ABSTRACT

THE STRUCTURE, EVOLUTION, AND ASSEMBLY MECHANISM OF THE BACTERIOPHAGE TAIL TUBE

Lisa Pell
Ph.D. Thesis, 2010
Department of Biochemistry
University of Toronto

Large multi-component structures play an essential role in many crucial cellular processes. The morphogenetic pathway of the long, non-contractile tail of bacteriophage λ provides a superb paradigm for studying the assembly of macromolecular complexes. This thesis describes the structural and functional characterization of two λ tail proteins, gpU and gpV, with the aim of improving our understanding of phage tail assembly and evolution, while also providing a starting point to answering some of the fundamental questions surrounding the assembly and function of other supramolecular structures.

Tail Terminator Proteins (TrPs) play an essential role in regulating the length of phage tails, and serve as the interaction surface for phage heads. To provide insight into the mechanisms by which TrPs exert their functions, I have determined the X-ray crystal structure of gpU, the TrP from phage λ, in its biologically relevant hexameric state. The gpU hexamer displays several flexible loops that are involved in head and tail binding. By comparing the hexameric crystal structure of gpU to its previously determined NMR solution structure I was able to identify large structural rearrangements in the protein, which are likely induced upon oligomerization. In addition, I have shown that the hexameric structure of gpU is very similar to the structure of a putative TrP from a contractile phage tail even though they display no detectable sequence similarity. This finding implies that the TrPs of non-contractile tailed phages are evolutionarily related to those of contractile-tailed phages.

To determine the mechanism by which tail tubes self-assemble prior to termination, I have determined the NMR solution structure of the N-terminal domain of gpV (gpVN), the protein comprising the major portion of the phage λ tail tube. I found that approximately 30% of
gpV\textsubscript{N} is disordered in solution and that some of these disordered regions are biologically important. Intriguingly, my gpV\textsubscript{N} structure is very similar to a previously solved tail tube protein from a contractile-tailed phage, once again suggesting an evolutionary connection between these two distinct tail types. A remarkable structural similarity is also seen to the hexameric structure of Hcp1, a component of the bacterial type VI secretion system. This finding, coupled with other similarities between phage and type VI secretion proteins support an evolutionary relationship between these systems. Using Hcp1 as a model, I proposed a mechanism for the oligomerization and polymerization of gpV involving several disorder-to-order transitions.

Further supporting the importance of unstructured regions, I have shown that the unstructured linker between the N- and C-terminal domains of gpV is crucial for protein function and that a complete truncation of the C-terminal domain (gpV\textsubscript{C}) results in a 100-fold decrease in activity compared to full-length gpV (gpV\textsubscript{FL}). To provide insight into the role of gpV\textsubscript{C}, I determined its NMR solution structure and showed that it possesses an Ig-like fold, however the function of gpV\textsubscript{C} remains unknown.

Interestingly, the gpV\textsubscript{C} structure revealed the location of two residues that when mutated were previously shown to either abrogate (G222D) or restore (G222D/P227L) function of gpV\textsubscript{FL}. In addition to being inactive, I demonstrated that the G222D mutation also exerts a temperature dependent dominant negative phenotype. My preliminary NMR data suggests that G222D causes gpV\textsubscript{C} to partially unfold and that this destabilized form of the domain interacts with gpV\textsubscript{N} in a region that is likely involved in both oligomerization and hexamer-hexamer interactions. To further our understanding of how these mutations exert their effect, I determined the NMR solution structure of gpV\textsubscript{C}-P227L. My structure reveals that the \(\beta7-\beta8\) region of gpV\textsubscript{C}-P227L is altered compared to gpV\textsubscript{C}-WT and suggests that the conformational changes in gpV\textsubscript{C}-P227L may protect the domain from protein-folding defects induced by the G222D mutation.
I am a wiser person today then when I first arrived at graduate school over 6 years ago, and for this I am thankful to many people: my supervisors, my colleagues, my friends, and my family.

I am grateful to my supervisors, Alan Davidson and Lynne Howell, for giving me the opportunity to work on a collaborative project that, although I didn’t immediately realize it, was perfect for me. I couldn’t have had two better mentors and I am sincerely thankful to both of them for their generosity, enthusiasm, and for having more faith in my abilities than I did myself.

I thank my supervisory committee Emil Pai and Jun Liu for their suggestions and insightful discussions. I am thankful to Dave Smith for patiently teaching me X-ray crystallography, and for keeping me on my toes during lab meetings. I will miss his stories, NMR taunts, and thought provoking questions. I thank Voula Kanelis for teaching me the ins and outs of NMR spectroscopy, lending an ear for the occasional rant, and making me run through blizzards.

I am also grateful to have had the opportunity to work with so many fantastic people in both the Howell and Davidson labs. I am especially thankful to Lia for being an amazing friend and teacher, and I admire her abilities as both a scientist and a writer. I thank Stephanie for her patience during our many “serious” discussions and for partaking in (and encouraging) many midday snack runs. I thank Shao-Yang and his pet llama for being mysterious and wise. I am thankful to Jason, Karen S., Lili, and Joel, for their genuine kindness and for all of the sound advice they have given me. I thank Carrie and Karen D. for their many delicious contributions to the sharing table, and Yura for all of his technical support. I am grateful to Jeff, whose kindness and accomplishments continue to inspire us all. I thank Patrick for being an excellent lab manager who ensured everything ran smoothly.

I am thankful to my parents, Terry and Sylvia, for their encouraging words. They have always been proud of my accomplishments and have stood by my side throughout this journey. Finally, I thank Chris for his patience and calm demeanor, for trying to keep me sane through the ups and downs of my project, and for giving me a good reason to leave the lab every night.
TABLE OF CONTENTS

CHAPTER 1 ........................................................................................................................... 1

INTRODUCTION .................................................................................................................. 1

1.1 Bacteriophages: A General Introduction .................................................................. 2

1.2 Bacteriophage Tails ................................................................................................. 3

1.3 Bacteriophage Evolution ......................................................................................... 4

1.4 Bacteriophage $\lambda$ ............................................................................................... 6

1.5 Wild-Type $\lambda$ Tail Assembly Pathway .................................................................. 9

1.5.1 gpV: The $\lambda$ Major Tail Protein ..................................................................... 13

1.5.1.1 Immunoglobulin-like Domains in Bacteriophage ...................................... 17

1.5.2 gpU: The $\lambda$ Tail Terminator Protein ............................................................... 18

1.6. Aberrant Tail Assembly Reactions ........................................................................ 21

1.7 Regulation of $\lambda$ Tail Assembly ......................................................................... 22

1.7.1 Regulation of Protein Synthesis .......................................................................... 22

1.7.2 Regulating the Order of Assembly ..................................................................... 22

1.7.2.1 Unstructured Regions Regulate Assembly .................................................. 23

1.7.3 Tail Length Regulation ....................................................................................... 27

1.8 Tube-Like Macromolecular Complexes: Structure and Assembly ...................... 29

1.8.1 Tobacco Mosaic Virus ....................................................................................... 29

1.8.2 The Bacterial Flagellum .................................................................................... 30

1.8.3 Bacterial Secretion Systems ............................................................................. 31

1.9 Thesis Objectives ..................................................................................................... 32

1.10 Organization of Thesis .......................................................................................... 33

CHAPTER 2 .......................................................................................................................... 35

THE X-RAY CRYSTAL STRUCTURE OF THE PHAGE $\lambda$ TAIL TERMINATOR PROTEIN REVEALS THE BIOLOGICALLY RELEVANT HEXAMERIC RING STRUCTURE AND DEMONSTRATES A CONSERVED MECHANISM OF TAIL TERMINATION AMONG DIVERSE LONG-TAILED PHAGES ........................................ 35
2.1 Overview .................................................................................................................. 35
2.2 Materials and Methods ............................................................................................ 36
  2.2.1 Sample Preparation and Crystallization .......................................................... 36
  2.2.2 Data Collection, Structure Determination, and Model Refinement .................. 39
  2.2.3 Functional Characterization ............................................................................. 40
  2.2.4 Electron Microscopy ......................................................................................... 40
  2.2.5 Sequence Alignment and Analysis .................................................................... 41
2.3 Results ...................................................................................................................... 41
  2.3.1 The X-ray Crystal Structure of gpU ................................................................. 42
  2.3.2 Conservation Analysis of an Alignment gpU Homologues .............................. 48
  2.3.3 Characterization of the Functional Surfaces of gpU through Site-Directed
      Mutagenesis ......................................................................................................... 49
  2.3.4 A Putative Role for the λ TrP in Genome Injection ........................................ 52
  2.3.5 Structural and Sequence Similarities between gpU and the TrPs of Contractile-
      Tailed Phages ........................................................................................................ 54
2.4 Discussion .................................................................................................................. 59

CHAPTER 3 ......................................................................................................................... 64

THE STRUCTURE OF THE PHAGE λ MAJOR TAIL PROTEIN REVEALS A COMMON
EVOLUTION FOR LONG-TAILED PHAGES AND THE TYPE VI BACTERIAL
SECRETION SYSTEM ....................................................................................................... 64

3.1 Overview .................................................................................................................. 64
3.2 Materials and Methods ............................................................................................ 65
  3.2.1 Plasmid Construction and Protein Purification ................................................ 65
  3.2.2 gpV Domain Mapping ...................................................................................... 67
  3.2.3 Crystallization Screening .................................................................................. 68
  3.2.3 NMR Spectroscopy and Structure Calculation .............................................. 68
  3.2.4 In vivo Assay for gpV Activity .......................................................................... 69
  3.2.5 Electron Microscopy ......................................................................................... 70
  3.2.6 Construction of a Sequence Alignment of gpV_N Homologues ...................... 70
3.3 Results .................................................................................................................................................. 71
  3.3.1 Biochemical Domain Mapping of gpV .............................................................................................. 71
  3.3.2 Crystallization Studies on gpV full length and gpV\textsubscript{N} ........................................................... 75
  3.3.2 The Solution Structure and Dynamics of the N-terminal Domain of gpV ............................... 77
  3.3.2 The Central Disordered Region of gpV\textsubscript{N} is Functionally Important ................................. 82
  3.3.3 Structural Similarity Between gpV\textsubscript{N} and a Tail Tube Protein from a Contractile-
          Tailed Phage ........................................................................................................................................ 84
  3.3.4 Phage λ Tails are Structurally Related to the Hcp1 Type VI Secretion System
          Protein .................................................................................................................................................... 86
  3.3.5 A Model for gpV Polymerization Based on Hcp1 ................................................................. 87
  3.4 Conclusions .......................................................................................................................................... 90

CHAPTER 4 ................................................................................................................................................ 92

STRUCTURAL AND PRELIMINARY FUNCTIONAL CHARACTERIZATION OF THE C-
TERMINAL DOMAIN OF THE λ MAJOR TAIL PROTEIN ......................................................................... 92
  4.1 Overview ................................................................................................................................................ 92
  4.2 Materials and Methods .......................................................................................................................... 93
    4.2.1 Plasmid Construction and Protein Purification .............................................................................. 93
    4.2.2 Biophysical Characterization ........................................................................................................ 94
    4.2.3 Solution Structure Determination .................................................................................................. 94
    4.2.4 \textit{In vivo} Functional Characterization ......................................................................................... 94
  4.3 Results .................................................................................................................................................... 94
    4.3.1 Biophysical Characterization .......................................................................................................... 95
    4.3.2 The C-terminal Domain and the Linker between the N- and C-terminal Domain is
          Functionally Important ...................................................................................................................... 96
    4.3.3 Solution Structure Determination .................................................................................................. 99
  4.4 Discussion .............................................................................................................................................. 103
    4.4.1 Putative Role of the Linker Between the N- and C-terminal Domains ............................................. 103
    4.4.2 Structural Features of the gpV C-terminal Domain ....................................................................... 104
    4.4.3 Putative Function of gpV\textsubscript{C} ............................................................................................... 105
CHAPTER 5

A PROTEIN FOLDING DEFECT IN THE C-TERMINAL DOMAIN OF THE λ MAJOR TAIL PROTEIN ABROGATES FUNCTION OF THE FULL LENGTH PROTEIN

5.1 Overview
5.2 Materials and Methods
  5.2.1 Plasmid Construction and Protein Purification
  5.2.2 In vivo Functional Studies
  5.2.3 NMR Spectroscopy
    5.2.3.1 HSQCs on gpV_C Mutants
    5.2.3.2 Structure Determination of gpV_C-P227L
    5.2.3.3 Interaction Studies between gpV_N and gpV_C
5.3 Results
  5.3.1 gpV_FL-G222D is Inactive and Dominant Negative
  5.3.2 gpV_C-G222D Interacts with gpV_N
  5.3.3 G222D Causes gpV_C to Partially Unfold at 25 °C
  5.3.4 P227L Induces Conformational Changes in gpV_C
5.4 Discussion

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

6.1 Summary
6.2 Future Directions
  6.2.1 The Structure of a gpV Hexamer
  6.2.2 Elucidating the Role of gpV_C
  6.2.2 Elucidating the Mechanism by which G222D exerts its Dominant Negative Behavior
    6.2.2.1 The Temperature Dependence of G222D
    6.2.2.2 The Effect of Time and Protein Concentration on Interaction Studies
    6.2.2.2 The Requirements to Produce a G222D Dependent Dominant Negative Phenotype
REFERENCES

ix
LIST OF TABLES

Table 1. The Amino Acid Composition of gpV ................................................................. 15
Table 2. The Amino Acid Composition of gpU ................................................................. 18
Table 3. Diffraction and Refinement Statistics For gpU-D74A and gpU-WT ..................... 44
Table 4. N-terminal Sequencing of Trypsin Proteolyzed Fragments from gpV FL and gpV N ...... 73
Table 5. Structural statistics for the ensemble of 20 structures for gpV N ............................ 80
Table 6. Structural Statistics for the Ensemble of 20 Structures for gpV C-WT .................... 100
Table 7. gpV N Residues Implicated in Interactions with gpV C-G222D .............................. 117
Table 8. Amide Chemical Shift Comparison Between gpV C-WT and gpV C-G222D .......... 123
Table 9. Chemical Shift Comparison between gpV C-WT and gpV C-P227L ........................ 126
Table 10. Structural statistics for the ensemble of 20 structures for gpV C-P227L ............... 129

LIST OF FIGURES

Figure 1. General Architecture of the Tailed Phage Families ........................................... 3
Figure 2. Tail Structure from Siphoviridae and Myoviridae ............................................. 4
Figure 3. Gene Order Conservation Between Myoviridae and Siphoviridae ....................... 6
Figure 4. Schematic Representation of a Bacteriophage λ Particle ..................................... 7
Figure 5. Electron Micrograph of Ur-λ ............................................................................. 8
Figure 6. The λ Morphogenic Genes ................................................................................ 9
Figure 7. The Bacteriophage λ Tail Assembly Pathway ..................................................... 13
Figure 8. Schematic Representation of a gpV Hexamer ....................................................... 16
Figure 9. The Monomeric Solution Structure of λ gpU ....................................................... 16
Figure 10. Unstructured Regions in λ Morphogenic Proteins ............................................ 20
Figure 11. Mass Spectrometry Analysis of Native and Se-Met Incorporated gpU-D74A ....... 38
Figure 12. Crystals of the Se-Met Incorporated gpU-D74A and Native gpU-WT ................. 39
Figure 13. The Quaternary Structure of gpU ................................................................. 43
Figure 14. The Oligomerization Interface of gpU ............................................................. 45
Figure 15. Comparison of the NMR and X-ray Structure of gpU ....................................... 47
Figure 16. Alignment of gpU Homologues ................................................................. 49
Figure 17. Functional Characterization of the Hexamerization, Head Binding, and Tail Termination Surfaces in gpU ................................................................. 51
Figure 18. Identification of the Head and Tail Binding Surfaces of gpU ........................................ 52
Figure 19. The DNA Injection Pore of λ gpU ........................................................................ 54
Figure 20. Comparison of Tail Terminator Proteins from Contractile and Non-Contractile Tailed Phages ......................................................................................... 56
Figure 21. Sequence Network of Tail Terminator Proteins from Contractile and Non-Contractile Tailed Phages ......................................................................................... 57
Figure 22. Sequence Comparison of Tail Terminator Proteins from Contractile and Non-Contractile Tailed Phages ......................................................................................... 58
Figure 23. The Purification of gpV FL and gpV N .................................................................. 67
Figure 24. Limited Trypsin Proteolysis of gpV FL and gpV N .................................................. 72
Figure 25. The Trypsin Cleavage Sites in gpV FL .................................................................. 74
Figure 26. Size Exclusion Profiles of Native and Proteolyzed gpV N ........................................ 75
Figure 27. Crystallization of gpV FL ..................................................................................... 76
Figure 28. Crystallization of gpV N ...................................................................................... 76
Figure 29. Crystallization of the Chymotrypsin-Resistant Core of gpV N .................................. 77
Figure 30. The Solution Structure and Dynamics of gpV N .................................................. 79
Figure 31. An Alignment of Diverse Homologues of gpV ....................................................... 81
Figure 32. The Role of Disordered Regions in λ Tail Assembly ............................................. 83
Figure 33. Structural Similarity between gpV N and a Tail Tube Protein from a Contractile-Tailed Phage ......................................................................................... 85
Figure 34. Structural Similarities between gpV N and Hcp1 .................................................. 87
Figure 35. A Model for gpV Oligomerization ...................................................................... 89
Figure 36. The Purification of gpV C ..................................................................................... 95
Figure 37. Spectroscopic Characterization of gpV C .............................................................. 96
Figure 38. The C-terminal Domain and the Flexible Linker Region Between gpV N and gpV C are Functionally Important ................................................................. 98
Figure 39. The Solution Structure of gpV C ......................................................................... 101
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST</td>
<td>basic local alignment search tool</td>
</tr>
<tr>
<td>BP</td>
<td>base plate</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CHAP</td>
<td>cysteine, histidine-dependent amidohydrolases/peptidases</td>
</tr>
<tr>
<td>cryoEM</td>
<td>cryo-electron microscopy</td>
</tr>
<tr>
<td>Da</td>
<td>dalton</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diamine tetraacetic acid</td>
</tr>
<tr>
<td>gp</td>
<td>gene product</td>
</tr>
<tr>
<td>gpV&lt;sub&gt;C&lt;/sub&gt;</td>
<td>C-terminal domain of gene product V</td>
</tr>
<tr>
<td>gpV&lt;sub&gt;FL&lt;/sub&gt;</td>
<td>full-length gene product V</td>
</tr>
<tr>
<td>gpV&lt;sub&gt;N&lt;/sub&gt;</td>
<td>N-terminal domain of gene product V</td>
</tr>
<tr>
<td>Hep1</td>
<td>hemolysin coregulated protein 1</td>
</tr>
<tr>
<td>HOC</td>
<td>highly immunogenic outer capsid</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>IC</td>
<td>initiator complex</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Ig-like</td>
<td>immunoglobulin-like</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>λ</td>
<td>lambda</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani medium</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>matrix-assisted laser desorption/ionization - time of flight</td>
</tr>
<tr>
<td>mRNA</td>
<td>messanger ribonucleic acid</td>
</tr>
<tr>
<td>MTP</td>
<td>major tail protein</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel-nitrilotriacetic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear overhauser effect</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PaPa</td>
<td>Paris Paseadena</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>pI</td>
<td>isolelectric point</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethlysulphonyl fluoride</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>PSI-BLAST</td>
<td>position specific iterative-BLAST</td>
</tr>
</tbody>
</table>
PVDF  polyvinylidene fluoride
RMSD  root mean square deviation
RNA  ribonucleic acid
RPM  revolutions per minute
SAD  single wavelength anomalous diffraction
SDS  sodium dodecyl sulfate
Se-Met  selenomethionine
ssDNA  single stranded deoxyribonucleic acid
Stf  side tail fiber
T3SS  type 3 secretion system
T6SS  type 6 secretion system
TCs  tail chaperones
TEM  transmission electron microscopy
TEV  tobacco etch virus
TF  tail fiber
Tfa  tail fiber assembly
TMP  tape measure protein
TMV  tobacco mosaic virus
Tris  tris(hydroxymethyl)aminomethane
TrP  terminator protein
TSP  tail sheath protein
<table>
<thead>
<tr>
<th>TTP</th>
<th>tail tube protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Macromolecular complexes mediate a diverse range of essential cellular processes including DNA replication, transcription, translation, and cellular motility. The assembly of these complexes proceeds along complicated, ordered pathways involving protein-protein and often protein-DNA and/or protein-RNA interactions. Therefore, in order to fully understand cellular function, one must precisely determine the specific interactions and regulatory mechanisms involved in building these large assemblies. Unfortunately, due to their intricacy and often-transient nature, many of these complexes are extremely difficult to work with, and therefore model systems that can easily be investigated in a laboratory are necessary for scientific advancement in a multitude of research areas.

The morphogenesis of the bacteriophage λ tail presents an advantageous model system for studying the assembly of complicated multi-component structures, as its morphogenic pathway has been well characterized and utilizes precisely coordinated interactions between hundreds of copies of 11 different proteins. While the roles of many of these proteins have been elucidated, there are still many questions regarding the regulatory mechanisms controlling tail assembly that need to be addressed: What types of protein-protein interactions are involved in building the phage tail? What mechanisms are in place to modulate the stringent timing of protein oligomerization and incorporation into the complex? What types of structural rearrangements occur upon assembly? Do similarities exist between the structure and assembly of the λ tail and other macromolecular machines? To address these questions, I have investigated two essential steps in the λ tail assembly pathway: the self-assembly and polymerization of the tail tube forming protein, gpV, and the cessation of tail tube assembly by the terminator protein, gpU.

The following sections provide a brief overview of bacteriophage structure and evolution, with emphasis on bacteriophage λ and the role of gpV and gpU in tail assembly. In addition, the current knowledge on the regulatory mechanisms involved in controlling tail assembly will be discussed and compared to those used by other large and presumably unrelated
tubular complexes. Finally, the specific aims of this research project are stated and an outline of each chapter is provided.

1.1 Bacteriophages: A General Introduction

Bacteriophages form a diverse group of viruses that infect bacteria. Some of the earliest observations of bacterial viruses date back to 1896 when waters from the Ganges and Jumna rivers were shown to have curative actions on cholera bacteria (Hankin, 1896; Thacker, 2003). It wasn’t until twenty years later that this phenomenon, noted by two scientists working independently, was recognized as a filterable entity capable of killing bacteria (d'Herelle, 1917; D'Herelle, 2007; Twort, 1915). Today, phages have been found in all environments where their hosts exist and are especially abundant in the oceans, with approximately $10^{10}$ particles per litre in surface waters (Bergh et al., 1989; Wommack and Colwell, 2000). The global population of phages is estimated to be on the order of $10^{31}$ individuals, making them the most abundant biological entities on the planet (Hendrix, 2003). With help from genome sequencing initiatives, the tremendous diversity in phage populations is becoming universally recognized.

Approximately 60% of all sequenced bacterial genomes contain at least one prophage (Casjens, 2003), a term given to a phage genome that has been incorporated into the bacterial genetic material. Prophages can constitute up to 20% of a bacterium’s genome and are major contributors to genomic diversity among species (Casjens, 2003). Notably, 75% of prophage genes have not yet been assigned a function (Edwards and Rohwer, 2005).

Currently, over 5500 different phages have been examined under an electron microscope, and of these, approximately 96% possess a specialized structure called a tail (Ackermann, 2007). Tailed phages are classified into the order of Caudovirales and are subdivided into three families: the Myoviridae (long, contractile tails), the Siphoviridae (long, non-contractile tails), and the Podoviridae (short tails) (Ackermann, 2007) (Figure 1). Strikingly, greater than 80% of these phages possess a long tail and hence belong to either the Siphoviridae (~61%) or Myoviridae (~25%) families. The majority of this study will focus on the structure, assembly and evolution of long-tailed bacteriophages.
Figure 1. General Architecture of the Tailed Phage Families.

A schematic diagram indicating the general structure of Podoviridae, Siphoviridae and Myoviridae phages. The number above each box indicates the percentage of each type observed in the phage population (Ackermann and Kropinski, 2007).

1.2 Bacteriophage Tails

Once assembled, bacteriophage phage tails play a critical role in host cell recognition, adsorption, and genome injection. Both long, non-contractile and long, contractile tails are composed of hundreds of copies of many different proteins that interact with one another in a strictly regulated pathway to form a functional tail. In general, long tails are comprised of one or more tail fibers, a baseplate/initiator complex that is located at the bottom end of the tail, a long tubular structure, and a tail ‘cap’, also known as the tail terminator protein (TrP), located at the top end of the tail (Figure 2). The tube of non-contractile tails is composed primarily of multiple copies of one protein, known as the major tail protein (MTP). During infection, the overall shape of the non-contractile tail is maintained while structural arrangements occur in the inner lumen of the tail (Plisson et al., 2007). In contrast, contractile tails include a central tail tube protein (TTP) that is surrounded by a tail sheath protein (TSP). Upon infection, the tail sheath contracts allowing the tail tube to penetrate the outer membrane and cell wall of the host bacterium (Leiman et al., 2004) (Figure 2). Despite striking functional differences, gross...
morphological similarities can be observed in electron micrographs between contractile and non-contractile tails, the most obvious of which lies in the tail tube region. This raises the question as to whether any evolutionary relationship exists between these two distinct tail types.

Figure 2. Tail Structure from *Siphoviridae* and *Myoviridae*.

A schematic diagram indicating the general structure of tails from *Siphoviridae* and *Myoviridae* phages. TrP, Tail Terminator Protein; MTP, Major Tail Protein; TF, Tail Fiber; IC, Initiator Complex; TSP, Tail Sheath Protein; BP, Baseplate; and TTP, Tail Tube protein.

### 1.3 Bacteriophage Evolution

At present, there are greater than 300 genome sequences available for long-tailed bacteriophages, and by comparing these sequences a wealth of information can be obtained regarding the function and evolution of individual bacteriophage components. In fact, this type of analysis has revealed that bacteriophage sequences have diverged so extensively that significant sequence similarity is often absent, even among proteins that are clearly evolutionarily related. For example, sequence similarities between experimentally verified MTPs from different *Siphoviridae* are often undetectable despite the high probability of a common origin (Rohwer and Edwards, 2002). It is also interesting to note that there are only two proteins encoded by virtually all tailed phages that are frequently recognized at the level of
their primary sequence: the portal protein and the large subunit of the terminase protein (Casjens, 2003; Casjens, 2008), both of which play a role in DNA packaging.

In the absence of sequence similarity, an evolutionary connection between different proteins with similar functions can be inferred if they possess similar three-dimensional structures (Bray et al., 2000). Notably, similar folds have been detected between the capsid proteins of phages HK97 (Siphoviridae), T5 (Siphoviridae) (Effantin et al., 2006), T4 (Myoviridae) (Fokine et al., 2005), P22 (Podoviridae) (Jiang et al., 2003) and λ (Siphoviridae) (Lander et al., 2008), strongly suggesting that the heads of tailed dsDNA phages have a common evolutionary origin. Furthermore, the similarity in capsid fold also extends to eukaryotic viruses such as herpes (Duda et al., 2006).

Similar structural comparisons between phage tails have not yet been possible as no structure of any major tail or tail tube protein has been reported and there are no structural homologues for the recently reported X-ray crystal structure of ~ 75% of the tail sheath protein from the Myoviridae phage T4 (Aksyuk et al., 2009). In spite of this, there are at least two pieces of evidence suggesting that contractile and non-contractile tails may share a common ancestor. First, regions of the λ side tail fiber (gpStf) and tail fiber assembly (gpTfa) proteins, exhibit sequence similarities to tail fiber proteins from contractile-tailed phages. For example, a portion of λ gpStf exhibits sequence similarities to tail fiber proteins from the Myoviridae phages P2 and T4 (Hendrix and Duda, 1992). Furthermore, gpTfa displays ~40% sequence similarity to the T4 tail protein, gp38 (George et al., 1983; Montag and Henning, 1987). In fact, the λ tfa gene can complement a gp38− mutant in T4 (Montag and Henning, 1987). In addition, gpTfa and parts of gpStf from λ can be substituted for the homologous regions in the T4 gene products (Montag et al., 1989).

A second indication of evolutionary relatedness is provided by the high conservation of gene order among many otherwise unrelated phages (Figure 3). Both Myoviridae and Siphoviridae phages possess a large gene (usually greater than 2 kbp) encoding a tape measure protein (TMP), which is responsible for precisely determining tail length. Genes encoding the tail tube or major tail proteins are generally upstream of the TMP gene, and are separated from it by a gene encoding a tail assembly chaperone. This gene is distinguished by a highly
conserved programmed translational frameshift that is observed in the genomes of both
Siphoviridae and Myoviridae (Xu, 2001; Xu et al., 2004).

![Gene Order Conservation Between Myoviridae and Siphoviridae](image)

**Figure 3. Gene Order Conservation Between Myoviridae and Siphoviridae.**

A conserved tail gene module is shown using λ (Siphoviridae), HK97 (Siphoviridae), Mu (Myoviridae), and P2 (Myoviridae) as examples. In each case the genes encoding the tail terminator protein (TrP), the major tail protein (MTP) or tail tube protein (TTP), the tail chaperones (TCs), and the tape measure protein (TMP) are coloured red, blue, yellow, and green, respectively. In Myoviridae, the gene encoding the tail sheath protein (TSP) is coloured purple.

The high conservation of this distinctive gene module among diverse phages is most easily explained by hypothesizing that they share a common ancestor however, without further mechanistic and structural data it is difficult to draw any definitive conclusions. To address the evolution and assembly mechanism of phage tails, I have carried out structural and functional studies on gpV, and gpU, the major tail protein and tail terminator protein, respectively, from phage λ.

### 1.4 Bacteriophage λ

Bacteriophage λ was discovered unintentionally in 1951 during genetic studies on *Escherichia coli* (*E. coli*) K-12, where λ can exist as a prophage (Lederberg, 1951). It has a linear dsDNA genome (~48 kbps) packaged within an icosahedral head that is attached to a long non-contractile tail (Figure 4). The head is ~ 60 nm in diameter and is composed of 20 equilateral triangular faces. At one unique 5-fold vertex, a structure referred to as the connector complex provides a bridge between the head and the tail. Once heads and tails join, a small
portion of the \( \lambda \) genome enters into the top third of the \( \lambda \) tail (Thomas, 1974). The \( \lambda \) tail is \( \sim 170 \) nm in length and is thought to be flexible based on how it appears in negatively stained electron micrographs. Structurally, the tail can be divided into 3 regions: a single tail fiber \( \sim 23\text{nm} \), a conical base plate complex \( \sim 15\text{nm} \), and the tail tube \( \sim 135\text{nm} \) (Figure 4). Negatively stained electron micrographs of the \( \lambda \) tail reveal that both the conical base plate and the tail tube are striated, indicating that ring-like structures stack together to form the tail (Kuhl and Katsura, 1975). The base plate striations correspond to 3 or 4 rings whereas the tail tube striations correspond to exactly 32 ring-like structures (Katsura, 1983).

![Figure 4. Schematic Representation of a Bacteriophage \( \lambda \) Particle.](image)

Notably, the tails of \( \lambda \) used in laboratory strains today and for the last \( \sim 40 \) years of phage research, are not identical to the strain of \( \lambda \) that was first isolated in 1951 (Dove, 1969). The strain in use today arose from a cross between two strains of \( \lambda \) from laboratories in Pasadena, California and Paris, France. The recombinant, \( \lambda \text{PaPa} \), more commonly referred to as \( \lambda \text{-WT} \), is thought to differ from the original isolate by a one base pair deletion in the gene encoding a side tail fiber (\( \text{stf} \)) protein (Hendrix and Duda, 1992). Restoration of the base pair deletion (\( \text{Ur-}\lambda \)) gives rise to approximately 6 thin, jointed tail fibers that protrude from the sides of the conical region of the \( \lambda \) tail (Figure 5).
Figure 5. Electron Micrograph of Ur-λ.

A representative image of a λ particle with restored side tail fibers. Black arrows indicate 4 visible jointed tail fibers protruding from the conical region of the tail. Figure adapted from http://www.asm.org/division/m/foto/UrLamMic.html (accessed on 02/19/09).

When compared to preparations of λ-WT, two protein products were identified in the ‘restored’ Ur-λ virion: a side tail fiber protein (77.4kDa) and tail fiber assembly protein (21.6kDa). Side tail fibers are not essential for the assembly or infectivity of λ, however they have been shown to play a role in host cell adsorption (Hendrix and Duda, 1992). Also, unlike λ-WT infections which yields medium size plaques, infections with Ur-λ give rise to tiny plaques (Hendrix and Duda, 1992). The strains of λ used for the experiments presented within this study lack side tail fibers and are referred to as λ-WT.

Genes encoding λ structural proteins are found on the left arm of the λ chromosome and are transcribed into one long mRNA by the host cell RNA polymerase. The amount of mRNA synthesized is essentially the same for all late genes, however the amount of protein produced is controlled at the translational level and gives rise to varying amounts of each morphogenic protein (Murialdo and Siminovitch, 1972; Sampson et al., 1988). The amount of protein produced is not always proportional to the number of protein copies found in the infectious virion.

The assembly of an infectious λ virion requires the action of multiple copies of 21 λ proteins. These proteins act in strictly regulated pathways involving many protein-protein and
protein-DNA interactions. Similar to the assembly of other long-tailed phages, λ heads and tails are assembled in independent pathways and join in the final assembly step to form an infectious virus. This study focuses strictly on proteins involved in λ tail assembly.

1.5 Wild-Type λ Tail Assembly Pathway

The assembly of the phage λ tail has been one of the most intensively studied morphogenic pathways of all phages. The products of 11 phage genes (J, I, K, L, M, H, G, GT, V, U, Z) are involved in the assembly of the λ tail. These genes are clustered together immediately downstream of the head assembly genes (Figure 6) and each of their gene products will be described below.

Figure 6. The λ Morphogenic Genes.

The λ head and tail assembly genes are represented schematically as they occur on the λ chromosome. Genes are represented as boxes in either red (head assembly) or blue (tail assembly) and the corresponding gene name is found in each box. An arrow above genes G and T indicates a programmed translational frameshift. A kilobase pair ruler is shown below the genes to demonstrate their approximate position within the 48kbp genome. Figure adapted from Hendrix et al. (Hendrix, 2006)

Genetic and in vitro complementation studies have established the order in which the tail genes act in the assembly pathway and the concentration dependence of at least two gene products, gpM and gpV (Katsura, 1976b; Katsura and Kuhl, 1975b; Kuhl and Katsura, 1975). Wild-type tail assembly begins with the product of gene J or gpJ, which encodes the tail fiber protein, of which 3 copies have been estimated to exist in the fully assembled tail (Casjens and Hendrix, 1974). gpJ is located at the tip of the tail (Buchwald and Siminovitch, 1969; Dove, 1966; Mount et al., 1968) and the C-terminal portion of the protein interacts directly with the
lamB receptor on the E. coli host cell (Wang et al., 2000). In phage lysates lacking gpJ, no tail structures are observed (Mount et al., 1968).

Proteins gpI, gpL, gpK, gpH, gpG, gpGT and gpM are then thought to interact in a stepwise manner, either directly or indirectly with gpJ, forming a structure referred to as the initiator complex (Katsura, 1976b; Katsura and Kuhl, 1975b) (Figure 7). It is currently not clear how many copies of each of these proteins participate in forming the complex and in many cases their specific functional roles remain to be elucidated. gpK, while largely uncharacterized, has recently been identified as a member of the CHAP (cysteine, histidine-dependent amidohydrolases/peptidases) superfamily of proteins (Rigden et al., 2003). CHAP proteins are capable of hydrolyzing amide bonds and are characterized by a conserved triad consisting of a cysteine, histidine and aspartic acid residue (Bateman and Rawlings, 2003; Rigden et al., 2003). This finding suggests that gpK may play an enzymatic role in host cell infection by potentially cleaving the amide bond between the N-acetylmuramic acid and L-alanine in the bacterial cell-wall peptidoglycan. This action would assist in penetration of the host cell wall and could be crucial for genome injection. Additionally, gpK may be the unidentified protease responsible for the proteolytic degradation of gpH into gpH* during tail assembly (discussed in more detail below) (Tsui and Hendrix, 1983). Intriguingly, lysates made from one particular mutant in gene K (Kam768) were shown to accumulate inactive particles that sediment as fast as normal phage (Katsura and Kuhl, 1975b). This result suggests that gpK may be comprised of two domains: a structural domain and a catalytic domain. Determining the position of the Kam768 mutant within the K gene, or testing additional K mutants may offer an explanation for this observation.

Both gpL and gpM have been implicated in the formation of a pseudo-initiator complex on to which gpV, the MTP, can polymerize to form polytubes (Katsura, 1976b). This observation suggests that within the initiator complex, proteins L and/or M may be localized at the top of the complex where they can be easily accessed by gpV. Since gpM acts at the last step of the pathway involved in forming the wild-type initiator complex, it is the most likely candidate for direct interactions with gpV (Katsura, 1976b) (Figure 7). It should be noted that the pseudo-initiator complex is only sufficient for gpV polymerization when a large number of free gpV molecules have accumulated in the absence of the wild-type initiator, suggesting that the wild-type initiator complex provides a more efficient platform for gpV polymerization.
There is currently little known about gpI except that lysates lacking gpI do not form tails (Weigle, 1966).

gpH is ~90kDa (853 amino acids) and is predicted to be largely alpha-helical. During tail assembly, at some point between tail termination and tail activation, gpH gets cleaved to gpH* (78kDa), and it is this proteolyzed form of gpH that is found in the infectious virion. gpH is thought to act, in part, as a phage encoded tape measure protein for determining tail length. Directly supporting this hypothesis are observations that viable in-frame insertions or deletions made to the middle of the $H$ gene altered the length of the tail by an amount that could be predicted by the change made to the gene size (Katsura, 1987; Katsura and Hendrix, 1984). There is also increasing evidence that gpH plays an additional role in $\lambda$ DNA injection. First, gpH gets injected into liposomes during the process of DNA injection (Roessner and Ihler, 1984) and second, mutations in gene $H$ can overcome injection deficiencies when $\lambda$ infects $pel$ E. coli strains, a strain that is resistant to $\lambda$ infections (Scandella and Arber, 1974, 1976).

While the exact mechanism by which gpH exerts its effect remains unclear, it is hypothesized to extend up the length of the tail acting as a template for gpV to polymerize around, and hence once the end of gpH is reached, gpV polymerization temporarily pauses. The extension of gpH up the length of the tail is thought to be assisted by two phage encoded chaperones, gpG and gpGT (Figure 7). The region encoding gene $G$ (upstream) and $T$ (downstream) contain overlapping reading frames that are related by a programmed translational frameshift (Levin et al., 1993; Xu et al., 2004). About 3.5 % of the time, a ‘slippery sequence’ in the mRNA (GGGAAAG) causes the ribosome to slip back one nucleotide (-1 reading frame). The -1 reading frame product is the gpGT fusion protein. The ratio of gpG:gpGT is crucial for the production of biologically active tails (Xu, 2001). The T portion of gpGT interacts with the MTP, gpV, and gpG and the G portion of gpGT binds to gpH (Xu, 2001). Neither gpG or gpGT are found in the infectious phage particle, further strengthening the idea that they are assembly chaperones.

In the absence of the initiator complex, the MTP, gpV, exists in an unoligomerized form (Katsura and Tsugita, 1977). Upon encountering the initiator, gpV oligomerizes on top of the conical initiator and around gpH to form hexameric rings. This process continues until there are exactly 32 gpV hexamers stacked on top of one another forming the tail tube (Figure 7).
incorporation of the correct number of gpV subunits into the tail, gpV polymerization pauses (Katsura, 1976b) and if gpU is present, it oligomerizes on top of gpV to form a gpU hexamer which permanently halts further polymerization (Figure 7). It is therefore speculated that gpH prevents gpU from interacting with gpV hexamers until the tail has reached the requisite length. After gpU terminates tail elongation, gpZ ‘activates’ the tail, and then the λ head, which is formed in an independent pathway, binds to gpU to form an infectious phage particle (Figure 7). While phage particles lacking gpZ resemble λ-WT, they have heads of normal morphology attached to wild-type looking tails, Z- phage are unable to inject their DNA (Thomas et al., 1978) (Figure 7). Additionally, unlike in λ-WT where the λ genome protrudes into the top portion of the phage tail, there is no evidence of DNA in the top portion of a Z- tail (Thomas et al., 1978). It has not yet been determined whether gpZ exists in the infectious virion. At some point between gpU binding and the activation of the tail by gpZ, gpH gets cleaved to gpH* (Tsui and Hendrix, 1983) by an unidentified protease. Since this study will focus on both gpV and gpU, they will be discussed in more detail in sections 1.5.1 and 1.5.2, respectively.
Figure 7. The Bacteriophage λ Tail Assembly Pathway.

In the wild-type tail assembly pathway (boxed in black), an initiator complex (green) is formed first followed by the polymerization of gpV (dark blue) on top of the initiator. This process is regulated by the TMP, gpH (orange), and the phage encoded chaperones gpG (light blue) and gpGT (light blue-green). Once the precise tail length is reached, gpV polymerization is terminated by gpU (purple) and the tail is activated by gpZ (pink). The question mark above gpZ indicates that it has not yet been identified in the infectious virion. Abnormal assembly intermediates and pathways (boxed in red) arise due to mutations in several tail genes.

1.5.1 gpV: The λ Major Tail Protein

gpV is the most abundant protein in the λ tail (Buchwald et al., 1970) and is found distributed along the entire length of the tail tube (Casjens and Hendrix, 1974; Katsura and Tsugita, 1977). Given its sheer abundance and position in the phage particle, it is not surprising...
that the majority of gpV mutants isolated to date lack tail tube structures (Katsura, 1976b; Kuhl and Katsura, 1975). Interestingly, several mutations in gpV have been identified that produce fully or partially assembled phages. For example, at least 13 different temperature sensitive mutants in gene V have been isolated that form phages capable of overcoming DNA injection deficiencies on *E. coli pel* strains (Scandella and Arber, 1976). *pel* Strains allow λ to bind to the cell but do not permit DNA injection (Scandella and Arber, 1974). The opposite scenario has also been observed as at least two different mutations in gene V have been isolated that form phage particles resembling wild-type λ but that are unable to inject their DNA into normal *E. coli* cells (Katsura, 1976a). Together, these findings suggest that gpV plays an important role in genome injection. This hypothesis is supported by the fact that gpV forms a conduit through which DNA is initially positioned (Thomas, 1974; Thomas et al., 1978) and then transported during infection. Surprisingly, 5 gpV mutants have been isolated that produce short tails that are not attached to heads (Katsura, 1976a). The short tails, which are on the average of half the length of the wild-type tail, may arise due to abnormal or inefficient polymerization properties in the protein itself. The mechanism by which each of these V mutants exerts their gain of function phenotypes remains largely uncharacterized.

gpV is 246 amino acids in length (~26kDa) and has a pI of ~ 5.28. It is abundant in both acidic and basic amino acids, and lacks histidine and cysteine residues (Table 1).
Table 1. The Amino Acid Composition of gpV

<table>
<thead>
<tr>
<th>Residue</th>
<th>Number Found</th>
<th>% of gpV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>27</td>
<td>10.98%</td>
</tr>
<tr>
<td>Arg</td>
<td>8</td>
<td>3.25%</td>
</tr>
<tr>
<td>Asn</td>
<td>8</td>
<td>3.25%</td>
</tr>
<tr>
<td>Asp</td>
<td>16</td>
<td>6.50%</td>
</tr>
<tr>
<td>Gln</td>
<td>7</td>
<td>2.85%</td>
</tr>
<tr>
<td>Glu</td>
<td>10</td>
<td>4.07%</td>
</tr>
<tr>
<td>Gly</td>
<td>23</td>
<td>9.35%</td>
</tr>
<tr>
<td>His</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Ile</td>
<td>6</td>
<td>2.44%</td>
</tr>
<tr>
<td>Leu</td>
<td>10</td>
<td>4.07%</td>
</tr>
<tr>
<td>Lys</td>
<td>14</td>
<td>5.69%</td>
</tr>
<tr>
<td>Met</td>
<td>6</td>
<td>2.44%</td>
</tr>
<tr>
<td>Phe</td>
<td>7</td>
<td>2.85%</td>
</tr>
<tr>
<td>Pro</td>
<td>13</td>
<td>5.28%</td>
</tr>
<tr>
<td>Ser</td>
<td>20</td>
<td>8.13%</td>
</tr>
<tr>
<td>Thr</td>
<td>31</td>
<td>12.60%</td>
</tr>
<tr>
<td>Tyr</td>
<td>5</td>
<td>2.03%</td>
</tr>
<tr>
<td>Val</td>
<td>29</td>
<td>11.79%</td>
</tr>
<tr>
<td>Trp</td>
<td>6</td>
<td>2.44%</td>
</tr>
<tr>
<td>Cys</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Circular dichroism experiments performed on unoligomerized gpV indicated that it was comprised primarily of random-coil with approximately 5% \( \alpha \)-helix, and 20% \( \beta \)-sheet content (Katsura and Tsugita, 1977). In these experiments unoligomerized gpV was easily obtained as the protein is highly expressed and exists in a monomer-dimer equilibrium prior to its incorporation into the \( \lambda \) tail (Katsura and Tsugita, 1977). Intriguingly, early studies revealed that the \( \lambda \) tail could be rapidly and reversibly dissociated into individual stable substructures at pH values below 2.8 or above 11.8 (Bleviss and Easterbrook, 1971; Easterbrook and Bleviss, 1969). At indiscriminate time points after dissociation, any step back towards neutrality drove the reassembly of the substructures into long tail-like structures. Both the dissociation and re-association reactions were completely inhibited in the presence of 0.7 M NaCl, suggesting a putative role for electrostatic interactions in the assembly of the \( \lambda \) tail. Remarkably, while the alkali subunits were irresolvable by electron microscopy, the acid induced subunits were visualized as one or a stack of a few rings that appear similar to the ring-like structures that stack to form the tail tube (Easterbrook and Bleviss, 1969). Examination of the acid-induced
ring-like structures revealed that they were each hexamers with a 9 nm diameter and a central 3 nm pore (Katsura, 1981) (Figure 8). The observation that monomers and dimers are involved in tail formation (Katsura and Tsugita, 1977), but hexamers are obtained upon acid-induced tail dissociation, suggests that gpV may undergo a stabilizing conformational change upon its oligomerization into the tail tube. Further inspection of the hexameric rings revealed six to twelve outer protrusions around the exterior of the inner ring, which are thought to be the C-terminal domain of the protein. The outer protrusions give rise to an overall ring diameter of 18 nm (Katsura, 1981) (Figure 8).

![Figure 8. Schematic Representation of a gpV Hexamer.](image)

The ‘N’ and ‘C’ labels indicate the N and C-terminal domains, respectively.

Several lines of evidence suggest that the C-terminal domain of gpV (gpVc) may not always be essential for function. Notably, a number of mutations that mapped to the end of the V gene were found to be leaky (Katsura, 1976a). Furthermore, mutant phage particles have been isolated where as much as 1/3 of the major tail protein was absent (Katsura, 1981). While the tails of these mutant phages appeared to be thinner and smoother compared to preparations of wild-type λ, the total shape and, under some conditions, the overall infectivity of these phage particles were not affected (Katsura, 1981). Intriguingly, attempts to obtain acid-induced hexamers from tails made from small major tail proteins were unsuccessful (Katsura, 1981), suggesting that the missing portion of gpV may act to stabilize the tail structure. Electron microscopy observations (Katsura, 1981) and protease sensitivity tests (Roessner and Ihler, 1984) indicated that the C-terminal part of gpV is ~8 kDa and exists as a protrusion at the outer
surface of the tail tube where it is susceptible to proteolytic removal (Roessner and Ihler, 1984). In addition to its removal, the C-terminus of gpV can also withstand both small and large additions without affecting its assembly into the virion (Dunn, 1995, 1996).

In contrast to these findings, a C-terminal defective mutant was isolated (VdefK244) and found to be temperature sensitive as it was viable at 30 °C but not at 37 °C (Katsura, 1976a, 1981). DNA sequencing of this mutant (carried out in the laboratory of Dr. Roger Hendrix) revealed that the only difference between VdefK244 and wild-type gpV was that glycine 222, within the C-terminal protein of gpV, was mutated to an aspartic acid (G222D) (personal communication, Dr. Bob Duda). Since the C-terminal region of gpV can be absent without altering the overall infectivity of the phage particles (Katsura, 1981), it is intriguing that a single point mutation within the C-terminal region of the protein can abrogate function. Interestingly, one pseudo-revertant that was isolated from VdefK244 was sequenced to reveal that in addition to the G222D mutation, a second site mutation was present whereby proline at position 227 was mutated to a leucine (P227L) (personal communication, Dr. Bob Duda). The molecular basis for the loss of function due to the G222D mutation, and the return of function in the G222D/P227L double mutant remains unclear.

Taken together, these findings imply that while gpVC may not be essential for tail assembly, conditions exist where the domain is in fact required for phage infection. Bioinformatic studies aimed at delineating the gpV domain boundaries revealed that gpVC displays significant sequence similarity to a variety of immunoglobulin-like (Ig-like) domains (Fraser et al., 2006).

### 1.5.1.1 Immunoglobulin-like Domains in Bacteriophage

Recently, Ig-like domains have been found to be abundant on the surface of tailed dsDNA phages. Specifically, 68 Ig-like domains located in 54 structural proteins in 41 different tailed dsDNA phage genomes were identified (Fraser et al., 2006). Ig-like domains were not found in ssDNA or RNA phages, or in dsDNA phages lacking a tail (Fraser et al., 2006). The gene encoding these domains can lie in-frame with the gene to which they are appended or they can be added to morphogenic proteins through a ribosomal frameshift. Interestingly, bacteriophage Ig-like domains were predominantly found in five classes of structural proteins:
tail fiber, baseplate/initiator, major tail, major head, and the highly immunogenic outer capsid (HOC) proteins (Fraser et al., 2006). Since proteins in this group are exposed on the surface of a mature virus, and because Ig-like domains are predominantly involved in binding reactions, it has been speculated that these domains within phage morphogenic proteins play an accessory role in host cell adsorption. There are currently no structures available of an Ig-like domain from any bacteriophage structural protein, nor is there any direct evidence of their specific functional role.

1.5.2 gpU: The λ Tail Terminator Protein

The product of the phage λ U gene (gpU) is the best-characterized tail terminator protein (TrP) from any long-tailed phage. gpU is a small (~14.8kDa) acidic protein (pI ~ 4.2) that lacks cysteine and asparagine residues (Table 2).

<table>
<thead>
<tr>
<th>Residue</th>
<th>Number Found</th>
<th>% of gpU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>16</td>
<td>12.21%</td>
</tr>
<tr>
<td>Arg</td>
<td>5</td>
<td>3.82%</td>
</tr>
<tr>
<td>Asn</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>Asp</td>
<td>16</td>
<td>12.21%</td>
</tr>
<tr>
<td>Gln</td>
<td>2</td>
<td>1.53%</td>
</tr>
<tr>
<td>Glu</td>
<td>11</td>
<td>8.40%</td>
</tr>
<tr>
<td>Gly</td>
<td>6</td>
<td>4.58%</td>
</tr>
<tr>
<td>His</td>
<td>3</td>
<td>2.29%</td>
</tr>
<tr>
<td>Ile</td>
<td>5</td>
<td>3.82%</td>
</tr>
<tr>
<td>Leu</td>
<td>12</td>
<td>9.16%</td>
</tr>
<tr>
<td>Lys</td>
<td>2</td>
<td>1.53%</td>
</tr>
<tr>
<td>Met</td>
<td>5</td>
<td>3.82%</td>
</tr>
<tr>
<td>Phe</td>
<td>5</td>
<td>3.82%</td>
</tr>
<tr>
<td>Pro</td>
<td>6</td>
<td>4.58%</td>
</tr>
<tr>
<td>Ser</td>
<td>9</td>
<td>6.87%</td>
</tr>
<tr>
<td>Thr</td>
<td>9</td>
<td>6.87%</td>
</tr>
<tr>
<td>Tyr</td>
<td>7</td>
<td>5.34%</td>
</tr>
<tr>
<td>Val</td>
<td>9</td>
<td>6.87%</td>
</tr>
<tr>
<td>Trp</td>
<td>3</td>
<td>2.29%</td>
</tr>
<tr>
<td>Cys</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
In the completed tail structure, gpU assembles at the top of the tail tube (Katsura and Tsugita, 1977), thereby terminating gpV polymerization, and providing the surface for interaction with the λ head (Katsura and Kuhl, 1975a). In the absence of gpU, λ infections result in the formation of normal heads and extremely long tails, which arise from the uncontrolled polymerization of gpV. In fact, gpU mutants have been isolated that produce long-tailed phages (Katsura, 1976a) and for at least one of these, the long-tailed phenotype is dependent upon the type of medium the strains are cultured in. Additionally, mutants in gpU have also been isolated that produce free ‘normal’ length tails that are not attached to heads. Of further interest, at least one gpU mutant isolated to date exhibits an inhibitory effect on the normal gene U product. It is known that the inhibitory mutant produces many polytails and a small amount of tails of normal length (Katsura, 1976a), however the exact mechanism of this inhibition remains unclear.

Together, these data emphasize the multi-functional role of gpU, such that it has one binding surface for the λ tail, and one for the λ head. Further strengthening this idea, weak intragenic complementation has been observed in vitro between head binding and tail binding mutants of gpU (Katsura, 1976a).

gpU is presumed to constitute a hexameric ring within the tail structure because it spontaneously forms such rings in the presence of Mg\(^{2+}\) that match the size of the hexameric rings of gpV that comprise the bulk of the tail tube (Katsura and Tsugita, 1977). While the formation of Mg\(^{2+}\) oligomers is intriguing, in vitro studies have demonstrated that Mg\(^{2+}\) is not required for gpU activity and in fact, the in vitro activity of gpU is hindered in the presence of 20mM Mg\(^{2+}\) (Katsura and Tsugita, 1977). In vivo, gpU is made in much higher amounts than is found incorporated into the virion (Murialdo and Siminovitch, 1972). Prior to tail assembly, gpU remains monomeric (Katsura and Tsugita, 1977) even at high (~mM) concentrations (Edmonds et al., 2007). This feature allowed for the determination of its monomeric solution structure using NMR spectroscopy (Edmonds et al., 2007). The solution structure of gpU revealed a mixed α-β fold flanked by several unstructured regions (Figure 9). NMR techniques were also used to identify the residues involved in Mg\(^{2+}\) induced oligomerization of gpU. Interestingly, while Mg\(^{2+}\) is not required for gpU function (Katsura and Tsugita, 1977), some of the residues involved in Mg\(^{2+}\) induced oligomerization were also implicated in biological
activity (Edmonds et al., 2007). This raises the question of how the mechanism of Mg\(^{2+}\) induced oligomerization differs from the formation of biologically relevant oligomers.

**Figure 9. The Monomeric Solution Structure of λ gpU.**

(A) The cartoon representation of the lowest energy gpU structure. (B) The superposition of the 20 lowest energy NMR structures. In each case helices are coloured red, strands are blue, and loops are gray. Panels A and B were generated using PyMOL (http://pymol.sourceforge.net/).

A strong similarity has been reported between the NMR structure of gpU and the structure of protein STM4215 (PDB ID: 2GJV) encoded in a prophage element of *S. Typhimurium* (Edmonds et al., 2007). Although this prophage element is uncharacterized, the gene encoding STM4215 lies adjacent to other genes encoding components of a contractile-tail, suggesting that it may be a TrP. Interestingly, STM4215 crystallized as a hexameric ring with dimensions similar to the ring of gpV seen in the λ tail, and to the dimensions observed for a Mg\(^{2+}\) induced oligomer of gpU-WT. Unfortunately, without atomic structural information on a biochemically confirmed TrP in its hexameric state and/or biochemical evidence proving that STM4215 is a true TrP, it is difficult to draw conclusions regarding the relationship between STM4215 and λ gpU.
1.6. Aberrant Tail Assembly Reactions

Mutations in a particular tail gene can lead to the accumulation of various tail precursors as well as abnormalities in downstream assembly steps. Notably, if gpU is not present, gpV will still pause at the correct tail length (Katsura, 1976b) but will then continue to polymerize forming aberrant polytail structures up to several microns in length (Katsura, 1976b; Katsura and Kuhl, 1975a) (Figure 7). These structures arise due to the uncontrolled polymerization of gpV after the correct tail length has been passed and is therefore a result of an error in tail termination. *In vitro*, a polytail can be attached to a λ head in the presence of gpU but these long-tailed phages are only 0.04% as active as λ-WT (Kuhl and Katsura, 1975). The commitment of tail assembly into the polytail pathway leads to additional assembly abnormalities, which may contribute to the low infectivity of a poly-tailed phage. First, in the polytail pathway gpZ can act on the polytail without the action of gpU. Second, gpH and gpH* (the cleavage product of gpH) are not found in polytail structures. Since gpH is required for the formation of a polytail, this suggests that gpH or gpH* is released during polytail assembly (Katsura, 1976b).

A second type of abnormally long tail-like structure has been observed in double mutants of gene *U* combined with mutations in the following tail components: *G, K, J, H, and I* (Katsura, 1976b) (Figure 7). These tail-like structures, called polytubes, reached lengths of up to several microns but lacked the conical initiator complex at the base of the tail. There were no tail related structures observed in *UV* lysates, consistent with the notion that gpV forms the tail tube. Additionally, tail related structures were absent in *UL* and *UM* lysates suggesting that gpL and gpM form a pseudo-initiator complex onto which gpV can prematurely polymerize. In order to prevent abnormal assembly reactions, a strict level of control must be enforced.
1.7 Regulation of λ Tail Assembly

In order for the λ tail to assemble into a functionally active structure, a precise level of regulation is required at several unique steps in the assembly pathway including: 1) protein synthesis; 2) the timing/order of the steps involved in assembly; and 3) tail length regulation.

1.7.1 Regulation of Protein Synthesis

The amount of each structural protein synthesized must be strictly regulated to ensure that enough of each component is available to assemble a functional structure. Of the tail proteins, gpG, gpV and gpU are made in large amounts: 470, 265, and 375 times as many copies as gpA (a protein involved in DNA packaging), respectively (Murialdo and Siminovitch, 1972). Since 192 copies of gpV get incorporated into the tail, it is easy to understand why a large amount of gpV is synthesized. On the other hand, gpU, of which only six copies get incorporated into each tail, is also made in huge excess. In fact, more gpU is made in vivo than gpV. Since the absence of gpU results in the uncontrolled assembly of gpV and hence the formation of inactive polytails, it is extremely important that enough U protein is available to terminate tail assembly. A similar explanation can be used to rationalize the large amount of gpG that is synthesized. gpG acts as a chaperone for the TMP, gpH, which is expressed at relatively low levels (12 times as many copies as gpA). Since gpG interacts with gpH to assist in its assembly (Xu, 2001), it is feasible that gpG completely surrounds the entire length of gpH in its extended form. This would require large amounts of gpG to be present. Regulating the amount of each protein that is synthesized may play a crucial role in preventing abnormal assembly intermediates from being formed.

1.7.2 Regulating the Order of Assembly

Once the assembly proteins are expressed to the correct levels, a strict control mechanism is in place to ensure that proteins do not act in the assembly pathway until ‘their turn’ has arrived. The order in which proteins are made in the cell may play a role in controlling
the early steps in assembly. Tail genes are transcribed from gene Z to gene J, opposite to the order in which they come together to assemble the tail. While this may play a role in the early steps of assembly, it is safe to assume that once all of the proteins have been translated, the order of their actions must be determined by some other mechanism. One hypothesis is that tail assembly is a highly cooperative process involving precisely triggered conformational changes within the proteins themselves. In this manner, each step in the pathway would be highly dependent on the previous step.

In fact, there are numerous examples in the literature of bacteriophage morphogenetic proteins that undergo conformational changes during assembly. Most noteworthy are the structural transitions observed during assembly and maturation of the phage HK97 capsid (Gertsman et al., 2009). These include the formation of stabilizing covalent crosslinks between the protein subunits that form the capsid (Gan et al., 2004), local refolding events (Lee et al., 2004), disorder-to-order transitions (Gertsman et al., 2009), and the twisting and bending of domains (Gertsman et al., 2009). Of these, perhaps the most commonly observed type of structural alteration during bacteriophage assembly involves the ordering of unstructured regions.

1.7.2.1 Unstructured Regions Regulate Assembly

The emergence of three-dimensional structures for many bacteriophage morphogenetic proteins reveals that unstructured regions, or regions that lack regular secondary structure, are a prominent feature in many of these proteins. Some examples include, the λ head assembly proteins gpW (Maxwell et al., 2001), gpFII (Maxwell et al., 2002), and gpD (Yang et al., 2000), the λ tail terminator protein gpU (Edmonds et al., 2007) (Figure 10), as well as tail proteins from the short tailed phage phi29 (Simpson et al., 2001) and the long-contractile tailed phage T4 (Aksyuk et al., 2009; Leiman et al., 2000; Leiman et al., 2006).

The disordered regions that occur within bacteriophage morphogenic proteins are generally quite large and found distributed over the entire primary sequence of the protein. For example, only 30% of the residues comprising the λ connector protein, gpFII, exist within regular secondary structure elements (Maxwell et al., 2002). Furthermore, 40% of the
unstructured regions in gpFII are highly dynamic in solution: a region at the N-terminus spanning residues 1-24, a seventeen residue loop in the middle of the protein comprising residues 46 to 62, and the last ten residues of the protein, 108-117 (Cardarelli et al., unpublished results). Similarly, the 1.1 Å crystal structure of the λ head decorator protein, gpD, revealed that only 27% of its primary sequence comprises regular secondary structure elements (Yang et al., 2000). In this case, the unstructured regions are distributed over the entire length of the protein and the first 14 residues were shown to be dynamic in solution (Iwai et al., 2005; Yang et al., 2000). Another interesting observation is that the unstructured regions within bacteriophage morphogenic proteins often exhibit the propensity to adopt secondary structure. For example, the disordered N-terminus of gpFII, is predicted to have high helical propensity (Maxwell et al., 2002).
In each case, strands are coloured blue, helices are red, unstructured regions are grey, and regions that are disordered are green. The disordered regions include the following residues: gpFII (1-24, 46-62, 108-117), gpD-NMR (1-14), gpU (1-4, 25-32, 45-60, 111-117, and 127-130). The arrows on the NMR structure of gpW and X-ray structure of gpD point to the C- and N-termini, respectively. In gpW, the 14 C-terminal residues were found to be unstructured and were excluded from the structure calculations (Maxwell et al., 2001). The first 14 residues in gpD were disordered and not observed in any of the three monomers in the asymmetric unit (Yang et al., 2000).

In a growing number of cases, the unstructured regions in phage proteins have been shown to be important for function. For example, the 14 C-terminal residues of λ gpW, a head completion protein, are unstructured in solution and the last three of these unstructured residues are critical for function (Maxwell et al., 2000). Similarly, the disordered regions in gpD (Yang et al., 2000), gpU (Edmonds et al., 2007), and gpFII (unpublished results Cardarelli et al.) also
have biological importance and it is believed that these regions may play a role in protein-protein and/or protein-DNA interactions, that ultimately lead to the ordering or folding of the unstructured region. Indeed, unstructured proteins or unstructured regions within proteins have been shown to undergo conformational ordering upon binding their biological partner (Dyson and Wright, 2002). One such example is the transcriptional anti-termination protein N from phage λ. NMR spectroscopy was used to assess the structure of gpN in the presence and absence of one of its binding partners, the boxB component of the RNA enhancer nut (Mogridge et al., 1998). These studies revealed that while gpN is disordered in solution, its arginine-rich amino terminus becomes folded upon binding its biological target. This study also revealed that different functional regions of the protein fold independently when their individual targets are contacted.

At least one well-documented example of a morphogenic phage protein undergoing a disorder-to-order transition upon interacting with its binding partner exists. A monomeric solution structure (Iwai et al., 2005), and trimeric crystal structure (Yang et al., 2000) exists for the λ head decorator protein, gpD. The NMR structure of gpD reveals that it is monomeric in solution and has 14 disordered residues at its N-terminus (Iwai et al., 2005). Crystallization of gpD resulted in the formation of trimers. While the N-terminus of gpD is also disordered in the crystal structure, a fit of the atomic structure into a high-resolution cryo-EM reconstruction of a mature λ head revealed that once incorporated into the λ head, the N-terminus of gpD is ordered and interacts with the capsid shell (Lander et al., 2008). Therefore, disorder-to-order transitions occur in gpD upon capsid binding.

These studies raise the question, what advantage is offered by coupling protein folding with the binding of a biological target during phage morphogenesis? First, the flexibility in disordered regions may allow proteins to interact with their partners more efficiently by presenting a greater surface area for the interaction to initially occur. In other words, the flexibility might allow for steric restrictions to be bypassed that would otherwise hinder the interaction (Garza et al., 2009). In addition, disordered regions in proteins may allow for the presentation of multiple ‘hidden’ binding surfaces that are only revealed after the protein interacts with its first partner and the unfolded region becomes folded. In this way, the order in which proteins interact with one another could be strictly controlled.
1.7.3 Tail Length Regulation

Within any particular species of phage, virtually every particle possesses a tail of precisely the same length. It is also interesting to note that tail length varies between different phage types, both within and between families. For example, the \( \lambda \) (\textit{Siphoviridae}) tail is shorter than tails from HK97 (\textit{Siphoviridae}), phi80 (\textit{Siphoviridae}), and T5 (\textit{Siphoviridae}) by 17\%, 19\%, and 27\% respectively; however it is 11\% and 27\% longer than the HK022 (\textit{Siphoviridae}) and T4 (\textit{Myoviridae}) tail, respectively (Abuladze et al., 1994; Katsura and Hendrix, 1984). Together, these observations suggest that all long-tailed phages employ a mechanism to precisely determine tail length.

Early studies aimed at elucidating the mechanism of tail length determination led to the proposal of three theories (reviewed in (Hendrix, 1988; Katsura, 1990; Katsura and Hendrix, 1984)). The first theory assumed that as individual subunits of the major tail protein are added to the growing tail, conformational deformities are introduced into the subunits that eventually prevent further polymerization. The second theory, called the “vernier theory”, proposed that two proteins must polymerize along side one another, and therefore the polymerization of one of the proteins is highly dependent on the other. In this manner, when one protein ceases to polymerize, the other would also stop. The final theory put forth was termed the “template model” and suggested that a molecular ruler is responsible for dictating tail length. The template model is the only model that has been supported by experimental evidence, the most direct of which was provided when viable in frame insertions and deletions made to \( \lambda \), gpH were shown to alter tail length by an amount that could be predicted by the alteration made to the gene size (Katsura, 1987; Katsura and Hendrix, 1984). Today, it is clear that tail length in both contractile and non-contractile tails, is regulated by the concerted action of two proteins, a tape measure protein and an additional protein termed the tail terminator protein.

While the exact mechanisms by which these proteins exert their functions are unknown, the current hypothesis for tail length determination in phage \( \lambda \) is as follows. First, the tape measure protein, gpH, plays a role in the formation of the initiator complex, presumably by interacting with the other proteins in the complex through its C-terminus (Katsura, 1990; Murialdo and Siminovitch, 1972). The remainder of gpH is thought to extend outward, away
from the initiator where it becomes surrounded by gpG and gpGT (Xu, 2001). The rationale behind the interaction between gpG and gpH is not clear, however it has been proposed that gpG may act to stabilize or shield the tape measure from proteolysis (Xu et al., 2004). In support of this hypothesis, the crystal structure of a gpG homologue from the Siphoviridae Lactococcus lactis phage p2 reveals a hydrophobic cleft on one surface of the protein that is predicted to interact with a 40 amino acid repeat of aromatic residues in the p2 tape measure protein (Siponen et al., 2009). Furthermore, through the application of crystallographic symmetry the gpG homologue forms two large spirals with an inner diameter of 36 Å. The docking of these two spirals onto a model of the helical tape measure protein repeat indicated that the two proteins have complementary surfaces (Siponen et al., 2009). In its extended form, gpH presents a template around which unoligomerized gpV can assemble. While the presence of gpH is not a requirement for gpV polymerization (gpV can polymerize into polytubes in the absence of gpH), the polymerization of gpV around gpH may present a more energetically favorable pathway for assembly. During the polymerization of gpV, gpH also seems to prevent gpU from prematurely terminating polymerization. It has therefore been hypothesized that the N-terminus of gpH protrudes from the growing end of the tail to inhibit gpU attachment (Katsura, 1990). Once gpV has polymerized such that it completely surrounds gpH, gpV polymerization pauses (Katsura, 1976b) and only then does gpU bind to gpV to permanently halt polymerization. After tail termination, gpH is cleaved to gpH*, a version of the protein that is missing approximately 10kDa from its C-terminus (Hendrix and Casjens, 1974; Tsui and Hendrix, 1983; Walker et al., 1982).

Since all long-tailed phages must avoid uncontrolled tail tube polymerization it is not surprising that several commonalities exist between the mechanisms regulating tail length determination in phage λ and those involved in regulating tail length in other phages. First, as described in section 1.3, all long-tailed phages possess a conserved cluster of genes encoding a tape measure protein downstream of a tail assembly chaperone. In addition, mutations have been identified in the contractile-tailed phages Mu, P2, T4, and SPO1 that lead to elongated tails similar to those observed in cells infected by phage λ U mutants (Grundy and Howe, 1985; Lengyel et al., 1974; Parker and Eiserling, 1983; Vianelli et al., 2000). However, the question of whether all of these phages use the same mechanism for tail termination remains open because
their putative tail terminator proteins have not been characterized, and no sequence similarity can be detected among them. As mentioned in section 1.5.2, one indication that the structure of TrPs may be conserved in diverse phages was provided by a strong similarity detected between the NMR structure of gpU and the structure of protein STM4215 (PDB ID: 2GJV) encoded in a prophage element of *S. typhimurium* (Edmonds et al., 2007). However, without knowledge of the biologically relevant quaternary structure of biochemically and genetically verified TrP, it is impossible to arrive at definitive conclusions pertaining to the function of STM4215 and the potential for conservation of tail terminator protein structure and function among diverse phages.

### 1.8 Tube-Like Macromolecular Complexes: Structure and Assembly

Despite a vast diversity in function, commonalities do exist between the structure and assembly of the λ tail and other long-tubular structures, which are presumably unrelated. These findings reinforce the idea that bacteriophage morphogenesis acts as an excellent paradigm for the study of macromolecular complex assembly and broadens the relevance of this research project. Described below is a small selection of systems, which exhibit similarities to phage λ in structure and/or assembly.

#### 1.8.1 Tobacco Mosaic Virus

Tobacco mosaic virus (TMV) is a rod-shaped virus formed from a helical array of over 200 copies of a single coat protein, which must assemble around its ssRNA genome. Much like many bacteriophage structural proteins, the TMV coat protein contains a long unstructured region spanning residues 90-110 (Jardetzky et al., 1978). During assembly, the coat protein interacts with the RNA genome, inducing folding of the loops, and suppressing the steric inhibition of coat protein assembly. This is not unlike the disorder-to-order transitions that occur during capsid morphogenesis in phage λ (Lander et al., 2008) and HK97 (Gertsman et al., 2009). The assembly of TMV is also favored due to a complementary charge distribution between the negatively charged phosphates on the RNA backbone and positively charged
residues in the coat protein (Namba and Stubbs, 1986). A similar role for complementary surface interactions has been proposed in the assembly of long-bacteriophage tails. For example, electrostatic interactions have been implicated in the stacking of gpV hexamers during λ tail assembly (Easterbrook and Bleviss, 1969), suggesting that the top and bottom surface of gpV may be oppositely charged. Atomic detail of the complementary surfaces remains largely unexplored due to the absence of structural data on any tail tube forming protein.

Furthermore, the assembly of the TMV coat protein around the RNA genome suggests that the RNA acts as a template for assembly. This ‘templating’ role in TMV assembly is reminiscent of the putative mechanism by which tape measure proteins function during bacteriophage tail assembly. The TMV utilizes internal repeats within the RNA genome to facilitate assembly around the template. This is similar to the internal repeats observed in the sequences of tape measure proteins of other long tailed phages (Boulanger et al., 2008; Chandry et al., 1997; Mc Grath et al., 2006; Siponen et al., 2009). It should be noted however that internal repeats have not yet been identified in the λ tape measure protein, suggesting that the mechanism by which tail chaperones and tail tube proteins interact with the tape measure protein may not be universal.

1.8.2 The Bacterial Flagellum

Bacterial flagella are specialized macromolecular complexes that allow for bacterial motility. Proteins involved in the assembly, export, and function of flagella form a complicated structure that spans the bacterial cytoplasm, inner membrane, periplasmic space, outer membrane, and the extracellular environment. The major extracellular portion of the flagellum, termed the flagellar filament, is formed from the globular protein flagellin. There are several similarities between the regulatory mechanisms controlling flagellum assembly and those involved in the assembly of long-tailed bacteriophages. First, the assembly of both systems occurs in a unidirectional manner. Much like the baseplate/initiator complex in long-tailed bacteriophages, flagellum assembly requires a structure referred to as the filament cap, which exists at the distal end of the growing filament and acts as a promoter for flagellin assembly (Yonekura et al., 2000). Thus, the filament cap and the baseplate/initiator complex both ensure
that assembly proceeds in only one direction while simultaneously preventing premature assembly. Second, the principal building blocks of each complex have been shown to contain regions of disorder in their monomeric states, which undergo conformational changes upon assembly. In its monomeric form, the flagellin proteins from *E. coli* and *Salmonella* each have ~65 unstructured residues at their N termini and 45 disordered residues at their C termini (Aizawa et al., 1990; Kostyukova et al., 1988; Vonderviszt et al., 1989). The disordered termini, which are highly conserved among bacterial species and strains, fold into compact helical structures only after they encounter the growing end of the filament cap (Aizawa et al., 1990). While the role of disordered regions is not entirely clear, it has been proposed that they play at least a partial role in preventing premature assembly. However, the fact that both flagellin and some bacteriophage morphogenic proteins including the T4 tail sheath protein can spontaneously polymerize *in vitro* (Asakura et al., 1964; Moody, 1967; Wakabayashi et al., 1969) suggests that disordered regions alone may not be sufficient to control the timing of self-association and that other regulatory mechanisms may be in place. This is in agreement with the observation that assembly chaperones are required for the assembly of both bacteriophages and bacterial flagellum. For example, a cytosolic chaperone protein interacts with the disordered C-terminus of flagellin *in vitro* and prevents its polymerization (Vonderviszt et al., 1989). Furthermore, a mutant in the flagellin subunit chaperone, leads to the production of short flagella (Yokoseki et al., 1995). Similarly, the putative λ tail assembly chaperones gpG/gpGT are required for the formation of functional tails.

### 1.8.3 Bacterial Secretion Systems

The type III secretion system (T3SS) is one of six bacterial secretion systems found in gram-negative bacteria that release virulence factors into the extracellular medium or directly translocate them into target cells. The T3S apparatus, called the injectisome, is formed from the concerted action of at least 20 different proteins that form a complex structure resembling the bacterial flagellum (Cordes et al., 2003; Kubori et al., 1998). Given the overall architectural similarities, it is not surprising that at least 10 out of 25 genes encoding the T3SS exhibit strong sequence similarities to components involved in assembly of the bacterial flagellum complex.
The external component of the T3S apparatus is ~60 nm long, hollow, and resembles a needle-like structure. The needle is comprised of hundreds of copies of a single protein that assembles into a helical polymer with an internal diameter of approximately 25 Å. Much like bacteriophage tails act as a conduit for genome delivery, the T3S needle is thought to act as a passageway through which unfolded effectors are transported. Intriguingly, both tube-like structures appear to use a similar mechanism to regulate tube length. In T3S, needle length is proportional to the size of a protein called YscP, suggesting that much like bacteriophage tape measure proteins, YscP acts as a protein ruler determining the length of the injectisome (Journet et al., 2003). Accordingly, needles of undetermined length are assembled in the absence of YscP and while long needles were able to secrete effector molecules as efficiently as wild-type needles, short needles displayed a marked decrease in secretion (Mota et al., 2005). The use of a similar mechanism to regulate tube length by both bacteriophage tails and the T3S injectisome raises the question of whether any similarities exist between the proteins that form the tubular portion of each structure. Monomeric structures of the protein building block of the T3S needle have been determined (Deane et al., 2006; Wang et al., 2007) and reveal a common two-helix bundle coiled-coiled motif that is linked by a conserved PxxP turn. While there are currently no structures of the tail tube forming component from bacteriophages available for comparison, circular dichroism studies on the λ major tail protein indicate that it is comprised of very little alpha-helical content (Katsura and Tsugita, 1977) suggesting that the building blocks of the T3S needle and the λ tail tube are not structurally similar. Remarkably, components of the newly discovered type VI secretion system (Mougous et al., 2006) have recently been shown to be structurally similar to some bacteriophage tail structures (Leiman et al., 2009; Pukatzki et al., 2007) suggesting that additional similarities are awaiting discovery between bacteriophages and bacterial secretion complexes.

1.9 Thesis Objectives

The absence of atomic structural data on a MTP or TTP from any long-tailed bacteriophage has greatly hindered studies aimed at providing mechanistic insight into phage tail assembly and the evolutionary relatedness of long-tailed bacteriophages and other long
tubular structures. In addition, while a monomeric solution structure of the $\lambda$ TrP is available, little insight has been provided into the quaternary structure of the biologically relevant hexameric form of the protein, making it difficult to determine the mechanism by which TrPs exert their function.

The experiments described in this thesis have been aimed at addressing the following questions:
1) What are the mechanisms involved in maintaining gpU and gpV as monomers until their turn in the assembly pathway is encountered?
2) What surfaces of gpU and gpV are involved in protein-protein interactions, both with themselves during oligomerization and with other proteins during incorporation into the $\lambda$ tail?
3) How does the structure and assembly of the $\lambda$ tail relate to other long-tailed phages and/or other large macromolecular complexes?
4) Is the C-terminal domain of gpV functionally important, and if so what is its role?

To address these questions, I have carried out detailed structural and functional studies on the $\lambda$ TrP, gpU, and the $\lambda$ MTP, gpV.

1.10 Organization of Thesis

This thesis has been organized into individual chapters aimed at addressing the questions mentioned above. CHAPTER 2 describes the X-ray crystal structure of gpU in its biologically relevant quaternary structure. This chapter outlines the surfaces of gpU involved in hexamerization, tail termination, and head binding. In addition, through structural and bioinformatic analyses, a conserved mechanism of tail termination among many long-tailed phages is discussed. In CHAPTER 3, the NMR solution structure of the N-terminal domain of the $\lambda$ MTP, gpV, is described. The structural similarity between gpV$_N$ and a previously solved but unpublished structure of a TTP from a contractile-tailed phage and a component from the bacterial type VI secretion system is discussed. In addition, a model for the oligomerization and polymerization of gpV is presented. CHAPTER 4 presents functional data indicating that the C-terminal domain of gpV is biologically important. Furthermore, the solution structure of gpV$_C$
in wild-type form is presented and structural similarities between gpV\textsubscript{C} and other Ig-like domains are discussed and a putative role for gpV\textsubscript{C} in host cell adsorption is proposed. In CHAPTER 5, a destabilizing and a stabilizing point mutant in gpV\textsubscript{C} are introduced. The destabilizing point mutant is shown to be lethal for WT \lambda infections and a potential mechanism by which this single point mutation exerts its dominant negative behavior is described. In addition, the structure of a stabilized form of gpV\textsubscript{C} is presented and compared to the gpV\textsubscript{C}-WT structure. The preliminary structural and functional characterization of each of the mutants is discussed. A summary of the results and conclusions of this study, and future directions are outlined in CHAPTER 6.

Parts of this chapter were reproduced with permission from the following sources:


CHAPTER 2

THE X-RAY CRYSTAL STRUCTURE OF THE PHAGE λ TAIL TERMINATOR PROTEIN REVEALS THE BIOLOGICALLY RELEVANT HEXAMERIC RING STRUCTURE AND DEMONSTRATES A CONSERVED MECHANISM OF TAIL TERMINATION AMONG DIVERSE LONG-TAILED PHAGES

2.1 Overview

Tail terminator proteins (TrPs) play an important role in tail assembly by precisely halting tail polymerization and presenting an interaction surface for phage heads. The λ TrP, gpU, is relatively small (14.8kDa) and possesses surfaces to mediate both self-self and non-self interactions. Interestingly, while gpU multimerizes upon contact with the appropriate tail assembly intermediate, it remains monomeric at high protein concentrations on its own. To fully understand the mechanisms by which gpU performs its multiple functions, knowledge of its biologically relevant quaternary structure is essential. In this chapter, the crystal structures of hexameric gpU-D74A and pentameric gpU-WT are presented and analyzed. Sequence alignment analysis and site-directed mutagenesis were used to show that the hexameric structure is biologically relevant, and its functional surfaces have been delineated. Comparison of the hexameric crystal structure with the previously solved NMR solution structure of gpU (Edmonds et al., 2007), shows that structural changes occur upon multimerization and suggests a mechanism that allows gpU to remain monomeric at high concentrations on its own, yet oligomerize efficiently upon contact with an assembled tail tube. Finally, similarities between the hexameric crystal structure of gpU and the X-ray structure of a putative TrP from a contractile-tailed phage are described. This finding coupled with further bioinformatic investigations has led to the conclusion that TrPs from non-contractile tailed phages, such as λ, are evolutionarily related to those of contractile-tailed phages, such as P2 and Mu, and that many long-tailed phages may utilize a conserved mechanism for tail termination.

Acknowledgments: Dr. Karen L. Maxwell generated the gpU-pET15b vector during her Ph.D. studies. Amanda Liu, a technician in the Davidson lab, performed all of the electron microscopy
experiments on gpU and some of the in vivo complementation assays. It should be noted that all of the in vivo complementation figures presented in this chapter are from experiments that I performed. Nicole Nironha, a summer student whom I supervised, assisted with setting up crystallization screens of gpU-WT. Dr. Matthew Cordes (University of Arizona) aided in detecting transitive homology among TrPs. Dr. Alan Davidson performed the PSI-BLAST searches.


2.2 Materials and Methods

2.2.1 Sample Preparation and Crystallization

The $U$ gene was expressed fused to an N-terminal hexahistidine (6-His) tag using the pET15b vector (Novagen) as reported previously (Edmonds et al., 2007). Protein expression was induced by the addition of isopropyl-$\beta$-D-thiogalactopyranoside (IPTG) to a final concentration of 1mM followed by an overnight incubation at 25 °C. Cells were harvested by centrifugation (4420 x g, 15 minutes), lysed by sonication in 10mM imidazole, 300mM NaCl, and 50mM NaH$_2$PO$_4$ $\cdot$H$_2$O (pH 8), and the cellular debris was removed by centrifugation (9800 x g, 15 minutes). After purification on Ni-NTA agarose, the eluted protein was dialysed overnight at 4 °C in 10mM imidazole, 50mM KCl, and 5mM Tris (pH 7.5) followed by cleavage of the tag with 2.5 units of thrombin from bovine plasma (Sigma) for 24 hours at room temperature. The cleavage reaction was stopped by the addition of phenylmethylsulphonyl fluoride (PMSF) to a final concentration of 0.1mM and was then subjected to a second round of Ni-NTA purification. The cleaved protein, no longer possessing a 6-His tag, flowed through the Ni-NTA resin leaving the uncleaved protein and cleaved 6-His tag bound. The untagged protein was then concentrated to approximately 4 mg/ml and loaded onto a Superdex75 column (Amersham Biosciences) pre-equilibrated in 50mM KCl, and 5mM Tris (pH 7.5). The column
was injected with 0.5ml of protein at a flow rate of 0.5ml/min. Protein fractions were pooled and concentrated using a YM10 centrifrep centrifugal device (Millipore) to 10 mg/ml.

The initial screening for crystallization conditions was performed in parallel on gpU-WT and a single point mutant, gpU-D74A, using commercially available sparse-matrix screens from Hampton Research (Crystal Screen I and II), and QIAGEN Inc (Cryos Suite, PEGs Suite, AmSO4 Suite, and Cations Suite). All screening was carried out at room temperature using the hanging drop phase equilibration technique at a protein concentration of 10 mg/ml and a drop size of 2 µl (1 µl of protein solution (50mM KCl and 5mM Tris (pH 7.5)) mixed with 1µl of precipitant). Diffraction quality crystals of gpU-D74A were obtained first in several conditions directly from the screen, each containing ammonium sulfate.

Selenomethionine (Se-Met) was incorporated into gpU-D74A as reported previously (Lemke and Howell, 2001). The Se-Met incorporated protein was purified as described above, with the addition of 0.5mM DTT to all purification buffers and the incorporation of Se-Met was verified using MALDI-TOF mass spectrometry at the Advanced Protein Technology Center, The Hospital for Sick Children (Figure 11). Crystallization conditions found for native gpU-D74A were not reproducible with the Se-Met incorporated protein and therefore additional screening was performed. Se-Met crystals were grown in 0.2M di-ammonium tartrate, and 2.2M ammonium sulfate (Figure 12A). Crystals were flash frozen in liquid nitrogen without the need for cryo-protection and stored in liquid nitrogen for future data collection. Wild-type gpU crystals were subsequently grown in 0.2M tri-sodium citrate, and 2.2M ammonium sulfate (Figure 12B) using the hanging drop phase equilibration method.
Figure 11. Mass Spectrometry Analysis of Native and Se-Met Incorporated gpU-D74A.

Mass spectra are shown for (A) native and (B) Se-Met incorporated purified samples of gpU-D74A. The expected masses, 14886.4 and 15120.9 Da, respectively, were very close to the experimentally determined masses of the major species in each spectra, 14887.3 and 15121.9 Da, respectively.
Figure 12. Crystals of the Se-Met Incorporated gpU-D74A and Native gpU-WT.  
(A) The gpU-D74A crystals were grown at room temperature in 0.2M di-ammonium tartrate, and 2.2M ammonium sulfate. The scale bar represents 0.38 mm. (B) gpU-WT crystals were grown at room temperature in 0.2M tri-sodium citrate, and 2.2M ammonium sulfate. The scale bar represents 0.15 mm.

2.2.2 Data Collection, Structure Determination, and Model Refinement

Single wavelength anomalous diffraction (SAD) data were collected at the selenium peak (0.9800 Å) on the Se-Met incorporated crystals of gpU-D74A on Beam line X12C National Synchrotron Light Source, Brookhaven, New York. A total of 440 images were collected with 0.5° oscillations and a 20 second exposure time. The data were initially processed on-site with HKL2000 (Otwinowski and Minor, 1997) and the positions of the Se atoms were found using SHELXD (Sheldrick, 2008). The data was subsequently re-indexed, integrated, and scaled using d*TREK (Pflugrath, 1999). After scaling, Bayesian statistics were applied using the DREAR program suite (Blessing, 1989). Data collection statistics are summarized in Table 3. An initial model was built using the model building module in PHENIX (Adams et al., 2002) followed by manual building in Coot (Emsley and Cowtan, 2004). Iterative rounds of refinement and model building were carried out using the programs CNS SOLVE (Brunger et al., 1998) and Coot, respectively. Water molecules were added using the water-pick feature in Coot using a 2.5 sigma peak in the $F_o-F_c$ electron density maps as initial criteria. Each water molecule was checked manually for at least one hydrogen bond to a protein or solvent molecule. Conjugate gradient refinement was used to refine the positions and temperature factors of the water molecules. The progress of refinement was monitored by the
reduction of $R_{\text{cryst}}$ and $R_{\text{free}}$ with 5% cross-validated data. The refinement statistics for the structure are summarized in Table 3.

Data on gpU-WT crystals were collected on an R-axis IV++ Rigaku RUH3R rotating anode generator with Osmic optics and processed with d*TREK (Pflugrath, 1999). The structure was solved by molecular replacement with Phaser (McCoy, 2007) using chain A of the gpU-D74A structure as the search model and refined as described above (Table 3). Each structure was validated with Molprobity (Davis et al., 2007) and PROCHECK (Laskowski, 1993). Unless otherwise noted, all molecular graphics in this chapter were generated using PyMOL (http://pymol.sourceforge.net/)

### 2.2.3 Functional Characterization

Site-directed mutations in the $U$ gene were constructed using the QuikChange mutagenesis method (Stratagene). BL21 (DE3-Δtail) (Gibbs et al., 2004) cells freshly transformed with either empty pET15b vector (negative control), or plasmids expressing WT or mutant gpU were suspended in molten 0.7% top agar and poured onto LB/Agar plates containing 10mM MgSO$_4$. Serial dilutions of a $\lambda$ Uam lysate were spotted in 4µl aliquots on top of the cells and the degree of complementation was scored by observing plaque formation after approximately 24 hours at 37 °C. To probe for a dominant negative phenotype, WT and mutant $U$ genes were sub-cloned into the vector pAD100 (Davidson and Sauer, 1994), which contains a $P_{\text{tac}}$-based, IPTG-inducible promoter. WT $\lambda$ phage was plated on lawns of bacteria containing these plasmids in a manner similar to that described above except that the top agar was supplemented with IPTG to a final concentration of 0.1mM.

### 2.2.4 Electron Microscopy

Samples were prepared by freshly transforming $E.\ coli$ strain 594 containing a $\lambda$Uam$^{413cI_{857}Sam_{7}}$ prophage with either empty pAD100 (negative control), or pAD100-based plasmids expressing WT or mutant gpU. A single colony was chosen and grown at 30 °C overnight in 5ml of LB. The starter culture was inoculated into 100 ml of LB to a final OD$_{600\text{nm}}$ of 0.1. Cells were grown to an OD$_{600\text{nm}}$ ~ 0.4, followed by induction in a 65° C water bath for 1
minute, followed by incubation in a 45 °C water bath for 15 minutes. After induction, cells were grown at 37 °C for 1.5 hours and then harvested by spinning at 1240 x g. Pellets were resuspended in 2 ml LB and 300 µl CHCl₃ was added. Samples were incubated at 37 °C for 20 minutes. Cellular debris was removed by centrifugation and the remaining lysate was applied to the surface of a continuous carbon film coated EM grid made hydrophilic by glow discharge (10 sec, 30 mA). After adsorption for 2 minutes, excess sample was blotted away, the grid was washed three times with water and stained with 2% (w/v) uranyl acetate. Grids were examined with a Hitachi H-7000 microscope at 100 kV with a 30 µA beam current. Images were captured with an AMT 6M 16 bit digital camera at 40000 X or 100000 X magnifications. To ensure that the observed populations were representative of the entire sample, several fields were randomly selected and the number of fully assembled phages, free tails of normal length, free heads, and polytails were counted.

2.2.5 Sequence Alignment and Analysis

The alignment of gpU homologues (shown in section 2.3.2, Figure 16) was generated by performing an iterated PSI-BLAST (Altschul et al., 1997) search on the non-redundant protein sequence database at NCBI using gpU as a query and searching until convergence was reached. This search yielded 54 putative homologues that were aligned in Jalview (Clamp et al., 2004) using the MAFFT (Katoh et al., 2002) sequence alignment program. Once aligned, the sequences were manually edited to eliminate all pairs of sequences that were more than 45% identical. Although most of these sequences were from uncharacterized prophages in a variety of bacterial species, the genes encoding seven out of the nine sequences in the final alignment lay adjacent to genes encoding homologues of gpV the MTP of λ, which supports their identity as TrPs. The search for transitive homology was performed on a database of 93 putative TrPs gathered by iterative PSI-BLAST searches with gpU, STM4215, and Phage Mu gp37, using methods described by (Newlove et al., 2004).

2.3 Results
2.3.1 The X-ray Crystal Structure of gpU

Even though the tertiary structure of gpU was solved previously using NMR spectroscopy (Edmonds et al., 2007), crystallization studies were initiated in the laboratory of Dr. Logan Donaldson with the hope that the conditions of crystallization would induce formation of a biologically relevant multimer as has been observed in other systems (Kim et al., 2002; Yang et al., 2000). Since diffraction quality crystals of gpU wild-type (gpU-WT) were not obtained initially, we also tested mutants that had been investigated previously (Edmonds et al., 2007). After removal of the N-terminal hexahistidine tag, I was able to obtain diffraction quality crystals with one of these mutants, gpU-D74A, in several different conditions containing ammonium sulfate. Ultimately, the structure of this protein was solved to 2.7 Å using selenomethionine-incorporation and the SAD method, and refined to an \( R_{\text{cryst}} \) and \( R_{\text{free}} \) of 0.233 and 0.283, respectively (Table 3).

The asymmetric unit of gpU-D74A contained two almost identical hexameric rings. These rings were 84 Å in diameter with a 30 Å diameter central pore (Figure 13), a size that is very close to the dimensions observed in EM studies of the hexameric rings of gpV that comprise the tail tube (Katsura, 1981), suggesting that they might be biologically relevant. Subsequent to solving the gpU-D74A structure, crystals of gpU-WT were grown and the structure was solved to 2.8 Å using molecular replacement (Table 3). The gpU-WT (PDB ID: 3FZB) and gpU-D74A (PDB ID: 3FZ2) monomeric structures were very similar with an overall RMSD of 0.55 Å over aligned \( \text{C}^{\alpha} \) atoms between chain A and chain C, respectively, using the program FATCAT (Ye and Godzik, 2003). Surprisingly, however, the asymmetric unit of the gpU-WT crystals contained two pentameric rings of gpU with a diameter of 74 Å and a 27 Å central pore (Figure 13B). Despite this difference in stoichiometry, the inter-subunit interfaces seen in the gpU-WT and gpU-D74A structures are almost identical (Figure 13C). It should be noted that in both the gpU-D74A and gpU-WT structures the two rings found in the asymmetric unit are arranged face-to-face across a two-fold non-crystallographic axis and an approximate two-fold non-crystallographic axis with a translation, respectively (Figure 13D). This arrangement is not believed to be biologically relevant because it would produce a closed
structure that would not allow the gpU ring to interact with two completely different surfaces (i.e. the head and the tail).

Figure 13. The Quaternary Structure of gpU.

The cartoon representation of (A) the gpU-D74A hexamer and (B) the gpU-WT pentamer. The subunits of each oligomer are coloured in varying shades of blue. (C) Ribbon representation of the superposition of a dimer from the gpU-WT (yellow) and gpU-D74A (blue) asymmetric units. Several residues buried in the inter-subunit interface in each structure are represented as sticks. (D) The asymmetric unit of gpU-D74A (at left) and gpU-WT (at right) contains two hexameric and pentameric rings, respectively. The subunits in each oligomer are coloured in varying shades of blue and green.
Table 3. Diffraction and Refinement Statistics For gpU-D74A and gpU-WT

<table>
<thead>
<tr>
<th></th>
<th>gpU-D74A</th>
<th>gpU-WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
<td>P 1 2 1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) (Å)</td>
<td>87.37</td>
<td>74.10</td>
</tr>
<tr>
<td>(b) (Å)</td>
<td>125.40</td>
<td>127.99</td>
</tr>
<tr>
<td>(c) (Å)</td>
<td>211.61</td>
<td>82.91</td>
</tr>
<tr>
<td>(\alpha) (°)</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>(\beta) (°)</td>
<td>90</td>
<td>108.92</td>
</tr>
<tr>
<td>(\gamma) (°)</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9800</td>
<td>1.5418</td>
</tr>
<tr>
<td>Resolution (Å)(^a)</td>
<td>46.84-2.70</td>
<td>70.10-2.80</td>
</tr>
<tr>
<td>No. Reflections</td>
<td>563982</td>
<td>141896</td>
</tr>
<tr>
<td>No. Unique Reflections</td>
<td>123413</td>
<td>36086</td>
</tr>
<tr>
<td>Redundancy(^a)</td>
<td>4.57 (4.57)</td>
<td>3.93 (3.93)</td>
</tr>
<tr>
<td>Completeness (%)(^a)</td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
</tr>
<tr>
<td>(R_{merge})(^b) (%)(^a)</td>
<td>0.083 (0.289)</td>
<td>0.099 (0.312)</td>
</tr>
<tr>
<td>Average I/(\sigma(I))(^a)</td>
<td>8.8 (3.7)</td>
<td>8.6 (3.4)</td>
</tr>
<tr>
<td>Refinement statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R_{cryst}/R_{free}) (%)(^c)</td>
<td>23.3/28.3</td>
<td>23.6/30.9</td>
</tr>
<tr>
<td># of protein atoms</td>
<td>12258</td>
<td>9748</td>
</tr>
<tr>
<td># of water molecules</td>
<td>246</td>
<td>109</td>
</tr>
<tr>
<td>RMSD from ideal geometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Estimated coordinate error (Å)(^d)</td>
<td>0.49</td>
<td>0.68</td>
</tr>
<tr>
<td>Average B factor (Å(^2))</td>
<td>42.55</td>
<td>47.39</td>
</tr>
<tr>
<td>Ramachandran (%)(^e)</td>
<td>86.6/13.2/0.1/0</td>
<td>82.6/16.8/0.6/0</td>
</tr>
</tbody>
</table>

\(^a\)Values in parentheses are for the outer resolution shell, 2.80-2.70 (gpU-D74A) and 2.90-2.80 (gpU-WT).

\(^b\)\(R_{merge} = \left(\frac{\sum \sum N \cdot I_{hkl} - I_{hkl}(j)}{\sum N \cdot I_{hkl}}\right)^{1/2}\), where \(I_{hkl}\) is the diffraction intensity of an individual measurement over N datasets.

\(^c\)\(R_{cryst} = \left(\frac{\sum |F_{obs}|-kF_{calc}|\sum |F_{obs}|}\right)^{1/2}\), where \(F_{obs}\) and \(F_{calc}\) are the observed and calculated structure factors, respectively and k is the scale factor.

\(^d\)Obtained from CNS-SOLVE (Brunger et al., 1998)

\(^e\)Determined with PROCHECK (Laskowski, 1993). (Percentage of residues found in the most favoured/additionally allowed/generously allowed/disallowed regions of the plot.)

The interaction of two gpU molecules in the gpU-D74A hexamer buries 1165 Å\(^2\) of its monomeric surface area (Brunger et al., 1998), a degree of burial that is quite typical for
biologically relevant protein-protein interactions (Janin et al., 2007). Eleven hydrophobic/aromatic and eight charged/polar amino acid side chains are significantly buried by hexamer formation as shown in Figure 14A. A potentially important network of electrostatic and hydrogen bond interactions occur between the side chains from Ser75, Asp78, and Tyr109 from one monomer, and His3, Thr4, Arg7, and Asp25 from an adjacent monomer (Figure 14B). The hexameric structure may also be stabilized by an inter-subunit antiparallel β-sheet formed between a short segment of β2 (residues 46-48) and a segment of β4 (residues 105-107) in a second monomer. An intermolecular backbone H-bond also occurs between positions 25 and 74. As can be seen in Figure 13C, the orientation of the side chains buried in the inter-subunit interface are almost identical in the gpU-WT and gpU-D74A structures.

**Figure 14. The Oligomerization Interface of gpU.**

(A) Nineteen residues are found buried in the inter-subunit interface of the gpU-D74A hexamer. Interface residues that are at least 80% conserved among gpU homologues are coloured red while residues that exhibit less than 80% conservation are coloured yellow. (B) An electrostatic and hydrogen bond network is found between residues His3, Thr4, Arg7, and Asp25 from one monomer (coloured blue), and Ser75, Asp78, and Tyr109 from a second monomer (coloured green). Residues participating in contacts at the interface were determined using CNS (Brunger et al., 1998) and PISA (Krissinel and Henrick, 2007)
Superposition of the 12 molecules of gpU-D74A in the asymmetric unit revealed minimal structural deviations (back bone RMSD 0.77Å ± 0.22Å) in the regions possessing regular secondary structure elements; however, three loop regions spanning residues 28 to 36 (β1-β2 loop), 50 to 57 (β2-β3 loop), and 111 to 117 (β4-β5 loop), did not overlay well (Figure 15A). Poorly defined electron density and higher than average temperature factors were also noted in these regions of the model. The moderate resolution and somewhat high R_{cryst} and R_{free} observed for this structure might be attributed to the presence of these flexible loops. The monomeric structure of gpU seen in both the wild-type and D74A crystal forms possesses the same mixed alpha and beta fold as was observed in the previous NMR study (Edmonds et al., 2007). This fold consists of a four-stranded anti-parallel β-sheet packed on one side against 3 α-helices (Figure 15A). Both the gpU X-ray and NMR structures also feature an unusual “loop-de-loop” motif (residues 22-37) joining α1 and β2 where the short β1 strand (residues 22-25) is joined to β2 with a circular loop resulting in a parallel interaction between β1 and β2 (Figure 15B). In spite of their overall similarity, significant deviations were observed between the NMR and X-ray structures such that the optimal overlay of all residues in the two structures yields a very high backbone atom RMSD of 5.4 Å (Figure 15C). Interestingly, almost all of the major deviations between these structures occur in the bottom quarter (as oriented in Figure 15C) with each of the secondary structure elements in this region (α1 and β2-β5) being shortened by between 3 and 5 residues in the NMR structure (Figure 15D). The most extreme structural deviations are seen just past the end of α3 where 15 Å separates the Thr100 position in the two structures, and in the loop between β2 and β3 where the NMR structure displays a large (residues 43-60) poorly constrained loop, while the X-ray structure possesses a much tighter turn in this region spanning far fewer residues (residues 49-56) (Figure 15C; Figure 15D). The differences between the NMR and X-ray structures are unlikely to be due simply to one of these structures being poorly determined since greater than 60% of their residues can be overlaid well to produce a backbone atom RMSD of only 1.9 Å, and the large structural changes are localized (Figure 15C). Interestingly, the regions of greatest deviation are involved in the formation of the inter-subunit β-sheet interaction mentioned above. It may be the oligomerization process and formation of these inter-subunit H-bonds that induces the structural differences seen between the monomeric NMR structure and the oligomeric crystal structure.
Figure 15. Comparison of the NMR and X-ray Structure of gpU.

(A) The superposition of the 12 monomers in the asymmetric unit of gpU-D74A (at left) with the β1-β2, β2-β3, and β4-β5 loops circled in orange. At right, the cartoon representation of a gpU monomer from the X-ray structure allows for easy visualization of the mixed α/β fold. The sheet (β1, β2, β3, β5, β4), primarily composed of anti-parallel strands with the exception of β1 which runs parallel to β2, is flanked on one surface by 3 α-helices and one 3₁₀ helix. Secondary structure elements are labeled on the gpU-D74A structure where strands are coloured in dark blue, the helices are in cyan, and loops are coloured grey. The two structural representations are related by ~180°. (B) The “loop de loop” motif of a gpU monomer is highlighted in red. (C) Overlay of the lowest energy NMR structure (red) with one chain from the X-ray crystal structure (blue). Thr100 is highlighted with spheres in each structure. The β2-β3 loop is coloured green in the X-ray structure and yellow in the NMR structure. (D) Secondary structure comparison between the NMR (red) and X-ray (blue) structures as defined by DSSP (Kabsch and Sander, 1983). The “loop de loop” motif is boxed. In panels A-C, the top and bottom is
noted with respect to how each molecule is oriented with respect to a hexamer. Structures should not be compared between panels as they are each shown in different orientations.

2.3.2 Conservation Analysis of an Alignment gpU Homologues

As a means to verify the biological relevance of our crystal structure and to identify key residues in gpU, an alignment of gpU homologues from a variety of prophages was constructed. The sequences in this alignment were selected for maximal diversity from a much larger collection of sequences identified through an iterated PSI-BLAST search (Altschul et al., 1997). Due to the increase in sequenced bacterial genomes, we were able to construct a much more diverse and informative alignment than was shown previously (Edmonds et al., 2007) such that the average pairwise identity between sequences in this alignment was 25%, and no pair of sequences was more than 45% identical. A strong correlation between the conserved features in this alignment and our structure of gpU can be detected. For example, more than 80% of the 29 hydrophobic core positions in our structure displayed high conservation of hydrophobic residues (Figure 16). Importantly, greater than 50% of the 19 positions occupied by residues significantly buried in the inter-subunit interface showed strong conservation. In contrast, conservation was seen at only 14% of the other surface exposed positions of gpU. The high degree of conservation of residues lying at the subunit interface provides strong support for the biological relevance of the hexameric structure of gpU. A smaller number of conserved positions were found on the top and bottom surfaces of the ring, and another conserved group of residues form a buried cluster including residues in the loop-de-loop region on the inner surface of the gpU ring (Figure 16). Strikingly, no conserved residues were found on the outside of the ring as would be expected since these residues are not part of any interaction surface.
Figure 16. Alignment of gpU Homologues.

The alignment was generated as described in Materials and Methods. The average pairwise identity of these sequences is 25% with no pair of sequences being more than 45% identical. Yellow and red circles above the alignment denote residues found in the hydrophobic core and the oligomerization interface, respectively. Additionally, several residues on the top (T) and bottom (B) surface of the gpU hexamer are noted. Positions that show conservation in 80% of the sequences are highlighted. Hydrophobic residues are coloured light blue, aromatic residues are green, positive residues are red, and negative residues are magenta. Aromatic residues at predominantly hydrophobic positions were considered to be conserved.

2.3.3 Characterization of the Functional Surfaces of gpU through Site-Directed Mutagenesis

In order to fulfill its function, gpU must be able to hexamerize on top of the tail tube, as well as interact with λ heads. Therefore, at least three surfaces of gpU are functionally relevant: the hexamerization interface, the tail binding surface, and the head binding surface. We used site-directed mutagenesis to further delineate the location of these functional surfaces.

A notable feature of the gpU inter-subunit interface is the complete conservation of both Arg7 and Asp78 (Figure 14; Figure 16). A gpU mutant carrying the D78A substitution was previously shown to be resistant to Mg$^{2+}$-induced oligomerization and was completely inactive in an in vivo functional assay (Edmonds et al., 2007). In this work, I substituted Arg7 with either Ala or Glu. Plasmids expressing either of these mutants were unable to complement a λ phage $U^\text{am}$ in vivo (Figure 17A) demonstrating that these substitutions abrogate biological
These data are consistent with Arg7 and Asp78 playing a key role in gpU oligomerization.

To determine which surface of the gpU ring binds to heads, and which to tails, residues located on the upper and lower surfaces of the gpU hexamer (as indicated in Figure 15B) were substituted. As can be seen in Figure 17A, substitutions to residues on both surfaces resulted in complete abrogation of their ability to complement the λ Uam mutant. Strikingly, the two substitutions on the top surface of the hexamer, Mut 29-33 (residues 29-33 AVFDE were substituted with KSGKS) and R111E, also inhibited the growth of WT λ phage (Figure 17C). This behavior suggested that these mutants were still able to interact with tails, but not with heads; thus, they could compete with WT gpU for incorporation into tails. Once incorporated, however, they were unable to mediate head-binding leading to an inactive tail. In contrast, the D54K substitution, which is on the other surface of the gpU ring, did not elicit an inhibitory effect over WT phage (Figure 17C), presumably because it prevents the interaction of gpU with tails.
Figure 17. Functional Characterization of the Hexamerization, Head Binding, and Tail Termination Surfaces in gpU.

(A) Plaque formation was used to assess the ability of gpU mutants (R7A, R7E, Mut29-33, R111E, and D54K) expressed from a plasmid to complement a λ Uam phage. Empty vector was used as a negative control and gpU-WT was used as a positive control. Four different dilutions of Uam lysates were used. (B) The cartoon representation of three monomers from the gpU-D74A hexamer is shown to demonstrate where each of the tested mutations is located. Residues 29-33 (Top) are displayed in yellow spheres, Arg111 (Top) in orange spheres, Asp54 (Bottom) in green spheres, and Arg7 (Inter-subunit) in red spheres. (C) Dilutions of a WT λ phage lysate were spotted on cells containing plasmids overexpressing gpU-WT, gpU-Mut29-33, gpU-R111E, and gpU-D54K. Empty vector and gpU-WT were used as controls.

To gain further insight into the effects of these gpU mutants, lysates of the λ Uam phage produced in the presence of plasmid-mediated expression of gpU were examined by transmission electron microscopy. As expected, when the λ Uam lysate was produced in the presence of a plasmid vector control, greatly elongated tails (polytails) were produced while expression of gpU-WT resulted in lysates containing normal length tails and complete phage (Figure 18). Mostly polytails were observed in lysates produced in the presence of the D54K mutant, which lies on the lower surface. In contrast, predominantly normal length tails were seen for the upper surface mutants (Figure 18). Very few intact phage were seen in any of the U mutant lysates as would be anticipated since these mutants are unable to mediate the assembly of infectious phage particles. These data support our hypothesis that mutants on the top surface are still able to bind to tails and thereby maintain proper tail length, but can’t bind to heads. On
the other hand, the D54K mutant cannot halt tail polymerization, most likely because it can’t bind to tails.

Figure 18. Identification of the Head and Tail Binding Surfaces of gpU.

Representative electron micrographs of phage particles produced when a λ Uam phage was complemented with gpU-WT, empty vector, gpU-Mut29-33, or gpU-D54K. The average percentage of intact phage and polytails observed over three independent sample preparations is displayed with each micrograph. The number of particles counted for gpU-WT, empty vector, D54K, and Mut-29-33 is 211, 298, 274, and 238 respectively. Four of the images were taken at 40000 X magnification while a higher magnification (100000 X) image is also displayed for gpU-Mut29-33. The scale bars represent 500nm or 100nm for images taken at magnifications of 40000 X or 100000 X, respectively.

2.3.4 A Putative Role for the λ TrP in Genome Injection

Since gpU is located at the head proximal end of the λ tail, and the λ genome is positioned within the top third of the tail prior to host cell infection (Thomas, 1974), λ DNA must therefore be poised within the gpU pore prior to genome injection and then passes through the pore and the tail tube during infection. In the crystal structure, the gpU-D74A central pore
has an approximate internal diameter of 34 Å. Interestingly, on the head proximal side of gpU, a loop formed from residues 111 to 118, protrudes into the DNA injection pore, diminishing its diameter to ~ 23 Å. Higher than average temperature factors are observed for the residues within this loop, suggesting that the loop may be dynamic. It is therefore feasible that the loop flips out of the way upon λ head binding. Since the diameter of a dsDNA helix is ~ 20 Å, λ DNA could be accommodated within the gpU pore with either the open or truncated diameter.

Remarkably, the electrostatic surface charge distribution of the injection pore clearly demonstrates that the head proximal side (top) of the pore is positively charged while the tail proximal side (bottom) of the pore is negatively charged (Figure 19A). The patch of positive charge is due to three arginine residues (Arg27, Arg110, and Arg111) found lining the top surface of the gpU injection pore (Figure 19B). This feature may attract the negatively charged phosphates on the DNA backbone to enter the top portion of the tail tube and may hold the DNA in place until the appropriate signal for DNA injection triggers genome injection. Consistent with this hypothesis is the presence of 12 sulphate ions (one ion per monomer), modeled in at 50% occupancy in my crystal structure, bound by Arg110 (Figure 19B). Interestingly, Arg110 is not required for gpU function but Arg 111, when mutated to a Glu abolishes gpU activity (Figure 17). Since Arg111 is located on the top surface of gpU and in the DNA injection pore, it is feasible that Arg111 holds a dual role in both head binding and DNA injection.

The signal for DNA injection in SPP1, a λ like phage, has been identified as structural rearrangements in the major tail protein which traverse the length of the tail, from the tail tip to the head, much like a domino reaction (Plisson et al., 2007). Since the overall structure of the tails of SPP1 and λ are similar, it is likely that a similar mechanism occurs during λ DNA injection. Conformational changes in the internal channel of the λ tail tube could alter the position of one or more of the three arginine residues lining the gpU pore, reducing the electrostatic attractions between the DNA and gpU. In addition, the structural changes could also alter the positioning of the acidic pore residues, moving them into close proximity of the DNA, allowing electrostatic repulsions to assist in the DNA injection process.
Figure 19. The DNA Injection Pore of λ gpU.
(A) Electrostatic surface charge distribution of DNA injection pore in gpU-D74A. A ring of positive charge is observed at the top surface of gpU-D74A, while a ring of negative charges is observed at the bottom surface. (B) Three arginine residues (Arg27, Arg110, Arg111) are found lining the top surface of the DNA injection pore. Interestingly, a sulphate ion is bound by Arg 110 in every monomer in the gpU-D74A asymmetric unit.

2.3.5 Structural and Sequence Similarities between gpU and the TrPs of Contractile-Tailed Phages

A VAST search (Gibrat et al., 1996) for structural neighbors of the gpU crystal structure in the PDB identified 22 similar structures in the medium redundancy dataset. However, the only two highly significant (i.e. an alignment length greater than 70 residues with a P value less than .001) hits were the gpU NMR structure (PDB ID: 1Z1Z), and the structure of a hypothetical protein, STM4215 from S. typhimurium LT2 (RefSeq ID: NP_463080; PDB ID: 2GJV). The high similarity of the STM4215 structure to the gpU NMR structure was previously reported (Edmonds et al., 2007). The gpU monomeric crystal structure and the STM4215 structure can be overlaid with an RMSD of 2.5 Å over 115 equivalent Cα positions (Figure 20A), which is striking given that they share only 10.4% sequence identity over the structural alignment. Among the 12 identical residues between gpU and STM4215 are 6 conserved hydrophobic core residues (Figure 20B). The most intriguing similarity between the structures lies in their quaternary structure. STM4215 crystallized as a ring-shaped hexamer with very similar dimensions (84 Å in diameter with a 32 Å diameter central pore) to that formed by gpU-D74A (Figure 20C). Most features of the gpU-D74A and STM4215 rings are very similar with each having β-sheets positioned in the interior of the ring and the helices flanking. In addition,
the inter-subunit interactions are predominantly made by residues in the same structural positions in both structures.

The striking tertiary and quaternary structural similarity between gpU and STM4215 suggests that STM4215 may also function as a TrP. This supposition is supported by our prior identification of ORFs encoded in the STM4215 genomic vicinity that are significantly similar to proteins found in the contractile tail of *E. coli* phage P2 (Edmonds et al., 2007). Further bioinformatic analysis showed that the proteins encoded in the genomic region adjacent to STM4215 are more similar in sequence and arrangement to the tail genes of BcepMu, a contractile tailed phage of *Burkholderia cepacia* (Summer et al., 2004) (Figure 20D). STM4215 is 28% identical to gp37 of BcepMu, which is the eponymous member of the “gp37 superfamily” found in the Pfam (Sammut et al., 2008) conserved domain database (pfam0964). Although the region around STM4215 appears to encode all the proteins required to construct a contractile tail, there are no genes nearby encoding homologues to any phage head morphogenesis protein. This situation is reminiscent of operons that encode phage-tail derived bacteriocins, such as those commonly found within *Pseudomonas aeruginosa* and known as R- or F-type pyocins (Nakayama et al., 2000). Interestingly, one of the most closely related homologues of STM4215 (42% identity) is found within a gene cluster encoding a characterized tail-like bacteriocin that is produced by *Erwinia carotovora*, called carotovoricin (Yamada et al., 2006; Nguyen et al., 1999). The other tail components encoded adjacent to STM4215 are also closely related to Carotovoricin subunits. Taken together, these data strongly suggest that STM4215 performs a TrP function in the production of a hitherto uncharacterized *S. typhimurium* contractile tail-like bacteriocin.
Figure 20. Comparison of Tail Terminator Proteins from Contractile and Non-Contractile Tailed Phages.

(A) Superposition of gpU-D74A chain C (blue) with STM4215 chain A (green). Residues that are 100% conserved over the structural alignment are coloured red. (B) Sequence alignment between gpU-D74A and STM4215 based on the VAST (Gibrat et al., 1996) structural overlay. Conserved hydrophobic core positions are coloured yellow, conserved interface positions are in red, and other conserved surface residues are coloured blue. (C) gpU-D74A (left) and STM4215 (right) form hexamers with similar dimensions and structural arrangements such that strands (blue) are on the interior of the ring, and helices (red) are on the exterior. (D) A comparison of gene arrangement between a phage tail-like entity in the Salmonella typhimurium chromosome with the tail gene regions from the contractile-tailed phages BcepMu, P2, and Mu. The names above the genes refer to P2 tail proteins. Tail terminator genes are colored black and other putative homologues as detected using Hidden Markov Models from the Pfam (Sammut et al., 2008) database are designated by colouring.

Although the similarities in structure and genomic position of gpU and STM4215 provide strong evidence for a common structure and evolution, there is still no significant sequence similarity between these two proteins. For this reason, we performed further bioinformatic studies aimed at connecting the sequences of these two proteins. Exhaustive iterative PSI-BLAST searches were performed using gpU and STM4215 as queries. Searches were also performed with bacteriophage Mu gp37 as a query since a mutation likely located within the gene encoding this protein led to a long-tail phenotype (Grundy and Howe, 1985; Morgan et al., 2002). Phage Mu was also of particular interest because the genes in the STM4215 region are arranged in a manner similar to this well characterized phage (Figure 20D). Once a diverse collection of putative TrPs was assembled using PSI-BLAST, bioinformatic connections were sought between these proteins by searching for transitive homology (Bolten et al., 2001; Newlove et al., 2004). This procedure yields a series of
sequences connected to one another between distantly related sequences (Figure 21). In this way, we were able to connect the sequences of STM4215 and gpU with five intermediate sequences. Each connecting intermediate sequence was at least 24% identical to the two adjacent sequences in the chain. Using the same approach Mu gp37 could be connected to gpU in four steps, and it could be connected to STM4215 in six steps. A sequence alignment containing all of these connecting sequences shows conserved positions among all of these putative homologues, particularly the hydrophobic positions within regions aligning to the secondary structure elements of gpU and STM4215 (Figure 22). The identification of all of these proteins as TrPs is supported by the positioning of all of them within clusters of genes encoding phage tail proteins.

Figure 21. Sequence Network of Tail Terminator Proteins from Contractile and Non-Contractile Tailed Phages.

The network of similar sequences connecting the sequences of λ gpU, STM4215, and phage Mu gp37. Sequences derived from non-contractile tailed phage are circled in red, contractile-tailed phage with P2-like gene arrangements are circled in green, and those with Mu-like gene arrangements are circled in blue. The sequence identity between direct connections in the network is noted as a percent above each line.
Figure 22. Sequence Comparison of Tail Terminator Proteins from Contractile and Non-Contractile Tailed Phages.

Sequence alignment of gpU, STM4215, phage Mu gp37 and phage P2 gpR. All of the linking sequences shown in Figure 21 were aligned and colored using ClustalX (Thompson et al., 2002) as implemented in Jalview (Clamp et al., 2004). An initial alignment was produced using MAFFT. This initial alignment contained the correct alignment of gpU and STM4215 according to their structures for all regions except for the two last strands. These regions were adjusted manually using the structures as a guide. Phage P2 gpR was aligned to the other sequences manually such that the predicted secondary structure of gpR aligned with the secondary structures of gpU and STM4215. The secondary structure elements of STM4215 and gpU are displayed above and below their sequences, respectively. The predicted secondary structure of phage P2 gpR is shown under its sequence. In each case, helices are in red and strands are in blue. The C-terminus is truncated in some sequences due to a lack of conservation.
A surprising feature of the sequences connecting gpU, STM4215, and Mu gp37 is that all except the one closest to gpU appear to be from contractile-tailed phages as judged by the fact that they are found adjacent to ORFs encoding homologues of tail sheath proteins. Most of these putative TrPs are from prophage regions where the tail genes are arranged as is seen in phage Mu with their TrP-encoding genes lying 2 ORFs upstream from the gene encoding the tail sheath protein. However, four proteins comprising the central hub of the connection network lie in prophage regions that are arranged like phage P2, which is very different from phage Mu (Figure 20D, Figure 21). In spite their P2-like arrangement none of these putative TrPs could be connected to a protein of phage P2 through exhaustive PSI-BLAST searches. However, gpR, the likely TrP of phage P2 as determined by mutational analysis (Lengyel et al., 1974) does possess a predicted secondary structure that is similar to gpU and STM4215, and a plausible sequence alignment could be made with gpR and the other putative TrPs in Figure 22. The only large gap in this alignment corresponds to the loop-de-loop region of gpU and STM4215, so gpR may not possess this feature. It is notable that all of the putative TrPs from contractile-tailed phages, including gpR, possess extra sequences at their C-termini that are variable.

2.4 Discussion

In this chapter, I have used X-ray crystallography to determine the structure of an oligomeric form of gpU, the TrP of phage λ. Although the oligomers obtained in this study were induced through the process of crystallization, I have strong evidence that they are biologically relevant and not an artefact of crystallization. The dimensions of hexameric gpU-D74A structure (~ 84 Å in diameter with a 32 Å central pore) match closely to the dimensions observed for hexameric gpV in the λ phage tail (~ 90 Å in diameter with a 30 Å central pore (Katsura, 1981)), the protein to which gpU binds to terminate tail polymerization. The veracity of the oligomeric arrangement of gpU seen in the crystal is also supported by our alignment of diverse gpU homologues where a high degree of conservation was observed for most residues in the inter-subunit interface. Furthermore, substitutions of two of these conserved interface residues abrogated biological function. Finally, the hexameric structure of gpU is very similar to the structure of STM4215, a S. typhimurium prophage-derived protein that is most probably also a TrP. It is improbable that two homologous proteins would fortuitously crystallize into such
similar hexameric ring structures. It is notable that other proteins that are monomeric in solution have been seen to form biologically relevant oligomers upon crystallization. For example, the phage λ head protein, gpD, is also monomeric at high concentrations in solution, yet was induced into its biologically relevant trimeric state by crystallization (Yang et al., 2000). An unexpected result was that while the gpU-D74A mutant crystallized in a biologically relevant hexameric ring, wild-type gpU crystallized in a pentameric ring even though the interactions at the inter-subunit interfaces in the two structures were almost identical. In the in vivo situation where gpU only oligomerizes upon contact with the hexameric gpV found in the tail tube, we expect that the templating role of the gpV rings would result in the formation of only hexameric gpU. However, when gpU oligomerizes on its own, there is clearly a small degree of plasticity within the inter-subunit interfaces that allows a smaller ring to be formed by the wild-type protein. This situation is similar to the portal protein of SPP1, which crystallized on its own as a 13-mer ring, but assembles as a 12-mer ring within the phage particle (Jekow et al., 1998; Lurz et al., 2001).

An unusual feature of gpU is its ability to convert from a monomeric form to a ring-shaped hexamer upon addition of Mg\(^{2+}\) (Katsura and Tsugita, 1977). This ring is similar in size to the ring observed in our crystal structure. In previously published work, several mutants were identified that did not display Mg\(^{2+}\)-induced oligomerization as assessed by NMR (Edmonds et al., 2007). As might be expected, two of the mutated positions (Asp74 and Asp78) map to the inter-subunit interface defined by the crystal structure, indicating that Mg\(^{2+}\)-induced oligomerization occurs through the same interactions as the biologically relevant hexamerization. However, the other mutants affecting this process did not map to the inter-subunit interface with two mutants (Mut 29-33 and D35A) located on the top surface of the ring. Since gpU is a very acidic protein (pI ~ 4.2), the role of Mg\(^{2+}\) in encouraging oligomerization may be to bind and neutralize clusters of acidic residues on the gpU surface, which would increase the likelihood of gpU molecules contacting one another. The substitutions at some positions that abrogate Mg\(^{2+}\)-induced oligomerization may do so by disrupting acidic clusters that can bind Mg\(^{2+}\). For reasons described previously (Edmonds et al., 2007), we do not believe that the gpU-Mg\(^{2+}\) interaction plays a physiologically relevant role.
Knowledge of the biologically relevant hexameric structure of gpU has allowed us to identify not only its inter-subunit binding interface, but also its likely head- and tail-binding surfaces. Since the λ tail is composed of stacked hexameric rings of gpV, it follows that the hexameric ring of gpU must stack upon the top ring of gpV in the tail. Thus, one side of the gpU ring must comprise a gpV binding surface while the other side must interact with the head (Figure 17B). By assessing the in vivo activity of amino acid substitutions on these two gpU surfaces, we found that substitutions on the top surface produced inactive proteins that were also able to inhibit wild-type phage (Figure 17C). This result implied that these mutants were still able to bind to tails, but blocked head binding. The bottom surface, which was assessed through characterization of the D54K substitution, is likely the surface interacting with gpV.

Interestingly, comparison of the gpU crystal structure to the NMR structure showed large structural deviations primarily in the region comprising the bottom surface of the gpU ring. It is possible that the altered structure of gpU in this region prevents it from multimerizing when in a monomeric form, but contact of this surface with hexamerized gpV in the tail induces a significant conformational change that primes gpU for further self-interaction. The gpU N-terminus, which is located near the bottom surface, provides a clear example of changes seen between the monomeric NMR structure and hexameric crystal structure. In the NMR ensemble the first 5 residues at the N-terminus of gpU are imprecisely determined, suggesting that they are likely disordered in solution. In contrast, the N-terminus of gpU in the crystal structure starting at residue 2 displays clearly defined electron density and lower than average temperature factors, and several residues in this region are found buried in the hexamerization interface. The disorder-to-order transition in this region upon hexamerization, which is implied by the differences seen in the NMR and crystal structures may play an important role in regulating this process.

Unusually large disordered regions have been observed in the structures of several λ morphogenetic proteins (Maxwell et al., 2002; Maxwell et al., 2001) and residues within these regions have been found to be crucial for function in several cases (Maxwell et al., 2000; Yang et al., 2000). In the gpU crystal structure, residues 28 to 36 (β1-β2 loop), 50 to 57 (β2-β3 loop), and 111 to 117 (β4-β5 loop) appeared to be relatively disordered as assessed by their variable conformations in different molecules of the asymmetric unit (Figure 15A), poorly defined
electron density, and higher than average temperature factors. Each of these regions also appeared disordered in the NMR ensemble (Edmonds et al., 2007). Nevertheless, residues within each of these gpU disordered regions are crucial for function (Edmonds et al., 2007) (Figure 15A), and they lie on the head- or tail-interacting surfaces of the gpU ring. We hypothesize that disordered regions in morphogenetic proteins play a crucial role in maintaining them in their monomeric states until their intended binding partners are encountered. In gpU, the disordered state of the head and tail interacting surfaces likely prevents the inappropriate binding of monomeric gpU to monomeric gpV, or to the phage head. I expect that the disordered loops of the gpU hexamer become ordered upon interaction with the phage tail and head just as the gpU N-terminus becomes ordered upon hexamerization. Consistent with this idea, the N-terminal 14 disordered residues of gpD, the head decoration protein of phage λ, become ordered upon interacting with the capsid shell (Lander et al., 2008).

In addition to providing insight into the mechanism of gpU function, the elucidation of the biologically relevant quaternary structure of gpU provides strong support for the hypothesis that TrPs in contractile and non-contractile tailed phages share a common evolution. The primary structural evidence is the very similar tertiary and quaternary structures of gpU and STM4215, the gene of which clearly lies within a cluster of genes encoding a contractile tail. However, the lack of detectable sequence similarity between gpU and STM4215 affords the possibility that their structural similarity could have arisen through convergent evolution. Therefore, to further bolster the divergent evolution of these proteins, we identified putative TrP sequences that could link the sequences of gpU and STM4215 through transitive homology (Newlove et al., 2004; Roessler et al., 2008). In this way, we identified a pathway of five sequences leading from gpU to STM4215 where the pairwise identities of sequences at each step was 24% or greater. The same procedure was also used to link gpU and STM4215 to gp37 of phage Mu. A mutation in gene K of phage Mu leading to a long-tailed phenotype likely maps within the open reading frame of gp37 (Grundy and Howe, 1985), implying that gp37 is a TrP. The gene encoding gp37 lies two ORFs upstream from the gene encoding the tail sheath protein just as is seen in the genomes of most of the other putative TrPs in our transitive homology network (Figure 20D; Figure 21). Although the sequences in this network are mostly distantly related to each other, their alignment shows conserved hydrophobic positions indicative of a
common fold. Taken together, the structural similarity of gpU and STM4215, the sequence similarity uncovered through transitive homology, and the similar genomic position of all the putative TrPs in Figure 20D, provide a compelling argument that all of these proteins are TrPs that diverged from a common ancestor. The common evolution of TrPs in the diverse contractile and non-contractile tailed phages and prophages investigated here suggests that the TrPs in all long-tailed phages may be evolutionarily related. This idea is supported by a previous demonstration that the predicted secondary structures of putative TrPs from a large number of very diverse non-contractile tailed phages and from phage T4 are similar to the secondary structure of gpU (Edmonds et al., 2007).
CHAPTER 3

THE STRUCTURE OF THE PHAGE λ MAJOR TAIL PROTEIN REVEALS A COMMON EVOLUTION FOR LONG-TAILED PHAGES AND THE TYPE VI BACTERIAL SECRETION SYSTEM

3.1 Overview

While chapter 2 has shed light on the mechanisms involved in terminating tail assembly, the factors controlling the self-assembly and polymerization of the tail tube prior to termination remain unclear. Studies aimed at addressing these questions have been greatly hindered by the absence of atomic structural data for a major tail protein from any long-tailed bacteriophage. Presented in this chapter is the solution structure and dynamic properties of the N-terminal domain of gpV (gpV_N), the protein comprising the major portion of the non-contractile phage λ tail tube. While the structure was ultimately obtained using NMR techniques, the efforts undertaken to obtain an X-ray crystal structure of full-length gpV (gpV_FL) and gpV_N will also be described. The gpV_N structure is primarily β-strand in composition and approximately 30% of the entire domain is disordered in solution. The most notable of the dynamic regions is a 29-residue loop connecting β-strands 2 and 3. In this chapter, the results demonstrate that the β2-β3 loop is critical for biological function and that it is likely involved in the stacking of gpV hexamers during tail tube polymerization. Furthermore, a structural similarity between gpV_N and a previously solved yet unpublished structure of a tail tube protein (TTP) from a contractile-tailed phage has been uncovered. This finding provides the first direct evidence of an evolutionary connection between these two distinct types of phage tails. In addition, a remarkable structural similarity is also seen to Hcp1, a component of the bacterial type VI secretion system. The hexameric structure of Hcp1 and its ability to form long tubes are strikingly reminiscent of gpV when it is polymerized into a tail tube. These data coupled with other similarities between phage and type VI secretion proteins support an evolutionary relationship between these systems. Based on structural comparisons, together with mutagenesis
and *in vivo* functional studies, a model for gpV polymerization involving several disorder-to-order transitions is proposed.

**Acknowledgements:** Dr. Karen Maxwell cloned \( gpV_{FL} \) into the pET15b vector during her Ph.D. studies. Amanda Liu, a technician in the Davidson lab, cloned \( gpV_{FL} \) into the pTrc99c vector. Dr. Tara Sprules, the facility manager at the Québec/Eastern Canada High Field NMR Facility collected all of the NMR data used in the determination of the \( gpV_N \) structure. Dr. Ranjith Muhandiram acquired the heteronuclear NOE data at the University of Toronto NMR centre. Dr. Voula Kanelis trained me to process and analyze all NMR data. Mr. Batista Calvieri and Mr. Steven Doyle trained me on the transmission electron microscope at the University of Toronto EM facility. Dr. John Rubenstein’s laboratory provided the carbon coated EM grids for the microscopy experiments and assistance with negative staining. Dr. Alan Davidson performed the PSI-BLAST search with \( gpV_N \).

Parts of the chapter are reproduced with permission from: Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L., Davidson, A. R. *The Phage \( \lambda \) Major Tail Protein Structure Reveals A Common Evolution for Long-Tailed Phages and the Type VI Bacterial Secretion System.* PNAS USA, 2009, Mar 17;106(11):4160-5.

### 3.2 Materials and Methods

#### 3.2.1 Plasmid Construction and Protein Purification

Portions of the \( V \) gene were amplified by polymerase chain reaction from bacteriophage \( \lambda \) DNA and cloned into a pET15b expression vector (*Novagen*), producing an N-terminal 6-His tagged protein. The 6-His tag in this construct could be removed by cleavage with thrombin. A region encoding residues 1-153 was sub-cloned into a modified pET32b expression vector (*Novagen*), in which the enterokinase protease cleavage site was replaced with one cleaved by the tobacco etch virus (TEV) protease. In this manner, \( gpV_N \) was produced as a 6-His tagged fusion with thioredoxin. The thioredoxin and 6-His tag could be removed by cleavage with the TEV protease. Shorter gpV fragments encompassing residues 1-75, 1-66, 75-153, and 75-246
were sub-cloned into pET28a (Novagen), producing N-terminal 6-His tagged protein fragments. For *in vivo* complementation assays, full-length gpV (gpV<sub>FL</sub>) was sub-cloned into a pTrc99c vector (Amann et al., 1988), where transcription of gpV was mediated by a Trp/Lac (Trc) fusion promoter. Quickchange mutagenesis (Stratagene) was used to make all mutations.

For purification, all constructs were expressed in BL21 CodonPlus (DE3)-RP (Stratagene) cells with overnight incubation at 25 °C after induction with 1mM IPTG. The cells were harvested by centrifugation and lysed by sonication in 10mM imidazole, 300mM NaCl, and 50mM NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O pH 8. The resulting supernatant was incubated with Ni-NTA agarose (Qiagen), pre-equilibrated in lysis buffer, for 1 hour at 4 °C while agitating. Subsequent to binding, the resin was washed three times with 10 ml of wash buffer (20mM imidazole, 300mM NaCl, and 50mM NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O pH 8) and the resin was transferred to a column where the protein was eluted with 250mM imidazole, 300mM NaCl, and 50mM NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O pH 8. The proteins were then dialyzed overnight in 10mM imidazole, 250mM KCl, 0.2mM EDTA, 10mM Tris pH 8 before tag cleavage. In the case of gpV<sub>FL</sub> and gpV<sub>1-75</sub>, the tag was removed by incubating the protein with 2.5U of Thrombin (Sigma) at room temperature for ~5 to 8 hours. The reaction was stopped by the addition of PMSF to a final concentration of 0.1mM and the untagged protein was separated from the cleaved tag and any remaining uncut protein by passing the reaction over a Ni-NTA column. The gpV<sub>N</sub> tag was cleaved with TEV protease in the presence of 1mM DTT at room temperature for 24 hours. After tag cleavage the DTT was removed by dialysis and the unwanted cleavage products and TEV protease were removed by another round of purification on Ni-NTA agarose (the TEV protease was also tagged with 6-His). gpV<sub>FL</sub>, gpV<sub>1-75</sub>, and gpV<sub>N</sub> were further purified through Superdex200 (gpV<sub>FL</sub>) and Superdex75 (gpV<sub>1-75</sub> and gpV<sub>N</sub>) columns. For crystallization and domain mapping studies, each column was pre-equilibrated in 10mM Tris pH 8, 250mM KCl, and 0.2mM EDTA. For NMR studies on gpV<sub>N</sub>, the Superdex75 column was pre-equilibrated in 50mM Na<sub>2</sub>HPO<sub>4</sub> pH 6.8, 200mM NaCl. Protein samples were concentrated using a Millipore centrifugal filter device with a 5K (gpV<sub>1-75</sub>) or 10K (gpV<sub>FL</sub> and gpV<sub>N</sub>) nominal molecular weight limit cut off. Figure 23 demonstrates the progress of each of the purification steps for gpV<sub>FL</sub> and gpV<sub>N</sub>.
3.2.2 gpV Domain Mapping

Trypsin and chymotrypsin were independently mixed with either gpV<sub>FL</sub> or gpV<sub>N</sub> in a molar ratio of either 1:1000 or 1:100 of protease:protein. It should be noted that the N-terminal tag on gpV<sub>FL</sub> was not removed prior to these experiments. Each of the samples were agitated at room temperature and aliquots were removed at varying time points and immediately mixed.
with 2X SDS sample buffer in a 1:1 ratio and boiled for 3 minutes. To separate and visualize the proteolyzed fragments, samples were run on SDS polyacrylamide gels and were subsequently stained with Coomassie blue dye or transferred to a polyvinylidene fluoride (PVDF) membrane. The PVDF membrane was stained with Coomassie blue dye, destained, and was then sent for N-terminal Edman sequencing at the Advanced Protein Technology Centre at The Hospital for Sick Children.

Alternatively, the gpVN proteolysis reaction was stopped at nearly 100% completion by adding PMSF to a final concentration of 1mM. Both undigested and digested gpVN samples were independently loaded onto a Superdex200 column pre-equilibrated in 10mM Tris pH 8, 250mM KCl, 0.2mM EDTA. The resulting chromatograms were overlayed for comparison.

3.2.3 Crystallization Screening

gpVFL, gpVN, and the protease-resistant core of gpVN were screened for crystallization conditions using commercially available sparse-matrix screens from Hampton Research (Crystal Screen I and II), QIAGEN Inc (formerly known as Nextal Inc. at the time these studies were carried out) (Cryos Suite, PEGs Suite, AmSO4 Suite, and Cations Suite), Emerald BioSystems (Wizard I and II, Cryo I and II). All screening was carried out using the hanging-drop vapor diffusion technique at protein concentrations ranging from 10 to 20 mg/ml.

3.2.3 NMR Spectroscopy and Structure Calculation

NMR studies on gpVN were carried out at protein concentrations of ~1.0mM in 50mM Na2HPO4, pH 6.8, 200mM NaCl or in 50mM Na2HPO4, pD 7.2, and 200mM NaCl made using 99.9% D2O. All NMR spectra were collected at 20 °C on a Varian INOVA 500 MHz or 800 MHz spectrometer equipped with pulsed field gradients at the Québec/Eastern Canada High Field NMR Facility. NMR data were processed using the NMRPipe software (Delaglio et al., 1995), and were analysed using NMRView version 5.2.2 (Johnson, 2004). Backbone and aliphatic resonance assignments were attained to greater than 90% completion using a combination of standard triple resonance experiments (Kanelis et al., 2001). Aromatic resonances were assigned using (Hβ)Cβ(CγCδ)Hδ, and (Hβ)Cβ(CγCδCε)Hε experiments.
(Yamazaki, 1993) together with a $^{13}$C-$^1$H HSQC centered at 125 ppm in the carbon dimension. Proline rich regions were assigned with $\text{H}\alpha\text{C}\alpha\text{N}$ and $(\text{H}\beta)\text{C}\beta\text{C}\alpha(\text{CO})\text{NC}\alpha\text{H}\alpha$ (Kanelis et al., 2000). Mixing times of 90 ms were used for both the $^{13}$C- and $^{15}$N-edited NOESY experiments. 1069 NOEs were manually assigned and calibrated using the CALIBA module of CYANA 2.1 (Gunert, 2004). Dihedral angle restraints were assigned using TALOS (Cornilescu et al., 1999). Initial structure calculations were performed using CYANA 2.1 (Gunert, 2004) for the automatic assignment of all NOE peaks and structure refinement. Hydrogen bond restraints were assigned manually based on both unambiguous NOE assignments and by assessing the ensemble of structures from the 37th round of CYANA calculations, where structures had been determined using 1283 unambiguous NOEs. In the final iteration, 100 structures were refined and the 20 lowest energy structures were kept for structural analysis. PROCHECK-NMR (Laskowski et al., 1996) was used for structure validation. The RMSD between the 20 lowest energy structures was calculated using MOLMOL (Koradi et al., 1996). Heteronuclear NOE experiments were collected on a 500 MHz spectrometer at the University of Toronto NMR center. The experiments were run twice in an interleaved fashion both with $^1$H saturation by scanning for 5s with a 7s delay between scans and without $^1$H saturation using a 12s delay between scans. NOEs were determined as the ratio of peak intensities between the two datasets (with saturation/without saturation). Errors in the peak intensities were estimated from the average noise levels in each spectrum.

### 3.2.4 In vivo Assay for gpV Activity

BL21 (DE3-Δtail) (Gibbs et al., 2004) cells freshly transformed with either empty pTrc99c vector (negative control), or plasmids expressing WT or mutant gpV were suspended in molten 0.7% top agar supplemented with IPTG to a final concentration of 0.5mM, then poured onto LB/Agar plates containing 10mM MgSO$_4$. Serial dilutions of a λ *Vam* lysate were spotted in 4µl aliquots on top of the cells and the degree of complementation was scored by observing plaque formation after approximately 24 hour incubation at 37 °C. The activity of each mutant was tested in the context of gpV$_{FL}$. 


3.2.5 Electron Microscopy

Samples were prepared by freshly transforming the *E. coli* strain 594 carrying a λ prophage with wild-type morphogenic genes (λcI857Sam7) with either empty pTrc99c, or pTrc99c plasmids expressing gpV-WT or gpV-D61A/D62A. A colony was grown at 30°C overnight in 5ml of LB and was inoculated into 100 ml of KH medium (1g/L NH₄Cl, 1g/L KH₂PO₄, 6g/L Na₂HPO₄·7H₂O, 0.5g/L NaCl, 0.6g/L MgSO₄·7H₂O, 2mg/L FeCl₃, 2g/L glucose, 15g/L Difco Casamino Acids). Cells were grown to a final density of 0.6 to 0.8. Phage production and protein overexpression were induced simultaneously in a 45 °C water bath for 20 minutes in the presence of 0.5mM IPTG. Cells were grown at 37 °C for 3 hours and then harvested by centrifugation at 3000 rpm. Bacterial pellets were resuspended in 1ml KH medium supplemented with 10mM putrescine. Cells were lysed by vortexing in the presence of CHCl₃. Phage particles were isolated from cellular debris by centrifugation at 13000 rpm for 10 min. 4μl of the resulting phage lysates were applied to the surface of a continuous carbon film coated EM grid made hydrophilic by glow discharge (10 sec, 30mA), and allowed to adsorb for 2 minutes. Excess sample was blotted away, washed three times with water and stained with 2% (w/v) uranyl acetate. Grids were examined with a Hitachi H-7000 microscope at 100 kV with 30μA beam current. All images were captured with an AMT 6M 16 bit digital camera at 100000X magnification.

3.2.6 Construction of a Sequence Alignment of gpV₇ Homologues

A PSI-BLAST (Altschul et al., 1997) search was initiated using residues 1-153 of gpV. Convergence was reached after 13 iterations with approximately 250 sequences showing significant similarity to gpV₇. Sequences were displayed in JalView (Clamp et al., 2004) and aligned using the MUSCLE (Edgar, 2004) sequence alignment program. After alignment, average distance trees were built in JalView to identify related groups of sequences within the alignment. The number of sequences in the alignment was iteratively reduced by taking single representatives of identified sequence subfamilies. Ultimately a set of sequences with maximal diversity that could still be aligned easily by automated procedures was produced. The final
alignment was optimized through manual adjustment using JalView. Secondary structure prediction was carried out using JPred (Cuff et al., 1998).

3.3 Results

3.3.1 Biochemical Domain Mapping of gpV

Since the success of crystallization is often dependent on the accuracy of domain boundaries and the absence of flexible linker regions, to commence structural studies it was first necessary to biochemically map the domain boundaries of gpV. Previously, using a bioinformatics approach the boundaries of gpV were estimated to encompass residues 1-153 for the N-terminal domain, and residues 160-246 for the C-terminal domain (Fraser et al., 2006). Here, limited proteolysis and N-terminal sequencing were used to biochemically verify these boundaries. Trypsin proteolysis of gpV_{FL} resulted in the production of four fragments (T1, T2, T3, and T4) (Figure 24). Based on their migration in an SDS polyacrylamide gel, bands T1 through T4 were estimated to be 18, 17, 13, and 10.5 kDa in size respectively (Figure 24). When gpV_{N} was digested with trypsin, two stable bands (gpV_{N}-T1 and gpV_{N}-T2) remained after 90 minutes. gpV_{N}-T1 and gpV_{N}-T2 were estimated to be 8.4 and 6.5 kDa, respectively.
Figure 24. Limited Trypsin Proteolysis of gpV\textsubscript{FL} and gpV\textsubscript{N}.

(A) 2.0 mg/ml gpV\textsubscript{FL} and (B) 3.6 mg/ml gpV\textsubscript{N} were proteolyzed with trypsin in a (A) 1000:1 and (B) 100:1 molar ratio of gpV to trypsin, respectively, for the amount of time indicated underneath the lanes in each gel. The digested fragments were resolved on a (A) 4-20\% SDS Tris-Glycine gel and a (B) 16\% SDS polyacrylamide Tris-Tricine gel.

Six rounds of Edman degradation determined the amino acid sequence of the N-terminus of each fragment (Table 4) and revealed that band T4 consisted of two overlapping bands of the same approximate size. For ease of visualization, each of the cleaved regions are highlighted within the full-length protein sequence in Figure 25. N-terminal sequencing results indicated that the bands T1, T3 and gpV\textsubscript{N}-T1 arose from proteolysis at the N-terminus of the protein.
within the engineered purification tag. Bands T2, T4A, and gpV\textsubscript{N}-T2 arose from cleavage within the middle region of the putative N-terminal domain, suggesting that a disordered region exists within gpV\textsubscript{N}, and band T4B arose from cleavage in the region that links the N- and C-terminal domains. These experiments were repeated using chymotrypsin and similar regions of the protein were found to be susceptible to proteolysis as were observed with trypsin. The digestion results were consistent with the domain boundaries estimated from bioinformatics analysis (Fraser et al., 2006) with the following exception: a region near the ‘middle’ of gpV\textsubscript{N} is sensitive to proteolysis by both trypsin and chymotrypsin.

Table 4. N-terminal Sequencing of Trypsin Proteolyzed Fragments from gpV\textsubscript{FL} and gpV\textsubscript{N}

<table>
<thead>
<tr>
<th>Band Name</th>
<th>Estimated Size* (kDa)</th>
<th>N-term. Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>18</td>
<td>GSHMPV</td>
</tr>
<tr>
<td>T2</td>
<td>17</td>
<td>SAGDTS</td>
</tr>
<tr>
<td>T3</td>
<td>13</td>
<td>GSHMPV</td>
</tr>
<tr>
<td>T4A</td>
<td>10.5</td>
<td>SAGDTS</td>
</tr>
<tr>
<td>T4B</td>
<td>10.5</td>
<td>STVTAA</td>
</tr>
<tr>
<td>gpV\textsubscript{N}-T1</td>
<td>8.4</td>
<td>GAMMPV</td>
</tr>
<tr>
<td>gpV\textsubscript{N}-T2</td>
<td>6.5</td>
<td>SAGDTS</td>
</tr>
</tbody>
</table>

* Size was estimated based on the bands migration in an SDS polyacrylamide gel
Figure 25. The Trypsin Cleavage Sites in gpVFL.

The gpVFL protein sequence is displayed along with GAM and GSH (boxed in black) which are a part of the engineered N-terminal tag on gpVN and gpVFL, respectively. The first 153 amino acids (the N-terminal domain) are shaded in light blue, and the last 87 residues of the protein (the C-terminal domain) are shaded in yellow. The N-terminus of the T1, T3, and gpVN-T1 fragments identified by N-terminal sequencing are in red text. The N-terminus of the T2, T4A, and gpVN-T2 fragments is in pink text. The T4B cut site is in green text and you can clearly see that it exists between the putative N- and C-terminal domains.

The proteolysis results can be interpreted in more than one way. First, gpVN may be comprised of two independently folded domains connected by a flexible loop or alternatively, a flexible loop may join two regions of gpVN that must remain associated to maintain proper folding of the entire domain. To address this question, constructs encompassing residues 1-75, 1-66, 75-153, and 75-246 were made to represent the possible domain structures identified by limited proteolysis. The vectors encoding residues 1-75 and 1-66 expressed well and were soluble when induced with 1mM IPTG overnight at 25 °C; however, no protein expression was observed for the constructs encompassing residues 75-153 and 75-246. The construct encoding residues 1-75 was $^{15}$N-labeled and a $^{1}H^{15}$N HSQC NMR spectrum was acquired. The resulting spectrum indicated that the protein was unfolded (data not shown), suggesting that the structural integrity of the N-terminal domain may be dependent upon the association of residues 1-75 and 75-153.

To test this hypothesis, gpVN was proteolysed to completion with trypsin and chymotrypsin independently and the digested fragments were separated on a size exclusion
column. The chromatograms of protease-digested gpV_N were superimposable with the chromatogram of undigested gpV_N (Figure 26), suggesting that gpV_N remains intact despite degradation of an internal loop.

Figure 26. Size Exclusion Profiles of Native and Proteolyzed gpV_N.
Undigested gpV_N (blue), and gpV_N digested with chymotrypsin (green) or trypsin (red) were loaded onto a Superdex200 column and the resulting chromatograms were overlaid. In each case, undigested gpV_N elutes at approximately the same volume as digested gpV_N. Fraction 2 from the gpV_N trypsin digested sample was run on a 15% tris-tricine gel (inset – lane 1). Uncut gpV_N is shown in lane 3.

3.3.2 Crystallization Studies on gpV full length and gpV_N

Based on the results obtained in the domain mapping studies, crystallographic studies were performed on gpV_FL, gpV residues 1-153 (gpV_N), and the stable core of gpV_N after proteolysis with trypsin and chymotrypsin. In each case, limited success was obtained. Crystallization of gpV_FL yielded only one promising condition for crystallization. Small crystals with a rod-like needle morphology were obtained in 30% v/v PEG-400, 100mM CAPS pH 10.5 (Wizard I – Condition # 5) (Figure 27). Unfortunately, these crystals were not reproducible and could therefore not be optimized for diffraction studies. Whether or not these crystals were protein or salt remains elusive.
Figure 27. Crystallization of gpVFL.

Crystals were grown 30% v/v PEG-400, 100mM CAPS pH 10.5 using the hanging drop vapor diffusion method.

Crystallization screening on gpVN yielded amorphous crystalline material in two similar conditions: 0.2M ammonium sulfate, 30% w/v PEG 8000, and 0.2M Ammonium sulfate, 20% w/v PEG 4000. While these ‘crystals’ were highly reproducible, their morphology could not be altered to make them amenable to diffraction studies. Figure 28 demonstrates the consistent morphology obtained for gpVN crystals.

Figure 28. Crystallization of gpVN.

gpVN ‘crystals’ were grown in (A) 30% (w/v) PEG 8000, 0.2 M Ammonium Sulfate (B) 28% (w/v) PEG 4000, 100mM Citrate pH 3, 0.2 M Ammonium Sulfate using the hanging drop vapor diffusion method.
A final attempt was made to crystallize gpV\textsubscript{N}, by subjecting the trypsin and chymotrypsin resistant core to crystallization screening. Only the chymotrypsin resistant core yielded crystals in one condition (0.1M sodium acetate trihydrate pH 4.6, 2.0 M sodium chloride). Unfortunately the crystals were small and of poor morphology (Figure 29). Since NMR studies on gpV\textsubscript{N} were being carried out in parallel to crystallization studies, little effort went into optimizing these particular crystals.

![Crystallization of the Chymotrypsin-Resistant Core of gpV\textsubscript{N}](image)

**Figure 29. Crystallization of the Chymotrypsin-Resistant Core of gpV\textsubscript{N}**
Crystals were grown in 0.1M sodium acetate trihydrate pH 4.6, 2.0 M sodium chloride using the hanging drop vapor diffusion method.

### 3.3.2 The Solution Structure and Dynamics of the N-terminal Domain of gpV

Since the crystallization of gpV\textsubscript{N} was largely unsuccessful and the proteolysis studies suggested the existence of unstructured regions within the N-terminal domain, NMR spectroscopy studies were pursued as an alternative tool to obtain atomic structural information on \(\lambda\) gpV, and the solution structure of a gpV\textsubscript{N} construct comprising gpV residues 1-153 was determined. An ensemble of the 10 lowest energy structures for gpV\textsubscript{N} are shown in Figure 30A. The structured regions of the protein are well defined having a backbone RMSD of 0.69 Å (Table 5). However three regions, spanning residues 1-14, 50-78, and 149-153, displayed no long-range NOEs and could be assigned no fixed structure. The disordered nature of these
regions was confirmed through an assessment of $^1$H-$^{15}$N heteronuclear NOEs, which indicated that they are dynamic on the ps-ns timescale (Figure 30B), similar to the behavior of an unfolded protein. Strikingly, these three regions account for ~30% of the entire domain.

The structured part of gpV$_N$ is composed of 7 β-strands and 1 α-helix as indicated in Figure 30A. The secondary structure assignments were confirmed by examination of short-range NOE data, and NMR chemical shift information. The 7 strands are arranged into two antiparallel sheets which fold into a twisted β-sandwich conformation. Sheet 1 is formed from strands 1, 2, 4 and 5, while sheet 2 is formed from strands 2, 3, 6, and 7. The α-helix is located between strands 3 and 4 (Figure 30A). The high quality of this structure is reflected by its well-packed core consisting of 16 hydrophobic residues that display less than 10% solvent accessibility (Figure 30C).
Figure 30. The Solution Structure and Dynamics of gpV\textsubscript{N}.

(A) The superposition of the backbone atoms (N, C\textsubscript{\alpha}, C') of the structured regions (16-20 (\beta_1), 37-46 (\beta_2), 80-86 (\beta_3), 93-102 (\alpha_1), 106-111 (\beta_4), 117-122 (\beta_5), 131-133 (\beta_6), 136-143 (\beta_7)) from the ten lowest energy gpV\textsubscript{N} structures. To the right is a ribbon representation of the lowest energy structure of gpV\textsubscript{N}. (B) Heteronuclear $^1$H-$^{15}$N NOEs for gpV\textsubscript{N} indicate three highly dynamic regions (highlighted in cyan). The y-axis ($^1$H-$^{15}$N NOE) is a ratio of peak intensities between datasets collected with and without $^1$H saturation. Low values are indicative of dynamics on the ns-ps timescale being present while values close to 1 indicate a lack of motion on this timescale. The backbone H-N bond vectors for amino acids in the N and C termini, and \beta_2-\beta_3 loop display dynamics on the ns-ps timescale, while those in secondary structure have very limited motion on this timescale. The secondary structure elements are displayed on top of the $^1$H-$^{15}$N NOE map. (C) The 16 hydrophobic core side chains are represented as yellow sticks. Strands are coloured in dark blue, the helix is in red, and loops are coloured grey.
Table 5. Structural statistics for the ensemble of 20 structures for gpV_N

<table>
<thead>
<tr>
<th>Distance restraints^a</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Range,</td>
<td>i-j</td>
<td>\leq 1</td>
</tr>
<tr>
<td>Medium Range, 1&lt;</td>
<td>i-j</td>
<td>&lt; 5</td>
</tr>
<tr>
<td>Long Range,</td>
<td>i-j</td>
<td>\geq 5</td>
</tr>
<tr>
<td>Hydrogen bond pairs (HN-O, N-O)</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

Dihedral angle restraints

φ/ψ pairs (160 angles) 80

Pairwise RMSD

Secondary structure^b

Backbone atoms 0.69 ± 0.13 Å
All heavy atoms 1.50 ± 0.17 Å

Assigned atoms^c

Backbone atoms 1.26 ± 0.23 Å
All heavy atoms 1.88 ± 0.20 Å

Ramachandran Statistics^d

Most favored regions 75%
Additional allowed regions 24.6%
Generously allowed regions 0.3%
Disallowed regions 0.1%

^a There were no distance violations > 0.5 Å. ^b RMSD for residues 16-20, 37-46, 80-86, 93-102, 106-111, 117-122, 131-133, 136-143. ^c RMSD for residues 16-46, 80-143. ^d Determined using PROCHECK-NMR (Laskowski et al., 1996).

In an alignment of a very diverse set of gpV_N homologues collected through multiple iterations of PSI-BLAST (Figure 31), an almost complete conservation of hydrophobicity is seen at 14 of the 16 core positions, which is remarkable in light of the low sequence identity among these proteins. In addition, secondary structure prediction performed on the whole alignment using the JPred program (Cuff et al., 1998) showed concordance at every secondary structure element in gpV_N except for the short β6-strand. The agreement between the gpV_N structure and the alignment data strongly suggests that all of these homologues adopt a tertiary structure similar to gpV_N.
Figure 31. An Alignment of Diverse Homologues of gpV.

The gpV\textsubscript{N} homologues shown here are representative of a large number of sequences collected through many iterations of PSI-BLAST (see Materials and Methods for details). These sequences were chosen to maximize diversity, and are derived mostly from prophages in a variety of bacteria (sequences are designated by NCBI GI numbers). Sequences taken from characterized phages are indicated. The average pairwise identity of these sequences to gpV\textsubscript{N} is 17% with no sequence being more than 23% identical. The average pairwise identity among all the sequences in the alignment is 18%. No two sequences in the alignment are more than 37% identical. Non-conserved sequences at the N termini of these proteins were truncated (gpV starts at Met 8). In the last two sequences, a large loop after β-strand 4 is truncated. The 14 highly conserved hydrophobic core positions are shaded yellow. Conserved non-polar positions in the structured regions of gpV\textsubscript{N} that are more 20% exposed in the gpV\textsubscript{N} monomeric structure are highlighted in red. Conserved residues within the middle unstructured loop are shaded orange and additional residues that are conserved and found at hexamer interface positions are shaded cyan. Experimentally derived secondary structure elements are indicated on top of the alignment in blue with the computationally predicted secondary structure (Cuff et al., 1998) shown in yellow.
### 3.3.2 The Central Disordered Region of gpV<sub>N</sub> is Functionally Important

Disordered regions are a hallmark of many unassembled bacteriophage morphogenic proteins and the functional importance of these regions has been demonstrated in several cases (Maxwell et al., 2001; Yang et al., 2000). A potentially important feature of the large unstructured region in the middle of gpV<sub>N</sub> (residues 50-78) is its high concentration of negatively charged residues, given that electrostatic interactions were previously implicated in the polymerization of gpV hexamers (Bleviss and Easterbrook, 1971). To investigate the role of this region and the charged residues within it, I substituted two neighboring pairs of Asp residues at positions 56 and 57 as well as 61 and 62 with Ala. In addition, to probe the function of hydrophobic amino acids within the β2-β3 loop I also mutated Tyr at position 59 to Ala. The biological activity of each of these mutants was assessed using an *in vivo* complementation assay whereby a λ phage that is unable to make gpV (*Vam* mutant) is complemented by gpV expressed from a plasmid. While the D56A/D57A had wild-type activity levels, it can be seen that the D61A/D62A and Y59A mutants had significantly reduced biological activity (Figure 32A) suggesting that this region of gpV is critical for its function. Furthermore, the D61A/D62A double mutant also displayed a dominant negative phenotype in that it inhibits the growth of wild-type λ phage (Figure 32B). This effect implies that the D61A/D62A mutant incorporates into tail assembly intermediates, but that its presence inhibits further tail formation. Since gpV remains monomeric until it is incorporated into the tail, this behavior is most simply explained if the mutant is still able to form hexamers, but is debilitated in the hexamer-hexamer interaction required for the formation of a tail tube. In order to test this hypothesis, either wild-type or mutant gpV were overexpressed, while simultaneously inducing a wild-type prophage and the resulting phage particles were examined using transmission electron microscopy. Fully assembled phages were observed in the empty vector negative control indicating that phage induction and assembly was successful (Figure 32C). Overexpression of wild-type gpV resulted in the presence of fully assembled phage and an abundance of free λ tails of wild-type length (Figure 32C). Remarkably, the overexpression of the D61A/D62A mutant resulted in the formation of short tails (Figure 32C). While fully assembled phage and tails of normal length were also observed, the most abundant specimen found on the grid was short tails. This
observation supports the idea that the D61A/D62A mutant inhibits the hexamer-hexamer interactions involved in tail polymerization.

Figure 32. The Role of Disordered Regions in λ Tail Assembly.

(A) The ability of gpV mutants expressed from a plasmid to complement a λ Vam lysate in vivo was scored by observing plaque formation. The “Empty Vector” experiment constitutes the negative control and the “Wild-type V” constitutes the positive control. A D56A/D57A and D61A/D62A double mutant and Y59A single point substitution was also tested. As indicated, three different dilutions of the Vam lysate were spotted onto the plates. (B) The effect of plasmid-expressed WT gpV and each of the mutants on the growth of WT λ phage was assayed. A WT phage lysate was spotted at the indicated dilutions on cells carrying these plasmids. (C) Representative electron micrographs of phage particles produced when a WT λ prophage was induced in the presence of empty vector, overexpressed wild-type gpV, or overexpressed D61A/D62A gpV. The scale bar in each image represents 100nm.
3.3.3 Structural Similarity Between $gpV_N$ and a Tail Tube Protein from a Contractile-Tailed Phage

A search for similar structures in the PDB using VAST (Gibrat et al., 1996) yielded many significant hits. Several of the similar structures were members of the PilZ domain, which binds bis-(3’-5’)-cyclic dimeric guanosine monophosphate (c-di-GMP) and is involved in bacterial signalling (Ramelot et al., 2007). Some other similar structures were members of the pyridoxamine 5'-phosphate oxidase families (Sammut et al., 2008). While these families of proteins do not appear to be functionally relevant to $gpV_N$, one significant hit (PDB ID: 2GUJ, unpublished structure by the Northeast Structural Genomics Consortium) corresponded to XkdM, a protein encoded within the well characterized $B. subtilis$ PBSX prophage, which produces $Myoviridae$ type particles when induced (Mauel and Karamata, 1984). XkdM comprises the tail tube of this particle (Romero et al., 2004). As would be expected for a tail-tube protein, the gene encoding XkdM is located between the gene encoding the tail sheath and tape measure protein of PBSX (Figure 33A). Surprisingly, the open reading frame immediately downstream of XkdM, XkdN, does not appear to encode the conserved frameshift seen in most long-tailed phage, suggesting that this frameshift may not be imperative for all phages.

$gpV_N$ and XkdM possess strikingly similar folds (Figure 33B), and can be overlaid with an RMSD of 1.9 Å over 57 equivalent C$\alpha$ positions. The topology of the two proteins is essentially the same except that $gpV_N$ displays disordered regions at its N and C termini where strands are present in XkdM (Figure 33C). The common core tertiary structure of each consists of two orthogonally packed sheets. Both proteins have large loops between strands 2 and 3. While this loop is completely disordered in $gpV_N$, it forms a structured loop with a helix in XkdM that folds back into the top of the structure (Figure 33B; Figure 33C). This conformation could be induced through the crystallization process, and this region may be disordered when XkdM is in solution. One other difference between the structures is the presence of a helix between strands 5 and 6 in XkdM where $gpV_N$ has a short strand.

The strong structural similarity between $gpV_N$, a non-contractile tail tube protein, and XkdM, a contractile tail tube protein, provides the first direct evidence that these two types of tail tubes are structurally related at the molecular level. Furthermore, this structural similarity combined with the gene order conservation seen in most $Siphoviridae$ and $Myoviridae$ genomes
(Figure 33A) provides compelling evidence of an evolutionary link between the distinctive tails utilized by these phages. The inability to trace the evolutionary link through sequence homology (gpV\textsubscript{N} and XkdM display no significant sequence similarity) might suggest that this structural similarity arises through convergent evolution; however, the lack of sequence similarity is not surprising given that tube proteins within the same phage families often display no detectable sequence similarity. The apparent common structure and evolution of the contractile and non-contractile tail tube proteins raises intriguing questions as to whether the sheath protein evolved to interact with the tail tube of a common progenitor phage, or whether a contractile-tailed progenitor lost its sheath at some point to give rise to the non-contractile tail family.

**Figure 33. Structural Similarity between gpV\textsubscript{N} and a Tail Tube Protein from a Contractile-Tailed Phage.**

(A) Gene order conservation among tail proteins from \(\lambda\) (*Siphoviridae*), HK97 (*Siphoviridae*), P2 (*Myoviridae*), Mu (*Myoviridae*), and PBSX (*Myoviridae*). In each case, the tail chaperones (gpG and frameshift product, gpGT, in \(\lambda\)), coloured purple, are located between the gene encoding the major tail protein (MTP) or tail tube protein (TTP), coloured blue, and the gene encoding the tape measure protein (TMP), coloured orange. In *Myoviridae*, the gene encoding the tail sheath protein (TSP), coloured red, is located upstream of the TTP. It should be noted that PBSX was previously reported to possess the conserved frameshift, but this was an error. (B) Tertiary and (C) Secondary structure comparison between gpV\textsubscript{N} and XkdM, the TTP from PBSX. Each secondary structure element in XkdM is colour-coded based on its tertiary alignment with gpV\textsubscript{N} (\(\beta_1\): salmon, \(\beta_2\): orange, \(\beta_3\): yellow, \(\beta_4\): blue, \(\alpha_1\): green, \(\beta_5\): sky blue, \(\beta_6\): purple, \(\beta_7\): deep pink). Loops and secondary structure elements that do not align are coloured in grey. In panel C, regions that are disordered in gpV\textsubscript{N} are denoted with a red star.
3.3.4 Phage λ Tails are Structurally Related to the Hcp1 Type VI Secretion System Protein

Although the structural similarity between gpV$_N$ and XkdM is striking, an even stronger structural similarity was found between gpV$_N$ and Hcp1 (PDB ID: 1Y12), which is an essential component of a type VI secretion system from *P. aeruginosa* (Mougous et al., 2006). Hcp1 and gpV$_N$ can be overlaid at 74 equivalent C$\alpha$ positions with an RMSD of 2.1 Å. The topology and tertiary structure of these proteins are the same with only minor deviations (Figure 34A; Figure 34B). The most prominent difference is the presence of three extra strands at the C-terminus of Hcp1 where gpV$_N$ has a disordered region. Remarkably, within its crystal lattice, Hcp1 is assembled into hexameric rings that stack upon each other to form tubes (Figure 34C). The dimensions of the Hcp1 hexameric rings (90 Å with a 40 Å central hole) are strikingly similar to the dimensions of the hexameric rings of gpV in λ tails (90 Å with a 30 Å central hole (Katsura, 1981)).

The structural similarity between gpV$_N$ and Hcp1 supports a growing body of data pointing to an evolutionary relationship between components of the type VI secretion system and phage tails. A close structural similarity for VgrG, another key component of the type VI system, and the tail spike proteins, gp27 and gp5, from bacteriophage T4 (*Myoviridae*) has recently been reported (Leiman et al., 2009; Pukatzki et al., 2007). In *V. cholerae*, the VgrG secreted proteins have been shown to interact with one another and it has been speculated that they may form a trimeric complex analogous to the (gp27)$_3$(gp5)$_3$ membrane-puncturing tail spike complex from bacteriophage T4 (Pukatzki et al., 2007). Hcp1 is absolutely required for both the assembly of the type VI secretion apparatus and for the localization of ClpV1, a AAA+ family protein that provides the energy for Hcp1 secretion (Mougous et al., 2006). The proclivity for Hcp1 to form tail-like tubes in solution has been demonstrated by engineering a disulfide bond at the hexamer-hexamer interface observed in the crystal structure (Ballister et al., 2008). In addition, at least one homologue of Hcp1 is able to form tubes spontaneously (Leiman et al., 2009). Due to its tube-forming ability and its high level of expression, it is believed that Hcp1 forms the injection needle for secretion. Since gpV does form a long tube through which DNA and proteins are transported, the hypothesis that Hcp1 may assemble into a
comparable structure for the intercellular transport of proteins and/or oligonucleotide-type substrates is strongly supported by the structural similarity between these two proteins.

Figure 34. Structural Similarities between gpVN and Hcp1.

(A) Tertiary and (B) Secondary structure comparison between gpVN and Hcp1. Each secondary structure element in Hcp1 is coloured-coded based on its alignment with gpVN (β1, salmon; β2, orange; β3, yellow; β4, blue; α1, green; β5, sky blue; β6, purple; β7, deep pink). Loops and secondary structure differences are coloured in grey, with the exception of the β2-β3 loop that is coloured red. In panel B, disordered regions in gpVN are denoted with a red star. (C) Superposition of gpV (blue) onto monomer A of the Hcp1 hexamer (yellow). Crystallographic symmetry was applied to build the stacked Hcp1 hexamers. The β2-β3 loop of Hcp1 is represented in red spheres, which are clearly located in the hexamer-hexamer interface.

3.3.5 A Model for gpV Polymerization Based on Hcp1

In light of the structural and putative functional similarities between gpVN and Hcp1, I believe that the oligomerization mode of gpV within the phage tail is similar to that seen for
Hcp1 within the crystal lattice. It is important to note, however, that gpV\(_N\) is strictly monomeric in solution whereas Hcp1 is hexameric in solution (Mougous et al., 2006). gpV\(_N\) also possesses significant regions of disorder. We predict that some or all of these disordered regions would gain structure upon gpV polymerization and play a key role in facilitating this process. For example, the large disordered region from residues 50-78 in gpV\(_N\) corresponds to a long loop in Hcp1 that interacts with the neighbouring subunit in the hexameric ring and also sits on the upper surface of the ring (Figure 34C) where it can interact extensively with the next hexamer sitting on top. The unstructured loop in gpV\(_N\) likely becomes ordered upon gpV polymerization in a manner similar to Hcp1, a supposition which is supported by my mutagenesis data (Figure 32A; Figure 32B) suggesting that residues within this loop may play a key role in the polymerization process. The importance of this loop in gpV function is also underscored by the occurrence of several highly conserved positions in this region (residues Tyr55, Leu60, and Gly73) and the minimal degree of loop length variation observed here compared to other loops in the alignment (Figure 31). The C-terminal residues of gpV\(_N\) are disordered. Since this region forms a \(\beta\)-strand lying on the outside of the ring in Hcp1, a strand may also form here in gpV upon assembly (Figure 35A). This idea is bolstered by the computational prediction of a \(\beta\)-strand in this region of the gpV\(_N\) alignment (Figure 31) and by the occurrence of a strand at this position in the structure of XkdM (Figure 35A). Finally, the C-terminal domain of gpV has been seen as a protrusion on the exterior of the \(\lambda\) tail (Katsura, 1981). The formation of the proposed strand at the end of gpV\(_N\) would bring the terminus of this domain to the outside surface of the hexameric ring where the C-terminal domain would be located. The role of the C-terminal domain and the linker between gpV\(_N\) and gpV\(_C\) will be discussed further in chapter 4.

My model for the oligomerization of gpV based on the Hcp1 structure is supported by the gpV\(_N\) sequence alignment. Five highly conserved hydrophobic positions are detected in structured portions of gpV\(_N\) that were more than 20% exposed in the monomeric structure (shaded red in Figure 31). Remarkably, the superposition of the gpV\(_N\) monomer onto a subunit of Hcp1 reveals that each of these residues is positioned close to the hexamerization interface (Figure 35B). Another eight positions display relatively high conservation in that at least half of the residues at each position are identical, or they maintain charge or aromatic character. Five of these positions lie in the large central unstructured loop (shaded orange in Figure 31), and three
others are found at the putative hexamerization interface (shaded cyan in Figure 31) (Figure 35B). The gpV alignment has two regions exhibiting large insertions and/or deletions. Both of these regions comprise loops that would lie on the outside of the putative gpV ring in positions where Hcp1 also possesses large loops. In summary, the structural and bioinformatics data available support the hypothesis that gpV forms a hexameric ring in a manner similar to Hcp1.

**Figure 35. A Model for gpV Oligomerization.**

(A) Structural superpositions of gpV with Hcp1 and XkdM suggest that the disordered C-terminal residues of gpV could form a strand. For ease of visualization, in the superposition of Hcp1 and gpV, residues 32-56 and 138-162 from Hcp1 and 49-82 from gpV have been removed, and in the superposition of XkdM with gpV, residues 33-59 from XkdM and 49-82 from gpV have been removed. The folding of the gpV C-terminus (deep pink) into the Hcp1 β8” (deep pink) or the XkdM β7 position (deep pink) are likely conformational changes. (B) A side view of 3 monomers from the Hcp1 hexamer where monomer A has been replaced with gpV (blue). Highly conserved hydrophobic residues are coloured red, while relatively conserved residues are coloured in cyan. Conserved residues in the β2-β3 loop are orange. Residues contributing to the hexameric interface are circled in black (Asp44, Trp86, Gln93, Trp123, Phe112, Phe120, Val136, Ile137).

The presence of large unstructured regions within gpV that I hypothesize become structured upon polymerization provides a putative mechanism for the strict control of tail
assembly during λ morphogenesis. Since gpV has the potential to polymerize into aberrantly long non-functional structures (Bleviss and Easterbrook, 1971), premature association between gpV monomers must be prevented. The unstructured regions could serve this purpose such that gpV\textsubscript{N} remains monomeric even at the very high concentrations used in this NMR study (~1mM). Upon encountering the tail initiator complex, gpV commences polymerization in a process that is expected to include the ordering of unstructured regions. This ordering is likely facilitated by the initiator complex, and results in the formation of a surface that can participate in further polymerization until the end of the tape measure protein is reached. Interestingly, when the λ tail is dissociated at low pH, individual hexameric rings of gpV are obtained, indicative of a major structural change occurring in gpV upon oligomerization that prevents dissociation into monomers. The conformational changes in gpV may involve multiple steps of disordered-to-order transitions as the three disordered regions may not all fold in the same morphogenetic step.

### 3.4 Conclusions

The three dimensional structure of gpV\textsubscript{N} provides the first direct evidence that contractile and non-contractile phage tail tubes evolved from a common ancestor. The structural similarities between the tail tube proteins described here indicate that all long phage tails may be assembled from a common repository of components and a common assembly mechanism may exist. The structural similarity observed between gpV\textsubscript{N} and the hexameric type VI secretion protein, Hcp1, strongly supports the hypothesis that phage tails and type VI secretion systems are structurally, functionally, and evolutionarily related, and that Hcp1 forms the tube of this bacterial secretion apparatus, just as gpV\textsubscript{N} forms the tubular structure of the phage tail. As bacterial secretion systems are attractive drug targets, understanding the structural components and mechanisms of their assembly are imperative to the development of successful inhibitors. New possibilities for the application of knowledge from studies of tailed phages to the investigation of type VI secretion will dramatically increase the rate of knowledge acquisition in this emerging field. The inability of Hcp1 and many of its homologues to readily form tubes under normal solution conditions (Leiman et al., 2009) implies that, just as in the phage system,
other factors, which remain to be discovered, are necessary for efficient polymerization of this protein.

Finally, it is interesting to note that type VI secretion systems are not the first known example of bacteria evolving to use phages or parts of phages for their own advantage. Most strains of *P. aeruginosa* produce phage tail-like entities, known as pyocins, that are able to selectively kill other *P. aeruginosa* strains (Nakayama et al., 2000). The operons encoding pyocins are undoubtedly derived from phage genomes as significant sequence similarity is observed between pyocin proteins and various phage tail proteins. In addition, phage-derived particles, called Gene Transfer Agents, play an important role in mediating lateral gene transfer in many species of alpha-proteobacteria (Lang and Beatty, 2007). Together, these observations emphasize the importance of phage-derived elements in the evolution and function of a variety of complex bacterial systems.
CHAPTER 4

STRUCTURAL AND PRELIMINARY FUNCTIONAL CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF THE λ MAJOR TAIL PROTEIN

4.1 Overview

The role of the C-terminal domain of gpV (gpV_C) in λ infections remains unclear. While one study has shown that the removal of gpV_C does not affect the overall shape of the λ tail and under some conditions the overall infectivity, a C-terminal temperature sensitive defective mutant has been isolated (Katsura, 1981), suggesting that conditions may exist where gpV_C is absolutely required for phage infection.

Bioinformatics (Fraser et al., 2006), and proteolysis studies presented in Chapter 3 were used to delineate the domain boundaries of gpV_C, which was found to span residues 160 through to 246. In this chapter, I have shown that these residues form an independently folded β-sheet domain and that the absence of this domain results in a 100-fold decrease in activity compared to gpV_FL. By investigating the function of multiple C-terminal truncations of varying lengths, I have shown that the linker between the N- and C-terminal domains is also crucial for gpV function. This finding is consistent with the model of hexameric gpV_N presented in chapter 3.

In an attempt to elucidate the function of gpV_C, I determined its three-dimensional solution structure using NMR spectroscopy. The structure possesses an Ig-like fold, comprised of eight β-strands arranged into two β-sheets. I observed strong structural similarities between gpV_C and the cellular adhesion proteins invasin (PDB ID: ICWV-Domain 3), and intimin (PDB ID: 1F00-Domain 1). Based on the fact that gpV_C is surface exposed in the λ phage tail and that it exhibits structural similarities to several proteins involved in cell-cell adhesion, I speculate that it may weakly interact with carbohydrates on the outer membrane of E. coli in order to position the viral tail within close proximity to its high affinity target receptor, lamB.
Acknowledgements: Amanda Liu, a technician in the Davidson lab, cloned gpV residues 160-246 into a pET-32b vector. Dr. Voula Kanelis assisted me with some of the NMR data processing. Dr. Tara Sprules, the facility manager at the Québec/Eastern Canada High Field NMR Facility collected all of the NMR data used in the determination of the gpV<sub>C</sub> structure. Although none of their data is presented in this study, Dr. Genevieve MC Gasmi-Seabrook and Dr. Logan W. Donaldson worked on the solution structure of gpV<sub>C</sub> prior to the work presented here and therefore their time and efforts need to be acknowledged.

4.2 Materials and Methods

4.2.1 Plasmid Construction and Protein Purification

Codons corresponding to gpV residues 160-246 (gpV<sub>C</sub>) were sub-cloned into a modified pET32b expression vector (Novagen) as described in chapter 3, section 3.2.1, producing gpV<sub>C</sub> with a TEV-cleavable N-terminal 6-His affinity tag. The fusion proteins were overexpressed in BL21 CodonPlus cells grown at 37 °C in LB media or M9 minimal media enriched with 1g of <sup>15</sup>NH<sub>4</sub>Cl and 2g <sup>13</sup>C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> for NMR structural studies. The protein was expressed and purified using the same protocol described for gpV<sub>N</sub> in chapter 3, section 3.2.1. The cleaved gpV<sub>C</sub> protein collected in the flow through off the Ni-NTA column was buffer exchanged into 50mM Na<sub>2</sub>HPO<sub>4</sub> pH 6.8, and 200mM NaCl. The protein was concentrated using an amicon centrifugation filter device (Millipore) with a nominal molecular weight cut off of 5000 Da to the desired concentration for structural or biophysical characterization.

For in vivo complementation assays, gpV<sub>FL</sub> and its truncations were cloned into a pTrc99c vector (Amann et al., 1988) (see chapter 3, section 3.2.1 for further details). gpV truncations were made in the pTrc99c vector by mutating a stop codon into the desired amino acid position. In this manner, gpV constructs were made to span the following residues: 1-153, 1-154, 1-156, 1-157, 1-159, and 1-160. Quikchange mutagenesis (Stratagene) was used to make all mutations.
4.2.2 Biophysical Characterization

Circular dichroism (CD) experiments were performed in an Aviv 62A DS circular dichroism spectrometer. Far-UV scans were collected at wavelengths from 200 to 260 nm in 1 nm increments at 25 °C. A one second averaging time was used and three independent scans were collected and averaged for analysis. Samples with protein concentrations of 20µM were used in a 0.1-cm cuvette.

4.2.3 Solution Structure Determination

NMR studies on gpV were carried out at protein concentrations of 2.5mM in 50mM Na₂HPO₄ pH 6.8, 200mM NaCl and 0.71mM in 50mM Na₂HPO₄, pD 7.2, and 200mM NaCl made using 99.9% D₂O. All NMR spectra were collected at 25 °C on a Varian INOVA 500 MHz spectrometer at the Québec/Eastern Canada High Field NMR Facility. All NMR data were collected, processed, and analyzed as described in section 3.2.3. 1349 NOEs were manually assigned and calibrated using the CALIBA module of CYANA 2.1 (Guntert, 2004). Dihedral angle restraints were assigned using TALOS (Cornilescu et al., 1999). Initial structure calculations were performed using CYANA 2.1 for the automatic assignment of all NOE peaks and structure refinement. In the final iteration, 100 structures were refined and the 20 lowest energy structures were kept for structural analysis. PROCHECK-NMR (Laskowski et al., 1996) was used for further validation. The RMSD between the 20 lowest energy structures was calculated using MOLMOL (Koradi et al., 1996).

4.2.4 In vivo Functional Characterization

The in vivo complementation assays on the gpV truncations were carried out as described in Chapter 3, section 3.2.4.

4.3 Results
4.3.1 Biophysical Characterization

To confirm that the protease-resistant region of gpV (Chapter 3, section 3.3.1) that displays sequence similarity to other Ig-like domains (Fraser et al., 2006) was in fact an authentic domain, we sub-cloned the DNA region encoding residues 160 to 246 into the pET32b expression vector with a cleavable 6-His tag at its N-terminus. The protein domain was highly expressed in E. coli and was purified to homogeneity (Figure 36).

Figure 36. The Purification of gpV C.

The purification of gpVC monitored on a 16% SDS polyacrylamide gel. The samples were loaded as follows: 1, Ladder; 2, Supernatant; 3, Pellet; 4, Flow through off Ni-NTA; 5, Wash number 1; 6, Wash number 2; 7, Wash number 3; 8, Pooled elution fractions off Ni-NTA column; 9, gpVC cut with TEV; 10, gpVC separated from tag; 11, Pure gpVC; 12, Pure concentrated gpVC.

To determine if gpVC possesses stably folded structural elements, circular dichroism (CD) spectroscopy was used to provide a measure of the type of secondary structure present. The far-UV CD spectrum of gpVC is typical of a β-strand protein, with a single minimum observed at 217 nm (Figure 37A). In order to further characterize the biophysical properties of gpVC, an 1H-15N HSQC spectrum was acquired. Based on the amino acid composition of gpVC, 98 peaks were expected in the HSQC. The HSQC of gpVC is excellent, with 96 well-dispersed peaks of approximately equal intensity (Figure 37B). The two unobservable amide peaks are attributed to the engineered N-terminal residues in the TEV recognition site that remain after tag
cleavage. Together, these biophysical data suggest that gpV_C is indeed an independently folded domain with β-sheet secondary structure.

Figure 37. Spectroscopic Characterization of gpV_C.
(A) Circular Dichroism (CD) far-UV wavelength spectra of gpV_C at 25 °C. The spectrum was obtained on 20µM protein in a 0.1cm cuvette. (B) ^1H-^15N HSQC of gpV_C. The HSQC spectrum was acquired at 25 °C.

4.3.2 The C-terminal Domain and the Linker between the N- and C-terminal Domain is Functionally Important

To determine whether the absence of gpV_C leads to a functional phenotype, the function of several N-terminal domain constructs of varying amino acid lengths were tested. The solution structure and dynamic studies on the first 153 residues of gpV (Chapter 3, section 3.3.2) revealed that the N-terminal domain is structured up to residue 148. In addition, my proteolysis studies revealed a trypsin cut site between the N- and C-terminal domains after residue 155 (Chapter 3, section 3.3.1). Together with previous bioinformatic analysis (Fraser et al., 2006), these data suggest that a flexible linker spans residues 148-159 to connect the N- and C-terminal domains. For these reasons, the constructs tested for biological activity encompassed residues 1-153, 1-154, 1-156, 1-157, 1-159, and 1-160 (referred to from this point forward as V153, V154, etc.) (Figure 38A). While V153 and V154 were each found to be functionally inactive, V156
and V157 showed a subtle increase in activity, and V159, and V160 partially restored gpV function but were still 100-fold less active when compared to gpV_{FL} (Figure 38B). Intriguingly, V154 exhibited a dominant negative phenotype such that it inhibited WT λ by nearly 1 million-fold (Figure 38C). V156 and V157 did not inhibit WT λ but visibly altered the appearance of the clearings produced by WT λ on the bacterial lawn. The clearings were ‘lighter’ and not as well defined as those observed when λ was plated on cells over-expressing other forms of gpV.
Figure 38. The C-terminal Domain and the Flexible Linker Region Between gpV
and gpVC are Functionally Important.

(A) Schematic representation of the boundaries of gpV used in the complementation assay. The residues that make up the linker region (residues 149-160) are indicated by their single letter code. (B) The activity of gpV constructs ending at residues 153, 154, 156, 157, 159, and 160 (V153, V154, V156, V157, V159, and V160 respectively) were tested by observing the ability of each protein, expressed from a plasmid, to complement a Vam phage lysate. (C) The effect of each of the truncations on a WT λ infection was also assayed.
Together, these data suggest that both the flexible region connecting the N- and C-terminal domains and the C-terminal domain itself hold a functional role. It should be noted, that with the exception of V153, the expression levels of each of the truncations have not yet been tested and compared to the expression of gpV_{FL}. In addition, the folding and stability of each truncation has also not yet been assessed. These issues will need to be addressed before definitive conclusions can be made. For example, the dominant negative behavior noted for V154 may be due to two neighboring negatively charged amino acids at the end of the construct (Glu and Asp). Electrostatic repulsions between these two residues may destabilize the domain, which could in turn be detrimental to the surrounding gpV_{FL}. Studies designed to address these concerns are described in chapter 6. Since some activity is observed for V159, and V160, this suggests that λ tails are still being assembled. The low activity may indicate that the efficiency of tail assembly is altered in the absence of gpV_{C} or that gpV_{C} may hold some other unidentified accessory role.

4.3.3 Solution Structure Determination

To help ascertain the functional role of gpV_{C}, I determined its three-dimensional structure by NMR spectroscopy. A combination of two- and three-dimensional NMR data were collected and used to identify and sequentially assign greater than 95% of the ¹H, ¹³C, and ¹⁵N resonances of the backbone and side chain atoms for each of the 87 residues in gpV_{C}. The ensemble of the 20 lowest energy structures calculated for gpV_{C} is presented in Figure 39A, and the statistical parameters of the structure determination are summarized in Table 6. Analysis of the 20 lowest energy structures reveals that there are no distance violations greater than 0.5 Å. The secondary structure elements were defined using a combination of strong dαN (i, i+1) NOEs, and the secondary structure defining program, DSSP (Kabsch and Sander, 1983).
Table 6. Structural Statistics for the Ensemble of 20 Structures for gpV\textsubscript{C}-WT

<table>
<thead>
<tr>
<th>Distance restraints\textsuperscript{a}</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Range,</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium Range, 1&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long Range,</td>
<td>i-j</td>
</tr>
<tr>
<td>Hydrogen bond pairs (HN-O, N-O)</td>
<td>16</td>
</tr>
</tbody>
</table>

Dihedral angle restraints
φ/ψ pairs (104 angles) | 52 |

Pairwise RMSD

<table>
<thead>
<tr>
<th>Secondary structure\textsuperscript{b}</th>
<th>RMSD (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone atoms</td>
<td>0.21 ± 0.05 Å</td>
</tr>
<tr>
<td>All heavy atoms</td>
<td>0.62 ± 0.11 Å</td>
</tr>
</tbody>
</table>

Assigned atoms\textsuperscript{c}

| Backbone atoms                          | 0.40 ± 0.09 Å |
| All heavy atoms                         | 0.82 ± 0.11 Å |

Ramachandran Statistics\textsuperscript{d}

| Most favored regions                    | 69.6% |
| Additional allowed regions              | 29.8% |
| Generously allowed regions              | 0.6%  |
| Disallowed regions                      | 0.0%  |

\textsuperscript{a}There were no distance violations > 0.5 Å. \textsuperscript{b}RMSD for residues 5-8, 12-17, 36-38, 47-51, 54-58, 63-66, 69-71, 76-78, 81-85. \textsuperscript{c}RMSD for residues 1-87. \textsuperscript{d}Determined using PROCHECK-NMR (Laskowski et al., 1996).

gpV\textsubscript{C} consists of a single domain formed from eight β-strands, with residues contributed by amino acids 5-8 (β1 - A), 12-15 (β2 - B), 19-26 (β3 - C), 36-38 (β4 - D), 47-51 (β5 - E), 54-58 (β6 - F), 63-66 (β7 - G), 69-71 (β7’ - G’), 76-78 (β8’ - H’), 81-85 (β8 - H). β7 and β8 break into β7’, and β8’, respectively, due to the presence of proline 68 (proline 227 in the context of gpV\textsubscript{FL}). Proline 68 introduces a kink in β7, which disrupts the hydrogen-bonding pattern between β7 and β8 (Figure 39BC). Interestingly, bioinformatics analysis revealed that gpV\textsubscript{C} is the only Ig-like domain found to possess a proline at position 68 (James Fraser, personal communication). In addition, proline 9 breaks β1 into β2. The β-strands form a classical Ig-like fold with two β-sheets, each comprised of 4 strands (Figure 39BC). The first β-sheet is formed
by $\beta_1$(A), $\beta_3$(C), $\beta_6$(F), and $\beta_5$(E) while, $\beta_2$(B), $\beta_8$(H), $\beta_7$(G), and $\beta_4$(D) form the second $\beta$-sheet (Figure 39C).

Figure 39. The Solution Structure of gpV$_C$.

(A) Best fit superposition of the backbone atoms (N, C$\alpha$, C(O)) of the well structured regions (residues 5-8, 12-15, 19-26, 36-38, 47-51, 54-58, 63-66, 69-71, 76-78, 81-85) from the twenty lowest energy gpV$_C$ structures. (B) Cartoon representation of residues 1-87 from the lowest energy structure. (C) Two-dimensional topology diagram of gpV$_C$. In each case, strands are coloured blue and loops are coloured grey. Structure figures were generated using PyMOL (http://pymol.sourceforge.net/).

It should be noted that the edge strand $\beta_4$ (strand D), formed from residues Phe 36, Arg 37, and Ala 38 was particularly difficult to assign and there is some evidence suggesting that this region of the structure may be dynamic. For example, the backbone amide peaks for neighboring residues Val 39 and Ser 40 in the HSQC are broad and weaker in intensity compared to other amide peaks in the spectrum (Figure 40). In addition, I was unable to assign chemical shifts to the side chain atoms of Val 39. Despite this, NOEs were observed that linked $\beta_4$ to $\beta_7$. 
Figure 40. Broadening in the HSQC of gpV<sub>C</sub>-WT

A small region of the $^1$H-$^15$N HSQC for gpV<sub>C</sub> containing is enlarged. The backbone amide resonance for residues 39 and 40 are labeled.

Intriguingly, searches with the DALI algorithm (Holm and Sander, 1995) reveal that the domains from proteins often involved in cellular adhesion, such as invasin (PDB ICWV-Domain 3) and intimin (PDB 1F00-Domain 1), are most structurally similar to gpV<sub>C</sub>. Invasin and intimin both share limited sequence identity to gpV<sub>C</sub>, 14% and 15%, respectively, however the best 3D superposition of the two domains yields an RMSD between C<sub>α</sub> atoms of matched residues of 2.70 Å and 3.00 Å, respectively (Figure 38AB). The invasin domain is a member of
the PFAM invasin domain 3 family, whereas the intimin domain is a member of the PFAM bacterial Ig-like domain family, Big 1 (Sammut et al., 2008). Previous bioinformatics studies showed that by sequence analysis gpVC is most closely related to the bacterial Ig-like domain, Big 2 family (Fraser et al., 2006).

**Figure 41. Superposition of a Representative Set of Proteins Displaying Structural Similarities to gpVC.**

Proteins were superimposed onto gpVC using COOT (Emsley and Cowtan, 2004). In each superposition, gpVC is coloured blue, (A) Intimin is coloured green, (B) invasin is red, (C) and MAAdCAM-1 is orange.

### 4.4 Discussion

#### 4.4.1 Putative Role of the Linker Between the N- and C-terminal Domains

Residues 149 through 160 form a flexible linker between the N- and C-terminal domains of gpV. In this chapter, a portion of this loop up to at least residue 154, was shown to be essential for gpV function. This result was not entirely surprising as it is consistent with several lines of evidence presented in chapter 3, suggesting that the linker region between the two
domains undergoes a disorder-to-order transition during gpV oligomerization. For example, in a secondary structure prediction carried out across a diverse alignment of gpV homologues, residues 148 to 152 were predicted to form a β-strand (Figure 31). A strand was also predicted to form in this linker region based on structural comparisons between gpV_N with both Hcp1 and XkdM. This would ultimately position gpV_C on the exterior surface of the tail tube (Figure 42), and would place the linker region at the interface between two gpV monomers within a hexameric model (Figure 42). Experiments designed to precisely determine the role of this linker region are described in chapter 6.

![gpV_N C-terminus (linker region)](image)

**Figure 42. The Domain Linker Folds into the Oligomeric Interface.**

Three monomers from an Hcp1 hexamer are shown with the middle Hcp1 subunit (yellow) overlayed with gpV_N (blue). The region lacking regular secondary structure at the C-terminus of gpV_N is coloured red. An arrow indicates the predicted direction of folding upon oligomerization, where it is expected to fold into a similar position as an Hcp1 strand formed from residues 131 to 137 (coloured red). In this way, the putative C-terminal strand in gpV_N would be positioned near the oligomerization interface and gpV_C would be located on the outside of the hexameric ring.

### 4.4.2 Structural Features of the gpV C-terminal Domain

The NMR solution structure of gpV_C provides conclusive biochemical evidence that the major tail protein from phage λ possesses an independently folded Ig-like domain at its C terminus, corroborating BLAST and sequence alignment data previously reported (Fraser et al., 2006). Ig-like domains are classified on the basis of number and location of strands, as well as sequence similarities (Bork et al., 1994). Typically, classical Ig-like domains have seven to nine anti-parallel β-strands arranged into two distinct sheets forming a β-sandwich (Bork et al.,
1994), with a common structural core of six strands (Halaby et al., 1999). Careful examination of the 2D-topology and three-dimensional structure of gpV$_C$, reveals a topology most similar to the I-set domains (type I1), an intermediate between the V and C1-type subtypes (Bork et al., 1994; Halaby and Mornon, 1998; Harpaz and Chothia, 1994). Similar to other I-set domains, strand A (β1) in the gpV$_C$, begins in sheet 1 however, a proline at position 9 induces the formation of a kink, causing the strand to continue in the same direction and finish in sheet 2 where it hydrogen bonds in a parallel fashion with strand H (β8). Many domains involved in cellular adhesion and cell surface receptors belong to the I-set structural family (Harpaz and Chothia, 1994). Interestingly, based on sequence alone gpV$_C$ is most similar to the Big2 family of Ig-like domains, and yet structurally it is most similar to the I-set subgroup. Whether this topology is conserved among any other Ig-like domains from dsDNA tailed bacteriophages remains unclear.

### 4.4.3 Putative Function of gpV$_C$

The Ig-like fold is one of the most abundant and widely dispersed structural arrangements seen in nature (Halaby and Mornon, 1998) and has been reported in an array of protein structures with immense functional diversity, ranging from cell adhesion molecules (Luo et al., 2000), to intracellular muscle proteins (Fong et al., 1996; Improta et al., 1996). Recently, an abundance of Ig-like domains among unrelated proteins in many dsDNA phages has been identified, and using a bioinformatic approach 68 Ig-like domains in 54 proteins in 41 different dsDNA tailed phages have been reported (Fraser et al., 2006). The abundance of Ig-like domains found in bacteriophage proteins may be indicative that they provide a selective advantage.

Although there is not yet any experimental data identifying the function of gpV$_C$, several lines of evidence suggest that the protein aids in the attachment of λ phage to its E. coli host cell. It is well documented that λ phage adsorbs to the surface of E. coli through a high affinity interaction between the C terminus of the tail fiber protein, gpJ, and the E. coli outer member protein, LamB (Wang et al., 2000; Werts et al., 1994). This high affinity interaction facilitates the first critical steps in the bacteriophage lambda life cycle: adsorption and DNA injection into
the host. The studies presented here suggest that weak interactions occur between the gpV$_C$ and carbohydrates, such as peptidoglycans and/or lipopolysaccharides (LPS), on the surface of the outer membrane of E. coli. These weak interactions may assist in positioning λ phage within close vicinity of the host cell and lamB receptor. In order for this hypothesis to hold true, the gpV$_C$ must be located on the surface of phage λ. In fact, gpV$_C$ has been shown to exist as an outer protrusion along the tail tube (Katsura, 1981; Roessner and Ihler, 1984), where 192 copies of the Ig-like domain span the surface of the tubular region of the tail. gpV$_C$ may interact with cell surface polysaccharides in a manner similar to which Selectin-mediated rolling of leukocytes occurs along the vessel wall at a site of stimuli (Kansas, 1996; Wild et al., 2001). A weak protein-carbohydrate interaction would be consistent with the abundance of gpV$_C$ copies displayed on the surface of the λ tail. One of the key players in selectin-mediated rolling is mucosal addressin cell adhesion molecule 1 (MAdCAM-1). In humans, MAdCAM-1 contains two Ig-like domains, a mucin-like region, a transmembrane domain, and a cytoplasmic domain (Shyjan et al., 1996). The Ig-like domains are responsible for binding to the integrin α4β7, which mediates slower transient rolling along the endothelium. Interestingly, gpV$_C$ is structurally similar to the first Ig-like domain from MAdCAM-1 (PDB ID 1bqs-A) with an RMSD of 2.7 Å between C$^{\alpha}$ atoms of matched residues despite sharing only 8% sequence identity (Tan et al., 1998) (Figure 41C). In addition, intimin and the integrin binding protein, invasin (Hamburger et al., 1999; Jones et al., 1995; Luo et al., 2000) are also structurally similar to gpV$_C$ with RMSDs of 3.0, and 2.7 Å, respectively. Although the specific mechanism by which these proteins exert their functions is different, a striking commonality exists in their general role as intercellular adhesion mediators. Intimin is a bacterial protein that mediates the attachment between a mammalian cell and a pathogen by binding to the secreted bacterial protein Tir that gets inserted into the host membrane (Liu et al., 1999). Invasin is highly related to intimin, however invasins bind to integrin proteins, ultimately leading to the uptake of bacteria into the host cell (Isberg and Tran Van Nhieu, 1994; Isberg et al., 1987). It is well known that structural similarities are often indicative of evolutionary relationships, and it is therefore interesting that the roles of many of the structurally most similar proteins to gpV$_C$ lie in cellular adhesion.
Ig-like domains have been implicated in interacting with cell surface carbohydrates (Buts et al., 2003; Zaccai et al., 2003) and in addition, non-Ig-like domains from other bacteriophages have been shown to bind to LPS on the surfaces of its hosts during infection. Since a large number of functionally diverse proteins contain Ig-like domains and Ig-like domains can interact with other molecules via any accessible surface of the protein (Bork et al., 1994), it is extremely difficult to definitively predict both the function of gpV
\textsubscript{C} and the residues which may be crucial to polysaccharide binding. Future studies described in chapter 6, will be aimed at assessing the ability of gpV
\textsubscript{C} to adsorb to \textit{E. coli} cells and its potential to bind polysaccharides. Alternatively, it is also possible that gpV
\textsubscript{C} plays a role in stabilizing the tail tube once it is assembled. In support of this hypothesis, studies have indicated that while fully formed wild-type tails can be dissociated at pH 2 to yield individual hexameric rings, investigators have been unable to isolate ring-like structures from tails formed from only the N-terminal domain (Katsura, 1981). Chapter 6 outlines a series of experiments designed to address these questions.
A PROTEIN FOLDING DEFECT IN THE C-TERMINAL DOMAIN OF THE λ MAJOR TAIL PROTEIN ABROGATES FUNCTION OF THE FULL LENGTH PROTEIN

5.1 Overview

Since the complete removal of gpV_C decreases the function of gpV by only approximately 100-fold (Chapter 4), it is intriguing that a point mutation, G222D, within gpV_C completely abrogates function of gpV_FL (Katsura, 1981). Furthermore, when G222D is combined with P227L to produce the double mutation G222D/P227L, gpV function is recovered (Katsura, 1981). The structure of gpV_C (Chapter 4) reveals that G222 is located at the beginning of β7 and P227 is found at the end of β7 just before β7’ (Figure 43).

Figure 43. Location of G222 and P227 within gpV_C.
The cartoon representation of the lowest energy solution structure of gpV_C-WT. Highlighted in red and represented as sticks are G222 and P227.
In an attempt to understand the mechanisms by which gpV function is abolished and restored by G222D and P227L respectively, this chapter presents the preliminary structural, biophysical, and functional characterization of the following mutations within both gpVFL and the isolated C-terminal domain: G222D, G222D/P2227, and P227L. These studies have revealed that the G222D mutation is not only functionally inactive, but that it exerts a temperature dependent dominant negative phenotype. Also, in addition to inhibiting WT λ, G222D exerts a dominant negative phenotype on phage phi80 infections and it also appears to negatively affect infections by phage HK97. The preliminary NMR data suggests that G222D destabilizes the C-terminal domain causing it to partially unfold. Furthermore, the destabilized form of the domain interacts with gpVN in a region that is localized to its ‘top’ and oligomerization surface.

In contrast to gpVC-G222D, gpVC-P227L is folded and I have determined its solution structure using NMR spectroscopy. To provide insight into how the P227L mutation can make the C-terminal domain resistant to the detrimental effects of the G222D mutation, the structure of gpVC-P227L was compared to gpVC-WT. The structural comparison reveals that the β7-β8 region of gpVC-P227L is altered compared to gpVC-WT and suggests that the P227L mutation may act to stabilize the domain and protect it from G222D induced unfolding.

**Acknowledgements:** Dr. Elliott Stollar collected the HSQC data on gpVC-WT, gpVC-G222D, gpVC-P227L, and gpVC-G222D/P227L. Dr. Voula Kanelis collected the HSQCs on samples of gpVN mixed with gpVC. Erik MacKinnon, a rotation student whom I supervised, tested the dominant negative effect of G222D on other phages and at 30 °C.

### 5.2 Materials and Methods

#### 5.2.1 Plasmid Construction and Protein Purification

gpVC in pET32b (*Novagen*) (described in chapter 4) and gpVFL in pTrc99c (described in chapters 2, 3 and 4) were used as templates to introduce the desired mutations into gpV. The defective mutant (G222D) and its pseudo-revertant (G222D/P227L) were isolated by I. Katsura.
and later sequenced in Roger Hendrix’s lab by his graduate student Jun Xu. Bob Duda (Hendrix lab) communicated the sequencing results to Alan Davidson. Quikchange mutagenesis (Stratagene) was used to make all mutations. \(\text{gpV}_C\) in pET32b was used for the expression and purification of all \(\text{gpV}_C\) mutations. \(\text{gpV}_{\text{FL}}\) in pTrc99c was used for \textit{in vivo} complementation assays.

### 5.2.2 \textit{In vivo} Functional Studies

\textit{In vivo} functional studies were performed as described in Chapters 3 and 4 with the following exceptions. First, these experiments were carried out in the absence of IPTG. Second, the dominant negative phenotype from G222D was probed at both 30 °C and 37 °C on \(\lambda\) infections and on two other \textit{E. coli} phages: HK97, and phi80. The activity of each of the mutants in this chapter was tested in the context of \(\text{gpV}_{\text{FL}}\).

### 5.2.3 NMR Spectroscopy

#### 5.2.3.1 HSQCs on \(\text{gpV}_C\) Mutants

HSQCs were acquired on \(\text{gpV}_C\)-WT, \(\text{gpV}_C\)-G222D, \(\text{gpV}_C\)-P227L, and \(\text{gpV}_C\)-G222D/P227L at protein concentrations of 1.56mM, 42uM, 1.47mM, and 0.137mM respectively in 50mM \(\text{Na}_2\text{HPO}_4\), pH 6.8, 200mM NaCl. All NMR spectra were collected at 25 °C on a 500 MHz magnet at the University of Toronto NMR centre. NMR data were processed and analyzed as described in Chapter 3, section 3.2.3.

#### 5.2.3.2 Structure Determination of \(\text{gpV}_C\)-P227L

NMR structural studies on \(\text{gpV}_C\)-P227L were carried out at protein concentrations of 1.9mM in 50mM \(\text{Na}_2\text{HPO}_4\), pH 6.8, 200mM NaCl and 1.9mM in 50mM \(\text{Na}_2\text{HPO}_4\), pD 7.2, and 200mM NaCl made using 99.9% \(\text{D}_2\text{O}\). All NMR spectra were collected at 25 °C on a Varian INOVA 500 MHz spectrometer at the Québec/Eastern Canada High Field NMR Facility. NMR data were processed and analyzed as described in Chapter 3, section 3.2.3. All resonance
assignments were attained as described in Chapter 4. 1047 NOEs were manually assigned and calibrated using the CALIBA module of CYANA 2.1 (Guntert, 2004). Dihedral angle restraints were assigned using TALOS (Cornilescu et al., 1999). Initial structure calculations were performed using CYANA 2.1 (Guntert, 2004) for the automatic assignment of all NOE peaks and structure refinement. In the final iteration, 100 structures were refined and the 20 lowest energy structures were kept for structural analysis. PROCHECK-NMR (Laskowski et al., 1996) was used for further validation. The RMSD between the 20 lowest energy structures was calculated using MOLMOL (Koradi et al., 1996).

5.2.3.3 Interaction Studies between gpV_N and gpV_C

An $^{15}$N$^{13}$C-labelled gpV_N sample that was purified for previous structural studies (Chapter 3) was mixed in a 1:1 ratio with either unlabelled gpV_C-G222D or unlabelled gpV_C-WT to a final volume of 500µl. The starting concentration of each protein in the gpV_C-G222D mixing experiment was 0.28mM and the starting concentration of each protein in the gpV_C-WT mixing experiment was 1.3mM. After mixing the two proteins, an HSQC spectrum was immediately acquired at 20 °C on a 600 MHz spectrometer at the University of Toronto Mississauga Campus NMR center. Prior to mixing, an HSQC of $^{15}$N$^{13}$C-labelled gpV_N was acquired as a reference. The resulting HSQC spectra were overlaid and compared for chemical shift changes.

5.3 Results

5.3.1 gpV_FL-G222D is Inactive and Dominant Negative

To confirm previous findings that the single point mutant G222D abrogates and the double mutation G222D/P227L recovers gpV_FL function, respectively (Katsura, 1981), I introduced each of these mutations into gpV_FL and assessed their biological activity using an *in vivo* complementation assay. As can be seen in Figure 44A, the G222D mutant within gpV_C, is completely void of biological activity, while the G222D/P227L and P227L mutants display activity at levels comparable to wild-type gpV (Figure 44). My results also indicated that in
addition to being inactive, the G222D mutant is also dominant negative as it inhibits the growth of wild-type λ at 37 °C by approximately one-million fold (Figure 44B).

Figure 44. The G222D Mutation within gpV<sub>C</sub> is Inactive and Dominant Negative.

(A) The ability of gpV mutants expressed from a plasmid to complement a λ<sub>Vam</sub> lysate in vivo was scored by observing plaque formation. The “Empty Vector” experiment constitutes the negative control and “Wild-type” refers to gpV<sub>FL-WT</sub> and constitutes the positive control. The G222D and P227L single point substitutions and a G222D/P227L double mutant were also tested. As indicated, three different dilutions of the Vam lysate were spotted onto the plates. The G222D mutant was unable to complement a Vam phage lysate. (B) The effect of plasmid-expressed gpV<sub>FL-WT</sub> and each of the mutants on the growth of WT λ phage was assayed. A WT phage lysate was spotted at the indicated dilutions on cells carrying these plasmids. Only the G222D mutant inhibited WT λ.

Since previous studies indicated that G222D was a temperature sensitive mutation as it was active at 30 °C but not at 37 °C (Katsura, 1981), I wanted to determine whether the dominant negative phenotype was also exerted in a temperature dependent manner. To accomplish this, Erik MacKinnon, a rotation student in the Davidson laboratory, repeated the dominant negative assay at 30 °C in parallel to a ‘control’ reaction at 37 °C (Figure 45). As can be seen in Figure 45A, the G222D mutation does not exhibit this detrimental gain of function phenotype at 30 °C.
Figure 45. gpV\textsubscript{FL}-G222D Exerts a Temperature Dependent Dominant Negative Phenotype on both λ and Phi80 Infections.

The effect of plasmid-expressed gpV Wild-type and G222D on λ, Phi80, and HK97 infections was assayed at (A) 30 °C and (B) 37 °C. The lysates were spotted at the indicated dilutions on cells carrying these plasmids. The “Empty Vector” experiment constitutes the negative control and “Wild-type” refers to gpV\textsubscript{FL}-WT and constitutes the positive control. The G222D mutant inhibited λ and phi80 infections by approximately 10,000 and 100-fold, respectively, at 37 °C but not at 30 °C. Furthermore, G222D also altered the appearance of the bacterial clearings made by HK97 at 37 °C.

Furthermore, to assess whether this phenotype was unique to λ infections, Erik also plated two other \textit{E. coli} phages, phi80 and HK97, onto cells bearing plasmids that express gpV\textsubscript{FL}-G222D (Figure 45). gpV\textsubscript{C}-G222D inhibited phi80 infections by approximately 100-fold at 37 °C (Figure 45B) but no inhibition was observed at 30 °C (Figure 45A). It is not yet known whether λ gpV and the phi80 MTP are functionally interchangeable however the N-terminal domains of each protein share 74% sequence identity. While gpV\textsubscript{FL}-G222D did not inhibit
HK97 infections, it visibly altered the appearance of the bacterial clearings at 37 °C but not at 30 °C (Figure 45). In the presence of the gpV\textsubscript{FL}-G222D, the HK97 clearings were very cloudy and ‘non-distinct’ compared to those produced on cells bearing an empty vector or a vector expressing gpV\textsubscript{FL}-WT (Figure 45B)

### 5.3.2 gpV\textsubscript{C}-G222D Interacts with gpV\textsubscript{N}

One possible mechanism by which gpV\textsubscript{FL}-G222D exerts its dominant negative effect is by interacting with gpV\textsubscript{N} in a manner that prevents it from assembling into tails. To test this hypothesis, $^{15}$N$^{13}$C-gpV\textsubscript{N} was mixed with either unlabelled gpV\textsubscript{C}-WT or unlabelled gpV\textsubscript{C}-G222D in a 1:1 ratio and an HSQC spectrum was immediately acquired. It should be noted that each sample was purified under the same buffer conditions (50mM Na\textsubscript{2}HPO\textsubscript{4}, pH 6.8, 200mM NaCl). The superposition of the HSQC spectrum of gpV\textsubscript{N} mixed with gpV\textsubscript{C}-WT onto the spectrum acquired for gpV\textsubscript{N} is nearly perfect. There are however 6 very subtle chemical shift changes (Figure 46). The subtle nature of the chemical shift changes makes it difficult to determine whether a weak and/or transient interaction exists between gpV\textsubscript{N} and gpV\textsubscript{C}-WT or rather, that minute differences exist between the pH of the gpV\textsubscript{N} solution and the pH of the final mixture. In contrast, after the addition of unlabelled gpV\textsubscript{C}-G222D to $^{15}$N$^{13}$C-gpV\textsubscript{N}, marked changes were observed for at least 16 peaks within the gpV\textsubscript{N} spectra, suggesting that gpV\textsubscript{C}-G222D interacts with gpV\textsubscript{N} (Figure 47).
Figure 46. The HSQC of $^{15}N^{13}C$-gpV$_N$ Mixed with unlabelled gpV$_C$-WT.

The spectrum of gpV$_N$ was acquired prior to mixing with gpV$_C$-WT and is shown in black. After combining gpV$_N$ with gpV$_C$-WT in a 1:1 ratio, an HSQC spectrum was immediately acquired (green open circles). Six subtle changes were noted after the addition of gpV$_C$-WT and are highlighted on the spectral overlay with a blue arrow. The chemical shifts in the gpV$_N$ spectra (black) that did not perfectly overlay with a peak in the gpV$_N$ plus gpV$_C$-WT spectrum (green) are labeled with the residue to which they belong (based on assignments obtained in chapter 3). Unless noted otherwise, all chemical shifts arose from backbone amides. A question mark indicates the appearance of a new peak in the gpV$_N$ plus gpV$_C$-WT spectrum (green).
Figure 47. gpV₅-G222D Interacts with gpV₆.

The spectrum of gpV₆ was acquired prior to mixing with gpV₅-G222D and is shown in black. After mixing gpV₆ with gpV₅-G222D in a 1:1 ratio, an HSQC spectrum was immediately acquired (red open circles). Sixteen changes were noted after the addition of gpV₅-G222D and are highlighted on the spectral overlay with a blue arrow. The chemical shifts in the gpV₆ spectra (black) that did not perfectly overlay with a peak in the gpV₆ plus gpV₅-G222D spectrum (red) are labeled with the residue to which they belong (based on assignments obtained in chapter 3). Unless noted otherwise, all chemical shifts arose from backbone amides. A question mark indicates the appearance of a new peak in the gpV₆ plus gpV₅-G222D spectrum (red).
Based on the chemical shift resonance assignments obtained in Chapter 3, I was able to map the observed chemical shift changes onto the structure of gpV$_N$. Table 7 summarizes each of the residues whose amide groups underwent a chemical shift change upon mixing with gpV$_C$-G222D. Interestingly, each of these residues was found localized to one region of gpV$_N$ (Figure 48). Furthermore, in an oligomeric model of gpV$_N$ built by superimposing one gpV$_N$ monomer onto each chain in the hexameric crystal structure of Hcp1 (described in chapter 3), these residues were found to be primarily localized onto one surface which I will arbitrarily call the ‘top’ surface of the gpV hexamer as well as the oligomerization interface (Figure 49).

Table 7. gpV$_N$ Residues Implicated in Interactions with gpV$_C$-G222D

<table>
<thead>
<tr>
<th>Res. #</th>
<th>Res. Name/Atom</th>
<th>Location in Hexamer</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>Phe</td>
<td>Hydrophobic Core</td>
</tr>
<tr>
<td>99*</td>
<td>Trp/ NH$_{ε1}$</td>
<td>Hydrophobic Core</td>
</tr>
<tr>
<td>100</td>
<td>Phe</td>
<td>Top/Interface</td>
</tr>
<tr>
<td>101</td>
<td>Asn</td>
<td>Top/Interface</td>
</tr>
<tr>
<td>103</td>
<td>Gly</td>
<td>Top</td>
</tr>
<tr>
<td>105</td>
<td>Thr</td>
<td>Top</td>
</tr>
<tr>
<td>123*</td>
<td>Trp</td>
<td>Top</td>
</tr>
<tr>
<td>123</td>
<td>Trp/NH$_{ε1}$</td>
<td>Top</td>
</tr>
<tr>
<td>126</td>
<td>Ser</td>
<td>Top/Interface</td>
</tr>
<tr>
<td>128*</td>
<td>Gly</td>
<td>Interface</td>
</tr>
<tr>
<td>131*</td>
<td>Val</td>
<td>Interface</td>
</tr>
<tr>
<td>141*</td>
<td>Val</td>
<td>Hydrophobic Core</td>
</tr>
<tr>
<td>144</td>
<td>Thr</td>
<td>Top</td>
</tr>
<tr>
<td>145</td>
<td>Asn</td>
<td>Top</td>
</tr>
</tbody>
</table>

* The chemical shift from the same residue/atom slightly shifted upon the addition of gpV$_C$-WT
Figure 48. Chemical Shift Changes in gpV\textsubscript{N} Upon Interacting with gpV\textsubscript{C}-G222D.

The cartoon representation of the lowest energy gpV\textsubscript{N} structure is shown in gray. Red sticks are used to highlight the residues for which backbone or side chain amide groups exhibited a perturbation in chemical shift upon the addition of gpV\textsubscript{C}-G222D.
Figure 49. gpV\textsubscript{C}-G222D Interacts with the ‘Top’ and Hexameric Interface of gpV\textsubscript{N}.

A cartoon representation of a gpV\textsubscript{N} hexameric model built based on Hcp1 is shown at two different angles. Each gpV\textsubscript{N} chain is coloured differently and for ease of visualization, residues 1-14, 20-36, 50-76, and 147-153 have been removed from the structure. The residues that exhibit chemical shift changes upon the addition of gpV\textsubscript{C}-G222D are represented as pink spheres. The majority of the residues are localized at the ‘top’ and hexameric interface of gpV.

### 5.3.3 G222D Causes gpV\textsubscript{C} to Partially Unfold at 25 °C

In order to provide insight into how the G222D mutation inactivates gpV function and the G222D/P227L double mutation recovers functions, I investigated whether any of the mutations within gpV\textsubscript{C} altered the structure of the C-terminal domain. To accomplish this, HSQC\textsubscript{s} were acquired for gpV\textsubscript{C}-WT, gpV\textsubscript{C}-G222D, gpV\textsubscript{C}-G222D/P227L, and gpV\textsubscript{C}-P227L (Figure 50), and the resulting spectra were overlaid for comparison (Figure 51).
Figure 50. HSQC Spectra of gpV\textsubscript{C} Mutants.

The HSQCs for (A) gpV\textsubscript{C}-WT (black), (B) gpV\textsubscript{C}-G222D (red), (C) gpV\textsubscript{C}-P227L (green), and (D) gpV\textsubscript{C}-P227L/G222D (magenta) are shown. HSQCs were acquired at protein concentrations of (A) 1.56mM, (B) 42uM, (C) 1.47mM, and (D) 0.137mM, respectively, in 50mM Na\textsubscript{2}HPO\textsubscript{4}, pH 6.8, 200mM NaCl at 25 °C.
Figure 51. A Comparison of the HSQC Spectra for each of the gpVc Mutants.

gpVc-WT (black) is shown superimposed on (A) gpVc-G222D (red), (B) gpVc-G222D/P227L (magenta), and (C) gpVc-P227L (green). For completeness, gpVc-P227L (green) is also compared to both (D) gpVc-G222D (red) and (E) gpVc-G222D/P227L (magenta). In panel (F), gpVc-G222D/P227L (cyan) is overlayed with gpVc-G222D (red). Chemical shift differences are observed in each case.

Upon comparison with gpVc-WT, chemical shift changes were apparent for each mutant. Notably, the peaks in the spectra acquired for gpVc-P227L and gpVc-G222D/P227L were well dispersed, similar to those observed in the gpVc-WT spectrum, while the chemical shifts in the gpVc-G222D HSQC were poorly dispersed with most of the peaks clustering at ~8.5 ppm in the $^1$H dimension (Figure 50B, Figure 51A), a property that is typical of an unfolded protein. In addition, approximately 96 peaks were counted in the gpVc-WT spectrum whereas 113 peaks were visible in the gpVc-G222D spectrum and while most of the gpVc-G222D chemical shifts were found clustered together in the middle of the spectrum, approximately 38 of the observed peaks for gpVc-G222D overlayed or partially overlayed with peaks in the gpVc-
WT spectrum (Figure 52). These peaks were termed ‘maintained’. On the other hand, there were approximately 31 shifts in the gpVc-WT spectrum that had no clear partner in the gpVc-G222D HSQC (Figure 52). These peaks were considered ‘not maintained’. Table 8 summarizes the residues for which peaks were ‘maintained’ and ‘not maintained’.

Figure 52. gpVc-WT and gpVc-G222D HSQCs Display Some Similarities.
The HSQC of gpVc-WT (black open circles) is superimposed with gpVc-G222D (red). Chemical shifts that are partially overlapped are boxed in blue. Overlapping peaks in the 8.5 ppm region were selected with caution. Due to the close proximity of some overlapping chemical shifts, some blue boxes surround more than one peak.
Table 8. Amide Chemical Shift Comparison Between gpV<sub>C</sub>-WT and gpV<sub>C</sub>-G222D

<table>
<thead>
<tr>
<th>Residues for which Amide Atom* Chemical Shift is:</th>
<th>Maintained</th>
<th>Not Maintained</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (tag)</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>18 sc</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>18 sc</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>66 sc</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>66 sc</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>27 sc</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>27 sc</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>58 sc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58 sc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Unless otherwise noted all numbers represent the residue number associated with the backbone amide chemical shift. ‘sc’ indicates that the amide group is from a side chain. Residue ‘0’ was introduced as part of a purification tag and remains after tag cleavage.
By mapping the residues for which chemical shifts were ‘maintained’ or ‘not maintained’ onto the structure of gpVC-WT (Chapter 4), a localization pattern was observed. ‘Maintained’ chemical shifts were localized to β1, β2, β3 and β6, while chemical shifts that were not maintained were localized to β4, β5, β7, and β8 (Figure 53). The G222D point mutation is located at the beginning of β7. Taken together, these preliminary data suggest that the G222D mutant may be partially unfolded.

Figure 53. Chemical Shift Comparison Between gpVC-WT and gpVC-G222D.

The chemical shift changes between gpVC-WT and gpVC-G222D were mapped onto the WT structure. Residues in Table 8 for which backbone amide chemical shifts were maintained and not maintained are coloured cyan and red, respectively. To visualize both sheets the structure is shown at two different angles and a black arrow indicates the position of glycine 222 (63 in the context of only gpVC).
5.3.4 P227L Induces Conformational Changes in gpVC

Since my results indicate that the G222D/P227L double mutant is functionally active and folded (Figure 44, Figure 50D), experiments were designed to further investigate the isolated P227L mutant in order to provide insight into how a single point mutation can prevent partial protein unfolding. To accomplish this, I first assigned the backbone chemical shifts of gpVC-P227L using standard triple resonance experiments. By comparing the chemical shifts of gpVC-WT and gpVC-P227L (Table 9), a significant difference was observed for residues 38, 41-42, 48-49, 67, 69 (area of point mutation), 75, 77-80, and 82 (Figure 51).

![Chemical Shift Comparison Between gpVC-WT and gpVC-P227L.](image)

The chemical shift difference between two peaks was calculated using both the $^1$H and $^{15}$N values for each peak as X and Y coordinates respectively, in the Pythagorean Theorem. The resulting chemical shift difference was plotted versus residue number to demonstrate where the greatest changes were taking place. The most striking changes (residue 67 and 69) occur in the region of the point mutation. Proline 227 is equivalent to residue 68 in the context of the isolated C-terminal domain.
Table 9. Chemical Shift Comparison between gpV<sub>c</sub>-WT and gpV<sub>c</sub>-P227L

<table>
<thead>
<tr>
<th>Residue Number</th>
<th>Chemical Shift Change</th>
<th>Residue Number</th>
<th>Chemical Shift Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.0284</td>
<td>46</td>
<td>0.1342</td>
</tr>
<tr>
<td>0</td>
<td>0.0440</td>
<td>47</td>
<td>0.0982</td>
</tr>
<tr>
<td>1</td>
<td>0.0303</td>
<td>48</td>
<td>1.0800</td>
</tr>
<tr>
<td>2</td>
<td>0.0323</td>
<td>49</td>
<td>0.6276</td>
</tr>
<tr>
<td>3</td>
<td>0.0955</td>
<td>50</td>
<td>0.3487</td>
</tr>
<tr>
<td>4</td>
<td>0.0377</td>
<td>51</td>
<td>0.0488</td>
</tr>
<tr>
<td>5</td>
<td>0.2860</td>
<td>52</td>
<td>0.1370</td>
</tr>
<tr>
<td>6</td>
<td>0.4131</td>
<td>53</td>
<td>0.1802</td>
</tr>
<tr>
<td>7</td>
<td>0.0694</td>
<td>54</td>
<td>0.0263</td>
</tr>
<tr>
<td>8</td>
<td>0.1040</td>
<td>55</td>
<td>0.3453</td>
</tr>
<tr>
<td>10</td>
<td>0.4890</td>
<td>56</td>
<td>0.1448</td>
</tr>
<tr>
<td>11</td>
<td>0.2160</td>
<td>57</td>
<td>0.1001</td>
</tr>
<tr>
<td>12</td>
<td>0.0922</td>
<td>58</td>
<td>0.1068</td>
</tr>
<tr>
<td>13</td>
<td>0.0091</td>
<td>59</td>
<td>0.0753</td>
</tr>
<tr>
<td>14</td>
<td>0.1008</td>
<td>60</td>
<td>0.1420</td>
</tr>
<tr>
<td>15</td>
<td>0.0796</td>
<td>61</td>
<td>0.0314</td>
</tr>
<tr>
<td>16</td>
<td>0.1082</td>
<td>62</td>
<td>0.1120</td>
</tr>
<tr>
<td>17</td>
<td>0.1695</td>
<td>63</td>
<td>0.0991</td>
</tr>
<tr>
<td>18</td>
<td>0.0762</td>
<td>64</td>
<td>0.1040</td>
</tr>
<tr>
<td>19</td>
<td>0.0141</td>
<td>65</td>
<td>0.1998</td>
</tr>
<tr>
<td>20</td>
<td>0.0641</td>
<td>66</td>
<td>0.2353</td>
</tr>
<tr>
<td>21</td>
<td>0.1136</td>
<td>67</td>
<td>1.7799</td>
</tr>
<tr>
<td>22</td>
<td>0.2734</td>
<td>69</td>
<td>4.0658</td>
</tr>
<tr>
<td>23</td>
<td>0.1372</td>
<td>70</td>
<td>0.2658</td>
</tr>
<tr>
<td>24</td>
<td>0.1233</td>
<td>71</td>
<td>0.4069</td>
</tr>
<tr>
<td>25</td>
<td>0.0609</td>
<td>72</td>
<td>0.5705</td>
</tr>
<tr>
<td>26</td>
<td>0.0673</td>
<td>73</td>
<td>0.3733</td>
</tr>
<tr>
<td>27</td>
<td>0.0400</td>
<td>74</td>
<td>0.1131</td>
</tr>
<tr>
<td>29</td>
<td>0.1011</td>
<td>75</td>
<td>0.8022</td>
</tr>
<tr>
<td>30</td>
<td>0.0700</td>
<td>76</td>
<td>0.1281</td>
</tr>
<tr>
<td>31</td>
<td>0.0502</td>
<td>77</td>
<td>1.4741</td>
</tr>
<tr>
<td>32</td>
<td>0.0480</td>
<td>78</td>
<td>0.8898</td>
</tr>
<tr>
<td>33</td>
<td>0.0434</td>
<td>79</td>
<td>0.6681</td>
</tr>
<tr>
<td>35</td>
<td>0.1281</td>
<td>80</td>
<td>1.4950</td>
</tr>
<tr>
<td>36</td>
<td>0.1228</td>
<td>81</td>
<td>0.2190</td>
</tr>
<tr>
<td>37</td>
<td>0.3680</td>
<td>82</td>
<td>0.6724</td>
</tr>
<tr>
<td>38</td>
<td>1.6280</td>
<td>83</td>
<td>0.0402</td>
</tr>
<tr>
<td>41</td>
<td>0.7444</td>
<td>84</td>
<td>0.0611</td>
</tr>
<tr>
<td>42</td>
<td>1.1875</td>
<td>85</td>
<td>0.0281</td>
</tr>
<tr>
<td>43</td>
<td>0.1454</td>
<td>86</td>
<td>0.0750</td>
</tr>
<tr>
<td>44</td>
<td>0.3117</td>
<td>87</td>
<td>0.0620</td>
</tr>
<tr>
<td>45</td>
<td>0.1140</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Residues that are shaded yellow exhibit a chemical shift difference greater than 0.6 ppm

Chemical shift differences were calculated as explained in the caption for figure 51. Residues ‘0’ and ‘-1’ were introduced as part of the purification tag and remain after tag cleavage.
Since both gpV_C-WT and gpV_C-P227L were purified in parallel using identical purification buffers and the NMR experiments were acquired at the same temperature, the chemical shift differences between the two domains are unlikely a result of sample conditions but rather due to the intrinsic properties of each protein. In fact, highlighting the residues that exhibit a chemical shift change of at least 0.6 ppm (Table 9) onto the gpV_C-WT structure (Figure 55), suggests that β7 and β8 may be elongated in gpV_C-P227L. The elongation of these two strands could stabilize the surrounding regions of the structure and protect against the detrimental effects of the G222D mutation.

**Figure 55.** gpV_C-P227L Chemical Shift Changes Mapped onto gpV_C-WT.

The structure of gpV_C-WT is shown in blue. Highlighted in red are residues for which backbone amides undergo a chemical shift change of at least 0.6 ppm in gpV_C-P227L when compared to gpV_C-WT. Proline 227 is coloured green in stick representation.

To further investigate the structural basis for the stabilization of gpV_C-P227L, I determined its solution structure using NMR spectroscopy (Figure 56, Table 10). The gpV_C-P227L structure is very similar to the gpV_C-WT structure with a backbone RMSD of 1.98 Å over 88 equivalent Cα positions. Not surprisingly, most of the differences between the two structures occur in the β7-β8 and surrounding regions (Figure 58). In the gpV_C-WT structure β7
is broken at Pro 227 (proline 68 in the context of just gpVc), which disrupts the hydrogen bonding between β7 and β8. In contrast, in the gpVc-P227L structure β7 and β8 are each extended (Figure 58).

Figure 56. The Solution Structure of gpVc-P227L.

(A) The ensemble of the twenty lowest energy gpVc-P227L structures. (B) Cartoon representation of residues 1-87 from the lowest energy structure. (C) Two-dimensional topology diagram of gpVc-P227L. In each panel, strands are coloured green and loops are coloured grey.
Table 10. Structural statistics for the ensemble of 20 structures for gpVC-P227L

<table>
<thead>
<tr>
<th>Distance restraints&lt;sup&gt;a&lt;/sup&gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Range,</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium Range, 1&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long Range,</td>
<td>i-j</td>
</tr>
<tr>
<td>Hydrogen bond pairs (HN-O, N-O)</td>
<td>18</td>
</tr>
</tbody>
</table>

Dihedral angle restraints

φ/ψ pairs (114 angles) | 57 |

Pairwise RMSD

<table>
<thead>
<tr>
<th>Secondary structure&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone atoms</td>
<td>0.53 ± 0.12 Å</td>
</tr>
<tr>
<td>All heavy atoms</td>
<td>1.04 ± 0.14 Å</td>
</tr>
</tbody>
</table>

Assigned atoms<sup>c</sup>

| Backbone atoms                  | 0.77 ± 0.15 Å |
| All heavy atoms                 | 1.39 ± 0.19 Å |

Ramachandran Statistics<sup>d</sup>

| Most favored regions            | 71.5% |
| Additional allowed regions      | 28.0% |
| Generously allowed regions      | 0.5%  |
| Disallowed regions              | 0.0%  |

<sup>a</sup>There were no distance violations > 0.5 Å. <sup>b</sup>RMSD for residues 5-8, 13-15, 19-26, 36-39, 46-51, 54-59, 63-71, 77-85. <sup>c</sup>RMSD for residues 1-87. <sup>d</sup>Determined using PROCHECK-NMR (Laskowski et al., 1996).

It should also be noted that unlike gpVC-WT, the position of the β4-edge strand was not difficult to determine in gpVC-P227L. In fact, upon comparing the HSQCs of the two proteins there is some evidence that the β4 region of gpVC-WT may be more dynamic than the same region in gpVC-P227L. More specifically, the backbone amide peaks for Val 39 and Ser 40 are more broadened in the gpVC-WT spectrum than in the gpVC-P227L HSQC (Figure 57).
Figure 57. Comparing the Broadening in the HSQC for gpV<sub>C</sub>-P227L and gpV<sub>C</sub>-WT. 

(A) A region of the HSQC for gpV<sub>C</sub>-P227L is enlarged and the backbone amide peaks from Val 39 and Ser 40 are labeled. (B) As a comparison, the region of the gpV<sub>C</sub>-WT HSQC containing the Val 39 and Ser 40 amide resonances is shown enlarged. The resonances from Val 39 and Ser 40 are significantly more broadened in the gpV<sub>C</sub>-WT spectrum.
Figure 58. Structural Comparison of gpVc-WT and gpVc-P227L.

(A) The superposition of the lowest energy structure of gpVc-WT (blue) with gpVc-P227L (green) is shown at two different angles. (B) Comparison of the two-dimensional topology of gpVc-WT (blue) with gpVc-P227L (green) allows for clear visualization of the break in β7 and β8 within gpVc-WT that does not occur in gpVc-P227L.

5.4 Discussion

In this work, biochemical and biophysical techniques have been used to assess the function, structure and stability of various mutations within gpVc. The single point mutation, G222D, was found not only to be inactive, but also to exert a temperature dependent dominant negative phenotype (Figure 44; Figure 45). A dominant negative phenotype often implies that while a mutant is functionally inactive, it is still capable of forming assembly intermediates but that these intermediates inhibit a step downstream to its first function. Since λ tail assembly can occur in the absence of gpVc and direct evidence of the role of gpVc in λ tail assembly is
lacking, this type of thought process cannot be easily applied to the dominant negative effects arising from mutations in gpV_C. An alternate explanation for the mechanism by which gpV_{FL-G222D} exerts its dominant negative behavior is that the G222D point mutation alters the conformation of the protein in such a way that enables it to interact with and sequester gpV_{FL-WT}, hence making it unavailable for interactions with another molecule of gpV_{FL-WT} during tail tube assembly. Preliminary data is presented that suggests gpV_C-G222D is partially unfolded at 25 °C and that this version of the domain is capable of interacting with gpV_N in a region that is predominantly localized to the ‘top’ and oligomerization surface of a gpV_N hexamer. Of the 13 residues implicated in interacting with gpV_C-G222D based on backbone chemical shift changes, 10 have a total residue exposure to solvent of greater than 20%. Furthermore, 50% of these residues are non-polar and hydrophobic. If we consider that partial unfolding of gpV_C-G222D would expose its hydrophobic core to solvent, one could speculate that the detrimental interactions may be in part driven by the hydrophobic effect.

Non-specific hydrophobic interactions can partially explain the inhibition of phi80 infections by gpV_C-G222D, however they cannot easily account for the differences in the amount of inhibition. gpV_C-G222D inhibits gpV_{FL-WT} by approximately 1,000,000 -fold, and yet it inhibits phi80 infections by only about 100-fold. gpV_N and the N-terminal domain of the phi80 major tail protein share 74% sequence identity. Among the identical residues are 10 hydrophobic core positions and if conservative substitutions are taken into consideration, 15 hydrophobic core positions are conserved strongly suggesting that the two proteins fold into similar structures. With such a high percentage of conservation between the two proteins it is not surprising that 9 of the 13 gpV_N residues implicated in interacting with gpV_C-G222D are also conserved in the phi80 major tail protein. The 4 non-identical interacting residues are each hydrophobic in the corresponding position of the phi80 MTP. This suggests that in addition to non-specific hydrophobic interactions there must be a certain degree of specificity involved in the mechanisms by which gpV_{FL-G222D} exerts its dominant negative effect.

The effect that gpV_{FL-G222D} had on HK97 infections supports the idea that the mutated protein utilizes more than one mechanism, both non-specific and specific, to exert its effects. First, HK97 infections were not inhibited by gpV_{FL-G222D}, a result that one would expect if specific interactions between the mutated protein and a wild-type tail protein were mediating the
dominant negative behavior; the N-terminal domain of the HK97 major tail protein shares only 7% sequence identity to gpV_N. While gpV_{FL}-G222D did not inhibit HK97 infections, it did visibly alter the appearance of the bacterial clearings, making them cloudier. The cloudiness of the plaques could indicate that gpV_{FL}-G222D is affecting the ability of the cell to lyse, decreasing the burst size of the infection, or driving more phage to enter the lysogenic cycle. If we once again consider that partial unfolding of gpV_C would expose its hydrophobic core, it is easy to imagine how it could interact non-specifically with other hydrophobic surfaces. In this manner, gpV_{FL}-G222D could potentially affect any protein in the cell with a hydrophobic surface patch.

While this hypothesis is in agreement with the scientific data presented, too many questions remain unanswered to draw a definitive conclusion regarding the mechanism by which gpV_C-G222D exerts its effect. For example, although it has been shown that gpV_C-G222D interacts with the monomeric N-terminal domain, it remains unclear as to whether gpV_C-G222D can also interact with the wild-type C-terminal domain or with hexameric gpV. In addition, it is still unclear whether the dominant negative effect is dependent on gpV_C-G222D being fused to gpV_N. Furthermore, it cannot be ignored that the partial unfolding of gpV_C-G222D and its interactions with gpV_N were observed at 25 and 20 °C, respectively. These temperatures fall below 37 °C, the temperature at which we observe the dominant negative phenotype in our biological assay. In order to complete this study, the folding and interaction studies will need to be repeated at 37 °C.

Previous bioinformatic studies on the λ gpV_C showed that a proline at position 68 (position 227 in full length) is unique to gpV and there are no other Ig-like domains known to have a proline at this position (James Fraser, personal communication). Typically, in other Big-2 domains position 68 is occupied by a serine or threonine residue (Fraser et al., 2006). It is therefore not surprising that gpV_C-P227L is more stable then gpV_C-WT. Since the side chain of residue 227 is oriented on the outside of the structure and the stabilization seems to arise simply from the introduction of the backbone amide group, I would speculate that many different residues can be substituted for Pro 227 to ‘protect’ the structure against G222D induced partial unfolding. It is intriguing to note that the conformational changes induced by P227L is not only localized to β7 and β8, but that the β4-edge strand also appears to be less dynamic in the P227L
structure compared to gpV<sub>C</sub>-WT. As described in Chapter 4, it was difficult to assess the location of β4 during determination of the gpV<sub>C</sub>-WT. In contrast, the location of β4 in gpV<sub>C</sub>-P227L was not difficult to determine. The HSQC of each protein suggests that the edge strand may be more dynamic in gpV<sub>C</sub>-WT than in gpV<sub>C</sub>-P227L.
6.1 Summary

In this thesis I have investigated two key steps in the λ tail assembly pathway: the self-assembly and polymerization of the major tail protein gpV, and the termination of tail tube polymerization by gpU, the tail terminator protein. More specifically, I have structurally and functionally characterized both gpV and gpU using a number of biophysical and biochemical techniques and these studies have led to significant advances in our understanding of phage tail assembly and evolution.

In Chapter 2, the X-ray crystal structure of gpU was presented. This was the first published structure of a biochemically characterized tail terminator protein from any long-tailed bacteriophage in its biologically relevant state. The structure helped guide our functional studies, which ultimately led to the identification of the residues involved in hexamerization, tail termination, and head binding. By comparing the hexameric crystal structure and monomeric solution structure (Edmonds et al., 2007), several differences were noted that appeared to be localized to one region of gpU. Furthermore, some of these conformational differences involved the ordering of unstructured regions in the monomeric structure. Taken together, I proposed that the unstructured regions in gpU may prevent the protein from prematurely interacting with its binding partners and that they undergo conformational ordering upon oligomerization and interaction with its binding partners. Last, based on structural similarities and a detailed bioinformatic analysis, a conserved mechanism of tail termination among many long-tailed phages, both Myoviridae and Siphoviridae, was identified.

My most significant scientific contribution in this thesis is the NMR solution structure of the N-terminal domain of the λ MTP, gpVN, presented in Chapter 3. Before the determination of the gpVN structure, atomic information on a MTP from any long-tailed phage was not available. My structure revealed that approximately 30% of the entire domain is disordered; the most notable disordered region is a 29 residue loop connecting β2 and β3. I demonstrated that the β2-β3-disordered loop is biologically important and plays a critical role in hexamer-hexamer
interactions during tail tube assembly. Interestingly, a structural similarity was identified between gpVN and a previously solved yet unpublished tail tube protein, XkdM, from the contractile-tailed phage PBSX. This finding provided the first direct evidence of an evolutionary relationship between the tail tubes of contractile and non-contractile phages and it is also consistent with my finding presented in Chapter 2 that TrPs from *Myoviridae* and *Siphoviridae* phages share common ancestry. In addition, a remarkable structural similarity was observed between gpVN and the hexameric structure of Hcp1, a component of the bacterial type VI secretion system. These data coupled with other similarities between phage and type VI secretion proteins (Pukatzki et al., 2007) support an evolutionary relationship between these systems. Furthermore, based on the structure of Hcp1 together with biochemical data, I was able to propose a model for the oligomerization and polymerization of gpVN involving several disorder-to-order transitions. One of the transitions proposed involves the C-terminus of gpVN, or in other words, the linker region between gpVN and gpVC, suggesting that the linker between the two domains holds a functional role. This hypothesis was validated in Chapter 4 by carrying out functional studies on various truncations of gpV that lacked gpVC and varying amounts of the linker that tethers the two domains together. In this manner I found that the linker between the N- and C-terminal domains is absolutely required for gpV function and that the absence of gpVC results in a 100-fold decrease in activity compared to gpVFL. In an attempt to provide insight into the biological role of gpVC, I solved the NMR solution structure of gpVC-WT. Structural similarities between gpVC and other Ig-like domains involved in cellular adhesion were identified suggesting a role in cell surface interactions; however, a functional role for gpVC has yet to be determined.

In Chapter 5, I showed that an amino acid substitution within gpVC, G222D, is inactive and exerts a temperature sensitive dominant negative phenotype. Biophysical techniques were used to determine that G222D causes the C-terminal domain to partially unfold and this unfolded form of the protein interacts with gpVN in a region that is primarily localized to its top and oligomerization surface. By combining gpVC-G222D with the P227L mutation, the functional and structural integrity of gpVC is restored. In addition, the structure of gpVC-P227L revealed that β7 and β8 are extended as compared to gpVC-WT and that these extensions also seem to stabilize the edge strand, β4.
6.2 Future Directions

While this study has provided significant insight into phage tail structure, assembly, and evolution, there are still many questions that remain unanswered. For example, what does a hexamer of gpV look like in atomic detail and how does it compare to the model proposed based on Hcp1? What is the functional role of gpVc? Does it play a role in stabilizing or shielding the tail once it has formed or does it play an accessory role in the infection process by interacting with the host cell? In addition, much of the data presented in chapter 5 is preliminary and many questions need to be addressed to precisely determine the mechanisms by which G222D exerts its temperature sensitive dominant negative phenotype and how P227L is able to overcome these effects. Experiments designed to address these questions are described below.

6.2.1 The Structure of a gpV Hexamer

My data strongly suggest that gpV undergoes one or more disorder-to-order transitions upon oligomerization and polymerization. By obtaining an X-ray crystal structure of gpV in its biologically relevant hexameric form, any conformational changes that take place upon oligomerization could be directly visualized. Since gpV contains a significant amount of disorder in its monomeric state (Chapter 3), the chance of obtaining diffraction quality crystals on this form of the protein is low. The isolation of gpV hexamers, and not monomers or dimers, from dissociated tail tubes at low pH (Bleviss and Easterbrook, 1971) strengthens the idea that the monomeric form of the protein undergoes one or more stabilizing transitions upon oligomerization. Based on this observation, it is likely that hexameric gpV is more amenable to crystallization studies than monomeric gpV. Therefore, in order obtain a hexameric X-ray crystal structure of gpV one must be able to purify the protein in its hexameric form. To accomplish this, a λ lysogen bearing an amber mutation in the V gene will be transformed with a plasmid encoding gpVFL with a C-terminal hexahistidine tag. After growth to the optimal cellular density, the lysogen will be induced to enter the lytic cycle while simultaneously inducing the overexpression of tagged gpVFL protein. In this manner, fully formed tails will be made. TEM can be used to assess the success of the complementation by observing the products
formed in each lysate. Since the construct of gpV that I am proposing to use is tagged at its C-terminus, the phage tails made by complementation can be purified from other cellular debris by passing the lysates over a Ni-NTA column using a protocol similar to those described in Chapters 2, 3 or 4. The purified tails can then be dialyzed into a buffer of pH 2, triggering the dissociation of the tail into individual hexamers of gpV (Bleviss and Easterbrook, 1971). One caveat of these experiments is that they will be very sensitive to both pH and salt concentration. For example, after the acid-induced dissociation of the tail, any step back towards neutrality results in the re-polymerization of gpV hexamers into long polypipe structures (Bleviss and Easterbrook, 1971). Fortunately, this re-polymerization reaction can be inhibited in the presence of 0.7M NaCl (Bleviss and Easterbrook, 1971). Therefore, in order to ensure sample homogeneity during crystallization screening, the dissociated gpV hexamers will have to be dialyzed into a buffer containing greater than 0.7M NaCl to ensure that once the protein sample is mixed with the crystallization buffer, the reduction in salt concentration does not alter the polymerization properties of the protein. Gel filtration can be used to further purify gpV hexamers and TEM can be used at each step in the purification protocol to monitor the protein behavior prior to crystallization screening.

### 6.2.2 Elucidating the Role of gpV<sub>C</sub>

My work has demonstrated that the removal of gpV<sub>C</sub> results in a 100-fold decrease in activity when compared to gpV<sub>FL</sub>. This result implies that while λ can still assemble in the absence of gpV<sub>C</sub> the domain clearly holds a yet unidentified advantage for infection. To provide insight into the role of gpV<sub>C</sub>, experiments will first need to confirm that phage tails are in fact being assembled in the absence of gpV<sub>C</sub>. This could be accomplished using an in vivo complementation assay whereby a gpV construct encoding residues 1-160 (Chapter 4, V<sub>160</sub>), is transformed into a λ lysogen bearing amber mutations in the V gene and expressed simultaneously with the induction of the lysogen. The success and products of the complementation reaction could easily be assessed by titrating out the resulting lysate on a suppressor E. coli strain and by directly visualizing the reaction products using negatively stained TEM. A complementation reaction with gpV<sub>FL</sub> and an empty vector will be required as a
positive and negative control, respectively. One possibility to take into consideration is that while V160 may still form tails, it may do so less efficiently than gpVFL. To address this issue, care will need to be taken to treat each sample identically so a direct comparison of phage titer can be made.

Based on the observed structural similarities between gpVC and other Ig-like domains involved in cellular adhesion, it has been speculated that the gpVC plays an accessory role in infection by weakly interacting with carbohydrates on the surface of the *E. coli* host cell. This interaction would presumably allow the phage to more readily interact with its host. In order to determine whether gpVC plays a role in host cell adsorption, λ phage particles that either possess or lack the C-terminal domain can be subjected to a simple kinetics adsorption assay as described by Hendrix *et al.* (Hendrix and Duda, 1992). In this assay, phages of a known concentration are mixed with an excess amount of *E. coli* cells and incubated at 37 °C. At various time points, a small amount of sample is removed and centrifuged to pellet the bacterial cells. The supernatant can then be assayed for phage activity on a suppressor strain, where activity is indicative of unadsorbed phages. The values obtained can be compared between the phages whose tails possess or lack the C-terminal domain.

Alternatively, the C-terminal domain may act to stabilize and/or shield the tail during infection. If this were true, then other domains of comparable size should be able to sufficiently restore function in λ phage lacking gpVC. To test this hypothesis, domain swap experiments can be carried out whereby gpVC is replaced with either Ig-like domains from other bacteriophage morphogenic proteins or completely unrelated domains that lack an Ig-like fold. For example, Ig-like domains from phages that exhibit similar host specificities as λ, or Ig-like domains from phages that infect unrelated gram-positive bacteria and hence are exposed to an entirely different cell surface, can be swapped with λ gpVC. I have already begun to generate the constructs required to carry out these experiments. In these constructs, residues 1-160 from gpV are fused to the Ig-like domains from phages phi80 (residues 161 to 247), N15 (residues 161-245), HK97 (residues 145-234), A2 (residues 214-295), as well as the SH3 domain from Abp1 (residues 1-59). Sequence alignment data was used to determine the boundaries of the C-terminal fusions.
6.2.2 Elucidating the Mechanism by which G222D exerts its Dominant Negative Behavior

6.2.2.1 The Temperature Dependence of G222D

My preliminary data presented in Chapter 5 suggests that the single point substitution G222D within gpV_C causes the domain to partially unfold at 25 °C and that this partially unfolded version of the domain, but not gpV_C-WT, is able to interact with gpV_N. This finding allowed me to propose a putative model for the mechanism by which the G222D mutant exerts its dominant negative behavior, however there are not yet enough data to draw a definitive conclusion.

It should be mentioned that G222D only exerts its detrimental phenotype at 37 °C and not 30 °C. It is therefore unlikely that G222D exerts a dominant negative phenotype at 25 °C, the temperature at which the NMR experiments were conducted. Therefore, in order to provide a thorough understanding of how the G222D mutation is functioning, each of the NMR experiments will need to be repeated at both 30 and 37 °C. First, HSQC spectra of gpV_C-WT, gpV_C-P227L, gpV_C-G222D/P227L, and gpV_C-G222D will be acquired at 25 °C, then at 30 °C, and 37 °C. As the temperature increases, I speculate that the unfolding of the domain will become more severe. Once the structural integrity of each of the mutants is assessed at elevated temperatures, the interaction studies between gpV_N and both gpV_C-G222D and gpV_C-WT will also be repeated at 30 °C and 37 °C. These experiments will require that gpV_N is also assessed at 30°C and 37 °C to ensure that its structure is not significantly affected at these elevated temperatures. Once this is accomplished, gpV_N will be mixed with gpV_C-WT or gpV_C-G222D and the mixture will be equilibrated at the desired temperature before an HSQC is acquired.
6.2.2.2 The Effect of Time and Protein Concentration on Interaction Studies

The interaction experiments carried out to date were acquired immediately on a sample of gpV_N mixed in a 1:1 ratio with either gpV_C-WT or gpV_C-G222D. It is therefore of interest to determine whether the outcome is altered either after prolonged periods of time or at varying concentrations of gpV_C. To test this, I will carry out titration experiments whereby increasing amounts of unlabelled gpV_C-WT or gpV_C-G222D will be mixed with gpV_N in a step-wise manner and after each addition an HSQC spectrum will be acquired. In this manner, the minimum amount of gpV_C required to interact with gpV_N will be determined. Furthermore, it will be interesting to determine if time has an effect on the interaction between the two domains. To address this, I will mix 15N-gpV_N with unlabelled gpV_C-WT or gpV_C-G222D and record a HSQC spectrum once every hour for up to or exceeding 12 hours and then overlay the resulting spectra for comparison.

6.2.2.2 The Requirements to Produce a G222D Dependent Dominant Negative Phenotype

Thus far, functional assays designed to assess the dominant negative phenotype caused by G222D were carried out using constructs encoding mutated full-length gpV (gpV_Fl-G222D). The fact that my preliminary interaction studies presented in Chapter 5 showed that gpV_C-G222D on its own interacts with gpV_N raises the question as to whether gpV_C-G222D alone is sufficient to observe a dominant negative phenotype or whether it must be tethered to gpV_N. To answer this question, the C-terminal domain and its mutants will be sub-cloned into the pAD vector used in the in vivo complementation studies presented in Chapter 2, 3, 4 and 5. Once they are successfully sub-cloned, each vector will be tested for its ability to inhibit wild-type λ. Furthermore, the experiments carried out to date have also failed to address whether gpV_C-G222D is also capable of interacting with gpV_C-WT. There are several ways one could go about addressing this question. First, the in vitro binding experiments can be repeated but using 15N-gpV_C-WT in place of 15N-gpV_N. Second, one could also test whether gpV_Fl-G222D is able
to exert a dominant negative phenotype on λ phages that lack the C-terminal domain; our lab possesses several variants of λ that lack the C-terminal domain. The absence of a dominant negative phenotype would suggest that these phages are resistant to the effects of gpV_{FL}-G222D and that part of the mechanism by which gpV_{FL}-G222D exerts its effect involves the C-terminal domain of gpV.
REFERENCES


150


