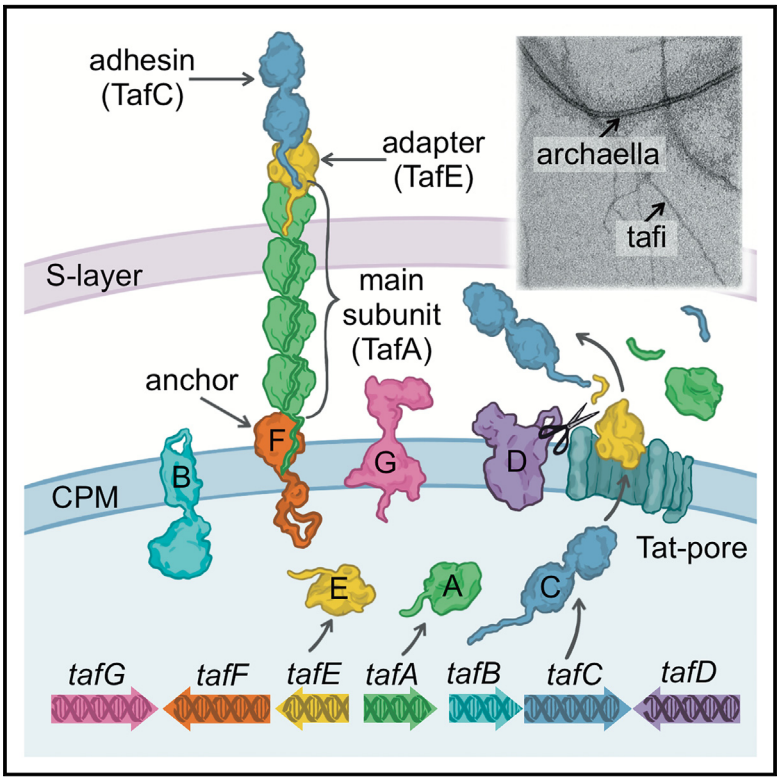
iScience

Article

Tat-fimbriae ("tafi"): An unusual type of haloarchaeal surface structure depending on the twin-arginine translocation pathway

Graphical abstract



Highlights

- Study reveals unusual prokaryote surface fimbrial appendages
- Fimbriae are composed of protein subunits secreted through the twin-arginine pathway
- Knocking out the key components of the fimbria gene cluster prevents fimbrial synthesis
- *In silico* predictions suggest fimbrial subunits to form oligomeric linear structures

Authors

Anna V. Galeva, Dahe Zhao, Alexey S. Syutkin, ..., Jingfang Liu, Hua Xiang, Mikhail G. Pyatibratov

Correspondence

xiangh@im.ac.cn (H.X.), bratov@vega.protres.ru (M.G.P.)

In brief

Molecular biology; Microbiology; Cell biology

Galeva et al., 2025, iScience *28*, 111793
February 21, 2025
https://doi.org/10.1016/j.isci.2025.111793

CellPress

iScience

CellPress
OPEN ACCESS

Article

Tat-fimbriae ("tafi"): An unusual type of haloarchaeal surface structure depending on the twin-arginine translocation pathway

Anna V. Galeva,[1,10] Dahe Zhao,[2,10] Alexey S. Syutkin,[1] Marina Yu Topilina,[1] Sergei Yu Shchyogolev,[3] Elena Yu Pavlova,[1] Olga M. Selivanova,[1] Igor I. Kireev,[4] Alexey K. Surin,[1,5,6] Gennady L. Burygin,[3,7] Jingfang Liu,[2] Hua Xiang,[2,8,9,*] and Mikhail G. Pyatibratov[1,11,*]

[1]Institute of Protein Research, Russian Academy of Sciences, Institutskaya st. 4, Pushchino, Moscow Region 142290, Russia
[2]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
[3]Institute of Biochemistry and Physiology of Plants and Microorganisms, Saratov Scientific Centre of the Russian Academy of Sciences, Prospekt Entuziastov 13, Saratov 410049, Russia
[4]A.N. Belozersky Institute of Physico-chemical Biology, M.V. Lomonosov Moscow State University, Leninskie Gori 1, Bldg 40, Moscow 119234, Russia
[5]Branch of the Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Prospekt Nauki 6, Pushchino, Moscow Region 142290, Russia
[6]State Research Center for Applied Microbiology & Biotechnology, Obolensk, Serpukhov District, Moscow Region 142279, Russia
[7]Vavilov Saratov State Agrarian University, 1 Teatralnaya Ploshchad, Saratov 410012, Russia
[8]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China
[9]Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
[10]These authors contributed equally
[11]Lead contact
*Correspondence: xiangh@im.ac.cn (H.X.), bratov@vega.protres.ru (M.G.P.)
https://doi.org/10.1016/j.isci.2025.111793

SUMMARY

The surface structures of archaeal cells, many of which exist at high temperatures, high salinity, and non-physiological pH, are key factors for their adaptation to extreme living conditions. In the haloarchaeon *Haloarcula hispanica*, we have discovered a thin filamentous surface appendage called tat-fimbriae ("tafi"), which were identified to be composed of three protein subunits, TafA, TafC, and TafE, among which TafA is the major fimbrial subunit. Molecular genetic evidence demonstrates TafA was transported through the twin-arginine translocation pathway (Tat-pathway). Based on protein structure prediction (including AlphaFold 3), tafi exhibits a linear structure: TafC at the tip, TafE acting as an adapter, TafA forming the core filament, and they link the fourth subunit TafF, anchoring tafi to the cell wall. To our knowledge, this is the first case that the Tat-pathway has been linked to the secretion of protein subunits forming prokaryotic filamentous structures.

INTRODUCTION

Interest in the study of Archaea, a third domain of life, has increased in recent years. Previously unknown species of archaea are discovered regularly, many of which are extremophiles that exist in the harshest environments on Earth, such as salt lakes, hot springs, and places with extreme pH levels. Archaea have now also been found in oceans, soil, and even the human gut and skin.[1–3] Archaea have developed various mechanisms to survive and adapt to harsh environments. Their surface structures help the archaea survive in extreme conditions by providing support, attachment, and communication with other cells and surfaces.[4–7]

Archaella and pili are widely distributed in Archaea, whereas other types of cellular appendages are less common. Specific surface structures, such as Iho670 fibers, Mth60 fimbriae, hami, cannulae, and bindosomes, have been found in only a few archaeal species. These structures perform various functions, including adhesion to biotic and abiotic surfaces and cell-cell interaction.[4,5] They also play a role in biofilm formation.[6] The names of these structures are either derived from the names of the archaeal species that produce them or from their structural and functional characteristics.

Archaella, similar to bacterial flagella, allows cell motility. These organelles enable cells to swim and move within their habitat. However, archaella and bacterial flagella differ fundamentally in their structure, assembly process, and origin. Archaella have many similarities with bacterial type IV pili (T4Ps).[4,8] In addition to its main function, archaellum can also mediate other functions such as adhesion to abiotic surfaces, formation of cell-to-cell contacts, and, possibly, intercellular communication.[8]

Archaeal pili, which have been discovered to date, have been shown to play a role in adhesion, cell aggregation, biofilm formation, and DNA transfer.[9] The two most studied types of

archaeal pili are adhesive pili (Aap) and UV-inducible pili (Ups pili). Ups pili are synthesized when cells are exposed to UV radiation and have been shown to play a role in cell aggregation. This process allows cells to exchange genetic material and repair damaged DNA.[6]

Archaella, Aap, and Ups pili are evolutionarily similar to T4P. The ArlI and ArlJ proteins responsible for archaellum assembly are homologous to the type IV pilus components PilB and PilC, respectively. The precursors of protein subunits in these filaments (archaellins and pilins) contain a type IV prepilin-like signal peptide that is processed by an archaeal type IV prepilin peptidase (ArlK/PibD), which is homologous to the bacterial type IV prepilin peptidase PilD.[6]

Iho670 fibers found in the archaeon *Ignicoccus hospitalis* are involved in adhesion and are composed of the major protein Igni_0670.[10] These fibers have been shown to be structurally similar to T4Ps.[11,12]

Bindosomes have been identified in *Sulfolobus solfataricus* cells. These structures are thought to be integrated into the cell wall, helping to bind saccharides from the outside environment. The proteins that bind sugar, which are produced as precursors, have a class III signal peptide, similar to that of archaellins and type IV pilins. The functional expression of these sugar-binding proteins on the cell surface depends on the bindosome assembly system (Bas), which is similar to the bacterial T4P assembly system.[13]

Mth60 fimbriae from *Methanothermobacter thermoautotrophicus* are approximately 5 nm in diameter and 10 μm in length. These structures are significant for adhesion, as planktonic cells typically have only 1–2 fibers, while cells attached to a surface are covered in many fimbriae. Mth60 fimbriae are composed of the 15 kDa Mth60 protein, which is likely the only component of the fimbrial structure, as it is able to form fibers similar to the native ones *in vitro*. It is likely that mature Mth60-fimbrin has been processed, as predicted by the SignalP program. Its sequence contains a Sec signal peptide, and the cleavage site is located between Ala-33 and Ala-34. No patterns characteristic of archaellin/pilin signal peptides or twin-arginine translocation (Tat) were found.[7,14]

Hami were found in uncultivated euryarchaeal coccus SM1 living in cold (~10°C) sulphidic springs as a part of archaeal-bacterial community. These filamentous structures have a tripartite, barbed grappling hook at their distal end, which has been proposed to be called the "hamus." These hami provide strong cell adhesion to both various abiotic surfaces and bacteria in their community.[15] The sequence of the hamus subunit protein (97 kDa) does not show any similarity to currently known proteins involved in the formation of microbial fibers, pili, or archaella. However, it does have some similarities with known archaeal S-layer proteins. It is hypothesized that individual hamus subunits are produced in the cytoplasm and transported across the inner membrane using the Sec pathway and then assembled in the periplasmic space between the inner and outer membranes.[16]

The hyperthermophilic archaeon *Pyrodictium* grows as a network of extracellular tubules called cannulae. These structures are highly resistant to high temperatures and denaturing agents such as 2% sodium dodecyl sulfate. Cryo-EM analysis has shown that cannulae enter the periplasmic space, but not the cytoplasm of cells.[17] It has been found that the cannulae protein subunit CanA has no known homologues in other organisms. CanA contains the Sec signal pattern and a characteristic signal sequence of 25 amino acids.[18]

Thus, archaella and other archaeal cellular appendages that have been described so far are assembled using T4P assembly mechanisms.[19] However, certain archaeal cell appendages, such as cannulae, hamii, and Mth60 fimbriae, are assembled using mechanisms that are not well understood and differ from the T4P assembly method. These mechanisms may utilize the Sec pathway for the translocation of subunits.[6]

The Tat-pathway is an alternative protein secretion pathway for bacterial and archaeal cells, which differs from the more common Sec-pathway. The main difference between these two pathways is that substrates for the Sec-pathway are transported into the periplasm in their unfolded state, whereas substrates for the Tat-pathway are already folded prior to crossing the cytoplasmic membrane.[20,21] The present study describes haloarchaeal filamentous structures with a diameter of approximately 3 nm. These filaments are composed of the main structural protein TafA, which contains a motif involved in the twin-arginine translocation pathway (Tat-pathway) and the Tat-translocation pathway was experimentally identified. This is the first report on the involvement of the Tat pathway in the secretion of protein subunits from filamentous structures, so we name these structures **ta**t-**fi**mbriae or **tafi** for short. Alternatively, the term "tafi" can be interpreted both as "**fi**mbriae that use the **Ta**t secretion pathway" or "**t**hin **a**dhesive/**a**rchaeal **fi**mbriae." Based on the molecular model of the tafi, we hypothesize that these structures are assembled using a "donor strand exchange" mechanism and have adhesive properties, similar to bacterial adhesive pili.

RESULTS AND DISCUSSION

Haloarcula hispanica DF60 synthesize thin filamentous surface structures

Through electron microscopy, we found that the auxotrophic *Har. hispanica* DF60 strain, which has a deletion of the *pyrF* gene responsible for pyrimidine synthesis, produced, along with typical archaella (10–11 nm in diameter), thinner filaments (approximately 3 nm in diameter) (Figure 1A). The synthesis of these thin filaments also occurs in the wild type strain, but it is much less noticeable than in DF60. In *Har. hispanica* DF60 cells thin filaments can be observed extending from the cell surface (Figure 1A).

On the SDS-PAGE electropherogram, the *Har. hispanica* archaellins appear as a set of ladder-like bands. The intensity of these bands decreases with decreasing mobility. Mass spectrometric analysis (data not shown) identified three archaellins ArlA1, ArlB, and ArlA2 in each of the sub-bands of the "ladder." This multiplicity is likely due to the presence of different archaellin glycoforms. In the samples from *Har. hispanica* DF60 cells taken after precipitation with polyethylene glycol (PEG), an additional major protein band with an apparent molecular weight of 37 kDa was observed on the SDS-gel electropherogram, in addition to the archaellin "ladder" (Figure 2B). The relative amount of this band increased with an increase in the number of thin

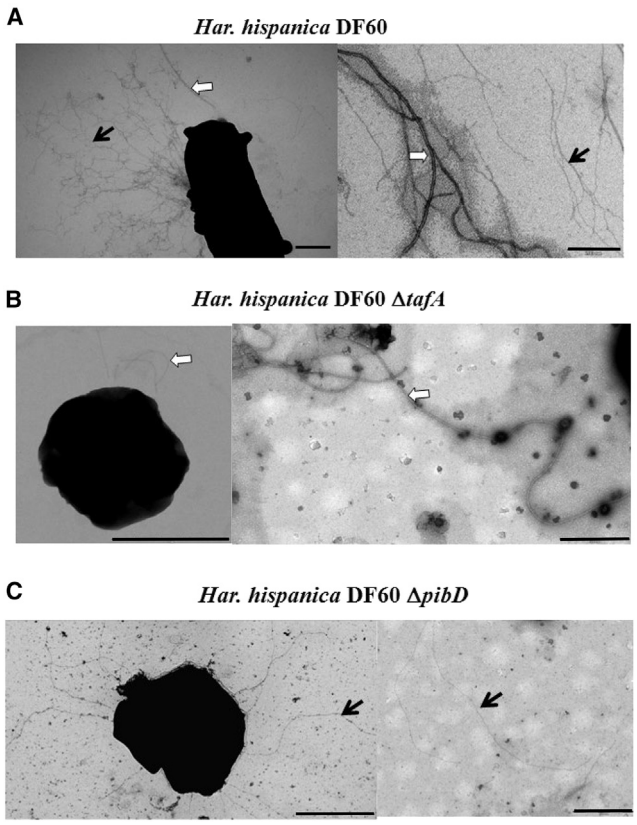


Figure 1. Synthesis of thin filamentous structures (tafi) by *Har. hispanica* cells

Transmission electron microscope micrographs of cells (left) and surface structures isolated by PEG precipitation (right) are shown. White arrow points to the archaella and black arrow points to the thin filaments.

(A) *Har. hispanica* DF60. Scale bars are 1 μm and 200 nm.

(B) *Har. hispanica* DF60 Δ*tafA* strain. Scale bars are 500 nm.

(C) *Har. hispanica* DF60 Δ*pibD* strain. There are only thin filamentous structures extending from the cell body. Scale bars are 1 μm.

filaments observed in the preparations by electron microscopy. The abundance of the 37 kDa protein was significantly higher in samples from the auxotrophic *Har. hispanica* DF60 strain compared to wild type strain (Figure 2B).

Mass spectrometry revealed that the additional 37 kDa protein band in the DF60 strain sample corresponds to the GenBank: WP_044951594.1 protein, which is annotated as the hypothetical *Har. hispanica* protein (the locus tag is HAH_0240 or HAH_RS01150). We will refer to this protein as TafA (Figure 2A and Figure S1).

Structural features of TafA (HAH_0240) protein

Gene *tafA* encodes a sequence of 208 or 210 amino acid residues, depending on which of the two neighboring methionine codons is used as the start. Its sequence shows no noticeable homology to known proteins, including those from other filamentous structures such as archaellins and pilins. Interestingly, by analyzing the TafA protein sequence using the TatFind,[20] SignalP 5.0[22] and TatP[23] prediction servers we found a pattern for the twin-arginine translocation (Tat) pathway (NRRSVL) (Figure 2A and Table S1). However, we did not find any Sec signal sequences.

The Tat-pathway is an alternative protein secretion pathway to the Sec-pathway, which is dominant for bacterial and archaeal protein export. The main difference between the two pathways is that the substrates for the Sec-pathway are transported into the periplasm in an unfolded state, while substrates for the Tat-pathway are already folded before being transported across the cytoplasmic membrane.[24] While the use of the Sec pathway for the assembly of archaeal[7,16,18,21] and bacterial[25] extracellular filamentous structures has been documented, the use of the Tat pathway for similar processes has not previously been reported. Therefore, we describe the first type of prokaryotic filamentous surface appendages, which are predicted to depend on the Tat system. These structures will be referred to as "**ta**t-**fi**mbriae" or "tafi" for short.

The Tat-system got its name because the sequences of their substrates contain two adjacent arginines in the signal peptide. Most haloarchaeal secreted proteins, whose homologs in non-haloarchaea were identified as potential Sec substrates, are predicted to be substrates of the Tat pathway.[22] This may be due to the high concentration of salts in the cytoplasm, which is necessary to balance the high concentration of sodium in the environment. Under high salinity conditions, unfolded proteins can easily aggregate, so the speed and quality of protein folding is vital for these microorganisms. Tat-translocation in haloarchaea is carried out by integral membrane proteins from the TatA and TatC families.[26] TatC recognizes a specific sequence at the N-terminus of the protein substrate and interacts with it, while TatA forms pores of the appropriate size through which the substrate is transported across the membrane.

The *Har. hispanica* genome contains genes that encode both main components of the Tat-secretion system: twin-arginine translocase proteins TatA (HAH_1119/GenBank: WP_01404 0012.1) and two preprotein translocase subunits TatC (HAH_0133/GenBank: WP_014039152.1 and HAH_0134/GenBank: WP_014039153.1). Using the TatFind server, we analyzed 3,860 protein sequences of *Har. hispanica* ATCC 33960 from the UniProt database and identified 122 potential Tat substrates, including TafA (UniProt: G0HRN5) and TafC (UniProt: G0HRN7).

The SignalP 5.0 server predicted the cleavage site for the signal sequence of the TafA protein to occur between Ala-35 and Thr-36 (Table S1). However, using mass spectrometry, we found that the N-terminal residue of mature TafA is the Phe-30 (Figure 2A and Figure S1).

It appears that not only the secretion mechanism but also the structure of the TafA differs significantly from known archaeal pilins and archaellins. As we know, the structures of type IV pili and archaella are supported by hydrophobic interactions between the N-terminal α-helices of archaellins and pilins. Based on the theoretical predictions of the secondary structure using the PsiPred program,[27] the mature TafA protein contains only unstructured regions and β-strands (Figure S2).

The SDS-PAGE analyses exhibit TafA bands located at a position of approximately 37 kDa (Figure 2B). However, the *tafA* gene encodes a protein with a predicted molecular mass of 21.6 kDa (18.8 kDa for the matured form). The discrepancy between the

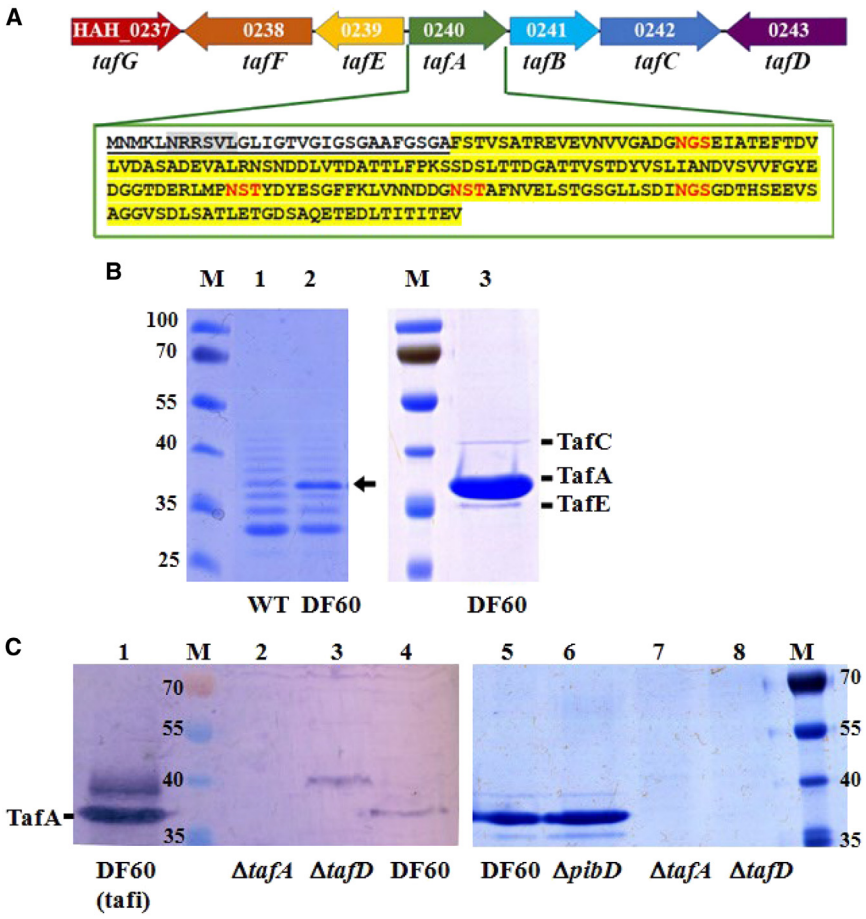


Figure 2. Identification of proteins involved in tafi assembly

(A) Genomic organization of the *Har. hispanica taf* gene cluster, *HAH_0237–0243*, located on chromosome 1 (NC_015948.1, 219076–225835) and amino acid sequence of the HAH_0240 (TafA) precursor protein. Codons encoding both the first and second methionines can be considered start codons. The Tat motif is shown in gray, and the mature TafA sequence starting with Phe30 is shown in yellow. Potential N-glycosylation sites are marked in red. See also Figures S1, S3, S6 and Table S1.

(B) SDS-PAGE analysis of surface structures from WT (lane 1) and DF60 (lane 2) *Har. hispanica* strains, isolated using PEG precipitation (without prior high-speed centrifugation). M - protein standards (kDa). In sample (1), the *Har. hispanica* archaellins look like a "ladder", with the intensity of the bands decreasing with a decrease in mobility. In sample (2), there is also an additional major band, marked with an arrow. Lane 3: Purified preparation of thin filaments (tat-fimbriae) from the DF60 strain, in which, in addition to HAH_0240 (TafA), also HAH_0239 (TafE) and HAH_0242 (TafC) proteins have been identified. See also Figure S7.

(C) Analysis of TafA synthesis in the parent and deletion strains of *Har. hispanica*. Lanes 1–4 show Western blot analysis of tafi sample obtained from the DF60 strain using PEG precipitation (1) and cell extracts from the deletion strains Δ*tafA* (2), Δ*tafD* (3), and the parent strain DF60 (4). In sample (3) there is no band at the TafA level, but an additional band corresponding to a protein with a larger mass is observed. This protein product is likely an unprocessed TafA, which accumulates in cells in the absence of the signal peptidase TafD. Lanes 5–8: SDS-electropherogram of samples obtained from the parent DF60 strain (5), as well as deletion strains, Δ*pibD* (6), Δ*tafA* (7), and Δ*tafD* (8) by PEG precipitation after the removal of archaella by high-speed centrifugation. The band corresponding to the protein TafA is observed only in preparations from the Δ*pibD* and parent DF60 strains. See also Figure 1B.

molecular masses estimated by SDS-PAGE and those predicted from the gene sequence is common among haloarchaeal proteins due to the high content of acidic amino acid residues[28,29] and glycosylation.[30] In fact, the TafA sequence is rich in acidic residues (Asp + Glu = 19.89%) and its calculated isoelectric point is 3.55.[31] In addition, there are four potential N-glycosylation sites in the TafA sequence (Asn-X-Thr/Ser) (Figure 2A) and Schiff staining of the protein was positive (see Figure S3).

We found that tafi could completely dissociate upon heating to 90°C when the NaCl concentration in the solution is below 5%. After this treatment tafi are likely to dissociate into separate subunits. Circular dichroism measurements have confirmed that the TafA protein is in an unfolded state at both 0% and 5% NaCl concentrations. As the salt concentration increases further, we observed spectra characteristic of proteins with a predominance of β-structure (Figure S4). These results are consistent with the theoretical predictions for secondary structure the TafA (Figure S2).

Microcalorimetric studies have confirmed that at the salt concentration same at the cultivation for *Har. hispanica* (20% NaCl) TafA has a domain structure. A single heat absorption peak was observed at a melting point of approximately 72°C and the melting was reversible. However, when TafA is dissociated, it is unable to spontaneously polymerize into fimbriae that resemble natural ones. Experiments conducted *in vitro* to polymerize TafA under various conditions yielded negative results (data not shown).

Bioinformatical analysis of the *tafA* gene and surrounding open reading frames

Currently, the product of the *tafA* gene has been annotated as a hypothetical protein. This gene is located between open reading frames (ORFs) that also encode "hypothetical proteins", and does not appear to be part of an operon. Using the protein-protein BLAST algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi), we found that the TafA protein is specific to halophilic archaea and its homologs are not present in other organisms. Homologs of this protein have been identified in many representatives of the genera *Haloarcula*, *Haloferax,* and *Halorubrum*, as well as in some other genera such as *Halorhabdus*, *Haloterrigena*, *Natrialba*, *Natrinema*, *Natronorubrum,* and *Salinadaptatus*. The species in which the TafA homologs were found are listed in Table S2.

The closest homologs (100-67% identitical) of the TafA protein were found in the genomes of approximately 20 known strains, including *Haloarcula amylolytica* JCM 13557, several strains of *Har. hispanica*, *Har. rubripromontorii* pws8, *Har. rubripromontorii* SL3, *Haloarcula* sp. Atlit-7R, *Haloarcula* sp. Atlit-47R, *Haloarcula* sp. Atlit-120R, *Haloarcula* sp. CBA1115, *Haloarcula* sp. CBA1127, *Haloarcula* sp. CBA1128, *Haloarcula* sp. CBA1131, *Haloarcula* sp. K1, *Har. taiwanensis*, *Haloferax elongans* ATCC BAA-1513, *Hfx. larsenii* CDM_5, *Hfx. larsenii* JCM 13917 and *Hfx. mediterranei* ATCC 33500. The genomes of the second group of haloarchaea (*Haloarcula argentinensis, Har. californiae, Har. marismortui, Har. sinaiiensis, Har. vallismortis, Halorubrum coriense, Hrr. ezzemoulense, Hrr. hochstenium, Hrr. litoreum*, and so forth) also contain genes encoding proteins similar to TafA with significantly less homology (about 40% identity) (Table S2). In addition, the genomes of some haloarchaea contain proteins that have approximately 40 amino acid residues at their N-termini that are homologous to the TafA protein. The central and C-terminal regions of these proteins do not share any homology with TafA.

The Weblogo representation of N-terminal motifs based on a comparison of 97 N-terminal sequences of TafA homologous proteins shows the stability of the RR motif and the high overall conservation of these sequences (Figure S5). The observed pattern is consistent with the Tat signal peptide regions from different classes of substrates[23] and the corresponding Tat motifs found in haloarchaea.[26]

The *Har. hispanica tafA* gene is located adjacent to six other genes (*HAH_0241*, *HAH_0242*, *HAH_0243*, *HAH_0239*, *HAH_0238,* and *HAH_0237*), which we have named *tafB-tafG*. These genes are illustrated in Figure 2A and Figure S6. These genes in the genome appear to co-occur with *tafA*. We assume that these genes, along with *tafA*, play a role in the formation of tat-fimbriae. Most of the proteins encoded by these genes (except for TafD, signal peptidase I) are annotated as hypothetical. According to our analysis using TatFind, SignalP 5.0 and TatP prediction servers, the proteins HAH_0240 (TafA), HAH_0239 (TafE) and HAH_0242 (TafC) contain twin-arginine motifs (see Table S1).

On the SDS-PAGE electropherogram of comparatively pure tafi preparations, two minor bands are visible above and below the main band of the TafA (Figure 2B). Mass spectrometric analysis revealed that the upper band corresponded to TafC and the lower band to TafE (Figure 2B and Figure S7). By scanning the electrophoresis gel and measuring the intensities of the bands corresponding to TafA, TafC, and TafE, which have estimated molecular masses of 18.8 kDa, 37.8 kDa, and 17.1 kDa respectively, we were able to estimate the approximate ratio of these proteins in the fimbria preparation. For every 100 TafA molecules, there are approximately one TafC molecule and 3–5 TafE molecules.

TafA and TafE have been annotated as hypothetical proteins. They are similar in size and their predicted secondary structures also show similarities (Figure S2).

It should be noted that the amino acid sequence of TafE (HAH_0239) as reported in the UniProt database (UniProt: G0HRN4) and NCBI database (GenBank: WP_014039244.1) may be inaccurate, as it appears to lack 14 N-terminal amino acid residues. This is our opinion based on the available information. In this work, we have used the HAH_0239 sequence from the MIST database, available on the page https://mistdb.com/mist/genes/GCF_000223905.1-HAH_RS01145. This database provides information about the source organism, *Haloarcula hispanica* ATCC 33960, and the protein accession number GenBank: WP_044951591.1. The link to the protein sequence can be found at https://www.ncbi.nlm.nih.gov/protein/WP_044951591.1. However, it reports that the sequence has been suppressed because it is no longer annotated in any genome. The active "Identical Proteins" link on this page leads to two identical protein sequences: GenBank: KAA9405492.1 hypothetical protein *Haloarcula* sp. CBA1131 NCBI: GCA_008728975.1 (189 aa) and GenBank: KAA9408628.1 hypothetical protein *Haloarcula hispanica* NCBI: CBA1121 GCA_008729095.1 (189 aa). Thus, the protein TafE (HAH_0239) from *Har. hispanica* ATCC 33960, which is linked to the MIST database, is encoded by the hidden entry GenBank: WP_044951591.1 and is identical to proteins GenBank: KAA9405492.1 and GenBank: KAA9408628.1. These latter proteins can be used as a "benchmark" to restore the missing N-terminal region of TafE from *Har. hispanica* ATCC 33960. A version of this protein that is truncated from the N-terminus is available in the UniProt and NCBI databases as UniProt: G0HRN4 and GenBank: WP_014039244.1, respectively. This truncated version of the protein consists of 175 amino acids and may have resulted from errors in gene sequencing or annotation. By examining the DNA sequences surrounding the gene that encodes for the protein GenBank: WP_014039244.1 in the original *Har. hispanica* ATCC 33960 genome, we can easily identify the section that codes for the 14 missing amino acids at the N-terminus in the same reading frame as the truncated version. Additionally, in the extended protein, AUG functions as a start codon, rather than the alternative GUG seen in the shortened version (Figure S8). Note that the N-terminal region of GenBank: WP_014039244.1 does not contain the double arginine motif, which is a feature of the leader peptide of the TafE protein. This differs from GenBank: KAA9408628.1 (GenBank: WP_044951591.1). In our opinion, it would be appropriate to accept the entry GenBank: KAA9408628.1 from the NCBI database for *Har. hispanica* CBA1121 as the full version of HAH_0239 and make appropriate changes in the UniProt:G0HRN4 and GenBank: WP_014039244.1 entries.

The N-terminal region of TafC (amino acid residues 1–146) structurally resembles that of TafA and TafE (Figure S2). However, its C-terminal portion includes a laminin G-like domain (residues 203–360), which belongs to the concanavalin A-like lectin/glucanase superfamily. It is known that laminins are a large family of adhesive glycoproteins, found in the basal membrane and perform various functions.[32] Proteins from this glucanases superfamily are also known to bind sugars and glycoproteins. The presence of this protein in the tafi may indicate its adhesive potential and possible involvement in the biofilm formation. The minor protein TafE is possibly used for the initiation/termination of tafi assembly or functions as an adapter between the main structural protein TafA and the adhesin TafC. The Weblogo representation of the N-terminal motifs of TafE and TafC homologs, such as TafA, demonstrates the stability of the two arginines, the high conservation of adjacent residues, as well as agreement with the consensus for known Tat substrates[23,26] (Figure S5).

In addition to the three protein subunits of tafi, we also analyzed the proteins encoded in the gene cluster. The TafB protein (HAH_0241, GenBank: WP_023843080.1) is annotated as hypothetical in the NCBI database, but it is annotated as a transcriptional regulator in the UniProt database (UniProt: G0HRN6). One of its close homologs (an identity is 64% and a positivity is 76%), GenBank: WP_094495324.1 from *Halorubrum ezzemoulense* Ec15, is also annotated as a Lrp/AsnC family transcriptional regulator. The Lrp/AsnC family of proteins that regulate transcription has been found in both archaea and bacteria. Leucine-responsive regulatory proteins (Lrp) are involved in the transport, degradation, and biosynthesis of amino acids. They have also been linked to the production of pili, porins, sugar transporters, and nucleotide transhydrogenases.[33] TafB, which is predicted to be transmembrane, has its functional domain located in the cytoplasm (Figure S9).

TafD (HAH_0243) belongs to the S24 and S26 superfamily of LexA/signal peptidases, which includes membrane-bound proteases that cleave the N-terminal signal peptide of membrane-secreting proteins.[34] Interestingly, the TafD peptidase has four predicted transmembrane regions and two domains located outside the cytoplasmic membrane (Figure S9). The N-terminal domain contains conservative residues (Ser-40, His-83, and Arg-84) that are responsible for peptidase activity. The function of the C-terminal domain is unclear.

TafG (HAH_0237) contains a DUF5305 domain (amino acid residues 132–349). The DUF5305 family includes several hypothetical proteins with unknown function and is mainly represented in archaea and some bacteria (https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=cdd&uid=407348). Interestingly, some bacteria, such as *Actinoplanes lutulentus*, *Haloplanus salinus,* and *Tepidiforma bonchosmolovskaya*) have signal peptidases (GenBank: WP_221402508.1, WP_147270848.1 and WP_158065712.1, respectively), which, along with the domain with peptidase activity, contain a DUF5305 domain. In some haloarchaea such as *Natrinema* sp. SLN56, there are duplicated pairs of *tafD* and *tafG* genes (GenBank: WP_226004277.1/WP_226004276.1, WP_226004489.1/WP_226004488.1 and WP_226005358.1/WP_226005357.1) (Table S2), which may suggest that TafG and the signal peptidase, TafD, function in a complementary manner.

TafF (HAH_0238) protein is annotated as hypothetical, and its close homologs have been found only in archaea with the *tafA* gene. One of such homologs, GenBank: WP_233127167.1 from *Halorubrum* sp. SD612 (Table S2), is annotated as ''MSCRAMM family adhesin SdrC.'' Its gene is located within a similar *taf* cluster adjacent to genes *tafE* (GenBank: WP_086219893.1), *tafA* (GenBank: WP_086219895.1) and *tafB* (GenBank: WP_086219894.1). The MSCRAMM family (microbial surface components recognizing adhesive matrix molecules) is a group of proteins that have two adjacent IgG-like subdomains. They are involved in the initial attachment of bacteria to host tissue during infection. For example, protein A that binds to IgG and factors from *Staphylococcus* and *Streptococcus* that bind to fibrinogen, and so forth.[35] Interestingly, for TafF, the SignalP-5.0 server predicts a high probability of the characteristic signal peptide for the Sec-pathway (Table S1), rather than for the Tat-pathway as it does for TafA, TafC, and TafE. However, for proteins TafB, TafD, and TafG, SignalP-5.0 does not detect any Tat or Sec signal peptides (Table S1).

Analysis of the protein sequences encoded by the seven *Har. hispanica taf* genes for the presence of transmembrane regions using the TMHMM-2.0 server[36,37] (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0), revealed that TafB (two transmembrane helices, THs), TafD (four THs), TafF (three THs) and TafG (three THs) are probably integrated into the cell membrane (Figure S9). However, only for TafB, the majority of its polypeptide chain appears to be localized inside the cell. In contrast, for proteins TafD, TafF, and TafG functional domains are likely to be located on the outside surface of the cell membrane.

All of these seven proteins, as we hypothesize, are related through a common function and play a role in the assembly and secretion of tat-fimbriae.

Deletion of *tafA* and *tafD* genes abolishes tafi biosynthesis

To confirm that the TafA protein is the main structural component of tafi, we generated a strain with a deletion of the *tafA* gene (Δ*tafA*), as described in our experimental procedures. As the signal peptidase gene, *tafD*, is located in close proximity to *tafA*, we hypothesized that it may play a role in the processing of TafA. To test this hypothesis, we created a strain lacking *tafD* (Δ*tafD*) and analyzed the resulting protein products.

We were unable to detect the TafA protein band on the SDS-PAGE gel in the samples prepared from Δ*tafA* and Δ*tafD* cells using PEG precipitation (Figure 2C). Additionally, the corresponding filamentous structures were not observed under electron microscopy (Figure 1B). This loss of TafA synthesis in Δ*tafA* cells supports our hypothesis that TafA is the main structural component of the tat-fimbriae.

To detect TafA in cell lysates, we used specific polyclonal antibodies against both purified tafi and a non-glycosylated His-tagged TafA protein expressed in *E. coli*. However, the polyclonal antibody raised against purified tafi had low specificity and was not suitable for our purposes. This is likely due to its cross-reaction with glycoproteins from *Har. hispanica*, whose oligosaccharides have a similar structure to those of glycosylated TafA. In contrast, the polyclonal antibodies generated against the non-glycosylated recombinant TafA demonstrated satisfactory specificity. Western blotting using these antibodies revealed the presence of TafA in the parental DF60 cells, but not in the Δ*tafA* mutants. In the Δ*tafD* cell lysate, a band corresponding to the TafA level was not detected. Instead, a band for a larger protein was observed (Figure 2C). This larger protein is likely the unprocessed form of TafA, which, in the absence of the signal peptidase TafD, cannot be secreted and accumulates within the cells. We hypothesize that the deletion of the *tafD* gene prevents the processing of TafA into its functional form, thereby preventing its participation in the tafi assembly.

Deletion of *pibD* gene does not affect tafi biosynthesis

For a more detailed study of tafi, it is useful to obtain a strain lacking archaella and pili. To achieve this, we have constructed a *Har. hispanica* strain with the deletion of *pibD* (*HAH_2955*) gene, which encodes a common prepilin peptidase III involved

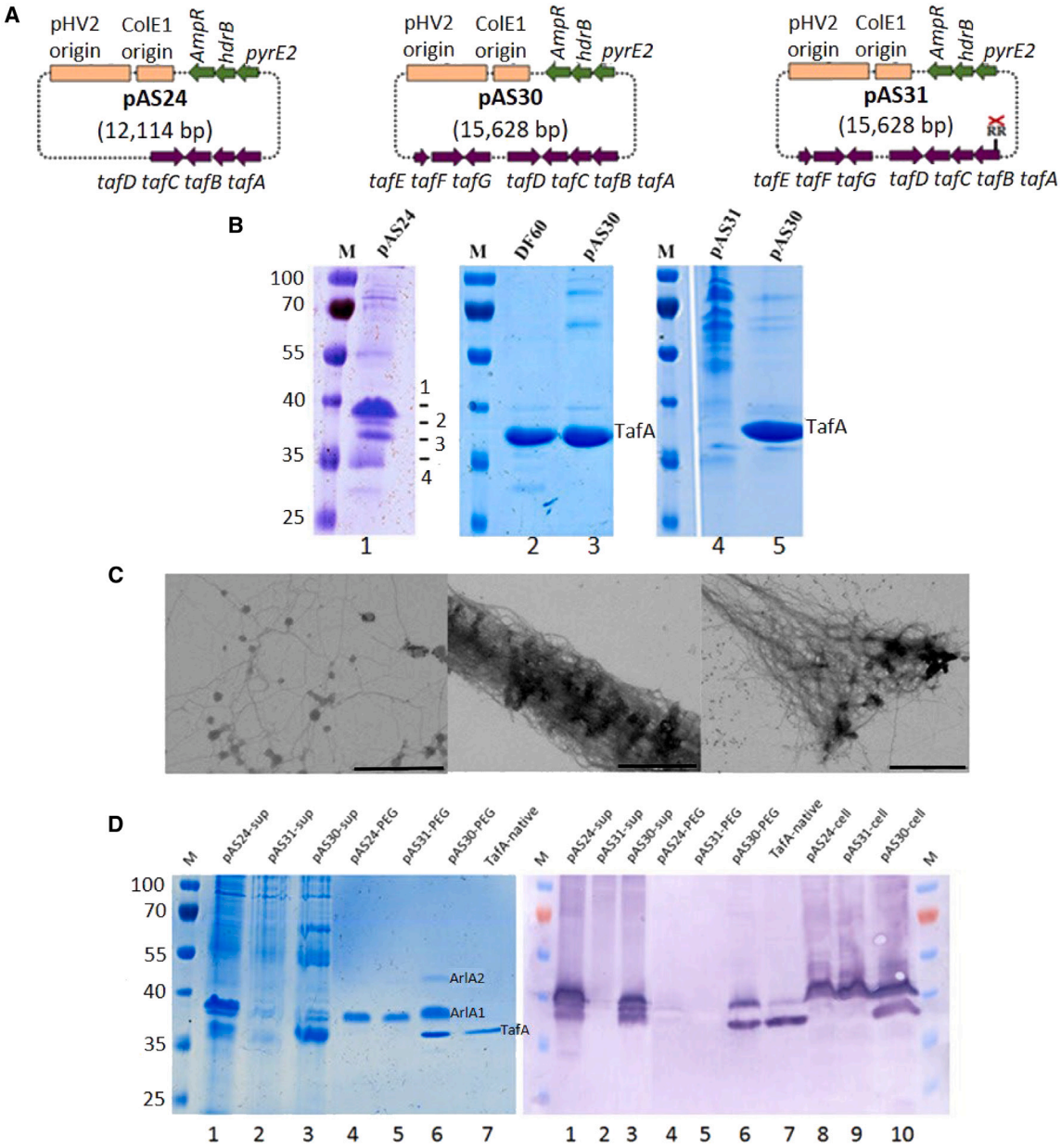


Figure 3. Assembly of tafi during the heterologous expression of *taf* genes in *Hfx. volcanii*

(A) Plasmid constructions: pAS24, pAS30, and pAS31, with *Har. hispanica taf* genes for heterologous expression in *Hfx. volcanii* MT78 and RE26 strains.

(B) SDS-electropherograms of preparations obtained through heterologous expression. Lane 1- concentrated extracellular supernatant from *Hfx. volcanii* MT78 with pAS24 plasmid containing an incomplete set of *taf* genes (*tafA* - *tafD*). Protein bands (1–4) were analyzed by mass spectrometry. Lane 2 – tafi sample obtained from *Har. hispanica* DF60. Lane 3 - sample of filamentous structures obtained using the heterologous expression of the pAS30 plasmid with the complete set of *taf* genes, compared to a sample of naturally occurring *Har. hispanica* tat-fimbriae (lane 2). Lanes 4 and 5 - samples obtained by precipitation with polyethylene glycol under identical conditions, using the pAS30 (5) and pAS31 (4), with the same *taf* genes, but with the region RR in the TafA signal peptide replaced by KK. See also Figures S12-S15.

(C) Electron microscope images of recombinant tafi structures obtained through the heterologous expression of the pAS30 plasmid in *Hfx. volcanii* MT78. The scale bar is 500 nm. Both individual filaments (left) and thicker filament bundles (middle andright) were observed.

(D) Analysis of the presence of TafA in *Hfx. volcanii* RE26 preparations upon the expression of plasmids pAS24, pAS30, and pAS31 using specific antibody. Left: SDS-PAGE stained with Coomassie Brilliant Blue G-260. Right: Immunoblot after SDS-PAGE stained using specific antibody. Preparations isolated from *Hfx. volcanii* RE26 strains expressing plasmids: pAS24, pAS31 and pAS30, respectively: (1–3) Supernatants obtained by high-speed centrifugation after the precipitation of polymeric structures; (4–6) Precipitates obtained by high-speed centrifugation; (8–10) Cell extracts. (7) Natural tat-fimbriae. The bands of the TafA protein and *Hfx. volcanii* archaellins ArlA1 and ArlA2 are marked on the electropherogram. It can be seen that after the replacement of arginine residues in the signal peptide of the TafA with lysine residues, this protein is no longer detectable in the extracellular space. This is true whether it is in the supernatant or in

(legend continued on next page)

in the N-terminal processing of both archaellins and pilins. It should be noted that this peptidase does not have any noticeable similarity to the TafD peptidase, either in amino acid sequence or predicted spatial structure. Despite the presence of at least five hypothetic type IV pilins genes in the *Har. hispanica* genome (*HAH_0921*, *HAH_1312*, *HAH_2274*, *HAH_2659,* and *HAH_2796*), we have not observed any filamentous structures that could be related to type IV pili in our experiments. As shown by our data, Δ*pibD* cells only contain long, thin fimbriae, and no archaella or pili. In samples from these cells, we detected TafA protein, but there were no archaellin bands visible (Figure 2C). Electron microscopy data confirmed that tafi are located on the cell surface and appear to be anchored there (Figure 1C).

Complementation analysis

No tafi was produced in the Δ*tafA* and Δ*tafD* mutant strains. When the *tafA* and *tafD* genes were reintroduced into the Δ*tafA* and Δ*tafD* mutant strains, respectively, as described in the experimental procedures (Figure S10), we verified that the synthesis of TafA and tat-fimbriae had been restored only in the strain with the *tafD* gene complementation (Figure S11A, C). Western blotting using TafA-specific antibodies revealed the presence of unprocessed TafA in cell extracts of both complemented strains. However, only in the strain with a *tafD* gene complementation, we observed not only the unprocessed form of the TafA but also its processed form (Figure S11B), confirming the role of the TafD signal peptidase in TafA processing. It remains unclear why, when the *tafA* gene is expressed from a plasmid rather than a chromosome, its protein product is not processed. Perhaps that, with the deletion of the *tafA* gene, the synchronous operation of the *taf* genes (particularly the *tafB* gene, which is adjacent to *tafA* and encodes a putative transcription regulator), is disrupted. It can be assumed that the potential goal of TafB is to activate the transcription of *tafD*. If this is the case, the processing of the TafA protein would be disrupted, and tat-fimbriae would not form. This issue can be further clarified through experiments on knocking out the *tafB* gene, which we plan to conduct in the near future. A non-motile Δ*pibD* strain was complemented with the expression of the *pibD* gene, resulting in the restoration of motility on semi-solid agar (Figure S11D).

Heterologous expression of *Har. hispanica tafA-tafG* genes in *Haloferax volcanii*

Previously, it has been shown that functional archaellar filaments can assemble upon the expression of the *Halorubrum lacusprofundi* or *Halorubrum saccharovorum* archaellin genes in a heterologous system using a *Hfx. volcanii* strain that is non-motile[38] and restores its motility.[39] For similar experiments, a special strain of *Hfx. volcanii* MT78 is used in which the archaellin operon (Δ*arlA1*Δ*arlA2*) and the genes necessary for pili formation (Δ*pilB3-C3*) have been deleted.[40] This strain is non-motile and does not posses any structures related to type IV pili. We used this strain to test the possibility of tafi synthesis in a heterologous system. Unlike its closest relatives, *Hfx. elongans*, *Hfx. larsenii* and *Hfx. mediterranei*, the *Hfx. volcanii* genome does not contain genes corresponding to *Har. hispanica tafA-tafG*.

For this purpose, a vector pAS24 containing a set of four genes from *tafA* to *tafD*, was constructed based on the pTA1228 vector.[41] This set includes the genes for the main tafi protein TafA and TafD, a putative signal peptidase, which processes the TafA precursor. In this vector, the *tafA* gene is positioned under a tryptophan-inducible promoter, while *tafB*, *tafC,* and *tafD* genes are under their natural promoters (Figure 3A). Since the *tafD* gene is the last in the tandem array and transcribed in the opposite direction, a non-coding sequence of 175 bp (NC_015948.1: complement: 225836–226010), between neighboring *HAH_0243* and *HAH_0244* genes, was cloned with the *tafD* gene. This sequence contains the putative promoter region of the *tafD* gene and precedes its start codon. In the case of native host *Har. hispanica*, we found that both TafA and TafD were required for tafi synthesis. However, when we expressed the *taf* genes from the pAS24 plasmid in *Hfx. volcanii*, we did not observe any tafi-like structures. We hypothesized that the translocation of TafA might be impaired and that it either accumulates inside the cell or is secreted but cannot form fimbriae.

Experimental data confirmed that TafA is secreted but not polymerized. The supernatant obtained after separating of the cell biomass after several rounds of high-speed centrifugation was concentrated by 100-fold using Amicon Ultra Filters. The collected product was analyzed using SDS-PAGE. As a control, we used the same preparation from *Hfx. volcanii* MT78 without the expression vector. As seen in Figure 3B, the concentrated supernatant from *Hfx. volcanii* MT78/pAS24 contained four additional protein bands. Mass spectrometry revealed that all four bands contained protein products from the *tafA* and *tafC* genes (Figures S12-S15). TafA was identified with the highest confidence (87% coverage) in band 3, and TafC (55% coverage) was found in band 1. Surprisingly, in all of the bands, we detected the *Hfx. volcanii* HVO_A0133 protein (GenBank: WP_013035207.1 with the highest confidence (83% coverage) in band 4 (Figure S15). This protein is a member of the glycine zipper family, and it contains a Tat-pathway signal sequence. It is encoded by the *Hfx. volcanii* pHV4 plasmid. This protein is not ubiquitous and has been identified in only a few haloarchaea species, including *Halorubrum* sp. Atlit-28R, *Halomicrobium katesii*, *Halorubrum lipolyticum, Halobacterium* sp. *NMX12-1*, *Haloferax larsenii*, and *Halobiforma haloterrestris*. The function of HVO_A0133 and the effect of pAS24 plasmid expression on its production remain unclear.

Interestingly, TafAs from different bands appear to be processed differently. In natural tafi, the mature form of TafA starts with Phe-30 (see Figure S1). However, in band 2 (see Figure S13), we detected six additional amino acid residues before Phe-30. In

polymeric structures. A protein detected in the cell extract using immunoblotting has a lower mobility compared to the naturally occurring TafA (lane 9) and probably corresponds to an unprocessed form of the protein that is unable to be secreted from the cells. Additionally, it is evident that upon the expression of plasmids pAS24 and pAS30, TafA can be produced in several forms. Upon the expression of pAS30, a significant proportion of TafA (likely due to excess production) appears not to be used in the formation of the tat-fimbriae (lane 3). See also Figure S16.

band 3 Phe-30 was preceded by one amino acid residue Ala (Figure S14). Finally, in band 4, the TafA started with Phe-30 again (see Figure S15). This may indicate that TafA is processed differently during the expression of a subset of *taf* genes. The heterogeneity observed in proteins TafC and HVO_A0133 may also be associated with variability in their post-translational processing.

These results demonstrate that TafA and TafC can be synthesized and secreted in the heterologous *Hfx. volcanii* system. However, the recombinant TafA protein is unable to polymerize and form tafi. We suspected that this inability of the TafA protein to polymerize may be due to the absence of additional accessory proteins encoded by *tafE*, *tafF,* and *tafG*. Therefore, our next step was to obtain an extended genetic construct, pAS30, which contains the full set of *taf* genes from *tafA* to *tafG* (Figure 3A). We have shown that the expression of all seven *taf* genes on the pAS30 plasmid in the *Hfx. volcanii* results in the production of recombinant tat-fimbriae that are similar to natural ones from *Har. hispanica* (Figure 3C). The recombinant tafi tend to form thick bundles (Figure 3C). The electrophoretic mobility of recombinant TafA does not differ from that of the native protein (Figure 3B). These findings confirm that the polymerization of the major tafi subunit, TafA, requires additional proteins encoded by the *taf* gene cluster. Further experiments will be needed to elucidate the roles of each of these genes in tafi assembly and function.

To confirm the key role of the Tat-pathway in TafA secretion and its subsequent assembly into tafi, we used nucleotide substitution mutagenesis to modify the pAS30 plasmid. This plasmid encodes a pair of arginine residues, which are a key component of the Tat pathway for TafA. We replaced these residues with a pair of lysine residues to block export via the Tat secretion pathway. The modified pAS31 plasmid (Figure 3A) was used to transform cells of the *Hfx. volcanii* MT78 strain. We used a technique that was previously effective for isolating tafi from the MT78/pAS30 strain to obtain similar preparations from the MT78/pAS31 strain. However, preparations from MT78/pAS31 did not contain any filamentous structures composed of TafA, compared to preparations from MT78/pAS30. This confirms that arginine residues (RR) at a specific position are critical for the assembly of tafi (Figure 3B).

The specific antibodies against His-tagged TafA were used to obtain additional information about its expression in a heterologous *Hfx. volcanii* system. The studies were conducted using the motile, archaellated strain *Hfx. volcanii* RE26,[42] which allows for the rapid selection of transformants through culturing on a semi-solid medium. The presence of archaella in the samples does not interfere with immunological assays, as the archaellin proteins do not interact with the antibodies being tested. Comparison of cell lysates from three strains, containing plasmids pAS24, pAS31, and pAS30, respectively, revealed the presence of a major protein band in each case. This protein was stained with specific antibodies and was localized upstream of the processed TafA, as shown in Figure 3D. This band was absent in lysates from the parental strain RE26 (Figure S16) and appears to correspond to the unprocessed form of TafA. An additional major antibody-reactive band at the same level as the mature TafA was observed in the lysate from the RE26/pAS30 strain only (Figure 3D and S16). These findings are consistent with the data on tafi synthesis during the expression of the pAS30 plasmid, which contains a complete set of *taf* genes. The replacement of two arginine residues in the Tat motif with lysines in the TafA sequence (RE26/pAS31) likely prevents its processing and inhibits the production of tafi. In the case of a partial set of *taf* genes (RE26/pAS24), as shown here, the TafA protein is produced and secreted, but it fails to form tat-fimbriae. As a result, unprocessed TafA can be observed in the cell extracts, while the mature protein is released into the extracellular space.

In addition to the cell lysates, we also tested the concentrated extracellular supernatants from these three strains for the presence of TafA. Interestingly, for strains expressing both the complete (RE26/pAS30) and incomplete (RE26/pAS24) sets of *taf* genes, we detected at least two additional antibody-reactive protein bands in the supernatants that migrated more slowly than the mature TafA (Figure 3D). This could be due to various post-translational modifications, such as glycosylation or heterogeneous processing, as suggested by the mass spectrometry data mentioned above (Figures S12-S15). Even when the full set of *taf* genes is expressed, a significant proportion of the synthesized TafA is secreted from the cell without participating in tafi assembly (Figure 3D). It should be noted that in the plasmids pAS24, pAS31, and pAS30, which are designed for heterologous expression, the *tafA*, unlike other *taf* genes is expressed under a tryptophan-inducible promoter, rather than its native promoter. This may lead to an imbalance in the amount of TafA compared to other proteins involved in tafi assembly, especially TafD, a peptidase. TafA alone may be unable to be processed by TafD. It is also possible that excess TafA could be cleaved by *Hfx. volcanii*'s own peptidases, which process other protein substrates in the Tat pathway. The heterogeneity of TafA may be linked to differences in its interaction with different types of signal peptidases.

When pAS31 is harbored, TafA is not detected either in the polymeric form or in the extracellular environment. However, it is present in the cell lysate in an unprocessed form (Figure 3D). Replacing the two critical arginine residues for the Tat pathway with lysine residues should not prevent secretion through the Sec pathway, as predicted by SignalP 5.0. Therefore, based on our data, we conclude that TafA is a substrate for the Tat pathway.

Molecular modeling of tafi structure (AlphaFold)

For all unprocessed single-stranded proteins considered in this work, the AlphaFold database https://alphafold.ebi.ac.uk contains the results of modeling their 3D structures. The AlphaFold Colab[43–45] and Dali[46] servers were used to predict and compare the spatial structures of mature TafA, TafC, TafE, and TafF proteins and their complexes. These proteins were transformed into mature ones by removing signal peptides from their sequences. It was found that the TafA, TafE, and N-terminal TafC domains have β-sandwich fold, which is similar to that of the main and auxiliary protein subunits of the bacterial adhesive pili type 1 (P1T) (Figure 4A). The FimA main pilus type 1 subunit from an enteroinvasive bacterium (PDB: 5NKT) is shown as an example in Figure 4A.

According to analysis using the Dali server, the TafE protein shows a higher structural similarity to the FimA from the PDB (PDB: 5NKT) (Figure S17A, Table S3) as well as FimA proteins

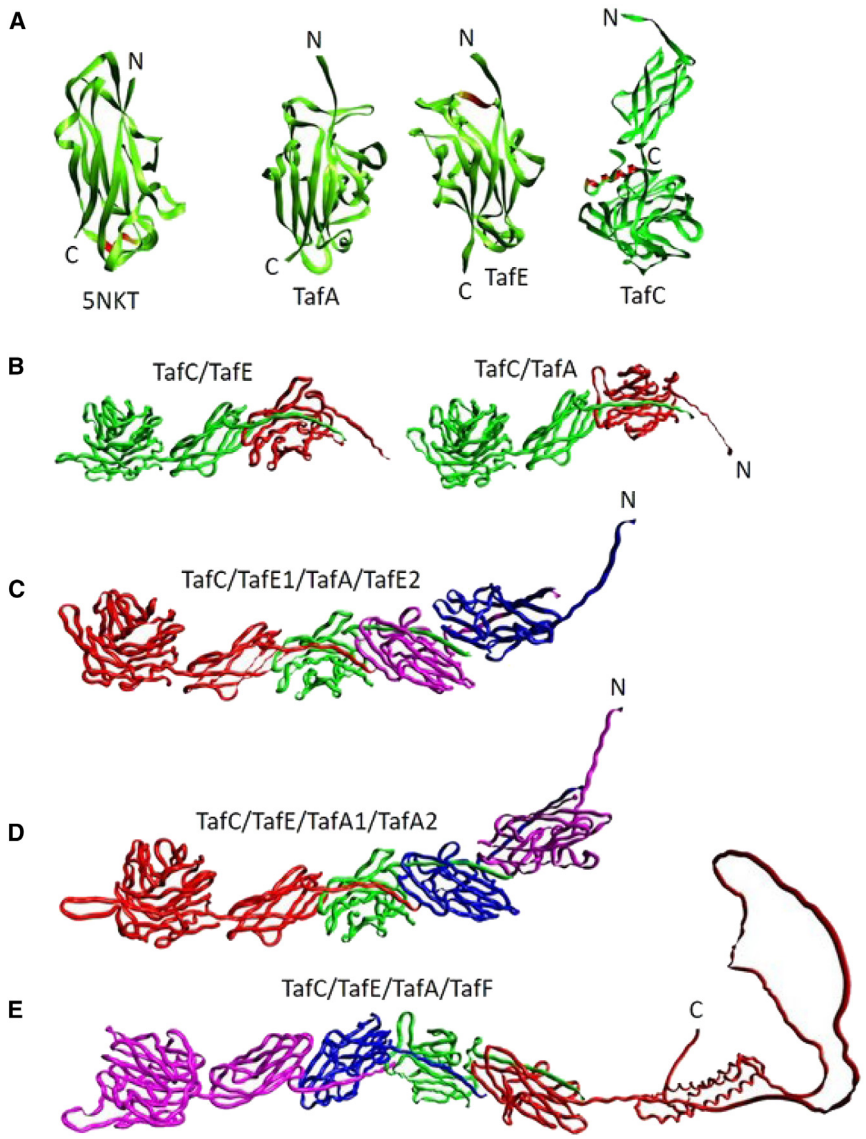


Figure 4. AlphaFold 2 molecular modeling of 3D structures for *Har. hispanica* tafi proteins

(A) Predictions for matured TafA, TafE, and TafC proteins, and experimental (X-ray) structure of FimA WT from *E. coli* (PDB: 5NKT). See also Figure S17A and Table S3.

(B) Predictions for dimeric complexes of matured TafA, TafE, and TafC proteins. It can be seen that the interaction between TafC and TafE (or TafA) leads to their activation and switching to the "open form" when the N-terminal donor strand is turned away from the globular region. See also Figure S18.

(C and D) Predicted structures of tetrameric complexes composed of TafC and either one or two copies of protein TafA/TafE. In both cases, it can be seen that TafC is the first subunit, followed by TafE. Then, TafE and TafA can possibly alternate.

(E) A prediction for a heterotetrameric complex composed of matured TafC, TafE, TafA, and TafF proteins.

from other PDB : 2M5G, 6R74, and 6R7E (Z-scores: 6.6–6.8). TafA has a higher level of structural similarity to major pilin protein from *Streptococcus pyogenes*, including both the major subunit (PDB: 3GLD and 3GLE) and minor subunit FimG (PDB: 3JWN) (Z-scores of 5.3 and 4.6, respectively). It also has a high level of similarity to CfaB, the major pilin subunit from enterotoxigenic *Escherichia coli* (PDB: 3F83) (*Z* score of 3.5). At the same time, TafA's structure obtained using AlphaFold fits well with TafE's structure (*Z* score: 8.3, %identity: 18). Pairwise comparisons of the TafA/TafE core structures with other Taf family proteins (see Table S3 for details) reveal that TafF, TafD, and the N-terminal domain of TafC have structurally similar regions to TafA. For TafA/TafF and TafA/TafD pairs, the Z-scores were 7.0 and 5.7, respectively. For the TafA-N-terminal domain of TafC pair, the *Z* score was 4.3. However, the structural similarity between TafB, TafD, and TafG should only be considered in terms of their transmembrane regions.

The bacterial adhesive P1T (or P-pili) is assembled through the chaperone-usher pathway of pilus biogenesis.[47–50] A more detailed investigation of the P1T composition has demonstrated that these structures are multicomponent and composed of several proteins. The individual pilus subunits, such as FimA and PapA associate through "donor strand complementation" when the incomplete immunoglobulin-like fold of each subunit is completed by the N-terminal extension of a neighboring subunit. The tip protein, FimH or PapG, is located at the distal end of the pilus and has two essential domains. The N-terminal domain is a lectin that is responsible for the adhesive properties of the protein. The C-terminal domain has a structure similar to the main core protein of the pilus, FimA or PapA. Furthermore, there may be one or more copies of the adapter subunits, such as FimG, FimF/PapF, PapE, and PapK, which connect FimH/PapG to FimA/PapA. Approximately 1000 copies of FimA/PapA make up the rod of the pilus. It is important to note, that all of these proteins pass through the inner membrane (IM) using the Sec-dependent pathway. Once inside the periplasm, the subunits form binary complexes with the FimC/PapD protein chaperone. To assemble into a pilus fiber and be secreted to the cell surface, these chaperone-subunit complexes need to interact with the protein usher on the outer membrane (OM). The usher catalyzes the switching of a chaperone-subunit interaction for a subunit-subunit interaction. Interactions between the subunits occur through a process called donor strand exchange, in which an N-terminal extension of the incoming protein subunit replaces the chaperone's donor β-chain in the previously bound subunit. Type 1 pili are formed

by the assembly of subunits, starting with the FimH adhesin, and the pilus is extended by the gradual addition of additional chaperone-subunit complexes to the base of the fiber. First, spacer proteins are added, followed by the major filament subunit FimA/PapA. Each subunit interacts specifically with its respective neighboring subunit in the filament, and binding specificity is determined by the mechanism of donor strand exchange.[47–50]

Based on our experimental and computational data, we suggested that the interaction between the tafi subunits might be through the donor strand exchange, as in P1T, However, the chaperone-usher mechanism cannot be used in archaeal cells, that have to assemble the tafi in the cytoplasmic membrane (CM) and not in OM.[21]

Using the AlphaFold 2 server, we generated hypothetical structures of oligomers composed of matured subunits TafA, TafC, and TafE subunits. Models of the structures of the TafC/TafE and TafC/TafA heterodimeric complexes show an interaction mediated by the complementarity nature of the N-terminal donor strand of TafC to TafE or TafA. This is accompanied by a displacement of donor chains of these proteins outward (Figure 4B). Modeling for tetramers of TafC, TafA, and two TafE (Figure 4C) and TafC, TafE, and two TafA (Figure 4D), revealed filamentous complexes formed by linearly interacting molecules. In these cases, TafC is at the end, with its laminin G-like domain directed outward. The N-terminal donor strand of TafC interacts with TafE to activate it. Activated TafE then interacts with TafE or TafA via its donor strand. When modeling heteromeric complexes from TafA, TafC, and TafE, the interaction of TafC with TafE through the donor strand is typically preferred over the interaction between TafC and TafA (except in the absence of TafE). This suggests that the direct interaction between TafC and the main subunit, TafA, may be less stable, and TafE is necessary to facilitate the polymerization of the TafA subunit. At the same time, PDBsum analyses[51] of inter-subunit interactions indicate stronger interactions between identical TafA subunits within the tetramer TafC/2xTafA/TafE, with more salt bridges and hydrogen bonds compared to other interactions (see Table S4).

It is interesting to note that the use of AlphaFold to model the structure of the Mth60 fimbriae from *Methanothermobacter thermoautotrophicus* also results in similar fimbriae in which subunits interact through the exchange of donor strands. As mentioned earlier, the Mth60-fimbrin is transported via the Sec pathway.[7,14] Modeling of the mature Mth60 monomer using AlphaFold yields a structure in which the potential N-terminal donor strand faces away from the globular region (“open conformation”) and is available for interaction with other monomers without requiring activation, forming a homomeric filament (Figure S18). This agrees with the reported polymerization of isolated Mth60 monomers *in vitro*.[7,14] In contrast, in the *Har. hispanica* TafA monomer, the potential donor strand (N-terminal fragment) is directed toward the globule (“closed conformation”). However, its activation by the TafC protein causes it to turn outwards and start the polymerization process (Figure S18 and Figure 4B). Therefore, TafC is an assembly activator for subunits that provides the initial donor strand during tafi formation, triggering an exchange of donor strands between TafE and TafA. This mechanism is similar to the one proposed for the formation of the main component of the *Bacillus subtilis* biofilm matrix, TasA fibers, when the auxiliary protein TapA nucleates TasA fibers formation by providing the initial donor strand for the major protein TasA.[52]

The AlphaFold analysis reveals that the most appropriate minimal model (pLDDT = 7, pTM = 0.654, ipTM = 0.632) for tafi is a linear structure composed of four proteins arranged in the following order: TafC at the tip, TafE as an adapter, TafA as the filament core, and TafF as a potential anchor to the cell wall (Figure 4E). As we mentioned previously, the N-terminal region of TafF has structural similarity to the key subunits, TafA, TafC and TafE. We can see that the complementary donor strand of TafA interacts with the N-terminal region of TafF. Because TafF contains a C-membrane-spanning region, it may be involved in tafi anchoring to the cell membrane. Based on the available data, it is possible to hypothesize an assembly mechanism for the tat-fimbriae (Figure 5), although we still do not fully understand the detailed mechanisms behind tafi assembly and its regulation.

Modeling of tafi structures involving TafA oligomers using the AlphaFold 3 web service and other computational resources

We used the AlphaFold 3 (AF3) server[53] and the ClusPro protein–protein docking method[54] to see whether the formation of tafi polymer chains proposed in Figure 5 is possible in principle. The visualization of the models was done using the Jmol program.[55] Carbohydrate-binding sites in the TafA AF3 model were predicted using a specific program.[56] Figure 6A shows an AF3 model of the TafA hexamer, which illustrates the self-assembly of a homooligomeric chain. In this model, the TafA monomer undergoes the transition from a closed conformation, in which the N-terminus is part of the protein globule (Figure 4A) to an open conformation where the N-terminus deviates by approximately 90° from the main globule and becomes incorporated into the oligomer as a donor chain. A possible explanation for this transition could be the structural disorder of the N-terminal region of the TafA protein. This is evident in the AlphaFold models, particularly in regions with pIDDT values less than 50–70 (data not shown). The disorder leads to conformational variability and flexibility in the N-terminus of TafA, which may allow it to adopt an open conformation during interaction with other proteins.

Our electrophoretic, mass spectrometric, and electron microscopic results allow us to consider two main types of tafi structures, corresponding to the Figure 5 scheme, which involves TafA homooligomers. We call the heterooligomeric tafi chains involving the TafF terminal protein, which has transmembrane properties and ensures the fixation of the chains in the cytoplasmic membrane, (Figures 6C–6E), “finalized chains.” The N-terminal domain of the mature protein, with a length of 140 amino acid residues (Figure 4E), was used as TafF in the modeling of the polymer chains shown in Figure 6. The tafi chains that do not contain the TafF-terminal protein are referred to as “nonfinalized chains” (Figures 6A and 6B). It can be assumed that these nonfinalized chains are secreted into the extracellular environment and play a role in the formation of the extracellular matrix and other physiological and biochemical processes.

Importantly, for a nonfinalized seven-membered TafA heterooligomer with TafC without TafE involvement (Figure S19A), AF3

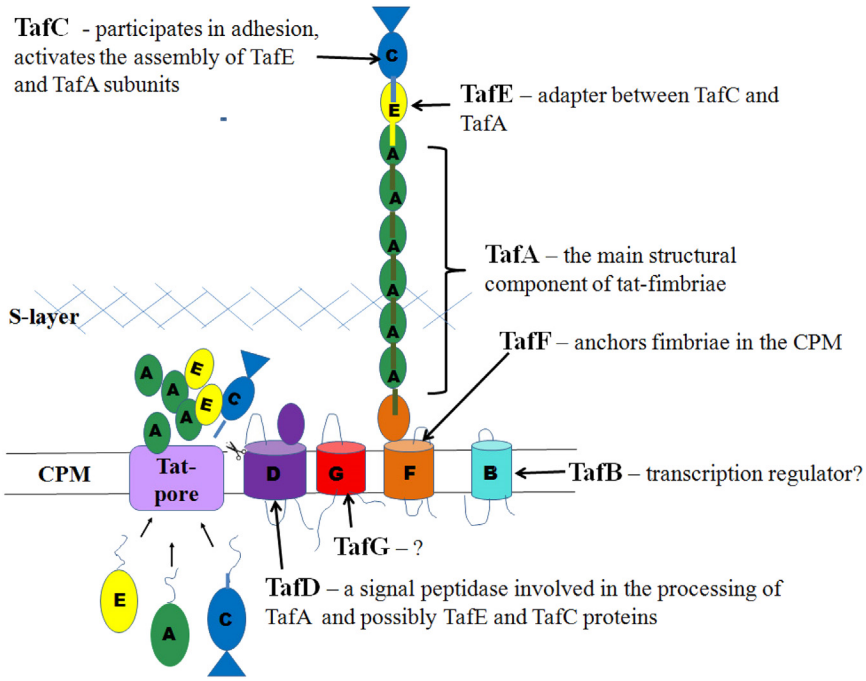


Figure 5. Hypothetical model for tat-fimbriae assembly

The protomers, TafA, TafC, and TafE, are translocated across the cytoplasmic membrane (CPM) through the Tat-pore. After the signal sequences are cleaved by peptidase TafD, these proteins are assembled using a donor strand exchange mechanism. TafC acts as an activator for assembly, while TafE serves as an adapter between TafC and TafA. The membrane protein, TafF, terminates assembly by anchoring tat-fimbriae in the cytoplasmic membrane. The ovals represent the TafA-like domains, which are similar to immunoglobulin folds, and the triangle represents the TafC laminin G-like domain. The mechanisms that coordinate the subunit assembly process have not yet been fully elucidated. For more information, please see the text. See also Figures 4 and 6.

predicts a structure that is implausible in terms of the scheme in Figure 5 and the experimental results of this work. However, when TafE is included in this set of proteins, it leads to the prediction of a linear, non-finalized chain (Figure 6B), consistent with the scheme in Figure 5. On the other hand, AF3 simulations with the complete set of proteins (TafC, TafE, TafA homo-oligomer, and TafF) again give an inconsistent result (Figure S19B), and only the exclusion of TafC from this set results in a consistent prediction of an incomplete finalized tafi chain (Figure 6C).

Thus, the generation of a complete nonfinalized tafi chain by AF3 (Figure 6B) requires the presence of the TafC/TafE pair in it, whereas a correct finalized chain (correct in terms of the Figure 5 scheme; Figure 6C) can be obtained only by excluding TafC from it. The existence and significance of these tafi variants lacking TafC, which possess adhesive properties, is the subject of our further studies.

To obtain a model reflecting the assumed finalization of the complete nonfinalized tafi chain (Figure 6B) by attaching the TafF terminal protein to it, we used the ClusPro server, which uses a method considered one of the best among the programs working in the so-called solid-state approximation.[54] It uses AF3-obtained .pdb files as the input 3D structures: a receptor (in our case, a nonfinalized tafi chain; Figure 6B) and a ligand (in our case, the N-terminal domain of a mature TafF protein with a length of 140 amino acid residues; Figure 4E). As a result, we arrived at a model for the finalized tafi chain (Figure 6D), that fully corresponded to the hypothetical scheme of tat-fimbriae formation in Figure 5.

Given that the main tafi protein, TafA, has been shown to be glycosylated (Figure S3), we have used the advanced capabilities of the AF3 web service to investigate the effect of glycosylation on the AF3's predictions of tafi structures. Because the precise structure of the glycosyl groups in the TafA glycoprotein is not yet known, we used the simplest model of N-glycosylation with the monosaccharide N-acetyl-D-glucosamine (GlcNAc, NAG). The residue of this sugar is located at the beginning of the glycan chain and is attached to an asparagine residue during N-glycosylation in various glycoproteins in archaea and eukaryotes.[57] We used the potential carbohydrate-binding sites marked in Figure 2A, which corresponded to predictions made by the program[56] with a prediction probability >85% (cutoff = 0.85). The results (Figure 6E) show that the predicted structure of the finalized tafi chain by AF3 and ClusPro was robust relative to the glycosylation of the TafA protein by N-acetyl-D-glucosamine.

The reliability and quality of the interprotein bonds in the models of the complete finalized tafi chain (Figures 6D and 6E) were characterized by using the results obtained from the PDBePISA program (https://www.ebi.ac.uk/pdbe/pisa).[58] Figure S20 visualizes the interfaces of interactions between the components of the tafi heterocomplex (Figure 6E). For each of the constituent proteins (except for TafA6), in addition to the main interface formed by the N-terminal donor chain (24–34 residues), an additional, relatively short interface (8–9 residues) was found that interacts with the protein following the immediate partner for this protein (see the right side of Figure S20). This additional interface contributes to the stabilization of the heterocomplex.

According to the work,[58] negative $\Delta^{i}G$ values correspond to hydrophobic interfaces and indicate the thermodynamic affinity of the protein structures being considered. The $\Delta^{i}G$ p-value <0.5, in contrast to $\Delta^{i}G$ p-value >0.5, indicates that the interactions are specific. However, the $\Delta^{i}G$ estimates do not take into account the effects of the hydrogen bonds (HBs; −0.44 kcal/mol per bond), salt bridges (SBs; −0.3 kcal/mol per salt bridge), and disulfide bonds (DS; −4 kcal/mol per bond) that may be present at the interface. The numbers of HBs and SBs are 17–34 and 6–9, respectively, for N-terminal donor chains and 1–10 and 0–2, respectively, for additional short interfaces. There is no DS for any of the interfaces. The $\Delta^{i}G$ p-value is 0.211–0.312 for N-terminal donor chains and 0.386–0.496 for the additional short

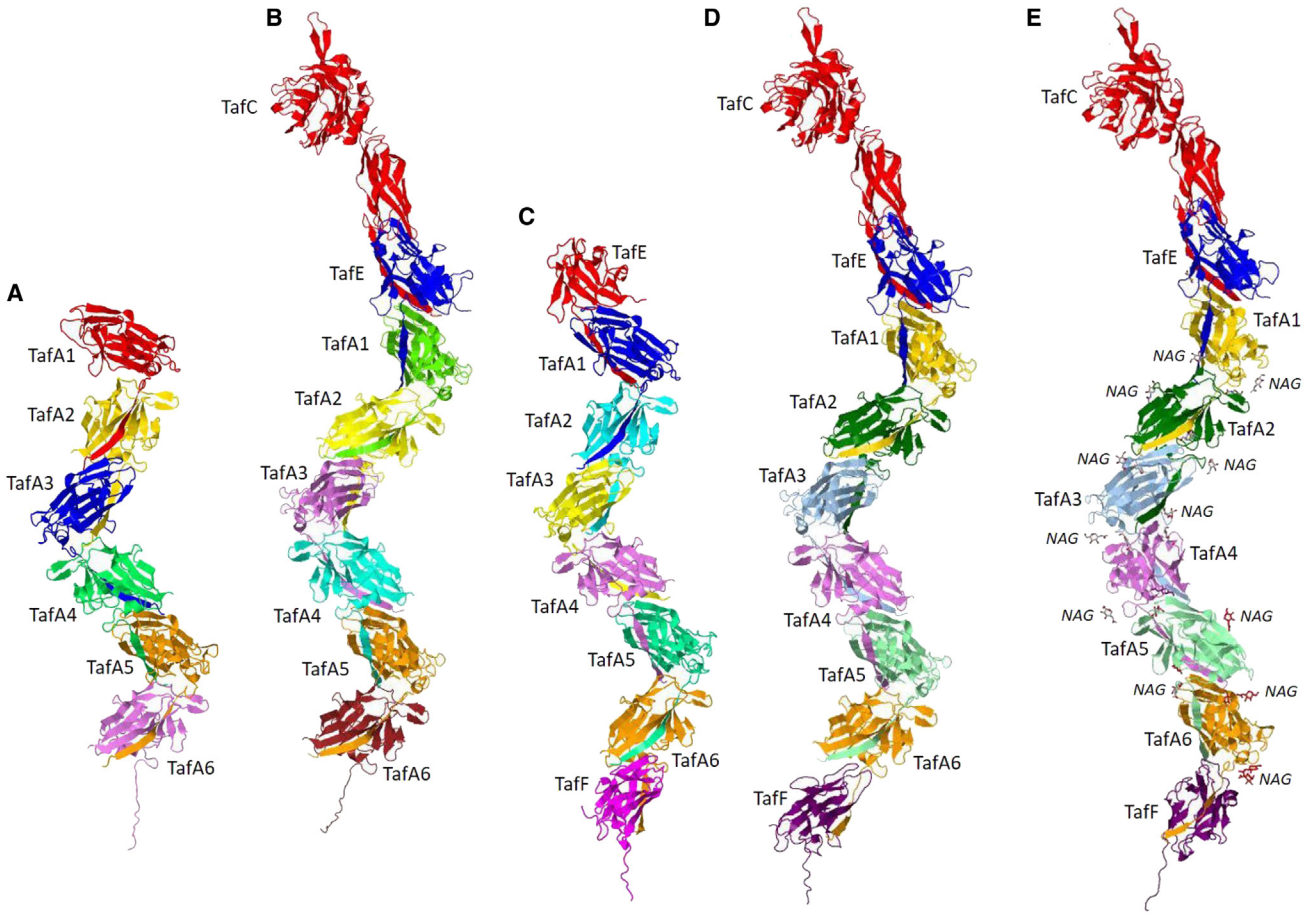


Figure 6. AF3-predicted results for the assembly of tat filaments corresponding to their formation scheme in Figure 5

(A) Self-assembly of the hexamer of the main TafA protein.

(B) Assembly of a nonfinalized (without TafF) TafA chain involving the TafC and TafE proteins.

(C) Assembly of an incomplete (without TafC) finalized TafA chain involving the TafE and TafF proteins.

(D) Assembly of a complete finalized TafA chain involving the TafC, TafE, and TafF proteins, which was obtained by combining the AF3 and ClusPro methods.

(E) Assembly of the complete finalized TafA chain involving TafC, TafE, and TafF proteins, which was obtained by combining the AF3 and ClusPro methods and taking into account the glycosylation of the TafA1–TafA6 proteins by N-acetyl-D-glucosamine (NAG). Carbohydrate-binding sites N20, N102, N121, and N140 were predicted by the program,[56] with a prediction probability >85% (cutoff = 0.85). See also Figure 5, S3, S19 and S20.

interfaces, indicating a higher specificity of interactions in the interfaces with N-terminal donor chains.

Tafi are a haloarchaeal-adapted kind of a broader class of archaeal filamentous structures that assemble via a donor strand exchange mechanism

Since the Tat pathway is the known as the primary protein secretion system in halophilic archaea,[20,21,26] it is reasonable to investigate whether tafi-like fimbriae also exist in non-halophilic archaea. A search for *Har. hispanica* TafA protein homologs using BLAST among non-halophilic archaea did not yield any results. Therefore, we looked for TafD signal peptidase homologues, as a more conserved protein. We found that TafD homologs with approximately 30% identities (45% positives) are present in the genomes of various archaea from the genera *Thermococcus* (over 60 species), *Pyrococcus* (seven species), *Archaeoglobus* (four species), *Palaeococcus* (two species) and one of *Geoglobus* species. These proteins, such as *Har. hispanica* TafD, are peptidases with two-domains and four transmembrane helices, with the N-terminal domain is responsible for peptidase activity, and the C-terminal region, similar to *Har. hispanica*, having an immunoglobulin-like structure. Figure S21 shows a simplified evolutionary tree obtained using the BLAST algorithm for *Har. hispanica* TafD and its homologs in selected representatives of non-halophilic archaea.

We examined the surrounding regions of the respective peptidase genes and found that they are typically part of a conserved cluster of 5–7 genes. These gene clusters for the four characteristic species, *Thermococcus sibiricus* MM 739, *Archaeoglobus fulgidus* DSM 4304, *Thermococcus kodakaraensis* KOD1, and *Pyrococcus horikoshii* OT3 are shown in Figure 7. These selected archaeal species are highlighted in yellow in Figure S21. Along with the gene that encodes the signal peptidase (TSIB_1249, AF_2078, TK_1703, and PHO563, respectively), in all cases,

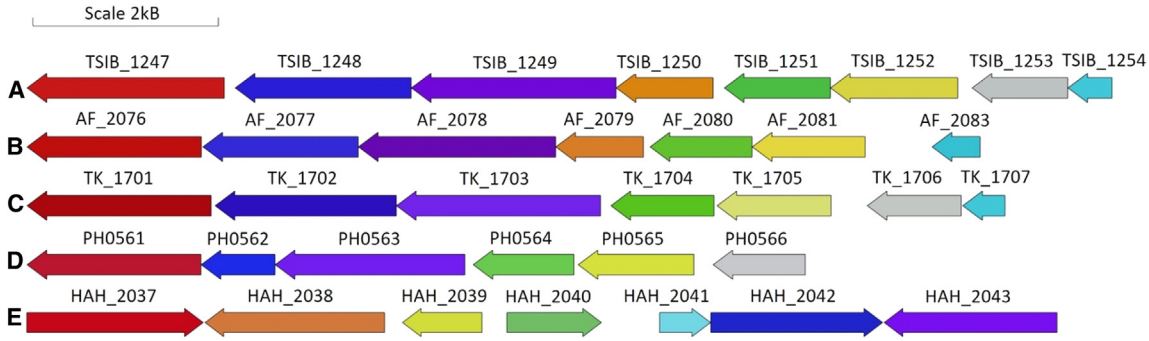


Figure 7. *taf*-like gene clusters in non-halophilic archaea compared to that in *Har. hispanica*

(A) *Thermococcus sibiricus* MM 739.

(B) *Archaeoglobus fulgidus* DSM 4304.

(C) *Thermococcus kodakaraensis* KOD1.

(D) *Pyrococcus horikoshii* OT3.

(E) *Har. hispanica* ATCC33960. Genes encoding hypothetical structural analogues are colored the same. The image was created using the Gene Graphics server (https://katlabs.cc/genegraphics/). See also Figure S21.

there are genes that encode the DUF5305 family proteins homologous to TafG (TSIB_1247, AF_2076, TK_1701, and PHO561, respectively). In addition, there are two or three genes that encode proteins with about 150–200 residues and annotated as DUF1102-domain containing proteins (the DUF1102 family consists of several hypothetical archaeal proteins of unknown function). These include proteins: TSIB_1250, TSIB_1251, TSIB_1252, AF_2079, AF_2080, AF_2081, TK_1704, TK_1705, PHO564 and PHO565, respectively. The N-terminal regions of all these DUF1102 proteins contain a Sec motif. A gene encoding a protein of approximately 300 amino acids, that contains either a Sec/SPI (TSIB_1248 and AF_2077) or a lipoprotein (Sec/SPII) (TK_1702) signal peptide has been found in this cluster in the genomes of *Thermococcus sibiricus*, *Archaeoglobus fulgidus*, *Thermococcus kodakaraensis,* and similar species. However, *Pyrococcus horikoshii* and the other seven *Pyrococci* and *Thermococcus chitinophagus* do not possess such a gene. Instead, there is a gene that encodes a small protein PH0562 of approximately 130 amino acids with a Sec/SPI pattern. In addition, the gene encoding a small protein annotated as a transcriptional regulator (probable TafB analog) also adjoins this gene cluster in *Thermococcus sibiricus*, *Archaeoglobus fulgidus* and *Thermococcus kodakaraensis* (TSIB_1254, AF_2083, and TK_1707). However, this gene is not present in *Pyrococcus horikoshii*.

AlphaFold modeling of the structures of the corresponding non-halophilic proteins and their complexes revealed striking similarities to Taf proteins (Table S3). Figure 8A and 8B show structures of heterotetramers obtained by AlphaFold for four mature proteins from *Thermococcus sibiricus* and *Archaeoglobus fulgidus*, whose precursors contain a Sec motif. These proteins are able to form tafi-like linear oligomers using the mechanism of donor strand exchange. In these heterotetrameric complexes, non-halophilic proteins containing the Sec motif may be structural analogs of TafC, TafE, TafA, and TafF. Two-domain proteins TSIB_1248 and AF_2077, are analogs of TafC and are located at the tip of the filament, and, apparently, they can initiate assembly. Proteins containing the DUF1102 domain are analogs of the major tat-fimbrial protein TafA and the adapter protein, TafE. According to the analysis using the Dali server, the *Thermococcus sibiricus* DUF1102 domain-containing protein structures generated using AlphaFold, fit well into the TafA structure (Z-scores: 8.8 to 9.9, and %identity: 10 to 13) (Table S3). The proteins TSIB_1250 and AF_2079 are likely functional analogs of TafF, which terminates fimbrial assembly. However, they do not contain transmembrane regions. Based on the hypothetical tetramer structures obtained from AlphaFold, both *Thermococcus kodakaraensis* and *Pyrococcus horikoshii* lack a TafF-like protein and *Pyrococcus horikoshii* protein PHO562 is a single-domain analogue of the TafC protein (Figure 8C and 8D). It is noteworthy that a similar set of structurally similar proteins is involved in the formation of haloarchaeal tafi, as well as hypothetical fimbriae in non-halophilic archaea. However, the most noticeable differences between these proteins are the different predicted secretion pathways for their protomers (Tat and Sec), and the absence of a transmembrane region in the TafF analog that would terminate the fimbriae assembly or the absence of the TafF protein itself.

We hypothesize that the tafi-like structures in non-halophilic archaea may be evolutionarily older than those in haloarchaea. The acquisition of these genes by haloarchaeal cells was accompanied by the shuffling of genes from a single operon and required a shift in the secretion pathway from Sec to the Tat, which is better adapted to high salt concentrations.

The development of unique methods for cryo-electron microscopy (cryo-EM) and modern bioinformatics servers, such as AlphaFold, have significantly improved our ability to study the structure and function of various biological macromolecules. Recently, two research teams reported on the atomic structures of two previously unknown archaeal surface filaments: the bundling pili from *Pyrobaculum calidifontis*[59] and the thread filaments from *Sulfolobus acidocaldarius*,[60] both structures were determined using the cryo-electron microscopy combined with AlphaFold analysis. Both filamentous structures exhibit a remarkable similarity to P1T bacterial adhesive pili and *Har. hispanica* tat-fimbriae. The assembly of these archaeal structures seems similar to that of chaperone-usher pili,[61] TasA fibers,[52] and type

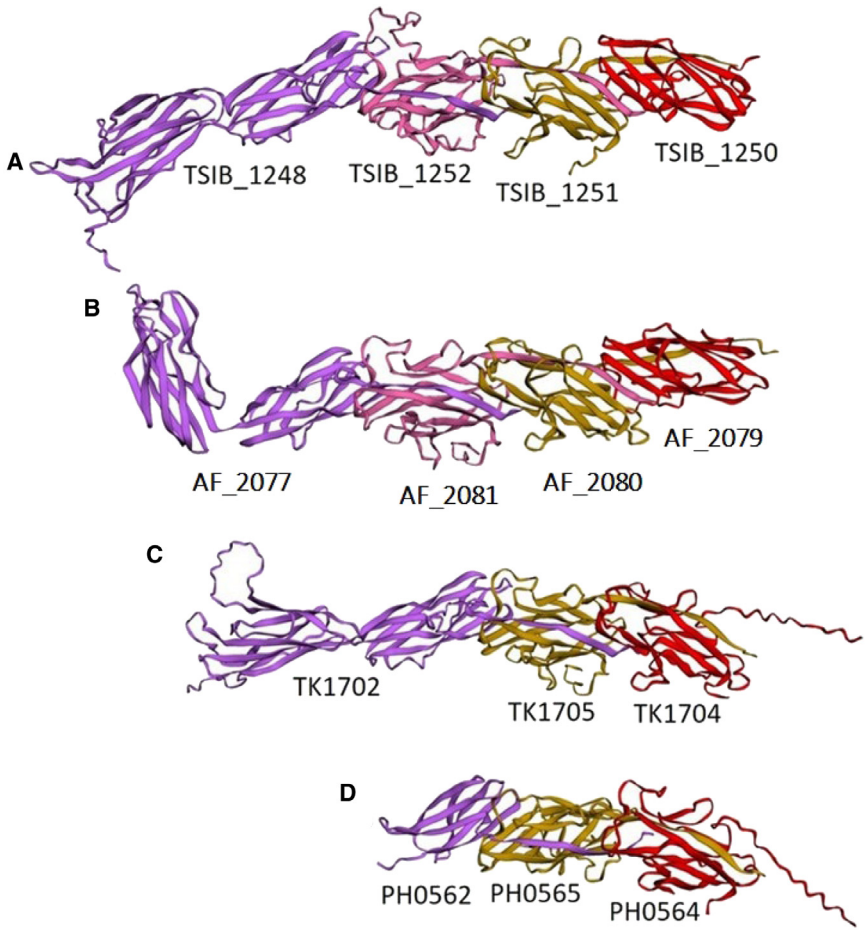


Figure 8. AlphaFold 2 predictions for heterotetrameric and heterotrimeric complexes of tafi-like proteins in non-halophilic archaea

(A) *Thermococcus sibiricus*, (B) *Archaeoglobus fulgidus*, (C) *Thermococcus kodakaraensis* and (D) *Pyrococcus horikoshii.* See also Table S3.

V pili[62] in that they all use the donor-strand exchange mechanism. However, the genomic organization of genes encoding the main subunits of surface structures in *Pyrobaculum calidifontis* and *Sulfolobus acidocaldarius* is significantly different from those of the *tafA* gene in *Har. hispanica* and *tafA*-like genes in *Thermococcus sibiricus*, *Archaeoglobus fulgidus*, *Thermococcus kodakaraensis*, and *Pyrococcus horikoshii*. This difference is primarily due to the lack of homologs for two genes, *tafD* and *tafG*, in corresponding gene clusters in these organisms.

Despite the lack of homology between the protein subunits of filamentous structures mentioned, the principle of their formation (common packaging of subunits through the mechanism of donor chain exchange) is similar in different species of bacteria and archaea, and it seems to be more widespread in nature than previously thought.

Limitations of the study

The present study demonstrates that the secretion of TafA, a major subunit of tat-fimbriae, occurs through the Tat pathway. It was shown that knockout of the *tafA* or *tafD* genes leads to a cessation of tat-fimbriae synthesis. Additionally, the involvement of signal peptidase TafD in processing the major fimbrial subunit TafA has been confirmed. However, the roles of other proteins encoded by the taf cluster, such as TafB, TafC, TafE, TafF, and TafG, in fimbrial assembly have not been directly investigated using genetic knockouts. Rather, assumptions about their functions were made based on indirect evidence. Therefore, further research is required to elucidate the molecular mechanisms involved in the assembly of tat-fimbriae.

Conclusion

In this study, we report on the discovery and initial analysis of an unusual type of haloarchaeal surface structure. We have named these structures ‘‘tat-fimbriae’’ – ‘‘tafi’’- to emphasize the uniqueness of their predicted secretory pathway. We found that the key structural proteins of tafi – TafA, TafC, and TafE - contain a specific ‘‘twin-arginine’’ motif, which is typically found in plant, bacterial and archaeal proteins using the Tat secretory pathway. Our experimental data demonstrated the use of the Tat-pathway in the formation of these haloarchaeal surface filamentous structures. This may be a preferred way of evolutionary adaptation in highly saline environments, considering that many surface proteins in haloarchaea (various extracellular enzymes, proteins responsible for adhesion, and redox processes) often use the Tat-pathway for secretion.[20,21,26,63]

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead contact, Dr. Mikhail G. Pyatibratov (bratov@vega.protres.ru).

Materials availability

All unique/stable reagents and antibodies generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

Data reported in this article will be shared by the lead contact upon request.

This study did not generate any unique datasets or code.

Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was funded by the National Key R&D Program of China (grant No. 2020YFA0906800) awarded to H.X., the Central Asian Drug Discovery and Development Center of Chinese Academy of Sciences grant No. CAM202202 awarded

to H.X. and the Russian Foundation for Basic Research (RFBR) grant No. 19-04-01327A awarded to A.S. We thank T. Allers for providing the pTA1228 plasmid, M. Pohlschröder for kindly presenting the *Haloferax volcanii* strains, M.Y. Suvorina for assistance with MS analysis, and K.F. Jarrell, V.A. Meshcheryakov and F. Pfeiffer for their thoughtful and critical reading of the article. The authors thank B.S. Melnik for CD measurements and T.N. Melnik for calorimetric measurements. Electron microscopy was carried out with the support of the Moscow State University Development Program (PNR 5.13) and mass spectrometric analyses were performed at the United Pushchino Center for "Structural and functional studies of proteins and RNA" (Project No. 584307). Part of the immunochemical and computational research was done at the IBPPM RAS Laboratory of Immunochemistry (research topic no. 1022040700963-8).

AUTHOR CONTRIBUTIONS

Conceptualization, A.V.G., D.Z., A.S.S., S.Yu.S., H.X., and M.G.P.; investigation, A.V.G., D.Z., A.S.S., M.Yu.T., S.Yu.S., E.Yu.P., A.K.S., G.L.B., O.M.S., I.I.K., J.L., and M.G.P.; formal analysis, A.V.G., D.Z., M.G.P., A.S.S., M.Yu.T., A.K.S., and S.Yu.S.; resources, H.X., A.S.S., M.G.P., S.Yu.S., I.I.K., and A.K.S.; writing – original draft, A.V.G., D.Z., A.S.S., S.Yu.S., H.X., and M.G.P.; writing – review and editing, D.Z., S.Yu.S., M.Yu.T., H.X,. and M.G.P.; visualization, A.V.G., D.Z., S.Yu.S., M.Yu.T., A.K.S., and M.G.P.; project administration, H.X., M.G.P., A.S.S., and S.Yu.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Haloarchaea
  - Bacteria
  - Antibody preparation
- METHOD DETAILS
  - Isolation of tafi and their separation from archaella
  - Mass spectrometry analysis
  - Electron microscopy
  - TafA (HAH_0240) purification
  - Construction and confirmation of the deletion mutants
  - Plasmid-based complementation for *tafA*, *tafD*, and *pibD* gene deletions
  - Heterologous expression of *Har. hispanica tafA-tafG* genes in *Haloferax volcanii*
  - Preparation of recombinant TafA (HAH_0240) by heterologous expression in *E. coli*
  - Preparation of rabbit anti-TafA antibody
  - Western blotting
  - Structural prediction
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.111793.

Received: August 10, 2023
Revised: October 13, 2023
Accepted: January 9, 2025
Published: January 10, 2025

REFERENCES

1. Chaban, B., Ng, S.Y.M., and Jarrell, K.F. (2006). Archaeal habitats – from the extreme to the ordinary. Can. J. Microbiol. *52*, 73–116. https://doi.org/10.1139/w05-147.
2. Mihajlovski, A., Alric, M., and Brugère, J.F. (2008). A putative new order of methanogenic Archaea inhabiting the human gut, as revealed by molecular analyses of the *mcrA* gene. Res. Microbiol. *159*, 516–521. https://doi.org/10.1016/j.resmic.2008.06.007.
3. Baker, B.J., De Anda, V., Seitz, K.W., Dombrowski, N., Santoro, A.E., and Lloyd, K.G. (2020). Diversity, ecology and evolution of Archaea. Nat. Microbiol. *5*, 887–900. https://doi.org/10.1038/s41564-020-0715-z.
4. Jarrell, K.F., and Albers, S.V. (2012). The archaellum: An old motility structure with a new name. Trends Microbiol. *20*, 307–312. https://doi.org/10.1016/j.tm.2012.04.007.
5. Jarrell, K.F., Ding, Y., Nair, D.B., and Siu, S. (2013). Surface appendages of Archaea: structure, function, genetics and assembly. Life *3*, 86–117. https://doi.org/10.3390/life3010086.
6. Lassak, K., Ghosh, A., and Albers, S.V. (2012). Diversity, assembly and regulation of archaeal type IV pili-like and non-type-IV pili-like surface structures. Res. Microbiol. *163*, 630–644. https://doi.org/10.1016/j.resmic.2012.10.024.
7. Thoma, C., Frank, M., Rachel, R., Schmid, S., Näther, D., Wanner, G., and Wirth, R. (2008). The Mth60 fimbriae of *Methanothermobacter thermoautotrophicus* are functional adhesins. Environ. Microbiol. *10*, 2785–2795. https://doi.org/10.1111/j.1462-2920.2008.01698.x.
8. Jarrell, K.F., Albers, S.V., and Machado, J.N.d.S. (2021). A comprehensive history of motility and archaellation in Archaea. FEMS Microbes *2*, xtab002. https://doi.org/10.1093/femsmc/xtab002.
9. Pohlschröder, M., and Esquivel, R.N. (2015). Archaeal type IV pili and their involvement in biofilm formation. Front. Microbiol. *6*, 190. https://doi.org/10.3389/fmicb.2015.00190.
10. Muller, D.W., Meyer, C., Gurster, S., Kuper, U., Huber, H., Rachel, R., Wanner, G., Wirth, R., and Bellack, A. (2009). The Iho670 fibers of *Ignicoccus hospitalis*: a new type of archaeal cell surface appendage. J. Bacteriol. *191*, 6465–6468. https://doi.org/10.1128/JB.00858-09.
11. Meyer, C., Heimerl, T., Wirth, R., Klingl, A., and Rachel, R. (2014). The Iho670 fibers of *Ignicoccus hospitalis* are anchored in the cell by a spherical structure located beneath the inner membrane. J. Bacteriol. *196*, 3807–3815. https://doi.org/10.1128/JB.01861-14.
12. Braun, T., Vos, M.R., Kalisman, N., Sherman, N.E., Rachel, R., Wirth, R., Schröder, G.F., and Egelman, E.H. (2016). Archaeal flagellin combines a bacterial type IV pilin domain with an Ig-like domain. Proc. Natl. Acad. Sci. USA *113*, 10352–10357. https://doi.org/10.1073/pnas.1607756113.
13. Zolghadr, B., Klingl, A., Rachel, R., Driessen, A.J.M., and Albers, S.V. (2011). The bindosome is structural component of the *Sulfolobus solfataricus* cell envelope. Extremophiles *15*, 235–244. https://doi.org/10.1007/s00792-010-0353-0.
14. Sarbu, C. (2013). Untersuchung der Mth60-Fimbrien von *Methanothermobacter thermoautotrophicus*. Thesis of the University of Regensburg (PhD), . Preprint at. https://doi.org/10.5283/epub.28103.
15. Moissl, C., Rachel, R., Briegel, A., Engelhardt, H., and Huber, R. (2005). The unique structure of archaeal "hami", highly complex cell appendages with nano-grappling hooks. Mol. Microbiol. *56*, 361–370. https://doi.org/10.1111/j.1365-2958.2005.04294.x.
16. Perras, A.K., Daum, B., Ziegler, C., Takahashi, L.K., Ahmed, M., Wanner, G., Klingl, A., Leitinger, G., Kolb-Lenz, D., Gribaldo, S., et al. (2015). S-layers at second glance? Altiarchaeal grappling hooks (hami) resemble archaeal S-layer proteins in structure and sequence. Front. Microbiol. *6*, 543. https://doi.org/10.3389/fmicb.2015.00543.
17. Nickell, S., Hegerl, R., Baumeister, W., and Rachel, R. (2003). *Pyrodictium* cannulae enter the periplasmic space but do not enter the cytoplasm, as

revealed by cryo-electron tomography. J. Struct. Biol. *141*, 34–42. https://doi.org/10.1016/S1047-8477(02)00581-6.

18. Kreitner, R., Munte, C.E., Singer, K., Stetter, K.O., Horn, G., Kremer, W., and Kalbitzer, H.R. (2020). Complete sequential assignment and secondary structure prediction of the cannulae forming protein CanA from the hyperthermophilic archaeon *Pyrodictium abyssi*. Biomol. NMR Assign. *14*, 141–146. https://doi.org/10.1007/s12104-020-09934-x.
19. Pohlschröder, M., Pfeiffer, F., Schulze, S., and Halim, M.F.A. (2018). Archaeal cell surface biogenesis. FEMS Microbiol. Rev. *42*, 694–717. https://doi.org/10.1093/femsre/fuy027.
20. Rose, R.W., Brüser, T., Kissinger, J.C., and Pohlschröder, M. (2002). Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. Mol. Microbiol. *45*, 943–950. https://doi.org/10.1046/j.1365-2958.2002.03090.x.
21. Pohlschröder, M., Giménez, M.I., and Jarrell, K.F. (2005). Protein transport in Archaea: Sec and twin arginine translocation pathways. Curr. Opin. Microbiol. *8*, 713–719. https://doi.org/10.1016/j.mib.2005.10.006.
22. Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. *37*, 420–423. https://doi.org/10.1038/s41587-019-0036-z.
23. Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005). Prediction of twin-arginine signal peptides. BMC Bioinf. *6*, 167. https://doi.org/10.1186/1471-2105-6-167.
24. Szabo, Z., and Pohlschröder, M. (2012). Diversity and subcellular distribution of archaeal secreted proteins. Front. Microbiol. *3*, 207. https://doi.org/10.3389/fmicb.2012.00207.
25. Dalbey, R.E., and Kuhn, A. (2012). Protein Traffic in Gram-negative bacteria – how exported and secreted proteins find their way. FEMS Microbiol. Rev. *36*, 1023–1045. https://doi.org/10.1111/j.1574-6976.2012.00327.x.
26. Ghosh, D., Boral, D., Vankudoth, K.R., and Ramasamy, S. (2019). Analysis of haloarchaeal twin-arginine translocase pathway reveals the diversity of the machineries. Heliyon *5*, e01587. https://doi.org/10.1016/j.heliyon.2019.e01587.
27. Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. *41*, W349–W357. https://doi.org/10.1093/nar/gkt381.
28. Matagne, A., Joris, B., and Frère, J.M. (1991). Anomalous behaviour of a protein during SDS/PAGE corrected by chemical modification of carboxylic groups. Biochem. J. *280*, 553–556. https://doi.org/10.1042/bj2800553.
29. Pyatibratov, M.G., Beznosov, S.N., Rachel, R., Tiktopulo, E.I., Surin, A.K., Syutkin, A.S., and Fedorov, O.V. (2008). Alternative flagellar filament types in the haloarchaeon *Haloarcula marismortui*. Can. J. Microbiol. *54*, 835–844. https://doi.org/10.1139/w08-076.
30. Wieland, F., Paul, G., and Sumper, M. (1985). Halobacterial flagellins are sulfated glycoproteins. J. Biol. Chem. *260*, 15180–15185. https://doi.org/10.1016/S0021-9258(18)95719-4.
31. Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S., and Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis *14*, 1023–1031. https://doi.org/10.1002/elps.11501401163.
32. Domogatskaya, A., Rodin, S., and Tryggvason, K. (2012). Functional diversity of laminins. Annu. Rev. Cell Dev. Biol. *28*, 523–553. https://doi.org/10.1146/annurev-cellbio-101011-155750.
33. Thaw, P., Sedelnikova, S.E., Muranova, T., Wiese, S., Ayora, S., Alonso, J.C., Brinkman, A.B., Akerboom, J., van der Oost, J., and Rafferty, J.B. (2006). Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. Nucleic Acids Res. *34*, 1439–1449. https://doi.org/10.1093/nar/gkl009.
34. Auclair, S.M., Bhanu, M.K., and Kendall, D.A. (2012). Signal peptidase I: Cleaving the way to mature proteins. Protein Sci. *21*, 13–25. https://doi.org/10.1002/pro.757.
35. Foster, T.J. (2019). The MSCRAMM family of cell-wall-ahchored surface proteins of gram-positive cocci. Trends Microbiol. *27*, 927–941. https://doi.org/10.1016/j.tim.2019.06.007.
36. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. *305*, 567–580. https://doi.org/10.1006/jmbi.2000.4315.
37. Hallgren, J., Tsirigos, K.D., Pedersen, M.D., Armenteros, J.J.A., Marcatili, P., Nielsen, H., Krogh, A., and Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at: bioRxiv. https://doi.org/10.1101/2022.04.08.487609v1
38. Tripepi, M., Imam, S., and Pohlschröder, M. (2010). *Haloferax volcanii* flagella are required for motility but are not involved in PibD-dependent surface adhesion. J. Bacteriol. *192*, 3093–3102. https://doi.org/10.1128/JB.00133-10.
39. Pyatibratov, M.G., Syutkin, A.S., Quax, T.E.F., Melnik, T.N., Papke, R.T., Gogarten, J.P., Kireev, I.I., Surin, A.K., Beznosov, S.N., Galeva, A.V., and Fedorov, O.V. (2020). Interaction of two strongly divergent archaellins stabilizes the structure of the *Halorubrum* archaellum. MicrobiologyOpen *9*, e1047. https://doi.org/10.1002/mbo3.1047.
40. Esquivel, R.N., and Pohlschröder, M. (2014). A conserved type IV pilin signal peptide H-domain is critical for the post-translational regulation of flagella-dependent motility. Mol. Microbiol. *93*, 494–504. https://doi.org/10.1111/mmi.12673.
41. Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. Appl. Environ. Microbiol. *76*, 1759–1769. https://doi.org/10.1128/AEM.02670-09.
42. Tripepi, M., Esquivel, R.N., Wirth, R., and Pohlschröder, M. (2013). *Haloferax volcanii* cells lacking the flagellin FlgA2 are hypermotile. Microbiology *159*, 2249–2258. https://doi.org/10.1099/mic.0.069617-0.
43. Cramer, P. (2021). AlphaFold2 and the future of structural biology. Nat. Struct. Mol. Biol. *28*, 704–705. https://doi.org/10.1038/s41594-021-00650-1.
44. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
45. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.
46. Holm, L., Kääriäinen, S., Rosenström, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics *24*, 2780–2781. https://doi.org/10.1093/bioinformatics/btn507.
47. Zavialov, A., Zav'yalova, G., Korpela, T., and Zav'yalov, V. (2007). FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens. FEMS Microbiol. Rev. *31*, 478–514. https://doi.org/10.1111/j.1574-6976.
48. Lillington, J., Geibel, S., and Waksman, G. (2015). Reprint of "Biogenesis and adhesion of type 1 and P pili". *Biochim. Biophys. Acta* (*BBA*) – *General Subjects 1850*, 554–564. https://doi.org/10.1016/j.bbagen.2014.07.009.
49. Hospenthal, M.K., Costa, T.R.D., and Waksman, G. (2017). A comprehensive guide to pilus biogenesis in Gram-negative bacteria. Nat. Rev. Microbiol. *15*, 365–379. https://doi.org/10.1038/nrmicro.2017.40.
50. Waksman, G. (2017). Structural and Molecular Biology of a Protein-Polymerizing Nanomachine for Pilus Biogenesis. J. Mol. Biol. *429*, 2654–2666. https://doi.org/10.1016/j.jmb.2017.05.016.
51. Laskowski, R.A. (2007). Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. Bioinformatics *23*, 1824–1827. https://doi.org/10.1093/bioinformatics/btm085.

52. Böhning, J., Ghrayeb, M., Pedebos, C., Abbas, D.K., Khalid, S., Chai, L., and Bharat, T.A.M. (2022). Donor-strand exchange drives assembly of the TasA scaffold in *Bacillus subtilis* biofilms. Nat. Commun. *13*, 7082. https://doi.org/10.1038/s41467-022-34700-z.

53. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

54. Desta, I.T., Porter, K.A., Xia, B., Kozakov, D., and Vajda, S. (2020). Performance and its limits in rigid body protein-protein docking. Structure *28*, 1071–1081.e3. https://doi.org/10.1016/j.str.2020.06.006.

55. Cohlberg, J.A. (2012). Exploring Proteins and Nucleic Acids with Jmol (W.H. Freeman and Company). https://web.csulb.edu/~cohlberg/Jmolmanual.pdf.

56. Wang, D., Liu, D., Yuchi, J., He, F., Jiang, Y., Cai, S., Li, J., and Xu, D. (2020). MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. Nucleic Acids Res. *48*, W140–W146. https://doi.org/10.1093/nar/gkaa275.

57. Jarrell, K.F., Ding, Y., Meyer, B.H., Albers, S.V., Kaminski, L., and Eichler, J. (2014). N-linked glycosylation in *Archaea*: a structural, functional, and genetic analysis. Microbiol. Mol. Biol. Rev. *78*, 304–341. https://doi.org/10.1128/mmbr.00052-13.

58. Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

59. Wang, F., Cvirkaite-Krupovic, V., Krupovic, M., and Egelman, E.H. (2022). Archaeal bundling pili of *Pyrobaculum calidifontis* reveal similarities between archaeal and bacterial biofilms. Proc. Natl. Acad. Sci. USA *119*, e2207037119. https://doi.org/10.1073/pnas.2207037119.

60. Gaines, M.C., Isupov, M.N., Sivabalasarma, S., Haque, R.U., McLaren, M., Mollat, C.L., Tripp, P., Neuhaus, A., Gold, V.A.M., Albers, S.-V., and Daum, B. (2022). Electron cryo-microscopy reveals the structure of the archaeal thread filament. Nat. Commun. *13*, 7411–7413. https://doi.org/10.1038/s41467-022-34652-4.

61. Hospenthal, M.K., Zyla, D., Costa, T.R.D., Redzej, A., Giese, C., Lillington, J., Glockshuber, R., and Waksman, G. (2017). The Cryoelectron Microscopy Structure of the Type 1 Chaperone-Usher Pilus Rod. Structure *25*, 1829–1838.e4. https://doi.org/10.1016/j.str.2017.10.004.

62. Shibata, S., Shoji, M., Okada, K., Matsunami, H., Matthews, M.M., Imada, K., Nakayama, K., and Wolf, M. (2020). Structure of polymerized type V pilin reveals assembly mechanism involving protease-mediated strand exchange. Nat. Microbiol. *5*, 830–837. https://doi.org/10.1038/s41564-020-0705-1.

63. Bolhuis, A. (2002). Protein transport in the halophilic archaeon *Halobacterium* sp. NRC-1: a major role for the twin-arginine translocation pathway? Microbiology *148*, 3335–3346. https://doi.org/10.1099/00221287-148-11-3335.

64. Liu, H., Wu, Z., Li, M., Zhang, F., Zheng, H., Han, J., Liu, J., Zhou, J., Wang, S., and Xiang, H. (2011). Complete genome sequence of *Haloarcula hispanica*, a model haloarchaeon for studying genetics, metabolism, and virus-host interaction. J. Bacteriol. *193*, 6086–6087. https://doi.org/10.1128/JB.05953-11.

65. Han, J., Lu, Q., Zhou, L., Zhou, J., and Xiang, H. (2007). Molecular characterization of the *phaECHm* genes, required for biosynthesis of poly(3-hydroxybutyrate) in the extremely halophilic archaeon *Haloarcula marismortui*. Appl. Environ. Microbiol. *73*, 6058–6065. https://doi.org/10.1128/AEM.00953-07.

66. Gerl, L., Deutzmann, R., and Sumper, M. (1989). Halobacterial flagellins are encoded by a multigene family. Identification of all five gene products. FEBS Lett. *244*, 137–140. https://doi.org/10.1016/0014-5793(89)81179-2.

67. Privalov, P.L., and Potekhin, S.A. (1986). Scanning microcalorimetry in studying temperature-induced changes in proteins. Methods Enzymol. *131*, 4–51. https://doi.org/10.1016/0076-6879(86)31033-4.

68. Tarasov, V.Y., Kostyukova, A.S., Tiktopulo, E.I., Pyatibratov, M.G., and Fedorov, O.V. (1995). Unfolding of tertiary structure of *Halobacterium halobium* flagellins does not result in flagella destruction. J. Protein Chem. *14*, 27–31. https://doi.org/10.1007/bf01902841.

69. Liu, H., Han, J., Liu, X., Zhou, J., and Xiang, H. (2011). Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. J. Genet. Genom. *38*, 261–269. https://doi.org/10.1016/j.jgg.2011.05.003.

70. Cai, S., Cai, L., Liu, H., Liu, X., Han, J., Zhou, J., and Xiang, H. (2012). Identification of the haloarchaeal phasin (PhaP) that functions in polyhydroxyalkanoate accumulation and granule formation in *Haloferax mediterranei*. Appl. Environ. Microbiol. *78*, 1946–1952. https://doi.org/10.1128/AEM.07114-11.

71. Xia, Y., Li, K., Li, J., Wang, T., Gu, L., and Xun, L. (2019). T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. Nucleic Acids Res. *47*, e15. https://doi.org/10.1093/nar/gky1169.

72. Piatibratov, M., Hou, S., Brooun, A., Yang, J., Chen, H., and Alam, M. (2000). Expression and fast-flow purification of a polyhistidine-tagged myoglobin-like aerotaxis transducer. *Biochim. Biophys. Acta* (*BBA*) – *General Subjects 1524*, 149–154. https://doi.org/10.1016/S0304-4165(00)00151-3.

73. Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., and Hutchison, G.R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. J. Cheminform. *4*, 17. https://doi.org/10.1186/1758-2946-4-17.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-HAH_0240	This paper	N/A
Anti-6X His tag antibody	Abcam	Cat # ab14923; RRID: AB_443105
Goat anti-rabbit IgG conjugated with alkaline phosphatase	Merck	Cat # A3687; RRID: AB_258103
Bacterial and virus strains		
Bacterial and archaeal strains are listed in Table S5		N/A
Chemicals, peptides, and recombinant proteins		
BamHI-HF®	NEB	R3136S
EcoRI-HF®	NEB	R3101S
KpnI-HF®	NEB	R3142S
NdeI	NEB	R0111S
PstI	NEB	R0140S
Q5® High-Fidelity DNA Polymerase	NEB	M0491S
SphI	NEB	R0182S
DNase I	Thermo scientific	EN0521
BCIP/NBT Color Development Substrate	Promega	S3771
IPTG	Thermo scientific	R0392
L-Tryptophan	Merck	T0254
Bacto™ Casamino Acids	Thermo scientific	223050
Brilliant Blue G	Merck	B0770
Schiff's fuchsin-sulfite reagent	Merck	S5133
DNA Ladder 1 kb	Evrogen	NL001
PageRuler™ Prestained Protein Ladder, 10 to 180 kDa	Thermo scientific	26616
5-Fluoroorotic Acid (5-FOA)	Thermo scientific	R0811
Uracil	Merck	U0750
Proteinase K	AppliChem	A4392
Trypsin	Merck	T4799
Polyethylene glycol (PEG) 6000	Panreac	203-473-3
Critical commercial assays		
GeneJET Plasmid Miniprep Kit	Thermo scientific	K0502
QIAquick PCR Purification Kit	Qiagen	28104
Amicon® Ultra Centrifugal Filter, 3 kDa MWCO	Merck	UFC9003
Oligonucleotides		
See Table S6 for cloning primers	This work	N/A
Recombinant DNA		
See Table S5	This work	N/A
Software and algorithms		
ImageJ	National Institutes of Health	N/A
Software used for protein structure prediction	AlphaFold2	https://alphafold.ebi.ac.uk
Software used for protein structure alignment	Dali server	http://ekhidna2.biocenter.helsinki.fi/dali/
Other		
Immobilon®-P^{SQ} PVDF Membrane	Merck	ISEQ00010
Poros MS/M 10/100 column	Applied Biosystems	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The archaeal and bacterial strains and plasmids used in this study are described in Table S5.

Haloarchaea

The haloarchaeon *Haloarcula hispanica* is a model organism for researching haloarchaea, because it is easy to grow and genetically manipulable.[64] *Haloarcula hispanica* B-1755 strain (DSM 4426, ATCC 33960, CGMCC 1.2049) was obtained from the All-Russian Collection of Microorganisms, (VKM) Pushchino, and we have designated it as the wild type (WT) strain. For genetic manipulations, the auxotrophic *Har. hispanica* DF60 strain, which lacks the *pyrF* gene encoding for orotidine 5′-phosphate decarboxylase and therefore cannot produce uracil, was used. *Har. hispanica* DF60 strain was constructed from WT by deleting the *pyrF* gene, as described in.[65] *Har. hispanica* strains were cultivated at 37°C or 42°C in liquid or agar (1.0% for solid and 0.30% for semi-solid) media. We used a AS-168 complex medium[65] of the following composition (per 1L): 5 g Bacto Casamino acids, 5 g yeast extract, 1 g sodium glutamate, 3 g trisodium citrate, 20 g $MgSO_4 \cdot 7H_2O$, 2 g KCl, 200 g NaCl, 50 mg $FeSO_4 \cdot 7H_2O$, 0.36 mg $MnCl_2 \cdot 4H_2O$, pH 7.0 or AS-168SY medium[65] (AS-168 without yeast extract). When required, AS-168SY medium was supplemented with uracil at a concentration of 50 μg/mL, and 5- fluoroorotic acid (FOA) at 150 μg/mL. *Haloferax volcanii* strains MT78[38] and RE26[42] were kindly provided by M. Pohlschröder. All *Hfx. volcanii* transformed strains were grown at 37°C in liquid or agar (1.0% for solid and 0.24% for semi-solid) medium (per 1L: 5 g Bacto Casamino acids, 200 g NaCl, 18 g $MgSO_4$, 4 g KCl, 0.33 g $CaCl_2$, 0.2 g $MnCl_2 \cdot 4H_2O$, pH 7.2).[39] Heterologous *tafA*-gene expression was induced by the addition of tryptophan (0.2–1 mg/mL). Other *taf* genes were expressed under their natural promoters.

Bacteria

Escherichia coli XL1-Blue was grown in LB medium (1% bactotryptone, 0.5% yeast extract, and 1% NaCl) containing 100 μg/mL of ampicillin added. *E. coli* strains XL1-Blue, BL-Rosetta, BL-21 and C41, which were used to express the C-His-tagged TafA protein, were grown in either LB medium or ZYP5052 (50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.5% glycine, 0 .05% glucose, 0.2% lactose, 0.5% yeast extract, and 1% tryptone) with the addition of 100 mg/mL ampicillin and chloramphenicol.

Antibody preparation

This study used specific antibodies produced by immunizing rabbits with a purified target protein, TafA, using a standard immunization procedure. The care and treatment of animals were carried out in accordance with ***Guide for the Care and Use of Laboratory Animals, Eight Edition***, as well as the European Convention on the Protection of Vertebrate Animals Used for Experimental and Scientific Purposes and the laws of the Russian Federation.

METHOD DETAILS

Isolation of tafi and their separation from archaella

Normally, we use polyethylene glycol (PEG 6000) precipitation to isolate archaellar filaments, as described in the literature.[66] However, we have found that *Har*. *hispanica* archaella and tafi are co-isolated using this method. Therefore, we have employed differential centrifugation to separate these two types of filaments. At first, we pelleted the cells by low-speed centrifugation at 8000 x g using a Beckman J2-HS centrifuge with a JA-14 fixed-angle rotor. This resulted in a supernatant containing released archaella and tafi. We then precipitated archaella from this supernatant using a Beckman Coulter ultra-centrifuge (Optima XE-90) with a fixed-angle titanium rotor (Type 45 Ti, Beckman Coulter), at 142,400 x g. The remaining tafi was either precipitated after many hours of high-speed centrifugation or within one hour at 46,500 x g after adding 3% PEG 6000. This yielded a relatively pure preparation of tafi with a yield of approximately 5 mg from one liter of liquid culture. Precipitated archaella and tafi were dissolved in 10 mM Tris–HCl buffer, pH 8.0, containing 20% NaCl. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze the protein preparations. Before applying the protein samples to the gel, they were mixed with the SDS-PAGE buffer and incubated for 2 minutes at 90°C. The proteins were then stained with Coomassie Brilliant Blue G-250 or Schiff's fuchsin-sulfite reagent following to the manufacturer's instructions. ImageJ software (NIH) was used to scan the stained acrylamide gels and calculate the TafE/TafA and TafC/TafA molar ratios based on the measured intensities of the corresponding bands and the known molecular masses.

A similar procedure using PEG precipitation was used to isolate the recombinant tafi from *Hfx. volcanii* MT78 and RE26 cells. Although the MT78 strain does not produce its own archaella or pili, a preliminary high-speed centrifugation step was helpful to remove debris from the cell membranes and other contaminants.

In addition, we examined the presence of TafA protein expressed in *Hfx. volcanii* MT78 and RE26 cells in culture media after the precipitation of cells and polymeric structures. The cells were pelleted by low-speed centrifugation at 8000 g using Beckman J2-HS and the supernatant was further centrifuged twice at 142,400 x g using the the Beckman Coulter ultracentrifuge (Optima XE-90) to remove any contaminating materials. The concentrated solution, obtained by using an Amicon Ultra Centrifugal Filter (Ultracel - 3K), was then concentrated further to approximately 500 μL (100-fold) using the same method as above. The concentrated supernatant

was then centrifuged for 90 minutes at 265,100 x g to precipitate any remaining polymeric structures using a Beckman Coulter ultracentrifuge TLA-100. Detection of TafA in the resulting supernatant was performed by SDS-PAGE, immunoblotting with specific antibodies and mass spectrometry.

Mass spectrometry analysis

Proteins were excised from the gel and treated with proteinase K and trypsin at 37°C in a Thermo Mixer (Eppendorf, Germany). To stabilize the proteinase K, $CaCl_2$ was added to the solution to a final concentration of 5 mM. The molar ratio of enzyme to protein was 1:50. The reaction was terminated by adding trifluoroacetic acid to the mixture. Prior to mass spectrometry analysis, the peptides were separated using reversed-phase high performance liquid chromatography on an Easy-nLC 1000 Nanoliquid chromatography system (Thermo Fisher Scientific). The separation was carried out using a homemade column that was 25 cm in length and had a diameter of 100 μm. The column was packed with a C18 adsorbent, which had particles with a size of 3 μm and pores with a diameter of 300 Å. The column was packed under laboratory conditions with a pressure of 500 atm. The peptides were eluted using a gradient of acetonitrile, ranging from 3% to 40%, for 180 min. The mobile phase flow rate was set at 0.3 μL/min. Mass spectra of the samples were acquired using an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Germany). Peptides were ionized by electrospray at nanoliter flow rates with a 2 kV ion spray voltage. Ion fragmentation was induced by collisions with helium in the high-energy cell. Mass spectra were processed, and peptides were identified using the Thermo Xcalibur Qual Browser and PEAKS Studio software versions 7.5, based on UniRef-100 sequences.

Electron microscopy

Archaella and tafi samples were prepared for transmission electron microscopy (TEM) by negative staining with 2% uranyl acetate on formvar-coated copper grids. A grid was placed on a 20 μL drop of archaella or tafi solution (about 0.01-0.05 mg/mL in 20% NaCl, 10 mM TrisHCl, pH 8.0), for 2 minutes. Then, it was blotted with filter paper and placed on top of another drop of 2% uranyl acetate for an additional 1 minute. The excess stain was removed by wiping the grid with filter paper, and the grid was allowed to dry. The samples were examined using a JEOL JEM-1400 transmission electron microscope (JEOL, Japan), operated at 120 kV. Digital images were captured using a high-resolution water-cooled bottom-mounted CCD camera.

TafA (HAH_0240) purification

Additional purification of the TafA protein was performed to remove traces of archaella, membrane proteins, and other impurities. This was done by depolymerizing the tafi and then using size-exclusion chromatography a HiLoad 26/60 Superose 12 gel-filtration column (Pharmacia, Sweden). After precipitation with PEG, the tafi preparation was diluted in 50 mM Tris-HCl buffer at pH 8.0 with 0-5% NaCl. The mixture was incubated at 90°C for 15 minutes. The preparation was then cooled and centrifuged at 213,400 x g for 30 minutes using a Beckman TLA-100 ultracentrifuge in order to remove any remaining high-molecular weight impurities. The supernatant was applied to a Superose 12 column that was equilibrated with a 10 mM Na-phosphate buffer at different NaCl concentrations. The elution rate was 0.2 mL/min throughout the experiment. Protein molecular masses were estimated from elution volumes using bovine serum albumin (BSA, 66 kDa, Sigma-Aldrich) and ovalbumin (OVA, 45 kDa, Sigma-Aldrich) as standard proteins. The presence of proteins in the eluate was confirmed using spectrometry and SDS electrophoresis. Most contaminant proteins, including archaellins, eluted behind the column's void volume, whereas TafA eluted with a delay.

The study of the secondary structure of the TafA protein was carried out using circular dichroism (CD) spectroscopy on a Chirascan spectropolarimeter (Applied Photophysics) in the wavelength range from 190 nm to 250 nm. The values of molar ellipticity were calculated following the equation: $[\theta] = [\theta]_m M_{res}/(LC)$, in which C is the protein concentration (mg/mL), L is the optical path length of the cuvette (mm), $[\theta]_m$ is the measured ellipticity (degrees) and M_{res} is the average molecular mass of the peptide residue (Da) calculated from its amino acid sequence. Measurements were carried out in a 0.1 mm cuvette.

Calorimetric measurements were conducted on a SKAL-1 differential scanning microcalorimeter (ZAO SKAL, Pushchino, Russia), at a heating rate of 1°C per minute. The working volume of the gold cell was 0.3 mL, and the protein concentration in the samples used in the experiments ranged from 1 to 0.5 mg/mL. For the experiments, 10 mM Na-phosphate buffer, at pH 8.0, containing 25% NaCl, was used. Measurements were carried out according to the previously described methods.[67,68]

Construction and confirmation of the deletion mutants

Deletions of the *tafA*, *tafD*, and *pibD* genes were generated in the *Har. hispanica* DF60 strain using the homologous recombination method (pop-in/pop-out), as previously described.[69] The sequences of the PCR primers used in this study are summarized in Table S6. To obtain the target strains, DNA fragment pairs were prepared using PCR with the following primers: pr1/pr2 and pr3/pr4 for AG1 (Δ*tafA*), pr5/pr6 and pr7/pr8 for AG2 (Δ*tafD*), and pr9/pr10 and pr11/pr12 for AG3 (Δ*pibD*). For AG1 and AG2, the amplified DNA fragments (upstream and downstream flanking sequences for each gene) were treated with corresponding restriction enzymes, and then mixed with the pHAR plasmid that had also been treated with the same enzymes. After ligation and transformation, *E. coli* XL1-Blue competent cells were used to screen for the presence of pHARΔ*tafA* and pHARΔ*tafD* using PCR. Colonies containing these genes were selected. For AG3, two DNA fragments were amplified using overlap PCR with primers pr9 and pr12. These fragments were then joined together to create a DNA fragment containing the necessary restriction sites (Table S6). The resulting fragment was inserted into the pHAR plasmid, creating the pHARΔ*pibD* construct. Out of all the competent cells, only those with the pHARΔ*pibD*

plasmid were selected for further study. The next step was to transform *Har. hispanica* DF60 with plasmids pHARΔ*tafA*, pHARΔ*tafD*, and pHARΔ*pibD*. The cells were then plated on AS-168SY solid medium, a nutrient-poor medium without yeast extract. Transformants were selected based on the integration of the knockout plasmid into the corresponding gene using PCR analysis. To verify the deletion of the *tafA*, *tafD*, and *pibD* genes, we used external (pr1, pr2, pr5, pr8, pr9, and pr12) and internal (pr13-pr18) primers. These primers annealed to regions surrounding and within each gene, respectively, to ensure accurate detection of the deletion. Cells containing the required pHARΔ*tafA*, pHARΔ*tafD*, or pHARΔ*pibD* plasmids integrated into their genomes were subcultured on AS-168 medium supplemented with yeast extract and 50 mg/mL uracil and 150 mg/mL 5-FOA to select for the double-crossover recombinant cells. The PCR products were then analyzed using agarose gel electrophoresis.

Plasmid-based complementation for *tafA*, *tafD*, and *pibD* gene deletions

In order to investigate the significance of gene deletions in *tafA*, *tafD*, and *pibD*, we obtained plasmids that could be used to complement these genes in deletion strains. These plasmids were derived from the pWL502 vector,[70] which contains an ampicillin resistance gene, a *pyrF* gene, and regions necessary for replication in both *E. coli* and *Har. hispanica* cells. Based on this, we created three plasmids: pAS102, pAS103, and pAS104, each containing the target genes *tafA*, *tafD*, and *pibD* respectively. In addition to the target genes, each plasmid also contained 200 base pairs on either side of the gene. These regions contain native promoters and transcription terminators. These inserts were created using PCR with primers pr25-pr30, and the plasmids were assembled using the TEDA method.[71] After creating the plasmids, *E. coli* cells were transformed with them, and individual colonies were obtained. The colonies were then screened for the presence of the inserts using PCR and the results were visualized using agarose gel electrophoresis. Based on the screening results, the colonies that contained plasmids with the expected insert size were identified. The sequence of the inserts in these plasmids was confirmed by sequencing performed by Evrogen (Russia). These resulting plasmids were then used to transform *Har. hispanica* strains that lacked the corresponding genes. The presence of plasmids in strains was confirmed using PCR with appropriate primers, as shown in Figure S10.

Heterologous expression of *Har. hispanica tafA-tafG* genes in *Haloferax volcanii*

For heterologous expression, we used a special immotile strain of *Hfx. volcanii* MT78[40] (Δ*arlA1*Δ*arlA2* Δ*pilB3-C3*), in which the genes encoding archaellins (*arlA1* and *arlA2*) and those critical for the assembly of type IV pili (*pilB3* and *pilC3*), have been deleted. In some experiments, we have also used the motile strain of *Hfx. volcanii* RE26,[42] in which only genes required for the assembly of type IV pili, *pilB3* and *pilC3*, have been inactivated, but archaellin genes have been retained. For inducible expression, we have used vectors based on pTA1228,[41] with tryptophan inducible promoter.

To generate the expression plasmid pAS24, we used PCR to create a DNA fragment containing the *tafA, tafB, tafC* and *tafD* genes using primers pr19 and pr20. We then cut the original pTA1228 vector with EcoRI and NdeI restriction enzymes and inserted the PCR fragment into it using the TEDA cloning method.[71] To create the pAS30 plasmid, we first cut the pAS24 vector at the NotI site and inserted a PCR fragment containing the *tafG, tafF*, and *tafE* genes using TEDA. For pAS31, we cut the pAS30 vector at NdeI and KasI sites and inserted another PCR product generated with primers pr23 and pr24. We confirmed the correct assembly of all plasmids using Sanger sequencing.

Preparation of recombinant TafA (HAH_0240) by heterologous expression in *E. coli*

To obtain highly specific antibodies for TafA, we used a recombinant protein produced through heterologous expression in *E. coli*. Antibodies generated against the isolated *Har. hispanica* tafi were not specific enough to detect the TafA protein in cell lysates. Therefore, we hypothesized that using an unmodified (non-glycosylated) protein for immunization would increase the specificity of the antibody response. The truncated *tafA* gene, without the sequence encoding the N-terminal signal peptide, was cloned into the pET22b vector from the pAS102 plasmid that we had previously constructed. The final construct included a region encoding six C-terminal histidine residues. A genetic construct, pMT1 (pET22b-*tafA-His6*), was assembled using primers pr31 and pr32 and then transformed into *E. coli* cells in order to express the C-His-tagged TafA protein. The sequence of the insert in plasmid pET22b was confirmed by sequencing performed by Evrogen (Russia) using standard primers T7uni and T7rev. The synthesis of the target protein in *E. coli* was confirmed by using an anti-6xhis tag antibody (Abcam), and the highest expression was observed in the BL-Rosetta strain grown in LB medium at a temperature between 28°C and 30°C. When analyzing the localization of the recombinant TafA protein in *E. coli* cells, we found that it was present in a soluble form.

Before purification of the target protein, *E. coli* cells were disrupted using ultrasound and the cell extract was clarified by low-speed centrifugation. The recombinant TafA protein was purified using nickel affinity chromatography on a Poros MS/M 10/100 column, following a previously described procedure.[72] A 10 mM sodium phosphate buffer pH 8.0 containing 0.2 M NaCl was used for cell extraction and chromatography. The same buffer containing 250 mM imidazole was used for gradient elution. The peak fractions containing recombinant TafA protein were concentrated using a Centricon 10 device and stored at +4°C.

Prior to further purification, the peak fractions were concentrated to approximately 1 mL and then loaded onto a HiLoad 26/60 Superose 12 gel-filtration column (Pharmacia, Sweden), which was connected to an FPLC system (Pharmacia). The column was equilibrated with the appropriate buffer. It was then eluted with the same buffer at a flow rate of 0.2 mL per minute. The peak fractions were then collected and confirmed using SDS-PAGE (15%), with Coomassie blue G 250 staining.

Preparation of rabbit anti-TafA antibody

The recombinant protein, TafA, in a concentration of 1.5 mg/mL, was used to hyperimmunize the rabbit. For the first injection, 0.5 mL of protein solution and 0.5 mL of complete Freund's adjuvant (CFA) were thoroughly mixed, and 0.5 mL was injected locally into both popliteal lymph nodes (PLN). After 14 days, 1.0 mL of protein solution was mixed with 1.0 mL of incomplete Freund's adjuvant (IFA), and 1.0 mL was injected back into the PLN. On the 28th day following the first immunization, 1.5 mL of the protein solution combined again with 1.5 mL of IFA and injected back into PLN in addition to 10 further injections of 0.1 mL each, intradermally along the spine. On the 8th day after the third immunization (36 days), blood (30 mL) was taken from the marginal ear vein of a rabbit. The blood was incubated at +37°C for 1 hour and at +4°C overnight. After the formation of a clot, the antiserum was discarded and the blood cells were removed by centrifugation at 5,000 g, +4°C for 20 minutes (using a centrifuge, model 5810 R, Eppendorf, Germany). The resulting supernatant was then mixed with phosphate-buffered saline (PBS) at pH 7.2 in a 1:1 ratio. The immunoglobulin fraction was then precipitated using a semi-saturated ammonium sulfate solution at +4°C in an ice bath, followed by centrifugation at 10,000 g for 10 minutes at +4°C. The precipitated immunoglobulins were then resuspended in a saturated ammonium sulfate solution and stored at −20°C. Before use, the antibodies were dialyzed in 2,000 times their volume of PBS to remove excess salts.

Western blotting

Samples for the analysis of the TafA protein in the cell lysates of *Har. hispanica* and *Hfx. volcanii* were prepared as follows: 300 μL of cell suspension with an optical density of 2 at 600 nm was pelleted. The cell pellet was then resuspended in 24 μL of an aqueous solution containing DNase 1. After that, 8 μL of a 4-fold buffer for SDS-PAGE was added and the mixture was heated at 90°C for 3 minutes. The proteins were separated using SDS-PAGE and then transferred from the gel to an Immobilon-P^{SQ} membrane using the Trans-Blot Turbo Transfer System (from Bio-Rad, USA) as described in the user manual. Polyclonal antibodies against TafA, diluted 1:500, were used to detect the presence of the TafA protein in the samples. Antibody binding was detected by incubating the membrane with goat anti-rabbit IgG, which was conjugated to alkaline phosphatase, followed by staining with nitroblue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP).

Structural prediction

The structures of the TafA-TafG protein precursors are known from the database https://alphafold.ebi.ac.uk/. To study the structure of mature proteins and their complexes, we used the AlphaFold 2,[43,44] ColabFold,[45] AlphaFold 3,[53] and ClusPro protein–protein docking[54] software tools. The Dali[46] server was used to align 3D protein structures. A comprehensive analysis of the obtained 3D protein models and protein-protein interactions was performed using the PDBsum Generate[51] (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and PDBePISA[58] (https://www.ebi.ac.uk/pdbe/pisa) programs. To visualize the protein models, the Avogadro[73] (http://avogadro.cc/) and Jmol[55] (https://jmol.sourceforge.net/) programs were used. The MusiteDeep[56] server (https://www.musite.net/) was used to predict the carbohydrate binding sites of N-glycosylated TafA protein.

QUANTIFICATION AND STATISTICAL ANALYSIS

In the mass spectrometry analysis, peptides were identified using the Thermo Xcalibur Qual Browser and PEAKS Studio software versions 7.5, based on UniRef-100 sequences. A parent mass error tolerance of 2.0 ppm and a fragment mass error tolerance of 0.1 Da were applied. Only peptides with a threshold value T = -10 lgP higher than 15 were considered which corresponds to a p-value less than 0.03. When comparing the AlphaFold-predicted TafA-TafG protein structures with each other and with other proteins using the Dali server, we took into account only results with Z-scores greater than 4. According to the reference[46] "significant similarities" have a Z-score above 2 and usually correspond to similar folds.