Structure and Membrane Topography of the Vibrio-Type Secretin Complex from the Type 2 Secretion System of Enteropathogenic Escherichia coli

Iain D. Hay,a Matthew J. Belousoff,a Rhys A. Dunstan,a Rebecca S. Bamert,a Trevor Lithgowa

aInfection and Immunity Program, Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, Australia

ABSTRACT The β-barrel assembly machinery (BAM) complex is the core machinery for the assembly of β-barrel membrane proteins, and inhibition of BAM complex activity is lethal to bacteria. Discovery of integral membrane proteins that are key to pathogenesis and yet do not require assistance from the BAM complex raises the question of how these proteins assemble into bacterial outer membranes. Here, we address this question through a structural analysis of the type 2 secretion system (T2SS) secretin from enteropathogenic Escherichia coli O127:H6 strain E2348/69. Long β-strands assemble into a barrel extending 17 Å through and beyond the outer membrane, adding insight to how these extensive β-strands are assembled into the E. coli outer membrane. The substrate docking chamber of this secretin is shown to be sufficient to accommodate the substrate mucinase SteC.

IMPORTANCE In order to cause disease, bacterial pathogens inhibit immune responses and induce pathology that will favor their replication and dissemination. In Gram-negative bacteria, these key attributes of pathogenesis depend on structures assembled into or onto the outer membrane. One of these is the T2SS. The Vibrio-type T2SS mediates cholera toxin secretion in Vibrio cholerae, and in Escherichia coli O127:H6 strain E2348/69, the same machinery mediates secretion of the mucinases that enable the pathogen to penetrate intestinal mucus and thereby establish deadly infections.

KEYWORDS protein secretion, bacterial pathogenesis, BAM complex, translocation, outer membrane, lipoproteins, secretin

Enteropathogenic Escherichia coli (EPEC) and enterohemorrhagic E. coli (EHEC) are pathogens causing diseases that range from diarrhea to hemorrhagic colitis and hemolytic-uremic syndrome in humans. In order to access the epithelial layer of the human gut, these pathogens secrete mucinases that cause a thinning of the mucus layer to accelerate pathogenesis (1–3). These large mucinases are folded into functionally active forms in the bacterial periplasm before being secreted across the outer membrane (1, 4–8). Secretion of these mucinases is mediated by the type 2 secretion system (T2SS), with a defining component of the T2SS being the secretin: an approximately 70-kDa protein that homo-oligomerizes to form an ~1-MDa secretin complex (9). The secretin complex is embedded in the outer membrane and spans the periplasm (10). Diversity in the sequence features of secretins led to the classification of two major forms of the T2SS: (i) the Klebsiella type, represented in a broad and diverse set of bacterial species and which has been functionally characterized best in Klebsiella oxytoca (11), and (ii) the Vibrio type, which is found only in a few species of Vibrio and pathovars of Escherichia coli, including EPEC and EHEC (7). Other secretins, including those from the T2SS expressed by species of Pseudomonas, Legionella, and Acinetobacter, are so diverse in sequence that their relationships have remained complex and
obscure (7, 12). As the number of genome sequences available for analysis grows, these relationships too are becoming more clear (13).

The targeting pathway that delivers secretin monomers to the outer membrane has been studied intensively, particularly in *Klebsiella*. In *K. oxytoca*, the secretin (PulD) is synthesized with a signal sequence to allow its passage as an unfolded monomer into the periplasm (14). Once in the periplasm, the C-terminal S-domain of the PulD monomer is bound by a pilotin lipoprotein (PulS) and the PulS-PulD complex is trafficked across the periplasm to the inside surface of the outer membrane, utilizing the LOL pathway (15–17). The same assembly pathway is evident in bacteria such as *Erwinia chrysanthemi* (the secretin is OutD, and the pilotin is OutS) and many strains and pathovars of *E. coli* (the secretin/pilotin pair is either GspD/GspS or EtpD/EtpO [18, 19]). The targeting of the *Vibrio*-type secretin to the outer membrane follows equivalent steps but, through a striking example of convergent evolution, depends instead on a structurally distinct pilotin referred to as AspS that evolved independently to perform the same function (7). Thus, the S-domains in the *Vibrio*-type secretins and *Klebsiella*-type secretins have strikingly different amino acid compositions that provide recognition signals (7). These distinctive signals in the respective S-domains have been shown to dictate pilotin specificity, with chimeric secretin constructs incorporating the S-domain sequences from PulD dictating reliance on PulS (20).

How the secretin monomers then assemble to form the outer membrane channel of the T2SS has remained unclear. Typically, integral proteins are assembled into the bacterial outer membrane by the β-barrel assembly machinery (BAM complex [21, 22]). However, depletion of BamA, the core component of the BAM complex from *E. coli*, does not decrease the rate of assembly for secretin monomers from *Klebsiella* or from EPEC (23, 24). In order to address the question of how the T2SS secretin is assembled into the bacterial outer membrane, we analyzed the structure of the secretin complex from EPEC O127:H6 strain E2348/69.

The secretin complex was purified and analyzed by cryo-electron microscopy (cryo-EM), revealing structural details down to a resolution of 3.3 Å. Comparisons to monomeric and homo-oligomeric β-barrel proteins that are assembled by the BAM complex revealed that all contain a highly conserved belt of hydrophobic residues stretching around a β-barrel domain, of sufficient height to sit within the lipid phase of the outer membrane. However, in the GspD oligomer, 15 secretin subunits oligomerize into a 60-stranded pentadecamer β-barrel with each subunit contributing two full-length β-strands, one segmented β-strand, and two incomplete β-strands which are keyed into β-strands of the preceding and succeeding subunits to give a 60-stranded barrel. The length of the β-strands thereby formed in the barrel domain of the secretin complex is approximately 78 Å with an average of 22 residues in each strand, far more extensive in length than any known substrate of the BAM complex. Sitting astride this belt is an aromatic girdle, a feature known to designate the line at which the lipid phase of the bacterial outer membrane engages integral membrane proteins. The secretin complex, however, extends 50 Å below the membrane boundary with a pair of short α-helices sitting against or in the bottom face of the outer membrane. Taken together with structural information on the mucinase substrates, the size of the barrel domain so formed is sufficient to enable the passage of these very large substrates across the outer membrane in a prefolded state.

**RESULTS**

**Architecture of the EPEC secretin complex.** Particles of GspD from EPEC O127:H6 strain E2348/69 form a pentadecameric secretin complex (Fig. 1A and B) with no other symmetries observed in the two-dimensional (2D) class averages. Figure 1B shows how the domains from each GspD monomer are situated within the architecture of the secretin complex. The upper chamber, inner gate, and cap could be clearly resolved, whereas the N-domains were less well resolved (Fig. 1A). The N-domains contribute to a large periplasmic chamber which is known to dock onto the pseudopilus structure emanating from the inner membrane (10, 25). This N-domain chamber was less well
resolved than the upper secretin domain, suggesting that there is relatively high flexibility/movement among the N-domains. Approximately 44,000 particles were used to construct a 4.4-Å map from the Falcon II data, and 8,800 particles were used to generate a 3.3-Å map from the K2 data (Fig. 1B). Individual β-strands could be easily identified in the 3.3-Å map, and much of the sequence from the N3-domain onward could be built into the map (Fig. 1C). Residues around the center of the inner gate (corresponding to residues 462 to 473) and cap

**FIG 1** Structure of the EPEC secretin. (A) Flow chart summarizing the expression and purification of the secretin complex is shown alongside a final size exclusion chromatography trace. The arrow indicates the elution position of the complex at ~1 MDa corresponding to the sample for electron microscopy: 2D class averages of “side” and “top” views from this sample show the pentadecamer arrangement of the multimer. (B) The 3.3-Å electron density map of the EPEC secretin derived from the K2 summit camera and the 4.3-Å electron density map derived from the Falcon II camera. The orange density corresponds to density that could be assigned to one GspD monomer. The blue density corresponds to density around the center of the symmetry axis that could not be assigned to any protein chains. (C) A slice through the K2 map showing how the derived model fits into the map. Regions 379 to 389 and 462 to 473, which could not be modeled, are circled. (D) The secretin gates and N3-domain were modeled from the K2 data at 3.3 Å, and the N1- and N2-domains were derived from the crystal structure (PDB accession number 3EZJ) with the N1- and N2-domains split and then individually assigned to density from the Falcon II data. The N0-domain could not be assigned. Two rotational views of a single GspD protomer extracted from the complex structure reveal the relative positioning of the characteristic secretin domains.
(residues 379 to 389) could not be assigned (Fig. 1B and C). High-resolution density/noise was apparent in these regions, but this is probably due to amplified artifacts at the axis of the \( C_{15} \) symmetry imposed during map refinement. Map generation without symmetry imposed yielded poor results.

The 4.4-Å map generated from the Falcon II data was not sufficiently detailed to build an atomic model into, but the crystal structure of the N2-domain (PDB accession number 3EZJ) was readily placed in the density map below the N3-domain. Likewise, since GspD from ETEC and that from EPEC have identical primary structures, the crystal structure of the N1-domain of ETEC GspD (PDB accession number 3EZJ) was placed into the density, immediately below the N2-domain based on the position of two visible \( \alpha \)-helices. No density for the N0-domain could be assigned. The relative position of the N1- and N2-domains was combined with the model of the N3-domain and secretin domain from the 3.3-Å map to produce a model of the complex (Fig. 1D and 2A). The domain architecture of secretin monomers (Fig. 1D) consists of four N-terminal domains (N0 to N3) followed by the highly conserved secretin domain (Pfam PF00263) and a C-terminal S-domain that is necessary and sufficient for pilotin binding (16, 17, 24). The disposition of the S-domain in the complex structure is detailed in Fig. S2 in the supplemental material.

Based on sequence analysis, we previously proposed that the secretin from EPEC O127:H6 strain E2348/69 was of the same subtype as that from Vibrio cholerae and coined the term Vibrio-type secretins to describe this protein subfamily (7). Recently, Yan et al. solved the structure of the secretin from V. cholerae (26), providing a basis to assess the similarities of the two secretins.

Among the known secretin structures, the EPEC GspD T2SS secretin (covering the N3-domain to the S-domain) shares 54.3% identity to that of EspD from V. cholerae and 47.7% identity to GspD from E. coli K-12. Overall, a similar architecture can be seen in all structures, though the overall architecture of the EPEC secretin domains appears to best resemble that of V. cholerae EspD with loops extending from the extracellular side to form an exterior gate/cap (Fig. S3A). More detailed root mean square deviation (RMSD) structural analysis reveals that as expected the EPEC GspD structure is very similar to Vibrio EspD, with the only minor differences apparent at the cap and poorly resolved S-domain. Compared to the E. coli K-12 (Klebsiella-type) T2SS structure, obvious differences are apparent in the cap structure and transmembrane region (Fig. S3A and B). EPEC GspD also shares 17.5% identity to the T3SS secretin InvG from Salmonella enterica serovar Typhimurium. The architecture of the T3SS secretin resembles that of the Klebsiella-type T2SS secretins with an open cap at the surface of the cell, though the T3SS lacks the N2-domain. Also, InvG has an additional \( \beta \)-hairpin protruding from the N3-domain pointing up into the lumen of the lower chamber, and the RMSD of the inner barrel shows slight differences between the structures (Fig. S3B). Overall, the structures of the N3 domain and the outer barrel are strikingly well conserved among the T2SS and T3SS secretins (Fig. S3B).

The GspD \( \beta \)-barrel chamber. The EPEC secretin complex is resolved into four important structural features: the cap gate, an extensive \( \beta \)-barrel chamber, a periplasmic constriction, and a central gate. A longitudinal cross section of the secretin complex highlights these features (Fig. 2A). The \( \beta \)-barrel chamber is a unique protein structure, consisting of two concentric barrels: a short inner barrel composed of 60 \( \beta \)-strands (4 antiparallel \( \beta \)-strands from each protomer) that are aligned with the long axis of the secretin complex and the large outer barrel with a diameter of approximately 110 Å (Fig. 2B). This “inner barrel” contributes to the central gate structure. The outer barrel has a complex \( \beta \)-strand arrangement comprised of 30 long “strands” running the full height of the barrel and 60 nonsequential \( \beta \)-strands interleaved among the long strands. To form this complicated outer barrel, each subunit contributes a total of 6 \( \beta \)-strands to the barrel. The walls of the barrel are hydrogen bonded together with the topographically first (short) strand of one subunit stacking onto the last \( \beta \)-strand of the preceding subunit (Fig. 2C).
FIG 2 Elongated β-strands in the EPEC secretin determine its cavity diameter. (A) Slice through the Falcon II map showing the fitted model of the complex. The dimensions of the complex are shown. (B) Zoom views of the double β-barrel. The gray ellipse (in the lower panel) represents the unassigned density in the middle gate corresponding to amino acids 462 to 473. (C) Properties of the giant β-barrel, including calculations of its shear number (S) and diameter. (D) How shear affects the properties of a β-barrel. The values presented here are derived from the formulas \(\tan \alpha = Sb/n\) and \(R = b/2 \sin (\pi/n)\) described in the work of Murzin et al. (72), where \(\alpha\) is the average slope of the strands in the barrel, \(R\) is the radius of the barrel, \(a\) is the Cα-to-Cα distance along the peptide, \(b\) is the interstrand distance, \(n\) is the number of strands, and \(S\) is the shear number. Shear numbers of \(S = 0\) to \(S = 3n\) are shown in the graphs. The maximum theoretical H bonds are calculated assuming that every other residue is bonded to the adjacent residue on the next strand.
The multimeric barrel is stabilized by a C-terminal $\alpha$-helix, folded down over a negatively charged patch on the outer wall of the preceding subunit. This can be further stabilized by at least one intrasubunit salt bridge between the bottom of the helix and wall of the barrel (Lys$^{598}$-Glu$^{326}$) and one intersubunit salt bridge between Arg$^{609}$ on the helix of one subunit and Glu$^{334}$ on the outer wall of the preceding subunit which locks the subunits together (Fig. S2). Analysis of other T2SS secretin structures (26) revealed that these inter- and intrasubunit salt bridges that tie the subunits together are conserved, thereby providing an explanation for the well-documented thermal and chemical quaternary stability of secretin complexes (9).

In transmembrane $\beta$-barrel proteins, two parameters dictate the luminal diameter of the barrel: the number of $\beta$-strands ($n$) and their shear number ($S$) (27). The shear number indicates the stagger of each $\beta$-strand relative to its neighboring strands. In transmembrane $\beta$-barrel proteins, $n$ is almost always an even number and must be greater than a value of 8 (28). $S$ must be greater than or equal to 0, with $S = 0$ describing $\beta$-strands that sit perpendicular to the membrane. The outer barrel consists of four very long and discontiguous $\beta$-strands with a shear number of 60 ($n = 4$ for each of the 15 subunits), resulting in an angle of $\sim$35° to the long axis of the secretin complex. These strands run the length of the complex, connecting the N-terminal domains to the cap and inner gate (Fig. 2C). For a given number of strands, a greater shear number will produce a greater diameter and volume of the resultant barrel at the expense of the height and theoretical stability of the $\beta$-barrel (Fig. 2D).

**Transmembrane topography of secretins.** In a $\beta$-barrel protein, the side chain of every second residue points outward. In a transmembrane $\beta$-barrel protein, these residues have side chains with high hydrophobicity scores in order to be accommodated in the lipid phase of the membrane. Calculations of hydrophobicity were graphically represented on the structure of the secretin complex (Fig. 3A) and show a hydrophobic belt which is 17.6 Å wide, forming a layer across the top of the long $\beta$-strands; this is slightly narrower than the width of other well-characterized outer membrane $\beta$-barrels (Fig. 3A to E). Included in this hydrophobic belt is an aromatic residue (Trp$^{419}$) whose placement forms the upper boundary of the hydrophobic belt and Phe$^{340}$ and Phe$^{354}$, which together form the lower boundary. An equivalent, though larger, hydrophobic belt and aromatic girdle are seen in the monomeric $\beta$-barrel proteins FhaC (Fig. 3B) and FimD (Fig. 3C) and also in the homo-oligomeric $\beta$-barrels ToIC (Fig. 3D) and CsgG (Fig. 3E). The aromatic girdle present in EPEC GspD is a highly conserved feature among the sequences of S81 Vibrio-type GspD homologues (Fig. 3F and G). The Klebsiella-type T2SS secretins as shown for the structure of GspD from E. coli K-12 strain DH5$\alpha$ have similar conserved aromatic residues (Fig. 4A and B). However, in addition to the residues that form the girdle, the Klebsiella-type secretins also have two additional and conserved aromatic residues (one outward facing and one inward facing) near the top of the second hairpin (Fig. 4A and B).

Members of the secretin protein family are also found in the needle complex of the type 3 secretion system (T3SS) (29–31). Despite showing some overall principles in common, the other components and the overall architectures of the T2SS and T3SS are quite distinct, as is the mechanism of substrate entry into the complex (Fig. 4C). In EPEC, the T3SS secretin is EscC, and the structure of the homologous protein InvG from Salmonella was recently solved (31). Comparative hydrophobicity analysis of the T3SS secretin complex revealed a major distinction from the T2SS. The T3SS secretin InvG/EscC has a hydrophobic band that could serve for membrane integration, but it is tilted inward toward the lumen of the barrel. This distinguishing transmembrane region is due to the 4 $\beta$-strands of the barrel transitioning to 3 strands around the hydrophobic band at the membrane interface (Fig. 4D and E). The transmembrane region of InvG/EscC has fewer conserved aromatic residues, with only two aromatic residues that are highly conserved in the transmembrane region (Fig. 4F).

**Structural distinctions in the T2SS subtypes found in E. coli.** As a result of lateral gene transfer, two different types of the T2SS secretin protein family are found in the
FIG 3 Topological comparison of the secretin complex and other β-barrel proteins in the bacterial outer membrane. (A) Surface rendering of the secretin complex from EPEC, where residues with hydrophobic side chains are colored red according to the Eisenberg hydrophobicity scale (73). (Inset) Zoomed view of the aromatic girdle, with the side chains of aromatic residues (Phe and Tyr) represented as sticks (in blue). (B to E) Hydrophobic belts and aromatic girdles of other bacterial outer membrane proteins, shown to scale. The measurements of the transmembrane sections for all proteins are from the Orientations of Proteins in Membranes (OPM) database and the PPM server (69); FhaC (PDB accession number 2QDZ) (B), FimD (PDB accession number 4J3O) (C), TolC (PDB accession number 1EK9) (D), and CsgG (PDB accession number 4UV3) (E). (F) The transmembrane regions of the EPEC GspD secretin (PDB accession number 5W68) representing the Klebsiella-type T2SS secretin; the conserved aromatic residues are shown in blue. (G) The sequence logo showing the conservation of the aromatic residues in the transmembrane β-strands was generated from 581 Vibrio-type T2SS secretin sequences with the GspD sequence from EPEC O127:H6 strain E2348/69 shown for reference. The locations of the putative transmembrane β-strands are shown in gray below the sequence logos.
various pathotypes and environmental isolates of *E. coli*: the *Vibrio*-type and *Klebsiella*-type secretins (7). The secretin from EPEC O127:H6 strain E2348/69 belongs to the *Vibrio*-type secretins, and we sought to rationalize the structural features that define these two different secretin types. In their recent work, Yan et al. (26) presented the
structures of the *Klebsiella*-type secretin from *E. coli* K-12 substrain DH5α (PDB accession number 5WQ7).

The secretin from EPEC O127:H6 strain E2348/69 has features that distinguish it from the *Klebsiella*-type secretin (Fig. 5A). Notably, the cap gate is longer: it protrudes above the line of the outer membrane, and it appears to be closed (based on the cryo-EM density). In the *Klebsiella*-type secretin from *E. coli* K-12 substrain DH5α, the cap gate is shorter and open to the extracellular environment. The electrostatic charge of the interior of the cavity is also a distinguishing feature, with the *Vibrio*-type secretins containing alternating positively and negatively charged bands, whereas the *Klebsiella*-type secretin is largely negatively charged with a periplasm-facing positively charged rim (Fig. 5B).

The *Vibrio*-type secretin complex found in EPEC and EHEC secretes substrates such as the mucinases StcE and SslE. The crystal structure of StcE (PDB accession number 3UJZ) is shown here represented to scale with the secretin complex structure (Fig. 5B). The *Klebsiella*-type secretin complex is well documented as secreting pullulanase, a starch-degrading enzyme that enables the bacterium to access sugars. The pullulanase structure from *K. oxytoca* (PDB accession number 2YOC) is represented to scale with the secretin complex structure (Fig. 5B). In both cases, the substrates can fit into the widest point of the barrel (the opening), a passage made even more penetrable given the apparent flexibility of the N-domain within the complex (observed in Fig. 1A).

**DISCUSSION**

The secretin protein superfamily has members which function in the T2SS, T3SS, type IV pili, and other secretion systems, in all cases forming the pore in the outer membrane that provides for gated transport (29, 32). We confirm here, with the structure of the T2SS secretin GspD from EPEC O127:H6 strain E2348/69, that there are defining three-dimensional (3D) features common to the *Vibrio* type of secretins previously described from sequence-based similarities. These features include the charge characteristics in the lumen of the complex and the specific shape and arrangements in the cap gate. Together with recently published high-resolution structures of the secretins from *Vibrio cholerae* (26) and the *Klebsiella*-type secretin found in *E. coli* K-12 substrain DH5α (26), these structural features suggest how these huge complexes might be assembled in the outer membrane and enable a refinement of our knowledge of how substrates are loaded into the T2SS.

**Protein secretion mechanism for T2SS.** Substrate proteins enter the T2SS from the periplasm, after being translocated across the cytoplasmic membrane by either the Sec translocon or the Tat transporter (12, 33, 34). Despite the highly specific nature of protein secretion via the T2SS, how the substrate proteins are selected for engagement with the secretion machinery remains unclear. The one prerequisite for secretion is that the substrate protein must be in a folded state. Identifying secretion “signals” within T2SS substrates has been complicated since these proteins show a large variation in sequence, structure, posttranslational modifications, and oligomeric states (35). Several studies have shown the importance of surface motifs for the secretion of target substrates, and yet no general features have been identified among all T2SS substrates bioinformatically or biochemically (reviewed by Thomassin et al. [36]). That being said, several structures of secretin complexes do provide further insight into the constraints on the passage of those proteins across the outer membrane.

In the T2SS, the process of protein secretion is an active process that depends on ATP hydrolysis in the cytoplasm promoting the assembly and driving upward movement of a pseudopilus into the N-domain chamber of the secretin complex (37–39). The nature of the interactions between the N-domains of each secretin protomer allows these domains to deform considerably with respect to each other (13), and this “mobility” was observed in the 2D class averages of the EPEC secretin particles when imaged by electron microscopy. After being “actively pushed” into this accommodating chamber, the major impediment to substrate passage remains the central gate.
FIG 5 Cryo-EM structures reconcile the sequence differences seen in the Vibrio-type and Klebsiella-type secretin protein families. (A) Comparison of the Vibrio-type (EPEC O127:H6 strain E2348/69) and the Klebsiella-type (E. coli K-12 substrain DH5α) T2SS secretin complexes. The structures of the aligned models are shown as an overlay with green representing EPECO127:H6 strain E2348/69 and red representing the E. coli K-12 substrain DH5α structure. (B) Rendering of the cavity space for the Vibrio-type secretin complex, alongside rendering of the cavity space for Klebsiella-type secretin complex. The electrostatic potentials of the surfaces are shown, with red being negatively charged and blue being positively charged. Structures for the T2SS substrates protease StcE (PDB accession number 3UJZ) and pullulanase PulA (PDB accession number 2YOC) are represented to scale below the structure of the Vibrio-type and Klebsiella-type secretin complexes. The widest opening of the chamber is shown in gray, and the narrowest point (the N3 constriction) is shown in darker gray.
The structuring of the central gate from short segments of β-strand allows for at least two scenarios by which it could open. Breaking the few, interstrand hydrogen bonds that structure this gate would allow for a lateral shift, thereby opening in a mechanism analogous to that of an iris diaphragm. Alternatively, breaking the interstrand hydrogen bonds that structure the gate could simply allow the upward displacement of the sectors of the gate to align against the wall of the β-barrel, just below the region embedded in the outer membrane. A G453A mutation in the secretin from *Vibrio cholerae* captured the central gate open (26) but did not distinguish between the two opening mechanisms. Considering that the size of the characteristic mucinase (this study) or cholera toxin (32, 40) would be barely accommodated in the upper chamber of the secretin complex, the opening mechanism might provide elastic energy to promote the final push of the substrate through the cap gate. By comparison to the central gate, the cap gate structure would be relatively easy to deform given that it is formed of interstrand loops in the EPEC structure (this study) and other *Vibrio*-type secretins (26). The cap gate structure is absent from the *Klebsiella*-type secretins (26).

**Assembly of secretin complexes into bacterial outer membranes.** GspD is not assembled into the bacterial outer membrane by the BAM complex (23, 24). Instead, secretin monomers are transferred from the SecYEG complex in the cytoplasmic membrane to a dedicated chaperone called the pilotin. For *Klebsiella*-type secretins, the pilotin is a homologue of the *Klebsiella* protein PulS, whereas *Vibrio*-type secretins use a structurally unrelated protein called AspS (7). While we know most about these two forms of secretin, it is becoming clear that others, such as the Hxc secretin from a subset of *Pseudomonas* species, might not need pilotins (41) and might represent a new subfamily of secretins (13). Pilotins selectively bind the C-terminal S-domain of their cognate secretin, through an induced fit mechanism that has been captured in biochemical assays and a crystal structure (17, 42, 43). Most of the known pilotins are lipoproteins and thus engage with the LOL pathway for transport across the periplasm to the outer membrane receptor LolB. A long standing question concerns how such a complicated protein structure as a secretin oligomer is assembled into the outer membrane. Two features common to all of the published secretin structures lead us to consider a transertion-based model for assembly of the secretin oligomer into the outer membrane.

The transertion model for protein translocation posits that the signal sequences of proteins will dictate their engagement with the closest Sec translocons in the cytoplasmic membrane. In *E. coli*, the RNA transcribed from the gsp operon has been mapped (44), and multiple translation initiation events (45) ensure that many molecules of GspD (and the other subunits of the T2SS) would be synthesized in close proximity, translocated into the periplasm in close proximity, and translocated to arrive at the same or a neighboring LolB receptor site within a short window of time (46–49). Computer simulations and protein localization studies confirm that transertion creates local regions of high concentration for a given membrane protein (reviewed by Matsumoto et al. [48]). In this scenario, the local concentration of GspD monomers in the outer membrane would be relatively high and would favor an allosteric activation of the monomers to polymerize. The extensive β-strand structure of the secretin complex reported here throws structure-informed support behind this model for secretin assembly. The extensive hydrogen-bond donor and acceptor contacts at the edge of an individual GspD subunit would serve as the template for β-augmentation reactions, providing favorable energetics for the assembly reaction. Each incoming monomer has to form at least 9 new hydrogen bonds between monomers in the main β-sheet region of GspD (Fig. 2C), which equates to a favorable energy change of $\Delta G \approx 45 \text{ kJ/mol}$ (50). As formation of hydrogen bonds is an enthalpically downhill process, taken together, it would appear that oligomerization of GspD monomers is energetically driven by extensive β-sheet formation.
Each protomer incorporated into the growing secretin oligomer would also add to the hydrophobic belt, creating a feature that would be increasingly costly to maintain out of a membrane environment. While not homologous, this mechanism for β-strand engagement is somewhat reminiscent of the means by which pore-forming toxins are thought to assemble into host cell membranes (51), wherein an initial engagement of monomers sets up allosteric events and conformational changes to penetrate a membrane bilayer. Past work with T2SS secretins has shown that intrinsic features in secretin protomers do allow for oligomerization without assistance from cellular machinery (52–54), and the LOL machinery may simply provide the means by which secretins can find each other for an otherwise unassisted integration into the outer membrane.

MATERIALS AND METHODS

Protein expression and purification. The sequence of GspD was amplified from E. coli O127:H6 strain E2348/69 with the addition of a C-terminal hexahistidine tag. The gene was cloned into the plasmid pET-20b (+). This plasmid was transformed into E. coli C43 cells (55). Transformers were grown overnight in LB (Lennox) supplemented with 150 μg ml−1 ampicillin and 0.5% (wt/vol) glucose. Cells were diluted 1:50 in Terrific broth (100 mM potassium phosphate buffer, pH 7.5, 12 g liter−1 trypthone, 24 g liter−1 yeast extract, 5% [wt/vol] glycerol) and incubated at 37°C with shaking until they reached an optical density at 600 nm of approximately 0.8, at which point 0.2 mM IPTG (isopropyl-β-D-thiogalactopyranoside) was added, and cells were incubated at 25°C for a further 16 h.

Cells were harvested by centrifugation at 7,000 × g, washed once in buffer A (20 mM HEPES, 300 mM NaCl), and resuspended approximately 1.5 (vol/vol) in buffer A. DNase I (10 μg ml−1) and then white lysozyme (100 μg ml−1) were added, and cells were incubated for 30 min before they were disrupted with an Avestin EmulsiFlex C3 homogenizer. Cell lysates were clarified by centrifugation at 7,000 × g for 15 min. Cell envelopes were isolated from the clarified lysate by centrifugation at 100,000 × g for 1 h. Cell envelopes were washed once with a glass homogenizer in buffer A and again centrifuged at 100,000 × g for 1 h. The washed envelope fraction was resuspended 1:10 (vol/vol) in buffer A with a glass homogenizer.

The isolated envelopes were mixed 1:1 with buffer A containing 5% Sb3-14 and stirred for 30 min. This solution was loaded onto a Hitrap nickel-nitrilotriacetic acid (Ni-NTA) column (GE Life Sciences). The column was washed with 4 column volumes of buffer A containing 50 mM imidazole and then 4 column volumes of buffer A containing 100 mM imidazole. The protein was eluted from the column with buffer A containing 400 mM imidazole. Amphipol A8-35 (Anatrace) was added to the elution fraction (4:1 volumes of buffer A containing 100 mM imidazole. The protein was eluted from the column was shown to be a pure GspD multimer when analyzed by SDS-PAGE or by electron microscopy (see Fig. S1 in the supplemental material). The GspD multimer was purified on a Superose 6 increase 10/300 GL size exclusion column running in buffer A (1 with Amicon Ultra 0.5-ml 100K centrifuge filtration units, flash frozen, and stored at −80°C until further use.

Protein complexes were flash frozen, and vitreous ice was imaged on a Titan Krios transmission electron microscope with an FEI Falcon II or Gatan K2 Summit direct electron detector (Fig. 1A).

Electron microscopy. Quantifoil R1.2/1.3 holey carbon grids were washed and coated in a layer of lauryldimethylamine oxide (LDAO) to prevent the protein complex preferentially adsorbing to the carbon support as previously described (56). Sample (4 μl) was applied to detergent-treated grids, and frozen-hydrated samples were prepared with a Vitrobot Mark IV (FEI, OR) in a 100% humidity atmosphere with 2-s blot time and no drain time before plunge-freezing into liquid ethane. Grids were imaged on a Titan Krios transmission electron microscope (FEI) at 300 kV. Falcon II data were collected as follows. Images were collected at ×127,000 magnification (1.1 Å/pixel) with 1-s exposures in 7-frame movie mode using a dose rate of 45 electrons per second. The corresponding 17 subframes were fractionated in 7 frames as follows. Subframe 1 was discarded. Subframes 2 to 7 were recorded as frames 1 to 6, respectively. Subframes 8 to 16 were pooled and integrated as frame 7. Subframe 17 was discarded. The defocus was set to a range of 0.6 to 3.5 μm in intervals of 0.2 μm. K2 data were collected as follows. Images were collected at ×130,000 magnification (0.53 Å/pixel in superresolution mode) with 18 frames at 0.4 s per frame with a total dose of 50 electrons/Å2. A random defocus between 0.8 and 3.5 μm was applied for each micrograph.

Data processing. RELION 2 (57) was used as a wrapper to motion correct the movies with dose weighting using MotionCorr 2.1 (58) and estimate CTF with CTFFIND4 (59). Micrographs with acceptable contrast transfer function (CTF) estimations were used to manually pick ~1,000 particles which were used to generate 2D class averages used for RELION autopicking. Particles were extracted with a 288×288-pixel box size and subjected to multiple rounds of 2D classification in RELION—only C1 symmetry could be observed in particles. Selected particles were imported into CryoSPARC (60) for ab initio three-dimensional (3D) classification and final refinement. A total of 45,526 particles were used to generate a 4.26Å-resolution map from the Falcon II data set, and 8,896 particles were used in the final 3.33Å-resolution map from the K2 data set. Resolution of the final reconstructions was determined using the gold standard Fourier shell correlation (FSC) = 0.143 criterion (see Fig. 54 in the supplemental material). Electron microscopy maps derived in this study are available on the Electron Microscopy Data Bank (EMDB) (EMD-8778 and EMD-8779).
Atomic model refinement. A model of the GspD monomer was created using Rosetta (61) and the Vibrio cholerae GspD model as a template (PDB accession number 5WQ8). The monomer was symmetrized around the electron microscopy (EM) electron density map using Situs pdbsymm (62). The resulting model was then subjected to energy minimization to remove any steric clashes. Fitting the model to the EM electron density map was achieved using the MDFF routine in namd (63). The fitted model was further refined by rounds of manual model building in Coot (64) and real space refinement as implemented in the Phenix software package (65).

Bioinformatic analysis. Protein sequences from the InterPro family “T2SS_GspD” (IPR013356) were downloaded and classified by an all-against-all BLAST search. Sequences were clustered based on pairwise similarities and visualized with CLANS (66). Sequences falling into the Vibrio and Klebsiella types (581 and 363 sequences, respectively) were extracted. Multiple sequence alignments were conducted with Clustal Omega (67) and visualized with WebLogo3 (68). For T2SS analysis, the InterPro “T3SS_OM_pore_YscC” (IPR003522) was downloaded and all proteins with a similar domain architecture (1,543 sequences, IPR005644) were used for further analysis in Clustal Omega. The predicted transmembrane regions of the proteins presented in Fig. 3 and 4 were calculated with the Orientations of Proteins in Membranes PPM server (69).

Image processing. All images were generated with PyMOL or UCSF Chimera (70). Electrostatics were calculated with APBS (71).

Accession number(s). The 3.3-Å model of the EPEC GspD secretin domain derived from this study is available at the PDB accession number 5W68.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JB.00521-17.

SUPPLEMENTAL FILE 1, PDF file, 9.1 MB.

ACKNOWLEDGMENTS

We acknowledge the support staff and facilities at the Monash Ramaciotti Centre for Cryo-Electron Microscopy. We thank Christopher Stubenrauch for critical comments on the manuscript.

I.D.H. was the recipient of an NVIDIA Hardware grant, and we gratefully acknowledge the support of NVIDIA Corporation for the donation of the TitanX GPU used for this research. Our work in this area is supported by NHMRC program grant 1092262 and the Australian Research Council (ARC; FL130100038). I.D.H. is an ARC Laureate Postdoctoral Fellow, and T.L. is an ARC Australian Laureate Fellow.

REFERENCES


