Expression and Characterization of Hyperthermophilic *Aquilfex Pyrophilus* Flagellin Protein in Mesophilic Bacterial Systems

Mukkamala

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Chemistry Commons

**Recommended Citation**

https://scholarworks.wmich.edu/masters_theses/4328
EXPRESSION AND CHARACTERIZATION OF HYPERTHERMOPHILIC *AQUIFEX PYROPHILUS* FLAGELLIN PROTEIN IN MESOPHILIC BACTERIAL SYSTEMS

by

Srivani V. Mukkamala

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Science
Department of Chemistry

Western Michigan University
Kalamazoo, Michigan
December 2006
In memory of my father
Shri. Apparao Mukkamala
ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Subra Muralidharan for his encouragement, sound advise, guidance and support through out my research. I would extend my gratitude to my co-advisor Dr. Brian C. Tripp for his valuable suggestions and guidance in my project. I thank Dr. Dongil Lee and Dr. Sherine Obare for serving on my research committee and contributing valuable insights to my thesis work. I owe my thanks to Dr. Narayanan Srividya for her priceless suggestions and discussions in my research and thesis and importantly her friendship.

I would like to thank Raghuram Malapaka, Thilak Kumara Mudalige and Venkat Reddy Guduru for their help in my work. Also I would thank my friend Jenny Powell for her help in my stay here. I would like to thank all my friends here for making my stay more memorable.

I thank W. M. Keck foundation and Department of Chemistry for the financial support. I would also thank the office staff of Chemistry department particularly Pam McCartney and Annie Dobbs who were available at all times to help.

Today I stand here only because of my parents, my husband and my son. The encouragement and support that my mother, Suseela Devi and my husband, Raman gave me strengthened me to achieve this goal. This is a success because of my cute son Satya Prasad. I dedicate my work and this thesis to my father Late Shri. Apparao.

Srivani V. Mukkamala
The structural protein flagellin, FlaA derived from hyperthermophilic *Aquifex pyrophilus*, a rod shaped eubacterial species that grows near hydrothermal vents at an optimum growth temperature of 85 °C forms elongated thermostable flagella nanotubes. FlaA was successfully over-expressed for the first time using pET28c plasmid in *E. coli* BL21 cells at 37 °C. The protein was isolated from inclusion bodies that were solubilized by alkaline pH, refolded by dialysis and purified by chromatographic techniques. The FlaA was characterized by fluorescence, light scattering and circular dichroism. The FlaA was not in the correct folded form to form flagella as efforts to form them employing temperature and salt resulted in random aggregates. Large rod shaped aggregates were obtained employing polyethylene glycol and were characterized by dynamic light scattering, fluorescence microscopy after staining with NanoOrange and transmission electron microscopy. The emission from the tryptophan residues were monitored in terms of intensity and anisotropy to understand the unfolding of the FlaA in the presence of guanidine hydrochloride (GdnHCl). The folded to completely unfolded state had an intermediate partially folded state with the completely unfolded state requiring 7 M GdnHCl. This work provides the basis for formation of thermally stable protein nanotubes for the generation of novel bionanostructures.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .............................................................................................................. ii

LIST OF TABLES..................................................................................................................... vii

LIST OF FIGURES..................................................................................................................... viii

CHAPTER

I. INTRODUCTION .................................................................................. 1

1.1 Classification of microorganisms......................................................... 1

1.2 Hyperthermophilic bacteria............................................................... 1

1.3 Aquifex pyrophilus ............................................................................ 2

1.4 Description of flagella....................................................................... 3

1.5 Function of flagella........................................................................... 4

1.6 Flagellin characteristics..................................................................... 5

1.7 Aquifex pyrophilus flagella and flagellin.......................................... 7

1.8 Sequence similarity of A. pyrophilus FlaA flagellin with Salmonella typhimurium FliC flagellin................................. 8

1.9 Flagellin monomer characterization.................................................. 9

1.10 Central hypothesis.......................................................................... 11

1.11 Specific aims of the research........................................................... 11

II. MATERIALS AND METHODS ....................................................................... 13

2.1 Bacterial strains.............................................................................. 13
## Table of Contents-Continued

### CHAPTER

2.2 Plasmids ................................................................. 14

2.3 PCR amplification for subcloning .................................. 15

2.4 Subcloning and DNA sequencing of the *flaA* gene .......... 19

2.5 Motility assay .......................................................... 22

2.6 Expression and purification of FlaA protein .................... 23

2.7 N-Terminal sequencing .............................................. 29

2.8 UV absorption measurements of FlaA monomer ............ 29

2.9 FlaA and FliC flagellin sample solutions for chemical denaturation ................................................................. 30

2.10 Fluorescence measurements ....................................... 31

2.11 Fluorescence anisotropy measurements ....................... 33

2.12 Circular dichroism spectroscopy measurements ................ 34

2.13 Chemical denaturation of FlaA and FliC with protein titration protocol ................................................................. 35

2.14 Dynamic light scattering ............................................ 36

2.15 Polymerization of FlaA ............................................. 39

2.16 Fluorescence microscopy ........................................... 41

2.17 Electron microscopy ................................................ 42
III. RESULTS AND DISCUSSION .......................................................... 44

3.1 Subcloning and sequencing ............................................. 44

3.2 Protein expression and purification .................................. 49

3.3 FliC and FlaA absorption and emission spectral characteristics 57

3.4 Protein folding .............................................................. 64

3.4.1 Fluorescence techniques to monitor folding of proteins 67

3.4.2 Unfolding studies of FlaA and FliC using steady-state fluorescence anisotropy 74

3.4.3 Characterization of FlaA and FliC monomer unfolding and oligomerization by light scattering 78

3.4.4 Characterization of FlaA monomer with circular dichroism 83

3.4.5 Analysis of folding data obtained by different techniques 88

3.5 Polymerization of flagellin FlaA monomer 100

IV. CONCLUSIONS AND FUTURE DIRECTIONS .............................. 106

4.1 The main conclusions from the research 106

4.2 Future directions ............................................................ 108

REFERENCES ................................................................. 110
<table>
<thead>
<tr>
<th>Table of Contents-Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPENDICES ................................. 122</td>
</tr>
<tr>
<td>A. DNA sequencing results ................. 122</td>
</tr>
<tr>
<td>B. N-terminal sequencing results ........... 132</td>
</tr>
<tr>
<td>BIBLIOGRAPHY ............................... 140</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1. Classification of microorganisms................................. 2
2. Polymerization of FlaA monomer with dry PEG......................... 40
3. Excitation and emission wavelengths to monitor FliC and FlaA........ 64
4. Summary of fitted equilibrium constants obtained from analysis of
   folding studies of flagellin FlaA and FliC monomers.................. 98
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Structure of the flagellin gene</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Electron micrograph of <em>A. pyrophilus</em> with polytrichous flagella</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>3D model of FlaA and FliC</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Map of 5367 bp pET-28c(+)</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>Techne Touchgene Gradient Thermal Cycler</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>Kodak EDAS 290 gel documentation system</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>MicroPulser™ Electroporator and cuvette electrode chamber</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>Comparison of swarming motility</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>Mini-Protean 3 Electrophoresis System Power Supply</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>Ultrasonic cell disrupter XL-2020 Sonicator® used for lysing bacterial cells</td>
<td>26</td>
</tr>
<tr>
<td>11</td>
<td>ÄKTA Fast Protein Liquid Chromatography™ (FPLC)</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>Beckman DU530 single cell module spectrophotometer</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>Shimadzu UV-Vis spectrophotometer</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>Varian Eclipse™ fluorescence spectrophotometer</td>
<td>32</td>
</tr>
<tr>
<td>15</td>
<td>Shimadzu RF-5301 fluorescence spectrophotometer</td>
<td>33</td>
</tr>
<tr>
<td>16</td>
<td>Close-up view of interior sample chamber</td>
<td>34</td>
</tr>
<tr>
<td>17</td>
<td>Aviv Model 202 Circular Dichroism Spectrometer</td>
<td>35</td>
</tr>
<tr>
<td>18</td>
<td>DynaPro Titan™ light scattering instrument</td>
<td>38</td>
</tr>
<tr>
<td>19</td>
<td>Wyatt QELS™ (Quasi-Elastic-Light-Scattering) instrument</td>
<td>39</td>
</tr>
</tbody>
</table>
List of Figures-Continued

20. Nikon TE 2000U inverted fluorescence microscope................................. 41
21. JEOL, JEM-1230EXII transmission electron microscope............................ 42
22. JEOL 3011 high resolution transmission electron microscope..................... 43
23. Agarose gel electrophoresis of 5' and 3' fragments of the
flaA flagellin gene from A. pyrophilus.................................................. 45
24. Agarose gel electrophoresis analysis of flaA gene insertion into
Into the pTrc99A S. typhimurium expression plasmid................................. 46
25. Agarose gel electrophoresis análise of flaA gene insertion
into the pET28c E. coli expression plasmid........................................... 47
26. SDS-PAGE analysis of FlaA expression and solubility in
E. coli BL21(DE3) cells............................................................................ 50
27. SDS-PAGE analysis of FlaA expression and solubility in
S. typhimurium strain SJW134................................................................. 51
28. SDS-PAGE analysis of solubilized and refolded FlaA flagellin
protein prepared from inclusion bodies.................................................. 52
29. Chromatogram of FlaA purification on a Superdex 75TM 60-70 cm
size exclusion column run on an ÄKTA Fast Protein Liquid Chromatography
instrument................................................................................................. 53
30. SDS-PAGE analysis of refolded FlaA purified by size exclusion
chromatography on a Superdex 75™ column......................................... 54
31. UV absorbance spectrum of FlaA monomer........................................... 58
32. Fluorescence emission spectra of FliC and FlaA.................................... 60
33. Fluorescence emission spectrum of FlaA monomer................................ 61
34. Fluorescence emission spectrum of FlaA monomer................................ 62
List of Figures-Continued

35. D3 domain of the predictive model of FlaA............................... 63
36. Structures of three common aromatic amino used for intrinsic fluorescence studies of proteins............................................. 66
37. Fluorescence emission spectra of native FlaA at pH 7.5................. 68
38. Fluorescence emission spectra of native FlaA at pH 7.5................. 73
39. Fluorescence spectrum of FlaA in the presence of GdnHCl............. 70
40. FlaA fluorescence titration curves in the presence of increasing concentration of GdnHCl....................................................... 71
41. Fluorescence emission spectra of native FliC at pH 7.5.................. 73
42. The fluorescence emission spectrum of FliC in the presence of GdnHCl...... 73
43. FliC titration curves in the presence of increasing concentration of GdnHCl... 74
44. Electronic absorption transitions $1L_a$ and $1L_b$ of the tryptophan indole side chain................................................................. 75
45. Fluorescence anisotropy of FlaA monomer as a function of GdnHCl concentration at 25 °C, excited at 295 nm.............................................. 76
46. Fluorescence anisotropy of FliC monomer as a function of GdnHCl concentration at 25 °C, excitation at 280 nm................................. 77
47. Fluorescence anisotropy of FlaA as a function of GdnHCl concentration at 25 °C, with excitation at 295 nm and 300 nm wavelengths............. 77
48. Light scattering of FlaA monomer, ~ 12.8 nm, determined by light scattering measurements................................................................. 78
49. Correlation function for FlaA monomer........................................... 79
List of Figures-Continued

50. Light scattering of FliC monomer, ~ 13.6 nm, determined by light scattering measurements ............................................................................................................ 79

51. Light scattering of FlaA aggregates formed in 1 M NaCl at 85 °C for 1 hour with a protein concentration of 2.0 mg/mL ................................................................. 80

52. Hydrodynamic radius of FlaA monomer as a function of GdnHCl concentration at 25 °C ........................................................................................................................................ 81

53. Hydrodynamic radius of FliC monomer as a function of GdnHCl concentration at 25 °C ........................................................................................................................................ 82

54. Translational diffusion coefficient of FlaA monomer with increasing concentration of GdnHCl ................................................................................................................................ .. 82

55. CD spectrum for FlaA flagellin ........................................................................................................... 84

56. Theoretical CD spectrum for FlaA flagellin monomer ........................................................................ 85

57. FlaA flagellin monomer CD titration curve in the presence of increasing concentration of GdnHCl denaturant monitored at 223 nm .............................................. 86

58. CD thermal melting curve of flagellin FlaA monitored at 220 nm ..................................................... 88

59. Fluorescence intensity titration curve for FlaA .................................................................................... 93

60. Fluorescence anisotropy of FlaA and FliC monomer as a function of GdnHCl concentration at 25 °C, excitation at 295 nm ....................................................................... 94

61. Hydrodynamic radius of FlaA monomer as a function of GdnHCl concentration at 25 °C ................. 95

62. FlaA monomer CD titration curve in the presence of increasing concentration of GdnHCl ......................... 96

63. Proposed folding pathway of thermophilic FlaA flagellin monomer .............................................. 99
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>64.</td>
<td>TEM image of polymerized of FlaA monomer in the presence of 1 M NaCl at 65 °C</td>
<td>101</td>
</tr>
<tr>
<td>65.</td>
<td>TEM image of polymerized FlaA monomer in the presence of 1 M NaCl at 85 °C</td>
<td>102</td>
</tr>
<tr>
<td>66.</td>
<td>Electron micrograph of negatively stained FlaA aggregate obtained after refolding and dialysis procedure from high pH solution</td>
<td>102</td>
</tr>
<tr>
<td>67.</td>
<td>Images of CongoRed stained FlaA aggregates under a polarized microscope</td>
<td>103</td>
</tr>
<tr>
<td>68.</td>
<td>Fluorescence images of NanoOrange stained polymerized FlaA in the presence of 20% PEG 4000 at room temperature and neutral pH</td>
<td>104</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

1.1 Classification of microorganisms

In nature microorganisms are classified as psychrophilic, mesophilic, thermophilic and hyperthermophilic depending on their optimal growth temperature. Psychrophilic microorganisms exhibit a growth optimum temperature of less than 20 °C. A mesophilic organism is considered to grow at an optimum temperature greater than 20 °C and less than 50 °C. As the name suggests, thermophilic microorganisms grow at high temperatures; their optimum growth temperature is between 50 and 80 °C. Those microorganisms which grow at extremely high temperatures, i.e., near the boiling point of water, are classified as hyperthermophilies. The optimal growth temperature for hyperthermophiles is typically between 80 and 120 °C. Unlike thermophiles, hyperthermophiles are unable to grow below 60 °C. They are very well adapted to grow above 80 °C.\textsuperscript{1, 2} The examples of each classification along with their optimal growth temperature are listed in Table 1.

1.2 Hyperthermophilic bacteria

Hyperthermophiles belong to phylogenetically distant groups; they represent ancient adaptations of life to conditions of extreme heat. Hyperthermophiles consist of anaerobic and aerobic chemolithoautotrophs and heterotrophs that grow at neutral or acidic pH. Hyperthermophiles include both bacteria and archaea, formerly known as archaeabacteria,
which represent the organisms at the extreme temperatures of life. Both bacteria and archaea are prokaryotes. Most hyperthermophiles belong to archaea except two genera of Bacteria, *Aquifex* and *Thermotoga*.

**Table 1. Classification of microorganisms**

<table>
<thead>
<tr>
<th>Nature of existence</th>
<th>Optimal temperature</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychrophilic</td>
<td>0-20 °C</td>
<td><em>Colwellia psychrerythraea,</em> <em>Moritella abyssi</em></td>
</tr>
<tr>
<td>Mesophilic</td>
<td>20-50 °C</td>
<td><em>Escherichia coli,</em> <em>Salmonella Typhimurium</em></td>
</tr>
<tr>
<td>Thermophilic</td>
<td>50-80 °C</td>
<td><em>Thermus Thermophilus,</em> <em>Thermotoga maritima</em></td>
</tr>
<tr>
<td>Hyperthermophilic</td>
<td>80-120 °C</td>
<td><em>Aquifex Pyrophilus,</em> <em>Thermotoga petrophila</em></td>
</tr>
</tbody>
</table>

**1.3 *Aquifex pyrophilus***

*Aquifex pyrophilus* is an example of hyperthermophilic hydrogen-oxidizing or “Knallgas” bacteria that was isolated from a hot marine thermal deposit (depth: 106 m) at the Kolbeinsey Ridge north of Iceland, which is a site of known volcanic activity. *Aquifex pyrophilus* is a Gram-negative, highly motile rod-shaped bacterium that belongs to genus *Aquifex*. It grows at temperatures between 67 and 95 °C, with an optimum temperature of 85 °C; at pH values ranging from pH 5.4-7.5, with an optimum pH of 6.8; and in the presence of 1 to 5% NaCl, with an optimum of 3% NaCl. This organism is a microaerophile and grows chemolithoautotrophically by using molecular hydrogen,
thiosulfate and sulfur as electron donors. Its unique property is its ability to utilize oxygen as an electron acceptor. It can also use nitrate as an electron acceptor under anaerobic growth conditions.\(^4\) Based on 16S rRNA sequence comparisons \textit{Aquifex pyrophilus} has been placed at the deepest phylogenetic branch within the eubacterial evolutionary tree.\(^5,6\)

1.4 Description of flagella

The locomotion of prokaryotic cells is usually provided by the special primary motility organelle, termed the flagellum (plural: flagella).\(^7-11\) Flagella are filamentous protein structures attached to the cell surface that are rotated to provide propulsion for most motile prokaryotes.\(^12,13\) The arrangement of the flagella on the cell surface could be lophotrichous (a “tuft” of flagella at one pole), monotrichous (a single flagellum at one pole), amphitrichous (a single flagellum or tuft of flagella at both poles), or peritrichous (flagella cover the entire bacterial cell). The ability of bacterial cells to perform rapid and directed movement depends on the presence of this motility organelle. Its structural and functional proteins are encoded by approximately 50 genes in bacteria.\(^14,15\) The bacterial flagellum is conserved at both the genetic and structural level throughout the bacterial kingdom, even though there is a wide diversity in the organisms that possess these flagella.\(^16\)

The bacterial flagellum is a complex structure which is composed of the following parts; a) a basal body localized within the cell wall, b) a trans-membrane rotary motor,\(^9,14,17-19\) c) an elongated external helical protein filament\(^20\) composed of flagellin subunits and d) a flexible universal joint-like, “hook” structure, connecting these two parts.\(^21-25\) The
filament is a long, thin cylindrical structure that is helical in shape with an outer diameter of 12-25 nm. Another important feature is the 2-3 nm diameter inner channel that extends all the way from the periplasmic face of the MS ring protein to the tip of the filament\textsuperscript{26, 27}; this channel functions as the flagellum-specific export pathway. The filament can be 1-15 µm or more in length.\textsuperscript{14, 20}

1.5 Function of flagella

The bacterial flagellar motor is a nanotechnological marvel which is located in the cell envelope at the flagellar base\textsuperscript{24}; it rotates clockwise or counterclockwise at speeds on the order of 100 Hz, thus rotating long thin helical filaments to enable the bacterial cells to swim.\textsuperscript{17} During counterclockwise rotation, the flagella associate together into a bundle to form a single helix,\textsuperscript{28, 29} thus resulting in steady forward motion of the cell, termed “running”. When the motors rotate clockwise the flagella bundle unravels and the flagella separate into individual fibers; the resulting cellular motion, in which the cell randomly spins in place with no net velocity, is known as “tumbling”. Tumbling results in a random change in the orientation of the bacterial cell, i.e., the direction at which it is pointed and will swim when flagella motors again reverse direction.\textsuperscript{17} The processes of running and tumbling together enable the organism to move in one direction and reorient itself to move in another direction, allowing a time-dependent sampling of the environment. The amount of time spent in tumbling mode vs. swimming process of bacterial movement may be a response to chemical stimuli (chemotaxis).\textsuperscript{7}
1.6 Flagellin characteristics

The number of flagellin genes present in every organism varies. Some bacterial genomes contain only one flagellin gene, while others contain multiple flagellin genes. The requirement of multiple flagellin genes in some bacteria is to provide functional flagella as in *Bacillus thuringiensis*\textsuperscript{30} and in others is thought to create phenotype variation to avoid host immune detection, as in *Salmonella typhimurium*.\textsuperscript{31} The amino acid composition of the N- and C- terminal regions of flagellin proteins is very consistent across species. Amino acids such as valine, leucine and isoleucine, which are mostly found in secondary structure elements, are common in the conserved regions of flagellin proteins, amino acids with a strong preference for forming α-helix secondary structures such as alanine are very common in flagellin protein. The amino acids proline, histidine and tryptophan are known to disrupt the α-helix secondary structure and consequently, are rare in the conserved N- and C- terminal regions of flagellin sequences. Proline is known to introduce sharp bends in the secondary structure and pairs of cysteine residues tend to form disulfide bonds in proteins, which are not found in any known flagellin proteins, to date.

Flagellin is a multi-domain protein. Earlier structural analysis of flagellin proteins indicated that they are composed of three distinct domains; an N-terminal domain, a central variable region and a C-terminal domain (Figure 1). More recent structural studies have shown that the *S. typhimurium* phase 1 flagellin protein is composed of four globular domains, D0, D1, D2 and D3. The flagellin sequence contains ~140 conserved residues from the start codon to the N-terminal D0, D1 and the β-turn.\textsuperscript{16} Another ~90 residues are conserved in the C- terminal D0 and D1 domains.\textsuperscript{16} The highly conserved N-
and C-terminal regions form the D0 and D1 domains, which form the α-helical, "coiled-coil" region of flagellin. These coiled-coil domain regions are critical for self-assembly of flagellins to form the helical flagella fiber and form the inner region surrounding the central-pore of flagella fibers.

Insertions and deletions in the N- and C-terminal residues disrupt the ability of flagellins to self-assemble and form flagella fibers. The N- and C-terminal sequences are known to have hydrophobic residues that are conserved: They are required for the formation of α-helical coiled-coils and intermolecular interactions with other D0 domains to form oligomeric flagella structure.32 The "spoke" region which connects D0 and D1 regions of flagellin is also conserved. The hydrophilic residues in C-terminal residues are known to extend into the pore of flagella.32 This is very important for the transport of unfolded flagellin monomers through the pore, prior to FliD/Hap2 chaperonin-mediated self-assembly at the distal end of the growing flagella fiber.

Given the high degree of sequence conservation observed in the D0 and D1 domains, this suggests that that the self-assembly of any eubacterial flagellin protein to form flagella fibers occurs in a similar manner. The central hypervariable region forms the D2 and D3 domains: this sequence region varies greatly in composition and size. i.e., the number of

---

**Figure 1.** Structure of the flagellin gene. Showing the highly conserved N- and C-terminal domain regions and the middle, hypervariable sequence region.
residues across species. This sequence variability of the hypervariable region results in flagellins having sequences that are unique to the organisms from which they are isolated.

1.7 *Aquifex pyrophilus* flagella and flagellin

*A. pyrophilus* is motile by the possession of eight polytrichously arranged flagella fibers on the cell surface (Figure 2). The morphology of *A. pyrophilus* flagella exhibits a sinusoidal curvature. Its average width is 19 nm, which is within the typical range of mesophilic bacterial filament diameters.\(^{33}\)

![Electron micrograph of A. pyrophilus flagella. Image obtained from Behammer et al.\(^{33}\)](image)

**Figure 2.** Electron micrograph of *A. pyrophilus* with polytrichous flagella. Image obtained from Behammer et al.\(^{33}\)

An *A. pyrophilus* flagella is a bio-polymer which is composed of helically arranged flagellin protein subunits. The 1506 base pair (bp) \(flaA\) gene of *A. pyrophilus* encodes a 501 residue flagellin protein (FlaA).\(^{33}\) The N-terminal methionine residue of FlaA is post-translationally removed.\(^{33}\) The corresponding molecular weight of the mature form of
FlaA is 53.9 kDa. A predictive model of the hyperthermophilic *A. pyrophilus* FlaA protein structure\(^3\) (PDB 1XGX) was built based on the experimentally determined structure of *S. typhimurium* flagellin (FliC)\(^1\) (PDB 1UCU) (Figures 3a and 3b). These structures show the previously described domains, D0, D1, D2 and D3.

**1.8 Sequence similarity of *A. pyrophilus* FlaA flagellin with *Salmonella typhimurium* FliC flagellin**

It was previously shown that the hyperthermophilic FlaA protein has \(~30\%\) sequence similarity with mesophilic FliC flagellin.\(^16\), \(^33\), \(^34\) FliC is a mesophilic flagellin protein which denatures at \(~65\ °C\), well below the optimum growth temperatures of *A. pyrophilus*, which is 85 °C.\(^35\) Comparison of the FlaA and FliC amino acid sequences revealed that the N- and C- terminal residues are highly conserved and that the middle domain sequence is different from that of other flagellins.

*Figure 3.* 3D model of FlaA and FliC (a) *A. pyrophilus* FlaA (PDB 1XGX), (b) Experimentally determined 3D structure of *S. typhimurium* FliC (PDB 1UCU).
The primary sequence of FlaA encodes the structure and unusually high heat stability of this hyperthermophilic protein. It is evident from both sequences that FlaA contains tryptophan residues which are absent in FliC. There is also an increase of proline residues in FlaA along with an increase in hydrophobic residues such as isoleucine and aromatic amino acids when compared to FliC.

Proline residues are known to introduce sharp turns in the alpha helix secondary structures of proteins; such a feature can lead to the formation of compact monomers and large interfaces between subunits (excluding water molecules) in the helical polymer. Behammer et al. suggested an increase of 7.9% ± 3.3% of the above amino acids in hyperthermophilic flagellin in comparison with mesophilic species. The percentage of hydrophobic residues in N-terminal, C-terminal, and middle domains according to Malapaka et al. are 45.2%, 50.6% and 40.4%. In the case of FliC the values are 35.5%, 43% and 36.6%. There is an increase in the overall hydrophobic character of FlaA, relative to FliC. Along with this increase in hydrophobicity, there is a decrease in the polar residue content of FlaA. This increase in hydrophobic character may result in a decrease in the average chain flexibility; thermostable FlaA flagella filaments isolated from *A. pyrophilus* were observed to be fragile, i.e., brittle and prone to breakage, at room temperature.

1.9 Flagellin monomer characterization

One aspect of this study is to characterize the stability aspects of hyperthermophilic FlaA flagellin in detail by chemical denaturation studies. FlaA is a thermostable protein, thermal denaturation studies may not be feasible; thus, the protein folding and unfolding
of FlaA were studied in presence of a chemical denaturant. Techniques such as fluorescence intensity, fluorescence anisotropy, light scattering and circular dichroism were used to detect changes in the FlaA tertiary structure that occur in the presence of a denaturant.

A major goal of research in our laboratory is to employ unique biological structures such as flagella as scaffolds for the generation of nanocomposites by bottom-up construction for protein monomers. The outmost D3 domain of FliC flagellin subunits are evenly spaced apart in *S. typhimurium* FliC flagella fibers at a distance of 5.4 nm. The D3 domain is comprised of the hypervariable region in flagellin. It has been demonstrated that peptide loops and other functional groups can be introduced into this region to obtain evenly spaced binding sites on the surface of flagella bionanotubes. Such sites can be employed to bind preformed nanoparticles or nucleate the growth of nanoparticles and nanotubes. Novel bionanostructures can be generated for applications in catalysis, sensing, molecular electronics and energy generation and storage devices. Research in our group has already demonstrated the usefulness of a bioengineered mesophilic *E. coli* flagellin for the generation of novel bionanostructures. The FlaA system is complementary to mesophilic FliC systems and could potentially be used to develop bionanostructures that are stable at high temperatures and pressures. Such structures could be especially useful for applications in catalysis, sensing and molecular electronic devices. Thus, the impetus for the expression and characterization of FlaA, in addition to its fundamental importance, is its potential application for the generation of useful bionanostructures. Flagella offer a unique advantage over peptide nanotubes in that they can be generated in large quantities employing bacteria in contrast to peptide
nanotubes which are purely synthetic systems. This is particularly important for the economical generation of useful bionanostructures for practical applications.

1.10 Central hypothesis

The central hypothesis of this research is that hyperthermophilic flagellin protein can be expressed in mesophilic systems such as *E. coli* and *Salmonella* and isolated for characterization and self-assembly to form flagella bionanostructures.

1.11 Specific aims of the research

1. Expression of hyperthermophilic FlaA in mesophilic *Escherichia coli* strain BL21 (DE3) which is a lambda lysogenic strain carrying T7 RNA polymerase (B F *dcm ompT* *hsdS* (rB− mB−) *gal l* (DE3)). This aim includes testing FlaA for soluble expression vs. insoluble (misfolded) inclusion body aggregates.

2. Expression of FlaA in mesophilic *Salmonella* strain SJW134 (*ΔflfC* and *ΔflfB*) and testing for its export and assembly into functional flagella via a motility assay. SJW134 was derived from the parent strain SJW806, which is wild-type except for the deletion of the phase-1 and phase-2 flagellin genes. This strain is non-motile, unless a functional flagellin gene is introduced, e.g., on an expression plasmid.

4. Characterization of purified FlaA monomer and oligomer structure using biophysical methods such as dynamic light scattering, circular dichroism spectroscopy, fluorescence spectroscopy and fluorescence anisotropy. 42,43

5. Characterization of FlaA folded stability as a function of guanidine hydrochloride chemical denaturant concentration and temperature, relative to mesophilic FliC flagellin. The resulting data will be modeled using standard thermodynamic models for the unfolding of proteins.

6. Conduct experimental studies to determine if FlaA can be self-assembled in vitro to form flagella nanotubes. Characterize any resulting self-assemblies by TEM, fluorescence and dark field microscopy.

7. Perform characterization of fibers with an established CongoRed dye assay to determine whether the polymers formed are flagella fibers or beta-amyloid like fibers.
CHAPTER II

MATERIALS AND METHODS

2.1 Bacterial strains

*Escherichia coli* BL21(DE3) is a lambda lysogenic strain. The genotype includes a prophage carrying T7 RNA polymerase (B F dcm ompT hsdS (rB- mB-) gal l (DE3)).

BL21(DE3) electrocompetent cells (Stratagene, La Jolla, CA) were used for overexpressing the flagellin protein (FlaA) of *Aquifex pyrophilus* from the flaA gene insert in pET28c. This strain carries an integration-deficient immunity 21 phage which contains the gene for T7 RNA polymerase under the control of LacUV5 promoter and is an all-purpose strain for high-level protein expression and easy induction by addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG) with T7 promoter plasmids, e.g. pET vectors. This strain lacks both the lon protease and the ompT outer membrane protease, which can degrade proteins during purification. These cells have a transformation efficiency greater than $1 \times 10^6$ transformants/µg of DNA. *E. coli* strain DH5α (Stratagene) was used for sub-cloning experiments. DH5α cells carry recA1 and endA1 mutations that increase insert stability and improve the quality of plasmid DNA respectively. The approximate transformation efficiency of these cells is $1.0 \times 10^7 \sim 1.0 \times 10^8$ transformants / µg of DNA. The following *Salmonella serovar Typhimurium* LT2 (common name *S. typhimurium*) strains were used for overexpression of flaA cloned into the pTrc99a vector. The strain was kindly provided by the late Dr. Robert Macnab (Yale University). *S. typhimurium* strain SJW1103, which can only express the phase 1 fliC flagellin (*i.e.*, fliC stable) was used as a wild-type flagellar motility control for the studies.
of modified flagellin proteins and motility assays in solution and on agar plates. *S. typhimurium* strain SJW134 (AfliC and AflijB), which was derived from parent strain SJW806, is wild-type except for the deletion of the phase 1 *fliC* and phase 2 *fljB* flagellin genes.\(^{41, 50}\) This strain is non-motile, unless a functional flagellin gene is introduced, e.g., on an expression plasmid.\(^{41}\) The electrocompetent cells of *E. coli* BL21(DE3), *E. coli* DH5α and *S. typhimurium* strain SJW134 were prepared using standard lab protocols.\(^{51}\) *S. typhimurium* strain SJW1103 was used as a wild-type flagellar motility control for the studies of modified flagellin proteins and motility assays in solution and on agar plates.

### 2.2 Plasmids

The *A. pyrophilus* flagellin gene, *flaA*, was previously cloned and sequenced by Behammer *et al.*\(^{33}\) The plasmids pRU1650 and pRU1651, each containing *A. pyrophilus* genomic fragments encoding partial 3' and 5' overlapping segments of the *flaA* gene were kindly provided by Dr. Rudiger Schmitt (University of Regensburg, Germany). The pET-28c(+) vector from Novagen was used for subcloning the flagellin gene and expression of flagellin protein. It is a 5367 base pair (bp) plasmid that carries an IPTG-inducible T7 promoter, a kanamycin (Kan) antibiotic resistance gene and the *lacI* gene encoding the lac repressor protein, which is a regulatory DNA binding protein that inhibits the expression of genes downstream of the T7 promoter region in the absence of IPTG or lactose (Figure 4). Its sequence is numbered by pBR322 convention.

The pTrc series of plasmids were constructed based on pKK233-2 for the regulated expression of genes in *E. coli*. The pTrc vectors carry a strong hybrid trp/lac, IPTG-
inducible promoter, the lacZ ribosome-binding site (RBS), the multiple cloning sites of pUC18, the rRNA transcription terminators and a β-lactamase ampicillin (Amp) resistance gene. The pTH890 plasmid is a derivative of the pTrc99A plasmid with the S. typhimurium fliC phase 1 flagellin gene cloned into the XbaI/HindIII restriction site. The pTH890 vector was kindly provided by the late Dr. Robert Macnab (Yale University).

![Figure 4](image_url)

**Figure 4.** Map of 5367 bp pET-28c(+). Showing kanamycin resistant gene and other important features. Reproduced from page 1 of Novagen pET vector manual.

### 2.3 PCR amplification for subcloning

The procedures for restriction digestion, ligation and recombinant DNA formation were followed as described by Sambrook and Russell. All restriction enzymes were obtained from New England Biolabs (Ipswich, MA), unless otherwise noted. All plasmid DNA and PCR product DNA purification kits were obtained from Qiagen Inc. (Valencia, CA). The DNA polymerase, nucleotides (dNTPs) and other polymerase chain reaction (PCR)
reagents were obtained from Stratagene (La Jolla, CA). All PCR primers were obtained from Invitrogen (Carlsbad, CA) and dissolved in deionized (DI) H$_2$O to give a stock concentration of 100 µM. The PCR amplification primers with overhangs encoding restriction sites were designed with the use of Primer3$^{53}$ and Fast PCR$^{54}$ programs for the amplification of the flaA gene fragment inserts.

**Figure 5.** Techne Touchgene Gradient Thermal Cycler (Burlington, NJ).

The cloning design strategy involved the generation of identical HindIII restriction sites in the two PCR products: when these two sites were cut with HindIII restriction enzyme, the ends of the two complementary 3' and 5' flaA PCR products could be ligated together to generate a full-length flaA gene. One pair of PCR primers were designed with Ncol and HindIII restriction sites for the pRU1650 plasmid and a second pair of PCR primers
were designed with HindIII and NdeI restriction sites for the pRU1651 plasmid. The *flaA*
inserts were amplified by PCR with a Techne Touchgene Gradient Thermal Cycler
(Burlington, NJ) (Figure 5); this instrument has a heated lid that does not require the use
of mineral oil overlays on the reactions to prevent sample evaporation.

The FASTA format of the nucleotide sequence was used for designing the primers using
the above software. The 5' *flaA* gene fragment was amplified from pRU1650 using the
following PCR primers, with restriction sites denoted by boldface and the complimentary
regions to the *flaA* gene underlined: forward primer 5'-CAT GAC ATG GCC ACG AGG
ATT AAT TAC A-3' with an NcoI restriction site and reverse primer 5'-AAA ATA TAAT
AGC TTT TGC CTT TAT TCC CAT-3' with a HindIII restriction site. Likewise, the 3'
*flaA* gene fragment was amplified from pRU1651 using the following PCR primers:
forward primer 5'-ACC AAC ACA CAA ATA GAG TCG-3' with a HindIII
restriction site and reverse primer 5'-CAT ATG CAT TCA CCT GAG AAG CTG
GAG AA-3' with an NdeI restriction site. PCR amplification reactions were prepared
with 1 µl *PfuTurbo®* DNA polymerase, 5 µL of 10x standard PCR buffer, 5 µL of 2.5
mM dNTP mix, 5 µL each of 10 µM forward and reverse PCR primers, 30 µL of DI H2O
and 1 µL of plasmid DNA (pRU1650 or pRU1651) in a 50 µL PCR tube. The PCR
reaction conditions used were: An initial cycle of 94 °C for 10 minutes to perform cell
lysis, followed by 30 cycles of 94 °C for 1 min; 55 °C for 1 min; 72 °C for 1.5 min,
followed by a final extension step of 72 °C for 10 min. After amplification, 5 µL of
samples were analyzed by agarose gel electrophoresis in a standard 1% agarose gel.
followed by staining with ethidium bromide (EtBr) to confirm the presence of the
amplified gene insert. The EtBr-stained DNA in the gel was visualized by fluorescence
excitation with a 300 nm UV light source and images were recorded with a Kodak EDAS 290 gel documentation system (Kodak Molecular Imaging, New Haven, CT) (Figure 6). The two PCR cloning products obtained from amplification of pRU1650 and pRU1651 were subjected to HindIII restriction digestion to yield cohesive ends. Ligation of two PCR products at their HindIII-digested cohesive ends was performed using a Fast Link DNA Ligation Kit (Epicentre Biotechnologies, Madison, WI) to obtain a complete flaA gene.

Figure 6. Kodak EDAS 290 gel documentation system (Kodak Molecular Imaging, New Haven, CT).

The full-length flaA gene thus obtained was amplified by PCR using primers specific for the 5' and 3' ends of the full-length gene; this step was intended to minimize the concentration of other fragments and incorrect ligation products. The PCR primers
contained overhangs encoding Ncol and Sall restriction sites: forward primer 5'-CAT GCA ACG AGG ATT AAT TAC A-3' with an Ncol site and reverse primer 5'- TAG CTG TCG ACT CAC CTG AGA AGC TGG AGA A-3' with a Sall site. The restriction sites in these primers were designed for cloning of the full-length flaA gene into corresponding restriction sites encoded in multiple cloning sites in the pET28c and pTrc99a vectors.

2.4 Subcloning and DNA sequencing of the flaA gene

The PCR-amplified flaA DNA fragment was simultaneously digested, i.e., “double digested” with Ncol and Sall restriction enzymes. A 50 µL reaction mixture was incubated at 37 °C for 1 hour which contained 5 µL of NEB buffer 3, 1 µL of 100X BSA, 1 µL each of Ncol and Sall, 10 µL of PCR amplified flaA DNA and 32 µL of DI H2O. The restriction digest reaction was terminated by heating the sample at 67 °C for 20 minutes, the recommended protocol for inactivation of both restriction enzymes. Following the double digestion with Ncol and Sall, the flaA gene in the reaction mixture was purified with a QIAquick PCR Purification Kit (Qiagen). A quantity of 200µg of the pET28c and pTrc99a plasmids of were also double digested with the same Ncol and Sall restriction enzymes by incubation at 37 °C for 3 hours using the same 50 µL reaction mixture as above. The restriction enzymes were inactivated by incubating at 67 °C for 20 minutes. The small internal fragments of the plasmids which were cleaved by restriction digestion were removed with a QIAquick PCR Purification Kit (Qiagen). Following the double digestions, the flaA gene was cloned into the Ncol and Sall sites of the plasmids pET28c and pTrc99a plasmids with an overnight ligation at 16 °C in a 20 µL reaction
mixture with T4 ligase. Confirmation of the ligation was performed with a double
digestion by Ncol and Sall. The size of insert (1,503 bp) from this digestion was
determined by 1% agarose gel electrophoresis. Electrocompetent *E. coli* BL21(DE3) and
*S. typhimurium* SJW134 cells were prepared using standard lab protocols. The
electrocompetent cells were then transformed with the recombinant pET28c and pTrc99a
vectors containing the *flaA* gene.

*Figure 7.* MicroPulser™ Electroporator and cuvette electrode chamber (BioRad,
Hercules, CA).

Electroporation was performed by applying an electric pulse with MicroPulser™
Electroporator (BioRad, Hercules, CA) (Figure 7) to a mixture containing 50 µL
electrocompetent cells and 1 µL plasmid DNA in a Gene Pulser® and MicroPulser™
Cuvettes, 0.1 cm gap. After electroporation, SOC media was added and cells were incubated at 37 °C for 45 minutes with shaking in an incubator. These transformed cells were centrifuged at 10,000 x g at 4 °C for 10 minutes to concentrate the cells. The supernatant was discarded and the cells were then plated on Luria-Bertani (LB) media plates containing 50 µg/mL kanamycin and 50 µg/mL ampicillin and cultured overnight at 37 °C. Single colonies from the plates were picked with sterile toothpicks and used to inoculate 5 mL test-tube cultures of LB-ampicillin-kanamycin media and grown overnight at 37 °C in a Cell-GRO Tissue culture rotator (Lab-Line instruments Inc. Melrose Park, IL, USA). These cultures were used to prepare purified plasmid DNA with a QIAprep Spin Miniprep Kit (Qiagen).

The presence of the full-length flaA insert in the plasmids was first analyzed by performing a double digestion of purified plasmid DNA with Ncol and Sall restriction enzymes. The approximate size of the insert (1,503 bp) from this digestion was determined by gel electrophoresis on a 1% agarose gel. In-frame ligation of the amplified DNA into the pET28c and pTrc99a vectors was then confirmed by DNA sequencing at the University of Michigan DNA Sequencing Core (Ann Arbor, MI), using T7 primers for pET28C and custom primers for pTrc99a. The DNA sequencing results indicated that the resulting pET28c plasmid contained the entire coding region of the cloned flaA gene downstream of the T7 promoter and the ribosomal binding site (Appendix A). The nucleotide sequence was analyzed by NCBI-BLASTN software and was compared with the previously deposited flaA gene sequence, GenBank no. U17575.
2.5 Motility assay

A swarming agar motility assay was performed to determine if the thermophilic FlaA flagellin protein could rescue motility in the non-motile mesophilic S. typhimurium strain SJW134. The S. typhimurium SJW134 electrocompetent cells were transformed with pTrc99a-flaA and pTH890 plasmids and plated on LB agar (1.5%, wt./vol. agar) followed by overnight growth at 37 °C. The media used for the motility assay was tryptone broth prepared using 10 g/L tryptone (EM Science, Gibbstown, NJ), 5 g/L NaCl (Fisher Scientific) and 0.3% (wt./vol.) agar (Sigma-Aldrich, St. Louis, MO). For the motility assay, sterile motility agar plates were streaked with overnight grown bacterial colonies on LB agar plates with a sterile, blunt-ended toothpick. The plates were then incubated at 30 °C for 6-8 hours in a humidified incubator to prevent dehydration of the media.

![Figure 8. Comparison of swarming motility. Comparison of S. typhimurium SJW134 cells transformed with plasmids that express FliC and FlaA flagellins. The image on the right shows that flaA expression in SJW134 cells could not restore the motility of the non-motile cells, while the left image shows restoration of motility observed with fliC gene.](image-url)
The low density of the agar allowed any motile, flagellated bacteria to move within the agar, forming a visible halo of growth around the point of inoculation, while non-motile cells remained in the region of initial inoculation. The swarming of the cells beyond point of inoculation was observed only for the cells transformed with the pTH890 plasmid encoding the functional fliC flagellin gene but not for cells transformed with the plasmid containing the flaA insert. This negative result indicated that the FlaA protein does not confer motility in mesophilic S. typhimurium cells (Figure 8).

2.6 Expression and purification of FlaA protein

The recombinant pET28c plasmid with the flaA gene insert was electroporated into E. coli BL21 (DE3) cells. The colonies were inoculated into 100 mL of LB broth containing 50 µg/mL kanamycin and were grown overnight at 37 °C with shaking at 300 rpm. The cells were used to inoculate 1 L of LB broth containing 50 µg/mL of kanamycin to give an initial OD$_{600}$ of 0.05. Expression of the FlaA protein was induced by the addition of 1 mL of 0.8 mM IPTG when the OD$_{600}$ reached 0.8, to give a final IPTG concentration of 0.8 µM. The cells were then grown for an additional 6 hours and were then harvested by centrifugation at 5000 x g, 4 °C for 30 minutes.

Confirmation of FlaA protein expression and investigation of protein solubility was determined as follows. A 1 mL volume of the pre- and post-induction samples were centrifuged at 10,000 x g for 10 min. using pre-weighed 1.5 mL microcentrifuge tubes. The supernatant was discarded and the wet weight of each cell pellet was determined. The cell pellet was resuspended by vortexing and the cells were lysed using 5 mL Bug Buster reagent (Novagen) per gram of wet cell paste at room temperature. The cell
suspension was incubated for 20 min at room temperature and was then centrifuged at 10,000 x g for 20 min at 4 °C.

**Figure 9.*** Mini-Protean 3 Electrophoresis System Power Supply (BioRad, Hercules, CA)

The resulting supernatant and pellet were analyzed with respect to the amount of FlaA present in soluble form and insoluble inclusion body form by 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with a Mini-Protean 3 Electrophoresis System (BioRad, Hercules, CA) (Figure 9) by the method of Laemmli. The gel was run for 90 min at 30 mA and 180 V. The gel was visualized by staining with Coomassie brilliant blue R-250 (EM Science). The 15% acrylamide gel was prepared by pouring a separating gel mixture which contained 7.5 mL of 30:0.8% acrylamide:bis-acrylamide, 3.75 mL of 4X Tris-Cl/sodium dodecyl sulfate (SDS) (pH 8.8), 3.75 mL of DI H₂O, 0.05 ml of 10% (wt./vol.) ammonium persulfate (APS) and 0.01 mL of N,N,N',N'-tetramethylethylenediamine (TEMED) into the casting frame and glass plate.
sandwich. Once the gel polymerized, a stacking gel mixture was carefully poured onto the top of the polymerized separating gel, followed by insertion of a plastic 10-well or 15-well comb to form the sample loadingwells. The stacking gel contained 0.65 mL of 30:0.8% acrylamide:bis acrylamide, 1.25 ml of 4X Tris-Cl/SDS (pH 6.8), 3.05 mL of DI H₂O, 0.025 mL of 10% (wt./vol.) APS and 0.005 mL of TEMED. The acrylamide gel staining solution contained 2 g of Coomassie brilliant blue R-250 dye dissolved in 100 mL of methanol, 100 mL of glacial acetic acid and 800 mL of DI H₂O. The acrylamide gel destain solution contained 100 mL of 100% methanol, 100 mL glacial acetic acid and 800 mL H₂O.

The cell pellet from 1 L of E. coli BL21(DE3) cell culture was processed using a Protein Refolding Kit (EMD Biosciences, Inc., San Diego, CA) to isolate and solubilize the inclusion bodies. The bacterial pellet was resuspended in 0.1X culture volume of 1X 1B wash buffer (200 mM Tris-HCl, pH 7.5, 100 mM EDTA, 10% Triton X-100). The suspension of cells was lysed by sonication on ice for 10 min using an ultrasonic cell disrupter XL-2020 Sonicator® (Misonix, Farmingdale, NY) (Figure 10). The sonicator power settings were set for alternating cycle of 10 sec of sonication (power on) followed by 10 sec of no sonication (power off) for 5 minutes. The cell lysate was centrifuged at 10,000 x g, 4°C for 10 min. The supernatant was discarded and the pellet was resuspended in one tenth the original cell culture volumes of 1X1B wash buffer. This wash cycle was repeated twice. The wet pellet weight was determined with an electronic balance and the volume of 1X1B solubilization buffer (500 mM CAPS, pH 11.0, 30% N-lauroylsarcosine) necessary to resuspend the washed inclusion bodies at a concentration
of 20 mg/mL was prepared from 10X1B solubilization buffer and 30% N-lauroylsarcosine.

Figure 10. Ultrasonic cell disrupter XL-2020 Sonicator® (Misonix, Farmingdale, NY) used for lysing bacterial cells.

The Protein Refolding Kit manual instructions were to resuspend the inclusion bodies at a concentration of 10-20 mg/mL in 1X1B solubilization buffer. The calculated volume of 1X1B solubilization buffer was added and the sample was incubated for 15 min at room temperature. After the incubation step, the suspension was centrifuged at 10,000 x g for 10 min. The supernatant containing the solubilized FlaA protein was dialyzed in 1X1B dialysis buffer (1 M Tris-HCl, pH 8.5), to remove the NaOH, lower the pH and allow the solubilized FlaA protein to refold. An overnight dialysis was performed with a buffer volume of 50 times the protein solution volume using Slide-A-Lyzer cassette™ 10 kDa
molecular weight cutoff (MWCO) dialysis cassettes (Pierce, Rockford, IL). The dialysis was repeated for 3 h. with a single buffer change. The dialyzed protein was further purified using size exclusion chromatography on a Superdex 75 prep grade 10/30 column (GE Healthcare Life Sciences Corp / Amersham-Biosciences, Piscataway, NJ), which has dimensions of 2.6 cm diameter, ~60 cm bed height. The column was connected to an ÄKTA Fast Protein Liquid Chromatography™ (FPLC) instrument with sample detection at 280 nm and a Frac-950 fraction collector (Figure 11). UNICORN™ version 3.1 software which runs on the Windows 2000 operating system was used to control the FPLC and analyze the absorbance data that was obtained from the chromatographic separation.

Figure 11. ÄKTA Fast Protein Liquid Chromatography™ (FPLC). Instrument equipped with sample detector at 280 nm and a Frac-950 fraction collector.
The Superdex 75 prep grade column has a size exclusion limit of 100 kDa) and an optimal size range between 3–70 kDa. The FPLC system alarm pressure was set to 0.5 MPa. A volume of 20 mL of the refolded FlaA inclusion sample was loaded onto the column, using a 50 mL capacity Superloop. The sample was then eluted from the column using a flow rate of 0.5 mL/min. The monomeric FlaA protein was eluted isocratically with 20 mM Tris-HCl, pH 7.5, 150 mM NaCl. The eluted protein fractions were pooled and concentrated using Centriprep concentrators (Amicon Inc., Beverly, MA) and analyzed by SDS-PAGE on 15 % polyacrylamide gels, with power settings of 30 mA and 180 V for 90 min and visualized with Coomassie brilliant blue R-250.

Figure 12. Beckman DU530 single cell module spectrophotometer (Beckman Coulter Inc., Fullerton, CA).
The concentration of DNA and protein samples was determined with a Beckman DU530 single cell module spectrophotometer (Beckman Coulter Inc., Fullerton, CA) (Figure 12). The protein concentrations were determined from the absorbance at 280 nm and the samples diluted 1: 100 in semi-micro quartz cuvettes using molar absorptivity values (extinction coefficients) of 37820 M$^{-1}$ cm$^{-1}$ for FlaA and 17880 M$^{-1}$ cm$^{-1}$ for FliC. These molar absorptivity values were computed from the aromatic amino acid residues present in FlaA and FliC amino acid sequences using the ProtParam tool at the ExPASy Proteomics Server at http://www.expasy.org/tools/protparam.html.

2.7 N-Terminal sequencing

The identity of the purified flagellin FlaA protein was confirmed by N-terminal sequencing. A sample of purified FlaA was submitted for N-terminal sequencing on an Applied Biosystems (Foster City, CA) Model 494 cLC at Michigan State University (East Lansing, MI). The sample was desalted on a C18 column by reverse phase chromatography.

2.8 UV absorption measurements of FlaA monomer

The UV absorption spectrum of soluble, monomeric FlaA was obtained with a Shimadzu UV-Vis spectrophotometer (Model: UV 1650PC) (Figure 13). The baseline spectrum was determined by scanning a solution of 20 mM sodium phosphate buffer, pH 7.5 from 200 to 800 nm.
2.9 FlaA and FliC flagellin sample solutions for chemical denaturation

A buffer solution consisting of 20 mM sodium phosphate at pH 7.5 was used in all protein denaturation experiments. A commercial stock of predissolved 8 M guanidine hydrochloride solution (GdnHCl) (catalog number- G7294) from Sigma-Aldrich was used for the chemical denaturation studies. Stock 10X protein solutions of 4.06 mg/mL FlaA and 0.77 mg/mL FliC flagellin proteins were prepared in 20 mM, pH 7.5 sodium phosphate buffer. GdnHCl stock solution and 100 µL of protein stock were added with the required volume of 20 mM phosphate buffer to give a final volume of 1 mL. This yielded final GdnHCl concentrations ranging from 0 to 7 M and the desired final protein concentration of 0.2 mg/mL. Samples were incubated overnight at 4 °C and also at room temperature (25 °C) for 1 hour. Longer incubation times at 4 °C resulted in identical fluorescence, anisotropy, circular dichroism (CD) and light scattering signals, therefore
the incubation temperature of 4 °C was chosen for denaturation experiments. Fluorescence, anisotropy, light scattering and CD experiments were performed with the above protein samples.

2.10 Fluorescence measurements

Measurement of protein intrinsic fluorescence as a function of denaturant concentration, e.g., urea and GdnHCl, is a standard method for investigating the folding and stability that are characteristic of globular proteins. FlaA contains all three types of aromatic amino acid residues (phenylalanine, Phe; tyrosine, Tyr; and tryptophan, Trp) as part of its peptide sequence, whereas the FliC sequence contains only tyrosine and phenylalanine residues.

Figure 14. Varian Eclipse™ fluorescence spectrophotometer (Varian, Palo Alto, CA).

Steady state fluorescence measurements for FlaA and FliC were performed with a Varian
Eclipse™ Fluorescence spectrophotometer (Varian, Palo Alto, CA) (Figure 14) equipped with a xenon flash lamp as light source and a four cuvette sample holder with thermoelectric temperature control. Different excitation wavelengths were used to collect the fluorescence emission spectra: 257 nm and 274 nm wavelengths are specific for excitation of phenylalanine and tyrosine, 280 nm for excitation of all aromatic amino acids and 295 nm is specific for exciting tryptophan residue. (http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk/projects/gmocz/fl uor.htm). Both emission and excitation slit widths were set to 5 nm. The experiments were performed at 25 °C with a 1 cm path length quartz cuvette.

Following this native scan, the emission and excitation maxima were recorded and were used as parameters for thermal denaturation studies. The thermal denaturation curve was recorded over a range of 25 – 98 to °C in the controlled temperature mode, where the FlaA sample was excited at 295 nm and emission was observed at 333 nm and FliC was excited at 280 nm and emission was observed at 305 nm. The change in fluorescence intensity and the shift in maximum emission wavelength were recorded to monitor the folding and unfolding transition of FliC and FlaA.\(^{42,43,58-61}\)

Specific emission due to the two tryptophan residues present in FlaA was obtained by selective excitation of FlaA monomer solution at 295 nm and 303 nm\(^{43}\) using a Shimadzu RF-5301 fluorescence spectrophotometer (Shimadzu Scientific Instruments, Columbia, MD) (Figure 15). The selective excitation should yield information regarding the extent of solvent exposure of tryptophan residues in the native, folded FlaA monomer.
2.11 Fluorescence anisotropy measurements

The steady state fluorescence anisotropy measurements of FlaA and FliC flagellin monomers as a function of denaturant concentration were measured with a Shimadzu RF-5301 fluorescence spectrophotometer equipped with a polarizer attachment (Figure 16). Samples of FliC and FlaA were excited by vertically polarized light at wavelengths of 280 nm and 295 nm, respectively. FlaA was specifically excited at 300 nm and the emission was viewed at 336 nm.

The emission intensity, $I$, of each sample was then measured by detection through a polarizer that was set to either parallel ($\parallel$) or perpendicular orientation ($\perp$) to the phase of the excitation light beam polarizer. The anisotropy, $r$, was calculated from the intensities $I_{\parallel}$ and $I_{\perp}$, using equation (1)

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$  (1)
Figure 16. Close-up view of interior sample chamber. Shimadzu RF-5301 fluorescence spectrophotometer (Shimadzu Scientific Instruments, Columbia, MD) equipped with a polarizer attachment.

Tryptophan has electronic absorption transitions $^1L_a$ and $^1L_b$. The specific excitation at 300 nm was performed to avoid complications of tyrosine to tryptophan energy transfer. At this longer excitation wavelength, only the $^1L_a$ state absorbs energy, which allows examination of the specific effects on emission from the $^1L_a$ state of tryptophan.

2.12 Circular dichroism spectroscopy measurements

The secondary structures of FlaA and FliC were characterized by circular dichroism (CD) spectroscopic measurements. CD spectra were recorded with an Aviv Model 202 Circular Dichroism Spectrometer equipped with a 450 watt Xenon arc lamp and thermostated cell holders (Figure 17). Protein concentrations of 0.2 mg mL$^{-1}$ and cuvettes
with a 0.2 cm path length were used for collection of far UV (180-260 nm) CD spectra.

**Figure 17.** Aviv Model 202 Circular Dichroism Spectrometer. Instrument equipped with a 450 watt Xenon arc lamp and thermostated cell holders.

The base-line was corrected by subtraction of the background spectra of protein-free buffer. The CD spectral data were analyzed with the K2D algorithm\(^{63}\), to estimate the percentages of \(\alpha\)-helix, \(\beta\)-sheet and random coil in each protein. Thermal melt experiments were performed using excitation wavelengths of 230 nm for FliC and 225 nm for FlaA over a range of 25 – 90 °C in the temperature mode, with a scan rate of 1 °C per min.

**2.13 Chemical denaturation of FlaA and FliC with protein titration protocol**

Chemical denaturation studies of FlaA with GdnHCl were also performed with the same Aviv Model 202 Circular Dichroism Spectrometer. Solutions of FlaA and FliC samples
were prepared in 20 mM, pH 7.5 phosphate buffer to give a final concentration of 0.2 mg/mL of protein and 5 mM sodium phosphate buffer. GdnHCl-protein solutions of FlaA and FliC were prepared using a stock solution of 8 M GdnHCl from Sigma-Aldrich, 20 mM, pH 7.5 sodium phosphate buffer and stock protein solutions. The final GdnHCl-protein solutions thus contained 7 M GdnHCl, 0.2 mg/mL protein and 5 mM sodium phosphate buffer. The instrument has a Microlab 500 series syringe pump (Hamilton Company, Reno, NV) that operates under the Aviv system control software to allow titration experiments in which the concentration of denaturant is varied. A volume of 15 mL of GdnHCl-protein solution in a 50 mL Falcon centrifuge tube was connected to the syringe pump. A volume of 2 mL of protein solution was placed in a quartz cuvette containing a magnetic stir bar. The parameters of titration were set as initial titrant concentration (GdnHCl denaturant) in the cuvette as 0 M and the final target titrant concentration of 6.6 M. Scans were performed for 1.0 sec collection time per 1.0 nm step with an increment of 0.2 M titrant per experiment. The stir time was set to 2.0 min after each titration step. A constant volume of 2.0 mL was maintained in the cuvette. The complete procedure was performed at 25°C.

2.14 Dynamic light scattering

The technique of dynamic light scattering (DLS) was used to determine the hydrodynamic radii of the soluble monomers of FlaA and FliC flagellin proteins, their aggregates and also the radii of FlaA and FliC flagellin monomers as a function of denaturant concentration. Measurements were made at 25°C. The data were analyzed to obtain the hydrodynamic radius of the sample. In DLS a monochromatic beam of light of 830 nm from the laser is used to illuminate the cuvette containing the sample and the
fluctuation of the intensity in the scattered light is detected by the photodiode close to 90°. The particles diffusing in and out of the path of light influences the scattering intensity measured by the detector. In the absence of any applied force, Brownian motion causes a change in the position of the particle. An autocorrelator analyzes the fluctuations in the scattering by performing a fourier analysis to these fluctuations giving it the first order autocorrelation function

\[ G(\tau) = 1 + \exp(-2D_Tq^2\tau) \]  

(2)

\( D_T \) is the translational diffusion coefficient of the molecule and \( q = (4\pi n/\lambda) \sin(\theta/2) \). The refractive index of the solvent \( n \) and \( \theta \) is the scattering angle and \( \lambda \) is the wavelength of the incident light. \( D_T \) is inversely proportional to the hydrodynamic radius and is obtained from the Stokes-Einstein equation (equation 3).

\[ D_T = \frac{k_B T}{6\pi \eta R_h} \]  

(3)

where \( k_B \) is the Boltzmann constant and \( T \) is the temperature in Kelvin. \( \eta \) is the viscosity of the solvent and \( R_h \) is the hydrodynamic radius of the scattering molecule.

Changes in the radii of the monomeric and aggregated proteins were observed as functions of seeding, temperature (16-85°C), pH (5.9 and 6.9) and salt concentrations (0.5 and 1.0 M). The pH was adjusted with 10 M NaOH and 1M HCl. Dynamic light scattering was performed with DynaPro Titan (Figure 18) and Wyatt QELS (Quasi-Elastic-Light-Scattering) instruments (Figure 19) (Wyatt Technology Corporation, Santa Barbara, CA). Stock solutions of monomeric FlaA and FliC were diluted to 0.51 and 0.56 mg/mL in 20 mM, pH 7.5 sodium phosphate buffer and passed through a 0.45 nm pore size low protein binding 13mm Whatman anodisc filter (Whatman Inc., Florham Park,
NJ) to remove any preformed protein aggregates. Buffer used for dilutions was passed through a 0.02 µm, 13-mm Whatman anodisc filter (Whatman Inc., Florham Park, NJ).

To determine the aggregation state of FlaA in solution, FlaA solutions with 0.5 M, 1 M NaCl were prepared and incubated at 16 °C, 25 °C, 37 °C, 65 °C and 85 °C. In addition a FlaA solution containing preassembled FliC flagella seed fragments was incubated at 16° C. Aggregates of FliC were prepared at varying pH values.

Aggregated solutions of FlaA and FliC were diluted to 0.01 and 0.02 mg/mL with 20 mM, pH 7.5 sodium phosphate buffer. Data from regularization was analyzed by ASTRA V and DynaLS software for the determination of size.

Figure 18. DynaPro Titan™ light scattering instrument (Dynamic light scattering) (Wyatt Technology Corporation, Santa Barbara, CA).
2.15 Polymerization of FlaA

The *in vitro* self assembly of FlaA into flagellar fibers in solution was investigated by using a variety of solvent conditions and promoters of protein nucleation and aggregation; these methods included techniques of salting out with ammonium sulfate, addition of nucleation-promoting flagella seeds and addition of polyethyleneglycol (PEG) polymers, either in the form of predissolved aqueous solutions and as a dry powder form.

**First approach:** Purified FlaA monomer solutions with 0.5 M and 1 M NaCl⁶⁴, ⁶⁵ at neutral pH were prepared and incubated at 16 °C, 25 °C, 37 °C, 65 °C and 85 °C.

**Second approach:** A purified FlaA monomer solution was incubated with FliC flagella seed fragments at 16 °C overnight.
Third approach: Polyethyleneglycol (PEG; \( \text{HOCH}_2(\text{OCH}_2\text{CH}_2)_n\text{CH}_2\text{OH} \)) a linear polymer of ethylene glycol monomers is available in a range of molecular weights. PEG was obtained from Sigma-Aldrich. A solution of FlaA monomer with 10% PEG 8000 (wt./vol.) and 1.25 M NaCl was prepared at pH 6 and pH 7. The samples of FlaA containing the PEG additive were incubated at 16 °C, 25 °C, 65 °C and 85 °C. Also samples of FlaA monomer containing only 10% PEG (no salt) at pH 7 were incubated at 16 °C, 25 °C, 65 °C and 85 °C.

Fourth approach: Self assembly studies of FlaA were undertaken in the presence of dry PEG.\(^{10, 66-69}\) PEG with molecular weights of 2000, 4000, and 6000 were previously shown to speed up polymerization of \( \text{Bacillus brevis} \) flagellin into flagella fibers.\(^{10}\) Dry PEG at concentrations of 10% and 20% (wt./vol.) was added to FlaA monomer (2.63 mg/mL) at pH 7.5 (Table 2). Samples were incubated at 20 °C for a minimum of 5 min. Control solutions with dry PEG at the same concentrations were prepared with 20 mM phosphate buffer (pH 7.5).

These polymerized FlaA samples were investigated with techniques of dynamic light scattering, electron microscopy and fluorescence microscopy.

Table 2. Polymerization of FlaA monomer with dry PEG

<table>
<thead>
<tr>
<th>FlaA monomer, mg/mL</th>
<th>PEG 2000, wt./vol.</th>
<th>PEG 4000, wt./vol.</th>
<th>PEG 6000, wt./vol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.06</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>4.06</td>
<td>20%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>
### 2.16 Fluorescence microscopy

Polymerized samples of FlaA prepared by the various approaches as described above in section 2.15 were analyzed by fluorescence microscopy. Fluorescence microscopy was performed with a Nikon TE 200U inverted fluorescence microscope (Figure 20). Samples of 50 µL FlaA polymer solution were stained with 2 µL of NanoOrange reagent from the NanoOrange® Protein Quantitation Kit (Molecular Probes, Invitrogen detection technologies, Catalog number N-6666, Eugene, OR). NanoOrange is merocyanine dye that produces an increase in fluorescence upon interaction with proteins. It is used for covalently labeling proteins. The images were obtained with Retiga EXi Fast 1394 a color cooled CCD camera, model: RET-EXi-F-CLR-12-C (QImaging, Burnaby, BC, Canada) attached to the microscope.

**Figure 20.** Nikon TE 2000U inverted fluorescence microscope.
2.17 Electron microscopy

The polymerized FlaA flagellin protein samples were prepared for imaging by transmission electron microscopy (TEM) in the following manner. Volumes of 10 µL of each polymerized flagellin sample were applied to carbon coated Formvar copper grids and negatively stained with 4% aqueous phosphotungstic acid (pH 7.5). TEM micrographs were recorded with a JEM-1230EXII electron microscope (JEOL, Tokyo, Japan) (Figure 21) located at the Western Michigan University Imaging Center and a JEOL 3011 High Resolution Electron Microscope (Figure 22) located at the University of Michigan.

Figure 21. JEOL, JEM-1230EXII transmission electron microscope. Located at Western Michigan University Imaging Center.
TEM micrographs were taken at an accelerating voltage of 80 kV for the JEM-1230EXII TEM and at 300 kV for the JEOL 3011 TEM. Images on the JEM-1230EXII electron microscope were captured as TIFF files by a 1 megapixel Gatan Bioscan digital camera (Gatan Inc., Pleasanton, CA). Images on the JEOL 3011 were captured as TIFF files by a Gatan 794 Slow Scan CCD TV system and Gatan 622SC image intensified TV rate camera.

Figure 22. JEOL 3011 high resolution transmission electron microscope. Located at the University of Michigan.
CHAPTER III

RESULTS AND DISCUSSION

3.1 Subcloning and sequencing

The 5' and 3' fragments of the flaA flagellin gene from *A. pyrophilus* in lanes 2 and 3 of an agarose gel are shown in Figure 23. These are the products of PCR amplification of 5' and 3' fragment inserts in the pRU1650 and pRU1651 plasmids, using primers described in Materials and Methods (Chapter 2). Lanes 4 and 5 show the full-length *flaA* gene fragment obtained by ligation of the two cohesive “sticky” ends of the PCR products in lanes 1 and 2, following restriction digestion with the HindIII restriction enzyme. It can be seen from the agarose gel image that the ligated product is about 1500 base pairs (bp) in size. The size of this cloned DNA fragment is consistent with the size of the *flaA* gene that was previously reported by Behammer et al.33

Following the ligation procedure to generate a full-length *flaA* gene, Ncol and SalI restriction sites were introduced near the 3' and 5' ends by performing PCR on the ligated *flaA* fragment with new primers. The purpose of this procedure was to insert new, unique “sticky end” restriction sites on both ends of the *flaA* gene fragment to facilitate directional cloning, i.e., insertion of the *flaA* gene into the multiple cloning sites of two protein expression plasmids, pET28c and pTnc99A.
Figure 23. Agarose gel electrophoresis of 5’ and 3’ fragments of the flaA flagellin gene from *A. pyrophilus*. Lane 1, 100 base pair DNA marker (Invitrogen). Lane 2, 737 base pair 3’ *flaA* gene fragment obtained by PCR amplification of pRU1650. Lane 3, 774 base pair 5’ *flaA* gene fragment obtained by PCR amplification of pRU1651. Lanes 4 and 5, ligation products of 3’ and 5’ *flaA* PCR fragments.

The PCR product and the two target plasmids were both simultaneously “double” digested with the Ncol and Sall restriction enzymes and the desired large molecular mass DNA fragment was purified with a PCR purification kit (Qiagen), prior to ligation with T4 ligase. Figure 24 shows the result of insertion of the *flaA* gene into the pTrc99A vector by ligation with T4 ligase and Figure 25 shows the result of insertion of the *flaA* gene into the pET28c vector by ligation. Insertion of the *flaA* gene into the two vectors was confirmed by performing PCR on the two plasmids with the same primers used to introduce the restriction sites for cloning and also by double digestion of both vectors with the Ncol and Sall restriction enzymes. Lane 6 in Figure 24 shows the 1.5 kbp fragment which was obtained by double digestion of pTrc99A and lane 7 in Figure 24 shows a fragment which is the product of PCR amplification of the inserted *flaA* gene in the pTrc99A plasmid. Both fragments were about 1500 base pairs in size, as expected.
Figure 24. Agarose gel electrophoresis analysis of flaA gene insertion into the pTrc99A S. typhimurium expression plasmid. Lane 1, 1 kb DNA marker (Invitrogen); Lane 2, uncut pTrc99A vector; Lane 3 and 4, Ncol and Sall digested pTrc99A vector and flagellin gene flaA; Lane 5, miniprep purified pTrc99A plasmid containing inserted flaA flagellin gene; Lane 6, Small and large DNA fragments of pTrc99A plasmid containing inserted flaA flagellin gene after digestion with Ncol and Sall restriction enzymes; Lane 7, PCR amplification product of inserted flaA in pTrc99A; Lane 8, 100 base pair DNA marker (Invitrogen). The two distinct bands observed for the miniprep plasmid in lane 5 are the supercoiled and relaxed “nicked” forms of the same circular plasmid DNA.

The pET28c-flaA plasmid was purified by a Qiagen miniprep kit procedure and was double-digested with Ncol and Sall for 3 hours. The resulting product was analyzed by gel electrophoresis (Figure 25), using a 1% agarose gel. Lane 2 shows the original pET28c plasmid without any insert. Lanes 3, 5, 7 and 9 in Figure 25 show the modified pET28c plasmid containing the flaA gene insert. These bands show a significant difference in their size and migration, compared to the original 5 kb plasmid band visible in lane 2. Lanes 4, 6, 8 and 10 show the resulting products obtained after double digestion of the pET28c-flaA plasmid with Ncol and Sall. These four lanes show the presence of much smaller DNA fragment with an approximate size of 1500 bp, corresponding to the expected size of the inserted flaA gene.
The previous methods of analysis of the flaA insert plasmids by restriction digestion and PCR amplification indicated that the flaA gene was correctly inserted into the pTrc99A and pET28c expression plasmids. However, the actual DNA sequences of the modified regions of each plasmid containing the putative flaA inserts were not known. Therefore, to confirm that the inserted DNA fragments contained the correct, full-length flaA DNA sequence, and that they were inserted in a correct, in-frame manner with respect to the plasmid expression promoter regions, the two modified pTrc99A-flaA and pET28c-flaA plasmids containing the flaA inserts were analyzed by DNA sequencing. This procedure was performed at the DNA Sequencing Core facility (University of Michigan, Ann Arbor, MI), using an ABI Model 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA).

![Agarose gel electrophoresis analysis of flaA gene insertion into the pET28c E. coli expression plasmid. Lane 1, 1 kb DNA marker (Invitrogen); Lane 2, pET28c plasmid; Lanes 3, 5 and 7, pET28c plasmid inserted with flaA flagelin gene; Lanes 4, 6 and 8, NcoI and Sall double-digested pET28c-flaA plasmid showing a 1500 base pair size DNA fragment containing the flaA gene.](image)

**Figure 25.** Agarose gel electrophoresis analysis of flaA gene insertion into the pET28c E. coli expression plasmid. Lane 1, 1 kb DNA marker (Invitrogen); Lane 2, pET28c plasmid; Lanes 3, 5 and 7, pET28c plasmid inserted with flaA flagelin gene; Lanes 4, 6 and 8, NcoI and Sall double-digested pET28c-flaA plasmid showing a 1500 base pair size DNA fragment containing the flaA gene.
A standard T7 forward primer was used to determine the 5' sequence of the putative \textit{flaA} insert in the pET28c plasmid. Although the capillary electrophoresis DNA sequencing results at this facility typically yielded reasonably long reads of 700-900 base pairs, this number of bases was not sufficient to read through the entire 1500 bp \textit{flaA} gene. Therefore, a second custom DNA sequencing primer, starting in the middle of the \textit{flaA} gene was designed and used for further DNA sequencing. This primer started at a location in the middle of the \textit{flaA} gene, located at 783 base pairs, and allowed DNA sequencing through the 3' end of the gene. The raw data sequencing chromatograms obtained from the DNA sequencing procedure were analyzed and checked for incorrectly assigned bases using Chromas 2.31 software (Technelysium Pty. Ltd., Tewantin, Australia). The final, complete \textit{flaA} nucleotide sequence was analyzed by NCBI-BLASTN software\textsuperscript{72}, which resulted in a close match with GenBank no. U17575. This corresponds to the \textit{flaA} flagellin gene of \textit{A. pyrophilus} that was originally reported by Behammer et al.\textsuperscript{33} No mutations, deletions or insertions were apparent in either of the \textit{flaA} sequences cloned into the pET28c and pTrc99A plasmids. Furthermore, the \textit{flaA} genes were inserted in-frame with respect to the T7 and Trc promoter regions in both plasmids. These sequencing results indicated that both modified plasmids were suitable for expression of the full-length FlaA flagellin protein in compatible bacterial strains, e.g., BL21(DE3) \textit{E. coli} and \textit{Salmonella typhimurium} SJW134 strains, as experimentally demonstrated in a later section.

The \textit{flaA} DNA sequencing data and the alignment results of the BLASTN search are attached in the appendix B.
3.2 Protein expression and purification

The *flaA* gene encoding the thermostable FlaA flagellin protein was previously cloned from the hyperthermophilic eubacterial species, *A. pyrophilus* into the pET28c and pTrc99A plasmids, as described in Materials and Methods. *E. coli* BL21(DE3) cells were transformed by electroporation with the pET28c plasmid containing the *flaA* gene and *S. typhimurium* SJW134 cells were electroporated with the pTrc99A plasmid containing the *flaA* gene. Overexpression of the FlaA protein was induced in both types of transformed bacterial cells by the addition 0.8 mM IPTG. Following expression and harvesting of the cells by centrifugation, the solubility of FlaA was investigated by cell lysis and SDS-PAGE analysis of the soluble supernatant and insoluble pellet. The harvested cells were lysed by treatment of cell pellet with Bug Buster reagent (Novagen), which contains non-ionic detergents that do not denature proteins. Figure 26 shows the FlaA expression obtained with the *E. coli* BL21(DE3) strain and Figure 27 shows FlaA expression obtained with *S. typhimurium* SJW134 cells.

The recombinant FlaA protein was expressed at very high levels in both bacterial strains, as indicated by the presence of a band of protein with a molecular mass of 55 kDa in SDS-PAGE gels of post-induction whole cell lysate prepared by heating post-induction cells with SDS-PAGE loading buffer and heating at 90 °C for 20 minutes. However, further analysis of the soluble versus insoluble cell fractions indicated that much of the expressed recombinant FlaA protein was produced in the form of insoluble inclusion bodies in both *E. coli* and *S. typhimurium* cells. This was indicated by the thick bands of expressed FlaA protein that are visible in lane 3 of Figure 26 and lane 2 of Figure 27. It is
also evident from the smaller amounts of soluble protein, apparent as lighter 55 kDa bands in Figure 26, lane 2 and Figure 27, lane 3, that a much larger fraction of the expressed recombinant protein was insoluble.

**Figure 26.** SDS-PAGE analysis of FlaA expression and solubility in *E. coli* BL21(DE3) cells (15% cross-linked acrylamide gel). FlaA has a computed molecular mass of 53.9 kDa. Lane 1, Mark12™ unstained standard molecular weight marker (Invitrogen). Lane 2, supernatant of BL21(DE3) cells transformed with pET28c-flaA plasmid after 3 hour induction with 0.8 µM IPTG. Lane 3, FlaA inclusion body pellet after 3 hour induction. Lane 4, Pellet of BL21(DE3) cell lysate before induction. Lane 5, supernatant of BL21(DE3) cell lysate before induction.

Inclusion bodies are a form of insoluble, misfolded aggregates of recombinant protein that are often observed for foreign proteins that were expressed at high levels in *E. coli* strains.\(^{73}\)\(^{74}\) The mechanisms by which these aggregates formed are unknown. The formation of inclusion bodies could be due to cellular stress which is associated with high level expression of recombinant protein.\(^{75}\)\(^{76}\)
Figure 27. SDS-PAGE analysis of FlaA expression and solubility in *S. typhimurium* strain SJW134 (15% cross-linked acrylamide gel). Lane 1, Mark12™ unstained standard molecular weight marker. Lane 2, FlaA inclusion body pellet after 3 hour induction. Lane 3, Soluble supernatant of SJW134 cells transformed with pTrc99A-*flaA* plasmid after 3 hour induction with 0.8 µM IPTG. Lane 4, insoluble pellet of SJW134 cells before induction. Lane 5, supernatant of SJW134 cells before induction.

Isolation of inclusion bodies from *E. coli* BL21(DE3) and *S. typhimurium* SJW134 cells, followed by solubilization and refolding of FlaA, was achieved with the use of the Protein Refolding Kit (Novagen). As given in the product manual, the Protein Refolding Kit uses a CAPS buffer at alkaline pH in combination with N-lauroylsarcosine detergent (included as a separate reagent) to achieve solubility of the inclusion bodies. This method of dissolving inclusion bodies in high pH buffer is one of several methods commonly used to solubilize inclusion bodies.\textsuperscript{77-80} N-lauroylsarcosine is a mild, biodegradable anionic surfactant derived from fatty acids and sarcosine, an amino acid; its presence increases the solubility of aggregates, yet it typically does not impair refolding and purification of recombinant proteins. After isolation of FlaA as inclusion bodies, they
were solubilized at high (alkaline) pH in the presence of 0.3% N-lauroylsarcosine detergent\(^8\). The solubilized aggregates were then dialyzed against 20 mM, pH 7.5 sodium phosphate buffer at 4 °C, to lower the pH of the FlaA solution and allow the FlaA protein to refold into a soluble form. At every step of isolation, solubilization and purification, the purity of FlaA was checked by analysis with SDS-PAGE. Lane 2 and lane 3 in Figure 28 correspond to the FlaA protein in solubilization buffer and dialysis buffer. The thick bands of solubilized and refolded FlaA protein are next to the 55.4 kDa marker (Mark12™ unstained standard, Invitrogen).

![SDS-PAGE analysis of solubilized and refolded FlaA flagellin protein prepared from inclusion bodies.](image)

**Figure 28.** SDS-PAGE analysis of solubilized and refolded FlaA flagellin protein prepared from inclusion bodies. Lane 1, Mark12™ unstained standard molecular weight marker. Lane 2, FlaA dissolved in alkaline pH 11 CAPS buffer. Lane 3, Soluble, refolded FlaA after dialysis for 15 hours in 20 mM, pH 7.5 phosphate buffer.

After dialysis of solubilized FlaA, a clear reduction of impurities was visible on the SDS-PAGE gel (Figure. 28, lane 3). FlaA obtained after dialysis was subjected to further
purification using size exclusion chromatography on a Superdex 75™ prep grade 26/60 column (GE Healthcare Bio-Sciences Corp. /Amersham Pharmacia, Piscataway, NJ). A protein sample volume of 20 ml was loaded onto the column and was eluted using 20 mM, pH 7.5 phosphate buffer. The flow rate was maintained at 0.5 mL/min.

![Elution of FlaA monomer](image)

**Figure 29.** Chromatogram of FlaA purification on a Superdex 75™ 60-70 cm size exclusion column run on an ÄKTA Fast Protein Liquid Chromatography instrument. The absorbance at 280 nm of the post-column eluted solution is indicated by the solid blue line; two elution peaks were observed, one centered at an elution volume of 100 ml (fractions A6, A7) and a second smaller elution peak centered at an elution volume of 160 ml (fractions B4-B7). The larger elution peak centered at 100 ml contained most of the soluble, refolded FlaA. The data plot was produced with UNICORN™ software.

Figure 29 shows an example chromatogram of FlaA purification by size exclusion chromatography on a preparative scale Superdex 75 column. The fractions at which there was an increase in the absorbance signal at 280 nm were collected and analyzed for the presence of FlaA using SDS-PAGE analysis. Figure 30 shows the analysis of the eluted
fractions by SDS-PAGE, in which lanes 2-6 of the gel correspond to fractions A5-A9 of the first elution peak and lanes 7-11 correspond to fractions B3-B7 of the second elution peak. Purified FlaA protein from fractions A6 and A7 was pooled and used for further biophysical characterization experiments.

![SDS-PAGE analysis of refolded FlaA purified by size exclusion chromatography on a Superdex 75™ column. Lanes 2-6 correspond to elution fractions A5-A9 from the first elution peak in Figure 29 and lanes 7-11 correspond to fractions B3-B7 from the second elution peak in Figure 29. FlaA from fractions A6 and A7 was pooled and used for further biophysical characterization.](image)

**Figure 30.** SDS-PAGE analysis of refolded FlaA purified by size exclusion chromatography on a Superdex 75™ column. Lanes 2-6 correspond to elution fractions A5-A9 from the first elution peak in Figure 29 and lanes 7-11 correspond to fractions B3-B7 from the second elution peak in Figure 29. FlaA from fractions A6 and A7 was pooled and used for further biophysical characterization.

The identity of the FlaA protein was confirmed by N-terminal sequencing. A sample of purified FlaA was submitted for N-terminal sequencing on an Applied Biosystems (Foster City, CA) Model 494 CLe at Michigan State University (East Lansing, MI). The resulting chromatograms of the fluorescently labeled products released at each cleavage step of the protein were analyzed with respect to a standard amino acid chromatogram and were compared to the published *A. pyrophilus* flagellin FlaA sequence (Swiss-Prot ID P46210). This analysis resulted in an N-terminal alanine (Ala) residue with a minor amount of methionine (Met) present, suggestive of incomplete cleavage of the N-terminal
methionine residue in some of the expressed protein; the methionine aminopeptidase enzyme catalyzes the removal of methionine from newly synthesized polypeptides in vivo\textsuperscript{82,83}. Following the alanine residue were threonine (Thr), arginine (Arg), isoleucine (Ile), asparagine (Asn) and tyrosine (Tyr) residues (Appendix 2). Post-translational modifications of proteins are known to occur in \textit{A. pyrophilus}; specifically, the N-terminal Met residue was shown to be cleaved from the mature form of FlaA.\textsuperscript{33} Similar observations were noticed in mesophilic expression of thermophilic FlaA. The chromatograms showed a larger proportion of Ala as the N-terminal amino acid residue, although some Met was present, as noted above. During overexpression of FlaA, trace amounts of protein may not have undergone this post-translational modification, due to an inability of the natural expressed levels of the methionine aminopeptidase enzyme\textsuperscript{84} to process all of the expressed protein. In the sequencing results Arg detection was low due to the fact that Arg might give lower yields. This is because the sample extraction is poor from the solid supports which results in such low yields on the chromatogram. (http://www.healthsystem.virginia.edu/internet/biomolec/seqguid4.cfm).

The terminal regions of bacterial flagellin sequences are highly conserved in nature. Earlier studies suggested a 54% similarity in N-termini and 55% similarity in C-termini between the \textit{A. pyrophilus} FlaA flagellin, and the \textit{S. typhimurium} FliC flagellin.\textsuperscript{33,34} A recent analysis of bacterial flagellin sequence homology by Beatson at al. indicated a very high degree of similarity in the N- and C-termini of 202 unique flagellins.\textsuperscript{16} This observed sequence conservation of flagellins is a consequence of the critical biological function of these regions. These sequence regions form several $\alpha$-helical coiled-coil secondary structure motifs that are a key part of the D0 and D1 domains; these domains
are essential for flagellin to self-assemble to form the helical flagella fibers. Considering the high sequence similarity, FlaA was expressed in non-motile mesophilic *S. typhimurium* strain SJW134 and the resulting cells were tested for motility on swarming agar plates.

The goal of this experiment was to test the thermostable FlaA flagellin protein for possible biological function; it might have the ability to be exported and self-assemble into functional flagella in mesophilic bacterial cells. The successful formation of flagella would render motility to the SJW134 cells. However, it was observed that the FlaA protein did not render motility in mesophilic *Salmonella* cells lacking their wild-type flagellin genes (Figure 8). Chaperones present in cytosol are responsible for the export of flagellin monomer. The incompatibility of mesophilic chaperones with the thermophilic flagellin, e.g. failure to bind FliS, and/or failure to be recognized by type III export complex at the base of the flagella structure, could be one reason for FlaA not rendering motility to *Salmonella* cells. The FlaA protein may also be unable to bind several rings of mesophilic adapter proteins at the base of the flagella that normally connect the bent hook structure to the terminal flagella fiber. The relatively low temperature of the mesophilic bacterial growth environment (37 °C) could also contribute to the observed failure of FlaA to form functional flagella; the low temperature may result in a lack of protein flexibility, i.e. a solid, wax-like physical state and result in the observed inability of FlaA to fold properly upon synthesis by ribosomes.
3.3 FliC and FlaA absorption and emission spectral characteristics

The UV absorption spectrum of soluble, monomeric FlaA was obtained with a Shimadzu UV-Vis spectrophotometer (Model: UV 1650PC). The baseline spectrum was determined by scanning a solution of 20 mM sodium phosphate buffer, pH 7.5. The FlaA protein concentration was 0.406 mg/mL. FlaA has a molecular weight of 53.9 kDa and contains 501 amino acids. FliC has a molecular weight of 51.4 kDa and contains 494 amino acids. The FlaA sample was scanned in the range of 200-400 nm. The spectrum showed an absorption maximum around 270 nm (Figure 31). This absorption due to phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) aromatic amino acid residues are sensitive to the environment present. Both FlaA and FliC are devoid of cysteine (Cys) residues which are sensitive to the environment to a lesser extent.

The thermostable *A. pyrophilus* FlaA flagellin protein has ~30% sequence homology with the mesophilic *S. typhimurium* FliC flagellin protein. In general structural proteins with ~30% sequence similarity are known to have structural similarity. Thus, a comparison of the biophysical properties of these two mesophilic and thermostable proteins can be made.

Proteins often exhibit intrinsic fluorescence; the specific fluorophores are the three types of aromatic amino acids Phe, Tyr and Trp. The FlaA amino acid sequence contains 10 Phe, 18 Tyr and 2 Trp residues. The FliC amino acid sequence consists of 6 Phe and 12 Tyr residues; unlike FlaA, the FliC sequence has no Trp residues. In all experiments, protein fluorescence was obtained by exciting at 280 nm or at longer wavelengths. At the excitation wavelength of 280 nm Phe has very low molar absorptivity, so its contribution
to total fluorescence is nil. This is because the quantum yield for Phe due to excitation at wavelength 280 nm is approximately 0.03. Therefore, the contribution from Phe is not considered further in the fluorescence experiments.

![UV absorbance spectrum of FlaA monomer (0.406 mg/ml)](image)

Figure 31. UV absorbance spectrum of FlaA monomer (0.406 mg/ml).

The absorption at 280 nm is primarily due to Tyr and Trp residues. The quantum yields of these residues are 0.14 and 0.13. Among the three aromatic amino acids, Trp is the dominant intrinsic fluorophore and its fluorescence is highly sensitive to its environment. Trp is known to absorb at wavelengths longer than 295 nm. Thus Trp can be selectively excited at 295-305 nm. It shows emission at a shorter wavelength around 317 nm when present in highly nonpolar environment or buried in the hydrophobic core of a folded globular protein. Depending on the polarity of the environment or exposure to solvent, Trp shows emission between 320-350 nm. The presence of Trp residues in FlaA
and absence of Trp residues in FliC results in a large difference in the observed fluorescence for excitation at 280 nm. The spectra in Figure 32 are due to fluorescence of FliC (a) and FlaA (b) at an excitation wavelength of 280 nm. The peak with a maximum of 304 nm in the FliC spectrum is characteristic for Tyr emission. The emission maximum observed at a higher wavelength of 324 nm for FlaA is due to the contribution of Trp residue emission. The emission from Tyr residues is not seen clearly due to the interference resulting from resonance energy transfer to Trp residues, which tend to function as the terminal acceptor of fluorescence energy emitted by Phe and Tyr residues, when present in a protein. Thus, in this study, the measurement of Tyr fluorescence was limited to FliC characterization and the measurement of Trp fluorescence was limited to FlaA characterization.

The presence of “natural protein fluorescence” i.e. intrinsic Trp residue fluorescence in FlaA makes it convenient to study the structural aspects of the monomer. Further characterization of FlaA was performed making use of the absorption and emission properties of tryptophan. Steady state fluorescence studies were performed by exciting a 0.406 mg/mL FlaA sample at 295-305 nm. This selective excitation was designed to separate the Tyr and Trp contributions to the total fluorescence emission.
The literature indicates that UV absorbance at wavelengths longer than 295 nm is only due to Trp\(^{43}\). Figure 33 shows the emission spectrum of FlaA at an excitation wavelength of 295 nm. This spectrum was obtained with a Shimadzu RF-5301 fluorescence spectrophotometer (Shimadzu Scientific Instruments, Columbia, MD) with the following instrument parameters: an instrument response time of 0.1 sec, excitation and emission slit widths of 5 nm and a “low” amplification voltage setting on the PMT detector. The sample was placed in a quartz cuvette with a pathlength of 10 mm. The scan was performed with a 1 nm emission wavelength step size at a constant temperature of 25 °C. The spectrum shows emission maxima at two wavelengths of 330 and 340 nm. This
indicates that both tryptophans are present in different environments in the FlaA protein structure. The Trp residues are 1.1 nm apart in the D3 domain.

![Fluorescence emission spectrum of FlaA monomer.](image)

**Figure 33.** Fluorescence emission spectrum of FlaA monomer. Samples excited at 295 nm, showing $\lambda_{\text{max}}$ at 330 nm with a shoulder at 340 nm. FlaA (0.406 mg/mL) dissolved in 20 mM phosphate buffer, pH 7.5 at 25 °C.

To more clearly characterize the environment of these two Trp residues, FlaA was selectively excited at 295 nm and 303 nm. Figure 34 shows the emission spectrum of FlaA for the above two excitation wavelengths. With excitation at 295 nm there was an emission maximum of 329 nm and with excitation at 303 nm there was an emission maximum of 339 nm, with the appearance of a shoulder at ~320 nm.
Figure 34. Fluorescence emission spectrum of FlaA monomer. Selective excitation at 295 nm and 303 nm showing \( \lambda_{\text{max}} \) at 329 nm and 339 nm in 20 mM phosphate buffer, pH 7.5 at 25 °C (FlaA 0.406 mg/ml).

These results indicate that one of the tryptophan residues is present in a more polar region when compared to the other residue in native, monomeric FlaA. Details of above explanation are tabulated in Table-3. The emission data were analyzed with respect to the positions of the two Trp residues in the predictive model of the folded FlaA structure (Figure 35). Figure 35a shows the extent of solvent exposed region of Trp 229, and Figure 35b shows that of Trp 254. The emission at longer wavelength is due to Trp 229 as it is in more polar region in native monomeric FlaA.
Figure 35. D3 domain of the predictive model of FlaA. Extent of solvent exposure of Trp residues (a) Trp 229, (b) Trp254.
Table 3. Excitation and emission wavelengths to monitor FliC and FlaA

<table>
<thead>
<tr>
<th>Flagellin Protein</th>
<th>Excitation wavelength (^a) (nm)</th>
<th>Emission wavelength (^b) (nm)</th>
<th>Intrinsic Chromophore</th>
</tr>
</thead>
<tbody>
<tr>
<td>FliC</td>
<td>280</td>
<td>304</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>FlaA</td>
<td>280</td>
<td>324</td>
<td>tyrosine and tryptophan</td>
</tr>
<tr>
<td>FlaA</td>
<td>295</td>
<td>330 and 340</td>
<td>Tryptophan</td>
</tr>
</tbody>
</table>

Concentration of FlaA monomer was 0.406 mg/mL, concentration of FliC monomer was 0.07 mg/mL in 20 mM phosphate buffer at 25 °C.

\(^a\) Excitation slit width of 5 nm.

\(^b\) Emission slit width of 5 nm.

3.4 Protein folding

The three dimensional structure of any protein is determined by its primary structure—amino acid sequence.\(^89\) There are different weak interactions involved in bringing the primary sequence together to form the three dimensional structure, including hydrogen bonds, electrostatic interactions, van der Waals interactions and the hydrophobic effect, thought to dominate the folding energetics of globular proteins.\(^90\) The stability of any protein depends on the kind of local interactions which holds the folded structure together. The biological function of a protein is also encoded by its folded structure. The unfolding pathway followed by proteins is often studied using chemical denaturants to gain an understanding of the molecular interactions that determine the folded structure and stability. During the unfolding process, proteins may undergo different conformational changes, before attaining a relatively “random coil”-like structure.\(^91\)
The protein unfolding/folding process is an equilibrium process. During this process some proteins pass through a stable intermediate with some elements of secondary structure present (termed a "molten globule") before they completely fold or unfold. In other cases the transition between the folded, native state and the unfolded, denatured state is a direct process. By exploiting biophysical properties of proteins such as fluorescence, anisotropy, circular dichroism and light scattering, the above mentioned equilibrium process can be monitored and understood. Each technique can yield different information about folded state of a protein. Fluorescence intensity and anisotropy can be used to monitor changes in polarity of local or overall structural regions of a protein. Light scattering (and gel filtration chromatography) measurements yield information about the size of intermediates in the unfolding process, while Secondary structural features, e.g., percentage of α-helix and β-sheet structures can be monitored by circular dichroism.

Globular proteins often have phenylalanine, tyrosine and tryptophan amino acid residues with aromatic side chains that can be excited by ultraviolet wavelengths of light to fluoresce. The amino acid Phe contains a benzene ring side chain, Tyr has a phenol (hydroxy benzene) side chain and Trp has an indole side chain (Figure 36). This property of intrinsic fluorescence is often exploited for biophysical measurements of protein folding and stability. It should be noted that the excitation energy can be silently transferred between aromatic residues located within a minimum distance of each other, without emission of photon, by Förster resonance energy transfer (FRET). In proteins this energy transfer process occurs in the hierarchical order from donor to acceptor: Phe → Tyr → Trp. The 494 amino acid residue FliC flagellin protein from S. typhimurium
(Swiss-Prot ID P06179) has a total of 6 Phe, 12 Tyr and no Trp residues. The 501 residue thermostable FlaA flagellin protein from *A. pyrophilus* (Swiss-Prot ID P46210) has 10 Phe, 18 Tyr and 2 Trp residues. Analysis of the experimentally determined FliC structure indicates that most of the aromatic residues are located in the outermost D2 and D3 domains. A predicted structure of the FlaA protein also indicates that the majority of the aromatic residues are also in the D2 and D3 domains, with the two Trp residues located in the outermost D3 domain.

\[
\text{HN-CH-C-OH} \quad \text{CH}_2 \quad \text{O} \\
\text{Phenylalanine, Phe (F)} \\
\]

\[
\text{CH}_2 \\
\text{OH} \\
\text{Tryosine, Tr (Y)} \\
\]

\[
\text{HN-CH-C-OH} \quad \text{CH}_2 \\
\text{Tryptophan, Trp (W)} \\
\]

**Figure 36.** Structures of three common aromatic amino used for intrinsic fluorescence studies of proteins. (a) Phenylalanine. (b) Tyrosine. (c) Tryptophan.
3.4.1 Fluorescence techniques to monitor folding of proteins

Fluorescence intensity has been widely used to characterize the structural changes that a protein undergoes in the presence of a denaturant. Fluorescence measurements were performed using 280 nm and 295 nm excitation wavelengths, which are the absorption maxima for tyrosine and tryptophan, as mentioned in section 3.3. Excitation at the higher wavelength of 295 nm was used for FlaA unfolding experiments because this wavelength excites only Trp residues and does not excite Tyr residues, and thus avoids any contribution from tyrosine residues. Samples of FlaA (0.406 mg/mL) and FliC (0.077 mg/mL) with varying concentrations of GdnHCl denaturant were incubated overnight at 4 °C to allow equilibrium states of folding/unfolding to be attained by both proteins. Long equilibration times were a necessary precaution; previous studies have shown that large proteins require longer incubation times, e.g., up to 18 hours.56 The overnight incubation time resulted in complete unfolding of both samples. Same samples were incubated at 4 °C for 180 days. No further change in the measured fluorescence anisotropy signal was observed when compared to samples that were incubated for 18 hours. This indicated that a minimum of 18 hours of incubation time was sufficient for complete unfolding to be achieved in the protein-denaturant mixture.

3.4.1.1 Fluorescence spectra of FlaA in the presence of GdnHCl

Samples of FlaA equilibrated with varying concentrations of GdnHCl (0-7.2 M) were excited at 280 nm in order to understand the contribution of both Tyr and Trp residues to the fluorescence signal. From figure 37 it can be seen that native, i.e., completely folded FlaA without any GdnHCl, showed a single peak with an emission maximum at 324 nm.
However, FlaA incubated with a high concentration of 6.75 M GdnHCl showed two fluorescence emission peaks, one at 306 nm and the other at 352 nm. The peak at 306 nm is very specific for Tyr and that at 352 nm is specific for Trp in a completely polar environment, as reviewed by Lakowicz. The native FlaA protein, when excited at 280 nm, showed an emission peak at 324 nm that was due to tryptophan interference.

The emission maximum of 324 nm indicated that the two Trp residue side chains were buried inside the folded FlaA structure, as might be expected due to the hydrophobic nature of these residues. The significant change in fluorescence of FlaA in 6.75 M GdnHCl is due to unfolding to yield a random coil polymer structure, with loss of FRET between the Tyr and Trp residues. The emission maximum at 352 nm is very specific for tryptophan residue which is present in a highly polar environment.

![Fluorescence emission spectra](image)

**Figure 37.** Fluorescence emission spectra of native FlaA at pH 7.5 (a), and FlaA unfolded in 6.75 M GdnHCl (b). Both samples (0.406 mg/mL) were excited at 280 nm.
Fluorescence unfolding studies of FlaA at 280 nm excitation were not performed because of the contribution of the other two aromatic amino acids, Trp and Tyr. Instead, an excitation wavelength of 295 nm was used for unfolding studies of FlaA; this wavelength selectively excites Trp residues but not Phe or Tyr residues. Figure 38 shows the fluorescence emission spectra of FlaA that were obtained with an excitation wavelength of 295 nm. The spectrum shows a maximum at 335 nm for the folded protein and when the tertiary structure was completely unfolded in 7.2 M GdnHCl, the wavelength of the emission maximum shifted to 356 nm. Further unfolding experiments were carried out with increasing concentrations of GdnHCl from 0 to 7.2 M.

Figure 38. Fluorescence emission spectra of native FlaA at pH 7.5 (a) and unfolded FlaA in 7.2 M GdnHCl (b). A concentration of 0.406 mg/mL of FlaA was excited at a wavelength of 295 nm.
The unfolding experiment in Figure 39 showed the following results when excited at 295 nm:

A. Native FlaA has a fluorescence maximum at about 329 nm that decreased in intensity and red shifted to 356 nm with treatment of 7.2 M GdnHCl.

B. The FlaA denaturation behavior was observed with intermediate concentrations of GdnHCl. Initially there was a decrease in the intensity with an increase in the GdnHCl concentration from 0 to 4.32 M. Along with the decrease in intensity there was a shift in the emission maximum from 335 to 341 nm.

Figure 39. Fluorescence spectrum of FlaA in the presence of GdnHCl (0, 1.44, 2.16, 2.88, 4.32, 5.04, 5.76, 6.12 and 7.2 M), conditions: 0.406 mg/mL FlaA, 25 °C, excitation at 295 nm, 5 nm excitation and emission slit widths, 20 mM phosphate buffer, pH 7.5. The spectra at 0 and 7.2 M GdnHCl were assumed to represent the spectra of the folded and unfolded states, respectively. (A) FlaA showing a decrease in fluorescence emission intensity with an increase of GdnHCl concentration from 0 to 4.32 M (—). (B) FlaA showing an increase in fluorescence emission intensity with an increase in GdnHCl
concentration from 4.32 to 7.2 M. These Spectra were recorded using the Varian fluorimeter.

C. With an increase of GdnHCl from 4.32 to 7.2 M, there was an increase in the FlaA fluorescence intensity and the fluorescence maximum showed a significant red shift to 356 nm.

D. Unfolding of FlaA resulted in increased exposure of two Trp residues to the polar aqueous environment, which was revealed by the red shift in fluorescence maximum.

**Figure 40.** FlaA fluorescence titration curves in the presence of increasing concentration of GdnHCl (0 to 7.2 M) at 317, 336, 356 and 380 nm emission wavelengths. Excitation wavelength of 295 nm. (Note: x-axis is a nonlinear scale)
The emission wavelengths at which the intensity differed to a maximum extent were chosen to investigate the unfolding pathway of FlaA. The intensity values obtained from the above fluorescence curves at different wavelengths were plotted against the concentration of GdnHCl. This resulted in titration curves which showed multiple transitions for FlaA. Figure 40 shows four such transition curves at 317, 336, 356 and 380 nm emission wavelengths.

The derivation of equilibrium constants for unfolding/folding equilibrium is discussed in a subsequent section (Section 3.4.5).

3.4.1.2 Comparison of fluorescence spectra of hyperthermophilic FlaA with mesophilic FliC flagellin

Fluorescence unfolding studies of FliC were monitored at 280 nm excitation. Due to the absence of Trp residues in FliC, only Tyr was used as a probe for unfolding studies. Figure 41 shows the emission maximum of native FliC and FliC unfolded in 6.75 M GdnHCl. A decrease in the fluorescence intensity was observed when FliC was completely unfolded. The intensity decreased until a concentration of 3.38 M GdnHCl was reached, after which it remained constant with further increases in the concentration of GdnHCl. Thus, the fluorescence emission signal of FliC was only recorded for 0-3.42 M GdnHCl.

The unfolding experiments of FliC in the presence of denaturant shown in figure 42 indicates that with an increase of GdnHCl concentration there was a gradual decrease in the emission intensity for FliC.
Figure 41. Fluorescence emission spectra of native FliC (0.077 mg/mL) at pH 7.5 (a), and of unfolded FlaA in 6.75 M GdnHCl b); both samples were excited at 280 nm.

Figure 42. The fluorescence emission spectrum of FliC in the presence of GdnHCl (0, 0.675, 1.35, 2.03 and 3.38 M), showing a decrease in intensity with increase of denaturant concentration. Experimental conditions: 0.077 mg/mL FliC, 25 °C, excitation at 295 nm, 5 nm slit widths, 20 mM phosphate buffer, pH 7.5. The spectra at 0 and 3.38 M GdnHCl were assumed to represent the spectra of the folded and unfolded states, respectively.
To analyze the unfolding pathway of FliC, the emission wavelengths at which the intensity differed to a maximum extent were chosen. The intensity values obtained from FliC fluorescence curves at these emission wavelengths were plotted against the concentration of GdnHCl. This resulted in titration curves which show a single transition. Figure 43 shows four transition curves at 295, 300, 307 and 320 nm emission wavelengths.

Figure 43. FliC titration curves in the presence of increasing concentration of GdnHCl (0 to 3.42 M) at 295, 300, 307 and 320 nm emission wavelengths. (Note: x-axis is a nonlinear scale)

3.4.2 Unfolding studies of FlaA and FliC using steady-state fluorescence anisotropy

The steady state fluorescence anisotropies of FlaA and FliC flagellin monomers were measured with a Shimadzu RF-5301 fluorescence spectrophotometer equipped with a
polarizer attachment. Samples of FliC and FlaA monomers were excited by vertically polarized light at wavelengths of 280 nm and 295 nm, respectively. The emission intensity, $I$, of each sample was then measured by detection through a polarizer that was set to either parallel ($\parallel$) or perpendicular ($\perp$) orientations to the phase of the excitation light beam polarizer, as given in equation 1.

Tyrosine and tryptophan residues were used as intrinsic molecular probes in the equilibrium unfolding study of the two flagellin proteins by anisotropy. The spectral properties of tryptophan are complex because the long wavelength absorption band (240-300 nm) consists of the two overlapping transitions. The two electronic absorption transitions $^1L_a$ and $^1L_b$ have vectors which are perpendicular to each other and are different in their directions, as shown in Figure 44. Thus, the anisotropy of a protein is dependent on the contribution of both electronic transition states. Only one of these two transitions, $^1L_a$, is sensitive to the polarity of the environment.

![Figure 44](image_url)

**Figure 44.** Electronic absorption transitions $^1L_a$ and $^1L_b$ of the tryptophan indole side chain.

The effects of GdnHCl on the anisotropy of FlaA and FliC are shown in Figures 45 and 46. Excitation wavelengths of 280 nm and 295 nm were used for the FliC and FlaA protein monomers in these anisotropy experiments.
Excitation at 300 nm was used to excite the FlaA sample for further unfolding and folding anisotropy experiments. Excitation at 300 nm does not excite tyrosine residues and also avoids any complications resulting from tyrosine to tryptophan energy transfer. The reason for this specific excitation at 300 nm was to examine the specific effects on emission from the $^1L_a$ state of tryptophan. At this longer excitation wavelength, only the $^1L_a$ state absorbs energy.

![Figure 45](image)

**Figure 45.** Fluorescence anisotropy of FlaA monomer as a function of GdnHCl concentration at 25 °C, excited at 295 nm. Protein concentration was 0.406 mg/mL. (Note: x-axis is a nonlinear scale)

Figure 47 shows the anisotropies obtained due to excitation at 295 nm and 300 nm for the FlaA monomer. Results from the specific excitation at 300 nm gave rise to titration curve which was similar to the titration curve obtained at excitation 295 nm.
**Figure 46.** Fluorescence anisotropy of FliC monomer as a function of GdnHCl concentration at 25 °C, excitation at 280 nm. Protein concentration was 0.077 mg/mL. (Note: x-axis is a nonlinear scale)

**Figure 47.** Fluorescence anisotropy of FlaA as a function of GdnHCl concentration at 25 °C, with excitation at 295 nm and 300 nm wavelengths. Protein concentration was 0.406 mg/mL. (Note: x-axis is a nonlinear scale)
3.4.3 Characterization of FlaA and FliC monomer unfolding and oligomerization by light scattering

Dynamic light scattering studies are useful for characterizing the size distribution of proteins in solution. The hydrodynamic radii of purified FlaA (0.406 mg/mL) and FliC (0.077 mg/mL) monomers were studied with DynaPro™ Titan and Wyatt QELS™ instruments. Both proteins were dissolved in 20 mM phosphate buffer, pH 7.5. The experiments were conducted at 25 °C. In both cases, 10 separate repeats of data acquisition were performed.

Analysis of the resulting light scattering data indicated that the FlaA monomer has a hydrodynamic radius, $R_H$, of ~ 12.8 nm and FliC has an $R_H$ value of ~ 13.6 nm $R_H$. Figures 48 and 50 show the size distributions determined for monomers of both flagellin proteins. Figure 49 shows the regularization fit for FlaA monomer.

![Figure 48. Light scattering of FlaA monomer. ~ 12.8 nm, determined by light scattering measurements. Protein concentration was 0.406 mg/mL.in 20 mM phosphate buffer at pH 7.5 and 25 °C.](image-url)
Figure 49. Correlation function for FlaA monomer.

Figure 50. Light scattering of FliC monomer, ~13.6 nm, determined by light scattering measurements. Protein concentration was 0.077 mg/mL, in 20 mM phosphate buffer at pH 7.5 and 25 °C.

A number of different conditions were tested to promote the formation of flagella fibers from refolded FlaA monomers. Aggregates of FlaA were prepared by heating in high salt
(NaCl) buffer to mimic the natural conditions under which this protein may self-associate into flagella fibers as explained in Materials and Methods. The aggregated form of FlaA showed a mean hydrodynamic radius of ~100 nm (Figure 51).

Unfolding studies of FlaA and FliC were performed as a function of concentration of denaturant; GdnHCl. Dynamic light scattering measurements of these proteins in the presence of increasing concentrations of GdnHCl were performed with the DynaPro instrument. The data was analyzed using Dynamics software (version 6.7.6). Figures 52 and 53 show the titration curves of the measured Stokes radius of FlaA and FliC as a function of GdnHCl concentration.

![Graph showing differential intensity fraction against hydrodynamic radius](image)

**Figure 51.** Light scattering of FlaA aggregates formed in 1 M NaCl at 85 °C for 1 hour with a protein concentration of 2.0 mg/mL.
Figure 52. Hydrodynamic radius of FlaA monomer as a function of GdnHCl concentration at 25 °C. The protein concentration was 0.406 mg/mL. (Note: x-axis is a nonlinear scale)

It can be seen from Figure 52 that the thermostable FlaA monomer undergoes unfolding as indicated by the increase in hydrodynamic radius through a stable intermediate, and two distinct transitions were observed. In the case of FliC, the unfolding behavior was different; FliC was completely unfolded without any apparent intermediate transition states. These light scattering results are in agreement with the previous fluorescence intensity and anisotropy behaviors of these proteins in the presence of GdnHCl. These results from light scattering depend only on the dynamic nature of protein in the surrounding solvent environment.
Figure 53. Hydrodynamic radius of FliC monomer as a function of GdnHCl concentration at 25 °C. The protein concentration was 0.077 mg/mL. (Note: x-axis is a nonlinear scale)

Figure 54. Translational diffusion coefficient of FlaA monomer with increasing concentration of GdnHCl.
The translational diffusion coefficient $D_T$ of FlaA in the solvent environment was determined using Equation 3. Figure 54 shows the extent of diffusion coefficient of FlaA with increasing concentrations of GdnHCl. With increase of GdnHCl changes occur in FlaA tertiary structure resulting in a decrease in the diffusion coefficient of FlaA.

### 3.4.4 Characterization of FlaA monomer with circular dichroism

Circular dichroism (CD) spectroscopy is an optical technique useful for determining whether a protein is in its native, folded conformation. CD can be used to study the different secondary structural features present in a folded protein. CD measures the difference in the absorption of left-handed polarized light and right-handed polarized light which arises due to asymmetry in the biomolecular structure. External factors such as pH, temperature, salts and denaturants will cause a change in the structural aspects of the folded protein. These changes can be detected from the signal obtained with the CD instrument.

The secondary structure of flagellin FlaA was determined by CD spectroscopy in the far-UV spectral region (190-250 nm) as mentioned in Materials and Methods (Chapter 2). The signal obtained was converted to mean residue weight ellipticity $\{\theta\}_{MRW}$ using Equation 4. Figure 55 shows the CD spectrum of FlaA obtained by plotting $\{\theta\}_{MRW}$ values against wavelength. This far-UV CD spectrum shows that native FlaA flagellin contains a substantial amount of secondary structure.

$$[\theta]_{MRW} = \frac{\theta_{obs}}{10 \times c \times \ell \times n} \quad (4)$$
where \( c \) is concentration of flagellin FlaA in mg/ml, \( \ell \) is path length and \( n \) is the total number of amino acids in FlaA.

**Figure 55.** CD spectrum for FlaA flagellin. Spectra obtained in 10 mM phosphate buffer (pH 7.5) containing 150 mM NaCl at 25 °C in a cell with a 1 cm path length. This spectrum was analyzed in the range of 200 to 240 nm to estimate the secondary structure by the method of the K2D program.

The mean residue weight ellipticity values in the range of 200 to 240 nm wavelength were analyzed to estimate the secondary structure, using the algorithm employed by the K2D program\(^6\). The result obtained from this analysis was that the FlaA flagellin monomer contained 48% \( \alpha \)-helix, 20% \( \beta \)-structure and 31% random coil structure.
Figure 56. Theoretical CD spectrum for FlaA flagellin monomer. CD spectrum was generated using a linear combination of the reference spectra in K2D program.

The percentages of α-helix and β-secondary structure obtained by the above analysis were again used with the K2D program to generate a theoretical CD spectrum using a linear combination of the reference spectra. Figure 56 shows the CD spectrum obtained as output from the combination of the reference spectra. The CD spectrum in Figure 56 shows minima at 208 nm and 220 nm. This indicates that flagellin FlaA has large fraction of α-helix secondary structure. This result is consistent with the results obtained in an earlier CD spectral analysis of *Salmonella* FliC flagellin.

CD spectroscopy is a very important tool in the study of protein folding. It allows us to understand the effect of denaturants on the secondary structure of a protein. The resulting data can be used to determine the relative stability of a protein in a particular external solvent environment. The major interactions that hold the protein structure together can
also be studied by this method. The Figure 57 shows the effect of GdnHCl on FlaA flagellin monomer as monitored by CD spectroscopy. This data was obtained from the titration of FlaA monomer with GdnHCl, the protocol for which has been previously explained in Materials and Methods (Chapter 2). From Figure 57 it can be interpreted that an increase in GdnHCl concentration results in the loss of all secondary structure in FlaA. It can also be concluded that the effect of denaturant on the FlaA structure is relatively weak, requiring high concentrations for unfolding. This could be due to the high stability of the thermostable *A. pyrophilus* FlaA folded structure, which evolved to function at extremely high temperatures.

![Figure 57](image)

**Figure 57.** FlaA flagellin monomer CD titration curve in the presence of increasing concentration of GdnHCl denaturant (0 to 6.8 M) monitored at 223 nm.

The effect of temperature on the thermostable FlaA monomer was studied by performing a thermal melt with the CD instrument. The effect of temperature on the CD signal was
monitored at 225 nm. Figure 58 shows the effect of temperature on FlaA. The temperature measurements were conducted in the temperature range 25-90 °C. There is no folding transition in this temperature range. This might suggest that the hyperthermophilic protein FlaA has a melting temperature of greater than 90 °C. This may also suggest that there are no transitions associated in the range of temperature studied. Further experiments with wider temperature ranges as well as in other conditions are required to confirm this condition. Increasing CD signal with temperature may indicate a complex behavior due to the thermal stability of FlaA. The thermal stability of FlaA is not surprising as the optimum growth temperature of *A. pyrophilus* is 85 °C. This thermal stability to temperature can be partially explained by the increase in the number of hydrophobic residues in the FlaA sequence and folded structure.

Since *A. pyrophilus* FlaA flagellin has ~30% sequence similarity with the mesophilic *Salmonella* FliC flagellin, the FlaA thermal melt data was compared with the FliC thermal melting curve obtained by Honda et al. From their analysis it was known that FliC had a melting temperature of 47 °C. This intermediate unfolding temperature is very characteristic for a mesophilic protein. When the behavior of FlaA was compared with FliC in terms of the effect of temperature, it becomes evident that FlaA is extremely stable to higher temperatures. It should be noted that at the melting temperature of the corresponding mesophilic protein FliC flagellin, FlaA retains most of its secondary structure.
Figure 58. CD thermal melting curve of flagellin FlaA (0.5 mg/ml) monitored at 220 nm.

3.4.5 Analysis of folding data obtained by different techniques

Some proteins structures may exhibit a simple two-state equilibrium between a native folded state and an unfolded random coil-like state. The conversion between these two states is a one step process. However, more complicated folding pathways may be observed for some proteins. For example, some proteins may exhibit a three-state folding pathway, where an intermediate partially unfolded “molten globule” state exists in equilibrium with both the native and unfolded states. In order to model the unfolding process as a two-step process, the population of any partially unfolded state needs to be considerably large and its signal should be different from that of the folded native and unfolded states.

This type of complicated behavior can be determined by the signal from fluorescence, anisotropy, circular dichroism and light scattering as a function of denaturant
concentration (GdnHCl). The folded to unfolded state can be obtained by analyzing the measured spectroscopic signal as a function of the change of denaturant concentration. The GdnHCl molecule is a well known chaotropic agent that disrupts the weak noncovalent interactions which are responsible for maintaining the native folded conformation of a protein.

The data obtained for FlaA folding using different techniques described earlier (sections 3.4.1 to 3.4.4) indicate the equilibrium in equation 5. FlaA has a highly populated intermediate in its unfolding pathway.

\[ A \rightleftharpoons B \rightleftharpoons C \]  \hspace{1cm} (5)

Here A is the folded, B is the partially unfolded and C is the completely unfolded state. A simpler equilibrium only involves folded and unfolded states A and B respectively as shown in equation 6.

\[ A \rightleftharpoons B \]  \hspace{1cm} (6)

From equation 5, two parameters such as the dissociation constant, \( K_d \) and Hill coefficient, \( n \) can be determined from denaturant titration curves. The transition midpoint is the concentration of denaturant (M) at which equal populations of both folded and unfolded species coexist. This transition mid point is related to the dissociation constant as shown in equation 7. The cooperative folding/unfolding transition can be obtained from the Hill coefficient, \( n \). Defining \( K_d \) as

\[ [\text{Den}_{1/2}]^n = K_d \]  \hspace{1cm} (7)
Here $[\text{Den}_{1/2}]$ is the concentration of denaturant at which equal concentrations of folded and unfolded species exist.

By the law of mass action, the dissociation constant and denaturant midpoint can be expressed as in equation 8.

$$K_d = (K_D)^n = \frac{[\text{Den}]^n[A]}{[B]} \quad (8)$$

This equilibrium constant $K_d$ represents the unfolded to folded equilibrium; when $[A] = [B]$, $K_d = [\text{Den}_{1/2}]^n$

The total signal observed at any point of the transition curve is due to the contribution of signal from initial state, $S_A$ and from the final state, $S_B$. The $S_A$ and $S_B$ values can be obtained from the transition curve by extrapolating the linear regions of the titration curves in the regions of low and high denaturant concentrations. In the above two-state mechanism at any given point in the titration curve the sum of the fractions of initial and final states is equal to 1, as given in Equation 9. The total signal observed in Equation 10 can be written as contributions from folded and unfolded species weighted by their fractions.

$$f_A + f_B = 1 \quad (9)$$

$$S_{\text{obs}} = S_A f_A + S_B f_B \quad (10)$$
The fractions initial \((f_A)\) and final \((f_B)\) can be determined from the signals of initial and final states obtained from techniques mentioned in sections 3.4.1 to 3.4.4.

\[
f_A = \frac{S_B - S_{obs}}{S_B - S_A} \quad f_B = \frac{S_{obs} - S_A}{S_B - S_A}
\]  

(11)

The equilibrium constant \(K_{eq}\) for the unfolded state is given by Equation 12.

\[
K_{eq} = \frac{[B]}{[A]}
\]  

(12)

From Equation 6, the equilibrium constant \(K_{eq}\) can be expressed in terms of the denaturant midpoint and dissociation constant.

\[
K_{eq} = \left(\frac{[Den]}{K_D}\right)^n
\]  

(13)

The signal as a function of denaturant for a single transition can be expressed as

\[
S_{obs} = S_A + \frac{[Den]^n}{[Den]^n + (K_D)^n}(S_B - S_A)
\]  

(14)

In case of a titration curve where two transitions are observed such as in Equation 1, the signal can be expressed as Equation 15

\[
S_{obs} = \frac{S_A + S_B(F_1) + S_C(F_1)(F_2)}{1 + F_1 + (F_1)(F_2)}
\]  

(15)
In the above equation, \( n_1 \) and \( K_{D1} \), \( n_2 \) and \( K_{D2} \) are the Hill coefficients and dissociation constants for the first and second transitions respectively. The signals obtained from initial, intermediate and final states are \( S_A \), \( S_B \) and \( S_C \). The values of \( n_1 \) and \( n_2 \) were determined from CD data by fitting the observed signal \( S \) and allowing all four parameters \( n_1 \), \( n_2 \), \( K_{D1} \) and \( K_{D2} \) to vary. The values of \( n_1 \) and \( n_2 \) were determined to be 3.5 and 12.1. These values were used to fit the observed signals in fluorescence intensity, anisotropy and dynamic light scattering using Equation 15. Data analysis was carried out using Origin software version 7.5.

Chemical denaturation studies of FlaA and FliC were also performed using intrinsic fluorescence, as described in Section 3.4.1 in Chapter 3. For the analysis of FlaA, the transition curves were obtained by plotting the fluorescence intensity at 317, 336, 356 and 380 nm emission wavelengths against increasing concentrations of GdnHCl. These curves were fit to Equation 15. The best fit was obtained for the titration curve obtained at 317 nm and is shown in Figure 59a. This curve clearly shows two transition midpoints. The corresponding data obtained for FliC was fit to Equation 14. The unfolding process of FliC follows a single transition, which can be seen in figure 59b. The values of the denaturant mid-points are tabulated in Table 3.
Figure 59. Fluorescence intensity titration curve for FlaA (0.406 mg/mL) (a) in the presence of increasing concentration of GdnHCl (0 to 7.2 M) at 317 nm emission wavelength showing two transitions. (b) Fluorescence titration curve for FliC (0.077 mg/mL) in the presence of increasing concentrations of GdnHCl (0 to 3.42 M) at 307 nm emission wavelength showing a single transition. (Note: x-axis is a nonlinear scale)
The fluorescence anisotropy of FlaA and FliC in response to increasing concentrations of GdnHCl also yielded titration curve data which were analyzed with Equations 15 and 14. Figures 60a and b show the fits for FlaA and FliC.

\[ K_{Gd1} = 1.06 \text{ M} \]
\[ K_{Gd2} = 5.1 \text{ M} \]

**Figure 60.** Fluorescence anisotropy of FlaA and FliC monomer as a function of GdnHCl concentration at 25 °C, excitation at 295 nm. The protein concentration of (a) FlaA was 0.406 mg/mL and (b) that of FliC was 0.077 mg/mL. (Note: x-axis is a nonlinear scale)
Figure 61. Hydrodynamic radius of FlaA monomer as a function of GdnHCI concentration at 25 °C. The protein concentration was 0.406 mg/ml. (b) Hydrodynamic radius of FliC monomer as a function of GdnHCl concentration at 25 °C. The protein concentration was 0.077 mg/ml. (Note: x-axis is a nonlinear scale)

The effect of GdnHCl on the hydrodynamic radius of the FlaA and FliC monomers was studied with dynamic light scattering. Analysis of the change in radius with denaturant
indicated two transitions for FlaA (Figure 61a) and a single transition for FliC (Figure 61b).

**Figure 62.** FlaA monomer CD titration curve in the presence of increasing concentration of GdnHCl (0 to 6.8 M). The CD signal was monitored at 223 nm. (a) x-axis is a nonlinear scale. (b) the x-axis has a linear scale.

The disruption of secondary structure with GdnHCl was monitored by far-UV CD signal at 223 nm. Titration of the FlaA monomer with GdnHCl gave rise to a two transition
titration curve. The CD data shows a cooperative unfolding transition between 0.4 and 2.4 M GdnHCl and another between 3.4 and 5.4 M GdnHCl (Figure 62). The analysis of this curve with Equation 15 yielded Hill coefficients $n_1$ and $n_2$ of 3.5 and 12.1, respectively. These values of $n_1$ and $n_2$ were kept constant for determining the denaturant midpoints.

The tertiary structure of FlaA was monitored by steady-state fluorescence spectroscopy. Sigmoidal plots are characteristic of cooperative unfolding transitions. This behavior was observed for both FlaA and FliC monomers that were denatured in GdnHCl solutions. Under native conditions, the FlaA fluorescence emission spectra showed a maximum intensity around 330 nm (Figure 38), indicating that both Trp residues are partially buried in hydrophobic environments in the folded protein. With increasing GdnHCl concentration, the observed decrease in emission intensity and shift of the spectrum emission peak to longer wavelengths implies that the two Trp residues are exposed to the polar solvent upon unfolding. The observation of more than one unfolding transition in the presence of GdnHCl suggests that FlaA contains more than one domain. This is in agreement with earlier studies of FliC monomer, which is known to be a multi domain protein with 4 distinct globular domains, termed D0, D1, D2 and D3. The results of the unfolding studies employing various biophysical techniques are summarized in Table 4. The results obtained by different methods show excellent agreement for the $K_{D1}$ and $K_{D2}$ values.
Table 4. Summary of fitted equilibrium constants obtained from analysis of folding studies of flagellin FlaA and FliC monomers

<table>
<thead>
<tr>
<th>Method</th>
<th>FlaA</th>
<th>FliC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( K_{Gdn1}, \text{M} )</td>
<td>( K_{Gdn2}, \text{M} )</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>1.8 ± 0.1</td>
<td>5.0 ± 0.05</td>
</tr>
<tr>
<td>Fluorescence</td>
<td>1.2 ± 0.01</td>
<td>5.1 ± 0.06</td>
</tr>
<tr>
<td>Anisotropy</td>
<td>1.3 ± 0.1</td>
<td>4.7 ± 0.07</td>
</tr>
<tr>
<td>Circular</td>
<td>2.2 ± 0.04</td>
<td>5.0 ± 0.08</td>
</tr>
<tr>
<td>dichroism</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Data obtained from single FliC monomer CD titration.

The mesophilic FliC monomer tends to unfold immediately upon exposure to even relatively low concentrations of GdnHCl. In contrast, the thermostable FlaA monomer requires the presence of significant amounts of GdnHCl to effect a significant change in the structure, as indicated by the titration curve patterns and by the change in emission maximum with an increase of GdnHCl concentration. In the titrations curves shown in Figures 59a, 60a, 61a and 62, it is apparent that some structural changes takes place, with attainment of a stable intermediate (un)founded state in the range of 3 – 4 M GdnHCl. Further increases in the concentration of GdnHCl resulted in a further change in the FlaA structure; a second transition was observed before complete unfolding of FlaA was attained at the highest concentrations of denaturant. A possible mechanism is shown in Figure 63.
The tertiary structural changes that occur in the unfolding pathway could be observed in the fluorescence emission studies. The initial decrease of intensity with a slight shift in emission maximum to a longer wavelength could be related to the disruption of interactions mainly in the D0 and D1 domains. But with an increase in GdnHCl concentration, a shift of the emission maximum to a much longer wavelength was observed. The occurrence of this characteristic red shift was probably because of the formerly buried Trp residues become exposed to the solvent upon unfolding of the D2-D3 domain region. These results indicate that the hydrophobic interactions within the hypervariable middle domain region (D2 and D3 domains) are disrupted, resulting in unfolding of the globular domain structure and complete exposure of the Trp residues to polar, aqueous solvent. Therefore, the FlaA unfolding pathway can be envisioned as the initial unfolding of the α-helical D0 and D1 domains, followed by unfolding of the hypervariable middle domain region, which encompasses the D2 and D3 domains in the FliC structure.

**Figure 63.** Proposed folding pathway of thermophilic FlaA flagellin monomer.
Table 4 shows the denaturant mid points for FlaA and FliC. From this table it can be seen that a small concentration of denaturant completely destabilizes FliC but a much higher concentration is required for FlaA. This higher chemical stability is consistent with the extreme environment in which *A. pyrophilus* lives. The high temperature, high pressure and potentially high salt environment in which FlaA functions requires that the interactions within FlaA should be very strong to keep the folded state of the monomer and resulting oligomeric flagella fibers. It was also observed that the FlaA monomer remains intact at the melting point of mesophilic FliC monomer; in fact the thermal melting temperature of FlaA was not experimentally accessible with the equipment used.

A theoretical model of FlaA$^{34}$ shows that there is an increase in the percentage of hydrophobic residues in the middle domain region. The increased thermal and chemical stability observed for FlaA in these folding studies are consistent with the previously noted increase in hydrophobic residues; increased hydrophobic interactions are a probable explanation for the increased stability of FlaA.

### 3.5 Polymerization of flagellin FlaA monomer

The flagellum is the primary organelle that provides propulsion in bacteria. The largest part of the flagellum, the elongated extracellular filament, is formed by the FliD/Hap2 chaperonin-mediated self-assembly of many thousands of flagellin monomers, e.g. 20,000 -30,000 monomers in one *S. typhimurium* flagella fiber. Flagella fibers may also be assembled *in vitro* by different approaches as described in Materials and Methods (Chapter 2). In the presence of high salt concentrations, the mesophilic FliC flagella monomer is known to polymerize spontaneously. Due to the sequence similarity of the
FlaA monomer with the FliC monomer, the process of FliC polymerization originally used by Asakura et al. was adapted and tried for polymerization of FlaA. FlaA monomer was incubated with 1 M NaCl at various temperatures, including higher temperatures approaching those at which the source organism, *A. pyrophilus*, typically lives (85 °C). These oligomerization experiments yielded the formation of helical protein structures for the samples incubated at 65 °C and 85 °C. These structures are visible in Figure 64 and 65, TEM images of negatively stained polymerized FlaA samples.

![Figure 64](image)

**Figure 64.** TEM image of polymerized of FlaA monomer in the presence of 1 M NaCl at 65 °C.

Following the dialysis of FlaA monomer during purification from inclusion bodies, the dialysis sample was centrifuged to eliminate any aggregates. The formation of protein aggregates during dialysis of refolded protein is common. This is because of the exposure of hydrophobic groups in unfolded protein. The pellet thus obtained was observed under a transmission electron microscope.
Figure 65. TEM image of polymerized FlaA monomer in the presence of 1 M NaCl at 85 °C.

Fibrous structures similar to those observed in Figure 66 were also observed in the dialysis pellet.

Figure 66. Electron micrograph of negatively stained FlaA aggregate obtained after refolding and dialysis procedure from high pH solution.
In order to confirm whether the protein structures formed were correctly assembled helical flagella fibers or misfolded β-amyloid aggregates, the FlaA oligomeric aggregate samples were stained with CongoRed dye. This dye is used in a characteristic method for staining amyloid-like proteins, which are entirely composed of stacked β-sheet structures. The samples of FlaA aggregate were stained with CongoRed (Sigma-Aldrich). When observed under a polarized microscope a red birefringence was observed (Figure 67). These observations suggested a possibility that polymerized FlaA fibers might be composed of β-amyloid aggregates.

Figure 67. Images of CongoRed stained FlaA aggregates under a polarized microscope.
Figure 68. Fluorescence images of NanoOrange stained polymerized FlaA in the presence of 20% PEG 4000 at room temperature and neutral pH. (Concentration of FlaA 4.06 mg/mL)
A number of published reports describe the self-assembly of bacterial and archaeal flagella from flagellin monomers in the presence of polyethylene glycol (PEG)\(^{10, 66, 69}\). When PEG was added to flagellin solutions in dry form, it increased the viscosity of the solution. Another function of PEG in solution is to absorb the solvent; it forms a sieve-like (gel-like) environment for the protein. The compounds PEG 2000, 4000 and 6000 were shown to catalyze the polymerization of flagellins; PEG polymers are also frequently used as “excluded volume” reagents to promote formation of protein crystals for use in X-ray diffraction studies.

The FlaA aggregates obtained in the presence of PEG were stained with NanoOrange dye and were viewed under a fluorescence microscope. The Figures 68 shows the structures formed in the presence of different concentrations of PEG. NanoOrange is a merocyanine dye that specifically binds and fluorescently labels proteins. The excitation wavelength for protein bound NanoOrange is 470-490 nm and it emits between 570-590 nm.
CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 The main conclusions from the research

1. Hyperthermophilic flagellin FlaA cloned from A. pyrophilus can be expressed in a mesophilic system. However, it does not export and assemble to form functional flagella, probably because it does not have the optimum temperature, pressure and appropriate chaperone proteins. Expression of FlaA in mesophilic bacteria resulted in formation of inclusion bodies, from which the purified protein was isolated and purified.

2. The sub-cloning of flaA into pET28c and pTrc99a plasmids resulted in recombinant vectors. The expression of protein from these vectors produced FlaA flagellin protein. DNA sequencing analysis of the insert in the recombinant vectors, along with N-terminal sequencing of the purified protein indicated the formation of the correct FlaA flagellin protein.

3. Attempts to self-assemble the purified FlaA monomers in vitro with 0.5 and 1 M NaCl in the temperature range 16-85 °C yielded globular aggregates that were probably not flagella fibers. TEM imaging and CongoRed polarization microscopy indicated that these aggregates were random globular and fibrous aggregates. The aggregation of FlaA monomer could be monitored by dynamic light scattering. Light scattering indicated that the hydrodynamic radius of the FlaA monomer was ~12.8 nm and that of the aggregate to be ~100 nm. Random aggregates were also detected by TEM. Staining with CongoRed and examining under the fluorescence microscope did not conclusively indicate the
presence of β-amyloid structures. These results indicated that while FlaA was isolated in a folded form, this structure was not optimum for self-assembly to yield flagella nanotubes, at least under the conditions studied.

4. Self-assembly of FlaA with polyethylene glycol (PEG) of different molecular weights yielded rod-like aggregates which need to be further characterized to determine if they are flagella.

5. Fluorescence spectroscopy of refolded FlaA monomers indicated the presence of two tryptophans in different environments. One Trp residue is present in a solvent accessible hydrophilic region (emission maximum: 339 nm) and the other Trp residue is present in comparatively solvent inaccessible hydrophobic region (emission maximum: 329 nm). In comparison the mesophilic FliC flagellin does not contain Trp residues.

6. Analysis of the circular dichroism spectrum of native FlaA monomer indicated that its structure contains 48% α-helix, 20% β-sheet and 31% random coil structures.

7. The stability of FlaA was determined by chemical denaturation studies. The unfolding of FlaA was examined by adding the denaturant guanidine hydrochloride (GdnHCl) in the concentration range 0-7.2 M. The unfolding of the protein was monitored by circular dichroism, fluorescence intensity, fluorescence anisotropy and dynamic light scattering. All of these biophysical studies indicated a two-step mechanism in which the folded FlaA partially unfolded at about 1 M GdnHCl, resulting in formation of a partially folded intermediate and completely unfolded at a GdnHCl concentration of 5 M. The FlaA, being a very stable protein, requires higher concentrations of denaturant to disrupt the
interactions that hold the tertiary structure together than the corresponding mesophilic FliC protein. The unfolding pathway of hyperthermophilic FlaA was quite different from mesophilic FliC which only exhibited folded and unfolded states.

8. Unfolding of FlaA was determined to be a cooperative process. Hill coefficients were derived from circular dichroism data to obtain information on cooperativity. The initial imbalance within the structure leads to structural changes in the total protein to form completely unfolded protein.

4.2 Future directions

There are a number of future directions in which this thermostable protein can be investigated.

1. Characterize the nature of rod like aggregates of FlaA that were formed in the presence of PEG. Characterize FlaA aggregates with thioflavin dye for differentiating helical fibers with beta-amyloid like fibers.99

2. Explore optimum conditions for the proper folding and aggregation of FlaA to form flagella. The conditions could include co-expression of FlaA with chaperones like FliS, and GroEL-GroES in a mesophilic system.

3. Explore the use of other anions such as sulfate, citrate, etc., which have been shown to promote aggregation of FliC and are used in crystallization trials of other proteins. One of these other ions could be useful for promoting FlaA aggregation. A number of protein crystallization and refolding kits with preformulated buffers and additives are
commercially available and could be used to explore different conditions with a “sparse matrix” experimental screening approach.

4. The environment in which each tryptophan is present in folded FlaA could be further determined by mutating individual tryptophans to other residues with different spectral properties, such as tyrosine.

5. Introduction of peptide loops in the variable D3 domain of FlaA by site directed mutagenesis towards forming bionanocomposites of hyperthermophilic flagella similar to mesophilic flagella and investigation of their fundamental properties.
REFERENCES


99. Omer, B. P.; Liu, L.; Murphy, R. M.; Kiessling, L. L., Phage Display Affords Peptides that Modulate <img src="http://pubs.acs.org/images/gifchars/beta2.gif">
APPENDIX A

DNA sequencing results

BLACK LETTERS – pTRc99a Plasmid

BROWN LETTERS - trc promoter from pKK233-2

BLUE LETTERS – flaA gene

HIGHLIGHTED SEQUENCES -

GREEN – Nco I site

RED - Sal I site

> pTH890 complete sequence

GTTCGACAGCTTTATCATCGAAGACTGCAACGGGTCGACCAATGCTTCTGCGTCA
GGCGACCCATCGGAGAAGCTGGT
GTATGGCTGTGCAGGTTCTGAGCTGTTGTCGCTGCTCAAGG
CGCACTCCCGTTCTGGAATAT
GTTTTTTGCAGCAGCAGATCATCATAACGGTTCTGGAATATCTCTGAAATGAGC
TGTTGACAAATTAATCATCCG
GCTCGTATAATGTGGAATTTGTGAGCGGAATACAAATTCACACAGGAA
ACAGA

CAGACGAGGATATATACGCTAATGACAGACTGCAACGGGTCGACCAATGCTTCTGCGTCA
GGCGACCCATCGGAGAAGCTGGT
GTATGGCTGTGCAGGTTCTGAGCTGTTGTCGCTGCTCAAGG
CGCACTCCCGTTCTGGAATAT
GTTTTTTGCAGCAGCAGATCATCATAACGGTTCTGGAATATCTCTGAAATGAGC
TGTTGACAAATTAATCATCCG
GCTCGTATAATGTGGAATTTGTGAGCGGAATACAAATTCACACAGGAA
ACAGA
GGGGTTTCGTGCACACAGCCAGCCAGTCTGGAAGCGAAACGACCTCAACCCGAACTGAG
ATACCTACAGCCTGAGCTATGAGAACCCGACGCCTTCCCCGAAGGGAGAAA
GGCGGACAGGTATCCGCTAAGCGGCGAGGGTTCGGAACAGGAGAGGCGACGAG
GGAGCTTCCAAGGGGAACGCGTGTATCTTATACGCTCTGTGTTGGTAGTACATGAT
ACCCGCTTTTGGATGAGCTGTGATAACCGCTGCCGCGACCCGAAACGACGAGC
AGCGAGTCAGTGAGCAGGAGAAGCGGAAAGGCCCTGATGCCTGTATTTCTCC
TTACGCATCTGCTGGGCGGTTATACCCCTGATTTCTGTAAGACAAATCTC
TGCTCTGATGCCGCAACGATTGTAACACTCCGCTATTACGAGTACGAGAC
TGGGTCTAGGGCAGCCCGGCTAGTGCGGTTAAGGGGTTTGACGACGCGGACG
AGATCCAGTCCGCGGCTAGCGGCAAGCGGAGGCCGTAAGAACACGACCATGT
GAATGCTGCAAAAAACCTTTCGCGGTATGCGATAGCGCCGAAAAAGAGAGTCA
ACATTCAACAACCGTTGCGAACAAACAACCTGCGGGCAAAACAGTCTGCTGAT
TGGCGGTTGCCACCTCAGTCTGGCCCTAGCCCGGTCAAGACGCGGGCATTGCAATC
CGATGAGCTGACGTGGCTGACGTCGCTGGCTGGCTGGCTGGCATAGCTGCGG
CGATGAAATCTCGCGCCGATCACCTGCGGCGTGAGCTGCTGCTGCTGCTGCTGCT
TGCGCGTCGCTGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
ACCAGACACCCATCAACAGTAATCTTTCTCCCATGAAAGCTGTTGCTGACTG
GGCGGTGAGCATCTCTGCTGCATTGGGTCAACGAAATCTCGCTGTATTAGCG
GGCCCCAATAGTTCTCTGCGCGCGCTCGCTGCTGGCTGGCTGGCTGGCTGCTG
TCTCAGTCGAATCAATCTGCGGAGATGGGAGAAGCGGCTAGGAGGTGAGGT
GGCATGCTGGCTTTCACAACAAACCAGCAAAATCTGGAATGAGGGCGTCTGTC
CCACTGCGAATGCTGGTGGTCCGAACAGATCAGATGGGGCTGGGCGCAAAGCGGC
CATTACCGAGTCCGGGTATTTGCGGATATCCTGCTGTAGGAGATAC
GACGATACCGAAGCAGACTCATGTTATATCCGCGCGTTAACACCACATCAAAC
AGGATTTCGCCCTGGGGCAAGCCAGCGTGAGCGCTTGGCAGCAACTTCTT
CAGGGCCACGCCGAGGGCAACATCGCTGGCAGCGCTTCCGGCTGAGGAGAA
GAAAAACACCTCTGGCCGACAAATCGGAAACGCCCTCTCCGCCCCGGCGCTTGGC
CGATTCAATATGCGACAGCGAGGTTTCCCGACTGGAAGACCGCCAGC
TGAGCGCAACCGCATTAAATGTTAGGTTACGCGGAAATTGATCTG
Results obtained from the BLASTN 2.2.14 for the sequenced \textit{flaA} gene insert in \textit{pTrc99A} plasmid

\texttt{> gi|596244|gb|U17575.1|APU17575  Aquifex pyrophilus Usf (usf) and flagellin (flaA) genes, complete
cds}

\texttt{Length=2803}

\begin{itemize}
  \item Score = 2985 bits (1506), Expect = 0.0
  \item Identities = 1506/1506 (100%), Gaps = 0/1506 (0%)
  \item Strand=Plus/Plus
\end{itemize}

<table>
<thead>
<tr>
<th>Query 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGGCAACGAGATTAATTACAACATATGAAGCAGCCGTTACTTTAATCTACCTTTAAAACAA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sbjct 1066</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGGCAACGAGATTAATTACAACATATGAAGCAGCCGTTACTTTAATCTACCTTTAAAACAA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 61</th>
</tr>
</thead>
<tbody>
<tr>
<td>AATGAGAGACTTATGAACAAGTCCCTCCTCAGGGCTTTCAACGGGACTAAGAATTTTATCC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sbjct 1126</th>
</tr>
</thead>
<tbody>
<tr>
<td>AATGAGAGACTTATGAACAAGTCCCTCCTCAGGGCTTTCAACGGGACTAAGAATTTTATCC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 121</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAGCGGATGACGCTTCGGGACTCTCTATAGCGGACCAGCTTTCCCTGTATCAACGGGA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sbjct 1186</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAGCGGATGACGCTTCGGGACTCTCTATAGCGGACCAGCTTTCCCTGTATCAACGGGA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 181</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCCAACAGGGAAACAGGAATATCCAGTTTGTATATCCGCACCTCCAGATTGCGGAGGGA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sbjct 1305</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCCAACAGGGAAACAGGAATATCCAGTTTGTATATCCGCACCTCCAGATTGCGGAGGGA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 241</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCTGGGCCAGTATACGACAAGCTAAAGACCATGTATCAAAAGCGGTAAGTGCCCGCG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sbjct 1365</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCTGGGCCAGTATACGACAAGCTAAAGACCATGTATCAAAAGCGGTAAGTGCCCGCG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query 301</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGCAGATTAACCGACCCCCAACGGAAGGGCGGCACCTCCAGAGGATATAGAAAAACCTGAGG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sbjct 1425</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGCAGATTAACCGACCCCCAACGGAAGGGCGGCACCTCCAGAGGATATAGAAAAACCTGAGG</td>
</tr>
</tbody>
</table>
Query 361
GACGCTATTCAGAAGATAGCGCAGGATACGGAGTATAACGGAATAAGGCTCTTAGACGGC 420

Sbjct 1426
GACGCTATTCAGAAGATAGCGCAGGATACGGAGTATAACGGAATAAGGCTCTTAGACGGC 1485

Query 421
ACACCTGGAACCGGTATTTACATACACTACCGAGCACGTTCAGAACAAAACTTAGTGTG 480

Sbjct 1486
ACACCTGGAACCGGTATTTACATACACTACCGAGCACGTTCAGAACAAAACTTAGTGTG 1545

Query 481
GCTATAGATGATGTTAGAGCTTCAAATTTAGGAGCTTATTTACTTGATTTGAAAGGGCA 540

Sbjct 1546
GCTATAGATGATGTTAGAGCTTCAAATTTAGGAGCTTATTTACTTGATTTGAAAGGGCA 1605

Query 541
TCAAATTCGCATATGATTCTCTTGCTAATCTGTTGACACAGACACAAATTTTGATATC 600

Sbjct 1606
TCAAATTCGCATATGATTCTCTTGCTAATCTGTTGACACAGACACAAATTTTGATATC 1665

Query 601
GCTTCCAGGAAACTATAAGCATAGCGGAGTAAACATCTCCACCCAACAAATATAGTA 660

Sbjct 1666
GCTTCCAGGAAACTATAAGCATAGCGGAGTAAACATCTCCACCCAACAAATATAGTA 1725

Query 661
ACGGATGCAAGGTATATTGCAGACTGGATAAACGGGGACCCGACTTTACAACAAATGGGA 720

Sbjct 1726
ACGGATGCAAGGTATATTGCAGACTGGATAAACGGGGACCCGACTTTACAACAAATGGGA 1785

Query 721
ATAAAGGCAAAGCTTCAAATAGACTGTAAGACCCCCCTGTTAAAATGTAGCTGTTGAT 780

Sbjct 1786
ATAAAGGCAAAGCTTCAAATAGACTGTAAGACCCCCCTGTTAAAATGTAGCTGTTGAT 1845

Query 781
ACAGGTGATTCCATTAACGGATAAAGTTCTATGTAGGAACAACTGCAATAACAAC 840

Sbjct 1846
ACAGGTGATTCCATTAACGGATAAAGTTCTATGTAGGAACAACTGCAATAACAAC 1905
Query 841
TTGACTTACGGTCCAGGTAAAACTATTACCTTGATAGATATTATAAGTGATATCAATTCA 900

Sbjct 1906
TTGACTTACGGTCCAGGTAAAACTATTACCTTGATAGATATTATAAGTGATATCAATTCA 1965

Query 901
AAAGCCACAGGTCTGAACTTAGTAGCCAAGGAAGAAAATGGAAGACTCGTTATTGAAACG 960

Sbjct 1966
AAAGCCACAGGTCTGAACTTAGTAGCCAAGGAAGAAAATGGAAGACTCGTTATTGAAACG 2025

Query 961
AGCGGCGAAACGGTAGGAGTAGAGGTTTCAAAATCTGGTACAGGAGGCACAACTTTTTCA 1020

Sbjct 2026
AGCGGCGAAACGGTAGGAGTAGAGGTTTCAAAATCTGGTACAGGAGGCACAACTTTTTCA 2085

Query 1021
TTAGACCAAATTATTTCTGGCGTAAACAAGACTGTGAACAATACAAACACCCAAGCATCA 1080

Sbjct 2086
TTAGACCAAATTATTTCTGGCGTAAACAAGACTGTGAACAATACAAACACCCAAGCATCA 2145

Query 1081
GCCATAAAAGTAGGTGATTTGAGAATTATTGATGATGAAATCATATACCTATGATTTCACG 1140

Sbjct 2146
GCCATAAAAGTAGGTGATTTGAGAATTATTGATGATGAAATCATATACCTATGATTTCACG 2205

Query 1141
GGTATTGCGGGTGCCTTTGGATTAACCAGCCCAAGCGGAACCGTAGGAATTACAGACCTT 1200

Sbjct 2206
GGTATTGCGGGTGCCTTTGGATTAACCAGCCCAAGCGGAACCGTAGGAATTACAGACCTT 2265

Query 1201
TACGCAATTGACGTTACCACAAATGAAGGGGCGGAGCTTGCGATGGATATTCTGACAATA 1260

Sbjct 2266
TACGCAATTGACGTTACCACAAATGAAGGGGCGGAGCTTGCGATGGATATTCTGACAATA 2325

Query 1261
GCCGCTCAAAAAGTTGATGAAATACGCTCCCAGATAGGCTCCACCATAATAAACCTCCAG 1320

Sbjct 2326
GCCGCTCAAAAAGTTGATGAAATACGCTCCCAGATAGGCTCCACCATAATAAACCTCCAG 2385
Query 1321
GCTATTTACGACGCTAAGGCTGTTGCCAAGGACAACACGGATAACGCGGAAAACATCATT 1380

Sbjct 2386
GCTATTTACGACGCTAAGGCTGTTGCCAAGGACAACACGGATAACGCGGAAAACATCATT 2445

Query 1381
AGGAATGTGGACTTTGCCAAGGAGATGACCGAGTTTACCAAGTATCAAATTAGGATGCAG 1440

Sbjct 2446
AGGAATGTGGACTTTGCCAAGGAGATGACCGAGTTTACCAAGTATCAAATTAGGATGCAG 2505

Query 1441
TCCGGTATAGCTATGCTTGCACAGGCAAATGCCCTACCTCAGCTCGTTCTCCAGCTTCTC 1500

Sbjct 2506
TCCGGTATAGCTATGCTTGCACAGGCAAATGCCCTACCTCAGCTCGTTCTCCAGCTTCTC 2565

Query 1501 AGGTGA 1506

Sbjct 2566 AGGTGA 2571
Results obtained from the BLASTN 2.2.14 for the sequenced *flaA* gene insert in pET28c plasmid

> gi|596244|gb|U17575.1|APU17575  Aquifex pyrophilus Usf (usf) and flagellin (*flaA*) genes, complete cds
Length=2803

Score = 1840 bits (928), Expect = 0.0
Identities = 1021/1046 (97%), Gaps = 8/1046 (0%)
Strand=Plus/Plus

Query 46
ATGGCAACGAGNNATTAATTACAACTATGAAGCAGCCGTTACTTATACTACTTTAAACA 105
|
Sbjct 1066 ATGGCAACGAG-ATTAATTACAACTATGAAGCAGCCGTTACTTATACTACTTTAAACA 1124

Query 106
AAATGGGAGACTTATGAACAAGTCCCTCCTCAGGCTTTCAACGGGACTAAGAATTTTATC 165
|
Sbjct 1125 AAATGAGAGACTTATGAACAAGTCCCTCCTCAGGCTTTCAACGGGACTAAGAATTTTATC 1184

Query 166
CGTAGCGGATGACGCTTCGGGACTCTTCATAGCGGACCAGCTTTCCCTCGTATCAAACGG 225
|
Sbjct 1185 CGCAGCGATGACGCTTCGGGACTCTTCATAGCGGACCAGCTTTCCCTCGTATCAAACGG 1244

Query 226
ACTTGAACAGGGAAACAGGAAATATCCAGTTTGTATACCTCAGAAGCCGAGATGCCGAGG 285
|
Sbjct 1245 ACTTCAACAGGGAAACAGGAAATATCCAGTTTGTATACCTCAGAAGCCGAGATGCCGAGG 1304

Query 286
AGGGTGGCCGGCGATATACGACAAGCTAAAGACCATGATCAAAAGGCGGTAAGTGCCGC 345
|
Sbjct 1305 AGGGTGGCCGGCGATATACGACAAGCTAAAGACCATGATCAAAAGGCGGTAAGTGCCGC 1364

Query 346
GAACGACATTAACGACCCCAACCGCAAGGGCGCAGCTCCAGAGGGATATAGAAAACCTGAG 405
|
Sbjct 1365 GAACGACATTAACGACCCCAACCGCAAGGGCGCAGCTCCAGAGGGATATAGAAAACCTGAG 1424
Query 406
GGACGCTATTCAGAAGATAGCGCAGGATACGGAGTATAACGGAATAAGGCTCTTAGACGG 465

Sbjct 1425
GGACGCTATTCAGAAGATAGCGCAGGATACGGAGTATAACGGAATAAGGCTCTTAGACGG 1484

Query 466
CACACTTAAGAACGGTATTTACATACACTACGGAGCACGTTCAGAACAAAAACTTAGTGT 525

Sbjct 1485
CACACTTAAGAACGGTATTTACATACACTACGGAGCACGTTCAGAACAAAAACTTAGTGT 1544

Query 526
GGCTATAGATGATGTTAGAGCTTCAAATTTAGGAGCTTATTTACTTGATTTGAAAGGGCA 585

Sbjct 1545
GGCTATAGATGATGTTAGAGCTTCAAATTTAGGAGCTTATTTACTTGATTTGAAAGGGCA 1604

Query 586
ATCAAAATTCCGCATATGATTCTTTTCTTAACCTGTGGACAACAGACACAAAAATTTGAAATT 645

Sbjct 1605
ATCAAAATTCCGCATATGATTCTTTTCTTAACCTGTGGACAACAGACACAAAAATTTGATAT 1664

Query 646
CGCTTCCACGGAAACTATAAGCATAGCCGGAGTAACACATTCTCCACCCAAACAATATAGT 705

Sbjct 1665
CGCTTCCACGGAAACTATAAGCATAGCCGGAGTAACACATTCTCCACCCAAACAATATAGT 1724

Query 706
AACGGATGCAAGGTATATTGCAGACTGGATAAACGGGGACCCGACTTTACAACAACCTGGG 765

Sbjct 1725
AACGGATGCAAGGTATATTGCAGACTGGATAAACGGGGACCCGACTTTACAACAACCTGGG 1784

Query 766
AATAAAGGCAAAAGCTTCAAATAGAGTCGTAGGAGACCCCTGGGTTAATGTAGCTGTTGA 825

Sbjct 1785
AATAAAGGCAAAAGCTTCAAATAGAGTCGTAGGAGACCCCTGGGTTAATGTAGCTGTTGA 1844

Query 826
TACAGGTGATTCATTAACGATAAAGTCTTCTATGTAGGAACACCTGCAATAAC 885

Sbjct 1845
TACAGGTGATTCATTAACGATAAAGTCTTCTATGTAGGAACACCTGCAATAAC 1904
Query 886
ATTGACTTACGGTCCAGGTAAACTATTACCTTGATAGAATTATAAGTGATATCAATTC 945

Sbjct 1905
ATTGACTTACGGTCCAGGTAAACTATTACCTTGATAGAATTATAAGTGATATCAATTC 1964

Query 946
AAAAGCCACACGNTCTGAACCTAGTAGCCAAGGAAGAAAATGGAAGACTCGTTNTTG-
AAG 1004

Sbjct 1965
AAAAGCCACACGNTCTGAACCTAGTAGCCAAGGAAGAAAATGGAAGACTCGTTATTGAAAC 2024

Query 1005
GAGCGGCG-AACNGTNGGAGTAGAGNTT--CAAATCTGGT-NNGGAGGC-
CCACTTTTTTC 1059

Sbjct 2025
GAGCGGCGAAACGGTAGGAGTAGGTTTCAAAATCTGGTACAGGAGGCACAACTTTTTTC 2084

Query 1060
NTTAGACCAAATT-TTNTGGCGTAA 1084

Sbjct 2085
NTTAGACCAAATT-TTNTGGCGTAA 2110
N-terminal sequencing results

Name: Sriyani Mikkamala
Sample ID: sm_uwnu_HRP-1
Date: 8/14/06

N-terminal Sequencing Results:

<table>
<thead>
<tr>
<th>A</th>
<th>T</th>
<th>I</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Standard 1, Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmol</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.32</td>
<td>D</td>
<td>c</td>
<td>9774</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6.79</td>
<td>N</td>
<td>c</td>
<td>8557</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.58</td>
<td>S</td>
<td>c</td>
<td>6323</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.94</td>
<td>Q</td>
<td>c</td>
<td>8792</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8.21</td>
<td>T</td>
<td>c</td>
<td>6328</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8.51</td>
<td>G</td>
<td>c</td>
<td>8681</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9.12</td>
<td>E</td>
<td>c</td>
<td>9843</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10.40</td>
<td>H</td>
<td>c</td>
<td>6845</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11.30</td>
<td>A</td>
<td>c</td>
<td>7101</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12.94</td>
<td>R</td>
<td>c</td>
<td>3856</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>13.94</td>
<td>Y</td>
<td>c</td>
<td>7516</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>16.59</td>
<td>P</td>
<td>c</td>
<td>7147</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>17.47</td>
<td>M</td>
<td>c</td>
<td>6570</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>17.83</td>
<td>V</td>
<td>c</td>
<td>7675</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>20.49</td>
<td>W</td>
<td>c</td>
<td>8317</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>21.30</td>
<td>F</td>
<td>c</td>
<td>7069</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>21.83</td>
<td>I</td>
<td>c</td>
<td>5978</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>22.15</td>
<td>K</td>
<td>c</td>
<td>7497</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>22.45</td>
<td>L</td>
<td>c</td>
<td>7002</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Residue 1, Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmol Ht</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.37</td>
<td>D</td>
<td>c</td>
<td>45541</td>
<td>4.68</td>
</tr>
<tr>
<td>3</td>
<td>6.88</td>
<td>N</td>
<td>c</td>
<td>28981</td>
<td>3.38</td>
</tr>
<tr>
<td>6</td>
<td>7.67</td>
<td>S</td>
<td>c</td>
<td>57136</td>
<td>9.04</td>
</tr>
<tr>
<td>7</td>
<td>8.04</td>
<td>Q</td>
<td>c</td>
<td>16623</td>
<td>1.91</td>
</tr>
<tr>
<td>8</td>
<td>8.31</td>
<td>T</td>
<td>c</td>
<td>43439</td>
<td>6.88</td>
</tr>
<tr>
<td>9</td>
<td>8.60</td>
<td>G</td>
<td>c</td>
<td>37016</td>
<td>4.52</td>
</tr>
<tr>
<td>10</td>
<td>9.22</td>
<td>R</td>
<td>c</td>
<td>15991</td>
<td>1.62</td>
</tr>
<tr>
<td>13</td>
<td>10.51</td>
<td>H</td>
<td>c</td>
<td>2013</td>
<td>0.29</td>
</tr>
<tr>
<td>15</td>
<td>11.39</td>
<td>A</td>
<td>c</td>
<td>156544</td>
<td>220.45</td>
</tr>
<tr>
<td>17</td>
<td>13.05</td>
<td>R</td>
<td>c</td>
<td>14210</td>
<td>3.69</td>
</tr>
<tr>
<td>20</td>
<td>14.02</td>
<td>Y</td>
<td>c</td>
<td>11306</td>
<td>1.50</td>
</tr>
<tr>
<td>25</td>
<td>16.67</td>
<td>P</td>
<td>c</td>
<td>16598</td>
<td>2.32</td>
</tr>
<tr>
<td>27</td>
<td>17.53</td>
<td>M</td>
<td>c</td>
<td>200441</td>
<td>30.51</td>
</tr>
<tr>
<td>28</td>
<td>17.89</td>
<td>V</td>
<td>c</td>
<td>15710</td>
<td>2.05</td>
</tr>
<tr>
<td>34</td>
<td>20.53</td>
<td>W</td>
<td>c</td>
<td>517</td>
<td>0.06</td>
</tr>
<tr>
<td>36</td>
<td>22.32</td>
<td>F</td>
<td>c</td>
<td>10564</td>
<td>1.49</td>
</tr>
<tr>
<td>37</td>
<td>22.88</td>
<td>I</td>
<td>c</td>
<td>12233</td>
<td>2.05</td>
</tr>
<tr>
<td>38</td>
<td>22.98</td>
<td>K</td>
<td>c</td>
<td>7098</td>
<td>0.95</td>
</tr>
<tr>
<td>39</td>
<td>22.49</td>
<td>L</td>
<td>c</td>
<td>33129</td>
<td>4.73</td>
</tr>
</tbody>
</table>
### Residue 2, Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmol</th>
<th>Mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6.37</td>
<td>D</td>
<td>c</td>
<td>60347</td>
<td>6.24</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6.88</td>
<td>N</td>
<td>c</td>
<td>71270</td>
<td>6.32</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.67</td>
<td>S</td>
<td>c</td>
<td>33374</td>
<td>5.28</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8.03</td>
<td>Q</td>
<td>c</td>
<td>59611</td>
<td>6.78</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8.31</td>
<td>T</td>
<td>c</td>
<td>1070172</td>
<td>169.15</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9.21</td>
<td>E</td>
<td>c</td>
<td>66317</td>
<td>6.74</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10.25</td>
<td>H</td>
<td>c</td>
<td>1459</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11.38</td>
<td>A</td>
<td>c</td>
<td>95020</td>
<td>13.38</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>13.04</td>
<td>R</td>
<td>c</td>
<td>14261</td>
<td>3.70</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>14.01</td>
<td>Y</td>
<td>c</td>
<td>33988</td>
<td>4.52</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>16.65</td>
<td>P</td>
<td>c</td>
<td>26430</td>
<td>3.70</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>17.52</td>
<td>M</td>
<td>c</td>
<td>21790</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>17.88</td>
<td>V</td>
<td>c</td>
<td>66601</td>
<td>8.68</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>20.53</td>
<td>W</td>
<td>c</td>
<td>4081</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>21.33</td>
<td>F</td>
<td>c</td>
<td>24452</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>21.86</td>
<td>L</td>
<td>c</td>
<td>68059</td>
<td>11.38</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>22.17</td>
<td>K</td>
<td>c</td>
<td>87786</td>
<td>11.38</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>22.48</td>
<td>L</td>
<td>c</td>
<td>107250</td>
<td>15.32</td>
<td></td>
</tr>
</tbody>
</table>

Applied Biosystems Precise - PROCISE-cLC
### Residue 3, Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmol Ht</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.38</td>
<td>D</td>
<td>c</td>
<td>71546</td>
<td>7.32</td>
</tr>
<tr>
<td>3</td>
<td>6.88</td>
<td>N</td>
<td>c</td>
<td>71655</td>
<td>8.37</td>
</tr>
<tr>
<td>6</td>
<td>7.67</td>
<td>S</td>
<td>c</td>
<td>37758</td>
<td>5.97</td>
</tr>
<tr>
<td>7</td>
<td>8.03</td>
<td>Q</td>
<td>c</td>
<td>53987</td>
<td>6.14</td>
</tr>
<tr>
<td>8</td>
<td>8.30</td>
<td>T</td>
<td>c</td>
<td>192719</td>
<td>30.45</td>
</tr>
<tr>
<td>9</td>
<td>8.60</td>
<td>G</td>
<td>c</td>
<td>98712</td>
<td>12.07</td>
</tr>
<tr>
<td>10</td>
<td>9.21</td>
<td>E</td>
<td>c</td>
<td>56634</td>
<td>5.65</td>
</tr>
<tr>
<td>13</td>
<td>10.51</td>
<td>H</td>
<td>c</td>
<td>1894</td>
<td>0.28</td>
</tr>
<tr>
<td>16</td>
<td>11.38</td>
<td>A</td>
<td>c</td>
<td>100437</td>
<td>14.14</td>
</tr>
<tr>
<td>18</td>
<td>12.02</td>
<td>R</td>
<td>c</td>
<td>114049</td>
<td>29.58</td>
</tr>
<tr>
<td>22</td>
<td>14.00</td>
<td>Y</td>
<td>c</td>
<td>43318</td>
<td>5.76</td>
</tr>
<tr>
<td>28</td>
<td>16.65</td>
<td>P</td>
<td>c</td>
<td>12796</td>
<td>4.59</td>
</tr>
<tr>
<td>30</td>
<td>17.51</td>
<td>N</td>
<td>c</td>
<td>16290</td>
<td>2.48</td>
</tr>
<tr>
<td>31</td>
<td>17.88</td>
<td>V</td>
<td>c</td>
<td>77364</td>
<td>10.08</td>
</tr>
<tr>
<td>37</td>
<td>20.52</td>
<td>W</td>
<td>c</td>
<td>5019</td>
<td>0.60</td>
</tr>
<tr>
<td>40</td>
<td>21.33</td>
<td>F</td>
<td>c</td>
<td>33932</td>
<td>4.80</td>
</tr>
<tr>
<td>41</td>
<td>21.85</td>
<td>I</td>
<td>c</td>
<td>101251</td>
<td>16.94</td>
</tr>
<tr>
<td>42</td>
<td>22.17</td>
<td>K</td>
<td>c</td>
<td>71112</td>
<td>9.49</td>
</tr>
<tr>
<td>43</td>
<td>22.47</td>
<td>L</td>
<td>c</td>
<td>152833</td>
<td>21.80</td>
</tr>
</tbody>
</table>
Residue 4, Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmol Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6.39</td>
<td>D</td>
<td>c</td>
<td>88466</td>
<td>9.05</td>
</tr>
<tr>
<td>4</td>
<td>6.88</td>
<td>N</td>
<td>c</td>
<td>88845</td>
<td>10.38</td>
</tr>
<tr>
<td>6</td>
<td>7.68</td>
<td>S</td>
<td>c</td>
<td>44741</td>
<td>7.08</td>
</tr>
<tr>
<td>7</td>
<td>8.04</td>
<td>Q</td>
<td>c</td>
<td>58652</td>
<td>6.67</td>
</tr>
<tr>
<td>8</td>
<td>8.31</td>
<td>T</td>
<td>c</td>
<td>122103</td>
<td>19.45</td>
</tr>
<tr>
<td>9</td>
<td>8.60</td>
<td>G</td>
<td>c</td>
<td>92891</td>
<td>11.35</td>
</tr>
<tr>
<td>10</td>
<td>9.22</td>
<td>R</td>
<td>c</td>
<td>69926</td>
<td>7.10</td>
</tr>
<tr>
<td>13</td>
<td>10.60</td>
<td>H</td>
<td>c</td>
<td>2036</td>
<td>0.30</td>
</tr>
<tr>
<td>16</td>
<td>11.38</td>
<td>A</td>
<td>c</td>
<td>109092</td>
<td>15.36</td>
</tr>
<tr>
<td>19</td>
<td>13.02</td>
<td>K</td>
<td>c</td>
<td>106372</td>
<td>27.59</td>
</tr>
<tr>
<td>22</td>
<td>14.01</td>
<td>Y</td>
<td>c</td>
<td>56877</td>
<td>7.57</td>
</tr>
<tr>
<td>29</td>
<td>16.65</td>
<td>P</td>
<td>c</td>
<td>33618</td>
<td>4.70</td>
</tr>
<tr>
<td>32</td>
<td>17.52</td>
<td>M</td>
<td>c</td>
<td>13271</td>
<td>2.04</td>
</tr>
<tr>
<td>33</td>
<td>17.88</td>
<td>V</td>
<td>c</td>
<td>84036</td>
<td>10.35</td>
</tr>
<tr>
<td>38</td>
<td>20.52</td>
<td>W</td>
<td>c</td>
<td>8664</td>
<td>1.04</td>
</tr>
<tr>
<td>41</td>
<td>21.33</td>
<td>F</td>
<td>c</td>
<td>56926</td>
<td>8.05</td>
</tr>
<tr>
<td>42</td>
<td>21.85</td>
<td>I</td>
<td>c</td>
<td>1035302</td>
<td>173.19</td>
</tr>
<tr>
<td>43</td>
<td>22.14</td>
<td>X</td>
<td>c</td>
<td>83478</td>
<td>11.13</td>
</tr>
<tr>
<td>44</td>
<td>22.47</td>
<td>L</td>
<td>c</td>
<td>153468</td>
<td>21.92</td>
</tr>
</tbody>
</table>
Residue 5. Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmol Mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.34</td>
<td>D</td>
<td>c</td>
<td>183187</td>
<td>15.67</td>
</tr>
<tr>
<td>3</td>
<td>6.83</td>
<td>N</td>
<td>c</td>
<td>895492</td>
<td>104.65</td>
</tr>
<tr>
<td>4</td>
<td>7.61</td>
<td>S</td>
<td>c</td>
<td>53915</td>
<td>8.53</td>
</tr>
<tr>
<td>5</td>
<td>7.97</td>
<td>Q</td>
<td>c</td>
<td>95866</td>
<td>10.90</td>
</tr>
<tr>
<td>6</td>
<td>8.24</td>
<td>T</td>
<td>c</td>
<td>148497</td>
<td>23.47</td>
</tr>
<tr>
<td>7</td>
<td>8.53</td>
<td>G</td>
<td>c</td>
<td>143710</td>
<td>17.57</td>
</tr>
<tr>
<td>8</td>
<td>9.13</td>
<td>K</td>
<td>c</td>
<td>109713</td>
<td>11.15</td>
</tr>
<tr>
<td>13</td>
<td>10.68</td>
<td>H</td>
<td>c</td>
<td>5601</td>
<td>0.82</td>
</tr>
<tr>
<td>15</td>
<td>11.29</td>
<td>A</td>
<td>c</td>
<td>173734</td>
<td>24.47</td>
</tr>
<tr>
<td>18</td>
<td>12.92</td>
<td>R</td>
<td>c</td>
<td>40180</td>
<td>10.42</td>
</tr>
<tr>
<td>21</td>
<td>13.89</td>
<td>Y</td>
<td>c</td>
<td>72289</td>
<td>9.62</td>
</tr>
<tr>
<td>28</td>
<td>16.52</td>
<td>P</td>
<td>c</td>
<td>43492</td>
<td>6.09</td>
</tr>
<tr>
<td>30</td>
<td>17.38</td>
<td>M</td>
<td>c</td>
<td>20960</td>
<td>3.19</td>
</tr>
<tr>
<td>31</td>
<td>17.75</td>
<td>V</td>
<td>c</td>
<td>115143</td>
<td>15.00</td>
</tr>
<tr>
<td>36</td>
<td>20.38</td>
<td>W</td>
<td>c</td>
<td>6062</td>
<td>0.73</td>
</tr>
<tr>
<td>39</td>
<td>21.18</td>
<td>F</td>
<td>c</td>
<td>65751</td>
<td>9.30</td>
</tr>
<tr>
<td>40</td>
<td>21.71</td>
<td>I</td>
<td>c</td>
<td>342039</td>
<td>57.22</td>
</tr>
<tr>
<td>41</td>
<td>22.02</td>
<td>K</td>
<td>c</td>
<td>85366</td>
<td>11.39</td>
</tr>
<tr>
<td>42</td>
<td>22.32</td>
<td>L</td>
<td>c</td>
<td>191173</td>
<td>27.30</td>
</tr>
</tbody>
</table>
### Residue 6, Interpolated baseline

<table>
<thead>
<tr>
<th>Peak No</th>
<th>RT</th>
<th>Peak ID</th>
<th>Type</th>
<th>Height</th>
<th>Pmfl Ht</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.33</td>
<td>D</td>
<td>c</td>
<td>129304</td>
<td>13.23</td>
</tr>
<tr>
<td>3</td>
<td>6.82</td>
<td>N</td>
<td>c</td>
<td>306439</td>
<td>35.81</td>
</tr>
<tr>
<td>4</td>
<td>7.61</td>
<td>S</td>
<td>c</td>
<td>51633</td>
<td>8.17</td>
</tr>
<tr>
<td>5</td>
<td>7.87</td>
<td>Q</td>
<td>c</td>
<td>104600</td>
<td>11.90</td>
</tr>
<tr>
<td>6</td>
<td>8.24</td>
<td>T</td>
<td>c</td>
<td>150154</td>
<td>23.73</td>
</tr>
<tr>
<td>7</td>
<td>8.53</td>
<td>G</td>
<td>c</td>
<td>155903</td>
<td>19.06</td>
</tr>
<tr>
<td>8</td>
<td>9.13</td>
<td>E</td>
<td>c</td>
<td>108807</td>
<td>11.05</td>
</tr>
<tr>
<td>9</td>
<td>9.69</td>
<td>H</td>
<td>c</td>
<td>5534</td>
<td>0.81</td>
</tr>
<tr>
<td>10</td>
<td>11.30</td>
<td>A</td>
<td>c</td>
<td>224676</td>
<td>31.64</td>
</tr>
<tr>
<td>11</td>
<td>12.94</td>
<td>R</td>
<td>c</td>
<td>35586</td>
<td>9.23</td>
</tr>
<tr>
<td>12</td>
<td>13.91</td>
<td>Y</td>
<td>c</td>
<td>1107025</td>
<td>147.29</td>
</tr>
<tr>
<td>13</td>
<td>16.56</td>
<td>F</td>
<td>c</td>
<td>44866</td>
<td>6.28</td>
</tr>
<tr>
<td>14</td>
<td>17.43</td>
<td>M</td>
<td>c</td>
<td>23339</td>
<td>3.55</td>
</tr>
<tr>
<td>15</td>
<td>17.80</td>
<td>V</td>
<td>c</td>
<td>144050</td>
<td>18.77</td>
</tr>
<tr>
<td>16</td>
<td>20.45</td>
<td>W</td>
<td>c</td>
<td>4504</td>
<td>0.66</td>
</tr>
<tr>
<td>17</td>
<td>21.25</td>
<td>T</td>
<td>c</td>
<td>75691</td>
<td>10.71</td>
</tr>
<tr>
<td>18</td>
<td>21.78</td>
<td>I</td>
<td>c</td>
<td>213209</td>
<td>38.67</td>
</tr>
<tr>
<td>19</td>
<td>22.09</td>
<td>K</td>
<td>c</td>
<td>90613</td>
<td>12.09</td>
</tr>
<tr>
<td>20</td>
<td>22.40</td>
<td>L</td>
<td>c</td>
<td>214649</td>
<td>30.66</td>
</tr>
</tbody>
</table>


