ABSTRACT

Dr. Ray A. Larsen, Advisor.

The first and most crucial step in a bacteriophage’s life cycle is the adsorption step. This step entails docking on to the phage’s host and inserting its genetic material into the host cell. Should this first important step not occur, the phage will remain an inert particle in the environment until a more suitable host is found. For the lambdoid phage, φ80, this is a two step process. An initial reversible adsorption results from docking to the outer membrane siderophore transporter FhuA by weak electrostatic interactions. Additionally, a second nonreversible step requires the presence of TonB and a functional proton motive force. Although adsorption process of φ80 has been determined, the phage protein that mediates this process has not been characterized. This thesis describes the means of finding and characterizing the open reading frame responsible for the production of the φ80 tail spike. Once found, the predicted product of this open reading frame was subjected to proteomic analyses to determine sequence similarities to other lambdoid phage and phage that require FhuA and/or TonB for their adsorption process. Like the phage λ, the putative spike protein of φ80 was predicted to have a functional domain near the carboxyl terminus that mediates receptor binding to FhuA. Additionally, a putative TonB box was predicted in the amino terminus that allows for the interaction with TonB, providing energy for the uptake of the phage genome.
AKNOWLEDGEMENTS

The completion of my Master’s degree could not have happened without the help of many important people. Their support and guidance along the way have been most appreciated and I would like to take a moment to thank these individuals.

First and foremost, I would like to thank my advisor, Dr. Raymond Larsen. His guidance and support helped me to learn various techniques used in the lab as well as in writing this thesis. He has been a crucial component towards my education in microbiology. Additionally, I would like to thank my committee, Dr. Tami Steveson, Dr. Stan Smith, and Dr. Jill Zeilstra for their guidance as well. Having known them from my undergraduate years, and in Dr. Smith’s case since high school, has been a great experience.

Secondly, I would like to thank my family. They’ve given me more than just a place to live throughout my education; their love and support kept me on the right track. My parents, Steve and Karen, have questioned me through this process to help me think deeper in my subjects and find aspects in my education that really matter to the outside world.

As for my peers and colleagues, I would like to thank those in my lab first, as they’ve been a big help towards my research in many ways. Todd Cramer and Krishna Mutgi have contributed to helping the lab understand the workings of bacteriophage φ80 as well as being great people to chat with. Previous lab members, Dr. Kerry Brinkman and Cari Smerk, have been a great influence on my learning as well as being entertainment in the lab. Tim South, Mary Teleha, and many of the other undergraduate researchers have shared many great times in the lab as well.
Lastly, I would like to thank Bowling Green State University. Not only has the university been my undergraduate alma mater, but also financially supported both the research and my living expenses through my Master’s career.
TABLE OF CONTENTS

CHAPTER ONE – INTRODUCTION..................................................................................1

CHAPTER TWO – AIM ONE: IDENTIFICATION OF THE GENE ENCODING THE PUTATIVE φ80 TAIL SPIKE PROTEIN.................................................................9
  Introduction............................................................................................................9
  Materials and Methods.........................................................................................10
  Results..................................................................................................................18

CHAPTER THREE – AIM TWO: IDENTIFYING FUNCTIONAL DOMAINS ON THE φ80 J HOMOLOG.................................................................28
  Introduction............................................................................................................28
  Materials and Methods.........................................................................................28
  Results..................................................................................................................29

CHAPTER FOUR – FUTURE RESEARCH DIRECTIONS.............................................39

References...............................................................................................................41
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. φ80 bacteriophage and the proposed mechanism of entry</td>
<td>3</td>
</tr>
<tr>
<td>2. pUC19 derivatives cloning inserts of φ80 genomic DNA</td>
<td>10</td>
</tr>
<tr>
<td>3. Partial digest of the φ80 genome using different relative quantities of Sau3AI</td>
<td>13</td>
</tr>
<tr>
<td>4. Relative position of φ80 clone 103 DNA to the known map of phage λ</td>
<td>21</td>
</tr>
<tr>
<td>5. Relative position of φ80 clone 341 DNA to the known map of phage λ</td>
<td>22</td>
</tr>
<tr>
<td>6. Relative position of φ80 clone 563 DNA to the known map of phage λ</td>
<td>23</td>
</tr>
<tr>
<td>7. Assembled φ80 open reading frame encoding the host attachment spike</td>
<td>26</td>
</tr>
<tr>
<td>8. Restriction analysis of PCR amplified φ80 host attachment spike DNA</td>
<td>27</td>
</tr>
<tr>
<td>9. Alignment of the first 150 amino acids of φ80 J protein with the amino terminus of similar phage attachment spikes</td>
<td>31</td>
</tr>
<tr>
<td>10. An alignment of the carboxyl terminus of the φ80, HK022, T1, Lambda and T5 host attachment spikes, generated using PRALINE</td>
<td>32</td>
</tr>
<tr>
<td>11. Percentage of amino acid similarity mapped on the φ80, λ and T1 predicted J protein sequences</td>
<td>33</td>
</tr>
<tr>
<td>12. The crystal structure of FhuA in stereo view generated using the Swiss PDB viewer</td>
<td>35</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION

Bacteriophage, also known as phage, are a group of viruses that specifically infect bacteria. Similar to eukaryotic viruses, phage are simply genetic material, either DNA or RNA, encapsulated in a protein coat. As such, they are obligate intracellular parasites that require a host for replication. Without a host to infect, phage act as inert particles in the environment until they encounter a potential host. All viruses share common needs for survival and propagation. First they must find a host cell and attach to it. Then the virus must get its genome to the correct area of the cell to be replicated. Next the viral genome must be copied, viral proteins produced, and the viral components assembled into new viral particles. Lastly, the virus must be able to escape the host in order to start the cycle over and infect another potential host. Although there are variations in this simple life cycle, viruses cannot reproduce alone.

Another aspect shared between eukaryotic viruses and phage is their diversity. Even within the tailed phage (order Caudoviridae) considered in this study, there is a great deal of diversity among phage populations. For example, it has been estimated that there are about 10 different phage species that can infect any given bacterial strain and total phage in the environment outnumber bacteria by a ratio greater than ten to one (Brüssow and Hendrix, 2002). Given this ratio, it is evident that phage are the most abundant entity on earth. In addition, many isolated phage have features from apparently distant families, a phenomenon fittingly termed “mosaicism” (Juhala et al., 2000). The degree of mosaicism seen within these phage renders classifying these biological agents somewhat problematic.
If we just consider the kinds of phage which *Escherichia coli* hosts as a well-studied model, the breadth of diversity is evident. Certain phage, such as Qβ, have ssRNA genome (Watanabe *et al.*, 1967). Some phage, like the filamentous M13 and the icosahedral φX174, store their genetic information in an ssDNA conformation (Sinsheimer, 1959; Ray, 1966). Lastly, the lambdoid and tailed T phage store their genetic information as dsDNA (Hershey *et al.*, 1963; Bresler *et al.*, 1967). Even though these phage infect the same organism, they use different methods to get their genome where it needs to be in their host to be replicated. Some phage such as T1 and φ80 are known to require energy from their host in order to inject their genome into the cell (Figure 1, Hancock and Braun, 1976), whereas other phage like T4 appear to use potential energy stored within the phage to deliver its genome into the host (Rossmann *et al.*, 2004).
Figure 1. φ80 bacteriophage and the proposed mechanism of entry. Initial attachment of φ80 at the outer membrane (OM) of *E. coli* involves binding to the ferric hydroxamate transporter, FhuA. Irreversible adsorption is dependent upon both the cytoplasmic membrane (CM) proton gradient (H⁺), and the energy transducer TonB. The ExBB/ExbD complex which couples TonB to the proton gradient is also depicted.

Of the large diversity of phage, several groups have been heavily studied and provide models for virus biology and genetics. One such group is the lambdoid phage, named for the member first discovered, λ. Beyond having the familiar head and tail morphology seen with the T phage (Ackerman, 2006), the λ phage was found to display two distinct life cycle patterns; a lytic and a lysogenic cycle (Lederberg, 1951). Indeed, the λ phage was first discovered in 1951 accidentally while isolating *E. coli* mutants induced by exposure to UV radiation (Lederberg, 1951).
The lytic or lysogenic life cycle decision and the genes behind this phenomenon is one of the central themes of studies of lambdoid phage. During the lytic cycle, the phage amplifies its genome, producing viral structure proteins, assembles, and finally lyses the host cell to release progeny. Conversely, the lysogenic cycle is characterized by the viral genome integrating into the host genome and passively replicating along with the host genome, with resultant transmission to each subsequent daughter cell. The factors that determine whether the phage productively infect cells or cryptically occupies the host are phage encoded gene products. This has been best studied in λ, where two proteins (N and Cro) compete to determine the course of the infection (Oppenheim et al., 1977). Though the proteins are slightly different for each lambdoid phage, the mechanisms and themes appear conserved.

In addition to having the lysogenic and lytic life cycles, the lambdoid phage display genomic mosaicism. This apparent product of lateral gene transfer is probably facilitated by lysogeny – increasing the probability of two distinct lambdoid phage inhabiting the same host cell, allowing homologous recombination. A clear example of this phenomenon came from the early observation that coinfection with λ and φ80 can produce phage particles that encode λ head proteins and φ80 tail proteins (Inokuchi and Ozeki, 1970).

The ability to recognize and adsorb to an appropriate host is critical to a virus’s replication success. Since viruses are too large to freely diffuse through cellular membranes, they must either inject their genetic material into the cell or be internalized by the cell. If the virus cannot adsorb to a potential host cell, then it cannot proceed to subsequent steps in its life cycle to produce progeny. Thus, an initial, critical component
of the host range specificity of a virus is dependent upon one or several proteins that recognize and bind to specific host cell molecules. For the lambdoid phage, the tail spike proteins are responsible for this initial host range specificity, mediating attachment to a specific molecule on the outermost membrane of the bacterium that it will infect. For example, λ’s tail spike protein, encoded by the J gene, is responsible for the λ phage adsorbing to the host cell. The J protein binds to the LamB protein on the outer membrane of its E. coli host cell (Randall-Hazelbaur and Schwartz, 1973). The LamB protein is a trimer and provides for the uptake of maltodextrins by facilitated diffusion (Szmelecman, 1975). As shown in Figure 1, the lambdoid phage φ80 infects E. coli by using FhuA, an outer membrane protein that provides for ferric hydroxamate transport (Braun et al. 1976). Unlike λ, φ80 requires a second protein to successfully infect the cell, TonB (Hancock and Braun, 1976). TonB functions to harness the proton motive force of the cytoplasmic membrane to energize certain outer membrane processes (Larsen and Postle, 2001). Because successful infection by φ80 is also dependent upon proton motive force (Hancock and Braun, 1976), it is likely that the TonB requirement reflects the need for host energy for successful injection of the φ80 genome. However, the φ80 protein(s) responsible for this host range specificity have not been characterized.

Another important characteristic of lysogenic bacteriophage is that they provide their host with immunity against superinfection by similar phage. Two common strategies of phage to protect against superinfection are to regulate transcription of newly infecting phage or to prevent the infection by limiting access to the phage’s host receptor. E. coli λ lysogens are protected from superinfection by the presence of the phage protein cl that prevents the transcription of lytic pathway genes (Reichardt and Kaiser, 1971).
Even though other λ phage can adsorb and inject their genome into the same host, the cI protein protects the cell from genes of the lytic cycle being transcribed. A second mechanism for immunity by φ80 involves the Cor protein, a similar system in φ80 which does not cross talk with the λ system (Ogawa et al., 1989). This protein is proposed to bind to TonB using a TonB box motif and thus prevent the normal function of TonB (Vostrov, 1996). By acting on the TonB system, φ80 Cor can prevent the entry of additional φ80 and other TonB dependent phage.

The main purpose of this study is to better understand the adsorption of φ80 and the phage protein involved in this function. As mentioned above, the energy required for a phage to inject its genome into a cell is either carried by the phage or provided by the host. For example, the T4 phage uses a contractile sheath to inject its DNA into the host similar to a syringe, with the required energy stored in protein conformation (Rossman, 2004). Conversely, T5 binds to the FhuA receptor and injects the first 8% of its genome, within which are genes that code for products that enable the uptake of the remainder of the genome (Maltouf and Labedan, 1983). On the other hand, phage T1 and φ80 also use the FhuA receptor but also absolutely require TonB for infection (Hancock and Braun, 1976). Although λ phage infection has been described in detail, how the energy required to inject its genome into the host is obtained is not yet known. Unlike other lambdoid phage, λ does not require the presence of TonB-transduced energy.

Another reason to study the adsorption of φ80 is to better understand the FhuA and TonB systems, which the phage pirates in order to obtain entrance into the host cell (Figure 1). With the knowledge of how this phage hijacks the system, one can better understand the structure and function of the components that transfer energy to the outer
membrane. The first protein that φ80 comes in contact with during its adsorption step is the FhuA protein. FhuA is an active transport protein on the outer membrane of many gram negative bacteria, such as *E. coli*, and is required for the binding and uptake of ferric hydroxamate siderophores (Wayne *et al.*, 1975; Carter, 2006). Structurally, FhuA consists of 22 antiparallel transmembrane strands that form a β-barrel structure, with an amino-terminally located globular domain that forms the floor of the barrel (Ferguson *et al.*, 1998). Binding of ligand to siderophore transporters such as FhuA results in a conformational change in the globular domain that exposes a short amino-terminal motif (conserved amongst siderophore receptors and termed a TonB box) that can specifically interact with TonB (Köck *et al.*, 1987; Locher, 1998). Such interactions drive a second set of conformational changes in the globular domain in FhuA, resulting in uptake of ligand (Moeck, 1997). How these changes mediate the release of ligand into the periplasm remains unclear, and how this might provide for transport of something as large as a phage genome is an even greater mystery. The first binding interaction is reversible, meaning that phage can be recovered after binding by agitation under higher salt concentrations, thus this initial adsorption probably represents weak electrostatic interactions (Hancock and Braun, 1976). In the presence of TonB, and an intact proton gradient at the cell membrane, phage are not similarly recovered. This irreversible adsorption step is thought to reflect the successful uptake of the phage genome (Hancock and Bruan, 1976).

Even though the host range and conditions that φ80 requires for entry into the host cell have been well documented, the phage protein responsible for adsorption has not been described. In this study, I have determined the DNA sequence of an open reading
frame that putatively encodes the φ80 tail spike gene. Additionally, I have analyzed features of the predicted tail spike protein to identify domains with potential to confer host range-specificity.

**Hypothesis:** The φ80 genome will encode a protein with features that allow for interactions with both FhuA and TonB, which will account for being a recipient of TonB-transduced energy required for the translocation of the φ80 genome.

To test this hypothesis, I have proposed two **Specific Aims.**

1. Determine the φ80 DNA sequence that encodes a putative tail spike protein.

2. Predict the homology and functional domains of φ80 through the application of proteomic tools on the putative φ80 tail spike sequence generated from the first specific aim.
CHAPTER TWO

AIM ONE: IDENTIFICATION OF THE GENE ENCODING THE PUTATIVE φ80 TAIL SPIKE PROTEIN

Introduction

As an initial step toward characterizing the molecular components responsible for the dependence of φ80 on the energy-transducing protein, TonB, a plasmid-based library was constructed in pUC19 by inserting fragments generated by limited Sau3A1 digestion of the φ80 genome into the BamHI site of pUC19 (Figure 2). Restriction analysis and subsequent sequence determinations identified a gene predicted to encode a protein with significant sequence homology to the J gene product of bacteriophage λ. Construction of this library, its screening and the identification of the putative J gene homolog of φ80 are presented.
Figure 2: pUC19 derivatives cloning inserts of φ80 genomic DNA. The location of the BamHI site used in library construction and the LacZα gene it interrupts are indicated. The location of sequences corresponding to the M13/pUC-20 and -24 reverse primers used for initial sequence determination are shown as well (Adapted from New England Biolabs).

Materials and Methods

Strains and plasmids. The bacterial strain used to propagate φ80 was Escherichia coli W3110 (Hill and Harnish, 1981). The φ80 genomic library was propagated in E. coli DH5α (Gibco. BLR; San Francisco, CA). Cultures were grown and maintained in Luria-Bertani (LB) broth and on LB agar plates (Miller, 1972) supplemented with ampicillin at 100 µg/ml where appropriate. Screening of transformants was done on ampicillin-supplemented LB plates containing 80 µg/ml 5-
bromo-4-chloro-3-inolyl-β-D-galactopyranoside (X-Gal) and 0.34μM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were incubated at 37 ºC, with short term storage on LB ampicillin plates at 4 ºC, and long term storage in 25% glycerol at -80 ºC. The strain of bacteriophage studied is the φ80vir strain which does not undergo lysogeny. This isolate was obtained from the laboratory of Dr. Kathleen Postle (Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA). The pUC19 plasmid (Yanisch-Perron et al. 1985) was used to generate the φ80 genomic library in E. coli.

**Bacteriophage growth.** φ80 was grown in 255 ml LB broth containing LB, 5 mM MgSO₄, 5 mM CaCl₂, and 1 ml of a fresh overnight LB broth culture of E. coli W3110. The E. coli were allowed to grow at 37 ºC for 30 min before adding 45 µl of φ80 containing either 4 x 10⁵ or 4 x 10⁶ plaque forming units (pfu), or no phage for a negative control. Cell density was monitored periodically by spectrometric absorbance readings at 550 nm (A₅₅₀). After 7 h (at which time lysis of the host cells in broth phage-inoculated cultures was evident), 2 ml of chloroform was added to the phage preparations and stirred with a stir bar. The lysates were then centrifuged for 15 min at 7500 x g and a sample titered, using a fresh W3110 culture. Both phage preparations were found to contain ~2 x 10¹¹ pfu of φ80.

**Recovery of the φ80 genome.** DNAse and RNAse were added to the lysate to a final concentration of 1 μg/ml. The lysate was then incubated at room temperature (RT) for 60 min to digest host genetic material. Then the DNAse was heat inactivated by incubation at 60 ºC for 20 min. Lysates were adjusted to 1 M NaCl and stirred at 4 ºC for 120 min. Polyethylene glycol (PEG8000) was added to 10% w/v, and the mixture stirred
at 4 °C overnight. Lysates were then centrifuged at 11,000 x g for 30 min and the resultant precipitate harvested in 8 ml of SM buffer (100 mM NaCl, 8 mM MgSO₄•H₂O, 50 mM Tris•Cl, pH 8.0), which was then extracted with an equal volume of chloroform.

To extract DNA, 1 ml of this φ80 concentrate was supplemented with 40 μl of 500 mM EDTA, 1 μg proteinase K, and 50 μl 10 % SDS and incubated at 60 °C for 1 h to remove the viral capsid and release the viral DNA. After incubation, the lysate was serially extracted with 1 ml Tris•Cl equilibrated phenol (pH 8.0); 1 ml phenol:chloroform:isoamyl alcohol (25:24:1), and 1 ml chloroform:isoamyl alcohol (24:1), with the aqueous phase retained at each step. The final aqueous phase was then dialyzed 3 times for 24 hours each against 1 L of 10 mM Tris•Cl (pH 8.0) and 2 mM EDTA.

**Construction of the φ80 genomic library.** To generate the φ80 genomic library, the extracted genome needed to be broken down into smaller pieces to be cloned. For this, I used *Sau3A1* since the restriction endonuclease generates overhangs that will ligate into *BamHI* restriction sites. To generate overlapping fragments, suboptimal quantities of the enzyme were used so that not all *Sau3A1* restriction sites were cut. To determine the appropriate amount of *Sau3A1* to use, the following experiment was performed. The φ80 genome was partially digested by adding the 50 μl of the purified φ80 DNA (at 460 μg/μl) to 50 μl of 10 x *Sau3A1* buffer (1 M NaCl, 100 mM Bis Tris Propane-HCl, 100 mM MgCl₂, and 10 mM dithiothreitol; New England Bioloabs, Ipswich, MA), 5 μl bovine serum albumin at 100 μg/ml, and 395 μl dH₂O, to a total of 500 μl. Samples of 50 μl were taken from this mixture and placed into 8 separate tubes. Then 2.5 μl of *Sau3A1* was diluted in the remaining 100 μl to produce a final concentration of 4.3 units
per μg of DNA. After the initial dilution, Sau3AI was serially diluted seven times to produce concentrations of 4.3 units/µl, 2.2 units/µl, 1.1 units/µl, 0.54 units/µl, 0.26 units/µl, 0.13 units/µl, 0.07 units/µl, and 0.03 units/µl. Next, 1 µl of each serial dilution was added to a 50 µl sample. The final mixtures were then incubated at 37 ºC for 90 min to allow for digestion of the DNA. A 10µl sample was removed from each reaction and resolved on a 1% agarose gel for evaluation (Figure 3).

The pUC19 vector was prepared to receive the Sau3AI inserts by restricting 5 µg of pUC19 with 40 units of BamHI, in 150 mM NaCl, 10 mM Tris•Cl (pH 7.9), 10 mM MgCl₂, and 1 mM dithiothreitol (New England Biolabs, Ipswich, MA), supplemented with bovine serum albumin at 100 µg/ml, for 90 min at 37 ºC. Next, to prevent pUC19
from recircularizing during the ligation reaction, the 5’ phosphates of the vector were removed by adding 5 μl of 10 x Antarctic phosphatase buffer (500 mM Bis Tris-Propane (pH 6.0), 10 mM MgCl₂, 1 mM ZnCl₂) and 5 units of Antarctic phosphatase to the vector reaction. To prevent the enzymes from reacting during the ligation step, the φ80 genome preparation and pUC19 vector preparation were heat inactivated by incubation at 65 ºC for 20 min. The final step to make the random clones was to mix 3 μl of the 0.07 Sau3AI units/μl restricted φ80 genome with 50 ng of the BamHI restricted pUC19, 10 μl of 2 x quick ligase buffer (132 mM Tris-HCl (pH 7.6), 20 mM MgCl₂, 2 mM dithiothreitol, 2 mM ATP, 15 % Polyethylene glycol), 8 μl dH₂O, and 1 μl quick ligase mix. The Sau3AI insert and vector mix then was used to transform DH5α competent E. coli cells. The competent cells were first thawed by placing 200 μl of cells on ice. Once thawed, the cells were divided into 6 tubes (for the ligation reaction and respective controls). Next, 1 μl of each reaction and control were added to the cells and gently mixed by lightly tapping the tube. The cells then were incubated on ice for 30 min, followed by a 60 sec heat shock at 37 ºC, and placed back on ice for 2 min. After the incubation on ice, 400 μl S.O.C. medium at RT (2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, and 20 mM glucose) was then added, and the mix incubated for 1 h at 37 ºC. Lastly, 100 μl samples were plated on a series of LB-ampicillin plates containing X-Gal and IPTG, and the plates were incubated at 37 ºC overnight. The ampicillin in the plates was used to select for plasmid transformants, whereas the X-Gal in the plates was used for identification of insert-containing plasmids by the absence of α-complementation. This takes advantage of the presence of the lacZα gene on pUC19 that allows the colonies to digest the X-Gal releasing a chromophore and therefore
causing the colony to appear blue. Since the *BamHI* site in pUC19 is located within the *lacZα* gene (Figure 2), insertion of φ80 DNA should disrupt the function of *lacZα* and cause the colonies to remain white.

**Library screening.** Screening plasmids for extracts was achieved by growing selected colonies in 1.5 ml LB ampicillin broth. Then plasmids from isolated colonies were extracted using an alkaline lysis method to recover the plasmids (Sambrook *et al.*, 2001). This was done by centrifuging 1.5 ml of the *E. coli* saturated overnight culture in a microcentrifuge at 14,000 x g for 5 min and then aspirating the supernatant. From there, the pellet was resuspended in 100 μl of ice cold solution I (50 mM Glucose, 25 mM Tris pH 8.0, and 5 mM EDTA). Next 200 μl of solution II (0.2 N NaOH, 1 % SDS) was added to the suspension, gently mixed by inversion of the tube, and incubated for 5 minutes on ice. After the incubation, 150 μl of solution III (3 M K, 5 M acetate) was added and mixed by inversion again. The addition of these three solutions was to break down the cellular membranes and denature the genomic DNA, leaving the plasmid in solution. After the three solutions were added, the samples were placed on ice for 5 min then centrifuged at 14,000 x g in a microcentrifuge for 5 min. Next the supernatant containing the plasmid DNA was transferred to new microfuge tubes. Then 200 μl Tris•HCl saturated phenol and 200 μl chloroform:isoamyl alcohol (24:1) was added to separate the plasmids from the proteins. From here, the aqueous phase, about 375 μl of solution, was collected and precipitated by the addition of 1000 μl 100 % ethanol. The solution was mixed, incubated at RT for 5 min, and centrifuged in a microcentrifuge at 14,000 x g for 5 min. After centrifugation, the supernatant was aspirated with precautions to not disturb the pellet. Then the plasmid DNA was rinsed with 1ml 70 %
ethanol and centrifuged again at 14,000 x g for 1 min. Next, the alcohol was aspirated off and the sample air dried. Lastly, the pellet was solublized in 40 μl dH₂O and stored at -20 ºC.

Once the plasmids were recovered from the cells, they were digested with both EcoRI and HindIII restriction endonucleases to excise the insert from the vector. This step was achieved by mixing 33 μl 10 x buffer 2 (100 mM Tris-HCl, 100 mM MgCl₂, 500 mM NaCl, 10 mM dithiothreitol; New England Biolabs, Ipswich, MA) with 126 μl dH₂O, 30 units of EcoRI, and 30 units of HindIII. From here the mixture was divided among 15 microfuge tubes with 10 μl of mixture in each. Next, 10 μl of recovered plasmid was added to each tube and the tubes were incubated at 37 ºC for 90 min. After this initial incubation, 1 μl of a 2 μg/ml RNAse A solution was added to each test tube to hydrolyze contaminating RNA and the tubes were subsequently incubated at 37 ºC for 5 min. Finally, 4 μl of gel loading buffer (0.0254 g bromophenol blue, 0.0254 g xylene cyanol, 30 % glycerol in ddH₂O/ddH₂O to 100 ml) was added to the 21 μl of the EcoRI and HindIII restricted plasmid preparation and the sample electrophoresed through a 1 % agarose gel.

**Sequence determination.** Recovered plasmids that were chosen to have their inserts sequenced were sent to the University of Iowa DNA facility. Initial sequencing reactions used the M13/pUC -20 and -24 reverse primers. Isolates whose sequenced regions did not completely overlap were sequenced further using primers synthesized based on the previously generated sequence (Table 1).
Table 1. List of primers used for sequencing.

<table>
<thead>
<tr>
<th>Primer Number</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>209</td>
<td>5’ CGG CGT ATC GTT GAG G 3’</td>
</tr>
<tr>
<td>210</td>
<td>5’ GAC GCT GAT CGT CGT 3’</td>
</tr>
<tr>
<td>212</td>
<td>5’ GGG CGA TAG ACA GTG 3’</td>
</tr>
<tr>
<td>213</td>
<td>5’ GTC GTC GTT AGC TGC 3’</td>
</tr>
<tr>
<td>219</td>
<td>5’ CAA CTT CTC CAT GGT GAA AAC 3’</td>
</tr>
<tr>
<td>229</td>
<td>5’ GCA TCA AGC TCA GCC 3’</td>
</tr>
<tr>
<td>230</td>
<td>5’ GTC AAC ATC GTC CTG 3’</td>
</tr>
<tr>
<td>231</td>
<td>5’ GGC GGT CAA TGG CCG 3’</td>
</tr>
<tr>
<td>232</td>
<td>5’ GGC AGC TAT GAC GTG 3’</td>
</tr>
<tr>
<td>234</td>
<td>5’ GGG TGC AGG TTG GCC 3’</td>
</tr>
<tr>
<td>238</td>
<td>5’ GGT CTA GAT CAG AAT ACT CCC GTT AAT CGG 3’</td>
</tr>
<tr>
<td>240</td>
<td>5’ CTC CCT CAG AAA CGG 3’</td>
</tr>
<tr>
<td>241</td>
<td>5’ CAG GAA AAC GCT GAC 3’</td>
</tr>
</tbody>
</table>

Polymerase chain reaction (PCR). To amplify the φ80 open reading frame (ORF) that is homologous to λ’s J protein, a primer pair representing the 5’ and 3’ end of the open reading frame were synthesized (5’ CAACTTCTCCATG GTGAAACC 3’ and 5’ GTCTAGATCAGAATACTCCCGTTAATCGG 3’). The amplification reaction mix consisted of 1 μl template φ80 genomic DNA (460 μg) or 1 μl of dH2O for a negative control, 5 μl 10 x Thermopol buffer (100 mM NaCl, 100 mM (NH₄)₂SO₄, 200 mM Tris-HCl pH 8.8, 20 mM MgSO₄, and 1 % Triton X-100; New England Biolabs, Ipswich, MA), 50 pmol 5’ primer, 50 pmol 3’ primer, 1 μl 200 μM mix of each dNTP, 1 μl of 2 U/μl Deep Vent Polymerase and 29 μl dH₂O. To determine the optimal magnesium ion content, three tests were set up: 10 μl dH₂O, 5 μl of 10 mM MgSO₄ and 5 μl dH₂O, and 10 μl of 10 mM MgSO₄. In addition, 5 μl of dH₂O and 10 mM MgSO₄ was added to the
negative control. For the reaction, a cycle of temperatures was set to allow for the appropriate activity of DNA polymerase to effectively replicate the open reading frame. The first step was used to denature the DNA and was set for 94 °C for 30 sec. Next, a step of 62 °C (30 sec) was used to allow the primers to anneal to the single stranded DNA. A step at 72 °C (4 min) was used to allow the polymerase to attach and amplify the DNA between the primers. In all, these three steps were cycled 35 times to allow for adequate DNA replication, followed by a final step at 72 °C for 5 min. Products then were evaluated by restriction and electrophoresis through 1 % agarose gels as described.

Results

**Preliminary screening of φ80 plasmid library.** The decision to use Sau3AI was based on the ability of Sau3AI to recognize a 4 base pair motif. Thus, the enzyme would cut stretches of random DNA on average every 256 base pairs. Partially digesting with Sau3AI would allow for some sites to be cut while others remain uncut (Figure 3), producing an overlapping library. This would be helpful when assembling the sequences. For efficiency, we chose to initially characterize only those clones that contained inserts of at least 1000 bp in length. To screen inserts for size, digestions were performed using EcoRI and HindIII, sites that flanks the BamHI cloning site in pUC19 (Figure 2).

Because preliminary restriction mapping identified 9 EcoRI and 7 HindIII sites in φ80vir (Larsen, unpublished data), we expected some digests to result in multiple bands.

Of the 590 plasmids screened, 72 were found to have inserts of 1 kb or longer. Many of the 1000 bp or longer fragments contained multiple cut sites from EcoRI, HindIII, or both (Table 3). From the data generated from screening plasmids with EcoRI
and *HindIII*, two clones were chosen for sequence determination, isolates 103 and 341 (data not shown).

**Table 2. Fragment sizes of φ80 DNA clones.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Band Size Produced</th>
<th>EcoRI/HindIII</th>
<th>Kpn I</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>750, 1400</td>
<td>600, 1000, 1200, 1500</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>800, 1000</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1500</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>300, 700, 1100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>300, 800, 1000, 1200, 2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>300, 1500, 2000, 3000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>300, 1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>600, 800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>750, 900</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>103</em></td>
<td>800, 1300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>300, 700, 900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>1300</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>700, 1000, 1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>188</td>
<td>500, 800, 1300, 1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>189</td>
<td>300, 800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>193</td>
<td>800, 1500</td>
<td>300, 400, 600</td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>1200</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>800, 1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>1000, 1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>800, 900, 1500, 2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>233</td>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>236</td>
<td>1000</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>600, 1000, 1200, 1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>242</td>
<td>2000</td>
<td>300, 400, 600, 1000, 1200, 1500</td>
<td></td>
</tr>
<tr>
<td>247</td>
<td>800, 2000</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>249</td>
<td>800, 2100</td>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>251</td>
<td>700, 900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>265</td>
<td>1200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>268</td>
<td>1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td>1200, 1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>303</td>
<td>700, 1200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Sequencing of initial isolates.** Plasmids from isolates 103 and 341 were submitted for sequence analysis using the M13/pUC -20 and -24 reverse primers, corresponding to sequences that flank the inserts (Figure 2). Because of the size of the inserts, this did not provide for sequence determination of the central portion of the inserts. Therefore, new primers were synthesized based on the sequences obtained (Table 1). Once overlaps were obtained, additional primers were synthesized and sequencing reactions performed to yield the complete nucleotide sequences of both strands.

The first clone that was sequenced was isolate 103, which had an insert of 2,163 bp in length. DNA sequence analysis revealed that the φ80 DNA present had no internal EcoRI restriction sites. However, one HindIII restriction site was located at position 1709 bp (Figure 4). Initial Basic Local Alignment Search Tool (BLAST) searches at National Center for Biotechnology Information (NCBI) using a nucleotide query
indicated a relationship to λ tail proteins (Altschul, 1991; Wheeler et al., 2005). The predicted probability value of this sequence being unrelated to that of the λ J gene was very low (1 to 8 x 10^{11}, data not shown).

Figure 4. Relative position of φ80 clone 103 DNA to the known map of phage λ. The sequence of 103 bore similarity to the λ genes J, I, and K, as shown (Hendrix and Casjens, 2006). The sequence is 2163 bp in length and contains one internal HindIII restriction site as indicated.

The second clone that was sequenced was isolate 341 (Table 2), having an insert of 1,616 bp in length. Sequence analysis identified an Ncol site at position 1021 bp, a PstI site at position 1239 bp, a KpnI site at position 1285 bp, and an EcoRI site at position 1606 bp (Figure 5). This restriction map provided for the use of other restriction endonucleases to screen for overlapping clones. Using the blastx function at NCBI, the search results revealed that the sequences bore a striking resemblance to the λ J gene (NP_718509.1), in this case the probability that the sequences were unrelated was 3 x 10^{-81} (data not shown).
Figure 5. Relative position of φ80 clone 341 DNA to the known map of phage λ. The sequence of 341 bore similarity to the λ gene J as shown (Hendrix and Casjens, 2006). The sequence is 1616 bp in length and contains internal EcoRI, KpnI, NcoI, and PstI restriction sites as indicated.

Identification of additional isolates of interest. Since the φ80 DNA contained on the clones did not overlap and did not encompass the entire coding sequence of the φ80 host attachment spike protein, additional measures were taken to identify clones that would overlap with isolate 341. Because isolate 341 had a KpnI site and total genome digestions suggested only six such sites occur in the φ80 genome (Larsen, unpublished data), the presence of KpnI sites among other isolates would be a good initial indicator of potentially overlapping clones. Thus, all 72 isolated clones of 1 kb or longer were digested with KpnI and revealed 15 isolates with KpnI restriction sites (Table 2). Since isolate 341 contains a PstI restriction site located 46 bp away from the KpnI site, the 15 isolates with KpnI sites were further restricted with PstI. Amazingly, one clone, isolate 563, contained both KpnI and PstI restriction sites separated by approximately 50 bp, suggesting an overlapping clone.
The inserted DNA in clone 563 was isolated and sequenced which revealed an insert of 3539 bp. Not only did isolate 563 contain the same DNA sequence in clone 341, but the fragment also contained overlapping segments on both ends. These overlapping segments contained a portion of sequence from clone 103 and completed the sequence for the φ80 host attachment spike homologous to λ’s J gene (Figure 6). Not all of the genes identified in the φ80 sequence outside of the host attachment spike protein were observed to be homologous to genes flanking J in λ. While the gene flanking J on one side had homology to the λ I gene, the gene on the opposite side of J was identified as the φ80 cor gene instead of the expected lom gene of λ.

![Figure 6: Relative position of φ80 clone 563 DNA to the known map of phage λ.](image)

**Final Assembly.** After determining that clones shared overlapping regions, the sequences of the clones were assembled *in silico*. Areas of ambiguity were resequenced using primers synthesized from the obtained sequence and used to directly sequence φ80
genomic DNA. Once fully sequenced, the translation tool available at ExPASY was used to identify open reading frames within the sequence (http://ca.expasy.org/tools/blast/; Altschul 1991; Gasteiger, 2003). A complete open reading frame was discovered starting in clone 103 and ending in clone 563, predicting a continuous coding region 3,840 bp in length (Figure 7). Predicted translation of this open reading frame indicates a putative protein of 1,279 amino acids. Initial BLAST searches using NCBI’s blastx program identified bacteriophage host specificity proteins, and most importantly the λ J gene with an expected value of less than $6 \times 10^{-178}$ was found (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi; Altschul 1991). Clearly, this expected value suggests that the similarity of the products of these two genes did not happen by chance.

```plaintext
ATGCAACTTCTCCATGGTAACCATCATACAGGGTGCAAAAGGGCGTGGAAGCGCGCG
MQLHGETIIQGAKGGGGASA
CATACTCCGGTGTGACCACTGAGATCTGCTGTCGTCGAAGATTAAATATGGCTCATT
HOTPVEQPDLLLSVAKLMILIGCCGTTTCTAGGAGAAATACACGGCGAGCTAGCCAAATATTTCTCAACGAT
AVSEGIELQGDLTAQNIPLND
ACGCGCTGGCAAACGACAGCGGAGTATAAACACTCACGCGGTCGGAAGAAGCTTCCGT
TPLANDSGEYNFSGVKEFWRAAGGGCAACAGGACCAGACCATATTGCCGGGATGCACCAAGTGACTGCG
KGTQDQTYIAAGMPQVDNELLAGTTGCGCAACCTGCACCCACCCGCGCTGCGACCCAGTTTACAACTTCTCCCGT
VGTDVVTTPAWTROFNTNSL
GATGTCATCCGATCAAGTCCACCTCCGCTGAGATCTCTTCTAAAGGAAAGGCGTGAT
DAIRIKLSLPLVPQLYKDNGD
ATGGGCGAGCGGCTACCGAGATTCGCTAGATTTATTACACGGACGCGCGCCTGAAA
MVGTVEYAIDLSTDDGAWK
ACCGTGTAAACGCGAAATTTGACCGAAAGACACACGAGAATATCAGCTGACCCACCGT
TVVNGKFPDKTTEYYQRDRHR
ATCGACTCGCAAATCCACGTCGGCTACTGACGGGTATTACGCGTAT
IDLPKSTSGWSVRVRRITAD
GCCAAGGTACAAATCGGTTAACCCCTTACGCTGGAAGTTTACCGGAAGTGACGC
ASGSNSKLVNAPKVFSPYAEV
ATCGACACCAAGCTCAGTTATCCTTAAACCAGCTCTCTGTATGCGAAATGGAGACGCG
IDSKRLYPLTALLLYVEVDSS
CAGTGTCGAAAGTGACCTGTAAGGATAAAGGCAAGCTGATTAAGGTT
QFNPSAPKVTKCKIKGKLIKVK
CGCGTTAATACAGATCGATAACCCGAATCTATTACGTCTGCGTCCGCGGTTTCAA
PDNYDPITRTYSGSWSGGFK
ATGGCGCTGTCGAAATCCCGCCCTGATTTACGCTGTTTCTGATGAAATTTAC
MAWSNPAWIFYDLVDLEIY
GGCATGACCGCGCGCCTGAGTGACCTCACATTGGAAGTTGCGCTATATTACATC
GMGTRVDSMAVMDKWALYISA
CATGACTGTCGAAATGTTCCGCAAGCGCGGCGTGGACCAACCCGGCTTACCTGC
QYCDEMVSDGAGGTEPRFTC
```
cagccatcggcgacaaaaaattaacaccctgactgttaaatctggcgaaaacagtgccgcg
Q A S A T K I N T L T V K S G E N S A A
ataaatgtcaacgcacaggtatcgcagatgtaaacctaaagtgcagatgtaacac
I N V N A Q A I A D V N G N L S A M Y N
atcaaggtgtgtcttcagcaatgacagatattacgcgccggatggtatctgcgccg
tK V G V S S N G Q Y Y A A G M G I G V
gagaatacgcacctcgcagcacgtagcagttctctctgtgtagctgcgcctgctgc
E N T P S G M Q S Q V I F L A D R F A V
accacgcagccggaataagctagtacattgctgctggtatccagaaacgggcagacattc
t T A A G N S V A L P F V I Q N G Q T F
atccggccagcttcctaatccagacgcacactataagcaacgaaagttgtaattttatc
I R A S F I Q D G T I S N A K I G N F I
cagtcgaaacattatgttgctgctctgtgctgctgagcttgataaaaggggacgttt
Q S N N Y V A G S A G W K L D K G G T F
gagaactacggtgtagcgcgctgctgctgctgctgagcctgatgataaaggggacgttt
E N Y G S T A G E G A M K L T N Q T I S
gtcaaaagatggcagtaatgttctttagggtcaggtttgggagattaacgggagtagtattctga
V K D G S N V L R V Q V G R L T G V F Stop

Figure 7. Assembled φ80 open reading frame encoding the host attachment spike. Shown is the determined sequence of φ80’s host attachment spike. The amino acid translation of the open reading frame is shown below the DNA sequence.

After sequencing and assembling the data, primers were developed to PCR amplify the open reading frame containing the φ80 host attachment spike from genomic φ80 DNA. Restriction mapping of the PCR product with previously used endonucleases was used to verify the assembly of the sequencing data (Figure 8). The restriction digest visually shows the predicted fragments generated by EcoRI cutting at 2718 bp, KpnI cutting at 2397 bp, and PstI cutting at 2352 bp. HindII was predicted to cut in two locations on the predicted sequence, at positions 670 bp and 3337 bp. Additionally, NcoI is predicted to cut in 3 locations: at positions 12 bp, 926 bp, and 2134 bp. Using the φ80 genome as a template for sequencing and PCR, it was determined that the predicted 3,840 bp stretch of DNA is present within the φ80 genome.
Figure 8. Restriction analysis of PCR amplified φ80 host attachment spike DNA. The first lane is the size standard, indicated in kbp at the left of the panel. The second lane is the undigested φ80 host attachment spike gene, the third lane is EcoRI digested φ80 host attachment spike gene, the fourth is the host attachment spike gene digested with HindIII, fifth is the φ80 host attachment spike gene digested with KpnI, sixth is the φ80 host attachment spike digested with NcoI, and the last lane is the φ80 host attachment spike digested with PstI. Conceptual map of the φ80 host attachment spike is presented to the right with the number of the primer used to sequence a given region of the DNA sequence. Primer sequences are listed in Table 2.
CHAPTER THREE

AIM TWO: IDENTIFYING FUNCTIONAL DOMAINS ON THE φ80 J HOMOLOG

Introduction

One of the more recent techniques for researching novel proteins is the application of proteomic approaches in silico. This entails querying a novel DNA or amino acid sequence against a database of previously described DNA sequences or proteins. The advantage of using in silico approaches to research proteins is to help classify an unidentified protein based on previously described models. Using this method, not only can one predict the function of a protein, but also identify potential functional domains. Two of the main databases used in this paper are from NCBI and ExPASY.

Materials and Methods

To obtain search results, the DNA sequence of the φ80 host attachment spike was first translated into an amino acid sequence using the DNA translation tool found at http://ca.expasy.org/tools/dna.html. The amino acid sequence was then used in BLAST searches against other proteins in the ExPASY database. Results with an expected value of less than 10^{-129}, lambdoid phage, and phage that are known to use either FhuA or TonB for entry were recorded. Additionally, BLAST searches using the database at NCBI were performed for comparisons.

The alignment of highly similar amino acid sequences was accomplished using two programs, ClustalW (Larken et al., 2007) and the PRALINE (Simossis and Heringa, 2005) multiple sequence alignment tool (http://130.37.199.5/programs/pralinewww/index.php). The results were very similar
using both programs. However, results generated using PRALINE were more readily interpreted visually. Therefore, this tool was primarily used in my analyses.

Viewing the FhuA protein crystal structure was done using the Swiss Protein Data Base (PDB) Viewer associated with ExPASY. The Swiss PDB viewer was downloaded at http://ca.expasy.org/spdbv/. Additionally, the protein crystal structures that can be viewed with this program were found at http://swissmodel.expasy.org/. Finding the crystal structure of interest was accomplished by querying the database mentioned above for the FhuA gene of E. coli. This structure is based on the original crystallographic analyses (Locher et al., 1998).

Results

The 3,480 base pair section sequenced directly from the φ80 genome encodes a polypeptide with an amino acid sequence that is homologous to the λ phage’s J protein. As stated previously, the J gene in λ functions primarily as a host attachment spike (Buchwald and Siminovitch, 1969). Other proteins with similarity to φ80 J protein include Gp24, tail fiber, and Gp21, proteins of bacteriophage HK022, HK97, and N15 respectively. The expected value scores for these proteins were all 9.5 x 10^{-57} or less, signifying that chance similarities between φ80 J and these proteins is extremely improbable. Since the predicted product of the open reading frame in φ80 showed a great deal of similarity to the J protein of λ and the other similar proteins noted above, it is reasonable to conclude that it has a similar function in that it adsorbs to the host cell. In addition, this open reading frame showed similarity to the attachment proteins of other related phage. Most notable are the λ phage and those phage that utilize the FhuA protein for access into the host cell.
An alignment of the amino acid sequences for the φ80 and HK022 tail spike proteins against each other generated by PRALINE (http://zeus.cs.vu.nl/programs/pralinewww/index.php), revealed 61 % sequence identity overall. More strikingly, the last 144 amino acids in the two sequences were entirely identical (Figure 9), which would suggest that they share common protein structure and in turn similar function within that domain of the tail spike. Since both φ80 and HK022 use FhuA and infect the same host range it was not too surprising that the two phage share very similar amino acid sequences in their tail spike proteins. Also, this suggested a common and recent origin. A recent origin would signify that there was a recombination event in the relatively recent evolutionary timeline of phage.

The J protein of φ80 and λ were less conserved, with 47 % sequence identity, and the greatest sequence similarity at the amino and carboxyl terminal regions (Figure 9 and Figure 10). This was consistent with the fact that φ80 and λ use different host receptors (LamB vs. FhuA). Most of the similarity between these two tail spike proteins can probably be attributed to a common lambdoid phage ancestor or from countless generations of recombination between the lambdoid phage, resulting in mosaicism.

Bacteriophage T1 had 44 % sequence identity with φ80’s tail spike protein even though the phage is both FhuA and TonB dependent, like φ80. The majority of this sequence similarity was confined to two locations within the proteins, the amino terminus region and the last 100 amino acids of the carboxyl terminus (Figure11). Another phage, T5, also uses FhuA as a route of entry into bacteria but does not require TonB. The T5 J protein had 20 % sequence identity with φ80’s J protein. As was true of the T1 phage, the majority of the amino acid identities between φ80 and T5 proteins occurred within the
carboxyl terminus (Figure 9). Thus, it would seem reasonable to propose that these phage are less similar to $\phi 80$ since they are not part of the lambdoid phage family and have evolved separately. But regions of similarity may indicate sites of interaction with FhuA (and TonB).

**Figure 9. Alignment of the first 150 amino acids of $\phi 80$ J protein with the amino terminus of similar phage attachment spikes.** The alignment was generated using PRALINE. The following accession numbers were used to generate this sequence alignment: NP_037684.1, NP_037717.1, NP_046916.1, NP_718509.1, YP_003912, and AAP75892. $\phi 80$, HK022, HK97, N15 and $\lambda$ are all in the lambdoid phage group. Of these $\phi 80$, HK022, HK97, are known to require both FhuA and TonB, whereas N15 is only known to require FhuA for adsorption. On the other hand, $\lambda$, requires LamB for adsorption. Phage T1 and T5 are not in the lambdoid phage group but both require FhuA for adsorption. Additionally, T1 also requires the presence of TonB. The outlined amino acids from 103-111 are proposed to be a TonB box motif. Warmer colors of reds and orange represent amino acids with greater similarity, whereas cooler colors of blues and greens represent less similar amino acids. The consistency gives a numerical value from 1 to 10, with 10 represented as a *, based on the similarity of the amino acids in that alignment position.
Figure 10. An alignment of the carboxyl terminus of the φ80, HK022, T1, Lambda and T5 host attachment spikes, generated using PRALINE. The alignment was generated using PRALINE. The following accession numbers were used to generate this sequence alignment: NP_037684.1, NP_718509.1, YP_003912, and AAP75892. As per Figure 9, the warmer colors of reds and oranges represent amino acids of greater similarity, whereas cooler colors of blues and greens represent less similar amino acids. The consistency gives a numerical value from 1 to 10, with 10 represented as a *, based on the similarity of the amino acids in that alignment position. Notably, the last 144 amino acids of the φ80 and HK022 sequences are identical.
Figure 11. Percentage of amino acid similarity mapped on the $\phi$80, $\lambda$ and T1 predicted J protein sequences. Using sequence alignment data from PRALINE and ClustalW, a relative map of amino acid similarity was made between $\phi$80 versus $\lambda$ and between $\phi$80 versus T1. Clearly, the amino terminals and carboxyl terminals show the greatest percent similarity. Accession numbers used were NP_718509.1 and YP_003912 for $\lambda$ and T1 respectfully.

As previously discussed, $\phi$80 requires the FhuA protein to reversibly adsorb to its host cell. Although it has not been shown directly to which amino acid $\phi$80 binds, studies have identified the amino acids on FhuA that are critical to $\phi$80’s adsorption. For FhuA to bind ligand efficiently, the aspartate at amino acid position 348 must be present. Without this aspartate residue, *E. coli* becomes resistant to $\phi$80 (Killman and Braun, 1992). However, this aspartate residue is on the outer surface of the outer membrane. Even though this amino acid is critical for function, it cannot span the outer membrane alone. Thus, Killman and Braun proposed the need for more areas of contact between $\phi$80 and FhuA required for adsorption of the phage (1992). They identified the following additional residues GVCSDPANAY (316-325), ALAPADKGYL (331-341), and VDDEKLQ (347-353). This was achieved by a technique called competitive peptide binding which uses short peptide sequences to bind to and inhibit specific binding
domains of a protein (Killman et al., 1995). However, one flaw that becomes apparent in Killman’s work is the locations of these peptides. Killman et al. proposed that these regions of FhuA reside entirely in an outer surface loop domain. The crystal structure data (Figure 12) clearly shows that these segments of the protein not only occur in the loop domain, but also extend into the β-barrel of FhuA, which suggests that the φ80 tail spike would be able to span the outer membrane. The crystal structure of the FhuA protein (Locher et al., 1998; P06971) was viewed in stereo and labeled by using the Swiss PDB Viewer (Peitsch, 1995; Guex and Peitsch, 1997; Schwede et al., 2003; Kopp and Schwede, 2004; Arnold et al., 2006).
How then does the $\phi 80$ tail spike interact with the FhuA protein? More specifically, does $\phi 80$’s tail spike transverse the outer membrane through the pore of FhuA or does it transverse the outer membrane by binding to the outside of FhuA?

Evidence for $\phi 80$ being able to transverse the outer membrane along the outside of the
FhuA protein is supported by a study conducted by Bohm and her colleagues’ on the T5 phage using cryo-electron tomography (2001). This demonstrated that T5 releases its genome into vesicles, even double membrane vesicles, by binding to the outside of FhuA and injecting its DNA like a syringe (Bohm, 2001). However, T5’s tail spike only showed 20% sequence identity and is not TonB dependent like φ80. Therefore, this model of adsorption to the host cell is probably not the most likely to occur.

There is other evidence that φ80 binds to the inner surfaces of FhuA β-barrel domain. For instance, the addition of ferrichrome to nutrient media specifically inhibits φ80 infection (Wayne, 1975). It becomes obvious that if φ80 must pass through the center of FhuA’s pore, then it would become considerably more difficult if other ligands were competing for the same entrance. If ferrichrome binds to FhuA first, then the phage cannot enter until the receptor becomes available again. Further experiments constructed deletion mutants of FhuA and showed that φ80 can infect a deletion mutant in which the FhuA cork domain is absent as long as this deletion mutation retains FhuA’s TonB box region, a region to interact with TonB (Killman et al., 1996; Braun, 1999). The cork domain of FhuA is on the amino terminal of FhuA and forms a globular domain inside the β-barrel. This is important because the cork domain acts like a gate to restrict the flow of incoming molecules.

Further justification for the φ80 J protein binding to the inner surfaces of the β-barrel of FhuA comes from experiments using FepA, a paralogous protein to FhuA. These studies showed that the cork domain is absolutely essential for binding of ligands, interacting with TonB, and translocation of the ligand to the periplasm (Vakharia, 2002). If φ80 does interact with the FhuA residues suggested by Killman et al., (1996) then it
must bind to an inner surface of the outer loop domain (Figure 12). The inner side of this loop region becomes easier to access to the phage tail spike if it were to adsorb to the host by binding on the inner walls of the FhuA β-barrel. Additionally, the residues suggested to bind φ80’s tail spike would be topologically inaccessible should the tail spike bind on the outside of the β-barrel.

More evidence for φ80 binding to the inside of the receptor comes from considerations of the λ structural homolog. First, it has been shown that the last 249 bp of the λ J gene interacts with the LamB trimer (Wang et al., 2000). Secondly, the λ J protein has been shown to bind to specific amino acid residues in LamB: 18, 148, 151, 152, 163, 164, 245, 247, 249, 250, 259, and 382 (Charbit et al., 1994; Charbit et al., 1984; Charbit et al., 1988; Clément, 1983). Not too surprisingly, all of these amino acid residues reside on the inner portion of the external loop regions of the LamB receptor.

In addition to φ80’s requirement for FhuA, the phage also needs a functional TonB system. Since the TonB system aids in providing energy to the outer membrane, it is reasonable to conclude that this energy is required by the phage to inject its genome into the host. Also, it is reasonable to suggest that since there is a requirement for TonB, that φ80’s tail spike would be able to directly interact with TonB. The interaction between an outer membrane protein and TonB is via an amino acid sequence known as a TonB box. One technique that was employed to find a potential TonB box within the φ80 tail spike was to align peptide sequences of previously shown TonB boxes to the φ80 tail spike amino acid sequence. Both PRALINE and ClustalW programs were used to align the sequences. As shown in Figure 9, a TonB box predicted within the φ80 J protein has the amino acid sequence ELAVGTTVT at positions 97-105. If this does indeed
constitute a TonB box, the φ80 tail spike protein would be capable of directly binding to the TonB protein, which would account for the ability of φ80 to irreversibly bind to its host when TonB is present.
CHAPTER FOUR

FUTURE RESEARCH DIRECTIONS

Even though this study on the φ80’s adsorption answers many questions, ultimately, even more questions arise. Fortunately, the data from this study will be able to help answer the questions in studies to come. More specifically, experiments generated from proteomics data will be able to empirically prove whether the proposed interactions between the phage and host proteins occur.

Questions pertaining to φ80’s relationships with other phage in the lambdoid phage group would be specifically helpful in being able to determine which phage most likely resemble the model lambdoid phage. Even though λ was discovered first and is regarded as the model phage for the group, should it really be the model when it seems to be more distantly related to other phage like φ80, HK022 and HK97? In order to accomplish research in this direction, more of the φ80 genome would have to be sequenced. Genomic comparisons between φ80, HK022 and HK97 would prove useful in understanding the relationships among lambdoid phage since HK022 and HK97 have already been sequenced. Since a φ80 genomic library has been generated with this study, further restriction mapping, sequencing and assembly of sequences can readily take place. Proteomic studies to follow would be based on the sequence data generated and the phylogenetic relations can be better placed together after the proteomic data is obtained.

Another great opportunity would be to test the interactions of FhuA and φ80’s tail spike further. This would test for φ80’s requirement to bind on the inside of FhuA pore as opposed to along the outside of the protein. Methods such as competitive peptide
binding and site directed mutagenesis would likely be the most helpful experiments in this regard. By using proteins that bind to the outside of the FhuA β-barrel, where the tail spike is predicted to bind, and introducing them into wild type cells one can limit access to the outside of the FhuA protein. For example, a purified T5 tail spike, which is known to bind FhuA on the outside of the FhuA β-barrel, could be used in combination with wild type FhuA to determine if φ80 infection is inhibited. Additionally, point mutations can precisely knock out key amino acids to prevent binding.

Another question that could be asked is whether or not the φ80 tail spike has a TonB box in the predicted region of the protein. One method that could be potentially useful in this study would be to apply site directed mutagenesis to the φ80 DNA corresponding to the predicted TonB box. Either completely deleting the nucleotides coding for the TonB box in the protein or introducing mutations that destroy the predicted function of the domain would work equally well. Assays of phage viability using the mutant phage can determine whether the TonB box is in the predicted region of the phage tail spike. In addition, sequencing φ80 mutants that no longer require TonB may prove to be useful to these studies in the interaction with TonB and other surface receptor proteins.
References


