Recombinant production of selected autotransporters from *Proteus mirabilis* for biochemical and biophysical studies

Liam Ryan

Supervisor: Dr. Tewfik Soulimane

Submitted for completion of MSc October 2012
1. Introduction and Literature review .............................................. 1
  1.1. Introduction..................................................................................... 1
  1.2. Literature review............................................................................ 3
    1.2.1 Bacterial protein secretion : An overview .................................. 3
    1.2.2 Secretion in Gram-negative bacteria........................................ 3
    1.2.3 Transport across the cytoplasm.................................................. 4
      1.2.3.1 The Sec pathway................................................................. 4
      1.2.3.1.1 The twin arginine pathway (Tat)............................ 4
      1.2.3.2 Signal recognition particle pathway................................... 5
    1.2.4 Secretion pathways ..................................................................... 6
      1.2.4.1 Type I system..................................................................... 6
      1.2.4.2 Type II system................................................................... 6
      1.2.4.3 Type III system.................................................................. 7
      1.2.4.4 Type IV system.................................................................. 7
      1.2.4.5 Type VI system.................................................................. 8
    1.2.5 Other systems of translocation................................................... 8
      1.2.5.1 Type VII system................................................................. 8
    1.2.6.1 Type V secretion system....................................................... 9
    1.2.6.2 Type Va.............................................................................. 10
    1.2.6.3 Type Vb ( two partner secretion TPS).................................... 10
    1.2.6.4 Oligomeric coiled-coil Adhesins. Type Vc............................ 12
      1.2.6.4.1 Trimeric Autotransporters............................................. 12
      1.2.6.4.2 Translocation of TAA................................................... 13
      1.2.6.4.3 Trimeric binding properties.......................................... 14
    1.2.7 The beta Domain........................................................................ 15
      1.2.7.1 Overview........................................................................... 15
      1.2.7.2 Mechanism of translocation............................................... 16
        1.2.7.2.1 The Hairpin model....................................................... 16
        1.2.7.2.2 Bam A model............................................................. 18
        1.2.7.2.3 Hexameric model....................................................... 19
        1.2.7.2.4 Passenger domain cleavage..................................... 21
1.2.8 Accessory Proteins for translocation .............................................. 23
  1.2.8.1 Overview ................................................................. 23
1.2.9 Families of autotransporters .......................................................... 24
  1.2.9.1 Ssp ................................................................. 24
  1.2.9.2 SphB1 ............................................................... 24
  1.2.9.1.3 AspA/NapP ...................................................... 25
1.2.9.2 *Helicobacter pylori* autotransporters ....................................... 25
  1.2.9.2.1 Overview ........................................................... 25
  1.2.9.2.2 VacA ............................................................ 26
1.2.9.3 Serine protease autotransporters ........................................ 27
  1.2.9.3.1 IgA1 protease .................................................. 27
  1.2.9.3.2 Hap ............................................................... 28
  1.2.9.3.3 SPATE .......................................................... 29
  1.2.9.3.4 Pic ................................................................. 30
1.2.9.4 AIDA autotransporter family ........................................... 30
  1.2.9.4.1 AIDA ............................................................. 30
  1.2.9.4.2 Antigen 43 ...................................................... 31
  1.2.9.4.3 Taah and AipA .................................................. 32
1.2.10.1 Overview of autotransporter structure ........................... 33
1.2.10.2 Extraction of membrane proteins ................................... 34
1.2.10.3 Inclusion bodies ............................................................. 34
1.2.10.4 Factors influencing refolding ........................................ 35

2. Production and characterisation of His-tagged full length autotransporters
2.1 Objectives .......................................................................................... 37
2.2 Materials and Methods ........................................................................ 37
  2.2.1 Reagents .......................................................................................... 37
    2.2.1.1 Bacterial strains ............................................................ 37
    2.1.2 Preparation of competent cells ............................................. 37
    2.2.1.3 Genomic DNA preparation ........................................... 38
    2.2.1.4 PCR amplification of DNA from *P. mirabilis* ............... 39
    2.2.1.5 Restriction Digestion of DNA ........................................ 39
    2.2.1.6 DNA Ligation ............................................................. 41
    2.2.1.7 Transformation ............................................................... 41
    2.2.1.8 Plasmid screening ............................................................ 42
2.2.2 Protein expression ............................................................................... 43
   2.2.2.1 Expression of recombinant protein in E. coli.......................... 43
   2.2.2.2 SDS-PAGE ........................................................................... 44
   2.2.2.3 Affinity Chromatography................................................... 44

2.3 Results and discussion
   2.3.1 Selection of autotransporters from P. mirabilis ....................... 45
      2.3.1.1 Cloning of targets........................................................... 46
      2.3.1.2 Expression Studies ......................................................... 48
      2.3.1.3 Purification of full length autotransporters .................... 49

2.4 Conclusion .......................................................................................... 51

3. Production and characterisation of His-tagged autotransporter passenger domains
   3.1 Objectives .......................................................................................... 52
   3.2 Materials and methods ..................................................................... 52
      3.2.2 PCR amplification of DNA from P. mirabilis ..................... 52
      3.2.3 Protein expression ................................................................. 54
         3.2.3.1 Expression of recombinant protein in E. coli............. 55
         3.2.3.2 SDS-PAGE and Western blot .................................... 55
         3.2.3.3 Soluble/Insoluble preparation ................................. 55
         3.2.3.4 Preparation of inclusion bodies ............................... 56
         3.2.3.5 Preparation of periplasmic proteins ....................... 56
         3.2.3.6 Purification of membrane proteins ......................... 57
      3.2.4 Protein Characterisation .......................................................... 57
         3.2.4.1 Acid Phosphatase assay ............................................. 57
         3.2.4.2 Affinity chromatography ......................................... 58
         3.2.4.3 Size exclusion chromatography ............................... 58
         3.2.4.4 Anion exchange chromatography ............................ 58
         3.2.4.5 Circular Dichroism ..................................................... 59
         3.2.4.6 Trp Fluorescence measurents ................................. 59
   3.3 Results and discussion ..................................................................... 60
      3.3.1 Introduction............................................................................. 60
      3.3.2 Cloning and Expression...................................................... 60
      3.3.3 Analysis of expressed protein .......................................... 62
3.3.4 Refolding of 2575p and 0844p .............................. 63
3.3.5 Purification of 2575p ............................................. 64
  3.3.5.1 Anion exchange chromatography .................. 64
  3.3.5.2 Upscaling of the production of 2575p ......... 66
  3.3.5.3 Analysis of refolding ................................. 67
  3.3.5.4 Circular dichroism ..................................... 68
  3.3.5.5 Molecular weight determination 2575p .......... 70
3.3.6 Refolding of 0844p ........................................... 71
  3.3.6.1 Purification of 0844p ................................. 73
3.4 Cloning and expression using p15 ara ................................. 74
  3.4.1 Cloning and expression of 2575 into p15 vector .... 75
  3.4.2 Cloning and expression of 2575p into p15 vector .... 75
    3.4.2.1 Optimisation of 2575p expression ............ 76
    3.4.2.2 Molecular weight determination ............... 78
    3.4.2.3 Refolding of 2575p on solid support ........... 78
    3.4.2.4 Trypsin digest .................................. 80
    3.4.2.6 Trp Fluorescence measurements ............... 80
3.5 Future work ..................................................... 83
3.6 Summary ......................................................... 84

4. Bibliography ........................................................................... 86
Acknowledgements

I am very grateful to a lot of people without the help of whom I would not have been able to submit this thesis.

Firstly, I would like to thank Dr. Tewfik Soulimane for all his help and support over the course of my work and for offering me the opportunity to come and work in his lab.

I would also like to thank Dr. Sylvain Robin for his guidance, help in designing experiments and time throughout the course of this project and generally being a sounding board for any ideas I had.

Thanks also to Dr. Jakki Cooney for her advice on crystallisation and for the use of her laboratory equipment. Thank you also to Professor Tony Pembroke and the Materials and Surface Science Institute for the unlimited use of their Labs and equipment.

Thanks to all the gang from the Soulimane, Cooney and Pembroke groups past and present; Sarah, Magdalena, David, Brighid, Marguerita, Olga, Patricia, Roibeard, Maurice, Mary, Elizabeth, Gosia and Sarah.

I would also like to thank the support staff from the CES department and the MSSI for all their help.

Thanks also to IRCSET for providing financial support.

And finally thanks to all my friends and family for all their support and putting up with me.
Abstract

Autotransporters are a family of outer membrane proteins in Gram-negative bacteria. This family of proteins has a conserved architecture consisting of the membrane domain and the extracellular domain which contains the effector function of the protein. The membrane domain anchors the protein to the cellular surface and has a pore-like structure. The extracellular domain is connected to the anchor domain by means of an α-helix, which permits translocation of the extracellular domain. This family of proteins are referred to as autotransporters due to their ability to be transported to the extracellular surface of the cell without the need of external factors such as ATP, ion gradients or other proteins suggesting that the membrane domain is solely responsible for translocation across the outer membrane.

Although highly conserved, autotransporters have numerous functionalities and are primarily involved in cell adhesion. The membrane domain is highly conserved across all bacterial species, however the extracellular domain has been found to vary greatly and have a diverse range of activities such as; protease, lipase, adhesin, phosphatase and haemolysin. Pathogenic Gram-negative bacteria contain typically several of these proteins to attach to host cells and evade an immune response. *Proteus mirabilis*, a causative agent in urinary tract infections of patients with indwelling catheters, is one such pathogenic bacteria.

This project will investigate a selection of autotransporters involved in virulence of *P. mirabilis* as outlined by Alamuri and Mobley 2008. The conditions needed for expression of five predicted autotransporters from *P. mirabilis* were investigated, specifically: 0844 an acid phosphatase, 2575 an adhesin, 2341 a serine protease, 2174 an adhesin and 2126 a serine protease. Expression and purification of 2575, as well as the passenger domains of 0844 and 2575, was successful, enabling characterisation of both of these proteins using a variety of techniques such as circular dichroism, florescence spectroscopy as well as activity assays in the case of the passenger domain of 0844.
1. Introduction and Literature review

1.1 Introduction

During infection, disease-causing bacteria secrete a variety of proteins that interact with the host organism to establish or maintain the pathogen in host tissues. These extracellular proteins are collectively called virulence factors, due to their function in promoting infection. Virulence factors may be enzymes, such as proteases that break down host proteins, toxins that cause damage to the host by interfering with some vital functions or adhesins that mediate the attachment of the pathogen to host cells and tissues. Virulence factors are often essential for the pathogenesis of disease-causing bacteria and so make attractive targets for specific antimicrobial drugs. For drug development, understanding how virulence factors are secreted and how they interact with the host are of central importance. Elucidation of the atomic structures of virulence factors, the proteins of the secretion pathways and the target molecules of the host gives valuable information about the mechanisms of secretion and the function of virulence factors. This would permit the rational design of inhibitor molecules that could either obstruct the interaction of a virulence factor with its target in the host or prevent its secretion from the bacterial cell.

Autotransporters are a family of virulence factors and consists of monomeric and trimeric autotransporters. Trimeric autotransporters comprise a group of virulence-related proteins in Gram-negative bacteria. These obligate homotrimeric proteins are found in the outer membrane (OM), where their main function is to act as adhesins. In addition to their primary adhesive activity, many trimeric autotransporters have secondary functions such as autoagglutination and protecting the bacterium from host immune responses (Grosskinsky et al., 2006). Trimeric autotransporters share a highly conserved C-terminal region, the translocation unit. In addition to acting as an integral membrane anchor, the translocation unit translocates the N-terminal portion, the “passenger” domain, of the protein into the extracellular space. This translocation event appears to be independent of any auxiliary factors such as adenosine triphosphate (ATP), ion gradients or even other proteins; hence the name autotransporter (Henderson et al., 2004). As these proteins are involved in virulence, and as they all seem to be translocated in a similar manner, the C-terminal translocation unit is a potential target for drugs that would inhibit the translocation across the outer membrane.
The aim of this project is the investigation of suitable autotransporters from *Proteus mirabilis* with a view for their use in structural studies. Although only involved in less than ten percent of normal urinary tract infections, *Proteus mirabilis* is a Gram-negative human uropathogen which causes complicated urinary tract infections (Alamuri and Mobley 2008). Infection may lead to the formation of kidney or bladder stones, catheter blockages, bacteraemia, fever and acute pyelonephritis.

*P. mirabilis* HI4320 was first isolated from the urine of a Nursing Home patient with a long term indwelling catheter. Unusually, it is not only able to degrade urine but analysis of its genome suggests it may also be able to synthesise it (Alamuri and Mobley 2008). Additionally analysis of the genome of *P. mirabilis* HI4320 found it encodes 17 different types of fimbriae signifying the importance of adhesion to this bacterium (Mobley 2008) as this is more than any other bacterium sequenced to date (Alamuri and Mobley 2008). Analysis of the genomic sequence of *P. mirabilis* provided six targets which were predicted to be involved in virulence. The primary objective of this project was to clone and express these genes and use the purified full length protein for characterisation. The focus of the project shifted to full length 2575 protein as this was the only protein that was found to cause the direct virulence effect and was capable of being produced in a soluble form. When sufficient quantities of this protein could not be attained, the focus then shifted to the passenger domain of this protein which was produced initially by refolding inclusion bodies and then in a soluble form using an arabinose promoter. The passenger domain of 0844 was also refolded and produced in an active form. The aim of this project therefore shifted to the production of both the passenger domains of 2575 and 0844 protein for use in characterisation studies and crystallisation with the ambition of solving the protein structure.
1.2 Literature Review

1.2.1 Bacterial protein secretion: An overview

In order to survive, bacteria must interact with their environment by secreting proteins into the extracellular space. These proteins have numerous functions such as cell-to-cell communication, sequestering of minerals important for protein function and virulence factors needed for bacterial infection by pathogenic Gram-negative bacteria. Virulence factors include adhensins, proteases, lipases, haemolysins and phosphatases which combine to evade the host cell response and invade the cell (Henderson and Nataro 2001).

In this section a brief outline will be given of the bacterial protein secretion mainly in Gram-negative bacteria with a particular focus on autotransporters as this project is a study of these proteins from *P. mirabilis*.

1.2.2 Secretion in Gram-negative bacteria.

Unlike other cells, Gram-negative bacteria contain an extra membrane that imbues the bacterial cell with a number of functions which assist the cell in evading the host defence mechanisms and allow the cell to endure harsh environmental conditions. Peptidoglycan in the outer membrane prevents osmotic cell lysis. The outer membrane is also semi-permeable and allows the retention of certain enzymes and prevents the entry of substances which are toxic to the cell such as penicillin G and lysozyme. Outer membranes can also produce vesicles containing toxins which allow communication with other bacterial cells while lipopolysaccharides provide strength to the structure (Bos *et al.*, 2007).

Although the additional membrane provides advantages it causes additional problems in the secretion of proteins. The periplasmic space is devoid of high energy transfer molecules such as ATP, which would be normally used in the translocation process, making it particularly difficult. To overcome these energetically unfavourable biological reactions, Gram-negative bacteria have evolved six different mechanisms that allow proteins to cross the energy deficient periplasm and either traverse or integrate into the outer membrane.
1.2.3 Transport across the cytoplasmic membrane.

1.2.3.1 The Sec pathway

The first step of protein secretion in Gram-negative bacteria is traversing the inner membrane, which maintains its integrity while allowing the transport of molecules. Most bacterial proteins are transported in an unfolded state across the inner membrane via a small gated channel. The most common mechanism for this process is the Sec system of translocation, which is formed by the highly conserved conserved proteins SecE and SecY interacting with the cytoplasmic membrane protein SecG (Scott and Barnett, 2006). The channel is gated by the protein SecY and contains the proteins SecD, SecF. Targeting of OMP proteins to the membrane channel can occur through the action of SecB, which binds proteins with a consensus motif consisting of more than twenty amino acids, SecB also acts as a chaperone and prevents folding of the protein prior to export (Scott and Barnett, 2006). Subsequently, secreted proteins are recognised by the presence of an N-terminal signal sequence which is cleaved by type I signal peptidase during the process of secretion though the inner membrane. By comparison, the signal sequence required for transport through Gram-positive cell membranes is much longer than that required for transport through the Gram-negative inner membrane.

1.2.3.1.1 The twin arginine pathway (Tat)

The twin arginine translocation pathway (Tat) is fundamentally different to any other pathway required to secrete protein across the cytoplasmic membrane. Firstly, the energy used to drive protein translocation is provided via a proton pump. It has been also been shown that the signal sequence required for translocation of proteins by this system has a distinct twin arginine motif which is located at the border of the N-terminal domain and the hydrophobic region of the signal sequence. This characteristic of the Tat signal sequence is different to that of Sec dependant pathways even though both pathways have a similar overall structure. This pathway also permits the passage of fully folded proteins (Schaerlaekens et al, 2004).
Homologues of genes of this pathway have been found in archaea, bacteria, chloroplasts, and cell mitochondria. The Tat pathway was first discovered in the early 1990s when it was discovered that a set of polypeptides could be translocated with the hydrolysis of ATP. The Tat system consists of three integral membrane protein families TatA, TatB and TatC. The components of this system can vary in bacteria with TatB not being required in some species for membrane transport while certain Gram positive bacteria contain multiple copies of TatA and TatC for the formation of multiple Tat channel for substrate specific export. The TatC protein is the most conserved of these proteins while TatA and TatB are similar in terms of structure and sequence making it difficult to distinguish between their respective homologues in some species (Lee et al., 2006).

The utilisation of this pathway of this pathway varies greatly across the organisms in which it is contained, with some only using it to transport a handful of proteins while others such as Streptomyces coelicolor encode 129 proteins which may be secreted via this system (Li et al., 2005). Bacterial secretion by this method is essential for bacterial infection in plants and animals. As mammals do not contain a system which is homologous to the Tat system, it has been suggested that inhibiting the Tat system of translocation may lead to a new generation of antimicrobial pharmaceuticals (Marschall et al., 2011). This pathway may also have useful biotechnological applications as most proteins produced utilise the Sec pathway which is not compatible with some proteins as it requires that proteins be unfolded while crossing the inner membrane. Instead, proteins could be targeted to the Tat secretion pathway which may improve the screening and display of protein libraries in the search for novel proteins and receptors (Lee et al., 2006).

1.2.3.2 Signal recognition particle pathway.

Another main system of targeting bacterial proteins to the inner membrane is via the signal recognition particle pathway (Scott and Barnett 2006). GTPase, along with small cytoplasmic RNA molecules, binds the signal sequence of the proprotein and transports it to the inner membrane. During protein synthesis the GTPase chaperone competes with trigger factor for binding to newly formed protein molecules (Scott and Barnett 2006). Once bound with the protein, it is then targeted to the membrane by association with another GTPase called FtsY. At this stage the protein can interact with SecA ATPase and translocate. In Gram-negative bacteria the SRP pathway is used for most proteins whose final destination is the inner
membrane while in Gram positive bacteria this is the main route for membrane localisation due to the lack of a SecB homolog (Scott and Barnett 2006).

1.2.4 Secretion pathways

1.2.4.1 Type I system

The Type I secretion pathway requires three different accessory proteins which form a porin complex spanning the inner and outer membrane (Henderson et al., 2001). This complex consists of an ATP binding cassette (ABC) protein on the inner membrane, a membrane fusion protein (MFP), and an outer membrane protein (OMP) (Holland et al., 2005). The ABC transporter uses energy from the hydrolysis of the high energy molecule ATP to transport proteins across the inner and outer membrane. MFP and OMP form a channel which traverses the cell envelope allowing protein secretion. The signal sequence of the protein is located in the C-terminus, however, little is known about how it causes the protein to translocate. Gram-negative bacteria utilise the type I pathway to transport an array of proteins while the ABC transporter alone is used to transport a wide array of molecules such as ions, carbohydrates and peptides across the membrane surface (Holland et al., 2005). The secretion pathway for *Escherichia coli* hemolysin (HlyA) has been found to utilise this pathway.

1.2.4.2 Type II system

The type II secretion pathway is characterised by approximately 14 different accessory proteins all encoded on a single operon (Arrieta et al., 2004). Like the type I and type II secretion pathways the composite proteins of this pathway form a channel connecting the inner and outer membrane, but unlike type I secretion, the type II secretion pathway depends on Sec to fulfil its function. In contrast to the type I, the type II secretion system is a two-step secretion system. The proteins transported by this mechanism form distinct periplasmic intermediates, which are partially or wholly folded and have their signal sequence removed. Transport across the outer membrane requires a complex of 12-14 proteins which form an outer membrane secretion channel (Johnson et al., 2006). Other proteins involved in this process are responsible for anchoring the pore to the inner membrane and providing energy for secretion and include outer membrane proteins, integral inner membrane proteins and a cytoplasmic inner membrane ATPase. Connecting the inner and outer membrane, spanning
the cell envelope in this system is a pseudopilus type structure which may drive secretion of the protein across the cell membrane. This pathway is exemplified by pullulanase (Pul A), a starch hydrolyzing protein from *Klebsiella oxytoca* (Henderson *et al.*, 2001).

1.2.4.3 Type III system

The type III system is comprised of a complex array of proteins which form a highly regulated multimeric structure that spans both the inner and outer membranes. However, unlike type I and II, the type III pathway is shaped like a syringe, with a slender needle structure extending from the cell surface and a cylindrical base which is connected to both the inner and outer membrane and outer membrane and has marked similarity to a flagellar basal body (Henderson *et al.*, 2001). This pathway has been found to be not entirely Sec independent, although the Sec system is not required for effector molecule secretion, it has been found to provide some of the Sec machinery required for translocation across the inner membrane. Most type III systems inject virulence factors directly into the plant or animal host cells which the pathogen attacks. (He *et al.*, 2004) These virulence factors include toxins, phagocytosis inhibitors, and stimulators of cytoskeletal reorganisation in the host cell and promoters of apoptosis. An example of this system is the type III secretion system from *Yersinia* which injects *Yersinia* outer membrane proteins into immune cells to illicit the anti-phagocytic and anti-inflammatory response (Cornelis, 2002). However it has been found in some cases that the virulence factor is simply secreted into the extracellular environment. The Type III systems also transport other proteins, including i) proteins which regulate the secretion process, ii) some of the proteins needed to form the type III secretion complex, and iii) proteins that aid in the insertion of secreted proteins into target cells.

1.2.4.4 Type IV system

Type IV secretion requires the coordinated action of at least nine proteins localised within the cytoplasm and periplasm that are associated with the inner and outer membranes and like the type III system form a syringe type structure. (Zechner *et al.*, 2012) Not all bacteria contain a type IV secretion system while many organisms that do have homologous type IV secretion systems have been found to contain one common protein, Trbl. This system is unique in that it is utilised to secrete proteins as well as to transfer DNA from a donor bacterium to a recipient during bacterial cell conjugation. Bacteria which are known to utilise this system are
*Agrobacterium tumefaciens* to secrete T-DNA and *Bordetella pertussis* to deliver the pertussis toxin. The main virulence mechanism of this pathway is to inject T-DNA into the host and cause cancerous growth or the formation of crown gall tumours which in turn produce opines as a carbon and energy source for the bacterium. (Escobar and Dandekar, 2003) (Zechner *et al.*, 2012) The major components of the Type IV secretion system in *Agrobacterium tumefaciens* are VirB2-VirB1 and VirB4. VirB4 have been found to remodel the peptidoglycan layer of the cell wall using its lytic transglycosylase activity, while other VirB proteins are involved in the formation of the secretory complex (Tseng *et al.*, 2009).

1.2.4.5 Type VI system.

The type VI is a composed of a phage-tail-spike-like injectosome that introduces virulence proteins directly into the cytoplasm of host cells much like the type III and type IV systems. The pathway was first discovered as a conserved group of pathogenicity islands in Gram-negative bacteria (Bingle *et al.*, 2008). It has been found that over a quarter of sequenced bacterial genomes contain genes which encode for proteins of the type VI pathway, the majority of which are found in proteobacteria. The type VI secretion is required for virulence in plant and animal pathogens such as *Vibrio cholerae*, *Pseudomonas aeruginosa* and *Agrobacterium tumefaciens*. This pathway is also responsible for efficient root colonisation in nitrogen fixing species *Rhizobium* and *Mesorhizobium*. The injectosome contains a cytoplasmic chaperone with ATPase activity, a channel bridging the inner and outer membrane and a needle pore forming protein (Tseng *et al.*, 2009).

1.2.5 Other systems of translocation

1.2.5.1 Type VII system

The type VII secretion system is present in Gram positive *Mycobacterium* species which contain a cell wall is heavily modified by lipids. The Type VII secretion system was predicted bioinformatically based on clustering of genes which encoded secreted proteins that lacked a signal sequence. The presence of this new secretion system was confirmed by mutational analysis of the ESX-1 cluster of *Mycobacterium bovis* BCG vaccine strain. Little is known about this secretion system however current models suggest that it is composed of an inner membrane translocation protein Rv3877 and a separate channel in mycomembrane
which is of unknown composition. Proteins are predicted to cross the channel by binding Chaperone like ATPases anchored to the inner membrane (Abdallah et al, 2007).

![Figure 1.1: Summary of the known bacterial protein secretion systems](image)

**Figure 1.1: Summary of the known bacterial protein secretion systems**


### 1.2.6.1 Type V secretion system

Type V secretion is the simplest of all the secretion systems, given the complexity of all the other systems which require numerous accessory proteins forming a pore complex, to deliver the protein directly from the cytoplasm to the extracellular environment. Each protein secreted by this mechanism has a signature structure which encodes i) a signal sequence, which allows the protein to cross the inner membrane, ii) the passenger, which encodes the active protein and iii) a C-terminal domain encoding a β-barrel protein which integrates into the outer membrane (Henderson et al, 1998).

Currently, it is thought that all of the requirements of the protein for traversing the membrane are contained within the autotransporter molecule and that secretion is energy independent. Although the Type V method of secretion is thought to be the simplest, the protein that is transported is subject to biophysical constraints. The strongest of these constraints is the size of the pore, which is relatively small and dictates the structural properties of the protein.

There are three reported forms of this pathway within this system of translocation. The first of these are the proteins which utilise the classical autotransporter mechanism as described by Pohlner et al, 1987 proteins secreted by this mechanism are known as Type Va secretion. The
second category is type Vb proteins which undergo secretion via the two partner secretion pathway and the third is known as Vc under which multiple passenger domains are secreted through the same pore.

1.2.6.2 Type Va

The Va method of translocation was the first method of bacterial secretion utilising the type V system of protein secretion reported by Pohlner et al, 1987 in the first described autotransporter IgA1 protease from Neisseria gonorrhoeae. The modular structure of autotransporter proteins was first described in this seminal paper as well as a model for its translocation in which the β-domain formed a pore which allowed passage of the passenger domain across the outer membrane. As it is crossing the outer membrane the protein then folds into an active conformation until it is cleaved by a membrane bound protease releasing the protease into the extracellular milieu. The IgA1 protease was produced in other bacterial species where it was found that, when expressed, the passenger domain of this protein was contained in the growth medium, indicating that the IgA1 protein contained all the necessary information to direct translocation of the protein form the periplasm to the extracellular environment without the need of accessory proteins. These proteins were subsequently termed autotransporters due to this unique characteristic. (Pohlner et al, 1987)

1.2.6.3 Type Vb (two partner secretion TPS)

As with the Va type secretion, the Vb secretion system contains a signal sequence which directs the protein across the cytoplasmic membrane and enters the periplasm before integrating into the outer membrane via a pore formed by the β-barrel. Once outside the cell, the passenger domain may undergo further proteolytic activity to achieve its intended function. However, unlike the Va system, the passenger domain and the translocator domain are expressed as two separate protein chains, referred to as TpsA and TpsB proteins respectively (Jacob-Dubuisson et al, 2001). The β-domain of the Va system is also different, the β- domain of Va normally contains 16 β- sheets while the β- domain of Vb has been shown to contain 19 β- sheets (Meli et al., 2005) both β-domains have also been shown to have different conductivity values due to this difference in size. Both TpsA and TpsB are arranged in an operon type structure. The effector domain TpsA requires the translocator domain TpsB for transport across the outer membrane, this would indicate that there needs to
be a recognition structure in TpsB for TpsA in order for translocation to occur, this adds an additional complexity to the translocation domain unseen in the Va system. In TpsA there exists an N-terminal proximal domain called the TPS domain which interacts directly with the TpsB protein, facilitating translocation. An example of a protein secreted using this pathway is *B. pertussis* adhesin FHA (Clantin *et al.*, 2004).

The TPS system shares many characteristics with the autotransporter system such as the export of large proteins in excess of 100 kDa (Jacob-Dubuisson *et al.*, 2006). It has been postulated that the TPS system also shares the same mechanism of crossing the inner membrane using the Sec machinery in an unfolded conformation in an energy independent manner. An investigation of the signal sequences of several of these proteins has shown that they are quite similar in comparison to that of proteins exported by the Va system. Further evidence of a relation between these two systems of translocation can be drawn from phylogenetic analysis where even though it has been suggested that both pathways arose independently, it has been suggested that ‘shuffling’ of domains between proteins of the TPS system and the autotransporter system occurred over the course of evolution (Henderson *et al.*, 2000).

![Figure 1.2: Systematic overview of the different types of autotransporter secretion systems.](image)

*The type Va secretion system is depicted on the left, the type Vb system in the middle and the type Vc system on the right (Henderson *et al.*, 2004).*
1.2.6.4 Type Vc oligomeric coiled-coil adhesins.

Oligomeric coiled-coil adhesins are a sub-family of the surface attached oligomeric autotransporters. The YadA protein from *Yersinis pestis* is the prototypical example of this model, the YadA protein contains five distinct domains i) N-terminal signal, ii) a head-domain iii) the neck domain, iv) a stalk domain and v) a C-terminal β-region consisting of only four β sheets (Hoiczyk *et al*., 2000). In a similar fashion as the Va system, deletion of the C-terminal domain causes this protein to be misinserted into the outer membrane, while deletion of the linker region results in degradation of the protein (Roggenkamp *et al*., 2003). Trimerisation of the β-region forms a β-barrel much like that which forms for the Va system, this trimer forms a β-barrel containing twelve β-sheets and confers an overall ‘lollipop’ structure to proteins displayed on the cell surface (Hoiczyk *et al*., 2000).

![Figure 1.3: Model of YeYadA showing binding regions for different ligands (Koretke *et al*., 2006)](image)

1.2.6.4.1 Trimeric autotransporters

Conventional autotransporters have numerous functional activities such as protease activity, lipase activity and numerous other functions involved in virulence. Trimeric autotransporters normally contain an adhesive activity which facilitates interaction with host cells and as a binding site for circulating host cell factors such as antibodies and complement inhibitory proteins. Adherence to host cells has been observed with trimeric autotransporter proteins such as YadA from the Yersinia species and Hia from *Haemophilus influenzae* (Cotter *et al*., 2005) Like conventional autotransporters the passenger domain of trimeric autotransporters
are translocated through a pore formed by the β-domain, which in the case of trimeric autotransporters is a pore formed by three individual polypeptide chains each contributing four β-strands. There are physical constraints associated with the translocation of conventional autotransporter which also apply for translocation of the trimeric protein. This would indicate that trimerisation of the passenger domain cannot occur in the periplasmic space as a folded trimeric protein would be unable to cross the pore formed by a trimeric β-barrel. This indicates that translocation of the passenger domain must occur in an unfolded state, as in the Type Va model (Dautin and Bernstein 2007).

1.2.6.4.2 Translocation of TAA

The mechanism of trimeric autotransportation is assumed to be widely similar to translocation of monomeric autotransporters, although the BamA model has been ruled out due to a lack of reports of oligomeric trimeric β-barrels (Oomen et al., 2004). The process of translocation occurs at the junction of the passenger domain and the beta-domain. Under this model, the C-terminus of the protein is secreted first forming a hairpin loop across the outer membrane. The pore formed by the β-domain in the outer membrane must be large enough to accommodate four to six unfolded polypeptides depending on whether the subunits of the trimeric protein are translocated sequentially or simultaneously (Oomen et al., 2004). In the other model for trimeric translocation, in which the N-terminus of the passenger domain is secreted first, the pore in the trimeric β-barrel must be large enough for three unfolded polypeptides to pass. This means that the pore must be larger than that of the conventional autotransporter, exemplified by the NalP pore which, at 1.25 nm, is only large enough for two polypeptides to pass (Khalid and Sansom 2006). Once all three subunits have been translocated, it has been postulated that the free energy associated with trimer formation which provides the basis for protein folding.

The first protein to be characterised as using the Vc pathway was Hia autotransporter, unlike many classical autotransporters the Hia passenger domain remains surface associated (Cutter et al., 2002). As with classical autotransporters, the C-terminal translocation domain contains the autotransporting activity of trimeric autotransporter. The β-barrel of trimeric autotransporters has been found to contain a conserved glycine residue, which has been found to be replaced occasionally by an alanine, serine, threonine or asparagine residue. Mutation of this residue in YadA has been found to result in reduced expression, increased periplasmic
degradation and destabilised trimers. The effect of mutation increased with the size of the residue side chain (Grosskinsky et al, 2007). Although the translocation domains of trimeric autotransporters have been shown to be highly similar, they have been found to not assemble with related autotransporters. When YadA was co-expressed with Hia, it did not affect the adhesive activity of either protein, suggesting that there was no trimerisation of the C-terminal domains as cross-trimerisation would have affected activity (Cotter et al, 2005). This study also suggested that trimerisation of the passenger domain occurred after trimerisation of the translocating domain, which has been found to be different to the process that occurs in classical autotransporters, where trimerisation of the passenger domain occurred at the same time as the translocating domain in the absence of surface proteases. Although translocating domains of trimeric autotransporters have been found not to cross-assemble, chimeric proteins of the translocating domains from Hia and the YadA passenger domain have been found to produce a translocated passenger domain (Ackermann et al, 2008).

The only major difference between classical and trimeric autotransporters is in the number of peptide chains secreted. In the case of classical autotransporters one polypeptide is secreted as opposed to three for trimeric autotransporters. The application of the hairpin model to trimeric autotransporters poses some difficulties, specifically that the size of the pore only permits the passage of a maximum of two unfolded peptides at any one time. Although the size of the Hia translocating pore can accommodate three hairpins, the fit would be extremely tight, questioning the plausibility of this model (Cotter et al, 2005).

1.2.6.4.3 Trimeric binding properties

When folded in native conformation, the passenger domains of trimeric autotransporters have three identical faces, giving trimeric symmetry and three potential binding pockets. This has the potential to provide a more stable interaction with the host with an increased binding affinity when compared with monomeric autotransporters (Cotter et al, 2005). This structure also has the potential to overcome mechanical forces in the host. The increase in binding affinity is exemplified by the adenovirus type 2 fiber which is a homotrimer containing three identical binding sites for binding adenovirus receptors. The trimeric form of this protein has been found to impart a 25-fold greater binding affinity than that of the monomeric protein. Homotrimeric proteins have also been shown to exhibit resistance to trypsin and detergent
degradation, which allows these proteins to remain functional under harsh conditions (Muller et al., 2005, Ackerman et al., 2008).

1.2.7 The beta domain

1.2.7.1 Overview

β-barrels can have different topologies, the simplest of which is when each of the β-sheets is connected to the adjacent β-strand in an antiparallel fashion forming a β-barrel. All membrane proteins with solved structures have been found to contain the β-barrel architecture while α-helical bundles are generally associated with cytoplasmic membrane proteins. It has been postulated that the reason for this is that outer membrane proteins need to cross the inner membrane before integrating into the outer membrane. If outer membrane proteins contained an alpha helical architecture, they would become ‘stuck’ in the inner membrane and would not be able to integrate into the outer membrane (Henderson et al., 2004).

The beta domains of autotransporter proteins have been found to be conserved in structure but very diverse in sequence. The normal size of this domain has been found to contain 200-300 amino acids. Sequence analysis of β domains has allowed the identification of a consensus sequence in the C-terminal domain of several diverse autotransporter proteins which has permitted the identification of new autotransporter proteins through sequence analysis (Muller et al., 2005).

The majority of autotransporter proteins contain a 12 antiparallel β-sheets, it has been predicted that an even number of antiparallel β-sheets are required to ensure the correct conformation and ensure closure of the barrel (Loveless and Saier 1997). In addition to this, the terminal amino acid at the C-terminus is always a phenylalanine or tryptophan preceded by alternating charged (or polar) and aromatic (or hydrophobic) residues (de Cock et al., 1997). Deletion or substitution of the C-terminal phenylalanine residue has an extreme effect on folding and stabilisation of the monomer, resulting in ineffective localisation and trimerisation (de Cock et al., 1997). Deletion or substitution of the three final amino acids (YSF) in the Hap protein of H. influenza results in a reduction of the quantity of protein
transported to the outer membrane. However, individual mutation of these residues has been shown to have no effect on the localisation of the protein (Hendrixson et al, 1997).

Once the autotransporter protein crosses the inner membrane, the proprotein inserts spontaneously into the outer membrane in a biophysically favoured β-barrel conformation as it interacts with the non-polar environment of the outer membrane. The C-terminal domain of the protein forms a porin-type structure, comprising amphipathic antiparallel β-sheets in a β-barrel conformation. The β-barrel forms a complex protein structure in which the first and last β-sheets form hydrogen bonds, closing the ring conformation. The amphipathic primary structure permits the establishment of a molecular pore, with the polar side chains embedded in the membrane with the polar side chains exposed to the aqueous environment at the centre of the pore (Henderson et al, 1998).

Investigations of the outer membrane proteins OmpA and PhoE suggest that this process is not entirely spontaneous and the periplasmic chaperone Skp and lipopolysaccharide (LPS) are required for the efficient autotransporter assembly in the outer membrane. The first model suggested for this process was suggested by Bulieris et al, 2003 using Omp A. This model suggests that after OmpA is transported across the periplasm it binds three molecules of Skp which prevent folding of the protein. The Omp A/Skp complex then binds two to seven molecules of LPS forming a folding competent intermediate which facilitates folding and insertion of OmpA into the outer membrane. It was postulated that the activity of LPS and Skp is involved in delivering a number of proteins to the outer membrane. However, deletion of the Skp protein has not been found to inhibit the passage of proteins to the outer membrane, indicating that this pathway is not the only mechanism by which proteins reach the outer membrane (Bulieris et al, 2003).

1.2.7.2 Mechanism of translocation

1.2.7.2.1 The Hairpin model

There are three models for the translocation of proteins using the type V secretion mechanism. The first of these and the most generally accepted is the hairpin model (Pohlner et al, 1987) as it is strongly supported by the elucidation of the NalP translocator structure. The hairpin model postulates that translocation of the passenger is initiated with the C-
terminal end of the passenger domain forming a hairpin structure inside the autotransporter β-barrel. This is then followed by movement of the rest of the passenger domain through the barrel’s pore in a C- to N- terminal direction (Ruiz-Perez et al., 2009).

This secretion was first described for the Neisseria gonorrhoeae IgA1 proteases. It has been postulated that during secretion the Iga precursor is processed at both the N- and C-terminals. Export of the protein occurs via the Sec pathway and is followed by the cleavage of the 27 amino acid N-terminus signal sequence. The majority of autotransporters contain a signal sequence that contains some of the features of a Sec dependent signal sequence, which are i) a N- domain of positively charged amino acid residues, ii) a hydrophobic region of neutral amino acids and iii) a signal peptidase recognition site on the C-terminus. It has been found that a number of autotransporters have unusually large signal sequences, normally consisting of at least 47 amino acids which typically occur between the first methionine and the positively charged N-domain. It is thought that this extended signal sequence permits secretion through the Sec apparatus or is involved in the recruiting of accessory proteins.

Translocation of the protein across the outer membrane is mediated by the carboxy terminal domain of the protein. Under the model proposed by Pohlner et al., 1987, the passenger domain forms a hairpin like structure and appears to be unfolded or partially folded as it crosses the membrane. The solved structure for the β-barrel for NalP showed that it was sufficiently wide enough for two polypeptide chains to pass through simultaneously. After the signal sequence is cleaved from the proprotein, the protein is transported through the β-barrel structure to the cell surface.
1.2.7.2.2 The BamA model

It has been proposed that the outer membrane protein BamA might have a role in the translocation of folded proteins and could possibly facilitate the folding of entire autotransporter proteins. BamA is a highly conserved protein, homologues of which are present in every Gram-negative bacteria which has been sequenced. The gene which encodes BamA is located very close to the skp gene and the lpxA and lpxB genes which form LPS. These genes, as outlined above, have a role in the integration of the β-barrel into the membrane. The BamA protein has a vital role in cell viability and for outer membrane protein assembly in bacteria and shows sequence relationship to Sam50 of mitochondria, and has sequence similarity to Toc75 which is involved in the chloroplast protein import machinery. Strains which contain deleted BamA have been shown to contain unassembled forms of outer membrane proteins (Voulhoux and Tommassen 2004).

BamA has been suggested to assist and protect the insertion of the β-barrel into the outer membrane and provide a pore for the translocation of the passenger domain prior to release into the outer membrane. (Robert et al., 2006) One of the main problems associated with the hairpin model is that the pore of the β-barrel structure formed to transport the passenger domain is insufficient to permit the transport of small folded elements and closely spaced disulphide bridges which have been reported to occur in the translocation of autotransporters. It has been postulated that the Bam complex contained in the outer membrane assists in the
initial translocation of the folded passenger domain of the autotransporter Hpb before the β-domain is fully assembled in the outer membrane (Sauri et al, 2009).

Figure 1.4.2: Diagrammatic representation of the BamA model of autotransporter secretion (Dautin and Bernstein, 2007).

1.2.7.2.3 The hexameric model

This model proposes that transport of the passenger domain occurs when six IgA1-protease monomers form a common pore which was proposed by Veiga et al, 2002. In this model, it was proposed that although members of the Type V do not share the same multimeric complexes needed to secrete proteins using the Type I-IV models, it was found that proteins which utilised the Type V model contained structural similarities on their C-terminus to most secretion systems. This was then responsible for the formation of a multimeric ring like structure with a central hydrophilic channel. Such a structure would be analogous to other membrane complexes found in other secretion systems such as that for secretin and fimrial ushers (Veiga et al, 2002).

It was found that the structural and functional properties of the IgA1 protease supported the idea that the transport of N-passenger domains of autotransporters occurs through a common membrane channel, assembled into a single complex from numerous subunits from different autotransporter proteins. The pore formed by the putative hexameric complex of IgA1, at 2
nm, was found to be larger than that found for other autotransporter proteins such as OmpC and OmpF (1.1 and 1.2 nm respectively) and found to be of a similar size to that of porins from other the secretion pathways such TolC and PapC oligomers which have a pore size of 2 nm and 3 nm. The size of the pore formed by the IgA1 protease is of sufficient size to permit the transport of fully folded proteins, which directly contradicts the hairpin model proposed by Pohiner et al, 1987. However, it was found that the IgA1 hexameric complex was of a smaller external diameter than that is found in other secretins, while the central pore is narrower than those found in the type II and type III systems. This indicates the proposed model is loosely based on that which is found in the other protein secretion models in Gram-negative bacteria.

Figure 1.4.3: Schematic representation of the multimeric model of autotransporter secretion (Dautin and Bernstein 2007).
1.2.7.2.4 Passenger domain cleavage

Some autotransporters, such as trimeric autotransporters, have been found remain intact after display on the cell surface, while many classical autotransporters are known to release their passenger domain into the extracellular environment. This process can occur via several different processes:

i) Cleavage by an outer membrane protease which is demonstrated by the *Shigella flexneri* autotransporter IcsA, which is cleaved by a dedicated IcsP protease.

ii) Cleavage by another autotransporter; the passenger domains of several *N. meningitis* autotransporters such as IgA protease and MspA are cleaved by NalP which encodes a serine protease.

iii) Cleavage by an intermolecular protease reaction; the passenger domains of *H. influenza* Hap and *B. pertussis* proteins are cleaved by their own serine protease activity.

iv) Intramolecular cleavage; other autotransporters such as EspP from *E. coli* are cleaved inside the pore by the β domain, a reaction that involves nucleophilic attack of the polypeptide backbone by a conserved asparagine residue (Dautin and Bernstein., 2007).

Figure 1.5.1: Cleavage of the extracellular passenger domain by an outer membrane protease (Dautin and Bernstein 2007).
Figure 1.5.2: Cleavage of the secreted passenger domain by another autotransporter (Dautin and Bernstein 2007).

Figure 1.5.3: (left) Cleavage of the extracellular passenger domain by an intermolecular protease reaction and Figure 1.5.4 (right) intramolecular cleavage by other autotransporters (Dautin and Bernstein 2007).
1.2.8 Accessory Proteins for translocation

1.2.8.1 Overview

The name ‘autotransporter’ was coined for proteins which utilise the type V pathway as it was thought that these proteins had the intrinsic ability to transport themselves without the need for accessory proteins. However, it has recently been shown that several proteins have been involved in targeting and assembly of extracytoplasmic proteins. These accessory proteins fall into three main categories:

i) Molecular chaperones, examples of which are DegP, PpiA and Skp, have been shown to stabilise non-native confirmation of target proteins and facilitate folding.

ii) Proteases, such as DegP and DegQ, which degrade misfolded or unproductive proteins.

iii) Peptidyl–prolyl cis-trans isomerase which catalyse the rate-limiting steps of isomerisation during folding. Examples of these proteins are SurA and Fkp A (Ruiz-Perez et al, 2009).

The protein EspP, which is a protease from \textit{E. coli} O157:H7, has been found to interact directly with periplasmic chaperones SurA and DegP. Interestingly, the identified chaperones have been found to bind to the unfolded EspP protein but not to the native folded protein. It was also noted that the observed interactions were specific as equimolar concentrations of negative control proteins did not bind to EspP. It was also shown that strains expressing mutated SurA, DegP and EspP exhibit greatly reduced growth rates with apparent cell lysis. Periplasmic accessory proteins have been found to bind proteins with a aro-X-ar (aromatic residue- any amino acid- aromatic residue) motif, fourteen of which are found in EspP while it has also been revealed that there is a high density of this motif across the SPATE family of autotransporter proteins (Ruiz Perez et al, 2009).
1.2.10 Families of Autotranporters

1.2.9.1 Ssp

The first subtilase protease was discovered in *Serratia marcesans* and had its secretion mechanism deduced at the same time as IgA1 protease. Sequencing of the gene encoding subtilin serine protease found that this gene contained an active site similar to that of trypsin. When this protein was expressed in *E. coli* it was found that: i) when expressed under control of the *lac* promoter, the 112 kDa precursor protein was found in an insoluble form in the periplasmic space; ii) the mutated gene lacking the C-terminal domain was localised in the periplasmic space; iii) a mutant protein S341T was found to be secreted into the extracellular space with abnormal processing and; iv) the cleavage of the Ssp passenger domain from the β-domain occurred via an autocatalytic event and occurred at several points in the junction between the passenger domain and the β-domain (Yanagida *et al.*, 1986).

1.2.9.2 SphB1

SphB1 is a *Bordetella pertussis* autotransporter with a 24% identity to Ssp and is necessary for the cell surface maturation of the Bvg-regulated filamentous hemaglutinin (FHA). As with Ssp, SphB1 contains a subtilase motif. Mutation S412A in this motif causes the processing of the passenger domain upstream of its β-domain which suggests that SphB1, like Ssp and BrKA is autocatalytic. This mutation also effects the processing of FHA. Sph1 has been found to be attached to the bacterial surface, but a small amount was found in the extracellular millilieu. SphB1 remains attached to the cell surface through the lipidation of the N-terminus of the protein. Removal of this lipid moiety through mutation of the conserved lipidation site or globomycin treatment inhibited the translocation of the protein (Coutte *et al.*, 2001; Coutte *et al.*, 2003).
1.2.9.3 AspA/NalP

AspA/NalP was found through bioinformatical analysis of the *N. meningitides* genome and contains an N-terminal which is homologous to other secreted subtilisin-like proteases and, like SphB1 and Ssp, contains a conserved subtilisin-like active site. The protein exhibits autocleavage activity which causes cleavage of the passenger domain and the release of this 96 kDa protein into the extracellular environment. The protein also contains a type II signal peptidase motif, which would indicate that the protein is a lipoprotein like SphB1 (Turner *et al.*, 2002). AspA/NalP was cloned and expressed and the β- domain structure was solved. Antibodies to AspA/NalP were found in the blood sera of patients, suggesting that AsaP/NalP is expressed in vivo during infection. Electron microscopy and cellular fractionation studies have shown that AsaP/NalP is translocated to the outer membrane and remains surface exposed. Western Blot analysis has shown that smaller 68 or 70 kDa fragments are secreted into the extracellular environment. Site directed mutagenesis of S426 abolished the secretion of *E. coli* AspA which confirms that AspA is autocleaved (van Ulsen *et al.*, 2003).

1.2.9.2. *Helicobacter pylori* autotransporters

1.2.9.2.1 Overview

*Helicobacter pylori* is a Gram-negative bacteria involved in development of gastric and peptic ulcers as well as stomach cancer, specifically gastric adenocarcinoma and gastric lymphoma. The mechanism by which *H. pylori* causes damage to gastric mucosa resulting in these diseases is unknown, however the bacterium is known to produce several virulence factors secreted by the type V secretion mechanism which include VacA, SabA, BabA and AlpA.
1.2.9.2.2 VacA

VacA contains the hallmark features of an autotransporters, including a 33 amino acid signal sequence, a 90 kDa passenger domain and a 33 kDa translocation unit. The passenger domain has been found to be cleaved and secreted into the extracellular environment. The VacA passenger domain has been shown to aggregate into large oligomeric complexes under non-denaturing conditions. VacA has been shown to form two different complexes when examined using electron microscopy: i) large ring-form complex and; ii) a flat form complex. The ring-form complexes have been found to be 30 nm in diameter and consist of a central ring surrounded by six or seven ‘petals’, leading to this structure being referred to as the flower structure. The flat form consists of six or seven ‘petals’ radiating from the centre of the complex, but unlike the flower form do not contain a central ring (Luppetti et al, 1996). This conformation has been found to be supported by atomic force microscopy which has found that VacA is arranged in hexagonal ring and are attached by connectors to peripheral domains. It has also been found that exposure to acid or alkaline environment has been found to cause dissociation of these oligomeric complexes and revert to the 90 kDa monomeric VacA domain. VacA in these environments has been found to have increased cytotoxicity and internalisation ability, indicating that this dissociation is important for the biological functionality of the protein (Molinari et al, 1998).

VacA has been found to be further processed after secretion forming a 33 kDa N-terminal domain and a 55 kDa host cell receptor binding domain. The two domains have been found to be separated by limited proteolysis or autodegradation during prolonged storage. Mutational analysis of the VacA protein, has shown that VacA which had much of the p33 domain removed was found to form water soluble dimers of a similar structure to that of the ‘petals’ of the VacA oligomers, this would indicate that the 55 kDa domain is responsible for the formation of the observed oligomerisation of the peripheral ‘petals’. In frame deletions of the 33 kDa domain have been found to produce truncated forms of the protein which are secreted but do not oligomerise and lack cytotoxic activity.
1.2.9.3 Serine protease autotransporters

1.2.9.3.1 IgA1 protease

Immunological defence on mucosal surfaces is primarily mediated by type IgA1 antibodies which have been found to inhibit microbial adherence and neutralise bacterial toxins and viruses. IgA1 is targeted by bacterial cells to evade immune cell responses and was found to be degraded by bacterial proteases which were called IgA1 proteases due to this action. IgA1 proteases cleave one of the peptide bonds within the hinge region segment of the IgA1 heavy chain region, producing Fabα and Fcα fragments (Kilian et al., 1980).

IgA1 proteases are produced by several major bacterial pathogens such as N. gonorrhoea, N. meningitis, H. influenza as well as Streptococcus pneumoniae (Kilian et al., 1980). IgA1 proteases are classified into three groups: the IgA1 proteases of H. influenzae and N. meningitidis are serine proteases, metalloproteases in S. Sanguis and S. pneumoniae and cysteine proteases in Prevotella melaninogenic (Mortensen et al and Kilian, 1980?). The IgA1 protein of Haemophilus and Neisseria species specifically attack the Proline-Serine or Proline–Threonine peptide bonds within the hinge region of IgA1. All IgA1 proteases from these species contain a specific consensus sequence GDSGSPLF, which contains a conserved serine within the active site characteristic of serine proteases. The exact peptide bond cleaved by the specific species differs between strains (Bachchovin et al, 1990).

Cleavage of the fully active protein results in the loss of Fcα mediated functions such as the inhibition of adherence. This occurs despite the antigen binding activity of the Fabα domains. It has been postulated that binding of the cleaved Fabα domains to bacterial targets inhibits binding of the IgA1 antibody and prevents activation of the immune response. It has also been suggested from in vitro studies that IgA1 protease is not required for adherence and invasion of epithelial cells in N. gonorrhoeae (Cooper et al., 1984). However, cleaved IgA1 antibodies have also been found in the cerebrospinal fluid of patients with bacterial meningitis and in vaginal washings of patients with gonorrhoea and other secretions from individuals with bacteria which encode IgA1 proteases (Ahl T and Reinholdt, 1991). Antibodies raised to IgA1 have been found in sera and secretions of patients with infections from bacteria which encode IgA1 protease.
IgA1 protease does not only have the IgA1 as a substrate, it has been shown to undergo autocleavage once exposed on the extracellular side of the cell. When in the fluid phase, the IgA1 undergoes another cleavage in its sequence to release the 15 kDa α peptide. The susceptibility of several isolated inner and outer membrane proteins was examined by Shoberg and Mulks (Shoberg and Mulks 1991) in an in vitro assay using both wild type gonococcal strains and IgA1 protease deficient strains. It was found that IgA1 protease is capable of hydrolysing both inner and outer membrane proteins and that this is not strain or species specific (Shoberg and Mulks 1991). A second target of IgA1 protease, called synaptobrevin II, was discovered in eukaryotic cells. Synaptobrevin is a fusion complex forming protein associated with vesicles, it is also cleaved by tetanus toxin (Binscheck et al., 1995). It has also been reported that the type II N. gonorrhoae IgA1 protease has been found to cleave LAMP-1 with is a lysosomal/phagosomal membrane protein. These proteins, in mammalian cells, are coated in glycoproteins which are highly conserved and are thought to protect the membrane from degradation. LAMP-1 contains an IgA1-like hinge region with putative cleavage sites for IgA1 protease. Internalised bacteria have been shown to alter their phagocysomal environment to prevent lysosomal death, cleavage of the LAMP-1 protein has been shown to promote the intracellular survival of pathogenic N. gonorrhoae (Hauck and Meyer 1997). IgA1 protease has also been shown to stimulate the release of tumour necrosis factor alpha (TNF-α), interleukin 1β (IL 1β), IL-6 and IL8 from peripheral blood mononuclear cells which were found in mucosal infection with N. gonorrhoae (Lorenzen et al., 1999).

1.2.9.3.2 Hap

*H. influenzae* is a common commensal organism which is involved in several diseases such as bronchitis and pneumonia while encapsulated strains are known to cause septicaemia and meningitis. Adherence of *H. influenzae* to mammalian cells depends on the expression of three proteins, i) Hia adhesin, ii) HMW1 and HMW2 adhesins and iii) Hap protease (St Geme et al., 1994).

*Haemophilus* adhesion and penetration protein or Hap is a surface protein associated with attachment to, and entry into, cultured epithelial cells and has been shown to act as an adhesin in strains which lack HMW and/or hia adhesions. Hap is synthesised as a 155 kDa outer membrane protein which is further processed into a 45 kDa β-domain and a 110 kDa
passenger domain and has been shown to have significant homology to IgA1 protease of *H. influenzae* with 60% identity and 80% similarity. Hap also directs its own autocleavage via its own protease activity, like IgA1 protease, releasing the passenger domain into the extracellular environment (St Geme *et al.*, 1994). Unlike the IgA1 protease the cleavage is not complete and thus the Hap exists in two forms; an extracellular form and a surface associated form. It was found that the cell associated protein-mediated adhesion of *H. influenzae* to epithelial cells and bacterial aggregation (Hendrixson and St. Geme 1998). Secretory leucocyte inhibitor which protects the host cell epithelium has been shown to inhibit autoproteolysis of Hap facilitating the formation of the membrane bound form of the protein and, therefore, its ability for adherence. It also suggests that the protease ability of Hap is not needed for adherence (Hendrixson *et al.*, and St. Geme 1998). In contrast lactoferritin has been shown to specifically degrade Hap and Hap-mediated adherence (Qiu J *et al.*, 1998). Release of Hap into the extracellular environment is thought to facilitate the evasion of the immune system, it has been postulated that Hap may undergo the same autoproteolytic cleavage at several sites which releases bioactive peptides as with IgA1 protease. The secreted extracellular protease may also have a role in degrading host immune system proteins and migration of individual bacterial cells within the respiratory tract (Cutter *et al.,* 2002).

1.2.9.3.3 SPATE- serine protease autotransporters of the Enterobacteriaceae.

Members of the SPATE family of proteins contain are serine proteases like IgA1 protease and Hap. The SPATE family of autotransporters have several common features: i) unlike the IgA1 protease, SPATE proteins do not cleave IgA1; ii) the inherent serine protease motif does not have a role in autoprocessing; iii) each member of the SPATE family is the predominant protein secreted by its respective pathogen; iv) SPATE proteins have not been identified in non-pathogenic bacteria; v) SPATE proteins are highly immunogenic (Dutta *et al.,* 2002).

The first member of the SPATE family to be discovered was Tsh which is a temperature sensitive haemaglutinnin from avian pathogenic *E. coli*. The *tsh* gene product is a 140 kDa protein which is processed into a 106 kDa extracellular domain and a 33 kDa outer membrane domain. This protein contains the same seven amino acid consensus serine protease motif as found in IgA1 protease (Stathopoulos *et al.,* 1999). However, unlike IgA1
protease, Tsh does not cleave IgA1 protein or show proteolytic activity in casein based assays. The relation of protein expression to hemaglutination assay is complex and is strain and environment association, it was also found that activity was conferred by the whole bacterial cell and not the cell-free supernatant indicating that the passenger domain remains associated with the cell and is not secreted in the extracellular environment. (Stathopoulos et al, 1999).

1.2.9.3.4 Pic

Pic protease was first discovered in entero-aggregative *E. coli* (EAEC) in a paediatric ward of a Mexican hospital during an outbreak of diarrhoea (Navarro–Garcia et al, 1998). The *pic* gene is chromosomally encoded and is identical to a mucinase from *Shigella*, the *pic* gene also has two oppositely orientated genes called *set1A* and *set1B* which encode the 7 kDa and 20 kDa subunits of the 55 kDa ShETI toxin (Rajakumar et al., 1997). Pic catalyses gelatin degradation which is abolished though mutation of the serine protease motif (Henderson et al, 2004).

1.2.9.4 AIDA autotransporter family

1.2.9.4.1 AIDA

The autotransporter involved in diffuse adherence was first was found on a chromosome in *E. coli* strain 2787 which was found to have a diffuse adherence phenotype. Fragments of 2787 plasmids which were thought to contain the phenotype for diffuse adherence were cloned and transformed into *E. coli* C600 strain. The clones were analysed for the diffuse adherence phenotype using a HeLa cell assay, after which the diffuse adherence phenotype was localised to a 6 kb DNA fragment encoding a 44.8 and a 132 kDa protein called Aah and AIDA respectively. Further evidence that this protein was involved in diffuse adherence was found when: i) mutant strains of this protein were found to not contain the diffuse adherence phenotype; ii) when specific antisera for AIDA were added to the HeLa adherence assay, it was found that adherence was inhibited; iii) it was demonstrated by electron microscopy that the diffuse adherence phenotype was surface localised (Benz and Schmidt 1992b).
It was found that the 132 kDa protein differentiated into two forms; a passenger domain with a molecular mass of 79.5 kDa and a translocating domain of 47.5 kDa. The passenger domain of AIDA was found have similar characteristics to that of antigen 43, a trimeric adhesin form *E. coli*:

i) After secretion, it was found that the passenger domain remains surface associated, but could be removed from the cell surface by heating to 60°C.

ii) When analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), it was found that the protein migrated at a higher molecular weight value than the predicted size (Benz and Schmidt 1992a).

It has also been found that during purification of the AIDA passenger domain using gel filtration the protein aggregates together eluting in the 450-600 kDa range, indicating a pentameric or hexamer structure. Bioinformatic analysis and circular dichroism studies have revealed that the passenger domain contains several repetitive β-sheets which is similar to the structure of pertactin from *B. pertussis*. Repetitive β-sheets are associated with proteins which have adhesive or receptor functions (Laarmann and Schmidt 2003).

1.2.9.4.2 Antigen 43

Antigen 43 was one of the first outer membrane proteins to be discovered in the 1970’s, when researchers first started looking at the proteins of the cell membrane using SDS-PAGE and cross-immunoelectrophoresis. Antigen 43 was discovered in one such experiment and was called antigen 43 because it was the 43rd antigen in the cross-immunoelectrophoresis profile of membrane vesicles prepared from *E. coli* (Henderson and Owen 1999). Antigen 43 was one of the first autotransporters to be characterised and it was discovered that the protein consisted of a complex of two proteins which were termed α<sup>43</sup> and β<sup>43</sup>. These proteins were found to exist in a non-covalent complex with a ratio of 1:1, further investigation revealed that the α domain was surface localised and was found to be released when heated to 60°C while the β domain was found to be an integral outer membrane protein (Henderson and Owen 1999).

The activity of antigen 43 was not determined until sometime after its discovery. Antigen 43 was found to contain two tri-peptide RGD motifs which have been shown to be the motif which mediates cell attachment within fibronectin. When expressed, antigen 43 was shown to
lead to flat, frizzy and irregular colony morphology, whereas cells not expressing antigen 43 were shown to have glossy circular colonies. Antigen 43 was also shown to cause a cell-cell interaction which resulted in autoaggregation. Type 1 fimbrial expression was found to inhibit this cell-cell interaction possibly due to steric hindrance. Unlike other members of the AIDA family, antigen 43 is found across all strains of the \textit{E. coli} species, while many pathogenic strains have been found to contain multiple homologues of the antigen 43 gene (Henderson and Owen 1999).

1.2.9.4.3 Taah and AipA

The genes encoding TaaP and AipA were first described after sequencing of the \textit{Proteus mirabilis} HI4320 genome. The genes which were known as PMI 2575 and PMI 2122 respectively were first shown to encode trimeric autotransporters with ‘agglutinating adhesion-like’ and adhesion-like properties respectively. The first 62 amino acids of the C-terminal of Aipa and the first 76 amino acids of TaaP are homologous to the translocator domain of Hia and YadA trimeric autotransporter adhesions. Models of these proteins were created using the solved structure of the Hia autotransporter and were predicted to contain four antiparallel β-sheets and form homotrimers. When expressed recombinantly, TaaP was found to have a molecular weight of 78 kDa, while AipA was found to have a molecular weight of 28 kDa. AipA and TaaP were found to bind extracellular matrix proteins with a 10-fold to 60-fold increase in binding affinity over the control strain. Inactivation of the AipA protein in \textit{P. mirabilis} was found to significantly decrease the ability of \textit{P. mirabilis} to bind HEK293 cell monolayers. It was found that a 51 amino acid long invasion region in Aida was responsible for this activity. A TaaP mutant of \textit{P. mirabilis} was also found to produce reduced aggregation than HI4320. The conserved glycine residue in the C-terminal domain of other trimeric autotransporters such as YadA, which is required for efficient trimerisation of the protein, was also found in both Aipa and TaaP (Alamuri \textit{et al}, 2010).
1.2.10.1 Overview of autotransporter structure.

Due to the diversity in chemistry between the domains of autotransporters it is extremely difficult to isolate full length autotransporters in a stable form. The first crystal structure of a full length autotransporter EstA from *Pseudomonas aeruginosa* was solved by Van den Berg (2010) at 2.5 Å resolution. EstA has a narrow porin structure consisting of 12-stranded β-barrel which is occupied by an alpha helix to which the passenger domain is covalently attached. The passenger with a globular fold hairpin model is different from that of other known passenger domains (Pohlner et al., 1987). Some of the autotransporter proteins from *P. mirabilis* used in this study have a different structure which is similar to the Hia autotransporter from *Haemophilus influenza* (figure 2.2) or the YadA autotransporter from *Yersinis pestis*. The Hia autotransporter is similar to the 2575 protein analysed in this study and its structure was solved at 2 Å resolution (Meng et al., 2006). The conformation of Hia is trimeric with each of the monomers donating 4 β-sheets to form the translocation domain.
1.2.10.2 Extraction of membrane proteins

Autotransporters are located in the outer membrane of Gram-negative bacterial cells. The study of autotransporters, like many other membrane bound proteins, is limited by the method used to extract the protein from the membrane before the advent of detergent usage. Detergents are amphiphatic molecules which associate with hydrophobic surfaces. They consist mainly of a hydrophilic head group and a hydrophobic tail group and generally fall into three categories depending on the nature of the hydrophilic head group: (i) cationic/anionic, (ii) non-ionic and (iii) Zwitterionic. The properties of each detergent are determined by the stereochemistry and character of the head and the tail group (Garavito and Ferguson-Miller, 2001). The property which makes detergents useful with membrane proteins is their ability to form micelles at CMC (critical micelle concentration).

1.2.10.3 Inclusion bodies.

Bacterial protein production in *E. coli* provides a fast, inexpensive method of generating high concentrations of expressed protein in comparison to mammalian cell expression. However, proteins produced by this method in *E. coli* frequently yield an inactive, aggregated form of the protein called inclusion bodies. Production of inclusion bodies is considered to be
somewhat of undesirable but with some advantages; the protein is of high purity and concentration and has increased protection from proteolytic degradation. Inclusion bodies can be isolated with low speed centrifugation of lysed *E. coli* cells (De Bernardez *et al.*, 1999). Inclusion bodies are then solubilised using chaotropic agents such as guanidium hydrochloride or urea. This process causes the protein to lose its native or aggregated conformation. In order to regain its native structure, the protein must be transferred to a non-denaturing environment. This can be achieved by a number of methods such as dilution where the denaturant is diluted to levels which are conducive to protein folding or membrane controlled denaturant removal which utilises dialysis or filtration techniques to remove denaturant (Vallejo and Rinas, 2004). This process may result in aggregates as folding intermediates form when the denaturant is gradually removed. Chromatography techniques such as size exclusion chromatography, which are frequently used to exchange denaturing buffering for refolding buffer, can also be used for refolding (Vallejo and Rinas, 2004). Matrix assisted protein refolding is also commonly used. This technique involves binding of the urea solubilised protein to a matrix which can withstand the use of chaotropic agents. It is achieved by binding of the His-tagged protein to an ion exchange matrix or using immobilised metal affinity chromatography techniques. Finally refolding using hydrophobic interaction chromatography (HIC) has also been used (Vallejo and Rinas, 2004). During this process, unfolded proteins are loaded onto the column at high salt concentrations and eluted using a low salt gradient. It is thought that HIC facilitates refolding as unfolded proteins absorb onto the matrix and prevent aggregation, also hydrophilic regions bind to the matrix creating micro-domains where localised folding can occur (Vallejo and Rinas 2004).

1.2.10.4 Factors influencing refolding

Refolding can be influenced by several factors. Temperature has been shown to influence refolding. Low temperature has been shown to decrease aggregation of hydrophobic domains while decreasing the rate of refolding. L-arginine is commonly used as an additive in refolding studies which utilise its ability to shield hydrophobicity of partially folded proteins (De Bernardez *et al.*, 1999).
Figure 1.8: Mechanism of action for an alkaline phosphatase (Holtz et al., 1999)
2. Production and characterisation of His-tagged full length autotransporters.

2.1 Objectives

1. Expression and purification of selected His-tagged autotransporters from *P. mirabilis*.
2. Characterisation of the expressed proteins using SDS-PAGE and determination of their activity.

2.2 Materials and Methods

2.2.1 Reagents

All of the chemicals, reagents and media used in this project were supplied by Sigma Aldrich, New England Biolabs, GE Healthcare, JT Baker, Takara, Roche, Thermo Fisher Scientific, Melford, BDH and Biorad unless otherwise stated. Vector pET 21a+ was kindly donated by Dr. Jakki Cooney, University of Limerick. *E. coli* Top10 and BL21(DE3) strains were produced from stock cultures available in the lab. All primers were ordered from Eurofins MWG Operon GmbH.

2.2.1.1 Bacterial Strains

*Proteus mirabilis* HI4320

For this study, cultures of *P. mirabilis* H14320 were grown in a class II EPA licensed laboratory at 37 °C with shaking in LB media.

2.2.1.2 Preparation of competent cells

Appropriate *E. coli* strains were grown overnight in 3 ml of LB broth for 16 h at 37 °C in a shaking incubator. 3 ml of the overnight culture was inoculated in 300 ml of LB broth (1:100) and grown for approximately 3.5 h at 37 °C and 250 rpm until an OD of 0.6 at 600 nm was
reached. The cell culture was then incubated on ice for 20 min and harvested at 4 °C at 3300 x g for 20 min. The supernatant was then discarded and the cell pellet was resuspended on ice with 300 ml of 1 mM HEPES [pH 7]. The cells were then recovered by centrifugation at 3300 x g at 4 °C for 20 min and the pellet was resuspended with 100 ml of 10 % glycerol, this step was repeated twice with 50 ml and 20 ml volumes of 10 % glycerol. Finally the cells were resuspended in 1 ml 10% glycerol, aliquoted (50 μl) into sterile Eppendorf tubes and frozen in liquid nitrogen and stored at –80 °C.

Competent cell lines used in this study.

E. coli TOP10

*Genotype:* F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^−

*Strain characteristics:* Non-expression host; used for cloning and plasmid propagation; streptomycin resistant.

E. coli BL21(DE3)

*Genotype:* F−ompT gal dcm lon hsdS^B(rB−mB−) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

*Strain characteristics:* Expression host with DE3, a λ prophage carrying the T7 RNA polymerase gene (under control of the lacUV5 promoter) and lacI^q; transformed plasmids containing T7 promoter driven expression are repressed until IPTG induction of T7 RNA polymerase from a lac promoter; deficient in *lon* and *ompT* proteases.

2.2.1.3 Genomic DNA preparation

Genomic DNA from *P. mirabilis* H14320 was prepared using the Qiagen “QIAGEN Genomic-tip 100/G Kit” according to the manufacturer’s instructions. The column contained anion-exchange resin which purifies high-molecular-weight DNA from a wide range of biological samples without the use of phenol or chloroform. DNA binds to the column while other cell constituents pass through.

5 ml of cells from a 16 h culture were harvested at 1,500 x g for 10 min. The cell pellet was washed twice in chilled PBS and resuspended using 3 ml of chilled C1 lysis buffer. The cell lysate was then centrifuged at 1300 x g for 15 min at 4 °C and the supernatant was discarded.
The pellet was then resuspended in 1 ml of ice-cold buffer C1 and distilled water to remove all cellular debris before centrifuging at 1300 x g for 15 min at 4 °C. The pellet was then completely resuspended in 5 ml buffer G2 and vortexed at maximum speed before addition of 95 μl of protein kinase K stock solution and incubating at 50 °C for 60 min.

A QIAGEN genomic tip 110/G was then equilibrated using 4 ml of buffer QBT and the resuspended pellet was loaded. The column was then washed twice with 2 x 7.5 ml of buffer QC, before elution with 5 ml buffer QF. The DNA was then precipitated using 3.5 ml of room temperature isopropanol and centrifuged at >5000 x g for at least 15 min. The pellet was then washed with 70 % ethanol, vortexed and centrifuged under the same conditions. The ethanol was then removed and the pellet was air dried and resuspended in 2 ml of distilled water. The purity of the extracted genomic DNA was then tested using the nanodrop 3300 system.

2.2.1.4 PCR amplification of autotransporter specific DNA from *P. mirabilis* H14320 genomic DNA

PCR reactions were conducted using ~ 50 ng of genomic DNA from *P. mirabilis* H14320. The DNA polymerase used was Phusion polymerase from Finnzyme. Primers used in the cloning process were designed using the Bioinformatics analysis software “Vector NTI” from Invitrogen. This software permits the user to view the gene of interest to be ‘ligated’ into any one of a number of plasmids contained in the database. Analysis using this programme allows the user to ensure that the primer design is accurate. Once suitable primers are chosen, the melting temperature (Tm) of the primers can be calculated using the oligo analysis tool, which detects palindromes in the primer sequence as well as repeating nucleotides. The oligonucleotides used in this study were purchased from “Eurofins MWG Operon GmbH”.

The Flexigene Techne thermocycler was used for all PCR amplifications. Primer synthesis and sequencing was performed and supplied by Eurofins MWG Biotech. The PCR reaction used was 98 °C for 30 s, followed by 10 cycles of 98 °C for 10 s, 50°C for 30 s, 72°C for 2 min. This set of cycles was then followed by another denaturing step at 98 °C for 30 s followed by 30 cycles of 98 °C for 30 s, 55 °C for 30 s and 72 °C for 2 min before a final extension step of 72 °C. The Phusion polymerase utilised throughout this study was used with optimised buffer solutions supplied with the enzyme, buffer HF and buffer GC. Buffer HF is
a high fidelity buffer and minimises non-specific amplification while buffer GC is intended for use in genes with a high GC content, i.e. proteins from thermophilic bacteria.

PCR products were purified using the Qiagen “QUAEX II Gel extraction kit” and “Nucleotide removal Kit “which separated DNA from unwanted reaction constituents. The QIAEX II Gel Extraction Kit consists of a column containing silica to which DNA fragments bind in the presence of chaotropic salts. The bound DNA is then washed to remove salts and other contaminants using buffers provided with the kit. DNA is finally eluted in Tris buffer or water.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI0844/pET21a+_fow</td>
<td>A CTG ACT CAT ATG GCT TAT TTT ATG AAA AAT ATC GTA ATA CCT ACC TCG</td>
<td>Forward primer for PMI0844 containing Nde I restriction site.</td>
</tr>
<tr>
<td>PMI0844/pET21a+_rev</td>
<td>CAG TGA GAA TTC TTA ATG ATG ATG ATG ATG ATG ATG GAA GCG GAT AGA GAG TTT CAT ATC AAT AC</td>
<td>Reverse primer for PMI0844 containing an Eco RI restriction site, stop codon and a His_6 tag.</td>
</tr>
<tr>
<td>PMI2126/pET21a+_fow</td>
<td>A CTG ACT CAT ATG AAA AAT AAA AAT TTA TTA TTA AGT GCT GCT GCT G</td>
<td>Forward primer for PMI2126 containing Nde I restriction site.</td>
</tr>
<tr>
<td>PMI2126/pET21a+_rev</td>
<td>CAG TGA CTC GAG GAA ACG CTG GCT AAC AGT AAA G</td>
<td>Reverse primer for PMI2126 containing Xho I restriction site.</td>
</tr>
<tr>
<td>PMI2575/pET21a+_fow</td>
<td>A CTG ACT CAT ATG AAA ACG ACG GGA GTT AAA GTA ACG CTT TTG</td>
<td>Forward primer for PMI2575 containing Nde I restriction site.</td>
</tr>
<tr>
<td>PMI2575/pET21a+_rev</td>
<td>CAG TGA CTC GAG CCA GCC CAC CGC AAA CCC G</td>
<td>Reverse primer for PMI2126 containing Xho I restriction site.</td>
</tr>
<tr>
<td>PMI2341/pET21a+_fow</td>
<td>TG AAC AAA GAA ATA GCT TTA AGT CAC CAT CC</td>
<td>Forward primer for PMI2341 containing a partial Nde I restriction site for blunt cloning.</td>
</tr>
<tr>
<td>PMI2341/pET21a+_rev</td>
<td>CAG TGA CTC GAG GAA ATT AAT TTT CAA TAT TGC ATT AAT ACC ACT CG</td>
<td>Reverse primer for PMI2341 containing Xho I restriction site.</td>
</tr>
<tr>
<td>PMI2174/pET21a+_fow</td>
<td>TG AAT TAT AAT AAA TTA TTT TCT ATT TCT TTC TCT TTA ATT TAT TCT GC</td>
<td>Forward primer for PMI2174 containing a partial Nde I restriction site for blunt cloning.</td>
</tr>
<tr>
<td>PMI2174/pET21a+_rev</td>
<td>CAG TGA CTC GAG CCA ACT AAT ACC GAT CCC CCC CGC AGA G</td>
<td>Reverse primer for PMI2174 containing Xho I restriction site.</td>
</tr>
</tbody>
</table>

Table 2.1 List of primers used in this study
2.2.1.5 Restriction digestion of DNA

Restriction digestions were carried out as described in Sambrook and Russell (2001) using the following restriction enzymes – *BamHI*, *NdeI*, *XhoI*, *EcoRI* and *BglII*. All enzymes were used according to the manufacturer’s instructions. Digestions for cloning reactions were carried out at 37 °C in a water bath for 16 h while digestion for screening reactions was carried out at 37 °C for 3 h. All enzymes used in this study were purchased from Roche Bioscience unless otherwise stated. Digestions were analysed on agarose gels and purified via electrophoresis and the QIAEX II Gel Extraction kit.

Restriction enzymes used in this study.

*BamHI* (source: *Bacillus amyloliquefaciens*; reaction conditions: 37 °C in Roche restriction buffer B), *NdeI* (source: *Neisseria denitrificans*; reaction conditions: 37 °C in Roche restriction Buffer H), *XhoI* (source: *Xanthomonas holecilcola*; reaction conditions: 37 °C in Roche restriction buffer H), *EcoRI* (source: *Escherichia coli* BS5; reaction conditions: 37 °C with Roche restriction buffer H), *BglII* (source: *Bacillus globigii*; reaction conditions: 37 °C in Roche restriction buffer.

2.2.1.6 DNA Ligation

Ligation was carried out as described in Sambrook and Russell (2001). Insertion of the digested insert into digested plasmid DNA was conducted using T4 DNA Ligase (New England Biolabs). An insert to vector ratio of 4:1 was generally used, but this was modified as necessary. Reactions were carried out at 16 °C for 16 h for optimum ligation.

DNA precipitation was used typically to remove salt from the ligation reaction mixture which affects the electro conductivity of the cells. The DNA was first transferred into a sterile Eppendorf tube, to which 1/10 of the volume of 3 M sodium acetate pH 5.2 and 2.5 times the volume of ice cold ethanol 100 % were added. The mixture was then inverted to ensure uniformity before incubation at -20 °C for 2 h. The DNA precipitate was then centrifuged at 24,500 x g for 30 min at 4 °C. The supernatant was then removed and the DNA pellet was washed with 200 μl of 70 % ethanol and centrifuged at 24,500 x g for 30 min at 4 °C. The
supernatant was then discarded and the pellet was dried at room temperature for 30 min to remove residual ethanol.

2.2.1.7 Transformation

Transformation was performed by electroporation which was carried out according to the method described in Sambrook and Russell 2001. An aliquot of electro competent cells (50 μl) was removed from -80 °C, thawed on ice and placed in a 1.5 ml sterile Eppendorf tube. 1 μl of plasmid DNA was then added to the cells and transferred into an electroporation cuvette followed by 1 min incubation on ice. The outside of the cuvette was then dried and placed in the electroporator where the cells were subjected to 1800 V of electricity. Once shocked, 200 μl of warm LB (37 °C) was added and the cells were transferred to a 15 ml sterile tube for incubation at 37 °C for 45 min. The transformed cells were then plated on LB agar petri plates containing the appropriate antibiotic. The plates were incubated at 37 °C overnight upside down.

2.2.1.8 Plasmid Screening

Transformants were screened for the presence of recombinant plasmids using the Qiagen “QIAprep Spin Miniprep Kit”. Plasmid purification using QIAprep Kits follows a simple bind-wash-elute procedure. Single colonies were picked from a transformation plate and grown overnight in 5 ml LB broth containing the appropriate antibiotic. 1 ml of each overnight culture was transferred into a 1 ml Eppendorf tube. The cells were then harvested by centrifugation for 1 min at 13000 x g. The cell pellet was first lysed and then centrifuged. The lysate was then applied to the QIAprep column where plasmid DNA adsorbs to the silica gel membrane. The bound DNA was then washed and pure DNA was eluted in a small volume of elution buffer or water. The QIAprep Spin MiniPrep Kit enables purification of up to 20 μg molecular biology grade plasmid DNA. The extracted plasmids were analysed on an agarose gel after restriction digestion to determine if the ligation reaction was successful and also to determine DNA purity and concentration. This was typically done on 1 % agarose gel prepared with Tris, Acetate and EDTA (TAE) buffer with 3 μl of SYBR safe DNA stain. DNA samples were blended with a 1/10 volume of loading buffer, mixed and then loaded into the gel. A molecular weight marker of appropriate size was also loaded into the gel. The
gels were usually run at 90 V for 1 h, depending on the size of the samples to be analysed. The samples were then visualised under a UV lamp and the gel was photographed using the Kodak Digital Science package (version 203) together with a digital camera (Eastman Kodak Company, 1997)

The purity and concentration of extracted DNA was also determined using the Nanodrop 3300 spectrophotometer. A 1 μl sample was pipetted onto the end of a fiber optic cable (the receiving fiber). Then a second fiber optic cable (the source fiber) was brought into contact with the liquid sample causing the liquid to bridge the gap between the fiber optic ends. A pulsed xenon flash lamp provided the light source and a spectrometer utilizing a linear CCD array was used to analyze the light after passing through the sample. The instrument was controlled by PC based software and the data was logged in an archive file on the PC. The amount of DNA in the sample was measured at 260 nm, while the purity of the sample was based on the 260/280 nm ratio. A ratio of 1.8 was generally accepted as being pure for a sample, if the ratio was lower it indicated the presence of protein or phenol contamination which absorbs strongly in the 280 nm range.

2.2.2 Protein expression

2.2.2.1 Expression of recombinant protein in *E. coli*.

Unless otherwise specified, induction of recombinant protein production was performed according to the protocol set out below. An isolated colony was picked and grown overnight in LB broth with 100 μg/ml ampicillin. Overnight cultures were diluted in 1:100 in 50 ml LB with 100 μg/ml ampicillin. Cultures were grown by shaking at 37 °C until an OD₆₀₀ of ~ 0.5 – 0.6 was reached. Cultures were adjusted to 1 mM IPTG for 3 h and further incubated at 25 °C. Growth behaviour was monitored by checking the OD₆₀₀ and taking a sample which contained the same amount of cells as would be contained in a 1 ml sample at OD₆₀₀ of 1.0. Protein expression was analysed using SDS-PAGE and Western blotting where applicable. Expression of the protein was verified by running an uninduced sample against induced samples which were taken each hour over a 4 h period. The size of the protein if contained on the gel was estimated using appropriate molecular weight markers.

Expression cells were harvested by centrifuging at 3000 x g for 20 min at 4 °C. The supernatant was discarded and the pellet frozen at – 80 °C. The cell pellets were thawed at 4
°C and then weighed. Cells were resuspended in ten times the volume of cells present with phosphate buffer pH 7.4. EDTA was then added to final concentration of 2 mM and left shaking for 10 min before the addition of lysozyme at 1 mg/ml. After the addition of lysozyme, the solubilisation mixture was left stirring for 10 min before being transferred to a 100 ml beaker. The suspension was then sonicated using 30 % power for 3 x 10 min intervals. After sonication, LDAO was added to a final concentration of 0.01 %. 1 μg/ml DNase and 3 mM MgCl₂ were then added to remove DNA from the lysed cell culture. The solubilisation was left at 4 °C with stirring for 1 h, before centrifuging at 11,000 x g for 10 min. The supernatant was then removed, filtered through a 0.22 micron filter and purified by immobilised metal affinity chromatography.

2.2.2.2 SDS-PAGE

Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to Lämmeli, (1970). The running gel was poured into the sealed gel cast (9 x 6 cm) and overlaid with distilled water. Polymerisation was allowed for at least 30 min. The water was then discarded after the gel had solidified and the stacking mixture was then poured on top of the running gel, the comb was then placed into the gel cast and the gel was allowed to set for an additional 30 min. The gel was then placed in the electrophoresis unit and 1 x running buffer was added to the electrophoresis unit. The protein samples were then loaded into the wells contained in the gel after being denatured at 90 °C for 10 min. The lid was then placed on the unit and the gel was run at 30 mA for 40 min until the dye migrated to near the bottom of the gel. After this process, the gel was removed from the tank and stained using Instant blue staining solution.

2.2.2.3 Affinity chromatography

Raw protein extract containing 10 mM imidazole was loaded onto Ni sepharose 6 fast flow resin (GE Healthcare). 10 ml of the resin was previously packed into a column and equilibrated using three times the resin volume of phosphate buffer pH 7.4, 500 mM NaCl, 10 mM imidazole. The protein extract was loaded at 1.5 ml/min, the resin was then washed with 10 mM and 60 mM imidazole with a final elution with 500 mM imidazole. Eluted fractions were then pooled and concentrated. The buffer was then exchanged by use of a PD10 column or gel filtration for a low salt phosphate buffer (20 mM phosphate pH 7.4, 150 mM NaCl).
2.3 Results and discussion

2.3.1 Selection of autotransporters from *P. mirabilis* H14320

Five autotransporter proteins from *P. mirabilis* H14320, with implications in virulence, were initially targeted for structural studies (Alamuri and Mobley 2008). The main target of this study was 2341, an outer membrane protease autotransporter, which was shown to be expressed in an active form and was localised at the cell surface. This protein was targeted for overexpression and structural characterisation along with two putative autotransporters annotated as proteases; 0844 and 2126, respectively and three others putative adhesins; 2122, 2174 and 2575 respectively. Secondary structure analysis of each of these proteins show that they are of two different subclasses: 0844, 2126, 2341 and 2174 are type Va autotransporters while 2575 and 2122 are type Vc autotransporters.

The gene sequence of these proteins was assessed in the sequenced genome of *Proteus mirabilis* strain H14320 (Alamuri and Mobley 2008). They were analysed for their suitability for cloning into pET21a+ vector using Vector NTI software from Invitrogen. 2122 was not chosen initially as a target as it was found to have an aberrant start codon. An Nde I site in the 2174 and 2341 genes posed difficulty meaning that if this restriction enzyme was utilised then blunt cloning had to be employed. In this method, the primer was designed with a partial NdeI site which formed a full NdeI site when the gene was ligated into blunted pET21a+ vector. The vector for blunt cloning was previously digested with NdeI and blunted using a blunting kit from Takara before being digested again with the other enzyme (XhoI) used for cloning.

0844 was cloned using NdeI and EcoRI as XhoI which was used in the design of the other cloning experiments was found to cut the gene sequence of 0844. The reverse primer of the gene was designed to encode a His-tag required for protein purification after expression; this was favoured instead of the enteric His-tag on the pET21a+ vector as it would have resulted in a shorter ‘tail’ than that would have occurred if the His-tag was plasmid encoded. The 2575 and 2126 genes were designed to be cloned in the classical manner using Nde I and XhoI enzymes into the plasmid pET21A+.
2.3.1.1 Cloning of targets from *P. mirabilis*

PCR products were obtained at the required size for the amplified targets: 1813 bp for 2174, 2326 bp for 2126, 2386 bp for 2575, 3149 bp for 0844 and 3415 bp for 2341 (figure 2.4). These amplicons were purified using the Qiagen “QUAEX II Gel extraction kit” which permitted desalting of the PCR products before digestion with the appropriate restriction enzyme. The resulting digestion was run on a 0.5 % agarose gel. The digested fragments were extracted before purification with the Qiagen “QUAEX II Gel extraction kit”.

<table>
<thead>
<tr>
<th>Name</th>
<th>Autotransporter description</th>
<th>MW(KDa)</th>
<th>PI</th>
<th>AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI O844</td>
<td>acid phosphatase</td>
<td>108.6</td>
<td>6.83</td>
<td>983</td>
</tr>
<tr>
<td>PMI 2126</td>
<td>serine protease</td>
<td>79.5</td>
<td>4.93</td>
<td>729</td>
</tr>
<tr>
<td>PMI 2341</td>
<td>Protease</td>
<td>120.5</td>
<td>5.49</td>
<td>1106</td>
</tr>
<tr>
<td>PMI 2174</td>
<td>Yada family agglutinin, adhesion</td>
<td>61.6</td>
<td>5.18</td>
<td>558</td>
</tr>
<tr>
<td>PMI 2575</td>
<td>HGia, Yada superfamly, adhesion, STEC autoagglutinin</td>
<td>82</td>
<td>5.61</td>
<td>749</td>
</tr>
</tbody>
</table>

Figure 2.4: Table of autotransporter targets from *P. miabilis*

Figure 2.5: PCR of selected genes from *P. mirabilis* analysed on a 1 % agarose gel under UV light. Lane M contains 1Kb molecular weight marker and lanes 1-5 contain the amplicons for the project targets; lane1: pmi 0844, lane 2: pmi 2341, lane 3: pmi 2126, lane 4: pmi 2174, lane 5: pmi 2575.
2341 and 2174 needed ligation into a blunted vector for successful ligation and expression to occur. Extracted pET21a+ plasmid was firstly digested using NdeI and then underwent a blunting reaction using a blunting kit from Takara. The blunted pET21a+ was then digested using the other enzyme required for cloning in which was XhoI. The other genes were cloned using ‘sticky end’ ligation using restriction enzymes NdeI and XhoI for 2575 and 2126, NdeI and EcoRI for 0844. After digestion, the PCR products were ligated with T4 ligase into pET21a+ vector, previously digested with the same enzymes. The ligation was precipitated using ethanol precipitation, transformed into *E. coli* TOP10 cells, and screened for plasmid uptake using ampicillin. Isolated colonies were grown overnight and plasmids were extracted and screened for insert uptake using XbaI and XhoI. Three of the initial five targets identified were cloned into pET21a+ vector (figure 2.5) - 0844, 2575 and 2174. The main target of the study, 2341 was not cloned. This may have been due to a number of factors *i.e.*, the low concentration of vector after blunting of the pET21a+ vector, combined with the low yield of blunt cloning and size of the gene at 3415 bp would have made cloning of this gene particularly difficult. The positive clones were transformed into *E. coli* BL21 for expression as TOP10 cells lacked the necessary intrinsic T4 polymerase.

![Figure 2.6: Screening for uptake of plasmid after transformation. Lane M contains 1 Kb molecular weight marker and lanes 1-5 depict the positive clones, lane 1 and 2: pmi 2575, lane 3: pmi 2174, lane 4: pmi 0844, lane 5: pmi 2174.](image)
2.3.1.2 Expression studies

Expression trials for all three clones were carried out in order to determine the best method of expression. Small scale expressions were performed with positive clones as described in section 2.3.2.1. Samples of the expression were collected at each hour for the length of the expression trial. Samples were normalised i.e., they were taken in such a manner that each sample in the time course contained the same amount of cells as at 1.0 OD\textsubscript{600}. The samples were run on a 10 % SDS-PAGE gel and Western blot for analysis. A band was detected on the SDS-PAGE (lane 1 figure 2.6, lane 1-3 figure 2.7.) corresponding to the expressed 0844 which at greater than 80 kDa was close to the actual value of 108 kDa. A band was also detected between 58 and 80 kDa for expression of the 2174 full length protein (lane 2 figure 2.6 and lanes 4-6 figure 2.7), the actual molecular weight of this protein is 62 kDa putting it within this range. Expression samples for 2575 also exhibited a band at 80 kDa (figure 2.6 lanes 7 and 8) when run on an SDS-PAGE gel which is close to the actual value of 82 kDa.

![SDS-PAGE gel with bands at 80kDa, 58-80kDa, and 82kDa for different clones](Image)

Figure 2.7: Expression of pET21a+-044, pET21a+-2174, pET21a+-2575 and pET21a+-2162 after three hours with 1 mM IPTG. Lane M: molecular weight marker, lane 1: pET21a+-0844, lane 2,3: pET21a+-2174, lanes 4-6: pET21a+-2134, lane 7,8: pET21a+-2575, lane 9 pET21a+ negative control.
Figure 2.8: Time course expression of 0844 and 2174. Lane M: Molecular weight marker, lane 1 -3: expression of pET21a+-0844 at 1,2, 3 h in LB broth with 1mM IPTG. Lanes 4- 6: Expression of 2174 in LB broth with 1 mM IPTG at 1, 2, 3 h.

None of the proteins gave a signal when run on a Western blot, indicating that the His-tag to which the antibody was raised was inaccessible. It was evident from expression trials that the protein was produced, but needed to be extracted for characterisation. As the protein is transported to the outer membrane, LDAO was used for solubilisation and purification.

2.3.1.3 Purification of full length autotransporters 0844, 2174 and 2575.

Cells expressing the protein of interest were subjected to solubilisation using LDAO. Samples were taken from each stage of the solubilisation for analysis. Each protein showed the same characteristic solubilisation (figure 2.8, 2.9, 2.10): (i) As shown in lane 1, all proteins were present in the initial protein pellet, (ii) little or none of the protein was found to be present in the cell lysis or flow through of the Ni-NTA chromatography (lanes 2 and 3), (iii) a large amount of the protein was contained in the lysis pellet (lane 4), indicating that the expressed protein was mainly composed of inclusion bodies.
Figure 2.9: 2174 Purification from whole cell extract after solubilisation with Triton X-100. Lane M: Molecular weight marker. Lane 1: before lysis, lane 2: after lysis, lane 3: flow through, lane 4: pellet, lane 5: 20 mM imidazole, lane 6: 40 mM imidazole, lane 7: 250 mM, lane 9: negative control.

Figure 2.10: 0844 Purification from whole cell extract after solubilisation with Triton X-100. Lane M: Molecular weight marker. Lane 1: before lysis, lane 2: after lysis, lane 3: flow through, lane 4: pellet, lane 5: 20 mM imidazole, lane 6: 40 mM imidazole lane 7: 250 mM imidazole, lane 8: control.

Figure 2.11: 2575 Purification from whole cell extract after solubilisation with Triton X-100. Lane M: molecular weight marker, lane 1: before lysis, lane 2: after lysis, lane 3: flow through, lane 4: pellet, lane 5: 20 mM imidazole, lane 6: 40 mM imidazole, lane 7: 250 mM imidazole, lane 8: control.
When the cell lysate was loaded on a nickel column it was found that there was no protein contained in the elution fraction, with the exception of 2575, it was thought however that this protein was an artifact of cell lysis and not the protein of interest (lane 7, figure 2.10). The protein may be produced however in very low quantities however it is difficult to tell, as the protein did not give a signal on a western blot.

2.4 Conclusion

All of the five targets were successfully amplified. Three of the five initial targets were successfully cloned into pET21a+. The positive clones containing 0844, 2575 and 2174 were transformed into BL21(DE3) cells and were expressed using IPTG in *E. coli* strain BL21(DE3). Each of the positive clones was shown to express protein of the correct molecular weight and so the project progressed into large scale expression. Pellets containing selected targets were solubilised in LDAO and purified on an IMAC column. The protein was found to be expressed in inclusion bodies. It was determined that the strategy used needed to be reviewed and it was decided to focus on the passenger domain for each of the targets.
3. Production and characterisation of His-tagged autotransporter passenger domains

3.1. Objectives

1. Expression and purification of selected His tagged passenger domains of autotransporters from *P. mirabilis* using nickel affinity chromatography.
2. Characterisation of the expressed protein using SDS-PAGE, florescence and determination of activity.
3. Crystallisation trials utilising purified proteins.

Experimental

3.2. Materials and methods

Materials, DNA manipulation and cloning techniques are the same as outlined in chapter 2, section 3.1; any variations will be outlined and discussed.

3.2.2 PCR amplification of DNA from *Proteus mirabilis*

Two different primer sets were used to amplify the gene encoding the passenger domain of the targets with subsequent protein production in *E. coli* with two distinct methodologies: (i) generation of inclusion bodies for refolding and (ii) export the folded nascent protein to the periplasm.

As described before, genes encoding autotransporters consist of three parts – the signal sequence, the passenger domain and the transmembrane domain. The method for producing the passenger domain as inclusion bodies requires removal of both the signal sequence and the transmembrane domain. Primers were designed to amplify the gene sequence of 0844, 2575 and 2174 excluding the gene sequences associated with the signal sequence and the transmembrane domain. As was previously used to express the full length protein, primers were designed for cloning of the gene of interest into pET21a+ vector. Both the 2575 and the 2174 genes were cloned in the classical manner, while 2174 required the use of blunt ended cloning due to an NdeI site within the sequence of the gene.
The second strategy used for production of the passenger domain required cloning of the 0844 and 2575 proteins into vector p15ara which is based on the pBAD arabinose inducible vectors. This method involved cloning the passenger domain and associated signal sequence, such that upon expression the protein would migrate to the periplasm. This vector was chosen to allow greater control of protein expression; it has a low copy number when compared to the T7 expression system of the pET system.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI2575pass/pET21a+_fow</td>
<td>A CTG ACT CAT ATG AAA AAC GTG TCA CTG GGG GAC ACG</td>
<td>Forward primer for PMI2575 passenger domain containing Nde I restriction site.</td>
</tr>
<tr>
<td>PMI2575pass/pET21a+_rev</td>
<td>CAG TGA CTC GAG TAA CAC CTG CCC GAA CGT CTC ATC G</td>
<td>Reverse primer for PMI2575 passenger domain containing Xho I restriction site and a stop codon.</td>
</tr>
<tr>
<td>PMI2174pass/pET21a+_fow</td>
<td>TG AAT ACT GAT AGC CCA TTA CCA ACA GAT AAT CC</td>
<td>Forward primer for PMI2174 passenger domain containing a partial Nde I restriction site for blunt cloning.</td>
</tr>
<tr>
<td>PMI2174pass/pET21a+_rev</td>
<td>CAG TGA CTC GAG TTA TAA CTT ATT TCC TGC ATA TAA AGA GCC AAG ATG</td>
<td>Reverse primer for PMI2174 passenger domain containing Xho I restriction site and a stop codon.</td>
</tr>
<tr>
<td>PMI0844pass/pET21a+_fow</td>
<td>ACTG ACT CAT ATG GAA ACC AAT ATT TCA CAA ACC AAA CAA CGA ATT ATC</td>
<td>Forward primer for PMI0844 passenger domain containing Nde I restriction site.</td>
</tr>
<tr>
<td>PMI0844pass/pET21a+_rev</td>
<td>CAG TGA GAA TTC TTA GTT TTT TGT TTT AGC CAC ACT TGA AAA GG</td>
<td>Reverse primer for PMI0844 passenger domain containing Xho I restriction site and a stop codon.</td>
</tr>
<tr>
<td>PMI0844pass/p15ara_rev</td>
<td>CAG TGA CTC GAG GTT TTT TGT GTC CAC ACT TGA AAA GGC GGT</td>
<td>Reverse primer for PMI0844 containing Xho I restriction site.</td>
</tr>
<tr>
<td>PMI2575pass/p15ara_rev</td>
<td>CAG TGA CTC GAG CAC CTG CCC GAA CGT CTC ATC</td>
<td>Reverse primer for PMI2575 passenger domain containing Xho I restriction site.</td>
</tr>
<tr>
<td>PMI2575/p15ara_rev</td>
<td>CAG TGA AGA TCT TTA ATG ATG ATG ATG ATG ATG GAA CAC CAC CGC AAA CCC G</td>
<td>Reverse primer for PMI2575 full length autotransporter containing BglII restriction site, a Histag and a stop codon.</td>
</tr>
</tbody>
</table>

Table 3.1: List of Primers used in this study
3.2.3 Protein expression

3.2.3.1 Expression of recombinant protein in *E. coli*.

Unless otherwise specified, initial recombinant protein production for 0844p and 2575p pET21a+ in *E. coli* was performed according to the protocol set out below. An isolated colony was picked and grown overnight in LB broth with 100 μg/ml ampicillin. Overnight cultures were diluted in 1:100 in 50 ml LB with 100 μg/ml ampicillin. Cultures were grown by shaking at 37 °C until an OD$_{600}$ of ~ 0.5 – 0.6 was reached. Cultures were adjusted to 1 mM IPTG for 3 h and further incubated at 25 °C. Growth behaviour was monitored by checking the OD$_{600}$ and taking a sample which contained the same amount of cells as would be contained in a 1 ml sample at OD$_{600}$ of 1.0. Protein expression was analysed using SDS-PAGE and Western blot. Expression of the protein was verified by running an uninduced sample against induced samples which were taken each hour over a 4 h period. The size of the protein if contained on the gel was then estimated using molecular weight markers.

Large scale expression for 2575p/pET21a+ and 0844p/pET21a+ was performed using autoinducing media according to the method outlined by Studier, (2005). This method expresses the protein of interest typically at an OD$_{600}$ of 5-6, the higher cell density produces a greater amount of protein for use in characterisation studies when compared with the classical IPTG induction. It is also advantageous in that the OD$_{600}$ of the protein does not have to be monitored and multiple expression trials can be run in parallel.

Expression of 2575p/p15ara and 0844p/p15ara was carried out using the following method unless otherwise stated. An isolated colony was picked and grown overnight in LB broth with 35 μg/ml chloramphenicol. Overnight cultures were diluted in 1:100 in 50 ml LB with 35 μg/ml chloramphenicol. Cultures were grown with shaking at 37 °C until an OD$_{600}$ of ~ 0.5 – 0.6 was reached. Cultures were adjusted to 0.025 % arabinose for 3 h and further incubated at 25 °C. Growth behaviour was monitored by checking the OD$_{600}$ and taking a sample which contained the same amount of cells as would be contained in a 1 ml sample at OD$_{600}$ of 1.0. Protein expression analysis was performed as described above.
3.2.3.2 SDS-PAGE and Western Blot

Samples were separated by SDS-PAGE electrophoresis as described in Sambrook and Russell, (2001). Protein from the expression were electrophoretically transferred from the SDS-PAGE gel to nitrocellulose filter membrane using an ATTO corporation AE-6670 semi-dry electroblotter in transfer buffer. Whatman 3 mm filter paper and nitrocellulose membrane, 0.2 μm pore size (Bio-Rad) were first cut in shape and size enough to cover the gel. Nitrocellulose membrane was left in transfer buffer for 5-10 min, the filter paper was also soaked in transfer buffer and everything was placed in the electro-blotter in the following order: 3 pieces of Whatman 3 mm paper, nitrocellulose paper, gel and finally the other 3 pieces of Whatman paper. Blotting was performed at 60 mA for 1 h.

After blotting the membrane was soaked in the blocking solution for 1 h at room temperature with gentle rocking at room temperature and left overnight at 4 °C. The blocking solution was removed and the membrane was washed 3 times with TBS-Ca\(^{2+}\) at room temperature, with gentle rocking. Anti-His-tag antibodies conjugated to horseradish peroxidase, diluted 1:1000 times in 30 ml TBS-Ca\(^{2+}\) with 12.5 μl of Tween 20, were poured on the membrane and incubated at room temperature for 1 h with gentle rocking. After additional washing steps, the membrane was stained by overlaying the membrane with the chromogenic TMB (3,3’, 5,5’tetramethylbenzudine peroxidase substrate for membrane detection): reaction was usually allowed to proceed for 5 min, then stopped by adding distilled water.

3.2.3.3 Soluble/Insoluble preparation

Frozen samples were resuspended in 20 mM Tris HCl pH 8.0, 5 mM EDTA and 1% LDAO. Then 2 μl of 50 mg/ml lysozyme was added to the above buffer and incubated at room temperature for 5 min. MgCl\(_2\) was added at a final concentration of 10 mM and DNase at a final concentration of 1 mg/ml and incubated for 5 min at RT. The cell lysate was then spun at 24,500 x g, for 20 min at 4 °C. The supernatant contained the soluble protein while the pellet contained insoluble inclusion bodies. The pellet was then washed using the lysis buffer and solubilised in 8 M urea in 50 mM Tris-HCl pH 7.4. Both fractions were then loaded on an SDS gel and Western blot for analysis.
3.2.3.4 Preparation of inclusion bodies for refolding

Protein was first expressed with autoinducing media overnight for 16 h at 25 °C. The cells were harvested at 3300 x g for 20 min at 4 °C, washed using 10 mM HEPES pH 7 and centrifuged again with the same conditions. The pellet was resuspended using ten times the volume of the cell mass with 50 mM Tris pH 8 before the addition of EDTA at 5 mM final concentration and lysozyme at 1 mg/ml final concentration. The mixture was then incubated at room temperature for 20 min before the addition of MgCl₂ at 10 mM final concentration and DNase. Finally Triton X-100 was added at 5% concentration to lyse the cells and left overnight at 4 °C with stirring. The cell lysis was then spun down at 11,000 x g for 20 min at 4 °C and the supernatant removed. The pellet was washed three times using 50 mM Tris-HCl pH 8 with 0.5% Triton X-100. The resulting pellet was then solubilised using 8 M urea in 50 mM tris pH 8 overnight at 4 °C.

Urea solubilised 0844p and 2575p inclusion bodies from expression of 0844p/pET21a+ and 2575p/pET21a+, were refolded using two different methods - by dilution and by surface bound refolding. Those processes were optimised to determine the best conditions for refolding. After refolding was completed, residual Urea was removed by dialysis against 50 mM Tris HCl pH 8 and 150 mM NaCl. The refolded protein was concentrated using sartorious 50 kDa cut off concentrators and “polished” on a Hi load Superdex 200 column (GE Healthcare, Germany)

Surface bound refolding was used for Histagged 2575p/p15ara. This process involves refolding where unfolded protein is bound to nickel resin. The protein was denatured overnight using 8 M urea in 20 mM phosphate, 150 mM NaCl and bound to nickel resin before being refolded against 20 mM phosphate in 150 mM using a linear gradient for one hour at 1 ml/min. The refolded protein was then eluted with 500 mM imidazole.

3.2.3.5 Preparation of periplasmic proteins

Periplasmic expression utilises a signal sequence which directs the newly expressed protein from the cytoplasm across the inner membrane and into the periplasmic space. After overnight expression of the passenger domain at 25 °C, the cell culture was centrifuged at 3300 x g for 20 min. The cells were then washed using PBS buffer 1x and centrifuged again.
at 3300 x g for 20 min. The pellet was then weighed and resuspended with 10 times of its volume with 0.75 M sucrose in 100 mM Tris pH 7.0 and left for 5 min with stirring at room temperature. Two times the volume of EDTA 1 mM was then added and left again for five min at RT, before addition of 1 mg/ml lysozyme in 0.75 M sucrose 100 mM Tris pH 7 buffer. This was then incubated at RT for 1 h with stirring. 25 mM of MgCl₂ was then added and left stirring at RT for a further 20 min. The resulting spherulites were then removed by centrifugation at 3300 x g for 10 min. The supernatant from this reaction contained the protein. Excess of sucrose was removed by dialysis.

3.2.3.6 Purification of membrane proteins

Cells were thawed from stock after a previous expression. The cell pellet was then weighed and resuspended using ten times the volume of phosphate buffer pH 7.4. EDTA was then added to final concentration of 2 mM and left shaking for ten min before the addition of lysozyme at 1 mg/ml. After the addition of lysozyme, the solubilisation mixture was incubated under stirring for 10 min before being transferred to a 100 ml beaker. The suspension was then sonicated using 30 % power for 3 x 10 min intervals. After sonication, LDAO was added to a final concentration of 0.01%. 1 μg/ml DNase and 3 mM MgCl₂ were then added to remove DNA from the lysed cell culture. The solubilisation was left at 4 °C with stirring for 1 h, before spinning down at 11,000 x g for 10 min, after which the supernatant was then removed and filtered through a 0.22 μ filter. The supernatant was purified by immobilised metal affinity chromatography.

3.2.4 Protein Characterisation

3.2.4.1 Acid Phosphatase assay.

This assay was used to determine the activity of refolded acid phosphatase 0844p. The activity assay is based on the cleavage of P-Nitrophenylphosphate (PNPP) to give P-Nitrophenol and a free phosphate. The cleavage of PNPP results in a colour change from colourless to yellow which is then detected at 410 nm. 0.5 ml of 90 mM Citrate buffer pH 4.8 was combined with 0.5 ml of 15.2 mM p-Nitrophenyl phosphate to which 0.1 ml of refolded protein solution (0.15 - 0.25 U) was added. The mixture was inverted before incubation at 37 °C. A negative control was also prepared using distilled water instead of protein solution. The
colour change was then measured at 410 nm every 10 min; the reaction was stopped by adding 4 ml of NaOH 100 mM solution.

3.2.4.2 Affinity Chromatography

Metal Affinity chromatography was performed as described in chapter 2, section 3.2.3. Eluted fractions were then pooled and concentrated. The buffer was then exchanged by use of a p10 column or gel filtration for a low salt phosphate buffer (phosphate pH 7.4 150 mM NaCl).

3.2.4.3 Size exclusion chromatography

Size exclusion chromatography was performed on a Hiload Superdex 200 column (GE Healthcare, Germany). The column was equilibrated using the desired buffer, normally 20 mM phosphate buffer with 150 mM NaCl pH 7.4 or Tris 20 mM with 150 mM NaCl pH 8.0. After equilibration, 2-2.5 ml of concentrated protein was injected onto the column and run at 1 ml/min. The eluate of the column was collected in 1.5 ml fractions, which were analysed using a 10 % SDS gel. Fractions corresponding to the protein of interest were pooled, concentrated using a Vivaspin 200 concentrator (100 kDa cut-off) and re-injected to analyse the homogeneity of the isolated population.

3.2.4.4 Anion exchange chromatography

A sample of 2575p in 50 mM Tris pH 8.0 was loaded onto a MonoQ 5/50 GL prepacked column (GE Healthcare) previously equilibrated with the same buffer. The column was loaded at 0.5 ml/min onto a pre-equilibrated column (GE healthcare). The column was washed with five column volumes of buffer A (20 mM Tris pH7.4) at a flow rate of 0.5 ml/min. Bound protein was eluted with a flow rate of 0.5ml/min from the column by applying a linear salt gradient from 0 – 100% Buffer B (20mM Tris pH 7.4, 1 M NaCl)
3.2.4.5 Circular Dichroism

The secondary structure of 2575p was evaluated using circular dichroism (CD). The Chirascan CD spectropolarimeter was equipped with a 150 W xenon arc lamp and Julabo AWC 100 water-circulation system for temperature control. The protein buffer was exchanged to 10 mM phosphate buffer pH 7.6 from Tris pH 8.0, as Tris interferes with UV and induces ‘noise’ at higher UV values. The protein sample concentration was 1 mg/ml. Spectra were recorded with 1 nm bandwidth and 2 s per time point at 20 °C, over the range from 260 nm to 180 nm. Four scans were collected for each sample as well as for the buffer in a 0.01 cm path-length Suprasil cuvette. The value of the baseline was subtracted from the average values from the protein spectra using the pro-data chiroscan software, and the curves were smoothed with a Savitsky-Golay factor of 4.

3.2.4.6 Trp Fluorescence measurements.

Possible changes in the conformation of the trimeric and multimeric 2575p were analysed through analysing florescence of the proteins tryptophan residues. The protein concentration was 0.1 mg/ml using a quartz cuvette with a light-path of 1 cm. The florescence of the protein was analysed over the range 280 to 450 nm at an excitation wavelength of 295 nm. An excitation slit width of 2.5 nm and an emission slit width of 2.5 was changed to 5 nm after four repeats to get a higher emission signal. The emission spectrum was analysed across a range of denaturing urea concentrations from 0 to 8 M urea at 0.5 M intervals for both trimeric and multimeric form, in order to detect subtle changes in the emission spectrum, indicative of a conformation change. The average was taken of the emission spectra of each of the urea concentration and the average of the blank spectra was subtracted. The resulting values were plotted on a graph of wavelength vs. emission intensity value to better visualise any shift in wavelength of the emission spectrum for each of the urea concentrations. The weighted mean was calculated for each of the urea concentrations and the results were plotted on a graph of urea concentration vs. weighted mean. From this graph one could detect changes in the unfolding of the protein as the unfolding occurred. A graph of the maxima of each peak from fluorescence was plotted against the increasing concentration of urea to determine if the protein was unfolding correctly.
3.3 Results and discussion

3.3.1 Introduction

Passenger domains have previously successfully been cloned independently of the full length autotransporter, such an approach was used for the passenger domain of 2341 (Alamuri and Mobley, 2008). Hence the focus of this project turned to the passenger domains of 2575, 2174 and 0844. The passenger domains were expressed in the form of inclusion bodies as was done for Hap protease (Liu et al., 1999). Inclusion bodies were formed by removal of the signal sequence - this ensured that the expressed protein remained unfolded in the cytoplasm. The site and size of the signal sequence was determined using bioinformatic signal peptide prediction software ‘SignalP’ from the Centre for Biological Sequence analysis at the Technical University of Denmark (www.cbs.dtu.dk).

The size of the passenger domain was determined using the secondary structure prediction server ‘SABLE’ from the University of Cincinnati Medical College (www.sable.cchmc.org). The secondary structure prediction software was used to predict secondary structure within the protein sequence and differentiate the β-sheet rich translocator domain from the passenger domain. The gene sequence was removed before the α-helix on the passenger side connecting the passenger domain and the translocator domain; primers were designed from the sequence on the passenger side of the cut as had been done for the passenger domain of 2341.

3.3.2 Cloning and Expression

The passenger domains of 2575, 0844 and 2174 were amplified using the same conditions as the full length autotransporters (see chapter 2, section 3.1). Amplification by PCR was successful with bands detected at the required size of 1851 bp for 2575p, 1794 bp for 0844p and 1284 bp for 2174. These amplicons were purified using gel extraction which permitted desalting of the PCR products before digesting with the appropriate restriction enzyme. The resulting digestions were run on an agarose gel; and the digested amplicons were extracted before purification from the gel. 2174p needed to be ligated into a blunted vector for successful ligation and expression to occur. The other genes were ligated using the classical “sticky end” ligation using the enzymes NdeI and XhoI for 2575p/pET21a+ and Nde I and EcoR I for 0844p. All three targets were designed with a stop codon in their reverse primer. After digestion, the PCR products were ligated into the pET21a+ vector, previously digested
with the same enzymes. The ligation was precipitated, transformed into *E. coli* TOP10 cells and screened for clone uptake using ampicillin. Isolated colonies were picked, grown overnight and used for plasmid extraction before screening for plasmid uptake using HindIII for 0844p and Xbal and XhoI for 2575p. Of the three targets, cloning was found to be successful with bands at 1541 and 1891 bp for 0844 and 2575 respectively (figure 3.2).

The positive clones were transformed into BL21 and expressed in the manner described in section 2.3.2.1, samples were taken each hour for the length of the expression trial. The samples were run on a 10 % SDS-PAGE gel. The expressed proteins 0844p and 2575p were found to be present after expression between 58 and 80 kDa, close to the actual molecular weight of both proteins at 66 kDa (figure 3.3 and 3.4). It was evident from expression trials that the protein was being produced.
3.3.3 Analysis of expressed protein

The expressed proteins were analysed using a soluble/insoluble assay in which the cell samples were lysed and the presence of inclusion bodies was detected. Both proteins were expressed using autoinducing broth over the course of 48 h and samples were taken at 18, 24, 48 h. The expressed protein for 0844p/pET21a+ and 2575p/pET21a+ was present at ~66 kDa in the normalised expressed cell samples in figure 3.8 and 3.9 (lane 1, 4 and 7). One can see clearly the same band at ~66 kDa in figure 3.8 and 3.9 (lane 3, 6 and 9) for both 0844p/pET21a+ and 2575p/pET21a+ which corresponds to the insoluble fraction. Those bands do not appear in figure 3.8 and 3.9 (lane 2, 5 or 8) indicating that none of the proteins were soluble. Therefore, refolding experiments were required.

Figure 3.3 Expression of the pET21a+-2575p after transformation into BL21(DE3 cells). Lanes 1-3 contain expression after 1, 2 and 3 h of induction using 1mM IPTG, Lane 4 contains a negative control

Figure 3.4: Expression of the pET21a+-0844p after transformation into BL21(DE3 cells). Lanes 1-3 contain expression after 1, 2 and 3 h after induction using 1 mM IPTG
3.3.4 Refolding of 2575p and 0844p

The inclusion bodies were resuspended in 8 M urea overnight. The protein was then refolded in dilution buffer. Refolding was optimised for 2575p: The protein was refolded using a dilution of 10, 50 and 100 times from an initial protein concentration of 1 mg/ml. Each of the dilutions was run on a HiPrep 26/60 Superdex 200 10/300 GL (figure 3.7). It was decided to use protein with a final dilution of 10x as this delivered the best yield. This was not performed for the 0844p protein as the protein concentration of this protein was not of sufficient concentration for optimisation.
3.3.5 Purification of 2575p

3.3.5.1 Anion exchange chromatography

Refolded 2575p was bound to a GE Hitrap capto Q anion exchange column, to determine the different populations of proteins in the sample. The chromatogram showed four different populations for the same sample of refolded protein, each of which eluted at different concentrations of NaCl on a linear gradient (Fig 3.11).
Each of the peaks was loaded on a HiPrep 26/60 Superdex 200 10/300 GL s200 HR gel filtration column to determine the size of each of the populations. When samples of each of these peak populations were run on a 10% SDS-PAGE (figure 3.9), it was found that each of these populations contained the 2575 passenger protein which indicated that the populations were most likely different conformations of the same protein. However this may also have been due to overlapping of the different peaks. The first peak at 10 ml of the gel filtration was shared between all four populations, with the last peak of the anion exchange exhibiting smaller proteins possibly as a result of degradation. Each of the populations exhibited unspecific degradation at ~20 ml of elution. The size of these proteins was too small to be an active protein, while none of the peaks from the elution profile from gel filtration corresponded to the sizes of the trimeric and hexameric proteins. Only one of the peaks in figure 3.8 appeared to be close to this range and that is the second peak for the first population.

Figure 3.9: SDS gel of peaks from anion exchange (Figure 3.8). Lane 1: before loading. Lane 2: flow through. Lanes 2-12; fractions from peaks 1-5 of anion exchange.
3.3.5.2 Up-scaling of the production of 2575 passenger domain.

The amount of protein to be refolded was increased ten times from 5 ml to 50 ml. A sample of the refolded protein was loaded on a HiPrep 26/60 Superdex 200 10/300 GL gel filtration column. The first peak of the gel filtration at 8 ml corresponded to the dead volume of the column, in which proteins too large to be resolved by the column are eluted (figure 3.10). The next peak after this at 13 ml is most likely protein of the native folding conformation as it laid in a range of molecular weight which corresponded to the predicted molecular weight of the native 2575p, while the peak at 20 ml corresponded to degraded protein. The amount of refolded protein was rather low; from this graph it appeared to be 10-15% of the total refolded protein with aggregates and degraded protein forming the majority of the refolded protein.
3.3.5.3 Analysis of refolding

In figure 3.12 the peak contained degraded protein which eluted at 20 ml. This was removed and the peaks which contained refolded 2575p were analysed in greater detail. The above elution profile (figure 3.14) was run on a HiLoad Superdex 200 column (GE Healthcare, Germany), the first peak, which eluted at 40 ml, corresponds to the dead volume of the column. This peak contains aggregated material, however the other peaks in this elution profile are most likely to be of a natively folded conformation. When samples from each of these peaks was run on an SDS-PAGE 10% gel (figure 3.13) it was found that peak 1 which was eluted at ~ 48 ml was the most stable of the conformations while peaks 2 and 3 were found to have higher levels of degradation.

![Gel filtration profile](image1.png)

Figure 3.12: Gel filtration on a HiLoad Superdex 200 column of the large scale refolding trial of 2575 using a dilution factor of 1:10. Peak 1 represents the dead volume which contained aggregated forms of the protein while peak two was found to be within the range for viable protein folding conformations.

![SDS-PAGE](image2.png)

Figure 3.13: SDS-PAGE of the gel filtration shown in figure 3.12. Each well contained a fraction over the molecular weight range of 2575p.
Figure 3.14: Gel filtration on a Hiload Superdex 200 column of the large scale refolding trial of 2575p with a dilution factor of 1:10.

3.2.5.4 Circular dichroism

Circular dichroism spectra of the proteins contained in peaks 1-3 showed that the protein was folded although it also showed that there was still excessive degradation in peaks 2 and 3 (figure 3.10). The CD spectra of peak 1 which corresponded to the native form of the protein showed that the protein was folded, although it was not known whether the protein was folded in the native form or in a misfolded form.

Figure 3.15: Circular dichroism of the 2575 peaks from Hiload Superdex 200 column. Each of the peaks exhibited a profile characteristic of folded protein.
Figure 3.16: secondary structure full length 2575 autotransporter

Figure 3.17: Gel filtration profile of the purified form of 2575p which was pooled after concentrating the protein from several runs, loaded on a Hiload Superdex 200 column (GE Healthcare, Germany).
Refolded 2575p from different runs was pooled and analysed by gel-filtration and demonstrated high stability and robustness. The process of refolding was found to give the same result each time when run on a Hiload Superdex 200 column (GE Healthcare, Germany). Purified 2575p protein was found to elute at ~ 48 ml each time, while the CD spectra was indicative of a folded protein which contained both α-helical and β-sheet moieties.

3.3.5.5 Molecular weight determination of 2575p

The molecular weight of purified 2575p was determined by comparative analysis with molecular weight standards apoferritin (443 kDa), alcohol dehydrogenase (150 kDa) and BSA (66 kDa). The gel filtration peaks (figure 3.18) of each of the above standards were plotted against the molecular weight of each of the individual proteins. Purified 2575p (figure 3.16) was also run on a HiPrep 26/60 Superdex 200 10/300 GL column to determine its molecular weight by regression analysis. Using this method, the molecular weight of this protein was found to be ~600 kDa, which would imply that the purified protein was a monomer as the monomeric value of the 2575p protein was predicted to be 66 kDa. The predicted conformation for this protein was a hexamer (Alamuri and Mobley 2008) as the native trimeric conformation should form a dimer with other 2575 passenger proteins. The purified form was thought to form a trimer of trimers. The purified protein did not cause inhibition of autoaggregation when added to the growth media of a full-length protein expression indicating that the passenger domain was misfolded and therefore was not active. Under this new model, the protein was expressed under the control of an arabinose promoter and was cloned with the signal sequence allowing the protein to be folded in the periplasmic space.
3.3.6 Refolding of 0844p

0844p was expressed in the same manner as 2575p under the control of a T7 promoter using autoinducing media. The cell pellet was solubilised in the normal manner using LDAO and lysozyme. The inclusion bodies were isolated and resuspended using 8 M urea overnight. The protein was refolded in the same manner as 2575p, however, it was found to precipitate heavily. The protein activity was determined by acid phosphatase assay and was found to be extremely low. The protein was then refolded in the presence of arginine (De Bernardez et al., 1999) (Figure 3.19). The molar concentration of arginine used was plotted against the activity of the refolded protein measured via a chromogenic assay. The amount of arginine used was found to be directly proportional to refolded 0844p, this relationship is limited however by the solubility of arginine in water which is limited to ~0.6 M (figure 3.19)

![Molecular weight determination](image)

**Figure 3.19:** Gel filtration elution profile on a HiPrep 26/60 Superdex 200 10/300 GL of molecular weight standards apoferritin 443 kDa, alcohol dehydrogenase 150 kDa, and BSA 66 kDa.

**Figure 3.20:** 0844p activity after refolding with increasing concentrations of arginine
The refolding process was further optimised by varying the concentration of salt and arginine at pH 7 and 8 (figure 3.20 and 3.21). The protein was refolded overnight on a small scale and assayed the following morning to see which condition was the best combination for refolding. After this process, the amount of refolded protein was found to increase with the concentration of arginine but was also affected by NaCl concentration. The amount of protein which was refolded at 150 mM NaCl was greater than that which was refolded without salt. Refolding at high salt concentration (0.25 M) was found to be detrimental to refolding in most cases. It was also found that there was no significant difference between folding at pH 7 and pH 8.

Figure 3.21: Activity assay of 0844p Refolding at pH 8. Activity of refolded protein at varying salt and arginine concentrations.

Figure 3.22: Activity assay of 0844p Refolding at pH 7. Activity of refolded protein at varying salt and arginine concentrations.
3.3.6.1 Purification of 0844p

Purification with 0844p was performed in the same manner as previously accomplished with 2575p, however it required the removal of arginine and urea. As the protein precipitated when dialysed, it was decided to remove the residual arginine by buffer exchange using a p10 column. The eluate was then loaded on Hiload Superdex 200 column (GE Healthcare, Germany), see figure 3.22. Samples were taken from the eluate fractions of the gel filtration chromatogram corresponding to the molecular weight of 66 kDa. The fractions were then assayed for phosphatase activity (figure 3.23). The peak which contained the purified 0844p protein at 90 ml also contained another form of the protein which was unlikely to be folded properly as no activity was observed. The protein was possibly affected by the refolding process. There was also a peak at 120 ml which was too small to be of significance and like figure 3.12 it most likely comes from protein which has been degraded by the unfolding process.

![Purification 0844p](image)

Figure 3.23: Gel filtration of refolded 0844, the enclosed blue area indicates fractions used for the activity assay in figure 3.23. Hiload Superdex 200 column.

![Graph of activity assay of gel filtration of 0844p fractions C10 to E3 from Figure 3.20](image)

Figure 3.24: Graph of activity assay of gel filtration of 0844p fractions C10 to E3 from Figure 3.20. This band represents the theoretical elution of monomeric 0844p.
3.4 Cloning and expression using p15ara

3.4.1 Cloning and expression of full-length 2575 into p15 vector

Expression of the full-length 2575 protein under the T7 promoter in pET21a+ was found to generate insufficient protein for characterisation studies. The full length autotransporter was therefore subcloned into an arabinose inducible p15 vector. The purpose of this was to slow expression of the protein allowing for greater folding time and integration into the outer membrane. The protein was cloned with a His-tag as the p15 vector did not contain an inherent C-terminal histag. The protein was then expressed in *E. coli* TOP10 cells.

Purification of the full-length 2575 autotransporter can be seen below in figure 3.24. In lanes 6 and 7 there was an 80 kDa band which corresponds to the predicted molecular weight of the full length protein. The size of the band indicated there was too low a yield of the protein to warrant further analysis. An outer membrane purification of the full length 2575 protein was also attempted. The 2575 protein was found to be present in the sonicated cell suspension (lane 8), the supernatant after centrifugation (lane 9) and the sarkosyl solubilisation (lane 10) which removed the inner membrane, however the protein was not found in the triton solubilisation or in the final pellet indicating the protein either had been lost in the purification process or the protein was not contained in the outer membrane.

![Figure 3.25: Lanes 1-7; Ni purification of full length protein solubilised using LDAO. Lane 1; loaded solubilised cell fraction, lane 2; Flow through/unbound proteins, lane 3; 20 mM imidazole, lane 4; 100 mM imidazole, lane 5-7; 500 mM imidazole. Lane 8 – 12; Outer membrane prep. Lane 8; after sonication, Lane 9; supernatant after ultra centrifugation, lane 10; after sarkosyl solubilisation, lane 11; after triton 2% solubilisation, lane 12; final pellet.](image)
3.4.2. Expression and cloning of 2575p into p15 vector

It was thought that the refolded 2575p from inclusion bodies was misfolded. Following this, the 2575p gene was cloned into p15 vector with some differences. The protein was cloned containing the signal sequence, and a His-tag. Positive clones were expressed as in section 3.2.3.1 and analysed on a 10 % SDS-PAGE gel and western blot (figures 3.25 and 3.26). Lanes 1-6 show a band at 66 kDa corresponding to the passenger domain of 2575.

![SDS-PAGE](image1)

**Figure 3.26:** SDS-PAGE of 2575p colony screening (lanes 1-6) and Ni purification of full length protein (lanes 7-11). Lane 7: loading, Lane 8: Flow through, lane 10: 100 mM imidazole, 500 mM imidazole.

![Western blot](image2)

**Figure 3.27:** Western blot of 2575p screening (lanes 1-6) and Ni purification of full length 2575 (lanes 7-11). Lane 7: loading, Lane 8: Flow through, lane 10: 100 mM imidazole, 500 mM imidazole.
3.4.2.1. Optimisation of 2575p expression

Expression of 2575p was optimised for temperature and concentration of arabinose. It was found that there was no difference in the amount of protein being expressed at concentrations of 0.025% and 0.0025% (figure 3.27). While increasing the temperature from 25 °C to 37 °C led to the formation of a higher amount of insoluble protein (figure 3.28). The 25 °C expression was allowed to continue overnight and was subjected to a periplasmic prep the following day. This process removed the outer membrane releasing the proteins of the periplasmic space while simultaneously forming spheroplasts which prevent the release of periplasmic proteins. A sample of the periplasmic extract was run on a 10 % SDS-PAGE gel (figure 3.27 lane 13). Unlike the protein expressed in the samples from the optimisation (figure 3.27 lanes 1-12), the periplasmic prep did not contain any degraded forms of the protein and exhibited a single clear band at 66 kDa. From this optimisation it was found that the best conditions for expression were at 25 °C with 0.025% arabinose. Cells from an overnight expression were harvested, washed with PBS and subjected to a periplasmic prep. The resulting extract was dialysed against phosphate buffer pH 7.4 containing 150 mM NaCl. The extract was then loaded on a nickel column and washed with 10 mM and 60 mM and finally eluted with 500 mM imidazole. It was found that a large amount of the protein remained in the flow through during nickel column loading (lane 2 figure 3.27). The flow through had to be loaded multiple times to extract the protein.

Figure 3.28: Western blot of the SDS-PAGE of soluble fractions from 2575p expression using p15ara vector after releasing the cell pellets using Triton X-100.
Figure 3.29: SDS-PAGE of soluble fractions from optimisation of expression using p15ara vector after lysing the cell pellets with Triton X-100.

Figure 3.30: SDS-PAGE of insoluble fraction from optimisation of expression using p15ara vector after lysing the cells using Triton X-100.

Figure 3.31: IMAC purification of 2575, lane 1; periplasmic extract, lane 2; flow through, lane 3; 10 mM imidazole, lane 4, 5; 60 mM imidazole, lane 6-11; 500 mM imidazole.
3.4.2.2 Molecular weight determination

A sample of the eluted protein was loaded on a Superdex 200 10/300 GL gel filtration. It was found that the size of the protein was the same as that which was purified for the refolded 2575p protein.

3.3.2.3 Refolding of 2575p on solid support

Both strategies used to purify 2575p were found to produce a protein of the same molecular weight. This protein produced by these methods was found to be 600 kDa which was consistent with a nonamer. The size of this protein was far too large to be native folded 2575p and was thought to be misfolded. To verify that the protein was misfolded, the purified protein was refolded on a solid support, in such a fashion as to mimic the folding of the
protein on the external bacterial surface. Under these conditions, the protein would stand a better chance of refolding. In this scenario, the protein buffer was firstly exchanged to remove imidazole after IMAC purification and was then denatured using 8 M urea in Tris buffer pH 8.0 at 4°C overnight. The protein was loaded on the IMAC column and processed as described in section 3.2.3.4 The eluted protein was then loaded on a HiPrep 26/60 Superdex 200 10/300 GL HR gel filtration column to determine the nature of the protein after this process. The elution profile shows a significant shift in the size of the protein from an elution to 10 to ~12 ml, corresponding to the MW shift from of 900 kDa to 300 kDa which correlates to the shift from the protein purified by the previous methods, ie. nonamer to a trimer. (Figure 3.33). It was found that the protein reverted back from the refolded trimeric state to the original nonamer, however high concentrations of NaCl, i.e 500 mM were found to inhibit this process. This can be seen with the yellow and orange chromatogram which were ran one and three hours respectively after the reduction of the salt concentration to 50 mM (figure 3.33). One can see that over time and decreased salt concentration, there is a definite shift in the conformation of the protein, as high salt concentration inhibits the formation of the nonameric form of the protein from the trimeric state However, in low salt conditions, the formation of the nonameric state is not inhibited and trimeric 2575p freely reverts to the nonameric state.

![Transition of nonamer to trimer](image)

**Figure 3.34:** Gel filtration profile using a HiPrep 26/60 Superdex 200 10/300 GL HR column. 2575p was refolded on a solid support. The shift in molecular weight compared to normal purification indicating a possible Trimer-Nonamer shift consistent with what was predicted.
3.4.2.4 Trypsin Digest

The purified 2575p protein was digested using trypsin to determine the protein’s stability. The protein contained numerous trypsin cleavage sites, however the conformation of the protein did not permit cleavage, and this can be seen in figure 3.34 depicting cleavage of the protein over the course of 30 minutes. This indicated that the protein had an intrinsic stability as has been demonstrated for other trimeric autotransporter proteins such as AIDA (Muller et al., 2005).

![Figure 3.35: Trypsin cleavage of purified 2575p over time for samples with and without trypsin. M; marker, lane 1,2; 0 min - , +, lane 3,4; 5; min - , +, lane 5,6; 10 min - , +, lane 7,8; 15 min - , +, lane 9,10; 20 min - , +, lane 11, 12 30; min - , +.]

3.4.2.5 Fluorescence

The 2575 passenger domain contains three tryptophan residues. These were used as a reference to determine conformational changes of the protein during unfolding experiment. Using 8 M urea, 0.1 mg/ml of the protein was denatured across 16 different concentrations of urea (0-8 M; 0.5 M increments). This process was performed for both the nonameric and trimeric forms of the protein. Figure 3.35 was calculated using the weighted mean of each of the fluorescence values for each of the urea concentrations. From this graph it can be determined that there is no difference in conformational change between the trimeric form and the nonameric form. However fluorescence is determined from the vantage point of the tryptophan residues and a conformational shift may be determined if a fluorescence tag was
added close to the 2575 binding site. Conformational changes between the nonameric and trimeric forms which may take place cannot be determined using this method. The denaturation process of both proteins is presented in figure 3.36. The broad peak for undenatured protein undergoes a shift to a higher wavelength when subjected to higher urea concentrations.

Figure 3.36: Fluorescence of the tryptophan residues. Unfolding of the trimeric and nonameric forms of the 2575 passenger domain.

Figure 3.37: Plot of fluorescence maxima vs wavelength at different concentrations of urea to determine whether there was any conformational change in the protein during refolding.
Figure 3.38: Plot of fluorescence maxima vs wavelength at different concentrations of urea to determine whether there was any conformational change in the protein during refolding
3.4 Future work

The work performed to date on the passenger domain of 2575 has yielded significant results, however, although the protein was purified it was inactive. Previous studies of the YadA protein from *Yersinis enterocolitica* have shown that a conserved glycine (G389) in the translocation channel is needed for trimerisation of the passenger domain, and the same may be true for 2575 which also contains a conserved glycine at G708 (Grosskinsky *et al*, 2007; Alamuri *et al*, 2010). In the absence of this highly conserved glycine, the protein may misfold and lose its activity. Instead of purifying the passenger domain, the full length protein could instead be purified and be used for inhibition assays. However this process has yielded only a very small amount of protein.

Refolding of 0844p has been shown to produce an active form of the protein. Future large scale production would allow protein characterisation using various biochemical and biophysical methods, such as circular dichroism, activity assays and trypsin cleavage. However the protein is prone to precipitation and the process would need to be optimised before undertaking large scale refolding experiments. The same approach could be applied to other targets in this study such as the 2126 protease or refolding on a solid support as performed for 2575p.
3.5 Summary

Expression of the full-length autotransporters (0844, 2575 and 2174) failed to yield viable amounts of protein to initiate characterisation studies and therefore the focus of this study shifted to production of the passenger domains. The corresponding genes for 0844 and 2575 were successfully cloned into pET21a+. 0844p and 2575p were cloned without a signal sequence and therefore formed inclusion bodies on expression. The expressed proteins were purified in the form of inclusion bodies and refolded. Both proteins were refolded successfully, but at a low yield. From analysis of the gel filtration chromatogram of 2575p, the majority of the protein formed either aggregates or degraded, with only a small amount of protein refolding within a feasible range to be native 2575p. 2575p was also characterised using circular dichroism to determine its secondary structure. The refolded protein was found to be of a size that was indicative of a nonamer and not a trimer or a hexamer which was the native state. 0844p was refolded in its active form. Its molecular weight determine by gel filtration was found to be within the required range.

Due to the low yield of protein purified from inclusion bodies, 2575p was cloned into the p15ara vector to change the expression level due to a better tuneable vector. The protein was also designed to have a histag and a signal sequence to permit folding in the periplasm. This approach proved successful with enough protein being generated for characterisation and structural studies. The molecular weight of this refolded protein proved to be the same as that which was purified for the refolded 2575p protein. In order to analyse the mechanism of folding, the protein was denatured using 8 M urea and was bound to a Ni resin by means of its histag. The protein was refolded on a linear gradient against 20 mM phosphate buffer. When the resulting eluted protein was run on a gel filtration column it was found that there was a shift in the molecular weight of the protein from 600 kDa to 200 kDa, however when the protein was left at 4°C for 24 h it was found that the trimeric 200 kDa form reverted back to the nonameric form from which it had been refolded, although a high salt concentration was found to inhibit this process. Both the nonameric and trimeric forms of the protein were denatured with different concentrations of urea and subjected to florescence analysis, to determine if there was any shift in florescence upon formation of the multimers contained in the protein sample. It was found that over the range which was analysed there was no significant difference between the samples. The nonameric from of the protein was also digested with trypsin to determine if the protein contained accessible trypsin sites or was
inaccessible to trypsin as was the case with other autotransporters, the latter was found to be the case.
Bibliography


Filloux A.(2011). *Protein secretion systems in Pseudomonas aeruginosa an essay on diversity evolution and function.* 2:155


Kilian, M., Mestecky J., Kulhavy R, Tomana M., and Butler W. T. 1980. *IgAl proteases from Haemophilus influenzae, Streptococcus pneumoniae, Neisseria meningitidis, and*


Sauri A., Soprova Z., Wickstrom D., de Gier J.W., Van der Schors R. C., Smit A. B., Jong W.S.P., and Luirink J. The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. Microbiology (2009), 155, 3982–3991


Vallejo L. F. and Rinas U. *Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins*. Microbial Cell Factories 2004, 3:11


