High-resolution archaellum structure reveals a conserved metal-binding site

Vladimir A Meshcheryakov, Satoshi Shibata, Makoto Tokoro Schreiber, Alejandro Villar Briones, Kenneth F Jarrell, Shin-Ichi Aizawa, Matthias Wolf

EMBO reports

20

5

e46340

2019-03-21

EMBO Press

(C) 2019 The Author(s).

https://creativecommons.org/licenses/by-nc-nd/4.0/

doi: info:doi/10.15252/embr.201846340
High-resolution archaellum structure reveals a conserved metal-binding site

Vladimir A Meshcheryakov1,†, Satoshi Shibata1, Makoto Tokoro Schreiber1, Alejandro Villar-Briones2, Kenneth F Jarrell3, Shin-Ichi Aizawa4 & Matthias Wolf4,*†

Abstract

Many archaea swim by means of archaella. While the archaellum is similar in function to its bacterial counterpart, its structure, composition, and evolution are fundamentally different. Archaella are related to archaeal and bacterial type IV pili. Despite recent advances, our understanding of molecular processes governing archaellum assembly and stability is still incomplete. Here, we determine the structures of Methanococcus archaella by X-ray crystallography and cryo-EM. The crystal structure of Methanocaldococcus jannaschii FlaB1 is the first and only crystal structure of any archaellin to date at a resolution of 1.5 Å, which is put into biological context by a cryo-EM reconstruction from Methanococcus maripaludis archaella at 4 Å resolution created with helical single-particle analysis. Our results indicate that the archaellum is predominantly composed of FlaB1. We identify N-linked glycosylation by cryo-EM and mass spectrometry. The crystal structure reveals a highly conserved metal-binding site, which is validated by mass spectrometry and electron energy-loss spectroscopy. We show in vitro that the metal-binding site, which appears to be a widespread property of archaellin, is required for filament integrity.

Keywords archaea; cryo-EM; electron energy-loss spectroscopy; helical reconstruction; Methanococcus

Subject Categories Microbiology, Virology & Host Pathogen Interaction; Structural Biology

DOI 10.15252/embr.201846340 | Received 26 April 2018 | Revised 16 February 2019 | Accepted 27 February 2019 | Published online 21 March 2019

EMBO Reports (2019) 20: e46340

Introduction

Motility is widespread throughout the prokaryotic world. The most common motility organelle in bacteria is a rotating flagellar filament ~ 20 nm in diameter, while in archaea it is the much thinner filament (typically 10–14 nm in diameter) now known as the archaellum (formerly archaeal flagellum [11]). Although bacterial flagella and archaella share similar functions, their structures, compositions, and modes of assembly are fundamentally different. Archaella are composed of a limited number of Fla proteins, most of which are encoded in a single operon. None of these proteins shows homology to any bacterial flagellar proteins, but several are homologs of proteins pertaining to type IV pili [2,3]. Archaea often thrive in extreme environments, where their external structures, including archaella, may be exposed to wide ranges of temperature, pH, and salt concentration. The enhanced resistance of archaellae filaments to extreme environmental conditions makes them attractive objects from both fundamental and biotechnological points of view. Archaellins are the main protein components of the archaellum and represent the structural analogs of bacterial flagellins. In euryarchaeotes, archaella are characterized by the presence of multiple (≥ 5) archaellin proteins. Such multiplicity is assumed to be important for assembly of functional archaella, since deletion of individual archaellin genes often leads to non-archaellated cells [4–6]. Archaellin amino acid sequences are unrelated to those of bacterial flagellins, but they have homology to type IV pilins at their N-termini [7]. Like type IV pilins, archaellins are synthesized as preproteins with type IV pilin-like (class III) signal peptides that are cleaved by a specific membrane-located signal peptidase (FlaK/PibD), homologous to prepilin peptidase in type IV pili systems [5,8,9]. Cleavage of the signal peptide is essential for incorporation of archaellins into the archaellum [8]. Despite significant conservation of the N-termini of archaellins and type IV pilins that form the cores of the corresponding structures, their packing is different [10–12]. In contrast to bacterial analogs, archaellin structures are not well studied. Until recently, the best available pioneering structures of archaella, from the extreme halophile Halobacterium salinarum and the thermoacidophile Sulfolobus shibatae, were determined at resolutions too low to resolve subunit domains, but they did reveal the absence of an internal channel found in bacterial flagella and thus supported a model of assembly from their base [13–16]. Recently, two cryo-EM reconstructions of archaellar filaments [11,17], as well as an archaella-like adhesion structure in Ignicoccus hospitalis [12], have clarified the general structure of archaellin and its biological assembly. These were the first cryo-EM-based structures of an archaellum at a level of resolution to

1 Molecular Cryo-Electron Microscopy Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Kunigami, Okinawa, Japan
2 Instrumental Analysis Section, Okinawa Institute of Science and Technology Graduate University, Onna, Kunigami, Okinawa, Japan
3 Department of Biomedical and Molecular Sciences, Queen’s University, Kingston, ON, Canada
4 Department of Life Sciences, Prefectural University of Hiroshima, Shobara, Hiroshima, Japan
*Corresponding author. Tel: +81 98 966 8987; E-mail: matthias.wolf@oist.jp
†Present address: Faculty of Health Sciences, University of Macau, Taipa, Macau SAR, China
trace secondary structure elements with confidence. They revealed the fundamentally different structure of archaecal filaments compared to bacterial flagella and type IV pili. Daum et al [17] further described the structure of the archaellum motor complex of *Pyrococcus furiosus* (PDB code 5O4U) by cryo-electron tomography. To understand archaellar structure at the atomic level, we have used a hybrid approach by combining X-ray crystallography and cryo-EM. We solved the crystal structure of an N-terminal truncation construct of archaellin FlaB1 from the hyperthermophilic methanogen, *Methanocaldococcus jannaschii*, at a resolution of 1.5 Å. Our data reveal a highly conserved metal-binding site, a feature not previously described for archaellins. Simultaneously, we determined the structure of the archaellar filament from the mesophile *Methanococcus maripaludis*, using cryo-EM at 4 Å resolution to provide the biological context.

## Results and Discussion

### Crystal structure of the globular domain of *Methanocaldococcus jannaschii* FlaB1 archaellin

Archaellins possess highly conserved and mostly hydrophobic N-terminal sequences. Their N-terminal regions form α-helices, which are important for filament formation. To prevent aggregation and improve protein solubility, we prepared a truncated version of archaellin FlaB1 of *M. jannaschii*, containing residues 39–205 of the mature protein with a C-terminal His₆ tag. FlaB1₃⁹–₂⁰⁵ crystallized in the *P*₂₁₀₁₂ space group with two molecules in its asymmetric unit. The structure was solved by multiple isomorphous replacement and refined to 1.5 Å resolution (Table EV1). Each molecule in the asymmetric unit comprises residues 47–205. No electron density was observed for residues 39–46 at the N-terminus. This region is predicted to form an α-helix, which was later confirmed in the biological assembly by our cryo-EM map. In contrast, in bacterial pilins, the C-terminal half of the hydrophobic α-helix is embedded in the β-sheet of its globular domain [18–20]. The globular domain consists of several β-strands that fold into a β-barrel (Fig 1A). Within this structure, a region in one of the β-hairpins had a strong anomalous signal. Structural analysis with ConSurf [21] showed that these residues are highly conserved (Fig 1B and C). No metal ions were present in the crystallization solution. The geometry of the binding site suggested that the binding metal in *M. jannaschii* FlaB1₃⁹–₂⁰⁵ is likely Ca (2F_o–F_c map in Figs 1D and EV1). Refinement of the crystal structure with Ca ions at full occupancy resulted in B factors for Ca ions of 10.1 Å² in chain A and 17.5 Å² in chain B, which are close to environmental B factors of coordinating atoms of 9.9 and 16.8 Å², respectively. The metal ion is coordinated by the side chains of residues D139, D141, S143, N150, and D153, and one water molecule (Figs 1D and EV1). Systematic sequence comparison shows that this metal-binding site exists in a wide range of *Euryarcheota* (Figs EV2A and 2B). A database search performed with DALI [22] revealed that the closest structural homolog of *M. jannaschii* FlaB1 is archaellin FlaB0 from *P. furiosus* (PDB code 5O4U; Z-score = 22; RMSD = 1.6; Fig EV2C). Although FlaB0 possesses all conserved metal-binding residues, metal binding was not incorporated in their cryo-EM-based model [17].

### Cryo-EM analysis

We prepared cryo-EM samples of archaella from the mesophile *Me. maripaludis*, because it is difficult to grow the hyperthermophile *M. jannaschii* and since all our attempts to obtain crystals of full-length archaellin from *Me. maripaludis* were unsuccessful. FlaB1 proteins of *Me. maripaludis* and *M. jannaschii* share high level of sequence homology, with identity/similarity of 56.2 and 70.5%, respectively. Archaella from *Me. maripaludis* embedded in amorphous ice were clearly visible in cryo-EM images (Fig 2A). Image segments were aligned and classified into 100 classes that displayed much improved detail and contrast (Fig 2B and Table EV2). The average filament diameter was ~100 Å. As reported also by Poweleit et al in their samples of archaella from *Methanospirillum hungatii* (PDB code 5TFY), we occasionally observed “minor” filaments with a diameter of approximately 5–7 nm (Fig 2A, arrow, Fig EV3A). After initial analysis, its structural features appeared to be most consistent with the architecture of type IV pili, another surface appendage present in *Me. maripaludis* [23,24]. We reconstructed the archaellum from *Me. maripaludis* in three dimensions (3D) using single-particle cryo-EM and helical analysis. Because a real-space symmetry search can lead to ambiguities, we used the classic method of analyzing the helical diffraction pattern to obtain the helical symmetry operator. This layer line pattern, which corresponds to the projected Fourier transform of the object, is an average of the power spectra of all aligned boxed filament segments, with much improved signal-to-noise ratio (SNR; Fig 2C). Helical analysis of this layer line pattern allowed us to construct a lattice, which corresponds to the unrolled pattern of features on the surface of a cylinder surrounding the tubular helix [25], that follows the selection rule l = 3π + 10m (Fig 2D). We determined the absolute handedness of the basic (or 1-start) helix, which is right-handed, using metal shadowing and verified the correct hand by docking the crystal structure into the reconstruction. The true pitch of the basic helix, which is the axial displacement after which the helical pattern repeats exactly, is 57 Å and contains a set of 10 subunits (rise = 5.7 Å per subunit, rotation 108°). Using this helical operator, we reconstructed the helix in 3D from aligned image segments, followed by iterative refinement using SPRING [26] (Fig 2E and Table EV2).

The archaellum appears enforced by a three-stranded helix (Fig 2E) consisting of tightly packed tadpole-shaped subunits (Fig 2F), the long α-helical tails of which form the core and the globular head domains of which form the outer layer of the filament (Fig 2E–G). As reported for other archaella [11,12,14,17], there is no internal lumen of significant diameter (Fig 2H). The strongest surface features, which determine the macroscopic appearance of the archaellum, are formed by its head domains that run in the 3-start and the 7-start directions, creating the visible deeply grooved appearance on the surface of the filament (Fig 2D and E). The archaellum also has a prominent pattern in the 10-start direction (vertical arrow in Fig 2D), which coincides precisely with the axial direction of the filament. This spatial relationship creates the 10-fold symmetry when observing a thick radial slab in axial direction (Fig 2H). Strong coupling between subunits 0 and 10, as indicated in the helical lattice, suggests an interaction between these remote structural features. The averaged helical diffraction pattern from all aligned image segments contained visible layer lines to 4 Å.
resolution. In the diffraction pattern (Fig 2C), numbered layer lines are listed together with their Bessel order.

Individual subunits of the *Me. maripaludis* archaellum consist of an ~ 80-Å tail formed by a single conserved α-helix attached to a tadpole-shaped head domain (Fig 2F). Throughout the structure, map resolution (Fourier shell correlation, Fig EV3B) allowed unambiguous placement of the polypeptide chain and in many cases there was clear side-chain density (Fig 2F, inserts). The structure of *M. jannaschii* FlaB1 was fitted as a rigid body into the reconstructed electron potential map of the *Me. maripaludis* archaellum. The initial model was further modified according to the sequence of *Me. maripaludis* archaellins and refined by real-space fitting at 4 Å resolution. The genome of *Me. maripaludis* encodes three archaellin proteins—FlaB1, FlaB2, and FlaB3 [6]. Although we tried to build models for all three proteins, only the sequence of FlaB1 could be fitted well into our cryo-EM density map, indicating the filament is composed primarily of this single archaellin. Each FlaB1 subunit comprises two domains: N-terminal α-helix and a C-terminal globular part, similar to that described for archaellins of *P. furiosus* and *Msp. hungatei* [11,17]. The density map for the linkage region between both domains is well defined, suggesting little flexibility between domains within archaellar filaments. Archaeal filaments are characterized by multiple contacts between subunits. The core of the filament is formed by hydrophobic interaction between the extended N-terminal α-helices (Fig 2E, G and H). These N-termini make contact with the inner parts of head domains in the 10-start helix direction. Furthermore, there are distinct protein–protein interactions between globular domains, reinforcing major directions (Fig 2D) of the helical lattices between their eight adjacent subunits (Fig 3). In the helical diffraction pattern, there are two pairs of prominent spots on layer line 1 that originate from the tightly packed α-helices in the filament core (Figs 2C and EV3C).

Figure 1. Archaellin fold and metal-binding site.

The soluble domain of archaellin forms a 9-stranded anti-parallel β-barrel adorned with three short helices (γ1 = 3_10 helix, α = α-helix). Metal atom depicted by yellow sphere.
B Same view of the folded polypeptide chain colored by conservation index among 100 archaellin homologs (purple: highly conserved, turquoise: variable sequence). Metal-binding site shows side-chain conformations.
C Aligned sequences from FlaB proteins (Methanocaldococcus jannaschii, Methanococcus maripaludis, Pyrococcus furiosus, Methanospirillum hungatei) containing the highly conserved metal-binding site (marked by red stars). Full alignment in Fig EV5.
D Fo/C0 electron density map around the metal-binding site of *M. jannaschii* FlaB1-38 solved by X-ray crystallography to 1.5 Å resolution (iso-electron density representation at 1σ). Metal-coordinating amino acid side chains are highlighted in green. Metal atom density at center. Stereo view in Fig EV1.
E Reconstructed cryo-EM electron potential map at 4 Å resolution (iso-potential surface contoured at 3.5σ above average, map low-pass filtered at 3.7 Å) of *Me. maripaludis* FlaB1 region displayed in the same orientation and at similar scale as (D). The single coordinated metal atom is clearly resolved in the cryo-EM map.
The surrounding stacked $\beta$-sheets in the head domains give rise to a circular set of spots arranged in a spoke-like pattern, which corresponds to the unusual even-numbered, 10-fold radial symmetry created by stacking interactions in the 10-start helix direction (Fig EV3C). As a consequence, in the diffraction pattern, there is an overlap of the spots on layer line 1.

Our model of the *M. maripaludis* archaellum contains only one archaellin protein. Previously, it was shown that in *M. maripaludis*, two of the three archaellins, FlaB1 and FlaB2, are essential for filament assembly, although the role of each archaellin in archaellation is unknown [6]. The third archaellin, FlaB3, apparently forms the hook region, since deletion of this gene results in archaella lacking this curved region. It is possible that all archaellin proteins are incorporated into the same filament. Thus, our model may represent only the archaellum region enriched in FlaB1. However, our mass spectrometry results indicate that FlaB1 is the dominant component in our sample (Table 1), consistent with our structure interpretation. Whether the distribution of archaellins in the filament is random or whether it is regulated by some mechanism remains to be investigated. Interestingly, in both the *Msp. hungatei* and *P. furiosus* archaella samples recently studied, filaments were assumed composed of a single archaellin, even...
though multiple archaellin genes are found in both organisms [11, 17].

Interactions within the biological assembly

The helical lattice of archaellin subunits is unique and remarkable, as its symmetry should result in a sevenfold repeat (the 3-start and 4-start directions are the most prominent features; Fig 2D). In general, most self-assembling helical structures contain an uneven number of subunits. Otherwise, polymerization would easily terminate by forming a closed ring. However, the archaellum features strong coupling in the 10-start direction, parallel to the helical axis. The result is an even-numbered 10-fold symmetry when viewing a thick radial slab of the archaellum in the direction of the helical axis (Fig 2H) — a unique geometry unknown in any other helical filaments. The true pitch of the helical assembly is marked by the 10-start interaction and suggests an interaction of subunits in the axial direction, translated by 57 Å. Indeed, the 75-Å α-helical tail of subunit 10 interacts with the innermost part of the head domain of subunit 0 in perfect axial alignment by means of hydrophobic interactions, indicating that the 10-start helix is the polymerization site of the filament (Figs 2D and 3, right). Furthermore, each of the major lattice directions comes with its own set of interactions, responsible for torsional and longitudinal rigidity and for imparting mechanical properties to the filament required for archaeal motility.

The major surface feature is created by the tightly packed head domains that wrap around the entire filament along the 3-start direction (Fig 3 left). We assume that this structural arrangement conveys greater torsional stability to the filament.

Structure comparison

Comparison of the structure of Me. maripaludis FlaB1 with archaellins of P. furiosus [17] and Msp. hungatei [11] shows that these archaellins have similar shapes: a long N-terminal α-helix and a globular C-terminal domain (Figs 2F and EV2C). The α-helix is highly conserved, because it is essential for the formation of the filament backbone. The key motif of the globular domain is the β-barrel, which varies among species due to insertions or deletions between β-strands. Globular domains of FlaB1 of Me. maripaludis and FlaB0 of P. furiosus show remarkably similar structures, with an RMSD of only 1.6 Å (Fig EV2C). In contrast to FlaB3 of Msp. hungatei, archaellins of Me. maripaludis and P. furiosus have an insertion of ~40 amino acids. A similar insertion is found in M. jannaschii FlaB1. In general, these insertions are not conserved. However, they all contain a highly conserved metal-binding site. Moreover, this region makes multiple contacts with neighboring subunits in the archaellum. It seems that the insertion is only partially stabilized by metal binding; the remainder of the insertion shows large deviations. This assumes some flexibility of this region
that might be important for better accommodation of protein subunits during filament formation. The archaellum of Me. maripaludis has a rotational symmetry (108 degrees) close to those of P. furiosus and Msp. hungatei, but a distinct axial rise per subunit (5.7 Å for Me. maripaludis and 5.4 and 5.3 Å for P. furiosus and Msp. hungatei, respectively). Such differences in helical parameters, despite structural similarities, suggest that differences in filament arrangement can be the result of small variations in amino acid sequence [12]. Interestingly, another structural homolog of archaellin is FlaF Sulfolobus acidocaldarius (PDB code 4ZBH), an archaenal protein suggested to function as the archaellum stator [27] (Fig EV2C). FlaF is not a filamentous protein nor is it related to archaellin. Whether this protein plays a role in archaellum assembly remains yet to be discovered.

**Posttranslational modification**

Posttranslational modification by covalent attachment of carbohydrates has been reported on many archaellins [11,28,29]. The difference map between the experimental cryo-EM electron potential map and the model-derived density contains some amino acid modifications that are not included in the model. Such extra density marks asparagines residues N103 and N116 (Fig 4: numbering refers to the processed archaellin with signal peptide removed), which are the most prominent modifications. We estimated signal-to-noise ratio (SNR) of the masked side-chain densities in the difference map for each visually identified putative modification site (N103, N110, N116, N128, N140, N156) and disregarded weaker difference densities with SNR < 3 as insignificant, leaving only N103, N116, and N128 as significant modifications (Fig 4, right). However, the high SNR in the difference density of N128, which is located inside a β-sheet, chiefly stems from the inclusion of neighboring protein β-strands within the masked density. Furthermore, N128 is not part of a consensus N-glycosylation sequon (NXS/T where X is not proline) and no glycosylation of this site was detected previously [28]. In that study, only canonical N-linked sequons containing N110, N116, and N156 were glycosylated. Therefore, we exclude N128 as post-translational modification site based on these observations. We did not find significant difference density at N156 in our present analysis, while Kelly et al [28] showed that this site was clearly glycosylated with a tetrasaccharide. Based on our tandem MS-MS data using an error-tolerant search with all possible modifications enabled, we can furthermore exclude phosphorylation, acetylation, lipidation, and O-linked glycosylation. Only N-linked glycans were positively identified, consistent with the study by Kelly et al [28]. Non-canonical N-linked glycosylation was recently suggested in analyses of archaella from Msp. hungatei [11]. In the case of P. furiosus archaellin, five out of 13 asparagine residues were found to have large adjacent densities, suggesting possible N-linked glycans. All five asparagines were located as part of canonical N-linked glycosylation sequons, while none of the remaining eight asparagines, which did not show a similar density nearby, were part of such a sequon [17]. The Msp. hungatei archaellin was also modified at six sites with O-linked glycan, so far, unique for archaellins [11]. No indication of possible O-glycosylation was found in the P. furiosus archaellin [17]. Based on our tandem mass spectrometry analysis with error-tolerant search, any other modifications with exception of N-linked glycans can be excluded in Me. maripaludis. For several archaea, interference with the N-glycosylation pathway often results in motility defects and even non-archaellation, suggesting a role for this posttranslational modification in assembly or/and stability of the filaments [29–32]. Typically, such modifications are found in the C-terminal portion of the subunits, exposed to the environment [11,17] where they may contribute to the adhesive properties reported for archaella in various archaea.

**Metal-binding site in Methanococcus maripaludis archaellin**

The novel metal-binding site identified in the anomalous difference map of M. jannaschii archaellin (Figs 1D and EV1) is also found in our cryo-EM map of the Me. maripaludis archaellum, which resolves the hydrated single metal atom with surprising clarity (Fig 1E). The density is surrounded by the same conserved amino acids as in the M. jannaschii crystal structure. We validated the presence and nature of the bound metal in the native Me. maripaludis archaella by four methods: (i) inductively coupled plasma mass spectrometry (ICP-MS), (ii) high-resolution tandem liquid chromatography mass spectrometry (LC-MS/MS), (iii) electron

| Table 1. Metal composition in Methanococcus maripaludis archaellum. |
|-----------------|-----------------|-----------------|
| Element        | LC-MS/MS        | ICP-MS          |
|                | Associated      | PA              | mg l⁻¹ | EELS |
|                | sequence        |                 |       |      |
| ⁵²S            | N.D.            | 8.9             |       | –    |
| ³⁴Mg           | YDDSAVNAVINKG   | D203            | 11.4  | +    |
| ⁶⁵Zn           | N.D.            | 2.8             |       | –    |
| ⁵⁶Fe           | YDDSAVNAVINKG   | D203            | N.D.  | 2.4  |
| ⁴⁰Ca           | DDDSAVNAVINKG   | D203            | 24    | N.D. |

High-resolution tandem mass spectrometry (LC-MS/MS): element, FlaB1 consensus sequence from up to 10 overlapping metal-associated peptide fragments, associated amino acid (aa) assigned by software, sum of peak areas (PA) of metal-associated fragments in million units as a measure of relative abundance. Although a Fe-associated fragment was identified, its PA could not be determined. The analysis used data from four separate samples. All identified peptides had a false discovery rate of < 0.01%. Inductively coupled plasma mass spectrometry (ICP-MS) analysis shows absolute concentration of five identified elements after excluding carbon, oxygen, and nitrogen (N.D.: not detected). Samples were injected in triplicate. Ratio sample/background was > 10. Consistent with the most abundant divalent cation detected by mass spectrometry, only Mg was found by electron energy-loss spectrometry (EELS) using a direct electron detector in electron counting mode (EELS spectra in Fig EV4), but not Fe or Ca. The EELS experiment was repeated seven times.
energy-loss spectroscopy (EELS) in amorphous ice using a direct electron detector, and (iv) exposure to chelators followed by negative stain electron microscopy. ICP-MS is capable of detecting metal concentrations as low as 1 ppq. The purified archaella sample from *Me. maripaludis* was dialyzed for 24 h against Q-POD-purified water (Millipore Inc, ultra-low metal release) with multiple solvent exchanges. The same water was used as a reference for background subtraction. The most abundant elements were S>Mg>Zn>Fe. Ca was not detected (Table 1). In a separate experiment, the dialyzed sample (using ion-free ultra-purified water) was analyzed on a hybrid orbitrap mass spectrometer after separation by HPLC and ionization by electrospray. Fragments were identified according to their mass and charge by comparison with the *Me. maripaludis* genomic sequence. Peptide fragments with a second, shifted chromatographic peak corresponding to the masses of Mg, Fe, and Ca adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of metal-associated peptide fragments are consistent with the sequence forming adducts were identified. The consensus sequences of meta...
addition to its charge, the tightly associated hydration shell may further increase its apparent diameter.

The role of metal ions in the archaeal motility system is not clear. The obvious one is stabilization of archaellin structure. A variety of stabilizing strategies has been described for bacterial pilin/pseudopilin, including Ca binding [39]. The metal-binding site (Figs 1, 3) stabilizing strategies has been described for bacterial pilin/pseudopilin, including Ca binding [39]. The metal-binding site (Figs 1, 3) its biological assembly was determined by cryo-EM at 4 Å resolution [37]. This suggests that the divalent metal-binding motif is a more common feature, which may be conserved through evolution of these related structures.

In conclusion, we have determined the structure of the archaellin in cartoon representation are colored by sequence number from blue (N-terminus) to red (C-terminus). Closest inter-subunit contacts (dotted ovals) are formed between strand β6 and a short anti-parallel loop (labeled by asterisks) near β6 and the metal-binding site (yellow spheres). Note how the sequence colored in green forms a "paper clip"-like structure reaching across each subunit in the 3-start direction (along the diagonal in the figure).

**Materials and Methods**

**Cloning, expression, and protein purification of Methanocaldococcus jannaschii archaellin**

DNA fragment encoding truncated *M. jannaschii* FlaB1Δ1–38 (numbers correspond to numeration in mature protein) fused to C-terminal His6 tag was amplified by PCR from genomic DNA of *M. jannaschii* as a template using the following primers containing NcoI and BamHI restriction sites (in bold): 5′-GGT GGT CCA TGG GGA TTA ATG ATG ATG ATG ATG ATG TTG TAA CTC AAT TAC and 5′-GGT GGT TCC CTTA AATG ATG ATG ATG ATG TTG TAA CTC AAT TAC AGT TTG TGT ATA TGC AG-3′. The PCR product was double-digested with NcoI and BamHI and then ligated into the pET52b(+) expression vector (Novagen).

The recombinant vector containing the gene of FlaB1Δ1–38 protein was transformed into *E. coli* strain Rosetta (DE3; Novagen). Transformed cells were cultured at 37°C to late exponential phase in 2 l Luria-Bertani medium containing 50 μg ml⁻¹ ampicillin and
34 µg ml⁻¹ chloramphenicol. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at an OD₆₀₀ of 0.7–0.8. After 3-h incubation at 37°C, cells were harvested by centrifugation at 8,000 g for 15 min and suspended in 200 ml buffer A (40 mM HEPES, pH 8.0, 0.5 M NaCl, 10 mM imidazole). Suspended cells were sonicated on ice. Cell debris was removed by centrifugation for 30 min at 30,000 g, 5°C. Clarified supernatant was applied to a 5-ml HisTrap HP column (GE Healthcare) equilibrated with buffer A. The column was washed with 50 ml buffer A. Bound proteins were step eluted with increasing concentration of buffer B (40 mM HEPES, pH 8.0, 0.5 M NaCl, 0.5 M imidazole). Fractions containing FlaB₁₋₁₃₈ were combined, dialyzed overnight against 20 mM HEPES, pH 8.0, 50 mM NaCl, and applied to a 5-ml HiTrap Q HP column (GE Healthcare) equilibrated with the same buffer. Protein was eluted with a linear gradient of NaCl from 0.05 to 1 M in 20 mM HEPES and pH 8.0. Fractions containing pure protein were pooled, and the protein was transferred by dialysis in 10 mM Tris–HCl, pH 7.5, and 50 mM NaCl.

Crystallization of Methanocaldoccus jannaschii FlaB₁₋₁₃₈

Purified protein was concentrated to 10 mg ml⁻¹ using an Amicon Ultra system (Millipore) with a molecular-weight cutoff of 10 kDa. Protein concentration was measured by UV spectroscopy at 280 nm using a calculated absorption coefficient of 0.79 mg ml⁻¹ cm⁻¹. Preliminary screening of crystallization conditions was performed by the sitting-drop vapor-diffusion method in 96-well plates using an automated nanolitre liquid-handling system (Mosquito, TTP Labtech) at two temperatures: 10 and 20°C. 150 nl protein solution was mixed with 150 nl reservoir solution. Diffraction-quality crystals were grown in 32.5% PEG 4,000.

Data collection, processing, and model building of Methanocaldoccus jannaschii FlaB₁₋₁₃₈

Prior to data collection, single crystals were briefly soaked in cryoprotectant solution. Subsequently, crystals were mounted on a nylon loop and flash-cooled in liquid nitrogen. X-ray diffraction data were collected on beamline BL38B1 at SPring-8 using an ADSC Quantum 315r CCD detector. Diffraction data were processed and scaled using MOSFLM and SCALA from the CCP4 package [40]. The protein structure was solved using the multiple isomorphous replacement method. Heavy atom derivatives were prepared by soaking crystals in stabilization solution (30% PEG 4,000) containing a heavy atom component. Soaking time varied from several hours to several days. Two isomorphous derivatives were prepared: K₂OsO₄ (1 mM; 18 h soak) and p-chloromercuribenzoic acid (0.1 mM; 4 days soak). Heavy atom search, phasing, and density modification were performed with AutoSol [41] in the PHENIX suite [42]. Automatic model building was performed with Buccaneer [43]. The model obtained was used as a search template for molecular replacement with Phaser [44] and high-resolution native X-ray data. The model was further refined through an iterative combination of refinement with Refmac5 [45] and manual model building in COOT [46]. The quality of the final model was validated with MolProbity outliers. Table EV2 lists relevant model statistics.

Isolation of Methanococcus maripaludis archaella

Archaella were extracted from Me. maripaludis cells using detergent OP-10, precipitated using PEG and banded on a KBr gradient as previously described [5].

Model building and structure refinement of Methanococcus maripaludis filament

First, the crystal structure of FlaB1 of M. jannaschii was fitted as a rigid body into the cryo-EM density map. The initial model was further modified accordingly to the sequence of Me. maripaludis archaellins, and missing parts were built manually in COOT [46]. Finally, the model was refined at 4.0 Å resolution with phenix.real_space_refine [48].

Bioinformatics

The alignment figure was prepared with Crystal Omega [49] and ESPript3 [50]. Estimation of evolutionary conservation of amino acids was done using ConSurf [21]. CheckMyMetal (CMM) server [51,52] was used for the validation of metal-binding site.

Cryo-EM

Sample preparation

3.5 µl purified filament solution was applied to holey carbon grids (Quantifoil 1.2/1.3) and vitrified on a Vitrobot Mark IV (Thermo Fisher) by plunge-freezing into liquid ethane cooled to liquid nitrogen (LN2) temperature. Frozen grids were transferred into the cryo-TEM for observation at LN2 temperature.

Data collection

Sample grids were observed on a Titan Krios TEM equipped with an energy filter (Gatan Quantum 968, slit width set to 20 eV) in EFTEM mode using nanoprobe illumination (50 µm C₂ aperture, 100 µm objective aperture). The microscope and camera were controlled using LEGINON software. In order to accelerate acquisition, image data collection was set up to take exposures of three target areas per hole using image shift from the center of the hole. Images were recorded as movies (3.5 el/superpixel/s, 250 ms/frame, 12-s exposure, underfocus of 1.5–2.5 µm) on a K2 direct electron detector in super-resolution electron counting mode (dose rate 2 el/A²/s, total dose 98 el/exposure). Movies were saved as uncorrected raw data in MRC file format.

Movie processing

Movies were corrected for gain variation and magnification anisotropy of the projector lens (≈ 1.4% as measured with Au/Pt cross-grating and magdistcor program [53]) and binned by a factor of 2 using Fourier cropping. These frame stacks were then aligned and summed in motioncor2 [54]. All frames were included by applying dose weighting.

Image processing

2,000 filaments were manually selected using e2helixboxer.py [55]. Filament length and selected straight segments were short (average 147 nm, standard deviation 127 nm). All subsequent operations
were carried out using the SPRING helical image processing package [26]. Filaments were segmented at a step size of 75 Å and extracted as boxed particles with dimension 700 × 700 Å. Principal component analysis of threefold binned boxed particle images using the SPARX K-means algorithm was followed by classification into 100 classes. Class-sum images revealed a helix diameter of 105 Å and structural detail at improved SNR. Some of these class averages were used for a correlation-based symmetry search (program segclassreconstruct) near the solution found by analysis of the helical diffraction pattern, but the helical parameter could not be optimized further. Segment alignment and 3D reconstruction were performed with SPRING’s segrefine3d program using the gold standard procedure. The final figure-of-merit-weighted electron potential map [56] was sharpened with a B-factor of −200, masked with a cylindrical mask and low-pass filtered at 3.7 Å resolution. Fourier shell correlation was determined from the two independently refined half sets (so-called gold standard FSC). FSC figure was created with gnuplot. Maps of iso-contoured potential maps were visualized in UCSF Chimera.

**Difference map and calculation of SNR**

The difference map to highlight posttranslational modification sites was produced with diffmap.exe (Grigorieff Lab) by scaling and subtracting a low-pass filtered synthetic potential map based on the refined model coordinates from the normalized cryo-EM reconstruction. The synthetic map, which was also used to calculate the Fourier cross-correlation, was created using SFALL and FFT [40]. Signal-to-noise ratio (SNR) in the difference map near the modified side chains was calculated using the following equation:

$$\text{SNR} = \frac{\sum (\Delta I)^2}{N \cdot \sigma_{\text{Noise}}}$$

where $\Delta I$ are the pixel intensities of the difference map inside a spherical mask with radius 4 Å centered at the amide group of the asparagine side chain connected to putative modifications, $N$ is the number of voxels inside the mask, and $\sigma_{\text{Noise}}$ is the average variance of the difference map outside the masked volume (in a location that contains some solvent noise). The spherical masks were created with MAMA [40], and the masked map density was cut out using MAVE [40]. Voxel intensities were read out with the SPIDER LI command [57]. Noise variance, sum of the squared intensities, and SNR within these masked volumes were then calculated by shell script.

**Mass spectrometry**

**Sample dilution**

Clean sample was filtered after dialysis using Amicon cutoff 30 kDa to remove aggregated proteins. Flow-through was diluted 1/10 with 0.1% formic acid in water and final volume 40 µl.

**Tandem LC-MS/MS analysis**

Diluted sample was analyzed using a Thermo Scientific Q-Exactive Plus Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with HPLC (Dionex Ultimate 3000 nanoRSLC), autosampler (HTC PAL, CTC Analytics), and nano-electrospray ion source. Injection volume was 5 µl, separated on a Zorbax 300SB C18 column (0.3 × 150 mm; Agilent, Agilent Technologies, Waldbronn, Germany) at 40°C. A 1-h solvent B (0.1% formic acid in acetonitrile) gradient in solvent A (0.1% formic acid in water) was employed (1% B to 32% B in 45 min, 32% B to 45% B in 15 min, with a final wash at 75% B for 5 min and reequilibration at 1% B for 10 min). A flow rate of 3.5 µl min⁻¹ was used for peptide separation. Temperature of the heated capillary was 300°C, and 1.9 kV spray voltage was applied to all samples. Mass spectrometer settings were as follows: full MS scan range 350–1,500 m/z, with a mass resolution of 70,000, 30 µs scan time, and AGC set to 1 × 10⁶ ions, and fragmentation MS2 of the 20 most intense ions.

**Protein identification**

Protein identification employed Proteome Discoverer software v2.1, and Mascot as a search engine. A database downloaded from UniProtKB. *Me. maripaludis* (1,142 reviewed and 7,965 unreviewed proteins) was used. Search parameters were as follows: no enzyme, with precursor and fragment mass tolerance set to 10 ppm and 0.02 Da, respectively. Asparagine and glutamine deamidation, cysteine Ca, Fe, and Mg were set as variable modifications on aspartic acid and glutamic acid. Results were filtered using a false discovery rate of < 1% as a cutoff threshold, determined with the percolator algorithm in Proteome Discoverer software.

**Modified peptide screening**

Identified peptides with metal cation modification were subjected to manual examination using the following criteria. Each modified peptide was confirmed to have a chromatographic peak with the same retention time as its unmodified partner; only search ranks “one peptide” and “minimum Mascot score of 25” were accepted.

**Structure and map deposition, and accession codes**

Model coordinate geometry and structure factors of *M. jannaschii* FlaB1 have been deposited in the Protein Data Bank under accession 5Z1L (statistics in Table EV2). The cryo-EM reconstruction was uploaded to the EMDB data bank with accession number EMD-6876. Map and data collection statistics are summarized in Table EV2.

**Acknowledgements**

We gratefully acknowledge Young-Ho Yoon, Hideyuki Matsunami, and Seiya Kitanobo for their support with X-ray data collection, and S.D. Aird for proof-reading the manuscript. We thank N. Grigorieff for advice on SNR calculation. The synchrotron data collection was conducted by V.A.M. on beamline BL13-1 at SPring-8, Japan (Proposal No. 2014B1341).
purified the M.m. archaellum sample. S-IA helped with helical diffraction analysis and provided general advice. MW collected and analyzed cryo-EM data together with SS. MW was responsible for overall project organization and planning. VAM and MW wrote the paper with the input of all authors.

Conflict of interest
The authors declare that they have no conflict of interest.

References


License: This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.