

Structure, Volume 23

Supplemental Information

**FlaF Is a β -Sandwich Protein
that Anchors the Archaellum in the Archaeal
Cell Envelope by Binding the S-Layer Protein**

Ankan Banerjee, Chi-Lin Tsai, Paushali Chaudhury, Patrick Tripp, Andrew S. Arvai, Justin P. Ishida, John A. Tainer, and Sonja-Verena Albers

SUPPLEMENTAL INFORMATION

β -sandwich protein FlaF anchors the archaellum in the archaeal cell envelope by binding to the S-layer protein

Ankan Banerjee^{1,4,£}, Chi-Lin Tsai^{2,£}, Paushali Chaudhury^{1,\$}, Patrick Tripp^{1,\$}, Andrew S. Arvai³, Justin P. Ishida², John A. Tainer^{2,3,#} and Sonja-Verena Albers^{1,\$,#}

¹Molecular Biology of Archaea, Max Planck Institute for terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany

²Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

³Department of Integrative Structural and Computational Biology, Skaggs Institute of Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

⁴Current address: AK Essen- FB-Chemie, Phillips Universität Marburg, Hans Meerwein Straße, 35043 Marburg.

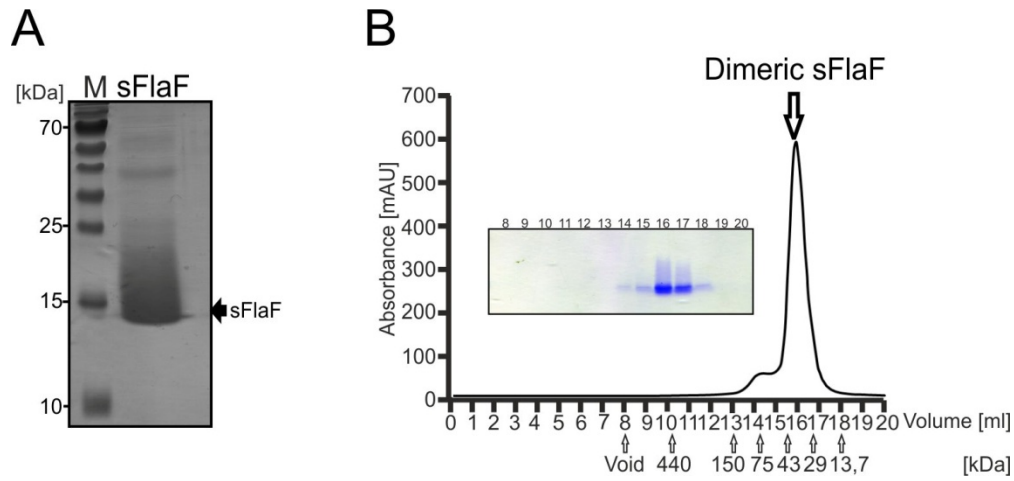
\$ Molecular Biology of Archaea, Institute of Biology, University of Freiburg, Schauenzlestrasse 1, 79211 Freiburg

Correspondence to: Sonja-Verena Albers, sonja.albers@biologie.uni-freiburg.de, tel: +49 761 2032630 or jatainer@lbl.gov

£ These authors contributed equally.

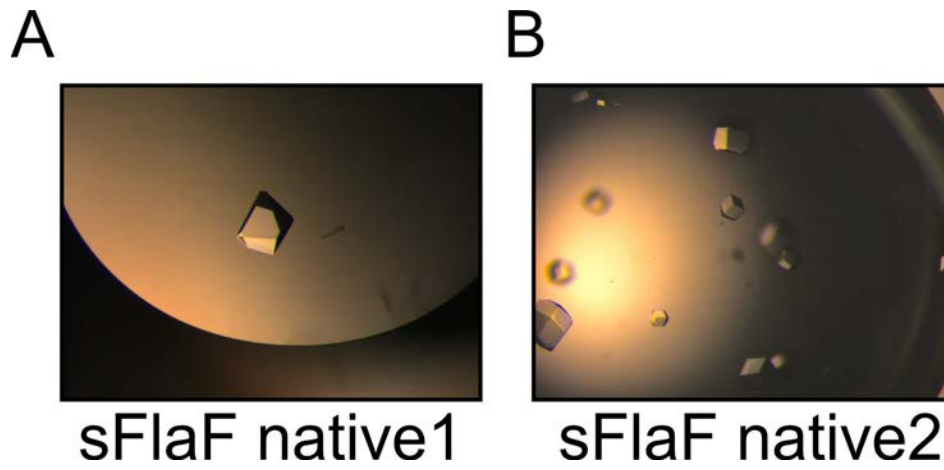
SUPPLEMENTAL FIGURES AND LEGENDS

Figure S1



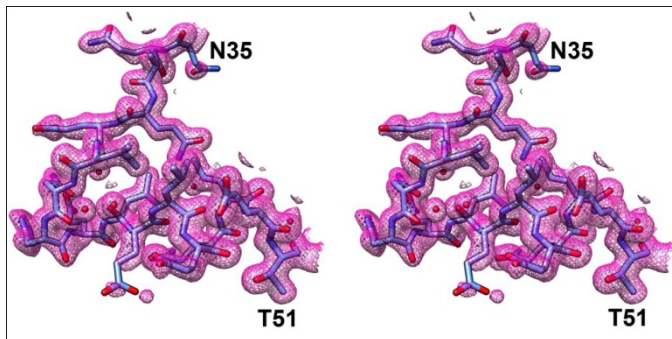
Related to Figure 3. Purification of sFlaF derivative. (A) Coomassie stained SDS-PAGE showing pure sFlaF, purified using Ni-NTA affinity chromatography. (B) Size exclusion chromatography of sFlaF showing a pure monodispersed fraction eluted as dimeric species, indicated with an arrow.

Figure S2



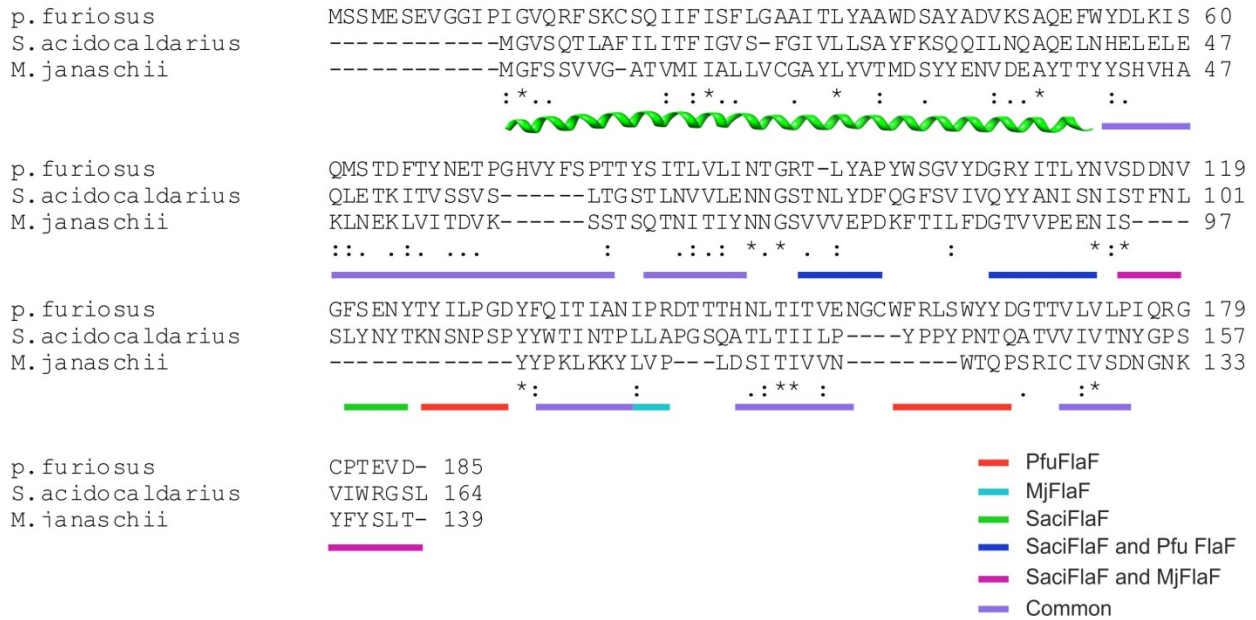
Related to Figure 2. Native sFlaF crystals. (A) The crystal grew in 3 M potassium chloride and 50 mM HEPES, pH7.0. (B) The crystals grew in 35% Tacsimate, pH 7.0.

Figure S3



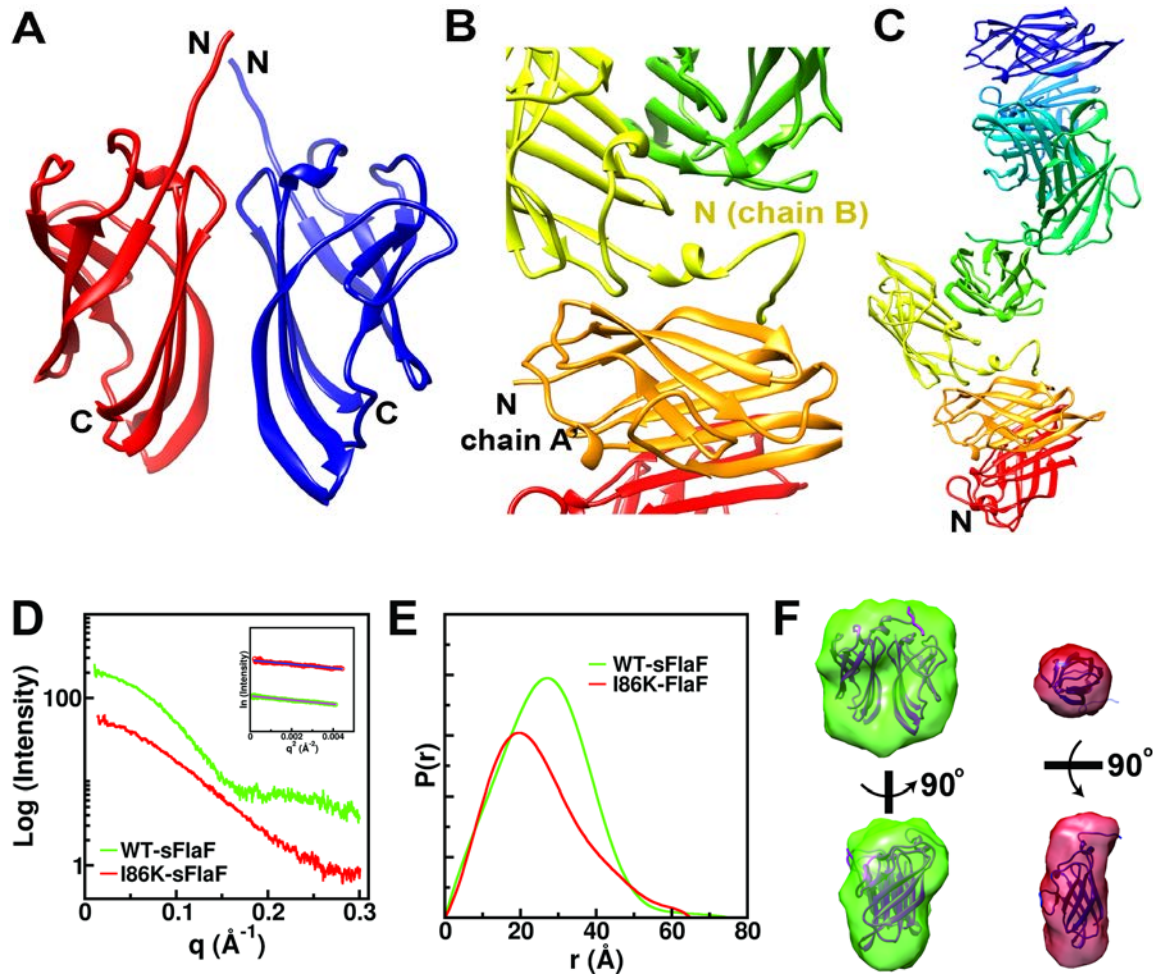
Related to Figure 2. The 2Fo-Fc electron density map of N-terminal region of sFlaF. The N-terminal residues 35-51 of sFlaF are shown in stick. The 2Fo-Fc electron density map is shown in magenta with 1σ contour level. The electron density is clear for the entire region.

Figure S4



Related to Figure 1. Predicted secondary structure and sequence alignment of FlaFs. ClustalW alignment of *Pyrococcus furiosus* (Pfu), *Sulfolobus acidocaldarius* (Saci) and *Methanocaldococcus janaschii* (Mj) FlaF showing less than 20% sequence similarities. The secondary structure prediction of these FlaF's using Psipred indicating the structural conservation. The N terminal extended α -helix is shown in green color and the β -sheets are indicated as colored lines. β -sheets colors are indicated as follows Pfu specific are red, cyan is Mj specific, green is Saci specific. Blue sheets are present in Saci and Pfu FlaF, pink are Saci and Mj specific. Violet β -sheets are common in all FlaFs.

Figure S5



Related to Figure 3. (A) Disordered N-terminal helix from native 1 sFlaF crystal structure. (PDB: 4LIO). The residues 35-45 are disordered and missing in the electron density map. (B) From native 2 sFlaF structure, N-terminal helix (residues 35-51) from chain B (yellow) is stabilized by the β -sheet of the neighboring subunit chain A' (orange) through electrostatic and hydrophobic interactions in the crystal lattice. (C) The possible filament formation of sFlaF protein from crystal lattice. Each dimer is stabilized by hydrophobic interactions (ex. red and orange subunits), The interactions between dimers are stabilized by N-terminal helix as described in (B). The length of four-dimer filament is about 15.5 nm, which can reach the S-layer from cytoplasmic membrane. The truncated transmembrane part is not included in this figure. (D) SAXS scattering profile of wt-sFlaF (green) and I86K-sFlaF (red). The inset

represents the Guinier region of both sFlaF proteins. The radius of gyration R_g is 21 Å and 19 Å for wt-sFlaF and I86K-sFlaF, respectively. (E) The pair distribution plot of wt-sFlaF and I86K-sFlaF. The maximum dimension D_{\max} of wt- and I86K-sFlaF are 75 Å and 65 Å, respectively. (F) *ab initio* shape reconstruction from SAXS. WT-sFlaF SAXS envelop (green) was fit with sFlaF dimer structure (pink ribbon) (PDB: 3P94); I86K-sFlaF SAXS envelop (red) was fit with sFlaF chain B monomer structure (blue ribbon).

Figure S6

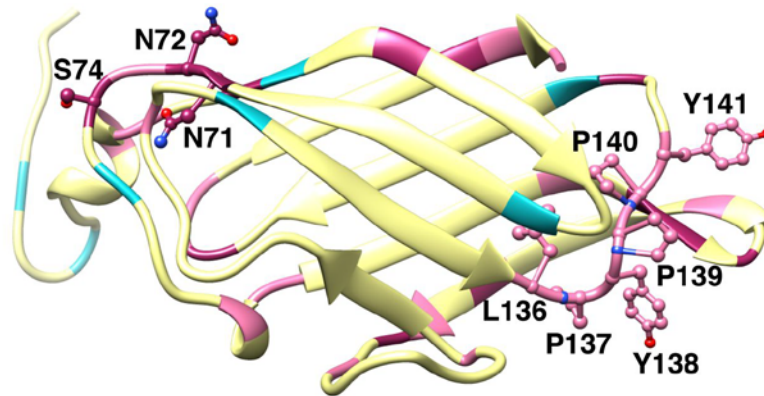
A

```
S. solfataricus    MGVSQVVA YVLIFFITIS-LGLIALEAYIKSQQLLHAENLRQNMELNQLNTRIFIKSIA 59
S. tokodaii      MSFSLILT YAFILIFISVS-LALILLS TYIRNQQLTYSEEIQQR IELNELNTKILIKSVY 59
S. acidocaldarius MGVSQTLA FILITFIGVS-FGIVLLSAYFKSQQLLNQAQELNHELELEQLETKITVSVS 59
M. sedula        MGVSNVIA FSLTIFIGLATFGSIFLVVTN-VVGTVEKSESVRQSLYLDQLNSKISIGSVS 59
                  *..*  :: :  ** :: : . : * .           :  ::::: : *::*::*: : *:

S. solfataricus    INGNL LYTITNGSTALFDFRDFAI IIKYYANI SNISTLI ISNYNS--TNLGPYKWS 117
S. tokodaii      VSNNLVYVTVTNGSTTFYDFKSFVIVKYYANISNVSTLILSQYNS--TVLAPYKWT 117
S. acidocaldarius LTGSTLNVVLENGSTNLYDFQGFVIVQYYANISNISTFNL SLYNYTKNSNPS PYYWT 119
M. sedula        LSGNDLILMVTNGSIPLWDFQHFAV I IQYYANISNKSTLLVSLYNFS--NSPSSYQWTS 117
                  :... : : : ****  ::*: * ::*:***** ** : * **:: . ..* *.

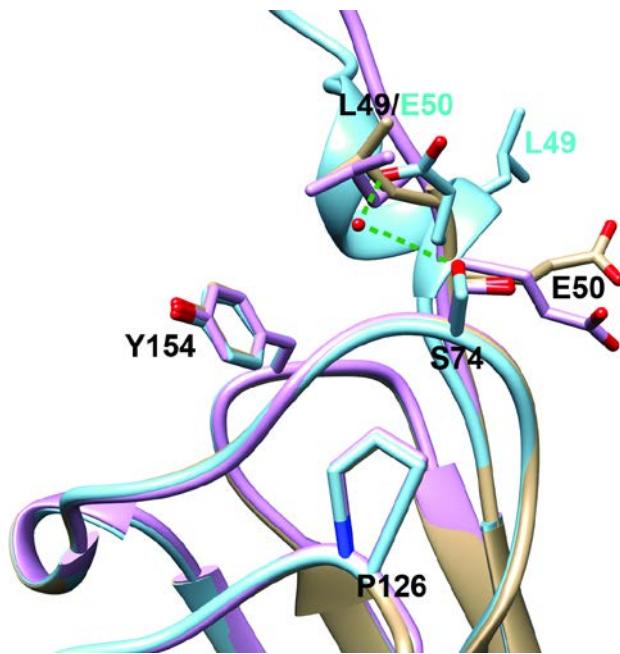
S. solfataricus    NTVLI NPNMVGTF IADLPYPPYPNTMATVVASNYGPEA IWRGIL 162
S. tokodaii      SAVVI PPDSNAVF TIDLPYPPYPNTKATI VI STNYGNEA IWRGIL 162
S. acidocaldarius NTPLLAPGSQATLTII LPYPPYPNTQATVVI VTNYGPSV IWRGSL 164
M. sedula        -SQILEPDGSS E FQIVLPYPPYPNTPAVAVI STNYGTSAVWRGTL 161
                  : :: *. . :  ***** *. * : :*** ..*** *
```

B



Related to Figure 6. Conserved regions of FlaF. (A) ClustalW alignment of FlaF from *Sulfolobales* sp. showing the most conserved regions in red. In bold red probable conserved residues involved in function. (B) The most conserved residues of sFlaF were present in the loop between $\beta 2$ and $\beta 3$ and the loop between $\beta 6$ and $\beta 7$. The sequence conservation was generated using the ConSurf webserver (1-3). The color gradient is from maroon (the most conserved), through yellow, to cyan (the least conserved).

Figure S7



Related to Figure 5. Structure alignment of sFlaF. sFlaF from two crystal forms are superimposed. PDB: 4P94 chain A (tan), chain B (cyan); PDB: 4LIO (magenta). Conserved residues L49, S74, P126, and Y154 and semi-conserved residue E50 are shown as stick. Residue S74 in chain B (cyan) forms H-bonding with E50 backbone and water-mediated H-bonding with E50 side chain that stabilizes the helix formation on N-terminus.

SUPPLEMENTARY TABLES

Table S1. Strains Used in Present Study

Strains	Relevant characteristics	Source
<i>E. coli</i>		
DH5 α	12 f80d/lacZ DM15 D(lacZYA-argF)U169 recA1 endA1 hsdR17 (rK2 mK1) supE44 thi-1 gyrA relA1	Gibco
B121(DE3)-RIL	B F-ompThsdS(rB – mB –) dcm + Tetr E. coli gal λ (DE3) endAHte [argUileYleuWCam ^r]	Stratagene
ER1821	F <i>glnV44 e14(McrA⁻) rfbD1? relA1? endA1 spoT1? thi-1</i> $\Delta(mcrC-mrr)114::IS10$	NEB
<i>S. acidocaldarius</i>		
DSM639	Wild-type <i>Sulfolobus acidocaldarius</i>	DSMZ
MW001(WT)	DSM 639 Δ <i>pyrE</i>	
Δ <i>flaI</i> Δ <i>aapF</i> , MW456	MW001, Δ <i>saci_1173</i> / Δ <i>saci_2318</i> , Δ <i>pyrE</i>	(9)
Δ <i>agl3</i>	MW001, Δ <i>saci_0423</i> , Δ <i>pyrE</i>	(31)
Δ <i>agl16</i>	MW001, Δ <i>saci_0807</i> , Δ <i>pyrE</i>	(39)

Table S2. Plasmids Used in Present Study

Plasmids	Relevant characteristics	Source
pETDuet1	Amp ^r , Cam ^r , expression plasmid containing replicon ColE1 (pBR322) and two MCS (MCS1 and MCS2)	Novagen
pSVA1921	Amp ^r , Cam ^r , pETDuet-1 carrying N-terminal His ₆ tagged <i>sflaF</i> in MCS1 using restriction sites EcoRI-HindIII	This study
pSVA1450	plasmid for expression in <i>S. acidocaldarius</i> based on pCMalLacS with <i>mal</i> promoter	(9)
pSVA1481	<i>E. coli</i> entry vector with <i>ara</i> -promoter and C-terminal Strep- and His-tag based on pGEM-T Easy backbone and pMZ1 cassette	(9)
pSVA1971	<i>flaF</i> gene in pSVA1450 backbone using NcoI-EagI	This study
pSVA1972	C-terminal His-StrepII tagged <i>flaF</i> gene in pSVA1450	This study
pSVA2801	<i>flaF N72A</i> gene in pSVA1481 using round PCR	This study
pSVA2802	<i>flaF S74A</i> gene in pSVA1481 using round PCR	This study
pSVA2803	<i>flaF Y138A</i> gene in pSVA1481 using round PCR	This study
pSVA2804	<i>flaF Y141A</i> gene in pSVA1481 using round PCR	This study
pSVA2807	<i>flaF L101Y</i> gene in pSVA1450 using round PCR	This study
pSVA2808	C-terminal His-StrepII tagged <i>flaF L101Y</i> gene in pSVA1450 using round PCR	This study
pSVA2809	Amp ^r , Cam ^r , pETDuet-1 carrying N-terminal His ₆ tagged <i>sflaF L101Y</i> in MCS1 using round PCR	This study
pSVA2813	Amp ^r , Cam ^r , pETDuet-1 carrying N-terminal His ₆ tagged <i>sflaF R161A</i> in MCS1 using round PCR	This study
pSVA2814	<i>flaF R161A</i> gene in pSVA1481 using round PCR	This study
pSVA2815	<i>flaF R161A</i> gene in pSVA1450	This study
pSVA2816	Amp ^r , Cam ^r , pETDuet-1 carrying N-terminal His ₆ tagged <i>sflaF I86K</i> in MCS1 using round PCR	This study
pSVA2817	<i>flaF I86K</i> gene in pSVA1481 using round PCR	This study
pSVA2818	<i>flaF I86K</i> gene in pSVA1450	This study
pSVA2831	<i>flaF</i> gene in pSVA1481	This study
pSVA2832	<i>flaF N72A</i> gene in pSVA1450	This study
pSVA2833	<i>flaF S74A</i> gene in pSVA1450	This study
pSVA2834	<i>flaF Y138A</i> gene in pSVA1450	This study
pSVA2835	<i>flaF Y141A</i> gene in pSVA1450	This study

Table S3. Primers Details

Primers	Sequence and characteristics	Source
P2131	CCCGGATCCTAGGCTTCCCCTCCATATTAC; reverse primer for <i>flaF</i> containing BamHI restriction site.	This study
P2155	CCCCGAATTCGAATCAAGCACAGAATTAAATC; forward primer $\Delta 34$ <i>flaF</i> containing EcoRI restriction site.	This study
P2156	CCCCAAGCTTTTATAGGCTTCCCCTCCATATTAC; reverse primer <i>flaF</i> containing HindIII restriction site.	This study
P3604	GGGGGGCCATGGATGGGAGTGTACAAACTTTG; forward primer <i>flaF</i> containing NcoI restriction site.	(9)
P3605	GGGGGGCGGCCGTCATAGGCTTCCCCTCCATATTAC; reverse primer <i>flaF</i> containing EagI restriction site.	(9)
P3763	GGGCGGCCGTTATGTTAATGAATAGAAATAC; reverse primer <i>flaF</i> containing EagI restriction site.	This study
P3727	GTGTTAGAAAATGCTGGATCTACTAATC; SaciFlaF N72A mutation forward Primer	This study
P3728	GATTAGTAGATCCAGCATTTTCTAACACAAC; SaciFlaF N72A mutation reverse Primer	This study
P3729	TAGAAAATAATGGAGCTACTAATCTGTATG; SaciFlaF S74A mutation forward Primer	This study
P3730	TCATACAGATTAGTAGCTCCATTATTTTCTAA; SaciFlaF S74A mutation reverse Primer	This study
P3731	TCAACTTTCAATTACTCGTTATATAACT SaciFlaF L101Y mutation forward Primer	This study
P3732	TAGTTATATAACGAGTAATTGAAAGTTGAGA SaciFlaF L101Y mutation reverse Primer	This study
P3733	TATCATTCTGCCAGCTCCTCCGTATCCAA; SaciFlaF Y138A mutation forward Primer	This study
P3734	ATTTGGATACGGAGGAGCTGGCAGAATGATAG; SaciFlaF Y138A mutation reverse Primer	This study
P3735	TGCCATATCCTCCGGCTCCAAATACTCAA; SaciFlaF Y141A mutation forward Primer	This study
P3736	GCTTGAGTATTTGGAGCCGGAGGATATGGCAG; SaciFlaF Y141A mutation reverse Primer	This study
P3737	TCGGTAATATGGGCGGGAAGCCTATGATT SaciFlaF R161A mutation forward Primer	This study
P3738	TAATCATAGGCTTCCCGCCATATTACCGAT SaciFlaF R161A mutation reverse Primer	This study
P3760	CAAGGTTTTTCCGTTAAAGTACAATACTATGC SaciFlaF I86K mutation forward Primer	This study
P3761	GCTAATGTTAGCATAGTATTGTACTTTAACGG SaciFlaF I86K mutation reverse Primer	This study

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Bioinformatics Analysis

In silico analyses were carried out using available online tools, e.g., Blast (Basic local alignment search tool; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>), Mutagen (Sulfolobus genome page; www.sulfolobus.org), TMHMM (Prediction of transmembrane using hidden markov model, <http://www.cbs.dtu.dk/services/TMHMM-2.0>), Flafind (Prediction of class-III signal peptide sequence in the protein; <http://signalfind.org/flafind.html>), SignalP (Prediction of class-I/II signal peptide sequence in the protein; <http://www.cbs.dtu.dk/services/SignalP>), SMART (a Simple Modular Architecture Research Tool; (<http://smart.embl-heidelberg.de>), MEME suite (Motif-based sequence analysis tools; http://meme.sdsc.edu/meme4_6_1/intro.html), the PSIPRED protein structure prediction server (<http://bioinf.cs.ucl.ac.uk/psipred/>), multicoil (two-and three-stranded coiled coil prediction) (<http://groups.csail.mit.edu/cb/multicoil/cgi-bin/multicoil.cgi>), Jpred (<http://www.compbio.dundee.ac.uk/www-jpred/>), Minnow prediction tool (Membrane protein Identification withOUt explicit use of hydrophathy profiles and alignments) and ExPASy- compute pI (http://expasy.org/tools/pi_tool.html), ConSurf server (Identification of functional regions of proteins) (1), NetNglyc 1.0.

Strains and Growth Conditions

S. acidocaldarius DSM639 was grown aerobically at 75°C in Brock's basal salts medium adjusted to pH 3.5 with sulphuric acid and supplemented with 0.1% (w/v) tryptone (Roth) or NZAmine AS (Sigma) and 0.2% (w/v) dextrin. The uracil auxotrophic *S. acidocaldarius* MW001 and $\Delta flaF\Delta aapF$ (MW453, $\Delta flaF$ in *S. acidocaldarius* $\Delta aapF$ background) (Lassak et al., 2012b) strains were grown in basal Brock medium supplemented with 10 mg ml⁻¹ uracil. To prepare plates, Brock medium was solidified by adding a final concentration of 0.6% (w/v) Gelrite and MgCl₂ and CaCl₂ to 10 mM and 3

mM, respectively. Plates were then incubated for five days at 75°C. For the propagation of plasmids, *Escherichia coli* strain DH5 α was used. Prior to electroporation into *S. acidocaldarius* plasmids were methylated in strain *E. coli* ER1821^{kan} was used. The *E. coli* BL21-DE3 containing RIL plasmid was used for heterologous expression of proteins.

Construction of Expression Plasmids

Full-length *flaF* gene (Saci_1175) was amplified from *Sulfolobus acidocaldarius* DSM639 genomic DNA using primers P3604 and P2131. To construct a heterologous expression clone in *E. coli*, the *flaF* gene product was ligated into the pSA4 backbone (Albers et al., 2003) using the *NcoI* and *BamHI* restriction sites. Primer pair P2172 and P2131 was used to construct the membrane domain truncation mutant Δ 32FlaF in the pSA4 backbone. However, this protein was insoluble. Therefore a construct was prepared in which the first 34 N-terminal amino acids were deleted, analogous to N-terminal truncations that produced soluble protein for type 4 pilin (T4P) without otherwise changing the structure (Craig et al., 2003; Hartung et al., 2011)

The pET based expression vector pETDuet1 was used to construct an N-terminal His₆-tagged Δ 34*flaF* using primers P2155 and P2156 to generate plasmid pSVA1921. A pRN1 based expression vector, pSVA1450, was used to construct plasmids for homologous expression in *S. acidocaldarius* (Wagner et al., 2009; Wagner et al., 2012). To generate a C-terminal His₆-StrepII tagged full-length FlaF (pSVA1972) in *S. acidocaldarius*, an intermediate cloning step was used where the full-length *flaF* gene product was amplified using P3604 and P2131 and cloned into pSVA1481, a high copy number plasmid derived from pMZ1, using the *NcoI* and *BamHI* restriction sites. To construct untagged full-length *flaF* gene product, primer pair P3604 and P3605 was used and the digested PCR product was ligated into the pSVA1450 backbone with *NcoI* and *EagI* restriction sites. Full-length *flaF* gene incorporated into

pSVA1481 generating pSVA2831 was used in all the *in vivo* mutation analyses. Strains and plasmids used in this study are given in Table S2-S3.

Small-angle x-ray scattering (SAXS)

The SAXS experiments were measured at the BL12.3.1 SIBYLS beamline at the ALS (Hura et al., 2009; Classen et al., 2013). The wavelength λ of the incident X-ray beam was 1 Å and the sample-to-detector distance was 1.5 meter. The scattering vector q range is from 0.01 Å⁻¹ to 0.32 Å⁻¹. The q is defined as $(4\pi\sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength. The three different concentrations of wt-sFlaF and I86K-sFlaF proteins were prepared with the matching buffer (50 mM Tris, pH 8.0, 150 mM NaCl, and 3% Glycerol). Each sample was exposed in 0.5, 1, 2, and 4 seconds. No radiation damage or aggregation was observed from the scattering data. All data were collected at room temperature (18–21 °C). The collected data were processed using scÅtter program (<https://bl1231.als.lbl.gov/scatter/>). The Guinier plot was used to calculate R_g (radius of gyration). The pair-distribution plot was converted by Gnom program (Svergun, 1992) and used to estimate D_{\max} (maximum dimension). 65 Å and 75 Å were used for the maximum dimension of wt-sFlaF and I86K-sFlaF in solution. The resulting pair-distribution function was used to determine *ab-initio* shape using GASBOR program (Svergun, et al., 2001). Ten models were determined and averaged by DAMAVER program (Volkov and Svergun, 2003). The shape and molecular structure presentation were visualized by Chimera program (Pettersen et al., 2004). The crystal structure of wt-sFlaF monomer (chain B) and dimer were converted to scattering curves and fitted to the experimental SAXS data using FoxS server (<http://salilab.org/foxs/>) (Schneidman-Duhovny et al., 2010). The fits to wt-sFlaF (dimer) and I86K-sFlaF (monomer) result $\chi^2_{\text{free}} = 2.54$ ($\chi^2 = 1.81$) and $\chi^2_{\text{free}} = 1.74$ ($\chi^2 = 1.47$), respectively. χ^2_{free} and χ^2 were calculated using scÅtter program.

SUPPLEMENTAL REFERENCES

1. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38, W529-533
2. Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25, 495–503
3. Svergun, D.I., Petoukhov, M.V. and Koch, M.H.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* 80, 2946-2953.
4. Volkov, V.V., Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36, 860–864.
5. Schneidman-Duhovny, D., Hammel M., Šali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38, W540–544.