J Mol Model (2006) 12: 481–493
DOI 10.1007/s00894-005-0075-x

ORIGINAL PAPER

V. Raghu Ram Malapaka · Brian C. Tripp

A theoretical model of *Aquifex pyrophilus* flagellin: implications for its thermostability

Received: 20 May 2005 / Accepted: 17 October 2005 / Published online: 13 January 2006

Abstract *Aquifex pyrophilus* is a flagellated hyperthermophilic eubacterial species that grows optimally at 85°C. The thermostable *A. pyrophilus* flagellar filament is primarily composed of a single protein called flagellin (FlaA). The N- and C-terminal sequence regions of FlaA are important for self-assembly and share high sequence similarity with mesophilic bacterial flagellins. We have developed a predictive 3D-structure of FlaA, using the published structure of mesophilic *Salmonella typhimurium* flagellin (FliC) as a template and analyzed it with respect to possible determinants of thermostability. A sequence comparison of FlaA and FliC revealed a +7.0% increase in FlaA hydrophobic residues, a +0.6% increase in charged residues and a corresponding decrease of −6.0% in polar residues. The FlaA N- and C-termini also have higher proportions of hydrophobic and charged residues at the expense of polar residues and higher non-polar surface areas. Thus, a predominant stabilizing factor in FlaA appears to be increased hydrophobicity, which often confers greater rigidity to proteins. Fewer intramolecular ion pairs were observed in FlaA than FliC, although an increase in the positive charge potential of the FlaA D0 and D1 domains was also observed; increased intermolecular salt bridges may also contribute to the thermal stability of the oligomeric flagellar fiber.

V. R. R. Malapaka
Department of Biological Sciences,
Western Michigan University
Mailstop 5410, 1903 West Michigan Avenue,
Kalamazoo, MI 49008-5410, USA

B. C. Tripp (✉)
Departments of Biological Sciences and Chemistry,
Western Michigan University
Mailstop 5410, 1903 West Michigan Avenue,
Kalamazoo, MI 49008-5410, USA
e-mail: brian.tripp@wmich.edu
Tel.: +1-269-3874166
Fax: +1-269-3875609

Keywords Hyperthermophile · Flagellin · *Aquifex* · Self-assembly · MODELLER

Introduction

The flagella is the primary organelle responsible for motility in prokaryotes [1–5]. It is a complex structure composed of a basal body, a trans-membrane rotary motor [3, 6–8], a universal joint-like hook structure and a hollow helical flagellar filament. This filament has an outer diameter of 12–25 nm, a 2–3 nm diameter inner channel [9, 10] and can be 1–15 μm or more in length [8, 11]. Flagella are rotated by the membrane rotary motor to enable movement of the bacterium in a process that is often regulated by chemotaxis [1]. Peritrichously arranged flagella either associate into a helical bundle of multiple flagella [12] that generates a net thrust along its axis or they separate into separate fibers, depending on the direction of rotation and resulting supercoiled state of the fiber.

Salmonella typhimurium is a bacterial pathogen that also frequently serves as a model organism for the investigation of bacterial flagellar structure and function. The flagellar fibers in *S. typhimurium* are primarily composed of up to 20,000 copies of a single globular protein called flagellin, which self-assembles *via* noncovalent forces to form a helical fiber composed of 11 protofilaments. The primary (phase-1) flagellin gene in *S. typhimurium* is designated as *fliC*, a secondary (phase-2) flagellin gene is named *fljB*; other bacterial flagellin genes are sometimes named as *flaA, hag* or *flaF*. Flagellin proteins are synthesized in the cytoplasm as soluble monomers and are prevented from self-assembly and protected from proteolytic degradation by binding of the FliS chaperone protein [13]. Flagellin is exported by a flagella-specific type III secretion system through the central pore of the flagellar fiber in a partially unfolded conformation, followed by complete folding and oligomer assembly at the distal end of the fiber [9, 10, 14, 15]. This terminal assembly process is aided by the flagellar chaperone cap protein, termed FliD or HAP2 [9, 10, 16–19].

Aquifex pyrophilus is a motile, hyperthermophilic, microaerophilic, chemolithoautotrophic, Gram-negative, rod-shaped bacterium that was originally isolated from the hydrothermal system at the Kolbeinsey Ridge north of Iceland [20]. It grows at temperatures between 67–95°C, with an optimum temperature of 85°C. Although it is an extremophile, *A. pyrophilus* is a eubacterial microbial species, not an archaeal species, and has been placed in the class Aquificae. Like many other species of motile eubacteria, *A. pyrophilus* is flagellated, with a polytrichous arrangement of up to eight flagella fibers on the cell surface, as previously described by Behammer et al. [21]. The diameter of *A. pyrophilus* flagella is 19 nm, which conforms to the general morphology of the mesophilic bacterial flagella [21]. *A. pyrophilus* flagella are primarily composed of a single flagellin protein that is encoded by the *flaA* gene [21]. The corresponding FlaA protein has 501 residues (53.9 kDa), with the first N-terminal Met residue removed post-translationally [21]. The sequence length of FlaA is nearly identical to that of some mesophilic bacteria, such as the 495-residue *S. typhimurium* FliC protein (51.4 kDa). However, *A. pyrophilus* flagella are unusual in that they are functional, i.e., remain self-assembled into fibers, at temperatures higher than 100°C and at low pH values [21]. This functional temperature range is much higher than that typically observed for the flagella of well-characterized mesophilic bacteria such as *S. typhimurium* and *Escherichia coli*, which typically dissociate at temperatures higher than 60–65°C [22, 23], a property often exploited in their purification for biophysical studies [19]. The *A. pyrophilus* flagellar fibers are prone to breakage during low-speed centrifugation, unlike the flagella of *S. typhimurium* and thus are more rigid than their mesophilic counterparts at room temperature [21]. Henceforth, we refer to the *S. typhimurium* flagellin protein as FliC and the *A. pyrophilus* flagellin protein as FlaA.

The partial and complete structures of mesophilic *S. typhimurium* FliC were determined in 2001 [11] and 2003 [24]. The complete FliC structure (PDB 1UCU) shows four distinct globular domains in the protein, termed D0, D1, D2 (with subdomains D2a and D2b) and D3. Domain D0 forms the inner core of the filament; D1 forms the outer core and domains D2 and D3 form a knob-like projection on the filament surface. Various sequence-analysis studies have shown that the N- and C-termini of FliC that form the D0 and D1 domains are essential for the export and assembly into flagella and are conserved across most bacterial species [25]. In contrast, the D2 and D3 domains in the middle are highly variable in sequence and size across different species. Flagellin molecular masses can range from 20–77 kDa [26–30], corresponding to amino-acid sequences ranging in size from under 300 amino acids to almost 700 amino acids (e.g., *Pseudomonas putida* has 688 residues), with the majority of flagellins containing ~500 amino acids [31]. These size differences are primarily due to high genetic variation in a "hypervariable" middle region of the protein encoding outer domains D2 and D3. These hypervariable domain regions can be deleted to a large extent while not affecting the self-assembly properties of the inner D0 and D1 domains [32].

The mechanisms of FlaA flagellar thermostability are fundamentally interesting, as this is a self-assembling structural protein, rather than a thermostable enzyme. Thermostable flagellin protein fibers based on *Aquifex* sp. flagellins may also have potential applications in biotechnology for the extracellular display of thermostable proteins and peptides with sensor or catalytic activity and in nanotechnology as templates for biomineralization and as bionanotubes [33] for the assembly of novel hybrid nanostructures. Previous reports have demonstrated the utility of genetically inserted fusion peptides and small proteins displayed on the immunologically reactive, solvent accessible, middle domain of mesophilic *E. coli* flagellin [34–37]. Thus, it is possible that FlaA could be adapted for use as similar peptide and protein display system for use in applications that require high temperature and chemical stability.

The complete amino-acid sequence of the thermostable *A. pyrophilus* FlaA flagellin is available in the Swiss-Prot database. It has a reasonable overall sequence homology of ~30% with the structurally characterized *Salmonella* FliC flagellin protein. Many hyperthermophilic and mesophilic enzymes and proteins typically have very similar secondary and tertiary structures and functional mechanisms [38]. Thus, the mesophilic FliC structure represents a possible folded structural template for the corresponding thermostable FlaA protein. These observations suggested that it would be possible to develop a predictive structural model of the *A. pyrophilus* FlaA protein using standard computational homology modeling approaches. This 3D structural model could then be analyzed with respect to its structural determinants of thermostability and might also serve as a basis for further protein engineering of thermostable flagella fibers for use in peptide display and other nanomaterial applications. Thus, we have used well-established comparative modeling software including MODELLER [39] to generate a hypothetical 3D structure for the FlaA protein and standard model evaluation software such as PROCHECK [40] to validate the theoretical model.

Computational methods

The amino acid sequences of *A. pyrophilus* flagellin (FlaA: P46210) and *S. typhimurium* phase-1 flagellin (FliC: P06179) were obtained from the Swiss-Prot [41–45] database of the *Swiss Institute of Bioinformatics* (SIB). The refined crystal structure coordinates of *S. typhimurium* flagellin FliC (PDB file 1UCU) were obtained from the RCBS Protein Data Bank [46, 47]. The ProtParam [48, 49] tool at the **Ex**pert **P**rotein **A**nalysis **Sy**stem (ExPASy) Proteomics Server [50] was used to compute various physical and chemical parameters for these two proteins from their 3D coordinate files and amino-acid sequences. The amino-acid composition of the proteins, aliphatic index and grand average of hydropathicity (GRAVY) were of

particular interest. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (Ala, Val, Ile and Leu). It may be regarded as a positive factor for the increase of thermostability of globular proteins. The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values [51] of all the amino acids, divided by the number of residues in the sequence. We analyzed the overall amino-acid flexibility of the proteins using the Average Flexibility Calculator [52] option in the ProtScale tool [48] at the ExPASy Proteomics Server. We also analyzed the propensity of residues in the proteins to be disordered using the DisEMBL [53] protein disorder prediction tool and the GlobPlot [54] protein disorder/order/globularity/domain predictor tool at the ExPASy Proteomics Server. The homology models for the human *A. pyrophilus* were generated using the computer program MODELLER (6v1) [39, 55, 56] of the Accelrys (San Diego, CA) InsightII software package. The input to the program is an alignment of the target sequence with the related three-dimensional structure of FliC. The InsightII visualization environment was used to analyze the structures and perform other structure-related operations. The **S**olvent **A**ccessible **S**urface **A**rea (SASA) [57] was measured using the Solvation module of InsightII. The Delphi module of InsightII was used to investigate the electrostatic charge distribution [58, 59] on the surfaces of the proteins. The protein model was validated using PROCHECK [40]. All Accelrys modeling software was run on an IBM IntelliStation M Pro PC workstation with the Red Hat Enterprise Linux WS 3 operating system. The multimer model of FlaA was constructed using the coordinates provided by Dr. Koji Yonekura (UCSF, San Francisco) and Dr. Keiichi Namba (Osaka University). The helical parameters are l=66 *n*+361 m and the repeat distance is 1,698.8 Å. The FlaA monomers were aligned with the FliC subunits using the Magic Fit option in Swiss-PdbViewer [60, 61]. We have calculated the molecular surfaces of each subunit to visualize the surface interactions.

The contribution of electrostatic pairing of ionized side chains, i.e., salt bridges/ion pairs, to the stability of FlaA was evaluated using the criteria that a salt bridge is present if two oppositely charged atoms of each neighboring side chain are closer than 6 Å. The residues Arg, Lys, His, Asp and Glu were all considered in the salt-bridge calculations [38]. Another school of thought is that proteins gain electrostatic stabilization by minimizing the number of repulsive contacts between like charged residues rather than by creating salt bridges [62]. Thus, we have also investigated the occurrence of like charged groups (+, + and −, −) within a 6 Å distance and compared this parameter for both proteins. Protein α-helices have a net dipole moment resulting from the alignment of the peptide backbone hydrogen bonds. It has been observed that negatively charged (Asp and Glu) and positively charged (Arg, Lys and His) residues are found preferentially at the N- and C-terminal ends of α-helices [63–66]. This charge helps to counteract and stabilize the charge dipole of an α-helix, resulting in increased stability of the folded structure. The N-capping box (Ncap) is a local motif that acts as a stop signal and gives stability to the folded protein. A similar Ccap motif was also identified at the C-terminus of α-helices [67]. The side chain of the residue at N3/C3 is hydrogen bonded to the amide group of Ncap/Ccap and the side chain of Ncap/Ccap is hydrogen bonded to the amide group of N3/C3. Possible N-capping boxes are characterized by the presence of any one of the following residues at both the Ncap and N3 positions: Ser, Thr, Asp, Asn, His, Glu or Gln. Similarly, C-capping boxes are characterized by the presence of any one of the following residues at both the Ccap and C3 positions: Ser, Thr, Asp, Asn, His, Glu or Gln. The cap regions where the above-mentioned residues are present but no hydrogen bonding was observed were designated as "potential" capping boxes. The FlaA model α-helices were evaluated for possible charge–α-helix dipole interactions in terms of the number of Asp and Glu residues at the Ncap, N1, N2 and N3 positions and the number of Arg, Lys and His residues at the C3, C2, C1 and Ccap positions.

Results and discussion

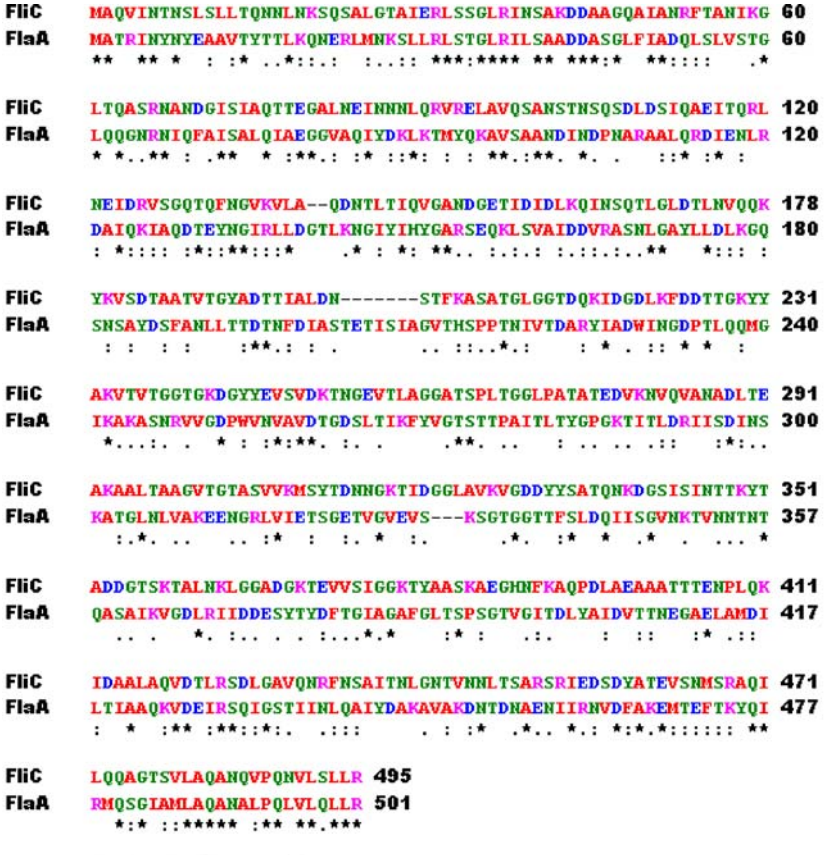
The properties of thermophilic proteins have been examined extensively over the past two decades. Many structural determinants for thermostable proteins have been postulated, as reviewed by Petsko [68], Vieille and Zeikus [38], Scandurra et al. [69, 70], Fields [71] and Sterner and Liebl [72]. Sequence analysis of hyperthermophilic bacterial genomes has detected some preferences of thermophilic proteins for particular amino acids but general design principles are not fully defined [70, 73]. Factors that have been reported to increase thermal stability of proteins include tighter internal packing of hydrophobic residues [69], increased structural rigidity [38, 71], increased numbers of hydrophobic residues with branched side chains [74], additional prolines [75], fewer glycines [76], deletion of flexible surface loops, fewer Asn and Gln residues [75, 77], fewer histidine residues [69] and greater numbers of charged residues and consequently, increased hydrogen bonds and salt bridges (ion pairs) on the protein surface, at the expense of uncharged polar residues [62, 69, 74, 78–80]. Other forms of electrostatic stabilization forces include increased stabilization of α-helices by intrahelical salt bridges, an increase in negative charge at the N-terminus (helix dipole stabilization) [80] and increased cation-π interactions [80]. Protein oligomerization has also been noted as a factor in increased protein stability [81, 82], in part due to decreased exposed surface area. Confinement of proteins in a small inert volume, i.e., an "Anfinson cage", is also known to stabilize the folded state of globular proteins [83]. Flagella, which are directly exposed to the external solvent environment, are a rather unique and ideal system to investigate several aspects of protein thermostability, in the context of a self-assembling, oligomeric structural system, rather than a catalytic enzyme system.

We have systematically compared the sequences and structures of FliC and FlaA to elucidate the factors responsible for the thermostability of FlaA.

Sequence comparison

Sequence alignment for modeling was done using the pairwise-alignment tool of InsightII. The alignment of the protein shows a 29.1% overall sequence similarity between the thermophilic and mesophilic flagellins, FlaA and FliC (Fig. 1).

The protein sequences were divided into three separate regions, based on their similarity and structure, for purposes of further sequence analysis: the N-terminus, the middle domain and the C-terminus. The FlaA N-terminus includes amino acid residues 1–180, the FlaA middle domain includes residues 181–398 and the FlaA C-terminus includes residues 399–500. The FliC protein was similarly divided into an N-terminal region of residues 1–175, a middle domain composed of residues 176–394 and a C-terminal region comprising residues 395–494. Both the N- and C-terminal regions have a higher sequence identity that the overall proteins; the FlaA and FliC N-termini have 54% similarity whereas the corresponding C-termini have 55% similarity. Conversely, the middle domain hypervariable regions have only 16% similarity. This is in contrast to an earlier estimate of 27% similarity for 246 residues, starting at residue 173, by Behammer et al. [21]. The highly conserved N- and C-terminal regions are essential for both recognition by the type III export apparatus and subsequent self-assembly into the flagellar fiber. Sequence analysis studies have shown that the N- and C-termini of flagellins are conserved across most bacterial species [25]. This sequence conservation also confers functional conservation; flagellins from different mesophilic bacterial species can be recombined into heterogeneous flagella in vitro [8, 84–86], due to complimentary interactions between their highly conserved N- and C-terminal domains. Thus, it is not surprising that a relatively high degree of similarity (~55%) is observed for the N- and C-terminal domains of FlaA and FliC. However, it remains to be determined experimentally if *Aquifex* flagellin will interact in a functional manner with mesophilic flagellins, mesophilic



Amino Acid	Color	Property
AVFPMILW	RED	Small (small + hydrophobic)
DE	BLUE	Acidic
RHK	MAGENTA	Basic
STYHCNGQ	GREEN	Hydroxyl + Amine + Basic - Q

* Indicates identical residues
: Indicates conserved substitutions
. Indicates semi-conserved substitutions

Fig. 1 A pairwise sequence alignment of FliC and FlaA. The overall sequence identity of the alignment was 29.1%. The sequence identity at the N- and C-termini was much greater (~55% identity) than the middle domain (16% identity)

FliS chaperone proteins and can be expressed and self-assembled into flagella in mesophilic *E. coli* or *Salmonella* bacteria.

The original investigation of *A. pyrophilus* flagella and FlaA sequence by Behammer et al. in 1995 [21] suggested some possible sequence and structural determinants for its thermostability. These authors noted increased levels of hydrophobic residues, including aromatic residues and proline in several heat-stable flagellins, with a corresponding decrease in polar hydrophilic residues. This increase in average hydrophobicity is frequently accompanied by a decrease in average chain flexibility at lower temperatures [87, 88], as possibly indicated by fragility of the thermostable FlaA flagellar filaments at room temperature. They also proposed that thermostable flagellins form compact monomer structures and large interfaces between subunits (excluding water molecules) in the helical polymer.

Comparison of the amino-acid compositions using the ProtParam tool was performed for the complete FlaA and FliC polypeptides, their more conserved N-termini and C-termini and their hypervariable middle domains. Residues were classified as hydrophobic: {Ala(A), Val(V), Ile(I), Leu(L), Met (M), Phe(F), Tyr(Y), Trp(W) and Pro(P)}; polar: {Asn(N), Gln(Q), Ser(S), Thr(T) and Cys(C)}; charged: {Arg(R), Lys (K), His(H), Asp(D) and Glu(E)}; and aromatic: {(Phe(F), Trp (W) and Tyr(Y)}.

Comparison of the FlaA and FliC residue proportions (percentages) for the total proteins show a +7.0% increase in the total hydrophobic residues and a decrease of −6.0% and a +0.6% increase in the polar and charged residues in FlaA (Table 1). The FlaA aromatic residues increased by 2.4%. According to one analysis, more charged residues (+3.2%), fewer polar residues (−5.0%) and slightly more hydrophobic residues are found in hyperthermophilic proteins, on average [38]. FlaA seems to be an exception to this rule as there is only a minor increase of +0.6% in the number of charged residues and a very high increase of +7.0% in the hydrophobic residues. The greatest contribution to the increase of hydrophobic residues in FlaA is from the branched hydrophobic residue, isoleucine (+4.5%).

A number of individual amino-acid compositions were analyzed in further detail. Isoleucine has greater thermostability at temperatures above 100°C [38]. The relative percentages of Ala, Asn, Asp, Gln, Lys, Ser, Thr and Val decreased in FlaA while those of Arg, Glu, His, Met and Pro increased, which conforms to the general trend observed among hyperthermophilic proteins. Several properties of Arg residues suggest that they are better suited to functioning at higher temperatures than Lys residues. The Arg δ-guanido moiety has a reduced chemical reactivity due to its high pK_a and significant resonance stabilization [38]. Arg also forms more stable ion-pair interactions at high temperatures [38]. The Arg/Lys ratio in FlaA is 0.9 and in FliC is 0.5. Hyperthermophilic proteins can have an increase in the total number of Pro residues in loop regions; Pro decreases the conformational entropy by conferring rigidity to loops. Thermophilic proteins can also have a decreased proportion of Gly residues because Gly has an opposite effect of Pro in terms of increasing conformational entropy of the polypeptide [38]. There are nine Pro residues in FlaA and five Pro residues in FliC, an increase of +0.8%. Most of the Pro residues in FlaA are located in the loop regions of the D2 and D3 domains except for residue Pro106, which is located in the α-helix of the D1 domain. There is also a −1.3% decrease in the Gly residue percentage in FlaA. The observed increase of Pro residues and decrease of Gly residues help to explain the thermostability

Table 1 Amino acid composition (percentages) of the full length *Aquifex pyrophilus* FlaA and *Salmonella typhimurium* FliC flagellin sequences and their middle domains

Amino acid	FlaA[a] (%)	FliC[b] (%)	FlaA–FliC[c] (%)	FlaA middle domain[d] (%)	FliC middle domain[e] (%)	FlaA–FliC middle domain[f] (%)
Ala (A)	10.8	12.3	−1.5	7.3	12.8	−5.5
Arg (R)	4.0	2.8	1.2	2.3	0.0	2.3
Asn (N)	6.8	8.5	−1.7	6.4	4.6	1.8
Asp (D)	7.4	7.5	−0.1	7.3	9.1	−1.8
Cys (C)	0.0	0.0	0.0	0.0	0.0	0.0
Gln (Q)	5.0	6.5	−1.5	1.8	2.3	−0.5
Glu (E)	3.8	3.4	0.4	3.2	2.7	0.5
Gly (G)	7.4	8.7	−1.3	10.1	13.2	−3.1
His (H)	0.4	0.2	0.2	0.5	0.5	0.0
Ile (I)	9.6	5.1	4.5	9.6	2.7	6.9
Leu (L)	9.0	8.5	0.5	5.5	5.0	0.5
Lys (K)	4.6	5.7	−1.1	4.1	10.0	−5.9
Met (M)	1.6	0.4	1.2	0.5	0.5	0.0
Phe (F)	2.0	1.2	0.8	2.8	1.4	1.4
Pro (P)	1.8	1.0	0.8	3.2	1.4	1.8
Ser (S)	6.8	7.7	−0.9	9.2	5.9	3.3
Thr (T)	9.6	11.5	−1.9	14.7	15.1	−0.4
Trp (W)	0.4	0.0	0.4	0.9	0.0	0.9
Tyr (Y)	3.6	2.4	1.2	2.8	5.0	−2.2
Val (V)	5.6	6.5	0.9	7.8	7.8	0.0

[a]The *Aquifex pyrophilus* FlaA flagellin protein consists of 500 amino acids
[b]The *Salmonella typhimurium* FliC flagellin protein consists of 494 amino acids
[c]Percentage difference in amino acid composition between FlaA and FliC for each individual amino acid
[d]The FlaA middle domain is composed of residues 176–394
[e]The FliC middle domain is composed of residues 181–398
[f]Percentage difference in amino acid composition between the middle domains of FlaA and FliC for each individual amino acid

and rigidity of the protein. Apart from the Gly and Pro composition, the flexibility of a protein also depends on its primary structure. By using the ProtScale tool of ExPASy, we have calculated the overall amino-acid flexibilities of FliC and FlaA and for their N- and C-termini and middle domains. The entire protein, N- terminus, C-terminus and middle region (D2 and D3 domains) have average flexibility values of 0.447, 0.446, 0.438 and 0.452 for FlaA and 0.450, 0.455, 0.444 and 0.450 for FliC. These values indicate a marginally higher relative flexibility for the entire FliC protein (+0.67%), N-terminal and middle domains (+2.0 and +1.4%) and a slightly lower flexibility for the C-terminal domain (−0.44%). DisEMBL predicted that FliC has 6% more residues than FlaA that may be disordered, whereas GlobPlot predicted a 3% increase in the disordered residues in FliC. Thus, all three algorithms yielded a consistent prediction of a slightly higher flexibility for mesophilic FliC than for thermostable FlaA, although the differences were statistically minor. Both FlaA and FliC proteins lack Cys; Cys is not observed in any wild-type flagellin sequences identified to date.

The overall aliphatic indexes of FlaA and FliC proteins with their N-terminal Met residues removed are 99.6 and 84.0 (Table 2). These aliphatic-index values indicate a higher overall aliphatic character for FlaA than FliC, a common sequence feature of thermostable proteins. The GRAVY values are −0.121 for FlaA and −0.400 for FliC (Table 2), which strongly indicates a much higher hydrophobic character for the FlaA sequence. Overall, both FlaA and FliC have very similar numbers of charged residues (Table 2), with 56 and 54 negative residues and 43 and 42 positive residues, respectively. However, the distribution of charged residues was very different in these two proteins, as discussed below.

The sequence compositions of FlaA and FliC were further analyzed at three different regions, as discussed above. The aliphatic indexes of the N-terminal regions of FlaA and FliC are 105.4 and 99.3 and the corresponding GRAVY values are −0.266 and −0.428 (Table 2). The FlaA N-terminus has an increase of +9.7% in hydrophobic residues, a decrease of −0.4% in charged residues and a decrease of −10.2% in polar residues (Table 3). The increase in charged residues was only due to an increase of seven additional positive residues (20 vs. 13, Table 2), as the numbers of negative residues were similar for both proteins, 19 vs. 18. The aliphatic index values of the FlaA and FliC C-terminal regions are 111.9 and 97.7 and the corresponding GRAVY values are −0.002 and −0.269 (Table 2). The FlaA C-terminal region has an increase of +7.6% in hydrophobic residues and a +5.4% increase in charged residues, relative to FliC (Table 3). The polar residues correspondingly decreased by −12.7% in FlaA. In contrast with the N-terminal region, 40% more negative residues were found in the C-terminal region of FlaA (14 vs. 10, Table 2), although two additional positive residues were also observed (9 vs. 7). The extensive increase in hydrophobic residues in the C- and N-terminal regions of FlaA leads to increased hydrophobicity and should contribute to its higher thermal stability. Thus, in spite of high sequence similarity between the FliC and FlaA N- and C-terminal regions, significant differences are apparent in the proportions of hydrophobic and polar amino acids. The structural implications of these differences are discussed in the next section.

The hypervariable middle domains of FlaA and FliC have aliphatic index values of 89.0 and 65.6, while the corresponding GRAVY values are −0.061 and −0.437 (Table 2). The individual amino-acid compositions for the middle domains

Table 2 Computed parameters of FlaA and FliC flagellin proteins and their N-terminal, C-terminal and middle hypervariable domain regions

Protein/domain[a]	Aliphatic index	GRAVY values[b]	pI[c]	Negative residues[d]	Positive residues[e]
FlaA protein	99.6	−0.121	4.82	56	43
FliC protein	84.0	−0.400	4.79	54	42
FlaA N-terminus[f]	105.4	−0.266	8.12	19	20
FliC N-terminus[f]	99.3	−0.428	4.65	18	13
FlaA C-terminus[g]	111.9	0.002	4.50	14	9
FliC C-terminus[g]	97.7	−0.269	4.61	10	7
FlaA middle domain[h]	89.0	−0.061	4.46	23	14
FliC middle domain[h]	65.6	−0.437	5.04	26	22

[a]All sequence analysis calculations were performed using the ProtParam tool of the ExPASy proteomics server, using the Swiss-Prot database files P46210 for *Aquifex pyrophilus* FlaA flagellin and P06179 for *Salmonella typhimurium* FliC phase-1 flagellin, with the N-terminal Met residue excluded from the calculations for both proteins

[b]Grand average of hydropathicity (GRAVY) values

[c]The theoretical isoelectric point of each protein or domain region

[d]Total number of negatively charged residues (Asp + Glu)

[e]Total number of positively charged residues (Arg + Lys). FliC has 1 His residue and FlaA has 2 His residues that were not included in this calculation

[f]The N-terminal domain regions were defined as residues 1–180 out of 500 total amino acid residues for FlaA and residues 1–175 out of 494 for FliC

[g]The C-terminal domain regions were defined as residues 399–500 out of 500 for FlaA and residues 395–494 out of 494 for FliC

[h]The FlaA hypervariable middle region encompassing domains D2 and D3 is defined as residues 181–398 out of 500 total residues. The FliC middle hypervariable region is defined as residues 176–394 out of 494 total residues

Table 3 Amino acid composition (percentages) of N-terminus, C-terminus and middle domains of FliC and FlaA

Protein/domain	Hydrophobic residues (%)[a]	Polar residues (%)[b]	Charged residues (%)[c]
FlaA N-terminus[d]	45.2	30.7	17.4
FliC N-terminus[d]	35.5	40.9	17.8
FlaA C-terminus[e]	50.6	24.3	22.4
FliC C-terminus[e]	43.0	37.0	17.0
FlaA middle domain[f]	40.4	32.1	17.4
FliC middle domain[f]	36.6	27.9	22.3

[a]The hydrophobic amino acids include Ala(A), Val(V), Ile(I), Leu (L), Met(M), Phe(F), Tyr(Y), Trp(W) and Pro(P)
[b]The polar residues include Asn(N), Gln(Q), Ser(S), Thr(T) and Cys(C)
[c]The charged residues include Arg(R), Lys(K), His(H), Asp(D) and Glu(E)
[d]The N-terminal domain regions were defined as residues 1–180 out of 500 total amino acid residues for FlaA and residues 1–175 out of 494 for FliC
[e]The C-terminal domain regions were defined as residues 399–500 out of 500 for FlaA and residues 395–494 out of 494 for FliC
[f]The FlaA hypervariable middle region encompassing domains D2 and D3 is defined as residues 181–398 out of 500 total residues. The FliC hypervariable middle region is defined as residues 176–394 out of 494 total residues

of FlaA and FliC are given in Table 1. The middle domain of FlaA has an increase of 3.8% in the hydrophobic side chains and a 4.9% decrease in the charged residues. There is an increase of 4.2% in the polar residue composition (Table 3). The FlaA middle domain has significantly fewer positively charged residues (14 vs. 22) and slightly fewer negatively charged residues (23 vs. 26). This indicates that this domain region is more acidic in FlaA and that it is not stabilized by formation of additional salt bridges, but rather by hydrophobic interactions.

Structure comparison

The 3D model for the FlaA protein was built using the coordinate file 1UCU from the Protein Data Bank as a structural basis using the MODELLER software. Altogether, ten models were generated with a high level of loop optimization and structurally aligned with FliC. The best-fit model had the lowest probability density function violations; its free energy and loops were refined further. The secondary structures were visualized using the Kabsch and Sander [89] method in InsightII (Fig. 2a and b). The FlaA secondary structure comprises nine α-helices and 20 β-sheets similar to the FliC structure. The three-dimensional model was verified using PROCHECK. The coordinates of this model have been deposited in the RCSB PDB theoretical model online database, under the ID 1XGX.

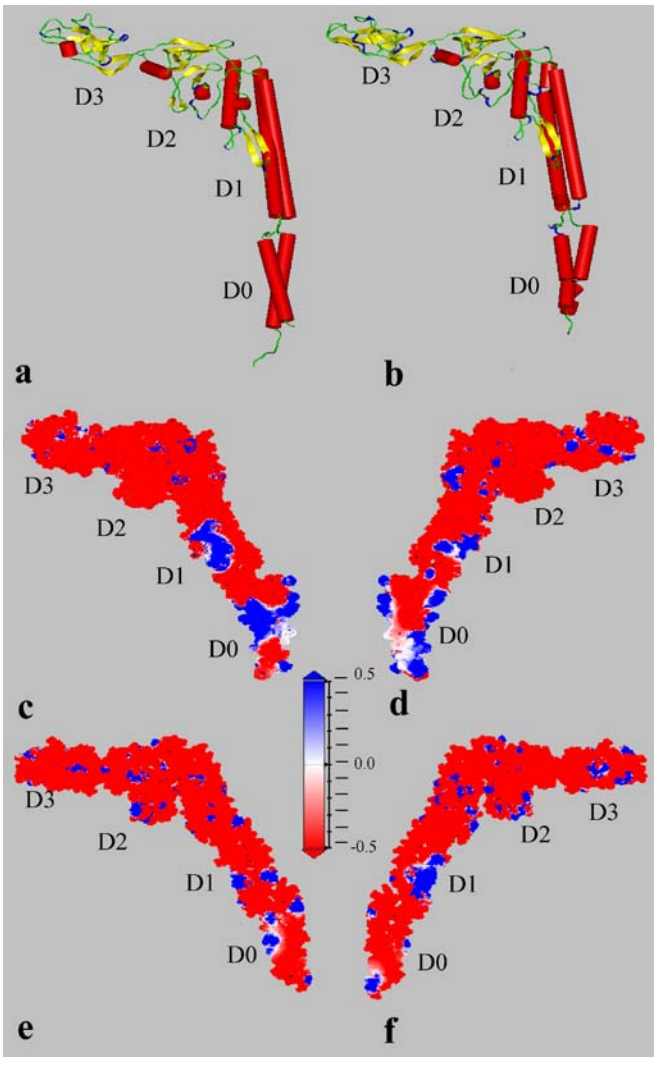


Fig. 2 **a** Model 3D structure of thermostable *A. pyrophilus* FlaA flagellin. **b** 3D structure of mesophilic *S. typhimurium* FliC flagellin. α-helices are shown in *red*, β-sheets are shown in *yellow*, turns are shown in *blue* and coils are shown in *green*. Both the structures show domains D0, D1, D2 and D3. **c, d** Stereo view of surface charge distribution of FlaA. **e, f** Stereo view of surface charge distribution of FliC. The legend shows the scale for coloring of the charge spectrum. *Blue* represents positive charge, *red* represents negative charge and *white* represents neutral

A Ramachandran plot (Fig. 3) shows that 87.2% of the total residues of the FlaA model occupy the most favored regions, 9.7% occupy additional allowed regions, 1.3% occupy generously allowed regions and only 1.1% fall in the disallowed region. Figure 4 shows the 3D-structure alignment of FlaA model and the experimentally determined FliC structure, PDB 1UCU. The average root-mean-square (rms) deviation at C^{α} positions between the FlaA model and PDB 1UCU is 1.9 Å.

The above observations indicate that the modeled FlaA structure is conformationally correct. However, it should be noted that the probability of the middle domain region having an alternate conformation is higher, due to the high variability of sequences observed in this region for bacterial flagellins. We have also modeled the structure

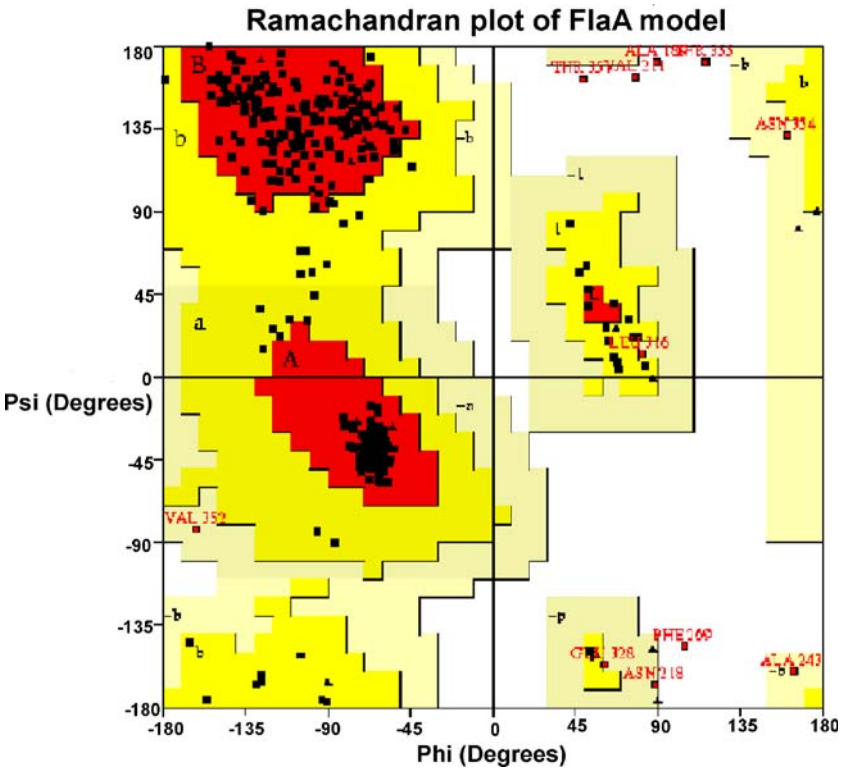


Fig. 3 Ramachandran plot of the FlaA model. Of the FlaA model residues, 87.2% occupy the most favored regions, 9.7% occupy additional allowed regions, 1.3% occupy generously allowed regions and only 1.1% are in the disallowed region

using the SWISS-MODEL server [90] and Rosetta algorithms [91, 92] implemented on the Robetta comparative modeling server [93]. SWISS-MODEL failed to build a comparative model, whereas Robetta gave a 3D-model (data not shown) comparable to the one generated by MODELLER. The gross topologies of the FlaA model and the FliC structure are very similar, as expected for a homology model. The α-helical coiled-coil conformations at the N- and C-termini of the FlaA model were conserved as these regions form the intricate intersubunit interactions in the flagellar filament structure. The C-terminal region of the D0 domain, which forms the inner surface of the channel, consists of mainly polar amino acids, as in the case of the mesophilic homolog, FliC. The polar nature of the surface is thought to help the diffusion of the unfolded monomers, because unfolded proteins with hydrophobic side chains exposed will be trapped on a hydrophobic surface [24]. It is interesting to consider that the C-terminus of FlaA that forms the inner core of the flagellum channel has a higher aliphatic index, 111.9 (Table 2), which may retard monomer movement in the channel. A conserved C-terminal Arg residue is present in both proteins, Arg500 in FlaA and Arg494 in FliC. This residue projects into the flagellar central channel formed by FliC oligomers and was previously hypothesized to have a functional role in transport of the unfolded flagellin monomers through the channel [24].

The proportions of residues in regular secondary structures were determined from the ratio of total number of helical, strand and loop residues to the total number of residues in FliC and FlaA. Loop residues include turns, bends and other irregular secondary structures. The proportion of residues in helical, strand and loop regions are 38.5, 19.6 and 41.9% for FliC compared to 42.5, 18.2 and 39.3% in FlaA. Although both proteins have the same number of secondary structures, the proportion of residues falling in the secondary structures in FlaA is greater, especially in the α-helical region. There is a 4.0% increase in the α-helical residues, a decrease of 1.5% in the β-strand residues and a 2.6% decrease in the loop residues in the FlaA model. The predicted increase of α-helical secondary-structure content and decrease in the loop residues in the FlaA proteins should contribute to the thermal stability [80].

Several studies have demonstrated that β-branched residues are α-helix-destabilizing due to their reduced conformational freedom in α-helices [38]. Thus, the proportions of β-branched residues Ile, Val and Thr were analyzed in the FlaA α-helices. We found results contrary to the general trend as there is an overall increase in the proportion of the β-branched residues in the FlaA α-helices. The overall percentages of Ile, Val and Thr in the α-helices are 13.0, 4.8 and 7.9%, with a total content of 25.7%. The corresponding Ile, Val and Thr percentages in FliC α-helices are 4.0, 5.2 and 6.1%, for a total content of 15.3%. Thus, there is a total of +10.6% increase in the β-branched residues in α-helices in FlaA. This suggests that the extensive increase in the hydrophobic β-branched residues leads to better packing of the α-helices, resulting in a more extensive hydrophobic environment in the coiled-coil structural motif and more stability in the D0 and D1 domains.

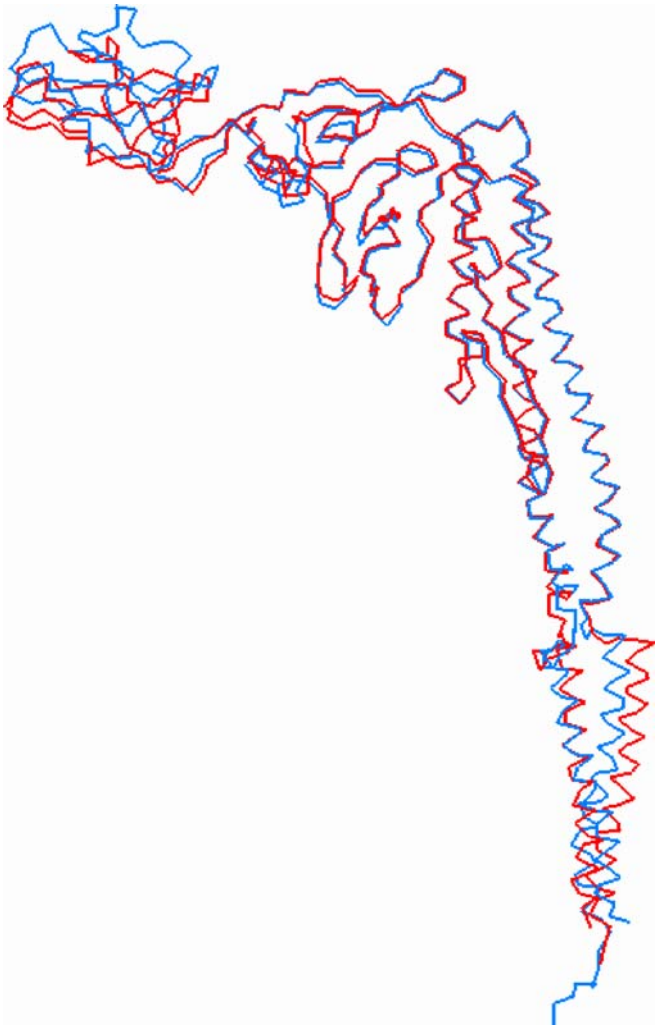


Fig. 4 3D structural alignment of FliC with the model of FlaA. The FliC structure is represented as a *red chain* and the FlaA model as a *blue chain*. The average root-mean-square (rms) deviation at C^{α} positions between the FlaA model and the FliC structure (PDB 1UCU) is 1.9 Å

In contrast with the highly conserved N- and C-terminal regions, the middle D2 and D3 domains of FlaA have relatively low homology with the same region in FliC. Given that the numbers of amino acids in each protein are similar and minimally acceptable sequence homology is apparent, the FlaA model was built with secondary structures and domain folds for the hypervariable D2 and D3 middle domains that are similar to the FliC structure (Fig. 2a). These conserved secondary structures include the three β-folium secondary structures observed for residues 220–260 in FliC domain D3; residues 308–345 in FliC domain D2a and residues 345–383 in FliC domain D2b [11]. The FlaA model shows β-folium structures at the same residues in the D2a, D2b and D3 domains.

Another conserved structural feature is the β-hairpin structure in the conserved N-terminal D1 domain region of FlaA that extends from residues 140–160, with beta strands from residues 141–146 and 154–159 and a β-turn in residues 151–152. This secondary structure motif was hypothesized to function as a two-state switch that abruptly changes conformation and as a result, changes the packing distance between monomers and alters the supercoiled state of the flagellar fiber. This conformational change occurs in response to a change in the direction of flagellar rotation and is ultimately responsible for the ability of motile bacteria such as *S. typhimurium* to change their swimming mode between running (moving) and tumbling (stationary) modes [11]. The previous study by Behammer et al. did not present any data that *A. pyrophilus* bacteria can alternate their direction of flagellar rotation, with resultant differences in their flagellar helical supercoiling and monomer spacing. However, it is not unreasonable to postulate that these thermophilic bacteria also use the same means of reversing flagellar rotation to either swim or remain stationary by forming or disassembling flagella into larger bundles, to enable sampling of temperature and chemical gradients over time, as observed for mesophilic *S. typhimurium* bacteria.

As discussed by Honda et al. [94], the secondary-structural elements in flagellin are well segregated radially, with inner domains that are rich in α-helices and outer domains that are rich in β-strands. Given the extremely harsh extracellular environment in which *Aquifex* flagella function, e.g., high temperature and low pH, it is not unreasonable to assume that the same secondary structural features that stabilize mesophilic FliC against proteases and denaturing chemicals could also be adapted for increased thermal stability in FlaA. Thus, it is also not unreasonable for the FlaA outer D2 and D3 domains to be largely composed of β-strands, as observed in the analogous FliC protein. These structures can be relatively rigid because of cooperative hydrogen-bonding networks. When combined with increased residue hydrophobicity and resulting tighter packing of the hydrophobic core regions, they can also be more stable against proteolysis due to minimal size β-turn loop regions that do not present themselves as flexible substrates for proteases.

A separate analysis of the secondary structure of FlaA indicates that the secondary structures in the FlaA model, including those in the hypervariable middle domain, are largely consistent with standard secondary-structure prediction algorithms (e.g., PSIPRED). Thus, in addition to having consistent structural similarity with FliC, the FlaA structure does not appear to violate any of the secondary-structure predictions that were independently derived by standard prediction algorithms. However, the relatively low degree of sequence homology (16%) observed for the FlaA and FliC hypervariable middle domains suggests that alternative conformations are possible for the D2 and D3 domains in the FlaA model. The contribution of this middle domain to the thermostability of the flagellar fiber remains to be investigated; this region also has very low sequence homology with that of thermostable *Aquifex aeolicus* flagellin.

The hydrophobic effect is considered to be one of the most important of the various molecular forces that determine the tertiary structure of proteins, and thus their stability [95], and the strength of this effect increases with

increasing temperature. Thus, the greater the magnitude of the hydrophobic effect, the more stable the protein [38]. An inverse correlation has been observed between flexibility of thermostable proteins and their hydrophobicity; they may be significantly flexible at elevated temperatures, allowing proper function, but may be too rigid, i.e. "solid" or "wax-like", to function at typical mesophilic temperatures (10–45°C) [69, 96]. The increased rigidity previously noted for FlaA flagella filaments at mesophilic temperatures is likely a result of their increased hydrophobic character and resulting increased compactness. Summation of the non-polar solvent accessible surface areas (SASA) of a folded chain yields a measure of the potential hydrophobic effect. The SASA values were computed using the Solvation module of InsightII. SASA values were calculated only for more conserved D0 and D1 domains that include N- and C-terminal regions as the lower sequence homology in the middle region may not give an accurate measurement (Table 4). The total SASA of the FliC D0 and D1 domains that include N- and C-termini is 1,673 Å^2, which includes a polar surface area of 718 Å^2 and a non-polar surface area of 955 Å^2. The FlaA monomer model D0 and D1 domains have a total SASA of 1,917 Å^2 that includes a polar surface area of 638 Å^2 and a non-polar surface area of 1,278 Å^2. Thus, the predicted total surface area of the more conserved regions of FlaA is 15% greater than that of FliC, suggesting that more interfacial area is available for interaction between flagellin subunits in the thermostable helical flagella fiber. There is also a decrease of −11% in the polar surface area and a corresponding large increase of +34% in the non-polar surface area. This is a strong indicator that increased hydrophobic interactions can form between FlaA monomers, resulting in a greater hydrophobic effect upon burial from the aqueous solvent and thus leading to a tighter packing of the protein.

A surface charge spectrum was constructed for each protein using a DelPhi grid to visualize the relative charge distributions on the surface of the proteins (Fig. 2c,d). The D0 and D1 domains of the FlaA model have a higher positive charge distribution on the surface. This is due to the fact that the number of negatively charged residues in the N-terminal D0 and D1 domains is the same for both FlaA and FliC, whereas FlaA has 6 more positively charged residues compared to FliC in these regions. The C-terminal D0 and D1 domains of FlaA have 13 negatively charged and nine positively charged residues while FliC has eight negatively charged and seven positively charged residues. The D0 and D1 α-helical coiled-coil domains interact with other subunits and are thought to form inter-subunit interactions and stabilize the fiber [24]. The increase of positively charged residues in the D0 and D1 domains of FlaA could have significance in terms of the increasing the stability of the fiber by forming inter-subunit salt-bridges.

Yonekura et al. [24] have shown that most of the intersubunit interactions found in the outer tube of *S. typhimurium* flagellum filament are polar-polar or charge-polar and contributions of hydrophobic interactions are relatively small, whereas those found within the inner tube and between the inner and outer tubes are mostly hydrophobic, contributing to the high stability of the filament structure. A pentamer of the FlaA subunit (Fig. 5) was constructed based on the FliC heptamer coordinates provided by the authors of the complete FliC structure [24], Dr. Koji Yonekura (UCSF, San Francisco) and Dr. Keiichi Namba (Osaka University). The multimer was visualized in Swiss-PdbViewer [60, 61]. The subunits were numbered according to the scheme shown in the Fig. 5a. The subunit numbered 1 and rendered in blue ribbon is used as a reference for analyzing the intersubunit interactions. Subunits 2, 3, 4 and 5 form the most extensive interactions along each direction of the major helical array with subunit 1 and are rendered in different ribbon colors. The lateral interactions of subunit 1 with subunits 2 and 3 can be seen in Fig. 5b. As we have noted earlier, there is an increase of hydrophobic and charged residues at the expense of polar residues in the N- and C-terminal regions of FlaA. The predominant lateral interactions between the D0 and D1 domains of subunits 1, 2 and 3 appear to be hydrophobic in nature. Many hydrophobic residues are exposed on the surface of D0 domain and to a lesser extent on D1 domain. Fig. 5c shows the interactions of subunit 1 with subunits 4 and 5 that lie immediately on top and bottom of the subunit in the protofilament. In this case both subunits show a mixture of hydrophobic and polar-charged interactions. Our prediction is that the FlaA subunits in the filaments are tightly packed due to stronger hydrophobic interactions around the D0 and D1 domains that make them rigid at room temperature and stable at higher temperatures.

Salt bridges/ion pairs were defined using the distance criteria that a salt bridge exists if two oppositely charged atoms of adjacent side chains are located closer than 6 Å.

Table 4 Solvent accessible surface area (SASA) calculations for conserved D0 and D1 domains of FlaA and FliC[a]

Protein domain	Total SASA[b] (Å^2)	Polar SASA[c] (Å^2)	Non-polar SASA[d] (Å^2)
FlaA D0 and D1[e]	1,917	638	1,279
FliC D0 and D1[f]	1,673	718	955
FlaA–FliC[g]	244	−80	324

[a]All SASA calculations were performed using the Solvation module of InsightII software (Accelrys)
[b]Total computed solvent accessible surface area for conserved D0 and D1 domains
[c]Computed solvent accessible surface area for polar surface regions of D0 and D1 domains
[d]Computed solvent accessible surface area for non-polar surface regions of D0 and D1 domains
[e]The D0 and D1 domains of FlaA include the more conserved N-terminal residues 1–180 and C-terminal residues 399–500 out of 500 total amino acids
[f] The D0 and D1 domains of FliC include the more conserved N-terminal residues 1–175 and C-terminal residues 395–494 out of 500 total amino acids
[g]Difference between the corresponding SASA values of D0 and D1 domains of FlaA and FliC

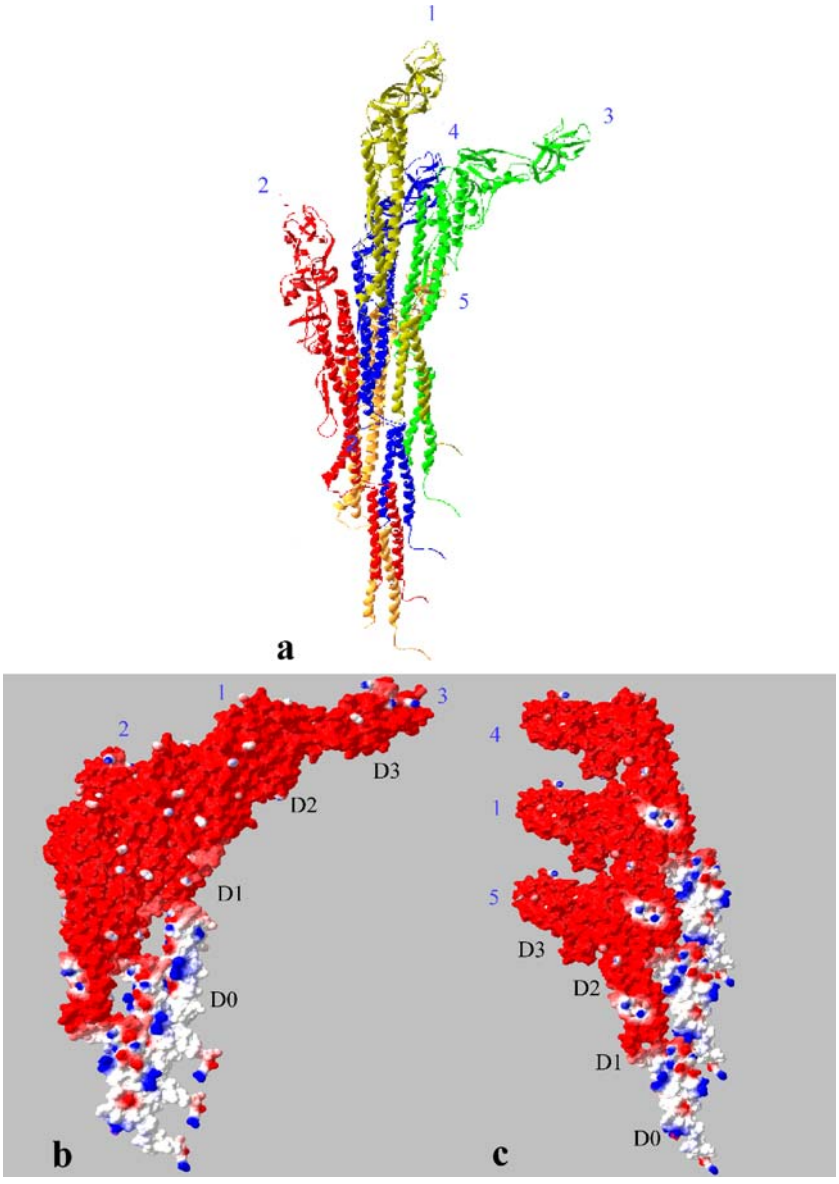


Fig. 5 **a** Stereo diagram of 5 FlaA subunits, viewed from inside of the flagellar filament. The subunit numbered 1 and rendered as a *blue ribbon structure* is used as a reference for analyzing the intersubunit interactions. Subunits 2, 3, 4 and 5 form the most extensive interactions with subunit 1 in the filament and are rendered in *different ribbon colors*. **b** Interaction of subunit 1 with subunits 2 and 3 that form immediate neighbors in the filament. The molecular surface of each subunit is shown. *Blue* represents positive potential, *red*, negative and *white*, zero charge potential. **c** Interaction of subunit 1 with subunits 4 and 5 that lie immediately above and below subunit 1 in the flagellar protofilament. These images were generated with Swiss-PdbViewer

Only the amino acids Arg, Lys, His, Asp and Glu were considered for salt-bridge calculations. Based on these criteria, FlaA has a total of 15 salt bridges, of which 12 are located in the α-helical regions of the D0 and D1 domains. In contrast, FliC has 24 salt bridges, of which nine are in the α-helical regions. These results suggest that intramolecular ion pairs do not play an important role in the increased thermostability of the FlaA monomer, relative to mesophilic FliC, in agreement with the previous results of the SASA calculations of polar surface area. However, these results do not necessarily indicate the importance of intermolecular ion pairs between FlaA monomers in stabilizing FlaA in the polymeric flagellar fiber, as noted above, the D0 and D1 domains have a positively charged region. We have also inspected the presence of like charged groups within the 6 Å cutoff distance visually. FliC has three potentially destabilizing interactions involving like charged groups where FlaA has ten such interactions. This result does not agree with the previously stated theory that minimizing the number of repulsive contacts between like-charged residues will increase the thermal stability of proteins [62].

We inspected the charge dipoles of FliC and FlaA α-helices visually to determine if any of them are stabilized further by the presence of negatively charged (Asp and Glu) and positively charged (Arg, Lys and His) residues at the N- and C-terminal ends. Neither FliC nor FlaA α-helices showed any preferential occurrence of these charged residues. FliC has four potential N-capping motifs and four actual C-capping boxes. FlaA has five potential N-capping boxes and two actual C-capping boxes. Thus, there is little apparent difference in the contribution of helix-dipole stabilization factors in FliC and FlaA.

As previously noted, protein oligomerization is another method for increasing the thermostability of proteins and flagellin functions in vivo by forming extended "20,000-mer" helical fibers. It has only recently been observed that the terminal FliD chaperonin cap complex contributes to

the thermal stability of *S. typhimurium* flagellar filaments [97]. Thus this additional oligomerization factor represents an additional potential stabilizing factor in thermostable *Aquifex sp*. flagella that may not be directly encoded in the FlaA protein itself.

Conclusions

In this work, a 3D-structure of *A. pyrophilus* FlaA protein was modeled based on the crystal structure of *S. typhimurium* FliC with reasonable accuracy. The overall secondary, tertiary and quaternary structure comparisons of FlaA and FliC indicate that electrostatic interactions do not play a major role in FlaA monomer thermostability, in contrast to some other thermophilic proteins. However, this does not preclude increased strength and numbers of intermolecular electrostatic interactions in the oligomeric flagella fiber. The predominant stabilizing force appears to be the hydrophobic effect, which leads to tighter packing of the fiber and its resulting rigidity.

Acknowledgements Financial support from the W. M. Keck Foundation and the Western Michigan University Nanotechnology Research and Computation Center is gratefully acknowledged. We also gratefully acknowledge the gift of the FliC multimer PDB coordinates by Dr. Koji Yonekura and Dr. Keiichi Namba.

References

1. Armitage JP (1992) Sci Prog 76:451–477
2. Bardy SL, Ng SY, Jarrell KF (2003) Microbiology 149:295–304
3. Berry RM, Armitage JP (1999) Adv Microb Physiol 41:291–337
4. Metlina AL (2004) Biochemistry (Mosc) 69:1203–1212
5. Moens S, Vanderleyden J (1996) Crit Rev Microbiol 22:67–100
6. Berg HC (2003) Annu Rev Biochem 72:19–54
7. DeRosier DJ (1998) Cell 93:17–20
8. Jones CJ, Aizawa S (1991) Adv Microb Physiol 32:109–172
9. Macnab RM (2003) Annu Rev Microbiol 57:77–100
10. Macnab RM (2004) Biochim Biophys Acta 1694:207–217
11. Samatey FA, Imada K, Nagashima S, Vonderviszt F, Kumasaka T, Yamamoto M, Namba K (2001) Nature 410:331–337
12. Flores H, Lobaton E, Mendez-Diez S, Tlupova S, Cortez R (2005) Bull Math Biol 67:137–168
13. Ozin AJ, Claret L, Auvray F, Hughes C (2003) FEMS Microbiol Lett 219:219–224
14. Blocker A, Komoriya K, Aizawa S (2003) Proc Natl Acad Sci USA 100:3027–3030
15. Minamino T, Namba K (2004) J Mol Microbiol Biotechnol 7:5–17
16. Ikeda T, Oosawa K, Hotani H (1996) J Mol Biol 259:679–686
17. Maki-Yonekura S, Yonekura K, Namba K (2003) Proc Natl Acad Sci USA 100:15528–15533
18. Maki S, Vonderviszt F, Furukawa Y, Imada K, Namba K (1998) J Mol Biol 277:771–777
19. Furukawa Y, Imada K, Vonderviszt F, Matsunami H, Sano K, Kutsukake K, Namba K (2002) J Mol Biol 318:889–900
20. Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, König H, Rachel R, Rockinger I, Fricke H, Stetter KO (1992) Syst Appl Microbiol 15:340–351
21. Behammer W, Shao Z, Mages W, Rachel R, Stetter KO, Schmitt R (1995) J Bacteriol 177:6630–6637
22. Asakura S (1970) Adv Biophys 1:99–155
23. Vonderviszt F, Kanto S, Aizawa S, Namba K (1989) J Mol Biol 209:127–133
24. Yonekura K, Maki-Yonekura S, Namba K (2003) Nature 424:643–650
25. Murthy KG, Deb A, Goonesekera S, Szabo C, Salzman AL (2004) J Biol Chem 279:5667–5675
26. Kondoh H, Hotani H (1974) Biochim Biophys Acta 336:117–139
27. Joys TM (1985) J Biol Chem 260:15758–15761
28. Lagenaur C, Agabian N (1976) J Bacteriol 128:435–444
29. Joys TM (1988) Can J Microbiol 34:452–458
30. Andrade A, Giron JA, Amhaz JM, Trabulsi LR, Martinez MB (2002) Infect Immun 70:5882–5886
31. Morgan DG, Khan S (2001) Nature ELS:1–8
32. Kuwajima G (1988) J Bacteriol 170:3305–3309
33. Martin CR, Kohli P (2003) Nature Reviews Drug Discovery 2:29–37
34. Lu Z, Murray KS, Van Cleave V, LaVallie ER, Stahl ML, McCoy JM (1995) Biotechnology (NY) 13:366–372
35. Westerlund-Wikstrom B (2000) Int J Med Microbiol 290:223–230
36. Lu Z, Tripp BC, McCoy JM (1998) Methods Mol Biol 87:265–280
37. Tripp BC, Lu Z, Bourque K, Sookdeo H, McCoy JM (2001) Protein Eng 14:367–377
38. Vieille C, Zeikus GJ (2001) Microbiol Mol Biol Rev 65:1–43
39. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
40. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–291
41. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2005) Nucleic Acids Res 33:D154–159
42. Bairoch A, Boeckmann B, Ferro S, Gasteiger E (2004) Brief Bioinform 5:39–55
43. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2004) Nucleic Acids Res 32:D115–D119
44. Apweiler R, Bairoch A, Wu CH (2004) Curr Opin Chem Biol 8:76–80
45. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) Nucleic Acids Res 31:365–370
46. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
47. Deshpande N, Addess KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, Green RK, Flippen-Anderson JL, Westbrook J, Berman HM, Bourne PE (2005) Nucleic Acids Res 33: D233–D237
48. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the expasy server. In: Walker JM (ed) Protein identification and analysis tools on the expasy server. Humana, Totowa, NJ, pp 571–607
49. Gill SC, von Hippel PH (1989) Anal Biochem 182:319–326
50. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) Nucleic Acids Res 31:3784–3788
51. Kyte J, Doolittle RF (1982) J Mol Biol 157:105–132
52. Bhaskaran R, Ponnuswamy PK (1988) Int J Peptide Protein Res 32:241–255
53. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Structure (Camb) 11:1453–1459
54. Linding R, Russell RB, Neduva V, Gibson TJ (2003) Nucleic Acids Res 31:3701–3708
55. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, Madhusudhan MS, Yerkovich B, Sali A (2003) Nucleic Acids Res 31:3375–3380
56. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Annu Rev Biophys Biomol Struct 29:291–325
57. Shrake A, Rupley JA (1973) J Mol Biol 79:351–371
58. Honig B, Nicholls A (1995) Science 268:1144–1149

59. Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, Honig B (2002) J Comput Chem 23:128–137
60. Kaplan W, Littlejohn TG (2001) Brief Bioinform 2:195–197
61. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
62. Karshikoff A, Ladenstein R (2001) Trends Biochem Sci 26: 550–556
63. Ptitsyn OB (1969) J Mol Biol 42:501–510
64. Richardson JS, Richardson DC (1988) Science 240:1648–1652
65. Chou PY, Fasman GD (1978) Adv Enzymol Relat Areas Mol Biol 47:45–148
66. Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL (1987) Nature 326:563–567
67. Harper ET, Rose GD (1993) Biochemistry 32:7605–7609
68. Petsko GA (2001) Methods Enzymol 334:469–478
69. Scandurra R, Consalvi V, Chiaraluce R, Politi L, Engel PC (1998) Biochimie 80:933–941
70. Scandurra R, Consalvi V, Chiaraluce R, Politi L, Engel PC (2000) Front Biosci 5:D787–D795
71. Fields PA (2001) Comp Biochem Physiol A Mol Integr Physiol 129:417–431
72. Sterner R, Liebl W (2001) Crit Rev Biochem Mol Biol 36: 39–106
73. van den Burg B, Eijsink VG (2002) Curr Opin Biotechnol 13:333–337
74. Kumar S, Nussinov R (2001) Cell Mol Life Sci 58:1216–1233
75. Sriprapundh D, Vieille C, Zeikus JG (2000) Protein Eng 13:259–265
76. Panasik N, Brenchley JE, Farber GK (2000) Biochim Biophys Acta 1543:189–201
77. Farias ST, Bonato MC (2003) Genet Mol Res 2:383–393
78. Cambillau C, Claverie JM (2000) J Biol Chem 275:32383–32386
79. Szilagyi A, Zavodszky P (2000) Structure Fold Des 8:493–504
80. Chakravarty S, Varadarajan R (2002) Biochemistry 41:8152–8161
81. Dams T, Auerbach G, Bader G, Jacob U, Ploom T, Huber R, Jaenicke R (2000) J Mol Biol 297:659–672
82. Goodsell DS, Olson AJ (2000) Annu Rev Biophys Biomol Struct 29:105–153
83. Zhou HX, Dill KA (2001) Biochemistry 40:11289–11293
84. Asakura S, Iino T (1972) J Mol Biol 64:251–268
85. Kamiya R, Asakura S, Yamaguchi S (1980) Nature 286:628–630
86. Lawn AM (1977) J Gen Microbiol 101:112–130
87. Karplus PA, Schulz GE (1985) Naturwissenschaften 72:212–213
88. Vihinen M (1987) Protein Eng 1:477–480
89. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
90. Schwede T, Kopp J, Guex N, Peitsch MC (2003) Nucleic Acids Res 31:3381–3385
91. Rohl CA, Strauss CE, Chivian D, Baker D (2004) Proteins 55:656–677
92. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Methods Enzymol 383:66–93
93. Kim DE, Chivian D, Baker D (2004) Nucleic Acids Res 32: W526–W531
94. Honda S, Uedaira H, Vonderviszt F, Kidokoro S, Namba K (1999) J Mol Biol 293:719–732
95. Dill KA (1990) Biochemistry 29:7133–7155
96. Zavodszky P, Kardos J, Svingor, Petsko GA (1998) Proc Natl Acad Sci USA 95:7406–7411
97. Dioszeghy Z, Zavodszky P, Namba K, Vonderviszt F (2004) FEBS Lett 568:105–109