Architecture of a bacterial type II secretion system

Authors: Anastasia A Chernyatina and Harry H Low*

Department of Life Sciences, Imperial College, London, SW7 2AZ, UK

*Corresponding author:

Harry H. Low

Department of Life Sciences, Imperial College, London, SW7 2AZ, UK

Phone: +44 207 594 3064

Email: h.low@imperial.ac.uk

Summary

Bacterial type II secretion systems (T2SS) translocate virulence factors, toxins and enzymes across the cell outer membrane (OM). An assembled T2SS has not yet been isolated in vitro. Here we use a fusion of negative stain and cryo-electron microscopy (EM) to reveal the core architecture of an assembled T2SS from the pathogen *Klebsiella pneumoniae*. We show that 7 proteins form a ~2.5 MDa complex that spans the cell envelope. The outer membrane complex (OMC) includes the secretin PulD with all domains modelled and the pilotin PulS. The inner membrane assembly platform (AP) components have an overall stoichiometry of six copies each. The AP includes PulE and the cytoplasmic domain of PulL, which combine to form a flexible hexameric hub consistent with an inactive state. Symmetry mismatch between the OMC and AP is overcome by PulC linkers spanning the periplasm with PulC HR domains binding independently at the secretin base. Our results show the T2SS to have a highly dynamic modular architecture with implication for pseudo-pilus assembly and substrate loading.

Keywords

Type II secretion system, bacterial secretion, assembly platform, secretin, cryo-EM, macro-molecular complex.

Introduction

The bacterial T2SS is found in human pathogens such as *Acinetobacter baumannii*, *Chlamydia trachomatis*, *Escherichia coli* and *Vibrio cholerae*[1]. It secretes a broad repertoire of substrates including digestive enzymes and infective agents like the cholera and heat-labile (LT) toxins[2]. Between 12-15 genes in a single operon usually encode the majority of T2SS components. Whilst the soluble domains for many of these proteins have been solved by X-ray crystallography[3,4], their relative stoichiometry, mode of association with binding partners, and temporal coordination for assembling a functional secretion apparatus is still poorly understood.

The protein GspD forms a 15-fold rotationally symmetric pore termed the secretin that inserts into the OM and provides a conduit for substrate into the external environment. OM insertion is usually dependent on a lipidated pilotin[5], which binds to the GspD C-terminal S-domain with 1:1 stoichiometry[6,7]. The pilotin gene is often chromosomally discrete from the main T2SS operon. Multiple recent high-resolution cryo-EM structures report the partial secretin architecture[8-10], and in complex with the pilotin[7]. However, the entire secretin has not yet been fully resolved due to instability in the periplasmic N0 and N1 domains.

Within the inner membrane (IM), GspL and GspM are bitopic and monotopic membrane proteins, respectively, that together form homo- and hetero-dimers[11]. Combined with the polytopic membrane protein GspF[12] and the ATPase GspE, these proteins constitute an assembly platform (AP) for the pseudo-pilus[13]. The pseudo-pilus constitutes a helical filament that extrudes fully-folded substrate through the secretin channel. The relative stoichiometry and overall ultrastructure of the AP is unknown. GspE is a cytoplasmic AAA+ ATPase that energises the T2SS and drives pilin assembly and pseudo-pilus formation. The active state is considered to be a hexameric ring as ATP

turnover is significantly upregulated in an artificially oligomerized GspE-Hcp1 fusion[14]. The homologous ATPases PilB and PilT in the closely related type IV pilus (T4P) system also function as hexamers[15,16]. Ultimately, the functional oligomeric state and stoichiometry of GspE within the T2SS apparatus has not yet been determined. The GspE N-terminal N1E domain connects to the N2E domain with an extended linker. A shorter but known flexible linker connects the N2E domain to the C-terminal CTE ATPase domain[17]. Such inherent flexibility within GspE is predicted to facilitate large-scale conformational changes. The N1E domain of GspE forms a 1:1 stoichiometric complex with the cytoplasmic domain of GspL[17,18]. These two proteins contact GspF[13,19], which is predicted to reside centrally within the AP. Concerted interplay between the GspE, GspL and GspF complex are thought crucial for coupling GspE conformational changes to the mechanical loading of pilin subunits within the pseudo-pilus assembly[20-21,22]. As GspL is a bitopic membrane protein, the direct contact between GspE and GspL also represents a mechanism for enabling cross-talk across the inner-membrane to other periplasmic components such as GspM. The coupling of the AP and OMC across the cell envelope is mediated by GspC, where the GspC N-terminus associates with GspL and GspM within the inner membrane[11]. The C-terminus of GspC must then span the periplasm as the GspC HR domain binds the GspD N0 domain[23]. Despite the known interaction between GspC HR domain and GspD N0 domain, the precise arrangement of GspC HR domain at the base of the secretin remains unclear[23]. Similarly, understanding how GspC HR domains bind to the secretin is likely critical for determining how the likely symmetry mismatch between AP and OMC components is overcome.

Here we isolate an assembled T2SS so that both OM and IM components are captured together. Using a fusion of cryo and negative stain EM, combined with stoichiometry measurements, we provide a reconstruction of the entire OMC and a model

for the cytoplasmic components of the AP. Combined they reveal the core ultrastructure of this cell envelope spanning nanomachine.

Results

Purification, EM and stoichiometry of $Pul_{CDELMNS}$

The T2SS from the human pathogen *K. pneumoniae* HS11286 strain comprises 12 genes in a single unidirectional operon termed PulC through to PulN (Figure 1A). Note that Pul and Gsp nomenclature relate to equivalent proteins in homologous T2SS systems. The pilotin PulS is located in a separate position within the chromosome. These 13 genes were cloned and over-expressed in *Escherichia coli*. Using affinity chromatography tags positioned on the cytoplasmic ATPase PulE and the periplasmic pilotin PulS, a complex containing 7 components was purified by two successive pulldowns. Glutaraldehyde stabilization was included after the initial pulldown. The complex comprised PulC, PulD, PulE, PulL, PulM, PulN and PulS, and is here termed $Pul_{CDELMNS}$ (Figure 1B). Visualisation of $Pul_{CDELMNS}$ by negative stain EM yielded particles ~40 nm long and 17-22 nm wide (Figure 1C). The OMC PulD secretin was readily identifiable within 2D class averages. Hanging beneath the OMC and separated by a 5-10 nm gap the inner membrane AP was observed. Highly flexible linkers connect the OMC and AP so that these two assemblies effectively constitute independent particles tethered together. The $Pul_{CDELMNS}$ complex was vitrified on thin carbon film and imaged by cryo-EM (Figure 1C). 2D class averages of $Pul_{CDELMNS}$ yielded well resolved side views of the OMC. All domains of PulD were identifiable with additional densities observed at the base of the secretin where the PulC HR domain was expected to bind to the N0 domain[23], and where PulS decorates the exterior of the secretin core[7] (Figure 1C). The AP was not observed here due to high flexibility and averaging effects. In the absence of glutaraldehyde, the OMC sometimes

separated from the AP and yielded top views, which confirmed PulD and PulS in 1:1 stoichiometric ratio[7] with C15 symmetry (Figure 1C). The stoichiometry of the other components within $Pul_{CDELMNS}$ was determined by SDS-PAGE band integration using PulS for calibration (Figure 1B). Band integration for PulD was avoided as its propensity to multimerize despite phenol treatment meant that some PulD sample did not enter the gel properly. The relative stoichiometry of $Pul_{C:E:L:M:N}$ was 7.5:6.1:6.0:5.6:5.6. Overall, the data supports a model where Pul_{CELMN} constitute a 6:6:6:6:6 complex (Figure 1C). PulC stoichiometry may be slightly elevated over the other components as additional copies of PulC may be bound as a monomer-dimer equilibrium within the inner membrane or be independently recruited to spare binding sites at the secretin base (see below).

PulD secretin structure determination

Focused refinement of the OMC yielded a reconstruction with an overall resolution of 4.3 Å (Figure 2 and Figure S1). All domains of PulD were resolved with sufficient map quality (Figure S2A) to build a complete model of the monomer and the secretin, excluding the amino acids (aa) in loops 288-303, 462-470 and 632-637 (Figure 3 and 4A, and Figure S3). The PulD fold is similar to partial *Escherichia coli* K12 and H10407 GspD models (RMSD Cα = 3.2 Å and 3.3 Å) where the secretin core and N3 domain have been described, along with homology modelled N2 and N1 domains[7,10] (Figure 4B). The entire secretin is 20 nm long with an external diameter of 15 nm at the base (Figure 2B and Figure 3). It includes an occluding central gate and N3 domain constriction sites within the secretin channel. It lacks a *Vibrio cholerae* cap gate[10]. The N1, N2 and N3 domains pack tightly (Figure 5) with a diagonal offset of 36˚. N0 is positioned almost directly below the N1 domain and does not maintain the diagonal offset. The N0 fold is similar to that described in multiple crystal structures[23-25] with a core of 2 helices flanked on each side by β-sheets.

However, its position relative to the N1 domain within the secretin is significantly different to these crystal structures where crystal contacts appear to have dominated domain arrangement (Figure S2B). The N0 and N1 domains are connected by loop 7, which constitutes a substantial 26 aa linker. The N-terminus of loop 7 forms a wedge that packs between neighboring N1 domains. Its C-terminus partially envelops the proximal N1 domain whilst making additional secondary contacts with N0 domain helix 2 (Figure 3A and Figure 5D). The N0 domains form a tightly packed ring with alternating stacked β-sheets sandwiched between helices 2 and 4 (Figure 5D). Failure to stabilize the N0 domain and to promote formation of the loop 7 wedge likely accounts for the previously reported N0, N1 and N2 domain flexibility in other systems[7-10,26]. Overall, a single PulD monomer has an azimuthal twist around the secretin long axis of 130° (Figure 3C).

PulC HR domain binds at the secretin base

Hanging beneath the N0 domains in the OMC map, additional globular densities at 7-8 Å resolution protrude from the secretin base (Figure 2 and Figure S1B). Focused refinements[27] failed to markedly improve resolution. These densities were predicted to be the PulC HR domain given its known interaction with the PulD N0 domain in homologous systems[23,28]. Using the GspC HR domain and GspD N0 domain crystal structure[23] as a reference, a homology model of PulC HR domain was fitted as a rigid body (Figure S2C). The model closely follows the surface envelope of the map in this region, with a pair of triple β-sheets opposed around a central cavity (Figure 2C). Based on the quality of this fit, the PulC HR domain was assigned to each globular protrusion. Importantly, each PulC HR domain binds to only a single PulD N0 domain and has no contact with neighbouring GspC HR domains. Given the stoichiometry of PulC (Figure 1B), not every PulD N0 domain has a PulC HR domain bound in the assembled $Pul_{CDELMNS}$

complex. This finding is important as it provides a natural mechanism for overcoming the symmetry mismatch between the OMC and the inner membrane AP, and does not require pseudo-C6 symmetry within the secretin periplasmic domains (Hay et al., 2017). Overall, the binding of the PulC HR domain to the PulD N0 domain appears to be important for the correct positioning of the N0 domain within the secretin and the subsequent stabilization of the N1 and N2 domains. Critically, substantial additional stabilization is derived from a plug that occludes the lumen of the secretin at the level of the N0-N1 domains (Figure 2B). The plug was observed in both the 2D class averages (Figure 1B) and the 3D reconstruction. Note that high resolution plug ultrastructure was not resolved due to likely symmetry mismatch with the C15 averaged OMC. Attempts to resolve the plug structure through refinement using lower symmetries yielded reconstructions of insufficient quality and resolution. A speculative candidate for plug formation is the PulC PDZ domain given its role in substrate recruitment[29] and positive modulation by the PulD N1 domain[30]. Truncation of the PulC PDZ domain destabilized the $Pul_{CDELMNS}$ complex leading to aggregated and poorly assembled particles.

The PulS pilotin decorates the secretin core

Decorating the outside of the secretin core proximal to the PulD S-domain in the map, globular densities were observed in a position consistent with the pilotin AspS relative to GspD in ETEC[7] (Figure 2). For these densities, map resolution was limited to ~7 Å (Figure S1) and focused refinements[27] did not markedly improve resolution. A homology model of the PulS pilotin in complex with the PulD S-domain helix 15 based on the equivalent structure in *Dickeya dadantii* (PDB 4K0U[31]) was fitted as a rigid body (Figure 4C). Compared to AspS[7,32], the position of PulS differs by a 12° azimuthal rotation around the secretin long axis (Figure 4B). Loop 38 between S-domain helix 14 and helix 15

bound to PulS constitutes the lone contact point between the secretin core and PulS (Figure 4C). No additional contacts were observed in contrast to AspS-GspD where the secretin core helix α11 forms extensive secondary contact with the pilotin[7]. The lack of equivalent secondary contacts between PulS and PulD likely accounts for the apparent flexibility between these proteins and may be a distinguishing feature between the structurally discrete *Klebsiella*-type and *Vibrio*-type pilotins.

PulC links the OMC to the inner membrane AP

Whilst the PulC HR domain binds to the base of the secretin, its N-terminus is located within the inner membrane AP[11] so that PulC is predicted to span the periplasm and link the AP and OMC. To verify the presence and positioning of the PulC N-terminus within the AP, a hexahistidine tag was inserted after aa 61 where PulC was predicted to exit the inner membrane and enter the periplasm. Ni-NTA gold labelling showed beads localize exclusively to the AP and not the OMC (Figure 6A and Figure S4). Given the PulC HR domains bind to the base of the secretin, PulC therefore spans the periplasmic gap between the OMC and AP (Figure 1C).

PulE and PulL form a flexible hexameric hub

Cryo-EM 2D class averages of the $Pul_{CDELMNS}$ complex revealed the ultrastructure of the AP positioned beneath the OMC. A 20-22 nm outer ring is coupled to a 10-12 nm inner ring by six radial linkers (Figure 6B). Focused alignments of the AP where the OMC was masked out show the outer ring to be comprised of weakly associating non-contiguous globular densities with the inner ring exhibiting overall C6 symmetry (Figure 6C and Figure S5A). This concentric ring structure is highly flexible and represents the preferred single orientation of the AP so that 3D structure determination was impeded. The

addition of non-hydrolysable ATP analogues made no obvious change to the $Pul_{CDELMNS}$ complex under conditions tested. In order to dissect the observed AP ultrastructure, a sub-assembly constituting PulE, PulL, PulM, and PulN (Pul_{ELMN}) was purified by 2-step affinity chromatography. GraFix [33] stabilized the complex and reduced significant particle heterogeneity. PulN bound weakly to the complex and only trace quantities were observed by SDS-PAGE within the Pul_{ELMN} complex after GraFix (Figure S5B). The resultant Pul_{ELM} complex was analyzed by cryo and negative stain EM on continuous carbon film (Figure 6D and Figure S5B-C). Under vitreous conditions, Pul_{ELM} yielded a preferred orientation concentric ring structure that was similar to the AP within the $Pul_{CDELMNS}$ complex, with equivalent dimensions and overall C6 symmetry. By negative stain, the same ultrastructure was observed although the sample was compacted so that the outer and inner rings have dimensions of 16-18 nm and 8-9 nm, respectively. Compaction was likely a consequence of sample flexibility and drying effects during the negative stain procedure. Purification and GraFix stabilization of PulE alone or PulE in complex with the cytoplasmic domain of PulL[17] (aa 1-235 and termed Pul_{ELcyto}), both isolated from the membrane fraction, yielded negative stain 2D class averages similar to Pul_{ELM} with the same single preferred orientation (Figure 6G-H and Figure S6A-B). PulE therefore constitutes the bulk of the observed AP concentric ring structure. The N2E and CTE ATPase domains of PulE homologues constitute flexible rings with 11-15 nm diameter depending on crystal packing[14-16]. Our data is consistent with a model of PulE where the N2E and CTE domains constitute the loosely packed outer ring, whilst the N1E domain locates to the inner ring (Figure 6G). Such architecture is reminiscent of the hexameric Hsp104 chaperone whose N and NBD1 domains form a similar concentric ring structure with equivalent dimensions[34]. The PulE model suggests that within the Pul_{ELcyto} complex, the $PulL_{cyto}$ subunits will be located within the inner ring given $PulL_{cyto}$ and

$PulE_{N1E}$ form a 1:1 stoichiometric complex[17,18] (Figure 6H). To demonstrate this, $PulE_{N1E}$ and $PulL_{cyto}$ were purified as a complex (termed $Pul_{N1E\text{-}Lcyto}$) by 2-step affinity chromatography with 1.1:1 stoichiometry (Figure S6C). Negative stain EM 2D class averages of $Pul_{E\text{-}N1E/Lcyto}$ showed 8-9 nm diameter rings with C6 and sometimes C5 symmetry, which was consistent with the diameter and symmetry of the inner ring in the Pul_{ELcyto} and Pul_{ELM} complexes when imaged by negative stain EM (Figure 6H and Figure S6C). $PulE_{N1E}$ and $PulL_{cyto}$ therefore constitute the inner ring of the AP concentric ring structure. This finding that $PulE_{N1E}$ and $PulL_{cyto}$ form a ring is key to AP architecture as it represents a central hub for inter-connecting PulE N2E and CTE ATPase domains, PulF, and other inner membrane components such as PulM and PulC.

Discussion

Model of an assembled T2SS in an inactive state

Our results support a model where the core cytoplasmic components of the T2SS AP constitute a $Pul_{E\text{-}N1E/Lcyto}$ ring-like hub with 6:6 stoichiometry and C6 symmetry. The PulE N2E and CTE domains form a dynamic and weakly interacting ring hanging from the $Pul_{E\text{-}N1E/Lcyto}$ hub (Figure 7). This conformation is consistent with a T2SS inactive state. In some instances, C5 symmetry class averages in the $Pul_{E\text{-}N1E/Lcyto}$ hub and in the Pul_{ELM} complex were observed suggesting that the AP may be able to assemble and function using both C5 and C6 symmetries. The periplasmic components of the $Pul_{CDELMNS}$ and Pul_{ELM} complexes, which include the periplasmic domains of PulL, PulM and PulN were not resolved here likely due to relatively small size and disorder induced by the absence of membrane. Other T2SS components such as the pseudo-pilus may be essential for stabilizing these periplasmic domains. The cryo-electron tomography structure of the

closely related T4P system[35] indicates that these periplasmic components will form a ring or shaft collar that acts as a support for the pseudo-pilus.

PulC mechanism for overcoming AP-OMC symmetry mismatch

PulC spans the periplasm to recruit the OMC with the PulC N-terminus located within the inner membrane and the HR domain bound to the PulD N0 domain at the secretin base. Stoichiometry measurements show that PulC exists with a copy number of 7.5 ± 1.0, which is consistent with a model where 6 copies of PulC will be bound in equal ratio with other AP components such as PulL, PulM and PulN. The slight excess of PulC may reflect a PulC monomer-dimer equilibrium or the binding of residual PulC to one of the unoccupied PulD N0 binding sites at the base of the secretin. Overall, the stoichiometry of PulC suggests that it constitutes a punctuated cage that spans the periplasm so that substrate has the potential to gain access to the secretin lumen in positions where PulC is absent. Such a cage is reminiscent of the virB10 N-terminus which spans the periplasm in the type IV secretion system[36]. The OMC structure shows how the PulD N0 domains are organized within the secretin relative to the N1 domain ring. Strikingly, this architecture is different to the predicted arrangement of these domains from partial crystal structures and EM models (Douzi et al., 2017; Hay et al., 2017; Korotkov et al., 2009; Korotkov et al., 2011; Van der Meeren et al., 2013). Furthermore, as each PulC HR domain binds a single PulD N0 domain independently with no lateral contacts between neighbouring HR domains, symmetry mismatch between the AP and OMC may be readily overcome. The PulC PDZ domain is closely linked to the PulC HR domain and therefore represents a speculative candidate for the plug domain. Given the PulC PDZ domain has a putative role in substrate recruitment[29], its potential positioning within the secretin

lumen may provide a natural mechanism for coupling secretin gating with substrate loading.

PulF likely resides within the $Pul_{E\text{-}N1E/Lcyto}$ ring lumen

Whilst the co-crystal structure of the GspE N1E domain in complex with the GspL cytoplasmic domain has previously been described[17], how this complex was arranged within the T2SS apparatus was unknown. Here we show that the $Pul_{E\text{-}N1E/Lcyto}$ complex oligomerizes as a dynamic ring with pseudo-C6 symmetry. In support of this model, the N1 domains in the GspE homologue PilF from *Thermus thermophilus* in the closely related T4P system, have recently been shown to form a pseudo-C6 symmetry ring (Collins et al., 2018). The $Pul_{E\text{-}N1E/Lcyto}$ ring is therefore a central structure to which other AP components such as GspC, GspM and GspN will bind. Importantly, the lumen of the $Pul_{E\text{-}N1E/Lcyto}$ ring, as observed under vitreous conditions, has a 3-4 nm diameter that is sufficient to accommodate the cytoplasmic domains of the polytopic membrane protein PulF[12]. This model is consistent with PulF homologues that are known to contact PulE and PulL in other T2SS systems[13,19]. It is also compatible with the PulF homologue PilC in the T4P system, which promotes pilus assembly through a possible rotary mechanism driven by large scale conformational changes in the PilB and PilT hexameric ATPases[15,16,37,38]. For the T2SS, PulE activates the secretion system by ATP hydrolysis, which requires association of the N2E and CTE domains and formation of a closely packed hexamer[14]. During the formation of this hexamer, our data supports a model where the closer packing between the N2E and CTE domains shifts them centrally beneath the $Pul_{E\text{-}N1E/Lcyto}$ ring[17] so that direct contact with PulF is facilitated (Figure 7). Such an arrangement is broadly consistent with the ultrastructure of the T4P system with PilB or PilT bound, as described by cryo-electron tomography[35]. However, the T4P system may

not form the T2SS-like inactive state as described here (Figure 7) as the equivalent PulE N1E domain is absent in T4P PilT, and interaction between PilB and the GspL homologue PilM may be indirect via PilC[39,40]. Overall, our results reveal the ultrastructure of an assembled T2SS and show the core architecture to be different to other known secretion systems (Figure S7).

Acknowledgements

We thank eBIC for cryo-EM data collection support particularly Kyle Dent and Yuriy Chaban. Tillmann Pape and Paul Simpson for in-house EM support. Francesca Gubellini for gold labelling advice. Arjen Jakobi and Carsten Sachse for LocScale support. This work was funded by a Wellcome Trust Career Development Fellowship Enhancement Award (200074/Z/15/Z) to H.L.

Author Contributions

A.C and H.L designed experiments. H.L initially cloned and purified the $Pul_{CDELMNS}$ complex. A.C and H.L cloned and purified proteins, collected and processed data, and solved structure. A.C built the OMC structure. A.C and H.L. determined stoichiometry. A.C. undertook gold labelling. H.L wrote the paper with contributions from A.C.

Competing interests

The authors declare no competing interests.

Data availability

3D cryo-EM density maps produced in this study have been deposited in the Electron Microscopy Data Bank with accession code EMD-0193. Atomic coordinates have been deposited in the Protein Data Bank (PDB) under accession code 6HCG.

References

1 Cianciotto, N. P. & White, R. C. Expanding role of type II secretion in bacterial pathogenesis and beyond. *Infect Immun* **85**, e00014-17 (2017).

2 Nivaskumar, M. & Francetic, O. Type II secretion system: a magic beanstalk or a protein escalator. *Biochim Biophys Acta* **1843**, 1568-1577 (2014).

3 Korotkov, K. V., Sandkvist, M. & Hol, W. G. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat Rev Microbiol* **10**, 336-351 (2012).

4 McLaughlin, L. S., Haft, R. J. F. & Forest, K. T. Structural insights into the Type II secretion nanomachine. *Curr Opin Struc Biol* **22**, 208-216 (2012).

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

6 Daefler, S., Guilvout, I., Hardie, K. R., Pugsley, A. P. & Russel, M. The C-terminal domain of the secretin PulD contains the binding site for its cognate chaperone, PulS, and confers PulS dependence on pIVf1 function. *Mol Microbiol* **24**, 465-475 (1997).

7 Yin, M., Yan, Z. F. & Li, X. M. Structural insight into the assembly of the type II secretion system pilotin-secretin complex from enterotoxigenic *Escherichia coli*. *Nat Microbiol* **3**, 581-587 (2018).

8 Hay, I. D., Belousoff, M. J., Dunstan, R. A., Bamert, R. S. & Lithgow, T. Structure and membrane topography of the *Vibrio*-type secretin complex from the type 2 secretion system of enteropathogenic *Escherichia coli*. *Journal of Bacteriology* **200**, 00521-17 (2018).

9 Hay, I. D., Belousoff, M. J. & Lithgow, T. Structural basis of type 2 secretion system engagement between the inner and outer bacterial membranes. *Mbio* **8**, 01344-17 (2017).

10 Yan, Z. F., Yin, M., Xu, D. D., Zhu, Y. Q. & Li, X. M. Structural insights into the secretin translocation channel in the type II secretion system. *Nat Struct Mol Biol* **24**, 177-183 (2017).

11 Lallemand, M. *et al.* Dynamic interplay between the periplasmic and transmembrane domains of GspL and GspM in the type II secretion system. *Plos One* **8**, e79562 (2013).

12 Abendroth, J. *et al.* The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*. *J Struct Biol* **166**, 303-315 (2009).

13 Py, B., Loiseau, L. & Barras, F. An inner membrane platform in the type II secretion machinery of Gram-negative bacteria. *Embo Rep* **2**, 244-248 (2001).

14 Lu, C. *et al.* Hexamers of the type II secretion atpase gspe from *Vibrio cholerae* with increased atpase activity. *Structure* **21**, 1707-1717 (2013).

15 Mancl, J. M., Black, W. P., Robinson, H., Yang, Z. M. & Schubot, F. D. Crystal Structure of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from Thermus thermophilus. *Structure* **24**, 1886-1897, doi:10.1016/j.str.2016.08.010 (2016).
16 Satyshur, K. A. *et al.* Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility. *Structure* **15**, 363-376, doi:10.1016/j.str.2007.01.018 (2007).
17 Lu, C., Korotkov, K. V. & Hol, W. G. J. Crystal structure of the full-length ATPase GspE from the *Vibrio vulnificus* type II secretion system in complex with the cytoplasmic domain of GspL. *J Struct Biol* **187**, 223-235 (2014).
18 Abendroth, J., Murphy, P., Sandkvist, M., Bagdasarian, M. & Holl, W. G. J. The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of *Vibrio cholerae*. *J Mol Biol* **348**, 845-855 (2005).
19 Arts, J. *et al.* Interaction domains in the *Pseudomonas aeruginosa* type II secretory apparatus component XcpS (GspF). *Microbiology* **153**, 1582-1592 (2007).
20 Lopez-Castilla, A. *et al.* Structure of the calcium-dependent type 2 secretion pseudopilus. *Nat Microbiol* **2**, 1686-1695 (2017).
21 Sandkvist, M. Biology of type II secretion. *Mol Microbiol* **40**, 271-283 (2001).
22 Shevchik, V. E., Robert-Baudouy, J. & Condemine, G. Specific interaction between OutD, an *Erwinia chrysanthemi* outer membrane protein of the general secretory pathway, and secreted proteins. *Embo Journal* **16**, 3007-3016 (1997).
23 Korotkov, K. V. *et al.* Structural and functional studies on the interaction of GspC and GspD in the type II secretion system. *Plos Pathogens* **7**, e1002228 (2011).
24 Korotkov, K. V., Pardon, E., Steyaert, J. & Hol, W. G. J. Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. *Structure* **17**, 255-265, doi:10.1016/j.str.2008.11.011 (2009).
25 Van der Meeren, R. *et al.* New insights into the assembly of bacterial secretins: structural studies of the periplasmic domain of XcpQ from *Pseudomonas aeruginosa*. *J Biol Chem* **288**, 1214-1225 (2013).
26 Douzi, B. *et al.* Unraveling the self-assembly of the *Pseudomonas aeruginosa* XcpQ secretin periplasmic domain provides new molecular insights into type II secretion system secreton architecture and dynamics. *Mbio* **8**, e01185-17 (2017).
27 Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol* **579**, 125-157 (2016).
28 Wang, X. H. *et al.* Cysteine scanning mutagenesis and disulfide mapping analysis of arrangement of GspC and GspD protomers within the type 2 secretion system. *J Biol Chem* **287**, 19082-19093 (2012).
29 Korotkov, K. V., Krumm, B., Bagdasarian, M. & Hol, W. G. J. Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from *Vibrio cholerae*. *J Mol Biol* **363**, 311-321 (2006).
30 Pineau, C. *et al.* Substrate recognition by the bacterial type II secretion system: more than a simple interaction. *Mol Microbiol* **94**, 126-140 (2014).
31 Rehman, S., Gu, S., Shevchik, V. E. & Pickersgill, R. W. Anatomy of secretin binding to the *Dickeya dadantii* type II secretion system pilotin. *Acta Crystallogr D Biol Crystallogr* **69**, 1381-1386 (2013).
32 Dunstan, R. A. *et al.* Assembly of the type II secretion system such as found in *Vibrio cholerae* depends on the novel pilotin AspS. *Plos Pathog* **9**, e1003117 (2013).

33 Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* **5**, 53-55 (2008).
34 Wendler, P. *et al.* Atypical AAA+ subunit packing creates an expanded cavity for disaggregation by the protein-remodeling factor Hsp104. *Cell* **131**, 1366-1377 (2007).
35 Chang, Y. W. *et al.* Architecture of the type IVa pilus machine. *Science* **351**, aad2001 (2016).
36 Low, H. H. *et al.* Structure of a type IV secretion system. *Nature* **508**, 550-553, (2014).
37 McCallum, M., Tammam, S., Khan, A., Burrows, L. L. & Howell, P. L. The molecular mechanism of the type IVa pilus motors. *Nat Commun* **8**, 15091 (2017).
38 Misic, A. M., Satyshur, K. A. & Forest, K. T. *P. aeruginosa* PilT structures with and without nucleotide reveal a dynamic type IV pilus retraction motor. *J Mol Biol* **400**, 1011-1021 (2010).
39 Bischof, L. F., Friedrich, C., Harms, A., Sogaard-Andersen, L. & van der Does, C. The type IV pilus assembly ATPase PilB of *Myxococcus xanthus* interacts with the inner membrane platform protein PilC and the nucleotide-binding protein PilM. *J Biol Chem* **291**, 6946-6957 (2016).
40 Takhar, H. K., Kemp, K., Kim, M., Howell, P. L. & Burrows, L. L. The platform protein is essential for type IV pilus biogenesis. *J Biol Chem* **288**, 9721-9728 (2013).
41 Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *Elife* **6**, e27131 (2017).
42 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65 (2014).
43 Wang, Z. *et al.* An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *Elife* **6**, e24905 (2017).
44 Worrall, L. J. *et al.* Near-atomic-resolution cryo-EM analysis of the *Salmonella* T3S injectisome basal body. *Nature* **540**, 597-601 (2016).

Figures

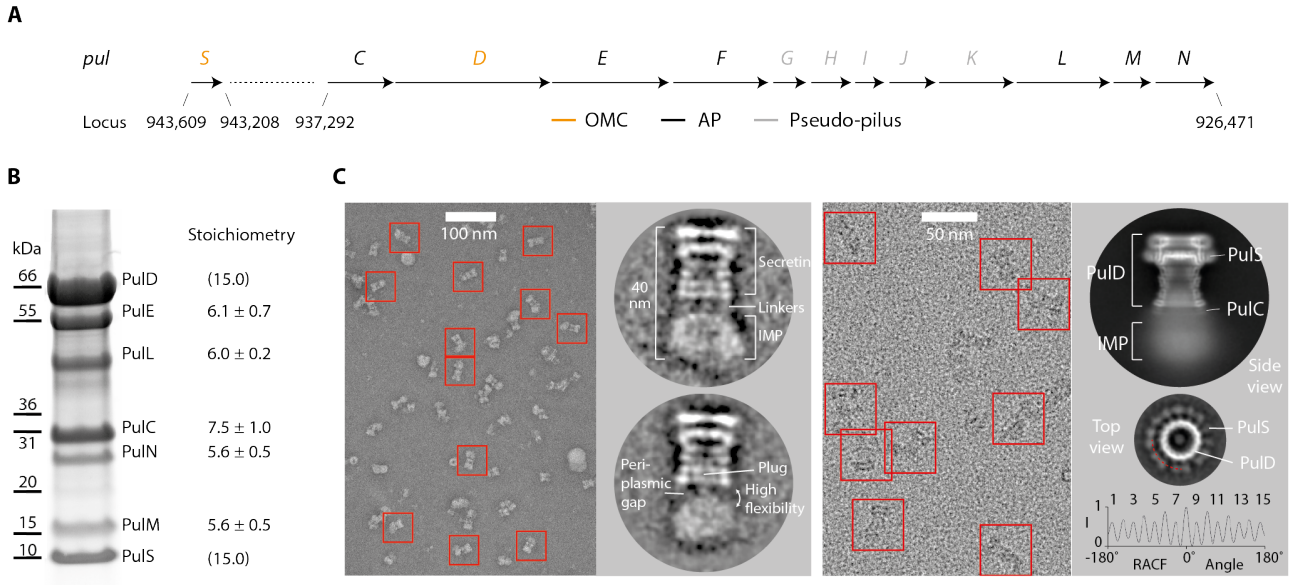

Figure 1. Purification, stoichiometry and EM analysis of the $Pul_{CDELMNS}$ complex. (A) Schematic showing the chromosomal location and gene arrangement for the *Klebsiella pneumoniae* T2SS. OMC - outer membrane complex, AP - assembly platform. **(B)** Representative Sypro stained SDS-PAGE analysis of the $Pul_{CDELMNS}$ complex with associated mean stoichiometry and standard deviation. Brackets indicate PulD and PulS stoichiometry were assigned from EM studies[7] and from the rotation auto-correlation function (RACF) in **C**. Stoichiometry measurements were determined from four independent purifications. **(C)** Left panel shows typical negative stain EM micrograph of the $Pul_{CDELMNS}$ complex with selected individual particles highlighted with a red box. Zoomed images of 2D class averages are also shown. Right panel shows typical cryo-EM micrograph of the $Pul_{CDELMNS}$ complex with associated side view 2D class average. Top view 2D class average shows PulS and PulD with C15 symmetry based on the RACF. Red dotted line indicates radial ring for RACF calculation.

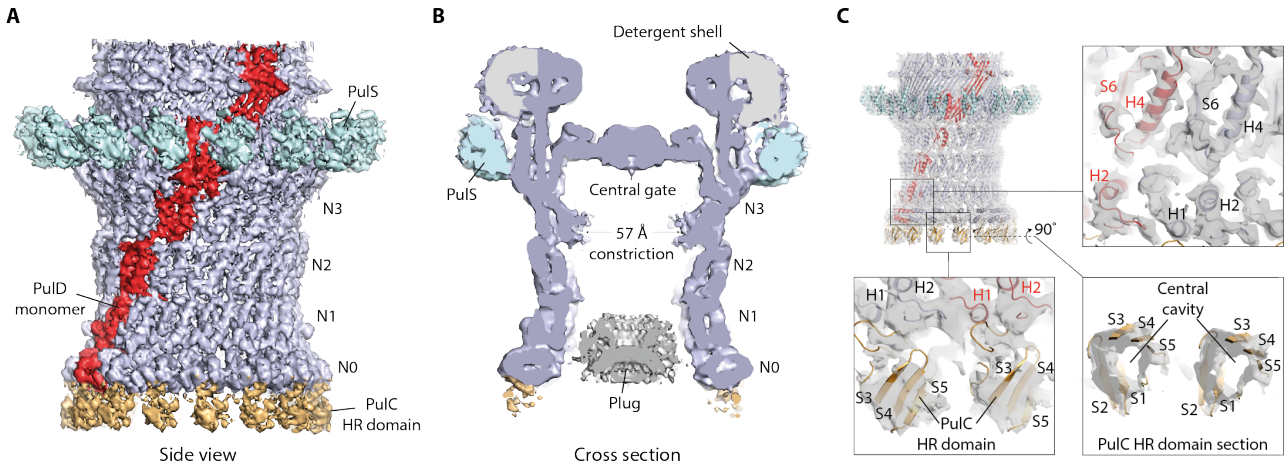

Figure 2. Cryo-EM density map of the outer membrane complex (OMC). (A) 4.3 Å resolution C15 symmetrized cryo-EM map of the OMC locally sharpened with LocScale[41] and contoured at 5.5σ. **(B)** Cross section view of the OMC revealing internal features of the PulD secretin. A plug occludes the secretin lumen. The map was unsharpened and contoured at 3σ. **(C)** Fit of molecular models within the OMC map. *De novo* model for the PulD secretin including the N1 and N0 domains (top right). A PulC HR domain homology model was rigid body fitted. The map was sharpened with LocScale and contoured at 4σ (bottom left). In this localized region of the map at 7-8 Å resolution, the central cavity of the PulC HR domain was clearly resolved (bottom right).

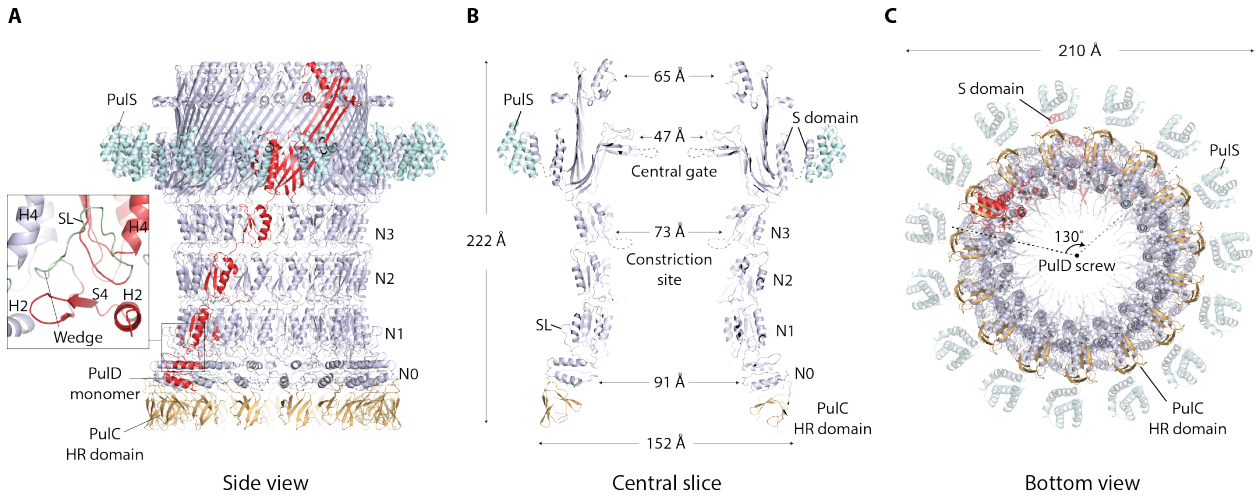

Figure 3. Structure of the outer membrane complex (OMC). (A) Cartoon representation of the OMC, which includes the C15 symmetrized secretin (light blue) with a single PulD monomer highlighted in red. PulS (cyan) and PulC HR domain (orange) homology models were fitted as rigid bodies. 15 PulC HR domains were modelled here although stoichiometry measurements indicate an occupancy of between 6-12. Zoom panel shows how loop 7 (green coil) constitutes a stabilizing loop (SL) that packs as a wedge between neighboring N1 domains. **(B)** Cross section view of the OMC structure with dimensions. **(C)** Bottom view of the OMC structure. Each PulD monomer has an azimuthal span twisting around the secretin long axis of 130˚.

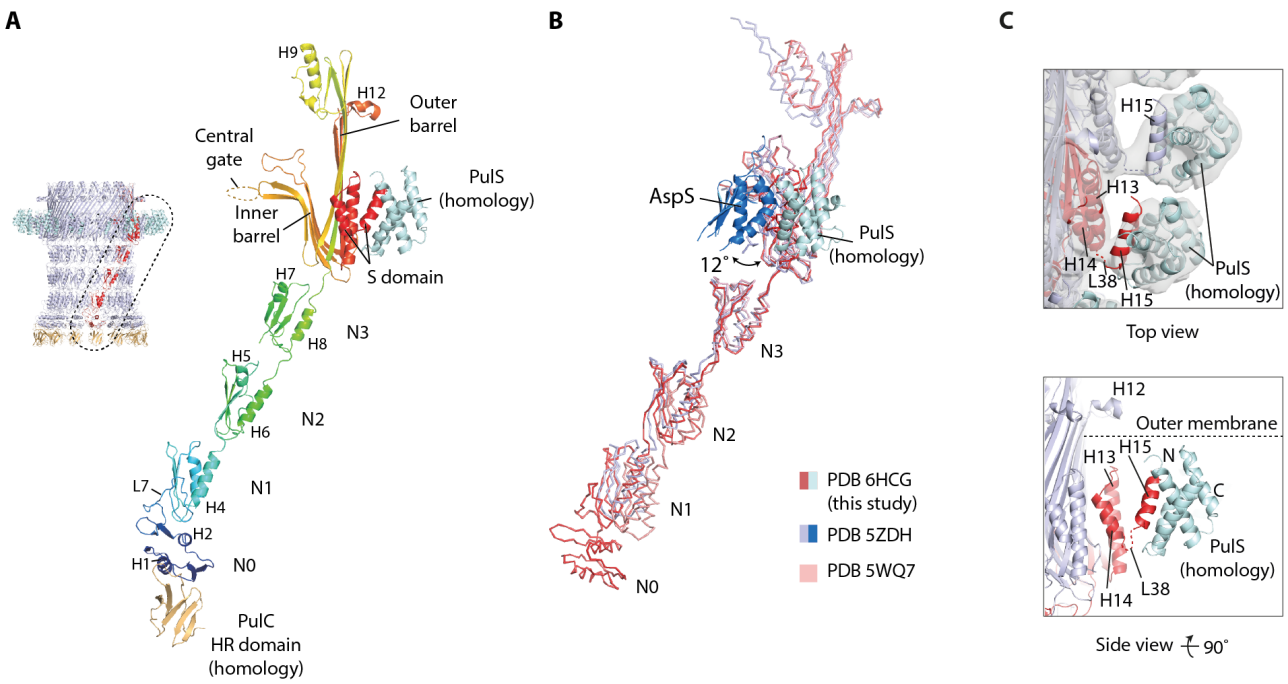

Figure 4. Structural comparison of PulD, PulS and PulC HR domain models. (A) A complex of PulD, PulS and PulC HR domain monomers extracted from the C15 OMC model (dotted region). For PulD, rainbow coloring highlights blue N-terminus through to red C-terminus. **(B)** Superposition of the PulD and PulS complex with the equivalent from enterotoxigenic *E. coli* PDB 5ZDH and *E. coli* K12 PDB 5WQ7. Compared to PulD, these PDBs differ by RMSD Cα = 3.3 Å and 3.2 Å, respectively. The relative position of AspS in PDB 5ZDH differs to PulS by a 12° azimuthal rotation around the secretin long axis. **(C)** (Top) Fit of PulS homology model with the OMC map. The map was low pass filtered to 8 Å and contoured at 5σ before rigid body fitting. (Bottom) Cartoon schematic showing the PulS model relative to the PulD S-domain (red helices). PulD helix 15 locates to a groove within PulS[31]. Flexible loop 38 represents the only attachment between the helix 15/PulS complex and the rest of the PulD monomer.

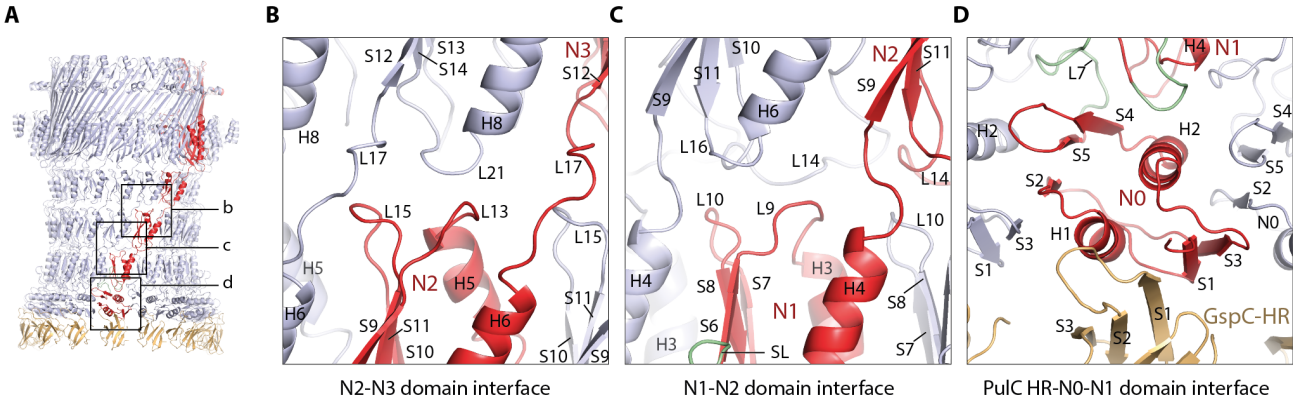

Figure 5. Structural analysis of the outer membrane complex (OMC) periplasmic domains. (A) Overview model of the OMC with PulS removed for clarity. Boxed regions are zoomed in **B-D**. **(B-D)** Cartoon representation of selected domain interfaces showing key structural details.

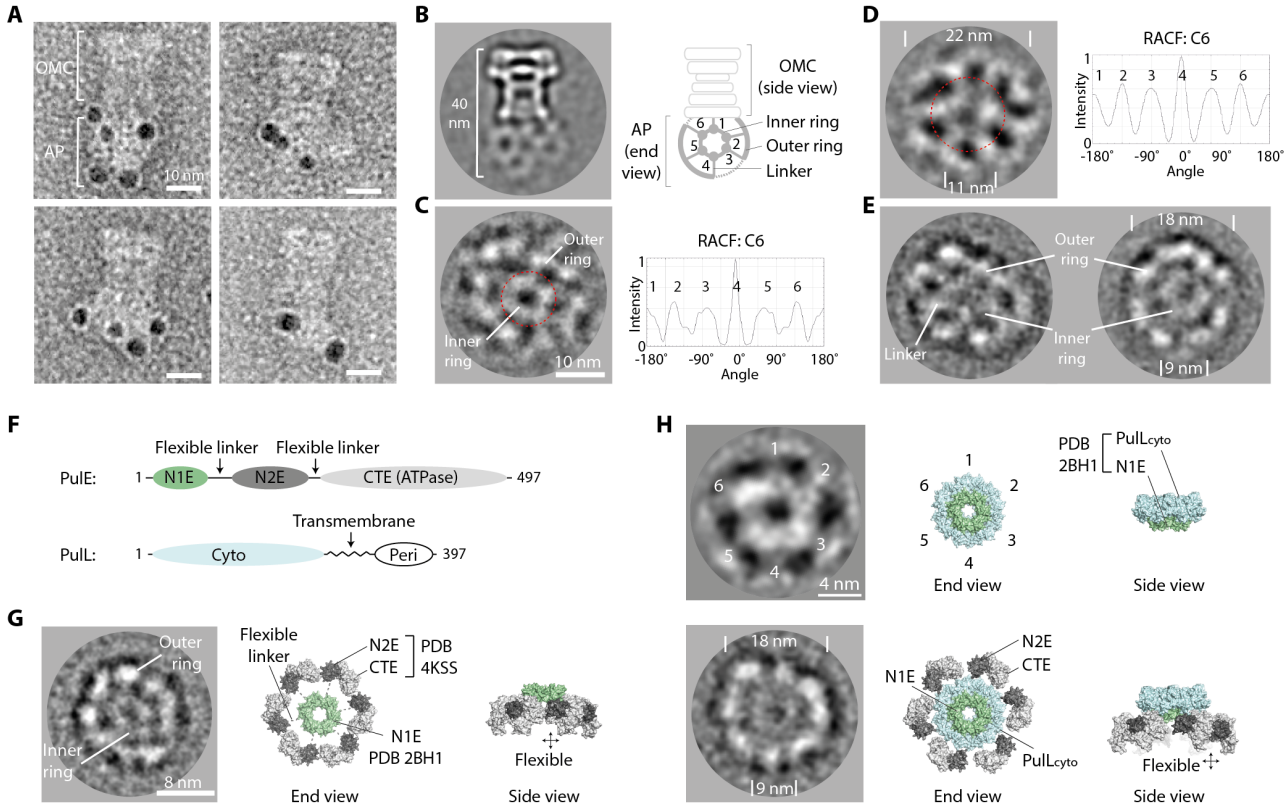

Figure 6. 2D EM and modelling of the assembly platform (AP). (A) Ni-NTA gold bead labelling of the PulC N-terminus. Beads locate exclusively to the AP. **(B)** Cryo-EM 2D class average of $Pul_{CDELMNS}$. The AP showed end views as a preferred orientation with C6 symmetry. **(C)** Cryo-EM 2D class average of $Pul_{CDELMNS}$ with the alignment and classification focused on the AP. RACF = rotation auto-correlation function calculated around dotted red radial ring. **(D)** Cryo-EM 2D class average of Pul_{ELM} showing C6 symmetry. **(E)** Negative stain EM 2D class average of Pul_{ELM}. **(F)** Schematic showing domains of PulE and PulL. **(G)** Negative stain EM 2D class average of PulE. A model of the PulE hexamer is generated where 6 copies of the N1E domain (PDB 2BH1) form the inner ring whilst 6 copies of the N2E-CTE domains (PDB 4KSS) comprise the outer ring. The PulE N2E/CTE domain outer ring is flexible with each monomer associating weakly to form a non-contiguous ring. The position of the outer ring relative to the N1E domain inner ring appears constrained by only the 44 aa flexible N1E-N2E inter-domain

linker. **(H)** Negative stain EM 2D class average of $Pul_{E\text{-}N1E/Lcyto}$ showed a ring with C6 symmetry. The $Pul_{E\text{-}N1E/Lcyto}$ ring model is based on 6 copies of PDB 2BH1. The relative orientation of each $Pul_{E\text{-}N1E/Lcyto}$ subunit within the ring is unknown (Top). Negative stain EM 2D class average of Pul_{ELcyto}. The Pul_{ELcyto} model constitutes the $Pul_{E\text{-}N1E/Lcyto}$ ring decorated by 6 PulE N2E/CTE domain monomers from PDB 4KSS (Bottom).

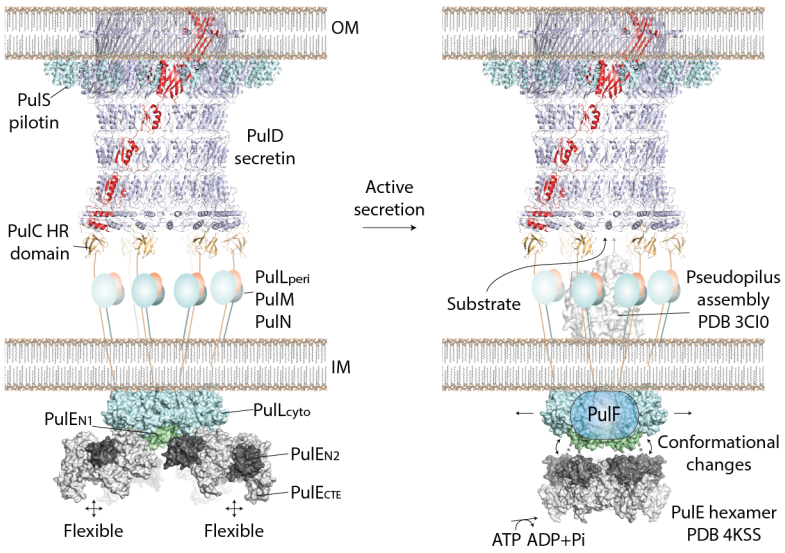

Figure 7. Stoichiometric model of an assembled T2SS. (Left) The $Pul_{CDELMNS}$ complex is consistent with a relaxed or inactive state. (Right) Speculative model of an active T2SS. The $Pul_{E\text{-}N1E/Lcyto}$ ring incorporates PulF within its central lumen. Substrate may be loaded through the horizontal gap between PulC HR domains and the putative Pul_{LMN} periplasmic shaft collar.

Supplementary Figures

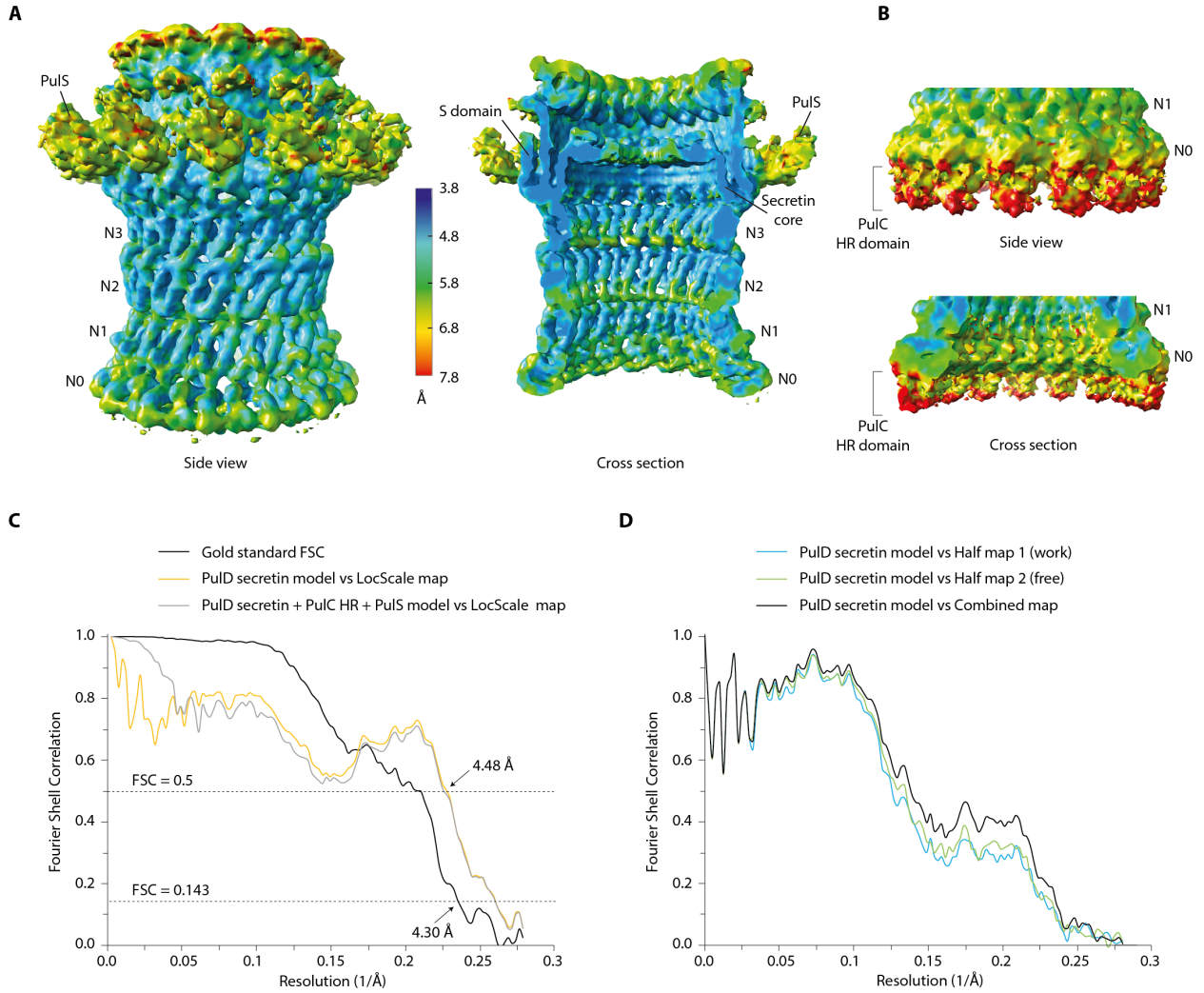

Figure S1. Outer membrane complex (OMC) local resolution map and FSC curves. (A) Unsharpened OMC map showing local resolution estimates calculated using ResMap[42] and contoured at 4.5σ. **(B)** Equivalent to **A** but contoured at 2σ so as to view the PulC HR domain densities at the secretin base. **(C)** Gold standard FSC curve of the OMC (black). FSC curves between the PulD secretin model and the LocScale[41] sharpened map (orange). FSC curves between the PulD secretin model with homology models of PulS and PulC HR domain fitted, and the LocScale sharpened map (grey). **(D)** Cross-validation for the PulD secretin model. PulD secretin model versus half map 1 (work- light blue), half map 2 (free- green), and combined (black). See Methods for further detail.

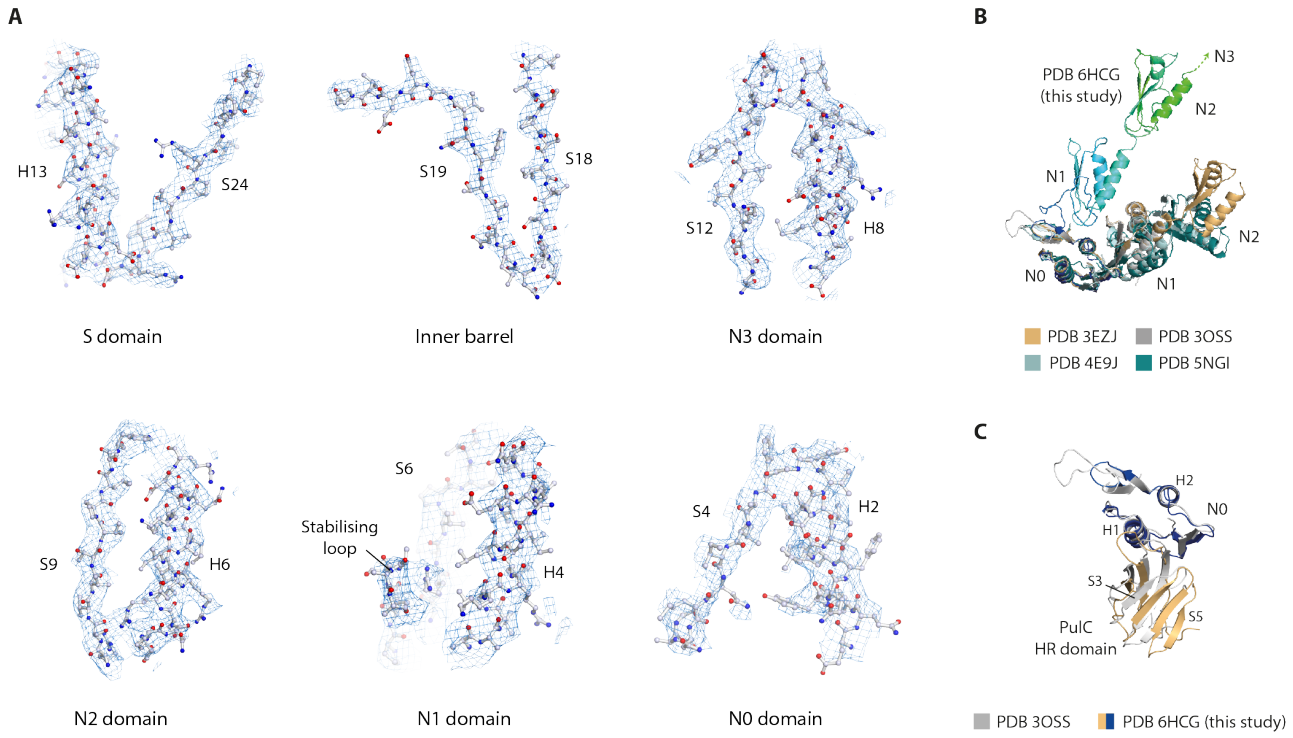

Figure S2. Map quality and PulS fit. (A) Selected regions of the OMC EM density map showing PulD secretin side chain detail and build. The map was sharpened with B factor = -142 Å^2 and contoured between 5-8σ. **(B)** Superposition of PulD N0, N1 and N2 domains with equivalent domains derived from various crystal structures including enterotoxigenic *E. coli* PDB 3EZJ and PDB 3OSS and *P. aeruginosa* PDB 4E9J and PDB 5NGI. **(C)** Superposition of PulD N0 domain in complex with PulC HR domain against the equivalent domains in the enterotoxigenic *E. coli* crystal structure PDB 3OSS. The complexes differ by RMSD Cα = 1.4 Å.

```
                 ------------------------------------------------------------------
PA               -MSQPLLRALFAPSS------RSYVPAVLLSLALGIQAAHAENSGGNAFVPAGNQQEAHW      53
EC_K12           -------MKGLNKIT--------CCL--LA---------------ALLMPCAGHAENEQY      28
KP               -------MKSLRKML--------PALLILT---------------PLL---FSPAAAEEF      27
KO               ----MIIANVIRPFS--------LMLLVFA---------------ALL---FKPAAAEEF      30
EC_H10407        MFWRDITLSVWRKKTTGLKTKKRLLPLVLA---------------AALCSSPVWAEEATF      45
VC               -------MKYWLKKSSW----------LLA---------------GSLLSTP-LAMANEF      27
                                                :                               :

                 N0 domain
                 S1      H1      S2          S3          H2
PA               TINLKDADIREFIDQISEITGETFVVDPRVKGQVSVVSKAQLSLSEVYQLFLSVMSTHGF     113
EC_K12           GANFNNADIRQFVEIVGQHLGKTILIDPSVQGTISVRSNDTFSQQEYYQFFLSILDLYGY      88
KP               SASFKGTDIQEFINTVSKNLNKTVIIDPSVRGTITVRSYDMLNEEQYYQFFLSVLDVYGF      87
KO               SASFKGTDIQEFINTVSKNLNKTVIIDPSVRGTITVRSYDMLNEEQYYQFFLSVLDVYGF      90
EC_H10407        TANFKDTDLKSFIETVGANLNKTIIMGPGVQGKVSIRTMTPLNERQYYQLFLNLLEAQGY     105
VC               SASFKGTDIQEFINIVGRNLEKTIIVDPSVRGKVDVRSFDTLNEEQYYSFFLSVLEVYGF      87
                   .::.:*::.*:: :.    :*.::.* *:* : : :   :.  : *.:**.::.  *:

                 N0 domain                                  N1 domain
                 S4      S5                                 S6          H3
PA               TVVAQGDQ-ARIVPNAEAKTEA------GGGQSAPDRLETRVIQVQQSPVSELIPLIRPL     166
EC_K12           SVITLDNGFLKVVRSANVKTSPGMIAD-SSRPGVGDELVTRIVPLENVPARDLAPLLRQM     147
KP               AVINMNNGVLKVVRAKDAKTSAVPVAS-AAAPGEGDEVVTRVVPLTNVAARDLAPLLRQL     146
KO               AVINMNNGVLKVVRSKDAKTAAVPVAS-DAAPGIGDEVVTRVVPLTNVAARDLAPLLRQL     149
EC_H10407        AVVPMENDVLKVVKSSAAKVEPLPLVGEGSDNYAGDEMVTKVVPVRNVSVRELAPILRQM     165
VC               AVVEMDNGVLKVIKSKDAKTSAIPVLS-GEERANGDEVITQVVAVKNVSVRELSPLLRQL     146
                 :*:   :   :::    .*.               *.: *::: : :  . :* *::* :

                 N1 domain                                  N2 domain
                          S7      S8      H4          S9          H5
PA               VPQYG--HLAAVPSANALIISDRSANIARIEDVIRQLDQKGSHDYSVINLRYGWVMDAAE     224
EC_K12           MDAGSVGNVVHYEPSNVLILTGRASTINKLIEVIKRVDVIGTEKQQIIHLEYASAEDLAE     207
KP               NDNAGAGSVVHYEPSNVLLMTGRAAVIKRLLTIVERVDNAGDRSVVTVPLSWASAAEVVK     206
KO               NDNAGAGSVVHYEPSNVLLMTGRAAVIKRLLTIVERVDNAGDRSVVTVPLAWASAADVVK     209
EC_H10407        IDSAGSGNVVNYDPSNVIMLTGRASVVERLTEVIQRVDHAGNRTEEVIPLDNASASEIAR     225
VC               IDNAGAGNVVHYDPANIILITGRAAVVNRLAEIIRRVDQAGDKEIEVVELNNASAAEMVR     206
                     .   :.    :* ::::.*:: : ::  ::.::*  * .    : *  . . : ..

                 N2 domain
                                 S10         S11         H6
PA               VLNNAMS----RGQAKGAAGAQVIADARTNRLIILGPPQARAKLVQLAQSLDTPTARSAN     280
EC_K12           ILNQLISESHGKSQMPALLSAKIVADKRTNSLIISGPEKARQRITSLLKSLDVEESEEGN     267
KP               LVTELN-KDTSKSALPGSMVANVVADERTNAVLVSGEPNSRQRIIAMIKQLDRQQAVQGN     265
KO               LVTELN-KDTSKSALPGSMVANVVADERTNAVLVSGEPNSRQRIIAMIKQLDRQQATQGN     268
EC_H10407        VLESLT-KNS-GENQPATLKSQIVADERTNSVIVSGDPATRDKMRRLIRRLDSEMERSGN     283
VC               IVEALN-KTTDAQNTPEFLKPKFVADERTNSILISGDPKVRERLKRLIKQLDVEMAAKGN     265
                 ::                  :.:** *** ::: *    * ::  : : **     ..*

                 N3 domain
                 S12         H7                                  S13         S14
PA               TRVIRLRHNDAKTLAETLGQISEGMKNNGGQGGEQTGGGRPSNILIRADESTNALVLLAD     340
EC_K12           TRVYYLKYAKATNLVEVLTGVSEKLKDEKGNARKPSSSGAMDNVAITADEQTNSLVITAD     327
KP               TKVIYLKYAKAADLVEVLTGISSSLQSDKQSARPVA--AIDKNIIIKAHGQTNALIVTAA     323
KO               TKVIYLKYAKASDLVEVLTGISSTMQSEKQAAKPVA--ALDKNIIIKAHGQTNALIVTAA     326
EC_H10407        SQVFYLKYSKAEDLVDVLKQVSGTLTAAKEEAEGTVGSG-REIVSIAASKHSNALIVTAP     342
VC               NRVVYLKYAKAEDLVEVLKGVSENLQAEKGTGQPTT-SK-RNEVMIAAHADTNSLVLTAP     323
                 .:*  *:: .*  *.:.*  :*  :      .         . : * *   :*:*:: *

                 N3 domain                  Secretin core domain
                          H8                 S15                 S16
PA               PDTVNALEDIVRQLDVPRAQVLVEAAIVEISGDIQDAVGVQWAINKGGMGGTKTNFANTG     400
EC_K12           QSVQEKLATVIARLDIRRAQVLVEAIIVEVQDGNGLNLGVQWANKNVGAQ---QF-TNTG     383
KP               PDVMNDLERVIAQLDIRRPQVLVEAIIAEVQDADGLNLGIQWANKNAGMT---QF-TNSG     379
KO               PDVMNDLERVIAQLDIRRPQVLVEAIIAEVQDADGLNLGIQWANKNAGMT---QF-TNSG     382
EC_H10407        QDIMQSLQSVIEQLDIRRAQVHVEALIVEVAEGSNINFGVQWASKDAGLM---QFANGTQ     399
VC               QDIMNAMLEVIGQLDIRRAQVLIEALIVEMAEGDGINLGVQWGSLESGSV---IQYGNTG     380
                  .  : :  :: :**: * ** :** *.*:       .*:**.  . *         .:
```

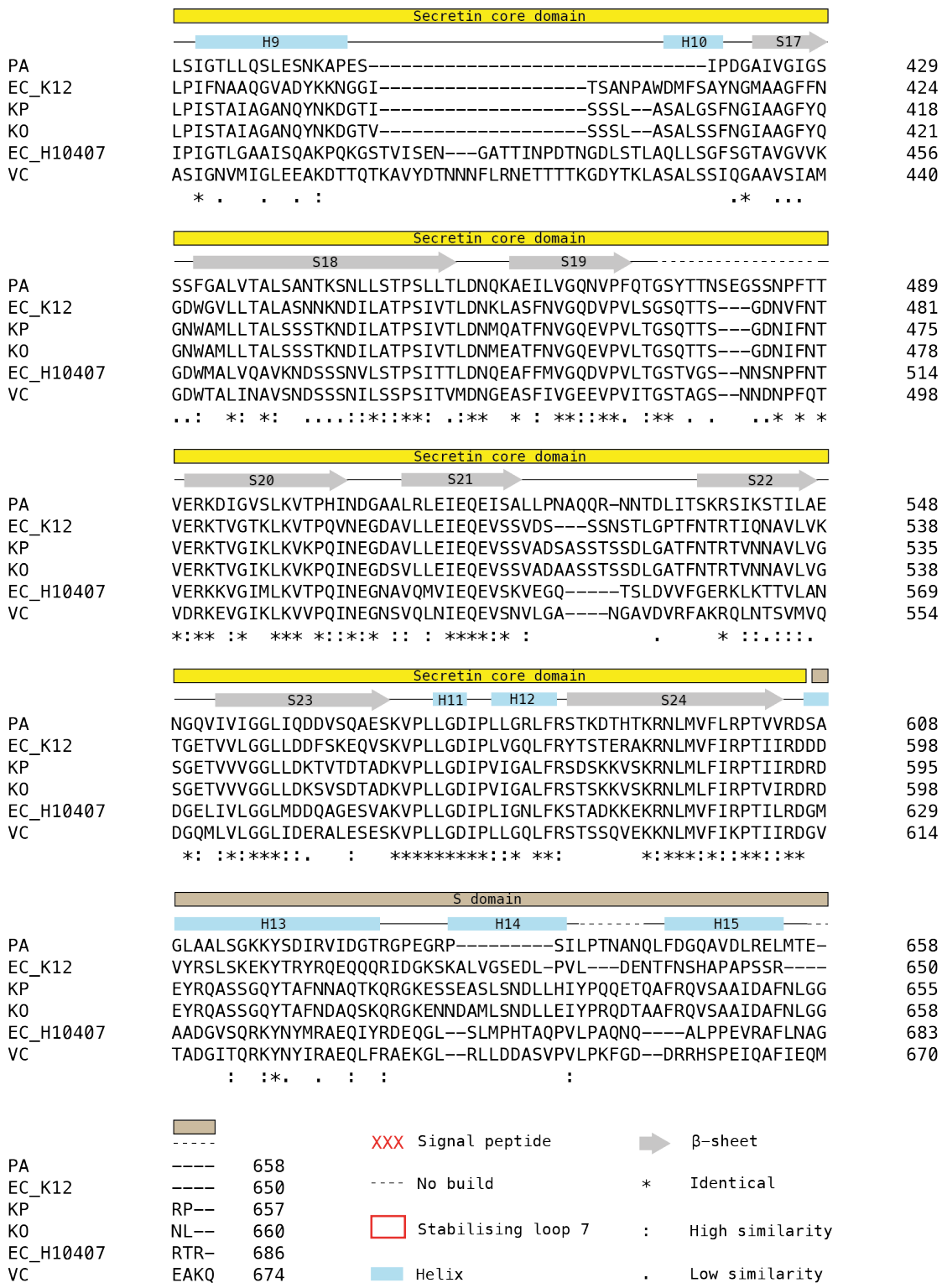

Figure S3. PulD secondary structure assignment and sequence alignment. Aligned PulD sequences include *Pseudomonas aeruginosa* (PA, Uniprot code P35818), *Escherichia coli* K12

(EC_K12, P45758), *Klebsiella pneumoniae* (KP, A0A0E1CJT4), *Klebsiella oxytoca* (KO, A0A0H3H6N4), *Escherichia coli* H10407 (EC_H10407, E3PJ86), *Vibrio cholerae* (VC, P45779).

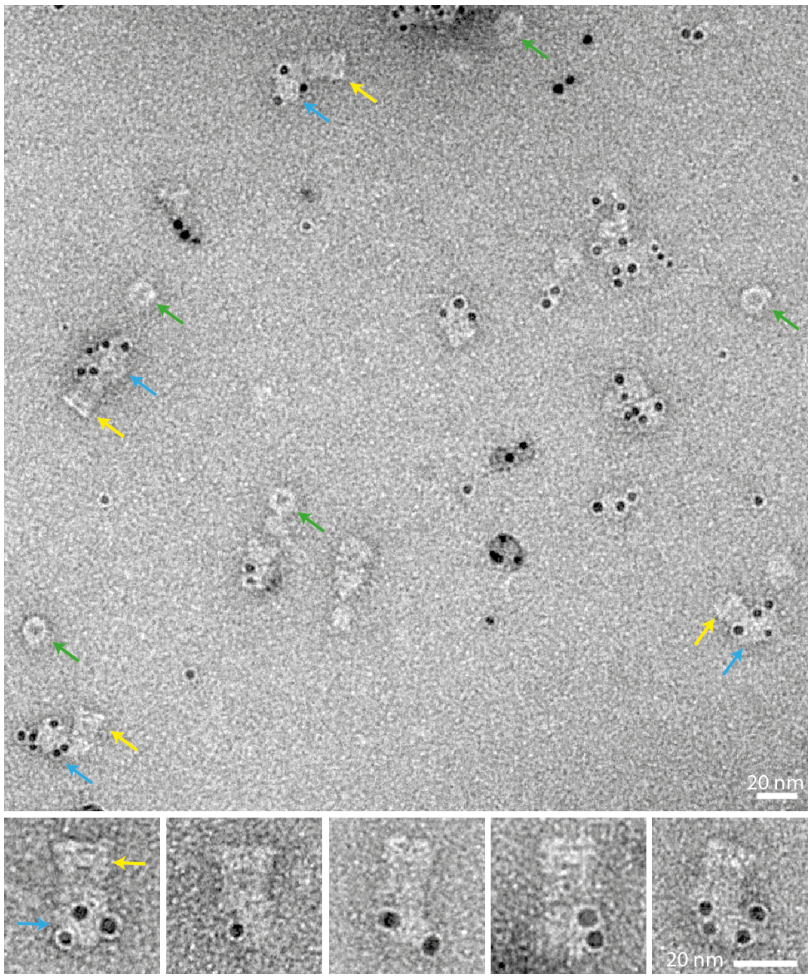

Figure S4. PulC N-terminus is located in the inner membrane assembly platform (AP). An overview negative stain EM image with a gallery of selected particles underneath. The PulC N-terminus was labelled with a hexahistidine tag at aa 61 within the $Pul_{CDELMNS}$ complex. Ni-NTA gold beads localize to the AP (blue arrow). Gold beads did not localize to the secretin when attached to the AP (side view, yellow arrow) or when dissociated from the AP (top view, green arrow).

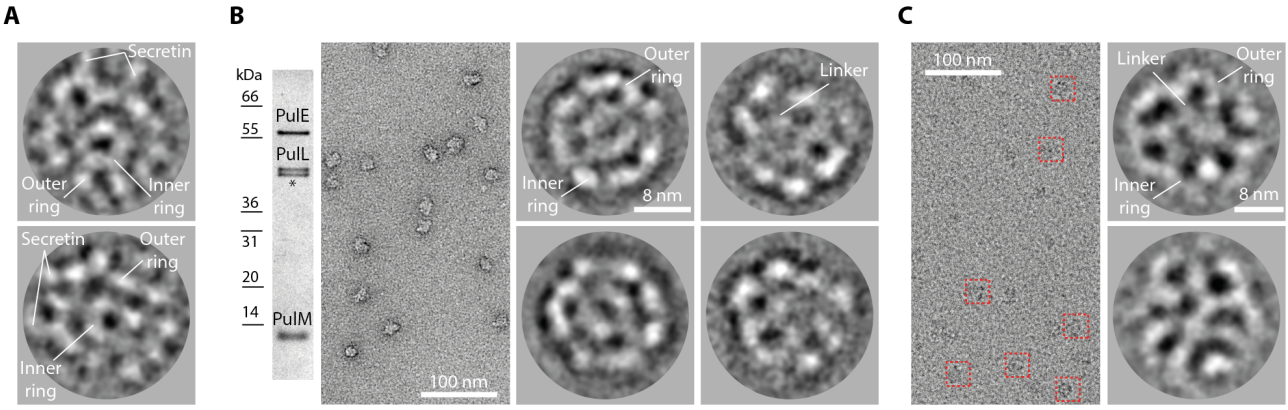

Figure S5. 2D EM analysis of the $Pul_{CDELMNS}$ assembly platform (AP) and the Pul_{ELM} complex. (A) Gallery of cryo-EM 2D class averages of the $Pul_{CDELMNS}$ complex where the alignment and classification were focused on the AP only. A single preferred orientation was observed consistent with a bottom or end view. The base of the masked-out secretin is visible in the class averages. **(B)** GraFix purified Pul_{ELM} complex and EM analysis. (Left) Silver stain SDS-PAGE gel showing purified Pul_{ELM} complex. PulN bound weakly to the complex and was observed in only trace quantities after the GraFix ultracentrifugation step. (Right) Gallery of negative stain EM 2D class averages of the Pul_{ELM} complex. The same single preferred orientation was observed as in **A**. **(C)** Gallery of cryo-EM 2D class averages of the Pul_{ELM} complex. Note how the concentric ring ultrastructure is equivalent in both negative stain (as in **B**) and under cryo conditions.

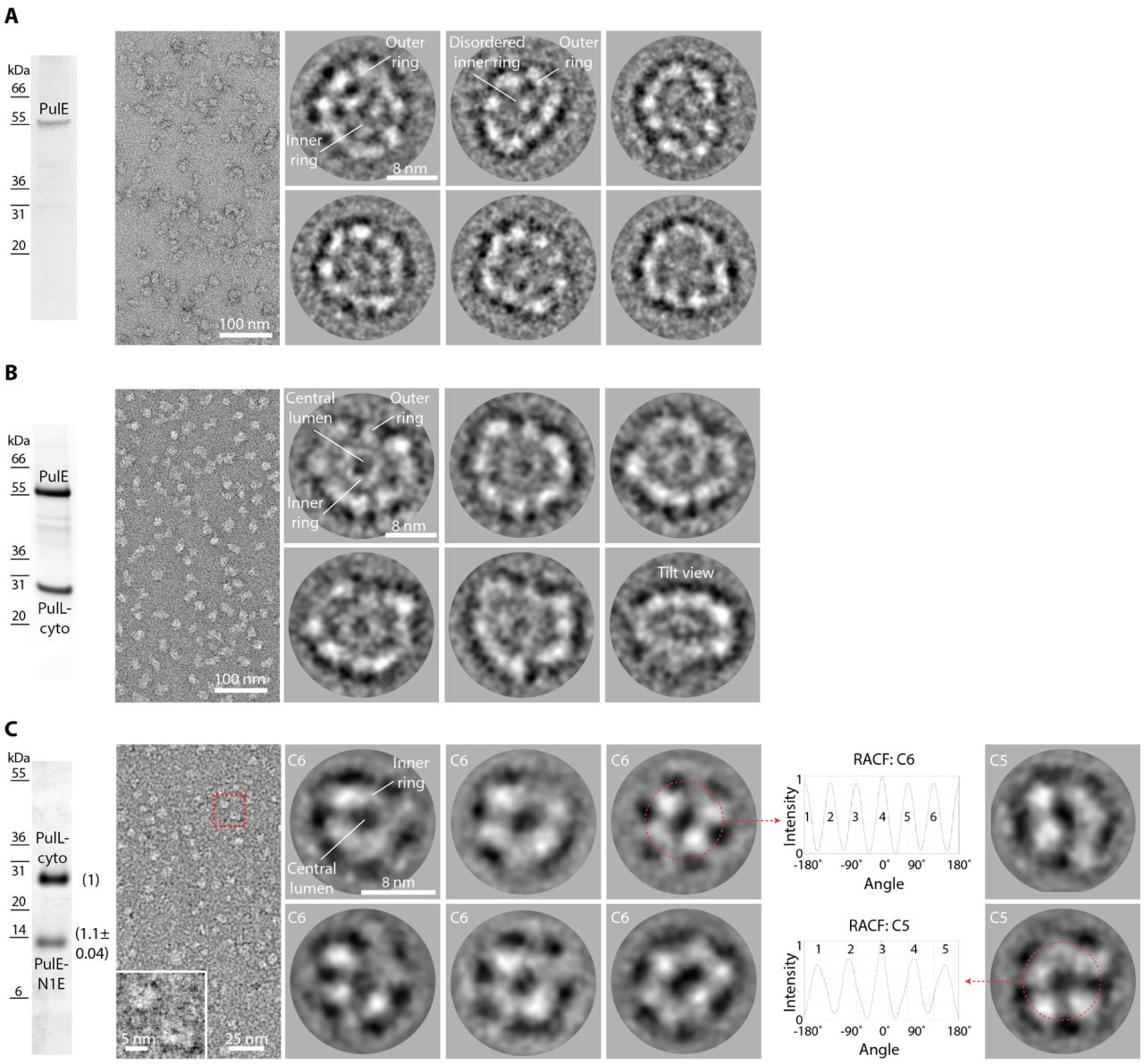

Figure S6. 2D EM analysis of assembly platform (AP) complexes. (A) GraFix purified PulE and EM analysis. (Left) SDS-PAGE of PulE purified from the membrane fraction. (Right) Negative stain EM micrograph of PulE particles with a gallery of typical 2D class averages. A single preferred orientation was observed. **(B)** GraFix purified Pul_{ELcyto} complex and EM analysis. (Left) SDS-PAGE of Pul_{ELcyto} complex purified from the membrane fraction. (Right) Negative stain EM micrograph of Pul_{ELcyto} complex particles with a gallery of typical 2D class averages in a single preferred orientation. **(C)** Purified Pul_{ELcyto} complex and EM analysis. (Left) SDS-PAGE of $Pul_{E\text{-}N1E/Lcyto}$ complex purified from the membrane fraction with associated

stoichiometry and standard deviation. (Right) Negative stain EM micrograph of Pul_{ELcyto} complex particles with a gallery of typical 2D class averages. Particles with C5 and C6 symmetries were observed suggesting both symmetries may be possible in an assembled T2SS. RACF = rotation auto-correlation function. Red dotted line indicates radial ring for RACF calculation.

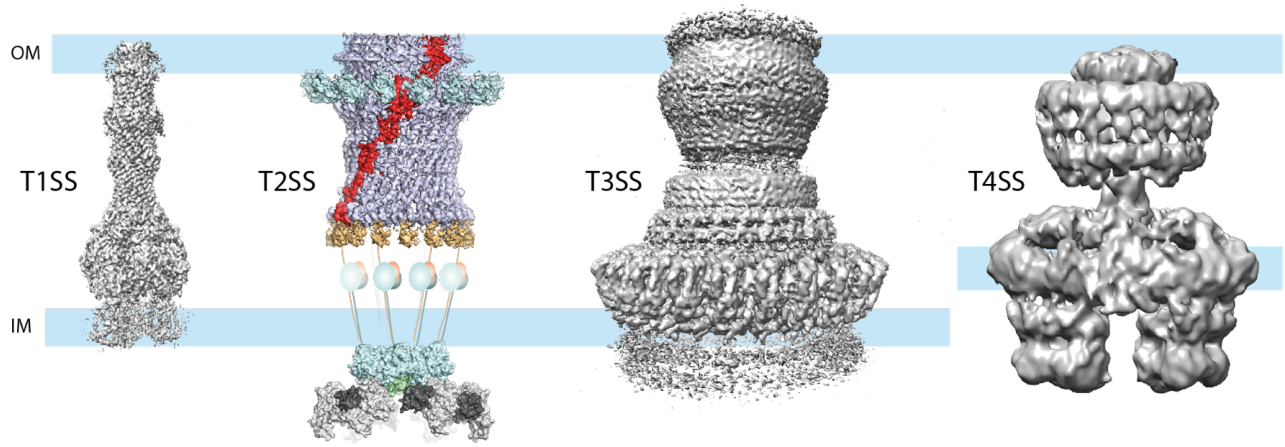

Figure S7. Comparison of bacterial type I-IV secretion systems. Substrate passes in an enclosed channel across the entire cell envelope in the type I (EMD 8636) and III (EMD 8400) secretion systems[43,44]. In contrast, in the T2SS the substrate must enter the periplasm before outer membrane translocation via the secretin channel. In the type IV secretion system[36], it is still unclear whether substrate passes across the cell envelope directly from the cytoplasm or is loaded first into the periplasm before being secreted.

Supplementary Table 1. 3D reconstruction and refinement statistics

Data collection		
Electron microscope	Titan Krios	
Voltage (kV)	300	
Pixel size (Å)	1.78	
Electron exposure ($e^-/Å^2$)	50	
Defocus range (μm)	2-4.5	
Images	3427	
3D reconstruction		
Final particles	7284	
Resolution (Å)	4.3	
FSC threshold	0.143	
B factor ($Å^2$)	−142	
Refinement		
Model	PulD	PulD-PulS-PulC
Homology model template		PulC HR- PDB 3OSS PulS- PDB 4K0U
Model composition		
Total atoms	64920	85170
Total residues	8655	11295
R.m.s. deviations		
Bond length (Å)	0.011	0.011
Angles (°)	1.4	1.3
Validation		
MolProbity score	2.66	2.58
Clashscore	38.01	36.54
Poor rotamers (%)	0	0
Ramachandran plot		
Favored (%)	88.07	90.01
Outliers (%)	0	0
B-factors/ADPs		
Minimum	36.5	36.5
Maximum	294.9	425.3
Mean	103.7	161.5
Real-space correlation		
Masked map	0.78	0.73
Unmasked map	0.76	0.71

Methods

Cloning, protein expression and purification. All clones were generated using a modified version of the Gibson isothermal DNA assembly protocol[45]. To obtain the $Pul_{CDELMNS}$ complex, the *Klebsiella pneumoniae* T2SS operon encoding genes from *pulC* to *pulO* was cloned into pASK3c vector (IBA-GO) with a StrepII tag at the C-terminus of PulE. The *pulS* gene was cloned into pCDF-duet vector with a C-terminal Flag tag. Uniprot codes for individual genes are as follows: *pulC* KPHS_08880, *pulD* KPHS_08870, *pulE* KPHS_08860, *pulF* KPHS_08850, *pulG* KPHS_08840, *pulH* KPHS_08830, *pulI* KPHS_08820, *pulJ* KPHS_08810, *pulK* KPHS_08800, *pulL* KPHS_08790, *pulM* KPHS_08780, *pulN* KPHS_08770, *pulO* KPHS_08760, *pulS* KPHS_08940. The clones were co-transformed into *Escherichia coli* C43 (DE3) electro-competent cells (Lucigen) modified here to incorporate a *pspA* gene knockout (PspA is a common contaminant induced by PulD over-expression). Cells were grown on selective LB-agarose plates with chloramphenicol (30 µg/ml) and spectinomycin (50 µg/ml). 2xYT media was inoculated and cells grown at 36°C until induction at OD_{600} = 0.5-0.6 with anhydrotetracycline (AHT, 0.2 mg/L) and isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.24 g/L). Cells were grown for ~15 hr at 19 °C and processed immediately. Pellets were re-suspended in ice-cold buffer 50 mM Tris-HCl pH 7.5, treated with DNase I, lysozyme and sonicated on ice. The lysate was clarified by centrifugation at 16,000g for 20 min. The membrane fraction was collected by centrifugation at 142,000g for 45 min. Membranes were mechanically homogenized and solubilized in 50 mM Hepes-NaOH pH 7.5, 150 mM NaCl, 1 % w/v DDM (Anatrace) and 5 mM EDTA at room temperature for 30-40 min. The suspension was clarified by centrifugation at 132,000g for 15 min. The supernatant was loaded onto a StrepTrap HP column (GE Healthcare) and washed with 50 mM Hepes-NaOH pH 7.5, 150 mM NaCl, 0.06 % w/v DDM and 5 mM EDTA (Buffer W) at 4 °C. All prior buffers were supplemented with

EDTA-free cOmplete protease inhibitor tablets (Roche). The protein sample was eluted in Buffer W supplemented with 2.5 mM desthiobiotin (IBA) but with protease inhibitors removed. Peak fractions were pooled, 0.05 % glutaraldehyde (Sigma- EM grade) added and incubated on ice for 10 min before quenching with 100 mM Tris-HCl pH 7.5. The sample was batch incubated with Flag resin (Sigma) for 1 hr. Flag resin was washed with Buffer W and then eluted with the same buffer supplemented with 3xFlag peptide. Peak fractions were collected and used immediately.

To obtain purified PulE, *pulE* was cloned into pASK3c vector to include an N-terminal StrepII tag and C-terminal Flag tag. The same initial purification strategy was then followed as for the $Pul_{CDELMNS}$ complex with the exception that no glutaraldehyde or Tris quenching buffer were added subsequent to elution from the Strep column. After elution from the Flag column, due to sample heterogeneity as judged by negative stain EM, GraFix[33] was undertaken. Using Beckman Ultra-Clear 4.2 ml 11x60 mm ultracentrifugation tubes 2.1 ml of 50 mM Hepes-NaOH, 150 mM NaCl, 0.06 % w/v DDM, 30 % v/v glycerol, 5 mM EDTA and 0.1 % glutaraldehyde was loaded under 2.1 ml of the equivalent but with 10 % v/v glycerol. A continuous gradient was made using a BioComp Gradient Master cycle set for 66 seconds at 83° tilt and 22 rpm. The PulE sample was loaded and spun at 71,000g for 16 hr at 4°C using a Beckman Ti 60.1 swing rotor. 150 µl fractions were collected manually and analysed.

To obtain the $Pul_{E\text{-}N1E/Lcyto}$ or the Pul_{ELcyto} complex, the N1E domain comprising aa 1-108 from *pulE*, or the full-length gene, were cloned into pASK3c vector to include a C-terminal StrepII tag. For Pul_{Lcyto}, aa 1-235 relating to the cytoplasmic domain of PulL were cloned into pCDF-duet vector with a C-terminal Flag tag. To obtain the Pul_{ELM} complex, the full-length *pulE* gene was cloned into pASK3c vector to include an N-terminal StrepII tag. The *pulL*, *pulM* and *pulN* region of the T2SS operon was cloned into pCDF-duet vector

with a PulN C-terminal Flag tag. Desired combinations of vectors were co-transformed, and the same 2-step affinity chromatography followed by GraFix purification strategy was then followed for all these complexes as for PulE.

Gold labelling. This was performed on the $Pul_{CDELMNS}$ complex modified to include a hexahistidine tag within PulC after aa 61. Purification was the same as for the $Pul_{CDELMNS}$ complex but without the addition of fixative. 5 nm Ni-NTA-Nanogold (Nanoprobes) pre-washed in Buffer W was added to 10 µl of the protein sample and incubated for 30 min at 4 °C. A homemade continuous carbon grid was deposited on the 10 µl sample for 3 min, blotted and washed 2 times in Buffer W supplemented with 10 mM imidazole, then 2 times in Buffer W before being stained with 3 drops of 2 % uranyl acetate.

Stoichiometry determination. Purified $Pul_{CDELMNS}$ complex from 4 independent purifications was extracted using phenol to disrupt PulD multimerization[46]. The sample was precipitated with an equal volume of ice cold phenol, vortexed and then 4 volumes of ice cold acetone were added. The mixture was kept overnight at -20 °C. The precipitate was pelleted in a bench-top centrifuge at 14,000 g for 30 min at 4 °C, washed once with ice cold acetone, dried under vacuum, resuspended in 50 mM Tris pH 7.5, 8 M urea, 2% CHAPS, 0.1% Triton X-100, 10 mM DTT, and analyzed by SDS-PAGE and Sypro Ruby stain. The stoichiometry was calculated by band integration using Image Lab software (BioRad) having normalized all intensities against the known stoichiometry of PulS. Despite phenol treatment and subsequent urea treatment, a proportion of PulD remained multimeric and so failed to transition through the gel properly. PulD was therefore under-represented relative to PulS and other components within the complex. PulD stoichiometry was assigned from EM studies undertaken in this work and carried out previously[7,10]. To determine the stoichiometry of the $Pul_{E\text{-}N1E/Lcyto}$ complex, gels from 2 independent

experiments with duplicate lanes run were stained with Simply Blue (Invitrogen) and then quantified as above.

Electron microscopy sample preparation and data collection. For outer membrane complex (OMC) structure determination, 4 µl of purified $Pul_{CDELMNS}$ complex solution was incubated for 30 seconds on glow discharged homemade continuous thin carbon grids before vitrification in liquid ethane using a Vitrobot Mark IV (FEI). Data was collected at 300 kV on a Titan Krios (M02 beamline at eBIC Diamond, UK) equipped with a Gatan Quantum K2 Summit detector. Images were acquired at a magnification of 28,090 yielding 1.78 Å/pixel using EPU software. Images were dose-weighted over 40 frames with 12 second exposures. Total dose was ~50 e/Å^2.

All other cryo and negative stain datasets were collected in-house at 200 kV on a Tecnai F20 microscope equipped with Falcon II direct electron detector. For cryo data, $Pul_{CDELMNS}$ complex (Figure 6B-C and Figure S5A) was vitrified as described above. Other GraFix treated samples required glycerol removal so that 4 µl of sample was loaded onto the glow discharged continuous carbon EM grid and after 1 min incubation was washed 4 times in Buffer W before plunge freezing. Images were acquired at a magnification of 90,909 yielding 1.65 Å/pixel using EPU software. Images were collected over 54 frames with 3 second exposures. Total dose was ~50 e/Å^2. For negative stain data, 4 µl of $Pul_{CDELMNS}$ complex was loaded onto the glow discharged continuous carbon EM grid, after 40 seconds the grid was washed with 3 drops of distilled water and stained with 3 drops of 2 % uranyl acetate. Other GraFix treated samples were similarly incubated on EM grids, washed iteratively with 4 drops of 15 µl Buffer W and then negatively stained as above. Images were acquired at a magnification of 90,909 using EPU software. Single frames were collected with 1 second exposure and total dose ~15 e/Å^2.

Image processing. For OMC structure determination, individual movie image frames were aligned with MotionCor2 [47] and the contrast transfer function estimated using Gctf 1.06[48]. Low quality images were discarded and 3427 micrographs used for subsequent reconstruction in Relion 2.1[49]. Initial manual particle picking was focused on the OMC/secretin region of the $Pul_{CDELMNS}$ complex. For particle extraction a box and mask diameter were chosen so that contributions from the inner membrane assembly platform (AP) were excluded. In this way, low-resolution 2D class averages of just the OMC were used as a template for auto-picking. OMC side views only were prevalent in this dataset, which provided a sufficiently even equatorial band distribution for a reliable reconstruction[50]. Low quality particles were removed by 4 rounds of 2D classification resulting in a stack of 36,240 particles. A single round of 3D classification was undertaken generating 10 classes. The *Vibrio cholerae* GspD reconstruction EMDB-1763 was used as an initial model filtered to 40 Å[51]. C15 symmetry was applied based on top views of the OMC (obtained in an alternative $Pul_{CDELMNS}$ complex purification) and an unambiguous 15 peaks observed from the rotation auto-correlation function calculation (Figure 1C). A single class containing 7284 particles was used for the final refinement, which attained 4.4 Å resolution. Post-processing yielded 4.3 Å resolution with an auto-estimated B-factor [52] of -142 $Å^2$ applied to sharpen the final 3D map for model building. A locally sharpened map was also generated using LocScale[41] once initial models were built. Further particle polishing and 3D refinement did not yield a marked increase in resolution. Resolutions reported are based on gold standard Fourier shell correlations (FSC) = 0.143. Statistics for data collection and 3D refinement are included in Supplementary Table 1. Local 3D refinements with various particle subtraction strategies focusing on PulS or PulC HR domain did not markedly improve resolution.

To generate all other 2D class averages both in cryo and negative stain conditions as for PulE, $Pul_{E\text{-}N1E/Lcyto}$, Pul_{ELcyto}, Pul_{ELM}, and $Pul_{CDELMNS}$ complexes, the following protocol was followed. Working initially within Relion 2.0 or 2.1, Gctf 1.06 was used for estimating the CTF. Negative stain micrographs were phase flipped. Low quality micrographs were discarded. Initial 2D class averages were generated from a manually picked stack to yield templates for autopicking. 3-4 rounds of 2D classification were then undertaken to remove low quality particles. Using the 'relion_stack_create' a cleaned image stack was generated for further processing. For cryo-EM images the stack was created from phase flipped particles. In Imagic[53], particles were normalised, band pass filtered, centred and subjected to reference-free MSA and classified. The best classes, typically judged by lowest variance, were used as references for multi-reference alignment (MRA) in Spider[54] followed by MSA and classification in Imagic. This cycle of MRA and MSA was typically iterated a further 2-3 times. For PulE, 805 micrographs yielded 92,507 extracted particles and a cleaned stack of 7061 particles for subsequent MSA and MRA. Note that in vitro PulE self-assembled loosely to yield particles with variable size and copy number. This combined with the inherent flexibility between N1E, N2E and CTE PulE domains yielded heterogeneous particle datasets. GraFix significantly improved the quality of the particle and facilitated particle averaging. Generally, for PulE, $Pul_{E\text{-}N1E/Lcyto}$, Pul_{ELcyto} and $Pul_{ELM(N)}$ complexes, relatively large datasets were therefore required to yield 2D class averages of sufficient quality to robustly determine 2D ultrastructure. Due to PulE, particles from all these datasets usually adhere to continuous carbon EM grids with a single preferred orientation consistent with a bottom or end view. For $Pul_{E\text{-}N1E/Lcyto}$, 363 micrographs yielded 108,324 extracted particles and a cleaned stack of 26,341 particles. For Pul_{ELcyto}, 333 micrographs yielded 101,245 extracted particles and a cleaned stack of 49,534 particles. For negative stain $Pul_{ELM(N)}$ data, 2746 micrographs yielded 81,727 extracted

particles and a cleaned stack of 63,014 particles. For cryo $Pul_{ELM(N)}$ data, 467 micrographs yielded 4937 hand-picked particles, and all particles were used for subsequent MSA and MRA. For $Pul_{CDELMNS}$, 3059 micrographs yielded 89,381 extracted particles and a cleaned stack of 66,855 particles for subsequent MSA and MRA (Figure 6B). From this same dataset, 1148 micrographs were then used to hand-pick 4020 AP-focused particles (Figure 6C and Figure S5A). Rotation auto-correlation functions were calculated using Imagic.

Model building. A PulD homology model was generated with I-Tasser[55] using the *E. coli* GspD PDB 5WQ7 as a template. This yielded a starting model for the secretin core, and N3, N2 and N1 domains. *E. coli* GspD shares 57 % sequence identity with PulD. A homology model for the N0 domain was generated using Swissmodel[56] and the relevant part of 3OSS as template (51 % sequence identity). Homology models were rigid body fitted into the map using Chimera[57] Fit in Map function. Using these models as a starting guide and the side chain detail from bulky residues to confirm sequence register, Coot[58] was used to manually build a complete model for PulD aa 27-652 excluding aa 288-303, 462-470 and 632-637. The model was further refined using real-space refinement in Phenix[59] with secondary structure, geometry and NCS restraints applied. For the low-resolution regions specific to the PulC HR domain, a homology model based on ETEC GspC HR domain PDB 3OSS (27 % sequence identity) was generated using Swissmodel. For fitting this homology model, the PDB 3OSS which includes the ETEC GspD N0 domain was first superimposed onto the PulD N0 domain. The PulC HR domain homology model was then superimposed onto the ETEC GspC HR domain from PDB 3OSS resulting in a near perfect fit within the map. The Chimera Fit in Map function was then applied to PulC HR domain resulting in a minor shift so that the PulC HR-PulD N0 domain complex has a RMSD Cα = 1.4 Å when aligned to PDB 3OSS (Figure S2C). For the low-resolution regions

specific to the PulD S-domain C-terminus (aa 638-652) in complex with the PulS pilotin, a homology model was generated using Swissmodel based on the equivalent structure from *Dickeya didantii* PDB 4K0U (>50 % sequence identity for both chains). The map was low pass filtered to 8 Å and the homology model initially fitted manually so that the PulS lipidated N-terminus orientated towards the membrane. The Chimera Fit in Map function was then used for final positioning. Cross-validations were carried out as previously described[10,60] using the auto-estimated B-factor sharpened map. Briefly, the PulD secretin model was displaced randomly by 0.2 Å and then refined against a map reconstructed from one of the independent data halves (Half map 1). FSC curves were then calculated using the resulting model and Half map 1 (FSC_{work}). FSC curves were also calculated between this same model and another reconstruction generated from the other independent data half (Half Map 2 and FSC_{free}). The similarity between FSC_{work} and FSC_{free} curves indicates an absence of overfitting within the PulD secretin model (Figure S1D). The final models were assessed using Molprobity[61] and statistics outlined in Supplementary Table 1.

45 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343-345 (2009).

46 Nouwen, N. *et al.* Secretin PulD: association with pilot PulS, structure, and ion-conducting channel formation. *Proc Natl Acad Sci USA* **96**, 8173-8177 (1999).

47 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

48 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).

49 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).

50 Orlova, S. S. Theory of three-dimensional reconstruction. 1. Conditions of a complete set of projections. *Sov phys, cryst* **20**, 312-314 (1975).

51 Reichow, S. L., Korotkov, K. V., Hol, W. G. & Gonen, T. Structure of the cholera toxin secretion channel in its closed state. *Nat Struct Mol Biol* **17**, 1226-1232 (2010).

52 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745 (2003).

53 van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J Struct Biol* **116**, 17-24 (1996).
54 Frank, J. *et al.* SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190-199 (1996).
55 Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
56 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).
57 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
58 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
59 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
60 Li, N. N. *et al.* Structure of the eukaryotic MCM complex at 3.8 angstrom. *Nature* **524**, 186-191, doi:10.1038/nature14685 (2015).
61 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).