nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0458-x

An extensively glycosylated archaeal pilus survives extreme conditions

Fengbin Wang[1], Virginija Cvirkaite-Krupovic[2], Mark A. B. Kreutzberger[1], Zhangli Su[1], Guilherme A. P. de Oliveira[1], Tomasz Osinski[1], Nicholas Sherman[3], Frank DiMaio[4], Joseph S. Wall[5], David Prangishvili[2], Mart Krupovic[2*] and Edward H. Egelman[1*]

Pili on the surface of *Sulfolobus islandicus* are used for many functions, and serve as receptors for certain archaeal viruses. The cells grow optimally at pH 3 and ~80 °C, exposing these extracellular appendages to a very harsh environment. The pili, when removed from cells, resist digestion by trypsin or pepsin, and survive boiling in sodium dodecyl sulfate or 5 M guanidine hydrochloride. We used electron cryo-microscopy to determine the structure of these filaments at 4.1 Å resolution. An atomic model was built by combining the electron density map with bioinformatics without previous knowledge of the pilin sequence—an approach that should prove useful for assemblies where all of the components are not known. The atomic structure of the pilus was unusual, with almost one-third of the residues being either threonine or serine, and with many hydrophobic surface residues. While the map showed extra density consistent with glycosylation for only three residues, mass measurements suggested extensive glycosylation. We propose that this extensive glycosylation renders these filaments soluble and provides the remarkable structural stability. We also show that the overall fold of the archaeal pilin is remarkably similar to that of archaeal flagellin, establishing common evolutionary origins.

Extremophiles are organisms that have evolved to thrive under extremely harsh environmental conditions, such as high temperature, acidic or alkaline pH, near-saturating salt concentrations, and so on. The question of how macromolecular assemblies remain stable in such settings has mystified scientists ever since the discovery that life can exist in such harsh habitats. Particularly puzzling is the robustness of proteinaceous assemblies, such as virus particles, which have to maintain stability while outside the host cell[1], and various appendages, which commonly decorate the surface of the cells and mediate diverse interactions with the environment[2]. Both types of structures are assembled from a limited number of symmetrically arranged protein components. Recent structural studies on several viruses infecting hyperthermophilic archaea have provided valuable insights into how these viruses protect their genomes and maintain structural integrity[3–7]. However, the principles underlying the stability of cellular appendages remain poorly understood.

Type IV pili (T4P) are among the most commonly found surface appendages in extremophilic archaea, with relatively small pilin proteins assembling into functionally diverse, stable structures of several micrometres in length[2,8,9]. T4P are conserved across both prokaryotic domains of life, suggesting their antiquity[10]. T4P have evolved to perform a range of functions, including motility, adhesion, aggregation, natural transformation, host cell signalling and more[8,11]. Although some functions of T4P are conserved in both bacteria and archaea, such as adhesion to various biotic and abiotic surfaces, there are also domain-specific T4P activities. The archaeal flagellum, which is evolutionarily unrelated to both bacterial and eukaryotic flagella, has evolved from T4P in the domain Archaea[12].

All structurally characterized T4P display a characteristic two-domain organization, with the highly hydrophobic amino (N)-terminal α-helix, which is typically melted in the approximate range of residues 14–23 in the filament[13–15] but remains a continuous helix in crystal structures of detergent-solubilized individual subunits[16–18], and the carboxy (C)-terminal globular domain, which can be quite variable or almost entirely absent[19]. N-terminal class III signal peptides target the precursors of T4P—prepilins—for transport across the plasma membrane in a Sec-dependent manner[20]. These are then processed by a prepilin peptidase (PilD in bacteria; PibD in archaea) and subsequently incorporated into the pilus[21,22]. The only other components necessary for T4P assembly appear to be the PilB ATPase, which powers the extrusion of pilin subunits from the membrane, and PilC, a transmembrane protein that is believed to anchor the assembled T4P to the membrane[8].

Here, we studied the structure of a pilus from *Sulfolobus islandicus* LAL14/1—a polyextremophilic (hyperthermophilic acidophile) archaeon growing optimally at ~80 °C and pH 3 (ref. [23]). A previous comprehensive comparative genomics analysis of the distribution of loci encoding flagella and other T4P in archaea showed that LAL14/1, unlike other *S. islandicus* strains, does not carry genes for a functional flagellum[9]. The filamentous appendages present on the surface of LAL14/1 cells are of only one discernible type, being ~10 nm in diameter when negatively stained or ~5 nm when observed using electron cryo-tomography, and their length is highly variable, reaching up to 12.5 μm[24]. Multiple pili are typically expressed on every cell (Supplementary. Fig. 1). These pili serve as the main receptor recognized by *S. islandicus* rod-shaped virus 2 (ref. [24])—a model virus in the family Rudiviridae[1].

Results

The pili resist degradation and characterization. The pili were purified from *S. islandicus* LAL14/1 cells. Initial attempts to

[1]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA. [2]Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, Paris, France. [3]Department of Microbiology, Immunology and Cancer Biology, University of Virginia, Charlottesville, VA, USA. [4]Department of Biochemistry, University of Washington, Seattle, WA, USA. [5]Brookhaven National Laboratory, Upton, NY, USA. *e-mail: krupovic@pasteur.fr; egelman@virginia.edu

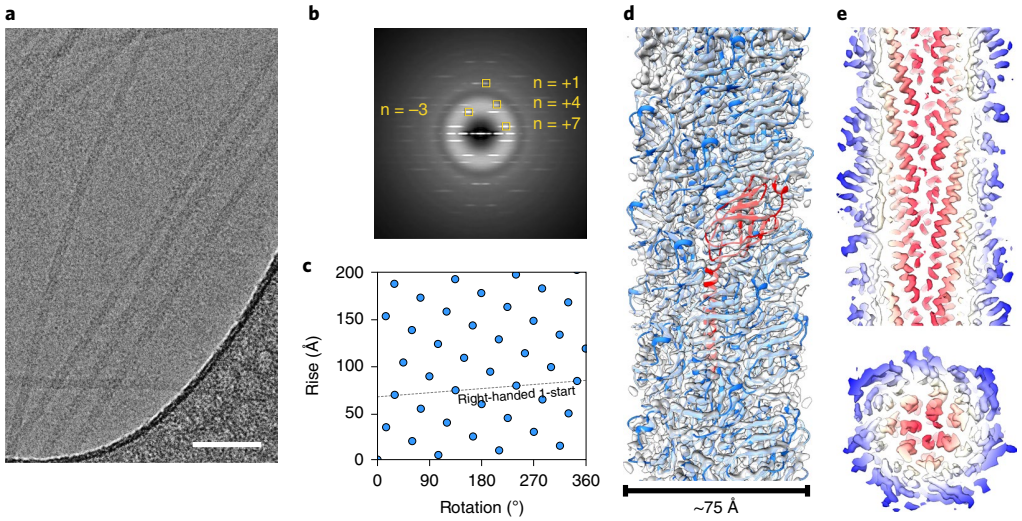


Fig. 1 | Cryo-EM of the LAL14/1 pilus. a, Representative cryo-EM image of the pili, taken from 1,824 micrographs recorded. Scale bar, 50 nm. **b**, Averaged power spectrum of the segments used in the final reconstruction. The layer lines that were used to determine the helical symmetry are labelled. **c**, Helical net of the LAL14/1 pilus using the convention that the surface is unrolled and we are viewing from the outside. One turn of the ~17-Å-pitch 1-start helix that passes through every subunit is shown. **d**, Cryo-EM reconstruction of the LAL14/1 pilus at 4.1 Å resolution with the fitted atomic model. One subunit is highlighted in red. **e**, Thin slices parallel (top) and perpendicular (bottom) to the helical axis of the pilus are shown, coloured by the helical radius. The tight packing of the all-α-helical core can be seen.

characterize these pili by mass spectrometry in both Paris and Charlottesville failed, as no proteins that were probable pilins could be identified after in-solution digestion using trypsin. Incubating the pili with a high concentration of pepsin (at pH 2.2, near the peak activity for pepsin) for 7 d resulted in a peptide mixture analysed by mass spectrometry in which over 90% were fragments of pepsin. No probable pilins were found in the remainder. Boiling the pili in 1% sodium dodecyl sulfate (SDS) failed to produce any bands seen by Coomassie Blue staining of gels, despite the high concentration of filaments, but trace amounts could be seen by silver staining. In fact, the pili were not depolymerized after boiling for 30 min in 5 M guanidine hydrochloride at pH 5.9 (Supplementary Fig. 2).

Electron cryo-microscopy (cryo-EM) of the pili. We used cryo-EM (Fig. 1a) to determine the structure of these filaments, and were able to generate a 4.1 Å-resolution reconstruction as judged by both a map:map Fourier shell correlation (FSC) and a subsequent map:model FSC (Supplementary Fig. 3). The subunits in the filament are related to each other by a rotation of ~105° about the helical axis and a translation of 4.9 Å along the helical axis, generating a right-handed 1-start helix with a pitch of ~17 Å (Fig. 1b,c). The filaments are ~75 Å in diameter and have a solid core (Fig. 1d). Normally, this resolution would immediately allow for the building of a de novo atomic model[25]. However, the sequence of the component pilin was unknown. We therefore attempted to use bioinformatics to determine it. The map revealed that the core consisted of α-helical segments each containing ~35 residues, and an atomic model without side chains for the α-helical core of the *Ignicoccus hospitalis* adhesion filament[26] fit remarkably well into this density.

It has previously been shown[27] that this core for the *I. hospitalis* adhesion filament is composed of a homologue of the bacterial T4P N-terminal domain. Such homology is also found in the archaeal flagellar filament N-terminal domain[28–32]. In contrast with the partial melting of this α-helix in the assembled bacterial T4P filaments[13–15], this region is all α-helical in the *I. hospitalis* adhesion filament[26], the archaeal flagellar filaments[28,29] and the density for the LAL14/1 pilus. The helix-breaking residues Gly14 and Pro22 that surround the unstructured region in the bacterial T4P filaments are not found in the archaeal sequences[27].

An unambiguous Cα trace was made through the density map, yielding ~125 residues (Fig. 2a). Given that the signal sequence might contain as many as 15 residues, and that there could be some uncertainty in the estimate of 125 residues, a search was made of the LAL14/1 proteome for all proteins containing between 120 and 145 residues with a single predicted transmembrane domain near the N-terminus and a β-strand-rich C-terminal domain. Only seven such sequences were found. RosettaCM[33] was then used to thread the first 50 amino acids of these 7 candidates through the α-helical core densities (Fig. 2b). This model-building approach screened the best 35 out of 50 residues and fit them into the map, thus eliminating the uncertainty of the length of the signal peptides. SiL_0372, SiL_1076 and SiL_1365 could not be fitted into the helical density; SiL_1862 and SiL_2610 could roughly fit into the helical density but had some side chain density mismatches; and SiL_2603 and SiL_2606 had identical N-terminal sequences and fit into the map of the helical core very well (Fig. 2b). A full-length model could not be built for either SiL_1862 or SiL_2610. In addition, neither of them had signal peptides present, assuming the helical fit was correct, which would be very unlikely for a pilus that needs to be transported by the conserved secretion system. A full-length model could be built for both SiL_2603 and SiL_2606, and their sequences only differed by eight residues (Fig. 2c). It was challenging to distinguish them by cryo-EM at the available resolution, as some of the amino acid changes involve leucine to valine or valine to alanine, and the two-amino-acid insertion/deletion occurs within a poorly defined loop in the density map. SiL_2606 matches the cryo-EM very well with an overall real-space correlation coefficient of 0.80 (Fig. 2d). If one assumes SiL_2603 to be the correct sequence, there would be a proline in the middle of a well-defined β-stand, which is extremely unlikely (Fig. 2e) as this would break the backbone hydrogen bonds holding the β-sheet together. Our structural model also indicates that the first ten residues in both SiL_2603 and SiL_2606 correspond to the signal sequence, which is cleaved from the mature pilin protein. The cleavage thus occurs between the Ala10 and Leu11 residues

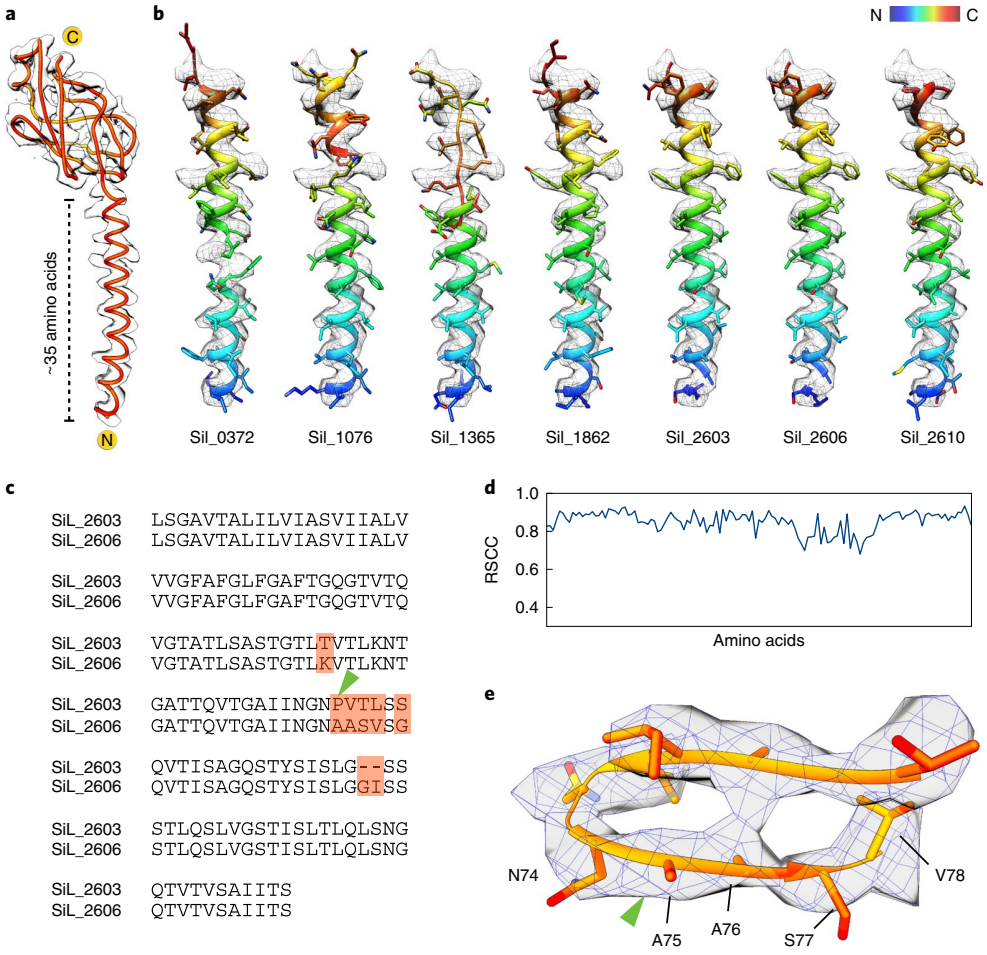


Fig. 2 | De novo atomic model building of LAL14/1 pilin. a, Protein Cα trace (orange), with the N- and C-termini of the pilin subunit labelled. **b**, RosettaCM models of the first 50 amino acids of 7 bioinformatically identified sequences that matched the Cα trace in **a**, fit into the N-terminal helix density. **c**, Sequence alignment of SiL_2603 and SiL_2606. The pink shading shows the regions where the two sequences differ. The green arrowhead indicates the proline present in Sil_2603. **d**, Per-residue real-space correlation coefficient (RSCC) plot of the atomic model from Sil_2606 against the 4.1 Å cryo-EM map. **e**, Cryo-EM map of the region indicated in **c** by the green arrowhead, with the SiL_2606 atomic model fit into the map.

and the cleavage site matches perfectly with the previously established[21] consensus recognition motif of the signal peptidase PibD ((Lys/Arg)-(Gly/Ala)-(Leu/Ile/Phe)-(Ser/Thr/Ala)). A BLASTP search has shown that orthologues of SiL_2603 and SiL_2606 are conserved in other *Sulfolobus* species. Notably, orthologues in *Sulfolobus acidocaldarius* have been characterized experimentally and shown to be involved in the formation of adhesive pili[34]. Consistent with our structural data, SiL_2606 is orthologous to the major pilin Saci_2319 (40% identity over an alignment of 140 amino acids).

Understanding the basis for the remarkable stability. The first question was whether these filaments contained amyloid, since amyloid filaments have also been shown to resist extremely harsh conditions[35]. The β-sheets in the pili appear to be typical β-sheets, and not the steric zippers associated with amyloid fibres[36]. Consistent with this, the filaments failed to stain with either Congo red or thioflavin T (Supplementary Fig. 4). Long range order (LRO) parameters have been invoked to explain the unusual thermal and kinetic stability of some proteins[37,38]. We compared the LRO score for the LAL14/1 pili with a number of bacterial T4P and archaeal flagellar filaments (Supplementary Fig. 5). It can be seen that while the predicted LRO for the LAL14/1 pilus is higher than most others in the comparison, it is comparable to that of an archaeal flagellar filament[28], which does not display such anomalous stability.

We asked whether other measures of stability could explain the anomalous properties of the pili. Using the Rosetta InterfaceAnalyzer[39], the LAL14/1 pili interfaces were compared (Supplementary Table 1) with the structures of seven archaeal and bacterial flagellar filaments, and T4P (Protein Data Bank (PDB) IDs: 5o4u, 5tfy, 5kyh, 5vxx, 5vxy, 5wjt and 5wk6). While the interface size is average, or even a little below average, when compared with these reference assemblies, the interfaces of LAL14/1 pili are markedly more hydrophobic (69% hydrophobic surface area, compared with a reference average and standard deviation of 60 ± 4%), and have a lower predicted interface energy per Å^2 (−0.037 kcal mol^{-1} Å^{-2}), compared with a reference average and standard deviation of −0.031 ± 0.002 kcal mol^{-1} Å^{-2}. While some of these metrics, such as interface energy per Å^2 or energy per residue, are lower for the LAL14/1 pilus than any of the references, the differences do not appear to be large enough to explain the anomalous properties.

Next, we looked at the overall sequence of the pilin, which is quite unusual in some respects. There are very few hydrophilic residues on the surface of the LAL14/1 pilus (Fig. 3a), in contrast with the surfaces for a bacterial T4P and an archaeal flagellar filament. The percentage of charged residues (Asp + Glu + His + Arg + Lys)

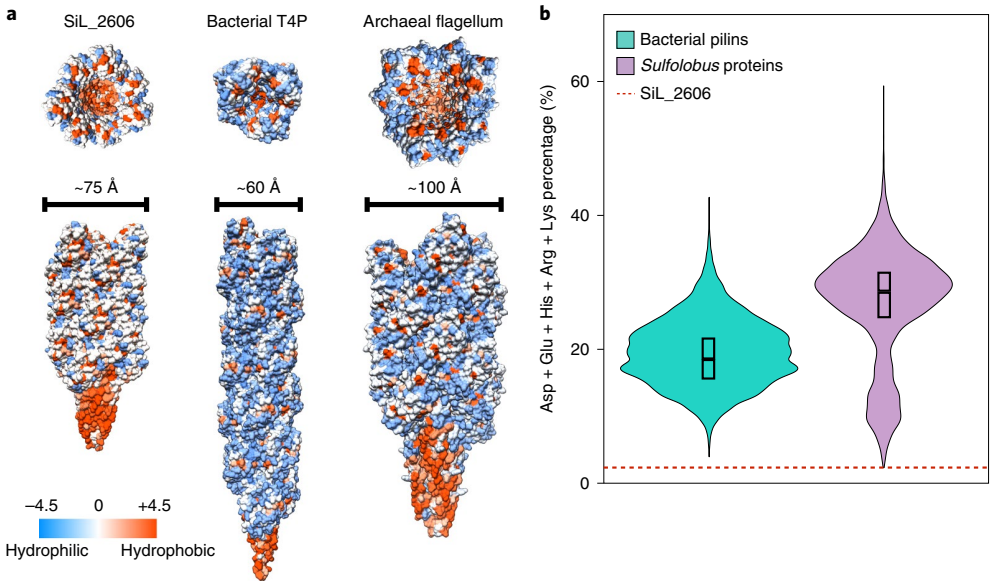


Fig. 3 | The LAL14/1 pilus contains an unusually high percentage of hydrophobic residues. a, Hydrophobicity surface of 21 subunits of the LAL14/1 pilus (left), bacterial (*N. gonorrhoeae*) type IV pilus (middle) and archaeal (*M. hungatei*) flagellar filament (right). **b**, Violin plot showing the distributions of the percentages of charged residues (Asp + Glu + Lys + Arg + His) in bacterial pilins (green, 100–200 amino acids long; $n = 52{,}248$) and *Sulfolobus* proteins (purple, 100–200 amino acids long; $n = 7{,}326$). The width of the distribution corresponds to the frequency of occurrence. The box indicates the interquartile range, and the central line in the box indicates the median of the data. The red dashed line indicates the percentage in LAL14/1 pilin SiL_2606.

within the mature protein (after cleavage of the signal sequence) is only 1.5% (Fig. 3b). This is lower than all other bacterial pilins and *Sulfolobus* proteins containing between 100 and 200 residues. This is also true when compared with archaeal flagellins and pilins (Supplementary Fig. 6a), but the number of such annotated entries is limited. Considering that many of these other 100- to 200-residue proteins are integral membrane proteins (while only the N-terminal α-helix in Sil_2606 is integrated into the membrane before filament formation), this is quite remarkable, and raises questions about how such a filament could be soluble.

Glycosylation of the filament. The cryo-EM map revealed additional mass associated with three surface residues: Thr83, Ser85 and Thr90 (Fig. 4a). Since these residues are potential targets for *O*-linked glycosylation, and it is known that many archaeal cell surface proteins are extensively glycosylated[40], the most reasonable assumption is that this extra mass is due to glycosylation. Other post-translational modifications[41], such as phosphorylation (+80 in mass), acetylation (+42 in mass) or methylation (+14 in mass), are much too small to explain the extra densities, while ubiquitination would generate a much greater density. The additional density for Thr83 (Fig. 4a) is best explained by two sugars, but at the available resolution we cannot distinguish among the many possible sugar groups. The potential targets for glycosylation on the pilus surface are extensive (Fig. 4b) given the huge number of threonine and serine residues on the surface. In fact, the frequency of threonine and serine in the SiL_2606 sequence is extremely anomalous when compared with similarly sized bacterial pilins or *Sulfolobus* proteins (Fig. 4c). This anomaly remains when comparing with archaeal flagellins and pilins (Supplementary Fig. 6b). When the transmembrane N-terminal domain is removed from the Sil_2606 sequence, 37% of the residues are threonines and serines. Looking at a projection of the reconstruction at a low threshold reveals a fuzzy coat surrounding the filament, which is absent when a higher threshold is used (Fig. 4d). As a control, projections from bacterial T4P reconstructions[13,14] were examined, which did not exhibit this behaviour (Supplementary Fig. 7). Such a fuzzy coat on the LAL14/1 pili would be consistent with extensive amounts of glycosylation.

To quantify the additional mass that is probably due to glycosylation, we employed scanning transmission electron microscopy (STEM)[42] to measure the mass per unit length of the pili (Fig. 4e). The average value was ~3.5 kDa Å^{-1}, while the predicted value of the protein alone (12.679 kDa and 4.9 Å) was only 2.6 kDa Å^{-1}. This roughly corresponds to a 35% increase in the mass of each subunit, presumably due to glycosylation. Given the huge amount of glycosylation, we asked why additional density was not seen in the three-dimensional reconstruction on many other sites for potential *O*- or *N*-linked glycosylation on the surface of the filaments. It is possible that many of the potential sites are only glycosylated on some subunits and not on others, so that this density is diminished after helical averaging. The three sites that are seen (Fig. 4a) must arise from both a very fixed orientation of these sugars and a glycosylation of these same three residues on almost every subunit.

We attempted to remove the glycosylation using a mix of deglycosylation enzymes, but without any effect as judged by mass spectrometry of the enzyme-digested products. This may be due to the fact that most of the commercial deglycosylases work mainly on *N*-linked glycans. Next, we used trifluoromethanesulfonic acid (TFMS), which has been described as a powerful method to remove both *N*- and *O*-linked glycans from proteins[43]. Conditions were found (Fig. 4f) under which a significant shift to lower mass could be obtained for a silver-stained band that ran at ~17 kDa before acid treatment and at ~13 kDa after acid treatment. This shift was almost exactly what would be predicted from the STEM analysis for the removal of glycosylation. These two bands were excised from the gel and analysed by mass spectrometry, which found that both contained SiL_2603 and SiL_2606 (Supplementary Fig. 8). A higher-mass band that was apparent before the acid treatment but absent after it cannot be due to a more heavily glycosylated SiL_2606 or SiL_2603 as neither protein was found within this band by mass spectrometry. This higher-mass band was shown by

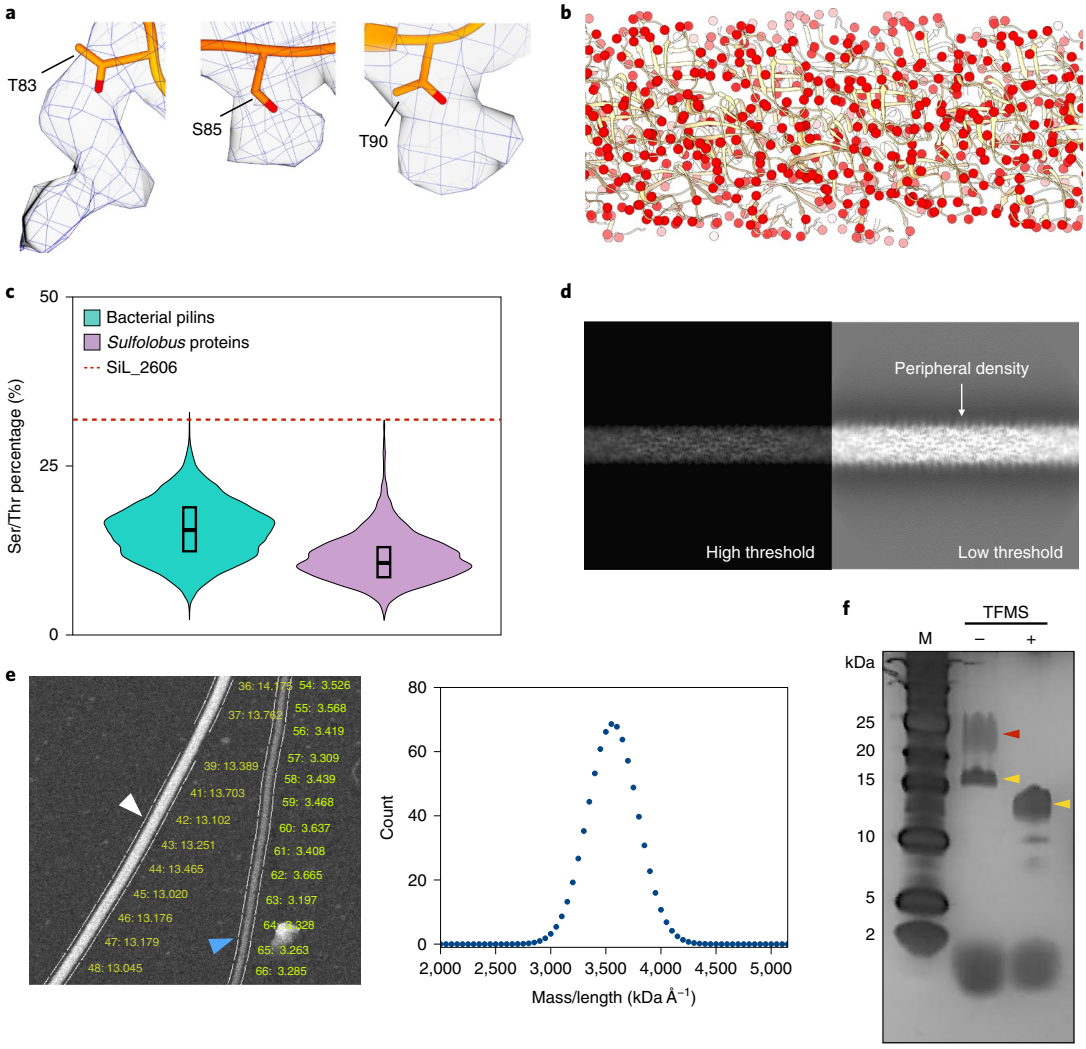


Fig. 4 | *O*-linked sugar modifications of the LAL14/1 pilus. a, Extra densities on the residues Thr83, Ser85 and Thr90 indicate post-translational modifications. **b**, Atomic model of the LAL14/1 pilus, with side chain oxygen atoms of serine and threonine shown as red spheres. **c**, Violin plot showing the distributions of the sum of serine and threonine percentages in bacterial pilins (green, 100–200 amino acids long; $n = 52{,}248$) and *Sulfolobus* proteins (purple, 100–200 amino acids long; $n = 7{,}326$). The red dashed line indicates the corresponding percentage in LAL14/1 pilin SiL_2606. The box indicates the interquartile range, and the central line in the box indicates the median of the data. **d**, Projections of the three-dimensional reconstruction under high (left) and low (right) thresholds. The low-threshold projection reveals a fuzzy coat of peripheral density. **e**, STEM analysis of LAL14/1 pili. Within one representative micrograph from 150 recorded (left), the white arrowhead indicates TMV (control) and the blue arrowhead indicates a LAL14/1 pilus. The mass-per-length measurement for each segment is shown. The distribution of all mass/length values for the pili (right) yields an average of ~3.5 kDa Å^{-1}. **f**, LAL14/1 pilus samples on silver-stained SDS-PAGE before and after deglycosylation using TFMS. This was performed twice, and the same bands were seen each time. Yellow arrowheads indicate the bands that contained SiL_2603 and SiL_2606 detected by tandem mass spectrometry analysis. The red arrowhead indicates a higher mass band that was shown by tandem mass spectrometry analysis to not contain either SiL_2603 or SiL_2606 (but contained a CdvB2 protein, SiL_1195) and thus could not be due to a more heavily glycosylated species (that is, SiL_2603 or SiL_2606). This panel is reproduced in Supplementary Fig. 9 for the purpose of comparison. M indicates the molecular weight markers lane.

mass spectrometry to contain the protein SiL_1195 (CdvB2)—a paralogue of cell division protein B (CdvB)—which probably came from membrane vesicles trapped between pili during sample preparation, consistent with the previous proteomic analysis of *Sulfolobus* vesicles[44]. Glycosylation of the protein band running at ~17 kDa was demonstrated directly using a stain specific for glycoproteins (Supplementary. Fig. 9). A consequence of the deglycosylation reaction was an aggregation of the pili, which was not surprising as the absence of many hydrophilic residues from the surface of the filament (Fig. 3a) previously raised questions about how such a filament could be soluble.

Evolutionary relationships. Sequence comparisons between SiL_2606 and the existing sequence databases of all proteins with known structures using BLASTP[45] failed to find any hits, consistent with the sparse sampling of proteins from organisms living in the most extreme environments[4]. We therefore employed structure-based comparisons using the Dali server[46] to find potential homologues of SiL_2606 (Fig. 5a). The most significant hit for SiL_2606 was with the FlaG protein from *S. acidocaldarius* (PDB ID: 5TUH), which yielded a *z* score of 6.3. The FlaG protein is part of the archaeal flagellum[29], and FlaG shows clear homology with the archaeal flagellin FlaB, with a *z* score of 11.3 (Fig. 5b). Thus, SiL_2606 shares a

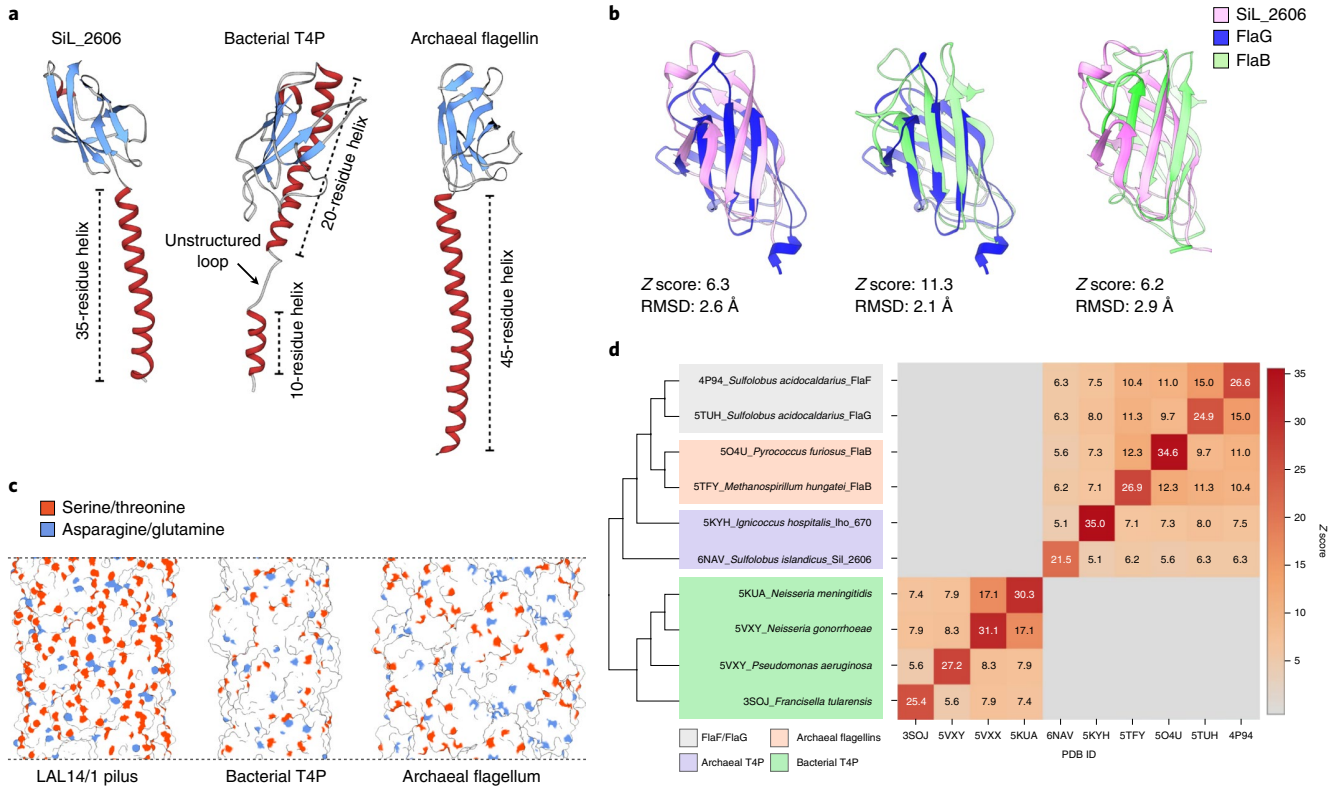


Fig. 5 | Comparison of the LAL14/1 pilus with bacterial T4P and the archaeal flagellar filament. a, Ribbon display of a single subunit of LAL14/1 pilin, T4P of *N. gonorrhoeae* (PDB ID: 5VXX) and archaeal flagellin of *M. hungatei* (PDB ID: 5TFY). **b**, Structural alignment of the globular β-sandwich domains: LAL14/1 SiL_2606, FlaG (PDB ID: 5TUH) and archaeal flagellin FlaB (PDB ID: 5TFY). *Z* scores were calculated by the Dali server. **c**, Surface displays of the filaments composed of the pilins/flagellins shown in **a**. The side chain oxygen atoms of serines and threonines are shown in orange, and the side chain nitrogen atoms of asparagines and glutamines are shown in blue. **d**, All-against-all comparison of the bacterial and archaeal structures with N-terminal T4P domains. The matrix (right) and cluster dendrogram (left) are based on the pairwise *z*-score comparisons calculated using the Dali server. The colour scale indicates the corresponding *z* scores. Based on their functions and origins, the proteins are partitioned into four groups, which are indicated with different colours.

highly similar fold with FlaG, and FlaG shares an even more similar fold with FlaB. Accordingly, SiL_2606 and FlaB also share similarities (*z* score: 6.2; Fig. 5b). However, due to differences in the angle between the β-sheets forming the β-sandwiches in all of these proteins, SiL_2606 is more similar to FlaG than it is to FlaB.

Discussion

Archaea and the viruses that infect them have evolved to survive in the most extreme environments, including nearly boiling acid[3,23]. Quite strikingly, the pili isolated from *S. islandicus* LAL14/1 were able to survive the conditions and treatments that are normally a first step to biochemical characterization of an assembly: boiling in 5 M guanidine hydrochloride, and digestion with trypsin or pepsin. The near-atomic resolution that we achieved by cryo-EM led to an atomic model for the filament, which was possible only by using a bioinformatics-based approach to search the LAL14/1 genome for all potential candidates and then using the cryo-EM map to exclude all but two proteins: SiL_2603 and SiL_2606. Both of these proteins were subsequently identified in the pilus preparation by mass spectrometry. Since it would be difficult to imagine how a proline in SiL_2603 within a β-strand would maintain the hydrogen bonding that holds the β-sheet together, we have assumed that the dominant pilin in the filament is actually SiL_2606. Subunits in a growing T4P must add at the base[16], and the disrupted β-sheet in SiL_2603 might allow the protein to bind at the base of a growing pilus but prevent additional subunits from being added. It might also be the first protein assembled by the secretory system and be found at the pilus tip.

The atomic model that we have generated for the *S. islandicus* pilus revealed an unusual surface (Fig. 3a) that raised questions about how such a filament would be soluble. In fact, knocking out a particular glycosylation pathway in an archaeon has previously been shown to cause the pili to aggregate into large bundles[47]. A similar phenomenon has been observed in bacteria, where eliminating glycosylation of pili led to greatly enhanced aggregation of these filaments[48]. The density map also showed extra mass consistent with glycosylation for three residues, and a cloud of low density surrounding the filament, suggestive of more extensive glycosylation. In addition, compared with bacterial T4P and archaeal flagella, there are significantly more serine and threonine residues present on the pilus surface (Fig. 5c)—potential targets for *O*-linked sugars. STEM showed that the mass per unit length of the filament was ~35% greater than expected for unmodified protein subunits.

Taken together, our results suggest that extensive glycosylation of the pili both allows for their solubility and provides the basis for their extreme stability. Such stability is not seen for homologous structures, which we now know to be the archaeal flagellar filaments, showing that the extensive glycosylation appears to be an evolutionary modification that gives rise to the unusual physicochemical properties. A previous study[49] showed that the glycosylation of a protein by a single fucose moiety led to stabilization

of the protein by ~1 kcal mol^{-1}, as judged by thermal denaturation. Another study[50] showed that the degree of thermal stabilization was directly proportional to the number of attached glycan chains. A review summarizes the extensive data on stabilization of proteins by glycosylation[51], which yields resistance to proteolysis, thermal denaturation, and unfolding by guanidine hydrochloride.

We have shown that the overall fold of this pilin is quite similar to that of archaeal flagellin, while previous work only suggested that this structural homology existed for the N-terminal domain[32]. The first indication that the archaeal flagellin fold might be more widespread came from the structure of an 'adhesion filament' from *I. hospitalis*[26]. This organism is non-motile and the protein (Iho_670) forming the adhesion filament showed no homology at the sequence level to archaeal flagellin outside the N-terminal T4P-like domain. It was suggested[26] that the fold of Iho_670 would be the same as in true archaeal flagellin, and this has subsequently been confirmed[28,29].

Pairwise structural comparisons of the bacterial and archaeal T4P proteins suggests a probable scenario of T4P evolution. Although the N-terminal α-helical domains of bacterial and archaeal T4Ps are homologous, the globular β-strand-rich domains do not display appreciable structural similarity (Fig. 5d). Indeed, whereas in bacterial T4P the β-strands form a rather flat sheet with a juxtaposed α-helix extending from the N-terminal domain, in all archaeal T4P, including flagellins and pilins, the C-terminal domain adopts an immunoglobulin-like fold—a β-sandwich consisting of two β-sheets that surround a central hydrophobic core (Fig. 5a). It is thus likely that the globular domains in bacterial and archaeal T4P proteins have evolved independently. Many more structures of bacterial and archaeal T4Ps will need to be determined to substantiate this hypothesis. However, it is already clear that FlaB, FlaF and FlaG components of archaeal flagella are more closely related to each other than they are to archaeal pilins (Fig. 5d), suggesting that these functionally different flagellar components have evolved from a single ancestor (probably a pilin) by gene duplication and subsequent subfunctionalization.

Remarkably, the features underlying the stability of filamentous viruses of hyperthermophilic archaea appear to be very similar to those of LAL14/1 pili. Namely, the contacts between the capsid protein subunits in the helically symmetrical virions are largely mediated by hydrophobic interactions and the virion surfaces are extensively glycosylated[3,5]. This suggests that at least some viruses and pili of hyperthermophilic archaea have converged on similar adaptive solutions, allowing them to maintain stability in extremely hot environments.

Methods

Strain cultivation and filament purification. A preculture of *S. islandicus* strain LAL14/1 (ref. [23]) was grown in rich medium[52] at 76 °C for 24 h, then diluted into fresh medium. The cells were grown until mid-logarithmic phase (OD_{600} = 0.5), collected by centrifugation (6,000*g*; 15 min; 15 °C) and resuspended in rich medium at one-tenth of the original volume. The pili were isolated as described previously[24]. Briefly, the filaments were mechanically sheared from the cells by vortexing at maximum speed for 15 min. The cells were removed by centrifugation (6,000*g*; 60 min; 15 °C). The cleared supernatant containing pili was ultracentrifuged (144,000*g*; 60 min; 15 °C) and the resultant pellet was resuspended in 20 mM Tris-acetate buffer (pH 6).

Cryo-EM and image processing. The LAL14/1 pilus sample (4 µl) was applied to discharged lacey carbon grids and plunge frozen using a Vitrobot Mark IV (FEI). Frozen grids were imaged in a Titan Krios at 300 keV and recorded with a Falcon III camera at 1.4 Å pixel^{-1}. Micrographs were collected using a defocus range of 1.5–2.5 µm, with a total exposure time of 2.4 s (amounting to ~56 electrons Å^{-2}) distributed into 24 fractions. A total of 1,824 micrograph movies were analysed. All of the micrographs were first motion corrected (ignoring the first fraction) using MotionCorr version 2.1 (ref. [53]), then used for contrast transfer function (CTF) estimation by the CTFFIND3 program[54]. After the images were corrected for the CTF through multiplication by the theoretical CTF, filament images were extracted from dose-weighted fractions 2–10 using the e2helixboxer program within EMAN2 (ref. [55]) (amounting to ~20 electrons Å^{-2}). A total of 881,252 overlapping 384-pixel-long segments (with a shift of 6 pixels between adjacent subunits, ~1.5× the axial rise per subunit) were generated. The helical symmetry was determined as a 4.9 Å rise and 105.0° rotation, after searching through a number of possible symmetries by trial and error. A reconstruction of ~4.5 Å was generated using the iterative helical real-space reconstruction method implemented in Spider[56], and this volume was subsequently filtered to 10 Å as the starting reference used in Relion[57]. The same micrographs and box coordinates used in the Spider reconstruction were imported into Relion. A comparable reconstruction of ~4.5 Å was generated after class2D and refine3D steps, which was then further improved to 4.1 Å after movie-refinement and particle-polishing steps. The final volume was estimated to have a resolution of 4.1 Å based on the model:map FSC and d_{99} (ref. [58]) (Supplementary Fig. 3 and Supplementary Table 1), and was sharpened with a negative *B* factor of −300.

De novo model building of the LAL14/1 pilus. First, the density corresponding to a single LAL14/1 pilin (pilin map) was segmented from the experimental filament density using Chimera[59]. Then, the density corresponding to the N-terminal helix (long-helix map) was further segmented from the pilin map. About 120 Cα atoms could be successfully traced from the pilin map, including an N-terminal α-helix of ~35 amino acids and a β-sandwich globular domain of ~85 amino acids. Those patterns were used to search for possible pilin candidates in the *S. islandicus* LAL14/1 proteome[23]. The proteome (n = 2,601) was downloaded from the National Center for Biotechnology Information GenBank and filtered to retain only proteins of 120–145 amino acids in length using the Galaxy platform maintained at Institut Pasteur (https://galaxy.pasteur.fr). The 263 extracted proteins were analysed for the presence of a transmembrane domain using TMHMM[60] and only those containing a single N-terminal membrane-spanning domain were retained. For the 12 proteins satisfying the above characteristics, the secondary structure was predicted using Jpred4 (ref. [61]), and the 7 proteins (SiL_0372, SiL_1076, SiL_1365, SiL_1862, SiL_2603, SiL_2606 and SiL_2610) containing β-strand-rich C-terminal domains were retained for further analyses. The first 50 amino acids of those 7 candidates were then built into the long-helix map using the RosettaCM protocol[62]. Candidates with good fit to the long-helix map were then built into the pilin map with their full sequences using the RosettaCM protocol. SiL_2603 and SiL_2606 could be successfully built into the pilin map, but SiL_2603 is extremely unlikely to be the correct candidate because it would have a proline in the middle of an ordered β-strand. Therefore SiL_2606 was selected to generate a filamentous model and refined against the full cryo-EM map using real-space refinement in PHENIX[63]. MolProbity[64] was used to evaluate the quality of the filament model. The refinement statistics are shown in Table 1.

LRO calculations. An LRO parameter was defined for a protein from the knowledge of potential contacts between two residues that are close in space (<8 Å) and far apart in sequence (>12 in amino acid register)[37]. Potential contacts between residues were computed with the pairdist tool from GROMACS package[65] using the equation $LRO = \Sigma n_{ij}/N$, where $n_{ij} = 1$ if $|i-j| > 12$ (otherwise $n_{ij} = 0$), i and j are two residues for which the Cα–Cα distance is ≤8 Å, and N is the total number of residues in a protein. The dataset consisted of two archaeal pili from *S. islandicus* LAL14/1 and *I. hospitalis*, two archaeal flagellins from *Pyrococcus furiosus* and *Methanospirillum hungatei*, two bacterial T4P from *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa*, and two bacterial flagellins from *Bacillus subtilis* and *P. aeruginosa*. Since these are all polymers, any residue in a different chain within 8 Å from a given residue was automatically assumed to be a distant residue.

STEM. STEM data were acquired at Brookhaven National Laboratory. The STEM instrument operates at 40 keV with a scanning probe of <0.3 nm diameter produced from a cold field-emission source. Specimen quality and mass calibration were checked by comparing the images with the known structure of tobacco mosaic virus (TMV; theoretical mass/length: 13.1 kDa Å^{-1}). Samples containing LAL14/1 pili were deposited on thin carbon. TMV was added to the grid first as an internal control, followed by injection buffer, then specimen solution (in 20 mM Tris/acetate buffer (pH 6.0)) for 1 min, and then 10 washes of 20 mM Tris/acetate buffer. Excess solution was wicked from the edge with filter paper between each injection. After the last wash, the grid was wicked to a thin layer, fast-frozen by plunging into liquid nitrogen slush and stored under liquid nitrogen. Grids were freeze-dried overnight in an ion-pumped chamber with an efficient cold trap and transferred under vacuum to the STEM cold stage (−160 °C). Imaging typically used a dose of 20 el Å^{-2} (causing <5% mass loss, corrected by comparison with TMV). Mass measurements were performed offline with the customized software PCMass[66].

LAL14/1 deglycosylation reaction. Initially, Protein Deglycosylation Mix II (New England Biolabs), which contains a mixture of different enzymes, was used unsuccessfully to remove the glycosylation from the LAL14/1 pili. Subsequently, deglycosylation of LAL14/1 pili was performed using the Glycoprofile IV Chemical Deglycosylation Kit (Sigma–Aldrich). Briefly, 50 µl of concentrated pili were lyophilized in a small glass tube. Then, 150 µl of TFMS was added to the tube and the mixture was incubated at 4 °C for 1 h. Following this, 150 µl of

Table 1 | Cryo-EM data collection, refinement and validation statistics

Parameter	LAL14/1 pilus model value/classification
Data collection and processing	
Magnification	59,000×
Voltage (kV)	300
Electron exposure (e^{-} Å^{-2})	20
Defocus range (μm)	−1.5 to −2.5
Pixel size (Å)	1.4
Symmetry imposed	Helical (rise: 4.94 Å; rotation: 104.97°)
Initial particle images (*n*)	881,252
Final particle images (*n*)	181,113
Map resolution (Å)	4.1
Model:map FSC (0.38)	4.0
'Gold-standard' map:map FSC (0.143) d_{99}	4.1
Refinement	
Initial model used (PDB code)	NA
Map-sharpening *B* factor (Å^2)	−300
Model composition	
Non-hydrogen atoms (*n*)	891
Protein residues (*n*)	131
Ligands	NA
***B* factors (Å^2)**	
Protein	61.4
Ligand	NA
Root-mean-square deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.823
Validation	
MolProbity score	1.52
Clashscore	4.1
Poor rotamers (%)	0
Ramachandran plot	
Favoured (%)	95.4
Allowed (%)	4.6
Disallowed (%)	0

The LAL14/1 pilus model used has the following IDs: Electron Microscopy Data Bank: 0397; PDB: 6NAV.

60% pyridine solution, cooled to ~15 °C with a methanol–dry ice bath, was added drop-wise to the reaction tube, neutralizing the TFMS acid. Using a Slide-A-Lyzer dialysis cassette with a 2,000 Da protein molecular weight cut-off (Thermo Fisher Scientific), the reaction solutions were removed from the deglycosylated pili sample with overnight dialysis at 4 °C into Tris/HCl (pH 7.8) buffer. After dialysis, small amounts of aggregates—presumably deglycosylated pili—were observed in the cassette. Centrifugation for 15 min at 4 °C and 20,000 r.p.m. was used to pellet the aggregates. The pellet was resuspended with 50 μl of Tris buffer (pH 7.8).

Analysis of deglycosylation reaction by SDS polyacrylamide gel electrophoresis (SDS-PAGE). The pellet from the deglycosylation reaction and the glycosylated LAL14/1 stock were separately diluted threefold into Tricine Sample Buffer (Bio-Rad: 200 mM Tris/HCl (pH 6.8), 40% glycerol and 0.04% Coomassie Blue G-250) containing 2% SDS. The samples were then boiled at 100 °C for 6–10 min. SDS-PAGE was then performed at a constant voltage of 120 V using 16.5% precast polyacrylamide Mini-PROTEAN Tris-Tricine precast gels (Bio-Rad) and Tris/Tricine/SDS running buffer (Bio-Rad). For each sample, 15 μl of boiled sample in SDS was added to separate wells in the gel, and 5 μl of Precision Plus Protein Dual Xtra protein standard (Bio-Rad) was added to a third well. No protein bands were detectable using standard Coomassie staining methods. As a result, more sensitive silver staining was performed using a Pierce Silver Stain for Mass Spectrometry kit (Thermo Fisher Scientific). The detected gel bands were then excised and mass spectrometry was performed to analyse the composition of each band.

Mass spectrometry. The gel pieces from the band were transferred to a siliconized tube and washed in 200 μl of 50% methanol. The gel pieces were dehydrated in acetonitrile, rehydrated in 30 μl of 10 mM dithiolthreitol in 0.1 M ammonium bicarbonate, and reduced at room temperature for 0.5 h. The dithiolthreitol solution was removed and the sample alkylated in 30 μl of 50 mM iodoacetamide in 0.1 M ammonium bicarbonate at room temperature for 0.5 h. The iodoacetamide reagent was removed and the gel pieces were dehydrated in 100 μl acetonitrile. The acetonitrile was removed and the gel pieces were rehydrated in 100 μl of 0.1 M ammonium bicarbonate. The pieces were dehydrated in 100 μl acetonitrile, the acetonitrile was removed and the pieces were completely dried by vacuum centrifugation. The gel pieces were rehydrated in 20 ng μl^{-1} chymotrypsin in 50 mM ammonium bicarbonate on ice for 30 min. Any excess enzyme solution was removed and 20 μl 50 mM ammonium bicarbonate was added. The sample was digested overnight at 37 °C and the peptides formed were extracted from the polyacrylamide in a 100 μl aliquot of 50% acetonitrile and 5% formic acid. This extract was evaporated to 15 μl for mass spectrometry analysis.

The liquid chromatography mass spectrometry system consisted of a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) with an EASY-Spray ion source (set to 40 °C) connected to a 75 μm × 15 cm (3 μm C18) EASY-Spray column. Some 3 μl of the extract was injected and the peptides were eluted from the column by an acetonitrile/0.1 M formic acid gradient at a flow rate of 0.3 μl min^{-1} over 1 h. The nanospray ion source was operated at 1.8 kV and 250 °C. The digest was analysed using the rapid-switching capability of the instrument, acquiring a full-scan mass spectrum (lock mass: 445.12006; 120 K resolution; 3e6 target; 60 ms max ion trap (IT)) to determine the peptide molecular weights, followed by product ion spectra (10 higher-energy collisional dissociation; 30 K resolution; 1e5 target; 60 ms max IT; 2.0 *m/z*; trigger 3e4; normalized collision energy 27) to determine the amino acid sequences in sequential scans. Exclusions were made for +1 and unassigned ions, as well as standard dynamic exclusion for 20 s. This mode of analysis produces approximately 20,000 tandem mass spectrometry spectra of ions ranging in abundance over several orders of magnitude.

The data were analysed by database searching using the Sequest search algorithm contained within Proteome Discoverer 2.2 against UniProt *S. Islandicus* (20 November 2018 with 16,206 entries). The parent mass was set to 10 ppm and fragment masses to 0.02 Da. The following parameters were used: chymotrypsin, fixed carbamidomethyl C, dynamic oxidation M and 1 missed cleavage. Raw data generated from the search were loaded into Scaffold 4.8.8 with initial filtering set to the following: xcorr +1 > 1.8; +2 > 2.0; +3 > 2.2; +4 > 3.0; peptide prophet > 60%; and protein prophet > 90%. The search and data analysis settings produced a false discovery rate of <1%. Any peptides meeting these criteria were manually examined by N.S. to further eliminate false positives before the data were reported. The Thermo RAW and Scaffold analysis files have been deposited into PRIDE (PXD012799).

Pro-Q Emerald 488 Glycoprotein staining. LAL14/1 pili in 20 mM Tris-acetate buffer (pH 6) were diluted threefold into Tricine Sample Buffer (Bio-Rad: 200 mM Tris/HCl (pH 6.8), 40% glycerol and 0.04% Coomassie Blue G-250) containing 2% SDS. The sample was then boiled at 100 °C for 8 min. SDS-PAGE was performed using the same conditions that were used to analyse the deglycosylation reactions. The CandyCane Glycoprotein Molecular Weight Standards (Invitrogen) were run on the gel with the LAL014/1 pili, in a separate lane. The gel was then stained with the Pro-Q Emerald 488 Glycoprotein Stain kit (Invitrogen). The Pro-Q Emerald stain was visualized using a FluorChem Q system (ProteinSimple) with a light source of 475 nm and the emission filter set to 537 nm.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The three-dimensional reconstruction has been deposited in the Electron Microscopy Data Bank with accession code EMD-0397. The atomic model has been deposited in the Protein Data Bank with accession code 6NAV. The mass spectrometry data have been deposited in PRIDE with accession code PXD012799.

Received: 23 December 2018; Accepted: 15 April 2019;
Published online: 20 May 2019

References

1. Prangishvili, D. et al. The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.* **15**, 724–739 (2017).
2. Chaudhury, P., Quax, T. E. F. & Albers, S. V. Versatile cell surface structures of archaea. *Mol. Microbiol.* **107**, 298–311 (2018).
3. DiMaio, F. et al. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* **348**, 914–917 (2015).
4. Liu, Y. et al. Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. *Nat. Commun.* **9**, 3360 (2018).
5. Ptchelkine, D. et al. Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nat. Commun.* **8**, 1436 (2017).

6. Kasson, P. et al. Model for a novel membrane envelope in a filamentous hyperthermophilic virus. *eLife* **6**, e26268 (2017).
7. Veesler, D. et al. Atomic structure of the 75 MDa extremophile *Sulfolobus* turreted icosahedral virus determined by CryoEM and X-ray crystallography. *Proc. Natl Acad. Sci. USA* **110**, 5504–5509 (2013).
8. Berry, J. L. & Pelicic, V. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol. Rev.* **39**, 134–154 (2015).
9. Makarova, K. S., Koonin, E. V. & Albers, S. V. Diversity and evolution of type IV pili systems in archaea. *Front. Microbiol.* **7**, 667 (2016).
10. Pohlschroder, M., Ghosh, A., Tripepi, M. & Albers, S. V. Archaeal type IV pilus-like structures—evolutionarily conserved prokaryotic surface organelles. *Curr. Opin. Microbiol.* **14**, 357–363 (2011).
11. Pohlschroder, M. & Esquivel, R. N. Archaeal type IV pili and their involvement in biofilm formation. *Front. Microbiol.* **6**, 190 (2015).
12. Albers, S. V. & Jarrell, K. F. The archaellum: an update on the unique archaeal motility structure. *Trends Microbiol.* **26**, 351–362 (2018).
13. 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 (2017).
14. Kolappan, S. et al. Structure of the *Neisseria meningitidis* type IV pilus. *Nat. Commun.* **7**, 13015 (2016).
15. Lopez-Castilla, A. et al. Structure of the calcium-dependent type 2 secretion pseudopilus. *Nat. Microbiol.* **2**, 1686–1695 (2017).
16. Craig, L. et al. Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Mol. Cell* **23**, 651–662 (2006).
17. Craig, L. et al. Type IV pilin structure and assembly: X-ray and EM analyses of *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol. Cell* **11**, 1139–1150 (2003).
18. Hartung, S. et al. Ultrahigh resolution and full-length pilin structures with insights for filament assembly, pathogenic functions, and vaccine potential. *J. Biol. Chem.* **286**, 44254–44265 (2011).
19. Reardon, P. N. & Mueller, K. T. Structure of the type IVa major pilin from the electrically conductive bacterial nanowires of *Geobacter sulfurreducens*. *J. Biol. Chem.* **288**, 29260–29266 (2013).
20. Szabo, Z. et al. Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases. *J. Bacteriol.* **189**, 772–778 (2007).
21. Albers, S. V., Szabo, Z. & Driessen, A. J. Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J. Bacteriol.* **185**, 3918–3925 (2003).
22. Strom, M. S., Nunn, D. N. & Lory, S. A single bifunctional enzyme, PilD, catalyzes cleavage and *N*-methylation of proteins belonging to the type IV pilin family. *Proc. Natl Acad. Sci. USA* **90**, 2404–2408 (1993).
23. Jaubert, C. et al. Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. *Open Biol.* **3**, 130010 (2013).
24. Quemin, E. R. et al. First insights into the entry process of hyperthermophilic archaeal viruses. *J. Virol.* **87**, 13379–13385 (2013).
25. DiMaio, F. et al. Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015).
26. Braun, T. et al. Archaeal flagellin combines a bacterial type IV pilin domain with an Ig-like domain. *Proc. Natl Acad. Sci. USA* **113**, 10352–10357 (2016).
27. 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 (2012).
28. Poweleit, N. et al. CryoEM structure of the *Methanospirillum hungatei* archaellum reveals structural features distinct from the bacterial flagellum and type IV pili. *Nat. Microbiol.* **2**, 16222 (2016).
29. Daum, B. et al. Structure and in situ organisation of the *Pyrococcus furiosus* archaellum machinery. *eLife* **6**, e27470 (2017).
30. Bayley, D. P. & Jarrell, K. F. Further evidence to suggest that archaeal flagella are related to bacterial type IV pili. *J. Mol. Evol.* **46**, 370–373 (1998).
31. 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 (2006).
32. Faguy, D. M., Jarrell, K. F., Kuzio, J. & Kalmokoff, M. L. Molecular analysis of archael flagellins: similarity to the type IV pilin-transport superfamily widespread in bacteria. *Can. J. Microbiol.* **40**, 67–71 (1994).
33. Song, Y. et al. High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).
34. Henche, A. L. et al. Structure and function of the adhesive type IV pilus of *Sulfolobus acidocaldarius*. *Environ. Microbiol.* **14**, 3188–3202 (2012).
35. Kryndushkin, D., Pripuzova, N., Burnett, B. & Shewmaker, F. Non-targeted identification of prions and amyloid-forming proteins from yeast and mammalian cells. *J. Biol. Chem.* **288**, 27100–27111 (2013).
36. Nelson, R. et al. Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
37. Gromiha, M. M. & Selvaraj, S. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J. Mol. Biol.* **310**, 27–32 (2001).
38. Broom, A. et al. Designed protein reveals structural determinants of extreme kinetic stability. *Proc. Natl Acad. Sci. USA* **112**, 14605–14610 (2015).
39. Stranges, P. B. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22**, 74–82 (2013).
40. Pohlschroder, M., Pfeiffer, F., Schulze, S. & Halim, M. F. A. Archaeal cell surface biogenesis. *FEMS Microbiol. Rev.* **42**, 694–717 (2018).
41. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255–261 (2003).
42. Wall, J. S. & Hainfeld, J. F. Mass mapping with the scanning transmission electron microscope. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 355–376 (1986).
43. Sojar, H. T. & Bahl, O. P. Chemical deglycosylation of glycoproteins. *Methods Enzymol.* **138**, 341–350 (1987).
44. Ellen, A. F. et al. Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. *Extremophiles* **13**, 67–79 (2009).
45. Altschul, S. F. & Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
46. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
47. Esquivel, R. N., Schulze, S., Xu, R., Hippler, M. & Pohlschroder, M. Identification of *Haloferax volcanii* pilin *N*-glycans with diverse roles in pilus biosynthesis, adhesion, and microcolony formation. *J. Biol. Chem.* **291**, 10602–10614 (2016).
48. Gault, J. et al. *Neisseria meningitidis* type IV pili composed of sequence invariable pilins are masked by multisite glycosylation. *PLoS Pathog.* **11**, e1005162 (2015).
49. Mer, G., Hietter, H. & Lefevre, J. F. Stabilization of proteins by glycosylation examined by NMR analysis of a fucosylated proteinase inhibitor. *Nat. Struct. Biol.* **3**, 45–53 (1996).
50. Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl Acad. Sci. USA* **105**, 8256–8261 (2008).
51. Sola, R. J. & Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J. Pharm. Sci.* **98**, 1223–1245 (2009).
52. Zillig, W. et al. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. *Syst. Appl. Microbiol.* **16**, 609–628 (1993).
53. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
54. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
55. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
56. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
57. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
58. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).
59. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
60. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
61. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
62. Wang, R. Y. et al. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* **12**, 335–338 (2015).
63. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
64. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
65. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
66. Wall, J. S. & Simon, M. N. Scanning transmission electron microscopy of DNA–protein complexes. *Methods Mol. Biol.* **148**, 589–601 (2001).

Acknowledgements

This work was supported by NIH grants GM122510 (to E.H.E.) and GM123089 (to F.D.), as well as l'Agence Nationale de la Recherche project ENVIRA (ANR-17-CE15-0005-01 to M.K.). M.A.B.K. was supported by NIH grant T32 GM080186. The cryo-EM imaging conducted at the Molecular Electron Microscopy Core facility at the

University of Virginia was supported by the School of Medicine and built with NIH grant G20-RR31199. The Titan Krios and Falcon II direct electron detectors were obtained with NIH grants S10-RR025067 and S10-OD018149, respectively. We thank V. Conticello for the suggestion of TFMS. We are also grateful to the Ultrastructural BioImaging (UTechS UBI) unit of Institut Pasteur for access to electron microscopes.

Author contributions

V.C.-K. isolated and purified the pili. J.S.W. performed the STEM analysis. N.S. performed the mass spectrometry. G.A.P.d.O. performed the amyloid assays. F.D. performed the interfacial analysis. F.W. performed the cryo-EM, three-dimensional reconstruction and model building, with assistance from T.O. and E.H.E. M.K. and Z.S. performed the bioinformatics analysis. M.A.B.K. performed the TFMS deglycosylation. D.P., M.K. and E.H.E. designed the project. F.W., D.P., M.K. and E.H.E. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0458-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.K. or E.H.E.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature microbiology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41564-019-0458-x

In the format provided by the authors and unedited.

An extensively glycosylated archaeal pilus survives extreme conditions

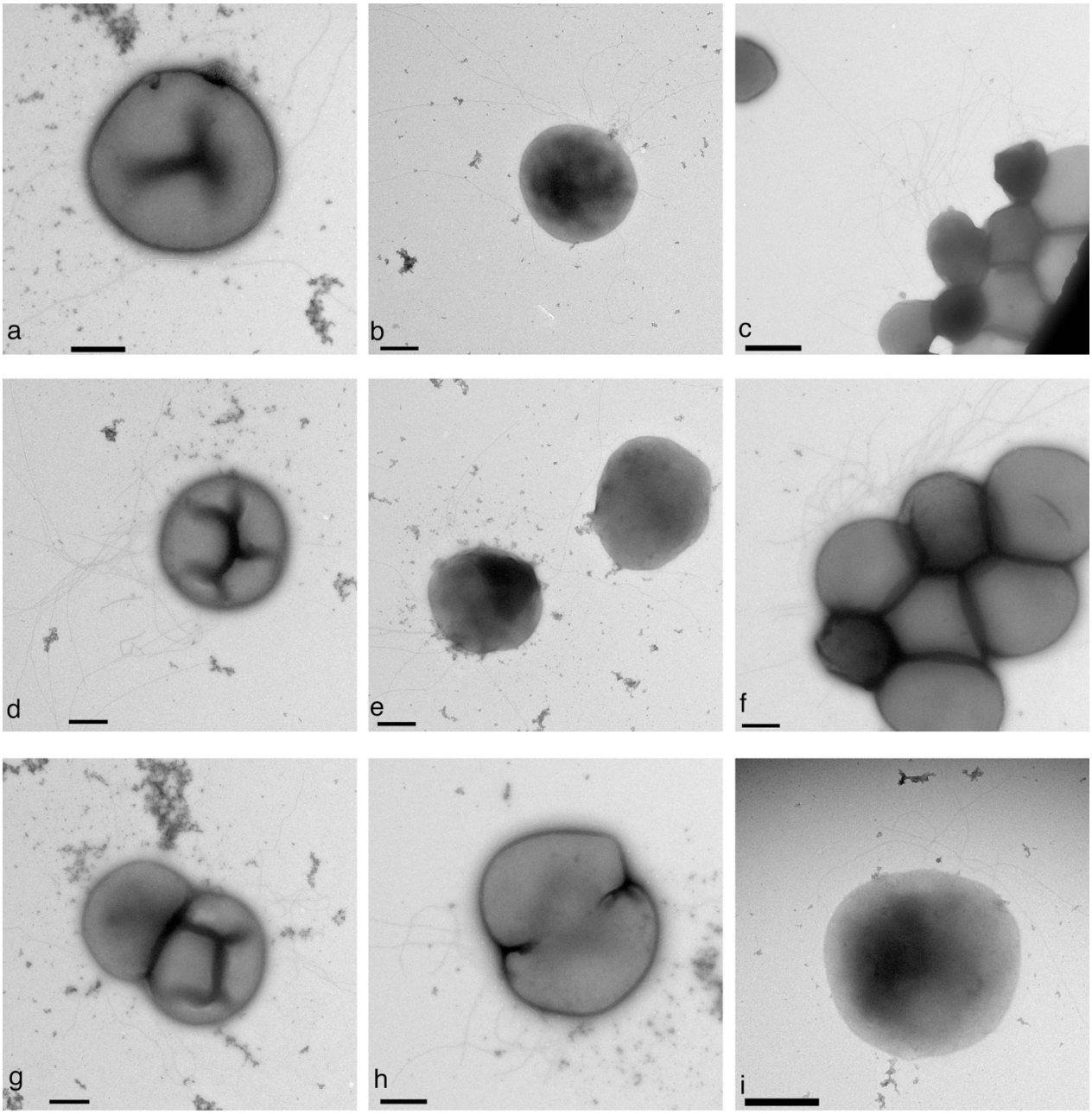
Fengbin Wang[1], Virginija Cvirkaite-Krupovic[2], Mark A. B. Kreutzberger[1], Zhangli Su[1], Guilherme A. P. de Oliveira[1], Tomasz Osinski[1], Nicholas Sherman[3], Frank DiMaio[4], Joseph S. Wall[5], David Prangishvili[2], Mart Krupovic[2*] and Edward H. Egelman[1*]

[1]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA. [2]Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, Paris, France. [3]Department of Microbiology, Immunology and Cancer Biology, University of Virginia, Charlottesville, VA, USA. [4]Department of Biochemistry, University of Washington, Seattle, WA, USA. [5]Brookhaven National Laboratory, Upton, NY, USA. *e-mail: krupovic@pasteur.fr; egelman@virginia.edu

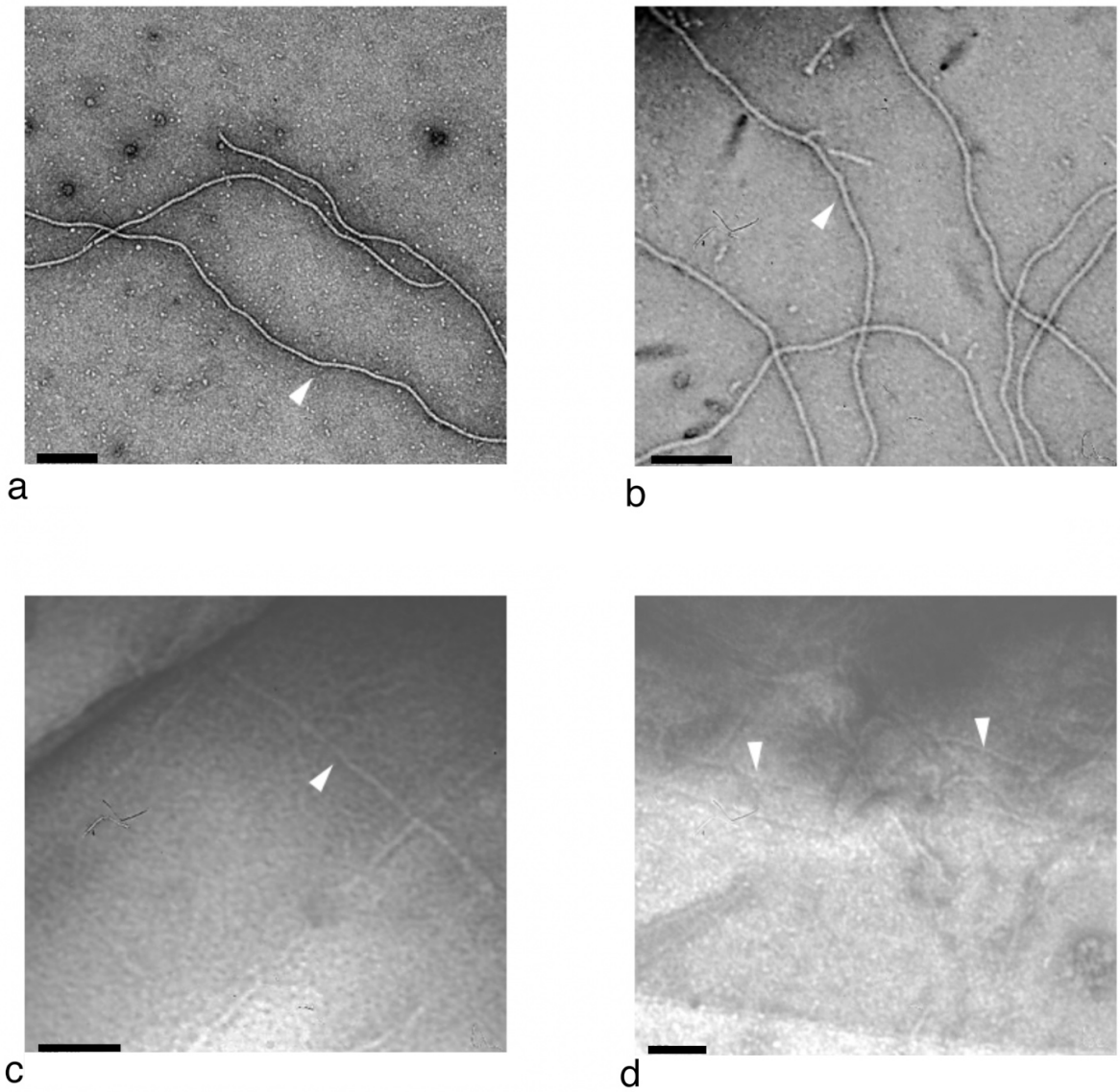
Supplementary Table 1. Comparative analysis of interfaces

	Interface SASA	Hydrophobic SASA	% Hydro-phobic	Interface Energy	Interface Energy/ SA	Interface Energy/ Residue	Interfacial H-bonds	Shape Compl.	Etotal	E/Res
Sil_2606	**9017**	**6244**	**0.692**	**-332.8**	**-0.0369**	**-4.200**	**14**	**0.659**	**-913.047**	**-6.970**
5o4u	13136	8018	0.610	-419.6	-0.0319	-3.576	34	0.682	-1168.204	-5.643
5tfy	12431	7847	0.631	-436.0	-0.0351	-4.188	28	0.620	-1092.933	-6.664
5kyh	11906	7393	0.621	-357.0	-0.0300	-3.393	40	0.548	-1480.307	-6.117
5vxx	8748	5411	0.619	-283.6	-0.0324	-3.335	22	0.592	-852.249	-5.463
5vxy	9234	5915	0.641	-252.5	-0.0273	-2.816	20	0.575	-774.445	-5.378
5wjt	17563	9307	0.530	-539.4	-0.0307	-3.429	80	0.671	-1610.120	-5.332
5wk6	15772	8565	0.543	-491.0	-0.0311	-2.786	66	0.645	-1442.300	-5.362

A comparison of structural features of the LAL14/1 Sil_2606 filament with seven other assemblies: archaeal flagellar filaments (5o4u, 5tfy), an archaeal flagellar-like filament (5kyh), bacterial T4P (5vxx, 5vxy) and bacterial flagellar filaments (5wjt, 5wk6). We report for each: (cols 2-3) the total and hydrophobic surface area covered in the assembled complex, per monomer; (col4) the fraction of the interface that is hydrophobic; (col. 5-7) the Rosetta predicted interface energy per subunit in kcal/mol, normalized by buried surface area and by number of interfacial residues; (col. 8) the number of interface-spanning hydrogen bonds; (col. 9) interface shape complementarity; and (col 10-11) the total complex energy, per subunit, normalized by the number of residues.



Supplementary Figure 1 Images of pili on the surface of *S. islandicus* cells, visualized by negative stain TEM. The scale bars are all 500 nm except in **c**, where it is 1 μ. The nine representative images shown were taken from hundreds of observations.



Supplementary Figure 2 Negative staining of the LAL14/1 pilus

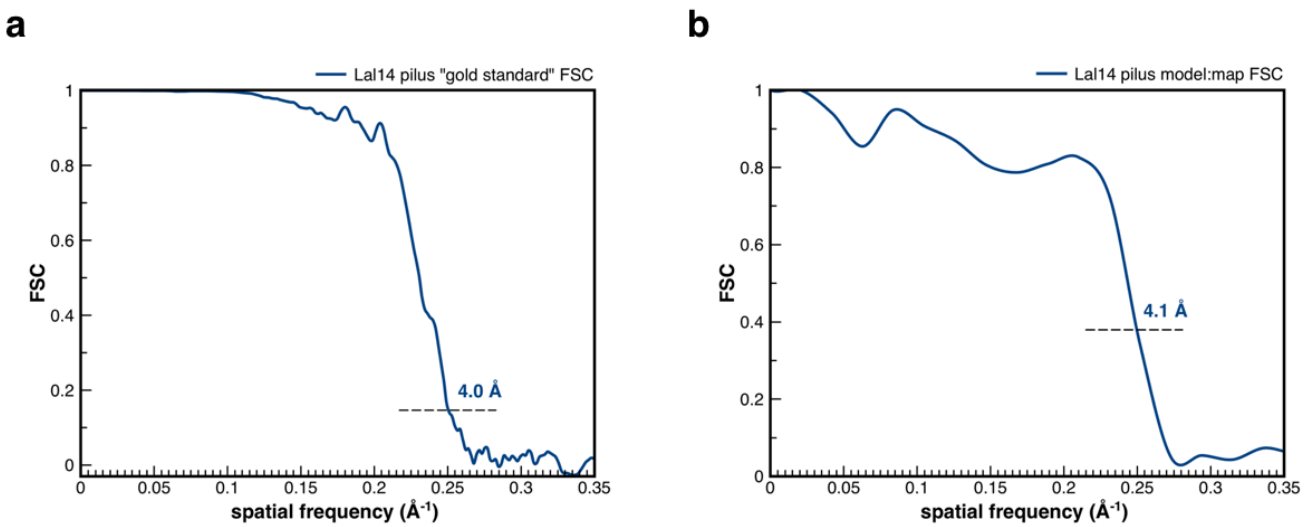
a, Negative staining of LAL14/1 pilus in 20 mM Tris-Acetate buffer pH 6.0. Scale bar, 200 nm.

b, Negative staining of LAL14/1 pilus after boiling in 5M guanidinium-HCl for 30 min. Scale bar, 200 nm.

c, Negative staining of LAL14/1 pilus after pepsin digestion for 7 days at room temperature. Scale bar, 100 nm.

d, Negative staining of LAL14/1 pilus after boiling in 100 mM NaCl, 1% SDS for 10 min. Scale bar, 100 nm.

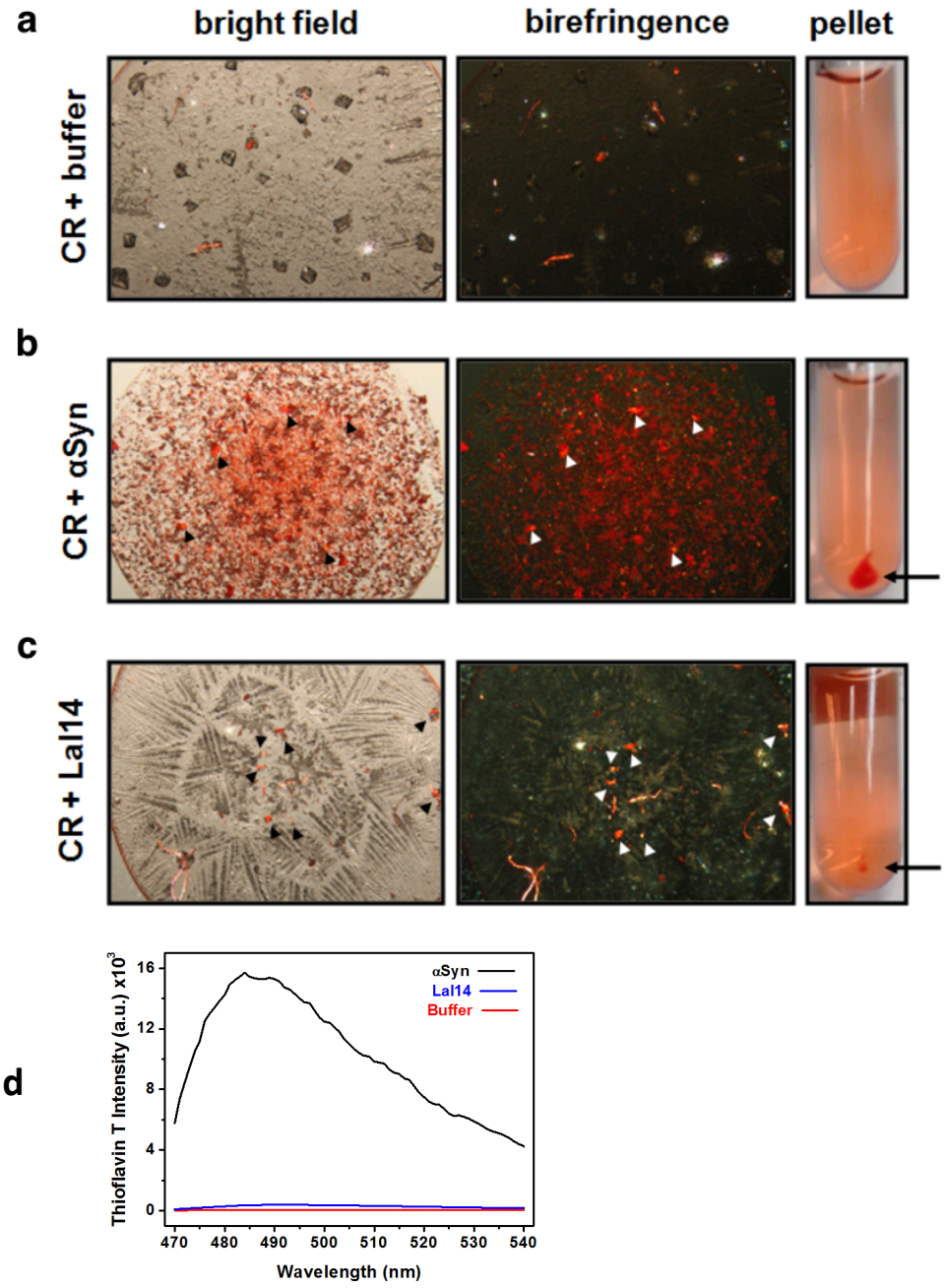
The experiments shown in (b-d) were each done once, while the control (a) is representative of hundreds of observations from multiple preparations.



Supplementary Figure 3 Fourier Shell Correlation (FSC) calculations

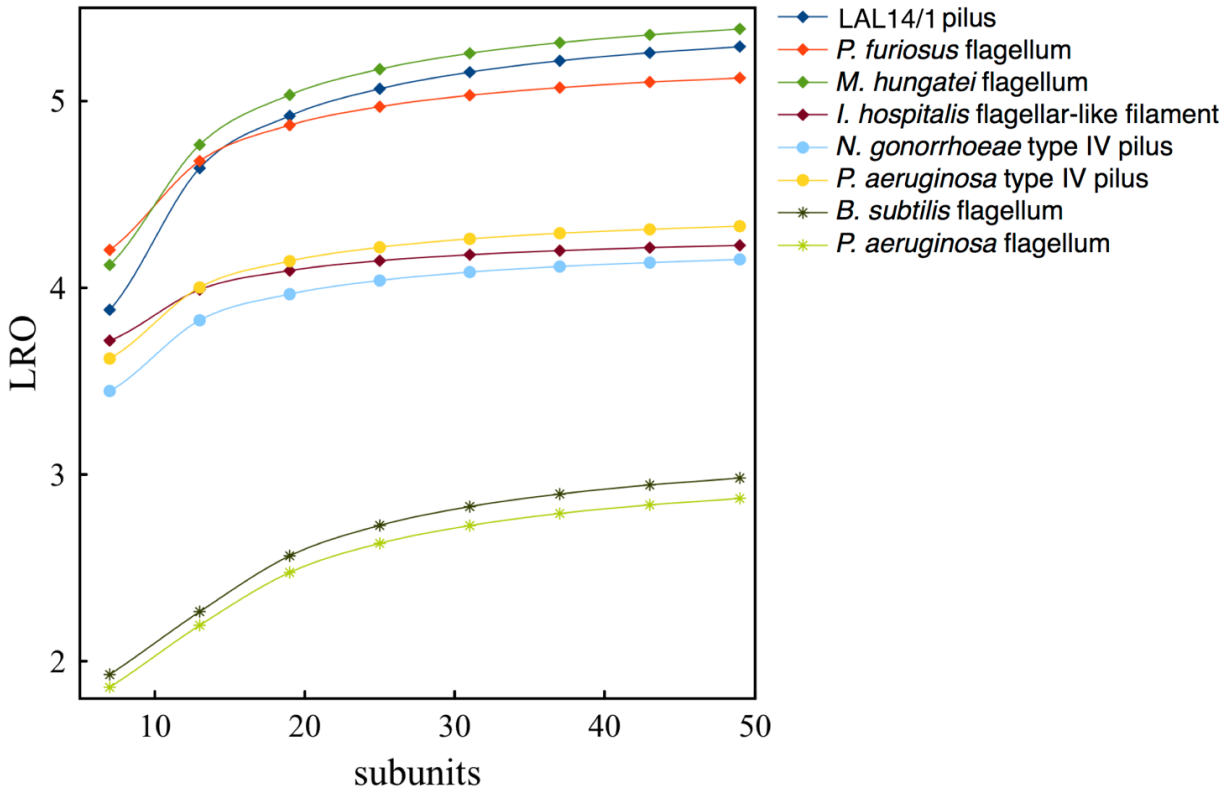
a, The map:map "gold standard" FSC using the 0.143 criterion estimates the final reconstruction map to have a resolution of 4.0 Å

b, The model:map FSC calculation using a 0.38 criterion, which is sqrt (0.143), estimates the final reconstruction map to have a resolution of 4.1 Å

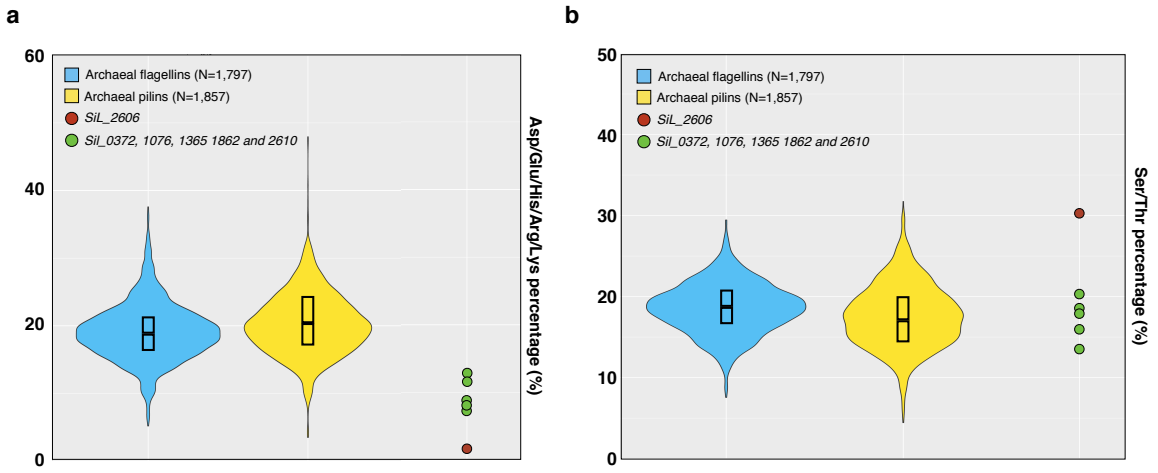


Supplementary Figure 4 LAL14/1 filaments exhibit neither Congo red (CR) birefringence nor Thioflavin T (ThT) binding.

a-c, CR binding examined under bright field, birefringence, and pellets after centrifugation. The three reactions are negative control (buffer, **a**), positive control (α-synuclein amyloid fibrils, **b**) and LAL14/1 pilus (**c**). For **a-c**, the experiments were each repeated at least twice. The binding of ThT (**d**) is not significantly greater than the buffer control (the results shown are the average of three independent experiments).



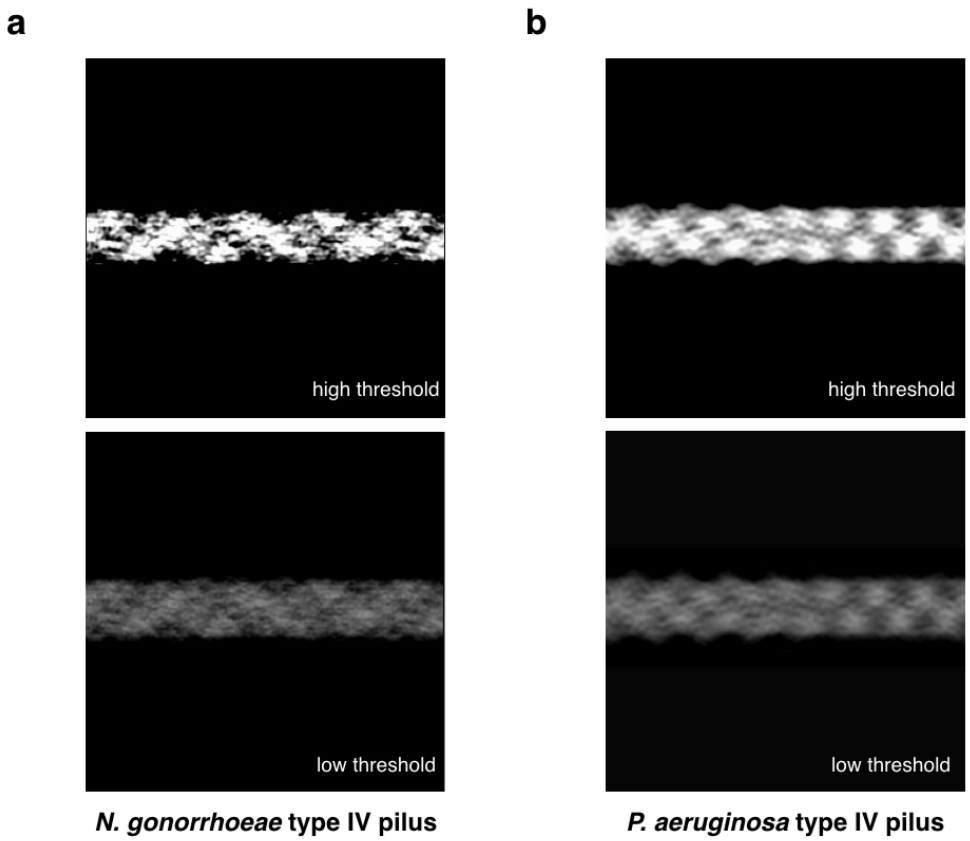
Supplementary Figure 5 Long range order score for LAL14/1 pilus and several other assemblies. Those assemblies include two bacterial type IV pili (*Neisseria gonorrhoeae* and *Pseudomonas aeruginosa*), two bacterial flagellar filaments (*Bacillus subtilis* and *Pseudomonas aeruginosa*), archaeal flagellar filaments (*Pyrococcus furiosus* and *Methanospirillum hungatei*) and an archaeal flagellar-like filament (*Ignicoccus hospitalis*). Because of the polymeric nature of the filaments, residues in one chain will be contacting residues in other chains, and the simple metric described for a single-chain protein cannot be applied. We therefore looked at the asymptotic values as one adds more and more subunits to a filament.



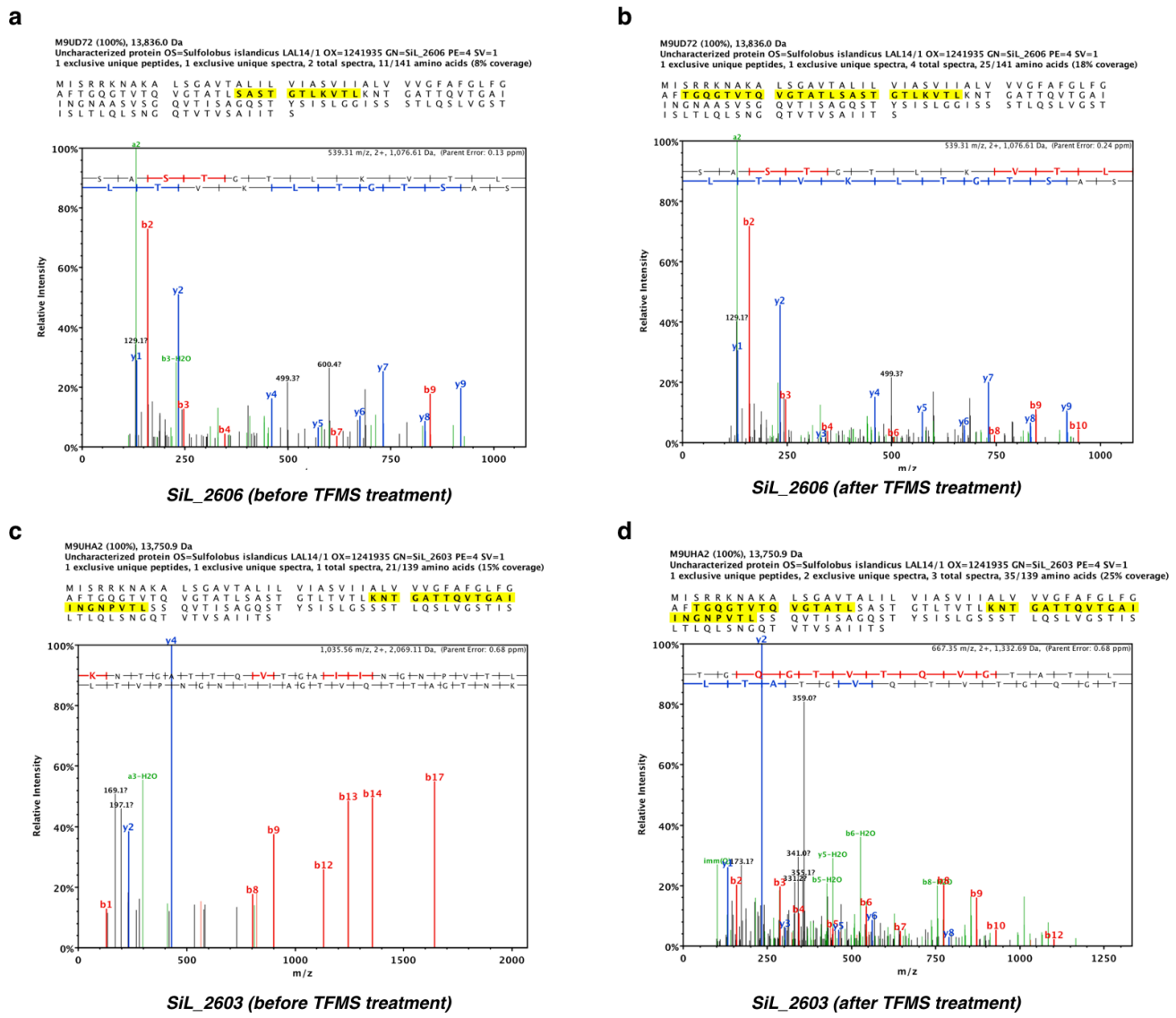
Supplementary Figure 6 Violin plots

a, Violin plot showing the distribution of the percentage of charged residues (Asp+Glu+Lys+Arg+His) in annotated archaeal flagellins (N=1,797, colored in blue) and archaeal pilins (N=1,857, colored in yellow). The width of the distribution corresponds to the frequency of occurrence. The box indicates the interquartile range, and the central line in the box indicates the median of the data. The red solid circle indicates the percentage in LAL14/1 pilin Sil_2606. The green solid circles indicate the percentage for other pilin candidates shown in Fig. 2A.

b, Violin plot showing the sum of serine and threonine percentage in annotated archaeal flagellins (N=1,797, colored in blue) and archaeal pilins (N=1,857, colored in yellow). The width of the distribution corresponds to the frequency of occurrence. The box indicates the interquartile range, and the central line in the box indicates the median of the data. The red solid circle indicates the percentage in LAL14/1 pilin Sil_2606. The green solid circles indicate the percentage of other pilin candidates shown in Fig. 2A.



Supplementary Figure 7 Projections of the 3D reconstruction of bacterial type IV pili. In contrast to the LAL14/1 pilus, these reconstructions do not show a fuzzy coat when displayed at low threshold.
a, Projections of the 3D reconstruction of *N. gonorrhoeae* type IV pilus under high threshold (top) and low threshold (bottom).
b, Projections of the 3D reconstruction of *P. aeruginosa* type IV pilus under high threshold (top) and low threshold (bottom).



Supplementary Figure 8 MS/MS analysis.

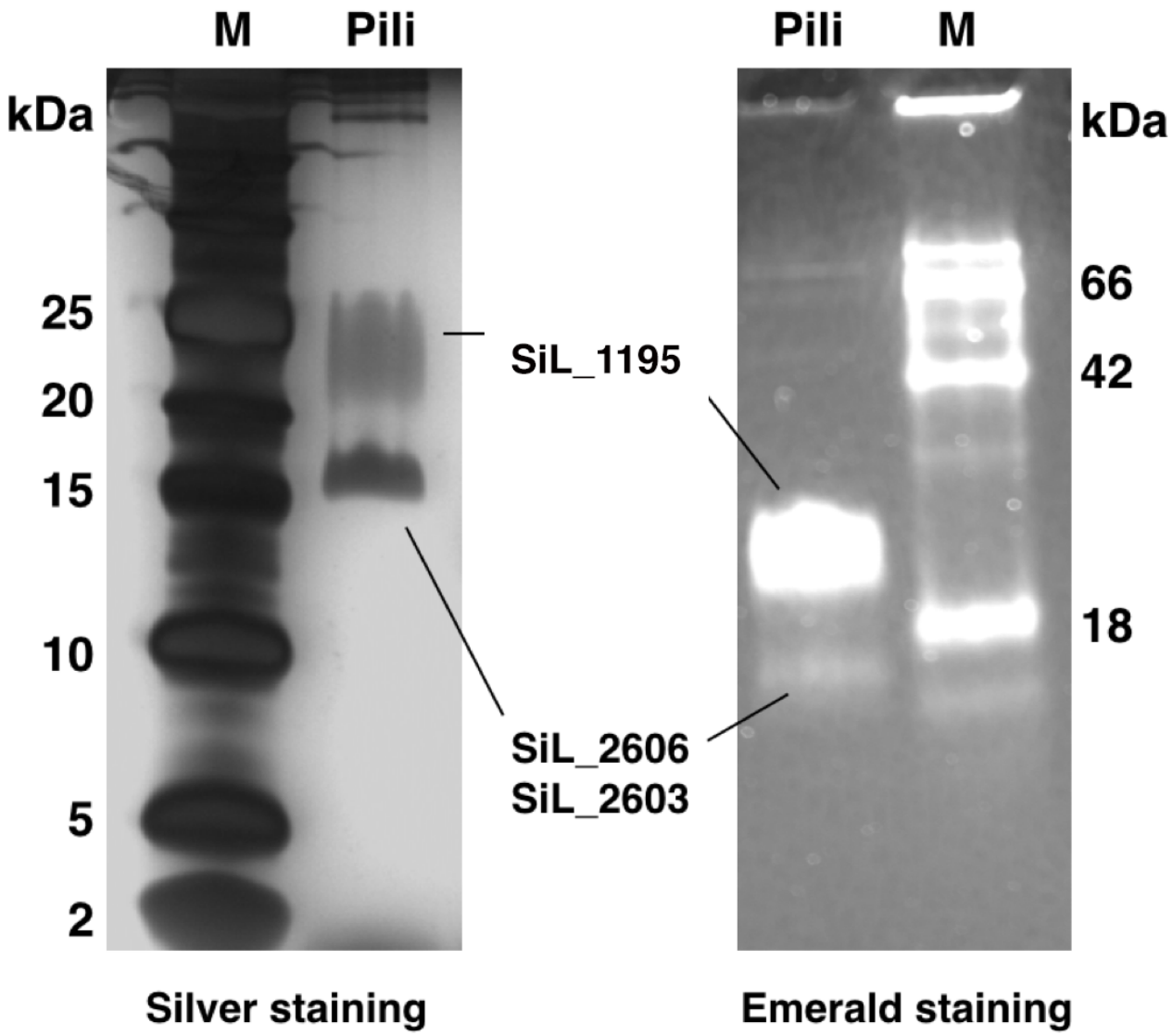
a-b, Unique peptide detection, coverage and spectrum of protein SiL_2606 before (**a**) and after (**b**) TFMS deglycosylation treatment.

c-d, Unique peptide detection, coverage and spectrum of protein SiL_2603 before (**c**) and after (**d**) TFMS deglycosylation treatment.

Representative spectra are shown.

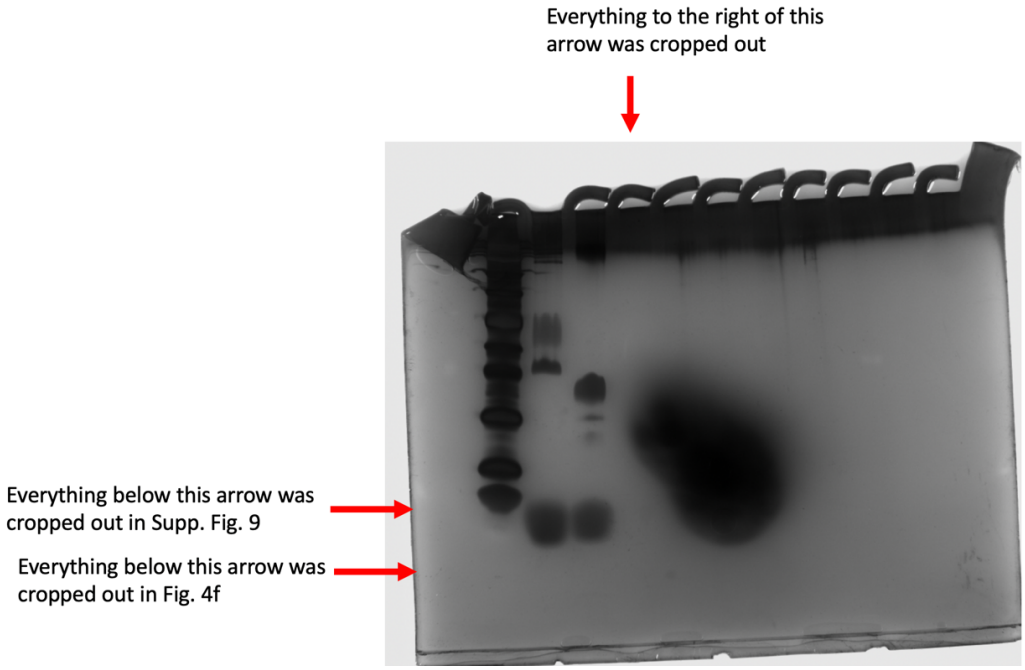
For gel band A – After TFMS treatment – 2603 (1 exclusive unique peptide, 2 exclusive unique spectra, 3 total spectra), 2606 (1 exclusive unique peptide, 1 exclusive unique spectra, 4 total spectra)

For gel band B – Before TFMS treatment – 2603 (1 exclusive unique peptide, 1 exclusive unique spectra, 1 total spectra), 2606 (1 exclusive unique peptide, 1 exclusive unique spectra, 1 total spectra)

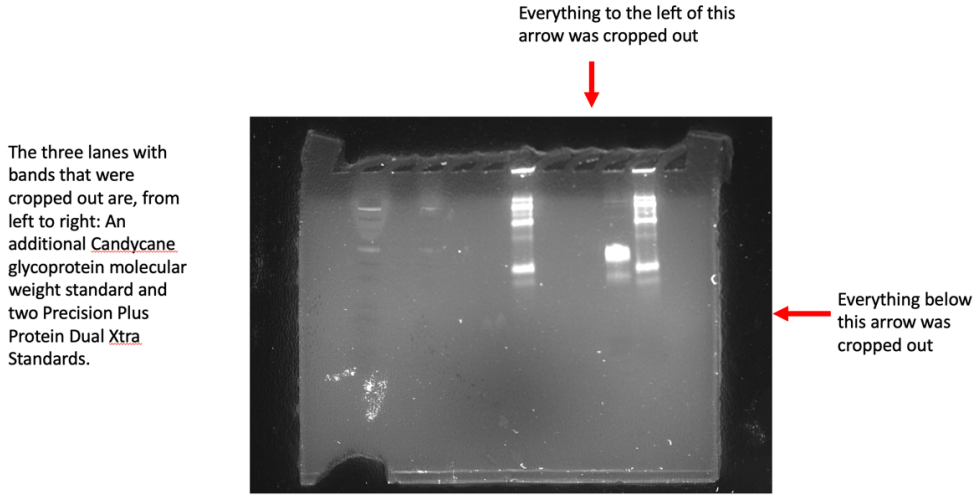


Supplementary Figure 9 Glycostaining of the pili.

A ProQ Emerald 488 stained gel is shown on the right, while a corresponding silver stained gel (from Fig. 4f, duplicated here for purposes of comparison) is shown on the left. The higher MW band was identified as protein SiL_1195 (CdvB2), a paralog of CdvB (ESCRT-III). This experiment was performed twice, with similar results each time.



For Fig. 4f and Supp. Fig. 9



For Supp. Fig. 9

Supplementary Figure 10 Full length gels