Core 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 PulC, PulE, PulL, PulM and PulN have a relative stoichiometric ratio of 2:1:1:1:1, respectively. The PulE ATPase, PulL and PulM combine to form a flexible hexameric hub. 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.

51 **Introduction**

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

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

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

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

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

101 for the cytoplasmic components of the AP. Combined they reveal a glimpse at the core
102 ultrastructure of this cell envelope spanning nanomachine.
103
104 **Results**
105 **Purification, EM and stoichiometry of $Pul_{CDELMNS}$**
106 The T2SS from the human pathogen *K. pneumoniae* HS11286 strain comprises 13 genes
107 in a single unidirectional operon termed PulC through to PulO (Figure 1A). Note that Pul
108 and Gsp nomenclature relate to equivalent proteins in homologous T2SS systems. The
109 pilotin PulS is located in a separate position within the chromosome. These 14 genes were
110 cloned and over-expressed in *Escherichia coli*. Using affinity chromatography tags
111 positioned on the cytoplasmic ATPase PulE and the periplasmic pilotin PulS, a complex
112 containing 7 components was purified by two successive pulldowns. Glutaraldehyde
113 stabilization was included after the initial pulldown. The complex comprised PulC, PulD,
114 PulE, PulL, PulM, PulN and PulS, and is here termed $Pul_{CDELMNS}$ (Figure 1B). SDS-PAGE
115 band identification was confirmed by LC-MS/MS. It is unclear why PulF and pseudo-pilus
116 components were not co-purified with $Pul_{CDELMNS}$. Their inclusion may require intact
117 membrane for stabilization. Co-expression of PulA substrate or *pulG* knockout did not
118 promote their inclusion. Trace quantities of GspJ and GspK were identified on the gel by
119 LC-MS/MS suggesting the pseudo-pilins were at least partially expressed. Visualisation
120 of $Pul_{CDELMNS}$ by negative stain EM yielded particles ~40 nm long and 17-22 nm wide
121 (Figure 1C). The OMC PulD secretin was readily identifiable within 2D class averages.
122 Hanging beneath the OMC and separated by a 5-10 nm gap the inner membrane AP was
123 observed. Highly flexible linkers connect the OMC and AP so that these two assemblies
124 effectively constitute independent particles tethered together. The $Pul_{CDELMNS}$ complex
125 was vitrified on thin carbon film and imaged by cryo-EM (Figure 1C). 2D class averages

126 of $Pul_{CDELMNS}$ yielded well resolved side views of the OMC. All domains of PulD were
127 identifiable with additional densities observed at the base of the secretin where the PulC
128 HR domain was expected to bind to the N0 domain[22], and where PulS decorates the
129 exterior of the secretin core[7] (Figure 1C). The AP was not resolved here due to high
130 flexibility and averaging effects. In the absence of glutaraldehyde, the OMC sometimes
131 separated from the AP and yielded top views, which confirmed PulD and PulS in 1:1
132 stoichiometric ratio[7] with C15 symmetry (Figure 1C). The relative stoichiometry of the
133 inner membrane AP components were determined by SDS-PAGE densitometry and
134 quantification of fluorescent emission using both Coomassie R250 and Sypro Ruby
135 dyes[23,24] (Figure 1B). $Pul_{C:E:L:M:N}$ relative mean ratios standardized around PulE were
136 observed as 2.13:1.00:0.97:0.99:0.95, which is consistent with an overall relative ratio of
137 2:1:1:1:1 for these components. The relative ratios of PulE and PulL were tightly
138 correlated (sigma = 0.07), which is important as it reflects their known 1:1 interaction
139 [17,18] and acts as an internal control for the gel densitometry. PulC stoichiometry was at
140 least twice that of the other AP components with a relatively broad variance (sigma =
141 0.19). The 15-fold copy numbers of PulD or PulS were not used as a reference to
142 determine an overall copy number for the AP components within the $Pul_{CDELMNS}$ complex.
143 PulD remained partially multimerized despite phenol treatment so that it failed to
144 consistently enter and migrate through the gel fully. PulS quantity was significantly
145 enriched as a consequence of the PulD secretin decoupling from the AP in the absence of
146 glutaraldehyde stabilization during purification.
147

148 **PulD secretin structure determination**

149 Focused refinement of the OMC yielded a reconstruction with an overall resolution of 4.3
150 Å (Figure 2 and Figure S1). All domains of PulD were resolved with sufficient map quality

151 (Figure S2A) to build a complete model of the monomer and the secretin, excluding the
152 amino acids (aa) in loops 288-303, 462-470 and 632-637 (Figure 3 and 4A, and Figure
153 S3). The PulD fold is similar to partial *Escherichia coli* K12 and H10407 GspD models
154 (RMSD Cα = 3.2 Å and 3.3 Å) where the secretin core and N3 domain have been described,
155 along with homology modelled N2 and N1 domains[7,10] (Figure 4B). The entire secretin is
156 20 nm long with an external diameter of 15 nm at the base (Figure 2B and Figure 3). It
157 includes an occluding central gate and N3 domain constriction sites within the secretin
158 channel. It lacks a *Vibrio cholerae* cap gate[10]. The N1, N2 and N3 domains pack tightly
159 (Figure 5) with a diagonal offset of 36˚. N0 is positioned almost directly below the N1
160 domain and does not maintain the diagonal offset. The N0 fold is similar to that described
161 in multiple crystal structures[22,25,26] with a core of two helices flanked on each side by β-
162 sheets. However, its position relative to the N1 domain within the secretin is significantly
163 different to these crystal structures where crystal contacts appear to have dominated
164 domain arrangement (Figure S2B). The N0 and N1 domains are connected by loop 7,
165 which constitutes a substantial 26 aa linker. The N-terminus of loop 7 forms a wedge that
166 packs between neighboring N1 domains. Its C-terminus partially envelops the proximal
167 N1 domain whilst making additional secondary contacts with N0 domain helix 2 (Figure
168 3A and Figure 5D). The N0 domains form a tightly packed ring with alternating stacked
169 β-sheets sandwiched between helices 2 and 4 (Figure 5D). Failure to stabilize the N0
170 domain and to promote formation of the loop 7 wedge likely accounts for the previously
171 reported N0, N1 and N2 domain flexibility in other systems[7-10,27]. Overall, a single PulD
172 monomer has a radial twist around the secretin long axis of 130˚ (Figure 3C).
173
174 **PulC HR domain binds at the secretin base**

175 Hanging beneath the N0 domains in the OMC map, additional globular densities at 7-8 Å
176 resolution protrude from the secretin base (Figure 2 and Figure S1B). Focused
177 refinements[28] failed to markedly improve resolution. These densities were predicted to
178 be the PulC HR domain given its known interaction with the PulD N0 domain in
179 homologous systems[22,29]. Using the GspC HR domain and GspD N0 domain crystal
180 structure[22] as a reference, a homology model of PulC HR domain was fitted as a rigid body
181 (Figure S2C). The model closely follows the surface envelope of the map in this region,
182 with a pair of triple β-sheets opposed around a central cavity (Figure 2C). Based on the
183 quality of this fit, the PulC HR domain was assigned to each globular protrusion.
184 Importantly, each PulC HR domain binds to only a single PulD N0 domain and has no
185 contact with neighbouring PulC HR domains. Overall, the binding of the PulC HR domain
186 to the PulD N0 domain appears to be important for the correct positioning of the N0
187 domain within the secretin and the subsequent stabilization of the N1 and N2 domains.
188 Additional stabilization is derived from a plug that occludes the lumen of the secretin at
189 the level of the N0-N1 domains (Figure 2B). The plug was observed in both the 2D class
190 averages (Figure 1B) and the 3D reconstruction. The high-resolution plug ultrastructure
191 was not resolved due to a likely symmetry mismatch with the C15 averaged OMC.
192 Attempts to resolve the plug structure through refinement using lower symmetries
193 yielded reconstructions of low quality and resolution. Further studies will be required to
194 resolve the source of the plug although the PulC PDZ domain is a speculative candidate
195 given the position of the PulC HR domain at the base of the secretin.
196
197 **The PulS pilotin decorates the secretin core**
198 Decorating the outside of the secretin core proximal to the PulD S-domain in the map,
199 globular densities were observed in a position consistent with the pilotin AspS relative

200 to GspD in ETEC[7] (Figure 2). For these densities, map resolution was limited to ~7 Å
201 (Figure S1) and focused refinements[28] did not markedly improve resolution. A homology
202 model of the PulS pilotin in complex with the PulD S-domain helix 15 based on the
203 homologous structure in *Dickeya dadantii* (PDB 4K0U[30]) was fitted as a rigid body (Figure
204 4C). Compared to AspS[7,31], the position of PulS differs by a 12° radial rotation around the
205 secretin long axis (Figure 4B). Loop 38 between S-domain helix 14 and helix 15 bound
206 to PulS constitutes the lone contact point between the secretin core and PulS (Figure 4C).
207 No additional contacts were observed in contrast to AspS-GspD where the secretin core
208 helix α11 forms extensive secondary contact with the pilotin[7]. The lack of equivalent
209 secondary contacts between PulS and PulD likely accounts for the apparent flexibility
210 between these proteins and may be a distinguishing feature between the structurally
211 discrete *Klebsiella*-type and *Vibrio*-type pilotins.
212
213 **PulC links the OMC to the inner membrane AP**
214 Whilst the PulC HR domain binds to the base of the secretin, its N-terminus is located
215 within the inner membrane AP[11] so that PulC is predicted to span the periplasm and link
216 the AP and OMC. To verify the presence and positioning of the PulC N-terminus within
217 the AP, a hexahistidine tag was inserted after aa 61 where PulC was predicted to exit the
218 inner membrane and enter the periplasm. Ni-NTA gold labelling showed beads localize
219 exclusively to the AP and not the OMC (Figure 6A and Figure S4). Given the PulC HR
220 domains bind to the base of the secretin, PulC therefore spans the periplasmic gap
221 between the OMC and AP (Figure 1C).
222
223 **PulE, PulL and PulM form a flexible hexameric hub**

224 Cryo-EM 2D class averages of the $Pul_{CDELMNS}$ complex revealed the ultrastructure of the
225 AP positioned beneath the OMC. A 20-22 nm outer ring is coupled to a 10-12 nm inner
226 ring by six radial linkers (Figure 6B). Focused alignments of the AP where the OMC was
227 masked out show the outer ring to be comprised of weakly associating non-contiguous
228 globular densities (Figure 6C and Figure S5A). This concentric ring structure is highly
229 flexible and represents the preferred single orientation of the AP so that 3D structure
230 determination was impeded. The addition of non-hydrolysable ATP analogues made no
231 obvious change to the $Pul_{CDELMNS}$ complex under conditions tested. In order to dissect the
232 observed AP ultrastructure, a sub-assembly constituting PulE, PulL, PulM, and PulN
233 (Pul_{ELMN}) was purified by 2-step affinity chromatography. GraFix[32] stabilized the complex
234 and reduced particle heterogeneity. PulN bound weakly to the complex and only trace
235 quantities were observed by SDS-PAGE within the Pul_{ELMN} complex after GraFix (Figure
236 S5B). The resultant Pul_{ELM} complex was analyzed by cryo and negative stain EM on
237 continuous carbon film (Figure 6D-E and Figure S5B-C). Under vitreous conditions,
238 Pul_{ELM} yielded a preferred orientation concentric ring structure that was similar to the
239 AP within the $Pul_{CDELMNS}$ complex, with equivalent dimensions and overall C6 symmetry.
240 By negative stain, the same ultrastructure was observed although the sample was
241 compacted so that the outer and inner rings have dimensions of 16-18 nm and 8-9 nm,
242 respectively. Compaction was likely a consequence of sample flexibility and drying effects
243 during the negative stain procedure. PulE alone, and PulE in complex with the
244 cytoplasmic domain of PulL (aa 1-235) or full-length PulL were also purified from the
245 membrane fraction. However, EM analysis yielded heterogeneous and flexible particles
246 that did not further resolve AP ultrastructure under conditions tested.
247
248 **Discussion**

249 **Model of the inner membrane AP**

250 Negative stain and cryo-EM studies of the $Pul_{CDELMNS}$ complex revealed the ultrastructure
251 of the AP positioned beneath the OMC to be a highly flexible hexameric hub comprising
252 two concentric rings. In vitreous ice, the inner ring had a diameter of ~11 nm and overall
253 C6 symmetry. The outer ring constituted non-contiguous globular densities with ~22 nm
254 diameter. Similar ultrastructure was observed with purified Pul_{ELM} complex meaning just
255 these components alone are sufficient to generate the bulk of the hexameric hub
256 structure. In both systems, the single preferred orientation captured was considered to
257 represent an end view projection of the AP. In the T2SS, the *V. cholerae* PulE homologue
258 GspE forms a quasi C6 ring with diameter 11-13 nm when combined with an Hcp1
259 fusion[14]. The related T4P system is energized by the ATPases PilB and PilT, which are
260 known to operate as highly dynamic hexameric rings. PilB has dimensions of 13.5 nm and
261 9 nm when in an elliptical C2 conformation[15] whilst the PilT ring has a diameter of 11.5
262 nm[33]. For PilB and GspE these dimensions are specific to the N2E and CTE ATPase
263 domains and do not include their N-terminal N1E domains. In situ reconstruction of the
264 T4P system in *Myxococcus xanthus* indicates the position of PilB and PilT to be located
265 centrally at the base of the AP[34]. Collectively, this data is consistent with a model for the
266 $Pul_{CDELMNS}$ AP and Pul_{ELM} complex in which the observed inner ring constitutes a PulE
267 hexamer comprising the N2E and CTE domains similar to that observed in *V. cholerae*
268 GspE[14] (Figure 6F).

269 The *V. cholerae* GspE N1E domain forms a direct 1:1 stoichiometric complex with
270 the cytoplasmic domain of GspL[17,18]. The equivalent domains in *K. pneumoniae* also form
271 a 1:1 stoichiometric complex termed here $Pul_{E\text{-}N1E/Lcyto}$ (Figure S5D). In the T4P system,
272 the $PulL_{cyto}$ homologue PilM forms a ring around the central PilB/PilT hexamer. This
273 architecture supports a model for the $Pul_{CDELMNS}$ AP and Pul_{ELM} complex where the bulk

274 of the hexameric hub outer ring constitutes $Pul_{E\text{-}N1E/Lcyto}$ complexes (Figure 6F). PulM, the
275 periplasmic domain of PulL (termed $PulL_{peri}$), and PulN and PulC when present, may also
276 contribute to the observed outer ring densities if ordered. Certainly $PulL_{peri}$, PulM and
277 PulN do not form a periplasmic ring here as might be expected based on the T4P system[34].
278 The retention of membrane or inclusion of other T2SS components such as the pseudo-
279 pilus may be required for the stabilization and visualization of any such ring.

280 Within the $Pul_{CDELMNS}$ complex, the mean relative ratio of PulC:PulE:
281 PulL:PulM:PulN was measured to be 2.13:1.00:0.97:0.99:0.95 indicating two copies of
282 PulC for each of the AP components. The globular densities that comprise the outer ring
283 of the hexameric hub do not form a contiguous well-ordered structure. This observation
284 combined with the stoichiometry supports a modular architecture where independent or
285 weakly associated sub-complexes containing two copies of PulC and a single copy of
286 PulM, PulN, and $Pul_{E\text{-}N1E}$ in complex with PulL decorate each subunit of the central PulE
287 hexamer (Figure 6G). This model does not preclude the possible formation of a $Pul_{E\text{-}N1E/Lcyto}$
288 ring positioned more centrally directly above the PulE N2E/CTE hexamer *in vivo*.
289 Such an arrangement would be similar to that observed for the N1 domains of *Thermus*
290 *thermophilus* PilF[35] and may promote the close contact and encircling of the putative pilin
291 spooling protein PulF[13,19].

292

293 **PulC mechanism for overcoming AP-OMC symmetry mismatch**

294 PulC spans the periplasm to bind the OMC with the PulC N-terminus located within the
295 inner membrane and the HR domain bound to the PulD N0 domain at the secretin base.
296 Based on six PulE subunits within the AP, stoichiometry measurements support a model
297 where the PulC copy number within the AP is 12. PulC therefore constitutes a cage that
298 spans the periplasm (Figure 6G) and is reminiscent of the virB10 N-terminus which spans

299 the periplasm in the type IV secretion system[36]. Substrate may have the potential to gain
300 access to the secretin lumen via PulL and PulM[37] in positions where PulC is absent. The
301 OMC structure shows that each PulC HR domain binds a single PulD N0 domain
302 independently with no lateral contacts between neighbouring HR domains. This is
303 important as it provides a natural mechanism for overcoming symmetry mismatch
304 between the 12 PulC subunits located within the hexameric AP and the 15-fold symmetric
305 OMC. In this model on average three secretin N0 domains remain unoccupied without
306 bound PulC HR domain. Whether full secretin N0 domain binding occupancy occurs with
307 up to 15 bound PulC HR domains cannot be entirely excluded as each of the N0 domains
308 has at least the potential to bind a PulC HR domain. T2SS assembly and secretion may
309 ultimately occur in a dynamic equilibrium with a range of N0 domain binding
310 occupancies.
311
312 **The secretin N0-N1 domains form a compact C15 ring**
313 Multiple T2SS secretin structures have been solved including *Klebsiella*-types from *E. coli*
314 K12[10] and *Pseudomonas aeruginosa*[9], and *Vibrio*-types from enterotoxigenic *E. coli*
315 (ETEC)[7], enteropathogenic *E. coli* (EPEC)[8] and *V. cholerae*[10]. Despite high quality 3-3.5 Å
316 resolutions achieved for the secretin core and N3 domains, all the structures have poorly
317 resolved densities towards the PulD/GspD N-termini with homology modelling required
318 for the N1 domains and entirely absent N0 domains. Docking of N0 and N1 domain co-
319 crystal structures[22,25-27] into the secretin maps failed to resolve the possible position and
320 orientation of the N0 domain due to significant steric clashes. Consequently, the PulC HR
321 domain orientation relative to the N0 domain remained unclear and it was modelled
322 decorating the inside of the secretin base[22]. To account for the high flexibility of the
323 secretin N0 and N1 domains and the symmetry mismatch between the 6-fold symmetric

324 AP and 15-fold OMC, a pseudo-6-fold arrangement comprising a hexamer of dimers had
325 been proposed for the N0-N2 domains[9]. In such a model the N0-N2 domains of three
326 secretin subunits would be displaced in a metastable arrangement. The OMC structure
327 derived from the $Pul_{CDELMNS}$ complex resolves now how the PulD N0 domains are
328 organized within the secretin relative to the N1 domain ring. The N-terminus of loop 7
329 wedges and stabilizes neighbouring N1 domains whilst the N0 domains form an intimate
330 and stable C15 ring. At least when PulC HR domain is bound, the N0 ring does not
331 immediately support a model where the secretin incorporates multiple symmetries along
332 its length. Instead the secretin forms a well-ordered barrel, which combined with the
333 PulC HR domains bound at the base protrudes from the outer membrane ~20 nm into the
334 periplasm.
335
336 **Concluding remarks**
337 In this study, the T2SS outer membrane components PulD and PulS are co-purified with
338 the core components of the inner membrane AP including PulC, PulE, PulL, PulM and
339 PulN. Overall, our results reveal a glimpse at the core architecture of an assembled T2SS
340 (Figure 6G) and show it to be different to other known secretion systems (Figure S6). The
341 T2SS AP does not constitute a stack of conjoined rings that seal directly to the secretin
342 base forming an enclosed central channel as observed in the T3SS[38]. Instead, the T2SS
343 has a highly dynamic and modular architecture where the OMC and inner membrane AP
344 have weakly associating and limited inter-connection via a PulC cage. This observation is
345 consistent with the requirement for substrate to gain access to the secretin lumen via the
346 periplasm. Whilst the resolution of the hexameric hub observed in the $Pul_{CDELMNS}$ and
347 Pul_{ELM} complexes is limited it shows the T2SS to have a six-fold arrangement for
348 assembled AP components. The observed AP architecture is consistent with that

349 described in the T4P system[34] providing evidence that the inner membrane
350 ultrastructure is broadly conserved between these different but related secretion
351 systems. Although the connection between inner and outer membrane complexes was
352 expected to be transient and dynamic[3], we show that it is possible to isolate an almost
353 fully assembled T2SS *in vitro*. Our work suggests that the purification of an entire T2SS
354 with the pseudo-pilus, PulF, and ultimately substrate incorporated is in reach. These
355 additional components may be necessary to help stabilize the inner membrane AP so that
356 ultimately full structural dissection and mechanistic understanding for the T2SS may be
357 achieved.
358

359 **Acknowledgements**
360 We thank eBIC for cryo-EM data collection support particularly Kyle Dent and Yuriy
361 Chaban. Tillmann Pape and Paul Simpson for in-house EM support. Francesca Gubellini
362 for gold labelling advice. Arjen Jakobi and Carsten Sachse for LocScale support. This work
363 was funded by a Wellcome Trust Career Development Fellowship Enhancement Award
364 (200074/Z/15/Z) to H.L.
365

366 **Author Contributions**
367 A.C and H.L designed experiments. H.L initially cloned and purified the $Pul_{CDELMNS}$
368 complex. A.C and H.L cloned and purified proteins, collected and processed data, and
369 solved structure. A.C built the OMC structure. A.C and H.L. determined stoichiometry. A.C.
370 undertook gold labelling. H.L wrote the paper with contributions from A.C.
371

372 **Competing interests**
373 The authors declare no competing interests.

374

375 **Data availability**

376 3D cryo-EM density maps produced in this study have been deposited in the Electron
377 Microscopy Data Bank with accession code EMD-0193. Atomic coordinates have been
378 deposited in the Protein Data Bank (PDB) under accession code 6HCG.

379

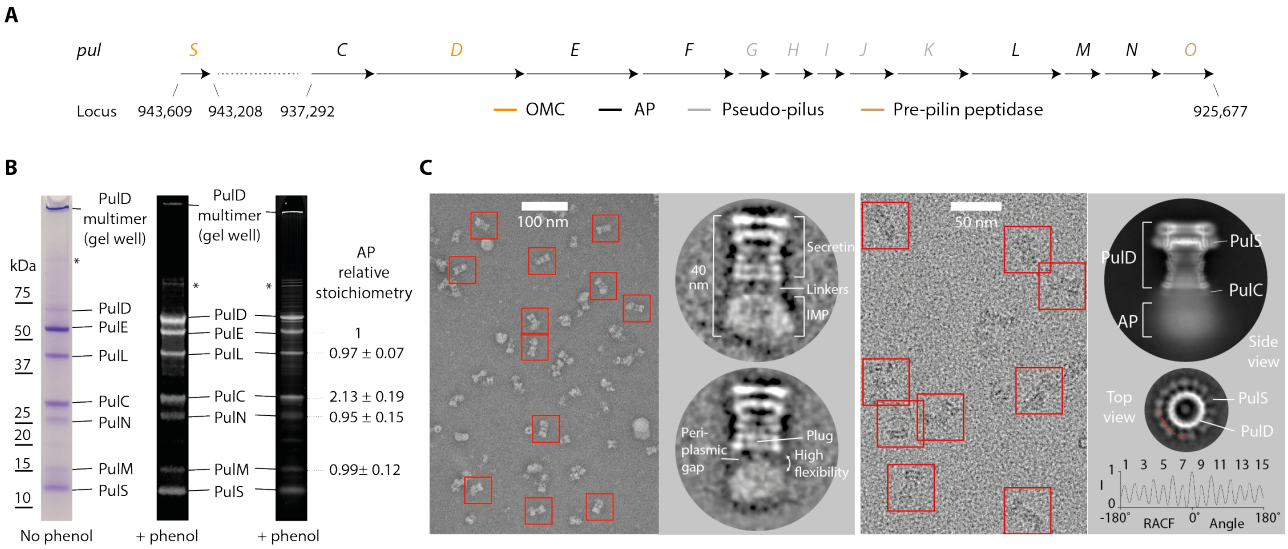
380 **References**

381 1. Cianciotto, N. P. & White, R. C. Expanding role of type II secretion in bacterial
382 pathogenesis and beyond. *Infect Immun* **85**, e00014-17 (2017).
383 2 Nivaskumar, M. & Francetic, O. Type II secretion system: a magic beanstalk or a
384 protein escalator. *Biochim Biophys Acta* **1843**, 1568-1577 (2014).
385 3 Korotkov, K. V., Sandkvist, M. & Hol, W. G. The type II secretion system: biogenesis,
386 molecular architecture and mechanism. *Nat Rev Microbiol* **10**, 336-351 (2012).
387 4 McLaughlin, L. S., Haft, R. J. F. & Forest, K. T. Structural insights into the Type II
388 secretion nanomachine. *Curr Opin Struc Biol* **22**, 208-216 (2012).
389 5 Guilvout, I., Chami, M., Engel, A., Pugsley, A. P. & Bayan, N. Bacterial outer
390 membrane secretin PulD assembles and inserts into the inner membrane in the
391 absence of its pilotin. *Embo J* **25**, 5241-5249 (2006).
392 6 Daefler, S., Guilvout, I., Hardie, K. R., Pugsley, A. P. & Russel, M. The C-terminal
393 domain of the secretin PulD contains the binding site for its cognate chaperone,
394 PulS, and confers PulS dependence on pIVf1 function. *Mol Microbiol* **24**, 465-475
395 (1997).
396 7 Yin, M., Yan, Z. F. & Li, X. M. Structural insight into the assembly of the type II
397 secretion system pilotin-secretin complex from enterotoxigenic *Escherichia coli*.
398 *Nat Microbiol* **3**, 581-587 (2018).
399 8 Hay, I. D., Belousoff, M. J., Dunstan, R. A., Bamert, R. S. & Lithgow, T. Structure and
400 membrane topography of the *Vibrio*-type secretin complex from the type 2
401 secretion system of enteropathogenic *Escherichia coli*. *Journal of Bacteriology* **200**,
402 00521-17 (2018).
403 9 Hay, I. D., Belousoff, M. J. & Lithgow, T. Structural basis of type 2 secretion system
404 engagement between the inner and outer bacterial membranes. *Mbio* **8**, 01344-17
405 (2017).
406 10 Yan, Z. F., Yin, M., Xu, D. D., Zhu, Y. Q. & Li, X. M. Structural insights into the secretin
407 translocation channel in the type II secretion system. *Nat Struct Mol Biol* **24**, 177-
408 183 (2017).
409 11 Lallemand, M. *et al.* Dynamic interplay between the periplasmic and
410 transmembrane domains of GspL and GspM in the type II secretion system. *Plos*
411 *One* **8**, e79562 (2013).
412 12 Abendroth, J. *et al.* The three-dimensional structure of the cytoplasmic domains of
413 EpsF from the type 2 secretion system of *Vibrio cholerae*. *J Struct Biol* **166**, 303-
414 315 (2009).

415 13 Py, B., Loiseau, L. & Barras, F. An inner membrane platform in the type II secretion
416 machinery of Gram-negative bacteria. *Embo Rep* **2**, 244-248 (2001).
417 14 Lu, C. *et al.* Hexamers of the type II secretion atpase gspe from *Vibrio cholerae* with
418 increased atpase activity. *Structure* **21**, 1707-1717 (2013).
419 15 Mancl, J. M., Black, W. P., Robinson, H., Yang, Z. M. & Schubot, F. D. Crystal Structure
420 of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB
421 from Thermus thermophilus. *Structure* **24**, 1886-1897,
422 doi:10.1016/j.str.2016.08.010 (2016).
423 16 Satyshur, K. A. *et al.* Crystal structures of the pilus retraction motor PilT suggest
424 large domain movements and subunit cooperation drive motility. *Structure* **15**,
425 363-376, doi:10.1016/j.str.2007.01.018 (2007).
426 17 Lu, C., Korotkov, K. V. & Hol, W. G. J. Crystal structure of the full-length ATPase
427 GspE from the *Vibrio vulnificus* type II secretion system in complex with the
428 cytoplasmic domain of GspL. *J Struct Biol* **187**, 223-235 (2014).
429 18 Abendroth, J., Murphy, P., Sandkvist, M., Bagdasarian, M. & Holl, W. G. J. The X-ray
430 structure of the type II secretion system complex formed by the N-terminal
431 domain of EpsE and the cytoplasmic domain of EpsL of *Vibrio cholerae*. *J Mol Biol*
432 **348**, 845-855 (2005).
433 19 Arts, J. *et al.* Interaction domains in the *Pseudomonas aeruginosa* type II secretory
434 apparatus component XcpS (GspF). *Microbiology* **153**, 1582-1592 (2007).
435 20 Lopez-Castilla, A. *et al.* Structure of the calcium-dependent type 2 secretion
436 pseudopilus. *Nat Microbiol* **2**, 1686-1695 (2017).
437 21 Sandkvist, M. Biology of type II secretion. *Mol Microbiol* **40**, 271-283 (2001).
438 22. Korotkov, K. V. *et al.* Structural and functional studies on the interaction of GspC
439 and GspD in the type II secretion system. *Plos Pathogens* **7**, e1002228 (2011).
440 23. Butt, R.H. & Coorssen, J.R. Coomassie Blue as a near-infrared fluorescent stain: A
441 systematic comparison with Sypro Ruby for in-gel protein detection. *Molecular &*
442 *Cellular Proteomics* **12**, 3834-3850 (2013).
443 24. Luo, S., Wehr, N.B. & Levine, R.L. Quantitation of protein on gels and blots by
444 infrared fluorescence of Coomassie blue and Fast Green. *Anal Biochem* **350**, 233-
445 8 (2006).
446 25. Korotkov, K.V., Pardon, E., Steyaert, J. & Hol, W.G.J. Crystal structure of the N-
447 terminal domain of the secretin GspD from ETEC determined with the assistance
448 of a nanobody. *Structure* **17**, 255-265 (2009).
449 26. Van der Meeren, R. *et al.* New insights into the assembly of bacterial secretins:
450 structural studies of the periplasmic domain of XcpQ from *Pseudomonas*
451 *aeruginosa*. *J Biol Chem* **288**, 1214-1225 (2013).
452 27. Douzi, B. *et al.* Unraveling the self-assembly of the *Pseudomonas aeruginosa* XcpQ
453 secretin periplasmic domain provides new molecular insights into type II
454 secretion system secreton architecture and dynamics. *Mbio* **8**, e01185-17 (2017).
455 28. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION.
456 *Methods Enzymol* **579**, 125-157 (2016).
457 29. Wang, X. H. *et al.* Cysteine scanning mutagenesis and disulfide mapping analysis of
458 arrangement of GspC and GspD protomers within the type 2 secretion system. *J*
459 *Biol Chem* **287**, 19082-19093 (2012).
460 30. Rehman, S., Gu, S., Shevchik, V. E. & Pickersgill, R. W. Anatomy of secretin binding
461 to the *Dickeya dadantii* type II secretion system pilotin. *Acta Crystallogr D Biol*
462 *Crystallogr* **69**, 1381-1386 (2013).

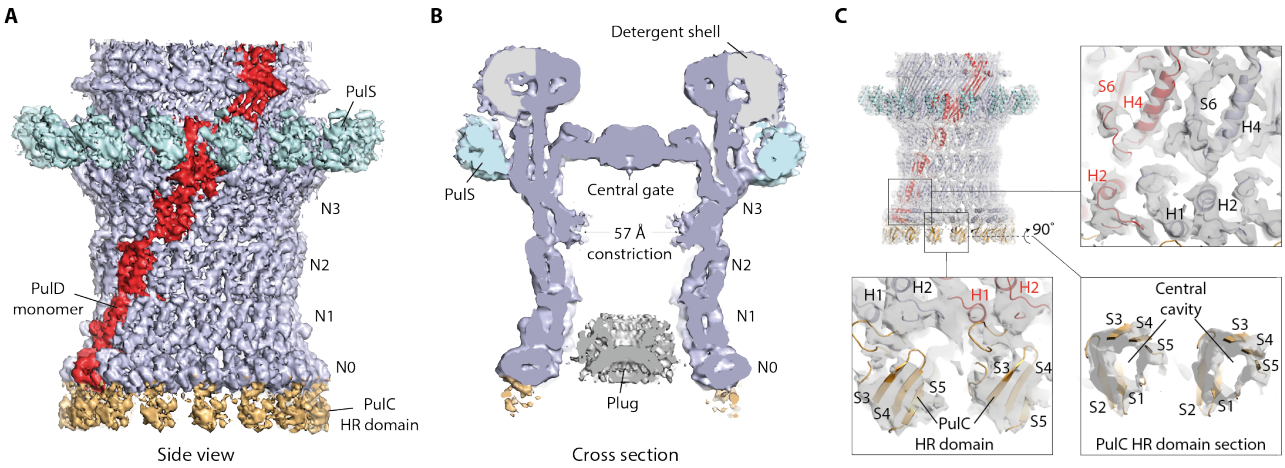
31. Dunstan, R. A. *et al.* Assembly of the type II secretion system such as found in *Vibrio cholerae* depends on the novel pilotin AspS. *Plos Pathog* **9**, e1003117 (2013).
32. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* **5**, 53-55 (2008).
33. 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).
34. Chang, Y. W. *et al.* Architecture of the type IVa pilus machine. *Science* **351**, aad2001 (2016).
35. Collins, R. et al. Structural cycle of the *Thermus thermophilus* PilF ATPase: the powering of type IVa pilus assembly. *Sci Rep* **8**, 14022 (2018).
36. Low, H.H. et al. Structure of a type IV secretion system. *Nature* **508**, 550-553 (2014).
37. Michel-Souzy, S. et al. Direct interactions between the secreted effector and the T2SS components GspL and GspM reveal a new effector-sensing step during type 2 secretion. *J Biol Chem* **293**, 19441-19450 (2018).
38. Worrall, L.J. et al. Near-atomic-resolution cryo-EM analysis of the *Salmonella* T3S injectisome basal body. *Nature* **540**, 597-601 (2016).
39. Jakobi, A.J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *Elife* **6**, e27131 (2017).
40. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-5 (2014).
41. Wang, Z. et al. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. *Elife* **6**, e24905 (2017).

501 **Figures**



502

503 **Figure 1. Purification, stoichiometry and EM analysis of the Pul$_{CDELMNS}$ complex. (A)**
504 Schematic showing the chromosomal location and gene arrangement for the *Klebsiella*
505 *pneumoniae* T2SS. OMC - outer membrane complex, AP - assembly platform. **(B)** Typical SDS-
506 PAGE analysis of the Pul$_{CDELMNS}$ complex. (Left) Coomassie R250 stained gel without phenol
507 treatment. (Middle) Fluorescent emission Coomassie R250 stained gel imaged at 680 nm with
508 phenol extraction. Residual PulD remains in the gel well. (Right) Fluorescent emission Sypro
509 Ruby stained gel imaged at 302 nm with phenol extraction. Relative mean stoichiometry and
510 standard deviation for all assembly platform (AP) components are indicated. Stoichiometry
511 measurements were determined from six independent purifications. * indicates PulD
512 multimer. **(C)** Left panel shows typical negative stain EM micrograph of the Pul$_{CDELMNS}$
513 complex with selected individual particles highlighted with a red box. Zoomed images of 2D
514 class averages are also shown. Right panel shows typical cryo-EM micrograph of the Pul$_{CDELMNS}$
515 complex with associated side view 2D class average. Top view 2D class average shows PulS
516 and PulD with C15 symmetry based on the RACF. Red dotted line indicates radial ring for RACF
517 calculation.



518

519

520 **Figure 2. Cryo-EM density map of the outer membrane complex (OMC). (A)** 4.3 Å
521 resolution C15 symmetrized cryo-EM map of the OMC locally sharpened with LocScale[39] and
522 contoured at 5.5σ. **(B)** Cross section view of the OMC revealing internal features of the PulD
523 secretin. A plug occludes the secretin lumen. The map was unsharpened and contoured at 3σ.
524 **(C)** Fit of molecular models within the OMC map. *De novo* model for the PulD secretin including
525 the N1 and N0 domains (top right). A PulC HR domain homology model was rigid body fitted.
526 The map was sharpened with LocScale and contoured at 4σ (bottom left). In this localized
527 region of the map at 7-8 Å resolution, the central cavity of the PulC HR domain was clearly
528 resolved (bottom right).

529

530

531

532

533

534

535

536

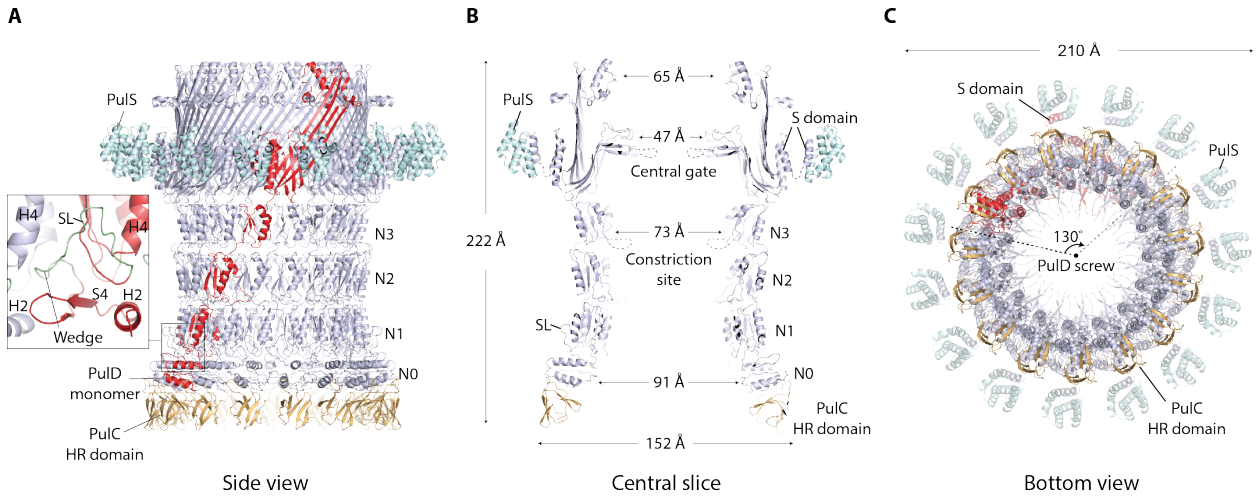


Figure 3. Structure of the outer membrane complex (OMC). (A) Cartoon representation of the OMC, which includes the C15 symmetrized secretin (light blue) with a single PulD monomer highlighted in red. PulS (cyan) and PulC HR domain (orange) homology models were fitted as rigid bodies. 15 PulC HR domains were modelled here although stoichiometry measurements indicate an occupancy of ~0.5 in the $Pul_{CDELMNS}$ complex. 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˚.

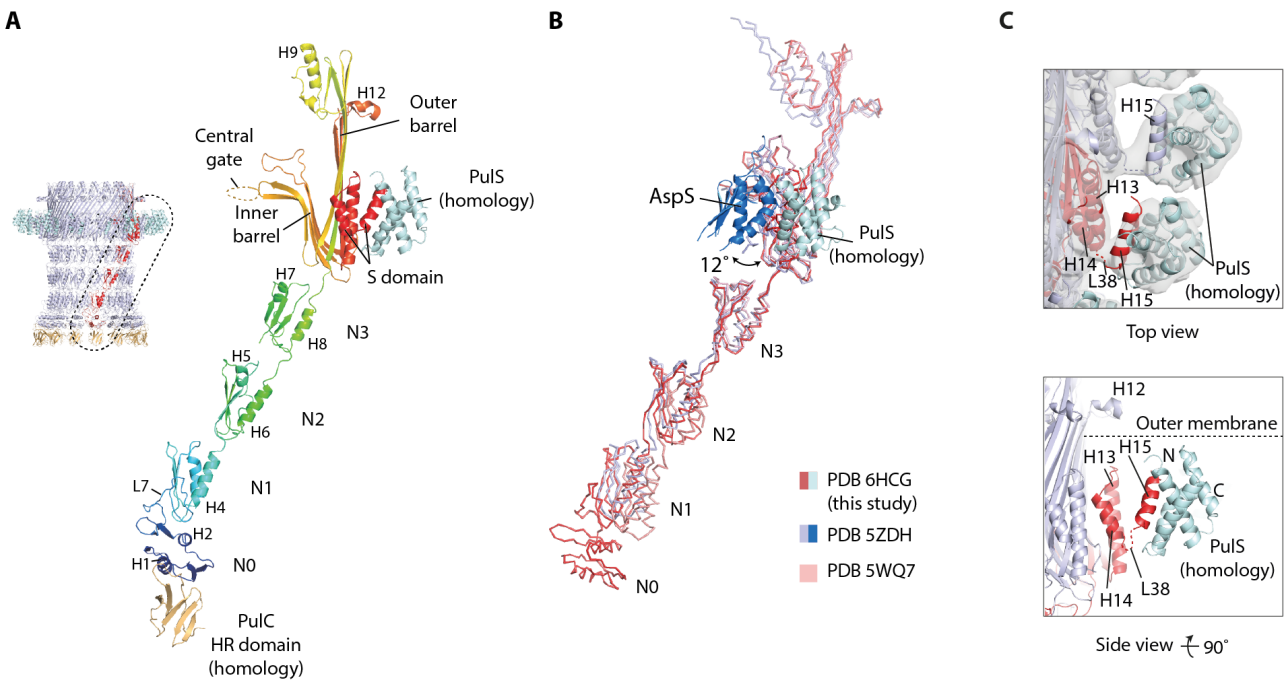


Figure 4. Structural comparison of PulD and PulS models. (A) A complex of PulD, PulS and PulC HR domain monomers extracted from the C15 OMC model (dotted region). For PulD, rainbow coloring highlights blue N-terminus through to red C-terminus. **(B)** Superposition of the PulD and PulS complex with the 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[30]. Flexible loop 38 represents the only attachment between the helix 15/PulS complex and the rest of the PulD monomer.

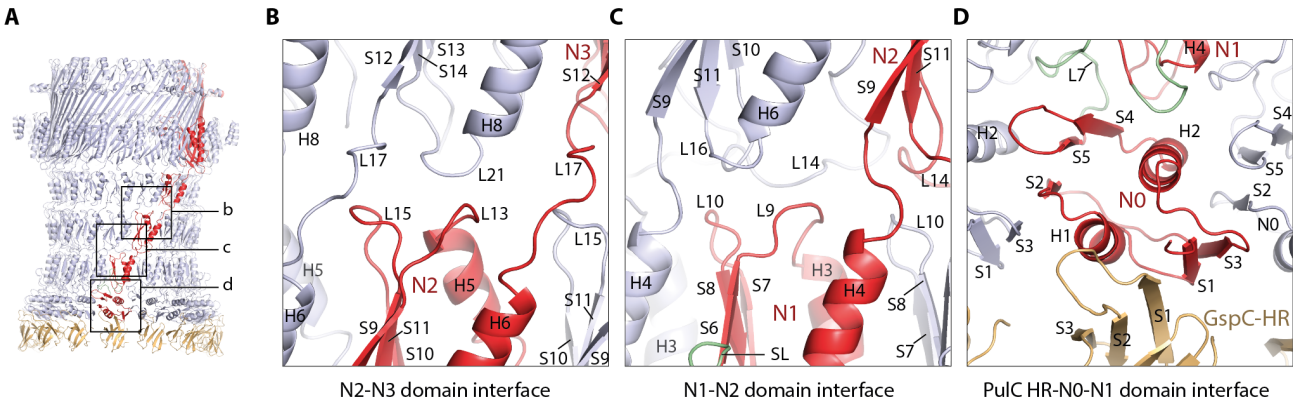
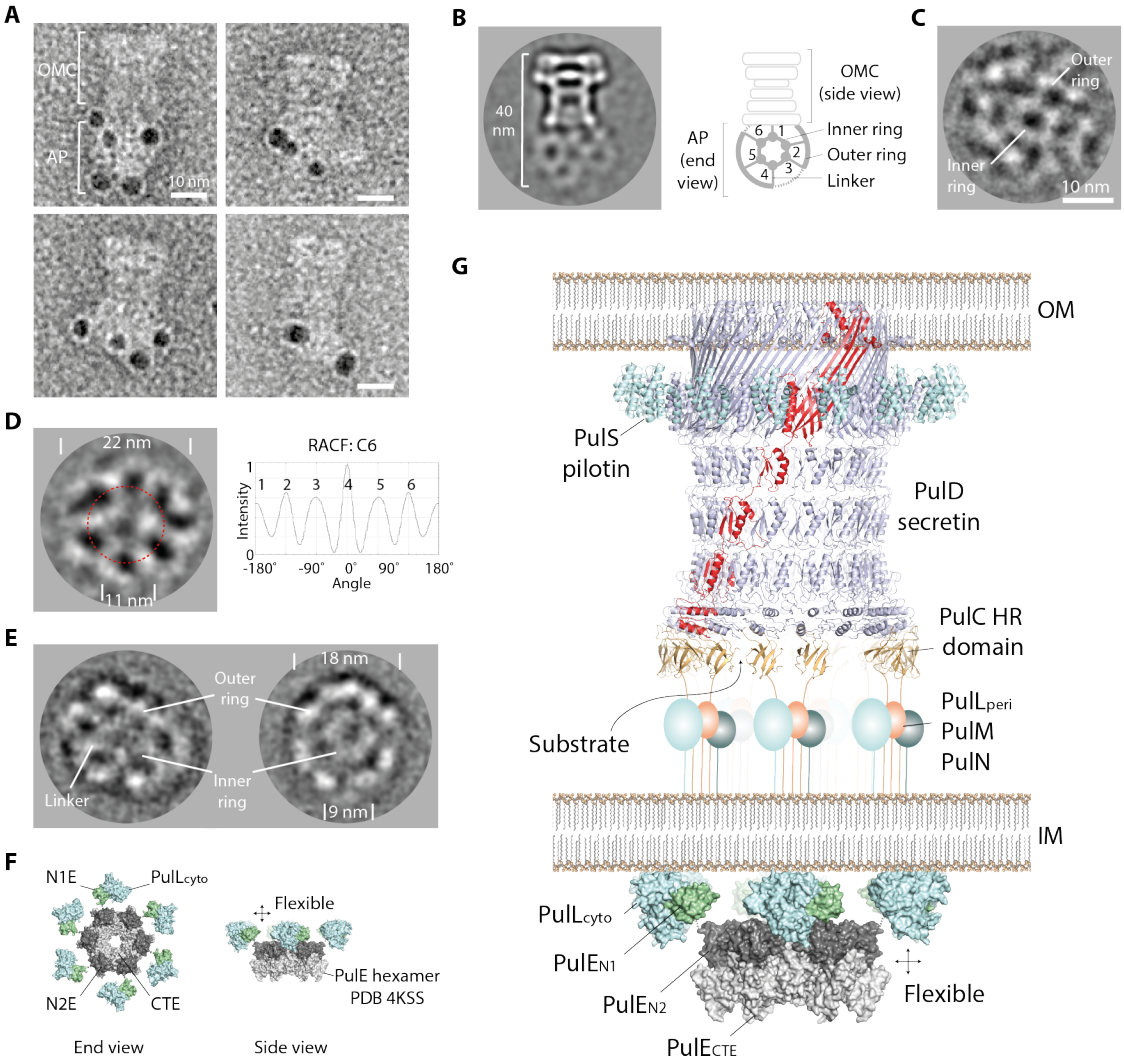


Figure 5. Structural analysis of the outer membrane complex (OMC) periplasmic domains. (A) Overview model of the OMC with PulS removed for clarity. Boxed regions are zoomed in **B-D**. **(B-D)** Cartoon representation of selected domain interfaces showing key structural details.

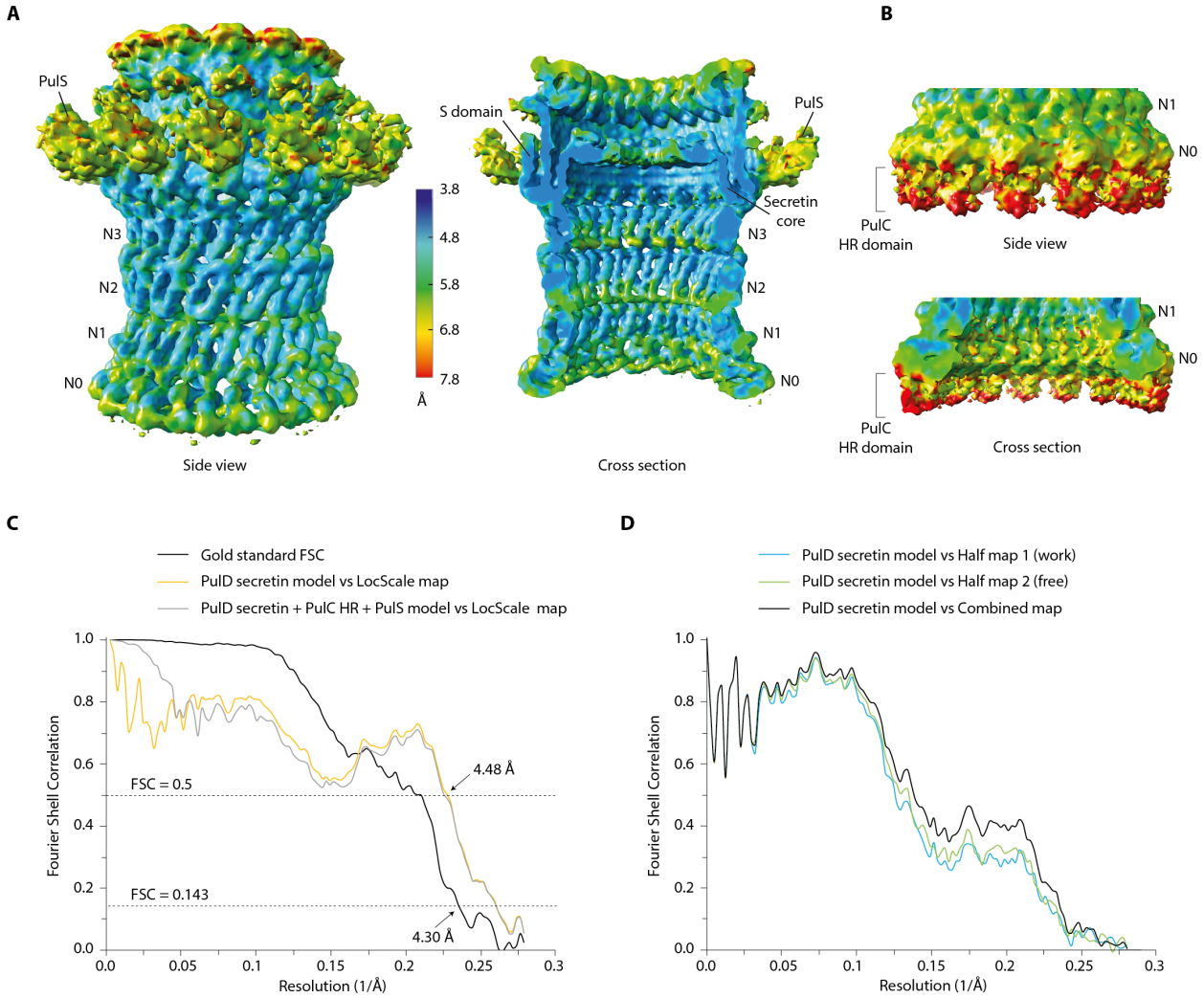


591

592 **Figure 6. 2D EM of the assembly platform (AP) and a T2SS core model. (A)** Ni-NTA gold
593 bead labelling of the PulC N-terminus. Beads locate exclusively to the AP. **(B)** Cryo-EM 2D class
594 average of $Pul_{CDELMNS}$. The AP showed end views as a preferred orientation with C6 symmetry.
595 **(C)** Cryo-EM 2D class average of $Pul_{CDELMNS}$ with the alignment and classification focused on
596 the AP. **(D)** Cryo-EM 2D class average of Pul_{ELM} showing C6 symmetry. RACF = rotation auto-
597 correlation function calculated around dotted red radial ring. **(E)** Negative stain EM 2D class
598 average of Pul_{ELM}. **(F)** Schematic showing modelled arrangement of the Pul_{EL} cytoplasmic
599 domains within the $Pul_{CDELMNS}$ and Pul_{ELM} hexameric hub. The central PulE hexamer (inner
600 ring, PDB 4KSS) is decorated by six copies of the $Pul_{E\text{-}N1E/Lcyto}$ complex (outer ring, PDB 2BH1).
601 Connection between inner and outer ring is mediated by the 44 aa flexible N1E-N2E inter-

602 domain linker. **(G)** Stoichiometric model of an assembled T2SS core apparatus. AP
603 components comprise a hexameric hub. PulC connects the AP to the OMC by binding to the
604 secretin base. Substrate may be loaded through the PulC cage in positions where GspC is
605 absent.
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626

627 **Supplementary Figures**



628

629 **Figure S1. Outer membrane complex (OMC) local resolution map and FSC curves. (A)**
630 Unsharpened OMC map showing local resolution estimates calculated using ResMap[40] and
631 contoured at 4.5σ. **(B)** Equivalent to **A** but contoured at 2σ so as to view the PulC HR domain
632 densities at the secretin base. **(C)** Gold standard FSC curve of the OMC (black). FSC curves
633 between the PulD secretin model and the LocScale[39] sharpened map (orange). FSC curves
634 between the PulD secretin model with homology models of PulS and PulC HR domain fitted,
635 and the LocScale sharpened map (grey). **(D)** Cross-validation for the PulD secretin model.
636 PulD secretin model versus half map 1 (work- light blue), half map 2 (free- green), and
637 combined (black). See Methods for further detail.

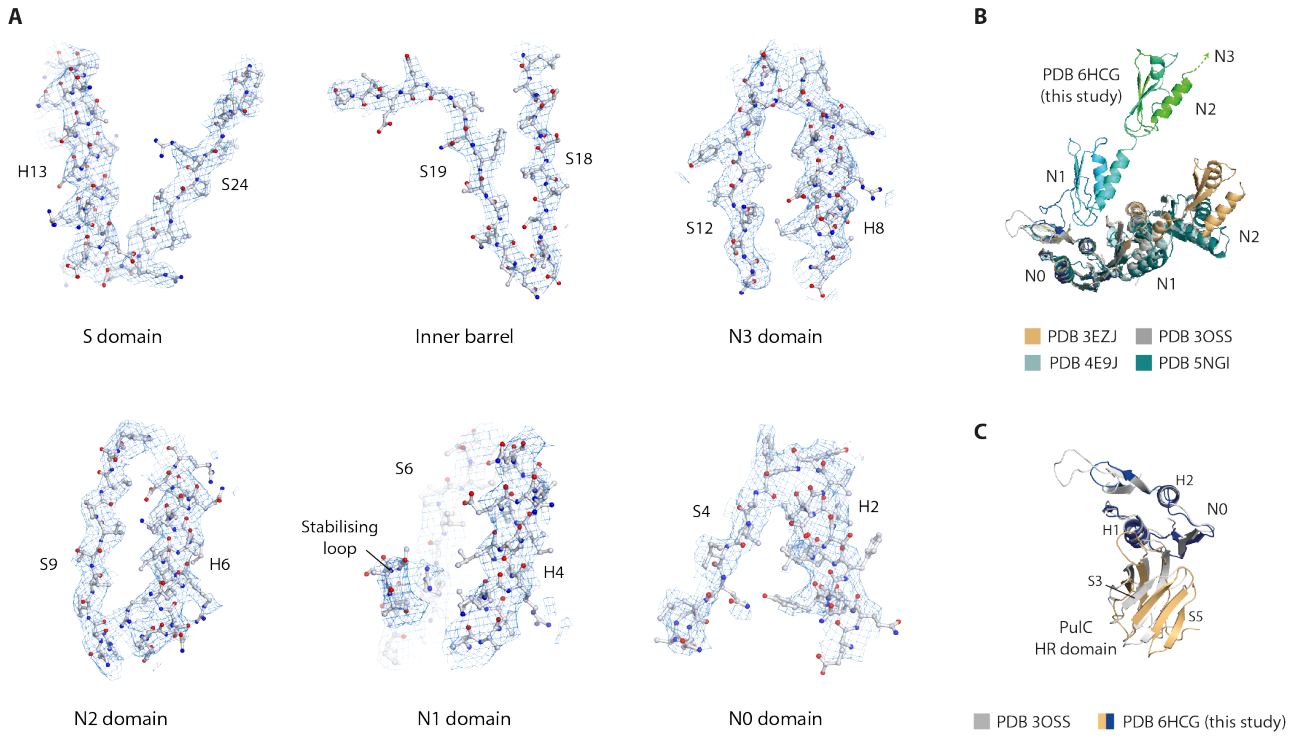


Figure S2. Map quality and PulS fit. (A) Selected regions of the OMC EM density map showing PulD secretin side chain detail and build. The map was sharpened with B factor = -142 Å^2 and contoured between 5-8σ. **(B)** Superposition of PulD N0, N1 and N2 domains with equivalent domains derived from various 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
                  .  : :  :: :**: * ** :** *.*:       .*:**.  . *         .:
```

654

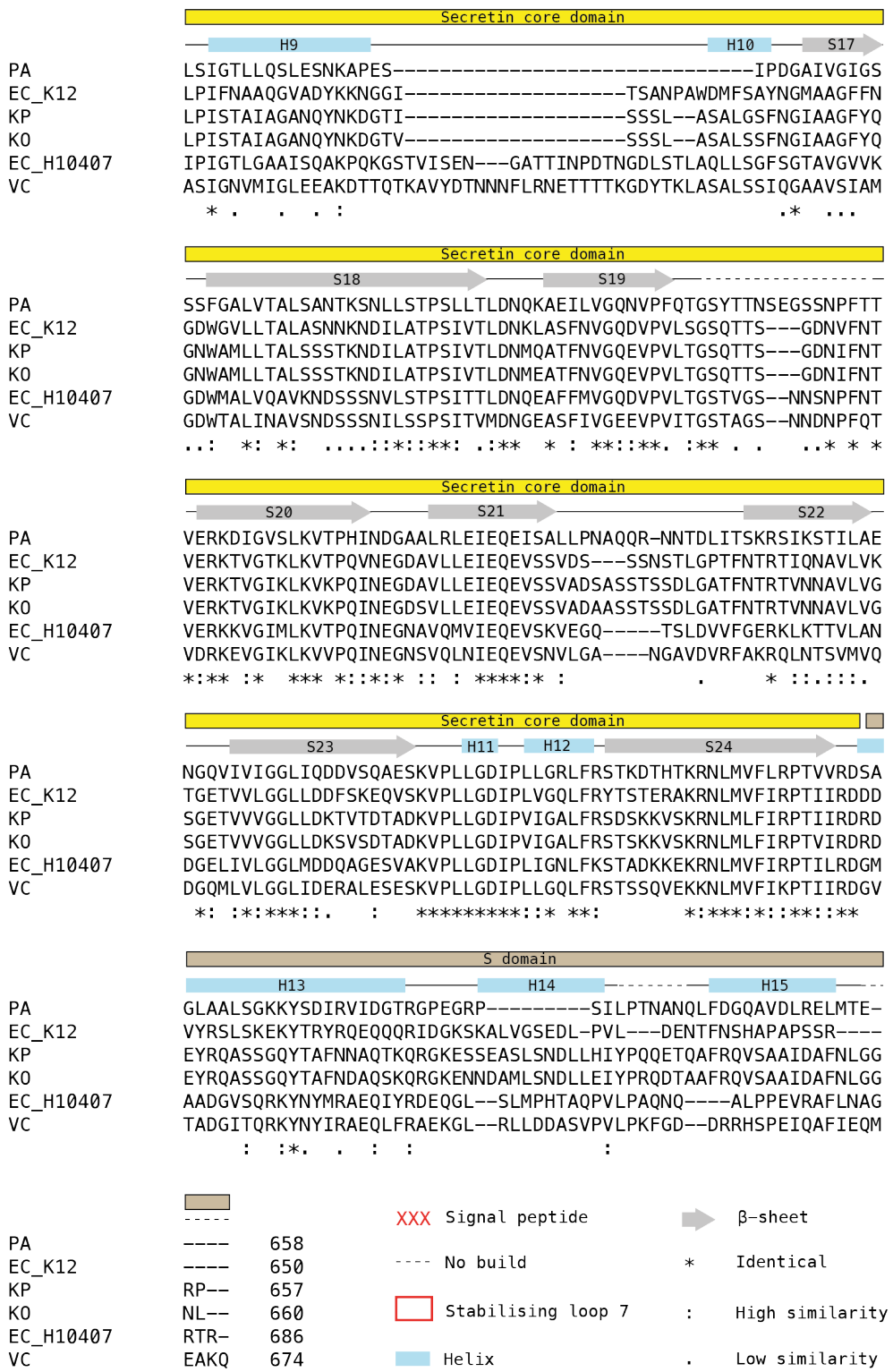


Figure S3. PulD secondary structure assignment and sequence alignment. Aligned PulD sequences include *Pseudomonas aeruginosa* (PA, Uniprot code P35818), *Escherichia coli* K12

658 (EC_K12, P45758), *Klebsiella pneumoniae* (KP, A0A0E1CJT4), *Klebsiella oxytoca* (KO,
659 A0A0H3H6N4), *Escherichia coli* H10407 (EC_H10407, E3PJ86), *Vibrio cholerae* (VC, P45779).
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680

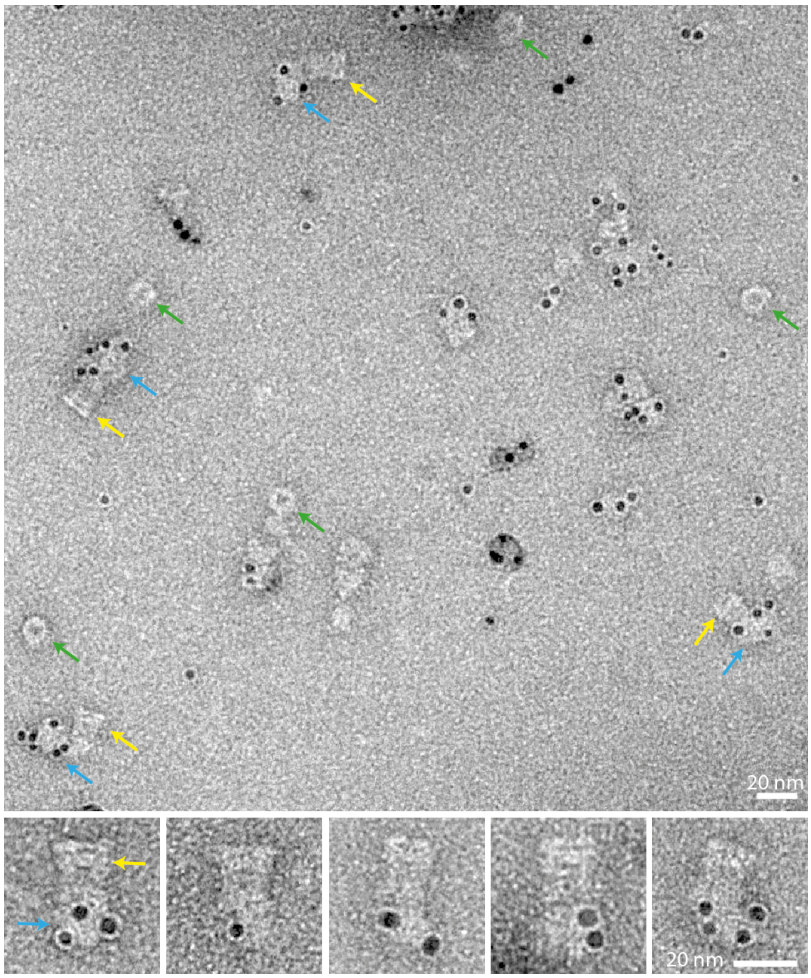


Figure S4. PulC N-terminus is located in the inner membrane assembly platform (AP). An overview negative stain EM image with a gallery of selected particles underneath. The PulC N-terminus was labelled with a hexahistidine tag at aa 61 within the $Pul_{CDELMNS}$ complex. Ni-NTA gold beads localize to 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).

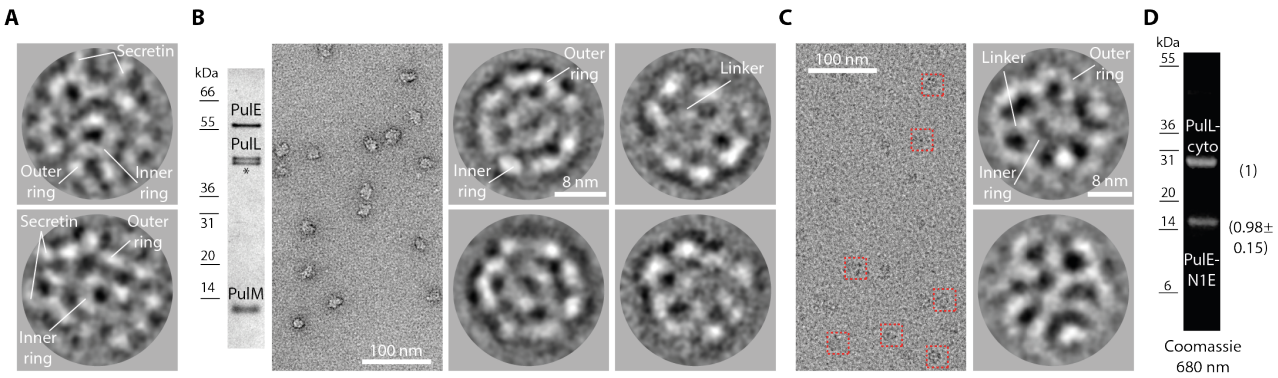


Figure S5. 2D EM analysis of the $Pul_{CDELMNS}$ assembly platform (AP) and the Pul_{ELM} complex. (A) Gallery of cryo-EM 2D class averages of the $Pul_{CDELMNS}$ complex where the alignment and classification were focused on the AP only. A single preferred orientation was observed consistent with a 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. **(D)** SDS-PAGE of purified $Pul_{E\text{-}N1E/Lcyto}$ complex. Fluorescent emission Coomassie R250 stained gel imaged at 680 nm with associated stoichiometry and standard deviation in parentheses. Stoichiometry measurements were determined from two independent purifications.

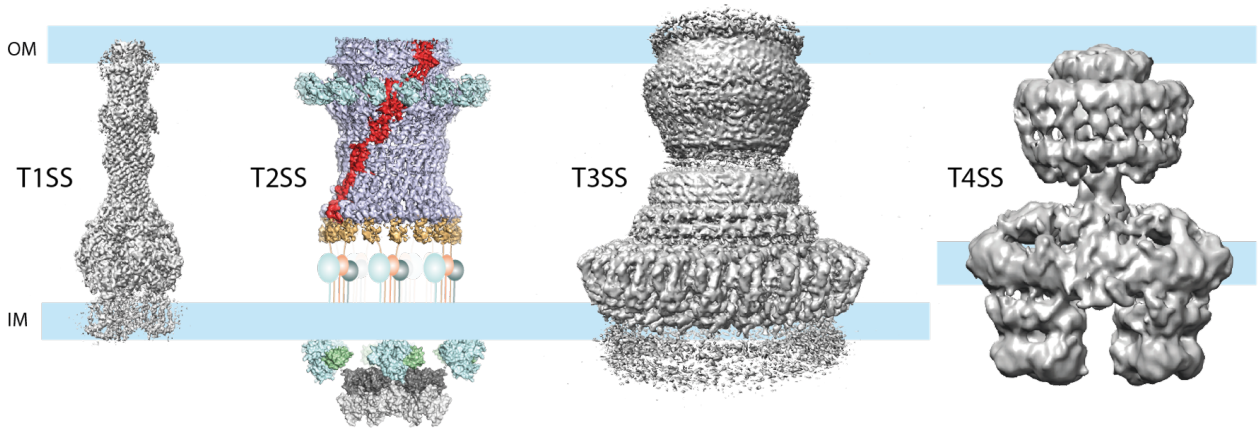


Figure S6. Comparison of bacterial type I-IV secretion systems. Substrate passes in an enclosed channel across the entire cell envelope in the type I (EMD 8636) and III (EMD 8400) secretion systems[38,41]. In the T2SS, the substrate must enter the periplasm before outer membrane translocation through 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

*identical models

782 **Supplementary Table 2. Cloning primers**
783

Primer pair 1	taacgagggcaaaaaatgccagtttctgtgatgaggttaacaaata aggtcaagcttattactagaccatctgctgccagagaaaaata
Primer pair 2 (pASK3c)	taataagcttgacctgtgaagtg tttttgccctcgttatctagatttttg
Primer pair 3	ccacccgcagttcgaaaaataagatggccctgtttcgctatcaggcg tcgaactgcgggtggctccagacgctttccgtcgtttgctgttcg
Primer pair 4	aggagatataccatgattcattattttgattactctatgatg tgttcgacttaagcattagtcgctgcgcaagctggc
Primer pair 5 (pCDF-duet)	tgcttaagtcgaacagaaagtaatc catggtatatctccttattaaag
Primer pair 6	actacaaagacgatgacgacaagtaatgcttaagtcgaacagaaagtaat catcgtctttgtagtcggatccgtcgctgcgcaagctggcaata
Primer pair 7	taacgagggcaaaaaatgacgccagccgccgaacgccgt gtggctccaagcgctgacgctttccgtcgtttgctgttcggtg
Primer pair 8 (pASK3c)	agcgcttggagccacccgcagttc tttttgccctcgttatctagatttttg
Primer pair 9	aggagatataccatgaacaagattaatgcttctccccaggcc tttgtagtcggatccttttttgctcctgtggtaaccattttccctg
Primer pair 10 (pCDF-duet/flag tag)	ggatccgactacaaagacgat catggtatatctccttattaaag
Primer pair 11	taacgagggcaaaaaatgacgccagccgccgaacgccgtcc gtggctccaagcgctgctgtcgagcagatcgtcggtgtccg
Primer pair 12	aggagatataccatgaacaagattaatgcttctccccaggcc tttgtagtcggatccggcgaactccccctgcagcagactga
Primer pair 13	catcaccatcatcaccactaattgccaaccatgaaaagcctgcgaaaaat gtgatgatggtgatggccttcatcccccagcgcgaaagcaatatcgtgtc

822

823

824

825

826

827

828

829

830 **Methods**

831 **Cloning, protein expression and purification.** All clones were generated using a
832 modified version of the Gibson isothermal DNA assembly protocol[42] where the *Taq* DNA
833 ligase was omitted from the one-step isothermal DNA assembly. To obtain the $Pul_{CDELMNS}$
834 complex, the *Klebsiella pneumoniae* T2SS operon encoding genes from *pulC* to *pulO* was
835 cloned into pASK3c vector (IBA-GO) using primer pair 1 and 2 (for all primers used see
836 Supplementary Table 2). A StrepII tag was then added at the C-terminus of PulE using
837 primer pair 3. The *pulS* gene was cloned into pCDF-duet vector using primer pair 4 and
838 5. A C-terminal Flag tag was then inserted using primer pair 6. Uniprot codes for
839 individual genes are as follows: *pulC* KPHS_08880, *pulD* KPHS_08870, *pulE* KPHS_08860,
840 *pulF* KPHS_08850, *pulG* KPHS_08840, *pulH* KPHS_08830, *pulI* KPHS_08820, *pulJ*
841 KPHS_08810, *pulK* KPHS_08800, *pulL* KPHS_08790, *pulM* KPHS_08780, *pulN*
842 KPHS_08770, *pulO* KPHS_08760, *pulS* KPHS_08940. The clones were co-transformed into
843 *Escherichia coli* C43 (DE3) electro-competent cells (Lucigen) modified here to
844 incorporate a *pspA* gene knockout using a Lambda Red recombinase strategy[43] (PspA is a
845 common contaminant induced by PulD over-expression). Cells were grown on selective
846 LB-agarose plates with chloramphenicol (30 µg/ml) and spectinomycin (50 µg/ml). 2xYT
847 media was inoculated and cells grown at 37°C until induction at OD_{600} = 0.5-0.6 with
848 anhydrotetracycline (AHT, 0.2 mg/L) and isopropyl β-D-1-thiogalactopyranoside (IPTG,
849 0.24 g/L). Cells were grown for ~15 hr at 19 °C and processed immediately. Pellets were
850 re-suspended in ice-cold buffer 50 mM Tris-HCl pH 7.5, 5 mM EDTA, treated with DNase
851 I, lysozyme and sonicated on ice. The lysate was clarified by centrifugation at 16,000g for
852 20 min. The membrane fraction was collected by centrifugation at 142,000g for 45 min.
853 Membranes were mechanically homogenized and solubilized in 50 mM Hepes-NaOH pH
854 7.5, 150 mM NaCl, 1 % w/v DDM (Anatrace) and 5 mM EDTA at room temperature for 30-

855 40 min. The suspension was clarified by centrifugation at 132,000g for 15 min. The
856 supernatant was loaded onto a StrepTrap HP column (GE Healthcare) and washed with
857 50 mM Hepes-NaOH pH 7.5, 150 mM NaCl, 0.06 % w/v DDM and 5 mM EDTA (Buffer W)
858 at 4 °C. All prior buffers were supplemented with EDTA-free cOmplete protease inhibitor
859 tablets (Roche). The protein sample was eluted in Buffer W supplemented with 2.5 mM
860 desthiobiotin (IBA) but with protease inhibitors removed. Peak fractions were pooled,
861 0.05 % glutaraldehyde (Sigma- EM grade) added and incubated on ice for 10 min before
862 quenching with 100 mM Tris-HCl pH 7.5. The sample was batch incubated with Flag resin
863 (Sigma) for 1 hr. Flag resin was washed with Buffer W and then eluted with the same
864 buffer supplemented with 3xFlag peptide. Peak fractions were collected and used
865 immediately. LC-MS/MS confirmed the identity of bands identified by SDS-PAGE.

866 To obtain purified Pul_{ELMN} complex, the full-length *pulE* gene was cloned into
867 pASK3c vector to include a C-terminal StrepII tag (primer pair 7 and 8). The *pulL*, *pulM*
868 and *pulN* region of the T2SS operon was cloned into pCDF-duet vector with a PulN C-
869 terminal Flag tag (primer pair 9 and 10). These vectors were co-transformed and the
870 same initial purification strategy was then followed as for the $Pul_{CDELMNS}$ complex with
871 the exception that no glutaraldehyde or Tris quenching buffer were added subsequent to
872 elution from the Strep column. Protease inhibitor tablets were included in all buffers.
873 After elution from the Flag column, due to sample heterogeneity as judged by negative
874 stain EM, GraFix[32] was undertaken. Using Beckman Ultra-Clear 4.2 ml 11x60 mm
875 ultracentrifugation tubes 2.1 ml of 50 mM Hepes-NaOH, 150 mM NaCl, 0.06 % w/v DDM,
876 30 % v/v glycerol, 5 mM EDTA and 0.1 % glutaraldehyde was loaded under 2.1 ml of the
877 equivalent but with 10 % v/v glycerol. A continuous gradient was made using a BioComp
878 Gradient Master cycle set for 66 seconds at 83° tilt and 22 rpm. An equivalent gradient
879 was also made but omitting glutaraldehyde so that the sample could be analysed by SDS-

880 PAGE after centrifugation. The Pul_{ELMN} sample was split in half and loaded onto the
881 gradients ±glutaraldehyde. Samples were spun at 71,000g for 16 hr at 4°C using a
882 Beckman Ti 60.1 swing rotor. 150 µl fractions were collected manually and analysed.
883 Note that PulN was generally lost from the Pul_{ELMN} complex during GraFix yielding just
884 Pul_{ELM}.

885 To obtain purified $Pul_{E\text{-}N1E/Lcyto}$ complex, the N1E domain comprising aa 1-108
886 from *pulE* were cloned into pASK3c vector to include a C-terminal StrepII tag (primer pair
887 8 and 11). For Pul_{Lcyto}, aa 1-235 relating to the cytoplasmic domain of PulL were cloned
888 into pCDF-duet vector with a C-terminal Flag tag (primer pair 10 and 12). The same 2-
889 step affinity chromatography purification strategy was then followed as for PulELMN but
890 excluding Grafix. Note that the $Pul_{E\text{-}N1E/Lcyto}$ complex readily purifies from the membrane
891 fraction despite the removal of the PulL trans-membrane domain.

892 **Gold labelling.** This was performed on the $Pul_{CDELMNS}$ complex modified to include a
893 hexahistidine tag within PulC after aa 61 (primer pair 13). Purification was the same as
894 for the $Pul_{CDELMNS}$ complex but without the addition of fixative. Protease inhibitor tablets
895 were included in all buffers. 5 nm Ni-NTA-Nanogold (Nanoprobes) pre-washed in Buffer
896 W was added to 10 µl of the protein sample and incubated for 30 min at 4 °C. A homemade
897 continuous carbon grid was deposited on the 10 µl sample for 3 min, blotted and washed
898 2 times in Buffer W supplemented with 10 mM imidazole, then 2 times in Buffer W before
899 being stained with 3 drops of 2 % uranyl acetate.

900 **Stoichiometry determination.** $Pul_{CDELMNS}$ complex was purified as above with the
901 exception that no glutaraldehyde or Tris quenching buffer were added subsequent to
902 elution from the Strep column. Protease inhibitor tablets were included in all buffers. Six
903 independent purifications were extracted using phenol to disrupt PulD
904 multimerization[44]. Before extraction samples were divided and phenol treated in

905 duplicate. The samples were precipitated with an equal volume of phenol and
906 immediately vortexed. Four volumes of ice cold acetone were then added and vortexed
907 again. The mixture was kept overnight at -20 °C. The precipitate was pelleted in a bench-
908 top centrifuge at 14,000 g for 30 min at 4 °C, washed once with ice cold acetone, dried
909 under vacuum and resuspended in loading buffer supplemented with 10 mM DTT. Both
910 duplicates for each independent purification were analyzed by SDS-PAGE densitometry
911 and quantification of fluorescent emission using both Coomassie Blue R250 and Sypro
912 Ruby dyes. The Coomassie R250 procedure was as described[24] but with extended stain
913 and destaining times. Briefly, gels were rinsed with water and then stained with 0.01%
914 Coomassie Blue R250 in 50% methanol and 10% acetic acid for 20 min. Gels were rinsed
915 with 40% methanol and 7% acetic acid and then destained twice with the same solution
916 for 20 min each time. Finally, gels were soaked in 2-3 times in water until fully destained,
917 usually 1-2 hours. Gels were imaged on a Licor Odyssey Fc at 680 nm. Sypro Ruby gel
918 staining was undertaken following the manufacturer's instructions (Biorad). Gels were
919 imaged on a Biorad Chemidoc XRS imaging system at 302 nm. Image Lab software
920 (BioRad) and ImageJ were used to calculate band intensities.

921 To determine the stoichiometry of the $Pul_{E\text{-}N1E/Lcyto}$ complex, two independent
922 purifications were undertaken. Samples were run in duplicate on gels and then stained
923 using the Coomassie Blue R250 procedure as outlined above. Gels were imaged on a Licor
924 Odyssey Fc at 680 nm.

925 **Electron microscopy sample preparation and data collection.** For outer membrane
926 complex (OMC) structure determination, 4 µl of purified $Pul_{CDELMNS}$ complex solution was
927 incubated for 30 seconds on glow discharged homemade continuous thin carbon grids
928 before vitrification in liquid ethane using a Vitrobot Mark IV (FEI). Data was collected at
929 300 kV on a Titan Krios (M02 beamline at eBIC Diamond, UK) equipped with a Gatan

930 Quantum K2 Summit detector. Images were acquired at a magnification of 28,090
931 yielding 1.78 Å/pixel using EPU software. Images were dose-weighted over 40 frames
932 with 12 second exposures. Total dose was ~50 e/Å^2.

933 All other cryo and negative stain datasets were collected in-house at 200 kV on a
934 Tecnai F20 microscope equipped with Falcon II direct electron detector. For cryo data,
935 $Pul_{CDELMNS}$ complex (Figure 6B-C and Figure S5A) was vitrified as described above. The
936 Pul_{ELM} GraFix treated sample required glycerol removal so that 4 µl of sample was loaded
937 onto the glow discharged continuous carbon EM grid and after 1 min incubation was
938 washed 4 times in Buffer W before plunge freezing. Images were acquired at a
939 magnification of 90,909 yielding 1.65 Å/pixel using EPU software. Images were collected
940 over 54 frames with 3 second exposures. Total dose was ~50 e/Å^2. For negative stain
941 data, 4 µl of $Pul_{CDELMNS}$ complex was loaded onto the glow discharged continuous carbon
942 EM grid, after 40 seconds the grid was washed with 3 drops of distilled water and stained
943 with 3 drops of 2 % uranyl acetate. Pul_{ELM} GraFix treated sample was similarly incubated
944 on EM grids, washed iteratively with 4 drops of 15 µl Buffer W and then negatively
945 stained as above. Images were acquired at a magnification of 90,909 using EPU software.
946 Single frames were collected with 1 second exposure and total dose ~15 e/Å^2.

947 **Image processing.** For OMC structure determination, individual movie image frames
948 were aligned with MotionCor2 [45] and the contrast transfer function estimated using Gctf
949 1.06[46]. Low quality images were discarded and 3427 micrographs used for subsequent
950 reconstruction in Relion 2.1[47]. Initial manual particle picking was focused on the
951 OMC/secretin region of the $Pul_{CDELMNS}$ complex. For particle extraction a box and mask
952 diameter were chosen so that contributions from the inner membrane assembly platform
953 (AP) were excluded. In this way, low-resolution 2D class averages of just the OMC were
954 used as a template for auto-picking. OMC side views only were prevalent in this dataset,

955 which provided a sufficiently even equatorial band distribution for a reliable
956 reconstruction[48]. Low quality particles were removed by 4 rounds of 2D classification
957 resulting in a stack of 36,240 particles. A single round of 3D classification was undertaken
958 generating 10 classes. The *Vibrio cholerae* GspD reconstruction EMDB-1763 was used as
959 an initial model filtered to 40 Å[49]. C15 symmetry was applied based on top views of the
960 OMC (obtained in an alternative $Pul_{CDELMNS}$ complex purification) and an unambiguous 15
961 peaks observed from the rotation auto-correlation function calculation (Figure 1C). A
962 single class containing 7284 particles was used for the final refinement, which attained
963 4.4 Å resolution. Post-processing yielded 4.3 Å resolution with an auto-estimated B-factor
964 [50] of -142 $Å^2$ applied to sharpen the final 3D map for model building. A locally sharpened
965 map was also generated using LocScale[39] once initial models were built. Further particle
966 polishing and 3D refinement did not yield a marked increase in resolution. Resolutions
967 reported are based on gold standard Fourier shell correlations (FSC) = 0.143. Statistics
968 for data collection and 3D refinement are included in Supplementary Table 1. Local 3D
969 refinements with various particle subtraction strategies focusing on PulS or PulC HR
970 domain did not markedly improve resolution.

971 To generate all other 2D class averages both in cryo and negative stain conditions
972 as for $Pul_{CDELMNS}$ and Pul_{ELM} complexes, the following protocol was followed. Working
973 initially within Relion 2.0 or 2.1, Gctf 1.06 was used for estimating the CTF. Negative stain
974 micrographs were phase flipped. Low quality micrographs were discarded. Initial 2D
975 class averages were generated from a manually picked stack to yield templates for
976 autopicking. 3-4 rounds of 2D classification were then undertaken to remove low quality
977 particles. Using the 'relion_stack_create' a cleaned image stack was generated for further
978 processing. For cryo-EM images the stack was created from phase flipped particles. In
979 Imagic[51], particles were normalised, band pass filtered, centred and subjected to

980 reference-free MSA and classified. The best classes, typically judged by lowest variance,
981 were used as references for multi-reference alignment (MRA) in Spider[52] followed by
982 MSA and classification in Imagic. This cycle of MRA and MSA was typically iterated a
983 further 2-3 times. For negative stain Pul_{ELM} data, 2746 micrographs yielded 81,727
984 extracted particles and a cleaned stack of 63,014 particles. For cryo Pul_{ELM} data, 467
985 micrographs yielded 4937 hand-picked particles, and all particles were used for
986 subsequent MSA and MRA. For $Pul_{CDELMNS}$, 3059 micrographs yielded 89,381 extracted
987 particles and a cleaned stack of 66,855 particles for subsequent MSA and MRA (Figure
988 6B). From this same dataset, 1148 micrographs were then used to hand-pick 4020 AP-
989 focused particles (Figure 6C and Figure S5A). Rotation auto-correlation functions were
990 calculated using Imagic.

991 **Model building.** A PulD homology model was generated with I-Tasser[53] using the *E. coli*
992 GspD PDB 5WQ7 as a template. This yielded a starting model for the secretin core, and
993 N3, N2 and N1 domains. *E. coli* GspD shares 57 % sequence identity with PulD. A
994 homology model for the N0 domain was generated using Swissmodel[54] with the relevant
995 part of 3OSS as a template (51 % sequence identity). Homology models were rigid body
996 fitted into the map using Chimera[55] Fit in Map function. Using these models as a starting
997 guide and the side chain detail from bulky residues to confirm sequence register, Coot[56]
998 was used to manually build a complete model for PulD aa 27-652 excluding aa 288-303,
999 462-470 and 632-637. The model was further refined using real-space refinement in
1000 Phenix[57] with secondary structure, geometry and NCS restraints applied. For the low-
1001 resolution regions specific to the PulC HR domain, a homology model based on ETEC GspC
1002 HR domain PDB 3OSS (27 % sequence identity) was generated using Swissmodel. For
1003 fitting this homology model, the PDB 3OSS which includes the ETEC GspD N0 domain was
1004 first superimposed onto the PulD N0 domain. The PulC HR domain homology model was

1005 then superimposed onto the ETEC GspC HR domain from PDB 3OSS resulting in a near
1006 perfect fit within the map. The Chimera Fit in Map function was then applied to PulC HR
1007 domain resulting in a minor shift so that the PulC HR-PulD N0 domain complex has a
1008 RMSD Cα = 1.4 Å when aligned to PDB 3OSS (Figure S2C). For the low-resolution regions
1009 specific to the PulD S-domain C-terminus (aa 638-652) in complex with the PulS pilotin,
1010 a homology model was generated using Swissmodel based on the equivalent structure
1011 from *Dickeya didantii* PDB 4K0U (>50 % sequence identity for both chains). The map was
1012 low pass filtered to 8 Å and the homology model initially fitted manually so that the PulS
1013 lipidated N-terminus orientated towards the membrane. The Chimera Fit in Map function
1014 was then used for final positioning. Cross-validations were carried out as previously
1015 described[10,58] using the auto-estimated B-factor sharpened map. Briefly, the PulD
1016 secretin model was displaced randomly by 0.2 Å and then refined against a map
1017 reconstructed from one of the independent data halves (Half map 1). FSC curves were
1018 then calculated using the resulting model and Half map 1 (FSC_{work}). FSC curves were also
1019 calculated between this same model and another reconstruction generated from the
1020 other independent data half (Half Map 2 and FSC_{free}). The similarity between FSC_{work} and
1021 FSC_{free} curves indicates an absence of overfitting within the PulD secretin model (Figure
1022 S1D). The final models were assessed using Molprobity[59] and statistics outlined in
1023 Supplementary Table 1.

1024

1025 42. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred
1026 kilobases. *Nature Methods* **6**, 343-345 (2009).
1027 43. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in
1028 Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97**, 6640-6645
1029 (2000).
1030 44. Nouwen, N. *et al.* Secretin PulD: association with pilotin PulS, structure, and ion-
1031 conducting channel formation. *Proc Natl Acad Sci USA* **96**, 8173-8177 (1999).
1032 45. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for
1033 improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

1034 46. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12
1035 (2016).
1036 47. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM
1037 structure determination. *J Struct Biol* **180**, 519-30 (2012).
1038 48. Orlova, S. S. Theory of three-dimensional reconstruction. 1. Conditions of a
1039 complete set of projections. *Sov phys, cryst* **20**, 312-314 (1975).
1040 49. Reichow, S.L., Korotkov, K.V., Hol, W.G. & Gonen, T. Structure of the cholera toxin
1041 secretion channel in its closed state. *Nat Struct Mol Biol* **17**, 1226-32 (2010).
1042 50. Rosenthal, P.B. & Henderson, R. Optimal determination of particle orientation,
1043 absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol*
1044 *Biol* **333**, 721-45 (2003).
1045 51. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of
1046 the IMAGIC image processing system. *J Struct Biol* **116**, 17-24 (1996).
1047 52. Frank, J. *et al.* SPIDER and WEB: Processing and visualization of images in 3D
1048 electron microscopy and related fields. *J Struct Biol* **116**, 190-199 (1996).
1049 53. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*
1050 **9**, 40 (2008).
1051 54. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures
1052 and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).
1053 55. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory
1054 research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
1055 56. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of
1056 Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
1057 57. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for
1058 macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21
1059 (2010).
1060 58. Li, N.N. et al. Structure of the eukaryotic MCM complex at 3.8 angstrom. *Nature*
1061 **524**, 186-191 (2015).
1062 59. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular
1063 crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
1064