Defining the Gate Domain of the Filamentous

Phage Secretin pIV

Julian Spagnuolo

Abstract

Secretins are a family of large outer membrane proteins with large-diameter lumens (5-10 nm). This allows them to transport bulky substrates, including folded proteins, or assembled macromolecular structures – filamentous phages and type IV pili. Many proteins exported by secretins are essential for virulence of Gram-negative pathogens. Such a large channel would ordinarily sensitize the bacterial cell to noxious agents. However, secretins do not - the presence of a mobile septum, or gate, across the lumen of the channel prevents access by noxious agents. Despite the importance of the gate in secretin function, the sequence identity of the gate residues is unknown. In this study, *in vivo* random mutagenesis was used to map amino acid residues involved in gating the filamentous phage secretin - pIV. This approach has identified 34 residues that are involved in the gating mechanism. These residues are predominantly located within the secretin homology domain and organised into two clusters; GATE1 (39 residues) and GATE2 (14 residues). A number of isolated point mutants sensitised *Escherichia coli* to bile salts and antibiotics. These findings allowed the construction of a site-directed deletion mutant of GATE2, confirming the gate function.

This thesis mapped, for the first time, a secretin gate. Given the success of the mutagenesis approach used in this thesis, the method here will be applicable to secretins of pathogenic bacteria and other outer membrane channels whose gate regions have not been determined as yet. Knowing the gate regions of "pathogenic" secretins in turn will help design secretin-targeting antimicrobials.

Foreword & Acknowledgements

"I almost wish I hadn't gone down that rabbit-hole – and yet–and yet – it's rather curious, you know, this sort of life" – Alice in Wonderland.

I, am an addict. I am addicted to many things, two of which are puzzles and learning. Like Alice, I find the mystery on the other side of the looking glass enticing. My 'Rubiks-complex' and curiousity was what drove me to research in the first place. Idealistic to say the least I wanted to 'save the world' and leave a mark on humanity. It is amazing what can happen during as a short time as two and a bit years to change a person. Layer by layer the harsh truths of life, society, and humanity are exposed, inducing reflection and change upon our core principles. Having observed the toxicification of idealism, I now fear for the survival of imagination. As children, we cannot wait to 'grow-up', at the same time, our imagination and idealism are at their peak, unbridled by the unwritten rules and nuance of the unwashed masses. Too much emphasis is placed on being normal; that we have stood silent while the evisceration of imagination takes place in our educational institutions. Instead, we are taught to sycophantically toe the party line because we have forgotten to teach ourselves how to learn; to think independently; to have an opinion and express it. These are seemingly simple things yet, without imagination, they have no hope of taking root and flourishing. What remains in their absence is an organic automaton, for without freedom of thought what are we but inconsequential meat puppets, blissfully ignorant of the ability to reach for greater purpose. To paraphrase Wayne Coyne; the only way forward for humanity is to love and let

love, if only to allow those few fearless freaks to survive and pass on the gift of imagination.

I am endebted to my supervisor, Dr. Jasna Rakonjac, for her openness and never-ending encouragement, even in the face of adversity or 'inside rain'. She gave me a home in her lab, and I can only hope she does not regret taking me on as a Masters student.

To my labmates; David Sheerin, Matt and Nick Bennett, and Dragana Gagic for their useful, humourous, and insightful discussions – thank you, I'd like to think you kept me sane, but somehow I don't think that quite fits. Without you I would have been lost.

Last, but not least, I'd like to thank my family for their support over the past few years.

Abbreviations

E. coli - *Escherichia coli*

OM - outer membrane

SPA - single particle analysis

cryo-EM - cryogenic electron microscopy

ORF - open reading frame

SOC - Super Optimal Catabolite repression media

2xYT - 2 times strength Yeast Extract, Tryptone media

IPTG - Isopropyl β-D-1-thiogalactopyranoside

X-Gal - 5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside

TAE - Tris-acetate-EDTA

EDTA - ethylene diamine tetraacetic acid

DMSO - Dimethyl sulfoxide

MIC - Minimal inhibitory concentration

psp operon - phage shock protein operon

Amp – Ampicillin

Cm – Chloramphenicol

BD – Becton, Dickinson and company

List of Tables

List of Figures

Table of Contents

1.0 Introduction

Secretion of folded proteins is a keystone to many bacterial lifestyles. However, the double-membrane envelope of Gram-negative bacteria poses a significant barrier to the secretion of such substrates. Evolution has provided these bacteria with several related (Figure 1) secretion systems or pathways for the secretion of proteins and macromolecular complexes through both the inner (IM) and outer (OM) membranes without disturbing the integrity or permeability of the cell envelope. There are more than half-a-dozen such secretion systems that differ in complexity, mode of function, and substrates they secrete. Some secretion systems assemble/secrete filamentous phage, exchange genetic material with other organisms or assemble specialised surface structures (Hueck, 1998, Russel, 1998, Gerlach & Hensel, 2007).

Among these secretion systems four have a homologous large OM channel, called secretin: type 2 secretion system (T2SS), type 3 secretion system (T3SS), type IV pilus assembly system (T4PS) and filamentous phage assembly/secretion system (FFSS). T3SS is essential to virulence processes in pathogenic bacteria such as *Salmonella enterica* Typhimurium, where perturbing the secretion of effector proteins abolishes the ability to invade host cells and as a result prevents cytotoxicity and dramatically reduces persistence in cell culture and murine models (Miki *et al.*, 2004). T3SSs are also required to inihibit activation of phagocytes, thereby preventing destruction of Salmonella within phagosomes and lysosomes (Galan & Collmer, 1999). A T2SS secretes pentameric protein cholera toxin, the key virulence factor of *Vibrio cholerae*

(Davis *et al.*, 2000). The same system is also required for the secretion of lipase (Sikora *et al.*, 2007). T2/3SS share similarities with the FFSS. Indeed, the cholera toxin encoding phage CTXφ of *V. cholerae* utilises the secretin of the host cell T2SS for its assembly and release (Davis et al., 2000). There is also some similarity between Type 4 pili assembly (T4PS) and phage assembly, in that phage structural subunits are localised to the periplasmic face of the IM, much the same way as the Type 4 pili subunit, pilin, does (Bayan *et al.*, 2006, Craig & Li, 2008). The main commonality between these secretion systems is in their OM component, the secretin.

The assembly machinery of filamentous phage can be thought of as a simplified version of T2/3 secretion systems, requiring fewer accessory proteins but utilising homologous and structurally similar gated OM-channels (Figure 1, Figure 2, Figure 3).
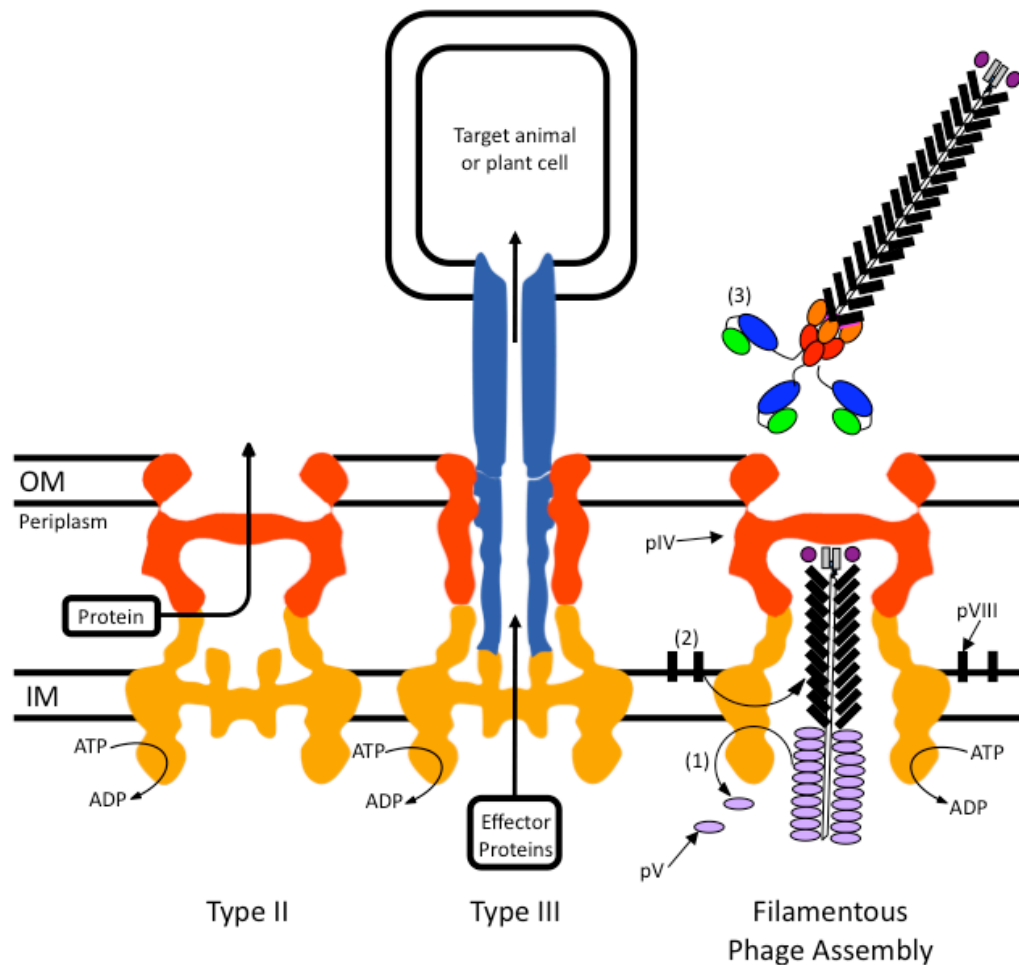
Figure 1: Overview of Gram-negative bacterial secretion systems. IM components vary between systems and species, however some generalities exist, that is export is driven by ATP hydrolysis and each utilise a similar OM component – the secretin (shown in red). Type 2 secretion substrates enter the secretion apparatus in the periplasm and are exported to the extracellular milieu. Type 3 secretion substrates enter a molecular needle (Blue) threaded through the centre of the secretion apparatus and are injected directly into target cells. Filamentous phage assembly is analogous to type 2 and 3 secretion systems. The assembly process is initiated when pV targets the phage genome to the pI/pXI complex which strips pV away (1), major coat protein pVIII is processively assembled onto the growing phage filament (2) as it is exported through an open pIV secretin channel (not shown), when assembly is terminated by the addition of pIII the phage is released into the environment (3). Source of silhouettes: (Opalka *et al.*, 2003, Marlovits *et al.*, 2004).

1.1 Filamentous Bacteriophage Assembly

The FFSS is comprised of an IM complex that incorporates phage structural components onto a growing filament, extruding it through a gated OM channel, the secretin pIV (Figure 1) (Russel *et al.*, 1997). This IM complex is composed of pI and pXI subunits. pI, the larger of the two IM proteins, has two functional domains; a large N-terminal cytoplasmic domain containing a highly conserved nucleotide-binding motif, and a membrane-spanning domain connecting the cytoplasmic portion of pI to a short periplasmic domain which interacts with the N-terminus of pIV.

During replication, the newly synthesised ssDNA phage-genome becomes a collapsed rod and is coated by the phage-encoded ssDNA-binding protein, pV. The only exposed segment of the genome is a hairpin loop, called packaging signal. The cytoplasmic domains of pI and pXI recognise the phage-packaging signal, guiding it into the pI/pXI assembly machine. It has been postulated that pI is an ATP-dependent transporter producing the requisite energy that drives the assembly and export process (Feng *et al.*, 1999). This is analogous to T2/3SS and T4 pilus assembly, all of which require ATP hydrolysis by their IM machinery to drive assembly and/or export and have an IM-associated ATPase (Figure 1).

In the initiation stage of filamentous phage assembly, minor virion proteins pVII and pIX are added first to the packaging signal, followed by the deposition of the major coat protein pVIII that replaces pV along the collapsed ssDNA genome.

4

During this process, Thioredoxin acts as a processivity factor, promoting the assembly of phage filaments (Russel & Model, 1986). Major coat protein pVIII is incorporated onto the growing phage filament via the assembly action of the IM pI/pXI machinery (Figure 1). pVIII is continuously added to the phage filament in an ordered helical fashion (Haigh & Webster, 1998).

The pI/pXI/pIV assembly complex exists as a functional unit. To maintain functionality, pI and pIV are only transferable into other filamentous phage assembly systems as a pair, not singularly (Feng et al., 1999). That is to say, pI and pIV can functionally replace respective components of related phage, but only when transferred to that assembly system together. This indicates pI has a role in the conformational changes required to open the gate-structure of pIV and/or in driving assembly and export through the OM.

A mature filamentous phage particle is 880 nm in length and 6 - 7 nm in diameter. Assembly of the growing filament coincides with its extrusion through the OM secretin pIV. Termination and release from the assembly channel is mediated by the addition of pVI and pIII (Endemann & Model, 1995, Rakonjac *et al.*, 1999).

1.2   pIV – A Model Secretin

The OM component of the filamentous phage assembly machine, pIV, is a member of the secretin family of homologous proteins, which are large multimeric gated channels embedded in the OM of Gram-negative bacteria

(Kazmierczak *et al.,* 1994, Linderoth *et al.*, 1997). Cryo-EM and gold-labelling experiments have revealed that the pIV multimer has 14-fold radial symmetry (Opalka *et al.*, 2003) and is composed of 14 identical subunits encoded by the genome of filamentous bacteriophage f1.

Cryo-electron microscopy (EM) and single particle analysis (SPA) of pIV has revealed the three-tiered ring structure. pIV has a large internal diameter at the channel entrance and exit, ranging from 6.0 nm at the N-ring to 8.8 nm at the C-ring (Opalka et al., 2003) – large enough to allow passage of bulky proteins and antibiotics, which cannot pass through the LPS leaflet of the OM. However, *E. coli* expressing pIV is not sensitive to these compounds. The impermeability of pIV to these compounds is attributable to a septum that blocks the lumen of the channel at the level of the M-ring (Figure 2A). Structures of the T3SS secretins from InvG and YscC of *Salmonella typhimurium* (Marlovits *et al.*, 2004) and *Yersinia enterocolitica* (Burghout *et al.,* 2004), respectively, and the T2SS secretin PulD of *Klebsiella oxytoca* (Nouwen *et al.*, 1999, Chami *et al.*, 2005) have also been obtained by Cryo-EM and SPA revealing very similar structures (Figure 2B & C).
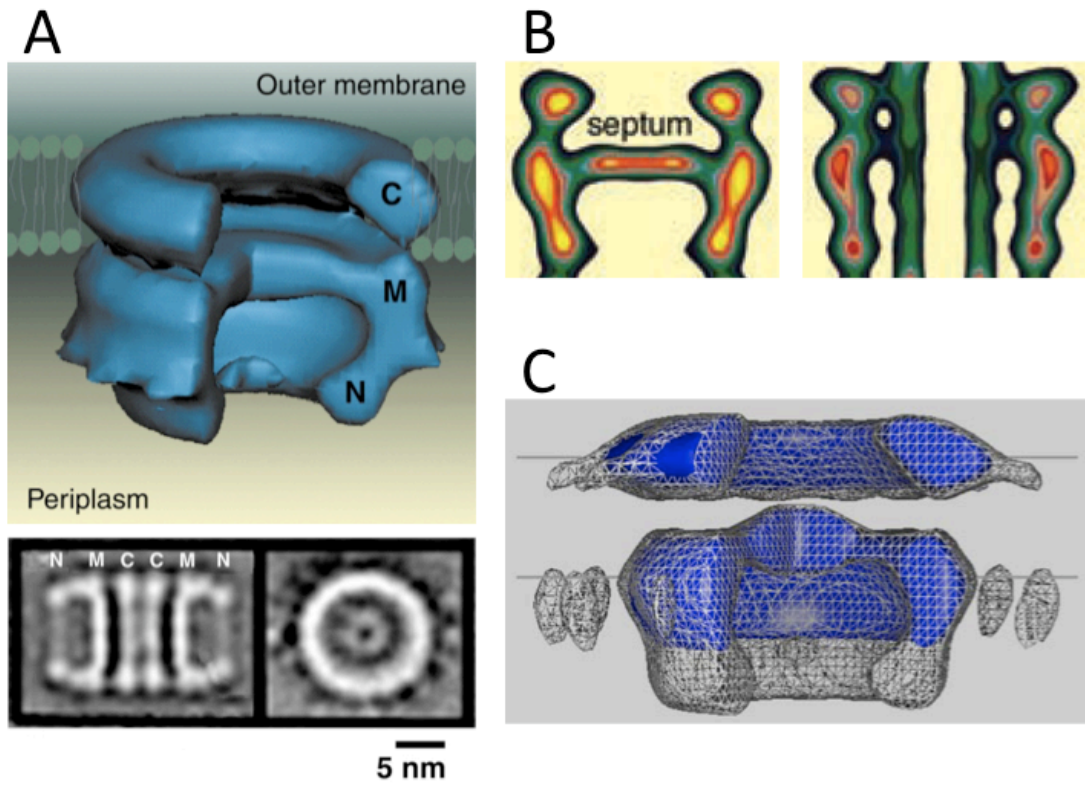
Figure 2: (A) 3D reconstruction of pIV multimer, with single-particle reconstruction beneath (Opalka et al., 2003). (B) Protein density mapping showing conformational change in the InvG secretin needle complex of *S. typhimurium* (Marlovits et al., 2004). (C) 3D reconstruction of the PulD multimer, the white mesh indicates the wildtype multimer. Highlighted in blue is the trypsin-resistant portion of PulD (Chami et al., 2005).

Sequence alignments of pIV homologs have identified a conserved C-terminal domain (Figure 3), which spans the outer membrane and forms the middle ring, including the septum or gate. The N-terminal domain is less well conserved; domain swapping experiments between pIV homologs from two related filamentous phages f1 and Ike, have suggested that this region may contain substrate specificity determinants (Daefler *et al.*, 1997a, Daefler *et al.*, 1997b, Linderoth et al., 1997). 3D SPA reveals an interesting structural feature of pIV common to other secretins; the lack of electron dense regions between the C and M rings, punctuated by connecting regions (Figure 2; (Opalka et al., 2003). Density mapping of contoured longitudinal sections of the InvG secretin reveal low-density regions correlating to the C-M connecting regions seen in pIV 3D reconstructions. Importantly, differences in the shape and density of InvG secretin rings can be seen by comparing Cryo-EM images in the presence and absence of the cognate substrate, the needle complex. When the needle complex is assembled the periplasmic N ring, and OM embedded C ring expands in diameter to accommodate the needle. The septum disappears completely, suggesting a complex gate opening/closing mechanism (Marlovits et al., 2004).

In addition to the septum revealed by SPA (Figure 2), it was also found in electrophysiology experiments that the application of high voltage (minimum of 200 mV) was required to open the pIV channel (Marciano *et al.*, 2001). This was later corroborated by similar electrophysiology experiments with wildtype Yersinia secretin YscC (Burghout et al., 2004). Together, these structural and physiological findings suggest the presence of a gate structure.

Electrophysiology experiments of purified leaky mutant S324G pIV (Marciano *et al.*, 1999) revealed a markedly-decreased voltage threshold required for opening the channel (80 mV), suggesting a lower resistance to electrical force and/or differences in protein conformation.

Earlier attempts to understand the pIV gate have created leaky-gate mutants defined as permissivity to maltopentaose in a Δ*lamB* background, or sensitivity to bile salts (deoxycholate) and/or the antibiotics Vancomycin and Bacitracin. These agents do not normally affect Gram-negative bacteria unless the integrity of the outer-membrane has been perturbed. Antibiotic and detergent sensitivity tests have been historically used to characterise leaky-mutations in OM porins (Wandersman *et al.*, 1979, Benson & Decloux, 1985, Benson *et al.*, 1988, Misra & Benson, 1988a, Misra & Benson, 1988b). Mutation of the highly conserved pIV residue G355 resulted in a leaky gate (Russel, 1994). This residue lies in a highly conserved segment (PFAM family alignment consensus TLVLGGLT). Other point mutations affecting gate integrity identified before the start of this work are N295S, S324G, A329T, N335I/S, and G367S, all within the secretin homology domain. Another group of mutations characterised in pIV affect stability of the multimer and the turnover of pIV in the periplasm, indicating folding and/or multimerisation defects. Several such mutants were found interspersed with the leaky mutants in the primary sequence of pIV (Russel, 1994, Linderoth *et al.*, 1996, Linderoth et al., 1997, Marciano et al., 1999, Marciano et al., 2001). Leaky mutants variably affect the function of pIV in filamentous phage assembly, whereas destabilizing mutants all render pIV inactive because they are unable to form channels in the OM.
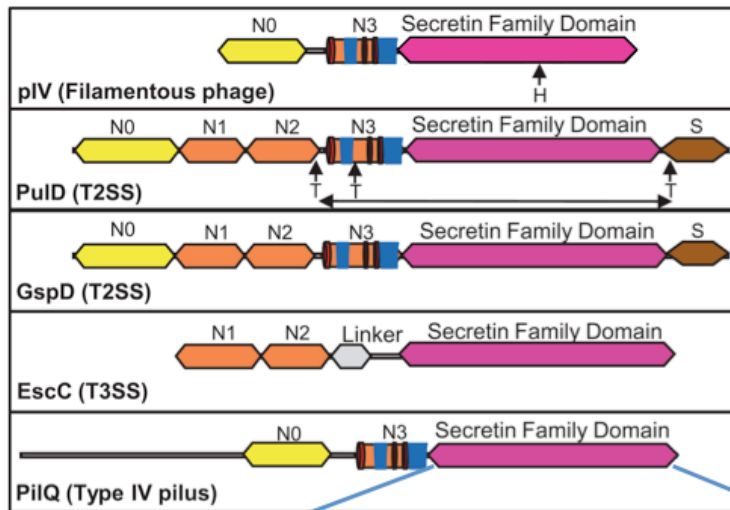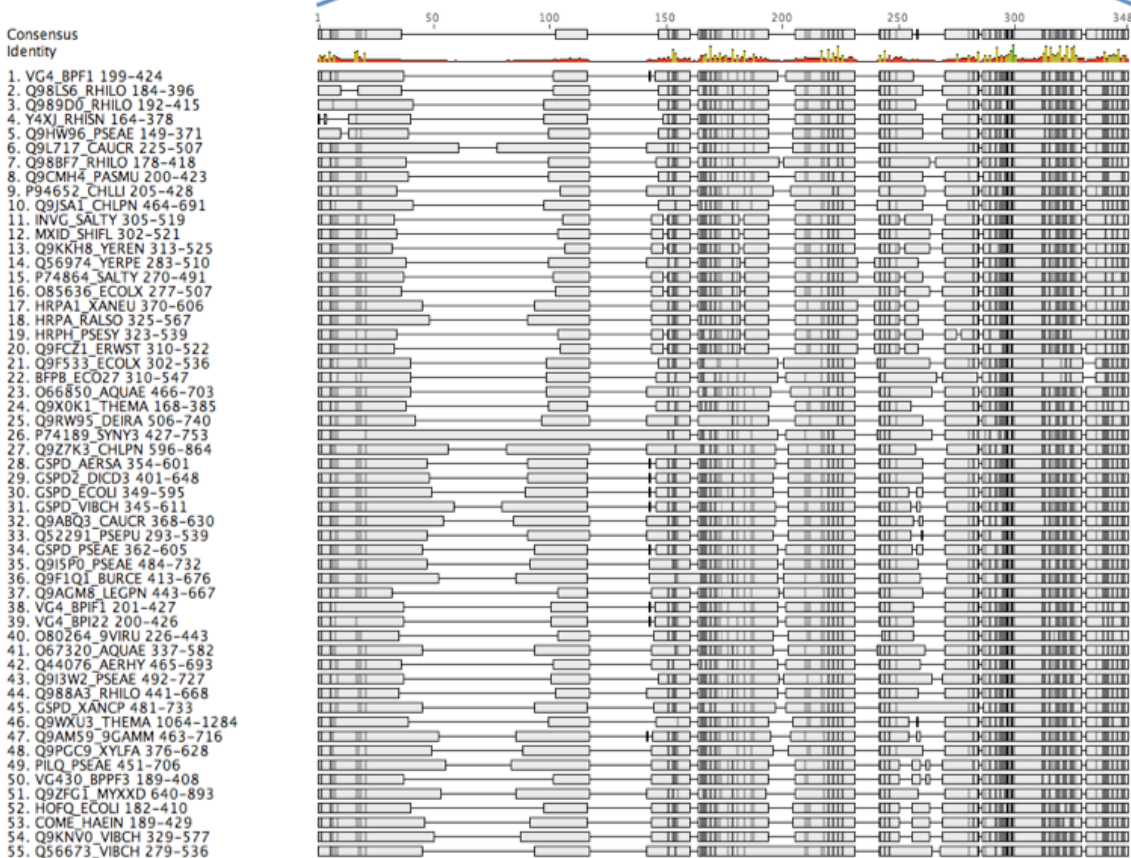
Figure 3: (A) Domain architecture of 4 secretins. pIV domain organization is compared to that of PulD (T2SS of *Klebsiella oxytoca*; (Chami et al., 2005), GspD (T2SS of EPEC; (Korotkov *et al.*, 2009) and EscC (T3SS of EPEC; (Spreter *et al.*, 2009)). Yellow diamond, N0 or FpvA-like domain; orange diamond, N1, N2 and N3, KH-fold domains; gray diamond, T3SS-specific N-terminal subdomain; pink diamond, secretin family domain; brown diamond, pilotin-binding domain; H; histidine tag; T, trypsin cut sites; Linker, variable linker. Double-headed arrow under PulD scheme indicates the portion that is found in the trypsin-resistant core of the channel

multimer. Shapes in the $N3_{pIV}$ and $N3_{PulD/GspD}$: secondary structures modelled based on the high-resolution structures of $N1_{GspD}$ (Korotkov et al., 2009) and $N1/N2_{EscC}$ (Spreter et al., 2009): red ovals, β strands; blue shapes, α helices. (B) PFAM seed alignment of the secretin family domain. pIV from filamentous phage f1 is listed at the top (VG4_BPF1 199-424), showing sequence homology of residues 199-424 of the nascent propeptide. Boxes indicate regions of homology; darker shading in the box indicates a higher level of conservation. Lines indicate breaks in homology.

1.3   Function of the C-terminal Secretin Homology Domain

The most conserved portion of secretins is the C-terminal secretin family domain (Figure 3). Several research groups have conducted truncation and protease digestion experiments to structurally and functionally characterise this region. Circular dichroism spectra of the trypsin-resistant PulD C-terminus (comprising one N sub-domain, and a complete C-terminal domain) indicated only 27% of the polypeptide is in β-strand conformation. This is a low β-strand content compared to most other integral OMPs (Chami et al., 2005). Although most OM proteins contain transmembrane β-strands, it is possible that trans-OM segments are α-helices as in the Wza and VirB10 crystal structures (Beis *et al.*, 2004, Dong *et al.*, 2006, Chandran *et al.*, 2009). Alternatively, the apparent 1:3 ratio of β-strands to α-helices could reflect the ratio of the membrane spanning vs. periplasmic portions in the analysed multimer.

Truncation of PulD to express only the C-terminal secretin homology domain (PulD-CS) revealed that the N-terminal domain in this secretin has no involvement in membrane insertion, multimerisation or localisation. Furthermore, negatively stained EM and single particle reconstruction showed that PulD-CS was still able to form 3-tiered barrels, showing that the C-terminal domain is structurally involved in gating and builds the major part of the N ring. The lack of significant difference in the transmembrane electrochemical potential of strains expressing wild-type PulD or PulD-CS is suggestive that there is no gating defect in PulD-CS, since it would be expected that a defective

gate would cause drastic dissipation of the proton motive force (PMF), if erroneously inserted into the IM (Guilvout *et al.*, 2006).

Attempts to His-tag the pIV secretin at the N- or C-termini prohibited multimerisation and/or destabilised the multimer. pIV His-tagged at positions A308 or D347 formed functional multimers, however stability was compromised in the absence of the stabilising spontaneous mutation S318I, which increased stability of the multimer in comparison to the untagged wild-type pIV (Linderoth et al., 1996). Transmembrane β-strand (TMBETA) prediction (Gromiha *et al.*, 2005) suggested that the least interfering his-tag site, A308, was situated between two predicted transmembrane β-strands (Appendix 4). Secretin family PFAM alignments (Figure 3) further showed that the A308 His-tag was located between two conserved regions, which correspond to the predicted β-strands. From this we can infer that sequence elements within the C-terminal domain of secretins are important for inter-subunit interactions.

It was noted that mutations affecting the gating of large OM channels sensitized *E. coli* to bile salts and normally impermeant antibiotics (Liu *et al.*, 1993, Marciano et al., 1999). This was used to screen an error-prone PCR mutant library of the *gIV* C-domain for sensitivity to deoxycholate (DOC) (Russel et al., 1997). Although this method isolated DOC sensitive mutants and identified 4 residues in pIV, which affected gate-permeability independent of function (N295, S324, N335, and G367), it was hampered by the lack of positive selection for leaky/permeable gate mutants. Work on one randomly generated DOC-permeable mutant (S324G) found that it enabled growth on maltodextrins in the

absence of the maltodextrin-specific import channel LamB (Marciano et al., 1999). LamB is essential for the import of maltodextrins larger than maltotriose (Wandersman et al., 1979), thus LamB-deletions provide a method to positively select for leaky-gate mutants in pIV. Furthermore, the rate of maltohexaose uptake by the pIV S324G mutant was measured (Marciano et al., 2001), showing the rate of maltodextrin uptake through the S324G mutant channel is much higher than in the wild-type pIV channel, which does not permit the uptake of maltohexaose. Additionally, the passage of phage filaments through pIV S324G multimers blocked the channel, preventing entry of maltohexaose. These results indicate that the C-terminal secretin homology domain has a role in both the structure and mechanism of the secretin gate.

1.4 Secretin Targeting to the Outer-Membrane, Folding, and Multimerisation

The processes underlying OM targeting of secretins is unclear. Secretin monomers are translocated to the periplasm via the Sec system; however, the subsequent folding and targeting pathway is not well understood. Some secretins are localised to the OM with the help of an OM lipoprotein, or pilotin, which interacts with a dedicated pilotin-binding domain at the C-terminus of the secretin. Representative secretin-pilotin pairs from T2SS and T3SS are PulD-PulS and InvG-InvH (Crago & Koronakis, 1998, Guilvout et al., 2006). Some secretins, like HxcQ, BfpB and TcpQ, contain a lipoprotein-like domain and they are targeted to the OM by fatty acylation – therefore they contain an "intramolecular" pilotin. However, BfpB and TcpQ require additional partner

proteins for OM localisation (Viarre *et al.,* 2009). Essentiality of these targeting signals was confirmed by the finding that deletion/inactivation of the pilotins in the pilotin-dependent secretins, or the HxcQ-lipid attachment motif, leads to the accumulation of secretin multimers exclusively in the IM.

The pilotin-secretin monomer complex is probably shuttled to the OM via the Lol lipoprotein-sorting pathway (Okon *et al.,* 2008). Given that multimerisation is not affected by the absence of the pilotin or fatty-acylation, thus localisation factors are likely to have no role in the formation of the multimer. However, it has been suggested for *N. meningitidis* PilQ secretin that there are additional cognate lipoproteins (besides pilotins) encoded by the respective secretion system gene clusters, which may have a function in multimerisation or stabilisation of the multimer (Okon et al., 2008). In contrast to the secretins that require accessory factors or domains for the OM targeting and/or multimerisation, pIV secretin of the FFSS has no known localisation factors and membrane fractionation shows the localisation of pIV to both the IM and OM (Russel & Kazmierczak, 1993). Similarly, *V. cholerae* T2SS secretin EpsD does not require additional localisation proteins. However, its IM/OM distribution remains unclear.

It is unknown whether secretins are transported through the murein layer of the periplasm as monomers, or as prefolded multimers ready for OM insertion. Elucidation of this step would require investigation of the folding and oligomerisation intermediates. The sheer size of secretin multimers (up to 1500 kDa) would make it exceedingly difficult to pass through the murein layer, even if a cognate transport protein exists. Even though pilotins efficiently target their

cognate secretins to the OM there is always a small IM-associated pool of secretin multimers (Daefler et al., 1997a, Crago & Koronakis, 1998), suggestive that pilotins sequester secretins, preventing IM association, rather than acting as chaperones.

For secretins that do not contain targeting domains or require accessory targeting proteins, such as pIV, it is unclear how they are targeted to the OM, given that pilotin-dependent secretins cannot reach the OM when expressed on their own (Crago & Koronakis, 1998, Daefler & Russel, 1998, Guilvout et al., 2006). One question of interest towards understanding of OM targeting of self-targeting secretins is the causality between multimerisation and OM insertion. Analyses of pIV mutants have shed some light on this. Substitution of pIV residue P375 (Russel, 1994) prevents efficient multimerisation of pIV. However, these substitutions were still found associated with both the IM and OM fractions, suggesting that association of monomers with membranes could precede multimerisation (Russel, 1994). Interestingly, when the pilotin-binding domain of *K. oxytoca* secretin PulD is fused to (pIV-PulD$_{65}$), subsequent coexpression of the cognate PulD pilotin (PulS) targets them to the OM, just like wild-type PulD. Furthermore, pIV-PulD$_{65}$ monomers are protected from proteolytic degradation by bound PulS (Daefler et al., 1997a). Proteolytic protection of PulD is independent of the lipidation of pilotin PulS, and of OM targeting, as long as the secretin-pilotin interaction is maintained (Daefler et al., 1997a, Carbonnelle *et al.*, 2005, Koo *et al.*, 2008). This suggests a model where secretin monomers are sequestered by their lipopeptide pilotin, or localisation factor, before they insert into the IM. Proteolysis of unsequestered secretins

16

secures the OM targeting pathway and at the same time protects the host from the formation of potentially toxic secretin channels in the IM.

The processes occurring between OM insertion and multimerisation of secretin monomers are less clear; PilQ from *N. meningitidis* T4PS requires the key OM protein assembly translocon Omp85 (YaeT/BamA homolog; (Voulhoux *et al.*, 2003, Ruiz *et al.*, 2006, Volokhina *et al.*, 2009)) for insertion, even in the presence of cognate pilotin PilW. However, PilQ is the only known case of a secretin requiring the Omp85 translocon for its OM insertion and multimerisation (Carbonnelle et al., 2005, Robert *et al.*, 2006, Collin *et al.*, 2007). From these observations a model can be proposed, whereby OM biogenesis of secretins occurs by co-targeting secretin monomers associated with a lipoprotein to the OM, followed by multimerisation and insertion into the membrane. Alternatively, lipoproteins may not act as pilots; rather they only prevent premature multimerisation/insertion by titrating monomers away from the IM. OM-targeting could be a consequence of the OM lipoprotein *itself* being targeted to the OM by the Lol pathway. The involvement of the Lol lipoprotein-sorting pathway in targeting the secretin to the OM has yet to be confirmed. Altogether, published work on several secretins from different secretion systems points to a basic common domain organisation and varied requirements for additional folding, multimerisation, membrane insertion and OM targeting domains and accessory factors.
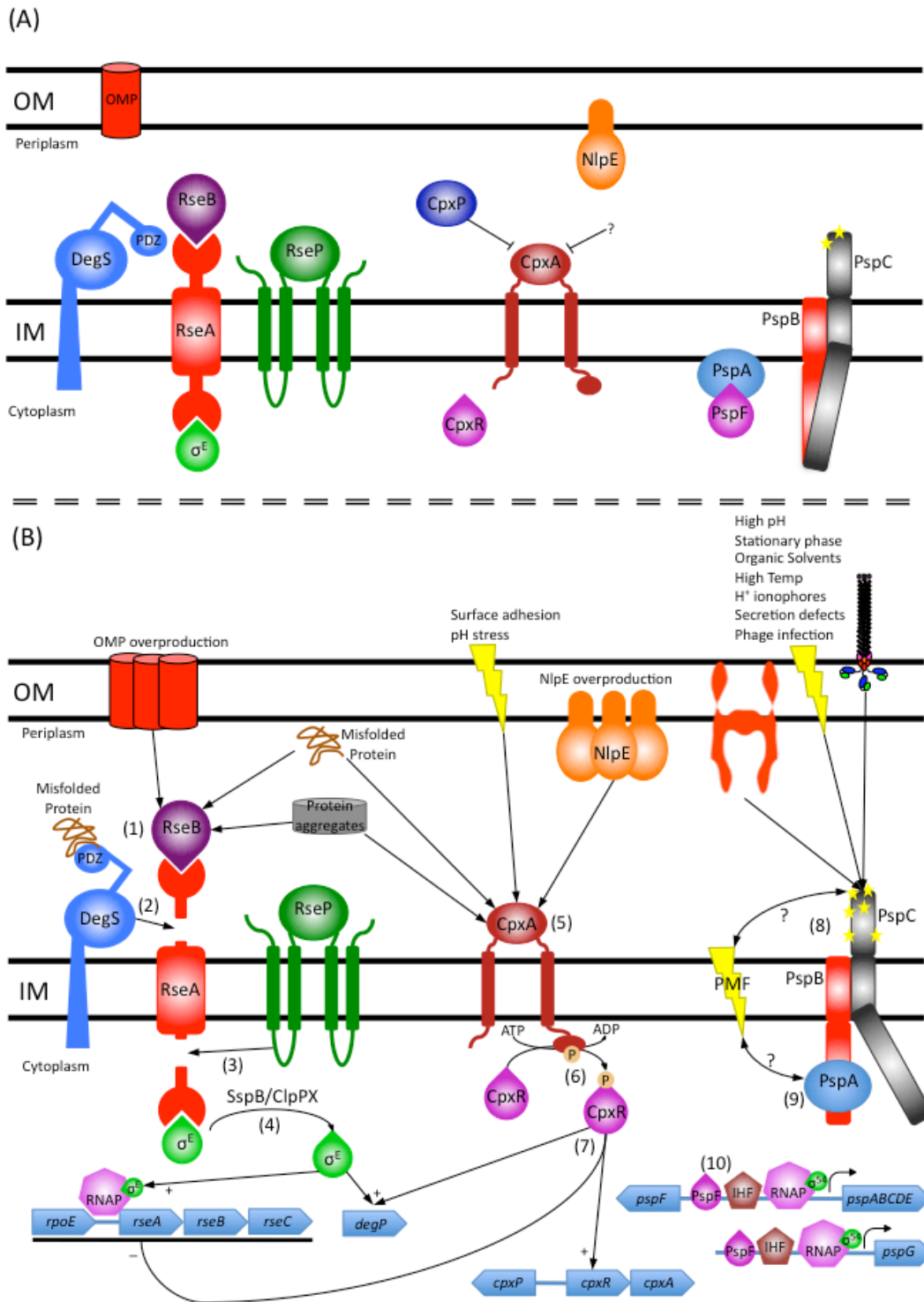
Figure 4: Overview of *E. coli* σ^E, Cpx, and Psp stress response pathways. When the bacterial cell is unstressed (A), stress responses are inhibited. The PDZ domain of DegS blocks the proteolytic domain (of DegS) from interacting with σ^E response regulator RseA. CpxP and other unknown factors inhibit the Cpx regulon. The Psp response activator PspF is inactivated by interaction with PspA. When extracytoplasmic stress occurs (B), specific stress activation pathways are activated, depending on the nature of the stress. The **RpoE/σ^E response** is induced by the accumulation of misfolded OM proteins and their subsequent aggregation

products in the periplasm, which is detected (1) by the PDZ domain of DegS. Detection initiates a proteolytic cascade; first, a conformational shift in DegS occurs allowing it to interact with and cleave a periplasmic site in RseA (2). This cleavage exposes the cytoplasmic domain of RseA to proteolysis by RseP (3), releasing a fragment (RseA complexed with σ^E) which is further processed (4) to remove the RseA fragment, allowing σ^E to interact with RNA polymerase (RNAP) and permit transcription of the *rpoE*/σ^F regulon. The **Cpx response** is induced by the overproduction of NlpE, misfolded/aggregated pilus subunits and proteins associated with surface attachment. The stress signal is transduced through regulator/detector, CpxA (5). This activates the ATPase domain of CpxA (6), allowing the phosphorylation and activation of response activator CpxR (7). Activation of CpxR allows transcription of the *cpx* regulon, *degP* and the inhibition of the *rpoE*/σ^F regulon. For induction of the ***psp* response** the periplasmic domain of PspC detects stressors inducing the *psp* response. When a stress threshold has been reached (8), a conformational shift in the cytoplasmic domain of PspC occurs, uncovering the binding domain of PspB. PspA then interacts with PspB, causing the release of response transcription factor PspF and subsequent transcription of the *psp* regulon (10) (Ruiz & Silhavy, 2005, Rowley *et al*., 2006, Gueguen *et al*., 2009).

1.5   *Periplasmic and Outer Membrane Stress Responses*

Bacteria have evolved a plethora of interacting pathways to cope with ever-changing environmental stressors, which can cause catastrophic damage to metabolism and cellular integrity if not circumvented by stress responses. *E. coli* K-12 has six documented extracytoplasmic stress response pathways, of which three have been well-characterised: the RpoE/σ^E, Cpx, and Bae pathways. These three pathways are induced by changes to OM homeostasis and misfolded proteins in the periplasm (Figure 4) (Raffa & Raivio, 2002, Duguay & Silhavy, 2004, Ruiz & Silhavy, 2005, Rowley et al., 2006). Another stress pathway, the phage shock protein (Psp) response, is most likely induced by agents and conditions that interfere with the homeostasis of the IM: filamentous phage assembly, extreme heat-shock, high salt concentrations, ethanol and perturbation of the proton-motive force (PMF) across the IM, amongst other signals (Figure 4). A signal-transduction pathway involving multiple components, of which at least one is IM-embedded, controls induction. Stress stimuli act via the periplasmic sensor domain of an IM protein controlling the release/activity of alternative sigma factors (RpoE/σ^E) or transcription factors, as in the Cpx, Bae and Psp responses (Ruiz & Silhavy, 2005, Rowley et al., 2006).

Expression of secretins, in the absence of other components of the secretion system, imposes stress onto the host cell (secretin-stress), which somehow induces the *psp* response. For secretins, induction correlates with the mislocalisation of a significant portion of the secretin multimer into the IM: with pilotin-guided secretins which are precisely targeted to the OM (like PulD and

InvG) being non-inducers, and pilotin-independent secretins that insert into both IM and OM (like pIV) being inducers. Recent microarray analyses of global transcription during secretin-stress in *Y. enterocolitica, E. coli* K-12, and *S. typhimurium* (Lloyd *et al.*, 2004, Becker *et al.*, 2005, Jovanovic *et al.*, 2006, Seo *et al.*, 2007, Bury-Mone *et al.*, 2009, Seo *et al.*, 2009) suggests that response to secretin-stress is limited to the *psp* operon; no other stress operons seem to be induced.

Originally identified in filamentous phage f1-infected *E. coli* (Brissette *et al.*, 1990), the *psp* regulon has been found to encode some 7 proteins (PspA-G). PspA-E forms the *psp* operon; PspF is located just upstream and is divergently transcribed, while PspG is not linked to the rest of the *psp* genes. The Psp response is induced by a range of stressors that are also known to induce other stress responses (RpoE, and Cpx). However, Psp is the sole response induced by secretin-stress. The Psp response has also been shown to play a role in maintaining the PMF (therefore protecting ATP-production) under PMF-dissipating conditions, such as SecYEG translocon overloading (Kleerebezem & Tommassen, 1993, Kleerebezem *et al.*, 1996), or secretin mislocalisation. Inhibition of LPS biosynthesis was also shown to induce the psp response (Bergler *et al.*, 1994). Defects in Lol, LppX, and LPS all result in incorporation of phospholipids into the outer leaflet and general disturbance of the OM (Ruiz & Silhavy, 2005). Thus, they may induce the *psp* regulon by indirectly increasing the level of IM stress.

Induction of the *psp*-regulon requires activation of the divergent IHF-enhanced σ^{54}-dependent *pspA* promoter by the binding of the specific transcriptional activator, PspF, and general DNA-bending protein, IHF, to the respective target sequences, UAS and IHF boxes. Psp-specific activation by PspF requires detection of an inducing signal by the periplasmic domain of integral IM protein PspC of the PspBC complex. An as yet unknown signal induces a conformational shift in the cytoplasmic domain of PspC – uncovering the cytoplasmic domain of its partner protein, PspB (Figure 4; (Gueguen et al., 2009)). PspB can then bind and inactivate PspA (the negative regulator), which normally binds and inactivates PspF (the positive regulator). Now liberated of PspA-mediated inactivation, PspF can interact with the UAS sequences upstream of the *pspA-E*/*pspG* -35/-10 sequences, and together with IHF, bend the DNA allowing interaction with RNA polymerase (σ^{54}-dependent) and transcription of the *pspA-E* and *pspG* operons. *psp* mRNA transcript and promoter analysis revealed that the *psp* gene, *pspE*, is independently transcribed from a temperature-sensitive σ^{70}–dependent promoter such that *pspE*-mRNA levels are not altered when the response is induced (Brissette et al., 1990, Brissette *et al.*, 1991, Jovanovic *et al.*, 1996, Jovanovic *et al.*, 1997).

The effector proteins of the Psp response are PspD, E, and G. PspD is a cytoplasmic protein associated with the IM but has no involvement in the regulation of Psp-induction. PspE localises to the periplasm where it is thought to act like a rhodanese, reconstituting Fe-S clusters required for respiration (Adams *et al.*, 2002, Adams *et al.*, 2003). The *pspG* gene, which is unlinked to the *psp*-operon, is co-regulated with the operon thanks to its PspF-σ^{54}-

dependent IHF-enhanced *pspA* promoter. Additionally, it is only found in organisms with *a psp* operon. Expression analysis showed that PspG had a role in regulating the motility of *E. coli* (Green & Darwin, 2004, Lloyd et al., 2004, Jovanovic et al., 2006).

Investigation into the protein-protein interactions of the Psp proteins indicated a PspBC complex transduces inducing signals from the periplasm to the PspAF complex in the cytoplasm (Adams et al., 2003). The involvement of PspBC is dependent on the nature of the inducing stimuli, as they are not required for *psp*-induction by heat-shock, but are essential for induction by secretin mislocalisation (the most severe inducer of the *psp* response). Adding to the confusion, PspBC are partially required for induction during overloading of the SecYEG IM translocon, known to dissipate the PMF (Kleerebezem & Tommassen, 1993, Kleerebezem et al., 1996). Thus, it seems the requirement for signal-transducing positive activators PspBC is conditional and dependent on the strength of the stimuli**.**

A wide range of stressors induces the Psp response and intermediary events that transduce stress signals to the *psp* regulon are not clear. Microarray analysis of transcription during secretin-stress showed the conditional involvement of the ArcAB, a two-component metabolic regulation system in inducing Psp under microaerobiotic conditions. The ArcAB system is thought to regulate the switch between aerobic and anaerobic metabolism (Jovanovic et al., 2006, Jovanovic *et al*., 2009).

Several studies have implicated the RpoE/σ^E-regulon in Psp responses, however, the outcome of these studies are different in three enteric bacteria, *Y. enterocolitica*, *E. coli*, and *S. typhimurium*. A transposon-insertion library screened for secretin-sensitivity in *Y. enterocolitica* identified an RpoE/ σ^E-null mutant containing an uncharacterised suppressor mutation (RpoE/σ^E is essential in *Y. enterocolitica*). Other mutants isolated included DegS-null and TrkA-null mutants (Seo et al., 2007). DegS is a periplasmic protease essential for activating the RpoE pathway. TrkA is a potassium transporter, probably implicated due to the involvement of potassium in regulation of osmotic pressure. The mutations in *rpoE*, *degS*, and *trkA* obtained in this screen did not affect Psp induction (Seo et al., 2007). Therefore, these proteins have a buffering role in the protection of *Y. enterocolitica* from secretin stress in the absence of the induction of the *psp* regulon. Conversely, microarray analysis of *E. coli rpoE* knock-down mutants displayed little effect on *psp* induction (Becker et al., 2005, Rhodius *et al.*, 2006, Seo et al., 2007). In contrast, analysis of *S. typhimurium rpoE* mutants revealed the upregulation of PspA protein due to PMF dissipation. Thus, depending on the organism, the RpoE/σ^E response may be implicated in maintaining PMF either directly or indirectly through PspA induction.

Interestingly, *P. aeruginosa*, a bacterium lacking a *psp* regulon, is not sensitive to expression of native secretins, and this secretin resistance does not require an AlgU response (an ortholog of RpoE) (Seo et al., 2009). Transcriptome analysis revealed no regulatory changes due to secretin expression in wild-type strains. However, random transposon mutagenesis isolated a secretin sensitive

mutant with an insertion at locus PA0943. This PA0943-null mutant was unable to export substrates of the Hxc T2SS, and displayed drastically reduced production of T4PS secretin XcpQ and its mislocalisation. Global OMP localisation, however, was not affected, and secretin mislocalisation could be suppressed by the expression of plasmid-borne *Y. enterocolitica* PspBC. PA0943 showed sequence similarity to LppX and Lol proteins involved in periplasmic trafficking, and fractionations showed that it was a periplasmic protein. Lpp is a murein lipoprotein involved with the Lol lipoprotein trafficking pathway, and was among those genes found to be up-regulated during psp-expression (Jovanovic et al., 2006).

In summary, although various stressors affecting the IM are shown to induce the *psp* operon, the inducing signal of secretin-stress has not been revealed. The data from *E. coli*, *Y. enterocolitica*, and *S. typhimurium* have not identified any proteins in the secretin-stress induction pathway other than those encoded by the *psp* regulon itself (Ruiz & Silhavy, 2005, Jovanovic et al., 2006, Rowley et al., 2006, Seo et al., 2007, Jovanovic et al., 2009). Establishing a quantitative correlation between an IM-localising secretin (whose gate is not tightly closed) leaky to protons with the extent of *psp* induction would help clarify the correlation between PMF dissipation and *psp* induction.

1.5 Hypothesis and Aims

Current data suggest that structural elements within pIV give rise to the body of the gate that emanates from the OM-embedded portion of the channel, and

must include a defined mechanism, by which the opening and closing of the gate is tightly regulated. The structural elements forming the gate would include a molecular "latch" specifically triggered by phage assembly; and a mobile gate-structure blocking the channel. Although low-resolution structures of several secretins are available, the major structural elements and the residues involved in the likely delicate opening mechanism have not been identified at the level of primary sequence for any of the secretins. Given the involvement of a large secretin family domain in both OM targetting and gating, it could be hypothesised that gate and trans-OM structural-elements would be clustered as discrete regions along the primary sequence of this domain.

To test this hypothesis, a collection of gate mutants was obtained by random mutagenesis and positive selection for pIV mutants with a leaky phenotype. This approach identified two clusters of leaky pIV mutations between predicted transmembrane segments. These clusters delineated, at the primary structure level, motifs that likely build the pIV gate.

To increase understanding of how the leaky pIV channels affect filamentous phage assembly, the set of leaky mutants was tested for the ability to complement an f1 gene-IV null mutant. The effect of a leaky secretin on the physiology of the host bacterium was assessed by testing *E. coli* cultures expressing the set of leaky gate mutants for sensitivity to detergent sodium deoxycholate, and antibiotics vancomycin and bacitracin.

The ability of an organism to respond effectively and efficiently to stress affects its ability to survive. The f1 phage secretin, pIV, is commonly used to analyse induction of the *psp* regulon under varying conditions. Random mutagenesis provided a large library of different pIV mutants, thus it was advantageous and opportunistic to survey the accumulation of PspA protein thoughout the mutant set. This determined, using the largest collection of secretin leaky-gate mutants to date, whether the severity of the leaky phenotype correlates with the induction of the *psp* regulon.

2.0   Materials & Methods

2.1   Bacterial Strains and Plasmids

All strains used were non-pathogenic laboratory *E. coli* strains and are listed in Table 1. Plasmids used in this project are listed in Table 2.

All bacterial cultures were grown in 2xYT rich media (Sambrook *et al.*, 1989), or M63 minimal media (Miller, 1972, Neidhardt *et al.*, 1974) supplemented with appropriate antibiotics (Amp for pUC19, and Cm for pPMR132), and – in the case of minimal media – micronutrients; Thiamine (15 µg/mL); Biotin (2 µg/mL); and as sole-carbon source either maltose or maltopentaose (Sigma-Aldrich) at 0.2%. pIV expression from pPMR132 was induced with IPTG at stated concentrations. All plates were solidified using Bacto-agar (BD) at a concentration of 1% (w/v). Transformed cells were recovered in SOC media (Sambrook et al., 1989). Incubations were performed at 37°C.

Table 1: List of *E. coli* K12 strains used

Strain	Parent	Genotype	Source
XL1-Red		F⁻ *endA1 gyrA96*(nal^R) *thi-1 relA1 lac glnV44 hsdR17*(rK- mK+) *mutS mutT mutD5* Tn10	Stratagene
XL1-Blue		*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^qZΔM15* Tn*10* (Tet^R)	Stratagene
K1508	MC4100	F⁻ *araD⁻ lacΔ* U169 *relA- thi- rpsL* (strep^R) Δ*LamB106*	
DH5α		*fhuA2 Δ(argF-lacZ)*U169 *phoA glnV44* Φ80 Δ(*lacZ*)M15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	
K2040	K1508	*degP41*(Δ*PstI*-Km^R)	(Baneyx & Georgiou, 1991)

Table 2: List of plasmids used

Name	Expressed protein	Reference
pPMR132	pIV	(Russel, 1994)
pGZ119EH	-	(Lessl *et al.*, 1992)
pUC19	LacZα	(Yanisch-Perron *et al.*, 1985)
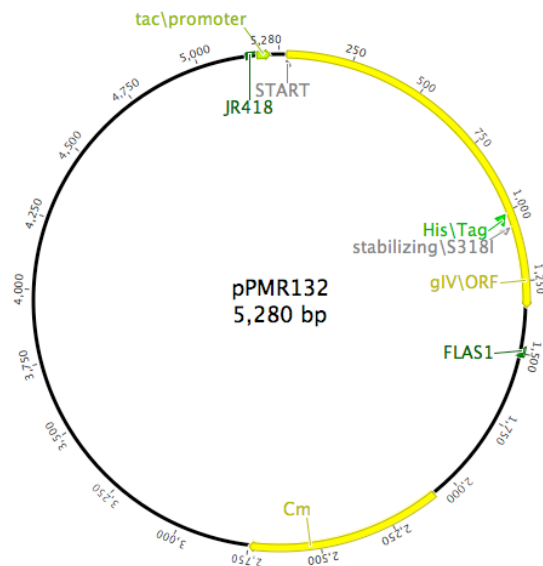
Figure 5: Plasmid map of pPMR132 showing position of *gIV* ORF (pIV) and sequencing primers FLAS1 and JR418.

2.2 Mutagenesis

Mutations were introduced at random into plasmid pPMR132 - which carries *gIV* under the control of the *tac* promoter – by passaging through a mutator strain. *E. coli* XL1-Red (Stratagene, La Jolla, CA) has mutations in *mutS*, *mutD*, and *mutT* of three major DNA repair pathways resulting in a mutation rate 5000-fold higher than that of wild-type *E. coli* (Greener & Callahan, 1994, Greener *et al.*, 1997). Commercial XL1-Red competent cells were transformed with 50 ng pPMR132. After recovery in SOC media, two aliquots of 450 μL were used to inoculate two 20mL 2xYT Cm (25 μg/mL) cultures. As a control to monitor mutagenic efficiency LacZα producing plasmid pUC19 was mutagenized under the same conditions applied to pPMR132, except that one 2xYT Amp (100 μg/mL) culture (20 mL) was inoculated with 900 μL of transformed cells. Plate counts and spectrophotometry (optical density (OD) at 600nm) were used to monitor growth of the XL1-Red cultures. After 20 generations of growth the mutagenized pPMR132 and pUC19 plasmids were extracted from XL1-Red, thus obtaining the libraries of mutant-pPMR132 and mutant-pUC19.

2.3 Assessing Mutagenic Efficiency

Mutagenic efficiency was assessed by the frequency of LacZα-negative pUC19 clones in the library of pUC19 mutants extracted from XL1-Red. To identify *lacZα* mutants, the mutant pUC19 library was electroporated into the strain DH5α (50 ng/80 μL of competent cells). 50 μL aliquots of 10^3-fold dilutions were spread onto 2xYT plates supplemented with Amp (100 μg/mL), IPTG (0.1 mM), and X-Gal (80 μg/mL). Mutation frequency was calculated as the number of

white colonies formed divided by the total colony count. Background mutagenesis of pUC19 in the non-mutator strain was also assessed by the same procedure outlined above except the primary host was the non-mutagenic strain DH5α that contains wild-type alleles of *mutS*, *mutD*, and *mutT*.

2.4   *Evaluation of Selection Plating*

E. coli selection strain K1508 lacks maltooligosaccharide-specific porin LamB (Klebba *et al.*, 1994, Hofnung, 1995) and therefore is unable to grow on maltopentaose as a sole-carbon source. Wild-type pIV does not permit entry of maltopentaose and does not permit growth on M63 media supplemented by maltooligosaccharides as a sole carbon source. In contrast, expression of leaky gate pIV mutants (S324G and A329T) allowed growth on minimal-media agar plates containing maltopentaose as the sole carbon source, as shown previously (Marciano et al., 1999), providing the means to select for leaky gate mutants. Therefore, M63 minimal-media agar plates supplemented with 0.2% Maltopentaose were used to select for pIV transformants containing mutations affecting the gate structure ("leaky mutants").

Aliquots (80 μL) of electrocompetent selection strain K1508 were transformed with 50 ng, 100 ng, and 200 ng of mutant-pPMR132 library, and recovered for 1 hour, at which point 2 mL of SOC medium supplemented with Cm (37.5 μg/mL) was added, bringing Cm concentration to 25 μg/mL. Incubation continued for a further 2-4 hours. The cells (1 mL) were harvested by centrifugation at 21000 × g for 2 min at room temperature. Cell pellets were washed twice in 0.9% NaCl to

remove nutrients before being finally resuspended in 1 mL 0.9% NaCl. 100 μL aliquots of washed transformed K1508 cells were serially diluted 10^3-, 10^4-, and 10^5-fold. 50 μL of these dilutions were subsequently plated onto two sets of control plates containing: (a) 2xYT medium supplemented with Cm (25 μg/mL) and IPTG (50 μM); and (b) M63 medium supplemented with Cm (20 μg/mL), IPTG (50 μM), and maltose (0.2%) as the sole carbon source. 50 μL aliquots of undiluted 50 ng and 200 ng transformations were plated onto selection plates: M63 medium supplemented with maltopentaose (0.2%), Cm (20 μg/mL), IPTG (50 μM).

Minimal media plates were incubated for 48 hours at 37°C in a box containing a damp towel to prevent the plates from drying out. 2xYT plates were incubated in a box overnight at 37°C. Plating efficiency on minimal plates was assessed as a ratio of transformants on M63 maltose plates *vs.* the number of transformants on 2xYT plates. Frequency of the leaky pIV mutants was calculated as a ratio of the number of transformants on M63 maltopentaose (selection) plates to the number of transformants on M63 maltose (control) plates.

2.5   Screening & Sequencing

A total of 103 colonies were individually picked off the maltopentaose selection plate. These colonies were passaged onto a maltopentaose selection master plate, taking care to obtain single colonies and thereby purify mutants away from the potential background of cells containing unmutated pPMR132 surrounding mutant colonies (due to cross-feeding effects common in

auxotrophic selection). Individual colonies from the master plates were used to inoculate 5 mL cultures in 2xYT broth supplemented with Cm (20 µg/mL). To 1 mL of each culture DMSO was added to 7% final concentration and they were then stored at –80°C. The pPMR132 DNA was recovered from the remaining 4 mL of each culture, using the High Pure Plasmid Isolation Kit (Roche Diagnostics, Indianapolis). In some cases cell pellets were harvested and frozen at –80°C until further processing. Purified DNA was quantified by spectrophotometry using a Nanodrop 1000 spectrophotometer.

gIV was sequenced in all 103 purified mutant plasmids using 3' reverse primer FLAS1 (5'-TCTTCTCTCATCCGCCAAAACAGC-3') and the forward primer JR418 (5′-CGACATCATAACGGTTCTGGCAA-3′) complementary to the vector sequences upstream and downstream of *gene IV*, respectively. The two sequence reads obtained using these primers covered the entire *gIV* sequence. The Alan Wilson Centre Genome Service performed sequencing reactions. Sequence data was analysed using the sequence analysis package Vector NTI (Invitrogen).

2.6   Vancomycin, Bacitracin, Deoxycholate Sensitivity Assays

Vancomycin (Van) and Bacitractin (Bac) *E*-test strips (AB Biodisk), which provide an antibiotic gradient ranging from 2 to 256 µg/mL, were used to assess minimal inhibitory concentrations. Deoxycholate diffusion discs used to test sensitivity to sodium deoxycholate (DOC) were prepared by pipetting 50 µL of 10% DOC solution (w/v) onto 6 mm cellulose discs in 10 µL aliquots, allowing

up to 1 hour drying time between applications. Assays were carried out on 2xYT agar plates containing Cm (20 µg/ml) and IPTG (0.1 mM). 100 µL of 10^2-fold diluted fresh overnight culture of each mutant was spread on two separate plates and allowed to dry. Van and Bac *E*-test strips were placed together on a single plate and DOC discs on a separate plate. Plates were incubated for 18 h at 37°C.

2.7   Functionality of pIV Mutants in Phage Assembly

As filamentous phage f1 requires F pili for infection, an F′ episome was introduced into the set of strains carrying the Cm^R vector pGZ119EH, pPMR132 or its *gene IV* mutant derivatives, by mating with XL1-Blue and selecting Cm^R Tet^R transconjugants on 2xYT plates containing Cm (20 µg/ml) and Tet (10 µg/ml). Transconjugants were colony-purified and used for phage plating assays. Aliquots from each mutant culture were added to soft 2xYT agar and overlaid onto 2xYT agar plates supplemented with Cm (25 µg/ml) and IPTG (0.1 mM). Serial dilutions of *gene IV* mutant phage, R484, carrying a 1053 nt in-frame deletion within *gene IV* (Brissette and Russel, 1990) and wild-type f1 were spotted on the cell lawn of each mutant. The extent of complementation was judged by comparing the efficiency of plating (e.o.p.) of phage R484 on a particular mutant with that on a lawn of cells expressing wild-type pIV encoded by pPMR132 and noting R484 plaque size in comparison with wild-type phage on the same mutant. R484 and f1 plaque sizes were identical on cells containing pPMR132.

2.8 Total E. coli Protein Extracts

Exponentially growing cultures of leaky pIV mutants (in pPMR132) or vector control (pGZ119EH) were prepared as described above and pIV expression was induced for 1 hour by 1 mM IPTG. Total *E. coli* protein extracts were obtained by precipitation in 5% Trichloroacetic Acid (TCA); 0.5 mL of culture was mixed with an equal volume of 10% TCA and incubated on ice for at least 30 min. Protein extracts were harvested by centrifugation at 16060 × g for 10 min at 4°C. The pellet was washed twice in 1 mL acetone, recovered by centrifugation between washes and finally allowed to air dry for 5-10 min. Pellets were resuspended in 4% SDS containing protease inhibitor (Complete, Roche); the resuspension volume was adjusted so that the lysates corresponded to a cell density of 10 OD_{600} units/mL. The resuspended pellets were boiled for 10 min and stored at –80°C until further analysis.

2.9 Protein Electrophoresis and Western Blotting

The TCA precipitated cell lysates prepared as above were thawed on ice and briefly spun down to remove any insoluble pellet. 18 µL of sample was mixed with an equal volume of Laemmli SDS-PAGE loading buffer as cited by Laemmli *et al* (1970). 10 µL of each sample was loaded onto a 12% polyacrylamide gel (for pIV immunoblotting) and two 14% polyacrylamide gels (for PspA and Trx immunoblotting). 5 µL of Novex Sharp pre-stained protein marker (Invitrogen, San Diego) was used as a molecular weight standard. Gels were run at 60 mA constant current in BioRad mini-PROTEAN 3 chambers until the loading dye-front reached the bottom of the gel.

Proteins were transferred to nitrocellulose membranes in Tris-Glycine buffer containing 20% Methanol at 300 mA for 1 hour in a Hoeffer transfer device. The membranes were then placed in 1x TBST buffer (Tris 50 mM, NaCl 150 mM, pH 8.0) containing 5% skim milk powder (Pams, New Zealand) blocking buffer and incubated at 4°C overnight with gentle rocking. The remaining steps of the procedure were carried out at room temperature.

To block binding of non-specific antibodies, pIV antiserum was preincubated with a Triton X-100 extract of K1508 (the host strain for pIV expression) not expressing pIV: For one 7 × 10 cm nitrocellulose membrane 4 μL of Rabbit αpIV serum was mixed with 100 μL of 1x TBST containing 5% skim milk powder (Pams, New Zealand), 100 μL K1508 cold competitor lysate and 2 μL 100x protease inhibitor in a 1.5 mL tube and rotated for 30-60 min. Rabbit primary antibodies were used at the following dilutions; αpIV (1:1000), αPspA (1:5000), and αTrx (1:1000). Alkaline phosphatase-conjugated Goat anti-Rabbit IgG (Sigma-Aldrich) was used as the secondary antibody at the same dilution as the corresponding primary antibody. All antibodies were diluted into 1x TBST containing 5% skim milk powder. Membranes were incubated with primary or secondary antibody in 50 mL Falcon tubes placed on a rotator for 90 min at room temperature. After primary antibody incubations, membranes were washed three times in 1x TBST, then three times in 1x TBST with 5% skim milk. After incubation with secondary antibody the membranes were washed four times in 1xTBST and the protein bands of interest were detected by an alkaline phosphatase assay in the appropriate buffers (Blake *et al.*, 1984).

3.0 Results

3.1   Mutagenesis

To construct the *gIV* mutant library, plasmid pPMR132 was passaged through the mutator strain XL1-Red (*mutD*, *mutS*, *mutT*). As a control to monitor the mutation frequency a mutant library of pUC19 was constructed in parallel. Growth of transformed XL1-Red was monitored over a period of 24 hours by spectrophotometry and plate count (data not shown).

The mutator strain XL1-Red is extremely unstable and upon passaging its mutator mutations revert to the wild-type alleles or they acquire compensatory mutations (XL1-Red Manual, Stratagene). To assess the effectiveness of commercial mutator strain XL1-Red to cause mutations above the background level we mutagenized plasmid pUC19 (as described in Materials & Methods). Transforming the mutant pUC19 library into blue-white reporter strain *E. coli* DH5α made it possible to determine the frequency of colonies with white phenotype and hence mutated *lacZα* sequence in pUC19. Spontaneous mutagenesis in the non-mutator (wild-type) strain DH5α was also investigated by growing pUC19-transformed cultures for 24 generations under the same conditions used for mutator strain XL1-Red. Table 3 displays the data obtained from the assessment of mutagenic efficiency. From this we can see that by passaging through XL1-Red we can cause an increase of mutation frequency from 0.2% to 11.1% of recovered plasmid pUC19. Background mutagenesis observed in DH5α therefore results in a ~60-fold lower frequency of LacZα-negative transformants than obtained in the mutator strain XL1-Red.

36

Table 3: pUC19-*lacZα* Mutation Frequencies

pUC19 Transformants	Strain for mutagenesis	
	XL1-Red	DH5α
Wild-type (LacZ$^+$)	834	2759
Mutant (LacZ$^-$)	105	5
TOTAL	939	2764
Frequency of LacZ$^-$ mutants	11.1%	0.2%

Table 4: Transformation and Plating Efficiency of K1508+pPMR132 on Rich, Minimal, and Selective Media

Amount of DNA Transformed (ng)	Transformation Frequency (e.o.p.)		
	2xYT	M63 Maltose	M63 Maltopentaose
50	3.83×10^8	7.92×10^6	1.98×10^3 (0.025%)
100	2.06×10^8	1.72×10^7	Not Plated
200	2.72×10^8	5.12×10^7	1.69×10^4 (0.033%)

The high frequency of *lacZα* mutants obtained in control experiments with pUC19 confirmed that commercial XL1-Red competent cells contained the mutator mutations and that the mutagenesis protocol described here is valid. Given the positive outcome of the control experiment mutagenesis of *gene IV* was undertaken. pIV expression tends to somewhat decrease the plating efficiency of host strains on minimal media (J. Rakonjac, unpublished observation). To monitor this effect, plating efficiency of pIV reporter strain *E. coli* K1508 transformed with mutagenized pPMR132 library was compared between rich (2xYT) and minimal (M63) media (as described in Materials & Methods). The transformants were also plated on selective M63 maltopentaose plates to select for leaky pIV mutants (Table 4). Plating efficiency of transformants on the M63 minimal plates was 2.06% relative to that on 2xYT plates. Given the pUC19 trial mutagenesis, it was expected that 10% of our mutant pPMR132 library would have mutations in pIV; only some of those are expected to exhibit a leaky phenotype. The plating efficiency of leaky mutants was then calculated by comparing the number of transformants obtained on the M63 maltopentaose plates with the numbers obtained on the corresponding M63 maltose plates. Plating efficiency on M63 maltopentaose selection plates was 0.025%-0.033% relative to plating on M63 maltose plates (Table 4). This means that plating efficiency is about 3 orders of magnitude lower on M63 Maltopentaose than on M63 Maltose plates. In comparison to the *lacZα* control mutagenesis (11% white mutants), this is a much lower frequency. This is expected, since only a small number of pIV mutants are likely to be leaky, following the general rule that mutations which result in a specific function are far less numerous in the sequence space than deleterious mutations.

Colonies on the M63 maltopentaose selection plates displayed variable mucoid morphologies and colony lysis occurred within 72 hours of plating. These observations suggest that expression of leaky pIV mutants leads to a decrease in organismal fitness. It has been shown that only about half of pIV produced in an *E. coli* cell integrates into the OM; the rest is mistargeted to the IM (Russel & Kazmierczak, 1993). Leaky mutant S324G of pIV was shown to "open" more frequently than the wild-type channel (Marciano et al., 1999), hence similar effects were expected in our maltopentaose-permissive leaky mutants. Mucoid morphology and colony lysis are indicators that the periplasmic and OM stress responses have been stimulated (Ruiz & Silhavy, 2005, Tam & Missiakas, 2005, Rowley et al., 2006).

3.2   Sequencing of Mutants and Analysis

To identify mutations in leaky mutants, 103 colonies were picked from maltopentaose selection plates; from which pPMR132 was purified and *gIV* sequenced using the flanking primers FLAS1 and JR418 (Figure 5). 36 individual missense single-mutations to 34 residues were identified in the secretin homology domain. Sequence analysis revealed that the majority (32 of 36 mutations) of these point mutations were grouped in 2 clusters located in the secretin homology domain. Four mutations were found outside the clusters: 3 in the N-terminal domain, and 1 just inside the secretin homology domain. Two positions were represented by two different mutational changes: G259S/D and A329T/V (Table 5 and Figure 6).

In addition to the single mutants, 6 double mutants were also identified, 5 of which (Y17C+T334A, M62I+S325P, N89D+A329V, G147V+A191T, I183M+V270A) have a mutation in the N-terminal domain as well as in the secretin homology domain. The remaining double mutant (A191T+T327A) had both mutations within the secretin homology domain (Table 5 and Figure 6). The individual mutant residues were not separated, so the independent effects of each of the two mutations, alone, are unknown. Furthermore, at least one of the two mutations was mapped to a residue that was already represented in the single-mutant cohort; hence the double mutants were excluded from further analysis.

The two regions in which the C-terminal mutations are predominantly clustered were named GATE1 and GATE2. GATE1 spans 39 amino acids, between residues 259 and 298 of the mature pIV sequence, and contains 22 mutations to 21 distinct residues. GATE2 is smaller, spanning just 13 amino acids between residues 322 and 334, of which 11 distinct mutations map to 10 residues. Transmembrane β-strand prediction algorithm (TMBETA) (Gromiha *et al.*, 2005) predicted the presence of two sets of short β-strands flanking each gate region. A His-tag used for pIV purification is located between predicted β-strands flanking the gates (within the inter-gate region), suggesting this region is exposed and is likely a short loop. PFAM sequence alignments show that the GATE1/2 regions and the inter-gate loop between the predicted β-strands are all located within regions of low conservation, seen as gaps in the alignment (Figure 3).

```
             N0              N3            GATE1   GATE2
                                          |—|  His tag  |—|
```

```
        C                                                    I
  1 QVIEMNNSPL RDFVTWYSKQ TGESVIVSPD VKGTVTVYSS DVKPENLRDF FISVLRANNF DMVGSNPSII QKYNPNNQDY
        |-a-helix- |    |-|                     |---a-helix--||--|        |---|
        ───────────────────────────────────────── N0 ─────────────────────────────────────

        D                          V Y                        V
 81 IDELPSSDNQ EYDDNSAPSG GFFVPQNDNV TQTFKINNVR AKDLIRVVEL FVKSNTSKSS NVLSVDGSNL LVVSAPKDIL
        |---b--|                |-a-helix -|       |--|       | --| |---
     N0                        ─────────────────────── N3 ───────────────────────────────

                 M            T
                 V            T
161 DNLPQFLSTV DLPTDQILIE GLIFEVQQGD ALDFSFAAGS QRGTVAGGVN TDRLTSVLSS AGGSFGIFNG DVLGLSVRAL
    --a-helix- --|  |------b-----|                       |--b--|         |----b---|
     N3            ───────────────── SECRETIN FAMILY DOMAIN ───────────────────────

           S          A
           D    F   D DI S V DCA D   PGKA   L   KC SIV            *
241 KTNSHSKILS VPRILTLSGQ KGSISVGQNV PFITGRVTGE SANVNNPFQT IERQNVGISM SVFPVAMAGG NIVLDITSKA
    |-b--| |----b----|                          |----b----|       |--b--|
               ───────────────── SECRETIN FAMILY DOMAIN ───────────────────────

            P    T  V   A I
          N GPFAGVN N VAS
321 DSLSSSTQAS DVITNQRSIA TTVNLRDGQT LLLGGLTDYK NTSQDSGVPF LSKIPLIGLL FSSRSDSNEE STLYVLVKAT
           |--b--| |-b--| |----b----|                    |---b---|         |---b----
               ───────────────── SECRETIN FAMILY DOMAIN ───────────────────────

401 IVRAL
    ----|
```

Figure 6: Schematic of pIV domain organisation and sequence-map of mature pIV.

Top: Linear schematic representation of pIV structure and domain organization. Orange, N-terminal domain; pink, C-terminal secretin family domain. Green and black crosshatched boxes, GATE1 and GATE2 regions; green lines across the box, leaky mutations A121V, D123Y, G147V and I183V. Red boxes, transmembrane β-strands predicted in the C-terminal secretin family domain using the TMBETA algorithm (Gromiha et al., 2005). The position of the His tag is indicated with an asterisk.

Bottom: Sequence of mature pIV. Residues highlighted in green, the positions of leaky mutations; residues above the sequence, amino acids replacing the wild-type residues (black letters and blue letters, respectively, indicate the missense residues obtained in this work and isolated previously). Boxed residues, amino acids deleted or replaced by site-specific mutagenesis: ΔGATE2 (ΔGATE→G), I312G and I316G. The β-strands are indicated using red lettering and α-helices are labelled in blue. Putative β-strands and α-helices in the N0 and N3 subdomains are derived from the model in figure 9, based on the co-ordinates of EscC (Spreter et al., 2009) and obtained using Swiss Model (Arnold *et al.*, 2006). Numbering of the residues is for the mature pIV (after the cleavage of the 21-residue signal sequence). Double mutants have been highlighted as pairs in yellow (Y17C+T334A), light blue (M62I+S325P), dark blue (N89D+A329V), dark yellow (G147V+A191T), grey (I183M+V270A) or black (A191T+T327A).

3.3 Characterisation of Leaky-Gate Mutants

3.3.1 Van, Bac, and DOC Sensitivity

To characterise leaky-gate mutants isolated in the genetic screen, the mutant set was examined for increased sensitivity to Van, Bac and DOC. Results of these tests are presented in Table 5. Of particular note is the overall insensitivity of leaky mutants to antibiotics. Of the GATE1 mutants; 16 had slightly increased sensitivity to DOC; 7 were sensitized to Van and/or Bac. However, only 5 GATE1 mutants displayed increased sensitivity to all 3 molecules tested, and of these only 3 are highly sensitive (N269D, E292K, N295S). Conversely, mutants with multiple sensitivities were more common in GATE2; of 14 mutants, 12 showed increased sensitivity to at least one molecule tested. The two most abundant mutants, A329T/V, showed only slight increases in sensitivity to Van and DOC. Interestingly, those mutants that are sensitive tend to be near the borders of the gate regions. A similar, although less striking pattern can be seen for DOC sensitivity.

3.3.2 Efficiency of Phage Production

Complementation of *gene IV* deletion mutant phage (R484; (Russel, 1994)) by the leaky pIV mutants was analysed to investigate whether they were functional for phage assembly/secretion (Table 5). In the absence of pIV (i.e. in K1508 cells transformed with the vector only) the e.o.p. of the R484 mutant is five orders of magnitude lower than in the presence of the complementing plasmid pPMR132. Only one of the single mutants, F288L, completely lacked pIV assembly/secretion function. It plated R484 at the same low efficiency as cells containing the vector. Cells expressing all the other leaky mutants plated R484

at efficiencies that varied between 0.1 and 1, at least four orders of magnitude higher than the negative control (Table 5). The phenotype of phage plaques varied amongst the mutants – diameter, turbidity and size varied between mutants. The size of plaques roughly correlated with e.o.p. Thus, they retain pIV function in phage assembly and export.

All mutants, including non-functional F288L, produced pIV, as determined by Western blots of total cell lysates obtained from exponentially growing cultures (Figure 7), but the amount of detected protein varied. The functionality in phage assembly/secretion implies that some portion of pIV mutant monomers fold and assemble correctly into functional multimers (Kazmierczak et al., 1994, Russel, 1994). However, because the extent of Δ *gene IV* complementation in the leaky mutants varied (e.o.p. of R484 ranged from 0.1 to 1), multimerisation could be quite variable among the mutants and hence the portion of functional pIV multimer could be different from the total amount detected.

Of the double mutants only Y17C+T334A was unable to complement phage assembly. The T334A mutation alone causes high sensitivity to Van, Bac and DOC, and assembles phage, albeit at a reduced efficiency. In contrast, the double mutant displays relatively high resistance to Van, Bac and DOC but is completely unable to form plaques. Comparison of the double mutant with the single mutant (T334A) suggests that the Y17C mutation causes the multimer to misfold or prevents passage of the phage. Alternatively, the gate mechanism could be functionally impaired jamming it in a position sufficiently open to allow

passage of maltopentaose but not Van, Bac, DOC or phage. These possibilities

remain to be tested.

Table 5: Summary of pIV mutants and their phenotypes

	Mutation	Number of isolates[a]	f1 assembly[b]	Van MIC[c]	Bac MIC[c]	DOC (mm)[d]
WT pIV	-	NA[e]	+++	R	R	2±0
Vector	NA[e] (no pIV)	NA[e]	-	R	R	0±0
N-domain mutants	A121V	2	+++	R	R	2±0
	D123Y	1	++	R	R	3±0
	G147V	1	+++	R	R	3±0
C-domain	I183V	2	+++	R	R	2±0
GATE1 mutants	G259S[f]	1	+++	R[-]	R	2±0
	G259D[f]	2	-/+	R	R	3±0
	S263F[f]	2	+++	12±6	48±0	3±0
	G267D	1	+++	R	R	3±0
	N269D	4	+	1±0	5±1	4±0
	V270I	6	++	3±0	21±3	3±0
	P271S[f]	1	+	R	R	3±0
	I273V	1	+++	R[-]	R[-]	3±0
	G275D	1	-/+	R	R	3±0
	R276C	1	+	R	R	2±0
	V277A[f]	1	+	R	R	3±0
	G279D[f]	7	-/+	R	R	3±0
	S281P	5	-/+	R	R	2±0
	A282G	1	+	64±0	R[-]	2±0
	N283K	1	+++	R[-]	R	3±0
	V284A	5	+++	R[-]	R	3±0
	F288L	1	-	R	R	2±0
	E292K[f]	5	+	1±0	3±1	12±1
	R293C	1	+++	R	R	2±0
	N295S[g]	0[g]	+	1±0	6±0	4±0
	V296I[f]	2	+++	R	R	3±0
	G297V	1	++	R	R	3±0
GATE2 mutants	S322N	1	+	4±1	21±5	3±0
	S324G[g]	0[g]	+++	2±0	7±1	4±0
	S325P	1	+++	6±2	19±4	3±0
	S326F	1	++	13±1	R[-]	2±0
	T327A	9	+++	85±11	R	2±0
	Q328G	1	+	4±2	24±6	3±0
	A329T	6	+	R	R	3±0
	A329V	12	+++	R[-]	R	2±0
	S330N	1	+	2±1	11±1	4±0
	D331N	3	++	48±24	R	2±0
	I333V	1	++	6±1	53±11	3±0
	T334A	1	+	1±0	2±0	6±0
	N335S[g]	0[g]	++	<3[g]	ND[h]	S[g]
	N335I[g]	0[g]	+	<3[g]	ND[h]	S[g]
Double Mutants[i]	Y17C+T334A	1	–	12	128	3
	M62I+S325P	1	++	R[-]	R[-]	0
	N89D+A329V[f]	1	–/+	R	R	2
	G147V+A191T[f]	1	+++	R	R	2
	I183M+V270A	1	++	1.5	12	4
	A191T+T327A	1	++	8	64	3
Site-directed mutants[j]	ΔGATE2	NA[e]	-	2±0	11±1	4±0
	ΔGATE2→ 10G	NA[e]	-	10±2	R[-]	3±0
	I312G	NA[e]	-	R	R	0±0
	I316G	NA[e]	-	R	R	0±0

[a] The number of sequenced isolates containing a given mutation.

[b] To assess f1 assembly function, dilutions of *gIV* deletion phage R484 were spotted on cells that expressed mutant or wild-type (WT) *gIV* and on vector-containing cells that lacked *gIV*. +++, e.o.p. 0.9-1 and normal plaque size compared to wt f1; ++, e.o.p. 0.5-0.8 and reduced plaque size; +, e.o.p. 0.3-0.4 and very small plaques; +/-, e.o.p. 0.1-0.3, barely detectable plaques; -, e.o.p $\sim$10^{-5} (equivalent to plating on cells containing the vector (pGZ119EH) that lacks *gIV*).

[c] Minimal inhibitory concentrations of vancomycin and bacitracin as measured by concentration gradient strips (in μg/mL) are indicated. R, completely resistant at the highest concentration tested (256 μg/ml); R$^-$, slight growth inhibition at 256 μg/ml. Each value was obtained from two cultures (three strips per culture) and includes the standard error of the mean.

[d] Mean annular radii (mm) of growth inhibition zones around discs pre-loaded with 50 μL of 10% DOC and dried prior to use. Each value was obtained from two cultures (three discs per culture) and includes the standard error of the mean.

[e] Not applicable.

[f] Some mutants showed a low-density zone of growth around the DOC disk, Van and Bac gradient strips. When sequenced, clones purified from this zone contained secondary mutations that put the C-domain out of frame and/or reversions to wild-type.

[g] Previously described mutants (Russel et al., 1997, Marciano et al., 1999, Marciano et al., 2001). N295S and S324G mutants were reconstructed and tested in this work. For N335S and N335I the data for Van and DOC sensitivity was taken from (Russel et al., 1997) who used different assays: <3 indicates a zone of growth inhibition around the 3 μg/mL vancomycin discs (annular radii of 4 and 5 mm for N335S and N335I, respectively); S indicates sensitivity to DOC as determined by inhibition of colony formation on plates containing 1% DOC (less than 10^{-4} relative to plates without DOC for both mutants).

[h] ND, not determined.

[i] Data for double mutants are the result of a single experiment, hence errors are unavailable.

[j] Construction of site-directed mutants is described in Spagnuolo *et al*, 2010 (Appendix 5). Data in the table refer to the *degP$^+$* host strain.

3.3.3 Psp Induction by Mutant pIV Stress

pIV has no specialised localisation factors to target it efficiently to the OM. As a consequence it inserts equally into both IM and OM (Russel & Kazmierczak, 1993). ATP synthesis and trans-membrane transport of molecules across the IM in *E. coli* requires the proton gradient across the IM. Even though the pIV gate appears impermeable to K^+ ions in electrophysiology experiments, voltages of ~200 mV cause temporary and complete opening of the gate (Marciano et al., 1999). The large pore of pIV (6-8 nm) would allow massive leakage of protons, thus IM-inserted pIV could cause disruption of the PMF. PMF dissipation has been linked to induction of the *psp* regulon (Kleerebezem & Tommassen, 1993, Kleerebezem et al., 1996), although the exact inducing stimuli (in terms of secretin-induced stress) have yet to be discerned. Given that *psp* induction by secretins could be due to their effect on PMF, it was hypothesised that the leaky pIV mutants would demonstrate stronger Psp induction relative to wild-type pIV. To test this, accumulation of the most abundantly accumulating Psp protein, PspA, was analysed during the expression of leaky pIV mutants. To that end, whole cell lysates were produced from cultures expressing leaky pIV mutants and PspA was detected by western blotting. pIV was also monitored to investigate the amount in the cells. Trx protein was monitored as a loading control. The leaky pIV mutants displayed a striking variety in the accumulation of PspA protein (Figure 7), indicating a variable ability (between different mutants) to induce the *psp* regulon.
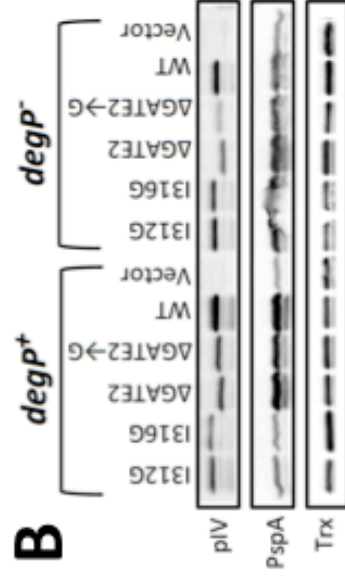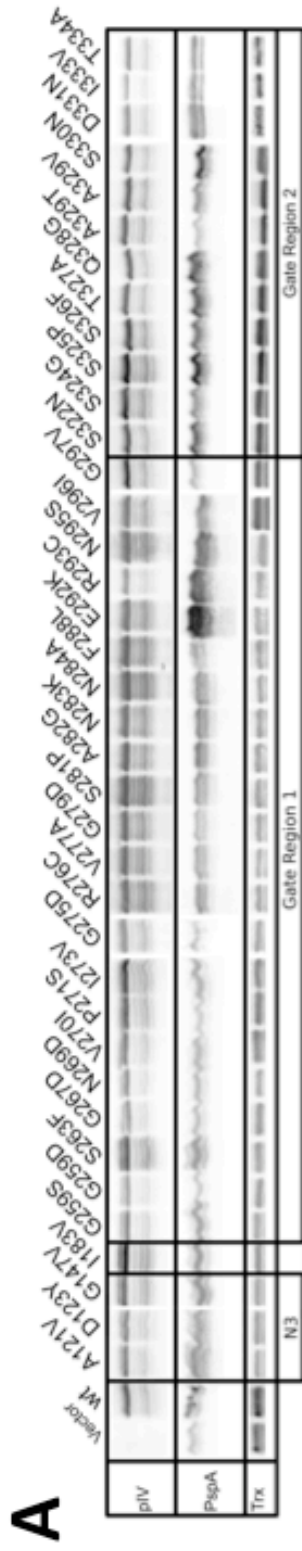
Figure 7: Western blots for pIV, PspA and Trx in whole-cell lysates of *E. coli* K1508. Vector control is K1508 transformed by pGZ119EH. wt is K1508 transformed by wild-type pPMR132. (A) Leaky pIV mutants obtained in this study. (B) Unstable pIV mutants (I312G, I316G), and site-directed leaky mutants ΔGATE2 and ΔGATE2→G in degP⁺ and degP⁻ backgrounds (Spagnuolo *et al.*, 2010)); Appendix 5.

48

The amount of PspA produced 1 hour after induction of pIV expression varied but generally correlated with the amount of pIV produced, and partially correlated with the severity of the leaky-gate phenotype (the severity being judged by the sensitivity to DOC, Van, and Bac). Several severely leaky pIV mutants (E292K, N295S, S322N, S325P and Q328G) caused elevated PspA production in comparison to wild-type pIV, despite producing similar amounts of pIV. There are, however, exceptions to the general rule that the amount of PspA is proportional to the amount of leaky pIV; two mutants that are not severely leaky (R293C and T327A) nevertheless induced high PspA production, even though they produced less pIV than wild-type. Another interesting finding was that several severely leaky mutants did not show elevated PspA production. These were N269D, D331N, I333V and T334A.

3.4 Site-Directed Mutants

Gate 2 deletion and glycine replacement mutants (ΔGATE2 and ΔGATE2→G, respectively) as well as destabilising point mutants I312G and I316G were constructed prior to commencing this project by the Post-Doctoral fellow Elodie Chabaud. These constructs were tested for sensitivity to Van, Bac, and DOC under the same conditions used for characterising the randomly acquired mutants, allowing direct comparison of results (Table 5). In addition, *pspA* mRNA accumulation was analysed by Wesley Wen using quantitative reverse transcription-PCR (qRT-PCR) (Figure 8; (Spagnuolo *et al.*, 2010)) extracted from K1508 cultures expressing pIV mutants. This approach directly measures

the amount of specific mRNA transcripts in a cell under set conditions, thus can directly measure induction of the *psp* regulon.

Not surprisingly, production of chronically unstable (or slow folding) site-directed mutants (I312G and I316G) did not induce a *psp* mRNA response beyond basal levels in qRT-PCR experiments (Figure 8). This was corroborated by western blotting whole-cell extracts produced from the same cultures (Figure 7).
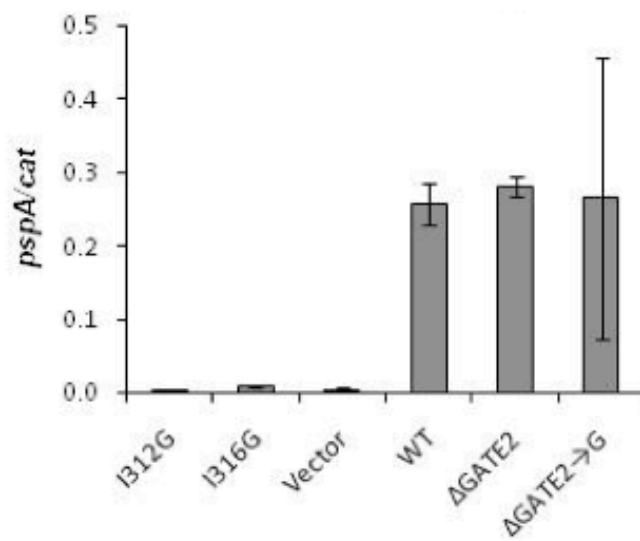
Figure 8: Induction of the *psp* regulon by secretin-stress after 1-hour expression of pIV constructs from pPMR132. Amount of *pspA* mRNA transcripts in whole cell lysates relative to *cat* gene transcripts expressed from the vector and pIV-expressing plasmids in strain K1508. Error bars represent the standard error of the mean.

3.5 N-Terminal Domain Mutants

Three single mutants in the N-Terminal portion of pIV were recovered in the maltopentaose-permissivity screen: A121V, D123Y, and G147V (Figure 6). The resistance of the 3 single N-terminal mutants to Van, Bac and DOC was comparable to that of wild-type pIV (Table 5). Western blotting for pIV (Figure 7) showed similar production of pIV to wild-type in all N-terminal mutants. Although it is not quantitative, western blotting of cells producing these mutants for PspA tentatively indicates that these mutations do not measurably increase secretin-stress in *E. coli* under these conditions.

Recently the crystal structures of N-terminal fragments of the T2SS/T3SS secretins EscC and GspD were published (Korotkov et al., 2009, Spreter et al., 2009). This allowed structural modelling of the pIV N-terminal domain. Using Swiss-Model alignment-based structural predictions (Arnold et al., 2006), pIV N-terminal residues 1-176 were modelled against the crystal structure of the N-terminal fragment of EscC. Using the resulting pIV model (Figure 9) the N-terminal mutations (A121V, D123Y and G147V single mutants only) were mapped to the same face of the N3 subdomain.
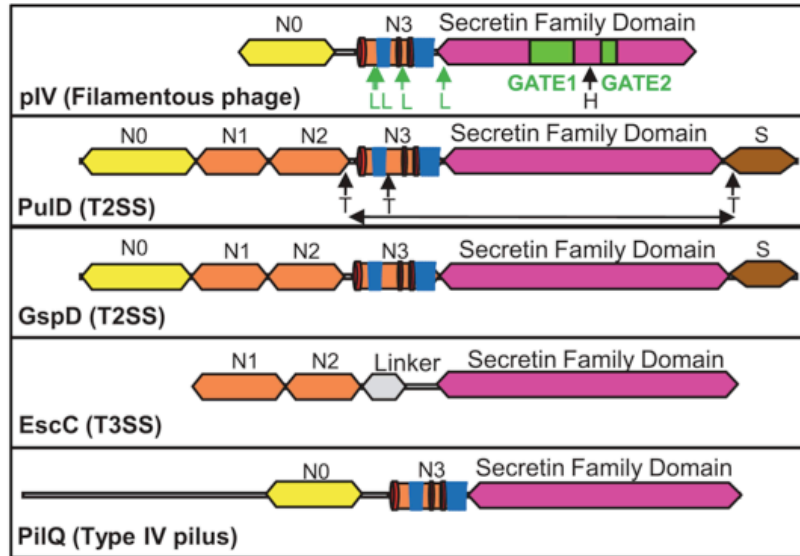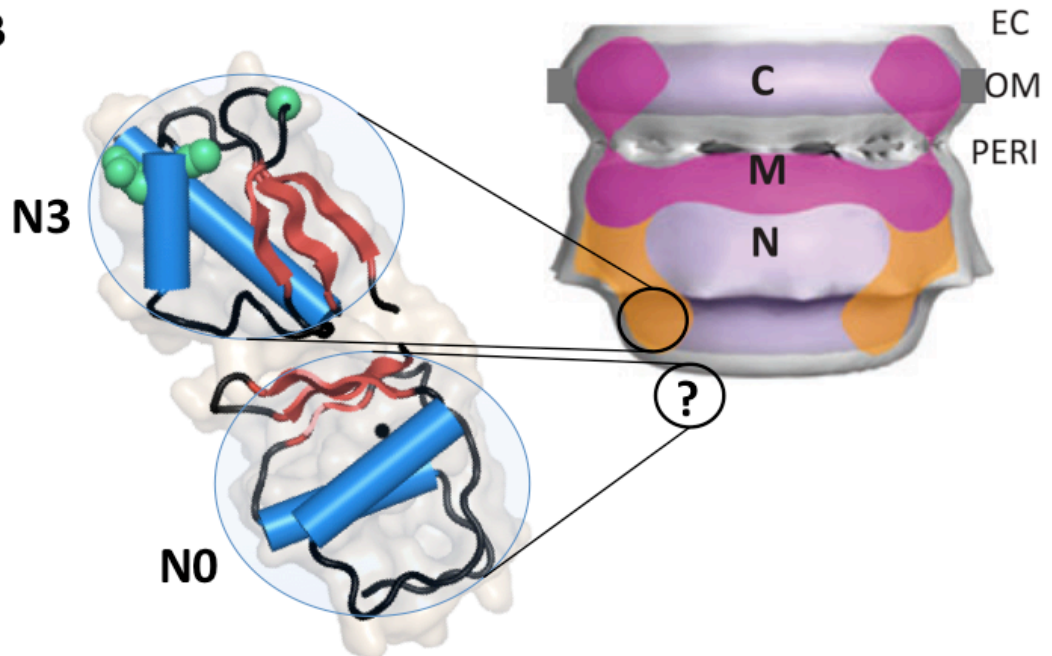
Figure 9: (A) pIV domain organization is compared to that of PulD (T2SS of *Klebsiella oxytoca*; Chami et al., 2005), GspD (T2SS of EPEC; (Korotkov et al., 2009), EscC (T3SS of EPEC; (Spreter et al., 2009), PilQ (type IV pilus assembly system; Bitter et al., 1998). Yellow diamond, N0 or FpvA-like domain; orange diamond, N1, N2 and N3, KH-fold domains; gray diamond, T3SS-specific N-terminal subdomain; pink diamond, secretin family domain; brown diamond, pilotin-binding domain; green boxes, GATE1 and GATE2 segments; L, leaky mutations; H; histidine tag; T, trypsin cut sites; Linker, variable linker. Double-headed arrow under PulD indicates the portion that is found in the trypsin-resistant core of the channel multimer. Shapes in the N3$_{pIV/PulD/GspD/PilQ}$: secondary structures modelled based on the high-resolution structures of N1$_{GspD}$ (Korotkov et al., 2009) and N1/N2$_{EscC}$ (Spreter et al., 2009): red pointed-boxes, β strands; blue shapes, α helices. Accession numbers: pIV, AAA32218; PulD,

AAA25126; GspD, ABV19019.1; EscC, CAG17523; PilQ, AAA16704.1. **(B)** Swiss-Model structural prediction of the N-terminal portion of pIV, modelled against the N-terminal EscC structure (Spreter et al., 2009). The likely position of the N3 and N0 subdomains have been circled in the low-resolution structure obtained by cryo-EM and single particle reconstruction (Opalka et al., 2003). The '?' indicates that the N0 subdomain is most likely not visible in the low-resolution structure, as reported for EscC (Spreter et al., 2009), possibly due mobility relative to the rest of the secretin. EC, extracellular; OM, outer membrane; PERI, periplasm.

4.0   Discussion

4.1   The GATE1–GATE2 region of pIV

Random mutagenesis coupled with positive selection for pIV mutants with compromised gate integrity identified 34 distinct residues involved in gating. Together with previously isolated mutants (Russel et al., 1997), a total of 38 distinct residues have a role in gating the pIV multimer. Most of the mutations (33 out of 38) cluster in two regions, GATE1 and GATE2, located within the secretin family domain.

The ability to complement assembly of Δ*gIV* phage R484 implies that a mutant pIV forms correctly folded multimers, sufficiently stable to permit efficient assembly and export of filamentous phage. Therefore, in the absence of data directly assessing stability of mutants, it can be said that all GATE mutants, except F288L, are able to form functional multimers. Following from this, the two GATE regions, likely to form part or all of the gate body or mechanism, appear to have no (or a limited) role in folding, membrane insertion, or multimerisation, which are all required to form the basic frame (walls) of the channel. The leaky mutant that cannot assemble phage, F288L, is interesting because this residue is conserved in all phage secretins, thus could have a role in substrate-triggered gate opening mechanisms. Besides the single mutant F288L, a double mutant Y17C+T334A is unable to export phage. A T334A mutation alone in pIV does not abolish phage assembly/export, but does cause relatively high sensitivity to Van, Bac, and DOC. The double mutant however, is much less sensitive to Van,

Bac and DOC, and does not complement ΔgIV phage R484, even though it allows entry of maltopentaose into the cells. The explanation of the difference between T334A and its double mutant (Y17C+T334A) requires the construction of a single Y17C mutant. Two possible characteristics could explain the observed phenotypes of the single and double mutants; (1) Y17C possibly delays folding or destabilises the multimer, leading to a lower number of functional channels per cell. This small number of channels still allows sufficient amounts of maltopentaose to enter the cell to allow colony formation, but not sufficient enough to allow killing by Van and Bac or phage assembly. (2) Alternatively, Y17C affects the gate opening/closing mechanism, causing a similar 'jammed' gate phenotype as the ΔGATE2 mutant that contains a partial GATE2 region deletion (Δ321-330, Figure 6, and Appendix 5) (Spagnuolo et al., 2010). Genetic evidence indicates the very N-terminal subdomain, in which Y17 is situated, interacts with the pI/pXI IM assembly complex (Russel & Kazmierczak, 1993, Daefler et al., 1997b). Thus, an additional explanation for the Y17C effect on phage assembly would be that it disrupts this interaction, thereby prohibiting phage assembly and export. Isolation of intergenic suppressors in gene I that restore phage assembly function in a Y17C mutant would argue in favour of the latter possibility.

Most OM channel proteins contain transmembrane (TM) β-strands that are arranged in β-barrel conformation. Recently, the crystal structures for two OM channels identified barrels of amphipathic α-helices in the OM channels of the capsular polysaccharide transporter, Wza (Beis et al., 2004, Dong et al., 2006);

and the OM channel, VirB of the Type IV secretion system, unrelated to the T4PS (Chandran et al., 2009). While these structures open the possibility for more α-helical OM channels, the TMPred predicted arrangement of loops (Appendix 1) in the secretin family domain is not likely to accommodate the transmembrane α-helices (Hofmann & Stoffel, 1993). One reason for this is that the minimal TM α-helix length of 12-14 residues (Bowie, 1997) does not allow a loop position for the pIV His-tag.

TM β-strands are predicted for the secretin family domain by the TMBETA algorithm designed for analysis of OM proteins (Russel et al., 1997, Guilvout *et al.*, 1999, Gromiha et al., 2005). However, without high-resolution structural data, the conformation of TM strands in the secretin family is subject to conjecture. Secretins have no homology to the α-helix containing OM channels VirB and Wza. However, circular dichroism spectrum of the protease resistant C-terminal fragment of PulD indicated a lower β-strand configuration than expected (Chami et al., 2005). JPred (Appendix 2) and PsiPred (Appendix 3) secondary structure and predictions suggest the presence of several β-strand segments flanking both GATE regions (Jones, 1999, Bryson *et al.*, 2005, Cole *et al.*, 2008). In addition to the agreement of TMBETA (Appendix 4) with PsiPred and JPred predictions in the locations of β-strands, neither algorithms predict α-helices flanking the GATE1/2 regions, suggesting that pIV may have a β-barrel conformation.

TMBETA (Gromiha et al., 2005) predicts 2 β-strands flanking each GATE region, two of which are predicted within the inter-GATE region, separated by a 4 residue linker (Appendix 4). The His-tag insert used to purify pIV is located within this linker, confirming that this region is a surface exposed loop. TMBETA also predicts the presence of TM segments N-terminal to GATE1, however to keep GATE1 and the N-terminal domain periplasmic they must cross the OM twice (Figure 10A). The secretin septum is located on the periplasmic side of the OM, hence the N-terminal domain, GATE1 and GATE2 are expected to be periplasmic. In contrast, the accessibility to affinity matrices and cross-linkers suggests that the His-tag and the C-terminus are exposed to the extracellular milieu. It follows then that TMBETA-predicted β-strands 298-306 and 311-320 traverse the OM in the in→out and out→in orientations, respectively, leaving GATE1 and GATE2 periplasmic and exposing the 306-311 His-tag loop to the cell surface (Figure 10A). The segment between the GATE2 and the C-terminus is expected to contain at least one TM strand to place the C-terminus at the cell surface (Opalka et al., 2003). TMBETA predicts the TM β-strands immediately flanking the C-terminal end of GATE2 (residues 335-359) (Appendix 4).

The PFAM secretin family domain alignment shows that both GATE regions (http://pfam.sanger.ac.uk/family?acc=PF00263, Figure 3) and the inter-GATE loop (residues 306-311) are poorly conserved, creating gaps in the alignment at these sites to accommodate the different loop sizes in various secretins. This is consistent with the JPRED/PsiPred predictions (Appendix 2 & 3) that the GATE2 region is a loop.
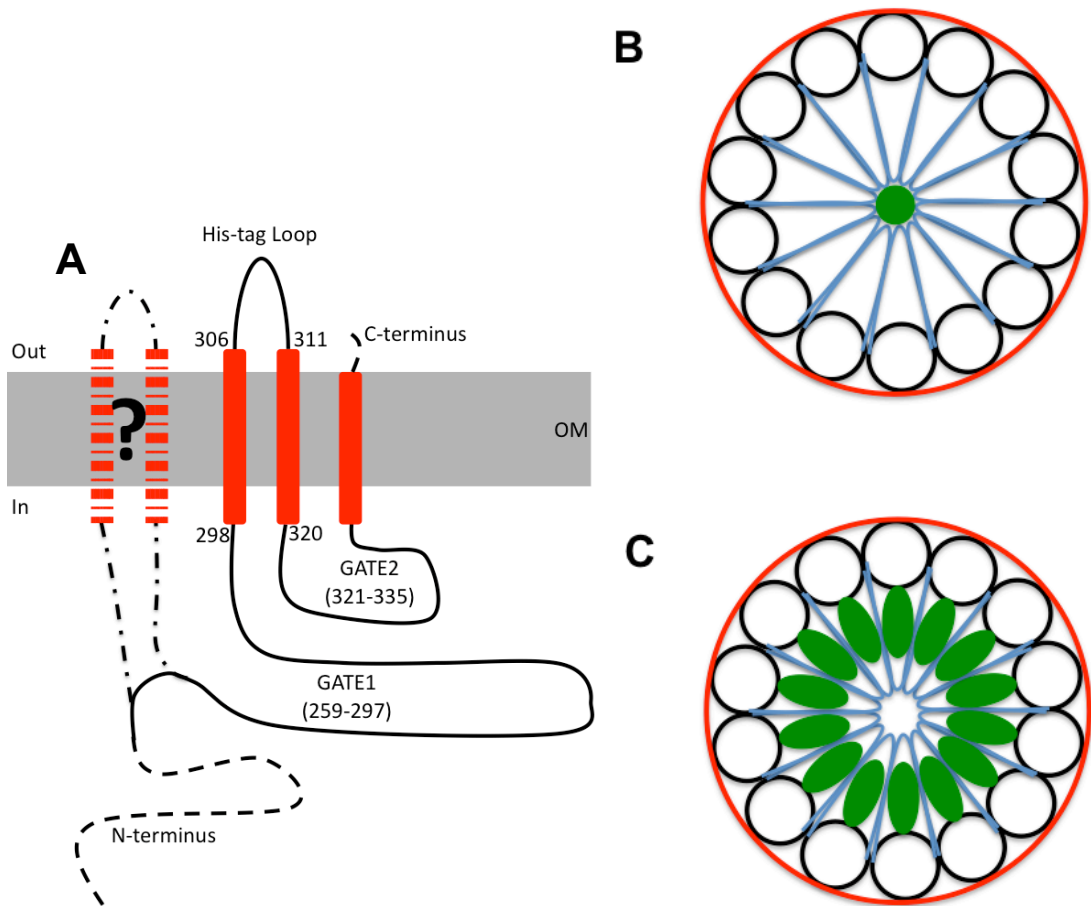
Figure 10: (A) Proposed topology model of pIV OM-spanning segments, surrounding the two GATE regions. Residue numbering is for the mature pIV. Dashed red and black lines with '?' indicate the presence of putative TM segments N-terminal of GATE1. Other black dashed lines indicate regions of unknown conformation at the C- and N-termini. (B, C) Top-down view of pIV models showing possible location and interactions of GATE1 (⋀) and GATE2 (●) loops. 14 GATE2 loops come together to form a plug in the centre of the septum (B), or overlay GATE1 loops (C) to fill gaps, prevent OM leakage and provide a structural scaffold for the larger GATE1 loop. GATE2 may also serve as a molecular latch/lever that keeps GATE1 closed.

The most severely leaky mutant in the collection, E292K, causes the greatest sensitivity to Van, Bac, DOC, and highest *pspA* induction. Furthermore, given its ability to assemble and export f1 phage effectively, it can be assumed the pIV^{E292K} multimer is stable. Thus, current data indicates this residue is important for gate integrity with a minor role in the gate mechanism. Because of the length of GATE1 (39 residues), and the break in homology, 14 radially arrayed GATE1 regions could extend towards the centre of the pIV channel to form the body of the gate structure (Figure 10). It is hard to predict whether GATE2 is also part of the gate-structure. However, Cryo-EM images of GATE2 deletion mutants, in which 10 of the 14 GATE2 residues are deleted (ΔGATE2), revealed a slight change in the septum of the mutant compared with the wild-type multimer (Spagnuolo et al., 2010) Supplemental; Appendix 6). The inability of ΔGATE2 pIV mutant to export phage, sensitivity to Van, Bac and DOC, in addition to the gap in the septum in low-resolution structure suggests that the GATE2 either contributes to the structural stability of the septum or that it serves as a latch/lever that controls the opening and closing of GATE1. If the latter were correct, the gate in the ΔGATE2 channel would be jammed in an incompletely closed state. Replacing the wild-type 10 GATE2 residues with 10 glycines (ΔGATE2→G) reduced sensitivity to Van, Bac, and DOC (compared to ΔGATE2) but did not restore the ability to assemble/export phage. These observations confirm that GATE2 has a major role in the gate opening/closing mechanism, and a minor role in forming part of the gate structure (Spagnuolo et al, 2010; Appendix 5).

The similarity of amino acid replacements created in random mutagenesis does not correlate well with severity of leakiness. The high frequency of alanine to valine, threonine or isoleucine substitutions, or a combination thereof, is interesting since the structural difference between these four amino acids is very small, yet the phenotypic - and by proxy the structural - consequences are immense. Some mutations that are expected to affect charge or folding (e.g. R293C and S281P, respectively) only have minor effects on permeability, allowing growth on maltopentaose but not sensitizing cells to Van, Bac or DOC. In other cases, relatively minor changes (e.g. V270I, and S324G) dramatically increase substrate diffusion through the pIV channel (Table 5). This indicates that interactions between gate-structural and gate-mechanism elements are complex. High-resolution structures of the secretin family domain would provide insight into the inner workings of the gate mechanism and the interactions between multimer subunits required to form the gate.

4.2 Induction of the psp regulon

Many secretins utilise cognate OM-targeting proteins called pilotins to prevent insertion into the IM and allow efficient transport to the OM (Crago & Koronakis, 1998, Guilvout et al., 2006). In the absence of pilotins, some secretins (like PulD) become mistargeted to the IM, with a very minor fraction making it to the OM (Daefler & Russel, 1998). pIV does not have a cognate pilotin; nevertheless, wild-type pIV inserts into the OM equally well as to the IM (Russel & Kazmierczak, 1993).

Insertion of secretins into the IM induces the *psp* response, up regulating effector proteins that protect bacteria from environmental and physiological stressors (Guilvout et al., 2006, Jovanovic et al., 2006, Seo et al., 2007). The nature of the inducing signal is unclear, but there is evidence that dissipation of the PMF, due to proton leakage through the secretin channel inserted into the IM, is the trigger (Weiner & Model, 1994, Kleerebezem et al., 1996).

Given the role of the *psp* response in maintaining homeostasis in the IM and periplasm, analysing the induction of the *psp* regulon by leaky pIV mutants was undertaken to assess the physiological impact of leaky secretin channels on the host cell. Monitoring accumulation of the PspA protein in the cells 1 hour after the induction of pIV, using western blotting, assessed the induction of the psp regulon for all screen and site-directed mutants. Although 1 hour should be ample time for accumulation of a protein in *E. coli* after an inducing signal, in our experiments it appeared that the accumulation of PspA in cells producing pIV relative to the cells lacking pIV (the vector control) was not very high. Upon careful inspection of publications in which the *psp* regulon was induced by pIV or other secretins, it became apparent that the induction times were most commonly 2-3h, the time at which the culture has reached the stationary phase. Nevertheless, our data were informative in that some of the mutants from our screen resulted in reproducibly higher accumulation of PspA relative to that observed in the wild-type pIV-producing cultures.

If dissipation of the PMF is the *psp*-inducing signal, then strength of *psp* induction should correlate with the leakiness of the mutant. That is to say, severely leaky mutants should induce the *psp* response to a greater extent (and more quickly) than mildly leaky mutants. Assuming that leakiness correlates with sensitivity to Van, Bac, and DOC; mutants N296D, E292K, N295S, S324G and T334A should show similar accumulation of PspA in western blots (Figure 7). However, this is not the case; of these severely leaky mutants only E292K and N295S show elevated PspA compared to wild-type pIV. Conversely, R293C is relatively resistant to Van, Bac, and DOC, thus is not thought to be severely leaky, however, it displays high accumulation of PspA. Since these mutants complement assembly and export of Δ*gIV* R484 phage, *psp* induction by slow folding or misfolded pIV subunits/multimers is not a likely signal. The discrepancies seen in our experiments are either relevant, or are due to the early time point in the induction relative to other publications, capturing the cultures at the point when the PspA is at the beginning of accumulation phase, leading to different behaviour within the "noise" of experiment.

pIV-induced PspA protein and mRNA accumulation was investigated in a group of site-specific mutants (I312G, I316G, ΔGATE2, ΔGATE2→G). In these experiments, the accumulation of PspA protein was much less dramatic than that of mRNA, pointing to regulation of Psp protein production at the post-transcriptional level. Current understanding of *psp* regulation explains the transcriptional regulation through the signal threshold described by Gueguen et al. (2009) (Figure 4), followed by the recruitment of PspF and IHF to the *psp*

UAS (Model *et al.*, 1997, Darwin, 2005). The combined results of our secretin-stress *psp* induction experiments hint at a second level of regulation, at the post-transcriptional stage of gene expression. Therefore, translation or turnover of the *psp* mRNA transcript are likely controlled by an, as yet, undiscovered mechanism of inhibitory RNAs. Multiple levels of regulation are common to biological systems, to provide specificity and regulation of response pathways.

4.3   Leaky Mutations in the N-Terminal Domain of pIV

Four of the leaky mutants isolated lie outside of the GATE1 and GATE2 regions. Three are located within the N-terminal portion of pIV (A121V, D123Y, G147V), and one in the very beginning of the secretin family domain (I183V). Sequence alignments show the N-terminal portion of pIV is much smaller than other secretins (Figure 3). Structural models of the pIV N-terminal region threaded from high-resolution structures of the N-terminal fragments of secretins GspD and EscC (Figure 9) indicate that pIV has only two N-terminal subdomains, the first corresponding to the most N-terminal (N0) subdomain of GspD which has structural homology to the periplasmic signalling domain of OM receptor FpvA. The second subdomain (N3) corresponds to the N-terminal KH fold subdomain of GspD and EscC (Arnold et al., 2006, Korotkov et al., 2009, Spreter et al., 2009). KH (hnRNP K homology) domains are usually found in nucleic acid binding proteins; however, the positively charged residues required to interact with nucleic acids are not conserved in pIV, GspD, or EscC. It is not unusual for proteins to share fold-structures but not the functions of those folds, as is seen in this case; this phenomenon is explained in (Goldstein, 2008).

Cryo-EM of the protease-resistant portion of PulD showed that it contained the C- and M-rings and a partial N-ring (Chami et al., 2005). Further analysis showed that it included the full N3 KH-fold subdomain, adjacent to the secretin homology domain (Figure 3). The three N-terminal leaky mutants obtained in this thesis are all located within the N3 subdomain corresponding to the protease resistant portions of PulD and pIV (Brissette & Russel, 1990, Chami et al., 2005). Structural modelling (Figures 6, 9, and 10) placed the leaky pIV mutations in the loop between the β6-strand and α3-helix (A121V); at the N-terminus of the α3-helix (D123Y); and in the loop between β7- and β8-strands (G147V); all three residues are predicted to cluster on the same face of the N3 subdomain (Figure 9). Although only the monomer structure of the N0/N1/N2 fragment of GspD was solved, fitting it into a multimer scaffold (Korotkov et al., 2009) places the loops and the mutations in a cluster along the lumen of the periplasmic portion of the channel.

It is known that interaction of pIV with the pI/pXI IM phage assembly complex involves the N0 subdomain (Russel & Kazmierczak, 1993, Daefler et al., 1997b). Therefore, if modelling is accurate, the residues of the three leaky mutants located in the N3 subdomain could be in position to interact with the leading tip of assembling phage, or a tip-N0 complex, to signal GATE opening.

5.0 Conclusion

This work has, for the first time, located the gate regions in the primary sequence of a secretin. Given the conservation of the secretin family domain, and their highly homologous structures in cryo-EM, it is likely that gate regions of other secretins will be similar to those of pIV. The map of pIV gate regions provides a framework for future structural, biochemical, and functional studies of secretin structure/function. In addition, this work has uncovered some unusual and unexpected questions about the regulation of the phage shock protein response induced by secretin-stress.

6.0 Future Directions

6.1 Secretin Mutagenesis

Random mutagenesis of *gIV* using mutator strain XL1-Red revealed the location of two gating regions within the primary sequence of pIV. Given the structural and sequence homology amongst members of the secretin family it is prudent to now complete similar random mutagenesis experiments with a range of other secretins. This will ascertain whether the location of the gating loops identified in pIV is conserved within the secretin family. In addition, the original pIV mutant library could be rescreened to isolate other low abundance and severely leaky mutants providing a greater diversity of different mutations to the gate structure, possibly revealing more information about the mechanism of the gate and the roles of GATE1/2.

A method for randomly deleting single amino acids has been developed (Jones, 2005). This technique allows the scarless removal of single amino acids, potentially creating physical holes in the secretin gate structure by shortening the gate loop. Using a modified Mu transposon, nucleotides can be randomly deleted in triplets. Because the system does not recognise reading frames, two types of mutations are produced; single amino acid deletions, and a deletion plus an adjacent substitution. The occurrence of substitutions could help stabilise alterations to loops if the accompanying deletion occurs near the edge of the loop or a TM segment (i.e. near a kink in the protein backbone).

6.2   Secretin Molecular and Structural Biology

To confirm structural and topology models of pIV (Figure 6 and 10), it is necessary to create deletion mutants of the GATE1 region in pIV for analysis by cryo-EM. However, this is complicated by the fact that expression of leaky secretins *in vivo* often causes severe toxicity issues (e.g. ΔGATE2; Spagnuolo *et al*, 2010; Appendix 5). To ameliorate toxicity and stress associated with expression leaky OMs, suppressor mutations arise within the expression host's stress-response pathways, allowing survival. This is problematic since suppressor mutations often go unnoticed if phenotypic changes are mild. Furthermore, tracking and characterising suppressor mutations is difficult because they can occur in many different pathways and have subtle effects that may not be characterisable or even noticeable. To avoid such issues many protein scientists turn to *in vitro* synthesis systems, which use cell lysates to provide the machinery to synthesise proteins from DNA templates. This method

has been used with some success to produce and purify high quantities of secretin PulD (Guilvout *et al.*, 2008). It is planned to use a similar *in vitro* protein synthesis system to produce a ΔGATE1 mutant of pIV for cryo-EM analysis.

To confirm the location and surface exposure of predicted loops and GATE regions a combination of cysteine insertion, site-specific labelling and cryo-EM can be used. Coupled with cryo-EM, potential surface-exposed loops can be probed by gold-maleimide labelling; similarly, gold-nickel binding can confirm the surface-exposed location of the pIV His-tag. If successful, cysteine-scanning experiments can then be undertaken using site-directed PCR to confirm the surface exposed nature of other predicted loops. Alternatively, a variation of Jones' (2005) MuDel transposon could be constructed to randomly insert His-tags (Hoeller *et al.*, 2008) and used to identify surface-exposed loops, in conjunction with gold-nickel binding and cryo-EM. Such experiments will further elucidate the OM topology of pIV, confirming or refuting the presence of TM segments N-terminal of GATE1. Two cysteine mutations located within GATE1 of pIV were identified in this thesis: R276C and R293C. The location of these cysteines can also be probed using maleimide-gold labelling and cryo-EM, providing evidence backing our model that GATE1 forms the body of the pIV septum. Similar site-specific labelling experiments have already been conducted by Opalka *et al* (2003) to confirm the number of pIV monomers that form the multimeric channel.

6.3   *Secretin-induced Stress Responses*

Given the importance of stress responses in maintaining cellular homeostasis during secretin-stress, it would be prudent to repeat western blotting experiments with a more quantitative slot-blot method which places all samples on the same blot. This will avoid the problem of comparing amounts of protein on different blots exposed for different amounts of time. In addition, qRT-PCR and Psp-reporter fusion experiments to measure induction by the entire set of pIV mutants could provide further clues about the nature of the putative post-transcriptional regulation of the *pspA* regulon, the existence of which was indicated by the comparison of the preliminary *pspA* qRT-PCR and PspA western blot analyses. To identify regulatory RNAs that may have a role in post-transcriptional regulation of the *pspA* regulon, DNA probes will be designed complementary to the various *psp* operon transcripts as described by Brissette et al (1991). Alternatively, BiaCore Chip assays to ascertain whether regulatory proteins bind *psp* transcripts could be performed. Next-generation sequencing (Solexa) of the small-RNA fraction of the total cell RNA in the presence or absence of secretin expression would be necessary to search for putative regulatory RNAs involved in the *psp* response.

When measuring PspA accumulation in response to secretin-stress, other research groups have used extended secretin expression times, ranging from 2 to 24 hours (Seo et al., 2007, Gueguen et al., 2009, Seo et al., 2009). This practice stems from the observation that PspA protein does not begin to accumulate in significant amounts until stationary phase. Thus, in order to

properly investigate *psp* regulon induction by leaky pIV mutants cell lysates, mRNA must be collected in a time-course experiment. This will show how and when PspA protein accumulates relative to the production of the mRNA transcripts.

To compare the dissipation of the PMF in the presence of leaky mutants relative to wild-type, pIV-expressing cultures can be treated with a cationic dye, JC-1, that fluoresces differently depending on membrane potential (i.e. fluorescence is affected by PMF). Measuring the fluoresence of the culture, in conjunction with confocal microscopy, can be used to determine how leaky pIV mutants affect the PMF (Becker et al., 2005, Jovanovic et al., 2006).

7.0 References

Adams, H., W. Teertstra, J. Demmers, R. Boesten & J. Tommassen, (2003) Interactions between phage-shock proteins in Escherichia coli. *J. Bacteriol.* **185**: 1174-1180.

Adams, H., W. Teertstra, M. Koster & J. Tommassen, (2002) PspE (phage-shock protein E) of *Escherichia coli* is a rhodanese. *FEBS Lett.* **518**: 173-176.

Arnold, K., L. Bordoli, J. Kopp & T. Schwede, (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195-201.

Baneyx, F. & G. Georgiou, (1991) Construction and characterisation of *Escherichia coli* strains deficient in multiple secreted proteases -

Protease III degrades high-molecular-weight substrates in vivo. *J. Bacteriol.* **173**: 2696-2703.

Bayan, N., I. Guilvout & A. P. Pugsley, (2006) Secretins take shape. *Mol. Microbiol.* **60**: 1-4.

Becker, L. A., I. S. Bang, M. L. Crouch & F. C. Fang, (2005) Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant Salmonella enterica serovar Typhimurium. *Mol. Microbiol.* **56**: 1004-1016.

Beis, K., R. F. Collins, R. C. Ford, A. B. Kamis, C. Whitfield & J. H. Naismith, (2004) Three-dimensional structure of Wza, the protein required for translocation of group 1 capsular polysaccharide across the outer membrane of Escherichia coli. *J. Biol. Chem.* **279**: 28227-28232.

Benson, S. A. & A. Decloux, (1985) Isolation and characterisation of outer-membrane permeability mutants in *Escherichia coli* K-12. *J. Bacteriol.* **161**: 361-367.

Benson, S. A., J. L. L. Occi & B. A. Sampson, (1988) Mutations that alter the pore function of the OmpF porin of *Escherichia coli* K-12. *J. Mol. Biol.* **203**: 961-970.

Bergler, H., D. Abraham, H. Aschauer & F. Turnowsky, (1994) Inhibition of lipid biosynthesis induces the expression of the *pspA* gene. *Microbiology-Uk* **140**: 1937-1944.

Bitter, W., M. Koster, M. Latijnhouwers, H. de Cock & J. Tommassen, (1998) Formation of oligomeric rings by XcpQ and PilQ, which are involved in protein transport across the outer membrane of *Pseudomonas aeruginosa*. *Mol. Microbiol.* **27**: 209-219.

Blake, M. S., K. H. Johnston, G. J. Russelljones & E. C. Gotschlich, (1984) A rapid, sensitive method for detection of alkaline-phosphatase conjugated anti-antibody on western blots. *Anal. Biochem.* **136**: 175-179.

Bowie, J. U., (1997) Helix packing in membrane proteins. *J. Mol. Biol.* **272**: 780-789.

Brissette, J. L. & M. Russel, (1990) Secretion of membrane integration of a filamentous phage-encoded morphogenetic protein. *J. Mol. Biol.* **211**: 565-580.

Brissette, J. L., M. Russel, L. Weiner & P. Model, (1990) Phage shock protein, a stress protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 862-866.

Brissette, J. L., L. Weiner, T. L. Ripmaster & P. Model, (1991) Characterization and sequence of the *Escherichia coli* stress-induced *psp* operon. *J. Mol. Biol.* **220**: 35-48.

Bryson, K., L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi & D. T. Jones, (2005) Protein structure prediction servers at university college london. *Nucleic Acids Res.* **33**: W36-W38.

Burghout, P., R. van Boxtel, P. Van Gelder, P. Ringler, S. A. Muller, J. Tommassen & M. Koster, (2004) Structure and electrophysiological properties of the YscC secretin from the type III secretion system of *Yersinia enterocolitica*. *J. Bacteriol.* **186**: 4645-4654.

Bury-Mone, S., Y. Nomane, N. Reymond, R. Barbet, E. Jacquet, S. Imbeaud, A. Jacq & P. Bouloc, (2009) Global Analysis of Extracytoplasmic Stress Signaling in *Escherichia coli*. *PLoS Genet.* **5**: 17.

Carbonnelle, E., S. Helaine, L. Prouvensier, X. Nassif & V. Pelicic, (2005) Type IV pilus biogenesis in *Neisseria meningitidis*: PilW is involved in a step occurring after pilus assembly, essential for fibre stability and function. *Mol. Microbiol.* **55**: 54-64.

Chami, M., I. Guilvout, M. Gregorini, H. W. Remigy, S. A. Muller, M. Valerio, A. Engel, A. P. Pugsley & N. Bayan, (2005) Structural insights into the secretin PulD and its trypsin-resistant core. *J Biol Chem* **280**: 37732-37741.

Chandran, V., R. Fronzes, S. Duquerroy, N. Cronin, J. Navaza & G. Waksman, (2009) Structure of the outer membrane complex of a type IV secretion system. *Nature* **462**: 1011-U1066.

Cole, C., J. D. Barber & G. J. Barton, (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**: W197-W201.

Collin, S., I. Guilvout, M. Chami & A. P. Pugsley, (2007) YaeT-independent multimerization and outer membrane association of secretin PulD. *Mol. Microbiol.* **64**: 1350-1357.

Crago, A. M. & V. Koronakis, (1998) Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol Microbiol* **30**: 47-56.

Craig, L. & J. Li, (2008) Type IV pilli: paradoxes in form and function. *Curr. Opin. Struct. Biol.* **18**: 267-277.

Daefler, S., I. Guilvout, K. R. Hardie, A. P. Pugsley & M. Russel, (1997a) The C-terminal domain of the secretin PulD contains the binding site for its cognate chaperone, PulS, and confers PulS dependence on pIV(f1) function. *Mol. Microbiol.* **24**: 465-475.

Daefler, S. & M. Russel, (1998) The Salmonella typhimurium InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Mol Microbiol* **28**: 1367-1380.

Daefler, S., M. Russel & P. Model, (1997b) Module swaps between related translocator proteins pIV(f1), pIV(IKe) and PulD: Identification of a specificity domain. *J. Mol. Biol.* **266**: 978-992.

Darwin, A. J., (2005) The phage-shock-protein response. *Mol. Microbiol.* **57**: 621-628.

Davis, B. M., E. H. Lawson, M. Sandkvist, A. Ali, S. Sozhamannan & M. K. Waldor, (2000) Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi. *Science* **288**: 333-335.

Dong, C. J., K. Beis, J. Nesper, A. L. Brunkan-LaMontagne, B. R. Clarke, C. Whitfield & J. H. Naismith, (2006) Wza the translocon for E-coli capsular polysaccharides defines a new class of membrane protein. *Nature* **444**: 226-229.

Duguay, A. R. & T. J. Silhavy, (2004) Quality control in the bacterial periplasm. *Biochimica Et Biophysica Acta-Molecular Cell Research* **1694**: 121-134.

Endemann, H. & P. Model, (1995) Location of Filamentous Phage Minor Coat Proteins in Phage and in Infected-Cells. *J. Mol. Biol.* **250**: 496-506.

Feng, J. N., P. Model & M. Russel, (1999) A trans-envelope protein complex needed for filamentous phage assembly and export. *Mol. Microbiol.* **34**: 745-755.

Galan, J. E. & A. Collmer, (1999) Type III secretion machines: Bacterial devices for protein delivery into host cells. *Science* **284**: 1322-1328.

Gerlach, R. G. & M. Hensel, (2007) Protein secretion systems and adhesins: The molecular armory of Gram-negative pathogens. *Int. J. Med. Microbiol.* **297**: 401-415.

Goldstein, R. A., (2008) The structure of protein evolution and the evolution of protein structure. *Curr. Opin. Struct. Biol.* **18**: 170-177.

Green, R. C. & A. J. Darwin, (2004) PspG, a new member of the Yersinia enterocolitica phage shock protein regulon. *J. Bacteriol.* **186**: 4910-4920.

Greener, A. & M. Callahan, (1994) XL1-Red: a highly efficient random mutagenesis strain. In: Stratagene Newsletter: Strategies in Molecular Biology. pp. 32-34.

Greener, A., M. Callahan & B. Jerspeth, (1997) An efficient random mutagenesis technique using an E.coli mutator strain. *Molecular Biotechnology* **7**: 189-195.

Gromiha, M. M., S. Ahmad & M. Suwa, (2005) TMBETA-NET: discrimination and prediction of membrane spanning beta-strands in outer membrane proteins. *Nucleic Acids Res.* **33**: W164-W167.

Gueguen, E., D. C. Savitzky & A. J. Darwin, (2009) Analysis of the Yersinia enterocolitica PspBC proteins defines functional domains, essential amino acids and new roles within the phage-shock-protein response. *Mol. Microbiol.* **74**: 619-633.

Guilvout, I., M. Chami, C. Berrier, A. Ghazi, A. Engel, A. P. Pugsley & N. Bayan, (2008) In vitro multimerization and membrane insertion of bacterial outer membrane secretin PulD. *J Mol Biol* **382**: 13-23.

Guilvout, I., M. Chami, A. Engel, A. P. Pugsley & N. Bayan, (2006) Bacterial outer membrane secretin PulD assembles and inserts into the inner membrane in the absence of its pilotin. *Embo J.* **25**: 5241-5249.

Guilvout, I., K. R. Hardie, N. Sauvonnet & A. P. Pugsley, (1999) Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? *J Bacteriol* **181**: 7212-7220.

Haigh, N. G. & R. E. Webster, (1998) The major coat protein of filamentous bacteriophage f1 specifically pairs in the bacterial cytoplasmic membrane. *J. Mol. Biol.* **279**: 19-29.

Hoeller, B. M., B. Reiter, S. Abad, I. Graze & A. Glieder, (2008) Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). *J. Microbiol. Methods* **75**: 251-257.

Hofmann, K. & W. Stoffel, (1993) TMBase - A database of membrane spanning protein segments. *Biological Chemistry Hoppe-Seyler* **374**.

Hofnung, M., (1995) An Intelligent Channel (and More). *Science* **267**: 473-474.

Hueck, C. J., (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**: 379-+.

Jones, D. D., (2005) Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 beta-lactamase to an amino acid deletion. *Nucleic Acids Res.* **33**: 8.

Jones, D. T., (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**: 195-202.

Jovanovic, G., J. Dworkin & P. Model, (1997) Autogenous control of PspF, a constitutively active enhancer-binding protein of Escherichia coli. *J. Bacteriol.* **179**: 5232-5237.

Jovanovic, G., C. Engl & M. Buck, (2009) Physical, functional and conditional interactions between ArcAB and phage shock proteins upon secretin-induced stress in Escherichia coli. *Mol. Microbiol.* **74**: 16-28.

Jovanovic, G., L. J. Lloyd, M. P. H. Stumpf, A. J. Mayhew & M. Buck, (2006) Induction and function of the phage shock protein extracytoplasmic stress response in Escherichia coli. *J. Biol. Chem.* **281**: 21147-21161.

Jovanovic, G., L. Weiner & P. Model, (1996) Identification, nucleotide sequence, and characterization of PspF, the transcriptional activator of the Escherichia coli stress-induced psp operon. *J. Bacteriol.* **178**: 1936-1945.

Kazmierczak, B. I., D. L. Mielke, M. Russel & P. Model, (1994) pIV, a Filamentous Phage Protein That Mediates Phage Export across the Bacterial-Cell Envelope, Forms a Multimer. *J. Mol. Biol.* **238**: 187-198.

Klebba, P. E., M. Hofnung & A. Charbit, (1994) A model of maltodextrin transport through the sugar-specific porin, LamB, based on deletion analysis. *Embo J.* **13**: 4670-4675.

Kleerebezem, M., W. Crielaard & J. Tommassen, (1996) Involvement of stress protein PspA (phage shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress conditions. *Embo J.* **15**: 162-171.

Kleerebezem, M. & J. Tommassen, (1993) Expression of the *pspA* gene stimulates efficient protein export in *Escherichia coli*. *Mol. Microbiol.* **7**: 947-956.

Koo, J., S. Tammam, S. Y. Ku, L. M. Sampaleanu, L. L. Burrows & P. L. Howell, (2008) PilF Is an Outer Membrane Lipoprotein Required for Multimerization and Localization of the Pseudomonas aeruginosa Type IV Pilus Secretin. *J. Bacteriol.* **190**: 6961-6969.

Korotkov, K. V., E. Pardon, J. Steyaert & W. G. J. Hol, (2009) Crystal Structure of the N-Terminal Domain of the Secretin GspD from ETEC Determined with the Assistance of a Nanobody. *Structure* **17**: 255-265.

Laemmli, U. K., (1970) Cleavage of structural proteins during assembly of head of bacteriophage T4. *Nature* **227**: 680-&.

Lessl, M., D. Balzer, R. Lurz, V. L. Waters, D. G. Guiney & E. Lanka, (1992) Dissection of IncP conjugative plasmid transfer - definition of the transfer region Tra2 by mobilization of the Tra1 region in trans. *J. Bacteriol.* **174**: 2493-2500.

Linderoth, N. A., P. Model & M. Russel, (1996) Essential role of a sodium dodecyl sulfate-resistant protein IV multimer in assembly-export of filamentous phage. *J Bacteriol* **178**: 1962-1970.

Linderoth, N. A., M. N. Simon & M. Russel, (1997) The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science* **278**: 1635-1638.

Liu, J., J. M. Rutz, J. B. Feix & P. E. Klebba, (1993) Permeability properties of a large gated channel within the ferric enterobactin receptor, FepA. *Proc Natl Acad Sci U S A* **90**: 10653-10657.

Lloyd, L. J., S. E. Jones, G. Jovanovic, P. Gyaneshwar, M. D. Rolfe, A. Thompson, J. C. Hinton & M. Buck, (2004) Identification of a new

member of the phage shock protein response in Escherichia coli, the phage shock protein g (PspG). *J. Biol. Chem.* **279**: 55707-55714.

Marciano, D. K., M. Russel & S. M. Simon, (1999) An Aqueous Channel for Filamentous Phage Export. *Science* **284**: 1516-1519.

Marciano, D. K., M. Russel & S. M. Simon, (2001) Assembling filamentous phage occlude pIV channels. *Proc Natl Acad Sci U S A* **98**: 9359-9364.

Marlovits, T. C., T. Kubori, A. Sukhan, D. R. Thomas, J. E. Galan & V. M. Unger, (2004) Structural insights into the assembly of the type III secretion needle complex. *Science* **306**: 1040-1042.

Miki, T., N. Okada, Y. Shimada & H. Danbara, (2004) Characterization of Salmonella pathogenicity island 1 type III secretion-dependent hemolytic activity in Salmonella enterica serovar Typhimurium. *Microb. Pathog.* **37**: 65-72.

Miller, J. H., (1972) *Experiments in Molecular Genetics,* p. 468. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Misra, R. & S. A. Benson, (1988a) Genetic identification of the pore domain of the OmpC porin of *Escherichia coli* K-12. *J. Bacteriol.* **170**: 3611-3617.

Misra, R. & S. A. Benson, (1988b) Isolation and characterisation of OmpC porin mutants with altered pore properties. *J. Bacteriol.* **170**: 528-533.

Model, P., G. Jovanovic & J. Dworkin, (1997) The Escherichia coli phage-shock-protein (psp) operon. *Mol. Microbiol.* **24**: 255-261.

Neidhardt, F. C., P. L. Bloch & D. F. Smith, (1974) Culture medium for Enterobacteria. *J. Bacteriol.* **119**: 736-747.

Nouwen, N., N. Ranson, H. Saibil, B. Wolpensinger, A. Engel, A. Ghazi & A. P. Pugsley, (1999) Secretin PulD: association with pilot PulS, structure, and

ion-conducting channel formation. *Proc Natl Acad Sci U S A* **96**: 8173-8177.

Okon, M., T. F. Moraes, P. I. Lario, A. L. Creagh, C. A. Haynes, N. C. J. Strynadka & L. P. McLntosh, (2008) Structural Characterization of the Type-III Pilot-Secretin Complex from Shigella flexneri. *Structure* **16**: 1544-1554.

Opalka, N., R. Beckmann, N. Boisset, M. N. Simon, M. Russel & S. A. Darst, (2003) Structure of the filamentous phage pIV multimer by cryo-electron microscopy. *J Mol Biol* **325**: 461-470.

Raffa, R. G. & T. L. Raivio, (2002) A third envelope stress signal transduction pathway in Escherichia coli. *Mol. Microbiol.* **45**: 1599-1611.

Rakonjac, J., J. N. Feng & P. Model, (1999) Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J. Mol. Biol.* **289**: 1253-1265.

Rhodius, V. A., W. C. Suh, G. Nonaka, J. West & C. A. Gross, (2006) Conserved and variable functions of the sigma(E) stress response in related genomes. *Plos Biology* **4**: 43-59.

Robert, V., E. B. Volokhina, F. Senf, M. P. Bos, P. Van Gelder & J. Tommassen, (2006) Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol* **4**: e377.

Rowley, G., M. Spector, J. Kormanec & M. Roberts, (2006) Pushing the envelope: extracytoplasmic stress responses in bacterial pathogens. *Nat Rev Microbiol* **4**: 383-394.

Ruiz, N., D. Kahne & T. J. Silhavy, (2006) Advances in understanding bacterial outer-membrane biogenesis. *Nature Reviews Microbiology* **4**: 57-66.

Ruiz, N. & T. J. Silhavy, (2005) Sensing external stress: watchdogs of the Escherichia coli cell envelope. *Curr Opin Microbiol* **8**: 122-126.

Russel, M., (1994) Mutants at Conserved Positions in Gene-IV, a Gene Required for Assembly and Secretion of Filamentous Phages. *Mol. Microbiol.* **14**: 357-369.

Russel, M., (1998) Macromolecular assembly and secretion across the bacterial cell envelope: Type II protein secretion systems. *J. Mol. Biol.* **279**: 485-499.

Russel, M. & B. Kazmierczak, (1993) Analysis of the Structure and Subcellular Location of Filamentous Phage-pIV. *J. Bacteriol.* **175**: 3998-4007.

Russel, M., N. A. Linderoth & A. Sali, (1997) Filamentous phage assembly: Variation on a protein export theme. *Gene* **192**: 23-32.

Russel, M. & P. Model, (1986) The role of thioredoxin in Filamentous Phage assembly - construction, isolation, and characterisation of mutant thioredoxins. *J. Biol. Chem.* **261**: 4997-5005.

Sambrook, J., E. F. Fritsch & T. Maniatis, (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Seo, J., A. Brencic & A. J. Darwin, (2009) Analysis of Secretin-Induced Stress in Pseudomonas aeruginosa Suggests Prevention Rather than Response and Identifies a Novel Protein Involved in Secretin Function. *J. Bacteriol.* **191**: 898-908.

Seo, J., D. C. Savitzky, E. Ford & A. J. Darwin, (2007) Global analysis of tolerance to secretin-induced stress in Yersinia enterocolitica suggests that the phage-shock-protein system may be a remarkably self-contained stress response. *Mol. Microbiol.* **65**: 714-727.

Sikora, A. E., S. R. Lybarger & M. Sandkvist, (2007) Compromised outer membrane integrity in Vibrio cholerae type II secretion mutants. *J. Bacteriol.* **189**: 8484-8495.

Spagnuolo, J., N. Opalka, W. X. Wen, D. Gagic, E. Chabaud, D. Bellini, M. Bennett, G. E. Norris, S. A. Darst, M. Russel & J. Rakonjac, (2010) Identification of the gate regions in the primary structure of the secretin pIV. *Mol. Microbiol.* **76**: 17.

Spreter, T., C. K. Yip, S. Sanowar, I. Andre, T. G. Kimbrough, M. Vuckovic, R. A. Pfuetzner, W. Y. Deng, A. C. Yu, B. B. Finlay, D. Baker, S. I. Miller & N. C. J. Strynadka, (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat. Struct. Mol. Biol.* **16**: 468-476.

Tam, C. & D. Missiakas, (2005) Changes in lipopolysaccharide structure induce the sigma(E)-dependent response of Escherichia coli. *Mol Microbiol* **55**: 1403-1412.

Viarre, V., E. Cascales, G. Ball, G. P. F. Michel, A. Filloux & R. Voulhoux, (2009) HxcQ Liposecretin Is Self-piloted to the Outer Membrane by Its N-terminal Lipid Anchor. *J. Biol. Chem.* **284**: 33815-33823.

Volokhina, E. B., F. Beckers, J. Tommassen & M. P. Bos, (2009) The beta-Barrel Outer Membrane Protein Assembly Complex of Neisseria meningitidis. *J. Bacteriol.* **191**: 7074-7085.

Voulhoux, R., M. P. Bos, J. Geurtsen, M. Mols & J. Tommassen, (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* **299**: 262-265.

Wandersman, C., M. Schwartz & T. Ferenci, (1979) *Escherichia coli* mutants impaired in maltodextrin transport. *J. Bacteriol.* **140**: 1-13.

Weiner, L. & P. Model, (1994) Role of an *Escherichia coli* stress-response operon in stationary-phase survival. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 2191-2195.

Yanisch-Perron, C., J. Vieira & J. Messing, (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

8.0   Appendix 1: TMPred Output

min=14 | max=41 | ascii | pIV+mature+wild-type+ | plain_text |
QVIEMNNSPLRDFVTWYSKQTGESVIVSPDVKGTVTVYSSDVKPENLRNFFISVLRANNFDMVGSNPSII
QKYNPNNQDYIDELPSSDNQEYDDNSAPSGGFFVPQNDNVTQTFKINNVRAKDLIRVVELFVKSNTSKSS
NVLSIDGSNLLVVSAPKDILDNLPQFLSTVDLPTDQILIEGLIFEVQQGDALDFSFAAGSQRGTVAGGVN
TDRLTSVLSSAGGSFGIFNGDVLGLSVRALKTNSHSKILSVPRILTLSGQKGSISVGQNVPFITGRVTGE
SANVNNPFQTIERQNVGISMSVFPVAMAGGNIVLDITSKADSLSSSTQASDVITNQRSIATTVNLRDGQT
LLLGGLTDYKNTSQDSGVPFLSKIPLIGLLFSSRSDSNEESTLYVLVKATIVRAL

   TMpred output for pIV mature wild-type

[ISREC-Server] Date: Thu Mar 25 1:04:04 2010

--
tmpred -par=matrix.tab -min=14 -max=41 -def
-in=wwwtmp/.TMPRED.15184.2039.seq -out=wwwtmp/.TMPRED.15184.2039.out
-out2=wwwtmp/.TMPRED.15184.2039.out2 -
out3=wwwtmp/.TMPRED.15184.2039.txt
>wwwtmp/.TMPRED.15184.2039.err

TMpred prediction output for : wwwtmp/.TMPRED.15184.2039.seq

Sequence: QVI...RAL   length: 405
Prediction parameters: TM-helix length between 14 and 35

1.) Possible transmembrane helices
==================================
The sequence positions in brackets denominate the core region.
Only scores above 500 are considered significant.

Inside to outside helices :   2 found
      from   to   score center
 214 (220) 240 (235)   723   227
 298 (298) 314 (314)   1054   306

Outside to inside helices :   3 found
      from   to   score center
 214 (220) 240 (236)   1098   227
 295 (298) 314 (314)   1067   306
 367 (367) 381 (381)   306   374

2.) Table of correspondences
============================
Here is shown, which of the inside->outside helices correspond
to which of the outside->inside helices.
  Helices shown in brackets are considered insignificant.
  A "+" symbol indicates a preference of this orientation.
  A "++" symbol indicates a strong preference of this orientation.

          inside->outside | outside->inside
   214- 240 (27) 723 | 214- 240 (27) 1098 ++
   298- 314 (17) 1054 | 295- 314 (20) 1067
                           |(367- 381 (15) 306 ++)

84

```
3.) Suggested models for transmembrane topology
===============================================
These suggestions are purely speculative and should be used with
EXTREME CAUTION since they are based on the assumption that
all transmembrane helices have been found.
In most cases, the Correspondence Table shown above or the
prediction plot that is also created should be used for the
topology assignment of unknown proteins.

2 possible models considered, only significant TM-segments used

-----> STRONGLY prefered model: N-terminus outside
 2 strong transmembrane helices, total score : 2152
 # from   to length score orientation
 1  214  240 (27)    1098 o-i
 2  298  314 (17)    1054 i-o

------> alternative model
 2 strong transmembrane helices, total score : 1790
 # from   to length score orientation
 1  214  240 (27)     723 i-o
 2  295  314 (20)    1067 o-i
```
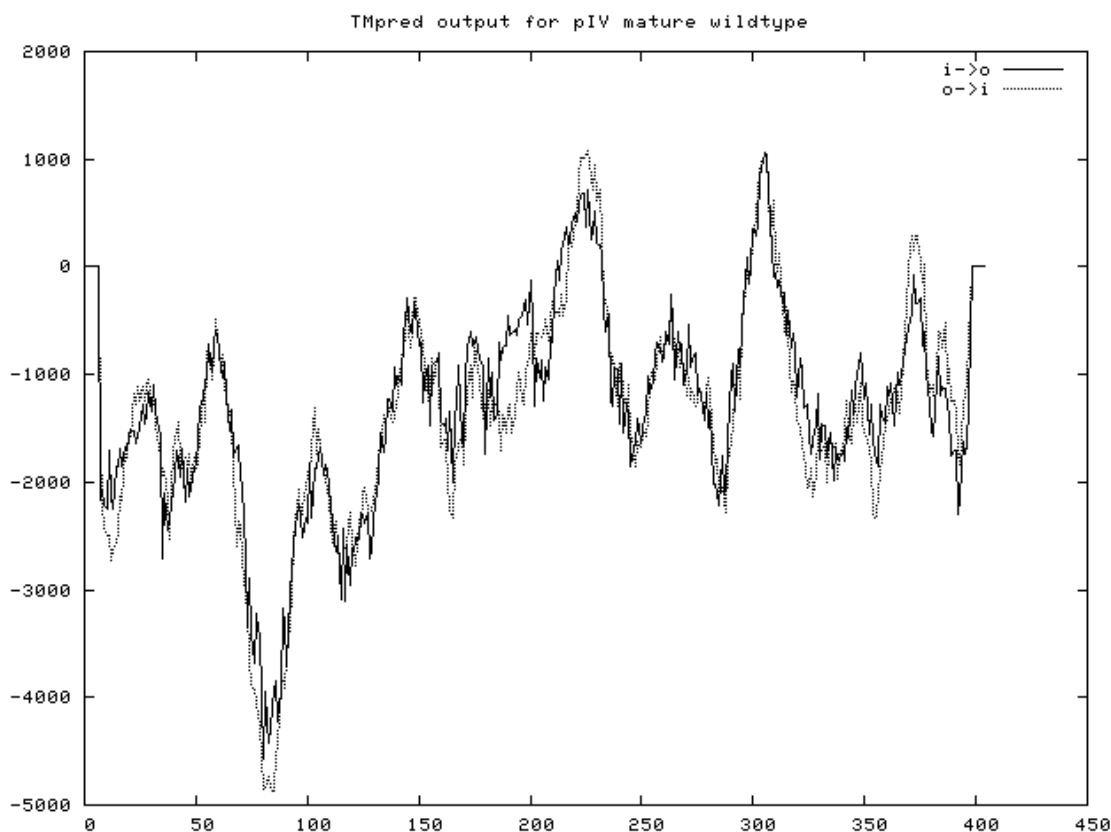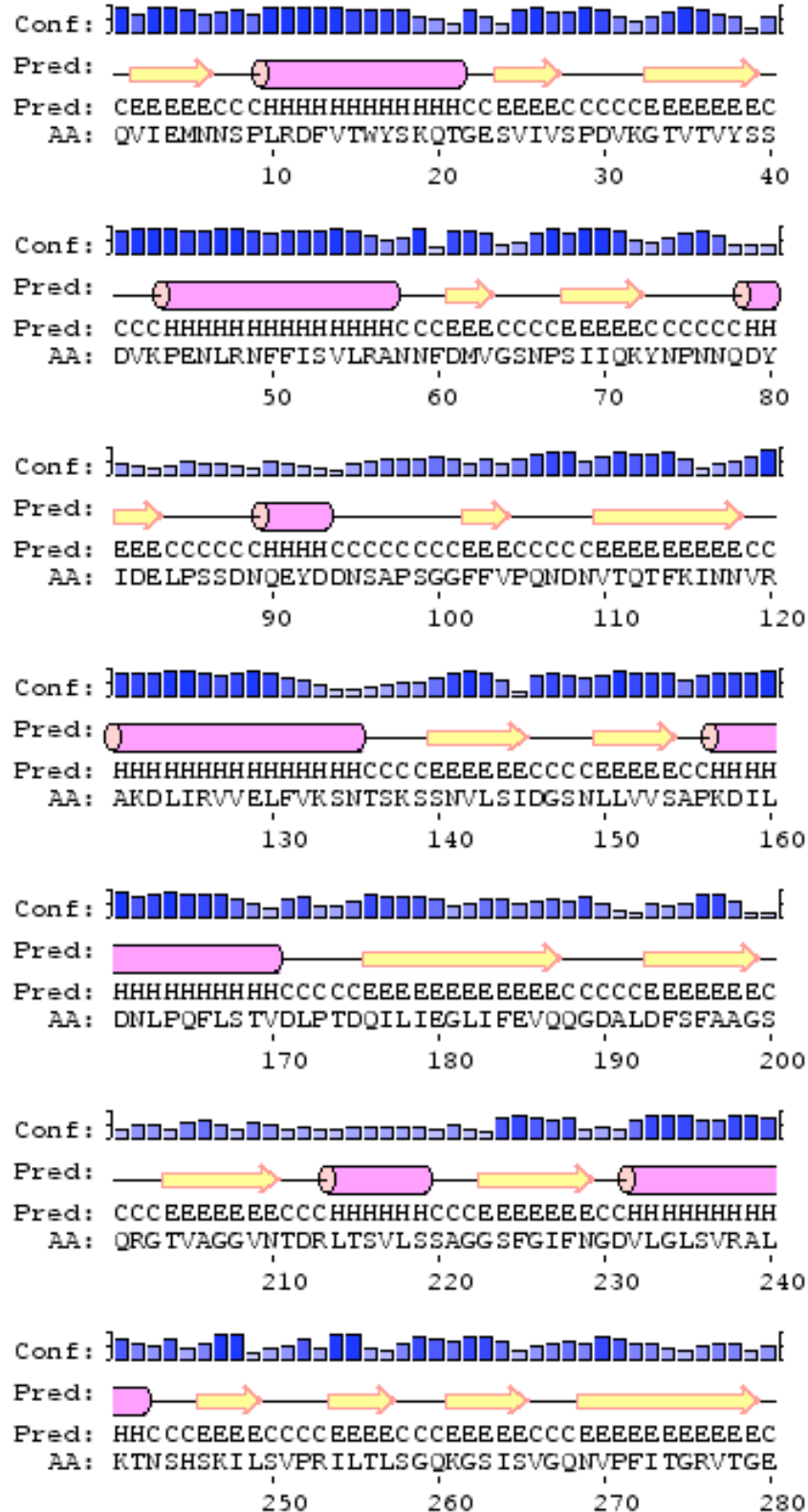

TMpred output for pIV mature wildtype

```
------------------------------------------------------------------------
--
/ Back to ISREC home page/ </index.html>
```

9.0 Appendix 2: JPred Output

```
QVIEMNNSPLRDFVTWYSKQTGESVIVSPDVKGTVTVYSSDVKPENLRNFFISVLRANNFDMVGSNPSII
-EEE-----HHHHHHHHHH---EEEE------EEEEE----HHHHHHHHHHHHH---EEEE---EEE

QKYNPNNQDYIDELPSSDNQEYDDNSAPSGGFFVPQNDNVTQTFKINNVRAKDLIRVVELFVKSNTSKSS
EE-----EEEH-----HHHHHHHH----------------EEEEE-----HHHHHHHHHHH-------E

NVLSIDGSNLLVVSAPKDILDNLPQFLSTVDLPTDQILIEGLIFEVQQGDALDFSFAAGSQRGTVAGGVN
EEEEE----EEEEE--HHHHHHHHHHHHHH-----EEEEEEEEEEEE--------EEEEEE---------

TDRLTSVLSSAGGSFGIFNGDVLGLSVRALKTNSHSKILSVPRILTLSGQKGSISVGQNVPFITGRVTGE
---------------EE-----HHHHHHHHHH---EEEEE--EEEEEE--EEEEEE--EEEEEEEEEE--

SANVNNPFQTIERQNVGISMSVFPVAMAGGNIVLDITSKADSLSSSTQASDVITNQRSIATTVNLRDGQT
-------EEEEEEEEEEEEEEEEEEEEE----EEEEEEEEEE------------EEEEEEEEEEEEEE----E

LLLGGLTDYKNTSQDSGVPFLSKIPLIGLLFSSRSDSNEESTLYVLVKATIVRAL
EEEEEEEEE-------EEEE-----EEEE----------EEEEEEEEEEEEE---
```
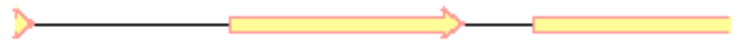
10.0 Appendix 3: PsiPred Output

```
Conf: [███████████████████████████████████████]
Pred: —▷—▷—◖██████◗—▷—▷—◁—
Pred: CEEEEECC CHHHHHHHHHHHHCC EEEEC CCCCEEEEEEEC
  AA: QVIEMNNSPLRDFVTWYSKQTGESVIVSPDVKGTVTVYSS
              10        20        30        40

Conf: [███████████████████████████████████████]
Pred: —◖██████████◗—▷—▷—◖█
Pred: CCCHHHHHHHHHHHHHHCCCEEECCCCCEEEECCCCCCHH
  AA: DVKPENLRNFFISVLRANNFDMVGSNPSIIQKYNPNNQDY
              50        60        70        80

Conf: [███████████████████████████████████████]
Pred: —▷—◖███◗—▷—▷—
Pred: EEECCCCCCHHHHCCCCCCCCEEECCCCCEEEEEEEEECC
  AA: IDELPSSDNQEYDDNSAPSGGFFVPQNDNVTQTFKINNVR
              90       100       110       120

Conf: [███████████████████████████████████████]
Pred: ◖████████████◗—▷—▷—◖█
Pred: HHHHHHHHHHHHHHHCCCCEEEEEECCCCCEEEEECCHHHH
  AA: AKDLIRVVELFVKSNTSKSSNVLSIDGSNLLVVSAPKDIL
             130       140       150       160

Conf: [███████████████████████████████████████]
Pred: ◖██████◗—▷—▷—
Pred: HHHHHHHHHHCCCCCEEEEEEEEEEEECCCCCEEEEEEEC
  AA: DNLPQFLSTVDLPTDQILIEGLIFEVQQGDALDFSFAAGS
             170       180       190       200

Conf: [███████████████████████████████████████]
Pred: —▷—◖███◗—▷—◖██████◗
Pred: CCCEEEEEEECCCHHHHHCCCCEEEEEEECCHHHHHHHHH
  AA: QRGTVAGGVNTDRLTSVLSSAGGSFGIFNGDVLGLSVRAL
             210       220       230       240

Conf: [███████████████████████████████████████]
Pred: ◖◗—▷—▷—▷—▷—
Pred: HHCCCEEEECCCCEEEECCCEEEEECCCEEEEEEEEEEEC
  AA: KTNSHSKILSVPRILTLSGQKGSISVGQNVPFITGRVTGE
             250       260       270       280
```

87

```
Conf: }▪▪▪▪▪▪▪▪▪▪▪▪▪▪▫▫▫▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪[

Pred: ──────────        ───────        ──────

Pred: CCCCCCCCEEEEEEECCEEEEEEEEEEECCCCEEEEEEEEEE
  AA: SANVNNPFQTIERQNVGISMSVFPVAMAGGNIVLDITSKA
                 290       300       310       320


Conf: }▫▪▫▪▪▪▪▫▫▫▫▫▫▪▪▪▪▪▪▪▪▪▪▫▫▪▪▪▪▪▪▫▫▫▪▪▪▪[

Pred: ▷────────        ──────────        ──────

Pred: ECCCCCCCCCCCCEEEEEEEEEEEEEECCCCCEEEEEEEEEEEE
  AA: DSLSSSTQASDVITNQRSIATTVNLRDGQTLLLGGLTDYK
                 330       340       350       360


Conf: }▫▫▫▫▪▫▫▫▫▪▫▫▪▫▪▫▫▫▪▪▪▫▫▪▫▫▫▫▫▫▫▫▫▪▪▪▫▫▫▫▫▫▪[

Pred: ▷───────        ───◖▬▬▷        ────────

Pred: ECCEEEEEEEECCCHHHHHHHEEEECCCCEEEEEEEEEEEEE
  AA: NTSQDSGVPFLSKIPLIGLLFSSRSDSNEESTLYVLVKAT
                 370       380       390       400


Conf: }▪▪▪▪▪[

Pred: ─▷──

Pred: EEECC
  AA: IVRAL
```
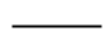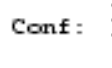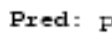
Legend:

◖▬▬▬ – helix   Conf: }▫▫▪▪▪[– confidence of prediction
                        -   +
▷ – strand   Pred: predicted secondary structure

───── – coil   AA: target sequence

11.0 Appendix 4: TMBETA Output

Thank you for using TMBETA.

Your prediction results are written below.

Amino acid composition, deviation from Globular proteins (GD)
and deviation from Outer Membrane Proteins (OD) in the sequence are as follows:

Residue	Composition(%)	GD(%)	OD(%)
A	4.44	4.03	4.51
C	0.00	1.39	0.47
D	6.67	0.70	0.76
E	2.96	3.36	1.82
F	4.69	0.78	1.01
G	7.41	0.41	1.13
H	0.25	2.01	1.00
I	6.67	0.96	1.90
K	3.95	1.81	0.98
L	9.38	0.90	0.60
M	0.99	1.22	0.57
N	8.15	3.61	2.41
P	4.20	0.43	0.46
Q	4.94	1.12	0.19
R	3.70	1.23	1.54
S	12.84	6.90	4.79
T	6.67	0.88	0.13
V	10.12	3.10	3.36
W	0.25	1.19	0.99
Y	1.73	1.85	2.40
Total	-	37.88	31.01

Amino acid composition of this sequence seems to be similar to outer-membrane proteins.
There are 18 predicted TM beta segments in your sequence.
Segment 1 : F13 to K19
Segment 2 : G33 to S40
Segment 3 : R48 to A57
Segment 4 : T111 to V119
Segment 5 : A121 to S134
Segment 6 : N149 to S154
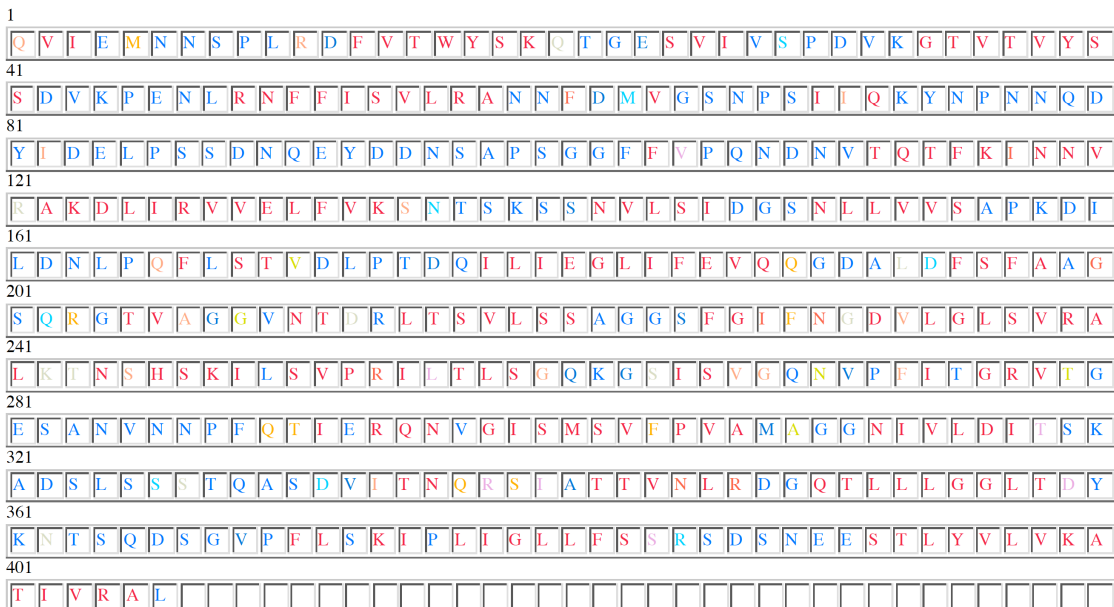Segment 7 : I177 to Q188
Segment 8 : L214 to S220
Segment 9 : D231 to L240
Segment 10 : N243 to I248

Segment 11 : S250 to G259
Segment 12 : G297 to A306
Segment 13 : N311 to T317
Segment 14 : I333 to I339
Segment 15 : T341 to R346
Segment 16 : Q349 to D358
Segment 17 : L376 to S383
Segment 18 : S391 to A404

Probability of residues to have beta conformation (Before refinements) are color coded as follows:

Probability
Probability 0.9 to 1.0
Probability 0.8 to 0.9
Probability 0.7 to 0.8
Probability 0.6 to 0.7
Probability 0.5 to 0.6
Probability 0.4 to 0.5
Probability 0.3 to 0.4
Probability 0.2 to 0.3
Probability 0.1 to 0.2
Probability 0.0 to 0.2

1
Q V I E M N N S P L R D F V T W Y S K Q T G E S V I V S P D V K G T V T V Y S

41
S D V K P E N L R N F F I S V L R A N N F D M V G S N P S I I Q K Y N P N N Q D

81
Y I D E L P S S D N Q E Y D D N S A P S G G F F V P Q N D N V T Q T F K I N N V

121
R A K D L I R V V E L F V K S N T S K S S N V L S I D G S N L L V V S A P K D I

161
L D N L P Q F L S T V D L P T D Q I L I E G L I F E V Q Q G D A L D F S F A A G

201
S Q R G T V A G G V N T D R L T S V L S S A G G S F G I F N G D V L G L S V R A

241
L K T N S H S K I L S V P R I L T L S G Q K G S I S V G Q N V P F I T G R V T G

281
E S A N V N N P F Q T I E R Q N V G I S M S V F P V A M A G G N I V L D I T S K

321
A D S L S S S T Q A S D V I T N Q R S I A T T V N L R D G Q T L L L G G L T D Y

361
K N T S Q D S G V P F L S K I P L I G L L F S S R S D S N E E S T L Y V L V K A

401
T I V R A L

13.0 Appendix 6

Supplementary Material

Identification of the gate regions in the primary structure of the secretin pIV

Julian Spagnuolo[1], Natacha Opalka[2], Wesley X. Wen[1], Dragana Gagic[1†], Elodie Chabaud[1], Pierdomenico Bellini[1€], Matthew D. Bennett[1‡], Gillian E. Norris[1], Seth A. Darst[2], Marjorie Russel[2] and Jasna Rakonjac[1*]

[1]Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

[2]The Rockefeller University, New York, NY, USA

Current address:

[†]AgResearch Grasslands, Palmerston North, New Zealand

[€]School of Biosciences, University of Liverpool, United Kingdom

[‡]Department of Chemistry and Biochemistry, Gene Center, Ludwig-Maximilians–University Munich, Germany

[*]Corresponding author

E-mail address of the corresponding author: j.rakonjac@massey.ac.nz

Supplementary Figure 1

Supplementary Figure 1. Alignment of the gate region of 55 secretins. The sequences of the secretin family domains (PF00263) of 55 secretins from the seed alignment were imported from the Sanger Institute Pfam database (http://pfam.sanger.ac.uk/family?acc=PF00263; (Finn *et al.*, 2006)) and aligned using the ClustalW algorithm within the Vector NTI package. Top, schematic diagram showing the location of segments of interest: red boxes labeled with the letter "β", transmembrane β strands predicted using the TMBETA algorithm (Gromiha *et al.*, 2005); green and black cross-hatched boxes – GATE1 and GATE2 regions as determined in this work; blue arrow, the position of the His tag; black arrows, positions of two conserved hydrophobic residues corresponding to pIV I312 and I316 whose site-directed mutation into glycines destabilized the multimer. Below, section of the secretin family domain seed alignment showing the segment spanning GATE1, GATE2, inter-gate region and the predicted flanking transmembrane β strands. The title of the f1 pIV sequence, as well as the GATE1 and GATE2 regions of the f1 pIV, are boxed. Color setup of the amino acid residues within the alignment: black letters on white background, non-similar; navy letters on aquamarine background, conservative; black letters on green background, block of similar; red letters on yellow background, identical; dark green letters on white background, weakly similar; black dashes, gaps introduced into alignment. The accession numbers are shown in the alignment.
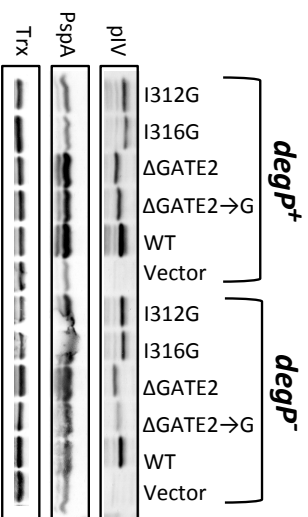
Supplementary Figure 2

A

Leaky point mutations in *degP*+ strain

B
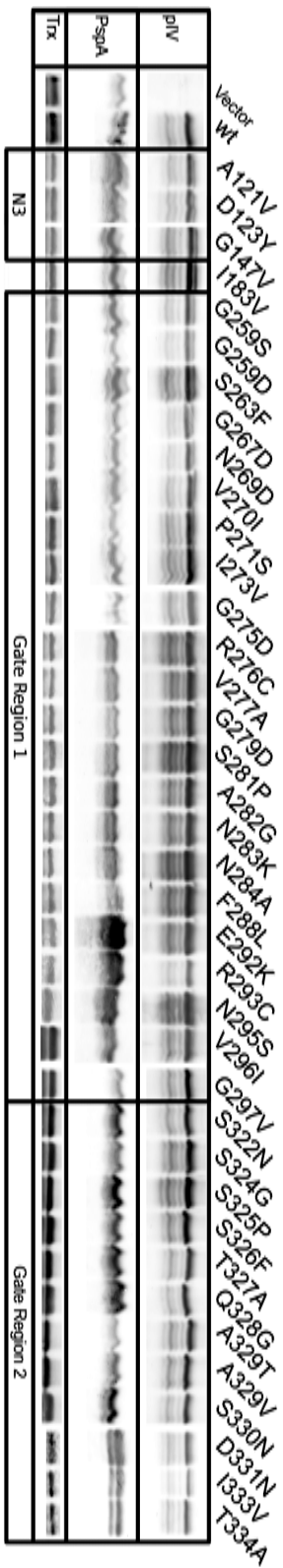
**Site-specific mutations in *degP*+ and *degP*- strain**

Supplementary Figure 2. PspA and pIV in the total cell lysates. A. Leaky point mutants obtained by random mutagenesis in this work and two previously isolated mutants (N295S and S324G), expressed in the selection strain K1508 (MC4100 *ΔlamB106*). **B.** Site-specific mutants constructed in this work. Lysates are derived either from K1508 or K2040 (K1508 *degP41(ΔPstI*-Km^R) transformed with the appropriate plasmids. WT, cells containing plasmid pPMR132 that expresses the wild-type pIV; Vector, cells containing the vector pGZ119EH (without *gene IV*). Proteins of interest were detected by western blotting. Trx (TrxA) blot represents the loading control. Lysates analyzed in this figure were obtained from the same cultures that were used as a source of RNA for the quantitative RT-PCR.

Methods:

Sample preparation and electrophoresis: The total cell lysates for protein analysis by SDS-PAGE in A and B were prepared using a TCA precipitation protocol which was found to dissociate otherwise SDS-resistant pIV multimer. Briefly, 0.5 mL of culture was mixed with an equal volume of ice-cold 10% TCA and incubated on ice for at least 10 min, centrifuged and the pellet was washed in 100% acetone, dried and resuspended in 4% SDS. The resuspension volume depended on the final OD of each culture and was adjusted to obtain an extract equivalent to 10 OD$_{600}$ units of cells per mL. The resuspended pellet was boiled for 5 min and the remaining insoluble material was removed by centrifugation. The supernatant was mixed with sample buffer (125 mM TrisHCl pH 6.8, 4% SDS, 10% β-mercaptoethanol, 0.01% bromophenol blue and 20% glycerol (Laemmli, 1970), boiled for another 3 min and separated by SDS-PAGE.

For detection of pIV by SDS-PAGE, proteins were separated on 12% and for PspA and TrxA on 14% acrylamide gels. After electrophoresis the proteins were transferred to the nitrocellulose filters and blotted by appropriate antibodies for detection of PspA, pIV and TrxA, as described in Experimental procedures.

114

Supplementary Figure 3
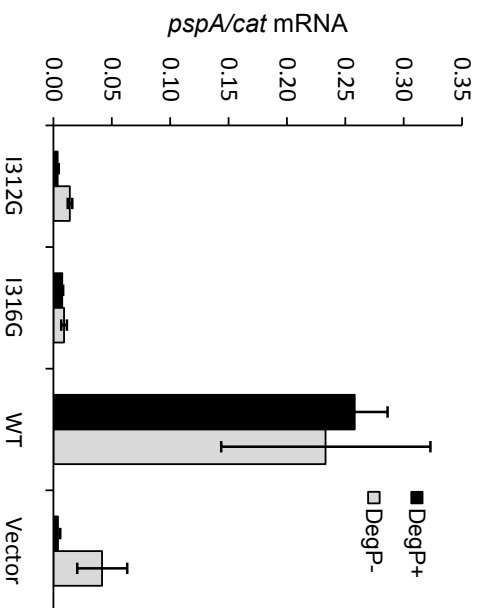
Supplementary Figure 3. **Induction of the *psp* promoter by inter-gate pIV mutants I312G and I316G.** The amount of *pspA* mRNA was determined by quantitative RT-PCR using as a reference the *cat* transcript (encoded by the plasmids). RNA was purified from the same cultures as those used in Figs. 3. and S2B. Each value represents the mean of 3 qPCR reactions; error bars indicate the standard error of the mean. For methods, please refer to the "Experimental procedures" section of the manuscript.
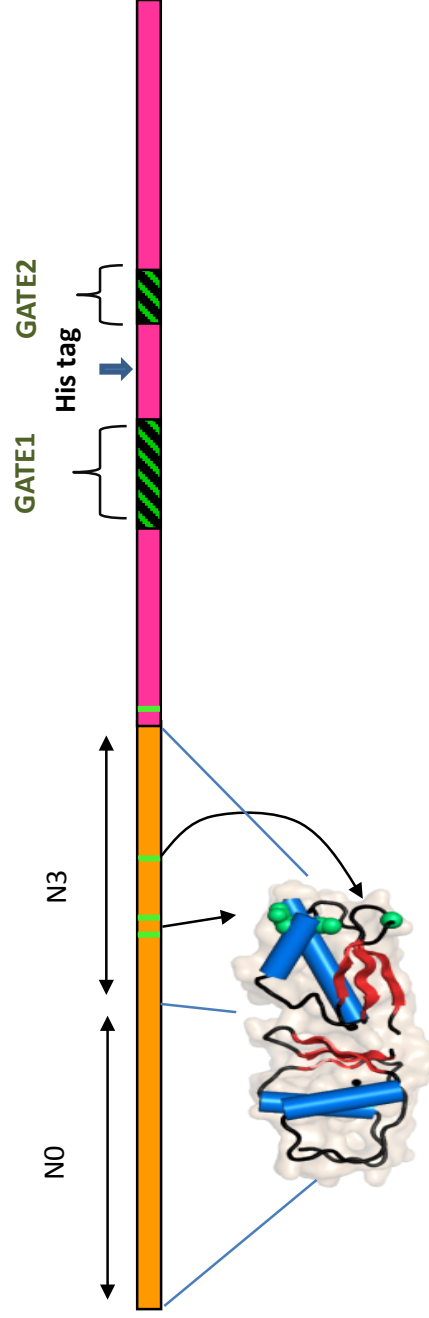
Supplementary Figure 4
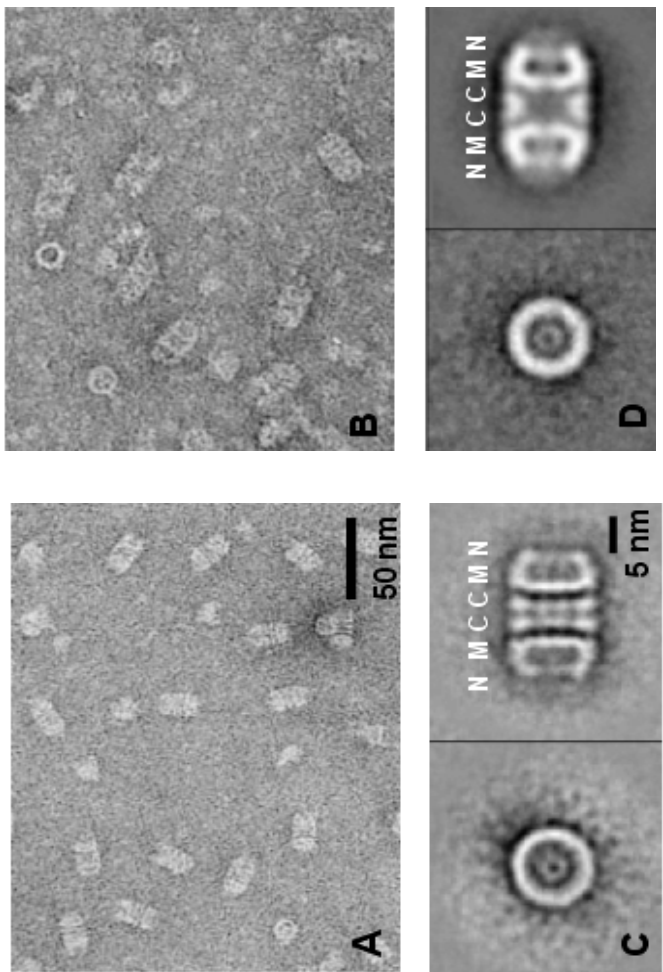
Supplementary Figure 4. Electron microscopy of negatively stained pIV multimer. A and **B**, micrographs of the wild-type (A) and the ΔGATE2 mutant (B) pIV. **C** and **D**, averaged images of side (left) and top (right) views of negatively stained particles of the wild type pIV (C) and the ΔGATE2 mutant (D). The densities corresponding to N, M and C-rings are indicated. Please note that pIV appears in the negative stain electron microscopy as a multimer-dimer, with the C-rings forming the interface (Linderoth *et al.*, 1997, Opalka *et al.*, 2003).

Methods:

Negative stain electron microscopy: A 5μL sample of purified pIV was applied to a glow discharged carbon film coated grid, washed once with a drop of water, stained with a drop of 1% (v/v) uranyl acetate followed by brief blotting, and then air dried. Images were recorded at a magnification of × 60,000 with a CM12 Philips transmission electron microscope (Philips Electron Optics, Eindhoven, The Netherlands) using low dose methods.

Image processing and 2D averaging: For analysis of negatively stained images, micrographs were selected and digitized at a pixel size of 21 μm, corresponding to 3.5Å on the image. Individual particles were selected by eye, windowed in 90 × 90 pixel images and subjected to reference free alignment to generate 2D averages (Penczek *et al.*, 1992).

Supplementary Figure 5

Supplementary Figure 5. The 3D model of the pIV N-terminal domain aligned to the linear diagram of the full length pIV. The model corresponds to the complete N-terminal domain of pIV which contains two subdomains, N0 and N3. It is a prediction obtained using Swiss Model (Arnold *et al.*, 2006), based on the coordinates of EscC (Spreter *et al.*, 2009) and. The mutated residues of the N1 subdomain, A121, D123 and G147, are emphasized in the model as green balls (each ball corresponding to a carbon atom in α, β and γ positions as appropriate). Top, a linear diagram of pIV, also shown in Fig. 1A.

References

Arnold, K., L. Bordoli, J. Kopp & T. Schwede, (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195-201.

Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer & A. Bateman, (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* **34**: D247-251.

Gromiha, M. M., S. Ahmad & M. Suwa, (2005) TMBETA-NET: discrimination and prediction of membrane spanning b-strands in outer membrane proteins. *Nucleic acids research* **33**: W164-167.

Laemmli, U., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Linderoth, N. A., M. N. Simon & M. Russel, (1997) The filamentous phage pIV multimer visualized by Scanning Transmission Electron Microscopy. *Science* **278**: 1635-1638.

Opalka, N., R. Beckmann, N. Boisset, M. N. Simon, M. Russel & S. A. Darst, (2003) Structure of the filamentous phage pIV multimer by cryo-electron microscopy. *J. Mol. Biol.* **325**: 461-470.

Penczek, P., M. Radermacher & J. Frank, (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* **40**: 33-53.

Spreter, T., C. K. Yip, S. Sanowar, I. Andre, T. G. Kimbrough, M. Vuckovic, R. A. Pfuetzner, W. Deng, A. C. Yu, B. B. Finlay, D. Baker, S. I. Miller & N. C. Strynadka, (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nature structural & molecular biology* **16**: 468-476.