EVOLUTION OF BACTERIOPHAGE HOST ATTACHMENT USING DET7 AS A MODEL

By

Robert H. Edgar

Bacherlor of Science, Geneva College, 2002
Bacherlor of Arts, Geneva College, 2002

Masters of Arts, University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This thesis was presented

by

Rob Edgar

It was defended on

November 19th, 2013

and approved by

Michael Grabe, Associate Professor, Department of Biological Sciences

Jim Pipas, Professor, Department of Biological Sciences

Craig Peebles, Professor, Department of Biological Sciences

Roger Hendrix, Distinguished Professor, Department of Biological Sciences

Evolution of Bacteriophage Host Attachment using Det7 as a Model
Robert H. Edgar, M.S.

University of Pittsburgh, 2013

Copyright © by Robert H. Edgar

Evolution of Bacteriophage Host Attachment using Det7 as a Model

Robert H. Edgar, M.S.

University of Pittsburgh, 2013

Bacteriophage fulfill a crucial role in maintaining bacterial population levels as well as being a driving force behind bacterial diversity. Bacteria display a huge variety of cell-surface molecules, and these cell surface antigens play a central role in bacterial virulence as well as being recognition sites of phage infection. Due to the wide range in variation of antigens, bacteriophages (phages) have adapted a variety of cell surface recognition systems.

All phage with contractile tails likely share a common ancestor and therefore share structural similarity. The best characterized contractile tail is that of phage T4 that shares remarkable similarity to the core tail genes of Det7. All of the major structural tail proteins and their T4 homologs have been biochemically identified in Det7. Despite the similarity in tail structure, there is no similarity in the host attachment proteins between Det7 and T4.

Most phage are equipped with only one type of primary cell adhesion protein, with the best characterized phage adhesion systems being those of P22 and T4. Phage Det7, a member of the recently identified Viunalikevirus family, carries five different cell attachment proteins.

Det7 gp5 shares amino acid and structural similarity in host attachment proteins with P22 and 9NA. The amino acid sequences of the active sites of these proteins are conserved, and the host range of Det7 includes the P22 and 9NA susceptible *Salmonella* serovars, despite these phage being morphologically different in tail structure: Det7 has a contractile tail (*Myoviridae*), P22 has a short tail (*Podoviridae*), and 9NA has a non-contractile long (Siphoviridae) tail.

We hypothesize that each of Det7's tail spike proteins are responsible for its ability to infect a particular subset of hosts. To investigate this, biochemical and genomic means were used to characterize the structure of the Det7 tail. Genome comparison of related phage in addition to host ranges allows for further characterization of tail spikes and determination of host specificities of tail spike proteins. This project provides insight into the function of these tail proteins and supports the hypothesis of evolution by recombination as the primary mechanism in adaptation of their host ranges.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Viruses are ubiquitous and the most abundant, populous biological entities on earth and bacteriophages (phages) represent the vast majority of viruses in the biosphere [1-4]. Phages have a major impact on food webs, act as agents of gene transfer between bacterial cells - and even play a crucial role in the non-host-derived immunity of metazoans from bacterial infection[6]. Phages follow at least two types of infection pathways: the lytic and lysogenic cycles, though a third pathway has been postulated [7, 8]. Temperate phages can undergo the lytic or lysogenic cycle. They are able to integrate their DNA into the chromosome of their bacterial hosts, forming lysogens, and silencing the expression of phage encoded genes that would lead to lysis of their host. The phage carrier state can be defined as the third lifecycle, where phage carrier cells harbor episomal phage DNA elements that segregate asymmetrically upon bacterial divisions [9]. Integrated temperate phage (prophage) always leads to lysis of the infected host but require an induction event to do so. Prophages provide immunity to their bacterial hosts from similar phages and often allow for over growth of lysogenic bacteria while virulent (lytic) phages lead to a decrease in susceptible bacterial populations [10]. Virulent phages promote bacterial diversity and the coexistence of competing bacteria [11, 12]. Additionally, the lysis of bacterial cells allows for the recycling of nutrients and the release of fixed carbon[13].

In addition to controlling bacterial populations, temperate phages represent the vast majority of lateral genetic material movement, transferring host derived genes between hosts and playing an important role in the evolution of bacteria [14]. Comparison of genomes suggests high rates of recombination; the observed recombination boundaries often correlate directly with gene boundaries [15]. Temperate phage genomes are typically highly mosaic in architecture. Virulent phage typically display a more linear ancestry and have defined areas or regions of recombination. Therefore, any particular phage can possess genes or groups of genes from diverse evolutionary lineages [16].

The majority of isolated phages show limited host ranges while a few have evolved elaborate mechanisms to allow for productive infection in a variety of host strains [17]. There are indications that hard-to-infect bacteria are more likely to be infected by generalist phages with broad host ranges while bacteria that are easy to infect are targeted by specialist phages [18]. Phage host-attachment proteins display a huge diversity in specificity that reflects the adaptation of phages to their target hosts, fostered in some cases by particularly high recombination rates in regions involved in host attachment [16, 18].

At a minimum, productive phage infection requires: 1. successful host attachment, 2. initial penetration of the host outer membrane and/or cell wall, 3. transfer of the phage DNA through the inner membrane leading to synthesis of phage- encoded proteins and nucleic acid, and 4. escape of progeny phage from the dead host cell. Host attachment is mediated by specialized proteins called tail spikes and tail fibers presented on the posterior tip of tailed phages. These proteins attach to specific structures on the outer surface of host bacteria. Lipopolysaccharides (O-antigens), flagellin, teichoic acids, capsular polysaccharides (K-antigens), and specific membrane proteins such as nutrient transporters can be used as

attachment sites [19]. The ability of a phage to attach to specific cell surface molecules is considered the primary determinant of host range. There are other physiological or genetic properties of cells that influence the ability of certain phage to replicate in them, for example the presence of a restriction enzyme or a CRISPR system, and these can in principle influence host range. However, for this study I have assumed that the binding interaction between the phage's host interaction protein and the cell's receptor is the primary and therefore most important component of determining host range. An argument supporting this view is detailed later.

The best studied host attachment protein is the tail spike of *Salmonella* phage P22 (P22TSP). The cell surface receptor of the P22TSP was shown to be the O-antigen of *Salmonella enteric* subsp. *enterica* , serovar Typhimurium [20]. Further investigation demonstrated that P22TSP recognizes the O-antigen α-D-Gal-(1-2)-α-D-Man-(1-4)-α-L-Rha-(1-3) repeats and cleaves the 3,6-dideoxyhexoses[21]. Each phage particle carries multiple homotrimeric tail spikes allowing for multivalent attachment and essentially irreversible binding [22]. O-antigen binding is much faster than hydrolysis suggesting the primary importance of binding the phage particle to the bacterial surface [23, 24].

P22 host range is primarily determined by its single tail spike protein. I have determined that the *Salmonella* phage Det7 used in this study is a Viunalike, differing from P22 in that it possesses a contractile tail similar to many of the T4 like phages but carrying multiple different tail spike proteins that are more or less similar to those of P22 like phages. I hypothesize that each specific Det7 tail spike protein is responsible for the phage's ability to infect a particular subset of hosts. To test this hypothesis, I have biochemically characterized the Det7 tail protein components and compared them to the well- studied protein components of the T4 contractile tail. Additionally, I have determined the host range of Det7 and those of related *Viunalikeviruses*

and compared their genomes using Phamerator [25]. Tail spike proteins of Det7 have been compared with homologous proteins present in *Viunalikeviruses* using bioinfomatic characterization. This investigation provides the ability to determine the host specificities of individual tail spike proteins. This project provides insight into the function of these tail proteins and supports the hypothesis of evolution by recombination as the primary mechanism in adaptation of their host ranges.

1.1 GENOME SEQUENCE DETERMINATION AND ANALYSIS

Det7 phage particles were purified using CsCl continuous gradients. A band of light scattering phage particles observed at a density of 1.503 g/ml was extracted. CsCl was dialyzed from the samples, phage nucleic acid was extracted, and the sequence of the phage genome was determined by Sanger sequencing. It was assembled as a circle using phredPhrap and Consed, but there was no evidence in the data for a terminal redundancy, which argues for circular permutation among the sequences of packaged DNA molecules. The genome contains 157,499 bps. DNAMaster software was used to annotate the genome. The genome contains 207 open reading frames (ORFs), apparently encoding protein, and five tRNA genes. The circular genome was linearized so that it was coterminous with similar phage genomes that had been previously annotated. This is consistent with its relationship to the T4superfamily of phages, as discussed below.

Probable structural genes are identified in most cases as genes encoding proteins with high sequence similarity to proteins that have been shown to be structural components of other phages, often in the T4 superfamily [26]. A total of forty four gene products have been identified through N-terminal sequencing and mass spectrometry as probable parts of the mature virion, as discussed below. Most of the structure-related genes are grouped together (gp1-gp40). Several additional probable structural genes are scattered throughout the genome, including six genes that putatively encode parts of the baseplate. Despite confirmation of nearly 50% of the predicted structural genes, and many other gene protein functions, the majority of open reading frames produce proteins of unknown function, as is common for large phage genomes. Roughly 25% of the coding regions in the Det7 genome can be confidently identified, leaving almost 150 unidentified protein encoding ORFs.

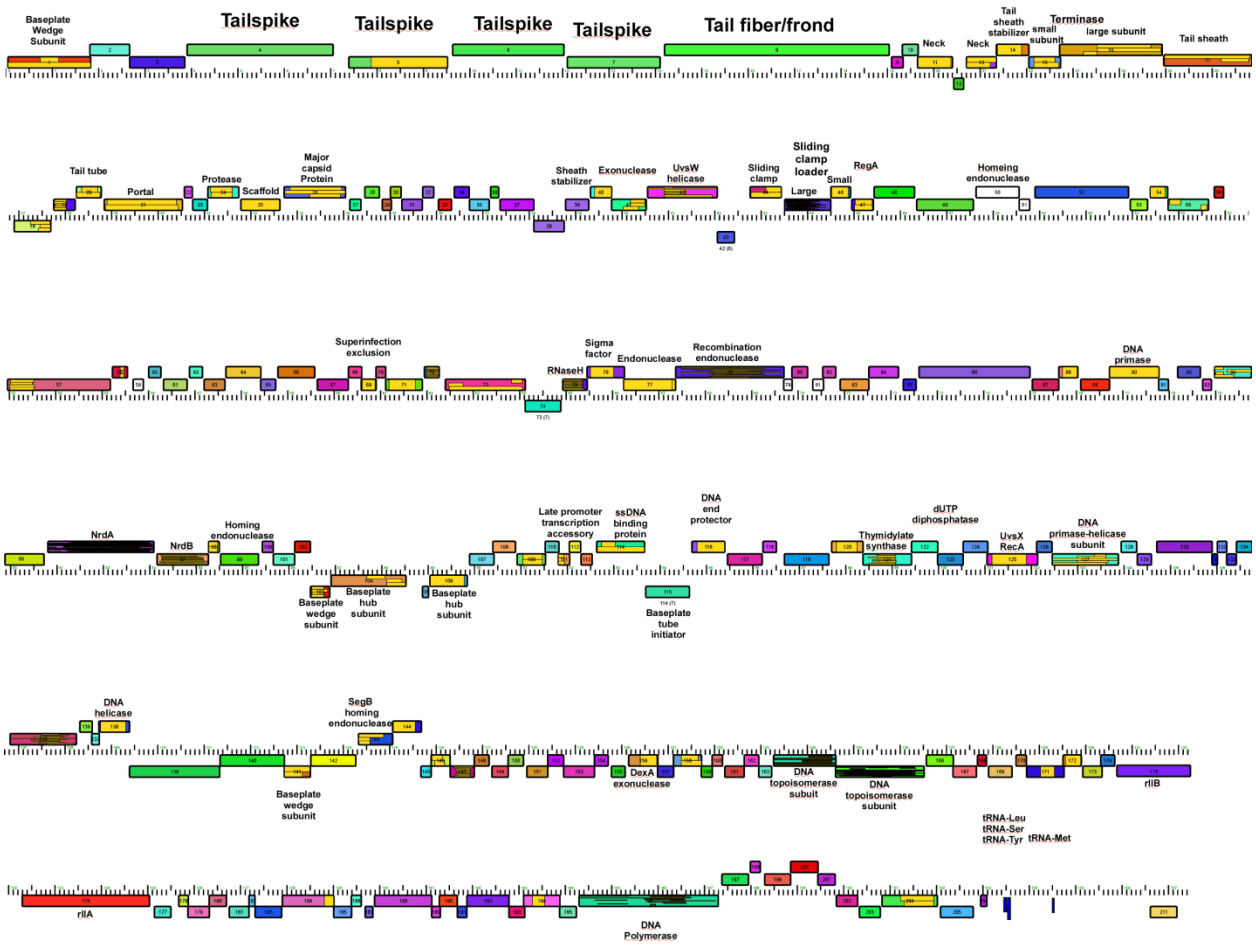


Figure 1. Det7 genomic map. Putative genes were identified from the completed Det7 sequence using GLIMMER, GeneMark and DNA Master. Hypothetical functions of encoded proteins were determined through comparison of amino acid sequences to the non-redundant databank using BLASTP. Genes are color-coded according to putative pham as determined by Phamerator when compared to a *Viunalikevirus* database.

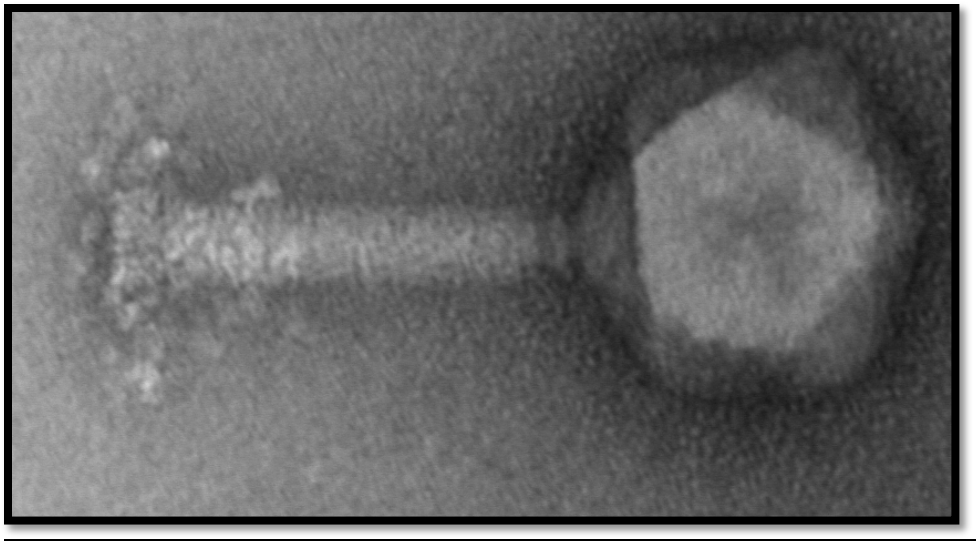


Figure 2.

Electron micrographs of *Salmonella* phage Det7. The phage was applied to a glow-discharged carbon/Formavar-coated 200-mesh copper grid and then stained with 5 % uranyl acetate. The grid was finally examined on a 120 KV Morgani BioTwin transmission electron microscope fitted with a Zeitz (CCD) camera

The Det7 phage particle contains an icosahedral capsid with T=16 symmetry and a diameter of 90nm as determined from CryoEM reconstruction (James Conway, personal communication). T=16 symmetry predicts that the mature capsid of Det7 consist of 955 copies of gp26, the major capsid protein, which was identified by nucleotide similarity to other phage major capsid proteins. Gp26 was confirmed as the major virion protein by SDS PAGE and N-terminal sequencing analysis. Protein sequencing revealed that in the mature Det7 virion particle the N-terminal 50 aa are cleaved from the major capsid protein (MCP). The MCP encoded by the

genome has a predicted molecular weight of 48.17kDa while the cleaved MCP runs true to its predicted weight of 42.4kDa on a 12% SDS PAGE (figure 4).

The uncontracted tail of Det7 is estimated to be 120nm long. The tail terminates in a large baseplate structure with “mop”-like appendages protruding from it (Figure 3). This large ornate baseplate structure is a defining element of the *Viunalikeviruses* [26].

The protein constituents of Det7 virion particles and purified tails have been assessed by N-terminal sequencing of protein bands from SDS-PAGE. Mass spectrometric analysis was performed on purified tails run into a single band extracted from SDS-PAGE. All major predicted structural proteins were identified by mass spectrometry, including several of unknown function (Table 1). N-terminal sequencing was able to confirm the presence of two tail spikes (gp4 and gp6) as well as the predicted tail fiber gp8. All four predicted tail spikes were identified from mass spectrometry with multiple peptides to each predicted protein. Additionally, the ability to identify tail spike proteins by N-terminal sequencing and high PSM (peptide Spectrum match) scores suggest that each protein is present in multiple copies per virion.

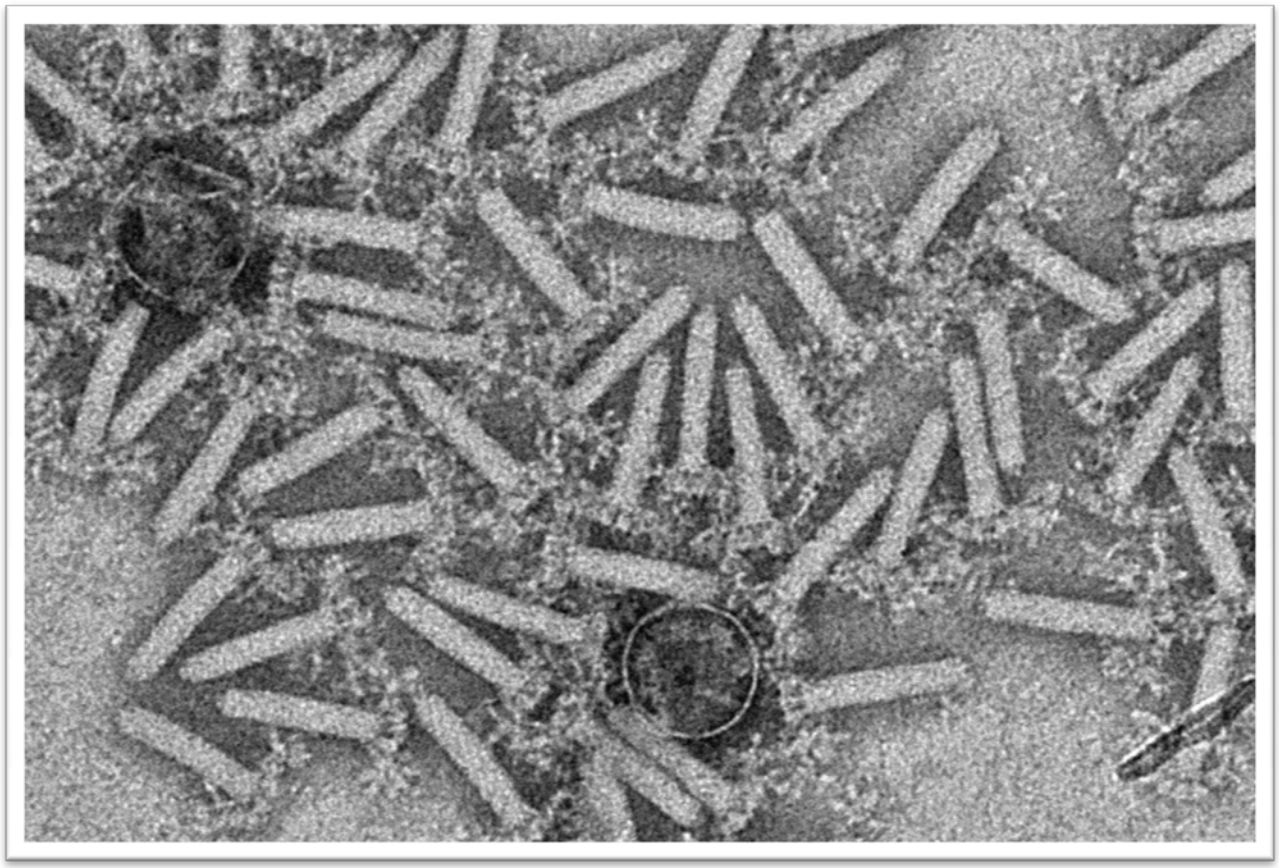
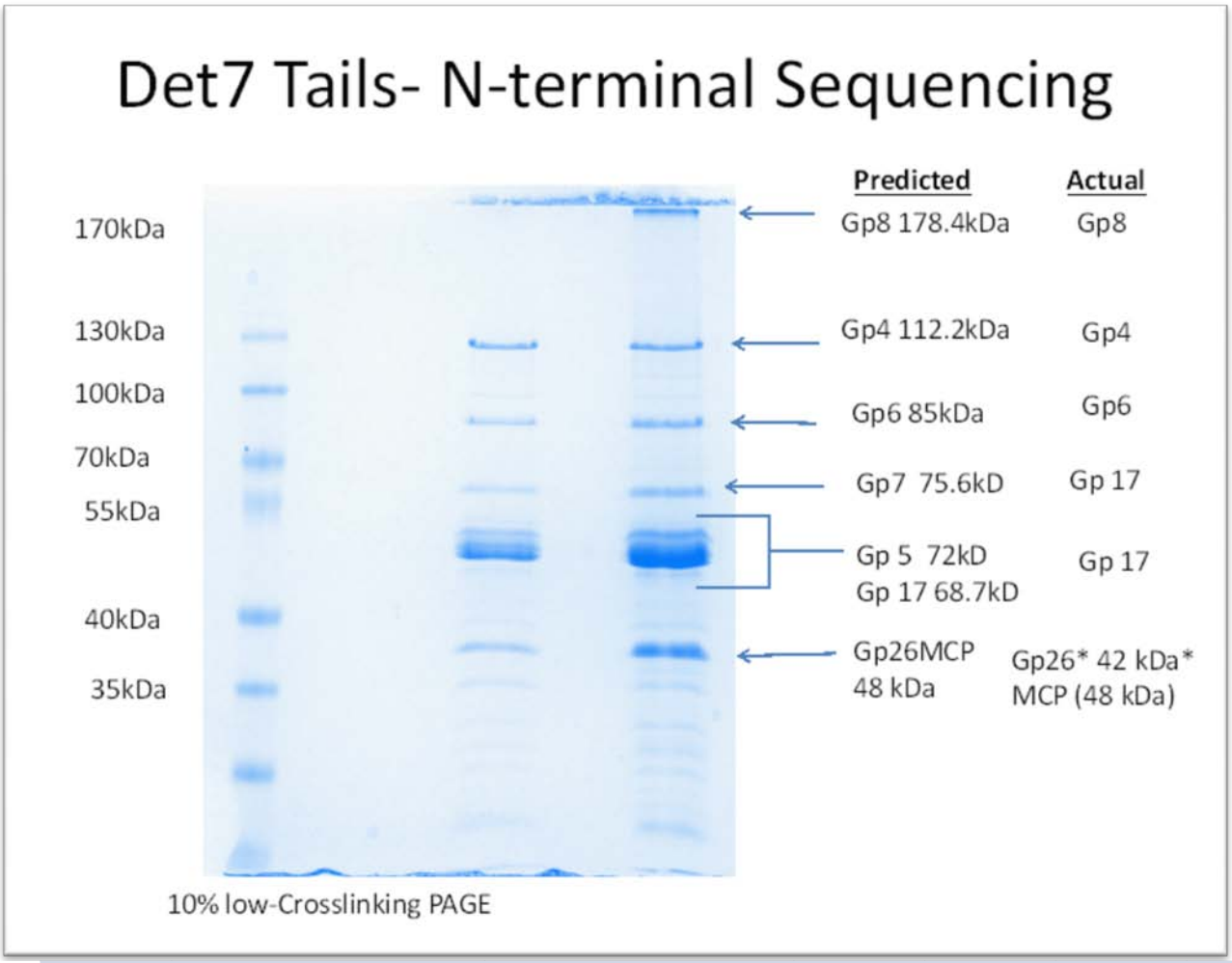


Figure 3

Electron micrographs of semi-purified tails of *Salmonella* phage Det7 used for mass Spectrometry analysis. The phage tails was applied to a glow-discharged carbon/Formavar-coated 200-



	Predicted	Actual				
gp4	MANKPTQPLF	MANxPtqPLF				
gp5	MISQFNQPRG	xxxxx				
gp6	MTRNVESIFG	MtRNVEsIFG				
gp7	MKPQFSQPKG	xxxxx				
gp8	MQSTNLNR	MqstNLNR				
gp17	MATQSFSVAP	MATqsFsVAP				
gp26		sDAP				

Figure 4. Predicted and actual results from N-terminal sequencing of purified Det7 phage. N-terminal sequence for gp5 and gp7 could not be determined due to bleed over from gp17 band. Results predicted from annotated genome and calculated molecular weights of proteins.

Coverage	PSMs	Peptides	# AAs	MW [kDa]	calc. pI	Score	Description	
69.89	39	35	31	8.3	.03	647.75	gp 17	tail sheath
54.84	71	64	612	77.7	.10	098.99	gp 8	tail fiber
51.81	42	35	048	11.7	.84	32.04	gp 4	tail spike
50.28	34	25	08	5.3	.38	73.13	gp 5	tail spike
58.41	02	25	66	1.6	.03	75.32	gp 7	tail spike
47.46	3	6	77	9.9	.88	25.99	gp 20	tail tube
75.00	7	19	40	8.0	.43	04.50	gp 26	MCP
58.84	4	19	96	2.4	.11	68.72	gp 3	hypothetical
46.37	2	22	93	5.5	.69	34.31	gp 17	baseplate wedge subunit
40.40	7	21	46	0.1	.19	18.95	gp 139	hypothetical
51.19	0	33	95	8.2	.83	80.56	gp 86	hypothetical
18.92	7	11	98	4.6	.92	78.28	gp 6	tail spike

2.60	**5**	5	2 1	62	2.6	.35	54.92	140	gp	hypothetical
6.47	**5**	8	1 1	55	8.4	.00	47.50	74	gp	hypothetical
3.26	**5**	7	2 1	37	8.5	.00	40.19	104	gp	baseplate hub subunit and tail lysozyme
6.01	**4**	6	1 3	13	5.0	.98	8.51	115	gp	baseplate tail tube initiator
0.00	**4**	5	7	85	1.5	.86	8.82	141	gp	baseplate wedge subunit
1.93	**3**	2	1 4	48	7.3	.94	6.27	ntaminant	co	groEL
3.23	**3**	0	8	22	6.1	.92	5.23	142	gp	putative tail tube associated base plate protein
8.79	**3**	1	7	32	6.9	.45	4.49	14	gp	tail sheath stabilizer
5.54	**3**	2	5	42	5.5	.54	5.19	37	gp	hypothetical
0.77	**2**	1	3	84	3.0	.86	7.68	2	gp	hypothetical
3.21	**2**	2	1 0	60	3.1	.10	1.64	21	gp	portal
2.23	**3**	1	9	94	3.3	.45	4.64	ntaminant	co	EF-Tu
5.61	**1**		4	69	7.9	.74	5.43	171	gp	putative tail fiber- ig like
2.12	**3**		3	65	8.5	.60	2.14	39	gp	tail completion and sheath stabilizer

Table 1. Details of mass spectrometric analysis of Det7 semi-purified tails. Top 26 hits are displayed. Full table in supplemental material. All predicted structural proteins were detected.

1.2 DET7 CONSERVED TAILSPIKE MODULE (GP5)

Analysis of the Det7 genome reveals five genes that likely encode tail spikes or tail fibers (Figure 1). Det7 was initially identified through sequence comparison as possessing a tail spike that shared similarity in the C-terminal portion with that of *Salmonella* phage P22. Det7 gp5 and P22 gp9 (P22TSP) share more than 60% aa identity in the 550 C-terminal residues. The structure of Det7 gp5 has been solved by x-ray crystallography and described [5] as being highly similar structurally to the phage P22 tail spike, gp9, and the 9NA TSP. In fact, the crystal structures only differ by 1.5 Å on average, and the active sites are identical between all 3 tail spikes (Figure 5). Antibodies against P22 gp9 as well as mass spectrometry were used by Walter *et al.* to identify Det7 gp5 as a virion structural protein. Comparison of the crystal structures (Det7-2V5I and P22- 3TH0) and ClustalX2 alignment now reveals conservation of the residues interacting with the carbohydrate ligand. Specifically, the residues that specify the O-antigen hydrolysis function in P22 gp9 and Det7 gp5 are absolutely conserved (Figure 5).

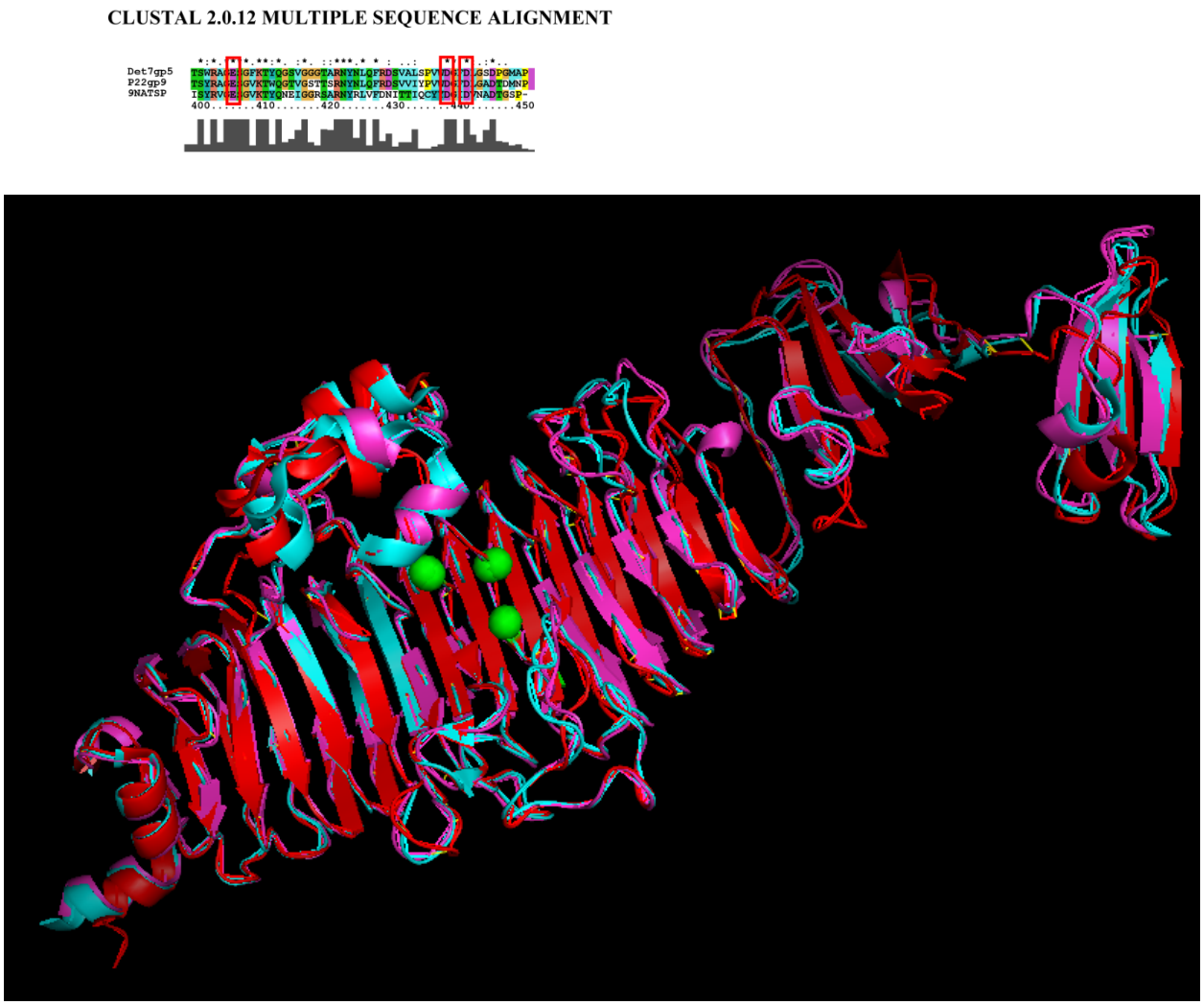


Figure 5. A. Clustal X alignment of Det7gp5, P22TSP (gp9), and 9NATSP. B. Comparison of the crystal structures of Det7-2V5I, P22-3TH0, and 9NA-3RIQ. Alignment of the Cα backbone shown R.M.S.D of <1.5Å. Pymol was used to perform alignments and R.M.S.D calculations

1.2.1 Det7 host range

Salmonella enterica has over 2500 serovars as defined by the World Health Organization Collaberating center for Refence and Research on *Salmonella* (WHOCC-Salm)[27]. Each of these has, by definition, a different antigenic specificity[28]. The O-antigen is the primary determinant of antigenicity of *Salmonella* and serves to disguise the *Salmonella* from immune responses [29]. O-antigens are also common attachment molecules for phages targeting *Salmonella.* A given serovar of *Salmonella* displays only a single type of O-antigen and the large variety of possible O-antigens suggests that phage will only infect a small subset of serovars that share a similar O-antigen, unless the phage possess multiple different LPS attachment proteins and/or the adhesion region has the ability to recognize multiple different structures, as with the case of T4, which recognizes OmpC in K12 strains, but B specific lipoplysaccharide on *E.coli* B, which has no OmpC.

Det7 host range was tested using *Salmonella* reference collections A, B, and C (SARA, SARB, SARC). SARA is a collection of 72 strains representing a variety of hosts and environmental samples that is designed to represent the full range of genotypic variation in *Salmonella enterica,* ssp. Typhimurium[30]. SARB contains 72 strains which represent 37 serovars of subspecies 1 and was designed to be used for research on genetic and phenotypic variation in natural *Salmonella* populations [31]. The genetic breadth of *Salmonella* is represented by the 16 strains with diverse numerical taxonomic parameters that compose SARC.

These strains represent previously described taxonomic groups of *Salmonella*[32] and epitomize the depth and breadth of *Salmonella* in all its glorious numerical, phonetic, and cladistic taxonomic variety. Det7, P22, and 9NA were tested on each reference strain in these collections.

Strain susceptibility to phage Det7 infection was tested using triplicated spot plaque assays. The titer was determined on each strain to compare Det7's relative efficiency of plating (eop). Det7 was found to plate with equal efficiency on 67% of SARA, 54% of SARB and 37% of SARC, representing a combined susceptibility of approximately 60% of *Salmonella* strains. Det7 was found to infect every P22-susceptible strain of *Salmonella*. Additionally, Det7 was also able to infect a large number of additional *Salmonella* serovars strongly suggesting use of multiple tailspikes for host attachment (Fig 6).

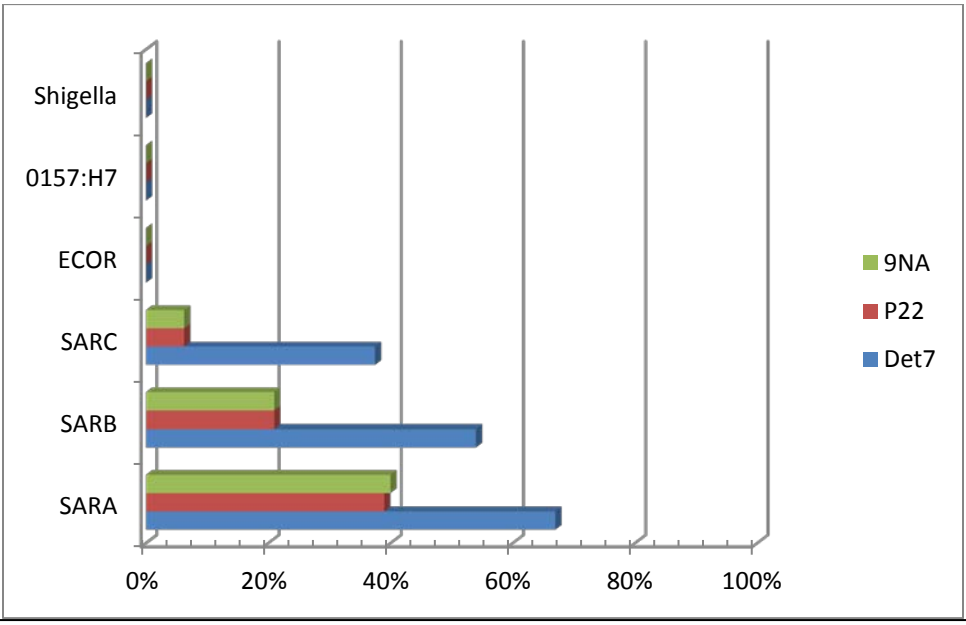


Figure 6. Host range summery of Det7, P22 and 9NA displayed as a percentage of the total number of serovars and strains tested. All three phage are exclusive for *Salmonella.*

1.2.2 Host range of Det7 compared to host range of P22 and 9NA

It is highly probable that Det7 gp5 binds and cleaves the same 1,4,12 O-antigen as the P22 tailspike based on the structural similarity of their active sites. To test whether Det7 was dependent upon the O-antigen, Det7 was tested for its ability to infect a mutant strain of *Salmonella* LT2 lacking any O-antigen. The rfbB deletion mutant (KAB37- provided by K. Butela) is unable to process the O-antigen sugars and therefore is unable to display the 1,4,12 O-antigen. Det7 was able to infect the LT2 parent strain but showed no ability to infect the naked KAB37 strain, demonstrating Det7's dependence upon O-antigen binding and specific use of the 1,4,12 O-antigen.

The host range of P22 (*Podoviridae*) and 9NA (*Siphoviridae*) was tested along with Det7 on *Salmonella* reference collections A, B, and C. The tailspikes of P22 and 9NA share 36% overall identity and have previously been described as infecting the same *Salmonella enterica* ssp. host strains and producing identical LPS cleavage products [16, 22, 33]. Det7 infected every serovar that P22 or 9NA infected and uniquely infected a variety of other serovars, which represent a wide variety of O-antigens (Figure 6).

1.2.3 Det7 similarity to viunalike viruses

Sequence comparison of Det7 to the non-redundant genome database reveals that Det7 shares similarity with members of the recently identified *Viunalikevirus* genus. As noted by Adriaenssens *et al.*, all the known members of this genus share similarities in virion morphology, gene regulatory elements, tRNA's, and DNA modification enzymes [26]. Det7 shares between 60 and 90% nucleotide identity with individual members of the genus. Pairwise comparison shows that Det7 shares the greatest degree of similarity (~85%) with *Viunalikevirus Salmonella*-tropic phages Vi1, ΦSH19, and SFP10. All members of the genus have genomes of approximately 157 kbps and share identical gene organization. All known members of this genus share similarities in virion morphology, gene regulatory elements, encoded tRNA complement, and an encoded hydroxymethyluracil transferase [26].

Despite the overall similarity in genome organization among viunalieviruses, the approximately15kb portion of the genome encoding the tail spikes and tail fiber genes diverges strikingly. The number of proteins is somewhat variable, with Phax1 possessing 3 tail spike proteins while LIMEstone1 has only 1 likely tail spike protein. All other known members of the genus have 4 probable tail spikes and share a highly conserved tail fiber.

The N-terminal portion of syntenous tail spike proteins show sequence similarity despite divergence in the C-terminal domains. These conserved N-terminal domains align with the head-binding domain identified in P22 TSP [34, 35]. The enzymatic activity of the tail spikes was demonstrated to reside in the C-terminal portion and be independent of the N-terminal capsid/baseplate binding domain in phage P22 [36].

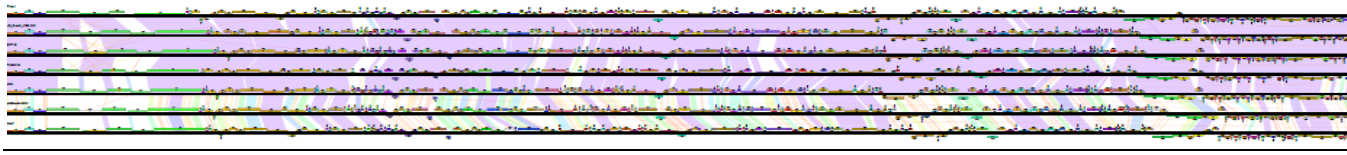


Figure 7. Phamerator database was constructed for the *Viunalikevurus* genus. Similarity on gene organization and overall nucleotide identity displayed suggesting evolution by linear descent. Area of recombination seen from dissimilarity in nucleotide identity encompassing tail spike genes.

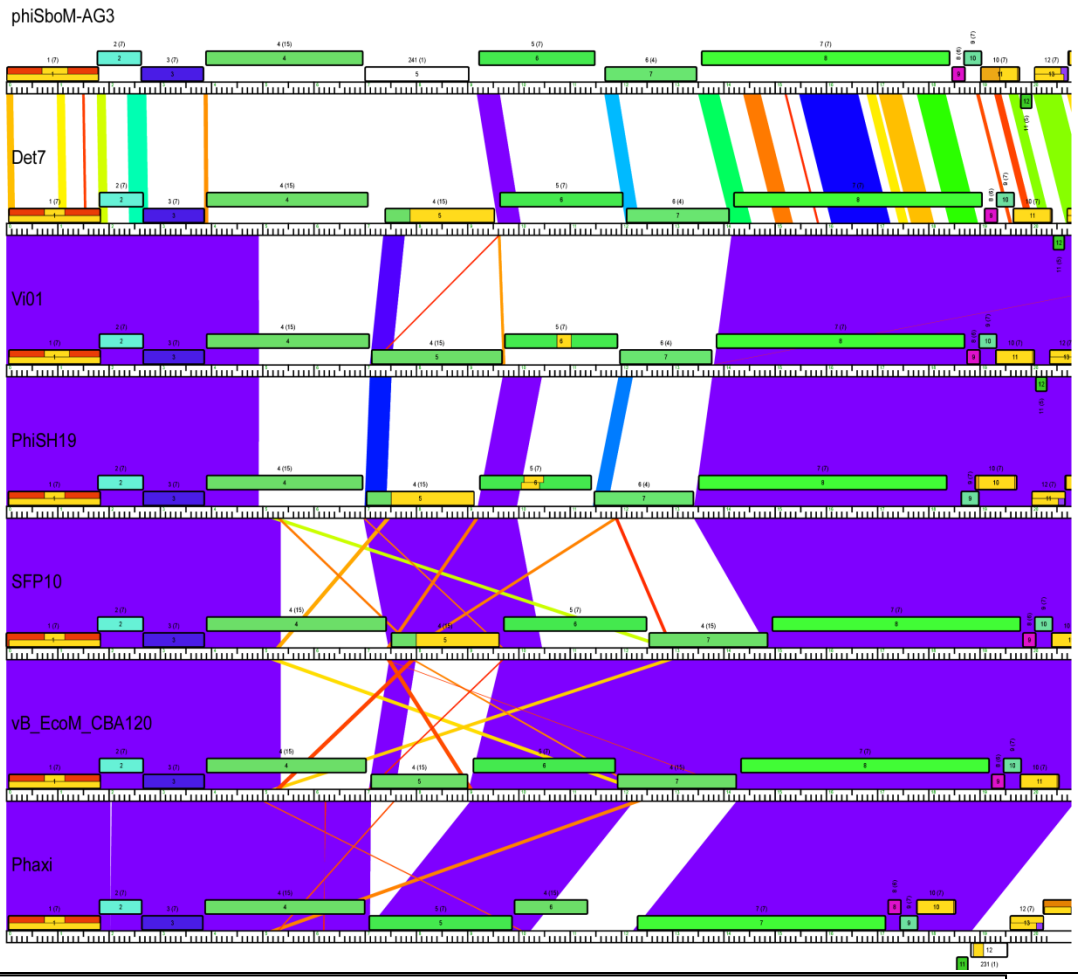


Figure 8. Phamerator map displaying area of recombination between *Viunalikevirus* genus. Gene organization is conserved despite divergence in gene products.

1.2.4 Tailspikes are primary host recognition systems

The *Salmonella* O-antigen has been shown to bind to the P22 TSP and the active site of the P22 TSP has been determined [21, 22, 24, 37]. P22 is dependent on the O-antigen for initiation of infection and unable to infect host strains that do not contain O-antigen. Walter *et al.* previously identified Det7 gp5 as being P22-like, and, upon crystallization, were able to show

that it contained the same active site [5]. As discussed above, it is highly probable that Det7 gp5 cleaves the same 1,4,12 O-antigen as the P22 tail spike based on the structural similarity of the active site. To test this possibility, Det7 was tested for its ability to infect a mutant strain of *Salmonella* LT2 lacking any O-antigen. An in-frame *rfbB* deletion mutant, KAB37, (provided by K. Butela) is unable to process and display the 1,4,12 O-antigen. Det7 was able to infect the LT2 parent strain but not KAB37, demonstrating Det7's dependence upon binding the 1,4,12 O-antigen.

1.2.5 Difference in tailspikes allows for divergence in host range

The viunalikeviruses share the trait of diversity among multiple tail spike genes, which presumably is the basis for their divergent but sometimes overlapping host ranges. Due to Det7's similarity to the described viunalikeviruses we tested its ability to infect several other enteric bacteria. Det7 was unable to infect any of the ECOR or tested O157:H7 strains, unlike the closely related viunalikeviruses SFP10, CBA6 and CBA120. SFP10 and CBA120 share two pairs of nearly identical tail spikes (orf161/Det7 gp6 and orf160/ Det7 gp7) one or both of which is therefore predicted to be confer infectivity on *Escherichia coli* O157:H7. The SFP10 host range also includes P22 susceptible *Salmonella* strains. SFP10 orf162/our gp5 is homologous to the P22 TSP and Det7 gp5, suggesting that this tailspike is responsible for SFP10's ability to infect P22 susceptible *Salmonella* serovars. Another viunalikevirus, phage PhiSboM-AG3, infects several strains of *Shigella*, suggesting them as possible hosts for Det7. Det7 was unable to infect a single *Shigella* strain tested, but the newly sequenced viunalikevirus CBA6 was able

to infect 10 of 23 strains (Hany Anany, personal communication). It worked especially well on *Shigella* flexnerei and sonnei, but not those strains hit by PhiSboM-AG3.

2.0 SECOND CHAPTER

2.1 DET7 TAIL STRUCTURE

Det7 has a contractile tail approximately 120 nm long. The tail shares many morphological characteristics with the well-studied contractile tail of T4. The proximal end of the tail is joined to the capsid by a neck surrounded by a collar much like T4. The tail tube is surrounded by a sheath with 24 transverse striations reminiscent of T4 [26]. The baseplate is positioned on the distal end of the tail and is the attachment site for the various tail spike and tail fiber proteins. Cryo-electron microscopy (Cryo-EM) modeling of the tail has confirmed six-fold symmetry of the tail sheath and baseplate (Alexis Huet, personal communication). Negatively stained virions captured in transmission electron micrographs display several different baseplate conformations that are similar to conformational changes in the T4 tail baseplate [38]. The tails and capsids of other *Viunalikevirus* members have been described by Adriaenssens *et al.* including the probable dimensions of each structural element [26].

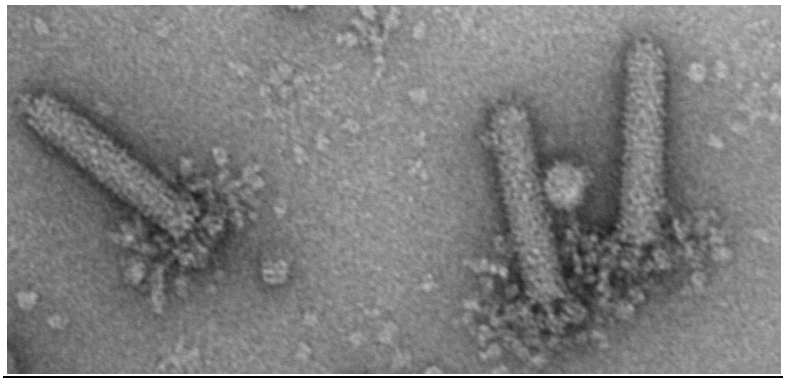


Figure 9. Electron micrographs of *Salmonella* phage Det7 BioCAD purified tails. The phage was applied to a glow-discharged carbon/Formavar-coated 200-mesh copper grid and then stained with 5 % Urinal acitate. The grid was finally examined on a 120 KV Morgani BioTwin transmission electron microscope fitted with a Tietz F415 charge-coupled-device (CCD) TemCam camera

2.2 SIMILARITY IN STRUCUTRE OF ALL CONTRACTILE TAILS

T4 has also been used as one of the archetype phages (others being T7, P22 and Lambda). Much of the structural detail of both the capsid and the tail has been determined for T4. All contractile tail phages share a great deal of structural similarity with the T4 tail. Det7 and the other *viunalikeviruses* are myoviruses that share much structural similarity to T4. Of the 19 known T4 gene products that comprise the T4 tail, 12 of them can be unambiguously identified in Det7. The major tail protein components, the tail sheath and tail tube, are identifiable in whole cell lysate as well as purified phage proteins run on SDS PAGE by both mass

spectrometry and N-terminal sequencing. These techniques have confirmed over 50% of T4 homologs identified in Det7.

The T4 tail assembly pathway is well studied and many of the tail components have had their structures determined by x-ray crystalography [38, 39]. The T4 tail components can be separated into at least three groups based on order of assembly. Six baseplate wedges assemble around a central hub to form the T4 baseplate. Each baseplate wedge is comprised of seven distinct proteins. Three of these proteins have homologs in Det7. The T4 central hub is composed of two proteins (gp5 and gp27) that can be identified in the Det7 genome based on synteny but demonstrate little sequence homology (Figure 1 and supplementary genome annotation of Det7). The tail tube proteins attach to the baseplate complex and tail sheath protein polymerizes around the tube. The tail tube and sheath are stabilized at the capsid end by tail completion protein (T4 gp3 & gp15) and neck proteins (T4 gp13 & gp14) that all share homology with Det7 (gp39 &gp14 and gp11 & gp13, respectively).

The Det7 and T4 contractile tails share overall similarity in structure except in components distal and peripheral to the baseplate. Det7 does not contain any detectable homolog to T4 gp7,9,11, and 12. T4 gp9 is on the periphery of the T4 baseplate and connects gp34 (T4 long fiber). Tail spikes fulfill distinctly different roles from T4 tail fibers in host recognition. The T4 long tail fibers bind to the host OmpC receptor. This has been described previously as the phage “walking” on the cell surface. Recognition of the receptor initiates contraction of the tail sheath [40]. In contrast, the tail spikes of Det7 bind and cleave the O-antigen and do not allow for migration on the cell surface [23, 24]. The tail spike proteins of Det7 will be described further in following section.

T4	Det7	Description	Crystal strucutre	Det7 n-terminal	Det7 - mass spec
25	104	baseplate wedge subunit			+
26	106	baseplate hub subunit			
5	105	baseplate hub subunit and tail lysozyme			
53	141	baseplate wedge subunit			+
6	1	baseplate wedge subunit	3H2T		+
13	11	neck protein			+
14	13	neck protein			
15	14	proximal tail sheath stabilization	4HUD		+
18	17	tail sheath monomer	3FOA , 3FO8	+	+

Table2. Homologous proteins of Det7 and T4 as determined by nucleotide and amino acid similarity. N-terminal sequencing and Mass Spectrometry results for Det7 homologous are displayed.

2.2.1 Identified Tail proteins from Mass spec and N-terminal sequencing

Det7 tails were purified and concentrated from lysates derived from infections of wild type *Salmonella enterica* by PEG precipitation and sucrose gradient sedimentation. Concentrated tails were then further purified using the BioCad profusion chromatography system. Concentrated and purified tails were analyzed using SDS-PAGE. SDS-PAGE bands that match the expected molecular weights of the potential tail spikes were observed as well as several other structural proteins. N-terminal protein sequencing (Edman degradation) performed by Dr. John Hempel confirmed the major capsid protein (with the first 51 AA cleaved off) and the tail sheath at bands corresponding to 42kDa and 68.7kDa, respectively. Additional N-terminal sequencing demonstrated that the three high molecular weight bands observed by SDS PAGE are gp8 (178kDa), gp4 (112kDa), and gp6 (85kDa). N-terminal sequencing was not possible for SDS-PAGE bands of gp5 (75kDa) and gp7 (72kDa) due to their proximity to the large band of tail sheath protein gp17 (68kDa) within a polyacrylamide gel.

Liquid chromatography and tandem mass spectrometry were performed on purified tails to determine the full complement of detectable Det7 tail proteins. Trypsin digested peptide fragments matching all putative tail proteins were identified in addition to peptides matching 14 proteins annotated with unknown functions. Peptides from each of the four tail spikes were identified in high abundance. Tail fiber protein gp8 (178 kDa) accounted for the second largest

number of recruited peptide fragments, with the largest number of peptide fragments recuiting to the tail sheath.

2.2.2 Purified tails denatured

Complete Det7 tails were isolated from wild type infections and purified. A BioCAD sprint high pressure liquid chromatography was used to further purify tails and remove residual contaminants. Det7 virion particles were purified from continuous Cesium Chloride gradients. The purity of the phage particles was verified with SDS-PAGE. Highly concentrated samples of purified phage virions and tails were then subjected to different concentrations of the denaturants -Guanidine HCL and Urea. Purified phage samples were titered to determine the point at which the particles lost infectivity or DNA. Concentration of guanidine HCL at 0.5M and Urea at 1M and lower showed no decrease in titer on LT2 or SARB2 host strains. Pellet and supernatant samples were run by SDS-PAGE to determine which proteins were stripped from the particles first. Tail spikes gp5, gp6, and gp7 are partially stripped from the particle first in approximately equal proportions, though they are not fully removed, as stronger bands are still observed in the pellet than in the supernatant samples at that level. Gp4 and Gp8 were also partially removed before a drop in titer was detected. The removal of these tailspike proteins demonstrates that they are in solvent exposed positions. Additionally, incomplete removal without significant loss of titer demonstrates that multiple copies of each protein are present on each virion particle and that a full complement is not a prerequisite for successful phage infection.

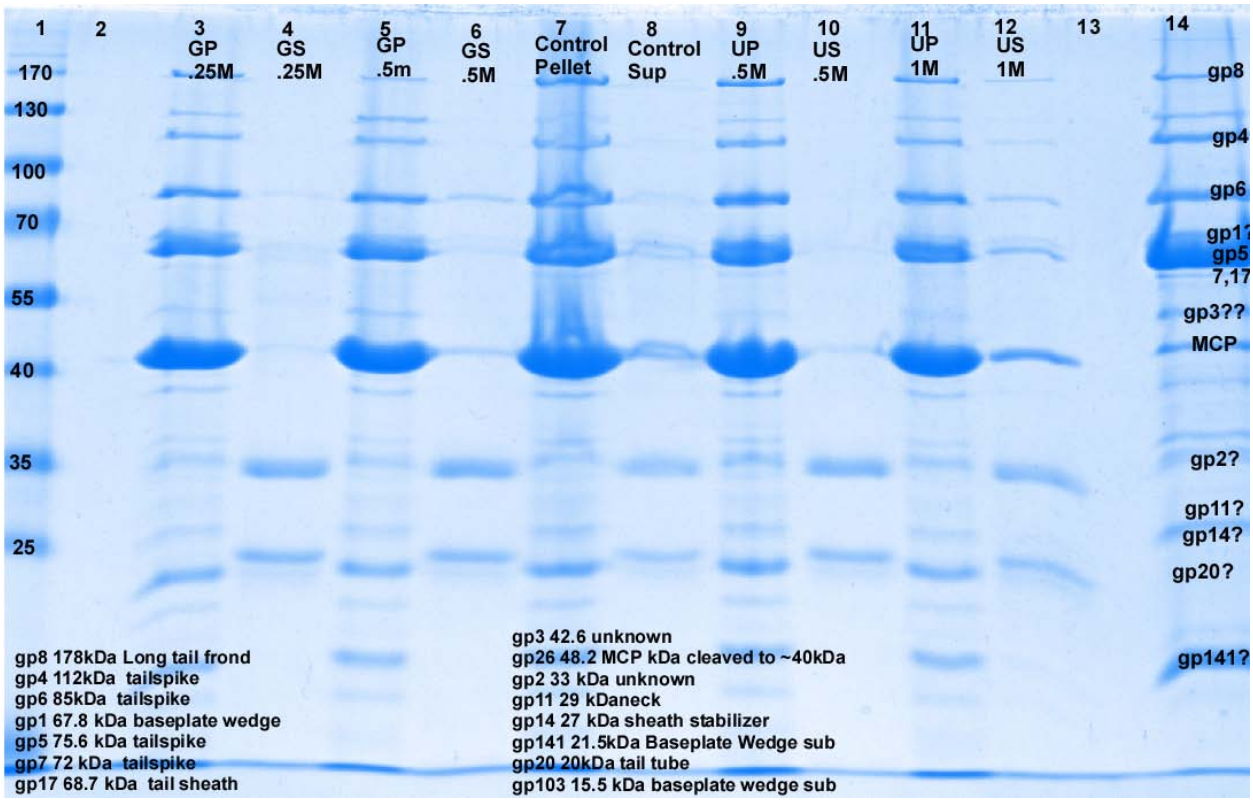


Figure 10. Det7 virion particles denatured with Guanidine and Urea for one hour at room temperature. Samples were centrifuged to pellet intact particles. Pellet (GP- Guanidine Pellet UP- Urea Pellet) and supernatant fractions (GS- Guanidine Supernatant UP- Urea Supernatant) were then run on SDS-PAGE. Bands corresponding to gp4-8 appear in the supernatant fraction indicating they are partially removed before a drop in titer is detected.

2.2.3 Bioinformatic characterization of Det7 gp4, gp5, gp6, gp7, gp8

BLAST was used to identify the tailspikes of Det7 through sequence comparison. Modeller 9.12 was used to build high quality homology models of each tail spike [41]. Structural homology was determined using Phyre2 and hhpred [42, 43]. CsBlast, phiBlast, and hhpred were used to further investigate each of the cell adhesion proteins.

Det7 gp5 has been crystallized and previously described as being highly similar to P22 gp9 [5]. Comparison of the crystal structures of Det7-2V5I, P22-3TH0, and 9NA-3RIQ and ClustalX2 alignment reveals conservation of the residues interacting with the carbohydrate ligand. Specifically, the residues that specify the O-antigen hydrolysis function in P22 gp9, 9NA TSP, and Det7 gp5 are absolutely conserved. This structurally well-conserved tail spike has now been observed in all three morphological tail types [5, 22].

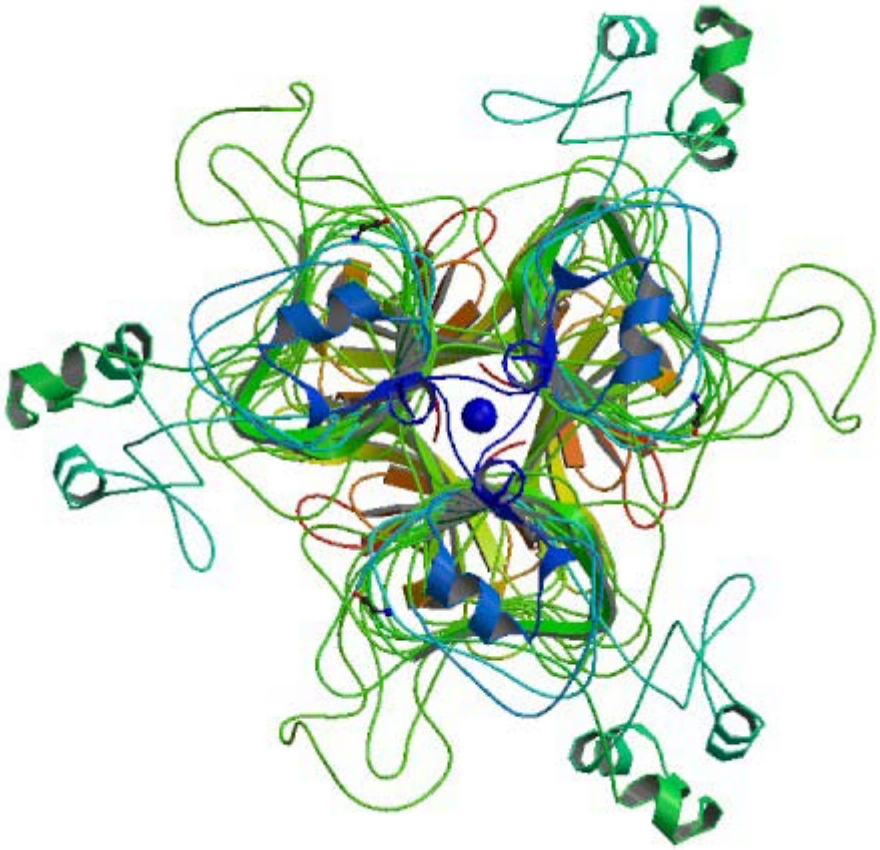


Figure 11. Assumed Biological assembly: Det7 gp5 Crystal structure 2V5I

[5]

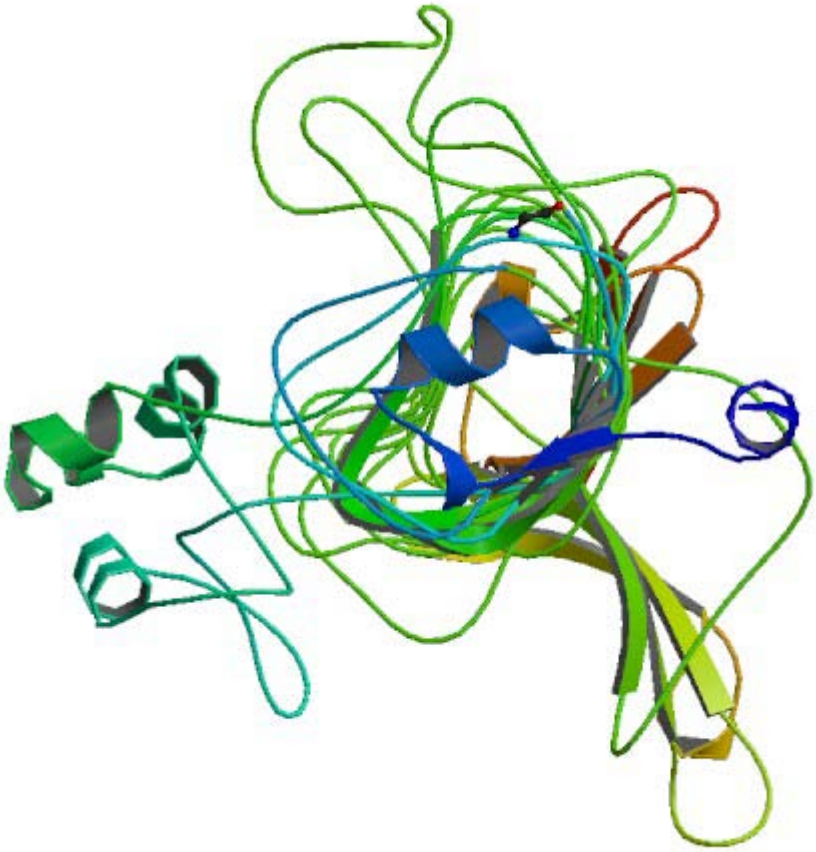


Figure 12. Asymmetric unit: Det7 gp5 Crystal structure 2V5I

[5]

The C-terminal portion of Det7 gp6 shares a high degree of structural similarity with tail spikes of HK620 and SF6 [44]. When compared to SF6 gp14 C-terminal portion, there is 99.5% probability of structural similarity with an e value of 2.9 e^{-14} [45]. Based on this similarity, a homology model of gp6 aa 300-700 was created. The SF6 gp14 (PBD ID: 2vbk_A) and HK620 (PBD IC: 2x6w) tail spikes were used as initial templates for the Det7 gp6 model. The N-terminal portion of Det7 gp6 shares 86% nucleotide identity and greater than 80% similarity with phage PhiSboM-AG3 orf 210/gp6. PhiSboM-AG3 infects *Shigella boydi* and phage SF6 infects *Shigella flexneri*. Despite this similarity, Det7 can not cleave the endo-1,3-α-L-rhamnosidase O-antigen nor infect Shigella like SF6 gp14 [44]. The structural proteins of these phages have been compared and shown to be very similar but mosaic in their structural genes sequences [46].

Figure 13. Homology model of Det7 gp7 made from SF6 gp14 (PBD ID: 2vbk A) and HK620 (PBD IC: 2x6w)

The O-antigen binding region of Det7 gp7 shares similarity with the Epsilon 15 tail spike that cleaves the (3,10 O-antigen) present on SARB2. HHpred search results identified structural similarity between Phi29 and HK620 tail spike crystal structures [45]. Det7 gp5 (PBD ID: 2V51) and Phi29 (PDB ID: 3gq8_A) crystal structures were used to build a high confidence homology model of Det7 gp7 using Modeller 9.12 [41]. Both Det7 gp5 and gp7 were confirmed using mass spectrometry but were unable to be N-terminally sequenced, because of the high signal bleed from the closely migrating band of the gp17 tail sheath in SDS-PAGE. Det7 gp4 and gp6 were confirmed both by N-terminal sequencing as well as mass spectrometry.

Det7 gp5, gp6, and gp7 share highly similar N-terminal domains. The corresponding region of P22 gp9 is designated as the capsid binding domain. Unlike P22 TSP, Det7 tail spikes bind to the baseplate. The similarity in baseplate binding domain and structural similarity of the C-terminal portion based on homology modeling of gp6 and gp7 suggest that these three tailspikes bind nearly identical or identical portions of the baseplate wedge [20, 35]. Further investigation of the N-terminal regions of Det7 gp5, gp6, and gp7 revealed similarity among identically positioned proteins in the other *Viunalikeviruses*.

Gp4 of Det7 is much larger than the other tailspike proteins at 112kDa. The N-terminal portion is conserved among the *viunalikevirus* genus. Sequence and structural comparison of the N-terminal portion reveals a high structural similarity with T4 gp9 despite only 18% amino acid identity. T4 gp9 serves as the attachment site on the T4 baseplate for the short tail fiber. The structural conservation suggests that this region is part of the baseplate structure. The c-terminal portion is identified as a tail spike possessing rhamnosidase activity based on nucleotide similarity to homologous proteins in related phages.

Det7 gp8 is the most highly conserved host recognition gene among the viunalikeviruses. Det7 gp8 is a 178kDa structural protein that has been identified from both N-terminal sequencing and mass spectrometry on purified tails. Due to gp8's size and genome position, it is most likely the long tail fiber frond easily observed in TEM on the periphery of the baseplate. Secondary structural prediction using PSIPRED and Quick2D on Det7 gp8 and suggests that this protein predominantly contains strands and coils with few helices [45, 47, 48]. The structural predictions for gp8 suggest it is not globular-like, but has a defined structure. Sequence analysis using BLAST suggest similarity with many cyanophage structural proteins. CsBlast also reveals the similarity with other general sugar binding proteins.

3.0 MATERIALS AND METHODS

3.1 STRAINS

(Source: Hendrix laboratory strain collection except as noted)

1). *Salmonella* LT2 –SARB3 (Brandenburg O: 1,4,12)
Isolation and propagation host of Det7. Non amber suppressor

2). Salmonella KAB37
LT2 donor strain lacking the rfb locus (Δ*rfbP-rfbB*-2773::*hph*)

3) BL21(DE3)plysS
Escherichia coli F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

4) *Salmonella* UB-0015 Provided by Dr. Sherwood Casjens, University of Utah, Division of Cell Biology and Immunology
(LeuA414, Fels2$^-$, sup^0)

5) *Salmonella* UB-0016 Provided by Dr. Sherwood Casjens, University of Utah, Division of Cell Biology and Immunology
(LeuA414, Fels2$^-$, sup^D(ser))

6) *Salmonella* UB-0017 Provided by Dr. Sherwood Casjens, University of Utah, Division of Cell Biology and Immunology
(LeuA414, Fels2$^-$, sup^E(Gln))

7) *Salmonella* UB-0018 Provided by Dr. Sherwood Casjens, University of Utah, Division of Cell Biology and Immunology
(LeuA414, Fels2$^-$, sup^F(Tyr))

8) *Salmonella* Reference Collection A (SARA) Collection Provided by Dr. Jeffrey Lawrence, University of Pittsburgh [30]

9) *Salmonella* Reference Collection B (SARB) Collection Provided by Dr. Jeffrey Lawrence, University of Pittsburgh [30, 31]

10) *Salmonella* Reference Collection C (SARC) Collection Provided by Dr. Jeffrey Lawrence, University of Pittsburgh [30]

11) *Ecoli.* Reference Collection (ECOR) Collection Provided by Dr. Jeffrey Lawrence, University of Pittsburgh [49]

12) *Shigella* and *Ecoli* 0157:H7 Collection listed in host range table. Strain susceptibility to Det7 infection conducted by Dr. Hany Anany of Canadian Research Institute for Food Safety, Food Science Dep., University of Guelph [50]

3.2 MEDIA AND BUFFERS

1) Luria Broth (LB)
   1% (w/v) tryptone (Difco), 0.5% (w/v) yeast extract (Difco) and 0.5% (w/v) NaCl in ddH2O, autoclaved for 25min. Salts, antibiotics and extra nutrients were added before using.
2) LB agar
   LB plus 1.5% (w/v) agar, autoclaved for 25min.
3) Soft agar
   1% (w/v) tryptone, 0.5% (w/v) NaCl and 0.7% (w/v) agar in ddH2O, autoclaved for 25min.
4) Phage dilution buffer
   20 mM Tris-HCl pH 7.5, 10 mM MgSO4 and 40 mM NaCl in ddH2O, 0.2 μm filtered.
5) Tail buffer
   20mM sodium phosphate monobasic, 20mM sodium phosphate dibasic, ddH20, 0.2 μm filtered.

3.3 METHODS

Phage particle preparation for Det7

1. Preheat 1 L of LB in a 2.8L Fernbach Flask at 37°C

2. Pick plaque from plate and place in test tube with 0.5ml of phage buffer
    a. Hold at 37°C for 30min -1hr or until plug is mostly dissolved
3. Add 0.5ml of overnight culture of LT2 to Fernbach
    a. Incubate for 15-20 min
4. Add contents of test tube into (phage pick) into Ferbach
    a. Incubate 4-6 hrs – can be allowed to go overnight but does not usually lead to a significant increase in phage
5. Add at least 5ml of $CHCL_3$ and mix thoroughly for 1 min
    a. Place on ice for 20 min
    b. Mix again 2-3 times during ice bath
6. Remove all $CHCL_3$ by tipping flask and using pipet
7. Add 1ml of 1M $MgSO_4$ then add DNase at about 0.2 µg/ml
    a. 100 µl of Hendrix lab stock at 1mg/ml
8. Raise temp to 30-37°C for 10-15 min
9. Spin out debris - Use 3 bottles in JA10 rotor at 7.5K for 15 min
10. PEG precipitation
    a. Add 0.5M NaCl to cleared lysate (29.2g/L)
    b. Once salt is dissolved add PEG 8000 to 10% (100g/L)
    c. Stir until dissolved - 45 min to overnight
        i. Slightly more phage are recovered with O/N
    d. Pellet at 8K for 10-15 min in JA10 rotor
    e. Decant of supernatant and drain pellets
    f. Resupend each tube using 20ml
    g. Spin out extra PEG using Ja 25.50 at 5k for 10-12 min
    h. Repeat as necessary to remove all PEG
11. Add triton X-100 to 60 ml of resuspended PEG pellet for final concentration
    a. Use at least 60 ul of 10% stock solution on TritonX-100
12. Pellet phage
    a. Place 60ml resuspension into Ti45 tube- total volume of tube is ~70ml
    b. Bring up to volume and make balance tube
    c. Spin at 20k for 90 min in Ti45
13. CsCl banding
    a. Resuspend phage pellet in 8ml of phage buffer overnight
    b. Add 8.4g of CsCl and bring volume up to 12ml
    c. Bring density to 1.5 using refractive index
    d. Place in heat seal tube and balance tube
    e. Spin in Ti70.1 for 16 hrs at 38K and 18°C

2) Preparation of Det7 tails

5% - 45% sucrose gradients in phage buffer were prepared using the GradientMaster. The concentrated particles were run through the gradient in a SW41 rotor at 35K rpm at 20°C for 30 minutes. The tails, identified by light scattering, were located on the top of the gradients. The sucrose in the tail sample was removed by dialysis against 2L phage buffer at 4°C overnight, and the sample was further purified by BioCAD Sprint ion-exchange chromatography.

3) Sequencing

The genome was sheared using a hydroshear and cloned into pBluescript II KS+. The genome was sequenced at the University of Pittsburgh Bacteriophage Institute by Sanger sequencing using an ABI 3730. A total of 1824 clones were used to assemble the complete circularly permuted genome. The genome was assembled using the Phred/Praf CONSED assembly package[51].

4) Phage DNA as adapted from Dale *et al.*[52]

5) Phage protein preparation for SDS-PAGE as adapted from Boulanger *et al.* [53]

6) Phage protein preparation for N-terminal sequencing as adapted from [53]

7) Phage proteins preparation for mass spectrometry as adapted from Lavigne *et al.*[54]

4.0 DISCUSSION

The tail of Det7 presents an excellent platform to investigate and explore the correlations between structure and evolution in regards to phage host attachment. The genome characterization places Det7 in the context of the T4 superfamily and specifically in the newly identified genus *Viunalikevirus*. The addition of the *Viunalikeviruses* allows a further characterization and a greater understanding of the evolutionary forces guiding bacteriophage host range selection.

The core of the Det7 tail is recognizably similar to that of T4. Homologues of twelve of the nineteen structural proteins of T4 have been unambiguously identified in Det7. The Det7 tail differs from T4 at the periphery and posterior end of the baseplate but, despite the remarkable similarity in the rest of the tail, there is no similarity on the host recognition proteins. Det7 employs a host recognition system more often seen in *Podoviridae* using tail spike proteins targeting O-antigens. The diversity in host range for Det7 is accomplished through its ability to display multiple different receptor binding proteins. These multiple tailspikes allow for the breadth of host range Det7 displays on *Salmonella* strains with wildly different O-antigen structures.

The multiple tail spike genes of Det7 present it with a selectable advantage over phages that can only attach to one type of surface receptor. The recombination module containing the tail spikes conveys an additional selectable advantage by allowing the exchange

of surface receptor proteins allowing an alteration and adjustment of host range. In real world mixed environments the ability to infect and replicate in multiple hosts provides the phage with sustainability and host reservoir reserves.

The best characterized tail spike is that of P22 which cleaves the 1,4,12 O-antigen at the 3,6-dideoxyhexoses [21]. Det7 gp5 shares structural homology and absolute conservation of the active site residues with P22TSP as well as 9NATSP. The identical host ranges of P22 and 9NA in addition to the structural comparison suggest strongly that they bind identical O-antigens. The host range of Det7 encompasses that of P22 and 9NA, also suggesting that Det7 gp5 binds and cleaves the same O-antigens. The presence of this tail spike in all three phage tail morphology lineages is an example of horizontal gene transfer. Det7 appears to have acquired an area of increased recombination that allows it to maximize its advantage in horizontal gene transfer that positively impacts its host range capabilities.

While this work was under way, Pickard *et al.* described a conserved acetyl esterase domain that targets the vi capsular receptor present on *Salmonella* Typhi. This group of seven morphologically diverse *Salmonella* typhi typing phages have exchanged, via recombination, their tail spike proteins that target the vi capsular antigen [55]. Additionally, the identification of the *Viunalikevirus* genus provides an example of multiple tail spikes horizontally transferred as recombination modules between nearly identical phage[26]. A genomic region with a defined framework allows these phages to alter their host ranges by swapping of tail spikes that bind to specific O-antigens. This group of phages have allowed the identification of host range specificities to specific tail spike proteins. By combining host range information with genomic and structural comparisons the work presented in this thesis has allowed for connections to be made between specific gene products and host ranges.

BIBLIOGRAPHY

1. Edwards, R.A. and F. Rohwer, *Viral metagenomics.* Nat Rev Micro, 2005. **3**(6): p. 504-510.
2. Angly, F.E., et al., *The Marine Viromes of Four Oceanic Regions.* PLoS Biol, 2006. **4**(11): p. e368.
3. Sullivan, M.B., et al., *Three <italic>Prochlorococcus</italic> Cyanophage Genomes: Signature Features and Ecological Interpretations.* PLoS Biol, 2005. **3**(5): p. e144.
4. Šimoliūnas, E., et al., *<italic>Klebsiella</italic> Phage vB_KleM-RaK2 — A Giant Singleton Virus of the Family <italic>Myoviridae</italic>.* PLoS ONE, 2013. **8**(4): p. e60717.
5. Walter, M., et al., *Structure of the Receptor-Binding Protein of Bacteriophage Det7: a Podoviral Tail Spike in a Myovirus.* J. Virol., 2008. **82**(5): p. 2265-2273.
6. Barr, J.J., et al., *Bacteriophage adhering to mucus provide a non–host-derived immunity.* Proceedings of the National Academy of Sciences, 2013. **110**(26): p. 10771-10776.
7. Scott, A.E., et al., *Bacteriophage influence Campylobacter jejuni types populating broiler chickens.* Environmental Microbiology, 2007. **9**(9): p. 2341-2353.

8. Cenens, W., et al., *Expression of a Novel P22 ORFan Gene Reveals the Phage Carrier State in <italic>Salmonella</italic> Typhimurium.* PLoS Genet, 2013. **9**(2): p. e1003269.

9. Cenens, W., et al., *Phage–host interactions during pseudolysogeny: Lessons from the Pid/<italic>dgo</italic> interaction.* Bacteriophage, 2013. **3**(1): p. e25029.

10. Jiang, S.C. and J.H. Paul, *Significance of Lysogeny in the Marine Environment: Studies with Isolates and a Model of Lysogenic Phage Production.* Microbial Ecology, 1998. **35**(3-4): p. 235-243.

11. Thingstad, T., *Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems.* Limnology and Oceanography [Limnol. Oceanogr.], 2000. **45**(6): p. 8.

12. Weigele, P.R., et al., *Genomic and structural analysis of Syn9, a cyanophage infecting marine Prochlorococcus and Synechococcus.* Environmental Microbiology, 2007. **9**(7): p. 1675-1695.

13. Wilhelm, S.W. and C.A. Suttle, *Viruses as regulators of nutrient cycles in aquatic environments*. 2000.

14. Lawrence, J.G. and H. Hendrickson, *Lateral gene transfer: when will adolescence end?* Molecular Microbiology, 2003. **50**(3): p. 739-749.

15. Hendrix, R.W., et al., *Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage.* Proceedings of the National Academy of Sciences, 1999. **96**(5): p. 2192-2197.

16. Casjens, S.R. and P.A. Thuman-Commike, *Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly.* Virology, 2011. **411**(2): p. 393-415.

17. Petty, N.K., et al., *A generalized transducing phage for the murine pathogen Citrobacter rodentium.* Microbiology, 2007. **153**(9): p. 2984-2988.

18. Flores, C.O., et al., *Statistical structure of host–phage interactions.* Proceedings of the National Academy of Sciences, 2011.

19. Thompson, J.E., et al., *The K5 Lyase KflA Combines a Viral Tail Spike Structure with a Bacterial Polysaccharide Lyase Mechanism.* Journal of Biological Chemistry, 2010. **285**(31): p. 23963-23969.

20. Seckler, R., *Folding and Function of Repetitive Structure in the Homotrimeric Phage P22 Tailspike Protein.* Journal of Structural Biology, 1998. **122**(1-2): p. 216-222.

21. Andres, D., et al., *An essential serotype recognition pocket on phage P22 tailspike protein forces Salmonella enterica serovar Paratyphi A O-antigen fragments to bind as nonsolution conformers.* Glycobiology, 2013. **23**(4): p. 486-494.

22. Andres, D., et al., *Tail morphology controls DNA release in two Salmonella phages with one lipopolysaccharide receptor recognition system.* Molecular Microbiology, 2012. **83**(6): p. 1244-1253.

23. Baxa, U., et al., *Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide.* Biophysical Journal, 1996. **71**(4): p. 2040-2048.

24. Andres, D., et al., *Carbohydrate binding of Salmonella phage P22 tailspike protein and its role during host cell infection.* Biochem Soc Trans, 2010. **38**(5): p. 1386-1389.

25. Cresawn, S., et al., *Phamerator: a bioinformatic tool for comparative bacteriophage genomics.* BMC Bioinformatics, 2011. **12**(1): p. 395.

26. Adriaenssens, E., et al., *A suggested new bacteriophage genus: "Viunalikevirus".* Archives of Virology, 2012. **157**(10): p. 2035-2046.

27. Brenner, F.W., et al., *Salmonella Nomenclature.* Journal of Clinical Microbiology, 2000. **38**(7): p. 2465-2467.

28. Popoff, M.-Y., J. Bockemühl, and F.W. Brenner, *Supplement 1998 (no. 42) to the Kauffmann-White scheme.* Research in Microbiology, 2000. **151**(1): p. 63-65.

29. Butela, K. and J.G. Lawrence, *Genetic Manipulation of Pathogenicity Loci in Non-Typhimurium Salmonella.* Journal of Microbiological Methods, 2012. **91**(3): p. 477-482.

30. Beltran, P., et al., *Reference collection of strains of the Salmonella typhimurium complex from natural populations.* Journal of General Microbiology, 1991. **137**(3): p. 601-606.

31. Boyd, E.F., et al., *Salmonella reference collection B (SARB): strains of 37 serovars of subspecies I.* Journal of General Microbiology, 1993. **139**(6): p. 1125-1132.

32. Boyd, E.F., et al., *Molecular genetic relationships of the salmonellae.* Applied and Environmental Microbiology, 1996. **62**(3): p. 804-8.

33. Wollin, R., U. Eriksson, and A.A. Lindberg, *Salmonella Bacteriophage Glycanases: Endorhamnosidase Activity of Bacteriophages P27, 9NA, and KB1.* Journal of Virology, 1981. **38**(3): p. 1025-1033.

34. Haase-Pettingell, C. and J. King, *Prevalence of temperature sensitive folding mutations in the parallel beta coil domain of the phage P22 tailspike endorhamnosidase.* Journal of Molecular Biology, 1997. **267**(1): p. 88-102.

35. Steinbacher, S., et al., *Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage.* Journal of Molecular Biology, 1997. **267**(4): p. 865-880.

36. Tang, L., et al., *Three-dimensional structure of the bacteriophage P22 tail machine.* EMBO J, 2005. **24**(12): p. 2087-2095.

37. Steinbacher, S., et al., *Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors.* Proceedings of the National Academy of Sciences, 1996. **93**(20): p. 10584-10588.

38. Leiman, P., et al., *Morphogenesis of the T4 tail and tail fibers.* Virology Journal, 2010. **7**(1): p. 355.

39. Rossmann, M.G., et al., *The bacteriophage T4 DNA injection machine.* Curr Opin Struct Biol, 2004. **14**(2): p. 171-80.

40. Bayer, M.E. and M.H. Bayer, *Fast Responses of Bacterial Membranes to Virus Adsorption: A Fluorescence Study.* Proceedings of the National Academy of Sciences of the United States of America, 1981. **78**(9): p. 5618-5622.

41. Eswar, N., et al., *Comparative Protein Structure Modeling Using MODELLER*, in *Current Protocols in Protein Science*. 2001, John Wiley & Sons, Inc.

42. Söding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction.* Nucleic Acids Research, 2005. **33**(suppl 2): p. W244-W248.

43. Kelley, L.A., Sternberg, Michael J E, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat. Protocols, 2009. **4**(3): p. 8.

44. Chua, J.E.H., P.A. Manning, and R. Morona, *The Shigella flexneri bacteriophage Sf6 tailspike protein (TSP)/endorhamnosidase is related to the bacteriophage P22 TSP and has a motif common to exo- and endoglycanases, and C-5 epimerases.* Microbiology, 1999. **145**(7): p. 1649-1659.

45. Biegert, A., et al., *The MPI Bioinformatics Toolkit for protein sequence analysis.* Nucleic Acids Research, 2006. **34**(suppl 2): p. W335-W339.

46. Casjens, S., et al., *The Chromosome of Shigella flexneri Bacteriophage Sf6: Complete Nucleotide Sequence, Genetic Mosaicism, and DNA Packaging.* Journal of Molecular Biology, 2004. **339**(2): p. 379-394.

47. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* Journal of Molecular Biology, 1999. **292**(2): p. 195-202.

48. Buchan, D.W.A., et al., *Protein annotation and modelling servers at University College London.* Nucleic Acids Research, 2010. **38**(suppl 2): p. W563-W568.

49. Ochman, H. and R.K. Selander, *Standard reference strains of Escherichia coli from natural populations.* Journal of Bacteriology, 1984. **157**(2): p. 690-693.

50. Anany, H., et al., *A Shigella boydii bacteriophage which resembles Salmonella phage ViI.* Virology Journal, 2011. **8**(1): p. 242.

51. Gordon, D., C. Abajian, and P. Green, *Consed: A Graphical Tool for Sequence Finis hing.* Genome Research, 1998. **8**(3): p. 195-202.

52. Dale, J.W. and P.J. Greenaway, *Preparation of Phage Lambda DNA*, in *Nucleic Acids*, J. Walker, Editor. 1985, Humana Press. p. 211-215.

53. Boulanger, P., *Purification of Bacteriophages and SDS-PAGE Analysis of Phage Structural Proteins from Ghost Particles*, in *Bacteriophages*, M.J. Clokie and A. Kropinski, Editors. 2009, Humana Press. p. 227-238.

54. Lavigne, R., P.-J. Ceyssens, and J. Robben, *Phage Proteomics: Applications of Mass Spectrometry*, in *Bacteriophages*, M.J. Clokie and A. Kropinski, Editors. 2009, Humana Press. p. 239-251.

55. Pickard, D., et al., *A Conserved Acetyl Esterase Domain Targets Diverse Bacteriophages to the Vi Capsular Receptor of Salmonella enterica Serovar Typhi.* J. Bacteriol., 2010. **192**(21): p. 5746-5754.