nature communications

Article

https://doi.org/10.1038/s41467-024-53923-w

Structural diversity and clustering of bacterial flagellar outer domains

Received: 2 March 2024

Accepted: 28 October 2024

Published online: 03 November 2024

Check for updates

Jessie Lynda Fields[1,5], Hua Zhang[2,5], Nathan F. Bellis[1,5], Holly A. Petersen[1], Sajal K. Halder[1], Shane T. Rich-New[1], Mart Krupovic[3], Hui Wu[2] ✉ & Fengbin Wang[1,4] ✉

Supercoiled flagellar filaments function as mechanical propellers within the bacterial flagellum complex, playing a crucial role in motility. Flagellin, the building block of the filament, features a conserved inner D0/D1 core domain across different bacterial species. In contrast, approximately half of the flagellins possess additional, highly divergent outer domain(s), suggesting varied functional potential. In this study, we report atomic structures of flagellar filaments from three distinct bacterial species: *Cupriavidus gilardii*, *Stenotrophomonas maltophilia*, and *Geovibrio thiophilus*. Our findings reveal that the flagella from the facultative anaerobic *G. thiophilus* possesses a significantly more negatively charged surface, potentially enabling adhesion to positively charged minerals. Furthermore, we analyze all AlphaFold predicted structures for annotated bacterial flagellins, categorizing the flagellin outer domains into 682 structural clusters. This classification provides insights into the prevalence and experimental verification of these outer domains. Remarkably, two of the flagellar structures reported herein belong to a distinct cluster, indicating additional opportunities on the study of the functional diversity of flagellar outer domains. Our findings underscore the complexity of bacterial flagellins and open up possibilities for future studies into their varied roles beyond motility.

More than 80% of all known bacterial species are motile at some stage of their life cycle[1]. These bacteria navigate toward nutrients or away from unfavorable environments through chemotaxis[2–4], using movements such as swimming, swarming, and twitching, among others[5]. The flagellum, extracellular machinery, facilitates swimming and swarming movements and comprises three main components: the basal body complex, functioning as the motor; the connecting rod and hook, acting as a universal joint; and the flagellar filaments, serving as a mechanical propeller[6,7]. Canonical flagellar filaments are long, supercoiled structures composed of around 20,000 flagellins. As the filament rotates, it generates thrust, acting similarly to an Archimedean screw. To date, all known bacterial flagellins share a conserved D0/D1 domain architecture, which is uniformly arranged across various bacterial species. These domains exhibit a helical rise of ~5 Å and a twist of 65.4°, aligning every 11 flagellins near vertically. This configuration presents distinct 11-start protofilaments on the filament's surface and its inner core. Recent research has revealed the molecular basis underlying flagellar supercoiling, identifying 11 different flagellin D0/D1 states that result in 11 unique protofilament conformations. Interestingly, similar supercoiling mechanisms have also been documented in archaeal flagella, despite the structural components of archaeal and bacterial flagella not sharing homology[8].

[1]Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL 35233, USA. [2]Department of Oral Rehabilitation & Biosciences, Oregon Health & Science University, Portland, OR 97239, USA. [3]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, Paris 75015, France. [4]Gregory Fleming James Cystic Fibrosis Research Center, University of Alabama at Birmingham, Birmingham, AL 35233, USA. [5]These authors contributed equally: Jessie Lynda Fields, Hua Zhang, Nathan F. Bellis. ✉e-mail: wuhu@ohsu.edu; jerrywang@uab.edu

For bacteria that colonize or infect other organisms, motility plays a crucial role in interactions between a bacterium and its host[5]. Beyond enabling movement, flagella are thought to possess several other functions, including adhesion to surfaces[9,10], colonization[11], biofilm formation[12], and potentially modulating host immune response[13], among others. Many of these properties are attributed to the flagellar outer domains, which are part of the central region of the flagellin, named D2–D4, and so forth. In several bacterial species, these outer domains are not essential for motility[14,15]. In fact, about half of the bacterial flagellins annotated in the UniProt database contain only the D0/D1 domain, lacking outer domains.

Several outer domain structures from a diverse range of species[10,16–26] that include soil-borne bacteria such as *Sinorhizobium meliloti*, opportunistic human pathobionts such as *Salmonella enterica* and *Pseudomonas aeruginosa*, as well as life-threatening primary human pathogens like *Burkholderia pseudomallei*, a high-priority biological agent responsible for melioidosis[27], have been reported in the previous studies. These structures encompass flagellin structures solved by X-ray crystallography and filamentous structures determined by cryo-electron microscopy (cryo-EM). Remarkably, the fold and architecture of the outer domains are highly variable. The lack of clear homology among the outer domains of flagellin across different species has historically made it challenging in the structural analysis of these domains. Recent advances in protein structure prediction methods, notably AlphaFold[28], have shown great promise in accurately predicting protein structures at the fold level, even when no analogous structure exists. With most UniProt sequences now predicted and available in the AlphaFold database[29], it is feasible to undertake large-scale structural analyses of flagellin outer domains.

In this study, we report the near-atomic resolution cryo-EM structures of supercoiled bacterial flagellar filaments from three diverse bacterial species. The first is from *Cupriavidus gilardii*, a Gram-negative, aerobic, opportunistic pathogen that has been increasingly associated with human infection and holds potential for bioremediation[30,31]. The second is from *Stenotrophomonas maltophilia*, a Gram-negative, aerobic bacterium known for its multidrug resistance and its ability to infect the lungs of individuals with cystic fibrosis[32,33]. The third is from *Geovibrio thiophilus*, a Gram-negative, non-sporulating bacterium residing and thriving in water sediments under anaerobic and microaerophilic conditions and that reduces sulfur and nitrate[34]. We discovered that the flagellar surface of *G. thiophilus* carries a significantly higher negative charge, suggesting a potential mechanism for its adhesion to positively charged minerals. Moreover, we observed remarkable diversity in the outer domains of the three elucidated structures in both size and fold. The only similarity was that the D2 domain of *S. maltophilia* and the D3 domain of *G. thiophilus* both exhibit an immunoglobulin-like (Ig-like) fold topology[35].

This discovery prompted us to conduct a comprehensive review of annotated bacterial flagellins using AlphaFold predications, classifying the predicted flagellin outer domains into 682 structural clusters. Our results indicated that nearly half of the flagellin sequences with outer domains contain at least one Ig-like domain, suggesting that the Ig-like domain is ubiquitous in flagellin proteins. Additionally, our analysis provided a profile of the most frequently observed outer domains and determined whether their structures have been experimentally resolved. Ranked by the abundance of outer domain architectures, prior cryo-EM studies have documented outer domain structures in clusters #2, 3, 5, 20, and 28, while X-ray studies have explored outer domains in clusters #1, 2, 5, 6, 10, and 17. The structures we reported belong to clusters #1, 27, and 41, addressing a significant gap in knowledge and highlighting the existence of numerous other fascinating clusters yet to be discovered in structural studies.

Results

Cryo-EM structures of three flagellar filaments with outer domains

Cryo-EM was utilized to determine the structures of three flagellar filaments: one peritrichous from *C. gilardii*[36], one lophotrichous from *S. maltophilia*[32], and one monotrichous from *G. thiophilus*[34]. The established helical symmetry of canonical flagellar D0/D1 domains, with a helical rise of ~5.0 Å and a twist of ~65.45°, was confirmed by examining the power spectra. When this symmetry was applied to 320-pixel box particles, it led to high-resolution 3D reconstructions: *C. gilardii* D0/D1 domains at 3.1 Å, *S. maltophilia* D0–D2 domains at 3.2 Å, and *G. thiophilus* D0–D3 domains at 3.4 Å. This indicates that the D2 domain of *C. gilardii* and D4/D5 domains of *G. thiophilus* diverge from canonical flagellar symmetry, featuring additional interfaces within the outer domain region (Fig. 1). In order to generate 3D reconstructions for supercoiled flagellar filament, particles were re-extracted using a larger 640-pixel box, followed by subsequent inspection of power spectra to identify a layer-line with a meridional intensity that does not exist in the canonical flagellar symmetry. A weak meridional layer-line near 1/(110 Å) was observed in *G. thiophilus* flagella, suggesting a repeating feature every 22 flagellin molecules, similar to the flagellar filament in *P. aeruginosa* PAO1[16,18]. Additionally, previously described intermediate layer-lines[18,37–39] were seen in both power spectra of *C. gilardii* and *G. thiophilus* flagella, indicating non-helical perturbations were present in the structure. As described in Supplementary Fig. 1, the reconstruction of the supercoiled filament was executed using helical refinement, applying the respective outer domain symmetry and post-processing with homogeneous and local refinements to relax the helical constraints. The 3D reconstructions of the supercoiled filaments for *C. gilardii*, *S. maltophilia*, and *G. thiophilus* achieved final, near-atomic resolutions of 3.4, 3.3, and 4.1 Å, respectively, as determined by map:map FSC (Supplementary Fig. 2).

The diameters of *C. gilardii* and *S. maltophilia* flagellar filaments were similar, ~165–170 Å, in contrast to the notably wider *G. thiophilus* filament, which measured around 240 Å (Fig. 1A–C). All three flagellar filaments exhibited a similar D0/D1 domain architecture, which is expected since all bacterial flagella share this homology. Notably, both *C. gilardii* and *S. maltophilia* have a single D2 domain. In *C. gilardii*, the D2 domain dimerizes, creating a screw-like feature on its surface (Fig. 1D). Conversely, in *S. maltophilia*, the D2 domain does not form an extra interface with adjacent domains, maintaining the same helical symmetry as the D0/D1 domain. The outer domains D2–D5 in *G. thiophilus* are the largest flagellin outer domains determined by cryo-EM (Fig. 1F, I), responsible for the diameter increase to 240 Å. Interestingly, three flagellin sequences are present in the *G. thiophilus* genome, quite similar in size and overall folds predicted by AlphaFold. Upon full-length modeling of all three flagellin sequences, the correct flagellin protein sequence was identified, as particular regions in the reconstruction were found that could not be explained by the other two sequences.

Outer domains in three flagellins share β-strand rich architecture

As expected, the D0/D1 domains among the three flagella filaments share an identical fold, with sequence identities ranging from 42% to 59%. However, the fold of their outer domains significantly diverges (Fig. 1G–I). Remarkably, the fold of the *C. gilardii* D2 domain (Fig. 2A) was not observed in known, experimentally determined structures, as indicated by Foldseek[40] and the DALI[41] server. Further investigation into sequence-level and AlphaFold prediction similarities revealed that similar outer domains are present in certain species of the *Betaproteobacteria* class, including those in the *Burkholderiales* order, and in species of the *Gammaproteobacteria* class, including those in the orders *Enterobacterales*, *Oceanospirillales*, and *Pseudomonadales*, many of which are opportunistic human and plant pathogens. It is

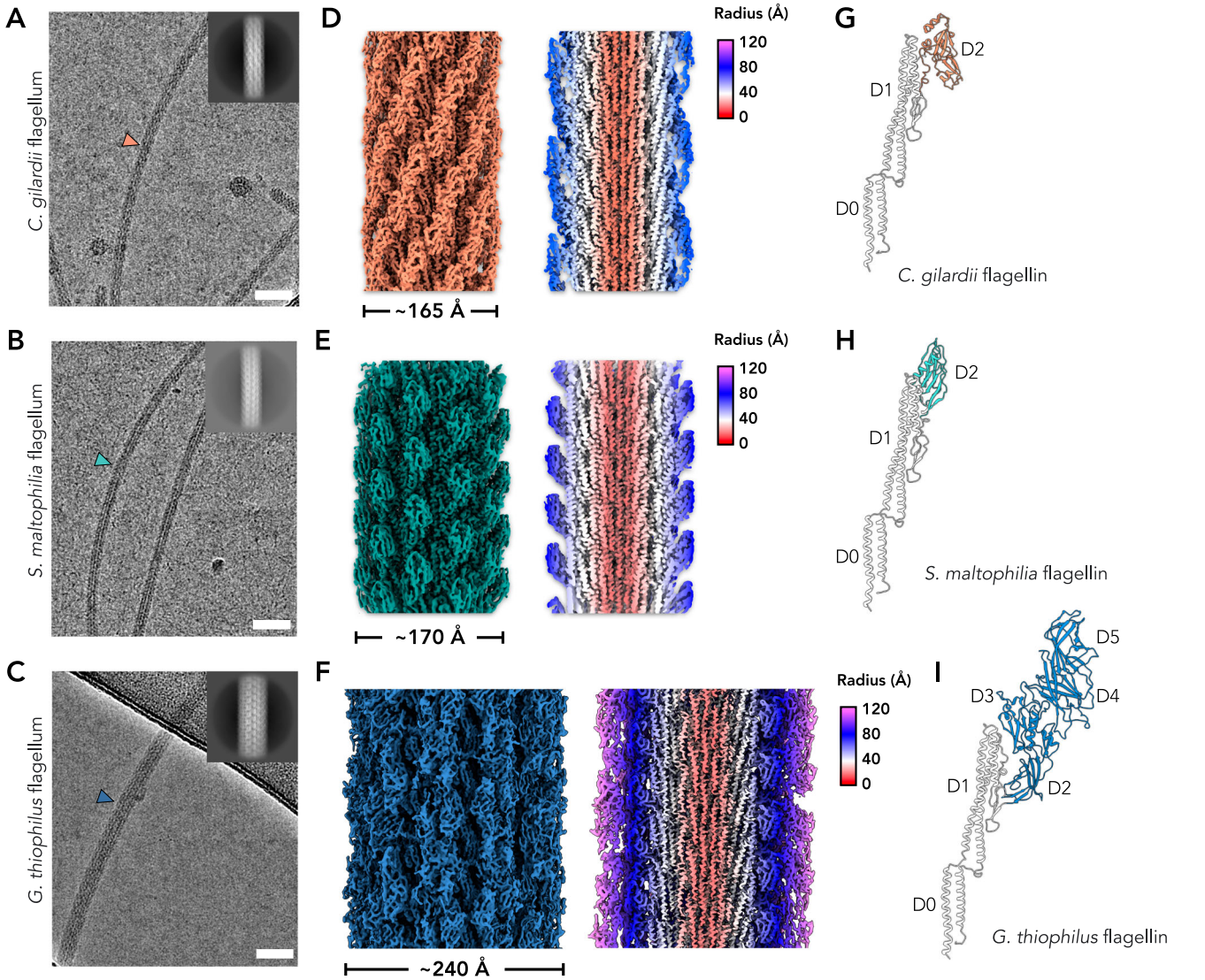

Fig. 1 | Cryo-EM of flagellar filaments from *C. gilardii*, *S. maltophilia*, and *G. thiophilus*. Representative cryo-electron micrographs of the *C. gilardii* flagella (**A**), *S. maltophilia* flagella (**B**), and *G. thiophilus* flagella (**C**). Scale bar, 50 nm in (**A**)–(**C**). Orange arrowhead points to the *C. gilardii* flagellum; green arrowheads point to the *S. maltophilia* flagellum; blue arrowhead points to the *G. thiophilus* flagellum. 2D averages of all three flagella are shown in the top right corner. Cryo-EM reconstructions of the *C. gilardii* flagellum at 3.4 Å resolution (**D**), the *S. maltophilia* flagellum at 3.3 Å resolution (**E**), and the *G. thiophilus* flagellum at 4.1 Å resolution (**F**). Thin sections parallel to the helical axis of the flagella are shown on the right of the 3D reconstruction, colored by the radius. **G**–**I** The single flagellin structures modeled into the cryo-EM densities are displayed on the right, with their respective domain names indicated. The conserved D0/D1 domains are shown in gray. The outer domains of *C. gilardii*, *S. maltophilia*, and *G. thiophilus* are consistently colored in accordance with the 3D reconstructions.

probable that the species belonging to those groups have been sparsely sampled in the past for structural studies. The *S. maltophilia* D2 domain exhibits an Ig-like fold (Fig. 2B), a common two-layer β-sandwich domain found in numerous proteins with diverse functions[35], including the globular domain in archaeal type IV pili[42–44]. Unsurprisingly and likely due to the ubiquitous nature of the Ig-like domain, similar outer domains were found in many AlphaFold-predicted flagellins across a diverse array of bacterial species, from anaerobes to aerobes and marine algae to human microbiota. A similar D2 domain was also experimentally captured in a partial flagellin crystal structure from *Sphingomonas sp.* A1[19], sharing approximately 28% sequence identity between the D2 domains. The *G. thiophilus* flagellum, notably the first flagellar structure from bacteria growing anaerobically, presents a large outer domain architecture with four domains D2–D5 (Fig. 2C). All four outer domains are rich in β-strands: the fold of D2 domain is mainly found in flagellar hook-associated proteins; the D3 domain is a variant of the Ig-like fold; D4 and D5 are β-sandwich domains similar to the Pfam[45] DUF992 (domain of unknown function). Similar flagellar outer domains are mostly identified in bacteria from environments like lake or ocean sediment, soil, and wastewater, across the classes *Chrysiogenetes*, *Clostridia*, *Deferribacteres*, and *Synergistia*.

Dimeric outer domain interactions in *C. gilardii* and *G. thiophilus* flagella

Next, we explored how the outer domains interact along the flagellar filaments. Interestingly, the outer domain of *S. maltophilia* maintains the same D0/D1 symmetry and does not form additional polymeric contacts with adjacent subunits. In contrast, the *C. gilardii* and *G. thiophilus* flagella display partial or entire outer domain dimerization, thereby disrupting the D0/D1 symmetry. The dimeric interface happens at the radius of ~80 Å in *C. gilardii* and ~120 Å in *G. thiophilus* (Fig. 3A). Given the canonical D0/D1 symmetry, adjacent flagellins along the 11-start have minimal twist (-65.45° multiplied by 11), rendering the protofilament nearly parallel to the helical axis. In both

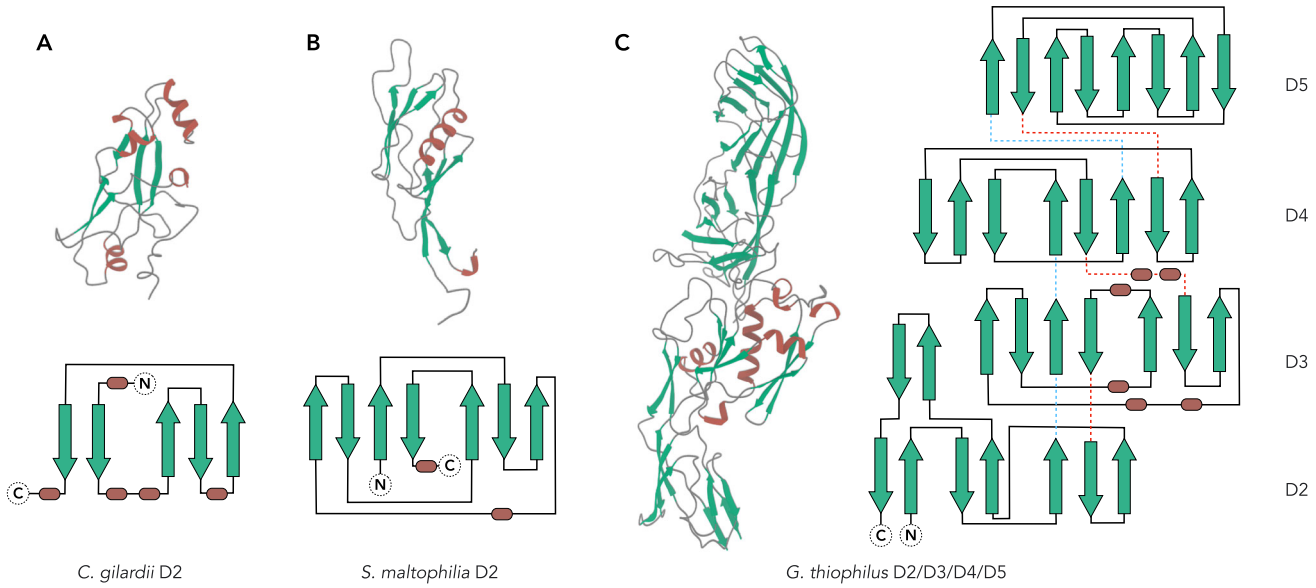

Fig. 2 | The fold of flagellar outer domains from *C. gilardii*, *S. maltophilia*, and *G. thiophilus*. Cryo-EM structures of the flagellar outer domains from *C. gilardii* (**A**), *S. maltophilia* (**B**), *and G. thiophilus* (**C**) were presented, with all α-helices colored brown and β-sheets colored green. These domains are also drawn in schematic representation, with β-sheets shown as green arrows, α-helices as brown cylinders, and loops as dark gray lines. When multiple domains are present, they are connected by blue dashed lines near the N-terminus and red dashed lines near the C-terminus.

C. gilardii and *G. thiophilus*, the outer domains—D2 in *C. gilardii* and D4/D5 in *G. thiophilus*—protrude from the 11-start and adopt either an "up" or "down" conformation.

This dimeric interface, formed between the "up" and "down" conformations of the outer domains, generates a screw-like feature in the *C. gilardii* flagellum, where the surface area buried between subunits is ~270 Å^2 (Fig. 3B) as calculated by PDB PISA[46]. Conversely, the *G. thiophilus* outer domain lacks a distinct screw-like feature, with the dimeric interface between the D5 domains of subunit S_0 and subunit S_{+11} enclosing a buried interface of about 220 Å^2. The interaction area between D4 domains along the 11-start protofilament is relatively small, at around 80 Å^2 (Fig. 3C). Notably, the D2/D3 domains in *G. thiophilus*, located at a smaller cylinder radius of about 80 Å, do not dimerize, preserving the approximate symmetry seen in D0/D1.

The dimerization interface occurs between subunits on the same 11-start protofilament, with dimer packing extending along the left-handed 5-start protofilament. Thus, a seam occurs between two 11-start protofilaments (Fig. 3D), similar to flagella previously studied in *P. aeruginosa* PAO1[16,18]. This seam, reminiscent of those observed in microtubule filaments, interrupts the continuity along the dimer extension, which is the left-handed 5-start of D0/D1 domains (Fig. 3D). Although such structures with a seam were termed "non-helical" in the past, they can be viewed as helical polymers with a large asymmetric unit comprising 22 flagellin subunits while ignoring flagellar supercoiling. This larger ASU reconstruction strategy failed to reach high resolution seven years ago in the reconstruction of *P. aeruginosa* PAO1 flagella data[18] recorded using a Falcon II camera in integrating mode and processed with legacy software SPIDER[47]. However, advancements in cryo-EM now make this strategy possible (Supplementary Fig. 1) using a Gatan K3 camera in counting mode, processed in CryoSPARC[48].

The flagellar surface of *G. thiophilus* is negatively charged

The flagellar outer domains of *C. gilardii* and *S. maltophilia* are significantly smaller compared to the outer domains in *G. thiophilus*, thereby failing to completely shield the inner D0/D1 domain from solvent exposure. In contrast, *G. thiophilus* possesses a two-layered outer domain, comprised of the D2/D3 middle layer and the D4/D5 outer layer, which effectively encapsulates the D0/D1 domain from solvent access. This observation prompted us to investigate whether such coverage impacts the surface charge by estimating the coulombic electrostatic potential for the flagellar surfaces of these three species. Remarkably, the surface of the *G. thiophilus* flagellum is significantly more negatively charged than those of the other two species, attributed to a higher presence of aspartic acid and glutamic acid as surface residues (Supplementary Fig. 3A); analysis of amino acid composition suggested the same conclusion. When examining the D0/D1 domains, the amino acid distribution among the three flagella was nearly identical. However, a comparison of the outer domains revealed that the *C. gilardii* flagellum, while having an overall amino acid composition similar to *S. maltophilia*'s, possesses slightly more negatively charged residues, resulting in a mildly negative surface charge. On the other hand, the outer domain of the *G. thiophilus* flagellum displays a notable difference, containing three times as many negatively charged residues compared to positively charged ones. Even though the *G. thiophilus* outer domain is ~3.5–4 times larger than those of the other two flagella, it has 6–7 times more negatively charged residues, resulting in a predominantly negatively charged surface that fully covers the D0/D1 domain (Supplementary Fig. 3B, C).

The Ig-like fold is ubiquitous in the bacterial flagellin outer domain universe

Flagellar outer domains are known to vary greatly in protein sequence and length[49], which is consistent with our findings. We observed that two of the three flagellins we report contain an Ig-like domain, which has been noted in several other previously reported flagellin structures[16,18–20,22]. This prompted us to question whether flagellin outer domains possess certain folds/domains more commonly than others. The advent of AlphaFold[28] has enabled this type of computational analysis. To date, all AlphaFold predictions have been clustered based on fold similarity[50]. These results, unfortunately, were biased by D0/D1 alignment and could not be directly applied to our study, which seeks domain clustering results independent of D0/D1 domains. Therefore, we downloaded all bacterial flagellin AlphaFold predictions for analysis, trimming the D0/D1 domains (263 residues) based on

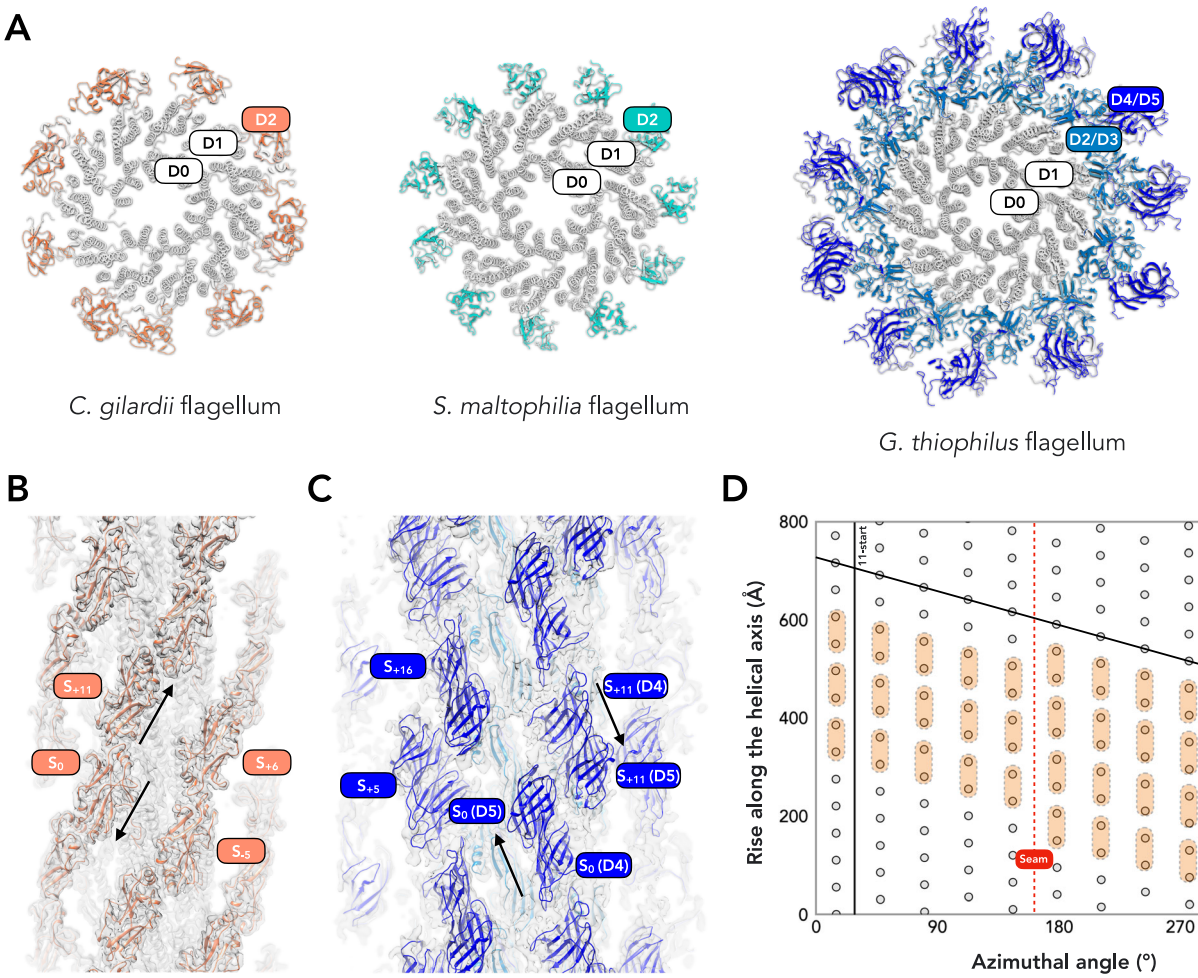

Fig. 3 | Outer domain arrangements on flagellar surface. A The top view of three flagellar filaments with the conserved D0/D1 domains in gray, the D2 domain of *C. gilardii* flagellum in orange, the D2 domain of *S. maltophilia* flagellum in green, the D2/D3 domains of *G. thiophilus* flagellum in light blue, and the D4/D5 domains of *G. thiophilus* flagellum in dark blue. **B** The D2 domain dimerization in *C. gilardii* is presented within the cryo-EM map density. **C** The D4/D5 domain dimerization in *G. thiophilus*, with the model shown within cryo-EM map density. **D** The helical net of the flagellar filament is illustrated using the convention that the surface is unrolled, providing a view from the outside. Gray dots indicate the approximate symmetry present in the D0/D1 domains of all flagellar filaments. The conventional 11-start and left-handed 5-start, originating from the D0/D1 domain, are indicated with black lines. The dimer interface that occurs in the flagellar outer domains of both *C. gilardii* and *G. thiophilus* is illustrated with transparent orange stadium shapes. The seam arising from this dimerization packed along the D0/D1 5-start is indicated by a red dashed line.

prior knowledge of experimental flagellin structures. Interestingly, about 60% of flagellins are shorter than 350 residues (Fig. 4A, B), containing either only the D0/D1 domain, such as the *Bacillus subtilis* flagellum[18], or the D0/D1 domain of *Borreliella burgdorferi* flaA[51], which contains an extra small domain/disordered loop (Fig. 4B). Thus, we set a cutoff at 350 residues for substantial outer domains analysis, creating a library of 16,948 bacterial flagellin outer domain predictions.

For initial clustering, we employed a protein sequence-based approach, similar to, yet more aggressive than, AFDB clustering, by grouping similar protein sequences with MMseqs2[52] at a 30% sequence identity and 80% overlap cutoff. This method categorized the 16,948 predictions into 1281 sequence-based groups. We noticed many flagellin outer domains had a variable number of domains present. We initially considered analyzing individual domains within multi-domain outer regions, but this proved impractical. The variability in protein domain folds makes it difficult to define clear matching criteria, and many domains are unique to bacterial flagellin, complicating their classification. Additionally, analyzing small domains is challenging, particularly given the limitations of current protein fold libraries. Therefore, we focused our analysis on entire flagellin outer regions. Using this strategy, we next applied a more conserved rigid-body alignment strategy by utilizing USalign[53] combined with graph-based community detection clustering via the walktrap[54] algorithm to further group entries with similar lengths and structures. This step organized 1281 sequence-based groups into 682 structural clusters. Each cluster contained a unique "representative" flagellin, distinct from others at the sequence or rigid-structure level (Fig. 4B). Lastly, we examined whether prevalent domain folds were maintained within this library of 682 structural clusters (Supplementary data 1). We used DALI, a tool for analyzing deep phylogenetic relationships and protein homology, known for its sensitivity to fold similarities, to perform an all-to-all analysis[41]. Strikingly, this analysis revealed significant clustering, as shown in the *Z*-score heatmap (Fig. 4C), with about 40% of the representatives carrying at least one Ig-like domain, comprising nearly half of the 16,948 AlphaFold predictions.

Diverse outer domains remain to be studied in bacterial flagella

Lastly, we asked what the most abundant folds/domains are in known bacterial flagellar outer domains and how they are related. To accomplish this, using the same community-based detection strategy for the USalign clustering discussed above, we grouped the remaining representatives using a DALI *Z*-score >10 as the cutoff for significance.

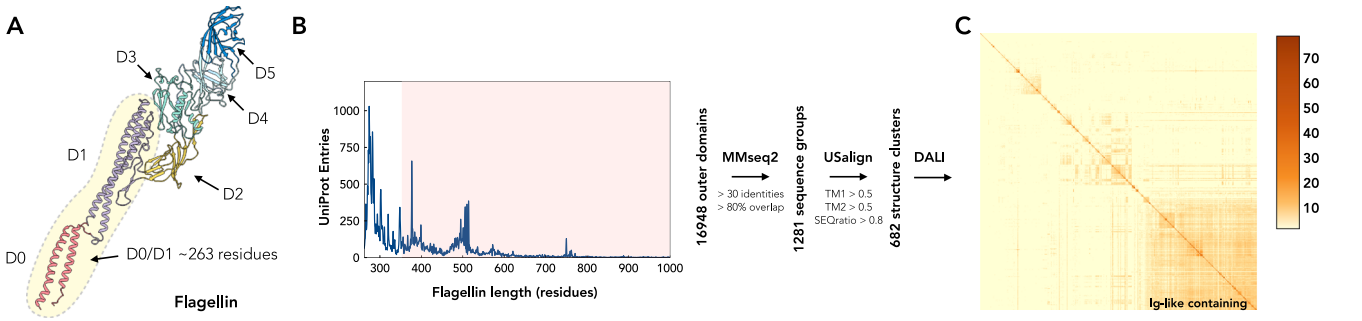

Fig. 4 | Analysis of AlphaFold predictions of bacterial flagellin outer domains. **A** The typical architecture of flagellin includes outer domains. The D0 and D1 domains, which have 263 or a bit more residues, are colored in red and purple, respectively. The other outer domains, extending from inner to outer diameter, are sequentially named D2, D3, D4, and so on. **B** This plot represents the number of UniProt entries corresponding to various lengths of flagellin proteins. Flagellin AlphaFold predictions equal and longer than 350 residues are selected for subsequent multistep analysis, as described in the method section. 682 structural clusters were generated. **C** The matrix visualizes a DALI all-to-all analysis of 682 structural cluster representatives. It is constructed using pairwise DALI *Z*-scores, with the corresponding *Z*-score color scale presented on the right. The most prominent cluster, located at the bottom right of the matrix, contains representatives with Ig-like domains.

We listed the most populated structural clusters, the number of predictions within those clusters, AlphaFold predictions of the representatives, and the known experimental structures belonging to a given cluster (Fig. 5A). The phylum with the highest abundance in each cluster is highlighted accordingly (Supplementary Fig. 4). Not surprisingly, the largest cluster in the center is comprised of representatives that have at least one Ig-like domain within the outer domains. Among 682 structural clusters, five out of the six most populated clusters contain one or more Ig-like domains: cluster #1 with 2304 entries, including *S. maltophilia* flagellin; cluster #2 with 1838 entries, featuring two Ig-like domains as observed in *P. aeruginosa* PAO1[16,18]; cluster #3 with 1094 entries, including another variant of two Ig-like domains, as in *Campylobacter jejuni* flagellin[20]; cluster #4 with 1094 entries, possessing one Ig-like and one β-barrel domain; and cluster #6, showing a single Ig-like domain variant previously identified in a different strain of *P. aeruginosa*[22] (Fig. 5B).

Apart from the Ig-like domain, we observed various other widespread folds in flagellin outer domains. For instance, several flagella structures in similar clusters like #5, #21, and #28 have been reported, including *Salmonella enterica*[21], *Salmonella enterica serovar typhimurium*[10], and *Escherichia coli* K12[17]. Intriguingly, reported flagellin structures FliC[25] and FljB[10] of *S. Typhimurium*, while sharing similar individual domains, are located in clusters #5 and #28, respectively, due to significant differences in the orientation between their two domains. Structural clusters mainly comprising α-helices, such as #7, #29, #115, etc., were also identified. Interestingly, no experimental structure has been documented in these α-helix rich clusters yet. Other clusters having experimental structures include cluster #8, featuring *E. coli* O157:H7, *E. coli* O127:H6, and *Achromobacter spp*. flagella[17], and cluster #10, containing *Salmonella dublin*[23] flagellum. These outer domains typically exhibit "skinny" architectures containing domain folds similar to the de novo-designed domain named foldit3[55]. Other less abundant, but intriguing, clusters were also observed; for instance, the outer domain in cluster #15 resembles the peptidase-M61 catalytic domain[56], referred to as flagella_HEXXH domain[57]; the fold in cluster #35 is similar to the pectate trisaccharide-lyase domain[58]; and the WD40-repeats in cluster #55, more commonly found in eukaryotes, is often thought to serve as a rigid scaffold for protein interactions[59]. Two of the three flagellar structures reported in this study belong to distinct outer domain clusters: *C. gilardii* (cluster #27) and *G. thiophilus* (cluster #41). Although these discovered families are not the most abundant in our analysis, there may be a bias introduced by commonly studied bacterial species that are more redundantly sampled in the database. Overall, this analysis suggests that many more folds exist in flagellar filaments that have yet to be sampled, and their functions remain to be discovered.

Discussion

The field of structural biology has rapidly advanced due to recent developments in cryo-EM[60,61] and structure prediction[28] techniques. These developments, when combined with experimental data, now enable the performance of large-scale[29] and robust structural analyses at the protein fold level. In this study, we determined the cryo-EM structures of three bacterial flagellar filaments at near-atomic resolution. These structures encompass both the conserved D0/D1 regions and the outer domains, providing a general 3D reconstruction strategy for wild-type supercoiled bacterial flagella. Among our findings, we identified distinct folds not previously observed in other bacterial flagella to the best of our knowledge. Of particular interest, was the prevalent Ig-like fold within the known flagellar outer domains. Our in-depth analysis of bacterial flagellar outer domains confirmed the widespread presence of the Ig-like domain; it is found in nearly half of all known flagellin proteins that are longer than 350 residues. The Ig superfamily is ubiquitous, present in various kingdoms, including eukaryotes, prokaryotes, bacteria, viruses, fungi, and plants, and encompasses hundreds of protein families with diverse functions[62]. These include antibodies[63], receptor tyrosine kinases[64], archaeal type IV pili[43], and many others. Given its widespread distribution, the Ig-fold was likely selected as the most abundant outer domain due to its structural stability, its ability to adapt functionally within the loop regions while maintaining the same fold, and its potential for duplication into multiple linear Ig-folds, enabling complex architectures[62]. Additionally, we discovered other interesting flagellar outer domains, which could spark future research to study flagellar filaments in unexplored species.

From the perspective of bacterial "economics", flagellin emerges as one of the most energetically demanding proteins for bacteria to produce, requiring an estimated 20,000 copies per flagellum. This high energy investment in flagella production was clearly demonstrated over 30 years ago in studies when bacteria were grown in stirred liquid cultures where motility provides no advantage but incurs additional energy expenditure. Bacterial cells lost their flagellar filaments resulting from spontaneous mutations in flagellar genes within just 10 days, highlighting an evolutionary advantage for bacteria that conserve energy by not producing flagella[65]. Recent estimates suggest that the energy cost of synthesizing flagellar filaments accounts for 0.5–40% of the total energy budget across different species[66]. The presence of outer domains in flagella, especially in species like *G. thiophilus*, where the flagellin size can reach up to 793 residues—

https://doi.org/10.1038/s41467-024-53923-w

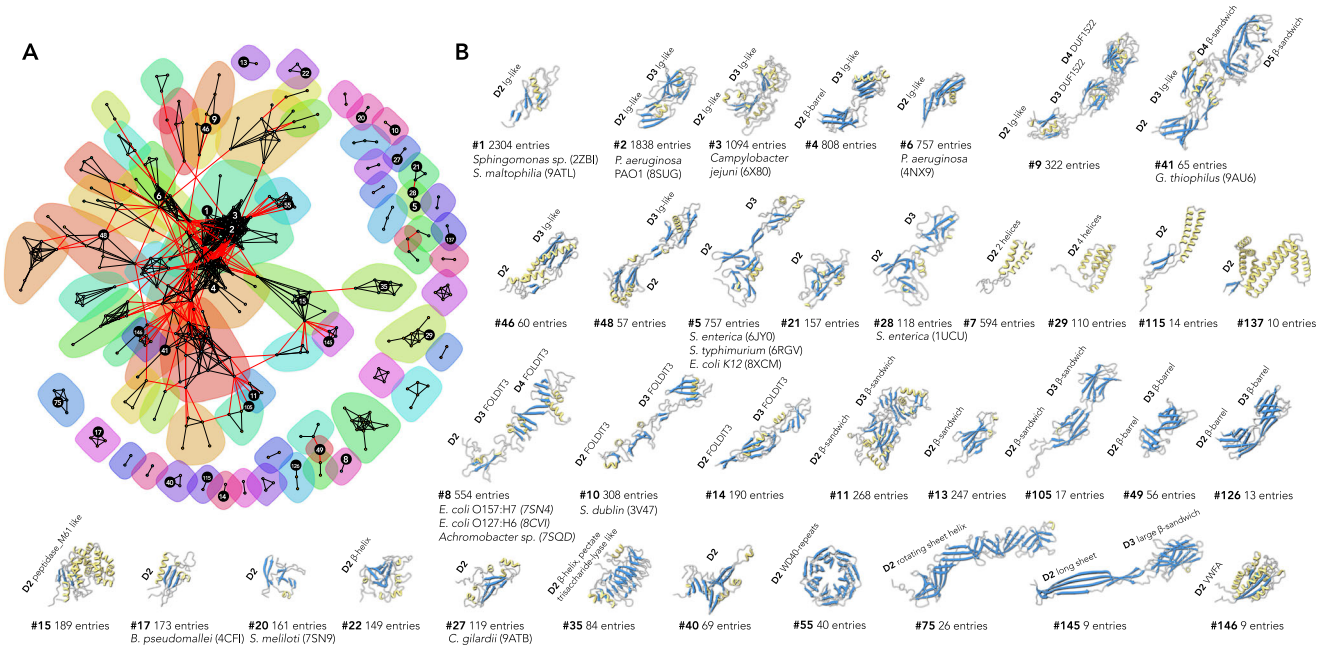

Fig. 5 | Populated flagellin outer domains in the bacterial domain. A The relationships among 682 structural clusters are depicted using the Fruchterman-Reingold algorithm, a force-directed layout algorithm from the R/igraph package. The communities, indicated with transparent color circles, were detected using the walktrap algorithm with a step size of 6. Additionally, some most populated clusters, as illustrated in (**B**), are highlighted with black circles. **B** AlphaFold predictions or experimentally determined structures from populous structural clusters are organized according to the population size of the first representative in a given community. Within these models, all α-helices are colored yellow, β-sheets blue, and loops gray. Known domain folds are labeled. For clusters with available cryo-EM or X-ray structures, the corresponding species and PDB IDs are provided.

approximately triple the size of a flagellin composed solely of the D0/D1 domain—underscores the hypothesis that these elaborate structures must confer substantial benefits to justify their high energy costs.

While the conserved D0/D1 domain is crucial for swimming motility, the necessity of the outer domains for this function has been questioned. In *S. typhimurium*, it has been shown that flagella lacking an outer domain remain motile, suggesting that this domain is not required for swimming motility[15]. Yet, studies suggest that the outer domain may enhance motility in several ways. For instance, in *E. coli*, the outer domains have been shown to extend tumbling time, thereby improving navigation efficiency[17]. Similarly, in *P. aeruginosa* PAO1, mutations in the outer domain disrupt the formation of supercoiled filaments, adversely affecting motility[16]. Conceptualizing the flagellar filament as a propeller, the outer domain could enhance the motility by increasing the flagellar diameter and adjusting the angle of attack, potentially aiding movement in high-viscosity environments[37]. Beyond motility, outer domains may serve diverse functions across different bacterial species. For example, the outer domains in cluster #15 resemble glycyl aminopeptidase, suggesting a role in digesting extracellular proteins to facilitate the uptake of essential amino acids[67]. In the case of *G. thiophilus* (cluster #41), which thrives in anoxic or microaerophilic environments, and has the largest flagellin structure identified, its flagellum features an unusually negatively charged surface. This characteristic might regulate motility modes in various environments. For instance, in the presence of positively charged surfaces like minerals, iron oxides, and most metal oxides, the flagellum may adhere more readily, potentially leading bacteria to interact with the surface. Conversely, in environments that are more negatively charged and devoid of minerals, the flagellum could experience reduced resistance, allowing for smoother rotation and enhanced motility.

Several studies have shown that flagellar outer domains can undergo additional surface structure organization, disrupting the D0/D1 symmetry and resulting in the formation of structures such as dimers, dimers with seams, tetramers, and tetramers with seams[16–18]. Notably, despite the similarity in flagellar outer domain architectures between *E. coli* O157:H7 and *E. coli* O127:H6, which both belong to cluster #8, they exhibit distinct surface organization patterns: O157:H7 forms dimers without a seam, whereas O127:H6 assembles into tetramers with a seam. On the other hand, dimers with seams are commonly observed in flagellar outer domains, with known cryo-EM structures identified in bacteria like *P. aeruginosa* PAO1 (cluster #2), *C. gilardii* (cluster #27), and *G. thiophilus* (cluster #41). This suggests that such surface organization reflects a strategy for optimized spatial packing, influenced more by the size and surface properties of the outer domains than by the protein fold or specific function. From the helical net (Fig. 3C), it becomes apparent that the linker connecting D1 and D2 domains in subunit S_0 is positioned close to S_5, S_6, and S_{11}. Should dimerization occur between S_0 and S_{11}, with the resulting protofilament extending along the left-handed 5-start (all the numbered start refers to the classic D0/D1 symmetry), a seam will form (Fig. 3C). Conversely, when dimerization occurs between S_0 and S_5, extending along the right-handed 6-start, as seen in *S. meliloti*[17], no seam is formed. Thus, it is reasonable to anticipate the discovery of other packing arrangements, such as trimerization or pentamerization, within the flagellar outer domain in the future. Such diversity in packing could further reveal the adaptability and complexity of flagellar structure and function.

The diversity seen in flagellin outer domains hints that these structural folds have been co-opted randomly or through horizontal gene transfer from other proteins. A similar phenomenon has been observed in bacterial type IV pili (T4P). One well-studied T4P from *Pseudomonas aeruginosa* PAO11 has a long N-terminal partially melted helix and a C-terminal globular domain[68]. Later, T4Ps with very different outer domains but conserved N-terminal helices were found in other bacteria, like *Thermus thermophilus*[69]. Additionally, a bacterial

T4P was found with two chains (two genes) for the N-helix and C-domain[70], and another with only an N-terminal helix[71]. Returning to bacterial flagella, all share a conserved D0/D1 domain, sufficient for supercoiled motility in many species, similar to the T4P N-terminal helix. The diverse D2/D3, etc., domains, akin to T4P outer domains, were likely acquired independently during evolution. Identifying their origins can be challenging. For instance, we could not pinpoint the D2 domain's origin in *C. gilardii*, possibly due to its small size. *S. maltophilia*'s D2 and *G. thiophilus*'s D2/D3 yielded many hits due to their Ig-domain nature, found in various protein families, including archaeal type IV pilin. The origin of *G. thiophilus*'s D4 and D5 is clearer. D5 is a duplication of D4, inserted within one of its loops. D4 is related to the D2 domain of FliD cap proteins, found at the flagellum's tip (Supplementary Fig. 5A). This suggests that the augmentation and diversification of this organism's flagellum's outer domains occurred through the gradual addition of preexisting domains encoded by flagellar genes via recombination. Finally, we examined additional clusters that could imply the origin of outer domains. For three clusters (peptidase, protease, WD40), we found cellular homologs with annotations and solved structures not labeled as flagellin (Supplementary Fig. 5B–D), suggesting they may be co-opted or horizontally transferred from other proteins.

Bacteria often harbor multiple flagellin genes, such as those seen in *G. thiophilus*, suggesting the potential for diverse functionality fulfilled by different flagellin variants, which are produced in response to environmental needs. The flagellin genes in *G. thiophilus* are situated in close proximity, with two of them being particularly near each other but not necessary in the same operon: EP073_12645 (which has a reported cryo-EM structure) and EP073_12655. The third flagellin gene, EP073_12695, is clearly in a different operon. The conditions under which these flagellin genes are expressed and the timing of their production remain unclear based on protein fold analysis using Alphafold predictions. All three flagellins have similar lengths (ranging from 793 to 801 residues), and all their outer domains belong to cluster #41. The most notable difference lies in the surface of the D3/D4 domains in EP073_12695, which has considerably fewer negatively charged residues (Asp and Glu combined) compared to the other two flagellins, with a decrease from 44 to 23 residues. This suggests that EP073_12695 might be produced under specific environmental conditions, such as varying pH levels. However, further investigation using mutagenesis will be necessary to confirm this hypothesis once genetic tools become available for this species.

Such complexity in flagellin expression and function is not unique to *G. thiophilus* and has been observed in various bacteria. For instance, bacteria like *Azospirillum lipoferum* and *Shewanella piezotolerans* have sophisticated adaptive behavior by producing polar flagella for swimming in liquid environments and multiple lateral flagella for swarming on surfaces. In some bacteria, including *Shewanella putrefaciens* and *Helicobacter pylori*, flagellar genes are regulated hierarchically, leading to a single flagellar filament composed of two different flagellins, one located near the hook and the other further away. This hierarchical regulation of appendages can be seen in the archaeal domain as well, with *Sulfolobales islandicus REY15A* utilizing the same pilin protein to generate two different type IV pili structures by modifying its Ig-like outer domain in response to varying environmental needs[42]. This underscores a broader principle of microbial adaptability and the strategic regulation of extracellular filament secretion. Clearly, despite these advances, our understanding of the regulatory mechanisms governing flagellin gene expression, extracellular filament secretion, and the specific roles of uncharacterized outer domains remains incomplete. Future research dedicated to unraveling these aspects promises to deepen our insights into microbial motility and interaction with their environments.

Methods

C. gilardii flagellar filament preparation

C. gilardii cells were grown aerobically at 37 °C for 48 h in chemical defined medium[72] in 5% CO_2 incubator. For isolation of filaments, cells were spun down at 7000 rpm (9120 × *g*, Beckman JLA, 12.500 rotor). Pellet was suspended with 4 ml phosphate-buffered saline (PBS) buffer pH 7.2, and filaments were sheared off from the cell using a homogenizer at 10,000 rpm for 10 min. Cells and debris were removed by centrifugation at 10,000 × *g* for 10 min. After this, filaments were pelleted from the supernatant by ultra-centrifugation (Beckman 50.3 Ti ultra-rotor, 35,000 rpm) at 4 °C for 1.5 h, and subsequently resuspended with 150 μl PBS buffer. DNase I (NEB) was added to the sample to remove possible extracellular DNA fibers.

S. maltophilia flagellar filament preparation

S. maltophilia cells were grown aerobically at 37 °C for 48 h in Tryptic Soy Broth (TSB, BD™) medium in a 5% CO_2 incubator. The resultant pellet was resuspended in 4 mL of PBS buffer, and the cell suspension was put under a homogenizer for 10 min to shear off the extracellular filaments as described above. The cells were then removed by centrifugation at 10,000 × *g* for 10 min. The supernatant was collected, and the filaments were pelleted by ultracentrifugation (Beckman 50.3 Ti ultra-rotor, 35,000 rpm, 1.5 h, 4 °C). After the run, the supernatant was removed, and the pellet was resuspended in 150 μL of PBS. DNase I (NEB) was added to the sample to remove possible extracellular DNA fibers.

G. thiophilus flagellar filament preparation

G. thiophilus cells were grown anaerobically at 30 °C in anaerobic freshwater medium 503 (DSMZ) in a volume of 10 mL. After this, cells were vortexed for 30 min to mechanically shear off the extracellular filaments, and cells were subsequently removed by centrifugation at 10,000 × *g* for 10 min. The supernatant was collected, and filaments were further enriched by an overnight 20% ammonium sulfate precipitation at 4 °C. The resultant flagellar filament pellet was collected and resuspended with 150 μL 100 mM ethanolamine buffer pH 10.5, incubated with Dnase I (NEB) prior to plunge freezing.

Cryo-EM conditions and image processing

The flagellar filament sample (4.5 μL) was applied to glow-discharged lacey carbon grids and then plunge-frozen using an EM GP2 Plunge Freezer (Leica). The cryo-EM micrographs were collected on a 300 keV Titan Krios with a K3 camera at 1.11 Å per pixel and a total dose of 50 $e^-/Å^2$. The cryo-EM workflow was initiated with patch motion corrections and CTF estimations in cryoSPARC[48,73,74]. Following this, an automated picking of particle segments was conducted using the 'Filament Tracer' function with a shift of 10 pixels between adjacent boxes. All auto-picked particles were subsequently 2D classified with multiple rounds, and all particles in bad 2D averages were removed. Next, the possible helical symmetries were calculated from averaged power spectra generated from the raw particles (640-pixel box)[75,76]. The identified parameters consistent with canonical flagellar filaments were then applied in 3D helical refinement to generate a high-resolution map of the D0/D1 domains. Subsequently, the power spectra's meridian area was carefully examined to detect layer-lines that might indicate interfaces of additional outer domains. Upon the identification of such layer-lines, a re-extraction of particles was performed, applying a shift slightly greater than the periodicity observed in the power spectra to avoid particle duplications. After that, 3D reconstruction was performed using "Helical Refinement" first, "Homogenous Refinement" next, then "Local Refine", Local CTF refinement, and another round of "Local Refine" using CTF-refined particles (Supplementary Fig. 1). The resolution of each reconstruction was estimated by Map:Map FSC, Model:Map FSC, and d_{99}[77]. Maps used in the resolution estimation

Table 1 | Cryo-EM and refinement statistics of flagellar filaments

Parameter	*C. gilardii* flagellum	*S. maltophilia* flagellum	*G. thiophilus* flagellum
Data collection and processing			
Voltage (kV)	300	300	300
Electron exposure (e^{-} Å^{-2})	50	50	50
Pixel size (Å)	1.11	1.11	1.11
Particle images (*n*)	223,582	275,407	21,618
Shift (pixel)	100	10	100
Approximate Helical symmetry before symmetry relaxation			
Point group	C1	C1	C1
Helical rise (Å)	110.25	5.03	111.12
Helical twist (°)	−0.61	65.40	−0.30
Map resolution (Å)			
Map:map FSC (0.143)	3.4	3.3	4.1
Model:map FSC (0.5)	3.6	3.5	4.5
d_{99}	3.8	3.6	4.5
Refinement and model validation			
Ramachandran favored (%)	95.8	94.1	93.0
Ramachandran outliers (%)	0.0	0.0	0.4
Real space CC	0.88	0.89	0.76
Clashscore	5.7	6.0	9.5
Bonds RMSD, length (Å)	0.008	0.005	0.010
Bonds RMSD, angles (°)	0.787	0.633	0.706
Deposition ID			
PDB (model)	9ATB	9ATL	9AU6
EMDB (map)	EMD-43829	EMD-43830	EMD-43868

were sharpened using Local Filter available in cryoSPARC. The statistics are listed in Table 1.

Model building of flagellar filaments

The first step in model building is to identify the correct flagellin protein from the experimental cryo-EM map, especially when there are multiple candidates. This was the case with the G. thiophilus flagellar sample, which had three similar flagellin sequences in its genome: A0A3R5UZF4, A0A3R5UW90, and A0A3R5V2W1. To address this, full-length modeling was carried out for each sequence until a region in the map was encountered that the sequences could not account for (see Supplementary Fig. 6). A0A3R5UZF4 was identified as the matching sequence through meticulous examination, fully agreeing with the cryo-EM density.

Regarding the modeling, the AlphaFold predicted structure of a single flagellin subunit was initially docked into the cryo-EM map. Domains (for instance, D0–D5 in *G. thiophilus* flagella) were docked individually since AlphaFold predictions for domain-domain orientation are frequently imprecise. The subunit was then manually adjusted and refined in Coot[78]. Given that the outer domain forms a dimer in *C. gilardii* and *G. thiophilus* flagella, this procedure was repeated for the other flagellin within the dimer. Following this, the refined single flagellin or flagellin dimers were docked into the other ten protofilaments within the supercoiled flagellar filament map and subjected to real-space refinement via PHENIX[79]. The filament model's quality was assessed using MolProbity[80], and the refinement statistics for all three flagellar filaments are detailed in Table 1. Cryo-EM map and model visualization were primarily done in ChimeraX[81].

Bioinformatic clustering of flagellin outer domains

First, AlphaFold[28] predictions of all bacterial proteins annotated in UniProt as "flagellin" were downloaded from the AlphaFold Protein Structure Database[29]. To conduct the analysis in the absence of D0/D1 domains, protein regions corresponding to D0/D1 domains were carefully trimmed from the prediction by excising the first 163 residues and the last 100 residues. This cutoff is selected based on existing experimental flagellin structures. Furthermore, flagellins shorter than 350 amino acids (outer structure shorter than 87 residues) were excluded. This is because the average size of a protein domain is approximately 100 residues, and analysis is based on AlphaFold predictions, which are very meaningful for such a short region. Next, DSSP[82] was used to estimate the secondary structure components of the remaining models, and models with <35% total secondary structures were removed, leaving 16,948 flagellin predictions for clustering analysis.

The initial clustering was performed at the amino acid sequence level using MMseqs2[52], with the conversion from PDB to FASTA format achieved via Pdb2Fasta. The Mmseqs2 easy-cluster workflow was used based on a minimum of 30% sequence identity, requiring the short sequence to be at least 80% the length of the other sequence. This process reduced the number of 16,948 predictions to 1281 "sequence groups", with representatives from each group chosen by MMSeqs2. Further investigation into rigid-body structural similarities among the 1281 groups was conducted using USalign[53] pairwise alignments using the semi-non-sequential alignment (sNS or -mm 6), resulting in 819,480 alignments from the 1281 representative structures. An undirected graph was built using the pairwise TM scores, TM1 and TM2, as described in the USalign documentation, and from an alignment length score calculated by taking the length of alignment (Lali in USalign) and dividing it by the total sequence length of the larger outer domain of the two. The cutoffs for edges to be included were TM1 and TM2 >0.5, and the alignment length score >0.8. This was done to include strong alignments yet eliminate edges that corresponded to high TM scores for single-domain structures aligning with a similar domain in a multi-domain structure. Using this graph, community detection-based clustering was performed using the walktrap algorithm with a step size of 6. From the 1281 sequence groups, 705 were included in the graph with 7115 edges and clustered into 107 communities, the remaining 575 had no significant connections. Merging clustering results from both MMSeqs2 and USalign left us with 682 total clusters encompassing all 16,948 input structures. Clusters were sorted and numbered by member number, largest to smallest. 142 clusters had 10 or more members, 251 clusters had 2–9 members, and 274 clusters had only one member, finding no significant structural or sequence similarity through either MMSeq2 or USalign. Within each group, a representative model was chosen for downstream steps by sorting by highest AlphaFold predicted score, pLDDT, and choosing the first structure within five residues of the average sequence length of the cluster. Finally, an all-to-all DALI[41] alignment was performed among those 682 representatives to identify similar domains within multi-domain flagellar outer structures, as well as similar structures that rigid-body methods failed to align due to different predictions in domain arrangements. A similar strategy to the USalign clustering was used with the cutoff for edge inclusion set at *Z*-scores above 10. We used a stricter cutoff of 10 here, rather than the standard 8, to prevent grouping larger domains that share only limited similarities. The walktrap algorithm with a step size of 6 was again used for community detection. The graph was drawn using the Fructerman–Reingold algorithm[83], a force-directed layout algorithm within the R/igraph package.

Reporting summary

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

Data availability

The three-dimensional reconstructions generated in this study have been deposited in the Electron Microscopy Data Bank under accession codes EMD-43829 (*C. gilardii* flagellum), EMD-43830 (*S. maltophilia* flagellum), and EMD-43868 (*G. thiophilus* flagellum). The atomic models generated in this study have been deposited in the Protein Data Bank under accession codes 9ATB (*C. gilardii* flagellum), 9ATL (*S. maltophilia* flagellum), and 9AU6 (*G. thiophilus* flagellum).

References

1. Soutourina, O. A., Semenova, E. A., Parfenova, V. V., Danchin, A. & Bertin, P. Control of bacterial motility by environmental factors in polarly flagellated and peritrichous bacteria isolated from Lake Baikal. *Appl. Environ. Microbiol.* **67**, 3852–3859 (2001).
2. Zhou, B. B., Szymanski, C. M. & Baylink, A. Bacterial chemotaxis in human diseases. *Trends Microbiol.* **31**, 453–467 (2023).
3. Colin, R., Ni, B., Laganenka, L. & Sourjik, V. Multiple functions of flagellar motility and chemotaxis in bacterial physiology. *FEMS Microbiol. Rev.* **45**, https://doi.org/10.1093/femsre/fuab038 (2021).
4. Spohn, G., and Scarlato, V. Motility, Chemotaxis, and Flagella. In *Helicobacter pylori*: *Physiology and Genetics* (eds Mobley, H. L. T., Mendz, G. L. & Hazell, S. L.) (2001).
5. Palma, V., Gutierrez, M. S., Vargas, O., Parthasarathy, R. & Navarrete, P. Methods to evaluate bacterial motility and its role in bacterial–host interactions. *Microorganisms* **10**, https://doi.org/10.3390/microorganisms10030563 (2022).
6. Macnab, R. M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).
7. Bardy, S. L., Ng, S. Y. M. & Jarrell, K. F. Prokaryotic motility structures. *Microbiology* **149**, 295–304 (2003).
8. Kreutzberger, M. A. B. et al. Convergent evolution in the supercoiling of prokaryotic flagellar filaments. *Cell* **185**, 3487–3500 e3414 (2022).
9. Erdem, A. L., Avelino, F., Xicohtencatl-Cortes, J. & Giron, J. A. Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. *J. Bacteriol.* **189**, 7426–7435 (2007).
10. Horstmann, J. A. et al. Methylation of Salmonella Typhimurium flagella promotes bacterial adhesion and host cell invasion. *Nat. Commun.* **11**, 2013 (2020).
11. Sevrin, G. et al. Adaptation of adherent-invasive *E. coli* to gut environment: impact on flagellum expression and bacterial colonization ability. *Gut Microbes* **11**, 364–380 (2020).
12. Lemon, K. P., Higgins, D. E. & Kolter, R. Flagellar motility is critical for biofilm formation. *J. Bacteriol.* **189**, 4418–4424 (2007).
13. Zhou, X. et al. Flagellin of enteropathogenic *Escherichia coli* stimulates interleukin-8 production in T84 cells. *Infect. Immun.* **71**, 2120–2129 (2003).
14. Malapaka, R. R., Adebayo, L. O. & Tripp, B. C. A deletion variant study of the functional role of the *Salmonella flagellin* hypervariable domain region in motility. *J. Mol. Biol.* **365**, 1102–1116 (2007).
15. Yoshioka, K., Aizawa, S. & Yamaguchi, S. Flagellar filament structure and cell motility of *Salmonella typhimurium* mutants lacking part of the outer domain of flagellin. *J. Bacteriol.* **177**, 1090–1093 (1995).
16. Nedeljkovic, M. et al. An unbroken network of interactions connecting flagellin domains is required for motility in viscous environments. *PLoS Pathogens* **19**, https://doi.org/10.1371/journal.ppat.1010979 (2023).
17. Kreutzberger, M. A. B. et al. Flagellin outer domain dimerization modulates motility in pathogenic and soil bacteria from viscous environments. *Nat. Commun.* **13**, 1422 (2022).
18. Wang, F. et al. A structural model of flagellar filament switching across multiple bacterial species. *Nat. Commun.* **8**, 960 (2017).
19. Maruyama, Y., Momma, M., Mikami, B., Hashimoto, W. & Murata, K. Crystal structure of a novel bacterial cell-surface flagellin binding to a polysaccharide. *Biochemistry* **47**, 1393–1402 (2008).
20. Kreutzberger, M. A. B., Ewing, C., Poly, F., Wang, F. & Egelman, E. H. Atomic structure of the *Campylobacter jejuni* flagellar filament reveals how epsilon Proteobacteria escaped Toll-like receptor 5 surveillance. *Proc. Natl Acad. Sci. USA* **117**, 16985–16991 (2020).
21. Yamaguchi, T. et al. Structural and functional comparison of flagellar filaments composed of FljB and FliC. *Biomolecules* **10**, https://doi.org/10.3390/biom10020246 (2023).
22. Song, W. S. & Yoon, S. I. Crystal structure of FliC flagellin from *Pseudomonas aeruginosa* and its implication in TLR5 binding and formation of the flagellar filament. *Biochem. Biophys. Res. Commun.* **444**, 109–115 (2014).
23. Yoon, S. I. et al. Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859–864 (2012).
24. Nithichanon, A. et al. Sequence- and structure-based immunoreactive epitope discovery for *Burkholderia pseudomallei* flagellin. *PLoS Negl. Trop. Dis.* **9**, e0003917 (2015).
25. Maki-Yonekura, S., Yonekura, K. & Namba, K. Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. *Nat. Struct. Mol. Biol.* **17**, 417–422 (2010).
26. Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643–650 (2003).
27. Bzdyl, N. M., Moran, C. L., Bendo, J. & Sarkar-Tyson, M. Pathogenicity and virulence of *Burkholderia pseudomallei*. *Virulence* **13**, 1945–1965 (2022).
28. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
29. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
30. Coenye, T. et al. Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *Int. J. Syst. Bacteriol.* **49**, 405–413 (1999).
31. Ruiz, C., McCarley, A., Espejo, M. L., Cooper, K. K. & Harmon, D. E. Comparative genomics reveals a well-conserved intrinsic resistome in the emerging multidrug-resistant pathogen *Cupriavidus gilardii*. *mSphere* **4**. https://doi.org/10.1128/mSphere.00631-19 (2019).
32. de Oliveira-Garcia, D. et al. Characterization of flagella produced by clinical strains of *Stenotrophomonas maltophilia*. *Emerg. Infect. Dis.* **8**, 918–923 (2002).
33. Adegoke, A. A., Stenström, T. A. & Okoh, A. I. *Stenotrophomonas maltophilia* as an emerging ubiquitous pathogen: looking beyond contemporary antibiotic therapy. *Front. Microbiol.* **8**, https://doi.org/10.3389/fmicb.2017.02276 (2017).
34. Janssen, P. H., Liesack, W. & Schink, B. *Geovibrio thiophilus* sp. nov., a novel sulfur-reducing bacterium belonging to the phylum Deferribacteres. *Int. J. Syst. Evol. Microbiol.* **52**, 1341–1347 (2002).
35. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320 (1994).
36. Fang, X. et al. Infection caused by *Cupriavidus gilardii* in a convalescent COVID-19 patient. *Int. J. Infect. Dis.* **134**, 287–289 (2023).
37. Cohen-Krausz, S. & Trachtenberg, S. The structure of the helically perturbed flagellar filament of *Pseudomonas rhodos*: implications for the absence of the outer domain in other complex flagellins and for the flexibility of the radial spokes. *Mol. Microbiol.* **48**, 1305–1316 (2003).
38. Trachtenberg, S., DeRosier, D. J., Zemlin, F. & Beckmann, E. Non-helical perturbations of the flagellar filament: *Salmonella typhimurium* SJW117 at 9.6 A resolution. *J. Mol. Biol.* **276**, 759–773 (1998).

39. Cohen-Krausz, S. & Trachtenberg, S. Helical perturbations of the flagellar filament: rhizobium lupini H13-3 at 13 A resolution. *J. Struct. Biol.* **122**, 267–282 (1998).
40. van Kempen, M. et al. Fast and accurate protein structure search with FoldSeek. *Nat. Biotechno.* **42**, 243–246 (2024).
41. Holm, L., Laiho, A., Toronen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, e4519 (2023).
42. Liu, J. et al. Two distinct archaeal type IV pili structures formed by proteins with identical sequence. *Nat. Commun.* **15**, 5049 (2024).
43. Wang, F. et al. The structures of two archaeal type IV pili illuminate evolutionary relationships. *Nat. Commun.* **11**, 3424 (2020).
44. Wang, F. et al. An extensively glycosylated archaeal pilus survives extreme conditions. *Nat. Microbiol.* **4**, 1401–1410 (2019).
45. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
46. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
47. Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
48. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
49. Beatson, S. A., Minamino, T. & Pallen, M. J. Variation in bacterial flagellins: from sequence to structure. *Trends Microbiol.* **14**, 151–155 (2006).
50. Barrio-Hernandez, I. et al. Clustering predicted structures at the scale of the known protein universe. *Nature* **622**, 637–645 (2023).
51. Zhang, K. et al. Analysis of a flagellar filament cap mutant reveals that HtrA serine protease degrades unfolded flagellin protein in the periplasm of *Borrelia burgdorferi*. *Mol. Microbiol* **111**, 1652–1670 (2019).
52. Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
53. Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).
54. Pons, P. & Latapy, M. Computing communities in large networks using random walks. *Lect. Notes Comput. Sci.* **3733**, 284–293 (2005).
55. Koepnick, B. et al. De novo protein design by citizen scientists. *Nature* **570**, 390–394 (2019).
56. Wang, X. et al. Discovery and characterization of a myxobacterial lanthipeptide with unique biosynthetic features and anti-inflammatory activity. *J. Am. Chem. Soc.* **145**, 16924–16937 (2023).
57. Eckhard, U. et al. Discovery of a proteolytic flagellin family in diverse bacterial phyla that assembles enzymatically active flagella. *Nat. Commun.* **8**, https://doi.org/10.1038/s41467-017-00599-0 (2017).
58. Alahuhta, M. et al. The catalytic mechanism and unique low pH optimum of *Caldicellulosiruptor bescii* family 3 pectate lyase. *Acta Crystallogr. D Biol. Crystallogr* **71**, 1946–1954 (2015).
59. Stirnimann, C. U., Petsalaki, E., Russell, R. B. & Muller, C. W. WD40 proteins propel cellular networks. *Trends Biochem. Sci.* **35**, 565–574 (2010).
60. Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **343**, 1443–1444 (2014).
61. Egelman, E. H. The current revolution in Cryo-EM. *Biophys. J.* **110**, 1008–1012 (2016).
62. Halaby, D. M., Poupon, A. & Mornon, J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng.* **12**, 563–571 (1999).
63. Feige, M. J., Hendershot, L. M. & Buchner, J. How antibodies fold. *Trends Biochem. Sci.* **35**, 189–198 (2010).
64. Linger, R. M., Keating, A. K., Earp, H. S. & Graham, D. K. TAM receptor tyrosine kinases: biologic functions, signaling, and potential therapeutic targeting in human cancer. *Adv. Cancer Res.* **100**, 35–83 (2008).
65. Macnab, R. M. Genetics and biogenesis of bacterial flagella. *Annu. Rev. Genet.* **26**, 131–158 (1992).
66. Schavemaker, P. E. & Lynch, M. Flagellar energy costs across the tree of life. *Elife* **11**, https://doi.org/10.7554/eLife.77266 (2022).
67. Byun, T., and Blinkovsky, A. Glycyl aminopeptidase (Sphingomonas). In *Handbook of Proteolytic Enzymes* 2nd edn (eds Barrett, A.J., Rawlings, N.D. & Woessner, J.F.) 470-471 (Academic Press, 2004).
68. Wang, F. et al. Cryoelectron microscopy reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* Type IV Pili at sub-nanometer resolution. *Structure* **25**, 1423–1435 e1424 (2017).
69. Neuhaus, A. et al. Cryo-electron microscopy reveals two distinct type IV pili assembled by the same bacterium. *Nat. Commun.* **11**, 2231 (2020).
70. Wang, F. et al. Cryo-EM structure of an extracellular Geobacter OmcE cytochrome filament reveals tetrahaem packing. *Nat. Microbiol.* **7**, 1291–1300 (2022).
71. Sonani, R. R. et al. Tad and toxin-coregulated pilus structures reveal unexpected diversity in bacterial type IV pili. *Proc. Natl Acad. Sci. USA* **120**, e2316668120 (2023).
72. Desai, K., Mashburn-Warren, L., Federle, M. J. & Morrison, D. A. Development of competence for genetic transformation of *Streptococcus mutans* in a chemically defined medium. *J. Bacteriol.* **194**, 3774–3780 (2012).
73. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
74. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
75. Egelman, E. H. Reconstruction of helical filaments and tubes. *Methods Enzymol.* **482**, 167–183 (2010).
76. Wang, F., Gnewou, O., Solemanifar, A., Conticello, V. P. & Egelman, E. H. Cryo-EM of helical polymers. *Chem. Rev.* **122**, 14055–14065 (2022).
77. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).
78. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr* **60**, 2126–2132 (2004).
79. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544 (2018).
80. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
81. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
82. Kabsch, W. & Sander, C. How good are predictions of protein secondary structure? *FEBS Lett.* **155**, 179–182 (1983).
83. Miguel, J., Caballé, S. & Xhafa, F. Trustworthiness in action: data collection, processing, and visualization methods for real online courses. In *Intelligent Data Analysis for e-Learning* (eds Miguel, J., Caballé, S. & Xhafa, F.) Ch. 7, 129–142 (Academic Press, 2017).

Acknowledgements

This research was, in part, supported by the National Cancer Institute's National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research under contract 75N91019D00024. Electron microscopy screening was carried out in the UAB Cryo-EM Facility, supported by the Institutional Research Core Program and O'Neal Comprehensive Cancer

Center (NIH grant P30 CA013148), with additional funding from NIH grant S10 OD024978. We are grateful to Dr. James Kizziah, Dr. Thomas Edwards, Dr. Tara Fox, Dr. Adam Wier, and Dr. Zhiqing Wang for assisting with the screening or data collection. We'd like to thank Dr. Ed Egelman, Dr. Mark Kreutzberger, and Dr. Daniel Bond for helpful discussions in the preparation of the manuscripts. The work in the F.W. laboratory was supported by NIH grant GM138756, DoE grant SC0024303, and a pilot grant from the UAB Global Center for Craniofacial Oral and Dental Disorders (GC-CODED). The work in the H.W. laboratory was supported by NIH grants R01 DE022350, R01 DE028329 and T90 DE030859.

Author contributions

J.L.F. and H.Z. performed sample preparation. J.L.F., H.A.P., S.K.H., and S.T.R. performed microscopy image analysis. H.Z. performed sequencing experiments to confirm flagellin sequences. N.F.B., M.K., and F.W. performed outer domains bioinformatic analysis. F.W. and H.W. obtained funding and supervised the research. F.W. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53923-w.

Correspondence and requests for materials should be addressed to Hui Wu or Fengbin Wang.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.