GENES AND STRUCTURAL PROTEINS OF THE PHAGE SYN5 OF THE MARINE CYANOBACTERIA, *SYNECHOCOCCUS*

by

WELKIN HAZEL POPE

S.B. Biology, Massachusetts Institute of Technology, 2000

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in Biology
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

September 2005

Signature of Author _____
Joint Program in Biological Oceanography
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
September 2005

Certified by _____
Jonathan King
Thesis Supervisor

Accepted by _____
John Waterbury
Chair, Joint Committee for Biological Oceanography
Woods Hole Oceanographic Institution

GENES AND STRUCTURAL PROTEINS OF THE PHAGE SYN5 OF THE MARINE CYANOBACTERIA, SYNECHOCOCCUS

By Welkin Hazel Pope

Submitted to the Departments of Biology of the Massachusetts Institute of Technology and the Woods Hole Oceanographic Institution on August 29, 2005 in partial requirement for the degree of Doctor of Philosophy in Biology

ABSTRACT

Bacteriophage have been proposed to be the most abundant organisms on the planet, at an estimated 10^{31} particles globally (Hendrix et al., 1999). The majority of bacteriophage isolates (96%) are double-stranded DNA tailed phages (Caudovirales). These phages possess a distinctive icosehedral head, with a protein tail structure protruding from a single vertex. This organelle determines host specificity and provides the mechanism of passage of the phage genome into the host cell. Phages infecting differing microbial hosts may have access to a global pool of genes, albeit at different levels.

Marine cyanobacteria of the genera *Prochlorococcus* and *Synechococcus* are numerically dominant photosynthetic cells in the large oligotrophic gyres of the open oceans, and contribute an estimated 30% to the oceanic photosynthetic budget. Cyanophages have been isolated which propagate on many strains of *Synechococcus* and *Prochlorococcus*. Cyanophages can effect community structure and succession through lytic infection of their hosts, and have implications in lateral gene transfer, mediated through lysogeny, mixed infections, pseudolysogeny, and transduction. The broad host ranges (between genera) observed in some phages indicates that lateral gene transfer is not confined to cells of the same strain. These phage/host interactions begin by host recognition by the tail of the infecting phage. Few studies have examined the structural proteins of cyanophage, partially due to the lack of a robust protocol for the growth and purification of phage particles.

Cyanophage Syn5 is a short-tailed phage isolated from the Sargasso Sea by Waterbury and Valois (1993) which infects *Synechococcus* strain WH8109. Methods of growing the host cells and the phage, and concentrating the phage by PEG precipitation were developed. These methods led to highly concentrated purified phage stocks, to titers of 10^{12} particles/ml. Preliminary characterization of the growth of Syn5 gave a burst size of approximately 30 phage/cell and a lytic period of approximately 10 hours when inoculated into exponentially growing host cells acclimated to a temperature of 26°C and a light intensity of 50μE m^{-2} s^{-1}. Isolation of the phage nucleic acid yielded dsDNA molecules of approximately 40kb. The Syn5 particles were comprised of twelve structural proteins, as determined by SDS-PAGE. The most intense band on the gel was assigned to the capsid protein of Syn5 (~35kDa). However, it was not possible to distinguish putative tail proteins via this method.

Purified Syn5 particles were sent to the Pittsburgh Bacteriophage Institute for genome sequencing. The completed Syn5 genome was 46,214 bp long with a 237bp terminal repeat. Annotation of the completed Syn5 genome identified 61 putative ORFs,

and revealed that Syn5 appeared closely related to the enteric phage T7 and cyanophages P-SSP7 and P60, as determined by gene similarity and synteny, although the genome was ~10kb longer than T7. Syn5 appeared to possess a more extensive DNA replisome that T7, containing copies of genes that encoded proteins of known T7 host co-factors, such as thioredoxin, utilized by the T7 DNAP. Several large ORFs were identified between the gene encoding the putative tail fiber and the gene encoding the putative terminase. These ORFs encoded proteins similar to some fibrous sequences within the NCBI non-redundant (nr) gene sequence database as of March, 2005; but had unknown functions within the phage. Unlike other recently sequenced cyanophages, Syn5 did not contain any photosynthetic genes.

The structural proteins of Syn5, as visualized by SDS-PAGE, were characterized by mass-spectroscopy and N-terminal sequencing. This allowed the assignment of sequences to putative ORFs within the Syn5 genome. The Syn5 particle was comprised of eleven discreet protein chains of molecular weight 152kDa, 139kDa, 99kDa, 90kDa, 66kDa, 60kDa, 47kDa, 35kDa, 22kDa, 21kDa, and 16kDa. The identified proteins included the portal, capsid, two tail tube proteins, and three internal virion proteins. Each of the genes encoding these proteins were found in the same gene order in the Syn5 genome as the corresponding genes were ordered in the T7 genome. There were three unidentifiable proteins within the particle (66kDa, 47kDa, and 16kDa). These mapped to the area of the Syn5 genome between the gene encoding the putative tail fiber and the gene encoding the putative terminase. No minor capsid or decorative capsid proteins were detected. The copy numbers of the corresponding protein chains were similar to those known for T7, with the exception of the tail fiber, which was present at a number of three chains per particle in comparison to T7's eighteen per particle. Polyclonal antibodies were raised against Syn5 particles. A Western blot with these antibodies showed that the tail fiber and the two unknown fibrous sequences were highly antigenic. This evidence implies that the unknown structures may act as host recognition proteins in addition to the tail fiber. Characterization of these novel proteins may provide insight to the host recognition abilities of cyanophages.

An additional study was also carried out, investigating the high temperature limit of the growth of phage P22. The results revealed that the production of infectious particles was limited by the temperature sensitivity of the folding and assembly of the P22 tailspike protein. This work has been published and is included in the Appendix.

ACKNOWLEDGEMENTS

First of all, many, many thanks to my advisor Jonathan King. Every successful Ph.D. thesis needs someone who can see the way through all the starts and stops, and I am incredibly fortunate to have had such a clear-sighted person to point the way---even when I didn't want to listen. Jon has taught me about being a good scientist, my responsibility as a scientist to the rest of the world, and maintaining a healthy balance in my life. Thanks for taking on a WHOI student.

Cammie Haase-Pettingell, Peter Weigele, and Jaqueline Piret have been my closest research partners during this whole process. Cammie got me started thinking experimentally again with the P22 work and later helped with protein induction and always with general moral support. She was perpetually ready lend an ear, whether it was for advice on a protocol, how to find something in the lab, or just in general. Thanks for all of those soy hot chocolates; I needed every one of them. Peter's help was invaluable throughout the entire process, providing the initial start-up boost on the Syn5 projects, training me on the EM, and generally acting as a sounding board for any ideas. I hope I wasn't too much of a pain. Jacqueline spent hours trying to help me clone phage genes... I know now that it is true that you learn things better when you get them wrong the first time. Matt Sullivan was a wonderful source of information on all things cyanophage...not to mention extra host cells if mine all happened to die. Debbie Lindell provided much advice and assistance. Vanessa Ward counted many SYBR stained frames.

This project would not have even been started if it wasn't for John Waterbury and Freddie Valois, the original isolators of Syn5. John provided advice in selection of the initial phage/host system, and then gave me the phage, the host cells, taught me how to grow them, and loaned me a microscope. Freddie taught me how to count. Thank you both!

The genome of Syn5 was sequenced at the Pittsburgh Bacteriophage Institute. Thanks to Roger Hendrix, Graham Hatfull, Marisa Pedulla, Jen Houtz, and Mike Ford who sequenced, assembled, and taught me how to annotate Syn5. Ian Molineux helped make it a *good* annotation. Thanks to Roger for hiring me.

Thanks to Patricia Reilly, who trained me in thin-sectioning. Thanks to Richard Cook, Alla Leshinsky, and Heather Amaroso of MIT Biopolymers performed the mass-spectroscopy and N-terminal sequencing on the Syn5 proteins, and spent many hours teaching me how to understand all the data. Thanks to Drew Endy's lab, who let me use their large epifluorescence microscope.

My committee members Penny Chisholm, John Waterbury, Graham Walker, and Eric Webb have been incredibly supportive and insightful. Thank you for all your input into this project.

Thanks to Cindy Woolley for teaching me how to lift weights, listening to all my rants, and taking care of nasty paperwork. Thanks to Ryan Simkovsky for acting as a buffer and taking me to Man Ray.... And printing out my final copies.

Thanks to Abby Bushman for being my copy editor, collator, and anchor of sanity at work in the turmoil of my deadlines.

The other King Lab members past and present: Melissa Kosinski-Collins, Shannon Flaugh, Ishara Mills, Jiejin Chen, Ligia Acosta, Steve Raso, Ajay Pande, Claire Ting, Veronica Zepeda, Kristen Cook, Phil Campbell, Cecilie Lin, Robin Nance, and visitors Sean Decatur, Katerina Papanikolopoulou, and Anait Seul. You guys are the reason why this is the best lab in the whole world. I will miss you.

Thanks to all my WHOI bio classmates: Kristen Gribble, Tin Klansjek, Joy Lapseritis, Gareth Lawson, Amanda McDonald, Eric Montie, and Sheri Simmons. I never would have made it through generals without you.

Thanks to the MIT Musical Theatre Guild and the Thirsty Ear Pub, for supplying such healthy distractions from science over the past five years.

My family is wonderful and always has been. I know that I would be nowhere near as successful and happy in my life as I am right now if it wasn't for all of you, and I am sorry this whole process meant that I couldn't see you nearly as much as I would like to. Thanks to Mom, Pop, Auren, Grandma, Grammie, Gramp, and everyone else for helping me in so many ways...art lessons, dance parties, Christmas Eve margaritas and unconditional love and support. Maybe next summer I can finally learn how to play bridge.

And to Pete, for everything.

This research was supported by NIH grant GM17980 and NSF grant EIA0225609 to Jonathan King, and by a Clare Booth Luce Fellowship, WHOI Education Fellowship and MIT SeaGrant travel grant to Welkin Pope

BIOGRAPHICAL NOTE
Welkin Hazel Pope

Education and Training

Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Ph.D., 2005, Biology

Massachusetts Institute of Technology, S.B., 2000, Biology

Research and Professional Experience

2001-2005	Massachusetts Institute of Technology Department of Biology, Cambridge MA, Graduate Research Assistant in the laboratory of Prof. Jonathan King, Ph.D.
2000-2001	Woods Hole Oceanographic Institution Department of Biology, Woods Hole MA, Graduate Research Assistant in the laboratories of Dr. Robert L. Olson, Ph.D., and Dr. Heidi M. Sosik, Ph.D.
1999-2000	Massachusetts Institute of Technology Departments of Biology and Civil and Environmental Engineering, Cambridge MA, Undergraduate Research Assistant in the laboratory of Prof. Sallie W. Chisholm, Ph.D.
1998-1999	Massachusetts Institute of Technology Department of Civil and Environmental Engineering, Undergraduate Research Assistant in the laboratory of Prof. John T. Germaine, Ph.D.
1997-1998	Massachusetts Institute of Technology Department of Brain and Cognitive Sciences, Undergraduate Research Assistant in the laboratory of Prof. Matthew A. Wilson, Ph.D.

Publications

Pope, W.H., Haase-Pettingell, C., King, J. 2004. "Protein folding failure sets high-temperature limit on growth of Phage P22 in *Salmonella enterica* serovar *typhimurium*" *Appl. Environ. Micro.* **70**:4840-4847

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATP	Adenosine triphosphate
CO_2	Carbon dioxide
CsCl	Cesium chloride
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
DNAP	Deoxyribonucleic acid polymerase
ds	Double-stranded
EDTA	Ethylenediaminetetraacetate
E. coli	*Escherichia coli*
FSW	Filtered seawater
HPLC	High pressure liquid chromatography
gp	Gene product
ICTV	International Committee on the Taxonomy of Viruses
LTQ	Linear trap quadrupole mass-spectrometer
LB	Luria-Bertani Broth
$MgCl_2$	Magnesium chloride
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
MW	Molecular weight
NaCl	Sodium chloride
NaOAc	Sodium acetate
NCBI	National Center for Biotechnology Information
NR	Non-redundant database
ORF	Open reading frame
PBS	Phosphate buffered saline
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PVDF	Polyvinylidene fluoride
RFLP	Restriction fragment length polymorphism
S. typhimurium	*Salmonella enterica* serovar *typhimurium*
S-D	Shine-Dalgarno sequence
SN	Natural seawater medium
ss	Single-stranded
RNA	Ribonucleic acid
RNAP	Ribonucleic acid polymerase
rRNA	Ribosomal ribonucleic acid
SDS	Sodium dodecyl sulfate
TAE	Tris-actetate-EDTA buffer
TE	Tris-EDTA buffer
TEM	Transmission electron microscopy
tRNA	Transfer ribonucleic acid
TRIS	Tris(hydroxymethyl)-amino methane
UV	Ultra violet light

CHAPTER I: INTRODUCTION

A: PHAGE ABUNDANCE AND MORPHOLOGY

Bacteriophages are highly abundant viruses which propagate on bacteria. These particles are ubiquitous throughout the world, recovered from the oceans, deserts, rivers, Antarctic, and sulfur hotsprings; as well as from within the guts of larger organisms and their feces. An exception to date, from which bacteriophage have not been isolated which do contain bacteria, are the hydrothermal vents in the deepest parts of the oceans, but it seems that this is a matter of time. Overall, these particles are believed to be most numerous organisms on the planet, at an estimated global population of 10^{31} phages (Hendrix et al., 1999), and are also believed to be the largest reservoir of genetic information on Earth (Brussow and Hendrix, 2002).

The first electron micrographs of phages, *Escherichia coli* (*E. coli*) T-phages, were taken by Luria and Anderson in 1942 (Luria and Andersen, 1942). These pictures revealed that the bacteriacidal particles were unique from any other known prokaryotic structure. T. F. Anderson wrote in his memoirs that this was not only the first pictures of the particles, but the first evidence that there was more than one type of particle (as shown by multiple morphologies), and that it took two weeks worth of repeating their experiments until Max Delbruck was convinced of it. (Anderson, 1975; Luria et al., 1943). Far larger than the largest cytoplasmic organelle, the ribosome, the phage particles measured 50-200 nm in diameter. The particles themselves were composed of a large

symmetric structure, called a "head", with a smaller structure protruding at one vertex disruptive of the overall symmetry of the particle, called a "tail". The later use of negative staining with uranyl acetate during electron microscopy revealed that the "head" or "capsid" of the phage was a highly regular structure with icosehedral symmetry, instead of spherical as previously supposed (Horne and Wildy, 1961; Horne and Wildy, 1963). The tail of the phage was the primary means of attachment of the particle to a host bacterium.

B: PHAGE TAXONOMY

Phage taxonomy is based both on morphology of the particle and on the nucleic acid of their genome (double-stranded (ds) DNA, single-stranded (ss) DNA, dsRNA, or ssRNA), as delineated by the International Committee on the Taxonomy of Viruses (ICTV) in their most recent report. (van Regenmortel et al., 2000). 96% of all characterized phage are "Caudovirales"; that is, possess a tail (Brussow and Hendrix, 2002). Tailed phages fall into three classes: the short-tailed phages, termed "Podoviridae" (P22, T3, T7, marine phages P60, P-SSP7, SIO1); the long non-contractile tailed phages, termed "Siphoviridae" (lambda, L5, TM4); and the long contractile tailed phages, termed "Myoviridae" (T4, P2, marine phages S-PM2, P-SSM4). Other phage morphologies include filamentous phages (fd, Pf3, CTXphi (McLeod et al., 2005; Wen et al., 2001; Wen et al., 1997)) and phages which maintain their apparent icosehedral symmetry by positioning tail fibers at all vertices (phiX174 (McKenna et al., 1992)), similar to eukaryotic viruses like adenovirus, herpes, and HIV. Phages have also been discovered

that contain a lipid core within their capids surrounding their genomes (PRD1 (Cockburn et al., 2004), marine phage PM2 (Huiskonen et al., 2004)).

With the advent of DNA sequencing, many bacteriophage genomes have been completely sequenced, with the result that the determination of viral taxonomy has become more complex (Hendrix et al., 2000; Hendrix et al., 1999; Paul and Sullivan, 2005; Paul et al., 2002). Studies of bacteriophage genomes have shown that the dsDNA phages exhibited a marked mosaicism with respect to each other. Clusters of genes, such as structural genes or genes involved in DNA replication, might appear closely related between two particular phages, while other separate clusters might appear more closely related to other phages (Hendrix et al., 1999; Pedulla et al., 2003). Host organism appeared to have little relation to gene content or synteny within the phage genomes. Morphology also appeared to have limited relevance to the genome content and organization of phages. For example, *Salmonella enterica* serovar *typhimurium* phage P22, a podoviridae, was shown to possess extensive genomic similarity to coliphage lambda, a siphoviridae (Botstein and Herskowitz, 1974; Brussow and Hendrix, 2002).

C. PHAGE INFLUENCES ON HOST POPULATIONS

Bacteriophages are known to mediate lateral gene transfer between microbial cells and to be a source of new genes; and therefore may confer new traits on the host organisms (Ochman et al., 2000; Weinbauer and Rassoulzadegan, 2004). Temperate phages may alter the sensitivity of the host cells to secondary infections (Wright, 1971),

and provide new genes and functions to the host cells (Calendar, 1988). *Vibrio cholera* becomes pathogenic by the infection and prophage integration of the temperate vibriophage CTXphi, as a phage-encoded fiber allows the lysogenic cells to attach to the mammalian stomach lining (McLeod et al., 2005; Waldor and Mekalanos, 1996).

During periods of host stress, temperate phages may excise imperfectly, thus including a portion of the host's genome in the new phage genome. After this newly packaged phage infects a new host cell, formation of a new prophage can result in the new host's acquisition of genes from the old host. During the prophage state, if enough integral phage genes are lost through mutation the ability of the phage to excise from the genome may be lost, resulting in a permanent gain of genes by the new host. Most microbial genomes show evidence of prophage integration (reviewed in (Casjens, 2003)).

Transducing phages package a portion of the host's genome, and then by further infection and recombination mediate the movement of sections of host DNA from one host to another. This can result in the permanent gain of new genes by the new host cell (Weinbauer and Rassoulzadegan, 2004). The rate of transduction in the marine environment is currently unknown; however, Jiang and Paul estimated that approximately 1.3×10^{14} transduction events occur annually in Tampa Bay (Jiang and Paul, 1998).

All of the effects that bacteriophage may have on host cells, either lateral gene transfer, lytic control of community structure or primary production; are mediated by phage infection. The initial particle structure required for infection in types of known

phages is the tail. Infection is initiated by the tail fiber recognition of the appropriate host cell receptors (Reanney and Ackerman, 1982). The tail fibers of T4 have been highly characterized (Cerritelli et al., 1996; King and Laemmli, 1971; King and Wood, 1969). The jointed T4 long tail fibers are comprised of four proteins: gps 34, 35, 36, and 37. Gp 34 and 37 comprise the major structural components of the fibers, with gp 34 comprising the entirety of the proximal half of the fiber and enabling attachment to the baseplate of the phage. Gp 35 forms the "knee" of the jointed fiber, connecting the distal end of gp 34 and the proximal end of gp 36. Gp 36 is responsible for about 20% of the length of the long tail fiber and has some host recognition properties while gp 37 comprises the majority of the distal half of the fiber and interacts directly with the host lippopolysaccharide at the distal tip of the fiber (Kikuchi and King, 1975a; Kikuchi and King, 1975b; Kikuchi and King, 1975c; Kikuchi and King, 1975d). Correct assembly of these fibers in T4 depends on gp 38 and gp 57, which are not part of the final structure. In phage T2, which appears morphologically identical to T4, gp 38 forms the extreme distal tip of the final tail fiber structure, conferring host specificity (Riede et al., 1986) and the C-terminal 120 amino acids of gp37 are proteolytically removed prior to fiber assembly (Drexler et al., 1986). T4 long fibers recognize host lippopolysaccharide and the outer-membrane protein (OMP) C (Henning and Jann, 1979). Several studies have shown that it is possible to swap sections of the tail fibers responsible for host-recognition within T-even like phages during mixed infections within hosts, conferring opposite host recognition on the newly formed phages (Tetart et al., 1998; Tetart et al., 1996).

Siphocoliphage lambda has short, side tail fibers when observed by TEM. Within the lambda genome, there are genes encoding a protein similar to gp 38 of T4 (named *tfa*, for "tail fiber assembly") and a protein similar to the c-terminal portion of gp 37 (named *stf*, for "side tail fiber"). These proteins were similar enough between the two phages such that expression of *tfa* could complement T4 gene 38 mutants and a hybrid phage in which the T4 gene 38 and the c-terminal portion of gene 37 were replaced by lambda *tfa* and *stf* resembled T4-like phages morphologically, but exhibited altered host specificity (George et al., 1983; Michel et al., 1986; Montag and Henning, 1987; Montag et al., 1989). Closer examination of the lambda genome revealed that *stf* and its neighboring open reading frame (ORF), ORF401, both exhibited similarity to the single tail fiber gene (gene H) of phage P2, with the product of *stf* resembling the proximal end of the fiber and ORF 401 resembling the distal (and host recognizing) end (Haggard-Ljungquist et al., 1992). These two ORFs in lambda could be converted to a single gene, resembling gene H, with the addition of one basepair between them. Examination of Ur-lambda, a lambda-like phage exhibiting long tail fibers, by Hendrix and Duda (1992), revealed that this phage did not possess the frameshift found in the common laboratory lambda. Interestingly, the tail fibers of Ur-lambda also contained the *tfa* protein, despite the ability of *tfa* expression to complement gene 38 in T4 (Hendrix and Duda, 1992).

P22 tailspikes have been highly characterized (Goldenberg et al., 1982; Goldenberg and King, 1981; King and Yu, 1986; Smith et al., 1980; Smith and King, 1981). The tailspike of P22 is a primarily beta-sheet homotrimer, which is comprised of three domains. The N-terminal domain binds to the neck proteins of the mature virion.

The middle of the elongated fiber is comprised of three parallel beta helices (one made out of each polypeptide chain). These domains have endorhamnosidase activity along their lateral surfaces (not the distal tip, like the long fibers of T-even phages) which binds and cleaves a specific octosaccharide motif in the O-antigen of lippopolysaccharide of the *S. typhimurium* cell surface (Goldenberg et al., 1982). The distal portion of the protein exhibits a triple-beta helix, in which all three polypeptide chains twine around each other. It is believed that this structure forms a molecular "clamp" conferring additional stability to the protein (Kreisberg et al., 2002).

A striking example of broad host range adaptation is found in the temperate myocoliphage mu. This phage alters its host range through the inversion of genome segment G genes which encode two sets of tail fiber genes, S and U, and S' and U' (Grundy and Howe, 1984; Morgan et al., 2002; van de Putte et al., 1980). Inversion of the segment is controlled by a phage encoded invertase, expressed by the gene *gin*, which is located on the non-inverting B segment of the genome. When the G segment is in the (+) orientation, the S tail fiber and U tail assembly protein are transcribed, and mu is highly efficient at absorbing to *E. coli* K12. When G is in the (-) orientation, S' and U' are transcribed, and allow mu to absorb to *E. coli* strains, *Citrobacter freundii*, and *Shigella sonnei (Kamp et al., 1978; van de Putte et al., 1980)*. S and S' share a common region in the non-inverting A region of the genome which includes the promoter and N-terminal encoding portion of the gene. During lytic infection of *E. coli* K12, almost 99% of mu particles were produced with G in the (+) orientation. In contrast, during mu prophage induction, approximately 50% of the particles produced had genomes with the G (+)

orientation and 50% had genomes with the G (-) orientation (Bukhari and Ambrosio, 1978). Subsequent studies revealed that *gin* is expressed during the prophage state, resulting in the equal distribution of the two phage types upon induction (Symonds and Coelho, 1978).

Phage tails perform two distinct functions: primary host recognition and secondary host receptor binding for phage DNA transfer. Host recognition occurs when a phage interacts with a primary protein or structure on the outer surface of the host cell, and this interaction identifies the cell as a host. An example of this is the P22 tailspike identification and cleavage of the *S. typhimurium* lippopolysaccharide. Upon the secondary phage binding to the host receptor, the tail fibers trigger a structural shift in the proteins responsible for plugging the portal hole in the capsid and trapping the phage genome inside. This structural shift opens this hole and allows passage of the phage genome to the host cell through the receptor (Makhov et al., 1993). The exact secondary receptor for P22 is not known. In some phages, like T4, the two stage binding process is much clearer. In T4, functions of host recognition and host receptor binding are performed by separate proteins; the long tail fibers recognize the correct host; then the short tail fibers bind to the lippopolysaccharide and trigger the rearrangement of the baseplate of the tail from a flat hexagon to an open star shape, allowing the tail tube to penetrate the cell surface (Crowther et al., 1977). This conformational shift also allows the baseplate-bound lysozyme access to the cell surface, the cell-wall degrading activity of which aids the penetration of the tail tube and the delivery of the phage DNA to the interior of the cell (Liao and Syu, 2002). Other known phage secondary receptors include

the LamB protein in the maltose transporter (lambda) (Szmelcman and Hofnung, 1975), which is recognized by the tip of the lambda tail tube.

D: VIRAL EVOLUTION

The evolution of viruses is poorly understood. Some unanswered questions are: Are new viruses are being continually generated out of cellular components?, or the converse Were the major classes of viruses evolved billions of years ago, and all viruses present now descended from the original types?. Evidence supporting the "continual new production of virus" lies in the examination of known bacterial structures which may transfer DNA. For example, a bacterial pilus closely resembles filamentous phages structurally, and is the means of a transfer of a short stretch of DNA from one cell to another. Phage portals are highly regular structures with an opening just wide enough for dsDNA to pass through, similar in organization to porins in cell membranes. One could propose that new phages could be formed by an opportune pinching off of a pilus, trapping inside the DNA which encodes the proteins necessary for pilus formation. On the other hand, overall similarity of the protein sequences and three dimensional structures of many phage structural proteins within the Caudovirales would seem to imply that these proteins are all related to a distant ancestor.

E: MARINE VIRUSES AND CYANOPHAGES

Virus-like particles were first discovered in the oceans in the late fifties (Spencer, 1955; Spencer, 1960; Spencer, 1963). Application of transmission electron microscopy (TEM) in the early nineties revealed substantially higher concentrations of viruses than previously expected (~10^8 particles/ml) (Bergh et al., 1989; Borsheim et al., 1990; Bratbak, 1990; Bratbak et al., 1992; Heldal and Bratbak, 1991). Despite their supposed adaptation to a radically different environment, many of the marine viral particles appear to resemble the enteric tailed bacteriophage, with icosehedral heads, and either short tails, long contractile tails or long non-contractile tails. These particles contain genes for structural proteins which are conserved between the marine isolates and the laboratory model organisms (Short and Suttle, 2005).

A subset of these phages have been shown to infect the cyanobacteria *Synechococcus* and *Prochlorococcus* (Suttle and Chan, 1993; Waterbury and Valois, 1993). Marine cyanobacteria are ubiquitous throughout the world's oceans. *Prochlorococcus* and *Synechococcus* are the numerically dominant photosynthetic organisms in the large oligotrophic gyres of the open and contribute a significant fraction of global photosynthesis (Li, 1994; Li, 1998; Liu et al., 1995; Liu et al., 1998; Liu and Landry, 1999; Liu et al., 1997; Partensky et al., 1999; Veldhuis et al., 1997). *Synechococcus* is found in higher concentrations nearer the coast, reaching densities of 10^5 or 10^6 cells/ml (Sullivan et al., 2003; Waterbury et al., 1986). These genera make up the base of the marine microbial food web, fixing CO_2 and inorganic nutrients, and allowing for the transfer of energy up through the trophic levels (Azam, 1998).

The viruses which infect these cyanobacterial species, termed "cyanophage", are ubiquitous throughout the euphotic zone like their hosts. Their natural abundances have been estimated to occur within an order of magnitude of the measured host cell concentration (Lu et al., 2001; Sullivan et al., 2003; Suttle and Chan, 1994; Waterbury and Valois, 1993). Current sampling methods, however, rely on the direct infection of host cells by collected particles, which means that only phages which are capable of infecting a laboratory strain are measured (Wommack and Colwell, 2000). So far, no one has managed to titer natural cyanophage populations on the numerically dominant strain co-existing at the same time (Muhling et al., 2005), and cyanophage host receptors remain unknown. Ecologic studies of cyanophages of both *Prochlorococcus* and *Synechococcus* have shown that many isolates are capable of infecting many strains of both genera, thus providing direct evidence for potential gene transfer between two genera (Sullivan et al., 2003; Waterbury and Valois, 1993).

Due to their abundance, cyanophages may have significant influences on ambient host populations, and therefore may affect the structure of the marine microbial food web (Jiang and Paul, 1998; McDaniel et al., 2002; Muhling et al., 2005; Suttle and Chan, 1993; Waterbury and Valois, 1993). Lytic actions may result in the loss of cyanobacteria, and therefore a loss of energy available for energetic transfer, as well as altering community structure by killing sensitive cells and encouraging strain succession. In coastal regimes, lytic cyanophages are estimated to be responsible for 3% of host cell mortality (Lu et al., 2001; Mann, 2003; Suttle, 2000; Suttle and Chan, 1994; Waterbury and Valois, 1993). Evidence for lytic phage control of *Synechococcus* succession was

found in a recent study of the seasonal succession of cyanophage and host cells. *Synechococcus* and cyanophage concentrations were monitored throughout the year in the Gulf of Aqaba, Red Sea (Muhling et al., 2005). *Synechococcus* and cyanophage marker genes were amplified by polymerase chain reaction (PCR) and screened by restriction fragment length polymorphism (RFLP) and denaturing gradient gel electrophoresis (DGGE) for specific strains. During the spring bloom of host cells, one strain of *Synechococcus* became dominant, with a following corresponding increase in the overall cyanophage titer and lower RFLP diversity. During the winter, the time with the lowest host cell numbers, the greatest diversity in phage and host RFLP was detected, showing that both phage abundance and diversity co-varied with that of the host cells. Unfortunately, it was not possible to culture the dominant *Synechococcus* strain during the spring and summer months, and therefore it was not possible to determine if the dominant phage type could infect the dominant host strain.

F: ENVIRONMENTAL INFLUENCES ON INFECTIOUS PHAGE PRODUCTION

There are several studies which indicate that it is possible to induce cyanophages from natural host populations (McDaniel et al., 2002; Ortmann et al., 2002). However the sequenced cyanobacterial genomes do not show the presence of any intact prophage. While phage may have significant influences on the host cell populations and community structure, environmental factors may have significant influences on the production of infectious phage. Stresses placed on lysogenic cells, such as ultra-violet light (UV), may trigger excision of the prophage and initiation of the lytic cycle. Integration of prophage

may be influenced by slow growth rates (Rohwer et al., 2000; Suttle, 2000; Williamson et al., 2002; Wilson et al., 1996), such as are found among cyanobacteria in the surface waters during the winter.

Temperature influences the number of infectious particles produced during lytic infection. In my initial work with phage, I investigated temperature controlled production of infectious viruses with the model system of phage P22 and *S. typhimurium*. Production of infectious phage particles declined sharply above 37°C, but in a range at which the host continued to grow and divide. The loss of infectivity was due to P22 particles produced at higher temperature which lacked sufficient tail fibers to infect the host cells. The lack of tail fibers was due to the failure of its folding or assembly at higher temperatures (Pope et al., 2004) see Appendix A

G: QUESTIONS EXAMINED BY THIS THESIS

Cyanophages have been detected within the surface layers of the oceans, and likely influence ambient host cell populations. Host/phage interactions require recognition and infection of the host cell, mediated by the tail organelles of the particle. Characterization of these protruding structures of the cyanophages have been limited in part by the lack of a robust method for the growth and purification of high titer cyanophage stocks. The cyanophage Syn5, a podocyanophage, was selected as a model organism to attempt to explore the structure of cyanophage particles. Its morphological simplicity implied that it would have fewer structural proteins for identification and

analysis than the myocyanophages. The second chapter of this thesis describes the concentration and purification of a cyanophage isolate to high enough titer to explore the phage/host system, examine the phage genome, and examine the structural proteins.

Once a robust method for host cell and phage growth as well as for phage purification had been developed, it was possible to address how the genome of this cyanophage compared with those previously sequenced phages. Chapter III reports the sequencing of the genome of Syn5, together with the annotation of the sequences. Among the questions explored in Chapter III are: Does Syn5 fit into the recognized groups of phage previously identified by genome sequencing and analysis? How related is Syn5 to known cyanophage sequences and known enteric phage sequences? Are there any identifiable novel genes present within the genome and what does the presence of these genes say about the lifestyle and environment to which the phage is adapted?

Finally, the structural proteins of Syn5 were examined by Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and eleven visible protein bands were assigned to predicted open reading frames (ORFs) within the Syn5 genome. Questions addressed in Chapter IV include: Do the structural proteins of Syn5 resemble those of the related enteric phages? Are novel structural proteins present within the particle, and what does the present of these proteins imply about the lifestyle and environment of the phage? Through the examination of these questions, this thesis attempts to begin the exploration the larger questions of viral evolution and the methods of cyanophage host recognition.

CHAPTER II:

PURIFICATION AND CHARACTERIZATION OF THE MARINE CYANOPHAGE SYN5

A: INTRODUCTION

Marine cyanobacteria are numerically dominant photoautotrophs in the global oceans (Partensky et al., 1999; Waterbury et al., 1986). Two genera, *Prochlorococcus* and *Synechococcus* have been shown to contribute a significant amount to global photosynthesis (Goericke, 1993; Li, 1994; Liu et al., 1998; Liu et al., 1997; Veldhuis et al., 1997). Marine viruses which infect these genera, termed "cyanophage", have been isolated from various environments ranging from estuarine to open ocean, with infectious particle concentrations in each regime ranging from similar to that of the host cells to within an order of magnitude lower (Sullivan et al., 2003; Suttle and Chan, 1994; Waterbury and Valois, 1993; Wilson et al., 1993). Due to their presence and prevalence within the upper surface layers, cyanophage have been implicated in the control of the cyanobacterial community structure and primary production; and therefore also in the flow of nutrient and energy throughout the marine ecosystems (Bergh et al., 1989; Bratbak et al., 1992; Proctor and Fuhrman, 1990; Suttle et al., 1990).

Recent analyses of enteric phage genomes have revealed a marked mosaicism among phages which infect the same host bacteria, as well as among phages which do not (Hendrix et al., 1999; Pedulla et al., 2003). Cyanophage genomes appear to exhibit a

similar mosaicism, including cassettes of genes with high levels of similarity to enteric phages, to host cyanobacteria, as well as genes of unknown function which are still found in multiple phages regardless of morphology (Chen and Lu, 2002; Mann et al., 2005; Sullivan et al., 2005) .

Despite their ability to infect photosynthetic host cells adapted to nutrient-poor regimes, all cyanophage isolated to date appear morphologically similar to three types of dsDNA tailed enteric phages (Sullivan et al., 2003; Suttle and Chan, 1993; Waterbury and Valois, 1993; Wilson et al., 1993): Myoviridae, long-contractile tailed T4-like phages; Siphoviridae, long-noncontractile tailed lambda-like phages; and Podoviridae, short extensible tailed T7-like phages (enteric designations refer to morphology only, as phage genomics is demonstrably more complicated). The presence of similar phage morphologies between enteric and cyanophages indicates that these phages may be descended from the same ancestral phage, yet adapted to suit their extremely different environments. As such, comparisons of the structural proteins may yield valuable information, including adaptation to certain environmental stresses, or host specificity.

While five cyanophage genomes have been sequenced and annotated to date, few biochemical studies have been performed to analyze their structural proteins. One preliminary study characterized five cyanophages of the morphologic types myoviridae and siphoviridae isolated on *Synechococcus* strain WH7803 (Wilson et al., 1993). Characterization included the isolation and digestion of phage genomic DNA by restriction enzymes, as well as analysis of particle proteins via SDS-polyacrylamide gels.

The cyanophage DNA was susceptible to some, but not all, of the restriction enzymes tested while the particle proteins exhibited similar banding patterns to enteric phage of the corresponding morphology.

A major barrier in the characterization of the molecular biology of cyanophages is the difficulty in obtaining significant amount of purified phage particles. Part of the reason for the difficulty in generating high quality cyanophage concentrates lies in the fact that cyanobacteria grow more slowly and do not reach the same cell densities as enteric bacteria, but rather seem to plateau around 10^8 cells/ml under optimal laboratory growth conditions, an order of magnitude lower than *E. coli* or *S. typhimurium*. Concentration of large quantities of crude phage lysate via methods developed with enteric phage has proved difficult (Sullivan et al., 2003; Wilson et al., 1993), resulting in low yields of particles.

In the development of a laboratory system for biochemical work on cyanophage proteins, we chose a short-tailed phage with fewer different protein chains within a given particle than a long tailed phage, to aid the protein identification process. The short-tailed phage also appeared to have a briefer lytic period (on the order of hours) than the long-tailed phage (on the order of days), which would speed the experimental periods. Syn5 is a podoviridae isolated by Waterbury and Valois (1993) from the Sargasso Sea on *Synechococcus* strain WH8109 (a member of the Marine A cluster of *Synechococcus*, clade II). At least 15 cyanophage isolates also propagate on WH1809, rendering it one of the more phage sensitive strains in collection as tested by Waterbury and Valois (1993)

and by Sullivan et al (2003). Syn5 did not propagate on at least 19 other *Synechococcus* and *Prochlorococcus* strains tested by Waterbury and Valois (1993) and by Sullivan et al (2003). This host specificity was common of the podoviridae isolates tested. As we were interested in the study of gene exchange among cyanophages, we selected this phage/host system for our initial study.

B: Materials and Methods

i. *Cell Growth*

Synechococcus strain WH8109 (herein after referred to as WH8109) was grown in SN media (Waterbury and Willey, 1988) at 26°C under continuous light at an irradiance of $50\mu Em^{-2}s^{-1}$ supplied by cool white fluorescent light (40W bulbs). WH8109 was grown in three types of containers with different aeration: the culture vessel with a glass fritted aerator, the culture vessel with bubbling without a fritted aerator, and a 2L Erlenmeyer flask swirled manually for several seconds daily. Large culture vessels were constructed from a 2L rectangular polycarbonate bottle (Nalgene) and the three-holed polypropelene cap (Nalgene) while the small vessels were constructed from 100ml rectangular polycarbonate bottles. Silicon tubing (Cole-Parmer) was attached to an aquarium pump, filtered through a 0.4μ bacterial air vent (VWR) and attached to the air input of the culture vessel. Opposite the air input, within the culture vessel, was attached a glass–fritted aeration tube (VWR) such that the aeration tube ended just above the bottom of the vessel. The air output was kept sterile via a piece of silicon tubing twisted

into a Pasteur loop. The final port in the cap was kept closed except during sampling, during which it provided easy access to the culture within the vessel.

Cells from a stock culture of WH8109 grown on the benchtop at room temperature in SN under ambient light were inoculated into each of the experimental culture vessels. Cells were allowed to adapt to each condition for three generations prior to growth rate measurement. Growth rates were determined by daily cells counts. Samples of each culture were removed, diluted appropriately in filtered seawater (FSW), and filtered onto a 0.4μ polycarbonate filter. The filter was quantitatively analyzed using epifluorescence microscopy. Ten microscope fields were counted per filter, with between 30 and 100 cells per field, and the average was used to determine cell concentration (Waterbury et al., 1986).

ii. *Phage Burst Size*

WH8109 was grown in SN media to a cell density of $7x10^5$ cells/ml. 200ml of cell culture was removed to a small culture vessel and Syn5 was added at a multiplicity of infection (moi) of 5. The remainder of the culture was allowed to continue growth. At $1x10^8$ cells/ml (late-log phase), a further 200ml of culture was removed to a small culture vessel and Syn5 from the same stock was added at a moi of 5. A final 200ml of culture was removed when the cells reached a concentration of $5x10^8$ cells/ml and Syn5 from the same stock was added at a moi of 5. Each small culture was incubated until lysis, approximately 10 hours after infection. During infection, the cell culture changed from

dark red to bright yellow after approximately eight hours and finally after approximately ten hours to light green. The final change to green was accompanied by a substantial loss of the initial turbidity of the culture, resulting in a clear, green culture. This clear green culture was identified as a "lysed" culture, as defined by a loss of at least 99% of initial cells when examined by epifluorescence microscopy. Lysate samples were appropriately diluted with SN and filtered onto 0.2μ polycarbonate filters to determine total cells left with in the lysate, as well as onto a 0.02μ Acrodisc filters for SYBR green staining for virus particles. Filters were SYBR green stained as according to methods outlined previously (Noble and Fuhrman, 1998; Wen et al., 2004) and quantitatively analyzed using epifluorescence microscopy. Ten fields were counted per filter, with 100-200 virus particles per field.

iii. *Phage Purification and Concentration*

Phage were concentrated and purified using a modified version of polyethelene glycol (PEG) precipitation (Yamamoto et al., 1970). WH8109 was grown to a density of $1x10^8$ cells/ml at 26°C in SN media at an irradiance of 50μE $m^{-2}s^{-1}$. Syn 5 was added at a moi of 0.1. The culture was incubated until lysis, approximately 10 hours after infection (noted by a change in turbidity and color). The crude lysate was spun at 8,000 g for 15 minutes at 4°C to pellet cellular debris. The supernatant was then filtered through a Whatmann glass fiber filter and a 0.4μ polycarbonate filter to further remove cellular debris, specifically cell membranes, which may bind free phage and prevent their recovery. After the incubation, the lysate was moved to a 4°C room, and NaCl was added

to a final concentration of 0.5M and stirred to dissolution. PEG 8000 was added to a final concentration of 10%w/v, mixed to dissolution, and incubated with stirring for two hours at 4°C. Phage were then pelleted by centrifugation at 9,000g for 30 minutes. The supernatant was poured off, and the pellet was resuspended by gentle stirring at 4°C in an appropriate amount of SN (10-15ml per liter of crude lysate) supplemented with 50mM Tris-HCl, pH 8.0 and 100mM $MgCl_2$.

The suspension was layered onto a step-gradient made of 20% sucrose (w/v), ρ=1.4 CsCl, and ρ=1.6 CsCl. Each gradient layer was made with SN. The gradients were ultracentrifuged at 40,000 rpm for 1.5 hours in a SW50.1 rotor or 28,000 rpm for 4 hours in a SW28 rotor. An opalescent phage band was visible at the interface between the ρ=1.4 CsCl band and ρ=1.6 CsCl steps. The phage band was collected and dialyzed for 30 minutes against 1M NaCl, 100mM $MgCl_2$, 50mM Tris-HCl, pH 8.0, and then against 100mM NaCl, 100mM $MgCl_2$, 50 mM Tris-HCl pH 8.0 in a Slide-A-Lyzer dialysis cassette (Pierce).

iv. *Phage DNA Extraction*

High quality phage DNA was successfully recovered using a modified phenol:chloroform:isoamyl extraction and ethanol precipitation. 400ul of CsCl-purified and dialyzed phage particles were mixed by inversion with 25ul 10% sodium dodecyl sulfate (SDS), 50ul 100mM Tris-HCl pH 8.0, and 25ul 0.5M EDTA, pH 8.0 and incubated at 65°-70°C for 20min. 500µl phenol:chloroform:isoamyl, pH 8 was added and

mixed by shaking until the solution appeared white. The organic and aqueous layers were separated by centrifugation at 14,000 rpm using a tabletop microcentrifuge. The aqueous layer (approximately 450 µl) was removed to a fresh tube. 55µl 3M NaOAc (pH 5.2) was added to the solution and mixed by inversion. The solution was incubated on ice for one houe and then 1 ml of 100% ice-cold ethanol was added to precipitate the genomic DNA. The cottony DNA was carefully wound onto a heat-sealed Pasteur pipette and carefully swirled in a solution of 70% ethanol to wash. The DNA was then air-dryed at room temperature on the pipette for approximate 15 minutes to allow the ethanol to completely evaporate. The DNA was then dissolved into 150 µl of TE. Dissolved DNA was digested with BamHI, HindIII and XbaI and electrophoresed through a 1% agarose gel in TAE buffer. The gels were then stained with ethidium bromide and DNA bands were visualized using UV light.

v. *Phage Protein Analysis*

CsCl-purified Syn5 particles were mixed with SDS-buffer, boiled for 3 minutes, and loaded onto a 10% polyacrylamide gel containing SDS. The samples were electrophoresed at 20mA constant current until the dye line ran off the end of the gel. The gels were stained with Coomassie blue. 12 protein bands were visible and their approximate molecular weights were determined by comparison to a protein standard (Broad Range Protein Marker, New England Biolabs).

vi. *Thin-sectioning of Infected Cells:*

Synechococcus cells were grown to a density of $7x10^5$ cells/ml, and Syn5 particles were added at a moi of 10. 25ml of infected culture were pelleted by centrifugation (10min at 8000g) and the supernatant was poured off. The pellet was fixed in a 1% gluteraldehyde solution buffered with 0.1M sodium cacodylate buffer on ice in the dark for 30 min. The pellet was then gently washed three times in buffer and fixed again with 1% osmium tetroxide in 0.1M sodium cacodylate buffer. After three more washes, the pellet was dehydrated with increasing concentrations of ethanol, 5 min at each of 50%, 75%, 90%, and finally 3 times at 100%. After dehydration, the pellet was infused first with 100% propylene oxide and then propylene oxide mixed 1:1 with Spurr's Low viscosity resin (prepared according to manufacturer's instructions) while being constantly rotated for even infusion. Finally, the pellet was infiltrated overnight in pure resin on a rotator, then transferred to a 2ml BEEM capsule in fresh resin. The capsule was baked at 60°C for 24 hours. The hardened resin was trimmed with razor blades and glass knives to expose the sample, and then 70 nm ultra-thin sections were cut using a diamond knife and a microtome. The sections were transferred onto 200 or 300 uncoated mesh copper grids and stained with 1% uranyl acetate for 15 minutes, washed three times in double-distilled water, stained in 1% lead citrate for 4 minutes, and washed a final three times in double-distilled water. The sections were then examined with an electron microscope.

C: RESULTS

i. *Cell Growth*

Synechococcus WH8109 cultures for phage propagation were grown in SN media at 26°C and an irradiance of $50\mu Em^{-2}s^{-1}$ in chambers constructed from 2L polycarbonate Nalgene containers fitted with aeration and sampling ports (Figure 2-1), for maximal light exposure and constant gentle aeration. Cell concentrations were monitored by filtering samples of cell culture onto polycarbonate filters and counting cells using epifluorescence microscopy.

Figure 2-1: *Synechococcus* culture vessels. Culture vessels were constructed from 2L polycarbonate containers, three-holed lids with silicon caps, silicon tubing, and glass-fritted aeration tubes. Red cultures contain unlysed WH8109. The green clear culture was inoculated with Syn5 and has lysed two hours earlier.

As seen in Figure 2-2, WH8109 achieved the most rapid doubling time (approximately three doublings per day) and the highest cell density (10^9 cells/ml) in the designed culture vessel with constant aeration via the glass fritted aerator. Aeration without the frit – delivered through an open tube of identical length and diameter— resulted in growth that was not as rapid (doubling time of approximately 11 hours) and the cells did not reach as high a density ($5x10^8$ cells/ml). Cell growth in the Erlenmeyer flask was the slowest (doubling time of 19 hours), and the cell density only reached 1e8 cells/ml. Based on these results, all further phage characterization and purification was carried out in the containers with fritted aerators.

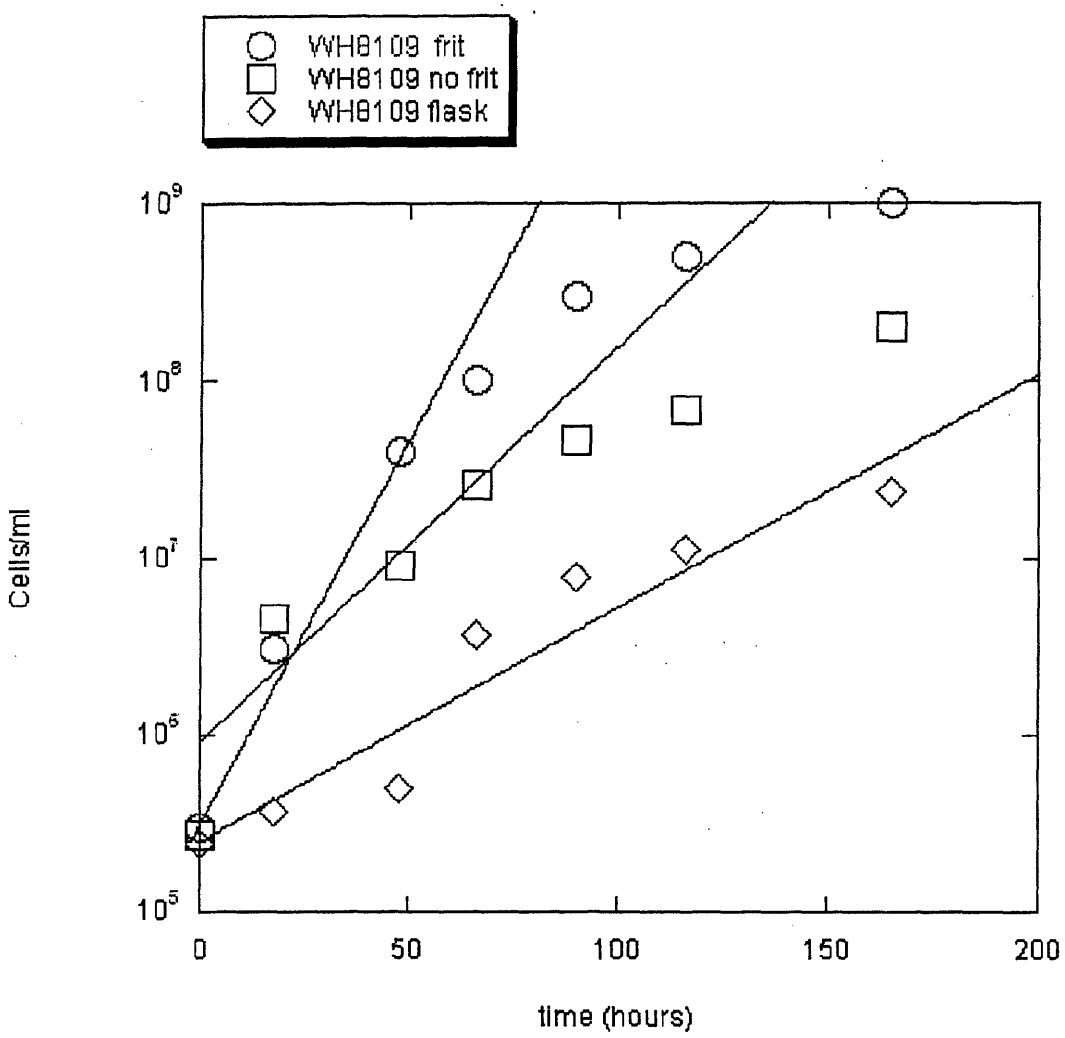


Figure 2-2: Growth curves of WH8109. Growth curves of WH8109 grown in 2L glass Erlenmeyer flask (diamonds) swirled twice daily, WH8109 grown in chamber with aeration from non-fritted glass tube (squares), and WH8109 grown in chamber with fritted aerator (circles). Best fit lines were determined using the log portion of the curves, and the growth rate was calculated from the slope of these lines, using the equation $N(t) = N_o e^{-\mu t}$, where N is the cell concentration, t is time, N_o is the initial cell concentration, and μ is the growth rate.

ii. *Phage Burst Size and Particle Yield*

Once the growth conditions of the host cells were determined, we investigated the best point during the growth cycle of *Synechococcus* to infect with Syn5 to maximize the burst size.

Rapidly growing host cells at concentrations of 10^7 cells/ml, $5x10^7$ cells/ml and 10^8 cells/ml (approximately 2-3 days after inoculation) were infected with Syn5 at a moi of 10. The high moi was used to ensure that every cell was infected by a phage during the first lytic period. Lysis occurred approximately 10 hours after infection, and was characterized by a change in color of the culture from the initial dark red to bright yellow after ~eight hours to light green after ~ten hours. Accompanying the change in color was a profound loss in the turbidity of the culture, from opaque to clear. After lysis, both a SYBR green count of phage particles and a cell count were performed on the crude lysate. This made it possible to determine the total number of phage particles in the crude lysate as well as the number of cells that remained unlysed. By subtracting the number of cells left in the lysate from the original cell number, only cells that were actually lysed were used to determine the burst size. Unlysed cells could be due to either a small resistant population within the host cells or the formation of lysogenic cells. The initial amount of phage added were also subtracted from the final phage titer.

As seen in Figure 2-3, the largest burst size of Syn5 in WH8109 grown under continuous light with aeration was approximately 30 particles per cell. This maximum

was achieved when the cells were infected during late exponential phase, at approximately 10^8 cells/ml. This burst size is about a fourth of the reported burst sizes for enteric phages like T7 (120 phage/cell) (Wang et al., 1999)and phi-YeO-3 (132 phage/cell) (Pajunen et al., 2000).

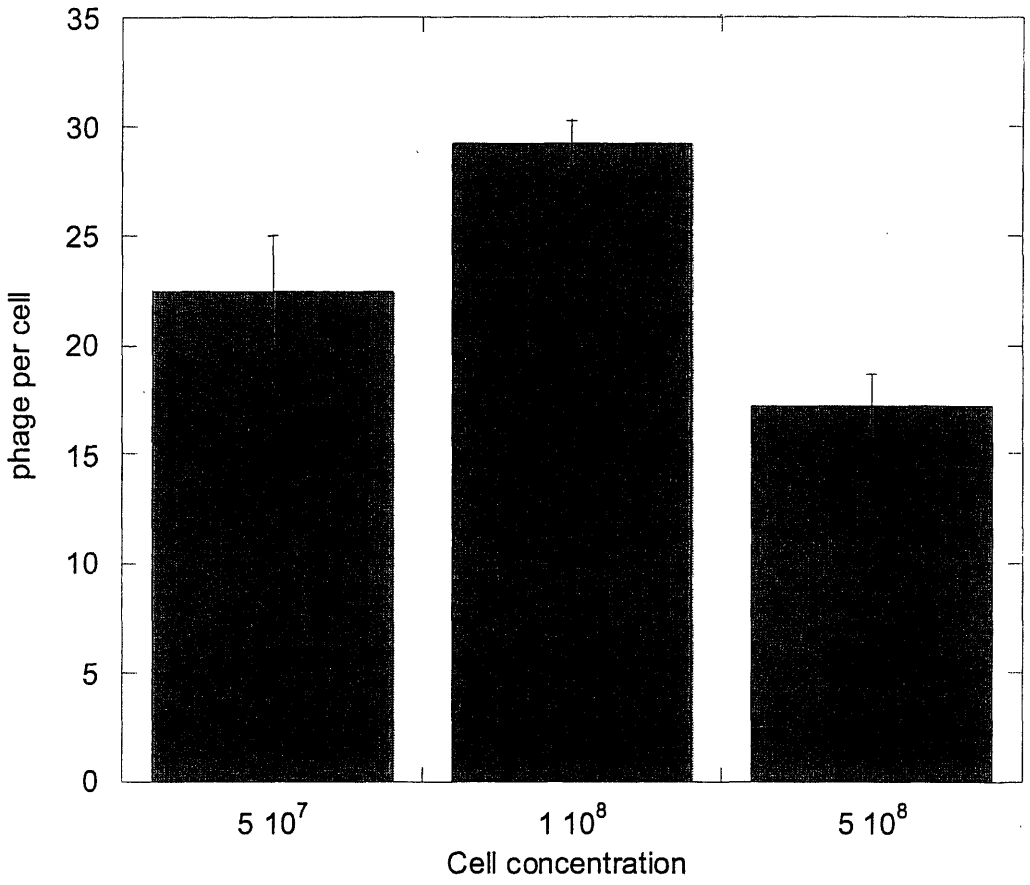


Figure 2-3 Burst Size of Syn5 during WH8109 growth. WH8109 was infected at $5x10^8$ cells/ml, 10^8 cells/ml, and $5x10^8$ cells/ml with Syn5 at a moi of 5. The samples were incubated until lysis (~10 hours) and the phage were titered using SYBR staining. Cell counts were performed on the crude lysate to assess how many cells remained unlysed. Phage titers were corrected for the input phage. Burst sizes were corrected for unlysed cells. Bar heights reflect the mean of three replicates. Error bars reflect standard error.

iii. *Phage Purification via PEG precipitation and CsCl gradient ultracentrifugation*

After the optimization of host and phage growth, conditions were developed to both purify and concentrate the phage particles, to allow for DNA and protein analysis. A well-known method for the concentration of enteric bacteriophage is precipitation from a crude lysate via polyethylene glycol (PEG) (Yamamoto et al., 1970). This method had previously been employed on cyanophage (Wilson et al., 1993).

We tested both a range of non-ionic detergents and a range of salt concentrations for their effects on cyanophage yield. Addition of detergent might minimize interactions between cyanophage and any remaining cellular membranes, and therefore allow more cyanophage to be recovered from the lysate. The variations in salt concentrations were tested to determine if 0.5M NaCl was still the optimal salt addition when working with phage in a seawater-based medium. Seawater alone is 0.5M salt and so it seemed possible that the PEG mediated phage precipitation might require more or less salt than a phage precipitation from LB.

In the detergent experiments, crude lysates were centrifuged and filtered to remove debris, and then split into 5 separate bottles. Each was treated with one of the following: 0.1% Triton-X 100, 0.2% Elugent, 0.2% NP40, 0.2% Tween 20, 0.2% Brij-58, or no detergent. NaCl was then added to 0.5M final concentration, and the samples were treated as in the Materials and Methods, with the exception that each solution used in the step-gradient was prepared with the same detergent at the same concentration as the

sample to be layered over it. Phage bands were harvested using a Coombs fractionator and dialyzed as in the methods. Phage yields were determined using SYBR staining. None of the detergents tested resulted in a significant increase or decrease in yield over no detergent (Figure 2-4), except perhaps Brij- 58. Though this experiment established that Syn5 was not inactivated by Triton-X 100, Elugent, NP40, or Tween20 detergents, subsequent purifications were carried out without added detergent.

The original PEG-phage precipitation protocol required an addition of 0.5M NaCl to the lysate prior to PEG precipitation to increase phage yield (Yamamoto et al., 1970). As SN media is 75% seawater and therefore already 0.375M salt, NaCl final concentrations (not including that of the SN) of 0M, 0.5M, 1M, 1.5M, 1.75M, and 2M were tested to further optimize our yield of cyanophage particles. The addition of 0.5M NaCl resulted in the highest particle yield, approximately 40% of the particles in the crude lysate (Figure 2-5). The maximum phage concentration achieved using the protocol listed in the methods was approximately 10^{12} particles/ml. Figure 2-6 shows phage yield at several purification steps, as well as a comparison in yield between samples prepared with Triton-X 100 and samples prepared without detergent.

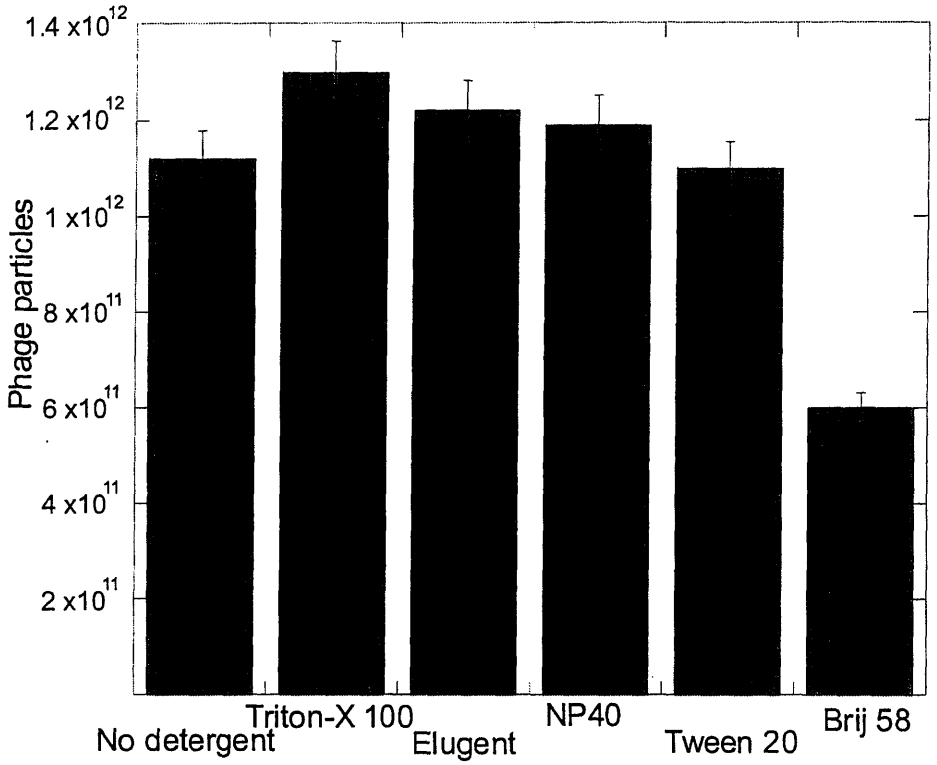


Figure 2-4: Syn5 particle yield during purification with non-ionic detergents. Syn5 was purified via PEG precipitation from a crude lysate. Prior to PEG addition, several detergents were added in an attempt to minimize phage/membrane interactions that could be preventing precipitation. Detergents tested were Triton-X 100, Brij 58, Elugent, NP40, and Tween20. After centrifugation, concentrated phage bands were harvested using a Coombs fractionator and dialyzed. Phage samples were then titered by filtration and SYBR green staining. Concentrations were determined by counting particles within at least 10 microscope fields. Error bars reflect standard error.

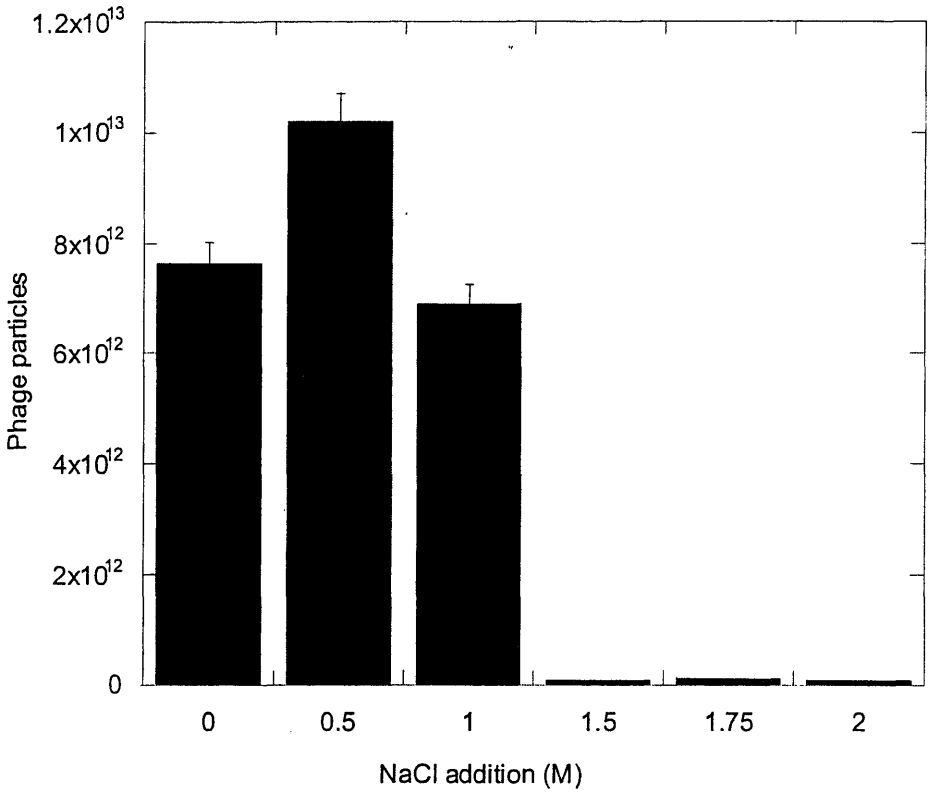


Figure 2-5: Syn5 particle yield when purified at varying salt concentrations
Syn5 was purified via PEG precipitation from a crude lysate. Prior to PEG addition, NaCl concentration was varied to determine maximal phage precipitation. Salt was added to final concentrations (not including initial SN salt concentration) of 0M, 0.5M, 1.0M, 1.5M, 1.75M, and 2M. After centrifugation, concentrated phage bands were harvested using a Coombs fractionator and dialyzed. Phage samples were then titered by filtration and SYBR green staining. Concentrations were determined by counting particles within at least 10 microscope fields. Error bars reflect standard error.

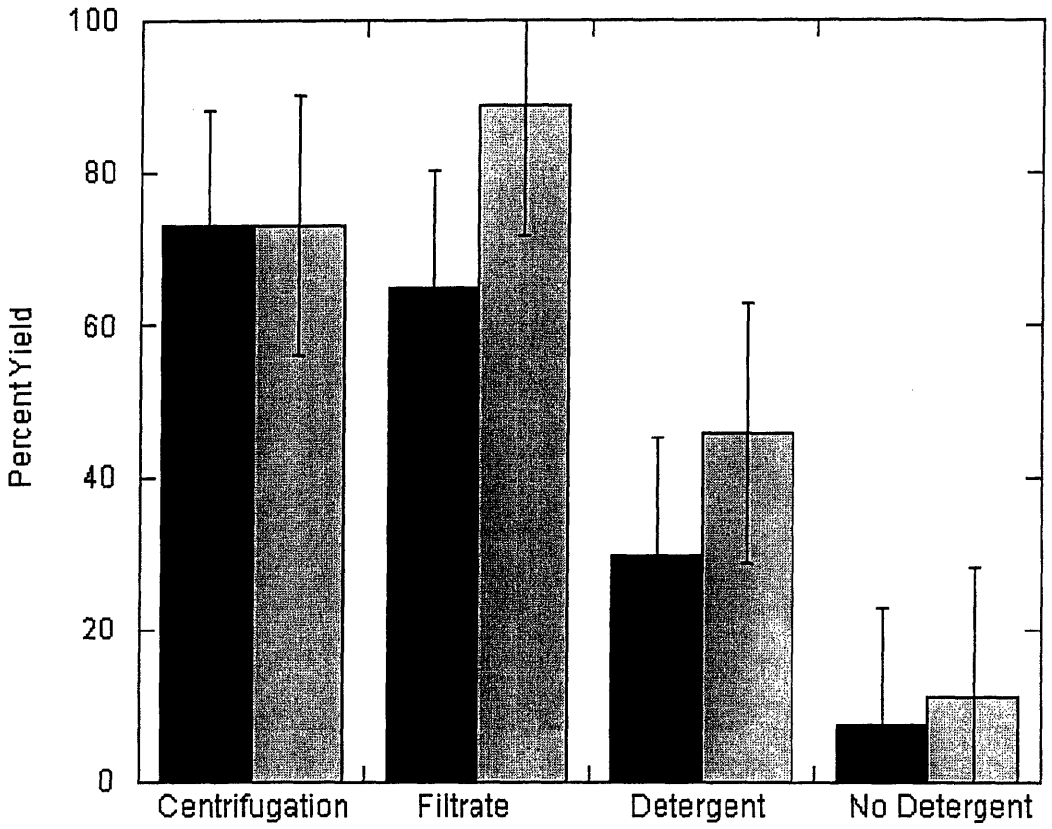


Figure 2-6: Syn5 particle yield during purification. Syn5 particle concentration was determined using SYBR green staining. The total particle concentration was calculated at different stages during the purification process, including the crude lysate, supernatant after centrifugation, supernatant after filtering, final concentration with detergent, and final concentration without detergent. Total yield (black) at each step is total number of particles remaining in the sample compared to the total number of particles present in the crude lysate. Specific yield (gray) at each step is the total number of particles remaining in the sample compared to the number of particles present at the prior stage of purification. Yields are the mean of three replicate samples. Error bars reflect standard error.

To confirm that the purified particles (shown as an opalescent band in a CsCl gradient in Figure 2-7) were identical in morphology to the Syn5 particles first characterized by Waterbury and Valois (1993), the phage particles were placed on carbon coated copper grids, stained with 2% uranyl acetate, and examined by TEM (Figure 2-8). The electron micrograph showed purified virus-like particles, with icosehedral heads and short tails similar to the phages identified as Syn5 at its initial isolation. No contaminating phages of different morphologies or intact cells were observed. Some of the phage particles were broken open, possibly due to the purification or staining process, but the majority appeared intact, with their genomic nucleic acid retained within the capsid.

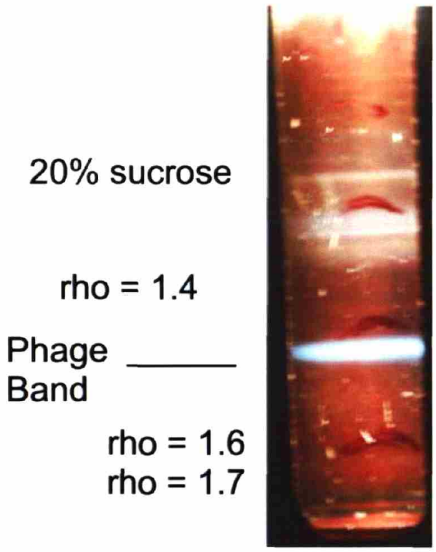


Figure 2-7: Opalescent phage band on CsCl gradient.
Purified Syn5 particles appear as a milky-white band between CsCl steps rho =1.4 and rho =1.6. Picture taken while CsCl gradient was illuminated in a Coombs Fractionator.

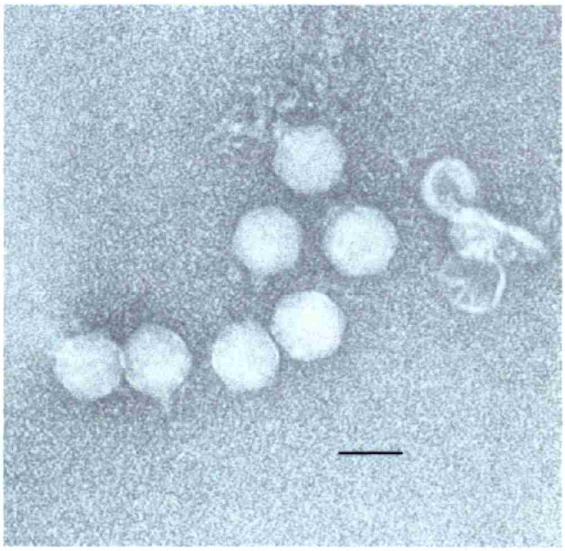


Figure 2-8: Electron micrographs of purified Syn5 particles.
Phage particles harvested from a CsCl gradient were dialyzed into 500mM NaCl, 100mM MgCl2, 50mM Tris pH 8.0. Particles were placed on glow-discharged carbon coated copper grids and stained with 2% uranyl acetate. Syn5 appears as a short-tailed phage, similar to T7 and P60. Some particles are broken open. Scale bar = 50nm

iv. *Double-stranded DNA as a genetic carrier*

To isolate Syn5 DNA, particles were purified as described in the Materials and Methods. Purified particles were gently broken open using heat and SDS, and the phage DNA was further purified from the particles proteins using a phenol:chloroform:isoamyl extraction. The DNA was then collected via an ethanol precipitation and centrifugation; and was resuspended in TE.

Earlier studies of cyanophage DNA indicated that some phage genomes are resistant to restriction by some enzymes (Chen and Lu, 2002; Wilson et al., 1993). Syn5 DNA does not appear to be modified, such as to confer resistance. It was possible to digest the Syn5 genomic DNA with restriction enzymes such as BamHI, HindIII, and XbaI (Figure 2-9). HindIII exhibited approximately six bands, XbaI generated five bands, and BamHI only two. Only XbaI appeared to have completely digested the genomic DNA. The size of the visible XbaI fragments indicated that the approximate genome length was 40-50kb. Both the BamHI and HindIII lanes appeared to contain undigested DNA. The fact that the genome DNA can be digested by these restriction enzymes also indicates that Syn5 particles utilized double-stranded DNA as a carrier of genetic information.

v. *Particle Protein analysis using SDS-Gel Electrophoresis*

The concentration of the phage particles was sufficient to examine the protein composition via SDS-polyacrylamide gel electrophoresis and Coomassie Brilliant Blue staining. Phage particles were boiled in SDS buffer to fully denature all phage proteins prior to electrophoresis. The samples were electrophoresed through a 10% polyacrylamide gel containing SDS. Syn5 exhibited 12 distinct polypeptide chains, ranging in molecular weight from ~140kDa to ~15kDa (Figure 2-10). As the intensity of each band is representative of the number of polypeptide chains within the particle, it is possible to estimate how many of each chain is present per particle. By inspection, the majority of the bands appear to be present in similar concentrations within the particle. However, one band (~31kDa) appeared far more intense, and probably represents the capsid protein, which in other phage is present in ~400-420 copies per particle.

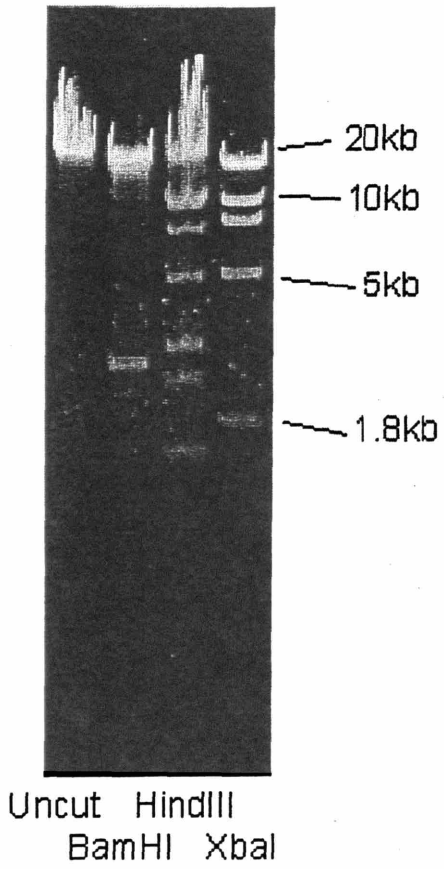


Figure 2-9: Agarose gel of digested Syn5 genomic DNA
Double-stranded genomic DNA was extracted from CsCl purified phage particles. The DNA was digested with BamHI, HindIII, or XbaI in the appropriate buffer for 1hr at 37°C and run on a 1% agarose gel in TAE buffer at 70v for 1hr. Undigested DNA is at the top of each lane.

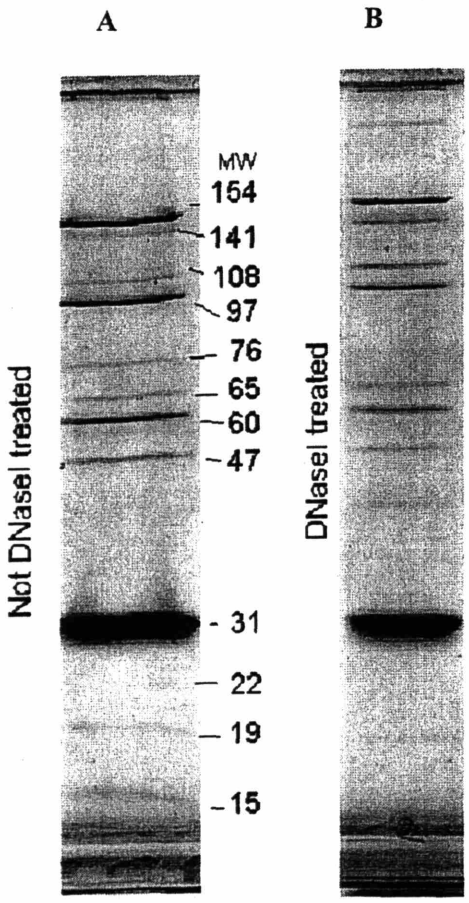


Figure 2-10: SDS-polyacrylamide gel of purified Syn5 particles
Purified Syn5 particles were denatured by boiling in SDS-sample buffer for three minutes and then electrophoresed through a 10% polyacrylamide gel. The proteins were stained with Coomassie Blue. 12 distinct bands are visible. Approximate molecular weights are indicated in kDa as calculated from movement through the gel. A) Particles purified without DNaseI. B) Particles purified with DNaseI.

Some cyanophage purification protocols have used a DNaseI treatment of the crude lysate prior to particle precipitation (Sullivan et al., 2005; Wilson et al., 1993) as a method of removing host DNA; however a distinct difference in protein banding patterns was discernible between particles purified with DNaseI and those without (Figure 9A and B). In the particles purified without DNaseI, a band with an approximate MW of 76kDa was visible, that was not visible in the DNaseI treated samples. The major capsid protein appeared to have suffered some proteolysis, as in the DNaseI treated samples, the MW was approximately 26kDa while in the non-treated samples the MW was approximately 31kDa. Also present in the DNaseI treated sample is a band of approximately 38kDa that has no analogous band in the un-treated sample; this band may be the result of proteolytic breakdown of the 76kDa band, the major capsid band, or both.

vi. *Thin-sections of Synechococcus infected with Syn5*

In an attempt to observe the interaction of Syn5 particles with the host, Syn5 infected cells were fixed, embedded, and ultra thin sectioned for electron microscopy (Figure 2-11A-E). Exponentially growing WH8109 was mixed with a 10-fold excess of phage and incubated for various time periods at room temperature prior to fixation. Infected cells were collected by low-speed centrifugation. Fixed pellets of phage and cells were embedded in resin, thin-sectioned, and examined by electron microscopy in an attempt to observe Syn5 particles both adhered to the surface of host cells and assembled in vivo.

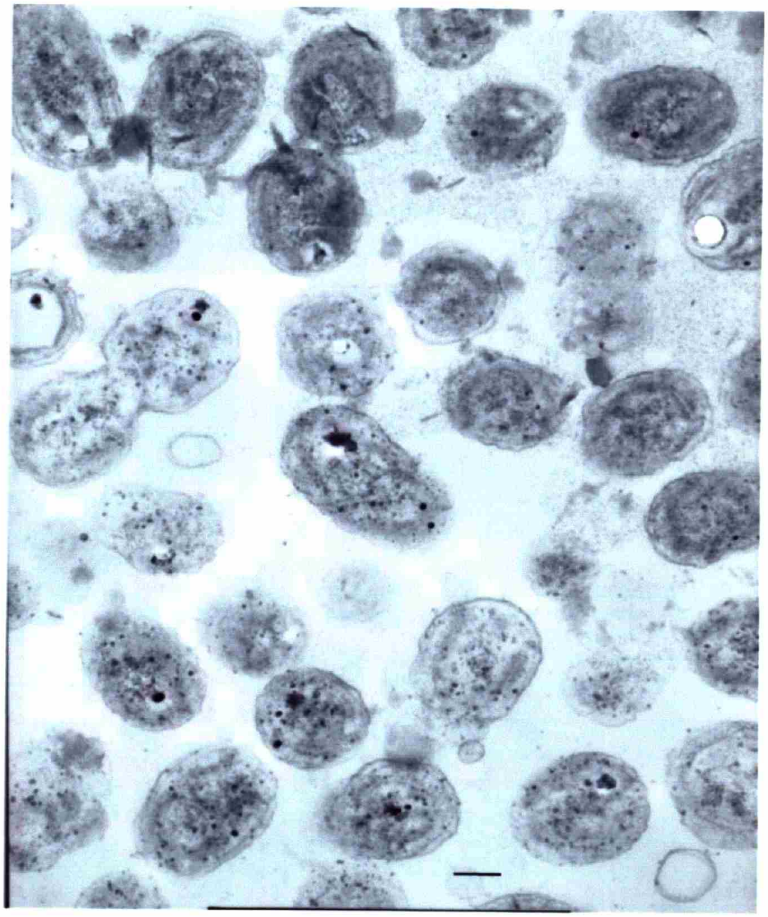
Figure 2-11 chronicles a time course of Syn5 infection. 2-11A shows uninfected control cells. Cells were intact and unruptured; the thylakoid membranes were visible within many of the cells. 2-11B shows the zero time point, in which phage and cells were mixed and immediately fixed with glutaraldehyde. The cells of the zero time point resembled the control cells, with intact visible cells. Only one phage particle was in evidence. The lack of visible phage may be an indication that fixation occurred prior to adhesion of most of the phage particles, as centrifugation to form the cell pellet was not sufficient to pellet free phage particles as well. 2-11C shows infected cells one hour after infection. Numerous phage particles were adhered to the surface of the host cells, indicating that sufficient time had passed for adsorption to occur. The thin sections were 70nm thick on average, and approximately $6.5 \times 10^6 nm^2$ of cell surface area was visible within these micrographs. 25 phage were observed adhered to cell surfaces. From the intensity of the staining, about 60% of the phage particles were empty of DNA while 40% still contained some DNA within their heads. No particles were yet assembled within the host cells, an indication that one hour is still within the eclipse period of the phage. At least one cell was lysed.

These electron micrographs of the thin-sections of WH8109 and Syn5 contained numerous black deposits which represent artifacts from the heavy metal staining process (Bozzola and Russell, 1992). These deposits were distinguishable from the particles attached to the cells both in terms of size (in some cases), stain density, and shape.

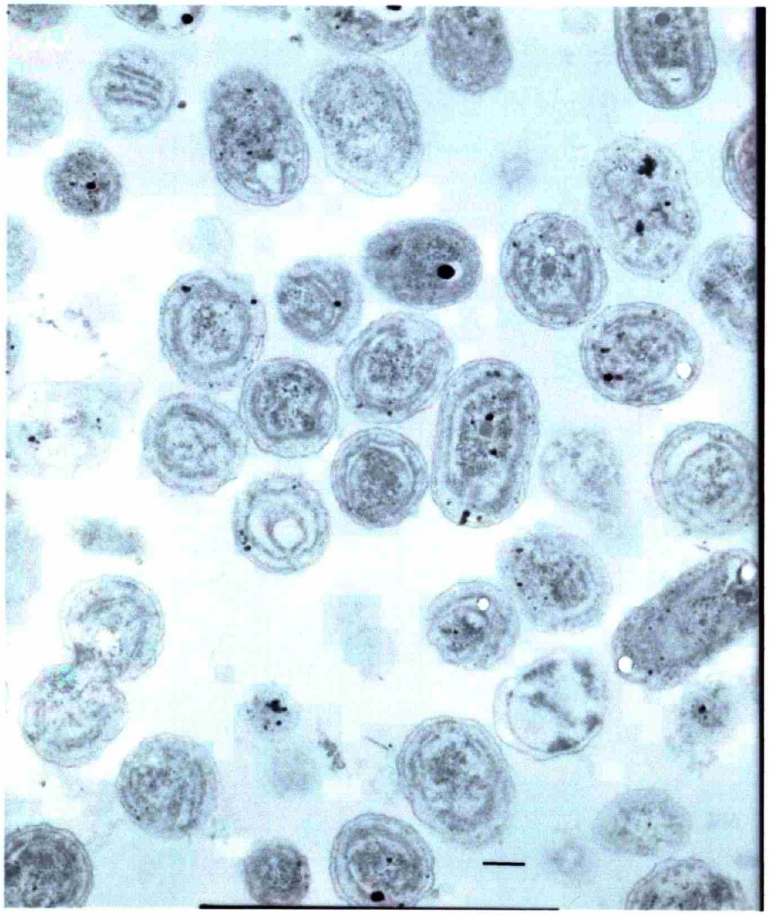
Nonetheless, their presence reduced the ability to monitor the development of new particles within the cells.

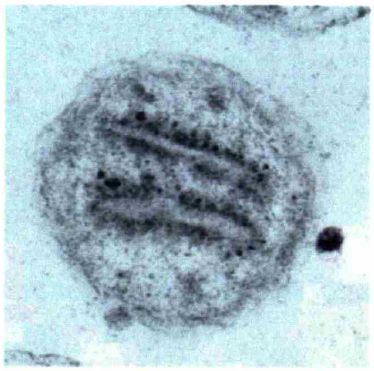
By two hours, 2-11D, 30% of the cells exhibited five or more assembled phage-like particles within the cells, distinguishable from the stain deposits noted above, with 70% of cells showing at least one internal phage particle. Some cells were visibly burst. By five hours, 2-11E, substantial numbers (~40%) of the cells within a frame were ruptured and particles that appeared to be free phage were associated with some of these membranes. Approximately 30 predominantly circular intact cells were visible within the frame. Three empty phage heads were visible attached to intact cells, no obvious phage particles were visible inside the intact cells. Given that full lysis did not occur until approximately 10 hours after infection, the ruptured cells found in the later time points may be partially due to sensitivity of the phage-containing host cells to the harsh fixation process.

Figure 2-11: Thin sections of WH8109 infected by Syn5
Time course of WH8109 cells infected by Syn5 in late exponential phase. A) Uninfected cells. B)Time zero: phage were mixed with cells and immediately fixed with gluteraldehyde. C-E) Phage and cells were incubated for one, two, and five hour periods prior to fixation.

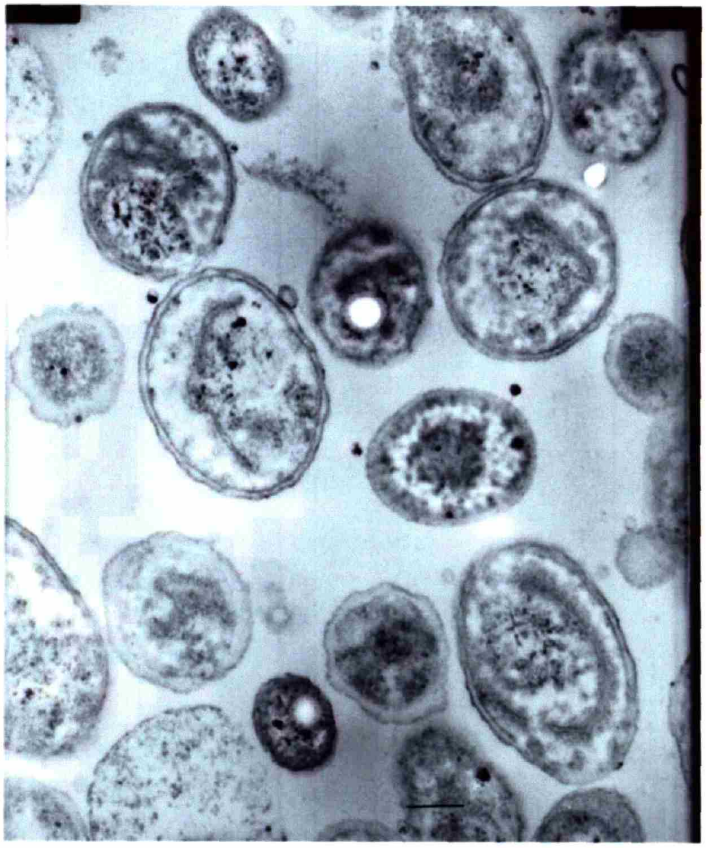


2-11A. WH8109. Scale bar = 200nm

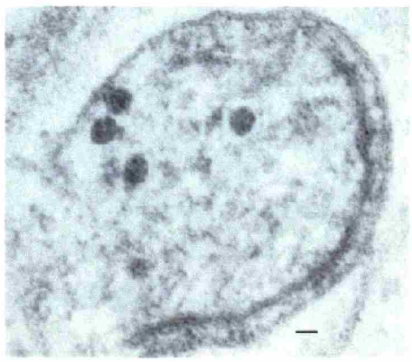




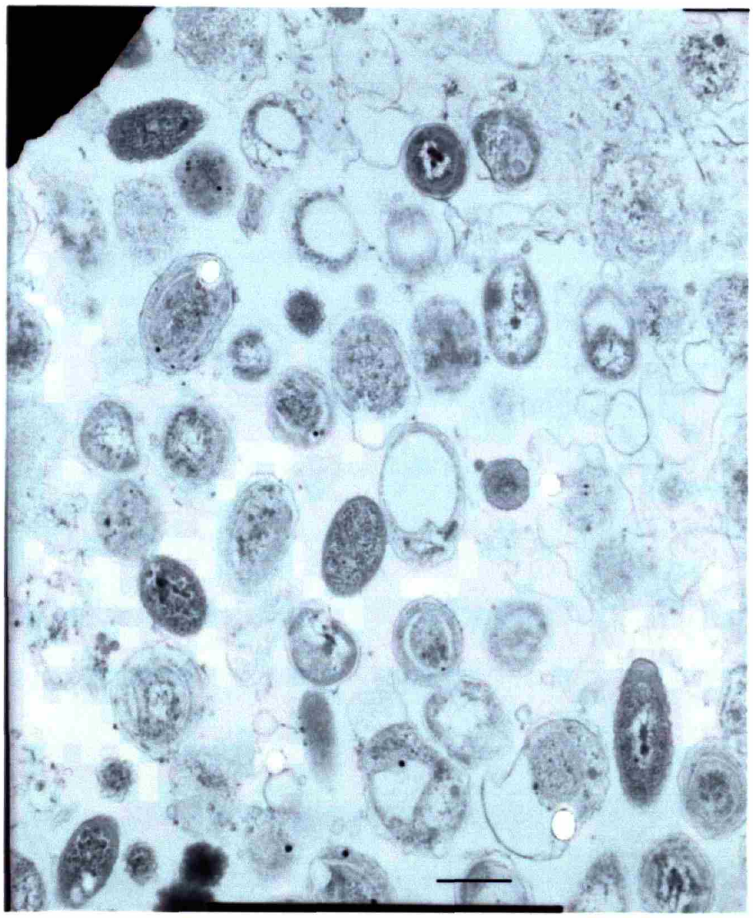
2-11B. WH8109 immediately after Syn5 infection (zero hours)
scale bar = 200nm

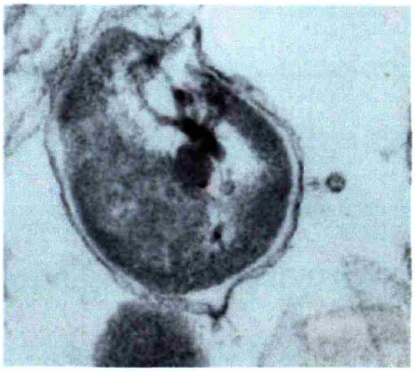


2-11C. One Hour after Syn5 Infection
Scale bar = 200nm



2-11D. WH8109 two hours after Syn5 infection
scale bar = 200nm





2-11E. WH8109 five hours after Syn5 infection
Scale bar = 500nm

D. DISCUSSION

Since the initial isolation of phages similar in morphology to enteric bacteriophage from the sea which infect cyanobacteria, numerous studies have been undertaken to determine their concentrations in the field, their genetic diversity, their host range, and their lifestyles. Further work has resulted in the genome sequences of five cyanophages, all of which exhibited some level of similarity to known enteric bacteriophage genes. However, difficulties in concentrating and purifying cyanophage particles has limited characterization of their proteins.

Our methods of *Synechococcus* growth and phage purification resulted in reliable production of significant quantities of highly concentrated cyanophage particles (10^{12} particles/ ml) that remained intact and infectious despite exposure to CsCl and various detergents. Using these samples, it was possible to initiate more synchronous infections and elucidate phage/host dynamics, such as burst size at various points within the growth curve of the host cell. The Syn5 burst size, ~30 phage per cell, under replete conditions from rapidly growing cells is less than that of enteric bacteriophage. This may simply be because the host cells were not growing as quickly as replete enteric cells, as phage burst size, eclipse period, and latent period have all been shown to be dependent on the growth rate of the host cell; and more specifically, dependent on the rates of DNA and protein synthesis and independent of cell size (Bohannan and Lenski, 2000; Hadas et al., 1997; Rabinovitch et al., 2002). In the field, under nutrient limited conditions where

Synechococcus cells are growing more slowly on a diel cycle, the Syn5 burst size may be smaller substantially smaller while the latent and eclipse periods may be longer than the calculated laboratory numbers. Observations supporting this have been made in the field by (Steward et al., 1996)

Electron micrographs of Syn5 show particles with an icosahedral head approximately 50nm in diameter and a short tail approximately 10nm in diameter and 20nm long; closely resembling T7-like podoviridae. The thin-sections of Syn5 interacting with its host, WH8109, demonstrate that these particles bind to the outer surface of the host cells within one hour. While bound, the phage particles appear to have a longer tail than unbound particles, which could indicate the presence of an internal T7-like extensible tail. During T7 infection, three proteins found inside the capsid of the free phage particles form a bridge from the capsid through the outer membrane of the host cell, facilitating the initial entry the phage DNA to the cell (Kemp et al., 2004). Similar proteins may exist within Syn5, and aid infection of WH8109 in the same manner.

Time course pictures of Syn5 infection appear to indicate that the eclipse period of Syn5 lies between one and two hours. Resolution of the latent period via the thin-sectioning time course is more difficult, as cells at the two hour time point appeared to contain assembled phage particles and cells at the five hour time point did not. This could indicate that under these conditions, the Syn5 latent period is between two and five hours, and that the method of titering Syn5 particles by SYBR green staining grossly overestimated the number of infectious particles within the phage stock by two or three

fold. Some phages have tail fibers or other structural proteins that are necessary for infection, yet are quite fragile and can be broken or damaged during ultracentrifugation. As SYBR green staining counts all DNA containing phage particles and does not differentiate between infectious and non-infectious particles, it is possible that the titer of the phage stock was substantially lower than we believed, in which case several rounds of infection would occur prior to complete lysis of the cell culture. An alternate explanation is that phage-containing cells are more likely to burst during the fixation process. This could concentrate the numbers of uninfected cells and/or phage resistant cells in the pellet, as phage burst from cells would likely remain soluble and would not be embedded. This can be resolved by thin-sections of more time points during the phage growth cycle, using a gentler fixation process.

Syn5 DNA was digestable by each three enzymes tested; appearing to have at least two, six, or eight recognition sites for the enzymes BamHI, XbaI and HindIII respectively. The number and approximate sizes of these bands indicate that the genome may be approximately 40-50kb long. However, the presence of uncut DNA at the top of each lane makes it difficult to resolve the true length of the genome. The phage genome appeared to be a linear molecule, as no super-coiled or nicked DNA was observed in the gel. Some cyanophage DNA has been demonstrated to be resistant to certain restriction enzymes (Wilson et al 1993; (Chen and Lu, 2002), including HindIII and EcoRI. This resistance is not limited to the phage, as some host strains have also demonstrated resistance to certain enzymes. This resistance may be the result of modifications made to the DNA. Syn5 DNA appears sensitive to restriction enzyme digestion, indicating that it

is likely unmodified. However, it is possible that partial modification has taken place, accounting for the fraction of DNA that remained undigested.

Syn5 particles exhibit 12 distinct bands when electrophoresed through a SDS-polyacrylamide gel, ranging from approximately 154kDa to 15 kDa. The majority of these bands appear to contain similar numbers of polypeptide chains, with the exception of an intense band of approximately 31kDa that likely represents the major capsid protein. T7 particles generate 9 protein bands on a polyacrylamide gel, including two forms of the capsid protein, three internal virion proteins, a tail fiber, a portal, and two tail tube proteins (Dunn and Studier, 1983). P22 particles are comprised of similar proteins: capsid, portal, tailspike, three internal virion proteins, and three neck proteins, although some of these are few in copy number and may not be detectable without a sensitive stain (Prevelige and King, 1993). The capsid band is the most intense band, as the capsid protein is present in 415 copies per particle, while the other structural proteins are present in 3-20 copies per particle. Syn5 has twelve bands, indicating that if it is truly a member of the T7 family, it contains three additional structural proteins.

Since the identification of the first cyanophage and the discovery of their prevalence within the surface waters of the ocean, researchers have primarily focused on the ecology of these phages. With the development of a robust phage/host laboratory system, biochemical characterization of the structural proteins of cyanophage and comparison to the morphologically identical proteins of enteric phage becomes possible.

Future directions include genome sequencing and annotation of Syn5 and identification of the proteins contained within the particle.

CHAPTER III:

GENOME SEQUENCE OF THE MARINE CYANOPHAGE SYN5

A. INTRODUCTION

Analyses of enteric phage genomes have demonstrated several conserved properties. Within enteric phage genomes, genes are organized into several classes, termed "early", "middle", and "late" based on the time during infection at which they are transcribed. Once the phage binds to a host cell during infection, the linear phage DNA enters the cell where some genes are immediately transcribed by the host's RNA polymerase. These genes, termed "early genes", influence the take-over of the host cell. In the case of temperate phages, genes which allow the phage genome to incorporate into the host's genome are transcribed. Whether a prophage is formed, or whether the lytic cycle progresses depends on many factors, both environmental and physiological. Some early lytic genes may encode enzymes which destroy host DNA or outer membrane proteins to prevent super-infection by another phage. "Middle" genes are transcribed partway through infection; these genes produce proteins involved in DNA replication, such as DNA polymerase, exonucleases, endonucleases, primases, and helicases. The "late" genes include the structural proteins which both form and aid the assembly of the phage particles, as well as genes that encode the DNA packaging genes. Each set of genes generally has its own promoter(s), thus enabling them to be transcribed during the proper time during infection and in the proper stochiometry. Therefore, these classes of genes generally appear in cassettes or modules with a given phage genome: for example,

DNA replisome genes are not interspersed with structural genes (Brussow and Hendrix, 2002) (Figure 3-1).

Double-stranded DNA phages have several elements in common during DNA packaging. At the initiation of assembly, phages form a "procapsid"; a DNA-less structure comprised of a full set of portal, capsid, and in some cases accessory proteins. The procapsid is smaller in diameter than a mature capsid, and appears spherical rather than icosehedral when examined by negative staining under EM. Scaffolding proteins aid in the nucleation and assembly of the procapsid, but are proteolytically cleaved or exit intact through the procapsid shell at or near the time prior to DNA packaging. DNA is highly ordered within the shell of the mature particle, forming a solenoid normal to the axis of the tail, and extremely dense (0.5 gm/cc). Within the head of the phage, and during transfer from capsid to host cell, the phage DNA is a linear molecule. Once the DNA is within the capsid, several proteins form a "plug" to block the hole in the portal ring. The tail fibers then attach to these neck proteins (Prevelige and King, 1993) (Figure 3-2).

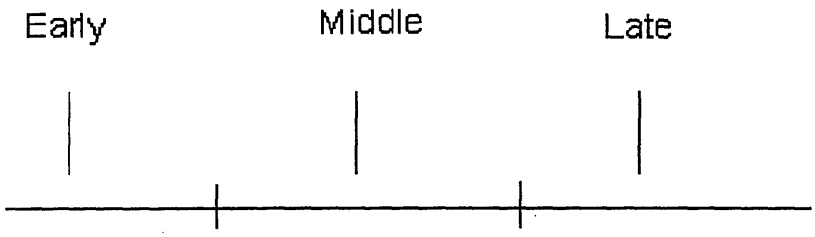


Figure 3-1. A Generalized Phage Genome. Double-stranded DNA tailed phages contain several conserved elements within their linear genomes. These "cassettes" are transcribed at different stages during the infection process. "Early" genes include genes which take over the host cell, "Middle" genes are involved in phage genome replication, and "Late" genes contain those that encode the phage structural proteins, and cell lysis proteins.

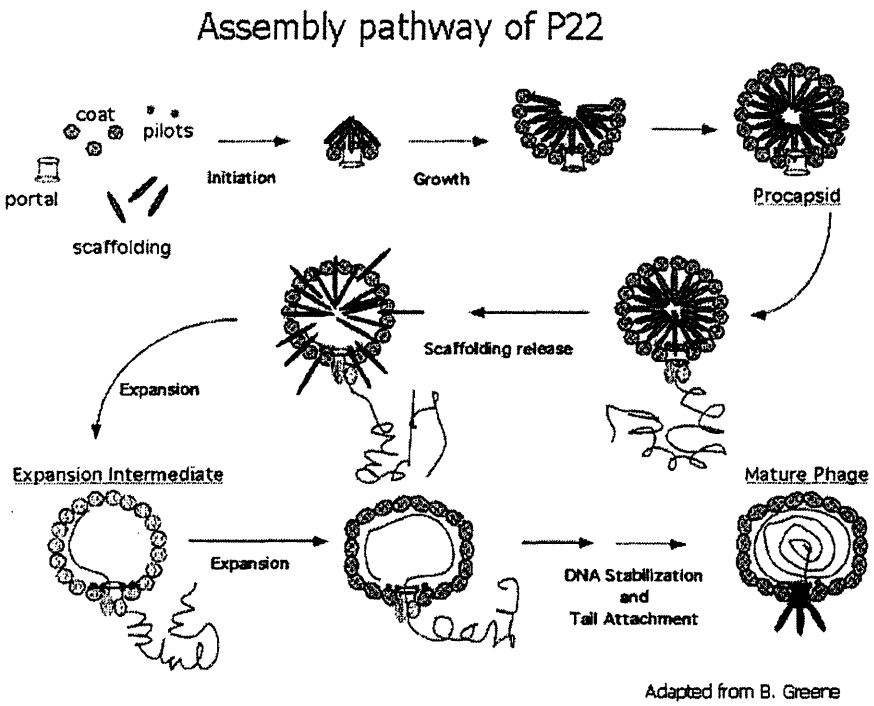


Figure 3-2. Assembly pathway of Phage P22. The assembly pathways of well-characterized Caudovirales are similar. P22 is shown as an example: First the capsid, scaffolding, pilot (or internal virion), and portal proteins form a nucleus. Assembly of the remainder of the procapsid is assisted by more scaffolding subunits. As the terminase binds the portal with the phage genome, the scaffolding exits through pores in the procapsid shell, and the capsid expands to its mature icosehedral shape. The terminase dissociates from the DNA, and the neck (or tail tube) proteins are added, forming a plug to prevent the DNA from spilling out through the portal. Lastly, the tail fibers are added.

Newly synthesized phage DNA is replicated from concatemers, which are formed early in infection through recombination of discreet terminal repeats or between chromosome ends and homologous sequences in partner molecules. During DNA packaging, accessory proteins called "terminases" bind to the phage DNA and may utilize ATP to pump the DNA into the phage head. The terminase then cuts the DNA, either at a specific site, or after the mature capsid is filled with DNA, termed "headful" packaging. In headful packaging, slightly more than "one" genome may be packaged, thus ensuring that the linear DNA molecule within the phage capsid exhibits some terminal redundancy. This circular permutation of the genome allows terminal end recombination and the formation of new concatemers in the next phage infection (reviewed in Earnshaw and Casjens, 1980).

Analyses of groups of bacteriophage genomes have indicated that all phage may be part of a global gene pool (Hendrix et al., 1999). Within the long non-contractile tailed phage that infect Mycobacterium, cassettes of similar genes appear within genomes of phages isolated from separate areas even though overall gene order is not conserved (Pedulla et al., 2003). This genome mosaicism appeared to be the result of both homologous and non-homologous recombination between phages and microbes. Similar mosaicism is apparent in the T4-like myoviridae (Hendrix et al., 1999). The T7-like podoviridae, in contrast, exhibit comparably high levels of synteny, from the presence of a phage encoded RNAP at the early end of the genome through the number, length, and order of the structural genes at the tail end of the genome. Cyanophage genomes exhibit similar mosaicism to the enteric myoviridae; containing cassettes of genes similar to

morphologically identical enteric phages, as well as genes similar to their photosynthetic hosts (Mann et al., 2005; Sullivan et al., 2005), indicating that the cyanophages are not isolated from enteric populations and therefore, may receive or distribute genes from sources other than their hosts.

To date, the genomes of five cyanophage have been sequenced and annotated; all of which infect either *Prochlorococcus* and/or *Synechococcus*. The genomes of the myoviridae P-SSM4, P-SSM2 and S-PM2 (the former two which were isolated on *Prochlorococcus* NATL2A and *Prochlorococcus* NATL1A, respectively, from the Sargasso Sea; and the latter on *Synechococcus* WH7803 from the English Channel) exhibited genes in common with the enteric bacteriophage T4, while the short-tailed phage P-SSP7 (isolated on *Prochlorococcus* MED4 from the Sargasso Sea) resembled T7 (Mann et al., 2005; Sullivan et al., 2005). The short-tailed *Synechococcus* phage P60 (isolated on WH7803 from a Georgia estuary), while morphologically identical to T7 and possessing some genes in common with T7, was lacking several of the T7 core genes as well as synteny within the phage structural genes (Chen and Lu, 2002).

Syn5 is a short-tailed cyanophage isolated from the Sargasso Sea on *Synechococcus* strain WH8109 (Waterbury and Valois, 1993). The phage has not been demonstrated to infect any other strains of *Synechococcus* or *Prochlorococcus* (Sullivan et al., 2003; Waterbury and Valois, 1993). Preliminary analysis indicated that it was a dsDNA tailed phage, morphologically similar to T7, P60, and P-SSP7 when negatively stained under EM. As described in Chapter II, it has been grown to high titer in cultured

WH8109 and purified in the lab thus enabling the study and analysis of phage DNA and particle proteins. Its lytic period of approximately 10 hours likewise resembles P60. Syn5 is the first open ocean short-tailed cyanophage isolate that appears to infect solely *Synechococcus* to be sequenced.

B. MATERIALS AND METHODS

i: *Phage Preparation*

Phage particles were concentrated and purified as in Chapter II. Briefly, *Synechococcus* strain WH8109 was grown in SN media at 26°C and 50μEm^{-2}s^{-1} supplied by cool white bulbs under constant aeration to a density of 10^8 cells/ml and infected with Syn5. After lysis, as determined by optical density and a change of culture color from orange to green, 2.5 liters of crude lysate was centrifuged and filtered to remove cellular debris. The supernatant was treated with additional 0.5M NaCl and polyethelene glycol (PEG) 8000. The phage were allowed to precipitate for 2 to 12 hours at 4°C. The precipitate was collected via centrifugation and then resuspended in ~15ml of 75% FSW. This solution was layered on top of a CsCl step gradient containing steps of ρ=1.6, ρ=1.4, and 20% sucrose and then ultracentrifuged for 4 hours at 28K at 4°C. The phage band visible at the interface between the ρ=1.4 and ρ=1.6 layers yielded approximately 1ml of phage at a concentration of 10^{12} particles/ml.

ii. *DNA Sequencing*

Syn5 was sequenced at the Pittsburgh Bacteriophage Institute at University of Pittsburgh (Pittsburgh, PA). Methods for DNA sequencing and contig assembly have been reported elsewhere (Dobbins et al 2004, Pedulla et al 2003). Briefly, 10μg genomic phage DNA was sheared hydrodynamically and repaired with T4 ligase to produce blunt ends. 1 to 3kb fragments of blunt-ended DNA were ligated into the EcoRV site of the pBluescript II KS+vector. Ligated plasmids were transformed into *E. coli* XL1-Blue cells. Plasmids were recovered from clones via QiaPrep plasmid purification kit (Qiagen, Santa Clarita, CA). Inserts were sequenced from both ends using the Applied Biosystems BigDye v3.0 dye terminator chemistry and universal sequencing primers. Labeled reaction mixtures were separated and analyzed using an ABI Prism 3100 DNA analyzer. Underrepresented areas of the genome were covered and assembly was achieved by design of oligonucleotide primers and whole genome template in further sequencing reactions. Approximately eight-fold coverage was obtained.

iii. *Sequence Assembly and Analysis*

Sequence chromatograms were assembled and analyzed using the Phred, Phrap and Consed sequence analysis software (Gordon et al., 1998). The Syn5 terminal repeat was verified by primer extension. The completed genome sequence was assembled in the proper orientation from the Consed assembly. Open reading frames were identified using GeneMark (Besemer et al., 2001), Glimmer (Delcher et al., 1999; Salzberg et al., 1998), and DNA Master (J. G. Lawrence) (http://cobamide2.bio.pitt.edu) software and visual

inspection. Translated ORFs were compared with known protein sequences using BLASTp (Altschul et al., 1990). Sigma70 promoter sequences were identified using the BPROM program; rho-independent terminators were identified using FindTerm; potential hairpins were identified with BestPal, all found at http://www.softberry.com. Transmembrane helix regions were identified with TMHMM (Krogh et al., 2001).

C. RESULTS

Syn5 particles were purified using CsCl gradients and centrifugation as outlined in Chapter II. Concentrated particles were broken open and the DNA isolated using a phenol-chloroform extraction as outlined in Chapter II. The DNA was hydrosheared and cloned into a vector for sequencing, and then sequenced until eight-fold coverage was obtained as stated in the Materials and Methods. The sequenced DNA was assembled using the Phred, Phrap, and Consed sequence analysis.

The genome of Syn5 is comprised of double-stranded DNA, with a 55% G-C content. It is 46,214 bp including a 237bp terminal repeat. Using the computer programs GeneMark, GLIMMER, and DNA Master, it was possible to identify 61 open reading frames (Figure 3-3, Table 3-1). Putative sigma70-like promoter sequences in the early portion of the genome were identified using the BPROM program (http://www.softberry.com) which is designed to find enteric RNAP promoter sites; however, cyanobacterial promoters may differ from enteric promoter sequences and thus

are difficult to predict with high accuracy. Putative Syn5 RNAP promoter sites were assigned based on identification of a conserved sequence found within the genome at locations similar to the locations of the T7 RNAP promoter sequences. Rho-independent terminator sequences were predicted using the FindTerm program (http://www.softberry.com) (see Transcription for details on promoter and terminator assignments).

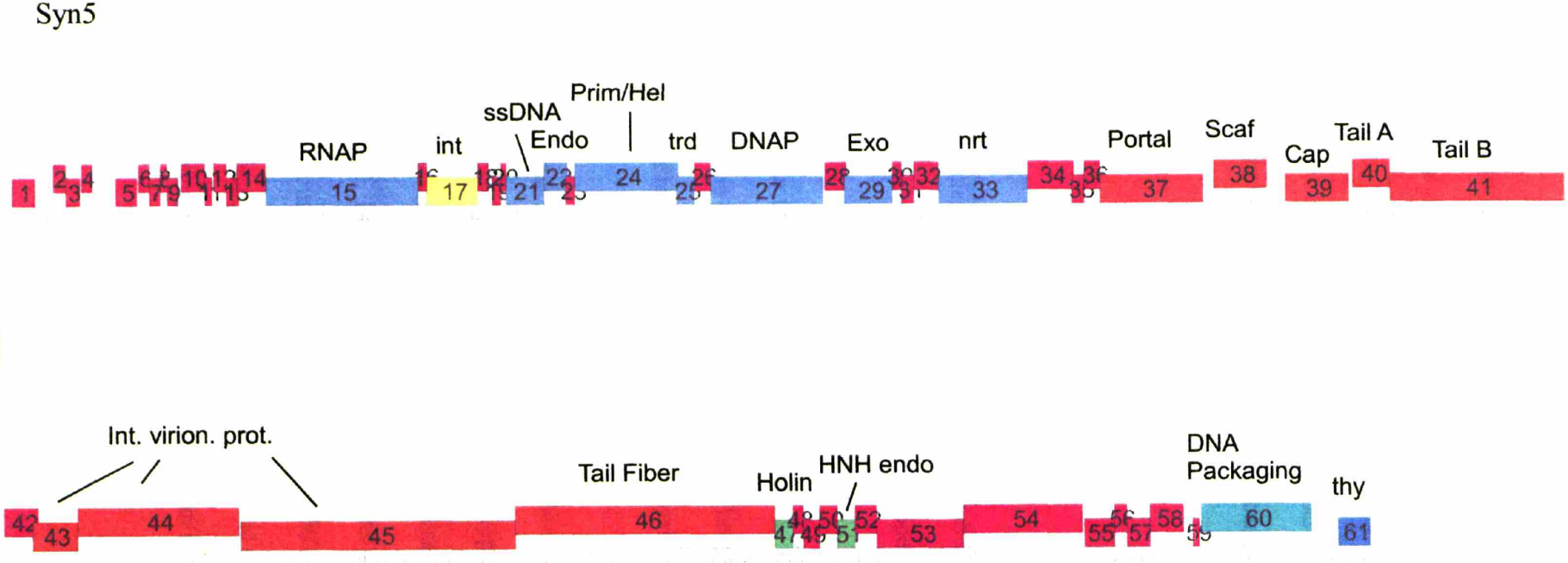


Figure 3-3: Genomic map of putative Syn5 ORFs. Syn5 is 46,214 bp with a 220 bp terminal repeat. 61 ORFs are labeled. Genes identified with BLASTp as having similarity to known phage DNA replication genes are in blue; genes with similarity to known phage structural proteins are in red, and non-structural late genes are in green. Unidentified genes are in magenta. Also labeled are the putative integrase (int) (yellow), ribnucleotide reductase (nrt), thioredoxin (trd), and thymidylate synthase (thy) genes.

i. *Gene Assignment*

Translated Syn5 ORFs were compared to known protein sequences within the non-redundant (NR) genomic database at NCBI using the BLASTp program. All ORFs are listed in Table 3-1; ORFs encoding sequences that matched known proteins with an e-value less than 0.001 are listed in Table 3-1 along with their putative function and the phage or microbe which contains the most similar sequence. Syn5 appears to contain similar genes to the core genes of enteric phage T7, including those used in DNA replication and those encoding structural proteins (Table 3-2). Additional ORFs that are similar to genes found in the cyanophages P-SSP7, P60, and S-PM2 include those that encode ribonucleotide reductase, thymidylate synthase, and thioredoxin, all thought to be used during phage DNA replication to scavenge host nucleotides in nutrient poor waters (Mann et al., 2005; Sullivan et al., 2005). The phage-encoded thioredoxin may have a dual role in DNA replication, both during ribonucleotide reduction, and secondly, to increase the binding affinity of the DNAP to template DNA. The T7 DNAP (as well as Syn5 DNAP) belongs to the Type1 DNAPs, which are generally involved in DNA repair and have low processivities. By utilizing host thioredoxin as "sliding clamps", the T7 DNAP processivity increases 1000 fold (Bedford et al., 1997). The phage-encoded thioredoxin in Syn5 may interact with the phage DNAP in a similar manner.

Holins are proteins that make holes in bacterial cell membranes during phage infection, to allow phage encoded lysozyme to degrade the cell wall. Their regulation

may play a role in the determination of the end of an infection. The putative holin gene, although lacking similarity to known holin sequences, has the physical characteristics of the lambda S holin. In lambda, two forms of the S holin (beginning at aa1 and aa 5) are expressed in a two to one ratio; with the longer form acting as an inhibitor of the shorter. The expression of the two forms is controlled by a stem-loop located directly over the ribosome binding site of the gene (Wang et al., 2000).

ORF47 of Syn5 exhibits all of these characteristics: a transmembrane domain at the N-terminus of the protein, positively charged residues immediately preceding the transmembrane domain, and a putative regulatory loop prior to the gene. Finally, the putative holin has a similar location in the genome of Syn5 as the T7 holin does in its genome (Dunn and Studier, 1983).

With the exception of P60, all of the other sequenced cyanophage contain at least one gene which encodes a photosynthetic protein (generally psbA) (Lindell et al., 2004; Millard et al., 2004); however, Syn5 contains no identifiable photosynthesis genes.

Table 3-1: Identified ORFs within the Syn5 Genome

ORF	Predicted Protein	Related Phage(s) or Microbes	E value
1	--	P60	0.004
2	--	--	--
3	--	--	--
4	--	--	--
5	--	*Neurospora crassa*	0.36
6	--	--	--
7	--	*Gloebacter violeceus*	9.0
8	--	--	--
9	--	--	--
10	--	--	--
11	--	--	--
12	--	--	--
13	--	--	--
14	--	P60, *Synechococcus elongatus*	4e-05
15	RNA Polymerase	P60, P-SSP7, Vibriophage VP4, T7	5e-161
16	ATPase, permease?	Lactococcus	5.2
17	Integrase	P-SSP7, *Agrobacterium, Geobacillus, Burkholderia*	9e-26
18	--	--	--
19	--	--	--
20	--	--	--
21	T7-like ssDNA binding protein	P-SSP7	9e-17
22	Endonuclease	P-SSP7, Vibriophage VP4, *Pseudomonas putida KT2440*	9e-26
23	--	--	--
24	Primase/Helicase	P60, P-SSP7, SIO1, phiYeO3-12, T3, *Pseudomonas putida*, T7	0.0
25	Thioredoxin	P60	2e-6
26	--	S-PM2	9.0
27	DNA polymerase	P60, Podovirus GOM, Podovirus SOG, *Pseudomonas putida*, Vibriophage VP4, T7	0.0
28	--	P-SSP7	4e-6
29	Exonuclease	P60, P-SSP7, *Pseudomonas putida*, phiYeO3, T3, T7	1e-86
30	--	--	--
31	--	P60; Staphylococcus aureus phages 47, 96, 3A, phiN315, ROSA, 85, 53; Listeria phage 2389	3e-15
32	ATPase w/chaperone activity?	P60, phiJL001	6e-25, (0.12 for protein function)
33	Ribonucleotide reductase	P-SSP7	9e-162
34	--	P-SSP7	5e-62
35	--	P60, P-SSP7	2e-07
36	--	--	--
37	Portal	P60, P-SSP7, *Yersinia pestis* phage phiA112, *Pseudomonas putida*, T7	0.0

38	Scaffolding	P-SSP7, P60, Vibriophage VP4, T3, K1-5, T7	5e-30
39	Capsid	P60, P-SSP7, *Pseudomonas putida*, T7	6e-74
40	Tail tube A	P-SSP7, *Pseudomonas putida*, P60, T3, phiYeO3-12, T7	6e-26
41	Tail tube B	P60, P-SSP7, *Pseudomonas putida*, Vibriophage VP4, T7	4e-140
42	--	P60, P-SSP7, epsilon15	2e-25
43	Internal Virion Protein (gp 14)	P60	1e-10
44	Internal Virion Protein (gp 15)	P60	3e-27
45	Internal Virion Protein (gp 16)	P60, P-SSP7	5e-66
46	Tail fiber	P-SSP7 (N-term 100aa)	4e-19
47	--	P60	3e-13
48	--	--	--
49	--	P60, P-SSP7	2e-22
50	--	P60	3e-15
51	HNH endonuclease	P60, P-SSM2, P-SSP7	8e-33
52	--		
53	Large protein involved in aggregation	*Microbulbifer degradans*	0.001
54	Outer membrane protein, putative RTX toxin	*Burkholderia cepacia* R18194, *Shewanella oneidensis* MR-1	3e-06
55	--	--	--
56	--	--	--
57	--	--	--
58	Fiber	P-SSM4, P60	0.054, 0.21
59	--	--	--
60	DNA maturase	P-SSP7, P60, *Pseudomonas putida*, T3, phiYeO3-12, T7	0.0
61	Thymidylate synthase	*Prochlorococcus* marinus strain 9313, P60, S-PM2, P-SSM2, P-SSM4, Mycobacterium phage L5	4e-51

Table 3-1: Identified ORFs within the Syn5 genome
Each putative ORF within the Syn5 genome was compared to known protein sequences within non-redundant database using BLASTp. The most likely putative function of each predicted protein is listed, as well as the phage(s) and/or microbe(s) which contained the most closely related sequence(s). The phage or microbe which contained the most closely related sequence, as determined by lowest e value, is listed first and that e-value is reported.

	%Gene Identity (%Similarity)			
	≥50%	**40-50%**	**30-40%**	**20-30%**
T7		Endonuc. 45 (62) Terminase 45 (63)	Prim/hel 34 (51) DNAP 32 (47) Portal 33 (54) n-term tail fiber 34 (43)	RNAP 25 (39) Exonuc 27 (45) Scaffold 26 (43) Capsid 26 (43) Tail Tube A 28 (51) Tail Tube B 23 (38)
P-SSP7	Portal 50 (67) Terminase 62 (77) Nrd 61 (74) ORFs 34, 49	DNAP 47 (64) Endonuc. 48 (65) Prim/hel. 47 (64) Exonuc 42 (62) Capsid 42 (59) n-term Tail fiber 47 (62) ORFs 31, 42	RNAP 35 (53) Int 32 (49) Scaffold 32 (48) Tail Tube A 36 (50 ORFs 21, 27, 28	Tail Tube B 27 (43)

Table 3-2: Similarity and Identity of Translated Syn5 ORFs to Translated T7 and P-SSP7 ORFs. Translated Syn5 ORFs were compared to the nr database using the program BLASTp. ORFs which displayed 50% or more similarity to T7 and P-SSP7 ORFs are listed. The first number is percent identity and the second number is percent similarity between the two ORFs. When the putative product of an ORF is unknown, the ORF number is used instead.

ii. *Transcription*

During T7 infection, the first 850bp of the genome, a stretch preceding the phage encoded RNAP gene, enter the cell first. Three *E. coli* RNAP promoters are located within these base pairs. These are recognized and transcribed by *E. coli* RNAP, facilitating the entrance of the rest of the genome to the cell (Dunn and Studier, 1983). Given the high level of synteny between the two phages, it seems plausible that the mechanism of infection of their respective host cells is likewise conserved. This would indicate that there may be *Synechococcus* RNAP promoters located within the first ~4000 bp of the Syn5 genome, prior to the phage-encoded RNAP. Cyanobacterial sigma70-like promoters exhibit high levels of similarity to the sigma70 promoters of *E. coli* (Curtis and Martin, 1994). Therefore, it was possible to screen for these types of promoters using programs designed to search for these sites in enteric sequences. Indeed, a number of early genes appear to have sigma70-like promoter sequences (Table 3-3), indicating that they may be transcribed by the host's RNAP.

The later T7 ORFs are transcribed by the phage-encoded RNAP and therefore are located behind one of phage's 17 RNAP promoters. These both aid in the faster translocation of the phage genome into the host cell and are responsible for the transcription of the phage genes involved in DNA replication and as well as the structural proteins. These promoter sites are highly conserved ~30bp stretches culminating in ribosome binding sites and are located directly before genes such as the capsid protein, the scaffolding protein, the ssDNA binding protein, as well as directly after the RNAP

gene (Dunn and Studier, 1983). By aligning the analogous sequences in Syn5, it is possible to identify an approximately 10 bp sequence: 5'-CCTTAATTAACT-3'. This sequence, or a similar one, appears at several other sites within the middle/late part of the Syn5 genome (Figure 3-4). The earliest copies of the putative promoter sequences appear before the T7-like ssDNA binding protein and within the DNA primase/helicase gene. Other sites include before some unidentified ORFs located between the DNA replication genes and the structural proteins, as well as in the middle of the tailfiber gene. There is no obvious protein that could be expressed from this last promoter, so it is unclear exactly what its putative function may be. However, both SP6 and T7 (Dobbins et al., 2004; Dunn and Studier, 1983) also have promoter-like sequences located within the middle of functional genes, so the presence of a promoter-like sequence in the middle of a putative gene does not indicate that the putative promoter sequence is incorrect.

ORF #	Promoter Sequence (-35 to -10)	Location (bp)
1	TTGCTGCGCCAAACCCACTGCAGCGTAAGGGATCTCACTAGATG	119
2	GTGGCACATGTGGCCCATTGCGTGTATGATATG	760
3	TTCGCGTGGAAGCCTGAGAACTTCACGATCGAGGAGATCTGATTG	932
4	TCTACTTCAGCTTCAACTGATCACCATCATCACCGCACCAACCATG	1187
6	CGCACACCTCGGCTTCTAATCACCATCACGCCATG	2071
7	GTCACTCACTCCTTCCCTTCCATCTGATCATG	2241
8	CTTCCACGAGATGTTCACCGCCTTCAACTGATG	2422
9	AAGCCCGAGTCCATCTACCGCAAGTACTGGAGGACAGCCAAGTG	2512
10	CTGGCCCTTATGGGCCTATTCACAATCATTTCGCACCGCAATG	2747
12	TTCGCCTACCACCTCGCTCTCCACAACGCTCATCCACCCATG	3244
13	TGGCCGAGCTCACCGCTCGTGGCCTTTGGCTCGGAGAGGAGTCATG	3436
14	TTGACGAACTCGAGGGCATCCACCAACAACAGTTCGACCCACCATG	3624
15	CCGAGATCGCTGGATTCCTGAACACCTACTGCAACTGATACATG	4079

Table 3-3: Putative sigma70 Promoter Sequences in the Syn5 Genome
Early ORFs with the Syn5 genome were examined with DNA Master to predict putative sigma70-like promoter sites. The best fit sequences as determined by the weighed scoring of DNA Master, which takes into account sequence, distance prior to the initian codon, and distance between putative -35 and -10 sites, are reported. The putative -35 and -10 sites are highlighted in gray.

Figure 3-4: Alignment of Putative Phage Promoter Sequences
DNA sequences upstream of the putative capsid, scaffolding, and ssDNA binding protein were aligned by inspection. Similar sequences to the consensus putative promoter sequence were then found in the latter half of the genome using the "find" command in MS Word, and aligned by inspection against the original three sequences. 16 potential promoter sites were identified in the latter part of the phage genome. The putative phage RNAP promoter sequence does not appear in the early part of the genome, with the initial site appearing before the ss-DNA binding protein. Bases in the putative promoter region are color-coded: guanine is magenta, cytosine is green, adenine is yellow, and thymine is blue.

Phage T7 has one rho-independent terminator (TE), which is located towards the beginning of the phage genome located immediately after the RNAP gene; and two other phage-encoded RNAP terminators, CJ, which is located to the left of the terminal repeat and has a putative role in DNA packaging, and Tphi which is located directly after the capsid gene (Dunn and Studier, 1983). Unlike T7, Syn5 has two rho-independent terminators, located between ORFs 4 and 5 (located at bp1424, 28 nucleotides long) and between ORFs 58 and 59 (located at bp 44852, 37bp long). There is no obvious analog to CJ in Syn5; however, it is possible that a potential Tphi-like terminator is located immediately post the capsid gene, as this area can make a terminator-like structure (Figure 3-5).

iii. *Translation*

58 of the 61 Syn5 ORFs begin with the start codon AUG, while only 3 of the 61 ORFs of Syn5 are predicted to begin with an unusual start codon (GUG or UUG). This is similar to many other enteric phages and the other short-tailed cyanophage, where start codons other than AUG are quite rare (Miller et al., 2003). In contrast, in at least one myocyanophage, S-PM2, 15% of predicted ORFs use an alternate start codon (Mann et al., 2005).

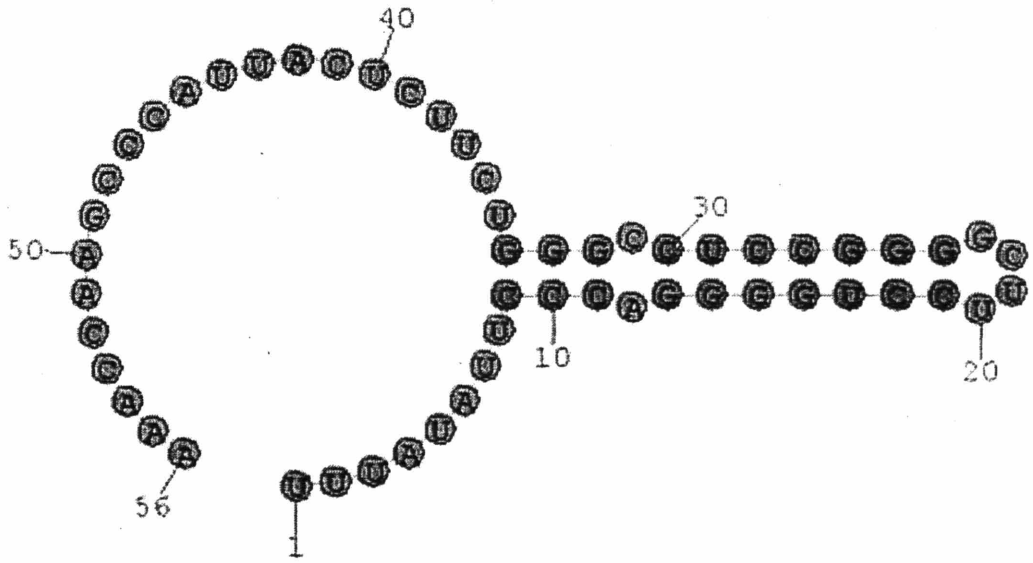


Figure 3-5: Secondary Structure of Putative Tphi-like Terminator
The sequence lying between the end of the capsid ORF and before the beginning of the next ORF was analyzed for secondary structure indicative of a possible termination sequence using the "Find Pal" software. The presence of the loop and trailing poly-U tail indicates that this sequence may code for a T-phi like termination sequence in Syn5. Position "1" is the base immediately after the "TGA" codon of the capsid gene (bp 21016).

Enteric bacteria and phages have highly conserved ribosome binding sites located directly upstream of the first codon translated of each gene called Shine-Dalgarno (S-D) sites. These sequences are conserved both in terms of location (9 to 5 bp before the initial codon) and sequence (GGAGG). The majority of cyanobacteria and chloroplast genes do not exhibit enteric S-D sites, despite the conserved anti-S-D sites found within the cyanobacterial 16S ribosomal RNA sequence (Fargo et al., 1998). To examine the quality of the potential S-D sites within the Syn5 genome, the nucleotide sequence was examined with program DNA Master. DNA Master allows the user to input potential ORFs and then choose a start codon for that ORF based on the quality of putative S-D sites. These sites are given a weighted score based on the number of nucleotides that match the enteric GGAGG, which of nucleotides match, what the replacement nucleotide is, and how far away the site is located from the initial codon within the ORF. Enteric bacteriophage T7 ORF S-D sites generally scored 300 or better. In Syn5, only two ORFs (ORFs50 and 54) exhibited S-D scores of 300+; the rest scored between 280-80; with an average of 168. The phage genes with the most similarity to enteric phage genes had some of the lowest scoring S-D sequences, including portal (180), DNAP (150), and RNAP (128).

The lack of identifiable promoters and strong S-D sequences along with the occurrence of genes with overlapping start and stop codons may be an indication of translational coupling. This phenomenon has been identified in T4 and suggested to occur in S-PM2, in which the translation of the downstream gene is dependent on the translation of the preceding genes. 35 of the Syn5 ORFs could be translationally coupled.

D: DISCUSSION

Through the use of genome analysis programs and similarity to known phage sequences, it was possible to place the marine cyanophage Syn5 within the T7-like family of phage. The annotation presented within this paper represents our best representation of where genes are likely located within the cyanophage genome. However, given the intrinsic differences between enteric and cyanophage with respect to promoter sequences, ribosome binding sites, and gene content; and our lack of evidence of gene expression, it is possible that we have mislabeled potential ORFs. Chapter IV examines particle proteins via mass-spectroscopy and N-terminal sequencing. Through these techniques, it is possible to confirm that the Syn5 particle is comprised of proteins similar to enteric phage structural proteins, and to validate our method of assignment of the initial codon of these proteins as well as the remaining ORFs.

One of the classes of phages that has emerged both through the classical studies of phage morphology and phage/host dynamics, as well as more recent genomic analyses is that of the T7 family. These phages were originally grouped due to the presence of certain broad traits: short tails, a genome of approximately 40kb, terminal repeats at the end of the genome, and a rifampicin resistant RNA polymerase (Hausmass, 1988). Access to multiple genomes of this family has further subdivided the phages of this group based on gene conservation and synteny (Scholl et al., 2004). The 24 "core" genes, as defined by those essential for T7 growth that are prevalent within related phages, include

the RNA polymerase gene at the left end of the genome, followed by genes involved in DNA replication, structural genes, and finally, towards the right end of the genome, genes involved in DNA packaging (Molineux, pers comm., Sullivan et al., 2005). The phages most closely related to T7 span a gamut of host strains, including *Pseudomonas* (Kovalyova and Kropinski, 2003), *Salmonella* (Dobbins et al., 2004), *Yersinia* (Garcia et al., 2003), and *Prochlorococcus* (Sullivan et al., 2003), while phages exhibiting T7 characteristics yet lacking a RNAP include vibriophage VpV262 (Hardies et al., 2003) and roseophage SIO1 (Rohwer et al., 2000).

Even within the closely related T7 phages, certain areas of the T7 genome exhibit more flexibility with regards to gene content and order. *Salmonella* phage SP6 contains an additional 5kb at the right end of the genome, including an ORF that is highly similar to the P22 tailspike without its N-terminal capsid binding domain. SP6 also contains a shorter ORF than T7's gp17 (tailfiber) in the appropriate place within the genome that does contain the capsid binding domain of gp17. It is hypothesized that these two proteins interact non-covalently to create the tailfibers of SP6 (Dobbins et al., 2004). The cyanophage P-SSP7 also contains an additional stretch of genes between the putative tailfiber and terminase genes; however, no functions have been assigned to these ORFs. More strikingly, P-SSP7 contains a putative integrase gene located directly after the RNA polymerase in the genome, suggesting that the T7 family may not be comprised solely of lytic phages (Sullivan et al., 2005). This flexibility within the closely related T7-phages as well as the difficulty in defining core genes indicates that it may not be possible to group phages into genomic classes easily (Lawrence et al., 2002)

Like P-SSP7 and SP6, the Syn5 genome possesses additional ORFs at the tail end of the genome between the tail fiber and the DNA packaging protein, including several that are quite large (ORFs 53 and 54). These ORFs have no obvious analog within known phage sequences, with the exception of a weak similarity of ORF 54 to the phage-encoded RTX toxin. The most similar known sequences appear to encode some fibrous microbial proteins. It is unclear if these ORFs in Syn5 encode putative structural proteins, toxins, or some other unrelated protein; or if they have another function entirely.

i. *ORFs located within the Terminal Repeat*

The Syn5 genome is flanked by a direct repeat of 237 bp. In T7, this repeat is used during concatemeric DNA replication during the middle stages of infection of the host cell. As of yet, there are no identified ORFs located within the T7 terminal repeat (Dunn and Studier, 1983), and it is unlikely that they exist (Molineux, pers comm.). The first ORF of Syn5 is predicted to begin at bp 134, within the 237bp terminal repeat, indicating that it is likely not transcribed except during DNA replication. There are several identifiable genes within Syn5 that are likely primarily used during DNA replication to scavenge host nucleotides, so it is possible that a gene who product expedites DNA replication could be located partially within the terminal repeat. Given that ORF1 has a high scoring S-D sequence and potential promoter sequence, it has remained in the genome despite its lack of similarity to known protein sequences.

ii. *Potential Prophage Attachment Site*

Syn5 is the second cyanophage of the T7-like phages that appears to have an integrase gene located within the early part of the genome. Within the genome of P-SSP7, the integrase gene is located directly before the RNAP gene with a small non-coding gap between the two ORFs. Sullivan et al (2005) identified a small, 42 bp stretch within this gap that is identical to a leucine tRNA encoded by the P-SSP7 host, MED4. This stretch, along with the presence of conserved amino acid motifs R-H-R-Y; LLGH, may indicate functionality of the putative integrase of P-SSP7 (Nunes-Duby et al., 1998; Sullivan et al., 2005).

Although the integrase of Syn5 does possess the conserved amino acid motifs identified by Nunes-Duby et al, 1998, it was not possible to perform the equivalent search between Syn5 and its host, as the genome of WH8109 is not sequenced. However, it was possible to compare the area of the Syn5 genome directly downstream of the integrase ORF to the genome of a related *Synechococcus* strain, WH8102. Using the SuperSearch program, we identified a small stretch of WH8102 that was highly similar to an area between ORFs 18 and 19 of Syn5 (Figure 3-5). In WH8102, this area was located at the end of an open reading frame that was predicted to encode an outer membrane protein, or specifically a "protective antigen". If a similar ORF exists within WH8109 and it is the site of Syn5 integration, it is possible that the Syn5 prophage directly alters an outer membrane protein of the host cell, and may thereby confer resistance to superinfection of

a new phage via alteration of a motif or protein, the recognition of which is crucial for Syn5 infection.

Potential att site... after ORF18

```
|       End 18, before ORF 19       | bp 7535-7571
ACTCAACTGAGCGCTACATTCGGATGCGGATCGCTGAGATTCGCGACGATCTGCTGAATCTACCCACAACCTACCCCGACACCT
|||  |  ..|  |  |  |.||||||.|   ..|  .  |  ..  |    ||..|..||     |   |  |   .||.|.
ACT-ACGAGGGTGATTCATTCGCACTGCGCAAG-ACCGG--CGGCAGCATTGCGTTCA-CGGGCCTCT-C
WH8102 546780-
```

Figure 3-5: Alignment of Putative Attachment Site observed between Syn5 and Synechococcus WH8102. The region of the genome lying between the putative integrase and before the T7-like ssDNA binding protein was compared to the entire genome of *Synechococcus* WH8102 using SuperMatcher (http://bioweb.pasteur.fr/seqanal/interfaces/supermatcher.html). Only one possible alignment was considered significant, and within this alignment, it was possible to identify a 30-bp region with higher similarity than the surrounding sequence. This 30-bp region mapped to a putative outer membrane protein in WH8102.

iii. *Recent Enteric Transfer*

Cyanobacteria do not appear to use S-D sequences as ribosome binding sites for translation. One study suggested that *E. coli* may also be capable of S-D independent translation, as it was possible to express chloroplast proteins within *E. coli* using only the leader sequences of the chloroplast genes. The inclusion of these sequences through cloning before some well-studied chloroplast genes did not increase levels of translation within the chloroplast; however these sequences did increase the protein expression when included before the chloroplast genes in *E. coli*. This suggests that mRNAs of these transcribed genes are recognized and translated by some S-D independent method (Fargo et al., 1998).. Given the lack of these sequences in cyanobacteria and their high level of conservation in enteric bacteria, it is possible that within a cyanophage genome the presence of a S-D site before a gene may indicate a more recent enteric ancestor than the other genes within the phage genome.

Within Syn5, the two ORFs with high scoring S-D sites were ORF50 and ORF54. ORF50 (S-D score of 340) exhibited no similarity to any known enteric bacteriophage sequence; however, ORF50 did exhibit similarity to an ORF with unknown function in cyanophage P60. This would seem to suggest that either ORF50 is not of recent enteric origin; or Syn5 and P60 are recent descendants of the same cyanophage which first acquired the enteric sequence. As Syn5 and P60 infect different *Synechococcus* strains and were isolates of the Sargasso Sea and Georgia Estuary, respectively, the latter

hypothesis seems unlikely. On the other hand, examination of ORF54 appears to support the theory of a strong S-D sequence indicating recent enteric transfer. ORF54 (S-D score of 430) exhibited no similarity to any known cyanophage or cyanobacterial proteins within the NR; resembling instead an enteric fibrous outermembrane protein. Both of these ORFs exhibited promoter sequences that were more similar to sigma70 sequences than to the late phage RNAP promoter sequence, which also could be an indication of recent lateral transfer from an enteric microbe. However, this theory of recent transfer is not supported or disproved by the examination of the G-C content of the ORFs, as the G-C content of both are not statistically different from the whole genome.

With the advent of phage genomics, the importance of role that cyanophage play in the ecology of the marine microbial community became even more evident. Our future studies include the biochemical study of putative phage attachment proteins through which all host/phage interactions must be mediated. These studies will be greatly facilitated by access to cyanophage genome sequences.

CHAPTER IV. IDENTIFICATION OF THE STRUCTURAL PROTEINS OF THE MARINE CYANOPHAGE SYN5

A. INTRODUCTION

Double-stranded DNA phages are a group of viruses which infect bacteria and utilized double-stranded DNA as their carrier of genetic information. These particles are grossly similar in size (nm in diameter) and morphology, consisting of a symmetric icosedral capsid or "head" filled with DNA, and a protruding structure at one vertex of the head which breaks its symmetry, termed a "tail". The tail is responsible for host recognition and facilitation of the entry of the phage DNA to the host cell. The relatively large size and unique shape of dsDNA phages distinguishes the particles from all other prokaryotic structures, including the largest cytoplasmic organelles, ribosomes, when infected cells are viewed via electron microscopy.

dsDNA phage particles are comprised of one copy of the phage genome tightly coiled within a protein shell. This shell is made up of multiple copies of distinct polypeptide chains, all of which are generally transcribed during the later stages of infection. During bacteriophage assembly, twelve copies of the portal protein associate with multiple copies of the capsid, scaffolding, and internal virion proteins to form a structure called the procapsid. The DNA packaging protein(s) then bind to the phage DNA and the dodecameric portal ring, facilitating the entry of the phage genome to the hollow particle head. At this stage, the scaffolding proteins either exit through the pores

in the procapsid head, becoming available for further rounds of procapsid assembly or are proteolytically cleaved to make room for the phage DNA. The capsid undergoes an irreversible expansion, forming the mature icosehedral shell. Neck proteins, which act as plugs to contain the DNA within the shell are added at the portal vertex. Lastly, the tail proteins bind at the same portal vertex, creating a mature infectious particle (Prevelige and King, 1993).

During infection, these tail proteins recognize the host cell and bind to a specific receptor on the surface of the cell. DNA ejection into the cell, through the portal vertex, is also facilitated by tail proteins: through the mechanical actions of the long contractile-tailed phages, or through more subtle means in the long non-contractile tailed phages. Some of the short-tailed phages appear to have extensible tails comprised of proteins located within the capsid which are capable of extending through the portal opening and into the host cell, across the cell membranes and wall directly into the cytoplasm, and which bring the leading edge of the phage genome with them (Kemp et al., 2004; Moak and Molineux, 2000; Struthers-Schlinke et al., 2000).

Classes of phages which utilize alternate nucleic acids as genetic carrier have also been identified. Single-stranded DNA phages, like phiX174, are icosehedral but do not have a unique tailed vertex. Their symmetry is maintained by the placement of tail fibers at every vertex (McKenna et al., 1992). Single stranded RNA phages sometimes do not appear icosehedral at all, but instead appear filamentous, like tobacco mosaic virus. Infection by these phages may not result in lysis of the host, but rather a continual

shedding of filamentous viruses out of the host cell without the death of the cell (Waldor and Mekalanos, 1996). Some phages possess a lipid envelope within the capsid surrounding the nucleic acid (Cockburn et al., 2004; Huiskonen et al., 2004).

Marine cyanobacteria are numerically dominant photoautotrophs in the global oceans (Partensky et al., 1999; Waterbury et al., 1986). Two genera, *Prochlorococcus* and *Synechococcus* have been shown to contribute up to 30% of photosynthesis in oligotrophic areas of the ocean (Goericke, 1993; Li, 1998; Liu et al., 1997; Veldhuis et al., 1997). Viruses capable of infecting *Synechococcus* and *Prochlorococcus*, termed "cyanophage", were first observed in 1990 (Proctor and Fuhrman, 1990; Suttle et al., 1990). Cyanophage isolates were first obtained in 1993 through infection of multiple *Synechococcus* strains (Suttle and Chan, 1993; Waterbury and Valois, 1993) although the ability of some of these phages as well as newer isolates to cross-infect multiple strains of both genera has been demonstrated (Sullivan et al., 2003).

Morphologically, cyanophage appear similar to three types of enteric bacteriophage: Myoviridae, Siphoviridae, and Podoviridae. Of the six cyanophage genomes sequenced to date, the three Myoviridae appear T4-like in organization and gene content, while the three Podoviridae appear T7-like (Chen and Lu, 2002; Mann et al., 2005; Sullivan et al., 2005). The similarity in genome organization, overall gene identity, and similar particle morphology seen within all dsDNA phages observed to date regardless of host is suggestive of an intimate relationship between cyanophage and the

enteric bacteriophage; namely, that all phages may have access to a global pool of genes through lateral gene transfer (Hendrix et al., 1999).

Preliminary investigations of the molecular biology of five cyanophage isolates was undertaken by Wilson et al (1993). These siphophages were concentrated from large volumes of crude lysate via PEG precipitation and centrifugation, and electrophoresed through SDS-gels. Banding patterns indicated that these phages contained structural proteins in the approximate number and proportion as the enteric siphoviridae. Phage S-PM2, a myoviridae, was concentrated in a similar fashion as the phages of Wilson et al (1993) and analyzed by SDS-gels. After the genome was sequenced, particle proteins were further examined via mass-spectroscopy. (Mann et al., 2005). Their results indicated that the SDS-banding patterns of S-PM2 were similar to the enteric T4-like phages.

Cyanophage Syn5 was isolated by Waterbury and Valois (Waterbury and Valois, 1993) from the Sargasso Sea through dilution onto *Synechococcus* strain WH8109. An excess of phage particles added to log phase WH8109 cultures resulted in lysis in approximately 10 hours. To date, only the original strain of isolation, WH8019, has shown sensitivity to Syn5 infection, although at least 15 phages are capable of infecting WH8109 (Sullivan et al., 2003; Waterbury and Valois, 1993). Morphologically, Syn5 appeared similar to the enteric phage T7. As reported in Chapter III, the genome sequence of Syn5 exhibited a marked similarity to the T7-like phages, including the arrangement of the genome as well as gene content. Our analysis of the structural proteins of Syn5 by SDS-PAGE, Western blot, Mass-spectrometry, and N-terminal

sequencing represents a more comprehensive biochemical study of the particle proteins in a cyanophage.

B. MATERIALS AND METHODS

i. *Cell Growth*

Synechococcus strains WH8109 and (herein after referred to as WH8109) was grown as described in Chapter II. Briefly, host cells were inoculated into SN media (Waterbury et al., 1986) at 26°C under continuous light in a culture vessel designed to optimize cyanobacterial growth. WH8109 was grown under cool white fluorescent light (40W bulbs) at an irradiance of $50\mu Em^{-2}s^{-1}$.

ii. *Phage Purification and Concentration*

Syn5 was purified and concentrated as described in Chapter II. Briefly, WH8109 was grown to a density of $1x10^8$ cells/ml at 26°C in SN media at an irradiance of $50\mu E$ $m^{-2}s^{-1}$ in an optimized culture vessel. Syn 5 was added at a moi of 10. The culture was incubated until lysis (noted by a change in turbidity and color) at about 10 hours after infection. The crude lysate was centrifuged at 8,000 g for 15 minutes at 4°C to pellet cellular debris. The supernatant was then filtered through a Whatmann glass fiber filter and a 0.4μ polycarbonate filter to further remove excess cellular debris, specifically cell membranes, which may bind free phage and prevent their recovery. After the incubation, the lysate was moved to a 4°C room, and NaCl was added to a final concentration of

0.5M and stirred to dissolution. Polyethelene glycol (PEG) 8000 was added to a final concentration of 10%w/v, mixed to dissolution, and incubated with stirring for two hours at 4°C. Phage were then centrifuged at 9,000g for 30 minutes. The supernatant was poured off, and the pellet was resuspended by gentle stirring at 4°C in an appropriate amount of SN (10-15ml per liter of crude lysate) supplemented with 50mM Tris-HCl, pH 8.0 and 100mM $MgCl_2$. The suspension was layered onto a step-gradient made of 20% sucrose (w/v), ρ=1.4 CsCl, and ρ=1.6 CsCl. Each gradient step was made with SN. The gradients were ultracentrifuged at 40,000 revolutions/min for 1.5 hours in a SW50.1 rotor or 28,000 revolutions/min for 4 hours in a SW28 rotor. An opalescent phage band was visible at the interface between the ρ=1.4 CsCl step and ρ=1.6 CsCl step. The phage band was collected and dialyzed for 30 minutes against 1M NaCl, 100mM $MgCl_2$, 50mM Tris-HCl, pH 8.0, and then against 100mM NaCl, 100mM $MgCl_2$, 50 mM Tris-HCl pH 8.0 in a Slide-A-Lyzer dialysis cassette (Pierce).

iii. *Phage Protein Analysis*

CsCl-purified Syn5 particles were mixed with SDS-buffer, boiled for 3 minutes, and loaded onto a 10% polyacrylamide gel containing SDS. The samples were electrophoresed at 20mA constant current until the dye line ran off the end of the gel. The gels were stained with Coomassie blue. 12 protein bands were visible and their approximate molecular weights were determined by comparison to a protein standard (Broad Range Protein Marker, NEB).

iv. *Mass-spectroscopy*

Purified Syn5 particles were electrophoresed through SDS-polyacrylamide gels, and the gels were stained with Coomassie blue. The bands were then excised with a razor blade and individually packaged in microcentrifuge tubes. A control slice of gel was included as a negative control and to provide background correction. The gel bands were digested with trypsin as according to the MIT Biopolymers trypsin digest protocol (proprietary). 2μL of the extracted sample was diluted to 50 μL with 0.1% acetic acid. 15 μL of this dilution was run through a reverse-phase capillary column using high-pressure liquid chromatography (HPLC) with an Agilent Model 1100 Nanoflow HPLC and into a linear trap quadropole (LTQ) mass-spectrometer (Thermo Electron Corporation, FL), where the peptides were ionized with 35kV. Detected peptides, as represented by well-resolved peaks with a high relative abundance at a specific mass/charge, were fragmented. Spectra of the resulting B- and Y- ions produced by peptide fragmentation were recorded every 0.2ms. Resulting spectra were compared to a database produced from the six-frame translation of the entire nucleotide sequence of the Syn5 genome and assigned a putative peptide sequence. Quality spectra indicative of true peptide matches were determined by visual inspection, and the scores (XC, Delta Cn, Sp, RSp) computed by the Sequest Database Search Software (MacCoss et al., 2002).

v. *N-terminal Sequencing*

Phage particles were electrophoresed through an SDS-polyacrylamide gel. The proteins bands were transferred to PVDF membrane (Millipore) via a plate-electrode transfer apparatus (BioRad) using a 10% methanol transfer buffer at a constant voltage of 100v for 1 hour at 4°C. The membrane was stained with Coomassie Blue to visualize the transferred protein bands. Visible bands were subjected to multiple cycles of Edman degradation. Amino acids recovered from each band after each cycle were identified using HPLC with a reverse phase column in conjunction with amino acid standards.

vi. *Antibody Production*

Purified Syn5 particles were prepared as in Chapter II. Immune sera was produced at Covance Research Products, Inc. by the following protocol: 150ul of a 10^{10} phage/ml solution was brought up to 1ml volume with phosphate buffered saline (PBS), split into 0.5ml volumes, and injected into two New Zealand White Rabbits. Two boosts of antigen at the same concentration were performed at two week intervals. After six weeks, serum was collected from each rabbit, and another antigen boost was injected. After eight weeks, more sera was collected and a final antigen boost was performed. Final bleeds were performed ten weeks and twelve weeks after initial inoculation.

vii. *Western Blotting*

Purified Syn5 particles were electrophoresed at 20mA through a 10% polyacrylamide gel containing SDS. The gel was immersed in transfer buffer (20%

methanol, 25mM Tris, 192mM glycine) for 15 min, and then electroblotted to PVDF membrane (Millipore) in transfer buffer using a Criterion electroplate transfer apparatus (BioRad) at 100V for 1hr at 4°C. The membrane was then stored in wash buffer (PBS + 0.1%Tween20, pH 7.5) overnight. The membrane was probed using an ECF anti-Rabbit antibody kit (Amersham) according to manufacturer's instructions. Briefly, the membrane was placed in blocking solution (Blocking Agent (Amersham)(5%w/v) in wash buffer) for 1 hour at 4°C, then washed 3X for 15min, then 2X for 5min in wash buffer at 4°C. The membrane was then probed with polyclonal Syn5 rabbit sera diluted 1:10 in wash buffer for 1 hr at 4°C, and washed as above. Finally, the membrane was probed with 2° anti-rabbit antibodies (Amersham) diluted 1:10000 in wash buffer for 1hr at 4°C and washed as above. The membrane was then incubated at room temperature for 5 min with ECF substrate (Amersham) and scanned using a FluroImager 595 (Molecular Dynamics).

C. RESULTS

i. *Particle Protein analysis using SDS-Gel Electrophoresis*

In Chapter II, we examined Syn5 particles with SDS-gel electrophoresis, and determined that the particles appeared to be composed of at least 12 discreet polypeptide chains with different relative mobilities. However, with the exception of the capsid protein, we were unable to assign any of the other detectable bands to specific putative structural proteins. With the sequence of the Syn5 genome, it became possible to

determine the putative molecular weights of phage structural proteins based on predicted amino acid sequence. Purified particles were electrophoresed through an SDS-polyacrylamide gel and the relative mobility of each band detectable by Coomassie Blue staining was calculated. Preliminary assignment of predicted specific phage structural proteins to bands observed by SDS-gel electrophoresis was then made by comparison of these mobilities to the predicted molecular weights of the putative structural proteins (Figure 4-1).

In Figure 4-1A, five of the twelve bands produced by Syn5 particles during SDS-gel electrophoresis were assignable to putative structural proteins by relative gel mobility. The capsid band (ORF 39), with a predicted weight of 35 kDa was assigned to the most intense band on the gel—calculated by gel mobility to weigh 31kDa. The most retarded band, calculated to weigh 157 kDa, was assigned to the longest ORF in the genome, an internal virion protein (encoded by ORF 45), with a predicted weight of 152 kDa. The next most retarded band was assigned to the tail fiber (encoded by ORF 46), the next to Tail Tube B (encoded by ORF 41), and the next to an internal virion protein (encoded by ORF 44).

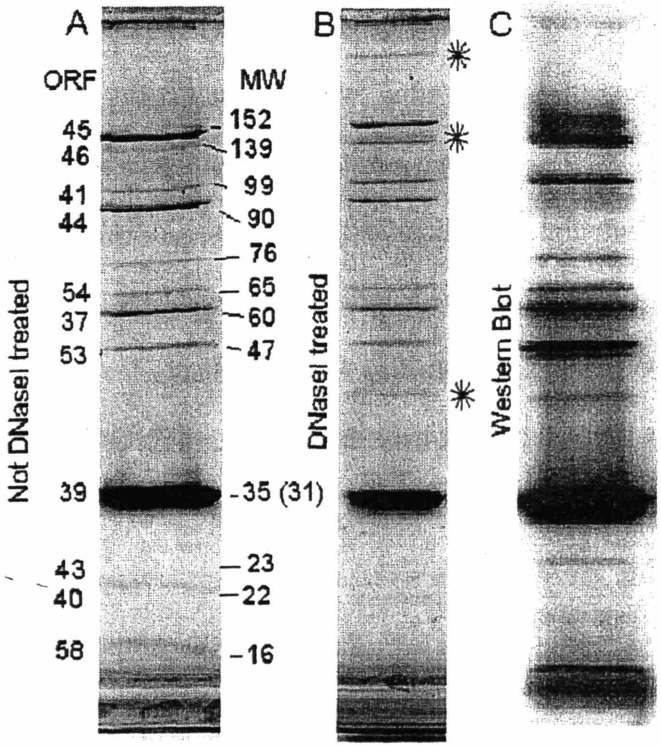


Figure 4-1: SDS-polyacrylamide gel and Western Blot of purified Syn5. Purified Syn5 particles were denatured by boiling in SDS-sample buffer for three minutes and then electrophoresed through a 10% polyacrylamide gel. The proteins were stained with Coomassie Blue. 12 distinct bands are visible. Approximate molecular weights are indicated as calculated from movement through the gel. Syn5 samples treated with DNaseI during purification exhibit marked proteolytic activity in comparison to samples that were untreated. A) SDS gel of Coomassie blue stained particles purified without DNaseI. B) SDS gel of Coomassie blue stained purified in the presence of DNaseI C) Western Blot of particles purified without DNaseI. Syn5 particles were electrophoresed through a 10% polyacrylamide gel containing SDS. The resulting bands were electroblotted to PVDF membrane, and the membrane was probed with polyclonal antiSyn5 rabbit sera. The bands were then visualized using the ECF anti-rabbit kit (Amersham).

The remaining bands were more difficult to assign, as the calculated weights of the bands either did not match any predicted ORFs, or matched ORFs that were not predicted structural proteins. For example, the portal protein was predicted to have a mobility of 60kDa, and although there was a ~60kDa band, there was also a ~65kDa band without a likely candidate among the remaining predicted structural proteins. Given the error inherent within the gel mobility calculations and potential for post-translational modifications, it was not possible to assign the portal protein to one band over the other by this method. Confirmation of these assignments as well as assignment of the remaining unknown bands was possible using mass-spectroscopy and N-terminal sequencing (see below).

ii. *Mass-Spectroscopy*

Although preliminary assignments of putative ORFs in the Syn5 genome could be made to some detected protein bands from purified particles based on molecular weight, final assignments were made using mass-spectroscopy and N-terminal sequencing of the protein bands. Mass-spectroscopic analysis was performed with a linear trap quadrupole mass-spectrometer (LTQ), which yielded internal amino acid sequence from each of the detected peptides as well as total molecular weight of each peptide. LTQ analysis in conjunction with HPLC and a reverse-phase column is limited to peptides between ~7-30 amino acids; thus sequence coverage of most proteins is not 100% as any peptide produced by tryptic digestion that falls out of the optimum range cannot be detected.

Purified phage particles were run on an SDS polyacrylamide gel and stained with Coomassie Blue. The protein bands visible in the sample lane, as well as an unstained portion of the gel to be used as a control, were excised and digested with trypsin as according to the MIT Biopolymers trypsin digest protocol (proprietary). The resulting peptides in each sample were then concentrated, run through a reverse-phase column using HPLC, and into a linear trap quadrupole mass-spectrometer (LTQ). Peptides were ionized with 35kV and ionized particles were detected by the LTQ. Putative peptides were detected by the presence of tall, well-resolved peaks on a graph of relative abundance vs specific mass/charge comprised of all material entering the LTQ. Detected putative peptides were fragmented. The LTQ then recorded the spectra of relative abundance vs. mass/charge (m/z) of the ionized peptide fragments every 0.2ms. Spectra of fragmented peptide ions which exhibited tall, well-resolved peaks were analyzed by Sequest Database Search Software to match the fragmented peptides to putative Syn5 peptides. These assignments were generated with several quality control variables (discussed below).

Mass-spectroscopic analysis was performed on every protein band labeled in Figure 4-1A as well as the bands indicated with asterisks in Figure 4-1B. Based on stringent conditions (as discussed below), it was possible to match eleven of the twelve visible protein bands in Figure 4-1A to a corresponding ORF based on the number and sequence of the recovered peptides (Table 4-1). As all visible bands contained some peptides from multiple structural proteins, an ORF assignment was made to a band when: a) the majority of the recovered peptides were encoded by one specific ORF, b) the full

length protein encoded by that specific ORF had a putative molecular weight that corresponded to the approximate mobility of the assigned band on the SDS gel, and c) no other band on the gel matched the protein encoded by that ORF more strongly. Multiple peptides detected within a single protein band excised from a gel lane loaded with a sample made up of multiple proteins is not unusual during LTQ analysis, as the instrument is highly sensitive and can detect peptides that are present within picogram amounts. The "true" peptides present in each visible band ("true" indicating that the peptide matched all stringent criteria and had a plausible spectrum) which did not match the protein encoded by the assigned ORF, were most likely the result of peptides from previous samples co-eluting from the column with the new sample or, as the LTQ ion trap is extremely sensitive, polypeptide chains which were present in very low quantities as leading or trailing edges of their distribution, in an area of the gel that did not correspond to the relative mobility of those chains.

iii. *Assessment of Mass-spectroscopic Data*

Analysis of the LTQ data in assignment of Syn5 ORFs to protein bands took into account several factors. First of all, the reverse-phase column ensured that the peptides entered the trap in order of least hydrophobic to most hydrophobic, and that only peptides between 7-30aa were accurately detectable by this method. Longer peptides become too difficult to elute from the column, while shorter peptides fly through the trap too rapidly for accurate analysis. This means that any calculated peptide coverage of the identified protein should be adjusted to account for unrecoverable peptides. Secondly, The LTQ ion

trap primarily detects B and Y ions from fragmented peptides (peptides produced by the breaking of the C-N bond between amino acids); A, C, X, and Z ions (peptides produced by breaking of the C-C or N-C bonds, where the C has the side chain attached) do not have a strong enough signal. In Figure 4-2, peaks that correspond to putative B ions are colored red, putative Y ions are colored blue, and unidentified peaks are colored black. In a quality spectrum, any unresolved peaks detected when the peptide is assigned a +1 charge (A) will become resolved if the peptide is assigned a +2 or +3 charge (B). A peptide with a charge of +1, nine amino acids long, such as GLYSAIGIR, produces 16 possible peptide ions when fragmented once: eight B ions and eight Y ions. Fragmentation happens more often within the middle of the peptide than towards the edges, so the middle peaks are higher, indicating a greater abundance. Inspection of a given peptide's m/z spectrum was the key determinant of detection of a "true" peptide. Quality spectra contained well-resolved peaks present at a high abundance over the background noise, while rejected spectra had poorly resolved peaks with a low abundance over background noise.

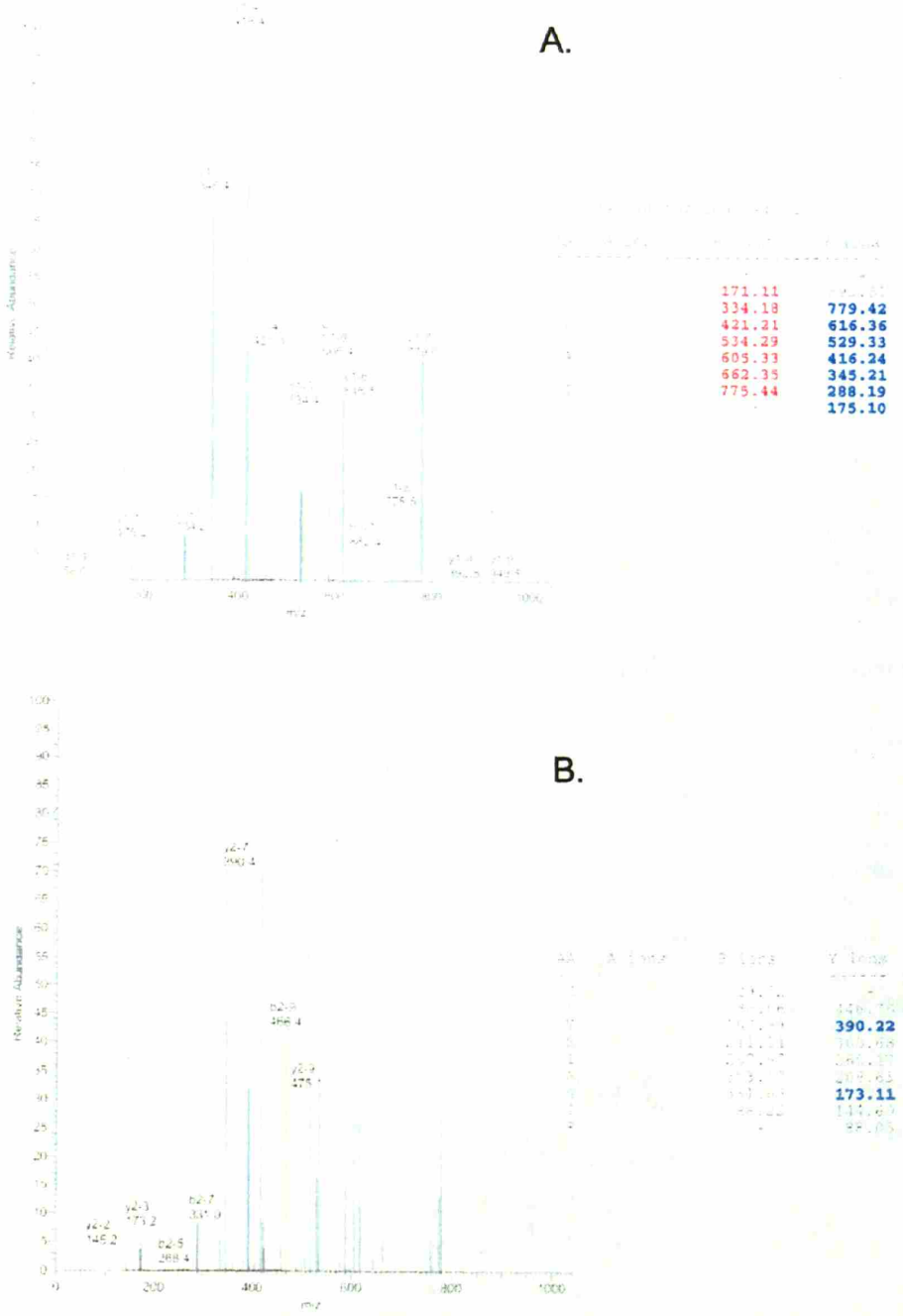


Figure 4-3: Spectra Produced by Fragmentation of the Peptide GLYSIAGIR
A peptide with an initial m/z of 745 was fragmented and the resulting spectrum analyzed.
A) Peaks (m/z) which can be assigned to the B (blue) and Y (red) ions with a +1 produced by the fragmentation of the peptide GLYSIAGIR. Black indicates unaccounted for peaks.
B) Peaks (m/z) which can be assigned to B (blue) and Y(red) ions with a +2 charge produced by fragmentation of the peptide GLYSIAGIR. Together, both graphs show that every peak (m/z) in the fragmentation spectrum can be accounted for by a match with the predicted peaks (m/z) in a fragmentation of the peptide GLYSIAGIR.

The spectrum produced by each fragmented peptide was analyzed with Sequest Database Search Software. This compares spectra to the entire translated Syn5 genome, and assigns a putative sequence based on a preliminary analysis. Sample assignments and parameters shown in Table 4-1. For all assignments see Appendix B. Preliminary analysis statistics in the column headings of Table 4-1 are:

- the Syn5 ORF the assigned peptide was encoded by (ORF)
- scan numbers by the machine that matched to the given peptide spectrum (scans)
- putative amino acid sequence (sequence)
- molecular weight (MH+)
- charge (Z)
- several factors which score the match of the peptide to the database (Sp, RSp, Δ cn, XC), and number of ions detected out of total putative ions per peptide.

RSp scored the match of the m/z spectrum of the peptide to the first choice potential protein in the database, without taking into account the relative abundance of the peaks. A "1" indicated that the peptide matches the first putative protein, a "2" the second putative protein, etc. Sp scored the match of the m/z spectrum of the peptide taking into account the relative abundance of the peaks. Sp's >750 indicated a quality match. The cross-correlation coefficient (XC) was a measurement of the potential that the m/z peaks produced by fragmentation of a peptide with given molecular weight could be produced by a different peptide with the same amino acids in a different order. A higher XC indicated a higher quality match. Peptides with a charge of 3 must have a XC score

>3.5 for highest stringency matches (good is >2.9, acceptable is >2.5); peptides with a charge of 2 must have a XC score >2.7 (good is >2.2, acceptable is >2.0); and peptides with a charge of 1 must have a XC score >1.9 (good is >1.5, acceptable is >1.2). The Delta cross-correlation (Δ Cn) was calculated by (XC of first match – XC of second match) / XC of first match. The Δ Cn must be above 0.1 for a quality match. The final column of Table 4-1 lists the number of detected fragmented B or Y peptide ions of total potential B and Y peptide ions for each of the matched peptides. A high ratio indicates better coverage of a specific peptide. Investigations of spectra which did not meet the criteria confirmed that these spectra contained poorly resolved peaks with low abundances above background.

In the generation of Table 4-1 and Appendix B, factors Δ Cn, Sp, and RSp were set at their respective limits, with XC set at the "acceptable" level as outlined above, and spectra/peptide assignments that made this preliminary cut-off were examined. However, these factors for individual spectra in the Syn5 were artificially improved, however, as the calculated value of each factor is dependent on the size of the database that the spectra were compared to, and the Syn5 database is substantially smaller than the NCBI NR database. As the Syn5 database was generated by a six-frame translation of the Syn5 genome, only peptides which could be encoded by the Syn5 genome appear in the resulting sequence to spectra assignments. To offset this artificial increase, we only considered peptides which met the "good" criteria and above in our data analysis. Next, the individual spectra of each peptide was examined by eye to determine a) if all the visible peaks were explainable by the m/z of putative B and Y ions and b) there was

minimal background detected in each spectra. All of the peptides which matched these criteria were considered "true" peptides, as opposed to background noise or a database mismatch. The preliminary cut-off factors appeared to accurately represent which spectra truly represented the putative fragmentation of Syn5 proteins.

Table 4-1: Syn5 Peptides Identified by LTQ Ion Trap Spectra
Protein bands produced by electrophoresis of Syn5 particles through an SDS polyacrylamide gel were eluted from the gel, digested with trypsin, and subjected to tandem MS using an LTQ Ion trap. Resulting spectra were analyzed with Sequest Database Search Software. Peptides which meet the criteria for the highest stringency are highlighted; light gray indicates peptides which support the eventual ORF assignment of the originating protein band, dark gray indicates peptides which do not support the eventual ORF assignment. (See Appendix B for peptides produced from all protein bands of Syn5 in order from highest molecular weight to lowest).

* indicates an oxidized methionine
# indicates an oxidixed cysteine
@ indicates a phosphorylated serine, tyrosine, or threonine

Table 4-1
Band #2 MW 139kDa: ORF46

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						284.68			29 (27 1 1 0 0)
46	5200 – 5215	R.YSVVVSPSEDSAPTSTLQAQVSTK.T	2481.24	2	6.66	0.83	1609.5	1	28/46
46	6066 – 6087	K.TDGQAFSADFEFVLAATNALPLK.G	2426.22	2	6.07	0.72	3002.6	1	29/44
46	5430 – 5440	K.TTTQFVVETVAVPGGGAANQFADR.G	2436.22	3	5.92	0.65	1721.2	1	44/92
46	5585 – 5603	K.VTTGTFDVVFTTAMPDAK.Y	1900.94	2	5.72	0.78	1790.0	1	23/34
46	6283 – 6296	K.TLVM*DDLLVSSQFVYSLDEIFSQYSTR.A	3172.54	3	5.54	0.84	1143.2	1	35/104
46	6330 – 6352	K.TLVMDDLLVSSQFVYSLDEIFSQYSTR.A	3156.55	3	5.33	0.63	2073.9	1	41/104
46	5370 – 5415	R.FNSAGTLEAGLNVASVTK.V	1778.93	2	5.31	0.80	2075.2	1	25/34
46	5461 – 5475	K.VTTGTFDVVFTTAM*PDAK.Y	1916.93	2	5.12	0.83	1462.1	1	21/34
46	5209 – 5240	R.YSVVVSPSEDSAPTSTLQAQVSTK.T	2481.24	3	4.88	0.71	2732.8	1	42/92
45	5713 – 5729	R.TGTGQYTVVLATAMPNANYSVSSGGEAALLLVSNK.T	3484.76	3	4.66	0.78	971.4	1	41/136
46	5434 – 5448	K.TTTQFVVETVAVPGGGAANQFADR.G	2436.22	2	4.13	0.71	640.6	1	28/46
46	5493 – 5552	K.ATLDGVATTEYTLANATTLR.F	2082.07	2	3.79	0.54	1366.3	1	22/38
39	3421 – 3434	R.EIGNSQGDMNSGK.G	1336.58	2	3.49	0.77	646.3	1	17/24
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
46	5069 – 5073	K.TASSFEVYVNK.T	1244.62	2	3.22	0.60	910.1	1	14/20

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39	4959 – 4967	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	3.11	0.60	1528.9	1	33/72
45	5258 – 5285	K.SGGVLNDELLK.G	1144.62	2	3.01	0.69	801.9	1	16/20
44	5157 – 5190	R.GDFISQDELK.S	1151.56	2	2.97	0.68	614.1	1	14/18
45	3467 – 3527	R.ILESGDTAR.F	961.50	2	2.61	0.50	649.9	1	14/16
46	5186 – 5192	K.ATFFPGSAVR.A	1052.55	2	2.56	0.69	872.9	1	17/18
45	3706 – 3710	K.SAEVDESALAK.E	1119.55	2	2.45	0.54	901.1	1	17/20
44	3548 – 3559	K.TLGDQYNR.Q	966.46	2	2.23	0.64	300.2	1	11/14
	5091 – 5103	R.PSISASAPSSTSR.E	1247.62	2	2.19	0.48	319.9	1	11/24
	5430 – 5440	R.C#EELGGCKSVPLGGCVACPYHTR.N	2436.09	3	2.10	0.66	468.5	3	27/88
46	5399 – 5416	K.TVSGFTVNIYHLTAGASR.D	1893.98	3	2.04	0.67	466.0	1	30/68
	5430 – 5440	R.CEELGGC#KSVPLGGCVACPYHTR.N	2436.09	3	2.01	0.69	474.9	2	27/88

The mass-spectroscopic analysis confirmed the preliminary assignments of the larger structural proteins and the capsid protein to specific bands in the gel. An internal positive control lay in the fact that the capsid protein was assigned by mass-spectroscopic analysis to the most intense band visible on the SDS-gel. It also became possible to assign the portal protein (encoded by ORF 37) to the 60kDa band, and the proteins encoded by ORFs 54, 53, 40, 43, and 58 to the remaining bands visible in Figure 2A, with one exception. The 65 kDa band in Figure 2A and the three bands in Figure 2B that do not appear in 2B (a highly retarded species, a doublet above the tail fiber, and an ~38kDa band, all indicated by asterisks) appeared to be mixtures of the detected structural proteins, either in some SDS-resistant aggregate or proteolytically cleaved peptides, and were therefore not assignable to a specific ORF within the Syn5 genome.

iv. *N-terminal Sequencing*

N-terminal sequencing of the visible protein bands produced by electrophoresis of Syn5 particles through an SDS-gel was performed to confirm that the N-terminal sequence of a particular band matched the predicted N-terminal sequence encoded by the assigned ORF. By determining these sequences, we could validate our method for the choice of a particular start codon for ORFs which may have had several putative start sites, in addition the N-terminal analysis assessed whether each visible protein band was the result of one ORF. Purified phage particles were electrophoresed through an SDS-polyacrylamide gel, then electroblotted to a membrane. The membrane was washed thoroughly to remove traces of glycine from the transfer buffer and stained with

Coomassie blue. Only eleven of the twelve particle proteins were visible after transfer, and these were subjected to multiple rounds of Edman degradation. The resulting amino acids were analyzed using HPLC and a reverse phase column to determine the N-terminal amino acid sequence (Table 4-2). Examples of graphs of amino acid elution are shown in Figure 4-3A and B. Amino acid assignment was based on quantity of each amino acid detected after each round compared to an amino acid standard; the highest quantity above background resulted in residue assignment.

Most of these sequences confirmed the ORF assignments made through mass-spectroscopy (Table 4-3). However, the protein encoded by ORF 41 was unassignable due to high background. Also noted in Table 4-3 is whether or not the N-terminal methionine of each protein was recovered during degradation. ORF/protein assignments of 9 of the 11 Syn5 bands were confirmed by N-terminal sequencing. One band was not visible after transfer (encoded by ORF 43), and one band (encoded by ORF 41) had high background signals making amino acid detection impossible.

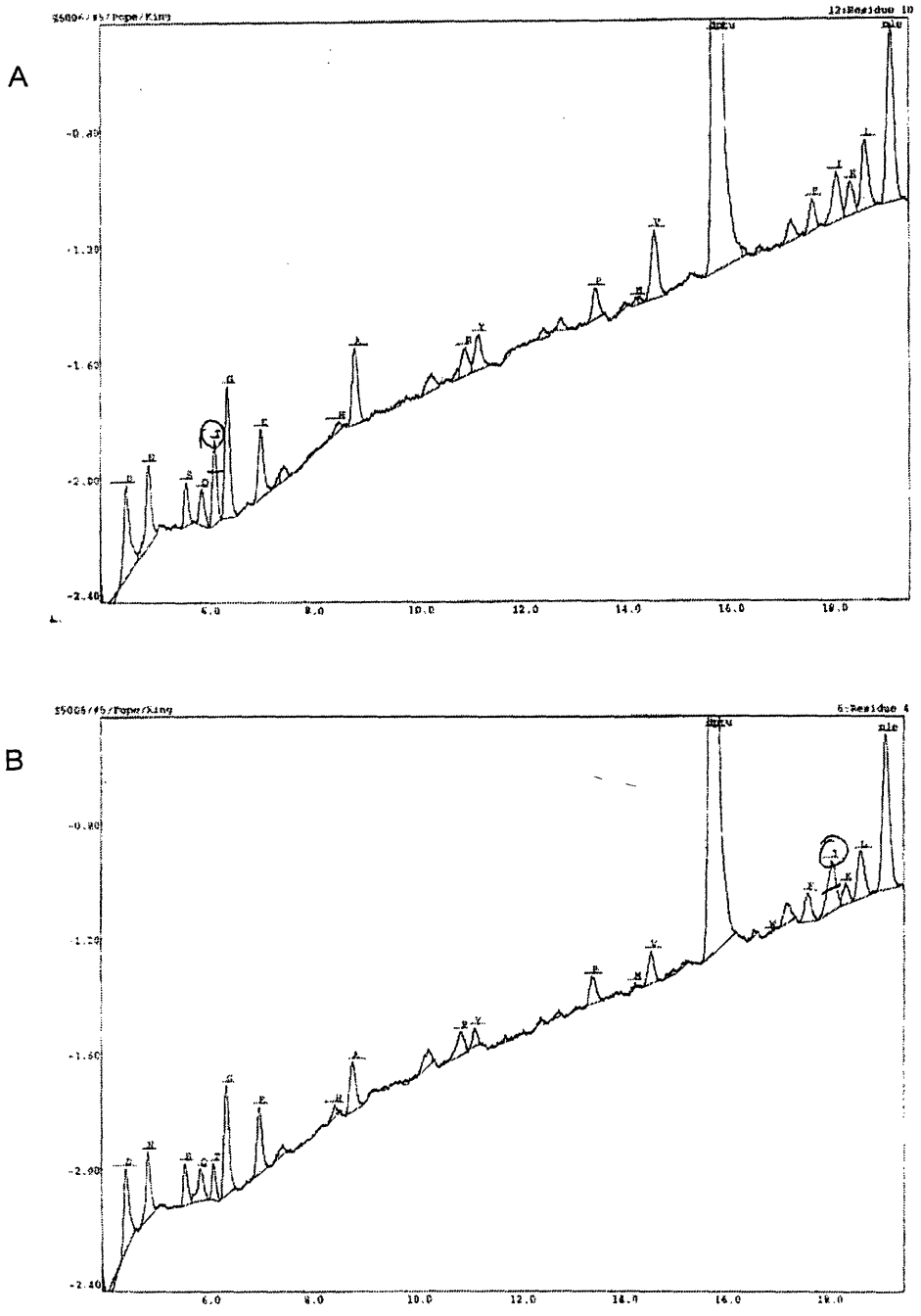


Figure 4-3: Amino acids detected during two different cycles of Edman Degradation
Protein bands produced by electrophoresis of Syn5 particles through an SDS-gel were electroblotted to PVDF membrane and stained with Coomassie Blue. Each band was subjected to multiple rounds of Edman Degradation and the resulting amino acids collected and analyzed using a reverse-phase column. A) The spectrum includes peaks of every amino acid in the standard, with the peak for threonine increased above normal. B) The peak for isoleucine is increased above background, and the threonine peak is normal.

Table 4-2 (A-J). Amino acids produced during Edman Degradation of Syn5 protein bands

Syn5 proteins were electroblotted to PVDF membrane and stained with Coomassie Blue. Ten of visible bands were subjected to rounds of Edman Degradation for either 25 rounds or until the signal was undetectable. The assigned ORF is listed above the table produced for each band, and detected amino acids are highlighted.

A) Band #1 MW 152

ORF45 – MDQLNLFQDGATEGA…

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	(MET)	0.4	LYS(0.1) TYR(0.1) ALA(0.2)
2	ASP	0.4	
3	GLN	0.2	
4	(ILE)	0.2	
5	ASN	0.3	
6	LEU	0.5	
7	PHE	0.3	
8	[GLN]	0.2	
9	-	-	
10	-	-	
11	[ALA]	0.2	

B) Band #2 MW 139

ORF46 – MATIENLYTGDGGTVLFS…

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	-	-	Met(0.1) Ser(0.7) Ala(1.0) Glu(0.2) Gly(2.1) Tyr(0.3)
2	THR	0.7	
3	ILE	0.4	
4	GLU	0.6	
5	ASN	0.3	
6	LEU	0.5	
7	TYR	0.3	
8	(THR)	0.4	
9	GLY	0.4	
10	(ASP)	0.2	
11	[VAL]	0.2	
12	-	-	
13	[THR]	0.2	
14	-	-	
15	[LEU]	0.2	
16	[PHE]	0.1	

17	-	-	
18	-	-	
19	-	-	
20	-	-	

C) Band #3 MW 99

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	-	-	High background
2	-	-	
3	-	-	
4	-	-	
5	-	-	

D) Band #4 MW 90

ORF 44- MARIYESNFSEGN...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	-	-	GLY(0.8) GLU(0.3) ALA(0.6) SSER(0.3)
2	GLN	0.2	
3	ILE	0.4	
4	TYR	0.2	
5	(GLU)	0.3	
6	[LEU]	0.1	
7	-	-	
8	-	-	
9	-	-	
10	-	-	

E) Band #5 MW 65

ORF 54- MSFNISNNVITQSGTDTDLSGLAGVT...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	(SER)	1.8	GLY(0.7) GLU(0.4)
2	PHE	1.0	
3	ASN	0.7	
4	ILE	0.5	
5	(SER)	0.2	
6	(ASN)	0.4	LYS(0.1)
7	ASN	0.4	
8	VAL	0.4	
9	ILE	0.4	
10	THR	0.4	
11	(GLN)	0.1	
12	-	-	
13	(GLY)	0.3	ALA(0.1
14	(ASN)	0.1	
15	(ASP)	0.1	

F) Band # 6 MW 60

ORF37-MKGLAQARYSAMRADREDFL...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	MET	1.0	
2	LYS	0.7	
3	GLY	0.7	
4	LEU	1.1	
5	ALA	1.0	
6	GLN	0.6	
7	ALA	0.6	
8	ARG	0.6	
9	TYR	0.7	
10	SER	0.3	
11	ALA	0.5	
12	MET	0.4	
13	[ARG]	0.1	
14	ALA	0.4	
15	ASP	0.3	
16	-	+	
17	GLU	0.2	
18	-	-	
19	-	-	
20	-	-	

G) Band #7 MW 48

ORF53-MALIVDPDDLTKDTEVVFDTAAK...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	(GLY)/(ALA)	2.4/2.3	GLU(0.4)
2	LEU	1.7	
3	ILE	2.0	
4	VAL	2.0	
5	ASP	1.3	
6	PRO	1.0	
7	ASP	0.4	
8	ASP	0.5	
9	LEU	0.8	
10	THR	0.6	
11	LYS	0.3	
12	-	-	
13	THR	0.5	
14	GLU	0.3	
15	VAL	0.4	
16	VAL	0.4	
17	PHE	0.2	
18	(ASP)	0.2	
19	-	-	
20	(ILE)	0.2	
21	(ALA)	0.3	
22	(GLY)	0.2	
23	-	-	
24	-	-	

H) Band #8 MW 35

ORF39-MTTLSNFSLPNQANGGARNADYDVRYATAL...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	THR	62.7	
2	THR	59.6	
3	LEU	59.2	
4	SER	28.4	
5	ASN	34.5	
6	PHE	43.4	
7	SER	19.3	
8	LEU	32.3	
9	PRO	23.1	
10	ASN	18.2	
11	GLN	14.9	
12	ALA	23.0	
13	ASN	11.2	
14	GLY	11.2	
15	GLY	10.3	
16	ALA	15.9	
17	ARG	8.0	
18	ASN	9.2	
19	ALA	10.0	
20	ASP	9.0	
21	TYR	9.5	
22	ASP	4.1	
23	VAL	7.4	
24	ARG	4.8	
25	TYR	5.3	
26	ALA	6.9	

I) Band # 9 MW 23

ORF40-MASKLTKLGAVNIVLTNIGMAPVTL...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	ALA	2.4	
2	(SER)	0.2	
3	LYS	1.1	
4	LEU	1.3	
5	THR	1.8	
6	LYS	0.8	
7	LEU	1.1	
8	GLY	1.0	
9	ALA	0.8	
10	VAL	0.7	
11	ASN	0.6	
12	ILE	0.6	
13	VAL	0.5	
14	LEU	0.8	
15	THR	0.6	
16	ASN	0.3	
17	ILE	0.4	
18	GLY	0.3	
19	MET	0.2	
20	(ALA)	0.2	
21	[SER]/[PRO]	0.2/0.1	
22	[PRO]	0.1	
23	[THR]/[VAL]	0.2/0.1	
24	[LEU]	0.2	

J) Band #10 MW 16

ORF58-MSATTSNASASSQPRKLYSLLSR...

Cycle	Assignment	pmoles (PTH AA)	Other residues present(pmol)/notes
1	SER	13.4	TYR(1.1) THR(1.4)
2	ALA	27.7	
3	THR	20.9	GLU(2.1)
4	THR	23.7	SER(1.2)
5	SER	7.0	PHE(1.8)
6	ASN	8.3	
7	ALA	11.5	
8	SER	4.0	
9	ALA	4.8	
10	(SER)	1.6	
11	(SER)	2.7	
12	GLN	2.7	
13	PRO	1.7	
14	ARG	1.3	
15	LYS	1.8	
16	LEU	2.0	
17	TYR	1.8	
18	(TYR)	1.3	
19	-	-	
20	(LEU)	1.4	
21	-	-	
22	-	-	
23	-	-	
24	-	-	
25	-	-	

ORF	MW	Protein	Chains/particle	Mass-Spec	N-term
45	152	Intern. Vir. Pro. (16)	7±2 (5)	95%	Y/Y
46	139	Tail Fiber (17)	3±1 (20)	64%	Y/N
41	99	Tail Tube B (12)	6±2 (6)	86%	N/-
44	91	Intern. Vir. Pro. (15)	8±2 (9)	96%	Y/N
54	65	??	4±2	75%	Y/N
37	60	Portal (9)	12 (12)	88%	Y/Y
53	48	??	6±2	44%	Y/N
39	35	Capsid (10)	400±50 (415)	100%	Y/N
43	23	Intern. Vir. Pro. (14)	8±3 (10)	66%	N/-
40	22	Tail Tube A (11)	6±3 (6)	50%	Y/N
58	16	??	12±3	83%	Y/N

Table 4-3. Assignment of Syn5 proteins by Mass-spectroscopy and N-terminal sequencing. 11 Syn5 proteins were assigned to putative ORFs within the Syn5 genome via analysis by mass-spectroscopy and/or N-terminal sequencing. Each protein band is labeled with its ORF number from the Syn5 genome, T7 identifier (if any), and molecular weight. Also included is the approximate number of peptide chains per particle (as determined by gel quantification, error bars reflect standard error) and the number of chains of the corresponding protein in T7 (if any), confirmation of protein/ORF assignment by N-terminal sequence data, and presence/absence of N-terminal methionine on protein.

v. *Polypeptide Chain Copy Number per Phage Particle*

Phage particles are made out of many discreet polypeptide chains which are generally present in conserved numbers of chains per particle throughout a phage population. Through a series of dilutions of Syn5 particles on SDS-gels stained with Coomassie Blue, we determined the rough copy number of each protein per Syn5 particle. The intensity of each protein band was determined using ImageQuant (Molecular Dynamics) software. Each band intensity, representative of the total concentration of protein, was normalized to the predicted molecular weight of the peptide chain. Determination the copy number of each chain was achieved by comparing the normalized intensity of each peptide chain to the normalized intensity of the portal band. The portal band was assumed to be present in a copy number of 12 chains per particle, a conserved number within all studied dsDNA phages. By using this factor, it was possible to assign a copy number to each of the other peptides (Table 4-3). Copy numbers were rounded to the nearest whole number, as was the standard error.

Each detectable Syn5 protein was similar in copy number to its corresponding protein in T7, with the exception of the tail fiber. T7 tailfibers are trimers, and each particle contains six homotrimers, for a total of 18 tail fiber peptide chains per particle. In contrast, the Syn5 particle appeared to contain only 3 tail fiber peptide chains, which may indicate that the Syn5 particle only has one tail fiber, or has a non-trimeric tail fiber, or is not homologous with the T7 tail fiber.

vi. *Western Blotting*

To prepare polyclonal antibodies, NZ white rabbits were immunized with purified Syn5 particles. A Western blot was then performed, using purified Syn5 particles. Particles were electrophoresed through a 10% polyacrylamide gel containing SDS, electroblotted to PVDF membrane, probed with anti-Syn5 polyclonal rabbit sera, and 2° anti-rabbit antibodies, and visualized using ECF substrate.

The resulting visible bands are shown in Figure 4-1C. Most of the major structural proteins were detectable, including those encoded by ORFs 45, 46, 41, 54, 37, 53, 39 and 40. Also visible were the discreet bands that were determined by mass-spectroscopy to be conglomerates of other structural proteins, including the bands at 76kDa, 38kDa, and 28kDa. Not detected were bands encoded by ORFs 43, 44, and 58. ORFs 43 and 44 may not have been detected as they are predicted to be internal virion proteins, and therefore may not have been exposed to the immune system of the rabbit during antibody production. ORF 45 is also predicted to be an internal virion protein; however, the corresponding protein found in T7 (gp 16) is the first protein that comes out of the particle after adhesion to a host cell to form the extensible tail that crosses the membrane of the host cell and facilitates DNA injection (Struthers-Schlinke et al., 2000). If the protein encoded by ORF45 acts in a similar manner as gp16 of T7, and cellular membranes were present within the preparation, free particles might bind to these membranes and begin the process of infection. The protein encoded by ORF 45 would

then travel out of the particle. Therefore, this protein could have been exposed to the immune system of the rabbit and have antibodies created against it. The protein encoded by ORF 58 has unknown function, but may also be located internally in the particle as it was not recognized by the rabbit antibodies.

The intensity of the bands of the Western blot varies in comparison to the intensity of the protein bands seen on the SDS-gel. Phage tail fibers are known to be highly antigenic and may stimulate antibody production at higher rates than other phage proteins (Edgar and Lielausis, 1965). The Syn5 tail fiber band (encoded by ORF 46) was more intense on the Western blot than it appeared on the SDS-gel, as were bands produced by Tail Tube B, and the proteins encoded by ORFs 53 and 54. This may indicate that these two unknown proteins are also highly antigenic and therefore may also play a role in host cell recognition.

D. DISCUSSION

We directly analyzed cyanophage proteins through SDS-polyacrylamide gel electrophoresis, Western blotting, mass-spectroscopy, and N-terminal sequencing. Through these methods, it was possible to assign 11 cyanophage structural proteins to predicted ORFs within the Syn5 genome.

The genome sequence of Syn 5 (as reported in Chapter III) revealed similarity to T7 and to P-SSP7, both in terms of organization of the genome as well as similarity of

genes at the amino acid level. In T7, the structural genes are located towards the end of the genome (Figure 4-4), as are the Syn5 genes. In T7-like phages, not only the location but the specific order of major structural genes appears to be conserved: portal, followed by scaffold, major/minor capsid, tail tube A, tail tube B, several internal virion proteins, and lastly, the tail fiber. In T7 nomenclature, the proteins which plug the portal hole and facilitate the binding of the tail fibers to the particle are called "tail tube" proteins. This term is slightly misleading, as the short-tailed phages do not possess the long tail tubes and sheathes that are present in the long tailed phages. Instead, these proteins might more appropriately be called "Tail Hub A" and "Tail Hub B", in terms of their function, and are more similar to the so-called "neck" proteins of P22, which open for DNA injection during infection.

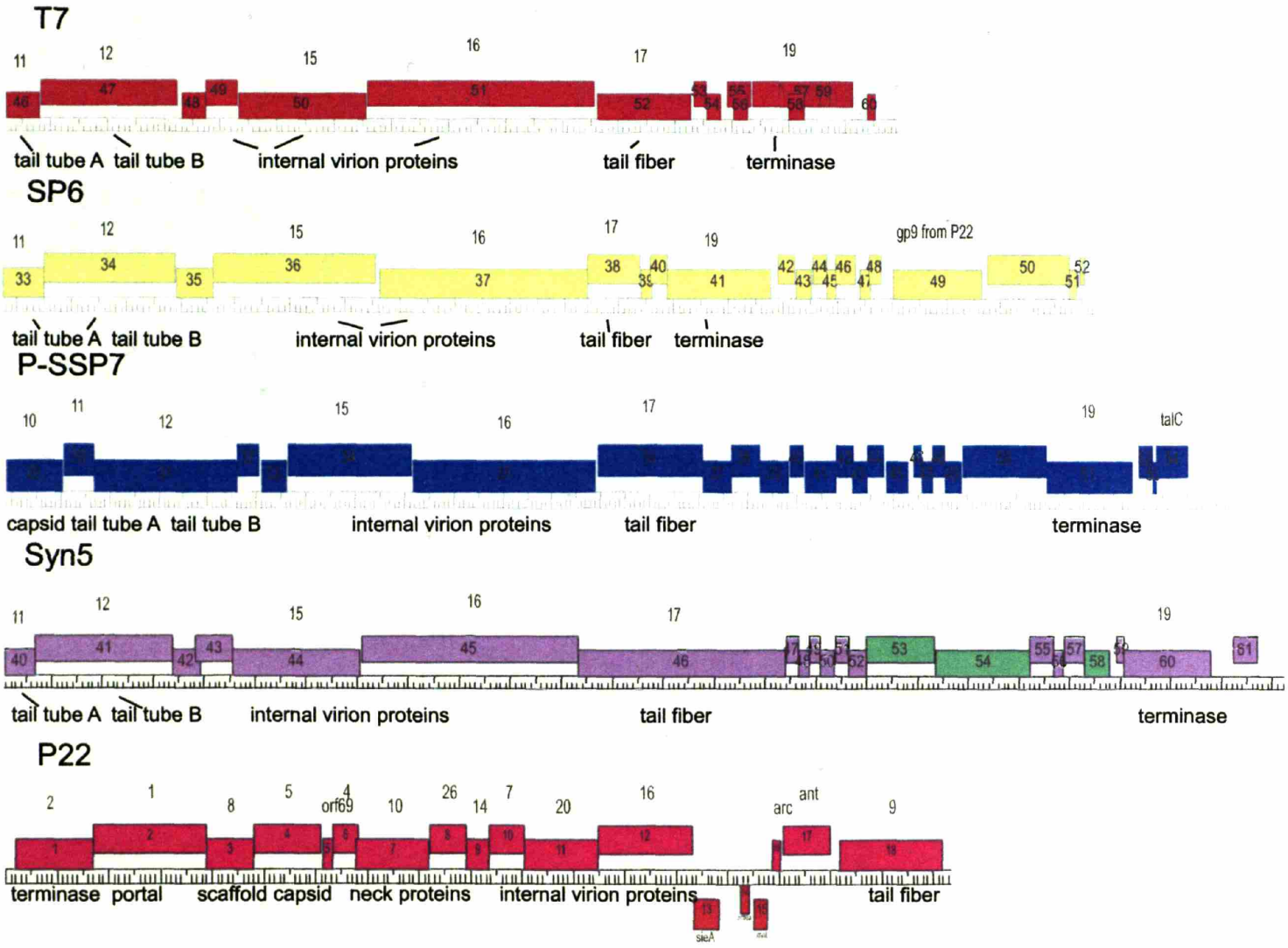
With regards to the structural protein genes, the Syn5 particle contained proteins similar to the T7 structural proteins: portal, capsid, tail tube A and B, internal virion proteins, and tail fiber (Figure 4-4). The genes encoding these proteins are approximately the same length and in the same order in the Syn5 genome as the genes that encode in the T7 proteins in the T7 genome.

Within the structural cassette of the Syn5 genome (between the portal and terminase genes) an additional 10kb of DNA more than in the T7 genome lay between the tail fiber and terminase genes. Several long ORFs (~1.2kb and ~1.8kb) were located between the tail fiber and the DNA packaging genes that were not present in T7, P-SSP7,

or other T7-like phages examined. It was not possible to determine putative functions for these genes through comparison to the NCBI NR database via the BLASTp program.

Figure 4-4: Structural arms of Syn5, T7, P-SSP7, SP6 genomes.

Comparison of structural arms of four T7-like phages and P22. Genes are labeled according to T7 nomenclature, with the exception P22 which is labeled according to P22 nomenclature. Novel Syn5 structural proteins (as determined by mass-spectroscopy) are indicated in green. Figure generated by DNA Master.


T7
tail tube A
tail tube B
internal virion proteins
tail fiber
terminase
SP6
gp9 from P22
tail tube A
tail tube B
internal virion proteins
tail fiber
terminase
P-SSP7
talC
capsid
tail tube A
tail tube B
internal virion proteins
tail fiber
terminase
Syn5
tail tube A
tail tube B
internal virion proteins
tail fiber
terminase
P22
orf69
arc
ant
terminase
portal
scaffold
capsid
neck proteins
internal virion proteins
tail fiber

In Syn5, three proteins have been identified via mass-spectroscopy and N-terminal sequencing that correspond to large ORFs located at the end of the genome, beyond the tail fiber. Two of these ORFs, 53 and 54, produce large proteins, 47 and 65 kDa respectively. The third ORF, 58, produces a small 16kDa protein, the smallest protein detected in the particle via SDS gel electrophoresis. The larger two proteins were both similar to known fibrous sequences, the protein encoded by ORF 53 to a large protein involved in aggregation in *Microbulbifer degradans*, and the protein encoded by ORF 54 to the RTX protein of *Cholera* and outermembrane proteins in *Shewanella* and *Burkholderia*. Both proteins were highly antigenic. These proteins may be alternate host recognition proteins in addition to the tail fiber. Analyses of SDS-gels of Syn5 particles indicated that all the assigned proteins were in similar copy numbers per particle as the corresponding T7 structural proteins with the exception of the tail fiber. As the tail fibers are primary host recognition proteins, the lack of Syn5 tail fibers also indicate that Syn5 may use an alternate host recognition protein in additional to the identifiable tail fiber. The role of the smaller protein remains unclear.

Further evidence that the proteins encoded by ORFs 53 and 54 may be fibrous proteins involved in host recognition was discovered with the use of a computer algorithm. This program, named "Beta-wrap"(Bradley et al., 2001a; Bradley et al., 2001b), recognizes the beta-helix motif common to many viral adhesins and rare in other known proteins (Weigele et al., 2003), and has been used to identify putative cyanophage tail fibers in S-PM2 (Mann et al., 2005). The amino acid sequence encoded by ORFs 53,

54, and the putative tail fiber (46) were all scored by Beta-wrap similarly to the tailspike protein of P22, which is known to have the parallel-beta helix fold.

Mass-spectroscopic analysis and N-terminal sequencing allowed the assignment of discreet Syn5 ORFs to eleven of the twelve bands visible on a SDS-gel. The remaining discreet band within the SDS-gel appeared to be a mixture of peptides encoded by multiple ORFs. This remaining band may be the result of an artifact of sample handling; for example, a specific part of the phage particle, like tail fibers, may be fragile and likely to break during the purification process; the broken fibers may then be more sensitive to proteolysis at a specific site and therefore produced a discreet band; or there may be a very stable structure on the phage particle made of multiple proteins which does not fully denature in the presence of SDS during boiling. As N-terminal sequencing of this band yielded no results and all of the protein bands on the gel yielded peptides encoded by multiple Syn5 ORFs when analyzed by mass-spectroscopy, it is difficult to determine the origin of the remaining band. One possibility is that the band is a host protein that somehow became incorporated into the phage particle, and therefore its sequence would not be in the Syn5 genome database.

Based on the positions of the encoding ORFs in the genomic map and the identification of these ORFs by BLASTp, it is likely that at least two other proteins may be involved in the assembly of the Syn5 particles but not found in the mature particle, the scaffolding protein and the DNA packaging protein. The scaffolding protein would aid in the formation of procapsid, comprised of the capsid units and portal protein. Scaffolding

proteins then exit through pores in the procapsid shell to be reused in further rounds of procapsid formation or are proteolytically cleaved as the capsid becomes filled with DNA and expands to the mature icosehedral shell. The DNA packaging proteins aid the filling of the phage head with DNA, generally by binding to an individual phage genome and to the empty procapsid. The phage DNA is then translocated into the head of the phage. Upon completion, the packaging protein dissociates from the head, and is therefore also not found in the mature particle.

Our future efforts involve the cloning and expression of several of the identified cyanophage proteins, including two of the "unknown" proteins encoded by ORFs 53 and 54, for *in vitro* analysis of native cyanophage proteins.

CHAPTER V: CONCLUDING REMARKS

The study of the cyanophage structural proteins, specifically the tail proteins, is the first step towards understanding cyanophage/host interactions and the rates at which these interactions occur in the oceans. Elucidation of these rates is critical towards determination of the amount of lateral gene transfer between seemingly unrelated host organisms. With an estimated 0.02% of the possible phage genomes sampled globally, and the presence of genes of unknown function within the majority of phage genomes, phages represent one of the largest reservoirs of unknown genomic sequences on Earth (Brussow and Hendrix, 2002). As there are numerous examples of prophages conferring new abilities on infected host cells with detrimental effects on mankind (Boyd et al., 2002; Waldor and Mekalanos, 1996), the potential exists for a novel phage/host combination to arise from promiscuous gene transfer with deleterious effects. The increases of *E. coli* and other sewage bacteria in the coastal areas of the oceans from human activities makes the questions of how much phage mediated transfer exists between these bacteria and the marine genera which produce a significant portion of our oxygen; and what effect, if any, this gene transfer may have very important.

A. PURIFICATION OF CYANOPHAGE SYN5

Biochemical studies of cyanophage structural have been hampered by the lack of a robust protocol for the purification and production of high titer stocks of these particles.

By development of a culture vessel to maximize host cell growth and optimization of an enteric phage concentration protocol via PEG precipitation, it was possible to produce milliliters of purified phage at a concentration of 10^{12} phage per ml. This was more than sufficient to elucidate some characteristics of the phage/host system, isolate double stranded genomic DNA, and to examine the particle proteins by SDS-PAGE. Syn5 appeared to exhibit lower burst sizes than its enteric counterpart, T7; as well as a longer lytic period. It was not possible to determine if these lowered parameters were a function of the lowered rate of cellular growth (i.e. S*ynechococcus* does not double as quickly as *E. coli*, and therefore, the phage exhibited a lowered burst size), or an intrinsic property of the cyanophage/cyanobacteria system (Syn5 will never achieve the burst size or rapid lytic period of T7, even if it was modified to infect *E. coli*). Further experimental work would be needed to distinguish between the two scenarios.

In terms of the physical properties of the Syn5 particles, the phage exhibited many similarities to the enteric phage T7. First of all, the Syn5 genome was comprised of dsDNA sensitive to digestion by several restriction enzymes. This indicated that the phage DNA was unmodified, similar to the genome of T7, and dissimilar to several known cyanophages. The genome appeared to exhibit a restriction pattern consistent with a genome size of ~40-60kb, as would be expected based on the diameter of the phage head as observed by TEM, and the likelihood that the phage genome completely fills the head.

Further supporting evidence that Syn5 was similarly constructed to T7 was observed in the examination of the particle proteins. Syn5 particles were comprised twelve discreet proteins bands; three more discreet structural proteins than T7. Although it was not an identical number of proteins as T7, twelve proteins comprising a particle was still a reasonable number of proteins to be found within a short-tailed phage. The longer tailed phages of myoviridae and siphoviridae may be comprised of ~30-40 structural proteins.

B. GENOME SEQUENCE OF SYN5

The complete genomic sequence of Syn5 revealed that it shared many genes in common and genome organization with the enteric phage T7 and the cyanophages P-SSP7 and P60. Analysis of the ~46kb genome revealed that the most conserved genes between the four phages belonged to the structural gene cassette and the DNA replisome. Additionally, the Syn5 genome possessed a gene encoding an RNAP, located at the beginning of the genome, which may be indicative that its method of infection is similar to that of T7. However, it also possessed an integrase gene similarly to P-SSP7, indicating that Syn5 may be a temperate rather than virulent phage.

The DNA replisome contained several genes that are not present in the genome of T7; such as thioredoxin and ribonucleotide reductase. The proteins encoded by these genes may aid the phage in the scavenging of the host's nucleotides, and therefore may be a result of adaptation to a nutrient poor environment. Furthermore, the T7 DNAP is

known to use the host's thioredoxin as a co-factor to increase processivity during DNA replication (Bedford et al., 1997). The presence of the Syn5 thioredoxin gene may indicate that Syn5 contains a complete DNA replisome, unreliant on host co-factors. This may be evidence that Syn5 and T7 were both evolved from a common ancestor; and T7 subsequently lost portions of its DNA replisome and its ability to form a prophage when it gained the ability to infect enteric host cells that could grow rapidly in rich environments to higher cell concentrations.

The Syn5 genome was approximately 7kb longer than the T7 genome. A large portion of the extra genome length in Syn5 was evident in a long stretch of DNA between the putative tail fiber gene and putative DNA packaging gene. This portion of the Syn5 genome contained several large ORFs of unknown function that are not similar to known phage proteins within the NCBI NR database. Several of these ORFs were preceded by enteric Shine-Dalgarno sequences, which may be evidence of more recent enteric transfer.

The Syn5 genome did not contain any recognizable photosynthetic genes or nutrient acquisition genes, as have been observed in other cyanophages (Lindell et al., 2004; Mann et al., 2005; Mann et al., 2003; Sullivan et al., 2005).

C. IDENTIFICATION OF THE STRUCTURAL PROTEINS OF SYN5

Assignment of the visible protein bands produced when purified Syn5 particles were electrophoresed through an SDS-polyacrylamide gel to specific ORFs within the annotated Syn5 genome was possible through mass-spectroscopy and N-terminal sequencing of the protein bands. By these methods, it was possible to determine that Syn5 particles were comprised of similar proteins as the enteric phage, including portal, capsid, two distinct tail proteins, three distinct internal virion proteins, and tail fibers. These proteins were present within the particle in similar copy numbers as the corresponding proteins in T7, with the exception of the putative tail fiber which was only present in a copy number of three chains per particle as opposed to T7's eighteen chains per particle. The lack of tail fibers was indication that Syn5 may use other proteins as host recognition proteins in addition to its tail fiber(s).

Three proteins were identified in the Syn5 particle which were not similar to any T7 proteins. Two of these proteins, encoded by ORFs 53 and 54 in the Syn5 genome, did exhibit some similarity to known microbial fibrous proteins, and also exhibited high antigenicity when probed with polyclonal antibodies to whole Syn5 particles. The amino acid sequences of these two proteins were likely to contain parallel beta-helix motifs, as determined by the algorithm, Beta-wrap, a fold which is uncommon in the protein data bank but quite common among viral adhesins. This evidence led to the conclusion that the proteins encoded by ORFs 53 and 54 may function as additional host recognition proteins, thus explaining the lack of tail fibers on the Syn5 particle.

D. CONCLUDING REMARKS

i. *Evolutionary relationship between Syn5 and T7*

With the relative ease and low cost of DNA sequencing today, many phage genomes have been completed (Paul et al., 2002), setting the stage for exploring how related are any given phages. The genomic and structural similarities observed between the enteric and cyanophages leads one to question how these phages evolved. Simply put, are Syn5 and T7 descendants of the same distant (or not so distant) viral ancestor? Or is it possible, given the amount of proposed lateral gene transfer and mosaicism observed in phage genomes, to create an enteric phage from a cyanophage or vice versa? These questions might be answered by an examination of the evolutionary relatedness of the host strains in comparison to the relatedness of the phages themselves. If the phages and hosts exhibit a similar evolutionary distance from each other, it may be likely that the two phage types are descended from a common ancestor and have co-evolved with their respective hosts. On the other hand, if the phages are more closely related than their hosts, one phage may have acquired the necessary genes to infect a different host more recently.

Ideally, the evolutionary comparison between the two types of organisms would be performed by the same method. However, the standard for bacterial evolutionary relationships is based on comparisons of sequence of the 16s ribosomal RNA, and phages (at least, no phages that have been isolated to date) do not encode their own ribosomes.

Therefore, these questions can not be addressed until a method of determining how related two individual phages are is decided upon that is also applicable to the host organisms.

The initial phage classifications proposed by International Committee on Virus Taxonomy (ICTV) were based on a Linnaenan system of classification that included the nucleic acid which carried the genetic information and the morphology of the phage particle (van Regenmortel et al., 2000). However, analyses of the complete genomes revealed that only limited relevance could be assigned to the morphologic classifications, as phages with little morphologic resemblance to each other could be so similar as to form DNA hybrids (P22 and lambda) (Botstein and Herskowitz, 1974), while phages which appeared morphologically very similar might possess comparatively few genes in common (P22 and *Salmonella* phage Sf6).

Three possible solutions have been outlined to solve the problems in current phage taxonomy and are discussed in a review by Nelson (Nelson, 2004). The phage proteomic tree, as proposed by Rohwer and Edwards, grouped each phage relative to its nearest neighbor as well as to all phages, using the BLASTp program. The resulting tree eliminated some problems with the ICTV taxonomy (placing P22 with the siphoviridae, etc) (Rohwer and Edwards, 2002). The Pittsburgh Bacteriophage Institute proposed an alternate solution, which took into account phage differences based on carrier of genetic information, and further suggested that classification into subgroups be based on whole genome sequence data, with individual phages belonging to multiple groups if necessary

(Lawrence et al., 2002). Lastly, Proux et al, classified dairy siphophages on the basis of their structural gene module, as it is believed to be the oldest and most conserved module in dairy phage (Proux et al., 2002). Similar conservation of phage structural gene sequences has been observed in the freshwater and marine phage populations (Short and Suttle, 2005).

With the limited sample size of cyanophage genomes available (6), and even more limited sample size of podocyanophage genomes (3), overall classification of cyanophages within a group may not yet be possible. The lack of a sequenced genome for WH8109 precludes the comparison of the whole genomes between the host cells. However, by using the third method of comparison—comparison of structural proteins---it may be possible to preliminarily determine the relatedness of Syn5 to the enteric phage T7 and compare it to the relatedness of *Synechococcus* WH8102 to *E. coli*, with the point of determining whether the phages or the hosts that they infect represent a larger evolutionary divergence. By using highly conserved structural proteins within the host cells, it is likely that the level of difference between WH8109 and WH8102 would be negligible as compared to the differences between WH8102 and *E. coli*.

As the hosts do not possess phage structural proteins, a comparison between several of the ribosome proteins can be made instead. Ribosomes are comprised of approximately fifty different proteins chains, however, several of these have been shown to possess high levels of sequence similarity, as these proteins are known to interact directly with RNA (Liljas and Garber, 1995). The protein sequences of ribosomal

proteins S5, S6, L1, L6, and L9 were examined. These proteins were all ~100-200 amino acids in length. When compared by BLASTp, the sequences of these proteins with the exception of L1, did not exhibit significantly different amino acid similarity or identity from the amino acid similarity and identity exhibited by the comparison of the phage structural proteins. S5 and L6 were approximately 40% identical (60% similar) between the two host strains, L9 and S6 were approximately 26-30% identical (50% similar); while the phage structural proteins were ~30% identical (48% similar). e-values produced by the BLASTp comparison ranged from 1e-73 (portal) to 0.002 (S6) (Table 5-1). If this type of evolutionary analysis is valid, this would indicate that the enteric and cyanophage structural proteins of T7 and Syn5 are descendants of a common ancestral phage, rather than the result of a one phage that recently acquired a new host recognition protein. This does not preclude the creation of an enteric phage from a cyanophage by a new host recognition protein, but is suggestive that it is not the case for Syn5 and T7.

	identical	Similar	e value
S5	40%	64%	1e-27
S6	26%	52%	0.002
L1	No similarity		
L6	40%	59%	4e-34
L9	30%	51%	2e-13
Portal	33%	54%	1e-73
Capsid	26%	43%	4e-20

Table 5-1: Relatedness of Host Ribosomal and Phage Structural Proteins. The protein sequences of five ribosomal RNA binding proteins, S5, S6, L1, L6, and L9 were compared between *Synechococcus* strain WH8102 and *E. coli* B. Two phage proteins, portal and capsid, were compared between Syn5 and T7 using BLASTp. Percent similarity, percent identity and the e-values of the comparisons are reported.

ii. *Lateral gene transfer and "enteric" sequences in Syn5*

The presence of S-D sequences that resemble the enteric S-D sequences and of structural proteins whose amino acid sequences most resemble enteric fibrous sequences, such as those of ORFs 53 and 54, could represent recent acquisition of enteric genes within the Syn5 genomes. If the presence of S-D sequences, as determined by DNA Master, in a cyanophage is a signal of newly acquired enteric sequences, how did cyanophage Syn5 acquire them? Currently, Syn5 is known to infect only WH8109. The phage may have a temperate lifestyle as evidenced by its putative integrase gene, and therefore may have gained some genes directly from the host through imprecise excision of the prophage. This would imply that WH8109 contained similar enteric sequences available for phage acquisition. Exploring this further will await the genome sequence of WH8109.

Conversely, T7 may have lost the corresponding fibrous sequences that Syn5 maintains. This would indicate that the presence of the enteric S-D sequences at the beginning of some of these genes is either by chance or that the S-D sequence enables some unknown function in cyanobacteria. Ting et al (2002) demonstrated how the smaller genome of *Prochlorococcus* may have lost some of the genes that are present in the larger genome of *Synechococcus* through differentiation and niche adaptation (Ting et al., 2002). The ability to replicate within an organism optimally adapted to a mammalian

gut would certainly indicate that the enteric phage may have undergone niche adaptation in a similar manner as *Prochlorococcus*.

WH8109 appeared to be particularly susceptible to infection by many different isolated cyanophages in comparison to other cultured *Synechococcus* and *Prochlorococcus* strains (Sullivan et al., 2003; Waterbury and Valois, 1993). This sensitivity may be due to loss of resistance genes during its years as a laboratory strain, or it may simply be a naturally highly phage-sensitive strain. A naturally phage sensitive strain could occur if it possessed some extremely common motif within its lippopolysaccharide that was recognizable by many phages, or lacked restriction enzymes capable of chewing up invading DNA, and was capable of growing faster in nature than the rate of lytic phage infection. In either case, there may exist phages which are capable of infecting both WH8109 and a heterotrophic host (or the phage may already be in collection, and the host is not yet cultured); and therefore Syn5 could have acquired the putative enteric genes through mixed phage infections.

iii. *Novel cyanophage host recognition proteins*

The lack of putative tail fibers detectable on the Syn5 particles raises some interesting questions. Given the dilute environment that the phage was isolated from, it seems highly unusual that the phage particle would contain only a few polypeptide chains designed to recognize and bind to a host cell (which, if the putative tail fiber protein is

truly a homotrimer as are other known viral adhesins, would indicated that the particle simply had one large spike with which it managed to detect and bind to the host).

To put this in perspective, the larger enteric phages of the myoviridae have both numerous long tail fibers which recognize the correct host cell strain(s), and six shorter tail fibers which recognize the specific receptor on the host cell surface for DNA injection. Siphoviridae newly isolated from the environment can have four long tail fibers which help recognize the correct host cell(s) and correctly position the phage on the host cell surface for DNA injection. Enteric podoviridae may have six tail fibers, but only made of one type of polypeptide chain; P22 requires at least three tailspikes (nine polypeptide chains) to retain infectivity. Cyanophages of the larger morphologic types appear to have similar numbers of tail fibers as the enteric phages of the same types. It would seem likely that phages isolated from a dilute environment would possess more, not less, fibers which aided in host recognition. So why does Syn5 have so few tail fibers?

One answer to this question is that Syn5 may be using multiple proteins as host cell detectors. This would be the first evidence of such a phenomenon in podoviridae, which are generally less complex structurally than the larger phages. The two larger ORFs at the end of the Syn5 genome are likely candidates for encoding such proteins, however, their exact structure and location on (or in) the particle are unknown. The putative tail-fiber contains a region similar to the portal binding domain of the T7 tail fiber; however, the other two proteins do not. Therefore, these two proteins could literally

be bound to anything anywhere on the outside of the phage particle, or immediately exposed upon receptor binding, such that they were detectable by the immune system of a rabbit. Further work is necessary to elucidate the answers to these questions.

iv: *Evolution of tailed phages*

The icosehedral symmetry of the head of a phage is quite striking when viewed by TEM. However, it is the asymmetric vertex which is actually more important in terms of mediation of phage/host interactions. The most highly conserved part of this asymmetric vertex is the portal: a dodecameric ring, comprised of a single polypeptide species, which is part of the initial nucleation of the assembly of the phage capsid, allows DNA to pass into the cell by docking with the terminase, and binds to the plug proteins which prevent the DNA from spilling back out of the capsid. These portal proteins have been isolated from phages phi29 and P22 and the structures determined by protein crystallography and X-ray diffraction (Badasso et al., 2000; Cingolani et al., 2002; Simpson et al., 2001). Discreet portal rings similar to those of the tailed phages have also been discovered in "symmetric" phages and viruses like PRD1 and the eukaryotic viruses adenovirus and herpes simplex 1 (Isidro et al., 2004; Jaatinen et al., 2004; Trus et al., 2004). However, it was not possible to closely observe the portal structure and an assembled asymmetric vertex within intact phage particles at high resolution until recently. Cryo-electron microscopists were able to reconstruct the asymmetric vertex of phage epsilon15, a *Salmonella* phage that exhibits high genome syteny with T7 (pers. comm.,Wen Jiang, Peter Weigele, Jonathan King, and Wah Chiu;). Through the examination of this

reconstruction, it became apparent that the asymmetric vertex of the phage is truly the "business" end, as it appeared that all the internal proteins and external proteins are clearly visible within the one asymmetric vertex.

These conserved structural features, the portal ring and single asymmetric vertex, found throughout viral isolates suggest that it was the development of these features---which allow for efficient DNA delivery--- that was important in the origin of viruses. As life on Earth is presumed to have begun in the oceans, one could extrapolate that the development of the phage and its tail also originated in the marine environment. The subsequent radiation and adaptation of bacteriophages to discreet environments---mammalian guts, hotsprings, soil---required the preservation of the DNA delivery mechanism, thus explaining the similarity in gross morphology found among dsDNA bacteriophage isolates.

A further necessary development of the evolution of the tail structure was its ability to rapidly change its host-recognition proteins. Some phages, like the P22-like podoviridae, are exquisitely specific with regards to their host organism, infecting only one or two specific microbial strains. However, this specificity is conferred through small portions of the tail fibers, as a section of a given tail fiber protein may be highly specific to a certain O-antigen, and the remainder of the protein exhibit high levels of similarity with other phage tail fibers. This similarity in the remainder of the fiber proteins enables high rates of gene recombination (sometimes even within the same phage, like mu) among the fiber genes and a broadening of phage host range. This would account for the

mosaicism present in sequenced phage genomes and the presence of phage gene cassettes within these genomes. Further study of the nature of viral evolution could be pursued by the determination of host cell receptors in cyanobacteria, and characterization of the features of the proteins involved in phage/host interactions.

CHAPTER VI. BIBLIOGRAPHY

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.

Anderson, T.F. (1975) Some personal memories of research. *Annu Rev Microbiol*, **29**, 1-19.

Azam, F. (1998) Microbial control of oceanic carbon flux: the plot thickens. *Science*, **280**, 694-696.

Badasso, M.O., Leiman, P.G., Tao, Y., He, Y., Ohlendorf, D.H., Rossmann, M.G. and Anderson, D. (2000) Purification, crystallization and initial X-ray analysis of the head-tail connector of bacteriophage phi29. *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 9)**, 1187-1190.

Bedford, E., Tabor, S. and Richardson, C.C. (1997) The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on Escherichia coli DNA polymerase I. *Proc Natl Acad Sci U S A*, **94**, 479-484.

Bergh, O., Borsheim, K.Y., Bratbak, G. and Heldal, M. (1989) High Abundance of Viruses Found in Aquatic Environments. *Nature*, **340**, 467-468.

Besemer, J., Lomsadze, A. and Borodovsky, M. (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res*, **29**, 2607-2618.

Bohannan, B. and Lenski, R. (2000) Linking genetic change to community evolution:insights from studies of bacteria and bacteriophage. *Ecol Letters*, **3**, 362-377.

Borsheim, K.Y., Bratbak, G. and Heldal, M. (1990) Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. *Appl Environ Microbiol*, **56**, 352-356.

Botstein, D. and Herskowitz, I. (1974) Properties of hybrids between Salmonella phage P22 and coliphage lambda. *Nature*, **251**, 585-589.

Boyd, E.F., Davis, B.M. and Musser, J.M. (2002) Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. *Trends Microbiol*, **9**, 137-144.

Bozzola, J.J. and Russell, L.D. (1992) *Electron Microscopy*. Jones and Bartlett Publishers, Sudbury, MA.

Bradley, P., Cowen, L., Menke, M., King, J. and Berger, B. (2001a) BETAWRAP: successful prediction of parallel beta -helices from primary sequence reveals an association with many microbial pathogens. *Proc Natl Acad Sci U S A*, **98**, 14819-14824.

Bradley, P., Cowen, M., Menke, J., King, J. and Berger, B. (2001b) Predicting the beta-helix fold from protein sequence data. *RECOMB, Fifth Annual International Conference on Molecular Biology on Computational Molecular Biology*, Montreal, Canada.

Bratbak, G. (1990) Viruses as partners in spring bloom microbial trophodynamics. *Appl Environ Microbiol*, **56**, 1400-1405.

Bratbak, G., Heldal, M., Thingstad, T.F., Riemann, B. and Haslund, O.H. (1992) Incorporation of viruses into the budget of microbial C-transfer: A first approach. *Marine Ecology Progress Series*, **83**, 273-280.

Brussow, H. and Hendrix, R.W. (2002) Phage genomics: small is beautiful. *Cell*, **108**, 13-16.

Bukhari, A.I. and Ambrosio, L. (1978) The invertible segment of bacteriophage Mu DNA determines the adsorption properties of Mu particles. *Nature*, **271**, 575-577.

Casjens, S. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol*, **49**, 277-300.

Cerritelli, M.E., Wall, J.S., Simon, M.N., Conway, J.F. and Steven, A.C. (1996) Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. *J Mol Biol*, **260**, 767-780.

Chen, F. and Lu, J.R. (2002) Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Applied and Environmental Microbiology*, **68**, 2589-2594.

Cingolani, G., Moore, S.D., Prevelige, P.E., Jr. and Johnson, J.E. (2002) Preliminary crystallographic analysis of the bacteriophage P22 portal protein. *J Struct Biol*, **139**, 46-54.

Cockburn, J.J., Abrescia, N.G., Grimes, J.M., Sutton, G.C., Diprose, J.M., Benevides, J.M., Thomas, G.J., Jr., Bamford, J.K., Bamford, D.H. and Stuart, D.I. (2004) Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature*, **432**, 122-125.

Crowther, R.A., Lenk, E.V., Kikuchi, Y. and King, J. (1977) Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J Mol Biol*, **116**, 489-523.

Curtis, S. and Martin, J. (1994) The transcription apparatus and the regulation of transcription. In Bryant, D. (ed.), *The molecular biology of cyanobacteria*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Delcher, A.L., Harmon, D., Kasif, S., White, O. and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, **27**, 4636-4641.

Dobbins, A.T., George, M., Basham, D.A., Ford, M.E., Houtz, J.M., Pedulla, M.L., Lawrence, J.G., Hatfull, G.F. and Hendrix, R.W. (2004) Complete genomic sequence of the virulent Salmonella bacteriophage SP6. *Journal of Bacteriology*, **186**, 1933-1944.

Drexler, K., Riede, I. and Henning, U. (1986) Morphogenesis of the long tail fibers of bacteriophage T2 involves proteolytic processing of the polypeptide (gene product 37) constituting the distal part of the fiber. *J Mol Biol*, **191**, 267-272.

Dunn, J.J. and Studier, F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol*, **166**, 477-535.

Earnshaw, W.C. and Casjens, S.R. (1980) DNA packaging by the double-stranded DNA bacteriophages. *Cell*, **21**, 319-331.

Edgar, R.S. and Lielausis, I. (1965) Serological studies with mutants of phage T4D defective in genes determining tail fiber structure. *Genetics*, **52**, 1187-1200.

Fargo, D.C., Zhang, M., Gillham, N.W. and Boynton, J.E. (1998) Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in Chlamydomonas reinhardtii chloroplasts or in Escherichia coli. *Molecular and General Genetics*, **257**, 271-282.

Garcia, E., Elliott, J.M., Ramanculov, E., Chain, P.S., Chu, M.C. and Molineux, I.J. (2003) The genome sequence of Yersinia pestis bacteriophage phiA1122 reveals an intimate history with the coliphage T3 and T7 genomes. *J Bacteriol*, **185**, 5248-5262.

George, D.G., Yeh, L.S. and Barker, W.C. (1983) Unexpected relationships between bacteriophage lambda hypothetical proteins and bacteriophage T4 tail-fiber proteins. *Biochem Biophys Res Commun*, **115**, 1061-1068.

Goericke, R. (1993) The marine prochlorophyte Prochlorococcus contributes significantly to phytoplankton biomass and primary production in the Sargasso Sea. *Deep Sea Research Part II*, **40**, 2283-2294.

Goldenberg, D., Berget, P. and King, J. (1982) Maturation of the tail spike endorhamnosidase of Salmonella phage P22. *Journal of Biological Chemistry*, **257**, 7864-7871.

Goldenberg, D.P. and King, J. (1981) Temperature-sensitive mutants blocked in the folding or subunit of the bacteriophage P22 tail spike protein. II. Active mutant proteins matured at 30 degrees C. *J Mol Biol*, **145**, 633-651.

Gordon, D., Abajian, C. and Green, P. (1998) Consed: A graphical tool for sequence finishing. *Genome Research*, **8**, 195-202.

Grundy, F.J. and Howe, M.M. (1984) Involvement of the invertible G segment in bacteriophage mu tail fiber biosynthesis. *Virology*, **134**, 296-317.

Hadas, H., Einav, M., Fishov, I. and Zaritsky, A. (1997) Bacteriophage T4 development depends on the physiology of its host Escherichia coli. *Microbiology*, **143 (Pt 1)**, 179-185.

Haggard-Ljungquist, E., Halling, C. and Calendar, R. (1992) DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J Bacteriol*, **174**, 1462-1477.

Hardies, S.C., Comeau, A.M., Serwer, P. and Suttle, C.A. (2003) The complete sequence of marine bacteriophage VpV262 infecting vibrio parahaemolyticus indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology*, **310**, 359-371.

Hausmass, R. (1988) The T7 Group. In Calendar, R. (ed.), *The Bacteriophage*. I Plenum, New York, NY.

Heldal, M. and Bratbak, G. (1991) Production and decay of viruses in aquatic environments. *Marine Ecology Progress Series*, **72**, 205-212.

Hendrix, R.W. and Duda, R.L. (1992) Bacteriophage lambda PaPa: not the mother of all lambda phages. *Science*, **258**, 1145-1148.

Hendrix, R.W., Lawrence, J.G., Hatfull, G.F. and Casjens, S. (2000) The origins and ongoing evolution of viruses. *Trends Microbiol*, **8**, 504-508.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E. and Hatfull, G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc Natl Acad Sci U S A*, **96**, 2192-2197.

Henning, U. and Jann, K. (1979) Two-component nature of bacteriophage T4 receptor activity in Escherichia coli K-12. *J Bacteriol*, **137**, 664-666.

Horne, R.W. and Wildy, P. (1961) Symmetry in virus architecture. *Virology*, **15**, 348-373.

Horne, R.W. and Wildy, P. (1963) Virus structure revealed by negative staining. *Advances in Virus Research*, **10**, 101-170.

Huiskonen, J.T., Kivela, H.M., Bamford, D.H. and Butcher, S.J. (2004) The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat Struct Mol Biol*, **11**, 850-856.

Isidro, A., Henriques, A.O. and Tavares, P. (2004) The portal protein plays essential roles at different steps of the SPP1 DNA packaging process. *Virology*, **322**, 253-263.

Jaatinen, S.T., Viitanen, S.J., Bamford, D.H. and Bamford, J.K. (2004) Integral membrane protein P16 of bacteriophage PRD1 stabilizes the adsorption vertex structure. *J Virol*, **78**, 9790-9797.

Jiang, S.C. and Paul, J.H. (1998) Gene transfer by transduction in the marine environment. *Appl Environ Microbiol*, **64**, 2780-2787.

Kamp, D., Kahmann, R., Zipser, D., Broker, T.R. and Chow, L.T. (1978) Inversion of the G DNA segment of phage Mu controls phage infectivity. *Nature*, **271**, 577-580.

Kemp, P., Gupta, M. and Molineux, I.J. (2004) Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Mol Microbiol*, **53**, 1251-1265.

Kikuchi, Y. and King, J. (1975a) Assembly of the tail of bacteriophage T4. *J Supramol Struct*, **3**, 24-38.

Kikuchi, Y. and King, J. (1975b) Genetic control of bacteriophage T4 baseplate morphogenesis. I. Sequential assembly of the major precursor, in vivo and in vitro. *J Mol Biol*, **99**, 645-672.

Kikuchi, Y. and King, J. (1975c) Genetic control of bacteriophage T4 baseplate morphogenesis. II. Mutants unable to form the central part of the baseplate. *J Mol Biol*, **99**, 673-694.

Kikuchi, Y. and King, J. (1975d) Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J Mol Biol*, **99**, 695-716.

King, J. and Laemmli, U. (1971) Polypeptides of the tail fibers of bacteriophage T4. *Journal of Molecular Biology*, **62**, 465-477.

King, J. and Wood, W. (1969) Assembly of bacteriophage T4 tailfibers: the sequence of gene product interactions. *Journal of Molecular Biology*, **39**, 533-601.

King, J. and Yu, M.H. (1986) Mutational analysis of protein folding pathways: the P22 tailspike endorhamnosidase. *Methods Enzymol*, **131**, 250-266.

Kovalyova, I.V. and Kropinski, A.M. (2003) The complete genomic sequence of lytic bacteriophage gh-1 infecting Pseudomonas putida--evidence for close relationship to the T7 group. *Virology*, **311**, 305-315.

Kreisberg, J.F., Betts, S.D., Haase-Pettingell, C. and King, J. (2002) The interdigitated beta-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization. *Protein Sci*, **11**, 820-830.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, **305**, 567-580.

Lawrence, J.G., Hatfull, G.F. and Hendrix, R.W. (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol*, **184**, 4891-4905.

Li, W.K.W. (1994) Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: Measurements from flow cytometric sorting. *Limnology and Oceanography*, **39**, 169-175.

Li, W.K.W. (1998) Annual average abundance of heterotrophic bacteria and Synechococcus in surface ocean waters. *Limnology and Oceanography*, **43**, 1746-1753.

Liao, C.P. and Syu, W., Jr. (2002) Analysis of the baseplate region of phage AR1 that specifically infects Escherichia coli O157:H7. *J Microbiol Immunol Infect*, **35**, 269-271.

Liljas, A. and Garber, M. (1995) Ribosomal proteins and elongation factors. *Current Opinion in Structural Biology*, **5**, 721-727.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F. and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc Natl Acad Sci U S A*, **101**, 11013-11018.

Liu, H., Campbell, L. and Landry, M.R. (1995) Growth and mortality rates of Prochlorococcus and Synechococcus measured with a selective inhibitor technique. *Marine Ecology Progress Series*, **116**, 277-287.

Liu, H., Campbell, L., Landry, M.R., Nolla, H.A., Brown, S.L. and Constantinou, J. (1998) Prochlorococcus and Synechococcus growth rates and contributions in the Arabian Sea during the 1995 Southwest and Northeast monsoons. *Deep Sea Research Part II*, **45**, 2327-2352.

Liu, H. and Landry, M.R. (1999) Prochlorococcus growth rates in the central equatorial Pacific: An application of the finax approach. *J Geophys Res*, **104**, 3391-3399.

Liu, H., Nolla, H.A. and Campbell, L. (1997) Prochlorococcus growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquatic Microbial Ecology*, **12**, 39-47.

Lu, J., Chen, F. and Hodson, R.E. (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting maring Synechococcus spp. in river estuaries. *Appl Environ Microbiol*, **67**, 3285-3290.

Luria, S.E. and Andersen, T.F. (1942) The identification and characterization of bacteriophages with electron microscope. *Proc Natl Acad Sci U S A*, **28**, 127-130.

Luria, S.E., Delbruck, M. and Andersen, T.F. (1943) Electron microscope studies of bacterial viruses. *J Bacteriol*, **46:57-76.**

MacCoss, M.J., Wu, C.C. and Yates, J.R., 3rd. (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem*, **74**, 5593-5599.

Makhov, A.M., Trus, B.L., Conway, J.F., Simon, M.N., Zurabishvili, T.G., Mesyanzhinov, V.V. and Steven, A.C. (1993) The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. *Virology*, **194**, 117-127.

Mann, N.H. (2003) Phages of the marine cyanobacterial picophytoplankton. *FEMS Microbiol Rev*, **27**, 17-34.

Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A. and Krisch, H.M. (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. *J Bacteriol*, **187**, 3188-3200.

Mann, N.H., Cook, A., Millard, A., Bailey, S. and Clokie, M.R. (2003) Bacterial photosynthesis genes in a virus. *Nature*, **424**, 741.

McDaniel, L., Houchin, L.A., Williamson, S.J. and Paul, J.H. (2002) Lysogeny in marine Synechococcus. *Nature*, **415**, 496.

McKenna, R., Xia, D., Willingmann, P., Ilag, L.L. and Rossmann, M.G. (1992) Structure determination of the bacteriophage phiX174. *Acta Crystallogr B*, **48 (Pt 4)**, 499-511.

McLeod, S.M., Kimsey, H.H., Davis, B.M. and Waldor, M.K. (2005) CTXphi and Vibrio cholerae: exploring a newly recognized type of phage-host cell relationship. *Mol Microbiol*, **57**, 347-356.

Michel, C.J., Jacq, B., Arques, D.G. and Bickle, T.A. (1986) A remarkable amino acid sequence homology between a phage T4 tail fibre protein and ORF314 of phage lambda located in the tail operon. *Gene*, **44**, 147-150.

Millard, A., Clokie, M.R.J., Shub, D.A. and Mann, N.H. (2004) Genetic organization of the psbAD region in phages infecting marine Synechococcus strains. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 11007-11012.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. and Ruger, W. (2003) Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews*, **67**, 86-+.

Moak, M. and Molineux, I.J. (2000) Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Mol Microbiol*, **37**, 345-355.

Montag, D. and Henning, U. (1987) An open reading frame in the Escherichia coli bacteriophage lambda genome encodes a protein that functions in assembly of the long tail fibers of bacteriophage T4. *J Bacteriol*, **169**, 5884-5886.

Montag, D., Schwarz, H. and Henning, U. (1989) A component of the side tail fiber of Escherichia coli bacteriophage lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. *J Bacteriol*, **171**, 4378-4384.

Morgan, G.J., Hatfull, G.F., Casjens, S. and Hendrix, R.W. (2002) Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. *J Mol Biol*, **317**, 337-359.

Muhling, M., Fuller, N.J., Millard, A., Somerfield, P.J., Marie, D., Wilson, W.H., Scanlan, D.J., Post, A.F., Joint, I. and Mann, N.H. (2005) Genetic diversity of marine Synechococcus and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environmental Microbiology*, **7**, 499-508.

Nelson, D. (2004) Phage taxonomy: we agree to disagree. *J Bacteriol*, **186**, 7029-7031.

Noble, R.T. and Fuhrman, J.A. (1998) Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquatic Microbial Ecology*, **14**, 113-118.

Nunes-Duby, S.E., Kwon, H.J., Tirumalai, R.S., Ellenberger, T. and Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res*, **26**, 391-406.

Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299-304.

Ortmann, A.C., Lawrence, J.E. and Suttle, C.A. (2002) Lysogeny and lytic viral production during a bloom of the cyanobacterium Synechococcus spp. *Microb Ecol*, **43**, 225-231.

Pajunen, M., Kiljunen, S. and Skurnik, M. (2000) Bacteriophage phiYeO3-12, specific for Yersinia enterocolitica serotype O:3, is related to coliphages T3 and T7. *J Bacteriol*, **182**, 5114-5120.

Partensky, F., Hess, W.R. and Vaulot, D. (1999) Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev*, **63**, 106-127.

Paul, J.H. and Sullivan, M.B. (2005) Marine phage genomics: what have we learned? *Curr Opin Biotechnol*, **16**, 299-307.

Paul, J.H., Sullivan, M.B., Segall, A.M. and Rohwer, F. (2002) Marine phage genomics. *Comp Biochem Physiol B Biochem Mol Biol*, **133**, 463-476.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Barbarov, S., Kriakov, J., Lawrence, J.G., Jacobs, W.R., Hendrix, R.W. and Hatfull, G.F. (2003) Origins of highly mosiac mycobacteriophage genomes. *Cell*, **113**, 171-182.

Pope, W.H., Haase-Pettingell, C. and King, J. (2004) Protein folding failure sets high-temperature limit on growth of Phage P22 in Salmonella enterica serovar typhimurium. *Appl Environ Microbiol*, **70**, 4840-4847.

Prevelige, P.E., Jr. and King, J. (1993) Assembly of bacteriophage P22: a model for ds-DNA virus assembly. *Prog Med Virol*, **40**, 206-221.

Proctor, L.M. and Fuhrman, J.A. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature*, **343**, 60-62.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., Desiere, F. and Brussow, H. (2002) The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. *J Bacteriol*, **184**, 6026-6036.

Rabinovitch, A., Fishov, I., Hadas, H., Einav, M. and Zaritsky, A. (2002) Bacteriophage T4 development in Escherichia coli is growth rate dependent. *J Theor Biol*, **216**, 1-4.

Reanney, D. and Ackerman, H. (1982) Comparative biology and evolution of bacteriophages. *Advances in Virus Research*, **27**, 205-280.

Riede, I., Drexler, K., Eschbach, M.L. and Henning, U. (1986) DNA sequence of the tail fiber genes 37, encoding the receptor recognizing part of the fiber, of bacteriophages T2 and K3. *J Mol Biol*, **191**, 255-266.

Rohwer, F. and Edwards, R. (2002) The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol*, **184**, 4529-4535.

Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F. and Azam, F. (2000) The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnology and Oceanography*, **45**, 408-418.

Salzberg, S.L., Delcher, A.L., Kasif, S. and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, **26**, 544-548.

Scholl, D., Kieleczawa, J., Kemp, P., Rush, J., Richardson, C.C., Merril, C., Adhya, S. and Molineux, I.J. (2004) Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup. *J Mol Biol*, **335**, 1151-1171.

Short, C.M. and Suttle, C.A. (2005) Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Applied and Environmental Microbiology*, **71**, 480-486.

Simpson, A.A., Leiman, P.G., Tao, Y., He, Y., Badasso, M.O., Jardine, P.J., Anderson, D.L. and Rossmann, M.G. (2001) Structure determination of the head-tail connector of bacteriophage phi29. *Acta Crystallogr D Biol Crystallogr*, **57**, 1260-1269.

Smith, D.H., Berget, P.B. and King, J. (1980) Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tail-spike protein. I. Fine-structure mapping. *Genetics*, **96**, 331-352.

Smith, D.H. and King, J. (1981) Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tail spike protein. III. Intensive polypeptide chains synthesized at 39 degrees C. *J Mol Biol*, **145**, 653-676.

Spencer, R. (1955) A marine bacteriophage. *Nature*, **175**, 690.

Spencer, R. (1960) Indigenous marine bacteriophages. *J Bacteriol*, **79**, 614.

Spencer, R. (1963) Bacterial viruses in the sea. In Oppenheimer, C.H. (ed.), *Symposium on Marine Microbiology*. Charles C. Thomas, Springfield.

Steward, G.F., Smith, D.C. and Azam, F. (1996) Abundance and production of bacteria and viruses in the Bering and Chukchi Sea. *Marine Ecology Progress Series*, **131**, 287-300.

Struthers-Schlinke, J.S., Robins, W.P., Kemp, P. and Molineux, I.J. (2000) The internal head protein Gp16 controls DNA ejection from the bacteriophage T7 virion. *J Mol Biol*, **301**, 35-45.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F. and Chisholm, S.W. (2005) Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *PLoS Biol*, **3**, e144.

Sullivan, M.B., Waterbury, J.B. and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature*, **424**, 1047-1051.

Suttle, C.A. (2000) Cyanophages and their role in the ecology of cyanobacteria. In Whitton, B.A. and Potts, M. (eds.), *The Ecology of Cyanobacteria*. Kluwer Academic Publishers, Netherlands.

Suttle, C.A. and Chan, A.M. (1993) Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. *Marine Ecology Progress Series*, **92**, 99-109.

Suttle, C.A. and Chan, A.M. (1994) Dynamics and distribution of cyanophages and their effects on marine Synechococcus spp. *Appl Environ Microbiol*, **60**, 3167-3174.

Suttle, C.A., Chan, A.M. and Cottrell, M.T. (1990) Infection of Phytoplankton by Viruses and Reduction of Primary Productivity. *Nature*, **347**, 467-469.

Symonds, N. and Coelho, A. (1978) Role of the G segment in the growth of phage Mu. *Nature*, **271**, 573-574.

Szmelcman, S. and Hofnung, M. (1975) Maltose transport in Escherichia coli K-12: involvement of the bacteriophage lambda receptor. *J Bacteriol*, **124**, 112-118.

Tetart, F., Desplats, C. and Krisch, H.M. (1998) Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J Mol Biol*, **282**, 543-556.

Tetart, F., Repoila, F., Monod, C. and Krisch, H.M. (1996) Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *J Mol Biol*, **258**, 726-731.

Ting, C.S., Rocap, G., King, J. and Chisholm, S.W. (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol*, **10**, 134-142.

Trus, B.L., Cheng, N., Newcomb, W.W., Homa, F.L., Brown, J.C. and Steven, A.C. (2004) Structure and polymorphism of the UL6 portal protein of herpes simplex virus type 1. *J Virol*, **78**, 12668-12671.

van de Putte, P., Cramer, S. and Giphart-Gassler, M. (1980) Invertible DNA determines host specificity of bacteriophage mu. *Nature*, **286**, 218-222.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B. (2000) *Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on the Taxonomy of Viruses*. Academic Press, San Diego, CA.

Veldhuis, M.J.W., Kraay, G.W., VanBleijswijk, J.D.L. and Baars, M.A. (1997) Seasonal and spatial variability in phytoplankton biomass, productivity and growth in the northwestern Indian Ocean: The southwest and northeast monsoon, 1992-1993. *Deep-Sea Research Part I-Oceanographic Research Papers*, **44**, 425-449.

Waldor, M.K. and Mekalanos, J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*, **272**, 1910-1914.

Wang, I.N., Smith, D.L. and Young, R. (2000) Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol*, **54**, 799-825.

Wang, W.F., Margolin, W. and Molineux, I.J. (1999) Increased synthesis of an Escherichia coli membrane protein suppresses F exclusion of bacteriophage T7. *J Mol Biol*, **292**, 501-512.

Waterbury, J.B. and Valois, F.W. (1993) Resistance to co-occurring phages enables marine Synechococcus communities to coexist with cyanophage abundant in seawater. *Appl Environ Microbiol*, **59**, 3393-3399.

Waterbury, J.B., Watson, S.W., Valois, F.W. and Franks, D.G. (1986) Biological and ecological characterization of the marine unicellular cyanobacterium Synechococcus. *Can Bull Fish Aquat Sci*, **214**, 71-120.

Waterbury, J.B. and Willey, J.M. (1988) Isolation and growth of marine planktonic cyanobacteria. *Methods in Enzymology*, **167**, 100-105.

Weigele, P.R., Scanlon, E. and King, J. (2003) Homotrimeric, beta-stranded viral adhesins and tail proteins. *J Bacteriol*, **185**, 4022-4030.

Weinbauer, M.G. and Rassoulzadegan, F. (2004) Are viruses driving microbial diversification and diversity? *Environ Microbiol*, **6**, 1-11.

Wen, K., Ortmann, A.C. and Suttle, C.A. (2004) Accurate estimation of viral abundance by epifluorescence microscopy. *Appl Environ Microbiol*, **70**, 3862-3867.

Wen, Z.Q., Overman, S.A., Bondre, P. and Thomas, G.J., Jr. (2001) Structure and organization of bacteriophage Pf3 probed by Raman and ultraviolet resonance Raman spectroscopy. *Biochemistry*, **40**, 449-458.

Wen, Z.Q., Overman, S.A. and Thomas, G.J., Jr. (1997) Structure and interactions of the single-stranded DNA genome of filamentous virus fd: investigation by ultraviolet resonance raman spectroscopy. *Biochemistry*, **36**, 7810-7820.

Williamson, S.J., Houchin, L.A., McDaniel, L. and Paul, J.H. (2002) Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. *Appl Environ Microbiol*, **68**, 4307-4314.

Wilson, W.H., Carr, N.G. and Mann, N.H. (1996) The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium Synechococcus sp WH7803. *Journal of Phycology*, **32**, 506-516.

Wilson, W.H., Joint, I.R., Carr, N.G. and Mann, N.H. (1993) Isolation and Molecular Characterization of 5 Marine Cyanophages Propagated on Synechococcus Sp Strain Wh7803. *Applied and Environmental Microbiology*, **59**, 3736-3743.

Wommack, K. and Colwell, R. (2000) Virioplankton: Viruses in aquatic ecosystems. *Microbiol Mol Biol Rev*, **64**, 69-114.

Wright, A. (1971) Mechanism of conversion of the Salmonella O antigen by bacteriophage epsilon34. *J Bacteriol*, **105**, 927-936.

Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L. and Treiber, G. (1970) Rapid bacteriophage sedimentation in the presence of Polyethylene glycol and its application to large-scale virus purification. *Virology*, **40**, 734-744.

APPENDIX A. PROTEIN FOLDING FAILURE SETS HIGH-TEMPERATURE LIMIT ON GROWTH OF PHAGE P22 IN SALMONELLA ENTERICA SEROVAR TYPHIMURIUM

Reprinted with permission from Applied and Environmental Microbiology

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Aug. 2004, p. 4840–4847
0099-2240/04/$08.00+0 DOI: 10.1128/AEM.70.8.4840–4847.2004

Vol. 70, No. 8

Protein Folding Failure Sets High-Temperature Limit on Growth of Phage P22 in *Salmonella enterica* Serovar Typhimurium

Welkin H. Pope,[1,2] Cameron Haase-Pettingell,[1] and Jonathan King[1]*

Massachusetts Institute of Technology, Cambridge,[1] and Woods Hole Oceanographic Institution, Woods Hole,[2] Massachusetts

Received 16 January 2004/Accepted 19 April 2004

The high-temperature limit for growth of microorganisms differs greatly depending on their species and habitat. The importance of an organism's ability to manage thermal stress is reflected in the ubiquitous distribution of the heat shock chaperones. Although many chaperones function to reduce protein folding defects, it has been difficult to identify the specific protein folding pathways that set the high-temperature limit of growth for a given microorganism. We have investigated this for a simple system, phage P22 infection of *Salmonella enterica* serovar Typhimurium. Production of infectious particles exhibited a broad maximum of 150 phage per cell when host cells were grown at between 30 and 39°C in minimal medium. Production of infectious phage declined sharply in the range of 40 to 41°C, and at 42°C, production had fallen to less than 1% of the maximum rate. The host cells maintained optimal division rates at these temperatures. The decrease in phage infectivity was steeper than the loss of physical particles, suggesting that noninfectious particles were formed at higher temperatures. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed a decrease in the tailspike adhesins assembled on phage particles purified from cultures incubated at higher temperatures. The infectivity of these particles was restored by in vitro incubation with soluble tailspike trimers. Examination of tailspike folding and assembly in lysates of phage-infected cells confirmed that the fraction of polypeptide chains able to reach the native state in vivo decreased with increasing temperature, indicating a thermal folding defect rather than a particle assembly defect. Thus, we believe that the folding pathway of the tailspike adhesin sets the high-temperature limit for P22 formation in *Salmonella* serovar Typhimurium.

Microorganisms have been detected in almost every climate on Earth, from the extreme heat and pressure of hydrothermal vents to the exposed rocks of frozen Antarctica. However, optimal growth is generally limited to a defined temperature regime characteristic of the habitat. Examination of the growth rates of microorganisms with respect to temperature reveals a clear pattern. The majority of organisms exhibit growth rates that slowly increase with rising temperature until a maximum growth rate is achieved. They then exhibit a sharp decline once the optimal temperature has been exceeded (24, 25). Many factors that contribute to the decline of the growth rate at higher-than-optimal temperatures have been identified, including lipid membrane stability (4, 33), rates of DNA synthesis and repair (10, 32), rate of protein synthesis (11), and protein misfolding (18, 28). However, it has generally been difficult to determine which factor sets the high-temperature growth limit. Environmental temperatures around the globe are increasing far more rapidly than documented rates of evolutionary change (see http://lwf.ncdc.noaa.gov/img/climate/research/anomalies/triad_pg.gif). Thus, it is becoming increasingly important to understand the processes that contribute to thermal limitation of growth.

A significant part of the thermal stress response is the induction of heat shock chaperones (12). These specialized proteins are capable of binding some species of misfolded and aggregated polypeptide chains, in some cases shifting folding from an unproductive aggregation pathway to the productive native pathway. However, as the temperature increases, chaperones become less capable of buffering the cell against the accumulation of misfolded proteins (12, 13, 24). Phage proteins, like cellular proteins, may or may not be recognized by the host's heat shock proteins. Temperature-sensitive-folding (*tsf*) mutants of the coat protein of phage P22 are rescued by overexpression of the GroEL and GroES chaperones (16, 17). However, P22 phage with similar *tsf* mutations in the tailspike do not appear to be aided by the GroEL/ES chaperone (8, 17). Some phage possess their own chaperones, which after infection are synthesized and provide nascent peptide chains with some measure of protection against aggregation induced by high intracellular protein concentrations and excessive heat (40). In *Escherichia coli*, phages T4 and λ both possess their own chaperones (23, 43). The ability of heat shock proteins to alleviate the aggregation of some polypeptide chains and not others increases the difficulty of determining which substrate pathway sets the growth limit at high temperatures.

Elucidation of the limiting process during high-temperature growth with respect to the folding of proteins cannot be attained by examination of the properties of mature proteins, since the properties of folding peptide chains are quite different from those of native folds. For example, the highly thermostable tailspike adhesin of phage P22 exhibits thermolabile folding intermediates (labeled "partially folded monomer" in Fig. 1A) (29, 39).

We have examined the issue of growth limitation at high temperature for *Salmonella enterica* serovar Typhimurium infected by phage P22. This model system has several characteristics that make it particularly appropriate for the study of this problem. The assembly of structural proteins of P22 has been

* Corresponding author. Mailing address: 77 Massachusetts Ave. 68-330, Cambridge, MA 02139. Phone: (617) 253-4700. Fax: (617) 252-1843. E-mail: jaking@mit.edu.

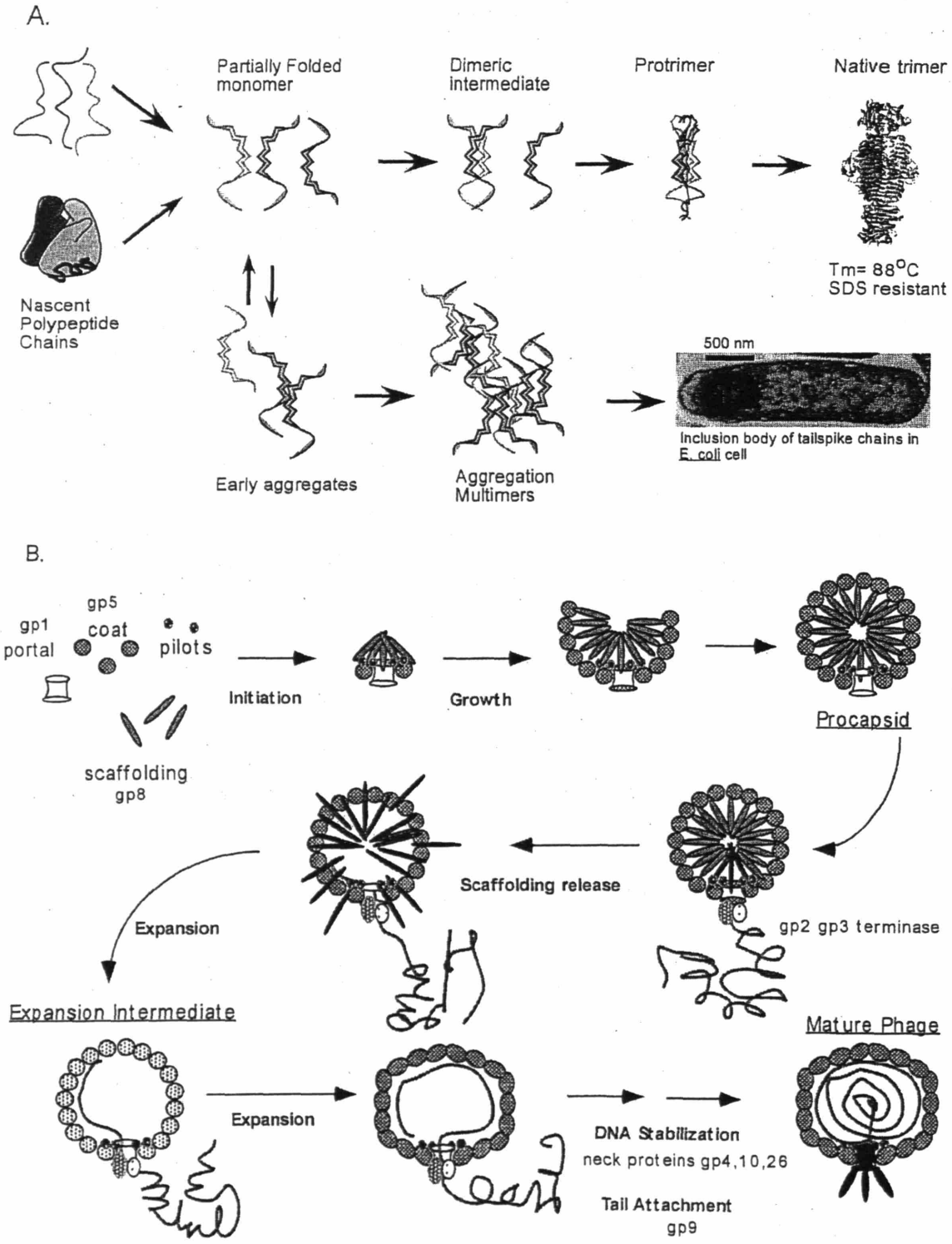


FIG. 1. Folding and assembly of the P22 tailspike and assembly of P22 virions. (A) Unfolded or nascent polypeptide chains proceed to some partially folded monomeric state. On the productive folding pathway, two such chains associate to form a partially folded dimer. The addition of a third chain creates a protrimer intermediate whose chains are associated but not fully folded. The protrimer undergoes chain rearrangement resulting in the native structure. At higher temperatures, the partially folded monomeric species is perturbed, preventing the formation of the productive dimer and allowing the formation of multimeric aggregates. These large aggregates form inclusion bodies within the cell. All species of tailspike, with the exception of the native trimer, are sensitive to SDS. Tm, thermal denaturation midpoint temperature. (B) After synthesis, the coat, scaffolding, portal, and pilot subunits form a nucleus of the procapsid shell. Coat and scaffolding subunits rapidly add on to the nucleus, forming the completed procapsid. Viral DNA, with the help of packaging proteins, is driven into the procapsid, and the scaffolding protein is ejected through the pores of the procapsid shell. The procapsid undergoes an irreversible expansion to the mature capsid, while DNA is pumped into the procapsid. After the encapsulation of the phage DNA, the neck and tailspike proteins attach to form the infectious mature virion.

well characterized (Fig. 1B). During the first stage, coat and scaffolding subunits and auxiliary proteins bind together to form a spherical procapsid. Phage DNA then enters the head through the portal complex, while the scaffolding protein exits the procapsid via pores in the shell. The DNA-filled procapsid undergoes an irreversible conformational change, expanding to the familiar icosahedral shape.

After attachment of the neck proteins, the tailspike trimer binds irreversibly to the capsid, conferring infectivity (26). The lateral surface of the tailspike recognizes and binds to the *Salmonella* lipopolysaccharide (LPS) projecting from the host cell surface (37). The intermediates in tailspike folding and assembly have been characterized (Fig. 1A). The tailspike adhesin, which has endorhamnosidase activity, binds the LPS in a cleft between two loop domains and cleaves it between rhamnose and galactose moieties (2, 27, 37).

Native tailspike homotrimers exhibit resistance to denaturation by sodium dodecyl sulfate (SDS) detergent, proteases, and temperatures up to 88°C, denaturing only in the presence of SDS at high temperatures. However, the tailspike folding process has been shown to be highly heat labile (15, 22, 36), resulting in the shifting of chains off the productive pathway at high temperatures to accumulate as inclusion bodies (Fig. 1A). These two properties, high thermostability of the native structure and thermolability of folding intermediates, make it possible to monitor the intracellular folding and assembly of the tailspike.

Early observations of P22 particles produced at high temperatures revealed a deficiency in the quantity of tailspike adhesins attached to phage heads (26). The in vivo and in vitro folding process of the tailspike's parallel β-helix motif has been demonstrated to be highly thermolabile, with significant loss of native tailspike protein beginning at temperatures as low as 35°C (22). This suggested that the loss of infectious particles at the high end of P22's physiological temperature range was due to the misfolding of the tailspike adhesin and its subsequent inability to assemble onto fully formed P22 heads.

MATERIALS AND METHODS

Salmonella serovar Typhimurium strain DB7155 [*sup*$^+$ *hisC525*(Am) *leuA414*(Am) SupE] was grown in M9 minimal medium supplemented with glucose (0.4%), yeast extract (0.01%), $MgSO_4$ (1 mM), $FeCl_3$ (1 μM), and $CaCl_2$ (1 μM); strain DB7136 [*hisC525*(Am) *leuA414*(Am)] was grown in the same medium additionally supplemented with histidine (0.0015%) and leucine (0.0015%). P22 strains 13_{H101}(Am) $C_1$7 (hereinafter referred to as P22) and 2_{H200}(Am)/13_{H101}(Am) $C_1$7 (hereinafter referred to as P22 2$^-$) were used. The amber mutation in gene 13 prolongs lysis of host cells, while the $C_1$7 mutation prevents lysogeny. The gene 2 amber mutation prevents packaging of DNA into newly formed capsids and upregulates tailspike synthesis (1).

Growth curves. An overnight (O/N) culture of strain DB7155 was inoculated (1:50) into minimal medium at 30, 37, 39, 40, 41, 42, or 43°C and was continuously aerated during incubation. Samples were withdrawn over a 6-h period, serially diluted with dilution fluid (tryptone [1%], NaCl [0.7%], and $MgSO_4$ [2 mM]), and plated on Luria-Bertani (LB) agar (1.2%) medium at 30°C. After O/N incubation, colonies were counted manually.

Burst size. An O/N culture of strain DB7155 was inoculated (1:50) ratio into medium at 37°C and were grown with constant aeration to a concentration of 2 × 10^8 cells/ml. Phage was added at a multiplicity of infection (MOI) of 10. After a 10-min adsorption period, infected cells were shifted to 30, 37, 39, 40, 41, 42, or 43°C. After cell lysis, which was evidenced by a significant loss of turbidity, phage were serially diluted in dilution fluid, mixed with soft LB agar and a few drops of plating bacteria, and plated on LB agar (1.2%) at 30°C. After O/N incubation, plaques were counted manually. Each titer was confirmed by three replicate experiments.

High-temperature P22 production and purification. An O/N culture of strain DB7136 (*sup*$^-$) was inoculated (1:10) into medium at 37°C and grown with constant aeration to a concentration of 2 × 10^8 cells/ml. Phage were added to the culture at an MOI of 10. After a 10-min adsorption period, the infected cells were moved to 30, 37, 39, 40, 41, 42, or 43°C and incubated for 2 or 3 h, depending on the length of time required to lyse strain DB7155 at the same temperature. Cells were harvested by centrifugation at 5,000 × *g* for 10 min and then resuspended in 1 ml of 50 mM Tris (pH 7.6)–100 mM $MgCl_2$. Cells were lysed with $CHCl_3$, and DNase I and phenylmethylsulfonyl fluoride were added to the lysates. Phage were pelleted by a low-speed spin at 8,000 × *g* overnight and then resuspended at 4°C in 50 mM Tris-HCl (pH 7.6)–100 mM $MgCl_2$. A 1-ml volume of each lysate was layered over a 16-ml CsCl step gradient (ρ = 1.3, 1.4, 1.5, and 1.7) and ultracentrifuged at 4°C and 28,000 rpm in an SW 28.1 rotor for 4 h. A milky-white phage-containing band was observed between ρ = 1.4 and ρ = 1.5. A 1-ml volume of purified phage preparation was harvested and, using a Pierce dialysis cassette, dialyzed against two changes of 100 mM Tris (pH 7.6)–50 mM $MgCl_2$.

Tailing assay. A 50-μl volume of purified P22 particle preparation (obtained as outlined above) was mixed with 50 μl (250 ng) of purified tailspike solution (purified as outlined by Haase-Pettingell et al. [20]) or 50 μl of dilution fluid and incubated at room temperature for 60 min. A 900-μl volume of dilution fluid was added. Samples were then serially diluted and plated for determination of PFU at 30°C. Each titer was confirmed by three replicate experiments.

SDS-polyacrylamide gel electrophoresis (PAGE) quantification. Purified P22 samples were boiled in SDS buffer and loaded onto a 10% acrylamide gel containing SDS. The gels were electrophoresed at a constant 20 mA until the dye front ran off the end of the gel. Gels were silver stained, and tailspike and portal bands were quantified using ImageQuant software (Molecular Dynamics).

DNA-packaging-deficient P22 lysate production and analysis. A 100-ml volume of minimal medium supplemented with histidine (0.0015%) and leucine (0.0015%) was inoculated with 5 ml of a strain DB7136 culture. Once the culture density reached ~2 × 10^8 cells/ml, P22 2$^-$ phage were added at an MOI of 10. Phage were allowed to adsorb for 10 min, and then 10-ml volumes of the culture were shifted to each of the following temperatures: 30, 37, 39, 40, 41, 42, and 43°C. Cultures were incubated for 3 h, and then cells were harvested by centrifugation in an SS-34 rotor for 10 min at 7,000 rpm and 4°C. Cells were resuspended in 600 μl of lysis buffer B (50 mM Tris-HCl [pH 8], 25 mM NaCl, 2 mM EDTA, 0.1% Triton-X) and frozen. Once thawed, samples were sonicated with a Microson XL benchtop sonicator (Misonix) for 30 s each to disrupt DNA and any cells remaining after the freeze-thaw step. SDS sample buffer was added, and samples were electrophoresed at a constant 20 mA through a 10% acrylamide gel containing SDS until the dye front ran off the gel. The gel was stained with Coomassie blue stain, and bands were quantified using the software ImageQuant version 1.2 (Molecular Dynamics).

RESULTS

Effects of temperature on growth of *Salmonella* serovar Typhimurium and production of P22. Yields of infectious particles of P22 [13_{H101}(Am) $C_1$7] phage have been reported to decrease with increasing temperature above a plateau range of 30 to 39°C (26). This might reflect breakdown of heat-sensitive biosynthetic processes within the host cells. Comparison of the growth rates of the *Salmonella* serovar Typhimurium host cells and phage P22 particles produced was performed at 30, 37, 39, 40, 41, 42, and 43°C.

The upper thermal limit of vegetative growth for *Salmonella* serovar Typhimurium in this experiment was 42°C. The temperature regime of P22 production did not parallel the range favored by its host but, rather, sharply declined at 40°C, 3°C lower than the temperature at which *Salmonella* serovar Typhimurium growth became undetectable. As seen in Fig. 2, the optimal growth rate of *Salmonella* serovar Typhimurium was at 37°C. At this temperature, the burst size of P22 was approximately 150 phage/cell. At 40 and 41°C, the host's growth rate was approximately 95% of the maximum growth rate. In contrast, at 40°C, P22 PFU decreased 10-fold, while at 41°C a 100-fold decrease in PFU was observed.

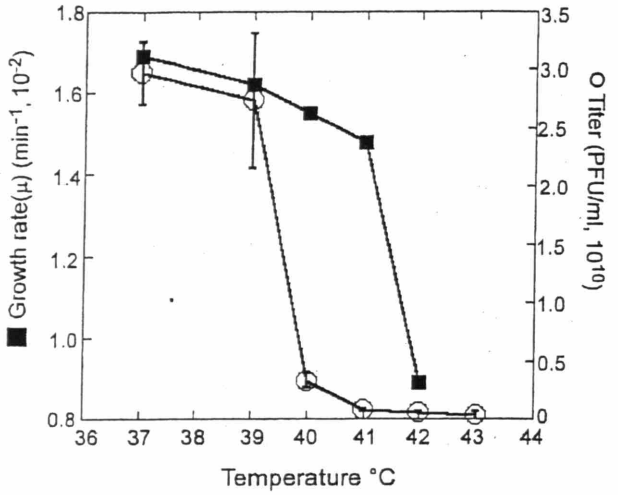


FIG. 2. *Salmonella* serovar Typhimurium growth rate and P22 titer versus temperature. The growth rate of *Salmonella* serovar Typhimurium at selected temperatures (closed squares) was monitored by enumeration of CFU. Growth rate (μ) was determined by applying the growth equation $N(t) = N_0e^{\mu t}$, where N_0 is the number of cells at time $(t) = 0$, N is the number of cells at time $(t) = t$, and solving for μ. R^2 values were all above 0.93, indicating a good fit for the regression. The titer of P22 infectious particles (open circles) was determined by plaque assay. Error bars reflect the standard error.

This result indicated that the observed deficit in phage production was not due to the thermally induced failure of some host process but rather resulted from the loss of an essential phage process, such as transcription, replication, protein production, DNA packaging, or virion assembly. As the initial steps of infection occurred at 37°C, with a subsequent shift to a target temperature, the loss of infectious particles at high temperatures was not due to a defect in the infection-absorption process.

Examination of phage formed at high temperatures. To identify the thermolabile process causing the loss of infectious particles at higher temperatures, a closer examination of particles formed at all temperatures was undertaken. Exponentially growing cultures at 37°C were infected with phage at an MOI of 10. After phage absorption, the cultures were shifted to 30, 37, 39, 40, 41, 42, or 43°C. Infected cells were collected by low-speed centrifugation and lysed, and phage particles were then purified by use of CsCl step gradients.

Figure 3a shows the results of SDS gel electrophoresis of purified particles. Coat protein, seen at a position corresponding to a molecular mass of 45 kDa, dominated the pattern. Also resolved were the products of gene 1 (portal protein), gene 9 (tailspike adhesin), and genes 16 and 20 (DNA injection proteins) (35). The ratio of these structural proteins was relatively constant at lower temperatures. However, samples produced from lysates of infected cells incubated at high temperatures exhibited decreased intensity of the tailspike band. Tailspike binding, the last step in phage assembly, yields from zero to six trimers per head (6, 26). To measure this process with greater sensitivity, the amount of tailspike per DNA-containing phage head was assessed.

The ring of P22 portal protein is required for DNA packaging and is present in the constant proportion of 12 protein subunits per DNA-containing phage head (3). Thus, the quantity of portal protein is directly proportional to the number of DNA-containing phage heads in the sample. The ratio of the amount of tailspike chains to the amount of portal protein chains, seen at 96 kDa, in each sample was determined. This ratio of tailspike chains to portal chains decreased as the temperature at which infection occurred increased, indicating that fewer tailspike adhesins were attached to the DNA-containing

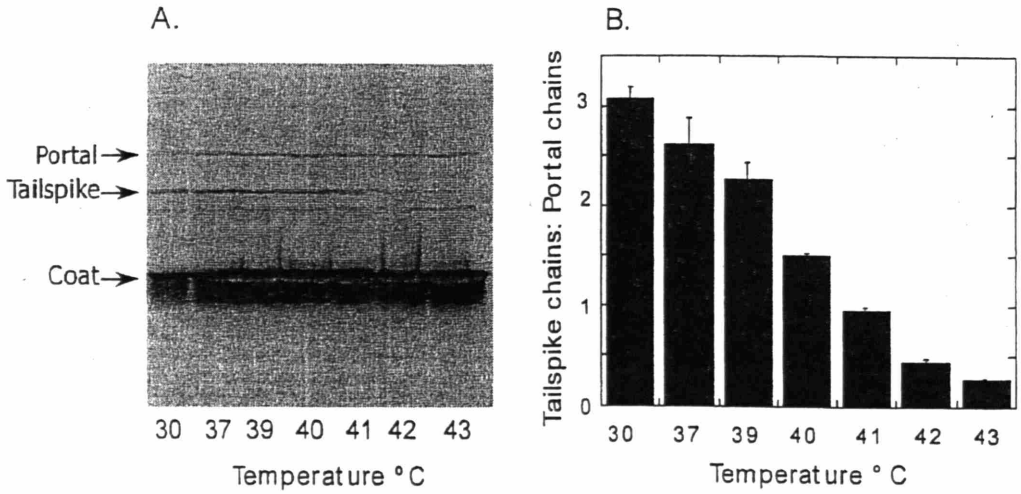


FIG. 3. SDS-PAGE of purified phage samples. (A) Phage particles, harvested by CsCl gradient purification from infected cells incubated at selected temperatures, were electrophoresed though an SDS–10% acrylamide gel. Protein bands were visualized with silver stain. Controls included previously purified P22 virions and P22 procapsids. (B) Analysis of tailspike chains versus portal chains in purified P22 samples. Relative quantities of tailspike chains and portal chains were determined by analyzing a silver-stained SDS-acrylamide gel with the ImageQuant software (Molecular Dynamics). Intensity of pixels was indicative of quantity of protein loaded in each lane. The ratio of the determined intensity of tailspike chains to portal chains in each sample was then plotted versus temperature. Error bars reflect the standard error as determined from three quantification replicates.

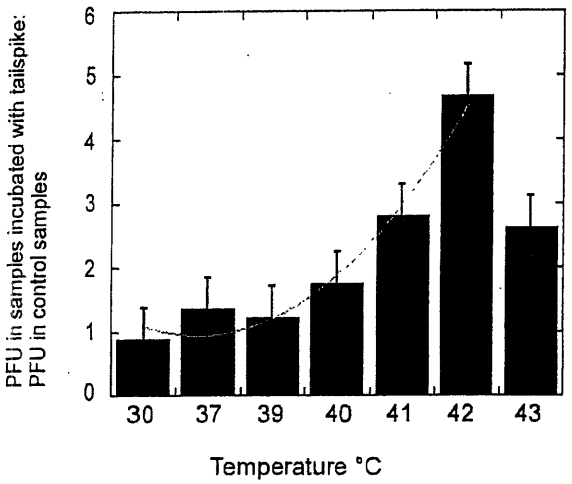


FIG. 4. Addition of purified exogenous native tailspike to samples. Native tailspike was added to purified P22 particles; after incubation for 60 min, the samples were plated to determine the phage titer. The ratio of PFU in P22 samples incubated with exogenous tailspike to PFU in P22 samples incubated with dilution fluid was determined and plotted versus the temperature at which the purified P22 particles were assembled. A trend line was added for clarity. Error bars reflect the standard error.

phage heads at high temperatures (Fig. 3B). This loss of bound tailspike was not due to thermal denaturation of mature tailspikes, as the thermal denaturation midpoint temperature for these structures is 88°C and they are resistant to intracellular degradation (39).

The tailspike deficit could reflect either a defect in their production at high temperatures or a defect in the portal vertex site to which they bind. This was resolved by determining whether the phage heads found at high temperatures could be converted to infectious particles by incubation with exogenous native tailspikes. In this tailing assay (Fig. 4), P22 infectious particles were generated upon mixing of tailspikeless DNA-containing phage heads with exogenous native tailspike (6, 26). The tailing assay indicated that addition of native tailspike to samples produced at high temperatures increased infectivity up to fivefold. Thus, the heads were competent to bind tailspike. The increase in number of infectious particles was an indication that many DNA-containing phage heads produced at higher temperatures were lacking sufficient tailspikes for infectivity.

In vivo, the lack of tailspikes at higher temperatures may be due to (i) misfolding and aggregation of the adhesins, (ii) slower synthesis of these adhesins, or (iii) a failure of these adhesins to successfully attach to the phage heads. To distinguish between translational and folding defects, we quantified the amount of native tailspike and partially folded tailspike intermediates produced within the cell at each temperature. This is possible through SDS-PAGE analysis, as native tailspike is not denatured by SDS unless it is boiled (14, 39). Thus, in unboiled samples on SDS gels, native tailspike electrophoreses with a mobility different from that of partially folded chains, which form conventional SDS-polypeptide chain complexes (14).

The lysates described above accumulated very little free native tailspike, presumably due to the rapid assembly of native trimer onto mature phage heads formed within infected cells. This suggested that the high-temperature-related problem was not failure of the assembly of native tailspike onto phage heads. Although native tailspike is thermostable, the presence of a thermolabile intermediate in the folding pathway has been well documented (21, 22). To examine chain folding and assembly in vivo, we used a P22 strain [2_{H200}(Am)/13_{H101}(Am) C_17] containing an amber mutation in the DNA-packaging protein encoded by gene 2 (1). This mutation prevents the packaging of phage DNA into newly formed procapsids, thereby preventing addition of the neck protein and prolonging lysis. As a result, all newly synthesized tailspike chains remain in the cytoplasm (30). In addition, unpackaged DNA also is involved in upregulating the total amount of tailspike synthesized.

Exponentially growing cells were infected with P22 2^- at 37°C. Samples of the culture were incubated at 30, 37, 39, 40, 41, 42, and 43°C, harvested by low-speed centrifugation, and lysed. Using SDS-PAGE analysis and Coomassie blue staining, we were able to quantify the amount of native tailspike, folding intermediates, and coat protein within each lysate (Fig. 5). The native trimers are easily distinguishable from partially folded species in unboiled samples electrophoresed through SDS-acrylamide gels, as only the native trimer is resistant to SDS denaturation and, consequently, migrates toward the top of the gel (14). All other tailspike chains—misfolded, aggregated, and folding intermediates—denature in the presence of SDS and migrate to a position corresponding to a molecular mass of 72 kDa (14, 22).

For each sample, the ratio of native tailspike to SDS-sensitive tailspike species and the ratio of total tailspike (native plus SDS sensitive) to coat protein were determined (Fig. 5B). The total-tailspike sample values were normalized to the amount of coat protein present. Similar quantities of tailspike were present in all lysates tested except for that produced at the highest temperature, 43°C. At that temperature, there was a substantial decrease in the amount of coat protein synthesized. The high ratio of total tailspike to coat protein may have been the result of this large decrease in coat protein. This result implies that the high-temperature (up to 42°C) defect in P22 formation is not the result of a decrease in the rate of tailspike polypeptide synthesis.

To rule out the possibility of an assembly defect at high temperatures, we examined the ratio of native tailspike to partially folded or assembled SDS-sensitive tailspike species in each lysate. An observed increase in this ratio in cells incubated at higher temperatures would indicate that the protein was capable of folding at higher temperatures but was unable to attach to phage heads, while a decrease in this ratio would indicate a thermally induced folding defect in the adhesin.

The results of an SDS-PAGE analysis of DNA-packaging-defective lysates are presented in Fig. 5B. A decrease in the ratio of native tailspike to SDS-sensitive tailspike species is evident. This indicates that a thermal folding defect is the cause of the loss of infectious particles at elevated temperatures. As the quantity of tailspike chains synthesized remained relatively constant at temperatures up to 42°C, a synthesis defect could also be ruled out. Therefore, it is likely that the thermolability of a tailspike folding intermediate is responsible for the decrease in infectious particles at high temperatures.

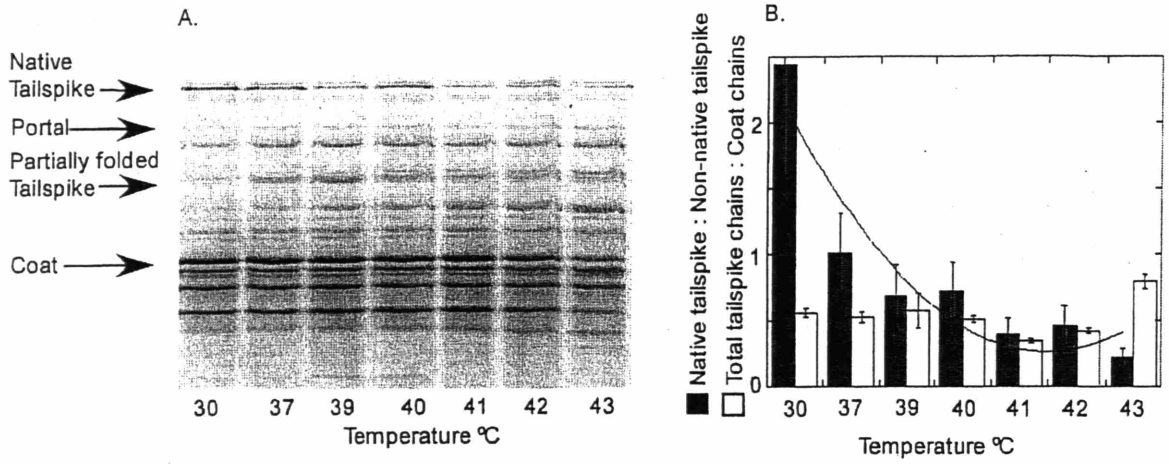


FIG. 5. SDS-PAGE analysis of P22 2⁻ DNA-packaging-defective lysates. (A) Cells were incubated for 2 to 3 h at selected temperatures, pelleted, and lysed by sonication. Lysates were DNase I treated and loaded, without boiling, on a 10% acrylamide gel containing SDS. Protein bands were visualized using Coomassie blue stain. (B) Analysis of native and partially folded tailspike chains (gray bars). Relative quantities of native tailspike, partially folded tailspike, and coat protein (white bars) were determined by analyzing a Coomassie blue-stained SDS-acrylamide gel with ImageQuant software (Molecular Dynamics). The ratio of native tailspike chains to partially folded tailspike chains versus temperature is in gray, and the ratio of total tailspike chains to coat chains versus temperature is in white. A trend line was added for clarity. Error bars reflect the standard error, as determined from three gel quantification replicates.

DISCUSSION

Identification of the thermolabile steps in those microbial processes which limit growth at high temperatures has been difficult. For phage P22 propagating in *Salmonella* serovar Typhimurium, the yield of infectious P22 particles produced declined at a lower temperature than did the reproductive capacity of the uninfected host cells. A significant loss of infectious P22 particles was observed for lysates incubated at temperatures of 40°C and above, while *Salmonella* serovar Typhimurium was capable of growth at temperatures up to 42°C. This suggested that the formation of some phage component was limiting the yield at higher temperatures. By infecting cells at the permissive temperature of 37°C and shifting them to higher target temperatures, we were able to control for potential thermal defects in adsorption of the phage.

Close examination of purified particles produced at high temperatures revealed a decrease in the amount of tailspike adhesin attached to DNA-containing phage heads. The heads that lacked tailspike were capable of binding exogenous active tailspike, indicating that the defect was not in the capsid structure.

SDS-PAGE analysis of crude lysates of cells infected with P22 terminase mutants showed that similar quantities of tailspike chains were produced at all temperatures, indicating that the loss of tailspike at high temperatures was not the result of a thermal defect in transcription or translation. The ratio of native tailspike to total tailspike chains decreased with increasing temperature, an indication that the lack of tailspike adhesins was due to misfolding of these polypeptide chains rather than to a defect in their synthesis or in the ability of the native trimer to bind the phage capsid.

The tailspike folding pathway, which has been well characterized experimentally, proceeds through several intermediates before achieving the native structure (7, 29). This pathway proceeds through a thermolabile folding intermediate, which aggregates into an inclusion body if folding occurs at high temperatures (21, 22) and prevents the production of native tailspike (Fig. 1A).

The native tailspike is a homotrimer. Each of the three chains has a parallel β-helix region that terminates in a triple β-helix formed by the wrapping of the three peptide chains around each other. Using the extended lateral surface of the parallel β-helix, native tailspike binds the host's LPS (37). Kreisberg et al. showed that the triple-β-helix motif gives the native trimeric protein extra stability (31). The most likely thermolabile motif within the tailspike adhesin is the parallel β-helix formed by each chain within the structure (21, 29, 34, 39).

It is not clear why the folding of the tailspike adhesin is the most thermally sensitive process in the production of mature P22 particles. The tailspike structure contains two relatively rare protein motifs: the parallel β-helix (38) and the triple-stranded β-helix (42). It is possible that the folding intermediates for these domains are intrinsically thermolabile. Mutational studies of the tailspike have revealed the existence of a substantial number of single-amino-acid, temperature-sensitive-folding mutations within the parallel β-helix (21).

An alternative explanation for the thermolability inherent in the folding pathway is the lack of a helper chaperone. Brunschier et al. (8) showed that tailspike intermediates, unlike the P22 coat protein, were not rescued by *E. coli*'s GroEL/ES chaperone system (8). Gordon et al. (17) examined several temperature-sensitive-folding mutants of both the tailspike and coat proteins in conjunction with overexpression of GroEL/ES. The coat protein mutants were rescued from thermally induced aggregation by chaperone overexpression, while the tailspike mutants could not be rescued in this manner (17). The 666-amino-acid tailspike chain may be too large for the lumen of the GroEL/ES chaperone. The decline in native tailspike yield is presumably a consequence of both the intrinsic ther-

molability of the partially folded intermediates and the inability of GroEL to chaperone the species. There is likely to be a strong selection for host recognition adhesin function within the host-phage ecosystem. We suspect that the stability and efficacy of the tailspike's native state balance the folding and assembly disadvantages of this motif at higher temperatures.

In the above discussion, we assume that the folding problem is intrinsic to the biochemistry of the parallel β-helix motif. Alternatively, this loss of tailspike yield could be an evolved response of P22 that allows for a reduction of the phage population when the host is experiencing stress. For example, under nutrient-poor conditions, the T4 coliphage Wac protein will bind the long tail fibers of the phage into an upright position, thereby preventing infection of the host (9). It is possible that the sensitivity of the folding of the parallel β-helix to elevated temperatures is a similarly evolved response preserved to prevent phage propagation in times of host stress.

It is likely that prior to the evolution of warm-blooded organisms, the early ancestors of *Salmonella* spp. were adapted to a soil or an aquatic environment. In the aquatic regime, temperatures rarely (if ever) reach 40°C, while a soil environment may experience large temperature fluctuations over brief periods of time. Both scenarios would select for a mature protein motif that is highly stable and resistant to external stresses such as proteases and temperature. Modern *Salmonella* spp. are capable of propagating outside the body and are exposed to diverse temperature regimes. With the evolution of warm-blooded organisms and subsequent adaptation of *Salmonella* spp. to the guts of these animals, the parallel β-helix may have been retained through selection for native properties.

In the marine environment, primary producers such as cyanobacteria are subjected to a very limited range of environmental temperatures. Phages whose morphologies resemble those of enteric phages have been isolated from marine synechococcal strains (41). We are presently examining whether the formation of phage structural proteins is a rate-limiting step in the production of cyanophages of the marine genus *Synechococcus*.

ACKNOWLEDGMENTS

We thank Peter Weigele for helpful discussions.

This work was funded by NIH grant GM17980 and NSF grant EIA0225609 to J.K.

REFERENCES

1. **Adams, M. B., H. R. Brown, and S. Casjens.** 1985. Bacteriophage P22 tail protein gene expression. J. Virol. **53**:180–184.
2. **Baxa, U., S. Steinbacher, S. Miller, A. Weintraub, R. Huber, and R. Seckler.** 1996. Interactions of phage P22 tails with their cellular receptor, *Salmonella* O-antigen polysaccharide. Biophys. J. **71**:2040–2048.
3. **Bazinet, C., J. Benbasat, J. King, J. M. Carazo, and J. L. Carrascosa.** 1988. Purification and organization of the gene 1 portal protein required for phage P22 DNA packaging. Biochemistry **27**:1849–1856.
4. **Beney, L., and P. Gervais.** 2001. Influence of the fluidity of the membrane on the response of microorganisms to environmental stresses. Appl. Microbiol. Biotechnol. **57**:34–42.
5. **Benton, C. B., J. King, and P. L. Clark.** 2002. Characterization of the protrimer intermediate in the folding pathway of the interdigitated beta-helix tailspike protein. Biochemistry **41**:5093–5103.
6. **Berget, P. B., and A. R. Poteete.** 1980. Structure and functions of the bacteriophage P22 tail protein. J. Virol. **34**:234–243.
7. **Betts, S., and J. King.** 1999. There's a right way and a wrong way: in vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. Struct. Fold. Des. **7**:R131–R139.
8. **Brunschier, R., M. Danner, and R. Seckler.** 1993. Interactions of phage P22 tailspike protein with GroE molecular chaperones during refolding in vitro. J. Biol. Chem. **268**:2767–2772.
9. **Conley, M. P., and W. B. Wood.** 1975. Bacteriophage T4 whiskers: a rudimentary environment-sensing device. Proc. Natl. Acad. Sci. USA **72**:3701–3705.
10. **Corry, P. M., S. Robinson, and S. Getz.** 1977. Hyperthermic effects on DNA repair mechanisms. Radiology **123**:475–482.
11. **Dean, R. G., and E. J. McGroarty.** 1979. Protein and ribonucleic acid syntheses in heat-damaged and heat-killed *Escherichia coli*. J. Bacteriol. **138**:492–498.
12. **Ellis, R. J., and S. M. van der Vies.** 1991. Molecular chaperones. Annu. Rev. Biochem. **60**:321–347.
13. **Fink, A. L.** 1999. Chaperone-mediated protein folding. Physiol. Rev. **79**:425–449.
14. **Goldenberg, D. P., P. B. Berget, and J. King.** 1982. Maturatión of the tailspike endorhamnosidase of *Salmonella* phage P22. J. Biol. Chem. **257**:7864–7871.
15. **Goldenberg, D. P., and J. King.** 1981. Temperature-sensitive mutants blocked in the folding or subunit of the bacteriophage P22 tailspike protein. II. Active mutant proteins matured at 30°C. J. Mol. Biol. **145**:633–651.
16. **Gordon, C. L., and J. King.** 1993. Temperature-sensitive mutations in the phage P22 coat protein which interfere with polypeptide chain folding. J. Biol. Chem. **268**:9358–9368.
17. **Gordon, C. L., S. K. Sather, S. Casjens, and J. King.** 1994. Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. J. Biol. Chem. **269**:27941–27951.
18. **Gragerov, A. I., E. S. Martin, M. A. Krupenko, M. V. Kashlev, and V. G. Nikiforov.** 1991. Protein aggregation and inclusion body formation in *Escherichia coli rpoH* mutant defective in heat shock protein induction. FEBS Lett. **291**:222–224.
19. **Greene, B., and J. King.** 1996. Scaffolding mutants identifying domains required for P22 procapsid assembly and maturation. Virology **225**:82–96.
20. **Haase-Pettingell, C., S. Betts, S. W. Raso, L. Stuart, A. Robinson, and J. King.** 2001. Role for cysteine residues in the in vivo folding and assembly of the phage P22 tailspike. Protein Sci. **10**:397–410.
21. **Haase-Pettingell, C., and J. King.** 1997. Prevalence of temperature sensitive folding mutations in the parallel beta coil domain of the phage P22 tailspike endorhamnosidase. J. Mol. Biol. **267**:88–102.
22. **Haase-Pettingell, C. A., and J. King.** 1988. Formation of aggregates from a thermolabile in vivo folding intermediate in P22 tailspike maturation. A model for inclusion body formation. J. Biol. Chem. **263**:4977–4983.
23. **Hashemolhosseini, S., Y.-D. Stierhof, I. Hindennach, and U. Henning.** 1996. Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and λ. J. Bacteriol. **178**:6258–6265.
24. **Herendeen, S. L., R. A. VanBogelen, and F. C. Neidhardt.** 1979. Levels of major proteins of *Escherichia coli* during growth at different temperatures. J. Bacteriol. **139**:185–194.
25. **Ingraham, J.** 1996. Effect of temperature, pH, water activity, and pressure on growth, p. 1543–1554. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society of Microbiology, Washington, D.C.
26. **Israel, J., T. Anderson, and M. E. Levin.** 1967. *In vitro* morphogenesis of phage P22 from heads and baseplate parts. Proc. Natl. Acad. Sci. USA **57**:284–291.
27. **Iwashita, S., and S. Kanegasaki.** 1973. Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. Biochem. Biophys. Res. Commun. **55**:403–409.
28. **Jaenicke, R.** 1991. Protein folding: local structures, domains, subunits, and assemblies. Biochemistry **30**:3147–3161.
29. **King, J., C. Haase-Pettingell, A. S. Robinson, M. Speed, and A. Mitraki.** 1996. Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates. FASEB J. **10**:57–66.
30. **King, J., C. Hall, and S. Casjens.** 1978. Control of the synthesis of phage P22 scaffolding protein is coupled to capsid assembly. Cell **15**:551–560.
31. **Kreisberg, J. F., S. D. Betts, C. Haase-Pettingell, and J. King.** 2002. The interdigitated beta-helix domain of the P22 tailspike protein acts as a molecular clamp in trimer stabilization. Protein Sci. **11**:820–830.
32. **Landry, J., and P. Chretien.** 1983. Relationship between hyperthermia-induced heat-shock proteins and thermotolerance in Morris hepatoma cells. Can J. Biochem. Cell Biol. **61**:428–437.
33. **Marr, A. G., J. L. Ingraham, and C. L. Squires.** 1964. Effect of the temperature of growth of *Escherichia coli* on the formation of β-galactosidase. J. Bacteriol. **87**:356–362.
34. **Miller, S., B. Schuler, and R. Seckler.** 1998. A reversibly unfolding fragment of P22 tailspike protein with native structure: the isolated beta-helix domain. Biochemistry **37**:9160–9168.
35. **Prevelige, P. E., Jr., and J. King.** 1993. Assembly of bacteriophage P22: a model for ds-DNA virus assembly. Prog. Med. Virol. **40**:206–221.
36. **Smith, D. H., P. B. Berget, and J. King.** 1980. Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tailspike protein. I. Fine-structure mapping. Genetics **96**:331–352.
37. **Steinbacher, S., U. Baxa, S. Miller, A. Weintraub, R. Seckler, and R. Huber.**

1996. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc. Natl. Acad. Sci. USA 93:10584–10588.

38. **Steinbacher, S., R. Seckler, S. Miller, B. Steipe, R. Huber, and P. Reinemer.** 1994. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. Science 265:383–386.
39. **Sturtevant, J. M., M. H. Yu, C. Haase-Pettingell, and J. King.** 1989. Thermostability of temperature-sensitive folding mutants of the P22 tailspike protein. J. Biol. Chem. 264:10693–10698.
40. **van der Vies, S. M., A. A. Gatenby, and C. Georgopoulos.** 1994. Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. Nature 368:654–656.
41. **Waterbury, J. B., and F. W. Valois.** 1993. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl. Environ. Microbiol. 59:3393–3399.
42. **Weigele, P. R., E. Scanlon, and J. King.** 2003. Homotrimeric, β-stranded viral adhesins and tail proteins. J. Bacteriol. 185:4022–4030.
43. **Wood, W. B., F. A. Eiserling, and R. A. Crowther.** 1994. Long tail fibers: genes, proteins, structure, and assembly, p. 282–290. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson, and E. S. Miller (ed.), Molecular biology of bacteriophage T4. American Society for Microbiology, Washington, D.C.

APPENDIX B: IDENTIFICATION OF PEPTIDES PRODUCED BY TRYPSIN DIGESTION OF SYN5 PROTEIN BANDS BY MASS-SPECTROSCOPY

"True" peptides, as determined by the stringent parameters outlined in Chapter IV, are highlighted in light gray; "False" peptides which met the stringent conditions for ORF assignment but did not match the ORF the protein band was eventually assigned to are high-lighted in dark gray.

- * indicates an oxidized methionine
- # indicates an oxidized cysteine
- @ indicates a phosphorylated serine, tyrosine, or threonine

A

Band #1; MW 152kDa; ORF 45

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accessio Sp	RSp	Peptides (Hits) Ions
						1237.06			129 (109 13 7 0 0)
45	4937 - 4952	R.IEPFGQIFSAVADIQYAFESGQMEEDR.A	3077.42	3	7.03	0.84	2713.8	1	41/104
45	4506 - 4557	K.IDWLDGSTITSGSGGFGNSILPMK.T	2453.20	2	6.53	0.77	2322.1	1	32/46
45	4633 - 4670	R.NFSPESFLAMGVETGNNFVPLAGAR.R	2625.28	2	6.35	0.84	1619.6	1	26/48
45	4629	K.AGSGFMHMLHDVANFPLFDWPSR.F	2632.22	2	6.10	0.76	1553.8	1	23/44
45	4693	R.NFSPESFLAMGVETGNNFVPLAGAR.R	2625.28	2	6.08	0.85	1230.1	1	24/48
45	3875	K.TGHNVLVYTGTHVPVLNMALK.E	2264.22	2	5.91	0.79	1827.2	1	25/40
45	4415 - 4447	K.DTWLTALAIDKEDNVFTAATK.T	2323.18	2	5.91	0.81	1691.1	1	23/40
45	3959 - 3978	R.AMM*ANTASAIYDAAFK.V	1691.77	2	5.88	0.01	2059.5	1	24/30
45	3959 - 3978	R.AM*MANTASAIYDAAFK.V	1691.77	2	5.83	0.80	2059.5	1	24/30
45	3849 - 3907	R.ALTNALHPYMQEFNEQYDR.A	2340.07	3	5.81	0.66	3129.8	1	36/72
45	4186 - 4210	R.AMMANTASAIYDAAFK.V	1675.78	2	5.76	0.73	2148.2	1	25/30
45	3816 - 3838	K.TALEGFCAGAAIETGVR.M	1619.84	2	5.67	0.76	2616.5	1	26/32
45	3876 - 3893	R.GIGETLLGAGEVAGDSGK.A	1630.83	2	5.65	0.79	1202.8	1	24/34
45	3072 - 3090	R.ALHSATFGLAGNTANK.I	1572.81	2	5.64	0.79	2422.5	1	24/30
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

45	3675 - 3699	R.LKQEYPELNAEMMMDK.Y	1969.91	2	5.58	0.81	1200.8	1	24/30
45	3173 - 3190	K.GKPAAEVAAEAAPVINR.T	1663.91	2	5.44	0.78	2842.1	1	27/32
45	3852 - 3873	R.AM*M*ANTASAIYDAAFK.V	1707.76	2	5.43	0.82	3014.9	1	25/30
45	3874 - 3895	R.ALTNALHPYMQEFNEQYDR.A	2340.07	2	5.31	0.76	876.1	1	23/36
45	3347 - 3363	R.RALTNALHPYM*QEFNEQYDR.A	2512.16	3	5.26	0.68	504.7	1	28/76
45	3719 - 3761	K.EAFSTTNYGDETILSR.Q	1803.84	2	5.19	0.76	1111.7	1	19/30
45	3646 - 3659	K.HEPAPVEAPASPSPEAAPAANVGTER.S	2552.24	3	5.14	0.70	1846.0	1	45/100
45	3707 - 3729	K.AIDLAGSEGESAVEETFKR.L	2008.98	2	5.12	0.80	1846.4	1	27/36
45	3844 - 3863	K.TGHNVLVYTGTHVPVLNMALK.E	2264.22	3	5.07	0.66	2643.8	1	44/80
45	3721 - 3775	R.TASMLQLTAGDVTK.M	1435.75	2	5.05	0.75	2450.6	1	22/26
45	3502 - 3519	R.LVEDGAEAEMLGR.L	1389.67	2	5.05	0.69	3116.6	1	22/24
45	4268 - 4274	R.LADYSINVGDMNLTM*EGFMK.A	2265.02	2	4.95	0.03	1179.8	1	26/38
45	4037 - 4058	R.LADYSINVGDM*NLTM*EGFMK.A	2281.01	2	4.90	0.20	1427.9	1	25/38
45	2974 - 2992	K.HEPAPVEAPASPSPEAAPAANVGTER.S	2552.24	2	4.87	0.76	1420.8	1	29/50
[redacted]									
45	3250 - 3266	K.VAAQQIEAATSLAKPANGGGGGR.L	2124.12	3	4.80	0.73	2044.9	1	37/88
45	3944 - 3962	R.TASMLQLTAGDVTK.M	1435.75	2	4.80	0.53	2094.4	1	20/26
45	4268 - 4274	R.LADYSINVGDM*NLTMEGFMK.A	2265.02	2	4.79	0.22	884.0	2	24/38
[redacted]									
45	3487 - 3510	K.VAAQQIEAATSLAKP	1400.77	2	4.73	0.82	1524.3	1	21/26
45	3373	R.EAAEPQKQGLEQAAEDFK.Q	1988.96	2	4.71	0.69	1139.9	1	28/34
45	4442 - 4470	R.EDVYLGSWVSEITGKPVVELSR.L	2463.28	2	4.69	0.84	1306.5	1	25/42
45	3160 - 3180	R.LKQEYPELNAEM*M*MDK.Y	2001.89	3	4.69	0.00	2134.8	3	30/60
45	3160 - 3180	R.LKQEYPELNAEM*MM*DK.Y	2001.89	3	4.67	0.02	2241.5	2	30/60
45	4351	R.LADYSINVGDMNLTMEGFM*K.A	2265.02	2	4.64	0.16	1649.4	1	25/38
45	4562 - 4594	K.SMFQGLEPLSGLLNSR.N	1748.90	2	4.63	0.72	1314.0	1	23/30
45	3160 - 3180	R.LKQEYPELNAEMM*M*DK.Y	2001.89	3	4.59	0.75	2297.5	1	30/60
45	4167 - 4215	R.DVVFTNTVEGAIMR.G	1551.78	2	4.53	0.66	2140.0	1	20/26
45	3932 - 3951	K.NAFVNNLTQETEQQDEQILAK.H	2433.19	2	4.52	0.81	1477.6	1	21/40
45	3336 - 3355	K.TAQEGIPSDMHIAQMK.D	1756.84	2	4.52	0.82	655.5	1	22/30
45	3101 - 3123	R.LVEDGAEAEM*LGR.L	1405.66	2	4.49	0.80	2009.8	1	20/24
45	3521	R.LKQEYPELNAEM*MMDK.Y	1985.90	3	4.47	0.01	1625.6	1	29/60

45	3627 - 3651	K.TGHNVLVYTGTHVPVLNM*ALK.E	2280.21	3	4.45	0.60	1114.8	1	32/80
45	3521	R.LKQEYPELNAEMM*MDK.Y	1985.90	3	4.42	0.03	1536.6	2	29/60
45	3236 - 3255	K.KAAIASYHNFFK.T	1396.74	2	4.40	0.69	1484.2	1	18/22
45	3521	R.LKQEYPELNAEMMM*DK.Y	1985.90	3	4.33	0.67	1455.1	3	28/60
45	3357 - 3378	R.EAAEPQKQGLEQAAEDFK.Q	1988.96	3	4.20	0.58	1364.8	1	30/68
45	3664 - 3684	K.SGGVLNDELLK.G	1144.62	2	4.16	0.68	968.9	1	17/20
45	3472 - 3489	R.LKQEYPELNAEMM*MDK.Y	1985.90	2	4.06	0.13	551.1	2	18/30
45	3390 - 3431	K.QGLEQAAEDFKQNQPEAVK.V	2130.05	2	4.00	0.78	471.6	1	15/36
45	3578 - 3600	R.ALTNALHPYM*QEFNEQYDR.A	2356.06	2	3.97	0.81	472.7	1	20/36
45	4037 - 4058	R.LADYSINVGDM*NLTMEGFM*K.A	2281.01	2	3.93	0.42	1025.3	2	22/38
45	3915	R.EIMPDFAQATYDR.L	1556.71	2	3.91	0.75	996.0	1	19/24
45	4351	R.LADYSINVGDMNLTM*EGFMK.A	2265.02	2	3.89	0.45	1012.8	2	21/38
45	4268 - 4274	R.LADYSINVGDMNLTMEGFM*K.A	2265.02	2	3.86	0.73	774.1	3	22/38
45	3520 - 3541	K.AAIASYHNFFKT	1268.64	2	3.84	0.69	1054.1	1	15/20
45	2770 - 2781	R.PITMAAGGNPKQ	1056.55	2	3.80	0.72	1180.5	1	18/20
45	3920 - 3931	K.QEYPELNAEMMMDK.Y	1728.73	2	3.78	0.67	905.6	1	19/26
45	3168	R.KAGESDDILLK.A	1188.65	2	3.77	0.69	1853.3	1	18/20
45	2965	R.KLDETIEAATK.E	1218.66	2	3.76	0.78	1623.8	1	18/20
45	4264 - 4308	K.FASFLNDVPALR.I	1349.72	2	3.74	0.68	781.8	1	16/22
45	3225	R.KAGESDDILLK.A	1188.65	2	3.73	0.72	1282.0	1	16/20
45	3472 - 3489	R.LKQEYPELNAEM*MMDK.Y	1985.90	2	3.73	0.78	411.6	3	16/30
45	3464 - 3479	R.DMVQNSVTDASLR.E	1435.69	2	3.72	0.79	2025.9	1	21/24
45	2819 - 2837	K.SAEVDESALAK.E	1119.55	2	3.64	0.65	1385.4	1	18/20
45	4391 - 4412	K.AGSGFMHM*LHDVANFPLFDWPSR.F	2648.21	3	3.61	0.03	823.6	1	31/88
45	3082 - 3105	R.DM*VQNSVTDASLR.E	1451.68	2	3.60	0.54	1737.7	1	21/24
45	2688 - 2708	K.TAQEGIPSDM*HIAQM*K.D	1788.82	2	3.55	0.61	621.8	1	20/30
45	3038 - 3050	R.KLDETIEAATK.E	1218.66	2	3.52	0.73	1364.5	1	18/20
45	2821 - 2838	K.DEYGAQTGIGK.A	1138.54	2	3.51	0.67	1412.5	1	17/20
45	3197 - 3213	R.LKQEYPELNAEMM*M*DK.Y	2001.89	2	3.51	0.71	643.0	2	19/30
45	4391 - 4412	K.AGSGFM*HMLHDVANFPLFDWPSR.F	2648.21	3	3.51	0.60	716.6	2	29/88
39	2366 - 2382	R.EIGNSQGDMNSGK.G	1336.58	2	3.41	0.73	511.4	1	15/24
39	3746 - 3764	K.GLYSIAGIR.I	949.55	2	3.31	0.53	1185.4	1	14/16
44	3599 - 3624	R.SNPAILDNLQER.N	1369.71	2	3.31	0.70	760.9	1	17/22

45	3010 - 3016	R.LKQEYPELNAEM*M*M*DK.Y	2017.88	2	3.14	0.71	333.3	1	16/30
45	2921 - 2939	K.KEDQYFYR.Q	1148.54	2	3.07	0.53	877.2	1	13/14
	3656 - 3674	K.YADEASFWM*PASFDYR.G	1971.82	3	2.99	0.41	448.7	1	19/60
45	2906 - 2926	R.KLDETIEAATK.E	1218.66	1	2.92	0.73	710.0	1	15/20
46	3442 - 3467	K.TASSFEVYVNK.T	1244.62	2	2.88	0.21	667.9	1	14/20
45	4037 - 4058	R.LADYSINVGDMNLTM*EGFM*K.A	2281.01	2	2.86	0.78	889.5	3	20/38
45	3990 - 4012	R.FLTTSDEFFK.V	1234.60	2	2.79	0.60	813.8	1	16/18
45	2925 - 2963	K.ALAEQVVR.F	885.52	2	2.75	0.58	728.3	1	13/14
45	2467 - 2474	R.ILESGDTAR.F	961.50	2	2.72	0.49	776.4	1	15/16
	3643 - 3665	R.EIM*PDFAQATYDR.L	1572.70	2	2.71	0.41	476.9	1	11/24
45	2959 - 2982	K.TAQEGIPSDM*HIAQMK.D	1772.83	3	2.71	0.26	545.1	1	23/60
45	3085 - 3103	R.KAGESDDILLK.A	1188.65	1	2.69	0.67	611.0	1	13/20
	3851 - 3865	R.SYASSNIESR.K	1113.52	2	2.60	0.41	294.4	1	11/18
45	4351	R.LADYSINVGDM*NLTMEGFMK.A	2265.02	2	2.57	0.74	293.0	3	13/38
	3608 - 3629	K.ATFFPGSAVR.A	1052.55	2	2.53	0.69	707.6	1	15/18
	2637 - 2655	K.NQDVLTR.E	845.45	2	2.52	0.54	608.8	1	11/12
45	3535 - 3564	K.KAAIASYHNFFK.T	1396.74	3	2.50	0.52	752.0	1	24/44
45	4013 - 4041	R.FLTTSDEFFK.V	1234.60	1	2.42	0.68	671.1	1	13/18
45	4154 - 4181	R.LVEDGAEAEMLGR.L	1389.67	2	2.41	0.67	600.2	1	15/24
44	3383 - 3392	K.GSGLFGTNR.S	908.46	2	2.39	0.59	328.5	1	11/16
45	2540 - 2557	K.HTSNLIGTR.L	998.54	1	2.37	0.50	224.9	1	11/16
45	2844 - 2864	K.DEYGAQTGIGK.A	1138.54	1	2.36	0.61	361.3	1	12/20
45	3043 - 3058	R.ILESGDTAR.F	961.50	2	2.36	0.54	638.9	1	14/16
45	4357 - 4379	R.IFFPFVK.T	897.52	2	2.27	0.44	327.4	1	10/12
46	3442 - 3467	R.FNSPPAVGVSIR.I	1243.68	2	2.26	0.52	465.0	2	13/22
45	3944 - 3962	R.DMVQNSVTDASLR.E	1435.69	2	2.26	0.73	803.6	2	15/24
45	3645 - 3668	K.HTSNLIGTR.L	998.54	2	2.26	0.73	100.7	1	10/16
45	3938 - 3960	R.GEGAGEAIDALIK.S	1243.65	1	2.25	0.69	895.3	1	15/24
45	2431 - 2447	R.LLKSEYSR.N	995.55	2	2.24	0.62	425.5	1	11/14
	2672 - 2689	R.S@YASSNIESRK.A	1321.58	2	2.22	0.24	389.9	1	11/20
45	1835 - 1843	K.GVESGDPK.A	788.38	2	2.21	0.47	561.3	1	11/14
	3120 - 3125	K.PSRSC#QAM*MVAR.R	1409.64	2	2.18	0.46	321.7	3	10/22
45	3803 - 3839	R.LFDQEGVFR.L	1110.56	2	2.18	0.64	828.4	1	14/16

45	3249 - 3270	K.EDQYFYR.Q	1020.44	2	2.17	0.57	401.2	1	9/12
	3525 - 3552	K.PSRSC#QAMMVAR.R	1393.65	2	2.17	0.54	591.4	1	11/22
45	2481 - 2495	K.EAVAEYQEK.V	1066.51	1	2.16	0.63	269.4	1	11/16
45	2676 - 2697	K.TAQEGIPSDM*HIAQM*K.D	1788.82	3	2.15	0.45	322.4	1	18/60
45	2899 - 2914	K.ALAEQVVR.F	885.52	1	2.12	0.44	338.6	1	10/14
45	2292 - 2313	K.TGVSNVGGTR.D	947.49	1	2.11	0.54	59.3	1	7/18
45	2476 - 2494	R.ILESGDTAR.F	961.50	1	2.07	0.60	245.8	1	9/16
45	2839 - 2859	K.SAEVDESALAK.E	1119.55	1	2.07	0.62	620.8	1	12/20
45	3462 - 3478	R.DMVQNSVTDASLR.E	1435.69	1	2.05	0.60	157.4	1	11/24
45	3070 - 3097	K.HTSNLIGTR.L	998.54	2	2.04	0.70	216.6	1	13/16
45	4102 - 4110	K.T@AQEGIPSDMHIAQMK.D	1836.81	2	2.04	0.13	453.7	1	13/30
45	2959 - 2982	K.TAQEGIPSDMHIAQM*K.D	1772.83	3	2.01	0.58	435.7	2	21/60
45	3856 - 3872	R.LFDQEGVFR.L	1110.56	1	2.01	0.73	217.0	1	9/16
45	1931 - 1938	K.YAAQADR.L	794.38	1	2.01	0.34	167.8	1	9/12
	2425 - 2442	R.SDNDLSNTGPDTR.I	1391.60	3	1.46	0.16	851.3	1	22/48

B
Band #2 MW 139kDa: ORF 46

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						284.68			29 (27 1 1 0 0)
46	5200 - 5215	R.YSVVVSPSEDSAPTSTLQAQVSTK.T	2481.24	2	6.66	0.83	1609.5	1	28/46
46	6066 - 6087	K.TDGQAFSADFEFVLAATNALPLK.G	2426.22	2	6.07	0.72	3002.6	1	29/44
46	5430 - 5440	K.TTTQFVVETVAVPGGGAANQFADR.G	2436.22	3	5.92	0.65	1721.2	1	44/92
46	5585 - 5603	K.VTTGTFDVVFTTAMPDAK.Y	1900.94	2	5.72	0.78	1790.0	1	23/34
46	6283 - 6296	K.TLVM*DDLLVSSQFVYSLDEIFSQYSTR.A	3172.54	3	5.54	0.84	1143.2	1	35/104
46	6330 - 6352	K.TLVMDDLLVSSQFVYSLDEIFSQYSTR.A	3156.55	3	5.33	0.63	2073.9	1	41/104
46	5370 - 5415	R.FNSAGTLEAGLNVASVTK.V	1778.93	2	5.31	0.80	2075.2	1	25/34
46	5461 - 5475	K.VTTGTFDVVFTTAM*PDAK.Y	1916.93	2	5.12	0.83	1462.1	1	21/34
46	5209 - 5240	R.YSVVVSPSEDSAPTSTLQAQVSTK.T	2481.24	3	4.88	0.71	2732.8	1	42/92
45	5713 - 5729	R.TGTGQYTVVLATAMPNANYSVSSGGEAALLLVSNK.T	3484.76	3	4.66	0.78	971.4	1	41/136
46	5434 - 5448	K.TTTQFVVETVAVPGGGAANQFADR.G	2436.22	2	4.13	0.71	640.6	1	28/46
46	5493 - 5552	K.ATLDGVATTEYTLANATTLRIF	2082.07	2	3.79	0.54	1366.3	1	22/38
39	3421 - 3434	R.EIGNSQGDMNSGK.G	1336.58	2	3.49	0.77	646.3	1	17/24
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
46	5069 - [redacted]	K.TASSFEVYVNK.T	1244.62	2	3.22	0.60	910.1	1	14/20

	5073	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39	4959 - 4967	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	3.11	0.60	1528.9	1	33/72
45	5258 – 5285	K.SGGVLNDELLK.G	1144.62	2	3.01	0.69	801.9	1	16/20
44	5157 – 5190	R.GDFISQDELK.S	1151.56	2	2.97	0.68	614.1	1	14/18
45	3467 – 3527	R.ILESGDTAR.F	961.50	2	2.61	0.50	649.9	1	14/16
46	5186 – 5192	K.ATFFPGSAVR.A	1052.55	2	2.56	0.69	872.9	1	17/18
45	3706 – 3710	K.SAEVDESALAK.E	1119.55	2	2.45	0.54	901.1	1	17/20
44	3548 – 3559	K.TLGDQYNR.Q	966.46	2	2.23	0.64	300.2	1	11/14
	5091 – 5103	R.PSISASAPSSTSR.E	1247.62	2	2.19	0.48	319.9	1	11/24
	5430 – 5440	R.C#EELGGCKSVPLGGCVACPYHTR.N	2436.09	3	2.10	0.66	468.5	3	27/88
46	5399 – 5416	K.TVSGFTVNIYHLTAGASR.D	1893.98	3	2.04	0.67	466.0	1	30/68
	5430 – 5440	R.CEELGGC#KSVPLGGCVACPYHTR.N	2436.09	3	2.01	0.69	474.9	2	27/88

C
Band #3 MW 99kDa: ORF 41

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						436.46			44 (43 0 1 0 0)
41	6000 – 6009	R.QADGNFTLEALNDEGTITGWAQR.E	2507.18	2	5.52	0.84	1402.8	1	23/44
41	5543 – 5562	R.SLAKPVSTGVSIAFVSEADTYSK.I	2357.22	3	5.44	0.63	2696.0	1	37/88
41	4966 – 4986	R.IVTMSGASEVDSNQR.A	1593.75	2	5.42	0.72	2356.9	1	23/28
41	5564 – 5584	K.IFEMSIDSVDNRPQVADITR.I	2306.15	3	5.20	0.71	1696.6	1	34/76
41	5555 – 5574	K.IPNLLGGVSQQPDPVK.L	1661.92	2	5.19	0.78	1320.0	1	21/30
41	4510 – 4517	R.IVTM*SGASEVDSNQR.A	1609.74	2	5.12	0.76	1684.9	1	21/28
41	1304 – 1319	R.IVTM*SGASEVDSNQR.A	1609.74	2	5.09	0.77	2157.7	1	21/28
41	5515 – 5529	R.YNVSVVLQNGGTGFR.K	1610.83	2	4.95	0.81	2817.0	1	25/28
41	5109 – 5130	R.VQC#AAYLENNEYR.S	1629.73	2	4.56	0.69	1687.9	1	19/24
41	1325 – 1346	K.GFAM*ADQM*AQQR.A	1385.58	2	4.16	0.65	1424.4	1	17/22
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
41	5228 – 5250	R.GFNTSTMPHALIR.Q	1444.74	2	4.04	0.67	1050.3	1	20/24
41	5880 – 5975	R.GISDMFFYNNR.L	1363.61	2	4.02	0.70	1381.2	1	16/20
41	4946 – 4964	R.EVGDDDTNPKPSFVGR.G	1762.81	2	3.97	0.62	722.5	1	19/30
41	5384 – [redacted]	K.FVGELATNLPSDTR.W	1519.78	2	3.93	0.68	970.2	1	18/26

	5414 – 5582								
41	5596	R.SAAEGIPGSGSWEETVAPGIER.G	2200.05	2	3.90	0.71	1115.5	1	27/42
	5001 – 5021								
41		R.GFNTSTM*PHALIR.Q	1460.73	2	3.77	0.56	1042.4	1	19/24
41	5546	R.GISDM*FFYNNR.L	1379.60	2	3.73	0.75	1052.2	1	15/20
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5144 – 5167								
41		R.KGDMITVNLNGR.D	1317.69	2	3.61	0.76	1294.2	1	18/22
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39	997 – 1018	R.EIGNSQGDMNSGK.G	1336.58	2	3.27	0.74	494.4	1	15/24
	5687 – 5700								
41		K.LTEISDYFYR.S	1306.63	2	3.25	0.59	927.5	1	15/18
	4984 –								
41	5004	R.KGDM*ITVNLNGR.D	1333.68	2	3.15	0.74	343.5	1	16/22
	4691 –								
41	4706	K.IKAQQAVANIK.T	1183.72	2	3.08	0.63	719.7	1	15/20
	5155 –								
41	5175	R.TDGRDFNLGVR.G	1249.63	3	3.07	0.67	820.2	1	20/40
	5613 –								
41	5661	R.GISDM*FFYNNR.L	1379.60	2	3.02	0.66	781.3	1	14/20
41	5162	R.YAVALYK.D	827.47	2	2.81	0.74	666.8	1	12/12
	5575 –								
41	5587	K.IFEMSIDSVDNRPQVADITR.I	2306.15	2	2.73	0.53	524.9	1	18/38
	4797 –								
41	4854	K.FFNQGNER.Q	1011.46	2	2.71	0.60	535.3	1	12/14
	5185 –								
44	5206	K.GSGLFGTNR.S	908.46	2	2.71	0.65	427.8	1	13/16
	5671 –								
41	5683	K.ITLPGFFTSEENKADR.V	1824.91	2	2.69	0.60	448.5	1	20/30
41	1152	K.AQQAVANIK.T	942.54	2	2.66	0.79	864.6	1	14/16
	5672 –								
44	5685	R.WVDAEDLIK.R	1088.56	2	2.64	0.68	654.9	1	15/16
	5048 –								
39	5061	K.QIGEALATHYDER.I	1502.72	2	2.54	0.66	595.9	1	17/24

41	5488 – 5497	K.WILPGEQR.M	998.54	2	2.51	0.61	849.9	1	13/14
41	5011 – 5031	R.YQIEVDR.I	922.46	1	2.49	0.39	402.3	1	10/12
41	5000 – 5008	R.YQIEVDR.I	922.46	2	2.42	0.65	835.0	1	12/12
41	1336 – 1382	R.IVTMSGAS@EVDSNQR.A	1673.72	3	2.41	0.34	700.6	2	22/56
41	4942	R.EVGDDDTNPKPSFVGR.G	1732.81	3	2.35	0.25	692.6	1	22/60
44	1331 – 1353	K.TLGDQYNR.Q	966.46	2	2.34	0.57	481.6	1	13/14
54	837 – 880	K.TIGGANEQTTR.Q	1147.57	2	2.23	0.69	215.1	1	11/20
44	4740 – 4759	K.TLGDQYNR.Q	966.46	2	2.08	0.51	462.5	1	13/14
41	5152 – 5164	R.YAVALYK.D	827.47	1	1.60	0.68	859.1	1	10/12

D

Band #4 MW 91kDa: ORF 44

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						582.70			59 (56 2 1 0 0)
44	7228 – 7232	R.IYESNFSEGNLFQGSQQSVGFNPVR.A	2804.33	3	6.69	0.76	2433.4	1	44/96
44	6234 – 6251	K.ATQAVQDYYDNPSGESK.A	1872.83	2	5.88	0.80	2541.7	1	24/32
44	7635 – 7637	K.QVISSAVQDGDVALLEQLAETPK.M	2411.27	2	5.81	0.76	1859.9	1	26/44
44	7152 – 7157	R.ELAPTM*M*TNNQNAIGQLVTAGIK.A	2447.22	3	5.71	0.67	2235.0	1	39/88
44	7198 – 7202	R.AQEDSYLDAIGFNEAAPAPR.Q	2135.00	2	5.33	0.81	1725.3	1	26/38
44	6776 – 6785	R.GAAYITSAISHESTWNGMR.E	2051.96	2	5.19	0.79	2030.6	1	26/36
44	6809 – 6815	K.IGFNNDPELEAELAAR.A	1645.78	2	5.18	0.77	1723.1	1	23/28
44	7206 – 7211	K.M*FQDELSQALAAEAR.S	1695.80	2	5.16	0.77	3253.7	1	25/28
44	7648 – 7652	R.AANLTQLQSNIYSAVSSDM*SAAEVWK.V	2800.34	3	5.15	0.74	1767.5	1	36/100
44	7334 – 7353	K.MFQDELSQALAAEAR.S	1679.81	2	5.07	0.76	3117.2	1	24/28
44	6373 – 6387	R.LIESLQQM*PQTEQVR.R	1815.92	2	5.01	0.75	1019.1	1	19/28
44	7497 – 7508	R.TFPEAQQLLQELNR.E	1686.88	2	4.94	0.69	871.8	1	22/26
44	6392 – 6411	R.NIAQINETDQLAYM*QHEM*R.T	2337.05	2	4.66	0.77	570.0	1	22/36
44	7628 – 7633	K.QVISSAVQDGDVALLEQLAETPK.M	2411.27	3	4.58	0.66	1400.0	1	37/88
44	6633 –	R.GAAYITSAISHESTWNGM*R.E	2067.95	2	4.57	0.66	1110.2	1	24/36

	6648 – 6629									
44	6635	R.LIESLQQMPQTEQVR.R	1799.93	2	4.53	0.64	1022.8	1	21/28	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	
44	6514 – 6522	K.GFAMADQMAQQR.A	1353.60	2	4.32	0.70	2193.0	1	19/22	
44	6214 – 6232	K.GFAM*ADQMAQQR.A	1369.59	2	4.29	0.04	1896.6	1	19/22	
44	5908 – 5915	K.GFAM*ADQM*AQQR.A	1385.58	2	4.20	0.76	1440.9	1	17/22	
44	6612 – 6620	R.GAAYITSAISHESTWNGM*R.E	2067.95	3	4.18	0.72	884.7	1	31/72	
44	6725 – 6734	K.FVGELATNLPSDTR.W	1519.78	2	4.14	0.68	1149.2	1	21/26	
44	6342 – 6364	R.IFNNPNASSADLR.W	1418.70	2	4.12	0.67	1226.7	1	20/24	
44	6214 – 6232	K.GFAMADQM*AQQR.A	1369.59	2	4.12	0.74	1675.5	2	18/22	
44	6854 – 6860	R.NGGLISWASWHNNSAR.L	1769.85	2	4.07	0.66	759.0	1	17/30	
44	5965 – 6022	R.QTQVSQADLSAQK.A	1403.71	2	3.93	0.73	1389.5	1	20/24	
44	6418 – 6446	R.NGKPSLSLLESQSPK.A	1584.86	3	3.89	0.63	1294.1	1	26/56	
44	7159 – 7163	R.ELAPTM*M*TNNQNAIGQLVTAGIK.A	2447.22	2	3.89	0.87	512.6	1	19/44	
44	5940 – 5954	K.IKAQQAVANIK.T	1183.72	2	3.85	0.65	969.9	1	17/20	
44	6723 – 6738	K.PSLSLLESQSPK.A	1285.70	2	3.81	0.73	1823.5	1	20/22	
44	5485 – 5526	K.GRADAIHAQR.A	1094.58	2	3.61	0.66	1154.8	1	16/18	
44	6626 – 6631	R.GDFISQDELK.S	1151.56	2	3.40	0.68	746.1	1	15/18	
44	6454 – 6470	K.GSGLFGTNR.S	908.46	2	3.26	0.71	649.1	1	14/16	

44	5851 – 5863	K.AQQAVANIK.T	942.54	2	3.25	0.70	860.4	1	14/16
44	6207 – 6218	R.PEYTLTNDPK.I	1177.57	2	3.17	0.72	818.6	1	15/18
44	7095 – 7100	R.WVDAEDLIK.R	1088.56	2	3.04	0.64	659.9	1	15/16
39	5765 – 5777	R.EIGNSQGDMNSGK.G	1336.58	2	3.04	0.71	611.0	1	17/24
44	5480 – 5501	R.AVDTSAQEK.I	948.46	1	2.82	0.62	631.6	1	12/16
44	6843 – 6849	R.SNPAILDNLQER.N	1369.71	2	2.80	0.65	563.3	1	15/22
44	6538 – 6547	R.AFVDAQLER.N	1048.54	2	2.76	0.55	1074.8	1	14/16
41	5448 – 5452	K.DGSGNTQVR.V	933.44	2	2.74	0.46	466.5	1	14/16
44	7175 – 7187	K.ELKPFIEQSLLGGVSM*R.D	1920.02	3	2.63	0.74	695.9	1	27/64
44	7214 – 7276	R.AQEDSYLDAIGFNEAAPAPR.Q	2135.00	3	2.57	0.59	800.5	1	27/76
44	6440 – 6459	R.YAVALYK.D	827.47	2	2.46	0.81	529.0	1	11/12
44	5415 – 5433	K.M*GTQPNGPK.L	945.44	2	2.38	0.69	673.8	1	14/16
	6591 – 6608	R.S@VS@ SLSACSR.A	1156.42	2	2.33	0.23	467.4	1	12/18
41	6740 – 6745	R.DFNLGVR.G	820.43	2	2.29	0.41	674.0	1	10/12
58	6651 – 6659	K.RHISQMGGGMDTEEVR.A	1802.83	2	2.27	0.02	90.3	4	9/30
44	6479 – 6492	R.AFVDAQLER.N	1048.54	1	2.26	0.52	185.9	1	10/16
	6651 – 6659	R.WSSPTRC#LM*IIMHR.S	1803.88	2	2.22	0.07	111.1	2	10/26
41	6138 – 6146	K.FFNQGNER.Q	1011.46	2	2.16	0.64	420.5	1	11/14
	6651 –	R.WSSPTRC#LMIIM*HR.S	1803.88	2	2.11	0.38	170.9	1	11/26

	6659								
	5866 –								
44	5875	K.AQQAVANIK.T	942.54	1	2.07	0.69	587.4	1	11/16
	5947 –								
44	5966	K.TLGDQYNR.Q	966.46	1	2.05	0.36	201.5	1	9/14
	6788 –								
	6800	R.DWLTLSMS@SVFLIR.S	1747.85	2	2.04	0.11	189.6	1	11/26
	7128 –								
	7141	R.PTRS@PADLR.L	1092.52	2	2.00	0.03	176.0	2	9/16
	5734 –								
44	5750	K.GNVYQAR.A	807.41	1	1.56	0.45	114.0	1	8/12
	5576 –								
44	5581	R.TM*ADISSK.G	868.40	1	1.54	0.63	147.3	1	10/14
	5918 –								
44	5928	R.TMADISSK.G	852.41	1	1.52	0.61	705.7	1	10/14

E

Band #5 MW 76kDa: Unmatched

	Reference					Score	Accession		Peptides (Hits)
	Scan(s)	Sequence	MH+	Z	XC	Δ Cn	Sp	RSp	Ions
ORF						60.5			0
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39	4799 – 4820	K.QIGEALATHYDER.I	1503.60	2	3.01	0.743	594.4	1	19/24
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39	5032 – 5040	K.QIGEALATHYDER.I	1503.60	2	2.64	0.714	644.5	1	19/24
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
53	4385 – 4430	R.FPLSTSSDLK.V	1095.23	2	2.22	0.489	509.9	1	12/18

F
Band #6 MW 65kDa: ORF 54

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn 358.63	Accession Sp	RSp	Peptides (Hits) Ions 36 (35 1 0 0 0)
54	5323 – 5329	K.VLVKEESDGNEATSAFLGNFSATDITHYATAR.N	3414.65	3	7.31	0.76	1989.0	1	43/124
54	5432 – 5439	R.NQIETTNASGVTPQIVTGLVVGYK.A	2489.33	3	6.23	0.69	2603.9	1	41/92
54	5435 – 5443	R.NQIETTNASGVTPQIVTGLVVGYK.A	2489.33	2	6.12	0.81	1047.3	1	26/46
54	5073 – 5078	R.VNTWTNYGGGVDLKDELSDTTSGDNAAR.G	2956.36	3	6.10	0.69	3049.1	1	41/108
54	5380 – 5391	K.QVQPTVVNTAGTLLQDVK.V	1911.06	2	4.92	0.67	1924.3	1	26/34
54	4138 – 4144	K.DELSDTTSGDNAAR.G	1451.63	2	4.80	0.64	2259.2	1	23/26
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
54	5457 – 5462	K.EESDGNEATSAFLGNFSATDITHYATAR.N	2975.33	3	4.32	0.68	851.5	1	31/108
54	5378 – 5387	K.QVQPTVVNTAGTLLQDVK.V	1911.06	3	4.18	0.63	1109.3	1	34/68
54	4620 – 4624	K.ATVDAYTTLESPEK.L	1524.74	2	4.09	0.64	1481.7	1	19/26
54	4687 – 4693	R.GIFRPWSTNNTGTK.N	1578.80	2	3.89	0.73	1050.0	1	19/26
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
54	4728 – 4732	R.STATLTLTGLKPNTEVR.V	1802.00	2	3.62	0.61	638.5	1	16/32
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
54	5253 –	R.IGFVNGDVDVDGIK.V	1447.74	2	3.29	0.58	1819.9	1	23/26

	5259								
	3903 –								
39	3908	R.EIGNSQGDMNSGK.G	1336.58	2	3.25	0.75	675.6	1	17/24
	3841 –								
54	3855	K.TIGGANEQTTR.Q	1147.57	2	3.24	0.75	676.5	1	17/20
	4711 –								
54	4719	R.STATLTLTGLKPNTEVR.V	1802.00	3	2.99	0.46	386.4	1	24/64
	5201 –								
54	5213	R.QFGYNYLLNEGR.I	1473.71	2	2.95	0.77	834.3	1	18/22
	4061 –								
54	4068	K.KIYNLGNAK.L	1020.58	2	2.83	0.72	814.0	1	14/16
	5240 –								
54	5249	K.QVQPTVVNTAGTLLQDVK.V	1911.06	2	2.65	0.59	544.9	1	17/34
	4387 –								
54	4397	R.SIITDSGGNIK.V	1104.59	2	2.57	0.52	1021.4	1	17/20
	4340 –								
37	4345	R.LQQPYQSLGSK.G	1248.66	2	2.50	0.56	414.1	1	13/20
	4414 –								
54	4424	R.SIITDSGGNIK.V	1104.59	1	2.49	0.69	615.0	1	13/20
	4134 –								
37	4142	R.AGTGAIIQGR.A	943.53	2	2.46	0.50	392.3	1	13/18
	4145 –								
37	4156	R.SEIDM*NLSK.M	1052.49	2	2.40	0.75	607.5	1	13/16
	3892 –								
37	3896	K.DSNAVGEDGPK.F	1088.49	2	2.38	0.71	453.6	1	13/20
	4951 –								
	4959	R.DWLTLSMS@SVFLIR.S	1747.85	2	2.30	0.13	202.5	1	11/26
	4124 –								
54	4129	R.QFNGSQAGIR.N	1077.54	2	2.24	0.58	215.1	1	13/18
	4303 –								
54	4320	K.IYNLGNAK.L	892.49	1	2.17	0.72	377.1	1	10/14
	4331 –								
37	4338	R.TVQEMIR.D	876.46	2	2.15	0.57	444.0	1	10/12
	4066 –								
37	4069	R.EVQM*ELDR.Q	1035.47	2	2.15	0.50	138.4	1	8/14
	4007 –								
39	4010	R.NADYDVR.Y	852.39	2	2.15	0.71	631.6	1	11/12

54	4174 – 4182	R.GIVIAQK.Q	728.47	1	2.11	0.51	597.9	1	8/12
	4951 – 4959	R.DWLTLSMSS@VFLIR.S	1747.85	2	2.01	0.34	202.5	1	11/26
54	4431 – 4448	R.AYLSGSLR.Q	866.47	1	1.70	0.49	43.3	3	6/14

G
Band #7 MW 60kDa: ORF 37

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						544.72			55 (52 3 0 0 0)
37	5529 – 5539	K.SPETM*ANEAQQAQQQQM*TASLM*GQAGQLAK.S	3224.45	3	7.59	0.74	1988.5	1	42/116
37	6486 – 6492	K.SPETMANEAQQAQQQQMTASLMGQAGQLAK.S	3176.48	3	6.51	0.75	1892.9	1	42/116
37	6794 – 6798	K.GLVMPTVVAGLGGVGR.G	1482.85	2	4.90	0.80	1849.3	1	26/30
37	6215 – 6256	K.GLVM*PTVVAGLGGVGR.G	1498.84	2	4.82	0.69	1400.5	1	23/30
37	6200 – 6205	R.LAAASGIDTLNLVK.S	1385.80	2	4.74	0.74	2217.5	1	21/26
37	6948 – 7003	R.ISDAFLILNVR.Q	1260.73	2	4.38	0.56	1977.6	1	18/20
37	5002 – 5006	R.AEDVSVVQANK.G	1159.60	2	4.34	0.70	1499.7	1	17/20
37	6147 – 6158	R.YVIERDGDGNVIEIITR.E	1962.03	3	4.34	0.58	1210.6	1	30/64
37	7887 – 7892	K.LMLSLFPIQTSFFK.L	1671.92	2	4.24	0.71	1283.8	1	23/26
37	5744 – 5751	K.VVFMVSPSATTKPQSLAR.A	1919.04	2	4.19	0.79	739.0	1	20/34
37	4842 – 4849	K.M*VM*QQIAESSDR.V	1426.62	2	4.16	0.78	2008.2	1	20/22
37	5280 – 5288	K.QSLLEGKDSNAVGEDGPK.F	1843.90	3	4.14	0.67	826.2	1	29/68
37	7694 – 7703	K.LM*LSLFPIQTSFFK.L	1687.91	2	3.93	0.55	1342.8	1	23/26
37	6221 – 6233	K.HLIVTGNVLVFAGK.K	1467.87	2	3.87	0.78	1288.2	1	21/26
37	6637 –	R.DGDGNVIEIITR.E	1301.67	2	3.65	0.44	1115.2	1	18/22

	6641								
	4734 –								
39	4740	R.EIGNSQGDMNSGK.G	1336.58	2	3.60	0.71	478.1	1	14/24
	5596 –								
37	5606	R.FNVVDGESYGR.G	1242.58	2	3.58	0.67	1028.4	1	17/20
	5261 –								
37	5267	K.QSLLEGKDSNAVGEDGPK.F	1843.90	2	3.49	0.72	622.7	1	20/34
	6600 –								
37	6607	K.LQINDAEIASVPELTPEVR.S	2094.11	3	3.46	0.64	659.0	1	29/72
	5146 –								
37	5152	R.SLIEGSAAAAK.V	1017.56	2	3.44	0.71	1295.5	1	18/20
	5876 –								
37	5880	R.ADREDFLDMAR.R	1338.61	2	3.36	0.65	814.5	1	15/20
	5572 –								
39	5584	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	3.36	0.74	1483.2	1	33/72
	5738 –								
54	5742	R.STATLTLTGLKPNTEVR.V	1802.00	3	3.25	0.63	558.7	1	24/64
	5466 –								
37	5472	R.ADREDFLDM*AR.R	1354.60	3	3.25	0.42	808.5	1	22/40
	5867 –								
37	5875	K.VVFMVSPSATTK.P	1266.68	2	3.09	0.85	832.1	1	16/22
	5580 –								
37	5585	K.VVFM*VSPSATTKPQSLAR.A	1935.03	2	3.01	0.73	509.3	1	19/34
	4673 –								
37	4680	K.DSNAVGEDGPK.F	1088.49	2	2.76	0.69	610.2	1	15/20
	5528 –								
37	5541	R.SEIDMNLSK.M	1036.50	1	2.72	0.64	637.3	1	12/16
	5031 –								
37	5043	R.AGTGAIIQGR.A	943.53	2	2.68	0.63	951.9	1	15/18
	5313 –								
37	5321	R.LQQPYQSLGSK.G	1248.66	2	2.66	0.63	533.4	1	15/20
	4852 –								
37	4856	K.GVNALSSK.L	775.43	2	2.66	0.57	518.9	1	12/14

37	7887 – 7892	K.GVNALSS@KLMLSLFPIQTSFFK.L	2508.30	3	2.63	0.01	239.3	2	20/84
37	7887 – 7892	K.GVNALS@SKLMLSLFPIQTSFFK.L	2508.30	3	2.60	0.25	218.0	3	19/84
37	5473 – 5479	R.ADREDFLDM*AR.R	1354.60	2	2.53	0.58	276.5	1	14/20
39	5403 – 5414	K.QIGEALATHYDER.I	1502.72	2	2.49	0.70	480.7	1	15/24
37	5587 – 5594	R.SEIDMNLSK.M	1036.50	2	2.48	0.78	710.4	1	14/16
	5626 – 5640	R.PAPPAS@SPTVSR.Q	1246.59	2	2.41	0.01	695.5	1	13/22
39	4882 – 4889	R.NADYDVR.Y	852.39	2	2.40	0.71	611.4	1	11/12
	5626 – 5640	R.PAPPASS@PTVSR.Q	1246.59	2	2.38	0.38	695.5	1	13/22
37	4878 – 4884	R.VQLTAAM*K.H	877.48	2	2.36	0.63	215.5	1	9/14
37	5683 – 5687	R.AGTGAIIQGR.A	943.53	2	2.35	0.60	281.1	1	12/18
37	5049 – 5061	R.SEIDM*NLSK.M	1052.49	2	2.33	0.72	791.8	1	14/16
39	4509 – 4523	R.EIGNSQGDM*NSGK.G	1352.57	2	2.17	0.64	205.9	1	11/24
37	5282 – 5289	R.TVQEMIR.D	876.46	1	2.07	0.61	204.7	1	9/12
	6637 – 6641	R.PSRAQLM*LLTR.@	1301.73	2	2.05	0.73	237.3	2	9/20
37	4875 – 4932	K.FGVAQGK.G	706.39	2	2.03	0.46	1106.3	1	11/12
	6041 – 6047	R.DWLTLSMS@SVFLIR.S	1747.85	2	2.01	0.13	173.9	2	10/26
37	4892 – 4901	K.FGVAQGK.G	706.39	1	1.99	0.49	594.7	1	10/12
37	4761 – 4769	K.GVNALSSK.L	775.43	1	1.90	0.51	327.8	1	9/14
37	5047 –	R.AGTGAIIQGR.A	943.53	1	1.75	0.60	240.3	1	11/18

	5057 – 5309 –									
37	5317	R.VQLTAAMK.H	861.49	1	1.64	0.63	552.5	1	10/14	
	5070 –									
37	5075	R.SEIDM*NLSK.M	1052.49	1	1.64	0.58	44.6	2	7/16	
	6031 –									
	6040	R.SLLPAEFQK.Q	1032.57	1	1.51	0.48	313.5	1	11/16	

H

Band # 8 MW 48: ORF 53

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn 406.78	Accession Sp	RSp	Peptides (Hits) Ions 41 (39 2 0 0 0)
53	6058 – 6075	K.FAQSSLADIGVTGDMTFQAYR.F	2278.08	2	6.63	0.82	2785.9	1	31/40
53	5724 – 5742	K.TIALVQTGNLSTDGVTLK.C	1831.02	2	6.02	0.83	3129.8	1	26/34
53	5575 – 5580	K.LKEGVFLTNFNATDTNR.I	1939.99	2	5.94	0.78	2561.0	1	25/32
53	5872 – 5877	K.FAQSSLADIGVTGDM*TFQAYR.F	2294.07	2	5.88	0.84	2314.7	1	30/40
53	6591 – 6596	R.GAASAGTDAPITVVAIGLATGQYVSATGTIAR.S	2960.57	3	5.40	0.74	2164.0	1	44/124
53	5962 – 5968	R.EQGDKFAQSSLADIGVTGDMTFQAYR.F	2835.33	3	5.32	0.83	2170.9	1	37/100
53	6932 – 6937	R.LGSDIDAGAGTQIGNVSSELASFLGDTLK.L	2836.42	3	5.13	0.80	1922.2	1	39/112
53	5004 – 5047	R.VFYTNDDAGDNTGR.D	1544.66	2	5.12	0.78	2554.2	1	23/26
53	5787 – 5800	K.EGVFLTNFNATDTNR.I	1698.81	2	4.76	0.77	1540.5	1	21/28
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
53	6141 – 6159	R.HPFPFTPITDESFELKDGWDFANDASR.Y	3139.44	3	4.05	0.70	686.9	1	30/104
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
53	5679 – 5686	K.DGWDFANDASR.Y	1253.52	2	3.65	0.59	844.1	1	15/20
53	5737 – 5747	R.STSNVVSLVSPLER.N	1487.81	2	3.58	0.60	1224.6	1	21/26
39	4590 – 4600	R.EIGNSQGDMNSGK.G	1336.58	2	3.55	0.76	676.1	1	17/24
39	5250 – 5258	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	3.53	0.71	1246.0	1	32/72
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
53	5959 – 5964	K.VTLDDSQVQTGTPATITITYSASPFAR.T	2840.43	3	3.03	0.45	303.0	6	19/104
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
53	5363 – 5385	K.DTEVVFDTAAK.T	1195.58	1	2.78	0.80	503.5	1	14/20
53	5406 – 5425	R.FPLSTSSDLK.V	1094.57	2	2.69	0.56	765.1	1	14/18
53	5395 – 5414	R.STLTVFAR.E	894.50	2	2.67	0.62	737.8	1	13/14
39	5115 – 5127	K.QIGEALATHYDER.I	1502.72	2	2.64	0.68	646.3	1	18/24

37	5506 – 5519	R.ADREDFLDMAR.R	1338.61	2	2.63	0.57	653.2	1	14/20
37	4857 – 4867	R.SEIDM*NLSK.M	1052.49	2	2.61	0.72	612.8	1	13/16
53	5268 – 5284	K.EQWKEEASLIR.H	1388.72	2	2.58	0.65	538.3	1	16/20
37	4555 – 4562	K.DSNAVGEDGPK.F	1088.49	2	2.57	0.77	514.5	1	14/20
37	5022 – 5056	R.LQQPYQSLGSK.G	1248.66	2	2.45	0.70	489.3	1	14/20
37	5031 – 5048	R.TVQEMIR.D	876.46	2	2.34	0.57	516.7	1	11/12
53	5339 – 5352	K.C#LYSFTK.E	918.44	2	2.33	0.65	576.7	1	12/12
53	5953 – 5958	R.HPFPFTPITDESFELK.D	1904.94	2	2.22	0.73	259.0	1	17/30
37	4715 – 4721	K.FGVAQGK.G	706.39	2	2.08	0.47	1291.2	1	12/12
37	4642 – 4645	K.GVNALSSK.L	775.43	1	1.86	0.57	405.3	1	10/14
53	5705 – 5726	K.DGWDFANDASR.Y	1253.52	1	1.78	0.60	127.0	1	8/20
53	4994 – 5020	K.EEASLIR.H	817.44	1	1.55	0.53	173.4	1	8/12

I
Band #9 MW 35kDa: ORF 39

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						230.37			23 (23 0 0 0 0)
39	18524 – 18573	K.TLVMDDLLVSSQFVYSLDEIFSQYSTR.A	3156.55	3	6.16	0.66	1164.1	1	33/104
39	17781 – 17804	K.LFSGEVFTAFNNASIFK.G	1891.96	2	5.67	0.75	1653.2	1	22/32
39	17507 – 17522	R.QYYSLISSVDTNILNR.E	1885.97	2	5.43	0.70	1487.0	1	23/30
39	18103 – 18136	R.QYYSLISSVDTNILNR.E	1885.97	2	5.03	0.72	1405.2	1	23/30
39	19068 – 19086	K.LFSGEVFTAFNNASIFK.G	1891.96	2	4.93	0.79	1573.4	1	23/32
39	17570 – 17582	R.VAVLSPRQYYSLISSVDTNILNR.E	2608.41	3	4.88	0.61	1087.4	1	28/88
39	18509	K.LFSGEVFTAFNNASIFK.G	1891.96	2	4.39	0.80	2065.3	1	25/32
39	16845 – 16858	R.QYYSLISSVDTNILNR.E	1885.97	2	4.26	0.77	1091.1	1	22/30
39	18413 – 18442	K.LFSGEVFTAFNNASIFK.G	1891.96	2	4.12	0.81	1648.1	1	23/32
39	12848 – 12853	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	3.91	0.65	1412.6	1	32/72
39	12890 – 12900	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	2	3.74	0.60	625.4	1	21/36
39	6538 – 6542	R.EIGNSQGDMNSGK.G	1336.58	2	3.40	0.74	497.0	1	15/24
39	17503 – 17567	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	3.14	0.66	1041.1	1	27/72
39	5299 – 5309	R.EIGNSQGDM*NSGK.G	1352.57	2	3.08	0.63	306.7	1	14/24
39	14963 – 15056	K.GLYSIAGIR.I	949.55	2	2.94	0.54	1149.9	1	14/16
39	10871 – 11845	K.QIGEALATHYDER.I	1502.72	2	2.89	0.72	739.7	1	19/24

39	16509 – 16513	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	2.80	0.48	906.0	1	27/72
39	10061 – 10065	K.SKQFMFTGK.L	1073.55	2	2.73	0.70	716.0	1	13/16
39	8542 – 8548	K.SKQFM*FTGK.L	1089.54	2	2.56	0.54	453.5	1	11/16
39	7117 – 8069	R.NADYDVR.Y	852.39	2	2.27	0.70	615.0	1	11/12
	14390 – 14397	R.DWLTLSMS@SVFLIR.S	1747.85	2	2.16	0.13	201.6	1	11/26
39	10228 – 10233	K.LAMGC#GSLR.T	964.47	2	2.15	0.61	829.7	1	15/16
39	14683 – 14690	K.GLYSIAGIR.I	949.55	1	1.59	0.62	61.5	1	7/16

J

Band #10 MW 23 kDa: ORF 43

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						140.50			14 (14 0 0 0 0)
43	4486 - 4490	K.VLASGATGQSVGLLVLDAER.Q	1956.08	2	5.42	0.82	2309.1	1	25/38
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
43	2675 - 2680	R.QADLTYQNAQR.Q	1307.63	2	3.26	0.70	879.9	1	17/20
39	2400 - 2405	R.EIGNSQGDMNSGK.G	1336.58	2	3.10	0.70	581.5	1	16/24
39	3196 - 3262	K.LSAGYHTPGTPIVGDAGIK.A	1853.98	3	2.79	0.40	687.1	1	24/72
43	7342 - 7366	K.ILQHQGQIR.A	1092.63	2	2.47	0.39	889.1	1	14/16
39	2130 - 2139	R.EIGNSQGDM*NSGK.G	1352.57	2	2.46	0.72	253.3	1	12/24
43	2912 - 2991	K.ILQHQGQIR.A	1092.63	2	2.33	0.30	1017.6	1	15/16
43	2938 –	R.AAYLDNVR.N	921.48	2	2.28	0.63	703.8	1	13/14

	2947								
43	2930 - 2935	R.AAYLDNVR.N	921.48	1	2.22	0.53	260.0	1	9/14
39	2551 - 2558	R.NADYDVR.Y	852.39	2	2.20	0.70	606.5	1	11/12
43	3026 - 3037	K.ALNLYAK.S	792.46	2	2.17	0.77	443.2	1	11/12
43	2830 - 2835	K.SYVAEQLK.L	937.50	2	2.07	0.38	560.8	1	12/14
43	2839 - 2844	K.SYVAEQLK.L	937.50	1	1.79	0.46	594.0	1	10/14

K

Band #11 MW 22kDa: ORF 40

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						308.58			32 (28 2 2 0 0)
40	9146 - 9162	R.AAIIEYETQQGDYNMLESESGR.D	2504.13	3	7.81	0.82	4122.7	1	41/84
40	8804 - 8821	R.AAIIEYETQQGDYNMLESESGR.D	2520.12	3	6.56	0.67	3705.9	1	39/84
40	9138 - 9145	R.AAIIEYETQQGDYNMLESESGR.D	2504.13	2	6.00	0.64	2310.3	1	25/42
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
40	9351 - 9363	R.IAIPSNVLSLDC#ASTSK.Y	1775.92	2	3.98	0.80	713.9	1	24/32
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
39	5912 - 5918	R.EIGNSQGDMNSGK.G	1336.58	2	3.13	0.80	411.5	1	14/24
	9154 - 9180	R.AC#PVWC#HLS@S@QEIS@VT@SAR.A	2507.91	2	2.82	0.00	293.7	2	11/36

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	9154 - 9180	R.AC#PVWC#HLS@S@QEIS@VTS@AR.A	2507.91	2	2.81	0.02	293.7	2	11/36
36	5969 - 5974	K.QQQQAASGTASLR.I	1345.68	2	2.81	0.83	401.3	1	17/24
	9154 - 9180	R.AC#PVWC#HLS@S@QEISVT@S@AR.A	2507.91	2	2.75	0.43	360.7	1	12/36
39	7718 - 7729	K.QIGEALATHYDER.I	1502.72	2	2.66	0.71	534.2	1	16/24
39	9581 - 9614	R.DGDGNVIEIITR.E	1301.67	2	2.62	0.42	513.5	1	15/22
40	9384 - 9406	R.DIYTYRPFDAVYR.F	1678.82	2	2.59	0.58	354.6	1	13/24
40	5544 - 5549	R.AVGSAEAVK.Y	831.46	2	2.58	0.68	470.9	1	14/16
	8787 - 8832	R.YTGNVEATIKGAIS@GNSC#PSTSAR.V	2521.14	2	2.40	0.19	182.2	1	11/46
58	8194 - 8200	R.KLYSLLSR.G	979.59	2	2.40	0.68	416.9	1	11/14
40	6350 - 6354	R.AVGSAEAVK.Y	831.46	2	2.28	0.61	309.5	1	12/16
57	9031 - 9044	R.EADHMLQYFVK.N	1380.66	2	2.28	0.65	334.0	1	15/20
40	7813 - 7852	R.AYPFIKDNTGR.I	1281.66	2	2.25	0.53	455.0	1	15/20
55	9138 - 9145	K.SVNM*T@FNVKT@IGGANEQT@T@R.Q	2503.95	2	2.17	0.66	306.2	2	13/38
	8904 - 8927	R.DWLTLSMS@SVFLIR.S	1747.85	2	2.12	0.14	172.4	1	10/26
40	9993 - 10003	R.DIYTYRPFDAVYR.F	1678.82	2	2.11	0.49	233.2	1	11/24
	9437 - 9448	R.AQVWPLWAYSQS@FR.T	1818.84	2	2.11	0.29	221.2	1	10/26
54	9138 - 9145	K.S@VNM*TFNVKT@IGGANEQT@T@R.Q	2503.95	2	2.05	0.75	274.7	3	12/38
40	7991 -	R.AANLFAGRA	819.45	1	1.92	0.39	226.3	1	10/14

	7996								
40	8181 - 8187	R.GGFLYDK.A	799.40	1	1.67	0.27	577.4	1	8/12
40	8686 - 8722	R.AANLFAGR.A	819.45	1	1.66	0.42	168.7	1	9/14
40	5553 - 5560	R.AVGSAEAVK.Y	831.46	1	1.56	0.49	408.6	1	10/16
40	7841 - 7850	K.NYITIR.A	779.44	1	1.50	0.40	160.4	1	8/10

L
Band #12 MW 16: ORF 58

ORF	Reference Scan(s)	Sequence	MH+	Z	XC	Score Δ Cn	Accession Sp	RSp	Peptides (Hits) Ions
						160.84			16 (16 0 0 0 0)
58	16308 - 16312	R.LGDVVSVSEGAGR.A	1245.64	2	5.07	0.75	1447.0	1	21/24
58	17251 - 17256	R.LGDVVSVSEGAGR.A	1245.64	2	4.24	0.75	1365.7	1	21/24
58	17642 - 18621	R.VLDATAVAAIK.A	1071.64	2	4.07	0.73	1372.1	1	18/20
58	15385 - 15396	R.LGDVVSVSEGAGR.A	1245.64	2	3.82	0.79	1347.4	1	20/24
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
58	19583 - 19587	R.VLDATAVAAIK.A	1071.64	2	3.70	0.76	1312.8	1	18/20
36	6157 - 6161	K.QQQQAASGTASLR.I	1345.68	2	2.78	0.82	540.2	1	18/24
36	4253 - 4259	K.QQQQAASGTASLR.I	1345.68	2	2.65	0.83	505.2	1	18/24
58	16634 - 17602	R.KLYSLLSR.G	979.59	2	2.62	0.70	458.0	1	11/14
50	2952 - 2956	R.AELNAAR.R	744.40	2	2.62	0.52	622.4	1	11/12
58	20468 - 20472	R.KLYSLLSR.G	979.59	2	2.45	0.69	441.0	1	11/14
58	19036 - 20004	K.LYSLLSR.G	851.50	2	2.40	0.46	487.3	1	11/12
58	18567 - 19536	R.KLYSLLSR.G	979.59	2	2.39	0.69	446.8	1	11/14
40	21093 - 21119	R.DIYTYRPFDAVYR.F	1678.82	2	2.37	0.61	242.8	1	11/24
58	21383 - 21387	R.KLYSLLSR.G	979.59	2	2.36	0.70	421.1	1	11/14
36	14168 - 14175	K.GGGLNIPK.@	755.44	1	1.65	0.43	232.0	1	9/14