Investigation of the function and secretion of CexE in enterotoxigenic *Escherichia coli*

By

Christopher Icke

This thesis is submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection

College of Medical and Dental Sciences

University of Birmingham

September 2017

Abstract

Enteroaggregative *Escherichia coli* (EAEC) utilises an atypical type I secretion system, AatPABCD, for the secretion of dispersin (Aap). AggR is known to regulate the transcription of the *aat* system and *aap* in EAEC 042. Using RNA-Seq, the transcriptomes of EAEC 042 and EAEC 042Δ*aggR* were compared for differential gene expression. As well as new targets for AggR activation, *aap* was found to have the greatest difference in transcription. Whilst the function of Aap is understood, the role of the homolog, CexE, in enterotoxigenic *E. coli* H10407 is not. It was found that CexE could not complement the *aap* mutant phenotype instead having the opposite effect, resulting in an increase in aggregation. Aggregation can indicate a role in biofilm. It was found that a *cexE* mutant could not form a biofilm. The investigation of CexE structure revealed that CexE could be secreted by the Aat system. Therefore, Aat-mediated secretion of CexE was investigated. A modification by AatD of Aap/CexE proteins was identified and a new model for Aat secretion is proposed. In summary, the function of Aap and CexE are different but they are both secreted by the Aat system.

Acknowledgements

First I would like to thank my supervisor, Professor Ian Henderson, for his support and guidance. Ian's influence has encouraged me to concentrate on my future career in science while making the group an exciting and fun place to work. Professor Jeff Cole has been a great source of advice on science, presentations, and writing. Dr. Timothy Wells has helped constantly from my start in T101 to the writing of my thesis.

I would like acknowledge the contribution of Dr. Timothy Knowles to this work. Tim collected the NMR experimental data for me, taught me how to interpret of the data, and calculated the structure of CexE. Sammi McKeand constructed the pCfaD plasmid for me and has been wonderful at orders and organising the laboratory. Finally, Charles Moore-Kelly who helped with the circular dichroism of CexE and entertainment in Birmingham.

I would like to thank the MIBTP program for funding my work and introducing me to some wonderful scientists. In particular, Richard Meek, who has kept me company in the library and is always keen to share his structural knowledge in the pub. Jack Connelly for his advice with plasmids also in the pub and Tamar Cranford-Smith for her friendship and help in the lab.

I would like to thank everyone in T101 past and present for their help and support over the last four years. It has been a wonderful place to come to everyday with great science and great friends. Especially my work colleagues, Jessica Rooke and Emily Goodall, who have been a constant source of laughter and entertainment. Georgia Isom, Emma Sheehan, and Jack Bryant have been a great help at all times with both science and fun.

My parents, Brian Icke and Susan Brockman, have been a constant source of inspiration support and help throughout my life. My aunt, Helen Brockman, has provided home cooked meals and hospitality and my sister, Elizabeth Brockman, has attempted to keep me sane and show me the delights of Baltimore. Before, during, and after my write up, Niamh McConnell has been wonderful thought-out and tolerant of my writing method.

The people I have come in contact with during my time here have made these past four years some of the best in my life.

Contents

List of figures

Figure **Page**

List of tables

CHAPTER 1:

Introduction

1.1. General introduction

Secretion systems are used by bacteria to affect their environment by translocation of molecules across their cell envelope. The substrates of these system range from simple proteins for nutrient acquisition to multi-protein complexes with the purpose of subverting eukaryotic cell physiology. Since the discovery of the first secretion system the field of bacterial secretion has grown. Currently, there are numerous different secretion systems that have been identified and confirmed experimentally in both Gram-negative and Gram-positive bacteria.

The complexity of the Gram-negative cell envelope requires multi-protein systems to secrete molecules across the hydrophobic membranes to the cell surface and the environment. This thesis will describe the six best characterised Gram-negative secretion systems before focusing on the type I secretion system (T1SS) found in many pathogenic *Escherichia. coli*. The focus of this thesis will be on an atypical T1SS, the Aat system, found in strains belonging to enteroaggregative (EAEC) and enterotoxigenic *E. coli* (ETEC) pathotypes. In EAEC 042 the Aat system is known to be required for the secretion of Aap (Nishi et al., 2003). However, there is limited knowledge on the function and secretion of the Aap related protein, CexE, in ETEC. Therefore, this thesis set out to investigate the CexE protein in ETEC strain H10407 and the Aat secretion system.

1.2. Membrane physiology of Gram-negative bacteria

1.2.1. Gram-negative cell envelope

The Gram-negative cell envelope consists of two lipid membranes separated by a periplasmic space containing the peptidoglycan. The outer leaflet of the outer membrane is composed of lipopolysaccharide (LPS). LPS molecules are large amphipathic molecules that are composed of a conserved hydrophobic lipid A attached to a long chain polysaccharide (Fig. 1.1). This

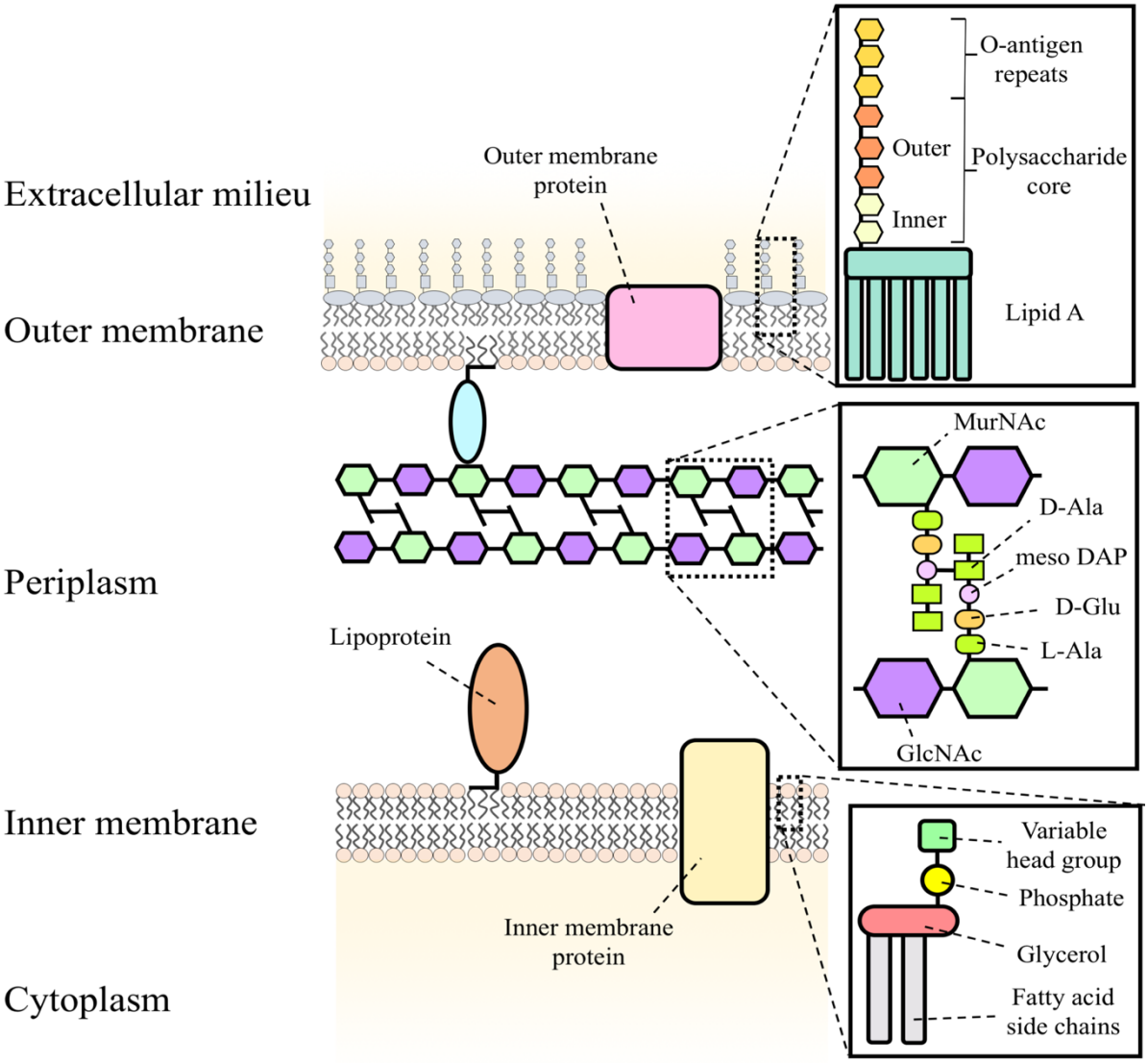


Figure 1.1: Gram-negative cell envelope. The outer membrane of a Gram-negative cell is an asymmetric barrier of phospholipids and lipopolysaccharide (LPS). The LPS is comprised of a lipid A molecule attached to a polysaccharide chain (top right). The length of the O-antigen repeats change between strains. Proteins in the outer membrane possess a β-barrel. The periplasm is a region between the outer and inner membrane. It contains the peptidoglycan of the Gram-negative cell (middle right). The peptidoglycan is covalently bound to Lpp, a lipoprotein in the inner leaflet of the outer membrane. The inner membrane also contains lipoproteins. This lipid bilayer is made of many different phospholipids, a schematic of which is present in the bottom left panel of this figure. Proteins in the inner membrane contain transmembrane helices.

layer prevents large and hydrophobic molecules from entering the cell. There are three regions of LPS: lipid A; polysaccharide core and O-antigen. The lipid A moiety anchors the LPS in the outer leaflet of the outer membrane. Lipid A contains 6 fatty acid chains instead of the 2 to 4 that are typical for other membrane lipids (Delcour, 2009). The hydrophilic O-antigen is antigenic. It is the most variable part of the LPS with a great diversity between strains. LPS molecules are negatively charged and electrostatically repel one another. This is overcome by the binding of divalent cations such as Ca^{2+} and Mg^{2+} (Amro et al., 2000). The outer membrane is asymmetric as the inner leaflet is populated by phospholipids.

Phospholipids are the major components of both the inner leaflet of the outer membrane and the inner membrane. These are amphipathic lipids that are commonly glycerophospholipids. They are composed of two fatty acids, a glycerol moiety, a phosphate group, and a variable head group (Fig. 1.1). In *E. coli,* there are three predominant phospholipids phosphatidylethanolamine, phosphatidylglycerol, and cardiolipin. Phosphatidylethanolamine is the predominant membrane phospholipid at around 75% of all membrane lipids while other phospholipids like phosphatidylglycerol and cardiolipin constitute 20% of membrane lipids (Raetz and Dowhan, 1990). The inner and the outer membrane are the major barriers for the import, export, or secretion of hydrophilic molecules. With the region between the inner and outer membrane termed the periplasm.

The periplasm is a gel-like matrix that contains a thin peptidoglycan layer. Peptidoglycan prevents the rupture of the cytoplasmic membrane from turgor pressure and gives the bacterial cell its structure. It is composed of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) monomers linked by a β-1,4 glycosidic bond. Peptide bridges are attached to MurNAc connecting the peptidoglycan chains (Fig. 1.1, middle right). The chirality of the

peptides in these bridges is unusual as they are rare D-amino acids. The peptidoglycan layer is anchored to the outer membrane by Lpp (Braun's lipoprotein); one of the most abundant proteins in *E. coli*. Its lipidated N-terminus resides in the inner leaflet of the outer membrane while it is covalently bound to the peptidoglycan layer (Braun and Wolff, 1970). Many proteins, like Lpp, reside in the periplasm or are associated with the outer membrane in *E. coli* and thus must pass through the inner membrane. These proteins are secreted via the general secretory (Sec) pathway.

1.2.2. General secretory pathway

The general secretory (Sec) pathway secretes proteins from the cytoplasm into the periplasm or inserts them into the inner membrane in Gram-negative bacteria. The majority of non-cytoplasmic proteins are translocated by the Sec pathway (Orfanoudaki and Economou, 2014). These proteins contain an N-terminal secretion signal sequence that targets them for secretion by the inner membrane complex SecYEG. This sequence is composed of one or more positive charges with a hydrophobic region of 8 to 12 amino acids (Papanikou et al., 2007). In general, the signal sequence is recognised either by ribonucleoprotein signal recognition particle (SRP) or the cytoplasmic chaperone SecB (Fig. 1.2). SRP is typically recruited by more hydrophobic signal sequences and targets the translating ribosome for co-translational secretion (Akopian et al., 2013). The SRP membrane receptor FtsY docks with the ribosome-protein-SRP complex and targets it to SecYEG (Luirink and Sinning, 2004). The translation of the protein from the ribosome drives the translocation of the protein. Through a lateral gate mechanism the transmembrane helix exits the SecYEG translocon (Gogala et al., 2014). For non-inner membrane proteins the cytoplasmic chaperone SecB prevents the aggregation of the translated protein in the cytoplasm (Randall and Hardy, 2002). The SecB-protein complex is targeted to

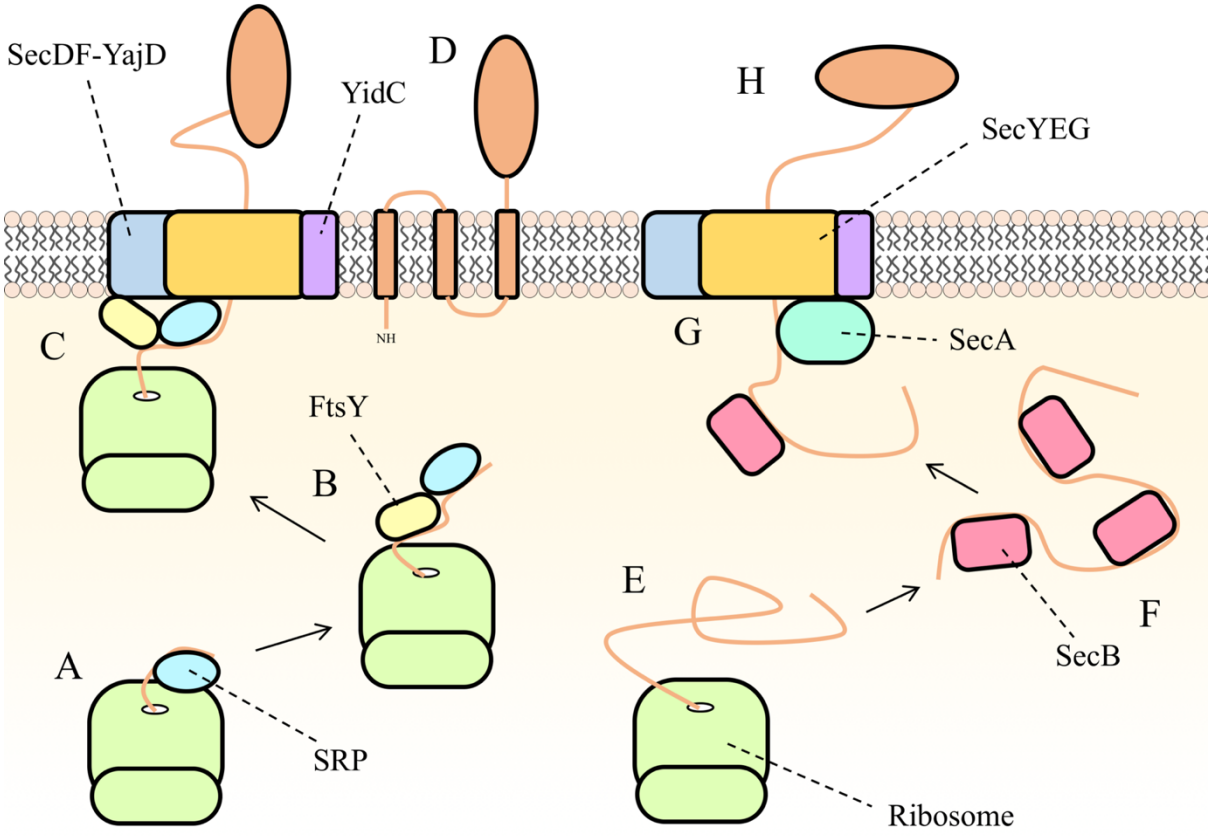


Figure 1.2: Secretion of proteins into the periplasm. The Sec machinery consists of SecYEG, SecDF-YajD, and YidC. SecYEG is the pore of the complex through which the peptides are secreted. SecDF-YajD and YidC have auxiliary role in peptide translocation. A-D; **Co-translational pathway**; A, SRP scans ribosomes for a signal peptide that is targeted for the Sec translocon. B, FtsY is recruited and the peptide-SRP-FtsY-ribosome complex is targeted to the Sec machinery. C, FtsY interacts with membrane and SRP feeds the nascent peptide into the Sec translocon. Proteins recognised by SRP are typically targeted to the inner membrane. E-G; **Post-translational pathway**; E, a protein is produced by a ribosome. F, the peptide is bound by the chaperone SecB to prevent folding. G, the SecB-peptide is targeted to the SecA-SecYEG translocon. SecA provides the energy for translocation. The protein is released into the periplasm.

SecA. SecA is bound to SecYEG and it guides the nascent protein into the SecYEG complex while providing the energy for translocation (Lecker et al., 1990).

SecYEG forms an hourglass like structure with a plug on the cytoplasmic side, thought to prevent undesired movement of molecules across the membrane (Berg et al., 2004). Accessory proteins such as SecDF-YajC and YidC improve translocation efficiency (Schulze et al., 2014). SecDF-YajC has been suggested to assist secretion by providing energy through the use of the proton motive force (Schulze et al., 2014). YidC is a catalyst for the insertion of hydrophobic α helices into the inner membrane (Dalbey et al., 2014). After translocation the signal peptides are cleaved, by signal peptidases, allowing the release of the translocated proteins (Paetzel et al., 2002). Proteins are then allowed to fold or continue to further maturation processes.

1.2.3. Transport of lipoproteins to the outer membrane

Lipoproteins are distinct membrane proteins as their N-terminus is attached to lipid moieties allowing them to localise to the inner and outer membrane (Fig. 1.3). Lipoproteins possess a characteristic signal peptide that contains a lipobox at the C-terminus; a characteristic set of residues that results in a cysteine at the N-terminus of the protein after signal sequence cleavage (Heijne, 1989). The signal peptide is inserted into the inner membrane by the Sec pathway (Fig. 1.3A). Prolipoprotein diacylglyceryl transferase (Lgt) catalyses the formation of a thioester bond between the sulfhydryl moiety of the N terminal cysteine and diacylglyercol, which is transferred from a phosphatidylglycerol (Sankaran and Wu, 1994). The signal peptide is cleaved by Lsp. The N-terminal NH group is made available for acylation by apolipoprotein N-acyltransferase (Lnt). Lnt catalyses the transfer of an acyl chain of phosphatidylethanolamine to the N-terminus of the substrate lipoprotein (Jackowski, 1986).

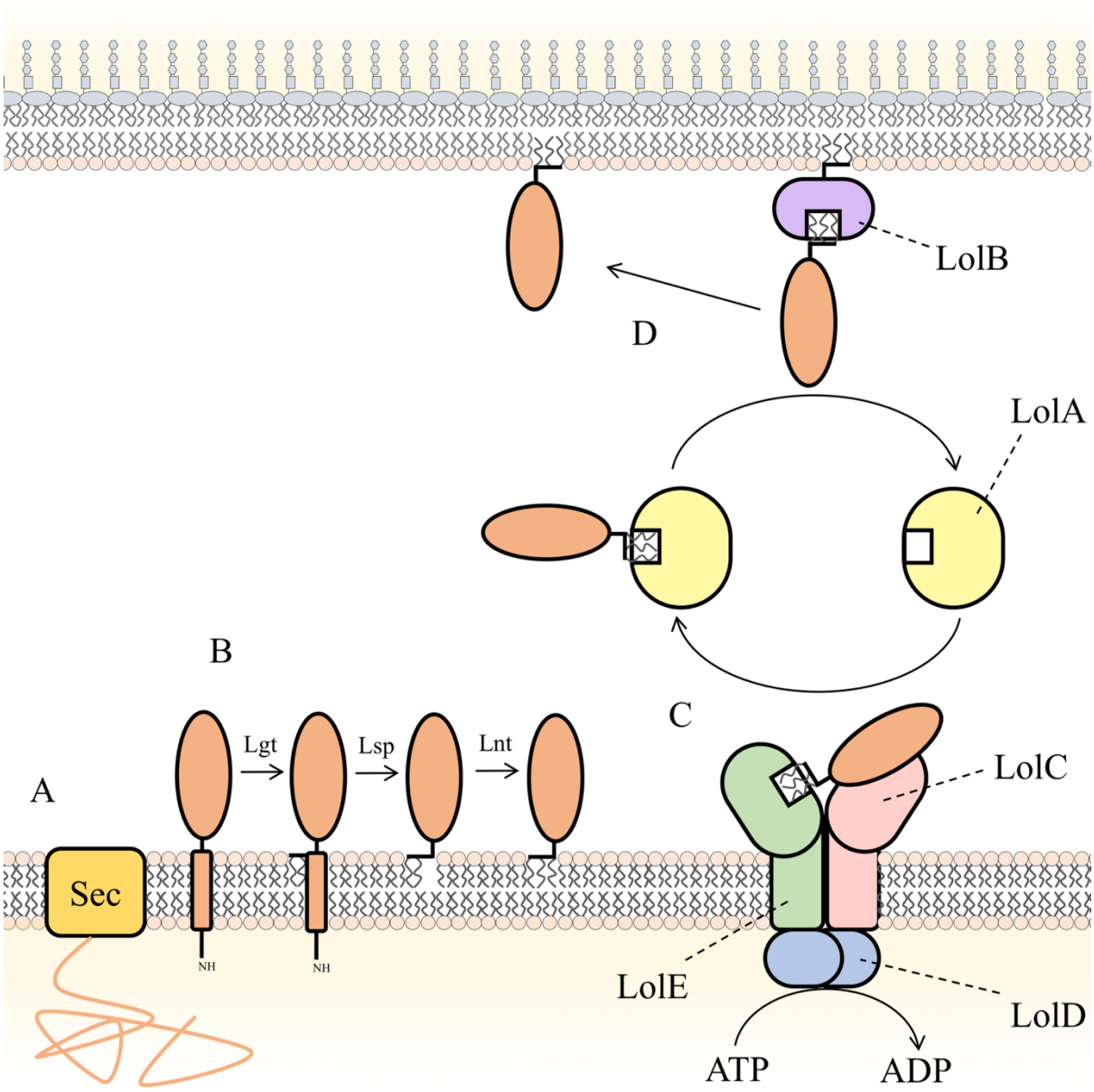


Figure 1.3: Maturation and transport of lipoproteins. A, a nascent peptide with a lipobox is secreted into the periplasm. B, Lgt adds two fatty chain groups to the +1 cysteine. Lsp cleaves the protein leaving the NH group of the cysteine. A final fatty acid from phosphatidylethanolamine is added to the NH group of the cysteine by Lnt. C, if the protein does not have a Lol avoidance signal it is sent to the outer membrane by the Lol pathway. LolE captures the lipoprotein from the membrane and with LolD and LolC transfers it to LolA. LolA transports the lipoprotein across the hydrophilic periplasm. D, LolB removes the lipoprotein from LolA and inserts it into the inner leaflet of the outer membrane.

The full maturation of a tri-acylated lipoprotein is essential for entry into the localisation of lipoproteins (Lol) pathway. The Lol pathway is responsible for the translocation of lipoproteins to the outer membrane based on the residue after the cysteine, known as the +2 rule. The presence of an asparagine residue at the +2 positions retains the protein in the inner membrane (Yamaguchi, 1988). The presence of any other residue in the +2 position results in a translocation to the outer membrane. The tri-acylated lipoprotein is removed from the inner membrane by the LolCDE complex and passed to LolA (Yakushi et al., 2000). This complex is composed of single copies of LolC and LolE and two copies of LolD. Together these proteins function as an ABC transporter. However, unlike typical ABC transporters they do not translocate molecules across a membrane. LolC and LolE have a similar sequence and topology but a different structure (Yasuda et al., 2009). LolE is suggested to capture the lipoprotein with a hydrophobic region, which is similar to LolA and LolB (Mizutani et al., 2013a). This causes a structural change that increases the affinity of LolD for ATP (Ito et al., 2006). LolC recruits a LolA that is not loaded with a lipoprotein (Okuda and Tokuda, 2009). The binding of the lipoprotein to LolE is weaken by the binding of ATP to LolD (Mizutani et al., 2013a). This leads to the transfer of the lipoprotein to LolA, by a mouth to mouth mechanism, and the release of LolA (Okuda and Tokuda, 2009). LolA with a lipoprotein is recruited to the outer membrane by LolB. The lipoprotein is passed from LolA to LolB and is inserted into the inner leaflet of the outer membrane. The insertion of a lipoprotein into the inner leaflet of the outer membrane is typically the end of lipoprotein translocation. However, some surface exposed lipoproteins have been described recently (Majdalani et al., 2005; Cowles et al., 2012; Webb et al., 2013).

1.2.4. β-barrel assembly machinery complex

Integral outer membrane proteins predominantly adopt a β-barrel conformation. The β-barrel assembly machinery (BAM) complex inserts these proteins into the outer membrane of Gram-

negative bacteria. These β-barrel proteins are characterised by anti-parallel β-sheets that form a barrel-like shape. Each β strand is composed so that the hydrophobic side chains face the outer membrane lipids while the hydrophilic side chains face the lumen of the barrel. The first and last β-strand form the junction for the β-barrel. Proteins that are inserted into the outer membrane are translated in the cytoplasm and transported across the inner membrane by the Sec pathway. In the periplasm, they are bound by chaperones, such as SurA and Skp, and targeted to the BAM complex (Sklar et al., 2007). The BAM complex is made of five proteins, BamA-E. BamA is the major component of the BAM complex, which is a β-barrel protein itself. The structure of BamA revealed a lateral gate that is likely used for the insertion of β-barrel proteins into the outer membrane (Noinaj et al., 2013). The BamA protein contains five polypeptide transport associated (POTRA) domains some of which act as the scaffold for the four associated lipoproteins BamB, BamC, BamD, and BamE. Although BamA and BamD are essential for the outer membrane protein insertion and cell viability while, the other Bam proteins may play a role in the insertion of more complex β-barrel proteins (Wu et al., 2005; Malinverni et al., 2006). The mechanism by which the BAM complex mediates the β-barrel insertion into the outer membrane is not yet fully understood.

1.3. Protein secretion systems of Gram-negative bacteria

The process of secretion across the two membranes of bacteria presents certain physical and biochemical challenges. Not only does the substrate protein have to traverse two membranes and the periplasmic space but the final steps must be accomplished without the expenditure of energy. To overcome these obstacles Gram-negative bacteria have developed secretion systems which make use of the Sec pathway, lipoproteins, and the BAM complex to create multimeric machines for the secretion of their substrates. There are many different secretion systems but

the best characterised are known as types I-VI. The type I secretion system (T1SS) will be considered in greater detail later on in this Chapter.

1.3.1. Type II secretion systems

The T2SS is a complex structure consisting of 12 to 15 different proteins that form a single machine totalling 40 to 70 proteins (Korotkov et al., 2012). T2SS consist of 4 major complexes the outer-membrane complex, inner-membrane platform, the secretion ATPase, and the pseudopilus (Korotkov et al., 2012) (Fig. 1.4). The type II secretion system (T2SS) secretes folded periplasmic proteins that were translocated by the Sec or Tat machinery. The Tat machinery is similar to the Sec pathway but it secretes folded proteins from the cytoplasm into the periplasm. The number of substrates of the T2SS varies, some systems are required for a single substrate while others secrete a number of different proteins. A well characterised example of a T2SS substrate is the heat labile toxin of ETEC. This is a AB_5 toxin, which is similar to the cholera toxin, that causes watery diarrhoea. Korotkov et al. (2012) suggested the "piston model" as the mechanism of T2SS (Fig. 1.4). In this model, the inner membrane platform complexes with the secretion ATPase, the outer membrane complex forms separately. The pseudopilins forms at the inner membrane platform. The substrate is then pushed out of the cell by the pseudopilins like a piston.

1.3.2. Type III secretion systems

The type III secretion system (T3SS) is a syringe like apparatus that is used by bacteria to secrete proteins, known as effectors, into host cells. The T3SS has a core of 9 proteins with 10 to 20 auxiliary proteins depending on the function of the system (Green and Mecsas, 2016). In general, the T3SS is comprised of three major components: the base complex; the needle; and the translocon (Abrusci et al., 2014) (Fig. 1.5). The base complex contains the cytoplasmic

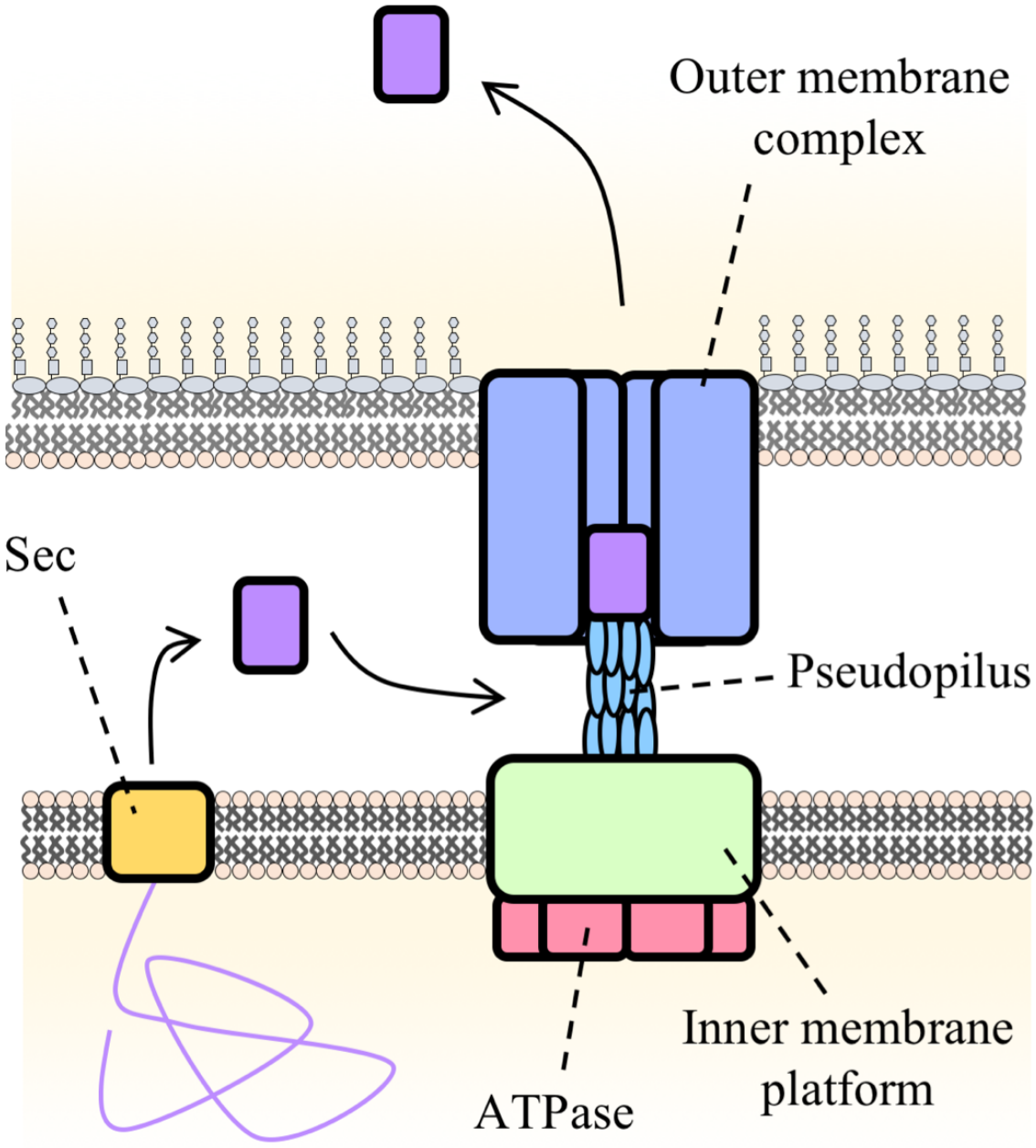


Figure 1.4: Type II Sectrion System. The T2SS makes use of either the Sec (pictured) or Tat pathways to secrete the substrate protein into the periplasm. The folded protein then enters the T2SS. The T2SS is proposed to use a "piston" mechanism to secrete proteins from the periplasm to the extracellular milieu; the psudopilins complex on the inner membrane platform and push the protein out of the cell like a piston.

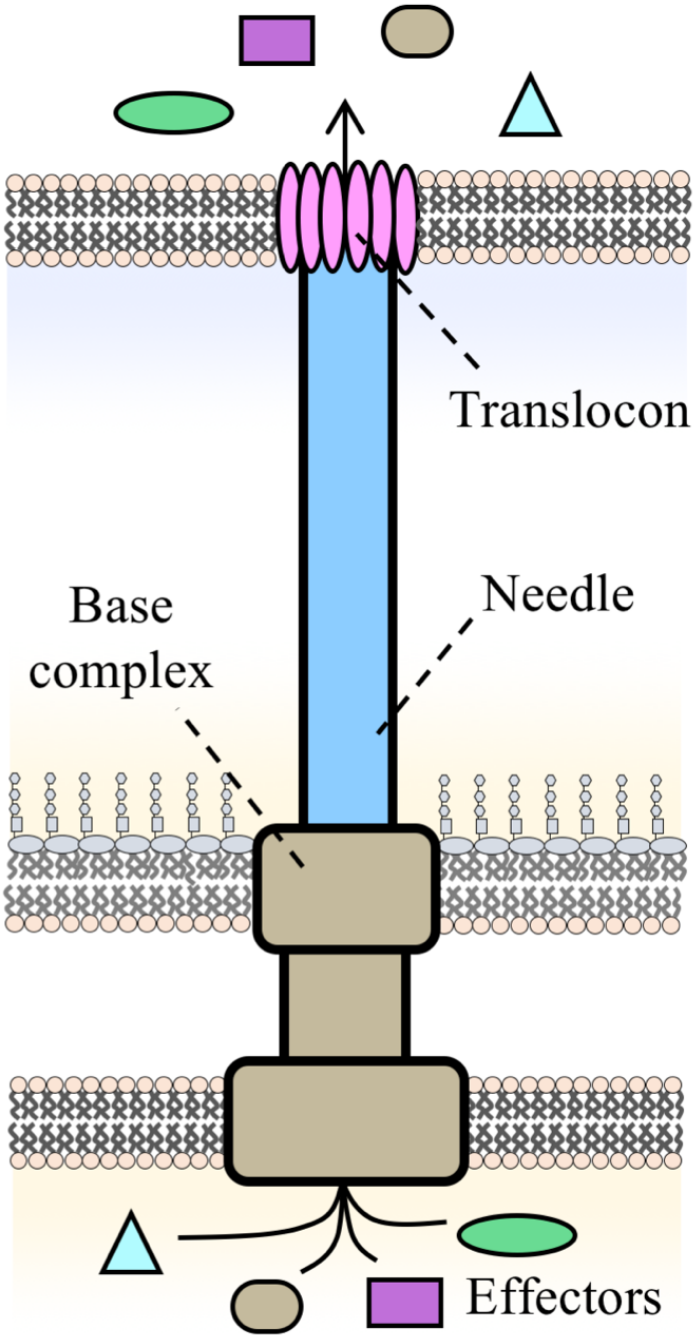


Figure 1.5: Type III Sectrion System. In the T3SS, effector proteins are directly secreted into the target cell. The base complex allows the formation and extention of the needle from the bacterial cell. When the the needle contacts the target cell the translocon forms. This makes a pore in the membrane of the target cell through which the effectors enter and subvert cellular processes in the target cell.

machinery and spans the two Gram-negative membranes (Murray et al., 1998). The base complex is the foundation for the extension of a filament like structure called a needle (Murray et al., 1998). The core of the needle is hollow allowing the passage of effectors (Demers et al., 2014). The tip of the needle is necessary for the sensing of host cells and the insertion of the translocon (Picking et al., 2005). The translocon is required for the final stage of the T3SS, upon host cell contact the translocon assembles forming a pore in the host cell membrane and the effector proteins are secreted (Hakansson et al., 1993; Olsson et al., 2001). The secretion of these effectors is used to modulate host cell biology by many pathogens. Enteropathogenic (EPEC) uses a diverse range of effectors to causes disease for example: Map affects mitochondrial structure and function (Ma et al., 2006); EspF disrupts eukaryotic tight junctions (Guttman et al., 2006); and Cif which prevents cell cycle progression (Samba-louaka et al., 2008).

1.3.3. Type IV secretion systems

Similar to the T3SS the type IV secretion system (T4SS) spans both the cell envelope and the host cell membrane (Fig. 1.6). There are many functions of the T4SS: conjugation for the transfer of plasmids and transposons (Grohmann and Espinosa, 2003); DNA uptake (Hamilton and Dillard, 2006); DNA delivery to host cells (Bundock et al., 1995); and secretion of virulence proteins into host cells (Backert and Meyer, 2006). Despite their variation in substrates and function all of the T4SS machinery are related. There are three major components of the T4SS, the inner membrane complex, the outer membrane complex and generally a pilus. As the names suggest the inner and outer membrane complexes span the respective inner and outer membranes. The Inner membrane complex provides the energy for the system and the outer membrane complex inserts into the inner membrane complex and provides a pore through which the pilus extends. The substrate is transferred from the

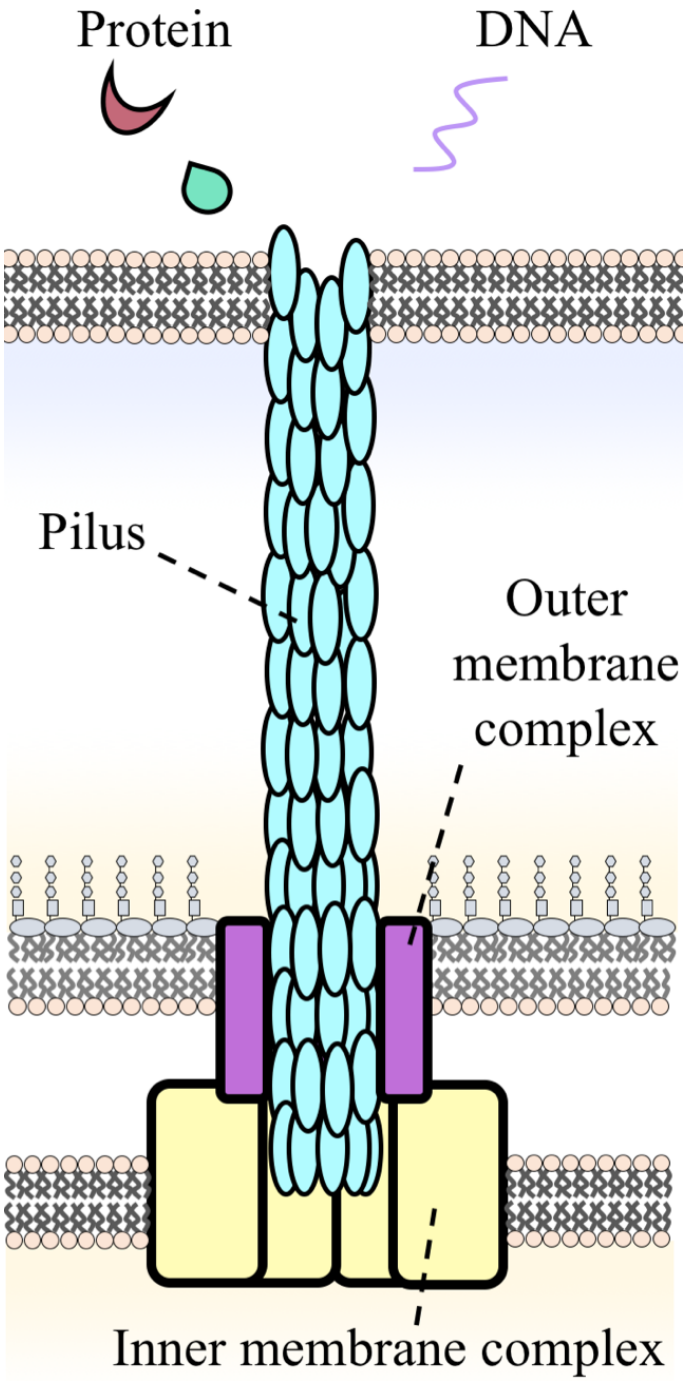


Figure 1.6: Type IV Sectrion System. The inner membrane complex and outer membrane complex form in the bacterium's envelope. This allows the extension of the pilus, which is made up of many copies of the pilin protien. When attached to the target cell the T4SS can secrete proteins or DNA into the target cell.

cytoplasm to the inner membrane complex, then to the outer membrane complex. It is not known if the pilus functions as an attachment device that eases substrate secretion or is necessary for the translocation of substrates to host cells (Fronzes et al., 2009; Babić et al., 2008).

1.3.4. Type V secretion systems

The type V secretion systems (T5SS) covers five sub-classes of secretion pathway, Type 5a-e (Fig. 1.7). Of these the best characterised is Type 5a, which were originally referred to as autotransporters as it was originally thought that the single primary protein sequence was sufficient for secretion to the cell surface. However, it is now known that multiple other protein complexes are required for the secretion of the T5SS proteins including the Sec machinery, Tam complex, periplasmic chaperones, and the BAM complex (Selkrig et al., 2012; Ruiz-perez et al., 2009; Rossiter et al., 2011b). All five sub-classes of T5SS use a β-barrel inserted in the outer membrane by the BAM complex for substrate secretion (Voulhoux et al., 2003). The substrate of a T5SS is known as the passenger domain. Type 5a, c, d, and e the passenger domain and β-barrel are encoded in the same gene. The passenger domain can be cleaved from the β-barrel domain and released into the extracellular milieu. In type 5b, also known as two partner secretion, the β-barrel and passenger domain are encoded as two genes in an operon. The β-barrel of the type 5b can be slightly more complex than type 5a and involved in modification of the passenger domain after transport (Henderson et al., 2004). Type 5c are obligate trimers that primarily form adhesins (Linke et al., 2006). Type 5d are similar to the type 5a but have an additional passenger domain homologous to regions in type 5b (Fan et al., 2016). The final subclass of autotransporter is the type 5e or inverse autotransporters. In contrast to type 5a, the passenger domain is C-terminal and the β-barrel is N-terminal (Leo et

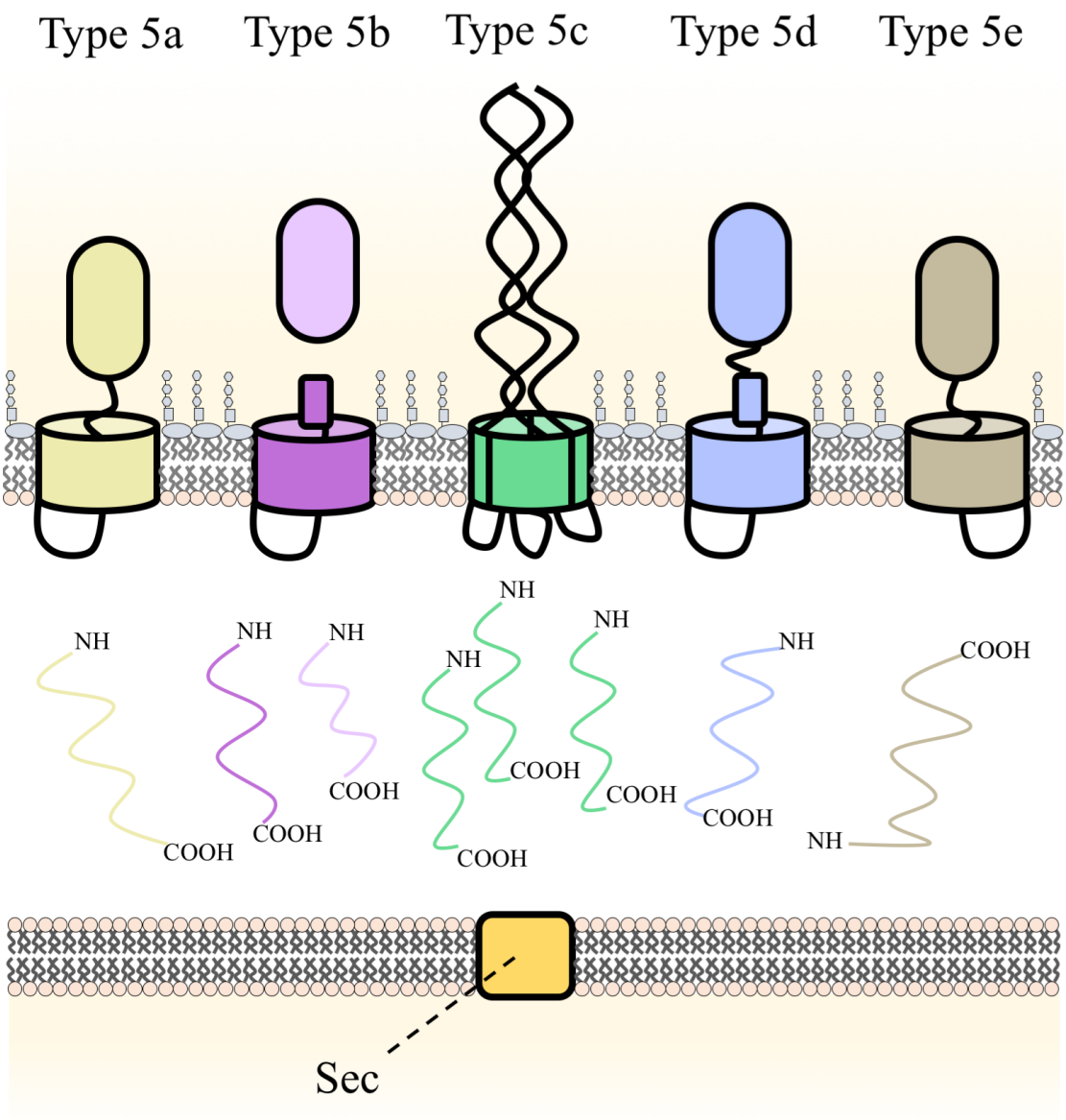


Figure 1.7: Type V Sectrion System. The T5SS is split into 5 subclasses. Type 5a is encoded by a single gene which has the β-barrel at the N-terminus and the passenger doamin at the C-terminus. This is typical of all T5SS except for type 5e, which is inverted compared to type 5a. The type 5b consists of two two proteins, a β-barrel and a passenger domain. The β-barrel of the type 5b may modify the passenger domain after transport. The type 5c consists of three proteins that form a trimeric β-barrel and long passenger domain that extends from the cell surface. Type 5d are similar to type 5a but they also have an extra domain that is invovled in the modification of the passenger domain.

al., 2012). The T5SS are used by commensal and pathogenic bacteria for the secretion of numerous proteins including enzymes, toxins, adhesins, and intracellular motility.

1.3.5. Type VI secretion systems

The type VI secretion systems (T6SS) is a dynamic system that uses a needle like element for the translocation of proteins. They can secrete proteins into both eukaryotic and prokaryotic cells. However, bacterial cells are the more common target (Russell et al., 2014). They are used by bacteria for a number of purposes including invasion/protection of a community, signalling, and killing of non-co-operators and phage infected cells (Russell et al., 2014). They possess a structure that is believed to be derived from that of a phage tail (Pukatzki et al., 2007). The T6SS is assembled at the inner membrane in an extended conformation (Fig. 1.8). In a contact dependent process, this conformation contracts quickly to "fire" into the target cell where the effector proteins are released (Hood et al., 2010). After the T6SS has "fired" the complex is disassembled. These steps are repeated for successive "firing" of the T6SS (Basler et al., 2013). These system are very common in Gram-negative organisms and are thought to aid the virulence of many pathogens (Green and Mecsas, 2016).

1.4. Type I secretion systems

Composed of only three proteins an outer membrane protein, membrane fusion protein, and ATP binding cassette (ABC) transporter, the T1SS seems relatively simple compared to the multi complexes of the T2SS, T3SS, T4SS, and T6SS. The membrane fusion protein links the outer membrane protein to the inner membrane localised ABC transporter. The connection of all three creates a pore which allows the substrate to exit the cell in an energy dependent manner. However, unlike the other secretion systems, T1SS can secrete a vast range of

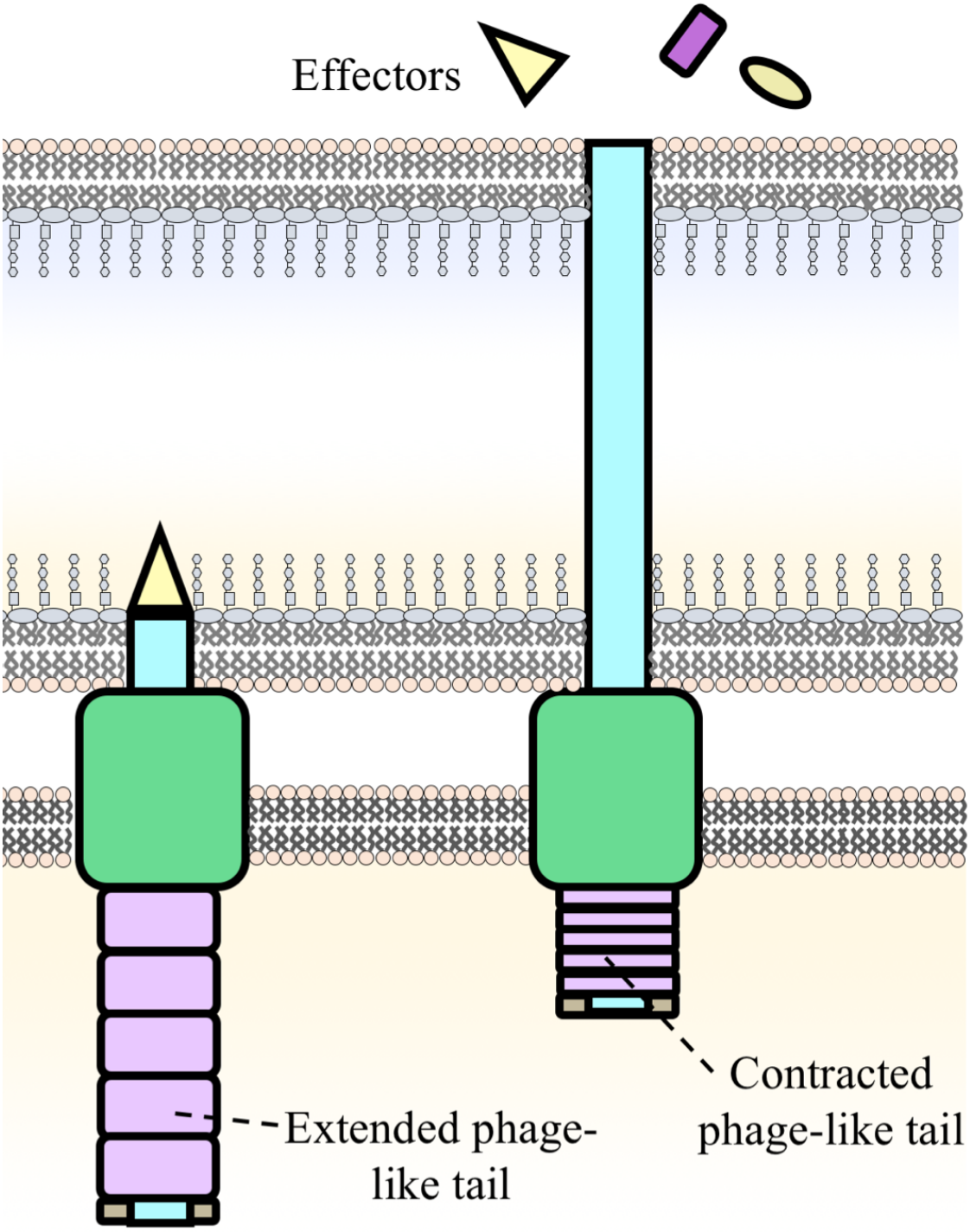


Figure 1.8: Type VI Sectrion System. The T6SS is similar to the phage tail. The T6SS forms in the cytoplasm. The phage-like tail contracts firing into the target cell. Effector proteins are then secreted directly into the target cell cytoplasm.

substrates from 100 to 9,000 amino acids including proteases, toxins, S-layer proteins, haemophores, lipases, and adhesins (Delepelaire, 2004).

Prior to the discovery of the T1SS the mechanism bacterial secretion was unknown. The first member of the T1SS was HlyBD-TolC, which is present in Uropathogenic *E. coli* (UPEC). HlyBD-TolC is required for the secretion of the pore forming toxin haemolysin encoded by *hlyA* (Welch, 1991). HlyA is part of the repeat in toxins (RTX) family which make up the majority of proteins secreted by the T1SS. Proteins in this family contain very similar repeats of nine residues that form a unique β-roll structure to bind Ca^{2+} ions with aspartate and glycine residues (Baumann et al., 1993). The RTX members remain unfolded in low Ca^{2+} environments such as the cytoplasm. The proteins can fold into their mature form at the higher Ca^{2+} concentration of the extracellular milieu (Chenal et al., 2009). HlyA is encoded in the four gene operon *hlyC*, *hlyA*, *hlyB*, and *hlyD* (Felmlee et al., 1985). HlyC is a cytoplasmic acyltransferase required for HlyA function but not secretion (Nicaud et al., 1985); whereas HlyB and HlyD are an ABC transporter and membrane fusion protein, respectively, that are required for secretion. The final part of the T1SS for HlyA is the outer membrane protein TolC, which is encoded separately on the chromosome as it is a promiscuous protein involved in efflux and secretion of many different substrates.

1.4.1. Mechanism of type I secretion systems

HlyB, which is the ABC transporter, is located in the inner membrane in a transient complex with the membrane fusion protein HlyD (Thanabalu et al., 1998). Both HlyB and HlyD recognise the signal sequence of substrate (HlyA). HlyA has been translated and remains unfolded in the cytoplasm (Fig. 1.9). HlyA is acylated at K564 and K690 by HlyC (Stanley et al., 1994). The C-terminus of HlyA contains the secretion signal sequence of around 50 amino

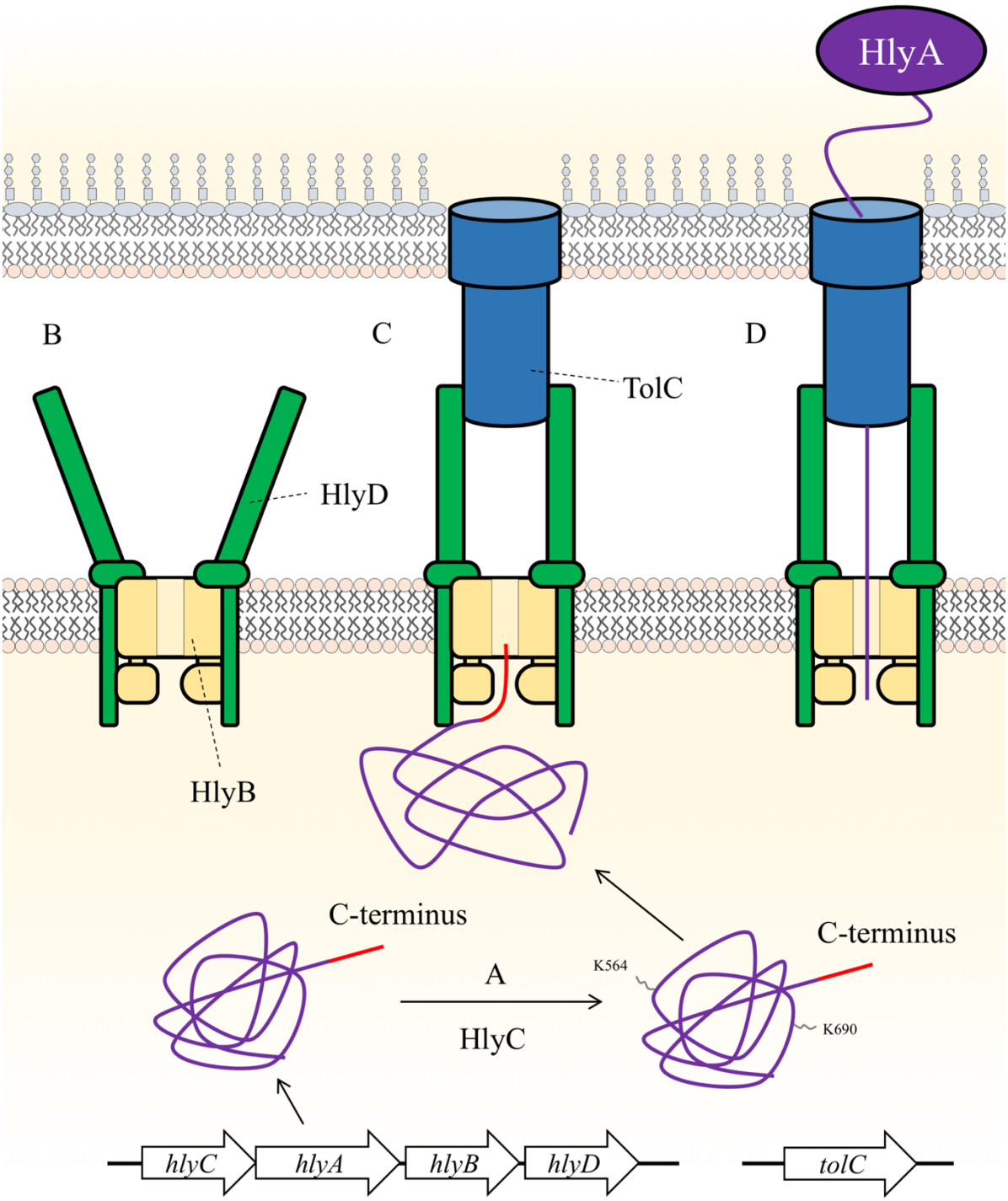


Figure 1.9: Secretion of HlyA by the prototypical T1SS HlyBD-TolC. A, the unfolded HlyA is acylated by HlyC at K564 and K690. This is required for function but not the secretion of HlyA. B, the HlyB and HlyD form a complex at the inner membrane. C, the C-terminus of HlyA is recognised by the HlyBD complex. This causes a conformational change that leads to the recruitment of TolC. D, HlyBD-TolC forms a single pore for the secretion of HlyA.

acids that targets it to the HlyBD complex (Koronakis et al., 1989). HlyB is composed of a transmembrane domain (TMD), a nucleotide binding domain (NBD), and a C39-like peptidase domain (CLD). HlyA interacts with both the CLD and NBD of HlyB and the cytoplasmic N-terminal region of HlyD. This causes a conformational change in HlyD leading to the recruitment of TolC (Balakrishnan et al., 2001). Interestingly the CLD contacts HlyA independent of the C-terminal secretion sequence (Lecher et al., 2012). The C-terminus of HlyA enters the HlyBD-TolC pore first (Lenders et al., 2015). The NBD of HlyB uses ATP to power the secretion of HlyA and the TMD is thought to provide the pore through the inner membrane for the secretion of HlyA. In the higher Ca^{2+} concentration environment of the extracellular milieu the RTX motifs are able to bind Ca^{2+} which enables the folding of HlyA as it exits the HlyBD-TolC pore.

Similar to the T1SS are the efflux pumps which require three proteins to export molecules across the cell membrane. They are comprised of: an inner membrane protein, which can be a resistance-nodulation-division (RND) transporter or an ABC transporter; a membrane fusion protein; and an outer membrane protein. Little information on the stoichiometry and protein-protein interactions of the T1SS and efflux pumps is known. For efflux pumps, the interaction of the resistance-nodulation-division (RND) transporter with the outer membrane protein and the point of contact of the membrane fusion protein with the outer membrane protein is still a matter of debate. The most studied efflux pump is that of AcrB (RND transporter), AcrA (membrane fusion protein), and TolC. Work by Du et al. (2014) using cryo-EM and structures of the efflux pump AcrAB-TolC suggest that there is no interaction of AcrB with TolC and that there is a slight overlap with AcrA and TolC (Fig. 1.10). The model showed a 3:6:3 stoichiometry for AcrB:AcrA:TolC. It has been suggested based on the interactions of AcrA and TolC that HlyD and TolC form a similar sealed channel (Holland et al., 2016). However,

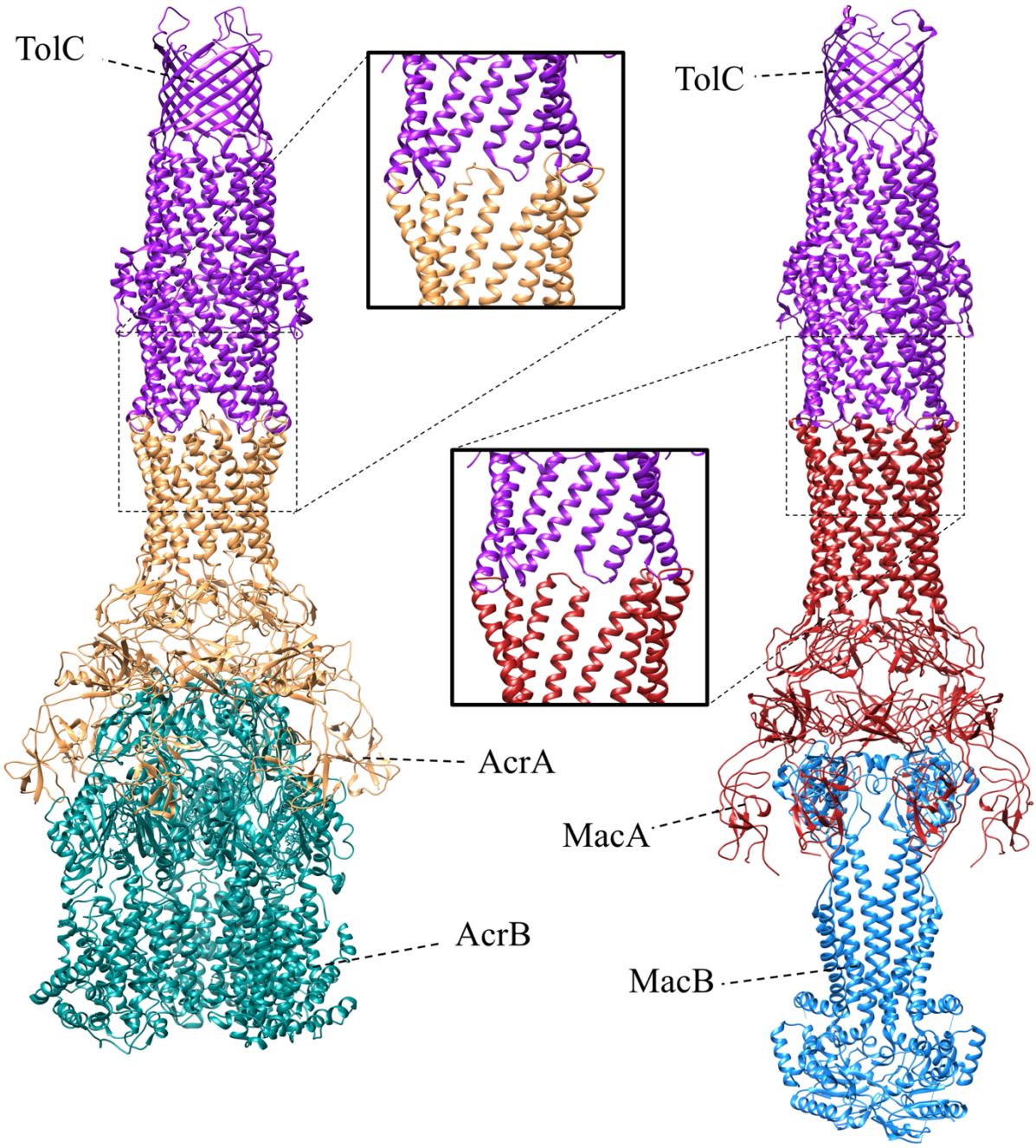


Figure 1.10: Cryo-EM structures of AcrAB-TolC and MacAB-TolC. The structures were downloaded from the 5NG5 and 5NIK PDB files for AcrAB-TolC and MacAB-TolC, respectively (Wang et al., 2017; Fitzpatrick et al., 2017). For the enlarged images, one TolC subunit and two membrane fusion subunits were removed to show the interaction between the TolC and membrane fusion proteins. Images were coloured in Chimera.

there can be no similarities drawn between the AcrB and HlyB as the functional region of AcrB is predominantly periplasmic while HlyB is cytoplasmic (Wang et al., 1991). A more relevant system to that of the T1SS maybe that of MacAB-TolC (Fig. 1.10), as instead of an RND transporter MacB is an ABC transporter (Modali and Zgurskaya, 2011). The stoichiometry of the complex was 2:6:3 for MacB, MacA, and TolC respectively (Fitzpatrick et al., 2017). MacB is a dimer similar to HlyB. The overlap overserved for AcrB and TolC was again observed for MacA-TolC (Fitzpatrick et al., 2017). While a comparison of MacAB-TolC complex to HlyBD-TolC is flawed, as the periplasmic domain of MacB is much larger than that of HlyB, it is currently the best model available for the T1SS.

1.4.2. ATP binding cassette transporters

ABC transporters use ATP to provide the energy to translocate a molecule across a membrane either as an importer or an exporter. They are ubiquitous throughout all domains of life with a large variety of substrates. There have been significant amounts of study devoted to the function of ABC transporters. An ABC transporter complex is comprised of two TMD and two NCB that enable the translocation of a target substrate against a gradient in an energy dependent manner. The sequence conservation of the TMDs is very low, however the NBD is typically conserved, as it contains motifs required for ATP hydrolysis such as Walker A, Walker B and C-loop (Walker et al., 1982; Ames et al., 1992). There is still debate over the mechanism of action for the ABC transporters. These mechanisms are typically focused on the secretion of small molecules with comparatively little investigation into the mechanism of those involved in T1SS.

In the T1SS, it is thought that the ABC transporter forms a pore in the inner membrane that allows the secretion of the unfolded substrate protein. Unlike other bacterial ABC transporters,

the TMD and NBD are encoded in the same gene in T1SS ABC transporters (Davidson et al., 2008). For this reason the ABC transporter of the T1SS are often referred to as "half-size transporters". Kanonenberg et al. (2013) split the T1SS ABC transporters into three groups based on their domains auxiliary to the TMD and NBD: C39 peptidases; C39-like peptidases (CLD); and neither. C39 peptidases are required for the secretion of bacteriocins. Bacteriocins are small peptides that are toxic to closely related bacteria. These are the smallest molecules secreted by the T1SS (Duquesne et al., 2007). Unlike the usual RTX the secretion signal sequence is present in the N-terminus of the substrate (Duquesne et al., 2007). The N-terminal leader peptide interacts the C39 peptidase to allow the cleavage of a GG motif in the C-terminus prior to mature protein release (Havarstein et al., 1995). Colicin V, which forms a pore in prokaryotic cells, is one such peptide secreted by these T1SS (Fath et al., 1994). Interestingly, unlike HlyBD-TolC, the colicin V secretion system, CvaA (ABC transporter), CvaB (membrane fusion protein), and TolC (outer membrane protein), can form a complex without the substrate (Hwang et al., 1997). The second group are the CLD ABC transporters which have an N-terminal C39-like domain. This domain is similar to that of the C39 peptidases however the catalytic residue is mutated (Lecher et al., 2012). For example, in the case of HlyB the catalytic cysteine of the C39 peptidases is mutated to a tyrosine. The third group contain neither a C39 peptidase or a CLD. While the substrates of this class can have RTX motifs, they typically do not, and are sustainably smaller than the proteins of the RTX family (Delepelaire, 2004). The most well studied example of this group of ABC transporters is HasD. HasD in conjunction with HasE (membrane fusion protein) and HasF (outer membrane protein) is responsible for the secretion of HasA, a small (19 kDa) haemophore from *Serratia marcescens* (Letoffe et al., 1994). Curiously, unlike the substrates of the RTX family, HasA will fold in the cytoplasm without the chaperone SecB thus preventing secretion (Debarbieux and Wandersman, 2001; Sapriel and Delepelaire, 2003).

1.4.3. Membrane fusion proteins

There is no evidence to suggest that the membrane fusion protein can fuse membranes in any way. Membrane fusion proteins were so named as they are similar to viral fusion proteins with the misnomer still persisting in the field (Dinh et al., 1994). The membrane fusion proteins of the T1SS are not as well investigated as their efflux pump counter parts, however it is clear that the N-terminal domain of HlyD extends into the cytoplasm and interacts with HlyA to recruit TolC. It appears that this extension of the N-terminus into the cytoplasm maybe common of T1SS membrane fusion proteins (Holland et al., 2016). The membrane fusion proteins of the efflux pumps are typically lipoproteins attached to the inner membrane. Other membrane fusion proteins such as EmrA also have a N-terminal transmembrane helix but, unlike HlyD, do not have a cytoplasmic domain (Borges-Walmsley et al., 2003). The recent structure of the periplasmic region of HlyB was very similar to that of EmrA (Kim et al., 2016). The authors identified a long α helical extension into the periplasm as well as a lipoyl domain that was not identifiable by sequence homology. A lipoyl domain is known to alter the function of a number of enzymes, typically fusing two halves together. An analogous membrane fusion protein AcrA has a lipoyl domain and forms a hexamer that creates a ring and is the narrowest point in the AcrAB-TolC pump (Kim et al., 2015). HlyD itself has been identified as a trimer and a hexamer (Thanabalu et al., 1998; Lee et al., 2012). A hexamer of the HlyD structure was found to be superimposable onto AcrA lipoyl domains forming a similar ring (Kim et al., 2016). The periplasmic region of HlyD is very flexible (Kim et al., 2016); this is true of all membrane fusion protein structures thus far determined. This is likely due to the function of the membrane fusion protein to connect proteins in the inner and outer membrane (Holland et al., 2016).

1.4.4. T1SS outer membrane proteins

There are many different outer membrane proteins that function as part of a T1SS. TolC has a bifunctional role in the T1SS and efflux pumps. Gram-negative bacteria typically contain a gene for *tolC* or a closely related homolog. The most important function of TolC appears to be in efflux with AcrAB and other efflux pumps. The deletion of *tolC* renders *E. coli* sensitive to low concentrations of detergents (Whitney, 1971). Interestingly, only a small fraction of 1,500 TolC copies are required to maintain efflux in *E. coli* (Krishnamoorthy et al., 2013; Tikhonova and Zgurskaya, 2004). This leaves TolC to interact with other systems outside of efflux.

The structure of TolC was determined to be a trimer consisting of a 12 stranded anti-parallel β-barrel inserted into the outer membrane with 12 α-helices extending into the periplasm (Fig. 1.11) (Koronakis et al., 2000). The TolC trimer forms an open water filled barrel and towards the periplasmic end this barrel tightens to 3.9 Å (Koronakis et al., 2004). This periplasmic closing means that TolC does not require a plug like many other β-barrel proteins found in the outer membrane. TolC is proposed to open by an "iris-like" mechanism which is mediated by the membrane fusion proteins (Andersen et al., 2002). This involves a twisting of the α helices to allow the periplasmic opening to widen to a diameter of around 20 Å (Andersen et al., 2002). This mechanism has been shown to be required for HlyA secretion (Eswaran et al., 2003).

The TolC protein is unusual for a T1SS outer membrane protein as other similar proteins do not appear to have the same promiscuity. However, there is very little investigation into the structure and mechanism of other T1SS outer membrane proteins. The majority of the knowledge about these pores is from analogous efflux pump outer membrane proteins. From the structures of other efflux pump outer membrane proteins the overall fold is very similar (Fig. 1.11). Characterised by a β-barrel with α helices extending into the periplasm.

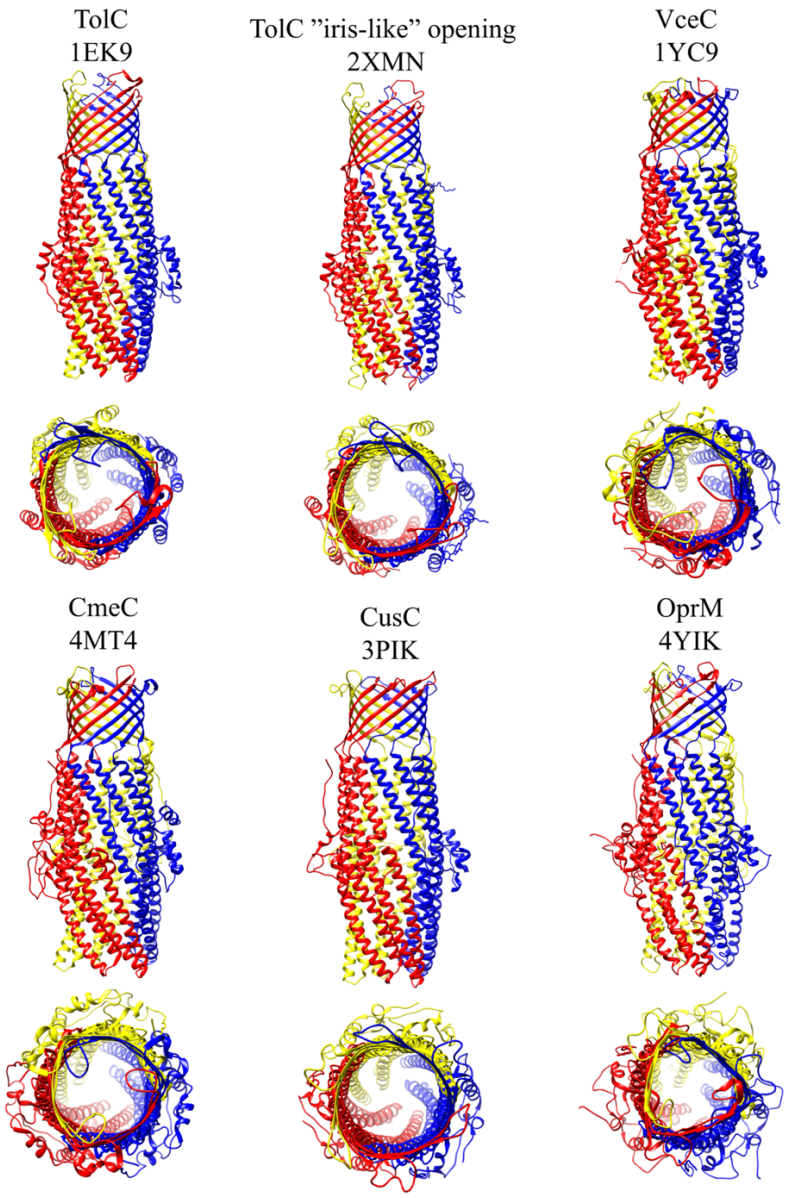


Figure 1.11: Structures of outer membrane proteins involved in efflux. The monomers in each structure are coloured separately. The PDB file for each structure is below the name of the protein. Structures are presented in profile with pore they form underneath. TolC is a versatile protein involved in both efflux and T1SS in *E. coli*. VceC, CmeC, and OprM form multidrug efflux pumps in *Vibrio cholera*, *Campylobacter jejuni*, and *Pseudomonas aeruginosa*, respectively. In *E. coli* CusC is required for the efflux of copper and silver ions. The structures were solved, from top left to bottom right, by Koronakis et al. (2000), Andersen et al. (2002), Federici et al. (2005), Su et al. (2014), Kulathila et al. (2011), and Monlezun et al. (2015).

Interestingly not all conform to the "iris-like" opening observed for TolC. VceC is an outer membrane efflux protein in *Vibrio cholerae* (Colmer et al., 1998). The structure of VceC was determined by Federici et al. (2005), who showed that the "iris-like" mechanism was not applicable to VceC. Instead, the authors showed a straightening of one of the helical pairs in the periplasmic portion of the protein was responsible for the opening of the channel. The functions and structures of TolC and VceC are similar but their mechanisms of action are different. Suggesting that while the fold of other T1SS outer membrane proteins maybe similar their mechanism may have diverged.

1.4.5. Atypical type I secretion systems

The paradigm of HlyA secretion by HlyBD-TolC is not true for all substrates of the T1SS. Perhaps due to the simplicity of the system it is easily malleable. There are very few rules that can extend across the entirety of the T1SS. The secretion of colicins and HasA does not follow the prototypical mechanism for T1SS but the construction of their respective translocon is consistent. That is to say, the proteins required for secretion are an ABC transporter, a membrane fusion protein and an outer membrane protein. However, other T1SS have evolved from this paradigm.

A slight variation on the common RTX repeat is present in a different group of substrates. The multifunctional-autoprocessing repeats in toxins (MARTX) are giant proteins, which can be up to 9,000 residues in length, that are multifunctional for pathogenic bacteria (Satchell, 2011). These proteins are like simple AB toxins, in that they contain both the cytotoxic element and mechanism to bind host cells. However, instead of a single effector, MARTX proteins contain multiple effectors in the same protein which are analogous to the effectors of the T3SS. Interestingly, the species encoding these MARTX proteins do not have a T3SS (Gavin and

Satchell, 2015). The MARTX proteins and their secretion system are encoded in two proximal but divergent operons (Lin et al., 1999). The first operon encodes three proteins: RtxH, a protein of unknown function; RtxC, an acyltransferase that does not appear to be required for toxin function; and RtxA the MATRX toxin (Lin et al., 1999). While the substrate itself is atypical the major difference in this system is the organisation of the translocon. The second operon encodes three other proteins: RtxB, an ABC transporter; RtxD, a membrane fusion protein; and RtxE, a second ABC transporter. The outer membrane protein used by this system is TolC. RtxB and RtxE are thought to form a heterodimer that is required for the secretion of the RtxA substrate (Boardman and Satchell, 2004). The mechanism and requirement for two ABC transporters for the secretion of RtxA is as yet unknown.

Another atypical T1SS is for the secretion of another large substrate SiiE. SiiE (595 kDa) is an adhesion in *Salmonella enterica*. SiiE does not contain the RTX motif but does require Ca^{2+} to fold it's 53 IgB-like repeats (Griessl et al., 2013). SiiE contains these many repeats so that it can extend past the *S. enterica* LPS (Gerlach et al., 2008). The secretion system of SiiE and the protein itself are encoded in a single operon, *siiABCDEF*. SiiC, SiiD, and SiiF function as the outer membrane protein, membrane fusion protein, and ABC transporter, respectively (Morgan et al., 2004). SiiA and SiiB are inner membrane proteins and use proton motive force to facilitate SiiE secretion (Wille et al., 2014). However, why these proteins are required for SiiE secretion and how they interact with the SiiCDF is not known.

The final example of an atypical T1SS is the Aat system. The Aat system is composed of five proteins, AatPABCD, that are required for the secretion of dispersin (Aap) in enteroaggregative *E. coli* (EAEC) (Nishi et al., 2003). While Aap has been well characterised, which will be discussed further in this chapter, its homolog CexE and the mechanism of Aat secretion has

not. CexE is a small protein encoded in enterotoxigenic *E. coli* (ETEC). EAEC and ETEC with other pathotypes, make up a large group of the *E. coli* species that have acquired pathogenic determinants.

1.5. Pathogenic *Escherichia coli*

The products secreted by these systems have a myriad of functions and mechanisms with many of them used by pathogenic bacteria to gain an advantage in the host. The majority of *E. coli* strains are mammalian gut commensals, however there are many strains that can cause a variety of disease including gastroenteritis, urinary tract infections, and meningitis. These pathogenic *E. coli* are grouped into pathovars, which do not follow evolutionary lineages but are instead dictated on their mechanism of pathogenesis (Wirth et al., 2006). The best described pathovars fall into two groups, diarrheagenic *E. coli* (DEC) and extraintestinal pathogenic *E. coli* (ExPEC). DEC is made up of the enteropathogenic *E. coli* (EPEC), enteroheamorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), ETEC, and EAEC. Other DEC pathovars exist but they are not as well characterised as these five. The best characterised *E. coli* pathovars will be briefly described with a focus on ETEC and EAEC.

ExPEC infect regions outside of the GI tract, the most notable member of this group is the uropathogenic *E. coli* (UPEC) (Fig. 1.12). UPEC are asymptomatically carried in the intestine. UPEC cells gain access to the urethra where they attach to epidermal cells using the fimbrial adhesin, FimH (Martinez et al., 2000). Binding triggers an actin remodelling resulting in the internalisation of the UPEC cell (Eto et al., 2007). Upon invasion, UPEC cells can form intracellular bacterial communities (IBC) that protect them from the host immune system and allow recurrent infection (Anderson et al., 2003). Motile UPEC cells can leave the IBC and enter into the lumen of the bladder to cause further infection (Justice et al., 2004). To avoid the

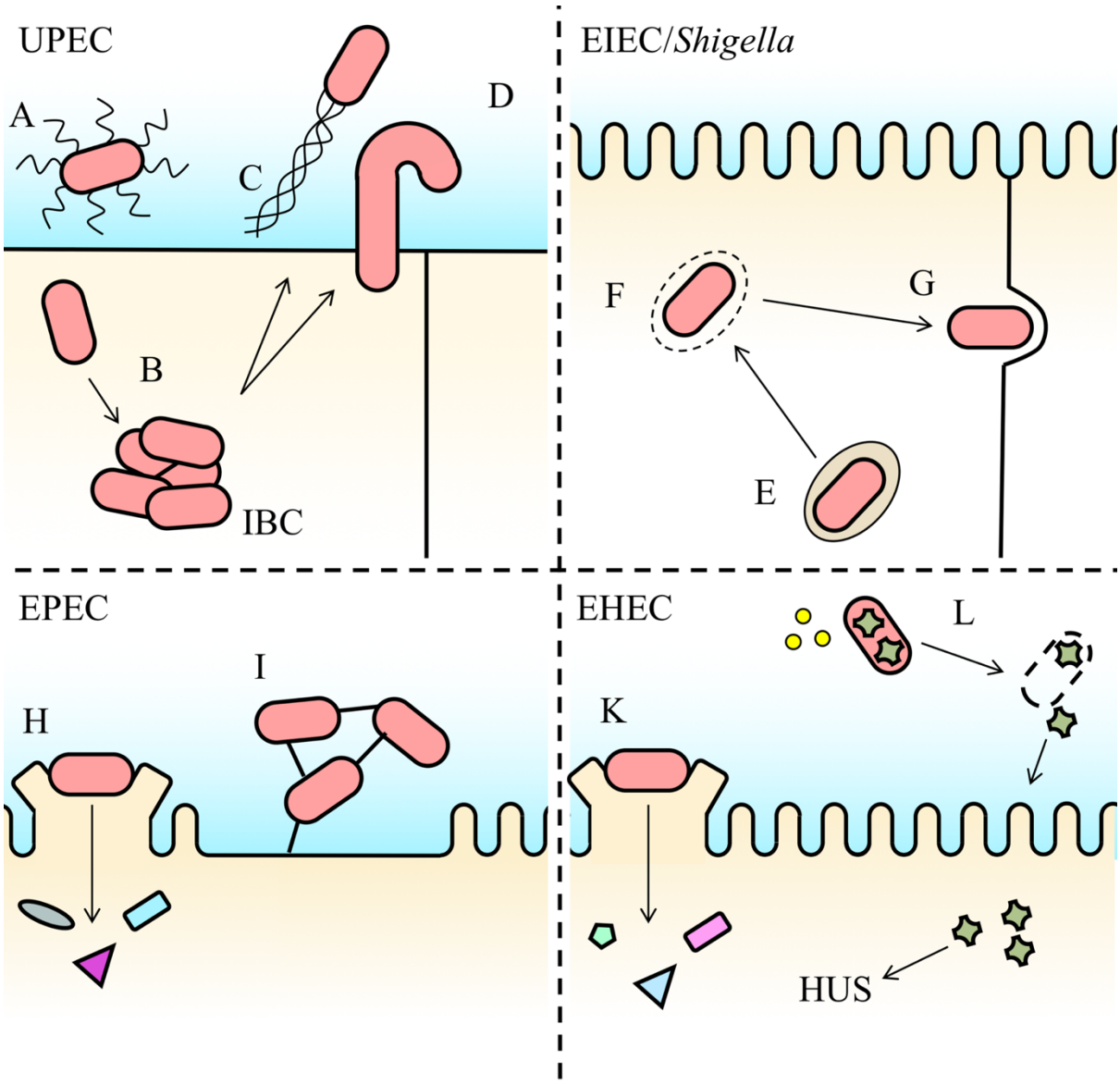


Figure 1.12: Major features of different pathovars pathogenesis. A, UPEC cells use FimH to attach to host cells in the urinary tract. B, internalised cells form intracellular bacterial communities (IBC). C, the UPEC cells spread using motility and filamentous growth. E, EIEC cells enter enterocytes via the basolateral membrane. F, they escape the phagosome using T3SS. G, they can spread between cells using actin polymerisation. H, EPEC cells create attaching and effacing lesions using T3SS. I, EPEC cells use bundle forming pili to create microcolonies. K, EHEC cause attaching and effacing lesions similar to EPEC but secretes different T3SS effectors. L, antibiotics kill EHEC cells containing the Shiga toxin. The toxin is released from the cells and enters the host causing haemolytic uraemic syndrome.

innate immune system, UPEC cells are capable of filamentous growth that allows the infection of nearby cells (Justice et al., 2006). In response to the immune system UPEC cells secrete haemolysin A (HlyA) using the previously described T1SS. The secretion of HlyA leads to host epithelial exfoliation, epithelial cell apoptosis, and suppression of inflammatory responses (Wiles et al., 2008; Dhakal and Mulvey, 2013).

The ExPEC do not represent the majority of *E. coli* pathovars, which are most commonly associated with the digestive tract. Unlike the other *E. coli* pathovars that will feature below, the EIEC include obligate intracellular bacteria that are non-motile nor have adherence factors. It is typically considered that EIEC and *Shigella* should form a single pathovar due to their similar mechanisms of pathogenesis and genetic relatability. However, *Shigella* is still retained due to the association with shigellosis. EIEC/*Shigella* transverse the epithelium via microfold (M) cells. They are taken up by macrophages. After replicating and causing apoptosis of the macrophage, EIEC/*Shigella* invade the basolateral side of epithelial cells (Fig. 1.12). These stages of infection are accomplished by the Mxi-Spa type III secretion system (T3SS) encoded on a virulence plasmid, pINV, carried by all members of this pathovar (Schroeder and Hilbi, 2008; Hale, 1991). The Mxi-Spa T3SS is required for the secretion of at least 25 effector proteins involved modulating many systems in the host cell (Buchrieser et al., 2000). The host cell pathways targeted by the effectors include invasion, phagosome escape, cell survival, intracellular motility and intercellular dissemination (Schroeder and Hilbi, 2008). While cell invasion by *E. coli* pathogens is not common the use of a T3SS is not limited to the EIEC/*Shigella* pathovar.

EPEC and EHEC use a T3SS, which is encoded in the locus for enterocyte effacement (LEE), for the formation of attachment and effacing lesions on intestinal epithelial cells (McDaniel et

al., 1995). These lesions are characteristic for a group of pathogens that includes, as well as EPEC and EHEC, rabbit diarrhoeagenic *E. coli* (RDEC) and *Citrobacter rodentium*. EPEC and EHEC are similar in their method of lesion formation but they differ on other aspects of pathogenesis. EPEC form micro colonies via the bundle forming pilus that is strengthened by the *E. coli* common pilus (Saldana et al., 2009). EPEC use the T3SS effector Tir and intimin to mediate close interaction with the host cell (Fig. 1.12). Tir is translocated and presents on the host cell membrane allowing attachment by the outer membrane protein intimin (Kenny et al., 1997). Other effectors secreted by the T3SS into the host cell cause cytoskeletal rearrangement, modulate the immune system, and aide the causation of diarrhoea (Croxen and Finlay, 2010).

EHEC strains are a commensal of cattle in agriculture. Like EPEC strains, EHEC causes attachment and effacing lesions with the use of Tir and intimin (Fig. 1.12). The main difference between EHEC and EPEC strains is the Shiga toxin. The Shiga toxin can increase the amount of nucleolin on the host cell which is then used by EHEC cells to adhere via intimin (Robinson et al., 2006). The Shiga toxin is responsible for the causation of haemorrhagic fever and haemolytic uraemic syndrome, which can lead to kidney failure. The Shiga toxin is AB_5 toxin that interestingly does not have a secretion system. Instead the toxin is released due to lysis of the cells by the lambdoid phage due to DNA damage (Toshima et al., 2007).

1.6. Enteroaggregative *E. coli*

EAEC were first described by Nataro et al., (1987) as forming a "stacked-brick" aggregative adherence pattern on the human laryngeal carcinoma (HEp-2) cells. The formation of "stacked-brick" biofilms is the gold standard the detection of EAEC strains (Fig. 1.13). EAEC strains are transmitted by the faecal-oral route and the main site of EAEC colonisation is the

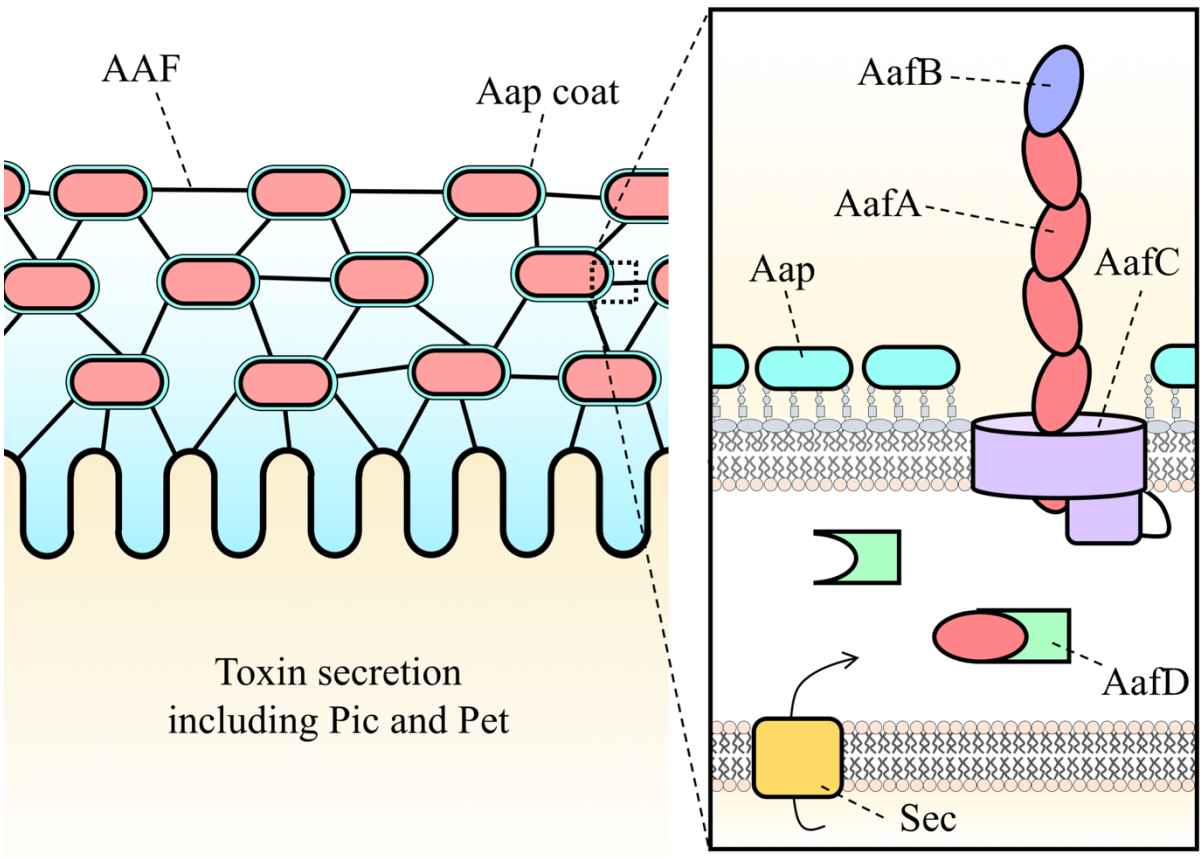


Figure 1.13: EAEC pathogenesis. Upon an increase of nutrient availability due to entering the GI tract Fis is produced. Fis activates the expression of *aggR* which activates the expression of the AAF and *aap*. EAEC cells use the AAF and Aap to form a "stacked-brick" biofilm from which they can release toxins that cause disease in the host. This leads to mucoid diarrhoea. Critical to the biofilm formation are the AAF. These are encoded by four genes *aafA*, *aafB*, *aafC*, and *aafD*. The AAF is produced using the chaperone-usher pathway. The proteins are secreted into the periplasm. AafC inserts into the outer membrane. The minor subunit AafB is at the tip of the fimbriae. The AafA form the main body of the organelle. AafD is the chaperone protein that transports the fimbrial subunits to AafC. In this manner, the fimbriae are extended. The negative charge of the outer membrane is counter acted by Aap. Aap binds to the outer membrane allowing the positively charge AAF to extend away from the cell. The genes encoding these proteins are

gastrointestinal tract (Hicks et al., 1996). EAEC strains are associated with an acute or persistent watery mucoid diarrhoea, with a low-grade fever, and little to no vomiting (Nataro and Kaper, 1998). Some outbreaks of EAEC are associated with bloody diarrhoea due the acquisition of the Shiga-toxin (Rohde et al., 2011). EAEC strains have also been suggested as the cause of some urinary tract infections (Boll et al., 2013). Since the early observations of EAEC pathogens they have emerged as a pathogen of global importance. Early in EAEC research, EAEC strains were considered to only be endemic in developing countries infecting children and travellers in these regions (Okeke et al., 2000; Adachi et al., 2001). However, recent evidence indicates that EAEC strains are becoming a common cause of food poisoning in industrialised nations (Clements et al., 2012).

The characteristic "stacked-brick" pattern of EAEC strains is different from many other *E. coli* pathogens. This is commonly referred to as aggregative adherence. The proteins required for aggregative adherence are encoded on a large plasmid, which is termed pAA. The pAA plasmid contains the majority of the virulence genes with a few pathogenicity islands present on the chromosome of most EAEC strains. The majority of the molecular characterisation of EAEC pathogenesis has been investigated with the strain EAEC 042. This strain was chosen as the prototypical EAEC strain due to its ability to cause diarrhoea in volunteer studies (Nataro et al., 1995). EAEC strains elicit disease by attaching to the intestinal cells using the aggregative adherence fimbriae (AAF) which are responsible for the characteristic aggregative adherence pattern. Once attached they form a biofilm where the cells secrete toxins that elicit host cell damage. The AAF appear to be the main mechanism for EAEC biofilm formation as other factors like curli, flagella, and Antigen 43 are not required (Sheikh et al., 2001). The toxins secreted by EAEC include plasmid encoded toxin (Pet) which is a T5SS serine protease that degrades of the host cell cytoskeleton (Canizalez-Roman and Navarro-García, 2003). Another

EAEC toxin, Pic, is associated with a number of roles including mucinase activity, serum resistance, and haemagglutination (Henderson et al., 1999). Pic has also recently been associated with the evasion of complement mediated killing (Abreu et al., 2015). The secretion of these toxins induces an inflammatory response leading to mucoid diarrhoea.

The AAF are the main mechanism by which EAEC attach and are able to cause disease. The AAF consist of 4 proteins. The major subunit and minor subunits are secreted proteins. The major subunits form a linear strand using donor strand complementation with the minor subunit as a cap on the linear strand (Berry et al., 2014). The AAF extend from the cell attaching to other cells and to host epithelial cells (Czeczulin et al., 1997). The AAF mediate attachment to host by the electrostatic binding of fibronectin (Berry et al., 2014). The genes encoding the AAF are activated by another important factor in EAEC pathogenesis, AggR.

1.6.1. AggR

One of the genes typically encoded on the pAA plasmid is *aggR*. AggR is required for the activation of expression of virulence genes in EAEC. Due to the importance of AggR, EAEC are broadly characterised as typical and atypical EAEC by the respective presence or absence of *aggR* (Nataro, 2003). AggR is a member of the AraC family of promoters (Gallegos et al., 1997). The members of this family are characterised by an 100 amino acid sequence that contains two helix-turn-helix motifs at the C-terminus which typical bind upstream of the promoters to activate expression (Egan, 2002). These family members contain two domains an N-terminal multimersation or co-factor binding domain and the C-terminal DNA binding domain that joined by a flexible linker (Mahon et al., 2010).

AggR is closely related to CfaD, RegA, and PerA which are all virulence regulators in ETEC,

C. rodentium, and EPEC, respectively. These are mostly responsible for the activation of transcription. The initial activation of AggR expression is by Fis (Sheikh et al., 2001). After initial production AggR is capable of autoactivation (Morin et al., 2010). Like the other virulence activators AggR is repressed by Aar (Fig. 1.14), which is a member of the AraC family of negative regulators (Santiago et al., 2014). Aar is able to bind AggR and prevent its dimerisation which is required for activation of gene expression (Santiago et al., 2016). AggR activates the expression of its own repressor.

The AggR regulon in EAEC 042 contains 44 genes (Morin et al., 2013). The genes that AggR activates the expression of encode proteins known to be required for EAEC 042 pathogenesis such as AAF. The genes that are activated by AggR are typically on the pAA plasmid. However, interestingly a number of pathogenicity islands on the EAEC 042 chromosome were also identified (Morin et al., 2013); including a T6SS of which the function is currently unknown (Dudley et al., 2006). Advances in transcriptome profiling techniques like RNA-Seq have identified more genes under the regulatory control of the related regulator, CfaD (Hodson et al., 2017). The results of the CfaD profiling suggest that the complete view of the AggR regulon in EAEC 042 may not be complete.

1.6.2. Dispersin

One gene known to be activated by AggR is *aap*. The *aap* gene encodes a small extracellular protein of 12.5 kDa with an N-terminal signal sequence, that is cleaved at residue 21 to form a 10.2 kDa mature protein. Aap ensures the correct orientation of the AAF (Sheikh et al., 2002). Insertional mutants of *aap* aggregate at a faster rate than the wild-type (Sheikh et al., 2002). When the *aap* mutants were inspected by electron microscopy the AAF were seen to be lying of the surface of the cell unlike the wild-type in which they were fully extended (Sheikh et al.,

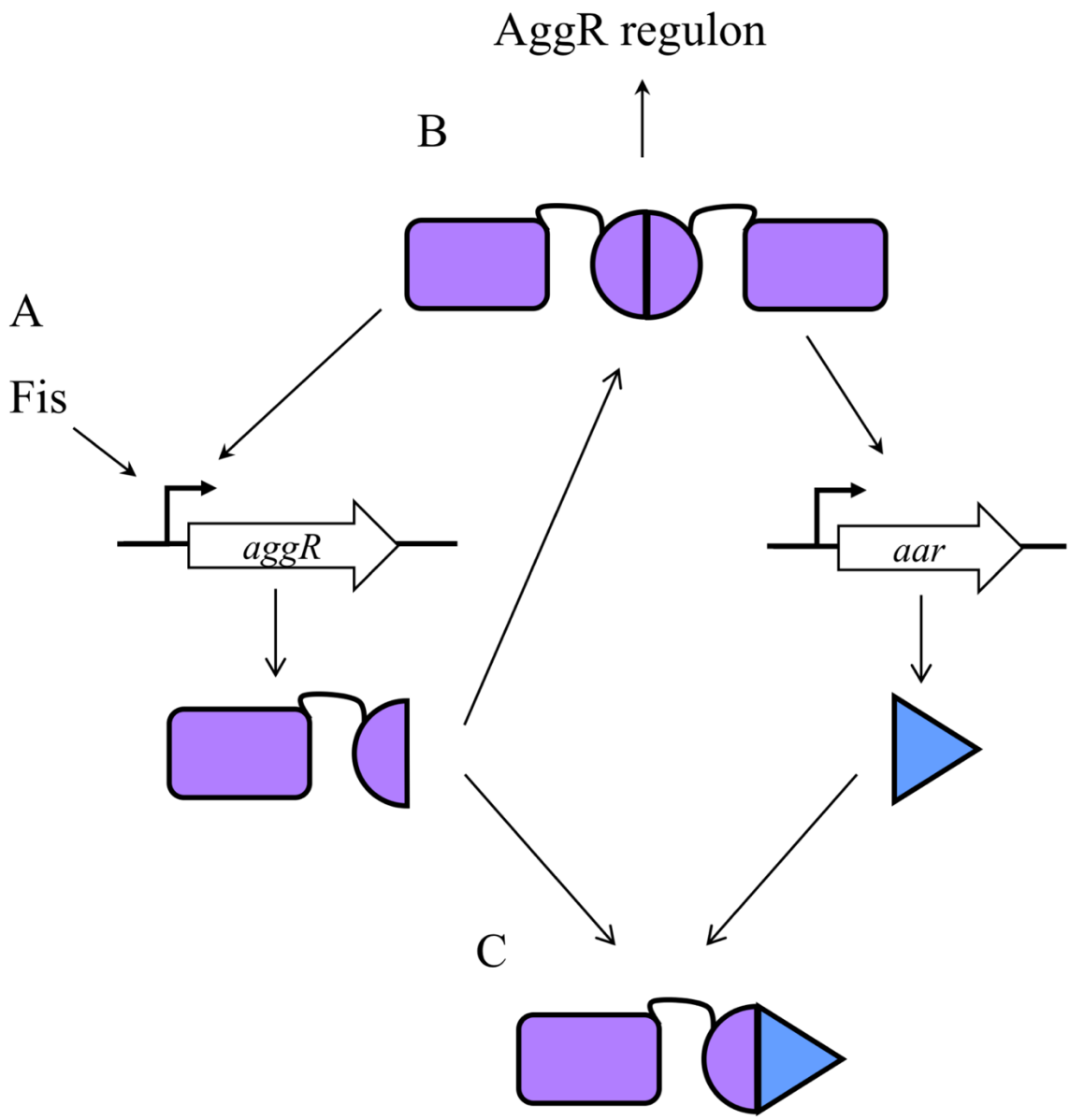


Figure 1.14: Regulation of AggR by Aar. A, Fis is transcribed in response to a change in nutrients when entering a new environment, such as ingestion into the mammalian gut. Fis activates the expression of *aggR*. The AggR protein consists of a C-terminal DNA-binding domain and an N-terminal signaling domain. B, AggR dimerises and activates the expression of proteins in its regulon. This includes activation of itself and of a regulator protein Aar. Aar binds to the N-terminal domain of AggR preventing it from dimerising and activating the expression of genes in its regulon.

2002). Aap does not form part of the AAF organelle (Sheikh et al., 2002). Instead it binds the outside of the outer membrane electrostatically thus altering the charge and allowing the positively charged AAF to extend from the cell (Velarde et al., 2007). The pI of the major subunit of the AAF is high, ranging from 8.92 to 9.74 therefore at physiologically pH ranges they are positively charged (Velarde et al., 2007). The correct orientation of the AAF is ensure by the modulation of the cell surface charge by Aap which coats the outside of the cell (Sheikh et al., 2002; Velarde et al., 2007). The main cause of the negative charge on the outer membrane is the LPS which Aap is able to bind (Velarde et al., 2007).

Fascinatingly, *aap* mutants are more resistant to ciprofloxacin, which inhibits DNA-gyrase and topoisomerase IV preventing DNA synthesis (Mortensen et al., 2011). In fact, the production of Aap increases the uptake of ciprofloxacin (Mortensen et al., 2013). The effect of Aap on the uptake of other fluoroquinolones is unknown. The uptake of ciprofloxacin is normally associated with outer membrane porins like OmpF (Hirai et al., 1986). This potentially suggests a disruption in the outer membrane caused by Aap. However, the mechanism by which an increased amount of ciprofloxacin is taken up by Aap producing cells remains cryptic.

Aap is not limited to EAEC strains. Sequencing has identified homologous genes to *aap* in aEPEC, *C. rodentium* ICC168 and ETEC (Srikhanta et al., 2013; Petty et al., 2010; Crossman et al., 2010). These strains also encode an atypical T1SS, the Aat system, which is required for the secretion of Aap.

1.6.3. Aat system

The secretion of Aap is dependent on an atypical T1SS composed of five proteins AatPABCD (Nishi et al., 2003). This system is drastically different from those of the other atypical T1SS

as Aap is first secreted into the periplasm by the Sec pathway (Fig. 1.15). Comparatively little is known about the Aat system, with very little characterisation of any of the proteins involved. Nishi et al. (2003) identified AatP as the transmembrane permease, AatA as a TolC homolog, and AatC as a ATPase. The authors went further and predicted the AatA structure based on that of TolC. After Nishi et al. (2003) first showed that AatA was a TolC homolog, Iwashita et al. (2006) confirmed this by mutagenesis. They showed that F381, L382, and L383 were required for the function of AatA. Single mutations could still secrete but the mutant of all three could not. They showed that it was important that the residues at 382 and 383 were non-polar and suggested this would be where the membrane fusion protein would interact. These mutants were in the equatorial domain of the predicted AatA structure, which in TolC is known to be important for the interactions with the membrane fusion proteins. TolC is known not to be required for the secretion of Aap but does contribute to aggregation of EAEC 042 (Imuta et al., 2008). A possible function of AatB and AatD was not identified.

After the first investigations into the Aat system by Nishi et al. (2003) there was no further understanding of the mechanism of Aap secretion. Recently the advances in sequencing technology have identified Aat homologs in a variety of strains and pathovars. Interestingly the probe used to identify EAEC strains by DNA hybridisation, CVD432, anneals to the *aat* region of the pAA2 plasmid (Jensen et al., 2014). While known to be in ETEC and EAEC, the Aat system was also observed in NCTC86, the strain first isolated by Theodor Escherich (Dunne et al., 2017). A CRP binding site has been identified in ETEC H10407 was identified at the end of the *aatC* gene (Haycocks and Grainger, 2016). This site was the promoter for a gene that encoded a small protein of 62 amino acids with homology to the DUF1602, which are genetically associated with genes encoding transport genes (Haycocks and Grainger, 2016).

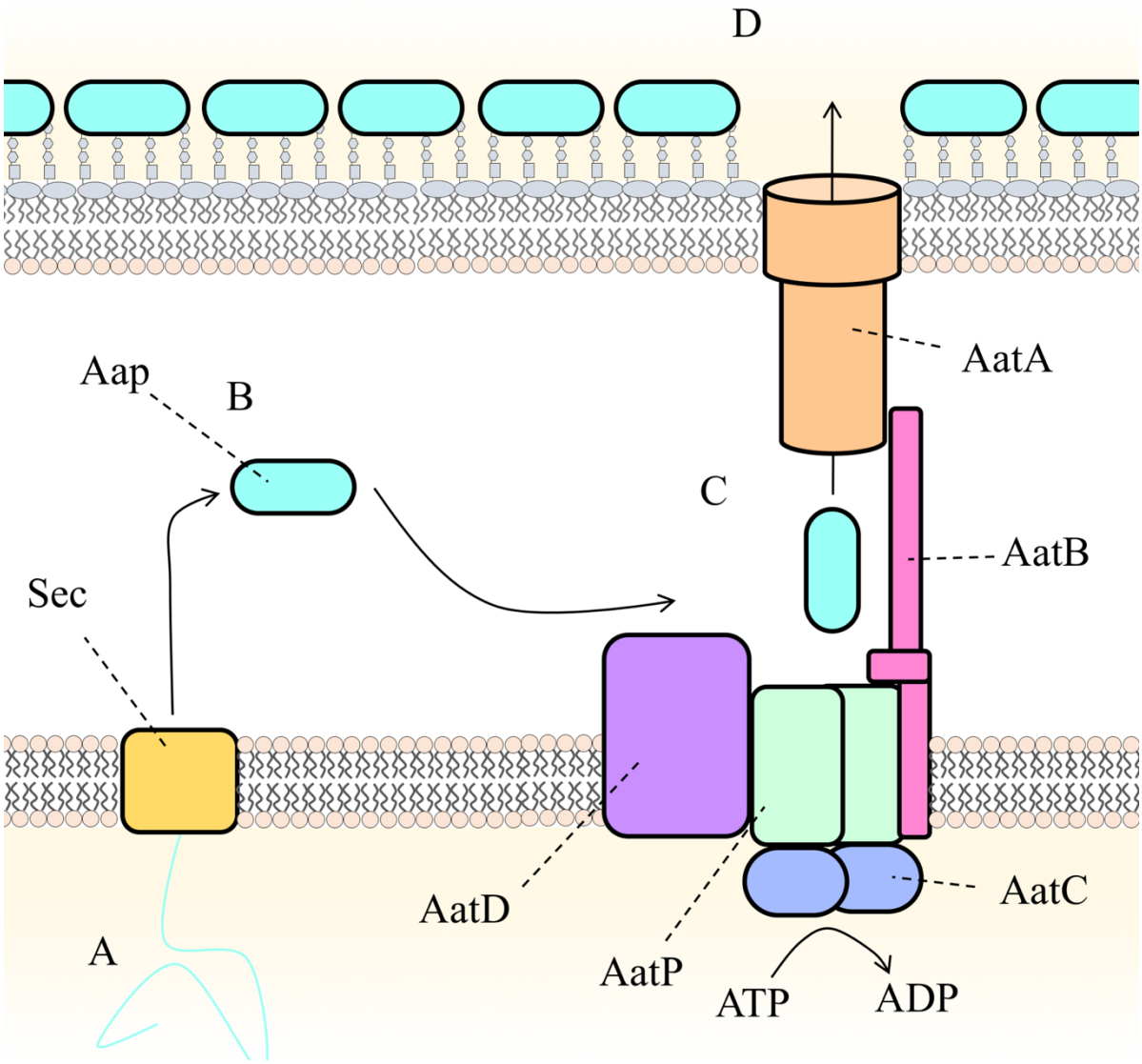


Figure 1.15: Alteration of the mechanism of Aat action proposed by Nishi et al. (2003). A, the nascent Aap peptide is secreted into the periplasm by the Sec machinery. B, the protein folds in the periplasm. C, in an as yet not understood mechanism Aap enters the Aat pore. D, Aap is secreted into the extracellular milieu. Aap attaches to the negatively charged LPS. This alters the charge allowing the extension of the AAF. Nishi et al. (2003) did not suggest that AatB was a membrane fusion protein. However, AatB is commonly listed as a membrane fusion protein in protein databanks. AatP is the transmembrane protein. AatC is the ATPase. AatA is the outer membrane protein. The function of AatD is not known.

The authors speculated that this may have a function in Aat-mediated secretion of Aap, however this ORF is not conserved in other *aatC* genes.

There has been no investigation into the function of these proteins other than they are known to be required for Aap secretion. Why an atypical secretion system is required for a small protein like Aap, which does not have any known catalytic activity, when the other examples of atypical secretion are large multi-domain proteins is not understood. There are still many questions left to be answered about the mechanism and requirement of this atypical T1SS which has been identified in many pathovars including ETEC.

1.7. Enterotoxigentic *E. coli*

ETEC is the pathovar most associated with traveller's diarrhoea and military personnel abroad (Nada et al., 2013). ETEC is the most commonly isolated enteric pathogen accounting for more than 210 million cases and 380,000 deaths each year, predominately in children under the age of 2 years in developing countries (Steffen et al., 2005). The majority of ETEC infections are acquired from contaminated water. ETEC colonises the small intestine as there is little competition from commensal bacteria, which are predominantly present in the colon (Swidsinski et al., 2005).

Once ETEC cells are ingested they must attach to host epithelial cells to cause disease (Fig. 1.16). ETEC strains are remarkable due to the huge variability in the adhesins they used to adhere to host cells. These are adhesins in ETEC are termed colonisation factors (CFs). ETEC CFs are divided into four main groups: fimbrial (pilus); fimbrillar, helical and afimbrail. Pili CFs, e.g. CFA/I which is also known as CS1 in ETEC H10407, are more ridged compared to the more flexible fimbrillar structures, e.g. CS3 in E24377A. The helical structures are two

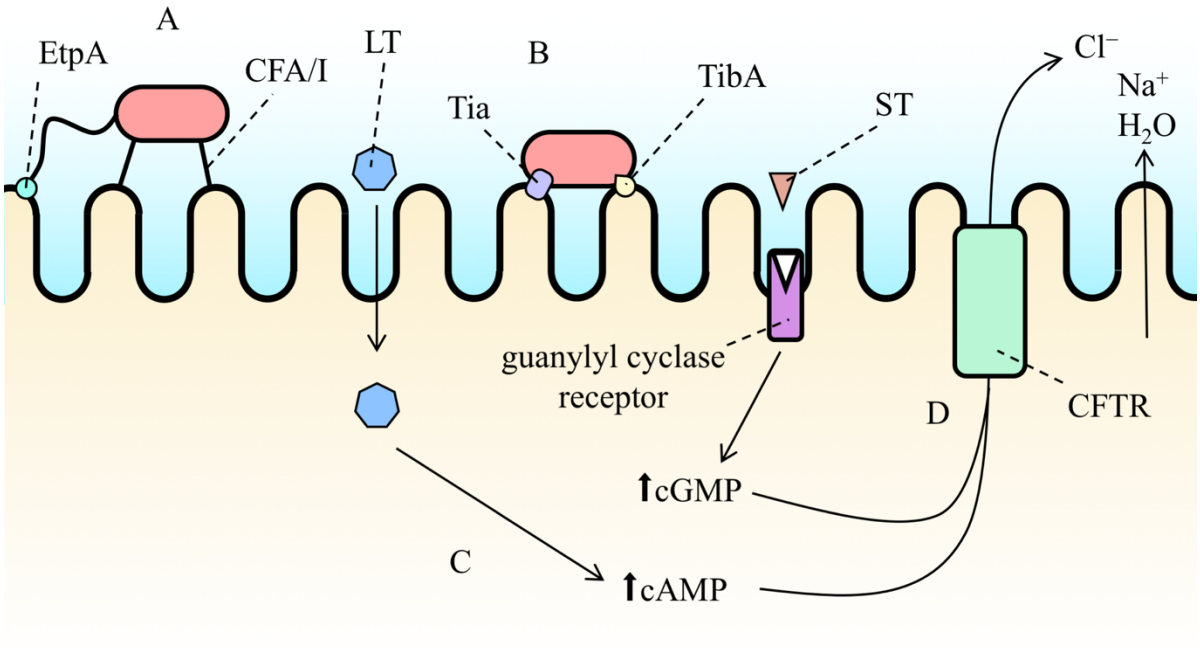


Figure 1.16: Pathogenesis of ETEC. A, ETEC cells adhere to the host epithelium using CFA/I (CS1) and EtpA at the tip of flagellum. B, closer adherence is mediated by the outer membrane proteins Tia and TibA. The LT and ST toxins are secreted. C, the LT toxin is internalised which leads to an increase in intracellular cAMP. The ST toxin activates the guanylyl cyclase receptor resulting in an increase in cGMP in the host cell. D, the increase in cGMP and cAMP in the cell opens the cystic fibrosis transmembrane conductance regulator (CFTR). Cl^- exit the cell. This causes the loss of Na^+ ions to the lumen. The difference in ion concentration in the lumen results in water loss. This causes the watery diarrhoea associated with ETEC infection.

helically intertwined filaments, e.g. CS5 and CS7. An example of the afimbrial CFs is CS6 encoded by B7A. CS6 does not form an observable macromolecular structure when produced. It is similar to CS3 and is thought to be a very short filament structure (Ludi et al., 2008). Another example of an afimbrail adhesion is EtpA. EtpA is incorporated onto the tip of the flagellin and when in contact with host cells can mediate attachment (Roy et al., 2009). The CFs are not equally prevalent with CFA/I, CS3, CS5 and CS6 accounting for 50-80% of ETEC clinical isolates (Bourgeois et al., 2016). These CFs are required for attachment but they do not cause disease as once the cell is adhered to human enterocytes it secretes toxins.

ETEC generally harbours one or more of its two major toxins the heat sable toxin (ST) and the heat labile toxin (LT). Infections with strains producing the ST are associated with a higher risk of death than those producing the heat labile toxin alone (Kotloff et al., 2013). The number of ST receptors in the GI tract decreases with age (Sears and Kaper, 1996). This is thought to be responsible for the predominant ST toxin association with children (Sears and Kaper, 1996). The LT is a AB_5 toxin closely related to the cholera toxin. Both the heat stable and heat labile toxin activate the opening of the cystic fibrosis transmembrane conductance regulator. This leads to the flow of water out of the cell, resulting in water diarrhoea (Sears and Kaper, 1996). While other ETEC factors can also influence disease such as EatA, which is a serine protease T5SS that increases toxin uptake, production of ST and/or LT is characteristic of ETEC strains (Kumar et al., 2014).

The prototypical ETEC strain is H10407. Similar to the prototypical EAEC strain 042, ETEC H10407 elicits diarrhoea in human volunteers (Evans et al., 1978; Coster et al., 2007). The sequencing of ETEC H10407 by Crossman et al., (2010) identified *cexE* as a homolog of *aap*. Like Aap, CexE encodes a predicted Sec signal sequence and is activated by a virulence

activator, CfaD, which is similar to AggR (Pilonieta et al., 2007). Like AggR and the genes encoding the AAF, CfaD is known to activate the genes encoding the CFA/I. CexE with LT and EtpA are associated with outer membrane vesicles (Roy et al., 2010). After immunisation of mice with outer membrane vesicles, the ability of ETEC H10407 to colonise those mice was impaired (Roy et al., 2011). An immune response to CexE was identified in the immunised mice. Thus, CexE appears to be a potential vaccine target for ETEC. The similarity of the protein coding sequence of CexE to Aap suggests a similar function. However, the major subunits of the ETEC CFs are not positively charged at a neural pH (Velarde et al., 2007). The pIs of the major subunit CFs of ETEC are much lower those of the EAEC AAF. Despite this the pI of the CexE protein is similar to that of Aap. Therefore, it is likely that CexE does not perform the same function in this pathovar.

1.8. Bacterial biofilms

The formation of biofilms by EAEC and ETEC strains aide their ability to infect, colonise, and cause disease in the host (Hicks et al., 1996; Ahmed et al., 2012). Biofilms are heterogeneous in nature, where many species of bacteria can compete or cooperate. The majority of biofilm dry weight is comprised of the extracellular matrix, around ~90%, with bacterial cells accounting for around ~10% (Flemming and Wingender, 2010). The extracellular matrix is composed of polysaccharides, proteins (such as fimbriae, pili, and flagella), and extracellular DNA (eDNA) (Flemming et al., 2016). The eDNA forms a major part of the biofilm providing structural support, nutrition, protection from antibiotics, and enabling horizontal gene transfer. There is currently three known mechanisms for eDNA release: cell lysis, outer membrane vesicles, and T4SS (Allesen-Holm et al., 2006; Baumgarten et al., 2012; Hamilton et al., 2005). Cell lysis and outer membrane vesicle release do not have a specificity for the genomic DNA released, however DNA secreted by the T4SS in *Neisseria gonorrhoea* is enriched for a region

of the genome that contains the origin of transfer (Jakubovics et al., 2013). eDNA is not released alone; proteins are secreted that interact with the eDNA.

DNase are secreted that modify the structure of the biofilm or degrade the DNA to be used as a nutrient source (Gödeke et al., 2011; Pinchuk et al., 2008). These DNase are either anchored to the cell or secreted into the extracellular milieu. Other proteins have a structural role as they bind eDNA in the biofilm. The DNABII family of proteins are typically intracellular; binding and condensing the nucleoid. However, in UPEC proteins belonging to the DNABII family are required for the formation of biofilms by binding eDNA (Devaraj et al., 2015). As well as proteins polysaccharides, such as Pel from *Pseudomonas aeruginosa*, can covalently bind eDNA thus strengthening the biofilm (Jennings et al., 2015). eDNA, with proteins and polysaccharides, allows bacteria to form multicellular communities for protection, absorption of nutrients, and colonisation.

1.9. Aims

There are still many questions to be answered about EAEC and ETEC pathogenesis. The development of new technologies can help further the understanding of the virulence transcriptome. Using new profiling techniques like RNA-Seq the AggR regulon will be further characterised. While it is already known that *aap* and the *aat* system are activated by the AggR virulence regulator other targets that were not apparent in the original microarray dataset may become clear.

The homolog of the virulence regulator AggR, CfaD, is known to activate the gene encoding the CexE protein (Pilonieta et al., 2007). However, the function of CexE is not understood. A low sequence identity of 18% to Aap, identified by Crossman et al. (2010), makes it unlikely

that they are functionally similar. Also the mechanism of charge modulation proposed by Velarde et al. (2007) for Aap would not be relevant for CexE due to the differences in charge of the fimbriae in ETEC. Additionally, the role of the Aat system in CexE secretion remains unclear. Crossman et al. (2010) hypothesised that the Aat system was likely to be required for the secretion of CexE, however no experimental conformation was provided. The differences in EAEC and ETEC pathogenesis and the low sequence identity suggest that this cannot be taken for granted.

While the substrates of the Aat system in ETEC are unknown, there are still further questions about its mechanism. Outside of the requirement of the Aat system for the secretion of Aap there is no knowledge of the actual mechanism of action (Nishi et al., 2003). The comparatively high transcription of these systems in response to virulence regulators suggests that they are important for bacterial pathogenesis (Morin et al., 2013; Hodson et al., 2017). Investigating these mechanisms can lead to possible ways to combat these strains which cause significant amounts of morbidity and mortality worldwide.

Therefore, the aims of this study were:

1. Further investigate the AggR regulon in EAEC 042
2. Identify a function for CexE in ETEC H10407
3. Determine the structure of CexE
4. Attempt to understand the mechanism of the Aat system

A mixture of approaches will be used to achieve these aims. RNA-Seq will be used to further characterise the AggR regulon. The role of CexE will be suggested by using an ETEC H10407 deletion mutant and biochemical techniques. The differences in the structure of Aap and CexE

will be assessed by determining the structure of CexE. Finally, the Aat system in EAEC 042 and ETEC H10407 will be investigated to try to understand the function and suggest a new mechanism of action for this atypical T1SS.

CHAPTER 2:

Materials and Methods

2.1. Bacterial strains, growth conditions and culture media

2.1.1. General growth conditions

Bacterial strains used in this study are listed in Table 2.1. Bacterial strains were regularly cultivated in lysogeny broth (LB) which consisted of 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl. Bacterial strains were routinely grown on solid media before inoculation of liquid cultures. In all cases, solid media consisted of LB with 1.2% (w/v) agar (LBA). An overnight culture was prepared by inoculating a single colony into 5 ml LB contained in a 20 ml glass universal which was then incubated at 37ºC, unless otherwise stated, with aeration overnight and the appropriate antibiotic. For the storage of bacteria strains, 800 µl of overnight culture was inoculated into 800 µl LB with 30% glycerol and stored at −80ºC. Strains were streaked onto LBA plates to single colonies. Overnight cultures were inoculated into Erlenmeyer flasks, containing a fifth of their total volume of LB, to an OD_{600} of 0.05 and grown at 37ºC, unless otherwise stated, with aeration and the appropriate antibiotic to the required optical density.

2.1.2. Antibiotic preparation and supplementation

Antibiotics were used to supplement the growth medium for plasmid retention. Kanamycin sulphate and carbenicillin disodium salt stocks were made to a concentration of 100 mg/ml in deionised H_2O and filtered using Millex-GP Syringe Filter Unit, 0.22 µm, polyethersulfone (Merck). The filtered stocks were aliquoted and stored at −30ºC. Final working concentrations were prepared to a concentration of 100 µg/ml in the appropriate medium. Kanamycin sulphate and carbenicillin disodium salt are routinely referred to as kanamycin and carbenicillin, respectively.

Table 2.1: Strains used in this study.

Strain	Description	Source
ETEC H10407	Prototypical enterotoxigenic Strain. CFA/1$^+$, LT$^+$, and ST$^+$	Evans and Evans (1973)
ETEC H10407Δ*aatP*	ETEC H10407 with *aatP* disrupted by a kanamycin cassette	This study
ETEC H10407Δ*aatA*	ETEC H10407 with *aatA* disrupted by a kanamycin cassette	This study
ETEC H10407Δ*aatB*	ETEC H10407 with *aatB* disrupted by a kanamycin cassette	This study
ETEC H10407Δ*aatC*	ETEC H10407 with *aatC* disrupted by a kanamycin cassette	This study
ETEC H10407Δ*aatD*	ETEC H10407 with *aatD* disrupted by a kanamycin cassette	This study
ETEC H10407Δ*cexE*	ETEC H10407 with *cexE* disrupted by a kanamycin cassette	This study
E. coli DH5α	Cloning strain	NEB
E. coli BL21(DE3)	Expression strain	Invitrogen
EAEC 042	Prototypical enteroaggregative strain. AAF/1+ and Pet+	Nataro et al. (1985)
EAEC 042Δ*aap*	EAEC 042 with pJB5603 inserted into *aap*	Sheikh et al. (2002)
EAEC 042Δ*aggR*	EAEC 042 with pJB5603 inserted into *aggR*	Sheikh et al. (2001)
EAEC 042Δ*aatA*	EAEC 042 with pJB5603 inserted into *aatA*	Nishi et al. (2003)
EAEC 042Δ*aatC*	EAEC 042 with pJB5603 inserted into *aatC*	Nishi et al. (2003)
EAEC 042Δ*aatD*	EAEC 042 with pJB5603 inserted into *aatD*	Nishi et al. (2003)
EAEC 042Δ*aap*::*cexE*:*aph*	EAEC 042 with disrupted *aap* by *cexE* and kanamycin cassette.	This study
C. rodentium ICC168	Murine pathogen	Wiles et al. (2004)

2.2. Molecular techniques

Unless otherwise stated the manufacturer's instructions were followed for all molecular techniques.

2.2.1. Plasmid growth and isolation

Strains containing plasmids were grown at 37ºC with aeration in LB supplemented with the relevant antibiotic. If plasmid replication was temperature sensitive, strains containing plasmids were grown at 30ºC with aeration in LB supplemented with the relevant antibiotic. Plasmids were isolated from the overnight culture via Qiagen Mini Prep Kit (Qiagen). Plasmids were eluted in nuclease free water and used the same day or stored at −30ºC. Table 2.2 details the plasmids used in this study. Unless stated in the description, all plasmids were grown at 37ºC. For maintenance, plasmids were transformed into *E. coli* DH5α and stored as previously described.

2.2.2. Polymerase chain reaction

Primers for PCR can be found in Table 2.3 these were used at a concertation of 10 μM. For cloning of genes or mutagenesis, Phusion® High-Fidelity DNA Polymerase (NEB) was used. In all other cases MyTaq™ Red Mix (Bioline) was used. All reagents were stored at −30ºC and thawed on ice prior to use. Unless otherwise stated, the manufactures' instructions were followed for all PCR.

2.2.3. Agarose gel electrophoresis

All PCR products and isolated plasmids were visualised by agarose gel electrophoresis. Agarose gels were prepared by melting 1% agarose in 1x TAE. The working concentration of 1x TAE was made from a 1 in 50 dilution of 50x TAE (2 M Tris-base, 0.05 M EDTA (pH8.0),

Table 2.2: Plasmids used in this study.

Plasmid	Description	Source
pDOC-K	Contains *bla* and *kanR* with FRT sites	Lee et al. (2009)
pKD4	Contains *bla* and *kanR* with FRT sites	Datsenko and Wanner (2000)
pKD46	Heat sensitive plasmid with λ Red recombinase genes expressed in response to L-arabinose	Datsenko and Wanner (2000)
pET26b(+)	T7 expression vector with C-terminal 6His tag, IPTG inducible, *kanR*	Novagen
pET26b-CexE$_{H10407}$	pET26b with *cexE* from ETEC H10407	This study
pET26b-CexE$_{ICC168}$	pET26b with *cexE* from *Citrobacter rodentium*	This study
pET26b-Aap	pET26b with *aap* from *EAEC* 042 into	This study
pCfaD	pBAD/myc-HisA with *cfaD* under PBAD from ETEC H10407	Made by Sammi McKean
pAggR	pBAD30 with *aggR* under PBAD from EAEC 042	Sheikh et al. (2001)
pBAD24	L-arabinose expression vector with *ampR*	Guzman et al. (1995)
pJB39	Constitutive expression of *gfp* with *ampR*	Browning et al. (2013)
pBAD/myc-HisA	L-arabinose expression vector with C-terminal Myc and 6His tag	Invitrogen
pSR*CC*(−41.5)	pBR322 derivative containing consensus DNA site for CRP	Rossiter et al. (2011)a

Table 2.3: Primers used in this study.

Name	Sequence	Description
C*cexE*-F	GGCGGCCATATGAAAAA ATATATATTAGGTGT	Primer for cloning *cexE* from ETEC H10407 into pET26b(+)
C*cexE*-R	CGAAGCCTCGAGTTTAT ACCAATAAGGGGTGT	Primer for cloning *cexE* from ETEC H10407 into pET26b(+)
ICC168/*cexE*/Clone-FW	GGCGGCCATATGAAACT TATAGGAAAATTCAT	Primer for cloning *cexE* from *Citrobacter rodentium* ICC168 into pET26b(+)
ICC168/*cexE*/Clone-RV	CGCGGCCTCGAGTGATT CCCACCTAAAAATTGGC A	Primer for cloning *cexE* from *Citrobacter rodentium* ICC168 into pET26b(+)
042/*aap*/Clone-FW	GGCGGCCATATGAAAAA AATTAAGTTTGTTAT	Primer for cloning *aap* from EAEC 042 into pET26b(+)
042/*aap*/Clone-RV	CGCGGCCTCGAGTTTAA CCCATTCGGTTAGAGCA C	Primer for cloning *aap* from EAEC 042 into pET26b(+)
GD*aatP*-F	GAATTCAAGCTTGCAAA GCGTATTGTTGGTGCAG GCTTGTATAAGTTATAT GGACCGGTCAATTGGCT GGAG	Forward primer for generating kanamycin cassette and disruption of *aatP* gene in ETEC H10407
GD*aatP*-R	GAGCTCCATATGAATAT AAATATAATCTTCATGA AAATCTTTCTTTTATTA AAATATCCTCCTTAGTT CC	Reverse primer for generating kanamycin cassette and disruption of *aatP* gene in ETEC H10407
GD*aatA*-F	GAATTCAAGCTTGATTT ATCAATCTTAATAAAAG AAAGATTTTCATGAAGA GACCGGTCAATTGGCTG GAG	Forward primer for generating kanamycin cassette and disruption of *aatA* gene in ETEC H10407
GD*aatA*-R	GAGCTCCATATGTATAT AATTCAATTTCATTTTC CTTTTTTATTAACTCTC AATATCCTCCTTAGTTC C	Reverse primer for generating kanamycin cassette and disruption of *aatA* gene in ETEC H10407
GD*aatB*-F	GAATTCAAGCTTGAATT ATATAAAATGCATATGT TTTTTATTGGTTAGTTT GACCGGTCAATTGGCTG GAG	Forward primer for generating kanamycin cassette and disruption of *aatB* gene in ETEC H10407
GD*aatB*-R	GAGCTCCATATGATCAA TACTTAATTTAATCATG TAGTGTTATCTCAAATG CTGAATATCCTCCTTAG TTCC	Reverse primer for generating kanamycin cassette and disruption of *aatB* gene in ETEC H10407

GD*aatC*-F	GAATTCAAGCTTCTTGGATACAGCATTTGAGATAACACTACATGATTAAATTAGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatC* gene in ETEC H10407
GD*aatC*-R	GAGCTCCATATGAAATCACATAAAATTTTTATTAAAAGGATATAACCTTCATAATATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatC* gene in ETEC H10407
GD*aatD*-F	GAATTCAAGCTTTCGTATATTGGCATTAAATTATCTTTTGAATTGATTAATGGACCGGTCAATTGGCTGGAG	Forward primer for generating kanamycin cassette and disruption of *aatD* gene in ETEC H10407
GD*aatD*-R	GAGCTCCATATGACCAAATGATGTATATGGTTTATACATCAATGACAAAAAAAATATCCTCCTTAGTTCC	Reverse primer for generating kanamycin cassette and disruption of *aatD* gene in ETEC H10407
cexE-F1	GGGTTAAGTGATAACAGGCG	Forward flanking primer of *cexE* gene in ETEC H10407
cexE-R1	GCACCAACAATACGCTTTGC	Reverse flanking primer of *cexE* gene in ETEC H10407
cexE-F2	GGTGTTATTCTGGCTATGGG	Forward primer with homology inside of *cexE* gene in ETEC H10407
cexE-R2	TTTACAGTCCGATGCATGGC	Reverse flanking primer with homology inside of *cexE* gene in ETEC H10407
aatP-F1	TACAATGTCGGGACTCAACC	Forward flanking primer of *aatP* gene in ETEC H10407
aatP-R1	ACATACGCAAATGCGGATGG	Reverse flanking primer of *aatP* gene in ETEC H10407
aatP-F2	GGGAGAATCCATATGCAACG	Forward primer with homology inside of *aatP* gene in ETEC H10407
aatP-R2	AGTGCCAACACATCATACCG	Reverse flanking primer with homology inside of *aatP* gene in ETEC H10407
aatA-F1	CATCAGGTTGAGTTACTCGC	Forward flanking primer of *aatA* gene in ETEC H10407
aatA-R1	CATTGTAAGCATTGCTGGCG	Reverse flanking primer of *aatA* gene in ETEC H10407
aatA-F2	ACTACATCATCTGGAGTGGG	Forward primer with homology inside of *aatA* gene in ETEC H10407

aatA-R2	GCAATCCACTTAGCGGAATG	Reverse flanking primer with homology inside of *aatA* gene in ETEC H10407
aatB-F1	CATTCCGCTAAGTGGATTGC	Forward flanking primer of *aatB* gene in ETEC H10407
aatB-R1	ATATCCGACACATCACTCCC	Reverse flanking primer of *aatB* gene in ETEC H10407
aatB-F2	TGGTATGGCATACTGAGAGG	Forward primer with homology inside of *aatB* gene in ETEC H10407
aatB-R2	ACGAACGATACCTGATAGCC	Reverse flanking primer with homology inside of *aatB* gene in ETEC H10407
aatC-F1	GTAATCCACATGTGCAAGGG	Forward flanking primer of *aatC* gene in ETEC H10407
aatC-R1	TGCCGTAGTTATTCGTGAGG	Reverse flanking primer of *aatC* gene in ETEC H10407
aatC-F2	GTTATAACTGGCGCTTCTGG	Forward primer with homology inside of *aatC* gene in ETEC H10407
aatC-R2	TCGCCACCAGATAGTAAAGC	Reverse flanking primer with homology inside of *aatC* gene in ETEC H10407
aatD-F1	CTCCCTCTATAGTGTGTAGC	Forward flanking primer of *aatD* gene in ETEC H10407
aatD-R1	TCTGGCGTCCATTCATTTCC	Reverse flanking primer of *aatD* gene in ETEC H10407
aatD-F2	ATTGGGTGGAAACACTACGG	Forward primer with homology inside of *aatD* gene in ETEC H10407
aatD-R2	TAAGTACAGTTCCTCCGTGG	Reverse flanking primer with homology inside of *aatD* gene in ETEC H10407
polA-1	TGGCGATAAAGATATGGC	Forward qRT-PCR primer to the *polA* gene in EAEC 042
polA-2	GTACTTATTCACCACCTC	Reverse qRT-PCR primer to the *polA* gene in EAEC 042
aap-1	TGGAACGCAGATAATGTG	Forward qRT-PCR primer to the *aap* gene in EAEC 042
aap-2	CATTAAGGCCTTGCATAC	Reverse qRT-PCR primer to the *aap* gene in EAEC 042
filC-1	TGATGGTGAAACCATCAC	Forward qRT-PCR primer to the *fliC* gene in EAEC 042

filC-2	AGTATTGGCTGTTTCACC	Reverse qRT-PCR primer to the *fliC* gene in EAEC 042
filA-1	ATATTGCCGATTATCGCC	Forward qRT-PCR primer to the *fliA* gene in EAEC 042
filA-2	ACCAGTTCAATGCTATCG	Reverse qRT-PCR primer to the *filA* gene in EAEC 042
bssS-1	TATGATGCGCTGATGTTG	Forward qRT-PCR primer to the *bssS* gene in EAEC 042
bssS-2	CAACATCAGTGGTTAACC	Reverse qRT-PCR primer to the *bssS* gene in EAEC 042
4803-1	ATCCACATTTGTGGGTAC	Forward qRT-PCR primer to the EC042_4803 gene in EAEC 042
4803-2	ATGTGATTCCATACCAGC	Reverse qRT-PCR primer to the EC042_4803 gene in EAEC 042
flgB-1	GACGATGACCTCAACGCAAC	Forward qRT-PCR primer to the *flgB* gene in EAEC 042
flgB-2	ATTACCGTCAAGCGAAGGCT	Reverse qRT-PCR primer to the *flgB* gene in EAEC 042
phoA-1	CGAACCAGCAAAAACCCCTG	Forward qRT-PCR primer to the *phoA* gene in EAEC 042
phoA-2	GGCTTGTCGATATTGCCGTG	Reverse qRT-PCR primer to the *phoA* gene in EAEC 042
Fwd*aap-aap*/*cexE*	GTGAAAAATACCTCTATATACATGGGGAATATCTAGAGAGAAGTCATATGAAAAAATATATATTAGGTGTTATTCTGGC	Forward primer for amplification of the *cexE* gene from ETEC H10407 with homology to the *aap* gene in EAEC 042
Rev*aap-kanR*/*cexE*	CTCCAGCCAATTGACCGGTCTTATTTATACCAATAAGGGGTGTCACCAC	Reverse primer for amplification of the *cexE* gene from ETEC H10407 with homology to the kanamycin cassette fragment
Rev*aap-aap*/*kanR*	GATAGGAATCCAGCGTGGAGAGTTGAAATTTTAGCTAGAGCTAGATATTAAATATCCTCCTTAGTTCC	Reverse primer for kanamycin cassette amplification with homology to the *aap* gene in EAEC 042
Fwd*aap-cexE*/*kanR*	GTGGTGACACCCCTTATTGGTATAAATAAGACCGGTCAATTGGCTGGAG	Forward primer for kanamycin cassette amplification with homology to the *cexE* PCR fragment
Fwd*aap*-PCREx	GTGAAAAATACCTCTATATA	Forward primer for amplification of the *cexE* kanamycin cassette fragment

Rev*aap*-PCREx	GATAGGAATCCAGCGTGGAG	Reverse primer for amplification of the *cexE* kanamycin cassette fragment
Fwd-*aap*	GTCGAAACGAGTAACACTCG	Forward flanking primer to the *aap* gene in EAEC 042
Rev-*aap*	TGTCCATATTACGAGGGTGG	Reverse flanking primer to the *aap* gene in EAEC 042
T7F	TAATACGACTCACTATAGGG	Forward primer for amplification of pET26b(+) multiple cloning site
T7R	GCTAGTTATTGCTCAGCGG	Reverse primer for amplification of pET26b(+) multiple cloning site

1 M glacial acetic acid in deionised water). The agarose was allowed to cool, when cool enough to be handled, Midori Green Advance (Nippon Genetics) was added. Agarose was poured into moulds to set. All DNA samples except for PCR products from a MyTaq™ Red Mix (Bioline) reaction were prepared by the addition of 5x DNA Loading Buffer Blue (Bioline). One lane was reserved for HyperLadder™ 1kb (Bioline). A voltage of 120 V was applied across the gel in 1x TAE buffer. Gels were electrophoresed until there was sufficient migration of the loading dye. Gels were visualised with UV light (300 nm) using a G:BOX (Syngene).

2.2.4. Cloning

Table 2.3 contains the primers used to amplify the genes for cloning. A primer pair was designed so that they contained 50 bp homology flanking the region to be cloned. At the 5′ end of each primer there was 6 bp sequence before a restriction enzyme cutting site. The strain containing the gene to be cloned was used as the template for the cloning PCR. The PCR product of the cloning primers was termed the insert. The plasmid into which the insert was to be cloned was isolated as previously described. The isolated plasmid was termed the vector. Both the vector and the insert were cut with NdeI (NEB) and XhoI (NEB) using the CutSmart® (NEB) protocol. The cut plasmid was treated with Antarctic Phosphatase (NEB). Cut DNA was purified by gel extraction from a TAE 1% agarose gel using the QIAquick Gel Extraction Kit (Qiagen). DNA was quantified using Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific). T4 DNA Ligase (NEB) was used to ligate the insert DNA into the vector DNA. The following equation was used to determine the amount of insert DNA to be used for 5 µl of the vector DNA.

A negative control with the insert volume replaced with nuclease-free water was used for the ligation. Ligation reactions were incubated for 1 hour at 16ºC. From the ligation reaction 10 µl

was used to transform 30 µl NEB® 5-alpha Competent *E. coli* (High Efficiency). Recovered cells were plated on LBA supplemented with the required antibiotic and incubated at 37ºC overnight. Colonies were screened for the successful insertion of the insert with colony PCR. PCR primers with homology to regions of the vector outside of the multiple cloning site were used. Colonies that produced PCR products larger than the empty vector were isolated. Plasmids were sequenced by the Biosciences Genomics DNA sequencing facility. Only strains containing plasmids that were 100% identical to the desired ligation of vector and insert were stored as previously described.

2.3. Gentenic engineering of bacterial strains

2.3.1. Production and transformation of chemically competent cells

Overnight cultures were inoculated into 50 ml LB to an OD_{600} of 0.05 and grown at 37ºC with aeration. When cultures had reached an OD_{600} of 0.6 they were cooled on ice for 10 minutes. The cells were harvested by centrifugation at 3,400 *g* for 10 minutes at 4ºC. The culture supernatant was aspirated and tubes were inverted to remove any remaining supernatant. The pellet was resuspended in 50 ml ice-cold 50 mM $CaCl_2$ and incubated on ice for 30 minutes. Cells were centrifuged at 3,400 *g* for 10 minutes at 4ºC. The pellet was resuspended in 500 µl ice-cold 0.1M $CaCl_2$ with 15% glycerol for 30 minutes. Cells were either used immediately or stored at −80ºC.

Chemically competent cells were incubated with the plasmid for 30 minutes on ice. Cells were heat-shocked at 42ºC for 90 seconds and returned to ice for 5 minutes. The cells were recovered in 1 ml of LB which was incubated for 1 hour at 37ºC before plating onto LBA supplemented with the selective antibiotic.

2.3.2. Production and transformation of electrocompetent cells

Overnight cultures were used to inoculate 50 ml LB. Cultures were grown to an OD_{600} of 0.6 they were then cooled on ice for 10 minutes before centrifugation at 3,400 *g* for 15 minutes at 4ºC. Pellets were resuspended in 50 ml ice-cold 10% glycerol. Resuspended cells were centrifuged at 3,400 *g* for 15 minutes at 4ºC. Pellets were washed twice with 20 ml ice-cold 10% glycerol. The pellet was resuspended in 500 µl ice-cold 10% glycerol. Electrocompetent cells were used immediately. Plasmid or linear DNA fragments in deionised water were incubated with 50 µl of cells on ice for 5 minutes. The mix of cells and DNA were transferred to 2 mm ice-cold electroporation cuvette and electroporated at 22,000 V. Cells were immediately suspended in 1 ml SOC media (20 g tryptone, 5 g yeast extract, 8.56 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM D-glucose in 1 L H_2O) and transferred to a 15 ml falcon tube. Cells were recovered for 1 hour at 37ºC, or 30ºC for temperature sensitive plasmids. Recovered cells were plated onto the appropriate antibiotic for selection on LBA and incubated overnight at 37ºC or 30ºC for temperature sensitive plasmids.

2.3.3. Gene deletions in *E. coli*

Gene deletions in *E. coli* were made using the protocol from Datsenko and Wanner (2000). The forward primer contained 50 bp homology to the 5′ of the target gene containing the start codon. The reverse primer contained 50 bp homology to the 3′ of the target gene starting from the 7^{th} last codon. At the 3′ of both of the primers, there was homology to the kanamycin cassette. Both the pDOC-K and pKD4 plasmids was used for the template of the kanamycin cassette (Lee et al., 2009; Datsenko and Wanner, 2000). The primers were used to amplify the kanamycin cassette with flanking 50 bp of homology to the target gene and the resulting PCR reaction was electrophoresed on a 1% agarose gel. The PCR product at 1.5 kb was purified

using the QIAquick Gel Extraction Kit (Qiagen). The linear fragment was eluted in 50 µl deionised water.

The strain in which the mutation was desired, was made electrocompetent as described above and transformed with pKD46. Cells were recovered at 30ºC and plated on LBA with 100 µg/ml carbenicillin. Plates were incubated overnight at 30ºC. Colonies that grew were transferred to fresh LBA. Overnight cultures of the colonies suspected of containing pKD46 were incubated at 30ºC with aeration. Plasmids were isolated as previously described from the overnights and visualised on 1% TAE agarose. Transformants that contained a band of the same size as pKD46 were stored as previously described. An overnight culture of a strain containing pKD46 was used to inoculate 50 ml of 0.2% L-arabinose 100 µg/ml carbenicillin LB. Cultures were grown to an OD_{600} of 0.6 and cells were made electrocompetent as previously described. The electrocompetent cells were electroporated with 5 µl of the purified linear fragment containing the kanamycin cassette as previously described. Cells were recovered as previously described. Recovered cells were plated onto 100 µg/ml kanamycin LBA and incubated at 37ºC overnight. Colonies that grew were checked for insertion of the kanamycin cassette by colony PCR. Colonies that produced a PCR product of the expected size for the kanamycin cassette insertion were streaked on LBA containing 100 µg/ml kanamycin and incubated overnight at 37ºC. Single colonies were checked by PCR for the removal of wild-type contamination. Colonies that were free of contamination were stored as previously described.

2.3.4. Sewing PCR for insertion of genes into *E. coli*

Sewing PCR and Datsenko and Wanner (2000) was used to introduce a gene into the genome of an strain and replace a gene already present. The fragment was designed in such a way that the new gene would be expressed in the same manner as the gene it replaced. A schematic of

the process is presented in Fig. 2.1. Phusion® High-Fidelity DNA Polymerase (NEB) was used throughout. In Fig. 2.1, the gene, *gene A*, from the source strain was introduced into the recipient strain at the same position as *gene B*. Thus, *gene A* was expressed in the same conditions as *gene B*.

In total 3 pairs of primers were designed, the colours referred to in the text match the colours of the primers in Fig.2.1. Primer pair 1, for the amplification of *gene A* were designed with homology to the 25 bp at the 5' of *gene A* (green) including the start codon and homology to the 50 bp upstream of the start codon of *gene B* (red). The primer partner was designed as the reverse complement of the 25 bp at the 3' of *gene A* (green) including the stop codon and priming site 1 of pKD4 (blue). Primer pair 2, for the amplification of the kanamycin cassette were designed with 25 bp homology to the 3' of *gene A* (green) including the stop codon and priming site 1 of pKD4 (blue). The primer partner was the reverse complement of priming site 2 of pKD4 and the 50 bp downstream of the stop codon of *gene B*. A final pair of primers, primer pair 3, were designed with 20 bp homology to the 5' and 3' ends of the desired linear fragment to *gene B* (purple). PCR was used to amplify *gene A* and the kanamycin cassette using primer pairs 1 and 2, repetitively. The PCR products were visualised by agarose gel electrophoresis, as previously described. The PCR products of the anticipated size were purified by gel extraction using the QIAquick Gel Extraction Kit (Qiagen). The extracted DNA was eluted in nuclease-free water and quantified. The purified PCR products were made to 100 ng/µl. For the sewing PCR, 4 reactions were prepared. The buffer for each reaction consisted of 6.5 µl nuclease-free water, 5 µl 5X Phusion HF Buffer (NEB) and 5 µl 10 mM dNTPs. Both of the PCR products of *gene A* and the kanamycin cassette amplification were added into separate reactions at a volume of 0 µl, 0.5 µl, 1 µl and 2.5 µl. Each reaction was made up to 22.25 µl with nuclease free-water. Prior to cycling 0.25 µl of Phusion DNA Polymerase (NEB)

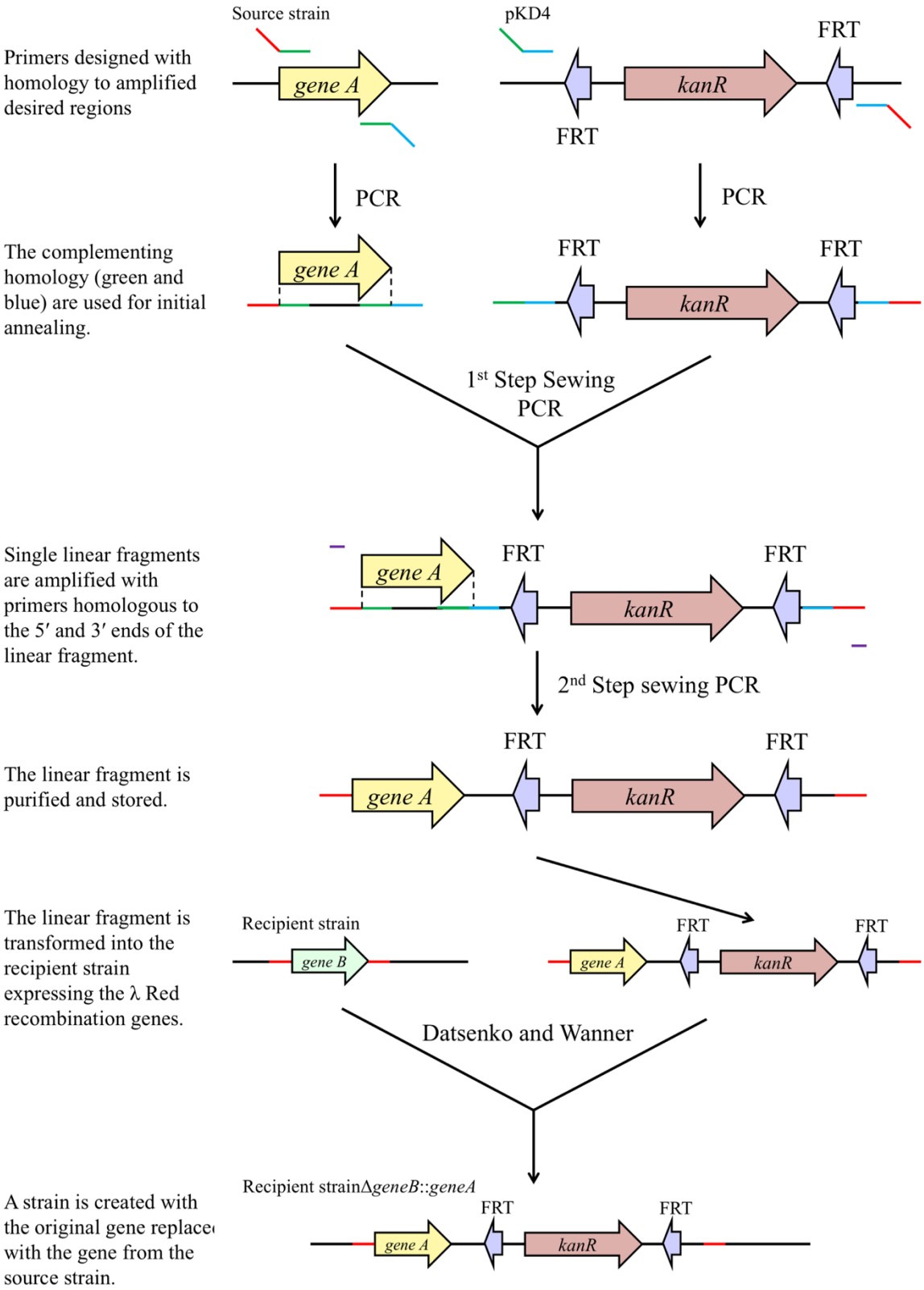


Figure 2.1: Schematic of sewing PCR. The colour represents their homology. Red is homologous to the recipient strain, green to *gene A*, blue to priming site 1 and 2 of pKD4 and purple to the 20 bp at the 5' and 3' of the linear fragment.

was added to each reaction. The following program was used, 98ºC for 5 minutes then 10 cycles of 98ºC for 10 seconds, 40ºC for 30 seconds and 72ºC for 2 minutes. The reactions were removed and 1.25 µl of each of primer pair 3 were added the reactions. The reactions were placed back in the thermocycler. The following program was used, 15 cycles of 98ºC for 10 seconds, 45ºC for 30 seconds and 72ºC for 2 minutes followed by 10 cycles of 98ºC for 10 seconds, 50ºC for 30 seconds and 72ºC for 2 minutes then a final extension of 72ºC for 5 minutes. The reactions were stored at 4ºC.

The reactions were separated by agarose gel electrophoresis. If a band at the size of the desired product was not present, reactions were set up again and the extension time for each cycle was increased by 30 seconds. This was repeated until the desired product was produced. The desired product was extracted using QIAquick Gel Extraction Kit (Qiagen). If a higher quantity of the product was required, a PCR using primer set 3 was set up using the manufacturer's instructions. The linear fragment of *gene A* and the kanamycin cassette flanked by the regions of homology was used as the linear fragment for Datsenko and Wanner as described above.

2.4. Protein analysis

Unless otherwise stated the manufacturer's instructions were used for all reagents.

2.4.1. Preparation of whole cell lysate

Cultures were normalised to 1 ml of culture volume at an OD_{600} of 0.8. Cells were harvested by centrifugation at 3,400 *g*. The culture supernatant was aspirated. The pellet was resuspended in 100µl Sample Buffer, Laemmli 2× Concentrate (Sigma). The protein sample was incubated at 100ºC for 5 minutes. The sample was cooled on ice. Undissolved particulates were removed

by centrifugation at 16,000 *g* for 2 minutes at room temperature. Sample were analysed by SDS-PAGE.

2.4.2. Precipitation of culture supernatant proteins

The culture supernatant was separated by centrifugation at 8,000 *g* for 10 minutes at 4 ºC. The culture supernatant was removed from the cell pellet and 4 ml of was filtered using Millex-GP Syringe Filter Unit, 0.22 µm, polyethersulfone (Merck). The culture supernatant was cooled on ice. Ice-cold 100% TCA was added to a final concentration of 20%. Samples were incubated on ice for 30 minutes. Proteins were pelleted by centrifugation at 14,000 *g* for 15 minutes at 4ºC. The supernatant was removed and discarded from the sample. The pellet was washed twice with 1 ml ice-cold 100% methanol with centrifugation at 14,000 *g* for 15 minutes at 4ºC. The supernatant was removed, residual methanol was evaporated by incubating the sample at 60ºC for 10 minutes with the lid off. The pellet was resuspended in 50 µl Sample Buffer, Laemmli 2× Concentrate (Sigma). If a colour change to yellow occurred saturated Tris-base was added until the original colour returned. Samples were incubated at 100ºC for 5 minutes. Samples were placed on ice. Undissolved particulates were removed by centrifugation at 16,000 *g* for 2 minutes at room temperature. Samples were analysed by SDS-PAGE.

2.4.3. SDS-PAGE

Typically gels proteins were analysed using Tris-tricine SDS-PAGE. However, where stated NuPAGE™ 10% Bis-Tris Protein Gels, 1.0 mm, 10-well (Thermo Fisher Scientific) were used. Tris-tricine SDS-PAGE were poured using Mini-PROTEAN® Tetra Handcast Systems (Bio-Rad) with 1 mm thickness. Gels were poured to consist of a resolving gel (10% acrylamide) and a stacking gel (4% acrylamide). The resolving gel was 17 ml Tris-tricine buffer concentrate, 17 ml ProtoGel 30% (GeneFlow), 12 ml deionised water and 5 ml 100% glycerol.

The stacking gel was 6 ml Tris-tricine buffer concentrate (3 M Tris-HCl, 0.3% sodium dodecyl sulphate (SDS) at pH 8.25), 3.2 ml ProtoGel 30% (GeneFlow) and 15 ml deionised water. Separate buffers were used for the anode and the cathode in the electrophoresis chamber. The 10x cathode buffer (1 M Tris-HCl, 1 M tricine, 1% SDS at pH 8.25) and the 10x anode buffer (2 M Tris-HCl at pH 8.9) were diluted with deionised water to the 1x working concentration. Electrophoresis chambers were assembled with the gels according to the manufacturer's instructions. A voltage of 150 V was applied to the gels for 35 minutes. The gels were visually inspected for the progress of the dye-front. If the dye front was not migrated satisfactorily further electrophoresis was applied. When the migration of the dye front was satisfactory, the gels were removed for analysis by Coomassie Brilliant Blue stain or western blotting.

2.4.4. Coomassie Brilliant Blue staining

Gels were stained using the Fast Coomassie (0.05% Coomassie Brilliant Blue, 20% Propan-2-ol, 0.4 M Citric acid). Gels were microwaved at full power for 40 seconds. Incubated rocking at room temperature for 5 minutes and microwaved at full power for 40 seconds. Fast Coomassie was removed and the gel washed with hot water. The gel was covered with hot water and microwaved at full power for 40 seconds. This was repeated until the desired intensity of stain was achieved.

2.4.5. Western blotting

Gels of samples that were used for western blotting were always prepared in duplicate. One gel was strained with Coomassie Brilliant Blue, as described above. The other gel was transferred to nitrocellulose using iBlot 2 Dry Blotting System (Thermo Fisher Scientific). After transfer, any extraneous membrane that the protein was not transferred to was removed. The membrane was covered in Blotto (50 g skimmed milk powder, 2.42 g Tris-base, 8 g NaCl,

1 g sodium azide, pH 8.4 in 1000 ml deionised water). The membrane was incubated for a minimum of 30 minutes with rocking at room temperature. Primary antibody was diluted in Blotto. Antibody specific to Aap was diluted at 1 in 5,000 (Sheikh et al., 2002). Antibody specific to CexE was diluted 1 in 1,000 (this study). Antibody specific to the β′ subunit of RNAP (E. coli RNA Polymerase beta Monoclonal Antibody, NeoClone) at 1 in 10,000 and 6x-His Tag Monoclonal Antibody (Thermo Fisher Scientific) at 1 in 5,000. Blotto was removed from membranes. Membranes were incubated with rocking in primary antibody for a minimum of 1 hour but typically overnight at 4ºC. Primary antibody was removed from the membranes. Membranes were washed three times in TBST (2.42 g Tris-base, 8 g NaCl, 1% Tween-20, pH 8.4 in 1000 ml) for 5 minutes at room temperature.

The binding of primary antibody to proteins on the membrane was detected by antibodies with conjugated Horseradish Peroxidase (HRP) or by alkaline phosphatase. When using HRP, the secondary antibody for Aap and CexE primary antibodies was Goat Anti-Rabbit IgG H&L (HRP) from abcam. The secondary antibody for RNAP and 6His tagged proteins was Rabbit Anti-Mouse IgG H&L (HRP) from abcam. All secondary antibodies were diluted 1 in 10,000 in Blotto. Membranes were covered and incubated for a minimum of 1 hour at 4ºC with rocking. Membranes were washed 4 times with TBST as previously described. Western blots were developed using Novex™ ECL Chemiluminescent Substrate Reagent Kit (Thermo Fisher Scientific). Westerns were visualised using the ChemiDoc™ XRS+ System (Bio-Rad) or GeneGnome XRQ (Syngene).

The secondary antibodies for alkaline phosphatase was Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat (Sigma) or Anti-Mouse IgG (whole molecule)–Peroxidase antibody produced in rabbit. Antibodies were diluted 1 in 10,000 in Blotto.

Membranes were covered with secondary antibodies and incubated at 4ºC for a minimum of 1 hour rocking. The secondary antibodies were discarded. Membranes were washed 4 times with TSBT as previously described. The TBST was removed and the secondary antibody binding was visualised using 1-Step™ NBT/BCIP Substrate Solution (Sigma). The reaction was stopped after the desired exposure by washing the membrane with water.

2.4.6. Protein production optimisation

Protein was produced in *E. coli* BL21(DE3) using the pET vector expression system. *E. coli* BL21(DE3) was made chemically competent as previously described. *E. coli* BL21(DE3) was transformed with pET26b encoding the protein to be purified. The gene was inserted into the plasmid so that it was produced with a 6 C-terminal histamine tag (6His). For protein production, overnight cultures were used to inoculate separate 50 ml LB. Cultures were grown to an OD_{600} of 0.3 at 37ºC with aeration. The optimal condition for protein production was determined for temperature and induction concentrations. Protein production was tested first for induction concentration. Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma) was used to induce the expression of *cexE* from pET26b. IPTG was dissolved to 0.5 M in 10 ml deionised water. IPTG was aliquoted and stored at −30ºC. Aliquots were thawed and separate cultures at an OD_{600} of 0.3 were induced by addition of IPTG to a concentration of 250 μM, 50 μM, 10 μM or 0 μM. Cultures were grown at 37ºC with 180 rpm. Protein sample were taken at 4 hour and 24 hours, post induction. Protein samples were analysed by Coomassie Brilliant Blue staining and western blotting with anti-His antibodies, as previously described.

2.4.7. Protein production and purification

The optimal conditions for protein production were identified. Overnight cultures of 50 ml LB were used to inoculate the desired amount of LB (typically 2 L). Cultures were grown at 37ºC

with aeration. When the cultures reached an OD_{600} of 0.3 the cultures were transferred to the optimal temperature for protein production. The cultures were cooled for 30 minutes with 180 rpm. IPTG was added to the optimal protein production concentration. The cultures were incubated overnight at the optimal protein production temperature. After overnight growth, cells were harvested by centrifugation at 6,000 *g* for 15 minutes at 4ºC. The cell pellet was resuspended in ice-cold binding buffer (50 mM NaP (77:33 ratio of Na_2HPO_4 to NaH_2PO_4), 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP at pH 7.3). The volume that the cell pellet was resuspended in was 3 ml of buffer to 1 g of wet cells. Cells were lysed using EmulsiFlex®-C3 (Avestin). The cell lysate was kept cold throughout the procedure. Cells that were not lysed were removed by centrifugation at 10,000 *g* for 10 minutes at 4ºC. The pellet was discarded. The cell membrane components were removed by centrifugation at 75,000 *g* for 1 hour at 10ºC. The pellet was discarded. The supernatant was filtered using Millex-GP Syringe Filter Unit, 0.22 μm, polyethersulfone (Merck). A HisTrap™ HP 5 ml column (GE Healthcare Life Sciences) was washed first with 50 ml of distilled H_2O then with 50 ml binding buffer using a peristatic pump. The filtrate was applied to the washed HisTrap™ HP 5 ml column. The filtrate was cycled through the column overnight at 4ºC. In the morning, the column was washed with 5 column volumes of binding buffer. The protein was eluted from the column using 10 column volumes of the elution buffer (50 mM NaP, 500 mM NaCl, 500 mM imidazole and 0.5 mM TCEP at pH 7.3). Fractions were taken of each of the wash volumes and the elution volumes. Protein samples were analysed using Tris-tricine SDS-PAGE. Samples containing purified protein were concentrated using Vivaspin® 6 5,000 MWCO columns (Sartorius Stedim). Protein was buffer exchanged using membrane filtration into 50 mM NaP, 250 mM NaCl and 0.5 mM TCEP at pH 7.3.

2.4.8. Protein concentration determination

The concentration of a protein sample was determined using the absorbance of the protein at 280 nm. All absorbance readings were taken using 1 ml glass quartz cuvettes. The spectrometer was blanked using the buffer that the protein was in. The protein sample was diluted and placed into 1 ml glass quartz curette. The absorbance was measured at 280 nm. The concentration of the diluted protein was determined using the following equation.

$$c = \frac{A}{\varepsilon \times b}$$

Where c is molarity, A is absorbance at a wavelength of 280 nm, ε is the extinction coefficient of the protein and b is the path length in centimetres. Protein extinction coefficients were calculated using ProtParam (ExPASy). Protein concentration was left in molarity or converted to mg/ml as required.

2.4.9. Primary antibody production

CexE-6His protein was purified as previously described. Purified CexE-6His protein was denatured with 6 M Urea (Sigma). The denatured CexE-6His protein was aliquoted into 200 µl at a concentration of 200 µg/100 µl. Primary antibodies were produced by Eurogentec using the 28-day speedy protocol. The serum received was used as the primary antibody for the detection of CexE.

2.5. Phenotypic analyses

2.5.1. Aggregation assay

Overnight cultures were grown in triplicate at 37ºC with aeration. The OD_{600} of the cultures was normalised to OD_{600} 3 in 5 ml with LB. After normalisation, the initial reading for each cell suspension was taken. The cells kept stationary and allowed to aggregate at room

temperature. Samples of 100 µl were taken from just under the meniscus of the cultures. The OD_{600} of the sample was measured. After the desired time course the OD_{600} measurements taken from each culture were compared to the initial reading and a percentage of the initial reading calculated.

2.5.2. Flow biofilm growth

The plasmid pJB39 was isolated as previously described. ETEC H10407 and ETEC H10407Δ*cexE* were transformed by electroporation with pJB39 as previously described. To create the flow cells, cover glass slides were attached using silicone to the top of Perspex flow cells and allowed to dry overnight at room temperature. The flow biofilm system was connected as in the schematic (Fig. 2.2). Clear Translucent Silicone Hose Pipe Tubing (Advanced Fluid Systems) with 1 mm diameter was used throughout except for the initial tubing into the media, which was Clear Translucent Silicone Hose Pipe Tubing (Advanced Fluid Systems) with 2 mm diameter. The media reservoir bottle was connected to a peristatic pump by 2 mm tubing that was adapted to 1 mm tubing before the pump. The periplasmic pump was a Watson Marlow Multi channel cassette pump - 205CA. A single cassette was set up for each line. The pump was connected to bubble traps. Bubble traps functioned by allowing liquid to flow in at the bottom. The inlet for the bubble trap was raised whilst the outlet was below that of the inlet. The liquid was allowed to fill the trap. The tap was closed. Thus, any bubble that were came into the trap were caught before they could pass through the flow cell. Connectors flanked the flow cell allowing the flow cell to be removed for microscopic inspection and re-connected for continuation of the experiment. The stopcock was used to remove bubbles that formed due to the re-connection of the connectors. The media that flowed through the flow cell was collected in waste before disposal. System was washed at 90 rpm for with 2 L deionised H_2O. System was sterilised for 1 hour at 20 rpm with 0.5% sodium

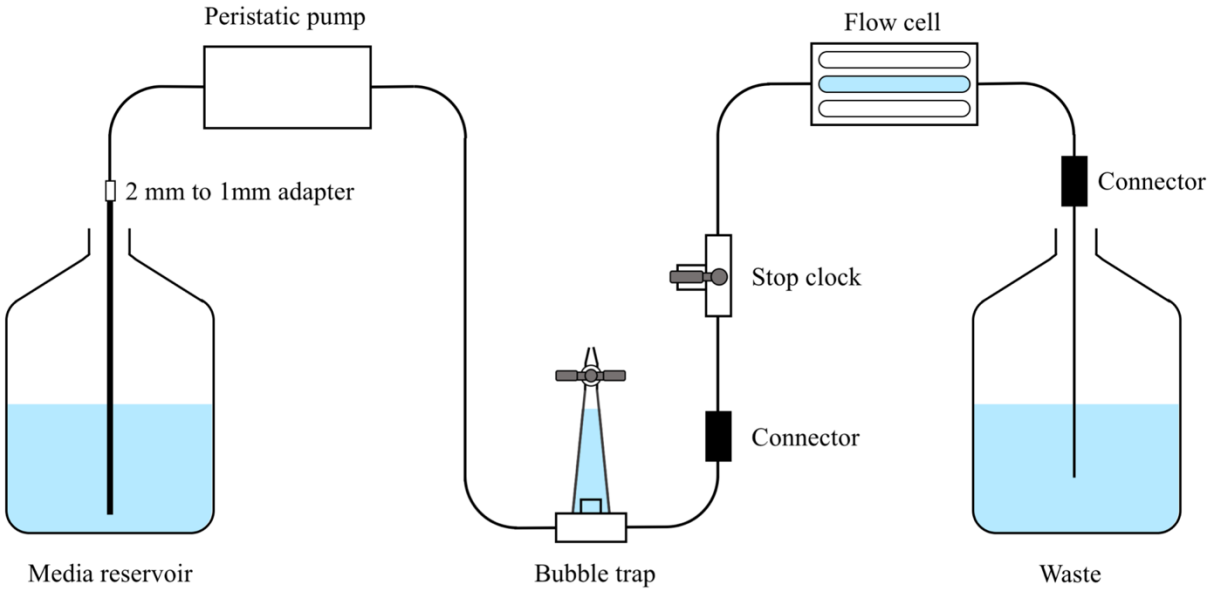


Figure 2.2: Schematic of flow biofilm system. The above schematic is for one chamber of a flow cell only. Each flow cell contained three chambers. The dimensions of one chamber is 40 mm x 4 mm x 50 μm (length x width x height). Each chamber is connected in the same way through a separate line. After autoclaving, the M9 media was incubated at 37ºC to reduce bubble formation. A bubble trap was used in each line to reduce the amount of bubbles that entered the flow cell chambers. The amount of bubbles entering the flow cell chambers needs to be at a minimum; as bubbles can destroy the bacterial biofilm that forms in the chambers. The connectors were placed so that the flow cell could be removed from the system for imaging and then reconnected for further growth if required. The stop clock has one inlet and two outlets. The tap on the stop clock opens and closes the outlets; this allows bubbles that have formed in the connector due to reconnection to be removed before they can enter the flow cell. Prior to use, the system was connected and water pumped through overnight to check for leaks. After the system was sterilised with 0.5% sodium hypochlorite, the tube into the media reservoir was washed with 70% ethanol during transfer to the M9 glucose media. This was to ensure sterility during transfer into the culture media.

hypochlorite made up to 1 L in deionised H_2O. The system was moved to the 37ºC incubator room. The system was washed at 20 rpm for 2 hours with 2 L deionised H_2O. M9 glucose for the flow biofilm experiment was prepared as follows. Before the media was cooled after autoclaving the M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl to pH 7.3) was incubated at 37ºC. The night prior to use, 4 ml 1M $MgSO_4$, 200 µl 1M $CaCl_2$ and 40 ml 20% D-glucose were added to the 2 L M9 salts. Media was incubated overnight at 37ºC. Just prior to use the required antibiotic was added. The system was filled with M9 glucose and left stationary for 2 hours. The cells were prepared for injection from triplicate overnight cultures of ETEC H10407 pJB39 and ETEC H10407Δ*cexE* pJB39, grown as previously described. In the morning strains were inoculated into fresh 5 ml LB. Cells were grown to an OD_{600} of 0.3. From the cultures, 1 ml of cells was taken and pelleted by centrifugation at 3,400 *g*. The cell pellet was washed three times with PBS. The cells were resuspended in M9 glucose, 300 µl was used to inoculate a flow cell chambers using a 23G needle. Holes in the tubing were sealed using silicone. The cells were left stationary in the flow cell for 1 hour. A flow rate of 1 ml/min was applied to the system. The system was left overnight at 37ºC. After 15 hours, the biofilm growth was viewed using the confocal microscope at the BALM facility in the University of Birmingham.

2.6. RNA Sequencing

Before handling RNA, areas were decontaminated and RNases were removed using RNaseZap™ RNase Decontamination Solution (Thermo Fisher Scientific). RNA was sequenced with the help of Dr. Pavelas Sazinas at the University of Warwick.

2.6.1. mRNA isolation

RNA was isolated from 50 ml Dulbecco's Modified Eagle's Medium - high glucose (Sigma),

referred to as DMEM high glucose from now on. Overnight night cultures were grown as previously described. From the overnight cultures 1 ml of cells was pelleted by centrifugation. The cell pellet was washed 3 times with PBS. The final suspension of the cell pellet was in 1 ml DMEM high glucose. Pre-warmed DMEM high glucose media was inoculated to an OD_{600} of 0.05. Cultures were grown at 37ºC with aeration to OD_{600} of 0.6. From each culture 2 ml of cells was harvested by centrifugation. RNA was isolated using RNeasy Mini Kit (Qiagen). RNA was eluted in 50 µl of the TE buffer supplied. Contaminating DNA was removed from the RNA samples using RNase-Free DNase Set (Qiagen). Complete removal of contaminating DNA was confirmed by PCR using the 27F and 1492R. RNA quantity was determined from 1 µl of each sample using the Qubit™ RNA BR Assay Kit (Thermo Fisher Scientific) on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). RNA quality was assessed from 1 µl of each sample using Agilent RNA 6000 Nano Kit (Agilent Genomics). The RNA 6000 Nano chip was run on the 2100 Bioanalyzer (Agilent Genomics). RIN values were determined. Any sample with a RIN value of less than 8 was discarded. From 3.5 µg of RNA from each sample ribosomal RNA (rRNA) was depleted using the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina®). Successful rRNA depletion was analysed using Agilent RNA 6000 Pico Kit (Agilent Genomics). The RNA 6000 Pico chip was run on the 2100 Bioanalyzer (Agilent Genomics). Successive rounds of rRNA depletion were used if the rRNA was not sufficiently depleted. Depleted RNA was quantified using Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific) on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific).

2.6.2. RNA Sequencing

RNA was reverse transcribed and prepared for sequencing using TrueSeq® Stranded mRNA LT Sample Prep Kit (Illumina®). The fragment size was checked using Agilent DNA 1000 Kit (Agilent Genomics). The DNA 1000 chip was run on the 2100 Bioanalyzer (Agilent

Genomics). Each sample was quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). The samples were pooled and sequenced on MiSeq Desktop Sequencer (Illumina®).

2.6.3. Sequencing analysis

Raw sequencing reads were downloaded in FASTQ format. Trimmomatic-0.36 was used to filter out poor reads and was called as follows LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50 (Bolger et al., 2014). Burrows Wheeler Aligner (BWA) was used to align the reads to the genomic sequence (Li and Durbin, 2010). Index files for BWA was created using bwa index from the EAEC 042 chromosome (FN554766.1) and the pAA2 plasmid (FN554766.1) sequence by Chaudhuri et al., (2010). The reads that passed the Trimmomatic-0.36 filtering were aligned to the EAEC genome using bwa mem. The suite SAMtools was used to convert SAM files to BAM files (Li et al., 2009). The SAM files generated by bwa mem were converted to sorted BAM files using samtools view. BAM files were sorted using samtools sort. Finally, BAM files were indexed using samtools index. Reads that were not mapped uniquely were removed using samtools -bq 1 on the indexed sorted BAM files. The number of reads that aligned to a region encoding a gene was counted using featureCounts (Liao et al., 2014). The .gtf file required for featureCounts was made from the .gff3 file downloaded from NCBI for FN554766.1 and FN554766.1 using gffread from Clufflinks (Trapnell et al., 2010). Differential gene expression was analysed in the R package DESeq2 (Love et al., 2014).

2.6.4. Quantitative reverse transcriptase PCR

Triplicate overnight cultures of EAEC 042 pBAD24, EAEC 042Δ*aggR* pBAD24 and EAEC 042Δ*aggR* pAggR were used to inoculate 4 ml DMEM high glucose to an OD_{600} of 0.05. Cells

were grown to an OD_{600} of 0.4 at 37ºC with aeration. The production of AggR was induced by the addition of 1 ml 10% L-arabinose dissolved in DMEM high glucose. Thus, the final concentration of L-arabinose in the culture was 2%. The cells were grown for a further hour at 37ºC with aeration. RNA was extracted using RNeasy Mini Kit (Qiagen). Contaminating DNA was removed using TURBO DNA-free™ Kit (Ambion). Isolated RNA was quantified using Qubit™ RNA BR Assay Kit (Thermo Fisher Scientific). From each sample 1.75 μg of RNA was used for reverse transcription. The reverse transcription reaction was set up with the Tetro cDNA Synthesis Kit (Bioline) using Random Hexamers.

2.7. Phylogenetic analysis of Aap and the Aat proteins

The protein sequences of the Aat proteins from EAEC 042 were used to search the NCBI non-redundant protein sequences database using PSI-BLAST. Significant hits were downloaded. The results were parsed using a custom python script so that a single sequence was represented by single code. Sequences were clustered using cd-hit (Li and Godzik, 2006). Either HHpred and HMMER (hmmer.org) were used to identify homologs (Alva et al., 2016; Finn et al., 2011). HMMER searches were against the UniprotKB or Swissprot databases. Protein sequences were aligned using Clustal Omega (Sievers et al., 2011). RAxML was used for the calculation of distances between sequences using the maximum likelihood method (Stamatakis, 2014). The correct model for the calculation of maximum likelihood based on the protein sequence alignment was determined using the ProteinModelSelection.pl script included in the RAxML distribution. Trees were drawn using iTOL (Letunic and Bork, 2016).

2.8. NMR

Dr. Timothy Knowles set-up and collected the experimental data at the Biomolecular NMR facility at the University of Birmingham. NMRFAM-Sparky was used for all NMR

experimental analysis (Lee et al., 2015).

2.8.1. Protein production and purification

Prior to the ^{13}C and ^{15}N labelled growth the optimal conditions for CexE-6His production were determined. *E. coli* BL21(DE3) was made chemically competent as previously described. Described below is the double labelled growth for structure determination. A single labelled growth, ^{15}N only, was previously conducted to determine CexE-6His viability for NMR structure determination. For this growth only 1 L was used and the ^{13}C was replaced with D-glucose (Sigma). All other steps were exactly the same. The plasmid pET26b-*cexE* was isolated as previously described. The plasmid was sequenced prior to transformation. *E. coli* BL21(DE3) was transformed by heat shock with pET26b-*cexE*. Successful transformants were streaked to single colonies on LBA100 μg/ml kanamycin. A single colony was inoculated into 50 ml LB and grown overnight at 37ºC with aeration. M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl to pH 7.3) were made to 2 L. Prior to growth, nutrient mix was added to the M9 salts at 20 ml per L of M9 salts. The nutrient mix contained 1 ml Thiamine (20 mg/ml), 0.4 ml 3mM $FeCl_3$, 1 ml metal mix (4 mM $ZnSO_4.7H_2O$, 1 mM $MnSO_4.5H_2O$, 4.7 mM H_3BO_3, 0.7 mM $CuSO_4.5H_2O$ made to 100 ml in deionised water), 2 ml 1M $MgCl_2$, 0.1 ml $CaCl_2$, 1 ml 100 mg/ml kanamycin sulphate, 1 g ^{15}N labelled ammonium chloride, and 2 g ^{13}C glucose made to 20 ml. Complete M9 media was warmed to 37ºC. For the inoculum, 10 ml of overnight culture was centrifuged at 3,400 *g* for 5 minutes at 20ºC. The culture supernatant was removed and the pellet was resuspended in 1 ml warmed M9 media. The resuspended cells were used to inoculate the M9 media. When the culture had grown to an OD_{600} of 0.3 the culture was moved to 18ºC. The culture was allowed to cool for 30 minutes and IPTG was added to a concentration of 50 μM. Cultures were grown at 18ºC with aeration overnight. The CexE-6His protein was

purified as previously described. The double labelled CexE-6His was further purified by size exclusion prior to NMR experiments.

2.8.2. NMR experiments

A 1.83 mM sample of purified CexE-6His was dialysed into 20 mM NaP, 150 mM NaCl, 50 mM L-arginine, 50 mM L-glutamine, 0.5 mM TCEP and 0.2 mM NaN_3 at pH 6.5. The following experiments, using a 600 MHz magnet, were used for the determination of the structure of CexE-6His; ^{1}H-^{15}N-HSQC, HNCA, HNCOCA, HNCACB, HN(CO)CACB, HNCO, HNCACO, CCONH and HCCONH. The 900 MHz magnet was used for the final experiments, ^{1}H-^{15}N-HSQC, ^{13}C- NOESY-HSQC and ^{15}N-NOESY-HSQC.

2.8.3. Shift assignments and structure calculation

The backbone of the protein consisting of the amide, Cα and the carboxyl groups were assigned using 1H-15N-HSQC, HNCA, HNCOCA, HNCACB, HN(CO)CACB, HNCO and HNCACO. The shifts for the Cα were assigned using the HNCA and HN(CO)CA spectrum. The chemical shifts for the Cβ were assigned using the HNCACB and HN(CO)CACB spectra. The carbonyl groups were assigned using the HNCO and HNCACO spectra. The carbon side chains of the amino acids were assigned using the CCONH spectrum. The ^{13}C- NOESY-HSQC was assigned using the chemical shifts determined from the experiments above. The ^{15}N-NOESY-HSQC was not assigned. The CexE-6His structure was modelled using CYANA by Dr. Timothy Knowles.

2.9. ^{32}P Radioactive labelled DNA

2.9.1. DNA fragment isolation

Overnight cultures of strains containing pSR*CC*(−41.5) were used for plasmid isolation as

previously described. For the excision of the 100 bp fragment from pSR*CC*(−41.5) plasmid the following reaction was set up: 40 µl pSR*CC*(−41.5), 2 µl EcoRI-HF® (NEB), 2 µl HindIII-HF® (NEB), 5 µl CutSmart® buffer (NEB) and 1 µl nuclease-free water. This reaction was set up in quintuplicate. The reaction was incubated at 37ºC for 2 hours. After 2 hours, restriction enzymes were heat inactivated by incubation of 85ºC for 5 minutes. After heat inactivation samples were cooled. Then 2 µl Alkaline Phosphatase, Calf Intestinal (CIP) (NEB), 1 µl CutSmart® buffer (NEB) and 7 µl nuclease-free water was added to each reaction. The reaction was incubated at 37ºC for a further two hours. For loading onto a 0.5x TBE 6% acrylamide gel 15 µl 5x DNA Loading Buffer Blue (Bioline) was added to each sample. A working concentration of 0.5x TBE was made from a 5x TBE stock (54 g/L Tris base, 27.5 g/L boric acid and 20 ml/L 0.5 M EDTA (pH 8.0)). The gels were electrophoresis at 100 V. When the dye front was towards the bottom of the gel, the gels were removed from their casts and incubated in 0.5x TBE with Midori Green Advance (Nippon Genetics) for 5 minutes at room temperature. The bands were visualised and excised. The DNA was electroeluted from the excised gel fragments by placing the gel into dialysis tubing. A 100 V current was put across the tubing for 30 minutes. The DNA fragments were purified by phenol-chloroform precipitation, detailed below.

2.9.2. Phenol-chloroform DNA precipitation

Phenol-Chloroform mix was added 1:1 to DNA sample and vortexed vigorously. DNA was spun at 20,000 *g* for 2 minutes. After centrifugation, the upper layer was removed by pipetting and placed into a new tube. 3M sodium acetate was added to one tenth of the volume. 1M $MgCl_2$ was added to one hundredth of the volume. Two volumes of 100% ice-cold ethanol were added to the DNA. DNA was incubated at −20ºC for 2 hours or −80ºC for 15 minutes. The DNA was then centrifuged at max speed for 15 minutes at 4ºC. Supernatant was removed

and the pellet was washed with 1 ml of 70% ice-cold ethanol. DNA was centrifuged at max speed for 15 minutes at 4ºC. The supernatant was removed and the remaining ethanol allowed to evaporate. Pellet was resuspended in 40 µl nuclease free water.

2.9.3. ^{32}P DNA labelling

All of the *CC*(−41.5) DNA fragments were incubated with 2 µl T4 Polynucleotide Kinase (NEB), 5 µl T4 PNK Reaction Buffer (10X) (NEB) and 3 µl ATP, [γ-32P] (PerkinElmer). The final volume was made up to 50 µl. The reaction was incubated at 37ºC for 15 minutes. T4 PNK was heat inactivated at 65ºC for 5 minutes. The DNA fragments were isolated using silicone columns.

2.9.4. Electrophoretic mobility shift assay

Electrophoretic Mobility shift assays were carried out as follows. Reactions were prepared in 10 µl with the following concentrations 0.2 µl ^{32}P labelled CC−41.5 DNA, 1 µl 10x Fis buffer (200 mM Tris-HCl (pH 7.0), 100 mM MgCl2, 1 mM EDTA, 1.2 M KCl and 20 mM DTT), 1 µl 50% glycerol, 0.5 µl BSA (10mg/ml making a final concentration of 7 µM), 1 µl protein diluted to the appropriate concentration and 6.3 µl H_2O. Reactions were left at room temperature for 15 minutes and run on 0.5x TBE 6 % acrylamide gel. A lane was reserved for 5x DNA Loading Buffer Blue (Bioline). When the dye front was sufficiently progressed, the gel was removed. The gel was placed onto filter paper. The other side was covered with cling film and the gel was dried. The dried gel was then placed into a cassette facing the white side of a phosphor screen and left overnight. The phosphor screen was visualised.

CHAPTER 3:

Characterisation of the AggR regulon

3.1. Introduction

A pathogen must be able to sense the host environment and change protein production accordingly. The expression of genes, encoding proteins required for virulence, is mainly modulated via transcription factors. Enteroaggregative *E. coli* (EAEC) are a major cause of disease worldwide (Huang et al., 2006; Rohde et al., 2011). EAEC form a characteristic 'stacked brick' biofilm on HEp-2 cells mediated by the aggregative adherence Fimbriae (AAF) (Nataro et al., 1987, 1992). The AAF and their transcription activator, AggR, are required for colonisation (Elias et al., 1999; Clements et al., 2012). In the prototypical EAEC strain, EAEC 042, the genes that encode AggR and AAF/II are present on a plasmid, pAA2 (Chaudhuri et al., 2010). The pAA2 plasmid contains many genes encoding proteins required for the pathogenesis of EAEC 042.

In previous studies on EAEC 042, Dudley et al. (2006) identified AggR as an activator for the transcription of a cluster of 25 genes termed *aaiA-Y* that are present on a pathogenicity island on the chromosome. These genes encode proteins that are homologous to a type VI secretion system (T6SS). AggR can induce transcription of genes on both the pAA2 plasmid and the chromosome. From this, Dudley et al. (2006) concluded that AggR is a global regulator. Using qRT-PCR and microarrays, Morin et al. (2013) identified 44 genes under positive transcriptional control of AggR in EAEC 042. These genes include the *aai* system, *aafABCD*, the *aat* system and *aap*. A number of other hypothetical genes were also shown to be under AggR control. In both of these studies, microarrays were used to provide information about the AggR regulon. Advances in sequencing technology have provided more sensitive techniques for the transcriptional profiling of an organism. In this study, RNA-Seq has been used to further characterise the AggR regulon.

RNA-Seq offers a number of benefits over previous methods of transcriptome profiling. Microarrays and qRT-PCR require hybridisation of probes to cDNA to identify transcripts. In contrast, in RNA-Seq the cDNA is sequenced and aligned to a reference genome. The alignment of reads to a reference genome is a quantitative method for the determination of abundance of individual transcripts. Microarrays and qRT-PCR use fluorescence to identify transcript hybridisation to specific probes. The reliance on fluorescence reduces the dynamic range of the technique. Transcripts with a low abundance will not be detected while transcripts with high abundance will be under-exposed (Cloonan and Grimmond, 2008). Non-specific hybridisation and probe bias can both influence the results from a microarray experiment but are not problematic for RNA-Seq (Kane et al., 2000; Cloonan and Grimmond, 2008). Thus, in RNA-Seq experiments, this bias is removed and the dynamic range is increased.

In an RNA-Seq experiment, RNA is extracted and contaminating DNA removed by DNase I treatment (Fig. 3.1). To improve the power of the experiment ribosomal RNA (rRNA) is removed from the sample. The high abundance of rRNA in a transcriptome (up to 95%) results in the majority of reads being aligned to genes for ribosomal RNA instead of mRNA (Giannoukos et al., 2012). Depletion of the rRNA from an RNA sample reduces the cost of sequencing and increases the power of the experiment for the identification of mRNA transcripts. cDNA is synthesised from the depleted RNA and is sequenced. The reads are aligned to the reference genome. The number of reads that align to each gene is counted. The count for each gene is compared between either growth conditions or a parental strain and a mutant. From this comparison genes that are differentially expressed can be identified. In this study, RNA-Seq was used to confirm the previously identified AggR regulon and identify new genes under the control of AggR (Dudley et al., 2006; Morin et al., 2013).

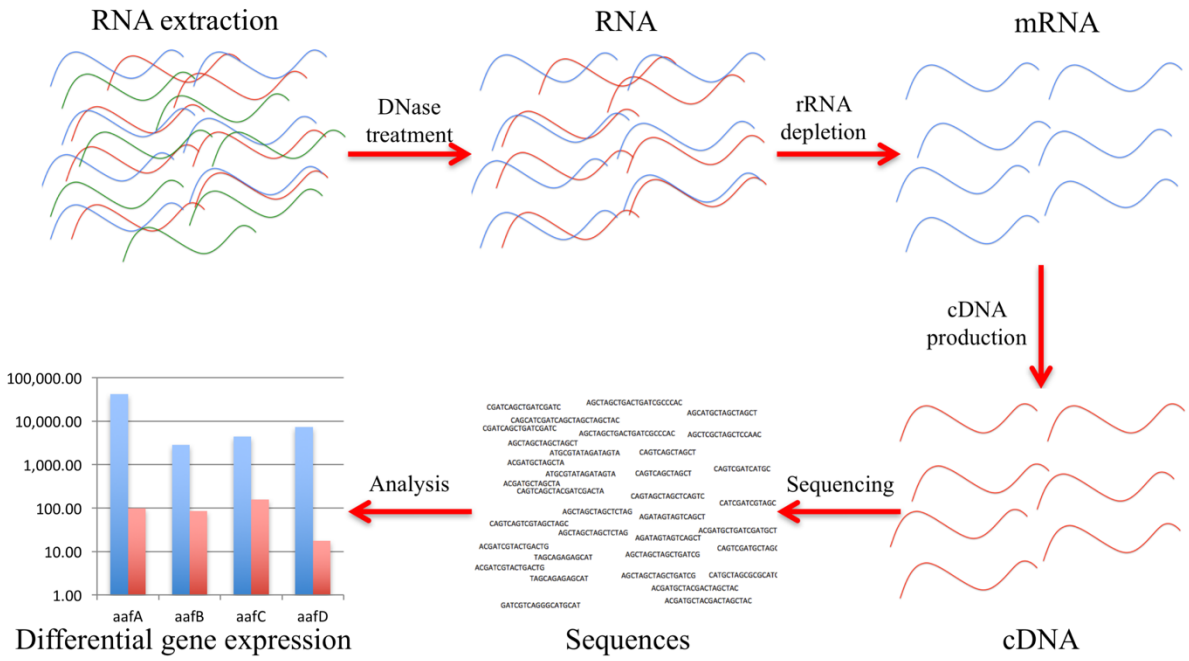


Figure 3.1: The general outline of an RNA-Seq experiment. RNA is harvested from a bacterial culture. Contaminating DNA is removed by DNase I treatment. Ribosomal RNA is depleted, cDNA is then synthesised and the library is sequenced. The resulting reads provide a quantitative method to measure the abundance of each transcript.

3.2. Results

3.2.1. Comparison of growth of an *aggR* mutant with the parental strain.

Ideally RNA for RNA-Seq analysis should be isolated from the test and control cultures of bacteria growing exponentially at similar rates. Initial experiments were therefore designed to determine the stage of growth at which samples of the *aggR* mutant and its parent strain should be harvested for RNA isolation. Triplicate overnight cultures were transferred to fresh DMEM high glucose (DMEM) and grown at 37°C with aeration. Samples were removed at 30 min. intervals, and a growth curve was constructed (Fig. 3.2). Both cultures grew exponentially at almost identical rates until the OD_{600} had reached 0.6. Growth of the parent strain then gradually stopped, reaching a final cell density of 0.8. In contrast, growth of the *aggR* mutant continued to a final cell density of about 1.3. It was concluded that in subsequent experiments, samples for RNA isolation should be harvested at an OD_{600} of 0.6.

3.2.2. RNA isolation and DNA removal

Triplicate overnight cultures of the *aggR* mutant and the parental strain were transferred to DMEM and grown at 37°C with aeration, as described above. At an OD_{600} of 0.6, a 2 ml sample from each culture was harvested by centrifugation. RNA was isolated from the bacterial pellet. Following removal of contaminating DNA from the samples using DNase I, PCR was used to check for any residual DNA contamination. PCR primers 27F and 1492R were used to amplify the DNA encoding 16S ribosomal RNA. Each of the following were used as the template for separate PCR reactions: 2 µl of each of the six RNA samples; a positive control of 2 µl of *Staphylococcus aureus* chromosomal DNA; and a negative control of 2 µl of nuclease free water. The PCR products were separated by agarose gel electrophoresis. No PCR product was amplified for any of the RNA samples or the nuclease free water. A product of 1,500 bp was

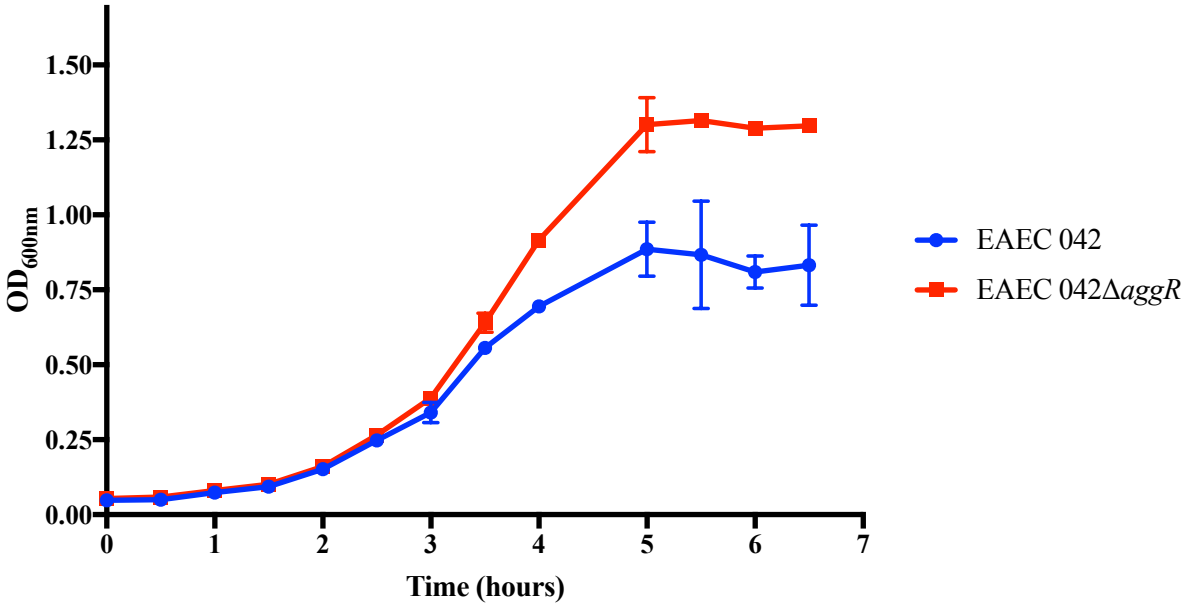


Figure 3.2: Growth curve of *aggR* mutant and parental strain in DMEM high glucose. Triplicate overnight cultures were inoculated in DMEM high glucose and grown at 37ºC with aeration. The growth curve was constructed with OD_{600} readings every 30 minutes. Each time point was the mean of the OD_{600} readings taken. Error bars represent the standard deviation between replicates.

amplified for *S. aureus* chromosomal DNA (Fig. 3.3). From this, it was concluded that there was no DNA present in the RNA samples that was detectable by PCR.

3.2.3. Quantity and quality of isolated RNA

Only concentrated, high-quantity RNA samples will provide useful information on the expression of genes. The quantity of each RNA sample was tested using Qubit™ RNA BR Assay kit. The total quantities of RNA in each 50 µl sample A to F were 8.1, 26.58, 14.64, 11.34, 7.02 and 11.22 µg, respectively (Table 3.1). To determine the quality of each sample 1 µl was loaded into a separate well of the Agilent 6000 Nano Chip. The RNA Integrity Number (RIN) for each sample was 9.5, 9.8, 9.1, 10, 8.1 and 8.5, respectively (Table 3.1). The first major peak at around 1,500 nucleotides (nt) is the 16S rRNA the second peak at 2,000 nt is the 23S rRNA. No signs of degradation in sample D resulted in a perfect RIN score of 10 (Fig. 3.4). Samples E and F were of a lower quality. Degradation products of the 16S were present at 1,000 nt. Although the RNA in samples E and F were slightly degraded, this was not substantial. A RIN of above 8 was required for RNA-Seq. Thus, the RNA samples were of a sufficient quantity and quality for use in RNA-Seq.

3.2.4. Ribosomal RNA depletion and depleted RNA quantification

Removal of rRNA from RNA samples prior to cDNA synthesis increases the dynamic range of the experiment. Ribosomal RNA was depleted from 3.5 µg of total RNA from samples A to F using the RiboZero rRNA Removal kit (Bacteria) from Illumina®. After depletion of rRNA, 1 µl of each sample was analysed using the Agilent RNA 6000 Pico kit. For the samples A, C, D, E and F the peaks for 16S and 23S rRNA were completely depleted (Fig. 3.5). In sample B, the rRNA peaks were still present, indicating that the rRNA depletion of sample B was therefore incomplete. However, there were new peaks for sample B that were not present in

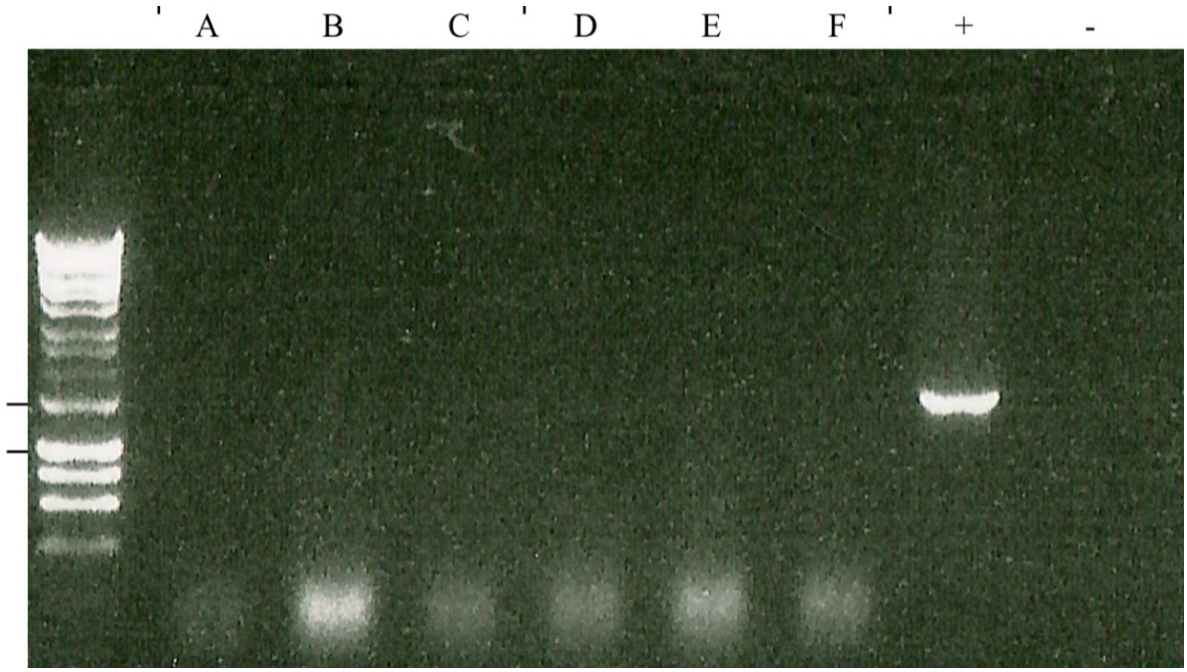


Figure 3.3: Detection of ribosomal DNA in RNA samples by PCR. No ribosomal DNA was detectable by PCR for any of the six samples. The size of the DNA fragments was estimated using Bioline 1kb Hyperladder. The product generated in the positive control (+) at 1,500 bp was the result of the successful amplification of the 16S gene from a genomic DNA sample of *S. aureus*. No band was present for the negative control (-) of nuclease free water.

Table 3.1: Summary of RNA isolation, cDNA synthesis and sequencing.

Strains	EAEC 042			EAEC 042Δ*aggR*		
RNA Samples	A	B	C	D	E	F
Total RNA isolated (μg in 50 μl)	8.1	26.58	14.64	11.34	7.02	11.22
RIN	9.5	9.8	9.1	10	8.1	8.5
Total RNA after depletion (ng in 20 μl)	3.29	27.2	23.5	29.3	26.3	34.8
Average fragment size (bp)	280	287	268	265	279	281
Concentration of cDNA (ng/μl)	3.56	82.14	18.01	21.59	18.12	21.45
Generated Reads	8,385,714	12,508,003	7,853,859	8,204,480	8,867,535	8,361,035
Filtered Reads	8,217,784	12,255,016	7,752,998	8,095,882	8,701,142	8,201,872
Percentage of reads successfully aligned (%)	96.66	93.26	92.57	90.48	90.78	93.26
Number of reads aligned to genes in the correct orientation	6,333,037	3,719,671	6,023,878	6,208,206	6,726,054	6,378,786
Percentage of aligned reads that aligned to genes in the correct orientation (%)	77.0	30.3	77.6	76.6	77.2	77.7

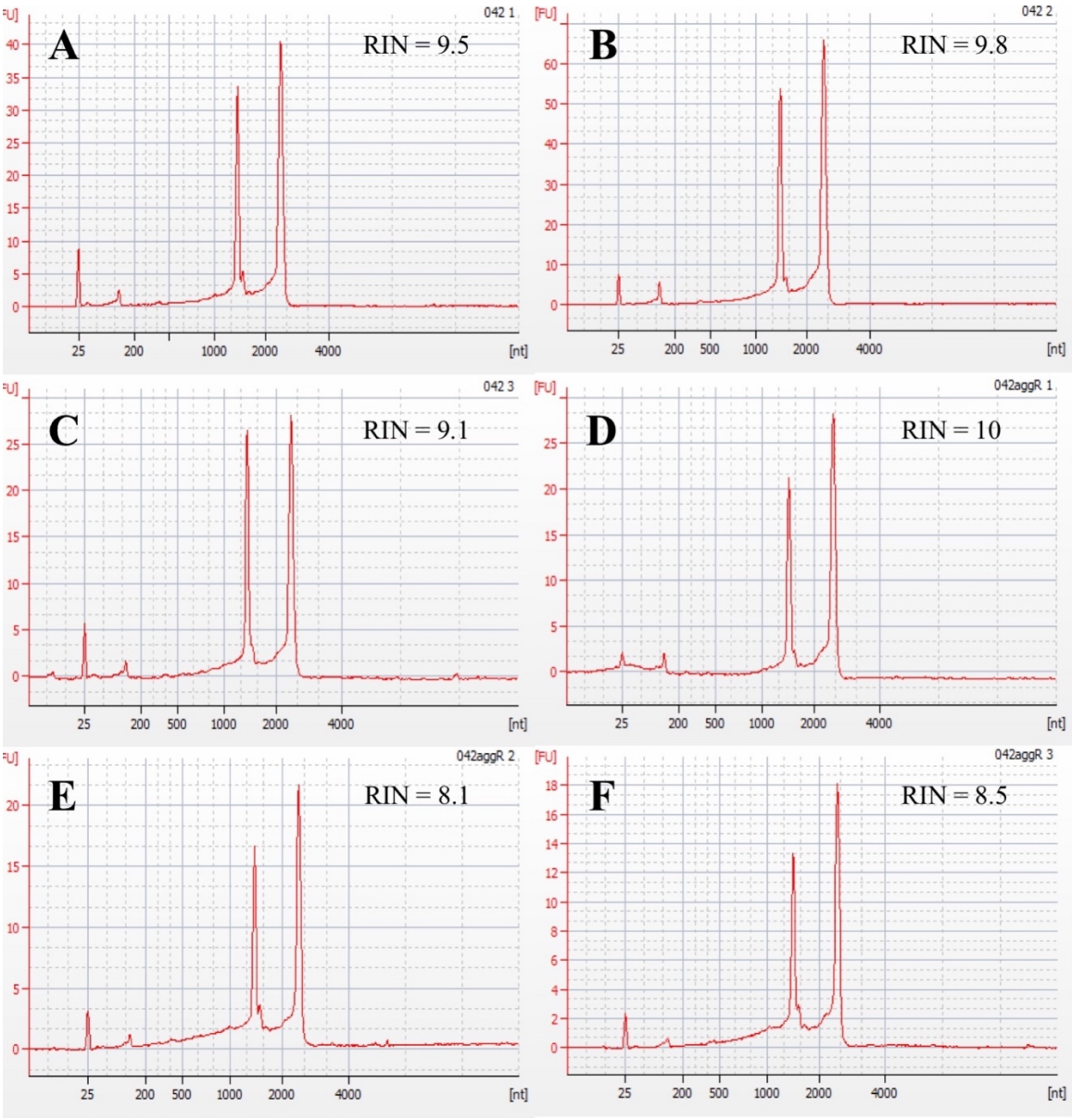


Figure 3.4: Total isolated RNA from EAEC 042 and EAEC 042Δ*aggR* replicates. The panels are labelled based on the sample. Three biological replicates of EAEC 042 (A, B and C) and EAECΔ*aggR* (D, E and F) are shown. For each sample, 1 µl was used to determine the RIN with the Agilent Bioanalyzer 6000 Nano Chip. The RIN is indicated on the plot for each sample.

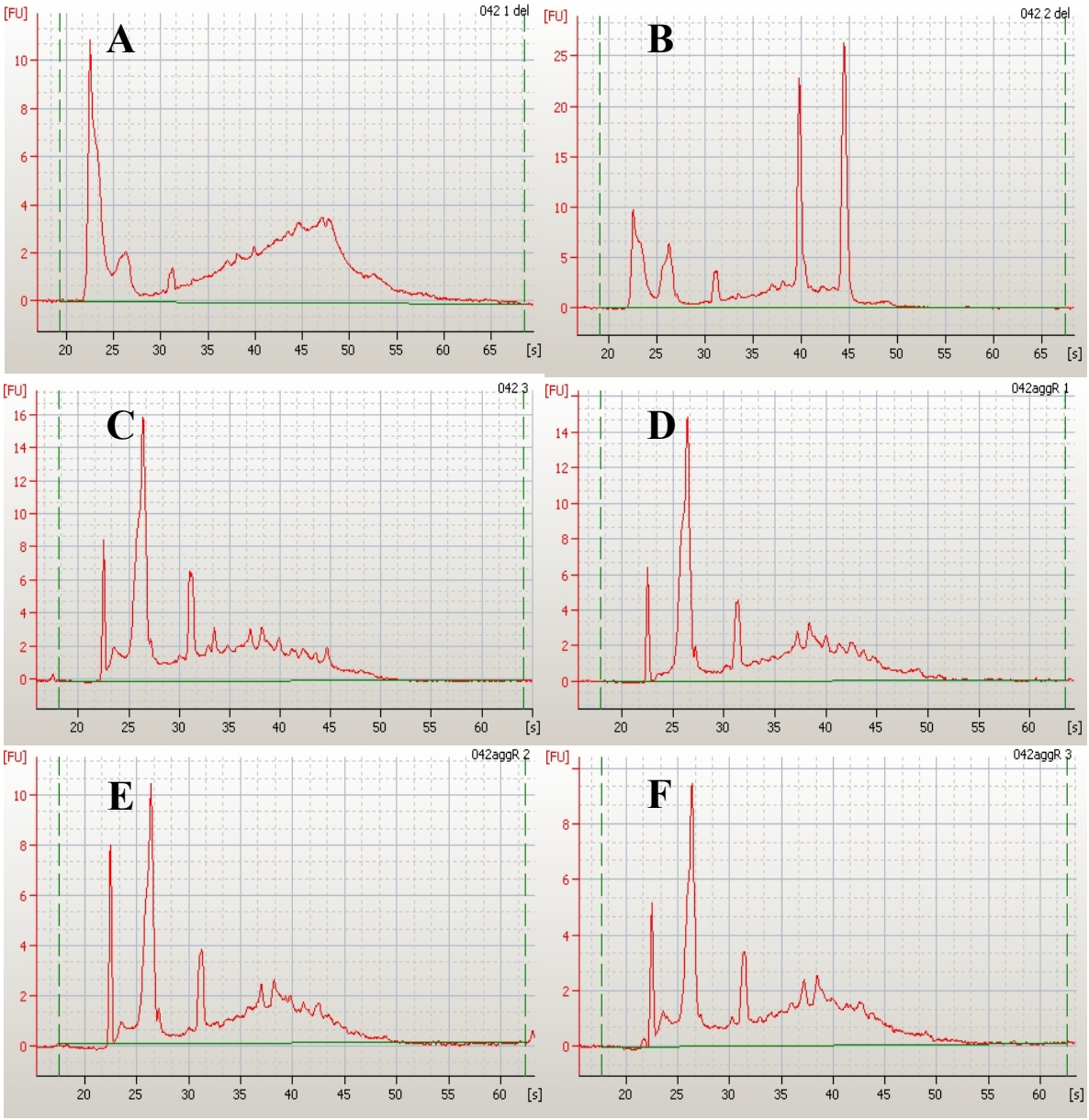


Figure 3.5: Post rRNA depletion of RNA of EAEC 042 and EAEC 042Δ*aggR*. Samples A to F on were analysed for rRNA depletion using the Agilent Bioanalyzer Pico 6000 RNA chip. In all of the samples, except sample B, rRNA was successfully depleted. Significant amounts of rRNA were still present in sample B. However, some rRNA had been depleted as peaks representing mRNA, that were not previously observable in Fig. 3.4, were visible.

Fig. 3.4. The fluorescence of the 16S and 23S peaks were also reduced from 53 and 66 FU to 22.5 and 25 FU (Fig. 3.4 and Fig. 3.5). Therefore, the rRNA has been partially depleted in sample B. The rRNA still present in sample B is not ideal but reads that align to the rRNA could be removed in the analysis and would therefore not invalidate the final result. Thus, the RNA samples were considered to be sufficiently depleted to continue with cDNA synthesis.

The amount of depleted RNA must be quantified for cDNA synthesis. RNA was quantified using the Qubit™ RNA HS kit. From the samples A to F, 1 µl was taken after rRNA depletion to assess the amount of RNA. The total depleted amount of RNA in 20µl was found to be 3.29, 27.2, 23.5, 29.0, 26.3 and 34.8 ng, respectively (Table 3.1). Despite the low quantity of RNA in sample A, it was still sufficient for cDNA synthesis.

3.2.5. Synthesis of cDNA

RNA must be converted into cDNA before it can be sequenced. From each of the samples B to F, 10 ng of RNA was used for cDNA synthesis. For sample A, the entire amount of 3.29 ng was used. RNA was reverse-transcribed into cDNA using True-Seq Stranded Total RNA LT Sample Prep Kit. For DNA sequencing, the average fragment size should be 260 bp. Using the Agilent High Sensitivity DNA Kit, the average fragment size was found to be 280bp, 287bp, 268bp, 265bp, 279bp and 281bp for samples A to F (Fig. 3.6, Table 3.1). The concentration of cDNA for each sample was 3.56, 82.14, 18.01, 21.59, 18.12 and 21.45 ng/µl, respectively (Table 3.1). For samples B, D and F a subpopulation of the DNA library appeared to have a larger fragment size (Fig. 3.6). However, the majority of the fragments were an acceptable size for sequencing. The low input amount of RNA for sample A resulted in a low yield of cDNA. There were a number of other peaks present in sample A compared to the other samples.

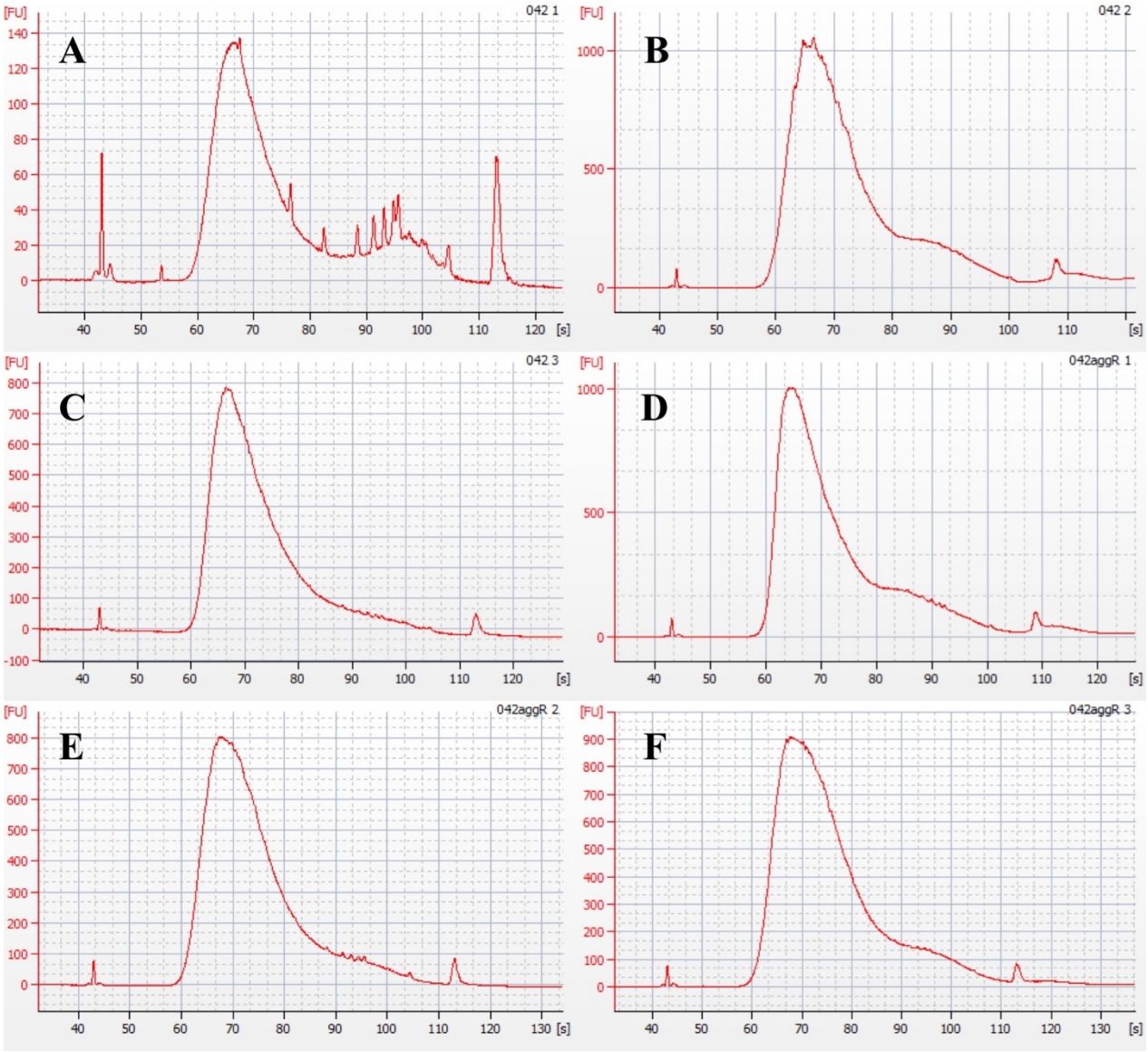


Figure 3.6: Reverse transcribed cDNA fragments of wild-type and *aggR* replicates. The average fragment size for each sample was determined using the Agilent High Sensitivity DNA Kit. The lower concentration of cDNA in sample A was due to low concentration of depleted RNA in the sample. Note the extra peaks present in sample A compared to the rest of the samples.

However, only a small amount of cDNA, 4 pM, is required for sequencing. Thus, the cDNA synthesised was sufficiently concentrated for sequencing.

3.2.6. Sequencing and alignment of reads to EAEC 042 genome

The cDNA was pooled and sequenced using the MiSeq Reagent Kit v3 (Illumina®). Pair-end reads of 75 bases were generated. The samples A to F generated 8,385,714, 12,508,003, 7,853,859, 8,204,480, 8,867,535 and 8,361,035 reads, respectively (Table 3.1). The higher number of reads generated from sample B was probably due to a higher concentration of cDNA. Despite the lower concentration of cDNA in sample A, a relatively similar number of reads was obtained after sequencing. Each sample generated enough reads to provide useful information about the AggR regulon.

Low quality reads were removed using Trimmomatic-0.36. The number of reads for samples A to F after filtering were 8,217,784, 12,255,016, 7,752,998, 8,095,882, 8,701,142 and 8,201,872, respectively (Table 3.1). The majority of the reads, 98%, were of a sufficient quality for genome alignment. The reads were then aligned to the EAEC 042 genome using Burrows-Wheeler Aligner. The percentage of reads successfully aligned for samples A to F to the genome was 96.66, 93.26, 92.57, 90.48, 90.78 and 93.26, respectively (Table 3.1). Thus, in conclusion, there was sufficient coverage of the transcriptome for differential gene expression analysis.

3.2.7. Differential gene expression analysis.

Differential expression of a gene is calculated by comparing the number of reads aligned to that gene in the mutant with the parental strain. The amount of reads that align to each gene was counted using featureCounts. Reads that aligned to ribosomal RNA were not counted. For

samples A to F the total number of reads that aligned to genes was 6,333,037, 3,719,671, 6,023,878, 6,208,206, 6,726,054 and 6,378,786, respectively (Table 3.1). The percentage of aligned reads that aligned to genes in the correct orientation for each sample was 77.07, 30.35, 77.70, 76.68, 77.30 and 77.77, respectively (Table 3.1). The lower number of reads mapped for sample B was due to the high level of rRNA after rRNA depletion (Fig. 3.5). Although, this was not ideal, this could be resolved by normalisation of transcripts per gene.

Differential gene expression was calculated using the R package DESeq2. Genes that contained no aligned reads in any of the samples were removed. The number of aligned reads to each gene was normalised. The number of reads that aligned to a gene was transformed onto a $\log_2$ scale. The amount of variance of genes with low but variable read counts between the different samples was reduced by transformation, using the DESeq2 rlog function. In Fig. 3.7, each gene is represented by a point, dictated by the number of reads after transformation and normalisation for EAEC 042 on the y-axis and EAEC 042Δ*aggR* on the x-axis. Genes that were not differentially expressed in the *aggR* mutant compared to the parent strain form a straight line out of the origin at an angle of 45º. Genes activated by AggR fall above this line; those repressed by AggR fall below this line. Therefore, AggR activates transcription for more genes than it represses.

The majority of genes, 4,908, were not differentially expressed in the *aggR* mutant compared to the parent strain. A total of 112 genes present on pAA2 and the chromosome were found to be differentially expressed, of which 4 were repressed by AggR: *phoA*; *flu1*; *flu2*; and EC042_4803, signified by a positive fold change in the wild-type compared to the mutant. Table 3.2 includes all of the genes that were identified by Morin et al. (2013) to be differentially

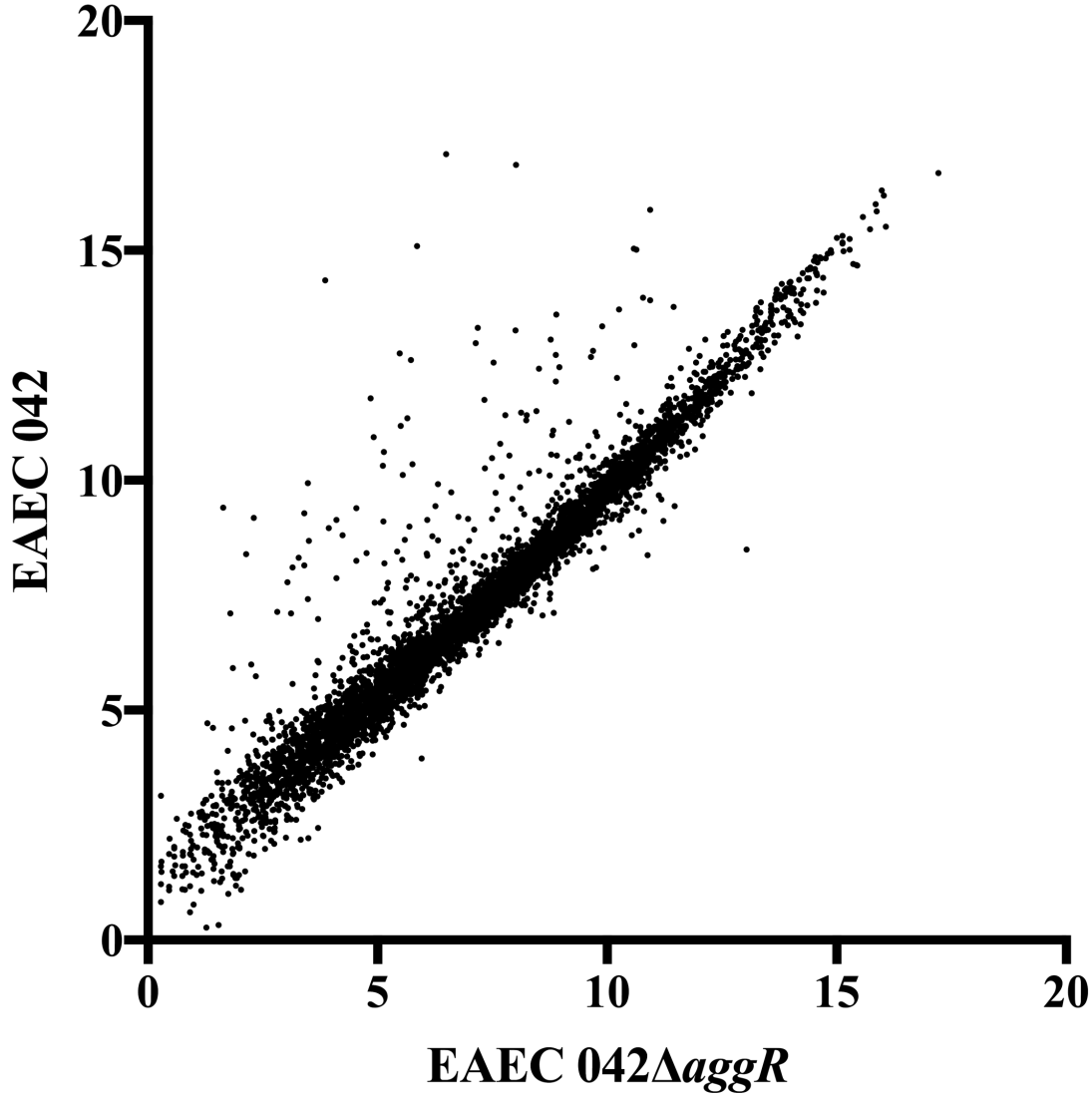


Figure 3.7: Read counts of genes in EAEC 042 compared to EAEC 042Δ*aggR*. Sequencing reads aligned to each gene were counted using FeatureCounts. The counts were normalised in DESeq2. The counts for each gene were averaged across EAEC 042 and EAEC 042Δ*aggR* replicates and plotted against each other. Each gene is represented by a point.

Table 3.2: Genes differentially expressed in the *aggR* mutant.

Chromosomal genes			
Gene Name	Function	$\log_2$ fold change	Adjusted p-value
phoA	alkaline phosphatase	1.44	4.97E−09
bssS	biofilm regulator	−2.76	2.29E−15
flgN	negative regulator of flagellin synthesis	−2.12	4.21E−09
flgM	anti-sigma factor	−1.97	4.60E−06
flgA	flagellar basal body P-ring formation protein	−3.29	2.92E−22
flgB	flagellar basal-body rod protein	−6.05	1.72E−51
flgC	flagellar basal-body rod protein	−5.60	2.40E−47
flgD	flagellar biosynthesis, initiation of hook assembly	−5.19	2.67E−42
flgE	flagellar hook protein	−5.23	1.06E−52
flgF	flagellar basal-body rod protein	−4.27	1.34E−31
flgG	flagellar basal-body rod protein	−4.33	6.68E−40
flgH	flagellar L-ring protein	−3.32	3.01E−21
flgI	flagellar P-ring protein	−2.74	3.81E−20
flgJ	putative peptidoglycan hydrolase	−2.78	6.54E−19
flgK	peptidoglycan hydrolase	−2.37	1.80E−14
flgL	flagellar hook-filament junction protein 1	−2.17	9.47E−11
cspI	cold shock-like protein	−2.02	4.47E−07
anmK	anhydro-N-acetylmuramic acid kinase	−1.18	6.58E−06
flhA	flagella biosynthesis protein	−2.13	1.40E−11
flhB	flagellar biosynthesis protein	−3.13	4.47E−15
cheW	chemotaxis protein	−1.86	6.29E−07
cheA	chemotaxis protein	−2.03	8.37E−11
flhC	flagellar biosynthesis protein	−2.13	1.34E−17
flhD	flagellar transcriptional activator	−2.05	3.35E−18
fliZ	DNA-binding transcriptional regulator	−3.35	6.62E−18
fliA	flagellar sigma factor	−5.09	3.97E−38
fliE	flagellar hook-basal body complex protein	−2.93	1.88E−17
fliF	flagellar M-ring protein	−4.99	2.38E−47
fliG	flagellar motor switch protein	−4.48	3.43E−36
fliH	flagellar assembly protein	−4.12	2.30E−31
fliI	flagellum-specific ATP synthase	−4.11	3.44E−28
fliJ	rod/hook and filament chaperone	−3.62	8.58E−22
fliK	flagellar hook-length control protein	−3.90	4.79E−22
fliL	flagellar basal body-associated protein	−4.08	1.13E−28
fliM	flagellar motor switch protein	−4.43	5.29E−42

fliN	flagellar motor switch protein	−4.19	2.08E−24
fliO	flagellar protein	−3.27	9.41E−14
fliP	flagellar biosynthetic protein	−2.63	1.32E−08
EC042_2219	conserved hypothetical protein	−2.07	1.07E−07
flu1	antigen 43	1.99	8.24E−20
EC042_2249	conserved hypothetical protein	−3.00	7.94E−12
EC042_3179A	conserved hypothetical protein	−1.57	6.27E−08
EC042_3180	conserved hypothetical protein	−2.88	1.09E−30
EC042_3181	putative transcriptional regulator	−2.44	4.95E−15
EC042_3182*	ParB-like nuclease	−3.38	4.19E−45
EC042_3183	conserved hypothetical protein	−4.35	2.26E−45
EC042_3184*	conserved hypothetical protein	−3.73	2.43E−25
EC042_3185	pseudogene	−3.99	6.55E−34
EC042_3187	putative helicase	−4.09	1.86E−78
EC042_3188	transposase	−4.91	6.15E−61
EC042_3476	conserved hypothetical protein	−2.27	5.13E−06
EC042_4006*	putative exported protein	−4.16	9.37E−43
EC042_4430	putative signal transduction protein	−2.89	1.59E−23
flu2	antigen 43	2.32	4.49E−14
EC042_4562*	putative type VI secretion protein	−4.41	2.72E−74
EC042_4563*	putative type VI secretion protein	−4.14	6.55E−59
EC042_4564*	putative type VI secretion protein	−3.99	2.88E−43
EC042_4565*	putative type VI secretion protein	−3.33	1.00E−34
EC042_4566*	putative type VI secretion protein	−3.14	2.04E−29
EC042_4568*	putative type VI secretion protein	−3.00	2.24E−27
EC042_4569*	putative type VI secretion protein	−3.21	1.51E−33
EC042_4570*	putative type VI secretion protein	−3.16	4.03E−41
EC042_4571*	putative type VI secretion protein	−2.75	1.40E−20
EC042_4572*	putative type VI secretion protein	−2.86	3.34E−32
EC042_4573*	putative type VI secretion protein	−2.99	1.05E−35
EC042_4574*	putative type VI secretion protein	−2.90	1.96E−27
EC042_4574A*	putative type VI secretion protein	−2.74	1.64E−28
EC042_4575*	putative type VI secretion protein	−2.86	1.46E−33
EC042_4576*	putative type VI secretion protein	−2.94	1.48E−26
EC042_4577*	putative type VI secretion protein	−2.78	6.92E−27
EC042_4578	pseudogene - transposase	−1.22	8.93E−06
EC042_4579A	pseudogene - transposase	−1.88	1.18E−12
EC042_4580*	conserved hypothetical protein	−1.21	1.39E−07
EC042_4581*	conserved hypothetical protein	−2.19	6.55E−18
EC042_4581A	pseudogene - transposase	−1.98	2.46E−16
EC042_4581B	pseudogene - transposase	−1.67	4.97E−09
EC042_4582*	hypothetical protein	−1.91	1.45E−17
EC042_4583*	hypothetical protein	−1.70	6.01E−16

EC042_4584	pseudogene - transposase	−1.67	1.86E−08
EC042_4585	transposase	−1.68	6.97E−09
EC042_4585A	pseudogene - transposase	−1.83	1.26E−08
EC042_4587	pseudogene - transposase	−1.57	3.54E−09
EC042_4803	antigen 43	4.29	3.63E−81

pAA2 genes

Gene Name	Function	log_2 fold change	Adjusted p-value
EC042_pAA003*	hypothetical protein	−4.54	1.51E−56
EC042_pAA004*	putative isopentenyl-diphosphate delta-isomerase	−4.13	4.81E−64
EC042_pAA005*	hypothetical protein	−5.34	1.53E−103
EC042_pAA005A*	hypothetical protein	−5.56	2.78E−80
EC042_pAA006	pseudogene - transposase	−1.58	9.06E−10
*aatP**	permease	−5.52	4.37E−50
*aatA**	TolC-like protein	−6.05	2.72E−85
*aatB**	membrane fusion protein	−4.68	2.68E−40
*aatC**	ATPase	−4.55	2.36E−48
*aatD**	required for Aap secretion	−3.47	2.06E−28
EC042_pAA019	conserved hypothetical protein pseudogene	−3.60	8.04E−50
EC042_pAA020*	conserved hypothetical protein	−3.28	5.15E−41
*shf**	polysaccharide deacetylase	−3.04	1.23E−32
EC042_pAA022*	glycosyl transferase	−3.24	1.42E−28
virK	virulence protein	−3.01	1.79E−29
*aafB**	afimbrial adhesin	−4.53	5.18E−43
*aafC**	aggregative adherence fimbria II usher protein	−4.53	6.08E−65
*afaB**	pseudogene	−6.68	2.91E−144
EC042_pAA033	transposase	−2.28	2.62E−06
*aafD**	chaperone protein	−7.49	8.21E−95
EC042_pAA047*	hypothetical protein	−9.19	3.65E−198
*aafA**	major fimbrial subunit of aggregative adherence fimbria II	−8.16	1.51E−208
EC042_pAA051	transposase	−4.94	1.37E−35
*aggR**	transcriptional activator	−5.47	2.40E−111
EC042_pAA053	pseudogene - transposase	−1.65	1.28E−10
*aap**	dispersin	−9.36	1.51E−208
*aar**	AggR antivirulence factor	−3.55	6.46E−33
EC042_pAA061	conserved hypothetical protein	−3.25	9.29E−24
EC042_pAA105	pseudogene - transposase	−2.05	5.12 E−10

Genes found to be differentially expressed by Morin et al. (2013) were marked with a *.

expressed in an *aggR* mutant compared to the parent strain except for EC042_pAA056, a hypothetical gene of unknown function.

3.2.8. Chromosomal genes differentially expressed in the *aggR* mutant compared to the parental strain

AggR is known to activate pathogenicity islands on the EAEC 042 chromosome (Dudley et al., 2006; Morin et al., 2013). A total of 112 genes were differentially expressed in the mutant compared to the parental strain, 83 of which are located on the chromosome and 39 on the pAA2 plasmid. Only 4 of the 83 genes appeared to be repressed by AggR; these were *flu1*, *flu2*, EC042_4803, and *phoA*. Three of these genes, *flu1*, *flu2*, and EC042_4803, encode Agn43 paralogs an autotransporter that promotes aggregation of cells in liquid broth (Diderichsen, 1980). Due to phase variation, these genes show variable levels of expression in randomly chosen populations of bacteria (Owen et al., 1996). Therefore, it is presumed that the difference in the number of transcripts for these genes, between the *aggR* mutant and the wild-type, is stochastic and simply represents phase variation events. PhoA is alkaline phosphatase. PhoA is produced in response to phosphate starvation (Torriani, 1960). No other genes involved in phosphate starvation were differentially expressed. Thus, the differential expression of *flu1*, *flu2*, EC042_4803, and *phoA* was not thought to be due to the presence of AggR.

Amongst the 83 differentially expressed genes on the chromosome, the products of 34 are involved in motility. A notable absence from the list of differentially expressed flagellar genes is *fliC*. The $\log_2$-fold change for *fliC* was −0.72. Differences in motility of the *aggR* mutant and wild-type have not been reported previously. The apparent activation of motility genes in the experiment is not thought to be due to the presence of AggR but due to the stochastic nature of bacterial motility expression (Spudich and Koshland, 1976; Korobkova et al., 2004).

In agreement with Dudley et al. (2006) and Morin et al. (2013) the 16 genes (EC042_4562 to EC042_4577) encoding a putative type 6 secretion system (T6SS), present in ROD69 (Chaudhuri et al., 2010), were differentially expressed. In fact, a total of 28 genes (EC042_4562 to EC042_4587) in ROD69 were differentially expressed in the *aggR* mutant compared to the wild-type. These 12 extra genes were 7 pseudogenes, 1 gene encoding a transposase, and 4 genes encoding hypothetical proteins. The 4 hypothetical proteins EC042_4580, EC042_4581, EC042_4582, and EC042_4583. EC042_4580, EC042_4581, and EC042_4583 are conserved in *E. coli* strains with no detectable homology to a family of proteins. The C and N-terminus of EC042_4582 are homologous to the pfam groups ImpA_N (PF06812) and VasL (PF12486), respectively. These pfam groups encode proteins of unknown function which are involved in T6SS and are commonly associated together.

Amongst the remaining differentially expressed chromosomal genes, 12 encode hypothetical proteins. Of the 12 genes, 8 formed a cluster on the chromosome. EC042_3179A to EC042_3187, which are in ROD48 (Chaudhuri et al., 2010), were differentially expressed in the presence of AggR (Fig. 3.8). The possible function of each of the proteins that the gene cluster encodes was investigated using a BLAST search. EC042_3179A, EC042_3180 and EC042_3184 are conserved proteins present in *E. coli* strains. There was no homology to a known family of proteins. EC042_3181 is a putative transcription regulator identified as homologous to PerC from enteropathogenic *E. coli* (EPEC). PerC directly activates the transcription of *LEE1* in EPEC (Knutton et al., 1997). EC042_3182 encodes a ParB-like nuclease domain. The N-terminal region of EC042_3183 is homologous to YbdN, a putative phosphoadenosine phosphosulfate sulfurtransferase. The C-terminal region of EC042_3182 contains a domain of unknown function, DUF3440. EC042_3185 is a pseudogene.

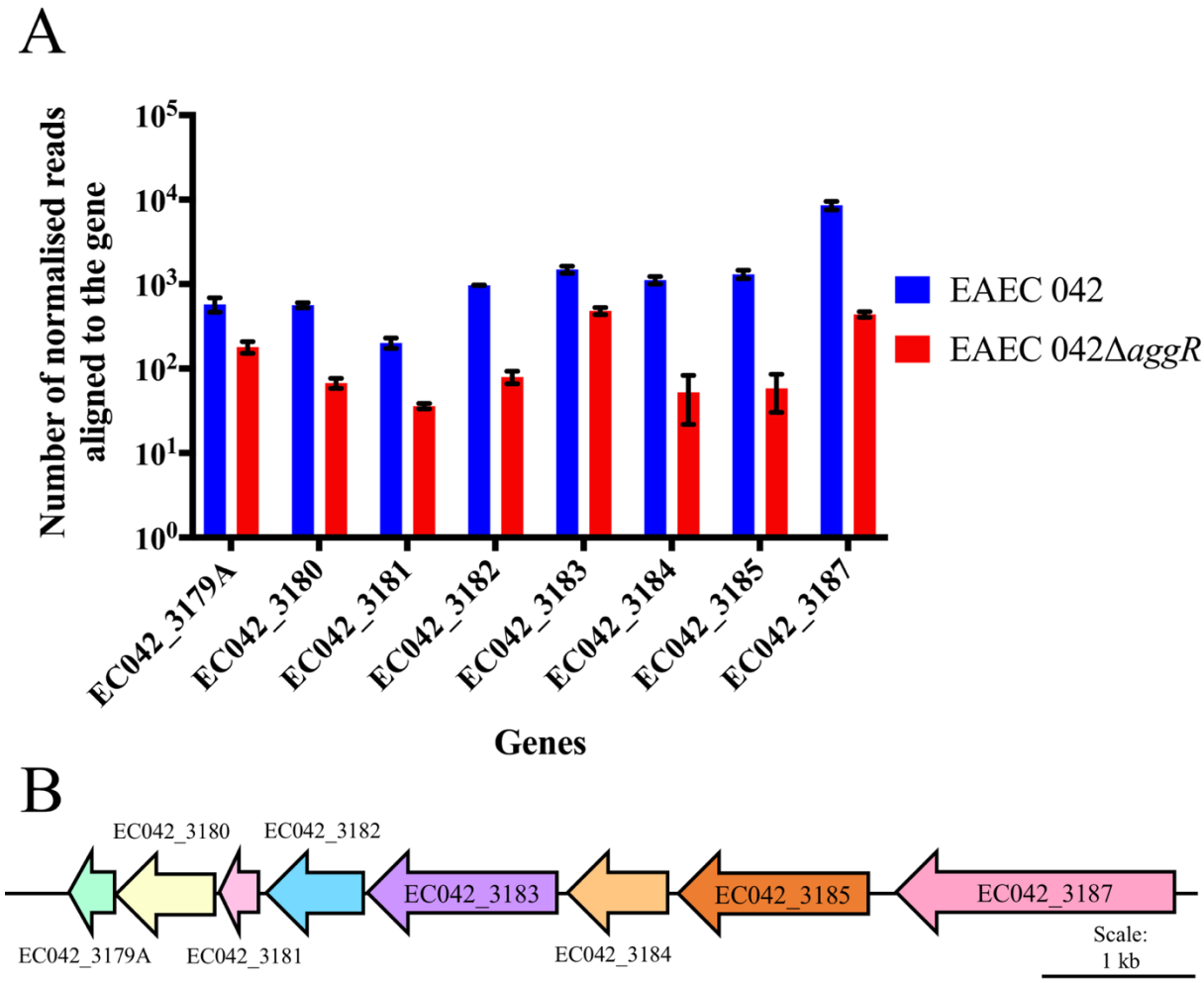


Figure 3.8: Expression of the genes EC042_3179A to EC042_3187. A, the number of normalised reads aligned to each gene in the *aggR* mutant and parental strain. The number of reads aligned to each sequence were normalised to the number of reads for each replicate. Thus, the normalised reads aligned to each gene are a measure the relative mRNA abundance of each gene. The error bars represent the standard deviation between the replicates of each strain. B, the genes EC042_3179A to EC042_3187 drawn to scale.

EC042_3187 was identified by BLAST as a UvrD-like helicase. Whether any of these proteins have a function in the colonisation of EAEC 042 is currently unknown. In conclusion, the expression of EC042_3179A to EC042_3187 was activated by AggR.

The remaining 4 differentially expressed hypothetical genes were not clustered together. The possible functions of the proteins encoded by these genes were investigated by BLAST searches. The C-terminus of EC042_2219 was homologous to the β domain of inverse autotransporters (pfam11924). EC042_2067 and EC042_3476 have no homology to proteins of known function. EC042_2067 is conserved in *E. coli* strains, both pathogenic and non-pathogenic strains. EC042_3476 is present in many *E. coli* pathogens. EC042_3476 is 99% identical to c3944 from UPEC CFT073 and EcE24377A_3672 from ETEC E24377A. However, while expression was activated by AggR their role in pathogenesis is unknown.

EC042_4006 is another differentially expressed gene encoding a hypothetical protein. The EC042_4006 product is a protein of 97 amino acids known as YicS in *E. coli* MG1655 K-12. EC042_4006 is adjacent to *nlpA*, *nlpA* is thought to be repressed by CfaD and AggR (Bodero et al., 2007). In this analysis, *nlpA* was not found to be differentially expressed. A $\log_2$-fold change of −0.74 was calculated for *nlpA*. It was concluded that, in agreement with previous studies, the expression of EC042_4006 was activated by AggR.

3.2.9. Differential expression of genes on the pAA2 virulence plasmid in the presence of AggR

On the pAA2 plasmid, 29 genes were differentially expressed. Of these genes, 9 encoded hypothetical proteins. There was a $\log_2$-fold difference of −4.54, −4.13, −5.34 and −5.56 for the genes EC042_pAA003 to EC042_pAA005A, respectively. Mutants of EC042_pAA003

and EC042_pAA004 affect the cell membrane and alter surface integrity, reducing EAEC 042 ability to form biofilms (Morin et al., 2013). EC042_pAA005 and EC042_pAA005A encode proteins of 50 amino acids. Using a BLAST search, no homology to a known family of proteins was identified. The function of EC042_pAA005 and EC042_pAA005A therefore remain unknown. Thus, in agreement with previous studies, EC042_pAA003 to EC042_pAA005A were confirmed to be activated by AggR.

The $\log_2$-fold differences of EC042_pAA019 to EC042_pAA023 were −3.60, −3.28, −3.04, −3.24 and −3.01, respectively. EC042_pAA019 is annotated as a pseudogene. However, an open reading frame of 107 residues is present for EC042_pAA019. The protein sequence of the open reading frame is conserved in *E. coli*, *Shigella* and *Salmonella*. EC042_pAA019 is homologous to YdfA, an uncharacterised conserved protein. No clear function has been associated with EC042_pAA019. EC042_pAA020 encodes a conserved hypothetical protein of 195 amino acids with no homology to a known family of proteins. Homologs of EC042_pAA020 were found in *E. coli*, *Salmonella* and *Klebsiella*. EC042_pAA021 is also known as *shf*. Shf is a polysaccharide deacetylase with homology to PgaB, an outer membrane N- deacetylase. EC042_pAA022 is a conserved glycosyl transferase similar to YqgM in *Bacillus licheniformis*. EC042_pAA023 is known as *virK*. VirK is a virulence protein in *Shigella flexneri* that is required for the expression or correct orientation of IcaA (VirG) on the outer membrane of the cell. This locus of *shf*, EC042_pAA022 and *virK* is found in the same order in *Shigella* and a large number of other *E. coli* strains. The locus is also present on the EAEC 042 chromosome. The chromosomal locus was not differentially expressed. Thus, it was concluded, that expression of *shf*, EC042_pAA022 and *virK* on the pAA2 plasmid was activated by AggR.

The $\log_2$-fold difference in expression of EC042_pAA061 was −3.25. EC042_pAA061 encodes a protein of 76 amino acids. A BLAST search was unable to identify homology to any known family of proteins. However, many hypothetical proteins were homologous to EC042_pAA061 including AWZ91_24990 from *Sh. sonnei*, CBL21_10600 from *Sh. flexneri* and BEH79_07420 from *K. pneumoniae*. EC042_pAA061 is 65% or more identical in amino acid sequence to these three proteins. The conservation of this protein in pathogenic strains suggests that EC042_pAA061 may have a role in pathogenesis. No clear function for the hypothetical proteins identified could be attained from homology searches. In conclusion, EC042_pAA061 is a gene with a cryptic function that is expressed in response to AggR in EAEC 042.

There were differences in expression of *aggR* and *aar* in the *aggR* mutant compared to the wild-type. The $\log_2$-fold difference of *aggR* was −5.47 in the mutant compared to the wild-type. The $\log_2$-fold difference in expression of the gene encoding AggR antivirulence factor, *aar*, was −3.55. Thus, it is concluded, consistent with previous reports that AggR activates expression of *aar*.

The production of the aggregative adherence fimbriae (AAF/II), encoded by *aafABCD*, are required for the characteristic aggregative adherence phenotype of EAEC. The $\log_2$-fold differences of *aafA*, *aafD*, *aafB* and *aafC* were −8.16, −7.49−4.54 and −4.54, respectively. The $\log_2$-fold difference of the pseudogene *afaB*, upstream of *aafB*, was −6.68. A gene encoding a hypothetical protein, EC042_pAA047, is positioned between *aafA* and *aafD*. The −9.19 $\log_2$-fold difference of EC042_pAA047 was the second greatest for differentially expressed genes. EC042_pAA047 is predicted to code for a protein of 54 amino acids with no significant homologous protein or nucleotide sequences identified by blastp or tblastn, respectively. No

signal sequence or function could be predicted for EC042_pAA047. The increase in aligned reads at the start of EC042_pAA047 suggests that its expression was not due transcriptional read-through from *aafD* (Fig. 3.9). In conclusion, in agreement with previous studies transcription of *afaB*, *aafABCD* and EC042_pAA047 were activated by AggR.

The greatest $\log_2$-fold difference, −9.36, was for the gene *aap*, that encodes the protein dispersin. In the wild-type strain, 2.42% of the normalised reads aligned to *aap*. A significant percentage of the mRNA transcriptome is used for the expression of *aap*. The genes that encode the system required for the secretion of Aap were also differentially expressed. The $\log_2$-fold change of *aatP*, *aatA*, *aatB*, *aatC* and *aatD* were −5.52, −6.05, −4.68, −4.55 and −3.47, respectively (Fig. 3.10A). The number of aligned reads across decreases across the operon in a 5′ to 3′ direction reflecting the orientation of the genes on the pAA2 plasmid (Fig. 3.10B). In conclusion, AggR activates expression of *aap* and the genes required for Aap secretion.

3.2.10. Validation of differentially expressed genes by qRT-PCR

In contrast to previous reports of the characterisation of the AggR regulon using microarrays, RNA-Seq indicated an activation of motility genes. The expression of 4 genes was also unexpectedly repressed in the presence of AggR. Triplicate overnight cultures of EAEC 042 pBAD24, EAEC 042Δ*aggR* pBAD24 and EAEC 042Δ*aggR* pAggR were used to inoculate DMEM. Cultures were grown to an OD_{600} of 0.4 at 37ºC with aeration. The culture medium was supplemented with L-arabinose a final concentration of 2%. Cultures were grown for a further hour at 37ºC with aeration. RNA was harvested and extracted as previously described. Contaminating DNA was removed by a DNase I treatment. The samples were quantified and 1.75 μg of total RNA was used to reverse transcribe cDNA. The template for each PCR was an equal amount of RNA prior to the reverse transcription of cDNA. The *polA* transcript was

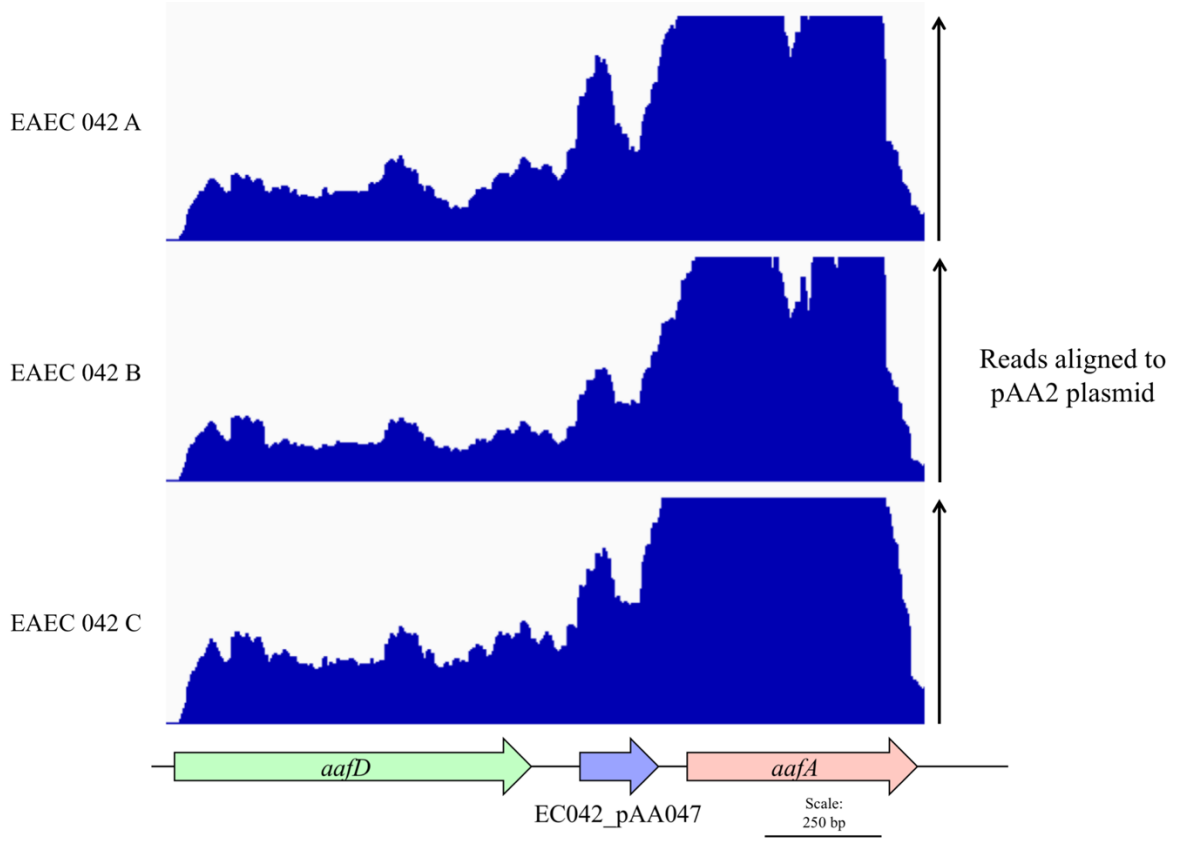


Figure 3.9: Aligned reads to pAA2 plasmid for the region between *aafD* and *aafA*. Each panel represents the number of aligned reads of each replicate of EAEC 042. Each panel is a histogram of the number of times a read aligns to a base in the sequence. The expression of *aafA* was substantially greater than that of *aafD* and EC042_pAA047. Thus, to visualise the expression of EC042_pAA047 the upper limit of the histogram was reduced. The position of the genes *aafD*, EC042_pAA047and *aafA* relative to the histogram is shown.

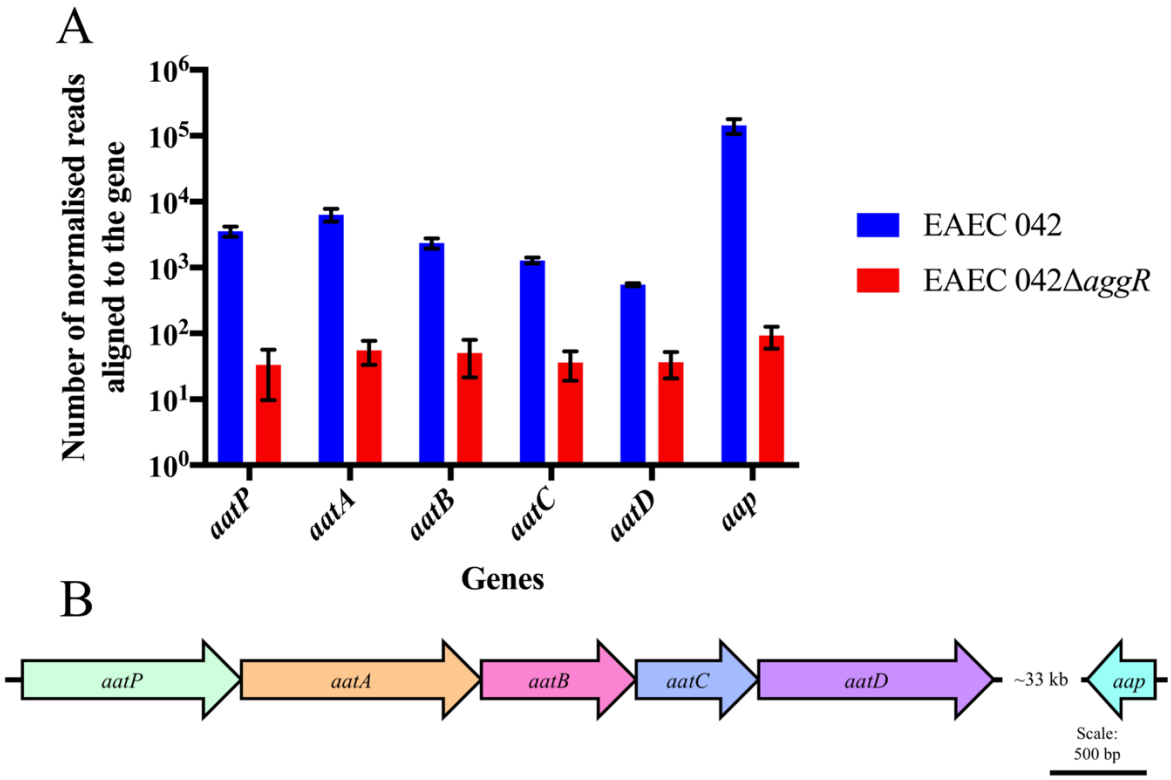


Figure 3.10: Normalised read counts and gene localisation for *aap* and the *aat* system. A, the normalised read counts for the genes. Error bars were drawn using the standard deviation between the three replicates. B, the *aat* genes are operonic with *aap* located 33 kb downstream of *aatD*.

used as the reference gene; the mRNA of the *aap* gene was used as the positive control, which is known to be activated by AggR. The primers used were *polA*-1, *polA*-2, *aap*-1, *aap*-2, *phoA*-1, *phoA*-2, *fliA*-1, *fliA*-2, *flgB*-1, *flgB*-2, *fliC*-1, *fli*-2, EC042_4803-1, EC042_4803-2, *bssS*-1 and *bssS*-2. For each gene, a total of 10 reactions were prepared, 1 for each sample and a no template control. Relative gene expression was calculated using the $2^{\Delta\Delta CT}$ method (Fig. 3.11, Table 3.3).

As expected, the relative gene expression of *aap* was reduced in the *aggR* mutant compared with the parental strain and complemented by pAggR above wild-type levels. There was a slight increase in the level of *phoA* transcripts in the *aggR* mutant, but was even higher in the strain transformed with pAggR. Although there was an increase in expression of *bssS* in the *aggR* mutant, complementation of the *aggR* mutant further increased the amount of *bssS* transcripts. Levels of *fliA*, *flgB*, *fliC*, and EC042_4803 expression were not significantly different in the wild-type compared to the mutant. With the exception of the *aap* mutant, none of the qRT-PCR results confirmed the RNA-Seq data. Therefore, it was concluded that there was no effect by AggR on the expression of *phoA*, *fliA*, *flgB*, *fliC*, *bssS*, or EC042_4803. Consequently, neither a repressor role for AggR nor a role in regulating motility was confirmed.

3.3. Discussion

Microarrays have previously been used to determine the extent of the AggR regulon (Morin et al., 2013; Dudley et al., 2006). Here, the AggR regulon was characterised using RNA-Seq. Using RNA-Seq the genes previously included in the AggR regulon were confirmed except for EC042_pAA056. In addition, 68 new candidates for AggR activation were identified. The

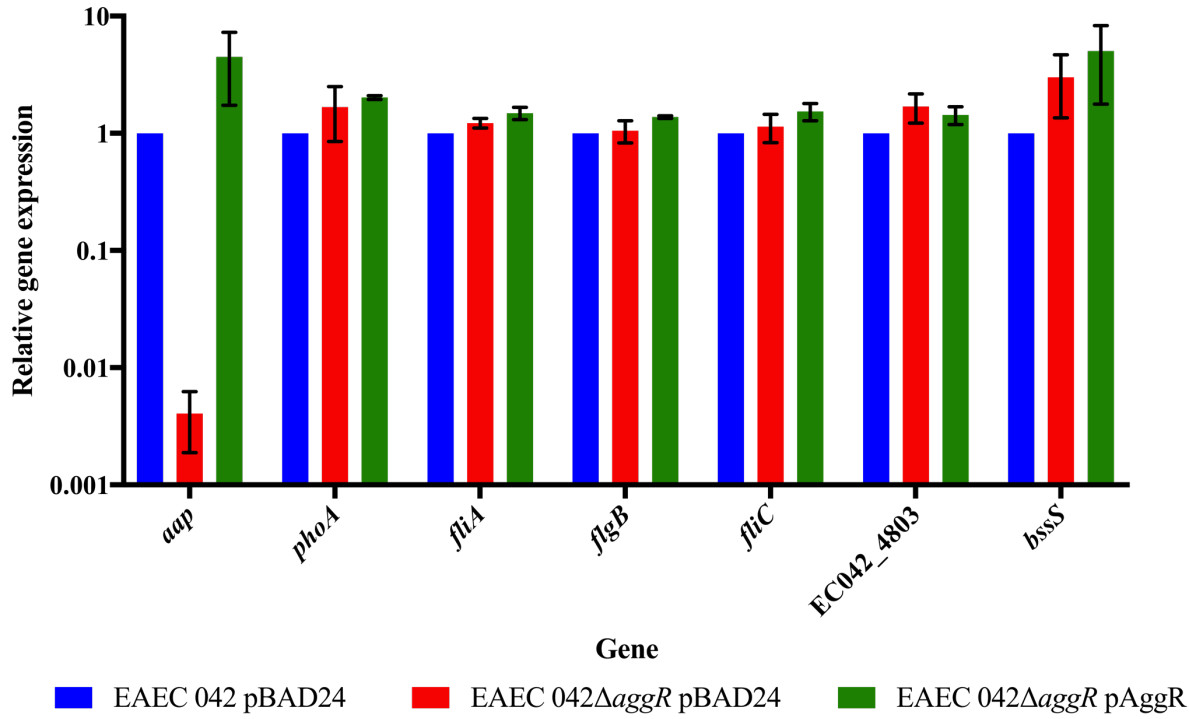


Figure 3.11: Validation of RNA-Seq results by qRT-PCR. The relative gene expression of each gene was tested against *polA* using the $2^{\Delta\Delta CT}$ method. Error bars represent standard deviation of relative gene expression between the replicates of each sample.

Table 3.3: Relative gene expression of qRT-PCR validated genes.

Relative gene expression	*aap*	*phoA*	*fliA*	*flgB*	*fliC*	EC042_4803	*bssS*
EAEC 042 pBAD24	1.0	1.0	1.0	1.0	1.0	1.0	1.0
EAEC 042Δ*aggR* pBAD24	0.0041	1.7	1.2	1.1	1.1	1.7	3.0
EAEC 042Δ*aggR* pAggR	4.5	2.0	1.5	1.4	1.5	1.4	5.0

genes encoding proteins associated with flagella and the genes that according to the RNA-Seq data appeared to be repressed could not be confirmed by qRT-PCR. Therefore, a total of 30 genes were potential new candidates for AggR activation. The function of AggR as a repressor has been previously suggested by Bodero et al. (2007), who concluded that AggR, CfaD and Rns are able to repress *nlpA*. These conclusions were based on data from *lacZ* promoter fusions and DNase I foot printing. The repression of *nlpA* was mediated by the binding of Rns downstream of the transcription start site of *nlpA*, which prevented the synthesis of *nlpA* mRNA. However, whether AggR represses *nlpA* was not tested in EAEC 042. Instead, AggR repression of *nlpA* was shown using the strain EAEC 17-2. As the full genome sequence is not available for EAEC 17-2, the similarity of the region of *nlpA* and EC042_4006 between EAEC 17-2 and EAEC 042 cannot be compared. Instead from the evidence presented here and that of Morin et al. (2013), it must be concluded that AggR does not repress the expression of *nlpA* in EAEC 042.

The initial site for EAEC 042 cells to form a biofilm in the host is a highly competitive microbial environment. Commensal bacteria provide a defensive barrier against pathogens. As T6SS are associated with bacterial killing (Russell et al., 2014), the expression of the *aai* T6SS during the formation of AA by EAEC 042 is potentially used to remove commensal or competing bacteria from the environment. Thus, allowing EAEC 042 to cause disease. The Aai T6SS likely does not have a direct role in the formation of the characteristic aggregative adherence. However, in the host environment it would be required for the initial stages of EAEC 042 pathogenesis.

There was a difference in expression of the duplicated *shf*, EC042_pAA022 and *virK* operons. On the chromosome EC042_pAA022 is termed EC042_4771. The nucleotide sequences of the

two operons are similar but not identical. During the analysis, reads that could be aligned to two separate regions in the genome were discarded. Therefore, the difference in regulation is likely real. The difference in regulation of these similar operons suggests that they are expressed in response to different conditions. The activation of this operon by AggR suggests that the proteins encoded are involved in EAEC 042 pathogenesis.

A total of 21 hypothetical genes were identified by RNA-Seq to be differentially expressed. The expression of 9 of these genes were not previously associated with AggR. The activation of these genes by AggR and homology to genes in other pathogens suggest that they affect pathogenesis. AggR is the activator of a cluster of hypothetical genes, EC042_3179A to EC042_3187. The function of EC042_3179A to EC042_3187 is not clear. Although their homology to other genes suggests a pathogenic role. This region presents a promising area for further research into EAEC 042 pathogenesis.

The AAF/II are required for the formation of aggregative adherence. The role of AggR in the activation of the genes encoding the AAF/II is well established. There has been no investigation into the function of EC042_pAA047. It is tempting to believe that due to the lack of conservation and the relatively small size of the protein that the expression observed is due to EC042_pAA047 position between *aafD* and *aafA*. The histogram of number of reads aligned to the *aafD* to *aafA* region indicates that EC042_pAA047 is not transcribed due to its position (Fig. 3.9). The second highest change in expression was observed in EC042_pAA047. EC042_pAA047 may have a role in the function of the AAF/II due to its position in the intergenic region of *aafD* and *aafA*. This gene presents itself as an interesting candidate for further research. However, the absence of homology suggests that this protein is only present in EAEC 042. Therefore, the function of EC042_pAA047 is not required for EAEC disease.

The role of AggR in the formation of aggregative adherence is well documented (Clements et al., 2012). The highest changes in transcription occur in the genes directly involved in aggregative adherence, such as *aafABCD*, *aatPABCD* and *aap*. The greatest difference in expression between the *aggR* mutant and the wild-type was for *aap*. The Aap protein has been investigated in EAEC 042 (Sheikh et al., 2002; Velarde et al., 2007).

AggR is a major transcription activator in EAEC. The mRNA transcriptome of an AggR producing strain is dominated by *aap* and *aaf*. AggR does not only affect the plasmid, on which it is encoded, by also the chromosome. AggR activates two major regions on the chromosome, the *aai* system and EC042_3179A to EC042_3187. The functions of these regions are yet to be understood. Thus, AggR is a global regulator. The study of the AggR regulon can lead to the identification of new mechanisms of EAEC 042 pathogenesis and present targets for vaccination and therapeutic treatments.

CHAPTER 4:

Characterisation of the CexE protein from the ETEC strain H10407

4.1. Introduction

The RNA-Seq experiment, in Chapter 3, concluded that the greatest difference in expression between the *aggR* mutant and the wild-type was for the *aap* gene. A substantial proportion of mRNA was transcribed in the wild-type strain for the expression of *aap* hinting at the importance of Aap for the pathogenesis of EAEC. Aap was characterised in EAEC 042; it is required for the correct orientation of the AAF/II in EAEC 042 (Sheikh et al., 2002). Aap is secreted out of the cell and coats the outer membrane (Nishi et al., 2003). There it modulates the cell surface charge allowing the positively charged AAF/II to extend away from the negatively charged outer membrane (Velarde et al., 2007). When prototypical ETEC H10407 was sequenced by Crossman et al., (2010) the *cexE* gene was identified as a homolog of *aap*. There is 18% identity between the protein sequences of CexE and Aap (Fig. 4.1).

Compared to Aap there is very little information on the CexE protein and other homologs. The *cexE* gene is regulated by CfaD, a transcription activator similar to AggR (Gallegos et al., 1993; Pilonieta et al., 2007). CexE elicits an immune response when murine hosts are injected with outer membrane vesicles of ETEC H10407, which is similar to that of the heat labile toxin and EtpA (Roy et al., 2010, 2011). However, these papers did not report a functional analysis of the CexE protein. Therefore, the aims of this study were to investigate the prevalence of Aap/CexE homologs and the possible function of CexE in ETEC H10407.

4.2. Results

4.2.1. Variation between Aap and CexE homologs

Other than reports from sequencing studies there is no information on the range and variation of the protein sequences of Aap and CexE homologs. Thus, a PSI-BLAST search was used to identify Aap and CexE homologs. The search identified 290 sequences that were homologous

```
CLUSTAL O(1.2.4) multiple sequence alignment


CexE      MKKY---ILGVILAMGSLSAIAGGGNSERPPSVAAGECVTFNSKLGEIGGYSWKYSNDAC
Aap       MKKIKFVIFSGILGI-SLNAFAGGS-GWNADNVDPSQCIKQSG-----VQYTYNSGVSVC
          ***    *:. **.: **.*:***. . .  .*  .:*:. ..       *::: . ..*

CexE      NETVAKGYAIGVAMHRTVNYEGGYSIQSSGIVKPGSDFIMKGGKTYKGHKKVSAGGDTPY
Aap       MQGLNEGKVRGVSVSGVFYYNDGTTSNFKGVVTPSTPVNTNQDINKTNKVGVQKYRALTE
           : : :* . **::  .. *:.* : : .*:*.*.: .  : . . ..:  *.

CexE      WYK
Aap       WVK
          * *
```

Figure 4.1: Alignment of CexE and Aap. The amino acid sequence of CexE and Aap from ETEC H10407 and EAEC 042, respectively, were aligned using clustal omega.

to the Aap protein sequence. The sequences were clustered to 90% identity, using cd-hit, forming 22 clusters. A phylogram, with bootstrapping analysis, was constructed using RAxML (Fig. 4.2). There were three major clades for Aap and CexE. These clades can be broadly characterised into Aap homologs in EAEC strains, CexE homologs in ETEC strains, and CexE homologs in *Yersinia enterocolitica* strains. A single sequence commonly represented a large number of sequences that were more than 90% identical in each clade. For the Aap clade, the WP_032283735.1 *E. coli* (EAEC 042) Aap sequence represented 65 sequences. All of the *Y. enterocolitica* sequences were represented by a single sequence. This clade was termed *Yersinia* CexE. This clade contains the CexE protein from the mouse pathogen *Citrobacter rodentium* ICC168. This strain is typically used as a mouse model for EPEC pathogenesis.

The majority of the Aap homologs were represented by a single sequence. The WP_032283735.1 *E. coli* (EAEC 042) tip contained many sequences of Aap homologs from EAEC O104:H4 that was responsible for the 2011 German outbreak of Shiga toxin producing EAEC. The AAF/III producing ancestor strain of the German outbreak strain, EAEC 55989 (Rohde et al., 2011), was included in this cluster. The tip representing the second highest number of sequences in the Aap clade contain the strain NCTC86. This is the first strain of *E. coli* that was isolated by Theodor Escherich in 1886. Interestingly the nucleotide sequence for *aap* in this strain is 100% identical to 10 other sequences in the NCBI nucleotide database (Appendix I). These were clinical isolates from different sites around the globe, every continent excluding Asia, from 2003 to 2016. One strain, HVH202 was isolated from a patient with bacteraemia caused by a urinary tract infection. Information on the strains is limited. However, this is an appropriate example of the spread and longevity of the *aap* gene. The Aap clade also contained a subclade including a *Salmonella enterica* strain isolated in Cuba. The sequence of Aap for *E. coli* DEC6A represented a *Shigella sonnei* sequence.

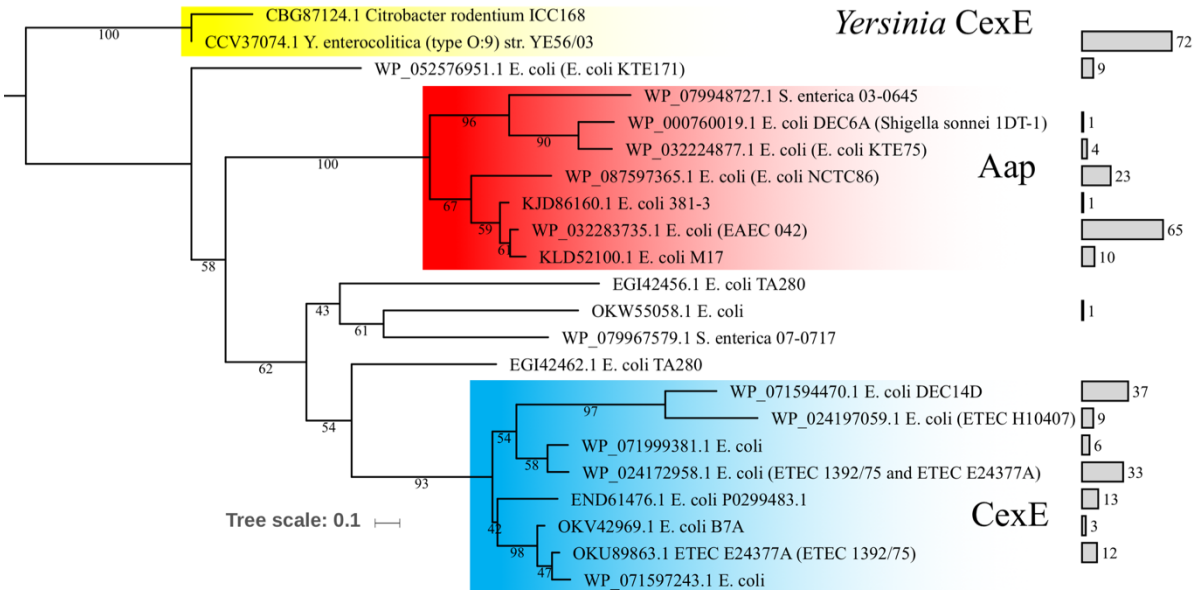


Figure 4.2: Phylogram of Aap/CexE homologs. Aap homologs were identified using 4 iterations of a PSI-BLAST search. The sequences were clustered based on their sequence identity using cd-hit. Clustering refers to the identification of protein sequences that are more than 90% identical to each other. These sequences are grouped together and the longest amino acid sequence is then used to represent the other sequences in the cluster. The longest sequence is used as it is considered to contain the most information about the amino acid sequences in that cluster. The end of each branch is the sequence used to build the tree, this is referred to as the tip. Thus, all sequences at the tips of the above phylogram are less then 90% identical. Sequences were aligned using Clustal Omega. From the sequence alignment, a tree was constructed with the branch lengths representing the number of changes in the amino acid sequence per site between each sequence using RAxML. The bootstrapping values associated with each branch were included. The tree was drawn using iTOL. The bars on the right represent the number of sequences, other than the sequence in the tree, that are represented by each cluster. These include identical sequences and sequences that were more than 90% identical. The names of strains referenced in the text that were present in a cluster were included in brackets.

The third major clade was CexE from ETEC strains. The prototypical ETEC, H10407, and a number of other ETEC strains including, 1392/75, E24377A and B7A, also fell into this clade. There were a number of duplications of the *cexE* gene in some strains identified in the CexE clade. There were two sequences that were homologues to CexE for strains ETEC 1392/75 and ETEC E24377A. These duplications were present on separate plasmids in the strains. For ETEC 1392/75, the *cexE* duplications were on p748 and p1018 and for ETEC E24377A they were present on pETEC_73 and pETEC_80. The identities of the protein sequences for the duplicates was 59% and 61% for ETEC 1392/75 and ETEC E24377A, respectively. The two homologous CexE sequences from ETEC 1392/75 were in separate clusters, the same was true of ETEC E24377A. In fact, the separate clusters of the CexE homologs in ETEC 1392/75 were the same as the two CexE homologs from ETEC E24377A. This perhaps suggests not a double duplication event by a single duplication before the strains diverged.

There were a small number of other sequences that did not fall within the three major clades. Little sequence identity is shared between these sequences. Another duplication event for the *cexE* gene was found in the strain TA280. Similarly, with ETEC 1392/75 and ETEC E24377A there is a difference between the sequence identity of the two CexE proteins of 28% in TA280.

The bootstrapping values for some of the clades in the analysis are quite low, which reflects the low identity between the protein sequences. However, the bootstrapping values associated with each major clade are quite high: 100 for *Yersinia* CexE and Aap and 93 for CexE. Therefore, it can be confidently stated that the definition of the major clades is correct. The sequences of CexE and Aap have diverged. Sequences in separate clades correlated closely with the pathovar of the host in which they are found. Therefore, it is suggested that the function of these homologs varies between pathovars.

4.2.2. Construction of a *cexE* deletion mutant in the ETEC strain H10407

It was hypothesised above that the function of the CexE protein is different to that of Aap. Thus, a *cexE* deletion was constructed in ETEC H10407 using the Datsenko and Wanner (2000) method. ETEC H10407 was transformed by electroporation with pKD46, a plasmid harbouring the λ-Red recombinase genes. The primers GD*cexE*-FW and GD*cexE*-RV were designed to amplify a kanamycin cassette with 50 bp of flanking homology directly upstream and downstream of the *cexE* gene. Instead of pKD4, a kanamycin resistance cassette was amplified from pDOC-K using the primers above. The PCR products were gel purified and eluted in nuclease-free water.

An overnight culture of ETEC H10407 pKD46 was used to inoculate LB supplemented with 0.2% L-arabinose to an OD_{600} of 0.05. The culture was grown to an OD_{600} of 0.6. Cells were harvested by centrifugation. The linear kanamycin cassette, with flanking homology to the *cexE* gene, was transformed into ETEC H10407 pKD46 using electroporation. Cells were recovered in SOC at 37ºC with aeration. Transformants were selected on LBA supplemented with kanamycin. Kanamycin resistant colonies were tested for the insertion of the kanamycin resistance cassette into the *cexE* gene using the primers *cexE*-F1 and *cexE*-R1, which annealed outside of the *cexE* gene. Colonies that contained a deleted *cexE* gene were isolated by streaking onto LBA supplemented with kanamycin.

The isolation of a ETEC H10407 strain with a disruption of the *cexE* gene was confirmed by colony PCR. The primers *cexE*-FW1 and *cexE*-RV1 were used to amplify the region containing the *cexE* gene by PCR. The template DNA for the PCR were single colonies of ETEC H10407 and ETEC H10407 with the deleted *cexE* gene. A negative control of nuclease-free water was

included. The approximate size of the PCR products was determined by agarose gel electrophoresis (Fig. 4.3). The size of the region containing the *cexE* gene in ETEC H10407 was around 950 bp, as expected for the wild-type, and 2 kb for the *cexE* deletion mutant. The difference in size between the two templates was accounted for by the insertion of the kanamycin resistance cassette into the *cexE* gene. There was no PCR product for the nuclease-free water. Therefore, it was concluded that the *cexE* gene had been deleted. Thus, the strain containing the deleted *cexE* gene was termed ETEC H10407Δ*cexE*.

4.2.3. Construction of a plasmid for the expression of *cexE*

In order to be able to detect CexE in protein samples specific antibodies were required. This required the production and purification of recombinant CexE protein. The *cexE* gene from ETEC H10407 was cloned into the expression vector pET26b using the primers C*cexE*-FW and C*cexE*-RV. The primers were designed so that when the *cexE* gene fragment was ligated into pET26b(+) the product would be CexE with a C-terminal 6His tag. The PCR product and pET26b(+) were digested with NdeI and XhoI. This retained the native signal sequence and removed the *pelB* signal sequence present in pET26b(+). The digested *cexE* PCR product and pET26b(+) were incubated with T4 ligase and then transformed into *E. coli* DH5α. The recovered cells were plated onto LBA supplemented with kanamycin and incubated overnight at 37ºC. Potential candidates for the insertion of *cexE* into pET26b(+) were screened using colony PCR with the primers T7F and T7R. A positive control of pET26b(+) and a negative control of nuclease-free water were included. The approximate size of the PCR products was determined (Fig. 4.4). The PCR products for two of the candidate colonies were larger than the PCR product of the pET26b(+) control. The difference in the size of the PCR products was equal to the expected difference for a successful insertion of the *cexE* gene. Therefore, these candidates were suspected of containing the *cexE* gene. Overnight cultures of each of the two

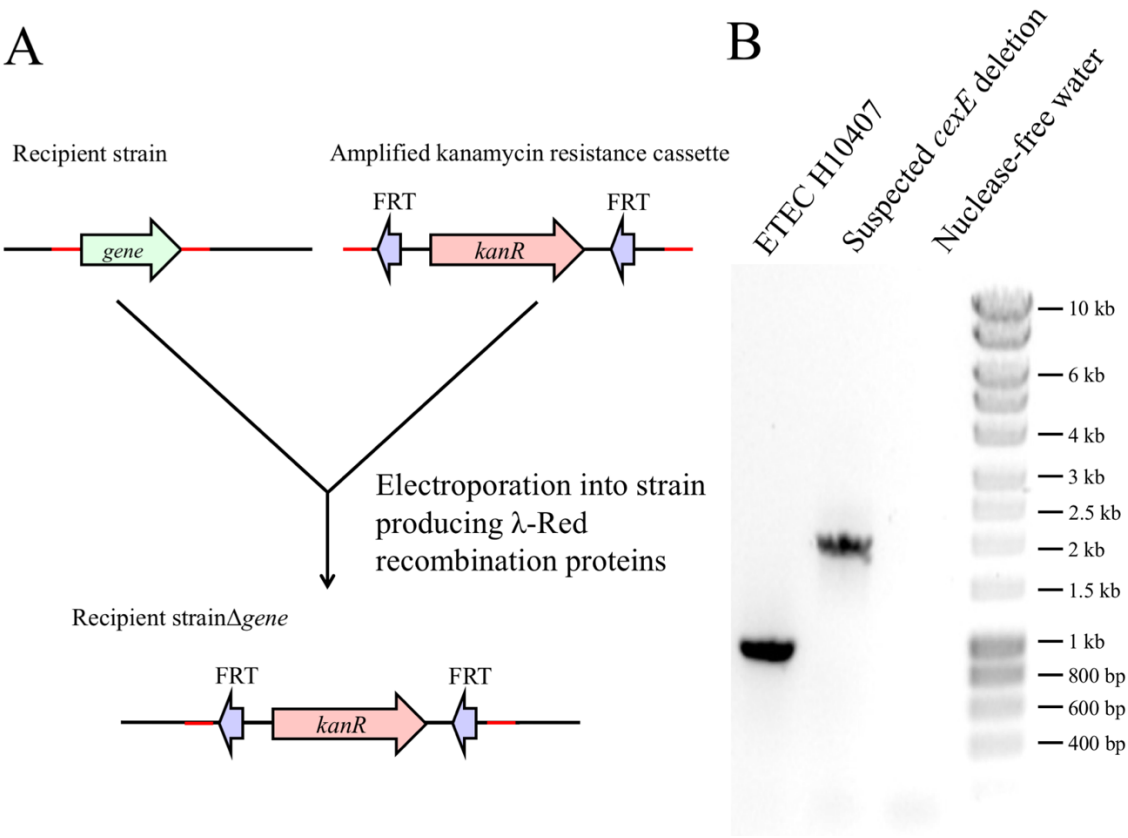


Figure 4.3: Construction of *cexE* gene deletion using λ-Red recombination. A, schematic of the λ-Red recombination system. The kanamycin resistance cassette is amplified using primers with homology regions of 50 bp (red) to the 5′ and 3′ ends of the target gene. B, PCR products of wild-type and suspected *cexE* deletion mutant. PCR products were separated on a 1% agarose gel by electrophoresis. The size of each band on the ladder is indicted on the right of the gel.

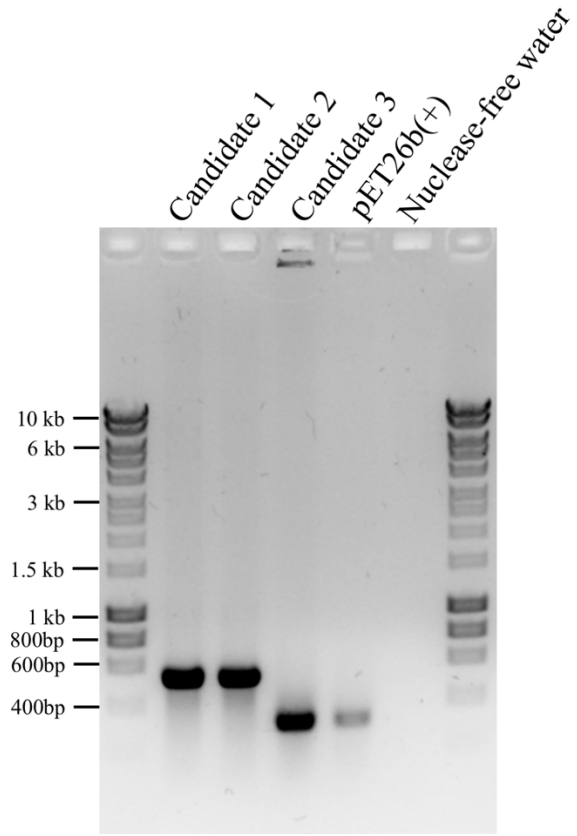


Figure 4.4: PCR products of candidates for *cexE* insertion into pET26b(+). An overnight culture of *E. coli* DH5α pET26b(+) was used to isolate the pET26b(+) plasmid. The digested plasmid and *cexE* PCR product were purified by gel extraction. The cut vector and insert were ligated using T4 ligase and transformed *E. coli* DH5α. After recovery, cells were harvested by centrifugation and plated onto LB agar supplemented with kanamycin. The resulting colonies were used as the template for the PCR with primers external to the multiple cloning site of pET26b(+). PCR products were separated on a 1% agarose gel. HyperLadder 1kb (Bioline) was used for PCR product size approximation. An empty vector produces a PCR product of 309 bp. A vector with the *cexE* gene inserted produces a PCR product of 543 bp. Two of the candidates produced a product under 600 bp and the empty vector produced a product under 400 bp.

candidates suspected of containing the *cexE* gene were used to isolate plasmid DNA. The plasmids were sequenced, only one of which was found to contain the correct nucleotide sequence for the correct insertion of the *cexE* gene into pET26b(+). Therefore, it was concluded that *cexE* was successfully cloned into the pET26b(+) expression vector. Thus, this plasmid was termed pET26b-*cexE*. Protein produced using this plasmid with a C-terminal 6 histidine tag was termed CexE-6His.

4.2.4. Optimisation of induction concentration for CexE-6His production

Protein production varies with the concentration of the inducer. Both pET26b(+) and pET26b-*cexE* were isolated and transformed into *E. coli* BL21(DE3). Cells were recovered and plated. PCR, with the primers T7F and T7R, was used to check for successful transformation. Overnight cultures of *E. coli* BL21(DE3) containing either pET26b-*cexE* or the empty vector were used to inoculate cultures of LB that were grown at 37ºC with aeration to an OD_{600} of 0.3. Protein production was induced in separate cultures of *E. coli* BL21(DE3) containing pET26b-*cexE* by the addition of IPTG to 250 μM, 50 μM and 10 μM. A non-induced control was included. For each culture of *E. coli* BL21(DE3) pET26b-*cexE* a control culture of pET26b(+) was induced at the same concentration of IPTG. Cultures were grown at 37ºC with aeration. After 4 hours, proteins in whole cell lysate samples were separated by electrophoresis on Tris-tricine gels. Gels were stained with Coomassie Brilliant Blue or used for western blotting to determine the production of CexE-6His (Fig. 4.5). In the Coomassie Brilliant Blue stained gel a band was present at the approximate size for CexE-6His in each of the whole cell lysates of cultures containing pET26b-*cexE*, but absent from the negative control cultures. The westernblot confirmed that these bands CexE were tagged with 6His. Note that there was production of CexE-6His without any IPTG induction. This was comparable to the amount produced with 250 μM IPTG induction. The greatest production of CexE-6His was with 10

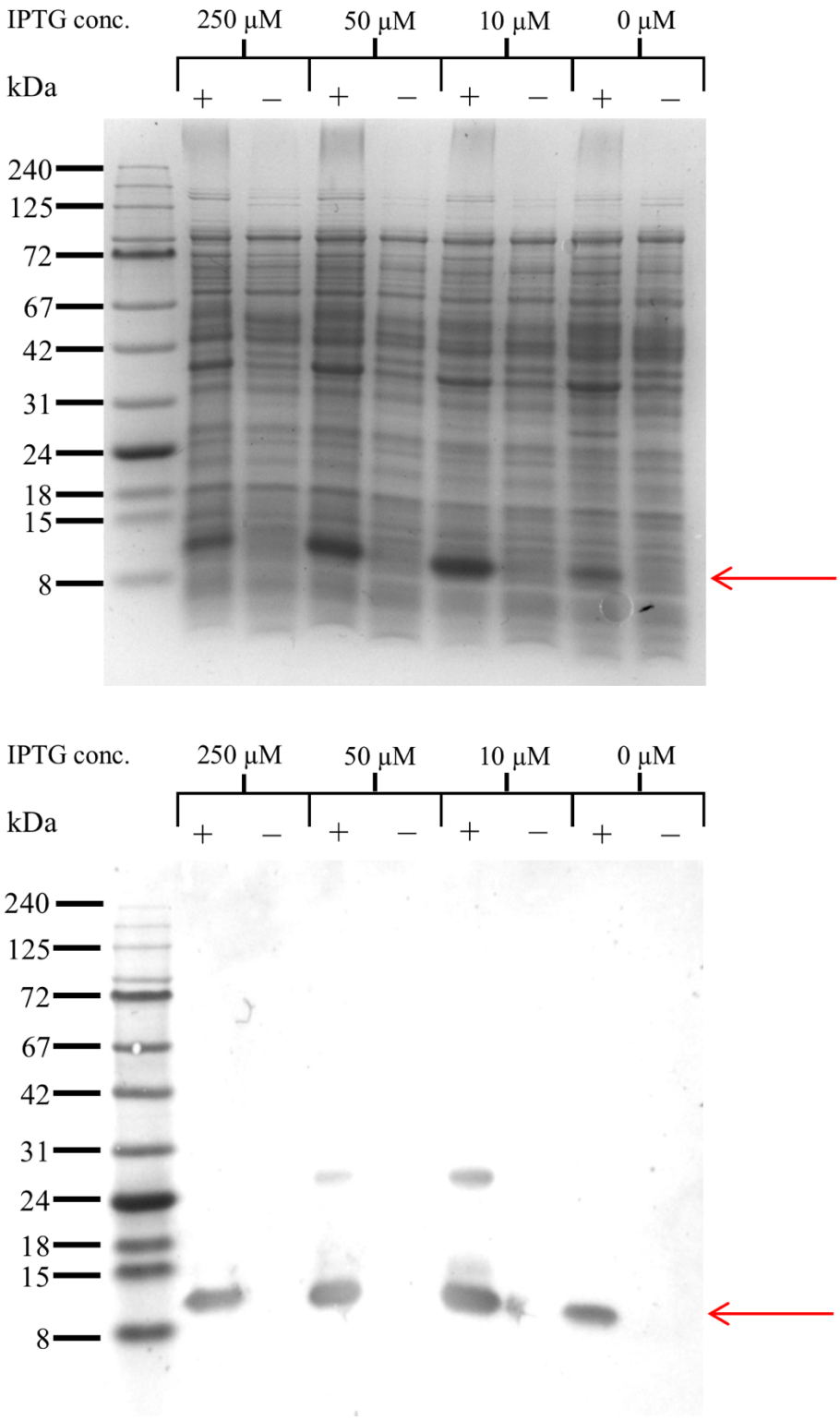


Figure 4.5: The effect of IPTG concentration on the production of CexE-6His. BL21 pET26b-*cexE* and BL21 pET26b(+) are represented by + and −, respectively. A, Coomassie stain of whole cell lysate sample separated on a Tris-tricine gel electrophoresis. B, western blot using anti-His antibodies. Arrows indicate the position of CexE-6His protein produced.

μM IPTG induction. However, it was not substantially more than the production of CexE-6His in the presence of 50 μM IPTG. At both of these concentrations there was a larger molecular weight band present on the western blot than for 250 μM IPTG and the non-induced whole cell lysate. The detection of this band by western blotting with anti-His antibodies, but not in the negative control proteins samples, confirmed that it is the CexE-6His protein. It was not clear what was responsible for the higher molecular weight band but it was proportional to the amount of CexE-6His produced. Thus, the larger band was believed to be a dimer of CexE-6His due to the high amount of production. The optimal concentration for CexE-6His production was concluded to be 10μM IPTG. However, it was decided that as the differences between the production of CexE-6His with 50 μM and 10 μM induction were not substantially different, both of these concentration of IPTG induction would be used.

4.2.5. Optimal temperature for the production of CexE

Lower temperatures can reduce the amount of protein in inclusion bodies and thus increase the yield of soluble protein available. Overnight cultures of *E. coli* BL21(DE3) containing either pET26b-*cexE* or pET26b(+) were used to inoculate LB supplemented with kanamycin. Cultures were grown at 37ºC with aeration to an OD_{600} of 0.3 and were cooled to 16ºC or 10ºC. After 30 minutes, cultures transformed with pET26b-*cexE* were induced with 50 μM or 10 μM IPTG. Negative control pET26b(+) cultures were not induced. Cultures were grown at 16ºC or 10ºC with aeration. Whole cell lysate samples were taken after 4 hours and overnight post-induction. The production of CexE-6His was determined as described above (Fig. 4.6). In the samples from three of the cultures, bands in the Coomassie Brilliant Blue stain were more intense than at the same position in the negative control. These bands were confirmed to be CexE-6His by detection by western blotting with the anti-His antibody. After 4 hours of induction at 16ºC the CexE-6His protein was produced only with 50 μM IPTG induction. After

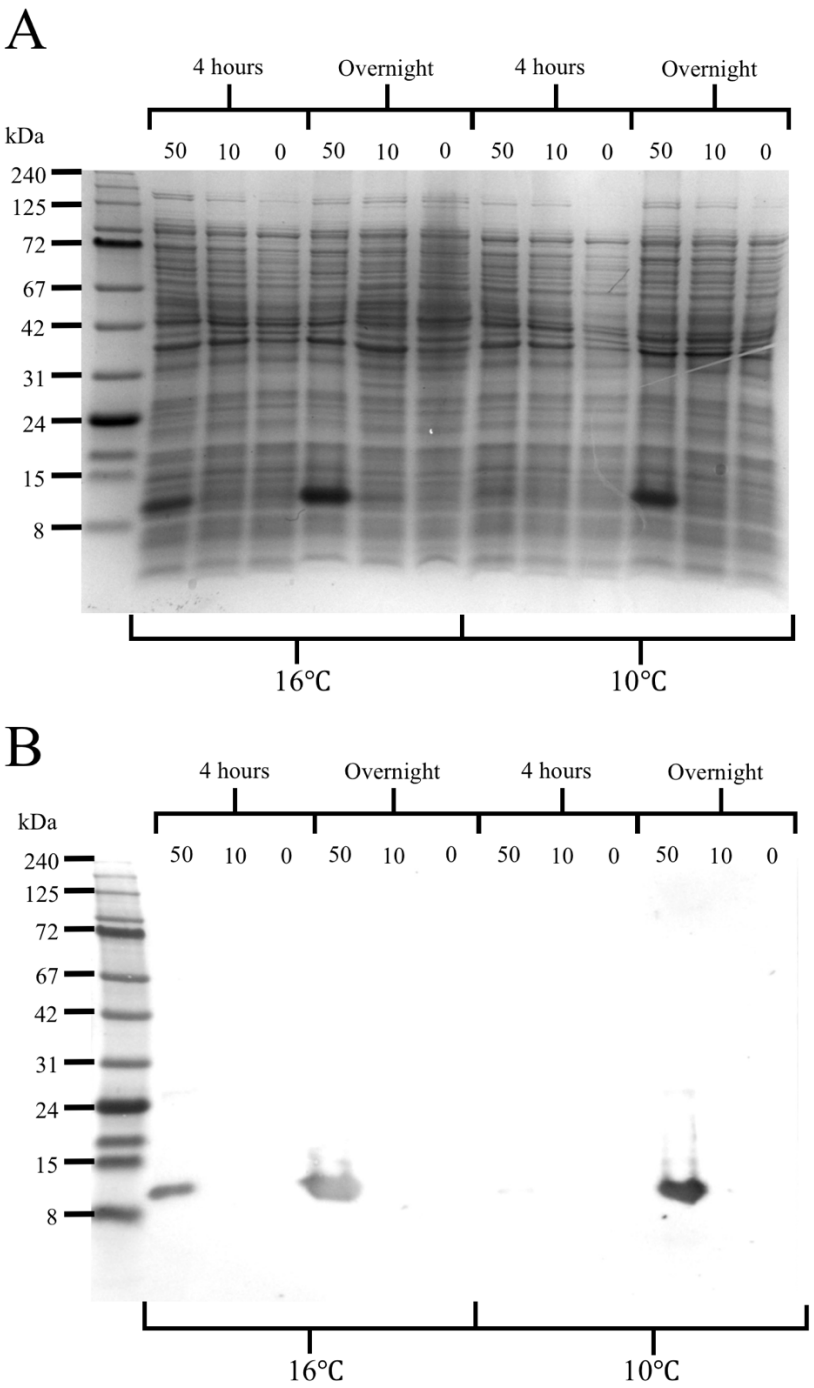


Figure 4.6: Production of CexE-6His at 16ºC and 10ºC. BL21 pET26b-*cexE* was grown for 4 hours and overnight in LB supplemented with 50 μM or 10 μM IPTG (labelled 50 and 10, respectively) at 16ºC or 10ºC. BL21 pET26b(+) was grown for 4 hours and overnight in LB without IPTG (labelled 0) at 16ºC or 10ºC. A, Coomassie stain of whole cell lysates separated on a Tris-tricine gel. B, the primary antibody for the western blot was against 6His tag a duplicate Tris-tricine gel.

overnight growth, CexE-6His was produced at both 16ºC and 10ºC with 50 μM IPTG induction. It was not produced with 10 μM IPTG induction at either 16ºC or 10ºC. The production of CexE-6His after overnight growth with 50 μM IPTG induction was relatively similar at 16ºC and 10ºC. However, the yield of biomass at 10ºC was less than at 16ºC. Therefore, it was concluded that the optimal conditions for the production of CexE-6His was at 16ºC with 50 μM IPTG induction.

4.2.6. Purification of the CexE-6His protein

The positively charged C-terminal histidine residues can be used for the purification of protein using Ni-Sepharose columns. An overnight culture of *E. coli* BL21(DE3) pET26b-*cexE* was used to inoculate six 1 L LB cultures. The cultures were grown at 37ºC with aeration to an OD_{600} of 0.3 and then cooled to 16ºC for 30 minutes. Protein production was induced with 50 μM IPTG. After overnight growth at 16ºC with aeration, cells were harvested by centrifugation. The cell pellet was resuspended in binding buffer. DNase I was added to ease lysis. Cells were lysed using a EmulsiFlex-C3. Undisrupted cells were removed by centrifugation. The remaining supernatant was subjected to further centrifugation to pellet the membrane components. The soluble protein fraction was removed, filtered and loaded into a 5 ml HisTrap™ HP column. The filtrate was left to flow through the column overnight at 4ºC. The column was washed with binding buffer, 50 mM and 100 mM wash buffer and the protein eluted using 500 mM imidazole. Fractions of 5 ml were collected throughout. From each fraction, a 10 μl sample was taken and mixed with an equal volume of 2x Laemmli sample buffer. Each fraction was analysed by electrophoresis using a Tris-tricine gel for the purification of the CexE-6His protein (Fig. 4.7). There was protein present in the binding, wash and elution fractions of expected molecular weight of the CexE-6His protein. In the binding buffer fractions the CexE-6His was heavily contaminated with larger sized proteins. In the 50

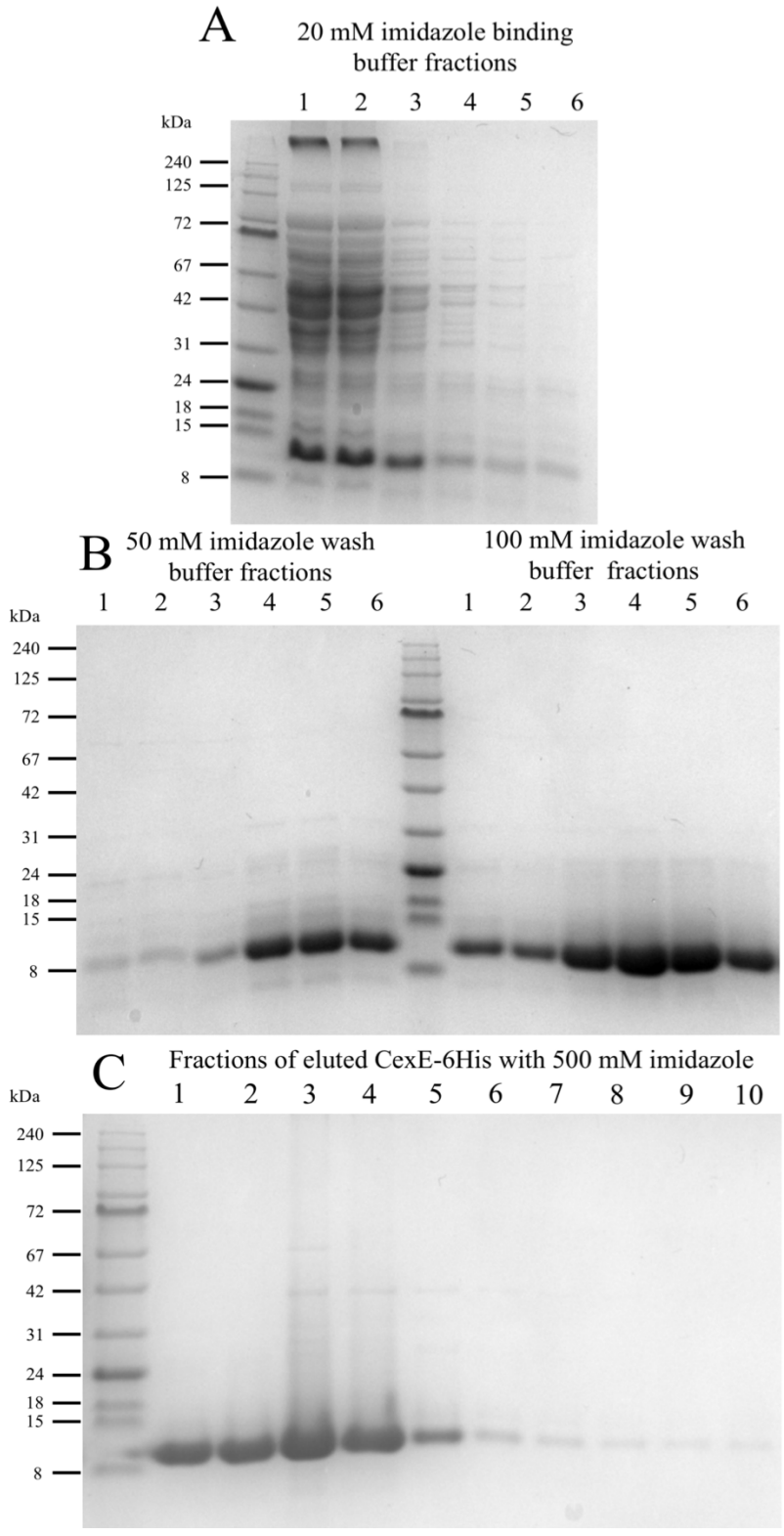


Figure 4.7: Purification of CexE-6His. A, column washed with binding buffer (20 mM imidazole). B, the column was first washed with 50 mM imidazole wash buffer and then with 100 mM imidazole wash buffer. C, the CexE-6His protein was eluted off of the column with 500 mM imidazole.

mM imidazole wash this was decreased, but there was still contaminating proteins present. The supernatant samples were prepared by centrifugation. The fractions 3, 4, 5, and 6 of the 100 mM imidazole wash contained substantial amounts of CexE-6His protein with little to no contamination. The larger molecular band present in these fractions was the same size as those observed in Fig. 4.5. Therefore, these bands were thought to be CexE-6His. Fractions 1 and 2 of the elution buffer did not contain any contamination. There were larger molecular weight bands present in elution fractions 3, 4, 5, and 6. However, there was large amounts of CexE-6His. Thus, the elution fractions and fractions 3, 4, 5, and 6 of the 100 mM imidazole wash were concentrated by membrane filtration. Imidazole was removed from the protein sample by dialysis. The protein sample was further purified by size exclusion. Fractions containing pure CexE-6His were concentrated again by membrane filtration. The concentration was determined using the absorbance at A280. The concentration of CexE-6His protein was 20.3 mM. The protein was diluted to 2 mM prior to storage. Thus, CexE-6His was successfully been purified.

4.2.7. Primary antibody production for CexE

Primary antibodies enable a protein to be detected specifically. Aliquots of purified CexE-6His protein at 200 μg/100 μl were sent to Eurogentec for 28-day speedy rabbit antibody production. Sera 872 and 873 were produced for the CexE protein. The specificity and ability of these sera to detect CexE was assessed. Overnight cultures of ETEC H10407 and ETEC H10407Δ*cexE* were used to inoculate LB. Cultures were grown at 37ºC for 6 hours. Whole cell lysate samples were prepared by centrifugation of the cells and resuspension in Laemmli buffer. Culture spent medium was aspirated and filtered. The proteins in the culture supernatant were precipitated by the addition of TCA. The precipitated proteins were harvested by centrifugation. The precipitated protein pellets were washed with methanol, pelleted and resuspended in Laemmli sample buffer. A dilution series of the purified CexE-6His protein was prepared. The whole

cell lysate, culture supernatant samples and CexE-6His dilution series were subjected to electrophoresis on a Tris-tricine gel. Two separate gels were transferred to nitrocellulose for western blotting with the serum 872 and 873 as the primary antibody for each membrane. The specificity and detection of the CexE protein by the serum was assessed (Fig. 4.8). The two sera were able to detect the purified CexE-6His protein. There was a band at around the size expected for CexE in the whole cell lysate sample for ETEC H10407 for both sera. This band was present in the *cexE* mutant as well. Therefore, it was a non-specific interaction and not the CexE protein. There was not a band for CexE in the supernatant for ETEC H10407. There was sustainably fewer non-specific bands for serum 872 compared to serum 873. There was greater signal of purified CexE-6His for serum 872 compared to 873. Thus, it was concluded that serum 872 would be used as the primary antibody for the detection of CexE.

4.2.8. Replacement of the *aap* gene in EAEC 042 with *cexE* from ETEC H10407 under transcriptional control of the *aap* promoter

A phenotype for the *cexE* mutant in ETEC H10407 could not be identified under typical laboratory conditions. Therefore, it was decided to investigate whether *cexE* could complement an *aap* mutant. A linear fragment containing the *cexE* gene and a kanamycin resistance cassette was constructed using overlapping PCR. The fragment containing the *cexE* gene and kanamycin resistance cassette was flanked by 50 bp homologous to 5′ and 3′ ends of the *aap* gene (Fig. 4.9). The *cexE* gene was cloned with ETEC H10407 as the template using the primers Fwd*aap*-*aap*/*cexE* and Rev*aap*-*kan*/*cexE*. The kanamycin cassette from pKD4 was used as the template for a PCR with the primers Fwd*aap*-*cexE*/*kan* and Rev*aap*-*aap*/*kan*. The two PCR products were purified. The PCR products were quantified. Separate PCR containing 50 ng, 100 ng and 250 ng of both the *cexE* gene fragment and the kanamycin resistance cassette fragment each were prepared. After the initial 10 cycles the primers Fwd*aap*-PCREx and

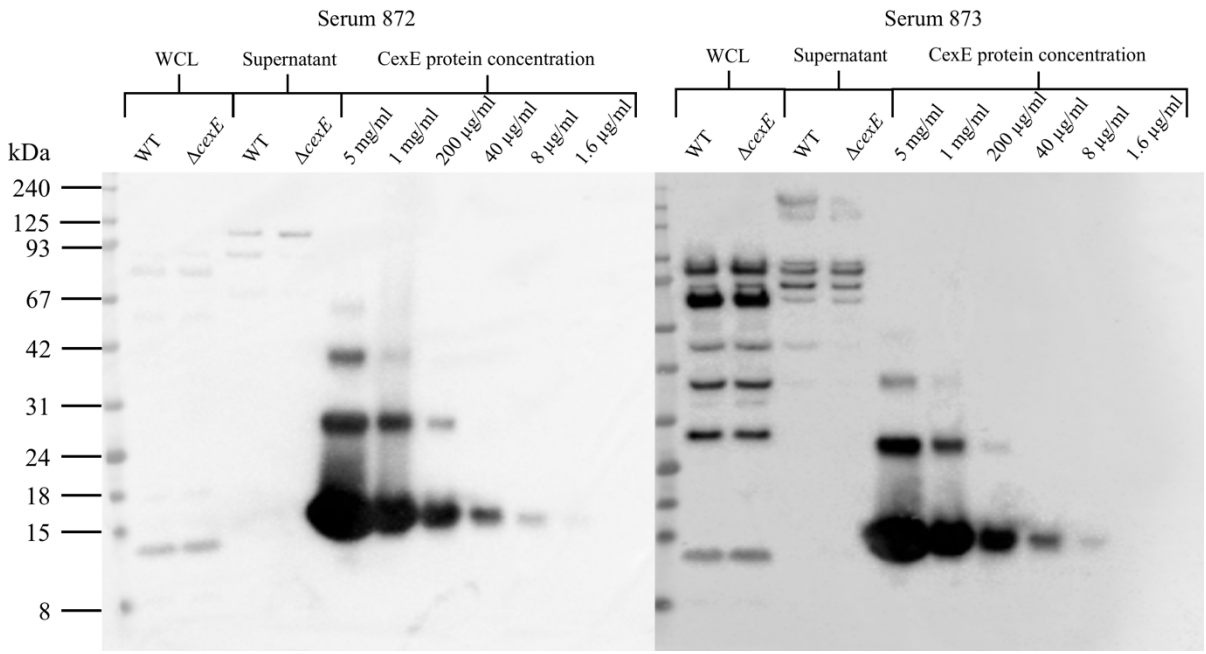


Figure 4.8: Testing of sera 872 and 873 for the detection of the CexE protein. The membranes were incubated in Blotto (50 g skimmed milk powder, 1 g sodium azide in 1 L TBS) prior to primary antibody application. The serum 872 or 873 was used as the primary antibody at a concentration of 1 in 1,000 in Blotto. The membranes were incubated fully immersed overnight at 4ºC with agitation. The membranes were washed with TSBT. Secondary antibody consisted of goat anti-rabbit IgG linked with HRP at a concentration of 1 in 10,000 in Blotto. Membranes were incubated with anti-rabbit secondary antibody conjugated with HRP for 2 hours with agitation and washed with TSBT. ECL was used for the detection of secondary antibody attachment to the membrane.

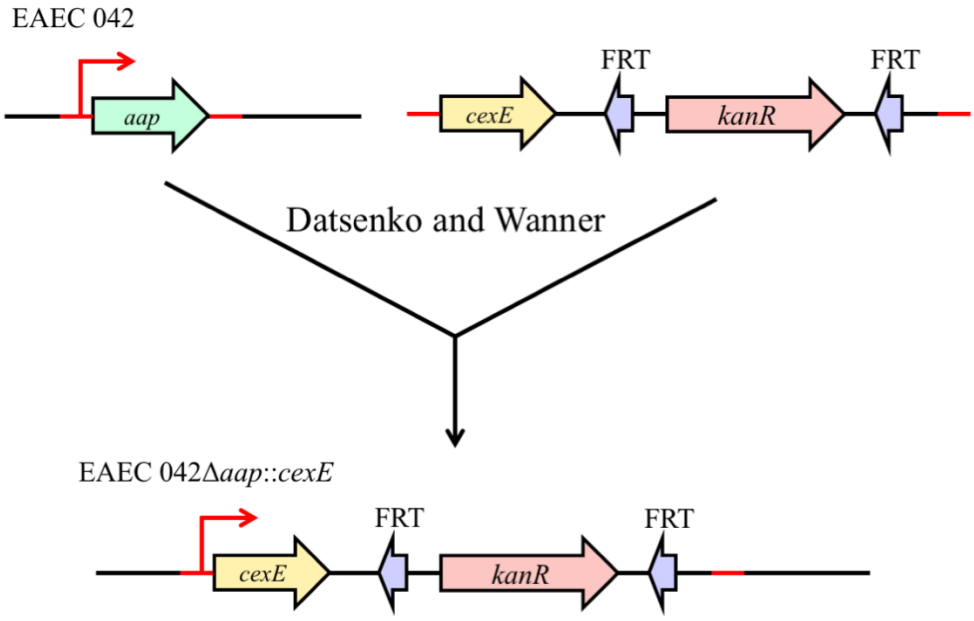


Figure 4.9: Construction of *cexE* gene under the *aap* promoter in EAEC 042. The linear fragment containing *cexE* and the kanamycin cassette is flanked by 50 bp regions of homology to the *aap* gene (red). The fragment is designed so that the start codon of *aap* was replaced with the start codon of *cexE*. The *cexE* was expressed at the same conditions in EAEC 042Δ*aap*::*cexE* as *aap* in EAEC 042. The linear fragment was incorporated into the EAEC 042 genome using the Datsenko and Wanner (2000) method.

Rev*aap*-PCREx were added. The product was purified and termed *cexE-kan*. A second PCR using Fwd*aap*-PCREx and Rev*aap*-PCREx was used to increase the yield. EAEC 042 was transformed by electroporation with pKD46. An overnight culture of EAEC 042 pKD46 was used to inoculate LB supplemented with 0.2% L-arabinose. The culture was grown to an OD_{600} of 0.6 and the cells harvested. The cells were made electrocompetent and transformed with the *cexE-kan* linear fragment. The cells were recovered and plated onto LBA containing kanamycin. Colonies that grew were checked for the insertion of the *cexE-kan* into the *aap* gene by colony PCR. PCR products with sizes corresponding to the expected size for the insertion of *cexE-kan* were termed EAEC 042Δ*aap*::*cexE*.

4.2.9. Effect of CexE production on EAEC 042 aggregation

Aap is known to disperse EAEC 042, thus preventing aggregation (Sheikh et al., 2002). It was hypothesised that if the function of CexE is similar to Aap, the production of CexE would be able to complement an *aap* mutant. Thus, triplicate overnight cultures of EAEC 042Δ*aap*, EAEC 042Δ*aap*::*cexE* and the parental strain were grown at 37ºC with aeration in LB. After overnight growth cultures were normalised to an OD_{600} of 3 in 5 ml LB. Cultures were kept stationary. OD_{600} readings were taken from just under the meniscus of the culture. The readings were taken every 5 minutes for the first 30 minutes and then every 15 minutes for the following 45 minutes. As a measure of aggregation, the average percentage of the starting OD_{600} for each culture was calculated. Each average was plotted against time (Fig. 4.10A). The rate of aggregation of the *aap* mutant was initially comparable to the wild-type. During the course of the experiment the rate of aggregation increased compared to the parental strain resulting in a greater amount of sedimented cells. Thus, the aggregation of EAEC 042Δ*aap* was greater than that of the wild-type as previously described by Sheikh et al. (2002). However, when *aap* was replaced by *cexE* in EAEC 042 the rate of aggregation greatly increased. The rate of

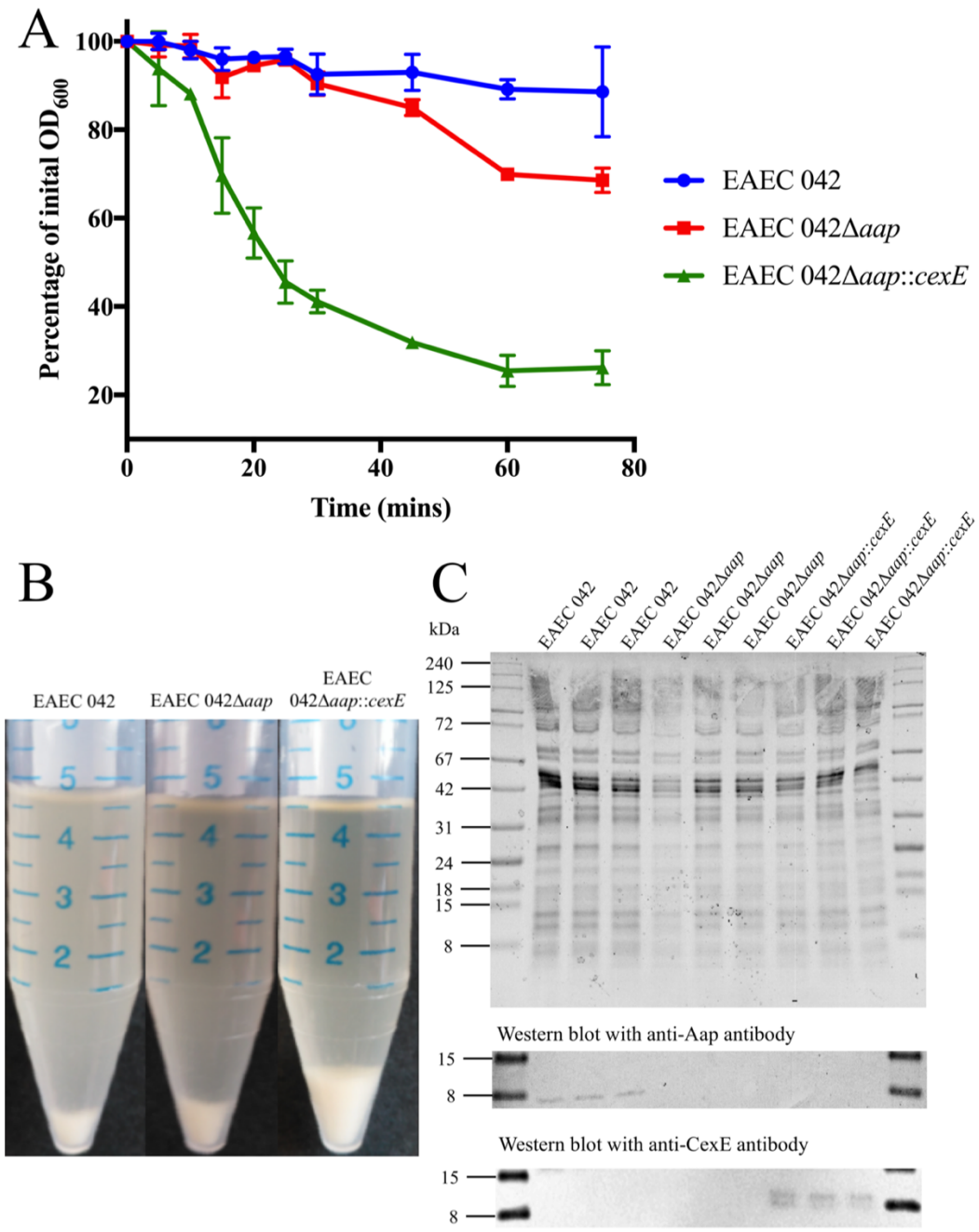


Figure 4.10: Aggregation of EAEC 042, EAEC 042Δ*aap* and EAEC 042Δ*aap*::*cexE*. A, aggregation assay of EAEC 042, *aap* and *cexE* insertional mutants. The amount of aggregation of each culture was measured by calculating the percentage of the initial OD_{600} using a 100 µl sample taken from the meniscus at each time point of three biological replicates. B, the aggregated cell sediments shown were representative of the three replicates of each strain after 75 minutes. C, Coomassie stained gel and western blot for Aap and CexE of whole cell lysates of the strains from A and B.

aggregation of EAEC 042Δ*aap*::*cexE* was higher in the first 30 minutes compared to the last 45 minutes, the majority occurring within the first 25 minutes. This was not comparable to the *aap* mutant or the parent strain. The final amount of sedimented cells was slightly greater for the *aap* mutant compared to the wild-type but this was substantially less than the *cexE* insertional mutant (Fig. 4.10B). The production of Aap and CexE was assessed from whole cell lysates of each replicate by Coomassie Brilliant Blue stain and western blotting (Fig. 4.10C). The production of Aap and CexE in the expected strains was confirmed by western blotting with anti-Aap and anti-CexE antibodies, respectively. There was a double band produced for CexE in EAEC 042Δ*aap*::*cexE*. The larger band was not comparable in size to that produced during CexE-6His production (Fig. 4.5). The cause of the double band for CexE was not clear. From these results, it was concluded that the production of CexE did not complement an *aap* mutant. However, it did increase the rate and amount of aggregation. Therefore, the function of CexE is not the same as that of Aap.

4.2.10. Comparison of ETEC H10407 and *cexE* mutant biofilm formation

Due to the aggregation phenotype of CexE produced in EAEC 042 it was suggested that CexE could play a role in biofilm formation. Strains ETEC H10407 and ETEC H10407Δ*cexE* were transformed by electroporation with the pJB39 plasmid that constitutively produced GFP. Prior to cell preparation, a flow biofilm system was incubated at 37ºC filled with M9 glucose medium. Overnight cultures of ETEC H10407 pJB39 and ETEC H10407Δ*cexE* pJB39 were used to inoculate LB. Cultures were grown at 37ºC with aeration to an OD_{600} of 0.3. The cells were harvested and injected into the chambers of the flow biofilm system. The cells were allowed to adhere to the flow cells prior to the commencement of the liquid flow. After overnight incubation at 37ºC with a constant flow, the flow chambers were inspected for the formation of biofilm using the Leica TCS SP8 confocal microscope (Fig. 4.11). There was less

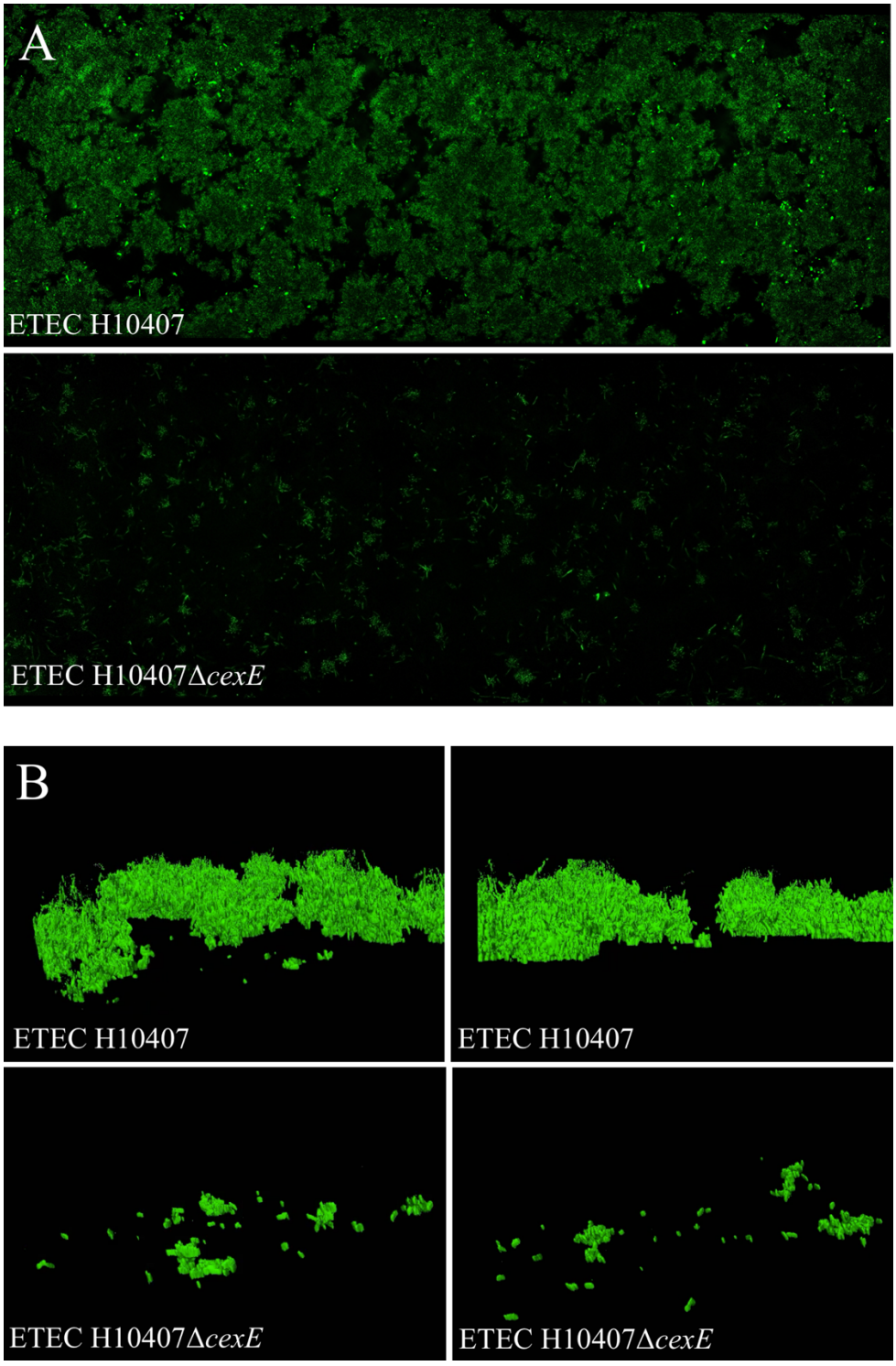


Figure 4.11: Confocal microscope images of ETEC H10407 biofilm. Bacterial biofilms were grown in triplicate; panel A and B are representative of biofilms grown in triplicate from separate experiments. A, 2D images taken from a vertical position of the ETEC H10407 pJB39 and ETEC H10407Δ*cexE* pJB39 biofilm. B, 3D images of ETEC H10407 pJB39 and ETEC H10407Δ*cexE* pJB39 biofilm from different angles.

intensity of the GFP signal present in each of the three flow cell chambers containing the *cexE* mutant compared to those containing the wild-type (Fig. 4.11A). Successive horizontal planes of the biofilm were imaged. These were compiled into a z-stack. Leica Application Suite X was used create a 3D image of the biofilm using the z-stack (Fig. 4.11B). The images presented are representative of the biofilm for the *cexE* mutant and the wild-type. The images are of the same section of biofilm from different angles. The 3D structure of the *cexE* mutant biofilm was different to that of the wild-type. The *cexE* mutant biofilm lacks the 3D structure of the wild-type. Therefore, it was concluded that CexE was involved in biofilm formation.

4.2.11. Electrophoretic mobility shift assay of CexE

Extracellular DNA is a major component of bacterial biofilms (Flemming et al., 2016). A protein with the ability to bind DNA would aid the formation of a biofilm. Thus, a dilution series of purified CexE-6His was made to examine whether CexE can bind DNA. Each dilution in the series was incubated separately with a ^{32}P labelled *CC*(−41.5) DNA fragment for 15 minutes at room temperature. The *CC*(−41.5) fragment is a derivative of the *melR* promoter that has a CRP site 41.5 base pairs upstream transcription start site (Gaston et al., 1990). A no protein control was included. Each sample was loaded onto an acrylamide TBE gel, and the DNA with or without protein attached were separated by electrophoresis. The gel was dried and the position of the radioactive DNA was visualised using a phosphor screen. The ability of CexE-6His to retard migration of the *CC*(−41.5) fragment in the gel was assessed (Fig. 4.12).

At the first concentration of CexE-6His (0.5 μM) there was a reduction in the intensity of the ^{32}P labelled DNA band compared to the no protein control. The intensity of the DNA fragment decreased proportionally with the increase in CexE-6His concentration. The retardation of the DNA by 12.5 μM CexE-6His was substantial and complete at 20 μM CexE-6His. At 12.5 μM

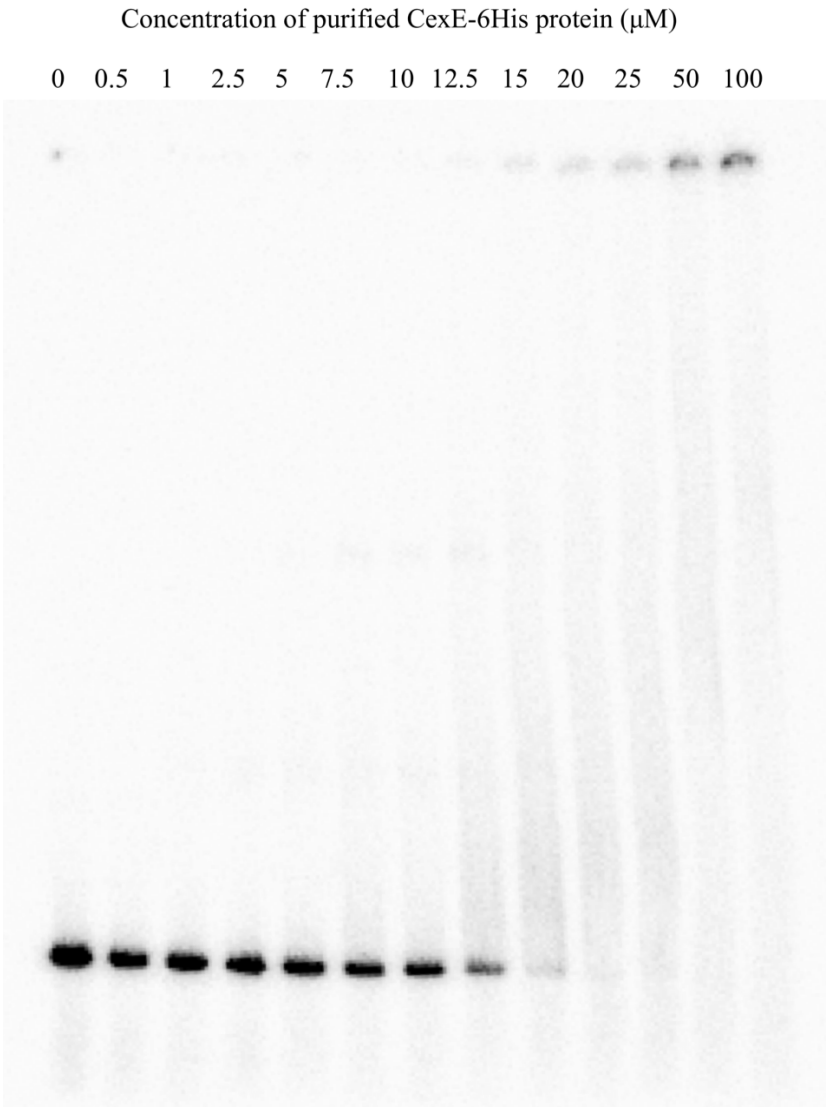


Figure 4.12: Retardation of ^{32}P labelled DNA after incubation with CexE-6His. CexE-6His was incubated for 15 minutes at room temperature with the CC(−41.5) DNA fragment. The concentrations of CexE-6His indicated were the final concentration of CexE-6His in each reaction. The above figure is representative of two replicates of the same experiment.

there was signal indicating retention of DNA fragment in the wells of the gel. This increased with CexE-6His concentration to the upper concentration of 100 μM. At 12.5 μM there was a smear of DNA in the lane. The start of the smear increased in height with concentration. The retardation of DNA in an EMSA increased with the concentration of CexE-6His. Therefore, it was concluded that CexE-6His binds to DNA.

4.2.12. Purification and assessment of CexE homologs to impair the migration of DNA in an electrophoretic mobility shift assay

To investigate whether the DNA binding was common to other CexE homologs, Aap and CexE from *C. rodentium* ICC168 (from now on termed $CexE_{ICC168}$), were purified. Expression plasmids for the production of Aap and $CexE_{ICC168}$ were made as described for CexE. In brief, *aap* and $cexE_{ICC168}$ were cloned from EAEC 042 and *C. rodentium* ICC168, respectively, using the primers 042/*aap*/Clone-FW, 042/*aap*/Clone-RV, ICC168/*cexE*/Clone-FW and ICC168/*cexE*/Clone-RV. The PCR products were purified and ligated into the pET26b(+) plasmid. The completed ligation reactions were transformed into *E. coli* DH5α. Successful ligations were sequenced. Plasmids with the correct sequence were termed pET26b-*aap* and pET26b-$cexE_{ICC168}$ for *aap* and $cexE_{ICC168}$, respectively. Plasmids were transformed into *E. coli* BL21(DE3). Overnight cultures of *E. coli* BL21(DE3) containing the respective plasmid were used to inoculate, separately, 2 L LB. Cultures were grown as described for the production of CexE-6His. Cells were harvested and lysed. Intact cells were removed by centrifugation. Membranes were removed by high speed centrifugation. The soluble protein fractions were filtered and the filtrate was cycled through a 1 ml HisTrap™ HP column at 4 ºC. Columns were washed and the protein was eluted with imidazole. The eluted samples of purified Aap-6His or $CexE_{ICC168}$-6His protein were concentrated. Thus, Aap and $CexE_{ICC168}$ were successfully purified.

The Aap protein modulates the cell surface charge to allow the extension of the AAF/II (Velarde et al., 2007). If the charge of the homologous CexE protein was responsible for DNA binding, it was hypothesised that Aap would also bind DNA. Thus, Aap-6His was incubated with ^{32}P labelled *CC*(−41.5) DNA fragments at room temperature as described for CexE-6His. The reactions were separated by electrophoresis on an acrylamide TBE gel. The gel was visualised using a phosphor screen for the retardation of the DNA fragment (Fig. 4.13). There was no retardation of the DNA fragment after incubation with Aap-6His up to 50 μM. There was a reduction in the intensity of the DNA fragment signal 100 μM Aap-6His. There was slight smearing and retention in the well at the same concentration. However, the concentration required for those slight effects was very high. It was not comparable to that of the CexE-6His protein. Therefore, it was concluded that Aap-6His did not bind DNA.

The $CexE_{ICC168}$ protein sequence did not clade with Aap or CexE. It was unknown if the $CexE_{ICC168}$-6His could bind DNA. Thus, as described for Aap and CexE, $CexE_{ICC168}$ was incubated with ^{32}P labelled DNA at room temperature. The reactions were separated by electrophoresis. The gel was visualised for retardation of the DNA fragment (Fig. 4.14). The intensity of the DNA band decreased slightly with an increase in protein concentration. There were faint bands present in samples that had been incubated with 10 μM $CexE_{ICC168}$-6His, that were clearer at 15 μM. However, there was only substantial retardation of the DNA band at 50 μM and 100 μM $CexE_{ICC168}$-6His. At 100 μM $CexE_{ICC168}$-6His there was still DNA present that was not retarded when compared to the no protein control. The retardation of DNA by $CexE_{ICC168}$-6His was not comparable to the CexE-6His protein. Therefore, it was concluded that $CexE_{ICC168}$ did not bind DNA.

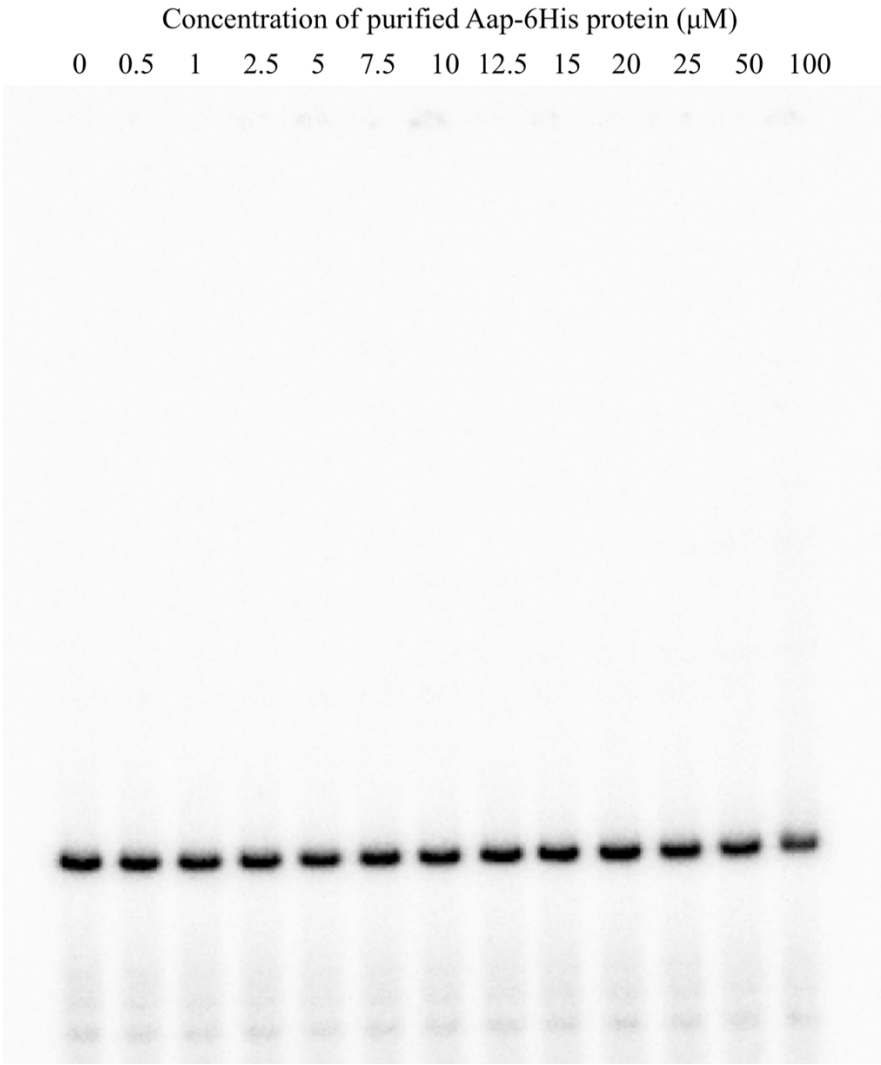


Figure 4.13: Retardation of ^{32}P labelled DNA after incubation with Aap-6His. Aap-6His was incubated for 15 minutes at room temperature with the DNA fragments. The above figure is representative of two replicates of the same experiment.

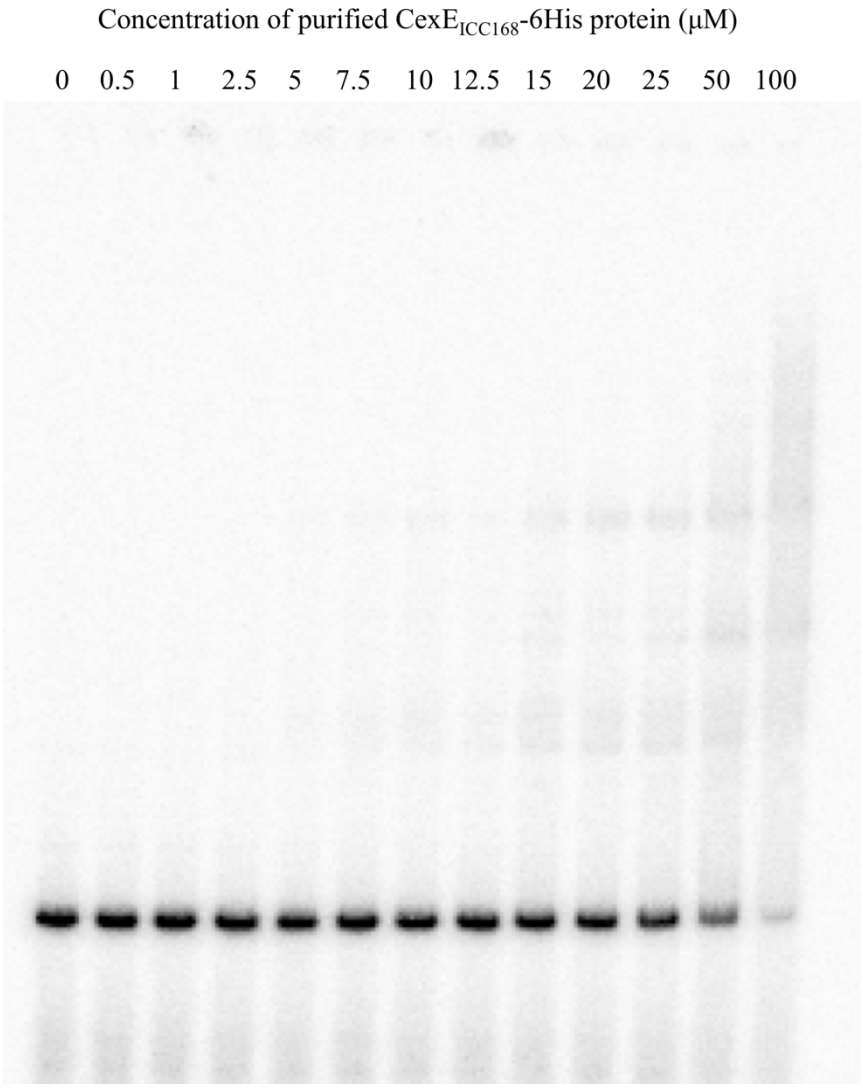


Figure 4.14: Retardation of ^{32}P labelled DNA after incubation with CexE$_{ICC168}$-6His. CexE$_{ICC168}$-6His was incubated as described for Aap. The above figure is representative of two replicates of the same experiment.

4.2.13. Comparison of the ability of Aap-6His, $CexE_{ICC168}$-6His and CexE-6His to retard DNA in an electrophoretic mobility shift assay

To confirm the results above the Ability of the three proteins, Aap-6His, $CexE_{ICC168}$-6His and CexE-6His, to bind DNA were compared in the same experiment. The DNA fragment incubated with the Aap-6His protein were not retarded in the gel at all (Fig. 4.15). There was no difference in the intensity of the DNA band of the no protein control to that incubated with 50 μM Aap-6His. There was a decrease in the signal intensity of the DNA fragment with an increase in concentration of $CexE_{ICC168}$-6His. There was a slight reduction of DNA signal at 12.5 μM compared to the no protein control and a substantial reduction at 50 μM $CexE_{ICC168}$-6His. The decrease in signal intensity due to an increase in protein concentration was substantially more for CexE-6His compared $CexE_{ICC168}$-6His. The DNA signal was decreased at 2.5 μM CexE-6His. When incubated with 5 μM CexE-6His, a smear was apparent indicating DNA retardation. The migration of the DNA fragment was completely retarded at 50 μM CexE-6His. The DNA was present as a smear. The retardation of the DNA fragments incubated with CexE-6His was not comparable to Aap-6His or $CexE_{ICC168}$-6His. Therefore, it was concluded that CexE-6His was able to bind DNA to a substantially greater affinity than the other proteins tested.

4.3. Discussion

Aap and CexE were distributed in three main pathovars: *Y. enterocolitica*; EAEC and ETEC. Multiple copies of *aap* or *cexE* in a single genome were common. Therefore, *aap* or *cexE* are highly mobile genes. The mobility of the *aap* or *cexE* genes lead to examples of single strains within a species containing the gene, such as those observed for *Salmonella* and *Shigella*. Sahl et al. (2011) failed to identify a *cexE* homolog in ETEC E24377A or ETEC B7A, possibly due to the problems with *aap* and *cexE* annotation. During the course of the analysis it was clear

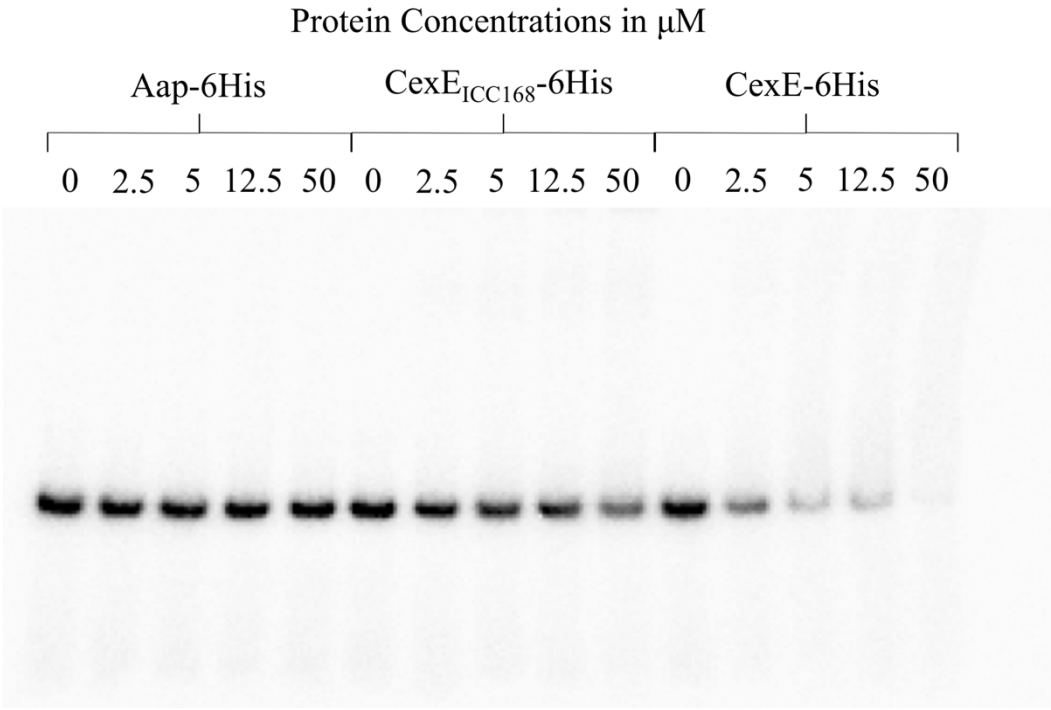


Figure 4.15: Comparison of DNA binding ability. Aap-6His, CexE$_{ICC168}$-6His and CexE-6His were incubated with ^{32}P DNA in 20 mM Tris-HCl (pH 7.0), 10 mM MgCl2, 0.1 mM EDTA, 120 mM KCl, 2 mM DTT, 5% glycerol and 7 μM BSA for 15 minutes at room temperature. The reaction mixes were separated on a 6% acrylamide 0.5x TBE gel by electrophoresis. The migration of the radioactive DNA was visualised using phosphor screens. The concentrations of the proteins are indicated above the gel image.

that current sequence annotation methods commonly missed genes encoding Aap or CexE. Thus, the full prevalence of *aap* and *cexE* is difficult to discern. If possible, an analysis including the unannotated *aap*/*cexE* genes is not thought to greatly affect the major clades in Fig. 4.2. It would instead change the numbers of *aap*/*cexE* genes attributed to each clade and sequence.

The function of Aap is characterised as modulating the cell surface charge to allow the extension of the AAF/II (Velarde et al., 2007). CexE does perform the same function when produced in EAEC. If CexE did have a similar function to allow the extension of the CFA/I from ETEC H10407 it would be expected to complement or not alter the *aap* mutant phenotype. Instead the production of CexE significantly increased the aggregation of EAEC 042. Thus, the CexE protein does not have a similar function to that of Aap in EAEC. It is inferred from the evidence presented here the CexE protein is involved in binding extracellular DNA, which in turn promotes the formation of biofilm an ETEC H10407. This is the first report of a *cexE* mutant not being able to form biofilm. Biofilm formation is not typically associated with ETEC pathogenesis. The biofilm investigated here was formed on an abiotic surface. It is not thought that CexE initiates the initial attachment to the abiotic surface. Instead it is thought to strengthen the biofilm formation. There is evidence of ETEC biofilms acting as a reservoir for infection (Ahmed et al., 2012). This suggests that CexE provides an advantage to ETEC strains to form biofilms on abiotic surfaces creating reservoirs for infection.

ETEC H10407 would not produce CexE under normal laboratory conditions. It required growth in a flow biofilm environment before a phenotype could be observed. However, when the gene was cloned into pET26b there was production of CexE-6His without induction, which was to such a significant amount as to be clearly visible on a Coomassie Brilliant Blue stain of

a whole cell lysate. Although CexE-6His was produced from a multi-copy number vector in an overexpression strain, it is not typical to see proteins produced to a visible concentration without induction. This suggests a strong mechanism for the prevention of CexE production in ETEC H10407. The low GC content of the upstream region of *cexE* could enable H-NS binding to prevent transcription (Lucchini et al., 2006). This likely repression of *cexE* expression is relieved by CfaD (Pilonieta et al., 2007; Hodson et al., 2017). Another mechanism preventing the production of CexE could be present at the translation level. The start codon of *cexE* in ETEC H10407 is flanked by 4 adenosines before the ATG and 6 thymines after. It appears likely that when transcribed, a loop would form in the mRNA limiting access to the start codon and thus preventing translation. The upstream adenosines were removed when the *cexE* gene was cloned into pET26b. This is likely to relieve repression of *cexE* expression in the absence of induction. It is not clear if this possible loop is responsible for the lack of production of CexE in ETEC H10407. From the RNA-Seq data it is known that *aap* is highly expressed under the correct conditions. Aap is produced when EAEC 042 is grown overnight in LB. Thus, the expression of *cexE* appears to be more tightly controlled than that of *aap*. While not within the remit of this study, an investigation into the mechanism of CexE production in ETEC H10407 is warranted.

The extremely high concentration of purified CexE-6His attained indicates the inherent stability of the protein. The ability to purify high quantities of CexE-6His lends to structural analysis. Velarde et al., (2007) elucidated the structure of Aap. However, due to the low sequence identity between the proteins, this structure is unlikely to be predictive of the structure of CexE.

CHAPTER 5:

Structural analysis of CexE from ETEC H10407.

5.1. Introduction

The structure of the enterotoxigenic *E. coli* (ETEC) CexE homolog, Aap, in enteroaggregative *E. coli* 042 (EAEC) was previously determined (Crossman et al., 2010; Velarde et al., 2007). The structure of Aap is a β-sandwich, which consists of 6 β strands forming two antiparallel β sheets with two polar α helices (Fig. 5.1). As Aap is secreted by the Aat system, which is an atypical type I secretion system, it must traverse the AatA outer membrane protein (Nishi et al., 2003). The bullet-like structure of Aap, 50x20x20 Å, is amenable to exiting the cell via the AatA open pore that is ~20 Å in diameter (Velarde et al., 2007; Nishi et al., 2003). The structural analysis of Aap indicated an abundance of positively charged residues in one hemisphere of the protein (Fig. 5.2), which led to the hypothesis that the aggregative adherence fimbriae (AAF) are electrostatically attracted to the cell surface. Aap binds to the LPS electrostatically and blocks the AAF attraction, thereby allowing their extension (Velarde et al., 2007). However, the major subunits of the ETEC colonisation factor antigen I (CFA/I) and AAF do not have the same charge at a neutral pH (Velarde et al., 2007). Therefore, the function of CexE cannot be the same.

In chapter 4, the CexE protein from the prototypical ETEC H10407 was found to bind DNA. The two homologous proteins tested, Aap and $CexE_{ICC168}$ from *C. rodentium* ICC168, could not bind DNA. Thus, CexE is functionally different from Aap. The determination of the structure of CexE can help elucidate the mechanism of DNA binding. The purification of high quantities of CexE-6His described in chapter 4 suggest that it might be possible to determine the structure of CexE. The aim of this work was to investigate the structure of CexE with circular dichroism and nuclear magnetic resonance (NMR) spectroscopy.

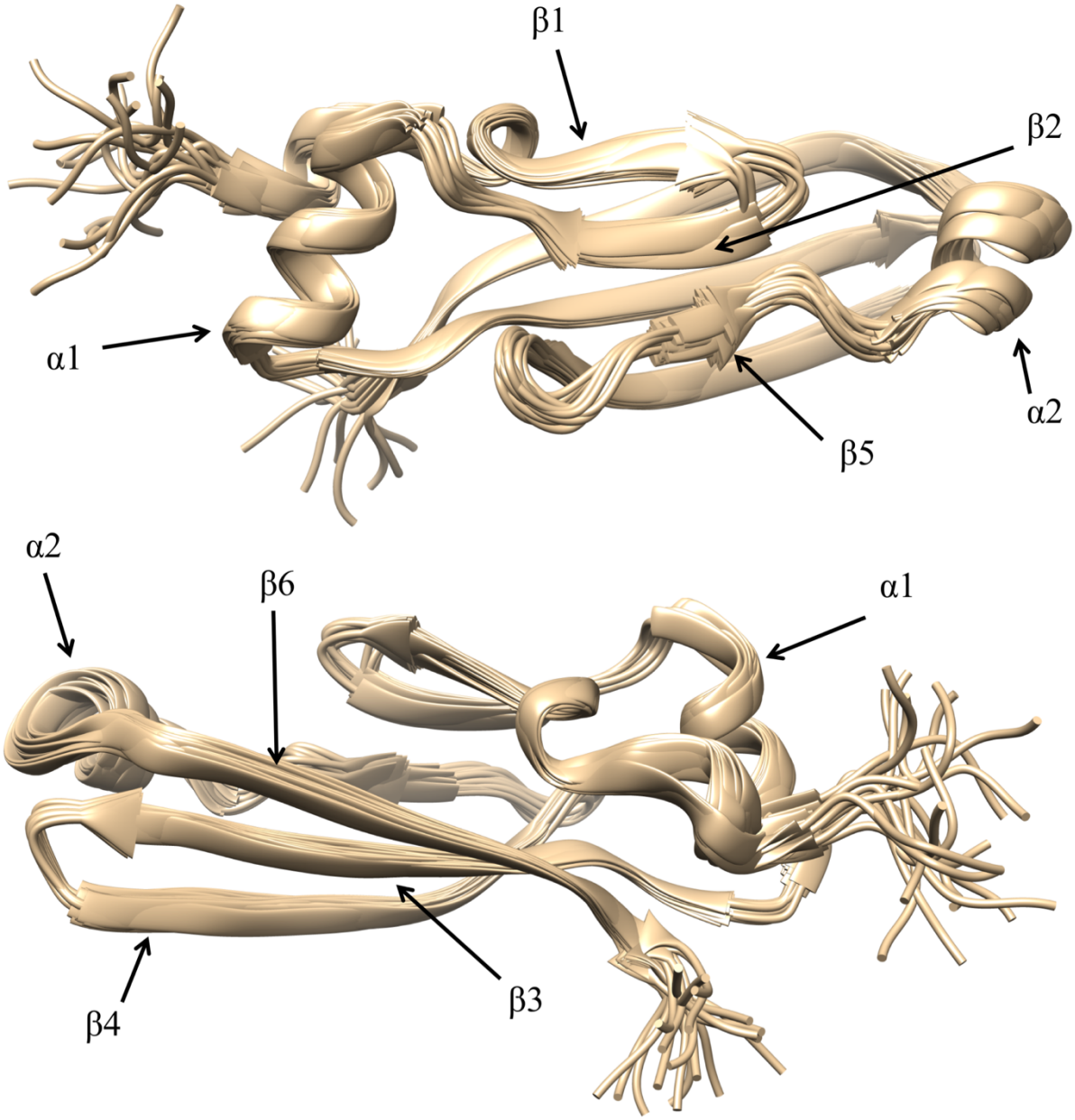


Figure 5.1: Ensemble of Aap solved by Velarde et al. (2007). The 20 Aap NMR solution structures were downloaded from the PDB file 2jvu. The two images were taken from 180º turns of the 20 Aap structures. In numerical order from the N-terminus the secondary structure is labelled α 1 and 2 for the two α helices and β 1 to 6 for the six β strands. The labelling is the same as that of Velarde et al. (2006). The images were taken using Chimera.

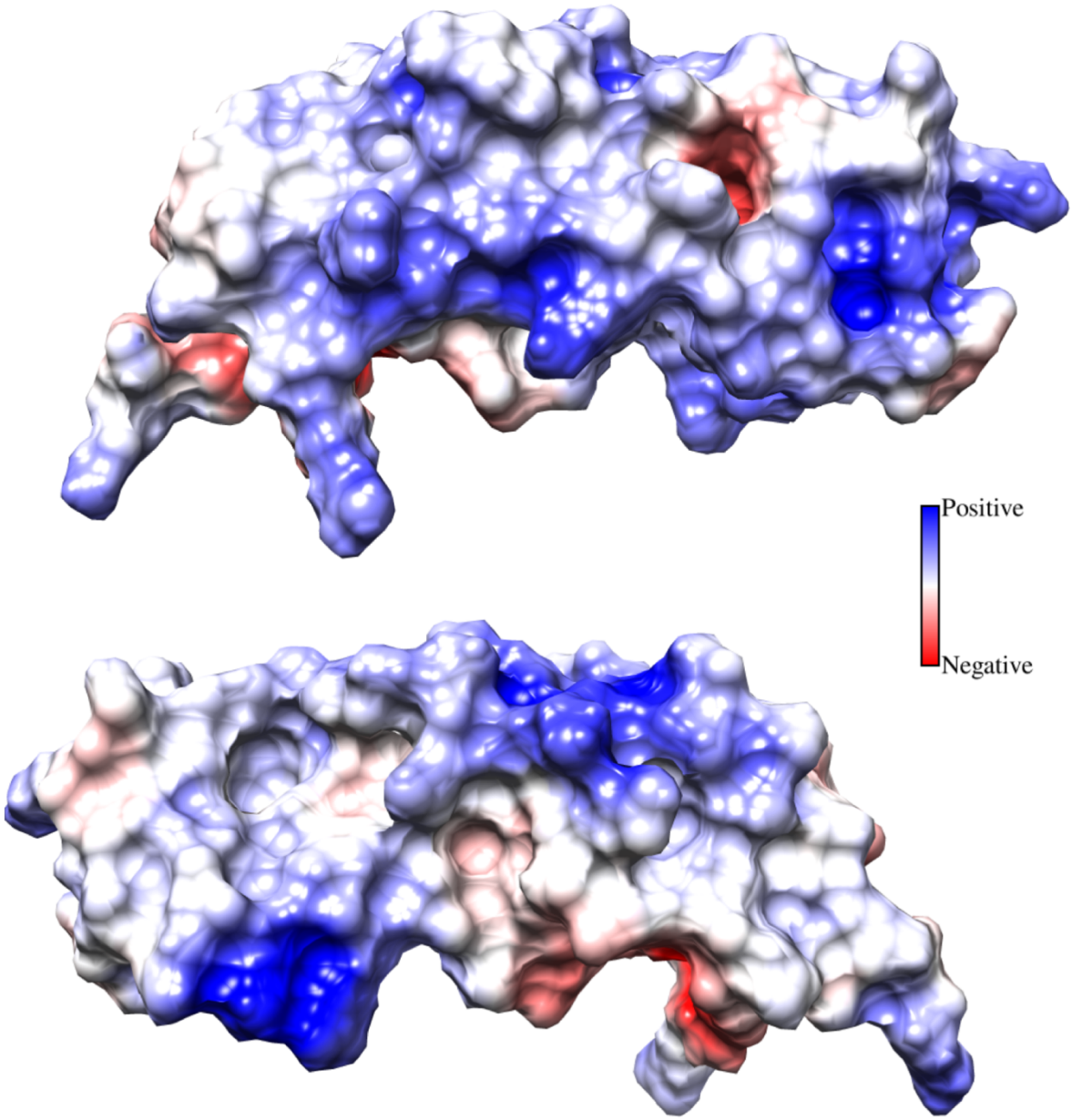


Figure 5.2: Electrostatic surface charge of Aap solved by Velarde et al. (2007). The electrostatic surface charge was calculated using PDB2PQR and APBS functions in Chimera from using the best representative conformer of Aap from the 2juv PBD file. The electrostatic surface charge calculated above is similar to that presented for Aap in Velarde et al. (2007). There is a localisation of positive charge in the lower hemisphere of Aap. The two images were taken from 180º turns of the Aap surface structure.

5.1.1. Principles of circular dichroism

Circular dichroism uses light polarisation to estimate the secondary structure of proteins (Kelly et al., 2005). The oscillations of linearly polarised light are in a single plane. Polarised light is the sum of two linearly polarised light waves that are in the same phase at right angles to each other. If one of the linearly polarised light waves is out of phase by a quarter turn the resultant light wave will be helical, termed circularly polarised light. The turns of the helices will either be clockwise, left circularly polarised light, or counter-clockwise, right circularly polarised light. The superposition of the left and right circularly polarised light can be viewed as a linear polarised wave. Left and right polarised light are absorbed by chiral molecules by different amounts. This is called circular dichroism. After passage through a chiral sample the superposition of the left and right polarised light is no longer linearly polarised. The superimposed left and right circularly polarised light is now elliptical. The amount of circular dichroism can be compared between samples by measuring their ellipticity. The major radius of the ellipse is defined by the amplitude of the left and right circularly polarised light, while the minor radius is the difference between the left and right polarised light. The degree of ellipticity is the tangent of the ratio of the minor to major elliptical axis. The size of the degree of ellipticity (θ) is very small and is typically measured in millidegrees (mdeg). Over a range of wavelengths from 180 nm to 240 nm, a characteristic circular dichroism spectrum arises from protein folds like α helices and β sheets. Using algorithms, such as BeStSel, the circular dichroism spectra of protein structures determined by X-ray crystallography can be used to estimate the folding of a protein (Micsonai et al., 2015).

5.1.2. Principles of nuclear magnetic resonance spectroscopy

NMR spectroscopy can be used to determine the structure of a protein. NMR makes use of spin, which is a fundamental physical property possessed by all protons, neutrons, and

electrons. Spin is a form of angular momentum, which occurs when a particle has a twisting or rotating trajectory. The spin of a nucleus is determined by the composition of protons and neutrons. Both protons and neutrons have spin. If the total number of neutrons and protons in a nucleus is even (^{2}H, ^{12}C, ^{14}N) the nucleus does not have spin. The spin of the protons and neutrons is paired and cancelled out, the nucleus is said to have a paired spin. When the nucleus of the atom contains an odd number of total protons and neutrons (^{1}H, ^{13}C, ^{15}N) the nucleus of the atom has an overall spin, due to their being an unpaired spin. The spin of a nucleus is used by nuclear magnetic resonance (NMR) spectroscopy to provide chemical information about that atom. A spinning nucleus generates a magnetic field, which like the earth's magnetic field is off centre. Nuclei with a spin producing a magnetic field have a nuclear magnetic moment in a magnetic field. When an external magnetic field is applied, it is energetically preferred for the magnetic moment to be aligned with the external magnetic field (Fig. 5.3A). These are referred to as the low-energy state, magnetic moment aligned with the external field, and the high-energy state, magnetic moment opposed to the magnetic field. The amount of energy required to move the nucleus from the low to the high-energy state is increased with the power of the magnet. A larger difference in energy is easier to observe. This is why NMR spectroscopy uses very powerful magnets. In an NMR experiment the energy required to align the magnetic moment against the magnetic field is comparatively low. Energy from collisions is sufficient to promote nuclei to the higher energy state. It is easier to consider the net magnetic moment of a sample or magnetisation vector. This is a sum of all the individual magnetic moments. The net magnetic moment of a sample will align with the external magnetic field along the z axis (Fig. 5.3B). If a pulse of electromagnetic radiation around the radio frequency is applied at 90º along the y axis to the external magnetic field the electromagnetic radiation with the required energy for an increase in energy state will be absorbed and the net magnetism will align with the y axis (Fig. 5.3C). When the pulse is removed the net magnetism with

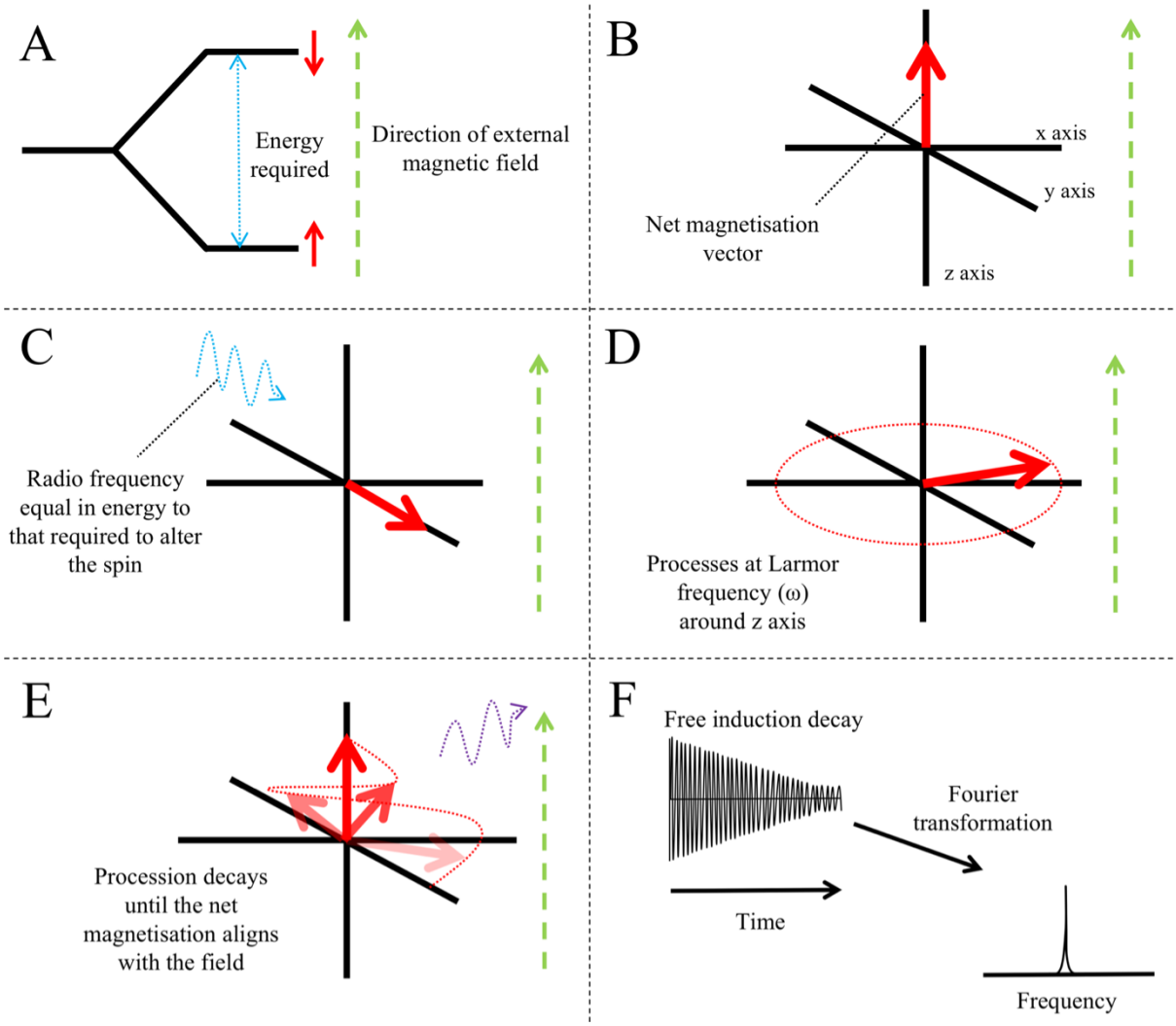


Figure 5.3: Principles of NMR spectroscopy. A, the energy required to change the state of the spin of a nucleus in an external magnetic field. B, the net magnetisation moment is comprised of all the spin states of the sample. The net magnetisation moment aligns with the external magnetic field. C, a pulse of radio waves at 90º to the z axis is applied. The radio waves with the correct energy for a change in spin state (A) are absorbed. The magnetisation moment moves from the z axis to the xy plane. D, after the pulse is removed the magnetisation moment processes around the z axis at the Larmor frequency. E, the procession of the magnetisation moment decays and reaches equilibrium with the external magnetic field. F, the decay of the magnetic moment causes a change in the magnetic field in the sample. This causes a current in the coil surrounding the sample resulting in the detection of the free induction decay. This is transformed from time vector to a frequency using the Fourier transformation.

process around the z axis at the Larmor frequency. This frequency is different for different nuclei (Fig. 5.3D). The frequency of the electromagnetic radiation absorbed by the nucleus is equal to the Larmor frequency. Therefore, it is said to resonate. This procession degrades as the net magnetisation vector aligns with the magnetic field (Fig. 5.3E). The degradation in the procession around the z axis changes the magnetic field. This change in magnetic field causes a current to flow in the coils surrounding the sample. The signal generated due to the flow of current is known as the free induction decay (Fig. 5.3F). This is a time vector that is transformed into a frequency vector by Fourier transformation.

The above example only considers a sample of one species of nuclei. This is not present in nature because nuclei are surrounded by electron density. The circulation of the electrons around the nucleus of the atom creates a magnetic force that opposes that of the external magnetism (Harris, 1983). The electron density around a nucleus is dependent on the molecule to which that atom is bound. For instance, the electron density of a hydrogen bound to the carbon atom on a methane molecule will be different to that on a methanol molecule. In fact, the three hydrogens bound to the carbon on the methanol molecule will be different to the one bound to the oxygen. The ^{1}H nuclei of the molecules will have different frequencies. However, the frequency observed is dependent on the NMR machine used, temperature, magnetic field strength, and other variables. Therefore, the frequency of nuclei is not reported in hertz but instead in chemical shift (parts per million). Chemical shift is calculated by comparing the frequency of the molecule tested to that of a standard, which is typically tetramethylsaline $Si(CH_3)_4$. This allows the comparison of NMR spectra between different machines. The composition of an unknown sample can be determined if the chemical shifts have been previously observed in a known sample. Using these principles, the chemical shifts of amino acids in a protein can be identified in a spectrum.

The spin of a nucleus can be altered not just by the nuclei it is chemically bound to but also to those to which it is spatially close. The nuclear Overhauser effect (NOE) is used to define nuclei that are close in space to each other either due to chemical bonds or tertiary protein structure. The NOE is caused when the spin of a nucleus is promoted to the higher energy state by irradiation. This results in a change in the spin of nuclei nearby, located within a distance of 5Å. The spin of these nuclei can be observed. Using the chemical shifts from these alterations in spin, a three-dimensional structure of the protein can be calculated.

5.1.3. NMR experiments for structure determination

NMR spectra for protein determination are either 2D or 3D. These spectra are generated by the alteration of the time between a second 90º pulse. This provides separate planes of chemical shifts that can be built into a 2D or 3D spectrum. Since the chemical shifts of ^{1}H are far easier to detect than ^{13}C or ^{15}N, a technique called insensitive nuclei enhanced by polarization transfer (INCPT) is used. In basic terms the magnetisation of the ^{1}H is transferred to the nuclei of interest, i.e. ^{13}C or ^{15}N. After an amount to time required to gain information about the nuclei the magnetisation is returned to the ^{1}H for detection.

A number of NMR experiments are used to find the position of the residues relative to each other in space. These provide information about the chemical environment of each nucleus. A schematic of NMR experiments used in this Chapter, and spectra they generate, is provided in Fig. 5.4. The Cα are termed i and i−1, this denote the position of the Cα in relation to the amine group. A i Cα is the Cα bound to the amine, and the i−1 Cα is the Cα connected to the amine via the carbonyl group. In the ^{1}H-^{15}N-HSQC magnetisation is transferred from the HN group, which is the ^{1}H bound to the ^{15}N in the amide group, to the nitrogen before it is returned for

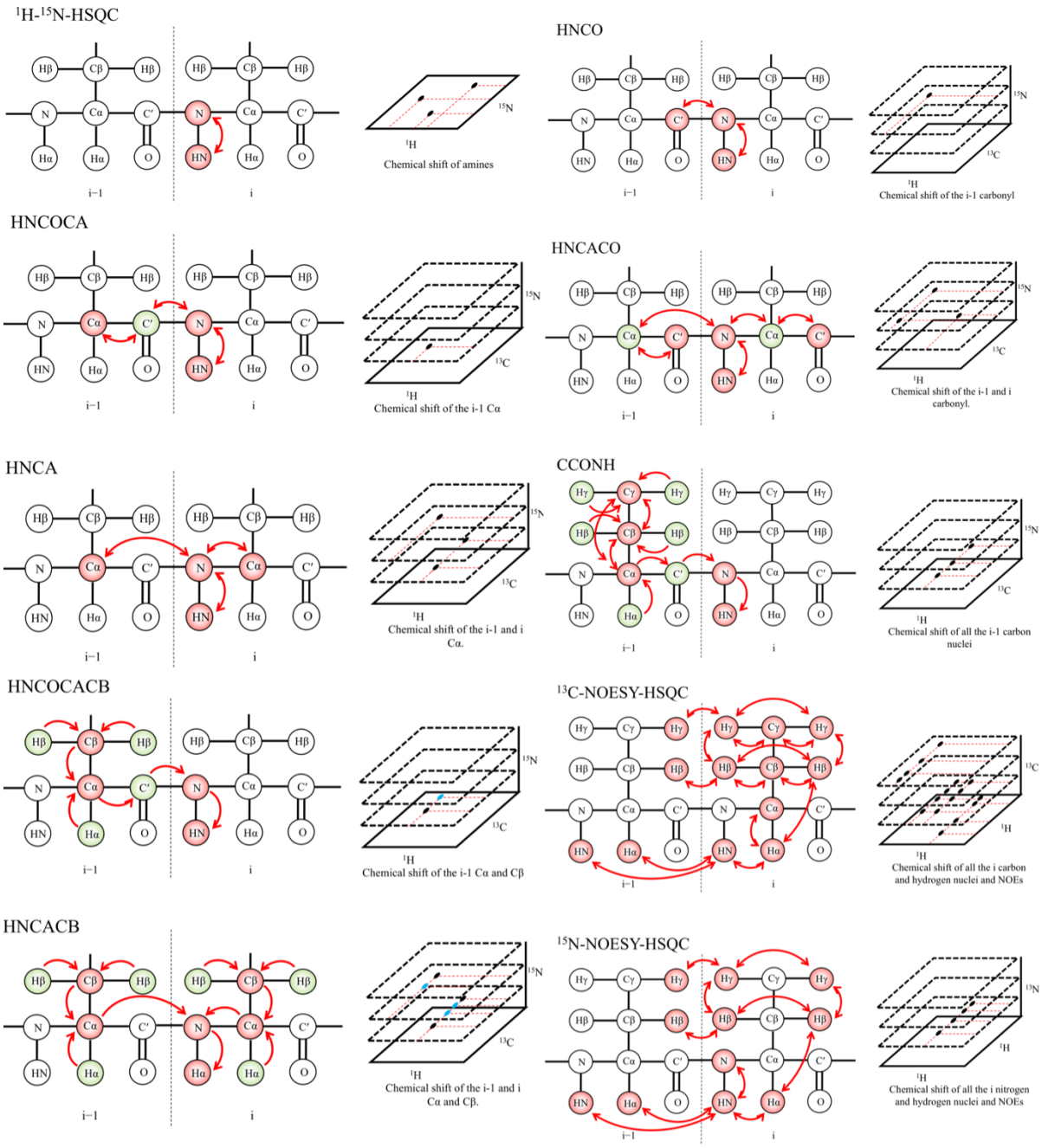


Figure 5.4: NMR experiments for protein structure determination. Red arrows represent magnetisation transfer. The red coloured nuclei are those that chemical shift information is gathered. No chemical shift information is collected about the green nuclei. However, they are involved in magnetisation transfer. Uncoloured nuclei are not involved in the experiment. The spectrum generated is on the right of the molecular view of the experiment. Positive chemical shifts are in black and negative chemical shifts are in blue.

detection (Cavanagh et al., 2006). This provides a 2D spectrum with the ^{1}H and ^{15}N chemical shifts on the x and y axis respectively. From the chemical shifts of the ^{15}N and ^{1}H the chemical shift of the Cα in the i-1 and i residue can be obtained using the HNCA and HNCOCA. In HNCA, the magnetisation is passed from the HN to the ^{15}N and to the Cα, a 3D spectrum is produced (Grzesiek and Bax, 1992). The HNCOCA is very similar however the magnetisation is passed from ^{15}N to the C′ and then to the i-1 Cα, the only carbon chemical shift is that of the i-1 Cα (Grzesiek and Bax, 1992). Using the HNCA and HNCOCA spectra the chemical shifts of the i-1 and i Cα can be determined. In the HNCACB experiment the magnetisation is transferred from the Hβ to the Cβ and Hα to the Cα, magnetisation is then transferred from the Cβ to the Cα, then to the ^{15}N and to the HN for detection (Grzesiek and Bax, 1992). The HNCOCACB similarly to the HNCOCA uses the carbonyl group to limit the chemical shifts to the i-1 residue (Grzesiek and Bax, 1992). Using these two spectra the chemical shift of the Cβ carbons can be obtained. In the HNCACO experiment the magnetisation is transferred to the ^{15}N from the HN and via the i-1 Cα to the C′ it is then returned along the same path for detection at the HN (Bax and Ikura, 1991). Thus, the chemical shifts of the i-1 and i carbonyls are collected. The HNCO transfers magnetism from the HN to the ^{15}N and then to the C′ (Muhandiram and Kay, 1994). The chemical shift of the i-1 carbonyl is defined.

From these seven experiments the chemical shifts associated with the protein backbone can be identified. These chemical shifts are used with the CCONH spectrum to identify the chemical shifts of the other ^{13}C in the amino acid side chain. Magnetisation is transferred from the associated ^{1}H to the ^{13}C nuclei of the side chain. This is then transferred down the side chain to the Cα. Magnetisation is then transferred to the carbonyl, then to the ^{15}N and finally to the HN for detection (Grzesiek et al., 1993).

Now the chemical shifts of the ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC can be assigned. The chemical shifts of only one of the aliphatic or aromatic nuclei can be collected for the ^{13}C-NOESY-HSQC in a single experiment. These NOSEY experiments provide spatial information of the ^{1}H in relation to one another in the protein. These experiments make use of the nuclear Overhauser effect to identify ^{1}H that are located within 5 Å. Due to their position in the residue or tertiary structure. Magnetisation is exchanged between all of the ^{1}H within 5 Å, and is then transferred to the ^{13}C or ^{15}N for detection on the Hα or HN for ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC, respectively (Marion et al., 1989). The structure is modelled using these chemical shifts, the sequence of the protein, and peptide bond contortion angles. NMR modelling algorithms provide 20 models that best fit the experimental data provided. The similarity of these models to each other are reported by the root-mean-square deviation of atomic particle (RMSD). This is a measure of the average distance between atoms of superimposed structures.

5.2. Results

5.2.1. N-terminal sequencing of CexE-6His

The protein for structural determination should have its native fold, which can be dependent on the compartment it is purified from. The CexE protein was predicted, by SignalP, to be secreted by the general secretory pathway into the periplasm resulting the removal of the first 20 N-terminal amino acids. The N-terminus of the CexE-6His was determined by sequencing five of the N-terminal amino acids. Purified CexE-6His, from Chapter 4, was separated by electrophoresis on Tris-tricine gels, then transferred to PVDF and stained with Ponceau S. There was a band at the expected molecular weight of CexE-6His. This region was excised, the stain was removed with NaOH and sent to Alta-Innovations at the University of Birmingham for N-terminal sequencing. The five N-terminal residues of CexE-6His were

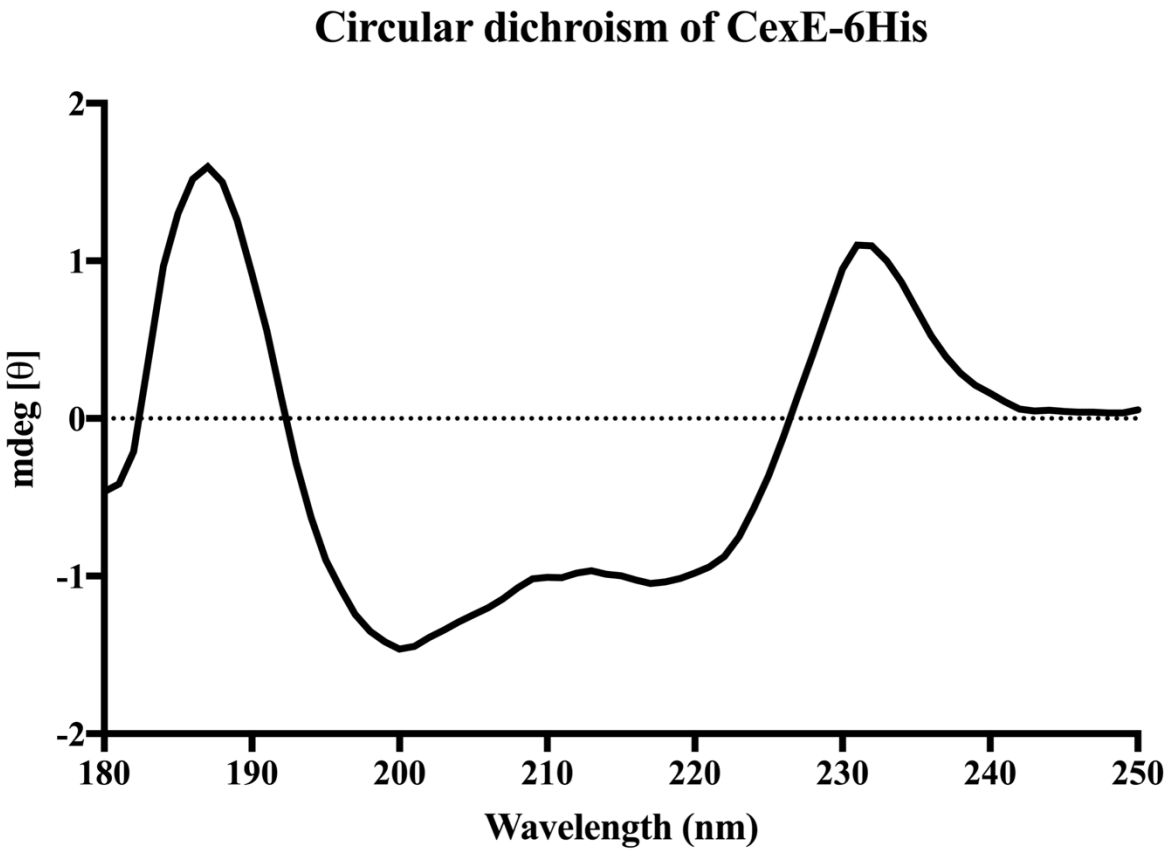
GGGNS. These were the N-terminal residues predicted by SignalP after signal peptidase I cleavage of the full length CexE-6His protein. Therefore, it was concluded that CexE-6His purified was secreted into the periplasm similar to its native mechanism.

5.2.2. Investigation of CexE-6His secondary structure using circular dichroism

From the circular dichroism spectra of experimentally determined protein structures, the unknown secondary structure of a protein can be estimated. The CexE-6His protein, which was purified in chapter 4, was subjected to circular dichroism analysis. CexE-6His was diluted to a concentration of 0.1 mg/ml and buffer exchanged by membrane filtration. The degree of ellipticity of CexE-6His was measured at a range of wavelengths from 180 to 250 nm with 1 nm increments and a curve between the degree of epilticity points was drawn (Fig. 5.5). The curve of the circular dichroism spectrum confirms that the protein is folded. The mdeg values were submitted to BeStSel for secondary structure estimation (Micsonai et al., 2015). Just below the majority, 47.9%, of the secondary structure of CexE-6His was estimated to be anti-β sheet (4.9% left twisted, 15.5% relaxed and 27.5% right twisted). A substantial amount, 38.9%, was not clearly estimated with the remainder estimated to be turns. These estimations suggest that the secondary structure of CexE is similar to that of Aap.

5.2.3. Purification of CexE-6His composed of ^{13}C and ^{15}N nuclei

The secondary structure of CexE was predicted to be similar to that of Aap but they have alternative functions. Therefore, to further investigate the structure of CexE the protein was produced for NMR spectroscopy. The carbon and nitrogen nuclei of proteins produced by cultures grown in LB have a paired spin, i.e. composed of ^{12}C and ^{14}N. For there to be an NMR signal the proteins must be made of atoms with unpaired spins, i.e. ^{13}C and ^{15}N. An overnight culture of *E. coli* BL21(DE3) freshly transformed with pET26b-*cexE* was pelleted by



Estimated secondary structure content by BeStSel

Secondary structure content	(%)
Helix1 (regular):	0.0
Helix2 (distorted):	0.0
Anti1 (left-twisted):	4.9
Anti2 (relaxed):	15.5
Anti3 (right-twisted):	27.5
Parallel:	0.0
Turn:	13.2
Others:	38.9

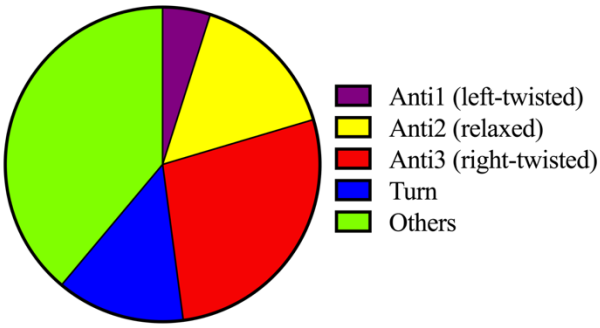


Figure 5.5: Circular dichroism and estimated secondary structure of CexE. The circular dichroism spectrum for CexE-6His is the curve at the top of the figure. Purified CexE-6His was buffer exchanged with 20 mM HEPES, 250 mM NaCl pH 7.3 and diluted to 0.1 mg/ml. The table included is the results of the secondary structure content estimated by BeStSel (Micsonai et al., 2015). The pie chart is a graphical representation of the estimation results.

centrifugation. The pellet was washed three times with PBS. The resuspended cells were used to inoculate 2 L of M9 salts supplemented with kanamycin and nutrient mix. The M9 medium was further supplemented with ^{13}C glucose and ^{15}N ammonium sulphate as the respective carbon and nitrogen sources. The cultures were grown to an OD_{600} of 0.3 at 37ºC with aeration, then cooled to 16ºC for 30 minutes, and the production of CexE-6His was induced with 50 μM IPTG. Cultures were grown overnight at 16ºC and cells were harvested by centrifugation. The cell pellet was resuspended in binding buffer and lysed using EmulsiFlex-C3. Unbroken cells were removed by centrifugation. Membranes were removed by high speed centrifugation and the soluble fraction was filtered. The filtrate was used to purify CexE-6His using a 5 ml HisTrap™ column, which was left to cycle through the column overnight at 4ºC. The column was washed with binding buffer and then successive wash buffers containing 50 mM and 100 mM imidazole. The protein was eluted with 500 mM imidazole. Fractions containing purified CexE-6His were combined and concentrated by membrane filtration. The imidazole was removed by dialysis. Thus, CexE-6His with unpaired ^{13}C and ^{15}N nuclei was purified. The purified CexE-6His composed of ^{13}C and ^{15}N was referred to as double labelled CexE-6His.

The removal of contaminating proteins from the sample reduces the noise in NMR spectra. The concentrated double labelled CexE-6His sample was filtered using a size exclusion column. The filtrate was collected in 4 ml fractions. The concentration of protein in the filtrate was measured using absorbance at 280 nm and plotted against the volume of the filtrate (Fig. 5.6). A single large peak was present corresponding to the double labelled CexE-6His protein. There were smaller amounts of larger proteins present but their peaks were not comparable to that of double labelled CexE-6His. The peak for double labelled CexE-6His was not defined. This was likely due to a high concentration of double labelled CexE-6His that prevented accurate absorbance readings. The fractions of the filtrate containing CexE-6His, which was 204 ml to

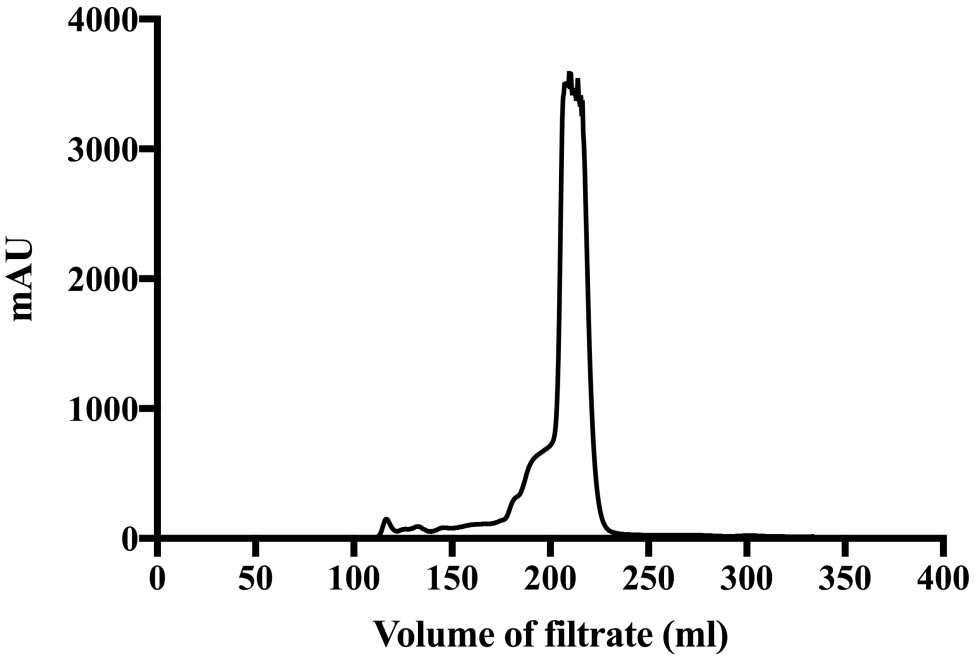


Figure 5.6: Size exclusion of the double ^{13}C and ^{15}N labelled CexE-6His purified protein. The CexE-6His protein was buffer exchanged by dialysis into 50 mM NaP (27:73 NaH_2PO_4 to Na_2HPO_4), 250 mM NaCl, 0.5 mM TCEP-HCl, pH 7.3. The CexE-6His protein was loaded onto a size exclusion column and the absorbance of the filtrate was measured at a wavelength of 280 nm.

232 ml elution volume, were combined and concentrated by membrane filtration. Thus, it was concluded that double labelled CexE-6His had been purified sufficiently for NMR spectroscopy.

5.2.4. Assignment of the chemical shifts to the backbone of CexE-6His

With NMR spectroscopy, information about the chemical environment of each nucleus in CexE-6His with an unpaired spin can be obtained. Assigning the chemical shifts to the backbone moieties is critical for the determination of protein structure. The salt concentration and pH in the CexE-6His buffer were reduced by dialysis in preparation for NMR spectroscopy. The final concentration of double labelled CexE-6His for NMR structural analysis was 1.83 mM. The ^{1}H-^{15}N HSQC, HNCA, HNCOCA, HNCACB, HNCOCACB, HNCACO, and HNCO spectra were collected by Dr. Timothy Knowles on the Bruker 600 Avance-IVDr at the Biomolecular facility at the University of Birmingham. The peaks for each spectrum were picked computationally with NMRFAM Sparky. There were 152, 213, 145, 764, 289, 429, and 268 peaks for the ^{1}H-^{15}N-HSQC, HNCA, HNCOCA, HNCACB, HNCACBCO, HNCACO, and HNCO spectra, respectively. The spectra were used to assign the chemical shifts of the amine, Cα, Cβ, and carbonyl moieties to the backbone of CexE-6His, in NMRFAM Sparky. The chemical shift of a random peak in the ^{1}H-^{15}N-HSQC spectrum was used to provide the ^{1}H and ^{15}N chemical shifts used to specify the plane containing the corresponding Cα peak in the HNCA and HNCOCA spectra. The peak in the HNCA spectrum that was not present in the HNCOCA spectrum was assigned as the i position Cα. Using the HNCACB and HNCOCACB spectra, the same process was used to assign the i Cβ. A peak for the chemical shift of the i Cα was also present in HNCACB was only assigned when the chemical shift of ^{1}H, ^{15}N and ^{13}C agreed with the HNCA spectrum. The ^{13}C chemical shift of the i Cα in the HNCA was used to search the spectrum to identify the chemical shift when that

Cα was in the i−1 position (Fig. 5.7, blue lines). The ^{1}H and ^{15}N chemical shift of this peak was used to identify the corresponding peak in the ^{1}H-^{15}N-HSQC (Fig. 5.8). The carbonyl groups were assigned in the same way with HNCACO and HNCO spectra. Once 3 to 4 residues were fully assigned, NMRFAM Sparky was used to find the string of residues in the backbone of CexE-6His that best fit the chemical shifts observed. This was based on a database of already known chemical shifts for amino acids. These steps were repeated until the backbone of the protein was assigned. Excluding moieties prior to a proline residue, all Cα, Cβ, amine, and carbonyl groups from the Cα of G21 to the carbonyl of H123 were assigned. The Cα, Cβ, amine, and carbonyl moieties of G20 and the other histidines in the 6His tag could not be assigned. Therefore, these residues likely were likely unstructured.

5.2.5. Assignment of the chemical shifts of the carbon atoms in the amino acid side chains of CexE-6His

From the chemical shifts assigned to the backbone of CexE-6His above, the chemical shifts of the Cα and Cβ were known. The CCONH spectrum was collected as above by Dr. Timothy Knowles. The peaks in the CCONH spectrum were used to assign the chemical shifts of the amino acid side chain carbons (Fig. 5.9). There were 287 peaks in the CCONH spectrum. Peaks with the same chemical shifts as the Cα and Cβ from the backbone assignment were assigned as the respective Cα and Cβ of that residue. Any other peaks, on the same ^{15}N plane with the same ^{1}H chemical shift, were the chemical shifts of the other ^{13}C nuclei in the amino acid side chain. These chemical shifts were assigned based on the comparison of their ^{13}C chemical shift to the average chemical shift of that nuclei in the Biological Magnetic Resonance Data Bank (Ulrich et al., 2008). If there was ambiguity regarding the chemical shift ^{13}C nucleus, assignment was completed in conjunction with the aliphatic ^{13}C-NOESY-HSQC spectrum. The

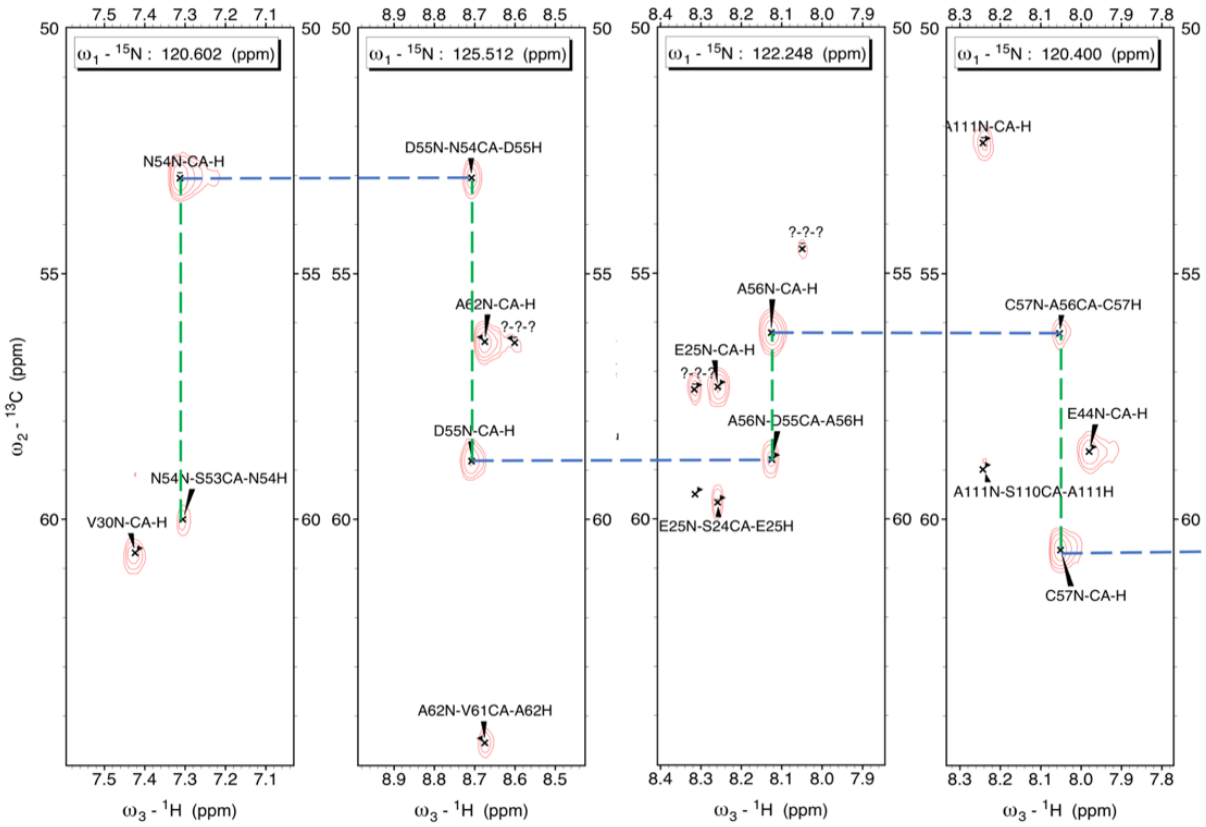


Figure 5.7: Backbone assignment of CexE-6His. HNCA, progressive backbone assignments of the Cα, the blue lines are match ^{13}C chemical shifts and the green lines are matching ^{1}H chemical shifts for the residues N54 to C57. The HNCA spectra provides the chemical shifts of the atoms associated with the backbone of the protein. The HNCA spectra is viewed with the nitrogen chemical shift on the z-axis, the hydrogen chemical shift on the x-axis, and the carbon chemical shift on the y-axis. Each carbon atom in the protein backbone has two peaks in the HNCA spectra; as the Cα has a peak associated with it in the C_i and C_{i-1} position, which change based on the associated amine group. The chemical shift of the Cα is exact same in every spectra. The HNCOCA spectra contains only the chemical shifts associated with the C_{i-1} carbons. Using the HNCA and HNCOCA spectra, a chain of the chemical shifts of the Cα carbons and amine groups in the protein backbone can be assigned.

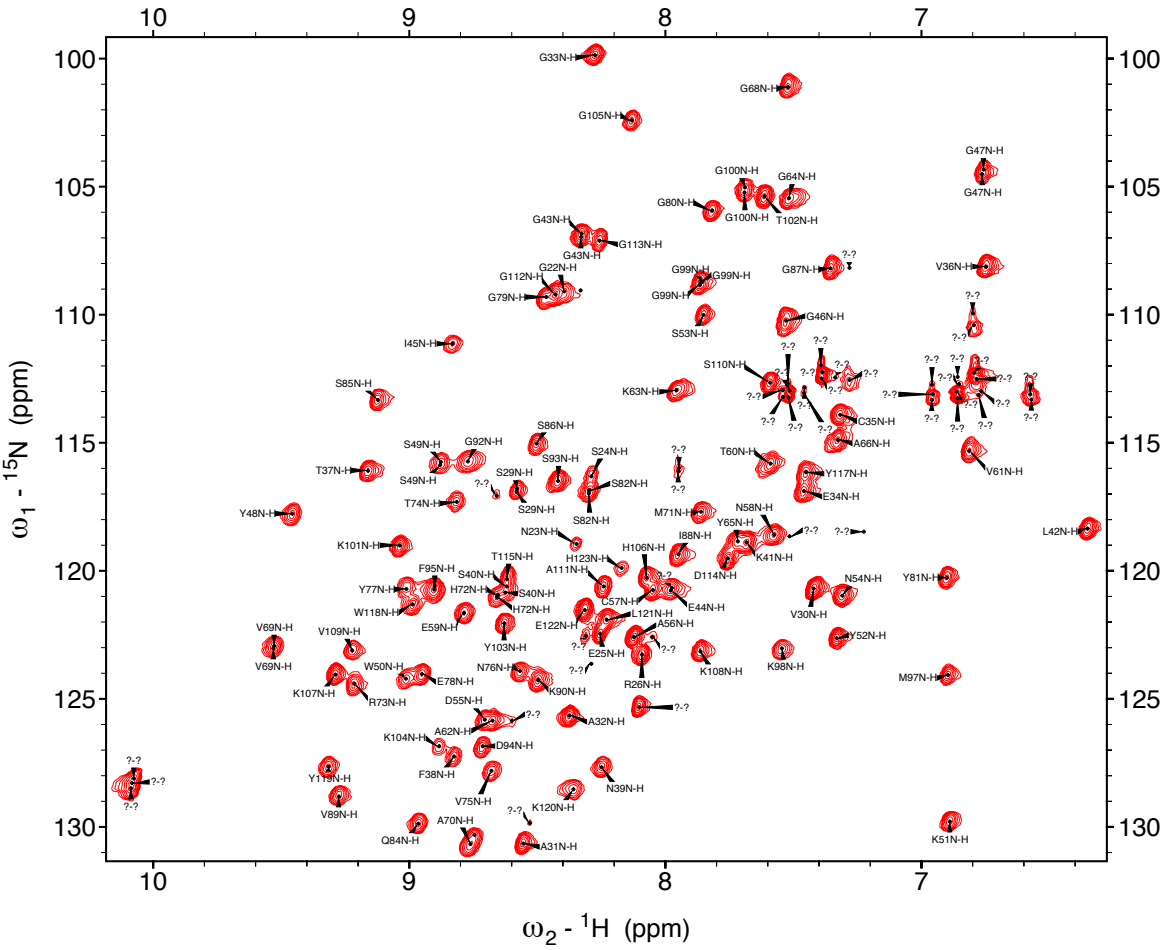


Figure 5.8: ^{1}H-^{15}N-HSQC of backbone assigned CexE-6His. The ^{1}H and ^{15}N chemical shifts for the amino acids in the ^{1}H-^{15}N-HSQC are the same as those in the HNCA (Fig. 5.7). The nitrogen chemical shift is on the y-axis with hydrogen on the x-axis, there is no z-axis in the ^{1}H-^{15}N-HSQC spectra. Each peak is an amine group. These can be in the protein backbone or side chains. The amine group of each amino acid in the protein backbone was assigned in conjunction with the HNCA and HNCOCA spectra. A peak is picked and assigned an amine group at random from the backbone of the protein. The chemical shifts that associate with this peak in the HNCA and HNCOCA spectra are assigned. After a chain of amino acid chemical shifts are assigned to peaks in the spectra NMRFAM-Sparky is used to search which chain of amino acids these chemical shifts are most likely to be based on a database of previously observed chemical shifts for amino acids. In this way the backbone of the protein can be assigned.

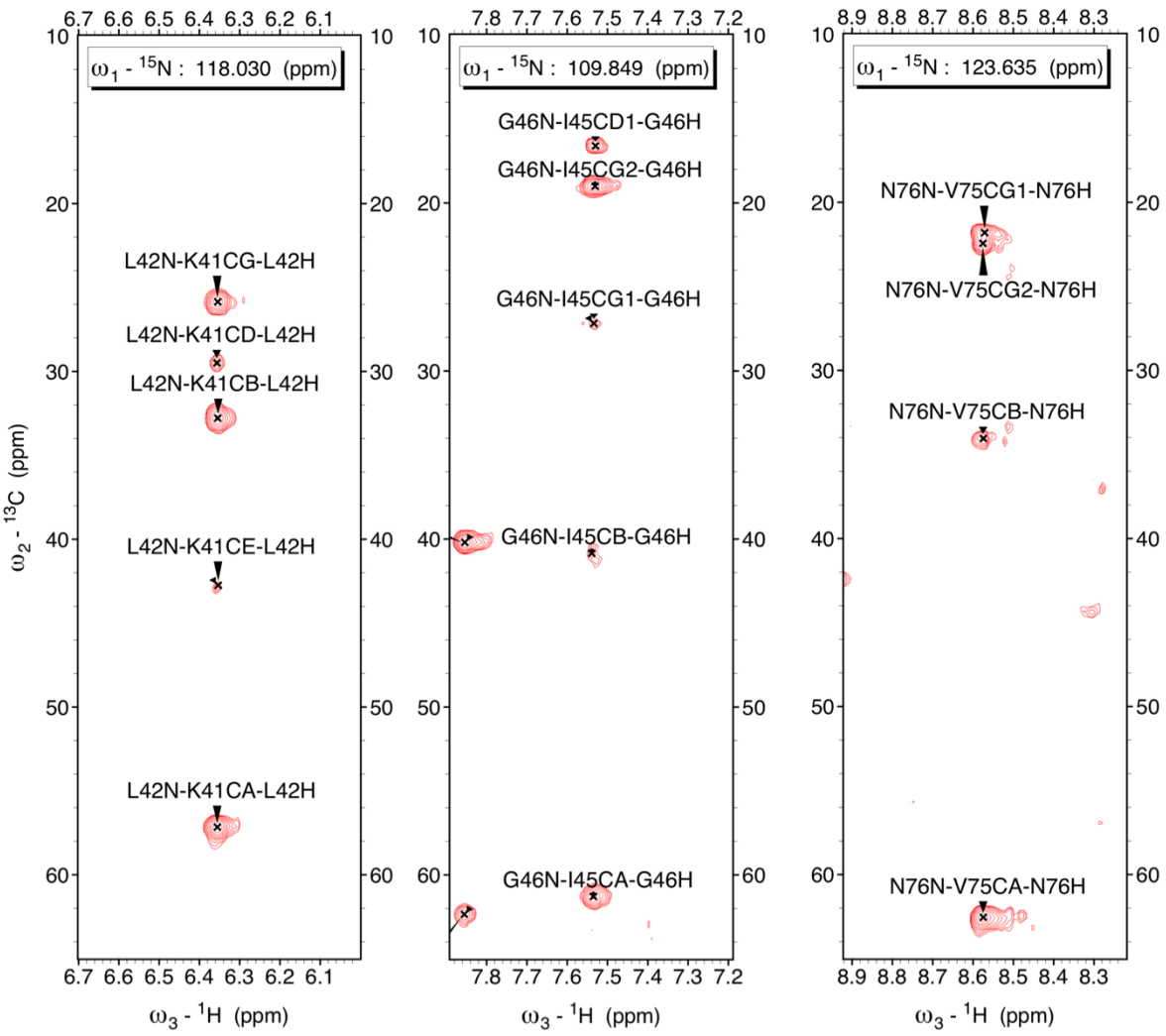


Figure 5.9: Examples of residue side chains in the CCONH spectrum. The example residues depicted are K41, I45, and V75. The chemical shifts of the Cα and Cβ are known for the assignment of the HNCA and HNCACB spectra. The Cδ, Cγ, and Cε if present were assigned using the average chemical shifts of the carbons in that residue form the Biological Magnetic Resonance Data Bank (Ulrich et al., 2008). The assignment of ^{13}C with very similar chemical shifts was completed in conjunction with the aliphatic ^{13}C-NOESY-HSQC spectrum.

vast majority of side chain carbon nuclei were assigned to a chemical shift in the CCONH spectrum.

5.2.6. Chemical shift assignment of spatially proximal residues in CexE-6His

A ^{13}C-NOESY-HSQC makes use of the nuclear Overhauser effect (NOE) to identify residues in a protein that are within 5 Å of each other due to their bonding or the tertiary structure of the protein. A ^{13}C-NOESY-HSQC spectrum can collect either the chemical shifts of aliphatic or aromatic moieties. The aliphatic ^{13}C-NOESY-HSQC spectrum was collected by Dr. Timothy Knowles on the Bruker 900 MHz NMR spectrometer at the Biomolecular facility at the University of Birmingham. Peaks of chemical shifts were picked manually. The peaks in the ^{13}C-NOESY-HSQC spectrum were assigned from the chemical shifts determined above and using the average chemical shift of each atom in the Biological Magnetic Resonance Data Bank repository (Ulrich et al., 2008). Once a chemical shift had been assigned to a nucleus of CexE-6His, all of the peaks with the same ^{1}H chemical shift were picked manually. The peaks that were not assigned were the chemical shifts of the ^{1}H that were located within 5 Å of the residue, these are known as NOEs. There were 3859 peaks in the ^{13}C-NOESY-HSQC spectrum. Chemical shifts that did not agree with the expected standard deviation of average chemical shifts for that atom in that amino acid were subjected to further interrogation until the issue was resolved. The 1616 peaks of the chemical shifts in the ^{15}N-NOESY-HSQC spectrum were picked computationally and not assigned. The peaks of the assigned ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC were exported with all of the shifts in the ^{13}C-NOESY-HSQC for use in structure calculation.

5.2.7. Structure calculation of CexE-6His

Dr. Timothy Knowles completed the final adjustments of aliphatic ^{13}C-NOESY-HSQC and

^{15}N-NOESY-HSQC spectra and then calculated the structure of CexE-6His using CYANA. CYANA used known torsion angles, the assigned chemical shifts provided, and the peaks of the aliphatic ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC spectra to calculate the best 20 structure models. The top 20 models for the CexE-6His structure were calculated (Fig. 5.10). The root-mean-square deviation of atomic particle for the backbone of the CexE-6His of the best 20 models was 0.44±0.06 Å and 1.06±0.08 Å for heavy atoms, which accounts for all atoms except hydrogen. The surface charge for the structure of CexE was calculated using the PDB2PQR and APBS functions in Chimera (Fig. 5.11). There is a clear cleft of positively charged residues that may be responsible for the DNA binding phenotype observed. However, an anomalous fold in the CexE protein, between N39 and I45, could be responsible for this cleft. The reason for this fold was investigated. It was found that positions of the aromatic amino acids in the CexE-6His structure were not correctly calculated by CYANA as the NOEs of these residues were not present in the experimental data. According to the model that the aromatic moieties of CexE-6His were exposed and not in the hydrophobic core of the protein (Fig. 5.12). As this is unlikely to be correct a further aromatic ^{13}C-NOESY-HSQC spectrum is required. Although the position of the aromatic residues in the structure was not correct, the structure of the other residues is not thought to substantially change with addition of the aromatic ^{13}C-NOESY-HSQC spectrum. At the N-terminal of CexE-6His there is a helical turn from A32 to C35. This cysteine forms a disulphide bond with C57. There is an α helix from N55 to G64. There were four β strands: G68 to Y77; S82 to V89; K108 to D114; and Y117 to Y119. The final β strand is composed entirely of aromatic amino acids. With the addition of the aromatic ^{13}C-NOESY-HSQC there will likely only be three β strands. These β strands are present on the other side of the protein from the N-terminus, where they form an anti-parallel β sheet. It was concluded, although subject to change with the addition of the aromatic 13C-NOESY-HSQC, that there were at least three β strands and one α helix in the structure of CexE.

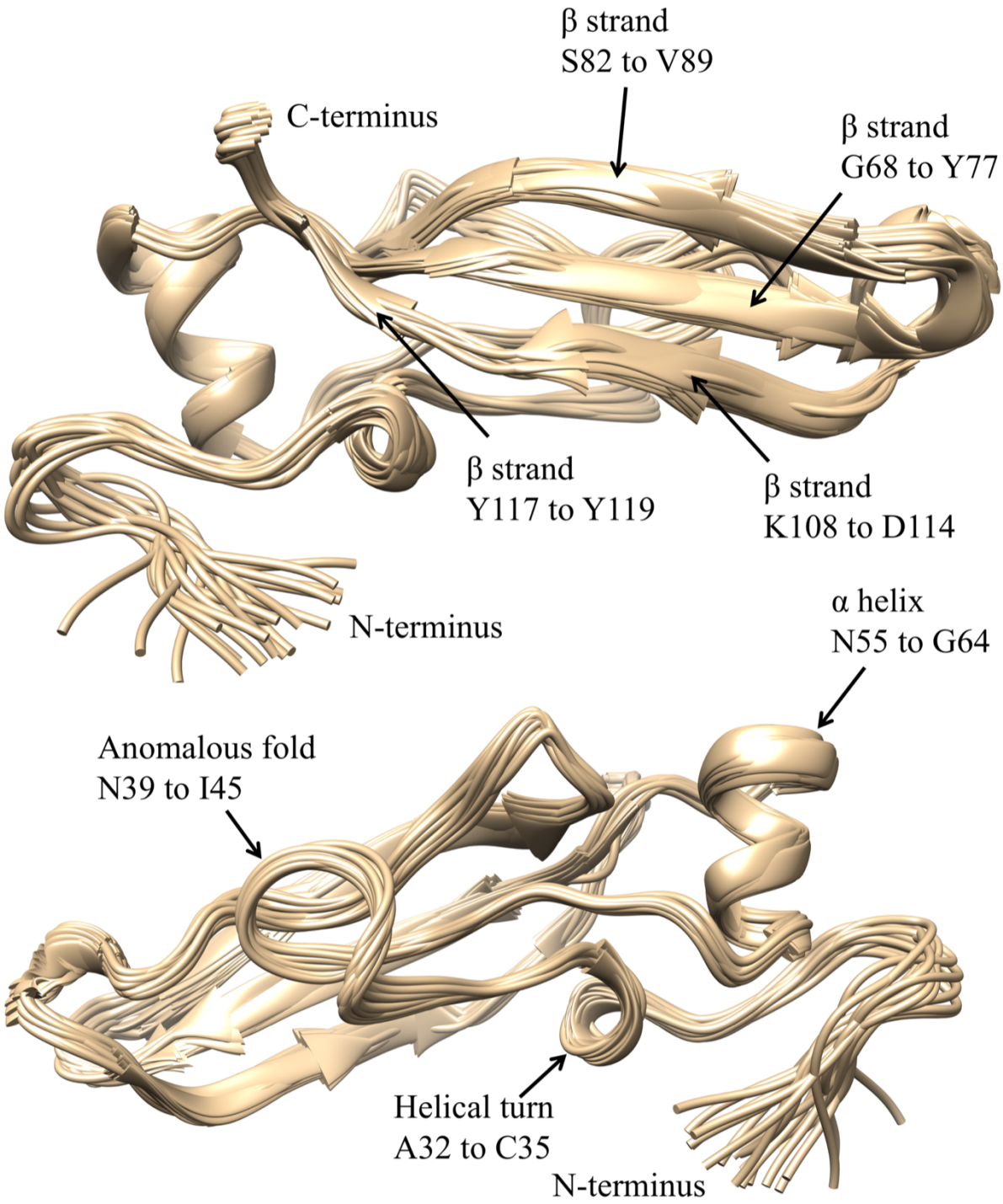


Figure 5.10: Ensemble of 20 models of CexE-6His as determined by CYANA. The unstructured 6His tag was removed from the structure calculation. Regions referred to in the text are labelled on the figure. The figure shows the solution structure of CexE and a 180º turn of the same structure. Amino acids are numbered from the N-terminus of the full length CexE protein before Sec signal cleavage. Images were collected in Chimera.

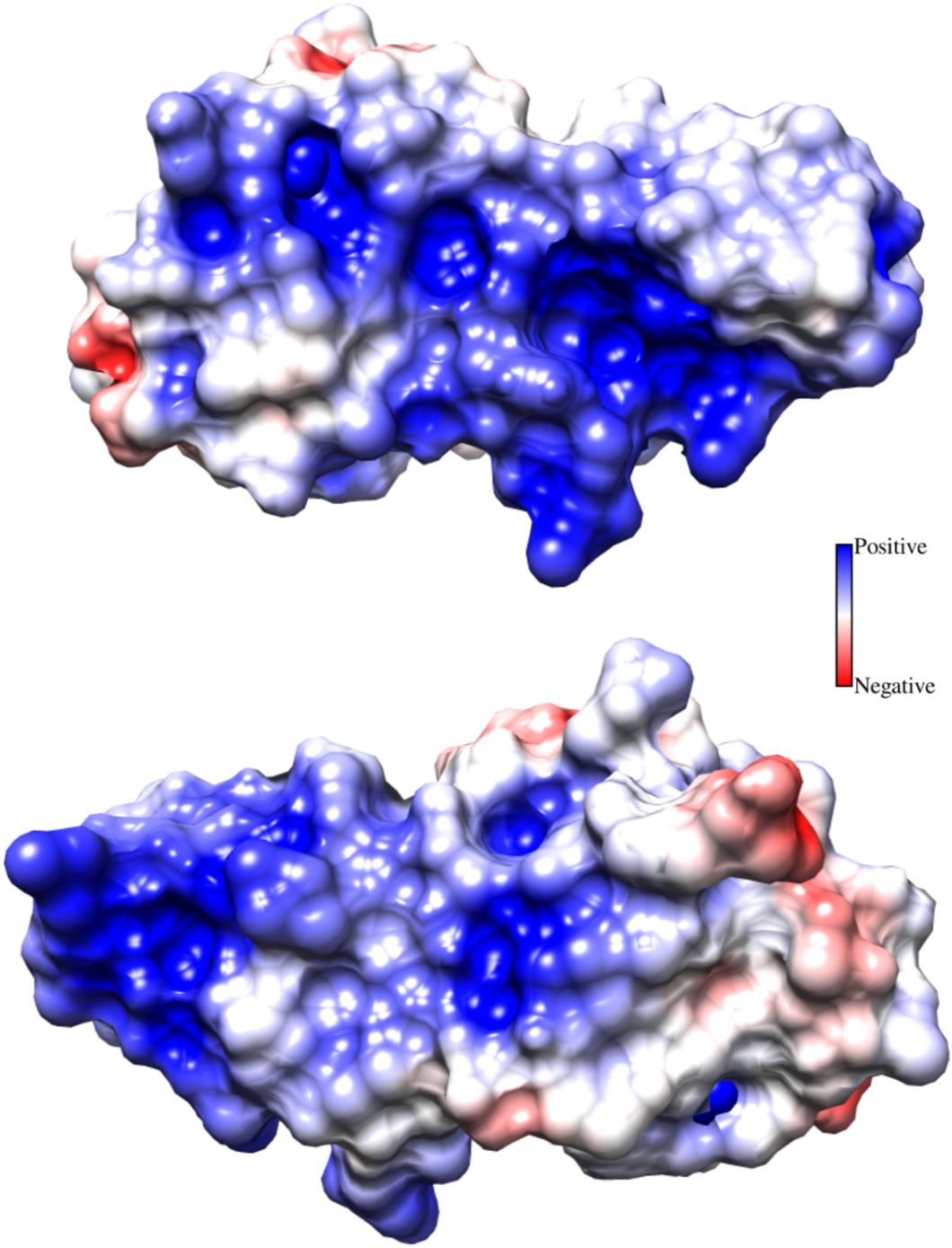


Figure 5.11: Electrostatic surface charge of CexE. The electrostatic surface charge of CexE was calculated using the PDB2PQR and APBS functions in Chimera. The figure shows a 180º turn of the surface of the CexE protein with electrostatic surface charge. There is a clear concentration of positive charge, however it is not clear if this is due to the analysis fold present in the CexE solution structure.

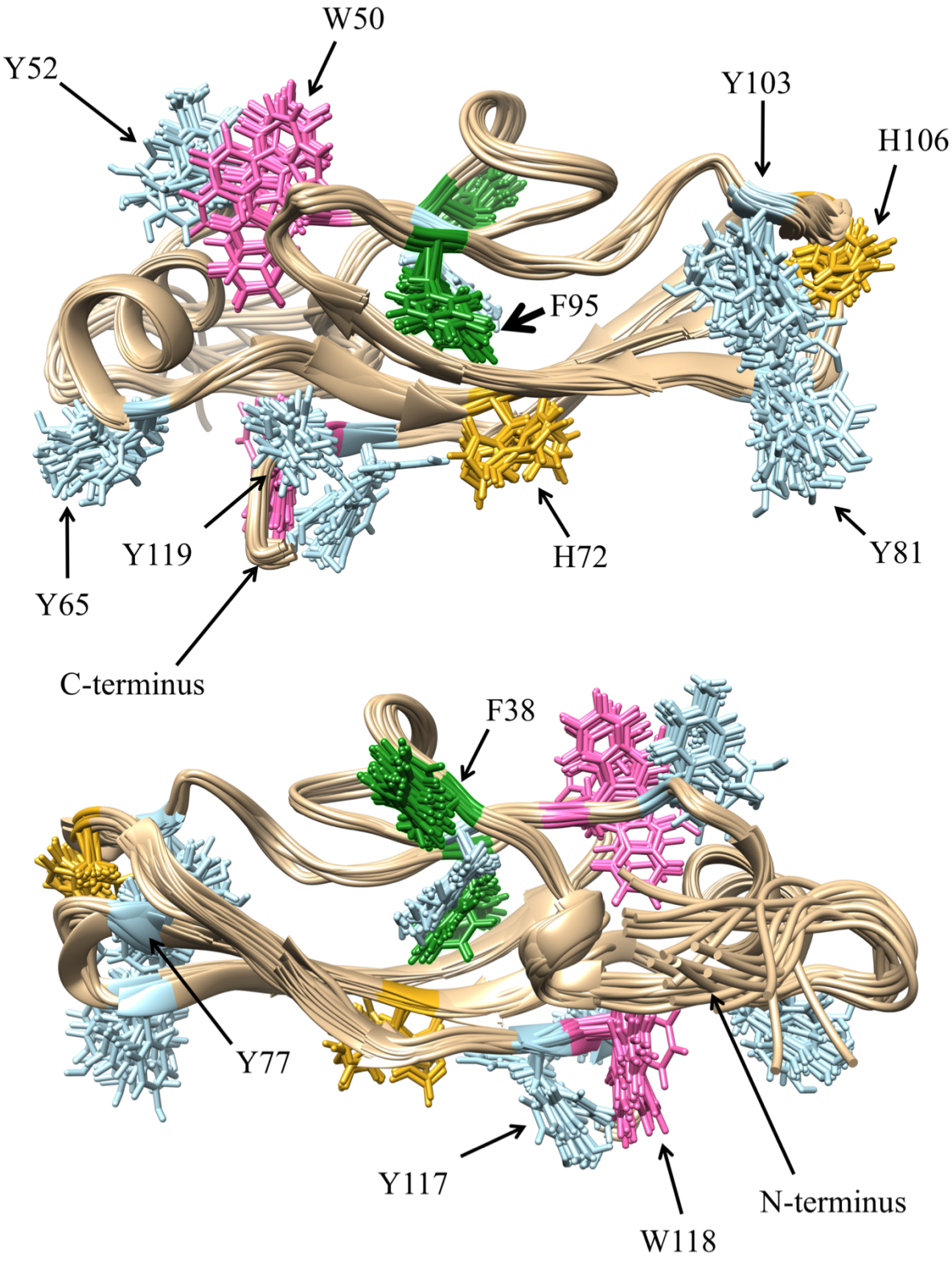


Figure 5.12: The aromatic residues exposed in the CexE structure. The bonds of the aromatic residues in CexE were represented as lines on the structure. They were coloured: pink for tryptophan; blue for tyrosine; yellow for histidine; and green for phenylalanine.

5.2.8. Comparison of preliminary CexE structure to the structure of Aap

Alignment of two similar structures can reveal information about which region of a protein are responsible for function. The best representative conformer structure of Aap and CexE were aligned in Chimera by their amino acid sequences using Matchmaker. The structures of Aap and CexE were very similar (Fig. 5.13). The α helices, on the right in Fig. 5.13, of Aap and CexE were superimposable. The anti-parallel β sheets of CexE were superimposable onto the β sheets in Aap. The other β sheets in Aap were not present in CexE. In addition, the other α helix at the pole of Aap, on the left in Fig. 5.13, does not appear to be present in the CexE model. It is thought that the addition of the aromatic ^{13}C-NOESY-HSQC spectrum will resolve anomalous folding calculated for the CexE protein. This misfolded region maybe the difference between CexE and Aap. The structure of CexE is bullet shaped. As described for Aap by Velarde et al. (2007). In conclusion despite the differences in amino acid composition the structure of CexE was almost identical to that of Aap.

5.2.9. Position of conserved residues on the Aap structure

Since CexE and Aap do not share function and there is comparatively little amino acid sequence identity for two structures that were superimposable, the residues conserved must have a structural role. The position of the conserved residues of Aap and CexE on the Aap structure, which was determined by Velarde et al. (2007), were identified. The predicted Sec signal sequence was removed from 70 non-identical Aap and CexE homologous sequences described in Chapter 4.2.1. by SignalP. The resultant sequences were aligned using T-coffee. Each residue in the best representative conformer structure of Aap was coloured according to its percentage of conservation in the Aap and CexE alignment. The amino acids were coloured in Chimera on a scale of red to white to blue, where red was 100%, white was 58%, blue was 15% conservation, respectively. The amino acid side chain structure was included if they were

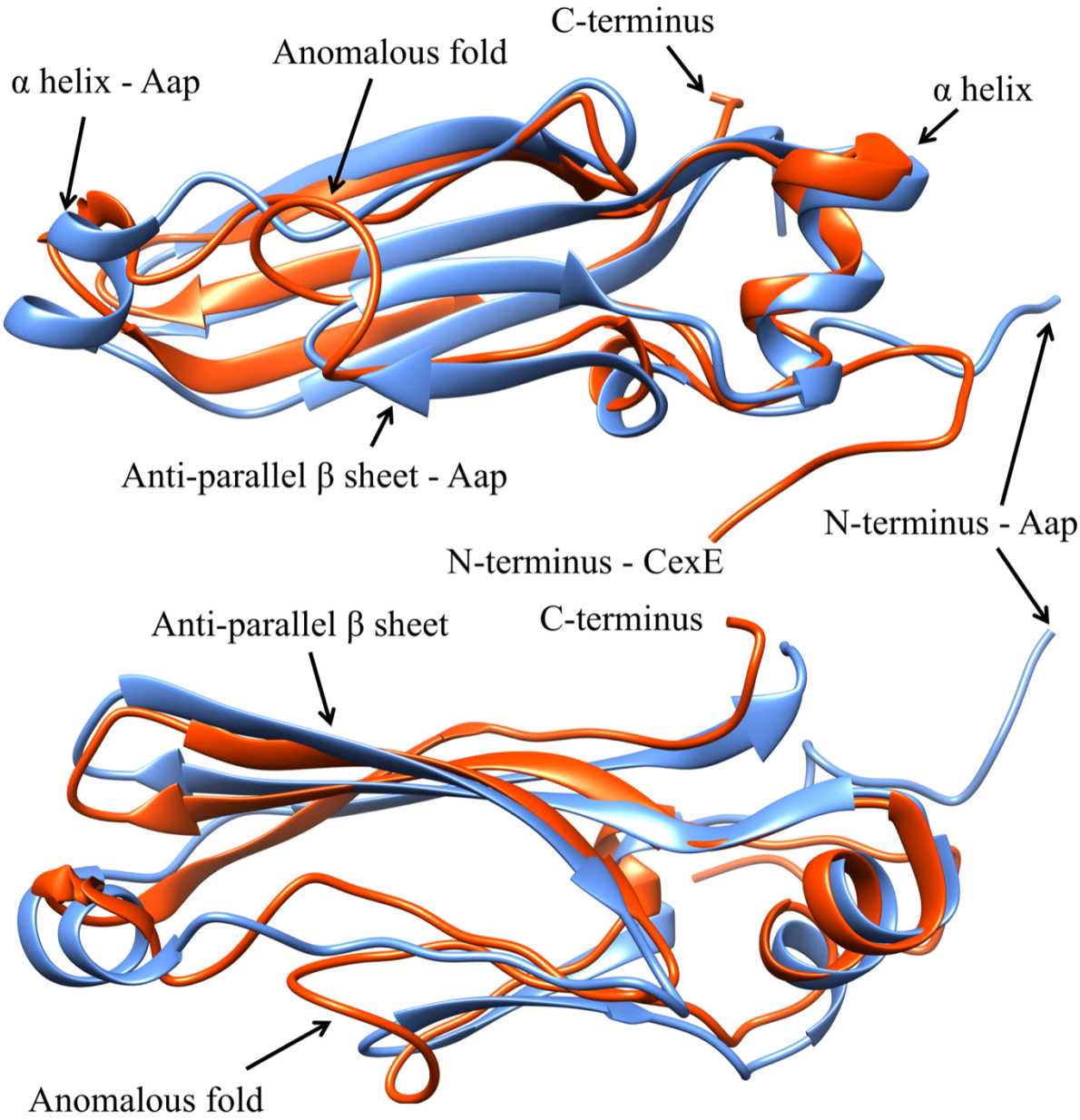


Figure 5.13: Comparison of Aap and CexE structures. The best representative conformer of Aap solved by Velarde et al. (2007) and a representative structure model of CexE were aligned by amino acid sequence in Chimera. The best representative conformer is a single structure that best represents the group of 20 Aap structures calculated. The structures were coloured blue and orange for Aap and CexE, respectively.

more than 80% conserved. The surface of Aap and CexE was not conserved (Appendix II). Instead the conserved residues were present in the hydrophobic core of Aap (Fig. 5.14). The hydrophobic core of Aap was previously described by Velarde et al. (2007). However, they did not identify the conservation of these residues. The disulphide bridge is conserved in the majority of Aap/CexE proteins (Fig. 5.14; C36 and C53 and B, Fig. 5.15; residues 17 and 41). Two conserved prolines, P33 and P87, cause turns in the structure. A hydrophobic region at this pole is very conserved, which consists of W26, A28, V65, V85, and W114. The amino acids at the other pole of Aap are not as well conserved. Only Y73 was present in over 80% of the sequences, likely forming another hydrophobic region. Interestingly the unstructured N-terminal residues of Aap were highly conserved. The initial amino acid was invariably glycine, the next was either a serine or a glycine. The first 6 amino acids of the N-terminus after the predicted signal peptidase cleavage site were glycine rich and to a lesser extent serine (Fig. 5.15). The major similarity between Aap and CexE is there suspected mechanism of secretion, which has not been confirmed for CexE. Therefore, it was concluded that the conserved residues between Aap and CexE are required for the specific structure of Aap- and CexE like proteins.

5.3. Discussion

The NMR structure of the CexE protein requires further refinement based upon additional experimental data. The inclusion of an aromatic ^{13}C-NOESY-HSQC spectrum would define the CexE structure. Dr. Timothy Knowles has collected the aromatic ^{13}C-NOESY-HSQC spectrum for CexE-6His and is currently assigning the peaks.

The preliminary structure of CexE presented here is very similar to that of Aap, which is notable due to their lack of sequence identity. It is not typical for two proteins with 18%

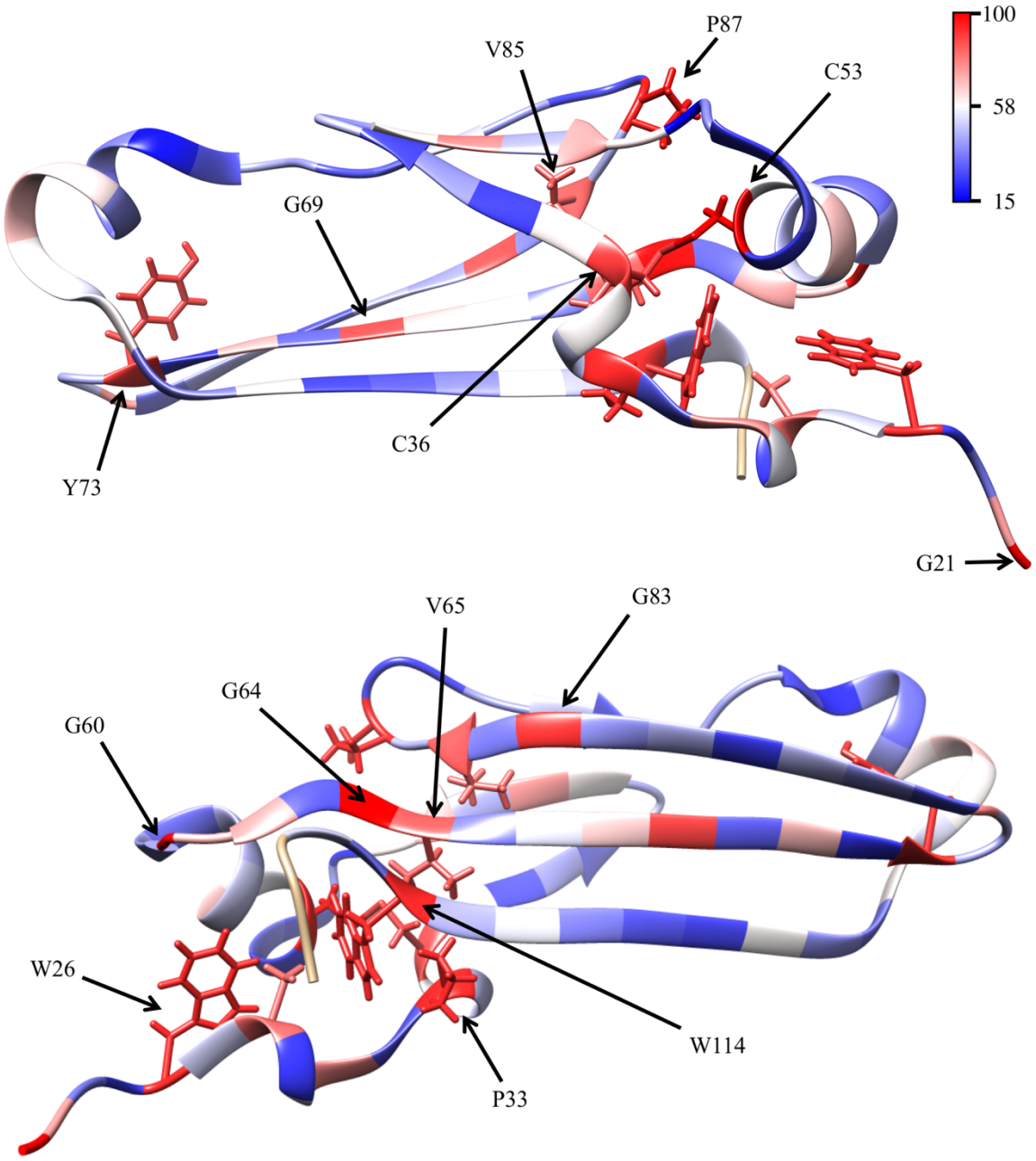


Figure 5.14: Structure of Aap coloured by percentage conservation. The signal sequence of non-identical sequences from Chapter 4 were removed by SignalP. The resulting sequences were aligned using T-coffee. The best representative conformer of the 20 Aap models specified in the file, 2jvu (Velarde et al., 2007), was coloured by percentage conservation according to the alignment. The scale for which is located in the top right of the figure. Residues coloured red are highly conserved while blue residues are not. The atoms of the residues with over 80% conservation were represented by lines and labelled with the amino acid.

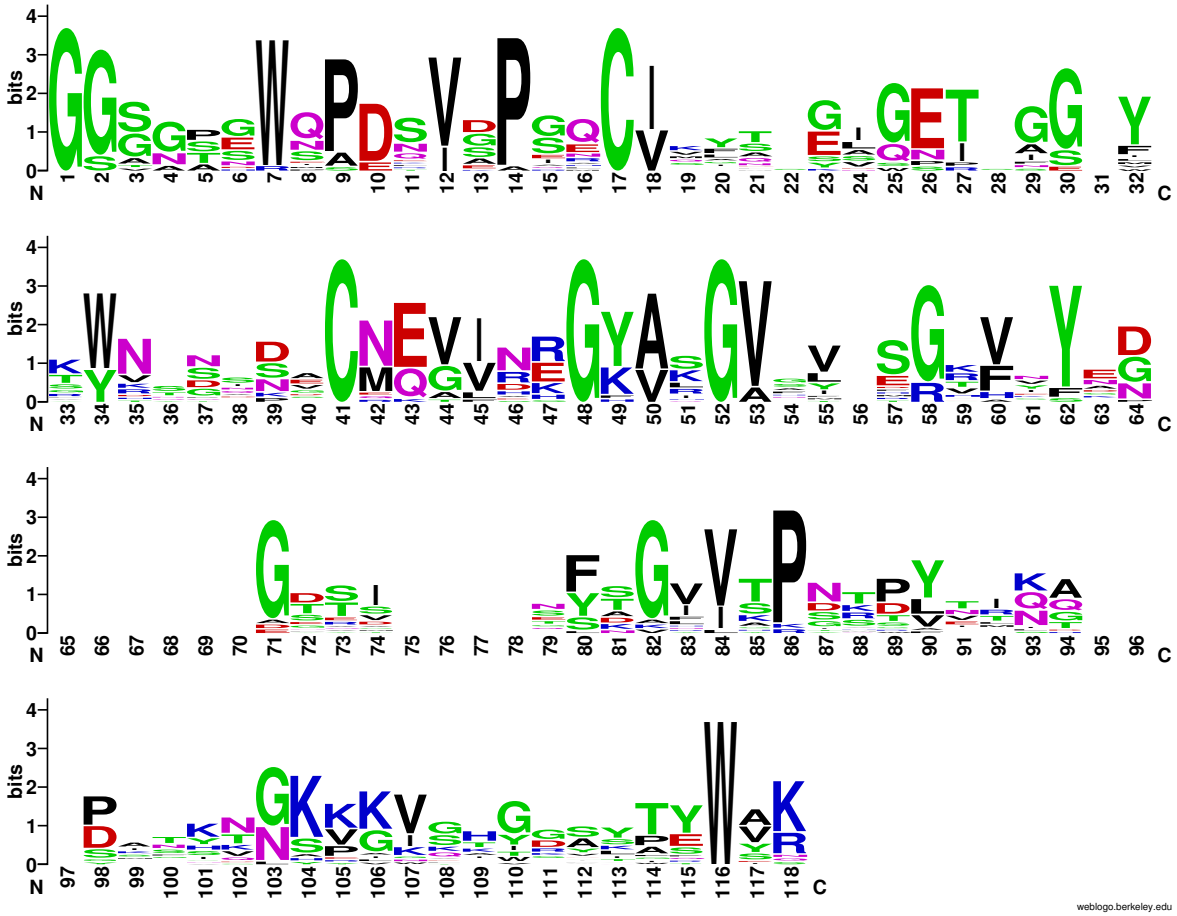


Figure 5.15: WebLogo of the residues in the Aap and CexE aligned sequences. The height of the letter indicates the level of conservation in the alignment. The numbering of the residues is not that of Aap instead each number is the position in the alignment of Aap and CexE sequences.

sequence identity to have superimposable structures (Rost, 1999). These structures are relatively unique. Therefore, it is very likely, as first hypothesized by Velarde et al., (2007), that the structure of Aap and now CexE is a requirement for secretion by the Aat system. Interestingly there were highly conserved residues that were not part of the structure. The N-terminus of the Aap/CexE proteins after the signal sequence cleavage is highly glycine rich. It is unlikely that the unstructured residues at the N-terminus have a role in Aap or CexE function. The rest of the conserved residues between Aap and CexE are required for a structure that can be secreted by the Aat system. Therefore, it is suggested that the conservation of the N-terminal glycine rich region is also required for Aat secretion.

The secretion of Aap is by the Aat system (Nishi et al., 2003). There is no information on the mechanism of CexE secretion. In the current work, it was confirmed that CexE is secreted into the periplasm by the general secretory pathway but the next steps in CexE secretion are unknown. It is postulated that due to the specific structure of CexE presented here and the position of *cexE* in the ETEC H10407 genome, that CexE is secreted by the Aat system.

CHAPTER 6:

Secretion of CexE and Aap in EAEC 042 and ETEC H10407.

6.1. Introduction

In the RNA-Seq experiments from chapter 3, like *aap*, transcription of the *aat* system was significantly lower in the *aggR* mutant compared to the wild-type. The Aat system is required for the secretion of Aap (Nishi et al., 2003). The Aat system is comprised of 5 proteins: AatP; AatA; AatB; AatC; and AatD. AatP and AatC are the permease and ATPase domain of an ABC transporter. AatA is a homolog of TolC (Nishi et al., 2003). A function for AatB was not suggested by Nishi et al. (2003). However, recent annotations of AatB have identified it as a membrane fusion protein of the type associated with T1SS (see Chapter 1). The function of AatD is unknown. Based on the homology of these proteins, Nishi et al. (2003) proposed that Aap requires an atypical T1SS for extracellular localisation (Fig. 6.1). Like Aap, CexE possesses a Sec-dependent signal sequence for translocation to the periplasm. Recent investigations have demonstrated that CexE can be found in the extracellular milieu of ETEC H10407 (Roy et al., 2011). However, a secretion mechanism for CexE is unknown.

Therefore, the aim of the work presented in this Chapter was to investigate the association between CexE and Aat-mediated protein secretion. Bioinformatic approaches were used to examine the prevalence and organisation of the *aat* gene cluster and its association with genes encoding Aap/CexE-like molecules. A role for Aat in CexE secretion was demonstrated.

6.2. Results

6.2.1. Prevalence of the *aat* genes

The Aat system is proposed to function together, if all proteins are required then they will be encoded in a single strain. A PSI-BLAST search identified 717, 725, 663, 470, and 689 sequences homologous to AatP, AatA, AatB, AatC, and AatD, respectively. The vast majority (81%) of *aat* loci were found in *E. coli*, predominantly ETEC and EAEC. Genes encoding *aat*

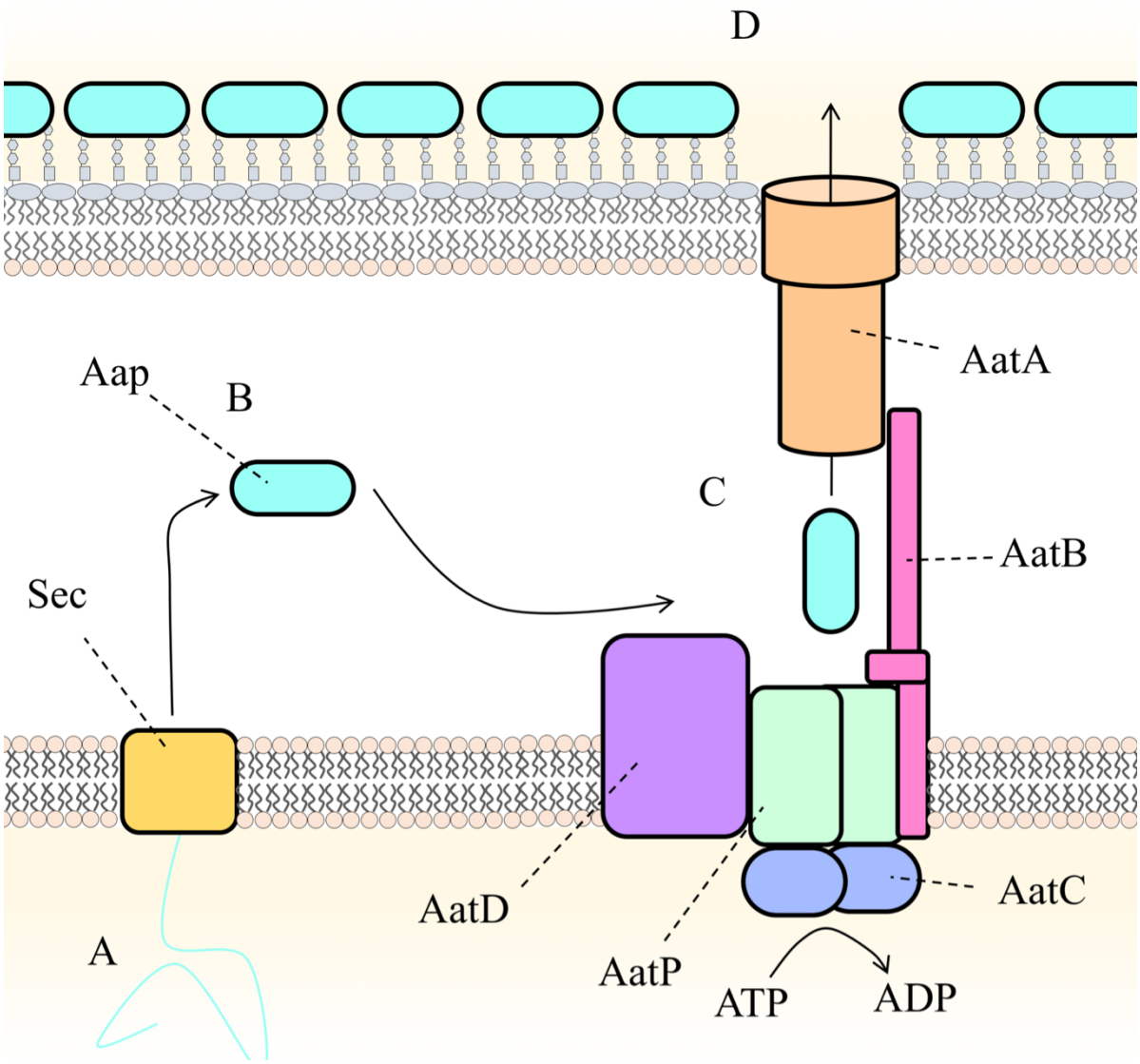


Figure 6.1: Mechanism of the Aat system proposed by Nishi et al. (2003). This figure was previously presented as Fig. 1.15, it is reproduced here. A, Aap is produced in the cytoplasm. B, Aap is secreted into the periplasm via the general secretory pathway. The signal sequence is cleaved by signal peptidase I. C, an as yet unknown function of AatD is required for the secretion of Aap. Aap access the Aat pore via the periplasm. D, Aap is secreted out of the cell and electrostatically attaches to the LPS.

loci were also found in *S. enterica*, *Providencia alcalifaciens*, *Y. enterocolitica*, and *C. rodentium*. The genomic regions encoding these proteins were inspected. It was noted that *aatP* and *aatA* are well annotated. In contrast, *aatB* and/or *aatC* were not annotated or annotated incorrectly. Furthermore, the majority (96%) of the *aat* genes identified in this study were associated with strains that do not have a complete genome. Therefore, there remains the possibility that intact *aat* loci exist in these strains but nucleotide sequence data for specific *aat* genes was not obtained during sequencing. This explains the lower number of sequences homologous to AatB and AatC that were identified by PSI-BLAST.

The functional association of Aap with the Aat system suggested the *aat* loci identified above were atypical T1SS. Therefore, the nucleotide sequences of these strains were examined for the co-occurrence of *aap* and *aat* loci. Only 290 of the strains contained a gene encoding a Aap homolog; this includes ETEC H10407 which possessed CexE and AatPABCD. However, manual inspection revealed that the *aat* loci frequently contained an unannotated *aap*/*cexE* gene, and the incomplete nature of the majority of the genomes meant that it could not be certain *aap* homologs were absent from these strains. Therefore, further analysis was limited to strains with complete genomes; 28 strains with complete genomes were found to contain homologs of the *aat* system. The presence or absence of each of the *aat* and *aap* homologs was visualised using a heatmap (Fig. 6.2). All of the genomes contained a complete *aat* system and an *aap* or *cexE* gene, except *E. coli* MRE600. An ORF that contained an *aap* or *cexE* homologous sequence could not be identified on any of the three plasmids or the chromosome in MRE600. However, it remains possible that a divergent *aap* is present in the genome but was undetected using the bioinformatic approaches employed in this investigation. There were two duplication events in the ETEC stains E24377A and FORC_31. Therefore, it is hypothesised that genes encoding *aap* or *cexE* homologs co-occur with the *aat* encoded T1SS.

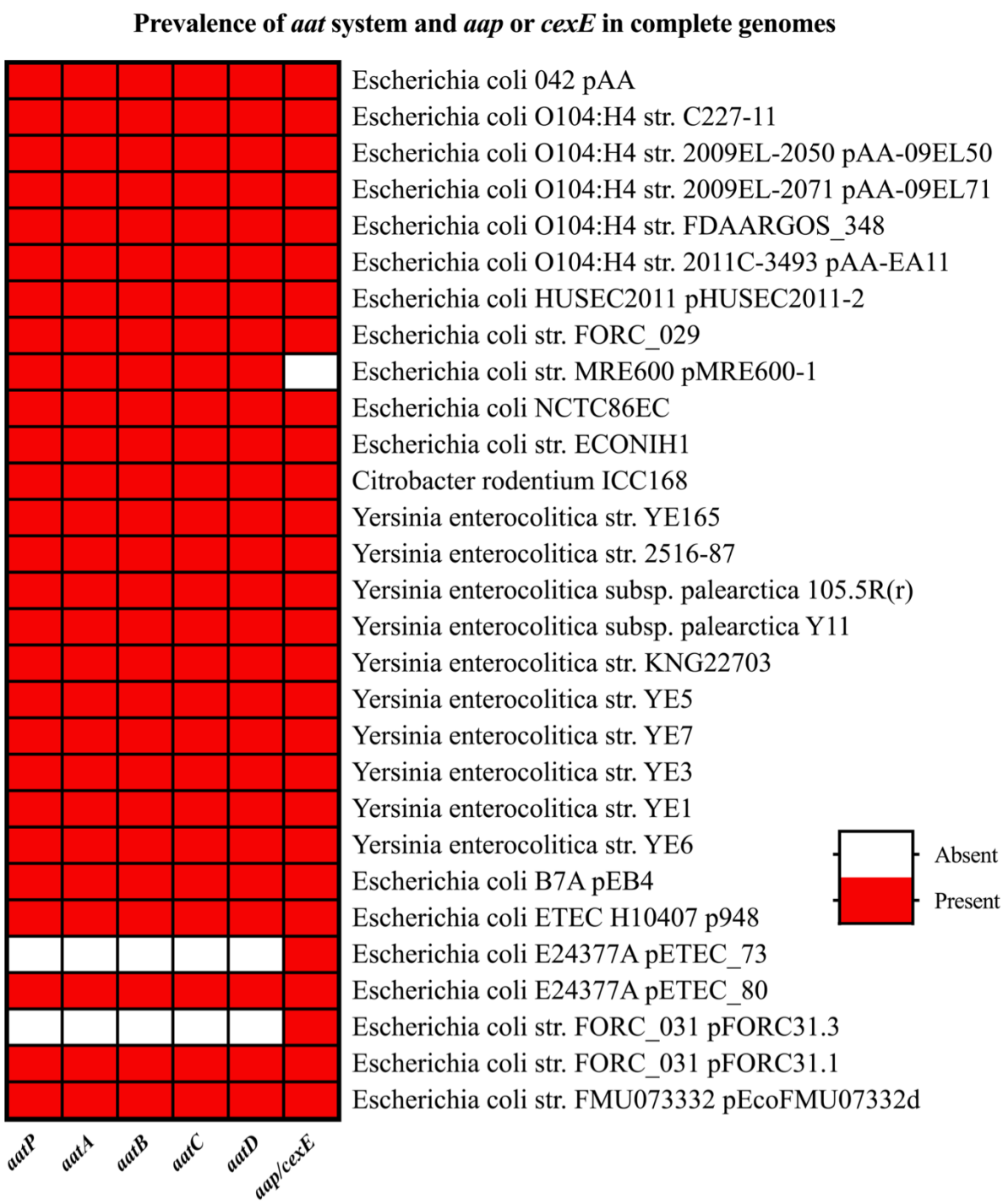


Figure 6.2: Prevalence of the *aat* system and *aap*/*cexE*. A red block indicates the gene indicated was present in the strain sequence. A white block indicates that a homologous gene could not be detected. Only strains with complete genomes were included in this analysis. Homology was detected using BLASTp.

6.2.2. Organisation of the *aat* cluster and *aap*/*cexE* in strains from the analysis of the prevalence of the Aat proteins

The operon structures of the *aat* cluster differ between EAEC 042 and ETEC H10407. Thus, the organisation of genes *aatP*, *aatA*, *aatB*, *aatC*, *aatD*, and *aap* or *cexE* in relation to each other was investigated (Fig. 6.3). The GC content of the regions encoding these genes was around 30%. This was typical for all of the genes in the *aat* cluster, but lower than the overall GC content of the genomes in which they were present. What was clear from the analysis is that the *aat* cluster is organised in three main ways. There is a full cluster, these contained all of the *aat* genes with *aap* or *cexE* upstream. Two examples are *S. enterica* subsp. *diarizonae* and *E. coli* E22. In these instances, the *aat* cluster is flanked by genes encoding hypothetical proteins. In the second group, the *aap* or *cexE* genes are distant from the *aatPABCD* operon. For example, in *E. coli* NCT86 and EAEC 042, the *aap*/*cexE* gene was 38 and 32 kb away, respectively. In the final organisation, there are significant gaps between *aatD* and *aap*/*cexE* and *aatPABC*, which are fully conserved as a single operon. The genes were flanked by transposable elements, similar to NCTC86 and EAEC 042. In summary, both *aatD* and *aap/cexE* are always present with an *aatPABC* operon, but they are not necessarily closely linked.

6.2.3. Possible mechanism of function of AatP inferred from protein homology

Homology to characterised proteins can suggest a biological function. Since BLAST did not detect any homology of AatP to domains of proteins with known function, a hidden Markov model (HMM) of AatP was used to search the Swissprot database. This search identified 126 protein sequences homologous to AatP. MacB was the major protein identified with 91 sequences. Other well characterised proteins such as LolC, LolD and FtsX were also identified. The MacB sequences were clustered to 70% identity using cd-hit.

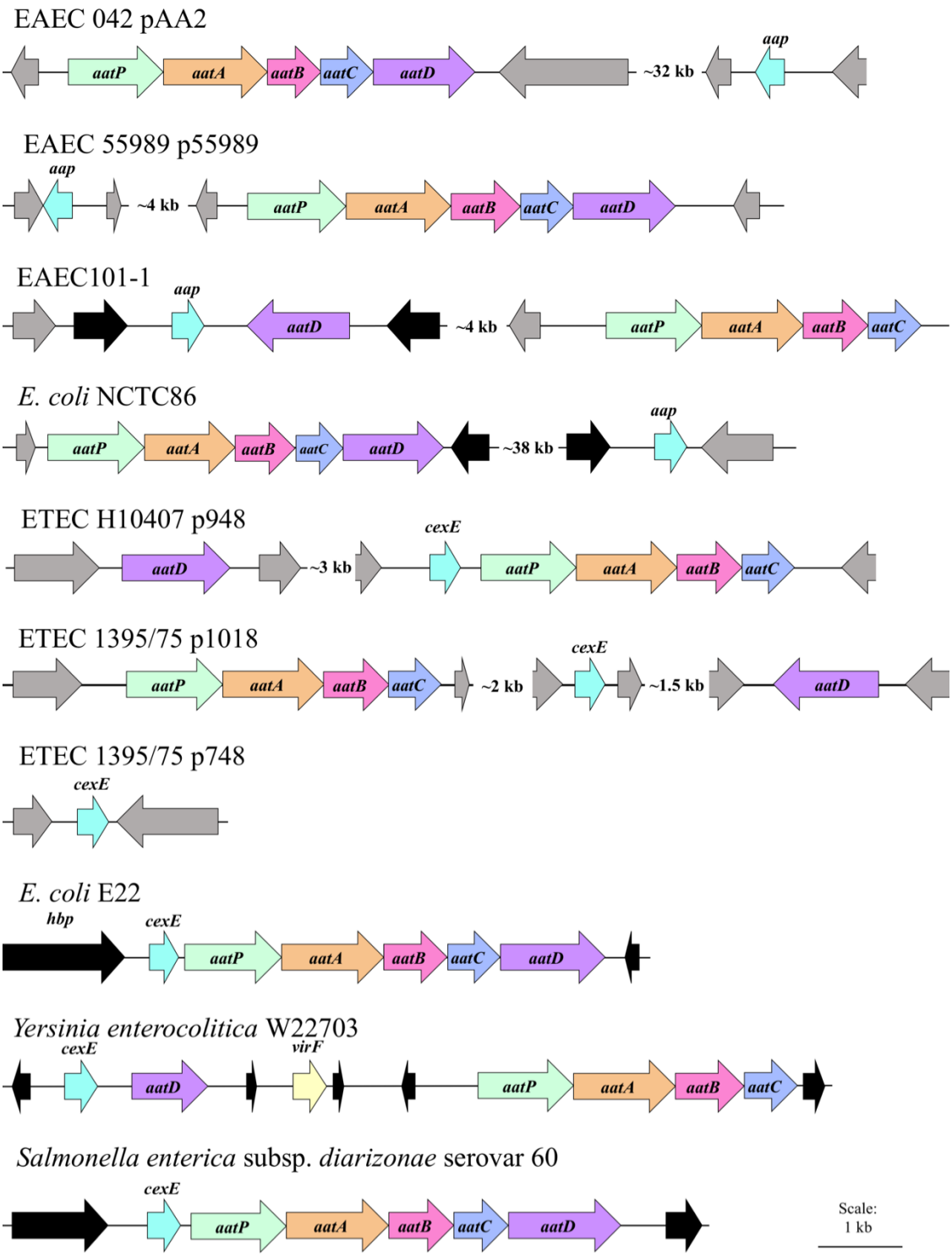


Figure 6.3: Examples of the *aat* and *aap*/*cexE* loci organisation. The genes were drawn to the scale indicated in the bottom right. Grey arrows indicate transposable elements, such as transposons or other insertion elements. The black arrows were ORF that were annotated as encoding hypothetical proteins. The *aat* cluster was present on the chromosome of the strain unless otherwise stated in the name.

Clustering refers to the grouping of sequences with more than 70% identity, the longest sequence is used to represent the cluster. The number of changes between each sequence was calculated using RAxML. A phylogram was constructed of the AatP homologs and AatP (Fig. 6.4). Further analysis revealed that the transmembrane and periplasmic regions of MacB, but not the nucleotide binding domain, are homologous to AatP (Appendix III). The MacB clade contained two uncharacterised proteins, YnzK and YvrN, both from *Bacillus subtilus* 168. These proteins contain an N-terminal MacB periplasmic domain (PF12704) and a C-terminal FtsX-like domain (PF02687). These two domains were again present in the uncharacterised proteins, Y1507_METJA and Y797_METJA, from *Methanocaldococcus jannaschii* and YCLI_BACSU and YTRF_BACSU from *B. subtilus*. A group of uncharacterised *Mycobacteria* proteins were identified by the search: Y072_MYCTO; Y2593_MYCBO and Y2563_MYCTO and genes encoding FtsX (Fig. 6.4). The closest homolog of AatP was FtsX of the *Mycobacteria*; FtsX of *E. coli* MG1655 was not a significant hit. However, the *E. coli* MG1655 proteins LolC and LolE were identified as homologs of AatP. These proteins are involved in the extraction of lipoproteins from the inner membrane. This suggests that AatP might be involved in the release of acylated proteins from the inner membrane.

6.2.4. Phylogenetic analysis of the AatA protein

AatA is homologous to TolC (Nishi et al., 2003). However, the relationship of AatA to the other outer membrane proteins is not understood. Homologs of AatA were identified using a HMM of AatA. The search of the Uniprot database with the AatA HMM identified homologous 3,563 sequences. As there were too many sequences for individual further analysis, an HHpred search was used to identify AatA homologs. This analysis revealed that AatA is homologous to outer membrane proteins involved in type I secretion and efflux. For each homolog of AatA identified by HHpred, amino acid sequences were compiled and aligned. The number of

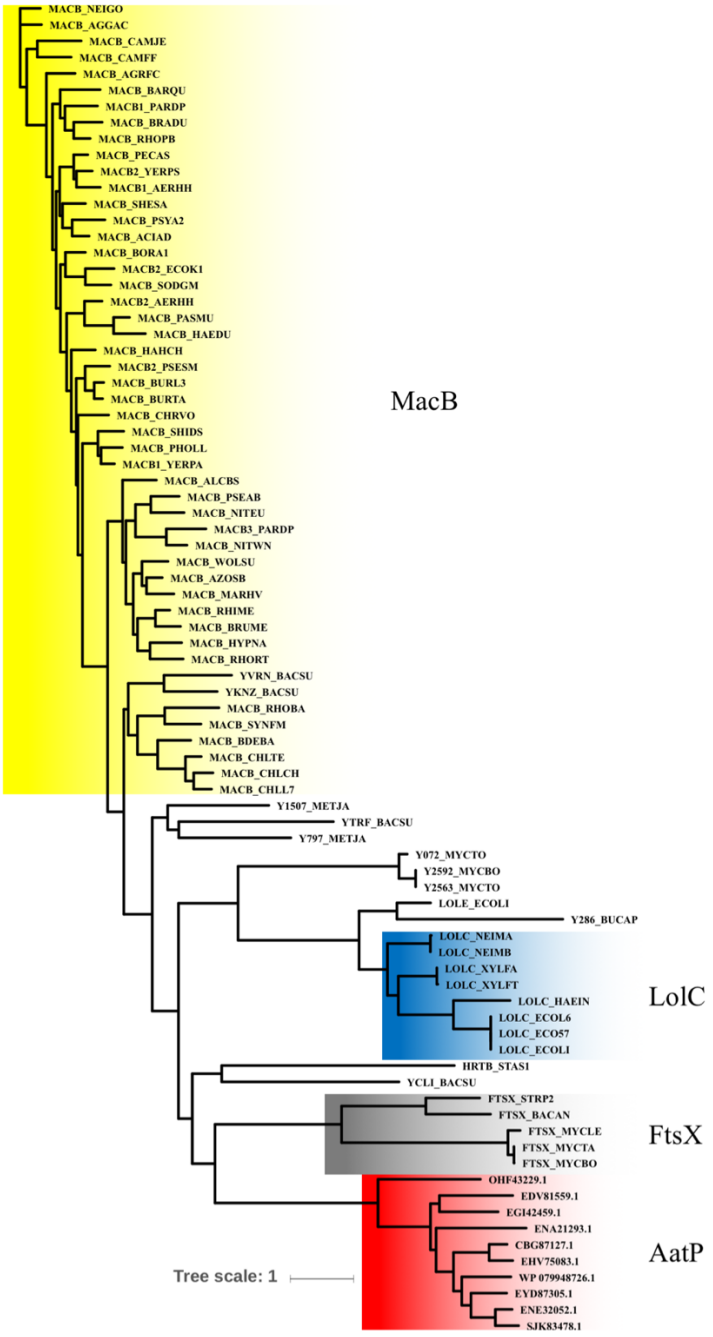


Figure 6.4: Results of search using AatP HMM. A HMM using AatP sequences was built in HMMER and used to search the Swissprot database. In the resulting homologous sequences there was a high proportion of MacB sequences (91 out of 126). To reduce the amount of analysis the MacB sequences were represented by clusters that had less than 70 identity to other clusters of MacB sequences. The remaining sequences were aligned using Clustal Omega. Branch lengths were calculated using RAxML. The tree was drawn using iTOL. The tree was left unrooted. Clades were assigned based on the function of the protein sequences present.

changes between each sequence was calculated and a phylogram was drawn (Fig. 6.5). The branch lengths between the outer membrane proteins involved in T1SS, AprF/PrfT, TolC, and CyaE, were small compared to the other outer membrane proteins. The same was true of the outer membrane proteins involved in efflux, CmeC, VceC, CusC, OprM, OprJ, and MtrE. However, the AatA clade is more divergent from any other clade in the analysis. Therefore, it was concluded that the function of AatA is likely to be distinct from the other characterised T1SS or efflux outer membrane protein.

6.2.5. Inference of the function of AatB from protein homology

The function of AatB is not understood. A HMM of AatB was created, which was then used to search the Uniprot database. In this search, 7,586 sequences were identified. These sequences were mostly membrane fusion proteins. However, the number of sequences identified was not conducive to further analysis. Therefore, HHpred was used to find homologs of AatB. The most significant hits were to membrane fusion proteins. An E-value less than 1e-16 was associated with the MFPs: AcrA, MacA, ZneB, CusB and HlyD. This suggests that AatB is the membrane fusion protein of the Aat system.

The relationship between AatB and other membrane fusion proteins is not understood. The distance between each sequence was calculated. A phylogram was drawn with the branch lengths indicating the distance between each sequence (Fig. 6.6). Similar to the outer membrane proteins, the membrane fusion proteins form clades based on their function in efflux or type I secretion (Fig. 6.5). AcrA, MacA, CusB and ZneB are involved in efflux systems and HlyD is required for haemolysin secretion. However, the distance for the AatB clade was greater than that for any other clade. Therefore, the mechanism of action of AatB is different from that of a typical T1SS membrane fusion protein.

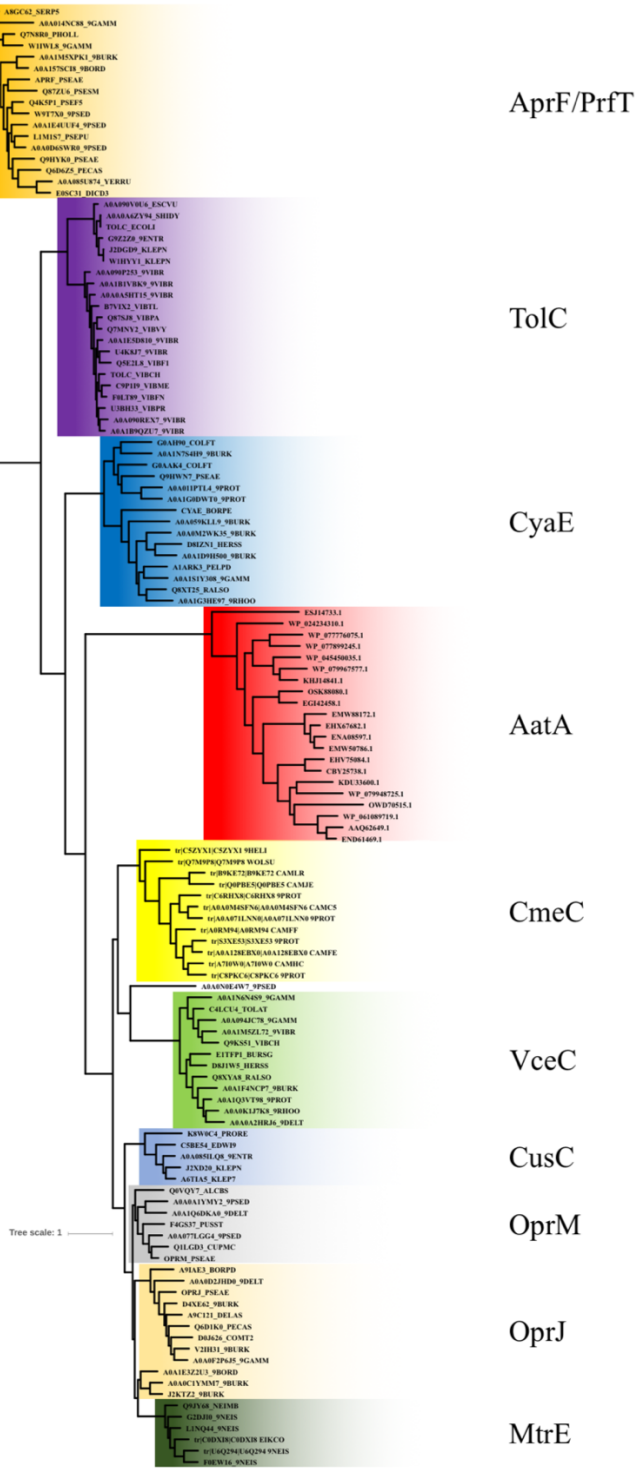


Figure 6.5: Phylogram of the relationship of AatA to other OMPs. HHpred was used to identify homologs of AatA using the EAEC 042 AatA protein sequence as the query. Representative sequences of each homolog were used to construct a phylogram. Sequences were aligned with Clustal Omega, the variation in sequences was calculated with RAxML and the tree was drawn with iTOL. The tree was left unrooted.

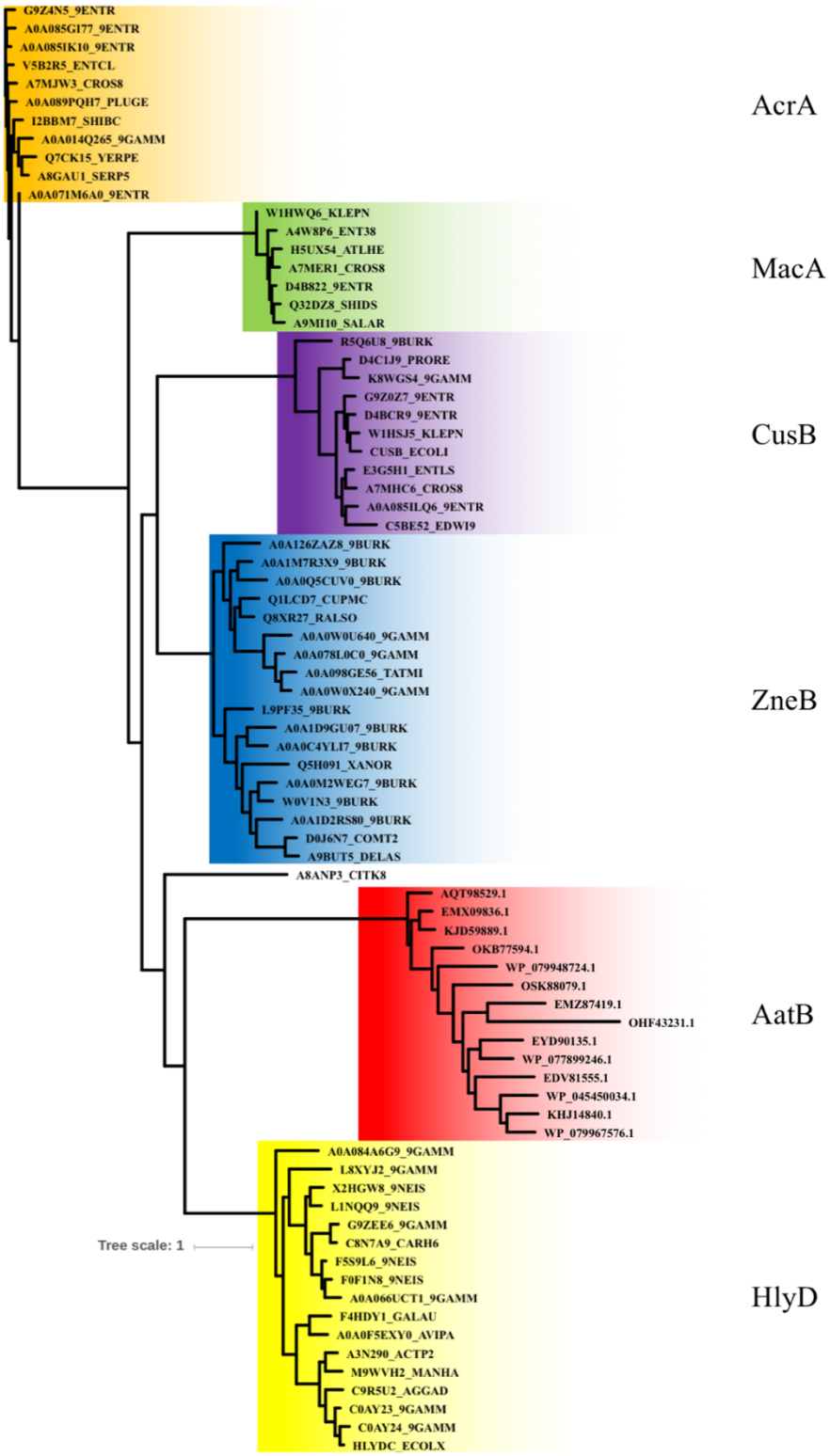


Figure 6.6: Relationship of AatB to membrane fusion proteins. HHpred was used to identify homologs of AatB using the EAEC 042 AatA protein sequence as the query. Representative sequences for each homolog were aligned using Clustal Omega, branch lengths were calculated with RAxML and drawn using iTOL.

6.2.6. Homology searches for AatC

Previously, it was suggested AatC is an ATPase (Nishi et al., 2003). A BLAST search revealed AatC possessed the COG1136, cd03255, TIGR03608, PRK10535, pfam00005 and smart00382 protein domains. All of these domains correspond to an ATPase domain. These domains are found in LolD (lipoprotein export), MacB (macrolide efflux), FtsE (cell division), and other ABC transporters. Although AatC is homologous to MacB, AatC is not homologous to the same region of MacB as AatP. The sequence of AatC aligns with the N-terminal nucleotide binding domain of MacB (Appendix IV). Thus, it is very likely that AatC is an ATPase.

6.2.7. Identification of characterised proteins homologous to AatD

The function of AatD in Aap secretion is unknown. AatD homologs were not found in association with other known T1SS. A hidden Markov model of AatD was built, which was then used to search the Uniprot database. From the hidden Markov model search of the Uniprot database 71 sequences, including AatD, were identified. The sequences were aligned. The number of changes between each sequence was calculated and a phylogram was drawn (Fig. 6.7). The non-AatD proteins were apolipoprotein N-acyltransferases (Lnt). While Lnt of *E. coli* MG1655 was identified by this search, it was not included in the phylogram analysis as it fell below the threshold for significance. Lnt in *E. coli* MG1655 is involved in the addition of acyl groups onto the C-terminus of lipoproteins prior to their LolE-mediated extraction from the inner membrane. This suggests that AatD is involved in protein acylation.

Proteins with similar functions often have conserved residues. Although, *E. coli* MG1655 Lnt was not a significant hit for AatD it is the best studied apolipoprotein N-acyltransferase. The residues required for the catalytic function of Lnt are known. Thus, the sequences of AatD above were aligned and a sequence logo of conserved residues was produced (Appendix V).

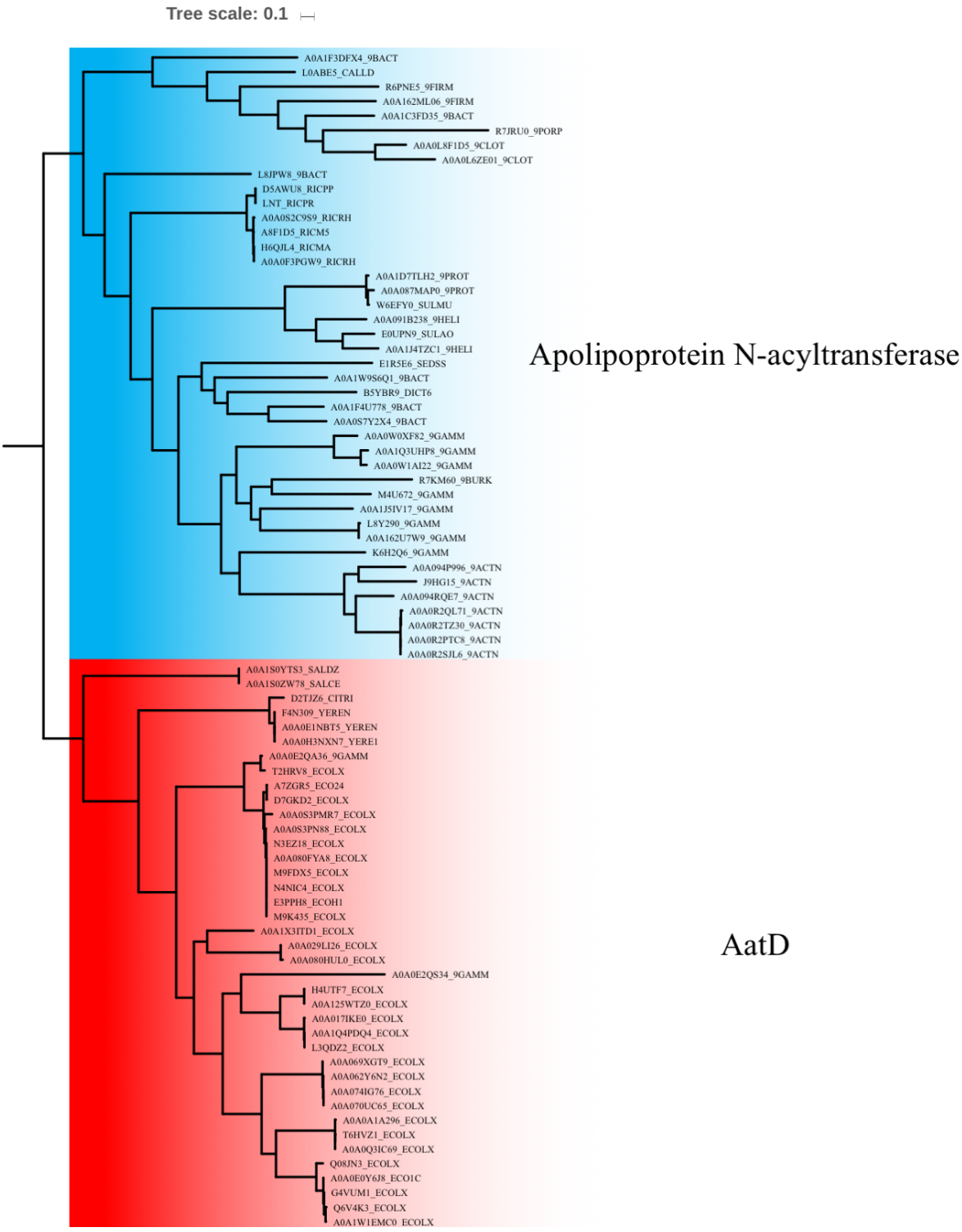


Figure 6.7: Phylogram of AatD and apolipoprotein N-acyltransferase. A HMM of AatD was constructed using the AatD sequences identified by PSI-BLAST. The results of the Uniprot database search were aligned using Clustal Omega. The phylogram was calculated using RAxML. The tree was drawn with iTOL. For easier interpretation of the data the tree was rooted between the AatD and apolipoprotein N-acyltransferase.

Lnt from *E. coli* MG1655, AatD from EAEC 042 and AatD from ETEC H10407 were aligned and compared with the conserved sequence logo of AatD. The residues known to be required for Lnt function were compared to the consensus sequence of AatD (Fig. 6.8). The catalytic residues of Lnt are known to be E267, K335 and C387 (Nakayama et al., 2012). The residues of the active site were all conserved in AatD. Thus, it was concluded that the catalytic residues required for the acylation of proteins by Lnt were conserved in AatD.

6.2.8. Aap production in EAEC 042

In the work by Nishi et al. (2003) the mutants were not complemented to show restoration of Aap secretion. Therefore, to confirm the previous observations. Overnight cultures of *aap*, *aatA*, *aatC* and *aatD* mutants with corresponding complements for the *aat* mutants and the parental strain were used to inoculate pre-warmed DMEM high glucose. The cultures were grown at 37ºC with aeration. After 5 hours of growth, Triton X-100 was added to cultures to remove Aap from the cell membrane. Cultures were grown for another hour. For whole cell lysates, samples were harvested by centrifugation and resuspended in Laemmli buffer. The culture supernatant was retained for protein precipitation. Whole cell lysates were separated on Tris-tricine gels to determine the production of Aap. Gels were stained with Coomassie Brilliant Blue or transferred for western blotting using anti-Aap antibodies (Fig. 6.9). On the Coomassie Brilliant Blue stained gel, Aap was not visible (Fig. 6.9A). On the western blot, Aap was detected in all of the strains except the *aap* mutant (Fig. 6.9B). The amount of Aap was relatively similar in each of the *aat* mutants. There was a slight increase in the amount observed for the *aatA* mutant. The *aatA* mutant was complemented with pJNW, a single copy number vector with the *aat* system under native expression. In the complemented *aatA* mutant strain the amount of Aap detectable was less than that of the mutant. In the *aatC* mutant a

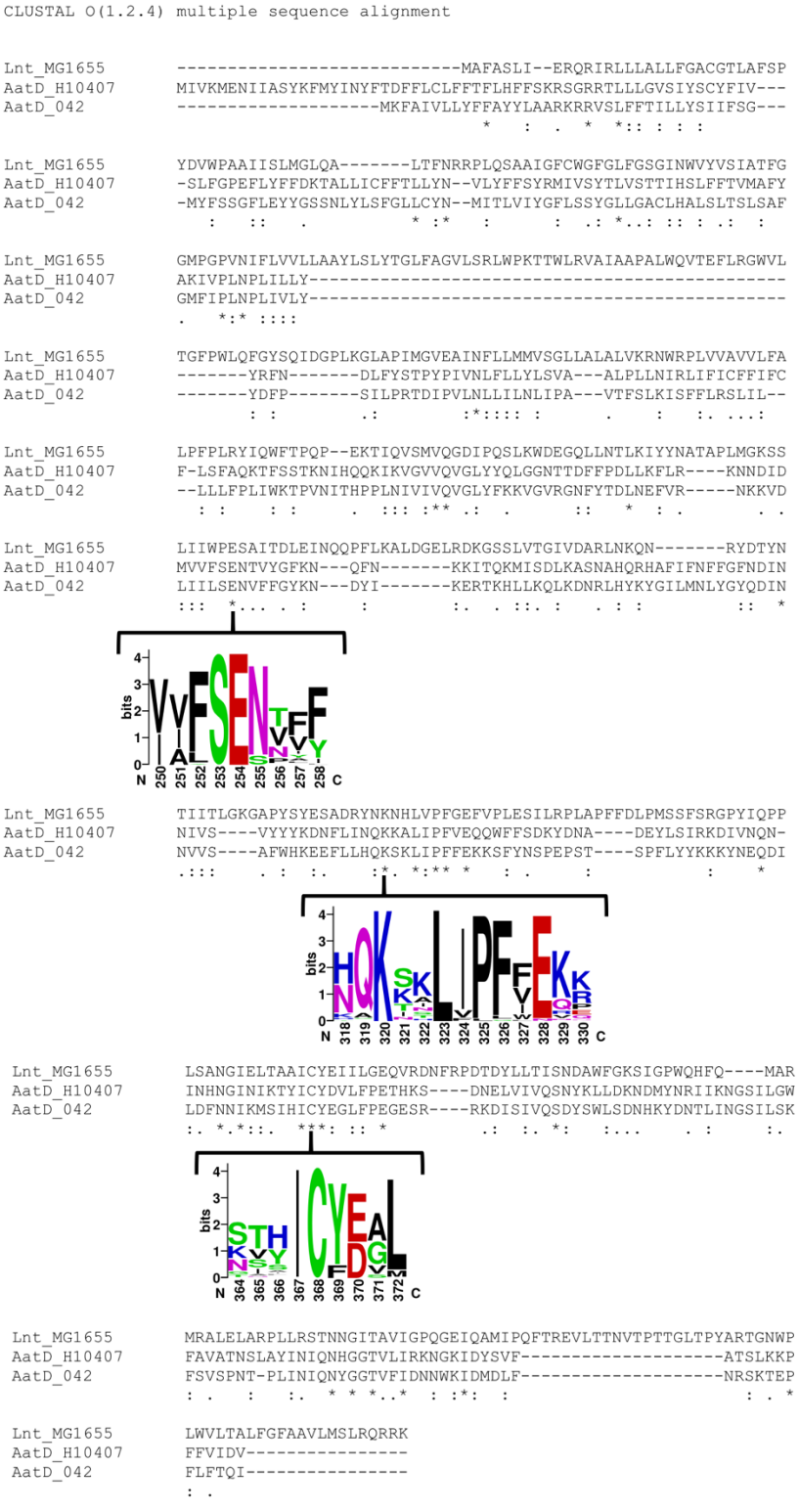


Figure 6.8: Conservation of Lnt catalytic residues in AatD. A sequence logo was created of AatD sequences identified by PSI-BLAST. The height of the residue indicates the level of conservation. The catalytic residues E267, K335, and C387 are indicated in this figure as E254, K320, and C368, respectively.

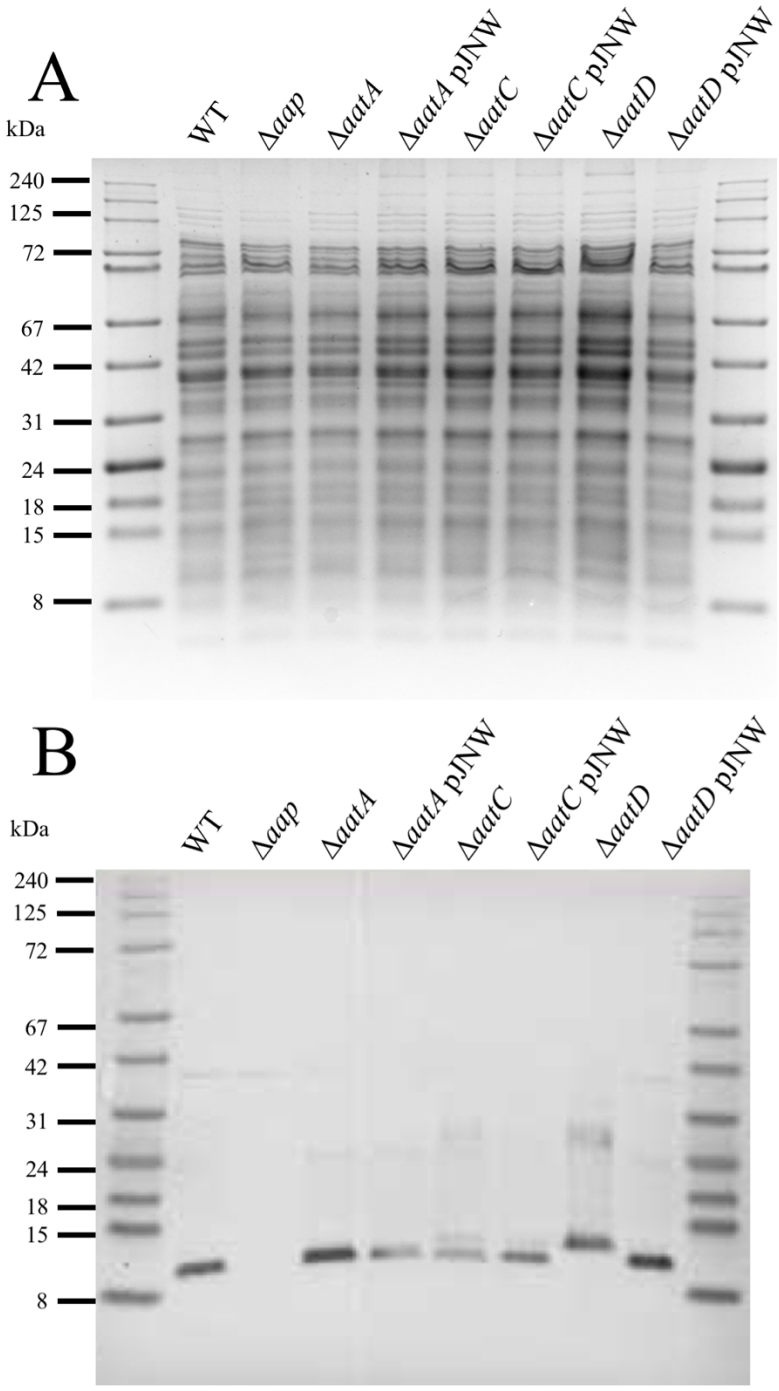


Figure 6.9: Whole cell lysates of EAEC 042 and *aat* mutants. EAEC 042, *aap* mutant, and *aat* mutants with and without the pJNW plasmids were grown in DMEM high glucose for 6 hours. The complementation plasmid pJNW contains the entire *aat* cluster on a single copy number plasmid, pZC320. Whole cell lysate samples were prepared and separated on Tris-tricine gels. A, the Coomassie stained gel. B, the western blot with primary antibodies specific for Aap. This experiment was repeated twice with the same results.

double band was present for Aap. The double band consisted of a wild-type sized Aap and a band that appeared to be of a higher molecular weight. Less Aap was present in the whole cell lysate in the *aatC* mutant than in the wild-type. Complementation of the *aatC* mutant reduced the amount of the retarded band present for Aap. In the *aatD* mutant, the Aap protein was retarded in the gel. Aap in the *aatD* mutant was the same size as the higher molecular weight band in the *aatC* mutant. In the complemented *aatD* mutant, Aap migration was comparable to the wild-type. The wild-type molecular weight Aap was only fully absent in the *aatD* mutant. It is hypothesised that the appearance of the high molecular weight protein species in the *aatC* mutant is due to polar effects on *aatD* expression. There was a reduced amount of AatD present in the *aatC* mutant due to the alteration of *aatD* transcription because of the insertion of the suicide vector into *aatC*. From this data, it is concluded that AatD was responsible for altering the electrophoretic mobility of Aap.

6.2.9. Secretion of Aap via the Aat system from EAEC 042

Nishi et al. (2003) reported that the *aat* system was required for the secretion of Aap. Proteins in the supernatant, from the cultures described above, were concentrated by TCA precipitation. Pellets were washed with methanol and resuspended in Laemmli buffer. Resuspended proteins, henceforth referred to as the supernatant fraction, were separated by electrophoresis on Tris-tricine gels and the presence of Aap was detected by staining with Coomassie Brilliant Blue and western blotting with Aap antibodies (Fig. 6.10). In the Coomassie Brilliant Blue stained gel, Aap was apparent in the parental strain and the complemented *aat* mutants but not the individual mutants (Fig. 6.10A). The western blot confirmed that the bands present in the Coomassie Brilliant Blue stained gel were the Aap protein (Fig. 6.10B). On the western blot, as expected, more Aap was observed in the wild-type samples and the complemented mutants than in the individual mutants. Nevertheless, a small amount of Aap could be observed in the

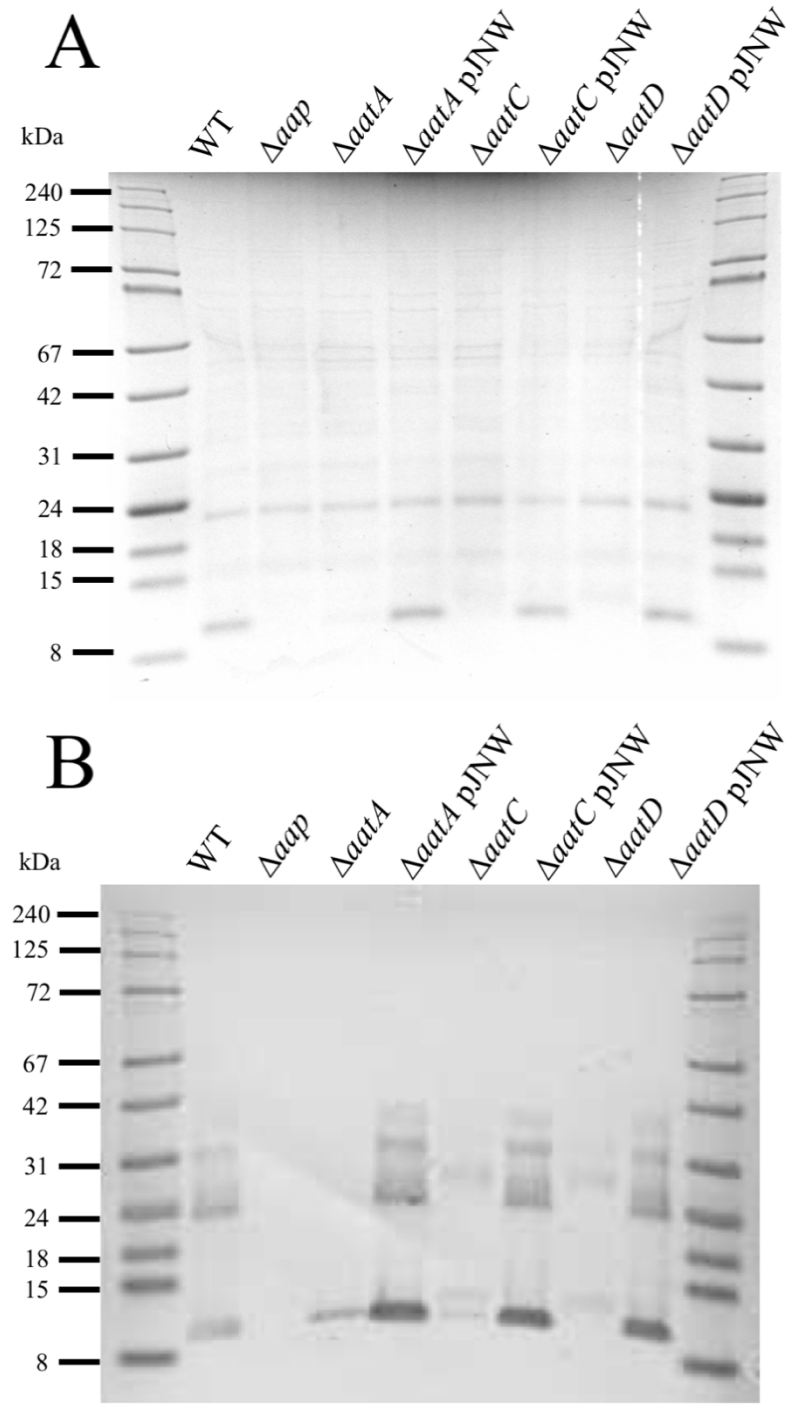


Figure 6.10: Supernatant of EAEC 042, *aap* and *aat* mutants. EAEC 042, *aap* mutant, and *aat* mutants with and without the pJNW plasmid were grown in DMEM high glucose for 6 hours. Culture supernatant fractions were prepared by TCA precipitation. Protein samples were separated on Tris-tricine gels. A, Coomassie Brilliant Blue stained gel. B, western blot with primary antibodies for the detection of Aap. This experiment was repeated twice with the same results.

aatA and *aatC* mutants. In the supernatant of the *aatC* mutant, a double band was present for Aap, similar to the double band observed in the whole cell lysate (Fig. 6.10). The complementation of the *aatC* resulted in a single band for Aap at a comparable size to that of the wild-type. The small amount of Aap present in the supernatant of the *aatD* mutant had an altered electrophoretic mobility; this was similar to the single retarded band in the *aatD* mutant whole cell lysate (Fig. 6.9). The complementation of the *aatD* mutant with pJNW restored Aap to a similar size and amount to the wild-type.

These data suggest that Aap exists in three forms. The first form is the fully translated protein containing the signal sequence (pre-pro-Aap). The second form occurs post Sec secretion after single sequence has been cleaved (pro-Aap). The third form is the mature Aap after AatD-mediated modulation of the protein. Modification to this latter form is required for secretion into the culture supernatant. As Aap was produced in each of the EAEC 042 mutant strains, except *aap* mutant, but was only secreted when the *aat* system was intact, it was concluded that the *aat* system is required for the secretion of Aap. However, the work presented here revealed a previously undocumented post-translational modification of Aap.

6.2.10. Production of CexE in ETEC H10407

Before a secretion of a protein can be attributed to a secretion system the protein must be produced. In Chapter 4, it was shown that the CexE protein was not produced in ETEC H10407 that had been grown aerobically at 37ºC (Fig. 4.8). Different conditions were tested to find the conditions under which ETEC H10407 produces CexE. CexE was detected not in any of the conditions tested.

It is known that the *cexE* expression is activated by CfaD (Pilonieta et al., 2007). The *cfaD* gene was cloned into pBAD/myc-His A expression plasmid under the transcriptional control of the *ara* promoter to drive the production of CexE, by Sammi McKean. The *cexE* mutant and the parental strain were transformed with the resulting recombinant plasmid, pCfaD. Cultures of ETEC H10407 pCfaD and ETEC H10407Δ*cexE* pCfaD were grown overnight at 37ºC with aeration in LB supplemented with 0.2% L-arabinose. Whole cell lysates were prepared by centrifugation and resuspension in Laemmli buffer. Culture supernatants were harvested by centrifugation and proteins were precipitated by TCA. Whole cell lysate and supernatant protein samples were separated by electrophoresis on Tris-tricine gels. CexE was detected by staining gels with Coomassie Brilliant Blue and transferring gels for western blotting with anti-CexE antibodies (Fig. 6.11). For the Coomassie Brilliant Blue stain, a band at around the size of the CexE protein (10.6 kDa) was clearly visible in the wild-type supernatant. There was more protein present in the culture supernatant of the wild-type than the mutant. However, there was no protein in the supernatant fraction of the mutant that potentially could be CexE. The band in the wild-type supernatant was confirmed to be CexE by the western blot. The apparent mass of the protein identified in the western blot as CexE in the whole cell lysate was smaller than that in the supernatant. There was no band at the same size in the *cexE* mutant. Thus, it was concluded that the expression of *cfaD* from pCfaD was able to drive expression and subsequent secretion of CexE in ETEC H10407.

6.2.11. Construction of *aat* gene deletions in ETEC H10407

Crossman et al. (2010) had previously reported an *aat* system in ETEC H10407. It was hypothesised that this Aat system was required for the secretion of CexE. Mutants defective in the *aat* system were constructed in ETEC H10407 using λ Red recombination to test this hypothesis. In brief, primers were designed to amplify a kanamycin fragment with 50 bp

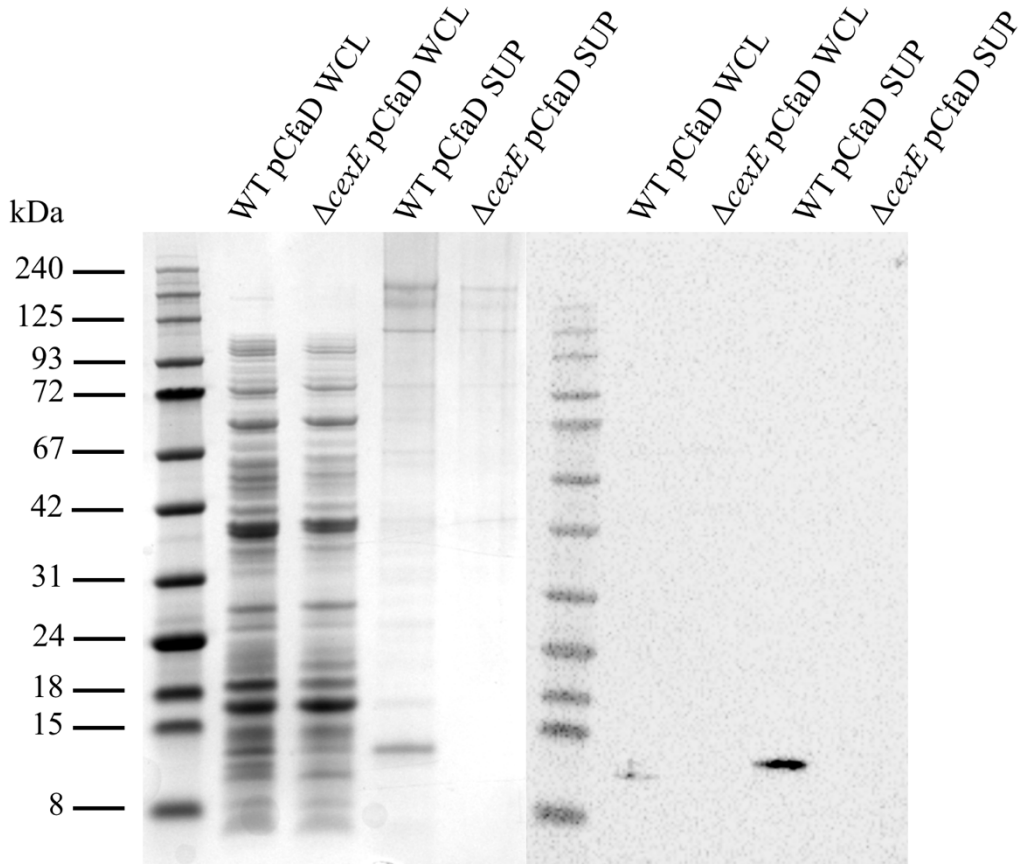


Figure 6.11: Production of CexE from pCfaD transformed ETEC H10407. A *cexE* mutant and wild-type strain transformed with pCfaD were grown in LB supplemented with 0.2% L-arabinose for 6 hours. Whole cell lysates and culture supernatant fractions were prepared from the same culture. Protein samples were separated on Tris-tricine gels. The Coomassie stained gel is on the left and the western probing for CexE is on the right. WCL were whole cell lysates and SUP were the culture supernatant. Anti-CexE antibodies with secondary HRP conjugate antibodies were used for the detection of CexE.

homology to the target gene. Homology regions were targeted to the start codon and the last 7 codons of the target gene. Linear fragments for each primer pair were amplified containing a kanamycin resistance cassette from pDOC-K. An overnight culture of ETEC H10407 pKD46 was used to inoculate LB supplemented with 0.2% L-arabinose. Cells were harvested by centrifugation. The cells were made electrocompetent. The linear fragments were transformed into ETEC H10407 pKD46 by electroporation. Cells were recovered in SOC at 37ºC with aeration and were plated on to LBA supplemented with kanamycin. After overnight growth at 37ºC, kanamycin resistant colonies were streaked to single colonies on LBA supplemented with kanamycin. A colony PCR using primers external to the gene of interest was used to confirm the successful mutation of the *aat* genes. For each candidate *aat* mutant, a wild-type colony was used as a positive control and nuclease free-water was used as a negative control. A single colony from the streaked plate was used to test for the presence of the mutation. For each of the five mutants, *aatP*, *aatA*, *aatB*, *aatC*, and *aatD*, the PCR product was smaller in the parent strain, confirming the construction had been successful (Fig. 6.12A). For extra confirmation of the external PCR, primers with homology to regions inside of the gene of interest were used as primers for a second PCR. The templates described above were used for a successive set of PCRs. The presence of a PCR product indicated that the gene was still present (Fig. 6.12B). No PCR product was present for the *aat* genes. As the strains streaked to single colonies were not contaminated with the wild-type, it was concluded that the mutation of the *aat* genes had been successful.

6.2.12. CexE production in ETEC H10407 *aat* mutants

Before proteins can be secreted they must be produced in the cell. Overnight cultures of *cexE, aat* mutants and the parental strain were used to inoculate LB supplemented with 0.2% L-arabinose. After 6 hours of growth at 37ºC with aeration, whole cell lysates were harvested by

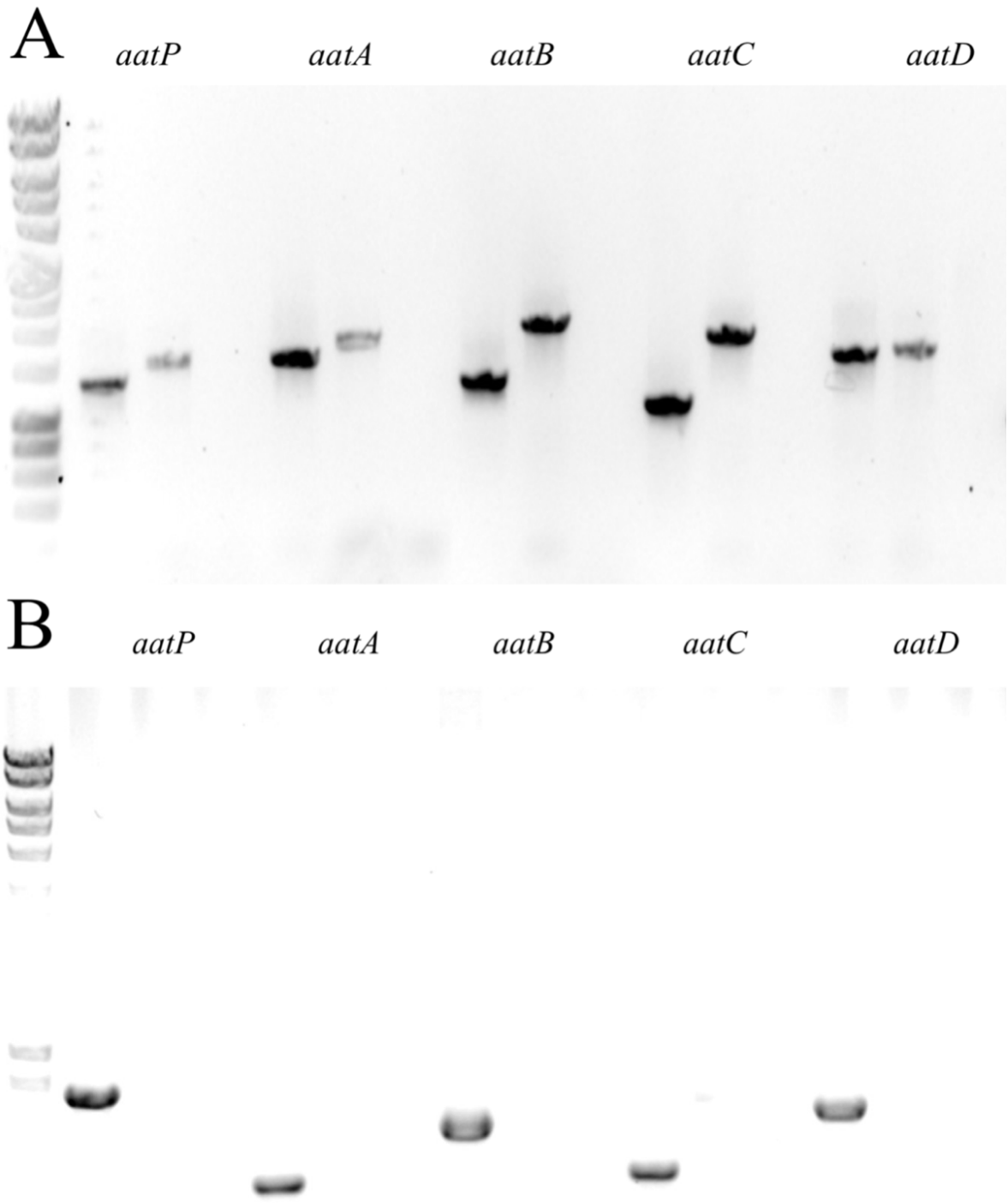


Figure 6.12: PCR products for the *aat* genes and *cexE*. For each gene, 3 reactions were set up; wild-type, insertion mutant and nuclease free water control. In section A, the primers are designed to anneal to positions outside of the target gene. In section B, the primers are design to anneal to regions inside the target gene.

centrifugation. Protein samples were assessed for the presence of CexE by western blot (Fig. 6.13). The β' subunit of RNAP was used as a loading control. The amount of β' subunit of RNAP in each lane was comparable. The CexE protein was produced in each of *aat* mutants and the parental strain. However, in all of the mutants, except the *aatD* mutant, CexE was present in two forms with differences in electrophoretic mobility. In the *aatD* mutant, all of the CexE protein migrated at the higher molecular weight of the two wild-type forms. Thus, the larger band was termed pro-CexE and the smaller, presumably mature form, was termed CexE. Therefore, it was concluded, that like Aap in EAEC 042, the Aat system is required for CexE secretion and AatD is responsible for the post-translational maturation of CexE in ETEC H10407.

6.2.13. Differences in the secretion of CexE between the *aat* mutants and wild-type in ETEC H10407

As cell lysis occurs during growth, proteins that are not secreted appear in the culture supernatant. Overnight cultures of ETEC H10407 pCfaD and ETEC H10407Δ*aatA* pCfaD were used to inoculate LB supplemented with 0.2% L-arabinose. Cultures were grown at 37ºC with aeration, culture supernatant were prepared as above from the overnight culture without induction and after 2 and 3 hours of growth with induction. The production of CexE in the samples was assessed by Coomassie Brilliant Blue stain and western blot as described above. No CexE protein detected in the overnight culture supernatant (Fig. 6.14AB). After 2 hours of growth the presence of 0.2% L-arabinose, CexE was produced. No CexE was detectable in the *aatA* mutant supernatant at this time. After 3 hours of growth, there was more CexE in the supernatant of the wild-type strain compared to the *aatA* mutant. However, the CexE that was present in the supernatant of the *aatA* mutant was pro-CexE. Thus, it was hypothesised that the secretion of CexE into the supernatant requires an intact Aat system. It was suspected that the

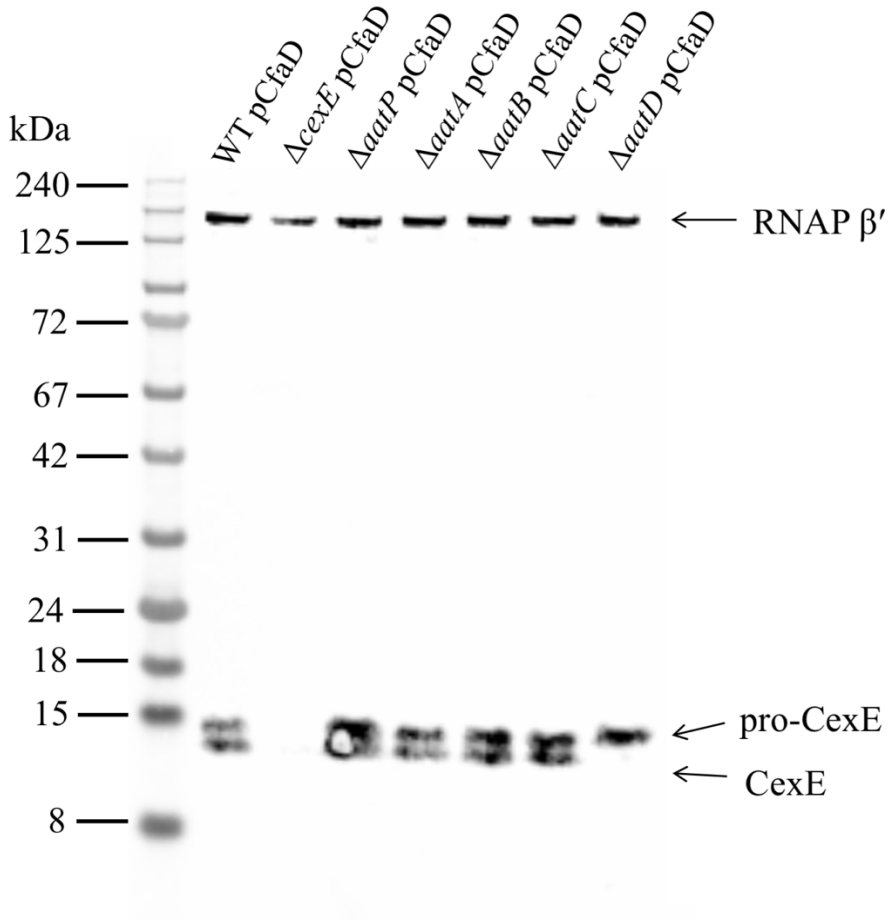


Figure 6.13: CexE production in ETEC H10407 and *aat* mutants. Cultures of *aat* and *cexE* mutants and the parental strain transformed with pCfaD were grown in LB supplemented with 0.2% L-arabinose. After 6 hours of growth, whole cell lysate samples for each culture were prepared. Protein samples were separated on NuPAGE Bis-Tris 4-12% protein gels. Proteins were transferred to nitrocellulose and western blotted with anti-CexE and β′ RNAP antibodies. For the secondary antibodies, a HRP conjugate was used. The presence of only the larger band in the *aatD* mutant was routinely observed across the three replicates.

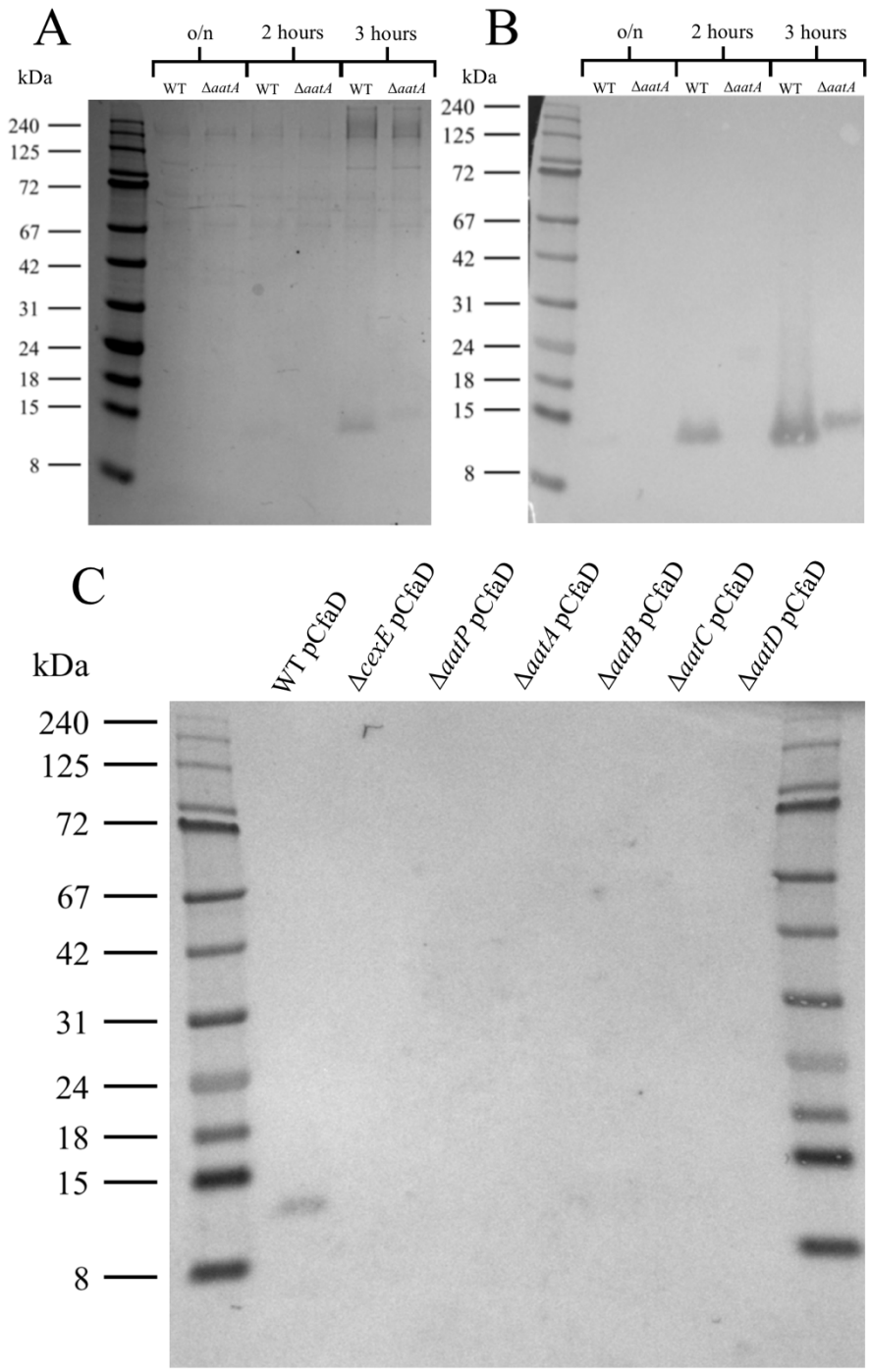


Figure 6.14: CexE secretion in ETEC H10407 and *aat* mutants. A, Coomassie Brilliant Blue stain of ETEC H10407 and *aatA* culture supernatants after overnight growth and 2 and 3 hours post L-arabinose induction. B, western blot for the detection of CexE for the samples in A. C, western blot for the detection of CexE in the culture supernatants of ETEC H10407, *cexE*, and *aat* mutants after 2 hours growth in LB supplemented with 0.2% L-arabinose.

presence of pro-CexE in the supernatant was due to cell lysis rather than active secretion. Therefore, supernatants for the examination of CexE secretion were taken after 2 hours of growth in LB supplemented with 0.2% L-arabinose. The Aat proteins were hypothesised to be required for the secretion of CexE from ETEC H10407. Overnight cultures of wild-type, *cexE* mutant and the *aat* mutants were used to inoculate LB supplemented with 0.2% L-arabinose. Strains were grown for 2 hours at 37ºC with aeration. Culture supernatants were harvested as previously described. CexE was detectable by western blot only in the culture supernatant of the parental strain (Fig. 6.14C). Therefore, it was concluded that the Aat system must be fully intact for the secretion of mature CexE into the supernatant.

6.3. Discussion

From the prevalence analysis, it appears that *aap* or *cexE* homologs co-occur with the *aat* system. There are few examples *S. enterica* strains harbouring the *aat* system, whereas there are many ETEC, EAEC and *Y. enterocolitica* strains that do. From the current analysis, it appears there are no *aat* systems outside of the *Enterobacteriaceae*. Therefore, the function of Aap/CexE is thought to be more important for these species.

The bioinformatic analysis of the Aat proteins indicates that the current model of Aap/CexE secretion by the Aat system requires modification. Nishi et al. (2003) proposed that Aap is secreted into the periplasm and then enters the Aat system in the periplasm (Fig. 6.1). From their bioinformatic analysis, they concluded that AatA was the outer membrane pore, AatC was the ATPase and AatP the permease. Sufficient similarity with other proteins in the database was not detected to assign possible functions for AatB and AatD. Improvements in computation power, homology search algorithms and database size enabled the detection of the homology of AatB to membrane fusion proteins and AatD to apolipoprotein N-acyltransferase. Therefore,

an alteration of the Aat secretion model is proposed, in which the function of AatP, AatA and AatC are retained, as the outer membrane protein, permease, and ATPase, respectively. AatB and AatD are assigned updated putative functions based on their homology. AatB is confirmed as the membrane fusion protein of the Aat system. Membrane fusion proteins interact with the OMP and ABC transporters to form a pore through which an unfolded protein can be secreted (Kanonenberg et al., 2013). AatD is thought to be present in the inner membrane. While the AatD-dependent alteration of Aap/CexE is yet to be determined it is known that it is required for secretion. However, the differences in size of both Aap and CexE in EAEC 042Δ*aatD* and ETEC H10407Δ*aatD* confirm that there is an as yet unknown modification. Previous studies of the Aat system did not report that Aap is post-translationally modified after Sec secretion. The reason why it was identified in this study is thought to be due to the change in protein gel systems. Previously high acrylamide concentration, 20%, Tris-glycine gels were used (Nishi et al., 2003). In this analysis, Tris-tricine gels were used. This was due to the improved resolution of Tris-tricine gel separation of low molecular weight proteins (Schägger, 2006). In EAEC 042, the mutation of *aatC* resulted in a double band for Aap similar to that observed for ETEC H10407. However, it is thought that the organisation of the *aat* cluster in EAEC 042 is responsible for the double band present in the *aatC* mutant (Fig. 6.9 and Fig. 6.13). The mutants used in this experiment were insertional mutants donated by Nishi et al. (2003). Thus, the insertion into the *aatC* gene resulted in the incorrect expression of *aatD*. In EAEC 042, *aatPABCD* are present as a single operon, the single-crossover insertion of plasmid pJP5603 is likely to have disrupted native *aatD* expression. Therefore, insufficient quantities of AatD were produced for full maturation of Aap. It was for this reason that the conclusion was that AatD was responsible for Aap post-translational modification. This conclusion was confirmed in ETEC H10407. The double bands for CexE in the ETEC H10407Δ*aatC* mutant were

comparable to the other CexE producing mutants and wild-type except for the *aatD* mutant (Fig. 6.13).

Experimental evidence for the alteration of CexE/Aap by AatD was presented here. The exact modulation of CexE/Aap was not clear. Further investigations into the modified CexE/Aap protein is required. Purification of the CexE and Aap protein from EAEC 042 and ETEC H10407 wild-type and *aatD* mutants will reveal the exact change. The suspected maturation event, implied from homology searches, is the addition of a lipid group onto the N-terminus of Aap/CexE. The mutation of *aatD* in EAEC 042 and ETEC H10407 results in what appears to be a smaller protein. The actual addition and position of attachment is yet to be determined. Attempts to isolate the matured Aap and pro-Aap protein from the supernatant of EAEC 042 and EAEC 042Δ*aatD* were not successful. Therefore, it is suggested that cultures of EAEC 042Δ*aatD*, ETEC H10407Δ*aatD*, and their parental strains are grown in conditions for the production of Aap and CexE, respectively. From the lysed cells the pro- and mature forms of Aap and CexE are purified using their isoelectric points. This approach allows the purification of the proteins without any additions. With the use of mass spectrometry, the actual size change and the position in which it is occurring can be defined.

If this speculation is correct this would also explain the homology of AatP to LolC and LolE in the lipoprotein release system. LolE in known to capture lipoproteins from the inner membrane after the final acylation by Lnt (Mizutani et al., 2013b). After this, the lipoprotein is passed to LolA (Okuda and Tokuda, 2009). The homology of AatD to Lnt and AatC to LolD support this theory, albeit ATPase protein sequences are very similar to each other. However, the next steps in lipoprotein transport do not bear any resemblance to proteins in the Aat system. These speculations currently do not explain the roles of AatA and AatB. Though these proteins

are distinct from their T1SS homologs. There is still much yet to understand about the secretion of Aap and CexE.

CHAPTER 7:

Final discussion

7.1. Remaining questions about the CexE protein

Prior to the investigation of CexE in ETEC H10407 the only information available was that the gene encoding it was activated by CfaD and it elicited an immune response in mouse colonisation experimets (Pilonieta et al., 2007; Roy et al., 2011). The results presented in Chapter 4 strongly indicated that CexE is an extracellular DNA binding protein required for ETEC H10407 biofilm formation. The main problem faced when investigating CexE function in ETEC H10407 was the inability to produce CexE in the wild-type. However, this issue was resolved in Chapter 6 by transformation of ETEC H10407 with pCfaD which enabled the control of *cfaD* expression in response to L-arabinose. The pCfaD plasmid was not constructed until late in the study. At the time a decision was made to prioritise the investigation of the secretion of CexE instead of the function in ETEC H10407. The immediate experiments following this thesis should be to compare the phenotypes of ETEC H10407 and the *cexE* mutant when *cfaD* is expressed.

The prioritisation of CexE secretion lead to a fascinating result. AatD was responsible for an apparent reduction in molecular weight of CexE which was clearly shown in the ETEC H10407 *aatD* mutant (Fig. 6.13). Interestingly this was also true when CexE was produced in EAEC 042 (Fig. 4.9). The causation of the double band in CexE was not investigated in EAEC 042, however it is very likely that the AatD protein was responsible. From this point on, in the interests of clarity $AatD_{042}$ and $AatD_{H10407}$ are the AatD proteins from EAEC 042 and ETEC H10407, respectively. The above result indicated a fascinating possibility that the $AatD_{042}$ recognises and alters the CexE protein from ETEC H10407. Therefore, $AatD_{042}$ and $AatD_{H10407}$ must be recognising the same signal from Aap and CexE. It is hypothesised that reverse of the experiment in Chapter 4 would result in $AatD_{H10407}$ modulating the Aap protein when produced in ETEC H10407. At the time, this experiment was not possible as there was no mechanism to

produce CexE in ETEC H10407. However, it is now viable with the pCfaD plasmid. The modulation of CexE by $AatD_{042}$ raises the interesting possibility that Aat proteins from different pathovars could function together. For instance, would $AatD_{042}$ complement a ETEC H10407 *aatD* mutant for the secretion of CexE?

While it is clear that $AatD_{H10407}$ and likely $AatD_{042}$ are responsible for the double bands of CexE it is not clear the underlying reason for this phenotype. There are thought to be two potential mechanisms responsible for this. The first, and simplest mechanistically, is that the double band of CexE is due to the inherent nature of the CexE protein. The second, and more likely, possibility is that a high production of CexE in both strains prevents the completion of the maturation process. In EAEC 042 the CexE protein is foreign and not designed to function with the Aat system that is present. While *cexE* is expressed similarly to *aap* in EAEC 042Δ*aap*::*cexE* the efficiency of the modulation of CexE by $AatD_{042}$ is unlikely to be equal to that of Aap. The result of acylating some but not all of the CexE produced is a double band. In ETEC H10407 the production of CexE is driven by the expression of *cfaD* from the pCfaD expression vector, which is unlikely to be similar to the native *cexE* expression as other factors that would typically activate *cfaD* expression are not present. Therefore, it is possible that the large amount of CexE being produced overloads the Aat system preventing complete modulation. Again, resulting in a double band. Altering of the L-arabinose induction concentration may remove the double band observed. This suggests a difference in the regulation of *aap*/*cexE* and the *aat* system which was confirmed in the RNA-Seq experiment.

7.2. Regulation of the *aat* genes and *aap*/*cexE*

From the RNA-Seq data of EAEC 042, presented here, and ETEC H10407, by Hodson et al. (2017), it is clear that *aap*/*cexE* are expressed differently to the *aat* system. Suggesting a reason

for the positioning of the *aap*/*cexE* genes in the genome. While frequently in close proximity *aap*/*cexE* never form part of the *aatPABC* operon (Fig. 6.3). This is clearly due to the difference in expression required of these genes likely because of a single Aat system being able to export multiple copies of Aap/CexE proteins. In EAEC 042 all of the genes in the *aat* operon were expressed relatively equally (Fig. 3.10). Interestingly this does not appear to be true for the *aat* genes in ETEC H10407. The data produced by RNA-Seq on ETEC H10407 by Hodson et al. (2017) indicated that the expression of *aatD* was higher than the genes in the *aatPABC* operon. This difference in the expression of *aatD* to the *aatPABC* genes is likely common amongst strains where *aatD* does not form part of an operon with the other *aat* genes. The separation of *aatD* from the other *aat* genes suggests that its expression needs to be different in these strains, likely indicating the importance of AatD in Aap/CexE secretion.

7.3. Hypothesised mechanism for the secretion of proteins by the Aat system

The function of the AatD is critical to understanding the Aat secretion mechanism. Unlike AatPABC, AatD is not homologous to any protein in the T1SS but is instead homologous to Lnt in the lipoprotein transport partway. While the actual modulation is still unknown it is likely that AatD functions similarly to Lnt. Like Lnt (Jackowski, 1986), AatD is presumably adding a fatty acid group onto the Aap/CexE protein resulting in the change in size (Fig. 6.9 and Fig. 6.13). Since this modulation of CexE can be catalysed by both $AatD_{042}$ and $AatD_{H10407}$, it is reasonable to speculate that both CexE and Aap have a similar signal which is recognised by AatD. The conserved residues of Aap and CexE seemed to be predominantly required for structure. However, the N-terminus after Sec signal cleavage was consistently glycine rich (Fig. 5.15). In fact, the N-terminal residue of these proteins after Sec signal cleavage was invariably a glycine. Being the only conserved residue that is no embedded in the structure it is thought that this glycine is the residue acylated by AatD. This forms the basis for the

proposed model for Aap/CexE secretion (Fig. 7.1). Aap/CexE is translated in the cytoplasm (Fig. 7.1A). The protein is translocated into the periplasm. Here is it able to fold into a bullet like structure (Velarde et al., 2007). AatD then acrylates the folded protein. This attachment of a fatty acid retains the protein at the inner membrane (Fig. 7.1B). This typically would not be conducive to secretion, however AatP, like its LolE homolog (Mizutani et al., 2013b), can capture this protein out of the membrane (Fig. 7.1C). ATP hydrolysis by AatC powers the system resulting in a release of the acylated Aap/CexE protein into the AatA, AatB pore. Once secreted, Aap/CexE can position itself in the outer leaflet of the outer membrane similar to a lipoprotein (Fig. 7.1D). This modulation and insertion into the outer leaflet of the outer membrane may explain the ciprofloxacin sensitivity observed for strains producing Aap (Mortensen et al., 2013). It might be possible that the insertion of the acylated Aap into the outer membrane disrupts the surface barrier enabling easier access to the cytoplasm.

This proposed mechanism of secretion is not like any other T1SS. The modification mediated by AatD is proposed to be the reason why Aap and CexE are secreted atypically. The structures of Aap (Velarde et al., 2007) and CexE have evolved to so that after this modification the protein can be secreted through AatA. The AatD modulation keeps the Aap and CexE proteins attached to the cell surface. In the mechanism hypothesised unmodulated Aap/CexE protein cannot be secreted. In other T1SS, like HlyBD-TolC, acylation is not required for secretion (Nicaud et al., 1985). Therefore, the acylation of Aap/CexE is a requirement for their function and not just a mechanism for secretion.

7.4. Future investigation into the Aat system

Conformation of AatD mediated acylation is a priority in future investigations of the Aat system. Understanding the role of AatD is critical to determining the mechanism of Aat

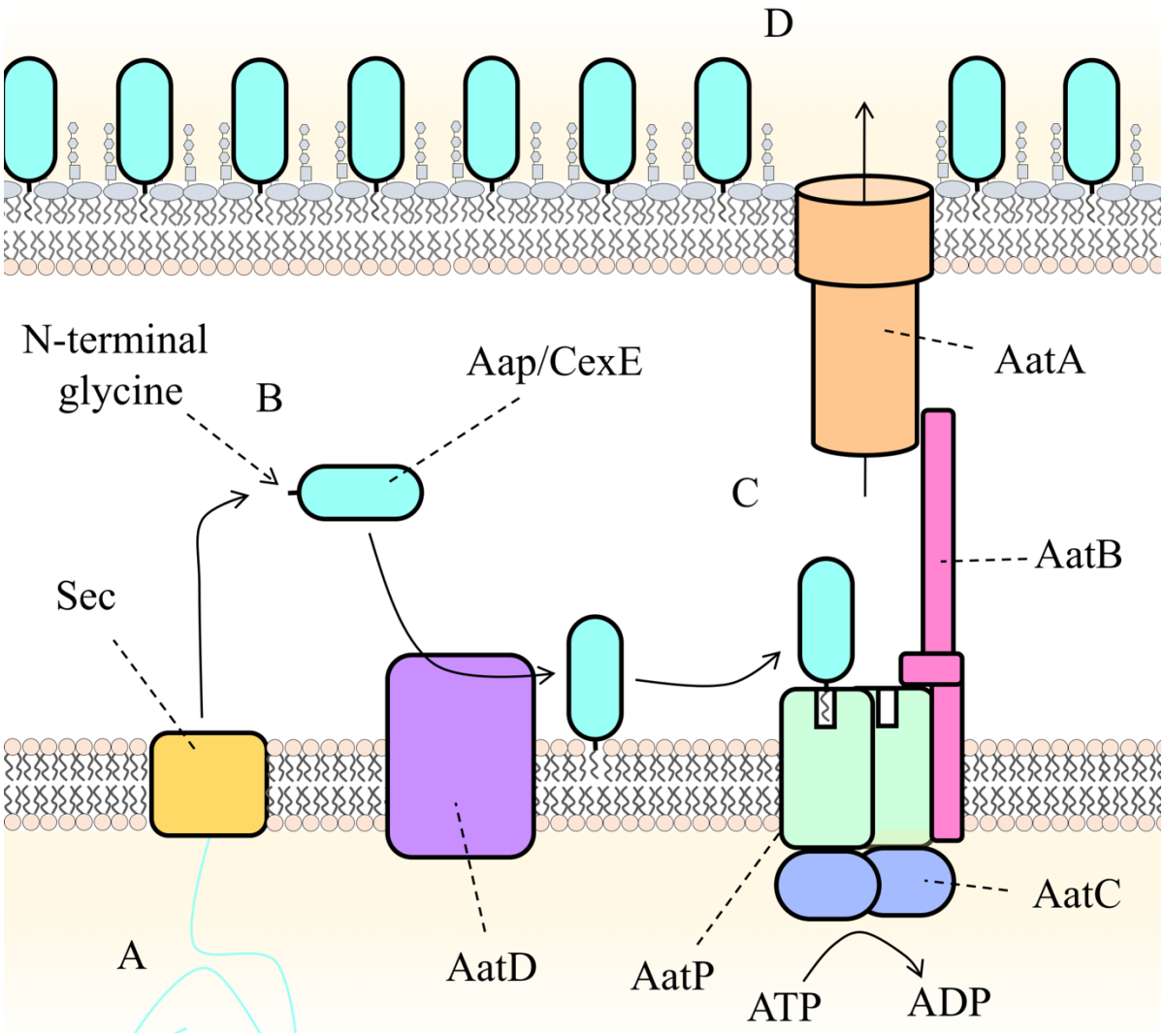


Figure 7.1: Proposed mechanism for the Aat secretion system. The distance between AatD and AatPABC is only for illustrative purposes, as the true interactions are not known. A, the substrate protein is translated in the cytoplasm and secreted by the Sec machinery into the periplasm. AatD transfers a fatty acid onto the N-terminal glycine residue. The fatty acid of the substrate protein is inserted into the inner membrane. C, AatP captures the substrate protein from the inner membrane. The protein is then secreted out of the cell by the AatA pore. D, the fatty acid addition allows it to insert into the outer leaflet of the outer membrane.

secretion. The next steps should be to identify how the disruption of the *aat* genes affects the position of Aap/CexE in the cell. For example, in an *aatD* mutant the Aap/CexE would be expected to be in the periplasm of the cell while in an *aatP* mutant the Aap/CexE would be expected to be predominantly in the inner membrane of the cell. Other experiments into the interactions of the Aat proteins with each other should be considered.

While these proposed experiments would better define the mechanism of Aat secretion there are still many questions remaining. Since the function of AatD is not the only oddity of the Aat system investigation of the other components would likely provide fascinating results. The structure of the Aat proteins is suspected to be different from many in the current databases as there are no other examples of this secretion system or anything similar. AatA would be of particular interest, as if the above speculation is correct the secretion of a hydrophobic region due to AatD acylation via a water filled barrel like TolC is curious. It is known from the bioinformatic analysis that the AatA protein is distinctly separated from the other TolC homologs potentially indicating a different mechanism. Similar to AatA, the AatB protein sequence was in a distinct clade from the other membrane fusion proteins. This is unsurprising as allowing access via the periplasm is unheard of for membrane fusion proteins of T1SS and efflux pumps. Therefore, AatB must have a peculiar operation which is worthy of further investigation.

This work has provided new insights and provided a foundation for further investigation into the Aat secretion system. The AggR virulence regulator was further characterised and its role *aap* expression clarified. A homolog of this highly-expressed gene was *cexE*. The investigation into CexE revealed that these two proteins with a similar structure had opposite functions. Here the first experimental evidence of CexE secretion by the Aat system was provided. Finally, the

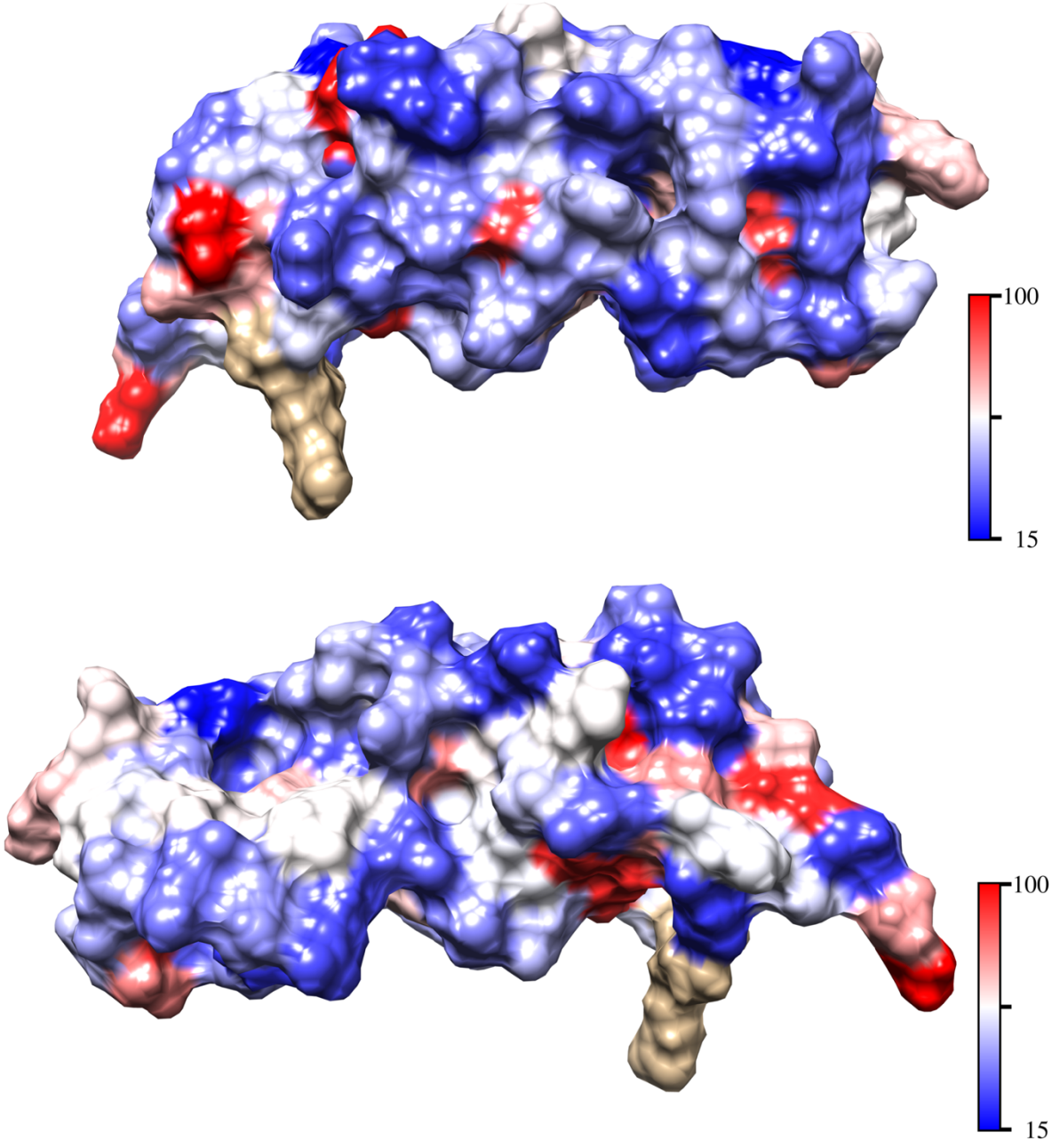
first steps necessary to determine the mechanism of the Aat system were completed and the importance of AatD in understanding Aap/CexE secretion was realised. Aap, CexE, and their respective Aat system are highly-expressed in many pathogenic strains in response to environmental signals that activate virulence. Therefore, understanding the secretion of Aap and CexE from the pathogenic *E. coli* can lead to a reduction of morbidity and mortality worldwide.

Appendix

Appendix I: Nucleotide alignment of *aap* gene from NCTC86 and other identical genes

```
CLUSTAL O(1.2.4) multiple sequence alignment


NCTC86              atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
13.1-R1_2009        atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
ICBEc10_2013        atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
ICBECAM5_2016       atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
HVH_202_2003        atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
BWH49_2014          atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
BWH50_2014          atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
221_2015            atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
ECOR50_2016         atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
E1526_P2017         atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
E390_P2017          atgaaaaaacttaatgtgtcaatttttctattgcggcttgttttagtcttaatgtattt
                    ************************************************************

NCTC86              gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
13.1-R1_2009        gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
ICBEc10_2013        gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
ICBECAM5_2016       gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
HVH_202_2003        gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
BWH49_2014          gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
BWH50_2014          gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
221_2015            gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
ECOR50_2016         gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
E1526_P2017         gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
E390_P2017          gctggtggaggaggatggaattcggatcttgtagatccgcaacaatgtgtaaaattgtca
                    ************************************************************

NCTC86              ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
13.1-R1_2009        ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
ICBEc10_2013        ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
ICBECAM5_2016       ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
HVH_202_2003        ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
BWH49_2014          ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
BWH50_2014          ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
221_2015            ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
ECOR50_2016         ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
E1526_P2017         ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
E390_P2017          ggtgcacagtatacatataatagctcttctaataaatgtatgcaaggtatcaacgaaggg
                    ************************************************************

NCTC86              aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
13.1-R1_2009        aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
ICBEc10_2013        aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
ICBECAM5_2016       aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
HVH_202_2003        aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
BWH49_2014          aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
BWH50_2014          aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
221_2015            aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
ECOR50_2016         aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
E1526_P2017         aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
E390_P2017          aaagtccatggtgtttcattatttggaacattctattatggcgatggcagccaagggact
                    ************************************************************

NCTC86              tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
13.1-R1_2009        tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
ICBEc10_2013        tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
ICBECAM5_2016       tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
HVH_202_2003        tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
BWH49_2014          tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
BWH50_2014          tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
221_2015            tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
ECOR50_2016         tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
E1526_P2017         tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
E390_P2017          tttaaaggaagagtttccccggggacaacactgaatacaaatcaagatatgaacaaaact
                    ************************************************************

NCTC86              aataaatatggcgtgaaatataaggttattactgagtgggttagataa
13.1-R1_2009        aataaatatggcgtgaaatataaggttattactgagtgggttagataa
ICBEc10_2013        aataaatatggcgtgaaatataaggttattactgagtgggttagataa
ICBECAM5_2016       aataaatatggcgtgaaatataaggttattactgagtgggttagataa
HVH_202_2003        aataaatatggcgtgaaatataaggttattactgagtgggttagataa
BWH49_2014          aataaatatggcgtgaaatataaggttattactgagtgggttagataa
BWH50_2014          aataaatatggcgtgaaatataaggttattactgagtgggttagataa
221_2015            aataaatatggcgtgaaatataaggttattactgagtgggttagataa
ECOR50_2016         aataaatatggcgtgaaatataaggttattactgagtgggttagataa
E1526_P2017         aataaatatggcgtgaaatataaggttattactgagtgggttagataa
E390_P2017          aataaatatggcgtgaaatataaggttattactgagtgggttagataa
                    ************************************************
```

Appendix II: Conserved surface residues of the Aap structure



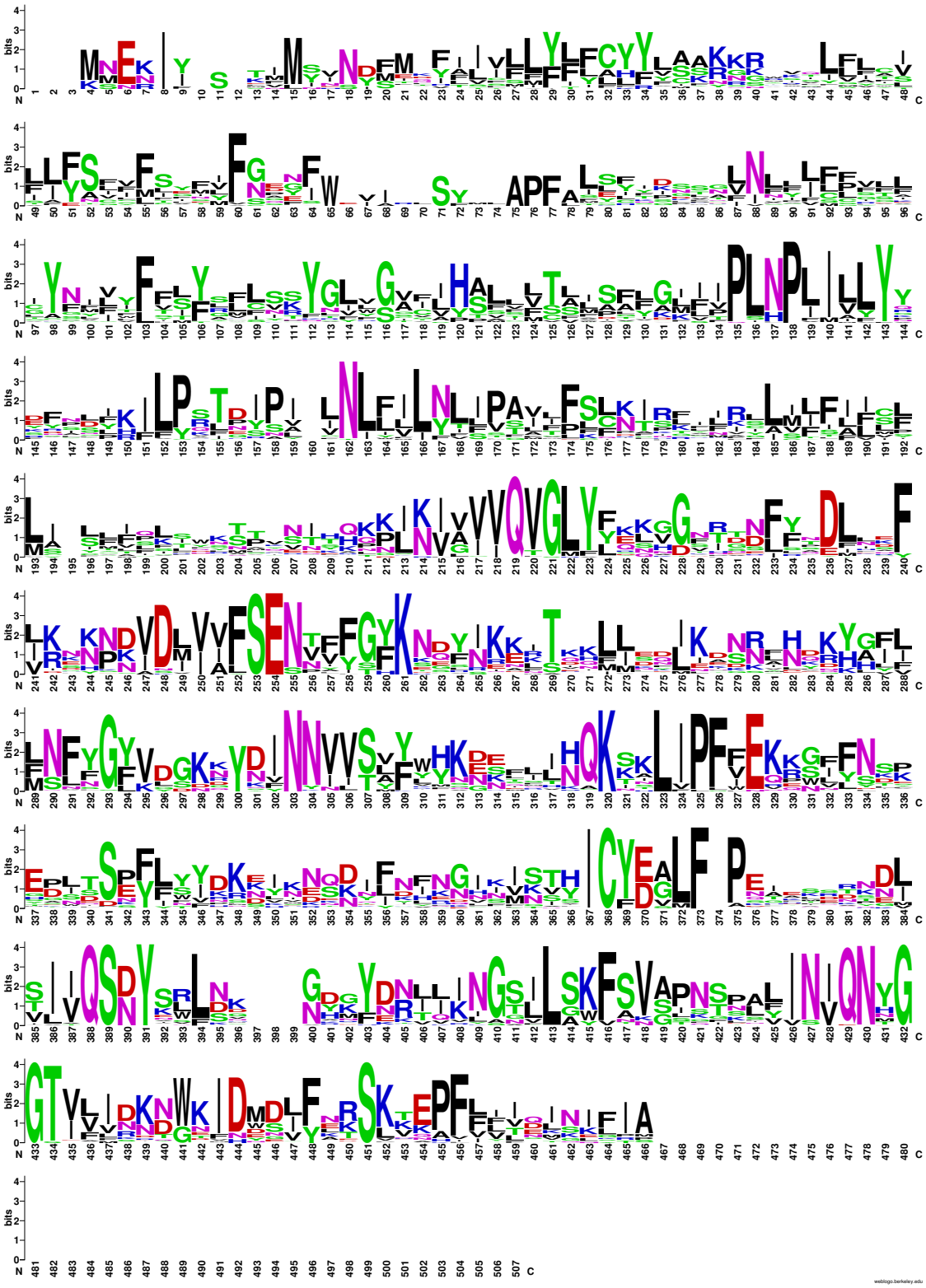
Appendix III: Alignment of AatP to MacB. The AatP sequence from EAEC 042 was aligned to the MacB sequence for *E. coli* MG1655 using Clustal Omega.

```
CLUSTAL O(1.2.4) multiple sequence alignment


MacB      MTPLLELKDIRRSYPAGDEQVEVLKGISLDIYAGEMVAIVGASGSGKSTLMNILGCLDKA
AatP      ------------------------------------------------------------


MacB      TSGTYRVAGQDVATLDADALAQLRREHFGFIFQRYHLLSHLTAEQNVEVPAVYAGLERKQ
AatP      ------------------------------------------------------------


MacB      RLLRAQELLQRLGLEDRTEYYPAQLSGGQQQRVSIARALMNGGQVILADEPTGALDSHSG
AatP      ------------------------------------------------------------


MacB      EEVMAILHQLRDRGHTVIIVTHDPQVAAQAERVIEIRDGEIVRNPPAIEKVNVTGGTEPV
AatP      ------------------------------------------------------------


MacB      VNTVSGWRQFVSGFNEALTMAWRALAANKM-------RTLLTMLGIIIGIASVVSIVVVG
AatP      ---MTTLHY---YLNEALL----NIIENRRQNFAFLVFLSLSFIGIIITDSLIYSVSLKA
             ::  :     :****      :  *:           *:::****  : : *: : .

MacB      DAAKQMVLADIRSIGTNTIDVYPGKDFGDDDPQYQQALKYDDLIAIQKQPWVASATPAVS
AatP      EEE--L---KVHSDKVIFVKLYRPKTVGYITEKFIT---VSKVLSFSKSAFL--------
          :    :   .::*  .  :.:*  * .*    ::      ..:::::.*. ::

MacB      QNLRLRYNNVDVAASANGVSGDYFNVYGMTFSEGNTFNQEQLNGRAQVVVLDSNTRRQLF
AatP      ------------YVSDTPFSGELFSVNGIDKLGLNTEYSGDLNDKYNGNVAIVNESSPFF
                     .* . .**: *.* *:     **  . :**.: :  *   *    :*

MacB      PHKADVVGEVILVGNMPARVIGVAEEKQSMFGSS-------KVLRVWLPYSTMSGRVMGQ
AatP      SK------KQIFINGVPFKIIGVRLNSKTDFLDSLGLKASQSDEHIFIPLETMFKVKLDN
           :      : *::..:* ::***  :.:: * .*       .  ::::* .**    :.:

MacB      SWLNSITVRVKEGFDSAEAEQQLTRLLSLRHGKKDFFTWNMDGVLKTVEKTTRTLQLFLT
AatP      R-VNAVKIFLDNIVTK-RDINNVKRVLY----DNDIRKFDIVTSLNAKEAVDRVLERFSL
            :*::.: :.: . . .  :::.*:*     .:*: .:::   *:: * . *.*: *

MacB      LVA---VISLVVGGIGVMNIMLVSVTERTREIGIRMAVGARASDVLQQFLIEAVL---VC
AatP      LTNSVYVILTLSASVTCFILSKRSFYSRRVELSLKIIHGTEKKEITVLIIIESLIMLSVC
          *.    **  : ..:  : :   *. .*  *:.:::  *:. .::   ::**:::   **

MacB      LVGGAL--GITLSLLIAFTLQLF--LPGWEIGFSPLALLLAFLCSTVTGILFGWLPARNA
AatP      LFISVIYAGVIMH-IIKYFLDVTISIRTTMITISLANVLLVFIS---ANIIFGRL----F
          *. ..:  *: :  :* : *::   :    * :*   :**.*:.   :.*:** *

MacB      ARLDPVDALARE--
AatP      FSINPVNAIKGKIE
            ::**:*:  :
```

Appendix IV: Alignment of AatC to MacB. The sequence of AatC from EACE 042 was aligned with the sequence of MacB from *E. coli* MG1655 using Clustal Omega.

```
CLUSTAL O(1.2.4) multiple sequence alignment


MacB      MTPLLELKDIRRSYPAGDEQVEVLKGISLDIYAGEMVAIVGASGSGKSTLMNILGCLDKA
AatC      ---MIRVKIH----KKPIENRTILNNSTIEIKEGSFNIITGPSGVGKTSLLNIIGLLDNA
             ::.:*          *:  :*:. :::*  *.:  *.* ** **::*:**:* **:*

MacB      TSGTYRVAGQDVATLDADALAQLRREHFGFIFQRYHLLSHLTAEQNVEV----PAVYAGL
AatC      FVGEYELFGKKVEIKDNSITTYIRRKYFGFIFQDSLINVKQNVLRNILCSVDSQNIIAAR
            * *.: *:.*   * .  : :**::******   :  : .. :*:        : *.

MacB      ERKQRLLRAQELLQRLGLEDRTEYYPAQLSGGQQQRVSIARALMNGGQVILADEPTGALD
AatC      ------ERINEVLVSVGLSNIN-NNVSFLSGGEKQRLALARALIKKPSILLADEPTASLD
                 * :*:*  :**.: .    : ****::**:::****::  .::******.:**

MacB      SHSGEEVMAILHQLRDRGHTVIIVTHDPQVAAQAERVIEIRDGEIVRNPPAIEKVNVTGG
AatC      IKNKKLVMNILSEYNNQGGTVVMVTHDLELIDENMTLIQLLNT-----------------
           :. : ** ** : .::* **::**** ::  :   :*:: :

MacB      TEPVVNTVSGWRQFVSGFNEALTMAWRALAANKMRTLLTMLGIIIGIASVVSIVVVGDAA
AatC      ------------------------------------------------------------


MacB      KQMVLADIRSIGTNTIDVYPGKDFGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNL
AatC      ------------------------------------------------------------


MacB      RLRYNNVDVAASANGVSGDYFNVYGMTFSEGNTFNQEQLNGRAQVVVLDSNTRRQLFPHK
AatC      ------------------------------------------------------------


MacB      ADVVGEVILVGNMPARVIGVAEEKQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRV
AatC      ------------------------------------------------------------


MacB      KEGFDSAEAEQQLTRLLSLRHGKKDFFTWNMDGVLKTVEKTTRTLQLFLTLVAVISLVVG
AatC      ------------------------------------------------------------


MacB      GIGVMNIMLVSVTERTREIGIRMAVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAF
AatC      ------------------------------------------------------------


MacB      TLQLFLPGWEIGFSPLALLLAFLCSTVTGILFGWLPARNAARLDPVDALARE
AatC      ---------------------------------------------------
```

Appendix V: Full sequence logo for AatD.



Bibliography

Abreu, A.G., Fraga, T.R., Martínez, A.P.G., et al. (2015) The Serine Protease Pic From Enteroaggregative Escherichia coli Mediates Immune Evasion by the Direct Cleavage of Complement Proteins. **J Infect Dis**, 212 (1): 106–15

Abrusci, P., Mcdowell, M.A., Lea, S.M., et al. (2014) Building a secreting nanomachine : a structural overview of the T3SS. **Curr Opin Struct Biol**, Apr (25): 111–7

Adachi, J.A., Jiang, Z., Mathewson, J.J., et al. (2001) Enteroaggregative Escherichia coli as a Major Etiologic Agent in Traveler's Diarrhea in 3 Regions of the World. **Clin Infect Dis**, 32 (12): 1706–9

Ahmed, D., Islam, M.S., Begum, Y.A., et al. (2012) Presence of enterotoxigenic Escherichia coli in biofilms formed in water containers in poor households coincides with epidemic seasons in Dhaka. **J Appl Microbiol**, 114 (4): 1223–9

Akopian, D., Shen, K., Zhang, X., et al. (2013) Signal Recognition Particle: An essential protein targeting machine. **Annu Rev Biochem**, 82: 693–721

Allesen-Holm, M., Barken, K.B., Yang, L., et al. (2006) A characterization of DNA release in Pseudomonas aeruginosa cultures and biofilms. **Mol Microbiol**, 59 (4): 1114–28

Alva, V., Nam, S., Johannes, S., et al. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure. **Nucleic Acids Res**, 44: W410-5

Ames, G., Mimura, C., Holbrook, S., et al. (1992) Traffic ATPases: A Superfamily of Transport Proteins Operating from Escherichia coli to Humans. **Adv Enzymol Relat Areas Mol Biol**, 65: 1–47

Amro, N.A., Kotra, L.P., Wadu-mesthrige, K., et al. (2000) High-Resolution Atomic Force Microscopy Studies of the Escherichia coli Outer Membrane : Structural Basis for Permeability. **Langmuir**, 16 (6): 2789–96

Andersen, C., Koronakis, E., Bokma, E., et al. (2002) Transition to the open state of the TolC periplasmic tunnel entrance. **Proc Natl Acad Sci USA**, 99 (17): 11103–8

Anderson, G.G., Palermo, J.J., Schilling, J.D., et al. (2003) Intracellular Bacterial Biofilm-Like Pods in Urinary Tract Infections. **Science**, 301 (5629): 105–7

Babić, A., Lindner, A.B., Vulić, M., et al. (2008) Direct Visualization of Horizontal Gene Transfer. **Science**, 319 (5869): 1533–7

Backert, S. and Meyer, T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. **Curr Opin Microbiol**, 9 (2): 207–17

Balakrishnan, L., Hughes, C. and Koronakis, V. (2001) Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by Escherichia coli. **J Mol Biol**, 313 (3): 501–10

Basler, M., Pilhofer, M., Henderson, P.G., et al. (2013) Type VI secretion requires a dynamic contractile phage tail-like structure. **Nature**, 483 (7388): 182–6

Baumann, U., Wu, S., Flaherty, K.M., et al. (1993) Three-dimensional structure of the alkaline protease of calcium binding parallel beta roll motif. **EMBO J**, 12 (9): 3357–64

Baumgarten, T., Sperling, S., Seifert, J., et al. (2012) Membrane vesicle formation as a multiple-stress response mechanism enhances pseudomonas putida DOT-T1E cell surface hydrophobicity and biofilm formation. **Appl Environ Microbiol**, 78 (17): 6217–24

Bax, A. and Ikura, M. (1991) An efficient 3D NMR technique for correlating the proton and15N backbone amide resonances with the α-carbon of the preceding residue in uniformly 15N/13C enriched proteins. **J Biomol NMR**, 1 (1): 99–104

Berg, B. Van Den, Jr, W.M.C., Collinson, I., et al. (2004) X-ray structure of a protein-conducting channel. **Nature**, 427 (6969): 36–44

Berry, A.A., Yang, Y., Pakharukova, N., et al. (2014) Structural Insight into Host Recognition by Aggregative Adherence Fimbriae of Enteroaggregative Escherichia coli. **PLoS Pathog**, 10

(9): e1004404
Boardman, B.K. and Satchell, K.J.F. (2004) Vibrio cholerae strains with mutations in an atypical trpe I secretion system accumulated RTX intracellularly. **J Bacteriol**, 186 (23): 8137–43
Bodero, M.D., Pilonieta, M.C. and Munson, G.P. (2007) Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic Escherichia coli. **J Bacteriol**, 189 (5): 1627–32
Bolger, A.M., Lohse, M. and Usadel, B. (2014) Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data. **Bioinformatics**, 30 (15): 2114–20
Boll, E.J., Struve, C., Boisen, N., et al. (2013) Role of Enteroaggregative Escherichia coli Virulence Factors in Uropathogenesis. **Infect Immun**, 81 (4): 1164–71
Borges-Walmsley, M.I., Beauchamp, J., Kelly, S.M., et al. (2003) Identification of Oligomerization and Drug-binding Domains of the Membrane Fusion Protein EmrA. **J Biol Chem**, 278 (15): 12903–12
Bourgeois, A.L., Wierzba, T.F. and Walker, R.I. (2016) Status of vaccine research and development for enterotoxigenic Escherichia coli. **Vaccine**, 34 (26): 2880–6
Braun, V. and Wolff, H. (1970) The Murein-Lipoprotein Linkage in the Cell Wall of Escherichia coli. **Eur J Biochem**, 14 (2): 387–91
Browning, D.F., Wells, T.J., França, F.L.S., et al. (2013) Laboratory adapted Escherichia coli K-12 becomes a pathogen of Caenorhabditis elegans upon restoration of O antigen biosynthesis. **Mol Microbiol**, 87 (5): 939–50
Buchrieser, C., Glaser, P., Nedjari, H., et al. (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. **Mol Microbiol**, 38 (4): 760–71
Bundock, P., Dulk-ras, A. Den, Beijersbergen, A., et al. (1995) Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae. **EMBO J**, 14 (13): 3206–14
Canizalez-Roman, A. and Navarro-García, F. (2003) Fodrin CaM-binding domain cleavage by Pet from enteroaggregative Escherichia coli leads to actin cytoskeletal disruption. **Mol Microbiol**, 48 (4): 947–58
Cavanagh, J., Skelton, N.J., Fairbrother, W.J., et al. (2006) **Protein NMR Spectroscopy – Principles and Practice**. Academic Press
Chaudhuri, R.R., Sebaihia, M., Hobman, J.L., et al. (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative Escherichia coli strain 042. **PLoS ONE**, 5 (1): e8801
Chenal, A., In, J., Raynal, B., et al. (2009) RTX Calcium Binding Motifs Are Intrinsically Disordered in the Absence of Calcium. **J Biol Chem**, 284 (3): 1781–9
Clements, A., Young, J.C., Constantinou, N., et al. (2012) Infection strategies of enteric pathogenic Escherichia coli. **Gut Microbes**, 3 (2): 71–87
Cloonan, N. and Grimmond, S.M. (2008) Transcriptome content and dynamics at single-nucleotide resolution. **Genome Biol**, 9 (9): 234
Colmer, J.A., Fralick, J.A. and Hamood, A.N. (1998) Isolation and characterization of a putative multidrug resistance pump from Vibrio cholerae. **Mol Microbiol**, 27 (1): 63–72
Coster, T.S., Wolf, M.K., Hall, E.R., et al. (2007) Immune Response , Ciprofloxacin Activity , and Gender Differences after Human Experimental Challenge by Two Strains of Enterotoxigenic Escherichia coli. **Infect Immun**, 75 (1): 252–9
Cowles, C.E., Li, Y., Semmelhack, M.F., et al. (2012) The free and bound forms of Lpp occupy distinct subcellular locations in Escherichia coli. **Mol Microbiol**, 79 (5): 1168–81
Crossman, L.C., Chaudhuri, R.R., Beatson, S.A., et al. (2010) A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. **J Bacteriol**, 192 (21): 5822–31
Croxen, M. a and Finlay, B.B. (2010) Molecular mechanisms of Escherichia coli pathogenicity.

Nat Rev Microbiol, 8 (1): 26–38
Czeczulin, J.R., Balepur, S., Hicks, S., et al. (1997) Aggregative Adherence Fimbria II , a Second Fimbrial Antigen Mediating Aggregative Adherence in Enteroaggregative Escherichia coli. **Infect Immun**, 65 (10): 4135–45
Dalbey, R.E., Kuhn, A., Zhu, L., et al. (2014) The membrane insertase YidC. **Biochim Biophys Acta**, 1843 (8): 1489–96
Datsenko, K. a and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. **Proc Natl Acad Sci USA**, 97 (12): 6640–5
Davidson, A.L., Dassa, E., Orelle, C., et al. (2008) Structure , Function , and Evolution of Bacterial ATP-Binding Cassette Systems. **Microbiol Mol Biol Rev**, 72 (2): 317–64
Debarbieux, L. and Wandersman, C. (2001) Folded HasA inhibits its own secretion through its ABC exporter. **EMBO J**, 20 (17): 4657–63
Delcour, A.H. (2009) Outer Membrane Permeability and Antibiotic Resistance. **Biochim Biophys Acta.**, 1794 (5): 808–16
Delepelaire, P. (2004) Type I secretion in gram-negative bacteria. **Biochim Biophys Acta**, 1694 (1–3): 149–161
Demers, J., Habenstein, B., Loquet, A., et al. (2014) High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. **Nat Commun**, 5: 4976
Devaraj, A., Justice, S.S., Bakaletz, L.O., et al. (2015) DNABII proteins play a central role in UPEC biofilm structure. **Molecular Microbiology**, 96 (6): 1119–35
Dhakal, B.K. and Mulvey, M.A. (2013) The UPEC Pore Forming Toxin α-Hemolysin Triggers Proteolysis of Host Proteins to Disrupt Cell Adhesion, Inflammatory and Survival Pathways. **Cell Host Microbe**, 11 (1): 58–69
Diderichsen, B. (1980) flu, A metastable gene controlling surface properties of Escherichia coli. **J Bacteriol.**, 141 (2): 858–67
Dinh, T., Paulsen, I.T. and Saier, M.H. (1994) A Family of Extracytoplasmic Proteins That Allow Transport of Large Molecules across the Outer Membranes of Gram-Negative Bacteria. **J Bacteriol**, 176 (13): 3825–31
Du, D., Wang, Z., James, N.R., et al. (2014) Structure of the AcrAB-TolC multidrug efflux pump. **Nature**, 509 (7501): 512–5
Dudley, E.G., Thomson, N.R., Parkhill, J., et al. (2006) Proteomic and microarray characterization of the AggR regulon identifies a pheU pathogenicity island in enteroaggregative Escherichia coli. **Mol Microbiol**, 61 (5): 1267–82
Dunne, K.A., Chaudhuri, R.R., Rossiter, A.E., et al. (2017) Sequencing a piece of history : complete genome sequence of the original Escherichia coli strain. **Microb Genom**, 3 (3): mgen000106
Duquesne, S., Destoumieux-garz, D., Peduzzi, J., et al. (2007) Microcins , gene-encoded antibacterial peptides from enterobacteria. **Nat Prod Rep**, 24 (4): 708–34
Egan, S.M. (2002) Growing Repertoire of AraC / XylS Activators. **J Bacteriol**, 184 (20): 5529–32
Elias, W.P., Czeczulin, J.R., Henderson, I.R., et al. (1999) Organization of biogenesis genes for aggregative adherence fimbria II defines a virulence gene cluster in enteroaggregative Escherichia coli. **J Bacteriol**, 181 (6): 1779–85
Eswaran, J., Hughes, C. and Koronakis, V. (2003) Locking TolC Entrance Helices to Prevent Protein Translocation by the Bacterial Type I Export Apparatus. **J Mol Biol**, 327 (2): 309–15
Eto, D.S., Jones, T.A., Sundsbak, J.L., et al. (2007) Integrin-Mediated Host Cell Invasion by Type 1–Piliated Uropathogenic Escherichia coli. **PLoS Pathog**, 3 (7): e100
Evans, D.G., Satterwhite, T.K., Evans, D.J., et al. (1978) Differences in Serological Responses and Excretion Patterns of Volunteers Challenged with Enterotoxigenic Escherichia coli with

and without the Colonization Factor Antigen. **Infect Immun**, 19 (3): 883–8
Evans, D.J. and Evans, D.G. (1973) Three Characteristics Associated with Enterotoxigenic Escherichia coli Isolated from Man. **Infect Immun**, 8 (3): 322–8
Fan, E., Chauhan, N., Udatha, D.B.R.K.G., et al. (2016) Type V Secretion Systems in Bacteria. **Microbiol Spectr**, 4 (1)
Fath, M.J., Zhang, L.H., Rush, J., et al. (1994) Purification and Characterization of Colicin V from Escherichia coli Culture Supernatants. **Biochemisty**, 33 (22): 6911–7
Federici, L., Du, D., Walas, F., et al. (2005) The Crystal Structure of the Outer Membrane Protein VceC from the Bacterial Pathogen Vibrio cholerae at 1.8 Å Resolution VceC. **J Biol Chem**, 280 (15): 15307–14
Felmlee, T., Pellett, S. and Welch, R.A. (1985) Nucleotide Sequence of an Escherichia coli Chromosomal Hemolysin. **J Bacteriol**, 163 (1): 94–105
Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. **Nucleic Acids Res**, 39: W29-37
Fitzpatrick, A.W.P., Llabrés, S., Neuberger, A., et al. (2017) Structure of the MacAB–TolC ABC-type tripartite multidrug ef flux pump. **Nat Microbiol**, 2: 17070
Flemming, H.-C., Wingender, J., Szewzyk, U., et al. (2016) Biofilms: an emergent form of bacterial life. **Nat Rev Microbiol**, 14 (9): 563–75
Flemming, H.C. and Wingender, J. (2010) The biofilm matrix. **Nat Rev Microbiol**, 8 (9): 623–33
Fronzes, R., Christie, P.J. and Waksman, G. (2009) The structural biology of type IV secretion systems. **Nat Rev Microbiol**, 7 (10): 703–14
Gallegos, M.T., Michán, C. and Ramos, J.L. (1993) The XylS/AraC family of regulators. **Nucleic Acids Res**, 21 (4): 807–10
Gallegos, M.T., Schleif, R., Bairoch, A., et al. (1997) Arac/XylS family of transcriptional regulators. **Microbiol Mol Biol Rev**, 61 (4): 393–410
Gaston, K., Bell, A., Kolb, A., et al. (1990) Stringent Spacing Requirements for Transcription Activation by CRP. **Cell**, 62 (4): 733–43
Gavin, H.E. and Satchell, K.J.F. (2015) MARTX toxins as effector delivery platforms. **Pathog Dis**, 73 (9): ftv092
Gerlach, R.G., Cláudio, N., Rohde, M., et al. (2008) Cooperation of Salmonella pathogenicity islands 1 and 4 is required to breach epithelial barriers. **Cell Microbiol**, 10 (11): 2364–76
Giannoukos, G., Ciulla, D.M., Huang, K., et al. (2012) Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. **Genome Biol**, 13 (3): r23
Gödeke, J., Heun, M., Bubendorfer, S., et al. (2011) Roles of two Shewanella oneidensis MR-1 extracellular endonucleases. **Appl Environ Microbiol**, 77 (15): 5342–51
Gogala, M., Becker, T., Beatrix, B., et al. (2014) Structures of the Sec61 complex engaged in nascent peptide translocation ormembrane insertion. **Nature**, 506 (7486): 107–10
Green, E.R. and Mecsas, J. (2016) Bacterial Secretion Systems : An Overview. **Microbiol Spectr**, 4 (1)
Griessl, M.H., Schmid, B., Kassler, K., et al. (2013) Structural Insight into the Giant Ca2+-Binding Adhesin SiiE : Implications for the Adhesion of Salmonella enterica to Polarized Epithelial Cells. **Structure**, 21 (5): 741–52
Grohmann, E. and Espinosa, M. (2003) Conjugative Plasmid Transfer in Gram-Positive Bacteria. **Microbiol Mol Biol Rev**, 67 (2): 277–301
Grzesiek, S., Angluster, J. and Bax, A. (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic mixing of 13C Magnetiszation. **J Magn Reson**, B (101): 114–9
Grzesiek, S. and Bax, A.D. (1992) Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein. **J Magn Reson**, 96: 432–40

Guttman, J.A., Li, Y., Wickham, M.E., et al. (2006) Attaching and effacing pathogen-induced tight junction disruption in vivo. **Cell Microbiol**, 8 (4): 634–45
Guzman, L., Belin, D., Carson, M.J., et al. (1995) Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose PBAD Promoter. **J Bacteriol**, 177 (14): 4121–30
Hakansson, S., Bergman, T., Vanooteghem, J., et al. (1993) YopB and YopD Constitute a Novel Class of Yersinia Yop Proteins. **Infect Immun**, 61 (1): 71–80
Hale, T.L. (1991) Genetic Basis of Virulence in Shigella Species. **Microbiol Rev**, 55 (2): 206–224
Hamilton, H.L. and Dillard, J.P. (2006) Natural transformation of Neisseria gonorrhoeae : from DNA donation to homologous recombination. **Mol Microbiol**, 59 (2): 376–85
Hamilton, H.L., Domínguez, N.M., Schwartz, K.J., et al. (2005) Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system. **Mol Microbiol**, 55 (6): 1704–21
Harris, R.K. (1983) **Nuclear Magnetic Resonance Spectroscopy**. Pitman, London
Havarstein, L.S., Diep, D.B. and Nes, I.F. (1995) A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. **Mol Microbiol**, 16 (2): 229–40
Haycocks, J.R.J. and Grainger, D.C. (2016) Unusually situated binding sites for bacterial transcription factors can have hidden functionality. **PLoS ONE**, 11 (6): 1–11
Heijne, G. Von (1989) The structure of signal peptides from bacterial lipoproteins. **Protein Eng**, 2 (7): 531–4
Henderson, I.A.N.R., Czeczulin, J., Eslava, C., et al. (1999) Characterization of Pic , a Secreted Protease of Shigella flexneri and Enteroaggregative Escherichia coli. **Infect Immun**, 67 (11): 5587–96
Henderson, I.R., Navarro-Garcia, F., Desvaux, M., et al. (2004) Type V Protein Secretion Pathway : the Autotransporter Story. **Microbiol Mol Biol Rev**, 68 (4): 692–744
Hicks, S., Candy, D.C.A. and Phillips A. D. (1996) Adhesion of Enteroaggregative Escherichia coli to Pediatric Intestinal Mucosa In Vitro. **Infect Immun**, 64 (11): 4751–60
Hirai, K., Aoyama, H., Irikura, T., et al. (1986) Differences in Susceptibility to Quinolones of Outer Membrane Mutants of Salmonella typhimurium and Escherichia coli. **Antimicrob Agents Chemother**, 29 (3): 535–8
Hodson, C., Yang, J., Hocking, D.M., et al. (2017) Control of Virulence Gene Expression by the Master Regulator , CfaD , in the Prototypical Enterotoxigenic Escherichia coli Strain , H10407. **Front Microbiol**, 11 (8): 1525
Holland, I.B., Peherstorfer, S., Kanonenberg, K., et al. (2016) Type I Protein Secretion-Deceptively Simple yet with a Wide Range of Mechanistic Variability across the Family. **EcoSal Plus**, 7 (1)
Hood, R.D., Singh, P., Hsu, F., et al. (2010) A Type VI Secretion System of Pseudomonas aeruginosa Targets a Toxin to Bacteria. **Cell Host Microbe**, 7 (1): 25–37
Huang, D.B., Mohanty, A., DuPont, H.L., et al. (2006) A review of an emerging enteric pathogen: enteroaggregative Escherichia coli. **J Med Microbiol**, 55 (10): 1303–11
Hwang, J., Zhong, X. and Tai, P.C. (1997) Interactions of Dedicated Export Membrane Proteins of the Colicin V Secretion System: CvaA, a Member of the Membrane Fusion Protein Family, Interacts with CvaB and TolC proteins . **J Bacteriol**, 179 (20): 6264–70
Imuta, N., Nishi, J., Tokuda, K., et al. (2008) The Escherichia coli efflux pump TolC promotes aggregation of enteroaggregative E. coli 042. **Infect Immun**, 76 (3): 1247–56
Ito, Y., Matsuzawa, H., Matsuyama, S., et al. (2006) Genetic Analysis of the Mode of Interplay between an ATPase Subunit and Membrane Subunits of the Lipoprotein-Releasing ATP-Binding Cassette Transporter LolCDE. **J Bacteriol**, 188 (8): 2856–64
Iwashita, M., Nishi, J., Wakimoto, N., et al. (2006) Role of the carboxy-terminal region of the

outer membrane protein AatA in the export of dispersin from enteroaggregative Escherichia coli. **FEMS Microbiol Lett**, 256 (2): 266–72
Jackowski, S. (1986) Transfer of Fatty Acids from the 1-Position of Phosphatidylethanolamine to the Major Outer Membrane Lipoprotein of. **J Biol Chem**, 261 (24): 11328–33
Jakubovics, N.S., Shields, R.C., Rajarajan, N., et al. (2013) Life after death: The critical role of extracellular DNA in microbial biofilms. **Lett Appl Microbiol.**, 57 (6): 467–75
Jennings, L.K., Storek, K.M., Ledvina, H.E., et al. (2015) Pel is a cationic exopolysaccharide that cross-links extracellular DNA in the *Pseudomonas aeruginosa* biofilm matrix. **Proceedings of the National Academy of Sciences**, 112 (36): 11353–11358
Jensen, B.H., Olsen, K.E.P., Struve, C., et al. (2014) Epidemiology and Clinical Manifestations of Enteroaggregative Escherichia coli. **Clin Microbiol Rev**, 27 (3): 614–30
Justice, S.S., Hung, C., Theriot, J.A., et al. (2004) Differentiation and developmental pathways of uropathogenic Escherichia coli in urinary tract pathogenesis. **Proc Natl Acad Sci USA**, 101 (5): 1333–8
Justice, S.S., Hunstad, D.A., Seed, P.C., et al. (2006) Filamentation by Escherichia coli subverts innate defenses during urinary tract infection. **Proc Natl Acad Sci USA**, 103 (52): 19884–9
Kane, M.D., Jatkoe, T. a, Stumpf, C.R., et al. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. **Nucleic Acids Res**, 28 (22): 4552–7
Kanonenberg, K., Schwarz, C.K.W. and Schmitt, L. (2013) Type I secretion systems - a story of appendices. **Res Microbiol**, 164 (6): 596–604
Kelly, S.M., Jess, T.J. and Price, N.C. (2005) How to study proteins by circular dichroism. **Biochim Biophys Acta**, 1751 (2): 119–39
Kenny, B., Devinney, R., Stein, M., et al. (1997) Enteropathogenic E . coli (EPEC) Transfers Its Receptor for Intimate Adherence into Mammalian Cells. **Cell**, 91 (4): 511–20
Kim, J.-S., Jeong, H., Song, S., et al. (2015) Structure of the tripartite multidrug efflux pump AcrAB-TolC suggests an alternative assembly mode. **Mol Cells**, 38 (2): 180–6
Kim, J., Song, S., Lee, M., et al. (2016) Crystal Structure of a Soluble Fragment of the Membrane Fusion Protein HlyD in a Type I Secretion System of Gram-Negative Bacteria. **Structure**, 24 (3): 477–85
Knutton, S., Bain, C., Phillips, A.D., et al. (1997) Down regulation of intimin expression during attaching and effacing enteropathogenic Escherichia coli adhesion. **Infect Immun**, 65 (5): 1644–1652
Korobkova, E., Emonet, T. and Vilar, J.M.G. (2004) From molecular noise to behavioural variability in a single bacterium. **Nature**, 428 (6982): 574–8
Koronakis, V., Eswaran, J. and Hughes, C. (2004) Structure and Function of TolC: The Bacterial Exit Duct for Proteins and Drugs. **Annu. Rev. Biochem.**, 73: 467–89
Koronakis, V., Koronakis, E. and Hughes, C. (1989) Isolation and analysis of the C-terminal signal directing export of Escherichia coli hemolysin protein across both bacterial membranes. **EMBO J**, 8 (2): 595–605
Koronakis, V., Sharff, a, Koronakis, E., et al. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. **Nature**, 405 (6789): 914–9
Korotkov, K. V, Sandkvist, M. and Hol, W.G.J. (2012) The type II secretion system: biogenesis, molecular architecture and mechanism. **Nat Rev Microbiol**, 10 (5): 336–51
Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., et al. (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. **Lancet**, 382 (9888): 209–22
Krishnamoorthy, G., Tikhonova, E.B., Dhamdhere, G., et al. (2013) On the role of TolC in multidrug efflux: the function and assembly of AcrAB-TolC tolerate significant depletion of

intracellular TolC protein. **Mol Microbiol**, 87 (5): 982–97
Kulathila, R., Kulathila, R., Indic, M., et al. (2011) Crystal Structure of Escherichia coli CusC , the Outer Membrane Component of a Heavy Metal Efflux Pump. **PLoS One**, 6 (1): e15610
Kumar, P., Luo, Q., Vickers, T.J., et al. (2014) EatA, an Immunogenic Protective Antigen of Enterotoxigenic Escherichia coli, Degrades Intestinal Mucin. **Infect Immun**, 82 (2): 500–8
Lecher, J., Schwarz, C.K.W., Stoldt, M., et al. (2012) An RTX transporter tethers its unfolded substrate during secretion via a unique N-terminal domain. **Structure**, 20 (10): 1778–87
Lecker, S., Schiebel, E., Hendrick, J.F., et al. (1990) The Binding Cascade of SecB to SecA to SecY/E Mediates Preprotein Targeting to the E . coli Plasma Membrane. **Cell**, 63 (2): 269–79
Lee, D.J., Bingle, L.E.H., Heurlier, K., et al. (2009) Gene doctoring: a method for recombineering in laboratory and pathogenic Escherichia coli strains. **BMC Microbiol**, 9: 252
Lee, M., Jun, S., Yoon, B., et al. (2012) Membrane Fusion Proteins of Type I Secretion System and Tripartite Efflux Pumps Share a Binding Motif for TolC in Gram-Negative Bacteria. **PLOS One**, 7 (7): e40460
Lee, W., Tonelli, M. and Markley, J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. **Bioinformatics**, 31 (8): 1325–7
Lenders, M.H.H., Weidtkamp-peters, S., Kleinschrodt, D., et al. (2015) Directionality of substrate translocation of the hemolysin A Type I secretion system. **Sci Rep**, 5: 12470
Leo, J.C., Heinz, N., Autenrieth, I.B., et al. (2012) Intimin and Invasin Export Their C-Terminus to the Bacterial Cell Surface Using an Inverse Mechanism Compared to Classical Autotransport. **PLoS ONE**, 7 (10): e47069
Letoffe, S., Ghigo, J.M. and Wandersman, C. (1994) Secretion of the Serratia marcescens HasA Protein by an ABC Transporter. **J Bacteriol**, 176 (17): 5372–77
Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3 : an online tool for the display and annotation of phylogenetic and other trees. **Nucleic Acids Res**, 44 (W1): W242-5
Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows – Wheeler transform. **Bioinformatics**, 26 (5): 589–95
Li, H., Handsaker, B., Wysoker, A., et al. (2009) The Sequence Alignment / Map format and SAMtools. **Bioinformatics**, 25 (16): 2078–9
Li, W. and Godzik, A. (2006) Cd-hit : a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**, 22 (13): 1658–9
Liao, Y., Smyth, G.K. and Shi, W. (2014) Sequence analysis featureCounts : an efficient general purpose program for assigning sequence reads to genomic features. **Bioin**, 30 (7): 923–30
Lin, W., Fullner, K., Clayton, R., et al. (1999) Identification of a Vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. **Proc Natl Acad Sci USA**, 96 (3): 1071–6
Linke, D., Riess, T., Autenrieth, I.B., et al. (2006) Trimeric autotransporter adhesins: variable structure, common function. **Trends Microbiol**, 14 (6): 264–70
Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biol**, 15 (12): 550
Lucchini, S., Rowley, G., Goldberg, M.D., et al. (2006) H-NS Mediates the Silencing of Laterally Acquired Genes in Bacteria. **PLoS Pathog**, 2 (8): e81
Ludi, S., Frey, J., Favre, D., et al. (2008) Morphological and immunocytochemical analysis of Escherichia coli-specific surface antigens in wildtype strains and in recombinant Vibrio cholerae. **Antonie Van Leeuwenhoek**, 93 (1–2): 185–92
Luirink, J. and Sinning, I. (2004) SRP-mediated protein targeting : structure and function revisited. **Biochim Biophys Acta**, 1694 (1–3): 17–35
Ma, C., Wickham, M.E., Guttman, J.A., et al. (2006) Citrobacter rodentium infection causes both mitochondrial dysfunction and intestinal epithelial barrier disruption in vivo : role of

mitochondrial associated protein (Map). **Cell Microbiol**, 8 (10): 1669–86
Mahon, V., Smyth, C.J. and Smith, S.G.J. (2010) Mutagenesis of the Rns regulator of enterotoxigenic Escherichia coli reveals roles for a linker sequence and two helix–turn–helix motifs. **Microbiology**, 156 (Pt 9): 2796–806
Majdalani, N., Heck, M., Stout, V., et al. (2005) Role of RcsF in Signaling to the Rcs Phosphorelay Pathway in Escherichia coli. **J Bacteriol**, 187 (19): 6770–8
Malinverni, J.C., Werner, J., Kim, S., et al. (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. **Mol Microbiol**, 61 (1): 151–64
Marion, D., Kay, L.E., Sparks, S.W., et al. (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. **JACS**, 111 (4): 1515–1517
Martinez, J.J., Mulvey, M.A., Schilling, J.D., et al. (2000) Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. **EMBO J**, 19 (12): 2803–12
McDaniel, T.K., Jarvis, K.G., Donnenbergt, M.S., et al. (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. **Proc Natl Acad Sci USA**, 92 (5): 1664–8
Micsonai, A., Wien, F., Kernya, L., et al. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. **Proc Natl Acad Sci USA**, 112 (24): E3095-103
Mizutani, M., Mukaiyama, K., Xiao, J., et al. (2013a) Functional differentiation of structurally similar membrane subunits of the ABC transporter LolCDE complex. **FEBS Letters**, 587 (1): 23–29
Mizutani, M., Mukaiyama, K., Xiao, J., et al. (2013b) Functional differentiation of structurally similar membrane subunits of the ABC transporter LolCDE complex. **FEBS Lett**, 587 (1): 23–9
Modali, S.D. and Zgurskaya, H.I. (2011) The periplasmic membrane proximal domain of MacA acts as a switch in stimulation of ATP hydrolysis by MacB transporter. **Mol Microbiol**, 81 (4): 937–51
Monlezun, L., Phan, G., Benabdelhak, H., et al. (2015) New OprM structure highlighting the nature of the N-terminal anchor. **Front Microbiol**, 6: 667
Morgan, E., Campbell, J.D., Rowe, S.C., et al. (2004) Identification of host-specific colonization factors of Salmonella enterica serovar Typhimurium. **Mol Microbiol**, 54 (4): 994–1010
Morin, N., Santiago, A.E., Ernst, R.K., et al. (2013) Characterization of the AggR regulon in enteroaggregative Escherichia coli. **Infect Immun**, 81 (1): 122–32
Morin, N., Tirling, C., Ivison, S.M., et al. (2010) Autoactivation of the AggR regulator of enteroaggregative Escherichia coli in vitro and in vivo. **FEMS Immunol Med Microbiol**, 58 (3): 344–55
Mortensen, N.P., Boisen, N., Carey, S., et al. (2013) Enteroaggregative Escherichia coli: surface protein dispersin increases bacterial uptake of ciprofloxacin. **Int J Antimicrob Agents**, 42 (5): 462–5
Mortensen, N.P., Fowlkes, J.D., Maggart, M., et al. (2011) Effects of sub-minimum inhibitory concentrations of ciprofloxacin on enteroaggregative Escherichia coli and the role of the surface protein dispersin. **Int J Antimicrob Agents**, 38 (1): 27–34
Muhandiram, D.R. and Kay, L.E. (1994) Gradient-Enchanced Triple-Resonace Three-Dimensioanl NMR Experiments with Improved Snesitivity. **J Magn Reson**, B (103): 203–16
Murray, A.W., Opin, C., Biol, C., et al. (1998) Supramolecular Structure of the Salmonella typhimurium Type III Protein Secretion System. **Science**, 280 (5363): 602–5
Nada, R.A., Armstrong, A., Shaheen, H.I., et al. (2013) Phenotypic and genotypic characterization of enterotoxigenic Escherichia coli isolated from U.S. military personnel

participating in Operation Bright Star, Egypt, from 2005 to 2009. **Diagn Microbiol Infect Dis**, 76 (3): 272–7
Nakayama, H., Kurokawa, K. and Lee, B.L. (2012) Lipoproteins in bacteria: structures and biosynthetic pathways. **FEBS J**, 279 (23): 4247–68
Nataro, J., Baldini, M., Kaper, J., et al. (1985) Detection of an adherence factor of enteropathogenic Escherichia coli with a DNA probe. **J Infect Dis**, 152 (3): 560–5
Nataro, J., Deng, Y., Maneval, D., et al. (1992) Aggregative adherence fimbriae I of enteroaggregative Escherichia coli mediate adherence to HEp-2 cells and hemagglutination of human erythrocytes. **Infect Immun**, 60 (6): 2297–304
Nataro, J., Yikang, D., Cookson, S., et al. (1995) Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers. **J Infect Dis**, 171 (2): 465–8
Nataro, J.P. (2003) "Enteroaggregative Escherichia coli." In W. M. Scheld, J. M. Hughes, and B.E.M. (ed.). **Emerging infections 6. ASM Press, Washington, D.C., in press. 2003**. Emerging infections 6. ASM Press, Washington, D.C., in press.
Nataro, J.P. and Kaper, J.B. (1998) Diarrheagenic Escherichia coli. **Clin Microbiol Rev**, 11 (1): 142–201
Nataro, J.P., Kaper, J.B., Robins-Browne, R., et al. (1987) Patterns of adherence of diarrheagenic Escherichia coli to HEp-2 cells. **Pediatr Infect Dis J**, 6 (9): 829–31
Nicaud, J., Mackman, N., Gray, L., et al. (1985) Characterisation of HlyC and mechanism of activation and secretion of haemolysin from E . coli 2001. **FEBS Lett**, 187 (2): 339–44
Nishi, J., Sheikh, J., Mizuguchi, K., et al. (2003) The export of coat protein from enteroaggregative Escherichia coli by a specific ATP-binding cassette transporter system. **J Biol Chem**, 278 (46): 45680–9
Noinaj, N., Kuszak, A.J., Gumbart, J.C., et al. (2013) Structural insight into the biogenesis of β-barrel membrane proteins. **Nature**, 501 (7467): 385–90
Okeke, I.N., Lamikanra, A., Czeczulin, J., et al. (2000) Heterogeneous Virulence of Enteroaggregative Escherichia coli Strains Isolated from Children in Southwest Nigeria. **J Infect Dis**, 181 (1): 252–60
Okuda, S. and Tokuda, H. (2009) Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. **Proc Natl Acad Sci USA**, 106 (14): 5877–82
Olsson, J., Cherepanov, P., Holmstro, A., et al. (2001) LcrV is a channel size-determining component of the Yop effector translocon of Yersinia. **Mol Microbiol**, 39 (3): 620–32
Orfanoudaki, G. and Economou, A. (2014) Proteome-wide Subcellular Topologies of E . coli Polypeptides Database (STEPdb). **Mol Cell Proteomics**, 13 (12): 3674–87
Owen, P., Meehan, M., Loughry-doherty, H. De, et al. (1996) Phase-variable outer membrane proteins in Escherichia coli. **FEMS Immunol Med Microbiol**, 16 (2): 63–76
Paetzel, M., Karla, A., Strynadka, N.C.J., et al. (2002) Signal Peptidases. **Chem Rev**, 102 (12): 4549–80
Papanikou, E., Karamanou, S. and Economou, A. (2007) Bacterial protein secretion through the translocase nanomachine. **Nat Rev Microbiol**, 5 (11): 839–51
Petty, N.K., Bulgin, R., Crepin, V.F., et al. (2010) The Citrobacter rodentium genome sequence reveals convergent evolution with human pathogenic Escherichia coli. **J Bacteriol**, 192 (2): 525–38
Picking, W.L., Nishioka, H., Hearn, P.D., et al. (2005) IpaD of Shigella flexneri Is Independently Required for Regulation of Ipa Protein Secretion and Efficient Insertion of IpaB and IpaC into Host Membranes. **Infect Immun**, 73 (3): 1432–40
Pilonieta, M.C., Bodero, M.D. and Munson, G.P. (2007) CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic Escherichia coli. **J Bacteriol**, 189 (14): 5060–7

Pinchuk, G.E., Ammons, C., Culley, D.E., et al. (2008) Utilization of DNA as a sole source of phosphorus, carbon, and energy by Shewanella spp.: Ecological and physiological implications for dissimilatory metal reduction. **Appl Environ Microbiol**, 74 (4): 1198–208
Pukatzki, S., Ma, A.T., Revel, A.T., et al. (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. **Proc Natl Acad Sci USA**, 104 (39): 15508–13
Raetz, C.R.H. and Dowhan, W. (1990) Biosynthesis and Function of Phospholipids in Escherichia coli. **J Biol Chem**, 265 (3): 1235–18
Randall, L.L. and Hardy, S.J.S. (2002) Cellular and Molecular Life Sciences SecB , one small chaperone in the complex milieu of the cell. **Cell Mol Life Sci**, 59 (10): 1617–23
Robinson, C.M., Sinclair, J.F., Smith, M.J., et al. (2006) Shiga toxin of enterohemorrhagic Escherichia coli type O157 : H7 promotes intestinal colonization. **Proc Natl Acad Sci USA**, 103 (25): 9667–72
Rohde, H., Qin, J., Cui, Y., et al. (2011) Open-Source Genomic Analysis of Shiga-Toxin–Producing E. coli O104:H4. **N Engl J Med**, 365 (8): 718–24
Rossiter, A.E., Browning, D.F., Leyton, D.L., et al. (2011a) Transcription of the plasmid-encoded toxin gene from Enteroaggregative Escherichia coli is regulated by a novel co-activation mechanism involving CRP and Fis. **Mol Microbiol**, 81 (1): 179–91
Rossiter, A.E., Leyton, D.L., Tveen-jensen, K., et al. (2011b) The Essential β-Barrel Assembly Machinery Complex Components BamD and BamA Are Required for Autotransporter Biogenesis. **J Bacteriol**, 193 (16): 4250–3
Rost, B. (1999) Twilight zone of protein sequence alignments. **Protein Eng**, 12 (2): 85–94
Roy, K., Bartels, S., Qadri, F., et al. (2010) Enterotoxigenic Escherichia coli elicits immune responses to multiple surface proteins. **Infect Immun**, 78 (7): 3027–35
Roy, K., Hamilton, D.J., Munson, G.P., et al. (2011) Outer membrane vesicles induce immune responses to virulence proteins and protect against colonization by enterotoxigenic Escherichia coli. **Clin Vaccine Immunol**, 18 (11): 1803–8
Roy, K., Hilliard, G., Hamilton, D., et al. (2009) Enterotoxigenic Escherichia coli EtpA mediates adhesion between flagella and host cells. **Nature**, 457 (7229): 594–8
Ruiz-perez, F., Henderson, I.R., Leyton, D.L., et al. (2009) Roles of Periplasmic Chaperone Proteins in the Biogenesis of Serine Protease Autotransporters of Enterobacteriaceae. **J Bacteriol**, 191 (21): 6571–83
Russell, A.B., Peterson, S.B. and Mougous, J.D. (2014) Type VI secretion system effectors : poisons with a purpose. **Nat Rev Microbiol**, 12 (2): 137–148
Sahl, J.W., Steinsland, H., Redman, J.C., et al. (2011) A comparative genomic analysis of diverse clonal types of enterotoxigenic Escherichia coli reveals pathovar-specific conservation. **Infect Immun**, 79 (2): 950–60
Saldana, Z., Erdem, L., Schu, S., et al. (2009) The Escherichia coli Common Pilus and the Bundle-Forming Pilus Act in Concert during the Formation of Localized Adherence by Enteropathogenic E. coli. **J Bacteriol**, 191 (11): 3451–61
Samba-louaka, A., Nougayrède, J., Watrin, C., et al. (2008) Bacterial cyclomodulin Cif blocks the host cell cycle by stabilizing the cyclin-dependent kinase inhibitors p21 waf1 and p27 kip1. **Cell Microbiol**, 10 (12): 2496–508
Sankaran, K. and Wu, H.C. (1994) Lipid Modification of Bacterial Prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. **J Biol Chem**, 269 (31): 19701–6
Santiago, A.E., Ruiz-perez, F., Jo, N.Y., et al. (2014) A Large Family of Antivirulence Regulators Modulates the Effects of Transcriptional Activators in Gram-negative Pathogenic Bacteria. **PLoS Pathog**, 10 (5): e1004153
Santiago, A.E., Yan, M.B., Tran, M., et al. (2016) A large family of anti-activators accompanying XylS/AraC family regulatory proteins. **Mol Microbiol**, 101 (2): 314–32

Sapriel, G. and Delepelaire, P. (2003) The SecB Chaperone Is Bifunctional in Serratia marcescens : SecB Is Involved in the Sec Pathway and Required for HasA Secretion by the ABC Transporter. **J Bacteriol**, 185 (1): 80–8
Satchell, K.J.F. (2011) Structure and Function of MARTX Toxins and Other Large Repetitive RTX Proteins. **Annu Rev Microbiol**, 65: 71–90
Schägger, H. (2006) Tricine-SDS-PAGE. **Nat Protoc**, 1 (1): 16–22
Schroeder, G.N. and Hilbi, H. (2008) Molecular Pathogenesis of Shigella spp.: Controlling Host Cell Signaling, Invasion, and Death by Type III Secretion. **Clin Microbiol Rev**, 21 (1): 134–56
Schulze, R.J., Komar, J., Botte, M., et al. (2014) Membrane protein insertion and proton-motive- force-dependent secretion through the bacterial holo-translocon SecYEG–SecDF–YajC–YidC. **Proc Natl Acad Sci USA**, 111 (13): 4844–9
Sears, C.L. and Kaper, J.B. (1996) Enteric Bacterial Toxins : Mechanisms of Action and Linkage to Intestinal Secretion. **Microbiol Rev**, 60 (1): 167–215
Selkrig, J., Mosbahi, K., Webb, C.T., et al. (2012) Discovery of an archetypal protein transport system in bacterial outer membranes. **Nat Struct Mol Biol**, 19 (5): 506–10
Sheikh, J., Czeczulin, J.R., Harrington, S., et al. (2002) A novel dispersin protein in enteroaggregative Escherichia coli. **J Clin Invest**, 110 (9): 1329–37
Sheikh, J., Hicks, S., Dall'Agnol, M., et al. (2001) Roles for Fis and YafK in biofilm formation by enteroaggregative Escherichia coli. **Mol Microbiol**, 41 (5): 983–97
Sievers, F., Wilm, A., Dineen, D., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Mol Syst Biol**, 7: 539
Sklar, J.G., Wu, T., Kahne, D., et al. (2007) Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. **Genes Dev**, 21 (19): 2473–84
Spudich, J. and Koshland, D.J. (1976) Non-genetic individuality: chance in the single cell. **Nature**, 262 (5568): 467–71
Srikhanta, Y.N., Hocking, D.M., Wakefield, M.J., et al. (2013) Control of bacterial virulence by the RalR regulator of the rabbit-specific enteropathogenic escherichia coli strain E22. **Infect Immun**, 81 (11): 4232–43
Stamatakis, A. (2014) RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. **Bioinformatics**, 30 (9): 1312–3
Stanley, P., Packman, L.C., Koronakis, V., et al. (1994) Fatty Acylation of Two Internal Lysine Residues Required for the Toxic Activity of Escherichia coli Hemolysin. **Science**, 266 (5193): 1992–6
Steffen, R., Castelli, F., Nothdurft, H.D., et al. (2005) Vaccination against Enterotoxigenic Escherichia coli, a Cause of Travelers' Diarrhea. **J Travel Med**, (12): 102–107
Su, C., Radhakrishnan, A., Kumar, N., et al. (2014) Crystal structure of the Campylobacter jejuni CmeC outer membrane channel. **Protein Sci**, 23 (7): 954–61
Swidsinski, A., Weber, J., Loening-baucke, V., et al. (2005) Spatial Organization and Composition of the Mucosal Flora in Patients with Inflammatory Bowel Disease. **J Clin Microbiol**, 43 (7): 3380–9
Thanabalu, T., Koronakis, E., Hughes, C., et al. (1998) Substrate-induced assembly of a contiguous channel for protein export from E . coli: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. **EMBO J**, 17 (22): 6487–96
Tikhonova, E.B. and Zgurskaya, H.I. (2004) AcrA , AcrB , and TolC of Escherichia coli Form a Stable Intermembrane Multidrug Efflux Complex. **J Biol Chem**, 279 (31): 32116–24
Torriani, a (1960) Influence of inorganic phosphate in the formation of phosphatases by Escherichia coli. **Biochim Biophys Acta**, 38 (1957): 460–469
Toshima, H., Yoshimura, A., Arikawa, K., et al. (2007) Enhancement of Shiga Toxin Production in Enterohemorrhagic Escherichia coli Serotype O157: H7 by DNase Colicins.

Appl Environ Microbiol, 73 (23): 7582–88
Trapnell, C., Williams, B.A., Pertea, G., et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nat Biotechnol**, 28 (5): 516–20
Ulrich, E.L., Akutsu, H., Doreleijers, J.F., et al. (2008) BioMagResBank. **Nucleic Acids Res**, 36: 402–8
Velarde, J.J., Varney, K.M., Inman, K.G., et al. (2007) Solution structure of the novel dispersin protein of enteroaggregative Escherichia coli. **Mol Microbiol**, 66 (5): 1123–35
Voulhoux, R., Bos, M.P., Geurtsen, J., et al. (2003) Role of a Highly Conserved Bacterial Protein in Outer Membrane Protein Assembly. **Science**, 299: 262–6
Walker, J.E., Saraste, M., Runswick, M.J., et al. (1982) Distantly related sequences in the alpha- and beta- subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. **EMBO J**, I (8): 945–51
Wang, R., Seror, S.J., Blight, M., et al. (1991) Analysis of the Membrane Organization of an Escherichia coli Protein Translocator , HlyB , a Member of a Large Family of Prokaryote and Eukaryote Surface Transport Proteins. **J Mol Biol**, 217 (3): 441–54
Wang, Z., Fan, G., Hryc, C.F., et al. (2017) An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. **Elife**, 6: e24905
Webb, C., Selkrig, J., Perry, A., et al. (2013) Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. **J Mol Biol**, 422 (4): 545–55
Welch, R.A. (1991) Pore-forming cytolysins of Gram-negative bacteria. **Mol Microbiol**, 5 (3): 521–8
Whitney, E.N. (1971) The tolC Locus in Escherichia coli K12. **Genetics**, 67 (1): 39–53
Wiles, S., Clare, S., Harker, J., et al. (2004) Organ specificity, colonization and clearance dynamics in vivo following oral challenges with the murine pathogen Citrobacter rodentium. **Cell Microbiol**, 6 (10): 963–72
Wiles, T.J., Dhakal, B.K., Eto, D.S., et al. (2008) Inactivation of Host Akt / Protein Kinase B Signaling by Bacterial Pore-forming Toxins. **Mol Biol Cell**, 19 (4): 1427–38
Wille, T., Wagner, C., Mittelstädt, W., et al. (2014) SiiA and SiiB are novel type I secretion system subunits controlling SPI4-mediated adhesion of Salmonella enterica. **Cell Microbiol**, 16 (2): 161–78
Wirth, T., Falush, D., Lan, R., et al. (2006) Sex and virulence in Escherichia coli : an evolutionary. **Mol Microbiol**, 60 (5): 1136–51
Wu, T., Malinverni, J., Ruiz, N., et al. (2005) Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in Escherichia coli. **Cell**, 121 (2): 235–245
Yakushi, T., Masuda, K., Narita, S., et al. (2000) A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. **Nat Cell Biol**, 2 (4): 212–8
Yamaguchi, K. (1988) A Single Amino Acid Determinant of the Membrane Localization of Lipoproteins in E. coli. **Cell**, 53 (3): 423–32
Yasuda, M., Iguchi-yokoyama, A., Matsuyama, S., et al. (2009) Membrane Topology and Functional Importance of the Periplasmic Region of ABC Transporter LolCDE of ABC Transporter LolCDE. **Biosci Biotechnol Biochem**, 73 (10): 2310–6