Structure and Function in Bacteriophage Phi6

James Carpino
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu(gc_etds
Part of the Molecular Biology Commons, and the Virology Commons

Recommended Citation
Carpino, James, "Structure and Function in Bacteriophage Phi6" (2014). CUNY Academic Works.
https://academicworks.cuny.edu(gc_etds/183

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
STRUCTURE AND FUNCTION IN BACTERIOPHAGE PHI6

JAMES CARPINO

Dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014
This manuscript has been read and accepted for the
Graduate Faculty in Biology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Date  Chair of Examining Committee
Dr. John Dennehy

Date  Executive Officer
Dr. Laurel Eckhardt

Dr. John Dennehy, Queens College

Dr. Cathy Savage-Dunn, Queens College

Dr. Stéphane Boissinot, Queens College

Dr. Paul Gottlieb, City College of New York

Dr. Dan Dykhuizen, Stony Brook University, SUNY
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

STRUCTURE AND FUNCTION IN BACTERIOPHAGE Φ6

by

JAMES CARPINO

Advisor: Professor John Dennehy

The present study of bacteriophage φ6 has been preceded by a great number of exploratory studies of its structure and function, and these studies have formed a basis for φ6’s development into a model organism. In this study, two aspects of the model organism have been examined. 1. There are several uncharacterized and presumed untranslated regions (UTRs) in φ6’s 13.3 kilobase-pair dsRNA genome. I examined the impact of specific modification to the 3’ UTR of the small segment of bacteriophage φ6. I determined that modification to the purported UTR of the small segment resulted in severe fitness costs, supporting a functional role for unidentified gene products, secondary RNA structure, or both. 2. Bacteriophage φ6 packages its dsRNA genomic segments selectively and sequentially through the function of the packaging motor P4 which occupies fivefold vertices of the φ6 procapsid, and studies support the functioning of one and only one P4 during packaging. The mechanism of this specific phenomenon is not known. I used computational reconstruction of cryoelectron microscopy and examined the occupancy of P4 on the φ6 procapsid, and acquired insight into the mechanism of assembly and packaging.
Acknowledgement and Dedication

I owe my sincerest thanks to a long list of people without whose help and support, my success would have been highly improbable. The list is too long to present here; it includes many advisors, professors, friends, classmates, students, teammates, and one true love, all of whom know who they are.

This work is dedicated to the loving memory of my father, Dr. Joseph James Carpino, October 15, 1930 – August 4, 1998. May he rest in peace, and watch over us all with his beloved son, John.

My father was the first to teach me to embrace knowledge, truth, and justice, and to seek them through patience, perseverance, and personal sacrifice. This lesson was one of many which he imparted passively, by example, by allowance, and never by force or imposition. He was a teacher, a philosopher, a soldier, a carpenter, and a Christian, and a father more deeply in his heart than anything else.

He advised me never to be stingy with ideas, lest the river from which they flow run dry.
TABLE OF CONTENTS

ABSTRACT........................................................................................................................... IV

ACKNOWLEDGEMENT AND DEDICATION............................................................................ V

CHAPTER 1: INTRODUCTION.......................................................................................... 1

1.1 BACTERIOPHAGE Φ6 AS A MODEL ORGANISM......................................................... 1

1.2 BACTERIOPHAGE Φ6 DESCRIPTION......................................................................... 4

1.3 BACTERIOPHAGE Φ6 LIFE HISTORY......................................................................... 5

1.4 GENETIC EXCHANGE, CYSTOVIRIDAE AND REASSORTMENT OF SEGMENTED GENOMES.................................................................................................................. 7

CHAPTER 2: EVIDENCE FOR A FUNCTIONAL ROLE IN THE 3’ UTR OF THE SMALL SEGMENT OF BACTERIOPHAGE Φ6................................................................. 11

ABSTRACT: .......................................................................................................................... 11

2.1 INTRODUCTION........................................................................................................ 12

2.1.1 φ6 Tagging Project...............................................................................................12

2.1.2 The 3’ UTR of the Small Segment of Bacteriophage φ6 .......................................14

2.1.3 The Case for Functional Secondary Structure in the 3’ UTR ..............................15

2.1.4 The Case for Genes............................................................................................15

2.1.5 The Curious Precedent of the Bullseye Plaque ..................................................20

2.2 EXPERIMENTAL METHODOLOGY ....................................................................... 25

2.3 MATERIALS AND METHODS ................................................................................ 28

2.3.1 Bacterial and Viral Strains and Plasmids............................................................28

2.3.2 Culture Conditions...............................................................................................29

2.3.3 Preparation of Plasmids......................................................................................29
2.3.4 Genetic Modification of RNA Segments of φ6 ..................................................29
  2.3.4.1 Mutagenesis Methods .................................................................................29
  2.3.4.2 Mutant Classes and Segment Variants .........................................................31
  2.3.5 Recombinant Bacteriophage Rescue .............................................................33
    2.3.5.1 Competent Cell Preparation .....................................................................33
    2.3.5.2 Electroporation and Plating .....................................................................34
  2.3.6 Plaque Isolation and Purification .................................................................35
  2.3.7 RNA Extraction ............................................................................................36
  2.3.8 cDNA Synthesis ..........................................................................................37
  2.3.9 PCR Detection of Inserts .............................................................................37
  2.3.10 Sequencing ..................................................................................................37
  2.3.11 Fitness Assay .............................................................................................37
  2.3.12 OD Depletion Assay ..................................................................................38
  2.3.13 Growth Curve of JC1917 ...........................................................................39
  2.3.14 RNA Secondary Structure Prediction .........................................................39
  2.3.15 Protein 3D Structure Prediction .................................................................39

2.4 RESULTS ........................................................................................................40
  2.4.1 Fitness Impact of Modification to the 3’ Region of φ6 Small Segment ........40
  2.4.2 Fitness of Small Tag Mutant .........................................................................40
  2.4.3 Identification of ORFs ..................................................................................40
  2.4.4 Prediction of Secondary Structure; Impact of Tagging on Secondary
      Structure ........................................................................................................40
  2.4.5 Construction of New UTR Mutants ............................................................43
2.4.6 DNA Synthesis and Sequence Verification ................................................................. 43
2.4.7 Fitness of Newly constructed Strains MUT1 through MUT6 ............................... 43
2.4.8 Relationship between Fitness and Optical Density ............................................... 45
2.4.9 Growth Curve of JC1917 ....................................................................................... 47

2.5 DISCUSSION .............................................................................................................. 49

2.5.1 Comparisons .......................................................................................................... 50

2.5.2 MUT2 vs. MUT5 Comparison Reflects Altered ssRNA Stability ..................... 50

2.5.3 MUT1 vs. MUT3: ORF J Impacts Plaque Morphology but not Fitness ............. 51

2.5.4 MUT2 and MUT3 demonstrate Different Impacts on Fitness, Supporting Function of One or Two Genes ................................................................. 51

2.5.5 Deletions vs. Insertions: Both Genes and Secondary Structure Impact Fitness ...... 53

2.5.6 Observation: Fitness and Plaque Size are Not Universally Correlated .......... 54

2.5.7 Gene Discovery ................................................................................................ ...... 54

2.5.8 JC1917 Appears to Enhance Bacterial Growth .................................................... 54

2.6 CONCLUSION ............................................................................................................ 55

CHAPTER 3: TRANSIENT OCCUPANCY OF P4 IN THE PORTALS OF THE Φ6 PROCAPSID ................................................................................................................. 56

3.1 ABSTRACT .................................................................................................................. 56

3.2 INTRODUCTION ......................................................................................................... 56

3.2.1 The Φ6 RNA Packaging Motor ........................................................................... 56

3.2.2 P4 Occupancy and Phage Particle Function ....................................................... 57

3.2.3 Cryoelectron Microscopy ....................................................................................... 59
3.2.4  Single Particle Analysis .................................................................59
3.2.5  Tomographic Reconstruction and Subtomogram Averaging ..........62
3.2.6  Homology Modeling .....................................................................65

3.3  MATERIALS AND METHODS .......................................................... 65
3.3.1  Sample Preparation and Electron Microscopy ...............................65
3.3.2  Single Particle Analysis .................................................................65
3.3.3  Measurement of Densities in Reconstructions ...............................66
3.3.4  Tomographic Reconstruction ........................................................66
3.3.5  Subtomogram Averaging ...............................................................66
3.3.6  Portal Averaging ...........................................................................67
3.3.7  Density Measurement and Image Production ...............................67
3.3.8  Homology Modeling ....................................................................67

3.4  RESULTS .......................................................................................... 69
3.4.1  P4 Occupancy on the Procapsid is Low .......................................69
3.4.2  P4 Occupancy is Correlated to P8 Occupancy in the Expanded
       Nucleocapsid ....................................................................................71
3.4.3  Observation of Interaction between P8 and P4 in the Expanded
       Nucleocapsid ....................................................................................73
3.4.4  Observation of Interaction Region between P4 and P1 in the Unexpanded
       Procapsid ........................................................................................76
3.4.5  Evidence for Asymmetric Attachment in the Unexpanded φ6 Procapsid ....77
3.4.6  Visualization of P4 Homology Model and C-Terminal Region Essential for
       Attachment ......................................................................................79
List of Tables

Table 1: Bacteriophage Families ................................................................. 2
Table 2: dsRNA Virus Families................................................................. 2
Table 3: Proteins of Bacteriophage φ6 ......................................................... 9
Table 4: Sequence of Insert into φ6 Medium Segment ............................ 21
Table 5: φ6 Small-Segment Mutation Classes, Methods, and Primers Used ........................................... 32
Table 6: Summary of Pi6 Small Segment Mutations ................................ 32
Table 7: Summary of Procapsid Occupancy Data ....................................... 70
Table 8: Summary of Reconstructions and Density Data .......................... 72
Table 9: Measurement of material in presumed unoccupied portal ............ 96
Table 10: Sequence Alignment of φ6 P4 and φ12 P4 ................................... 97
## List of Figures

Figure 1: Bacteriophages and their related Viruses ................................................................. 3  
Figure 2: Schematic Representation of Bacteriophage φ6 ..................................................... 4  
Figure 3: Life Cycle of Bacteriophage φ6 .............................................................................. 5  
Figure 4: Maps of the small, medium and large segments of φ6 ........................................ 10  
Figure 5: Plaques of wild-type, medium-segment tagged, and small-segment tagged φ6 .... 13  
Figure 6: Schematic of φ6 Small Segment and Uncharacterized ORFs ............................... 17  
Figure 7: Autoradiogram of Protein Gel of Bacteriophage φ6 Nonsense Mutants .............. 20  
Figure 8: Medium segment of φ6 Tagged with Signature ................................................... 21  
Figure 9: Bullseye plaques ..................................................................................................... 22  
Figure 10: Fitness of Bullseye strain relative to Wild-type and Wild-type Rescue strain..... 22  
Figure 11: Turner lab Bulseye Plaques ................................................................................... 24  
Figure 12: Location of 1bp insertion for MUT2 ..................................................................... 27  
Figure 13: Schematic Representation of φ6 Small Segment Mutations .............................. 31  
Figure 14: Plaque Picking Procedure ................................................................................... 36  
Figure 15: Fitness Assay in Liquid Culture (Schematic) ........................................................ 38  
Figure 16: Secondary structure of φ6 Small Segment ............................................................ 41  
Figure 17: Secondary structure of φ6 Small Segment 3’ UTR ............................................... 41  
Figure 18: Impact of Small segment Insert on Secondary Structure .................................... 42  
Figure 19: Predicted structures of ORF I, ORF J ................................................................. 42  
Figure 20: Fitness of Mutants of Bacteriophage φ6 Small Segment 3’ Region .................... 44  
Figure 21: Plaque morphology of selected mutants ............................................................. 45  
Figure 22: Relationship between Fitness and OD Depletion ............................................... 47
Figure 23: Growth Curve of JC1917 vs. Control................................................................. 48
Figure 24: Excavated Dodecahedral Structure of φ6.......................................................... 58
Figure 25: Cryo-electron microscopy cross-section of the φ6 Nucleoapsid......................... 61
Figure 26: Isosurface Representation of Single Particle Reconstruction............................. 62
Figure 27: Cryoelectron tomography tilt series .................................................................. 63
Figure 28: Closeup of Tomogram, φ6 Procapsids with Occupied Portals............................... 63
Figure 29: Density measurements of P1, P4 in the procapsid ............................................. 70
Figure 30: Examples of Density Measurements in Nucleocapsid ....................................... 72
Figure 31: Density of P4 vs. P8 ......................................................................................... 73
Figure 32: Contact points visible in nucleocapsid reconstruction ....................................... 74
Figure 33: Contact points between P1 and P4 in Nucleocapsid ........................................... 74
Figure 34: Coordinates of Peak Densities in Nucleocapsid Interaction Zone .................... 75
Figure 35: Single particle reconstruction of procapsid ........................................................ 76
Figure 36: Interaction between P4 and P1 Core in φ6.......................................................... 77
Figure 37: P4 Occupancy in Relaxed-Symmetry Reconstruction ......................................... 78
Figure 38: Reconstruction of Occupied Portals with Relaxed Symmetry ............................. 79
Figure 39: φ6 P4 and φ12 P4 structural alignment ............................................................... 80
Figure 40: Hypothetical model of P4 hexamer ................................................................. 80
Figure 41: Three views of P4 monomer in context............................................................. 81
Figure 42: Model for Sequential Packaging and P4 Stabilization ....................................... 85
Figure 43: Asymmetrical anchoring of P4 on the Procapsid .............................................. 88
Figure 44: “Wobble” of the P4 hexamer ............................................................................ 88
Figure 45: Fitness of 35 isolated φ6 mutants ....................................................................... 95
Figure 46: Fitness of 30 strains vs. Depletion in Optical Density ........................................ 95

Figure 47: Observation of material in presumed unoccupied portal ..................................... 96

Figure 48: Procapsid or Polymerase Complex (PC) vs. Nucleocapsid (NC) ......................... 98
Chapter 1:
Introduction

1.1 Bacteriophage φ6 as a Model Organism

Bacteriophage φ6 is an obligate intracellular parasite of *Pseudomonas* bacterial hosts. φ6 was the first lipid-coated, dsRNA bacteriophage to be isolated. φ6 is now a member of a broad family of lipid-coated dsRNA bacteriophages, Cystoviridae. Cystoviridae bears several features in common to the human pathogen reoviridae, including a 120-subunit protein inner core, and a subdivided dsRNA genome (Mindich 2004). Cystoviridae owes some of its utility as a model organism to this similarity (Mindich 2012).

Taxonomy in viruses suffers from the same challenges as taxonomy in all other realms of biology: Structure and morphology-based classification is limited, and is gradually revised over time by analyses of sequence similarity (Krupovic et al. 2011). New studies in similarity between viruses have been more quantitative because of available sequence and structure information. Some recent studies have examined the structural relationship between distantly related viruses (Bamford et al. 2002; El Omari et al. 2013) and structural analysis in other viruses suggest common ancestry between viruses and bacteriophages (Selvarajan Sigamani et al. 2013).

Table 1 represents current known taxonomic groups of bacteriophages (Krupovic et al. 2011). Figure 1 illustrates the similarity between examples of bacteriophages and viruses (Bamford et al. 2002). Tailed, dsDNA bacteriophages are among the most reported in the literature (in contrast with RNA or ssDNA phages), but this bias has not yet been verified in nature and may be influenced by the laboratory culturability of hosts, as well as limitations of
methods of gene discovery favoring DNA vs. RNA (Krupovic et al. 2011). However dsRNA viruses have been found to infect a wide variety of hosts (Table 2) (Mertens 2004).

Table 1: Bacteriophage Families
Classical taxonomic breakdown of bacteriophage by structure and nucleic acid composition (Krupovic et al. 2011)

<table>
<thead>
<tr>
<th>Family</th>
<th>Number of genome segments</th>
<th>Type of virus particle</th>
<th>Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypoviridae</td>
<td>1 (unpackaged)</td>
<td>~50–80 nm diameter, pleomorphic vesicles (no capsid)</td>
<td>Fungi</td>
</tr>
<tr>
<td>Temoviridae</td>
<td>1 (packaged singly)</td>
<td>~30–40 nm diameter, icosahedral</td>
<td>Fungi</td>
</tr>
<tr>
<td>Birnaviridae</td>
<td>2 (co-packaged)</td>
<td>~60 nm diameter icosahedral, single shell</td>
<td>Fish, insects, birds, molluscs</td>
</tr>
<tr>
<td>Paramyxovirus</td>
<td>2 (separately packaged)</td>
<td>~18 × 320–360 nm, rod shaped</td>
<td>Plants</td>
</tr>
<tr>
<td>Partitiviridae</td>
<td>2 (separately packaged)</td>
<td>~30–40 nm diameter, icosahedral protein capsid</td>
<td>Fungi plants</td>
</tr>
<tr>
<td>Cystoviridae</td>
<td>3 (co-packaged, equimolar)</td>
<td>~85 nm diameter, three layer structure, with an envelope surrounding a two-layered icosahedral nucleocapsid</td>
<td>Bacteria (<em>Pseudomonas</em>)</td>
</tr>
<tr>
<td>Chrysoviridae</td>
<td>4 (packed separately)</td>
<td>~30–40 nm diameter icosahedral protein capsid</td>
<td>Fungi</td>
</tr>
<tr>
<td>Reoviridae</td>
<td>10, 11 or 12 (co-packaged, equimolar)</td>
<td>~70–90 nm diameter icosahedral (one, two or three layered protein capsid)</td>
<td>Insects, plants, fish, reptiles, birds, mammals, arachnids, fungi, arthropods, crustaceans</td>
</tr>
</tbody>
</table>

Table 2: dsRNA Virus Families
(Mertens 2004)
Bamford (2002) reflects groupings of viruses by structure; this distinction sheds light on the apparent relatedness of eukaryotic viruses (red) and prokaryotic viruses (green). Top panel: dsDNA lipid-containing viruses are represented in the bacteriophage realm by PRD1 and in the eukaryotic realm by Adenovirus and PBCV-1. Middle panel: dsRNA bacteriophage are represented by φ6, a member of the Cystoviridae family; and dsRNA eukaryotic viruses are represented by Bluetongue virus, a member of the Reoviridae family. Bottom panel: Tailed viruses in prokaryotes, archaea (yellow). While no known tailed eukaryotic virus is known, Herpesvirus is proposed to be related (Bamford et al. 2002).
1.2 Bacteriophage φ6 Description

The φ6 genome is divided into three dsRNA segments, named S, M, L, which are 2948, 4063, and 6374 base pairs in length, respectively (Figure 4). These segments are contained in a symmetric protein shell consisting primarily of protein P1, which forms an excavated dodecahedron, the 12 “portals” of which may contain P4, the packaging motor. The inner core consists of the main structural protein P1, the polymerase P2, the packaging motor P4 and the auxiliary assembly catalyst P7 (whose exact function is not fully understood), together form what is called the procapsid (Figure 48). Inclusion of the RNA genome and the structural protein P8 define the nucleocapsid; addition of the lipid bilayer and all membrane proteins defines the complete phage particle or virion (Figure 2) (Jäälinoja et al. 2007).

![Schematic Representation of Bacteriophage φ6](image)

**Figure 2: Schematic Representation of Bacteriophage φ6**
The Bacteriophage φ6 particle contains a protein core or procapsid (PC) consisting of P1, P2, P4, and P7. A second protein layer consists of P8 and interacts with P4 and also contains murein peptidase P5 and P11 (which are transcript variants of the same gene). The outer layer is a lipid envelope and retains the host attachment protein P3 and its membrane-bound anchor P6, as well as P9, lysin P10, and P13. Note that P12 and P14 are not pictured, as they are not integral to the particle. (Mindich 1999)
Figure 3: Life Cycle of Bacteriophage φ6
The Phi6 particle (a) attaches to the host pilus through interaction between the major pilus protein (pilin) and P3. Retraction of the pilus and subsequent interaction with viral protein P6 facilitate fusion of the Phi6 envelope with the bacterial outer membrane (om). Murein peptidase P5 digests peptidoglycans in the periplasmic space and interaction with the inner membrane results in invagination. Inner membrane vesicle is shed, and the nucleocapsid becomes transcriptionally active (d), synthesizing strands of large, medium, and small segments, of which the large segment is translated first, providing proteins for nucleation of the empty procapsid (pc). Strands of ssRNA are packaged sequentially into the procapsid, resulting in its incremental expansion (e) and eventual acquisition of the protein layer P8. Protein P9 facilitates formation of intracellular vesicles (g) which are precursors to the mature virion (mv) envelope; vesicles envelop the nucleocapsid by interaction of P9 with P8. Buildup of mature viruses is followed by rupture of the cell facilitated by P10 and P5, and the release of phage particles into the environment. Adapted from Mindich 1999.

1.3 Bacteriophage φ6 Life History

Bacteriophage φ6 host attachment is facilitated by its surface protein P3. P3 attaches to the pilin protein of a Pseudomonas host’s pilus, and the host’s retraction of the pilus facilitates entry (Romantschuk & Bamford 1985) (Figure 3). P6, which anchors P3 in the φ6 envelope, is a factor in fusion of the φ6 envelope with the Pseudomonas outer membrane (Bamford et al. 1987;
After fusion, φ6 enters the periplasmic space and passes through the peptidoglycans layer, where protein P5, a murein peptidase, functions to break down peptidoglycans (Mindich & Lehman 1979; Caldentey & Bamford 1992; Hantula & Bamford 1988). The phage nucleocapsid (NC) passes through and out of the periplasmic space and passes the inner membrane, where P8 mediates budding into the inner membrane (Romantschuk et al. 1988). The newly acquired inner membrane coat is subsequently shed along with P8. The release of P8 activates the “polymerase complex” (PC). The core proteins P1, P2, P4, and P7 act to transcribe plus-strands using internal genomic φ6 RNA as a template. Strands are synthesized by the RNA-dependent RNA polymerase P2, and transited passively into the cytoplasm by the packaging motor P4 (Lísal et al. 2006; Kainov et al. 2004; Koonin et al. 1989; Gottlieb et al. 1990).

Free large-segment RNA is translated, leading to the translation of proteins for new procapsids (PC) (Figure 48). Procapsid are assembled from P1, P2, P4, and P7; the assembly proceeds hierarchically, and is dependent on magnesium (Mg++) and NTP or ADP (Poranen & Tuma 2004; Mindich 1999; Paatero et al. 1995). Empty procapsids accumulate, and are able to accept newly synthesized single-stranded RNA (ssRNA) genomes. ssRNA is packaged into procapsids sequentially, specifically in order from small to large segment (Mindich 2004; Huiskonen et al. 2006; Mindich 1999). The process of precise packaging is fairly well studied, and is believed to involve sequential recognition of strands by P1, the capsid’s primary structural constituent (Mindich 2004; Huiskonen et al. 2006; Onodera et al. 1998b; Gottlieb et al. 1992; Onodera et al. 1995; X. Qiao et al. 2003; Qiao & Mindich 2013; Qiao, Casini, et al. 1995). The capsid expands as RNA is packaged, and the structure of empty, partially expanded, and expanded procapsids has been elucidated by cryo-electron tomography (Nemecek et al. 2011).
Packaging of all three plus-strand ssRNA molecules is believed to be completed prior to synthesis of complementary minus strands by P2 (Frilander et al 1992). Separately, pre-envelope vesicles form in the cytoplasm, directed by the membrane-bound P9, which interacts with P8 (Sarin et al. 2012). These vesicles envelop procapsids once packaging is complete, and the addition of P8 and the addition of the lipid envelope with viral proteins marks the formation of the complete virion. After the accumulation of as many as 200 virions, P5 and P10 mediate the rupture of the cell membrane and the release of virions into the environment (Mindich & Lehman 1979).

1.4 Genetic Exchange, Cystoviridae and Reassortment of Segmented Genomes

One of the advantages of Cystoviridae as a model organism is its similarity to mammalian dsRNA viruses with segmented genomes. The similarity is not only structural but behavioral, with respect to genetic exchange (Silander et al. 2005). RNA viruses generally do not undergo homologous recombination as do bacteria (however infrequently) and eukaryotes (in germ line cells) (Duffy et al. 2008; Silander et al. 2005). Recombination events are unlikely because the Cystoviridae life cycle constrains replication to the interior of the precisely packaged capsid. However, two other forms of genetic exchange occur; non-homologous recombination, and segment reassortment.

Non-homologous recombination in φ6 has been shown to occur after modification of the 3’ UTR of the medium segment. Significant modification results in spontaneous substitution of the modified strand with a portion of the 3’ UTR from another segment, also resulting in the loss of the insert, suggesting that there is a stabilizing role for the 3’ UTR of that segment (Onodera et al. 1993; Mindich et al. 1992). Non-homologous recombination is believed to function in recovery of the essential 74bp functional region in the 3’ UTR (Mindich 2004).
Reassortment occurs during infection of a host cell by multiple viruses. Under these circumstances, plus-strands of ssRNA are manufactured for both viruses and subsequently packaged without preference as to the parental source. This results in hybrid genomes. The advantage of this exchange is the same advantage imparted by sex in sexually reproducing species, and the phenomenon has been compared to sex in several studies (Chao 1988; Poon & Chao 2004; Froissart et al. 2004; Silander et al. 2005; O’Keefe et al. 2010).
<table>
<thead>
<tr>
<th>Segment</th>
<th>Gene</th>
<th>Protein</th>
<th>Size (kDa)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>1</td>
<td>P1</td>
<td>85.0</td>
<td>Major capsid protein (Poranen &amp; Tuma 2004)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>P2</td>
<td>74.8</td>
<td>RNA dependent RNA polymerase responsible for plus and minus strand synthesis (Mindich 2004)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>P4</td>
<td>35.0</td>
<td>Responsible for genomic packaging (Mindich 2004).</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>P7</td>
<td>17.2</td>
<td>Stabilizes the procaspid and takes part in regulation of plus strand synthesis (Juuti &amp; Bamford 1997).</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>P14</td>
<td>6.8</td>
<td>Nonessential non-structural, regulatory (Casini &amp; Revel 1994)</td>
</tr>
<tr>
<td>Medium</td>
<td>3</td>
<td>P3</td>
<td>69.2</td>
<td>Adsorption to host cell (host attachment) (Gottlieb et al. 1988).</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>P6</td>
<td>17.2</td>
<td>Adsorption to host cell (anchor) (Gottlieb et al. 1988).</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>P10</td>
<td>4.3</td>
<td>Necessary for host lysis (Gottlieb et al. 1988).</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>P13</td>
<td>7.6</td>
<td>Minor function, nonessential in recent studies (Gottlieb et al. 1988; Mindich et al. 1992).</td>
</tr>
<tr>
<td>Small</td>
<td>5</td>
<td>P5/P11</td>
<td>24.0</td>
<td>Lysozyme, murein peptidase, acts in entry and lysis (McGraw et al. 1986).</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>P8</td>
<td>15.9</td>
<td>Comprises the membrane where the φ6 membrane is assembled (McGraw et al. 1986).</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>P9</td>
<td>9.5</td>
<td>Membrane protein also responsible for membrane formation (McGraw et al. 1986).</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>P12</td>
<td>20.3</td>
<td>Responsible for viral membrane formation (McGraw et al. 1986).</td>
</tr>
</tbody>
</table>

Table 3: Proteins of Bacteriophage φ6  
Derived from (Sarah Butcher; Dennis Bamford 2002)
Figure 4: Maps of the small, medium and large segments of φ6
Restriction maps of the three segments of bacteriophage φ6 (Mindich 1995)
Chapter 2:

Evidence for a Functional Role in the 3’ UTR of the Small Segment of Bacteriophage φ6

Abstract:

Bacteriophage φ6 serves as an important model organism for the study of viruses, for many reasons including well established working procedures, a well studied genome, a predictable life history, and the availability of tools for genetic manipulation, including a system of reverse genetics. However, significant gaps exist in our knowledge regarding life cycle as well as several uncharacterized and presumed untranslated regions (UTRs) in φ6's 13.3 kilobase-pair dsRNA genome. I examined the impact of specific modification to the 3’ UTR of the small segment of bacteriophage φ6. I found that modification to the purported UTR of the small segment resulted in severe fitness costs, supporting a functional role for as yet unidentified gene products.
2.1 Introduction

Bacteriophages serve as a model organism in evolution experiments for a number of reasons. Their smallness leads to a number of major advantages. They are easy to work with due to low materials cost. They may be filtered from host bacteria, providing a great advantage for such a small host-parasite system. Their short life cycle and large population sizes allow for laboratory evolution experiments which may demonstrate adaptation and emergence in a matter of days. Bacteriophage may be frozen for storage with a high recovery rate, are relatively stable in refrigeration, are easily handled compared to animal models, and can be genetically manipulated easily. φ6 has been used as a model for a number of evolutionary experiments (Dennehy 2012).

2.1.1 φ6 Tagging Project

A disadvantage of using bacteriophages for ecology and evolution experiments is that it is not trivial to analyze mixed populations, where bacteriophage strains are allowed to compete and have few phenotypic differences to distinguish them. Animal and plant individuals or genotypes can typically be tracked with phenotypic differences, genetic markers, or physical tags. Bacteriophage φ6 can accommodate genetic markers, but only at a cost to fitness and, in some cases, at risk of losing the marker downstream (Turner et al. 1999; Onodera et al. 1993). The present study grew out of an attempt to mark the small, medium and large segments with small nucleotide inserts which could be detected with PCR, rtPCR, or hybridization probes such as molecular beacons, in order to quantify the relative frequencies of different genotypes in competition.

The approach to marking bacteriophage φ6 proceeded as follows: plasmids containing the recombinant small, medium, and large segments of bacteriophage φ6 were obtained from the
Mindich Laboratory (Sun et al. 2004; J. Qiao et al. 2003). The phage genomic regions on these plasmids were independently modified to contain 20-34 base pair inserts in their respective 3’ untranslated regions. The modified segments were electroporated into *Pseudomonas phaseolicola*, in combination with wild-type segment plasmids, producing bacteriophage. This “rescue” system allows for genetic modification of bacteriophage φ6 and other cystoviruses.

In this case, the goal was to leave a genetic signature in nonessential regions of bacteriophage φ6, with little or no impact on the bacteriophage’s fitness. The results were mixed. Plaques produced after modifying the medium segment were slightly smaller than wild-type, and plaques produced after modifying the small segment were much smaller and more turbid (Figure 5). The genetic markers were successfully detected by PCR of cDNA from plaques, and sequencing results confirmed that the inserted nucleotides matched the intended sequences. Both small and medium segments had acquired tags, but in both cases there were significant fitness costs attributable to the markers.

![Figure 5: Plaques of wild-type, medium-segment tagged, and small-segment tagged φ6](image)

a) Plaques of wild-type φ6 (PT522). b) Plaques of φ6 modified with an insert in the 3’ UTR of the medium segment (JC1920). c) Plaques of φ6 modified with an insert in the 3’ UTR of the small segment (JC1922).

Although previous studies asserted that these regions were unessential and did not contain open reading frames (Mindich et al. 1992; Onodera et al. 1992; McGraw et al. 1986; Paul Gottlieb et al. 1988), the molecular biologist’s and the evolutionary biologist’s definitions
of essential may differ. The molecular biologist may deem anything not directly required for reproduction to be unessential, whereas evolutionary biologists hold that only those aspects that do not impact fitness are unessential. This finding prompted an investigation into the significance of the 3’ UTR, focusing on the 666 bp UTR of the small segment.

2.1.2 The 3’ UTR of the Small Segment of Bacteriophage φ6

All three segments of bacteriophage φ6 and of other members of the cystoviridae family are structured similarly. Plus RNA strands contain a 5’ untranslated packaging (pac) region which is structured and essential to packaging of the genome (Gottlieb et al. 1994), followed by densely packed genes, followed by a 3’ untranslated region of variable length whose terminal ~74bp are structured and essential for minus-strand synthesis (Mindich et al. 1994). In the case of the small segment, this ~74bp essential 3’ region is preceded by 592bp which neither code for any known gene nor have any known function in φ6.

Much is understood about the 5’ and 3’ structured ends of the small segment. The 5’ end with its 230 to 305bp packing-essential pac region is not identical from segment to segment (Qiao et al. 2005), whereas the 3’ ~74bp contains near sequence identity among all segments of φ6 (Mindich et al. 1994; Mindich 2004), even in some cases in common with other Cystoviridae (Qiao et al. 2000). In φ8, the 5’ pac regions may even overlap with genes (Qiao et al. 2005). It is proposed that the sequence similarity in the 3’ region has come about by recombination, and non-homologous recombination can be demonstrated when the 3’ region is damaged (Mindich 2004).

There is no apparent known functional role for the small segment 3’ 592bp region of interest for this study (Mindich et al. 1992; Onodera et al. 1992; McGraw et al. 1986; Paul Gottlieb et al. 1988). The extreme compactness and high mutation rates of the genome suggest
that such a large region would not exist without imparting some benefit to the organism. It’s possible that the sequence stabilizes upstream RNA elements protecting them from attack by cellular nucleases (Mindich 2004), but structurally functional RNA sequences can overlap with coding regions and, especially in consideration of the reduced significance of 3rd bases, stabilizing sequence complementarity does not require dedicated sequence space. Indeed, φ7, φ9, and φ10 demonstrate similar folding to φ6 in the 5’ region but maintain this folding through alternate base pairing (Mindich 1999). This provides an example in which significantly different sequences can provide the same function because base pairing is conserved. In the context of the fast mutagenic rate of RNA viruses (Duffy et al. 2008), the option of non-homologous recombination (Mindich 2004), and the compactness of the 3’ UTRs of the other two segments, the 592bp region has a suspicious staying power. The basis for this and the fitness effects observed could be secondary RNA structure, uncharacterized protein-coding regions, or both.

2.1.3 The Case for Functional Secondary Structure in the 3’ UTR

Although no previous study directly considered secondary structure in the 3’ UTR region of interest (ROI), several studies have demonstrated that modification of the 3’ UTR of the medium segment results in spontaneous non-homologous recombination and/or loss of inserts, indicating a stabilizing role for the 3’ UTR of that segment (Onodera et al. 1993; Mindich et al. 1992). These examples, taken with the aforementioned complexity of structure in the 5’ region and conservation of the 3’ ~74bp essential region, and the compactness of the genome, open up the possibility that secondary structure plays a role in the 3’ UTR of the small segment.

2.1.4 The Case for Genes

Fourteen protein gene products are documented in bacteriophage φ6 (Sinclair et al. 1975; Paul Gottlieb et al. 1988; Casini & Revel 1994).
The original publication of the sequence of the small segment described a computational search for open reading frames (ORFs), and only one ORF was found within the 3’ UTR, and possibly discredited due to its overlap with gene 5 (McGraw et al. 1986). Several things should be noted with respect to this negative result:

1. The PERCEPTRON algorithm used at the time was applied to the whole small segment and, although four genes were known to be functioning at the time, only two were detected algorithmically from sequence data. Not all genes have canonical Shine-Dalgarno sequences. Some genes exhibit a phenomenon known as polarity. Polar genes are those in which a missense mutation in a gene adversely affects expression in a downstream gene (Jacob & Monod 1961; Sinclair et al. 1976; Oppenheim & Yanofsky 1980). This effect is unexpected when the start codon and Shine-Dalgarno regions of the downstream gene are unaffected by the mutation. In particular, this phenomenon exists on the small segment, where the stop codon for gene 8 (UAA) overlaps by 1bp with the start codon for gene 12 (AUG), yielding the sequence UAAUG. Protein synthesis of P12 is approximately 1/10 of that of P8. This polarity, combined with gene overlap, suggests a departure from classical ribosome binding and functioning of the Shine-Dalgarno region (McGraw et al. 1986).

2. Consideration was only given to ORFs producing proteins at least 50 amino acids in length (McGraw et al. 1986).

3. Sequence error or discrepancy in this publication may have obscured a >50AA ORF from later discovery (Figure 6), by introducing a frameshift in the reference sequence.
a) small segment reference NC_003714

![Diagram](image)

b) small segment NC_003714-corrected

![Diagram](image)

c) Corrected sequence

\[ 2651 \quad 5' \quad UA\underline{U}CAAGCUCUCAUAGCUUGGAAGGAAAGGCCGCAUCAUG 3' \]

**Figure 6: Schematic of φ6 Small Segment and Uncharacterized ORFs**

a) Reference sequence for small segment of bacteriophage φ6, with detectable ORFs in the 3’ UTR region. Note four ORFS: ORF I2260, ORF I, ORF J1 and ORF J2. b) Sequence corrected by 1bp insertion found to be ubiquitous in sequenced samples. Note three ORFs, ORF I2260, ORF I, and the lengthened ORF J. c) Sequence surrounding the erroneous deletion. Inserted uracil at base pair 2653 (U2653) is indicated in blue.

A characterization of the distantly related bacteriophage φ8 revealed no ORFs in the corresponding UTR of Phi 8’s small segment, but a later φ8 study identified two open reading frames; they were named ORF I and ORF J, and have no sequence similarity to the corresponding regions in φ6 (Hoogstraten et al. 2000; Qiao et al. 2005). An ORF in the 3’ UTR
of bacteriophage φ13 is named P5b but was not visible on a protein gel of whole bacteriophage particles (Qiao et al. 2000).

Several ORFs are detectable by examination of the sequence of the small segment of bacteriophage φ6 (Figure 6b). A 333bp ORF starting at nucleotide 2260 has been previously reported (McGraw et al. 1986); in the original text it was described as starting at nucleotide 2290 but all other descriptions point to the ORF starting at nucleotide 2260. For this study it will be referred to as ORF I2260. ORF I2260 overlaps by 20bp with the 3’ end of P5, which does not preclude the possibility that it is expressed. Downstream from P5 and in the same reading frame as ORF I2260, at nucleotide 2308, is a 285bp ORF which will be referred to as ORF I, following the convention in previous Cystoviridae studies.

Another ORF occurs at nucleotide 2538 and contains 252bp, and will likewise be referred to as ORF J. ORF I and ORF J overlap by 55bp. Overlap in genes is not unusual in Cystoviridae; several gene pairs overlap by 1 base, and some more; for example, genes 12 and 9 on the small segment of bacteriophage φ12 overlap by 8bp. In what may be the most extreme prokaryotic example known, coliphage φX174 contains genes D and E which overlap by the entire length of E, 276bp, in different reading frames (Barrell et al. 1976), and examples have been found in eukaryotes as well (Makalowska et al. 2005). If expressed, the products of ORF I2260, ORF I, and ORF J would be 111, 94, and 83 amino acids (aa’s) in length.

Why there would be genes not yet discovered in a bacteriophage with so small a genome 40 years after the discovery? A number of studies had the potential to reveal genes if they did exist (Sinclair et al. 1975; Sinclair et al. 1976; Paul Gottlieb et al. 1988; Casini & Revel 1994). This situation is certainly not without precedent, however, and in φ6 not all known genes were discovered in the initial 1975 analysis by Sinclair et al. Protein P13 was discovered in 1988 and
protein P14 in 1994 (Paul Gottlieb et al. 1988; Casini & Revel 1994). P14 is a particularly
difficult target for two reasons: 1. it is small (62aa) and was found to migrate similarly to P9
(91aa) on a polyacrylamide gel; and 2. it acts only within the host and is not incorporated into the
\( \phi_6 \) particle, so it is not visible in gels of whole bacteriophage protein extract (Casini & Revel
1994). If ORF I and ORF J do indeed code for proteins, there has been no evidence for or
against their product(s)’ incorporation into the \( \phi_6 \) particle. The ORF I and J products, measuring
94 and 83 aa’s in length, would be easily obscured by the 90-aa P9. In fact a faint duplicity can
be seen in the P9 of various gels published from extracts including those of Vidaver et al. 1976,
as well as a faint band visible in the vicinity of P9 for a P9-deficient mutant (Figure 7) (Sinclair
et al. 1976).

The two ORFs I and J constitute the region of interest for this study. The region of
interest excludes the 74bp 3’-most nucleotides which are almost 100% sequence-identical
between strands and are known to be essential for minus-strand synthesis (Mindich et al. 1994).
Figure 7: Autoradiogram of Protein Gel of Bacteriophage φ6 Nonsense Mutants
 Autoradiogram of protein products of cells infected with mutants, after pulse of [14C]leucine. Lane (a): whole protein of uninfected bacteria. Lane (j): P9/P5 nonsense mutant, whole protein. The indicated P9 region is expected to be clear as in lane (a). Box marks the faint band colocalizing with P9 in adjacent lanes. This band may reflect a product of ORF I2260, ORF I, or ORF J. Modified from Sinclair et al. 1976.

2.1.5 The Curious Precedent of the Bullseye Plaque

A modified strain of bacteriophage φ6 was constructed with a genetic marker in the 3’ UTR of the medium segment (Figure 8, Table 4). The modification was a 34bp insert more than 100bp downstream from the 3’ end of the P13 coding region, and more than 70bp upstream from
the 74bp 3’ region. The modification procedure took advantage of a PstI restriction site centrally located in the UTR.

Figure 8: Medium segment of φ6 Tagged with Signature

a) Schematic of the initial tagging modification to the medium segment of φ6. Red arrow indicates the insertion location at the PstI restriction site within the 3’ UTR of the medium segment.

b) Sequence of modified untranslated region (portion). Modified region indicated with red bar.

Table 4: Sequence of Insert into φ6 Medium Segment

Numbers relative to φ6 medium segment start, restriction site overhangs in red, insert in blue.

Upon storage and re-plating of a sample of the phage, an interesting phenomenon became apparent. Plaques formed bullseyes, or alternating rings of turbid and clear regions in a plaque (Figure 9). The sample was not homogeneous, but contained a mix of bullseye and clear plaques; however, bullseye and clear plaques were each true-breeding. Samples of true-breeding clear and bullseye plaques of this medium-segment marked strain were tested for fitness vs. wild-type φ6, and unexpectedly, both strains demonstrated similar fitness to wild-type in a liquid-culture fitness assay (Figure 10).
Figure 9: Bullseye plaques
Early discovery of bullseye plaque phenomenon in bacteriophage φ6 medium-segment mutants. Arrows indicate bullseye plaques. All plaques show impaired morphology (reduced size vs. wild-type) but some exhibit clear plaques and others exhibit turbid phenotypes, and a small number exhibit bullseye patterns.

Figure 10: Fitness of Bullseye strain relative to Wild-type and Wild-type Rescue strain
C1A is a clear plaque mutant from Figure 9. B1C1 is a bullseye mutant from Figure 9. Pt522 is Wild-type bacteriophage φ6, and JC257 is bacteriophage φ6 constructed from the recombinant rescue system described in this chapter. S5C and L7C are mutants tagged on the small segment, demonstrating measurable loss of fitness with small-segment modifications other than P5. Series 1, 2 and 3 are fitness assay replicates.
The obvious phenotypic change intuitively implies that there was some fitness cost to whatever mutation occurred; however, fitness in liquid culture remained the same. This reflected the possibility that the overall long-term productivity of the phage could in some circumstance be unaltered, while a significant impact on life history was in effect. Additionally, sequencing of the medium and small segments revealed a mutation in the murein peptidase P5, involved in entry into and exit from the host; in this case, a frame shift mutation induced a premature STOP codon.

This change reflected an unexpected consequence while modifying the medium segment; the P5 mutation occurred on the small segment. It isn’t clear why this happened, however, I have recorded three such events. In all three cases, there was an insert to the medium segment, and bullseye formation indicated damage to the small segment P5 gene. Also in all three cases, samples showed no bullseyes initially, but the bullseye phenotype appeared after isolates had been stored at 4°C for several days and were subsequently plated. Each sample demonstrated a frequency of bullseye mutation between 5 and 80%. This curious behavior warrants further investigation.

Independently, Dessau, Goldhill, McBride, and Turner discovered bullseye plaques in φ6 in a heat-shock experiment (Dessau et al. 2012). Bullseyes in this case were also a result of an adaptive mutation in P5, however in this case they were not the result of a targeted mutation to the medium segment, and occurred to unmodified wild-type bacteriophage φ6. The Turner Lab exposed wild-type φ6 to heat-shock, and repeatably discovered that bullseye-inducing mutations in P5 were selected for by the heat-shock conditions.
Figure 11: Turner lab Bulseye Plaques
Heat-shock induced mutants of φ6, mixed with wild-type φ6. Bulls-eye patterns are the mutants; wild-type plaques are clear. (Dessau et al. 2012)
2.2 Experimental Methodology

The goal of this experiment was to understand the impact of modifying the untranslated region (UTR) of the small segment of bacteriophage φ6 and, in turn, shed light on the region’s relevance to life history. Two possibilities were tested; a) whether the secondary structure of this region was important, and b) whether previously uncharacterized ORFs are expressed in nature; i.e. whether there are protein-coding genes in the region of interest.

Several approaches to understanding the significance of the UTR were pursued. A library of mutants was constructed to ask specific questions of the system, and measurements were made in the form of fitness assays, plaque morphology observations, growth curves, and depletion of optical density (OD) of the bacterial host (used as a proxy for lysis of bacteria, although as will be explained, clearing of cell matter is not always consistent with fitness).

Mutants constructed fell into three overlapping categories, and the implications of possible outcomes are considered here:

1. **Major sequence modification.** These specifically included MUT4, MUT5, and MUT6. These mutations were intended to measure the impact of removing the majority of the 592bp region, in order to test for both possibilities – genes and secondary structure – without discriminating between the two. In addition, since different regions were ablated in different mutants, there is the potential to test for function or impact on fitness of different regions or ORFs. MUT4 deleted only a portion of ORF I; MUT5 deleted both ORF I and a large portion of ORF J, and MUT6 deleted a region spanning the two.

2. **Fusion of two ORFs.** These deliberately included JC1922 and MUT2, and incidentally included MUT6. The initially discovered JC1922 suffered a severe
fitness decrease vs. wild-type. MUT2 was constructed as a positive control, to replicate the scenario of a small modification between the two ORFs I and J, but contained only a single base pair insertion in order to minimize the possible impact on RNA secondary structure (Figure 18). If fitness was severely impacted by this mutation (MUT2), it was hypothesized that due to the fusion of two ORFs, a significant handicap was imposed on some protein function, either a product of ORF I, or of ORF J, or of both.

3. **Specific disruption and expression mutants.** MUT1 was constructed to modify ORF I by modifying its Shine-Dalgarno sequence (SD); the SD from P14 was inserted in the normal location for an SD ahead of the P14 gene. It was assumed because of the proximity to the functioning P5, that a small insert would not be relevant to RNA secondary structure effects; so, if modifying the SD changed fitness in this case, it would be presumed to have impacted expression of ORF I. MUT3 was constructed to insert a STOP codon early in ORF J, with little or no effect due to unaltered reading frame in the overlapping ORF I. If this modification resulted in altered fitness, this evidence would support the hypothesis that ORF J is a functioning gene.

All modifications, even the 1bp insertion of MUT2, have the potential to impact secondary structure in RNA, however, it is hypothesized that a major role of secondary structure in RNA is stabilization of ssRNA against attack from host nucleases (Mindich et al. 1994). Although it is possible that the RNA carries a catalytic or otherwise sequence-specific function, this is assumed to be less likely given the lack of sequence similarity among ORF I/J in φ6 and some related cystoviridae species, and the lack of ORF I/J entirely in others. For stabilization of
ssRNA strands, the majority of this function would be expected to remain intact with small modifications, but expected to be destroyed with large modifications. If small modifications alter fitness drastically, this supports the hypothesis that either one or both of ORF I and ORF J codes for a gene.

**Figure 12: Location of 1bp insertion for MUT2**

a) Sequence schematic of small segment, with location of 1bp cytosine insert at nucleotide 2573.  
b) Predicted secondary structure of the portion of the 3’ UTR where a single base-pair insertion was performed.  
c) Predicted secondary structure of the portion of the 3’ UTR where a single base-pair insertion was performed. Structures obtained via the MFOLD web server (Zuker 2003).
2.3 Materials and Methods

2.3.1 Bacterial and Viral Strains and Plasmids

The plant pathogen *Pseudomonas syringae* pathovar *phaseolicola* (ATCC # 21781) was used as a host for bacteriophage φ6, and will be referred to as PP (Vidaver et al. 1973). Bacteriophage φ6 strain PT522 provided by Paul Turner, Yale University, New Haven, CT, was used as a wild-type control (Vidaver et al. 1973).

A recombinant system for genetic manipulation of bacteriophage φ6 was developed by Mindich et al. as described in Sun et al. 2004, and consists primarily of a set of plasmids each containing a recombinant cDNA copy of an RNA segment of φ6 in the vector pT7T3 19U vector. These plasmids were provided by Paul Gottlieb, City College of New York CUNY, New York City, NY. These plasmids are designated pLM659, pLM656, and pLM687, and contain the small, medium, and large segments (S, M, L hereafter) of bacteriophage φ6, respectively (Gottlieb et al. 1992; Qiao et al. 1997; Sun et al. 2004).

The system is expressible in hosts expressing T7 RNA polymerase. *Pseudomonas* strain LM2691 is a derivative of PP containing plasmid pLM1086, which is a derivative of pRK290 and pAR1219 and expresses T7 RNA polymerase to facilitate expression of these recombinant φ6 segment plasmids, and was also provided by Paul Gottlieb (Davanloo et al. 1984; Ditta et al. 1980; Sun et al. 2004; Qiao et al. 2008).

Plasmids were stored and propagated in XL1-Blue chemically competent cells (Stratagene, Agilent catalog #200249) (endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F' [::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+) or XL10-Gold (Agilent catalog #200315) (endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB lacIqZΔM15 Tn10(TetR Amy CmR)]).
Plasmids stored in cells were frozen to -80°C by mixing equal volumes of 80% glycerol with cells cultured overnight in Luria-Bertani “lysogeny broth” (LB) medium (Sambrook & Russell 2001).

2.3.2 Culture Conditions

Working stocks of *Pseudomonas* host were prepared as follows: Frozen (-80°C) stock was streaked onto appropriate media and incubated at 25°C for 24-36 hours to obtain single colonies. PP was plated on LB, and LM2691 was plated on LB+tetracycline (15µg/mL).

Single colonies were placed in 10 mL of LB in 50mL covered Erlenmeyer flasks and incubated with shaking at 140 RPM overnight to produce stationary phase bacteria. LM2691 was cultured in LB+tetracycline (15µg/mL).

Colonies and working stocks for XL1-Blue and XL10-Gold containing all φ6 segment recombinant plasmids were prepared in the same manner as *Pseudomonas*, but incubated at 37°C with carbenicillin (CB) (50ug/mL).

2.3.3 Preparation of Plasmids

For any plasmid-carrying strain, 5 mL of overnight culture was centrifuged 7 minutes at 4400 RPM (3000 rcf) in an Eppendorf 5702 benchtop centrifuge, in a 14mL polypropylene conical tube. Pellet was resuspended in 250 µL of Qiagen Buffer P1; material was processed according to Qiagen Spin Miniprep Kit protocol (Qiagen catalog# 27106). Samples intended for downstream electroporation were eluted with 40 µL molecular-grade water.

2.3.4 Genetic Modification of RNA Segments of φ6

2.3.4.1 Mutagenesis Methods

Plasmids pLM659, pLM656, and pLM687 serve as templates for modification of wild-type bacteriophage φ6 segments S, M, and L (Gottlieb et al. 1992). Several molecular methods
were used to modify these segments in the production of mutants of S, M, and L. 1) Restriction followed by ligation to an oligomeric insert, 2) PCR to append restriction sites flanking an insertion/deletion region followed by restriction and self-ligation, and 3) the QuickChange II XL complementary mutagenic primer method (Agilent Catalog #200521, Agilent, Santa Clara, CA).

All custom primers were ordered from Life Technologies (Carlsbad, CA) or Invitrogen (Carlsbad, CA). The three methods are described individually; referring to Table 5: φ6 Small-Segment Mutation Classes for details not specified, such as primers and loci:

1. **Restrict/Ligate**: Oligo pair inserts indicated for a given mutant were annealed using a thermocycler program (Anon n.d.). The template plasmid was cut by restriction enzyme(s) listed. The sample was mixed in molar ratios 1:9, 1:3 or 1:1 with annealed complementary pair inserts (designed to provide suitable overhangs for the given restriction). The sample was then ligated using T4 DNA ligase (NEB part# M2200S) to produce the mutation indicated. This procedure was used to produce small inserts.

2. **PCR/Restrict/Ligate**: PCR was performed using the template and primer pair indicated. PCR resulted in the addition of the restriction site indicated, plus any additional bases included in the primer sequences. Samples were restricted using the enzyme indicated, and subsequently self-ligated. This procedure was used to produce small inserts and large deletions.

3. **Agilent QuickChange II XL**: Primers were designed using the QuickChange II XL primer design website to generate primers able to produce the indicated changes. These primers were used in a Quickchange reaction as specified in the Quickchange II XL protocol. This procedure was used to produce small inserts and modifications.
Products generated in 1, 2, or 3 were subsequently transformed into XL-10 Gold cells using the provided transformation protocol, and plated onto LB/CB plates to obtain single colonies. Single colonies were screened for the desired mutation using PCR, restriction, and sequencing, and plasmids were used in subsequent φ6 rescue procedures.

2.3.4.2 Mutant Classes and Segment Variants

Table 5 and Table 6 list names of the mutant classes of plasmids which were produced with the modifications indicated, which were later incorporated into bacteriophages by recombinant phage rescue (Sun et al. 2004). Figure 13 graphically illustrates the modifications performed to the small segment.

Figure 13: Schematic Representation of φ6 Small Segment Mutations
a) Schematic of wild type φ6 small segment, for reference.  b) JC1922; original small-segment mutation intended as an inert genetic marker, generating a fusion of two possible ORFs. c) MUT1; 18bp insertion to disrupt the Shine-Dalgarno sequence prior to ORF I.  d) MUT2; 1bp insertion designed to mimic JC1922 with negligible impact on RNA secondary structure. e) MUT3; premature STOP codon in ORF J, minor disruption to ORF I. f) MUT4; 128bp deletion in the region of ORF I.  g) MUT5; 349bp deletion in the region of ORF I.  h) MUT6; 164bp deletion between ORF I and ORF J, resulting in a fusion of two reading frames. Red stripes indicate deleted region in f, g, h.  Fusion of ORF I, J is indicated with orange in  b, d, h.
### Table 5: φ6 Small-Segment Mutation Classes, Methods, and Primers Used

The primer sequences, restriction enzymes, and the type of method used to create these mutants. Method 1) restriction, ligation 2) PCR, restriction, ligation 3) QuickChange II XL.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Primer 1 Sequence</th>
<th>Primer 2 Sequence</th>
<th>Restriction Enzyme</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC1922</td>
<td>TCCACCTCTGAATTTCGCCG GCCGCCCAGGCCTGCA</td>
<td>GCGGCCGCCCCGCGTAAATT CAGAGGTGGATGCA</td>
<td>NsiI</td>
<td>1</td>
</tr>
<tr>
<td>MUT1</td>
<td>AGATCTAAGCTTCATATGTTC CCAGATGTCACGAAAGG</td>
<td>CTGCAGCATATGAAGCTTTCC ATTTCAGTTTACGATGCAGTTCAGG</td>
<td>N/A (blunt ligate)</td>
<td>2</td>
</tr>
<tr>
<td>MUT2</td>
<td>CCGTTGTCCAGCCCCCTCCCAT GCTCTGAGATTG</td>
<td>CAAAATCTGAGATGCAATGGAA GGGGCTGGAACAGG</td>
<td>None</td>
<td>3</td>
</tr>
<tr>
<td>MUT3</td>
<td>TCCAGCCCCCTTCATGCAATGG CTTCTGAGATGGTACGTTAAG</td>
<td>CTCTACGAAAATCTGAAAGG CTATGCATGAGGGGCTGGG</td>
<td>None</td>
<td>3</td>
</tr>
<tr>
<td>MUT4</td>
<td>CTGCAGAAGCTTCAGTGTCAGG ATGCAGTTCAGGGG</td>
<td>AGATCTAAGCTTTAGCTATCG AGTITTTGATTAAAC</td>
<td>HinDIII</td>
<td>2</td>
</tr>
<tr>
<td>MUT5</td>
<td>CTGCAGAAGCTTCAGTGTCAGG ATGCAGTTCAGGGG</td>
<td>AGATCTAAGCTTTAGCTCTCT CATAGGCATTGGAAG</td>
<td>HinDIII</td>
<td>2</td>
</tr>
<tr>
<td>MUT6</td>
<td>CTGCAGAAGCTTCAGTGTCAGG ATGCAGTTCAGGGG</td>
<td>AGATCTAAGCTTTAGCTCTCT CATAGGCATTGGAAG</td>
<td>HinDIII</td>
<td>2</td>
</tr>
</tbody>
</table>

### Table 6: Summary of φ6 Small Segment Mutations

Locations on the small segment of modifications listed in Table 5 and Figure 13.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Type of Mutation</th>
<th>Start</th>
<th>End</th>
<th>Insertion/Deletion</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC1922</td>
<td>Small-tag mutant, fused ORF I and ORF J</td>
<td>2578</td>
<td>2578</td>
<td>Insertion</td>
<td>54 bp</td>
</tr>
<tr>
<td>MUT1</td>
<td>18 bp insertion at ORF I</td>
<td>2308</td>
<td>2308</td>
<td>Insertion</td>
<td>18 bp</td>
</tr>
<tr>
<td>MUT2</td>
<td>Fusion of ORF I and ORF J</td>
<td>2573</td>
<td>2573</td>
<td>Insertion</td>
<td>1 bp</td>
</tr>
<tr>
<td>MUT3</td>
<td>Stop codon on ORF J</td>
<td>2580</td>
<td>2580</td>
<td>Insertion</td>
<td>6 bp</td>
</tr>
<tr>
<td>MUT4</td>
<td>128 bp deletion in ORF I</td>
<td>2305</td>
<td>2433</td>
<td>Deletion</td>
<td>128 bp</td>
</tr>
<tr>
<td>MUT5</td>
<td>349 bp deletion in ORF I</td>
<td>2305</td>
<td>2654</td>
<td>Deletion</td>
<td>349 bp</td>
</tr>
<tr>
<td>MUT6</td>
<td>164 bp deletion in ORF I-ORF J</td>
<td>2490</td>
<td>2654</td>
<td>Deletion</td>
<td>164 bp</td>
</tr>
</tbody>
</table>
2.3.5 Recombinant Bacteriophage Rescue

After plasmids were modified and verified to have the desired mutation, electroporation was used to introduce the plasmids into Pseudomonas host and express viral RNA and proteins. This procedure is described in two parts; a) competent cell preparation, and b) electroporation and plating.

2.3.5.1 Competent Cell Preparation

LM2691 host, which is PP containing a plasmid expressing T7 transcription machinery, was cultured in quantity and washed to remove salts as follows, in preparation for electroporation:

1. A single colony of LM2691 was incubated with shaking approximately 48h in LB+tetracycline (15µg/mL) to produce stationary phase bacteria.

2. Exponential growth culture was started using 1 mL of stationary phase LM2691 into 200 mL LB+tetracycline (15µg/mL), in duplicate. This culture was grown 8-12 hours overnight. This typically yields optical density (OD) of 0.2 to 0.5.

3. In order to obtain OD 0.4 exponential phase LM2691, the exponential culture was subcultured into 200 mL x 2 LB+tetracycline (15µg/mL) to produce a calculated OD of approximately OD 0.2, and grown about 1 hour to reach OD 0.4. Cells were placed on ice and kept on ice or chilled by refrigeration through the whole preparation procedure.

4. OD 0.4 culture was pelleted by centrifugation in pre-chilled 50mL conical polypropylene tubes, by centrifugation for 15 minutes at 1000 rcf in an Eppendorf 5430R refrigerated centrifuge with refrigeration at 4°C.
5. Supernatant was poured off and cells were resuspended in chilled (0°C-4°C) sterile water, 50% to 100% of original volume.

6. Cells were pelleted again by centrifugation for 20 minutes at 1000 rcf.

7. Supernatant was carefully poured off and cells were resuspended in chilled sterile 10% glycerol, 25-50% of original volume.

8. Cells were pelleted again by centrifugation for 20 minutes at 1000 rcf.

9. Supernatant was carefully poured off and resuspended with 3 mL of 10% glycerol per 50 mL original volume, and combined into pre-chilled 14mL conical polypropylene tubes.

10. Cells were pelleted again by centrifugation for 20 minutes at 1000 rcf.

11. Supernatant was carefully removed and cells were resuspended in 200 µL 10% glycerol per 400 mL original volume.

12. Optical density was taken of a 1/100 dilution of this preparation, and volume was adjusted with 10% glycerol to obtain an OD of between 0.8 and 1.2 of the 1/100 dilution.

13. Electrocompetent cells were divided into 42 µL aliquots in pre-chilled Eppendorf 1.5mL tubes and used in electroporation experiments immediately or frozen by immersion in liquid nitrogen and stored at -80°C.

This protocol is derived from Sambrook & Russell 2001 protocol for preparation of competent E. Coli cells (Sambrook & Russell 2001).

2.3.5.2 Electroporation and Plating

The Mindich rescue system was used to allow production of bacteriophage from three recombinant plasmids each containing a segment of bacteriophage φ6, one or more with modifications. For each intended mutant, one µg of each of the plasmids was added to 42 µl of competent cells previously prepared, in a 1.5mL microcentrifuge tube. The mixture was left on
ice for 1 minute. The mixture of cells and plasmids was transferred to a cold 0.2 cm electroporation cuvette (Bio-Rad, Hercules, CA, USA) and electroporated in a Bio-Rad Gene Pulser at 2500V. The cuvette was removed from the chamber and 1 mL of SOC (super optimal broth with glucose) medium was immediately added for recovery. Cells were resuspended and transferred to a 1.5mL microcentrifuge tube and allowed to recover for 60 minutes at 25°C with rapid shaking (250 RPM). Dilutions were plated using the previously described plating protocol for bacteriophage, onto LB medium using permissive host (PP), and incubated from 16 to 48 hours at 25°C, to observe plaques.

2.3.6 Plaque Isolation and Purification

Plaques were picked using a pipette tip to minimize variation in sample volume; a 1000μL long pipette tip was used as illustrated in Figure 14. The tip was attached to a pipetting tool. For each plaque sampled, the tip was pressed into top agar but only slightly beyond the top agar into the bottom agar (about 1mm). The tip was moved sideways to capture the plug of top agar, and the pipette and tool were used to mix the plug into 100 μL of LB media in a 200 μL PCR tube, by pipetting several times. The sample was allowed to sit at room temperature for 30 minutes to ensure phage diffusion from the agar plug.
Subsequent to plaque picking, samples were serially diluted and plated for plaque isolation and purification. All mutants were again picked from an isolated plaque, for an additional round of purification, and dilutions were plated to determine plaque-forming units per mL (PFU/mL); PFU/mL was then used to calculate the appropriate dilution for a lacy lawn and a stock was produced by plating and incubating overnight at 25°C. The plate was scraped using a modified spatula, to collect the top agar (leaving the bottom agar). The top agar was resuspended in 4 mL of LB in a 14mL Falcon polypropylene tube and centrifuged at 300 rpm for 10 minutes. The supernatant was filtered through a Millex 0.22 µm syringe filter. A sample of filtrate was mixed 1:1 with 75% glycerol, and stored at -20°C for later use. In some cases, a working stock was stored at 4°C without glycerol.

2.3.7 RNA Extraction

A lysate was prepared by plating a lacy lawn on LB agar plates. The plate was scraped to isolate top agar; top agar was added to 4 mL of LB in a 14mL Falcon polypropylene tube and centrifuged at 3000 rpm for 10 minutes. The supernatant was filtered using a Millex 0.22 µm syringe filter.
syringe filter. 3 mL of lysate was placed in a polycarbonate ultracentrifuge tube and centrifuged at 50,000 rpm for 3 hours. Supernatant was poured off and the pellet was resuspended in 150 µL of distilled water. RNA was extracted using the Qiagen Viral RNA Mini Kit (catalogue number 52904, lot number 145026316) according to the protocol in the Qiagen Viral RNA Mini Handbook.

2.3.8 **cDNA Synthesis**

Complementary DNA (cDNA) was synthesized using SuperScript™ II RT, according to the protocol provided by the manufacturer.

2.3.9 **PCR Detection of Inserts**

Standard TAQ and/or high-fidelity polymerase was used for detection of inserts and, in some cases, production of samples for sequencing. A standard PCR protocol was used according to the instructions provided by the manufacturer, with annealing temperatures adjusted to the primers used and extension times adjusted for the maximum length product.

2.3.10 **Sequencing**

Sequencing was performed at the DNA Analysis Facility at Yale University. Samples were prepared either from plasmid preparations or from PCR product. Plasmid preparations were prepared using Qiagen Miniprep kit according to manufacturer’s instructions. PCR product was cleaned using EXOSAP-IT.

2.3.11 **Fitness Assay**

A fitness assay in liquid culture was conducted as follows: PFU/mL for each sample was determined by dilution plating, and 1000 of each phage was cultured with 100 µL of exponential phase PP and incubated 18 hours at 25°C with shaking, for each strain tested (Figure 15). Three replicates were performed for each strain. Two controls were used for the first fitness assay of
the small tagged mutant; one generated through rescue, the other environmental wild type PT522. For further assays, PT522 was used as a control. Samples were filtered and diluted serially, then plated to determine output of each assay.

**Figure 15: Fitness Assay in Liquid Culture (Schematic)**
For each sample tested, 1000 bacteriophage are combined with bacteria in excess in 10ml of LB media and shaken overnight for 18 hours. The product is filtered to isolate resulting bacteriophage. The original sample and the assay filtrate are serially diluted and plated on a lawn of compatible bacterial host to obtain PFU/ml counts; the resulting counts are used in calculation of fitness.

### 2.3.12 OD Depletion Assay

In order to determine the extent to which each mutant was able to clear a culture of bacteria, an inoculation and overnight growth was performed. A fitness assay was conducted. PFU/mL for each sample was determined by dilution plating, and 1000 of each phage was
cultured with 100 µL of exponential phase PP and incubated 18 hours at 25°C with shaking, for each strain being tested. Three replicates were performed for each strain. OD600 of samples was measured at 18 hours to determine the depletion of OD relative to bacteria-only control.

2.3.13 Growth Curve of JC1917

100 µL of exponential-phase PP in 10 mL of LB were inoculated into a 50mL flask along with 100 and 1000 PFU of small-segment tagged strain JC1917, which is a variant of MUT6 previously mentioned. Negative control was PP with no bacteriophage. The test was repeated three times and growth curves were recorded.

2.3.14 RNA Secondary Structure Prediction

Sequences of all segments as well as the 3’ UTRs of the segments were submitted to the MFOLD website (Zuker 2003). Images were recorded for traditional structure representations and circle graphs of submitted mutants.

2.3.15 Protein 3D Structure Prediction

Sequences for the two hypothetical open reading frames represented by ORF I and ORF J were submitted to the I-TASSER web application (Roy et al. 2010; Zhang 2008).
2.4 Results

2.4.1 Fitness Impact of Modification to the 3’ Region of φ6 Small Segment

2.4.2 Fitness of Small Tag Mutant

I had initially constructed a pilot study measuring fitness of constructed, marked strains of bacteriophage φ6. Modifications to small and medium segments produced a difference in phenotype and plaque morphology. A series of 3 fitness assays of the small tag mutant was conducted to verify this finding. In this study, fitness ($W$) is defined as $W = N_T / N_0$ where $N_T$ is the final population and $N_0$ is the initial population (Orr 2009), measured by plaque assay.

The assay showed that the small tag mutant was approximately 4 orders of magnitude less fit compared to wild type as seen in Figure 10. This prompted an investigation as to the significance of the UTR in the small segment.

2.4.3 Identification of ORFs

Potential open reading frames were found using the NCBI ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). ORFs >25 codons in length were considered for this study (Figure 6).

2.4.4 Prediction of Secondary Structure; Impact of Tagging on Secondary Structure

The secondary structures were found using the Mfold program (Zuker 2003). Secondary structures are presented in Figure 16 and Figure 17.

The initial small-tag insert may disrupt some probable secondary structure, pictured in Figure 18.
Figure 16: Secondary structure of φ6 Small Segment
Structures were obtained by submitting sequence to the Mfold web server (Zuker 2003).

Figure 17: Secondary structure of φ6 Small Segment 3’ UTR
Structures were obtained by submitting sequence to the Mfold web server (Zuker 2003).
Figure 18: Impact of Small segment Insert on Secondary Structure
Structures were obtained by submitting sequence to the MFOLD web server (Zuker 2003). Numbers are relative to the 3’ UTR. Tag is 294-331 in the UTR. Insertion site NsiI is at 294-297 in φ6 wt 3’ UTR.

Protein 3D Structure Prediction

Figure 19: Predicted structures of ORF I, ORF J
2.4.5 Construction of New UTR Mutants

The series of small-segment mutants described had been constructed to investigate the possibility that the difference in fitness was a result either of protein coding genes in the detected ORFs or an impact of secondary structure. The initial JC1922 clearly had a slight impact on secondary structure (Figure 18); therefore, a more elaborate test was required. The methodology of MUT1 thru MUT6 is described in the “Methodology” section above; to summarize: MUT1, MUT2, and MUT3 contain small modifications to test specific hypotheses, and mainly preserve secondary structure, whereas MUT4, MUT5, and MUT6 obliterate large regions of ORF I and ORF J to test more crudely for fitness changes that could result either from protein-coding gene disruption or from disruption of functional or stabilizing secondary structure.

MUT1 modifies the initiation region of the first ORF in the untranslated region; MUT2 brings the majority of ORF J into reading frame with ORF I creating a fusion; MUT3 inserts an early stop in ORF J. Sequences were verified at the plasmid level as well as the bacteriophage level after rescue and isolation.

2.4.6 DNA Synthesis and Sequence Verification

cDNA derived from the RNA of the purified φ6 mutant strains was sequenced at the DNA Analysis Facility at Yale University. Mutant sequences were aligned to wild-type φ6 using various tools including SeaView with MUSCLE and Clustal Omega sequence alignment algorithms, and the proprietary algorithms in CLC Bio Workbench 5.1 were used to confirm intended mutations.

2.4.7 Fitness of Newly constructed Strains MUT1 through MUT6

A series of fitness assays was carried out to test small-segment mutants of φ6. The results indicate that MUT 1 and MUT3 were nearly as fit as wild type. MUT2, MUT4, MUT5, MUT6
measured orders of magnitude less fit than the wild type, as seen in Figure 20. Plaques from each strain are presented in Figure 21.

Figure 20: Fitness of Mutants of Bacteriophage φ6 Small Segment 3’ Region
JC1922: Original tagged small-segment resulting in fusion of ORF I, J. MUT1: 18bp insertion mutant disrupting the Shine-Dalgarno sequence prior to ORF I. MUT2: 1bp insertion designed to mimic JC1922 with negligible impact on RNA secondary structure. MUT3: Premature STOP codon in ORF J, minor disruption to ORF I. MUT4: 128bp deletion in the region of ORF I. MUT5: 349bp deletion in the region of ORF I. MUT6: 164bp deletion between ORF I and ORF J, resulting in a fusion of two reading frames. Standard Error of the Mean is indicated.
Samples of all mutants were grown simultaneously under identical conditions. WT1A and WT1B are identical treatments of the same control strain.

2.4.8 Relationship between Fitness and Optical Density

Bacteriophage biologists routinely look at a cleared culture to determine that bacteriophage have been effective, and the depletion in OD is taken as a measure of fitness of the bacteriophage in a sample. In order to test the relationship between optical density and fitness, the results from the fitness assay and the OD assay were compared (Figure 22). In order to relate OD to fitness, OD was converted to “OD Depletion” by subtracting OD from the OD of a pure bacterial culture grown for the same length of time, i.e. untreated bacterial control. The logarithm was taken for graphs and analysis.
This allowed for graphing of fitness vs. OD depletion, and previous studies capitalize on the principle that the two properties are related (Turner et al. 2012). Although there was an obvious linear relationship between the two, there were three noteworthy exceptions. JC1909 and JC1922 exhibited lower OD, and thus higher depletion, than would be expected given their low fitness (see also Figure 45). JC1917 exhibited much higher OD (and lower depletion) than its fitness would predict. These opposite cases are intriguing; the former examples are able to clear a lysate but output fewer bacteriophage than would be expected for a given ability to kill bacteria. This may suggest they are able to kill bacteria but are rendered ineffective or fail to function in the assembly pathway. The latter example, JC1917, ultimately exhibited higher OD during some tests than the untreated bacteria, suggesting that it enhanced bacterial growth, and this prompted an investigation into its behavior.
2.4.9 Growth Curve of JC1917

100 μL of exponential-phase bacteria in 10 mL of LB were inoculated into a 50mL flask along with 100 and 1000 PFU of small-segment tagged strain JC1917, which is a variant of MUT6 previously mentioned. Observed OD after a fitness assay was higher than bacterial control, so the test was repeated three times and a growth curve was recorded.
Figure 23: Growth Curve of JC1917 vs. Control
Dependent axis is OD600. Independent axis is time (Hours). “1000 phage” (magenta) represents inoculation with 1000 PFU of JC1917.
2.5 Discussion

This study examined the impact of modifying a 592bp region near the 3’ end of the small segment of φ6. Previous studies had not found justification to characterize any open reading frames in this region as genes. This study provides evidence for the impact of these ORFs on fitness and in turn their likelihood of coding for polypeptides active in the φ6 life cycle.

The genome of bacteriophage φ6 is composed of 3 segments: small, medium and large. Genes appear to be organized according to function, with structural/procapsid genes on the large segment, and membrane/envelope proteins on the medium and small segments (Mindich 1999). Each gene of the genome encodes for a specific protein; genes are packed tightly to the extent that some genes overlap; raising the question of why a region as large as ORF I and ORF J would lack functioning protein-coding genes. However, past studies have overlooked genes and open reading frames that are now known to code for proteins. The first comprehensive study of the proteins in φ6 was done in 1975 by Sinclair et al. and found 12 proteins (Sinclair et al. 1975). Over the next 20 years, two more protein-coding genes were found; one, P13, was found by Gottlieb et al. in 1988, and the other, P14, was found by Casini and Revel in 1994 (Paul Gottlieb et al. 1988; Casini & Revel 1994). In both cases, a gene was found in a region of RNA that was previously believed to lack expressed proteins. The current study identified and modified ORFs in the 3’ end of the small segment which also have no previous indication of function in the record.

Although most of the RNA in bacteriophage φ6 codes for proteins, noncoding regions may function through their secondary structure, either with specific roles in life history, or through general stabilization of ssRNA against degradation. An example of a specific function can be seen in pac sequences previously mentioned. The preservation of these pac sequences and
secondary structures in related strains suggests that these conserved regions on the three segments occurred through recombination. Recombination allows for recovery of functional regions in ssRNA and protects φ6 in situations where the 3’ segments are damaged (Mindich 2004).

Several studies demonstrate molecular modification of φ6 (Onodera et al. 1998a; Onodera et al. 1993; Sun et al. 2004). In an attempt to incorporate a nucleotide insert for tracking in evolutionary experiments in the uncharted 3’ end of the small segment, an unexpected effect on plaque size and fitness was discovered. The present study investigated the cause of this impact. Two hypotheses were tested: 1. whether the 3’ end of the small segment carries protein-coding genes, and 2. whether secondary structure functions in the 3’ region. The results support both of the hypotheses, which are not mutually exclusive, and heavily favor the former.

2.5.1 Comparisons

There are several comparisons to be made among the small-segment mutant results.

2.5.2 MUT2 vs. MUT5 Comparison Reflects Altered ssRNA Stability

In support of the role of secondary structure in the 592bp region of interest, a comparison of plaque morphology can be considered. MUT2, a single base pair insertion, would be expected to have only a small impact on secondary structure (see Figure 12). MUT5, in contrast, suffered a large deletion in the region, and should be expected to have altered secondary structure. Both mutants suffered by orders of magnitude in fitness (Figure 20), however, plaque morphology tells an interesting story. MUT2 shows no visible variation in morphology, whereas MUT5 has significant variability in plaque morphology (Figure 21). Recombination is known to be triggered by deletions in the 3’ end of segments of φ6 (Mindich 2004). Two possible
explanations for the plaque size variations in MUT5 are recombination events and point mutations enhancing complementarity in ssRNA secondary structure, both of which would impact stabilization of the small segment against degradation and cellular nuclease attack.

MUT2 shows no such variation; this supports the hypothesis that the mutant, with only a minor 1bp insertion, is stable and undergoes little or no spontaneous compensatory mutation, despite its extreme loss of fitness.

2.5.3 MUT1 vs. MUT3: ORF J Impacts Plaque Morphology but not Fitness

MUT1 and MUT3 were almost as fit as wild type in liquid culture assay (Figure 20). MUT3, however, exhibits a greater change in plaque morphology as compared to MUT1; MUT3 plaques were smaller and more turbid (Figure 21). MUT1 also exhibited inconsistent plaques, some plaques looking nearly as large as wild type, with others looking as small as MUT3’s.

MUT1 and MUT3 inserts should not induce a frameshift, being multiples of 3bp in both cases. If ORF I and ORF J code for proteins, it is possible that the function of ORF I was not largely affected in MUT1, or that they are compensated for by point mutations which arise during plaque formation. In contrast, MUT3 inserts a STOP codon early in ORF J, and MUT3 plaques are consistent and greatly reduced in size. This suggests that ORF J impacts an aspect of life history affecting plaque size. The counterexample of MUT1 is a modification with a similar size insert in a different location, but radically different impact on plaque morphology. Likely explanations for MUT1’s behavior include a) impact on secondary structure, and b) compensatory mutations in the 18bp insert region upstream of ORF I which alter either a Shine-Dalgarno region of ORF I, or secondary structure, or both.

2.5.4 MUT2 and MUT3 demonstrate Different Impacts on Fitness, Supporting Function of One or Two Genes
MUT2 contains a 1bp insertion resulting in the fusion of ORF I and ORF J. MUT 3 contains a STOP codon early in ORF J. If ORF I is not expressed as a protein, it would be expected that fusion with MUT2 would result in the inhibition of ORF J, and the impact should be similar to MUT3. However, MUT2 exhibits the most severe fitness impact of all mutants studied. The most parsimonious explanation for this is that both ORF I and ORF J code for proteins, that MUT2 results in a fusion of the two proteins, and that this fusion protein is expressed at ORF I’s natural expression rate.

The effect of such a fusion on the natural function of two adjoined protein domains should be carefully considered. Routine cell biology methods include adding the 25kDa GFP tag to a protein of interest, allowing for localization of the protein, without apparent inhibition of function (Chudakov et al. 2010; Kremers et al. 2011). But this is not always the case, and GFP tags also have been found to alter protein localization and function (Skube et al. 2010; Hanson & Ziegler 2004; Palmer & Freeman 2004; Arai et al. 2001), and the length of the linker in a protein-marker fusion, for the purpose of reducing steric hindrance caused by a marker such as GFP or luciferase, has been a recurring subject of investigation (Maeda et al. 1997; Arai et al. 2001; Silacci et al. 2014). There may be a difference in the impact of such a fusion depending on a) the length of the linker, and b) the mode of function of the proteins, specifically, whether they are structural or catalytic in nature. Steric conflict between two fused proteins can alter localization, and it can also alter enzymatic function. If the two domains of a fusion protein naturally localize by association with different binding partners from each other, the resulting fusion has the potential to make matters worse than if the two nonessential proteins were not manufactured at all.
The small-segment genes tend to bear lipid-associated functions and localize in the lipid coat of $\varphi 6$. Several proteins of $\varphi 6$ function as multimers in close proximity to one another (P3 believed to be a trimer, P6 a hexamer, P4 a hexamer, P1 intricately interacting 120 copies in the structural shell). In contrast with a bacterial or eukaryotic genome with an abundance of enzymatic genes, the $\varphi 6$ genome consists largely of structural genes which integrate into the phage particle in close proximity with intricately ordered binding partners.

JC1922 exhibits a fitness reduction to a lesser extent than MUT2 (Figure 20). In the case of JC1922, the linker between the two ORFs I and J is longer (a 34bp insertion vs. the 1bp insertion of MUT2, between two ORFS which naturally overlap by 55bp), potentially liberating the two domains to function with less steric interference from one another.

These considerations easily account for the great disparity in the fitness of MUT2 vs. MUT3, and also account for the preservation of MUT3 fitness despite an obvious impact on plaque morphology.

2.5.5 Deletions vs. Insertions: Both Genes and Secondary Structure Impact Fitness

Comparing the fitness of the deletion mutants (MUT4, MUT5 and MUT6) to that of the insertion mutants (JC1922, MUT1, MUT2, and MUT3) supports the idea that both genes and secondary structure are at play. The deletion mutants are less fit than most of the insertions (with the exception of MUT2), which supports the idea that secondary structure is a factor, if not the only factor. Furthermore, the recovery of plaque sizes in the deletion mutants MUT4, MUT5 and MUT6 are most easily explained by restoration of secondary structure through recombination or compensatory base changes, since it’s unfathomable to reconstruct deleted genetic material. If ORF I and ORF J code for proteins, they are clearly nonessential, so deletion mutants MUT4, MUT5 and MUT6 need not impact fitness as drastically as MUT2, which may disrupt function
more negatively than these deletions. Looking only at MUT4, MUT5 and MUT6, the largest impact on fitness is seen in MUT5, which is the largest deletion of the three. These observations support the hypothesis that the 3’ 592bp region’s secondary structure has an impact on fitness. MUT5 and MUT6 nearly obliterate both ORFs, and MUT4 obliterates a large part of ORF I, so it is not expected that these mutations retain function in these ORFs; however, the larger the deletion, the larger the fitness cost. This does not refute the possibility that ORFs I and J code for proteins, but in the context of MUT1 and MUT3 which each have small insertions and low impact on fitness, it supports the idea that secondary structure is a factor in this region.

2.5.6 Observation: Fitness and Plaque Size are Not Universally Correlated

The experiment demonstrates that fitness and phenotype are not universally correlated. Some mutants (MUT1 and MUT3) exhibited smaller plaque size, but measure similar to wild-type in a liquid culture fitness assay; furthermore, they differ greatly in impact on plaque size (Figure 20, Figure 21). MUT4 and MUT6 show only subtle differences in fitness, but also exhibit stark differences in plaque morphology.

2.5.7 Gene Discovery

The study applies a simple means of detecting the function and purpose of an unknown region of RNA in a virus, which can be applied to other viral studies. By targeting specific regions of RNA and observing the effects on plaque morphology and fitness, evidence was provided to suggest the existence of genes and secondary structure in a previously unmapped region. More study is needed to provide a more comprehensive understanding of the roles this region plays in the fitness and plaque morphology of bacteriophage φ6.

2.5.8 JC1917 Appears to Enhance Bacterial Growth
Strain JC1917 provides a special example illustrating that depletion of optical density is not universally related to fitness. In a JC1917 growth assay, phage are productive, but bacterial density appears higher than untreated bacteria (no bacteriophage) under otherwise identical conditions. This wholly unexpected result indicates any of several possibilities: 1. Bacterial debris are being formed which do not inhibit cell growth; 2. Productivity of the host is increased by infection with JC1917; 3. Quorum sensing is inhibited by JC1917; 4. Absorption of light by bacteria is altered by JC1917 infection, including the possibility that bacteria are forming aggregates; 5. Bacteriophages are forming aggregates which impact absorption. This is an intriguing problem and it demands further investigation.

2.6 Conclusion

Data suggest that alterations in the 3’ UTR of the φ6 small segment may impact plaque formation through several means, including changing rates of diffusion, formation of multiphage aggregates, and changing phage affinities and attachment rates for environment and/or bacteria. These factors impact fitness, and plaque morphology can be used as an indicator for a change in life cycle parameters. Additionally, unresolved questions in bacteriophage φ6’s life history are potentially answered by the function of proteins proposed by this study. Further work is needed to isolate these proteins or exhaustively disprove their existence and explain the dramatic fitness effects demonstrated here. The unusual impact of the fusion mutation represents a convenient scheme for diagnosing the function of, or zeroing in on difficult-to-find genes; the mutants constructed here can be used in further studies, including complementation studies, to determine whether and where products of ORF I and ORF J function in the cell, in the virion, and in the life cycle of bacteriophage φ6.
Chapter 3:

Transient Occupancy of P4 in the Portals of the φ6 Procapsid

3.1 Abstract

Much is understood about the life history of φ6 but numerous specific details are still unknown. P4 is essential for nucleation of the 120 copies of the procapsid core protein P1, and 12 P4 hexamers are included in the assembled nucleocapsid. Only one P4 hexamer functions to package three consecutive ssRNA strands into the expanding procapsid. What mechanism prevents distal P4 hexamers from concurrently packaging additional ssRNA is not known. Here I present evidence for low occupancy of P4 on the unexpanded procapsid, for the role of P8 in stabilizing P4 on the expanded nucleocapsid, and for the tenuous connection between P4 and P1. A hypothetical atomic model of P4 is presented. Taken together, these observations support a model for φ6 packaging and assembly which describes the association between P4 and P1 as tenuous, stabilized only after the arrival of P8 on the nucleocapsid. This model explains inhibition of simultaneous multiple strand packaging as a stochastic process governed by the kinetics of P1-P4 interaction.

3.2 Introduction

3.2.1 The φ6 RNA Packaging Motor

Bacteriophage φ6 ssRNA is packaged into assembled procapsids by the P4 packaging motor, which is a hexameric NTP-ase. The hexamer occupies a 5-fold vertex of the excavated dodecahedron formed by the P1 shell. This represents a mismatch in symmetry (de Haas et al. 1999). The anchoring of P4 to P1 is not well understood, although the C-terminus of P4 is known to be essential to assembly of P4 to P1 (Paatero et al. 1998). The disordered nature of
this C-terminal region in crystal structures (Mancini et al. 2004) suggests that it is not bound tightly to the P4 hexamer.

3.2.2 P4 Occupancy and Phage Particle Function

Procapsid assembly begins with production of the four core proteins P1, P2, P4, and P7. P1 is the major constituent of the procapsid, occurring in 120 copies. The RNA-dependent RNA polymerase P2 and the assembly cofactor P7, as well as protein P4, the hexameric ATP-ase responsible for packaging, are all cofactors in assembly. P4 resides at the fivefold vertex (Figure 24) of the excavated-dodecahedral core, and there are twelve such fivefold vertices, which will be referred to as “portals” for their role in ssRNA translocation. The hexameric P4 is known to be required at almost every fivefold vertex during assembly (Sun et al. 2012). However, extensive study has determined that the three genomic segments pack sequentially, indicating that only one portal is active during packaging (Gottlieb et al. 1992; Onodera et al. 1995; Qiao, Casini, et al. 1995; Mindich et al. 1995; Qiao, Qiao, et al. 1995; Onodera et al. 1998b; Onodera et al. 1998a; Mindich 1999; X. Qiao et al. 2003; Mindich 2004; Huiskonen et al. 2006).
Several explanations are plausible for how a single portal is active while eleven others are not:

I. Intramolecular communication between the active portal and all other portals results in inhibition of all other portals. In this model, electrostatic interactions, structural changes, and/or cofactors are activated when P4 is present and/or functioning at a given portal. When one P4 is active, all others are either released or rendered inactive.

II. It is a stochastic process, and the kinetics of P4 attachment to the P1 core make concurrent portal activation unlikely. In this model, it is possible but extremely rare for two of the same segment to package.

III. The procapsid does not have true icosahedral symmetry, and some portals are different from others.
a. It has been reported that only 8 copies of the P2 polymerase are present in the procapsid; also, P4 has been observed to bind asymmetrically in the portal (Nemecek et al. 2010).

b. If the procapsid is not truly symmetrical, and exhibits differences in polar vs. equatorial regions, structural differences between portals may lead to functional differences.

These explanations are for the most part not mutually exclusive. For instance, it may occur that there is a polarity to the procapsid (III) induced by intramolecular communication (I), or that stochastic attachment (II) results in partial inhibition of other portals (I) and that this inhibition may work though asymmetry (III). It may be that asymmetry governs differences in binding affinity at the different portals.

This study focuses on explanation II (kinetically controlled activity), and offers evidence for III (asymmetry of the procapsid). The frequency at which a P4 hexamer occupies a φ6 procapsid, nucleocapsid, or virus particle, will be referred to as “occupancy”; it may be expressed as a percentage or as a number of portals per particle containing 12 portals.

3.2.3 Cryoelectron Microscopy

Cryoelectron microscopy reduces deterioration of a sample, allowing higher resolution and longer exposure times than a sample examined at room temperature (Guerrero-Ferreira & Wright 2013).

3.2.4 Single Particle Analysis

In single particle analysis, electron micrographs of multiple replicates of identical samples are processed so as to take an average of the densities of the samples. This allows for higher resolution and reduction of noise, and the construction of a three-dimensional (3D) high
resolution density map from two-dimensional micrographs, and reconstructions may reflect a near-atomic resolution of the subject, approaching the quality of X-ray crystallographic methods (Zhang et al. 2008). The resulting density map may be then used to construct a 3D surface map, called an isosurface, reflecting the consensus densities of a single particle representing the collection. Computed density maps generally do not have sharp “edges” or “sides”; density of an object is generally a gradient. An isosurface is used to represent the surface at which density makes the transition from one density value to another. This is analogous to contour lines on a geological map; the lines do not represent any physical structure in nature, but rather a path along which altitude remains the same; this contour can be used to interpret the shape of a geological feature. In the same way, an isosurface may be used to interpret the shape of three-dimensional objects without sharp boundaries.

Figure 25 illustrates a single-particle reconstruction. Figure 26 illustrates an isosurface view of the same reconstruction.
Figure 25: Cryo-electron microscopy cross-section of the φ6 Nucleoapsid
Distinct layers of the nucleocapsid are shown in a reconstruction of the nucleocapsid. A) Protein layer P8, which mates P1 with the lipid envelope. (Note that there is no lipid envelope in the nucleocapsid.) B) P4, the RNA-translocating motor. C) P1, the major component of the icosahedral lattice forming the major structure of φ6. D) The internal space containing ssRNA. Note that order reflected in this region is not necessarily defined due to imposed icosahedral symmetry.
3.2.5  Tomographic Reconstruction and Subtomogram Averaging

To study the procapsid of φ6, cryoelectron microscopy micrographs of the procapsid were used to reconstruct a three-dimensional model. In tomographic reconstruction, generally cryotomography, cryoelectron microscopy is performed on a sample and the sample is rotated in the electron microscope chamber for multiple exposures at different angles. Figure 27 depicts two images in a tilt series used for tomographic reconstruction.
Figure 27: Cryoelectron tomography tilt series
Two views of a tilt series are presented; left, the sample significantly tilted; right, the sample more level. Note the difference in contrast.

Figure 28: Closeup of Tomogram, φ6 Procapsids with Occupied Portals

Computational methods allow for the construction of a three-dimensional image, a density map, reflecting the densities of the subject at all points in the sample area. The use of multiple angles provides information about depth of features and also serves to average out and diminish noise in the electron micrographs. Member of the Cystoviridae family φ12’s host-
attachment proteins and other surface proteins have been studied by cryo-electron tomography (Hu et al. 2008) and the method has been used by many studies to understand viral structures (Guerrero-Ferreira & Wright 2013).

In Subtomogram Averaging, the averaging process is taken one step further. Multiple subjects or particles in the sample field are identified. Computational methods are used to verify their similarity and average them. If the particles are symmetrical, additional averaging may take place, increasing the resolution of the resulting density map, assuming the symmetry is reflected in nature.

Taken yet another step further, subvolumes of a tomographic reconstruction may be averaged with symmetry or without, or with different symmetry from what is expected. In some cases, reconstruction with different symmetries or relaxed symmetry leads to a better understanding of the symmetry of parts of a structure. For instance, in φ12, reconstruction of the host attachment proteins at different symmetries confirmed that the complex is hexameric (Leo-Macias et al. 2011).

In the case of the φ6 nucleocapsid, reconstruction with different symmetries is important because of the symmetry mismatch between the fivefold vertex of the P1 shell and the sixfold symmetry of the P4 hexamer (de Haas et al. 1999). It is also important because imposed symmetry may mask asymmetric features in a nearly symmetric structure such as the procapsid. Although the P1 shell is believed to exhibit 5-3-2 symmetry and this is confirmed in numerous reconstructions (Sen et al. 2008; Nemecek et al. 2011; Katz et al. 2012; Nemecek et al. 2013), P4 has been shown to occupy the fivefold vertex of the φ6 capsid in an asymmetric fashion (Nemecek et al. 2010). A closer examination of this occupancy and of the attachments between P1 and P4 would require relaxed symmetry in the reconstruction.
3.2.6 Homology Modeling

Homology modeling, pioneered in the 1990’s, is based on the structural similarity between proteins with similar sequence. The term “homology” is often misused to refer to sequence similarity, but refers to commonality in ancestry of similar proteins in different organisms. In many cases, structure, sequence and function coincide and the nature of functional building blocks of structural and catalytic proteins can be recognized from sequence alone.

In this study, homology modeling was used to understand the character of the connection between the P4 packaging motor and the P1 shell. Because a model of φ12’s packaging motor was available (Mancini et al. 2004) and had significant sequence similarity to that of φ6, P4 from φ12 was used as a template to produce a model of the φ6 P4 hexamer.

3.3 Materials and Methods

3.3.1 Sample Preparation and Electron Microscopy

Electron micrograph data for both single particle analysis and cryoelectron tomography of bacteriophage φ6 procapsids and nucleocapsids was kindly provided by Alex Wei and Paul Gottlieb for all studies presented here.

3.3.2 Single Particle Analysis

Micrographs stored in MRC format were tested for acceptable power spectra using utilities from EMAN and SPIDER, and inspected for other anomalies; the highest quality micrographs were included in the analysis (Ludtke et al. 1999; Shaikh et al. 2008). Particles were located and coordinates were recorded using IMOD (Kremer et al. 1996; Katz et al. 2012). An initial model was provided by previous studies (Katz et al. 2012). A refined model was created with AUTO3DEM and resolutions for each map were determined by Fourier Shell Correlation (FSC) (Yan et al. 2007).
3.3.3 Measurement of Densities in Reconstructions

Features of the φ6 procapsid were located visually on the central slice of the reconstructed procapsid’s density map. The application 3DMod, part of the IMOD toolkit (Kremer et al. 1996), facilitated visualization of density maps and measurement of densities. The 3DMod option “Pixel View” was used to search for the highest density value in the region of the feature measured. P1, P4 and P8 densities for each reconstruction were measured. Because of variation in the amplitude of different reconstructions, P4 and P8 measurements were normalized to P1 density for comparison.

3.3.4 Tomographic Reconstruction

Tomographic reconstruction was used to study the unexpanded procapsid of bacteriophage φ6. Procapsids were previously produced by expression of the plasmid pLM687 in an Escherichia coli host (Katz et al. 2012; P Gottlieb et al. 1988), whereupon samples were frozen cryogenically and imaged on a JEOL 2100 microscope to produce a tilt series. The tilt series data was kindly provided for this study by Alex Wei and Paul Gottlieb.

Raw electron micrographs in a tilt series were normalized and aligned, and assembled using the eTomo software and workflow from the IMOD toolkit (Kremer et al. 1996; Mastronarde 1997), and using the BSOFT software package (Heymann et al. 2008). Each tilt series was then used to compute a tomogram using the respective software.

3.3.5 Subtomogram Averaging

Tomograms of unexpanded procapsid were examined for resolution and particle quantity and condition, and the best tomograms were used in subtomogram averaging. Individual procapsid particles in the tomogram were marked (“boxed”) using the 3dMOD tool from the IMOD toolkit. A script was generated for particle extraction; coordinates of individual particles
were used to extract regions from the larger tomogram and generate separate density maps for each particle. Alignment, averaging and reconstruction of particles was performed in MATLAB using the TOM software toolbox and AV3 (Nickell et al. 2005; Förster et al. 2005).

3.3.6 Portal Averaging

Each φ6 particle harbors twelve fivefold vertices, or “portals”, each of which may or may not be occupied by a P4 hexamer. In order to contrast occupied and unoccupied portals, occupancy was determined visually and portals were processed in a separate subtomogram averaging process. MATLAB scripts were used to generate new volumetric selections containing the portal regions, centered on the location of the P4 hexamer. Two sets of portals were studied; occupied and unoccupied. Subtomogram averaging was performed for both sets of portals with relaxed symmetry, and a representative volume was produced for occupied and unoccupied portals.

3.3.7 Density Measurement and Image Production

For tomographic reconstructions, densities were measured using 3DMod as described above, or by direct data observation in MATLAB. Images were generated using 3DMod and Chimera (Pettersen et al. 2004; Yang et al. 2012; Kremer et al. 1996).

3.3.8 Homology Modeling

The protein sequence for bacteriophage φ6 packaging motor P4 was obtained by translation of the reference sequence at the National Center for Biotechnology Information database at the National Library of Medicine (NCBI/NLM), accession # NC_003715.1 (Mindich et al. 1988). The sequence was submitted with default parameters to the I-TASSER submission form (Zhang 2008; Roy et al. 2010; Roy et al. 2012). A hypothetical model for a φ6 P4 monomer was returned and its references were analyzed. The principal structural reference was
the φ12-P4 crystal structure previously determined, Protein Data Bank (PDB) ID:1w44 (Mancini et al. 2004; Bernstein et al. 1978). 1w44 is a crystal structure representing a crystallized trimer (half hexamer) of φ12-P4 monomers, and was used as a reference for the construction of a hypothetical hexamer.

A reference hexamer was constructed from the 1w44 trimer obtained from the Protein Data Bank (Bernstein et al. 1978). A MATLAB script (Appendix C) was used to rotate the resulting three monomers into symmetric positions to produce a complete φ12-P4 hexameric motor.

To align the I-TASSER P4 monomer model with the φ12-P4 reference, the “Iterative Magic Fit” function of Swiss PDB Viewer was applied for six copies of the P4 monomer against each of the φ12-P4 monomers present in the reference (Guex & Peitsch 1997). The six models were combined using Chimera (Pettersen et al. 2004).
3.4 Results

3.4.1 P4 Occupancy on the Procapsid is Low

Procapsids were studied by tomographic reconstruction and subtomogram averaging. Data for averaging was produced by expression of pLM687, a plasmid producing only the essential core proteins of the procapsid. To understand P4 occupancy of the portals in the unexpanded procapsid, the frequency of occupancy was counted using two methods; 1. direct observation of tomogram images, 2. density measurements of a reconstructed subtomogram average. A total of 2328 particles were included in this analysis.

Occupancy measurements by counting were made by classifying portals as either occupied or unoccupied, by visual inspection of subsections of the tomogram. A sample region of a tomogram counted is provided in Figure 28. By counting, roughly 10% of portals were occupied, on average 1.19 P4 hexamers per procapsid.

Density measurements provide an indirect measurement of occupancy. A subtomogram average is an average of many particles (2328 in this case) and with imposed symmetry it is also an average of the many possible symmetric orientations of each particle. Also, the averaging process increases signal to noise ratio dramatically. (Each particle has 12 portals, each of which may be rotated 5 ways, so there are at least 60 x 2328, or 139680 contributors to the symmetrically averaged particle.) As such, averaging in this manner produces a linear relationship between the frequency of a feature’s presence in the sample and the density, or voxel value at the feature’s location (relative to the ambient density). For instance, if ten selected particles have P4 hexamers at thirty portals, one would expect density at the portal location to be three times greater than if there were only ten hexamers per ten particles.
By density measurement, numbers reflected slightly higher occupancy; the density ratio revealed 2.34 particles per procapsid. The results are summarized in Table 7.

Table 7: Summary of Procapsid Occupancy Data

<table>
<thead>
<tr>
<th>Method</th>
<th>Total particles</th>
<th>Unoccupied</th>
<th>Occupied</th>
<th>Occupancy / Particle</th>
<th>Occupancy %</th>
<th>Density P4/P1</th>
<th>P4</th>
<th>P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Counting</td>
<td>2328</td>
<td>2118</td>
<td>210</td>
<td>1.19</td>
<td>9.9%</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Density</td>
<td>2328</td>
<td>455¹</td>
<td>1873¹</td>
<td>2.34</td>
<td>19.5%</td>
<td>0.20</td>
<td>-1.69</td>
<td>-0.33</td>
</tr>
</tbody>
</table>

¹ Count is inferred from densities
² Direct counting was not used to predict density

The discrepancy between direct counts and density measurements may be accounted for by the difficulty in recognizing occupied portals in the tomogram, as in the example shown in Figure 28. To verify that some presumed unoccupied portals contained P4 matter, the 2118 unoccupied portals were averaged to produce a subtomogram average, and densities were observed closely to determine average occupancy in the dataset. Estimates of the occupancy in...
the “unoccupied” portals vary due to noise. Occupancy estimates range between 3% and 20%. The median measurement, 10%, accounts for the discrepancy (Figure 47).

3.4.2 P4 Occupancy is Correlated to P8 Occupancy in the Expanded Nucleocapsid

The low occupancy of P4 in the procapsid motivates questions as to how it is stabilized in the expanded nucleocapsid, where it occupies all 12 portals. To examine structural interactions between P4 and P8, and between P4 and P1 where all portals are occupied, nucleocapsids were stripped of their envelopes and examined using cryoelectron microscopy. Micrographs of the nucleocapsid provided by Alex Wei and Paul Gottlieb facilitated single particle reconstruction.

Particles were selected visually for similarity, but not all particles contain the same complement of outer components. (Particles with residual envelopes were excluded, but variations still occur and are not visually discernable.) Because of the variation in particles, different reconstructions represent an average skewed by the differences in each set’s constituents.

It was advantageous to ask whether P8 and P4 occupancy are correlated. By comparing reconstructions, it was observed that in some cases, P8 and P4 density was lower than P1 density, and that lighter P4 tended to coincide with lighter P8 (Figure 30). A regression was performed to demonstrate that decreased presence of P8 on the expanded P1 core was associated with decreased presence of P4 (Figure 31, $R^2=0.6208$). Data and reconstructions are summarized in Table 8.
Figure 30: Examples of Density Measurements in Nucleocapsid
a) Reconstruction with low density of P4 corresponding to low density in P8. b) Reconstruction with high density of P4 corresponding to high density of P8. P1, P4, and P8 protein layers are indicated. Note lighter outer shell in (a).

<table>
<thead>
<tr>
<th>file</th>
<th>ID</th>
<th>ID</th>
<th>P1</th>
<th>P4</th>
<th>P8</th>
<th>nP4</th>
<th>nP8</th>
<th>n</th>
<th>Res</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>A</td>
<td>ffca</td>
<td>23.104</td>
<td>18.386</td>
<td>20.04</td>
<td>0.796</td>
<td>0.867</td>
<td>54</td>
<td>31.13</td>
<td>1</td>
</tr>
<tr>
<td>02</td>
<td>B</td>
<td>482c</td>
<td>23.75</td>
<td>16.67</td>
<td>19.33</td>
<td>0.702</td>
<td>0.814</td>
<td>65</td>
<td>30.98</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>C</td>
<td>a55f</td>
<td>874.56</td>
<td>606.2</td>
<td>643.6</td>
<td>0.693</td>
<td>0.736</td>
<td>676</td>
<td>33.98</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>D</td>
<td>da75</td>
<td>833.4</td>
<td>636</td>
<td>650.4</td>
<td>0.763</td>
<td>0.780</td>
<td>676</td>
<td>30.63</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>E</td>
<td>7a61</td>
<td>759.7</td>
<td>549</td>
<td>546.4</td>
<td>0.723</td>
<td>0.719</td>
<td>297</td>
<td>30.38</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>F</td>
<td>bac1</td>
<td>789.3</td>
<td>505.9</td>
<td>529.7</td>
<td>0.641</td>
<td>0.671</td>
<td>179</td>
<td>32.47</td>
<td>2</td>
</tr>
<tr>
<td>08</td>
<td>G</td>
<td>29fe</td>
<td>667</td>
<td>467</td>
<td>519.7</td>
<td>0.700</td>
<td>0.779</td>
<td>118</td>
<td>41.92</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>H</td>
<td>c21c</td>
<td>23.9</td>
<td>16.31</td>
<td>17.73</td>
<td>0.682</td>
<td>0.742</td>
<td>65</td>
<td>31.67</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Summary of Reconstructions and Density Data
n is the number of particles used in the reconstruction. *Resolution determined by Fourier Shell Correlation (FSC). Normalized values are normalized to P1 density. Note 1: Highest occupancy measured. Note 2: Lowest occupancy measured.
3.4.3  **Observation of Interaction between P8 and P4 in the Expanded Nucleocapsid**

Single-particle reconstructions of φ6 nucleocapsids revealed contact points between P8 and P4, potentially responsible for stabilization of the P4 hexamer. The contact point is a bridge of density between the P8 shell and each P4 hexamer; five contact points are visible per hexamer. No contact between P4 and P1 is visible at this threshold; P4 appears to be “hovering” (Figure 32). However, at a threshold representing lower density, five contact points are visible.
descending from the P4 hexamer to the now expanded P1 core (Figure 33). This indicates that P1-P4 interactions have less density than P1-P8 interactions.

Figure 32: Contact points visible in nucleocapsid reconstruction
a) Isosurface view of φ6 nucleocapsid, threshold adjusted to illustrate contact points between P4 and the P8 matrix, indicated in blue lettering. Blue arrows indicate two of five attachment regions surrounding the P4 hexamer. b) Partially rotated central slice. c) Centrally sliced view of nucleocapsid. Note space between P1 layer and P4 hexamer. Blue arrow indicates an attachment point between P4 and P8. Orange arrows indicate contact points between the P8 layer and the P1 core.

Figure 33: Contact points between P1 and P4 in Nucleocapsid
Blue arrow indicates contact between P4 and P8. Black arrow indicates one of five contacts between P4 and P1.
Exact measurements were made of the coordinates high density regions indicated in Figure 33. The attachment regions span a radial gap of approximately 12 Angstroms from the sharp drop in density of P4 to the sharp rise in density of P1, and the peak densities turn counterclockwise as they approach P1 (counterclockwise from the “top view”, the vantage point of viewing the outside of the PC). The radius steadily decreases along the descending dense region. Taking into account all distance changes, the region spans approximately 18 Angstroms. The diameter of these regions is approximately 52 Angstroms at the P4 end, roughly half the diameter of P4. At P1, these regions fit a circle of approximately 64 Angstroms in diameter.

Figure 34: Coordinates of Peak Densities in Nucleocapsid Interaction Zone
The densities observed in the symmetry-imposed reconstruction of P4 fit a spiral pattern 52 to 64 Angstroms in diameter, smallest near P4 and increasing in diameter as they descend toward P1.
3.4.4 Observation of Interaction Region between P4 and P1 in the Unexpanded Procapsid

Assembly of the procapsid requires P4 at every vertex (Sun et al. 2012). To study the interaction between P4 and P1, the major constituent of the procapsid, two reconstructions of cryo-EM were examined. A single particle reconstruction provided 14 angstrom resolution and reveals regions of high density between P1 and P4, similar to those in the nucleocapsid.

![Figure 35: Single particle reconstruction of procapsid](image)
a) Central slice showing P1 shell and P4 hexamer. Note low occupancy of hexamer.  
b) Isosurface view of the procapsid reconstruction, in the same orientation and scale as the central slice.

Close inspection reveals that contact between P1 and P4 occurs at or very near the vertices in the concave region of the portals of the unexpanded procapsid (Figure 36).
Figure 36: Interaction between P4 and P1 Core in φ6
a) Cutaway closeup of a P4 hexamer contact region at low density threshold. b) Measurement of overall thickness of the contact region. Measurement was performed on isosurface view and on density map cross-section. c) Angular closeup of two interaction points.

3.4.5 Evidence for Asymmetric Attachment in the Unexpanded φ6 Procapsid

To understand whether the contact between P4 and P1 is symmetrical in the procapsid, a tomographic reconstruction was performed on preparations of procapsids. Subtomogram averaging was performed on 194 particles with relaxed symmetry. Because of the relaxed symmetry and moderate particle count, noise in the reconstruction is much higher than in the single particle analyses, which have 5-3-2 symmetry imposed. P4 density was not easily resolved in the whole-particle subtomoram average due to these factors, but is represented in
Figure 37. Densities of P4 at the twelve portals varied (by measurement of the maximum density). This is expected as a consequence of relaxed symmetry. The particle alignment procedure preferentially aligned occupied portals with occupied portals.

**Symmetry-free P4 Occupancy**

The alignment of tomogram data to the reconstruction makes it possible to know the location of each portal in each particle in the tomogram. In other words, the reconstruction process provides the angles and coordinates of P4 in every particle sampled from the tomogram. This information was used to generate a subtomogram average of the occupied and unoccupied portals selected. There are twelve portals in each procapsid; this yields a total of 2328 portals from 194 particles. Portals were visually assessed for occupancy and grouped for reconstruction.

Alignment and averaging produced density maps representing occupied and unoccupied portals. Unoccupied reconstruction represented 2118 portals and occupied reconstruction represented 210 portals.
Figure 38: Reconstruction of Occupied Portals with Relaxed Symmetry
a) Three views of the reconstructed occupied portal with P4 centrally located. b) View of the reconstructed unoccupied portal. c) Profile view of the portal, cutaway, and central slice of density map of the occupied portal. d) Central cross-sectional slice of the unoccupied portal reconstruction.

3.4.6 Visualization of P4 Homology Model and C-Terminal Region Essential for Attachment

The P4 sequence was modeled using the I-TASSER server, which combines homology modeling, energy-minimization, de-novo, and several other techniques (Zhang 2008; Roy et al. 2010; Roy et al. 2012). The primary structural template for φ6 P4 was the crystal structure of P4 of Bacteriophage φ12, PDB ID 1w44 (Mancini et al. 2004; Bernstein et al. 1978). Alignment of φ6 P4 to φ12 P4 is reported in Table 10. Figure 39 illustrates the alignment between φ12-P4 and φ6-P4. The aligned φ6-P4 model is presented with two views in Figure 40. The monomer is also illustrated in context within a P4 hexamer surface representation (Figure 41), where the C terminus essential for binding P1 is shown in red. Outside, profile and bottom views allow for visualizing potential binding partners for P1 and P8.
Figure 39: φ6 P4 and φ12 P4 structural alignment
φ12 P4 hexamer from crystal structure shown in blue. φ6 hypothetical structure shown in tan.

Figure 40: Hypothetical model of P4 hexamer
a) Top view. Coloring Blue for N-terminus to Red for C-terminus.
b) Side view. Red c-terminal region represents a disordered region in the crystal structure of φ12 P4, implicated in attachment to P1.
3.4.7 Measurement of Tether and Distance from P1

The overall diameter of the hypothetical P4 hexamer is 109 angstroms by measurement of the farthest sidechain atoms in the model. In density maps, P4 often appears to “hover” over the P1 layer, presumably because the connection between them is very low in density. The C-terminal end (residues 322-331) as well as several earlier residues (300-306) have been found to be disordered in the crystal structure of φ12 P4 (Mancini et al. 2004). To gauge what components of the P4 C-terminus may be involved in attachment to the procapsid, regions of the C-terminus were measured. Measurements are summarized in section 3.4.3 (page 73) and in Discussion: Measurements (page 90).
3.5 Discussion

The tenuous connection between P4 and its capsid proposes a model to explain poorly understood steps in the \( \varphi 6 \) assembly pathway. P4 appears to be shed and transiently attached in order to facilitate limiting the number of portals used during packaging. Previous studies have observed incomplete occupancy but stopped short of explaining the process over multiple steps. Here we present a model that accounts for asymmetry and a transient restriction on the portals allowing packaging of ssRNA into the \( \varphi 6 \) procapsid.

Through cryo-electron microscopy, the interaction between P4 and neighboring molecules was observed. The connection between P4 and P8 is distinct from the connection between P4 and P1. The latter connection is present in both the unexpanded procapsid and the nucleocapsid. In the expanded procapsid, the connection between P8 and P4 appears to be stronger than the connection between P1 and P4. Reconstruction of the portal with relaxed symmetry illustrates an asymmetrical binding of P4 to the P1 fivefold vertex.

The geometry of the observed contact regions between P1 and P4 is consistent with results indicating that a C-terminal region of P4 is required for binding the procapsid (Paatero et al. 1998) and with the modeled spatial location of the disordered C-terminal region of P4, relative to the known portion of the \( \varphi 6 \) P4 crystal structure (El Omari et al. 2013). Taken together, these findings establish a framework for developing a model of transient attachment and release of the P4 hexamer during the \( \varphi 6 \) life cycle.

3.5.1 Life Cycle: Packaging and Maturation: Transient Occupancy of P4

P4 is believed to be required at almost every vertex during assembly (Sun et al. 2012), but only one portal is active during packaging (Gottlieb et al. 1992; Onodera et al. 1995; Qiao, Casini, et al. 1995; Mindich et al. 1995; Qiao, Qiao, et al. 1995; Onodera et al. 1998b; Onodera
et al. 1998a; Mindich 1999; X. Qiao et al. 2003; Mindich 2004; Huiskonen et al. 2006). The explanation for this may lie in the tenuous nature of the connection between P4 and P1, the kinetics of which govern the activity of one portal during initial packaging and expansion. If one P4 hexamer is bound and packaging begins, it may simply be unlikely for another P4 hexamer to bind, for ssRNA to bind, and for the translocation process to begin at another portal simultaneously.

In the expanded nucleocapsid, density at the P4-P1 interaction is lower than the density at the P4-P8 interaction (Figure 32). This suggests that the P4-P8 connection is more stable. Moreover, when P8 is absent in the procapsid, P4 occupancy is low (Table 7), and in the nucleocapsid where P4 occupancy is high, P4 occupancy correlates to P8 presence on the surface (Figure 31). The presence of P8 seems to dictate the presence of P4, and this agrees with a model in which the P4-P1 connection is very unstable.

It is tempting to speculate that P4 promotes the nucleation of P8 on the nucleocapsid or vice versa, however, P8 associates the envelope with the nucleocapsid, and P9 has been found to nucleate P8 into vesicle-like pre-envelopes during the assembly pathway of φ6 (Stitt & Mindich 1983; Sarin et al. 2012), which occurs separately and concurrently to procapsid assembly. Translocation of RNA after the acquisition of the lipid-containing envelope is not implied by any model (it would be blocked by lipid), so it is also not clear whether P8, which has a greater affinity for P4, incorporates P4 into its matrix during formation of that envelope. The higher density interaction between P4 and P8 (than that between P4 and P1) suggests that P8 has a higher affinity for P4 than does P1, as does the concomitant loss of P4 along with P8 during sample preparation. A higher affinity for P4 by P8 in turn suggests that free P4 is taken out of
solution by the P9 and P8 vesicle formation process, thus inhibiting further transcription and translocation, and that a P9-P8-P4-lipid complex envelops the inner core of the nucleocapsid.

A study of φ8 also reports that in the fully packaged nucleocapsid, P4 is not tightly connected to the procapsid shell; this is consistent with a relatively fragile connection between P4 and P1 (Huiskonen et al. 2007).

In the model presented here, P4 binds transiently as needed for packaging and translocation of transcripts, and when sufficient P8 and P9 are manufactured to catalyze formation of envelope precursor vesicles, P8 binds P4 and incorporates it into the vesicle.
Figure 42: Model for Sequential Packaging and P4 Stabilization

a) Hexameric P4 binds transiently to the unexpanded procapsid (PC). b) Recognition of specific binding sites on P1 by the pac region of the small segment, coincident with the binding of P4, results in initiation of the P4 packaging machinery, and subsequent retention of P4 by a one-way valve mechanism. C) After packaging of the small segment, P4 is released. d) Expansion exposes different binding patterns on P1 which recognize the medium segment pac region. e) Packaging proceeds through the same means for the medium and large segments, each time altering the binding affinity for the different segments by exposing different residues of P1 through morphological changes which occur during expansion. The full procapsid does not require P4 binding in the absence of P8. f) P8 preferentially binds the full capsid, and the nucleocapsid forms by addition of P8. g) P8 stabilizes P4 on the nucleocapsid.
3.5.2 Asymmetry

The φ6 core complex is generally taken as symmetrical, presumably due to its symmetrical appearance in micrographs and the success of symmetry-based reconstruction methods. This study observes asymmetric components and considers the implication of asymmetry in the procapsid.

Asymmetry in cystoviridae has been seen before. In φ8, the symmetry detected in densities representing the P4 hexamer under imposed fivefold symmetry appears to reflect asymmetry (Jäälinoja et al. 2007). P2 is shown to occupy asymmetrical sites (Sen et al. 2008). In φ6, P4 has also been visualized as occupying portals in an asymmetric fashion (Figure 43) (Nemecek et al. 2010), with great similarity to the symmetry-relaxed reconstruction of the portal presented here (Figure 38). One study proposed that one fivefold portal and P4 hexamer of φ6 is characteristically and structurally different from all other portals (Pirttimaa et al. 2002).

In this study, two asymmetries were observed: 1. The lack of P4 in every portal, which implies an overall asymmetric procapsid configuration (Figure 37), and 2. the asymmetric location of the P4 hexamer in the symmetry-free reconstruction (Figure 38).

The incomplete occupancy of portals in the procapsid directly implies that P4 localization is not symmetric, and although the PC may be taken as very stable, it is plausible that conformational changes in P1 are produced by the binding of P4. Whether such changes result in altered affinity for P4 or ssRNA has not been considered by this study. Nevertheless, the asymmetric binding of P4 supports a model of transient occupancy.

Asymmetry observed in the portal supports the transient nature of the connection between P4 and P1. All other reconstructions demonstrating contact between P1 and P4 have relied on imposed symmetry, obscuring any possible tendency of P4 to bind partially or asymmetrically.
Nemecek et al. also observed this asymmetry and used it to justify the unexpected density distribution of the P4 hexamer in reconstructions with imposed symmetry (Figure 44) (Nemecek et al. 2010). This observation is not sufficient to explain the uniqueness of one portal per procapsid, as Figure 19 illustrates particles with multiple hexamers bound asymmetrically in their respective portals (Figure 19, A, row 4, 6.).
**Figure 43: Asymmetrical anchoring of P4 on the Procapsid**
Non-central attachment of the P4 hexamer in φ6 procapsid portals. The top row is a subtomogram average for reference, representing P4 if bound centrally. The remaining rows represent individual particle given as examples of asymmetrical binding. In (A), the particle is viewed along the twofold axis; in (B) the particle is viewed along the fivefold axis. White arrows indicate P4 and attachment points in (A) and (B) respectively. (Nemecek et al. 2010).

**Figure 44: “Wobble” of the P4 hexamer**
Free movement of the P4 hexamer in individual portals of the φ6 procapsid obscured by symmetric reconstruction (averaging) is represented in this figure. (Nemecek et al. 2010)
In this study, the observed asymmetry justifies the transient connection between P4 and P1. Multiple studies observe low occupancy of P4 in the P1 fivefold vertex of a symmetric reconstruction of the procapsid or present data illustrating this (de Haas et al. 1999; Huiskonen et al. 2006; Sen et al. 2008; Katz et al. 2012; Nemecek et al. 2012; Nemecek et al. 2013). In all cases, symmetry was imposed, and wherever reported, five symmetrical attachment points could be visualized in the reconstruction. Here, two asymmetrical reconstructions added to our understanding of those attachment points. In a symmetry-relaxed reconstruction of the procapsid, low occupancy was confirmed. Also, because the reconstruction aligned occupied portals with each other, it was possible to ask whether portals were occupied preferentially. No preference was observed (Figure 37), as all possible 2nd portals were occupied with roughly the same frequency. The lack of preference for a 2nd P4 binding site suggests that binding is random, and not due to any conformational changes in the procapsid.

Additionally, a symmetry-relaxed reconstruction of occupied and unoccupied portals was observed. Attachment in the occupied portal average was observed to be heavily weighted toward one of the five twofold vertices. This weighting is the expected result for random, transient binding of the P4 monomer constituents of the P4 hexamer in the portal. Again, the reconstruction process aligns the densest regions of all individuals- and high density occurs where the first P4 monomer is bound. P4 retains its structure as a unit whether it is bound or unbound (de Haas et al. 1999). One source suggests that if the P4 hexamer binds to one attachment point, it may no longer bind to another attachment point in the same way due to the mismatch in symmetry (Nemecek et al. 2010); regardless, if additional attachments do occur, the other P4 monomers presumably bind stochastically in the same manner as the first. Since the initial binding was low in frequency, it would be expected that subsequent bindings are even
lower, taking into consideration the geometrical challenge of the symmetry mismatch. In the symmetry-relaxed reconstruction, no significant density was observed beyond the initial attachment point. This suggests any of several possibilities: 1. secondary attachments in the procapsid do not occur (possibly due to symmetry mismatch), 2. attachments occur at a very low rate such that densities are too low to detect, but are present, 3. multiple attachments are dependent on ssRNA substrate or other factors, and do not occur in the sample conditions.

3.5.3 Measurements

The 12 angstrom gap between P4 density and P1 density observed in the procapsid reflects the greater distance between P4 and P1 in the procapsid than that which occurs in the packaged nucleocapsid (observed at 8 Angstroms in φ8) (Huiskonen et al. 2007). The spiral nature of the observed densities suggests that the tether apparatus is at least 18 angstroms in length.

To span that gap, a peptide tether must contain roughly 18 Angstroms of residues plus enough residues for binding. Straight peptide chains contain roughly one residue per 1.5 angstroms; an 18 Angstrom chain would contain 12 residues. The disordered region of φ12-P4’s C-terminal end revealed by crystallization contains 10 disordered residues (322-331), preceded by an ordered region of 15 AA’s (307-321), preceded by another 7 disordered residues (300-306) (Mancini et al. 2004). It is within reason to speculate that residues 300-331 form a tether which extends flexibly (between residues 300-306) from P4, controls distance in the 15 AA region, and binds P1 flexibly using some subset of the remaining 10 residues. In the energy-minimized reconstruction of the P4 hexamer presented, no apparent secondary structure occurs in this C-terminal region. The flexibility implied by a disordered region in the crystal structure would afford P4 the ability to bind under the range of geometries the capsid would undergo during...
expansion, and the range of motion afforded by the extra length may partially overcome the
symmetry mismatch, although the sixfold axis of the P4 hexamer might not be expected to
coincide with the fivefold symmetry axis of the portal, consistent with previous conjecture
(Nemecek et al. 2010) and reports in φ8 (Huiskonen et al. 2007). The measurements of diameter
and length of the apparent tether in EM density maps are consistent with the features of the C-
terminus of the homology-modeled structure derived from the crystal structure of φ12’s P4
hexamer. As the P4 C-terminus is known to be required for binding the P1 shell but not for P4’s
catalytic activity (Paatero et al. 1998), a measured C-terminus region disordered in crystal
structures may be hypothesized to provide P1-binding residues with sufficient intervening
residues to allow for a flexible tether apparatus, which in turn permits binding of the hexameric
P4 to the symmetry-mismatched fivefold vertex.

3.5.4 Conclusion

The P4 hexamer appears to bind transiently to the P4 procapsid. The contacts between
P4 and P1 as well as those between P4 and P8 are visualizable in EM reconstructions, and P4
occupancy coincides with presence of P8 on the nucleocapsid, indicating that P4 binds
transiently and is stabilized by P8. Taken together with recent findings that the P8 shell forms
separately from the packaged expanded procapsid, it is unclear whether P4 first binds the
nucleocapsid to facilitate recruitment of the pre-envelope vesicles or is taken from solution by P8-
P9-lipid envelopes to form a complete shell which then binds the nucleocapsid (sans P8). The
visualized attachment points can be measured, and measurements are consistent with a model
identifying features of the C-terminus as having a flexible tether and an anchoring region binding
the P1 major capsid protein.
Chapter 4:

Summary

4.1 Chapter 2: Evidence for a Functional Role in the 3’ UTR of the Small Segment of Bacteriophage φ6

By targeting specific regions of RNA and observing the effects on plaque morphology and fitness, evidence was provided to suggest the existence of genes and secondary structure in a previously unmapped region. More study is needed to give a conclusive understanding of the roles these factors play in the fitness and plaque morphology of bacteriophage φ6.

Strain JC1917 provides a special example illustrating that depletion of optical density is not universally related. Phage are productive in this assay, but bacterial density appears higher than untreated bacteria (no bacteriophage) under otherwise identical conditions. This wholly unexpected result indicates any of several possibilities: 1. Bacterial debris are being formed which do not inhibit cell growth; 2. Fitness of the host is increased by JC1917; 3. Quorum sensing is inhibited by JC1917; 4. Absorption of light by bacteria is altered by JC1917 infection, including the possibility that bacteria are forming aggregates; 5. Bacteriophages are forming aggregates which impact absorption. This is an intriguing problem and it demands further investigation.

Data suggest that alterations in the 3’ UTR of the φ6 small segment may impact plaque formation through several means, including changing rates of diffusion, formation of multi-phage aggregates, and changing phage affinities and attachment rates for environment and/or bacteria. These factors impact fitness, and plaque morphology can be used as an indicator for a change in life cycle parameters. Additionally, unresolved questions in bacteriophage φ6’s life history are potentially answered by the function of proteins proposed by this study. Further work
is needed to isolate these proteins or exhaustively disprove their existence and explain the
dramatic fitness effects demonstrated here. The unusual impact of the fusion mutation represents
a convenient scheme for diagnosing the function of, or zeroing in on difficult-to-find genes; the
mutants constructed here can be used in further studies, including complementation studies, to
determine whether and where products of ORF I and ORF J function in the cell, in the virion,
and in the life cycle of bacteriophage φ6.

4.2 Chapter 3: Transient Occupancy of P4 in the Portals of the φ6 Procapsid

The tenuous connection between P4 and its capsid proposes a model to explain poorly
understood steps in the φ6 assembly pathway. P4 appears to be shed and transiently attached in
order to facilitate limiting the number of portals used during packaging. Previous studies have
observed incomplete occupation but stopped short of explaining the process over multiple steps.
Here we present a model that accounts for asymmetry and a transient restriction on the portals
allowing packaging of ssRNA into the φ6 procapsid.

4.3 Conclusion

This paper sought to align evidence from several sources: 1. The body of knowledge
regarding the life history of bacteriophage φ6, and unknown aspects of such; 2. Fitness and other
observational data on mutants of the small RNA segment of bacteriophage φ6, and 3.
Observations on data from cryoelectron microscopy of bacteriophage φ6.

Quantitative and qualitative data suggest that alterations in the 3’ UTR of the φ6 small
segment may impact plaque formation, clearing of a lysate, and fitness. It is believed that these
changes represent alterations to life history parameters. Additionally, unresolved questions in
bacteriophage φ6’s life history are potentially answered by the function of proteins proposed by
this study. Further work is needed to isolate these proteins or exhaustively disprove their
existence and explain the dramatic fitness effects demonstrated here. The unusual impact of the fusion mutation represents a convenient scheme for diagnosing the function of, or zeroing in on difficult-to-find genes; the mutants constructed here can be used in further studies, including complementation studies, to determine whether and where products of ORF I and ORF J function in the cell, in the virion, and in the life cycle of bacteriophage φ6.
Appendix A: Fitness and OD Depletion of Additional Mutant Strains

Figure 45: Fitness of 35 isolated φ6 mutants
Fitness of additional strains constructed during the preparation of MUT1 thru MUT6.

Figure 46: Fitness of 30 strains vs. Depletion in Optical Density
Fitness and OD Depletion Assay results of additional strains constructed during the preparation of MUT1 thru MUT6.
Appendix B: Supplementary Figures

Figure 47: Observation of material in presumed unoccupied portal

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>P1</th>
<th>P4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Occupancy</td>
<td>Density</td>
<td>P4/P1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimmest ambient</td>
<td>3.0%</td>
<td>0.03</td>
<td>2.326</td>
<td>0.069</td>
<td>0.08</td>
</tr>
<tr>
<td>Median Ambient</td>
<td>10.9%</td>
<td>0.11</td>
<td>2.534</td>
<td>0.277</td>
<td>0.29</td>
</tr>
<tr>
<td>Brightest Ambient</td>
<td>18.4%</td>
<td>0.18</td>
<td>2.767</td>
<td>0.510</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Table 9: Measurement of material in presumed unoccupied portal
Portals were selected visually for occupancy and vacancy, but some portals may appear unoccupied due to noise in the tomogram. Measurements were taken to estimate how many presumed unoccupied portals were occupied. Because of variation in the ambient density, the estimate varies from 3% to 18%.
<table>
<thead>
<tr>
<th>PDB Hit</th>
<th>Identity 1</th>
<th>Identity 2</th>
<th>Coverage</th>
<th>Normalized Z Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1W44</td>
<td>0.23</td>
<td>0.20</td>
<td>0.83</td>
<td>3.06</td>
</tr>
</tbody>
</table>

Table 10: Sequence Alignment of φ6 P4 and φ12 P4

Identity 1 is the percentage sequence identity of the templates in the threading aligned regions with the query sequence. Identity 2 is the percentage sequence identity of the whole template chains with the query sequence. Coverage represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of the query protein. Normalized Z-score: Alignment with a Normalized Z-score >1 suggests a good alignment (Roy et al. 2010; Zhang 2008).
Figure 48: Procapsid or Polymerase Complex (PC) vs. Nucleocapsid (NC)

(Jäälinoja et al. 2007)
Appendix C: Matlab code for P4 rotation

```matlab
p4=pdbread('1W44.pdb');
% molviewer(p4);
p4a=p4;

xcenter=p4.Cryst1.a;
zcenter=0;
count=max(size(p4.Model.Atom));
% 6867 is 2289*3, btw.
for n=1:count
    x=p4.Model.Atom(n).X;
y=p4.Model.Atom(n).Y;
z=p4.Model.Atom(n).Z;

    % Rotate 180 degrees around Y:
    x=-x;
z=-z;

    % Translaje:
    x=x+xcenter;
    p4a.Model.Atom(n).X=x;
p4a.Model.Atom(n).Y=y;
p4a.Model.Atom(n).Z=z;
end

% Apply the same transformation to the HETATMs:
count=max(size(p4.Model.HeterogenAtom));
for n=1:count
    x=p4.Model.HeterogenAtom(n).X;
y=p4.Model.HeterogenAtom(n).Y;
z=p4.Model.HeterogenAtom(n).Z;

    % Rotate 180 degrees around Y:
    x=-x;
z=-z;

    % Translaje:
    x=x+xcenter;
    p4a.Model.HeterogenAtom(n).X=x;
p4a.Model.HeterogenAtom(n).Y=y;
p4a.Model.HeterogenAtom(n).Z=z;
end

pdbwrite('1W44-rot.pdb',p4a);
```
References


control. *Journal of virology*, 58(1), pp.142–51. Available at:


Nemecek, D. et al., 2010. Cryo-electron tomography of bacteriophage phi6 procapsids shows random occupancy of the binding sites for RNA polymerase and packaging NTPase.


