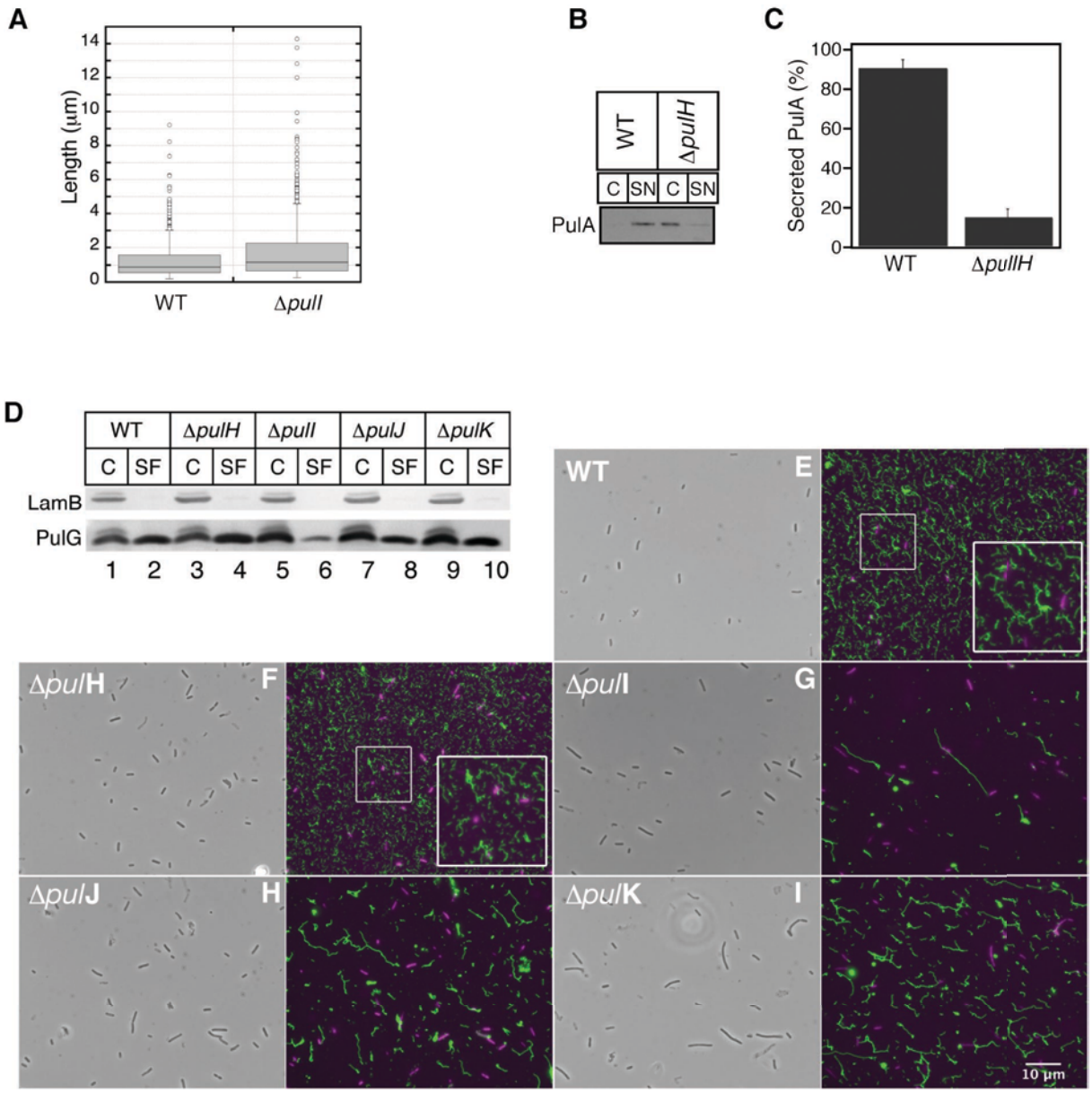
**Supplementary Information**

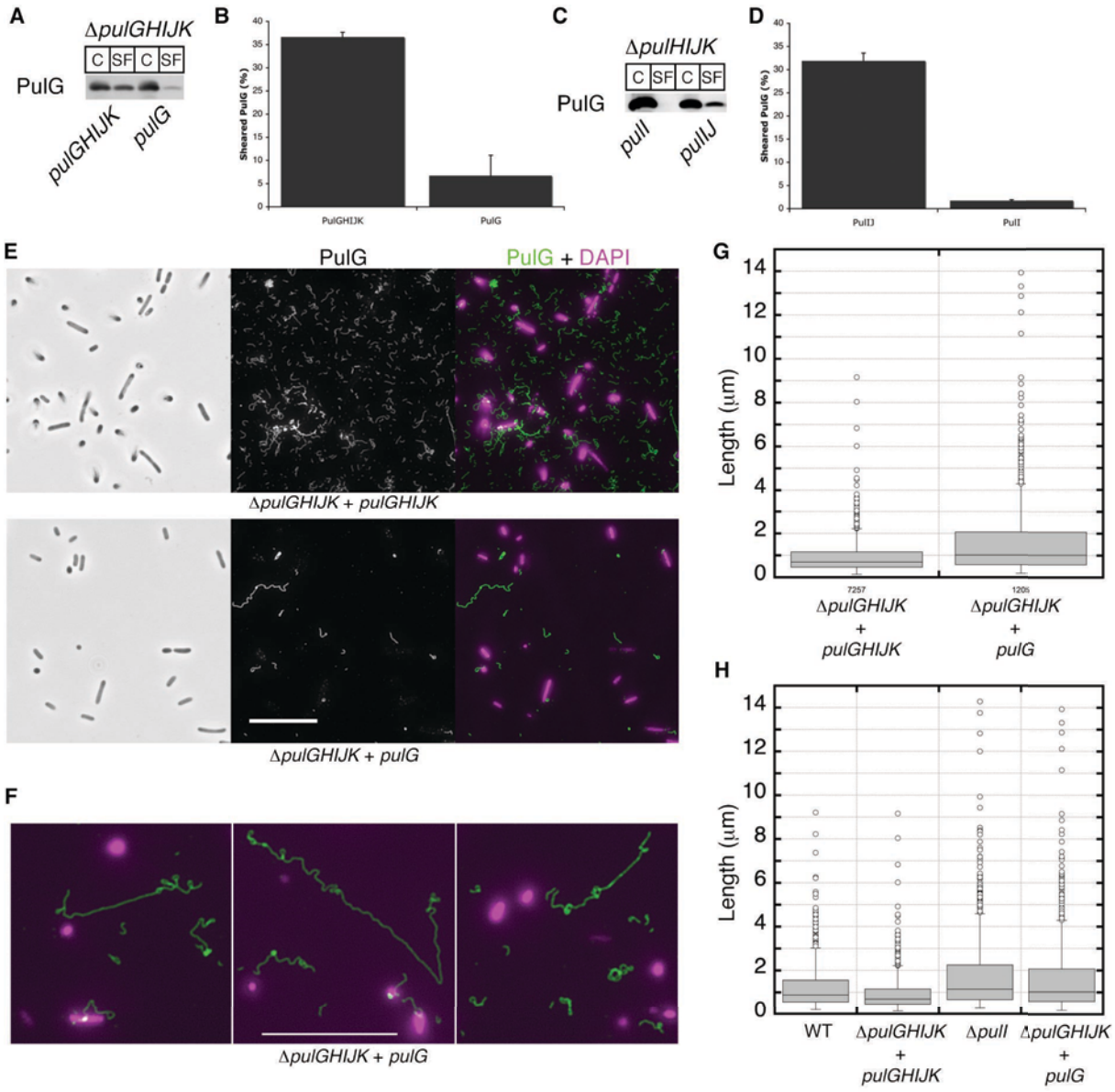
# Minor pseudopilin self-assembly primes type II secretion pseudopilus elongation

David A. Cisneros[1,2*], Peter J. Bond[3], Anthony P. Pugsley[1,2], Manuel Campos[1,2] and Olivera Francetic[1,2*]

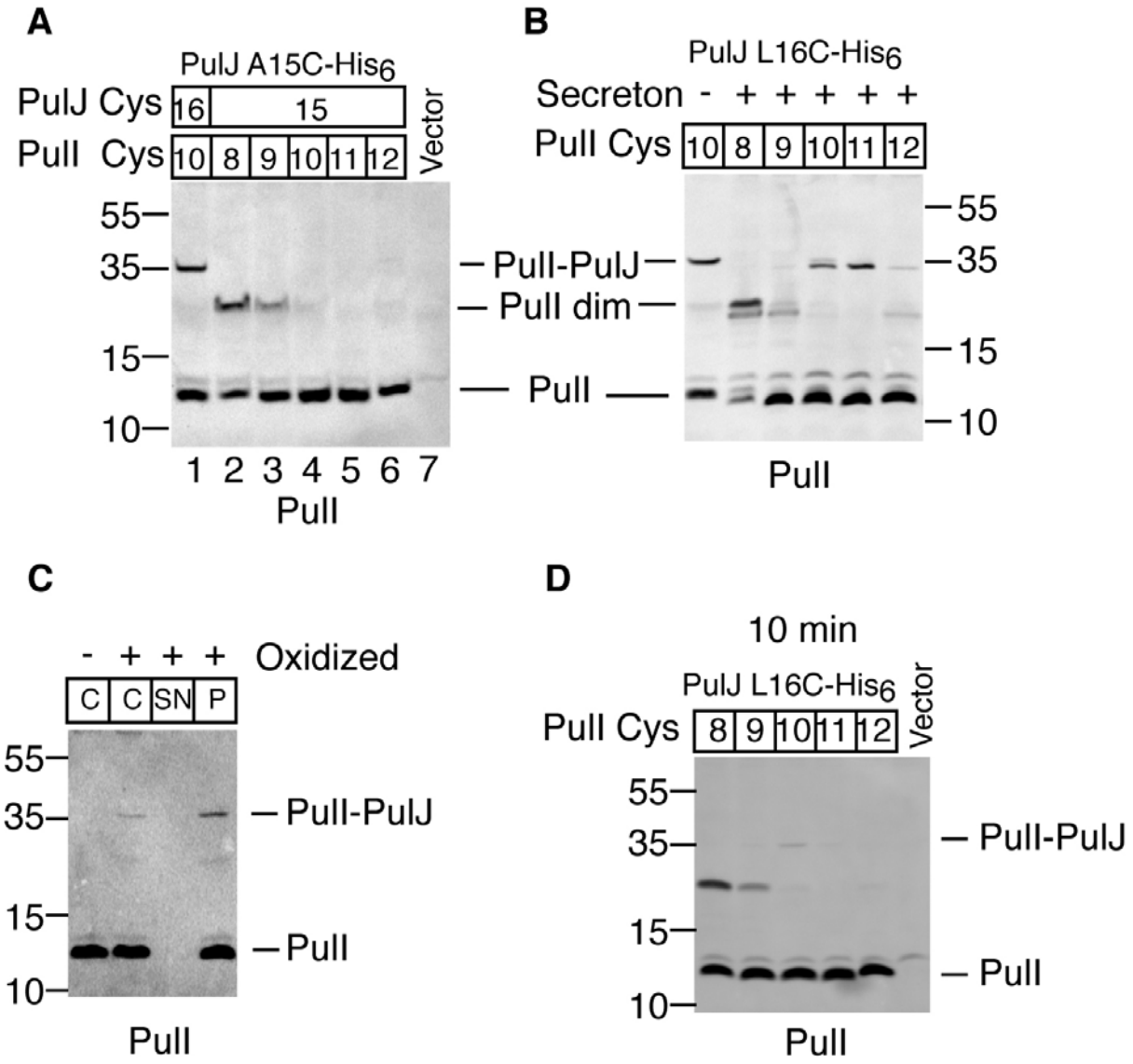
[1]Institut Pasteur, Molecular Genetics Unit, 25, rue du Dr. Roux, Paris CEDEX 15, France.
[2]CNRS URA 2172, 25, rue du Dr. Roux, Paris CEDEX 15, France
[3]The Unilever Centre for Molecular Science Informatics, Department of Chemistry, Lensfield Road, University of Cambridge, Cambridge, CB2 1EW, UK



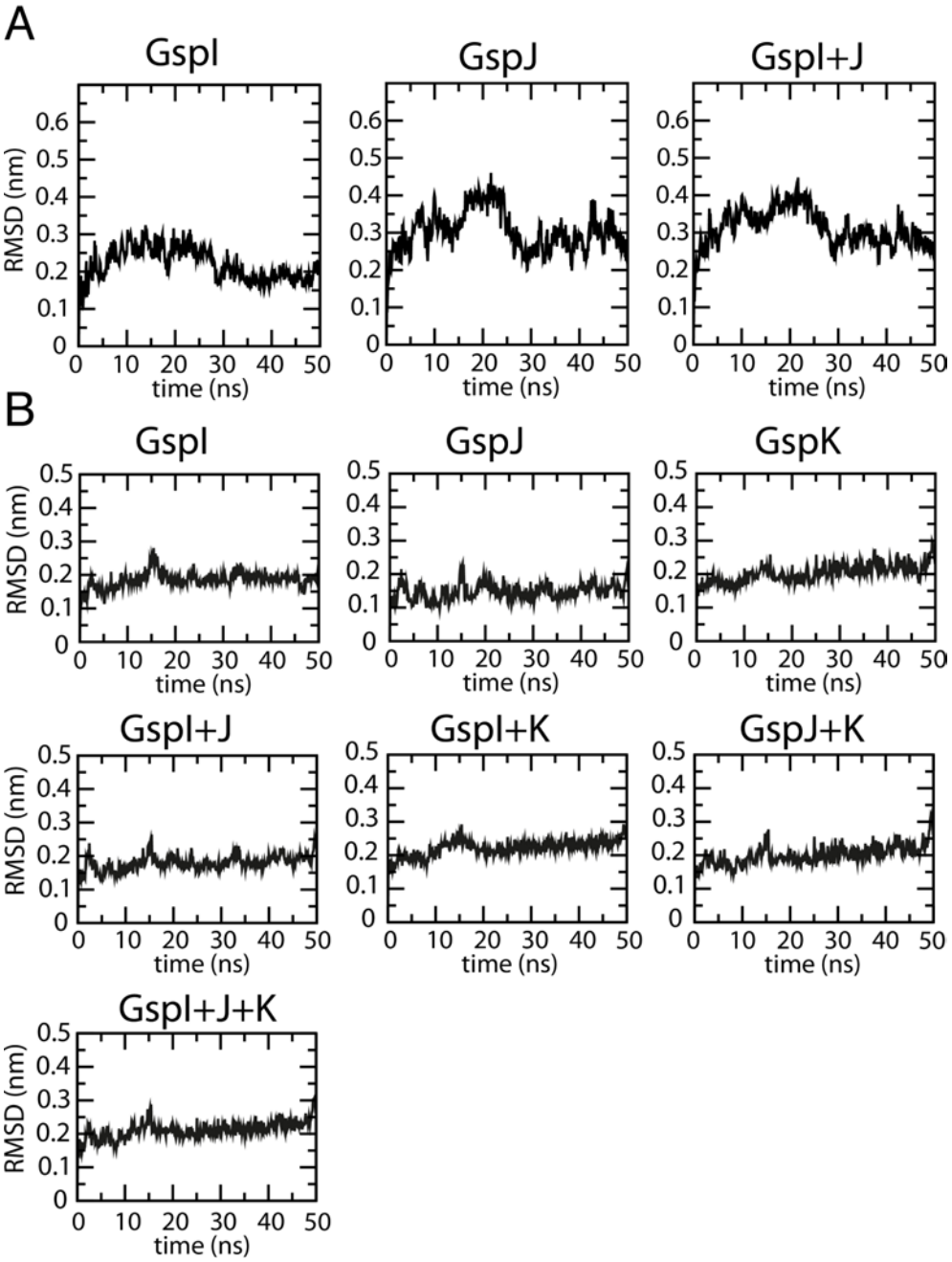

**Supplementary Figure S1. Assembly of PulG pili in single deletion mutants of the minor pseudopilins. (A)** Length analysis of pili observed in Fig. 1C (main text). The data correspond to 484 and 455 pili measured from the WT and the *ΔpulI* mutant respectively from two independent experiments. The difference in length is statistically significant (Mann-Whitney test, $p<0.0001$). **(B)** Secreted PulA immunodetection in WT and *ΔpulH* bacteria expressing low *pul* genes at near chromosomal levels in strain PAP5207. **(C)** Percentage of secreted PulA (mean + SD) quantified from four independent experiments like the one shown in (B). **(D)** PulG immunodetection in total cell extract (C) and sheared fractions (SF) of *E. coli* expressing the wild type *pul* operons (WT) on a plasmid or its minor pseudopilin single deletion derivatives with an additional copy of the *pulG* gene in a high-copy number plasmid. **(E-I)** Immunofluorescence and phase contrast microscopy showing strains analyzed in (B). PulG staining is shown in green and DAPI staining in false magenta color. Scale bar corresponds to 10 µm.



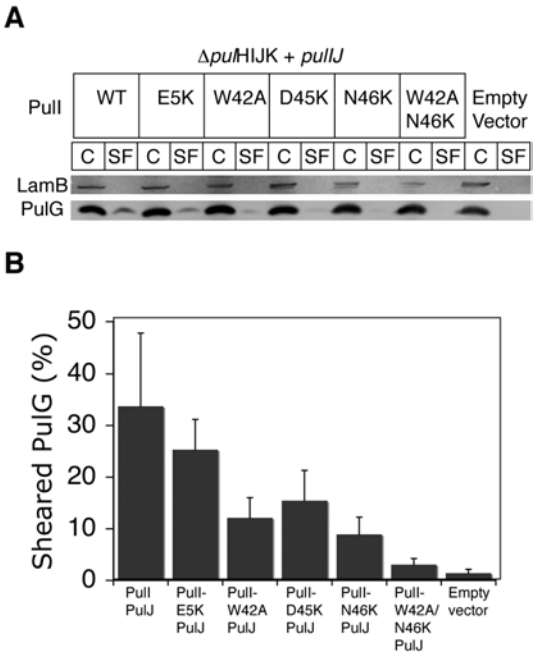

**Supplementary Figure S2. Assembly of PulG pili in a Δ*pulGHIJK* mutant overproducing PulG. (A)** PulG immunodetection in total cell extract and sheared fractions (C and SF) of *E. coli* expressing the *pul* genes on a plasmid with a pseudopilin deletion mutation (*ΔpulGHIJK*) complemented with the *pulG* or *pulGHIJK* genes in a high-copy number plasmid. **(B)** Percentage of PulG in SF (median) like the one shown in (A). The error bar corresponds to the difference between two independent samples. **(C)** PulG immunodetection in total cell extract and sheared fractions (C and SF) of *E. coli* expressing the *pul* genes on a plasmid with a minor pseudopilin deletion mutation (*ΔpulHIJK*) complemented with the *pulI* or *pulIJ* genes in a high-copy number plasmid. **(D)** Percentage of PulG in SF (mean + SD) from four independent samples like the one shown in (C). **(E)** Phase contrast and immunofluorescence microscopy of *E. coli* expressing a *ΔpulGHIJK* Pul secreton from a plasmid and *pulG* (top) or *pulGHIJK* (bottom) from a high copy-number plasmid. PulG staining is shown in green and DAPI staining in false magenta color. Scale bar corresponds to 10 µm. **(F)** Selected immunofluorescence images of pili assembled by the Δ*pulGHIJK* mutant in bacteria over-producing PulG. Scale bar corresponds to 10 µm. **(G)** Length analysis of pili observed in (A). The data correspond to 1197 and 1231 pili measured from the *ΔpulGHIJK* mutant complemented with *pulGHIJK* and *pulG* respectively from three independent experiments. The difference in length is statistically significant (Mann-Whitney test, $p<0.0001$). **(H)** Length analysis of pili observed in (A) compared to *ΔpulI* and WT shown in Supplementary Fig. S1A.



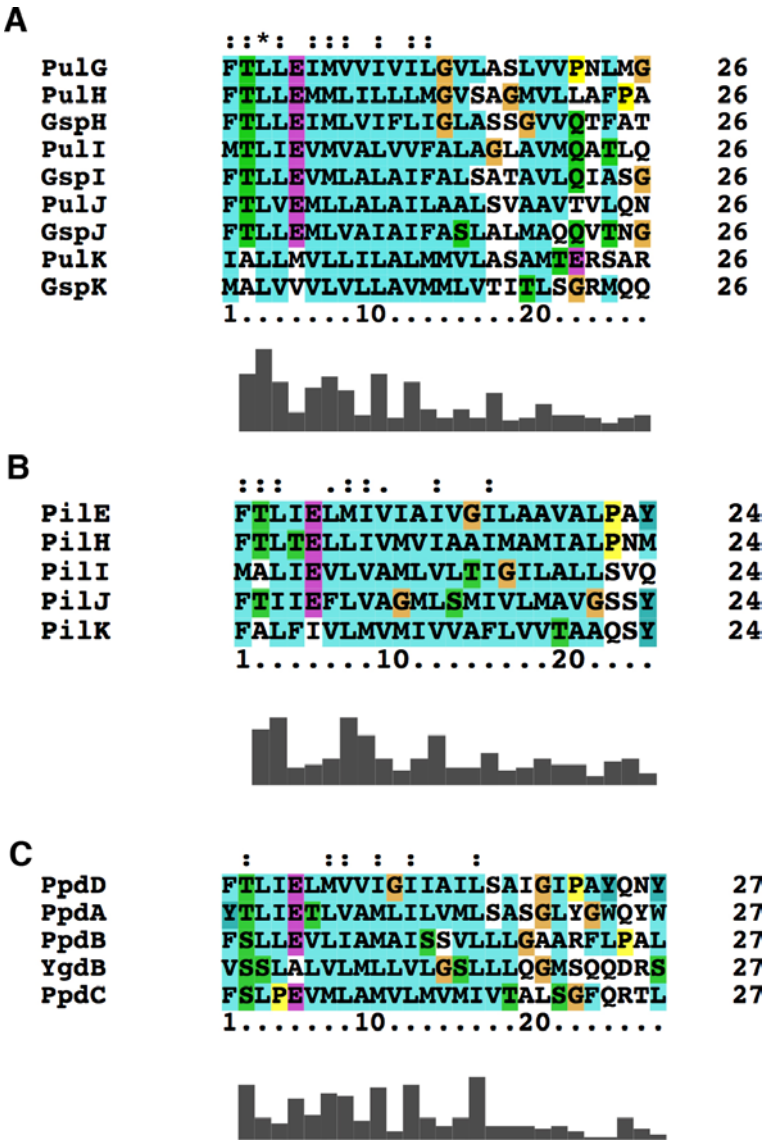

**Supplementary Figure S3. PulI and PulJ cross-link in the membrane independently on the secretion system. (A)** Total extracts of *E coli* co-producing PulJA15C-His6 with PulIV8C, PulIA9C, PulIL10C, PulIV11C, PulIV12C or empty vector (pSU18) were chemically oxidized for 60 min and analyzed using an anti-PulI antibody. **(B)** Total extracts of bacteria co-producing the *ΔpulHIJK* Pul secreton (plasmid pCHAP8296) and PulJL16C-His6 with PulIV8C, PulIA9C, PulIL10C, PulIV11C and PulIV12C chemically oxidized for 60 min and analyzed using an anti-PulI antibody. **(C)** Bacteria co-producing PulJL16C-His6 with PulIL10C were chemically oxidized for 10 minutes, harvested, sonicated and ultracentrifuged; cell extract (C), soluble fraction (SN) and membrane fraction (P) analyzed with an anti-PulI antibody. **(D)** Total extracts of *E coli* co-producing PulJA16C-His6 with PulIV8C, PulIA9C, PulIL10C, PulIV11C, PulIV12C or the empty vector were chemically oxidized for 10 min and analyzed using an anti-PulI antibody.




**Supplementary Figure S4. Stability of GspI, GspJ and GspK complexes during MD simulations. (A)** The C-α root mean square deviation (RMSD) with respect to the starting structure for both individual chains (GspI or GspJ) and for the their complex (GspI+J) during the GspI-GspJ heterodimer simulation 1. **(B)** The C-α RMSD with respect to the starting structure for individual chains (GspI, GspJ or GspK) and complexes thereof (GspI+J+K, GspI+J, GspI+K, etc.) during the GspI-GspJ heterotrimer simulation. In both cases, the fact that RMSD of the complexes are similar in magnitude to those of the individual chains indicates that neither the heterodimeric nor the heterotrimeric complexes dissociated.

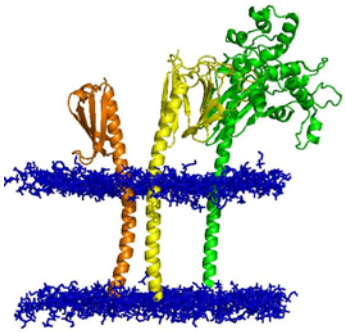



**Supplementary Figure S5. Initiation of pilus assembly by PulJ with PulI mutant variants. (A)** PulG immunodetection in total cell and sheared fractions (C+SF) (from 0.05 $OD_{600nm}$) of WT and the Δ*pulHIJK* mutant complemented with a plasmid carrying *pulIJ* genes encoding either WT PulI or its variants PulIE5A, PulIW42A, PulID45K, N46K or PulIW42A/N46K. **(B)** Percentage of PulG in the SF (mean + SD) from 4 independent experiments like the one shown in (A).

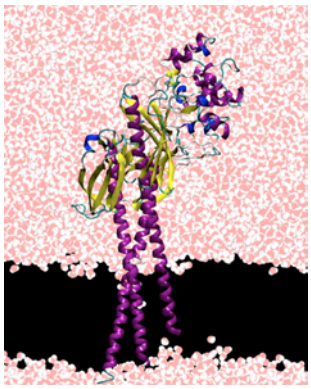



**Supplementary Figure S6. Primary sequence alignment of transmembrane segments in pilins of T2SS and T4P. (A)** Conservation of the TM segments between the minor pseudopilins of the Pul secreton and the minor pseudopilins of the ETEC Gsp T2SS. **(B)** Alignment of the TM segments of the major pilin (PilE) and the minor pilins of the T4P of *Neisseria meningitidis*. **(C)** Alignment of the TM segments of the major pilin (PpdD) and the minor pilins of the T4P of Enterohaemorrhagic *E. coli* (Sauvonnet *et al*, 2000a; Xicohtencatl-Cortes *et al*, 2007).

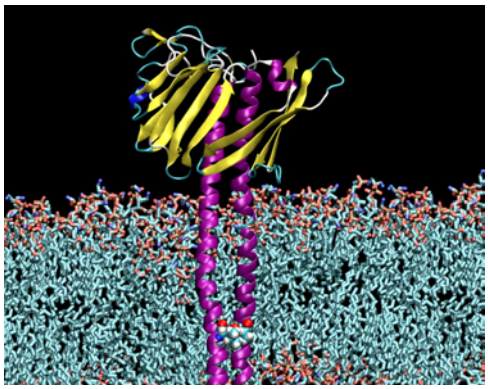
**Supplementary videos**



**Supplementary Video S1.** Superposition of three molecular dynamics simulations of ETEC's GspI (orange), GspJ (yellow) and GspK (green) in palmitoyl-oleoyl ethanolamine (POPE) modeled membranes. Water and membranes are deleted for visual purposes. Only the phospholipid headgroup atoms are shown in blue.



**Supplementary Video S2.** Molecular dynamics simulation of ETEC's GspI-GspJ-GspK complex in a POPE modeled membrane. Water molecules are shown in pink (oxygen) and white (hydrogen). The POPE membrane was deleted from this video for visual purposes to show the accumulation of water close to the GspK N-terminus, indicating a membrane deformation. The black area denotes the hydrophobic area of the membrane.



**Supplementary Video S3.** Molecular dynamics simulation of ETEC's GspI-GspJ complex in a POPE modeled membrane. Residues L10 in GspI and L16 in GspJ are shown as spheres. Water was deleted from this video for visual purposes.

**Supplementary materials and methods.**

***Molecular dynamics simulations***

To generate full-length models of GspI, GspJ and GspK, we extended the structures of their globular domains (pdb code: 3CI0) by modeling the missing N-terminal hydrophobic segments as straight alpha helices. The Gsp minor pseudopilins share approximately 50% similarity with the corresponding Pul pseudopilins. Simulations were performed using the GROMACS package (Hess *et al*, 2008) version 4.5 (Bjelkmar *et al*, 2010). The protein was treated using the CHARMM22/CMAP force field (MacKerell *et al*, 1998), and the lipid molecules using the new CHARMM36 parameter set (MacKerell *et al*, 2010). Equations of motion were integrated using the leapfrog method with a 2 fs time step, and the LINCS algorithm was used to constrain bond lengths (Hess *et al*, 1997). Electrostatic interactions were computed using the Particle-Mesh-Ewald (PME) algorithm (Essmann *et al*, 1995) and the real-space sum was cut off at 12 Å. van der Waals interactions were switched off between 10 Å and 12 Å. Neighbor list updates were performed every 10 steps. Simulations were performed using conditions of constant temperature (310 K) and pressure (1 atm), using the thermostat of Bussi *et al*. (Bussi *et al*, 2007) and semi-isotropic pressure-coupling to a Parrinello-Rahman barostat (Parrinello and Rahman, 1981) with a coupling constant of 1 ps, under periodic-boundary conditions. At each stage of set-up, Steepest Descent energy minimization were performed to relax the protein geometry and to remove steric clashes. All ionisable groups were assigned their most probable charged states at neutral pH. The dimeric protein complexes were placed in pre-equilibrated membrane systems with TIP3P water and a ~0.1 M concentration of NaCl, with dimensions ~9x9x16 $nm^3$ (dimer) or ~9x9x21 $nm^3$ (trimer), and overlapping lipid or solvent molecules were removed. The final dimer system contained ~29,000 water molecules and 308 lipids; the final trimer system contained ~37,000 water molecules and 298 lipids. Each system was equilibrated over 50 ns, during which position restraints, applied to all non-hydrogen protein atoms, were gradually removed to relax the protein structure, membrane and solvent. Finally, 50 ns production MD simulations were carried out. Visual analysis and preparation of molecular graphics figures was performed using VMD (Humphrey *et al*, 1996). Graphs were prepared with Grace (http://plasma-gate.weizmann.ac.il/Grace/) and Gnuplot (http://www.gnuplot.info/). Further analysis was performed using GROMACS and additional in-house code. Systems containing the individual GspI, GspJ, and GspK monomers within a lipid bilayer were also simulated. The monomeric proteins were placed in pre-equilibrated membrane/solvent systems so that

they had the same initial position and orientation as in the trimeric complex simulation systems. The same system size was used as for the trimer simulation, in order to avoid any initial bias. The final monomer systems contained ~38,000 water molecules and ~310 lipids. Following position-restrained equilibration, 20 ns production MD simulations were carried out. To characterize kinking in the helix-hinge-helix units of GspI, GspJ, and GspJ, clustering analysis of each simulation was carried out with GROMACS 4.5 (Hess *et al*, 2008) using the algorithm described in (Daura *et al*, 1999), with a neighbour-list cut-off in the pair-wise RMS difference of 0.25 nm. Trajectories were first least-squares fitted to the backbone atoms of the upper regions of each N terminal helix (residues 41-53, 36-57, and 29-58 in GspI, GspJ, and GspK, respectively). Clustering was then performed on the backbone atoms of the bottom regions of each N terminal helix (residues 5-41, 5-46 and 5-29 in GspI, GspJ, and GspK, respectively). The three main clusters of the GspI-GspJ, GspI-GspJ-GspK and GspK alone were present ~50% of the time in each simulation, while the 6 main clusters of the GspI alone and GspJ alone were present 50% of the time.

**Supplementary Table S1.** Plasmid list.

| Plasmid name | Plasmid number* | Genes/description | Resistance** Origin | Reference. |
|---|---|---|---|---|
| pPulIT18 | pCHAP8245 | *T18- pulI* | Ap, ColE1 | This study |
| pPulHT25 | pCHAP8257 | *T25-pulH-* | Km, p15A | This study |
| pPulIT25 | pCHAP8248 | *T25-pulI* | Km, p15A | This study |
| pPulJT25 | pCHAP8249 | *T25-pulJ* | Km, p15A | This study |
| pPulKT25 | pCHAP8250 | *T25-pulK* | Km, p15A | This study |
| pT25-Zip | - | GCN4 (zipper region) fused to T25 | Km, p15A | (Karimova *et al*, 1998) |
| pT18-Zip | - | GCN4 (zipper region) fused to T18 | Ap, ColE1 | (Karimova *et al*, 1998) |
| pKT25 | - | Cloning vector for T25 gene fusion | Km, p15A | (Karimova *et al*, 1998) |
| pUT18 | - | Cloning vector for T18 gene fusion | Ap, ColE1 | (Karimova *et al*, 1998) |
| pCHAP231 | pCHAP231 | All *pul* genes: *pulS-pulAB-pulCDEFGHIJKLMNO* | Ap, ColE1 | (d'Enfert *et al*, 1987) |
| pPulΔGHIJK | pCHAP7248 | All *pul* genes except *pulGHIJK* | Ap, ColE1 | This study |
| pPulΔHIJK | pCHAP8296 | All *pul* genes except *pulHIJK* | Ap, ColE1 | This study |
| - | pCHAP8403 | All *pul* genes except *pulD* and *pulHIJK* | Ap, ColE1 | This study |
| - | pCHAP1226 | All *pul* genes except *pulD* | Ap, ColE1 | (Sauvonnet *et al*, 2000b) |
| - | pCHAP1324 | All *pul* genes except *pulH* | Ap, ColE1 | (Sauvonnet *et al*, 2000b) |
| - | pCHAP1357 | All *pul* genes except *pulI* | Ap, ColE1 | (Sauvonnet *et al*, 2000b) |
| - | pCHAP1323 | All *pul* genes except *pulJ* | Ap, ColE1 | (Sauvonnet *et al*, 2000b) |
| - | pCHAP1325 | All *pul* genes except *pulK* | Ap, ColE1 | (Sauvonnet *et al*, 2000b) |
| - | pCHAP1205 | *pulG* | Cm, p15A | (Sauvonnet *et al*, 2000b) |
| pPulHIJK | pCHAP7273 | *pulHIJK* | Cm, p15A | This study |
| pPulHJ | pCHAP6108 | *pulHJ* | Cm, p15A | This study |
| pPulHI | pCHAP7269 | *pulHI* | Cm, p15A | This study |
| pPulIJ | pCHAP7267 | *pulIJ* | Cm, p15A | This study |
| pPulIK | pCHAP7295 | *pulIK* | Cm, p15A | This study |
| - | pCHAP6134 | *pulI8C-pulJ16C-his* | Cm, p15A | This study |
| - | pCHAP6135 | *pulI9C-pulJ16C-his* | Cm, p15A | This study |
| - | pCHAP6136 | *pulI10C-pulJ16C-his* | Cm, p15A | This study |
| - | pCHAP6155 | *pulI11C-pulJ16C-his* | Cm, p15A | This study |
| - | pCHAP6138 | *pulI12C-pulJ16C-his* | Cm, p15A | This study |
| - | pCHAP6157 | *pulI8C-pulJ 15Chis* | Cm, p15A | This study |

| | | | | |
|---|---|---|---|---|
| | pCHAP6158 | *pulI9C pulJ15C-his* | Cm, p15A | This study |
| - | pCHAP6147 | *pulI10C-pulJ15C-his* | Cm, p15A | This study |
| - | pCHAP6137 | *pulI11C-pulJ15C-his* | Cm, p15A | This study |
| - | pCHAP6159 | *pulI12C-pulJ15C-his* | Cm, p15A | This study |
| - | pCHAP6160 | *pulI10C-pulJ16C* | Cm, p15A | This study |
| pPulGHIJK-His | pCHAP6104 | *pulGHIJK-His* | Cm, p15A | This study |
| pPulGHI10C-J16C-K-His | pCHAP6186 | *pulGHIJK-His, with pulI10C-pulJ16C* | Cm, p15A | This study |
| pPulGHI10C-J16C-K | pCHAP6185 | *pulGHIJK-His, with pulI10C-pulJ16C* | Cm, p15A | This study |
| - | pCHAP8375 | *pulK* | Ap, ColE1 | This study |
| - | pCHAP6175 | *pulI16C* | Cm, p15A | This study |
| - | pCHAP6190 | *pulK –I9C* | Ap, ColE1 | This study |
| - | pCHAP6191 | *pulK –L10C* | Ap, ColE1 | This study |
| - | pCHAP6192 | *pulK –A11C* | Ap, ColE1 | This study |
| | pCHAP6178 | *pulI L10C/ E5K, pulJL16C* | Cm, p15A | This study |
| | pCHAP6181 | *pulI L10C/ N46K, pulJL16C* | Cm, p15A | This study |
| | pCHAP6182 | *pulI L10C/ D45K, pulJL16C* | Cm, p15A | This study |
| | pCHAP6139 | *pulIE5K, pulJ* | Cm, p15A | This study |
| | pCHAP6187 | *pulI D45K, pulJ* | Cm, p15A | This study |
| | pCHAP6188 | *pulI N46K, pulJ* | Cm, p15A | This study |
| | pCHAP8432 | *pulAsolpulΔHIJK* | Ap, ColE1 | This study |
| | pCHAP8185 | *pulAsol* all *pul* genes | Ap, ColE1 | This study |
| | pCHAP8201 | pCHAP8185*ΔpulH::kan* | Ap, ColE1 | This study |
| | pCHAP8184 | pCHAP8185 *ΔpulG* | Ap, ColE1 | Campos et al |
| | pCHAP8218 | pCHAP8185 *ΔpulI::kan* | Ap, ColE1 | This study |
| | pCHAP8209 | pCHAP8185 *ΔpulJ::kan* | Ap, ColE1 | This study |
| | pCHAP8212 | pCHAP8185 *ΔpulK::kan* | Ap, ColE1 | This study |
| | pCHAP8255 | pulA+ | Cm, p15A | This study |
| | pSU18 | | Cm, p15A | (Bartolome *et al*, 1991) |
| | pCHAP6184 | *pulI W42A, pulJ* | Cm, p15A | This study |
| | pCHAP6189 | *pulI W42A/N46K, pulJ* | Cm, p15A | This study |
| | pCHAP 6210 | *pulIJK* | Cm, p15A | This study |
| | pCHAP 6211 | *pulI N46K, pulJ, pulK* | Cm, p15A | This study |
| | pCHAP 6212 | *pulI W42A/N46K, pulJ,pulK* | Cm, p15A | This study |
| | pCHAP6179 | *pulI L10C/ E35K, pulJL16C* | Cm, p15A | This study |

| | | | | |
|---|---|---|---|---|
| | pCHAP6180 | *pulI L10C/ W42A, pulJL16C* | Cm, p15A | This study |
| | pCHAP6183 | *pulI L10C/ W42A/N46K, pulJL16C* | Cm, p15A | This study |

*Lab collection number
**Ap=ampicillin, Cm=chloramphenicol, Km=kanamycin, Ze=zeocin

**Supplementary Table S2.** Oligonucleotides used for the bacterial two-hybrid constructs.

| Primer name | Primer sequence |
|---|---|
| PulH5 | CGACAggtaccCTTTACGTTGCTGGAGATGATGC |
| pulH3 | GGACTACgaatTCATTGCGCCTCCTGCGGTTCG |
| pulI5 | GCAggtaccTATGACGCTGATTGAAGTCATGG |
| pulI3 | GCTTTAGCGCgaattcATGGCGATGTCACGTAGGTGC |
| pulJ5 | GCTCggtaccCTTTACCCTCGTCGAAATGCTG |
| pulJ3 | CTGTCAggtaccATCGGCTGTCTCCCGGCGTAAGC |
| pulK5 | GCGAggtaccCATCGCCCTGCTCATGGTGCTG |
| pulK3 | GGTTTAgaattcATTCATCGGCTACCCAGTAG |
| PulIE5K5 | GCAAggtaccTATGACGCTGATTAAAGTCATGG |

**Supplementary references**


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

Bjelkmar P, Larsson P, Cuendet MA, Hess B and Lindahl E (2010) Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput*, 6: 459-466.

Bussi G, Donadio D and Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys*, 126.

d'Enfert C, Ryter A and Pugsley AP (1987) Cloning and expression in *Escherichia coli* of the *Klebsiella pneumoniae* genes for production, surface localization and secretion of the lipoprotein pullulanase. *EMBO J*, 6: 3531-3538.

Daura X, Antes I, van Gunsteren WF, Thiel W and Mark AE (1999) The effect of motional averaging on the calculation of NMR-derived structural properties. *Proteins*, 36: 542-555.

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys*, 103: 8577-8593.

Hess B, Bekker H, Berendsen HJC and Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*, 18: 1463-1472.

Hess B, Kutzner C, van der Spoel D and Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*, 4: 435-447.

Humphrey W, Dalke A and Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graphics*, 14: 33-&.

Karimova G, Pidoux J, Ullmann A and Ladant D (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci USA*, 95: 5752-5756.

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D and Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, 102: 3586-3616.

MacKerell AD, Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I and Pastor RW (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J Phys Chem B*, 114: 7830-7843.

Parrinello M and Rahman A (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J Appl Phys*, 52: 7182-7190.

Sauvonnet N, Gounon P and Pugsley AP (2000a) PpdD type IV pilin of *Escherichia coli* K-12 can be assembled into pili in *Pseudomonas aeruginosa*. *J Bacteriol*, 182: 848-854.

Sauvonnet N, Vignon G, Pugsley AP and Gounon P (2000b) Pilus formation and protein secretion by the same machinery in *Escherichia coli*. *EMBO J*, 19: 2221-2228.

Xicohtencatl-Cortes J, Monteiro-Neto V, Ledesma M, Jordan D, Francetic O, Kaper J, Pente J and Giron J (2007) Intestinal adherence associated with type IV pili of enterohemorrhagic *Escherichia coli* O157:H7. *J Clin Invest*, 117: 3519-3529.