The metal binding site

Even though metals were not added to the crystallization mixture, a strong positive peak (56 sigma) in the fo-fc electron density map clearly shows the tetrahedral coordination of a metal ion in the centre of the dimeric structure (Fig. S1).

The only metal that was used during purification was the affinity chromatography ligand cobalt. Refinement of the structure with a cobalt ion (full occupancy) yielded a significantly lower temperature factor (B-factor) for the cobalt than for the coordinating atoms, however. Consequently, the coordinated metal in the crystallized PulG dimer must be of a higher order than cobalt. The best fit to the calculated fo-fc and 2fo-fc maps was obtained with a zinc ion and revealed nearly the same B-factors for the zinc (20.9 $Å^2$) and the four coordinating oxygens (OE2 E44 chain A: 19.5 $Å^2$, OE2 E44 chain B: 20.4 $Å^2$, water oxygens: 19.7 $Å^2$ and 20.5 $Å^2$) after refinement. This metal binding site was also occupied in the selenomethionine derivative crystal. Refinement of this structure with a zinc ion with full occupancy again revealed the best result (assessed by B-factor comparison of the metal ion and the coordinating oxygens and visual inspection of fo-fc and 2fo-fc maps). A higher order metal with partial occupancy does not seem plausible, since a missing metal would result in electrostatic repulsion of the negatively charged carboxyl groups of glutamates 44 of both PulG monomers, which would abolish the dimerization.

Furthermore the peak values in an electron density map calculated from the anomalous differences in the Se high energy remote dataset combined with the calculated phases of the final model should be proportional to the f''-values of the respective anomalous scatterer. With the mean peak height at the selenium positions as a reference, the anomalous electron density map revealed an experimental value of 0.75 for the peak of the coordinated metal ion. This fits best to a theoretical ratio of 0.65 calculated from the f'' values of zinc and selenium at the corresponding wavelength. Other metals with a similar number of electrons that commonly occur in proteins (iron, cobalt, nickel, copper) have even lower ratio for the f''s (0.39-0.57). Consistent with these considerations and the B-value comparison, the distances in the metal donor complex (1.99 Å and 1.93 Å to the carboxylate oxygens; 1.96 Å and 1.99 Å to the water oxygens) are typical distances for a zinc coordination compared with values from other high resolution X-ray structures (Harding, M. M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Cryst.* **D57**: 401-411).

Structure based alignments of PulG and type IV pilins

Structure-based alignments of all type IV pilins and PulG (Fig. S2) revealed a high structural conservation of the long α-helix and the four β-strands, despite the low sequence similarity. Most of the conserved residues are involved in hydrophobic contacts in the core of the proteins. Interestingly, even though the fold of PulG is more similar to the type IVa pilins, it clearly shows more matched residues in the superposition with TcpA and twice as many identical residues (16) compared to the best fit between PulG and a type IVa pilin (PilE with 8 identical amino acids) (Table S1). Moreover, the calculated root mean square distances for the matched CA atoms in PulG and TcpA are lower compared than for all other PulG-pilin super-positions. This similarity between the core structure of PulG and TcpA is further emphasized by the fact that only three (instead of four for the type IVa pilins) of the β-strands could be aligned due to the different topology of the antiparallel β-sheet in the TcpA structure. A comparison of all combinations of pilin structures confirms that the structures of type IVa pilins are more similar to each other than to the pseudopilin PulG or to the type IVb pilin TcpA.

TcpA is processed by PulO prepilin peptidase

TcpA is a member of the type IVb group of pilins that have a longer *N*-terminal prepeptide than PilE, PilA and the major pseudopilins and in which the first amino acid of the processed protein is a methionine instead of the usual phenylalanine. To demonstrate preTcpA processing by the type II secreton prepilin peptidase PulO, pCHAP1418(*tcpA*$^{+}$) was introduced into *E. coli* K-12 strain PAP105 (F'lacI^{q1} Tn*10*) carrying either pUC18 (an empty vector) or pCHAP576 (pUC18 carrying *pulO* gene under *lacZ*p control). The bacteria were grown in LB broth containing ampicillin (100 µg/ml), chloramphenicol (25 µg/ml) and IPTG (1 mM) at 30°C, harvested when the OD_{600nm} reached 1.0 and resuspended in SDS sample buffer for analysis by SDS-PAGE and immunoblotting with TcpA-specific antiserum. The results (Fig. S3) show that TcpA in the strain with PulO migrates faster that TcpA in the strain without PulO, indicating prepilin processing.

Amino acid substitutions in the N-terminal region of PpdD

Despite the fact that PpdD itself is assembled into pili by the secreton, none of the G-D chimeras were assembled. The *N*-terminal hydrophobic region was therefore analysed in order to identify amino acids that might be responsible for this phenomenon. PulG and all assembled PulG derivatives have a specific pattern of amino acids with hydrophobic or hydrophilic characteristics / side chains at position +11, +14, +17 and +20 (with respect to the G-1/F+1 processing site in prePulG). These positions in PpdD are occupied by amino acids of opposite characteristics, suggesting that this might hamper the assembly of the G-D hybrids (Fig. S5). Therefore, these 4 amino acids in the *N*-terminal hydrophobic domain of PulG were converted into those found in PpdD (V+11G, G+14A, A+17S, V+20G) at the same positions and the modified PulG derived *N*-terminus was fused to position 21 of the *C*-terminus of PpdD. Even the $G_{21(modified)}$-D hybrid was still not assembled into pili (Fig. S6). This suggests that, besides the overall hydrophobic character of the *N*-terminal domains, specific amino acid contacts between the hydrophobic *N*-terminal region and rest of the pseudo- or type IV pilins are needed to achieve the correct interfacial interactions for pilus assembly.

Table S1 Similarity of atomic structures of type IV pilins and PulG. The structures were superimposed with the program LSQMAN as described in **Fig. S2**. Blue numbers (top right) show the rms. deviation for C_{α} atoms in a 5 Å cut-off. Red numbers (down left) correspond to the number of matched (and identical) residues for each protein pair. The number in boldface on the diagonal is the total number of amino acids being compared. The structures are ordered according to their similarity to PulG.

	PulG	**TcpA**	**PilA^{PAK}**	**PilE**	**PilA^{k122-4}**
PulG	**107**	2.04	2.12	2.33	2.63
TcpA	77(16)	**171**	2.47	2.51	3.88
PilA^{PAK}	77(4)	67 (7)	**120**	2.09	2.81
PilE	73(8)	70 (8)	93(20)	**158**	2.55
PilA^{k122-4}	69(4)	61(5)	86(17)	81(5)	**129**

Table S2. Oligonucleotide primers used to amplify *pilE* and *tcpA* genes

Oligonucleotide	**Gene/ position**	**Sequence**
2H7	*pilE*, 5’	5’ AGAGATGAATTCCCTTCAAAAAGGCTTTA 3’
2H8	*pilE*, 3’	5’ AGAGGCAAGCTTAAGGCCTAATTTGCCTCA 3’
3A3	*tcpA*, 5’	5’ AGAGATAAGCTTGCAATTATTAAAACAGCTTTTTAAGA 3’
3A4	*tcpA*, 3’	5’ AACGGCTCTAGAACTTATTTGAATTAGCTGTTACCA 3’

Restriction endonuclease sites are underlined

Table S3. Oligonucleotide primers used to construct pilin/pseudopilin chimeras in which the *N*-terminal domains were interchanged.

plasmid / chimera	primer name	oligonucleotide sequence
pCHAP7015 (D17-G(His)^{6})	1: PpdD5'	5'-CTA TTC GAA TTC AAA GTA GCG CCA ACC AAA TC-3'
	2: PpdD(NT3')	5'-GGG AAT ACC AAT GGC GCT GAG CAC GCC GAG GAT-3'
	3: PulG(CT5')	5'-ATC ATT GCC ATT TTA GCC AGC CTC GTG GTG CCC-3'
	4: PulG3b	5'-CAA GAG AAG CTT TAA AGC CGC GCT GTC GCA-3'
pCHAP7024 (D21- G(His)^{6})	1: PpdD5'	5'-CTA TTC GAA TTC AAA GTA GCG CCA ACC AAA TC-3'
	2: PpdD(NT3')GnewD	5'-GTT GCC CAT CAG GTT GGG AAT ACC AAT GGC GCT-3'
	3:PulG(CT5')PnewD	5'-AGC GCC ATT GGT ATT CCC AAC CTG ATG GGC AAC-3'
	4: PulG3b	5'-CAA GAG AAG CTT TAA AGC CGC GCT GTC GCA-3'
pCHAP7023 (G17-D)	1: PulG5'	5'-TTG ACG AAT TCA GCT CCA TAC CTT GAT GAG-3'
	2: PulG(NT3')	5'-GGG AAT ACC AAT GGC GCT GAG CAC GCC GAG GAT-3'
	3: PpdD5'(CT)	5'-ATC CTC GGC GTG CTC AGC GCC ATT GGT ATT CCC-3'
	4: PpdD4	5'-CTA TAG GGA TCC GCA GAC ACA TTT CAG TGA GC-3'
pCHAP7019 (G21-D)	1: PulG5'	5'-TTG ACG AAT TCA GCT CCA TAC CTT GAT GAG-3'
	2: PulG(NT3')prl	5'-GTA GTT TTG ATA AGC GGG CAC CAC GAG GCT GGC-3'
	3: PpdD(CT5')prl	5'-GCC AGC CTC GTG GTG CCC GCT TAT CAA AAC TAC-3'
	4: PpdD4	5'-CTA TAG GGA TCC GCA GAC ACA TTT CAG TGA GC-3'
pCHAP7028 (G21$_{modified}$-D)	1: PulG5'	5'-TTG ACG AAT TCA GCT CCA TAC CTT GAT GAG-3'
	2: PulG(NT3')N-modi	5'-CAG GTA GTT TTG ATA AGC GGG CAC ACC GAG GCT GCT GAG CAC GGC GAG GAT GCC AAT CAC-3'
	3: PpdD(CT5')N-modi	5'-AGC CTC GGT GTG CCC GCT TAT CAA AAC TAC CTG-3'
	4: PpdD4	5'-CTA TAG GGA TCC GCA GAC ACA TTT CAG TGA GC-3'
pCHAP7016 (E17-G(His)^{6stran})	1: PilE5'	5'-AGA GAT GAA TTC CCT TCA AAA AGG CTT TA-3'
	2: PilE(NT3')	5'-GGG CAC CAC GAG GCT GGC CAA AAT GCC GAC GAT-3'
	3: PulG(CT5')E	5'-ATC GTC GGC ATT TTG GCC AGC CTC GTG GTG CCC-3'
	4: PulG3b	5'-CAA GAG AAG CTT TAA AGC CGC GCT GTC GCA-3'
pCHAP7017 (G17-E)	1: PulG5'	5'-TTG ACG AAT TCA GCT CCA TAC CTT GAT GAG-3'
	2: PulG(NT3')E	5'-GGG AAG GGC GAC TGC CGC GAG CAC GCC GAG GAT-3'
	3: PilE(CT5')	5'-ATC CTC GGC GTG CTC GCG GCA GTC GCC CTT CCC-3'
	4: PilE3'	5'-AGA GGC AAG CTT AAG GCC TAA TTT GCC TCA-3'
pCHAP7026 (G17-A^{PAK})	1: PulG5'	5'-TTG ACG AAT TCA GCT CCA TAC CTT GAT GAG-3'
	2: PulG17(NT3')A	5'-AAT GGC AAT TGC AGC GAG CAC GCC GAG GAT AAC-3'
	3: PilA(CT5')G17	5'-ATC CTC GGC GTG CTC GCT GCA ATT GCC ATT CCT-3'
	4: PilA3'(BH1)	5'-CGA CTG AGC GGA TCC GGC GGC GGC-3'
pCHAP7027 (G21-A^{PAK})	1: PulG5'	5'-TTG ACG AAT TCA GCT CCA TAC CTT GAT GAG-3'
	2: PulG21(NT3')A	5'-ATT CTG ATA CTG AGG CAC CAC GAG GCT GGC GAG-3'
	3: PilA(CT5')G21	5'-GGC AGC CTC GTG GTG CCT CAG TAT CAG AAT TAT-3'
	4: PilA3'(BH1)	5'-CGA CTG AGC GGA TCC GGC GGC GGC-3'

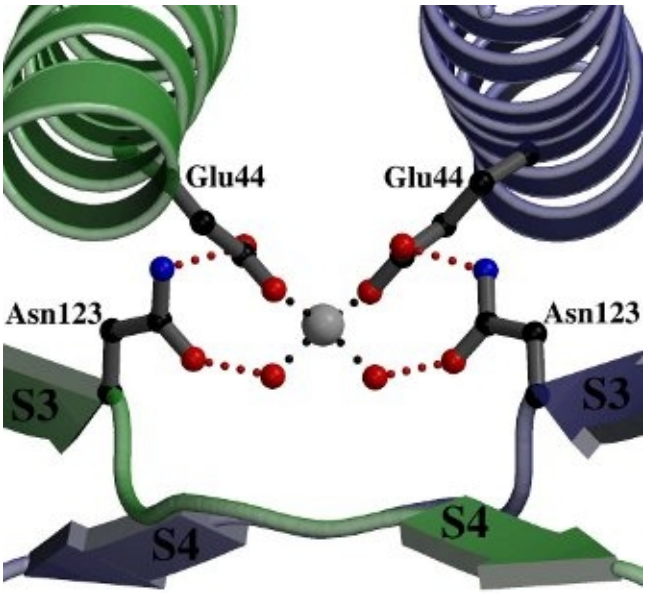

Figure S1: The metal binding site in the dimeric crystal form of $PulG_{25\text{-}134}(His)^6$. The metal ion is tetrahedrally coordinated between E44 residues in the middle of the N-terminal a-helices of each PulG monomer and by two water molecules. N123 at the end of S3 stabilizes the complex via H-bonds to the water molecules and the non-coordinating ε-oxygens of the E44 residues. Dashed black lines indicate coordinating bonds and dashed red lines the H-bonds.

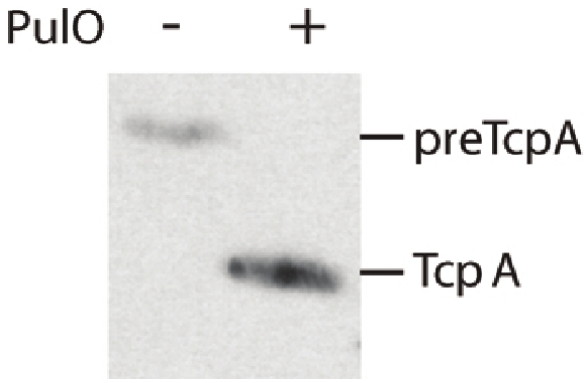

Figure S2: Processing of preTcpA in *E. coli* strain carrying the prepilin peptidase structural gene *pulO*. proteins were separated by SDS-PAGE and electroblotted onto a nitrocellulose membrane that was probed with antibodies against TcpA (a gift from R. Taylor).

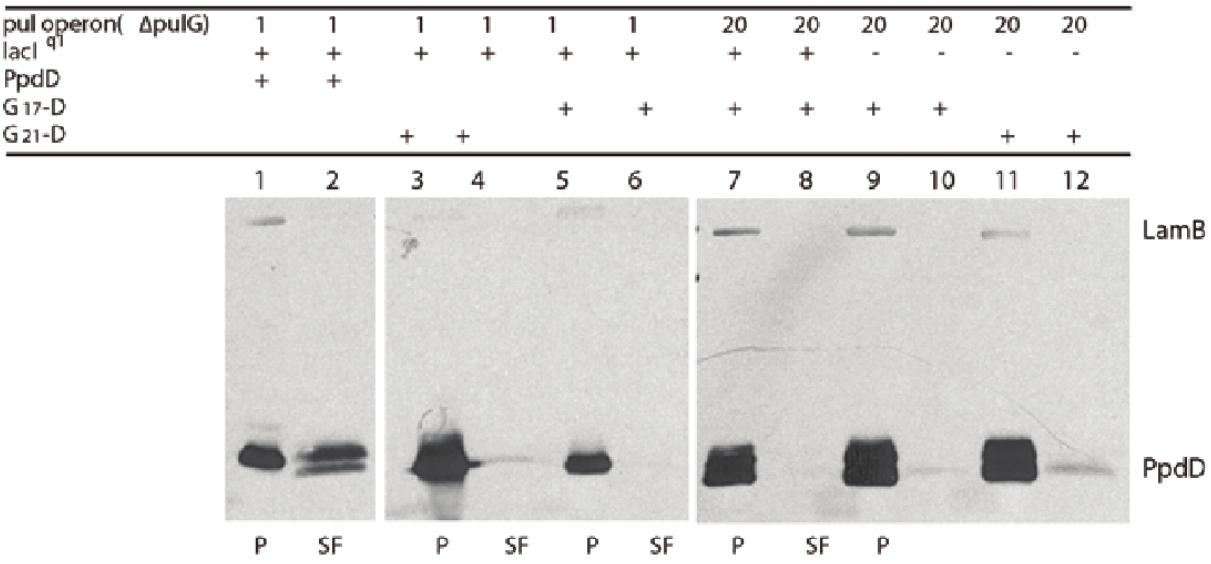

Figure S3: Analysis of pellet (P) and sheared fractions (SF) of *E. coli* stains PAP7000BG, PAP7501 (pCHAP1216) and PAP9001 (pCHAP1216) carrying pCHAP3100 (*ppdD*), pCHAP7023 and pCHAP7019 (encoding G_{17}-D and G_{21}-D, respectively) by immunoblotting. The relative amounts of secreton components present in the cells are indicated by numbers (chromosomal expression level = 1, expression from plasmid pBR322 = 20). Whether genes under *lac*Zp control are repressed or not is determined by the presence or absence, respectively, of the lac^{q1} repressor gene (+ or -). Membranes were probed with anti-PpdD and anti-LamB antibodies. P; pellet fraction after shearing; SF, sheared fraction.

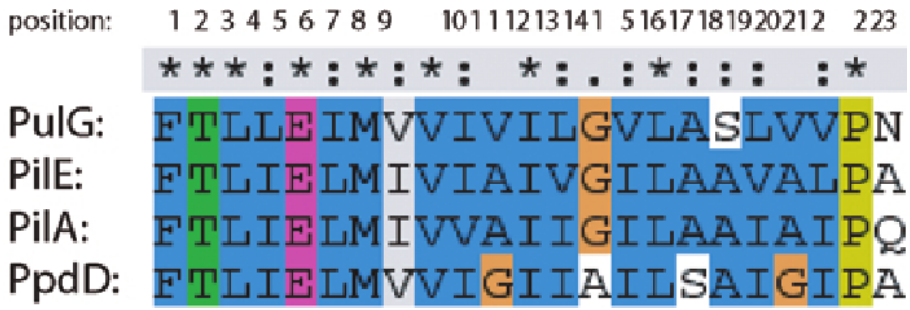

Figure S4: Multiple sequence alignment of the first 23 N-terminal amino acids of PulG, PilE, $PilA^{PAK}$ and PpdD. Positions 11, 14, 17 and 20 of the PulG, PilE and $PilA^{PAK}$ sequences are occupied by amino acids of equal characteristics (+), whereas the same positions in PpdD are occupied by amino acids of opposite characteristics (-). Sequence alignment was done using the MultiAlign software (http://prodes.toulouse.inra.fr/multalin/multalin.html).

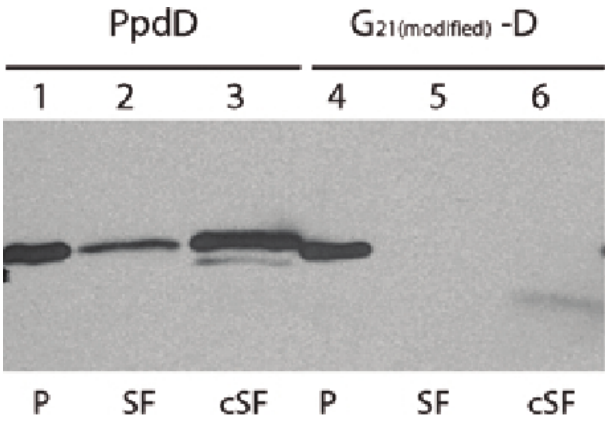

Figure S5: Western blot analysis of bacterial pellet (p), sheared fractions (SF) and concentrated sheared fractions (cSF) from *E. coli* K-12 strain PAP7500BG(Δ*pulG*) harbouring pCHAP3100 (encoding PpdD) (lanes 1-3) and pCHAP7028 (encoding $G21_{(modified)}$-D) (lanes 4-6), respectively. The blotted membrane was probed with antisera against PpdD. P; pellet fraction after shearing; SF, sheared fraction; cSF, concentrated sheared fraction.

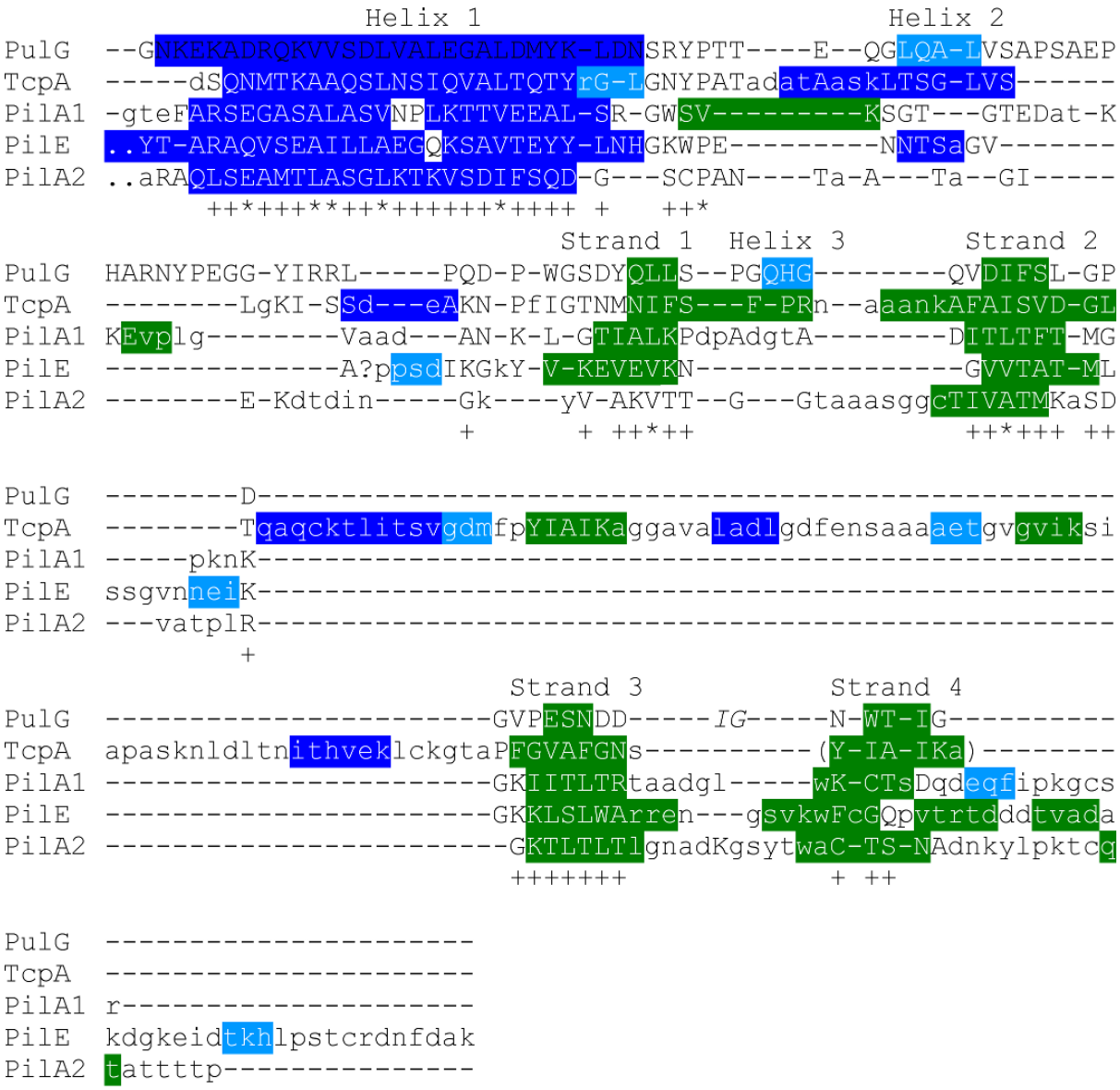

Figure S6: A structural alignment of the C-terminal globular domain of the pseudopilin PulG with known type IV pilin structures (the type IVb pilin TcpA from *V. cholerae* and the type IVa pilins PilA(PilA1) from *P. aeruginosa* strain PAK, PilE from *N. gonorrhoeae* strain MS11 and PilA (PilA2) from *P. aeruginosa* strain k122-4. The superposition was performed by the brute force fit option in LSQMAN (Kleywegt, G.J. and Jones, T.A. 1995. A super position. *CCP4/ESF-EACBM Newsletter on Protein Crystallography* 31:9-14) using a cut-off distance of 5 ?. The transfer matrix was calculated for the first 100 amino acids of the PulG model since amino acids N128-G132 forming S4 are exchanged between both monomers in the asymmetric unit. These residues were taken from the other monomer. I126 and G127 (in italics) were excluded from the alignment since they form a putative flexible hinge region. Matched residues within the cut-off distance are printed in capital letters. Secondary structure elements calculated with DSSP (Kabsch, W and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen bond and geometrical features. *Biopolymers*, **22**: 2577-2637.) are highlighted with colours: blue, α-helix; light blue, 3_{10}-helix; green, β-strand. Crosses indicate positions that are occupied by structurally related amino acids in all 5 structures and stars that these residues are similar.