

Supporting Information for
Archaic and alternative chaperones preserve pilin folding energy by providing
incomplete structural information

Natalia Pakharukova, Sophie McKenna, Minna Tuittila, Sari Paavilainen, Henri Malmi, Yingqi Xu,
Olena Parilova, Steve Matthews, and Anton V. Zavalov

Supplementary Tables S1-S3

Supplementary Figures S1-S10

Supplementary references

Table S1. Diffraction data and refinement statistics

Crystal parameters	CsuC-CsuA/B	CsuC-CsuA/B	CsuA/Bsc	
			Native dataset	Se-Met SAD dataset
Space group	<i>P</i> 6 ₅ 22	<i>H</i> 3	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2
Cell dimensions, (Å, °)	<i>a</i> =94.379, <i>b</i> =94.379, <i>c</i> =390.745, α =90.00, β =90.00, γ =120.00	<i>a</i> =111.419, <i>b</i> =111.419, <i>c</i> =167.745, α =90.00, β =90.00, γ =120.00	<i>a</i> =49.21, <i>b</i> =92.02, <i>c</i> =34.11, α =90.00, β =90.00, γ =90.00	<i>a</i> =48.913, <i>b</i> =92.486, <i>c</i> =34.113, α =90.00, β =90.00, γ =90.00
Number of molecules per asymmetric unit	2	2	1	1
Data collection				
Beamline	ID23-1	ID23-1	ID23-1	ID29
Wavelength (Å)	0.8731	0.968	0.8725	0.979
Resolution (Å)	47.19-2.85 (3.00-2.85)	55.71-2.50 (2.64-2.50)	49.21-1.47 (1.52-1.47)	49.31-1.60 (1.66-1.60)
Unique observations	24880 (3546)	24003 (3555)	27069 (2591)	19295 (1889)
<i>R</i> _{merge}	0.157 (0.866)	0.041 (0.245)	0.055 (0.637)	0.051 (0.689)
<i>R</i> _{sym}	0.157	0.041	0.054	0.058
$\langle I \rangle / \sigma I$	11.3 (2.0)	20.2 (5.0)	12.1 (1.8)	12.7 (1.8)
Completeness (%)	99.2 (99.5)	89.3 (90.2)	99.7 (99.5)	92.4 (93.4)
Redundancy	7.9 (6.4)	3.3 (3.1)	4.3 (4.09)	3.8 (3.78)
Overall <i>B</i> factor from Wilson plot (Å)	47.5	39.5	14.67	16.64
Refinement				
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.97/25.68	17.17/22.64	19.67/21.91	
Number of protein residues	718	668	157	
Number of ligands/ions	-	-	-	
rmsd stereochemistry				
Bond lengths (Å)	0.003	0.008	0.011	
Bond angles (°)	0.661	1.161	1.370	
Ramachandran analysis*				
Residues in outlier regions (%)	1.59	0.47	0	
Residues in favored regions (%)	94.93	96.06	97.39	
Residues in allowed regions (%)	98.41	99.53	100	
PDB code	6FQA	6FQ0	6FM5	

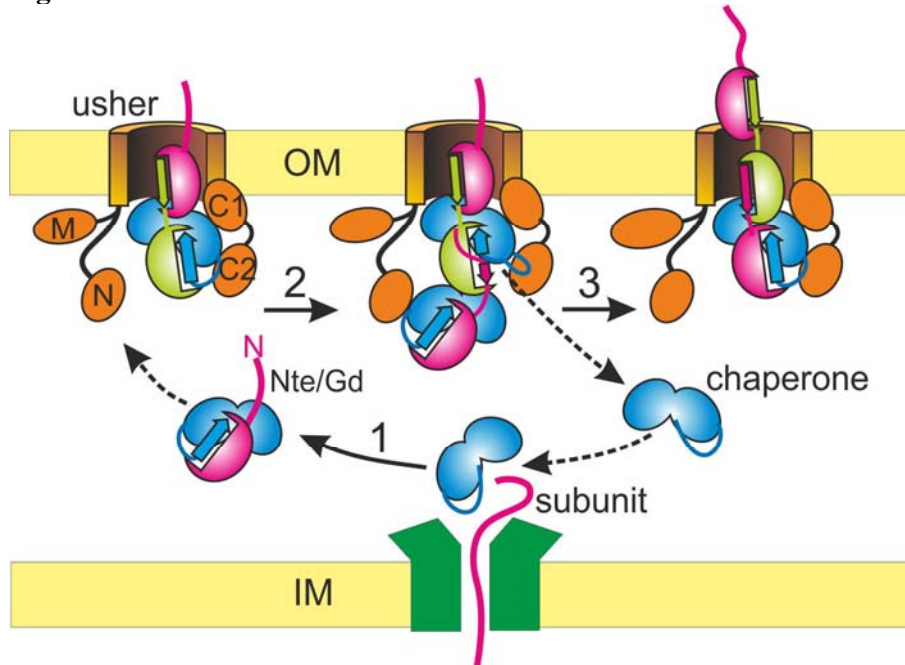
Table S2. Structural homologues of CsuA/B

Structure	PDB code	Number of Cα atoms	RMSD	Z-score	% of identity
CfaE, pilin domain, CFA/I fimbriae	2HB0	133	2.7	11.7	8
LpfD, adhesin domain, LPF	5AFO	130	3.1	10.6	19
CssB, CS6 colonization factor	4B9G	120	2.7	10.0	13
PsaA, pH6 antigen	4F8N	120	2.6	10.0	10
FimG, type 1 pili	5IQM	115	2.4	9.7	16
CsuA/B in complex with CsuC chaperone	5D6H	93	3.3	9.5	84
CooA, CS1 colonization factor	4HJI	127	3.0	9.5	10
FimA, type 1 pili	2JTY	127	2.9	9.3	15
CfaB, CFA/I fimbriae	3F83	126	3.2	9.2	10
-----	-----	-----	-----	-----	-----
EcpA, Ecp	3QS3	125	3.0	8.8	13
FimH, lectin domain, type 1 pili	4ATT	126	3.0	7.4	9.0
Caf1, F1-antigen	1Z9S	103	2.6	6.6	14.0

Table S3. Oligonucleotides

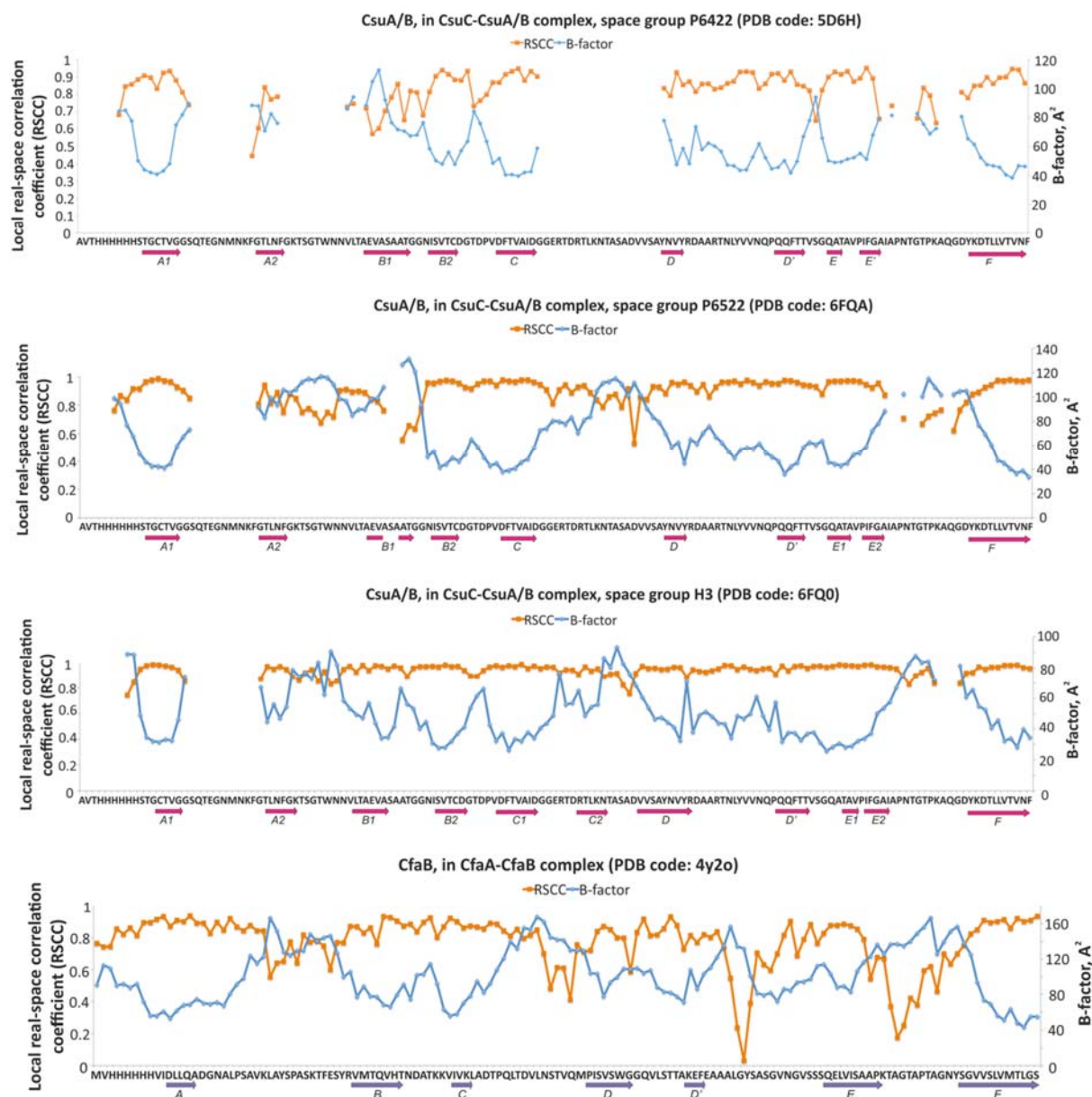
Name	Sequence (5'→3')
CsuAB_F-GR	CCT TGT TCA TAT TTC CTT CAG TTT GAC
CsuAB_F-GF	GTG GTA CTT TAA ATT TTG GTA AAA CTT CC
CsuAB_V-GR	CTT CAG CTG TTA ATA CGT TGT TCC AAG
CsuAB_V-GF	GTG CTT CAG CAG CAA CAG GTG
CsuC_LN-R	TTA AAG AAG CAT CTT GCT CGT TGC C
CsuC_IN-R	TTA TAG AAG CAT CTT GCT CGT TGC C
CsuC_NI-F	ATA TAA AAG TAA GTT TCC AAA TGC GTT ACT C
CsuC_NL-F	ATT TAA AAG TAA GTT TCC AAA TGC GTT ACT C
CsuAB_F-GR	CCT TGT TCA TAT TTC CTT CAG TTT GAC
CsuAB_F-GF	GTG GTA CTT TAA ATT TTG GTA AAA CTT CC
CsuAB_V-GR	CTT CAG CTG TTA ATA CGT TGT TCC AAG
CsuAB_V-GF	GTG CTT CAG CAG CAA CAG GTG

Figure S1



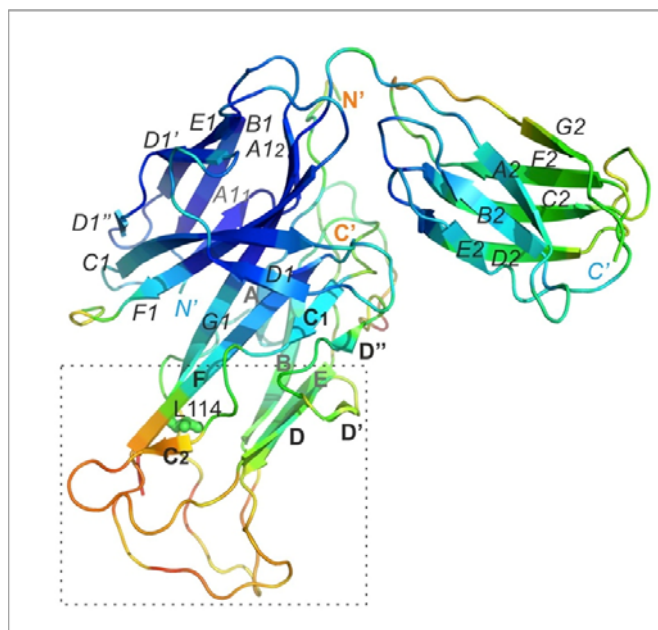
Schematic illustration of pilus assembly via the classical chaperone-usher pathway. Step 1: Periplasmic chaperones (blue) bind to pilus subunits (red or green) to form pre-assembly complexes. **Step 2:** The pre-assembly complex binds to the N-terminal domain of the usher (N) and the N-terminal extension (Nte or donor strand Gd) of the subunit in the pre-assembly complex replaces the donor strand G₁ of the chaperone in the acceptor cleft of the subunit at the base of the fiber. Strand G₁ and Gd runs in opposite directions in the cleft and the donor strand exchange (DSE) occurs gradually in a zip-out-zip-in fashion. Strand G₁ is shown wider than strand Gd to reflect the fact that it has larger side chains of its donor residues. **Step 3:** The complex becomes incorporated into the fiber, the former fiber-capping chaperone is released, the fiber dissociates from the usher N-terminal domain, binds at usher C-terminal domains (C1 and C2) and translocates to the cell surface. Some details of the assembly mechanism were omitted for clarity: **(a)** in addition to donor strand G₁, β strand A₁ of the chaperone plays role in the subunit binding (1); **(b)** the chaperone-subunit binding causes a conformational change in the chaperone (“proline lock” opening) that enables its binding to the N-terminal domain of the usher (2); **(c)** DSE is initiated in pocket P5 in the subunit acceptor cleft, which is either fully or partially/transiently exposed (3,4); **(d)** upon DSE, subunits collapse into a more compact structure due to a movement of two β -sheets towards each other, releasing energy that drives the assembly process (1,5).

Figure S2



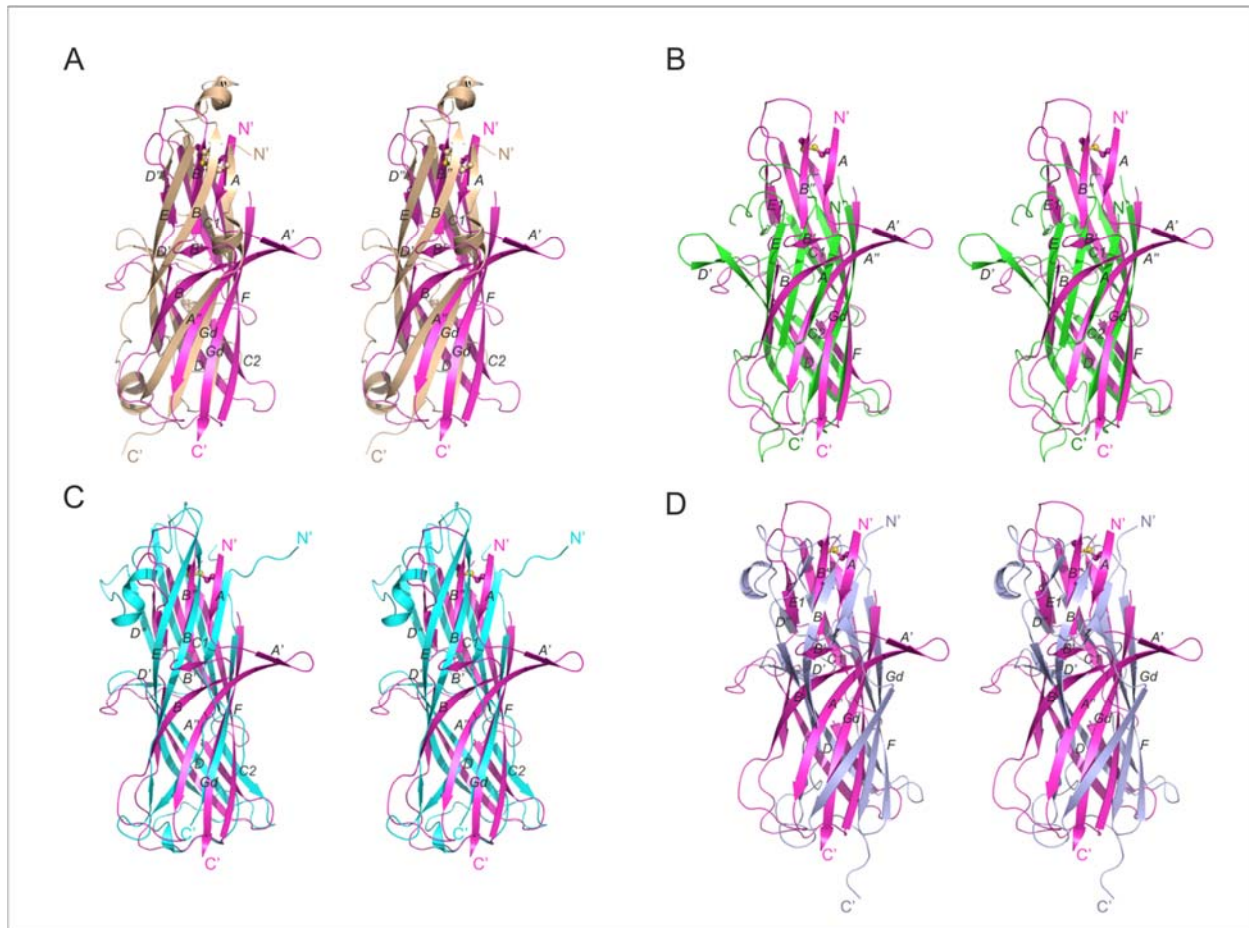
B-factors and local real-space correlation coefficients between the models and the electron density (RSCC) as a function of residue number. The protein sequence is depicted along abscissa with secondary structure elements shown below the sequence. B-factors and RSCC are shown in blue and orange, respectively.

Figure S3



Cartoon diagram of the crystal structure of the CfaA-CfaB chaperone-subunit complex (PDB code 4Y2O). The structure is colored by B-factor of Ca atoms with the color ranging from blue (lowest) to red (highest). High-B-factor region in CfaB is framed in a rectangle. Leucine 114 that was introduced instead of native threonine to stabilize the complex is shown as a ball-and-stick diagram. β -Strands are labeled.

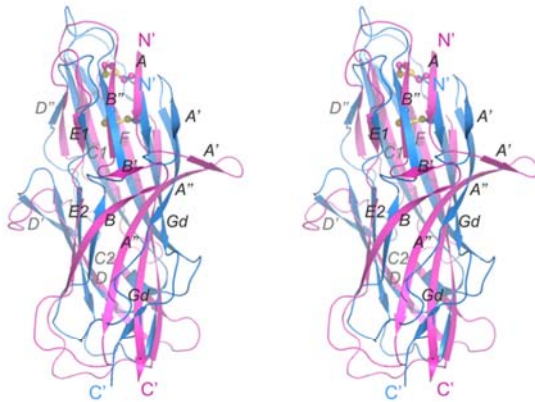
Figure S4



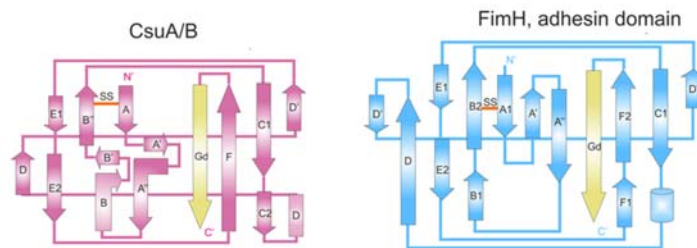
Superposition of the archaic *Csua/Bsc* (magenta) and representative pilin subunits from the classical FGS (PapA, PDB code 2UY6, wheat) (A) and FGL (Caf1, PDB code 1Z9S green) (B) and alternative Ecp-like (EcpA, PDB code 3QS3, cyan) (C) and Cfa-like (CfaB, PDB code 3F83, lilac) (D) CUPs (cartoon representations, stereo views). β -strands, N and C termini are labeled.

Figure S5

A

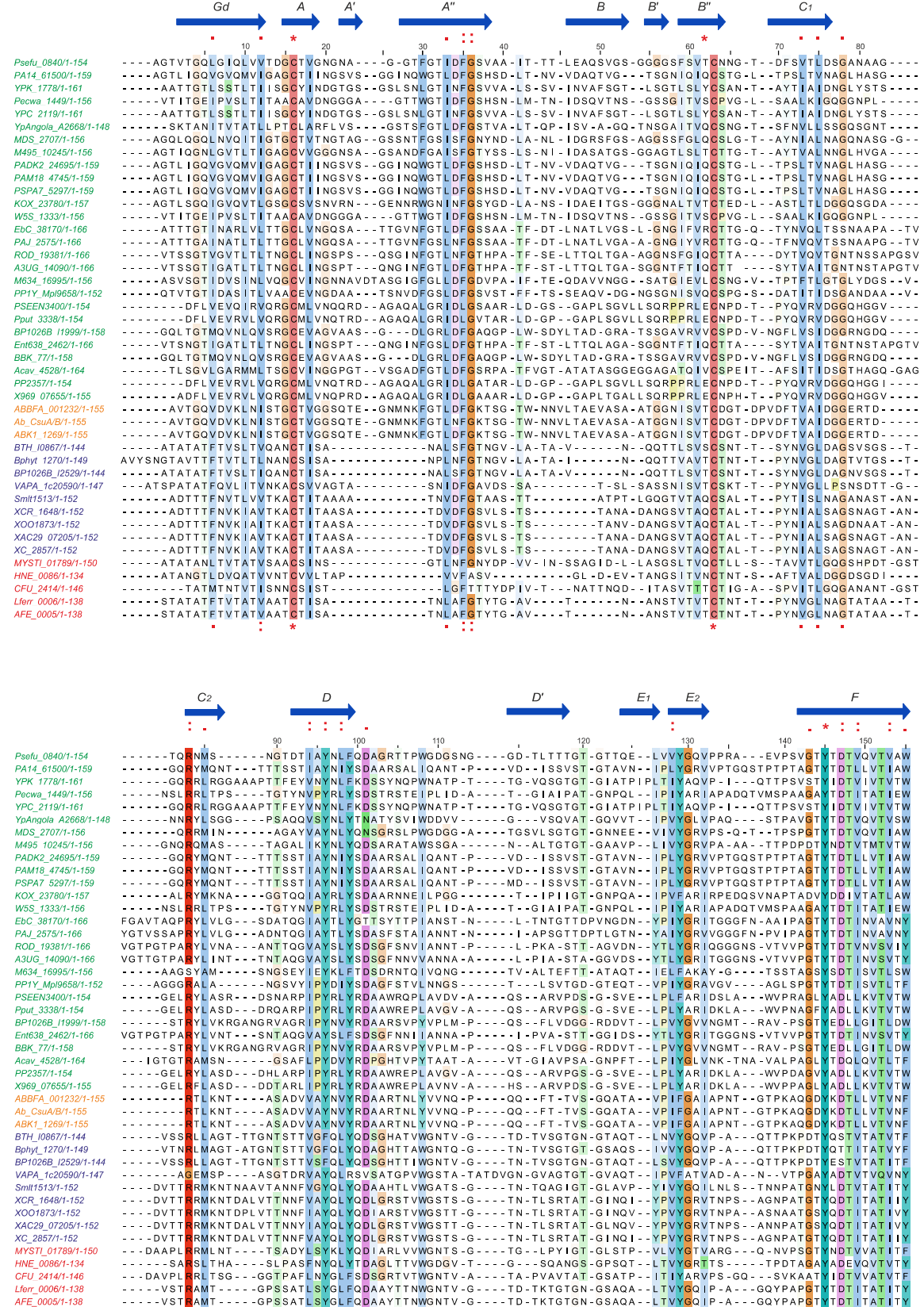


B



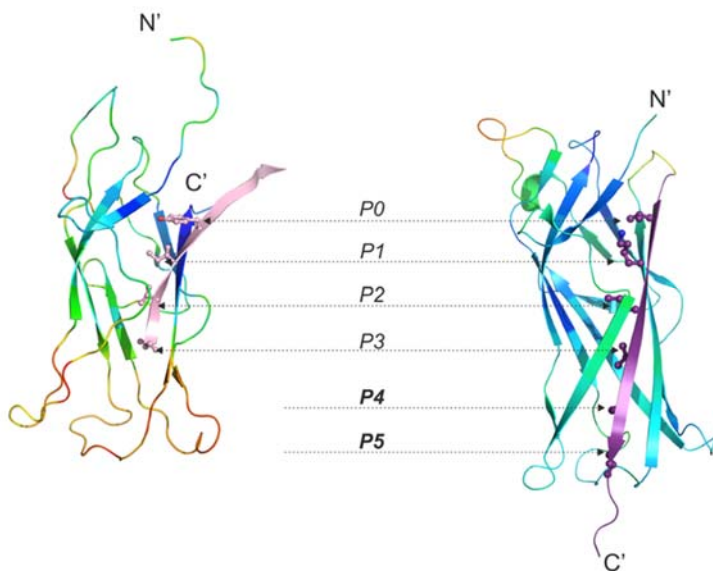
Comparison of Csua/Bsc with the lectin domain of FimH adhesin subunit. (A) Structural superposition of Csua/Bsc (magenta) and the lectin domain of the FimH adhesin subunit (blue, PDB code 4ATT) from the classical type 1 fimbriae (cartoon representation, stereo view). Note the presence of the A-A' hairpin in the structure of the lectin domain of FimH. β -strands and N and C termini are labeled. (B) Topology diagrams of Csua/Bsc and FimH. Donor strand Gd in Csua/B and strand G in FimH are shown in yellow.

Figure S6



Alignment of sequences of single-domain subunits from the archaic pathway. Single-domain major subunit dataset derived from σ_1 , σ_2 , $\sigma_{2/3}$ and σ_3 family clusters is indicated in red, violet, orange and green, respectively. The sequences of single-domain subunits were analyzed with SignalP 4.1 to exclude secretion signal peptides. Structural-based sequence alignment was produced with the program *EXPRESSO* [<http://tcoffee.org.cat/apps/tcoffee/do:expresso>]. The CsuA/Bsc structure was assigned as a template for the sequence dataset. The aligned sequences were annotated using the package JalView (6). Completely invariant residues are indicated with an asterisk, highly conserved regions are marked with a colon and residues essential for semi-conservation are dotted. Color-coding scheme of amino acid residues has been incorporated above conservation threshold according to their chemical properties: blue — hydrophobic, green — hydrophilic, red — basic and magenta — acidic. Conserved cysteine residues are colored in pink. Glycine residues are shaded in orange, proline in yellow, and aromatic residues in cyan. The secondary structure of CsuA/B is shown above the amino acid sequences. β -Strands and α -helices are depicted by arrows and bars, respectively, above the alignment table.

Figure S7



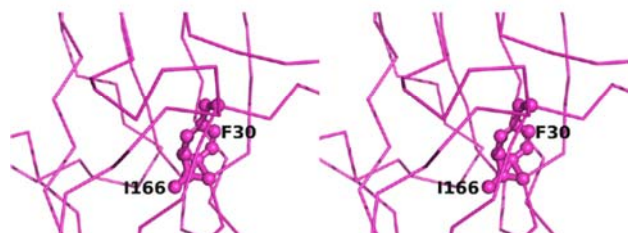
CfaA chaperone from the alternative pathway provides only partial information of the CfaB subunit folding. Cartoon representation of the chaperone-bound CfaB (left, PDB code 4Y2O) and CfaBsc (right, PDB code 3F83). Subunits are colored by B-factors. The chaperone G_1 and subunit G_d donor strands are colored in magenta and orange, respectively. The donor residues are shown as balls-and-sticks. The acceptor pockets P0-P5 are indicated and β -strands and N and C termini are labeled.

Figure S8

A



B

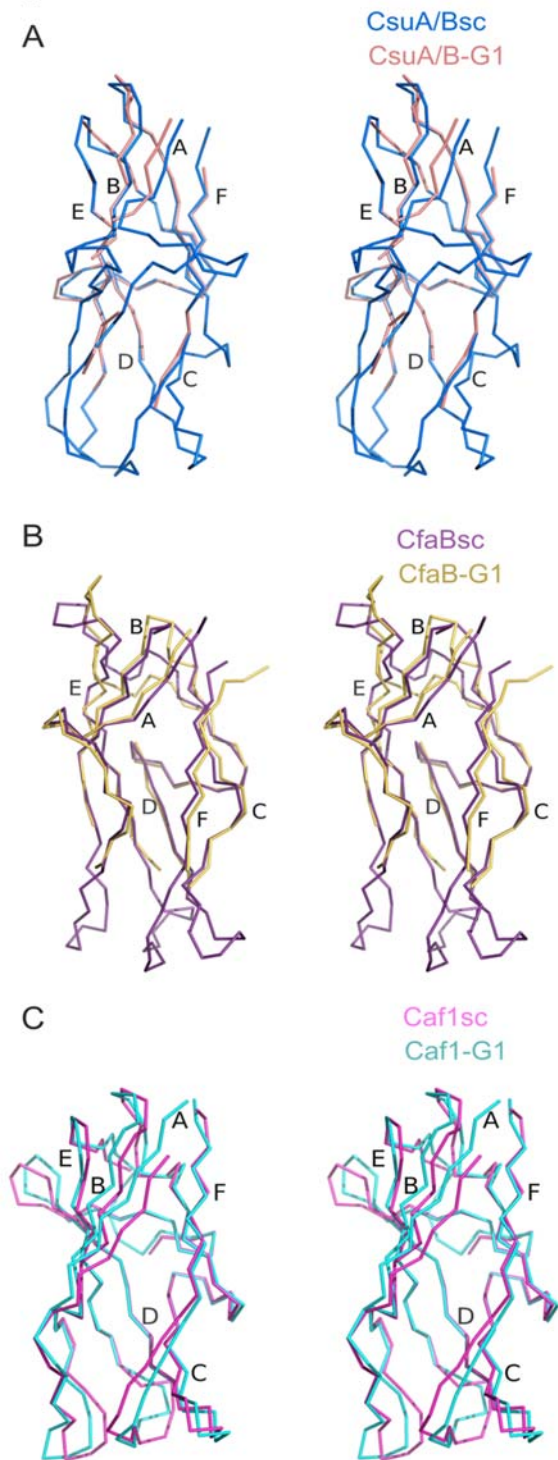


C



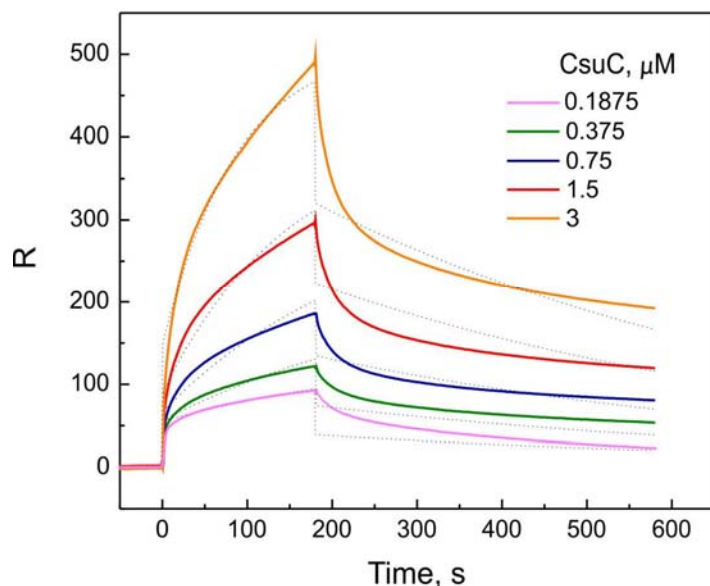
If the double-hairpin were ordered in the CsuC-CsuA/B complex, side chain of Phe30 would form steric clashes with residues of either CsuA/B or CsuC in any of its possible rotamer conformation. Fragments of the CsuA/Bsc structure (A and B) and the CsuA/Bsc and CsuC-CsuA/B superposition as in Fig. 4C, but with only CsuA/Bsc and donor strand G₁ shown for clarity (C) (ribbon representation, stereo view). Phe30 in three possible rotamer conformations: t30 with 33% probability (A), p90 rotamer with 13% probability (B), and m30 rotamer with 9% probability (C) and residues with which it collides (labeled) are shown as balls and sticks.

Figure S9



Superpositions of the chaperone-bound and self-complemented pilin subunits (ribbon diagrams, stereo views). **A.** Archaic system. The chaperone-bound and self-complemented CsuA/B are shown in pink and blue, respectively. **B.** Alternative system. The chaperone-bound and self-complemented CfaB are colored yellow and purple, respectively. **C.** Classical system. The chaperone-bound and self-complemented Caf1 are colored cyan and magenta, respectively.

Figure S10



Association and dissociation kinetics of CsuC and Csua/B. A Biacore X100 system (Biacore, GE Healthcare, Uppsala, Sweden) was used for biosensor experiments. The purified CsuC:Csua/B complex carrying a His-tag on Csua/B ($0.075\mu\text{M}$) was loaded on an NTA sensor chip. CsuC was removed from the chip by continuous washing of the chip with loading buffer. To minimize the time for Csua/B aggregation on the chip, we immediately performed binding experiments. Resonance signals were plotted as a function of time for several concentrations ($0.138 - 3\mu\text{M}$) of CsuC. The association phase was followed through 180 s after which time infusion of the soluble CsuC was stopped, and the dissociation phase in continuous buffer flow was monitored for an additional 400 s. The chip was regenerated at each cycle of the experiment. The software supplied with the device was used to determine the k_{on} and k_{off} rate constants assuming a 1:1 binding model. The model accounted for the bulk effect observed in the data.

References

1. Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S., and Knight, S. D. (2003) Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* **113**, 587-596
2. Yu, X., Dubnovitsky, A., Pudney, A. F., Macintyre, S., Knight, S. D., and Zavialov, A. V. (2012) Allosteric Mechanism Controls Traffic in the Chaperone/Usher Pathway. *Structure* **20**, 1861-1871
3. Yu, X. D., Fooks, L. J., Moslehi-Mohebi, E., Tischenko, V. M., Askarieh, G., Knight, S. D., Macintyre, S., and Zavialov, A. V. (2012) Large is fast, small is tight: determinants of speed and affinity in subunit capture by a periplasmic chaperone. *J Mol Biol* **417**, 294-308
4. Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E., and Waksman, G. (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell* **22**, 831-842
5. Zavialov, A. V., Tischenko, V. M., Fooks, L. J., Brandsdal, B. O., Aqvist, J., Zav'yalov, V. P., Macintyre, S., and Knight, S. D. (2005) Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem. J.* **389**, 685-694
6. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191