6-2007

Structure-Function Studies of Self-Assembling Flagellin Proteins from Salmonella typhimurium and Aquifex pyrophilus

Venkata Raghu Ram Malapaka
Western Michigan University

Follow this and additional works at: http://scholarworks.wmich.edu/dissertations

Part of the Biochemistry, Biophysics, and Structural Biology Commons, and the Biology Commons

Recommended Citation
http://scholarworks.wmich.edu/dissertations/893

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
STRUCTURE-FUNCTION STUDIES OF SELF-ASSEMBLING FLAGELLIN
PROTEINS FROM SALMONELLA TYPHIMURIUM AND AQUIFEX
PYROPHILUS

by

Venkata Raghu Ram Malapaka

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Biological Sciences
Dr. Brian C. Tripp, Advisor

Western Michigan University
Kalamazoo, Michigan
June 2007
ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Brian C. Tripp, a wonderful mentor I had during my tenure at Western Michigan University. Dr. Tripp always valued my creative spirit and had a sense of faith in me. His guidance in terms of designing a research project and writing manuscripts have been invaluable. He is always willing to try novel ideas and never afraid of going in a new direction. His critical analysis of my thoughts gave me tremendous confidence in pursuing various ideas during my graduate study. I sincerely believe the various skills that I have learned under him will help me launch my scientific career in a successful way.

I would also like to thank my committee; Dr. Bruce Bejcek, Dr. John Geiser, and Dr. Yirong Mo. Their commitment towards their students and colleagues is exemplary. Their suggestions and thoughtful insights have given me an opportunity to critically think about my work and approach it in a rational way. I would like to especially thank Dr. Subra Muralidharan for his financial and intellectual support during my study at WMU. I would like to thank all my lab members for maintaining a healthy work environment in the lab.

I would like to thank my father, Mr. Subrahmanyam, for providing me with the best possible means to fulfill my dreams. He has successfully played the role of both the parents after my mother’s demise and made numerous sacrifices for us. I would like to thank my wife Sandhya, for being a wonderful and patient companion. She has been a pillar of strength for me during my study. I would like to acknowledge

ii
my beloved sisters Vandana and Shilpa, with out whom my life wouldn’t be complete.

Finally, I would like to dedicate this work to my mother Suseela and my grandmother Mahalakshmi, who are not here today to share my happiness, but will always be there in our thoughts.

Venkata Raghu Ram Malapaka
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................................... ii
LIST OF TABLES ........................................................................................... x
LIST OF FIGURES .......................................................................................... xi

CHAPTER

I. INTRODUCTION TO THE DISSERTATION .............................................. 1

References ................................................................................................. 13

II. INVESTIGATION OF THE HYPERVERSIBLE DOMAIN REGION
OF THE SALMONELLA TYPHIMURIUM FLAGELLIN PROTEIN
IN MOTILITY .......................................................................................... 20

Abstract ................................................................................................... 20

Introduction ............................................................................................... 21

Materials and Methods ................................................................................ 25

Bacterial strains ......................................................................................... 25

Plasmids and construction of flagellin internal
deletion variant library ................................................................................ 26

Motility assays ........................................................................................... 31

Colony PCR ............................................................................................... 32

DNA sequence determination and analysis ................................................. 33

Flagellin protein expression, flagella fiber
purification and SDS-PAGE ......................................................................... 33

Electron microscopy ................................................................................... 34

Light microscopy and swimming velocity analysis ...................................... 35

iv

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# Table of Contents – Continued

## CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement of power required to stall cells using optical tweezer</td>
<td>35</td>
</tr>
<tr>
<td>Structure analysis and homology modeling</td>
<td>38</td>
</tr>
<tr>
<td>Results</td>
<td>38</td>
</tr>
<tr>
<td>Description of flagellin hypervariable domain structural features</td>
<td>38</td>
</tr>
<tr>
<td>Identification of functional fliC internal deletion variants</td>
<td>39</td>
</tr>
<tr>
<td>Analysis of functional deletion variant polypeptide sequences</td>
<td>42</td>
</tr>
<tr>
<td>Purification and transmission electron microscopy analysis of flagella</td>
<td>70</td>
</tr>
<tr>
<td>Swarming agar motility assay</td>
<td>74</td>
</tr>
<tr>
<td>Video analysis of swimming motility in dilute aqueous media</td>
<td>77</td>
</tr>
<tr>
<td>Measurement of stall force in viscous aqueous media by optical trapping</td>
<td>78</td>
</tr>
<tr>
<td>Comparison of motility results determined by different methods</td>
<td>80</td>
</tr>
<tr>
<td>Discussion</td>
<td>81</td>
</tr>
<tr>
<td>Effects of internal deletions on cellular motility and stability of flagella</td>
<td>81</td>
</tr>
<tr>
<td>Implications for folding, export and assembly</td>
<td>83</td>
</tr>
<tr>
<td>Implications for design of flagellin fusion proteins</td>
<td>85</td>
</tr>
</tbody>
</table>
# Table of Contents – Continued

## CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenBank accession numbers</td>
<td>85</td>
</tr>
<tr>
<td>References</td>
<td>88</td>
</tr>
<tr>
<td>Appendices</td>
<td>94</td>
</tr>
<tr>
<td>A. DNA sequences of the \textit{fliC} gene deletion variants</td>
<td>94</td>
</tr>
<tr>
<td>B. A multiple sequence alignment of all the 39 unique internal deletion variants of FliC</td>
<td>132</td>
</tr>
</tbody>
</table>

### III. A THEORETICAL MODEL OF \textit{AQUIFEX PYROPHILUS} FLAGELLIN: IMPLICATIONS FOR ITS THERMOSTABILITY

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>133</td>
</tr>
<tr>
<td>Introduction</td>
<td>134</td>
</tr>
<tr>
<td>Computational Methods</td>
<td>138</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>141</td>
</tr>
<tr>
<td>Sequence comparison</td>
<td>142</td>
</tr>
<tr>
<td>Structure comparison</td>
<td>154</td>
</tr>
<tr>
<td>Conclusions</td>
<td>169</td>
</tr>
<tr>
<td>References</td>
<td>170</td>
</tr>
</tbody>
</table>

### IV. INVESTIGATION OF INSERTION OF GREEN FLUORESCENT PROTEIN IN \textit{SALMONELLA TYPHIMURIUM} FLAGELLIN PROTEIN TO CREATE A FUNCTIONAL BIONANOTUBE

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>177</td>
</tr>
</tbody>
</table>
### Table of Contents – Continued

**CHAPTER**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials and Methods</td>
<td>183</td>
</tr>
<tr>
<td>Bacterial strains</td>
<td>183</td>
</tr>
<tr>
<td><em>fliC</em> and GFP transposon plasmids</td>
<td>184</td>
</tr>
<tr>
<td>Transposase reaction</td>
<td>185</td>
</tr>
<tr>
<td>Removal of Xmal restriction site from pTH890 plasmid</td>
<td>187</td>
</tr>
<tr>
<td>Screening of the GFP inserts</td>
<td>189</td>
</tr>
<tr>
<td>Removal of <em>Kan</em>&lt;sup&gt;R&lt;/sup&gt; gene</td>
<td>190</td>
</tr>
<tr>
<td>Screening for functional insertions using bacterial motility assay</td>
<td>190</td>
</tr>
<tr>
<td>Results</td>
<td>191</td>
</tr>
<tr>
<td>Transposase reaction</td>
<td>191</td>
</tr>
<tr>
<td>Screening for the fluorescence</td>
<td>192</td>
</tr>
<tr>
<td>Removal of <em>Kan</em>&lt;sup&gt;R&lt;/sup&gt; gene</td>
<td>193</td>
</tr>
<tr>
<td>Motility screening</td>
<td>195</td>
</tr>
<tr>
<td>Discussion</td>
<td>195</td>
</tr>
<tr>
<td>References</td>
<td>198</td>
</tr>
</tbody>
</table>

**V. EXPERIMENTAL INVESTIGATION OF THE HYPOTHETICAL**

*S. TYPHIMURIUM* **FLAGELLIN MECHANICAL SWITCH BY INSERTION OF A DISULFIDE BRIDGE**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>205</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>207</td>
</tr>
</tbody>
</table>
Table of Contents – Continued

CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strains and plasmid</td>
<td>207</td>
</tr>
<tr>
<td>Construction of cysteine disulfide bond</td>
<td>208</td>
</tr>
<tr>
<td>Partial purification of fibers and SDS-PAGE</td>
<td>208</td>
</tr>
<tr>
<td>Ellman’s assay</td>
<td>209</td>
</tr>
<tr>
<td>Motility assay</td>
<td>210</td>
</tr>
<tr>
<td>Video microscopy and swimming analysis</td>
<td>210</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>212</td>
</tr>
<tr>
<td>References</td>
<td>213</td>
</tr>
<tr>
<td>VI. HIGH-THROUGHPUT SCREENING FOR ANTIMICROBIAL COMPOUNDS</td>
<td>215</td>
</tr>
<tr>
<td>USING A 96-WELL FORMAT BACTERIAL MOTILITY ABSORBANCE ASSAY</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>215</td>
</tr>
<tr>
<td>Introduction</td>
<td>216</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>217</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>219</td>
</tr>
<tr>
<td>References</td>
<td>232</td>
</tr>
<tr>
<td>VII. RATIONAL DESIGN OF METAL BINDING SITES AND TESTING</td>
<td></td>
</tr>
<tr>
<td>THE FEASIBILITY OF CATALYTIC SITES IN FLAGELLIN</td>
<td>233</td>
</tr>
<tr>
<td>Introduction</td>
<td>233</td>
</tr>
</tbody>
</table>
Table of Contents – Continued

CHAPTER

Materials and Methods ................................................................. 237

Bacterial strains ........................................................................ 237

Design of the metal binding sites .............................................. 238

Site-directed mutagenesis ....................................................... 238

Partial purification of fibers and SDS-PAGE ....................... 241

Dialysis of mutants ................................................................ 241

Enzyme assay ......................................................................... 242

Results and Discussion .............................................................. 242

References .............................................................................. 243

VIII. FUTURE STUDIES ................................................................. 246

Study of the conformational state of Salmonella typhimurium flagellin protein during type III export and assembly into flagella ... 246

Self-assembly of GFP-flagellin fusion proteins in vitro .......... 248

Conformational study of various flagellin deletion variants using circular dichroism and fluorescence spectroscopy ......... 249

Engineering of catalytic sites in flagellin protein .................. 249

Insertion of single domain antibodies in flagellin hypervariable region ................................................................. 250

APPENDICES .................................................................................. 251

A. Recombinant DNA Biosafety Approval .............................. 251

B. Letters and Permission .......................................................... 258
**LIST OF TABLES**

2.1 Strains and plasmids used in this study ................................................................. 48

2.2 Sequence parameters of functional internal deletion variants .......................... 51

2.3 Summary of deleted secondary structures internal deletion variants............ 59

2.4 Motility of *S. typhimurium* SJW134 cells expressing FliC and deletion variants as determined by different methods .................................................. 75

2.5 The deletion variant plasmids and their Genbank accession numbers ......................... 86

3.1 Amino acid composition of the full length FliC and FlaA ......................... 145

3.2 Computed parameters of FlaA and FliC flagellin proteins and their N-terminal, C-terminal and middle hypervariable domain regions .......... 150

3.3 Amino acid composition (percentages) of N-terminus, C-terminus and middle domains of FliC and FlaA ................................................................. 151

3.4 Solvent accessible surface area (SASA) calculations for conserved D0 and D1 domains of FlaA and FliC ................................................................. 163

5.1 Ellman’s assay of the wild-type and cysteine mutants ........................................ 212

6.1 Complete list of antibacterial compounds detected in screen of Genesis Plus library with offset motility detection method .............................. 228

7.1 Parameters used for the construction of metal binding sites in FliC protein ................................................................. 236

7.2 Mutagenesis primers ............................................................................................. 240

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LIST OF FIGURES

1.1 General structure of eubacterial flagella ............................................................... 2
1.2 A cartoon representation of *S. typhimurium* FliC .................................................. 9
2.1 An Electron Micrograph of *Salmonella typhimurium* ........................................... 22
2.2 A pictorial representation of pTH890 plasmid ....................................................... 27
2.3 Construction of fliC gene deletion library ............................................................ 30
2.4 A schematic representation of a laser tweezer ...................................................... 37
2.5 DNA agarose gel of the pTH890 deletion library .................................................... 41
2.6 Motility assay ......................................................................................................... 42
2.7 Colony PCR of the deletion variants ................................................................. 43
2.8 Multiple alignment of FliC and internal deletion variants ...................................... 50
2.9 Homology models of flagellin variants ............................................................... 56
2.10 Plot of normalized relative cell motility ............................................................. 67
2.11 TEM images of flagella fibers .............................................................................. 72
2.12 SDS-PAGE of the FliC and deletion variant proteins ........................................ 73
2.13 Motility assay of internal deletion library ........................................................ 75
3.1 A pairwise sequence alignment of FliC and FlaA ...................................................... 143
3.2 A comparison of 3-D structures of FliC and FlaA ............................................... 153
3.3 Ramachandran plot of the FlaA model ............................................................... 155
3.4 A 3-D structural alignment of FliC with the model of FlaA ................................ 156
3.5 Alignment of the secondary structures of FliC and FlaA ...................................... 161
List of Figures – Continued

3.6 Structure of a flagellin pentamer .......................... 167
4.1 Crystal structure of a dimeric GFP (PDB:1GFL) ........... 181
4.2 A map of the pTH890 plasmid ................................. 186
4.3 A map of the pBNJ11.7 plasmid ............................... 186
4.4 Insertion of GFP using transposase enzyme ................ 188
4.5 Colony PCR of the fluorescent inserts ...................... 192
4.6 Gel picture of the removal of Kanamycin gene ............ 194
4.7 Ligation of the fliC-GFP inserts ............................... 194
5.1 A pictorial depiction of the mutated Cys residues ........ 206
5.2 SDS-PAGE and motility analysis of the wild-type flagellin and two cysteine mutation variants .................. 211
6.1 Comparison of motile and non-motile bacterial growth patterns in 96-well plate-formatted motility agar .......... 222
6.2 Comparison of average absorbance readings for motile and non-motile S. typhimurium bacterial cells .............. 223
6.3 Growth curves for motile and non-motile S. typhimurium bacterial cells inoculated in 96-well plate-formatted motility agar .......... 226
7.1 Location of the three rationally designed metal binding sites in the flagellin (FliC) structure (PDB 1UCU) .................. 235
7.2 Pictorial depiction of three rationally designed metal binding sites in the FliC protein ................................. 239
CHAPTER I

INTRODUCTION TO THE DISSERTATION

The ability to provide movement in living organisms plays an important role in terms of fulfilling their need for nutrition, evading harmful conditions and colonization of new environments. Specialized membrane-bound appendages have evolved in prokaryotes and eukaryotes in the form of flagella and cilia. These motility structures must have evolved in conjunction with other sensor components involved in sensing the environment, such as chemotaxis, thermotaxis, magnetotaxis, geotaxis, etc.

The flagellum is the primary organelle responsible for motility in eubacteria. It is a complex structure composed of a basal body, a trans-membrane rotary motor, a universal joint-like hook structure and a hollow helical flagellar filament (Figure 1.1). The flagella of *E. coli* and *S. typhimurium* have been extensively studied and serve as models for the study of eubacterial flagella. The flagella filament has an outer diameter of 12–25 nm, a 2-3 nm diameter inner channel and can be 1-15 micrometers or more in length. Flagella are rotated by the membrane rotary motor to enable movement of the bacterium, in a process that is often regulated by chemotaxis. Peritrichously arranged flagella either associate into a helical bundle of multiple flagella that generates a net thrust along its axis or they separate into individual fibers, depending on the direction of rotation and resulting

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
supercoiled state of the fiber. Figure 1.1 shows the basic structure of typical eubacterial flagella.

**General structure of Flagella**

![Diagram of flagella structure](image)

**Figure 1.1 General structure of eubacterial flagella.** The MS (membrane/supramembrane), L (lipopolysaccharide) and P (peptidoglycan) rings anchor the structure into the cell wall. The MotA-MotB studs represent the site of the motor complexes which drive flagellar rotation. The C (cytoplasmic) ring complex is involved in controlling the direction of flagellar rotation. The C (cytoplasmic) rod may control export of flagellar proteins. The filament cap and hook–filament junction sites contain minor proteins not mentioned in the text. The FliC (flagellin) protein that forms the bulk of the filament is synthesized in the cytoplasm; the FliS protein binds to the C-terminus of the monomers to prevent degradation and premature
polymerization of the protein. The FliC protein is exported through the central channel to the tip where it folds into proper structure with the help of the FliD/Hap2 filament cap. (The above figure has been adopted and modified from an Encyclopedia of Life Sciences article titled “Bacterial Flagella”.)

Depending on their location, the flagellar complexes have been divided into MS (membrane/supramembrane), L (lipopolysaccharide) and P (peptidoglycan) that anchor the structure into the cell wall. The motor complexes are represented by MotA and MotB complexes and the cytoplasmic ring complex (C) is involved in controlling export of flagellar proteins and the direction of rotation (Figure 1.2). The C component comprises a peripheral ring and an axial rod. The FliG protein forms the cytoplasmic component of the switch complex that also includes FliM and FliN proteins. The MS ring forms a connection between the C component and the cell wall component and is predominantly made up of several copies of the FliF protein. Flil is the ATPase of the export apparatus that aids in the export process of other type III proteins. A recent study shows structural similarities Flil and F1-ATPase despite the apparently different functions of these proteins. It is central to the structure and function of the flagella. The MS ring forms an axial rod that emerges through the outer wall and is connected to the external hook. Due to the presence of the outer membrane in Gram-negative bacteria, the L and P ring structures serve as bushings while rotation occurs.

The rotary motor is composed made up of two components, a stator and a rotor. The MotA and MotB proteins constitute the stator and the FliG protein.
constitutes the rotor. The motor is driven by movement of either protons or sodium ions across the membrane. The rotary motor of flagella is the most powerful of the known macro molecular machines, generating torques of ~4000 pN nm near stall and rotates at ~300 Hz. The sodium-driven motor rotates at up to ~1700 Hz. For each motor revolution, ~1200 protons pass through it. The flagella hook formed by FlgE protein plays an important function in the functioning and assembly of the filament. The FlgD protein forms the hook cap structure that forms prior to the assembly of the hook. The FliK/Hap1 protein functions as a molecular ruler that controls the length of the hook. The FlgK and FlgL/Hap3 proteins form the hook-filament junction.

A striking feature of the different prokaryotic and eukaryotic flagella is that they don't appear to have a common evolutionary origin. Eukaryotic flagella differ remarkably from their prokaryotic counterparts. The basic unit of eukaryotic flagella and ciliary systems, the axoneme, is predominantly composed of microtubules, dynein proteins and up to 250 associated proteins. The basic scheme of flagella architecture is represented as 9+2 (2 central microtubules and 9 peripheral tubes connected by spokes). The dynein proteins hydrolyze ATP molecules to drive rotary motion of the axoneme.

At first inspection, eubacterial and archaeal flagella appear to be more similar in structure and function. However, they are actually not similar in the makeup of their genetic components and are not derived from the same type of secretion system; bacterial flagella are based on a type III export system, similar to that used by...
pathogens such as *Yersinia pestis* to deliver toxin proteins to their host cells. In contrast, archaeal flagella appear to have more sequence similarity with pili, a type IV secretion system \(^2\); \(^{26}\); \(^{27}\); \(^{28}\). The archaeal flagellar filament has significant structural differences when compared to the eubacterial flagella. The archaeal flagella has been found to be extensively glycosylated, which could be a stabilizing factor under the extreme conditions that many archaea live in \(^{29}\). Furthermore, the archaeal flagella filament is a nearly solid structure \(^4\); \(^{27}\); \(^{28}\); \(^{30}\) whereas the eubacterial filament is hollow with distinct a central channel. The archaeal filament is also more narrow and ranges from 10-14 nm in diameter, compared to 18-24 nm diameter for the eubacterial counterpart. The protein components of the flagella follow a different mode of export, similar to that of type IV pili of eubacteria. There is also growing evidence about the similarities between the ATPases and archaeal flagella motor proteins. The archaeal flagellin proteins have an N-terminal signal peptide similar to that of pili that is cleaved by a membrane-bound peptidase. Recently, pili have also been shown to be the source of propulsion in twitching motility of bacterial cells \(^{30}\); \(^{31}\); \(^{32}\).

This study primarily involves understanding the structural and functional aspects of the *S. typhimurium* flagellin protein that forms the flagella filament. The flagella filament in *S. typhimurium* is primarily composed of up to 30,000 copies of a single globular protein called flagellin (FliC), which self-assembles via noncovalent forces to form a helical fiber composed of 11 protofilaments \(^9\). The primary (phase-1) flagellin gene in *S. typhimurium* is designated as *fliC*, a secondary (phase-2) flagellin gene is named *fljB*, while other bacterial flagellin genes are named as *flaA*, *hag*, or...
flaF. Flagellin molecular masses can range from 20-77 kDa \cite{33, 34, 35, 36, 37}, corresponding to amino acid sequences ranging in size from under 300 amino acids to almost 700 amino acids (e.g., *Pseudomonas putida* has 688 residues), with the majority of flagellins containing 500 amino acids \cite{13}. The N- and C-termini of flagellin proteins are conserved across various species \cite{38}. These size differences are primarily due to high genetic variation in a "hypervariable" middle region of the protein encoding outer domains, D2 and D3 \cite{38}. Flagellins have long been characterized with respect to their serotype antigenicity in eliciting production of antibodies in mammals and are classified as H-antigens \cite{34, 39, 40, 41, 42}. The outer domains are the primary immunological determinant of this protein; the sequence of the middle region has long been correlated with variations in the H antigenicity \cite{42, 43}. The flagellin proteins are so variable that the *Salmonella sp.* has at least 70 different serological strains \cite{44}.

The export of flagellin monomers is facilitated by the cytosolic chaperone FliS, which binds to FliC and inhibits its polymerization \cite{45, 46}. FliS binds specifically to the C-terminal 40 amino acid component of the disordered D0 domain in a 1:2 ratio, which prevents self-assembly and formation of flagellin fiber aggregates within the cell \cite{45, 46, 47}. Without FliS binding, the C-terminus is degraded. However, it is not clear what the conformational states of the N-terminus and middle hypervariable regions of the FliC are prior to assembly and during export through the flagella central pore. No study has been performed to understand the interactions of these domains before they are exported through the flagellar central channel. A new report published by Muskotal et al. \cite{48} suggests that the FliS protein binds to the C-terminal
region of the FliC in the ratio of 1:1. The report suggests that FliS does not act as an unfolding factor. Instead it helps in the formation of alpha-helical secondary structure in the region of FliC, where it binds.  

Eubacterial flagellins are known to undergo several post-translational modifications including glycosylation and N-methylation of lysines, yet no functional significance has been attributed. The flagellin export in bacteria is very much similar to the type III secretion system (TTSS) of several pathogenic Gram-negative bacteria including Yersinia enteritidis. The TTSS also employs a chaperone (e.g. SycE) that prevents the folding of the pathogenic proteins (Yops) before secretion. The secretion apparatus, i.e., the “injectisome”, includes ~25 proteins and forms a needle like projection that penetrates the host cytoplasmic membrane to release the pathogenic factors. Several studies have been performed to determine the extent of the folding of the secreted proteins, as the needle diameter is only 2-3 nm. A recent study involving a fusion protein of YopE-DHFR suggests that the proteins can be folded before export, but they are exported with the help of an unfoldase activity. This terminal assembly process is aided by the flagellar chaperone cap protein, termed FliD or HAP2.

The partial and complete structures of S. typhimurium FliC were determined in 2001 and 2003. The complete FliC structure (PDB 1UCU) shows four distinct globular domains in the protein, termed D0, D1, D2 (with subdomains D2a and D2b) and D3. Domain D0 forms the inner core of the filament; D1 forms the outer core and domains D2 and D3 form a knob-like projection on the filament surface. Extensive
biophysical studies performed *in vitro* on *S. typhimurium* flagellin monomers, using proton nuclear magnetic resonance (NMR), Fluorescence Resonance Energy Transfer (FRET) and Circular Dichroism (CD) indicate that monomeric flagellin contains
**D3 domain**
Structure: mainly β-sheet.
Outer part of dispensable, solvent accessible “hypervariable” region, extracellular display region.
Function: unknown; propulsion, host immune evasion.

**D2 domain**
Structure: Mainly β-sheet.
Inner part of dispensable, solvent accessible “hypervariable” region, extracellular display region.
Function: folding, linker, conformational flexibility.

**D0 domain.**
Structure: α-helical.
Function: type III export signal, FlIS chaperonin binding & flagellar self-assembly via intermolecular coiled-coil interactions.

**C-terminus**

**N-terminus**

---

Figure 1.2 A cartoon representation of *S. typhimurium* flagellin protein (PDB: 1UCU). The D0 domain is formed by the N-and C-terminal regions and is composed of α-helices. The D1 domain is consists of mostly α-helices and a β-hairpin. The D2 domain consists mostly of β-strands and falls in the hypervariable region of the protein. The D3 domain forms the solvent accessible hypervariable outer region and consists of a unique β-folium structure.
regions both of low mobility (broad resonances) and of high mobility (sharp peaks)\textsuperscript{64};\textsuperscript{65,66}. These high mobility regions are generally considered to be highly disordered and can be assigned to specific residues in the N- and C-terminal regions of the sequence. CD studies show that flagellin contains a mixture of $\alpha$-helix, $\beta$-structure and random coil and appear to form two discrete folding domains as measured using microcalorimetry. Upon polymerization, the amount of $\alpha$-helix increases, while the amount of random coil decreases. Sequence analysis work based on the \textit{E. coli} and \textit{Salmonella} species indicated that the N- and C-terminal regions of flagellins are more highly conserved than the interior sequences. Among the eubacterial flagellins, $\sim$180 N-terminal and $\sim$100 C-terminal residues are highly conserved and form the D0 and D1 domains\textsuperscript{67}. The region between two conserved termini is highly variable, both in length and sequence. Our study of internal deletions of FliC protein of \textit{S. typhimurium} indicates that the hypervariable region can be deleted to a large extent without interfering with the export and assembly of the flagellin\textsuperscript{68}. A similar study that was performed earlier on \textit{E. coli} indicated that the minimal sequence capable of forming filaments contained little more than these conserved terminal regions\textsuperscript{69}. The 272 residue flagellin of \textit{Bacillus} is slightly smaller than this ‘minimum-length flagellin’, and may represent the minimum residues required for filament formation and function.

 Structural features of the D2 and D3 domain regions of \textit{S. typhimurium} phase 1 flagellin PDB structure (1UCU), as defined by Samatey, \textit{et al.},\textsuperscript{70} are summarized here. The central, hypervariable region of \textit{S. typhimurium} flagellin that encodes the
D2 and D3 domains is composed of residues Lys177–Ala401. Subdomain D2a consists of the N-terminal residues Lys177–Gly189 and the C-terminal residues Ala284–Ile344. Subdomain D2b consists of residues Asn345–Ala401. Domain D3 is formed by residues Tyr190–Val283. The structure has nine helical structural motifs, with an α-helical region and short adjoining 3_10 helical region, collectively termed helix 6 (Leu288–Ala298), present in subdomain D2a. A second, largely helical region that is neither in 3_10 nor α-helical form is present in domain D3 (residues Asn199–Gly210) and was named helix 4-5 in accordance with the PDB structure notation. The wild-type structure also has 20 defined β-strands, of which 18 are present in the outer D2 and D3 domains. The Ca backbone of the D2 and D3 domains contains three examples of a unique folding motif, termed the β-folium, comprising a series of four successive β-strands that form three β-hairpins all pointing away from one another. The three β-folium structural motif regions are composed of residues 220–260 in domain D3, residues 308–345 in subdomain D2a and residues 345–383 in subdomain D2b. Due to the large structure of the flagellin and solvent exposed dispensable hypervariable region, flagellin serves as an excellent scaffold for protein and peptide display.

Flagella display, similar to phage display, has evolved as an alternative method of peptide display, in part due to the ease of purification and tolerance for insertion of foreign peptides. Flagella display is based on the genetic fusion of foreign loop peptides into the surface-exposed non-essential “hypervariable” central
region of flagellin, which is the major subunit present in thousands of copies per flagella filament. Previous reports have demonstrated the utility of genetically inserted fusion peptides and small proteins displayed on the immunologically reactive, solvent accessible, middle domain of mesophilic E. coli flagellin.

Expression of these constructs in flagellin-deficient host strains results in hybrid flagella carrying the heterologous peptides in thousands of copies displayed in a regular array on the flagella surface. This approach has been successfully used for expression of foreign peptides/proteins to be displayed on flagella. A versatile variant of flagellar display is the FliTrx hybrid display system created by the insertion of the entire E. coli thioredoxin gene into the central region of flagellin. Peptide libraries have been genetically introduced into a disulphide loop of thioredoxin, creating a conformationally constrained library that is readily accessible on the flagellar surface. This FliTrx random peptide display library system has been used with great success for epitope-mapping purposes. Direct flagella display has also proven to be applicable also in bacterial adhesion technology, as large fragments, up to 302 amino acid residues in length, of bacterial adhesins can be functionally expressed as fusions to flagellin. Several studies indicate that tolerance to internal domain fusions may be a general property of many proteins and represents another means by which protein structures may evolve. A number of foreign peptides have been successfully inserted into the hypervariable region of flagellin, including simultaneous insertion of 115 and 302 amino acid peptides. These engineered hybrid peptide-flagellin proteins are most frequently used as...
immunological reagents to generate vaccine antibodies against the inserted peptides when injected into animals. Only two known reports describe the insertion of full-length proteins into flagellin, the FliTrx system first described in 1995 and a second 1998 publication. Hybrid flagella are easily purified and can easily be analyzed for binding to various targets, such as immobilized proteins, tissue sections, as well as cell cultures. Flagella have also been used as a scaffold for composite materials insertion and display of other fusion peptides and proteins on flagella as a "bionanotube", as demonstrated in other recent publications by Kumara, et al.

One specific aim of the research described in the following chapters was to understand the functional and structural role of the hypervariable domains of flagellin and to what extent they can be removed. We have also modeled and studied the structure of a thermostable flagellin protein from *Aquifex pyrophilus*, with the goal of engineering this thermostable protein for use as a thermostable flagella nanotube in peptide display and nanotechnology applications. Motility being central to the virulence of many pathogens, we have also designed and tested a high-throughput screening assay to find bacterial motility inhibitors.

References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


CHAPTER II

INVESTIGATION OF THE HYPERVARIABLE DOMAIN REGION OF THE SALMONELLA TYPHIMURIUM FLAGELLIN PROTEIN IN MOTILITY

Abstract

The eubacterial flagellum is a complex structure with an elongated extracellular filament that is primarily composed of a single protein termed flagellin. The highly conserved N- and C- termini of flagellin are important in its export and assembly, whereas the middle “hypervariable” region is highly variable in size across different species and can be deleted to a large extent. In Salmonella typhimurium phase 1 flagellin, this hypervariable region encodes two solvent-exposed D2 and D3 domains that form a knob-like feature on flagella fibers. The functional role of this structural feature in motility remains unclear. We investigated the structural and physiological role of the hypervariable region in flagellar assembly, stability and cellular motility. A library of random internal deletion variants of S. typhimurium phase 1 flagellin was constructed and screened for functional variants using a swarming agar motility assay. The relative cellular motility and propulsive force of ten representative functional variants were determined in semi-solid and liquid media using swarming motility assays, video microscopy and optical trapping of single cells. All ten variants investigated exhibited diminished motility, with varying extents of partial motility observed for internal deletions less than 75 residues and nearly complete loss of ...
motility for deletions greater than 100 residues. The mechanical stability of flagella fibers also decreased with increasing size of deletions. Analysis of deletion variant sequences with respect to the wild-type structure indicated that all deletions involved both loss of hydrophobic core residues and loss of both partial and complete segments of β-strand secondary structure in D2 and D3 domains. Homology modeling also predicted disruptions of secondary structures in each variant. The hypervariable region middle domain appears to have an important role in stabilizing the folded conformation of the flagellin protein and resulting mechanical stability of the flagella fibers and the resulting propulsion.

Introduction

The eubacterial flagellum is a helical protein filament that is rotated by a membrane-bound motor to propel the cell through liquid environments. Other flagellar biological functions include host adhesion, colonization and virulence. The flagellum is a complex structure composed of more than 20 component proteins and has several major features: a basal body, a trans-membrane motor, a hook structure and an elongated helical filament. The filament has an outer diameter of 12–25 nm, a 2–3 nm diameter inner channel and can be 1–15 μM or more in length. It is composed of as many as 30,000 subunits of a single flagellin protein.

Salmonella enterica serovar Typhimurium, commonly named Salmonella typhimurium, is a peritrichously flagellated (6-8) bacterial pathogen that serves as a model organism for the investigation of bacterial flagellar structure and function.
(Figure 2.1). The mature *S. typhimurium* phase 1 flagellin protein, encoded by the *fliC* gene, consists of 494 amino acids; the first methionine residue is removed post-

---

**Figure 2.1 An Electron Micrograph of *Salmonella typhimurium*.** An electron micrograph of a wild-type *Salmonella typhimurium* strain SJW1103, negatively stained with phospho-tungstic acid. The figure shows six peritrichously arranged flagella.

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
translationally. The structure of *S. typhimurium* flagellin was recently determined (PDB 1UCU)\textsuperscript{12,15} and shows four distinct globular domains in the protein, termed D0, D1, D2 (with subdomains D2a and D2b) and D3. Domain D0 forms the inner core of the filament; D1 forms the outer core and domains D2 and D3 form a knob-like projection on the filament surface. The N-terminal polypeptide chain starts at domain D0, proceeds through domains D1 and D2 and reaches domain D3, and then returns through domains D2 and D1, with the C-terminal chain ending in domain D0.\textsuperscript{15}

The D0 and D1 domain regions are formed from the N- and C-terminal regions and are highly conserved across different bacterial species.\textsuperscript{16,17,18,19} These terminal domain regions are essential for flagellar self-assembly and are largely composed of α-helical coiled-coils that further associate to form helical bundles in the interior of the filament. The middle region of the flagellin sequence encodes the surface-exposed D2 and D3 domains that are largely composed of β-sheet structures. This “hypervariable” sequence region can vary greatly in sequence composition and number of residues across different strains and species, resulting in observed flagellin molecular masses that range from 20-77 kDa.\textsuperscript{20,21,22,23}

\textsuperscript{24} The D2 and D3 middle domains have been attributed with functions such as aiding in increase of traction while swimming and evasion of host defenses due to its sequence variability.\textsuperscript{23,26,27} Flagellins have long been characterized with respect to their serotype antigenicity in eliciting production of antibodies in mammals and are classified as H-antigens.\textsuperscript{21,28,29,30,31} The outer D2 and D3 domains are the primary

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
immunological determinant of this protein; the sequence of the middle region has long been correlated with variations in the H antigenicity.\textsuperscript{31, 32} Large portions of the D2 and D3 hypervariable domain regions can be deleted to a large extent without completely disabling the export and self-assembly functions of flagellin.\textsuperscript{18, 32, 33, 34} Furthermore, the deletion tolerance of the D2 and D3 domains has been exploited by a number of researchers to insert peptides and proteins in this extracellular protein for use as a peptide display tool for selection of constrained peptides with affinity for other ligands and immunogenic epitope display.\textsuperscript{35, 36, 37, 38, 39}

The biophysical function of the middle domain of flagellin in propulsion remains unclear. Further understanding of the evolution and function of the hypervariable region of the flagellin protein and its resulting physiological impact on cellular motility would be useful, in light of the recently determined structure. Structural information about the protein folding rules that allow deletion of segments of the middle region may also be useful in developing rational approaches for insertion and display of other fusion peptides and proteins on flagella as a "bionanotube", as demonstrated in other recent publications by Kumara, et al.\textsuperscript{40, 41, 42, 43}

We investigated the function of the hypervariable region of flagellin on cellular motility by deleting portions of the outermost D2 and D3 domains and screening for variants with functional motility. Three complementary biophysical methods were used to characterize the motility of cells expressing the wild-type flagellin and ten representative internal deletion variants, in environmental conditions
that ranged from viscous agar to dilute aqueous solution. The structures of these variants and two previously described internal deletion variants were analyzed with respect to the wild-type protein structure and predictive homology models were developed for most of these variants.

Materials and Methods

Bacterial strains

*E. coli* XL-1 Blue electrocompetent cells (Stratagene, La Jolla, CA) were used for the preparation of plasmid libraries of *fliC* internal deletion variants. These cells have a transformation efficiency greater than $1 \times 10^{10}$ transformants/µg of DNA, are tetracycline resistant, endonuclease (*endA*) deficient, which greatly improves the quality of miniprep DNA and are recombination (*recA*) deficient, improving insert stability. The *S. typhimurium* strains used for the experiments were kindly provided by the late Dr. Robert Macnab (Yale University). *S. typhimurium* strain SJW1103,44; 45; 46 a derivative of serovar Typhimurium LT2 that 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 (*ΔfliC* and *ΔfljB*), 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.47; 48 This strain is non-motile, unless a functional flagellin gene is introduced, e.g., on an expression plasmid.47 Thus, it serves as an ideal strain to screen engineered or mutated *fliC* genes, present in the introduced plasmids, for
functional flagellin protein expression, folding, export and flagellar fiber assembly, as indicated by a simple agar microbial motility assay. *S. typhimurium* strain JR501, a stable restriction-deficient (*r*), methylation-proficient (*m*⁺) galE strain,⁴⁹,⁵⁰ was used to convert *E. coli* grown plasmids to *S. typhimurium* compatibility. Electrocompetent cells of *S. typhimurium* strains JR501 and SJW134 were prepared using standard lab protocols.⁵¹,⁵²

**Plasmids and construction of flagellin internal deletion variant library**

The pTrc series of plasmids were constructed using plasmid pKK233-2, for the regulated expression of genes in *Escherichia coli*.⁵³,⁵⁴ These vectors carry a strong hybrid trp/lac, isopropyl-β-D-thiogalactopyranoside (IPTG) inducible promoter, the lacZ ribosome-binding site (RBS), the multiple cloning site of pUC18 and the rnb transcription terminators. The pTrc plasmids have been frequently used for cloning purposes. The pTH890 plasmid is a derivative of pTrc99A plasmid with the *S. typhimurium fliC* phase 1 flagellin gene cloned into the XbaI/HindIII site (Figure 2.2). The pTH890 plasmid was kindly provided by the late Dr. Robert Macnab (Yale University). This plasmid has a unique BsrGI restriction site at the nucleotide position 843, located approximately in the middle of the *fliC* gene. We exploited this fortuitous unique site to generate the internal deletion library. Unless otherwise noted, all restriction enzymes and PCR reagents were obtained from New England Biolabs (Beverly, MA). First, the pTH890 plasmid was cut at the unique BsrGI site with the corresponding restriction enzyme to generate a linear plasmid with intact 5' and 3'
Figure 2.2 A pictorial representation of pTH890 plasmid. The pTH890 plasmid is a derivative of pTrc99A plasmid with the *S. typhimurium* *fliC* phase 1 flagellin gene cloned into the XbaI/HindIII site. This plasmid has a unique BsrGI restriction site at the nucleotide position 843, located approximately in the middle of the *fliC* gene. A *Pvu*I restriction site is present in the middle of *β-lactamase* gene.
ends of the \textit{fliC} gene. This was followed by digestion with Bal31 slow exonuclease time course reactions for 30 s, 60 s, 90 s, 120 s, 150 s, 180 s, 210 s, 240 s, 300 s and 20 min. The varied time periods were used to generate diversity in the length of deletions in the \textit{fliC} gene present at both ends of the linear plasmid. After each time interval equal volumes of the reaction mixture were transferred into the phenol:chloroform:isoamyl alcohol (25:24:1) mixture to denature Bal31 nuclease and stop the reaction. The DNA was precipitated using 5 µl of 3M sodium acetate and 150 µl of 100% ethanol (Figure 2.3).

The Bal31 nuclease creates rough, single-stranded ends with overhangs. These rough ends were repaired by DNA polymerase I Large Fragment (Klenow Fragment), in the presence of deoxy nucleotide tri-phosphates (dNTPs). This enzyme either digests or fills in any overhangs on the ends of the linearized plasmid DNA fragment. To generate additional diversity in the \textit{fliC} internal deletion library, \textit{i.e.}, allow short 5' fragments of \textit{fliC} to recombine with long 3' \textit{fliC} gene fragments and vice-versa, a unique \text{PvuI} restriction site at nucleotide position 419 in the plasmid-encoded β-lactamase gene (\textit{amp'}) was cut following variable digestion with the slow Bal31 nuclease and the resulting used to shuffle the plasmid fragments. The resulting plasmid fragments were ligated at 16 °C overnight with T4 DNA ligase, thus producing a circular plasmid library of \textit{fliC} genes with internal deletions of variable length. The ligated plasmid library was then digested a second time with BsrGI restriction endonuclease before transformation into SJW134 cells. This procedure was designed to linearize plasmids containing the unique restriction site present only in
the wild-type fliC gene, thus preventing their transformation into the intermediate E. coli cloning strain. A proportion of the library was transformed into E.coli XL-1 Blue electrocompetent cells and all the transformants were inoculated into 50 ml of LB media (Shelton Scientific, Peosta, IA) with 50 µl of 100 µg/ml ampicillin (Shelton Scientific) for overnight growth. The direct transformation of the midiprep plasmid library obtained from E. coli XL-1 Blue cells into the non-motile S. typhimurium strain SJW134 was highly inefficient. The intermediate step of transformation into S. typhimurium strain JR501, plasmid recovery and subsequent transformation into SJW134 strain yielded good results. After the SJW134 cells were electroporated with the plasmid library, the plates were incubated overnight at 37°C. The colonies were picked using a blunt ended toothpick and smeared onto the surface of the motility agar and incubated at 30 °C in a humidifying incubation chamber for 6-8 hrs. The plasmid library from the cell culture was isolated using a QIAprep Plasmid Midi Kit (Qiagen Inc., Valencia, CA). The library was transformed into S. typhimurium strain JR501 for methylation of the plasmids and the transformants were inoculated into 50 ml of LB media with ampicillin for overnight growth. The plasmid library was isolated using a QIAGen Plasmid Midi Kit.
Construction of *fliC* gene deletion library

**Figure 2.3 Construction of fliC gene deletion library.** a) The pH890 plasmid was cut with BsrGI restriction enzyme. b) The linear plasmid was digested using BsaI nuclease for varying time periods and the ends were repaired with Klenow fragment. c) The digested fragments were further cut with PvuII restriction endonuclease. d) The fragments were ligated using T4 DNA ligase and transformed into XL-1 blue *E. coli* cells.
The identified motile colonies were streaked on fresh LB ampicillin plates and incubated overnight at 37 °C to obtain single isolated colonies. Colony PCR was performed with primers that flanked the full length gene, to identify the extent of deletion within each *fliC* gene that conferred motility to the cells. The PCR results were analyzed on 2% agarose gels. A control PCR reaction on full-length *fliC* formed a band just above the 1.5 kb DNA marker. Colony PCR performed on the variably motile cells indicated *fliC* genes with molecular masses equal to or smaller than the full length control. After screening for motile cultures using a motility assay, colony PCR was performed to identify the approximate extent of deletion in each *fliC* gene. The results were initially analyzed using agarose gel electrophoresis, followed by DNA sequencing of the isolated plasmids. The sequencing results were verified using nucleotide BLAST and the extent of deletion was analyzed. The plasmids were named according to the sequence in which they were identified with the clones and hence have a suffix C with a number.

**Motility assays**

*S. typhimurium* strain SJW134 electrocompetent cells were transformed with the pTH890 *fliC* internal deletion variant plasmid library or purified pTH890 plasmids of interest and plated on LB agar (1.5%, wt/vol) for overnight growth at 37 °C. Motility agar media was composed of tryptone broth (10 g/liter tryptone, EM Science, 5 g/liter NaCl, Fisher Scientific) that contained 0.3% (wt/vol) agar (Sigma-Aldrich, St. Louis, MO). For the initial screening, agar plates were inoculated with bacteria from an overnight culture in LB agar (1.5%, wt/vol) plates at 37 °C with a sterile, blunt-ended
toothpick. The plates were then incubated at 30 °C for 6-8 h in a humidified incubator to prevent dehydration of the media. The low density of the agar allowed the bacteria to move within the agar, forming a visible halo of growth around the point of inoculation. For further analysis of swarming diameters of the 10 deletion mutants, the motility agar was inoculated with 2 μl of overnight culture at an OD550 of 0.5. Optical imaging measurements were taken after six h with a Kodak DC290 digital camera gel documentation system. A controlled motility assay was performed using the deletion mutants and the negative and positive control strains SJW134 (non-motile) and SJW1103 (wild-type). Five different sets of measurements were performed on each clone and the average swarming diameter was calculated.

**Colony PCR**

Colony PCR was performed on motile *S. typhimurium* colonies using a Techne Touchgene Gradient Thermal Cycler (Techne, Inc., Burlington, NJ). PCR reagents were obtained from NEB. Bacterial colonies were picked using 10 μl micropipette tips and added to 50 μl of PCR master mix (0.25 μl *Taq* DNA Polymerase (NEB), 5 μl 10x standard PCR buffer, 5 μl of 2.5 mM dNTP mix, 5 μl of 10 μM forward and reverse primers and 30 μl of DI H2O). The PCR reaction conditions used were: an initial cycle of 94°C for 10 min 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. A forward oligonucleotide primer with DNA sequence 5’-AATTAATCCGGCTCGT-3’ started at nucleotide position 197 of the pTH890 plasmid. A reverse oligonucleotide primer with DNA sequence 5’-
ATTTAGTCTTGTGGTCCT-TG - 3' started at nucleotide position 1958. These two primers were designed to completely flank the fliC gene in the pTH890 plasmid. The PCR products were analyzed by electrophoresis on 2% agarose gels.

**DNA sequence determination and analysis**

Plasmids containing functional fliC genes with internal deletions were purified using a QIAprep plasmid miniprep kit (Qiagen, Valencia, CA). A forward sequencing primer (5'-TCCATCCAGGCCTGAAATC-3') was designed starting at fliC nucleotide position 330 to analyze the extent of deletion. DNA sequences were analyzed with the BLASTN software\(^5\) using the default parameters, to identify the extent of deletion within each gene. The ORF (Open Reading Frame) finder tool at the NCBI server was used to convert the nucleotide sequences into protein sequences. Multiple sequence alignment of the mutant flagellin proteins was performed using the ClustalW software package.\(^6\)

**Flagellin protein expression, flagella fiber purification and SDS-PAGE**

Purified pTH890 plasmids encoding fliC internal deletions of interest were transformed into competent SJW134 *S. typhimurium* cells and plated on LB-Amp agar. Colonies were picked and inoculated into 10 ml LB broth with 5 μl of 100 mg/ml ampicillin and allowed to grow overnight at 37 °C. 10 ml of the bacterial cultures were transferred to a 15 ml polypropylene tube and placed on a heat block at 60 °C for 10 min. The cells were pelleted at 6000 x g and the supernatant was transferred to 15 ml Amicon Ultra centrifugal filtration tubes (Millipore Corporation, Billerica, MA) for concentrating the flagella fibers. The concentrates were resolved.
by SDS-PAGE on 10% (wt./vol.) polyacrylamide gels with power settings of 30 mA and 180 V and visualized by staining with Coomassie brilliant blue R-250.

Isolation of flagella fibers from bacteria for TEM studies was performed by an established technique\(^1\)^\(^7\),\(^3\)^\(^2\),\(^5\)^\(^7\) of vortexing 1.0 ml of cells for 30 sec. in a 1.5 ml microcentrifuge tube, followed by selective pelleting of the cells by centrifugation at 5,000 x g for 5 min. 0.5 ml of supernatant containing suspended flagella was transferred to an Amicon Centricon centrifugal filter (Millipore, Billerica, MA) with a molecular weight cutoff of 30 kDa and centrifuged at 15,000 x g for 30 min. The concentrated solution was collected into the retentate vial by brief centrifugation.

Partial purification of flagella by this method yielded a significant amount of protein even with a small amount of culture.

**Electron microscopy**

Flagellar fiber samples were prepared for TEM imaging by shearing and concentration in the manner described above. For TEM imaging of bacterial cells, 20 \(\mu l\) of overnight grown culture samples were applied to carbon coated copper grids and negatively stained with 4% aqueous phosphotungstic acid (pH 5.2). Both flagella attached to the bacterial cells as well as detached flagella were observed. TEM micrographs were taken using a JEM-1230EXII electron microscope (JEOL, Tokyo, Japan), at an accelerating voltage of 80 kV. Digital images were captured as TIFF files by a 1 megapixel Gatan Bioscan digital camera (Gatan Inc., Pleasanton, CA).
Light microscopy and swimming velocity analysis

A Nikon Eclipse E600 epifluorescence stereo microscope (Mager Scientific Inc., Dexter, MI) with 10X, 40X and 100X (oil immersion) Planfluor objectives, transmission and darkfield condenser light sources and a QImaging Fast 1394 cooled mono (greyscale) CCD digital camera was used to record visible images and movies of non-motile *S. typhimurium* SJW134 bacteria transformed with original or modified pTH890 expression plasmids encoding wild-type and internal deletion variant flagellins. The video images of swimming bacteria were captured at 13 fps speed using a 100X oil immersion lens. Bacterial cultures were grown to mid-log phase in LB media without shaking to minimize breakage of potential brittle flagella fibers. 30 µl of each culture at OD595 0.05 was loaded onto a concave glass slide and covered with a glass cover slip. The Manual Tracking plug-in of the ImageJ software (NIH, Bethesda, MD) (Version 1.36, http://rsb.info.nih.gov/ij/) was used to calculate the velocity with which the bacteria were moving. The swimming speeds of the cells were measured for 20 frames. The swimming speeds of 20 individual bacterial cells were determined and averaged for wild-type and each deletion variant. In general, the more the cells tumbled and changed direction, the slower was the observed cell speed. In order to measure the approximate swimming speed, cells were chosen for analysis that stopped and changed direction the least in two dimensions.

**Measurement of power required to stall cells using optical tweezer**

A BioRyx 200 optical trap using a 1064 nm (Arryx, Chicago, IL) high power 2 Watt laser source and proprietary computer-controlled holographic optical trapping
software was used to trap and monitor the propulsive forces generated by bacterial cells (Figure 2.4). The software can create up to 200 independently maneuverable optical traps in three dimensions and can trap objects from 150 nm to 300 μm in size. A Nikon Eclipse TE2000U inverted epifluorescence microscope with 4X, 10X and 60X (oil immersion) objectives, fitted with a digital 640 x 480 pixel resolution CCD digital camera was used to record visible images and movies of trapped *S. typhimurium* bacteria with different flagellin mutant plasmids. Cell cultures were grown to the mid-log phase, where they produce maximum number of flagella. The cells were grown without shaking in order to reduce potential breakage of flagella. The cultures were then diluted to an OD$_{550}$ 0.05. Cells observed in the LB growth media were difficult to follow as their movement was very rapid, especially the wild-type, and also due to the limitations of the laser power. Addition of 6.25% glycerol to the media considerably reduced the speed of the swimming cells and enabled their trapping with relative ease. Individual cells were initially trapped using the maximum laser power of 2.0 W.

The laser power was then decreased from 2.0–0.2 W by 0.1 W increments and the power at which cells started to move out of the trap was noted. The minimum power of the laser was 0.2 W. The average stall force for each variant was calculated from measurements performed on 20 different single cells, with similar dimensions, from the same culture preparation. The stall force each cell was measured multiple times for consistency.
Figure 2.4 A schematic representation of a laser tweezer. a) The trapping laser is an Nd-YAG continuous-wave operated at 1064 nm. The laser power can be varied from 0.2 W to 2 W. The tweezer is optimized to work with the 60X objective. b) The optical tweezers system has been constructed around an inverted microscope (Nikon TE2000U) with a high numerical aperture of 1.4 and a working distance of 0.21 mm (oil immersion objective).
Structure analysis and homology modeling

The deleted regions of the flagellin protein were visualized and analyzed with respect to end-to-end distances using the SwissPDBViewer software version 3.7. The structures of different flagellin protein sequences were modeled based on the available 3D structure of *S. typhimurium* flagellin 1UCU at the RCSB Protein Data Bank (PDB). The MODELLER (6v1) software module of the Accelrys (San Diego, CA) InsightII software package was used to generate and the models and analyze them. The Accelrys modeling software was run on an IBM IntelliStation M Pro PC workstation with the Red Hat Enterprise Linux WS 3 operating system.

Results

Description of flagellin hypervariable domain structural features

Structural features of the D2 and D3 domain regions of *S. typhimurium* phase 1 flagellin PDB structure (1UCU), as defined by Samatey, et al., are summarized here for reference in the following analysis. The central, hypervariable region of *S. typhimurium* flagellin that encodes the D2 and D3 domains is composed of residues Lys177–Ala401. Subdomain D2a consists of the N-terminal residues Lys177–Gly189 and the C-terminal residues Ala284–Ile344. Subdomain D2b consists of residues Asn345–Ala401. Domain D3 is formed by residues Tyr190–Val283. The structure has nine helical structural motifs, with an α-helical region and short adjoining $3_10$ helical region, collectively termed helix 6 (Leu288–Ala298), present in subdomain D2a. A second, largely helical region that is neither in $3_10$ nor α-helical form is
present in domain D3 (residues Asn199–Gly210) and was named helix 4-5 in accordance with the PDB structure notation. The wild-type structure also has 20 defined β-strands, of which 18 are present in the outer D2 and D3 domains. These were numbered as: β-strand 3, Lys 179–Ala184; β-strand 4, Ala191–Thr193; β-strand 5, Ile195–Leu197; β-strand 6, Asp213–Asp217; β-strand 7, Leu220–Asp223; β-strand 8, Lys228–Thr236; β-strand 9, Gly243–Asp250; β-strand 10, Glu255–Ala259; β-strand 11, Glu276–Val283; β-strand 12, Ser305–Asp313; β-strand 13, Lys317–Val327; β-strand 14, Asp330–Asn337; β-strand 15, Ser341–Asn344; β-strand 16, Tyr349–Ala351; β-strand 17, Thr355–Lys357; β-strand 18, Asn361–Gly364; β-strand 19, Thr370–Ile375; β-strand 20, Lys378–Ala382. Subdomain D2a includes the five β-strands 3 and 12–15. Subdomain D2b includes the five β-strands 16–20. Domain D3 includes the eight β-strands 4–11.

The Cα backbone of the D2 and D3 domains contains three examples of a unique folding motif, termed the β-folium, comprising a series of four successive β-strands that form three β-hairpins all pointing away from one another.12 The three β-folium structural motif regions are composed of residues 220–260 in domain D3, residues 308–345 in subdomain D2a and residues 345–383 in subdomain D2b.

**Identification of functional fliC internal deletion variants**

An internal deletion library of the *S. typhimurium* phase 1 flagellin gene, *fliC*, was constructed by a procedure involving restriction digestion and variable nuclease digestion of the *fliC* middle region in a pTrc expression plasmid (Table 2.1, Materials
and Methods). Agarose gel analysis of the purified plasmid library yielded a broad smear, indicating the presence of a diverse range of deletions in the \(fliC\) gene (Figure 2.5). This plasmid library was then transformed into non-motile SJW134 \(S.\) _typhimurium_ cells and screened for functional internal deletion variants on 0.3% swarming motility agar. The colonies were initially classified into non-motile, low, medium and strongly motile (Figure 2.6). The initial colony motility screens yielded a ratio of roughly 1:10 motile to non-motile cells. PCR and DNA sequencing analysis of a preliminary set of selected colonies showed that most of the strongly motile cells had an intact \(fliC\) gene without any internal deletions. Therefore, the plasmid library was again digested with BsrGI restriction endonuclease to remove any remaining uncut, wild-type plasmids before transformation into SJW134 cells. This step considerably reduced the number of false positive results and the ratio of motile vs. non-motile cells was about 1:30. Approximately 12,000 single colonies were manually screened by this method and approximately 200 functional deletion variants were identified.
Figure 2.5 DNA agarose gel of the pTH890 deletion library. DNA agarose gel (1%) depicting the pTH890 plasmid and its deletion library. The plasmid library was constructed as described in the materials and methods section.

The relatively small number of functional flagellin variants isolated in the screen, ca. 2%, suggests that internal deletion and recombination events that result in a functional flagellin protein are relatively rare. Colony PCR indicated that all isolated fliC genes had molecular masses equal to or smaller than the full length fliC control (Figure 2.7). A total of 46 functional deletion variants with more substantial deletions were further characterized by DNA sequencing, yielding 39 unique clones (Table 2.2).
Figure 2.6 Motility assay. A picture depicting variation in the swarming abilities of different deletion variants on a 0.3% motility agar media. The deletion variants were initially classified as strong swimmers, medium-strong swimmers, weak swimmers and non-motile.

Analysis of functional deletion variant polypeptide sequences

All deletions in functional variants were centered on the location of the initial BsrGI restriction site in the \textit{fliC} gene, corresponding to a cut in the wild-type peptide sequence after residue 281 in the subdomain D2a α-helix 6. These results are a consequence of the molecular biology procedures and this may not be the most permissible deletion region in the D2 and D3 domains. Conversely, two regions were conserved in all variants in this study: β–strand 3 segment (Lys179–Ala184) connecting domain D1 to D3 and the two successive β–strands 4 and 5 on domain D3; residues Ala191–Thr193 and Ile195–Leu197. This sequence region is furthest 42
from the BsrGI restriction site, and may represent another bias in the molecular biology procedures, rather than an intrinsic stability of this region of the protein.

Figure 2.7 Colony PCR of the deletion variants. A 2% agarose gel of colony PCR products of several deletion variants. Lane 1: 100 bp Marker; Lane 2: pTH890 fliC gene PCR; Lanes 3-24: Colony PCR of different deletion fliC gene deletion variants.

The extent of functional deletions in the flagellin fliC gene ranged from 12–177 amino acids, although the majority of functional deletions were in the range of 40–80 amino acids, with an average value of $43.0 \pm 36.3$ deleted residues. The average N-terminal start position of the deleted polypeptide region was $260.2 \pm 18.4$ residues and the average C-terminal end position was $304.3 \pm 23.3$ residues. This indicates a preference in the position of functional deletions that is centered towards the C-terminal region of the hypervariable region, especially for larger deletions. This result is consistent with prior analysis of flagellin sequence homology. For example, the Pfam Protein Families Database entries for the bacterial flagellin N-terminus

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
and C-terminus, pfam00669 and pfam00700, indicate a more extended region of amino acid sequence conservation in the N-terminal region encompassing residues 28–163 (136 residues) than for the conserved C-terminal sequence encompassing residues 401–472 (72 residues).

It was theoretically possible for deletions to occur in all four domains of flagellin. However, no functional variants were isolated that contained any deletions of the largely α-helical D0 and D1 domains essential for self-assembly. This result was consistent with numerous prior observations indicating that the highly conserved sequences encoding the D0 and D1 domains are responsible for the majority of inter-subunit interactions in the filament, vs. the dispensable D2 and D3 domains. It was theorized that smaller functional internal deletions were more likely to be found in the outermost D3 domain and the D2a subdomain region; more extensive deletions extended further into the D2b subdomain region. However, as discussed below, more extensive deletions only resulted in marginally functional proteins, as indicated by swarming agar motility measurements.

Analysis of the sequence data in Table 2.2 also indicated that the start and stop positions of the deleted peptides were not randomly distributed. Many different variants had deletions that started or ended at the same residue position, suggesting the presence of deletion-tolerant “hot spots” in the flagellin sequence and structure. Deletions that started at residue Asp250, located at the C-terminal end of D3 domain β-strand 9, were the most frequent, with five variants encoding deletions that started at this position: C37, C39, C79, C88, and C89. Four variants, C108, C118, C120 and
C170, had deletions starting at residue Asp277 near the N-terminal end of D3 domain β-strand 11. Three variants, C21, C30 and C42, had deletions starting at residue Leu266, a D3 loop region; three other variants, C27, C33 and C115, had deletions starting at residue Lys279, in the middle of D3 β-strand 11. Six other pairs of variants had deletions that started in various positions in the D3 domain at residues Val247 (C61, C167), Thr257 (C14, C85), Gly261 (C4, C35), Gly269 (C110, C112), Thr275 (C48, C116) and Asn280 (C47, C174). Deletions that ended at residues Ala293 and Gly299, located in the middle and C-terminal end of helix 6, were the most frequent. Four variants had deletions ending at these two residue positions: C3, C4, C39, C167; C6, C11, C14, C21. Two groups of three variants encoded deletions ending at residues Leu288 (C42, C55, C112) and Ala304 (C33, C37, C62), which are located at the N-terminal start region of subdomain D2a helix 6 and the N-terminal start region of β-strand 12. Five other pairs of variants had deletions ending at identical residues: Val281 (C61, C70) and Gln282 (C5, C30), located in domain D3 C-terminal β-strand 11; Ala291 (C9, C182) and Ala298 (C89, C118), located in subdomain D2a helix 6; and Thr312 (C116, C174), located near the C-terminal end of subdomain D2a β-strand 12. To summarize, all of the most frequent functional deletion start positions were located in the outermost D3 domain and nearly all of the most frequent deletion end positions were located near the C-terminal end of the D3 domain or the N-terminal start region of the D2a subdomain.

The deleted sequence regions of the functional variants in Table 2.2 were similar to those previously observed by Kuwajima for a series of eight internal
deletion variants of a homologous, 497 amino acid residue *E. coli* flagellin protein. The largest deletion in the *E. coli* flagellin encoded 187 residues (194–380) in the middle region, compared with 177 deleted residues (227–403) for the C150 *S. typhimurium* flagellin variant in this study. It is difficult to make further inferences without more detailed structural information about the homologous *E. coli* flagellin protein, which may have a number of sequence substitutions, insertions and deletions relative to the *S. typhimurium* flagellin, as indicated by a Blast sequence homology search (data not shown).

There were also some similarities between the deletion variants isolated in this study and two *S. typhimurium* flagellin deletion variants previously described by Yoshioka, *et al.* Selection for *S. typhimurium* motility in the presence of an antibody specific for the hypervariable region yielded two spontaneous deletion variants, one containing a deletion of 89 residues encompassing Ala204–Lys292 (SJW46 strain) and a second deletion of 97 residues encompassing Thr183–Lys279 (SJW61 strain). The starting point in the deleted sequence for SJW46 (Ala204) was similar to the deletion start residue Thr201 for the C131 variant in this study (Table 2.2), which corresponds to the N-terminal helix 4-5 of the D3 domain. The SJW46 deletion ending at Lys292 was identical or similar to the deletion end hot spot at residues 291–293 in the C3, C4, C9, C35, C39, C167 and C182 variants in this study (Table 2.2), corresponding to the D2a subdomain helix 6. The N-terminal sequence region for the SJW61 variant deletion included part of the interdomain β-strand 3 region between the D2 and D3 domains and was 18 residues earlier in the sequence than any
of the variants isolated in this study. Thus, no obvious similarities were apparent
between the N-terminal deletion start region of the SJW61 variant and the 39 unique
variants in this study. The end of SJW61 deletion region was similar to that observed
in the C61 and C70 variants in this study (Table 2.2).

A subset of ten representative deletion variants with a wide range of internal deletions
(12–177 residues) were chosen for more detailed structural analysis, predictive
modeling and experimental characterization of physiological motility function (Table
2.1). The published sequences of the two SJW46 and SJW61 variants were also
included in the theoretical structural analysis. All twelve variant amino acid
sequences were analyzed with respect with respect to the wild-type flagellin structure
(PDB 1UCU), to determine deleted hydrophobic core residues, secondary structures
and domain regions and deduce possible structural trends. The deleted regions in the
primary sequence of each variant are depicted in Figure 8. The deletions were
predominantly in the D3 domain with some deletions extending into the D2 domain;
more of the D2b subdomain was deleted in variants with the most extensive deletions.
Table 2.1 Strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Description</th>
<th>Source / Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Strains.</strong></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td></td>
</tr>
<tr>
<td>XL1-Blue</td>
<td>Cloning strain <em>recA1 endA1 gyrA96 thi-1 hsdR17</em> <em>supE44 relA1 lac {F' progAB lacIqZΔM15 Tn10</em> (Tetr)}</td>
</tr>
<tr>
<td><strong>S. typhimurium</strong></td>
<td></td>
</tr>
<tr>
<td>SJW1103</td>
<td>LT2 derivative that is wild-type for motility and chemotaxis, phase 1 flagellin stable (<em>fliC</em> gene)</td>
</tr>
<tr>
<td>SJW134</td>
<td>Non-motile, flagellin deficient strain, Δ<em>fliC</em> and Δ<em>fliB</em>, derived from parent strain SJW806</td>
</tr>
<tr>
<td>JR501</td>
<td>Stable restriction-deficient (<em>r</em>'), methylation-proficient (<em>m</em>') galE cloning strain for providing restriction compatibility with <em>E. coli</em> plasmids</td>
</tr>
<tr>
<td><strong>B. Plasmids</strong></td>
<td></td>
</tr>
<tr>
<td>pTH890</td>
<td>pTrc99A-derived <em>Amp</em> expression plasmid for wild-type <em>S. typhimurium fliC</em> gene</td>
</tr>
<tr>
<td>pTH890-C3</td>
<td>Derivative of pTH890, expressing FliC Δ282-293</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The deleted regions in the wild-type flagellin tertiary structure are shown in Figure 2.9 and key structural features of the deleted regions are summarized in Table 2.3. All representative variants in this study had deletions of hydrophobic core residues that would otherwise contribute to the stability of the outer D2 and D3 domain structures (Table 2.3). For example, the C3 variant encoded removal of two residues (Ala286 and Ala291) from the hydrophobic core of the D2a domain, the moderately functional C9 variant had additional deletions of eight core residues from the D3 domain, and the non-functional C12 variant had deletions of 11 D3 domain residues.
core residues and 8 D2a subdomain core residues. Thus, it is not surprising that the deletion variant flagella structures were progressively less likely to fold and assemble properly, with negative consequences for the resulting oligomeric flagella structures, as more residues were deleted from the D2 and D3 domains.

![Multiple alignment of FliC and internal deletion variants](image)

**Figure 2.8 Multiple alignment of FliC and internal deletion variants.** The five different domains and subdomains are represented with different patterns. The dotted line indicates the region of the protein deleted in each variant. Domain D0 comprises the N-terminal residues from Ala1 to Ser32 and C-terminal segment of residues from Ala459 to Arg494; Domain D1 comprises the N-terminal residues from Asn44 to Glu1176 and C-terminal residues from Thr402 to Arg450; Subdomain D2a consists of the N-terminal residues from Lys177 to Gly189 and the C-terminal residues from Ala284 to Ile344. Subdomain D2b consists of residues from Asn345 to Ala401. Domain D3 is formed by residues Tyr190 to Val283.
A naïve prediction was that most functional deletions should start and end at flexible loop regions or at the ends of secondary structures, rather than in the middle of extensively hydrogen bonded α-helix and β-strand secondary structures. Analysis of the variant deleted sequences with respect to the wild-type flagellin structure indicated that a number of deleted peptide regions started and stopped near the ends of secondary structures (Figure 2.8, Table 2.3). The C4 peptide deletion started immediately after the end of a D3 domain β-strand. The C11 deletion encoded complete removal of the D2a α-helix and ended in a loop region. The C37, C79 and C89 variants had deletions that started at an apparent “hot-spot” (Val 249) near the C-terminal end of a D3 β-strand and ended at or near the N-terminal end of several D2a

### Table 2.2 Sequence parameters of functional internal deletion variants. *

<table>
<thead>
<tr>
<th>Clone</th>
<th>Deleted nucleotides</th>
<th>Number of deleted amino acid residues</th>
<th>Deleted peptide N-and C-terminal residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5</td>
<td>36</td>
<td>12</td>
<td>271-282</td>
</tr>
<tr>
<td>C3</td>
<td>36</td>
<td>12</td>
<td>282-293</td>
</tr>
<tr>
<td>C108</td>
<td>39</td>
<td>13</td>
<td>277-289</td>
</tr>
<tr>
<td>C55</td>
<td>48</td>
<td>16</td>
<td>273-288</td>
</tr>
<tr>
<td>C30</td>
<td>51</td>
<td>17</td>
<td>266-282</td>
</tr>
<tr>
<td>C110</td>
<td>54</td>
<td>18</td>
<td>269-286</td>
</tr>
<tr>
<td>C115</td>
<td>54</td>
<td>18</td>
<td>279-296</td>
</tr>
</tbody>
</table>

* Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 2.2 – Continued

<table>
<thead>
<tr>
<th>Clone</th>
<th>Deleted nucleotides $^b$</th>
<th>Number of deleted amino acid residues $^c$</th>
<th>Deleted peptide N-and C-terminal residues $^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C70</td>
<td>54</td>
<td>18</td>
<td>264-281</td>
</tr>
<tr>
<td>C112</td>
<td>60</td>
<td>20</td>
<td>269-288</td>
</tr>
<tr>
<td>C118</td>
<td>66</td>
<td>22</td>
<td>277-298</td>
</tr>
<tr>
<td>C42</td>
<td>66</td>
<td>22</td>
<td>266-288</td>
</tr>
<tr>
<td>C182</td>
<td>75</td>
<td>25</td>
<td>267-291</td>
</tr>
<tr>
<td>C27</td>
<td>75</td>
<td>25</td>
<td>279-303</td>
</tr>
<tr>
<td>C48</td>
<td>78</td>
<td>26</td>
<td>275-300</td>
</tr>
<tr>
<td>C33</td>
<td>78</td>
<td>26</td>
<td>279-304</td>
</tr>
<tr>
<td>C62</td>
<td>87</td>
<td>29</td>
<td>276-304</td>
</tr>
<tr>
<td>C120</td>
<td>90</td>
<td>30</td>
<td>277-306</td>
</tr>
<tr>
<td>C35</td>
<td>96</td>
<td>32</td>
<td>261-292</td>
</tr>
<tr>
<td>C4</td>
<td>99</td>
<td>33</td>
<td>261-293</td>
</tr>
<tr>
<td>C174</td>
<td>99</td>
<td>33</td>
<td>280-312</td>
</tr>
<tr>
<td>C21</td>
<td>103</td>
<td>34</td>
<td>266-299</td>
</tr>
<tr>
<td>C61</td>
<td>105</td>
<td>35</td>
<td>247-281</td>
</tr>
<tr>
<td>C47</td>
<td>114</td>
<td>38</td>
<td>280-317</td>
</tr>
<tr>
<td>C116</td>
<td>114</td>
<td>38</td>
<td>275-312</td>
</tr>
<tr>
<td>C6</td>
<td>123</td>
<td>41</td>
<td>259-299</td>
</tr>
<tr>
<td>C14</td>
<td>129</td>
<td>43</td>
<td>257-299</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 2.2 – Continued

<table>
<thead>
<tr>
<th>Clone</th>
<th>Deleted nucleotides $^b$</th>
<th>Number of deleted amino acid residues $^c$</th>
<th>Deleted peptide N- and C-terminal residues $^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C39</td>
<td>132</td>
<td>44</td>
<td>250-293</td>
</tr>
<tr>
<td>C88</td>
<td>138</td>
<td>46</td>
<td>250-295</td>
</tr>
<tr>
<td>C167</td>
<td>141</td>
<td>47</td>
<td>247-293</td>
</tr>
<tr>
<td>C9</td>
<td>141</td>
<td>47</td>
<td>245-291</td>
</tr>
<tr>
<td>C89</td>
<td>147</td>
<td>49</td>
<td>250-298</td>
</tr>
<tr>
<td>C170</td>
<td>162</td>
<td>54</td>
<td>277-330</td>
</tr>
<tr>
<td>C37</td>
<td>165</td>
<td>55</td>
<td>250-304</td>
</tr>
<tr>
<td>C85</td>
<td>186</td>
<td>62</td>
<td>257-318</td>
</tr>
<tr>
<td>C11</td>
<td>207</td>
<td>69</td>
<td>231-299</td>
</tr>
<tr>
<td>C79</td>
<td>225</td>
<td>75</td>
<td>250-324</td>
</tr>
<tr>
<td>C12</td>
<td>327</td>
<td>109</td>
<td>220-328</td>
</tr>
<tr>
<td>C131</td>
<td>504</td>
<td>168</td>
<td>201-368</td>
</tr>
<tr>
<td>C150</td>
<td>531</td>
<td>177</td>
<td>227-403</td>
</tr>
</tbody>
</table>

$^a$ Functional flagellin variants were identified in a swarming agar motility screen with non-motile *S. typhimurium* SJW134 cells expressing randomly generated internal deletion variants of phase 1 flagellin (Swiss-Prot ID number P06179). The amino acid sequence numbering system does not include the first N-terminal methionine that is post-translationally removed and ends at residue 494.

$^b$ Total number of deleted nucleotides in the mutated *fliC* gene encoded in each
Table 2.2 – Continued

<table>
<thead>
<tr>
<th>modified pTH890 plasmid.</th>
</tr>
</thead>
<tbody>
<tr>
<td>c Total number of internally deleted amino acid residues in each corresponding flagellin internal deletion variant that is encoded by the mutated fliC gene.</td>
</tr>
<tr>
<td>d Sequence identity of N- and C-terminal amino acid residues in the internally deleted peptide region of flagellin.</td>
</tr>
</tbody>
</table>

β-strands. The larger C12 deletion both started and ended at or near loop regions in the D3 and D2a domains. The larger C131 deletion started after a D3 β-strand, at the start of the D3 helical region and stopped at the N-terminal end of a D2b β-strand. The largest C150 variant deletion started near the N-terminal end of a D3 β-strand and ended in a loop region near the N-terminal end of a D1 domain α-helical region.

However, examples of deletions that started or ended in the middle of α-helix and β-strand secondary structures were also observed. The C3, C4 and C9 variants, with relatively small deletions, had only partial removal of the subdomain D2a α-helix. The C3 and C11 variants had significant partial deletions of β-strands in the deletion start regions. The C9 variant deletion started in the N-terminal region of a β-strand following a loop region in the D3 domain. The C79 and C89 deletions started in the C-terminal region of a D3 β-strand. The SJW46 variant deletion started...
in the middle of the D3 domain helix 4-5 and ended in the middle of the subdomain D2a helix 6. The SJW61 variant deletion started in the middle of subdomain D2a β-strand 3 and ended in the middle of domain D3 β-strand 11. Thus, some of the functional deletions started and ended in peptide regions with extensive hydrogen bonding.
Figure 2.9 Homology models of flagellin variants. *S. typhimurium* wild-type phase 1 flagellin structure (1UCU) and homology model structures of representative flagellin internal deletion variants developed with Modeler software (6v1). Structures are color coded by secondary structure; α-helices are red, β-strands are yellow and random coils are green. The insets show the hypervariable D2 and D3 domain region of wild-type flagellin with the deleted sequence for each variant colored in black.
Figure 2.9 – Continued
Figure 2.9 – Continued
It should also be noted that in some variants, adjacent anti-parallel \( \beta \)-strands were removed in pairs, e.g., the D2a domain \( \beta \)-hairpin turn motif in the C12 and C79 variants and another pair of \( \beta \)-strands in the D3 domain of the C9 variant. The C3 and C4 variants did not have any disruption of the \( \beta \)-folium motifs (Figure 2.8, Table 2.3). The C9, C89, C37 and C11 variants had partial deletions of the D3 domain \( \beta \)-folium. The C79 variant had deletions of parts of both the D3 and D2a \( \beta \)-folium motif. The C12 variant had complete deletion of the D3 \( \beta \)-folium and partial deletion of the D2a \( \beta \)-folium. The C131 variant had complete deletion of the D3 and D2a \( \beta \)-folium motifs and partial deletion of the D2b \( \beta \)-folium. The C150 variant had

<table>
<thead>
<tr>
<th>Deletion variant</th>
<th>Deleted residues (sequence)</th>
<th>End-to-end distance (Å)</th>
<th>Deleted secondary structure ( ^d )</th>
<th>Deleted core hydrophobic residues in each domain ( ^e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>12 (282-293)</td>
<td>21.4</td>
<td>D3: C-terminal segment of ( \beta )-strand 11. D2a: N-terminal segment of ( \alpha )-helix 6.</td>
<td>D2a: A287, A292</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Deletion variant</th>
<th>Deleted residues (sequence)</th>
<th>End-to-end distance (Å)</th>
<th>Deleted secondary structure</th>
<th>Deleted core hydrophobic residues in each domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>C89</td>
<td>49 (250-298)</td>
<td>32.0</td>
<td><strong>D3</strong>: Short segment of β–strand 9, β–strands 10–11.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a</strong>: Entire α–helix 6.</td>
<td></td>
</tr>
<tr>
<td>C37</td>
<td>55 (250-304)</td>
<td>25.2</td>
<td><strong>D3</strong>: Short segment of β–strand 9, β–strands 10–11.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a</strong>: Entire α–helix 6, C-terminal loop region.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a</strong>: Entire α–helix 6, short loop segment.</td>
<td></td>
</tr>
<tr>
<td>C79</td>
<td>75 (250-324)</td>
<td>28.2</td>
<td><strong>D3</strong>: Short segment of β–strand 9, β–strands 10–11.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a</strong>: Entire α–helix 6,</td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Deletion variant</th>
<th>Deleted residues (sequence)</th>
<th>End-to-end distance (Å)</th>
<th>Deleted secondary structure</th>
<th>Deleted core hydrophobic residues in each domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>C12</td>
<td>109 (220-328)</td>
<td>44.1</td>
<td>β-strand 12, long segment of β-strand 13.</td>
<td>A304, V306, M309</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D3:</strong> β-strands 7-11.</td>
<td><strong>D3:</strong> L220, A231, V233, V247, V249, V256, L258, A259, A262, V278, V281</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a:</strong> Entire α-helix 6, β-strands 12-13.</td>
<td><strong>D2a:</strong> A286, A291, L295, L298, V300, A304, V306, M309</td>
</tr>
<tr>
<td>C131</td>
<td>168 (201-368)</td>
<td>50.9</td>
<td><strong>D3:</strong> Majority of D3 domain, including β-strands 6-11 and most of helix 4-5.</td>
<td><strong>D3:</strong> L220, A231, V233, V247, V249, V256, L258, A259, A262, V278, V281</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a:</strong> Majority of subdomain, including α-helix 6, β-strands 12-15.</td>
<td><strong>D2a:</strong> A286, A291, L295, L298, V300, A304, V306, M309, A334, I344</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2b:</strong> β-strands 16-18.</td>
<td><strong>D2a:</strong> A231, V233, V247, V249, V256, L258, A259, A262, V278, V281</td>
</tr>
<tr>
<td>C150</td>
<td>177 (227-403)</td>
<td>54.0</td>
<td><strong>D3:</strong> β-strands 8-11.</td>
<td><strong>D2a:</strong> A231, V233, V247, V249, V256, L258, A259, A262, V278, V281</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>D2a:</strong> Most of subdomain, including α-helix 6, β-strands 12-15.</td>
<td><strong>D2a:</strong> A231, V233, V247, V249, V256, L258, A259, A262, V278, V281</td>
</tr>
</tbody>
</table>

61

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 2.3 – Continued

<table>
<thead>
<tr>
<th>Deletion variant (sequence)</th>
<th>Deleted residues (sequence)</th>
<th>End-to-end distance (Å)</th>
<th>Deleted secondary structure&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Deleted core hydrophobic residues in each domain&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJW46 (204-292)</td>
<td>89</td>
<td>35.4</td>
<td><strong>D3</strong>: Majority of D3 domain, including β-strands 6–11 and part of helix 4-5. <strong>D2a</strong>: N-terminal segment of α–helix 6.</td>
<td>L258, A259, A262, V278, V281</td>
</tr>
</tbody>
</table>

<sup>a</sup> Secondary structure regions are defined in the Results section. *S. typhimurium* phase 1 flagellin internal deletion variants include ten representative variants identified in this study and two variants (SJW46, SJW61) previously described by Yoshioka, *et al.*<sup>33</sup>.  

<sup>b</sup> Total number of deleted amino acid residues in each internal deletion variant and N- and
Table 2.3 – Continued

<table>
<thead>
<tr>
<th>C-terminal ends of deleted region (in parentheses).</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^c$ End-to-end distances between the $\alpha$-carbons of remaining N- and C-terminal residues flanking internal deletions of each deletion variant in the wild-type phase 1 flagellin structure (PDB 1UCU), as measured with the SwissPDBViewer 3.7 software distance measurement tool.</td>
</tr>
<tr>
<td>$^d$ Description of deleted regions of secondary structure in D2 and D3 domain regions of each flagellin variant, corresponding to inset regions in Figure 7.</td>
</tr>
<tr>
<td>$^e$ Deleted core hydrophobic residues in D2 and D3 domains of each variant.</td>
</tr>
</tbody>
</table>

partial, almost complete deletion of the D3 $\beta$–folium and complete deletion of the D2a and D2b $\beta$–folium motifs. The SJW46 and SJW61 variants both had complete removal of the D3 $\beta$–folium, with no disruption of the other D2a and D2b $\beta$–folium motifs.

Larger deletions in the marginally functional C12, C131 and C150 variants encoded complete removal of the D2a helix 6, most of the D3 domain and large sections of the D2a (C12) and D2b subdomains (C131, C150) (Table 2.3). Variant C12 had deletion of subdomain D2a helix 6 and $\beta$–strands 12 and 13 in a $\beta$–hairpin motif, similar to the C79 variant, but also had an extensive deletion of most of the outermost D3 domain, except for an N-terminal segment comprising D3 domain.

63
β–strands 4, 5, the helix 4-5 and β–strand 6. The C131 variant had deletions of additional β–strands on both the D2 and D3 domains, with almost complete removal of the outermost D3 domain, except for three successive β–strands 3–5 that extend from subdomain D2a into domain D3. Most of the D2a subdomain and part of the D2b domain were also removed, leaving only a partial D2b subdomain including β–strands 19 and 20. The C150 variant had nearly complete deletion of the D2a and D2b domains; the remaining peptide included subdomain D2a β–strand 3 and part of the D3 domain that included β–strands 4-7 and helix 4-5. None of the ten representative variants in this study had complete deletion of the D3 domain and corresponding preservation of the D2 domain. However, the previously described SJW46 mutant had most of the outermost D3 domain deleted (Figure 2.8), similar to the C131 variant, but with only a minimal deletion of the D2a domain, including partial deletion of the D2a α–helix. The SJW61 variant had the most complete deletion of the D3 domain and a relatively minimal disruption of the D2 domain region, with complete preservation of the subdomain D2a α–helical region (Figure 2.8). Both of these spontaneous mutations resulted in deletions that appear to clip off the majority of the D3 domain and part of the D2a subdomain, with little disruption to the D2b subdomain. It is straightforward to envision how the remaining D3 domain β–strand 11 segment could flip over to join the other end of the remaining β–strand 3 segment in the SJW61 mutant, with minimal disruption to the D2 domain region, as observed in the predictive model. These two prior “domain-clipping” mutations suggest a rational engineering approach to deleting and inserting other proteins in the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
hypervariable region of flagellin; however, this approach remains to be experimentally validated.

The end-to-end distances of remaining, undeleted peptide segments, with respect to the wild-type flagellin structure, were determined for the 12 variants (Table 2.3). These measurements can be considered to be a rough estimate of the possible degree of structural distortion and rearrangement required to join together the two remaining peptide segments to yield an intact protein. The end-to-end distances of the seven relatively motile variants (C3, C4, C9, C11, C37, C79 and C89) ranged from 21.4–38.4 Å, with an average value of 31.1 ± 6.4 Å. The end-to-end distances for the motile SJW46 and SJW61 variants were also similar, 35.4 Å and 33.6 Å. It is also interesting to note that the average end-to-end distance is similar to the end-to-end distance of 26 Å for the N- and C-termini (Ser1-Ala108) of E. coli thioredoxin, previously inserted in an E. coli flagellin internal deletion variant to yield the hybrid FliTrx protein.37 The three relatively non-motile C12, C131 and C150 variants, encoding the largest deletions, had correspondingly larger end-to-end distances of 44–54 Å (Table 3). Thus, a rough correlation was apparent for the end-to-end distance between remaining residues in the original wild-type structure and the number of deleted residues (Figure 2.10); larger internal deletions tended to result in larger end-to-end distances. However, this correlation was not entirely linear; the C4 and C9 variants had a local maximum in their end-to-end distances of 36 Å that coincided with a relatively minimal change in the motility for these same variants. A second local maxima of 38 Å in the end-to-end distance was apparent for the C11
variant. Variation in the deletion start and stop positions and intrinsic topological constraints of the folded protein structure probably account for these variations. For example, removal of an entire β-hairpin could result in a smaller end-to-end distance between remaining peptide segments than removal of only one strand of this same structure.

Given that deletions should alter the remaining protein conformation, we attempted to develop predictive structural models of the twelve deletion variants. Homology modeling of all 12 variants was performed with the MODELER software package, using the *S. typhimurium* flagellin PDB structure (1UCU) as the template (See Supplemental Data for the model PDB coordinate files). We were able to model structures of the ten C3, C4, C9, C11, C12, C37, C79, C89, SJW46 and SJW61 variants often with reasonable accuracy (Figure 2.8), but were unable to model the C131 and C150 variant structures that contained extensive deletions of both the D2 and D3 domains. Thus, one preliminary result from this *in silico* experiment was that the ability of this software algorithm to predict a compact folded domain structure decreased as the size of the internal deletion increased.
Figure 2.10 Plot of normalized relative cell motility. Plot of normalized relative cell motility of wild-type and ten internal deletion variants of S. typhimurium phase 1 flagellin and corresponding end-to-end distances of remaining peptide segments in the wild-type structure (PDB 1UCU). Normalized motility values and end-to-end distances (Table 4) were plotted as a function of the number of deleted residues for each variant. Three types of cellular motility were determined by agar swarming motility; video microscopy swimming velocity and optical trapping stall force measurements, as described in Materials and Methods. The motility data in Table 3 were normalized by computing the ratio of each value relative to the baseline value determined for the SJW134 strain (transformed with pTH890) expressing wild-type flagellin. The larger of the two normalized motility values for zero deleted residues are for the wild-type motile SJW1103 strain. The plotted normalized swimming velocity and trapping force values were vertically offset by 0.4 and 0.8 for clarity.
Conservation of wild-type flagellin secondary structures and hydrophobic cores in the outer domains of the ten converged variant models was analyzed using the Swiss-PdbViewer software. All ten of the deletion variant models had an intact D2b domain with respect to its secondary structures. Helix 4-5 in the D3 domain was intact in most of the models, except for the C12 variant model, where it assumed more of a loop structure. The model of the C3 variant, with only 12 deleted residues, indicated that the subdomain D2a α-helix 6 was completely replaced by an extended loop connecting the D3 domain and D2a subdomain. The corresponding β-strands 4, 5 and 11 in the D3 domain were replaced by loops. The subdomain D2a β-strands 3, 12 and 13 were also replaced by loops. The model of the C4 variant, with 33 residues deleted primarily from the D3 domain (261-293), shows disruption of the hydrophobic core region in the D3 domain. The remaining D3 domain β-strands 4, 5, 10 and 11 were disrupted, while the D2a domain secondary structures remained intact. The C9 variant model, with a deletion of 47 residues (244–291), showed disruption of β-strands 4, 5 and 11 in the D3 domain and disruption of β-strands 3, 12 and 13 and α-helix 5 in the D2a domain. The C89 variant model with 49 deleted residues (250–298) had disruption of β-strands 4, 5, 9, 10 and 11 in the D3 domain and α-helix 6 in the D2a domain. The C37 variant model (55 deleted residues; 249–304) had disruption of all remaining β-strands in the D3 domain and β-strands 3, 12 and 13 and α-helix 6 in the D2a domain. The C11 variant model with 69 deleted residues (230–299) had deletions that were predominantly in the D3 domain.
causing either disruption of the secondary structures or their complete elimination.
The β-strands in the D2a domain appeared to be intact but α-helix 6 was replaced by a loop. The C79 variant, with 75 residues deleted (249–324), had deletions that occurred in most of the D3 and D2 domains, disrupting the hydrophobic cores in both domains. The C79 variant model indicated that β-strands 4, 5, 10 and 11 were disrupted in the D3 domain and no secondary structures were present in the D2a domain. The model of the C12 variant did not have preservation of any secondary structures in the D3 domain while β-strands 12 and 13 and α-helix 6 were disrupted in the D2a domain.

The remaining outer domain peptides in the C11, C12, and C37 models were predicted to form large segments of extended, random-coil loops, with minimal hydrogen bonding. These loop regions also contained significant numbers of solvent-exposed hydrophobic side chains that did not collapse to form buried hydrophobic cores. The lack of compact globular structure in the outer domain regions of these three models suggests that they may not be very accurate. Thus, it cannot be ruled out that these predicted loop regions may assume other conformations that would result in more compact, domain-like structures. The model of the SJW46 variant indicated complete removal of the D3 domain region. The β-strands 12-15 and α-helix 6 in the D2a region and β-strands 16–20 in the D2b region are preserved. The SJW61 model also indicated complete removal of the D3 domain region and preservation of the β-strands 12-15 and α-helix 6 in the D2a region and β-strands 16–20 in the D2b region are preserved.
These preliminary structural models suggest that even a deletion of a few residues in the outer domains can perturb the folded structure, resulting in decreased protein and fiber stability and consequently, lowered cellular motility. The modeling predictions involving extended loop regions and the failure to model the C131 and C150 variants were probably due to the lack of an appropriate template protein structure; the MODELER algorithm is very dependent on existence of a known protein structure to predict unknown structures. More detailed modeling studies with a different structure prediction algorithm that is not dependent on a single template are underway.

**Purification and transmission electron microscopy analysis of flagella**

Although the models and sequence analysis were revealing, we also pursued experimental measurement of the macroscopic, physiological effects of the deletions on oligomeric flagellar structure and motility for the ten representative flagellin internal deletion variants. The presence of intact flagella fibers that were still attached to cells was verified using transmission electron microscopy (TEM) of both *Salmonella typhimurium* cells and media samples in which cells were grown. TEM images of the wild-type and deletion variant flagella revealed subtle differences (Figure 2.11). TEM imaging indicated that the C3, C4 and C9 flagella fibers exhibited near wild-type mechanical stability in cell culture. However, the other variants were substantially more brittle than the wild-type protein and were prone to breakage. When SJW134 cells expressing the C3, C4, C9, C89, C37, C11 and C79 flagellin variants were grown with rigorous shaking (225 rpm), flagella were mostly found in the media as
detached, isolated fibers, indicating that they were more readily sheared off the cells than the wild-type protein. The C12 and C131 variants formed shorter fibers which were also mostly found as isolated fibers in the media. However, all variants formed flagella that were still attached to the cells, similar to the wild-type, under shear-free growth conditions. Flagella filaments with larger internal deletions have previously been observed to be less mechanically stable than the wild-type flagellin, e.g., the SJW46, SJW61 and SJW46S *S. typhimurium* flagellin internal deletion variants. Large internal deletions may destabilize the conformation of flagellin, leading to weaker interactions between flagellin subunits and between flagellins and other flagellar hook proteins. No significant aggregation of the C3, C4, C9, C89, C37, C11 and C79 variants into bundles was observed, in contrast to the previously described SJW46 and SJW61 variants, which aggregated into flagellar bundles in the presence of more than 0.1 M NaCl.
Figure 2.11 TEM images of flagella fibers. Electron micrographs of flagella fibers or protein aggregates isolated from *S. typhimurium* SJW134 cells. (a) wild-type cells. (b) C79 variant; (c) and (d) C12 variant and (e) and (f) C150 variant. The scale bars indicated in the images are as follows: (a) 100 nm, (b) 100 nm, (c) 100 nm, (d) 100 nm, (e) 200 nm and (f) 2 μm.

Flagellin protein monomers and flagellar fibers of each variant were isolated from bacterial cells by heating or shearing, as described in Materials and Methods and were characterized by SDS-PAGE and TEM imaging. Analysis by SDS-PAGE revealed similar extracellular amounts of each flagellin variant (Figure 2.12), with different molecular masses corresponding to the smaller PCR products for each internally deleted *fliC* gene. These results indicated that similar amounts of each flagellin
deletion variant were exported for all ten variants. TEM images of the C12 and C131 variant flagella (Figure 2.10(c), (d)) indicated sharper edges compared to the wild-type and other variants (Figure 2.10(a), (b)), suggesting a deletion of most of the outer domains. Electron microscopy of the shear isolated C150 variant did not indicate intact flagella fibers but instead revealed large amorphous aggregates of protein (Figure 2.10(e), (f)) that were not observed for the other nine variants. These aggregates, along with the SDS-PAGE results, indicate that correct export of the smallest C150 variant was followed by aggregation and failure to assemble into fibers, resulting in accumulation in the media. This result is consistent with a previous result in which expression and secretion of *E. coli* flagellin internal deletion variants were similar to wild-type, but assembly into filaments was minimal for the smallest variant.32

![Figure 2.12 SDS-PAGE of the FliC and deletion variant proteins. A 10% SDS-PAGE of FliC proteins from wild-type SJW1103 *Salmonella* strain, pTH890 plasmid and other deletion variants.](image)
Swarming agar motility assay

The physiological function of the ten flagellin internal deletion variants was first characterized by a timed swarming motility assay performed on dilute 0.3% agar (Table 2.4). The swarming diameter of each variant expressed in SJW134 S. typhimurium cells was determined after 6 h; both wild-type SJW1103 and non-motile SJW134 S. typhimurium cells transformed with the pTH890 plasmid were measured as controls (Figure 2.13). The measured motility for both controls were very similar, suggesting that the rate of flagellin expression from the pTH890 plasmid was not a significantly rate-limiting factor in the production of functional flagella in the non-motile SJW134 strain. The seven C3, C4, C9, C89, C37, C11 and C79 variants showed a detectable swarming radius while the C12, C131 and C150 variants did not show any significant swarming motility within six h, similar to the non-motile SJW134 strain. These latter three variants only showed weak motility after longer periods of growth, e.g., 12–16 h. All seven of the relatively motile variants showed significant decreases of 23–58% in their swarming diameters relative to SJW134+pTH890 cells expressing the wild-type flagellin (Table 2.4, Figure 2.9). Thus, a detectable loss of function was observed even for the least perturbed C3 variant with a small deletion of 12 residues. Furthermore, the swarming radius was largely proportional to the extent of deletion (Figure 2.13), although the C4 and C9 variants had similar motilities, despite deletion of an additional 14 residues for the C9 variant. These results were largely consistent with a prior study of E. coli flagellin deletion variants.32
Figure 2.13 Motility assay of internal deletion library. Images of swarming motility agar assays for wild-type controls and selected internal deletion variants of S. typhimurium flagellin. Images were collected after 6 h growth at 30 °C.

Table 2.4 Motility of S. typhimurium SJW134 cells expressing FliC and deletion variants as determined by different methods.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Number</th>
<th>Swarming diameter (mm)</th>
<th>Swimming speed (μm/s)</th>
<th>Optical trapping stall force (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJW1103</td>
<td>n.a.</td>
<td>36.5 ± 6.5</td>
<td>13.7 ± 3.4</td>
<td>&gt; 2.0</td>
</tr>
<tr>
<td>(wild-type)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SJW134 + pTH890</td>
<td>n.a.</td>
<td>34.4 ± 4.3</td>
<td>11.2 ± 3.0</td>
<td>1.7 ± 0.3</td>
</tr>
<tr>
<td>C3</td>
<td>12</td>
<td>26.5 ± 1.9</td>
<td>10.5 ± 4.3</td>
<td>1.3 ± 0.3</td>
</tr>
<tr>
<td>C4</td>
<td>33</td>
<td>22.5 ± 2.6</td>
<td>9.0 ± 2.8</td>
<td>1.2 ± 0.3</td>
</tr>
</tbody>
</table>
Table 2.4 – Continued

<table>
<thead>
<tr>
<th>Variant</th>
<th>Number</th>
<th>Swarming diameter (mm)</th>
<th>Swimming speed (μm/s)</th>
<th>Optical trapping stall force (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>47</td>
<td>22.5 ± 1.7</td>
<td>10.1 ± 3.8</td>
<td>1.3 ± 0.3</td>
</tr>
<tr>
<td>C89</td>
<td>49</td>
<td>19.5 ± 1.9</td>
<td>9.3 ± 3.6</td>
<td>1.2 ± 0.3</td>
</tr>
<tr>
<td>C37</td>
<td>55</td>
<td>19.2 ± 1.7</td>
<td>8.3 ± 3.6</td>
<td>1.0 ± 0.3</td>
</tr>
<tr>
<td>C11</td>
<td>69</td>
<td>17.2 ± 4.7</td>
<td>9.0 ± 3.0</td>
<td>0.8 ± 0.2</td>
</tr>
<tr>
<td>C79</td>
<td>75</td>
<td>14.5 ± 2.1</td>
<td>6.0 ± 2.6</td>
<td>0.6 ± 0.3</td>
</tr>
<tr>
<td>C12</td>
<td>109</td>
<td>9.5 ± 1.7</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>C131</td>
<td>168</td>
<td>9.2 ± 1.0</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>C150</td>
<td>177</td>
<td>9.2 ± 0.5</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>SJW134 (control)</td>
<td>n.a.</td>
<td>8.2 ± 1.0</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

The non-motile S. typhiumurium strain SJW134 was transformed with pTH890 expression plasmids encoding either wild-type phase 1 flagellin or internal deletion variants shown in Figure 12. The untransformed SJW134 strain was used as a negative control for swarming agar motility measurements. Experiments in liquid media were performed using cell cultures grown to mid-log growth phase without shaking.

The number of deleted amino acid residues in each flagellin internal deletion variant, corresponding to the sequence schematic shown in Figure 7.
Table 2.4 – Continued

\(^c\) Swarming diameter of *S. typhimurium* SJW134 cell expressing the wild-type and internal deletion variants of phase 1 flagellin on 0.3% agar media at 30 °C for 6 hours. Each value is the average and standard deviation of five replicates.

\(^d\) Swimming velocity in aqueous LB media of SJW134 cells expressing wild-type and internal deletion variants of phase 1 flagellin. Values are the average and standard deviation of video microscopy measurements performed on 20 different cells, as described in Materials and Methods.

\(^e\) Minimum stall force required to optically trap individual SJW134 cells expressing wild-type and internal deletion variants of phase 1 flagellin in 6.25% glycerol LB media, as described in Materials and Methods. Values are the average and standard deviation of 20 measurements.

*n.a.* Not applicable.

*n.d.* No data.

---

**Video analysis of swimming motility in dilute aqueous media**

A second experimental technique of video microscopy was used to measure the effects of internal deletions on the swimming ability of cells in aqueous solution. The swimming rates of both wild-type SJW1103 and non-motile SJW134 *S. typhimurium* cells expressing wild-type flagellin and the ten representative variants were determined in liquid LB media (Table 2.4). For any given population of cells the

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
swimming rates can considerably vary from individual to individual and the swimming speed of the same species of bacteria can vary depending on different conditions. Hence, the average velocity was determined for 20 different cells expressing each variant under otherwise identical media conditions. Stereotypical motility behavior was observed for individual cells, with periods of smooth swimming in one direction interrupted by tumbling, followed by a change in the direction of movement. The SJW1103 wild-type cells had an average swimming speed of 13.7 ± 3.2 μm/s, while the SJW134 cells transformed with the pTH890 expression plasmid encoding wild-type flagellin had an average swimming speed of 11.2 ± 3.0 μm/s. The deletion variants showed significant reductions in swimming speed, with measured values ranging from 10.5 μm/s for C3 (6.3% reduction in velocity) to 6.0 μm/s for C79 (46% reduction in velocity), relative to the SJW134 cells expressing wild-type flagellin (Table 2.4, Figure 2.9). The C12, C131 and C150 variants did not show any measurable swimming ability in liquid media apart from random Brownian motion, in agreement with the previous swarming agar motility studies.

Measurement of stall force in viscous aqueous media by optical trapping
A third technique of optical trapping was employed to measure the relative propulsive force generated by cells expressing the flagellin deletion variants. Optical traps ("laser tweezers") allow manipulation of individual cells through the forces exerted on a particle by a focused laser beam, as first demonstrated by Ashkin. Optical traps have been employed in understanding different bacterial motions and

78
quantitation of the forces generated by their motility structures such as type IV pili and endoflagella, and the torque generated by single flagellum in *E. coli*. Optical traps have also been employed to measure the force generated by motile eukaryotic cells such as human sperm cells and *Chlamydomonas* sp. algae. One of the major concerns while using optical traps on living systems is the extent of photodamage to the specimen. The high energy laser light source can cause damage to the specimens and could eventually lead to their death, a process termed "optication". However, studies conducted on living systems have shown that the near-IR (790–1064 nm) wavelengths are relatively harmless for trapping live cells. Neuman et al. studied the effects of near-IR lasers of different wavelengths on *E. coli*, and concluded that the most damaging wavelength to be 870 nm and least damaging to be 970 nm. Thus, the wavelength of 1064 nm used in this study is considered to be relatively safe for using with the live specimens.

The relative propulsive forces exerted by wild-type *S. typhimurium* SJW1103 cells and SJW134 cells expressing wild-type flagellin and each of the seven motile deletion variants (C3, C4, C9, C11, C37, C79 and C89) were characterized by optical trapping (Table 2.4). These experiments were based on the assumption that all cells expressing flagellin deletion variants, and grown without shaking, produce similar numbers and lengths of flagella, as indicated by the TEM images and the similar amounts of extracellular flagellin protein indicated by SDS-PAGE of the isolated flagella (Figure 2.10). The average stall force required to stop movement of the cells in media containing 6.25% glycerol was measured by decreasing the power of the
laser until cells were able to escape the optical trap. Once the cells were optically trapped, they aligned vertically with their long axis perpendicular to the image plane, parallel to the laser beam, and rapidly rotated about the Z-axis. As the laser power was reduced, the motile cells started to move in different directions. Repeated trapping and releasing did not appear to damage the cells, as indicated by consistent stall forces measured over time for the same cell. The majority of the wild-type SJW1103 cells were difficult to trap, even in the relatively viscous glycerol media and had an average stall force of more than 2.0 W (Table 2.4). The SJW134 cells transformed with the pTH890 plasmid had an average trapping force of 1.7 W. Conversely, the non-motile SJW134 cells that lack flagella were not able to escape the trap even after reducing the laser power to the lowest setting of 0.2 W. Cells expressing the functional internal deletion variants exhibited decreases in their measured stall forces ranging from 24–65% for the C3 and C79 variants, relative to the SJW134 cells expressing wild-type flagellin (Table 2.3). A steep initial decrease in the stall force was followed by a slower rate of decrease with respect to number of deleted residues (Figure 2.9); the stall force was relatively constant for the C3, C4, C9 and C89 variants despite deletions ranging from 12–49 residues. These observations were consistent with the prior swimming velocity and agar motility measurements.

Comparison of motility results determined by different methods

A strong correlation was observed between the number of deleted residues and the cellular motility, for variants with deletions less than 100 residues, as determined by three different experimental methods (Figure 2.9). All three methods yielded
comparable results for relative decreases in cellular motility as a function of deleted residues, although optical trapping appeared to be slightly more sensitive to changes in the protein-encoded motility function than the other two methods. For example, the following changes were observed for the least perturbed C3 variant vs. the control SJW134 strain expressing wild-type flagellin: a 23% decrease in agar swarming diameter, a 6.3% decrease in the aqueous swimming speed and a minimum stall force of 24% by optical trapping in 6.25% glycerol media. The C79 variant, with a much larger deletion of 75 residues, had decreases of 58% in agar swarming diameter, 46% in the aqueous swimming speed, and 65% in the optical trapping stall force.

Following the large initial decrease in motility observed for the C3 variant, the motility values were relatively constant for the C3, C4 and C9 variants with deletions of 12–47 residues (Table 2.4, Figure 2.9). This analysis suggests that the C9 variant may have a more optimal folded protein structure than the other variants with slightly smaller or larger deletions. The motility decreased much more rapidly as a function of number of deleted residues for the C89, C37, C11 and C79 variants encoding deletions of more than 47 residues. Deletion of more than 100 residues in the C12, C131 and C150 variants resulted in nearly complete loss of cellular motility.

Discussion

Effects of internal deletions on cellular motility and stability of flagella

The remarkable sequence variations in the hypervariable region of bacterial flagellins make this class of protein an ideal candidate for investigating the structure-function
relationships of internal domains. Previous studies have shown that the hypervariable middle sequence region of *E. coli* and *S. typhimurium* flagellin proteins can be deleted to varying extents without complete loss of Type III export, folding, self-assembly and motility functions. The recent availability of a complete structure of the *S. typhimurium* flagellin protein\(^{15}\) enabled a more detailed molecular approach to investigating the structure-motility function relationships of flagellin internal deletion variants. This study demonstrated that many different functional internal deletion variants are possible, as long as the deletions are confined to regions of the solvent-exposed D2 and/or D3 domains, which are not critical for export and self-assembly. Although this was not an exhaustive study of all possible functional variants, the D0 and D1 domains, including the putative β-switch region in the D1 domain, were not altered in any of the functional sequences identified. This result is in agreement with previous studies indicating that the conserved N- and C-terminal D0 and D1 domain regions are critical for export and flagellar assembly.

Three different physiological measurements indicated that the motility of cells generally decreased as the size of the outer domains decreased, under a wide range of viscosity conditions, suggesting that the D2 and D3 domains are important in stabilizing the proper conformation of the flagellin protein and contributing to its role in propulsion. Removal of even small peptides from the hypervariable domain resulted in a measurable decrease in motility, although deletions of up to 75 residues, or 15% of the entire sequence, still resulted in partial motility. Lowered motility resulting from internal deletions may be a permissible evolutionary compromise
between colonization of the environment and evasion of host immune responses to the wild-type flagellin.

The mechanical stability of the flagella fibers decreased as more of the D2 and D3 domain regions were deleted. The lowered mechanical stability observed as may be due both to decreased stability of the folded monomer and decreased strength of intermolecular interactions between the D2 domains in the oligomeric flagella fibers. It should also be noted that thermostable flagellins from extremophilic Aquifex pyrophilus and A. aeolicus eubacterial species\textsuperscript{78, 79, 80} have similar numbers of amino acid residues (~500) to the wild-type S. typhimurium phase 1 flagellin. These thermostable flagellins have hypervariable regions that may also encode structurally-stabilizing domains. It can be concluded that the outermost D3 domain “knob” is important for protein stability and contributes to mechanical stability of oligomeric flagella and consequently, motility function, under all conditions. Depending on the species, the outer D2 and D3 domain sequence regions can be greatly reduced in size. It remains to be determined what structural mechanisms other flagellins with smaller hypervariable regions use for stability and how they compare in motility relative to the S. typhimurium flagellin.

**Implications for folding, export and assembly**

While large deletions significantly destabilized flagellins and their final assembly into oligomeric flagella, they did not appear to perturb efficient extracellular export. The formation of large extracellular aggregates instead of flagellar fibers observed for the C150 variant with most extensive deletions also raises the question of possible
molecular interactions between the outer D2 and D3 domains and the terminal FliD/Hap2 chaperone complex during assembly in vivo.

The globular D2a, D2b and D3 domains have hydrophobic cores at their centers which contribute to the folded stability of the protein. Removal of protein segments also removes portions of these hydrophobic cores, contributing to lowered stability. The variant end-to-end distances in the wild-type structure, on the order of the outer diameter of the D3 domain (30 Å), clearly indicate that some form of conformational rearrangement is required for the remaining polypeptide ends to be joined together to form a functional protein. Thus, in addition to disruption or complete loss of hydrophobic core regions, joining the ends of remaining peptide segments may introduce strain into the variant structures. This may also lower the folding stability of flagellin monomers and may also interfere with the packing of the D1 and any remaining D2b domain segments in the oligomeric flagella fiber.

The wild-type D2 and D3 domains have a large proportion of β-strand-rich structures that may be relatively flexible. These domains probably rearrange from some type of precursor structure, after export through the 20 Å diameter inner pore of the flagellum, to fold and assemble into the final oligomeric wild-type structure. Given the tendency of β-strands to pack together, some collapse and rearrangement of formerly distant segments may occur to give new β-sheet structures. It is also well known that some β-sheet proteins can undergo conformational rearrangements, e.g., silk proteins and β-amylid proteins. Thus, some of the remaining peptide segments regions may also assume alternative conformations. Rearrangements of the
remaining peptide may also result in generation of relatively unstructured loops that may further destabilize the folded state. However, it is not obvious from the preliminary simulated structures for many of the variants with larger deletions how the remaining segments in each D2 and D3 domain fold, or whether there is even a single folding solution for these modified domains.

**Implications for design of flagellin fusion proteins**

Given the extracellular display of flagellins, the hypervariable region would be a logical place to insert other proteins domains of interest, as previously demonstrated with thioredoxin. Some of the variants analyzed in this study suggest possible locations for rational insertion of foreign proteins as N- and C-terminal fusions in the hypervariable region of flagellin. One obvious approach would be to remove the D3 domain and genetically insert other proteins at the exposed ends of the D2a domain. Other approaches would involve insertion between segments of the D3 domain and D2a or D2b subdomains. Recovery of motility with otherwise non-functional variants could potentially be used as a simple selection method to identify functional protein insertion variants.

**GenBank accession numbers**

The sequences of the 39 unique internal deletion variants of the *S. typhimurium* fliC gene have been deposited in the GenBank database (Table 2.5). The corresponding DNA sequences can be found in the Appendix I and a multiple sequence alignment of all the 39 unique clones can be found in Appendix II.
<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Genbank Acc. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTH890-C5</td>
<td>EF057752</td>
</tr>
<tr>
<td>pTH890-C3</td>
<td>EF057753</td>
</tr>
<tr>
<td>pTH890-C108</td>
<td>EF057754</td>
</tr>
<tr>
<td>pTH890-C55</td>
<td>EF057755</td>
</tr>
<tr>
<td>pTH890-C30</td>
<td>EF057756</td>
</tr>
<tr>
<td>pTH890-C110</td>
<td>EF057757</td>
</tr>
<tr>
<td>pTH890-C115</td>
<td>EF057758</td>
</tr>
<tr>
<td>pTH890-C70</td>
<td>EF057759</td>
</tr>
<tr>
<td>pTH890-C112</td>
<td>EF057760</td>
</tr>
<tr>
<td>pTH890-C118</td>
<td>EF057761</td>
</tr>
<tr>
<td>pTH890-C42</td>
<td>EF057762</td>
</tr>
<tr>
<td>pTH890-C182</td>
<td>EF057763</td>
</tr>
<tr>
<td>pTH890-C27</td>
<td>EF057764</td>
</tr>
<tr>
<td>pTH890-C48</td>
<td>EF057765</td>
</tr>
<tr>
<td>pTH890-C33</td>
<td>EF057766</td>
</tr>
<tr>
<td>pTH890-C62</td>
<td>EF057767</td>
</tr>
<tr>
<td>pTH890-C120</td>
<td>EF057768</td>
</tr>
<tr>
<td>pTH890-C131</td>
<td>EF057789</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Genbank Acc. Number</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>pTH890-C35</td>
<td>EF057769</td>
</tr>
<tr>
<td>pTH890-C4</td>
<td>EF057770</td>
</tr>
<tr>
<td>pTH890-C174</td>
<td>EF057771</td>
</tr>
<tr>
<td>pTH890-C21</td>
<td>EF057772</td>
</tr>
<tr>
<td>pTH890-C61</td>
<td>EF057773</td>
</tr>
<tr>
<td>pTH890-C47</td>
<td>EF057774</td>
</tr>
<tr>
<td>pTH890-C116</td>
<td>EF057775</td>
</tr>
<tr>
<td>pTH890-C6</td>
<td>EF057776</td>
</tr>
<tr>
<td>pTH890-C14</td>
<td>EF057777</td>
</tr>
<tr>
<td>pTH890-C39</td>
<td>EF057778</td>
</tr>
<tr>
<td>pTH890-C88</td>
<td>EF057779</td>
</tr>
<tr>
<td>pTH890-C167</td>
<td>EF057780</td>
</tr>
<tr>
<td>pTH890-C9</td>
<td>EF057781</td>
</tr>
<tr>
<td>pTH890-C89</td>
<td>EF057782</td>
</tr>
<tr>
<td>pTH890-C170</td>
<td>EF057783</td>
</tr>
<tr>
<td>pTH890-C37</td>
<td>EF057784</td>
</tr>
<tr>
<td>pTH890-C85</td>
<td>EF057785</td>
</tr>
<tr>
<td>pTH890-C11</td>
<td>EF057786</td>
</tr>
<tr>
<td>pTH890-C79</td>
<td>EF057787</td>
</tr>
<tr>
<td>pTH890-C12</td>
<td>EF057788</td>
</tr>
</tbody>
</table>
Table 2.5 – Continued

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Genbank Acc. Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>pTH890-C150</td>
<td>EF057790</td>
</tr>
</tbody>
</table>

References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Chemistry of Materials accepted.


91
*Brief Bioinform* 2, 195-7.


Appendix A. DNA sequences of the \textit{fliC} gene deletion variants.

\begin{verbatim}
> pTH890-C5

1 atggccacaagtcattaatacaaaacagccttgctgctgttgacccagaataaactgtgaacaa
MA Q I N T S L L LT Q N L N K
61 tccgcactcctgtgcagcaccctacgctcgcctgctcttcgctgctgtatcacacgc
SQ S A L G T A I E R L S S G L R I N S
121 gcgaagacgatgcgcaggtacagccgtgtactgtaaccgttttacgccgaacataaanagtgt
AK D D A A Q G A I A N R F T A N I K G
181 ctgactcagccctcctgcgaacgcgtacctcctcctcccagctgttaaccgcacatcagacagc
L T Q A S R N A N D G I S I A Q T T E G
241 gcgcgtacagcacaatacacaacgacctgctgacgctcaccggtgcaaatgcccagcgcctg
AL N E I N N L Q R V R E L A V Q S A
301 aacacacaccactcctgctgacactctcatcaccggtgcaaatgcccagcgcctg
N S T N Q S D L D S I Q A E I T Q R L
361 aagtaacagctgacagctgtattacgcagcagccggtgctaacacatcgccggtgcaaatgcccagc
NE I D R V S G Q T Q F N G V K V L A Q
421 gcaacaccctgacacccagctgttacgctgtaaccgcagcagccggtgcaaatgcccagc
D N T L T I Q V G A N D G E T I D I D L
481 aagcagcagcagcactattcctcagccggtgcagctgctgatcagctgtaaccgcagcagc
K Q I N S Q T L G L D T L N V Q Q K Y K
541 gtcagcgatcagcgtcgcattcaagggatgtgctgccgatactacgcggtgtctgcatacacac
V S D T A A T V T G Y A D T T I A L D N
601 aagtaccttaaacgctgctactgttgctgtacggctgacacaggtatctggggtgtcactgttacgc
S T P K A S A T G L G T D Q K I D G D
661 ttataattttgtagatagcactgggaattaactgcacacagctgtttgcttcacgctACTTTG
L K F D D D T T G K Y K A V K T V T G T G T
721 ggttaaagtaggtgctattattagagttctggtgtaagcgacagctgtagtcatctttgct
G K D G Y Y E S V D K T N G E V T L A
781 ggcgcggctgagtttccccctgtcatggttaggtgatgctatgctctgtctccagctgttgccgt
G G A T S P L T G G Q V A N A D L T E A
841 aagtcgcctatgagcacagcagctggtttacgcgcacagctgtttgcttgatagctttctat
K A L T A A G V T G T A S V K M S Y
901 actgataaataacggttaaatatattgtagttggttaactcaggttacagttacgcgctattac
T D N N G K T I D G G L A V K G D D Y
961 tatcttcgcaatactaatagttcgctataattactacagaatatatactcagctacagcg
Y S A T Q N K D G S I S I N T T K Y T A
1021 gatgacgcgtacatccaaacctgcataaaacacctggtgtgccccagccgaacacgccaa
D D G T S K T A L N K L G G A D G K T E
1081 gtctgtttctattgtggttaaaactacctgcgcaagttgaatagccgcaacagttaaaactattcagcttg
V S I G G K T Y A A S K A E G H N F K
1141 gcacacagctgactgcgcgaaacgcgctcactaaccacagcgaacacacccgctgcagaaatt
A Q P D L A A A A A T T T E N P L Q K I
1201 gatgcctgtttgcacacagctttctcactgcttgctgcgctacagcgcgattgtgtaaccgc
D A A L A Q V D T L R S D L G A V Q N R
1261 ttcacccgctattacaacacgcttggccacacacgttaaaacacactgctctcgcctgatgc
F N S A I T N L G N T V N N L T S A R S
1321 cgtagctacagcgtactacgcgacgacagggtaacactcctttcacaagttgcctgcgtgatgc
R I E D S D Y A T E B V S N M S R A Q I L
\end{verbatim}

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
>pTH890-C30

1  atggcacaagtcattaatacacaacaacagcctgtcgtgttgaccacaaatatacctgaacaa
MAQVININTNSLSSLTLTTQNLLNK
121  ggcgaagacgatgccccaggtctgtccacgaggtgtgtactgtcctgtccgtgtacacacgc
SQAALGTAIERLSSGLRINS
241  gcgtctgaacgatatcacaacaacactgccaagcgtgtggtaccgaggtgaacctctggcagagctgtcgctctgtgtctacagcggagcgtgtatctcagtgacgctg
NLNEIINRNQRVRVELAVQSA
361  gacacacactcatcagctgacctcaacctagcaggtgtccacacgacgtgaataatcgattcggcagagct
NEIDRVSINGFTQFGNGVKVLQ
481  gaacacacactcatcagctgacctcaacctagcaggtgtccacacgacgtgaataatcgattcggcagagct
DNTLTIQVGANDGETIDIDL
601  gacacacactcatcagctgacctcaacctagcaggtgtccacacgacgtgaataatcgattcggcagagct
GGATSPVANADLTEAKAALT
841  gacacacactcatcagctgacctcaacctagcaggtgtccacacgacgtgaataatcgattcggcagagct
AAGV TGTVNATSVKYMSTDNNNG
961  gacacacactcatcagctgacctcaacctagcaggtgtccacacgacgtgaataatcgattcggcagagct
KTDGGLAVKVDHYYSATQ
1141  gacacacactcatcagctgacctcaacctagcaggtgtccacacgacgtgaataatcgattcggcagagct
AEAAATTTENPLQKIDAAAL

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>pTH890-C115</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;pTH890-C115</td>
</tr>
<tr>
<td>1 atggcaacagtcatattaataaacaacgccgctgtgtgaccccaagatataacgtgaaacaa</td>
</tr>
<tr>
<td>MAQVINSLSSLTTQNLNPK</td>
</tr>
<tr>
<td>10 tccagctccgctctggccacacgctctcagctctgtcctccgtctcgtatcaacagc</td>
</tr>
<tr>
<td>SQASLGTAIERSLSGLRINS</td>
</tr>
<tr>
<td>15 gcgaagacgtagcgccgaggtcagggtactgtcataacccgtttacgccgaccaatcaaggt</td>
</tr>
<tr>
<td>AKDAAAGQAIAANRFTANIK</td>
</tr>
<tr>
<td>20 ctgacttcagctttcctgtaacgcgtaactctctgctactgcaacagctaggaagggc</td>
</tr>
<tr>
<td>LTQASRNANDGISIAQTTEG</td>
</tr>
<tr>
<td>25 gcgcgtgaaccaaatcaacacaacacactgcagctgtggaactgcttcgtcctgct</td>
</tr>
<tr>
<td>ALNEINNLNRQVRRELAVQSA</td>
</tr>
<tr>
<td>30 acgcacaccacactccgctgcagcagtcctcaaczctggatgaatcaccacgacgcc</td>
</tr>
<tr>
<td>NSTNSQSDLSIDISAEITQR</td>
</tr>
<tr>
<td>35 acgaacaatcagccgactgtatccgacacagctagtcaacaccggttaacgtgagacg</td>
</tr>
<tr>
<td>NEIDRVSQGQTQFNGVQVLAQ</td>
</tr>
<tr>
<td>40 gcacacaccctgacacccataggtgtgtgtaaccacggtgcaacactcgatagctagt</td>
</tr>
<tr>
<td>DNTLTQSVGANDGETIDIDL</td>
</tr>
<tr>
<td>45 aacgcagatcataactcgcagctgggtctgatactgatagctgcaacacaaatataag</td>
</tr>
<tr>
<td>KOINSQTLGLDTLNQVQKYK</td>
</tr>
<tr>
<td>50 gtacacgataaggctacactgttactagatccgataactcgatcttgattagacaaat</td>
</tr>
<tr>
<td>VSDTAAATTVGITYADTTIALDN</td>
</tr>
<tr>
<td>55 agtcacctttaaaagcetctggactgctcttgtgtgtactacctgaccaagaaatggctag</td>
</tr>
<tr>
<td>STFKASATGLGGTDQKIDGD</td>
</tr>
<tr>
<td>60 ttataataattgtatgacactggagaaataataaccccaaaatgtacgggtgac</td>
</tr>
<tr>
<td>LKFPDDTGTKGYAKVKTGVTQTTG</td>
</tr>
<tr>
<td>65 gtatgactcttttatgtatgatgcttttgataacagcaacggtgaggtgactcttgct</td>
</tr>
<tr>
<td>GKDGYKEVSVDKTNSTNGVL</td>
</tr>
<tr>
<td>70 ggccgtgcacctccccgccgcttacggtgactacctgaccaacagctaggtggtggtc</td>
</tr>
<tr>
<td>GATSPLTGGGLPATATEDVA</td>
</tr>
<tr>
<td>75 gcagcgttacccggacagcatctggtttgaaggtgtttctgttactgtataataacggtaa</td>
</tr>
<tr>
<td>AGVTGTASVKNKMYSYTDNNGK</td>
</tr>
<tr>
<td>80 actattggttgctttagccttaaggttggctgtattactactctccaataaat</td>
</tr>
<tr>
<td>TIDGGGLAVKVGDYYSATQN</td>
</tr>
<tr>
<td>85 aaagacttgccataagatattaataactacaaataaacctgcaagttggtacacttcaa</td>
</tr>
<tr>
<td>KDSISINTTKYTADDTGSK</td>
</tr>
<tr>
<td>90 actgcactaacaaccacactgggtgagccgacagaaaccggaaagttggtttcttggtggt</td>
</tr>
<tr>
<td>TALKLQCADGKTEVVSIGG</td>
</tr>
<tr>
<td>95 aaaactacgtgcatgtaaagccgaaggtcacaactttaaaacccagcagctgtatctggcgc</td>
</tr>
<tr>
<td>KTYAASKAEHNFKAQPDLA</td>
</tr>
<tr>
<td>100 gaagcggctgctcaacaccacggagaaccacgctgacgaattgtactgtctttggccacag</td>
</tr>
<tr>
<td>BAAATTENPLQKIDAALAQ</td>
</tr>
<tr>
<td>105 gttgcaacagttaactttggtgtccgtactgatcaacacctcgctgatattaccc</td>
</tr>
<tr>
<td>Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.</td>
</tr>
</tbody>
</table>
>pTH890-C182

1  atggcacaagtcattataataaacacagctgtctgggtgaccagaaaaaataaacattgaacaaca
2  maqvintnsllstlrotqnlnlnk
3  61  tcccaagtctgcttggcaccagctatcgagctcttgcttttccggtctcggtatataacagc
4  sqsalgtaierrlssgrlns
5  121  gcgaagaacagtgacggcaggcaggttgctataacgctttattccogcaacaacaaggt
6  akddaaagqaianrptanikgg
7  181  cttgactcaggctctccctgtaaccgtaaaacaggattcttcatgcttggcagacactcagaaggc
8  ltqasrnandgisaqaottek
9  241  ggcgctgaaacgaatcaacaaacaacactgacgctgtgctgaacttgccgttcatctctgtct
10  aclnearnlnlorolvrelavqsa
11  301  aaacagcacaactccagctcgactcctgactcctacccagcgtgaaatccacccagcgcgtg
12  nstnstsdsldslsqaqaeitqrl
13  361  acaagtaacgacccgtgtacctccgacagactcagttcaccagcggtaagacgcttgcgcag
14  neihrvsgqfngvkvqlaq
15  421  gcaacaccctgacattcatcaggtgctgccaaacagcggtagttaactctgcatttcgct
16  dntltlqvgandgtdildl
17  481  aacagcatcaactctcagacccttggtcttgtgatactgatgcacaaaaaaatatagaq
18  kqinsqtlgldtlnvqkyk
19  541  gtcagcgtacgctgtccacactgtacagattgctccgactacgtgtctgttagacatat
20  vsdtaaatatvtyadttialdn
21  601  agtactttttaaacgcctcgctactgtcttttggtactgcaccaaaatagtggcgcgtat
22  sfkatasatgllggtdqkidg
23  661  ttaaaatcttgatagctacgatactggaaataataacgctgatgataacgttcaggggaaact
24  lkgfddttkgkyyakvtytvgt
25  721  ggttaaaatggtcatattatgaagtttcgtgtagatgataacggtgtagtctttgct
26  gkgdyevsvdktngtevlt
27  781  gcgcggcgtgacctccccgctttaaagcgcctgtgacacagcgcaggtgtacctcgcacagca
28  gatospklaaltaavgtgta
29  841  tctgttgatcagatctcttaactactgataaaccgtaaaatattgtgatttggaca
30  ssvkmsytdnngktidgga
31  901  gttaaagtgcgggatgattactttacttctcccaactaaatatttttcattacatattgctcaggttt
32  vkgddvyysatznkddsisi
33  961  aatactacagaaataacactcagagttccgtacctccacaaatataagattggtttcatataagtt
34  nttktayaddgtksttalnklg
35  1021  ggcgcagaagccaaaaacaggtgattttgtttcattgtggtttgaaatctagtctgcaagtaaa
36  gadgktevvsvissggkgtyaask
37  1081  gcgcggaaggtctacaacttttaaagcagcagcctgtctcgccaggaacgcgtctacaacacc
38  aegnhfkaqfpldaleaaadttt
39  1141  gaaaaaccctgcagaaaaattgtagctgtcctttgcacaggttcagctctgcgtac
40  enplqkidadlaqvdlrds
41  1201  ctggggtgccgtgctgagacacctttcaactcgcgtttctatattccacacctggaaccggtt
42  lgarvqnrnsaitnlngtv
43  1261  aacctgagctctcgccgagtgacaggtagatacgagctcgcagccagttttccac
44  nltarsriedstydatvsn
45  1321  atgcctcgcgctggcattttcgcagacggtgcacacgtctacccgcttcggcgcagggagccag
46  msraqilqqqagtsvlaqan

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
1381 gtttcgcaaaacgtctctctttactgcgttaaa 1413
V P Q N V L S L L R *

> pTH890-C27

1 atggccacaagtcattaataataaaacagctgcgtgtggcaccagaataaacctgaacaa
MA Q V I N T N S L S L L T Q N L N K
61 tcccgatctccgtcggccacgctatcagcgctgttctctccgtctgctgatcaacgc
SQSALGTAIERSLSSGRLNS
121 gcgcgaagagcatgcgcggcgagcaggtgtacacgtttactacgcaacacattgaacggt
AK D D A A G Q A I A N R F T A N I K G
181 ctgactcagcttcttcgcctaaagctaagcaggtatatctccattgcccagaccaactgaaaggc
LTQASNANDGSIIAQTTG
241 gcgcgttgaagactatcaacaccaacacactgcagctgtgcgtgaactggcgttcagtcgtctg
ALNEINNNLQVRERLAVOSA
301 aacagcaacaactc tactccgtctcgaacctaccatccagctgtgaaatcaccacagcgcctgt
NSTNSQSDLDLDSIQAEITRL
361 aacgaaatcgcacctgtatcgcgcagcactatgctcagaagctgtgaaacgtgtggcgcag
NEIDRVSQGTQFNFGVVKLQA
421 gcacaaccccgacccatccagctgtggtgtgcccaacagcgtgtaaatcataatcgatcgatctg
DNTLTIQVGA N D G T I D D L
481 aacgcatataacactctccagacctggtatcgtgatagttgctgaaactgcaacaaatatag
KQINSQTLGLDLTTLNVQQKYK
541 gtcagcgtacgctgcaactgttacaggtatatgcgcataactacagtgtttagtaacat
VSDTAAATVTVYADTTIALDN
601 agtacttttaaaaagcttcgcgtactgtttgtgtctgcagcagaaaaatgatgcgtgcgat
SFKAASATATGGLGGTDKQIGD
661 tttaaaattggtagatcagatcggaaatattacgccaaagttactcccgtaggcgggaact
LKFD T T T G K Y A K V K T V G T
721 ggttaagaggtcattttagataagttcgcgtgataagagacagctgggtcgactcttgctt
GKDGYYEYSVEVDKNTNBVLA
781 ggccgtggcagctctccocgccctacaggttggtactacgtgcacaggctgtgagcgcag
GGATSPLTGLPATA TEDV
841 tcgtgtagtttaagactcttcatattctctgataaaacgcgtaacagguatgatgtggctgcagc
SVVKMSYTDNNGKTIDCLLA
901 gttctagaggccaggtgattactttactctgcacactcacaattaagatgttgctataagttatt
V K V G D D D Y S T A Q N K D G S I S I
961 aaatactcagaaataactcgcagatgacgctatcccacaacacttaaaactactgggt
NTT K Y T A D D D G T S K T A L N K L G
1021 ggccgagaggccggaaaccacgaagttttcattggtgtgttaaagctctctcggacagtataaa
GADGTKEVVTVSISGKTTYAASK
1081 gcgcagagggccacaaatatagccagctgctgtgcagcagcggacggcgctgtcataccccaccc
AE GHNFKQAQPDLAAEAATTT
1141 gaaacaccggtcgcagaaatattgtgcctttggtgccagctgagcaactgtctctgcac
ENPLQKIDAALAAVQVDTLRSR
1201 ctgcggccggttcagacactttatccagcttatctcaaaccctggccaaacacactgggg
LGAVQNRFSNASAITNLNGLNTVN
1261 aacgctgacttctctccgagccggtatccagagagtcgcactctggccagacaccttac
N L T S A R S I E D S D Y A T E V S N
1321 atgtcctgcgccgagattttcgcagccctgtccttggcgcctgccgacaccag
M S R A Q I L Q Q A G T S V L A Q A N Q
1381 gtttcgcaaaacgtctctctttactgcgttaaa 1413

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
>pTH890-C48

1  atgccccacagctattatacaaacacagccctgtcgtgtgtgacccgagaataacactggacaaa
  m a q v i n t n s l s l l l t a q n n l n k
51  tcccgctgctggcggacacgttacgacgctgtgctttccggtgctgatgacagacgq
  s q s a l g t a i e r l s s g l r i n s
101  ggcgaaagcagagcggcgaggccgagttgctaatagcttttaccgcaaacatcgaaggt
  a k d d a a q q a i a n r f t a n i k g
151  ctcgactcgccgcacagctataacgcaggttaatccacgtgacgacagcgcacaggg
  l t q a s r n a n d g i s i a q t t e g
201  gccgctgaaacagaaatcaacaaacagcacgtcagctgtgcattgactgcggctacttgctgct
  a l n e i n n n l q r v r e l a v q s a
251  aacacgacaccaccacctccacgtgctcactccacccagctgaatcaccaccacgcgtg
  n s t n s q s d l d s i q a e i t r l
301  aacgaaatcgacccgtgsatccggcaggactctaggcttacagccgagaatttagtgacacgt
  s t f k a s a t g l g g t d q k i d d
351  ttaaaatggatgtagacgcaggaaaaatttacgacacccgtggaatggctagacgtcga
  l k f d d t t t g k y y a k v t v t g t g t
401  ggataagagctggcttatattgaagtttctgtagaataggacacggagttaggtgactcttgc
  g k d g y v e v s v d k t n g e v t l a
451  ggccgctgcgactcccccccgttcacagcttgactcctgcgacacagccacagcagctct
  g g a t s p l t g g l p a t a t g t a s
501  gtttgttaagagtctttatactgataataacggttaaaactatgtgtggtttacagtt
  v v k m s y t d n g g l a v
551  aaggttagcgcagctactattctcgagactcataacacagcataaagacgtctttacagttat
  a g c c c c g c t g c a g a a a a t t g a t g g c t g
601  acactacgaaatacactcgcagtgacgtagataccaaacagctgcataaacaacagctgggtggc
  t t k y t a d d d g t s k t a l n k l g g
651  gcagcagcagcagaaacctggtttctctatttgtgtgtaaaaacttacgtgcgaagctaagcc
  a d g h n f k a q p d l e a a a a t t k a
701  gaaggttcacacgtaaataacagcagctctgtctccgagacagcagttacagctctgctgct
  n p l k i d a a l a q v d t l r s d l
751  ggtgctggtaacaggaacccttttttaactctctgttctgagcaagcgttacagctgttactgcgctg
  g a v q n r f n s a i t n l g n t v n n
801  ctcgactttccgctgcctccgccctcagagcctgatccggcaccagcagttatcaacccgtgaaacacagc
  l t s a r s i e d s d y a t e v s n m
851  ttcgogccgagattttgtcgacacgctgccgtactccgctgctgcggcagggacaacgtt
  s r a q i l q q a g t s v l a q n q
901  cccgaaacagctctctctttactcgcgtaaa 1410
  p q n v l s l l r *
>pTH890-C33

1  atggcacaagtcattaataacaacagctgctgctgtgaccagaataaacctgaacaa
MA Q V I N T N S L S L L L T Q N N L N K
61  tccacagtggctctggcaccgctatagcagcgtgctttcgggtcgtgatatcaacacgc
SQ S A L G T A I E R L S S G L R I N S
121  ggcaaaagcaagtggcaggtcaggcgttattcaggttaaccggttttacccgaacacattgaaggt
AK D D A A G E Q A I A N R F T A N I K G
181  ctagctcaggtttccgtaaacgcgtatcctccattgcgacagcaccactgaagggct
LT Q A S R N A N D G I S I A Q T T E G
241  ggcgtgaacagataataacacaacacccgtgacgcgtgtgcgaactgcggccattccgtctgt
AL N E I N N N N L Q V R E L A V Q S A
301  aacacgaacacttccagctgacctgagactctccatccaggtgaaatcccacccgcccctg
N S T N S D L D S I Q A E I T R L
361  aacgaatacgaccgtagctatccggccagactcagttcaagcgggtgaagattctggccacag
NE I D R V S G Q T F N G V K V L A Q
421  gcaacaacctcggcaccactcgggttgcgcagactttccgtagaatgcgcagatccgactctg
D N T L T I Q V G A N D G E T I D I D L
481  aacacgtagcactcggccacgtacggtagatgcgcacccgtgtaaataaacatataag
K Q I N S Q T L G L D T L N V Q K Y K
541  gtcgacgtagcagggctcgaacctgtagccgataccgtagatctggttagacacagt
V S D T A A T V T G Y A D T T I A L D N
601  agtacttttaaacagctcggtactgtttggtgtactgaccagaataatgtggtccgat
S T F K A S A T G L G G T D Q K I D G D
661  ttaaaatttgtgatgacacgtgaaataatattaccgggaattgacctgacggttagggaact
L K P D D T T G K Y A X V K T V T G G T T
721  ggttaagatggtcattattgaaagttcctcggtgataagaacgacacctggtagtactctgct
G K D G Y Y E V S V D K T N G B V L A
781  ggccgtgctagcctcccccgttacaggtgactctgccagcaactcaggtatggtctgt
g a a t s p l t g g l p a t a t e d v s
841  gttggttaagatgctttatactccgataataagctgtaaatattgtggtgttagccggt
V K M S Y T D N N G K T I D G G L A V
901  aaggtgacgtagtacattctccgagactaccaaaaaaacatggtctctataagttaaat
K V G D D Y Y S A T Q N K D G S I S I N S
961  actacgaaatatacctgcagatgagcgtcactcccacagcataacccaggtctgtgagc
T T K Y T A D D G T S K T A L N K L G G
1021  gcacgacgaaaaacaggatgtttcttgtatgcgtaaaactactccgctagtcgaagttc
A D G K T E V P S I G G K T Y A A S K A
1081  gaagttctccaaacttttaaacagcagctgctctggcaggcttctcactcagcctgaacaa
E G H N F K A Q P D L A E A A A A T T E
1141  acgcgtcgacgaagaaaattgtactgtctttgacagaccgttacaggtctctgccctg
N P L Q K I D A A L A Q V D T L R S D L
1201  ggcgtgacagaaaccgcccgtctaccgcctctagttacaaccctgcgcacacgccgttaaaacacacgc
GA V Q N R F S A I T N L G N T V N N
1261  ctaggtttgcctcgcctagctgtagttacagggctcagctcagcgcacgagtttcaccacagt
L T S A R S I E D S D Y A T E V S N M
1321  tctcgccgacagttttccgctagcttcctcccgtctggtccgacagggccccaggttt
S R A Q I L Q A G T S V L A Q A N Q V
1381  ccggaaaaacagtccctctttactgctgttaa 1410
P Q N V L S L L R *

108

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
+pTH890-C62

1  atggcacaagctctattaataacaacagccctgtctgttatgccagaataacctgcaaaa
2  MAQVINTNSLSSLTTQTNNLK
3  61  tccceagtccgtcttggaaccgctatcagccctgtcttttccgggtctcgtctacgatcacaagc
4  SQSALGTAIERLSSGLRINS
5  121  gccaagacagtgccggaagtcagggcatgttatcagcaggtttctgccaagacaatcaaggt
6  AKKDAAGQAIAANRFTANIKG
7  181  cttgactcagctctccggtaaacgctatcagcaggtatatttcccagctggcagaccactgagaagc
8  LTQASRNANDGISIAQTTTEG
9  241  gccgtcgaacggaatacacaacaaacctgcagctgtcgaactgcgctgtctgtctctgtctctgtctc
10  A L N E I N N N L Q R V R E L A V S A
11  301  aacagcaccacactcccagctctcagctcatccacaccggtctaaacccagccccctg
12  N S T N S Q S D L D S I Q A E I T Q R L
13  361  aacgaaatccgacccgtgatctccgccacagacctgtaacaggctgaaggtctctggcgcggcag
14  N E I D R V S G Q T F N G V K V L A Q
15  421  gacacaccctcagccatcagctgtgctcggcccacaagcgctgtaacactcgtatctcgtct
16  D N T L T I Q V G A N D G T I D I D L
17  481  aagcagatactacccctcagccgctgtgctcagctgttacgtatgtgcaacaacaaataataag
18  K Q I N S Q T L G L D T L N V Q K Y K
19  541  gtacagcagatacggcctcaacgtcattacagttgtctctcggcagcaaaatgtagggcat
20  S T F K A S A T G L G T D Q K I D G D
21  601  aaatttttttaaagctcctcgtcattctcgttgggtactcgtgacctcgaagactgtggatcgat
22  V S D T A A T V T G Y A D T T I A L D N
23  661  agacttttttaaagctcctcgtcattctcgttgggtactcgtgacctcgaagactgtggatcgat
24  S T F K A S A T G L G T D Q K I D G D
25  721  ttaaaattttgtagatccacgtaactgtaatcagctatactgcaaaatgtgacggtcataactg
26  L K P D D T G T G K Y Y A K V T G T V T G T
27  781  ggttaaagtaggtctattataagttttttctgtatgaaacggactgcggtaggtgactcttctgtc
28  G G A T S P L T G G L P A T A S V K
29  841  ggtacatccttgccagctctcagtccagttgactcctgcagcataaactctggttatagtatataactcagaa
30  D D Y Y S A T Q N K D G S I S I N T T K
31  901  tacactgcagatgcgcgatcttttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
32  Q N R F N S A I T N L G N T V N N L T S
33  961  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
34  G G A T S P L T G G L P A T A S V K
35  1021  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
36  D D Y Y S A T Q N K D G S I S I N T T K
37  1081  tacactgcagatgcgcgatcttttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
38  Q N R F N S A I T N L G N T V N N L T S
39  1141  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
40  G G A T S P L T G G L P A T A S V K
41  1201  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
42  D D Y Y S A T Q N K D G S I S I N T T K
43  1261  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
44  G G A T S P L T G G L P A T A S V K
45  1321  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
46  D D Y Y S A T Q N K D G S I S I N T T K
47  1381  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
48  G G A T S P L T G G L P A T A S V K
49  1441  ggtacatccttttctattttttctgtatgaaacggactgcggtaggtgactcttctgtc
50  D D Y Y S A T Q N K D G S I S I N T T K

> pTH890-C120

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
421 gacaacacccctgaccacctcaggttgtgtgccaaagctgaaatctatgcatactgcgtg
dntltqvgandgetididl
481 aagcgagtaacacttcacgccctgggttgattatatgcgtaagctgtaacaaaataagaq
kqinsqtqlgldtlvqokxyk
541 gtcacgctgataagcttgcaacttgattacgagatctgaacctagcttattagcagat
vsdtaatvtyadttialdn
601 agtatattttaaagctctggctactgcctttgtgagttacgcaaaatattgtagggcag
stfkasatgllggtddqkidgd
661 tttaaaatgtgatagctggataaggaaatataacgccccaggattaccctttgggggaaact
lkfddttttckyyakvttvtgt
721 gtaaaagttggtaatttaagctttcctgattgacacgcaggtgttaaccggcaaca
gkdgyyevsvvaltaagvgt
781 gcaccttgttgaagatgtcttatactgataataacgtttaaaactttgtgtgtttta
asvkvkmsytdnngktidgl
841 gcacgtaaggtagcogcatgactactattctgaactacatataaaagtaggtgcattaa
tavkvgdddyysatatqkdgsis
901 attaatactacgaatataacttgacagtgacatcacaattaatcataaaggctcattatct
inttkytaddgtstktaalk
961 ggtgaggcagacgcgcaaaaaacgagttgattccattatggtaaaaccttactgtgcaagt
ggadgktvevsviikgyaas
1021 aacgccgaaggttcacaacatttaaaagcagccctgatctgacgggaaccgtgtacaaacc
kaeghfkaqvpdflabaaatth
1081 acocggaaaccccctgacgaaaattgtacgcttttgccacagttgacacgttagctctct
tenplqkidaaalaqvdlrs
1141 gacctgggtcgggtagcaaccgctttcaactctgagttattaccacctcgggcaacccgta
dlgavqnrpnsaitnlgnv
1201 aacaccttgactctctgccccgactgtcataaggatttcgaactacgcacggacagtct
nnltsarsriedsdyatevs
1261 aacacatctgctggcgccagattctgacgcagcgtactctcggtgccagggcaac
mnsraaqilcqagtsvlqan
1321 caggttccgcacaaagctctctctttactgctttaa1356
qvqnpnvlsslrr*

>pth890-c88

1 atggcacaagttcatatattaacacaaacagctgctgtgttgaggccacagataaaccctgaacaa
maqvintnssllltqonnlnk
61 tccacgctcgctgggacacgctatcagcgctcttgttcctgcgtctcgatcacaacgc
sqyalgtailerlssglrins
121 gccaaggacagcgccaggttcagctggcattgataacccgttttacccggacactaagaagt
akdddaaqgaianrf坦ikg
181 ctgcctaggcttccgtaaacgctacaacgcaggtactctctcgttgcacacagtgaacgcc
ltqasrmdangisiaqtteg
241 gcggctgaacaaataaccaacaaacacgcgcagcgctgtgtagctgaacgcgttcagctgtc
alneinnnnlqrvrvelavqsa
301 aacgcaccacactcctcgctgaccctgcacctcccaagctgtaataacccgccaacggctc
nstnssdldisiqaeraetqrl
361 aacgtaatgcaccgtgtatccggccagactcgtccaaacgggtgtggagaagtctcgggcag
neidrvsgqtqfngvkvlaq
421 gacaacacccctgaccacctcaggttgtgtgccaaagctgaaatctatgcatactgcgtg
dntltqvgandgetididl
481 aagcgagtaacacttcacgccctgggttgattatatgcgtaagctgtaacaaaataagaq
kqinsqtqlgldtlvqokxyk
541 gtcacgctgataagcttgcaacttgattacgagatctgaacctagcttattagcagat
vsdtaatvtyadttialdn
601 agtatattttaaagctctggctactgcctttgtgagttacgcaaaatattgtagggcag
stfkasatgllggtddqkidgd
661 tttaaaatgtgatagctggataaggaaatataacgccccaggattaccctttgggggaaact
lkfddttttckyyakvttvtgt
721 gtaaaagttggtaatttaagctttcctgattgacacgcaggtgttaaccggcaaca
gkdgyyevsvvaltaagvgt
781 gcaccttgttgaagatgtcttatactgataataacgtttaaaactttgtgtgtttta
asvkvkmsytdnngktidgl
841 gcacgtaaggtagcogcatgactactattctgaactacatataaaagtaggtgcattaa
tavkvgdddyysatatqkdgsis
901 attaatactacgaatataacttgacagtgacatcacaattaatcataaaggctcattatct
inttkytaddgtstktaalk
961 ggtgaggcagacgcgcaaaaaacgagttgattccattatggtaaaaccttactgtgcaagt
ggadgktvevsviikgyaas
1021 aacgccgaaggttcacaacatttaaaagcagccctgatctgacgggaaccgtgtacaaacc
kaeghfkaqvpdflabaaatth
1081 acocggaaaccccctgacgaaaattgtacgcttttgccacagttgacacgttagctctct
tenplqkidaaalaqvdlrs
1141 gacctgggtcgggtagcaaccgctttcaactctgagttattaccacctcgggcaacccgta
dlgavqnrpnsaitnlgnv
1201 aacaccttgactctctgccccgactgtcataaggatttcgaactacgcacggacagtct
nnltsarsriedsdyatevs
1261 aacacatctgctggcgccagattctgacgcagcgtactctcggtgccagggcaac
mnsraaqilcqagtsvlqan
1321 caggttccgcacaaagctctctctttactgctttaa1356
qvqnpnvlsslrr*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ATQNKDGSISINTTKYTADD

901 ggtacatccaaactgcatataaacaactgggtgtggcgcagacggcaaaaccgaagttgtt

GTSKTA LNK LGGADGK T

961 ttcaattgtgtgtaaactacgctgcaatgtaaagcgcaagtctacacatattaaagcaacg

S IG GKTY A ASKAEBGHNFKAQ

1021 cctgatctgccccacgctctgcataccaccaacggaaaaacccgctcgagaaactagct

PDL A E A AAATTENPL Q

1081 gttttgcaacaggtgacacgttgtcgtgctgctggaattcctgccctgtagctatc

AlAQLQAVNRFN

1141 tccgtagattaccaacctgccccacacgctaaacactagtctctgtgaccctgtagctatc

SAITNLGNTVNNLTSARSI

1201 gagattttcagactgacacggaagatatcccaacatctgctcgcggcagattctgcgcag

EDSDYATEVSNMSRAQILQQ

1261 ggcgcgtagctctcggtttgcggcagagctgcttcacgagttactccgcagagctgc

ACTS V LAQANQVPQNVNLSS

1321 cgttaa 1326

R *

> pTH890- C37

1 atggcacaagtcattaataaccaacagctgctgctgtggccagaaaaataacttgaaacaa

MAQVINTSLSL LTTQNLLNK

61 tcgattcgtccgctcggaactgatctagcagcgtctgctctctggtctgagataacaagc

QSALGTAIERSLSGLRINS

121 gcggagaagctggtggaaggcttaggtttgacacagcttattccccgtagacacttaacgc

AKDAAAG QAIAANRTANIKG

181 ctgtctgcaggtctccgctagattcctgctagctttctgatctgagttacacagttagcggc

LTQAS R N A N D G ISIAQT T E

241 gcggctgacagaaattcaaaaaaactctgagctgtctgtctagagattctgcgggtcgcgct

ALNEINNQLRVRRELAVQA

301 aacacgacactccgctctgcaacctctacccacggctgaataactccagccctg

N S T N S QSDLDSD S IQAEITQLR

361 aacacgagatgcttattgcgttcagggccagcagctaggcttgctcggcag

NEIDRVSQTQFNGVKVLAQ

421 gacaacccacttcaggcttaggtgcaccaagcggtgaaactctgatctagctgct

DNTLTVQVGANGEDTIIDL

481 aacacgagatcctactgcagccctctcgcaactctgaagctggcggtaaaaagcgggccacggt

KDINSQ T LG DTLN VQQKYK

541 gttcagctagcgcctgtggaactgatgctgaaattggcacaactttttataaagc

5SVD TAAAT TVGTYADTTIA LDN

601 aagtactttattagctggctactgctctcttgggtttgaactggagaggtgggaacttgaatc

STFKSA T AT GLGGTDQ KIDGD

661 tttaaattttgtagactacgctggggactgttcttgtagctgatatgttattacgaatcag

LKFDDDTT GKY YAK A KVT VTTG

721 ggttaagagtttattagcttctcctgctatatctgtgaaaggtgaactcttctactgataaat

G K DGY EBSVSVV VKMSYTDN

781 aacggtaaaatactgtatggttttagctagtggagttaccttttttctgaca

NGKTD IGG GLAVK VGDYDDYS A

841 accaaaaataataagttggctccagataattatataactctgaaatctcagatcagcggt

TQN KNDGSISINTTKYTADD G

901 acctccaaactgccacttccccacacaaaactgggtgtggcgcagacggcaaaaccgaagttgttct

TSKTA LNK LGGAD G KTEVVS

125

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
1201 cagattctcgacagcggcgttaacctgctttggcggcagcgccgaacggattccgcaaaac
Q I L Q Q A G T S V L A Q A N Q V P Q N
1261 gtctctcttttagtctgtaa 1281
V L S L L R *

>pTH890-C79

1 atggcacaagtcattataacccagcctgtcgtggtgaccagaataacctgaacaaa
M A Q I N S L S L L T Q N N L K
61 tctcagctcgcctgagcaggattacggacgttcgctttggcctggcctgtcgtcgaacgc
S Q S A L G T A I E R L S S G L R I N S
121 gcgaagacagttcgcaggtctgggcagtagttgcttacaccgttttaacccgaacataaaggt
A K D D A A Q I A I N R F T A N I K
181 cttgcacagctctcgccttacgctaagcaggtatatctccatgctgcacagactgagacgc
L T Q A S R N A N D G I S I A O T T E G
241 ggcgtgaacaggaatacaacaaacacaactcgacgtggtcgtgaacctgcgggttcagctgtcgt
A L N E I N N L Q R V R E L A V Q S A
301 aaccagaccaacctcaccagcctgtcattccgacacagctgttaaatcaccagcggcctgt
N S T N S Q S D L D S I Q A E I T Q R L
361 aaccgaaatcgcagggtatagctcgcagctttcaacggtgaaagtctcggcagcag
N E I D R V S G Q T Q F N G V K V L A Q
421 gcaacacacctgacctcaggtgtgcaccaacggaggtaaatctagctaatctgcgtg
D N T L T I Q V G A N D G E T I D I L
481 aacgagatacaacctcgcacctggtggtcgtgatactggtgaagacaaacaaaaataaag
K Q I N S I Q T L G L D T L N V Q Q Y K
541 gtcagcggatacggcctgctacacgctttacaggatatgcgcgataactacgattgttatagcactat
V S D T A A T V T G Y A D T T I A L D N
601 agtaacctttttaaagcctcggctactgtcgtcttggtactgcacgaaatagttgttgccagat
S T F K A S A T G L G G T D Q K I D G D
661 ttaaaattttagtagatcagctggaggaatattacccagaaattacccagatccgtaacgggggaact
L K F D D D T T G K Y Y A K V T V T G G T
721 ggtagaagattggtcattatagaagcccctgccttggaaggtagaggttactattcttcga
G K D G Y Y E V S V U V K V G D D Y S A
781 actcacaataaagatggttccataaagattataactacaggaatatagcagtagcggaggt
A C T C A A A T A T G T A A T T A T A T C G A T G
841 acatccaaaaatcactaaacccattgggtgacgcagccgaaacagaagttgctcttct
T S K T A L N K L G G A D G K T E V V S
901 atgtgggtggaataacactcagtgaagtaacggcagagtccacatttaataaccaacagctttct
T I G G K T Y A A S K A E G H N F K A Q P
961 gatctgagcggaaagcggtctgtcaacaccagcggaaacacctgtcagagattgtgcgcgctct
D L A E A A A A T T E N P L Q K I D A A
1021 ttgcacacggttgcacagttcagtctgctgtcgcgtgcacgcggagttcgctcaacttaccccct
L A Q V D T L R S D L G A V Q N R F N S
1081 gctttacaacgctgggcaacacgtaataacacgacacgtcctgcgtaggtccggtcgtcgttcgcctcgc
A I T N L G N T V N N L T S A R S R I E
1141 gattcgcgtcagcgcggcagcgtttccagctatgctcggcgcgcgttcgctccggtcgtcgtccgcctttg
D S D Y A T E B V S N M S R A Q I L Q A
1201 gttacccctctgcctttggcgcagagcgctctccggttcgcagcttcgtagctcggcgcacggcc
G T S V L A Q A N Q V P Q N V L S L L R
1261 taa 1263

*pTH890-C12
Appendix B. A multiple sequence alignment of all the 39 unique internal deletion variants of FliC.
CHAPTER III

A THEORETICAL MODEL OF AQUIFEX PYROPHILUS FLAGELLIN: IMPLICATIONS FOR ITS THERMOSTABILITY

Abstract

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

133

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Introduction

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

*Salmonella typhimurium* is a bacterial pathogen that also frequently serves as a model organism for the investigation of bacterial flagellar structure and function.

The flagellar fibers in *S. typhimurium* are primarily composed of up to 20,000 copies of a single globular protein called flagellin, which self-assembles via noncovalent forces to form a helical fiber composed of 11 protofilaments. The primary (phase-1) flagellin gene in *S. typhimurium* is designated as *fliC*, a secondary (phase-2) flagellin gene is named *fljB*; other bacterial flagellin genes are sometimes named as *flaA*, *hag* or *flaF*. Flagellin proteins are synthesized in the cytoplasm as soluble monomers and are prevented from self-assembly and protected from proteolytic degradation by binding of the FliS chaperone protein\(^13\). Flagellin is exported by a flagella specific type III secretion system through the central pore of the flagellar fiber in a partially...
unfolded conformation, followed by complete folding and oligomer assembly at the
distal end of the fiber \textit{9, 10; 14; 15}. This terminal assembly process is aided by the
flagellar chaperone cap protein, termed FliD or HAP2 \textit{9, 10, 16, 17; 18, 19}.

\textit{Aquifex pyrophilus} is a motile, hyperthermophilic, microaerophilic,
chemolithoautotrophic, gram-negative, rod-shaped bacterium that was originally
isolated from the hydrothermal system at the Kolbeinsey Ridge north of Iceland \textit{20}. It
grows at temperatures between 67–95 °C, with an optimum temperature of 85 °C.
Although it is an extremophile, \textit{A. pyrophilus} is a eubacterial microbial species, not
an archaeal species and has been placed in the class Aquificae. Like many other
species of motile eubacteria, \textit{A. pyrophilus} is flagellated, with a polytrichous
arrangement of up to eight flagella fibers on the cell surface, as previously described
by Behammer, \textit{et al.} \textit{21}. The diameter of \textit{A. pyrophilus} flagella is 19 nm, which
conforms to the general morphology of the mesophilic bacterial flagella \textit{21}. \textit{A.}
\textit{pyrophilus} flagella are primarily composed of a single flagellin protein that is
encoded by the \textit{flaA} gene \textit{21}. The corresponding FlaA protein has 501 residues (53.9
kDa), with the first N-terminal Met residue removed post-translationally \textit{21}. The
sequence length of FlaA is nearly identical to that of some mesophilic bacteria, such
as the 495 residue \textit{Salmonella typhimurium} FliC protein (51.4 kDa). However, \textit{A.}
\textit{pyrophilus} flagella are unusual in that they are functional, i.e., remain self-assembled
into fibers, at temperatures higher than 100 °C and at low pH values \textit{21}. This
functional temperature range is much higher than that typically observed for the
flagella of well-characterized mesophilic bacteria such as \textit{S. typhimurium} and
*Escherichia coli*, which typically dissociate at temperatures higher than 60-65 °C
22,23, a property often exploited in their purification for biophysical studies19. The *A.
pyrophilus* flagellar fibers are prone to breakage during low speed centrifugation,
unlike the flagella of *S. typhimurium* and thus are more rigid than their mesophilic
counterparts at room temperature21. Henceforth, we refer to the *S. typhimurium*
flagellin protein as FliC and the *A. pyrophilus* flagellin protein as FlaA.

The partial and complete structures of mesophilic *S. typhimurium* FliC were
determined in 200111 and 200324. The complete FliC structure (PDB 1UCU) shows
four distinct globular domains in the protein, termed D0, D1, D2 (with subdomains
D2a and D2b) and D3. Domain D0 forms the inner core of the filament; D1 forms the
outer core and domains D2 and D3 form a knob-like projection on the filament
surface. Various sequence analysis studies have shown that the N- and C-termini of
FliC that form the D0 and D1 domains are essential for the export and assembly into
flagella and are conserved across most bacterial species25. In contrast, the D2 and D3
domains in the middle are highly variable in sequence and size across different
species. Flagellin molecular masses can range from 20-77 kDa26,27,28,29,30,
corresponding to amino acid sequences ranging in size from under 300 amino acids to
almost 700 amino acids (e.g., *Pseudomonas putida* has 688 residues), with the
majority of flagellins containing ~500 amino acids31. These size differences are
primarily due to high genetic variation in a "hypervariable" middle region of the
protein encoding outer domains D2 and D3. These hypervariable domain regions can

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
be deleted to a large extent while not affecting the self assembly properties of the inner D0 and D1 domains.

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

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

**Computational Methods**

The amino acid sequences of \textit{A. pyrophilus} flagellin (FlaA: P46210) and \textit{S. typhimurium} phase-1 flagellin (FliC: P06179) were obtained from the Swiss-Prot\textsuperscript{41, 42, 43, 44, 45} database of the Swiss Institute of Bioinformatics (SIB). The refined crystal structure coordinates of \textit{S. typhimurium} flagellin FliC (PDB file 1UCU) were obtained from the RCBS Protein Data Bank\textsuperscript{46, 47}. The ProtParam\textsuperscript{48, 49} tool at the Expert Protein Analysis System (ExPASy) Proteomics Server\textsuperscript{50} was used to compute various physical and chemical parameters for these two proteins from their 3-D coordinate files and amino-acid sequences. The amino acid composition of the proteins, aliphatic index and grand average of hydropathicity (GRAVY) were of particular interest. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (Ala, Val, Ile and Leu). It may be regarded as a positive factor for the increase of thermostability of globular proteins. The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values\textsuperscript{51} of all
the amino acids, divided by the number of residues in the sequence. We analyzed the overall amino acid flexibility of the proteins using the Average Flexibility Calculator option in the ProtScale tool at the ExPASy Proteomics Server. We also analyzed the propensity of residues in the proteins to be disordered using the DisEMBL protein disorder prediction tool and the GlobPlot protein disorder/order/globularity/domain predictor tool at the ExPASy Proteomics Server.

The homology models for the human *A. pyrophilus* were generated using the computer program MODELLER (6v1) of the Accelrys (San Diego, CA) InsightII software package. The input to the program is an alignment of the target sequence with the related three-dimensional structure of FliC. The InsightII visualization environment was used to analyze the structures and perform other structure related operations.

The Solvent Accessible Surface Area (SASA) was measured using the Solvation module of InsightII. The Delphi module of InsightII was used to investigate the electrostatic charge distribution on the surfaces of the proteins. The protein model was validated using PROCHECK. All Accelrys modeling software was run on an IBM IntelliStation M Pro PC workstation with the Red Hat® Enterprise Linux WS 3 operating system. The multimer model of FlaA was constructed using the coordinates provided by Dr. Koji Yonekura (UCSF, San Francisco) and Dr. Keiichi Namba (Osaka University). The helical parameters are \( l = 66 \, n + 361 \, m \) and the repeat distance is 1698.8 Å. The FlaA monomers were aligned with the FliC subunits.
using the Magic Fit option in Swiss-PdbViewer\textsuperscript{60,61}. We have calculated the molecular surfaces of each subunit to visualize the surface interactions.

The contribution of electrostatic pairing of ionized side chains, i.e., salt bridges/ion pairs, to the stability of FlaA was evaluated using the criteria that a salt bridge is present if two oppositely charged atoms of each neighboring side chain are closer than 6 Å. The residues Arg, Lys, His, Asp and Glu were all considered in the salt bridge calculations\textsuperscript{38}. Another school of thought is that proteins gain electrostatic stabilization by minimizing the number of repulsive contacts between like charged residues rather than by creating salt bridges\textsuperscript{62}. Thus, we have also investigated the occurrence of like charged groups (+, + and -, -) within a 6 Å distance and compared this parameter for both proteins. Protein α-helices have a net dipole moment resulting from the alignment of the peptide backbone hydrogen bonds.

It has been observed that negatively charged (Asp and Glu) and positively charged (Arg, Lys and His) residues are preferentially found at the N- and C-terminal ends of α-helices\textsuperscript{63,64,65,66}. This charge helps to counteract and stabilize the charge dipole of an α-helix, resulting in increased stability of the folded structure. The N-capping box (Ncap) is a local motif that acts as a stop signal and gives stability to the folded protein. A similar Ccap motif was also identified at the C-terminus of α-helices\textsuperscript{67}. The side chain of the residue at N3/C3 is hydrogen bonded to the amide group of Ncap/Ccap and the side chain of Ncap/Ccap is hydrogen bonded to the amide group of N3/C3. Possible N-capping boxes are characterized by the presence of any one of the following residues at both the Ncap and N3 positions: Ser, Thr, Asp,
Asn, His, Glu or Gln. Similarly, C-capping boxes are characterized by the presence of any one of the following residues at both the Ccap and C3 positions: Ser, Thr, Asp, Asn, His, Glu or Gln. The cap regions where the above-mentioned residues are present but no hydrogen bonding was observed were designated as “potential” capping boxes. The FlaA model α-helices were evaluated for possible charge α-helix dipole interactions in terms of the number of Asp and Glu residues at the Ncap, N1, N2 and N3 positions and the number of Arg, Lys and His residues at the C3, C2, C1 and Ccap positions.

Results and Discussion

The properties of thermophilic proteins have been examined extensively over the past two decades. Many structural determinants for thermostable proteins have been postulated, as reviewed by Petsko 68, Vieille and Zeikus 38, Scandurra et al. 69, 70, Fields 71 and Sterner and Liebl 72. Sequence analysis of hyperthermophilic bacterial genomes has detected some preferences of thermophilic proteins for particular amino acids but general design principles are not fully defined 70, 73. Factors that have been reported to increase thermal stability of proteins include tighter internal packing of hydrophobic residues 69, increased structural rigidity 38, 71, increased numbers of hydrophobic residues with branched side chains 74, additional prolines 75, fewer glycines 76, deletion of flexible surface loops, fewer Asn and Gln residues 75, 77, fewer histidine residues 69 and greater numbers of charged residues and consequently, increased hydrogen bonds and salt bridges (ion pairs) on the protein surface, at the
expense of uncharged polar residues \(^6\); \(^2\); \(^9\); \(^7\); \(^4\); \(^7\); \(^8\); \(^7\); \(^9\); \(^8\). Other forms of electrostatic stabilization forces include increased stabilization of \(\alpha\)-helices by intrahelical salt bridges, an increase in negative charge at the N-terminus (helix dipole stabilization) \(^8\), and increased cation-pi interactions \(^8\). Protein oligomerization has also been noted as a factor in increased protein stability \(^8\); \(^1\); \(^8\), in part due to decreased exposed surface area. Confinement of proteins in a small inert volume, i.e., an "Anfinson cage", is also known to stabilize the folded state of globular proteins \(^8\). Flagella, which are directly exposed to the external solvent environment, are a rather unique and ideal system to investigate several aspects of protein thermostability, in the context of a self-assembling, oligomeric structural system, rather than a catalytic enzyme system. We have systematically compared the sequences and structures of FliC and FlaA to elucidate the factors responsible for the thermostability of FlaA.

**Sequence comparison**

Sequence alignment for modeling was done using the pairwise-alignment tool of InsightII. The alignment of the protein shows a 29.1% overall sequence similarity between the thermophilic and mesophilic flagellins, FlaA and FliC (Figure 3.1).

The protein sequences were divided into three separate regions, based on their similarity and structure, for purposes of further sequence analysis: the N-terminus, the middle domain and the C-terminus. The FlaA N-terminus includes amino acid residues 1-180, the FlaA middle domain includes residues 181-398 and the FlaA C-terminus includes residues 399-500. The FliC protein was similarly
Figure 3.1 A pairwise sequence alignment of FlcC and FlaA. The overall sequence identity of the alignment was 29.1%. The sequence identity at the N- and C-termini was much greater (~55% identity) than the middle domain (16% identity).

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
divided into an N-terminal region of residues 1-175, a middle domain composed of residues 176-394 and a C-terminal region comprising residues 395-494. Both the N- and C-terminal regions have a higher sequence identity that the overall proteins; the FlaA and FliC N-termini have 54% similarity whereas the corresponding C-termini have 55% similarity. Conversely, the middle domain hypervariable regions have only 16% similarity. This is in contrast to an earlier estimate of 27% similarity for 246 residues, starting at residue 173, by Behammer, et al. 21.

The highly conserved N- and C-terminal regions are essential for both recognition by the type III export apparatus and subsequent self-assembly into the flagellar fiber. Sequence analysis studies have shown that the N- and C-termini of flagellins are conserved across most bacterial species 25. This sequence conservation also confers functional conservation; flagellins from different mesophilic bacterial species can be recombined into heterogeneous flagella in vitro 8, 84, 85, 86, due to complimentary interactions between their highly conserved N- and C-terminal domains. Thus, it is not surprising that a relatively high degree of similarity (~55%) is observed for the N- and C-terminal domains of FlaA and FliC. However, it remains to be experimentally determined if Aquifex flagellin will interact in a functional manner with mesophilic flagellins, mesophilic FliS chaperone proteins and can be expressed and self-assembled into flagella in mesophilic E. coli or Salmonella bacteria.

The original investigation of A. pyrophilus flagella and FlaA sequence by Behammer et al. in 1995 21 suggested some possible sequence and structural determinants for its thermostability. These authors noted increased levels of
Table 3.1 Amino acid composition (percentages) of the full length FliC and FlaA.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>FlaA</th>
<th>FliC</th>
<th>FlaA - FliC</th>
<th>FlaA Middle domain</th>
<th>FliC Middle domain</th>
<th>FlaA - FliC Middle domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>10.8%</td>
<td>12.3%</td>
<td>-1.5%</td>
<td>7.3%</td>
<td>12.8%</td>
<td>-5.5%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>4.0%</td>
<td>2.8%</td>
<td>1.2%</td>
<td>2.3%</td>
<td>0.0%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>6.8%</td>
<td>8.5%</td>
<td>-1.7%</td>
<td>6.4%</td>
<td>4.6%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>7.4%</td>
<td>7.5%</td>
<td>-0.1%</td>
<td>7.3%</td>
<td>9.1%</td>
<td>-1.8%</td>
</tr>
<tr>
<td>Cys (C)</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>5.0%</td>
<td>6.5%</td>
<td>-1.5%</td>
<td>1.8%</td>
<td>2.3%</td>
<td>-0.5%</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>3.8%</td>
<td>3.4%</td>
<td>0.4%</td>
<td>3.2%</td>
<td>2.7%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>7.4%</td>
<td>8.7%</td>
<td>-1.3%</td>
<td>10.1%</td>
<td>13.2%</td>
<td>-3.1%</td>
</tr>
<tr>
<td>His (H)</td>
<td>0.4%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>9.6%</td>
<td>5.1%</td>
<td>4.5%</td>
<td>9.6%</td>
<td>2.7%</td>
<td>6.9%</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>9.0%</td>
<td>8.5%</td>
<td>0.5%</td>
<td>5.5%</td>
<td>5.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>4.6%</td>
<td>5.7%</td>
<td>-1.1%</td>
<td>4.1%</td>
<td>10.0%</td>
<td>-5.9%</td>
</tr>
<tr>
<td>Met (M)</td>
<td>1.6%</td>
<td>0.4%</td>
<td>1.2%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>2.0%</td>
<td>1.2%</td>
<td>0.8%</td>
<td>2.8%</td>
<td>1.4%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>1.8%</td>
<td>1.0%</td>
<td>0.8%</td>
<td>3.2%</td>
<td>1.4%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>6.8%</td>
<td>7.7%</td>
<td>-0.9%</td>
<td>9.2%</td>
<td>5.9%</td>
<td>3.3%</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>9.6%</td>
<td>11.5%</td>
<td>-1.9%</td>
<td>14.7%</td>
<td>15.1%</td>
<td>-0.4%</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.4%</td>
<td>0.9%</td>
<td>0.0%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 3.1 – Continued

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>FlaA</th>
<th>FliC</th>
<th>FlaA - FliC</th>
<th>FlaA Middle domain</th>
<th>FliC Middle domain</th>
<th>FlaA - FliC Middle domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr (Y)</td>
<td>3.6%</td>
<td>2.4%</td>
<td>1.2%</td>
<td>2.8%</td>
<td>5.0%</td>
<td>-2.2%</td>
</tr>
<tr>
<td>Val (V)</td>
<td>5.6%</td>
<td>6.5%</td>
<td>0.9%</td>
<td>7.8%</td>
<td>7.8%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

hydrophobic residues, including aromatic residues and proline in several heat-stable flagellins, with a respective decrease in polar hydrophilic residues. This increase in average hydrophobicity is frequently accompanied by a decrease in average chain flexibility at lower temperatures\(^{87,88}\), as possibly indicated by fragility of the thermostable FlaA flagellar filaments at room temperature. They also proposed that thermostable flagellins form compact monomer structures and large interfaces between subunits (excluding water molecules) in the helical polymer. Comparison of the amino acid compositions using the ProtParam tool was performed for the complete FlaA and FliC polypeptides, their more conserved N-termini and C-termini and their hypervariable middle domains. Residues were classified as hydrophobic: \{Ala(A), Val(V), Ile(I), Leu(L), Met(M), Phe(F), Tyr(Y), Trp(W) and Pro(P)\}; polar: \{Asn(N), Gln(Q), Ser(S), Thr(T) and Cys(C)\}; charged: \{Arg (R) Lys(K), His(H), Asp(D) and Glu(E)\}; and aromatic: \{(Phe(F), Trp(W) and Tyr(Y)\}.  

A comparison of the FlaA and FliC residue proportions (percentages) for the total proteins show a +7.0% increase in the total hydrophobic residues and a
decrease of -6.0% and a +0.6% increase in the polar and charged residues in FlaA (Table 3.1). The FlaA aromatic residues increased by 2.4%. According to one analysis, more charged residues (+3.2%), fewer polar residues (-5.0%) and slightly more hydrophobic residues are found in hyperthermophilic proteins, on average. FlaA seems to be an exception to this rule as there is only a minor increase of +0.6% in the number of charged residues and a very high increase of +7.0% in the hydrophobic residues. The greatest contribution to the increase of hydrophobic residues in FlaA is from the branched hydrophobic residue, isoleucine (+4.5%).

A number of individual amino acid compositions were analyzed in further detail. Isoleucine has greater thermostability at temperatures above 100 °C. The relative percentages of Ala, Asn, Asp, Gln, Lys, Ser, Thr and Val decreased in FlaA while those of Arg, Glu, His, Met and Pro increased, which conforms to the general trend observed among hyperthermophilic proteins. Several properties of Arg residues suggest that they are better suited to functioning at higher temperatures than Lys residues. The Arg δ-guanido moiety has a reduced chemical reactivity due to its high pKₐ and significant resonance stabilization. Arg also forms more stable ion pair interactions at high temperatures. The Arg/Lys ratio in FlaA is 0.9 and in FliC is 0.5. Hyperthermophilic proteins can have an increase in the total number of Pro residues in loop regions; Pro decreases the conformational entropy by conferring rigidity to loops. Thermophilic proteins can also have a decreased proportion of Gly residues because Gly has an opposite effect of Pro in terms of increasing conformational entropy of the polypeptide. There are 9 Pro residues in FlaA and 5...
Pro residues in FliC, an increase of +0.8%. Most of the Pro residues in FlaA are located in the loop regions of the D2 and D3 domains except for residue Pro106, which is located in the α-helix of the D1 domain. There is also a -1.3% decrease in the Gly residue percentage in FlaA. The observed increase of Pro residues and decrease of Gly residues help to explain the thermostability and rigidity of the protein. Apart from the Gly and Pro composition, the flexibility of a protein also depends on its primary structure. By using the ProtScale tool of ExPASy, we have calculated the overall amino acid flexibilities of FliC and FlaA and for their N- and C-termini and middle domains. The entire protein, N-terminus, C-terminus and middle region (D2 and D3 domains) have average flexibility values of 0.447, 0.446, 0.438 and 0.452 for FlaA and 0.450, 0.455, 0.444 and 0.450 for FliC. These values indicate a marginally higher relative flexibility for the entire FliC protein (+0.67%), N-terminal and middle domains (+2.0% and +1.4%) and a slightly lower flexibility for the C-terminal domain (-0.44%). DisEMBL predicted that FliC has 6% more residues than FlaA that may be disordered, whereas GlobPlot predicted a 3% increase in the disordered residues in FliC. Thus, all three algorithms yielded a consistent prediction of a slightly higher flexibility for mesophilic FliC than for thermostable FlaA, although the differences were statistically minor. Both FlaA and FliC proteins lack Cys; Cys is not observed in any wild-type flagellin sequences identified, to date.

The overall aliphatic indexes of FlaA and FliC proteins with their N-terminal Met residues removed are 99.6 and 84.0 (Table 3.2). These aliphatic index values indicate a higher overall aliphatic character for FlaA than FliC, a common sequence
feature of thermostable proteins. The GRAVY values are -0.121 for FlaA and -0.400 for FliC (Table 3.2), which strongly indicates a much higher hydrophobic character for the FlaA sequence. Overall, both FlaA and FliC have very similar numbers of charged residues (Table 3.2), with 56 and 54 negative residues and 43 and 42 positive residues, respectively. However, the distribution of charged residues was very different in these two proteins, as discussed below.

The sequence compositions of FlaA and FliC were further analyzed at three different regions, as discussed above. The aliphatic indexes of the N-terminal regions of FlaA and FliC are 105.4 and 99.3 and the corresponding GRAVY values are -0.266 and -0.428 (Table 3.2). The FlaA N-terminus has an increase of +9.7% in hydrophobic residues, a decrease of -0.4% in charged residues and a decrease of -10.2% in polar residues (Table 3.3). The increase in charged residues was only due to an increase of seven additional positive residues (20 vs. 13, Table 2), as the numbers of negative residues were similar for both
Table 3.2 Computed parameters of FlaA and FliC flagellin proteins and their N-terminal, C-terminal and middle hypervariable domain regions.

<table>
<thead>
<tr>
<th>Protein / Domain</th>
<th>Aliphatic index</th>
<th>GRAVY valu</th>
<th>pl</th>
<th>Negative residues</th>
<th>Positive residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlaA Protein</td>
<td>99.6</td>
<td>-0.121</td>
<td>4.82</td>
<td>56</td>
<td>43</td>
</tr>
<tr>
<td>FliC Protein</td>
<td>84.0</td>
<td>-0.400</td>
<td>4.79</td>
<td>54</td>
<td>42</td>
</tr>
<tr>
<td>FlaA N-terminus</td>
<td>105.4</td>
<td>-0.266</td>
<td>8.12</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>FliC N-terminus</td>
<td>99.3</td>
<td>-0.428</td>
<td>4.65</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>FlaA C-terminus</td>
<td>111.9</td>
<td>0.002</td>
<td>4.50</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>FliC C-terminus</td>
<td>97.7</td>
<td>-0.269</td>
<td>4.61</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>FlaA middle domain</td>
<td>89.0</td>
<td>-0.061</td>
<td>4.46</td>
<td>23</td>
<td>14</td>
</tr>
<tr>
<td>FliC middle domain</td>
<td>65.6</td>
<td>-0.437</td>
<td>5.04</td>
<td>26</td>
<td>22</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 3.3 Amino acid composition (percentages) of N-terminus, C-terminus and middle domains of FliC and FlaA.

<table>
<thead>
<tr>
<th>Protein / Domain</th>
<th>Hydrophobic residues (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Polar residues (%)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Charged residues (%)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlaA N-terminus&lt;sup&gt;d&lt;/sup&gt;</td>
<td>45.2%</td>
<td>30.7%</td>
<td>17.4%</td>
</tr>
<tr>
<td>FliC N-terminus&lt;sup&gt;d&lt;/sup&gt;</td>
<td>35.5%</td>
<td>40.9%</td>
<td>17.8%</td>
</tr>
<tr>
<td>FlaA C-terminus&lt;sup&gt;e&lt;/sup&gt;</td>
<td>50.6%</td>
<td>24.3%</td>
<td>22.4%</td>
</tr>
<tr>
<td>FliC C-terminus&lt;sup&gt;e&lt;/sup&gt;</td>
<td>43.0%</td>
<td>37.0%</td>
<td>17.0%</td>
</tr>
<tr>
<td>FlaA middle domain&lt;sup&gt;f&lt;/sup&gt;</td>
<td>40.4%</td>
<td>32.1%</td>
<td>17.4%</td>
</tr>
<tr>
<td>FliC middle domain&lt;sup&gt;f&lt;/sup&gt;</td>
<td>36.6%</td>
<td>27.9%</td>
<td>22.3%</td>
</tr>
</tbody>
</table>

<sup>a</sup> The hydrophobic amino acids include Ala(A), Val(V), Ile(I), Leu(L), Met(M), Phe(F), Tyr(Y), Trp(W) and Pro(P).

<sup>b</sup> The polar residues include Asn(N), Gln(Q), Ser(S), Thr(T) and Cys(C).

<sup>c</sup> The charged residues include Arg(R), Lys(K), His(H), Asp(D) and Glu(E).

<sup>d</sup> The N-terminal domain regions were defined as residues 1-180 out of 500 total amino acid residues for FlaA and residues 1-175 out of 494 for FliC.

<sup>e</sup> The C-terminal domain regions were defined as residues 399-500 out of 500 for FlaA and residues 395-494 out of 494 for FliC.

<sup>f</sup> The FlaA hypervariable middle region encompassing domains D2 and D3 is defined as residues 181-398 out of 500 total residues. The FliC hyper-variable middle region is...
The hypervariable middle domains of FlaA and FliC have aliphatic index values of 89.0 and 65.6, while the corresponding GRAVY values are -0.061 and -0.437 (Table 3.2). The individual amino acid compositions for the middle domains of proteins, 19 vs. 18. The aliphatic index values of the FlaA and FliC C-terminal regions are 111.9 and 97.7 and the corresponding GRAVY values are -0.002 and -0.269 (Table 3.2). The FlaA C-terminal region has an increase of +7.6% in hydrophobic residues and a +5.4% increase in charged residues, relative to FliC (Table 3.3). The polar residues correspondingly decreased by -12.7% in FlaA. In contrast with the N-terminal region, 40% more negative residues were found in the C-terminal region of FlaA (14 vs. 10, Table 3.2), although two additional positive residues were also observed (9 vs. 7). The extensive increase in hydrophobic residues in the C- and N-terminal regions of FlaA leads to increased hydrophobicity and should contribute to its higher thermal stability. Thus, in spite of high sequence similarity between the FliC and FlaA N- and C-terminal regions, significant differences are apparent in the proportions of hydrophobic and polar amino acids. The structural implications of these differences are discussed in the next section.

Table 3.3 – Continued

| Residues | Defined as residues 176-394 out of 494 total residues. |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 3.2 A comparison of 3-D structures of FliC and FlaA. 2a) Model 3-D structure of thermostable *A. pyrophilus* FlaA flagellin. 2b) 3-D structure of mesophilic *S. typhimurium* FliC flagellin. α-helices are shown in red, β-sheets are shown in yellow, turns are shown in blue and coils are shown in green. Both the structures show domains D0, D1, D2 and D3. 2c, 2d) Stereo view of surface charge distribution of FlaA. 2e, 2f) Stereo view of surface charge distribution of FliC. The legend shows the scale for coloring of the charge spectrum. Blue represents positive charge, red represents negative charge and white represents neutral.
FlaA and FliC are given in Table 1. The middle domain of FlaA has an increase of 3.8% in the hydrophobic side chains and a 4.9% decrease in the charged residues. There is an increase of 4.2% in the polar residue composition (Table 3.3). The FlaA middle domain has significantly fewer positively charged residues (14 vs. 22) and slightly fewer negatively charged residues (23 vs. 26). This indicates that this domain region is more acidic in FlaA and that it is not stabilized by formation of additional salt bridges, but rather by hydrophobic interactions.

**Structure comparison**

The 3-D model for the FlaA protein was built using the coordinate file 1UCU from the Protein Data Bank as a structural basis using the MODELLER software. Altogether, 10 models were generated with a high level of loop optimization and structurally aligned with FliC. The best-fit model had the lowest probability density function violations; its free energy and loops were further refined. The secondary structures were visualized using the Kabsch-Sander method in InsightII (Fig. 3.2a and b). The FlaA secondary structure comprises 9 α-helices and 20 β-sheets similar to the FliC structure. The three dimensional model was verified using PROCHECK. The coordinates of this model have been deposited in the RCSB PDB theoretical model online database, under the ID 1XGX. A Ramachandran plot (Fig. 3.3) shows that 87.2% of the total residues of the FlaA model occupy the most favored regions, 9.7% occupy additional allowed regions, 1.3% occupy generously allowed regions and only 1.1% fall in the disallowed region. Figure 4 shows the 3-D structure alignment of FlaA model and the experimentally determined FliC structure, PDB.
1UCU. The average root-mean-square (rms) deviation at C\textsuperscript{α} positions between the FlaA model and PDB 1UCU is 1.9 Å.

Figure 3.3 Ramachandran plot of the FlaA model. 87.2\% of the FlaA model residues occupy the most favored regions, 9.7\% occupy additional allowed regions, 1.3\% occupy generously allowed regions and only 1.1\% are in the disallowed region.
Figure 3.4 A 3-D structural alignment of FliC with the model of FlaA. The FliC structure is represented as a red chain and the FlaA model as a blue chain. The average root-mean-square (rms) deviation at C\(^\circ\) positions between the FlaA model and the FliC structure (PDB 1UCU) is 1.9 Å.
The above observations indicate that the modeled FlaA structure is conformationally correct. However, it should be noted that the probability of the middle domain region having an alternate conformation is higher, due to the high variability of sequences observed in this region for bacterial flagellins. We have also modeled the structure using the SWISS-MODEL server\textsuperscript{90} and Rosetta algorithms\textsuperscript{91} implemented on the Robetta comparative modeling server\textsuperscript{93}. SWISS-MODEL failed to build a comparative model, whereas Robetta gave a 3-D model (data not shown) comparable to the one generated by MODELLER. The gross topologies of the FlaA model and the FliC structure are very similar, as expected for a homology model. The \(\alpha\)-helical coiled-coil conformations at the N- and C-termini of the FlaA model were conserved as these regions form the intricate intersubunit interactions in the flagellar filament structure.

The C-terminal region of the DO domain, which forms the inner surface of the channel, consists of mainly polar amino acids, as in the case of the mesophilic homolog, FliC. The polar nature of the surface is thought to help the diffusion of the unfolded monomers, because unfolded proteins with hydrophobic side chains exposed will be trapped on a hydrophobic surface\textsuperscript{24}. It is interesting to consider that the C-terminus of FlaA that forms the inner core of the flagellum channel has a higher aliphatic index, 111.9 (Table 3.2), which may retard monomer movement in the channel. A conserved C-terminal Arg residue is present in both proteins, Arg500 in FlaA and Arg494 in FliC. This residue projects into the flagellar central channel.
formed by FliC oligomers and was previously hypothesized to have a functional role in transport of the unfolded flagellin monomers through the channel.  

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

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

In contrast with the highly conserved N- and C-terminal regions, the middle D2 and D3 domains of FlaA have relatively low homology with the same region in FliC. Given that the numbers of amino acids in each protein are similar and minimally acceptable sequence homology is apparent, the FlaA model was built with secondary structures and domain folds for the hypervariable D2 and D3 middle domains that are similar to the FliC structure (Fig 3.2a). These conserved secondary structures include the three beta-folium secondary structures observed for residues 220–260 in FliC domain D3; residues 308–345 in FliC domain D2a and residues 345–383 in FliC domain D2b. The FlaA model shows beta-folium structures at the same residues in the D2a, D2b and D3 domains.

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

As discussed by Honda, et al., the secondary structural elements in flagellin are well segregated radially, with inner domains that are rich in α-helices and outer domains that are rich in β-strands. Given the extremely harsh extracellular environment in which Aquifex flagella function, e.g., high temperature and low pH, it is not unreasonable to assume that the same secondary structural features that stabilize mesophilic FliC against proteases and denaturing chemicals could also be adapted for increased thermal stability in FlaA. Thus, it is also not unreasonable for the FlaA outer D2 and D3 domains to be largely composed of β-strands, as observed in the analogous FliC protein. These structures can be relatively rigid because of cooperative hydrogen-bonding networks. When combined with increased residue hydrophobicity and resulting tighter packing of the hydrophobic core regions, they
Figure 3.5 Alignment of the secondary structures of FliC and FlaA. The α–helices are represented as red cylinders, β–strands as yellow arrows and random coils as green lines.
can also be more stable against proteolysis due to minimal size β-turn loop regions that do not present themselves as flexible substrates for proteases.

Figure 3.5 shows an alignment of the secondary structures of the FlaA model and the FliC structure. The above observations indicate that the modeled FlaA structure is conformationally correct. However, it should be noted that the probability of the middle domain region having an alternate conformation is higher, due to the high variability of sequences observed in this region for bacterial flagellins. A separate analysis of the secondary structure of FlaA indicates that the secondary structures in the FlaA model, including those in the hypervariable middle domain, are largely consistent with standard secondary structure prediction algorithms (e.g., PSIPRED).

Thus, in addition to having consistent structural similarity with FliC, the FlaA structure does not appear to violate any of the secondary structure predictions that were independently derived by standard prediction algorithms. However, the relatively low degree of sequence homology (16%) observed for the FlaA and FliC hypervariable middle domains suggests that alternative conformations are possible for the D2 and D3 domains in the FlaA model. The contribution of this middle domain to the thermostability of the flagellar fiber remains to be investigated; this region also has very low sequence homology with that of thermostable *Aquifex aeolicus* flagellin.
Table 3.4 Solvent accessible surface area (SASA) calculations for conserved D0 and D1 domains of FlaA and FliC

<table>
<thead>
<tr>
<th>Protein Domain</th>
<th>Total SASA $^b$ (Å$^2$)</th>
<th>Polar SASA $^c$ (Å$^2$)</th>
<th>Non-polar SASA $^d$ (Å$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FlaA D0 and D1 $^e$</td>
<td>1,917</td>
<td>638</td>
<td>1,279</td>
</tr>
<tr>
<td>FliC D0 and D1 $^f$</td>
<td>1,673</td>
<td>718</td>
<td>955</td>
</tr>
<tr>
<td>FlaA - FliC $^g$</td>
<td>244</td>
<td>-80</td>
<td>324</td>
</tr>
</tbody>
</table>

$^a$ All SASA calculations were performed using the Solvation module of InsightII software (Accelrys).

$^b$ Total computed solvent accessible surface area for conserved D0 and D1 domains.

$^c$ Computed solvent accessible surface area for polar surface regions of D0 and D1 domains.

$^d$ Computed solvent accessible surface area for non-polar surface regions of D0 and D1 domains.

$^e$ The D0 and D1 domains of FlaA include the more conserved N-terminal residues 1-180 and C-terminal residues 399-500 out of 500 total amino acids.

$^f$ The D0 and D1 domains of FliC include the more conserved N-terminal residues 1-175 and C-terminal residues 395-494 out of 500 total amino acids.

$^g$ Difference between the corresponding SASA values of D0 and D1 domains of FlaA and FliC.
The hydrophobic effect is considered to be one of the most important of the various molecular forces that determine the tertiary structure of proteins and thus their stability and the strength of this effect increases with increasing temperature. Thus, the greater the magnitude of the hydrophobic effect, the more stable the protein. An inverse correlation has been observed between flexibility of thermostable proteins and their hydrophobicity; they may be significantly flexible at elevated temperatures, allowing proper function, but may be too rigid, i.e. "solid" or "wax-like", to function at typical mesophilic temperatures (10-45 °C). The increased rigidity previously noted for FlaA flagella filaments at mesophilic temperatures is likely a result of its increased hydrophobic character and resulting increased compactness.

Summation of the non-polar solvent accessible surface areas (SASA) of a folded chain yields a measure of the potential hydrophobic effect. The SASA values were computed using the Solvation module of InsightII. SASA values were calculated only for more conserved D0 and D1 domains that include N- and C-terminal regions as the lower sequence homology in the middle region may not give an accurate measurement. The total SASA of the FliC D0 and D1 domains that include N- and C-termini is 1,673 Å², which includes a polar surface area of 718 Å² and a non-polar surface area of 955 Å². The FlaA monomer model D0 and D1 domains have a total SASA of 1,917 Å² that includes a polar surface area of 638 Å² and a non-polar surface area of 1,278 Å². Thus, the predicted total surface area of the more conserved regions of FlaA is 15% greater than that of FliC, suggesting that more interfacial area is available for interaction between flagellin subunits in the...
thermostable helical flagella fiber. There is also a decrease of -11% in the polar
surface area and a corresponding large increase of +34% in the non-polar surface
area. This is a strong indicator that increased hydrophobic interactions can form
between FlaA monomers, resulting in a greater hydrophobic effect upon burial from
the aqueous solvent and thus leading to a tighter packing of the protein.

A surface charge spectrum was constructed for each protein using a DelPhi
grid to visualize the relative charge distributions on the surface of the proteins (Fig.
3.2c, d). The D0 and D1 domains of the FlaA model have a higher positive charge
distribution on the surface. This is due to the fact that the number of negatively
charged residues in the N-terminal D0 and D1 domains is the same for both FlaA and
FliC, whereas FlaA has 6 more positively charged residues compared to FliC in these
regions. The C-terminal D0 and D1 domains of FlaA have 13 negatively charged and
9 positively charged residues while FliC has 8 negatively charged and 7 positively
charged residues. The D0 and D1 α-helical coiled-coil domains interact with other
subunits and are thought to form inter-subunit interactions and stabilize the fiber.\textsuperscript{24}
The increase of positively charged residues in the D0 and D1 domains of FlaA could
have significance in terms of the increasing the stability of the fiber by forming inter-
subunit salt-bridges.

Yonekura, et al.\textsuperscript{24} have shown that most of the inter-subunit interactions found
in the outer tube of \textit{S. typhimurium} flagellum filament are polar–polar or charge–
polar and contributions of hydrophobic interactions are relatively small, whereas
those found within the inner tube and between the inner and outer tubes are mostly

165

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
hydrophobic, contributing to the high stability of the filament structure. A pentamer of the FlaA subunit (Figure 3.6) was constructed based on the FliC heptamer coordinates provided by the authors of the complete FliC structure, Dr. Koji Yonekura (UCSF, San Francisco) and Dr. Keiichi Namba (Osaka University).

The multimer was visualized in Swiss-PdbViewer. The subunits were numbered according to the scheme shown in the Figure 5a. The subunit numbered 1 and rendered in blue ribbon is used as a reference for analyzing the intersubunit interactions. Subunits 2, 3, 4 and 5 form the most extensive interactions along each direction of the major helical array with subunit 1 and are rendered in different ribbon colors. The lateral interactions of subunit 1 with subunits 2 and 3 can be seen in Fig 3.5b. As we have noted earlier, there is an increase of hydrophobic and charged residues at the expense of polar residues in the N- and C-terminal regions of FlaA. The predominant lateral interactions between the D0 and D1 domains of subunits 1, 2 and 3 appear to be hydrophobic in nature. Many hydrophobic residues are exposed on the surface of D0 domain and to a lesser extent on D1 domain. Figure 5c shows the interactions of subunit 1 with subunits 4 and 5 that lie immediately on top and bottom of the subunit in the protofilament.
Figure 3.6 Structure of a flagellin pentamer. a) Stereo diagram of 5 FlaA subunits, viewed from inside of the flagellar filament. The subunit numbered 1 and rendered as a blue ribbon structure is used as a reference for analyzing the intersubunit interactions. Subunits 2, 3, 4 and 5 form the most extensive interactions with subunit 1 in the filament and are rendered in different ribbon colors. b) Interaction of subunit 1 with subunits 2 and 3 that form immediate neighbors in the filament. The molecular surface of each subunit is shown. Blue represents positive potential, red; negative and white; zero charge potential. c) Interaction of subunit 1 with subunits 4 and 5 that lies immediately above and below subunit 1 in the flagellar protofilament. These images were generated with Swiss-PdbViewer.
interactions of subunit 1 with subunits 4 and 5 that lie immediately on top and bottom of the subunit in the protofilament. In this case both subunits show a mixture of hydrophobic and polar-charged interactions. Our prediction is that the FlaA subunits in the filaments are tightly packed due to stronger hydrophobic interactions around the D0 and D1 domains that make them rigid at room temperature and stable at higher temperatures.

Salt bridges/ion pairs were defined using the distance criteria that a salt bridge exists if two oppositely charged atoms of adjacent side chains are located closer than 6 Å. Only the amino acids Arg, Lys, His, Asp and Glu were considered for salt bridge calculations. Based on these criteria, FlaA has a total of 15 salt bridges, of which 12 are located in the α-helical regions of the D0 and D1 domains. In contrast, FliC has 24 salt bridges, of which 9 are in the α-helical regions. These results suggest that intramolecular ion pairs do not play an important role in the increased thermostability of the FlaA monomer, relative to mesophilic FliC, in agreement with the previous results of the SASA calculations of polar surface area. However, these results do not necessarily indicate the importance of intermolecular ion pairs between FlaA monomers in stabilizing FlaA in the polymeric flagellar fiber, as noted above, the D0 and D1 domains have a positively charged region. We have also visually inspected the presence of like charged groups within the 6 Å cutoff distance. FliC has 3 potentially destabilizing interactions involving like charged groups where FlaA has 10 such interactions. This result does not agree with the previously stated theory that
minimizing the number of repulsive contacts between like charged residues will increase the thermal stability of proteins.\cite{62}

We visually inspected the charge dipoles of FliC and FlaA \(\alpha\)-helices to determine if any of them are further stabilized by the presence of negatively charged (Asp and Glu) and positively charged (Arg, Lys and His) residues at the N- and C-terminal ends. Both FliC and FlaA \(\alpha\)-helices did not show any preferential occurrence of these charged residues. FliC has four potential N-capping motifs and four actual C-capping boxes. FlaA has five potential N-capping boxes and two actual C-capping boxes. Thus, there is little apparent difference in the contribution of helix dipole stabilization factors in FliC and FlaA.

As previously noted, protein oligomerization is another method for increasing the thermostability of proteins and flagellin functions \textit{in vivo} by forming extended “20,000-mer” helical fibers. It has only recently been observed that the terminal FliD chaperonin cap complex contributes to the thermal stability of \textit{S. typhimurium} flagellar filaments\cite{97}; thus this additional oligomerization factor represents an additional potential stabilizing factor in thermostable \textit{Aquifex sp.} flagella that may not be directly encoded in the FlaA protein itself.

\section*{Conclusions}

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

References


45. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., 
   Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. 
47. Deshpande, N., Addess, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-
   Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z., Green, R. 
   K., Flippen-Anderson, J. L., Westbrook, J., Berman, H. M., and Bourne, P. E. 
48. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, 
   571-607. Humana Press, Totowa, NJ.
50. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, 
   *Structure (Camb)* 11, 1453-9.
   Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 
   31, 3701-8.
   C., Marti-Renom, M. A., Madhusudhan, M. S., Yerkovich, B. & Sali, A. 


68. Petsko, G. A. (2001). Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol* 334, 469-78.


CHAPTER IV

INVESTIGATION OF INSERTION OF GREEN FLUORESCENT PROTEIN IN
SALMONELLA TYPHIMURIUM FLAGELLIN PROTEIN TO CREATE A
FUNCTIONAL BIONANOTUBE

Introduction

Over the last decade several bacterial surface proteins have been used to display foreign peptides as fusion proteins. Flagella display, similar to phage display, has evolved as an alternative method of peptide display, in part due to the ease of purification and tolerance for insertion of foreign peptides. Flagella display is based on the genetic fusion of foreign loop peptides into the surface-exposed nonessential: “hypervariable” central region of flagellin, which is the major subunit present in thousands of copies per flagella filament. Previous reports have demonstrated the utility of genetically inserted fusion peptides and small proteins displayed on the immunologically reactive, solvent accessible, middle domain of mesophilic E. coli flagellin.

Expression of these constructs in flagellin-deficient host strains results in hybrid flagella carrying the heterologous peptides in thousands of copies displayed in a regular array on the flagella surface. This approach has been successfully used for expression of foreign peptides/proteins to be displayed on flagella. A versatile variant of flagellar display is the hybrid display system created by the insertion of the entire E. coli thioredoxin gene into the central region of flagellin. Peptide libraries have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
been genetically introduced into a disulphide loop of thioredoxin creating a conformationally constrained library that is readily accessible on the flagellar surface. This FliTrx random peptide display library system has been used with great success for epitope-mapping purposes. Direct flagella display has also proven to be applicable also in bacterial adhesion technology since large fragments, up to 302 amino acid residues in length, of bacterial adhesins can be functionally expressed as fusions to flagellin. Several studies indicate that tolerance to internal domain fusions may be a general property of many proteins and represents another means by which protein structures may evolve. A number of foreign peptides have been successfully inserted into the hypervariable region of flagellin, including simultaneous insertion of 115 and 302 amino acid peptides. These engineered hybrid peptide-flagellin proteins are most frequently used as immunological reagents to generate vaccine antibodies against the inserted peptides when injected into animals. Only two known reports describe the insertion of full-length proteins into flagellin, the FliTrx system first described in 1995 and a second 1998 publication. Hybrid flagella are easily purified and can easily be analyzed for binding to various targets, such as immobilized proteins, tissue sections, as well as cell cultures. In order to explore the feasibility of using flagella as a bio-nanotube we randomly inserted the Green Fluorescent Protein (GFP) into flagellin.

GFP is one of the most widely used biological reporter proteins today. The wild-type GFP was isolated from the pacific jellyfish *Aequorea victoria* that dwells in the Pacific ocean, off the North American coast. Its biological role is to
transduce, by energy transfer, the blue chemiluminescence of another protein, aequorin, into green fluorescent light. GFP-like proteins with high sequence homology to GFP have also been identified in several Anthozoans (Corals). These proteins are homologous to GFP and the spectral varieties are due to minor variations in the chromophore and surrounding region in the folded protein. Since its initial cloning and expression, GFP has revolutionized several fields of biology. GFP is fusion tolerant on both its N- and C-termini and has been used as a fusion tag in various applications including protein targeting, plant cellular transport, Drosophila cellular studies, gene expression studies in bacteria, Fluorescence Resonance Energy Transfer (FRET), Bioluminescence Resonance Energy Transfer (BRET), Fluorobodies, etc. GFP has also been used as a biosensor in numerous applications by mutating the residues in the vicinity of the chromophore.

GFP is a protein with 238 residues that forms a "beta-can" motif structure, as described by Yang et al. GFP forms a monomer in solution, although other fluorescent proteins have been identified in other related organisms (corals) that are present as stable dimers and tetramers. The crystal structure of the protein (PDB:1GFL) shows a cylindrical structure formed by 11 beta strands enclosing an alpha helix at the center and some short helices at the periphery (Figure 4.1). The functional fluorophore is formed at the central α-helix by the three residues Ser65-Tyr66-Gly67. An auto-cyclization process generates the fluorophore that requires no co-factors or other components. It occurs by a rapid cyclization of Ser65 and Gly67.
followed by a slower oxygenation of Tyr66 on a timescale of hours. The Ser65 and Gly67 form an -4-(p-hydroxybenzylidene) - imidazolidin-5-one structure. The Glycine residue plays an important role in the formation of the fluorophore and cannot be replaced by any other residue. The deprotonated phenolate form of residue Tyr66 provides the major contribution to the fluorescence of GFP. The GFP has a major excitation peak at a wavelength of 395 nm and a minor one at 475 nm, with extinction coefficients of roughly 30,000 and 7,000 M⁻¹ cm⁻¹, respectively. The emission peak is at 509 nm resulting in the observation of green light. Mutational studies have shown that the N- and C- termini are important for the formation of a functional fluorophore and the minimal domain length required is from the residues 7-229; truncation of more than 7 amino acid residues from the C-terminus or more than the N-terminal Met residue lead to total loss of fluorescence.
Figure 4.1 Crystal structure of a dimeric GFP (PDB:1GFL). The crystal structure of the protein shows a cylindrical structure formed by 11 β-strands enclosing a α-helix at the center and some short helices at the periphery. b) The functional fluorophore is formed at the central α-helix by Ser65-Tyr66-Gly67.
GFP is a conformationally rigid protein due to the strong interactions between its secondary structures. The fluorophore residues have been mutated to produce various spectral and enhanced variants of GFP 21; 39; 40.

It has been demonstrated that GFP protein insertion in various proteins preserves the function of the host protein. Studies have also shown that GFP will fold and form a fluorophore when inserted into virtually any domain of another protein and is relatively indifferent to N- or C-terminal fusions 41; 42; 43. A complementation system was recently developed in which an N-terminal fragment of the Enhanced GFP (EGFP), a weak fluorophore, becomes brightly fluorescent upon complementation with the corresponding small, C-terminal EGFP fragment 44; 45. The ability of the EGFP system to respond quickly to DNA hybridization is useful for detecting the pairwise interactions of oligonucleotides \textit{in vitro} and \textit{in vivo} 45.

Apart from its natural fluorescence, GFP has been found to be functional in almost all cell types 21. The protein does not require an extensive post-translational modification to fluoresce, hence it can be used in various cell types including prokaryotic systems 21. With the above desirable characteristics, GFP serves as an excellent protein to test the feasibility of understanding the insertion tolerance level of flagellin protein. We have used a GFP transposon system created by Sheridan \textit{et al.} to randomly insert the protein into a \textit{Salmonella typhimurium} flagellin gene. We used a wild-type flagellin gene and a flagellin internal deletion variant, pTH890-C12 46 (Constructed as described in the previous chapter) for the creation of fusion proteins. Preliminary experiments with this system have failed to yield any functional GFP-
flagellin fusion variants. It is not yet known if GFP can be exported when inserted into flagellin; the high stability, low flexibility and larger physical size of GFP may prevent its export as a fusion protein. One study suggests that the type III secretion process can unfold previously folded proteins during the secretion process and another recent report describes the secretion of GFP using a modified flagellar type III secretion apparatus. Wild-type mouse dihydrofolate reductase (DHFR; 21.6 kDa) was not exported by another type III system, although a destabilized variant of DHFR was.

Materials and Methods

Bacterial strains

E. coli XL-1 Blue electrocompetent cells (Stratagene, La Jolla, CA) were used for the preparation of plasmid libraries of fliC internal deletion variants. These cells have a transformation efficiency greater than $1 \times 10^{10}$ transformants/µg of DNA, are tetracycline resistant, endonuclease (endA') deficient, which greatly improves the quality of miniprep DNA and are recombination (recA') deficient, improving insert stability.

The S. typhimurium strains used for the experiments were kindly provided by the late Dr. Robert Macnab (Yale University). S. typhimurium strain SJW1103, a derivative of serovar Typhimurium LT2 that 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 (*Afl*C and *Afl*B), 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.\(^{54,55}\) This strain is non-motile, unless a functional flagellin gene is introduced, e.g., on an expression plasmid.\(^{54}\) Thus, it serves as an ideal strain to screen engineered or mutated *fliC* genes, present in the introduced plasmids, for functional flagellin protein expression, folding, export and flagellar fiber assembly, as indicated by a simple agar microbial motility assay. *S. typhimurium* strain JR501, a stable restriction-deficient (r\(^{-}\)), methylation-proficient (m\(^{+}\)) galE strain,\(^{56,57}\) was used to convert *E. coli* grown plasmids to *S. typhimurium* compatibility. Electrocompetent cells of *S. typhimurium* strains JR501 and SJW134 were prepared using standard lab protocols.\(^{58,59}\)

**fliC** and **GFP transposon plasmids**

The pTrc series of plasmids were constructed using plasmid pKK233-2, for the regulated expression of genes in *Escherichia coli*\(^{60,61}\). These vectors carry a strong hybrid trp/lac, isopropyl-β-d-thiogalactopyranoside (IPTG) inducible promoter, the lacZ ribosome-binding site (RBS), the multiple cloning site of pUC18 and the rnb transcription terminators. The pTrc plasmids have been frequently used for cloning purposes. The pTH890 plasmid is a derivative of pTrc99A plasmid with the *S. typhimurium fliC* phase 1 flagellin gene cloned into the XbaI/HindIII site (Figure 4.2). The pTH890 plasmid was kindly provided by the late Dr. Robert Macnab (Yale University).
The Enhanced GFP (EGFP) transposon plasmids were kindly provided by Dr. Douglas L. Sheridan (Yale University). The EGFP-pBNJ11.7 transposon plasmids (Figure 4.3) were constructed using different spectral variants Green, Cyan, Yellow and Venus strains as in the expression plasmid. The GFP gene has been incorporated into an EZ::TN™ transposon system which also carries a kanamycin-resistance (KanR) at the 3' end of the GFP gene for selection of the inserts. The KanR gene is flanked with SrfI restriction endonuclease sites that can be removed after the selection of the inserts.

**Transposase reaction**

The transposons with *gfp* and KanR gene inserts were amplified from their host plasmids by performing PCR with a single 19 bp long oligonucleotide primer complementary to the Tn5 ME sequence (5'-CTGTCTCTTTACACATCT-3'). (1 μl PfuTurbo® DNA Polymerase, 1μg of pBNJ11.7 plasmid DNA, 5 μl 10x PCR buffer, 1 μl of 2.5 mM dNTP mix, 2.5 μl of 100 ng/μl primer and 36 μl of DI H2O). The PCR product was purified using a QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA). The transposition reaction was performed *in vitro* (1 μl of 200 ng/μl pTH890 DNA, 1 μl of 200 ng/μl GFP transposon, 1 μl of 10x EZ::TN™ buffer, 1 μl of EZ::TN™ Transposase enzyme and 6 μl of DI H2O) using an EZ::TN™ transposase enzyme (Epicentre Biotechnologies, Madison, WI).
Figure 4.2 A map of the pTH890 plasmid. The pTH890 plasmid for expression of *S. typhimurium* flagellin *fliC* was constructed by inserting the *fliC* gene into the pTrc99a plasmid.

Figure 4.3 A map of the pBNJ11.7 plasmid. Plasmid pBNJ11.7 carries Enhanced Green Fluorescent Protein (EGFP) gene along with Kanamycin resistance gene. The *EGFP* and *Kan*<sup>R</sup> genes are flanked with Tn5 ME transposon sequences.
The transposase reaction mixture was incubated at 37 °C for 2 hrs and stopped by heating it at 70 °C for 15 minutes. Electrocompetent XL-1 blue Gold E. coli (Stratagene, La Jolla, CA) were transformed with 0.5 μL of the transposition reaction and plated on LB agar with ampicillin (100 μg/mL) and kanamycin (50 μg/mL).

**Removal of Xmal restriction site from pTH890 plasmid**

The removal of Kan\textsuperscript{r} from the transposon after the identification of the mutants using the 8-cutter SrfI restriction enzyme proved to be very expensive. Hence, we used Xmal endonuclease, a relatively inexpensive enzyme. SrfI is a type II restriction enzyme isolated from *Saccharomyces* species with an 8 bp long recognition sequence (5'-GCCCGGGGC) and produces blunt ends\textsuperscript{63}. The Xmal restriction endonuclease has a 6 bp cut site (5'-CCCGGG) that falls within the SrfI site. Analysis of the pTH890 plasmid sequence shows that there is only one site (286\textsuperscript{th} bp) for the Xmal restriction site and does not fall in any of the open reading frames. Hence, we mutated the site to enable the use of Xmal endonuclease to remove the Kan\textsuperscript{R} gene after selection. The oligonucleotide primers used for site-directed mutagenesis are as follows; 5'-GAG CTC GGT ACC CGG AGA TCC TCT AGA AAT AAT TTT G -3' and 5'-CAAAAT TAT TTCTAGAGG ATCTCCGGGTACCGA GCTC-3'. Site-directed mutagenesis of the pTH890 plasmid (1 μl *PfuTurbo*® DNA Polymerase, 50 ng of pTH890 plasmid DNA, 5 μl 10x PCR buffer, 1 μl of 2.5 mM dNTP mix, 1.25 μl of 100 ng/μl forward and reverse primers and 36 μl of DI H\textsubscript{2}O) was performed using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA).
Figure 4.4 Insertion of GFP using transposase enzyme. a) PCR product of the EGFP gene with kanamycin resistance gene. b) The transposon was mixed with pTH890 plasmid along with the transposase enzyme for the random insertion. c) The transposon insert plasmids were isolated and cut with SrfI or XmaI enzymes to remove the KanR gene and ligated.
Screening of the GFP inserts

The transformants after an over night growth were left at 4 °C for 24 hrs and screened for fluorescence under UV light for fluorescence. The fluorescent colonies were then streaked onto fresh LB plates with ampicillin and kanamycin to confirm the fluorescence. Colony PCR was performed on the fluorescent colonies using a Techne Touchgene Gradient Thermal Cycler (Techne, Inc., Burlington, NJ). PCR reagents were obtained from New England Biolabs (Beverly, MA). Bacterial colonies were picked using 10 μl micropipette tips and added to 50 μl of PCR master mix (0.25 μl Taq DNA Polymerase (NEB), 5 μl 10x standard PCR buffer, 5 μl of 2.5 mM dNTP mix, 5 μl of 10 μM forward and reverse primers and 30 μl of DI H₂O). The PCR reaction conditions used were: an initial cycle of 94°C for 10 min 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. A forward oligonucleotide primer with DNA sequence 5’-AATTAATCCGGCCTCGT-3’ started at nucleotide position 197 of the pTH890 plasmid. A reverse oligonucleotide primer with DNA sequence 5’- ATTTAGTCTTGCGTCTTCGC - 3’ started at nucleotide position 1958. These two primers were designed to completely flank the fliC gene in the pTH890 plasmid. The PCR products were analyzed by electrophoresis on 2% agarose gels. A control PCR reaction on full-length fliC formed a band just above the 1.5 kb DNA marker. The transposon inserts should yield PCR products with molecular masses significantly larger than the wild-type PCR product.
Removal of Kan\textsuperscript{R} gene

The *E. coli* colonies with *fliC*-gfp inserts identified using colony PCR were grown as 5 ml cultures and the plasmids were extracted via a miniprep procedure. The isolated plasmids were cut with either SrfI (Stratagene, La Jolla, CA) or XmaI (NEB) restriction endonucleases to remove the Kan\textsuperscript{R} gene from the inserts. Gel electrophoresis was performed with 2% agarose gels to confirm that the Kan\textsuperscript{R} gene was removed from the plasmid. The linear plasmids from the gel were isolated using a QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA) and ligated at 16 °C overnight with T4 DNA ligase (NEB), thus producing circular plasmids. These plasmids were then transformed into the *S. typhimurium* JR501 cells to provide methylation that is compatible with the restriction-modification genes present in the other *S. typhimurium* strains. The methylated plasmids were isolated from the JR501 cells and the *S. typhimurium* SJW134 non-motile cells were transformed with the methylated form of the plasmids.

Screening for functional insertions using bacterial motility assay

The positive *fliC*-transposon insert plasmids were purified using a QIAprep Plasmid Miniprep kit (Qiagen Inc., Valencia, CA). First, the *S. typhimurium* strain JR501 electrocompetent cells were transformed with *fliC*-transposon insert plasmids before transforming into the SJW134 strain for methylation. The transformed SJW134 cells were streaked on motility agar media to test for motility. Motility agar media was composed of tryptone broth (10 g/liter tryptone, EM Science, 5 g/liter NaCl, Fisher Scientific) that contained 0.3% (wt/vol) agar (Sigma-Aldrich, St. Louis, MO). The
plates were then incubated at 30 °C for 6-8 hrs in a humidified incubator to prevent dehydration of the media. The low density of the agar allowed the bacteria to move within the agar, forming a visible halo of growth around the point of inoculation.

**Results**

**Transposase reaction**

We have performed a transposon reaction on the pTH890 plasmid with the wild-type \textit{fliC} gene and the \textit{fliC} internal deletion variant pTH890-C12 (Please refer to the previous chapter for the description of the construction of C12 deletion variant). The variant pTH890-C12 has 107 amino acids deleted from its middle domain region and hence serves as a good model to study the tolerance of large protein insertions in a smaller functional flagellin variant. The transposon insertion is a widely used genetic technique to insert, knock-out and knock-in genes in various organisms \textsuperscript{64, 65, 66, 67}. They serve as carriers of DNA fragments for both homologous and heterologous insertions. The GFP transposon system developed by Sheridan et al. \textsuperscript{62} serves as an excellent method to randomly insert the gene with an effective selection mechanism. The \textit{in vitro} transposase reaction performed along with the pTH890 gene yielded good results, as indicated by colony PCR screening (Figure 4.5). Due to the random nature of the transposon reaction, it can be incorporated into virtually any site in the plasmid. The insertions in the ampicillin resistance gene will be eliminated due to the loss of their resistance to the antibiotic on the LB plates. We did not expect any GFP expression unless it was inserted in-frame with the other two open reading frames.
(ORFs), the lacR and fliC genes. After the in vitro transposase reaction, the E. coli XL-1 blue cells were electroporated with the transposed plasmids.

**Screening for the fluorescence**

After overnight incubation, the XL-1 blue transformants were transferred to the 4 °C refrigerator and left there for 24 hrs to allow the proper folding of GFP inside the cell. A long wave ultraviolet (UV) light (365 nm) was held over the LB plates with colonies to screen for fluorescence. The fluorescent colonies were picked and freshly streaked onto new LB Amp+Kan plates. The colony PCR performed as described in the materials and methods section, shows the presence of inserts in the fliC gene of the plasmid.

![Figure 4.5 Colony PCR of the fluorescent inserts.](image)

**Figure 4.5 Colony PCR of the fluorescent inserts.** A 2% agarose gel of the colony PCR of the fluorescent transposon inserts. Lane 1 is the control PCR reaction product obtained with the C12 plasmid. Lanes 2,3,6,7 and 13 show fliC-GFP inserts. Other lanes do not show any inserts in the fliC gene.
The cells that are fluorescent but do not show any insert in the PCR product must do so due to the insertion in the *lacR* gene. The fluorescence of the GFP inserts suggests that the GFP protein does form a fluorophore in the large fusion protein. Figure 5.6 shows a 2% agarose gel of the PCR products of the colonies with inserts and no inserts in the *fliC* gene.

**Removal of Kan\(^R\) gene**

The *E. coli* colonies with *fliC-gfp* inserts identified using colony PCR were grown as 5 ml cultures and the plasmids were extracted. The isolated plasmids were cut with SrfI / Xmal restriction endonucleases to remove the Kan\(^R\) gene from the inserts (Figure 4.6). An agarose gel was run to confirm that the Kan\(^R\) gene has been removed from the plasmid. The linear plasmids from the gel were isolated, ligated (Figure 4.7) and transformed into the *S. typhimurium* JR501 cells. The plasmids were isolated from the JR501 cells and the *S. typhimurium* SJW134 non-motile cells were transformed.
Figure 4.6 Gel picture of the removal of Kanamycin gene. A 2% agarose gel of the fliC-GFP insert plasmids treated with Srlf / Xmal restriction endonucleases. Lane 1 has a C12 plasmid. Lanes 2-8 show the small and large fragments corresponding to the cut kanamycin resistance gene and the rest of the linearized plasmid.

Figure 4.7 Ligation of the fliC-GFP inserts. A 2% agarose gel with the linear plasmids ligated using T4 DNA ligase. Lane 1 shows the 1 kb marker, lane 2 shows the linear plasmid, lanes 3-6 show the ligation products.
Motility screening

The SJW134 cells with the fliC-gfp gene inserts were screened for the motility on motility media. We have used S. typhimurium wild-type strain SJW1103 as a positive control and untransformed S. typhimurium SJW134 cells as a negative control for the motility experiments. Due to the low percentage of agar in the motility media, the cells producing functional flagella can swarm across the media, creating a halo around the inoculation zone. Along with the motility we also tested for fluorescence of expressed, folded GFP with a long wave UV light (365 nm). We did not identify any of the fliC-gfp inserts with swarming ability in either the wild-type or in the C12 inserts. Altogether ~100 wild-type flagellin-GFP insertion variants and ~200 C12 flagellin-GFP insertion variants were screened for motility, with no positive identification of any functional internal GFP insertions.

Discussion

GFP is a highly stable protein; the purified wild-type form is still functionally fluorescent in aqueous buffer solution after years of storage at 4 °C. GFP serves as a very effective reporter protein for the expression of the fusion protein in many types of cells. Due to its rigid core structure, GFP folds into its native structure irrespective of the length and conformation of the attached fusion protein. Flagellin is a relatively large bacterial protein with 494 amino acid residues. The flagellin export process is similar to other type III bacterial pathogen export systems, e.g. (Yersinia sp. YoP proteins) that release protein to the extracellular environment through a protein tube
Following translation, the C-terminal region of FliC flagellin protein is bound to the FliS protein, in a 1:2 ratio, which prevents self-assembly and formation of flagellin fiber aggregates within the cell. A new report published by Muskotal et al. suggests that the FliS protein binds to the C-terminal region of the FliC in the ratio of 1:1. The report suggests that FliS does not act as an unfolding factor. Instead it helps in the formation of alpha-helical secondary structure in the region of FliC where it binds. The FliC protein is exported to the distal end of flagella fibers through the interior channel of the flagella fiber, which has a diameter of only 2 nm (20 Å) wide. In order to pass through the flagella tube, the flagellin monomers are thought to be in a semi-folded or completely unfolded state, due to the severe size constraints posed by the pore diameter. None of the ~100 wild-type and ~200 C12 variant GFP inserts with detectable fluorescence identified showed any motility, indicating a failure of the hybrid flagellin-GFP variants to be exported and assembled into functional flagella. However, the observed fluorescence of the transformed E. coli cells suggests that the inserted GFP properly folded as an internal fusion protein in flagellin.

There are several possible explanations for these results. The GFP protein, when folded in a native state, is about 50 Å long and 30 Å wide. Thus, the inserted GFP domain may simply be too large to fit through the flagella pore and export apparatus in its native, folded state. Furthermore, GFP may be too stable to be unfolded by the FliI export protein and other proteins in the type III export complex that did not evolve to unfold GFP. The cytoplasmic FliS chaperonin protein only

196

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
binds one C-terminal flexible end of the FliC protein, and should have minimal, if any, effect on the folding or subsequent prerequisite unfolding of the inserted GFP domain in the middle of the FliC protein. Because the channel diameter through the filament is only 2 nm, proteins to be exported must be largely unfolded for entry into and translocation through the channel. The flagellar type III protein export apparatus is a complex made of six membrane proteins, FlhA, FlhB, FliO, FliP, FliQ, and FliR, and three soluble proteins, FliH, FliI, and FliJ. The FliI protein is thought to be an ATP-dependent chaperonin that exports the FliC protein. Salmonella InvC, the FlII homolog of the virulence type III secretion system, has been shown to induce chaperone release from and unfolding of the cognate protein to be secreted in an ATPase-dependent manner, suggesting that FliI may have a similar function and mechanism.

Although the negative results of functional screening for motility were not based on an exhaustive study, this study indicates that GFP does not readily serve as a good model protein for insertion into flagellin proteins to enable the creation of a flagella nanotube with genetically encoded fluorescence. Anecdotal evidence indicates that other attempts to export GFP through some types of trans-membrane export systems may have also resulted in failure, although publications that describe these failed attempts are not available. There is evidence suggesting an efficient intra- and intercellular transport of GFP fusion proteins in eukaryotic cells. This could be due to a more efficient unfolding and exporting systems present in the eukaryotic cells and their membrane bound organelles. A discussion with a visiting scientist,
Prof. Guy Cornelis from Biozentrum, University of Basel, indicated that GFP has been tried as a fusion tag in other type III export systems by various researchers, with no success. Furthermore, a recent seminar at a 2006 Protein Society meeting given by Prof. S. Choe, Salk Institute, indicated that GFP-fusion proteins were not always successfully exported from one cellular compartment to another. This study also suggests that, due to the high stability and structural constraints of the GFP protein, it does not serve as a good internal fusion tag for labeling and possible sensor functions in flagellin and other type III export systems. One study suggests that the type III secretion process can unfold previously folded proteins during the secretion process and other recent reports describe the secretion of GFP using a modified flagellar type III secretion apparatus and a destabilized variant of dihydrofolate reductase.

Perhaps a destabilized form of GFP, the opposite in terms of a recently described “superfolder” GFP variant could be yield a feasible method for developing GFP fusions that are more amenable to functional insertion and export across various cellular membranes.

References


200

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


CHAPTER V

EXPERIMENTAL INVESTIGATION OF THE HYPOTHETICAL S. TYPHIMURIUM FLAGELLIN MECHANICAL SWITCH BY INSERTION OF A DISULFIDE BRIDGE

Introduction

The eubacterial flagellum, a helical membrane-bound protein filament, helps in propelling the cells through their liquid environment\(^1\textsuperscript{,2}\). The flagellum is a protein complex that is composed of at least 20 different proteins\(^2\textsuperscript{,3}\textsuperscript{,4}\textsuperscript{,5}\textsuperscript{,6}\) and has several major features: a basal body, a trans-membrane motor, a hook structure and an elongated helical filament. The filament has an outer diameter of 12–25 nm, a 2–3 nm diameter inner channel\(^6\textsuperscript{,7}\) and can be 1–15 μM or more in length\(^5\textsuperscript{,8}\). It is composed of as many as 30,000 subunits of a single flagellin protein\(^9\textsuperscript{,10}\). The mature S. typhimurium phase 1 flagellin protein, encoded by the fliC gene, consists of 494 amino acids; the first methionine residue is removed post-translationally. The structure of S. typhimurium flagellin was recently determined (PDB 1UCU)\(^8\textsuperscript{,11}\) (Figure 5.1) and shows four distinct globular domains in the protein, the conserved D0 and D2 domains and the variable D2 and D3 domains (Fig. 5.1). Previous studies show that the helical flagellum filament exists in two supercoiled forms, the L- and R-forms. The transition of the flagella filament into either of the supercoiled forms leads to the tumbling and running motions of bacteria\(^12\textsuperscript{,13}\). Electron cryomicroscopy studies and X-ray fiber diffraction studies have shown that the flagellin subunit repeat

\[205\]
distances of the L- and R-forms of the fibers are 52.7 Å and 51.9 Å, a difference of

Figure 5.1 A pictorial depiction of the mutated Cys residues. A three dimensional figure of PDB: 1UCU showing the conserved D0 and D1 domains and the variable D2 and D3 domains. The inset shows the mutated cysteine residues and the disulfide bridge. Swiss-Pdb viewer (http://www.expasy.org/spdbv/) software was used to visualize the structure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
0.8 Å. Simulation studies performed by Samatey et al., by fixing the top molecule and pulling the bottom molecule in 0.1 Å steps predicted that the abrupt downward shift of the beta-hairpin turn (140-160 residues) in the D1 domain leads to the change in the repeat distance of the L- and R-form protofilaments. In order to test this hypothesis we have mutated the residues Asn56 and Ala149 to cysteine, to allow formation of a covalent disulfide bond. This disulfide bond cross-link could effectively lock the hairpin turn into place at one location, and thereby prevent any conformational changes in the flagellin protein that are predicted to occur upon conversion between the L- and R-forms. Any observed differences in the running and tumbling motility behavior of *S. typhimurium* bacteria expressing the engineered disulfide loop variant would help to confirm or disprove the original hypothesis that this loop region changes conformation.

**Materials and Methods**

**Strains and plasmid**

The *S. typhimurium* strain and the *fliC* plasmid used for the experiment were kindly provided by the late Dr. Robert Macnab (Yale University). *S. typhimurium* strain SJW134 (*A*fliC and *A*fliB), 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. This strain is non-motile, unless a functional flagellin gene is introduced. The pTH890 plasmid is a derivative of pTrc99A plasmid with the *S. typhimurium fliC* phase 1 flagellin gene cloned into the XbaI/HindIII site.
Construction of cysteine disulfide bond

The cysteine disulfide bond was selected by using the program Disulfide by Design. Site-directed mutagenesis of pTH890 plasmid (1 μl PfuTurbo® DNA Polymerase, 50 ng of pTH890 plasmid DNA, 5 μl 10x PCR buffer, 1 μl of 2.5 mM dNTP mix, 1.25 μl of 100 ng/μl forward and reverse primers and 36 μl of DI H2O) was performed using QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). The primers for generating the N56C mutation in the fliC gene were: 5'-TAA CCG TTT TAC CGC GTG CAT CAA AGG TCT GAC TC- 3' and 5' –GAG TCA GAC CTT TGA TGC ACG CGG TAA AAC GGT TAG- 3'. Primers for generating the A149C mutation are 5'-CCA TCC AGG TTG GTT GCA ACG ACG GTG AAA C- 3' and 5'-GTT TCA CCG TCG TTG CAA CCA ACC TGG ATG G- 3'. The pTH890 plasmid that has the mutation N56C is designated as pTH890-Cys1 and the plasmid that has both Cys residues (N56C+A149C) mutated is designated as pTH890-Cys2.

Partial purification of fibers and SDS-PAGE

To test whether the mutated flagellin subunits are exported and self-assembled into flagella fibers, the flagella fibers were partially purified. Bacterial cultures were inoculated into 15 ml LB broth with 15 μl of 100 mg/ml ampicillin and allowed to grow overnight at 37 °C. A 15 ml volume of each bacterial culture was transferred into a 15 ml polypropylene tube and placed in a heat block at 60 °C for 10 min to dissociate the flagella into monomers. The heated cells were pelleted by centrifugation at 6000 x g and the supernatant was transferred to 15 ml Amicon Ultra
centrifugal filter units with a 30 kDa MWCO membrane (Millipore Corp., Billerica, MA) for concentrating the flagella fibers. The concentrates were then resolved by reducing 10% SDS-PAGE, followed by staining with Coomassie Brilliant Blue R-250 dye.

Ellman’s assay

The Ellman’s reagent assay \(^{18}\) was performed on the mutated flagellin protein to test for the presence of unreacted cysteine residues in the flagellin variants. The generation of the yellow reaction product that is characteristic of reaction with free thiols in their reduced state would indicate the presence of a single cysteine residue in either of the N56C and A149C single mutation variants. Furthermore, the formation of a disulfide bond between the two cysteine residues in the A149C+N65C double mutation variant should result in no reaction with DTNB; conversely, reaction of DTNB to form a significant quantity of yellow reaction product would indicate a failure of the two introduced cysteine residues to form the desired disulfide bond.

The flagella fibers were from the wild-type plasmid and both the cysteine variants were partially purified as described above. Ellman’s reagent, 5,5’-Dithiobis(2-nitrobenzoic acid) (DTNB) (Sigma-Aldrich, St. Louis, MO) was dissolved in 50 mM sodium acetate solution to yield a 2 mM stock and a stock solution of 1 M Tris-HCl (pH 8.0) was prepared and stored at 4 °C. The assay was performed by addition of 100 µl of 1 M Tris-HCl and 100 µl of 2 mM Ellman’s reagent to 800 µl of 1 mg/ml of protein solution at 25 °C. The assay mixture was incubated for 5 minutes and the
absorbance at 412 nm wavelength ($A_{412}$) was measured at times of 0 minutes and 5 minutes with a SpectraMax384 spectrophotometer.

**Motility assay**

SJW134 strain *S. typhimurium* cells were transformed by electroporation with either the wild-type *fliC* control plasmid pTH890, or the pTH890-Cys1 and pTH890-Cys2 plasmids, and were grown overnight at 37 °C. Volumes of 2 µl of each overnight culture, diluted with LB to an OD$_{550}$ of 0.5, were inoculated on motility agar media (10 g/liter tryptone (EM Science), 5 g/liter NaCl, (Fisher Scientific) and 0.3% (wt/vol) agar (Sigma-Aldrich, St. Louis, MO) to test for variations in the cellular motility. The plates were then incubated at 30 °C for 8 hr in a humidified incubator to prevent dehydration of the media.

**Video microscopy and swimming analysis**

A Nikon Eclipse E600 epifluorescence stereo microscope (Mager Scientific Inc., Dexter, MI) with 10X, 40X and 100X (oil immersion) Planfluor objectives, transmission and darkfield condenser light sources and a QImaging Fast 1394 cooled mono (grayscale) CCD digital camera was used to record visible images and movies of non-motile *S. typhimurium* SJW134 bacteria transformed with pTH890 expression plasmids encoding either wild-type or cysteine mutation variants of flagellin. The video images of swimming bacteria were captured at 13 fps speed using a 40X oil immersion lens. Bacterial cultures were grown to mid-log phase in LB media without shaking to minimize breakage of potential brittle flagella fibers. A 30 µl volume of
Figure 5.2 SDS-PAGE and motility analysis of the wild-type flagellin and two cysteine mutation variants. a) A 10% reducing SDS-PAGE protein gel of partially purified flagella fibers. Lane 1 has the sample from the pTH890 transformed cells, lane 2 with pTH890-Cys1 and lane 3 with pTH890-Cys2 transformed cells. b) A picture depicting the swarming of SJW134 cells transformed with different plasmids. Each culture at OD_{550} 0.05 was loaded onto a concave glass slide and covered with a glass cover slip.
The SDS-PAGE analysis of the fibers from pTH890, pTH890-Cys1 and pTH890-Cys2 show that the cysteine mutant flagella fibers were efficiently exported (Figure 5.2a). Furthermore, both single and double cysteine mutation variants were as functional as the wild-type in providing motility, as indicated by the swarming agar motility assay. It has been previously demonstrate with the *E. coli* FliTrx flagellin that disulfides can be tolerated to some extent, allowing for successful export and assembly of flagellins containing a thioredoxin domain \(^{19,20}\). The results of Ellman’s assay (Table 5.1) indicate the successful formation of a disulfide bond in the pTH890-Cys2 variant.

### Table 5.1 Ellman’s assay of the wild-type and cysteine mutants. \(^{1,2}\)

<table>
<thead>
<tr>
<th></th>
<th>Wild-type flagellin</th>
<th>pTH890-Cys1</th>
<th>pTH890-Cys2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 minutes</td>
<td>0.140</td>
<td>0.167</td>
<td>0.145</td>
</tr>
<tr>
<td>5 minutes</td>
<td>0.163</td>
<td>0.221</td>
<td>0.152</td>
</tr>
</tbody>
</table>

\(^{1}\) The Ellman’s assay was performed as described in the Materials and Methods section.

\(^{2}\) The numerical values indicate the measured O.D\(_{412}\) values of each protein variant after the given time periods.

The swarming diameters measured for the three cultures were almost identical (pTH890; 31 ± 2 mm, pTH890-Cys1; 30 ± 1.3 mm and pTH890-Cys2; 31 ± 2.2 mm),
suggesting that the formation of a disulfide bond did not alter the swarming ability of bacteria in a relatively viscous media (Fig 5.2b). Video microscopy analysis of the swimming behavior of the cells with wild-type protein and cysteine variants did not show any variation in the swimming ability of the mutants. These results suggest that the formation of the disulfide bond formation at the proposed mechanical switch region does not alter the transformation of the fibers from L-form to R-form (or vice versa).

**References**


CHAPTER VI

HIGH-THROUGHPUT SCREENING FOR ANTIMICROBIAL COMPOUNDS USING A 96-WELL FORMAT BACTERIAL MOTILITY ABSORBANCE ASSAY

Abstract

There is a pressing need to develop new antimicrobial drugs, due to the increasing resistance of pathogenic bacteria to existing antibiotics. The preliminary development and validation of a novel methodology for the high-throughput screening of antimicrobial compounds and inhibitors of bacterial motility is described. This method uses a bacterial motility swarming agar assay combined with the use of both centered and offset inoculation of the wells in a standard clear 96-well plate to enable rapid screening of compounds for potential antibiotic and anti-motility properties with a standard absorbance microplate reader. Thus, the methodology should be compatible with 96-well and 384-well plate laboratory automation technology typically used in drug discovery and chemical biology studies. To validate the screening method, we screened a structurally diverse library of 960 biologically-active compounds, the Genesis Plus compound library, against a motile strain of the gram-negative bacterial pathogen, Salmonella typhimurium, to validate the assay. A collection of 60 compounds with well-known antibiotic properties were successfully identified using this assay.
Introduction

Antibiotic resistance in bacteria is currently a growing, major public health problem; bacteria may soon evolve resistance to all known antibiotics.\textsuperscript{1, 2, 3} Commercial research in this area has waned over the last several decades and consequently, few new classes of antibiotics have recently been developed.\textsuperscript{1, 3, 4} Bacterial signal transduction components, efflux pumps and genes involved in motility and biofilm formation represent some of the possible targets for antibiotic drug discovery research.\textsuperscript{5, 6} Bacterial motility frequently involves the chemotaxis-controlled rotation of extracellular protein fibers called flagella,\textsuperscript{7, 8} which are also involved in evasion of host immunity,\textsuperscript{9} host colonization and biofilm formation.\textsuperscript{10} Thus, the bacterial flagellar system represents a possible target for antimicrobial drugs.

Motility inhibitors could potentially be used with other antimicrobial compounds to prevent the spread of infectious pathogens in the environment or the host organism. A high-throughput screening (HTS) assay for detecting inhibition of flagellar motility or other types of bacterial motility, e.g. pili-mediated motility, could be used for screening compounds for antimicrobial activity and also in fundamental structure-function studies of genes and proteins involved in motility. Here, we report a simple screening method for the identification of small molecules that inhibit either bacterial growth or bacterial motility.
Materials and Methods

Two basic techniques were combined to enable rapid screening for motility inhibitors; a bacterial swarming agar motility assay and the use of 96-well microplates common to HTS assay methods. Initial proof of concept was demonstrated with two Salmonella typhimurium strains: S. typhimurium strain SJW1103,¹¹ which is wild-type for flagellar motility and non-motile S. typhimurium strain SJW134,¹² which has both the fliC and fljB flagellin genes deleted and cannot produce flagella fibers. A volume of 200 µl of freshly prepared motility agar, composed of 1% tryptone (EM Science, Gibbstown, NJ), 0.5% NaCl (Sigma-Aldrich, St. Louis, MO) and 0.3% agar (Calbiochem, La Jolla, CA), was pipetted into each well of a sterile 96-well U-bottom plate (Linbro, Flow Laboratories McLean, VA) using a Matrix Impact² P1250 programmable 8-channel pipettor (Matrix Technologies Corp., Hudson, NH) and allowed to solidify at 4 °C overnight. This low agar media remains in a liquid state until the temperature is lowered to approximately 35 °C; hence, it can be effectively mixed with thermolabile screening compounds. Liquid 50 ml culture volumes of motile SJW1103 and non-motile SJW134 S. typhimurium strains were grown to saturation (overnight) in Luria-Bertani (LB) media at 30 °C, with shaking at 150 rpm. Cultures were poured into a sterile lid designed for 96-well microtitre plates. A plastic sterile disposable 96-pin tool (V&P Scientific, Inc., San Diego, CA) was first dipped into the liquid culture, removed, and
then carefully lowered into the agar media to inoculate it. The inoculated agar plates were incubated for 12-15 hrs at 30 °C in a humidified incubator.

A slight modification of this screening method was used to screen the Genesis Plus (GenPlus) library of 960 biologically active, structurally diverse compounds (MicroSource Discovery Systems, Inc., Gaylordsville, CT) for inhibitors of bacterial growth and motility. A 3 μl volume of each 2 mM stock compound in 100% DMSO was pipetted into each well of a 96-well tissue culture plate with a CCS Packard PlateTrak robotic pipetting station (Perkin Elmer, Shelton, CT), followed by addition of 197 μl of liquefied motility agar media with a Matrix Impact2 P1250 programmable 8-channel pipettor. This yielded a final screening compound concentration of 30 μM, within the typical range used for screening assays. The compound and media were then mixed by pipetting up and down 5-6 times to insure even distribution of the compound within the well and allowed to solidify overnight. Bacterial cultures (50 ml) of motile and non-motile S. typhimurium strains were grown to saturation overnight in LB media at 30-37 °C, with shaking at 225 rpm. Each culture was pipetted into a sterile 96-well plate and a sterile plastic disposable 96 pin tool (V&P Scientific, San Diego, CA) was dipped into the plate containing each culture and then used to inoculate the motility agar media in each well of the 96-well microplates. The plates were incubated for 12-15 hrs at 30 °C in a humidified incubator. Plates were then inoculated with motile S. typhimurium SJW1103 cells and incubated overnight for 12-15 hrs at 30 °C under humid conditions.
The initial agar growth and motility screening procedure was further modified by varying the position where each well of the agar motility plates were inoculated. Overnight liquid cultures of motile SJW1103 and non-motile SJW134 *S. typhimurium* control strains in LB media were transferred into sterile 25 ml disposable reservoirs and 200 μl was pipetted into each sample or control well of a sterile 96-well tissue culture plate with a Matrix Impact® P250 8-channel pipettor. The PlateTrak instrument was equipped with a P250 (250 μl x 96 tip capacity) pipetting head with horizontal X- and Y-axis plate offset capability for dispensing liquids into 384-well plates. The sterile pipette tips were first lowered into the 96-well source plate containing the overnight liquid cultures and then raised. A 96-well plate containing sterile motility agar was then placed under the pipetting station and laterally offset by the PlateTrak to allow the 96 pipette tips to access the upper right quadrant of each well, corresponding to one fourth of the wells in a 384-well plate. The tips were lowered approximately halfway into the agar in each well of the motility agar plate to inoculate the media and then removed. The plates were then incubated for 12-15 hrs at 30 °C in a humidified incubator, as described previously.

**Results and Discussion**

The wild-type *S. typhimurium* motile cultures spread in all directions of the motility media within the wells after overnight incubation, while the non-motile cultures only grew in the immediate vicinity of the inoculation zone, leaving clear surroundings (Figure 6.1). Thus, the difference in the motile vs. non-motile cells was
readily apparent to the human eye; furthermore, this method might be readily adapted for simple high throughput screens of potent antibacterial compounds that completely inhibit growth of the inoculated cells. However, questions remained about the practicality of adapting this method for high-throughput screening of compounds against bacterial strains with a microplate spectrophotometer to identify more subtle differences in inhibition of growth and motility. Therefore, the absorbance of each well at 550 nm wavelength (OD$_{550}$) was measured with a SpectraMax Plus$^{384}$ spectrophotometer (Molecular Devices Corporation, Sunnyvale, CA). However, differences between the motile and non-motile $S$. typhimurium cells were not readily discernable by absorbance measurements with the plate reader, as all OD$_{550}$ readings yielded similar high values. This was due to the growth of cells in the inoculated center region of each well to high densities, regardless of motility. Thus, the results of this first method were not amenable to differentiating between motile and non-motile strains with a spectrophotometer (Figure 6.2).

Given that visual inspection was successful in identifying hits, a “high-content” screening digital camera image analysis system could probably be adapted to detect hits with this assay format; however, this approach may require a complex and expensive system that may not be available in many smaller academic and commercial labs. Visual inspection of the plates, a simple form of high content screening, yielded a total of 64 hit compounds that significantly inhibited bacterial growth. All of these compounds possessed known or suspected antimicrobial activity, e.g., amoxicillin, ciprofloxacin and rifampin. Furthermore, several different types of
growth patterns in the wells were observed: a complete spread of the cells in the well, a total absence of growth and very faint growth at the region of inoculation but no spreading of the cells (Figure 6.1B). One well indicated some growth but incomplete swarming of the cells and contained the compound hexachlorophene (Figure 6.1B), a chlorinated bisphenol compound with known bacteriostatic action against gram-positive organisms, but less effective against gram-negative organisms.

We also attempted to perform a more quantitative, HTS compatible analysis of the center-inoculated plates using a plate reader. Following overnight growth, the absorbance (OD$_{550}$) was measured for each of two duplicate plates with the SpectraMax$^{384}$ plate reader. The plate reader failed to detect significant differences in the OD$_{550}$ values of the controls, which were all close to a value of 1.0, as noted above with the motile and non-motile controls. For example, the average OD$_{550}$ values of the 80 compound sample wells in two duplicate plates of the GenPlus plate 2 were $0.88 \pm 0.22$ and $0.90 \pm 0.22$. The signal to noise ratios for the averaged motile negative controls vs. non-motile positive controls were 0.89 and 0.97. The statistical assay quality was determined by calculating the $Z'$ factor, given in Eq. 1:

$$Z' = 1 - (3\sigma_+ + 3\sigma_-)/(\mu_+ \sigma \mu_-)$$ (1)
Figure 6.1 Comparison of motile and non-motile bacterial growth patterns in 96-well plate-formatted motility agar. (A) Image of entire 96-well plate containing motility agar that was inoculated with motile and non-motile strains of *S. typhimurium* bacteria and enlarged images of two example wells. A 96-well, U-bottom plate containing sterile Luria-Bertani (LB) 0.3% agar media was inoculated with non-motile SJW134 and wild-type SJW1103 *S. typhimurium* strains in the upper right quadrant of each well and incubated at 30 °C for 12 hours. Columns 1-6 were inoculated with motile SJW1103 cells and columns 7-12 contained non-motile SJW134 cells. The non-motile cells grew only around the region of inoculation, well

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
away from the center of each well where optical absorbance would typically be measured by a 96-well microplate reader. In contrast, the wild-type cells swarmed across the whole well. (B) Enlarged images showing three types of growth patterns observed for *S. typhimurium* cells inoculated in a 24-well tissue culture plate. Left image: wild-type SJW1103 cells in plain LB agar. Center image: wild-type SJW1103 cells grown in the presence of 30 μM hexachlorophene, showing partial inhibition of growth and motility. Right image: non-motile SJW134 cells grown in plain LB agar.

### Average absorbance values of *Salmonella* controls

![Graph](image)

**Figure 6.2** Comparison of average absorbance readings for motile and non-motile *S. typhimurium* bacterial cells. When the cells were inoculated in the center of each well the absorbance readings always yielded values close to 1.0. However, when the location of inoculation was offset, i.e., moved to the upper right, the spectrophotometer was able to discriminate between the two strains by large differences in the absorbance readings.
Where $\sigma_{c+}$ is the standard deviation of the positive control for the assay, $\sigma_{c-}$ is the standard deviation for the negative control, and $(\mu_{c+} - \mu_{c-})$ define the mean values for the positive and negative controls. Analysis of the positive and negative control data for the GenPlus plate 2 OD$_{550}$ data yielded anomalously high $Z$ values of 5.4 and 38.6. If screening strictly for antimicrobial growth compounds, the $Z$ values would be greatly improved by using an antibiotic for the positive control instead of the non-motile strain, using this center-well inoculation method.

Although the positive controls did not provide a useful reference, we were able to statistically analyze the absorbance data with respect to the deviation of individual sample readings from the average value of each plate. The average ($\bar{x}$) and standard deviation ($\sigma$) of the OD$_{550}$ readings were computed for each of the two duplicate plates using a spreadsheet; the absorbance data in each well were analyzed with respect to the difference between each well and the average plate value. Strong inhibitors were defined by the criteria that the OD$_{550}$ of a well containing a strong inhibitor was less than the mean plate value by two times the standard deviation \{OD$_{550} < (\bar{x} - 2\sigma)\}. This cut-off criterion yielded a total of 60 compounds that significantly inhibited the growth of the bacteria, or 93.8% of those identified by the prior visual inspection method. Although no previously unknown antibiotic compounds were identified among the hits, these results indicate the general validity of the screening method for detecting antimicrobial compounds, although a further improvement was possible, as demonstrated below.
The modified "offset inoculation" method was used to inoculate one quadrant of each well instead of the center well position in motility agar plates. Using this method, we were able to visually differentiate between the wells inoculated with motile and non-motile cells after 9-10 hrs of incubation; visual inspection of the plates readily discriminated between the motile wild-type control and non-motile controls (Figure 1A), with 100% repeatability. The wild-type cultures spread in all directions of the motility media in the wells, while non-motile cultures only grew in the immediate vicinity of the inoculation zone. This resulted in clear media in the center of each well for non-motile cells; this location is where absorbance measurements by a 96-well plate reader would normally occur, a key advantage for using automated screening approaches. To demonstrate compatibility with HTS laboratory automation, growth curves in each well were monitored after inoculation by absorbance measurements at 550 nm (OD\textsubscript{550}), using a SpectraMax Plus\textsuperscript{384} spectrophotometer. This approach consistently detected a significant difference between the growth of motile and non-motile \textit{S. typhimurium} strains (Figure 6.3).
Average Growth of Motile vs. Non-motile Salmonella

Figure 6.3 Growth curves for motile and non-motile S. typhimurium bacterial cells inoculated in 96-well plate-formatted motility agar. S. typhimurium cells, either wild-type for motility (strain SJW1103) or non-motile (strain SJW134), were inoculated in the upper right quadrant of each well, as shown in Figure 1. The time-dependent absorbance readings (OD_{550}) from two identical 96-well plates containing six columns (48 wells) for each bacterial strain were averaged to generate each data point; the error bars represent the computed standard deviation of each set of measurements. The difference in a non-motile strain vs. a motile strain and the rate of growth of the motile strain can be readily determined with this assay method.

The hit compounds previously identified in the visual screen were used to prepare a master hit plate and retested for inhibition of growth and motility by this offset inoculation method. The absorbance at 550 nm of each well (OD_{550}) was
measured with the SpectraMax® Plus384 spectrophotometer. The reproducibility of the assay controls was determined by calculating the Z' factor; the average Z' value was found to be 0.80 ± 0.08, an acceptable value. Both the type of controls (motile and non-motile) and the location of inoculation contributed to the quality of the assay. Because of the different strains of *Salmonella* used as controls, the assay was able to consistently detect a wide range of antibiotic activity (Table 6.1).

Encouraged by the initial results of the offset inoculation method with previously identified hits, the entire 960 compound GenPlus library was rescreened in duplicate, using the offset inoculation method, to verify the consistency of the results. The hits identified in this more quantitative assay were 100% identical to the previous screening methods. The average Z' value for all 12 compound plates was 0.62 ± 0.19, an acceptable value for HTS format assays. These preliminary screening results, obtained with a model compound library, indicate that the HTS-compatible "offset inoculation" method can be used to screen compounds for bactericidal, bacteriostatic, and motility inhibition activity. Alternatively, a microplate spectrophotometer might be programmed to read a 96-well plate in 384-well detection mode, with a subsequent statistical analysis of the four offset quadrants in each well to monitor well-to-well changes in absorbance resulting in bacterial growth and spreading.
Table 6.1 Complete list of antibacterial compounds detected in screen of Genesis Plus library with offset motility detection method.

<table>
<thead>
<tr>
<th>Compound name</th>
<th>Functional classification</th>
<th>Antibiotic effectiveness a, b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cephaloridine</td>
<td>antimicrobial</td>
<td>106.13% (±5.66%)</td>
</tr>
<tr>
<td>2. Ofloxacin</td>
<td>antimicrobial</td>
<td>105.68% (±5.41%)</td>
</tr>
<tr>
<td>3. Tetracycline</td>
<td>antibacterial</td>
<td>105.24% (±5.01%)</td>
</tr>
<tr>
<td>hydrochloride</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Lomefloxacin</td>
<td>antimicrobial</td>
<td>105.21% (±4.30%)</td>
</tr>
<tr>
<td>hydrochloride</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Flumequine</td>
<td>antibacterial</td>
<td>105.04% (±5.36%)</td>
</tr>
<tr>
<td>6. Methacycline</td>
<td>antibacterial</td>
<td>104.94% (±5.73%)</td>
</tr>
<tr>
<td>hydrochloride</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Ciprofloxacin</td>
<td>antibacterial, fungicide</td>
<td>104.92% (±5.12%)</td>
</tr>
<tr>
<td>8. Thimerosal</td>
<td>anti-infective</td>
<td>104.90% (±5.23%)</td>
</tr>
<tr>
<td>9. Cefuroxime sodium</td>
<td>antimicrobial</td>
<td>104.72% (±5.22%)</td>
</tr>
<tr>
<td>10. Cefamandole nafate</td>
<td>antimicrobial</td>
<td>104.60% (±6.13%)</td>
</tr>
<tr>
<td>11. Bacampicillin</td>
<td>antibiotic</td>
<td>104.39% (±3.85%)</td>
</tr>
<tr>
<td>hydrochloride</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Trimethoprim</td>
<td>antibacterial</td>
<td>104.28% (±6.82%)</td>
</tr>
<tr>
<td>13. Ceftriaxone sodium</td>
<td>antibiotic</td>
<td>104.14% (±1.91%)</td>
</tr>
<tr>
<td>14. Polymyxin b sulfate</td>
<td>antibacterial</td>
<td>103.80% (±2.81%)</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Compound name</th>
<th>Functional classification</th>
<th>Antibiotic effectiveness&lt;sup&gt;a,b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>15. Doxycycline hydrochloride</td>
<td>antibacterial</td>
<td>103.73% (±3.03%)</td>
</tr>
<tr>
<td>16. Chloroxine</td>
<td>chelating agent</td>
<td>103.14% (±2.26%)</td>
</tr>
<tr>
<td>17. Broxyquinoline</td>
<td>antiseptic; disinfectant</td>
<td>103.07% (±2.50%)</td>
</tr>
<tr>
<td>18. Enoxacin</td>
<td>antibacterial</td>
<td>102.72% (±2.06%)</td>
</tr>
<tr>
<td>19. Minocycline hydrochloride</td>
<td>antibacterial</td>
<td>102.53% (±1.74%)</td>
</tr>
<tr>
<td>20. Cephapirin sodium</td>
<td>antibacterial</td>
<td>102.31% (±1.11%)</td>
</tr>
<tr>
<td>21. Cefoperazone sodium</td>
<td>antimicrobial</td>
<td>102.28% (±4.61%)</td>
</tr>
<tr>
<td>22. Meclocycline sulfosalicylate</td>
<td>antibacterial</td>
<td>101.97% (±1.30%)</td>
</tr>
<tr>
<td>23. Norfloxacin</td>
<td>antibacterial</td>
<td>101.82% (±0.07%)</td>
</tr>
<tr>
<td>24. Phenylmercuric acetate</td>
<td>fungicide</td>
<td>101.81% (±1.69%)</td>
</tr>
<tr>
<td>25. Furazolidone</td>
<td>antimicrobial</td>
<td>101.63% (±0.32%)</td>
</tr>
<tr>
<td>26. Alexidine hydrochloride</td>
<td>antibacterial</td>
<td>101.57% (±0.86%)</td>
</tr>
<tr>
<td>27. Nalidixic acid</td>
<td>antibacterial</td>
<td>101.46% (±0.16%)</td>
</tr>
<tr>
<td>28. Oxytetracycline</td>
<td>antibacterial</td>
<td>101.34% (±0.31%)</td>
</tr>
<tr>
<td>29. Hetacillin potassium</td>
<td>antibacterial</td>
<td>101.28% (±1.02%)</td>
</tr>
</tbody>
</table>
### Table 6.1 – Continued

<table>
<thead>
<tr>
<th>Compound name</th>
<th>Functional classification</th>
<th>Antibiotic effectiveness&lt;sup&gt;a,b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>30. Bergaptene</td>
<td>antipsoriatic</td>
<td>101.12% (±0.83%)</td>
</tr>
<tr>
<td>31. Mitomycin C</td>
<td>antineoplastic</td>
<td>101.04% (±1.72%)</td>
</tr>
<tr>
<td>32. Hexachlorophene</td>
<td>topical anti-infective</td>
<td>100.79% (±1.15%)</td>
</tr>
<tr>
<td>33. Rifampin</td>
<td>antibacterial</td>
<td>100.77% (±2.84%)</td>
</tr>
<tr>
<td>34. Aureomycin</td>
<td>antibacterial</td>
<td>100.38% (±3.24%)</td>
</tr>
<tr>
<td>35. Nitrofurantoin</td>
<td>antibacterial</td>
<td>100.20% (±1.97%)</td>
</tr>
<tr>
<td>36. Zidovudine</td>
<td>antiviral</td>
<td>99.92% (±5.42%)</td>
</tr>
<tr>
<td>37. Cefotaxime sodium</td>
<td>antibacterial</td>
<td>99.62% (±4.81%)</td>
</tr>
<tr>
<td>38. Chloramphenicol</td>
<td>antibacterial</td>
<td>99.36% (±1.74%)</td>
</tr>
<tr>
<td>39. Oxolinic acid</td>
<td>antimicrobial</td>
<td>98.76% (±3.39%)</td>
</tr>
<tr>
<td>40. Amoxicillin</td>
<td>antibacterial</td>
<td>98.74% (±3.23%)</td>
</tr>
<tr>
<td>41. Cefazolin sodium</td>
<td>antibacterial</td>
<td>98.44% (±1.26%)</td>
</tr>
<tr>
<td>42. Pyrithione zinc</td>
<td>antibacterial; antifungal</td>
<td>98.27% (±5.00%)</td>
</tr>
<tr>
<td>43. Cephalothin sodium</td>
<td>antibacterial</td>
<td>97.57% (±5.34%)</td>
</tr>
<tr>
<td>44. Moxalactam disodium</td>
<td>antibacterial</td>
<td>97.48% (±4.70%)</td>
</tr>
<tr>
<td>45. Piperacillin sodium</td>
<td>antibacterial</td>
<td>97.37% (±3.75%)</td>
</tr>
<tr>
<td>46. Chlorhexidine</td>
<td>topical antibacterial</td>
<td>97.23% (±2.60%)</td>
</tr>
<tr>
<td>47. Demeclocycline</td>
<td>antibacterial</td>
<td>94.96% (±8.16%)</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 6.1 – Continued

<table>
<thead>
<tr>
<th>Compound name</th>
<th>Functional classification</th>
<th>Antibiotic effectiveness $^a, ^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>48. Erythromycin</td>
<td>antibacterial</td>
<td>94.77% (±2.13%)</td>
</tr>
<tr>
<td>49. Cinoxacin</td>
<td>antibacterial</td>
<td>93.86% (±4.32%)</td>
</tr>
<tr>
<td>50. Penicillin g potassium</td>
<td>antibiotic</td>
<td>91.30% (±2.84%)</td>
</tr>
<tr>
<td>51. Fluorouracil</td>
<td>antineoplastic</td>
<td>89.50% (±0.59%)</td>
</tr>
<tr>
<td>52. Erythromycin ethylsuccinate</td>
<td>antibacterial</td>
<td>83.04% (±15.13%)</td>
</tr>
<tr>
<td>53. Cetylpyridinium chloride</td>
<td>topical anti-infective</td>
<td>78.99% (±5.13%)</td>
</tr>
<tr>
<td>54. Thiamphenicol</td>
<td>antibacterial</td>
<td>71.83% (±2.69%)</td>
</tr>
<tr>
<td>55. Benzalkonium chloride</td>
<td>topical anti-infective</td>
<td>60.32% (±55.55%)</td>
</tr>
<tr>
<td>56. Pipemidic acid</td>
<td>antimicrobial</td>
<td>59.08% (±59.54%)</td>
</tr>
<tr>
<td>57. Floxuridine</td>
<td>antineoplastic</td>
<td>57.74% (±61.87%)</td>
</tr>
<tr>
<td>58. Metampicillin sodium</td>
<td>antimicrobial</td>
<td>52.90% (±71.02%)</td>
</tr>
</tbody>
</table>

$^a$ The percentage of motility inhibition for each compound, as normalized by comparison with motile and non-motile *S. typhimurium* controls on each plate.

$^b$ Only compounds with at least 50% inhibition of the motility are listed.
References

CHAPTER VII

RATIONAL DESIGN OF METAL BINDING SITES AND TESTING THE FEASIBILITY OF CATALYTIC SITES IN FLAGELLIN

Introduction

Computer simulations play an increasingly important role in our understanding of protein folding, stability, activity, and the specificity of protein-ligand interactions. The methodologies being developed play a significant role in understanding the basics of protein architecture as well as drug development. Design methods are being developed which can be used to rationally modify the structure and function of a protein. A great challenge today is to rationally design specific properties for proteins that can be experimentally put to test. Significant progress has been made at enhancing the catalytic properties of existing enzymes; however, the design of proteins with novel catalytic properties has met with relatively limited success. These are being applied to building proteins with novel metal centers and the rational manipulation of their chemical reactivity.

Several software programs such as Rosetta, ORBIT etc. have been successfully put to use in designing proteins with novel functions. We believe that flagellin protein can play a potential role in enhancing the understanding of the basics of protein catalytic functions that can be tested experimentally with relative ease. The current study describes a preliminary engineering of metal binding sites and catalytic sites to modify flagellin, a structural protein, into a functional protein.
Flagella, due to their large size, abundance on the cell surface, and ease of purification, can serve as a model for protein engineering.

Carbonic anhydrase (CA) are metalloenzymes that catalyze the reversible reaction of the conversion of carbon dioxide to bicarbonate (Equation 1)\(^9\),\(^10\),\(^11\),\(^12\),\(^13\). There are 5 known CA structural families, the α, β, γ-classes, and the recently discovered δ and ε classes\(^9\). The primary activity of the enzyme in humans is to maintain the acid-base balance in blood and other tissues and to help transport carbon dioxide out of tissues.

\[
\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+ \quad \text{(Equation 1)}
\]

The human CA II crystal structure was determined by Eriksson et al\(^14\) (PDB 1CA2). The active site of the enzyme consists of the three histidine residues coordinately bound to the Zn\(^{2+}\) metal ion. The zinc ion is ligated to three histidyl residues and one water molecule in a nearly tetrahedral geometry. The neighboring residue Glu106 is deprotonated and the Oε1 atom of Thr-199 donates its proton to the Oε1 atom of Glu-106 and can function as a hydrogen bond acceptor only in additional hydrogen bonds. As a preliminary study we have only considered the three active site histidine residues for the design of metal binding regions in the flagellin protein. The different parameters for the active site of the CA II histidine residues are given in Table 1.

The structure of Salmonella typhimurium flagellin protein (FliC) is described in the Chapter I in detail. We have designed metal binding sites with 3 histidines, similar to the CA II site, in three different regions of the flagellin protein, labeled Site...
I, Site II and Site III (Figure 7.1). Site I is located in an outer loop region of the D2b domain. Site II is located at the interface of D2 and D3 domains, and Site III is located at the interface of the D1 and D2b domain (Figure 7.1). These locations were predicted to have active site-like features by the Computed Atlas of Surface Topography of proteins (CASTp) software server http://sts.bioengr.uic.edu/castp/, which analyzes protein structures for structural pockets and cavities that are often potential binding sites and active sites of proteins. As a preliminary experiment, two of the three potential sites were individually constructed in FliC by site-directed mutagenesis, (Site II and Site III) and tested for metal binding ability with the transition metal ion, Co$^{2+}$ and for catalytic activity with both Zn$^{2+}$ and Co$^{2+}$ ions.

Figure 7.1 Location of the three rationally designed metal binding sites in the flagellin (FliC) structure (PDB 1UCU). Site I is present in the loop region of D2b domain, Site II is located at the interface of D2 and D3 domains and Site III is located at the interface of the D1 and D2b domain.
Table 7.1 Parameters used for the construction of metal binding sites in FliC protein.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Distances between Zn$^{2+}$ ion-coordinated N atoms in histidine imidazole sidechains</th>
<th>Side Chain Angles</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA II parameters (PDB 1CA2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His94 NE2 - His19 ND1</td>
<td>3.27 Å</td>
<td>His119: 30.51 Degrees</td>
</tr>
<tr>
<td>His94 NE2 - His96 NE2</td>
<td>3.31 Å</td>
<td>His94: 35.30 Degrees</td>
</tr>
<tr>
<td>His96 NE2 - His119 ND1</td>
<td>3.11 Å</td>
<td>His96: 35.88 Degrees</td>
</tr>
<tr>
<td><strong>FliC Site I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His396 NE2 - His393 NE1</td>
<td>3.00 Å</td>
<td>His396: 32.29 Degrees</td>
</tr>
<tr>
<td>His396 NE2 - His388 NE2</td>
<td>3.18 Å</td>
<td>His388: 33.19 Degrees</td>
</tr>
<tr>
<td>His393 NE1 - His388 NE2</td>
<td>4.00 Å</td>
<td>His393: 32.27 Degrees</td>
</tr>
<tr>
<td><strong>FliC Site II</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His190 ND2 - His282 NE1</td>
<td>2.71 Å</td>
<td>His190: 32.65 Degrees</td>
</tr>
<tr>
<td>His190 ND2 - His222 NE1</td>
<td>4.19 Å</td>
<td>His282: 32.66 Degrees</td>
</tr>
<tr>
<td>His282 NE1 - His222 NE1</td>
<td>4.82 Å</td>
<td>His222: 32.29 Degrees</td>
</tr>
<tr>
<td><strong>FliC Site III</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His116 NE2 - His120 ND1</td>
<td>2.45 Å</td>
<td>His116: 32.65 Degrees</td>
</tr>
<tr>
<td>His116 NE2 - His384 NE2</td>
<td>3.83 Å</td>
<td>His384: 32.65 Degrees</td>
</tr>
<tr>
<td>His384 NE2 - His120 ND1</td>
<td>3.97 Å</td>
<td>His120: 32.66 Degrees</td>
</tr>
</tbody>
</table>
Materials and Methods

Bacterial strains

*E. coli* XL-1 Blue electrocompetent cells (Stratagene, La Jolla, CA) were used for the transformation of the mutant plasmids. These cells have a transformation efficiency greater than $1 \times 10^{10}$ transformants/µg of DNA, are tetracycline resistant, endonuclease (*endA*) deficient, which greatly improves the quality of miniprep DNA and are recombination (*recA*) deficient, improving insert stability. *S. typhimurium* strain SJW134 (*Δ*fliC and ΔfljB), 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.\(^{17,18}\) This strain is non-motile, unless a functional flagellin gene is introduced, e.g., on an expression plasmid.\(^{17}\) Thus, it serves as an ideal strain to screen engineered or mutated *fliC* genes, present in the introduced plasmids, for functional flagellin protein expression, folding, export and flagellar fiber assembly, as indicated by a simple agar microbial motility assay. *S. typhimurium* strain JR501, a stable restriction-deficient (*r*) , methylation-proficient (*m*) *galE* strain,\(^{19,20}\) was used to convert *E. coli* grown plasmids to *S. typhimurium* compatibility. Electrocompetent cells of *S. typhimurium* strains JR501 and SJW134 were prepared using standard lab protocols\(^{21,22}\). The pTH890 plasmid is a derivative of pTrc99A plasmid with the *S. typhimurium fliC* phase 1 flagellin gene cloned into the XbaI/HindIII site (Figure 7.2). The pTH890 plasmid was kindly provided by the late Dr. Robert Macnab (Yale University). This plasmid was used for the mutagenesis.
Design of the metal binding sites

The metal binding sites were designed by using the PDB structure of CA II 1CA2 as a template. The metal coordinating histidine residues have the following coordinates. Distances between Zn$^{2+}$ ion-coordinated N atoms in histidine imidazole side chains are as follows; His94 NE2 - His119 ND1 atoms: 3.27 Å; His94 NE2 - His96 NE2 atoms: 3.31 Å; His96 NE2 - His119 ND1 atoms: 3.11 Å. Distances between Zn$^{2+}$ ion and Zn$^{2+}$ ion-coordinated N atoms in histidine imidazole side chains are as follows; Zn$^{2+}$ - His94 NE2 atoms: 1.99 Å; Zn$^{2+}$ -His96 NE2 atoms: 2.10 Å; Zn$^{2+}$ -His119 ND1 atoms: 1.91 Å. The side chain angles of CA II; His119: 30.51 degrees; His94: 35.30 Degrees; His96: 35.88 degrees (Table 7.1). Using this structural data, we chose three possible sites were designed in the FliC protein. The SwissPDBViewer software version 3.7 was used to manually mutate different residues in the predicted regions to histidine and select the side chain angles with least steric hindrance for the design of the sites.

Site-directed mutagenesis

Site-directed mutagenesis of pTH890 plasmid (1 μl PfuTurbo® DNA polymerase, 50 ng of pTH890 plasmid DNA, 5 μl 10x PCR buffer, 1 μl of 2.5 mM dNTP mix, 1.25 μl of 100 ng/μl forward and reverse primers and 36 μl of DI H$_2$O) was performed using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA).

Please refer to the Table 2 for the primers.
Figure 7.2 Pictorial depiction of three rationally designed metal binding sites in the FliC protein. Please refer to the Table 2 for the description of the parameters for each metal binding site.
Table 7.2 Mutagenesis primers.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Primer Sequence</th>
</tr>
</thead>
</table>
| Q393H    | 5' CACAACTTTAAAGCACACCCCTGATCTGGCCGGAAG 3'  
          | 5' CTTCCGCCAGATCGGGTGCTTTAAAGTTTG 3' |
| L396H    | 5' CTTAAAGCCACAGGCTGATCATGCCGGAAGCGGCTGCTACAAC 3'  
          | 5' GTTGTAGCAGCGCGCTTCCGATGACGGCTGCTTTAAAG 3' |
| Y190H    | 5' GCTGCAACTGTTACAGGATCGACTACCGATATG 3'  
          | 5' GCAATCGTAGTATCGGCTGCTTTAAAGCTGCTGACTGC 3' |
| F222H    | 5' GATGGCGATTTTAAACATGTATGAGATACGACTGG 3'  
          | 5' CCAGTCATATCATCTATTTTTAAATCGCCATC 3' |
| Q282H    | 5' GAGGATGTGAAAATGTAACACGTTGCAAATGCTG 3'  
          | 5' CCAGATTTGCAAACGTCATACATTTTACATCTC 3' |
| T116H    | 5' CTCCATCCAGGCTGAAATCCACCACCCGCTGAACGAAATC 3'  
          | 5' GATTTCCAGGCGCTGGATTTCATGCTGAGGATT 3' |
| N120H    | 5' CTGAAATCACCACCCGCTGATGAAATCGACCGGTATCCG 3'  
          | 5' CGGATACACGCGCTGATTCATCGAGCGCTGGGTGTCCAG 3' |
| K384H    | 5' GTTAAAACTTACGCTGCAAGTCACTGACCACGGAAGGTCAACAAC 3'  
          | 5' GTTGTGACCTCCGCTGACTGACCGTGAAAGTTTACC 3' |
Partial purification of fibers and SDS-PAGE

Engineered active-site flagellins, in the self-assembled form of flagella fibers, were partially purified by shearing. The successful isolation of flagellins by this method indicated that the mutated flagellin subunits were successfully exported and assembled to form functional flagella fibers. Bacterial cultures were inoculated into 15 ml LB broth with 15 μl of 100 mg/ml ampicillin and allowed to grow overnight at 37 °C. 15 ml of the bacterial cultures were transferred to a 15 ml polypropylene tube and vortexed for 30 seconds to 1 minute. The cells were pelleted at 6000 x g and the supernatant was transferred to 15 ml (30 kDa cutoff) Amicon Ultra centrifugal filtration tubes (Millipore Corp., Billerica, MA) for concentrating the flagella fibers. The concentrates were resolved by 10% SDS-PAGE. The concentration of the protein was standardized at 2 mg/ml.

Dialysis of the mutants

The partially purified flagella fibers were dialyzed overnight with EDTA buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to remove any preexisting bound or unbound metal atoms. The overnight dialysis was performed with a buffer volume of 50 times the protein solution volume, using Slide-A-Lyzer cassette™ 30 kDa molecular weight cutoff (MWCO) dialysis cassettes (Pierce, Rockford, IL). The dialysis step was repeated with a buffer change for 6 hrs. After this step the fibers were dialyzed 24 hrs with 25 mM MOPS (pH 7.5) containing 150 mM KCl and 2 mM CoCl₂. This dialysis step was followed by dialysis with metal-free MOPS KCl buffer to remove unbound

241

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
metal ions for 4 hrs. After dialysis, the protein samples were scanned for absorbance with a Beckman DU-530 spectrophotometer from wavelengths 200-800 nm.

Enzyme assay

The catalytic esterase activity of the engineered flagellins was monitored by the hydrolysis of the nonphysiological but commercially available ester, 4-nitrophenyl acetate (4-NPA) (Sigma-Aldrich, St. Louis, MO). The products of the hydrolysis reaction are acetate and nitrophenolate, which ionizes to give a bright yellow anion that is detected by measuring its absorbance at 348 to 400 nm.

Results and Discussion

The parameters of designed metal binding sites are listed in Table 1 and described in Figure 2. Spectroscopic scans were taken for the native flagella fibers and the engineered FliC Site II and FliC Site III fibers. No differences were observed in the absorbance wavelengths and values in the pre- and post-dialyzed samples. This preliminary result suggests that the metal ions did not coordinate with the histidine residues. The enzyme assay performed with the cobalt chloride-dialyzed native and mutated fibers also yielded a negative result for increased esterase activity.

The above results suggest that the cobalt ion did not coordinate with designed histidine residues, indicating a failure of this simple preliminary approach to active site design in a protein. The distances and geometries of the histidine metal ligands that were anticipated to form may not have been attained. In a related example, a histidine residue was predicted to participate in transition metal binding in a rationally
designed "histidine-patch" thioredoxin mutant. However, this histidine residue, His6, failed to do so for reasons that are not well-understood.

One of the major impediments for predicting the conformation of protein structures is the limitation in the computational power. The side chain structures were designed using various available software that only considers the rigid crystal structures. The amino acid side chains in the solution state can assume numerous conformations that may not be accurately described in the rotomers available in the free software used in this study. A more detailed rational design approach to this problem, involving the use of protein design software, might yield a functional metal binding site. Alternatively, techniques of random mutagenesis and screening, common to "directed evolution" methods of enzyme engineering, might also be another pathway to achieving the creation of a functional flagellin protein with metal binding and catalytic activity.

References


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER VIII

FUTURE STUDIES

Study of the conformational state of *Salmonella typhimurium* flagellin protein during type III export and assembly into flagella

The specific aim of this study is to determine the extent of folding and the conformational state(s) of the *Salmonella typhimurium* phase 1 flagellin protein, both *in vivo*, after complete polypeptide synthesis and prior to export and assembly, and while it is being transported through the central 20 Å diameter channel of the flagella fiber. The phase 1 flagellin protein does not have any cysteine (Cys) or tryptophan (Trp) residues. These two residues can be introduced by site-directed mutagenesis of the corresponding pTH890 *fliC* expression plasmid. A Trp residue can be used as an intrinsic fluorescent probe (Trp), or used as a fluorescence energy donor, along with another introduced fluorescence acceptor, e.g. IADANS dye, in Forster fluorescence resonance energy transfer (FRET) studies of protein folding.

Cys residues can also be introduced that can either spontaneously cross-link to form covalent bonds in oxidizing environments or provide a uniquely reactive point of attachment for synthetic fluorescent dye molecules, for use as acceptor probes in fluorescence resonance energy transfer (FRET) studies on the folded state of the protein. Therefore, one or more of Cys and/or Trp residues can be introduced into any desired location of the flagellin protein. One proposed set of experiments will involve introducing pairs of Cys residues into various locations throughout the four D0, D1, D2 and D3 domains. Assuming that they are appropriately positioned next to each
other in the folded or unfolded flagellin structure, these pairs of Cys residues should spontaneously form covalently bonded disulfide bridges in the oxidizing environment of the bacterial periplasmic space. Previous in vitro studies have shown that the flagellin proteins are present in a highly disordered state as monomers and assume a globular shape with α-helical coiled-coil conformation at the N- and C-termini, upon polymerization. These interactions can be further studied by designing disulfide bridges in different domains of the flagellin protein and testing the resulting variants for successful export and assembly. The hypothesis is that disulfide bonds that are formed prior to export.

The disulfide bridges would be designed by using Disulfide by Design (need to cite ref for this software), a software application for the rational design of disulfide bonds in proteins, created by Dr. Alan Dombkowski, Wayne State University (Detroit, MI). The algorithm takes a Protein Data Bank (PDB)-format protein structure (e.g., determined by X-ray diffraction or NMR) as the input. Then all possible pairs of amino acid residues are rapidly assessed for proximity and geometry consistent with disulfide formation, assuming that the residues were mutated to cysteines. The output displays all the possible residue pairs meeting the appropriate criteria. The Disulfide by Design algorithm has been successfully used for disulfide engineering to study the protein dynamics and interactions. The PDB structure of S. typhimurium flagellin, 1UCU, was used as the input for the Disulfide by Design program. The algorithm revealed 36 potential amino acid pairs, that when substituted with Cys residues, have a strong potential for forming disulfide bonds. As an example
of the proposed preliminary experiments, five different substitution pairs were identified in the set of 36 disulfide pairs that are located in different domains of the protein. The selected substitutions are L17C and L479C (D0 domain); A50C and I453C (D1 domain); T171C and A400C (D2a domain); Y178C and Y311C (D2b domain) and G189C and G340C (D3 domain).

**Self-assembly of GFP-flagellin fusion proteins *in vitro***

As described in Chapter IV, approximately 100 flagellin-GFP fusion proteins, with GFP as an internal fusion, were successfully generated by transposon insertion, as identified by fluorescence screening of colonies for green fluorescence and confirmed by colony PCR screening. However, none of the GFP-flagellin variants identified in this study yielded functional flagella, as indicated by non-functional motility assays. Although the negative results of functional screening for motility were not based on an exhaustive study, this study indicates that GFP does not readily serve as a good model protein for insertion into flagellin proteins to enable the creation of a functional flagella nanotube with genetically encoded fluorescence. The limiting factor in this case could be the pore size of the flagella export apparatus. Purified monomeric flagellin proteins from mesophilic bacteria can rapidly polymerize *in vitro*, especially if incubated with small flagella fiber seed fragments (need in vitro polymerization refs). This hurdle could potentially be overcome by isolating the flagellin-GFP fusion proteins from the cells and polymerizing them *in vitro*. The proteins can be purified after over-expression in *Salmonella* or *E. coli* cells and standard flagellin polymerization methods can be used to assemble the fibers. Fiber
formation can then be verified using transmission electron microscopy (TEM), light scattering and, possibly, techniques of gel filtration chromatography with a large pore size resin.

**Conformational study of various flagellin deletion variants using circular dichroism and fluorescence spectroscopy**

In the Chapter II, a description is given of the creation of an internal deletion variant library of *Salmonella typhimurium* flagellin and the predictive modeling of some of the structures. Although the models predict the possible conformational states of each deletion variant, they need to be experimentally verified. Here, we propose the study of at least five different variants using circular dichroism spectroscopy (CD) to examine secondary structure formation and content and fluorescence spectroscopy to examine global folding of the flagellin structures. The variant proteins can also be studied for their chemical denaturation and thermal melting properties using the above mentioned studies; fluorescent probes could also be attached, using the previously described methods of site-directed mutagenesis to introduce reactive Cys residues in the flagellin variant sequences.

**Engineering of catalytic sites in flagellin protein**

Even though our preliminary study of engineering catalytic sites was not successful, this study can be further explored by designing sites by better algorithms and testing more and diverse active sites. Also, a directed evolution approach can be used to generate catalytic flagellin protein.
**Insertion of single domain antibodies in flagellin hypervariable region**

In our previous study we have demonstrated that flagellin hypervariable region can be removed to a large extent and created shorter functional flagellin variants. These short flagellins can be used to insert single molecule antibodies that can have implications in diagnostics, vaccine research etc.
Appendix A
Recombinant DNA Biosafety Approval
Western Michigan University

Recombinant DNA Biosafety Committee

Registration for Recombinant Research - 2007

This form must be submitted for all research involving recombinant DNA molecules. Renewal of approval is required annually.

In this form, Guidelines means: Guidelines for Research Involving Recombinant DNA Molecules (NIH Guidelines).

General Registration Information

Principal Investigator: Brian C. Tripp, Ph.D.
Office Phone: 387-4186 Laboratory Phone: 387-1955 Department: BIOS
Electronic Mail Address: brian.tripp@wmich.edu
Office Address: Room 3435, Building: Wood Hall

Laboratory Where Research is to be Conducted:
   Building: Haenelke Hall and WMU Engineering Campus

Project Title: Engineering and Display of Enzymes and Proteins on Bacterial Flagella Fibers

Project Start Date: 04/01/02

Current status (check one):
   □ Initiated Date:
   □ Will be initiated Date:
   □ Continuing (no changes) Current Project Number: 06-04-B7a
   □ Continuing (modifications) Current Project Number: 06
   □ Will not be initiated or will be discontinued
   □ Completed Date:

WMU Proposal Tracking Number (if applicable):

Revised 05/06 WMU-RDRC
All other copies obsolete.
A. Pathogenic Risk Group 2 Salmonella Strains.

The experiments covered under this registration form involve the use of several strains of the Risk Group 2 (RG2) bacterial agent, *Salmonella typhimurium*. *Salmonella typhimurium* is a known bacterial pathogen known to cause food poisoning when ingested and is widely present in the environment, sometimes associated with reptiles, birds and amphibians.

The section entitled Appendix B-II - Risk Group 2 (RG2) Agents of the April 2002 version of the NIH Guidelines gives the definition of RG2 agents: “RG2 agents are associated with human disease which is rarely serious and for which preventive or therapeutic interventions are often available.”

Appendix B-II of the NIH Guidelines lists Risk Group 2 (RG2) - Bacterial Agents including *Chlamydia*, which includes the statement: “*Salmonella* including *S. arizonae*, *S. cholerasuis*, *S. enteritidis*, *S. gallinarum-pulmonum*, *S. meleagridis*, *S. paratyphi A, B, C, S. typhi*, *S. typhimurium*.

B. Non-Pathogenic E. coli Strains.

Other experiments in these recombinant DNA protocols involve the use of standard, non-pathogenic E. coli K-12 Host-plasmid systems such as DH5-alpha and BL21 DE3 strains, which are considered exempt from the NIH Guidelines. As stated in the NIH Guidelines, Appendix C-II: Escherichia coli K-12 Host-Vector Systems: “Experiments which use *Escherichia* coli K-12 host-vector systems, with the exception of those experiments listed in Appendix C-II-A, are exempt from the NIH Guidelines provided that: (1) the *Escherichia* coli host does not contain conjugation proficient plasmids or generalized transducing phages; or (2) lambda or lambdoid or P1 bacteriophages or non-conjugative plasmids (see Appendix C-VII, Footnotes and References of Appendix C, Footnotes and References of Appendix C) shall be used as vectors. However, experiments involving the insertion into *Escherichia* coli K-12 of DNA from prokaryotes that exchange genetic information (see Appendix C-VII, Footnotes and References of Appendix C) with *Escherichia* coli may be performed with any *Escherichia* coli K-12 vector (e.g., conjugative plasmid). When a non-conjugative vector is used, the *Escherichia* coli K-12 host may contain conjugation-proficient plasmids either autonomous or integrated, or generalized transducing phages. For these exempt laboratory experiments, Biosafety Level (BL) 1 physical containment conditions are recommended. For large-scale fermentation experiments, the appropriate physical containment conditions need be no greater than those for the host organism unmodified by recombinant DNA techniques; the Institutional Biosafety Committee can specify higher containment if deemed necessary.”

C. Bacterial Cell Culture Volumes.

The experimental procedures described in this protocol involve bacterial cell culture of exempt or potential RG1 *E. coli* K-12 strains and RG2 *Salmonella typhimurium* cells in volumes ranging from 0.1 ml to 5 liters. As noted in Appendix C-II-A, Exceptions of the Guidelines, “The following categories are not exempt from the NIH Guidelines: ...(v) large-scale experiments (e.g., more than 10 liters of culture), ”. Thus, the cell culture procedures in this protocol do not use volumes of cell culture that exceed more than 50% of the volumes considered to be in the category of “large-scale” by the NIH Guidelines.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Description of Recombinant DNA Experiments

1. Summary of project (1 or 2 paragraphs). Include purpose and manner in which recombinant DNA will be used in the project. Please define acronyms and abbreviations.

I am continuing a research project that involves the investigation of the export and folding pathways of bacterial flagella fibers, with the goal of displaying foreign enzymes and proteins as genetic fusions to flagellin proteins attached to the surface of bacteria, for possible use as sensors and in nanotechnology applications. This flagellin-oriented research also involves a collaboration with Dr. Subra Muralidharan, WMU Chemistry Dept., on the construction of novel nanomaterials using flagellin proteins as a novel nanomaterial fiber to bind or induce precipitation of other inorganic nanomaterials. Some gene fusion experiments involve the well-characterized green fluorescent protein originally cloned from Aequorea victoria, Pacific jellyfish, which will be inserted in various ways into Salmonella, E. coli, and Aquophilus pyrophilus flagellin proteins. Additional cloning experiments will involve the same techniques and approaches applied to the extracellular display of human carbonic anhydrase II on bacterial flagellin. Carbonic anhydrase II is a well-characterized enzyme with easily detected catalytic activity. Another area of current research involves the engineering and optimization of a particular structural class of bacterial left-handed beta helical acyltransferase and carbonic anhydrase enzymes to perform novel catalytic functions, and the catalytic engineering of flagellins to a) mimic the carbonic anhydrase active site and catalytic function and b) contain an internal fusion of the human carbonic anhydrase enzyme. These areas of research will involve investigating the effect of various mutations, foreign gene insertions and gene rearrangements on function of the corresponding proteins. The catalytic activity of mutated enzymes, level of recombinant protein production in E. coli, extracellular protein export, and the flagella-mediated bacterial motility of E. coli and Salmonella typhimurium strain will be studied as a function of mutations in the corresponding genetic DNA. Thus, recombinant DNA will be used as a template for the introduction of amino acid and polypeptide substitutions into the corresponding proteins. Genes encoding various proteins and enzymes will be individually subjected to random mutations, site-directed mutagenesis, gene insertion and gene shuffling. Standard molecular biology techniques, such as polymerase chain reaction (PCR) will be used to introduce mutations in the DNA, while other cutting and ligation procedures will be performed with restriction enzymes and T4 ligase enzyme. Molecular biology and cell culture procedures are performed in the P's laboratory in 4007 and 4009 Haenick Hall. Other physiological experiments involve the use of a laser tweezer-optical microscope instrument to measure the motility, i.e. ability to swim, of E. coli and Salmonella bacteria. This instrument is installed as part of a W. M. Keck Nanoscience facility in room A110 in the WMU Engineering Campus building.

An interrelated enzymology research project that investigates the catalytic function and screening small molecules for inhibition of mammalian, bacterial and archaeal carbonic anhydrases and several structurally related left-handed beta helical acyltransferase enzymes. Cloning experiments will involve structure-function studies of the relationships between catalytic rate, substrate recognition and binding by various types of inhibitors. Initial studies involve human carbonic anhydrase II, which is a well-characterized enzyme with easily detected catalytic activity. Another area of my current research involves the engineering and optimization of alpha and gamma-class class carbonic anhydrase enzymes to perform novel catalytic functions. These areas of research will involve investigating the effect of various mutations, foreign gene insertions and gene rearrangements on function of the corresponding proteins. The catalytic activity of wild type and mutated enzymes, inhibition by small molecules, and the level of recombinant proteins production in E. coli will be studied as a function of mutations in the corresponding genetic DNA. Thus, recombinant DNA will be used as a template for the introduction of amino acid and polypeptide substitutions into the corresponding proteins. Genes encoding various carbonic anhydrase and related enzymes may individually be subjected to random mutations, site-directed mutagenesis, gene insertion and gene shuffling. Standard molecular biology techniques, such as polymerase chain reaction (PCR) will be used to introduce mutations in the DNA, while other cutting and ligation procedures will be performed with restriction enzymes and T4 ligase enzyme.

2. Host strain(s) used, including genus, species, parent strains, and class of each agent.

Only prokaryotic bacterial microorganisms will be used in this research. E. coli and Salmonella enterica serovar Typhimurium LT2 (Salmonella typhimurium) Host strains to be used include: E. coli K-12 strains (exempt class) XL-1 Red, XL-1 Blue, DH5-alpha, BL21(DE3)Gold (Stratagene, La Jolla, CA), TOP10, C3H, GB226, BL21(DE3)star, BL21(DE3)pLYsS, BL21(DE3)pLyzE, BL21(DE3)M4 (Introngen, Carlsbad, CA), NEB 5-alpha (New England Biolabs, Beverly, MA). Three Salmonella typhimurium strains (Class III D, risk group 2BL-2) will also be used: SJW1103 (wild-type for flagellar-mediated motility and chemotaxis), SJW134, which has the two genes for flagellin deleted (MIC and M69) and is therefore non-motile unless foreign flagellin genes are added via plasmid transformation, and (RS501 (rav)) which is used to convert E. coli grown plasmids for compatibility in other Salmonella strains. These bacterial strains will be used for host-vector plasmid production and also for bacterial protein expression.

Revised 05/06. WMU RMB: 4
All other copies obsolete
3. Source(s) and nature of inserted DNA sequences. Include size, gene name(s) and function of gene(s), and sequence(s), if known.

The PI, Dr. Trip, has a number of genes encoded in plasmid vectors suitable for transformation into E. coli or Salmonella bacteria. These include:

1. The Aerobacter aerogenes green fluorescent protein (gene size: 720 b.p.), an autofluorescent protein used as a biological prohibitory. It was encoded in the plasmid purchased from Bio-Rad, Hercules, CA.

2. Three plasmids encoding wild-type Salmonella IC flagellin gene pTH10 (gene size: 1491 b.p.), pTH890 (gene size: 1501 b.p.), and pMM1501 (gene size: 1501 b.p.), were obtained from Prof. Robert Macnab at Yale University, New Haven, CT.

3. Four plasmids containing green fluorescent protein variants were obtained from Doug Sheridan & Thomas R. Hughes, Yale University: pBU246, pBU263, pBU328, pBU237 (GFP gene size 720 b.p.).

4. Two plasmids containing fragments of the marine hyperthermophilic bacterium Aquifex pyrophilus flagellin gene (approx. 1500 b.p., pRU1650 and pRU1651) were obtained from Prof. Dr. Rudiger Schmitz at the University of Regensburg, Germany.

5. One plasmid, pUC8-52, containing the full-length gene encoding the biotin-binding protein, streptavidin, cloned from Streptomyces avidius, was obtained from Dr. Takashi Sano, Harvard Medical School.


7. Five plasmids encoding the left-handed helical enzymes: gamma-class carbonic anhydrase (CAm) from the archaeon Methanosarcina thermophila (gene size: 639 DNA base pairs), E. coli UDP-N-acetylglucosamine 3-O-acyltransferase (LPX, gene size: 788 b.p.), Mycobacterium bovis tetrahydrodicarboxolinate N-acetyltransferase (DapA), gene size: 872 b.p.), stearoyl-CoA desaturase from Endothiaparinga tarda (ValD), gene size: 624 b.p.), and N-acetylphosphatase: 1-phosphate phosphomonoesterase (GimU), gene size: 1368 b.p.). The latter four genes were obtained from Prof. Steven Rocketts at the Albert Einstein College of Medicine, Bronx, NY.

4. Vector(s) to be used. Include source.

Vectors to be used include pUC18, pUC19, pET15b, pET28a,b,c, pET34, pET35b, pET37b, pET38b (purchased from Newagen, Bedfordin, W), pLEX, pLEX-incZ, pRT1 X (purchased from Invitrogen, Carlsbad, CA), pCGO (pBAD-GFPuv), (purchased from Bio-Rad, Hercules, CA), pACA (p7 promoter vector obtained from Prof. Carol Fierke, Dept. of Chemistry, University of Michigan), three vectors obtained from Prof. Robert Macnab, Dept. of Molecular Biophysics and Biochemistry, Yale University: pTH10 (IC flagellin gene in pTH90), pTH890 (N-region IC gene in pTH90 vector) and pMM1501 (N-region IC gene in pTH10b vector). Four vectors obtained from Doug Sheridan & Thomas R. Hughes, Yale University: pBU246, pBU263, pBU328, pBU237. Two vectors pRU1650 and pRU1651 (Aquificx pyrophilus Flok flagellin gene fragments) were obtained from Prof. Dr. Rudiger Schmitz at the University of Regensburg, Germany. These gene fragments have been subcloned into the pUC and pET18b expression plasmids for use in E. coli. Plasmid vectors containing the other eleven human carbonic anhydrase enzymes are either pCMV/SPORT6, pOTB7 or pBlueScriptK and were obtained from Open Biosystems, Huntsville, AL.

5. Will an attempt be made to obtain expression of a foreign gene? ☑ Yes ☐ No

If so, what protein will be produced?

Eubacterial E. coli, Salmonella typhimurium, and Aquifex pyrophilus wild-type, partially deleted flagellin proteins, and hybrids of flagellin and green fluorescent protein, hybrids of E. coli threonin and flagellin (FMTrX), hybrids of streptavidin and flagellin and hybrids of flagellin and mammalian (human) carbonic anhydrase enzymes will be expressed in both E. coli and Salmonella typhimurium. Streptavidin, arabin carbonic anhydrase and acetyltransferase enzymes including Cam, LPX, DapA, DapD, GimU, and ValD will also be expressed in soluble proteins in E. coli. CA II and other previously described isoforms of human carbonic anhydrases CA I, CA III, CA IV, CA VA, CA VII, CA IX, CA XI, CA XII, CA XIII, CA XIV will be cloned into pET vectors and expressed in E. coli for use as controls in flagellin fusion-protein experiments and in catalytic engineering of flagellin.

Statements of Certification

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
I certify that I have read and understood the Western Michigan University Policy for Recombinant DNA Biosafety including the description of my role and responsibilities as Principal Investigator on this project. I agree to abide by the Guidelines for Research Involving Recombinant DNA Molecules (NIH Guidelines) in conducting all work using recombinant DNA molecules.

Principal Investigator Signature Date

I have read the description of this proposed research involving recombinant DNA molecules. I have determined that the facilities and procedures proposed in this project registration form are adequate for the safe conduct of this research and the safety of other personnel - faculty, staff, and visitors - using the facilities within which this research is to be conducted.

Department Chair/Unit Head Signature Date

Submit the original plus 7 copies to the research compliance coordinator, 251W Walwood Hall. (Be sure to make and keep another copy for your files.) Contact the research compliance coordinator at 387-8293 for additional information.
Recombinant DNA Biosafety Committee

Project Approval Certification

For rDNA Biosafety Committee Use Only

Project Title: Engineering and Display of Enzymes and Proteins on Bacterial Flagella Fibers

Principal Investigator: Brian Tripp

IBC Project Number: 06-BTa

Date Received by the rDNA Biosafety Committee: 11/15/06

☐ Reviewed by the rDNA Biosafety Committee

☒ Approved
☐ Approval not required

Chair of rDNA Biosafety Committee Signature 12/13/06
Appendix B
Letters and Permission
Dear Dr. Enders,
I have a question regarding the copyright permission for the article published in the Journal of Molecular Modeling, titled

"A theoretical model of Aquifex pyrophilus flagellin: implications for its thermostability"
Volume 12, Number 4 / March, 2006
Pages 481-493
Authors V. Raghu Ram Malapaka and Brian C. Tripp

I am the first author of the paper and currently I am writing my Doctoral Dissertation. I will need the copyright permission to use the figures and article as a whole in my Dissertation. If you could please guide me about how to obtain the copyright permission, I will be grateful.

Thanking you,
Raghu Ram V. Malapaka.

---

Dear Mr. Malapaka,
Springer hereby grants you the permission requested. We wish you much success with your PhD thesis and are looking forward to receiving more interesting papers from you.

Yours sincerely,
Peter Enders
Springer
Senior Editor Chemistry

---
Tiergartenstrasse 17 | 69121 Heidelberg | Germany
tel: +49 (0)6221 487-8541, fax: +49 (0)6221 487-8139
peter.enders@springer.com
www.springer.com <http://www.springer.com>
Dear Venkata Raghu Ram Malapaka,

Thank you for your reply. Our grant of permission is below. Where indicated, the word "Copyright" may be replaced with the usual symbol if desired.

Permission is hereby granted for the use requested.

If any of the material in question appears within our work with credit to another source, authorisation from that source must be obtained.

This permission does not include the right to grant others permission to photocopy or otherwise reproduce this material except for versions made by non-profit organisations for use by the blind or handicapped persons.

We are waiving our fee on this occasion.

Proper credit must be given to our publication.

Credit must include the following components:
Title of the Work, Author(s) and/or Editor(s) Name(s), Copyright year. Copyright [symbol] John Wiley & Sons Limited. Reproduced with permission.

Yours sincerely,

Duncan James
Permissions Coordinator
John Wiley & Sons, Ltd.
1 Oldlands Way
Bognor Regis
West Sussex
PO22 9SA
UK
Tel: +44 (0) 1243 843356
Fax: +44 (0) 1243 770620
Email: permreq@wiley.co.uk

Wiley Bicentennial: Knowledge for Generations
1807-2007

Venkata Raghu Ram Malapaka <venkata.malapaka@wmich.edu>
Dear Duncan,

I have the article details below and the figure number.

Article title: "Bacterial Flagella". Copyright (c) 2001 John Wiley & Sons, Ltd

Authors: David Gene Morgan, Shahid Khan

Figure 3. Schematic of the bacterial flagellar base.

Thank you,

Raghu Malapaka.

----- Original Message -----

From: Permission RequestsUK <permreq@wiley.co.uk>

Date: Friday, March 2, 2007 7:42 am

Subject: Re: Copyright permission

> Dear Raghu Malapaka
> 
> Thank you for your enquiry.
> 
> In order to process your request, we require details of the
> article
> containing the figure, and the figure number from within the article.
> 
> Please include this message and the previous correspondence below
> when
> replying.
> 
> Yours sincerely
> 
> Duncan James
> Permissions Coordinator
> John Wiley & Sons, Ltd.
Dear Editor,

I am a PhD student and am in the process of completing my dissertation. My work is primarily based on the eubacterial flagellin structure. I find the Figure 1 of the article in ELS titled "Bacterial Flagella" to be extremely relevant to my thesis. I would like to know if it is possible to get the copyright permission to use the particular figure in my thesis.

Thank you,

Raghu Malapaka.
Re: Copyright Permission [REF:7862491896]

Dear Mr Malapaka,

Thank you for your e-mail.

We wish to advise that you do not need to obtain a copyright permission if you will use the article or a part of the article in your dissertation.

As an author, you retain rights for large number of author uses, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier including "the right to include the article in full or in part in a thesis or dissertation (provided that this is not to be published commercially)."

Please ensure that the reference number remains in the subject line if responding to this e-mail.

Please feel free to contact us with any additional questions or concerns at the e-mail address or telephone numbers shown below.

Yours sincerely,

Micha Magante
Elsevier Customer Support
E-mail: support@elsevier.com

Helping you get published http://www.elsevier.com/authors

Global telephone support is available 24/7:
For The Americas: +1 888 834 7287 (toll-free for US & Canadian customers)
For Asia & Pacific: +61 3 5961 5032
For Europe & rest of the world: +353 61 709 190

If you have any feedback on our service, please e-mail customerfeedback@elsevier.com.

To ensure delivery of our e-mails to your inbox (not bulk or junk folders), please add support@elsevier.com to your address book or safe senders list.

Copyright 2005 Elsevier Limited. All rights reserved.

--- Original Message ---
From: mvranguram@gmaB.com
Sent: 07-Mar-2007 3:58:26
To: authorsupport
Subject: Copyright Permission

A Deletion Variant Study of the Functional Role of the Salmonella Flagellin Hypervariable Domain Region in Motility
Journal of Molecular Biology, Volume 365, Issue 4, 26 January 2007, Pages 1102-1116
Raghu Ram V. Malapaka, Leslie O. Adebayo and Brian C. Tripp

I am the first author of the paper and currently I am writing my Doctoral Dissertation. I will need the copyright permission to use the figures and article as a whole in my Dissertation. If you could please guide me about how to obtain the copyright permission, I will be grateful.

Thanking you,

Raghu Ram V. Malapaka.