



HAL
open science

Structure and Assembly of the Enterohemorrhagic Escherichia coli Type 4 Pilus

Benjamin Bardiaux, Gisele Cardoso de Amorim, Areli Luna Rico, Weili Zheng, Ingrid Guilvout, Camille Jollivet, Michael Nilges, Edward Egelman, Nadia Izadi-Pruneyre, Olivera Francetic

► **To cite this version:**

Benjamin Bardiaux, Gisele Cardoso de Amorim, Areli Luna Rico, Weili Zheng, Ingrid Guilvout, et al.. Structure and Assembly of the Enterohemorrhagic Escherichia coli Type 4 Pilus. Structure, 2019, 27 (7), pp.1082-1093.e5. 10.1016/j.str.2019.03.021 . hal-02326222

HAL Id: hal-02326222

<https://hal.science/hal-02326222>

Submitted on 5 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Structure and assembly of the Enterohemorrhagic *Escherichia coli* type 4 pilus**

2
3
4 3 Benjamin Bardiaux^{1,4}, Gisele Cardoso de Amorim^{2,4,6}, Areli Luna Rico^{1,3,4}, Weili Zheng⁵,
5
6 4 Ingrid Guilvout³, Camille Jollivet³, Michael Nilges¹, Edward H Egelman⁵, Nadia Izadi-
7
8 5 Pruneyre^{1,2*} and Olivera Francetic^{3,7*}

9
10 6 ¹ Structural Bioinformatics Unit, Department of Structural Biology and Chemistry, C3BI,
11
12 Institut Pasteur; CNRS UMR3528; CNRS USR3756; Paris, France,
13
14

15 8 ² NMR of Biomolecules Unit, Department of Structural Biology and Chemistry, Institut
16
17 9 Pasteur, CNRS UMR3528, Paris, France

18
19
20 10 ³Biochemistry of Macromolecular Interactions Unit, Department of Structural Biology and
21
22 11 Chemistry, Institut Pasteur, CNRS UMR3528, Paris, France.

23
24 12 ⁵Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville,
25
26 13 VA22908, USA

27 14 ⁴Equal contribution

28
29
30 15 ⁶Present address: Núcleo Multidisciplinar de Pesquisa em Biologia - NUMPEX-BIO,
31
32 16 Universidade Federal do Rio de Janeiro, Estrada de Xerém, RJ, Brazil.

33 17 ⁷Lead contact: olivera.francetic@pasteur.fr

34
35 18 ****correspondence:** nadia.izadi@pasteur.fr and olivera.francetic@pasteur.fr
36
37

38 19

39
40 20

41
42 21 Key words 10: type 4 pili, EHEC, NMR, cryo-EM, pilin structure, pilus assembly, molecular
43
44 22 modeling, hemorrhagic coli pilus (HCP)

45
46 23

47
48 24

49 25 **Summary**

50
51 26

52
53
54
55 27 Bacterial type 4a pili are dynamic surface filaments that promote bacterial adherence, motility
56
57 28 and macromolecular transport. Their genes are highly conserved amongst enterobacteria
58
59 29 and their expression in Enterohemorrhagic *Escherichia coli* (EHEC) promotes adhesion to
60
61
62

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

intestinal epithelia and pro-inflammatory signaling. To define the molecular basis of EHEC pilus assembly, we determined the structure of the periplasmic domain of its major subunit PpdD (PpdDp), a prototype of enterobacterial pilin subfamily containing two disulfide bonds. The structure of PpdDp, determined by NMR, was then docked into the electron density envelope of purified EHEC pili obtained by cryo-electron microscopy (cryo-EM). Cryo-EM reconstruction of EHEC pili at ~8 Å resolution revealed extremely high pilus flexibility correlating with a large extended region of the pilin stem. Systematic mutagenesis, functional and interaction analyses identified charged residues essential for pilus assembly. Structural information on exposed regions and interfaces between EHEC pilins is relevant for vaccine and drug discovery.

Introduction

Type 4 pili (T4P) are thin filamentous polymers that bacteria use to bind to their substrates, move across surfaces by twitching or gliding, and transport macromolecules in and out of the cell (Berry and Pelicic, 2015). T4P also promote formation of bacterial aggregates and biofilms, and mediate signaling to host cells during infection. The Enterobacteriaceae family of gamma proteobacteria comprises many commensal and environmental species, but also important human, animal and plant pathogens from the genus *Klebsiella*, *Yersinia* or *Dickeya*. Most enterobacteria contain gene clusters encoding pili of the T4a subclass in conserved chromosomal loci (Pelicic, 2008) and share high sequence conservation of major pilin subunits, designated PpdD in *E. coli* (Luna Rico et al., 2019). However, their expression and function have been described only in few species. In the plant pathogen *Erwinia amylovora*, PpdD pili promote biofilm formation and mutants deleted for the *ppdD* operon show reduced virulence (Koczan et al., 2011). In specific nutrient-limiting conditions, PpdD pili are also assembled by enterohemorrhagic *E. coli*

57 (EHEC) (Xicohtencatl-Cortes et al., 2007). EHEC is an important human pathogen causing
58 outbreaks of severe intestinal infections, with hemolytic uremic syndrome (HUS) as a life-
59 threatening complication of the disease. Sera of patients recovering from HUS contain
60 antibodies against PpdD, suggesting that T4P represent one of the virulence factors in the
61 arsenal of this pathogen (Monteiro et al., 2016).

62 The EHEC pili, also named HCP for *hemorrhagic coli pili*, promote adhesion to
63 intestinal epithelia in bovines, the main reservoir of this pathogen (Xicohtencatl-Cortes et al.,
64 2007) and induce pro-inflammatory signaling (Ledesma et al., 2010). Expression of pilin
65 genes in a laboratory *E. coli* strain HB101 promotes biofilm formation and twitching motility
66 (Xicohtencatl-Cortes et al., 2009), phenotypes typically associated with T4P dynamics.
67 Transcriptional studies in *E. coli* K-12 revealed that PpdD pilus assembly genes are co-
68 regulated with DNA uptake genes *via* Sxy (TfoX) competence activator (Sinha et al., 2009),
69 suggesting a role of these pili in natural transformation and horizontal gene transfer. PpdD
70 pilus assembly was achieved in a heterologous type 2 secretion system (T2SS) from
71 *Klebsiella oxytoca* (Sauvonnet et al., 2000; Cisneros et al., 2012), and more recently in the
72 reconstituted EHEC pilus assembly system in *E. coli* K-12 (Luna Rico et al., 2019). The
73 highly similar symmetry parameters of T4aP assembled in these different systems indicate
74 the key role of major pilins as determinants of pilus structure (Luna Rico et al., 2019).

75 EHEC pili are polymers of the major subunit PpdD, which is a prototype of the
76 enterobacterial class of pilins, present in many important human, animal and plant pathogens
77 (Luna Rico et al., 2019). In order to understand the molecular and structural basis of
78 assembly of these major subunits into a pilus, we determined the solution NMR structure of
79 the periplasmic domain of the EHEC PpdD (PpdDp). We then combined this structure with
80 the sub-nanometer resolution cryo-EM reconstruction of EHEC pilus to build an atomic model
81 of the assembled fiber. In addition, site-directed mutagenesis, functional assays and
82 interaction studies identified PpdD residues crucial for pilus assembly and stability.

83

84 **Results**

85

1
2 86 *Solution structure of the periplasmic PpdD domain*
3

4 87 In addition to the well-conserved N-terminal hydrophobic segments (α 1 helix), PpdD
5
6 88 homologues share similar periplasmic domains stabilized by two disulfide bonds at
7
8 89 conserved positions (de Amorim et al., 2014; Luna Rico et al., 2019). To gain molecular
9
10 90 insight into this group of pilins, we solved the structure of the soluble periplasmic domain of
11
12 91 EHEC PpdD (hereafter designated PpdDp) comprising residues 26 to 140 of the mature
13
14 92 protein. To this end we used solution-state NMR, which is particularly adapted for
15
16 93 determination of structure and dynamics of pilin proteins by providing structural insights into
17
18 94 loops and flexible regions, potentially involved in protein-protein and protein-ligand
19
20 95 interactions. The PpdDp structure is very well defined from 2219 NOE-derived distance, 216
21
22 96 dihedral angle and 45 RDC (residual dipolar coupling) restraints as well as 30 hydrogen
23
24 97 bonds. The details of restraints and structural statistics of the 20 lowest energy conformers
25
26 98 representing the solution structure of PpdDp are summarized in Table 1. The structures
27
28 99 show high convergence with a mean pairwise root mean square deviation (RMSD) of 0.55
29
30 100 and 1.06 Å for the backbone and the heavy atoms of ordered regions, respectively. The 3-D
31
32 101 structure of PpdDp displays a canonical alpha-beta pilin fold with two disulfide bonds (Figure
33
34 102 1A,B), largely supported by the dense network of NOE-derived restraints (Figure S1). It is
35
36 103 composed of two α -helices at both extremities (α 1C as the C-terminal part of α 1: N26-H54
37
38 104 and α 3: D123-F134) that are separated by a long α/β loop and a tilted 4-stranded antiparallel
39
40 105 β sheet (strands β 1- β 2- β 3- β 4) facing the α 1C helix. The structure is highly stabilized by two
41
42 106 disulfide bridges (Figure 1A,D); one is conserved in type 4 pilins (C118-C130) and connects
43
44 107 α 3 with β 4. The second (C50-C60) is characteristic for enterobacterial pilins and connects
45
46 108 α 1C to the beginning of the α/β loop leading into a short 3_{10} helix (α 2: G56-C60). This loop is
47
48 109 enriched in His, Gly and Asp residues conserved in the PpdD family.
49
50
51
52
53
54
55

56 110 The ^1H - ^{15}N heteronuclear NOE data for the backbone confirm an overall well-ordered
57
58 111 structure (Figure S2A). Besides the C-terminal tag, which is highly flexible and without any
59
60
61
62
63
64
65

112 stable interaction with the rest of the protein, two regions with low ^1H - ^{15}N NOE value are
113 present, including the $\beta 3/\beta 4$ (G104-G113) and $\beta 4/\alpha 3$ (I120-S122) loops. The former includes
114 the two residues whose signals were not observed in the NMR spectra (D106 and V111)
115 indicating the presence of intermediate exchange on the chemical shift time scale (μs - ms)
116 between different conformations. Consequently, very few long-range NOEs could be
117 assigned in the $\beta 3/\beta 4$ loop, consistent with the high variability of this region in the PpdDp
118 structure ensemble (Figure S1, S2B,C). Surprisingly, residues of the long α/β loop display
119 ^1H - ^{15}N NOE values representative of a well-ordered region with low flexibility and slow
120 internal motions, confirmed by chemical shifts analysis (Figure S2D,E). Numerous short- and
121 long-range NOE restraints are found within the α/β loop and between the loop and the $\beta 1$
122 strand. The loop is anchored at both ends by polar interactions to the extremities of the $\beta 1$
123 and $\beta 2$ strands, while short-range backbone hydrogen bonds stabilize the turns at the N- and
124 C-terminal parts of the α/β loop (Figure S1C).

Structure of the PpdD pilus filament

127 We determined that T4PS PpdD pili and T2SS PpdD pili have the same morphology and
128 indistinguishable helical parameters (Luna Rico et al., 2019). We therefore combined the
129 segments extracted from cryo-EM images of PpdD pili assembled by T4PS and T2SS for
130 helical reconstruction. The variability in the helical parameters was evident (Figure S3A,B)
131 from a reference-based classification for sorting the 99,678 segments in terms of variable
132 twist and rise. The mode of this distribution corresponded to an axial rise of 10.8 Å and twist
133 of 96°, accounting for ~25% of the whole dataset, and this group was used for the helical
134 reconstruction. After multiple cycles of IHRSR (Egelman, 2000), the symmetry of this
135 reconstruction converged to a rise of 11.2 Å and a rotation of ~96° per subunit. The overall
136 resolution of the reconstruction was ~8 Å according to the half-map FSC criterion (Figure
137 S3C,D), allowing us to identify the location of the well-separated globular pilin domains
138 forming the outer shell of the filament (Figure 2A). However, the inner core of the pilus has
139 less density. At high contour level ($\sim 5\sigma$), small rod-shaped densities are observed along the

140 filament axis, but remain disconnected from the main densities of the globular pilin heads
1 (Figure 2B). At lower contour level, weaker densities appear, connecting the outer globular
2 141 heads to the inner rods (Figure 2C). Since the PpdDp NMR structure fitted very nicely into
3
4 142 the cryo-EM density at the expected location but did not explain the inner rod-shape
5
6 143 densities, we inferred that they actually correspond to the N-terminal transmembrane
7
8 144 segments (TMS). We thus built an atomic model of the PpdD pilus by docking a homology
9
10 145 model of the helical TMS, connected by an extended linker to the fitted PpdDp structure.
11
12 146 After symmetrization and refinement in real-space, the structure of the PpdDp was
13
14 147 conserved with a C α RMSD of 2.3 Å between the NMR structure and the structure in the
15
16 148 pilus (Figure 2D,E). Upon refinement, most, if not all, structural changes are limited to loops.
17
18 149 The largest conformational change is the quasi-rigid displacement of the α/β loop towards
19
20 150 the pilin core. In the context of the pilus, the N-terminal part of $\alpha 1$ ($\alpha 1N$) stops at G11 and is
21
22 151 followed by an unstructured region until around P22, where the $\alpha 1C$ helix begins (Figure 2D).
23
24 152
25
26 153
27
28
29
30

31 154 The PpdD pilus has a diameter of ~60 Å with somewhat loosely packed pilin subunits
32
33 155 forming a right-handed 1-start helix (Figure 3A). Each PpdD subunit P has contacts with six
34
35 156 neighbors, forming three different interfaces with subunits $P_{\pm 1}$, $P_{\pm 3}$ and $P_{\pm 4}$ (Figure 3), with
36
37 157 respective buried surface areas of 569 Å², 943 Å² and 1529 Å². Potential salt-bridges are
38
39 158 present in the P- P_{+1} and P- P_{+3} interfaces, while the P- P_{+4} interface mostly involves
40
41 159 hydrophobic contacts via $\alpha 1N$ (Figure 3C). The $\beta 3/\beta 4$ loop, well visible and ordered in the
42
43 160 cryo-EM density, folds as an elongated extension of the β -sheet, making inter-subunit
44
45 161 contacts with the D-region and with the N-terminal region at the P- P_{+4} interface (Figure 3C).
46
47 162
48
49
50

51 163 Co-evolutionary analysis performed on the PpdD pilin sequence family provided intra- and
52
53 164 inter-pilin contact predictions that support the atomic model of PpdD pilus (Table S1; Figure
54
55 165 S4).
56
57
58
59

60 167 *PpdD mutagenesis and effects on pilus assembly*
61
62
63
64
65

168 The mature PpdD, N-terminally processed by the prepilin peptidase has 7 positively
1 and 13 negatively charged residues. The surface of the EHEC pilus is highly negatively
2 169 and 13 negatively charged residues. The surface of the EHEC pilus is highly negatively
3 charged (Figure 4A). To assess the role of electrostatic interactions in assembly and
4 170 charged (Figure 4A). To assess the role of electrostatic interactions in assembly and
5 stabilization of PpdD pili, we introduced charged residue substitutions that create
6 171 stabilization of PpdD pili, we introduced charged residue substitutions that create
7 electrostatic repulsion between PpdD subunits or between PpdD and its assembly
8 172 electrostatic repulsion between PpdD subunits or between PpdD and its assembly
9 components. We examined pilus assembly in an *E. coli* strain producing the *K. oxytoca* T2SS
10 173 components. We examined pilus assembly in an *E. coli* strain producing the *K. oxytoca* T2SS
11 but lacking its cognate major pseudopilin PulG (Figure 4D,E). The mean piliation efficiency,
12 174 but lacking its cognate major pseudopilin PulG (Figure 4D,E). The mean piliation efficiency,
13 expressed as a fraction of total PpdD detected on the bacterial surface, was determined for
14 175 expressed as a fraction of total PpdD detected on the bacterial surface, was determined for
15 each of these variants from four independent experiments. About 50% of native PpdD
16 176 each of these variants from four independent experiments. About 50% of native PpdD
17 (PpdD^{WT}) was assembled into pili in this system, a fraction comparable to the assembly
18 177 (PpdD^{WT}) was assembled into pili in this system, a fraction comparable to the assembly
19 efficiency for the cognate T2SS substrate, the major pseudopilin PulG assembled into pili
20 178 efficiency for the cognate T2SS substrate, the major pseudopilin PulG assembled into pili
21 under the same conditions (Campos et al., 2010). As expected, piliation was abolished for
22 179 under the same conditions (Campos et al., 2010). As expected, piliation was abolished for
23 the PpdD^{E5A} variant lacking the conserved E5 residue, as in *P. aeruginosa* (Strom and Lory,
24 180 the PpdD^{E5A} variant lacking the conserved E5 residue, as in *P. aeruginosa* (Strom and Lory,
25 1991) and *Neisseria gonorrhoeae* (Aas et al., 2007). The E5 residue has been implicated in
26 181 1991) and *Neisseria gonorrhoeae* (Aas et al., 2007). The E5 residue has been implicated in
27 an early step of subunit recruitment to the assembly machinery (Nivaskumar et al., 2016), as
28 182 an early step of subunit recruitment to the assembly machinery (Nivaskumar et al., 2016), as
29 well as in inter-subunit contacts in assembled pili (Kolappan et al., 2016; Wang et al., 2017).
30 183 well as in inter-subunit contacts in assembled pili (Kolappan et al., 2016; Wang et al., 2017).
31 Similar to the behavior of PulG^{E5A}, which is arrested in the membrane-embedded
32 184 Similar to the behavior of PulG^{E5A}, which is arrested in the membrane-embedded
33 preassembly state, (Nivaskumar et al., 2014; Santos-Moreno et al., 2017) cellular PpdD^{E5A}
34 185 preassembly state, (Nivaskumar et al., 2014; Santos-Moreno et al., 2017) cellular PpdD^{E5A}
35 levels were consistently higher compared to PpdD^{WT}. On the other hand, several PpdD
36 186 levels were consistently higher compared to PpdD^{WT}. On the other hand, several PpdD
37 variants showed reduced stability, including PpdD^{R29E} and, more dramatically, PpdD^{R44D},
38 187 variants showed reduced stability, including PpdD^{R29E} and, more dramatically, PpdD^{R44D},
39 consistent with the intra-protomer interaction of R44 with E48 identified by NMR. Charge
40 188 consistent with the intra-protomer interaction of R44 with E48 identified by NMR. Charge
41 inversion of residues mapping onto the fiber surface had little effect on pilus assembly
42 189 inversion of residues mapping onto the fiber surface had little effect on pilus assembly
43 (Figure 4B,C,D,E). With the exception of the surface-exposed residue K83, all positively
44 190 (Figure 4B,C,D,E). With the exception of the surface-exposed residue K83, all positively
45 charged residues were essential for piliation (Figure 4C,D,E). Several charge inversions
46 191 charged residues were essential for piliation (Figure 4C,D,E). Several charge inversions
47 localized at the interfaces with other subunits abolished pilus assembly, including K30E,
48 192 localized at the interfaces with other subunits abolished pilus assembly, including K30E,
49 E48K, D61R, R74D, R116D and R135E. Nearly all residues in the periplasmic part of the
50 193 E48K, D61R, R74D, R116D and R135E. Nearly all residues in the periplasmic part of the
51 alpha helical PpdD stem appear to be very important for piliation (Figure 4D,E).
52 194 alpha helical PpdD stem appear to be very important for piliation (Figure 4D,E).
53
54
55
56
57
58
59
60 195
61
62
63
64
65

196 Among the inter-protomer contacts detected in the cryo-EM structure of the EHEC
197 pilus, R74_{P+1} is in close proximity to both D137_P and D138_P (Figure 3D). Whereas the charge
198 inversion of residue R74 (R74E/D) abolished piliation, the double charge inversion D137R-
199 D138K did not. However, piliation of PpdD^{R74E} and PpdD^{R74D} variants was restored in the
200 presence of the D137R-D138K substitution, supporting the proximity of these residues and
201 their importance at the P-P₊₁ interface (Figure 3E). Similarly, in a non-piliated PpdD variant
202 R135E, pilus assembly could be partially rescued by an additional charge inversion of a non-
203 essential residue E92_{P+1}, facing the R135_P (Figure 3F). This effect was specific since piliation
204 was not restored in another double mutant variant, PpdD^{D35K-R135E} and it is consistent with the
205 relative proximity of R135 and E92 in the pilus structure (Figure 3F).

207 *Role of PpdD interactions with fiber assembly factors*

208 Recent functional reconstitution of the EHEC T4P system (T4PS) (Luna Rico et al.,
209 2019) allowed us to compare assembly of PpdD and its variants *via* the T2SS (Figure 5A, top
210 panel) with assembly *via* its cognate T4PS machinery (Figure 5A, bottom panel). The
211 majority of PpdD variants with residue substitutions in the N-terminal stem region were
212 defective for assembly in both systems. Charge inversions at these positions affected
213 homotypic interactions with PpdD subunits in the membrane, prior to pilus assembly, as
214 indicated by the bacterial two-hybrid (BACTH) analysis (Figure 5B). A notable exception was
215 the PpdD^{E5A} variant, whose ability to form dimers was indistinguishable from wild type
216 (Figure 5B). While fully assembly-defective in T2SS, PpdD^{E5A} was still assembled in the
217 cognate T4PS, albeit less efficiently than native PpdD. Several other variants, including
218 PpdD^{R74D} and PpdD^{R116D}, produced pili with T4PS but not with T2SS, indicating that
219 assembly *via* the cognate system was more robust and less sensitive to mutations.

220 In addition to their effect on docking to the pilus or dimerization of pilins in the
221 membrane, PpdD residue substitutions might differentially affect its interactions with specific
222 assembly factors in the two systems. Heterologous assembly of PpdD *via* the T2SS is likely
223 a result of its ability to interact with several T2SS components, including the assembly

224 protein PulM (Luna Rico et al., 2019). The PulM equivalent HofN is the major PpdD
225 interacting partner in the EHEC T4PS (Luna Rico et al., 2019). To test this possibility, we
226 examined the effect of charged residue substitutions in PpdD on its interactions with HofN
227 and PulM in the BACTH assay (Karimova et al., 1998). Although PpdD^{E5A} retained the ability
228 to interact with the assembly factor HofN, this interaction was significantly reduced compared
229 to PpdD^{WT} (Figure 5C). The charge inversion R74D, which abolished piliation in the T2SS
230 had a limited effect in T4PS; this differential effect apparently correlated with the ability of
231 PpdD^{R74D} variant to interact with HofN, but not with PulM in the BACTH assay (Figure 5C).
232 The assembly-defective PpdD^{D35K} variant, which interacted very poorly with PpdD^{WT} (Figure
233 5B) was also affected for interaction with the assembly factors HofN and PulM. Another
234 assembly-defective variant PpdD^{D135K} showed reduced interaction with HofN and PulM, while
235 retaining the ability to interact with PpdD^{WT} (Figure 5B,C). Together, these data highlight an
236 important role of electrostatic contacts in interactions with other protomers and with the inner
237 membrane assembly factors.

239 *Dynamic behavior of different T4aP*

240 We used normal mode analysis to obtain information on the dynamic behavior of
241 EHEC, *N. meningitidis* (Nm), *N. gonorrhoeae* (Ng) and *P. aeruginosa* strain PAK T4aP. For
242 each T4aP structure, the first 3 non-trivial modes reveal similar dynamic behavior. Modes 1
243 and 2 correspond to the bending of the filaments in two orthogonal directions, while
244 movements along mode 3 represent twisting of the filament around the helical axis (Figure
245 6). For each of the 3 modes, the EHEC T4P structure displays the largest mean square
246 fluctuations of the pilin head per subunit, while Nm, Ng and PAK T4P display similar
247 behaviors of lower fluctuations (Figure 6). As expected, the largest fluctuations are observed
248 at the extremities of the filaments, and in the central region for the bending modes. In the
249 twisting mode (mode 3), the fluctuation pattern along the pilus axis oscillates, with a
250 periodicity of 3 subunits correlating with the left-handed 3-start helix present in all T4aP
251 structures. This observation may indicate the importance of the P-P₊₃ interface in the

252 plasticity of T4aP filaments. Interestingly, the EHEC T4P structure displays a significantly
1 smaller P-P₊₃ interface compared to Nm, Ng and PAK (Figure S5).
2
3

4 254
5

6 255
7

8 256 **Discussion**

10 257
11

12
13 258 The fold of the periplasmic domain of EHEC PpdD is very similar to known structures
14
15 259 of T4a pilins with RMSD for C α atoms ranging from 2.7 to 3.2 Å (Figures 1 and S8 and Table
16
17 S2). Major structural differences are restricted to the characteristic hypervariable regions in
18 260 T4a pilins: the α/β loop, the $\beta 3/\beta 4$ loop and the D-region, encompassed by the two cysteines
19
20 261 forming the C-terminal disulfide bridge (Figure S6). Variable regions confer different surface
21
22 262 shape and charge properties to assembled pili, presumably adapted to their specific binding
23
24 263 properties and biological function. The most notable variation of the D region reflects the
25
26 264 selection pressure in the face of the immune system (Cehovin et al., 2010).
27
28
29 265
30

31 266 Among the pilins, the major pilin PilE from *F. tularensis* (PilE_{Ft}) is the closest structural
32
33 267 homologue to PpdDp despite sharing the lowest sequence identity (only 16% for the
34
35 268 periplasmic domain) (Table S2), indicating the high sequence variability in the structurally
36
37 269 conserved T4a pilins. The PilE_{Ft} X-ray structure shows two molecules in the asymmetric unit
38
39 270 (Hartung et al., 2011) with secondary structure features quite similar to that of PpdDp (Figure
40
41 271 1C). The most pronounced structural difference is an extra strand present in PilE_{Ft} in the α/β
42
43 272 loop involved in the crystallographic dimer interface (Hartung et al., 2011). When the two
44
45 273 molecules of PilE_{Ft} in the crystallographic dimer are superimposed (Figure 1C), the slightly
46
47 274 different conformation in the $\beta 3/\beta 4$ loop between the two molecules suggests some flexibility.
48
49 275 For PpdDp, the flexibility in this region is evidenced by the lower heteronuclear NOE values
50
51 276 (Figure S2). It was shown recently that a short segment in the α/β loop of the N-terminally
52
53 277 truncated PilA pilin of *P. aeruginosa* PA14 is highly dynamic in solution while essential for
54
55 278 piliation (Nguyen et al., 2018). Interestingly, increased backbone dynamics of both the α/β
56
57 279 and $\beta 3/\beta 4$ loops was also observed in monomeric *P. aeruginosa* K122-4 pilin (Suh et al.,
58
59
60
61
62
63
64
65

280 2001). In PpdD, a charge inversion at position D106 within the dynamic $\beta 3/\beta 4$ loop affected
1 pilus assembly specifically in the T2SS, presumably by altering contacts with another PpdD
2 281
3
4 282 protomer (Figure 5) or with an assembly factor.

5
6 283 Flexible regions might be involved in overall pilus dynamics and in the switching
7
8 284 between assembly and retraction modes, required for twitching motility (Clausen et al.,
9
10 285 2009). Identifying all interacting partners of the dynamic loops may shed light on the fine
11
12 286 tuning of these modes, which are driven by two strong and antagonistic ATPase motors at
13
14 287 the base of the system (McCallum et al., 2017). The IM assembly factors PilN and PilO likely
15
16 288 participate in this modulation, although their precise role and mechanism remain unclear
17
18 289 (Leighton et al., 2015).

19
20
21
22 290 So far, detailed cryo-EM reconstructions of bacterial T4aP have been achieved for *N.*
23
24 291 *meningitidis* (Kolappan et al., 2016), *N. gonorrhoeae* and *P. aeruginosa* (Wang et al., 2017).
25
26 292 In all cases, pilin subunits show a similar axial rise of approximately 10 Å, whereas the twist
27
28 293 angle is more variable and reflects specific packing of pilin globular domains. The new cryo-
29
30 294 EM reconstruction of EHEC pili shows very similar subunit organization and helical
31
32 295 parameters to that of other T4aP (Figure 7). As observed for the other T4aP and the related
33
34 296 T2SS pseudopilus structures recently solved, the $\alpha 1$ helix of PpdD is interrupted by a melted
35
36 297 and extended region. However, this extended region is longer in PpdD, going from G11 to
37
38 298 P22, whereas in Ng, Nm and PAK pilins and in *K. oxytoca* pseudopilin, it starts at G14, a
39
40 299 residue which is not conserved in PpdD. We note that the interruption of the $\alpha 1$ helix
41
42 300 correlates well with the T4a pilin-specific sequence-based secondary structure predictions
43
44 301 (Figure S7). This extended region at the $\alpha 1$ helix seems to be a structural feature of T4aP. It
45
46 302 has been proposed that the interruption of the helical part may have a role in the ability of
47
48 303 T4P to extend under force, and relax back to a native state when the force is removed (Wang
49
50 304 et al., 2017). This flexibility, while crucial for pili functions, is an impediment for their structure
51
52 305 determination at high resolution. The long, extended region of PpdD can explain the limited
53
54 306 resolution of the pilus structure. The enhanced flexibility of EHEC pilus towards bending and
55
56 307 twisting from normal mode analysis also indicates the crucial role of the P-P₊₃ interface in
57
58
59
60
61
62
63
64
65

308 pilus rigidity. In EHEC, this reduced interface mainly involves the extended region of the pilin
1 stem, thus supporting its implication in plasticity of pili.
2
3

4 310 Interestingly, when mapped onto the PpdD pilus structure, the location of
5
6 311 hypervariable pilin regions forms a triangular junction between neighboring pilins at the
7
8 312 filament surface (Figure 4F,G). Their exposed residues can define the specificity of
9
10 interactions with different partners.
11

12
13 314 Overall, our mutagenesis and functional analyses show that the majority of charged
14
15 315 PpdD residues required for pilus assembly mapped onto the side of the protein facing the
16
17 316 fiber core, consistent with their involvement in inter-protomer interactions in the polymerized
18
19 317 fiber. In addition, we show that some of these residues affect binding to fiber assembly
20
21 318 factors in the plasma membrane, suggesting their role in early assembly steps. Identifying
22
23 319 and characterizing these steps in dynamic terms should shed new light on substrate
24
25 320 specificity of fiber assembly systems. The complex network of electrostatic interactions that
26
27 321 ensures quality control of pilus assembly and correct pilin docking to the assembly site
28
29 322 requires further structure-function analysis. In the related *Klebsiella* T2SS, binding of the
30
31 323 major pseudopilin PulG to PulM has been implicated in subunit targeting to the assembly
32
33 324 machinery (Nivaskumar et al., 2016; Santos-Moreno et al., 2017). PpdD binding to PulM
34
35 325 probably allows for its assembly *via* the T2SS, whereas the PulM equivalent HofN in the
36
37 326 EHEC T4PS might play a similar role, as suggested by the results of BACTH-based
38
39 327 interaction studies.
40
41
42
43

44 328 EHEC pili have been implicated in colonization of mammalian host epithelia
45
46 329 specifically under starvation conditions and there is evidence for their synthesis *in vivo*
47
48 330 (Xicohtencatl-Cortes et al., 2007). Antibodies against PpdD are produced in patients and
49
50 331 they inhibit pro-inflammatory signaling in a dose-dependent manner (Ledesma et al., 2010).
51
52 332 Therefore, targeting T4P is a possible therapeutic strategy for EHEC infections, which are
53
54 333 life-threatening, and where antibiotic treatment is contraindicated.
55
56

57
58 334 In addition, the EHEC pili share common regulatory and structural features with the
59
60 335 *Vibrio* T4aP that are involved in natural competence (Matthey and Blokesch, 2016). In a
61
62
63
64
65

336 recent elegant study, Ellison and colleagues showed that *Vibrio* T4aP capture DNA *via* their
1 tip and promote its uptake to the periplasm through pilus retraction (Ellison et al., 2018).
2 337
3
4 338 Removing several positively charged residues of two minor pilin subunits, presumably
5
6 339 localized at the tip of the pilus, significantly affected DNA binding and transformation
7
8 340 efficiency (Ellison et al., 2018). Although DNA binding has been attributed to major pilins in
9
10 341 some bacteria like *P. aeruginosa* (van Schaik et al., 2005) or *Neisseria* spp. (Craig et al.,
11 342 2006; Imhaus and Duménil, 2014), direct biochemical evidence supports the role of
12
13 343 specialized minor pilin ComP in this binding in *N. meningitidis* (Cehovin et al., 2013).
14
15 344 Consistent with this model, we did not observe any interaction of PpdDp with double
16
17 345 stranded DNA by NMR. Moreover, the highly negatively charged surface of the EHEC pilus
18
19 346 makes its direct interaction with DNA unlikely. PpdD is highly similar to another pilin involved
20
21 347 in natural competence, the *Acinetobacter baumannii* PilA, whose recent structural analysis
22
23 348 revealed similar features including two disulfide bridges and a highly negatively charged
24
25 349 surface (Ronish et al., 2019). Surface charge affected the capacity of PilA to form bundles
26
27 350 and lead to microcolony formation, presumably through electrostatic repulsion (Ronish et al.,
28
29 351 2019). In comparison, PAK, Ng and Nm pili display more mixed surface electrostatics, with
30
31 352 almost continuous positive (PAK) and negative (Ng) patches (Figure S8). These might favor
32
33 353 pilus bundling, but also interactions with specific surfaces and substrates.
34
35
36
37
38
39
40
41

42 354
43
44 355 Specific nutritional or other signals so far unknown may activate DNA uptake *in vivo*,
45 356 as suggested by the capacity of *E. coli* to take up exogenous DNA (Sinha and Redfield,
46 357 2012). Given the high conservation of PpdD and T4aP systems in enterobacteria, this
47 358 capacity might be widespread among commensal strains, and among pathogens of the
48
49 359 ESKAPE group like *Klebsiella*, *Salmonella* or *Yersinia* that represent a particular threat due
50
51 360 to their antibiotic resistance. Further studies will be required to determine whether and how
52
53 361 T4P in enterobacteria participate in DNA uptake and contribute to exchange of genetic
54
55 362 material in the gut. Detailed structural information will be instrumental to gain a better
56
57
58
59
60
61
62
63
64
65

363 understanding of the mechanisms underlying these processes *in vivo*, and is crucial for drug
1
2 364 and vaccine discovery.
3

4 365

6 366 **Data availability**

8 367

10 368 The PpdDp structure and NMR restraints have been deposited in the Protein Data Bank
11
12 369 under the accession code 6GMS. The PpdD pilus structure and cryo-EM map are available
13
14 370 from the Protein Data Bank (accession code 6GV9) and Electron Microscopy Data Bank
15
16 371 (accession code EMD-0070), respectively.
17

18 372

20 373 **Acknowledgments**

21 374

22 375 This work was funded by the Institut Pasteur, the Centre National de la Recherche
23
24 376 Scientifique (CNRS), the French Agence Nationale de la Recherche (ANR-14-CE09-0004 to
25
26 377 O.F), the Fondation pour la Recherche Médicale (Equipe FRM 2017M.DEQ20170839114 to
27
28
29 378 M.N), the European Union FP7-IDEAS-ERC 294809 (to M.N) and the NIH R35GM122510 (to
30
31 379 E.H.E.). ALR was funded by the Pasteur Paris University (PPU) international PhD program.
32
33 380 We are grateful to Muriel Delepierre and Daniel Ladant for support and interest in this work.
34

35 381

37 382 **Author contributions**

38 383

40 384 Conceived and designed the experiments: BB, ALR, MN, EHE, NIP and OF. Performed the
41
42 385 experiments: BB, GCA, ALR, WZ, CJ, OF. Analyzed the data: GCA, ALR, BB, WZ, MN, EHE,
43
44 386 NIP and OF. Wrote the manuscript: BB, ALR, WZ, MN, EHE, NIP and OF.
45

46 387

47 388

50 389 **Declaration of interest**

51 390

52 391 The authors declare no competing interests.

53 392

54 393

55 394 **References**

56 395

57 396

59 397 Aas, F.E., Winther-Larsen, H.C., Wolfgang, M., Frye, S., Lovold, C., Roos, N., van Putten,
60 398 J.P., and Koomey, M. (2007). Substitutions in the N-terminal alpha helical spine of *Neisseria*
61
62
63
64
65

399 *gonorrhoeae* pilin affect Type IV pilus assembly, dynamics and associated functions. Mol
1 400 Microbiol *63*, 69-85.

2 401 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., and Echols, N., et al.
3 402 (2010). PHENIX: a comprehensive Python-based system for macromolecular structure
4 403 solution. Acta Crystallogr D, Biol Crystallogr, *66*, 213–221.

5 404 Bakan, A., Dutta, A., Mao, W., Liu, Y., Chennubhotla, C., Lezon, T.R., and Bahar, I. (2014).
6 405 Evol and ProDy for bridging protein sequence evolution and structural dynamics.
7 406 Bioinformatics *30*, 2681–2683.

8 407 Baker, N., Sept, D., Joseph, S., Holst, M., and McCammon, J. (2001). Electrostatics of
9 408 nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*,
10 409 10037-10041.

11 410 Bardiaux, B., Bernard, A., Rieping, W., Habeck, M., Malliavin, T.E., and Nilges, M. (2008).
12 411 Influence of different assignment conditions on the determination of symmetric homodimeric
13 412 structures with ARIA. Proteins *75*, 569-585.

14 413 Bartolome, B., Y. Jubete, E. Martinez & F. de la Cruz. (1991). Construction and properties of
15 414 a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives.
16 415 Gene *102*, 75-78.

17 416 Bernard, A., Vranken, W.F., Bardiaux, B., Nilges, M., and Malliavin, T.E. (2011). Bayesian
18 417 estimation of NMR restraint potential and weight: a validation on a representative set of
19 418 protein structures. Proteins *79*, 1525–1537.

20 419 Berry, J.L., and Pelicic, V. (2015). Exceptionally widespread nanomachines composed of
21 420 type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol Rev *39*, 134-154.

22 421 Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat Protoc *2*,
23 422 2728-2733.

24 423 Campos, M., Nilges, M., Cisneros, D.A., and Francetic, O. (2010). Detailed structural and
25 424 assembly model of the type II secretion pilus from sparse data. Proc Natl Acad Sci U S A
26 425 *107*, 13081–13086.

27 426 Cehovin, A., Simpson, P.J., McDowell, M.A., Brown, D.R., Noschese, R., Pallett, M., Brady,
28 427 J., Baldwin, G.S., Lea, S.M., Matthews, S.J., and Pelicic, V. (2013). Specific DNA recognition
29 428 mediated by a type IV pilin. Proc Natl Acad Sci U S A *110*, 3065-3070.

30 429 Cehovin, A., Winterbotham, M., Lucidarme, J., Borrow, R., Tang, C.M., Exley, R.M., and
31 430 Pelicic, V. (2010). Sequence conservation of pilus subunits in *Neisseria meningitidis*. Vaccine
32 431 *28*, 4817–4826.

33 432 Cisneros, D.A., Pehau-Arnaudet, G., and Francetic, O. (2012). Heterologous assembly of
34 433 type IV pili by a type II secretion system reveals the role of minor pilins in assembly initiation.
35 434 Mol Microbiol *86*, 805-818.

36 435 Clausen, M., Koomey, M., and Maier, B. (2009). Dynamics of type IV pili is controlled by
37 436 switching between multiple states. Biophys J *96*, 1169-1177.

38 437 Craig, L., Volkman, N., Arvai, A.S., Pique, M.E., Yeager, M., Egelman, E.H., and Tainer,
39 438 J.A. (2006). Type IV pilus structure by cryo-electron microscopy and crystallography:
40 439 implications for pilus assembly and functions. Mol Cell *23*, 651-662.

41 440 Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in
42 441 *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640–6645.

43 442 Dautin, N., G. Karimova, A. Ullmann and D. Ladant, (2000) Sensitive genetic screen for
44 443 protease activity based on a cyclic AMP signaling cascade in *Escherichia coli*. J Bacteriol
45 444 *182*, 7060-7066.

46 445 Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., and al., e.
47 446 (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.
48 447 Nucl. Acids Res. *35*, W375–383.

49 448 de Amorim, G.C., Cisneros, D.A., Delepierre, M., Francetic, O., and Izadi-Pruneyre, N.
50 449 (2014). ¹H, ¹⁵N and ¹³C resonance assignments of PpdD, a type IV pilin from
51 450 enterohemorrhagic *Escherichia coli*. Biomol NMR Assign *8*, 43-46.

52 451 Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a
53 452 multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-
54 453 293.

454 Egelman, E.H. (2000). A robust algorithm for the reconstruction of helical filaments using
1 455 single-particle methods. *Ultramicroscopy* 85, 225-234.
2 456 Ellison, C.K., Dalia, T.N., Ceballos, A.V., Wang, J.C., Biais, N., Brun, Y.V., and Dalia, A.B.
3 457 (2018). Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural
4 458 transformation in *Vibrio cholerae*. *Nat Microbiol* 3, 773-780.
5 459 Francetić, O., Pugsley A.P. (2005). Towards the identification of type II secretion signals in a
6 460 nonacylated variant of pullulanase from *Klebsiella oxytoca*. *J Bacteriol.* 187, 7045-55.
7 461 Frank, J. *et al.* (1996). SPIDER and WEB: Processing and visualization of images in 3D
8 462 electron microscopy and related fields. *J Struct Biol* 116, 190-199.
9 463 Hartung, S., Arvai, A.S., Wood, T., Kolappan, S., Shin, D.S., Craig, L., and Tainer, J.A.
10 464 (2011). Ultrahigh resolution and full-length pilin structures with insights for filament assembly,
11 465 pathogenic functions, and vaccine potential. *J Biol Chem* 286, 44254-44265.
12 466 Hubbard, S.J., and Thornton, J.M. (1993). NACCESS Computer Program. Department of
13 467 Biochemistry and Molecular Biology, University College London.
14 468 Imhaus, A.-F., and Duménil, G. (2014). The number of *Neisseria meningitidis* type IV pili
15 469 determines host cell interaction. *EMBO J* 33, 1767-1783.
16 470 Jones, D. (1999). Protein secondary structure prediction based on position-specific scoring
17 471 matrices. *J Mol Biol* 292, 195-202.
18 472 Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-
19 473 based residue-residue contact predictions in a sequence- and structure-rich era. *Proc Natl*
20 474 *Acad Sci U S A* 110, 15674–15679.
21 475 Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system
22 476 based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* 95, 5752-
23 477 5756.
24 478 Koczan, J.M., Lenneman, B.R., McGrath, M.J., and Sundin, G.W. (2011). Cell surface
25 479 attachment structures contribute to biofilm formation and xylem colonization by *Erwinia*
26 480 *amylovora*. *Appl. Environ. Microbiol.* 77, 7031–7039.
27 481 Kolappan, S., Coureuil, M., Yu, X., Nassif, X., Egelman, E.H., and Craig, L. (2016). Structure
28 482 of the *Neisseria meningitidis* Type IV pilus. *Nat Commun* 7, 13015.
29 483 Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein
30 484 side-chain conformations with SCWRL4. *Proteins* 77, 778–795.
31 485 Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a
32 486 program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-
33 487 291.
34 488 Ledesma, M.A., Ochoa, S.A., Cruz, A., Rocha-Ramírez, L.M., Mas-Oliva, J., Eslava, C.A.,
35 489 Girón, J.A., and Xicohtencatl-Cortes, J. (2010). The Hemorrhagic Coli Pilus (HCP) of
36 490 *Escherichia coli* O157:H7 is an inducer of proinflammatory cytokine secretion in intestinal
37 491 epithelial cells. *PLoS One* 5, e12127.
38 492 Leighton, T.L., Dayalani, N., Sampaleanu, L.M., Howell, P.L., and Burrows, L.L. (2015).
39 493 Novel role for PilNO in type IV pilus retraction revealed by alignment subcomplex mutations.
40 494 *J. Bacteriol.* 197, 2229-2238.
41 495 Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M., and Nilges, M. (2003). Refinement of
42 496 protein structures in explicit solvent. *Proteins* 50, 496-506.
43 497 Lopez-Castilla, A., Thomassin, J.-L., Bardiaux, B., Zheng, W., Nivaskumar, M., Yu, X.,
44 498 Nilges, M., Egelman, E.H., Izadi-Pruneyre, N., and Francetic, O. (2017). Structure of the
45 499 calcium-dependent type 2 secretion pseudopilus. *Nat Microbiol* 2, 1686-1695.
46 500 Luna Rico, A Thomassin, J.-L. and Francetic, O. (2018). Analysis of bacterial pilus assembly
47 501 by shearing and immunofluorescence microscopy. *Methods Mol Biol* 1764, 291-305.
48 502 Luna Rico, A., Zheng, W., Petiot, N., Egelman, E.H., and Francetic, O. (2019). Functional
49 503 reconstitution of the type IVa assembly system from enterohemorrhagic *Escherichia coli*.
50 504 *Mol. Microbiol.* 2018 Dec 18. doi: 10.1111/mmi.14188. [Epub ahead of print]
51 505 Matthey, N., and Blokesch, M. (2016). The DNA-uptake process of naturally competent
52 506 *Vibrio cholerae*. *Trends Microbiol.* 24, 98-110.
53 507 McCallum, M., Tammam, S., Khan, A., Burrows, L.L., and Howell, P.L. (2017). The molecular
54 508 mechanism of the type IVa pilus motors. *Nat Commun* 8, 15091.
55
56
57
58
59
60
61
62
63
64
65

509 Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and
1 510 specimen tilt in electron microscopy. *J Struct Biol.* 142, 334-347.
2 511 Monteiro, R., Ageorges, V., Rojas-Lopez, M., Schmidt, H., Weiss, A., Bertin, Y., Forano, E.,
3 512 Jubelin, G., Henderson, I.R., Livrelli, V., *et al.* (2016). A secretome view of colonisation
4 513 factors in Shiga toxin-encoding *Escherichia coli* (STEC): from enterohaemorrhagic *E. coli*
5 514 (EHEC) to related enteropathotypes. *FEMS Microbiol Rev* 363, pii: fnw179.
6 515 Nguyen, Y., Boulton, S., McNicholl, E.T., Akimoto, M., Harvey, H., Aidoo, F., Melacini, G.,
7 516 and Burrows, L.L. (2018). A highly dynamic loop of the *Pseudomonas aeruginosa* PA14 type
8 517 IV pilin is essential for pilus assembly. *ACS Infect Dis.* 4, 936-943.
9 518 Nilges, M., Bernard, A., Bardiaux, B., Malliavin, T., Habeck, M., and Rieping, W. (2008).
10 519 Accurate NMR structures through minimization of an extended hybrid energy. *Structure* 16,
11 520 1305-1312.
12 521 Nivaskumar, M., Bouvier, G., Campos, M., Nadeau, N., Yu, X., Egelman, E.H., Nilges, M.,
13 522 and Francetic, O. (2014). Distinct docking and stabilization steps of the pseudopilus
14 523 conformational transition path suggest rotational assembly of type IV pilus-like fibers.
15 524 *Structure* 22, 685-696.
16 525 Nivaskumar, M., Santos-Moreno, J., Malosse, C., Nadeau, N., Chamot-Rooke, J., Tran Van
17 526 Nhieu, G., and Francetic, O. (2016). Pseudopilin residue E5 is essential for recruitment by
18 527 the type 2 secretion system assembly platform. *Mol Microbiol* 101, 924-941.
19 528 Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and dipolar couplings from
20 529 simplified two-dimensional NMR spectra. *J Magn Reson* 131, 373-378.
21 530 Pelicic, V. (2008). Type IV pili: e pluribus unum? *Mol Microbiol* 68, 827-837.
22 531 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C.,
23 532 and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and
24 533 analysis. *J Comput Chem* 25, 1605–1612.
25 534 Possot, O.M., Vignon, G., Bomchil, N., Ebel, F., Pugsley, A.P. (2000) Multiple interactions
26 535 between pullulanase secreton components involved in stabilization and cytoplasmic
27 536 membrane association of PulE. *J Bacteriol.* 182, 2142-52.
28 537 Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative
29 538 protein sequence searching by HMM-HMM alignment. *Nat Methods* 9, 173–175.
30 539 Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007).
31 540 ARIA2: automated NOE assignment and data integration in NMR structure calculation.
32 541 *Bioinformatics* 23, 381-382.
33 542 Ronish, L., Lillehoj, E., JK, F., Sundberg, E., and Piepenbrink, K. (2019). The structure of
34 543 PilA from *Acinetobacter baumannii* AB5075 suggests a mechanism for functional
35 544 specialization in *Acinetobacter* type IV pili *J Biol Chem.* 294, 218–230.
36 545 Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial
37 546 restraints. *J Mol Biol* 234, 779–815.
38 547 Santos-Moreno, J., East, A., Bond, P.J., Tran Van Nhieu, G., and Francetic, O. (2017). Polar
39 548 N-terminal residues conserved in type 2 secretion pseudopilins determine subunit targeting
40 549 and membrane extraction during fibre assembly. *J Mol Biol* 429, 1746-1765.
41 550 Sauvonnet, N., Gounon, P., and Pugsley, A.P. (2000). PpdD Type IV Pilin of *Escherichia coli*
42 551 K-12 Can Be Assembled into Pili in *Pseudomonas aeruginosa*. *J Bacteriol* 182, 848–854.
43 552 Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel
44 553 electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*
45 554 166, 368-379.
46 555 Schrödinger, L. (2015). The PyMOL Molecular Graphics System, Version 1.8.
47 556 Seemayer, S., Gruber, M., and Söding, J. (2014). CCMpred—fast and precise prediction of
48 557 protein residue-residue contacts from correlated mutations. *Bioinformatics* 30, 3128–3130.
49 558 Shapovalov, M.V., and Dunbrack, R.L. (2011). A smoothed backbone-dependent rotamer
50 559 library for proteins derived from adaptive kernel density estimates and regressions. *Structure*
51 560 19, 844–858.
52 561 Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for
53 562 predicting protein backbone torsion angles from NMR chemical shifts. *J Biomolec NMR* 44,
54 563 213-223.
55
56
57
58
59
60
61
62
63
64
65

564 Sinha, S., Cameron, A.D.S., and R.J., R. (2009). Sxy induces a CRP-S regulon in
1 565 *Escherichia coli*. *J Bacteriol* *191*, 5180–5195.
2 566 Sinha, S., and Redfield, R.J. (2012). Natural DNA uptake by *Escherichia coli*. *PLoS One* *7*,
3 567 e35620.
4 568 Strom, M.S., and Lory, S. (1991). Amino acid substitutions in pilin of *Pseudomonas*
5 569 *aeruginosa*. Effect on leader peptide cleavage, amino-terminal methylation, and pilus
6 570 assembly. *J Biol Chem* *266*, 1656-1664.
7 571 Suh, J.Y., Spyropoulos, L., Keizer, D.W., Irvin, R.T., and Sykes, B.D. (2001). Backbone
8 572 dynamics of receptor binding and antigenic regions of a *Pseudomonas aeruginosa* pilin
9 573 monomer. *Biochemistry* *40*, 3985-3995.
10 574 Tang, G. *et al.* (2007). EMAN2: an extensible image processing suite for electron
11 575 microscopy. *J Struct Biol* *157*, 38-46.
12 576 Toth-Petroczy, A., Palmedo, P., Ingraham, J., Hopf, T.A., Berger, B., Sander, C., and Marks,
13 577 D.S. (2016). Structured states of disordered proteins from genomic sequences. *Cell* *167*,
14 578 158–170. e112.
15 579 van Schaik, E.J., Giltner, C.L., Audette, G.F., Keizer, D.W., Bautista, D.L., Slupsky, C.M.,
16 580 Sykes, B.D., and Irvin, R.T. (2005). DNA binding: a novel function of *Pseudomonas*
17 581 *aeruginosa* type IV pili. *J. Bacteriol.* *187*, 1455–1464.
18 582 Toth-Petroczy, A., Palmedo, P., Ingraham, J., Hopf, T.A., Berger, B., Sander, C., and Marks,
19 583 D.S. (2016). Structured states of disordered proteins from genomic sequences. *Cell* *167*,
20 584 158–170.e112.
21 585 Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L.,
22 586 Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR
23 587 spectroscopy: development of a software pipeline. *Proteins* *59*, 687-696.
24 588 Wang, F., Coureuil, M., Osinski, T., Orlova, A., Altindal, T., Gesbert, G., Nassif, X., Egelman,
25 589 E.H., and Craig, L. (2017). Cryoelectron Microscopy Reconstructions of the *Pseudomonas*
26 590 *aeruginosa* and *Neisseria gonorrhoeae* Type IV Pili at Sub-nanometer Resolution. *Structure*.
27 591 *25*, 1423-1435.
28 592 Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L.,
29 593 and Sykes, B.D. (1995). ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J*
30 594 *Biomol NMR* *6*, 135-140.
31 595 Wriggers, W. (2012). Conventions and workflows for using Situs. *Acta Crystallogr D, Biol*
32 596 *Crystallogr*, *68(Pt 4)*, 344–351.
33 597 Xicohtencatl-Cortes, J., Monteiro-Neto, V., Ledesma, M.A., Jordan, D., Francetic, O., Kaper,
34 598 J.B., Puente, J.L., and Girón, J.A. (2007). Intestinal adherence associated with type IV pili of
35 599 enterohemorrhagic *Escherichia coli* O157:H7. *J Clin. Investig.* *117*, 3519–3529.
40 600 Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldana, Z., Ledesma, M.A., Puente, J.L., and
41 601 Giron, J.A. (2009). The type 4 pili of Enterohemorrhagic *Escherichia coli* O157:H7 are
42 602 multipurpose Structures with pathogenic attributes. *J Bacteriol* *191*, 411-421.
43 603 Yanisch-Perron, C., Vieira, J., and J., M. (1985). Improved M13 phage cloning vectors and
44 604 host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* *33*, 103-119.
45 605 Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based
46 606 on the TM-score. *Nucleic Acids Res* *33*, 2302–2309.
47 607 Zweckstetter, M., and Bax, A. (2000). Prediction of Sterically Induced Alignment in a Dilute
48 608 Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.*
49 609 *122*, 3791-3792.

51 610 52 611 53 612 **Main figure titles and legends**

54 613 55 614 **Figure 1. Solution NMR structure of PpdDp.**

57 615 **(A)** Backbone representation of the NMR ensemble of PpdDp colored by secondary structure
58
59 616 elements. Helices α 1C, α 2 and α 3 are shown in red and the β 1- β 2- β 3- β 4 sheet is shown in

617 cyan. Disulfide bridges are colored in yellow. **(B)** Ribbon representation of the lowest energy
618 model of the PpdDp NMR ensemble. Characteristic hypervariable regions in T4a major pilins
619 are colored in B, C and D as follows: α/β loop in green, $\beta 3/\beta 4$ loop in cyan and D-region in
620 magenta. Disulfide bridges C50-C60 and C118-C130 are shown as yellow sticks. **(C)**
621 Superposition of the two molecules in the asymmetric unit of the *Francisella tularensis* Pile
622 crystallographic structure (PDB 3SOJ). The conformationally variable $\beta 3/\beta 4$ loop is
623 highlighted by a dotted circle. **(D)** Sequence alignment of full-length PpdD from EHEC, Pile
624 pilins from *F. tularensis* (Ft) and *Neisseria gonorrhoeae* (Ng). Positions of helix breaking
625 residues (G, P) in the N-terminal region are highlighted with red arrows. Cysteine residues
626 involved in disulfide bridges are connected by blue lines. See also Figures S2, S6 and Table
627 S2.

630 **Figure 2. Cryo-EM reconstruction of the EHEC T4P filament at ~ 8 Å resolution and** 631 **PpdD pilin structure fitting.**

632 **(A)** Surface view of the EHEC T4P cryo-EM reconstruction at $\sim 4.2\sigma$ contour level with the
633 refined atomic model shown in ribbons (red). The diameter of the pilus is ~ 60 Å. **(B-C)** Cross-
634 sections of the cryo-EM reconstruction at $\sim 5\sigma$ (B) and $\sim 3.8\sigma$ (C) contour levels and with a
635 single PpdD subunit shown in ribbons. **(D)** Structure of a PpdD pilin subunit in the refined
636 EHEC T4P structure colored by secondary structures and showing the extended region
637 between the short $\alpha 1N$ helix and the PpdDp domain. The positions of helix breaking between
638 residues G11 and P22 are indicated. **(E)** Superposition of the lowest-energy conformer of
639 monomeric PpdDp in solution (blue) with the refined pilus structure (red) shown in two
640 orientations. See also Figure S3.

642 **Figure 3. Structure of the EHEC PpdD pilus.**

643 **(A)** Surface representation of the PpdD pilus structure with subunits P, P_{+1} , P_{+2} , P_{+3} and P_{+4}
644 (along the 1-start helix) colored in orange, cyan, green, red and magenta, respectively. **(B)**
645 Structure of PpdD pilus shown in ribbons where subunits are colored as in (A). The helical
646 rise of the 1-start helix is 11.2 Å. **(C)** Topological arrangement of neighboring subunits in the
647 PpdD pilus structure, colored as in (A). The positions of charged residues involved in inter-
648 subunit interactions are labeled with colored outline. Residues making hydrophobic contacts
649 between subunits P and P_{-3}/P_{-4} are labelled in black. **(D)** Close-up view of the interface
650 between protomer P and P_{+1} showing the side-chains of D137/D138 and R74 making
651 potential salt-bridges in the PpdD pilus structure. The cryo-EM reconstruction is shown as
652 surface at $\sim 3.2\sigma$ contour level. **(E-F)** Piliation assay with single charge inversion PpdD
653 variants. Cells (top) and sheared pilus fractions (Pili, bottom) corresponding to 0.05 OD_{600nm}

654 of bacteria were separated on SDS-PAGE and analyzed by immunoblot using anti-PpdD
655 antibody. PpdD residue substitutions are indicated above each lane. Migration of PpdD is
656 indicated on the right. See also Figure S4 and Table S1.

657
658 **Figure 4. Mapping of residues essential for pilus assembly and hypervariable regions**
659 **on the PpdD pilus surface.**

660 **(A)** Surface electrostatic potential of the PpdD pilus **(B)** Surface representation of the PpdD
661 pilus structure showing the distribution of charged residues essential for pilus assembly
662 (positively charged in blue and negatively charged in red). Non-essential charged residues
663 are colored in green. Surface of a single PpdD subunit is outlined in black. **(C)** Surface
664 representation of a PpdD pilin subunit in the pilus colored as in (A). The surface-exposed
665 side of the subunit (front view) displays predominantly residues permissive for charge
666 inversions. The back view (180° rotation) shows the clustering of charged residues essential
667 for pilus assembly on the buried face of the pilin. **(D)** Piliation assay with single charge
668 inversion PpdD variants. Cells (top) and sheared pilus fractions (Pili, bottom) corresponding
669 to 0.05 OD_{600nm} of bacteria were separated on SDS-PAGE and analyzed by immunoblot
670 using anti-PpdD antibodies. **(E)** Quantification of the fraction of PpdD assembled into pili.
671 The bars represent the mean values and error bars represent standard deviation from 4
672 independent experiments. Statistical analysis was performed by using ANOVA and Kruskal-
673 Wallis multiple comparison test to compare datasets from native PpdD and its variants: ns:
674 non-significant difference; **: p<0.01; ****: p<0.0001. **(F)** Sphere representation of the PpdD
675 pilus structure where the location of hypervariable regions is highlighted (α/β loop in green,
676 $\beta 3/\beta 4$ loop in cyan and D-region in magenta). A single PpdD subunit is outlined in black. **(G)**
677 Schematic representation of the topological arrangement of neighboring subunits in the
678 PpdD pilus structure. The location of hypervariable regions is indicated and colored as in (F).
679 The size of the buried surface area in each inter-protomer interface is proportional to the
680 width of the grey line.

681
682 **Figure 5. Effect of charged residue substitutions on pilus assembly and on**
683 **interactions with assembly factors in T2SS and T4PS. (A)** Pilus assembly of PpdD^{WT} and

684 indicated variants in the T2SS (top) and T4PS (bottom panel). Cells and sheared fractions
685 (Pili) corresponding to 0.05 OD_{600nm} of bacteria were separated on SDS-PAGE and analyzed
686 by immunoblot using anti-PpdD antibodies. **(B)** Bacterial two-hybrid (BACTH) analysis of
687 interactions between PpdD^{WT} and PpdD mutant derivatives containing charge inversion at
688 indicated positions. Beta-galactosidase activity of *E. coli* strains DHT1 (*cyaA*) co-transformed
689 with plasmids pCHAP8501 encoding the T18-PpdD hybrid protein and their derivatives with

690 charge inversion mutations at indicated positions and pCHAP8504 encoding T25-PpdD^{WT}
691 hybrid protein. (C) Interactions of PpdD and its variants with the assembly component HofN
692 of the T4PS and PulM of the T2SS in the BACTH assay. Beta-galactosidase activity of *E. coli*
693 DHT1 strain co-transformed with indicated plasmids encoding the T18- and T25- hybrid
694 proteins and their mutant derivatives (listed in Table S3). BACTH analysis was performed as
695 described previously (Nivaskumar et al., 2016). Bars represent mean values and error bars
696 standard deviations from at least 5 independent clones. Comparison of mean values
697 indicated for selected pairs was performed using ANOVA and Kruskal-Wallis multiple
698 comparison test: **, p<0.01; ***, p<0.001; ****, p<0.0001.

Figure 6. Normal mode analysis of T4aP filament structures. Left: Average of the atomic
701 mean square fluctuations per subunit (pilin head) along the pilus axis for the first 3 non-trivial
702 slowest modes (A, mode 1; B, mode 2; C, mode 3) in the EHEC (orange), Nm (red), Ng
703 (blue) and Pa PAK (green) T4aP structures. For each mode, the overall direction is
704 represented by the black arrows surrounding the pilus structure. (D) Representations of the
705 individual modes per atom (orange arrows) on the EHEC T4aP structure (top: top view,
706 bottom: side view). See also Figure S5.

Figure 7: Comparison of T4a pili and T2SS pseudo-pilus structures.

709 From left to right, pilus structure of EHEC (present work), *N. gonorrhoeae* (Wang et al.,
710 2017), *P. aeruginosa* (Wang et al., 2017), *N. meningitidis* (Kolappan et al., 2016) and *K.*
711 *oxytoca* (Lopez-Castilla et al., 2017). Top row: Pilus organization with major pilin colored
712 along the 1-start helix (subunits P in orange, P₊₁ in cyan, P₊₂ in green, P₊₃ in red and P₊₄ in
713 magenta). Bottom row: Pilin structures in the pili. The positions of conserved Glycine
714 residues (G11 for EHEC, G14 for others) and Proline 22 (P22) are labeled. See also Figures
715 S7 and S8.

719 **Table 1: PpdDp NMR structure statistics and restraints** See also Figure S1.

720	Number of restraints	
721	NOE Distance restraints	
722	Intra-residue ($ i-j = 0$)	570
723	Sequential ($ i-j = 1$)	293
724	Medium-range ($2 \leq i-j < 5$)	363
725	Long-range ($ i-j \geq 5$)	294
726	Ambiguous	499
727	<i>Total</i>	2219
728	Dihedral angle restraints (φ/ψ)	216 (108/108)
729	Hydrogen bonds restraints	30
730	RDC restraints (D^{NH})	45
731	Restraints statistics^a	
732	RMS of distance violations	
733	NOE restraints	$0.15 \pm 0.05 \text{ \AA}$
734	H-bonds restraints	$0.22 \pm 0.04 \text{ \AA}$
735	RMS of dihedral violations	$1.47 \pm 0.11^\circ$
736	RMS of RDC violations	$1.67 \pm 0.16 \text{ Hz}$
737	Q-factor RDC	0.098 ± 0.003
738	RMS from idealized covalent geometry	
739	bonds	$0.003 \pm 0.001 \text{ \AA}$
740	angles	$0.484 \pm 0.007^\circ$
741	impropers	$1.509 \pm 0.055^\circ$
742	Structural quality^a	
743	<i>Ramachandran statistics^b</i>	
744	Most favoured regions	$86.1 \pm 1.7 \%$
745	Allowed regions	$12.9 \pm 1.6 \%$
746	Generously allowed regions	$0.7 \pm 0.6 \%$
747	Disallowed regions	$0.3 \pm 0.4 \%$
748	<i>WHAT-IF Z-score^c</i>	
749	Backbone conformation	-0.17 ± 0.40
750	2 nd generation packing quality	-1.78 ± 0.29
751	Ramachandran plot appearance	-2.46 ± 0.28
752	χ_1/χ_2 rotamer normality	-2.99 ± 0.54
753	Coordinates precision^d	
754	All backbone atoms (26-138)	$1.51 \pm 0.35 \text{ \AA}$
755	All heavy atoms (26-138)	$1.93 \pm 0.25 \text{ \AA}$
756	All backbone atoms (secondary structures)	$0.55 \pm 0.27 \text{ \AA}$
757	All heavy atoms (secondary structures)	$1.06 \pm 0.13 \text{ \AA}$

721 ^a Average values and standard deviations over the 20 conformers.722 ^b Percentage of residues in the Ramachandran plot regions determined by PROCHECK (Laskowski et

723 al., 1993).

724 ^c Z-scores values reported by WHAT-IF (Vriend, 1990).725 ^d Average root mean square deviation (RMSD) over the 20 conformers' atomic coordinates with

726 respect to the average structure.

730 **STAR Methods**

1
2 731

3 732 *Bacterial strains and plasmids*

4
5
6 733 *Escherichia coli* strain DH5 α (F' *lacI*^Q Δ *lacZM15 pro+*) was used for DNA cloning purposes
7
8 734 and strains PAP7460 (Δ (*lac-argF*)*U169 araD139 relA1 rpsL150 Δ malE444 malG501* [F' (*lacI*^Q
9
10 735 Δ *lacZM15 pro+ Tn10*)] and BW25113 [Δ (*araD-araB*)*567, Δ lacZ4787(::rrnB-3), λ^- , rph-1,*
11
12 736 Δ (*rhaDrhaB*)*568, hsdR514*)] F' *lacI*^Q (Datsenko and Wanner, 2000) were used for functional
13
14 737 assays. Strain PAP5171 (PAP7460 *degP::Km*^R) was used as a host for pilus purification.
15
16 738 Bacteria were cultured in LB or M9 0.5% glycerol plates supplemented with antibiotics as
17
18 739 required: Ap (100 μ g.mL⁻¹), Cm (25 μ g.mL⁻¹) or Km (25 μ g.mL⁻¹). Maltose (0.4 %) or
19
20 740 isopropyl-thio- β -D galactoside (IPTG) was added to induce *pul* gene expression.
21
22
23

24 741

25
26 742 *Purification of PpdDp*

27
28 743 pCHAP6154 was transformed into *E. coli* BL21(DE3) star cells (Novagen). The transformants
29
30 744 were grown at 37°C in minimal medium (M9) containing 0.1 % ¹⁵N-ammonium chloride and
31
32 745 0.4 % ¹³C-glucose. Protein expression was induced with 1 mM IPTG (Isopropyl β -D -1-
33
34 746 thiogalactopyranoside) when the culture reached an absorbance of 0.7–0.9 at 600 nm. After
35
36 747 3 h of induction at 30°C, the cells were centrifuged and the cell pellet was resuspended in 50
37
38 748 mM Tris–HCl buffer, pH 8.0 and disrupted by sonication. The
39
40 749 clarified cell lysate was applied to HisTrap FF column (GE Healthcare Life Sciences)
41
42 750 equilibrated with the same buffer containing 10 mM imidazole and 100 mM NaCl. Bound
43
44 751 proteins were eluted with a linear 10–300 mM imidazole gradient. The fractions containing
45
46 752 PpdD were then applied to Sephacryl S-100 column (GE Healthcare Life Sciences)
47
48 753 equilibrated with 50 mM phosphate buffer, pH 7.0, 50 mM NaCl. After the purification, the
49
50 754 fractions containing PpdD were concentrated up to 0.9 mM and supplied with 12 % D₂O for
51
52 755 NMR experiments. The total protein yield was 14 mg L⁻¹. The protein concentration was
53
54 756 estimated from its absorbance at 280 nm assuming a calculated ϵ_{280} of 15,720 M⁻¹ cm⁻¹. All
55
56
57
58
59
60
61
62
63
64
65

1 757 purification steps were performed at 4°C in the presence of the protease inhibitor cocktail
2 758 (Roche).

3
4 759

5
6 760 *Purification of pili*

7
8 761 Bacteria of strain PAP5171 containing plasmid pCHAP8565 encoding PpdD and one of the
9 762 two plasmids encoding the pilus assembly machinery, pCHAP8184 (for the T2SS) or pMS41
10 763 (for T4PS), were densely inoculated on LB Cm Ap IPTG plates and cultured at 30°C for 5
11 764 days. Bacteria were harvested in LB and pili were sheared by extensive vortexing and ten
12 765 passages through a 26-Gauge needle. Bacteria were pelleted for 10 min at 14000 g and the
13 766 supernatant was further pelleted in the same conditions in 1.5-ml eppendorf tubes. The
14 767 collected cleared supernatant was subjected to ultracentrifugation for 1 hr in Ti60 Beckman
15 768 rotor at 100 000 g. Pellets were resuspended in 50 mM HEPES pH 7.2, 50 mM NaCl and
16 769 stored on ice.

17
18
19
20 770

21
22 771 *Plasmid constructions*

23
24 772 Plasmid pCHAP8656 contains *ppdD* gene, amplified from the genomic DNA of strain
25 773 EDL933 using the high-fidelity Pwo DNA polymerase and placed under the control of *lacZ*
26 774 promoter (Table S3). This plasmid, as well as the plasmid pCHAP8501 encoding the hybrid
27 775 protein between the CyaA T18 fragment and PpdD, were used as templates for site-directed
28 776 mutagenesis to create charged residue substitutions in PpdD listed in Table S3. For
29 777 mutagenesis, five cycles of PCR amplification were performed in parallel with reverse and
30 778 forward complementary oligonucleotides (custom-synthesized by Eurofins, listed in Table
31 779 S4). The reaction mixtures were then combined and amplifications were continued for
32 780 another 13 cycles (30 s at 96°C, 30 s at 50°C and 3 min at 72°C). After *DpnI* digestion, 10-20
33 781 µL of the reactions were introduced into ultra-competent DH5α F'*lacI*^R ultracompetent cells
34 782 and transformants were selected on LB Cm plates. The resulting plasmids were purified
35 783 using the Qiagen miniprep kit and verified by DNA sequencing (GATC and Eurofins).

36 784

785 *Piliation assays*

1
2 786 The piliation assay was performed as described previously (Luna Rico et al., 2018). Bacteria
3
4 787 of strain PAP7460 transformed with plasmid pCHAP8184 and either vector alone or
5
6 788 pCHAP8565 derivatives carrying *ppdD* were grown 48 hours at 30°C on LB agar containing
7
8 789 Ap, Cm and 0.4% maltose. Bacteria were harvested and normalized to OD_{600nm} of 1 in LB.
9
10
11 790 Pili were detached by a 1-min vortex treatment and bacteria were spun for 5 min at 16000g.
12
13 791 The bacterial pellet was resuspended in SDS sample buffer at 10 OD_{600nm}. mL⁻¹ and the
14
15 792 supernatant was cleared from the remaining bacteria in a second 10-min centrifugation. The
16
17 793 cleared supernatant was transferred to a new eppendorf tube and precipitated with 10% tri-
18
19 794 chloro-acetic acid for 30 min on ice. Pellets were collected by 30-min centrifugation at 16000
20
21 g, washed twice with acetone, air-dried and taken up in SDS sample buffer at a
22 795 concentration of 10 OD_{600nm} equivalents per mL. Equivalent volumes of bacteria and pili
23
24 796 fractions were analyzed by denaturing sodium dodecyl sulfate polyacrylamide gel
25
26 797 electrophoresis (SDS-PAGE) with a Tris-Tricin buffer system (Schagger and von Jagow,
27
28 798 1987). Proteins were transferred to a nitrocellulose membrane (ECL Amersham) and probed
29
30 799 with antisera generated against MalE-PpdD fusion protein described previously (Sauvonnet
31
32 800 et al., 2000).
33
34
35 801
36
37
38 802

39
40 803 *NMR experiments*

41
42 804 NMR data were acquired at 298 K on a Varian spectrometer operating at a proton frequency
43
44 805 of 600 MHz and equipped with a cryogenically cooled triple resonance ¹H{¹³C/¹⁵N} PFG
45
46 806 probe. Proton chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate
47
48 807 (DSS) as 0 ppm. ¹⁵N and ¹³C chemical shifts were referenced indirectly to DSS (Wishart et
49
50 al., 1995). The pulse sequences were from VnmrJ Biopack. NMR data were processed with
51 808 NMRPipe/NMRDraw (Delaglio et al., 1995) and analyzed with the CcpNmr Analysis software
52
53 809 package (Vranken et al., 2005).
54
55 810
56
57 811 Residual dipolar couplings (RDCs) were determined by measuring the difference in J_{NH}
58
59 812 splitting between in-phase/antiphase (IPAP) NMR experiments under isotropic and
60
61
62
63
64
65

813 anisotropic conditions (Ottiger et al., 1998). The protein sample for RDC determination was
1
2 814 at 1 mM in 25 mM sodium phosphate, pH 7.5, 50 mM NaCl, 12% D₂O. For protein alignment
3
4 815 in the anisotropic sample, 10 mg/mL of the bacteriophage Pf1 (*ASLA Biotech*) was added
5
6 816 and the sample was incubated 1 hour in the magnetic field prior to acquisition.
7
8
9 817 The backbone ¹⁵N dynamics experiments were performed with a 1 mM sample in 50 mM
10
11 818 sodium phosphate, pH 7.5, 50 mM NaCl, 12% D₂O. ¹H-¹⁵N NOE values were determined as
12
13 819 the ratio between the intensities of corresponding peaks in the spectra recorded with and
14
15 820 without pre-saturation of ¹H.
16
17
18 821
19
20 822 *NMR structure determination*
21
22 823 The structure determination strategy consisted in several rounds of automated iterative
23
24 824 NOESY assignment and structure calculation with the ARIA 2.3 (Rieping et al., 2007);
25
26 825 (Bardiaux et al., 2008) and CNS software packages (Brunger, 2007). Backbone dihedral
27
28
29 826 angles were predicted with TALOS+ (Shen et al., 2009), and predictions classified as “good”
30
31 827 or “dynamic” were converted into ϕ and ψ dihedral angle restraints. In addition, 45 D_{NH}
32
33 828 residual dipolar coupling (RDC) were introduced as restraints in the structure calculation.
34
35
36 829 Axial and rhombic components of the alignment tensor were first estimated from an
37
38 830 intermediate structure and further refined with the software PALES (Zweckstetter and Bax,
39
40 831 2000). A log-harmonic potential was employed (Nilges et al., 2008) during the cooling phase
41
42 832 of the simulated annealing, for NOE derived and hydrogen bonds distances restraints. This
43
44 833 potential was combined with an automated estimation of the optimal weighting of the
45
46 834 distance restraints (Bernard et al., 2011; Nilges et al., 2008). For NOE restraints, the final
47
48
49 835 average weight was 14.9 kcal/mol, and 11.9 kcal/mol for hydrogen bonds restraints. The
50
51 836 disulfide bonds that had been confirmed by cysteine C β chemical shifts as reported
52
53 837 previously (de Amorim et al., 2014), were imposed in the structure calculation. In the final
54
55 838 ARIA iteration, 100 conformers were calculated and the 20 lowest energy structures were
56
57
58 839 refined in a shell of water molecules (Linge et al., 2003). Table 1 gives a summary of NMR-
59
60 840 derived restraints and statistics on the final ensemble of NMR structure.
61
62
63
64
65

841

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

842 *Cryo-electron microscopy*

843 Three microliters PpdD pili samples were applied to plasma-cleaned lacey carbon grids and
844 vitrified using a Vitrobot Mark IV (FEI). The grids were imaged on a Titan Krios microscope
845 operating at 300 kV equipped with a Falcon III camera with a sampling of 1.09 Å/px. Images
846 were collected with a defocus range of 1.0 μm to 3.0 μm. Motioncor2 was used for motion
847 correcting all the images, followed by the CTFFIND3 program (Mindell and Grigorieff, 2003)
848 for the defocus and astigmatism estimation. Long filament boxes were extracted from the
849 cryo-EM images (after correction of phases and amplitudes through multiplying by the CTF)
850 using the e2heliboxer program in EMAN2 (Tang, 2007). The SPIDER software package
851 (Frank, 1996) was used for most other image processing. 384-px long overlapping boxes
852 were cut from the long filament boxes (with a shift of 4% of the box size, which is ~ 1.5 times
853 the axial rise), yielding 99,678 segments in total. A reference-based classification was used
854 to sort the segments in terms of the axial rise and azimuthal rotation. After sorting, 25,669
855 segments from the dominant group, which accounted for ~25% of the whole dataset, were
856 processed by IHRSR (Egelman, 2000) to produce the final reconstruction.

857
858 *Model building and refinement*

859 We first generated a full-length model of PpdD by adding the alpha-helical stem (F1-Q25) to
860 the PpdDp structure by using the available structure of the full-length gonococcal PilE (PDB
861 1AY2) as template with the Modeller program version 9v8 (Sali and Blundell, 1993). Next, the
862 full-length PpdD model was docked as a rigid-body in the cryo-EM density map with Situs
863 version 2.7 (Wriggers, 2012). However, only the region A23-N140 (corresponding to PpdDp
864 plus Y24 and Q25) could be fitted inside the cryo-EM density. After symmetrization by
865 making use of the helical symmetry parameters of the PpdD pilus cryoEM map, a theoretical
866 density map was generated from the atomic coordinates of the fitted PpdD₂₃₋₁₄₀ model. A
867 difference map was created by subtracting the theoretical PpdD₂₃₋₁₄₀ pilus map from the cryo-
868 EM density map. Next, we rigidly docked the modelled PpdD₁₋₁₆ segment in the difference

869 map with Situs (Wriggers, 2012). An initial real-space refinement was performed with 14
1
2 870 copies of the docked PpdD₁₋₁₆ and PpdD₂₃₋₁₄₀ segments with PHENIX version 1.11.1 (Adams
3
4 871 et al., 2010). A full-length model of PpdD was then constructed with Modeller (Sali and
5
6 872 Blundell, 1993) by connecting the refined PpdD₁₋₁₆ and PpdD₂₃₋₁₄₀ segments by an extended
7
8 873 linker (S17-P22). Finally, full-atom real-space refinement, including morphing, minimization
9
10 874 and simulated-annealing, was performed with PHENIX (Adams et al., 2010) to improve
11
12 875 model geometry and the correlation with the experimental cryo-EM map. The final model of
13
14 876 PpdD pilus was validated with Molprobitry version 4.1 (Davis et al., 2007). Refinement
15
16 877 statistics are given in Table S5. Structure figures were generated with PyMOL (Schrödinger,
17
18 878 2015) and UCSF Chimera (Pettersen et al., 2004).

879 *Analysis of residue co-evolution and secondary structure propensities*

880
881
882 Long-range contact predictions from residue co-evolution were obtained with the Gremlin tool
883 (Kamisetty et al., 2013). Using the EHEC PpdD sequence as query, we generated a Multiple
884 Sequence Alignment of 1473 homologous sequences with HHblits (Remmert et al., 2011),
885 for a final ratio of ~11 sequences per query position. Secondary structure propensities were
886 derived from local residue co-evolutions predicted with CCMpred (Seemayer et al., 2014),
887 following the strategy proposed recently by Toth-Petroczy *et al.* (Toth-Petroczy et al., 2016).
888 Briefly, for each position in the query sequence, a propensity score for α helix or β strand (S_α
889 or S_β) is calculated based on the average prediction score of short-range contacts ($S_{i,i+1}$ to
890 $S_{i,i+4}$ for helices and $S_{i,i+1}$ to $S_{i,i+2}$ for strands) involving the given position or its flanking
891 residues (2 for helices and 1 for strands on both sides). Then, scores were normalized with
892 regards to their correlation with $S_{i,i+1}$ scores. The final scores were obtained by subtracting
893 scores of non-specific contacts from scores specific to a particular secondary structure
894 element, *i.e.* $S_\alpha \simeq S_{i+4} + S_{i+3} - S_{i+2} - S_{i+1}$ and $S_\beta \simeq S_{i+2} - S_{i+1}$.

895

896 *Normal Mode Analysis of T4P structures*

897 Normal modes were computed with Prody (Bakan et al., 2014) with an Anisotropic Network
1
2 898 Model on C α atoms of T4P filament structures and a cutoff of 15 Å for pairwise interactions in
3
4 899 the elastic network. The different helical pitch values of the EHEC, Nm, Ng and PAK T4P
5
6 900 filaments lead to structures with significantly variable length for a fixed number of pilin
7
8 901 subunits, that may influence the outcome of the normal mode analysis. Consequently,
9
10 902 filament structures of equivalent length were generated using different number of subunits
11
12 903 (30 for Nm and Ng, 29 for PAK and 27 for EHEC), the respective helical symmetry and the
13
14 904 pilin structure. The following PDB entries were used: 5KUA for Nm, 5VXX for Ng and 5VXY
15
16 905 for PAK pili. For each of the first 3 non-trivial lowest frequency modes, the average of the
17
18 906 atomic mean square displacements was calculated for each subunit, excluding the first 24
19
20 907 residues.
21
22
23
24
25
26
27

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Supplemental tables and figures

Table S1; Related to Figure 3: Agreement of high confidence (Prob > 0.7) predicted contacts from residue co-evolution with PpdD pilus model.

Table S2; Related to Figure 1: Sequence and structure similarity between EHEC PpdD and other T4a pilins

Table S3; Related to STAR Methods: Plasmids used in this study.

Table S4; Related to STAR Methods: Oligonucleotides used in this study.

Table S5; Related to STAR Methods: Refinement statistics of the PpdD pilus structure

Figure S1; Related to Table 1: NOE restraints of the PpdDp NMR structure.

Figure S2; Related to Figure 1: (A) ^1H - ^{15}N heteronuclear NOE of PpdDp in solution.

Figure S3; Related to Figure 2: (A,B) Large variability in terms of the helical symmetry of PpdD pili.

Figure S4; Related to Figure 3: EHEC PpdD pilus contact map and evolutionary contact predictions.

Figure S5; Related to Figure 6: Buried surface areas of inter-subunit interfaces in T4Pa structures.

937 **Figure S6; Related to Figure 1:** Structural comparison of major T4a pilins (periplasmic
1 938 domains).

2 939

3 940 **Figure S7; Related to Figure 7:** Secondary structure elements in T4a pilins.

4 941

5 942 **Figure S8; Related to Figure 7:** Surface electrostatics potential of T4P pilus structure.

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

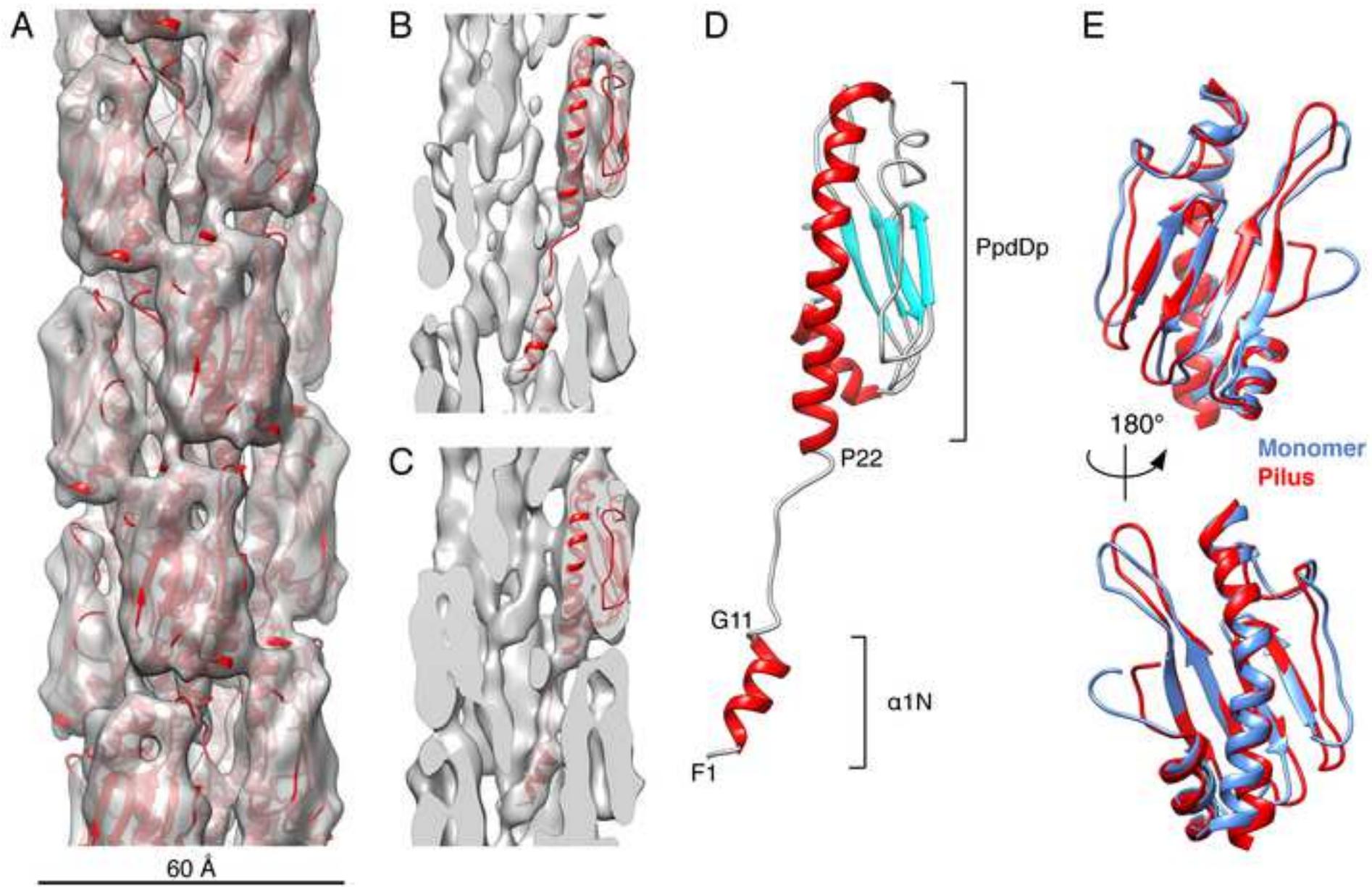
61

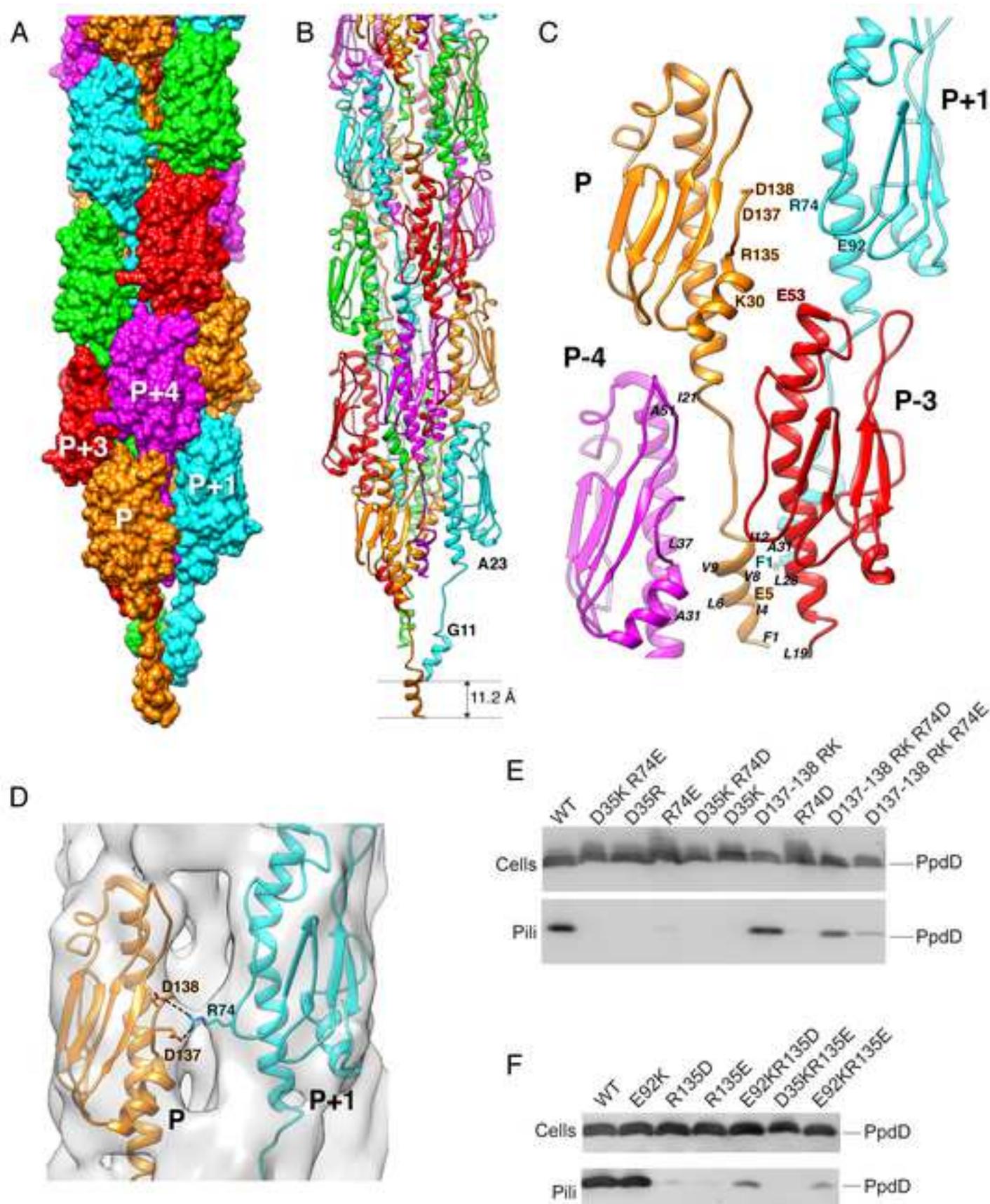
62

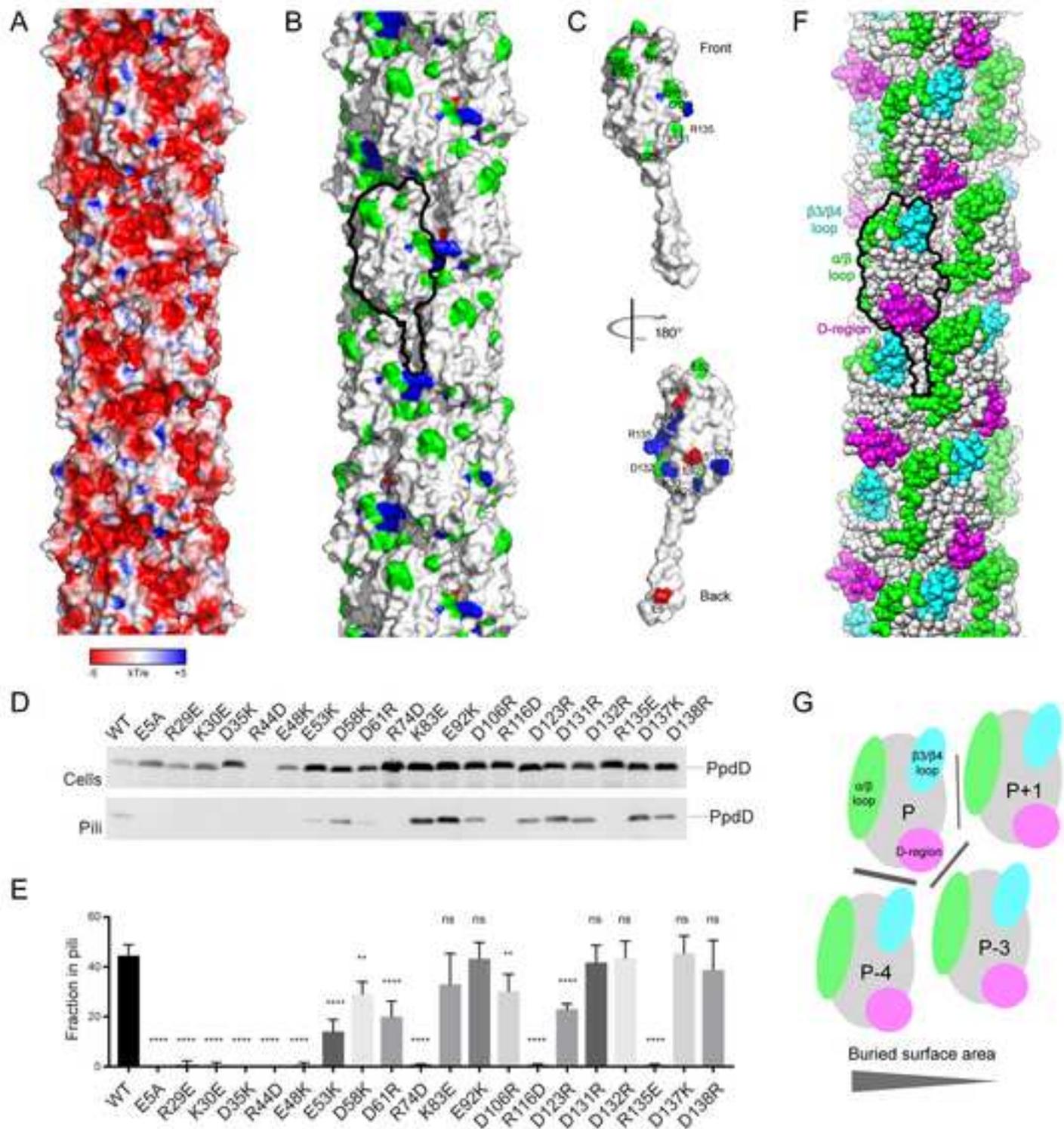
63

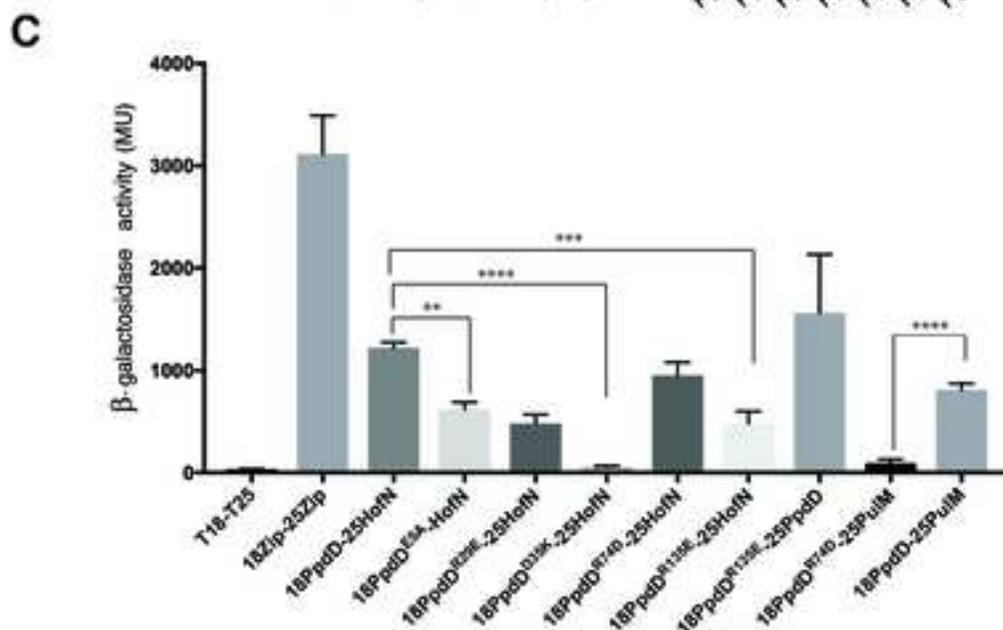
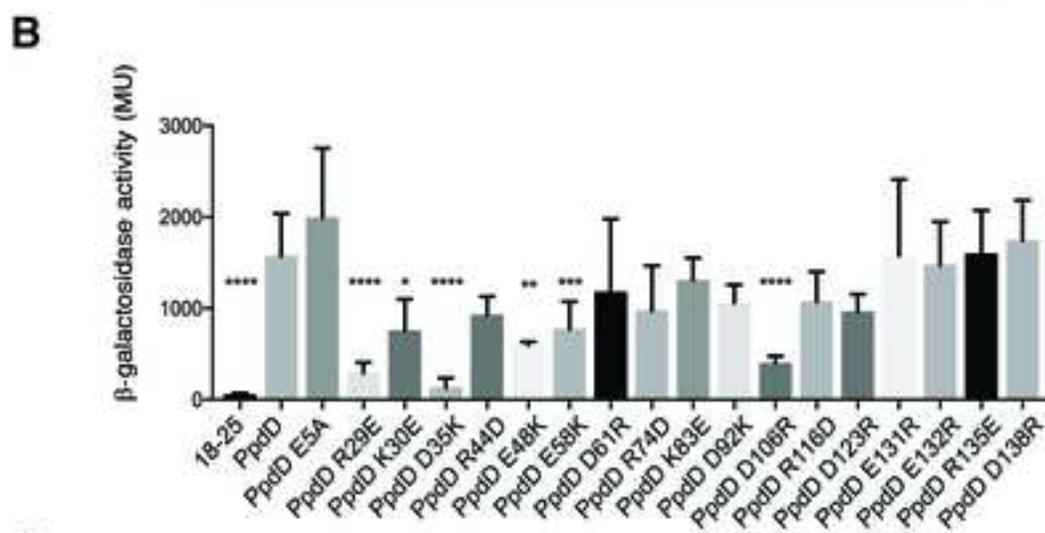
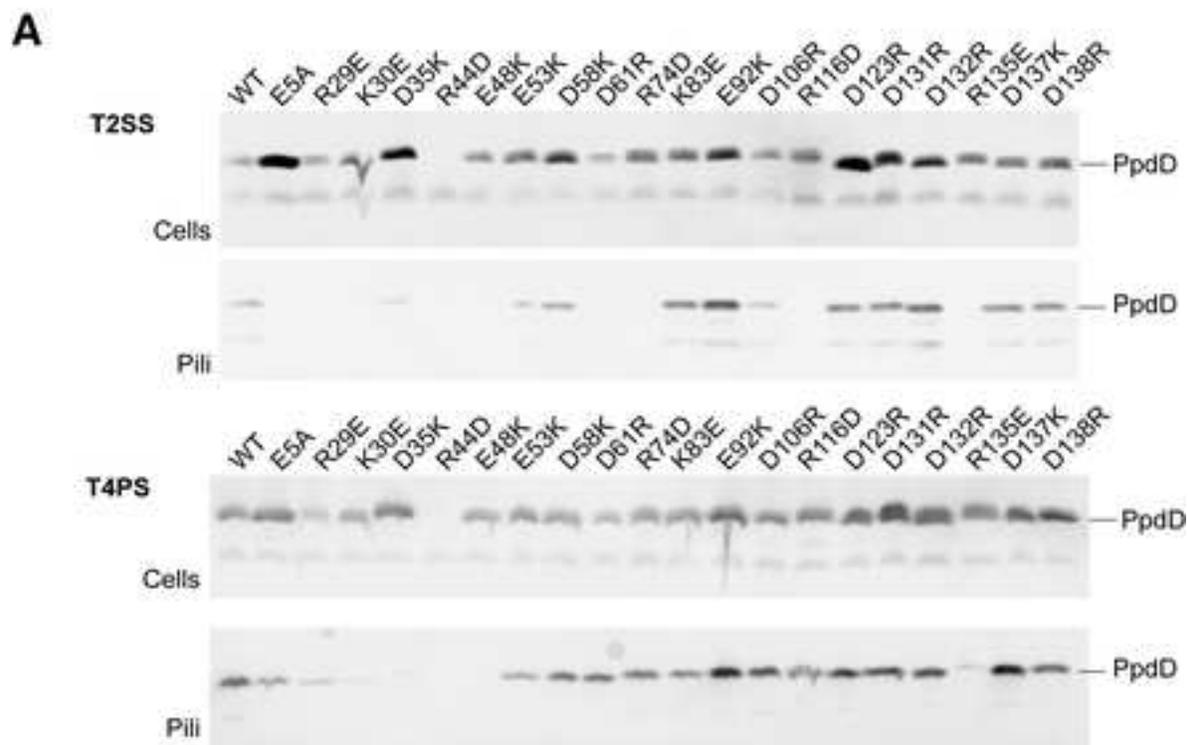
64

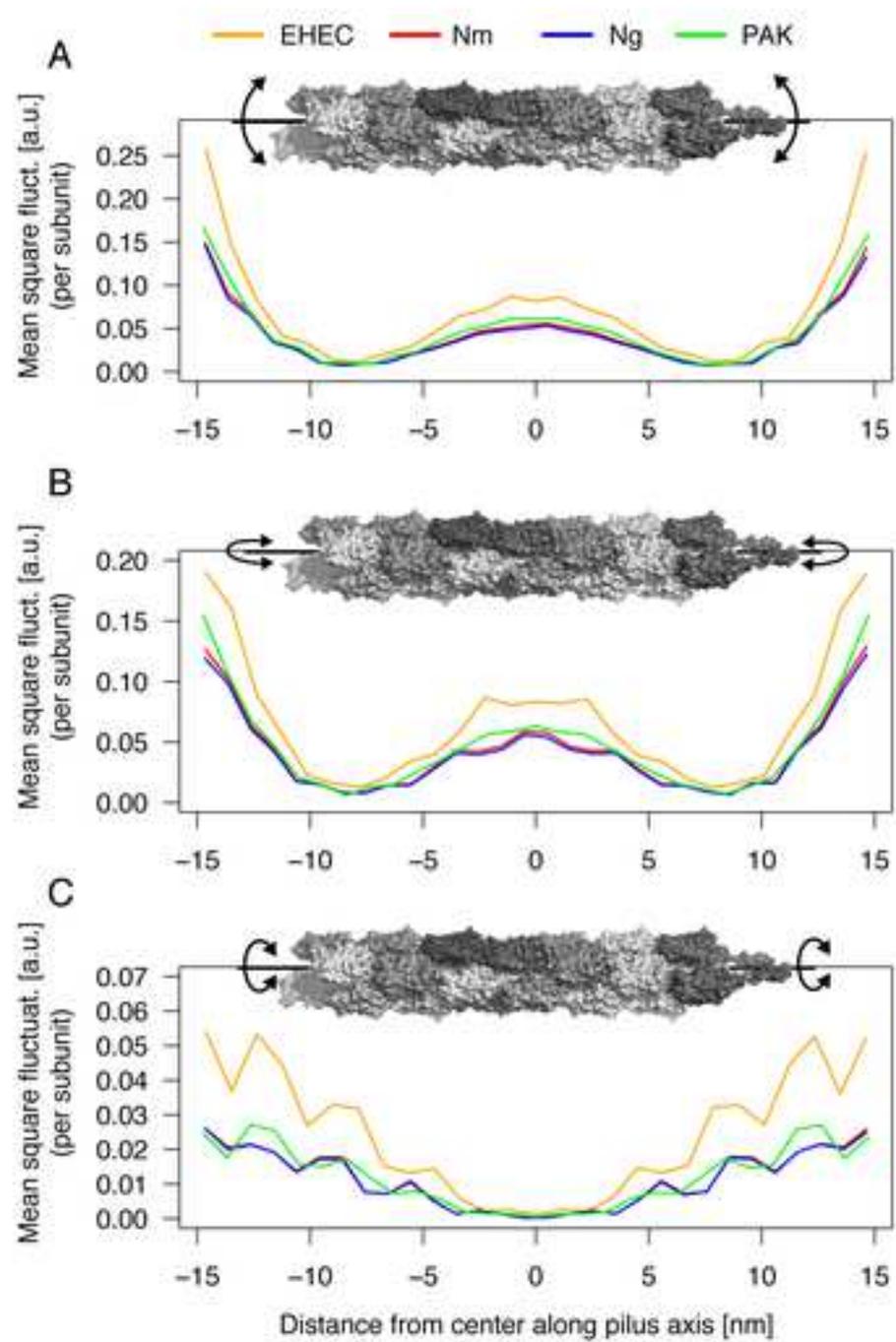
65











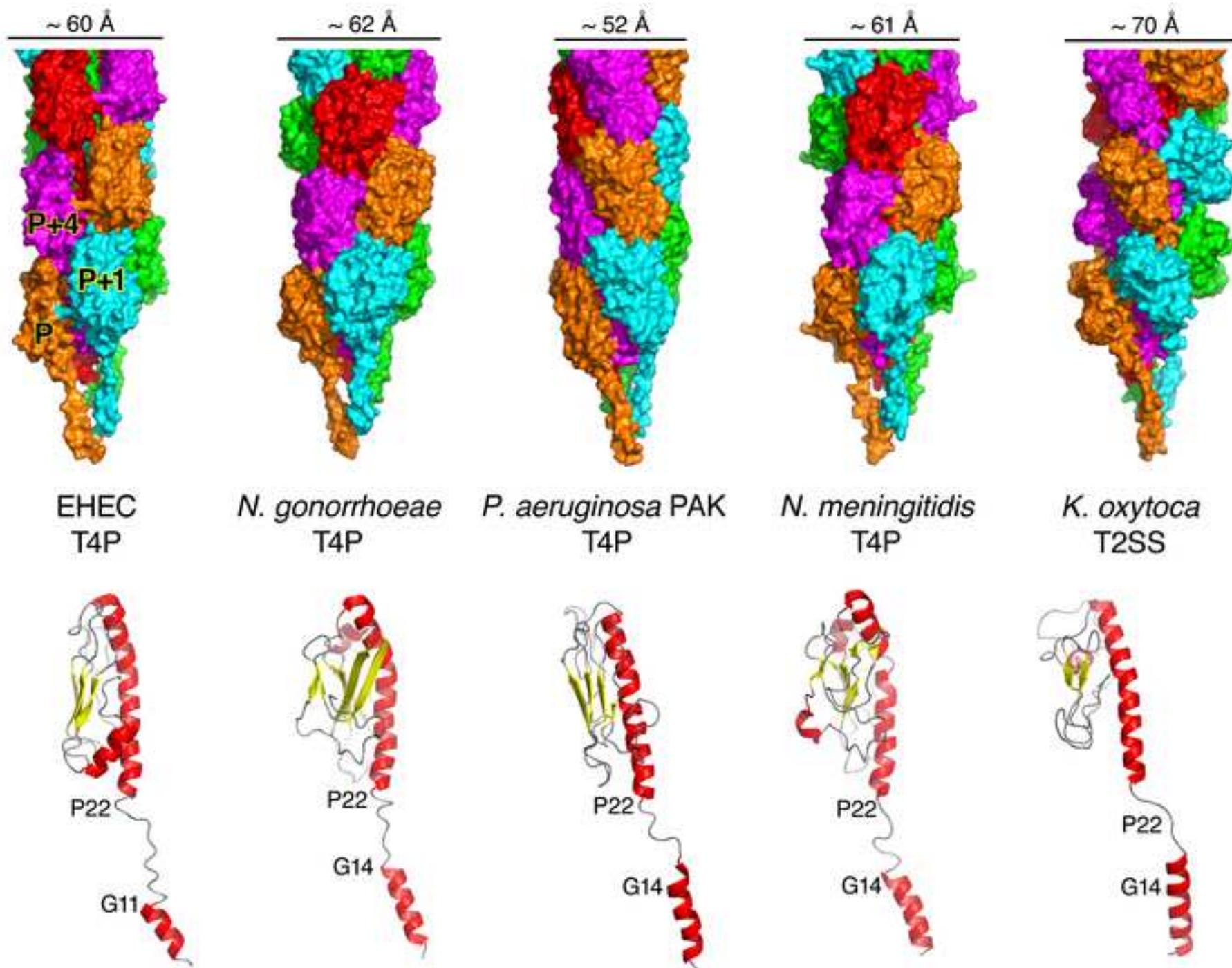


TABLE FOR AUTHOR TO COMPLETE

Please upload the completed table as a separate document. **Please do not add subheadings to the Key Resources Table.** If you wish to make an entry that does not fall into one of the subheadings below, please contact your handling editor. (NOTE: For authors publishing in *Current Biology*, please note that references within the KRT should be in numbered style, rather than Harvard.)

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-MalE-PpdD rabbit polyclonal antibodies	(Sauvonnet et al. 2000)	N/A
ECL™ anti-rabbit IgG linked to horse radish peroxidase from donkey	Amersham	Cat# NA934V
Bacterial Strains		
<i>Escherichia coli</i> DH5α F' <i>lacI</i> ^Q	Laboratory collection	N/A
<i>Escherichia coli</i> BW25113 F' <i>lacI</i> ^Q	(Datsenko and Wanner, 2000)	N/A
<i>Escherichia coli</i> BL21(DE3) Star	Novagen	Cat# 69450
<i>Escherichia coli</i> PAP7460	(Possot et al., 2000)	N/A
<i>Escherichia coli</i> DHT1	(Dautin et al., 2000)	N/A
<i>Escherichia coli</i> PAP5171	(Francetic and Pugsley, 2005)	N/A
Biological Samples		
Sheared pilus fraction from strain PAP5171 carrying plasmids pCHAP8565 and pCHAP8184	This study	N/A
Sheared pilus fraction from strain PAP5171 carrying plasmids pCHAP8565 and pMS41	This study	N/A
Chemicals, Peptides, and Recombinant Proteins		
DNA polymerase Pwo	Roche	Cat#11644947001
T4 DNA ligase	New England Biolabs	Cat# M0202
PpdD(25-140)Thr-His ₆ purified protein	(de Amorim et al, 2014)	N/A
Q5 High Fidelity Polymerase	New England Biolabs	Cat# M0491
<i>KpnI</i> -HF restriction enzyme	New England Biolabs	Cat# R3142L
<i>EcoRI</i> -HF restriction enzyme	New England Biolabs	Cat# R3010L
<i>Complete protease inhibitor</i>	Roche	Cat#11836153001
Bacteriophage Pf1	ASLA biotech	Cat# P-50-P
D ₂ O	Eurisotop	Cat# D214F
Critical Commercial Assays		
Qiaprep Spin plasmid purification kit	Qiagen	Cat#: 27104
Qiaquick PCR purification kit	Qiagen	Cat#: 28104
Pierce®ECL 2 Western Blotting Substrate	Fisher Scientific	Cat# : 993PT
Deposited Data		
NMR resonance assignments	(de Amorim et al., 2014)	BMRB: 18823
NMR structure ensemble of truncated PpdD pilin of EHEC	This paper	PDB: 6GMS
Cryo-EM reconstruction of the EHEC type IV pilus	This paper	EMDB-0070
Atomic model of the EHEC type IV pilus	This paper	PDB: 6GV9

Oligonucleotides		
See Table S4 for list of primers used in this study	N/A	NA/
Recombinant DNA		
See Table S3 for list of plasmids used in this study	N/A	N/A
Software and Algorithms		
PRISM		
Image J		
VnmrJ Biopack	Agilent Technologies, Inc.	N/A
NMRPipe/NMRDraw	(Delaglio et al., 1995)	https://spin.niddk.nih.gov/NMRPipe/
CcpNmr Analysis 2.4	(Vranken et al., 2005)	https://www.ccpn.ac.uk/
ARIA 2.3	(Rieping et al., 2007)	http://aria.pasteur.fr
CNS 1.2	(Brunger, 2007)	http://cns-online.org
TALOS+	(Shen et al., 2009)	https://spin.niddk.nih.gov/NMRPipe/talos/
PALES	(Zweckstetter and Bax, 2000)	https://spin.niddk.nih.gov/bax/software/PALLES/
CTFFIND3	(Mindell and Grigorieff, 2003)	http://grigoriefflab.janelia.org/ctf
EMAN2	(Tang, 2007)	https://blake.bcm.edu/emanwiki/EMAN2
SPIDER	(Frank, 1996)	https://spider.wadsworth.org/
IHRSR	(Egelman, 2000)	https://cryoem.ucsd.edu/wikis/software/start.php?id=ihrsr
Modeller 9v8	(Sali and Blundell, 1993)	https://salilab.org/modeller/
PHENIX 1.11	(Adams et al., 2010)	https://www.phenix-online.org/
Situs 2.7	(Wriggers, 2012)	https://situs.biomachina.org/
Molprobit 4	(Davis et al., 2007)	http://molprobity.biochem.duke.edu/
Pymol 2	(Schrödinger, 2015)	https://pymol.org/
UCSF Chimera	(Pettersen et al., 2004)	https://www.cgl.ucsf.edu/chimera/
WHAT-IF	(Vriend, 1990)	https://swift.cmbi.umcn.nl/whatif/
Procheck	(Laskowski et al., 1993)	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
Gremlin	(Kamisetty et al., 2013)	http://gremlin.bakerlab.org/
CCMpred	(Seemayer et al., 2014)	https://github.com/soedinglab/CCMpred
Prody	(Bakan et al., 2014)	http://prody.csb.pitt.edu/

Structure and assembly of the Enterohemorrhagic

Escherichia coli type 4 pilus

Benjamin Bardiaux^{1,4}, Gisele Cardoso de Amorim^{2,4,6}, Areli Luna Rico^{1,3,4}, Weili Zheng⁵,
Ingrid Guilvout³, Camille Jollivet³, Michael Nilges¹, Edward H Egelman⁵, Nadia Izadi-
Pruneyre^{1,2*} and Olivera Francetic^{3*}

¹ Structural Bioinformatics Unit, Department of Structural Biology and Chemistry, C3BI,
Institut Pasteur; CNRS UMR3528; CNRS USR3756; Paris, France,

² NMR of Biomolecules Unit, Department of Structural Biology and Chemistry, Institut
Pasteur, CNRS UMR3528, Paris, France

³Biochemistry of Macromolecular Interactions Unit, Department of Structural Biology and
Chemistry, Institut Pasteur, CNRS UMR3528, Paris, France.

⁵Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville,
VA22908, USA

⁴Equal contribution

⁶Present address: Núcleo Multidisciplinar de Pesquisa em Biologia - NUMPEX-BIO,
Universidade Federal do Rio de Janeiro, Estrada de Xerém, RJ, Brazil.

⁴Equal contribution

*corresponding authors: nadia.izadi@pasteur.fr, olivera.francetic@pasteur.fr

Supplementary Information:

Supplementary Tables 1-5.

Supplementary Figures 1-8.

Supplementary references.

Table S1; Related to Figure 3: Agreement of high confidence predicted contacts from residue co-evolution with PpdD pilus model. The accuracy (number of contacts with Prob > 0.7 consistent with the model/total number of contacts) is reported for increasing distance thresholds and for different rotameric states of the PpdD pilus model. A contact is considered as consistent with the model if the shortest distance between any non-hydrogen atoms of the two involved residues is smaller than the threshold. Optimized side-chain rotamers were built with SCWRL4.0 (Krivov et al., 2009). Alternatively, the at-most 10 best (most frequent) rotameric states from the Dunbrack library (Shapovalov and Dunbrack, 2011) were considered simultaneously to measure inter-residue distances in the PpdD pilus model.

Distance threshold (Å)	PpdD pilus model (PDB 6GV9)	PpdD pilus model + optimized side-chain rotamers	PpdD pilus model + 10 best side-chain rotamers
5	46.8 %	47.5 %	60.3 %
8	79.4 %	79.4 %	83.0 %
10	88.7 %	89.4 %	90.1 %
12	91.5 %	92.9 %	94.3 %

Table S2; Related to Figure 1: Sequence and structure similarity between EHEC PpdD and other T4a pilins

Pilin*	PDB id	% identity (full-length)	% identity (from Pro 22)	RMSD (Å)#
Ft PilE	3SOJ	25	16	2.7
Pa K122-4 PilA	1QVE	31	23	2.8
Pa PAK PilA	1OQW	30	22	2.8
Dn FimA	3SOK	28	20	3.0
Pa 110594 PilA	3JYZ	26	18	3.1
Ab PilA	5IHJ	30	20	3.1
Nm PilE	5JW8	32	26	3.2
Ng PilE	1AY2	32	26	3.2

* Ft: *Francisella tularensis*; PaK122-4: *Pseudomonas aeruginosa* strain K122-4; PAK: *Pseudomonas aeruginosa* strain PAK; Dn: *Dichelobacter nodosus*; Pa110594: *Pseudomonas aeruginosa* strain Pa110594; Ab: *Acinetobacter baumannii*; Nm: *Neisseria meningitidis*; Ng: *Neisseria gonorrhoeae*.

Computed with TM-align (Zhang and Skolnick, 2005).

Table S3; Related to STAR Methods: Plasmids used in this study.

Plasmid name	Ori/resistance ¹	Relevant markers	Source/reference
pCHAP8184	ColE1/Ap ^R	<i>pulS,pulAB pulCDEFHIJKLMNO</i>	(Campos <i>et al.</i> , 2010)
pSU18	p15A/Cm ^R	<i>placZ-lacZα</i>	(Bartolome <i>et al.</i> , 1991)
pUC18	ColE1/Ap ^R	<i>placZ-lacZα</i>	(Yanisch-Perron <i>et al.</i> , 1985)
pCHAP6154	ColE1/Ap ^R	<i>pT7-peIBsp-ppdD(25-140)Thr His₆</i>	(de Amorim <i>et al.</i> , 2014)
pCHAP8565	p15A/Cm ^R	<i>ppdD EHEC</i>	(Luna Rico <i>et al.</i> , 2019)
pCHAP6263	p15A/Cm ^R	<i>ppdD^{E5A}</i>	This study
pCHAP6250	p15A/Cm ^R	<i>ppdD^{D35R}</i>	This study
pCHAP6254	p15A/Cm ^R	<i>ppdD^{R29E}</i>	This study
pCHAP6256	p15A/Cm ^R	<i>ppdD^{E131K}</i>	This study
pCHAP6257	p15A/Cm ^R	<i>ppdD^{E92K}</i>	This study
pCHAP6258	p15A/Cm ^R	<i>ppdD^{K83E}</i>	This study
pCHAP6260	p15A/Cm ^R	<i>ppdD^{R74D}</i>	This study
pCHAP6261	p15A/Cm ^R	<i>ppdD^{K30E}</i>	This study
pCHAP6264	p15A/Cm ^R	<i>ppdD^{R44D}</i>	This study
pCHAP6265	p15A/Cm ^R	<i>ppdD^{D106R}</i>	This study
pCHAP6266	p15A/Cm ^R	<i>ppdD^{R116D}</i>	This study
pCHAP6267	p15A/Cm ^R	<i>ppdD^{D132R}</i>	This study
pCHAP6268	p15A/Cm ^R	<i>ppdD^{E48K}</i>	This study
pCHAP6269	p15A/Cm ^R	<i>ppdD^{E53K}</i>	This study
pCHAP6317	p15A/Cm ^R	<i>ppdD^{D58K}</i>	This study
pCHAP6271	p15A/Cm ^R	<i>ppdD^{D61R}</i>	This study
pCHAP6272	p15A/Cm ^R	<i>ppdD^{D123R}</i>	This study
pCHAP6316	p15A/Cm ^R	<i>ppdD^{D35K}</i>	This study
pCHAP6320	p15A/Cm ^R	<i>ppdD^{K30D}</i>	This study
pCHAP6333	p15A/Cm ^R	<i>ppdD^{R74E}</i>	This study
pCHAP6319	p15A/Cm ^R	<i>ppdD^{D106K}</i>	This study
pCHAP6321	p15A/Cm ^R	<i>ppdD^{R29D}</i>	This study
pCHAP6253	p15A/Cm ^R	<i>ppdD^{D137K}</i>	This study
pCHAP6262	p15A/Cm ^R	<i>ppdD^{D138R}</i>	This study
pCHAP6280	p15A/Cm ^R	<i>ppdD^{D137K D138R}</i>	This study
pCHAP6259	p15A/Cm ^R	<i>ppdD^{R135E}</i>	This study
pCHAP6327	p15A/Cm ^R	<i>ppdD^{D35K R74D}</i>	This study
pCHAP6337	p15A/Cm ^R	<i>ppdD^{D35K R74E}</i>	This study
pCHAP8968	p15A/Cm ^R	<i>ppdD^{D35R R74D}</i>	This study
pCHAP8969	p15A/Cm ^R	<i>ppdD^{D137R D138K R74D}</i>	This study
pCHAP8970	p15A/Cm ^R	<i>ppdD^{D137R D138K R74E}</i>	This study
pCHAP6314	p15A/Cm ^R	<i>ppdD^{E92R}</i>	This study
pCHAP6257	p15A/Cm ^R	<i>ppdD^{E92K}</i>	This study
pCHAP6259	p15A/Cm ^R	<i>ppdD^{R135E}</i>	This study
pCHAP6322	p15A/Cm ^R	<i>ppdD^{R135D}</i>	This study
pCHAP6277	p15A/Cm ^R	<i>ppdD^{E92K R135E}</i>	This study
pCHAP6285	p15A/Cm ^R	<i>ppdD^{E92K R135D}</i>	This study
pCHAP8956	p15A/Cm ^R	<i>ppdD^{D35K R135E}</i>	This study
pUT18c	ColE1/Ap ^R	<i>placUV5-cyaA</i> T18 fragment in pUC18	(Karimova <i>et al.</i> , 1998)
pKT25	p15A/Km ^R	<i>placUV5-cyaA</i> T25 fragment in pSU38	(Karimova <i>et al.</i> , 1998)
pCHAP8501	ColE1/Ap ^R	<i>T18-PpdD</i>	(Luna Rico <i>et al.</i> , 2019)
pCHAP8504	p15A/Km ^R	<i>T25-PpdD</i>	(Luna Rico <i>et al.</i> , 2019)

pCHAP8973	ColE1/Ap ^R	T18-PpdD ^{E5A}	This study
pCHAP8768	ColE1/Ap ^R	T18-PpdD ^{R29E}	This study
pCHAP8770	ColE1/Ap ^R	T18-PpdD ^{K30E}	This study
pCHAP8772	ColE1/Ap ^R	T18-PpdD ^{D35K}	This study
pCHAP8774	ColE1/Ap ^R	T18-PpdD ^{R44D}	This study
pCHAP8776	ColE1/Ap ^R	T18-PpdD ^{E48K}	This study
pCHAP8916	ColE1/Ap ^R	T18-PpdD ^{D58K}	This study
pCHAP8780	ColE1/Ap ^R	T18-PpdD ^{D61K}	This study
pCHAP8782	ColE1/Ap ^R	T18-PpdD ^{R74D}	This study
pCHAP8784	ColE1/Ap ^R	T18-PpdD ^{R116D}	This study
pCHAP8786	ColE1/Ap ^R	T18-PpdD ^{R135E}	This study
pCHAP8920	ColE1/Ap ^R	T18-PpdD ^{K83E}	This study
pCHAP8922	ColE1/Ap ^R	T18-PpdD ^{E92K}	This study
pCHAP8924	ColE1/Ap ^R	T18-PpdD ^{D106R}	This study
pCHAP8926	ColE1/Ap ^R	T18-PpdD ^{D123R}	This study
pCHAP8928	ColE1/Ap ^R	T18-PpdD ^{E131R}	This study
pCHAP8930	ColE1/Ap ^R	T18-PpdD ^{D132R}	This study
pCHAP8936	ColE1/Ap ^R	T18-PpdD ^{D138R}	This study
pCHAP8155	p15A/Km ^R	T25-PulM	(Nivaskumar et al., 2016)
pCHAP8910	p15A/Km ^R	T25-HofN	(Luna Rico et al., 2019)
pMS41	ColE1/Ap ^R	EHEC T4PS (Δ <i>ppdD</i> , <i>hofBC</i> , <i>hofMNOPQ</i> , <i>ppdAB</i> <i>BygdBppdC</i> <i>gspO</i>)	(Luna Rico et al., 2019)

¹Ap^R, ampicillin-resistance; Cm^R, chloramphenicol-resistance, Km^R, kanamycin-resistance).

Table S4; Related to STAR Methods: Oligonucleotides used in this study.

Oligonucleotide name	Sequence (5'-3')
PpdD-SD-EcoRI	CTATTCTGAATTCAAAGTAGCGCCAACCAAATC
PpdD-Hind-3	cacAAGCTTGCATGCCTGCAGGTCGACTCTAG
HofC EcoRI 5	CACGAATTCTGAAGAGTTAATCCCGGTATTG
HofC Hind 3	GACTGCAAGCTTCGTTATCCCATCCCACTCATCG
HofM Pst 5	CGCCTGCAGGGTAGTATAAAGGCAAGC
HofQ Sph 3	CACGCATGCCACGTTTCAGCGTAAAAAC
PpdD 5 Kpn	GCAGGTACCTATGACACTTATCGAACTGATGGTG
PpdD 3 Eco	CACGAATTCCATTTTCAGTGAGCTGTGGAAT
PpdDE5A 5 Kpn	GCAGGTACCTATGACACTTATCGcACTGATGGTG
PpdD D35R 5	CAAAGCCGCACTCACCcgCATGCTACAAAC
PpdD D35R 3	GTTTGTAGCATGcgGGTGAGTGCGGCTTTG
PpdD E5A 5	GTTTTACACTTATCGcACTGATGGTGG
PpdD E5A 3	CCACCATCAGTgCGATAAGTGTA AAC
PpdD R29E 5	CAA AACTACCTGgaaAAAGCCGCACTC
PpdD R29E 3	GAGTGCGGCTTTtcCAGGTAGTTTTG
PpdD R29D 5	TATCAA AACTACCTGgaCAAAGCCGCACTCACC
PpdD R29D 3	GGTGAGTGCGGCTTTGtcCAGGTAGTTTTGATA
PpdD K30E 5	ACTACCTGCGCgAAGCCGCACTC
PpdD K30E 3	GAGTGCGGCTTcGCGCAGGTAGT
PpdD K30D 5	CAA AACTACCTGCGCgAtGCCGCACTCACCGAC
PpdD K30D 3	GTCGGTGAGTGCGGCaTcGCGCAGGTAGTTTTG
PpdD R44D 5	CCTTTGTGCCTTACgaTACCGCCGTAGAG
PpdD R44D 3	CTCTACGGCGGTAtcGTAAGGCACAAAGG
PpdD K30D 5	CAA AACTACCTGCGCgAtGCCGCACTCACCGAC
PpdD K30D 3	GTCGGTGAGTGCGGCaTcGCGCAGGTAGTTTTG
PpdD D35K 5	AAAGCCGCACTCACCaAaATGCTACAAACCTTTG
PpdD D35K 3	CAAAGGTTTGTAGCATtTtGGTGAGTGCGGCTTT
PpdD R74D 5	CCTACCACCACCgaCTATGTTTCAGCC
PpdD R74D 3	GGCTGAAACATAGtcGGTGGTGGTAGG
PpdD R74E 5	CCCTCGCCTACCACCACCgaaTATGTTTCAGCCATGAGTG
PpdD R74E 3	CACTCATGGCTGAAACATAttcGGTGGTGGTAGGCGAGGG
PpdD D106R 5	CATGACGCCAGGTTGGcgTAACGCAAACGGCGTCAC
PpdD D106R 3	GTGACGCCGTTTGCCTTAacgCAAACCTGGCGTCATG
PpdD D106K 5	ATGACGCCAGGTTGGaAaAACGCAAACGGCGTC
PpdD D106K 3	GACGCCGTTTGCCTTtTtCCAACCTGGCGTCAT
PpdD D137K 5	GTCTTCCGCTTTaAaGACGCCAACTAAG
PpdD D137K 3	CTTAGTTGGCGTctTtAAAGCGGAAGAC
PpdD D138R 5	GTCTTCCGCTTTGATcgCGCCAACTAAGG
PpdD D138R 3	CCTTAGTTGGCGcgATCAAAGCGGAAGAC
PpdD D137K D138R 5	AAGATGTCTTCCGCTTTaAacgGCCAACTAAGGAGC
PpdD D137K D138R 3	GCTCCTTAGTTGGCgctTtAAAGCGGAAGACATCTT
PpdD E131K 5	CAGCAAGCCTGCaAAGATGTCTTCC
PpdD E131K 3	GGAAGACATCTTtGCAGGCTTGCTG
PpdD E92K 5	CTGACCGGGCAaAgAGTCTCAATGG
PpdD E92K 3	CCATTGAGACTcTtTTGCCCGGTCAG
PpdD K83E 5	CATGAGTGTGGCAgAGGGCGTGGTG
PpdD K83E 3	CACCACGCCCTcTGCCCACTCATG

PpdD D132R 5	GCAAGCCTGCGAACGTGTCTTCCGCTTTGATGACG
PpdD D132R 3	CGTCATCAAAGCGGAAGACACGTTTCGCAGGCTTGC
PpdD D123R 5	GCAATATTCAAAGTCGCAGCGCATTGCAGC
PpdD D123R 3	GCTGCAATGCGCTGCGACTTTGAATATTGC
PpdD D61R 5	GGATTAGATACCTGCCGCGGTGGCAGCAATGG
PpdD D61R 3	CCATTGCTGCCACCGCGGCAGGTATCTAATCC
PpdD R135E 5	GAAGATGTCTTCgaATTcGATGACGCCAAC
PpdD R135E 3	GTTGGCGTCATCgAATtcGAAGACATCTTC
PpdD E48K 5	CCTTACCGTACCGCCGTAAAGTTGTGCGCGCTGG
PpdD E48K 3	CCAGCGCGCACAACTTTACGGCGGTACGGTAAGG
PpdD E53K 5	GGTATCTAATCCACCATGTTTCAGCGCGCACAACTCTACG
PpdD E53K 3	GCCACCGTCGCAGGTtTtTAATCCACCATGTTC
PpdD D58K 5	GAACATGGTGGATTAaAaACCTGCGACGGTGGC
PpdD D58K 3	CCACCGTCGCAGGTACGTAATCCACCATGTTCC
PpdD R116D 5	CGTCACCGGCTGGGCGgaCAACTGCAATATTCAAAG
PpdD R116D 3	CTTTGAATATTGCAGTTGtcCGCCCAGCCGGTGACG
PpdD R44E E48K 5	CCTTACgaaACCGCCGTaAAGTTGTGCGCGCTGG
PpdD R44E E48K 3	CCAGCGCGCACAACTtTACGGCGGTttcGTAAGG
PpdD R44D E48R 5	GCCTTACgaTACCGCCGTAcgtTTGTGCGCGCTGGAACATG
PpdD R44D E48R 3	CATGTTCCAGCGCGCACAAacgTACGGCGGTAtcGTAAGGC

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table S5; Related to STAR Methods: Refinement statistics of the PpdD pilus structure

Deviation from ideal geometry	
RMS ^a for bond lengths	0.02 Å
RMS for bond angles	2.40 °
Cβ deviations	0 %

Molprobit statistics	
Ramachandran favored	92.0 %
Ramachandran allowed	6.5 %
Ramachandran outliers	1.4 %
Rotamer outliers	0 %
Clashscore	28.7
Overall score	2.43

^aRMS, root mean square

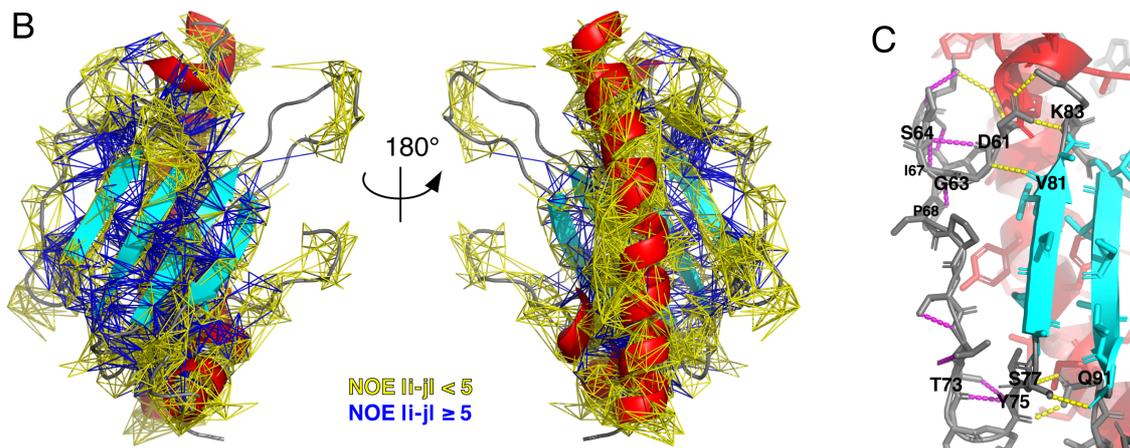
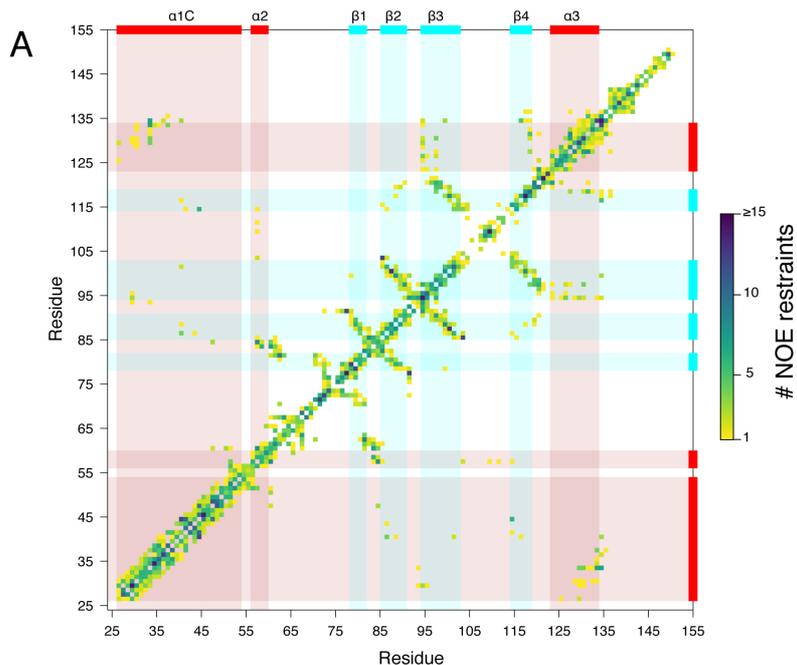


Figure S1; Related to Table 1: NOE restraints of the PpdDp NMR structure. (A) Map of inter-residue NOE restraints used to calculate the PpdDp structure. The color scale corresponds to the number of unique NOE restraints for a pair of residues. Secondary structures are shown in red (helix) and cyan (strands). (B) NOE restraints shown on the lowest energy conformer of the PpdDp NMR structure. Long range NOE restraints are shown with blue lines and other NOE restraints with yellow lines. (C) Hydrogen-bonds detected in the PpdDp structure stabilizing the α/β loop.

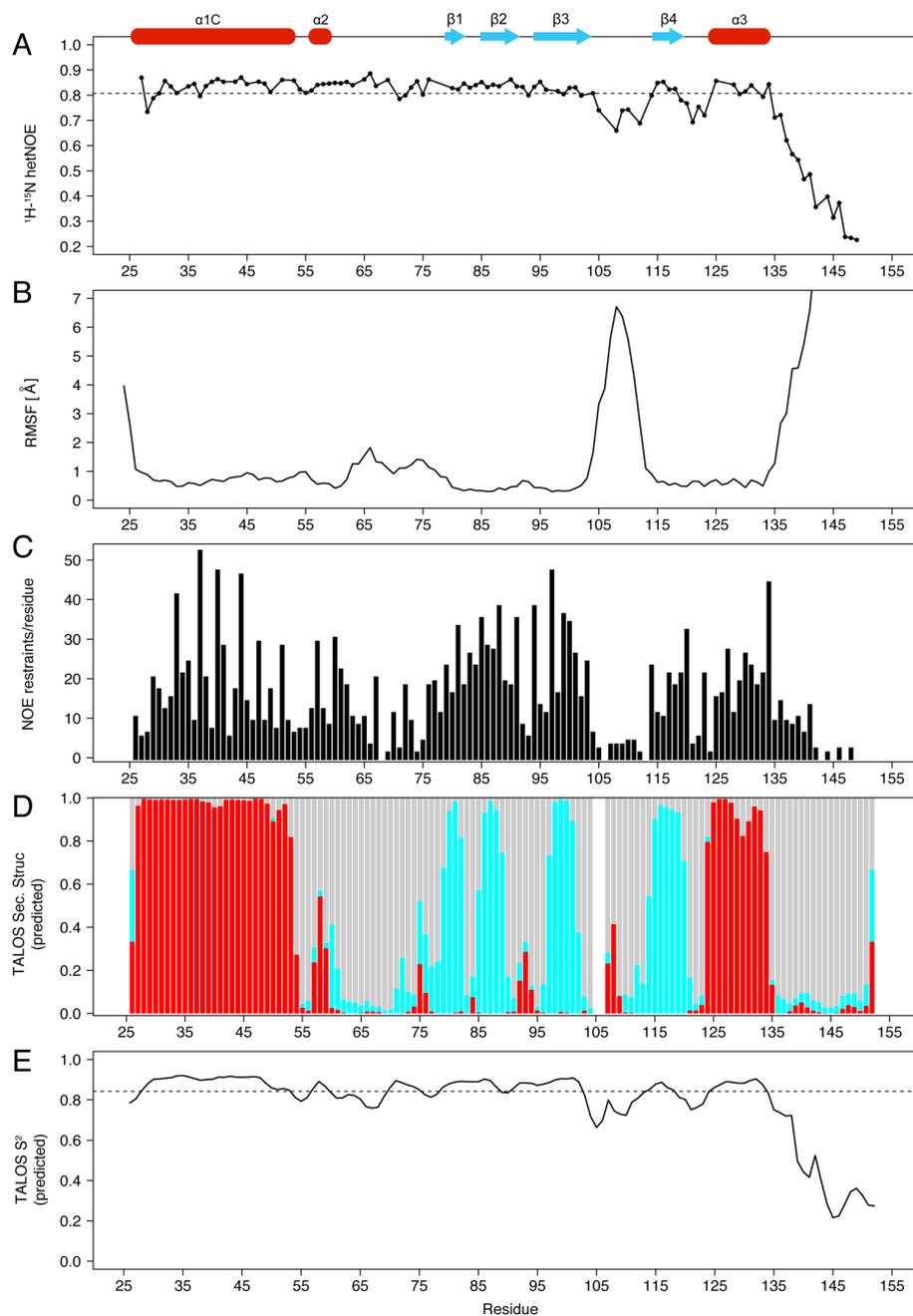
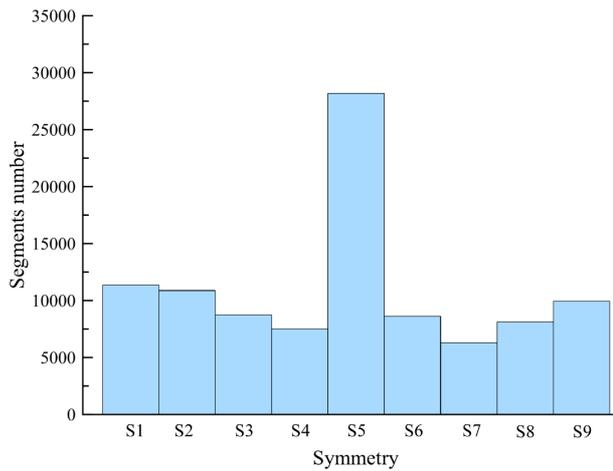


Figure S2; Related to Figure 1: Dynamics and flexibility of PpdDp in solution (A) ^1H - ^{15}N heteronuclear NOE of PpdDp in solution. Secondary structure elements are shown on top (helix in red, strand in blue). (B) Ensemble Root Mean Square Fluctuation (RMSF) of the NMR PpdDp bundle. (C) Number of NOE ($|i-j| > 1$) restraints per residue as assigned by ARIA. (D) Secondary structure (probabilities) predicted from chemical shifts using TALOS+ (red: helix, cyan: strand, grey: loop). (E) TALOS+ predicted order parameter S^2 .

A**B**

Symmetry	Twist	Rise
S1	94°	8.8Å
S2	96°	8.8Å
S3	98°	8.8Å
S4	94°	10.8Å
S5	96°	10.8Å
S6	98°	10.8Å
S7	94°	12.8Å
S8	96°	12.8Å
S9	98°	12.8Å

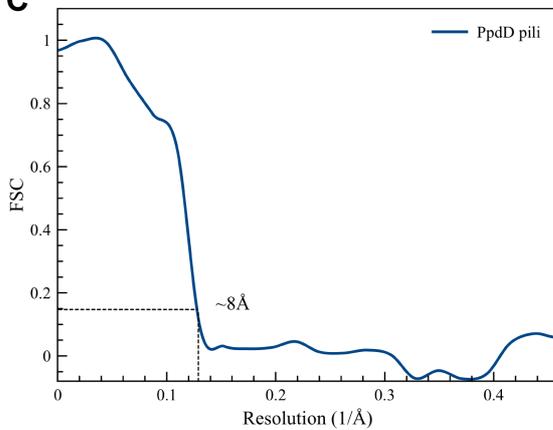
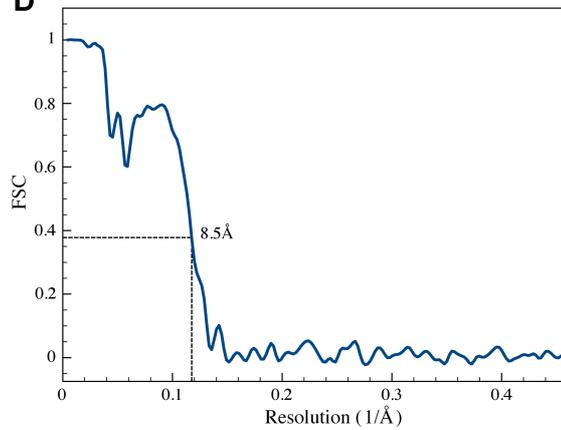
C**D**

Figure S3; Related to Figure 2: Large variability in terms of the helical symmetry of PpdD pili. (A) Histogram showing the symmetry distribution of PpdD pili segments. Nine reconstructions were generated with symmetry parameters shown in (B). These volumes were used for classifying the PpdD pili segments into different groups using a multi-reference alignment. (C,D) Fourier Shell Correlation (FSC) curves for the PpdD pili reconstruction. (C) The FSC of two half-maps shows a resolution of 8 Å at FSC=0.143. (D) The FSC curve for the refined PpdD pili model and the PpdD density map shows a resolution of 8.5 Å at FSC=0.38 ($=\sqrt{0.143}$).

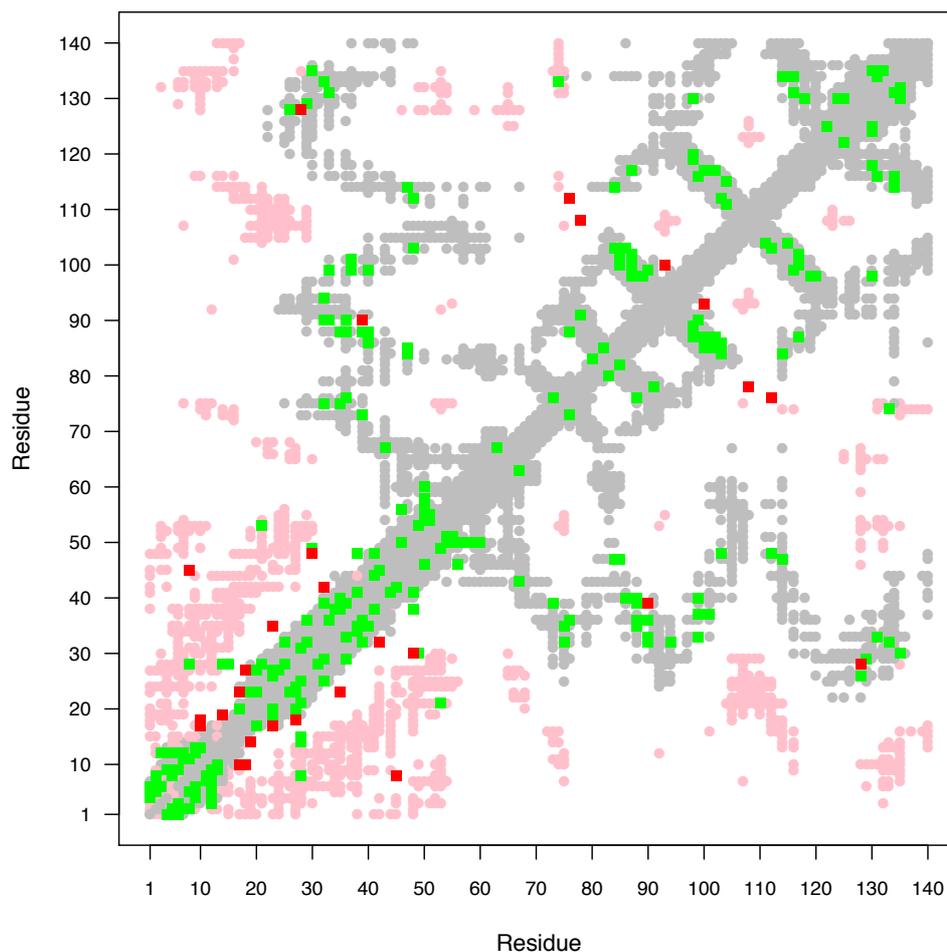


Figure S4; Related to Figure 3: EHEC PpdD pilus contact map and evolutionary contact predictions. Inter-residue contacts in PpdD pilus ($d_{ij} < 8 \text{ \AA}$) are shown in grey (intra-protomer) and pink (inter-protomer). High confidence evolutionary contact predictions from Gremlin (Kamisetty et al., 2013) are shown in green (consistent with PpdD pilus model with $d_{ij} < 10 \text{ \AA}$) and red (inconsistent).

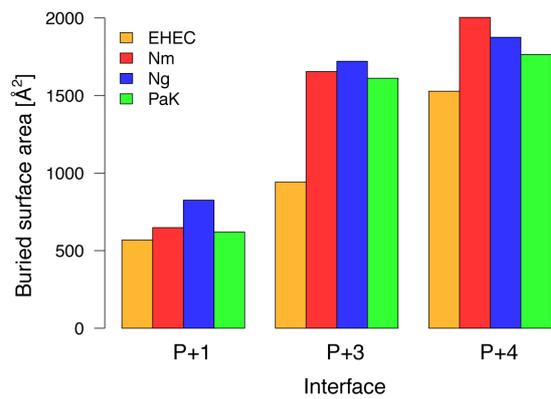
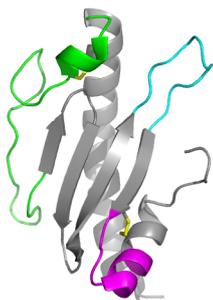
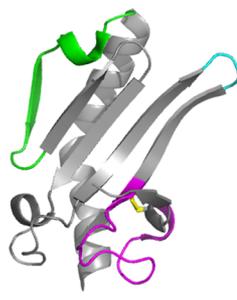


Figure S5; Related to Figure 6: Buried surface areas of inter-subunit interfaces in T4Pa structures. Surface area was determined using atomic solvent accessible areas of isolated and complexed pilin subunits computed with the NACCESS program (Hubbard and Thornton, 1993).

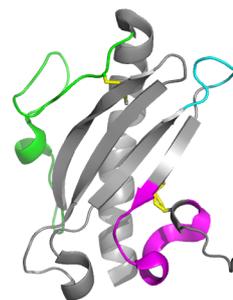
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



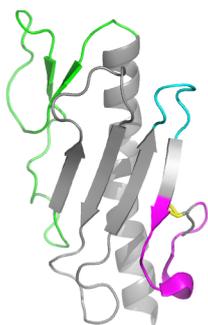
EHEC PpdD



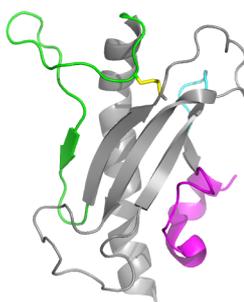
F. tularensis PilE
(3SOJ)



P. aeruginosa K122-4 PilA
(1QVE)



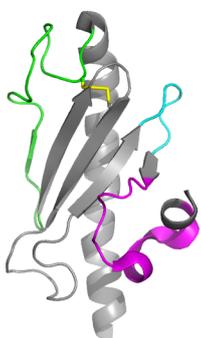
P. aeruginosa K PilA
(1OQW)



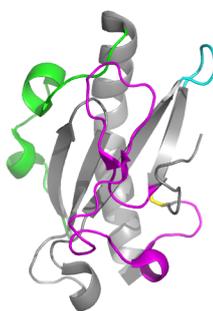
D. nodosus FimA
(3SOK)



P. aeruginosa Pa110594 pilA
(3JYZ)



A. baumannii PilA
(5IHJ)



N. meningitidis PilE
(5JW8)



N. gonorrhoeae PilE
(1AY2)

Figure S6; Related to Figure 1: Structural comparison of major T4a pilins (periplasmic domains). α/β loop is shown in green, β_3/β_4 loop in cyan and D-region in magenta.

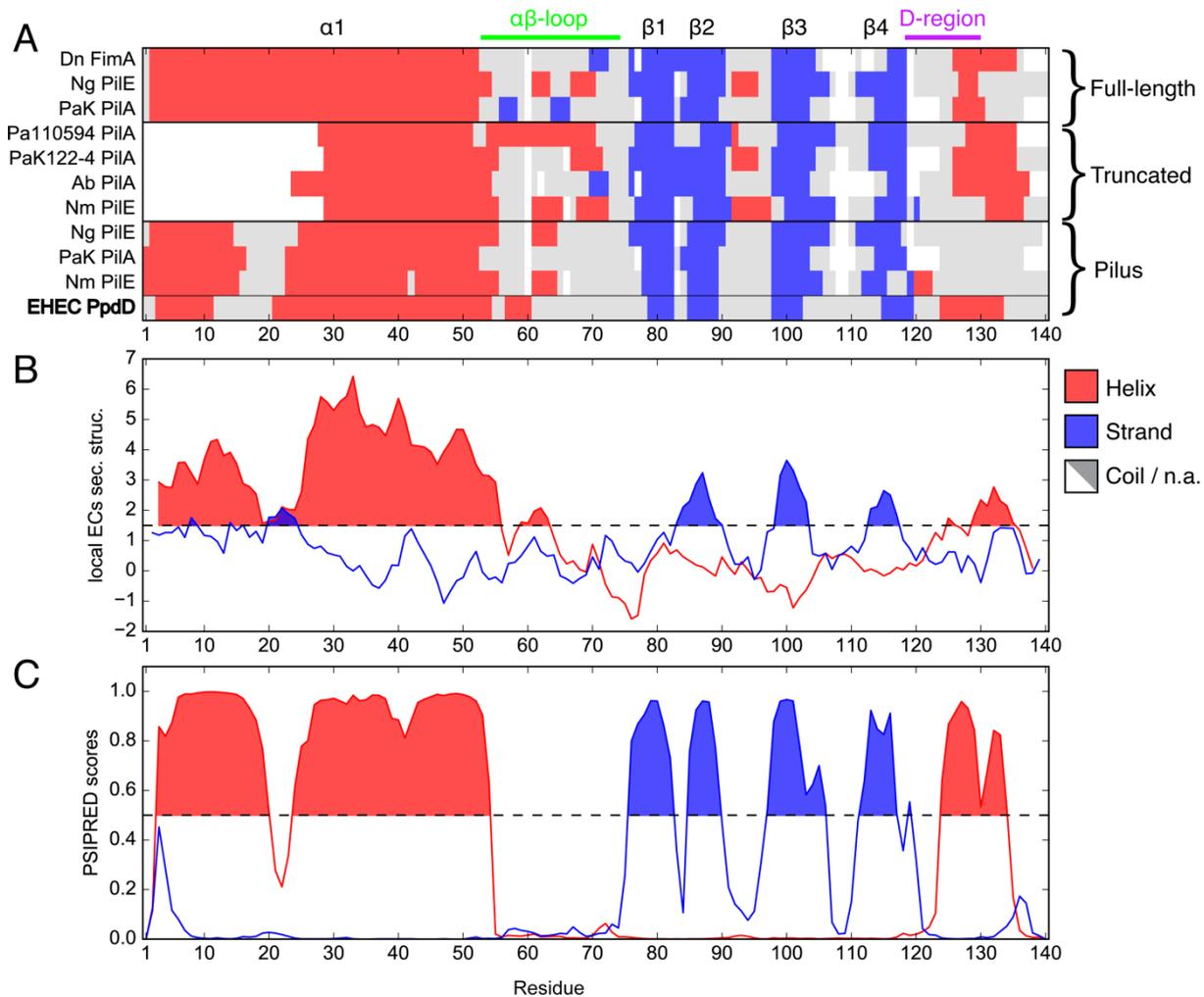


Figure S7; Related to Figure 7: Secondary structure elements in T4a pilins. (A) Position of secondary structures elements of T4a pilin structures aligned on EHEC PpdD sequence and grouped according to the constructs and techniques used to solve the structure (Structures of full-length pilins solved by X-ray, truncated pilins missing the transmembrane segment and pilus structures solved by cryoEM). (B) Secondary structure score predicted from local evolutionary contacts within the T4a pilin family (Toth-Petroczy et al., 2016). (C) Secondary structure confidence score predicted with PSIPRED (Jones, 1999) on the EHEC PpdD pilin sequence.

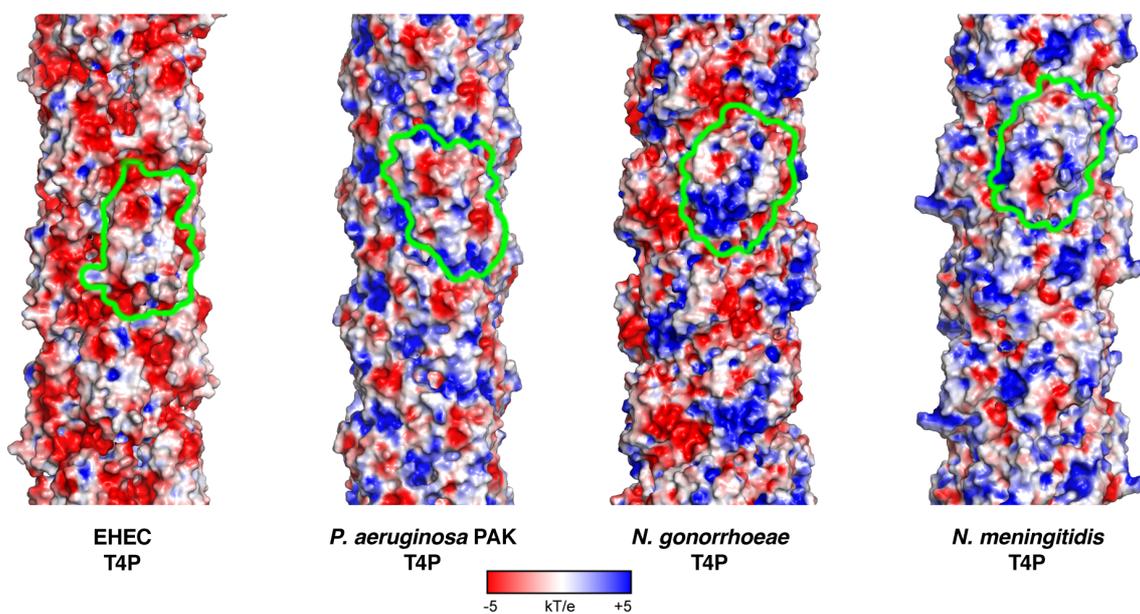


Figure S8; Related to Figure 7; Surface electrostatics potential of T4P pilus structure.

Electrostatics potential calculations were performed using APBS (Baker et al., 2001). The green lines delineate the surface of a single pilin subunit.

Supplementary references

- 1
2
3 Baker, N., Sept, D., Joseph, S., Holst, M., and McCammon, J. (2001). Electrostatics of
4 nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98,
5 10037-10041.
- 6 Bartolome, B., Jubete, Y., Martinez, E., and de la Cruz, F. (1991). Construction and
7 properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its
8 derivatives. *Gene* 102, 75-78.
- 9 de Amorim, G.C., Cisneros, D.A., Delepierre, M., Francetic, O., and Izadi-Pruneyre, N.
10 (2014). ¹H, ¹⁵N and ¹³C resonance assignments of PpdD, a type IV pilin from
11 enterohemorrhagic *Escherichia coli*. *Biomol NMR Assign* 8, 43-46.
- 12 Hubbard, S.J., and Thornton, J.M. (1993). NACCESS Computer Program. Department of
13 Biochemistry and Molecular Biology, University College London.
- 14 Jones, D. (1999). Protein secondary structure prediction based on position-specific scoring
15 matrices. *J Mol Biol* 292, 195-202.
- 16 Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-
17 based residue-residue contact predictions in a sequence- and structure-rich era. *Proc Natl*
18 *Acad Sci U S A* 110, 15674–15679.
- 19 Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system
20 based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* 95, 5752-
21 5756.
- 22 Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein
23 side-chain conformations with SCWRL4. *Proteins* 77, 778–795.
- 24 Luna Rico, A., Zheng, W., Petiot, N., Egelman, E.H., and Francetic, O. (2019). Functional
25 reconstitution of the type IVa assembly system from enterohemorrhagic *Escherichia coli*.
26 *Mol. Microbiol.* 111, doi: 10.1111/mmi.14188, *in press*.
- 27 Nivaskumar, M., Santos-Moreno, J., Malosse, C., Nadeau, N., Chamot-Rooke, J., Tran Van
28 Nhieu, G., and Francetic, O. (2016). Pseudopilin residue E5 is essential for recruitment by
29 the type 2 secretion system assembly platform. *Mol Microbiol* 101, 924-941.
- 30 Shapovalov, M.V., and Dunbrack, R.L. (2011). A smoothed backbone-dependent rotamer
31 library for proteins derived from adaptive kernel density estimates and regressions. *Structure*
32 19, 844–858.
- 33 Toth-Petroczy, A., Palmedo, P., Ingraham, J., Hopf, T.A., Berger, B., Sander, C., and Marks,
34 D.S. (2016). Structured states of disordered proteins from genomic sequences. *Cell* 167,
35 158–170.e112.
- 36 Yanisch-Perron, C., Vieira, J., and J., M. (1985). Improved M13 phage cloning vectors and
37 host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103-119.
- 38 Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based
39 on the TM-score. *Nucleic Acids Res* 33, 2302–2309.
- 40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65