

# Archaeal flagellin combines a bacterial type IV pilin domain with an Ig-like domain

Tatjana Braun<sup>a</sup>, Matthijn R. Vos<sup>b</sup>, Nir Kalisman<sup>c</sup>, Nicholas E. Sherman<sup>d</sup>, Reinhard Rachel<sup>e</sup>, Reinhard Wirth<sup>e</sup>, Gunnar F. Schröder<sup>a,f,1</sup>, and Edward H. Egelman<sup>g,1</sup>

<sup>a</sup>Institute of Complex Systems, Forschungszentrum Jülich, 52425 Jülich, Germany; <sup>b</sup>Nanoport Europe, FEI Company, 5651 GG Eindhoven, The Netherlands; <sup>c</sup>Department of Biological Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; <sup>d</sup>Biomolecular Analysis Facility, University of Virginia, Charlottesville, VA 22903; <sup>e</sup>Department of Microbiology, Archaea Center, University of Regensburg, D-93053 Regensburg, Germany; <sup>f</sup>Physics Department, Heinrich Heine University Düsseldorf, 40225 Duesseldorf, Germany; and <sup>g</sup>Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22903

Edited by Wolfgang Baumeister, Max Planck Institute of Biochemistry, Martinsried, Germany, and approved July 20, 2016 (received for review May 16, 2016)

The bacterial flagellar apparatus, which involves ~40 different proteins, has been a model system for understanding motility and chemotaxis. The bacterial flagellar filament, largely composed of a single protein, flagellin, has been a model for understanding protein assembly. This system has no homology to the eukaryotic flagellum, in which the filament alone, composed of a microtubule-based axoneme, contains more than 400 different proteins. The archaeal flagellar system is simpler still, in some cases having ~13 different proteins with a single flagellar filament protein. The archaeal flagellar system has no homology to the bacterial one and must have arisen by convergent evolution. However, it has been understood that the N-terminal domain of the archaeal flagellin is a homolog of the N-terminal domain of bacterial type IV pilin, showing once again how proteins can be repurposed in evolution for different functions. Using cryo-EM, we have been able to generate a nearly complete atomic model for a flagellar-like filament of the archaeon *Ignicoccus hospitalis* from a reconstruction at ~4-Å resolution. We can now show that the archaeal flagellar filament contains a  $\beta$ -sandwich, previously seen in the FlaF protein that forms the anchor for the archaeal flagellar filament. In contrast to the bacterial flagellar filament, where the outer globular domains make no contact with each other and are not necessary for either assembly or motility, the archaeal flagellin outer domains make extensive contacts with each other that largely determine the interesting mechanical properties of these filaments, allowing these filaments to flex.

archaea | flagellar filaments | helical polymers | cryo-EM

The bacterial flagellar system has been an object of intense study for many years (1–4). It has helped to elucidate issues of assembly, motility, and chemotaxis at a molecular level in a relatively simple system, typically containing ~40 different proteins. It has also been the icon for creationists in the United States who deny evolution (5–7). The bacterial flagellar filament, largely composed of a single protein, flagellin, has been fascinating from a structural point of view. In an ideal helical homopolymer, all subunits (excluding those at ends) have identical environments, and the minimum energy conformation of such a filament is a straight rod. However, the rotation of a straight rod generates no thrust, and bacterial flagellar filaments supercoil so as to behave as an Archimedean screw when rotated. The explanation for this supercoiling (8–12) is based on the notion that protofilaments in the filament can exist in two states: long and short. The short protofilaments will form the inside of a supercoil, whereas the long protofilaments will be on the outside. Structural studies of the flagellar filament using X-ray crystallography, fiber diffraction, and cryo-EM have provided a detailed picture of the switching between these two states (13–18).

The proteins that form the bacterial flagellar system have no known homologs in eukaryotic cells. The eukaryotic flagellar, based on a microtubule-containing axoneme, is vastly more complicated. In fact, the current estimate for the number of different

proteins in the axoneme is ~425 (19). In contrast, the archaeal flagellar system appears simpler than the bacterial one and can contain as few as 13 different proteins (20). As with the eukaryotic flagellar system, the archaeal one does not have homology with the bacterial one and must have arisen by means of convergent evolution. In some archaea, the flagellar filament contains mainly one protein, whereas in other archaea, several related proteins are found (21). All of these archaeal flagellins contain an N-terminal domain that is a homolog of the N-terminal domain found in bacterial type IV pilin (T4P) (22), and all contain a short signal sequence at the extreme N terminus that is cleaved by a conserved peptidase to form the mature protein, similar to what exists in T4P. As with the T4P, the highly hydrophobic and conserved N-terminal  $\alpha$ -helix exists as a transmembrane helix before the polymeric protein filament is formed. Thus, both T4P and archaeal flagellin are integral membrane proteins, assembling into a filament by a presumed common mechanism where subunits add on at the basal end. In contrast, bacterial flagellar filaments assemble in a completely different manner, where largely unfolded subunits diffuse through the hollow lumen of the filament to add on at the distal end (23).

In addition to the highly conserved N-terminal domain, bacterial T4Ps have a globular domain that shows much more sequence variation. However, in some bacterial pilins, this globular domain can be almost completely absent (24). In archaeal flagellins, however, no homology has yet been found outside of the N-terminal domain with any bacterial or eukaryotic proteins. We

## Significance

**Bacterial motility has been studied for many years, but much less is known about the flagellar system in archaea that provides motility. We have determined the structure of a flagellar-like filament from an archaeon using cryo-EM and can show how it has evolved combining domains from two different protein families. The mechanical properties of the filament are now explained from a computational analysis of the atomic model that we have been able to build. These results provide insights into how motile systems can arise by convergent evolution.**

Author contributions: G.F.S. and E.H.E. designed research; T.B., M.R.V., N.K., N.E.S., G.F.S., and E.H.E. performed research; N.K., N.E.S., R.R., and R.W. contributed new reagents/analytic tools; T.B., N.K., G.F.S., and E.H.E. analyzed data; and G.F.S. and E.H.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID code 5KYH) and the Electron Microscopy Data Bank (EMDB accession no. EMD-8298).

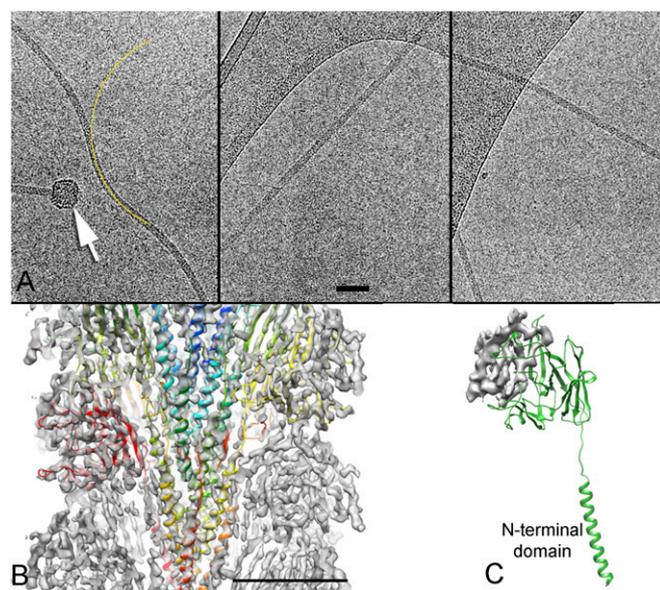
<sup>1</sup>To whom correspondence may be addressed. Email: [gu.schroeder@fz-juelich.de](mailto:gu.schroeder@fz-juelich.de) or [egelman@virginia.edu](mailto:egelman@virginia.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1607756113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1607756113/-DCSupplemental).

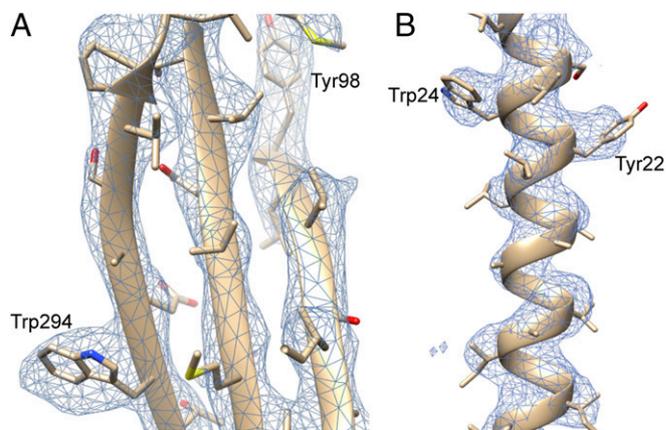
have previously described the structure of the *Ignicoccus hospitalis* Iho670 filament at 7.5-Å resolution (25). Because *I. hospitalis* has been shown to be nonmotile, these filaments were not called true flagellar filaments and have been called adhesion filaments (26). At a resolution of 7.5 Å, the N-terminal helix was clearly seen, establishing that it is packed differently in these adhesion filaments than in several different packings seen in bacterial T4P filaments (27–30). However, no information about the large globular domain was obtained at this limited resolution, and the sequence of Iho670 showed no homology with any other protein. We have now been able to take advantage of a direct electron detector to reconstruct by cryo-EM at ~4-Å resolution the Iho670 filament. This reconstruction has allowed us to trace most of the protein chain and establish that the globular domain is a  $\beta$ -sandwich and has the same fold expected for true archaeal flagellins. The Iho670 filament is, thus, a flagellar-like filament and gives us insights at the atomic level into the interfaces that hold this filament together.

## Results

Because the preparation of Iho670 filaments was very dilute, we needed to image from the same grid for 10 d to collect ~30,000 images, from which ~300 micrographs were selected that contained one or more filaments (Fig. 1A). No difference was found between images collected on the 1st day and the 10th day, showing that contamination of the grid was not an issue over this long timespan. Using the iterative helical real space reconstruction (IHRSR) method (31), a 3D reconstruction was generated (Fig. 1B), which was estimated (Fig. S1) to have an overall resolution of ~4 Å. However, the resolution was higher in the core and deteriorated at the very outside of the filament. The overall resolution was high enough (Fig. 2) that an atomic model could be built for most of the 303-residue sequence, except for a region on the periphery (residues 135–195). The model was built mostly



**Fig. 1.** Cryo-EM and 3D reconstruction of Iho670 filaments. (A) Images of frozen hydrated Iho670 filaments. The arrow points to a bulbous structure sometimes seen at one end of the filaments, likely the basal body (60). The yellow line is an arc with a radius of curvature of 1,500 Å. (Scale bar: 500 Å.) (B) A cutaway view of the 3D reconstruction, with a ribbon representation of the atomic model shown in a different color for each subunit. The core of the filament is entirely  $\alpha$ -helical, with each subunit contributing a single highly hydrophobic helix. (Scale bar: 50 Å.) (C) A single subunit from the model. The gray density shown is the region where we were unable to build a unique model, accounting for residues 135–195.



**Fig. 2.** Secondary structure is clearly resolved. (A) A  $\beta$ -sheet with two bulky aromatic residues. The presence of such large side chains allows for an unambiguous threading of the sequence through the density map. (B) A portion of the N-terminal  $\alpha$ -helix, again showing two clearly resolved bulky aromatics.

automatically with minor manual adjustments using a unique protocol (*Materials and Methods*). The structure of one subunit consists of an N-terminal helix that is connected by a linker of 8 residues to the globular outer domain containing 264 residues (Fig. 1C). The linker region has very weak density, presumably because of the flexibility in this region. The globular domain contains a  $\beta$ -sandwich fold. The results for the final refined subunit model are shown in Table 1. A Dali server (32) search with our Iho670 structure found closest similarity ( $Z$  score = 6.6) to FlaF [Protein Data Bank (PDB) ID code 4P94] from *Sulfolobus acidocaldarius* (33), an archaeal flagellar protein that has recently been shown to contain an Ig-like  $\beta$ -sandwich in addition to the conserved N-terminal transmembrane helix (Fig. S2), with no other proteins having significant similarity. FlaF is not the

**Table 1. Refinement and model statistics**

Category	Result
<b>Data collection</b>	
No. of overlapping segments	146,696
No. of unique molecules	~300,000
Pixel size (Å)	1.08
Defocus range ( $\mu\text{m}$ )	0.5–3.5
Voltage (kV)	300
Electron dose ( $e^- \text{Å}^{-2}$ )	20
<b>Refinement</b>	
Helical symmetry	106.65°/5.45 Å
Resolution (Å)	4.0
Free shell (Å)	4.3–4.0
$FSC_{\text{avg}}$ (work)/ $FSC_{\text{avg}}$ (free)	0.81/0.45
$C_{\text{work}}/C_{\text{free}}$	0.81/0.43
$R_{\text{work}}/R_{\text{free}}$	31.8/48.5
<b>Deviations from idealized geometry</b>	
Bond lengths (Å)	0.0018
Bond angles (°)	0.47
<b>Model quality</b>	
Molprobrity score	2.15 (100th percentile)
Clash score, all atoms	7.38 (100th percentile)
Good rotamers (%)	91.7
EMRinger score	3.42
<b>Ramachandran statistics</b>	
Favored (%)	96.2
Outlier (%)	1.7

**Table 2. Results from the Rosetta InterfaceAnalyzer and Alanine Scanning**

	Right-handed seven start	Left-handed three start
No. of residues in interface	20	44
Buried interface area (Å <sup>2</sup> )	1,044	1,623
Hydrophobic interface area (Å <sup>2</sup> )	514	754
Polar interface area (Å <sup>2</sup> )	550	931
Interface energy (Rosetta units)	-5.92	-7.50
Alanine scanning hotspot residues	Ser26 Val37 Ser41 Asp235 Arg236 Gln280 Ser301 Ile302	Ile13 Gln84 His105 Asp107 Tyr215 Arg226 Arg236 Gln262

actual FlaB flagellar filament protein in *S. acidocaldarius* but a homolog of FlaB that has been assumed to function as a stator protein that anchors the rotating flagellar filament to the archaeal cell envelope (33). We have compared the observed secondary structures of FlaF and Iho670 with those predicted for a number of true archaeal flagellins, such as FlaB and Mhun\_3140 (34) (Fig. S3). This comparison shows that all of these proteins are expected to have a similar globular domain that is largely composed of  $\beta$ -strands.

Interestingly, there is significant additional density in the reconstruction at Asn227 (Fig. S4) that suggested glycosylation. This finding would be consistent with previous observations (35, 36) that such glycosylation is widespread in archaeal flagellins, although for the *I. hospitalis* filament, it was not detected biochemically (26). To test this possibility, we did MS (Fig. S5), which showed that no unmodified protein was present and that the extra mass was from 1.5 to 3.5 kDa. The single largest species present (35,794 Da) corresponded to an additional 3.3 kDa above the predicted molecular mass. Fragmentation analysis of a number of ions showed that hexose molecules were being removed, establishing that some, if not most, of the additional mass was caused by glycosylation. Because the extra mass seen in Fig. S4 might account for only ~0.5 kDa if filled with sugars, the additional glycosylation is most likely in the region of residues 135–195 that we were not able to trace [specifically in residues 143–160, where there was no peptide coverage (Fig. S5C)], providing additional explanation of why this density is poorly resolved. The modification of Asn227 is likely important for the stability of the flagellar filaments, because it occurs at a pocket between three subunits. There is also a potential intramolecular disulfide bond between Cys104 and Cys271, but the resolution of the map is not good enough to see if this bond actually exists.

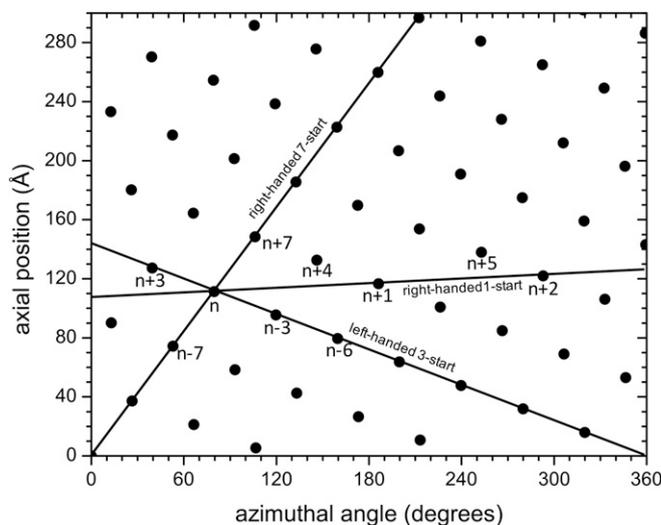
**Interface Analysis.** Although the model is missing 61 residues in the peripheral region, it is clear from the density that those residues are not involved in any interactions between the subunits. Therefore, the interfaces between neighboring subunits are fully described in the model and were analyzed using the Rosetta InterfaceAnalyzer as used in the work in ref. 37 (Table 2). For the globular domains, all interfaces are either along the right-handed seven-start helices or the left-handed three-start helices (Fig. 3). The left-handed three-start contacts are formed by subunit  $n$  with subunits  $n + 3$  and  $n - 3$ , whereas the right-handed seven-start contacts are formed by subunit  $n$  with subunits  $n + 7$  and  $n - 7$ . The subunit interface along the seven-start helices (Fig. 4) has a larger area, and its interaction energy is stronger than along the left-handed three-start helices. To identify residues in the interfaces that contribute strongly to the interaction, alanine scanning

was performed using Rosetta (38, 39). The determined hotspot residues are listed in Table 2, and details of specific subunit interactions are shown in Fig. 4.

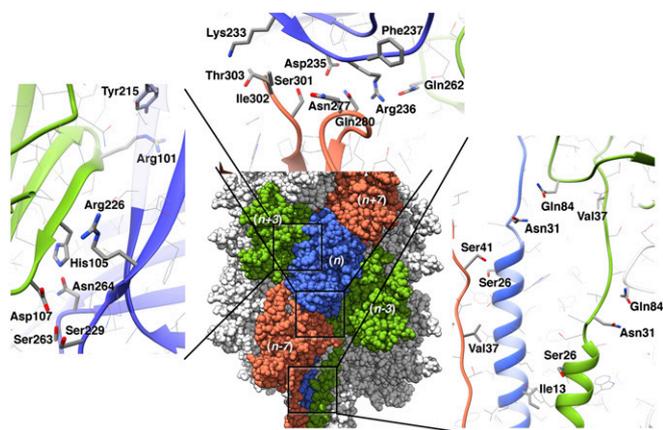
The right-handed seven-start interface is stabilized by hydrogen bonds between Gln280 and Ser301 ( $n$ ) with Asp235 ( $n + 7$ ). Also, the C-terminal carboxyl group of Ile302 ( $n$ ) forms a salt bridge with Lys233 ( $n + 7$ ). There is a potential backbone hydrogen bond between the carbonyl group of Asn277 ( $n$ ) and the amino group of Phe237 ( $n + 7$ ). Arg236 was identified as a hotspot that is relevant for both interfaces and interacts with Gln262 ( $n - 3$ ) and Gln280 ( $n - 7$ ). The left-handed three-start interface is stabilized by a hydrogen bond between Asn264 and His105 ( $n + 3$ ), and Asp107 ( $n + 3$ ) forms hydrogen bonds with both Ser229 and Ser263 ( $n$ ). In addition, there is potential hydrogen bonding between Try215 ( $n$ ) and Arg101 ( $n + 3$ ).

The hydrophobic core of the filament consists of a bundle of N-terminal helices. Each helix in this bundle contacts eight other helices:  $n - 7$ ,  $n - 4$ ,  $n - 3$ ,  $n - 1$ ,  $n + 1$ ,  $n + 3$ ,  $n + 4$ , and  $n + 7$ . Thus, each helix makes contacts along the one-, three-, four-, and seven-start helices in contrast to the globular domains that only make contacts along the three- and seven-start helices. The results from the interface analysis of one N-terminal helix (residues 1–36) with its neighboring subunits are shown in Table 3. The alanine scanning suggests a particularly stabilizing role of Ile13 in this bundle. Val37 forms a hydrophobic interaction with the N-terminal helix of subunit  $n + 7$ , in particular with Leu20 and Leu21. Ser26 stabilizes the interaction of the N-terminal helix with the outer domain of a neighboring subunit via a hydrogen bond with Ser41 on subunit  $n - 7$ . Another important interaction of the N-terminal domain is formed by Gln84 in the globular domain, which makes contact with Asn31 of the neighboring subunit ( $n + 3$ ) along the left-handed three-start helix.

**Motion Analysis.** Although bacterial flagellar filaments can generate a fixed supercoil by means of switching between two subunit states (13), the absence of any long-pitch protofilaments in the Iho670 filament raises questions about whether similar switching could exist in archaeal flagellar filaments. However, both theoretical and experimental studies have suggested that



**Fig. 3.** The helical net of Iho670 using the conventions that the surface is unrolled and that we are looking from the outside. There is a right-handed one-start helix that passes through every subunit, and subunits are labeled along this helix (e.g.,  $n - 1$ ,  $n$ ,  $n + 1$ , etc.). The outer globular domains only make significant contacts along the right-handed seven-start helices and the left-handed three-start helices, whereas the inner N-terminal helices also make contacts along the one-start helix and a four-start helix.



**Fig. 4.** Details of the subunit–subunit interfaces, with residues highlighted that were identified as hotspots by the alanine scanning. A close-up view of the interface with the  $n + 3$  subunit (blue and green) is shown in *Left*. (*Upper*) The  $n/n + 7$  interface (blue and orange) involves the C terminus. Arg236 interacts with two subunits ( $n + 7$  and  $n + 3$ ). The major interactions of the helical core with subunits  $n + 3$  and  $n + 7$  involve hydrophobic contacts (Val37 and Ile13) as well as Ser–Ser hydrogen bonds.

helical waveforms can be induced at low Reynold’s number in rotating, semiflexible filaments, and the rotation of these waveforms will generate thrust (40–42). To study the potential conformational motions of the filament, we used the program CONCOORD (43) to generate an ensemble of 400 models. The average rmsd between subunits in this ensemble is  $\sim 4$  Å; however, we used only a short filament with 21 subunits, and therefore, the conformational variability could be overestimated, because the motion of a longer filament would be more restrained. The distribution of vertical separations between subunits in the ensemble had an SD of  $\sim 3$  Å. We focused on the interaction between the globular domains (because these would dominate the mechanical properties of the filament), and the structures generated in the CONCOORD ensemble were aligned onto a central subunit. This aligned subunit together with the directly contacting four neighbors ( $n - 3$ ,  $n + 3$ ,  $n - 7$ , and  $n + 7$ ) were then subjected to principal component analysis. The first eigenvector shows an elongation along the right-handed seven-start helices coupled to a compression of the subunit along the left-handed three-start helices and vice versa (Fig. 5A and Movie S1). Because the left-handed three-start and right-handed seven-start helices have different pitches, compression along one with a similar amount of extension along the other does not cancel, and this motion along the first eigenvector will lead to an overall extension and compression of the filament.

The second eigenvector describes a motion that is related to the bending of the filament. The globular part of the subunit that points toward the core of the filament has a wedge shape (Fig. 5B and Movie S2). The second eigenvector mainly describes a

change of the angle of this wedge; the smaller the angle, the more the filament bends.

The radius of curvature of the typical filament shown in Fig. 1A is  $\sim 1,500$  Å. The transition from a straight filament to one with such a curvature requires the spacing between subunits on the inside to be compressed and the spacing between subunits on the outside to be expanded by only about 1 Å at a helical radius of 35 Å (the radius of the strongest contacts), which is well within the fluctuations of 3 Å observed by the CONCOORD analysis.

## Discussion

We had previously shown (25) that, although the Iho670 filaments contain an N-terminal domain that is a homolog of the N-terminal domain of bacterial T4P, the packing of these hydrophobic helices is different in Iho670 than in either *Neisseria gonorrhoeae* (27) or PulG (28, 29). We can now go farther and show that, although this region (residues 2–28) is completely  $\alpha$ -helical in Iho670, the corresponding region in *Neisseria meningitidis* (and by extension, *N. gonorrhoeae*, which is identical in sequence in this region) is partially melted when the bacterial filament is formed (44). This observation shows that homology at the level of tertiary structure does not dictate conserved quaternary structure because of the sensitivity of quaternary interactions to small changes in sequence (45, 46). Thus, although we expect that the packing of the N-terminal helices will be conserved across many archaeal flagellar filaments, it is also possible that small sequence changes in this family of proteins may result in different arrangements.

The high-resolution structures obtained for the *Salmonella* flagellar filament (13–15, 17) arose from using flagellin mutants that failed to supercoil. In general, rotating such a straight and rigid mutant filament does not generate thrust, and bacteria with such a mutation are nonmotile (47). Because archaeal flagellar filaments do not have the long-pitch protofilaments found in bacterial flagellar filaments (48), it has been a mystery how supercoiling might be induced in these filaments. Theoretical and experimental studies of semiflexible filaments (40–42) in highly viscous media have suggested that thrust can be generated by the rotation of such filaments. Our atomic model for the flagellar-like filament from *I. hospitalis* allows us to predict some of the conformational dynamics of such a filament (Fig. 5), and the curvature seen in micrographs (Fig. 1A) is easily within the accessible predicted conformational fluctuations without requiring large conformational changes. Thus, our structure provides the basis for future studies to understand archaeal flagellar motility in atomic detail.

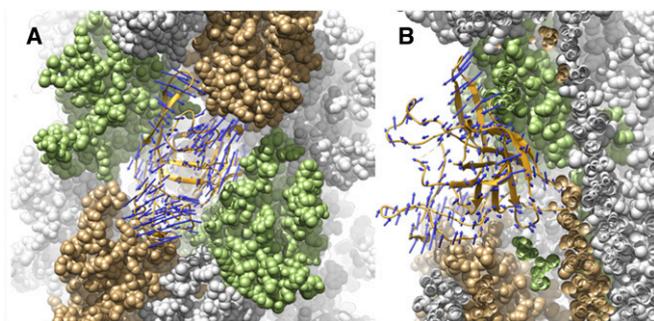
## Materials and Methods

**EM.** Samples were prepared as described (25) and applied to plasma-cleaned lacey carbon grids. These grids were verified using a Virobot Mark IV (FEI) and imaged on a Titan Krios operating at a voltage of 300 keV with a Falcon II direct electron detector. The magnification used yielded 1.08 Å per pixel. No dose fractionation was used, and the integrated images corresponded to a dose of  $\sim 20$  electrons per 1 Å<sup>2</sup>. Because the sample was very dilute and because most images did not contain any filaments, >30,000 images were collected, from which 302 were chosen that had one or more filaments.

**Table 3. Results from the Rosetta InterfaceAnalyzer for the helical core**

	Interface of helical core $n$ with							
	$n - 7$	$n - 4$	$n - 3$	$n - 1$	$n + 1$	$n + 3$	$n + 4$	$n + 7$
No. of residues in interface	39	15	29	10	10	21	14	11
Buried interface area (Å <sup>2</sup> )	1,187	553	1,054	352	373	692	568	399
Hydrophobic interface area (Å <sup>2</sup> )	946	508	837	271	294	611	559	392
Polar interface area (Å <sup>2</sup> )	241	44	217	81	78	81	9	7
Interface energy (Rosetta units)	−4.91	−5.91	−2.79	−2.38	−1.91	−4.04	−5.35	−1.24

The N-terminal helix was cut from one subunit, and the interface with all neighboring (complete) subunits was calculated.



**Fig. 5.** Potential deformations of the filament from the CONCOORD analysis. (A) The first eigenvector shows an expansion of a subunit along the right-handed seven-start helix along with a compression along the left-handed three-start helix (and vice versa). This motion induces a change of the length of the filament (Movie S1). (B) The second eigenvector describes an opening and closing of the wedge-shaped subunits, which affect the bending of the filament (Movie S2).

From these images, 220 were selected that showed a good contrast transfer function (CTF). The defocus was measured using CTFIND3 (49). The phase reversals caused by the CTF were corrected by multiplying each image by the calculated CTF, which is a Wiener filter in the limit of a very poor signal-to-noise ratio. Long filament boxes were extracted using the e2helixboxer routine within EMAN2 (50), and most other operations used the SPIDER software package (51). Overlapping boxes 400 pixels long, with a shift of 10 pixels between adjacent boxes, were cut from the long boxes, yielding 172,230 segments. The IHRSR method (31) was used for the helical reconstruction, and the final volume was from 146,696 segments. The CTF was corrected by dividing the amplitudes of the reconstruction by the sum of the squared CTFs, because the images had been previously multiplied by the CTF twice (once by the microscope and once when phases were corrected).

**MS.** Three 5- $\mu$ L aliquots of the sample were taken for digestion with Lys-C, Asp-N, and chymotrypsin. Each aliquot was heated to 95 °C for 10 min. The sample was reduced with 10 mM DTT in 0.1 M ammonium bicarbonate at room temperature for 0.5 h. The sample was then alkylated with 50 mM iodoacetamide in 0.1 M ammonium bicarbonate at room temperature for 0.5 h. The sample was again heated to 95 °C for 10 min and then, digested overnight at 37 °C with 1  $\mu$ g enzyme in 50 mM ammonium bicarbonate. The sample was acidified with acetic acid to stop digestion.

The liquid chromatography–MS system consisted of a Thermo Electron Velos Orbitrap ETD Mass Spectrometer System with an Easy Spray Ion Source connected to a Thermo 3- $\mu$ m C18 Easy Spray Column (through precolumn); 300 fmol each digest was injected, and the peptides were eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.25  $\mu$ L/min over 0.5 h. The nanospray ion source was operated at 2.3 kV. The digest was analyzed using the rapid switching capability of the instrument acquiring a full-scan mass spectrum to determine peptide molecular masses followed by product ion spectra to determine amino acid sequence in sequential scans. The data were analyzed by database searching using the Sequest search algorithm against the known sequence.

**Model Building.** The reconstructed density was sharpened using the VISDEM procedure (52) and then, filtered to 3.8 Å before the atomic model building. A map for a single subunit was segmented in Chimera (53) and used for initial model building and refinement of a single chain. To build the backbone trace, beads were placed using the tool dxbeadgen, which is part of the DireX package. The bead positions were refined using DireX, and some outliers were manually adjusted in Chimera. The connectivity was determined using e2pathwalker.py (54) from the EMAN2 package (50), where the density score was slightly modified to enforce connections to be within density. Because the peripheral region of the density had significantly lower resolution, a reliable backbone trace could not be obtained in that area. The final model is, therefore, missing 61 residues (142–202). The full atomic backbone model was built using a unique protocol. Briefly, a backbone fragment library (sequence independent) was built from seven-residue fragments by clustering of all seven-residue fragments from high-resolution X-ray structures mutated to alanine. The fragments from this library were then docked in the neighborhood of each bead position, and the best-fitting fragment was selected. These fragments were merged to yield a complete [all-alanine (all-ALA)] backbone.

The assignment of the sequence to the backbone was done by testing the fit of each amino acid at each residue position. From the resulting scores, a sequence profile was built, which was then aligned to the sequence using the Needleman–Wunsch algorithm to obtain an assignment. The all-ALA model was mutated according to the sequence assignment, and the full model was built using MODELLER, version 9 (55). As an additional test, all 326,668 sequences in SWISSPROT that are longer than 241 residues (the number of residues built in our model) were threaded onto the backbone of our model. For each sequence, a fit score was calculated. The best score obtained from the SWISSPROT sequences is 181.4, whereas the Iho670 sequence yielded a score of 232.3, which both confirms the high quality of the map and provides additional confidence in the correctness of the sequence assignment.

Additional model optimization was performed using the cryo-EM refinement protocol of Rosetta (56). The 20 best models were manually combined with Coot (57) to contain those parts of the models that fit best to the density. The final refinement of the single-chain model was done with CNS (crystallography and NMR System), and the results from this refinement are reported in Table 1. For cross-validation, only the Fourier shells 130.0–4.3 Å were used for the refinement, and the shells 4.3–4.0 Å were used for calculating the  $C_{free}$ ,  $R_{free}$ , and Fourier shell correlation [FSC;  $FSC_{avg}(free)$ ] values as described (58). To include subunit interactions, a partial filament structure was built for 21 subunits. The refined subunit models were placed into the unmasked filament density by rigid body docking in Chimera. Rosetta was used to optimize rotamers at the interfaces, and those side chains for which no clear density was visible were adapted to these predicted rotamers. The EMRinger score (59) was calculated (Table 1) to assess the map to model agreement, and it was excellent.

The atomic model has been deposited in the PDB (ID code 5KYH), and the map has been deposited to the Electron Microscopy Data Bank (accession number EMD-8298).

**ACKNOWLEDGMENTS.** We thank Michael Thomm, Harald Huber, Thomas Hader, and Konrad Eichinger for support in using the fermentation facility of the Archaea Centre Regensburg to grow *Ignicoccus* cells at large scale. This work was initially supported by Deutsche Forschungsgemeinschaft Grant WI 731/10-1 (to R.R. and R.W.) and also supported by NIH Grant EB001567 (to E.H.E.). The W. M. Keck Biomedical Mass Spectrometry Laboratory is funded by a grant from the University of Virginia's School of Medicine.

- Aizawa SI (2009) What is essential for flagellar assembly? *Pili and Flagella: Current Research and Future Trends*, ed Jarrell KF (Caister Academic Press, Poole, UK), pp 91–98.
- Toft C, Fares MA (2008) The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol Biol Evol* 25(9):2069–2076.
- Larsen SH, Reader RW, Kort EN, Tso WW, Adler J (1974) Change in direction of flagellar rotation is the basis of the chemotactic response in *Escherichia coli*. *Nature* 249(452):74–77.
- Berg HC, Anderson RA (1973) Bacteria swim by rotating their flagellar filaments. *Nature* 245(5425):380–382.
- Miller KR (2009) Deconstructing bacteria: A strategy for defending science. *Cold Spring Harb Symp Quant Biol* 74:463–468.
- Forrest B, Gross PR (2004) *Creationism's Trojan Horse: The Wedge of Intelligent Design* (Oxford University Press, New York).
- Egelman EH (2010) Reducing irreducible complexity: Divergence of quaternary structure and function in macromolecular assemblies. *Curr Opin Cell Biol* 22(1):68–74.
- Kamiya R, Asakura S, Yamaguchi S (1980) Formation of helical filaments by copolymerization of two types of 'straight' flagellins. *Nature* 286(5773):628–630.
- Calladine CR (1975) Construction of bacterial flagella. *Nature* 255(5504):121–124.
- Kamiya R, Asakura S (1976) Helical transformations of *Salmonella* flagella in vitro. *J Mol Biol* 106(1):167–186.
- Asakura S (1970) Polymerization of flagellin and polymorphism of flagella. *Adv Biophys* 1:99–155.
- Asakura S, Eguchi G, Iino T (1966) *Salmonella* flagella: In vitro reconstruction and over-all shapes of flagellar filaments. *J Mol Biol* 16(2):302–316.
- Maki-Yonekura S, Yonekura K, Namba K (2010) Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. *Nat Struct Mol Biol* 17(4):417–422.
- Yonekura K, Maki-Yonekura S, Namba K (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* 424(6949):643–650.
- Samatey FA, et al. (2001) Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* 410(6826):331–337.
- Hasegawa K, Yamashita I, Namba K (1998) Quasi- and nonequivalence in the structure of bacterial flagellar filament. *Biophys J* 74(1):569–575.
- Yamashita I, et al. (1998) Structure and switching of bacterial flagellar filaments studied by X-ray fiber diffraction. *Nat Struct Biol* 5(2):125–132.
- Mimori-Kiyosue Y, Vonderviszt F, Yamashita I, Fujiyoshi Y, Namba K (1996) Direct interaction of flagellin termini essential for polymorphic ability of flagellar filament. *Proc Natl Acad Sci USA* 93(26):15108–15113.
- Pazour GJ, Agrin N, Leszyk J, Witman GB (2005) Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 170(1):103–113.

20. Thomas NA, Bardy SL, Jarrell KF (2001) The archaeal flagellum: A different kind of prokaryotic motility structure. *FEMS Microbiol Rev* 25(2):147–174.
21. Cohen-Krausz S, Trachtenberg S (2002) The structure of the archaeobacterial flagellar filament of the extreme halophile *Halobacterium salinarum* R1M1 and its relation to eubacterial flagellar filaments and type IV pili. *J Mol Biol* 321(3):383–395.
22. Bayley DP, Jarrell KF (1998) Further evidence to suggest that archaeal flagella are related to bacterial type IV pili. *J Mol Evol* 46(3):370–373.
23. Evans LD, Poulter S, Terentjev EM, Hughes C, Fraser GM (2013) A chain mechanism for flagellum growth. *Nature* 504(7479):287–290.
24. Reardon PN, Mueller KT (2013) Structure of the type IVa major pilin from the electrically conductive bacterial nanowires of *Geobacter sulfurreducens*. *J Biol Chem* 288(41):29260–29266.
25. Yu X, et al. (2012) Filaments from *Ignicoccus hospitis* show diversity of packing in proteins containing N-terminal type IV pilin helices. *J Mol Biol* 422(2):274–281.
26. Müller DW, et al. (2009) The Iho670 fibers of *Ignicoccus hospitis*: A new type of archaeal cell surface appendage. *J Bacteriol* 191(20):6465–6468.
27. Craig L, et al. (2006) Type IV pilus structure by cryo-electron microscopy and crystallography: Implications for pilus assembly and functions. *Mol Cell* 23(5):651–662.
28. Nivaskumar M, et al. (2014) Distinct docking and stabilization steps of the Pseudopilus conformational transition path suggest rotational assembly of type IV pilus-like fibers. *Structure* 22(5):685–696.
29. Campos M, Nilges M, Cisneros DA, Francetic O (2010) Detailed structural and assembly model of the type II secretion pilus from sparse data. *Proc Natl Acad Sci USA* 107(29):13081–13086.
30. Li J, Egelman EH, Craig L (2012) Structure of the *Vibrio cholerae* Type IVb Pilus and stability comparison with the *Neisseria gonorrhoeae* type IVa pilus. *J Mol Biol* 418(1–2):47–64.
31. Egelman EH (2000) A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* 85(4):225–234.
32. Holm L, Rosenstrom P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545–W549.
33. Banerjee A, et al. (2015) FlaF is a  $\beta$ -sandwich protein that anchors the archaeum in the archaeal cell envelope by binding the S-layer protein. *Structure* 23(5):863–872.
34. Poweleit N (2016) The structure of the *Methanospirillum hungatei* flagellum as determined by cryo electron microscopy. *Biophys J* 110(3):22a.
35. Ng SY, Chaban B, Jarrell KF (2006) Archaeal flagella, bacterial flagella and type IV pili: A comparison of genes and posttranslational modifications. *J Mol Microbiol Biotechnol* 11(3–5):167–191.
36. Logan SM (2006) Flagellar glycosylation - a new component of the motility repertoire? *Microbiology* 152(Pt 5):1249–1262.
37. Lewis SM, Kuhlman BA (2011) Anchored design of protein-protein interfaces. *PLoS One* 6(6):e20872.
38. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci USA* 99(22):14116–14121.
39. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein-protein interfaces. *Sci STKE* 2004(219):pl2.
40. Coq N, Du Roure O, Marthelot J, Bartolo D, Fermigier M (2008) Rotational dynamics of a soft filament: Wrapping transition and propulsive forces. *Phys Fluids* (1994) 20(5):051703.
41. Wolgemuth CW, Powers TR, Goldstein RE (2000) Twirling and whirling: Viscous dynamics of rotating elastic filaments. *Phys Rev Lett* 84(7):1623–1626.
42. Tony SY, Lauga E, Hosoi A (2006) Experimental investigations of elastic tail propulsion at low Reynolds number. *Phys Fluids* (1994) 18(9):091701.
43. de Groot BL, et al. (1997) Prediction of protein conformational freedom from distance constraints. *Proteins* 29(2):240–251.
44. Kolappan S, Coureuil M, Yu X, Nassif X, Craig L, Egelman EH (2016) Structure of the *Neisseria meningitidis* Type IV pilus. *Nat Commun*, in press.
45. Egelman EH, et al. (2015) Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure* 23(2):280–289.
46. Galkin VE, et al. (2008) Divergence of quaternary structures among bacterial flagellar filaments. *Science* 320(5874):382–385.
47. Hyman HC, Trachtenberg S (1991) Point mutations that lock *Salmonella typhimurium* flagellar filaments in the straight right-handed and left-handed forms and their relation to filament superhelicity. *J Mol Biol* 220(1):79–88.
48. Trachtenberg S, Galkin VE, Egelman EH (2005) Refining the structure of the *Halobacterium salinarum* flagellar filament using the iterative helical real space reconstruction method: Insights into polymorphism. *J Mol Biol* 346(3):665–676.
49. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–347.
50. Tang G, et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157(1):38–46.
51. Frank J, et al. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116(1):190–199.
52. Spiegel M, Duraisamy AK, Schröder GF (2015) Improving the visualization of cryo-EM density reconstructions. *J Struct Biol* 191(2):207–213.
53. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
54. Baker MR, Rees I, Ludtke SJ, Chiu W, Baker ML (2012) Constructing and validating initial  $C\alpha$  models from subnanometer resolution density maps with pathwalking. *Structure* 20(3):450–463.
55. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
56. DiMaio F, et al. (2015) Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods* 12(4):361–365.
57. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
58. Falkner B, Schröder GF (2013) Cross-validation in cryo-EM-based structural modeling. *Proc Natl Acad Sci USA* 110(22):8930–8935.
59. Barad BA, et al. (2015) EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* 12(10):943–946.
60. Meyer C, Heimerl T, Wirth R, Klingl A, Rachel R (2014) The Iho670 fibers of *Ignicoccus hospitis* are anchored in the cell by a spherical structure located beneath the inner membrane. *J Bacteriol* 196(21):3807–3815.
61. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858.