New insights into the architecture and dynamics of archaella

Lavinia Gambelli[1,2]♣, Michail N. Isupov[3]♣, Rebecca Conners[1,4], Mathew McLaren[1,4], Annett Bellack[5], Vicki Gold[1,4], Reinhard Rachel[5], Bertram Daum[1,4]*

[1] Living Systems Institute, University of Exeter, Exeter, EX4 4QD, United Kingdom.

[2] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, United Kingdom.

[3] Henry Wellcome Building for Biocatalysis, Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QD, United Kingdom.

[4] College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4QD, United Kingdom.

[5] Institute of Microbiology and Archaea Centre, University of Regensburg, Regensburg, 93053, Germany.

♣ These authors contributed equally to this work.

* Corresponding author. Email: b.daum2@exeter.ac.uk

Keywords: *Methanocaldococcus villosus*, archaea, archaellum, electron cryo-microscopy, single particle analysis, helical reconstruction, type IV pili.

Abstract

Archaea swim by means of a unique molecular machine called the archaellum. The archaellum consists of an ATP-powered intracellular motor that drives the rotation of an extracellular filament, allowing the cell to rapidly propel itself through liquid media.

The archaellum filament comprises multiple copies of helically organised subunits named archaellins. While in many species several archaellin homologs are encoded in the same operon, structural studies conducted to date have suggested that archaella consist of only one protein species. Thus, the role of the remaining archaellin genes remains elusive.

Here we present the structure of the *Methanocaldococcus villosus* archaellum filament at 3.08 Å resolution. We find that the filament is composed of two alternating archaellins - ArlB1 and ArlB2, suggesting that the architecture and assembly of archaella is more complex than previously thought. Moreover, we identify two major structural elements that enable the archaellum filament to move.

Our findings provide new insights into archaeal motility and challenge the current view on the archaellum architecture and assembly.

Introduction

Archaea are ubiquitous microorganisms that successfully colonise diverse environments partially due to their ability to swim by means of archaella [1-3]. The archaellum is a propulsive nanomachine consisting of an intracellular motor that drives the rotation of an extracellular filament [4]. A clockwise rotation of the archaellum moves the cell forward, whereas a counterclockwise rotation moves the cell backward [5]. While archaella, like bacterial flagella, mediate swimming, the two systems are not homologous, suggesting their independent evolution. Instead, archaella are structurally homologous to filaments of the type IV filament (TFF) superfamily. The TFF superfamily includes a vast range of filamentous nanomachines, such as bacterial type IV pili (T4P) [6], associated with diverse functions [7]. Within the TFF superfamily, archaella are unique in providing swimming motion by means of a rotary propeller. Other functions of TFF filaments in archaea include surface adhesion [8-11], cell-cell contacts, biofilm formation [8,12] and electrical conductance [13].

The archaellum machinery consists of a membrane-embedded motor complex and a ~10 μm long filament, which together are formed by ~10 component subunits, depending on the species [14]. The filament is a helical array of ArlA and/or ArlB archaellins [15] that assemble proximal to the cell surface [16]. Filament assembly and rotation is an ATP-dependent process

driven by the motor complex, consisting of the platform protein ArlJ, the hexameric AAA + ATPase ArlI, and a putative regulator ArlH [6,17-19]. The motor is surrounded by a cytosolic ring of ArlX proteins (in Crenarchaeota) [20] or ArlC, D/E (in Euryarchaeota) [21,22]. In the periplasm, the motor complex is anchored to the S-layer by the ArlF and ArlG stator proteins [23,24].

Most genes encoding the archaellum machinery components are organised in one *arl* operon. The first genes in this operon (*arlA* or *arlB*) encode for the archaellin subunits that make up the archaellum filament. The number of archaellin genes is species-specific, usually varying from one to seven [25,26]. When multiple archaellins are encoded, several or all can be transcribed at the same time [27,28]. Conventionally, ArlA and ArlB archaellins are divided into major and minor structural components of the filament. Archaellins that are thought to form the bulk of the filament are referred to as major. Biochemical and genetic experiments suggest that in *Methanococcus voltae* [29], *Methanococcus maripaludis* [30], *Methanothermococcus thermolithotrophicus* [31], *Halobacterium salinarum* [32], and *Halorubrum lacusprofundi* DL18 [33] the archaellum filament consists of two major archaellins, hypothesised to be equally distributed along the filament. Nevertheless, the three high-resolution structures of archaella that are available to date show homopolymeric filaments consisting of only single (major) archaellin [21,34,35].

The roles of minor archaellins mostly remain enigmatic. Genetics and molecular biology experiments suggest that in *M. voltae* [36], *M. maripaludis* [30] and *H. salinarum* [37], minor archaellins form a region at the base of the archaellum filament that is reminiscent of the flagellar hook. However, not all archaea show this feature, including those that encode only one archaellin.

Here we employed electron cryo-microscopy (cryoEM) to solve the structure of the archaellum filament from *Methanocaldococcus villosus*, a hyperthermophilic archaeon [11] and one of the fastest swimming organisms known [38]. We report the first structure of a heteropolymeric archaeal filament composed of two alternating subunits, ArlB1 and ArlB2 and provide first insights into the dynamics of archaella. Our results provide a shift in our understanding of how archaella assemble and function.

Materials and Methods

Cell culture

Anaerobic media were prepared according to the technique described by Balch and Wolfe [39], and the enrichment and culture method was published by Bellack *et al.* [11]. *M. villosus*

was originally isolated from hot sediments sampled during the RV Poseidon cruise in 1997 to the Kolbeinsey Ridge [40]. Cells were grown in a 50 l bioreactor (Bioengeneering, Wald, Switzerland) using MMC medium designed for the fermentation process. The bioreactor was inoculated with 2.5 ml of *M. villosus* cells and supplied continuously with 2.5 bar (absolute pressure) H_2/CO_2 (80:20, v/v) as gas phase. Temperature was kept constant at 80°C and the pH at 6.5. Cell densities of grown cultures were determined by direct cell counting using a Thoma counting chamber (depth 0.02 mm). Standard reagents were purchased from VWR or Merck. The medium consisted of (per l in distilled water): NaCl (430.0 mM), $NaHCO_3$ (10.0 mM), $MgCl_2 \times 6H_2O$ (38.0 mM), KCl (3.6 mM), NH_4Cl (22.1 mM), $CaCl_2 \times 2H_2O$ (2.5 mM), K_2HPO_4 (0.3 mM), KH_2PO_4 (4.1 mM), $(NH_4)2Fe(SO_4)_2 \times 6H_2O$ (30.6 µM), $(NH_4)2Ni(SO_4)_2 \times 6H_2O$ (43.5 µM), $Na_2MoO_4 \times 2H_2O$ (10.1 µm), $Na_2WO_4 \times 2H_2O$ (9.9 µM), Na_2SeO_4 (42.3 µM), 10-fold trace mineral solution (1.00 ml), $Na_2S \times 7\text{-}9H_2O$ (0.40 g). The pH was adjusted to 6.5 with 50 ml H_2SO_4 (diluted 1+1 in MilliQ). The trace mineral solution 10-fold [41] contained per litre: $MgSO_4 \times 7H_2O$ (121.7 mM), NaCl (171.1 mM), $MnSO_4 \times H_2O$ (29.6 mM), $(NH_4)_2Ni(SO_4)_2 \times 6H_2O$ (7.1 mM), $CoSO_4 \times 7H_2O$ (6.4 mM), $ZnSO_4 \times 7H_2O$ (6.3 mM), $FeSO_4 \times 7H_2O$ (3.6 mM), $CaCl_2 \times 2H_2O$ (6.8 mM), $AlK (SO_4)_2 \times 12H_2O$ (0.4 mM), $CuSO_4 \times 5H_2O$ (0.4 mM), H_3BO_3 (1.6 mM), $Na_2MoO_4 \times 2H_2O$ (0.4 mM), $Na_2WO_4 \times 2H_2O$ (0.3 mM), Na_2SeO_4 (0.5 mM). Trace element solution was titrated at pH 1.0 with 1 M H_2SO_4 before dissolving chemicals, and it was filter sterilized and stored in the dark at 4 °C.

Preparation of archaella

The archaella preparation protocol was adapted from Kalmokoff *et al.* [29] and Näther *et al.* [10]. The complete bioreactor was used, and cells were harvested at early stationary phase (7,000–8,000 × g) and concentrated (3,500 × g, 30 min, 4 °C; Sorvall RC 5C plus, rotor GS3). A total of 17.5 g of cells (wet weight) were obtained. Then archaella were sheared (Ultraturrax T25; 1 min at 13,000 rpm and 10 sec at 22,000 rpm) and cell debris was removed by centrifugation (34,500 × g, 4 °C; Sorvall RC 5C plus, rotor SS34) for 20 min. Archaella were pelleted from the supernatant by ultracentrifugation (60,000 × g, 90 min, 4 °C; Beckman Optima LE-80K, rotor 70Ti), resuspended in 150 µl 0.1 M HEPES/NAOH pH 7.0 and purified for 48 h (250,000 × g, 4 °C; Beckman Optima LE-80K, rotor SW60-Ti) using a CsCl gradient (0.45 g/ml; Biomol). Fractions were taken by puncturing the ultracentrifuge tubes with sterile syringes, dialyzed against aerobic ½ SME/5 mM HEPES/NaOH pH 7.0, and analysed by SDS-PAGE and transmission electron microscopy [11] to identify the archaella-containing band. After addition of sodium azide to a final concentration of 0.001% (w/v), samples were stored at 4 °C.

CryoEM of archaella and helical processing

Prior to freezing, a 3 µl drop of undiluted suspension containing purified archaella was applied to glow-discharged 300 mesh copper R2/2 Quantifoil grids. The grids were blotted with 597 Whatman filter paper for 4 sec, using -1 blot force, in 95 % relative humidity, at 21 °C and plunge-frozen in liquid ethane using a Mark IV Vitrobot (FEI). Dose-fractionated movies were collected in linear mode using a 300 kV Titan Krios equipped with a Falcon III detector (Thermo Fisher Scientific) (suppl. table 1). 39-fraction movies were recorded using EPU 1.5 (Thermo Fisher Scientific) with an exposure time of 1 sec and a total electron dose of 37 e$^-$/Å^2 at a defocus range of -2.3 to -1.1 µm (0.3 µm increment). A total of 2,759 movies were collected and processed using the Relion 3.1 pipeline [42] and finalised using cryoSPARC 3.1.0 [43]. Briefly, the movies were motion corrected using MotionCorr 2 [44] and CTF-corrected using CTF-find 4 [45]. Helical segments were manually picked using Helixboxer from the EMAN-2 package [46]. The coordinates of the picked particles were imported into Relion 3.1, where particle extraction, 2D classification, 3D classification, 3D refinement, B-factor sharpening, CTF-refinement and Bayesian polishing were performed. Each filament was subdivided into 256 x 256 pixel size boxes at an offset of 6 Å, resulting in 929,165 segments upon particle extraction. The dataset was then 3D classified (8 classes) and the best class (containing 399,178 segments) was selected for 3D refinement using the helical parameters published for the *P. furiosus* archaeallum [21]. The refined consensus map was improved using CTF-refinement and Bayesian polishing in Relion 3.1. The shiny particles obtained from the Bayesian polishing were further refined using the optimised helical parameters 5.57 Å rise and 108 ° twist after helical refinement. The resulting map had a resolution of 3.29 Å. The particles of the final Relion 3.1 helical refinement were imported into cryoSPARC 3.1.0 and a new refinement was performed using the helical refinement (BETA) algorithm without imposing helical parameters. This resulted in a reconstruction at a resolution of 3.28 Å by gold-standard FSC in which two distinct subunits could be distinguished. To corroborate these results, another round of 3D refinement with no helical symmetry was performed in Relion 3.1, which provided a comparable result. The two subunits alternate along the left-handed 3-start helices as shown on the archaellum net diagram (fig. 4h, suppl. fig. 4a,b). An attempt to perform a helical refinement with (n+2) helical parameter 11.14 Å rise and 216 ° twist resulted in a filament reconstruction with lower resolution and poorly resolved features. Further inspection of the net diagram allowed to establish the minimal transformation of the *M. villosus* archaellum in which each monomer superimposes onto its equivalent, and which proved to be n+6. A final helical refinement was performed in cryoSPARC 3.1.0, this time imposing a (n+6) helical parameters 33.4 Å rise and

-71.8° twist (see results below). This produced a final map of 3.08 Å resolution by gold-standard FSC 0.143. DeepEMhancer was used for final map sharpening [47].

Model building and validation

Initially, the *P. furiosus* ArlB0 archaellin structure (PDB ID: 5O4U) was positioned into the Relion 3.1 cryoEM map, refined using REFMAC5 [48] and rebuilt to match the *M. villosus* ArlB2 sequence using *Coot* [49]. Both candidate sequences (ArlB1 and ArlB3) were attempted to be fitted by Buccaneer [50] into the density of the second archaellin found in the higher resolution cryoSPARC 3.1.0 cryoEM map. The sequence match to the cryoEM map features and glycosylation pattern unambiguously identified ArlB1 as the second protein forming the archaellum. Both ArlB1 and ArlB2 atomic models were then validated using Isolde [51]. Two full turns of the archaellum filament were built in UCSF Chimera [52] using ArlB1-ArlB2 3-start dimers. Optimised helical parameters with the position of each monomer was rigid body refined in *Coot* to adjust for local variations in the non-helically refined map. A script containing a number of CCP4 [53] programs was prepared to propagate any modification to a structure of a single monomer into a full filament. The resulting model was refined using REFMAC5 in CCP-EM interface [54]. Glycans were modelled in *Coot* using the sugar unit identity and connectivity reported for the glycan structure of *M. thermolithotrophicus*. This structure was recently reported as α-d-glycero-d-manno-Hep3OMe6OMe-(1–3)-[α-GalNAcA3OMe-(1–2)-]-β-Man-(1–4)-[β-GalA3OMe4OAc6CMe-(1–4)-α-GalA-(1–2)-]-α-GalAN-(1–3)-β-GalNAc-Asn [31]. Restraints dictionaries for non-standard sugars were prepared using JLIGAND [55]. The calcium ion in ArlB1 and 2 was modelled on the basis of the metal coordination and distances observed in our structure and agreed with the metal assignment in [35].

3D variability analysis and molecular flexibility

3D variability analysis (3DVA) [56] was performed in cryoSPARC 3.1.0 using the 399,178 helical segments from the cryoSPARC 3.1.0 helical refinement. First, the symmetry expansion step was run. Then, the particles stack created by the symmetry expansion was used for 3DVA. Different lowpass filter values were tested (3.5, 5, 8 and 10 Å) in simple mode with 20 intermediate frames and with 10 variability components as output. All components showed the same type of motion.

In order to investigate the filament's flexibility at the molecular level, we selected a 3DVA component obtained at 3.5 Å low-pass filter and used the first frame (number 0) and

last frame (number 19) of the component corresponding to two cryoEM maps at the two extremes of the filament motion range. The filament's atomic model was positioned into the two cryoEM maps using phased molecular replacement in MOLREP [57] and refined using Phenix [58], thereby producing two atomic models fitting the two cryoEM maps. The two atomic models were finally morphed using the morphing function in UCSF Chimera to display the full conformational space of the filament (suppl. movie 4) , in particular the tail domains (fig. 5a; suppl. movie 5). To show the local flexibility of the head domains, one ArlB1 and one ArlB2 subunits were isolated from the two atomic models fitting the two cryoEM maps at frame 0 and frame 19 and r.m.s.d-coloured vectors were calculated linking the Cα of the head domains (fig. 5c). To investigate the flexibility of the head domains along the filament, two homopolymeric strands of ArlB1 and ArlB2 were isolated from the atomic model fitting the cryoEM map at frame 0. Then for each strand the tails of its constituting subunits forming a full filament turn were aligned (fig. 5b; suppl. movie 6, 7).

STEM tomography

Cells were cryo-immobilized using high-pressure freezing (Leica HPM 100, Munich), then freeze-substituted in Ethanol containing 0.5% glutardialdehyde, 1% formaldehyde, 0.5% uranyl acetate, 5% water, and finally embedded into Spurr resin. Semithin sections with a thickness of 800-900 nm were cut using a diamond knife (Diatome) on an ultramicrotome (Leica UC6). Sections were mounted on formvar-coated grids (100 mesh Cu grids; Plano, Wetzlar). For STEM tomography at 200 kV [59], a 3 nm carbon layer (Cressington turbo 208 Carbon) was evaporated onto one side of the sections. Fiducials (proteinA-gold, 15 nm; University Medical Center, Utrecht, NL) were added onto both sides of the sections, to aid the alignment and reconstruction process. STEM tomography was done on a JEM-2100F (JEOL GmbH, Freising, Germany) using an electron beam with a small semi-convergence angle, essentially as described [59]. Tiltseries (STEM bright field) were recorded from +66 deg to -66 deg, using the EM-TOOLS software (TVIPS, Gauting, Germany). Reconstruction of 3D volumes was done in IMOD (Univ. of Colorado at Boulder, USA) using the weighted back projection, incl. a SIRT-like filter equivalent to 15 iterations.

Sequence analyses

The evolutionary conservation of amino acids positions in AlrB1 and ArlB2 (fig. 5d) and bacterial T4P (suppl. fig. 14c) was estimated using the ConSurf server [60]. The sequence alignment and secondary structure assignment of *M. villosus* ArlB1 and ArlB2 were performed

with BLASTp and visualised by ESPript3 [61] (suppl. fig. 7). The archaellar operons (suppl. fig. 15) were visualised using ARTEMIS [62] and the automatic genome annotations were downloaded from NCBI. The sequence comparison and alignment between archaellins and archaeal T4P (suppl. fig. 14a,b), and *M. maripaludis* ArlB3, *M. voltae* ArlB3 and *M. villosus* ArlB1, 2, 3 archaellins (suppl. fig. 16) were performed with BLASTp and the Praline server [63].

Structure analysis and presentation

The structure of the *M. villosus* archaellum filament was visualised using UCSF Chimera, Chimera-X [64], *Coot* and CCP4mg [65]. The topology diagrams were prepared using Pro-origami [66] and TopDraw [67]. Interfaces between different ArlB1 and ArlB2 monomers in the archaellum were analysed using the PISA software [68]. The Molprobity score in suppl. table 1 was calculated using [69].

Data deposition

The archaellum atomic coordinates and electron density map were deposited in the Protein Data Bank (https://www.rcsb.org/) with accession number 7OFQ and in the EM DataResource (https://www.emdataresource.org/) with the accession number EMDB–12875.

Results

High resolution cryoEM and helical reconstruction of *M. villosus* archaella

The structure of the *M. villosus* archaellum was determined from cryoEM movies of isolated filaments using Relion 3.1 [42] and cryoSPARC 3.1.0 [43] (suppl. fig. 1, 2). Helical parameters from previously published archaella structures were applied and optimised by helical search to a rise of 5.6 Å and a twist of 108°. This approach yielded a 3.28 Å resolution map in which the filament appeared to consist of one archaellin homologue (suppl. fig. 3a). Surprisingly, further refinement of the cryoEM map without imposing helical symmetry produced two distinct and alternating protein densities (fig.1; suppl. movie 1; suppl. fig. 3b,c). Careful analysis of the filament's heteropolymeric structure revealed that the helical symmetry did not comply with the initially applied rise and twist. Instead, the minimal transformation that superimposes each archaellin subunit onto one of the same type is n+6 (6 subunits along the filament long axis) (suppl. fig. 4e). Recalculating the helical symmetry of the archaellum based on this observation resulted in a helical rise of 33.4 Å (5.6 Å *6=33.6 Å, refined to 33.4

Å) and a twist of -71.8° (108°*6=648=720-72°). Refining our map with these parameters improved the resolution to 3.08 Å (fig. 1; suppl. fig. 5).

The *M. villosus* archaellum consists of the subunits ArlB1 and ArlB2

The archaellum operon of *M. villosus* encodes three archaellin homologs (ArlB1, 2 and 3). To investigate which two constitute the filament, we built atomic models based on their primary structures for both distinct protein densities. ArlB1 and 2 could each be modelled unambiguously into alternating positions in the map (fig. 2; suppl. movie 2), guided by aromatic side chains and glycosylation sites (fig. 3).The amino acid sequence of ArlB3 did not match the map density features and was too long to fit into either position.

The proteinaceous part of the *M. villosus* archaellum filament has a ~9.8 nm diameter. ArlB1 and ArlB2 alternate throughout the filament, following left-handed helical strands. The archaellum is composed of three of these helical strands, which by convention are defined as 3-start helices (fig. 1d-g; fig. 2a-c; fig. 4a,b; suppl. fig. 6; suppl. movie 2).

ArlB1 and 2 are homologous by structure and sequence (fig. 2d-f; suppl. fig. 7, 8). Their genes encode proteins of 221 and 225 amino acids respectively which share 56.5 % sequence identity (75.9 % similarity). As is typical for other members of the TFF superfamily, ArlB1 and 2 consist of an N-terminal hydrophobic α-helical "tail" and a globular, C-terminal β-strand rich "head" domain. In the assembled archaellum, the tails constitute the core of the filament, whereas the head domains are exposed to the periphery (fig. 2a-c). The first 12 amino acids are not present in the mature ArlB1 and 2 proteins due to N-terminal processing by a class-III signal peptidase prior to their insertion into the filament [70]. While the α-helical tail domains are identical in ArlB1 and 2 (I_{14}-A_{59}), the head domains show differences with respect to sequence and structure (fig. 2f; suppl. fig. 7, 8). ArlB1 consists of 12 β-strands and three α-helices, whereas ArlB2 has 13 β-strands and two α-helices. The most striking difference is a glycosylated outward-facing loop (K_{131}-A_{137}) in ArlB2, which is absent in ArlB1 (fig. 2f).

N-glycosylation and metal-binding sites

Refinement of the helical parameters was crucial to clearly resolve the glycosylation pattern of the archaellum filament. Glycan densities were identified as dead-end protrusions from the peptide backbone that were larger than side chains and co-localised with the canonical N-glycosylation consensus sequon (N-X-S/T) (fig. 3a-e). Including the density occupied by the glycan moieties, the diameter of the archaellum is ~110 Å. All glycosylation sites are located

on loops or sites of alternating secondary structure motifs and are equally distributed along the filament's length.

While the head domain of ArlB1 harbours three glycosylation sites (N_{112}, N_{133} and N_{176}) (fig. 3b,d), ArlB2 features six (N_{70}, N_{113}, N_{118}, N_{133}, N_{134} and N_{180}) (fig. 3c,e). Interestingly, two of these glycosylated asparagine residues (N_{133} and N_{134}) are consecutive and located in the NNTT motif of the ArlB2 glycosylation loop (K_{131}-A_{137}).

Since no information on sugar connectivity and identity is available for N-glycans of *M. villosus*, we modelled the branched heptameric glycan from the related archaeal species *Methanocaldococcus thermolithotrophicus* [31] in order to aid model building and refinement. We were able to fit four of the seven sugar units into our map, no further residues were resolved. The fitted glycans provide a detailed model of the location and distribution of glycans on the surface of the archaellum (fig. 2a-c; fig. 3a-e; suppl. fig. 9).

As featured by other archaellins [35], ArlB1 and 2 also contain a metal-binding site located at the periphery of the head domain, in a groove between β6 and β7 in ArlB1 and β7 and β8 in ArlB2 (fig. 3f). In both archaellins, the metal ion is coordinated by five conserved residues, which differ slightly in their position along the backbone (D_{154}, D_{156}, S_{158}, N_{166}, D_{169} in ArlB1 and D_{157}, D_{159}, S_{160}, N_{170}, D_{173} in ArlB2). The metal was modelled as a calcium ion. However, due to prevalence of magnesium over calcium in our growth culture, a mixed population of calcium and magnesium ions cannot be excluded. The metal binding residues are highly conserved between ArlB1, ArlB2 and the *M. jannaschii* archaellin (suppl. fig. 10), for which the site was resolved by X-ray crystallography [35].

Heteropolymeric assembly of ArlB1 and ArlB2 introduces a screw asymmetry in the archaellum

To understand the assembly of ArlB1 and 2 in the archaellum structure, we analysed their organisation and contacts within the filament. Consistent with published structures of homopolymeric archaella [18,21,34] we find that ArlB1 and 2 undergo hydrophobic interactions via their N-terminal tails in the filament's core. In particular, each tail interacts with neighbours at positions n+3, n-3, n+7, n-7, n-1, n+1, n-4 and n+4 (suppl. fig. 11).

In the outer sheath of the archaellum the head domains of ArlB1 and 2 interact via hydrogen bonds and salt bridges. These occur between alternating ArlB1 and 2 subunits along left-handed 3-start helical strands (fig. 4a-c; suppl. table 2) and along right-handed 7-start strands (fig. 4d-f). The alternating sequence in 3-start direction appears to be favoured by high shape and charge complementarity between ArlB1-2 and ArlB2-1 heterodimers. 13 hydrogen

bonds and 3 salt bridges with a ΔG_{diss} of 20.4 kcal/mol are formed at the ArlB1-2 contact zone and 8 hydrogen bonds and 2 salt bridges with a ΔG_{diss} of 16.2 kcal/mol generate the ArlB2-1 interface (fig. 4c; suppl. table 2). In contrast, poor charge and shape complementarity would not allow the formation of ArlB1-1 and ArlB2-2 in the 3-start direction. Interestingly, the sequence of alternating subunits is not always in sync in adjacent 3-start helices. One of the helical strands is shifted by one subunit, moving it out of register (fig. 4b). This results in an anisotropic distribution of ArlB1 and 2 throughout the archaellum, meaning that ArlB1 and ArlB2 neither alternate, nor form perfect homopolymeric strands in any direction but 3-start.

Along the left-handed 7-start helical strands the two subunits undergo all four possible contacts (ArlB1-2, ArlB2-1, ArlB1-1, ArlB2-2) (fig. 4d-f). The interactions in the right-handed 7-start direction are by an order of magnitude weaker than those in the 3-start direction (ΔG_{diss} of 1.7-4.7 kcal/mol), as they are stabilised by only 2-5 salt bridges per interunit interface and limited hydrophobic contact between the tails (suppl. table 2).

Strikingly, when tracing subunits of the same type (ArlB1 or ArlB2 only) throughout the filament, the archaellum appears to be defined by homopolymeric right-handed pseudo-strands with broken helical symmetry (fig. 4g,h). Here, the symmetry of the pseudo-strands is disrupted at the interface between every third and fourth subunit, giving the pseudo-strands a "stepped" appearance, in contrast to a hypothetical isotropic filament (with n+2 geometry) (suppl. fig. 4c,d,g).

Flexibility of the archaellum filament

The archaellum rotates prompted by an ATP-driven intracellular motor that generates torque. To propagate the motion along its length, the filament transitions from a curved filament in resting cells [21] (suppl. fig. 12) into a rotating superhelix when cells swim [71]. However, due to the nature of helical image processing of cryoEM data, all available structures of archaella suggest idealised helical filaments that are infinitely straight.

In order to obtain a truer representation of the structure of the archaellum and to probe the molecular basis for its flexibility, we performed 3D variability analysis in cryoSPARC, which estimates molecular motion from micrographs [56]. This analysis resulted in a series of 20 maps, displaying a distinctly bent conformation of the archaellum (see suppl. movie 3 for a morphing between these conformations). Refining our atomic model based on individual conformations, we visualised the dynamic properties of the archaellum at the molecular level (fig. 5; suppl. movies 4-7) .

Our analysis reveals that as the filament bends, the subunits are compressed at its concave side and stretched apart along its convex side (suppl. movies 3, 4). To allow for this flexibility, the tails in the backbone rotate around and slide along the filament axis (fig. 5a2; suppl. movie 5), whereas the head domains swing diagonally to the filament axis (fig. 5c; suppl. fig. 13). Along pseudo-helical ArlB1 and ArlB2 strands (fig. 5b1), the head domains of both archaellins are flexible, and their trajectories differ depending on the subunit location within the filament (fig. 5b2-5). Each head domain is free to swing in all directions within the cylinder surface of the outer sheath and with respect to the tail domain. An example of the head trajectories of two homopolymeric strands is shown in fig. 5b3, b5; suppl. movies 6, 7. Notably, we do not see substantial flexibility or conformational changes within the head or the tail domains themselves, indicating that they move relative to each other as rigid bodies. Moreover, no significant difference in flexibility between ArlB1 and ArlB2 was observed.

The flexibility of each archaellin is enabled by a hinge of three residues (Ala_{59}, Ser_{60}, Gly_{61}) linking the head and tail of each subunit. This hinge is highly conserved in ArlB1 and 2 sequences within the archaeal kingdom (fig. 5d; suppl. fig. 14a). Interestingly, the sequence of this region is partially conserved, albeit different in archaeal and bacterial T4P (suppl. fig. 14b,c).

Discussion

Many archaea encode for multiple archaellins, which likely originated from gene duplication events [72]. Studies employing genetics, biochemistry and molecular biology have shown that while some archaella contain a single major archaellin, others are assembled from two homologues [27,30-33]. The three structures of archaellar filaments that have been published so far suggest homopolymers [21,34,35], and until now it was unknown how two or more distinct proteins can assemble into a single archaellum filament.

Here, we present the first structure of a heteropolymeric archaellum. The filament contains two archaellin homologues (ArlB1 and ArlB2), which became distinguishable when the cryoEM map was reconstructed without imposing helical symmetry. This finding highlights the importance for relaxing the symmetry during processing of any helical data, as structural heterogeneities may otherwise be overlooked.

Impact of heteropolymeric assembly on archaellar function

The evolutionary advantages of a heteropolymeric over homopolymeric archaella are not fully understood. However, accumulating evidence suggests that these include increased cellular motility and filament stability. For example, while individual *arl*B1[-] and *arl*B2[-] knockout mutants in *H. lacusprofundi* DL18 still assemble archaella, they are less salt-resistant than wildtype [33]. In addition, micrographs of negatively stained filaments suggested that ArlB1[-] archaella are more flexible and ArlB2[-] archaella are stiffer compared to wild type. This is in line with the considerably different binding energies that we have calculated for the *M. villosus* archaellum, where ArlB2-2 interfaces in 7-start direction are significantly more stable than the corresponding ArlB1-1 contact (suppl. table 2). Provided that *M. villosus* could form homopolymeric archaella, an ArlB2-only filament would thus likely be more stable but stiffer than an ArlB1 homopolymer. Notably, knocking out one of the two archaellar subunits in *H. salinarum* results in reduced motility [32] and ablating either of the two archaellins in *M. maripaludis* abolishes archaella and swimming motility entirely [30]. Furthermore, in *Haloarcula marismortui* the two archaellins ArlA2 and ArlB are ecoparalogues, assembling different archaella depending on the level of salinity in the environment [73]. This suggests that the ability to form heteropolymeric archaella can, at least in some species, add to the adaptability of the cell to varying habitat conditions.

In the heteropolymeric archaellum of *M. villosus*, ArB1 and 2 are not isotopically organised. We propose that the resulting screw axis asymmetry, together with the helical variation of rigidity along the 3-start direction, enhances the supercoiling propensity of the filament. This would explain how the archaellum forms the observed stable right-handed supercoil independent of the direction of gyration [71], as a prerequisite for cellular propulsion in opposite directions.

The published structures of T4P in both archaea and bacteria report homopolymeric filaments constituted by one major pilin [74-78]. In one example however, two different homopolymeric T4P filaments are assembled simultaneously and by the same machinery [78]. Only one study published to date provides evidence of a heteropolymeric T4P in a Gram-positive bacterium, but the arrangement of the two subunits within the filament is speculative [79]. These findings suggest that the specialised structures of pilins/archaellins may allow for unique patterns of assembly within filaments, and differences within the TFF superfamily may be more widespread than originally thought.

Intermolecular interactions suggest a heterodimeric ArlB1-ArlB2 building block

Archaella are assembled by the platform protein FlaJ, powered by the ATPase FlaI [6]. For homopolymeric archaella, the filament is thought to be built from monomers. However, building the complex n+6 symmetry observed in the *M. villosus* archaellum from archaellin monomers would need to be selective for alternating subunits, and thus require a more complex assembly mechanism compared to one that assembles homopolymeric archaella. Therefore, it is conceivable that the filament is instead assembled by preformed ArlB1-2 heterodimers, which are then added to the 3-start strands of the growing archaellum (fig. 4i). This is in line with our observation that ArlB1-2 interfaces show by far the strongest interactions across the filament, which suggests that ArlB1-2 may be the fundamental building block of the *M. villosus* archaellum. Notwithstanding this, knockout mutants of one of two major archaellins still assemble archaella [33]. This shows that the assembly platform FlaJ is also capable of assembling monomers of the same kind into homolopolymeric archaella, potentially adding to the functional versatility of the archaellum machinery.

Role of minor archaellins

While the function of minor pilins in bacteria have been more extensively investigated, the role of minor archaellins remains elusive. In bacteria, minor pilins have been shown to initiate pilus assembly, by forming a cap at the tip of the pilus, as well as promote pilus functions, such as DNA binding, aggregation, and adherence [80]. For some archaeal species, including *M. maripaludis* [30], *M. voltae* [36], *M. thermolithotrophicus* [31], *H. salinarum* [37], and *M. hungatei* [81] (suppl. fig. 15), it has been shown that the archaellum filament is linked to the basal body through a distinct structure described as reminiscent of the flagellar hook found in bacteria.

It is conceivable that the remaining minor archaellin, ArlB3, assembles into a terminal structure of the archaellum, which could either act as a cap or basal structure. Corroborating this notion, sequence alignments show a significant degree of conservation between *M. villosus* ArlB3 and the ArlB3 homologs from *M. voltae* (48.3% identity and 78.5% similarity) and *M. maripaludis* (50.4% identity and 75.4% similarity). In contrast, a lesser degree of conservation is observed between *M. villosus* ArlB1 / ArlB2 with ArlB3 from *M. voltae and M. maripaludis* (suppl. fig. 16).

Glycosylation patterns in heteropolymeric archaella

One of the most obvious differences between the two archaellin components of the heteropolymeric *M. villosus* archaellum is the number of glycosylation sites. Interestingly, ArlB2 shows glycosylation sites on two consecutive Asn residues within the NNTT motif in its glycosylation loop. The sharply bent structure of the loop minimises the steric hindrance between the adjacent glycans. The NNTT motif is common in viral proteins [82] but simultaneous glycosylation of two neighbouring residues has so far been shown only for the Gp3 protein of the Equine Arteritis Virus, in which glycosylation may prevent processing of an adjacent cleavage site [83].

Glycosylation plays a key role in archaellum filament assembly and motility [84-88]. In addition, it has been suggested that glycosylation increases the resilience of archaeal type-IV filaments against acidic habitats [74,75], as well as thermal stability, with the latter being directly proportional to the degree of glycosylation [89]. *M. villosus* thrives at an optimal temperature of 80 °C. The glycosylation that we observe in the *M. villosus* archaellum could be an adaption to hot environments. Moreover, because *M. villosus* can grow between 55-90 °C, an interesting hypothesis to test would be to observe if at different temperatures archaella can be assembled by one or the other archaellin homologues independently. If this were the case, assembling archaella with different homologues could be a way to modulate the glycosylation level of the archaellum filament in response to different temperatures.

Flexibility of the archaellum

Archaella and bacterial flagella propel the cell in a process that starts from the motor complex, where torque is generated, and is transmitted through into the filament [6]. The bacterial flagellum filament supercoils due to its constituting subunits switching between two discrete states, long and short [90]. In archaella the conformational changes of the filament subunits are widely unexplored.

Here we provide first direct insights into the molecular flexibility of the archaellum at near-atomic detail and identify two major structural elements that govern the filament's flexibility. At the monomer level, the loop region between the head and the tail domain acts as a hinge, allowing the two domains to move relative to each other. The amino acid sequence of the hinge region is highly conserved in archaellins and differs from that of archaeal and bacterial T4P. This indicates that the hinge region of archaellins is specialised to filaments that perform a propulsive motion. The amino acids that make up the hinge of non-motile archaeal pili contain larger side chains (QQT/QVT/QGT) than those of archaella (ASG). This may infer

more steric hindrance and less flexibility of pili-borne compared to archaellar hinges. Based on this, the hinge region could be used to distinguish between archaellins and pilins when predicting their function at the sequence level.

At the core of the archaellum, the tail domains of individual subunits slide along and rotate around each other, further increasing the flexibility of the archaellum. This sliding motion is enabled by the hydrophobic interactions that solely hold together the tail domains. This type of intermolecular interaction is known to provide "molecular grease" in various flexible or rotary protein assemblies [91]. Together, the hinges within each subunit, as well as the molecular grease within the filament's core provide the basis for the flexibility of archaella that is essential in their function as molecular propellers.

The flexibility of the *M. villosus* archaellum filament supports recent insights into the motion of the flagellar hook from the bacterium *S. enterica* [92,93]. Similar to archaellins, the three domains of the flagellar hook subunits (D1, D2, D3) behave as rigid bodies connected by two flexible hinges. The authors identified eleven distinctive conformations of the subunits along a full helical turn. In contrast, our data suggest that archaellins change their shape in an indiscrete manner. While in the flagellar hook discrete conformations of the subunits may be important to maintain its hooked shape, the continuity of motion seen in archaellins is likely a prerequisite for continuously undulating and supercoiling archaella.

Acknowledgements

Author contributions

LG: Performed the research, provided methodology, wrote the manuscript.

MI: Performed the research, provided methodology, wrote the manuscript.

RC: Provided methodology.

MM: Provided methodology.

VG: Provided resources and funding for RC, contributed to writing of the manuscript.

AB: Provided resources and methodology, contributed to writing of the manuscript.

RR: Provided resources and methodology.

BD: Provided funding for LG, performed research, provided resources, wrote the manuscript.

The authors declare no competing interests.

References

1   Woese, C. R., Magrum, L. J. & Fox, G. E. Archaebacteria. *J Mol Evol* **11**, 245-251, doi:10.1007/bf01734485 (1978).

2   DeLong, E. F. & Pace, N. R. Environmental diversity of bacteria and archaea. *Syst Biol* **50**, 470-478 (2001).

3   Schleper, C., Jurgens, G. & Jonuscheit, M. Genomic studies of uncultivated archaea. *Nat Rev Microbiol* **3**, 479-488, doi:10.1038/nrmicro1159 (2005).

4   Albers, S. V. & Jarrell, K. F. The archaellum: how Archaea swim. *Front Microbiol* **6**, 23, doi:10.3389/fmicb.2015.00023 (2015).

5   Marwan, W., Alam, M. & Oesterhelt, D. Rotation and switching of the flagellar motor assembly in Halobacterium halobium. *J Bacteriol* **173**, 1971-1977, doi:10.1128/jb.173.6.1971-1977.1991 (1991).

6   Albers, S. V. & Jarrell, K. F. The Archaellum: An Update on the Unique Archaeal Motility Structure. *Trends Microbiol* **26**, 351-362, doi:10.1016/j.tim.2018.01.004 (2018).

7   Berry, J. L. & Pelicic, V. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol Rev* **39**, 134-154, doi:10.1093/femsre/fuu001 (2015).

8   Jarrell, K. F., Stark, M., Nair, D. B. & Chong, J. P. Flagella and pili are both necessary for efficient attachment of *Methanococcus maripaludis* to surfaces. *FEMS Microbiol Lett* **319**, 44-50, doi:10.1111/j.1574-6968.2011.02264.x (2011).

9   Yu, X. *et al.* Filaments from *Ignicoccus hospitalis* show diversity of packing in proteins containing N-terminal type IV pilin helices. *J Mol Biol* **422**, 274-281, doi:10.1016/j.jmb.2012.05.031 (2012).

10   Näther, D. J., Rachel, R., Wanner, G. & Wirth, R. Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts. *J Bacteriol* **188**, 6915-6923, doi:10.1128/jb.00527-06 (2006).

11   Bellack, A., Huber, H., Rachel, R., Wanner, G. & Wirth, R. *Methanocaldococcus villosus* sp. nov., a heavily flagellated archaeon that adheres to surfaces and forms cell-cell contacts. *Int J Syst Evol Microbiol* **61**, 1239-1245, doi:10.1099/ijs.0.023663-0 (2011).

12   Pohlschroder, M. & Esquivel, R. N. Archaeal type IV pili and their involvement in biofilm formation. *Front Microbiol* **6**, 190, doi:10.3389/fmicb.2015.00190 (2015).

13   Walker, D. J. F. *et al.* The archaellum of *Methanospirillum hungatei* is electrically conductive. *mBio* **10**, doi:10.1128/mBio.00579-19 (2019).

14   Beeby, M., Ferreira, J. L., Tripp, P., Albers, S. V. & Mitchell, D. R. Propulsive nanomachines: the convergent evolution of archaella, flagella and cilia. *FEMS Microbiol Rev* **44**, 253-304, doi:10.1093/femsre/fuaa006 (2020).

15   Cohen-Krausz, S. & Trachtenberg, S. The flagellar filament structure of the extreme acidothermophile *Sulfolobus shibatae* B12 suggests that archaeabacterial flagella have a unique and common symmetry and design. *J Mol Biol* **375**, 1113-1124, doi:10.1016/j.jmb.2007.10.048 (2008).

16   Jarrell, K. F., Bayley, D. P. & Kostyukova, A. S. The archaeal flagellum: a unique motility structure. *J Bacteriol* **178**, 5057-5064, doi:10.1128/jb.178.17.5057-5064.1996 (1996).

17   Reindl, S. *et al.* Insights into FlaI functions in archaeal motor assembly and motility from structures, conformations, and genetics. *Mol Cell* **49**, 1069-1082, doi:10.1016/j.molcel.2013.01.014 (2013).

18   Meshcheryakov, V. A. & Wolf, M. Crystal structure of the flagellar accessory protein FlaH of *Methanocaldococcus jannaschii* suggests a regulatory role in archaeal flagellum assembly. *Protein Sci* **25**, 1147-1155, doi:10.1002/pro.2932 (2016).

19   Chaudhury, P. *et al.* The nucleotide-dependent interaction of FlaH and FlaI is essential for assembly and function of the archaellum motor. *Mol Microbiol* **99**, 674-685, doi:10.1111/mmi.13260 (2016).

20   Banerjee, A. *et al.* FlaX, a unique component of the crenarchaeal archaellum, forms oligomeric ring-shaped structures and interacts with the motor ATPase FlaI. *J Biol Chem* **287**, 43322-43330, doi:10.1074/jbc.M112.414383 (2012).

21   Daum, B. *et al.* Structure and in situ organisation of the *Pyrococcus furiosus* archaellum machinery. *Elife* **6**, doi:10.7554/eLife.27470 (2017).

22   Briegel, A. *et al.* Morphology of the archaellar motor and associated cytoplasmic cone in *Thermococcus kodakaraensis*. *EMBO Rep* **18**, 1660-1670, doi:10.15252/embr.201744070 (2017).

23   Banerjee, A. *et al.* FlaF is a β-sandwich protein that anchors the archaellum in the archaeal cell envelope by binding the S-Layer protein. *Structure* **23**, 863-872, doi:10.1016/j.str.2015.03.001 (2015).

24   Tsai, C. L. *et al.* The structure of the periplasmic FlaG-FlaF complex and its essential role for archaellar swimming motility. *Nat Microbiol* **5**, 216-225, doi:10.1038/s41564-019-0622-3 (2020).

25   Tschitschko, B. *et al.* Genomic variation and biogeography of Antarctic haloarchaea. *Microbiome* **6**, 113, doi:10.1186/s40168-018-0495-3 (2018).

26   Ng, S. Y., Chaban, B. & Jarrell, K. F. Archaeal flagella, bacterial flagella and type IV pili: a comparison of genes and posttranslational modifications. *J Mol Microbiol Biotechnol* **11**, 167-191, doi:10.1159/000094053 (2006).

27   Kalmokoff, M. L. & Jarrell, K. F. Cloning and sequencing of a multigene family encoding the flagellins of *Methanococcus voltae*. *J Bacteriol* **173**, 7113-7125, doi:10.1128/jb.173.22.7113-7125.1991 (1991).

28   Näther-Schindler, D. J., Schopf, S., Bellack, A., Rachel, R. & Wirth, R. *Pyrococcus furiosus* flagella: biochemical and transcriptional analyses identify the newly detected flaB0 gene to encode the major flagellin. *Front Microbiol* **5**, 695 (2014).

29   Kalmokoff, M. L., Jarrell, K. F. & Koval, S. F. Isolation of flagella from the archaebacterium *Methanococcus voltae* by phase separation with Triton X-114. *J Bacteriol* **170**, 1752-1758, doi:10.1128/jb.170.4.1752-1758.1988 (1988).

30   Chaban, B. *et al.* in *Mol Microbiol* Vol. **66**, 596-609 (2007).

31   Kelly, J. F. *et al.* Identification of a novel N-linked glycan on the archaellins and S-layer protein of the thermophilic methanogen, *Methanothermococcus thermolithotrophicus*. *J Biol Chem* **295**, 14618-14629, doi:10.1074/jbc.RA120.012790 (2020).

32   Tarasov, V. Y., Pyatibratov, M. G., Tang, S. L., Dyall-Smith, M. & Fedorov, O. V. Role of flagellins from A and B loci in flagella formation of *Halobacterium salinarum*. *Mol Microbiol* **35**, 69-78, doi:10.1046/j.1365-2958.2000.01677.x (2000).

33   Pyatibratov, M. G. *et al.* Interaction of two strongly divergent archaellins stabilizes the structure of the *Halorubrum archaellum*. *Microbiologyopen* **9**, e1047, doi:10.1002/mbo3.1047 (2020).

34   Poweleit, N. *et al.* CryoEM structure of the *Methanospirillum hungatei* archaellum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nat Microbiol* **2**, 16222, doi:10.1038/nmicrobiol.2016.222 (2016).

35   Meshcheryakov, V. A. *et al.* High-resolution archaellum structure reveals a conserved metal-binding site. *EMBO Rep* **20**, doi:10.15252/embr.201846340 (2019).

36   Bardy, S. L., Mori, T., Komoriya, K., Aizawa, S. & Jarrell, K. F. Identification and localization of flagellins FlaA and FlaB3 within flagella of *Methanococcus voltae*. *J Bacteriol* **184**, 5223-5233, doi:10.1128/jb.184.19.5223-5233.2002 (2002).

37   Beznosov, S. N., Pyatibratov, M. G. & Fedorov, O. V. On the multicomponent nature of *Halobacterium salinarum* flagella.

38   Herzog, B. & Wirth, R. Swimming behavior of selected species of Archaea. *Appl Environ Microbiol* **78**, 1670-1674, doi:10.1128/aem.06723-11 (2012).

39   Balch, W. E. & Wolfe, R. S. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium ruminantium in a pressureized atmosphere. *Appl Environ Microbiol* **32**, 781-791, doi:10.1128/aem.32.6.781-791.1976 (1976).

40   Fricke, H. *et al.* Hydrothermal vent communities at the shallow subpolar Mid-Atlantic ridge.

41   Huber, H. & Stetter, K. O. Desulfurococcales. *Prokaryotes* **3**, 16, doi:10.1007/0-387-30743-5_4 (2006).

42   Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr D Struct Biol* **76**, 94-101, doi:10.1107/s2059798319016577 (2020).

43   Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

44   Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-590, doi:10.1038/nmeth.2472 (2013).

45   Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

46   Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).

47   Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv*, 2020.2006.2012.148296, doi:10.1101/2020.06.12.148296 (2020).

48   Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367, doi:10.1107/s0907444911001314 (2011).

49   Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/s0907444910007493 (2010).

50   Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011, doi:10.1107/s0907444906022116 (2006).

51   Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519-530, doi:10.1107/s2059798318002425 (2018).

52   Pettersen, E. F. *et al.* UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.2008410.1002/jcc.20084. (2004).

53   Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/s0907444910045749 (2011).

54   Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol* **73**, 469-477, doi:10.1107/s2059798317007859 (2017).

55   Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68**, 431-440, doi:10.1107/s090744491200251x (2012).

56   Punjani, A. & Fleet, D. J. 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol*, 107702, doi:10.1016/j.jsb.2021.107702 (2021).

57   Vagin, A. A. & Isupov, M. N. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr* **57**, 1451-1456, doi:10.1107/s0907444901012409 (2001).

58   Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367, doi:10.1107/s0907444912001308 (2012).

59   Rachel, R. *et al.* Dual-axis STEM tomography at 200 kV: Setup, performance, limitations. *J Struct Biol* **211**, 107551, doi:10.1016/j.jsb.2020.107551 (2020).

60   Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350, doi:10.1093/nar/gkw408 (2016).

61   Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324, doi:10.1093/nar/gku316 (2014).

62   Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464-469, doi:10.1093/bioinformatics/btr703 (2012).

63   Heringa, J. Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment. *Comput Chem* **23**, 341-364, doi:10.1016/s0097-8485(99)00012-1 (1999).

64   Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

65   McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67**, 386-394, doi:10.1107/s0907444911007281 (2011).

66   Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **27**, 3315-3316, doi:10.1093/bioinformatics/btr575 (2011).

67   Bond, C. S. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311-312, doi:10.1093/bioinformatics/19.2.311 (2003).

68   Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

69   Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).

70   Szabó, Z. *et al.* Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases. *J Bacteriol* **189**, 772-778, doi:10.1128/jb.01547-06 (2007).

71   Kinosita, Y., Uchida, N., Nakane, D. & Nishizaka, T. Direct observation of rotation and steps of the archaellum in the swimming halophilic archaeon *Halobacterium salinarum*. *Nat Microbiol* **1**, 16148, doi:10.1038/nmicrobiol.2016.148 (2016).

72   Desmond, E., Brochier-Armanet, C. & Gribaldo, S. Phylogenomics of the archaeal flagellum: rare horizontal gene transfer in a unique motility structure. *BMC Evol Biol* **7**, 106, doi:10.1186/1471-2148-7-106 (2007).

73   Syutkin, A. S. *et al.* Salt-dependent regulation of archaellins in *Haloarcula marismortui*. *Microbiologyopen* **8**, e00718, doi:10.1002/mbo3.718 (2019).

74   Wang, F. *et al.* The structures of two archaeal type IV pili illuminate evolutionary relationships.

75   Wang, F. *et al.* An extensively glycosylated archaeal pilus survives extreme conditions. *Nat Microbiol* **4**, 1401-1410, doi:10.1038/s41564-019-0458-x (2019).

76   Wang, F. *et al.* Cryoelectron microscopy reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* Type IV Pili at sub-nanometer resolution. *Structure* **25**, 1423-1435.e1424, doi:10.1016/j.str.2017.07.016 (2017).

77   Kolappan, S. *et al.* Structure of the *Neisseria meningitidis* Type IV pilus. *Nat Commun* **7**, 13015, doi:10.1038/ncomms13015 (2016).

78   Neuhaus, A. *et al.* Cryo-electron microscopy reveals two distinct type IV pili assembled by the same bacterium. *Nat Commun* **11**, 2231, doi:10.1038/s41467-020-15650-w (2020).

79   Berry, J. L. *et al.* Global biochemical and structural analysis of the type IV pilus from the Gram-positive bacterium *Streptococcus sanguinis*. *J Biol Chem* **294**, 6796-6808, doi:10.1074/jbc.RA118.006917 (2019).

80   Jacobsen, T., Bardiaux, B., Francetic, O., Izadi-Pruneyre, N. & Nilges, M. Structure and function of minor pilins of type IV pili. *Med Microbiol Immunol* **209**, 301-308, doi:10.1007/s00430-019-00642-5 (2020).

81   Cruden, D., Sparling, R. & Markovetz, A. J. Isolation and ultrastructure of the flagella of *Methanococcus thermolithotrophicus* and *Methanospirillum hungatei*. *Appl Environ Microbiol* **55**, 1414-1419, doi:10.1128/aem.55.6.1414-1419.1989 (1989).

82   Karamyshev, A. L. *et al.* Mapping the interaction of the STT3 subunit of the oligosaccharyl transferase complex with nascent polypeptide chains. *J Biol Chem* **280**, 40489-40493, doi:10.1074/jbc.M509168200 (2005).

83   Matczuk, A. K., Kunec, D. & Veit, M. Co-translational processing of glycoprotein 3 from equine arteritis virus: N-glycosylation adjacent to the signal peptide prevents cleavage. *J Biol Chem* **288**, 35396-35405, doi:10.1074/jbc.M113.505420 (2013).

84   Meyer, B. H., Birich, A. & Albers, S. V. N-Glycosylation of the archaellum filament is not important for archaella assembly and motility, although N-Glycosylation is essential for motility in *Sulfolobus acidocaldarius*. *Biochimie* **118**, 294-301, doi:10.1016/j.biochi.2014.10.018 (2015).

85   Zaretsky, M., Darnell, C. L., Schmid, A. K. & Eichler, J. N-Glycosylation is important for *Halobacterium salinarum* archaellin expression, archaellum assembly and cell motility. *Frontiers in Microbiology* **10**, doi:10.3389/fmicb.2019.01367 (2019).

86   Ding, Y. *et al.* Effects of N-glycosylation site removal in archaellins on the assembly and function of archaella in *Methanococcus maripaludis*. *PLoS One* **10**, e0116402, doi:10.1371/journal.pone.0116402 (2015).

87   Tripepi, M. *et al.* N-glycosylation of *Haloferax volcanii* flagellins requires known Agl proteins and is essential for biosynthesis of stable flagella. *J Bacteriol* **194**, 4876-4887, doi:10.1128/jb.00731-12 (2012).

88   Chaban, B., Voisin, S., Kelly, J., Logan, S. M. & Jarrell, K. F. Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol Microbiol* **61**, 259-268, doi:10.1111/j.1365-2958.2006.05226.x (2006).

89   Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* **105**, 8256-8261, doi:10.1073/pnas.0801340105 (2008).

90   Wang, F. *et al.* A structural model of flagellar filament switching across multiple bacterial species. *Nat Commun* **8**, 960, doi:10.1038/s41467-017-01075-5 (2017).

91   Yamaguchi, T. *et al.* Structure of the molecular bushing of the bacterial flagellar motor. *bioRxiv*, 2020.2011.2012.379347, doi:10.1101/2020.11.12.379347 (2020).

92   Shibata, S., Matsunami, H., Aizawa, S. I. & Wolf, M. Torque transmission mechanism of the curved bacterial flagellar hook revealed by cryo-EM. *Nat Struct Mol Biol* **26**, 941-945, doi:10.1038/s41594-019-0301-3 (2019).

93   Kato, T., Makino, F., Miyata, T., Horváth, P. & Namba, K. Structure of the native supercoiled flagellar hook as a universal joint. *Nat Commun* **10**, 5295, doi:10.1038/s41467-019-13252-9 (2019).

Figures
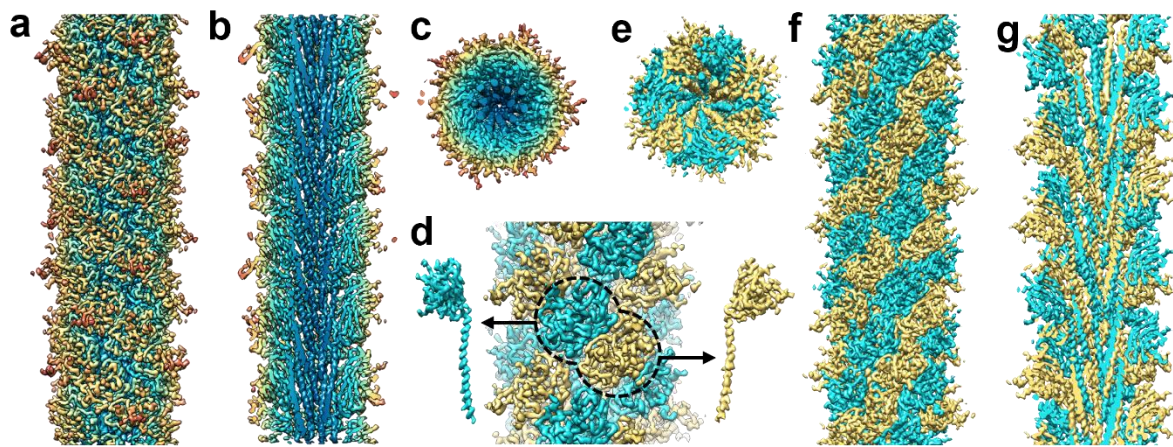
Fig. 1| CryoEM map of the *M. villosus* archaellum. **a** and **f**, surface view; **b** and **g** cross-section parallel to the filament's long axis and **c** and **e** cross sections perpendicular to the long axis of the filament. Colours in (a), (b) and (c) range radially from the core (blue) to the periphery (red) of the filament. Grooves on the surface in (a) follow the left-handed 3-start helix . In (d-g), the two alternating archaellin densities are coloured in cyan and yellow. In (d), the head domains of the two archaellins are highlighted with a dashed line and the density maps of each subunit are shown left (ArlB1, cyan) and right (ArlB2, sand) of the filament (arrows). Scale bars 50 Å.
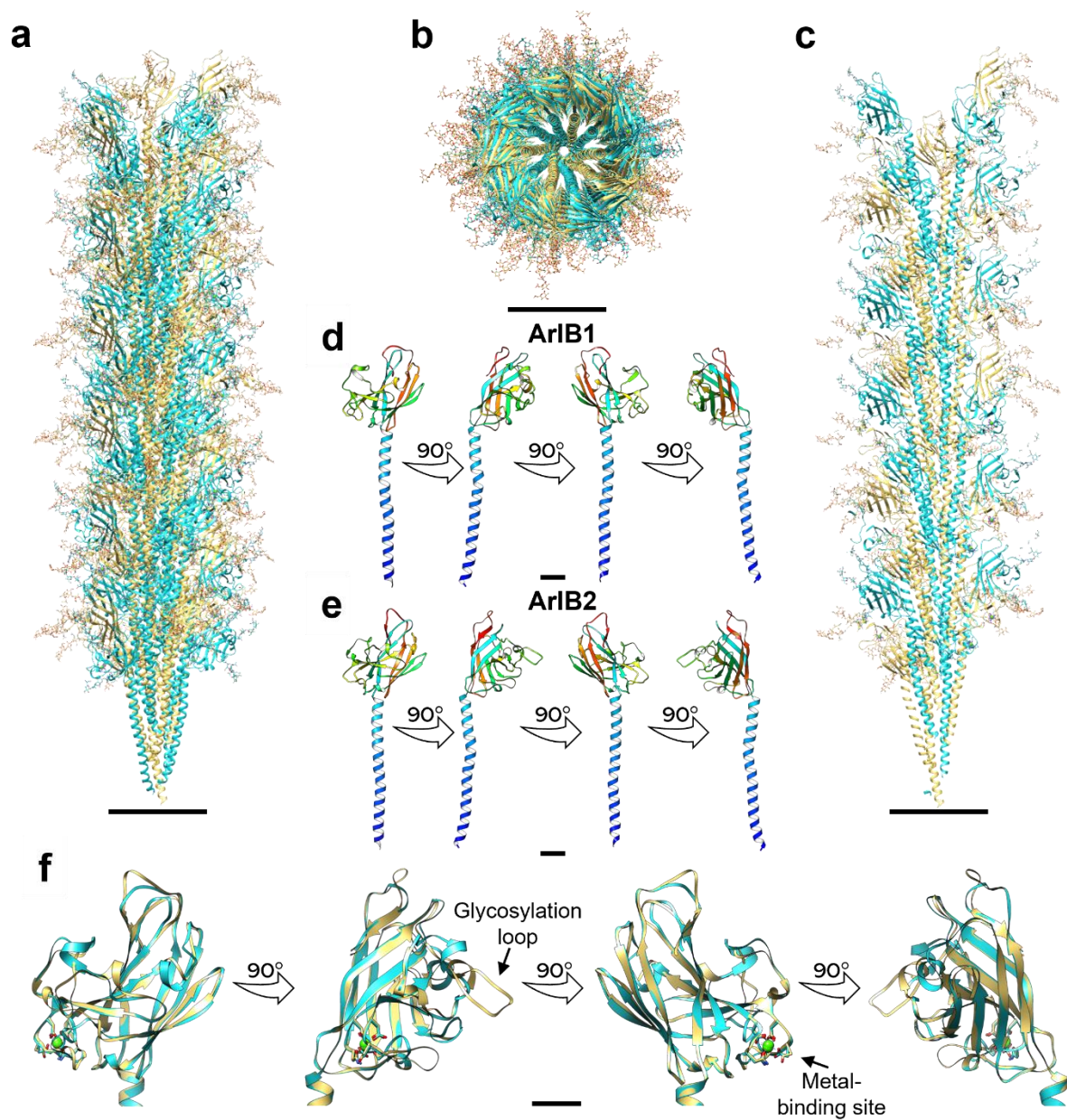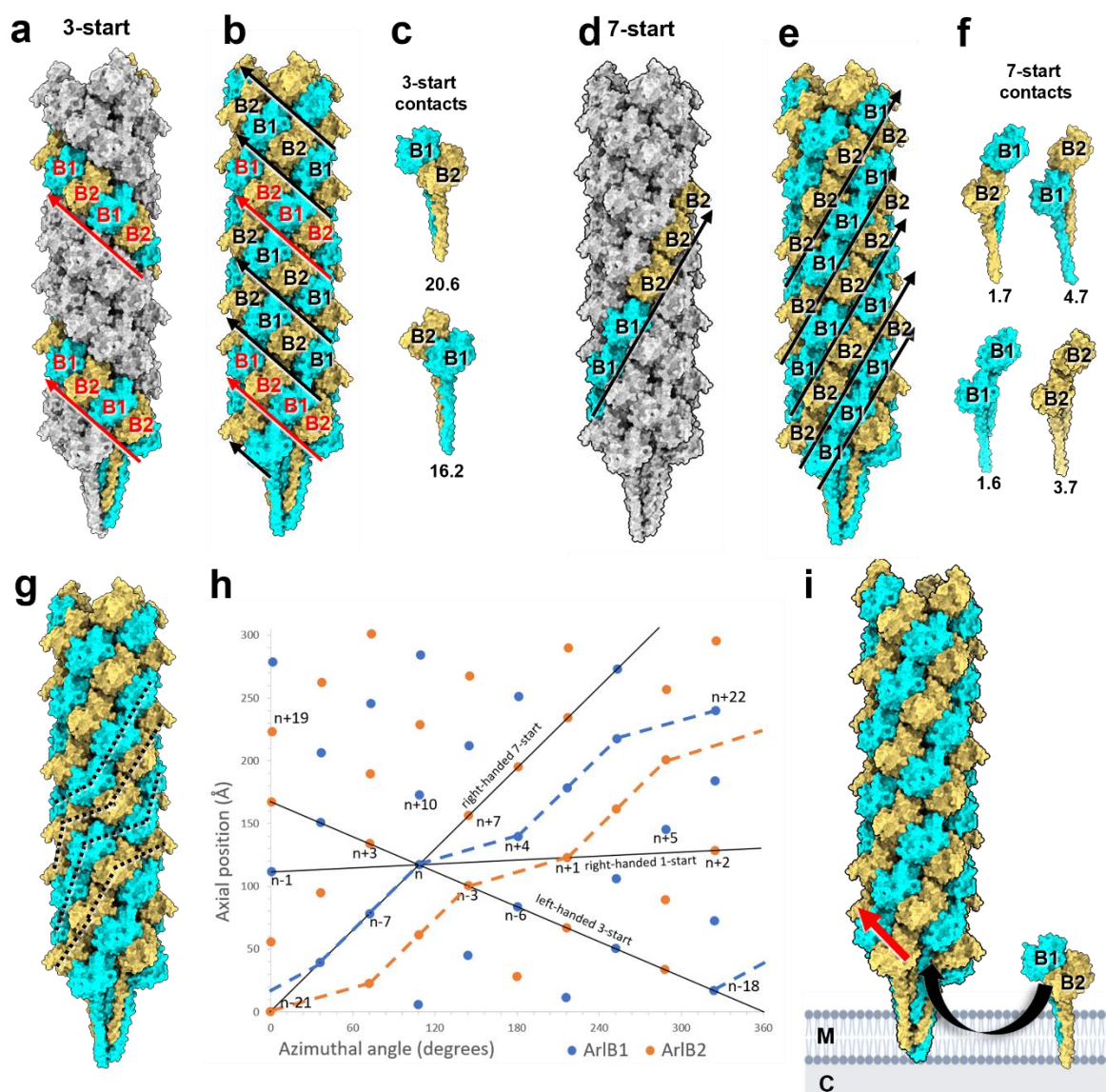
Fig. 2| Atomic model of the *M. villosus* archaellum filament. **a-c,** atomic model of the archaellum filament with glycosylation in side view (a) and cross-section perpendicular (b) and parallel (c) to the filament's axis. Scale bar, 50 Å. **d** and **e**, atomic models of ArlB1 (d) ArlB2 (e) in ribbon representation and rainbow colour scheme (N-terminus, blue; C-terminus, red). **f**, superposition of the ArlB1 (cyan) and ArlB2 (sand) head domain with metal binding highlighted. Scale bar in (d), (e) and (f), 10 Å.
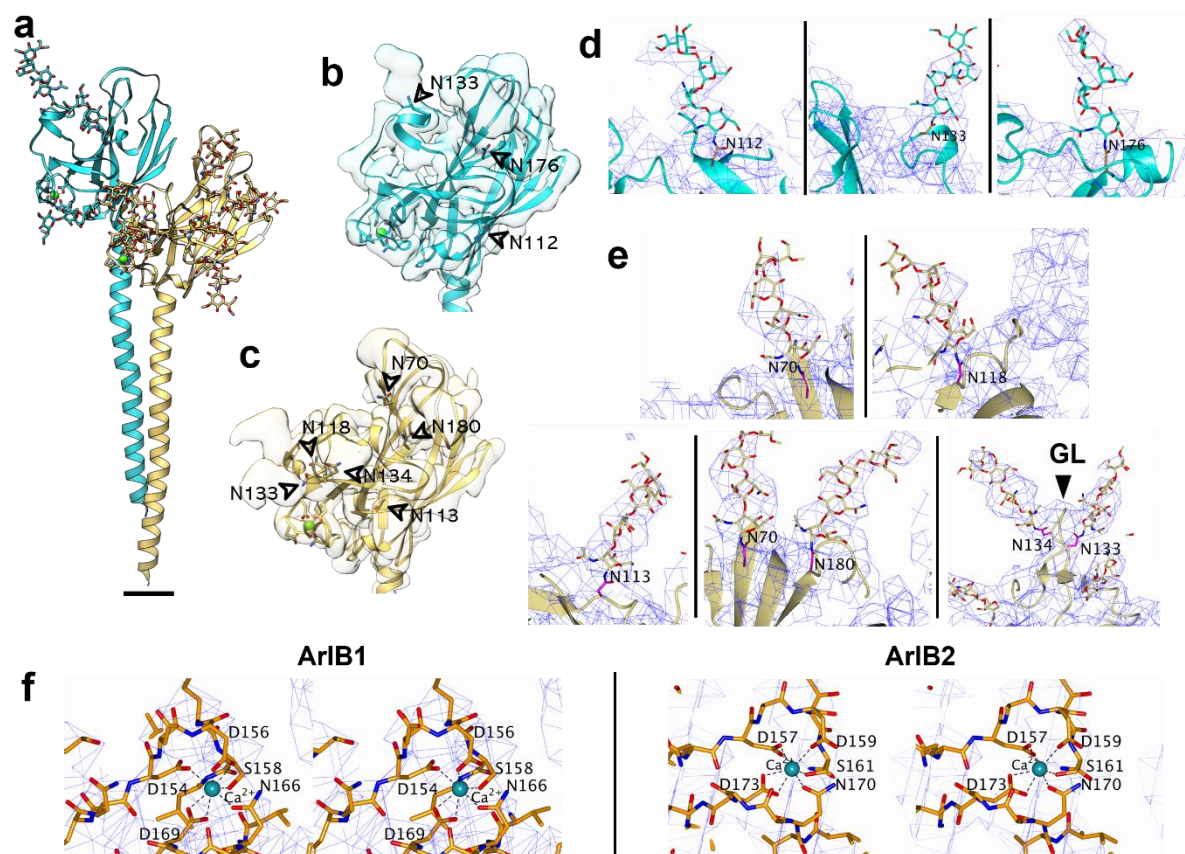
Fig. 3| Glycosylation and metal binding sites in ArlB1 and ArlB2. **a**, atomic models of ArlB1 (cyan) and ArlB2 (sand) are shown in ribbon representation with modelled glycan molecules, side chains of metal and glycan binding residues shown as stick models and metal ions highlighted as green spheres. Scale bar in (a), 10 Å. **b** and **c**, ribbon diagrams of head domains of ArlB1 (b) and ArlB2 (c) with their cryoEM density maps superimposed. Glycan densities are indicated by arrow heads. **d**, close-ups of the three glycosylation sites in ArlB1. **e**, close-ups of the six glycosylation sites in ArlB2. GL, glycosylation loop. **f**, stereo views of ArlB1 and ArlB2 metal binding sites. The cryoEM map is represented as a blue mesh and contoured at 2.8 σ. Protein chain is shown as a stick model, Ca^{2+} ion is shown as a cyan sphere with metal-coordinating residues labelled.

Fig. 4| Screw axis asymmetry in the *M. villosus* archaellum. a, the archaellum atomic model viewed parallel to the filament's axis. ArlB1 (cyan) and ArlB2 (sand) subunits are highlighted along one left-handed 3-start helical strand. ArlB1 and ArlB2 alternate in the 3-start direction (indicated by the red arrows). **b**, subunits within all three 3-start strands are highlighted. Every third 3-start strand (red arrow) is out of register with respect to the other two 3-start strands (black arrows). **c**, intermolecular contacts in 3-start direction between ArlB1-2 (top) and ArlB2-1 (bottom) dimers. **d**, one right-handed 7-start strand and **e**, all 7-start strands are highlighted (direction of the 7-start strands indicated with black arrows). **f**, four types of contacts with various binding strengths occur within the 7-start strands. The dissociation energies (ΔG_{diss} calculated by PISA) for each interaction in 3-start direction (c) and 7-start direction (f) are indicated below each dimer (unit: kcal/mol). **g**, tracing subunits of the same kind through the filament indicates non-helical homopolymeric pseudo-strands (dashed lines). **h**, helical net diagram showing the positions of ArlB1 (blue dots) and ArlB2 (orange dots) in a 2-dimensional plot. Solid black lines show various component helices. Blue and orange lines trace non-helical homopolymeric pseudo-strands of ArlB1 and ArlB2. **i**, model of the hypothetical assembly of the heteropolymeric archaellum from pre-formed ArlB1-2 dimers that are added to the growing filament in 3-start direction. M, membrane; C, cytosol; red arrow, 3-start direction. For simplicity, the archaellum machinery and the S-layer have not been drawn.
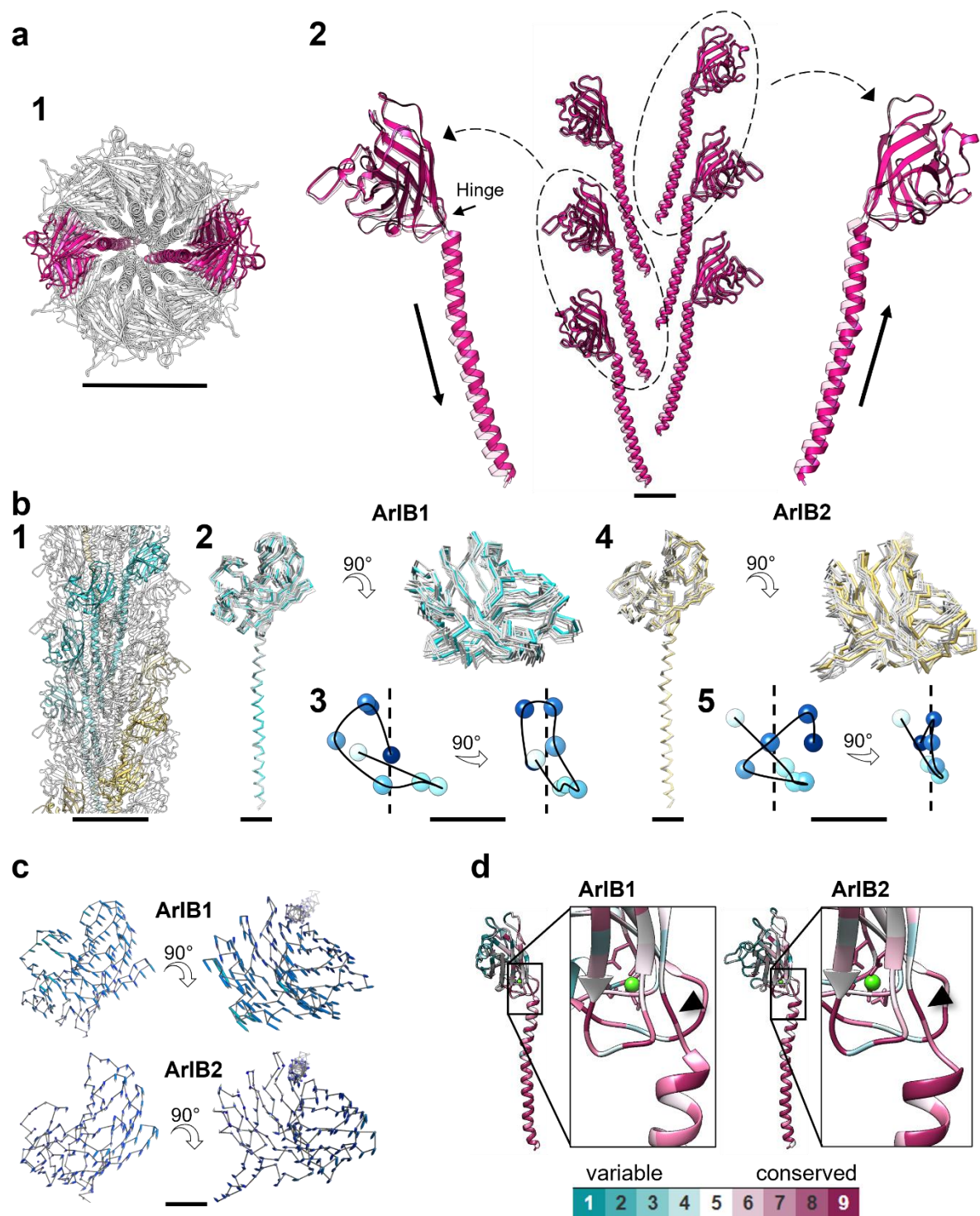
Fig. 5| Flexibility of the *M. villosus* archaellum. **a**, atomic model showing the motion of the tail domains. Two opposite n+10 sets of protein monomers (a1) were selected and the head domains of the constituting monomers were aligned. The solid and transparent magenta models (a2) were fitted into the frame0 and frame19 maps of the cryoSPARC 3DVA respectively. The arrows along the tail domains in (a2) highlight the up-down motion of the tail domains of opposite n+10 strands. Scale bar in (a2), 20 Å. **b**, atomic models and motion trajectories showing the flexibility of the head domains along a full filament turn. Eight ArlB1 and eight ArlB2 monomers forming a full helical turn were selected (b1) and their tail domains aligned. (b2) and (b4) show the flexibility of the head domains of ArlB1 and ArlB2 respectively. (b3) and (b5) represent the motion trajectories of the centroids (shades of blue) of each head domain. The black curve highlights the trajectory, and the dotted black line represents the filament axis. **c**, atomic models of one ArlB1 and one ArlB2 head domains showing their displacement

during filament motion. The motion is visualised by vectors linking the backbone atoms of the atomic models fitting into the frame0 (grey line) and frame19 (not shown for better clarity) maps of the cryoSPARC 3DVA. **d**, estimation of the evolutionary conservation of AlrB1 and ArlB2. Insets focus on the hinge region. Scale bars in (a1) and (b1), 50 Å; in (a2), (b2), (b4), (c), 10 Å; in (b3) and (b5), 2 Å.