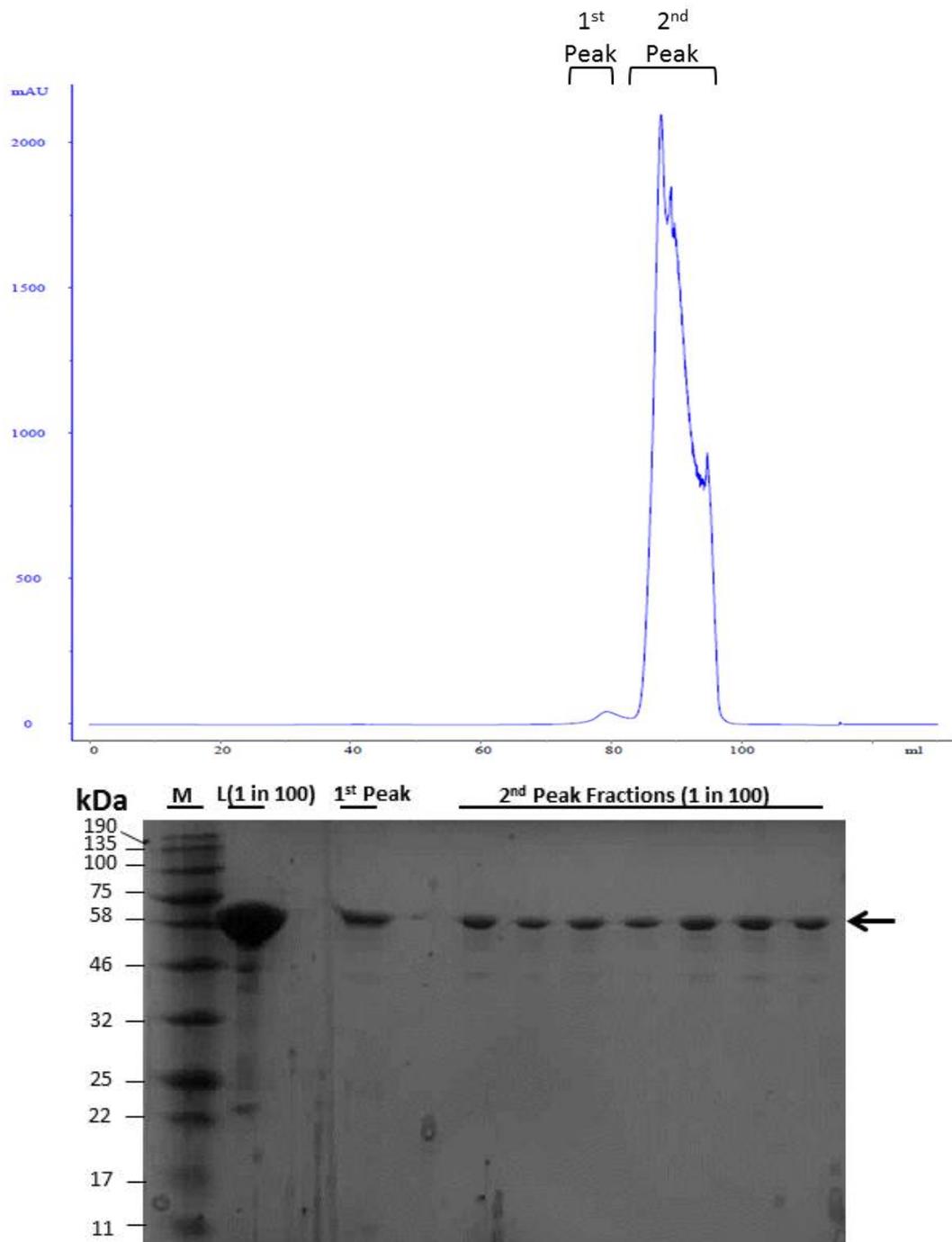
Figure 3.15. EcYejA can be further purified by Size Exclusion Chromatography.

12% polyacrylamide SDS-PAGE gels stained with Coomassie blue dye. Top shows the chromatogram of Size Exclusion purification of EcYejA, bottom shows the SDS-PAGE from the same purification. M = Marker; L = Load (1 in 100 dilution); Peak indicates fractions from the peak in the above chromatogram. The gel shows that the peak samples are purer than the load sample and are at the expected molecular weight for EcYejA. The 1st peak could be aggregation of EcYejA or a dimeric form.

Chapter Four

Biochemical Analysis and Structure Determination of EcYejA

4. Biochemical Analysis and Structure Determination of EcYejA

4.1 Biochemical analysis of EcYejA

After cytoplasmic expression and purification of EcYejA via a Nickel affinity column followed by an S200 SEC column EcYejA consisted of an N-terminal Methionine, residues 2-586 of the protein and a remaining LEVLFQ sequence at the C-terminus produced by HRV 3C protease cleavage. EcYejA was then subjected to further biochemical analysis and crystallisation to understand the properties of the protein in more detail.

Initially EcYejA was analysed by denaturing electrospray mass spectrometry, carried out by Dr. Andrew Leech, to establish that authentic and intact protein had been purified. The mass spectrometry equipment was calibrated using horse heart myoglobin and has an error of ± 1.5 Da. A clearly defined peak at 68375.6 Da can be seen (Figure 4.1), this peak was corrected using the horse heart myoglobin as a standard and gave a corrected value of 68383.6 Da. The expected molecular weight of EcYejA is 68382.42 Da, which is within the ± 1.5 Da error range of the instrument away from the corrected mass value. This indicates that the EcYejA protein has been purified and no alterations have been made during the expression or purification process.

To determine whether EcYejA was monomeric, dimeric or otherwise SEC-MALLS was carried out, under the supervision of Dr. Andrew Leech, on EcYejA using two different concentrations of EcYejA, 1 mg/ml and 3 mg/ml (Figure 4.2). The data indicate that the majority of EcYejA is monomeric. The majority of substrate binding proteins are monomeric and there was nothing to indicate that EcYejA would be any different to this. There is a small proportion of the protein that is dimeric.

During purification EcYejA was treated with 2 M guanidinium-HCl whilst bound to the Nickel affinity column, to determine whether EcYejA was folded following the treatment Circular Dichroism (CD) was used under the supervision of Dr. Andrew Leech. Proteins with α-helices produce CD spectra with negative troughs at 208nm and 222nm and a positive peak at 193nm. Proteins containing β-sheets have a positive peak at 195nm and a negative trough at 218nm. Disordered proteins however have a negative trough at 195nm and very low ellipticity above 210nm (Greenfield, 2006). A sample of untreated EcYejA was used as a control, as the spectra shows (Figure 4.3) both the treated and untreated EcYejA are within the range of α-helices and β-sheets and neither show any indication of random disordered coils. Both of the traces are more or less indistinguishable, the only difference being in the
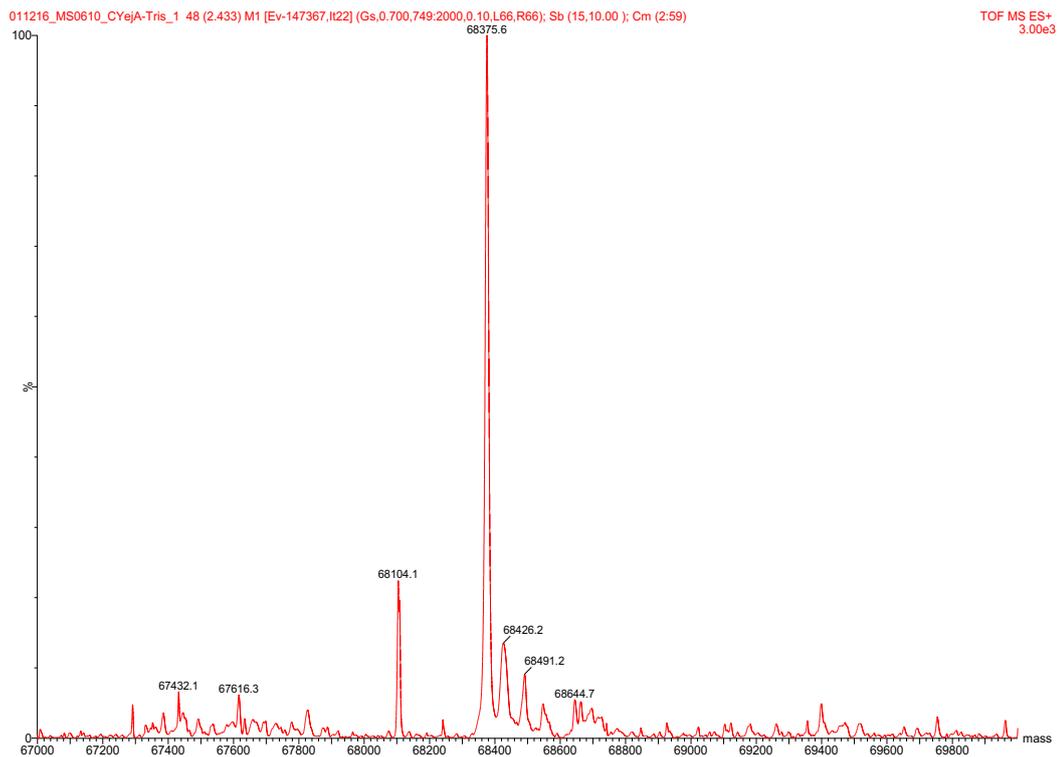
Figure 4.1. Denaturing electrospray mass spec shows EcYejA has been purified successfully.

The mass of the largest peak shown in the figure when corrected using horse heart myoglobin as a standard is 68383.6 Da, very close to the predicted molecular weight for EcYejA at 68382.42 Da and within the error range of the instrument.

111
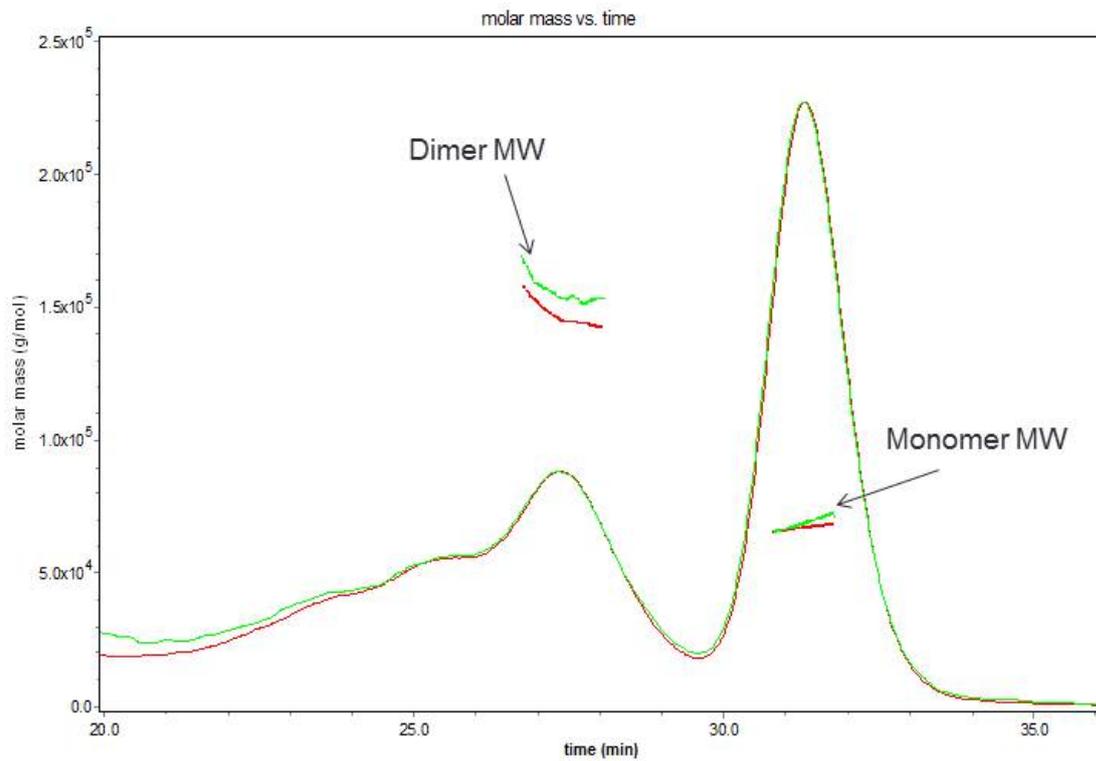
Figure 4.2. SEC-MALLS data shows EcYejA is mainly monomeric.

Green line indicates 1 mg/ml EcYejA, red line indicates 3 mg/ml EcYejA.
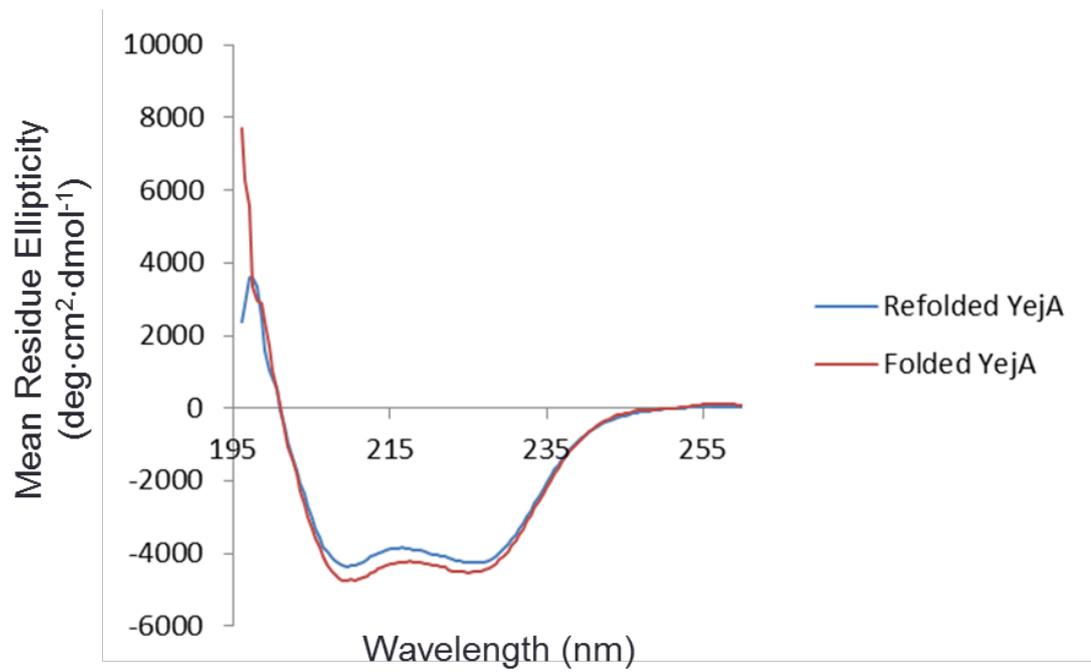
Figure 4.3. EcYejA treated with 2 M guanidinium-HCl is the same as untreated EcYejA.

EcYejA treated with 2 M guanidinium-HCl is shown by a blue line and EcYejA not treated

with 2 M guanidinium-HCl is shown via a red line.

concentrations of the two samples. This indicates that the 2 M guanidinium-HCl treatment has not had an effect on the structure of EcYejA.

4.2 Crystallisation of EcYejA

To allow X-ray diffraction and structure determination of EcYejA the protein first had to be crystallised. Purified EcYejA protein was put into crystallisation trials using a 96-well plate sitting drop format. The crystallisation screens used were JCSG+, PDB Min and PEG/ION with 10 mg/ml and 20 mg/ml EcYejA. Crystals appeared (Figure 4.4) from 0.2 M ammonium formate and 20% w/v polyethylene glycol 3350 at both 10 mg/ml and 20 mg/ml EcYejA. Before crystals were captured for X-ray diffraction and data collection they were preserved using a cryo protectant (30% glycerol, 14 µl mother liquor to a total of 20 µl). 1 µl of cryo protectant was added to each drop before fishing crystals from the drop using a nylon loop. The crystals were rapidly submerged in liquid Nitrogen for storage.

The crystals grown in 0.2 M ammonium formate and 20% w/v polyethylene glycol 3350 did not diffract well when sent to DIAMOND and gave two different crystal lattices, this is possibly due to the crystals growing in an overlapping way, which are very difficult to separate when fishing the crystals.

Several variations of the crystallisation conditions were tested in 24 well hanging drop plates with a reservoir volume of 1 ml. All of the resulting crystals had a very similar morphology (Figure 4.5), long thin overlapping rod like shapes. These crystals were very difficult to separate and mount for X-ray diffraction. A variety of different strategies were tested to try and slow the growth of the crystals and to improve their quality, these included adding DMSO and ethylene glycol to the reservoirs.

Approximately 1 month after setting up the 96-well sitting drop crystallisation plate of EcYejA, crystals were observed (Figure 4.6) in 12% w/v Polyethylene glycol 20,000 and 0.1 MES pH 6.5 at both 10 mg/ml and 20 mg/ml EcYejA. These crystals were usually single and of "chunkier" morphology. They were easy to mount and send to the DIAMOND Light Source for X-ray diffraction.

4.3 Structure determination of EcYejA

Data was collected remotely by Dr. Johan Turkenburg and Sam Hart on the i04-1 beamline at the DIAMOND Light Source for the crystal grown in 12% w/v Polyethylene glycol 20,000
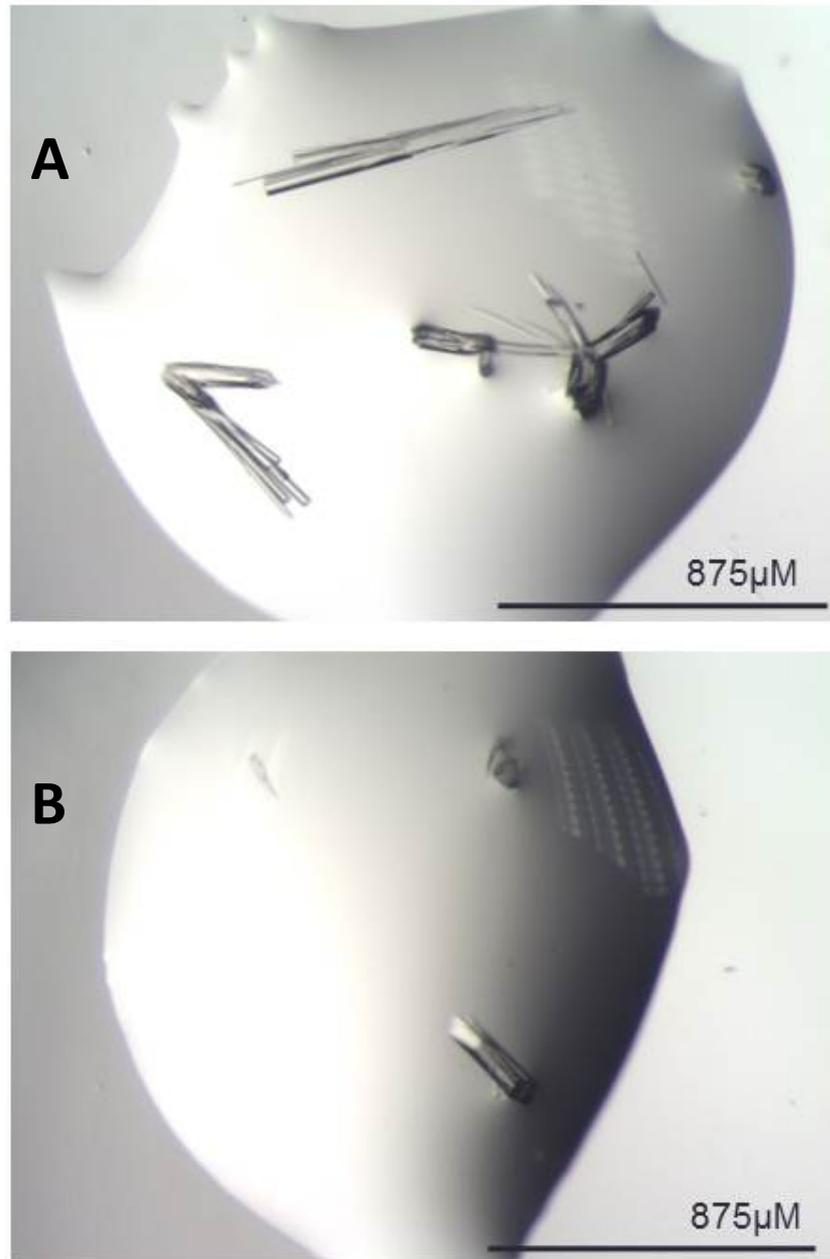
Figure 4.4. Crystals of EcYejA in 0.2 M ammonium formate and 20% w/v polyethylene glycol 3350.

(A) shows 20 mg/ml EcYejA and (B) shows 10 mg/ml EcYejA.
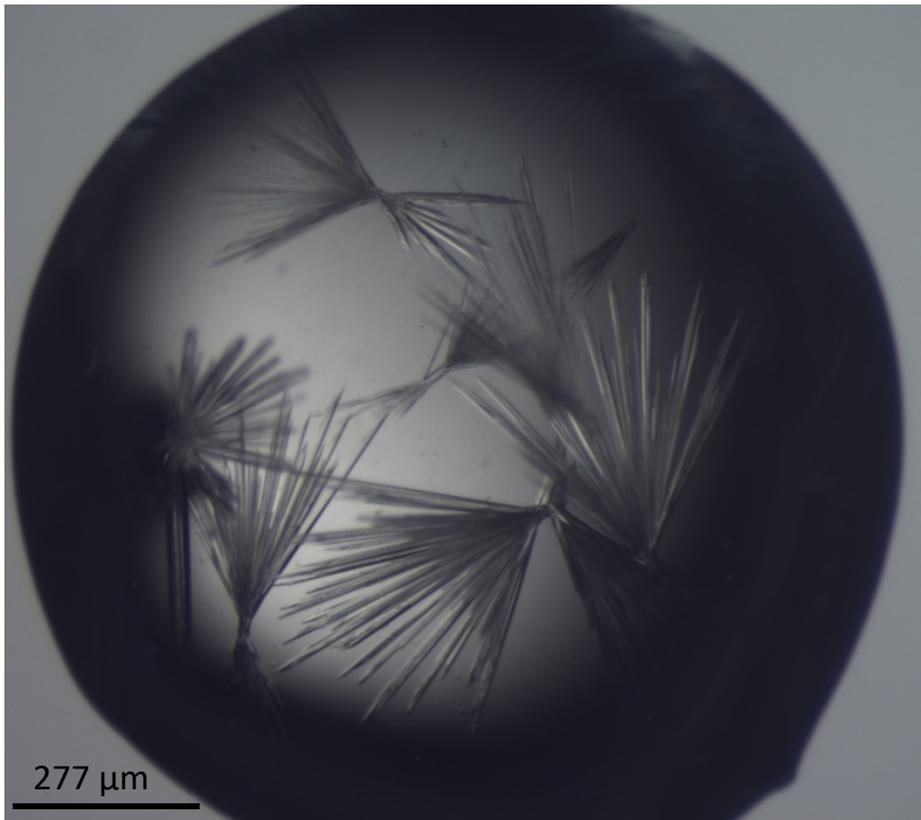
Figure 4.5. Crystals of EcYejA in 0.2 M ammonium formate, 1% DMSO and 25% PEG 3350.

The crystals have grown in a needle like formation.
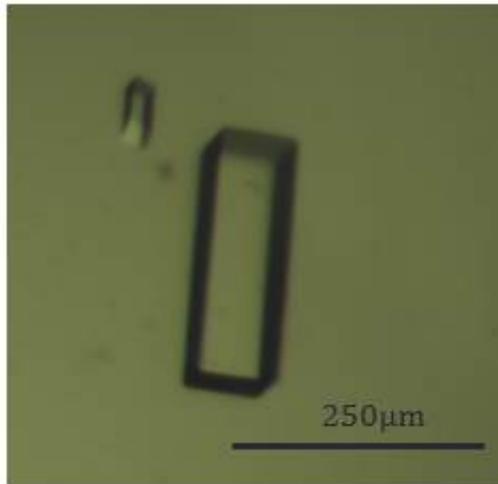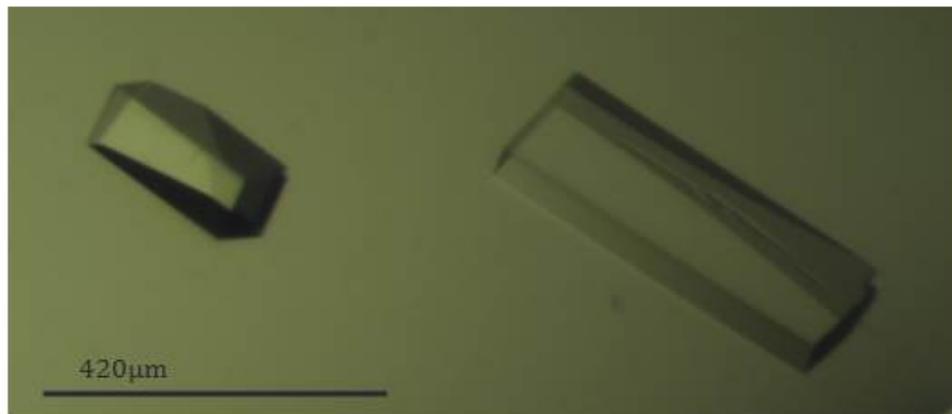
Figure 4.6. Crystals of EcYejA in 12% w/v Polyethylene glycol 20,000 and 0.1 M Mes 6.5.

(A) shows 20 mg/ml EcYejA and (B) shows 10 mg/ml EcYejA.

and 0.1 M Mes 6.5. Data processing was carried out via Xia2 and the space group was found to be I_{222} (Table 4.1).

The initial aim was to solve the structure of EcYejA by molecular replacement using the coordinate set with the PDB entry 4ONY. 4ONY was found by using BLAST and searching with the protein sequence of EcYejA whilst setting the results to search for only things deposited in the PDB, therefore meaning their structure is known. 4ONY was the top hit from this search meaning it had the most sequence similarity (32% identity) with EcYejA and therefore was a good choice for molecular replacement. In the PDB 4ONY is annotated as an SBP from *B. melitensis* and is in the apo form, again meaning it was a good choice for EcYejA as EcYejA was also expected to be in the apo form. SBPs go through large conformational changes upon ligand binding therefore meaning the difference between the apo/open and ligand bound/closed forms can be large so having a molecular replacement model in the same conformation can aid molecular replacement. As later discovered EcYejA was actually in the closed form and therefore molecular replacement was carried out in a slightly different way. As SBPs have two clear domains with the binding cleft residing in the interface between the two domains it was relatively simple to separate the two domains of 4ONY and use the two domains separately in molecular replacement calculations. Domain one of 4ONY was defined as residues 1-261 and 545-580, domain two was residues 262-544. Using Molrep (Vagin and Teplyakov, 2010), a molecular replacement tool in CCP4i2, it was possible to carry out molecular replacement with one domain of 4ONY, fix that domain in place and then carry out molecular replacement on the second domain. This approach was successful and following molecular replacement BUCCANEER (Cowtan, 2006) was used to build peptide chains between the two domains of the protein that were cut to carry out the molecular replacement. Following on from this step iterative rounds of REFMAC (Murshudov, Vagin and Dodson, 1997) and model building in COOT (Emsley and Cowtan, 2004) were carried out. The structure was judged to be solved when the R value was 0.16 and the Rfree was 0.19 (Table 4.1), at this point it was clear that EcYejA was in the closed conformation (Figure 4.7). The structure shows EcYejA has two clear domains with a binding cleft in the middle, as expected and is common with all Cluster C SBPs. It also has the extra domain which is characteristic of Cluster C SBPs. The only thing initially noticed as being slightly different from other Cluster C SBPs was the extended disordered loop in the N-terminal region, which is highlighted in Figure 4.7.

Data collection	
X-ray source	DLS beamline i04-1
Wavelength (Å)	0.92819
Resolution range (Å)	29.91-1.60
Space group	I_{222}
Unit-cell parameters (Å/°)	91.44, 105.27, 145.34/ 90.0, 90.0, 90.0
Matthews coefficient (Å^3/Da) /solvent content (%)	2.56/ 51.95%
Number of observations overall/outer shell[a]	754721/38115
Number of unique observations overall/outer shell[a]	91935/4473
Completeness (%), overall/outer shell[a]	99.6/99.1
$I / \sigma (I)$, overall/outer shell[a]	17.4/3.4
Rmerge[b], overall/outer shell[a]	0.064/0.550
CC $_{\frac{1}{2}}$[c]	0.999/0.930
Refinement and model statistics	
R-factor[d]/R-free[e]	0.160/0.191
Reflections (working/free)	91932/4545
Molecules/asymmetric unit	1
No. of atoms: Protein	4736
No. of atoms: Ligand	91
No. of atoms: Water	592
No. of atoms: Glycerol	18
R.m.s deviations[f]: Bonds (Å); Angles (°)	0.0262; 2.158
Average B-factor (Å^2): Protein	21.89
Average B-factor (Å^2): Ligand	27.6
Average B-factor (Å^2): Water	34.77
Average B-factor (Å^2): Glycerol	31.93
Ramachandran plot: preferred regions/ allowed regions/ outliers (%)	97.09/2.05/0.85

Table 4.1. X-ray data collection and refinement statistics obtained for EcYejA.

[a]The outer shell corresponds to 1.63-1.60 Å. [b]Rmerge = $\sum_{hkl}\sum_i |I_i - <I>|/\sum_{hkl}\sum_i <I>$ where I_i is the intensity of the ith measurement of a reflection with indexes hkl and $<I>$ is the statistically weighted average reflection intensity. [c]CC$_{1/2}$ is the correlation coefficient between two randomly selected half data sets as described in Karplus & Diederichs (2012). [d]R-factor = $\sum ||F_o|-|F_c||/\sum |F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. [e]R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from the refinement. [f]Root-mean-square deviation of bond lengths and bond angles from ideal geometry.
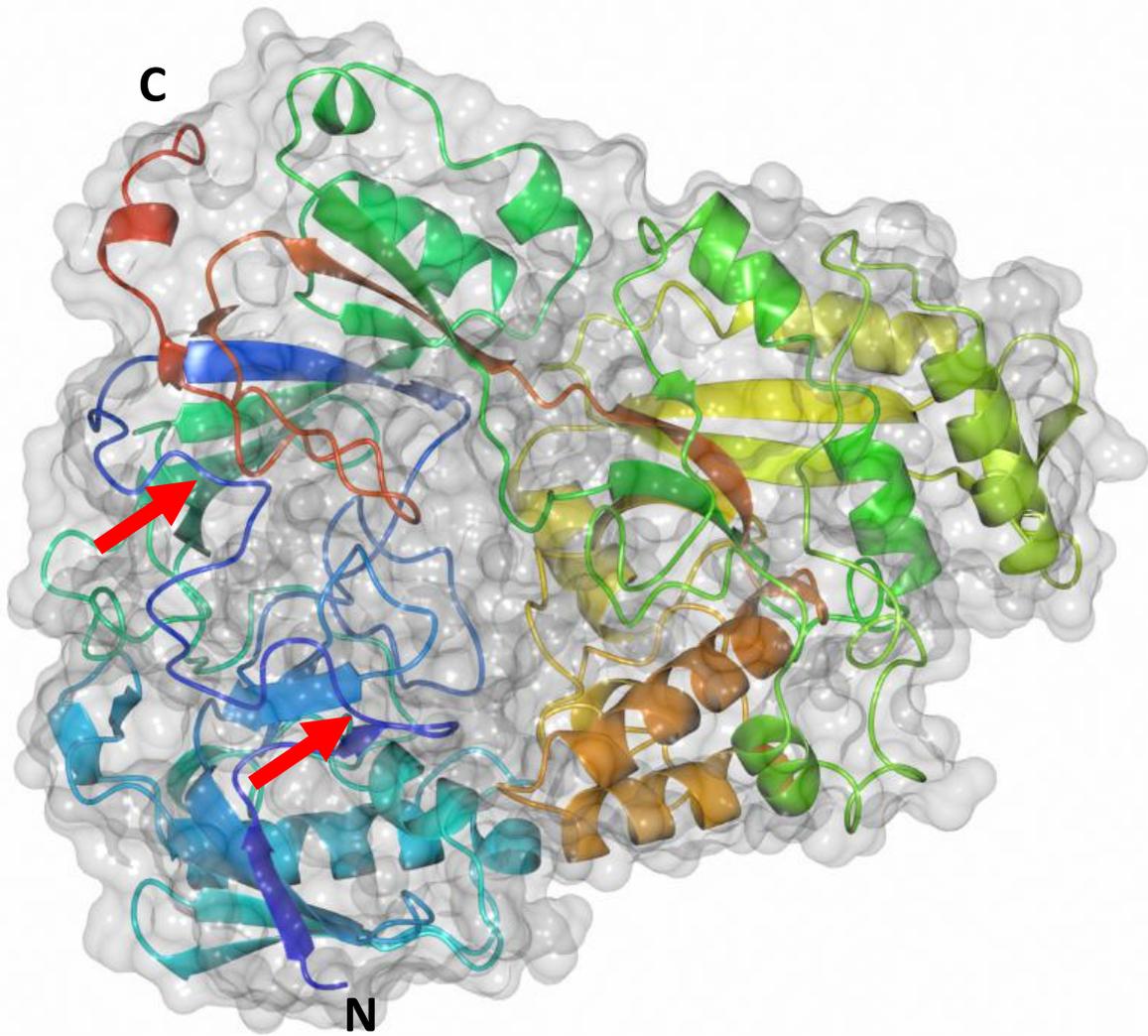
Figure 4.7. Structure of EcYejA.

EcYejA is shown in ribbon form coloured from N to C terminus in blue to red. The surface of EcYejA is shown in grey. The dark blue disordered section of EcYejA between the two red arrows is the disordered loop in the N-terminal region, as described in the text, which is different from other Cluster C SBPs.

4.4 Ligand determination

In the course of refining the structure of EcYejA, extended electron density was discovered in the protein interior which could not be accounted for by the protein sequence of EcYejA. Once the modelling of EcYejA chain had been completed, it was possible to determine specific features of the density. The density strongly resembled a peptide chain with very clear density for some of the side chains of the peptide chain (Figure 4.8) however, some of the side chains were difficult to pick out (Figure 4.9). As these Cluster C SBPs are often peptide binding proteins this peptide chain was assumed to be a ligand of EcYejA. Due to the density for the peptide ligand being so clear it was possible to build the peptide ligand backbone and add in certain side chains. A BLAST search using the inferred peptide sequence was carried out using the identified ligand sequence XXPRYXFXFX, the results came back showing the disordered loop in the N-terminal region of EcYejA as the source of the ligand and identified the ligand as LGEPRYAFNFN (GEP). When the unknown residues of GEP were fitted into the density they fitted the density well and therefore confirmed the ligand sequence as LGEPRYAFNFN (Figure 4.10). EcYejA having a ligand bound was highly unexpected as it was assumed that the 2 M guanidinium-HCl treatment of EcYejA whilst bound to the Nickel affinity column would release any pre-bound ligand and EcYejA would be in the apo form. This combined with the highly unusual source of the ligand led to much consideration of how the ligand could have been produced and available for EcYejA binding. It is possible that during the month it took for the crystals to grow that a proportion of the EcYejA was proteolytically cleaved in the mother liquor and was bound by full length EcYejA. It is also important to note at this point that 100% occupancy was seen for the full length EcYejA and additional 100% occupancy density for the ligand (Figure 4.11), meaning a single molecule of EcYejA was not binding its own N-terminal region.

To ensure that EcYejA was in a truly closed conformation, as opposed to a partially closed conformation, the software rdock and rbcavity were used to create a binding pocket mesh. If this mesh was visible when a surface view was used on the EcYejA structure then EcYejA was in a partially closed conformation, but if the mesh was not visible then it shows EcYejA is in the truly closed conformation (Figure 4.12). As Figure 4.12 shows only a small amount of mesh can be seen in a surface view and therefore EcYejA is in the truly closed conformation.
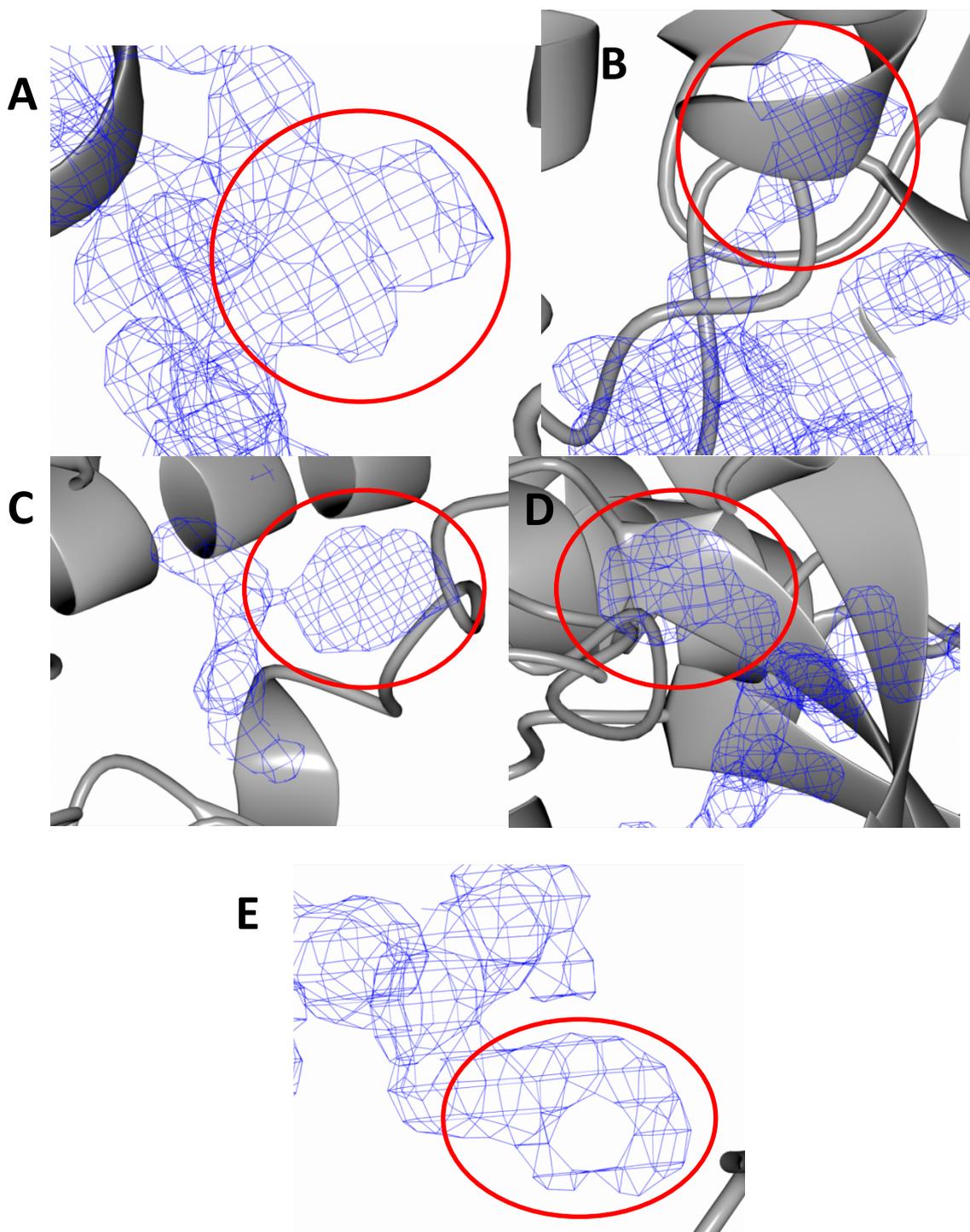
Figure 4.8. Electron density shows the presence of side chains in the crystal ligand.

EcYejA is shown in grey ribbon form and electron density is shown in blue mesh. The red circle highlights the side chains. (A) Shows a Proline at position 4 in the ligand, contour level 1.0, (B) shows an Arginine residue at position 5 in the ligand, contour level 1.0, (C) shows a Tyrosine residue at position 6 in the ligand, contour level 1.5, (D) shows a phenylalanine residue at position 8 in the ligand, contour level 1.5 and (E) shows a phenylalanine residue at position 10 in the ligand, contour level 1.0.
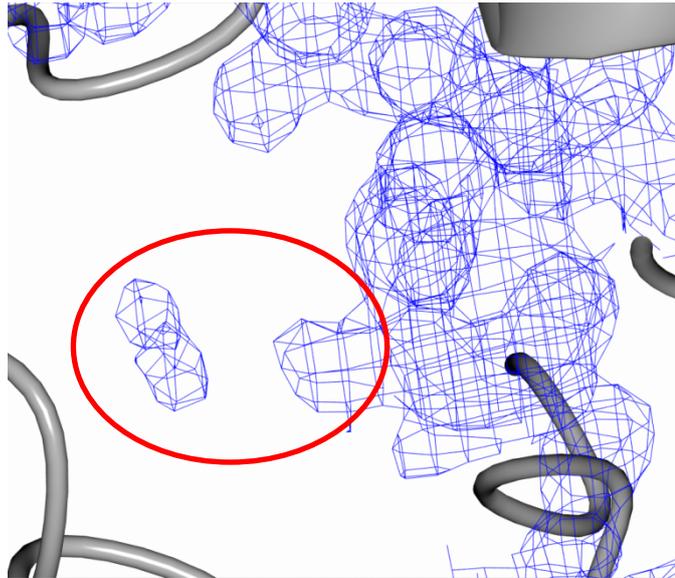
Figure 4.9. Example of electron density where it was not possible to determine the side chain.

EcYejA is shown in grey ribbon form and electron density is shown in blue mesh at a contour level of 1.0. The red circle highlights the side chain that could not be determined by the electron density.
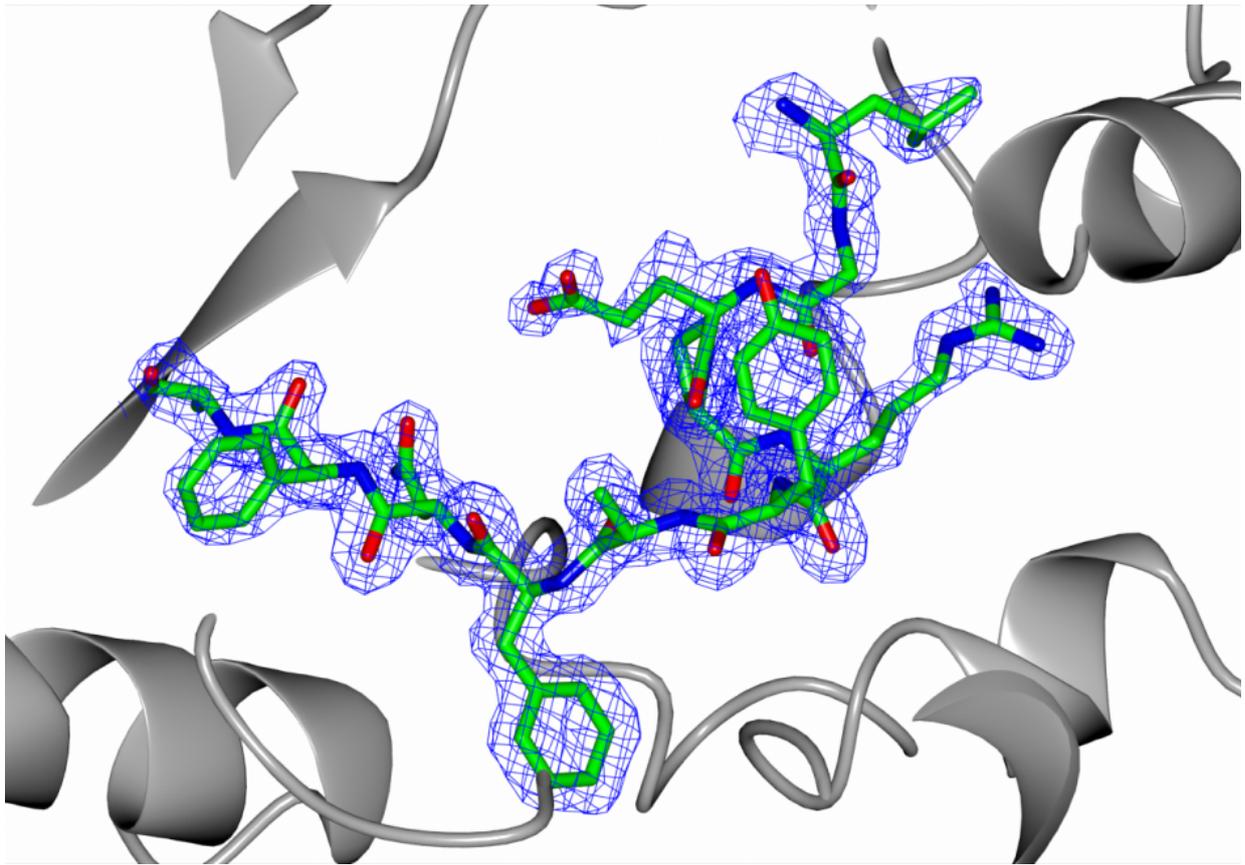
Figure 4.10. GEP ligand fully built with final density after refinement.

EcYejA is shown in grey ribbon form, GEP ligand is shown in cylinder form with carbon atoms shown in green, nitrogen atoms shown in blue and oxygen atoms shown in red. Electron density is shown in blue mesh at a contour level of 1.0.
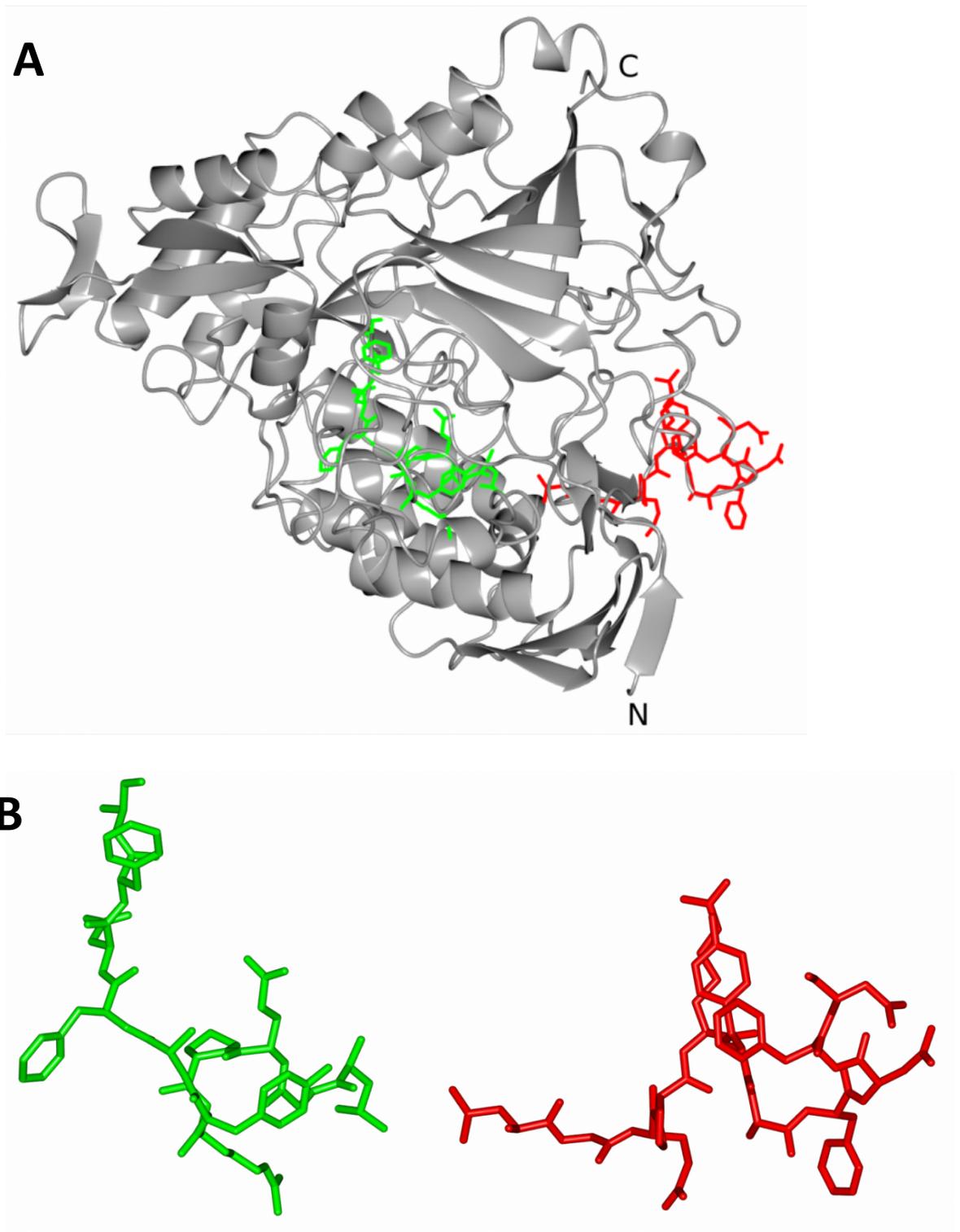
Figure 4.11. GEP ligand and corresponding section of EcYejA highlighted in structure.

(A) EcYejA is shown in grey ribbon form with N and C termini labelled. GEP ligand in the binding pocket is shown in green bond format and corresponding GEP peptide section of EcYejA is shown in red bond format. (B) Same orientation as (A) but EcYejA has been removed.
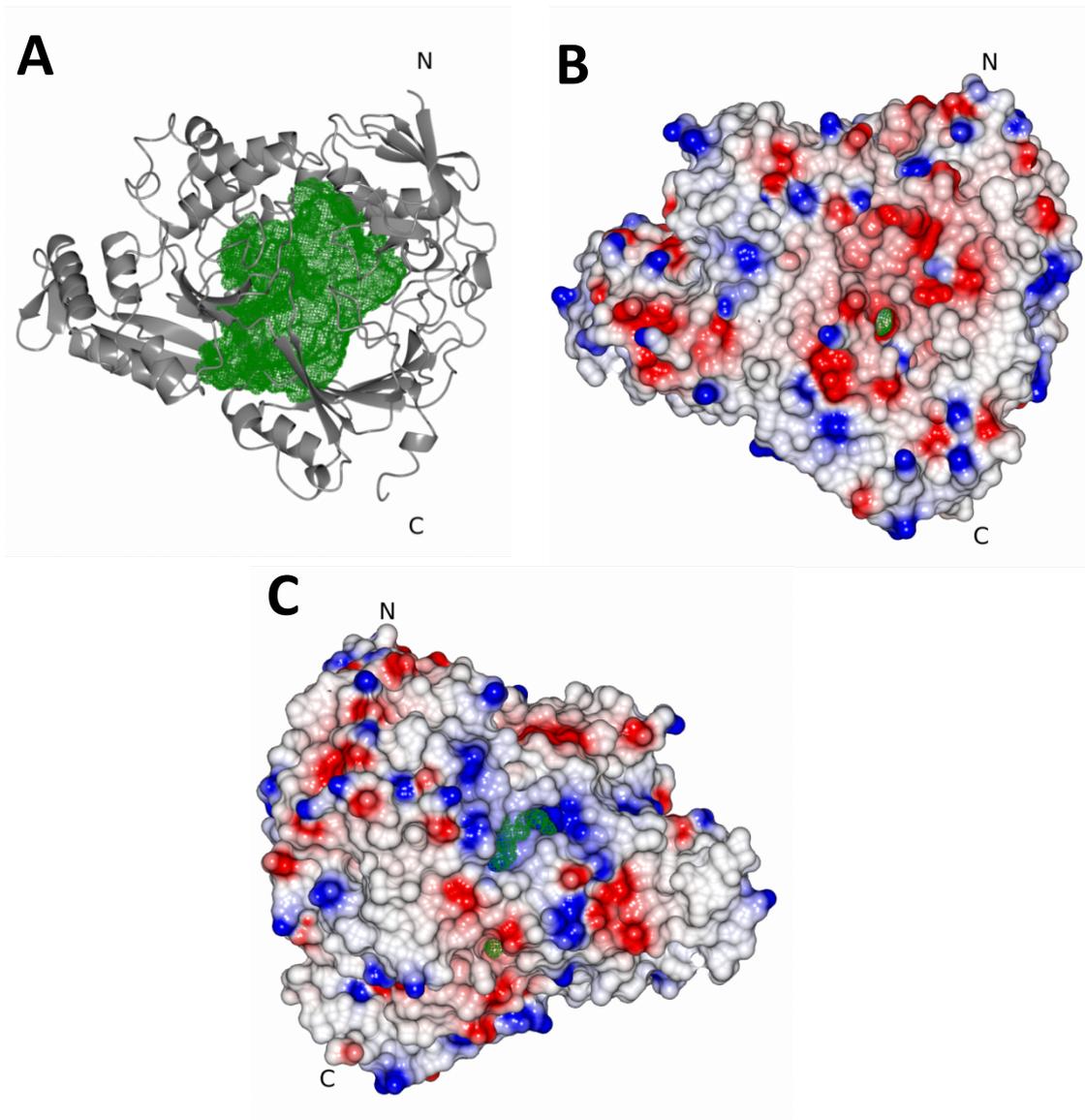
Figure 4.12. EcYejA with binding pocket mesh shown in different views.

EcYejA is shown in grey ribbon form or electrostatic surface view, N and C refer to EcYejA N and C termini. Calculated EcYejA binding pocket volume is shown in green mesh.

4.4.1 Analysis of the interactions between LGEPRYAFNFN and EcYejA

Leucine-1 makes two backbone hydrogen bonds to surrounding water molecules, one from the alpha amino group nitrogen and one from the backbone oxygen (Figure 4.13). Interestingly there appears to be more space on the N-terminus of GEP to build further into binding pocket, indicating that EcYejA may be able to bind longer peptides. As calculated by PDBePISA, 80% of the accessible surface area of Leu1 from GEP is buried upon binding. PDBePISA is an online tool from EMBL-EBI which determines the interface interactions between proteins and the properties of specific residues, this tool was used to further understand GEP binding. The solvation energy effects, in kcal/mol, of each GEP residue was also calculated using PDBePISA. A positive solvation energy of a residue makes a negative contribution to the solvation energy of the interface and therefore corresponds to hydrophobic effects. All solvation energy effects calculations made in PDBePISA do not include the effects of Hydrogen bonds (-0.44 kcal/mol per bond) or salt bridges (-0.15 kcal/mol per salt bridge). PDBePISA calculates the solvation energy effect of Leucine from GEP as +1.93 kcal/mol.

Glycine-2 again contributes two hydrogen bonds from the peptide backbone, one from the nitrogen to the carboxylate side chain of Asp477 (Figure 4.14A). The other one is interestingly between the backbone oxygen and the backbone nitrogen of Arg5 from GEP, this interaction is helping to hold GEP in a twisted almost α-helical form. 100% of the accessible surface area of Gly2 is buried upon EcYejA binding of GEP and the solvation energy effect is calculated as +0.1 kcal/mol.

Glutamate-3 which makes a single hydrogen bond with water molecule 292 via an oxygen on the side chain (Figure 4.14B). PDBePISA predicts that a salt bridge is also formed between a side chain oxygen and Arg51 in EcYejA. Unlike Leu1 and Gly2 beforehand there are no backbone interactions with EcYejA. PDBePISA calculated that 40% of the accessible surface area is buried upon EcYejA binding of GEP and the solvation energy effect is -0.45 kcal/mol.

Proline-4 forms a single hydrogen bond between the backbone oxygen and Gln132 from EcYejA (Figure 4.14C). 80% of Pro4 accessible surface area is buried upon EcYejA binding of GEP and the solvation energy effect is +0.86 kcal/mol.
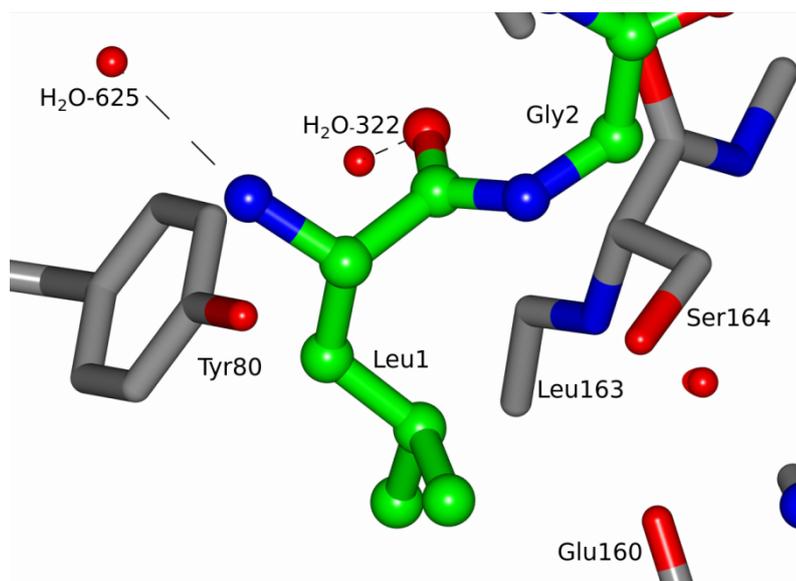
Figure 4.13. Neighbourhood and interactions of the first residue of GEP, Leucine, with EcYejA.

GEP is depicted in green and EcYejA in grey, water molecules are shown as red spheres and bonds are shown as black dashed lines. Hydrogen bonds can be seen between GEP Leu1 and water molecules 322 and 625.
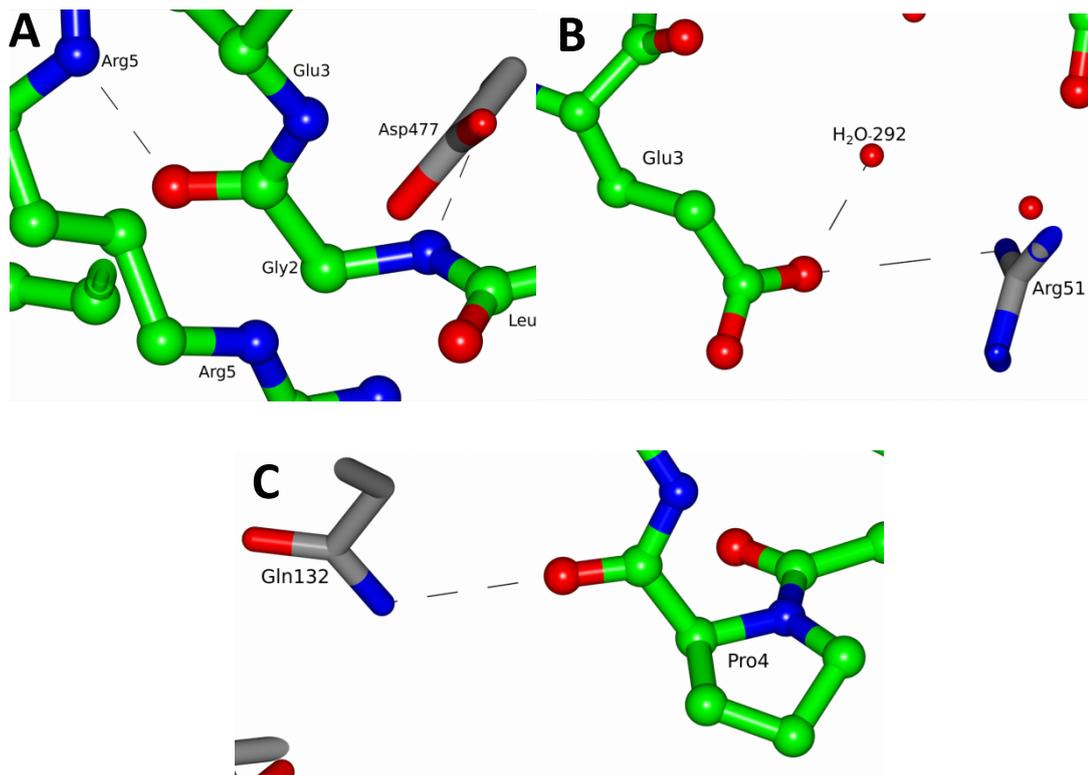
Figure 4.14. Neighbourhood and interactions of GEP residues with EcYejA.

GEP is depicted in green and EcYejA in grey, water molecules are shown as red spheres and bonds are shown as black dashed lines. (A) is a view of Gly2 from GEP. Hydrogen bonds can be seen between Gly2 and Asp477 from EcYejA and Arg5 from GEP. (B) is a view of Glu3 from GEP. A single Hydrogen bond is formed between a side chain oxygen on Glu3 and water molecule 292. A salt bridge is also formed between Glu3 and Arg51. (C) is a view of Pro4 of GEP, a hydrogen bond is formed between the oxygen from the peptide backbone and Gln132 in EcYejA.

The arginine-5 side chain in GEP makes two hydrogen bonds with surrounding residues, one with Ser476 and one with Tyr137 (Figure 4.15A). PDBePISA predicts Arg5 to make 4 salt bridges with EcYejA, two to Asp477 and another two to Asp161. As well as the side chain interactions there is also a hydrogen bond formed between the backbone nitrogen of Arg5 and the backbone oxygen of Gly2, as mentioned previously. Upon EcYejA binding of GEP 100% of Arg5 is buried and the solvation energy effect is -0.81 kcal/mol.

Tyrosine-6 makes a single hydrogen bond from the backbone oxygen to a water molecule (60) (Figure 4.15B). There are possible stacking interactions between the Tyr6 side chain and the rest of the GEP peptide. 90% of accessible surface area of Tyr6 is buried upon EcYejA binding of GEP and the solvation energy effect is +1.13 kcal/mol.

The seventh residue in GEP, Alanine-7, again makes a single hydrogen bond from the backbone oxygen to Tyr491 in EcYejA (Figure 4.15C). 100% of Ala7 is buried upon EcYejA binding of GEP and the solvation energy effect is +0.66 kcal/mol.

Phenylalanine-8 from GEP makes three hydrogen bonds, two of which are from the backbone oxygen to Arg51 and the other is from the backbone nitrogen to water molecule 34 (Figure 4.16A). 100% of Phe8 is buried upon EcYejA binding of GEP and the solvation energy effect is +2.44 kcal/mol.

Asparagine-9 makes a total of 6 hydrogen bonds to EcYejA, three from the side chain and three from the backbone (Figure 4.16B). Two are made from the side chain oxygen to water molecules 288 and 385, whilst one is made from the sidechain nitrogen to Met456. Interestingly this hydrogen bond is only formed when Met456 is in one of two conformations. The backbone oxygen makes two Hydrogen bonds to Arg457 and the backbone nitrogen makes one Hydrogen bond to Thr490. Upon EcYejA binding of the GEP ligand 80% of Asn9 is buried and the solvation energy effect is -0.27 kcal/mol.

There are two hydrogen bonds from the backbone of Phe10 to water molecule 288, one from the nitrogen and one from the oxygen (Figure 4.16C). There also seems to be stacking interactions between the side chain of Phe10 and Pro58 from EcYejA. 70% of Phe10 is buried upon EcYejA binding of GEP and the solvation energy effect is +2.05 kcal/mol.

Although the sidechain of asparagine-11 cannot be seen in the density it is possible to identify a hydrogen bond between the backbone nitrogen and water molecule 541 (Figure 4.16D). Only 20% of Asn11 is buried upon EcYejA binding of GEP and the solvation energy
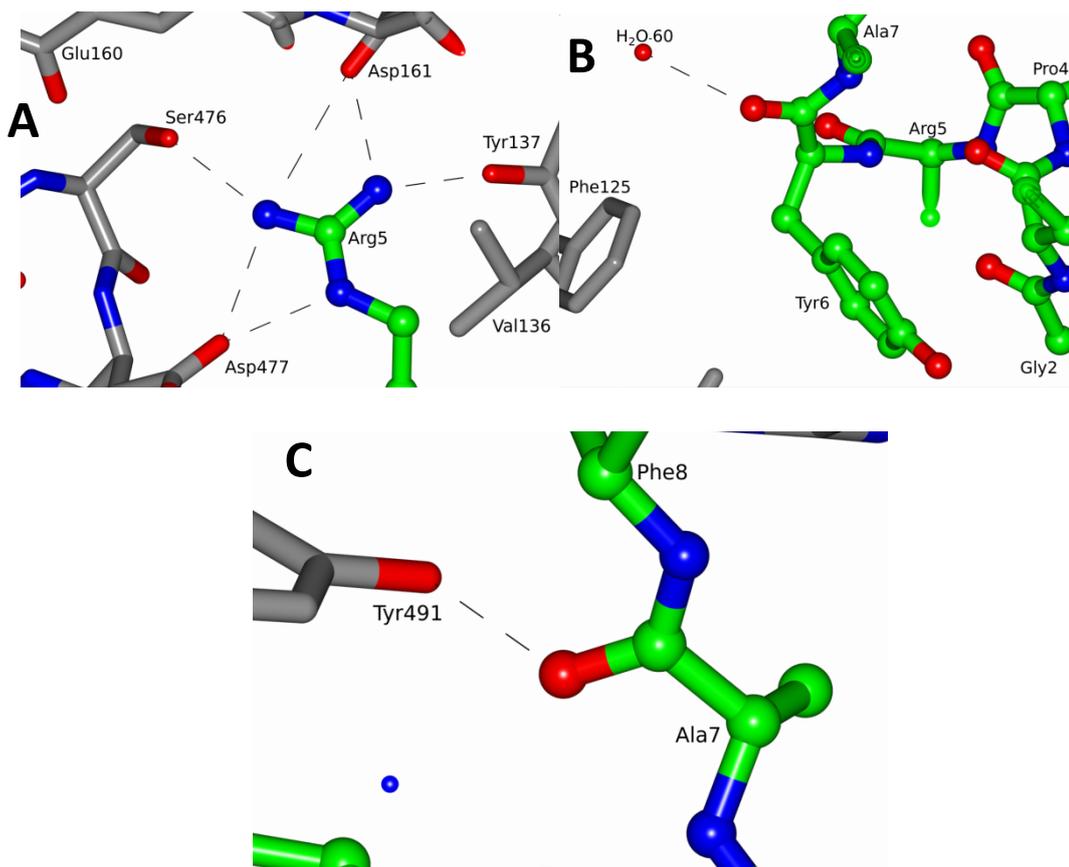
Figure 4.15. Neighbourhood and interactions of GEP residues with EcYejA.

GEP is depicted in green and EcYejA in grey, water molecules are shown as red spheres and bonds are shown as black dashed lines. (A) shows Arg5 of GEP. Hydrogen bonds are made between Arg5 and Ser476 and Arg5 and Tyr137. Four salt bridges are also made, two with Asp477 and two with Asp161. (B) is a view of Tyr6 of GEP. There is single hydrogen bond between the oxygen from the back bone of Tyr6 and water molecule 60. (C) shows Ala7 of GEP. A single hydrogen bond is made between the backbone oxygen and Tyr491.
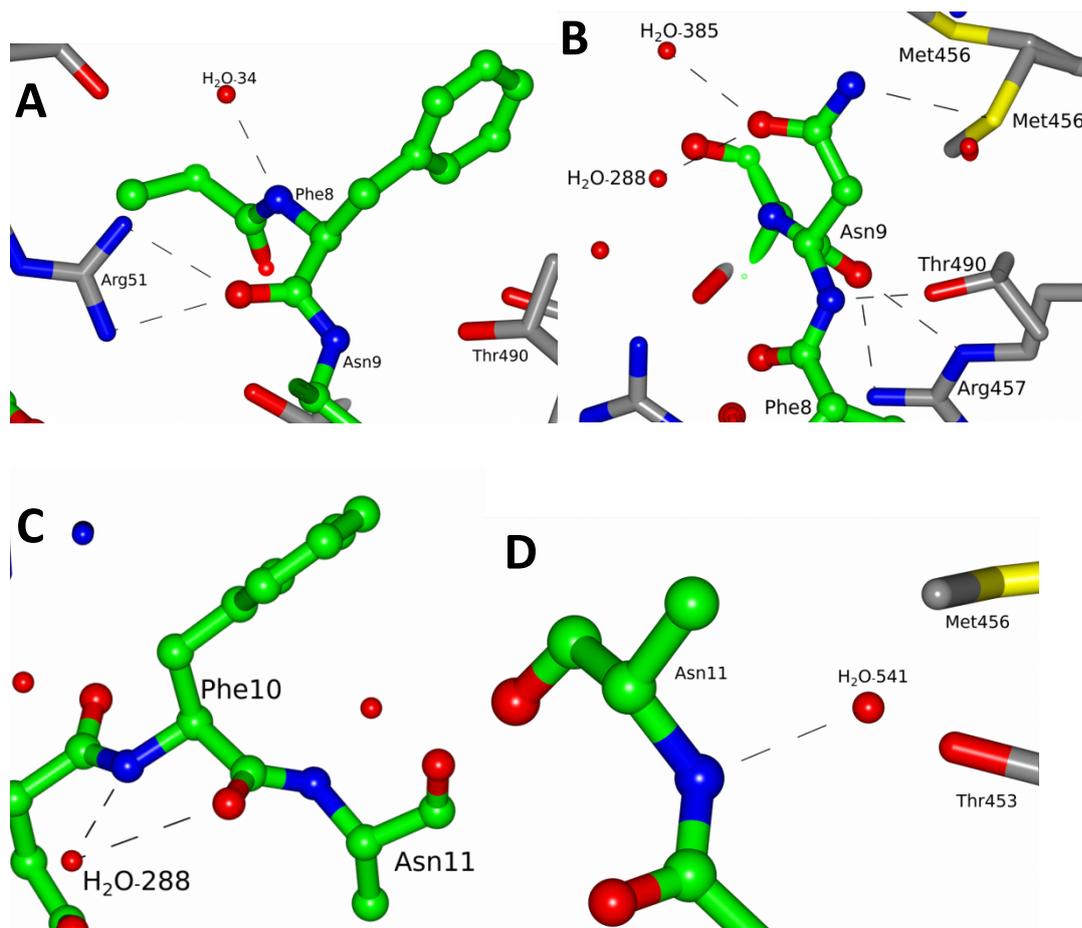
Figure 4.16. Neighbourhood and interactions of GEP residues with EcYejA.

GEP is depicted in green and EcYejA in grey, water molecules are shown as red spheres and bonds are shown as black dashed lines. (A) shows Phe8 of GEP. Two hydrogen bonds are made between the backbone oxygen and Arg51 and one is made between the backbone nitrogen and water molecule 34. (B) is a view of Asn9 of GEP. The side chain of Asparagine makes three hydrogen bonds, one to water molecule 288, one to water molecule 385 and one to Met456. The main chain backbone makes two from the backbone oxygen to Arg457 and one from the backbone nitrogen to Thr490. (C) shows Phe10 of GEP. Two hydrogen bonds are formed between the backbone and water molecule 288, one from the backbone oxygen and one from the backbone nitrogen. (D) is a view of Asn11 of GEP. A single hydrogen bond is shown between the backbone nitrogen of Asn11 and water molecule 541.

effect is +0.26 kcal/mol. This may be due to the side chain of Asn11 not being built into the structure.

Between GEP and either EcYejA or water molecules there are a total of 24 hydrogen bonds, 15 of which are from the backbone of the GEP peptide and 9 are from the side chains of GEP. 13 of hydrogen bonds formed are between GEP and EcYejA and 11 are between GEP and surrounding water molecules. Asn9 from the GEP peptide makes the most hydrogen bonds, 6 in total, of any of the GEP residues, however Arg5 makes the most side chain hydrogen bonds.

There are a total of 5 salt bridges between EcYejA and GEP, 4 of which are formed from the Arginine residue to various EcYejA residues. The other is between Glu3 from GEP and Arg51 of EcYejA.

When all the hydrogen bonds and salt bridges formed between GEP and EcYejA, i.e. not including those between GEP and water molecules, are taken into account the solvation energy effects of each of the GEP residues are as shown in Table 4.2. From this analysis it is possible to see that Arg5 is the residue with the best solvation energy effects, closely followed by Asn9, both of these residues also made the most hydrogen bonds with EcYejA, indicating that these residues are important for EcYejA binding of GEP. It can also be seen that Phe10 and Leu1 are the residues with the worst solvation energy effects and therefore are the residues that are contributing the least to EcYejA binding of GEP. These results indicate that EcYejA is binding the central portion of GEP and the ends of the peptide are not as important for binding as the middle region of the peptide. This means it could be possible for EcYejA to bind peptides that are longer than 11 residues so long as the central portion of the peptides remain the same as or similar to the central portion of GEP.

Due to the Guanidinium-HCl treatment step during the purification process, the only source of ligands available in the crystallisation tray was the crystal screen components or EcYejA. Therefore, it isn't surprising that the ligand is from EcYejA as these proteins are peptide transporters and the only source of peptides was EcYejA itself.

GEP is located 12 residues from the N-terminus of EcYejA (Figure 4.17), which begs the questions, where is the rest of the N-terminus of EcYejA? Why doesn't EcYejA bind that 12 residue peptide as well as the GEP? To what extent is EcYejA broken down? Is EcYejA selective for GEP?

133

Residue	Buried Surface Areas (%)	Solvation Energy Effects (kcal/mol)
Leu1	80	+1.93
Gly2	100	-0.34
Glu3	40	-1.04
Pro4	80	+0.42
Arg5	100	-3.61
Tyr6	90	+0.69
Ala7	100	-0.22
Phe8	100	+1.56
Asn9	80	-2.47
Phe10	70	+2.05
Asn11	20	+0.26

Table 4.2. Buried surface area and solvation energy effects of residues of GEP.

The numbers were calculated using PDBePISA, a web based tool which analyses the interface and interactions between two molecules, in this case EcYejA and GEP.

```
          10              20              30
MQAIKESYAF AVLGEPRYAF NFNHFDYVNP
```

Figure 4.17. GEP ligand is located 12 residues from the N-terminus of EcYejA.

GEP ligand is underlined and shown in red.

Once the ligand had been identified as a portion of the N-terminus of EcYejA, it was decided that truncating EcYejA at various points to remove the N-terminal region containing the ligand would be a good course of action. It was decided that truncating the N-terminus of EcYejA would help to prevent the GEP ligand from binding to EcYejA in future ligand binding experiments and therefore prevent skewing of future data. Three different truncations were made in EcYejA to remove GEP from the protein but limit the removal of any secondary structure features of EcYejA (Figure 4.18). The constructs were cloned into pETFPP_30 and created using the primers STEcYejA-F and EcYejA-R; MTEcYejA-F and EcYejA-R; and LTEcYejA-F and EcYejA-R for the truncations at positions 1, 2 and 3 in Figure 4.18 respectively. It was thought that this would prevent truncations in EcYejA from drastically altering the structure or function of EcYejA. However, when all three truncated EcYejA proteins were expressed following the same protocols as the expression of full length EcYejA, none were expressed (Figure 4.19). In the interests of time it was decided that pursuing the truncated versions of EcYejA would not be beneficial, instead fresh preparations of full length EcYejA were made regularly and stored at 4 °C for up to a month whilst being used in experiments. It was thought that the crystallisation solution was causing the breakdown of EcYejA, therefore storing it at 4 °C for short periods of time wouldn't cause the breakdown of EcYejA and release of the ligand seen in the crystal structure.

4.5 Comparisons between EcYejA structure and other Cluster C SBP structures

4.5.1 Overall structure of EcYejA as compared to other Cluster C SBPs

A number of other Cluster C SBPs, whose structures have been solved or are relevant to this work, were chosen to make comparisons with the structure of EcYejA, they included; *E. coli* MppA (EcMppA) which binds murein tripeptide (PDB:3O9P); *S.* Typhimurium OppA (StOppA) which binds peptides 2-5 residues in length, but in this structure is complexed with KLK (PDB:1B9J); *B. subtilis* AppA (BsAppA) that binds a nonapeptide (PDB:1XOC); *L. lactis* OppA (LlOppA) which binds a 35 residue peptide, but in this structure is complexed with a nonapeptide (PDB:3DRG) and *E. coli* SapA (EcSapA) where the ligand and structure are unknown. Further protein information is detailed in Table 2.1 Chapter 2.

A noticeable thing with the EcYejA structure when overlaid with other Cluster C SBPs is that EcYejA is a longer protein, there are extra elements of secondary structure and disordered loops scattered throughout EcYejA as compared to other Cluster C SBPs (Figure 4.20).

Figure 4.18. Truncations of EcYejA.

EcYejA is shown in grey ribbon form. The N and C refer to the N and C termini of EcYejA. The section of EcYejA which corresponds to the GEP ligand is shown in turquoise cylinder form. The residues at which truncations were made are shown in red cylinder form and numbered and circled in red.
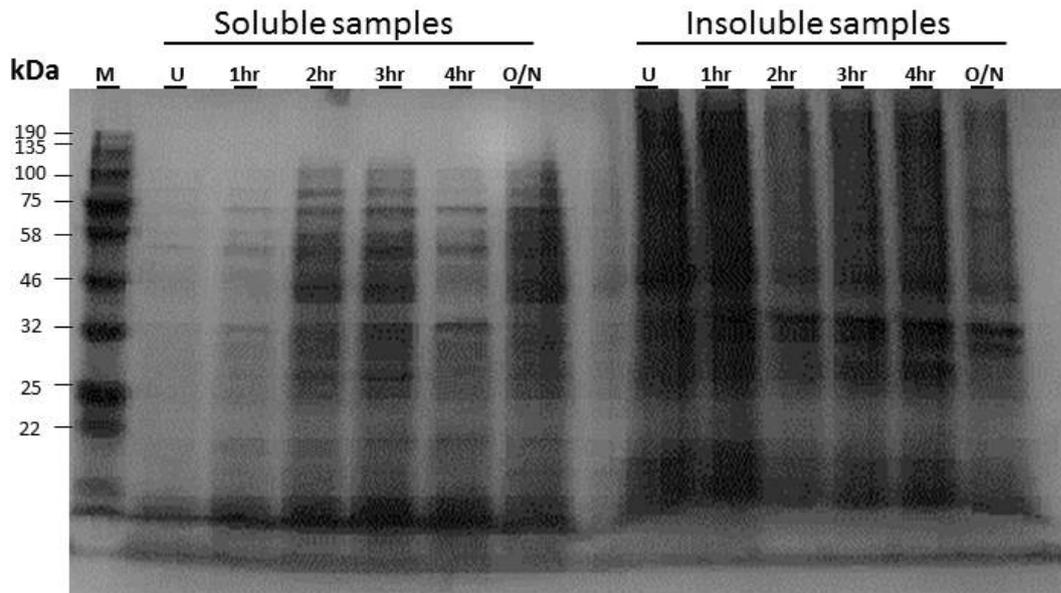
Figure 4.19. Truncations of the N-terminus of EcYejA are not expressed.

12% polyacrylamide SDS-PAGE gel stained with Coomassie blue dye. Cells were grown at 30 °C and recombinant protein production was induced with 1 mM IPTG, all samples were boiled at 95 °C in SDS for 5 mins before loading on the gel. M = Marker; U = Uninduced sample; 1, 2, 3, 4hr = Number of hours after induction that a sample was taken; O/N = Overnight sample. This gel is representative of all three EcYejA truncations made.
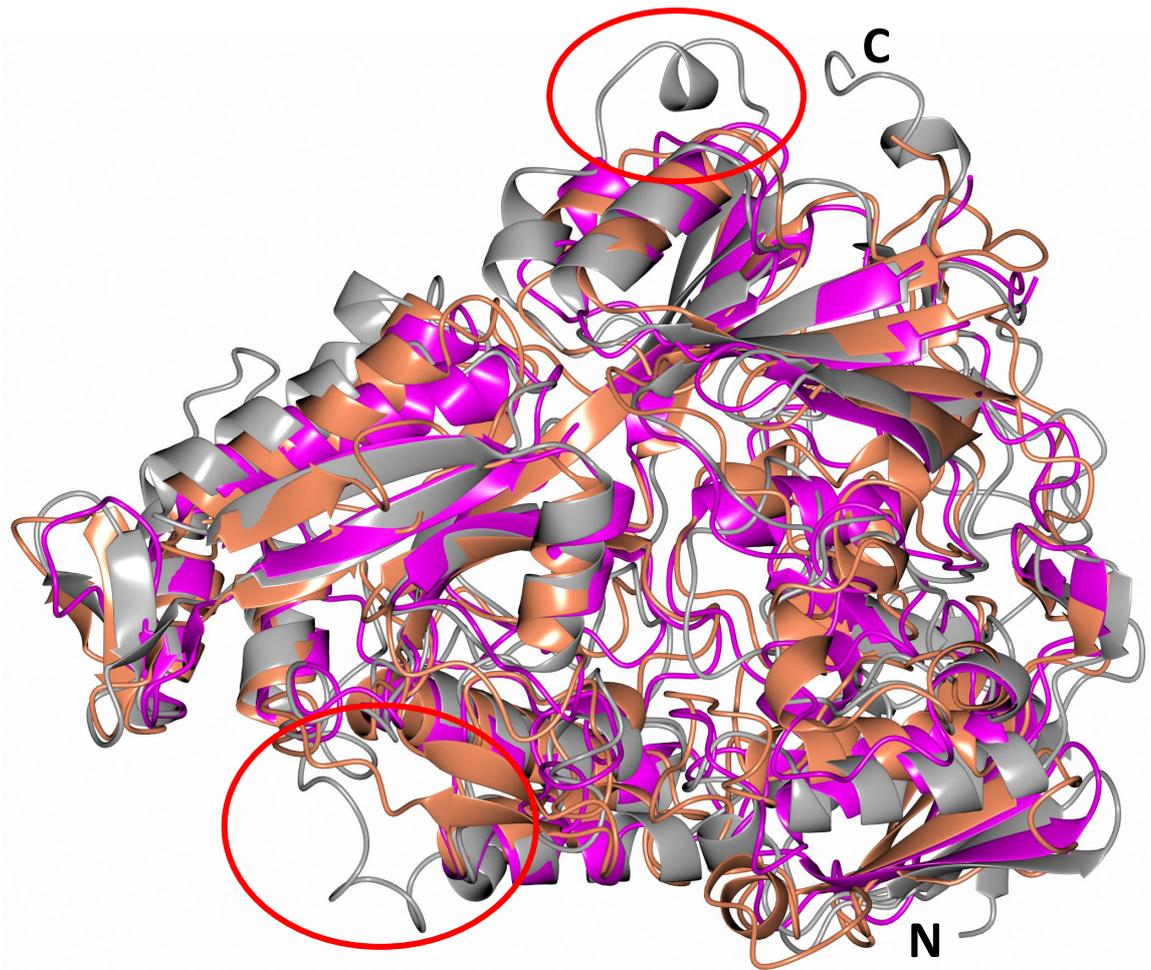
Figure 4.20. EcYejA superposed with BsAppA and StOppA.

EcYejA is shown in grey, BsAppA (PDB:1XOC) is shown in pink and LlOppA (PDB:3DRG) is shown in coral. The N and C refer to the N and C termini of EcYejA. The red circle indicates an area of EcYejA that is not present in either BsAppA or LlOppA.

EcYejA without its signal sequence is 585 residues, LlOppA is 570, BsAppA is 520, StOppA is 517 and EcMppA is 515. We hypothesise that EcYejA and LlOppA are longer proteins to allow them to create a larger binding pocket and therefore the ability to bind longer and bulkier substrates.

Interestingly GEP is not present in other Cluster C substrate binding proteins (Figure 4.21). Sequence and structural alignments of this family of substrate binding proteins show omissions in the protein sequences where GEP is present in EcYejA. This suggests the release of this specific GEP might have a role in EcYejA function such as autoregulation.

PDBePISA was used to better analyse and understand the structure of EcYejA as compared to other Cluster C SBPs. Residues in Cluster C SBP structures that were interacting with the bound ligands were identified using PDBePISA. 52 residues from BsAppA interact with the ligand, 50 from LlOppA, 42 from EcYejA, 31 from EcMppA and 30 from StOppA. Naturally the larger a ligand is the more residues there are that will interact with that ligand, however this is not the case with BsAppA, LlOppA and EcYejA. EcYejA has been shown to bind an 11 residue peptide whereas BsAppA and LlOppA, in these structures, are only complexed with 9 residue peptides. This gives an indication that there could be some differences in the binding pockets of EcYejA as compared to these proteins.

4.5.2 Binding pocket comparisons of EcYejA with other Cluster C SBPs

As mentioned in 3.1.2, some Cluster C SBPs have an aspartate residue in the binding pocket which "caps" the end of the binding pocket and serve as an anchor for the amino terminus of peptide ligands by forming a salt bridge with the alpha amino group of the peptide ligand. This prevents a longer peptide binding therefore "capping" the binding site. In sequence alignments (Figure 4.22) this aspartate can be seen in EcMppA, StOppA and is predicted to be present in the binding site of EcSapA. However, in Cluster C SBPs which have been shown to bind longer peptides, such as BsAppA, LlOppA and EcYejA, this aspartate residue is not conserved and is not present in the binding site. Therefore, the presence or absence of this capping aspartate can be used as a method to predict whether Cluster C SBPs bind longer or shorter peptides. Using this method it is possible to predict that even if GEP is not the natural ligand for EcYejA and is possibly only a crystal artefact, EcYejA has the ability to bind longer peptides. Again, the indication is that EcSapA is only able to bind shorter peptides due to the predicted presence of the capping aspartate.

```
                              β2
EcYejA/27-48        →              TT            TT    T
                           3 0         4 0
EcYejA/27-48        A F A V L G E P R Y A F N F N H F D Y V N P
LlOppA/32-41        S T K K L K A . . . . . . . . . . . . G N F
BsAppA/34-40        A K K S A G K . . . . . . . . . . . . . . .
EcSapA/25-29        . . S P P H A . . . . . . . . . . . . . . .
StOppA/27-33        . A D V P A G V . . . . . . . . . . . . . .
EcMppA/23-29        . A E V P S G T . . . . . . . . . . . . . .
```
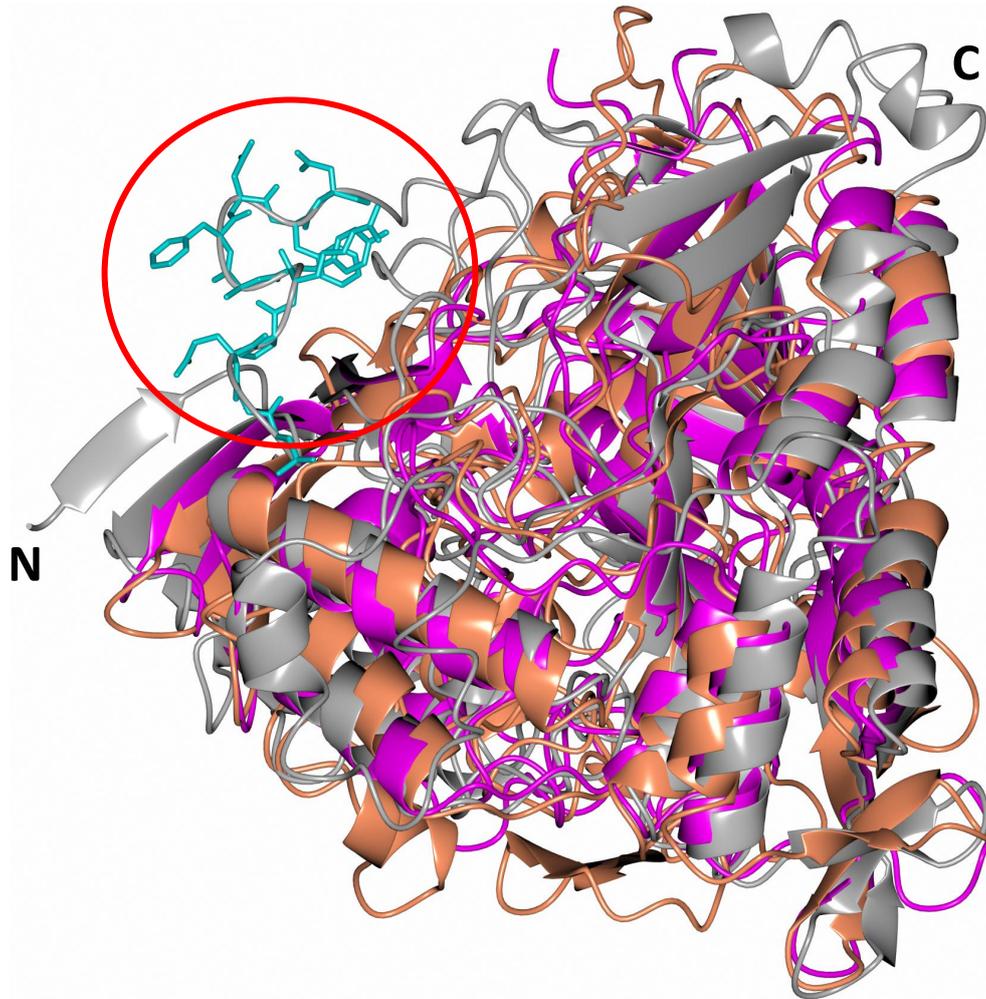

Figure 4.21. GEP ligand is not present in other Cluster C SBPs.

Top, the GEP peptide is underlined in red, EcYejA secondary structure is shown along the top of the alignment and the residue numbers shown in the alignment are detailed next to the sequence name. The majority of the GEP peptide is only present in the EcYejA protein sequence. Bottom, EcYejA is shown in grey, BsAppA (PDB:1XOC) is shown in pink and LlOppA (PDB:3DRG) is shown in coral. The N and C refer to the N and C termini of EcYejA. The section of EcYejA which corresponds to the GEP ligand is shown in turquoise cylinder form. The red circle indicates an area of EcYejA that is not present in either BsAppA or LlOppA which contains the GEP ligand.

```
EcYejA/1-604
                                          β1       β2
                                        ──────►  ──────►      TT         TT   T
                  1        10        20        30        40
EcYejA/1-604  .............MIVRILLLFIALFTFGVQAQAIKESYAFAVLGEPRYAFNFNHFDYVNP
LlOppA/1-600  ....MNKLKVTLL.ASSVV.LAATLLSACG.SNQSSSTSTKKLKA.............GNF
BsAppA/1-543  ....MKRRKTALMMLSVLM.VLAIFLSACSGSKSSNSSAKKSAGK...............
EcSapA/1-547  ..MRQV......LSSLL.VIAGLVSGQAIAAPE.......SPPHA...............
StOppA/1-543  MSNITKKS......LIAAG.ILTALIAASA.....ATAADVPAGV...............
EcMppA/1-537  ....MKHS......VSVTC.CALLVSSISL.....SYAAEVPSGT...............


                       β3                               η1       β4       β5       β6
EcYejA/1-604  T      ──────►              TT           ℓℓℓ      ───►    ──────►  ──────►
                 50        60        70        80        90        100
EcYejA/1-604  AAPKGGQITLSALGTFDNFNRYA.LR.GNPGARTEQLYDTLFTTSDDEPGSYYPLIAESA
LlOppA/1-600  DVA.YQNPDKAIKGGNLKVAYQSDSPMKAQWLSGLSND.ATFATMSGPGGGQDGLFFTDS
BsAppA/1-543  ..P..QQGGDLVVGSIGEPTLFNSLYSTDDASTDIENMLYSFLTKTDEKLNVKL..SLAE
EcSapA/1-547  DIRDSGFVYCVS.GQV.NTFNPSKASSG.......LIVDTLAAQFYDRLLDVDP......
StOppA/1-543  QLADKQTLVRNN.GSEVQSLDPHKIE.G.......VPESNVSRDLFEGLLISD......V
EcMppA/1-537  VLAEKQELVRHIKDEP.ASLDPAKAV.G.......LPEIQVIRDLFEGLVNQN......E


              ───►TT     β7       TT      T............T              α1              ℓℓ
                        ──────►                                 ℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
                  110       120                    130       140       150
EcYejA/1-604  RYADDYSWVEVAINPRARFH.............DGSPITARDVEFTFQKFMTEGVPQ....
LlOppA/1-600  GFKFIKGGAADVALDKESKTATITLRKDLKWSDGSEVTAKDYEFTYETIANPAYG..SDR
BsAppA/1-543  SIKELDGGLAYDVKIKKGVKFH..........DGKELTADDVVFTYSVPLSKDYK..GER
EcSapA/1-547  .YTYRLMPELAESWEVLDNGATYRFHLRRDVPFQKTDWFTPTRKMNADDVVFTFQRIFDR
StOppA/1-543  EGHPSPGVAEKWENKDF.KVWTFHLRENAKWSDGTPVTAHDFVYSWQRLADPNTA.....
EcMppA/1-537  KGEIVPGVATQWKSNDN.RIWTFTLRDNAKWADGTPVTAQDFVYSWQRLVDPKTL.....


                   α2                            β8        β9          α3
              ...ℓℓℓ..ℓ                    ────────·······►TT ──────►  ℓℓℓℓℓℓℓ
                                      160       170       180
EcYejA/1-604  ...FRL..VYKG.............TTVKA........IAPLTVRIELAKPGKEDMLSL.
LlOppA/1-600  WTDS.....LANIVGLSDYHTGKAKT.ISGITFPDGENGKVIKVQFKEMKPGMTQSGNGY
BsAppA/1-543  GSTYEM...LKSVE...................KKGDYEVLFKLKYKDG....NFYNNAL
EcSapA/1-547  NNPWHNVNGSNFPYFDSLQFA.............DNVKSVRKLD...NHTVE..FRLAQPD
StOppA/1-543  .SPYASYLQYGHIANIDDIIAGKKPATDLGVKALD...DHTFE..VTLSEPVPYFYKLL.
EcMppA/1-537  .SPFAWFAALAGINNAQAIIDGKATPDQLGVTAVD...AHTLK..IQLDKPLPWFVNLT.


                    α4                    TT                      β10       β11
              ℓℓℓℓ                                            ──────►  ──────►    TT
                  190       200       210       220       230
EcYejA/1-604  FSLPVFPEKYWKDH........KLSDPLATPPLASGPYRVTSWKMGQNIVYSRVKDYWAA
LlOppA/1-600  FLETVAPYQYLKDVAPKDLA...SSPKTTTKPLVTGPFKPENVVAGESIKYVPNPYYWGE
BsAppA/1-543  DSTAILPKHILGNVPIADLE...ENEFNRKKPIGSGPFKFKEWKQGQYIKLEANDDYFEG
EcSapA/1-547  ASFLWHLATHYASVM....SAEYARKLEKEDRQEQLDRQPVGTGPYQLSEYRAGQFIRLQ
StOppA/1-543  VHPSVSP....V..PKSAVEKFGDKWTQPANIVTNGAYKLKNWVVNERIVLERNPQYWDN
EcMppA/1-537  ANFAFFP....V..QKANV.ESGKEWTKPGNLIGNGAYVLKERVVNEKLVVVPNTHYWDN


                 η2         β12         α5        β13       α6         α7
              ℓℓℓ        ──────►   ℓℓℓℓℓℓℓℓℓ   ──────►  ℓℓℓℓℓℓ     ℓℓℓℓ       ─
                  240       250       260       270       280       290
EcYejA/1-604  NLPVNRGRWNFDTIRYDYYLDDNVAFEAFKAGAFDLRMENDAKNWATRYTG.KNFDKKYI
LlOppA/1-600  KPK.......LNSITYEVVSTAKSVAAL.SSSKYDII.NGMVSS...QYKQVKNLKGYKV
BsAppA/1-543  RPY.......LDTVTYKVIPDANAAVAQLQAGDINFF..NVPAT...DYKT.AEKFNNLK
EcSapA/1-547  RHDDFWRGKPLM...|..PQVVVDLG.SGGTGRLSKLLTGECDVLA..WPAA...SQLSI
StOppA/1-543  AKTV......INQVTYLPISSEVTDVNRYRSGEIDMTYNNMPIE...LFQK.LKKE....
EcMppA/1-537  AKTV......LQKVTFLPINQESAATKRYLAGDIDIT.ESFPKN...MYQK.LLKD....


                β14          β15                   η3       α8             α9
              ──────►     ──────►              TT ℓℓℓ  ℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓ
                  300       310             320       330       340
EcYejA/1-604  IKDEQKNESAQDTRWLAFN........IQRPVFSDRRVREAITLAFDFEWMNKALFYNAW
LlOppA/1-600  LGQQAMYISLMYYNLGHYDAKNSINVQDRKTPLQDQNVRQAIGYARNVAEVDNKFSNGLS
BsAppA/1-543  IVTDLA....LSYVYIGWN.E.......KNELFKDKKVRQALTTALDRESIVSQVLDGDG
EcSapA/1-547  LRDD....PRLRLTLRPGMNVAYLAFNTAKPPLNNPAVRHALALAINNQRLMQSIYYGTA
StOppA/1-543  IPNEVRVDPYLCTYYYEIN........NQKAPFNDVRVRTALKLALDRDIIVNKVK.NQG
EcMppA/1-537  IPGQVYTPPQLGTYYYAFN........TQKGPTADQRVRLALSMTIDRRLMTEKVL.GTG


                β16                   α10       η4      α11
              ──►            TT      ℓℓℓℓℓℓℓℓℓℓℓ ℓℓℓℓ  ℓℓℓ                    ℓℓℓℓ...
                  350       360       370       380       390       400
EcYejA/1-604  SRTNSYFQNTEYAARNYPDAAELVLLAPMKKDLPSEVFTQIYQPPVSKGDGYDRDNL...
LlOppA/1-600  T.P................ANSLIPPIFKQFT.........SSSVKGYEKQDLD..
BsAppA/1-543  E.V................AYIPESPLSWNYPK.........DIDVPKF.EYNEK..
EcSapA/1-547  ET.................AASILPRASWAYDN.........E...AKITEYNPAKS
StOppA/1-543  DLP................AYSYTPPYTDGAKL.........V..EPEWFKWSQQKR
EcMppA/1-537  EKP................AWHFTPDVTAGFTP.........E..PSPFEQMSQEEL
```

```
                  α12              β17     β18           β19              α13
EcYejA/1-604   .llllllllll  ....→      →TT  →TTT      →      →TT  l........llllllll.
                    410        420         430         440          450

EcYejA/1-604   .LKADKLLNEAGWVLKGQQRVNATTGQPLSFELLLPASSNSQWV........LPFQHSL.
LlOppA/1-600   ..KANKLLDEDGWKLNKSTGYREK..DGKELSLVYAARVGD....ANAETIAQNYIQQW.
BsAppA/1-543   ..KAKQMLAEAGWKDTNGDGILDK..DGKKFSFTLKTNQGN....KVREDIAVVQEQL.
EcSapA/1-547   REQLK...SLGLE............NLTLKLWVPTRSQAWNPSPLKTAE.LIQADM..
StOppA/1-543   NEEAKKLLAEAGFTADK..........PLTFDLL.......YNTSDLHKKLAIAVASIWK
EcMppA/1-537   NAQAKTLLSAAGYGPQK..........PLKLTLL.......YNTSENHQKIAIAVASMWK


                      β20       α14            β21              α15   η5
EcYejA/1-604   ll   ...→    llllllllll        →         llllll llll
                     460       470         480         490          500

EcYejA/1-604   QRLGINMDI...RKVDNSQITNRMRS....RDYDMMPRVWRAMP.WPSSDLQISWSSEYI
LlOppA/1-600   KKIGVKVSLYNGKLMEFNSWVDHMTTPPGANDWDITDGSWSLAS.EPSQ..QDLFSAAAP
BsAppA/1-543   KKIGIEVKT...QIVEWSALVEQMNPP..NWDFDAMVMGWSLST.FPDQ..YDIFHSSQI
EcSapA/1-547   AQVGVKVVI...VPVEGRFQEARLMD....MSHDLTLSGWATDSNDPDSFFRPLLSCAAI
StOppA/1-543   KNLGVNVNL...ENQEWKTFLDTRHQ....GTFDVARAGWCADYNEPTSFLNTMLSD...
EcMppA/1-537   KNLGVDVKL...QNQEWKTYIDSRNT....GNFDVIRASWVGDYNEPSTFLTLLTST...
                                                        ▲

                              α16             α17            β22      β23
EcYejA/1-604   llllllllllll    llll.lllllllllll         →         →
                     510       520         530         540          550

EcYejA/1-604   NSTYNAPGVQSPVIDSLINQTIAAQGNKEKL.LPLGRALDRVLTWNYYMLP.....MWYM
LlOppA/1-600   ...YNFGHFNDSEITKDLNDIDSAKSENPTYRKAAFVKYQEDMNKKAYVIPTNFMLNYTP
BsAppA/1-543   KKGLNYVWYKNAEADKLMKDAKSI.SDRKQYSKEYEQIYQKIAEDQPY......TFLYYP
EcSapA/1-547   HSQTNLAHWCDPKFDSVLRKALSSQ..QLAARIEAYDEAQSILAQELPILP.....LASS
StOppA/1-543   .SSNNTAHYKSPAFDKLIADTLKVA..DDTQRSELYAKAEQQLDKDSAIVP.....VYYY
EcMppA/1-537   .HSGNISRFNNPAYDKVLAQASTEN..TVKARNADYNAAEKILMEQAPIAP.....IYQY


                  β24     β25          η6  β26  α18
EcYejA/1-604   →       →            lll  →   llll
                    560       570         580         590          600

EcYejA/1-604   AEDRLAWWDKFSQPAVRPIYSLGIDTWWYDVNKAAKLPSASKQGE
LlOppA/1-600   VNKRVV.....GM....TLDYGAMNTWSEIGVSSAKLATK.....
BsAppA/1-543   NNHMAMPENLEGYKYHPKRDLYNIEKWWLAK.............
EcSapA/1-547   LRLQAYRYDIKGLVL.SPFGNASFAGVYREKQDEVKKP.......
StOppA/1-543   VNARLVKPWVGGYTGKDPLDNIYVKNLYIIKH.............
EcMppA/1-537   TNGRLIKPWLKGYPINNPEDVAYSRTMYIVKH.............
```

Figure 4.22. Alignment showing capping aspartate in the binding pocket of some Cluster C SBPs.

The capping aspartate residue (residue 489) is indicated by the red arrow below the alignment, EcYejA secondary structure is shown along the top of the alignment and the residue numbers shown in the alignment are detailed next to the sequence name. An aspartate residue, which caps the end of the binding pocket and limits peptide ligand size, can be seen in the binding pocket of StOppA and EcMppA, it has also been predicted to be in the binding site of EcSapA. No aspartate capping residue can be seen in the binding pocket of EcYejA, LlOppA or BsAppA.

It is clear to see from the structure of EcYejA that there is a large internal binding pocket that is negatively charged (red on an electrostatic surface image) (Figure 4.23). As CAMPs are positively charged this large negatively charged binding pocket fits with the hypothesis that EcYejA binds CAMPs. It is also interesting to note that GEP does not actually sit in the large negatively charged region of the binding pocket, it sits further down on the edge of the binding pocket, taking up very little space in the cavity (Figure 4.24).

By superposing some of the known Cluster C SBP structures onto that of EcYejA it is possible to view the binding pockets and whether the ligands are located in a similar place in the proteins. From Figure 4.25 it is clear to see that the GEP peptide is not located in a similar position in EcYejA as the ligands in other Cluster C SBPs. The other ligands superpose onto each other well even though the ligands are a range of different lengths and some of the binding pockets contain the capping aspartate residue. This indicates that there is something different about the binding pocket of EcYejA as compared to other Cluster C SBPs binding pockets.

It is also interesting to note that the binding pockets of EcMppA, StOppA and BsAppA are more enclosed than those of LlOppA and EcYejA (Figure 4.26 and 4.27). One side of the GEP peptide in EcYejA is being coordinated by water molecules as there is a vast space available in the binding cavity of EcYejA that is filled by water molecules. This seems to indicate that there can be some variability in length and sequence in the Cluster C SBPs that bind longer peptides as the binding pocket does not seem as enclosed and specific.

Using rbcavty within rdock it was possible to calculate the volume of the binding pockets, with water molecules removed, of some Cluster C SBPs (Table 4.3). From these results it is clear that as the size of the ligand increases the size of the binding pocket increases accordingly, as can be expected. However, currently EcYejA has only been shown to bind an 11 residue peptide but it has the largest binding pocket volume by a large margin. This suggests that EcYejA's natural ligand is much larger than 11 residues, possibly even larger than 35 residues as the binding pocket volume is larger than that of LlOppA.
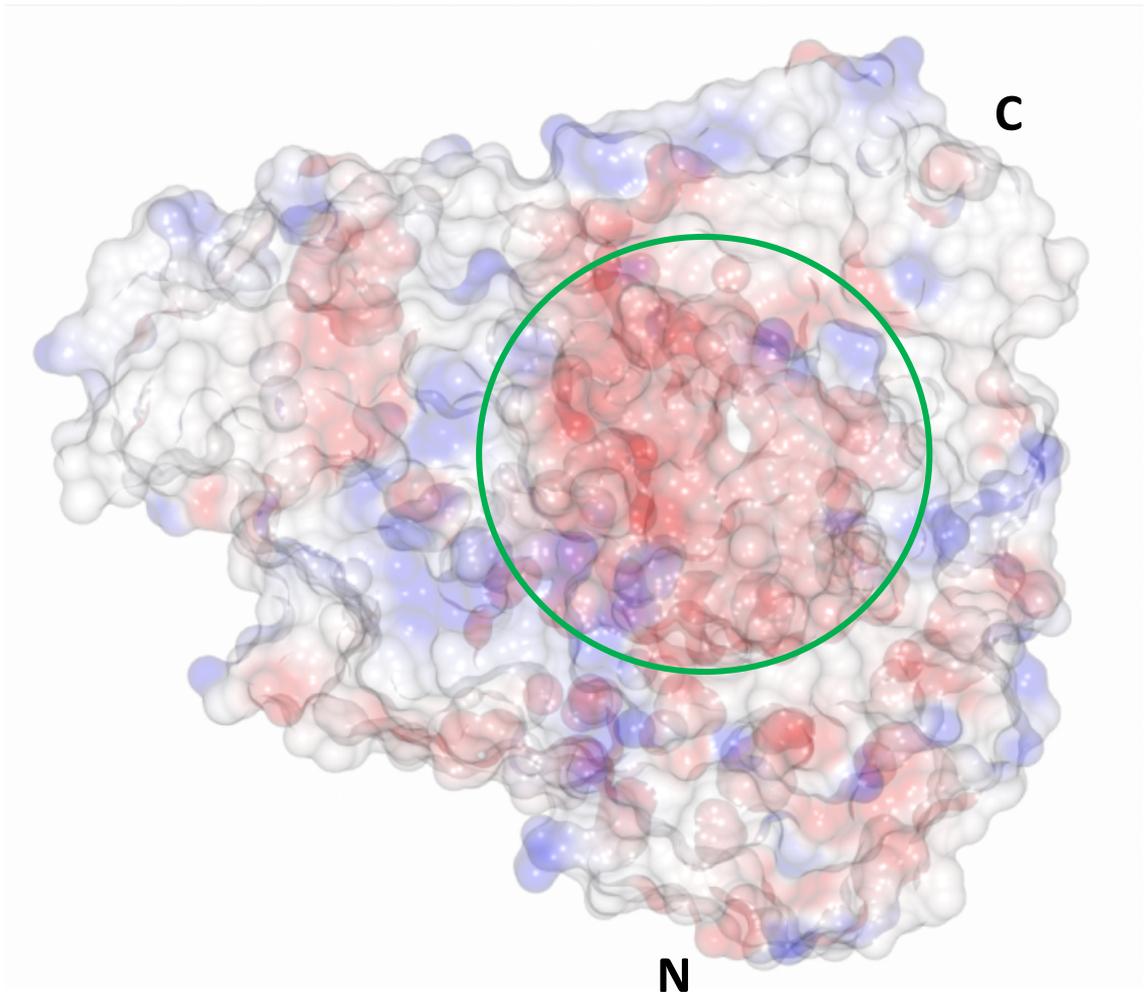
Figure 4.23. EcYejA has a large negatively charged binding pocket.

EcYejA is shown as an electrostatic surface, red indicates negative charge and blue indicates a positive charge, white is neutral. The green circle indicates the binding pocket of EcYejA.
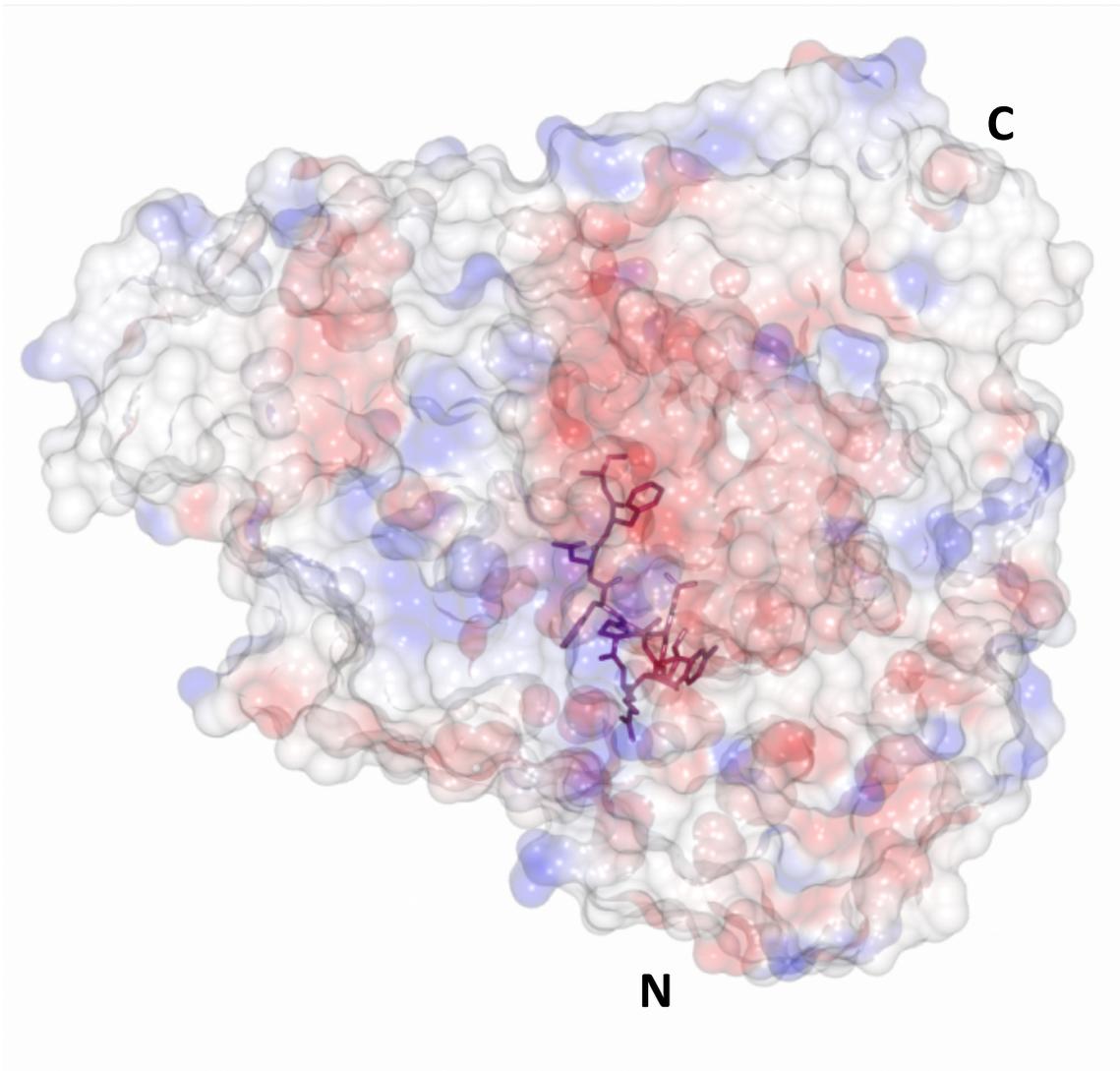
Figure 4.24. GEP does not sit in the centre of the EcYejA binding pocket.

EcYejA is shown as an electrostatic surface, red indicates negative charge and blue indicates a positive charge, white is neutral. GEP is shown in purple.
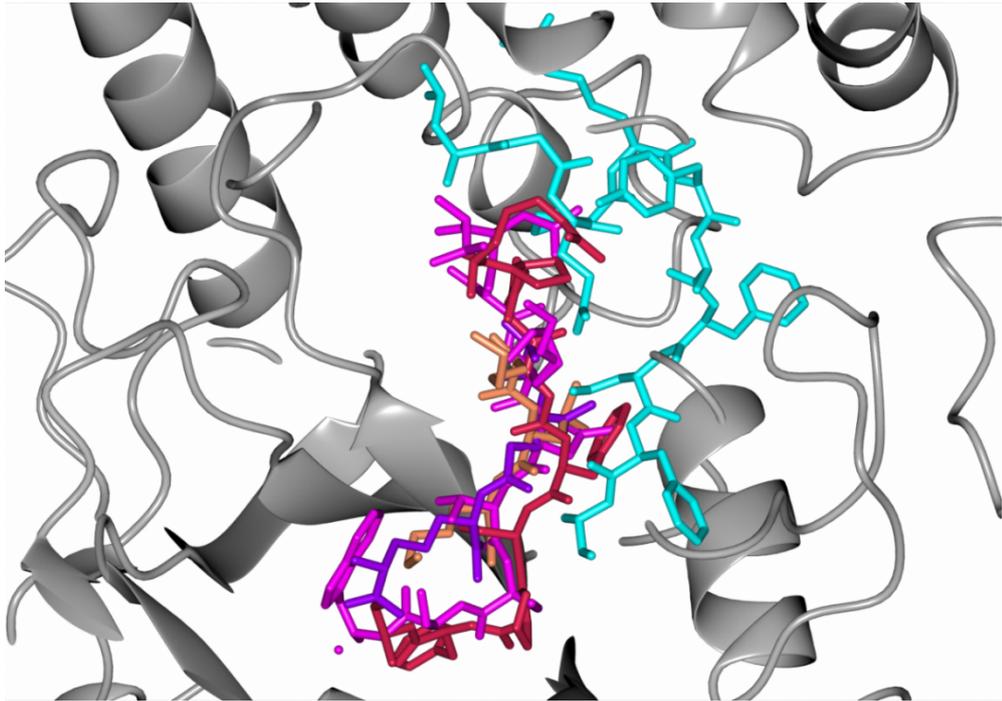
Figure 4.25. GEP ligand is not located in the same position in the binding pocket as other Cluster C SBP ligands.

EcYejA is shown in grey ribbon form, the GEP peptide is shown in blue, the BsAppA nonapeptide (PDB:1XOC) is shown in pink, the LlOppA bradykinin ligand (PDB:3DRG) is shown in red, the EcMppA murein tripeptide ligand (PDB:3O9P) is shown in purple and the StOppA KLK tripeptide (PDB:1B9J) is shown in coral. Cluster C SBPs were superposed onto EcYejA to create figures.
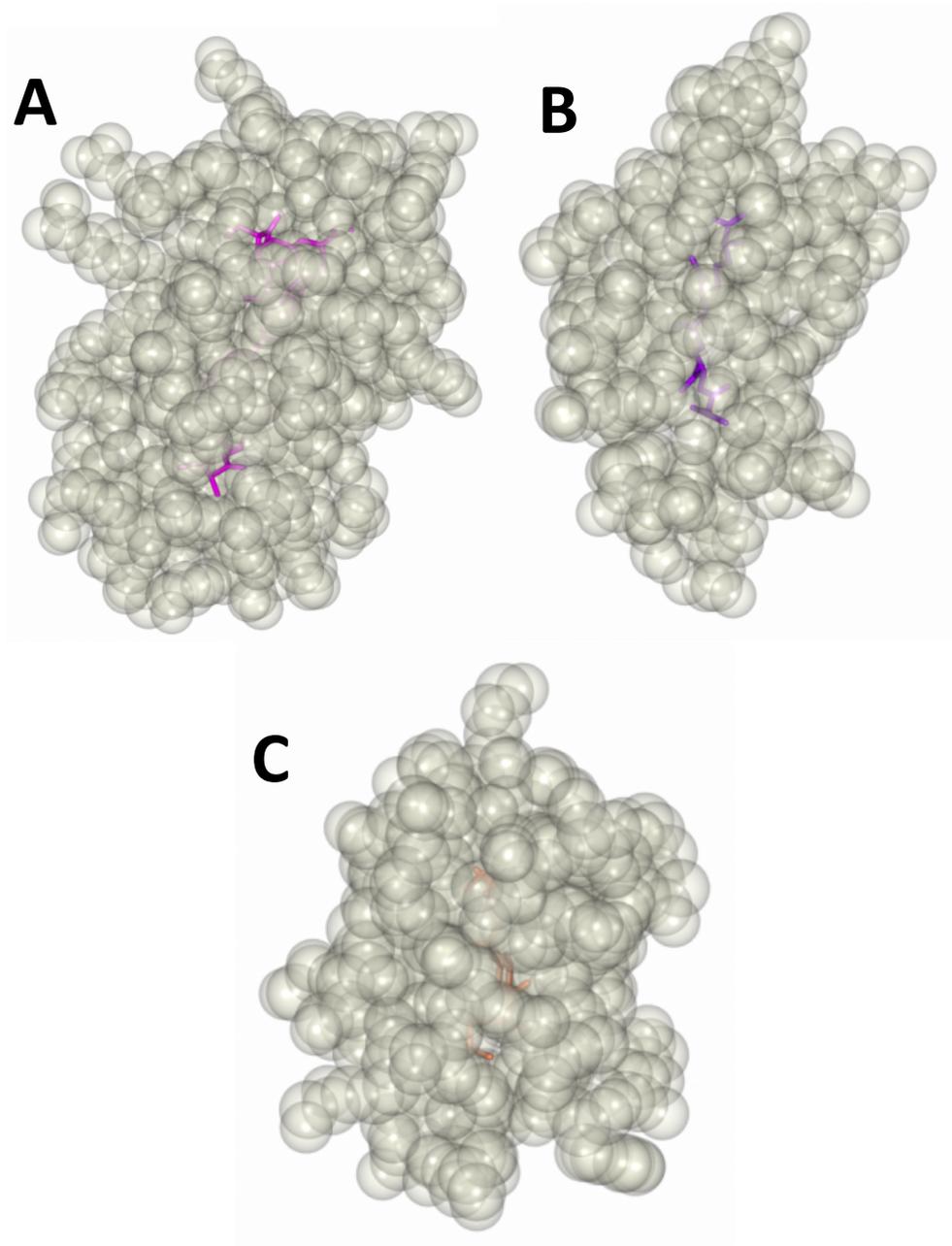
Figure 4.26. The binding pockets of EcMppA, StOppA and BsAppA are enclosed.

EcMppA, StOppA and BsAppA are shown in beige sphere form with transparency set to
25%. (A) BsAppA nonapeptide ligand is shown in pink cylinder form, (B) EcMppA murein
tripeptide ligand is shown in purple cylinder form and (C) StOppA KLK ligand is shown in
orange cylinder form.
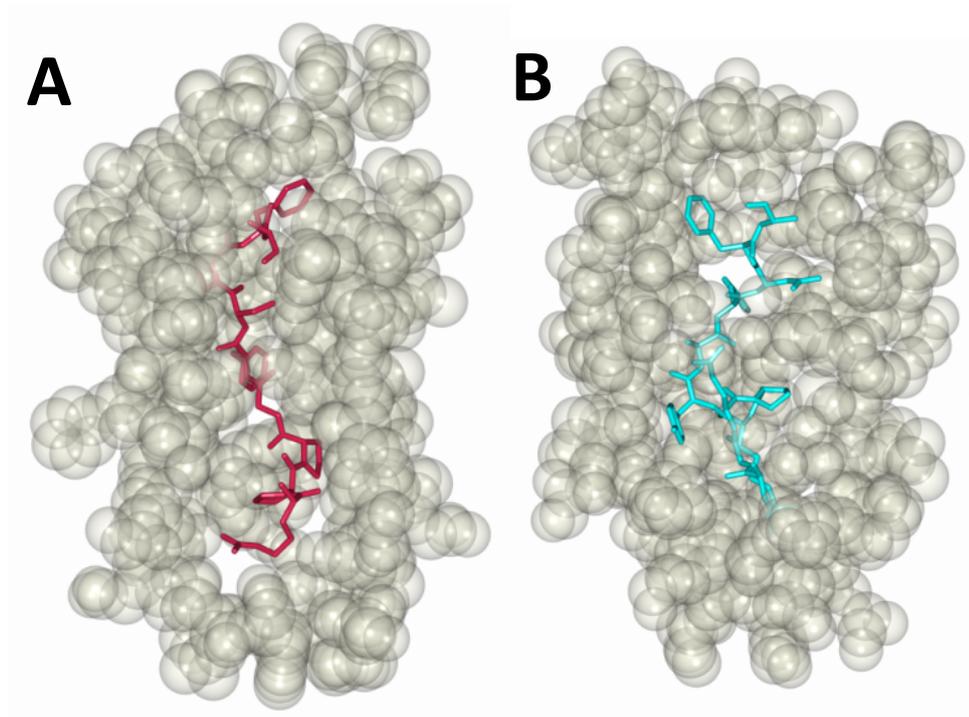
Figure 4.27. The binding pockets of LlOppA and EcYejA are more open than EcMppA, StOppA and BsAppA.

LlOppA and EcYejA are shown in beige sphere form with transparency set to 25%.

(A)LlOppA nonapeptide ligand is shown in red cylinder form and (B) EcYejA GEP ligand is

shown in blue cylinder form.

Protein	Known Length Ranges of Ligands	Binding Cavity Size ($Å^3$)
EcDppA	2 residues	1434
StOppA	2-5 residues	1874
BsAppA	9 residues	3773
LlOppA	Up to 35 residues	5955
EcYejA	11 residues	8253

Table 4.3. Binding cavity sizes of various Cluster C SBPs.

EcDppA is *E. coli* DppA, an SBP which binds dipeptides. The size of the binding pocket increases as the length of the peptide bound increases, as expected. This indicates that EcYejA is able to bind much larger peptides than other Cluster C SBPs as it has the largest binding pocket.

Chapter Five

Ligand Determination of EcYejA

5. Ligand Determination of EcYejA

The crystal structure of EcYejA revealed the protein to be in the closed conformation with a ligand bound. From the electron density maps this ligand was identified as the peptide LGEPRYAFNFN (GEP), and identical in sequence to a segment of the EcYejA polypeptide close to the amino terminus. There is no evidence to suggest that GEP is a natural ligand of EcYejA, and our initial interpretation is that its presence is the result of a crystallisation artefact. To better understand the specificity of EcYejA a series of binding experiments were conducted with various CAMPs and peptides.

5.1 Screening peptides for EcYejA ligands

Thermal shift assays are a screening method for protein ligands based on the melting temperature of the protein. As the protein is heated and denatures a fluorescent dye binds to the hydrophobic surfaces exposed as the protein unfolds and the melting temperature of the protein can be determined. The melting temperature of a protein is defined as the midpoint of the unfolding transition and can be read from the thermal shift assay graphs as the temperature at which the fluorescence value is half of the total fluorescence observed. In theory, if the protein is bound to a ligand the protein will be stabilised and the melting temperature will increase, and this can be measured. A number of different potential ligands were screened for EcYejA binding using thermal shift assays, some of which were carried out by project student Rebecca Lees under the direct supervision of the author. For the thermal shift binding experiments the concentration of EcYejA was kept constant at 0.5 mg/ml (7.3 µM). This value was reached by carrying out a series of thermal shift assay optimisations with different concentrations of EcYejA, 0.5 mg/ml EcYejA gave a clear fluorescence reading and did not use excessive amounts of protein per run. Potential peptide ligands were tested at as high a concentration as possible, with the aim of having a one molar excess over EcYejA, however solubility issues meant some peptides were tested at lower concentrations.

All EcYejA protein used in this chapter was purified without the 2 M guanidine hydrochloride step as this step was deemed unnecessary due to EcYejA crystallising with a ligand bound even after 2 M guanidine hydrochloride treatment. A thermal shift assay was carried out with a sample of EcYejA which had undergone 2 M guanidine hydrochloride treatment and a sample of EcYejA which had not undergone 2 M guanidine hydrochloride treatment, both samples were tested with and without GEP. The melting temperature increase of both EcYejA samples with GEP was the same, indicating that omission of the 2

M guanidine hydrochloride treatment step during purification was not compromising ligand binding studies.

CAMPs were the obvious target for initial screening of EcYejA to identify potential ligands. Previous literature has identified a number of CAMPs, in particular polymyxin B, melittin and LL-37 as being involved with the Yej transporter and possibly bound by YejA (Eswarappa *et al.*, 2008; Wang *et al.*, 2016). The CAMPs LL-37, as well as shorter fragments of LL-37, and polymyxin B were tested here using thermal shift assays.

Another potential EcYejA ligand is the antibiotic Microcin C, a heptapeptide attached to a nucleoside ring that the Yej transporter has been shown to transport into the cytoplasm (Novikova *et al.*, 2007). The heptapeptide section of Mircocin C, MRTGNAD, was synthesised and tested as a ligand for EcYejA. N-terminally formylated Microcin C has also been shown to be transported via the Yej transporter at a higher rate than the deformylated version of Microcin C. Therefore fMRTGNAD was also synthesised via solid phase peptide synthesis to be tested as a ligand for EcYejA. fMRTGNAD(Dansyl-K)G was also synthesised as a mimic of the whole structure of Microcin C, not just the heptapeptide section. All of these potential ligands were screened via thermal shift assays.

GEP and commercially available random peptides of varied sequence and chain length were bought in and also screened via thermal shift assays to try and identify any kind of "binding motif".

P. aeruginosa PhnD, a binding protein for 2-AEP, was used as a control for the thermal shift assays (Sophie Rugg). The GEP ligand was also used as a positive control for the assays. A thermal shift in melting temperature of anything equal to or more than +3 °C was deemed as binding to EcYejA. This value was chosen arbitrarily as variations in melting temperature of up to ±1/2 °C were seen across different thermal shift runs, probably due to different plates being used in different runs and slight differences in protein and dye concentrations. Via thermal shift assays GEP was shown to bind EcYejA (Figure 5.1), however of all the ligands tested GEP was the only one whose presence led to a marked increase in the melting temperature of EcYejA (Table 5.1), indicating that GEP alone is binding to EcYejA.

Unfortunately, GEP was insoluble at the concentrations required (>0.56 mM) for binding affinity testing using Isothermal Titration Calorimetry (ITC), which is the preferred method for obtaining a K_D with these proteins. Therefore, thermal shift assays were used as a way
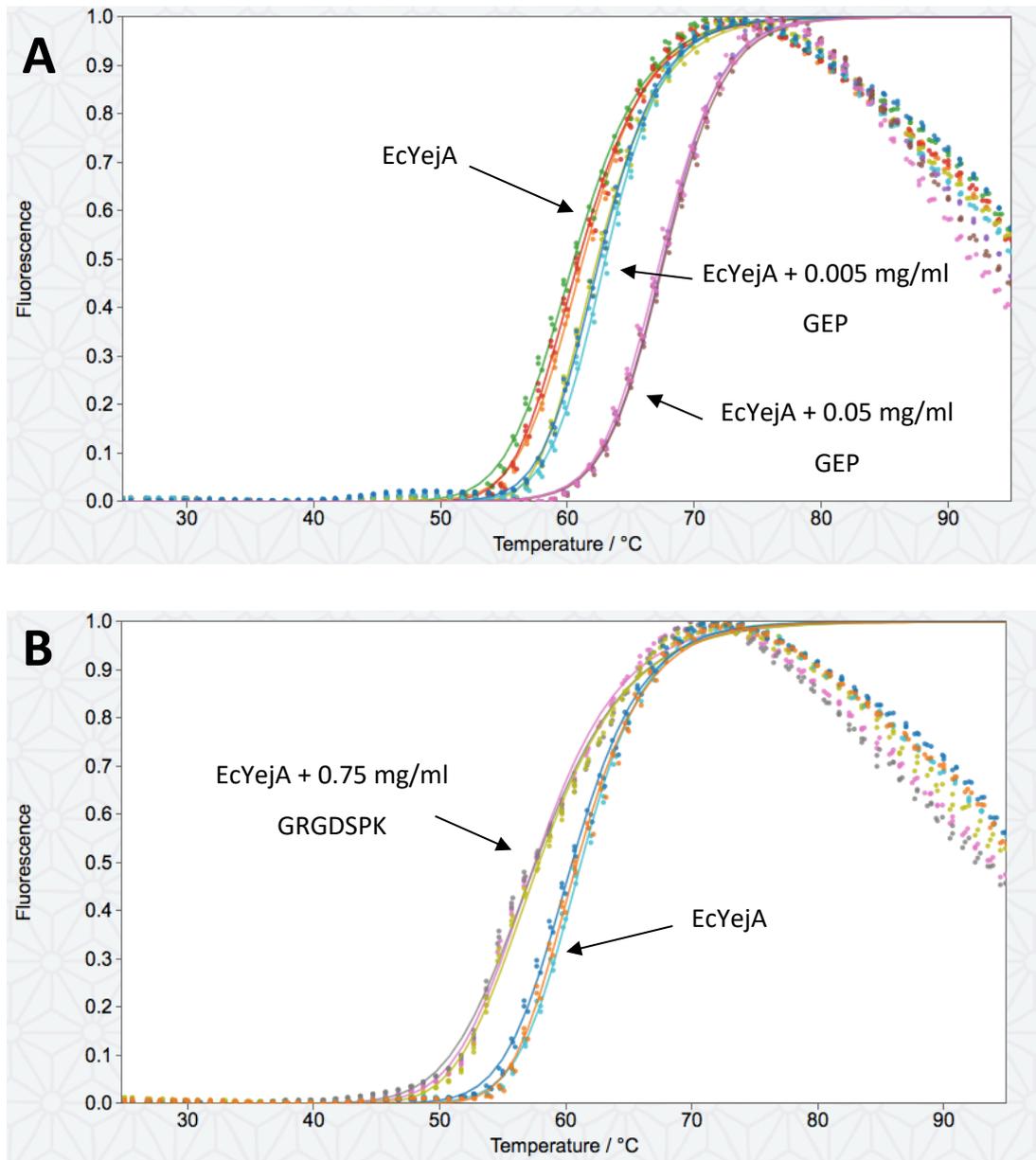
Figure 5.1. Thermal shift assays confirm EcYejA binding of LGEPRYAFNFN.

Both graphs show normalised data with 0.5mg/ml (7.3 µM) EcYejA. (A) shows EcYejA in the absence of added peptide (red, orange, green) T_m=60.8 °C; 0.05mg/ml LGEPRYAFNFN (pink, grey, brown) T_m=67.5 °C; 0.005mg/ml LGEPRYAFNFN (blue, light blue, lime green) T_m= 62.6 °C. (B) shows EcYejA with no ligand (blue, light blue, orange) T_m=60.8 °C; 0.75mg/ml GRGDSPK (grey, lime green, pink) T_m=57.7 °C.

of estimating the binding affinity of EcYejA to GEP. However, GEP was not soluble at high enough concentrations to saturate EcYejA, determined as the stabilisation of the shift in melting temperature, meaning an accurate estimate of the binding affinity could not be made.

5.2 Ligand binding in EcYejA examined by Mass Spectrometry

Identification of ligands bound to EcYejA was carried out using native electrospray ionisation mass spectrometry (native ESI-MS) and matrix-assisted laser desorption/ionisation mass spectrometry mass spectrometry (MALDI-MS/MS) in collaboration with Dr. Adam Dowle who collected and analysed the data. Native ESI-MS is a technique where the biological samples are sprayed from a nondenaturing solvent, in this case 1 M ammonium acetate, to allow the transfer of intact proteins and biomolecular complexes from solution to the gas phase. It is possible to see a difference in mass between a sample of liganded protein and a sample of unliganded protein, the difference in mass between these two samples is the mass of the ligand. The liganded protein sample after native ESI-MS can then be taken and analysed via MALDI-MS/MS. MALDI-MS/MS strips the proteins of ligands and identifies peptide ligands within the range 800-4000 Da, as the mass of the ligand bound to the peptide is known from native ESI-MS that particular peptide ligand mass can be selected. The selected peptide ligand is then fragmented into different smaller pieces which can be searched against a database, in this case the Mascot database, to identify the sequence of the ligand (Figure 5.2). It is important to note that the MALDI-MS/MS data alone shows that the ligands are present in the sample, it does not prove that they are bound to EcYejA. Only the positive mass change in the native ESI-MS signifies binding of that specific ligand. All samples of EcYejA in these experiments had not undergone the 2 M Guanidinium-HCl treatment step during purification.

5.2.1 Confirmation of EcYejA binding of GEP ligand

To confirm binding of EcYejA to GEP an initial native ESI-MS experiment was carried out on a sample of EcYejA, at a concentration of 1.46 µM. The expected mass of EcYejA is 68,381.73 Da, a peak can be seen at a mass of 68,381.89 Da (Figure 5.3) which gives a mass difference of 0.16 Da. A further native ESI-MS experiment was then conducted on a mixture of EcYejA (1.46 µM) and GEP (1.51 µM). The expected mass of an EcYejA-GEP complex is 69,709.17 Da (GEP mass = 1,326.64Da), a peak can be seen on the spectra at 69,709.31 Da (Figure 5.4). This is consistent with a GEP-EcYejA protein complex with a 0.14 Da mass error.

Ligand	Ligand Concentration	Melting Temperature Change (°C)
Polymyxin B	5 mg/ml (3.8 mM)	-1.3
LL-37	0.75 mg/ml (0.17 mM)	-0.8
LL 13-37	3 mg/ml (0.99 mM)	-1.1
LL 19-29	3 mg/ml (2.1 mM)	+0.3
LL 17-32	1 mg/ml (0.49 mM)	-3.6
RGDSPASSKP	0.75 mg/ml (0.75 mM)	+0.2
KKK	0.75 mg/ml (1.7 mM)	-0.2
RGDS	0.75 mg/ml (1.7 mM)	+2
GRGDSPK	0.75 mg/ml (1.0 mM)	-3.17
DWKDDDK	0.25 mg/ml (0.27 mM)	+0.5
Melittin	0.75 mg/ml (0.26 mM)	-5.4
KGG	3 mg/ml (10.1 mM)	-0.7
AK Hydrochloride	5 mg/ml (19.7 mM)	0
AQ	5 mg/ml (21.3 mM)	-0.2
AG	3 mg/ml (18.3 mM)	-0.1
GPRP amide	5 mg/ml (11.8 mM)	-0.3
AP Hydrate	3 mg/ml (16.1 mM)	0
Bradykinin	6 mg/ml (5.7 mM)	-1.4
LGEPRYAFNFN	0.75 mg/ml (0.56 mM)	+13
fMRTGNAD(Dansyl-K)G	1.5 mg/ml (1.2 mM)	+0.3
fMRTGNAD	1.5 mg/ml (1.89 mM)	-0.1
MRTGNAD	1.5 mg/ml (1.96 mM)	-0.3

Table 5.1. Ligands binding to EcYejA monitored by thermal shift assays.

Potential peptide ligands were tested at as high a concentration as possible, with the aim of having a one molar excess over EcYejA, however solubility issues meant some peptides were tested at lower concentrations. The EcYejA concentration in all experiments was 0.5 mg/ml (7.3 µM). A thermal shift in melting temperature of anything equal to or more than +3 °C was deemed as binding to EcYejA. This value was chosen arbitrarily as variations in melting temperature of up to ±1/2 °C were seen across different thermal shift runs.
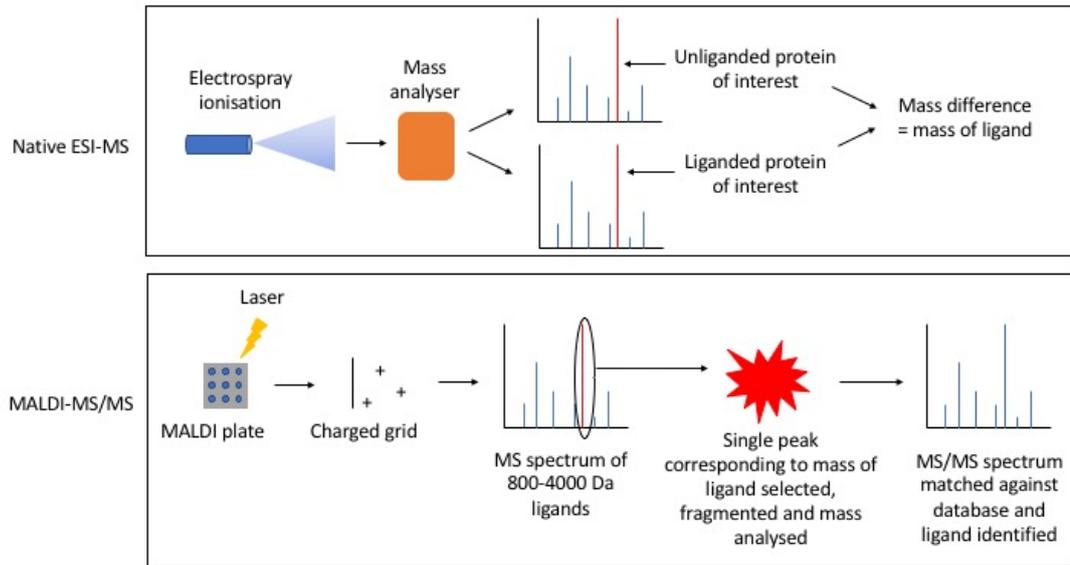
Figure 5.2. Schematic diagram of native ESI-MS and MALDI-MS/MS protein analysis.

Native ESI-MS measures the mass of intact proteins and complexes by spraying the samples from a nondenaturing solvent. In MALDI-MS/MS a single mass peak can be selected, fragmented, mass analysed and then matched against a database to identify the sequence.
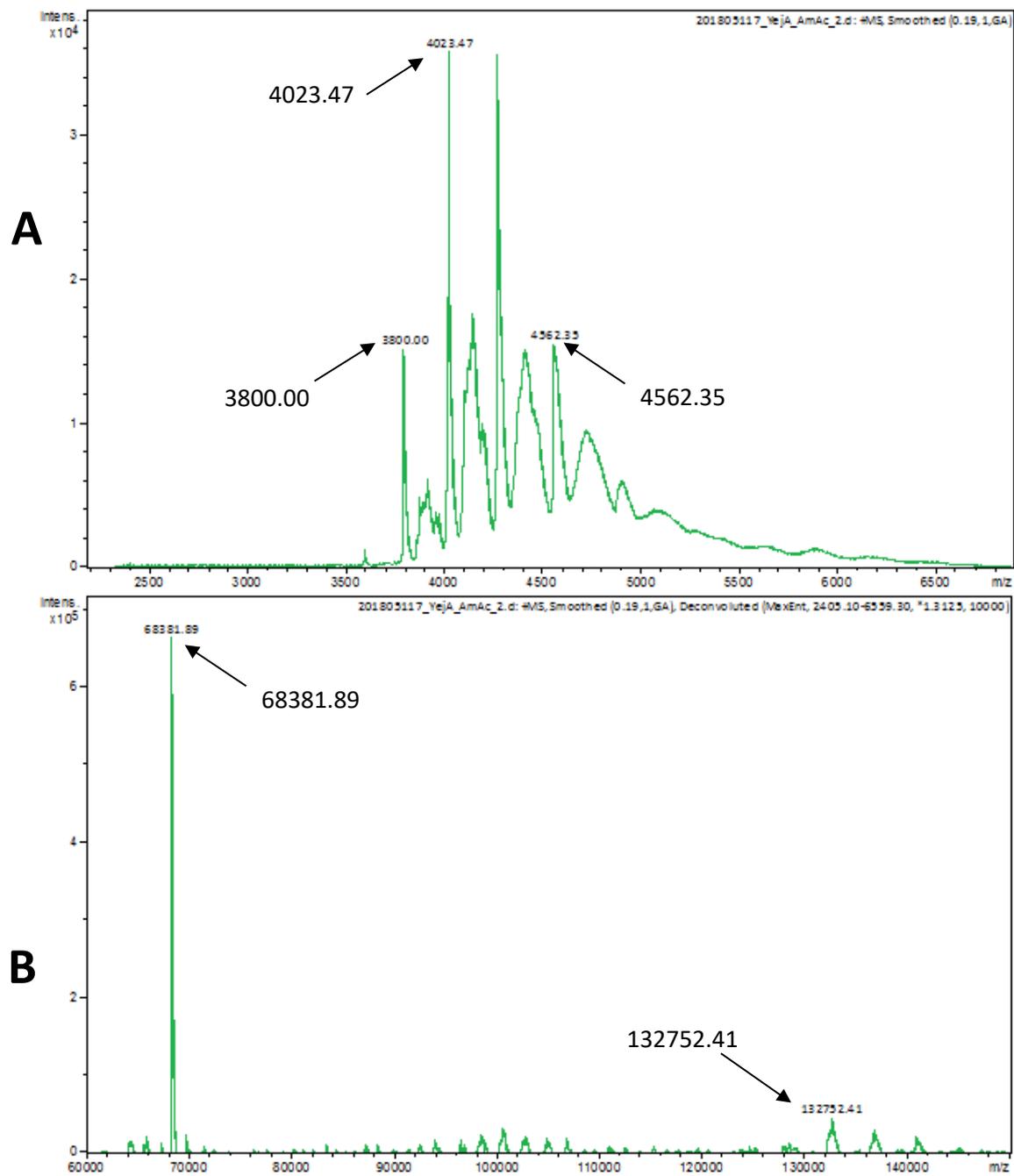
Figure 5.3. Mass spectra of EcYejA.

The above spectra were produced by the infusion of EcYejA. (A) shows the convoluted spectra. (B) shows the deconvoluted spectra of EcYejA and shows a peak at 68,381.89 Da.
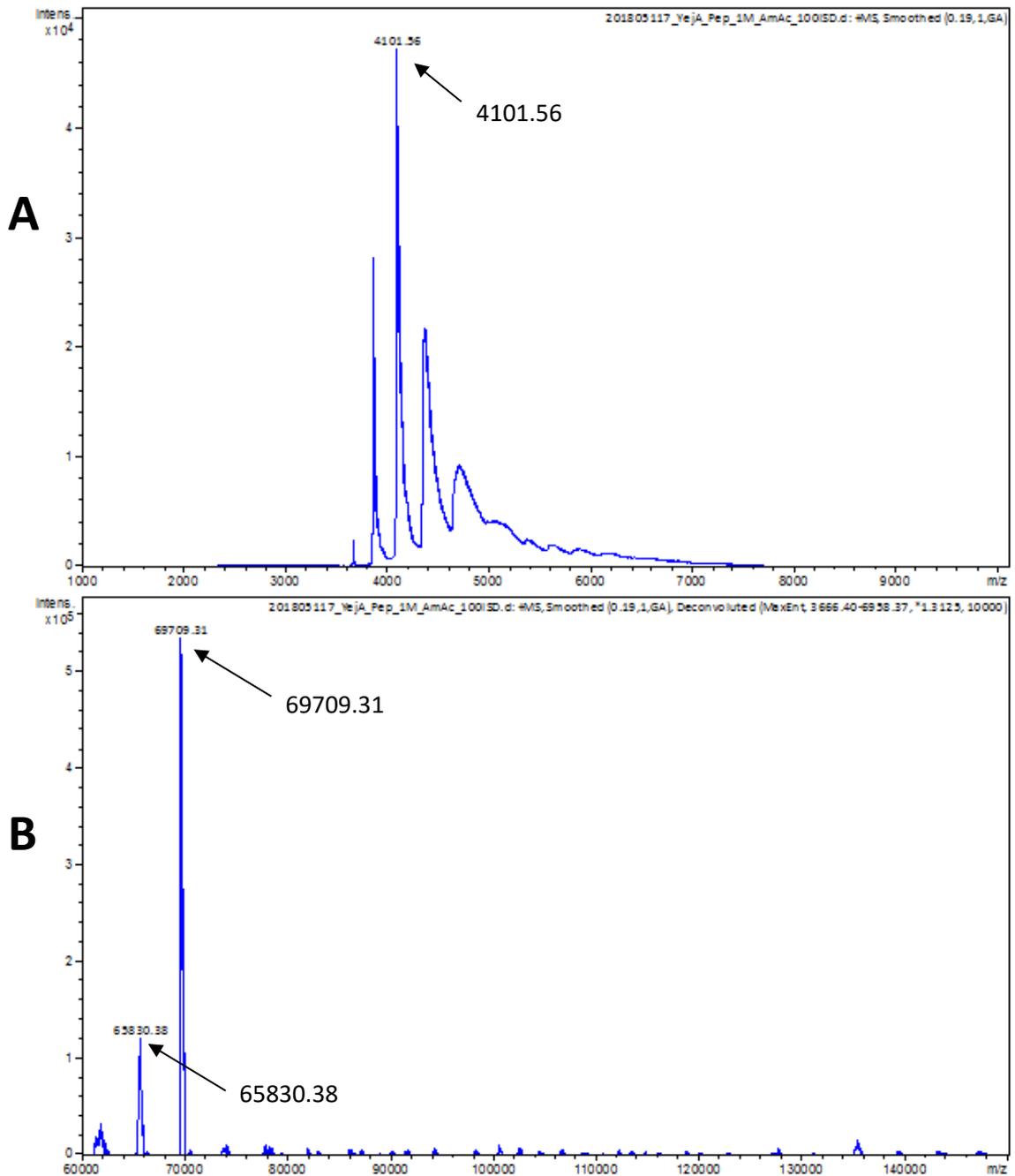
Figure 5.4. GEP ligand binds EcYejA.

Above spectra were acquired from an EcYejA and GEP peptide mix. (A) shows the convoluted spectra. (B) shows the deconvoluted spectra and shows a peak at 69,709.31 Da.

Due to the inability to carry out other techniques such as ITC, native ESI-MS was used as a way to try and estimate the binding affinity of EcYejA to the GEP ligand. This was achieved by titrating GEP ligand into the EcYejA sample until saturation of EcYejA. EcYejA was used at a final concentration of 1.46 µM and the GEP ligand was added from 0-1.51 µM. For this experiment a new sample of EcYejA had been produced and purified and from the mass spectrometry data appeared to come pre-bound with an endogenous ligand with a mass of 1424.72 Da (Figure 5.5). On the spectra acquired from the 1.46 µM EcYejA and 0 µM GEP infusion a dominant peak can be seen at a mass of 68,381.96 Da, this peak corresponds to the unliganded form of EcYejA. However, there is another clear peak on the spectra at a higher mass of 69,806.68 Da. Although the intensity of this peak is lower it does indicate a ligand bound state of EcYejA with a ligand of mass 1424.72 Da. This ligand could not be completely competed off by the addition of 1.51 µM GEP ligand, indicating that it is tightly bound to EcYejA (Figure 5.6). When the GEP concentration is 0.755 µM, there is no free EcYejA. At this point all EcYejA is either in complex with the GEP peptide or in complex with the 1424.72 Da ligand. From the addition of 0.755 µM to 1.51 µM of GEP peptide, 42.3 % of the 1424.72 Da ligand was competed off EcYejA and replaced with GEP peptide. Approximately 30 % of the initial EcYejA sample from this preparation was in complex with the endogenous 1424.72 Da ligand. An attempt was made to estimate the binding affinity of EcYejA to GEP but it was deemed too unreliable to make any real estimate. This was because approximately 70% of EcYejA (1.02 µM) was unliganded before any GEP titration and was completely saturated by 0.755 µM GEP, which is less than a 1:1 ratio. It is possible that liganded EcYejA flies better in the mass spectrometer than unliganded EcYejA and therefore a higher percentage of the liganded EcYejA is detected, which could give the data the appearance that EcYejA was saturated at a less than 1:1 ratio.

To identify the 1424.72 Da ligand MALDI-MS/MS was used in tandem with the Mascot database to identify the ligand as VLGEPRYAFNFN. This ligand is GEP with an additional valine residue on the N-terminus of the peptide, VLGEPRYAFNFN is again located in the N-terminus of EcYejA. In the crystal structure of EcYejA, there is volume within the binding pocket of EcYejA to add the valine residue onto the GEP ligand. It is possible that the actual ligand in the crystal structure is VLGEPRYAFNFN, and that the electron density for the valine residue is too diffuse to allow its building and refinement (Figure 5.7).
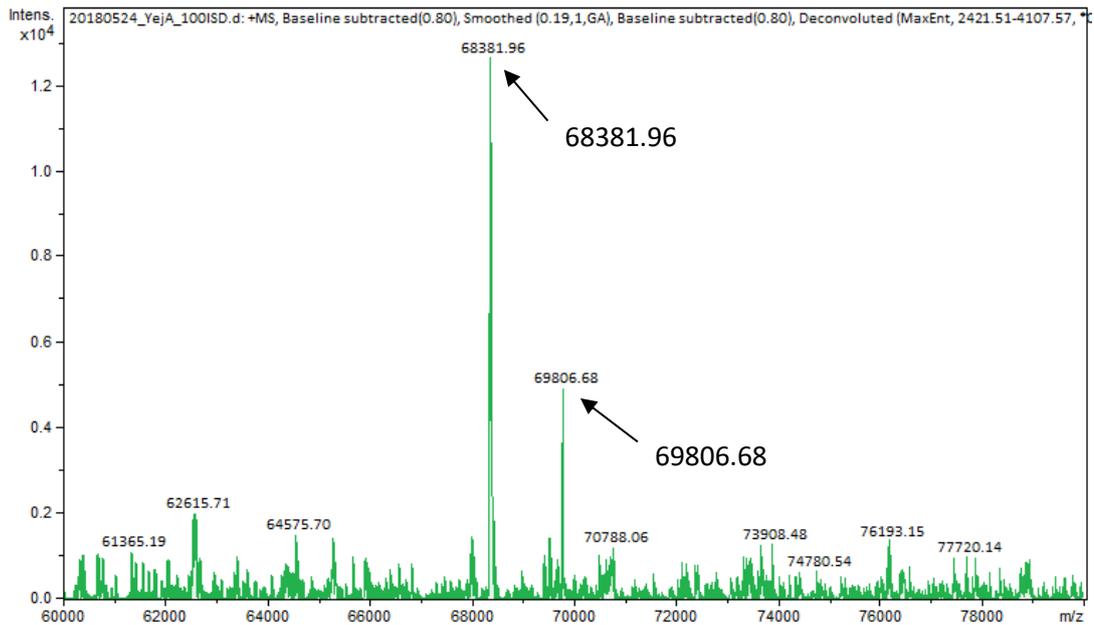
Figure 5.5. EcYejA is co-purified with an unknown ligand.

Mass spectra shows peaks of 68,381.96 Da and 69,806.68 Da, which corresponds to un-liganded EcYejA and EcYejA in complex with a ligand of 1424.72 Da respectively.
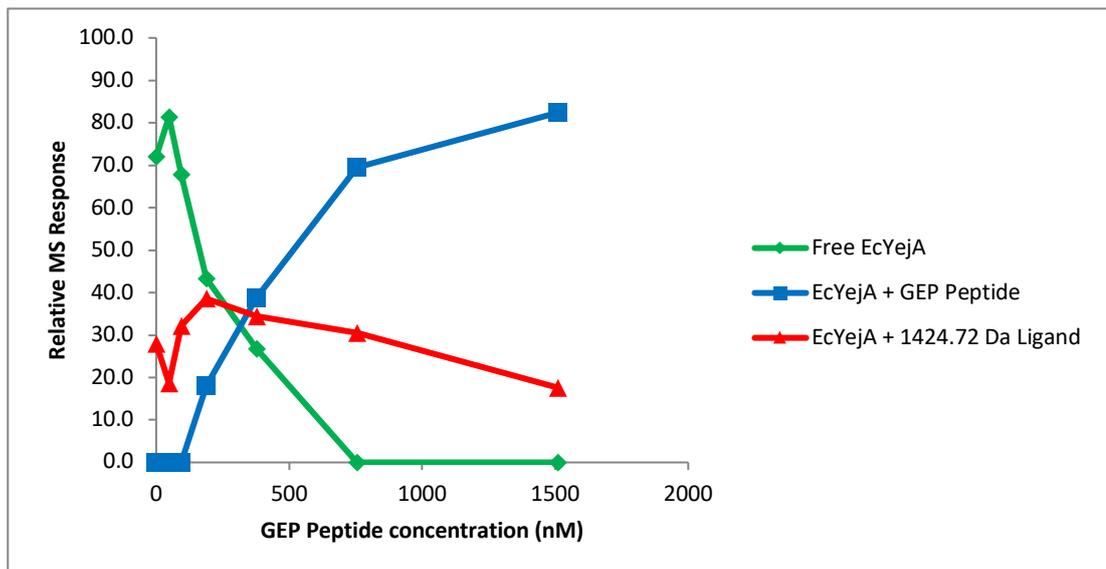
Figure 5.6. GEP ligand has a high binding affinity for EcYejA.

EcYejA was used at a final concentration of 1.46 µM and the GEP ligand was added from 0-1.51 µM. The relative intensities of the peaks corresponding to the free EcYejA, EcYejA + GEP and EcYejA + 1424.72 Da ligand were calculated from the spectra at each GEP ligand concentration.
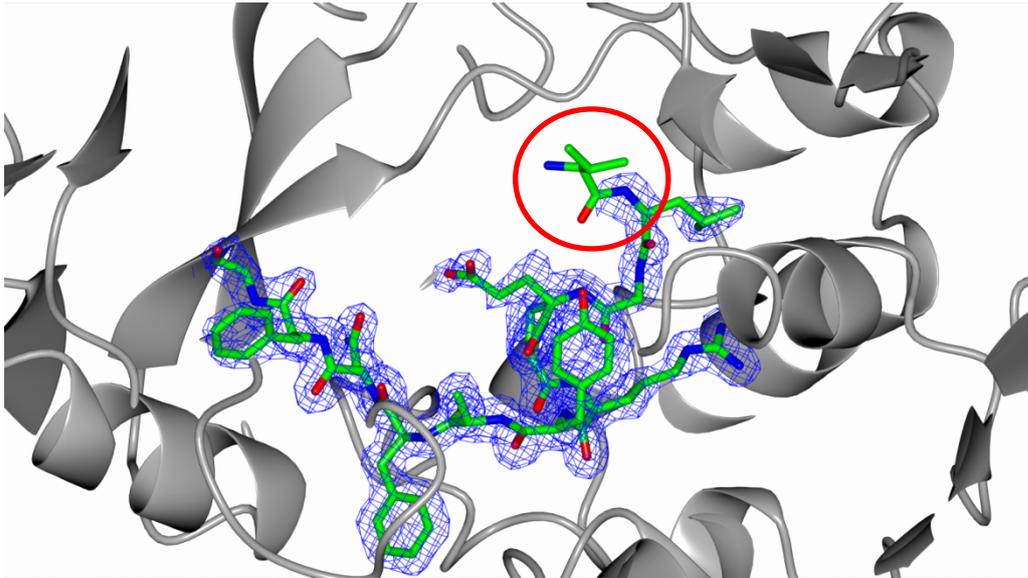
Figure 5.7. There is space in the binding pocket of EcYejA for VLGEPRYAFNFN.

EcYejA is shown in grey ribbon form, electron density is shown in blue, LGEPRYAFNFN with an additional valine on the N-terminus is shown in green, red and blue cylinder form. The red circle shows an additional valine residue on the N-terminus of GEP in the binding pocket of EcYejA. The additional valine residue is in space in the binding pocket of EcYejA and not clashing with any EcYejA residues.

5.2.2 Identification of ligands co-purifying with EcYejA

With the aim of identifying any endogenous ligands bound to EcYejA a fresh sample of EcYejA was prepared and native ESI-MS and MALDI-MS/MS was carried out on the sample. It was thought that due to the lack of the 2 M guanidinium-HCl treatment step during purification that any endogenous ligand acquired from either the LB growth media or the *E. coli* BL21(DE3) expression cells would remain bound to EcYejA throughout the purification process and should be seen in the data obtained. This sample of EcYejA was purified and then immediately subjected to native ESI-MS and MALDI-MS/MS. The MALDI-MS/MS identified species with a range of molecular masses as potential EcYejA ligands. As may be seen from Table 5.2, all are overlapping peptides derived from the N-terminus of EcYejA.

The native ESI-MS data were collected at ion cone voltages of 100 eV and 200 eV. A higher cone voltage delivers more energy to the protein complex and this is expected to increase the level of ligand dissociation from the complex during the experiment. Spectra collected at 100 eV are therefore more likely to represent the solution state, while spectra from samples collected at 200 eV will inform on the unliganded protein. The difference in mass then informs on the presence and nature of the ligand. At 200 eV the mass of the uncomplexed EcYejA is 1,427 Da lower than that recorded at 100 eV indicating the presence of bound VLGEPRYAFNFN (Figure 5.8). Two other peaks were also observed with masses 1,824 Da and 1,989 Da greater than uncomplexed EcYejA, consistent with the presence of EcYejA complexes with the peptides VLGEPRYAFNFNHFD and VLGEPRYAFNFNHFDY respectively. These species are present even in 200 eV spectra, suggesting these peptides are tightly bound.

At the lower ion cone voltage of 100 eV, little or no uncomplexed EcYejA protein is observed (Figure 5.9). The dominant species present was identified as EcYejA in complex with VLGEPRYAFNFN. The other peaks in this spectrum correspond to EcYejA in complex with: VLGEPRYAFNFNHFD; VLGEPRYAFNFNHFDY; AVLGEPRYAFNFNHFDY and AFAVLGEPRYAFNFNHFDY.

A further sample of EcYejA was also analysed by native ESI-MS and MALDI MS/MS. This sample was exposed overnight to a periplasmic extract from *E. coli* BW25113. Although EcYejA is a periplasmic protein when it is natively expressed and produced, during this work it has been produced within the cytoplasm for ease of production and purification.

Ligand	Calculated mass (Da)
VLGEPRYAFNF	1311.661
LGEPRYAFNFN	1326.636
VLGEPRYAFNFN	1425.704
VLGEPRYAFNFNH	1562.763
AVLGEPRYAFNFNH	1633.8
FAVLGEPRYAFNFN	1643.81
VLGEPRYAFNFNHF	1709.831
FAVLGEPRYAFNFNH	1780.869
VLGEPRYAFNFNHFD	1824.858
VLGEPRYAFNFNHFDY	1987.922
AFAVLGEPRYAFNFNHF	1998.974
AVLGEPRYAFNFNHFDY	2058.959
AFAVLGEPRYAFNFNHFDY	2277.064

Table 5.2. MALDI-MS/MS identified several peptides from the N-terminus of EcYejA as potential EcYejA ligands.

This data was generated from a sample EcYejA that had been purified and then immediately subjected to native ESI-MS and MALDI-MS/MS analysis.
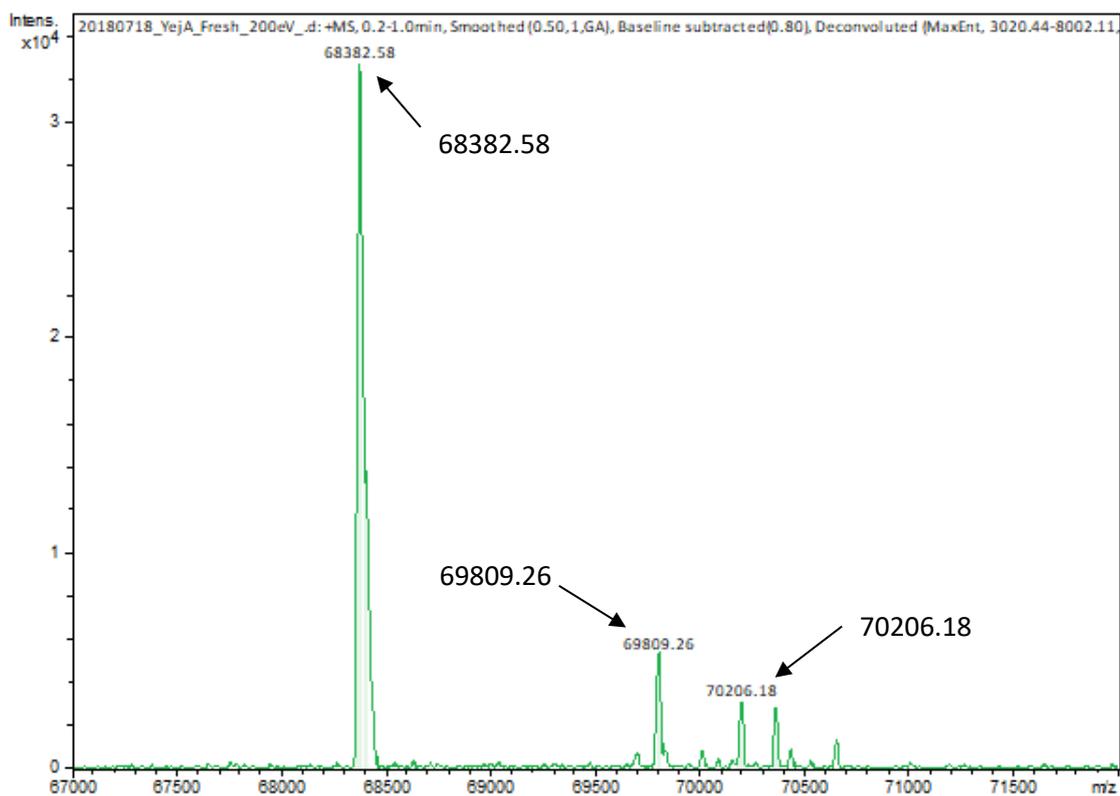
Figure 5.8. Native ESI-MS at 200 eV shows EcYejA in complex with several ligands.

The peak at 68,382.58 Da indicates the un-complexed form of EcYejA. Peaks can also be seen at 1,426 Da, 1,824 Da and 1,989 Da greater than un-complexed EcYejA, these correspond to the ligands VLGEPRYAFNFN, VLGEPRYAFNFNHFD and VLGEPRYAFNFNHFDY respectively.
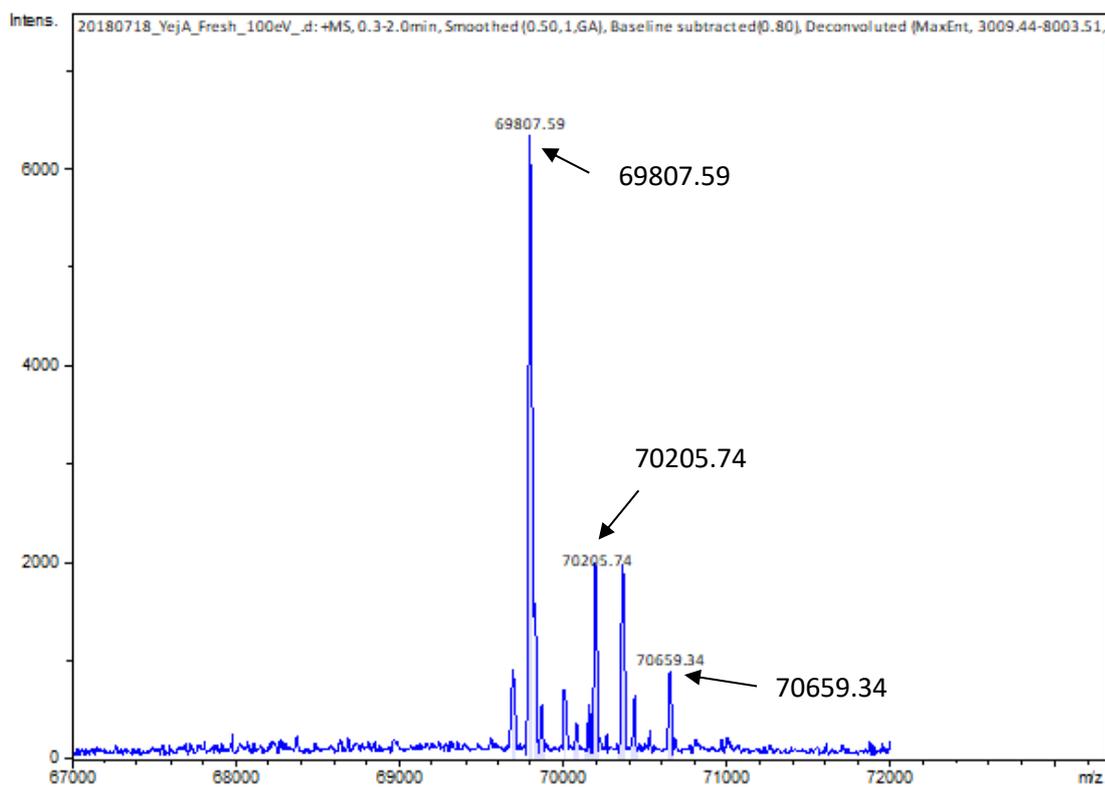
Figure 5.9. Native ESI-MS at 100 eV shows EcYejA in complex with several ligands.

Little to none of the un-complexed form of EcYejA can be seen in the spectra. The peak at 69,807.59 Da is the dominant peak in the spectra and corresponds to EcYejA complexed with VLGEPRYAFNFN. Other peaks with lower intensities can also be seen in the spectra, they correspond to EcYejA in complex with the flowing ligands: VLGEPRYAFNFNHFD; VLGEPRYAFNFNHFDY; AVLGEPRYAFNFNHFDY and AFAVLGEPRYAFNFNHFDY.

Therefore, EcYejA may not be coming into contact with its natural ligand as EcYejA has little interaction with the periplasmic fraction during production and purification. Exposing EcYejA to periplasmic extract from *E. coli* BW25113 might enable EcYejA to bind its natural ligand. *E. coli* BW25113 was chosen as the strain for the periplasmic preparation as it is a wild type strain with no gene deletions. *E. coli* BL21(DE3) cells, which EcYejA is usually produced in, contains gene deletions for the Lon and OmpT proteases. OmpT is a protease which is located in the outer membrane of bacteria and therefore influences the peptide makeup of the periplasm, which could affect the natural ligand of EcYejA. By exposing EcYejA to a periplasmic prep in this way it is hoped that any other EcYejA ligands will be picked up. Again, MALDI-MS/MS identified a number of different peptides as potential EcYejA ligands, all of which were from the N-terminus of EcYejA (Table 5.3).

At 200 eV in this sample a large peak can be seen which corresponds to the ligand free EcYejA (Figure 5.10). The only other peak seen in the spectra at this ion cone voltage is 1,426 Da greater than the mass of the free EcYejA, again this indicates that VLGEPRYAFNFN is bound at this higher ion cone voltage.

At the lower ion cone voltage of 100 eV, again very little of the ligand free EcYejA is observed (Figure 5.11). Once again the dominant ligand bound form of EcYejA is in complex with VLGEPRYAFNFN. The other two ligand bound states of EcYejA indicate binding of peptides that are slightly shorter than those seen in the fresh sample of EcYejA: VLGEPRYAFNF and FAVLGEPRYAFN.

A selection of ligands identified as being in complex with EcYejA, and GEP variants, were tested for binding via thermal shift assays to gain more information on the specificity of EcYejA (Table 5.4). The data shows a general trend of an increase in the melting temperature of EcYejA as the length of the ligand increases. The longest ligand shown to bind EcYejA via thermal shift assays is the 15 residue peptide VLGEPRYAFNFNHFD with an increase in melting temperature of 26.4 °C. The thermal shift data indicates that LGEPR and YAFNFN do not bind EcYejA, this implies that the necessary components for GEP (LGEPRYAFNFN) binding are contained across the middle of the peptide, not at either end.

These results clearly identify the N-terminus of EcYejA as the only source of ligands for EcYejA prepared in this way. All peptides shown to bind EcYejA contained a core motif of EPRYAFN, demonstrating some specificity for EcYejA. When exposed to a periplasmic sample from *E. coli* BW25113, EcYejA did not bind any ligands other than those from the N-

Ligand	Calculated mass (Da)
VLGEPRYAFN	1164.5927
VLGEPRYAFNF	1311.6612
FAVLGEPRYAFN	1382.6983
VLGEPRYAFNFN	1425.7041
VLGEPRYAFNFNH	1562.763
AFAVLGEPRYAFNF	1600.8038
AVLGEPRYAFNFNH	1633.8001

Table 5.3. In the periplasm mixed EcYejA sample MALDI-MS/MS identified several peptides from the N-terminus of EcYejA as potential EcYejA ligands.

This data was generated from a sample EcYejA that was exposed overnight to a periplasmic prep from *E. coli* BW25113.
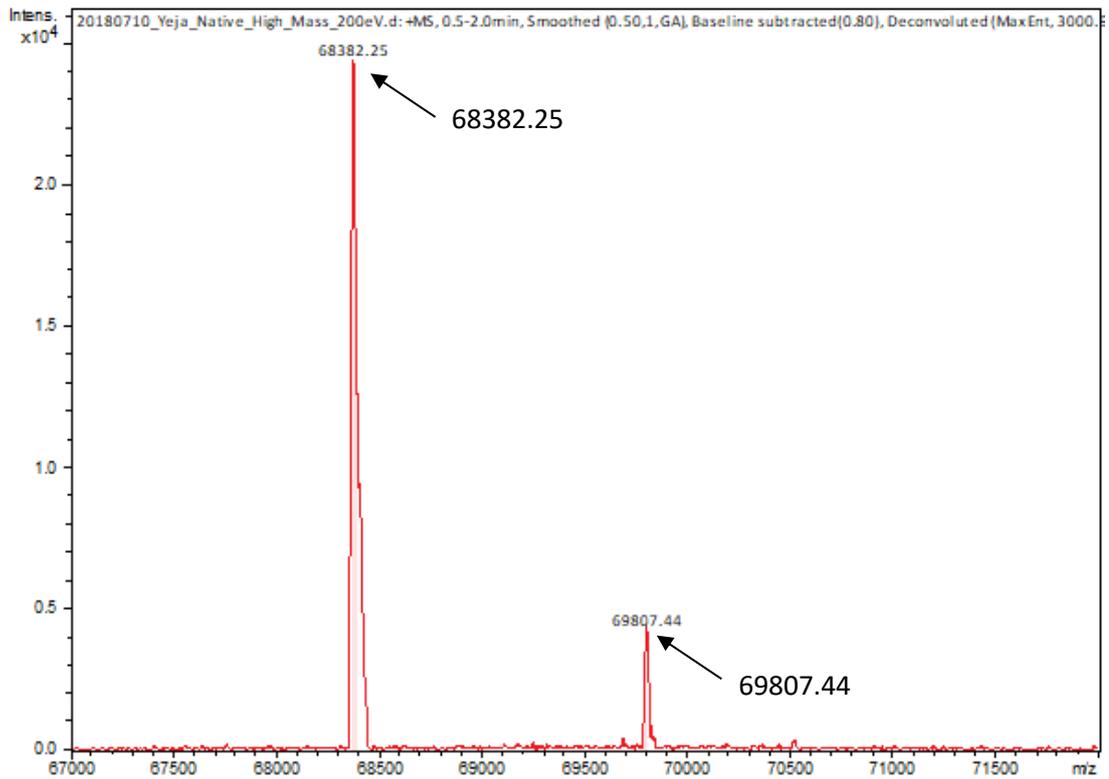
Figure 5.10. Native ESI-MS at 200 eV shows EcYejA exposed to periplasm in complex with a dominant ligand.

A peak at 68,382.25 Da can be seen which corresponds to the un-complexed form of EcYejA, this is the dominant peak. The only other peak that can be seen is a peak at 69,807.44 Da which corresponds to EcYejA in complex with VLGEPRYAFNFN.
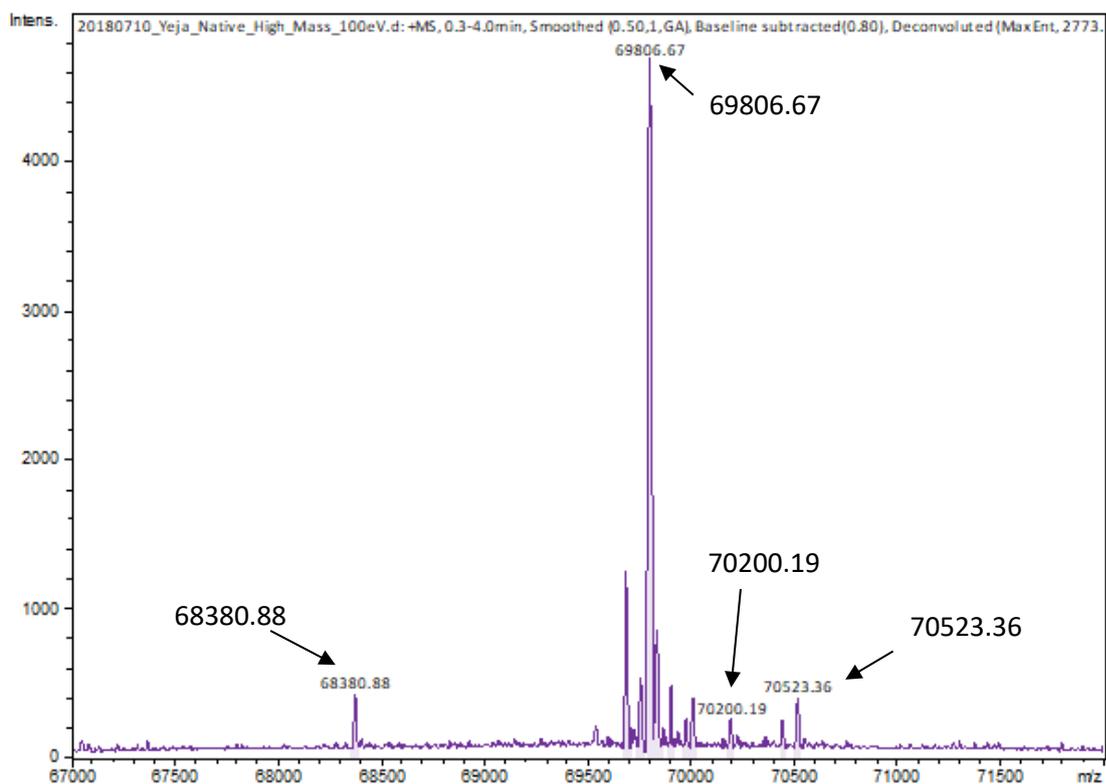
Figure 5.11. Native ESI-MS at 100 eV shows EcYejA exposed to periplasm in complex with several ligands.

A small peak can be seen at 68,380.88 Da, the un-complexed version of EcYejA. Again the dominant peak at 69,806.67 Da corresponds to EcYejA in complex with VLGEPRYAFNFN. Other peaks for different EcYejA complexes can be seen, these ligands are VLGEPRYAFNF and FAVLGEPRYAFN.

Ligand	Ligand Concentration	Melting Temperature Change (°C)
LGEPR	5 mg/ml (8.76 mM)	0
YAFNFN	2.5 mg/ml (3.23 mM)	+2.7
EPRYAFNFN	2.5 mg/ml (2.16 mM)	+7.9
GEPRYAFNFN	2.5 mg/ml (2.06 mM)	+14.4
LGEPRWAFNFN	2.5 mg/ml (1.85 mM)	+22
LGEPAYAFNFN	2.5 mg/ml (2.01 mM)	+6.9
LGEPRYAFN	1.25 mg/ml (1.17 mM)	+17.3
LGEPRYAFNF	2.5 mg/ml (2.06 mM)	+19.9
VLGEPRYAFNF	2.5 mg/ml (1.9 mM)	+22.1
VLGEPRYAFNFN	2.5 mg/ml (1.75 mM)	+20.7
FAVLGEPRYAFN	5 mg/ml (3.61 mM)	+19.1
VLGEPRYAFNFNHFD	2.5 mg/ml (1.37 mM)	+26.4

Table 5.4. GEP variants tested via thermal shift assays.

The EcYejA concentration in all experiments was 0.5 mg/ml (7.3 µM). A thermal shift in melting temperature of anything equal to or more than +3 °C was deemed as binding to EcYejA. This value was chosen arbitrarily as variations in melting temperature of up to ±1/2 °C were seen across different thermal shift runs.

terminus of EcYejA. This implies that there is a biological role for the binding of these N-terminal peptides by EcYejA, although further investigation is needed before that role can be elucidated. Longer peptides are observed in the freshly prepared sample of EcYejA as opposed to the EcYejA sample that was exposed to the periplasmic extract. This may be due to the longer storage time of the periplasmic extract sample of EcYejA, which may have resulted in the break-down of some of the peptide ligands to shorter peptides.

5.3 *In vivo* CAMP resistance assays

Previous literature implicated both the *S.* Typhimurium and the *B. melitensis* Yej transporters in resistance to CAMPs, including polymyxin B, melittin and LL-37, via genetic studies (Eswarappa *et al.*, 2008; Wang *et al.*, 2016). The binding experiments detailed in this chapter give no indication that the *E. coli* Yej transporter can bind, transport or provide resistance to CAMPs. Therefore, genetic studies were carried out to determine whether the Yej transporter phenotypes seen previously could be replicated in *E. coli* in the presence of CAMPs.

E. coli BW25113 is the parent strain for the Keio collection, a library of knockouts of the non-essential genes in *E. coli* (Baba *et al.*, 2006). The *E. coli* Δ*yejA,* Δ*yejB,* Δ*yejE,* Δ*yejF* and Δ*tolC* mutants referred to in this section were taken from the Keio collection and *E. coli* BW25113 was used as the wild type in experiments.

Initially disc diffusion assays were carried out with Yej transporter knockout mutants to test susceptibility of the different mutants to the CAMP polymyxin B. However, this method was not practical to continue with due to the zones of inhibition being very similar in size regardless of the concentration of polymyxin B. This indicated that polymyxin B was not diffusing through the agar plate particularly well and therefore the technique was not suitable for testing susceptibility of different mutants.

Following the unsuccessful disc diffusion assays, liquid culture was tried to test the susceptibility of the different Yej mutants, this was carried out using a plate reader. A 96 well plate was filled with LB containing different concentrations of polymyxin B and the antibiotic colistin and the wells were inoculated with *E. coli* BW25113. The plates were incubated at 37 °C for 48 hours with constant shaking and an OD_{600} reading was taken every 30 mins. Once a suitable concentration for polymyxin B and colistin were identified, roughly decreases in OD_{600} of 40-50% in stationary phase, these concentrations were tested on the Keio collection knockout mutants.

Using the same setup in the plate reader polymyxin B was tested at 10 and 15 µg/ml and colistin was used at 40 and 45 µg/ml with either *E. coli* BW25113, *E. coli* Δ*yejA* or *E. coli* Δ*tolC*. TolC is an outer membrane efflux protein that is a major factor in the protection against antibiotics that damage membranes (Zgurskaya *et al.*, 2011) and therefore made it a good control in these experiments.

The *E. coli* Δ*yejA* and *E. coli* Δ*tolC* mutants grown in LB without any polymyxin B or colistin grew in a very similar manner to the wild type, *E. coli* BW25113 (Figure 5.12). This indicates that the mutations are not having any effect on the growth of the *E. coli* in these conditions. However, at 10 and 15 µg/ml polymyxin B and 40 and 45 µg/ml colistin growth of the *E. coli* Δ*yejA* and *E. coli* Δ*tolC* mutants was completely abolished, a very strong phenotype.

An unusual phenotype displayed in the plate reader graphs is the increased lag phase in the growth of the *E. coli* BW25113 wild type to 10-12hrs after inoculation (Figure 5.12). This was thought to be the time required to upregulate the expression of the Yej transporter.

In an attempt to replicate the plate reader results susceptibility tests were carried out with 10 µg/ml polymyxin B in 100 ml shake flasks. Shake flasks were incubated at 37 °C for 48 hrs with shaking at 150 rpm, OD_{600} readings were taken at the stated times.

Interestingly the lag in the growth phase of the *E. coli* BW25113 wild type could not be replicated in the shake flasks (Figure 5.13). However, a new phenotype was seen where there is a small lag in the growth of the *E. coli* Δ*yejA* mutant on LB alone with no added polymyxin B. Although this result could not be repeated in subsequent experiments (Figure 5.14).

The difference seen between the plate reader and shake flask assays could be explained by the different amounts of aeration in each technique. The shake flask assays naturally aerate the liquid to a greater extent than the plate reader does, it is possible that this lower aeration in the plate reader induces a stress response allowing the *E. coli* BW25113 wild type strain to overcome the negative effects of the polymyxin B.

The plates used in the plate reader experiments are coated to give them a slightly negative charge, as CAMPs are cationic their function could be impaired by the negatively charged plates. This could be another reason that the *E. coli* BW25113 wild type could withstand the polymyxin B in the plate reader but not in the shake flasks.
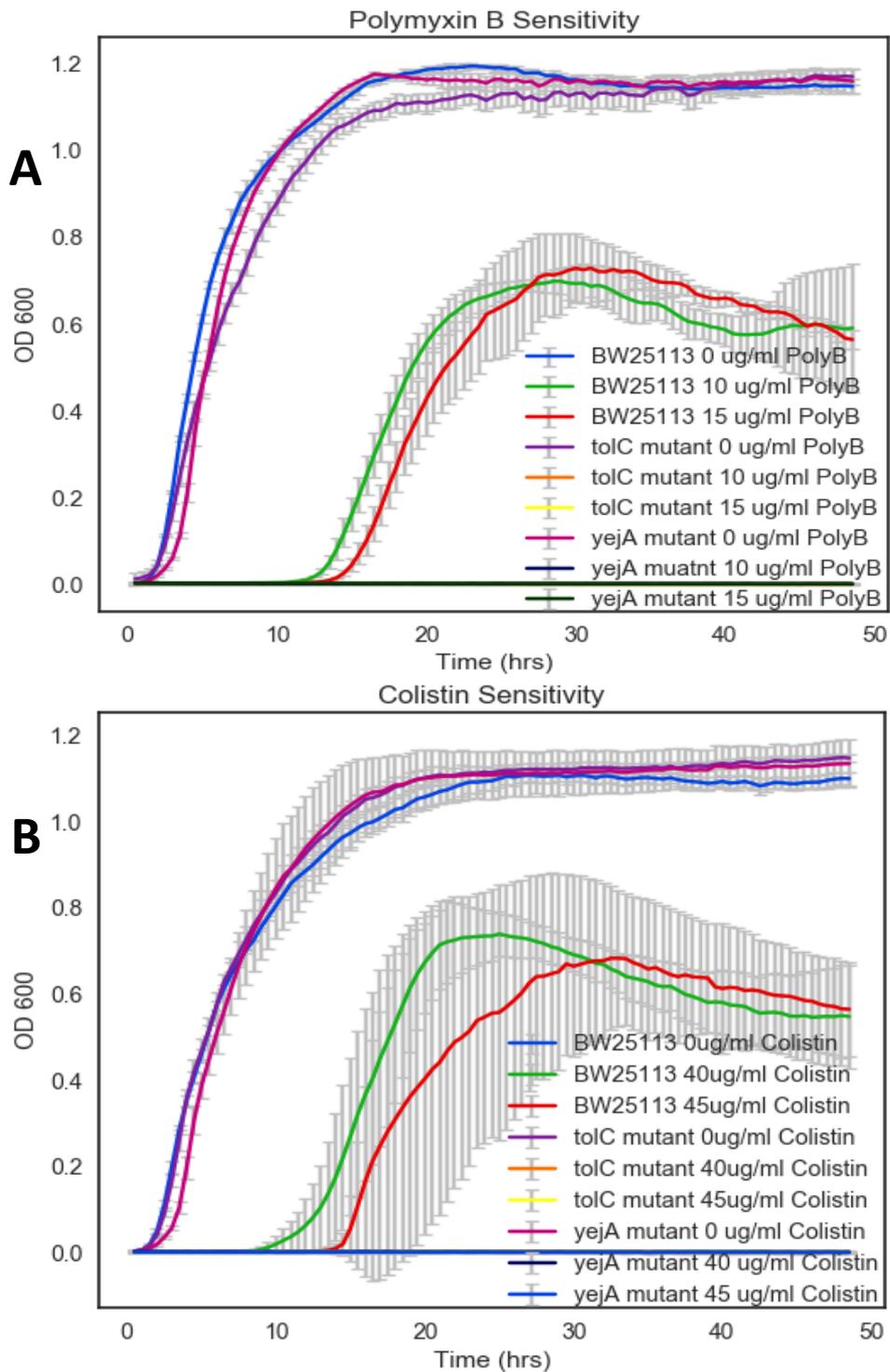
Figure 5.12. Growth of Δ*yejA* and Δ*tolC* mutants is completely abolished in the presence of polymyxin B or colistin.

(A) shows cell growth with 0, 10 or 15 μg/ml polymyxin B, (B) shows cell growth with 0, 40 or 45 μg/ml colistin. BW25113 is the wild-type control strain, there is a lag in growth of 10-12 hrs in the presence of both polymyxin B and colistin. Δ*yejA* and Δ*tolC* mutants are unable to grow at all in polymyxin B or colistin.

176
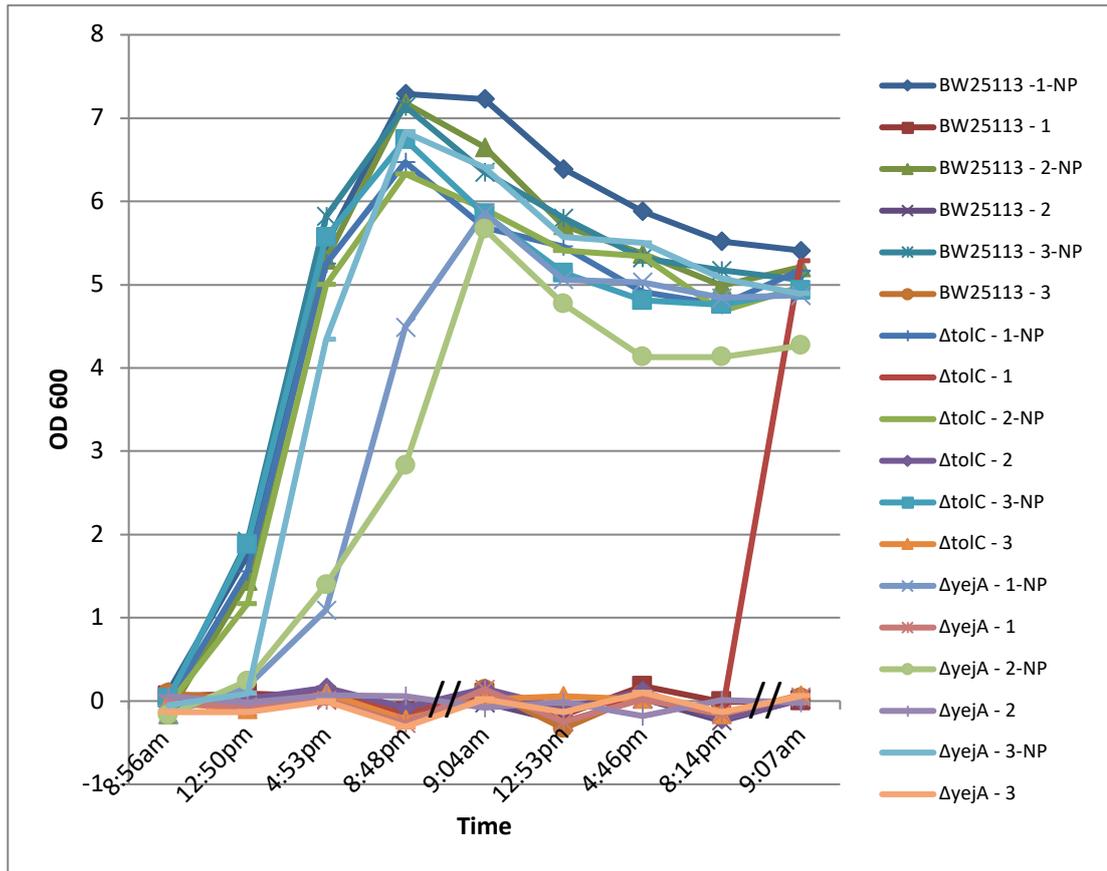
Figure 5.13. 10 μg/ml polymyxin B abolishes growth of *E. coli* BW25113, *E. coli* Δ*yejA* and *E. coli* Δ*tolC*.

1, 2, 3 = repeats; NP = No polymyxin B. BW25113-1 was thought to be contaminated at 8:14pm. There was no cell growth in any of the flasks which contained 10 μg/ml polymyxin B. There was a 4 hour lag in growth of the *E. coli* Δ*yejA* mutants without polymyxin B.
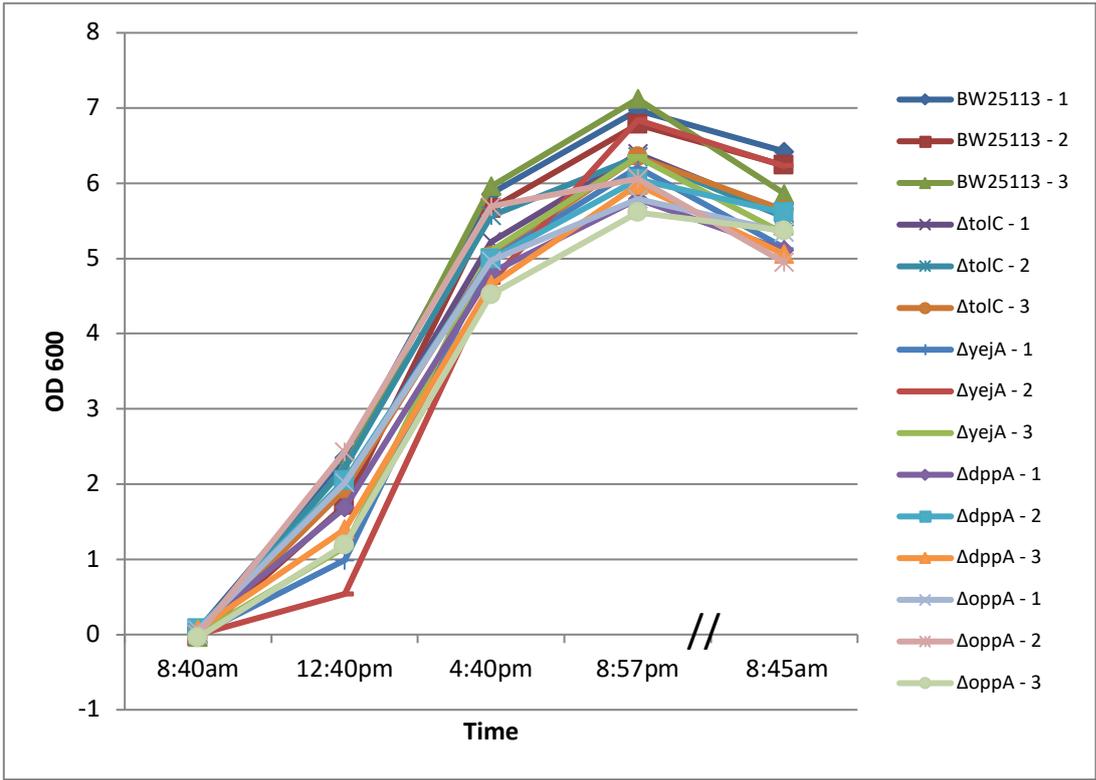
Figure 5.14. *E. coli ΔyejA* **phenotype seen in plate reader assays could not be replicated in shake flasks.**

1, 2, 3 = repeats. *E. coli* BW25113, *E. coli ΔyejA*, *E. coli ΔdppA*, *E. coli ΔoppA* and *E. coli ΔtolC* were grown for 24 hours at 37 °C.

Chapter Six

Discussion and

Future Work

6. Discussion and Future Work

CAMPs are important components of the innate immune system, the first line of defence against invading pathogens, and are conserved throughout biology (Pasupuleti, Schmidtchen and Malmsten, 2012). CAMPs are able to disrupt the integrity of bacterial cells by forming pores in the membrane. This causes the bacterial cell to leak and the bacteria to die, however bacteria have developed a number of resistance mechanisms to these CAMPs, including the use of the ABC transporters Sap and Yej (Parra-Lopez, Baer and Groisman, 1993; Eswarappa *et al.*, 2008; Wimley, 2010; Shelton *et al.*, 2011; Wang *et al.*, 2016). ABC transporters which import substrates contain SBPs, in the case of the Sap and Yej transporters, these are SapA and YejA respectively. The SBPs define the specificity of the transporter as they capture the substrates in the extracellular environment and make them available to the transmembrane components for transport (Locher, 2009). SapA and YejA are members of the Cluster C SBPs, these SBPs are structurally distinct from others as they contain an extra domain, giving rise to larger binding pockets which can accommodate larger substrates (Berntsson *et al.*, 2010). Genetic and/or biochemical data have shown that Sap and Yej are important in the defence of bacteria against CAMPs. It is believed that SapA and YejA are able to recognise and bind CAMPs, transporting them to the cytoplasm before they have the chance to integrate into the bacterial cell membrane (Groisman *et al.*, 1992; Parra-Lopez, Baer and Groisman, 1993; Mason *et al.*, 2006; Eswarappa *et al.*, 2008; Shelton *et al.*, 2011; Wang *et al.*, 2016). However, CAMPs are large (20-30 residue) peptides that often contain secondary structure. How then do SapA and YejA accommodate these large and potentially structured CAMPs?

We aimed to determine whether SapA and YejA bind CAMPs and if so how. A number of different strategies were employed to express soluble SapA at high enough concentrations to carry out biochemical and structural work, unfortunately this was not possible. YejA from *E. coli* on the other hand was cloned, expressed and purified successfully and was therefore taken forward in this study and its specificity investigated.

6.1 SapA is an insoluble protein

Several different constructs were created with the aim of producing high yields of soluble SapA. An *E. coli* SapA construct, which targeted production of SapA to the cytoplasm of the cell, was trialled with no success. Due to the predicted disulphide bonds in the SapA protein the next constructs contained a vector derived PelB leader sequence which targeted SapA to the periplasm where disulphide bond formation occurs. SapA from different species, *S.*

Typhimurium LT2 and *H. influenzae*, were also included to broaden the scope of the trials. Again, there was no success. Finally, an *E. coli* SapA construct was produced which contained the native leader sequence, again targeting the production of SapA to the periplasm with no soluble protein being produced.

As well as creating different constructs different growth and protein production conditions were tested. Many different IPTG concentrations and growth temperatures were trialled, but little to no soluble SapA was produced. Lowering the IPTG concentration can increase the production of soluble protein in the cell by reducing the burden on the cell and therefore helps to ensure the protein being produced is properly processed (Donovan, Robinson and Click, 1996). Reducing the growth temperature of the cells can slow the rate of protein production and increase protein solubility, again by ensuring the correct processing of the protein (Schein and Noteborn, 1988).

Although SapA was not taken forwards in this work, phylogenetic and bioinformatic analysis was carried out on the protein. The phylogenetic tree (Figure 3.1) shows SapA and DppA proteins positioned close together, indicating similarity between the two sets of proteins. SapA was also predicted to contain two disulphide bonds (Figure 3.2, 3.3), which are conserved in DppA. From sequence alignments (Figure 3.2) an aspartate residue is predicted to "cap" the binding site in *E. coli* SapA, this residue limits the length of the peptides bound by the SBPs which contain the capping aspartate and is a characteristic feature of SBPs binding shorter peptides. This analysis strongly suggested to us that SapA was a dipeptide binding protein, similar to DppA, and was therefore incapable of binding full length CAMPs.

The difficulties of producing soluble SapA could be caused by protein misfolding. Misfolded proteins can aggregate into larger structures which are often insoluble, this type of aggregation can be caused by environmental stress, chemical modifications and destabilising mutations (Vendruscolo, 2012). Newly synthesised proteins, such as SapA upon addition of IPTG to the cells, are particularly vulnerable to misfolding and aggregation (Winkler *et al.*, 2010). Aggregation can be caused by exposed hydrophobic regions of misfolded proteins interacting with hydrophobic regions of other misfolded proteins which leads to aberrant protein-protein interactions (Vabulas *et al.*, 2010). Recombinant SapA was produced in the cytoplasm a location in which the formation of disulphide bonds is compromised. As disulphide bonds are often important for structure it is possible that producing SapA in the cytoplasm led to exposed hydrophobic regions, contributing to

protein misfolding, aggregation and insolubility. The misfolding could also have resulted from the large quantity and high rate of protein production as aggregation is a concentration dependent process (Vabulas *et al.*, 2010).

Molecular chaperones interact with most proteins as they are being translated and prevent premature folding/misfolding and aggregation (Vabulas *et al.*, 2010). This is important for SapA with either a PelB leader sequence or native leader sequence as it passes through the Sec translocase to enter the periplasm. The Sec translocase only allows the passage of unfolded proteins, therefore SapA must be unfolded to pass through. SecB, the chaperone of the Sec translocase, maintains proteins in their unfolded state (du Plessis, Nouwen and Driessen, 2011). However, the rate of SapA production is vastly increased by the addition of IPTG to the cells which could overwhelm SecB as SecB production has not been increased accordingly with SapA production. This could lead to an accumulation of misfolded SapA in the cytoplasm as it cannot pass through the Sec translocase. Overproduction of SapA therefore overloads the capacity of the cell to support protein folding to the native structure.

If the work on SapA had continued a number of other techniques would have been trialled in the quest for soluble protein. Fusion to maltose binding protein, a solubility tag, would have been used to try and increase the quantity of soluble SapA. Maltose binding protein is able to promote the solubility of aggregation-prone proteins and so would be a good option to try with SapA (Pryor and Leiting, 1997). There are other solubility tags which could have been tested in addition to the maltose binding protein, for example the glutathione S-transferase (GST) tag and the SUMO tag (Marblestone *et al.*, 2006).

6.2 EcYejA has a large binding pocket and binds peptides with an EPRYAFN motif

EcYejA was crystallised and its structure determined by X-ray diffraction methods. This led to the discovery of a ligand bound in the binding site of EcYejA. This was unexpected as during the purification process EcYejA had been treated with 2 M guanidine hydrochloride which is commonly used to remove any pre-bound ligands (Lanfermeijer *et al.*, 1999). This suggested that the ligand was bound post-purification. The ligand was discovered to be LGEPRYAFNFN, and therefore derived from the N-terminal region of EcYejA itself. From the crystal structure of EcYejA it was clear to see that this N-terminal region is a loop packed

onto the surface of EcYejA and is not present in other Cluster C SBPs, such as BsAppA, StOppA and EcMppA, when superposed onto each other.

The interactions between EcYejA and LGEPRYAFNFN (GEP) feature key side chain residues, mainly Arg5 and Asn9 of GEP. These residues form a number of hydrogen bonds with residues of EcYejA that serve to anchor GEP in the binding pocket (Figure 4.15A, 4.16B). Other side chains of GEP appear not to contribute significantly to the binding of GEP to EcYejA, it is backbone interactions that contribute most to binding. An emphasis on backbone interactions is not unusual with Cluster C peptide transporter SBPs as this allows the SBP to accommodate peptides of varying sequence but usually of a defined length. This strategy allows the cell to produce a single transporter to transport all the dipeptides encountered in the periplasm rather than producing a transporter for each specific dipeptide available, drastically reducing the work load of the cell. However, some Cluster C peptide transporter SBPs are highly specific, for example the transport of murein tripeptide by MppA. In this case hydrogen bonds and salt bridges are formed between side chains of the murein tripeptide and MppA, helping to make the interaction highly specific (Maqbool *et al.*, 2011). To reduce the work load of the cell in this case MppA is an orphan SBP, meaning it is not found in an operon with a cognate transporter and so uses the OppBCDF transporter to transport murein tripeptide (Park *et al.*, 1998). EcYejA contains components of both of these systems, specific side chain interactions, some backbone interactions and its own cognate transporter.

LlOppA has the ability to bind a large range of peptide chain lengths (4-35 residues) but prefers peptides containing branched chain amino acids and at least one isoleucine. In the structure hydrogen bonds are mainly formed between LlOppA and the backbone of the bound peptides demonstrating that LlOppA has limited preference for sequence order, components and length. However, the isoleucine side chain sits in a hydrophobic pocket and is well defined in the electron density, indicating that it anchors the peptides to LlOppA and that specific position of isoleucine is highly important, the peptide chain can extended either way and for however long as long as the isoleucine of the peptide is anchored in the hydrophobic pocket (Berntsson *et al.*, 2009). It is possible that Arg5 and Asn9 of the GEP peptide are anchoring it to EcYejA in a similar mechanism to that of LlOppA, therefore as long as those residues are present in the peptide at the correct spacing, EcYejA will bind the peptide.

EcYejA is very similar to LlOppA structure wise as it contains a large internal cavity, lacks the capping aspartate residue and substrates are possibly anchored by key side chain residues. However, specificity wise, EcYejA is more similar to MppA as it binds one specific peptide but can tolerate slight variations on the ends of that peptide.

To determine other ligands of EcYejA, thermal shift assays were used where several different potential ligands were tested including; CAMPs, peptides of varying length, Microcin C like peptides and LGEPRYAFNFN based peptides. Only LGEPRYAFNFN was able to produce a large increase in the melting temperature of EcYejA that was indicative of binding.

CAMPs have many different structures and sequences, it is possible that EcYejA does not have the ability to bind and recognise all of the different CAMPs (Pasupuleti, Schmidtchen and Malmsten, 2012). However, every effort was made to carefully select CAMPs that had been previously linked with YejA and resistance, for example Polymyxin B, Melittin and LL-37 (Eswarappa *et al.*, 2008; Wang *et al.*, 2016). These were all tested for binding in thermal shift assays and none were shown to bind EcYejA. This indicates that EcYejA probably does not bind full length CAMPs. Although in this study no biochemical evidence has been found for EcYejA binding CAMPs, the genetic studies previously carried out clearly show the Yej transporter as a key component in resistance to CAMPs (Eswarappa *et al.*, 2008; Wang *et al.*, 2016). The work carried out here and the data from previous studies therefore indicate that the Yej transporter is important in resistance to CAMPs via a mechanism that does not involve the binding of full length CAMPs by YejA.

It is unsurprising that the heptapeptide (MRTGNAD) analogues of Microcin C did not bind EcYejA in thermal shift assays. There has been work trying to identify the determinants of the specificity of YejA for Microcin C. It was found that Microcin C analogues must have a minimum peptide chain length of 6 residues together with an N-terminal formyl-methionyl-arginyl sequence for transport to take place (Gaston *et al.*, 2011). As a result, it was unlikely that any of the Microcin C analogues tested in this work would show binding to EcYejA, but they were tested to add breadth to the ligands trialled.

Native ESI mass spectrometry and MALDI MS/MS was used to show binding of LGEPRYAFNFN to EcYejA. Related ligands from the N-terminus of EcYejA were also shown to bind to EcYejA, all of which contained a core motif of VLGEPRYAFN, this indicated some specificity of EcYejA. This was then probed further with more thermal shift assays which

determined the core motif to actually be EPRYAFN. The thermal shift assays also indicated, in general, that as the length of the ligand increased the shift in melting temperature increased accordingly. The largest peptide shown to bind EcYejA in this work was 15 residues.

It is not surprising that Arg5 of GEP is present in the conserved motif in view of its large contribution, 5 side chain hydrogen bonds, to the anchoring of GEP to EcYejA. Asn9 was also conserved throughout the ligands identified via mass spectrometry, again from the structure of EcYejA it is evidently a key residue in anchoring GEP to EcYejA with 3 side chain hydrogen bonds to the protein. It is possible that if shorter peptides had been present in the mass spectrometry solution containing Arg and Asn at the correct spacing, they would have bound. For example, the peptide RYAFN or RXXXN might have bound as they both contain Arg and Asn in the correct register. Interestingly, LGEPR and YAFNFN were shown to not bind EcYejA via thermal shift assays. Neither of these peptides contains both the Arg or Asn residues, lending more weight to the hypothesis that Arg and Asn are essential for EcYejA binding.

From the mass spectrometry data it is clear that there are a range of EcYejA derived peptides in the solution. However, to produce any of these peptides EcYejA must have been cut at least twice (Figure 6.1). Figure 6.1 shows all of the different cut sites on EcYejA to produce all of the different peptides identified via native ESI-MS as being in complex with EcYejA.

A known periplasmic peptidase, DegP, has been found to cleave between paired hydrophobic residues which could account for some of the cleavage sites indicated in Figure 6.1 (Jones *et al.*, 2002). However, not much is known about the specificity of other periplasmic proteases and whether they would cleave YejA to produce the seen peptides.

The transmembrane glycoprotein MUC1 undergoes self-cleavage in its extracellular domain to generate two subunits that specifically recognise and bind each other in a strong noncovalent interaction (Levitin *et al.*, 2005). It is possible that YejA does something similar and self-cleaves in its N-terminal region to produce EPRYAFN containing peptides which it then goes on to bind.

Inteins are another example of self-cleaving proteins, they excise themselves from a native host protein by self-cleaving and ligating the flanking peptide bonds

MQAIKESY|A|F|A|VLGEPRYAFN|F|N|HFD|Y|VNP

Peptides shown to be in complex with EcYejA
VLGEPRYAFNF
VLGEPRYAFNFN
FAVLGEPRYAFN
VLGEPRYAFNFNHFD
VLGEPRYAFNFNHFDY
AVLGEPRYAFNFNHFDY
AFAVLGEPRYAFNFNHFDY

Figure 6.1. Cleavage sites on EcYejA to produce VLGEPRYAFN containing peptides.

The sequence of the N-terminal region of EcYejA can be seen at the top of the figure, red lines indicate the cleavage sites required to produce all peptides found in complex with EcYejA. At the bottom of the figure is a table detailing all of the VLGEPRYAFN containing peptides found in complex with EcYejA. All peptides in the table have been aligned via the VLGEPRYAFN sequence.

(Warren, Coolbaugh and Wood, 2013). It is possible that the EPRYAFN containing peptides are actually excising themselves from YejA, rather than a peptidase or YejA self-cleaving to remove them.

It would be interesting to obtain accurate K_D information for EcYejA and the different ligands identified via the mass spectrometry work, this would help further pinpoint the specificity of EcYejA. For example, it could help determine whether both the anchoring Arg and Asn contribute equally to the binding of ligands, whether there is an optimal peptide length and whether there is an optimal sequence.

Whilst other ABC peptide importers are made up of 5 protein components, the Yej transporter is only made up of 4, with the YejF protein being a fused homodimeric NBD. This indicates that the role of the Yej transporter might be atypical, with the YejF component carrying out an additional perhaps unusual function such as signalling or protection. The ABC-F proteins contain two NBDs on a single polypeptide chain, much like YejF, but are not always associated with a transporter nor are they in operons encoding transporters. However, some ABC-F proteins do associate with TMDs to transport substrates, but it is clear that this is not the role of all ABC-F proteins. ABC-F proteins have been shown to drive dissociation of antibiotics bound to the ribosome to rescue translation in the cell via a mechanism of protein-mediated drug displacement (Sharkey, Edwards and O'Neill, 2016). It is possible that the YejF is able to dissociate from the transporter and carry out some similar role within the bacterial cell, possibly triggering resistance to CAMPs after they have been sensed in the periplasm.

6.3 New model for stress-based sensing in *E. coli*

From the data collected during this body of work we hypothesise that the transport of Microcin C by YejA is a subsidiary function of the transporter and that the true substrate of YejA is peptides containing the EPRYAFN motif. We propose that CAMPs enter the periplasm where they activate peptidases as a stress response to the presence of the CAMPs (Figure 6.2). These peptidases cleave the N-terminus of YejA to produce peptides with the core EPRYAFN motif which are bound by YejA and transported to the cytoplasm through the Yej transporter. The transport of these peptides activates YejF which then dissociates and possibly activates an LPS modifying enzyme, or other CAMP resistance system, which combats CAMP attack.
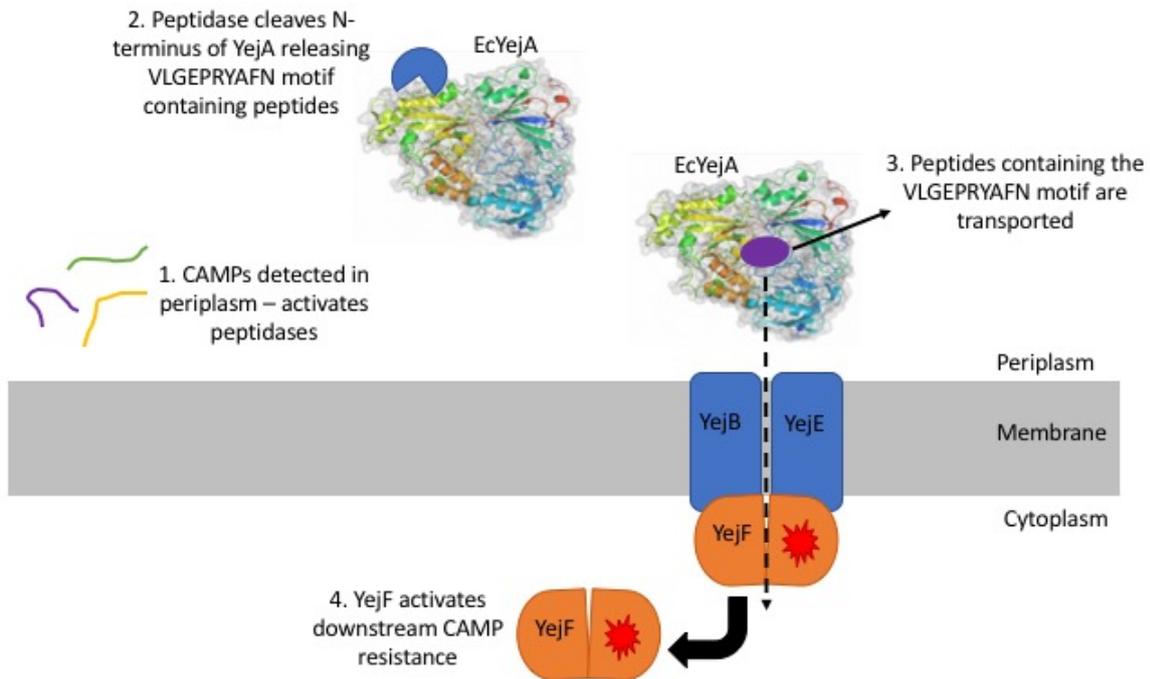
Figure 6.2. Hypothetical mechanism of action for the Yej transporter.

CAMPs are detected in the periplasm which activates peptidases. The peptidases then cleave the N-terminus of YejA releasing EPRYAFN containing peptides which are bound by YejA. Transport of these peptides activates YejF which then dissociates from the membrane components and activates downstream CAMP resistance mechanisms such as LPS modification.

When harvested the cells expressing EcYejA would have been in the stationary phase of growth. Stationary phase cells commonly induce a cellular stress response due to the exhaustion of the external environment and the accumulation of toxic side products from catabolism (Pletnev *et al.*, 2015). This stress response could have activated peptidases and could explain the presence of the EPRYAFN containing peptides in the binding site of EcYejA without a CAMP stimulus. A way of testing this hypothesis would be to harvest the cells in the log phase of growth and carry out native ESI-MS and MALDI-MS/MS to identify any bound ligands.

Pull down assays and tandem affinity purification could be used to investigate the interactions YejF makes with other proteins in the cell. These techniques work by fusing a tag to the protein of interest, in this case YejF, binding that protein to a column and then flowing cellular extract over the column. Proteins which interact with the protein of interest elute with the protein of interest when it is removed from the column (Puig *et al.*, 2001). Mass spectrometry could then be carried out on the eluted proteins to identify them.

Activation of YejF could be caused by binding of ATP or ATP hydrolysis as it is a nucleotide binding domain and already contains two ATP binding sites which bind and hydrolyse ATP as the substrate is transported (Locher, 2009). ATP therefore seems like an obvious choice of activation molecule. To test this hypothesis non-hydrolysable ATP analogues could be used and downstream effects monitored.

6.4 Conclusion

The initial aims of this project were to characterise proteins which are thought to bind CAMPs with a view to developing new drugs to help tackle antibiotic resistance. Although it was not possible to show any involvement of either SapA or YejA with CAMPs in this work, solving the structure of YejA in a closed conformation and further binding experimentation with this protein has provided useful information which could be used in drug design.

A new peptide drug could be designed based on the knowledge gained from the structure of YejA, possibly containing the EPRYAFN motif, which binds very tightly to YejA and is not transported or does not activate YejF. This peptide drug would then bind the majority of YejA, preventing YejA from binding EPRYAFN containing peptides which are transported and do activate YejF. By preventing the stress-based signalling system functioning correctly

CAMP resistance mechanisms would not be activated and would therefore leave naturally produced CAMPs available to carry out their bactericidal functions.

Abbreviations

ABC – ATP Binding Cassette

CAMP – Cationic Antimicrobial Peptide

CD – Circular Dichroism

CV – Column Volume

DNA – Deoxyribonucleic Acid

EcYejA – *E. coli* YejA

HRV – Human rhinovirus

IPTG – Isopropyl β-D-1-thiogalactopyranoside

LB – Luria-Bertani broth

MALDI-MS/MS – matrix-assisted laser desorption/ionisation mass spectrometry mass spectrometry

Native ESI – Native electrospray ionisation

NBD – Nucleotide Binding Domain

PAGE – Polyacrylamide Gel Electrophoresis

PCR – Polymerase Chain Reaction

SBP – Substrate Binding Protein

SEC – Size exclusion chromatography

SEC-MALLS – Size Exclusion Chromatography Multi-Angle Laser Light Scattering

SDS – Sodium Dodecyl Sulphate

TMD – Transmembrane Domain

Bibliography

Alexander, H. E., Hahn, E. and Leidy, G. (1956) 'On the specificity of the desoxyribonucleic acid which induces streptomycin resistance in Hemophilus', *J Exp Med*, 104(3), pp. 305–320.

Andersson, D. I., Hughes, D. and Kubicek-Sutherland, J. Z. (2016) 'Mechanisms and consequences of bacterial resistance to antimicrobial peptides', *Drug Resistance Updates*. Elsevier Ltd, 26, pp. 43–57. doi: 10.1016/j.drup.2016.04.002.

Antal, M. *et al.* (2005) 'A small bacterial RNA regulates a putative ABC transporter', *Journal of Biological Chemistry*, 280(9), pp. 7901–7908. doi: 10.1074/jbc.M413071200.

Anunthawan, T. *et al.* (2015) 'Biochimica et Biophysica Acta Cationic amphipathic peptides KT2 and RT2 are taken up into bacterial cells and kill planktonic and biofilm bacteria', *BBA - Biomembranes*. Elsevier B.V., 1848(6), pp. 1352–1358. doi: 10.1016/j.bbamem.2015.02.021.

Baba, T. *et al.* (2006) 'Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection', *Molecular Systems Biology*, 2, p. 2006.0008-2006.0008.

Balhorn, R. (2007) 'The protamine family of sperm nuclear proteins', *Genome Biology*. BioMed Central, 8(9), p. 227. doi: 10.1186/gb-2007-8-9-227.

ter Beek, J., Guskov, A. and Slotboom, D. J. (2014) 'Structural diversity of ABC transporters', *The Journal of General Physiology*, 143(4), p. 419 LP-435.

Bengoechea, J. A. and Skurnik, M. (2000) 'antiporter system mediates resistance to cationic antimicrobial peptides in Yersinia', *Molecular Microbiology*, 37(1), pp. 67–80.

Berntsson, R. P.-A. *et al.* (2009) 'The structural basis for peptide selection by the transport receptor OppA', *The EMBO Journal*. Nature Publishing Group, 28(9), pp. 1332–1340. doi: 10.1038/emboj.2009.65.

Berntsson, R. P. A. *et al.* (2010) 'A structural classification of substrate-binding proteins', *FEBS Letters*. Federation of European Biochemical Societies, 584(12), pp. 2606–2617. doi: 10.1016/j.febslet.2010.04.043.

Blair, J. M. A. *et al.* (2015) 'Molecular mechanisms of antibiotic resistance', *Nat Rev Micro*.

Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 13(1), pp. 42–51.

CDC (2018) *CDC: About Antimicrobial Resistance*. Available at: https://www.cdc.gov/drugresistance/about.html (Accessed: 17 December 2018).

Cowtan, K. (2006) 'The {\it Buccaneer} software for automated model building. 1. Tracing protein chains', *Acta Crystallographica Section D*, 62(9), pp. 1002–1011. doi: 10.1107/S0907444906022116.

Daugelavičius, R., Bakienė, E. and Bamford, D. H. (2000) 'Stages of Polymyxin B Interaction with the Escherichia coli Cell Envelope', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology, 44(11), pp. 2969–2978.

Davidson, A. L. *et al.* (2008) 'Structure, function, and evolution of bacterial ATP-binding cassette systems.', *Microbiology and molecular biology reviews : MMBR*, 72(2), p. 317–364, table of contents. doi: 10.1128/MMBR.00031-07.

Detmers, F. J. M. *et al.* (2000) 'Combinatorial peptide libraries reveal the ligand-binding mechanism of the oligopeptide receptor OppA of Lactococcus lactis', *Proceedings of the National Academy of Sciences*, 97(23), p. 12487 LP-12492.

Dintner, S. *et al.* (2014) 'A Sensory Complex Consisting of an ATP-binding Cassette Transporter and a Two-component Regulatory System Controls Bacitracin Resistance in Bacillus subtilis', *The Journal of Biological Chemistry*. 9650 Rockville Pike, Bethesda, MD 20814, U.S.A.: American Society for Biochemistry and Molecular Biology, 289(40), pp. 27899–27910. doi: 10.1074/jbc.M114.596221.

Donovan, R. S., Robinson, C. W. and Click, B. R. (1996) 'Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter', *Journal of Industrial Microbiology*, 16(3), pp. 145–154. doi: 10.1007/BF01569997.

Drousiotis, K. (2017) *Investigation of bacterial transport for improved uptake of hemicellulose-derived sugars and oligosaccharides*. University of York.

Emsley, P. and Cowtan, K. (2004) '{\it Coot}: model-building tools for molecular graphics', *Acta Crystallographica Section D*, 60(12 Part 1), pp. 2126–2132. doi:

Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 13(1), pp. 42–51.

CDC (2018) *CDC: About Antimicrobial Resistance*. Available at: https://www.cdc.gov/drugresistance/about.html (Accessed: 17 December 2018).

Cowtan, K. (2006) 'The {\it Buccaneer} software for automated model building. 1. Tracing protein chains', *Acta Crystallographica Section D*, 62(9), pp. 1002–1011. doi: 10.1107/S0907444906022116.

Daugelavičius, R., Bakienė, E. and Bamford, D. H. (2000) 'Stages of Polymyxin B Interaction with the Escherichia coli Cell Envelope', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology, 44(11), pp. 2969–2978.

Davidson, A. L. *et al.* (2008) 'Structure, function, and evolution of bacterial ATP-binding cassette systems.', *Microbiology and molecular biology reviews : MMBR*, 72(2), p. 317–364, table of contents. doi: 10.1128/MMBR.00031-07.

Detmers, F. J. M. *et al.* (2000) 'Combinatorial peptide libraries reveal the ligand-binding mechanism of the oligopeptide receptor OppA of Lactococcus lactis', *Proceedings of the National Academy of Sciences*, 97(23), p. 12487 LP-12492.

Dintner, S. *et al.* (2014) 'A Sensory Complex Consisting of an ATP-binding Cassette Transporter and a Two-component Regulatory System Controls Bacitracin Resistance in Bacillus subtilis', *The Journal of Biological Chemistry*. 9650 Rockville Pike, Bethesda, MD 20814, U.S.A.: American Society for Biochemistry and Molecular Biology, 289(40), pp. 27899–27910. doi: 10.1074/jbc.M114.596221.

Donovan, R. S., Robinson, C. W. and Click, B. R. (1996) 'Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter', *Journal of Industrial Microbiology*, 16(3), pp. 145–154. doi: 10.1007/BF01569997.

Drousiotis, K. (2017) *Investigation of bacterial transport for improved uptake of hemicellulose-derived sugars and oligosaccharides*. University of York.

Emsley, P. and Cowtan, K. (2004) '{\it Coot}: model-building tools for molecular graphics', *Acta Crystallographica Section D*, 60(12 Part 1), pp. 2126–2132. doi:

10.1107/S0907444904019158.

Eswarappa, S. M. *et al.* (2008) 'The yejABEF operon of Salmonella confers resistance to antimicrobial peptides and contributes to its virulence', *Microbiology*, 154(2), pp. 666–678. doi: 10.1099/mic.0.2007/011114-0.

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 69(Pt 7), pp. 1204–1214. doi: 10.1107/S0907444913000061.

Ganz, T. *et al.* (1985a) 'Defensins. Natural peptide antibiotics of human neutrophils.', *Journal of Clinical Investigation*, 76(4), pp. 1427–1435.

Ganz, T. *et al.* (1985b) 'Defensins. Natural peptide antibiotics of human neutrophils.', *Journal of Clinical Investigation*, 76(4), pp. 1427–1435.

Gaston,   *et al.* (2011) 'Characterization of Peptide Chain Length and Constituency Requirements for YejABEF-Mediated Uptake of Microcin C Analogues', *JOURNAL OF BACTERIOLOGY*, 193(14), pp. 3618–3623. doi: 10.1128/JB.00172-11.

Greenfield, N. J. (2006) 'Using circular dichroism spectra to estimate protein secondary structure', *Nature protocols*, 1(6), pp. 2876–2890. doi: 10.1038/nprot.2006.202.

Griffith, F. (1928) 'The Significance of Pneumococcal Types', *Journal of Hygiene*, 27(2), pp. 113–159. doi: 10.1017/S0022172400031879.

Groisman, E. A. *et al.* (1992) 'Resistance to host antimicrobial peptides is necessary for Salmonella virulence.', *Proceedings of the National Academy of Sciences*, 89(24), pp. 11939–11943. doi: 10.1073/pnas.89.24.11939.

Hancock, R. E. W. (1997) 'Peptide antibiotics', 349, pp. 418–422.

Hancock, R. E. W. and Lehrer, R. (1998) 'Cationic peptides : a new source of antibiotics', 6(1994), pp. 10747–10751.

Harms, C. *et al.* (2001) 'Identification of the ABC protein SapD as the subunit that confers ATP dependence to the K+-uptake systems TrkH and TrKG from Escherichia coli K-12', *Microbiology*, 147(11), pp. 2991–3003.

Heddle, J. *et al.* (2003) 'Crystal Structures of the Liganded and Unliganded Nickel-binding Protein NikA from Escherichia coli', *Journal of Biological Chemistry*, 278(50), pp. 50322–50329. doi: 10.1074/jbc.M307941200.

Hollenstein, K., Frei, D. C. and Locher, K. P. (2007) 'Structure of an ABC transporter in complex with its binding protein', *Nature*. Nature Publishing Group, 446, p. 213.

Jerke, K. H., Lee, M. J. and Humphries, R. M. (2016) 'Polymyxin Susceptibility Testing: a Cold Case Reopened', *Clinical Microbiology Newsletter*. Elsevier, 38(9), pp. 69–77. doi: 10.1016/j.clinmicnews.2016.04.003.

Jones, C. H. *et al.* (2002) 'Escherichia coli DegP Protease Cleaves between Paired Hydrophobic Residues in a Natural Substrate: the PapA Pilin', *Journal of Bacteriology*. American Society for Microbiology, 184(20), pp. 5762–5771. doi: 10.1128/JB.184.20.5762-5771.2002.

Koch, A. L. (1996) 'The Permeability of the Wall Fabric of Escherichia coli and Bacillus subtilis', 178(3), pp. 768–773.

Lai, Y. and Gallo, R. L. (2009) 'AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense', *Trends in immunology*. 2009/02/13, 30(3), pp. 131–141. doi: 10.1016/j.it.2008.12.003.

Lanfermeijer, F. C. *et al.* (1999) 'Kinetics and Consequences of Binding of Nona- and Dodecapeptides to the Oligopeptide Binding Protein (OppA) of Lactococcus lactis', *Biochemistry*. American Chemical Society, 38(44), pp. 14440–14450. doi: 10.1021/bi9914715.

Larrick, J. W. *et al.* (1995) 'Human CAP18: a novel antimicrobial lipopolysaccharide-binding protein.', *Infection and Immunity*, 63(4), pp. 1291–1297.

Lee, J.-K. and Park, Y. (2014) 'Mechanism of Action of Antimicrobial Peptides Against Bacterial Membrane', *Journal of Bacteriology and Virology*, 44(2), p. 140.

Levdikov, V. M. *et al.* (2005) 'The Structure of the Oligopeptide-binding Protein, AppA, from Bacillus subtilis in Complex with a Nonapeptide', *Journal of Molecular Biology*, 345(4), pp. 879–892. doi: https://doi.org/10.1016/j.jmb.2004.10.089.

Levitin, F. *et al.* (2005) 'The MUC1 SEA module is a self-cleaving domain', *Journal of Biological Chemistry*, 280(39), pp. 33374–33386. doi: 10.1074/jbc.M506047200.

Li, J. *et al.* (2006) 'Colistin : the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections'.

Locher, K. P. (2009) 'Structure and mechanism of ATP-binding cassette transporters', *Philosophical Transactions of the Royal Society B: Biological Sciences*. London: The Royal Society, 364(1514), pp. 239–245. doi: 10.1098/rstb.2008.0125.

Locher, K. P., Lee, A. T. and Rees, D. C. (2002) 'The E. coli BtuCD Structure: A Framework for ABC Transporter Architecture and Mechanism', *Science*, 296(5570), p. 1091 LP-1098.

Lodish H, Berk A, Zipursky SL, et al. (2000) 'Molecular Cell Biology', in *4th edition*.

Luque-Ortega, J. R. *et al.* (2008) 'Human antimicrobial peptide histatin 5 is a cell-penetrating peptide targeting mitochondrial ATP synthesis in Leishmania', *The FASEB Journal*. Federation of American Societies for Experimental Biology, 22(6), pp. 1817–1828. doi: 10.1096/fj.07-096081.

Malanovic, N. and Lohner, K. (2016) 'Biochimica et Biophysica Acta Gram-positive bacterial cell envelopes : The impact on the activity of antimicrobial peptides ☆', *BBA - Biomembranes*. The Authors, 1858(5), pp. 936–946. doi: 10.1016/j.bbamem.2015.11.004.

Manniello, J. M., Heymann, H. and Adair, F. W. (1978) 'Resistance of Spheroplasts and Whole Cells of Pseudomonas cepacia to Polymyxin B', *Antimicrobial Agents and Chemotherapy*, 14(3), pp. 500–504.

Manson, M. D. *et al.* (1986) 'Peptide chemotaxis in E. coli involves the Tap signal transducer and the dipeptide permease', *Nature*. Nature Publishing Group, 321, p. 253. Available at: http://dx.doi.org/10.1038/321253a0.

Maqbool, A. *et al.* (2011) 'Compensating stereochemical changes allow murein tripeptide to be accommodated in a conventional peptide-binding protein', *Journal of Biological Chemistry*, 286(36), pp. 31512–31521. doi: 10.1074/jbc.M111.267179.

Marblestone, J. G. *et al.* (2006) 'Comparison of SUMO fusion technology with traditional

gene fusion systems: Enhanced expression and solubility with SUMO', *Protein Science : A Publication of the Protein Society*. Cold Spring Harbor Laboratory Press, 15(1), pp. 182–189. doi: 10.1110/ps.051812706.

Marrs, B. (1974) 'Genetic Recombination in Rhodopseudomonas capsulata', *Proceedings of the National Academy of Sciences of the United States of America*, 71(3), pp. 971–973.

Mason, K. M. *et al.* (2006) 'The non-typeable Haemophilus influenzae Sap transporter provides a mechanism of antimicrobial peptide resistance and SapD-dependent potassium acquisition', *Molecular Microbiology*, 62(5), pp. 1357–1372. doi: 10.1111/j.1365-2958.2006.05460.x.

Mason, K. M., Munson, R. S. and Bakaletz, L. O. (2005) 'A Mutation in the sap Operon Attenuates Survival of Nontypeable Haemophilus influenzae in a Chinchilla Model of Otitis Media ', *Infection and Immunity*. American Society for Microbiology, 73(1), pp. 599–608. doi: 10.1128/IAI.73.1.599-608.2005.

McCoy, A. J. *et al.* (2001) 'Identification of Proteus mirabilis Mutants with Increased Sensitivity to Antimicrobial Peptides', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology, 45(7), pp. 2030–2037. doi: 10.1128/AAC.45.7.2030-2037.2001.

Min, L. *et al.* (2007) 'The antimicrobial peptide-sensing system aps of Staphylococcus aureus', *Molecular Microbiology*. Wiley/Blackwell (10.1111), 66(5), pp. 1136–1147. doi: 10.1111/j.1365-2958.2007.05986.x.

Modi, S. R. *et al.* (2014) 'Ecological Network of the Phage Metagenome', *Nature*, 499(7457), pp. 219–222. doi: 10.1038/nature12212.Antibiotic.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) 'Refinement of Macromolecular Structures by the Maximum-Likelihood Method', *Acta Crystallographica Section D*, 53(3), pp. 240–255. doi: 10.1107/S0907444996012255.

Nickitenko, A. V, Trakhanov, S. and Quiocho, F. A. (1995) '2 .ANG. Resolution Structure of DppA, a Periplasmic Dipeptide Transport/Chemosensory Receptor', *Biochemistry*. American Chemical Society, 34(51), pp. 16585–16595. doi: 10.1021/bi00051a006.

Norman, A., Hansen, L. H. and Sorensen, S. J. (2009) 'Conjugative plasmids: vessels of the communal gene pool', *Philosophical Transactions of the Royal Society B: Biological Sciences*,

364(1527), pp. 2275–2289. doi: 10.1098/rstb.2009.0037.

Novikova, M. *et al.* (2007) 'The Escherichia coli Yej transporter is required for the uptake of translation inhibitor microcin C', *Journal of Bacteriology*. doi: 10.1128/JB.01028-07.

Oldham, M. L. *et al.* (2007) 'Crystal structure of a catalytic intermediate of the maltose transporter', *Nature*. Nature Publishing Group, 450, p. 515.

Park, J. T. *et al.* (1998) 'MppA, a Periplasmic Binding Protein Essential for Import of the Bacterial Cell Wall Peptide l-Alanyl-γ-d-Glutamyl-meso-Diaminopimelate', *Journal of Bacteriology*. American Society for Microbiology, 180(5), pp. 1215–1223.

Parra-Lopez, C., Baer, M. T. and Groisman, E. a (1993) 'Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in Salmonella typhimurium.', *Embo J.*, 12(11), pp. 4053–62.

Pasupuleti, M., Schmidtchen, A. and Malmsten, M. (2012) 'Antimicrobial peptides: key components of the innate immune system', *Critical Reviews in Biotechnology*. Taylor & Francis, 32(2), pp. 143–171. doi: 10.3109/07388551.2011.594423.

Paustian, M. L. *et al.* (2002) 'Transcriptional Response of Pasteurella multocida to Defined Iron Sources', *Journal of Bacteriology*. American Society for Microbiology, 184(23), pp. 6714–6720. doi: 10.1128/JB.184.23.6714-6720.2002.

du Plessis, D. J. F., Nouwen, N. and Driessen, A. J. M. (2011) 'The Sec translocase', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1808(3), pp. 851–865. doi: https://doi.org/10.1016/j.bbamem.2010.08.016.

Pletnev, P. *et al.* (2015) 'Survival guide: Escherichia coli in the stationary phase', *Acta Naturae*. A.I. Gordeyev, 7(4), pp. 22–33.

PORATH, J. *et al.* (1975) 'Metal chelate affinity chromatography, a new approach to protein fractionation', *Nature*. Nature Publishing Group, 258, p. 598.

Potterton, L. *et al.* (2018) 'CCP 4 i 2: the new graphical user interface to the CCP 4 program suite', *Acta Crystallographica Section D Structural Biology*, 74(2), pp. 68–84. doi: 10.1107/S2059798317016035.

Pryor, K. D. and Leiting, B. (1997) 'High-Level Expression of Soluble Protein inEscherichia

coliUsing a His6-Tag and Maltose-Binding-Protein Double-Affinity Fusion System', *Protein Expression and Purification*, 10(3), pp. 309–319. doi: https://doi.org/10.1006/prep.1997.0759.

Puig, O. *et al.* (2001) 'The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification', *Methods*, 24(3), pp. 218–229. doi: https://doi.org/10.1006/meth.2001.1183.

Quan, S. *et al.* (2013) 'Bacterial Cell Surfaces', 966, pp. 359–366. doi: 10.1007/978-1-62703-245-2.

Rijnkels, M. *et al.* (2003) 'Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins', *Genomics*, 82(4), pp. 417–432. doi: 10.1016/S0888-7543(03)00114-9.

Rinker, S. D. *et al.* (2012) 'Permeases of the sap transporter are required for cathelicidin resistance and virulence of haemophilus ducreyi in humans', *Journal of Infectious Diseases*, 206(9), pp. 1407–1414. doi: 10.1093/infdis/jis525.

Saito, T. *et al.* (1995) 'A Novel Big Defensin Identified in Horseshoe Crab Hemocytes: Isolation, Amino Acid Sequence, and Antibacterial Activity', *The Journal of Biochemistry*, 117(5), pp. 1131–1137.

Schein, C. H. and Noteborn, M. H. M. (1988) 'Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature', *Bio/Technology*. Nature Publishing Company, 6, p. 291.

Schroeder, M., Brooks, B. D. and Brooks, A. E. (2017) 'The Complex Relationship between Virulence and Antibiotic Resistance', *Genes*. Edited by H. J. Wing. MDPI, 8(1), p. 39. doi: 10.3390/genes8010039.

Selsted, M. E. *et al.* (1985a) 'Primary structures of three human neutrophil defensins.', *Journal of Clinical Investigation*, 76(4), pp. 1436–1439.

Selsted, M. E. *et al.* (1985b) 'Primary structures of three human neutrophil defensins.', *Journal of Clinical Investigation*, 76(4), pp. 1436–1439.

Shafer, W. M. *et al.* (1998) 'Modulation of Neisseria gonorrhoeae susceptibility to

vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences, 95(4), pp. 1829–1833.

Sharkey, L. K. R., Edwards, T. A. and O'Neill, A. J. (2016) 'ABC-F Proteins Mediate Antibiotic Resistance through Ribosomal Protection', *mBio*. Edited by G. D. Wright, 7(2). Available at: http://mbio.asm.org/content/7/2/e01975-15.abstract.

Shelton, C. L. *et al.* (2011) 'Sap transporter mediated import and subsequent degradation of antimicrobial peptides in Haemophilus.', *PLoS pathogens*, 7(11), p. e1002360. doi: 10.1371/journal.ppat.1002360.

Shepherd, M., Heath, M. D. and Poole, R. K. (2007) 'NikA Binds Heme: A New Role for an Escherichia coli Periplasmic Nickel-Binding Protein', *Biochemistry*. American Chemical Society, 46(17), pp. 5030–5037. doi: 10.1021/bi700183u.

Silhavy, T. J., Kahne, D. and Walker, S. (2010) 'The Bacterial Cell Envelope', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 2(5), p. a000414. doi: 10.1101/cshperspect.a000414.

Sobrero, P. and Valverde, C. (2012) 'The bacterial protein Hfq: much more than a mere RNA-binding factor', *Critical Reviews in Microbiology*. Taylor & Francis, 38(4), pp. 276–299. doi: 10.3109/1040841X.2012.664540.

Stumpe, S. *et al.* (1998) 'Identification of OmpT as the Protease That Hydrolyzes the Antimicrobial Peptide Protamine before It Enters Growing Cells of Escherichia coli Identification of OmpT as the Protease That Hydrolyzes the Antimicrobial Peptide Protamine before It Enters Growi', *J. Bacteriol.*, 180(15), pp. 4002–4006.

Sugiyama, Y., Nakamura, A. and Matsumoto, M. (2016) 'A novel putrescine exporter SapBCDF A novel putrescine exporter SapBCDF of Escherichia coli . A novel putrescine exporter SapBCDF'. doi: 10.1074/jbc.M116.762450.

Tame, J. R. *et al.* (1995) 'The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands.', *Structure (London, England : 1993)*, 3(12), pp. 1395–1406. doi: 10.1016/S0969-2126(01)00276-3.

Terwilligert, T. C. and Eisenbergg, D. (1982) 'The Structure of Melittin', 257(11), pp. 6016–

6022.

Thomassin, J.-L. *et al.* (2012) 'OmpT Outer Membrane Proteases of Enterohemorrhagic and Enteropathogenic Escherichia coli Contribute Differently to the Degradation of Human LL-37', *Infection and Immunity*. Edited by B. A. McCormick. 1752 N St., N.W., Washington, DC: American Society for Microbiology, 80(2), pp. 483–492. doi: 10.1128/IAI.05674-11.

Thomma, B. P., Cammue, B. P. and Thevissen, K. (2002) 'Plant defensins', *Planta*, 216(2), pp. 193–202. doi: 10.1007/s00425-002-0902-6.

Vabulas, R. M. *et al.* (2010) 'Protein Folding in the Cytoplasm and the Heat Shock Response', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 2(12), p. a004390. doi: 10.1101/cshperspect.a004390.

Vagin, A. and Teplyakov, A. (2010) 'Molecular replacement with {\it MOLREP}', *Acta Crystallographica Section D*, 66(1), pp. 22–25. doi: 10.1107/S0907444909042589.

Vendruscolo, M. (2012) 'Proteome folding and aggregation', *Current Opinion in Structural Biology*, 22(2), pp. 138–143. doi: https://doi.org/10.1016/j.sbi.2012.01.005.

Viljanen, P. and Vaara, M. (1984) 'Susceptibility of gram-negative bacteria to polymyxin B nonapeptide.', *Antimicrobial Agents and Chemotherapy*, 25(6), pp. 701–705.

Wang, Z. *et al.* (2016) 'The ABC transporter YejABEF is required for resistance to antimicrobial peptides and the virulence of Brucella melitensis', *Scientific Reports*. The Author(s), 6, p. 31876.

Warren, T. D., Coolbaugh, M. J. and Wood, D. W. (2013) 'Ligation-independent cloning and self-cleaving intein as a tool for high-throughput protein purification', *Protein expression and purification*, 91(2), pp. 169–174. doi: 10.1016/j.pep.2013.08.006.

Wellcome Trust (2016) 'TACKLING DRUG-RESISTANT INFECTIONS GLOBALLY : FINAL REPORT AND RECOMMENDATIONS THE REVIEW ON', (May).

Wiesner, J. and Vilcinskas, A. (2010) 'Antimicrobial peptides: The ancient arm of the human immune system', *Virulence*. Taylor & Francis, 1(5), pp. 440–464. doi: 10.4161/viru.1.5.12983.

Wilkens, S. (2015a) 'Structure and mechanism of ABC transporters', *F1000Prime Reports*.

Faculty of 1000 Ltd, 7, p. 14. doi: 10.12703/P7-14.

Wilkens, S. (2015b) 'Structure and mechanism of ABC transporters', *F1000Prime Reports*. Faculty of 1000 Ltd, 7, p. 14. doi: 10.12703/P7-14.

Wimley, W. C. (2010) 'Describing the Mechanism of Antimicrobial Peptide Action with the Interfacial Activity Model', *ACS chemical biology*, 5(10), pp. 905–917.

Winkler, J. *et al.* (2010) 'Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing', *The EMBO Journal*. Nature Publishing Group, 29(5), pp. 910–923. doi: 10.1038/emboj.2009.412.

Von Wintersdorff, C. J. H. *et al.* (2016) 'Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer', *Frontiers in Microbiology*, 7(FEB), pp. 1–10. doi: 10.3389/fmicb.2016.00173.

World Health Organisation (2018) *Antimicrobial resistance*. Available at: https://www.who.int/en/news-room/fact-sheets/detail/antimicrobial-resistance (Accessed: 17 December 2018).

Yang, L. *et al.* (2001) 'Barrel-Stave Model or Toroidal Model ? A Case Study on Melittin Pores', 81(September), pp. 1475–1485.

Zaman, S. Bin *et al.* (2017) 'A Review on Antibiotic Resistance: Alarm Bells are Ringing', *Cureus*. Edited by A. Muacevic and J. R. Adler. Palo Alto (CA): Cureus, 9(6), p. e1403.

Zanetti, M. and Gennaro, R. (1995) 'Cathelicidins : a novel protein family with a common proregion and a variable C-terminal antimicrobial domain Margherita Zanetti a ' b '*, Renato Gennaro b , D o m e n i c o', 374, pp. 1–5.

Zgurskaya, H. I. *et al.* (2011) 'Mechanism and function of the outer membrane channel TolC in multidrug resistance and physiology of enterobacteria', *Frontiers in Microbiology*, 2(SEP), pp. 1–13. doi: 10.3389/fmicb.2011.00189.

Zhang, P. (2013) 'Structure and mechanism of energy-coupling factor transporters', *Trends in Microbiology*, 21(12), pp. 652–659. doi: https://doi.org/10.1016/j.tim.2013.09.009.