ABSTRACT

Title of Dissertation: STRUCTURAL AND FUNCTIONAL STUDIES OF TAILSPIKE PROTEINS FROM *ESCHERICHIA COLI* O157:H7 PHAGE CBA120

Julia Y. Greenfield, Doctor of Philosophy, 2019.

Dissertation Directed by: Dr. Osnat Herzberg, Professor, Department of Chemistry and Biochemistry

Bacteriophage CBA120, a member of the *Ackermannviridae* family, was isolated, sequenced, shown to selectively infect the food pathogen *E. coli* O157:H7. The CBA120 genome encodes four separate tailspike proteins (TSPs), TSP1-4 corresponding to ORFs 210 through 213. TSPs bind and degrade or modify the lipopolysaccharide (LPS) on the bacterial cell surface as part of the adsorption apparatus of tailed bacteriophages. Electron microscopy revealed that phage CBA120 possesses a long contractible tail with distinct star-like structures attached to the baseplate of the tail.

Sequence analysis of the four CBA120 TSPs suggests two distinct classes; one class comprising TSP1 and TSP3 corresponds to two-domain structures. The second class comprising TSP2 and TSP4 have longer amino acid sequences and contain additional N-terminal regions responsible for the assembly of the TSP star-like complex. Crystal structure of TSP1 at 1.8 Å resolution and TSP3 at 1.85 Å resolution revealed two-domain

homotrimers displaying conserved N-terminal head domain structures and C-terminal receptor binding domains with an overall β-helical fold. Sequence analysis of the assembly regions of TSP2 and TSP4 reveal remote homology to gp10 of phage T4, which is involved in the assembly of the phage baseplate. The crystal structure of TSP2 reveals the head domain and receptor binding domain, but the N-terminal assembly region is structurally disordered. In contrast, the assembly region of TSP4 has been purified without the head and receptor domain; pull-down experiments showed that this region binds TSP1.

Bacterial halo assays of TSPs 1-3 revealed that all three proteins produced circles of clearing corresponding to polysaccharide depolymerase activity. Site-directed mutagenesis coupled with the halo assay confirmed the identity of TSP3's catalytic residues as Asp383 and Asp426, which, based on the structure, suggests that the enzyme employs an acid/base mechanism to degrade lipopolysaccharide. The structure-based putative active site of TSP2 suggests that the catalytic machinery comprises Asp506 and Asp571. It was also shown that TSP2, but not TSP1 or TSP3, diminishes CBA120's ability to infect *E. coli* O157:H7. Thus, TSP2 is the phage specificity determinant that binds to the lipopolysaccharide of *E. coli* O157:H7, whereas the bacterial targets of TSP1 and TSP3 remain unknown.

STRUCTURAL AND FUNCTIONAL STUDIES OF TAILSPIKE
PROTEINS FROM *ESCHERICHIA COLI* O157:H7 PHAGE CBA120

By

Julia Y. Greenfield

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Osnat Herzberg, Chair
Professor Daniel Nelson
Professor John Orban
Professor Paul Paukstelis
Professor Kevin McIver, Dean's representative

Dedication

I dedicate this thesis to my husband Robert, for his support, encouragement, and patience. I also dedicate it to my Mom, Dad and my sisters: Erica and Mya for believing in me when I doubted myself; and to my friend, Lola, for challenging me and motivating me to complete this work. Last but not least, I dedicate this thesis to my grandmother, Nina, for her support and for never letting me hide under the bed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BLAST	Basic Local Alignment Search Tool
BRIG	BLAST Image Ring Generator
BSA	bovine serum albumin
E. coli	*Escherichia coli*
EPS	Extracellular polymeric substances
LPS	lipopolysaccharide
SeMet	selenomethionine
STEC	Shiga toxin-producing *E. coli*
TSP	tailspike protein

Chapter 1: Introduction

As more viral genomes are sequenced, and DNA sequencing costs decrease, we can unlock the many unique and specialized mechanisms bacteriophages use in their fight against bacteria. The use of phages as antimicrobial agents has been practiced as early as 1920; however, the rise of antibiotics has made the development of phage and phage product therapies a low priority[1]. More recently, the growing prevalence of drug-resistant bacteria and the rapid growth of the number of phage genome sequences have revived interest in using these organisms as a viable and mostly untapped resource for new biocontrol and therapeutic agents. This dissertation focuses on structural and biochemical studies of bacteriophage CBA120, a virus that infects pathogenic *Escherichia coli* O157:H7 and a cause of widespread gastrointestinal disease. Understanding the mechanism it utilizes to invade the pathogen is the first necessary step for developing the phage or phage components as antibacterial agents.

1.1   *E. coli O157:H7*

Escherichia coli is a rod-shaped, Gram-negative bacterium that can inhabit the intestines of people and animals. Most strains are harmless; however, some are pathogenic, including *E. coli* O157:H7. First recognized in 1982, *E. coli* O157:H7 is one of the most common Shiga toxin-producing *E. coli* (STEC) pathogens identified in North America[2]. Infection by Enterohemorrhagic *E. coli* (EHEC) such as O157:H7 can cause hemorrhagic colitis, characterized by bloody diarrhea and extreme abdominal discomfort. In a small percentage of cases, *E. coli* O157:H7 infection can progress to hemolytic-uremic syndrome, where red

1

blood cell damage is accompanied by acute kidney failure[2, 3]. Antibiotics are not usually prescribed to treat *E. coli* O157:H7 infection because the rapid killing of the bacteria can cause Shiga toxin to be released from dead cells and worsen symptoms, and increase the risk of developing hemolytic-uremic syndrome. Therefore, there is no standard treatment for EHEC infections that target the bacteria directly; however, intravenous fluids can be given to combat dehydration caused by diarrhea associated with the infection[4]. For the small population of patients who develop hemolytic-uremic syndrome, treatment to reduce the symptoms of the illness include dialysis for those experiencing renal failure and blood transfusions for the hemolytic anemia caused by the destruction of red blood cells[4, 5]. Despite the medical consensus of avoiding antibiotics to treat *E. coli* O175:H7 infection, resistance to antibiotics has been described[5]. The Center for Disease Control reports 3,704 confirmed cases of *E. coli* O157:H7 annually; however, many cases are undiagnosed or misdiagnosed, suggesting the actual number of instances to be closer to 63,000[6].

In addition to the physical suffering of those infected with the pathogenic microbe, there is also an associated economic burden. The cost related to contaminated food includes hospitalization, loss of wages, legal fees, recall expenses and, loss of profits for the food industry. The United States Department of Agriculture (USDA) monitors contaminated food recalls; in 2017, recalls caused by STEC pathogens reached 144,000 lbs[7]. This number is lower than the 10 million pounds of food recalled during 2013[8], and food companies still endured a heavy financial burden as the result of a product recall. For example, I.M. Healthy company has filed for bankruptcy since its 2017 recall of SoyNut Butter that was linked to contamination with *E. coli* O157:H7[9]. The CDC estimated that over 48 million

Americans suffered from a foodborne illness in 2011 and Dr. Robert Scharff, of Ohio State University's Food Innovation Center, estimates that the total cost of foodborne illness in the United States reaches $55 billion annually[10].

Despite the precautions taken by the food industry, *E. coli* O157:H7 remains a persistent food pathogen. Ungulates such as cattle and deer can serve as a reservoir for *E.coli* O157:H7 and be asymptomatic[5, 11]. Infected animals shed the pathogen in their feces, which can contaminate recreational and drinking water used by humans[5]. During the slaughtering process, bacteria from infected animals can be incorporated into meat intended for human consumption. Moreover, bacterial contamination can spread to food processing equipment and surfaces increasing the risk of cross-contamination with other food products. Thoroughly cooking the infected meat can kill the bacteria and reduce the chance of infection; however, many outbreaks with minimally processed food products such as apple cider, salad, and sprouts have been recorded[12]. Produce contamination is primarily attributed to cross-contamination from infected meats at restaurants or facilities that process both meat products and minimally processed foods[12]. Unlike meat products, consumption of minimally processed food products is more concerning because consumers do not necessarily cook these products before eating, and it takes <100 bacteria to cause illness[5].

E. coli O157:H7, as many other bacteria, have defense mechanisms for survival in hostile environments. One such defense strategy is the production of a structurally complex and

heterogeneous matrix called a biofilm. Bacterial biofilm is formed by secreting polysaccharides, proteins, and DNA, which comprise the extracellular polymeric substances (EPS)[13]. Biofilms can grow on a variety of surfaces from natural products such as vegetables and industrial materials[14]. The biofilm structure provides a local environment where a subset of the bacterial community persists in a metabolically inactive state while being protected from desiccation and other environmental hazards. Biofilms also provide a physical barrier from bacteriophages and antimicrobial agents[13, 15]. These metabolically inactive bacteria are not susceptible to many antimicrobial agents, but recover their activity when the biofilm structure is compromised[15]. The released bacteria can then colonize a new area. The cyclic growth of biofilm formation can cause chronic infection inside a host organism or persistent contamination on external surfaces.

Gram-negative bacteria contain external lipopolysaccharides (LPS) that extend from the cell surface. Also known as endotoxins, because they can produce hemorrhagic shock when released from dead cells, these molecules consist of three components: lipid A, core oligosaccharide, and O-antigen see Figure 1.1. The lipid A contains phosphates, glucosamine, and fatty acid chains that anchor the LPS to the bacterial membrane. Lipid A is connected to the core, which can be further divided into two, the inner core and the outer core. The inner core is bound to lipid A by several keto-deoxyoctulosonate (KDO) and heptose moieties and is conserved over bacterial genus. In contrast, the outer core contains more variability within a genus. There are five known outer core types associated with *E. coli* including the R3 core type, which is found in many virulent isolates including O157:H7[16]. The core oligosaccharide connects the O-antigen, a long repetitive

carbohydrate polymer that extends out from the bacterial surface. O-antigens vary across different strains of *E. coli*. The genes encoding enzymes that synthesize these structures are located on the O-antigen gene cluster. *E. coli* O157 gene cluster includes 12 genes involved in the production of N-acetyl-D-perosamine, L-fucose, D-glucose, and N-acetyl-D-galactose, the components of the O157: H7 antigen; see Figure 1.1[17, 18].

A

O-antigen   common core   lipid A

B

$$[\alpha\text{-D-PerNAc} \xrightarrow{1,3} \alpha\text{-L-Fuc} \xrightarrow{1,4} \beta\text{-D-Glc} \xrightarrow{1,3} \alpha\text{-D-Gal2NAc}] \xrightarrow{1,2}$$
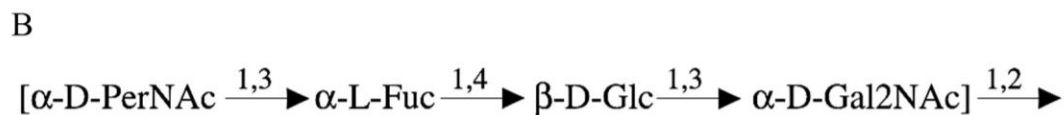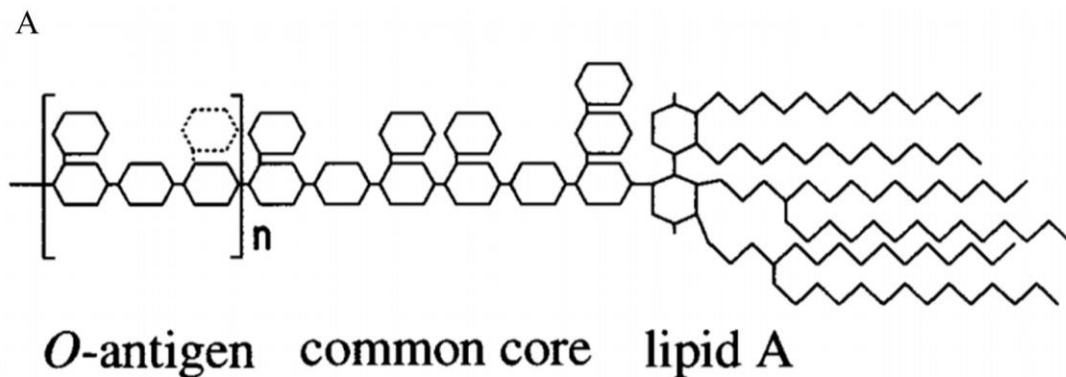
Figure 1.1 Lipopolysaccharide composition.
(A)The hydrophobic lipid A anchors the LPS in the cell membrane. Inner and outer core make up the common core structure of the LPS, and O-antigen is a repeating unit of four sugars attached to the core (figure adapted from Hughes et. al[19]). (B) The O-antigen structure of *E. coli* O157 includes the following four sugars; D-4-*N*-acetylperosamine (D-Per4NAc), L-fucose (L-Fuc), D-glucose (D-Glc), and D-2-*N*-acetylglucosamine (D-Glc2NAc,). (Figure adapted from Samuel et al. 2004[18]).

1.2 *Bacteriophages*

Caudovirales, the order of double-stranded tailed DNA bacteriophages, represent a diverse group of viruses with an estimated ten million species[20]. Tailed phages have one of the most recognizable virus morphology; they are characterized by their icosahedral heads

(capsid) containing their dsDNA genomes, and their tail appendage. The *Caudovirales* order is further divided into families based on their tail morphology; the *Podoviridae*, which have short non-contractile tails*;* the *Siphoviridae,* which have long non-contractile tails; and the *Myoviridae,* which have long contractile tails illustrated in Figure 1.2[13]. The tail machinery is responsible for host recognition before adsorption and transport of the DNA to the host during infection. *Caudovirales* have a baseplate connected to the end of the tail to which tailspikes or tail fibers are attached. The tailspikes (TSPs), the focus of this investigation, are responsible for binding, selective host recognition, and in contrast to tail fibers, tailspikes can also possess enzymatic activity that degrades the LPS[20]. Various TSPs have been studied, revealing several characteristics: their lack of amino acid sequence homology, their conserved three-dimensional fold, and the affinity for lipopolysaccharide (LPS) of their specific host organism[13]. In addition to their LPS selectivity, some TSPs also hydrolytically cleave the polysaccharides of the LPS, allowing the phage unhindered access to the cell surface for infection[13].
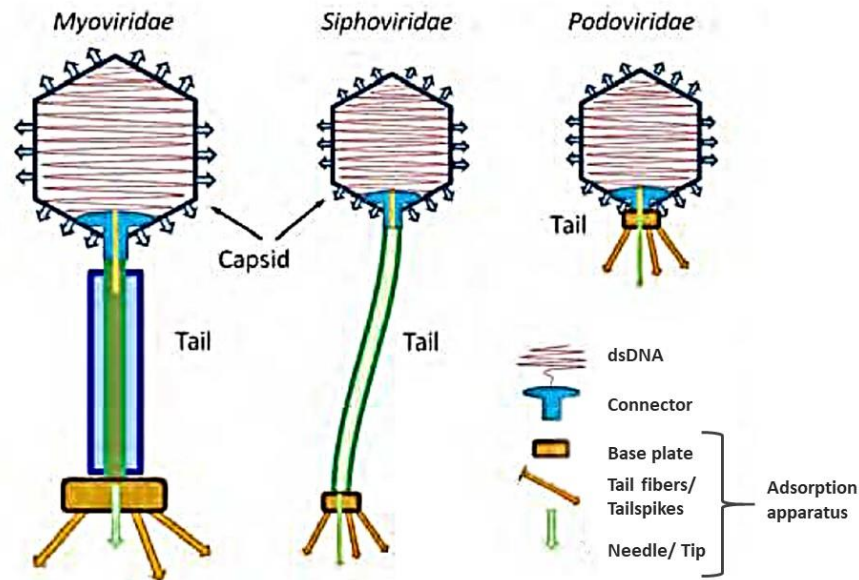
Figure 1.2 Illustration of tailed bacteriophages.
Myoviridae, Siphovirdae, and *Podoviridae* are the three families of tailed bacteriophages. At the distal end of the tail is the adsorption apparatus consisting of a baseplate, tail fibers/tailspikes, and a tail needle to inject their dsDNA. Figure adapted from Casjens et al. 2005[21].

Host infection by the bacteriophage is a three-step process: 1) the phage binds reversibly to the polysaccharide secondary receptor such as the LPS, 2) the phage uses its depolymerase activity to degrade the polysaccharide of the LPS, and 3) the phage recognizes and binds irreversibly to the primary receptors on the cell surface and invades the bacterial cell[13, 22].The long contractile tails of the *Myoviridae* are essential for their efficiency during the infection process. These tails consist of many proteins including long tail fibers, which can interact with the primary receptors on the bacterial cell surface; the tail needle required for the phage to inject its DNA; and the TSPs, which recognize the secondary receptor and degrade polysaccharides[23, 24]. The ability of phages to recognize

7

specific bacterial targets and degrade the LPS makes them drug candidates for bacterial control. Specifically, a phage or a phage component that selectively targeted *E. coli* O157:H7 offers opportunities to develop biologics to treat or prevent *E.coli* O157:H7 contamination in food products, thereby preventing illness.

The *Viunalikevirus* genus belongs to the Myoviridae family; they are distinguished by several traits including genomes of similar size (approximately 150,000+ base pairs), T4-like regulation of late transcription, the use of a modified form of uracil rather than thymine, and the use of tailspikes rather than tail fibers for host recognition[24]. Electron microscopy of members of this genus uncovered an unusual tail structure that differs from the long spinally tail fibers of the T4-like myovirus. Instead, *Viunalikevirus* members have an elaborate organization of branched star-like projections coupled with bulbous prongs protruding from the base plate, illustrated in Figure 1.3. Most members of the genus encode four TSPs, unlike other phages, which encode only a single TSP[24]. Because of their multiple TSPs, it has been suggested that perhaps these phages can infect a broader range of bacteria[25]. Indeed, infection of multiple bacteria species has now been identified with the *Salmonella* phage SFP10[25].
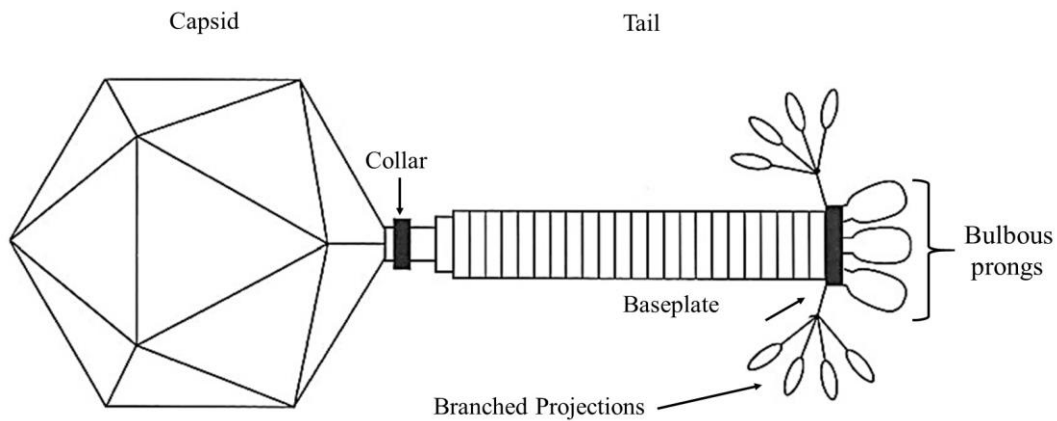
Figure 1.3 Schematic illustration of *Viunalikevirus* Genus Phage.
Cartoon adapted from "A suggested new bacteriophage genus: *Viunalikevirus*'' by Adriaenssens et al.[25]. The tail apparatus shows the prongs at the bottom of the baseplate. Also connected to the baseplate are branched projections.

Escherichia phage vB_EcoM_CBA120 (CBA120), a member of the *Viunalikevirus* genus, has a 157,000 base pair genome and four TSPs[24], denoted here TSP1-4. CBA120 was isolated from a cattle feedlot and was shown to infect the enterohemorrhagic *E. coli* O157:H7[3, 26]. Bacteriophage CBA120 has specificity towards multiple *E. coli* O157: H7 stereotypes, infecting 13 of the 17 O157 strains tested, but only 1 out of 77 non-O157:H7 strains tested from the *E. coli* Reference Collection[27]. CBA120's specificity towards *E. coli* O157:H7 is of particular interest because it may be developed as a potential new bio-indicator and biocontrol agent.

CBA120 phage has morphology consistent with the reported morphology of other *Viunalikeviru*s genus phages (see Figure 1.3). CBA120's icosahedral capsid, which measures 90 nm is connected to the contractile tail by a collared neck[24, 27]. The tail sheath

consists of 24 transverse striations connecting the tail to the base plate. The adsorption apparatus is anchored to the baseplate at the distal end of the tail; branched star-like structures emanate from the sides of the baseplate, and bulbs prongs protrude from the bottom of base plate; see Figure 1.3. To date, this genus includes seven sequenced members: *Salmonella* phages ViI, SFP10 and ΦSH19; *Escherichia* phages CBA120 and PhaxI; *Shigella* phage phiSboM-AG3; and *Dickeya* phage LIMEstone1. Their shared myovirus morphology, with comparable capsid sizes and tail dimensions, and genome organization are considered distinguishing features as detailed, in Table 1.1. They appear to have conserved regulatory sequences, a horizontally acquired tRNA set, and the probable substitution of an alternate base for thymine in the DNA. A close examination of the tailspike encoding region of the genome revealed four distinct tailspike proteins, and it was proposed that they assemble into an arrangement that comprises the star-like structures of the tails visible on electron micrographs. The star-like structures that fan out from the baseplate of CBA120 appear morphologically different from the long spinally tail fibers present in T4 phages, and it remains unknown whether it corresponds to the four tailspike encoded in CBA120's genome. These morphological properties set the *Viunalikevirus* genus apart from the recently ratified subfamily *Tevenvirinae*, although a significant evolutionary relationship can be observed[23].

Table 1.1 List of reported *Viunalikevirus* Genus Members

Virus	Head size (nm)	Tail size (nm)	Genome size (bp)	Open reading frames	Genome similarity and homologous genes to CBA120 #	reference
Dickeya phage LIMEstone1	91.4	113.8x14	152,427	201	76% / 155	[28]
Escherichia phage CBA120	90	105x17	157,304	210-213	N/A	[27]
Escherichia phage PhaxI	85-86	115x15	156628	209	95.1% / 194	[29]
Salmonella phage Sfp10	*86	*109x18	157950	201	93.1% / 190	[25]
Salmonella phage ΦSh19	83	110x14	157,785	166	77.9%/ 159	[30]
Salmonella phage ViI	88-90	110x18	157,061	208	88.5% / 184	[24, 31]
Shigella phage Ag3	83	110x14	158,006	216	85.3 / 174	[32]

*Dimensions calibrated by Adriaenssens et al. 2012[24].
Genome similarity determined by CoreGene[33] against the genome of CBA120.

Overall, the seven identified phages in *Viunalikevirus* genus have similar genomes supporting the classification of a distinct genus in 2012 by Kutter et al.[24]. The CoreGenes3.5[33] webserver was used to compare the genomes of the *Viunalikevirus* members against CBA120 to identify sequence homology and overall similarity amongst the phages. *Viunalikevirus* members share a significant number of homologs; at least 150 homologous genes were identified in ΦSh19, while PhaxI, with 194 genes, showed the largest number of homologs to CBA120. According to this analysis, all annotated *Viunalikevirus* genomes demonstrated better than 75% similarity when compared to CBA120, phage ViI (88.5% similarity), PhaxI (95%), SFP10 (93%), AG3 (85%), LIMEstone1 (76%), and ΦSh19 (78%). To visualize the similarities, I used a genome-wide

BLASTn search of *Viunalikevirus* members and depicted the results using the BLAST

Image Ring Generator (BRIG) tool[34]. As seen in Figure 1.4, the region around 110kBP-

130kbp, where the genes for TSPs are encoded, show significant gaps in similarity. An

expansion of this region is depicted in Figure 1.5.
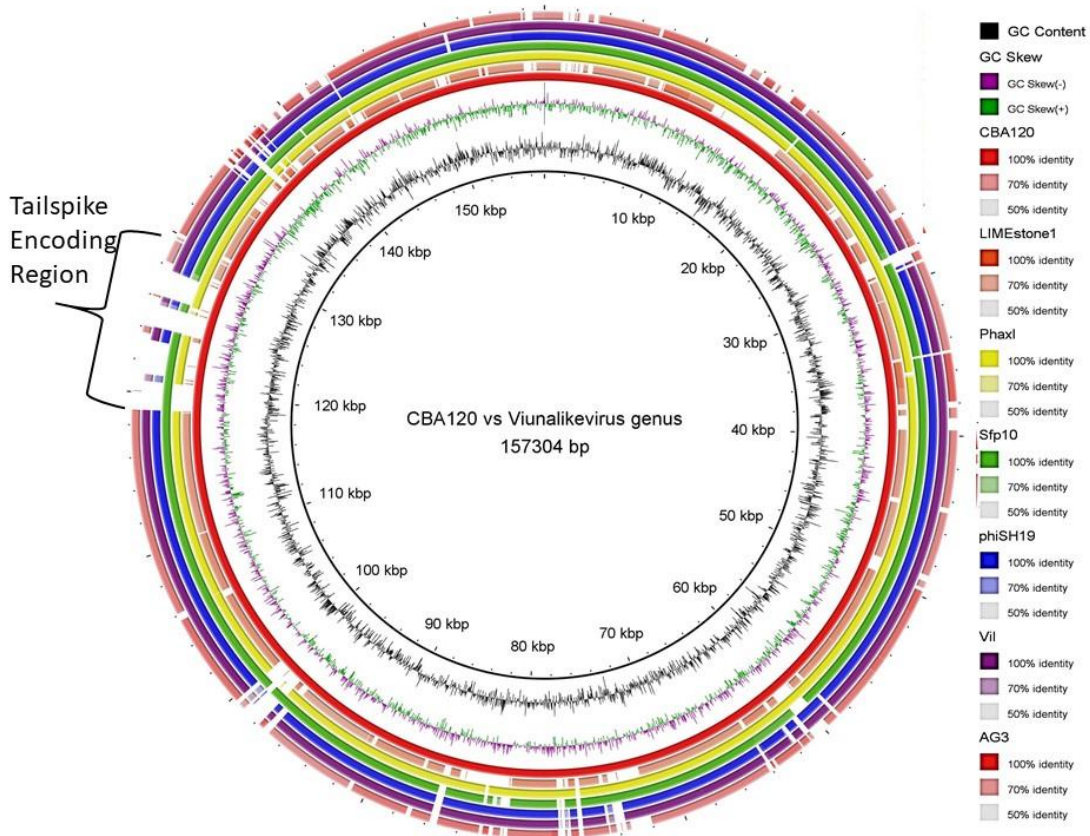
Figure 1.4 BLASTn Genome Comparison of *Viunalikevirus* members against Phage CBA120.
The innermost rings show the GC content (black) and GC Skew (purple/green). The third innermost ring shows the CBA120 genome in red. The remaining rings show the remaining members of the *Viunalikevirus* genus: LIMEstone1, PhaxI, Sfp10, phiSH19, ViI, and AG3. The coloring scheme is indicated on the right. The area with the most dissimilarities occurs in the region encoding the CBA120 TSP1-4 (orf 210-213). This image was created using BLAST Ring Image Generator (BRIG)[34].

The BLASTn comparison shows four gaps for ΦSh19, ViI, and AG3 corresponding to the

genes encoding TSP1-4. However, their TSP genes are significantly different from

CBA120, hence the gaps. PhaxI and SFP10 only have two gaps in this region of their genome when compared to CBA120. The gaps can be explained by the fact that PhaxI and SFP10 encode two tailspikes that are homologs to CBA120's (PhaxI encodes for TSP2 and TSP4 homologs, and SFP10 encodes for TSP1 and TSP2 homologs). LIMEstone1 only encodes for 3 TSPs, and they are all significantly dissimilar from CBA120 to cause gaps to be displayed in this area. In addition, two of LIMEstone1's tailspikes are missing the catalytic domain.
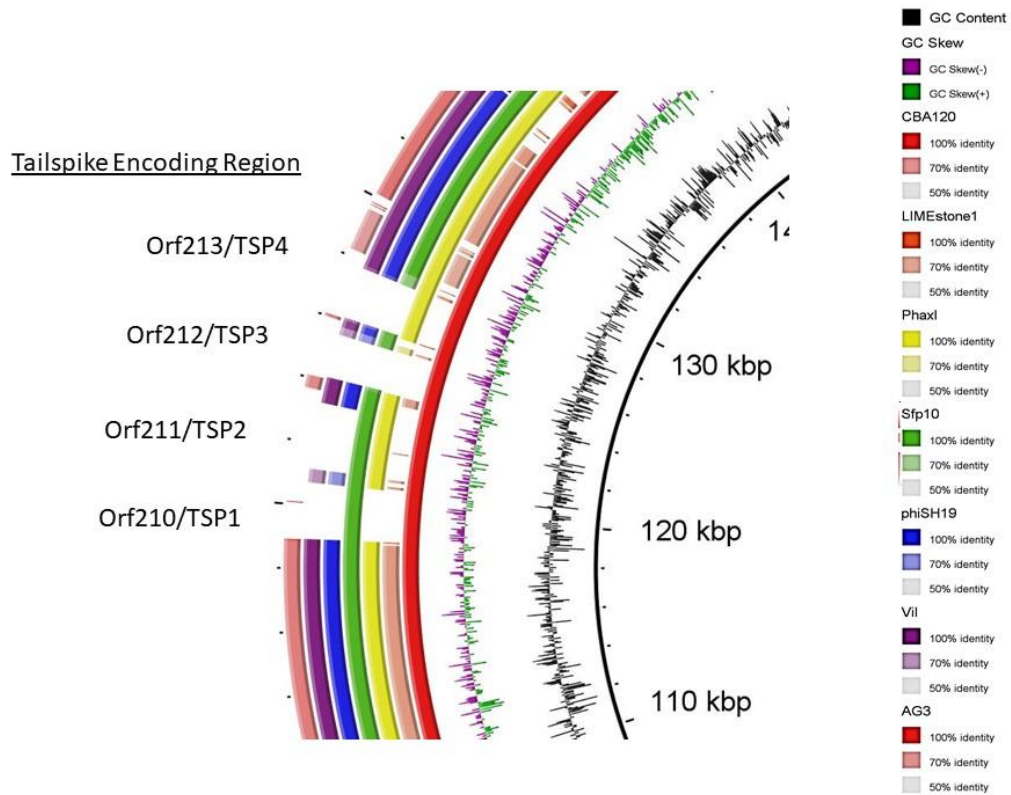
Figure 1.5 Expansion of the tailspike encoding regions of *Viunalikevirus* genus.

The gaps correlate to TSP1-4, specifically the receptor binding domain of each gene. Preceding each area of low identity, the conserved regions represent the head domains of TSPs. The receptor binding domains of the tailspike protein regions have the most genetic

13

diversity because this region of the tailspike determines host specificity; *Viunalikevirus* members have a variety of hosts and therefore variability in their host receptor proteins.

1.3 *The Family of Ackermannviridae*

During the writing of this thesis, the taxonomy regarding *Viunalikevirus* genus has been changed. Since several key references refer to the original genus classification, and I will continue to use it for this thesis. The current classification information was proposed and accepted by the International Committee on Taxonomy of Viruses (ICTV) after a "reassessment of the genus and phages related to its members"[35]. The result requires an expansion of the order of *Caudovirales* to include a new family, *Ackermannviridae*[35-37], see Figure 1.6. Under the *Ackermannviridae* family, there are two subfamilies *Aglimvirinae* and *Cvivirinae*. *Escherichia* phage CBA120 is the type species of the newly formed genus *Cba120virus* which is under the *Cvivirinae* subfamily. Members of the genus *Cba120virus* are characterized as having genomes that average 158.1 kbp in length and 44.5 mol % G+C content[35]. Most members of this genus encode approximately 201 proteins; however, there a few exceptions including *Salmonella* phage PhiSH19 and *Salmonella* phage 38[35]. Every member of the genus *Cba120virus* has a DNA sequence identity of 86% or better to CBA120[35]. Along with the creation of *Cba120virus* genus three other genera were created; the *Ag3virus*, *Limestonevirus*, and the *Vi1virus*. Each of these new genera contains a former member of the *Viunalikevirus* genus as type species: *Shigella* phage phiSboM-AG3, *Dickeya* phage vB_DsoM_LIMEstone1, and *Salmonella* phage Vi01[35, 36], see Table 1.3.
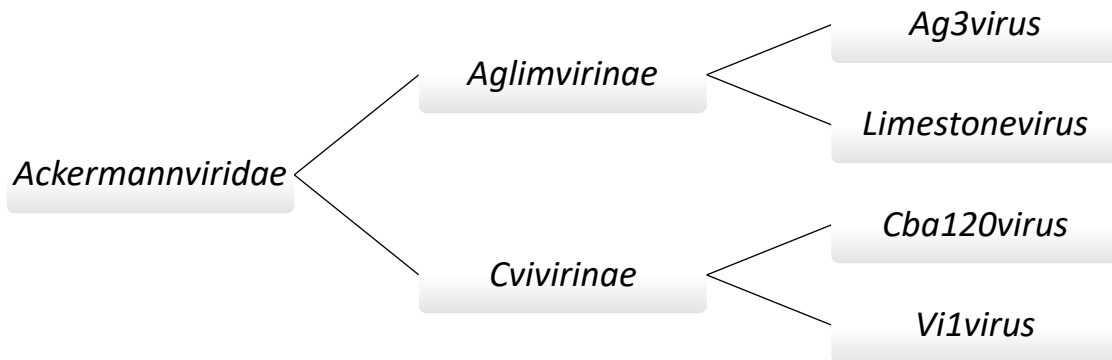
Figure 1.6 Taxonomy of the *Ackermannviridae* family.
The order of *Caudovirales* contains a new family of bacteriophage, the *Ackermannviridae*.
Escherichia phage CBA120 serves as the genus type representative of the *Cba120virus*
genus which contains four members from the former *Viunalikevirus* genus.

Table 1.2 Members of the Cba120virus genus

Phage Name	GenBank accession No.	RefSeq No.	Genome length (kb)	%G+C	% DNA sequence identity*	# proteins
Escherichia phage CBA120§	JN593240	NC_016570	157.3	44.5	100	204
Salmonella phage SFP10	HQ259103	NC_016073	158.0	44.5	92	201
Salmonella phage GG32 #	KX245012	NC_031045	157.9	44.5	89	202
Escherichia phage PhaxI	JN673056	NC_019452	156.6	44.5	93	209
Salmonella phage Det7 #	KP797973	NC_027119	157.5	44.6	86	210
Salmonella phage PhiSH19	JN126049	NC_019530	157.8	44.7	86	166
Salmonella phage 38 #	KR296692	NC_029042	156.8	44.6	265***	265***
Salmonella phage vB_SalM_PM10 # *	KX438380	NC_031128	158.1	44.6	209	209
Salmonella phage vB_SalM_SJ3	KJ174318	NC_024122	162.9	44.4	210	210

*** sequence contains numerous errors; # new species, § type species in the genus
Table adapted from Kropinski et al. 2017[35].

Table 1.3 Former Members of *Viunalikevirus* genus

Phage Name	GenBank accession No.	RefSeq No.	Genome length (kb)	%G+C	# of proteins	New Subfamily	New Genus
Dickeya phage vB_DsoM_ LIMEstone1§	HE600015	NC_019925	152.4	49.3	201	*Aglimvirinae*	*Limestonevirus*
Shigella phage phiSboM-AG3§	FJ373894	NC_013693	158.0	50.4	216	*Aglimvirinae*	*Ag3virus*
Salmonella phage Vi01§	FQ312032	NC_015296	157.1	45.2	208	*Cvivirinae*	*Vi1virus*

§ genus type species

Table adapted from Kropinski et al. 2017[35]

1.4 *Tailspike Proteins (TSPs)*

In order for phage adsorption to take place, the phage receptor binding proteins must interact with bacterial receptors. TSPs serve as the receptor binding proteins that facilitate the initial reversible contact between the phage and the bacterial cell. TSPs form stable trimeric proteins, with each protomer containing two main domains see Figure 1.7 [28]. The N-terminal domain includes the head domain that attaches the tailspike to the baseplate of the virus particle or other tailspikes. The head domain is divided into two subdomains, D1 and D2, following the notation used to describe TSP1 from phage CBA120[38]. The C-terminal domain, also known as the receptor binding domain, is further divided into two subdomains, a large β-helix domain (designated thereafter the D3 subdomain following the notation of TSP1 from phage CBA120), followed by a smaller C-terminal domain, which varies in fold among tailspikes and has been designated as the D4 subdomain, see Figure 1.8. The β-helix domain serves two functions: recognizing and cleaving the host bacteria's receptor[24, 39]. The activity that degrades the polysaccharide receptor is often, but not always, an endoglycosidase of the O-antigen, explaining the host specificity of the

bacteriophage. Although the receptor binding domains of known TSP structures exhibit the same overall fold, they lack significant amino acid sequence homology. The C-terminal subdomain (D4) has proven to be essential for tailspike trimerization of Phage P22's TSP[40], but this role has not been confirmed in other TSPs.

N-terminal Domain   C-terminal Domain
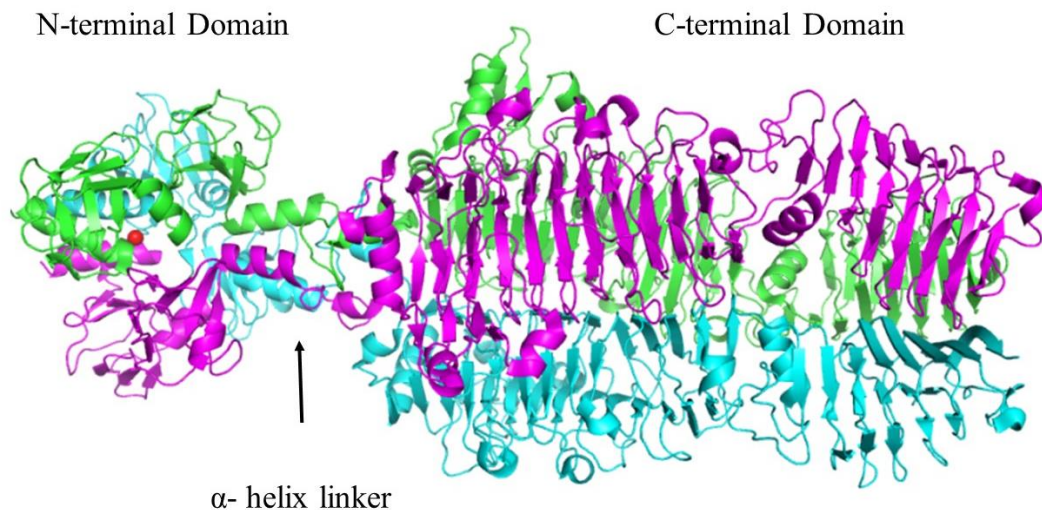
α- helix linker

Figure 1.7 Tailspike trimer of CBA120's TSP1.
Full-length TSP1 is depicted here as a highly integrated homotrimer. Two distinct domains connected by alpha-helix linker are observed[38]. PDB 4OJ5.

Instead of glycosidase activity, some TSPs merely modify the LPS or act on others substrates altogether. For example, TSPs of bacteriophages G7C, K1F or P22 have diverse enzymatic activity and substrates. Phage G7C's TSP deacetylates LPS[41], K1F TSP is an endosialidase that cleaves the polysialic acid of its host's capsule[42], and P22 is a hydrolase acting on a sugar branching from the repeating subunit of the *Salmonella* O-antigen[43]. The C-terminal domain of the 3-D structures of G7C and K1F TSPs do not adopt a β-helix fold, however, P22's tailspike does. The diversity of TSP sequences makes it impossible to identify the specific enzymatic activity of a particular TSP based on sequence alone.
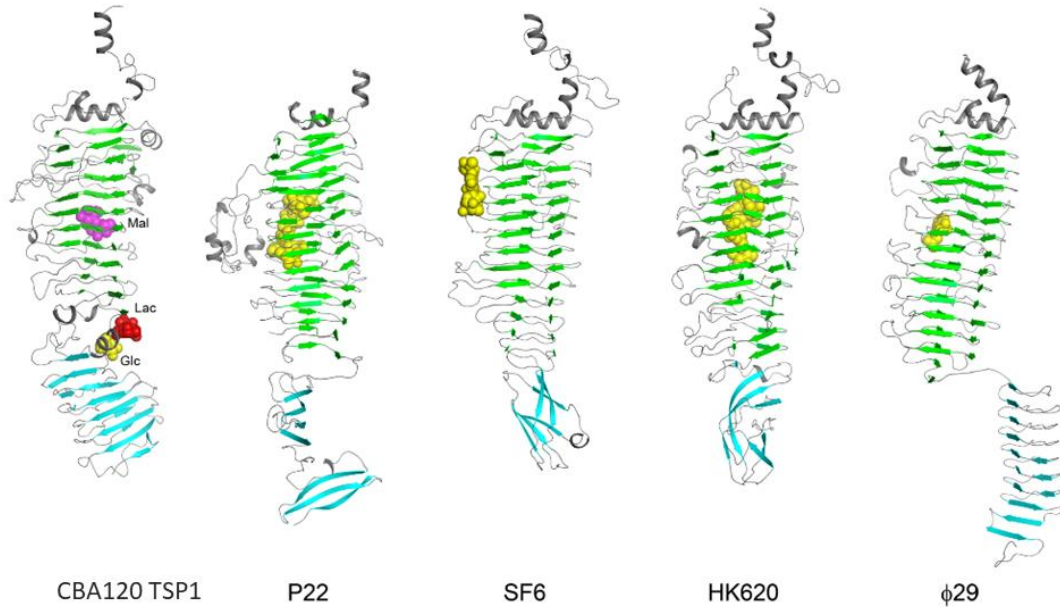
Figure 1.8 Structures of TSP receptor binding domain monomers sharing a β-helix catalytic domain.
Following the notation used to describe the phage CBA120 TSP1, the catalytic domain will be designated the D3 subdomains (green). The D4 subdomains (cyan) fold varies between TSPs. Turns and loops are colored gray. Bound sugars are shown as space-filling spheres. Figure adapted from Chen et al. 2014[38].

Table 1.4 Features of β-helical TSPs with determined crystal structures.

Virus	PDB code	Monomer mass (kDa) /oligomeric state	Activity
CBA120 TSP1	4OJ5	82/trimer	Unknown
P22	1TYW	72/ trimer	endorhamnosidase
SF6	2VBE	67/trimer	endorhamnosidase
HK620	2VJI	77/trimer	endo-N-acetylglucosaminidase
Φ29	3GQ7	92/trimer	teichoicase

1.5   *Salmonella Phage P22*

The TSP of *Salmonella* phage P22 has been studied most extensively, and the knowledge derived from these studies has provided the context for understanding other TSPs. Phage P22 TSP's active site serves as both receptor binding site for host recognition and a receptor endoglycosidase. It may seem as nonproductive for the receptor binding protein to also destroy the receptor. However, the slow activity of P22's TSP, one cleavage a minute at 10 ° C, is fine-tuned for the initial host recognition, and the slow enzymatic activity enables the phage to take advantage of the binding-affinity without destroying the receptor too quickly. Although not demonstrated, it was suggested that the endoglycosidase activity might also play a role in releasing phage progeny from the lysed bacterial cell debris, freeing them to infect other cells. Such receptor-destroying activity is seen in influenza type A neuraminidase, which is responsible for releasing the virus from bacterial surface sugars. It is important to note that unlike TSPs, the influenza type A neuraminidase is not essential for host recognition, a function carried by a hemagglutinin[43] .

The structure of bacteriophage P22, as determined by cryogenic electron microscopy (Cryo-EM), is shown in Figure 1.9, identifying the location of the TSP. The crystal structure of the P22 TSP was determined in 1996 by Steinbacher *et al.*[43]. The Cryo-EM structure of P22 shows six homotrimers are present on the Phage P22 tail structure[44, 45]. Each protomer exhibits a right-handed β-helix motif similar to the one first seen in pectate

lyase, an enzyme with glycosidase activity[46]. Thus, the presence of the β-helix motif in CBA120 TSP1, as well as in the TSPs of P22, HK620, Sf6, and Φ29, suggests, although it does not prove, a glycosidase activity. Consistent with the fold of its receptor binding domain, the P22 TSP acts on the O-antigen of several related *salmonella* serotypes, including serotype A, B, and D1 which share a common trisaccharide repeat[43].
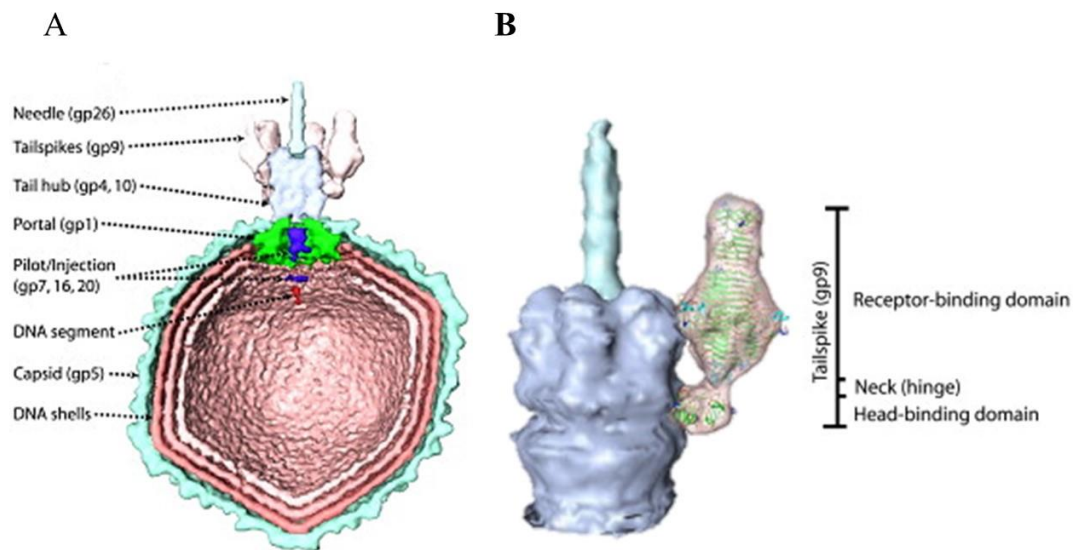
Figure 1.9 Cryo-EM reconstruction of *Podoviridae* phage P22.
(A) A cross section of the Phage P22 structure shows the capsid and the tail apparatus. The tail is made up of several essential proteins including TSP, as labeled. (B) The tail hub proteins and needle make up the bulk of the tail apparatus. Two distinct domains are visible on the TSP, the head domain and the receptor-binding domain. The crystal structure of P22 tailspike fits well in this space. Figure adapted from Chang et al. 2006[44].

The saccharide binding site of P22 tailspike is located in a shallow depression on the surface of the β-helical domain and flanked by extended loops on either side. The active site of P22 TSP contains three carboxylic acid residues, E359, D392 and D395, all located within the cleft. Single mutations of E359, D392, and D395 to Gln or Asn cause a 10,000

to 30,000 fold reduction in enzymatic activity[43]. The binding site cleft can accommodate

eight saccharide units or two repeats of the O-antigen, as seen in Figure 1.10[47].
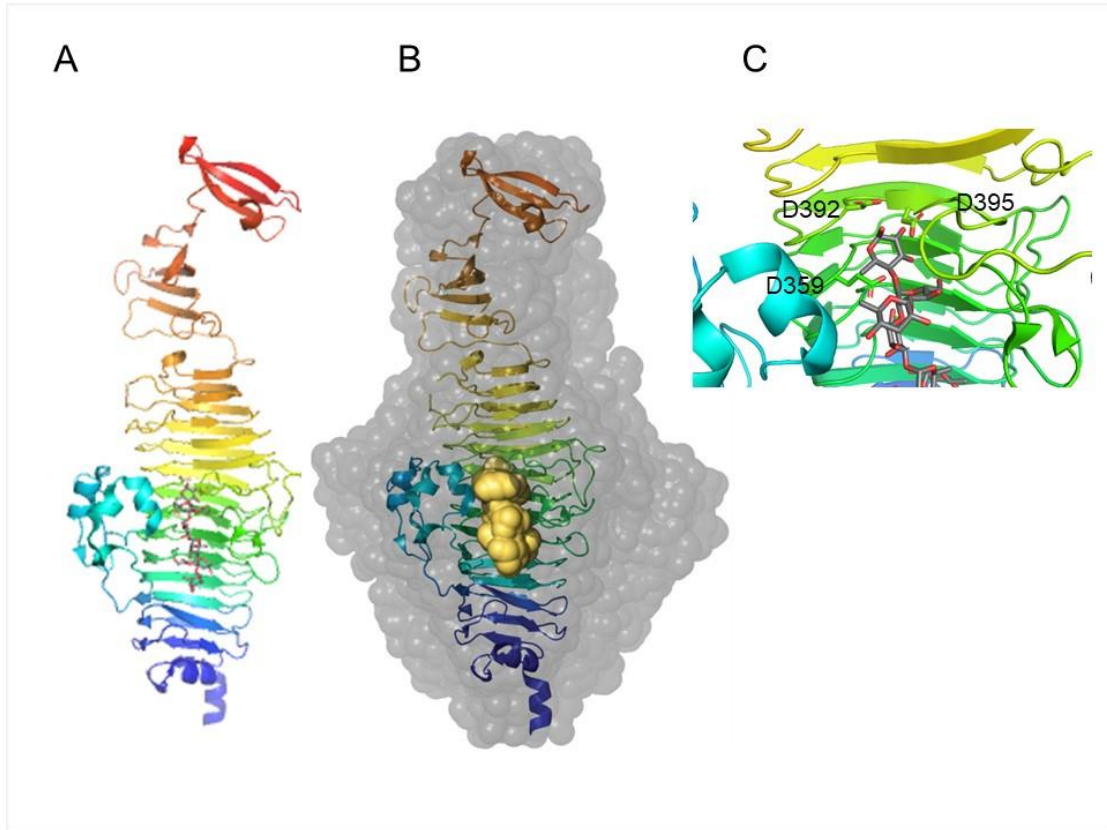
Figure 1.10 Phage P22 saccharide binding Site (PDB:3TH0)
(A) Cartoon representation of a full length P22 tailspike monomer with bound *S. Paratyphi A* octasaccharide (gray) (B) A surface representation of P22 TSP trimer with one monomer in cartoon representation showing The *S. Paratyphi A* octasaccharide (yellow spheres). (C) The active site of P22 has three potentially catalytic residues E359, D392, and D395; interacting with the octasaccharide substrate. Figure adapted from Andres et al. 2013[48].

1.6 *Phage CBA120 TSPs*

Since TSPs and tail fibers are known to convey host specificity, the roles of CBA120's

four TSPs are intriguing. As mentioned above, it was suggested that the multiple TSPs

expand the range of hosts the phage infects, but this has not been proven[24]. The amino acid

sequences of the four TSPs exhibit remarkable diversity. The only conserved region

between the four TSPs corresponds to the head domains with 39-70% amino acid sequence identity over 155 residues. TSP2 also shows sequence homology to the maturation-adhesion protein of Phage ViI tailspike protein 171c. Based on their domain structure, TSP1 and TSP3 resemble TSPs from the *Podoviridae* family[24]. However, TSP2 and TSP4 contain additional domains N-terminally to the head domain. It is tempting to speculate that these domains are involved in the assembly of the branched projections seen in electron micrographs (Figure 1.3)[24].

Chapter 2: Electron Microscopy of CBA120

2.1 _Introduction_

The morphology of _Viunalikevirus_ genus along with genomic sequencing provided significant support for the genus classification Adriaenssens et al.[24]. The distinctive structural feature on the tail of _Viunalikevirus_ genus members along with genomes that encode for multiple TSPs suggests that these phages have an unusual arrangement of TSPs. To investigate the positions of the TSPs on the virus tail, Transmission Electron Microscopy (TEM) was used. The TEM images of Vil, LIMEstone1, PhaxI, SFP10, ΦSH19, Ag3, and CBA120 shown in Figure 2.1[24, 25, 27, 29, 30, 32] demonstrate tail arrangements similar to the illustration in Figure 1.3. Capsid shape, contractile tails, and the "star-like" projections that extend from the bottom of the base plate have been reported in all the TEM images. The star-like projections are seen in many of the images; however, there is some variation amongst the members of the genus. Most notably TEM images of _Dickeya_ phage LIMEstone1 lacks the elaborate branched structure present on the other phages. This difference can be attributed to the fact that LIMEstone1 phage encodes for only one complete tailspike (orf158), and truncated versions of two other tailspikes (orf159 and 160), which have sequence homology to the N-terminal domains of TSP2 and TSP4 respectively[24]. Because the published images of the CBA120 phage were not clear, (Kutter et al. 2011[27]) I obtained new images from phages I cultured, showing the branched structures thought to comprise the tailspike assemblies more clearly.
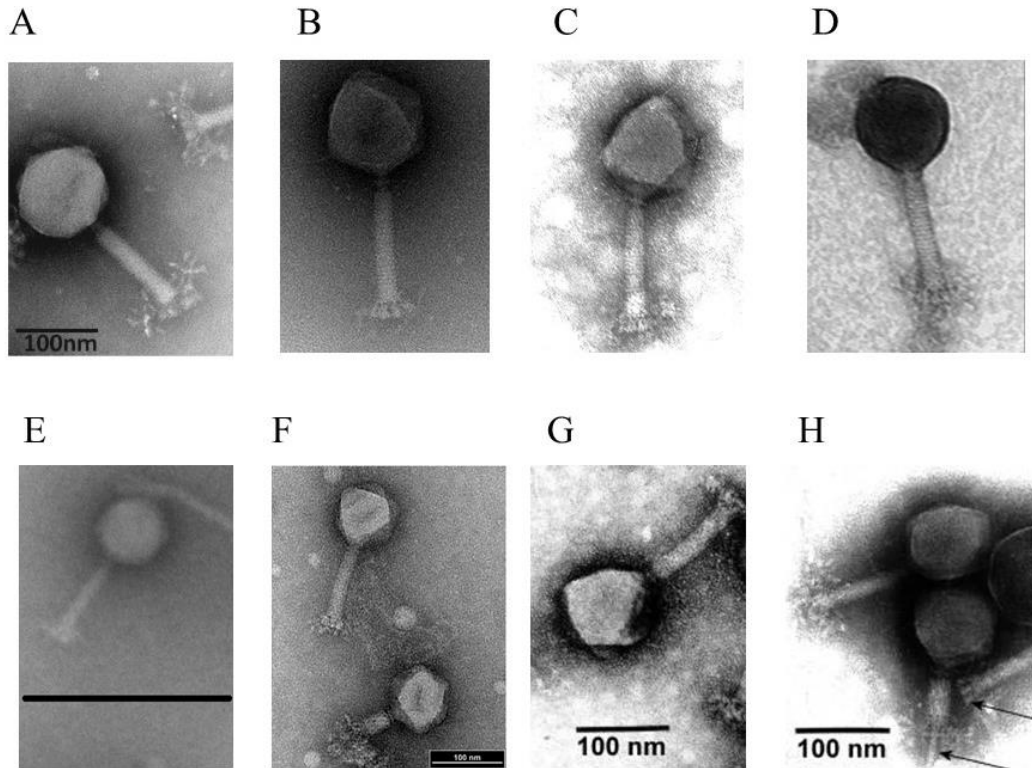
Figure 2.1 Electron micrographs of *Viunalikevirus* genus phages
(A) *Salmonella* phage Vil demonstrates a complex tail structure with many branched projections[24]. (B) *Dickeya phage* LIMEstone1 phage lacks the elaborate branched tail structure but contains the prong-like projections[24]. Other members of the genus have a structure consistent with *Salmonella* phage ViI, including (C) *E. coli* Phage PhaxI[29], (D) *Salmonella phage* SFP10[25], (E) *Salmonella phage* ΦSH19[30], (F) *Shigella Phage* Ag3[32], (G, H) *E.coli* phage CBA120[27].

2.2 *Material and Methods*

CBA120 bacteriophage, obtained from Dr. Elizabeth Kutter at Evergreen State College in

Olympia, W.A., was used to inoculate *E. coli* O157:H7 ATCC700728 and the culture was

grown in Luria-Bertani broth for four hours, centrifuged at 2500 rpm for 15 min at 4 °C.

The *E. coli* was pelleted, and the supernatant was filtered with a 2μm filter to remove any

bacterial cells. The CBA120 present in the supernatant was then concentrated using a 50

kDa concentrator and washed with distilled water. A 1:1 solution of CBA120 in water and

0.5% uranyl acetate were mixed in a microcentrifuge tube. The mixture spotted on a carbon-covered copper 200 mesh grid (Ted Pella, Inc.), and examined on a Joel 2100 Transmission Electron Microscope (TEM) at 200kv with a CCD camera.

2.3   *Results*

The electron micrographs of CBA120 are consistent with other *Viunalikevirus* genus members except for LIMEstone1 phage, which only encodes for one complete tailspike (Figure 2.2). The capsid diameter is consistent with 90 nm as reported by Kutter et al.[27]. The tail apparatus depicts a collar (Figure 2.2 arrow A) and a segmented tail sheath of approximately 100 nm. The baseplate is obscured by the elaborate star-like projections extending away from the tail (Figure 2.2 arrow B). Also visible are the stubby projections extending from the bottom of the baseplate (Figure 2.2 arrow C).
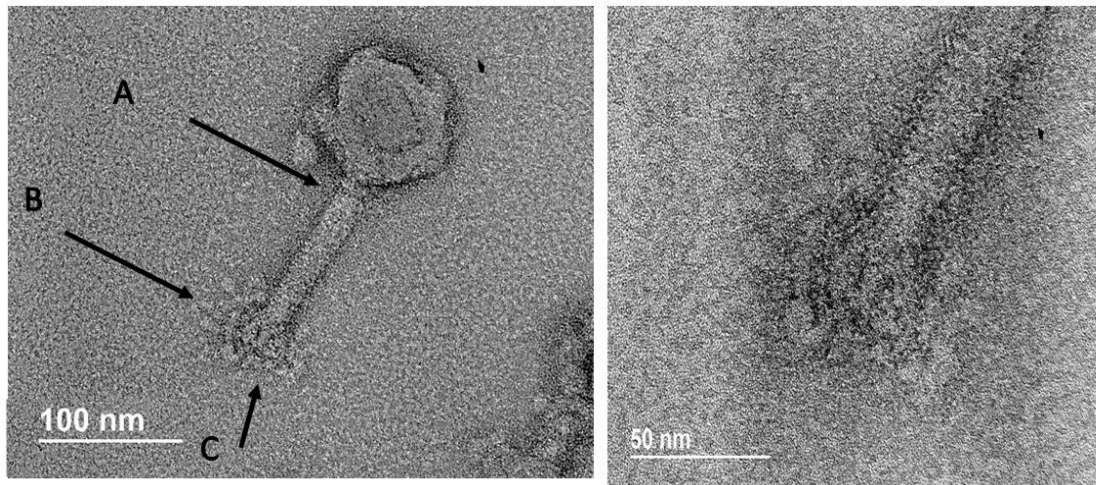
Figure 2.2 TEM image of CBA120.
(Left) *E. coli* Phage CBA120 phage from the current work. (Right) Enlargement of CBA120's tail showing both the "stubby" prongs and the star-like structures around the bottom of the tail.

2.4 *The Proposed Analogy Between G7C and CBA120 Tailspike Assemblies*

While *Salmonella* Phage P22 has revealed a lot about the structure of a tailspike, another member of the *Podoviridae* family has provided insight that allows us to speculate about the arrangement and interactions of receptor binding proteins in phages that encode multiple TSPs. This phage is N4-like phage vB_EcoP_G7C (phage G7C), which infects *E. coli* 4s, a bacteria that infects the equine intestinal tracts[49]. The genome of phage G7C has 71,759 base pairs and 78 open reading frames, much smaller than the CBA120 genome, which contains 150,000 base pairs and 213 orf[24, 49]. The electron micrograph of phage G7C (Figure 2.3) shows that the tail includes tailspikes extending from the base in a branched arrangement reminiscent of the one seen in *Viunalikevirus* genus. Also seen in the figure are appendages extending from the tail collar, as reported for N4 and other N4-like viruses[50].
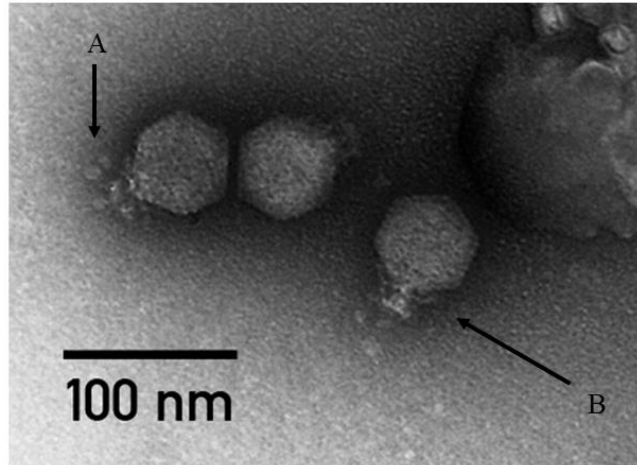
Figure 2.3 EM image of phage G7C.
Star-like structures reminiscent of the ones seen in the *Viunalikevirus* genus are indicated by arrow A. Thin projections extending from the collar are indicated by arrow B. Figure adapted from Kulikov et al. 2012 [49].

Phage G7C has two distinct receptor binding proteins, tailspikes gp63.1 and gp66. The former is a tailspike that deacetylates *E. coli* 4s's LPS. Gp63.1 forms an elongated homotrimer of 851 amino acids per monomer, with each monomer containing a head domain and a receptor binding domain[41, 49]. The N-terminal head domains of gp63.1 have been shown to be essential in the assembly of the branched tailspike structure in G7C[41]. Consisting of the first 155 amino acids, the head domain of gp63.1 exhibits 74%, 60% and 63% sequence identity to the head domains of TSP1, TSP3, and TSP4 head domains, respectively, see Figure 2.4. Adopting the same overall fold, gp63.1's head domain has a 0.5Å (RMSD) of α-carbon atoms to TSP1's head domain. Gp63.1 receptor binding domain has demonstrated deacetylation activity against *E. coli* 4s LPS. Interestingly gp63.1 does not cleave the LPS of *E. coli* 4s; instead, it deacetylates a carbohydrate moiety on the O-antigen[41]. The receptor binding domain of gp63.1 does not show any homology to the receptor binding domains of CBA120's tailspikes.

```
gp63.1    15 ----------------------------------------ILTNKQAIARHFGVKQSEVV
TSP1      15 ----------------------------------------ILTNKQAVARHFGVKQSEVV
TSP3      14 ----------------------------------------IEVNKQSIARNFGVKEDEVI
TSP4     301 YKNLGFTFDPLTSKITLAQELDAEDEVVVIINGTPNIYNQIDYTLREVARVTNVKDTEVI


gp63.1    35 YFSVGAVLSGYKVIYDKGTQRAYSLPANIGSGVTAISLSPAGVLVHSAGSVDLGALAVTR
TSP1      35 YFSVGVDLGGYKVIYDKETQRAYSLPVGIASGTTAVSLSTAAVLVHSAGSVDLGSLAVSR
TSP3      34 YFTAGIDLSGFKVIYDESTQRAYSLPFGIVSGTTAISLDERAILTHSAGSVDLGELAVSR
TSP4     361 YFSVGAVLSGYKVIYDKVTQRSYFIPE-LPTGTTAVSLSSSAILVHSAGSVDLGALAVSR


gp63.1    95 KEYVTLPDTFTSGSVIQTKNELLTHNGTQYRWAGGLPKSVPLNSTPVSAGGISPTAWVIA
TSP1      95 EEYVTLPGSFDSGSTLNVKNELLTYTDGKYRWDGILPKTVAPGSTPASTGGVGLGAWISV
TSP3      94 EEYVTLPGSFNFGHTINVKNELIVHDDKKYRWDGSLPKVVAAGSTPDSSGGVGLGAWLSV
TSP4     420 EEYVTLSGTFDSGAVINTKNELLTHTDGKYRWDGTLPKTVAAGSTPATTGGVGSGAWLSV


gp63.1   155 NDELIRQELNNGL--IPPVG-STSVYDVPG----IVVNT-T--TDNRAAAYAFPGKIF--
TSP1     155 GDASLRTQLANGD--GSLIG-IHPQGTLNN----VI---------------------
TSP3     154 GDAALRAELNTKVSDGTFPATIKYKYGLPSVIDGAIYRTVQDKLDDFVFLEDFGGKDDAG
TSP4     480 GDASLKSNLNKPNG-LSYIGTVSSVSELSS-IAGLIGDSII--LDSYVDGFNLGGGVM--
```

Figure 2.4 Sequence alignment of the head domain of phage G7C's gp63.1 vs. CBA120's TSP1, TSP3, and TSP4.
Multiple sequence alignment was produced by T-coffee alignment program [51].

The similarity of the head domains of G7C's gp63.1 and TSP1 suggest that they may have originated from a common ancestor. However, the divergent receptor binding domains indicate that the "spare parts" theory may apply here, where protein domains assemble through horizontal gene transfer and recombination [41]. Accordingly, gp63.1 does not adopt the glycosidase β-helix fold; instead, it contains a SGNH hydrolase-type esterase domain responsible for the deacetylation activity[41].

Gp66, the second and larger tailspike of phage G7C comprises 1,063 amino acids. The first 330 amino acids have been divided into three domains. The first 138 amino acids may be

responsible for binding the tailspike to the virus particle. The second and third domains, comprising residues 138-294 are responsible for binding to gp63.1[41]. Domains four and five span residues 295-460 and exhibit 37% identity to the head domains of gp63.1 and, 32%, 61%, and 39% to the CBA120's TSP1, TSP2, TSP4 head domains, respectively (Figure 2.5). Gp66's C-terminal domain is predicted to be a β-helical receptor binding domain, but its structure has not been elucidated. Figure 2.6 illustrates the proposed structure of G7C tailspike assembly and the interaction between gp63.1and gp66[41]. The model is based on pull-down experiments of gp66 domain deletion mutants[41].


```
gp66   233  GGEQVIDIPFTFQAVPAVYKNGLRFHKGLTNNSYTIDSDNNRILLTEPLATNDRLIVQLG
TSP1    13  ------------------------------------------------------------
TSP2   102  GGETEFTPDFTSYGVQSLYIGGEYKTP---EIEYSYDSTTGKVSLQTALSAGVWVVAEMS
TSP4   276  GGETEITLDIVVDDVPAIDINGSRQYK---NLGFTFDPLTSKITLAQELDAFDEVVVIIN

gp66   293  -GEAKVLEITDHTIQEVARAANVKDSEVILSTDITQVLNGKKVIYDVVTQHIYGLPT-LP
TSP1    13  -------TGILTNKQAVARHFGVKQSEVVYFSV-GVDLGGYKVIYDKETQRAYSLEVGIA
TSP2   159  VKQPNISPAFDRSIQEIARSANVKDSEVIVSTDTISLLDGKKVVYDIATQTSYGLPT-IP
TSP4   333  -GTPNIYNQIDYTIREVARVTNVKDTEVIYFSV-GAVLSGYKVIYDKVTQRSYFIPE-LP

gp66   351  TNVYINAV-SNCQLTYSPGNITVTLLDSYQQR--NTRELWRRSLA----------------
TSP1    65  SGTTAVSLSTAAVLVHSAGSVDLGSLAVSR------------------------------
TSP2   218  DGSVISSV-SAGKLNYNPGDVQVDLLPLEDSFINVINTLGRNDGAKYIGECHSVADLRNT
TSP4   390  TGTTAVSLSSSAILVHSAGSVDLGALAVSR------------------------------

gp66   393  --------------EAGLTLVDGSFEEGATVTSKLDVVWHVAAGQCYAWGGTLPKTVGAN
TSP1    95  --------------EEYVTLPGSFDSGSTLNVKNELLTYTD--GKYRWDGILPKTVAPG
TSP2   277  EPTMDGQRIILKQHTAGTLLGGGVFRALIDGTGKTD------------NNGTVIKTV---
TSP4   420  --------------EEYVTLSGTFDSGAVINTKNELLTHTD--GKYRWDGTLPKTVAAG

gp66   439  STPINTGGVSANAWVEYNS--FTTNI-PSGFYGG-SCFPT----ENNRSLS--VGDV--I
TSP1   138  STPASTGGVGLGAWISVGDASLRTQL-ANGD--G-SLI---------------------
TSP2   322  ---------GGAAWLRVNA--DRVNPFMFGALGG-S--------NDDTIP--VQSC--V
TSP4   463  STPATTGGVGSGAWLSVGDASLKSNL-NKPN--GLSYIGTVSSVSELSSIAGLIGDSIIL
```

Figure 2.5 Sequence alignment of the head domain phage G7C's gp66 and CBA120's TSP1, TSP2, and TSP4.
Multiple sequence alignment was produced by T-coffee alignment program [51].
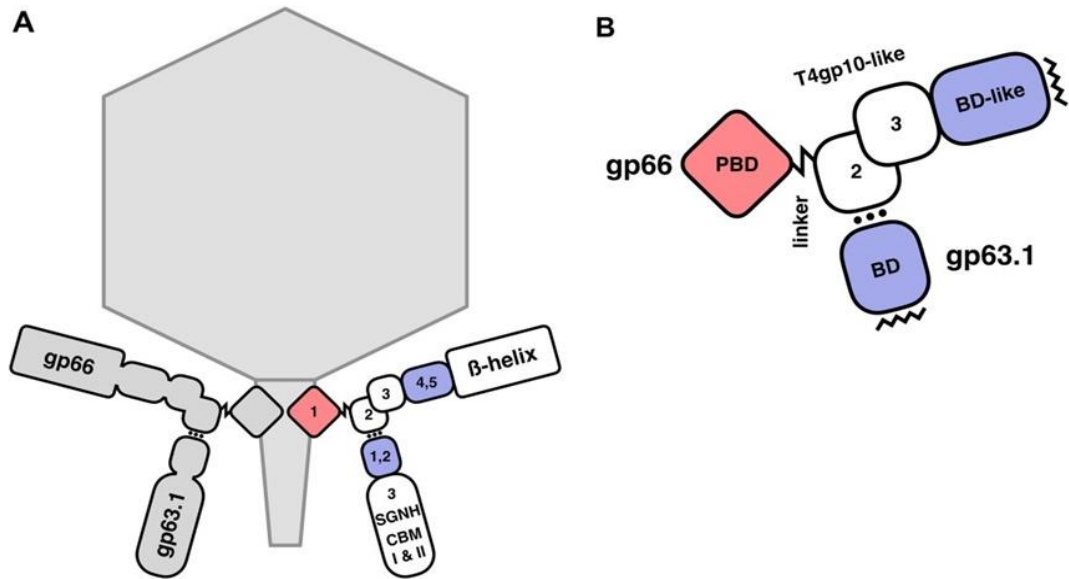
Figure 2.6 A Diagram of the proposed G7C's tailspike assembly structure
Tailspikes and gp66 associate with the rest of the virus particle via its particle binding domain (red). Tailspike gp63.1's N-terminal binding domain interacts with domain 2 of gp66 (Figure adapted from Prokhorov et al. 2017[41]).

By analogy to gp66 of phage G7C, the large N-terminal insertions preceding the head domains of CBA120's TSP4 and perhaps TSP2 may mediate the assembling of the branched tailspike structures observed in the EM images (~330 and 170 additional amino acids for TSP4 and TSP2, respectively (Table 2.1)). According to this model, the smaller TSP1 and TSP3 (~600-800 amino acids) would be analogous to gp63.1 because they contain only head and catalytic domains (Table 2.2). The CBA120 tailspike interactions are currently being tested in the laboratory by ultracentrifugation studies.

Table 2.1 Proposed domains of Assembly TSPs

Phage	Host	Protein	Residue ranges			
			Particle binding domain	Assembly domain	Head domain	Receptor binding domain (with neck and C-ter. domain)
G7C[41]	*E. coli* 4s	gp66	1-138	139-294	295-467	467-1063
CBA120	*E. coli* O157:H7	TSP2	-	1-167	168-245	246-921
		TSP4	1-81	82-329	330-480	481-1043

Table 2.2 Proposed domains of Branching TSPs

Phage	Host	Protein	Residue ranges	
			Head domain	Receptor binding domain (with neck and C-ter. domain)
G7C[41]	*E. coli* 4s	Gp63.1	1-155	156-851
CBA120	*E. coli* O157:H7	TSP1	1-155	156-770
		TSP3	1-154	155-627

The head domain of TSP4 begins at approximately amino acid 330, as inferred by the 74% sequence identity with the head domain of TSP1. By analogy to gp66, the preceding residues (1-329) may be responsible for interactions with other TSPs that assemble into the branched structures, and for attachment to the phage baseplate. For gp66, residues 1-138 have been proposed to be responsible for attachment to the phage particle and residues 139-294 for anchoring gp63.1[41]. BLAST sequence alignment of the TSP4 residues 1-335 with residues 1-294 of the gp66 reveals homology between ~70 residues of TSP4 (264-335) and two consecutive ~70-residue regions of gp66. The first gp66 region spans residues 138-

208 (38% identity and e-value of 2×10^{-11}) and the second spans residues 221-293 (37% identity of and e-value of 3×10^{-11}). These gp66 domains were reported to be related to D2 and D3 domains of gp10 from phage T4, and required for binding of gp63.1 to gp66 [41]. The homology of residues 264-335 of TSP4 to the two related gp66 domain supports the hypothesis that TSP4 residues preceding the head domain are involved in assembling other TSPs to form the branched structure seen in the EM images. In contrast, there is no sequence homology between TSP4 and the first 140 amino acids of gp66 that were proposed to be involved in the attachment of the tailspike to the main viral particle.

In addition, Hidden Markov model analysis using the program HHpred[52] reveals sequence relationship of TSP4 to phage T4's gp10 and gp9 that are components of the phage T4 baseplate and are involved in multiple protein-protein interactions[53-56]. The probability of sequence relationship between TSP4 residues 82-329 and gp10 residues 49-383 is 98% and the e-value is 2.2×10^{-5}, spanning a region that comprises the first three gp10 domains (designated D1-3). For gp9 residues 67-221, the alignment probability of TSP4 residues 82-227 is 97% with e-value of 1.5×10^{-4}, spanning residues that comprise two domains of gp9 analogous to D1 and D2 of gp10. Each of these T4 domains adopts a β-sandwich fold, forms homotrimers, and engages in complex interactions with other protein domains. As CBA120 contains three more TSPs in addition to TSP4, it is tempting to speculate that the TSP4 residues 82-329 contain three β-sandwich domains, which are required to mediate interactions with the three partner TSPs. Thus perhaps the first 82 residues are involved in the attachment to the base plate.

Amino acid residues 1-167 of TSP2 remain structurally undefined, as they are disordered in the crystal structure (section 4.5). The head domain (residues 168-245) adopts the same fold as subdomain D1 of TSP1 and TSP3 but lacks the D2 subdomain. Interestingly, BLAST alignment reveals sequence homology between residues 100-154 of TSP2, and residues 147-201 of G7C's gp66, corresponding to the predicted gp66 assembly domain. The sequence identity between TSP2 and G7c's gp66 is 32%, and the e-value is 1×10^{-3}. Also, HHpred reveals the similarity between the residues 16 – 161 of TSP2 and residues 160 – 389 of T4 gp10, spanning domains D2 and D3, with HHpred probability score of 96% but with rather high e-value of 0.034. Therefore, TSP2 N-terminal domain may also be involved in assembling the branched tailspike complex seen in the EM.
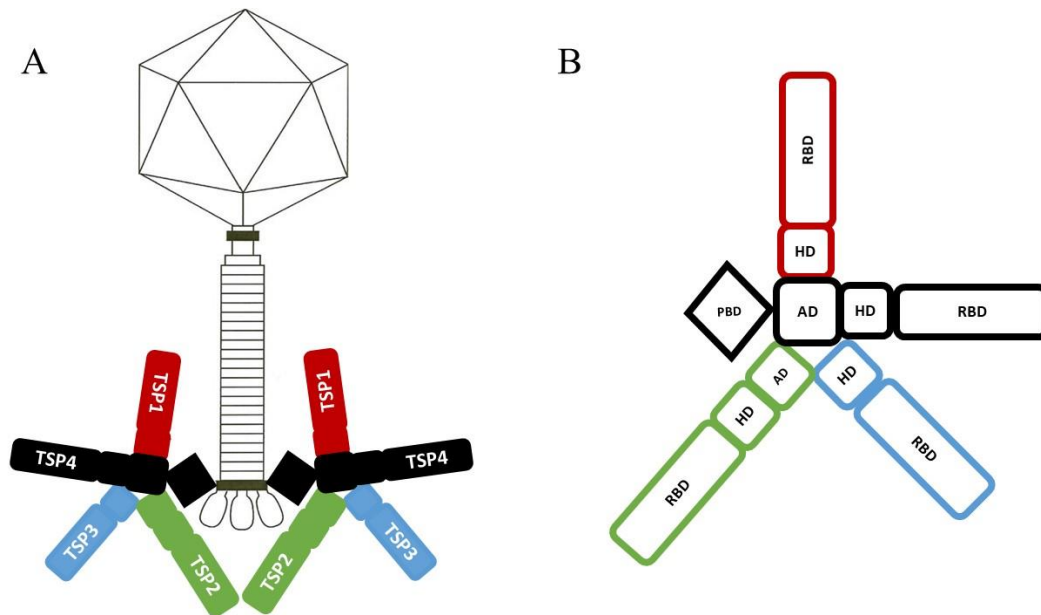
Figure 2.7 Proposed assembly of CBA120's TSPs.
(A) Proposed assembly of the star-like projections on CBA120 depict TSP4 (black) as the major assembly tailspike and TSP1(red), TSP2(green), and TSP3(blue) interacting with its assembly domain(s) (AD). (B) The tailspike apparatus is attached to the rest of the virus with the particle binding domain (PBD) of TSP4. TSP1 and TSP3 interact with the ADs via their head domains while TSP2 interacts with TSP4 via its AD. We hypothesized that the AD of TSP2 might facilitate the interaction of TSP3 (which will be discussed later). The C-terminal receptor binding domains (RBD) remain available to interaction with O-antigens.

Figure 2.7 depicts a proposed model for the arrangement of CBA120's TSPs based on sequence analysis and comparison of G7C's tailspikes and T4's gp9 and gp10. Future experiments to validate the model will include structural studies of TSP complexes and characterization of domain deletion constructs using analytical ultracentrifugation.

Chapter 3: Sequence Analysis of CBA120 Tailspikes

3.1 *Introduction*

In nature, proteins that exhibit sequence identity of ~30% tend to share the same overall three-dimensional structure. The sequences of phage TSPs are often unrelated, yet the catalytic domains of most form β-helix trimers. However, when there is no significant level of identity, the β-helix structures are quite diverse in detail, as evidenced by the inability to determine the structure of one TSP by molecular replacement using a search model of another TSP lacking significant amino acid sequence homology. The rapidly increasing number of phages sequenced raises an interesting question: To what extent new structures of TSPs resemble the structures of other known TSPs and how they differ. Sequencing of the bacteriophages from the *Viunalikevirus* genus has demonstrated that all members have substantial gene sequence similarities over their entire genome and encode similar proteins. The one area of significant divergence occurs in the region of the genome that encodes sequences assigned as TSPs[24].

3.2 *TSP1 and TSP3*

A Basic Local Alignment Search Tool (BLAST) protein search of the non-redundant database, against TSP1, retrieved four phage proteins with better than 93% sequence identity and 97% coverage. These include an unnamed protein product of the *Viunalikevirus* genus member *Salmonella* phage SFP10 (YP_004895331.1) and three non-*Viunalikevirus genus* members including a tail fiber from *Salmonella* phage S8 (ARB06442.1), tailspike protein of *Salmonella* phage STP07 (ARB06442.1), and a tail

fiber protein from *Escherichia* phage FEC14 (ATW66762.1). The homologous regions for the proteins mentioned above covered a majority (>90 %) of TSP1. However, there are other proteins with significant sequence similarity over distinctly smaller coverage range that all demonstrated an e-value of 5×10^{-19} or better. These areas of similarity can be divided into two groups with respect to the crystal structure of TSP1[38], one group showing sequence homology to the head domain (approximately 155 residues) and the other to the receptor binding domain (approximately 615 residues)[38].

Head domains homologous to that of TSP1 are present in numerous phage proteins including the tailspike protein of *Salmonella* phage vB_SalM_SJ3 (YP_009030454.1) which has 96% identity to TSP1 over the first 163 amino acids. In all, a BLAST search against the entire TSP1 amino acid sequence yielded about 85% of the top 100 hits with sequence identities to the head domain, including the TSP3 and TSP4 head domains of phage CBA120, which share ~70% sequence identity with TSP1 head domain, respectively. This level of similarity indicates that these domains share the same fold. The top 100 BLAST hits also contained 10% sequences that showed sequence identity to the receptor binding domain of TSP1, including *Salmonella* phage S115's tailspike (AXC40512.1), which has 93% sequence identity to TSP1 covering the first 480 residues of *Salmonella* phage S115's TSP. When the BLAST search is performed with the TSP3 sequence, the results are similar. About 84% of the top 100 hits exhibit sequence identity only to the N-terminal domain of TSP3, including the *Salmonella* phage 38 protein annotated as tail fiber (YP_009220801.1), which has 95% identity to the first 170 N-terminal amino acids of TSP3. About 16% of the BLAST results exhibited homology to

TSP3 receptor binding domain including *E. coli* bifunctional tail protein [WP_001544389.1], which shows 66% identity.

3.3 *TSP2 and TSP4*

A BLAST search reveals that the entire TSP2 protein sequence, including the N-terminal domain of unknown function, is 99% identical to that of several proteins from other bacteriophages: Proteins annotated as tail fibers of *Salmonella* phage vB_SalM_SJ3 (YP_009030453.1), *Salmonella* phage BSP101 (ARM69919.1), *Escherichia* phage EP75 (AVZ45056.1), *Salmonella* phage S8 (ARB06441.1), *Salmonella* phage S117 (AXC40876.1), *Escherichia* phage EP75 (AVZ45056.1), *Escherichia* phage PhaxI (YP_007002808.1), along with the tailspike 2 of *Salmonella* phage S115 (AXC40514.1), hypothetical proteins from *Escherichia* phage FEC14 (ATW66764.1), *Salmonella* phage GG32 (YP_009283775.1), and unnamed protein product of *Salmonella* phage SFP10 (YP_004895332.1). Although some of these proteins are annotated as tail fiber proteins, the enzymatic activity of TSPs from CBA120 described later suggests that these annotations might need revision.

Additionally, *Salmonella* ViI phage's tailspike Vi01_171c, a maturation-adhesion protein, and TSP2 share 91% sequence identity over the first 247 amino acids. This sequence identity suggests that the two proteins have a similar 3D structure for the domains they share. One can hypothesize that the N-terminal domain of TSP2 is important in the protein-protein interactions that assemble the multiple tailspike "star-like projections" seen at the base of *Viunalikevirus* phages. Interestingly, other Vi-type maturation-adhesion proteins

do not contain the N-terminal domain shared by ViI and TSP2; however, ViI is the only one of the *Salmonella Vi* phage type that encodes for three tailspike proteins.

Conversely, there is no homology between the two proteins beyond the first 240 residues, where Vi01_171c's tailspike encodes for a conserved acetyl esterase domain present in all the maturation-adhesion proteins of *Salmonella* Phage Vi types 1- 7. The tailspike for ViI phage, Vi01_171c is a maturation-adhesion protein, that has an acetyl esterase domain that modifies the acetyl groups on the exopolysaccharide capsule specifically, α-1,4-linked N-acetyl galactosaminuronate[57]. Maturation-adhesion proteins with acetylesterase activity disrupt the EPS capsule of *Salmonella* by destabilizing the carbohydrate structure and reducing hydrogen binding.

```
TSP2       1   MTRNVEELFGGVITAPHQIPFTYKSNVGGETFLSLPFYPVTGVVTINGGMQVPLDNFEIE
Vi01_171c  1   MTRNVEELFGGVITAPHQIPFTYKSSVGGETFISLPFYPVTGVVTINGGMQVPLDNFEID

TSP2      61   GNTLNLGRALSKGDVVYCLFDKILSPEDTAKGIRIYKFQAVGGETEFTPDFTSYGVQSLY
Vi01_171c 61   GNTLNLGRALSKDDVVYCLFDKILSPEDTSKGIRIYKFQAVGGETEFTPDFTSYGVQSLY

TSP2     121   IGGEYKTPEIEYSYDSTTGKVSLQTALSAGVWVVAEMSVKQPNISPAFDRSIQEIARSAN
Vi01_171c121   IGGEYKTPDIEYSYNSTTGKVSLQTALIAGVWVVAEMSVKQPNISPAFDRSIQEIARSAN

TSP2     181   VKDSEVIVSTDTISLLDGKKVVYDIATQTSYGLPTIPDGSVISSVSAGKINYNPGDVQVD
Vi01_171c181   VKDSEVIVSTDTISLLDDKKVVYDSAAQIIYGLPTIPDGSVISSVSDGKLTYNPGDVQVD

TSP2     241   LLPLEDSFINVINTLGRNDGAKYIGECHSVADLRNTEPTMDGQRIILKQHTAGTLLGGGV
Vi01_171c241   LIPVPTSAGALEIKLASEIGANGVGV--------------------GDTTVGEILKQDI
```

Figure 3.1 Alignment of the N-terminal domains of TSP2 and *Salmonella* phage maturation-adhesion protein.
The first 240 amino acids exhibit 91% sequence identity; thus they should adopt a similar structure. Multiple sequence alignment was produced by T-coffee alignment program [51].

38

TSP4 contains a longer N-terminal domain than TSP2, 335 amino acids long. *Salmonella* S8 phage (ARB06439) encodes a full-length TSP4 homolog with 99% sequence identity. Other *Salmonella* phages (SP1, vB_SalM_PM10, SFP10, GG32, vB_SalM_SJ3, Det7, SH19, vB_SalM_SJ2, SFL-SP-029, SFL-SP-063) exhibit 90-98% identity to TSP4 only within the N-terminal and head domains, underscoring the diversity of phage sequences and perhaps the exchange of genetic material. A goal of this study is to discover whether the N-terminal regions of TSP2 or TSP4 play a role in the assembly of the branched structures around the base plate. I have begun addressing this by producing the N-terminal domain of TSP4 for structural and biophysical characterization.

3.4 *The Sequence Diversity Within CBA120 TSPs*

Despite the difference in the TSP-encoding region, some similarities make identifying these TSPs easier. The head domain of CBA120's TSPs, spans 15-25% of the overall polypeptide and exhibit sequence identity in this domain. The head domain precedes the catalytic domain of the TSPs and consists of approximately 155 amino acids. The crystal structure of TSP1 from CBA120 published previously (Chen *et. al*[38]) revealed a head domain structure unlike any previously published structure. Comparison of the head domain sequences of TSP1-4 shows 70% or greater sequence identity between TSP1 and TSP3 and TSP4. Whereas the sequence identity between the head domain of TSP1 and TSP2 is limited to only one subdomain of the head domain (D1), of which they share 39 % sequence identity. The sequence identity of the TSP3 and TSP4 head domains to that of TSP1 is sufficient to conclude that their overall three-dimensional structures are similar. With respect to TSP2, the sequence identity of 39% is also sufficient to expect a similar

fold of the D1 domain only. The lack of sequence homology between the D2 subdomain of TSP1 and TSP2 indicated that TSP2's head domain is different. In section 4.5 we will learn that TSP2 lacks the D2 subdomain completely.

```
TSP1    1  MNEMFSQGGKGSTGILTNKQAVARHFGVKQSEVVYFSVGVD-LGGYKVIYDKETQRAYSL
TSP2  161  ------Q-PNISPAFDRSIQELARSANVKDSEVIVSLDTISLLDGKKVVYDIATQTSYGL
TSP3    1  MISQFNQ-ERGSTSIEVNKQSIARNFGVKEDEVIYFTAGID-LSGFKVIYDESTQRAYSL
TSP4  331  ING---T-PNIYNQIDYTLREVARVTNVKDTEVIYFSVGAV-LSGYKVIYDKVTQRSYFI

TSP1   60  PVGIASGTTAVSLSTAAVLVHSAGSVDLGSLAVSREEYVTLPGSFDSGSTLNVKNELLTY
TSP2  215  PT-IPDGSVISSV-SAGKLNYNEGDVQVDLLPLE-------------DSFINVINTLGRN
TSP3   59  PFGIVSGTTAISLDERAIILTHSAGSVDLGELAVSREEYVTLPGSFNFGHTINVKNELLVH
TSP4  387  PE-LPTGTTAVSLSSSAILVHSAGSVDLGALAVSREEYVTLSGIFDSGAVINTKNELLTH

TSP1  120  TDGKYRWDGILPKTVAPGSTPASTGGVGLGAWISVGDASLRTQLANG--DG-SLIG----
TSP2  260  DGAKYI----------------------GECHSVADLRNTEPTM----DG---------
TSP3  119  DDKKYRWDGSLPKVVAAGSTPDSSGGVGLGAWLSVGDAALRAELNTKVSDG-TFPATIKY
TSP4  446  TDGKYRWDGILPKTVAAGSTPAITGGVGSGAWLSVGDASLKSNLNKP--NGLSYIGTVSS
```

Figure 3.2 Sequence alignment of TSP 1-4 head domains.
Sequence alignment reveals the area of the most sequence similarity spans the head domain. The location of the head domain along the polypeptide chain varies depending on whether the TSP is an assembly TSP or a branching TSP. Multiple sequence alignment was produced by T-coffee alignment program [51].

Table 3.1 Characteristics of CBA120 Tailspike proteins

	orf 210	orf 211	orf 212	orf 213
Name	TSP1	TSP2	TSP3	TSP4
Size (amino acid)	770	922	628	1037
Start of the head domain (residue number)	1	173	1	345
% Sequence Identity to orf210 head domain	100	39*	70	72

*Sequence identity to TSP1's D1 domain only

40
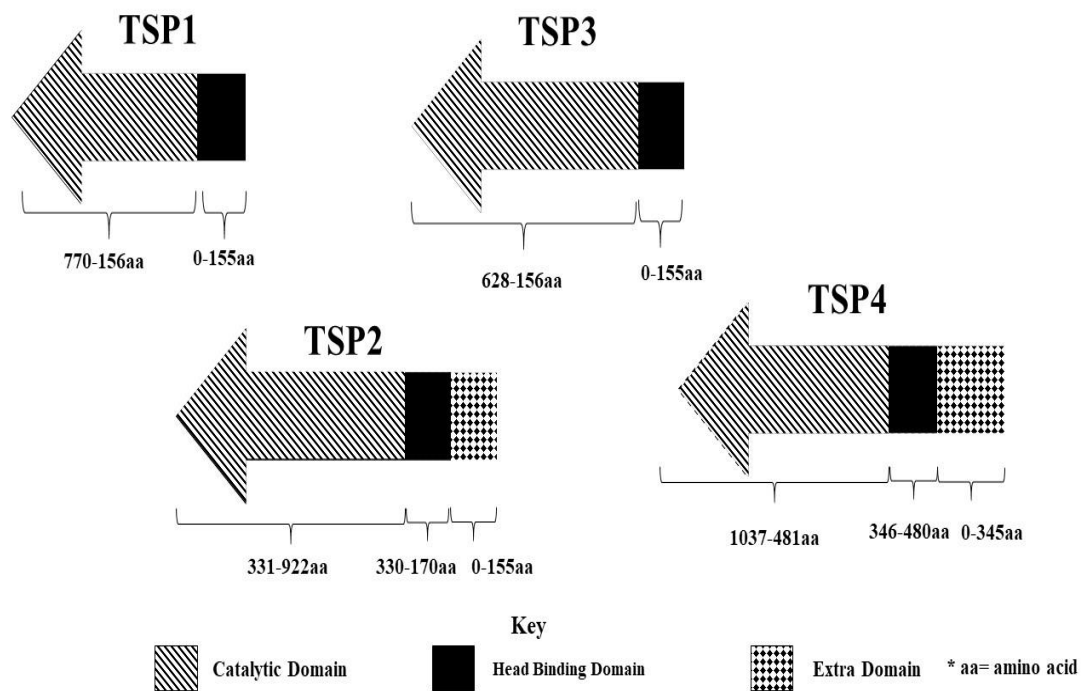
Figure 3.3 Domain constructs of CBA120's TSP's
This figure shows two structural domains for TSP1 and TSP3 and the additional N-terminal region for TSP2 and TSP4.

Chapter 4: Biophysical Characterization and Crystal Structures of CBA120 Tailspikes

4.1 *Introduction*

Chapter 4 describes tailspike stability studies, early investigation of interactions between the proteins, and crystallographic studies. I have contributed to the crystallographic studies of all four CBA120 TSPs. For TSP1, I obtained the crystals that were used in data collection. However, the structure determination was primarily carried out by another group member (Chen et al. 2014 [38]). Although I co-authored the TSP1 paper, this dissertation does not describe the work in detail. Included here are the structural studies of the remaining TSPs, TSP2, TSP3, and TSP4, performed by myself, acknowledging the contributions of others as appropriate.

4.2 *Materials and Methods*

4.2.1 Cloning, Expression, and purification of TSP2, TSP3, and TSP4

The nucleic acid sequences of TSP2-4 were codon optimized for expression in *E. coli* and synthesized with a C-terminal 6X-His tag using GeneArt. The genes were sub-cloned into a pBAD24 expression vector, and the resulting clones were used to transform Rosetta-gami™ two- competent cells[38]. The transformed cells were used to inoculate 2x YT medium supplemented with 200 μg/mL ampicillin at 37º C for 4-8 hours until the OD_{600} reached 0.8. The temperature was lowered to 18ºC, and the culture was induced with 0.25% arabinose overnight. The biomass was collected by centrifugation and resuspended in 1x PBS buffer and lysed by sonication. The TSP-enriched supernatant was separated by

centrifugation at 14,000 rpm for 45 minutes. The resulting supernatant was incubated in Ni-NTA agarose for one hour at 4º C after which, the TSP was purified by gravity-flow chromatography and dialyzed in Tris-HCl buffer pH 7.5 and concentrated. The expression of selenomethionine (SeMet) containing TSP3 followed the same protocol except that the cells were grown in SeMet™ medium (Molecular Dimensions Limited) supplemented with 40 mg/L SeMet.

4.2.2   Site-directed Mutagenesis
TSP3 mutants were produced by the Nelson Lab using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). Primers (**Error! Reference source not found.**) were designed to include mutations in the middle of a 30 nucleotide primer with the phosphorylated reverse primers complementary to the next 30 nucleotides upstream. The resulting PCR products were digested by *DpnI* to remove the methylated templates, ligated and transformed into *E. coli* DH5α. The mutations were confirmed by nucleotide sequencing (Macrogen, USA) before being transformed into *E. coli* for protein expression.

4.2.3   Size-exclusion Chromatography
The multimeric state of CBA120's TSPs was assessed by analytical size-exclusion chromatography. The Superdex 200 column (GE Healthcare) was equilibrated with PBS and TSP was applied to the column. The samples were run under isocratic conditions in PBS for 1.5 column volumes on an AKTA FPLC system (GE Healthcare). The molecular mass of TSPs was estimated from a standard curve generated by linear regression of log (molecular mass) vs. retention volume using gel filtration standards (Bio-Rad).

4.2.4   Protein Stability
The thermal stability of TSP3 and TSP2 were investigated by the Nelson lab using a Chirascan CD Spectrometer (Applied Photophysics). 0.1 mg/mL of a TSP sample in 20

mM sodium phosphate buffer, pH 7.0, was gradually heated at a rate of 1 °C/min from 20 °C to 95 °C. The mean residue ellipticity (MRE) of the sample contained in a quartz cuvette of 1 mm path length was monitored every 0.5 °C at 218 nm with 5 s signal averaging per data point. The data was then smoothed, normalized, and fit with a Boltzmann sigmoidal curve using the Pro-Data software (Applied Photophysics). The melting temperature (Tm) was determined by calculating the first derivative, which was determined from the minimum value of the derivative graph.

TSP3 sensitivity to LDS was assessed by incubating 0.25 mg/mL of TSP3 in lithium dodecyl sulfate (LDS) sample buffer (Invitrogen) (0.5% final LSD concentration) for 10 mins at room temperature or boiled. The resulting sample was analyzed on a 7% SDS-PAGE. TSP3 resistance to proteolysis was analyzed using 0.25 mg/mL samples incubated at 37 °C for one day with 0.5 mg/mL trypsin (Sigma Aldrich) or chymotrypsin (Sigma Aldrich) at a ratio of 1:25 (w/w) protease and TSP3. The protease reaction buffer contained 20 mM sodium phosphate, pH 7.0, and 1 mM $CaCl_2$. Bovine serum albumin BSA (Sigma Aldrich) served as control and was incubated overnight with and without protease. The resulting samples were analyzed on 7% SDS-PAGE.

4.2.5 Pull-down Experiment with TSP4-N and Branching TSPs

In order to test TSP4-N's association with TSP1 and TSP3, a pull-down experiment was conducted by the Nelson group. His-tagged TSP4-N was equilibrated on a Ni-NTA agarose column and washed to remove any unbound protein. Then tag-free TSP1 or TSP3 enriched lysate was applied to the column and washed again to remove any unbound tailspike. The column was eluted with a step gradient of increasing imidazole in PBS.

4.3 *Protein Purification and Characterization*

4.3.1 Purification of TSPs

Cloning and purification of TSP2 and TSP3 followed the standard protocol described in the Material and Methods section 4.2.1. Several milligrams of full-length TSPs were produced and purified using affinity chromatography (see Figure 4.2A-B). However, full-length TSP4 was not expressed using this expression protocol. Several TSP4 constructs were designed for cloning and expressing smaller constructs of TSP4 (Figure 4.1). The constructs were designed using the sequence identity between TSP4 and TSP1; and the secondary structure prediction by PSIPRED[58, 59] as a guide in determining the domain boundaries. Of the seven constructs three were cloned (options 1, 3, and 7), and option 1 expressed sufficient protein for further investigation. Option 1 contains the first 335 residues of TSP4 and spans the predicted particle binding domain and the assembly domain and hereby referred to as TSP4-N. Similar to TSP2 and TSP3, several milligrams of soluble TSP4-N was produced and purified Figure 4.2C.
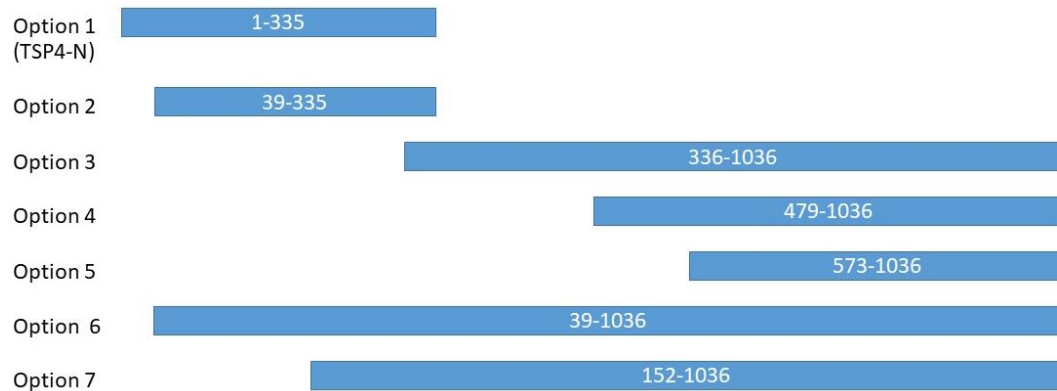
Figure 4.1 Constructs of TSP4.

Seven contracts of TSP4 were designed based on predicted domain boundaries. Option 1, referred to in this work as TSP4-N, is predicted to contain two domains: the particle binding domain and the assembly domain.
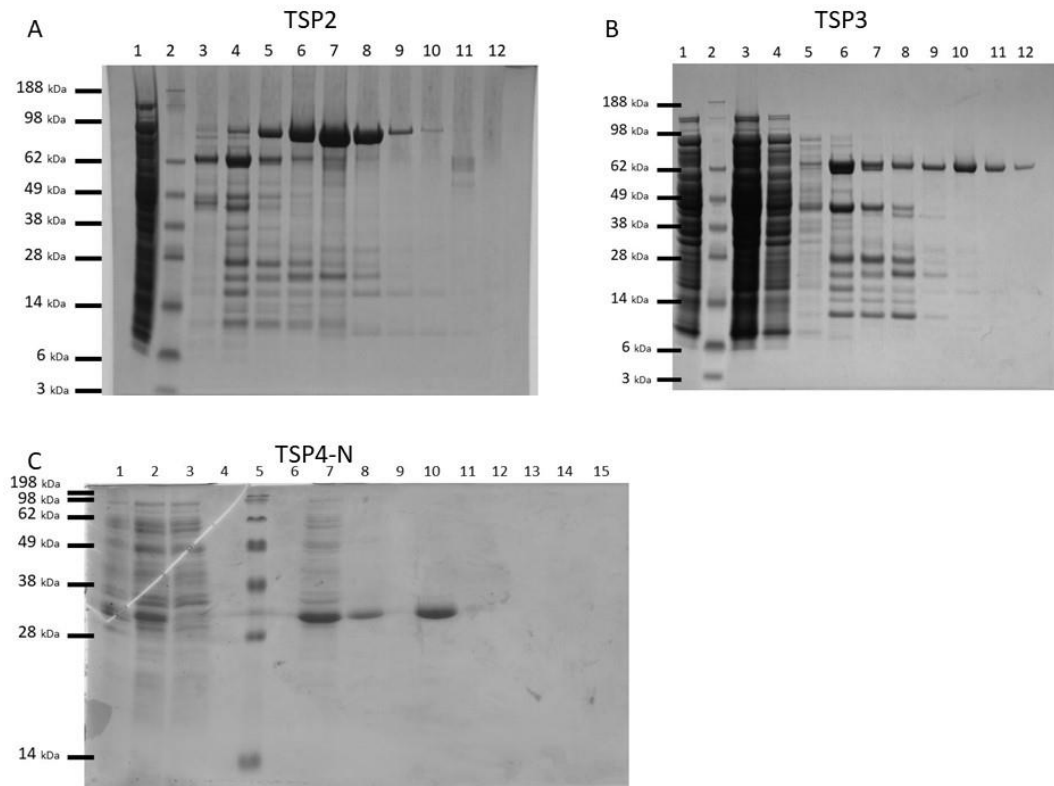
Figure 4.2 SDA-PAGE gel of TSP2-4 purification.
SDS-PAGE following the purification of 6X-his C-terminal tagged TSP2-4 after Nickel-NTA agarose binding and elution using an imidazole step gradient. (A) SDS-PAGE of TSP2 purification, Lanes: flow through (1); protein ladder (2); 50 mM imidazole (3); 50 mM imidazole (4); 100 mM imidazole (5); 100 mM imidazole (6); 150 mM imidazole (7); 150 mM imidazole (8); 200 mM imidazole (9); 200 mM imidazole (10); 250 mM imidazole (11); 250 mM imidazole (12).

(B) SDS-PAGE of TSP3 purification, Lanes: flow through (1); protein ladder (2), 20 mM imidazole (3), 20 mM imidazole (4); 50 mM imidazole (5); 50 mM imidazole (6); 100 mM imidazole (7), 100 mM imidazole (8); 150 mM imidazole (9); 150 mM imidazole (10); 200 mM imidazole (11); 200 mM imidazole (12).

(C) SDS-PAGE of TSP4-N purification, Lanes: Dilute whole cell lysate (1); supernatant (2), flow through (3), 20 mM imidazole (4); protein ladder (5); blank (6); 50 mM imidazole (7), 50 mM imidazole (8); blank (9); 250 mM imidazole (10); 250 mM imidazole (11); blank (12); 500 mM imidazole (13); 500 mM imidazole (14); blank (15).

4.3.2 Size-exclusion Chromatography

One common feature of tailspike proteins is their quaternary structure which forms trimers. Size-exclusion chromatography of TSP2 reveals a single homogenous peak near the void volume indicating a multimeric protein, (see Figure 4.3). The calculated weight according to the chromatogram is 470 kDa, which is far than greater the calculated trimeric weight of TSP2 (297 kDa); due to the poor accuracy of this method, we can only conclude that TSP2 is not a monomer in solution. TSP3 triple mutant (E362Q, D383N, and D426N) expressed and produced two peaks on the chromatogram. The earliest peak (8 mL) represents higher order aggregates and the second peak (12.5 mL) is consistent with a multimeric protein of ~270 kDa. Again, the exact molecular weight cannot be determined accurately using this method, but TSP3 triple mutant is expected to form a trimer since wild-type TSP3 is a trimer in solution and the crystal structure with a predicted molecular weight of 205 kDa. TSP4-N monomeric weight is calculated to be (~36 kDa including the 6x-His tag). The first peak at (10.5 mL) indicates higher order oligomers. The second peak in the TSP4-N chromatogram, (14.5 mL) represents a smaller oligomer of TSP4-N. The calculated molecular weight for the two TSP4-N species is 590 kDa for the first peak, and 66 kDa for the second according to the chromatogram. These results indicate that all TSPs examined form oligomers; however, due to the low accuracy of this method partly, due to the elongated shape of the TSPs, more information cannot be elucidated. In order to obtain a more accurate estimation of the oligomeric state, Dr. Chao, in unpublished data, performed analytical ultracentrifugation experiments. She determined that TSP2 and TSP3 form trimers in solution and TSP4-N forms both trimers and hexamers in solution.
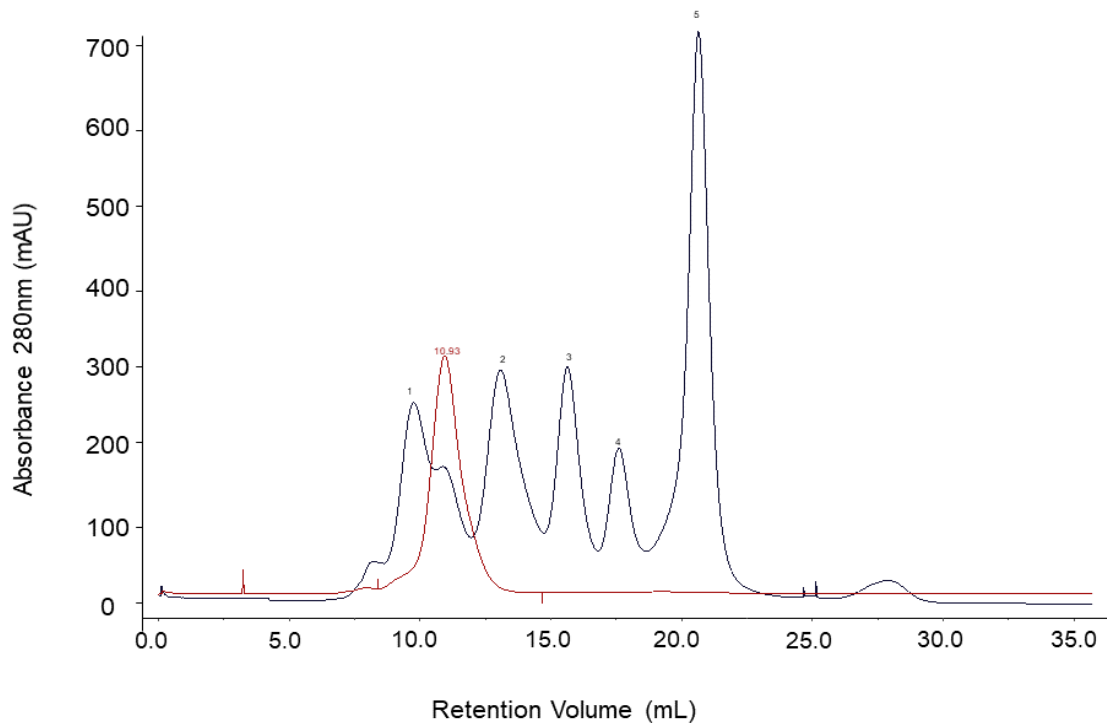
Figure 4.3 TSP2 Size-exclusion chromatography.
The peak of purified TSP2 is shown in red. The actual molecular weight for TSP2 is 99
kDa for a monomer and 297 kDa for a trimer. Molecular weight standards (black) are as
follows: (1) Thyroglobulin, 670 kDa (void volume); (2) Gamma globulin, 158 kDa; (3)
Ovalbumin, 44 kDa; (4) Myoglobin, 17 kDa; (5) Vitamin B_{12}, 1.35 kDa
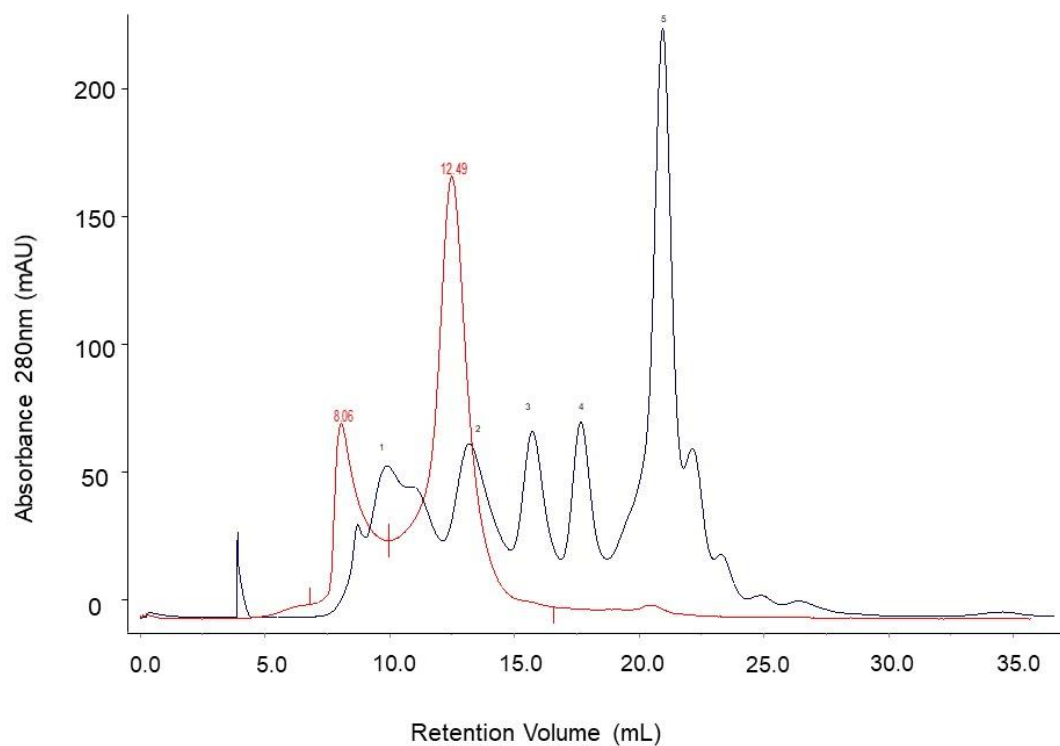
Figure 4.4 TSP3 Triple Mutant size-exclusion chromatography.
The peak of purified TSP3 triple mutant is shown in red. The actual molecular weight for TSP3 triple mutant is 69 kDa for a monomer, 205 kDa for a trimer. Molecular weight standards (black) are as follows: (1) Thyroglobulin, 670 kDa (void volume); (2) Gamma globulin, 158 kDa; (3) Ovalbumin, 44 kDa; (4) Myoglobin, 17 kDa; (5) Vitamin B$_{12}$, 1.35 kDa.
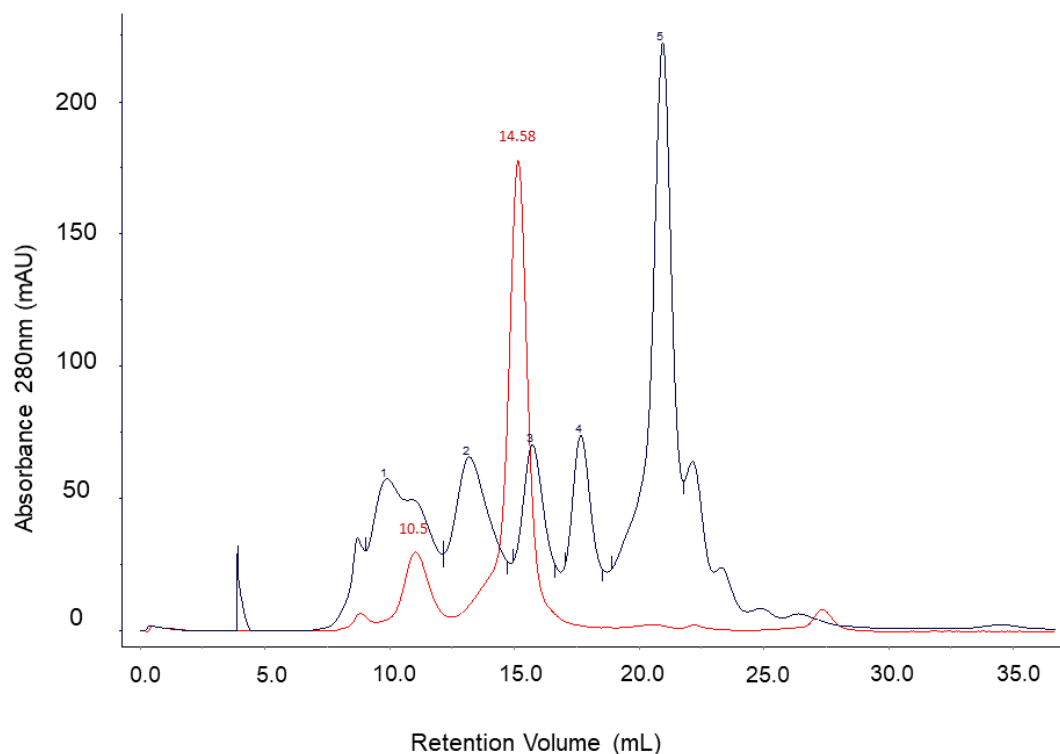
Figure 4.5 TSP4-N size-exclusion chromatography.
The peak of purified TSP4-N is shown in red. The actual molecular weight for TSP4-N is 36 kDa for a monomer, 108 kDa for a trimer, and 216 kDa for a hexamer. Molecular weight standards (black) are as follows: (1) Thyroglobulin, 670 kDa (void volume); (2) Gamma globulin, 158 kDa; (3) Ovalbumin, 44 kDa; (4) Myoglobin, 17 kDa; (5) Vitamin B_{12}, 1.35 kDa.

4.3.3 Protein Stability of TSP2 and TSP3

The thermal stability of TSP2 and TSP3 was determined by the Nelson Lab using CD Spectroscopy at 218 nm to monitor the loss in β-sheet content as the temperature increases from 20°C to 95°C. The melting temperatures (T_m) for TSP2 and TSP3 were determined to be 82.8 °C and 61.8 °C, respectively. The single inflection point of the derivative curve indicates that both proteins unfolded cooperatively concomitant with the trimer dissociation. Interestingly, the T_m for TSP3 is significantly lower than other T_m values

51

reported for other tailspike proteins. TSPs are usually characterized as extremely stable proteins, presumably because they are exposed to harsh environmental conditions. Published T_m values for other beta helical tailspike proteins have all been over 80 °C including; TSP1 (T_m=80.7 °C), Phages P22 TSP (T_m =88.4 °C) and HK620 TSP (T_m = 80 °C) [38, 40, 60].

A   B
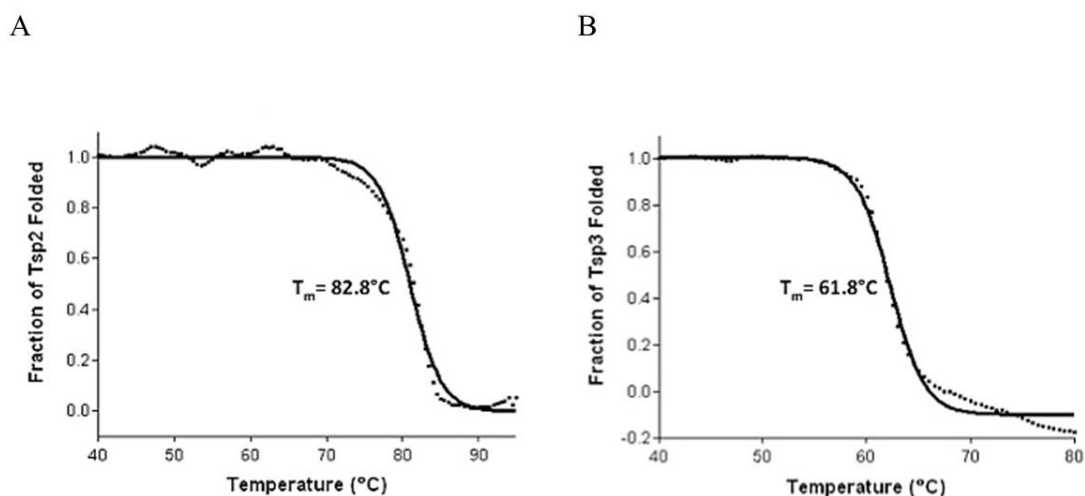
Figure 4.6   Unfolding curve of TSP2 and TSP3 as a function of temperature monitored by CD spectroscopy.
(A) Tm of TSP2 is 82.8°C which is consistent with the reported Tm values of other TSPs. (B) The Tm of TSP3 is 61.8°C, significantly lower than other TSPs. The experimental values are indicated by circles and the continuous line corresponds to the theoretical curve.

Lithium dodecyl sulfate (LDS) was used to test the stability of the TSP3 trimer further. TSP3 maintained its trimeric association when exposed to the anionic detergent, as shown in lane 6 of Figure 4.7. The trimer dissociates only when the protein was boiled as shown in lane 5. To determine TSP3 resistance to proteolytic degradation, the protein was incubated with trypsin and chymotrypsin for one day, and the corresponding SDS-PAGE of the samples are shown in (Figure 4.7) lanes 7-10. Lanes 7 and 9 show the boiled sample

52

of TSP3 incubated with trypsin and chymotrypsin respectively; the TSP3 band corresponds to a monomer similar to the band shown in lane 5 which was not treated with protease. Lanes 8 and 10, which correspond to the un-boiled samples, show an additional band ~50 kDa lower than the molecular weight of the TSP3 trimer, but still higher than the molecular weight of a monomer. In contrast to TSP1, which did not exhibit any proteolytic degradation, TSP3 was partially degraded as indicated by band below the full-length trimer.
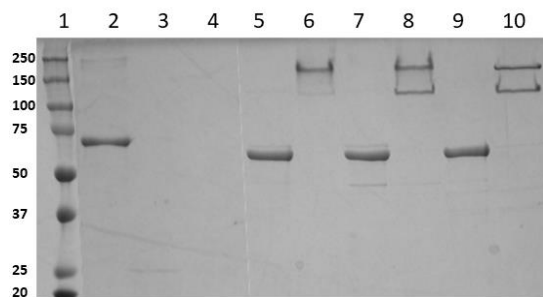
Figure 4.7 TSP3's Resistance to Proteolysis
TSP3 undergoes limited proteolysis by trypsin and chymotrypsin upon 24 hr. incubation. Lanes: protein molecular weight ladder (1); Bovine serum albumin (BSA) control (2); BSA +trypsin (3); BSA+chymotrypsin (4); TSP3, boiled (5); TSP3, not boiled (6); TSP3+trypsin, boiled (7); TSP3+trypsin, not boiled (8); (9-10) same as (7-8) but TSP3 incubated with chymotrypsin.

N-terminal sequencing of the protein extracted from the lower molecular weight bands from the non-boiled samples treated with trypsin and chymotrypsin (Figure 4.7 lanes 8 and 10) indicated that the N-termini was intact. When these samples are boiled, only one protein band is observed representing the full-length TSP3 monomer. A single cleavage per trimer on a loop after the neck but before the β-helix of the D3 subdomain can explain the presence of the lower band in the non-boiled protease samples. Cleavage at the position mentioned above results in the dissociation and subsequent degradation of the released

catalytic domain ~ 50kDa. The structure of TSP3 does show two potential cleavage sites at Lysine 176 and Tyrosine 190 on a solvent-exposed linker.

4.3.4   Interactions between TSPs

For unknown reasons, the expression of full-length TSP4 was unsuccessful. However, a construct of the N-terminal domain of TSP4 was created to probe the role of TSP4 in assembling the tailspike complex. Sequence identity between TSP1 and TSP4 indicates that the head and catalytic domains of TSP4 begin after residue 335. Using G7C's gp66 as a model along with HHpred's structure prediction suggests that residues 82 – 350 comprise a gp10-like module, we predict that the first 335 amino acids contain two structural features: a particle binding domain and an assembly domain(s) that mediates the interaction with TSP1, TSP2, and TSP3. This TSP4 construct (amino acids 1-335), now called TSP4-N was expressed, purified, and used to determine its ability to bind the remaining TSPs of phage CBA120. TSP4-N primarily forms a trimer in solution and a small amount of hexamers as measured by size exclusion chromatography.

Pull-down experiments, performed by the Nelson Lab, showed that TSP4-N and TSP1 co-eluted together at 100 mM imidazole indicating a binding affinity between the two proteins. Conversely, TSP4-N and TSP3 pull-down experiments did not show interaction. All the TSP3 is found in the TSP3 flow through, and the absorbed TSP4-N elutes alone (Figure 4.8). Recent analytical ultracentrifugation studies performed by Dr. Chao confirmed the interaction between TSP4-N and TSP1 and also showed binding of TSP2 to TSP4-N. TSP3

binds to the complex only in the presence of TSP4-N and TSP2; explaining the negative

pulldown result for TSP3.
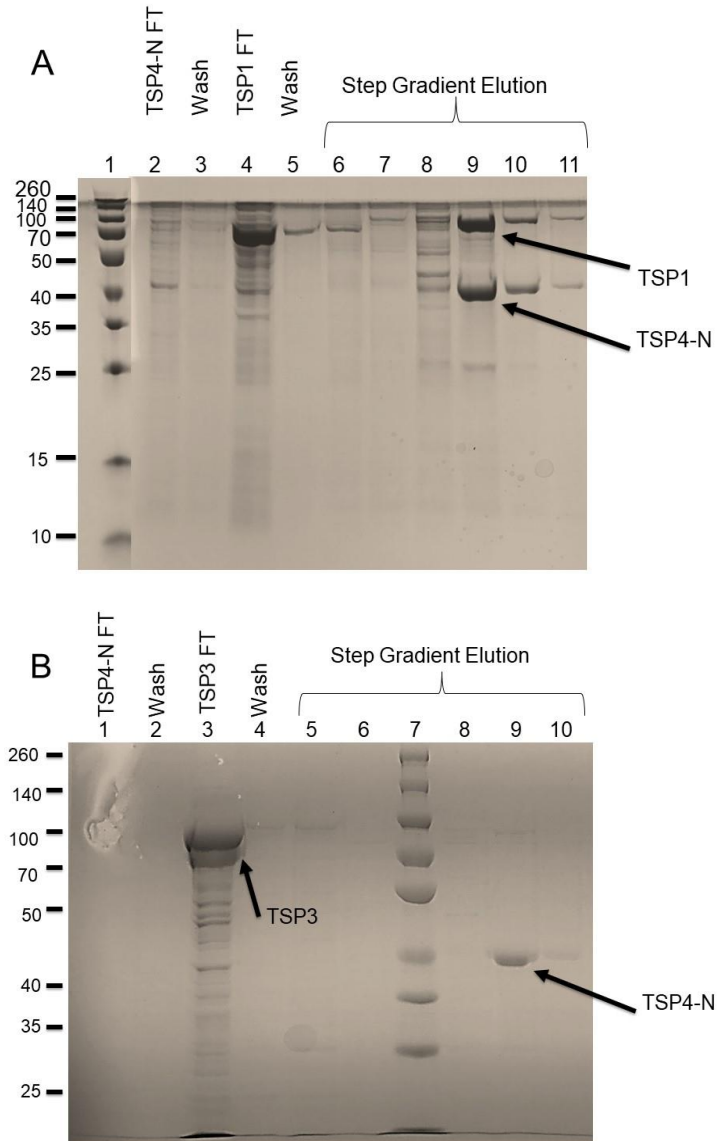
Figure 4.8 Pull-down with His-tagged TSP4-N and partner TSPs.
SDS-Page of pulldown experiment performed with His-tagged TSP4-N and non-his-tagged TSP1 or TSP3.

 SDS-Page analyses of TSP1 pulldown (A). Lanes: (1) protein molecular weight markers, (2) TSP4-N lysate flow through, (3) wash, (4) TSP1 lysate flow through, (5) Wash, 6-11 elution with increasing imidazole concentration: (6-7) 20 mM, (8) 50 mM, (9) 100 mM, (10) 250 mM, (11) 500 mM.

SDS-Page analyses of TSP3 pulldown (B). Lanes: (1) TSP4-N lysate flow through, (2) wash (3) TSP3 flow through, (4) Wash, (5-6) 20 mM Imidazole, (7) molecular weight markers (8) 100 mM Imidazole, (9) 250 mM Imidazole, (10) 500 mM Imidazole.

4.4 *TSP1 Crystal Structure - Overview*

Chen et al.[38] determined the structure of full-length TSP1. TSP1 assembles into an elongated homotrimer with two distinct domains, the head and the receptor binding domains (see Figure 4.9). The three monomers display extensive interaction burying almost a quarter of the surface area[38]. TSP1's head domain includes residues 12-155, which can be further divided into two subdomains; D1 (residues 12-96) and D2 (residues 97-154). A Zn^{2+} ligand was identified by anomalous diffraction at the zinc absorption edge. The zinc ion interacts with TSP1 by coordinating the His25 imidazole group on each monomer and a water molecule in a tetrahedral geometry. His25 is buried close to the C-terminus of an α-helical bundle formed by the three monomers.

TSP1 C-terminal domain, the receptor-binding domain, spans residues 166-770. The primary structural motif of this domain is a right-handed three-sided β-helix present in both subdomains D3 (residues 166-562) and D4 (residues 624-769). The D3 subdomain was predicted to contain the active site, situated in the groove between two monomers[38]. The proposed active site residues include two carboxylic residues Glu456 and Glu483 about 2.8 Å apart on the same monomer and Asp313 which resides on an adjacent monomer. Asp 313 is ~11 Å and ~10 Å apart from Glu456 and Glu483 respectively, and between the two monomers is a groove that could accommodate the polysaccharide substrate. Based on

the arrangement of the carboxylic residues ~10 Å apart, the inverting mechanism for glycosyl hydrolysis could explain TSP1 enzymatic activity, discussed in more detail in chapter 5. Also, several aromatic residues surrounding the postulated active site may be necessary for stabilizing the sugar ring(s) of the polysaccharide substrate including Tyr411 and Trp380. The D4 subdomain is unique to the TSP1 structure. It forms a β-helix, which has a 30-degree bend relative to the D3 subdomain. The bend creates a solvent assessable hole in the center of the trimer not seen in other tailspike structures Figure 4.9.
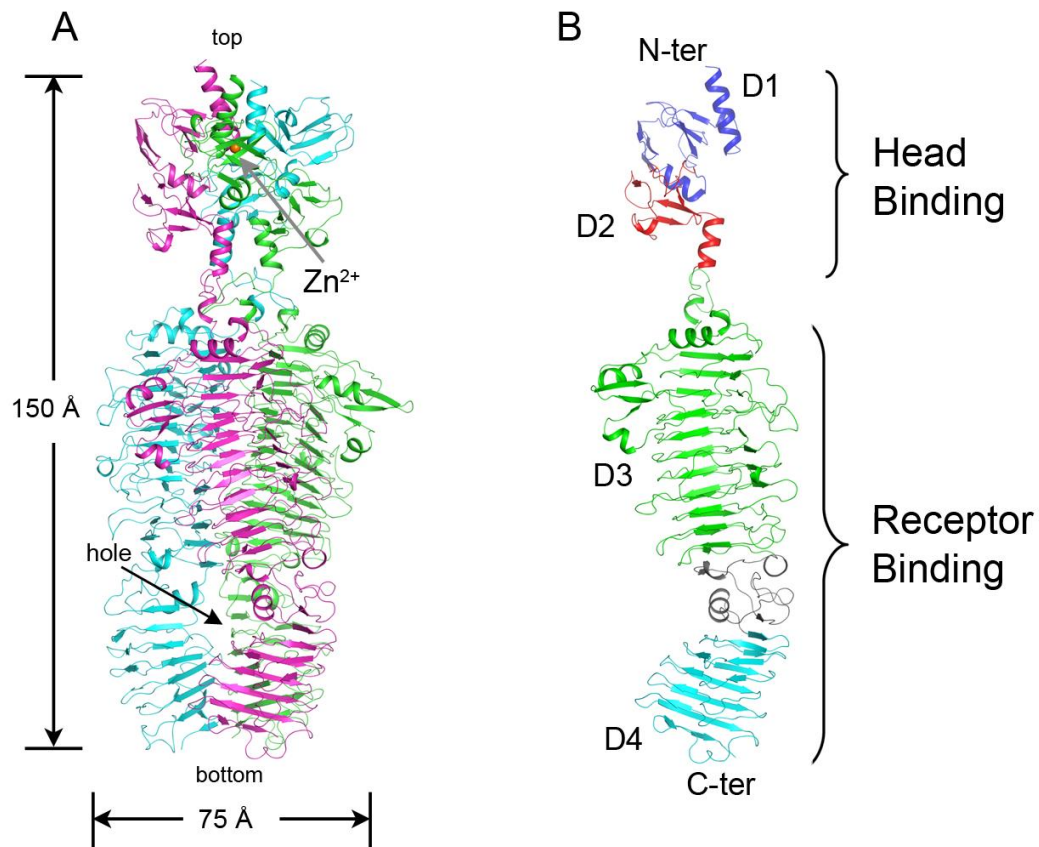
Figure 4.9 The structure of TSP1.
Cartoon representation of TSP1 homotrimer (A) and monomer (B)[38].

4.5 *TSP2 Crystal Structure*

4.5.1 Crystallization, Data Collection, and Structure Determination

A TSP2 crystal was obtained by the vapor diffusion method at room temperature in sitting drops. The reservoir solution contained 0.8 M ammonium sulfate and 0.1 M HEPES, pH 7.0. Only one single crystal appeared after five weeks (Figure 4.10). Further attempts to obtain more crystals yielded small crystals unsuitable for data collection, and the SeMet protein did not yield crystals at all. The TSP2 crystal was cryoprotected by adding 3µL of the reservoir solution supplemented with 30% glycerol to the drop. The crystals were then flash cooled with liquid nitrogen. A data set of a TSP2 crystal was collected at Argonne National Laboratory, General Medical Sciences and National Cancer Institute (GM/CA) beamline 23-ID_B equipped with a MARmosaic MX-300 detector (Marresearch). The data was processed using XDS[61] and scaled using Aimless[62]. The structure was determined by molecular replacement using Phenix[63] and the structure of the TSP2 receptor binding domain as the search model (kindly provided by Dr. Petr Lieman). Completion of the polypeptide chain and further modification were performed with Coot[64], and the structure was refined using Phenix[65] and Refmac[66].
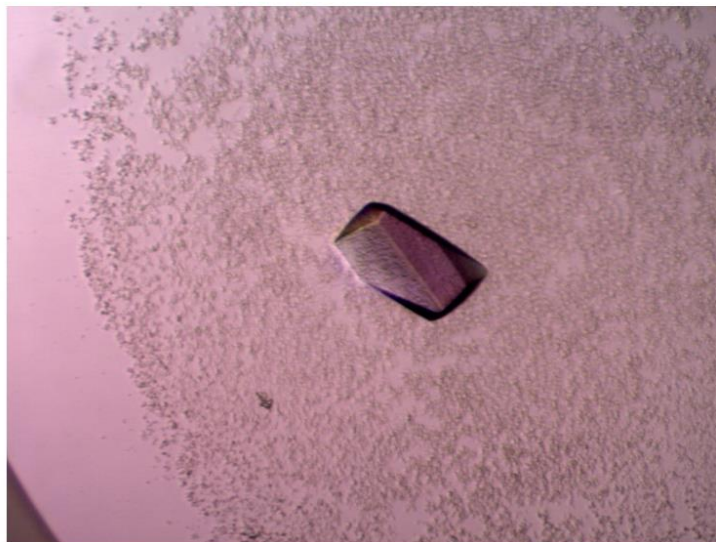
Figure 4.10 Protein crystal of TSP2.
Crystal of TSP2 crystalized 0.8 M ammonium sulfate and 0.1 M HEPES, pH 7.0.

4.5.2   Overall Structure of TSP2

TSP2 crystal structure was determined and refined at 1.9 Å resolution yielding an R-factor

of 0.168 and Free-R of 0.188 with good geometry. Table 4.1 provides a summary of the

data collection and refinement quality. Analogous to other tailspike proteins, TSP2 crystal

structure assembles into a homotrimer. The first 168 residues have no associated electron

density and therefore are not included in the model. As the crystals were not reproduced,

it is unknown whether this region was cleaved by an *E. coli* protease that contaminated the

sample or whether this region is structurally disordered. The C-terminal 6x-His affinity tag

is mostly disordered, and only a few histidine residues are seen in the final structure. The

asymmetric unit contains two homotrimers, providing two independent views of the

biological trimer. The trimer adopts an elongated shape, 170 Å long by 70 Å at the widest

point. The trimer association buries over 22,000 Å^2 surface area. Two structural domains are clearly distinguished in Figure 4.11. The head domain (residues 168-257), consists of three lobes, one from each monomer. The second domain comprises the receptor binding domain (258-921).
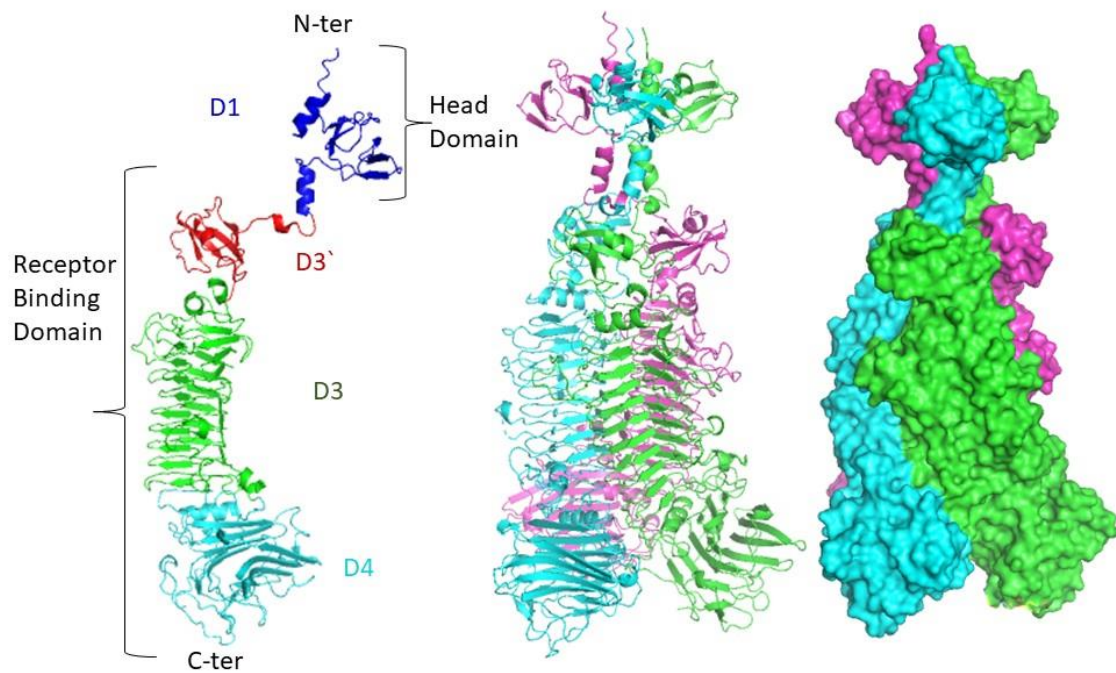
Figure 4.11 Structure of TSP2.
A cartoon representation of the TSP2's monomer (left) and trimer (middle). The surface representation of the TSP2 homotrimer is depicted on the right.

Table 4.1 Statistics of data collection and refinement of CBA120 TSP2.

Data Collection	TSP2
Wavelength (Å)	1.0332
Resolution (Å) [a]	48.36 - 1.9 (1.968 - 1.9)
Space group	P 21 21 21
Unit cell dimensions (Å)	a=83.0 b=259.0 c=269.7
No. of unique molecule in the asymmetric	6
No of unique reflections [a]	430672 (45098)
Multiplicity [a]	5.0 (5.0)
Completeness (%)[a]	99.74 (99.98)
Mean I/σ(I)[a]	9.58 (1.41)
Rmerge[a]	0.126(1.711)
Refinement	
Resolution (Å) [a]	48.36 - 1.9 (1.949 - 1.9)
Total No. of Reflections	430672 (31614)
R*work*/R*free*	0.168 (0.319)/ 0.188 (0.316)
No of Protein Residues	4533
Ligands	117
solvent	3887
RMSD from ideal geometry	
bond length(Å)/bond angles (°)	0.011/1.74
Ramachandran plot:	95.13/4.62/0.24
Molprobity overall score & percentile	1.90/76th

[a]The values in parentheses are for the highest resolution shell.

[b]$R_{merge} = \Sigma_{hkl}\Sigma_j|I_j(hkl) - <I(hkl)>| / \Sigma_{hkl} \Sigma_j I_j(hkl)$.

$R_{work} = \Sigma_{hkl} | F_o - F_c | / \Sigma_{hkl} F_o$, where F_o and F_c are the observed and calculated structure factors, respectively.

R_{free} is computed from 4.94% of randomly selected reflections (8672) and omitted from the refinement.

[c]Correlation coefficient between E_c and E_o in SHELXD[67].

[d]Molprobity geometry score and percentile correspond to PDB structures within the refinement resolution range[68].

4.5.3   D1 Head Domain

Consistent with the amino acid homology, the head domain of TSP2, referred to here as

the D1 domain, has the same fold as the D1 subdomain of TSP1 (and as described later,

the D1 of TSP3's head) with an RMSD of 2.2 Å for α-carbon atoms in the D1 domain.

TSP2's D1 begins with an α–helix; the α–helices of the three monomers interact to form

a bundle. Unlike TSP1, TSP2 does not coordinate a Zn^{2+} ion or any other metal at the base

of the α–helix bundle. Proceeding the α–helix, TSP2's D1 domain forms an anti-parallel β–sandwich similar to the of the D1 domains of TSP1, TSP3, and gp63.1. However, TSP2 lacks the D2 subdomain of the homologous TSP head domains. Instead, another α-helical bundle (residues 248- 257,Figure 4.12) follows the Dl domain. Common to many tailspikes, the α–helix bundle connects the head domain to the receptor binding domain and has an appearance of the "neck" of the trimeric structure.
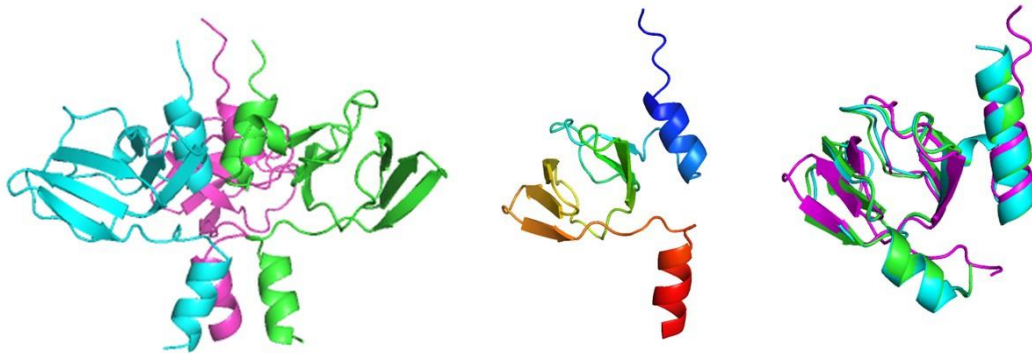
Figure 4.12 TSP2 Head + neck regions.
A cartoon representation of the head domain and neck trimer of TSP2 (left), rainbow representation of the head domain traced in rainbow colored from blue at the N-terminus to red at the C-terminus (middle). An alignment of D1 of TSP1 (cyan), TSP2 (magenta), and TSP3 (green) exhibiting the same fold.

4.5.4 Receptor Binding Domain

The receptor binding domain of TSP2 consists of three subdomains, unlike many other tailspikes counterparts, which only have two receptor binding subdomains. By analogy to other β–helical TSPs, this domain is presumed to possess glycolytic enzymatic activity. In contrast to TSP1 (and as described later also TSP3), whose receptor binding domain comprises two subdomains (D3 and D4), TSP2 receptor binding domain can be divided into three distinct subdomains. The first subdomain is made up of 4 anti-parallel β–strands

twisted around a short α–helix. Only 57 amino acids, the first subdomain, thereby called D3`, is not seen in other β–helical tailspikes and the function of this domain is not yet known (Figure 4.13B). The second subdomain is the D3 subdomain, analogous to other TSPs. D3 of TSP2 is 349 amino acids long, comprising a right-handed β–helix, and the three monomer wrap around one another to form a left-handed coiled coil. The C-terminal subdomain corresponding to D4 of other TSPs extends out from the main "body" of the receptor binding domain to form a propeller-like shape, see Figure 4.13.

Ensuing the neck, the additional subdomain, herein termed D3` (residues 271-328), adopts a "hotdog" fold (an α-helix that packs against a β-sheet) precedes the β-helix D3 subdomain that carries the glycosides catalytic function. D3' consists of a one and a half turn α-helix tucked into a twisted four stranded anti-parallel β-sheet. Tailspikes Sf6, HK620, and CBA120's TSP3 do not have such a subdomain; instead, they make an irregular extended loop in this position. Dali[69] structural comparison of the D3` revealed only a weak structural homolog with the highest z score of 2.6 for a protein kinase. The D3 subdomain (residues 329-678) begins with an α-helix cap followed by the β-helix. D3 consists of a three-sided right handed β-helix. The three β-sheets, labeled herein β1, β2, and β3, are intervened by turns of varying length to a total of 10 rungs. Finally, the D4 subdomain includes residues 679-921 and adopts a twisted antiparallel β-sandwich fold of complex topology. Dali structural comparison revealed 43 structurally similar protein domains with a Z score of 10 or greater. Despite a lack of significant amino acid identity, many of these structural homologs are carbohydrate-binding proteins, for example, the

GH43 extracellular arabinanase (PDB entry 5HO2, Z score = 13.9, RMSD = 3.2 over 181

Cα equivalent atoms) SAP-like Pentraxin (PDB entry 3FLP, Z score = 12.9, RMSD of

3.5Å over 176 Cα equivalent atoms) (Figure 4.13 C). Interestingly, an α-helix close to the

C-terminus of D4 caps the C-terminus of D3. In contrast to the N-terminal helical cap of

D3, helical capping of the C-terminus of D3 has not been seen in other tailspike proteins.

It remains unknown whether the D4 subdomain is involved in LPS binding even though its

fold is shared with many carbohydrate-binding proteins.
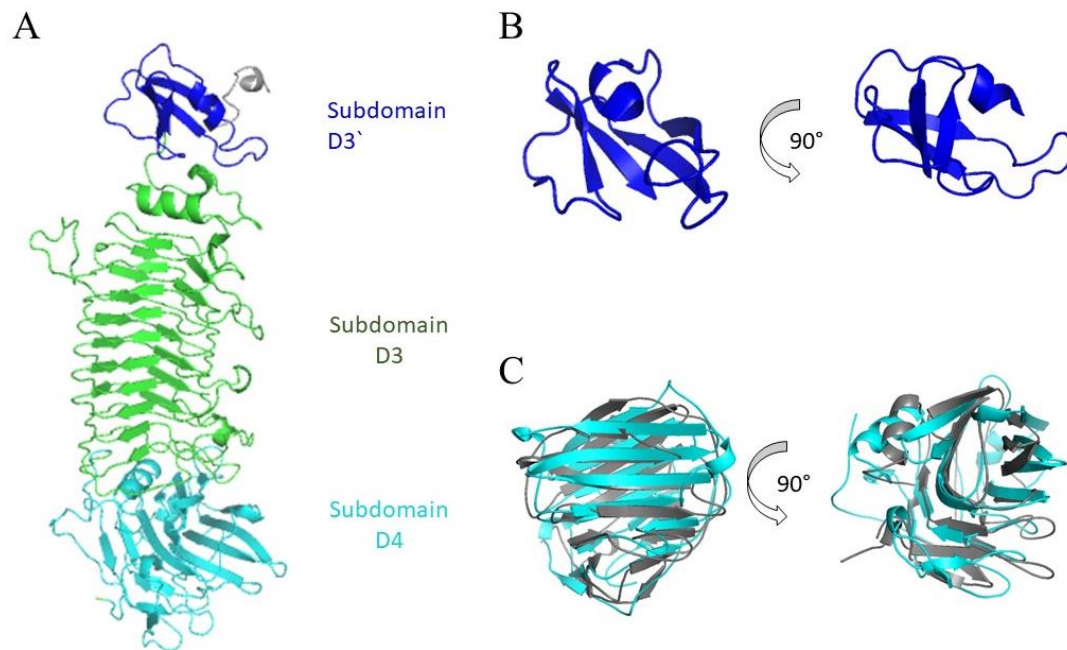
Figure 4.13 Receptor Binding Domain of TSP2
(A) Cartoon representation of the monomer highlighting D3` in blue, D3 catalytic
subdomain in green, and D4 in cyan. (B) Two views of the D3` subdomain, which
adopts a hotdog fold. (C) Cartoon representation of the D4 subdomain (cyan)
superimposed on SAP-Like Pentraxin (gray, PDB:3FLP).

4.5.5   Putative Active Site for TSP2

Receptor binding proteins such as TSP2 possess an active site where the LPS antigen can

bind and be cleaved. TSP2 possess one potential active site situated in the groove between

two monomers. The potential active site residues include three carboxylic acid residues; Asp506 on one molecule and Glu568, Asp571 on the adjacent molecule. Glu568 and Asp571 are 6Å apart and on the same monomer, while Asp506 resides on the adjacent monomer and measures 3.4Å away from Glu568 and 6.1Å form Asp571. Glu568 is buried deep within the binding cavity and is unlikely to be involved in substrate cleavage. The putative active site contains various aromatic F483, W539, and Y623 which may be important in sugar stacking. In addition, charged residues such as R458, E426, and K536, and hydrophobic residue L455 are also found in binding pockets to stabilize the carbohydrates of the O-antigen[70, 71].
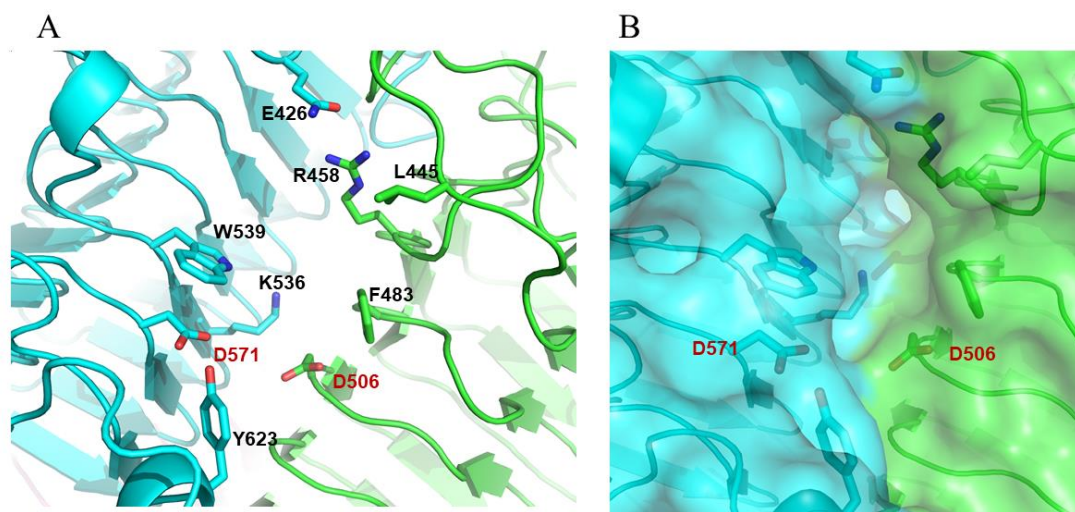
Figure 4.14 Proposed TSP2 active site.
(A) A cartoon representation of TSP2's putative active site situated on the groove between two monomers contains putative catalytic residues, D506 and D571. The active site also contains various aromatic residues (F483, W539, and Y623), charged residues (R458, E426, and K536), and a hydrophobic residue (L455). (B) A surface representation of TSP2's putative active site with important residues.

The presumed mechanism of glycosyl hydrolysis involves two carboxylate residues catalyzing the enzymatic reaction. The retaining mechanism requires the two active

65

residues to be ~5.5 Å apart while the inverting mechanism requires a larger distance of ~10Å[72, 73] (see Figure 5.1 or section 5.1 for more detail on inverting and retaining mechanisms). Both of the potential active site combinations Glu568/Asp571 and Asp506/Asp571 are approximately 6Å apart, suggesting the retaining mechanism. The structure of TSP2ΔN co-crystalized with O157 antigen (PDB ID: 5W6S) revealed that the O157 antigen is bound above residues Asp506 and Asp571 so that the carboxylate residues are positioned on either side. The binding of the O-antigen above the potential carboxylate residues supports that this location on the trimer is important in TSP2 activity and TSP2's affinity for the O157 antigen. The binding pocket also contains aromatic, hydrophobic, and charged residues such as Q426, F483, W539, K536, Q571, and Y623 which help stabilized the carbohydrate-binding sites [70, 71].
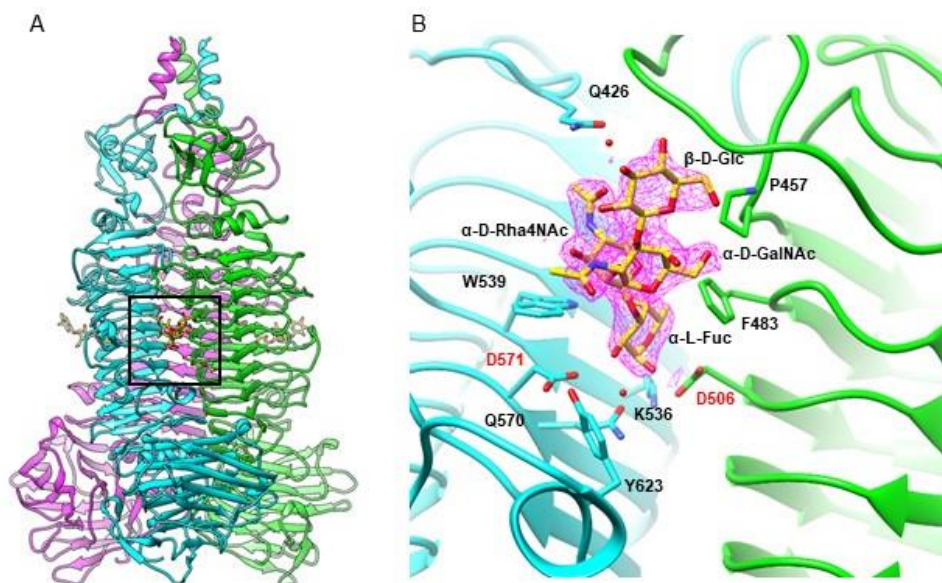
Figure 4.15 TSP2ΔN with O157 O-antigen tetrasaccharide.
(A) A cartoon representation of TSP2ΔN with the O157 antigen, indicated by the black box, bound in the groove between two monomers. (B) An enlarged image of TSP2ΔN (cartoon) with important binding pocket residues (Q426, F483, W539, D506, K536, D571, Q571, and Y623) and O157 antigen. Figure adapted from Plattner et al. (unpublished work)[74]. (PDB ID: 5W6S).

4.6 *TSP3 Crystal Structure*

4.6.1 Crystallization, Data Collection, and Structure Determination

Crystals of native and SeMet containing TSP3 were obtained by the vapor diffusion method in sitting drops at room temperature. The reservoir solution for both native and SeMet TSP3 crystals contained 18% w/v polyethylene glycol 8000, 0.2M NaCl, and 0.1M CHES, pH 9.0-9.8. Rod-shaped crystal appeared within a few days (Figure 4.16). The crystals were cryoprotected by adding to the drop 3 μL reservoir solution supplemented with 30% glycerol. Crystals were flash-cooled in liquid nitrogen. X-ray diffraction data were collected at Argonne National Laboratory, General Medical Sciences and National Cancer Institute (GM/CA) beamline 23-ID_B equipped with a MARmosaic MX-300 detector

(Marresearch). SAD data for a SeMet TSP3 crystal was acquired at the Se absorption edge

(0.98 Å) and extended to a resolution of 1.85 Å. The data was processed using

IMOSFLM[75]. The phases were determined by the Single Anomalous Diffraction method

(SAD) using the HKL3000 software, which incorporates the programs SHELXD and

SHELXE for heavy atom search[67, 76, 77]. The quality of the electron density map enabled

most of the polypeptide chains to be built automatically using Autobuild[78]. Structure

refinement was performed with the Phenix Refine in the Phenix software suite[63, 79].
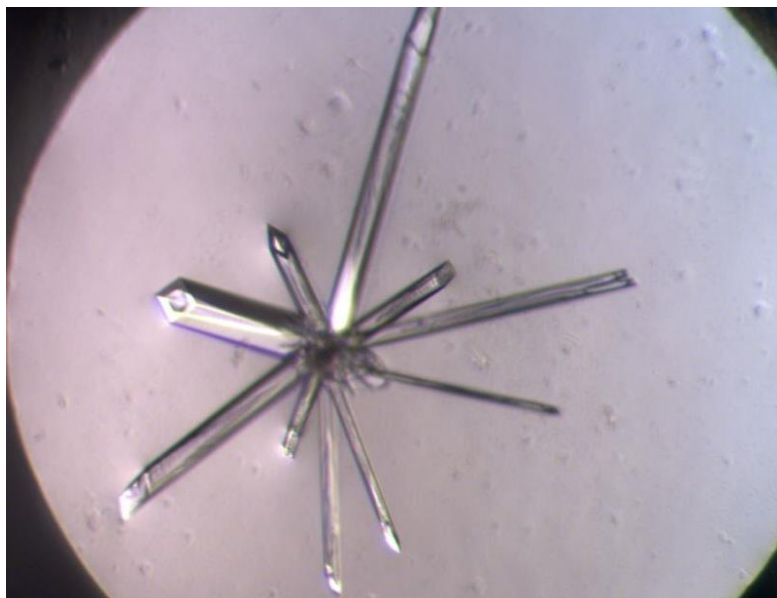
Figure 4.16 TSP3 Protein Crystal.
Long rod-shaped protein crystals were formed after a few days.

4.6.2 Overall Crystal Structure

TSP3 polypeptide consists of 627 amino acids with an additional six histidine residues at

the C-terminus engineered to facilitate protein affinity purification. The protein forms a

biological homotrimer and the asymmetric unit (Figure 4.17). The first 13-14 amino acids

of each monomer could not be modeled due to a lack of electron density in the map. TSP3

trimer is approximately 165 Å in length and 65 Å in diameter across the widest area and contains two domains, with the C-terminal domain making up roughly 75% of the overall structure.

The globular head domain, which is believed to mediate TSP attachment, is composed of 2 subdomains per chain, termed D1 and D2 (residues 13-96 and 97-154, respectively). An ensuing three α-helix bundle (residues 155-168) form the "neck" of the trimer. Next, the receptor binding domain is divided into the larger D3 subdomain (169-574) and a smaller D4 subdomain (598-627), which are connected by a 23 amino acid linker. The D3 subdomain forms a right-handed β-helix comprising 13 complete turns. Similar to TSP1, the interface between pairs of monomers in this region contains a negatively charged pocket that might bind and cleave the LPS substrate.
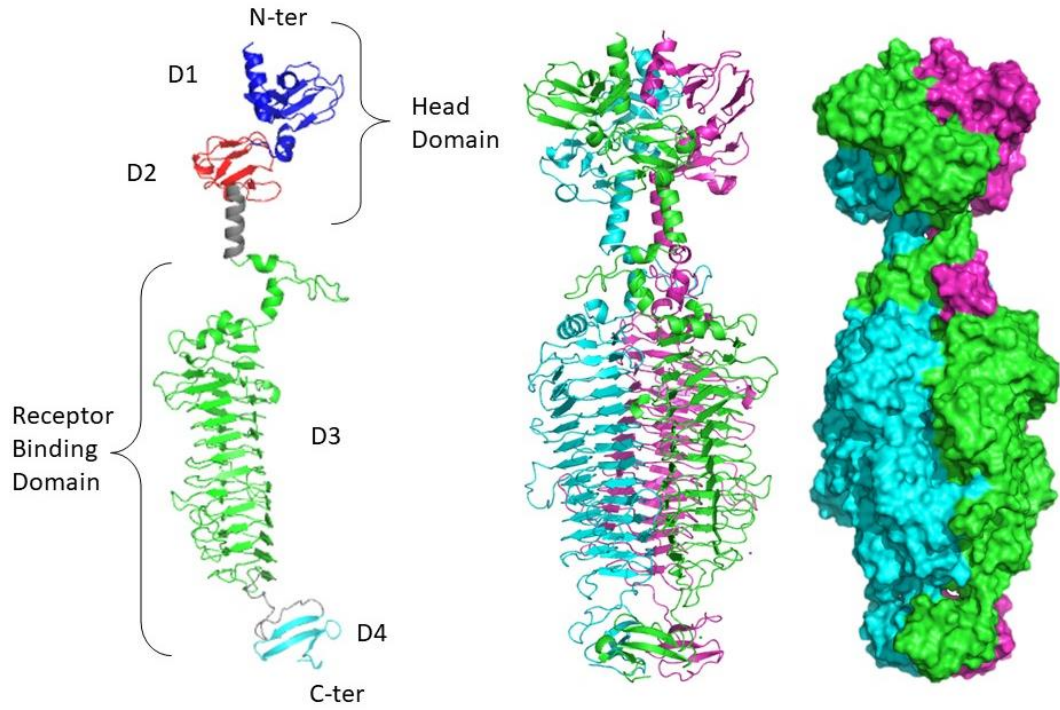
Figure 4.17 Structure of TSP3.
A cartoon representation of the monomer (left), the trimer (middle), and surface representation of the trimer (right).

Table 4.2 Statistics of data collection, phasing, and refinement of CBA120 TSP3.

Data Collection	Wild-type	Se-Met (absorption edge peak)
Wavelength (Å)	.97932	.97939
Resolution (Å) [a]	54.2-1.85 (1.92-1.85)	29.6-1.91 (1.94-1.91)
Space group	P 21	P 21
Unit cell dimensions (Å)	a=99.4, b=66.8, c=161.6, β=103.6	a=67.7, b=117.6, c=121.0, β=100.8
No. of unique molecule in the asymmetric unit	3	3
No of unique reflections [a]	175806 (17301)	140860 (6814)
Multiplicity [a]	2.0 (1.9)	2.6 (2.6)
Completeness (%)[a]	99.7 (99.1)	98.1 (95.8)
Mean I/σ(I)[a]	12.5 (4.8)	12.2 (1.6)
Rmerge[a]	0.048 (0.217)	0.058(0.687)
SAD-Phasing		
Resolution (Å)		2.1
No. of Se atoms (found/correct)		29/19
CC (%)c		37.1
Refinement		
Resolution (Å) [a]	1.85 (1.871 – 1.850)	
Total No. of Reflections	175535 (5468)	
R*work*/R*free*	0.152 (0.206)/0.184 (0.233)	
No of Protein Residues	1851	
Ligands	95	
solvent	1829	
RMSD from ideal geometry		
bond length(Å)/bond angles (°)	0.012/1.4	
Ramachandran plot: favoured/allowed/outliers (%)	96.4/3.32.99/0.3	
Molprobity overall score & percentile	1.4, 98%	

[a]The values in parentheses are for the highest resolution shell.

[b]$R_{merge} = \Sigma_{hkl}\Sigma_j |I_j(hkl) - <I(hkl)>| / \Sigma_{hkl} \Sigma_j I_j(hkl)$.

$R_{work} = \Sigma_{hkl} | F_o - F_c | / \Sigma_{hkl} F_o$, where F_o and F_c are the observed and calculated structure factors, respectively.

R_{free} is computed from 4.94% of randomly selected reflections (8672) and omitted from the refinement.

[c]Correlation coefficient between E_c and E_o in SHELXD[67].

[d]Molprobity geometry score and percentile correspond to PDB structures within the refinement resolution range[68].

4.6.3 Head Domain

The head domain of TSP3 shares 71% sequence identity with that of TSP1. Both head domains adopt the same overall fold with an 0.5 Å RMSD of α-carbon atoms (see Figure 4.18). The head domain region of several *Cba120virus* genus members demonstrates sequence identity: *Salmonella* phage Det7 TSP (97% identity), *Escherichia* phage PhaxI (80% identity), and Salmonella phage PhiSH19 (97 % identity) indicating a conserved head structure. There are examples of TSP's with homology to TSP3's head domain in both subfamilies of the *Ackermannviridae* family, the *Aglimvirinae* (*Dickeya* phage LIMEstone1 - 73% identity) and the *Cvivirinae* (*Salmonella* Phage ViI - 94% identity). Lastly, *E. coli* phage G7C's TSP gp63.1 possess 70% identity to CBA120's TSP3, despite belonging to the *Podoviridae* family and demonstrating an entirely different receptor binding structure that deacylates the LPS[41]. The shared head domain of TSP1 and TSP3 is found not only in multiple genera but also in multiple families suggesting that these proteins may have a common ancestor.

There are only a few examples of the TSP head domains structures in the Protein Data Bank because they are often truncated to facilitate crystallization. TSP3's head domain shares the same fold as CBA120's TSP1 and G7C's gp63.1. Although the TSP3 head domain residues 13-154 have higher temperature factors than those of the receptor binding domain, nevertheless the associated electron density map is well defined. The D1 subdomain begins with an α-helix where Asn24 is coordinated to an Mg^{2+} ion at the trimer center. In all, the Mg^{2+} coordinates six ligands in octahedral geometry; three Asn24 side

chains and three water molecules, as shown in Figure 4.18. The cation coordination at the

center of the trimer may contribute to the stability of the head domain.
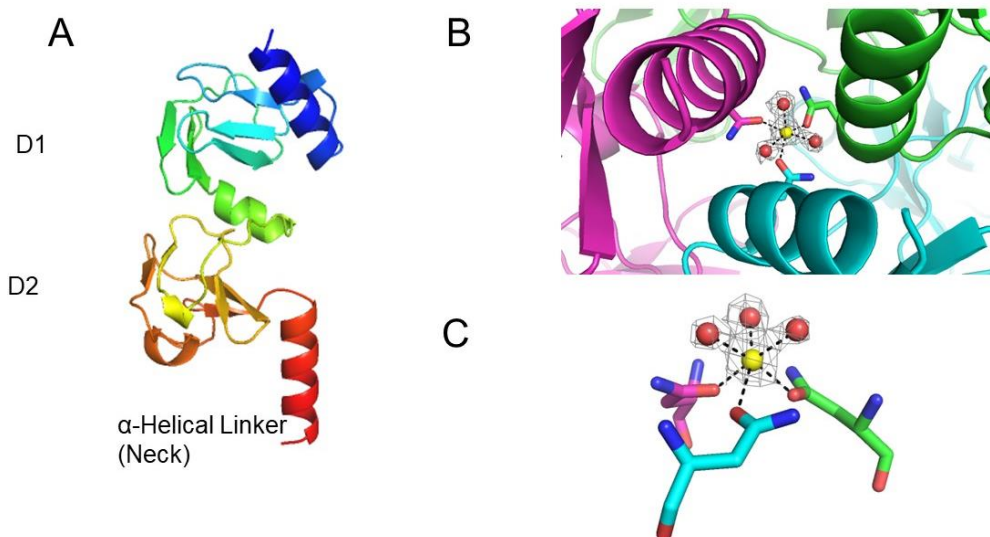
Figure 4.18 The head domain of TSP3
 (A) A rainbow cartoon representation of the two subdomains. (B) Top view of TSP3 trimer showing the octahedral Mg^{2+} coordination (yellow) to 3 Asn24 side-chains, one from each monomer, and three water molecules (red). The cation binds at the base of the N-terminal 3-helix bundle. (C) Stick representation of Mg^{2+} (yellow) coordination with an Mg-ligand distance range of 2.1-2.2 Å.

4.6.4   Receptor-Binding Domain

The receptor binding domain begins with a helical neck linker which is followed by two

splayed loops. The 61 amino acid residues (residues 169-230) which form the loops

contain 1-turn, 2-turns, and 3-turns α-helices, which wrap below the neck. The loop

containing the first two α-helices interdigitates between the neck's α-helix and the receptor-

binding domain of a neighboring monomer to enhance protein-protein interactions. The 2-

turn α-helices of the trimer (Val193- Asp199) form a 3-helix bundle along the axis of the

neck with translational displacement and tilt relative to the neck helices. The loop

containing the 3-turn α-helix caps the ensuing β-helix domain, traversing the center of the

β-helix cross-section. Similar, although not identical splayed loops and capping α-helix

have been observed in other tailspikes, including CBA120 TSP1[38] and the SF6[80] and HK620 tailspikes[60]. The α-helix capping of the β-helix also occurs in pectate lyase[60, 80, 81].

The three-sided β-helix, beginning at residue 231, is comprised of three repeating β-strands units, B1, B2, and B3 separated by three turns T1, T2, and T3. The length of the turns varies; with T3 facing the trimer interface being the shortest (1-3 residues) and several of the T1 and T2 containing extended loops. Notably, two large loops (residues 402-415 on T3 and residues 522-533 on T2) interact with one another across monomer, see Figure 4.19 Figure 4.13. The interface between two monomers defines a crevice along the two neighboring β-helices, and the intermolecular interaction mediated by the above two loops towards the C-termini of the β-helices block the crevice. As discussed in the next section and the following chapter, the active site residues are located close to this edge. This suggests an endoglycosidase binding/cleaving activity for TSP3.

As seen with other β-helical molecules, the core of each β-helix is stabilized by hydrophobic and aromatic residues including Phe294, Tyr318, Phe338, and Phe366. Other hydrophobic residues are also involving (Ile386, Ile420, Ile443). Within the hydrophobic core, three cysteines (Cys341, Cys369, Cys389) are positioned sequentially along three rungs of the internal β-sheet ladder, yet no disulfide bond is observed, possibly because the geometry is not appropriate. Strikingly, one positively charged residue in the center of each β-helix, Arg343, lacks an accompanying negatively charged residue. Instead, the charge is countered by electrostatic interactions with three backbone carbonyl groups and a water

molecule.   Intermolecular interactions along the trimer axis are mediated by both hydrophilic and hydrophobic interactions with many internal water molecules that solvate the hydrophilic residues.   Notably, three His367 side chains stack edge to face along the longitudinal axis of the trimer. An exquisite salt bridge network exists along the rungs N-terminal to the His367 cluster includes: Arg340, Asp319, Arg320, Asp296, Arg295, and glu254. Asp317 and Glu254 on the neighboring monomer are also involved, thus supporting the trimer assembly, see Figure 4.20. In addition, A PO_4^{-3} ion coordinates three Arg295 and three Arg233 at the N-terminal end of the β-helix. Multiple water molecules surround the charged network (see Figure 4.19). Another striking structural feature below the histidine cluster is the stacking of 4 threonine residues (Thr387, Thr421, Thr444, Thr470) on sequential rungs of the internal β-sheet ladder, each on the C-terminus of a β-strands, see Figure 4.20.
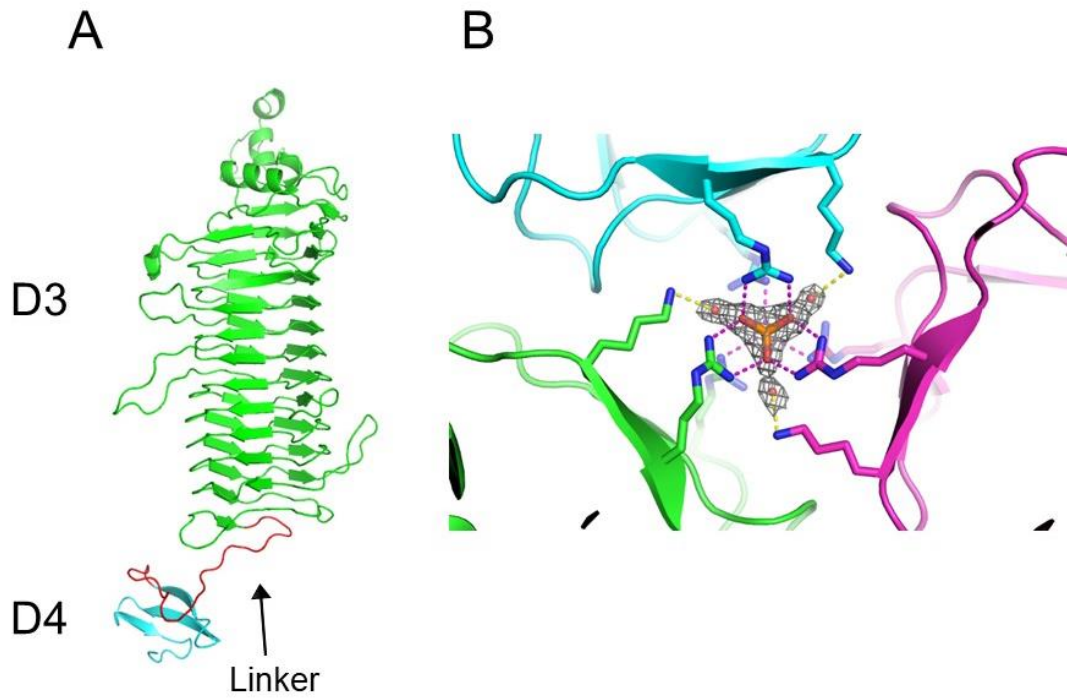
Figure 4.19 The receptor binding domain of TSP3.
The receptor binding domain of TSP3 monomer with the linker region highlighted in red, resulting in the displacement of D4 relative to D3 (left). In the D3 domain of TSP3 there is a phosphate ion with positively charged residues at the trimer symmetry axis including direct interactions with Arg233, Arg295, and interaction with Lys231 via a bridging water molecule (right).
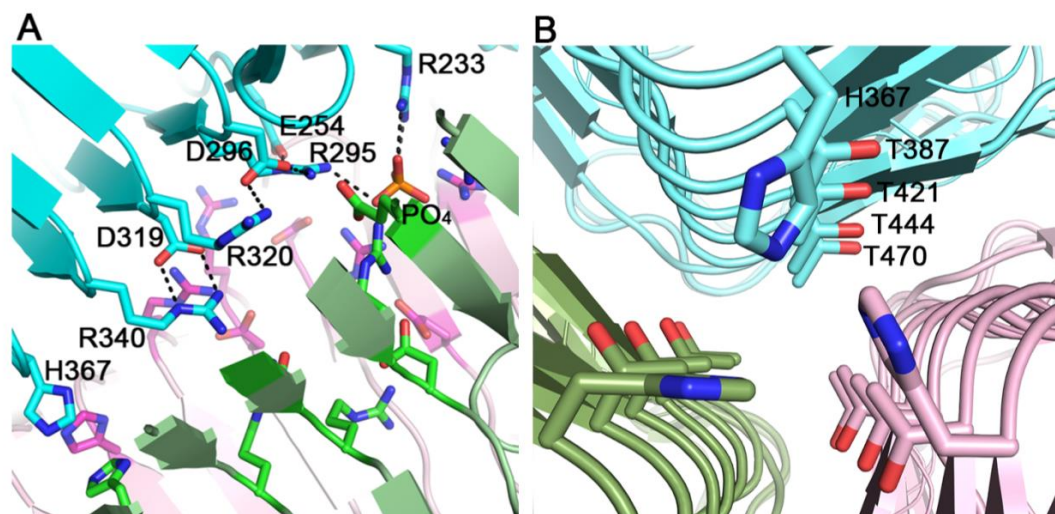
Figure 4.20 Intermolecular interactions along TSP3's 3-fold axis.
Intramolecular charge-charge interaction along the 3-fold axis of TSP3 catalytic domain (A). The side chains of His367 stack edge to face followed by the stacking of threonine residues along the β-helix rungs (B).

4.6.5 Proposed TSP3 Active Site

Similar to TSP2, the proposed active site of TSP3 resides in the groove between two monomers in the receptor binding domain, see Figure 4.21. There are three carboxylic acid residues present in the active site cleft; Glu362 and Asp383 on one monomer, and Asp426 on the adjacent monomer. All three of these residues are located 6.1 to 6.9 Å apart. In addition to the proposed catalytic residues, several aromatic residues decorate the active site including Phe322, Tyr324, Trp344, Tyr335, Phe370, and Phe392. Aromatic residues are known to stack against sugar pyranose rings, which may assist with orienting the LPS substrate. Site-directed mutagenesis (discussed in chapter 5) revealed that Glu362 and Asp426 are the two carboxyl residues involved in TSP3's enzymatic activity consistent with using an acid/base mechanism.
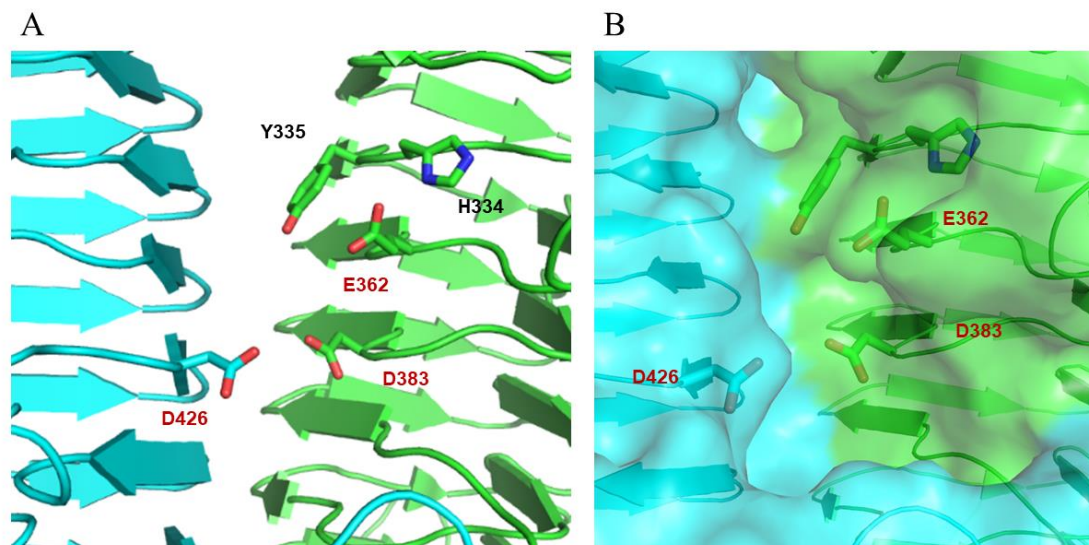
Figure 4.21 Proposed TSP3 active site.
(A) A cartoon representation of TSP3's putative active site situated on the groove between two monomers, potential catalytic residues include E362, D383, and D426. The putative active site contains aromatic residue Y335 and polar residue H334. (B) A surface representation of TSP3's putative active site with important residues.

4.7   *Crystallization and Diffraction Data Collection of TSP4*

Full-length TSP4 was not expressed; however, the N-terminal domain comprising residues 1-335, herewith termed TSP4-N, was overexpressed and purified in sufficient quantities to begin structural studies. Diffracting crystals of TSP4-N were obtained by myself and by Dr. Chao using the vapor diffusion method in sitting drops at room temperature (Figure 4.22). TSP4-N crystals were obtained in 0.1 M Bis-Tris pH 6.0, 1.6 M ammonium sulfate, 0.1 M TCEP hydrochloride after a few days. Data was first collected at the Institute for Bioscience and Biotechnology Research home source using Rigaku MicroMax007 X-ray generators and RAXIS IV++ detector; however, the crystal diffracted poorly. Dr. Chao also obtained crystals of TSP4-N with the following crystallization conditions 1 M potassium sodium tartrate, 0.2 M NaCl, and 0.1 M imidazole-HCl pH 8.0, as well as with 1.6 M

LiSO$_4$, 0.1 M NaCl, and 0.1 M Hepes 7.5; in both cases, crystals appear after a few weeks. For heavy atom derivatives crystals were soaked in crystalizing solution supplemented with 40 mM solution JBS Magic Triangle I3C [82], after which the crystals were back soaked in crystallizing solution supplemented with 30% glycerol and then flash cooled in liquid nitrogen. Diffraction data was collected GM/CA Beamline 23-IDB at Agronne National Laboratory. The data extended to ~3 Å resolution and processed with XDS[61]. Unfortunately, the crystals did not exhibit a strong anomalous signal, and the structure could not be determined by SAD method. Currently, work is underway to determine the TSP4-N structure by molecular replacement methods using models derived from the unpublished structure of phage G7C's gp66 provided by Dr. Lieman using the Modeller program [83] and Itasser [84] program.
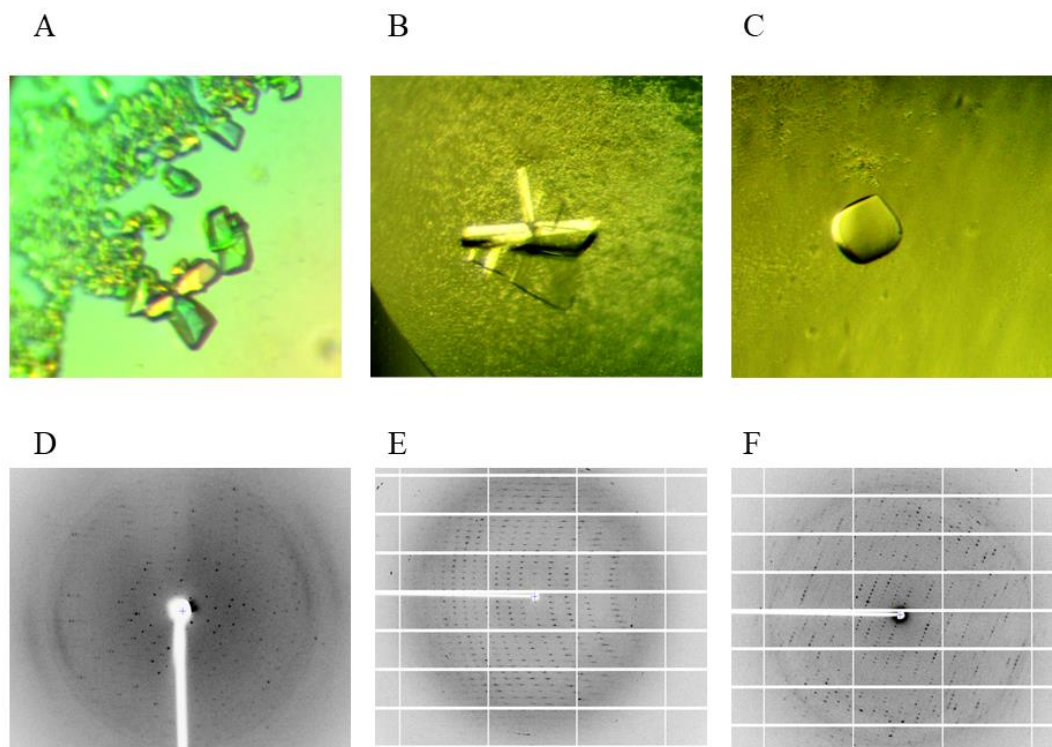
Figure 4.22 Crystals and Diffraction Images of TSP4-N.
Protein crystals and diffraction images of TSP4-N (A, D) 0.1M Bis-Tris pH 6.0, 1.6 M ammonium sulfate, 0.1 M TCEP hydrochloride (B, E) 1.6 M LiSO$_4$, 0.1 M NaCl, and 0.1 M Hepes 7.5 (C, F) 1 M K$^+$ Na$^+$-tartrate, 0.2 M NaCl, and 0.1 M imidazole-HCl pH 8.0.

4.8 *Structural comparison of CBA120 TSPs*

The head and receptor binding domains of tailspikes 1-4 all have a similar elongated morphology of ~150 Å in length, with a "head", "neck", and "body" (the crystal structure TSP4 lacking the N-terminal 340 amino acid residues was recently deposited by Petr Leiman group in the PDB, entry code 5W6H). The smaller TSP1 and TSP3 structures are more similar to one another because their C-terminal D4 domain is small and they contain only two receptor binding subdomains. In contrast, the C-terminal D4 subdomains of TSP2 and TSP4 are larger and splay away from the "body," and the receptor binding domains

80

contain three subdomains (D3', D3, and D4). In addition, TSP2 and TSP4 contain N-terminal regions that mediate the assembly of the branched structure attached to the baseplate of phage CBA120, whose structures are not yet available. Nevertheless, all four phage CBA120 TSPs share a common shape with glycosidase tailspike proteins from other bacteriophages, and they form homotrimers containing a catalytic domain that adopt the right-handed β-helix fold.

4.8.1 Head domains

There are only a few tailspike protein crystal structures that include the head domains, perhaps because some full-length molecules have inherent flexibility that impedes crystallization. Thus, the TSPs structures from phage CBA120 enriches the set of known head domain architectures substantially. When first published, the fold of TSP1 head domain was novel (TSP1 PDB ID: 4OJ5). The fold did not resemble that of the head domain of phage P22 TSP, the only other available head domain structure at the time. Subsequently, the TSP head domain of tailspike gp63.1 from phage G7C revealed the same fold as that of TSP1 even though the catalytic domain has an entirely different structure and catalytic activity[41]. Consistent with their sequence homology, the structures of TSP2, TSP3, and TSP4 head domains share the same fold except that the D2 subdomain of TSP2 is missing. The first 10-15 residues of TSP1 and TSP3 are disordered in the crystal structure. This is consistent with the Psipred Protein sequence analysis[58, 59] that predicts these areas to be disordered and perhaps involved in protein binding. For TSP2, the entire 170 residue N-terminal region, which is likely to mediate the assembly with other phage TSPs, is not seen in the electron density map, also suggesting structural flexibility.

Moreover, the structure of the N-terminal 335 amino acid residues of TSP4, which probably mediates both the attachment to the phage baseplate and assembling the partner TSPs is still unknown. Beyond these flexible regions, the head structure of TSP4ΔN begins with an N-terminal head three α-helix bundle, which is largely buried within the head domain and is followed by the three α-helix bundle neck seen in all TSPs with known structure.

TSP1 and TSP3 both have heads that adopt the same fold (Figure 4.23) with two subdomains and share a Cα atom positions RMSD of 0.48 Å. The head domain for TSP1 and TSP3 is approximately 42 Å long and 55Å across. Similarly, TSP4 shares the same fold as that of TSP1 and TSP3 consistent with the 70% sequence identity of these regions. TSP4 exhibits RMSD of .53 Å and 0.54 Å with TSP1 and TSP3 for the α-carbon atoms of the head domain. In contrast, TSP2's head domain contains only the D1 subdomain (39% and 29% sequence identity with the D1 domains of TSP1 and TSP3), is considerably shorter at 35 Å and a Cα atom positions RMSD of 1.2 Å with the D1 subdomain of TSP1 that indicates some structural divergence (Figure 4.23). The missing D2 subdomain suggests that D1 and D2 evolved independently as folding units and perhaps that the D2 subdomain is not essential for interaction with partner proteins.

The head domain of phage G7C gp63.1 adopts the same fold as TSP1, TSP3, and TSP4, and has a Cα atom positions RMSD value of 0.5Å to TSP1[41] head even though the G7C gp63.1 catalytic domain adopts an entirely different fold and catalyzes a different reaction

than the TSPs of phage CBA120. From an evolutionary perspective, this raises the possibility of horizontal transfer of genetic material consisting of DNA that encodes independent folding units. Finally, the head domain of phage P22 TSP adopts an entirely different fold comprising two β-sheets[85], and accordingly, it shares no sequence homology with the amino acid sequences of phages CBA120 and G7C TSPs.

| | | | |
| TSP1 | TSP3 | TSP2 | P22 |
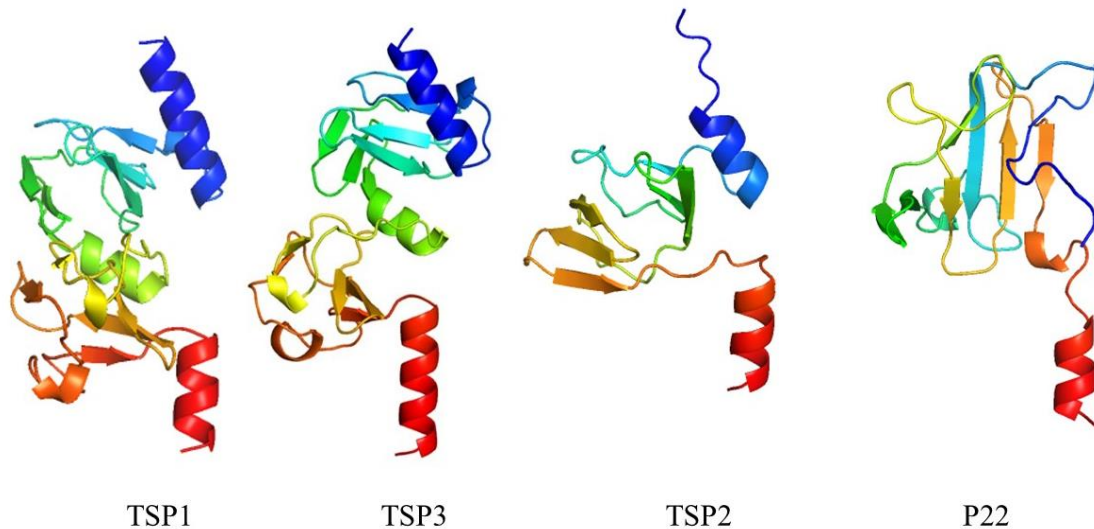
Figure 4.23 Head domains of phage CBA120 and phage P22 TSPs.
A rainbow cartoon representation of the TSP 1-3 and P22's tailspike. The polypeptide chain is colored progressively from blue at the N-terminus to red at the C-terminus. The red helices form the trimeric necks of all the structures. TSP1, TSP3, and TSP4 have two subdomains, and TSP2 has only one subdomain. P22 TSP (PDB:2VNL) contains a single all-β subdomain.

4.8.2 Receptor binding domain

The receptor binding domains of phage CBA120 TSPs follow the α-helical bundle comprising the "neck." For the TSP of phage P22, the neck was thought to allow flexibility between the head domain and receptor binding domain[85]. The postulated flexibility was hypothesized to facilitate the ability of multiple tailspikes to contact the target O-antigen, a hypothesis that was supported by a study showing that at least three TSPs must make

contact with O-antigen receptors for infectivity[86]. Thus, the necks of CBA120 TSPs may also play a similar role.

Ensuing the neck but preceding the β-helix is a conserved capping motif that stacks above the β-helix. Such a cap, reminiscent of that reported in pectate lyase[81], is commonly seen in β-helices. Prokhorov and colleagues suggested that the region between the head domain and the receptor binding domain may be a junction site for horizontal gene transfer, explaining cases of non β-helical receptor binding domains that possess a similar motif[41]. As mentioned above, the catalytic domain of gp63.1 adopts a different fold (SGNH-esterase), yet it contains a β-helix cap[41]. Indeed, the structures of TSP2 and TSP4 contain the D3' subdomain inserted within this junction site, further supporting this proposed hotspot for horizontal transfer. The D3'subdomain, observed so far only in TSP2 and TSP4, adopts in both proteins the hotdog fold, with a 0.5 Å RMSD over Cα positions of the subdomain and 33 % sequence identity.

A striking characteristic of the glycosidase-class of TSPs including, TSP1-4 of phage CBA120 and the TSPs of phages Sf6, P22 and HK62 is that despite their common β-helical fold of the D3 catalytic domain, they lack significant sequence homology. Moreover, the number of rungs along 3-faced β-helix domains varies from protein to protein. Accordingly, these domains do not align well with one another. Yet, they contain common features: the triangular cross-section of the β-helix is concave along the exterior edge of the triangle and flat along the edges that are involved in subunit-subunit interactions. In addition, the TSPs have large loops are located on the exterior vertices whereas tight loops

are located on the interior vertex of the triangles. These tight loops tend to have the same conformation along the rungs of the β-helix ladder yielding a regular packing of the trimer subunits and a triangular internal channel.

So far, the D4 subdomain exhibit the highest variability among β-helical TSPs, as most TSPs adapt different D4 folds. With phage CBA120, TSP1's D4 subdomain forms a right-handed 3-stranded β–helix. That of TSP2 forms a β-sandwich of complex topology similar to the carbohydrate binding protein pentraxin. D4 of TSP3 forms a 3-stranded β-sheet that the trimer assembles into a β-prism II, a fold associated with carbohydrate binding. Moreover, the D4 subdomain of TSP4 (determined by the Leiman's group) folds into a β-sandwich similar to a carbohydrate domain 61. Thus, all phage CBA120 D4 domains could potentially be involved in carbohydrate binding of the appropriate substrate or other carbohydrates on the bacterial surface. Interestingly, TSP3 has the shortest D4 subdomain (29 amino acids). The structure is broadly similar to the last 30 residues of the D4 subdomain of phage P22 TSP; however, D4 of phage P22 TSP is much larger. P22 TSP also forms a beta prism motif that has been proven important for trimer stabilization[87, 88]. It is tempting to speculate that the small TSP3 D4 subdomain contributes to the relatively low thermal stability of TSP3.

4.8.3   Structure-based putative locations of the catalytic sites
Vacuum surface electrostatic potentials of these trimers were calculated using PyMol to examine global charge trends (Figure 4.24). Although vacuum calculations are only

qualitative, a consistent trend emerges for all four TSPs showing a prominent negatively charged patch in grooves corresponding to interfaces between trimer subunits. These patches contain carboxylate groups that may function as the glycosidase catalytic machinery as discussed earlier. The location of the TSP3 catalytic site has now been confirmed by site-directed mutagenesis, as discussed in Chapter 5. TSP1 and TSP2 catalytic sites are currently being tested in the lab by site-directed mutagenesis as well. The TSP1 negative electrostatic potential is formed by Glu456 and Glu483 on one molecule and Asp313 on the adjacent molecule, as discussed earlier in this chapter. The negative patch on TSP2 surface is formed by Asp506 on one molecule and Glu568, Asp571 on the adjacent molecule, an arrangement that is also discussed in the previous section. TSP3 has a negatively charged pocket involving the carboxylate groups of Glu362 and Asp383 on the same monomer and Asp426 of the neighboring monomer. Finally, based on the coordinates deposited in the PDB (entry code 5W6H), the negatively electrostatic potential corresponding to TSP4 active site candidate includes Glu675 on one monomer and Asp737 on the neighboring monomer, which are located 10 Å apart and are surrounded by two tyrosine, and tryptophan, residues that are often encountered within catalytic sites of glycosidase. Table 4.3 summarizes the distances between catalytic residues candidates of phage CBA120 TSPs, which shows that both the short (~6 Å) and the long distances (~10 Å) between carboxylate groups corresponding to retaining and inverting stereochemistry acid/base hydrolytic mechanism, respectively, may be employed by the TSPs of phage CBA120[73]. Once all TSPs catalytic residues are identified, it will be possible to assess the diversity of the catalytic mechanisms employed by bacteriophages.

Table 4.3 Proposed active sites of TSP1-3.

	Proposed Active Site Residues	Distance apart (Å)	Suggested Mechanism
TSP1	E456/D313	11.2	Inverting
	E483/D313	10.2	Inverting
TSP2	D506/D571	6.1	Retaining
TSP3	E362/D426	6.7	Retaining
TSP4*	E675/D737	10	Inverting

* PDB ID: 5W6H
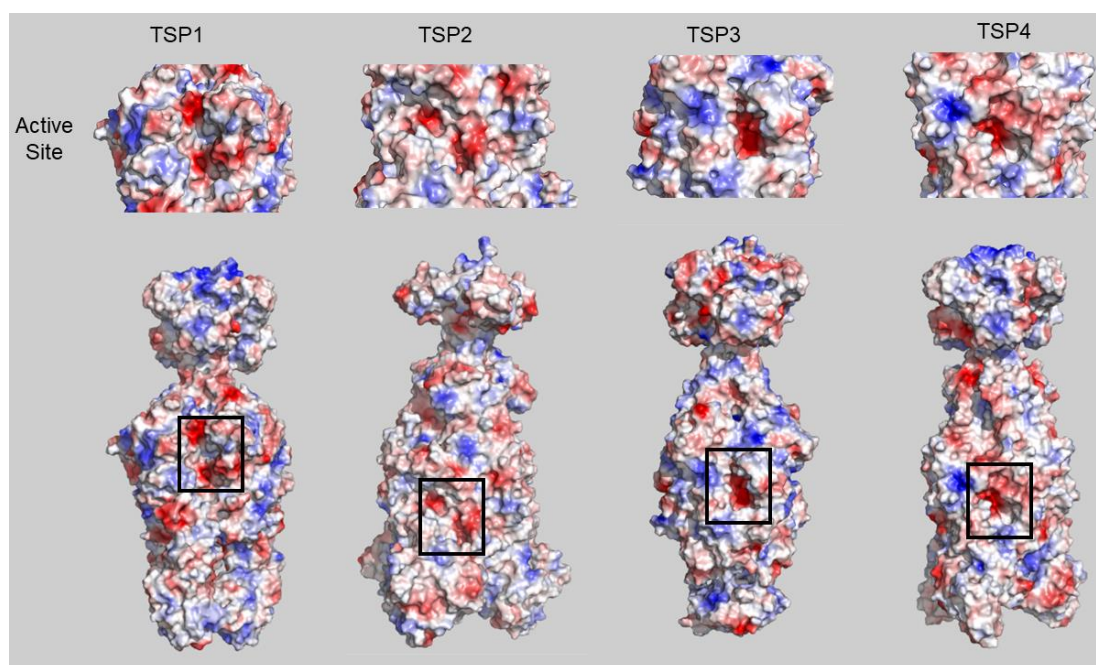
Figure 4.24 Vacuum surface electrostatic potential of TSP1-3.
Calculated using PyMol, positively charged region are colored blue and negatively charged regions are colored red. Boxes indicate the location of the proposed active sites along grooves formed by molecular interfaces.

Chapter 5: Tailspike activity

5.1   *Introduction*

For bacteriophages to replicate, they must be able to infect their host bacteria and commandeer its machinery to create new progeny. During the infection process, receptor binding protein engages the receptors present on the bacteria before adsorption can occur. Many species of bacteria have developed extensive LPS and Extracellular polymeric substances (EPS) that surround the bacteria and provide a physical barrier obstructing the bacteriophage's access to the host cell surface. In response, bacteriophages have evolved specialized proteins that allow them to degrade these carbohydrate barriers and infect their host bacteria. One group of proteins that enable phages to traverse these obstacles are depolymerases (reviewed by Hughes et al. [13, 21]).

The two main classes of phage depolymerases known to target extracellular bacterial carbohydrate structures are lyases and hydrolases[1]. Lyases cleave glycosidic bonds by the elimination mechanism and can be subdivided into several classes based on their target molecule including hyaluronidases, alginate lyases, and pectin/pectate lyases[1]. Hydrolases, catalyze the breaking of chemical bond using water molecules, hydrolases can also be divided into several subclasses including sialidases, levanases, xylosidases, dextranases, rhamnosidases, peptidases, etc[1]. So far many annotated tailspike proteins have been identified as hydrolases; including those from bacteriophages P22, HK620, Sf6, Phi29, K1F, and G7C.

Some hydrolases degrade the O-antigen, for example, phage P22's tailspike has endorhamnosidase activity specific to several *Salmonella* serogroups[89]. Conversely,

hydrolases may modify the O-antigen without cleaving the polysaccharide backbone, for example, phage G7C's tailspike gp63.1 has esterase activity towards an acetyl substituent of the O-antigen. Furthermore, phage φ29's tailspike is a teichoicase that degrades the poly (glycerol phosphate) teichoic acid that makes up part of the Gram-positive bacteria *Bacillus subtilis* cell wall[90]. In this chapter, I describe investigations to determine the activities of TSP1-3.

Glycosyl hydrolases have two primary mechanisms to catalyze the hydrolysis of a glycosidic bond, as first described by Koshland in 1953[73]. The retaining mechanism is a two-step reaction, catalyzed by two carboxylate-containing residues, where one residue acts like an acid (proton donor) and the other as a base (nucleophile). Each step inverts the stereochemical configuration at the sugar's anomeric carbon; thus a double inversion results in the final product retaining its original stereochemistry. First, the base attacks the anomeric center, creating a glycosyl-enzyme intermediate and releasing the terminal sugar, which has been protonated by acid. In the second step, the deprotonated carboxylate group deprotonates a water molecule which mounts a nucleophile attack to hydrolyze the glycosyl-enzyme intermediate (Figure 5.1). This mechanism requires approximately 5.5 Å between the two catalytic residues because the hydrolytic water molecule replaces the leaving saccharide fragment.

The second glycosidase mechanism inverts the stereochemical configuration of the anomeric carbon and is a one-step reaction. Similar to the retaining mechanism, the active

residues are two carboxylate residues, one acts as an acid and the other as a base. This is a concerted reaction, with the base deprotonating the nucleophilic water that attacks the anomeric center. Concurrently, the leaving group deprotonates the acid (Figure 5.1B). The inverting mechanism requires a larger distance between the catalytic residues of ~10 Å because the hydrolytic water molecule binds between the base and the saccharide substrate.
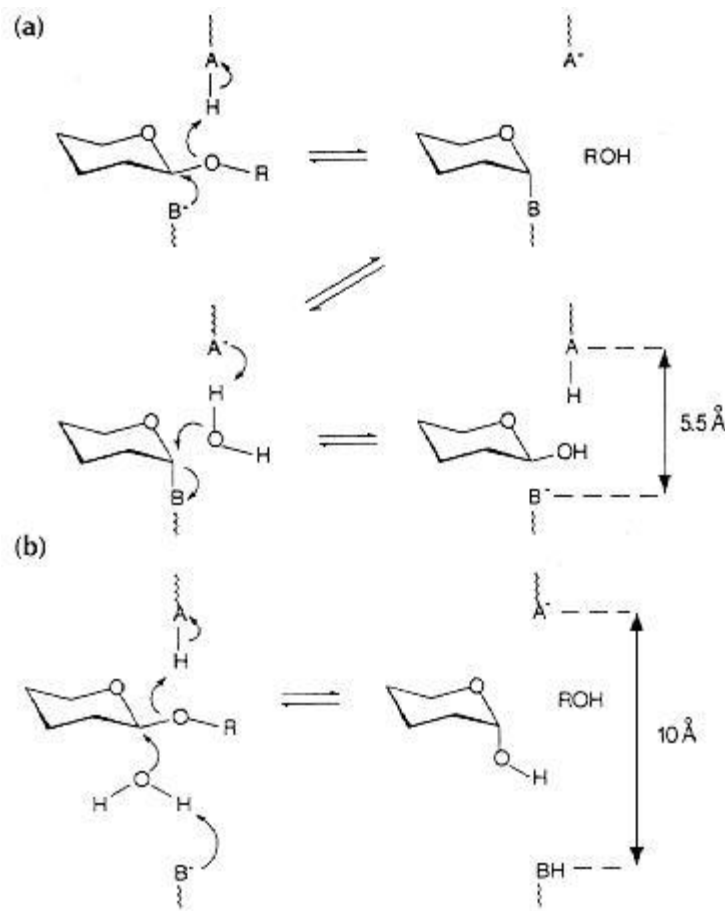
Figure 5.1 Mechanism of glycosyl hydrolases.
(A) The retaining mechanism involves a two-step reaction. (B) The inverting mechanism occurs in a single step reaction. Figure adapted from Davis & Henrissat, 1995.

5.2   *Materials and Methods*

5.2.1   Bacterial strains

Bacterial strains that were purchased from the American Type Culture Collection (ATCC) include toxigenic (EC43894) and non-toxigenic (EC700728) strains of *E. coli* O157:H7, and negative control non-O157:H7 *E. coli* strains (EC35218) that produces O6 O-antigen. *E. coli* O157:H7 (EDL933) was obtained from Matthew Waldor's laboratory where it was genetically modified to create three gal mutants, TEA023, TEA026, and TEA028, which are deficient in O157 antigen[91]. Deletion-insertion mutations to genes responsible for modifying galactose to N-acetyl-D-galactose, a component in O157 antigen, were introduced in these knockout strains. *E. coli* O157 mutant TEA023(ΔgalU::aad-7) replaced all but the first 33 bp and the last 30bp of the galU gene open reading frame with the spectinomycin gene (aad-7)[91]. For the ΔgalETKM mutant TEA026(ΔgalETKM::aad-7), all the genes in the gal ETKM operon where replaced with the spectinomycin gene, except the first 36bp and the last 30bp of the galE and galM genes, respectively. TEA028(ΔgalETKM::tetA), is similar to TEA026(ΔgalETKM::aad-7) but it contains the tetracycline resistance gene (tetA) instead of the spectinomycin gene[91].

5.2.2   Plaque assay

Host cultures were prepared by growing an overnight culture of strains EC35218, EC700728, TEA023, TEA026, and TEA028 in sterile LB media at 37 ºC and 220 rpm. The inoculation culture was prepared by making a 1:100 dilution of the overnight culture in fresh LB media and incubating at 37 ºC and 220 rpm until the culture reached OD_{600} of 0.30. Petri dishes were prepared by dissolving 7.5 g of agarose into 500 mL of sterile LB media. 10 mL of molten agar mixture was poured into a 100 mm petri dish and allowed to

solidify. Top agar was prepared by dissolving 2.5 g of agarose in 500 mL sterile LB Media and kept at 55 ℃ in a hot water bath to prevent the agar from solidifying. 0.05 mL of 5 x 10^4 CFU/mL of phage was used to inoculate 250 µL of the host culture and incubated for 10 min at 37 °C. The inoculated culture was then transferred to 3 mL of molten top agar and poured onto a prepared petri dish. After the top agar layer has solidified, the plates were transferred to 37 ℃ incubator overnight and observed the next day. The formation of discrete clear spots in the bacterial lawn indicated the presence of bacteriophage capable of infecting and lysing the bacteria in that location.

5.2.3   Adsorption assay

E. coli 700728 culture was grown at 37 ℃ to OD_{600} of 0.1. 500 µL EC700728 was mixed with ~100 µg tailspike protein or BSA or 23 µg inactive PlyC (negative controls) and 500 µL of $5x10^4$ CFU/mL phage CBA120 in SM buffer (100 mM NaCl, 8 mM MgSO4, and 50 mM Tris-HCl). The mixture was incubated for 8 min at 37 °C and then centrifuged for 1 min at 14,000 g. The supernatant was removed and used in a serial dilution and added to 250 µL of $9x10^7$ CFU/mL *E. coli* 700728. 3 mL molten top agar was added to the mixture and poured onto a plate. After the agar solidified, the dish was transferred to a 37 ℃ incubator overnight. This procedure was adapted from Prokhorov et al. 2017[41].

5.2.4   Turbidity assay

E. coli 700728 culture was grown at 37 ℃ to OD_{600} of 0.2. 250 µL was mixed with 100 µg/mL, 50 µg/mL, 25 µg/mL, or 5 µg/mL of tailspike protein in SM buffer. The mixture was incubated for 30 min at 37 ℃. 125 uL of $5x10^4$ CFU/mL phage CBA120 in SM buffer was added to the mix of the *E. coli* and tailspike protein. Control conditions included *E. coli* 700728 without the CBA120 phage or tailspike protein, CBA120 phage without

tailspike protein, and tailspike protein without the CBA120 phage. 150uL of the final mixture was transferred into the 96 well plates in triplicate. The plates were incubated at 37 ºC, and the OD_{600} was measured every 30 min for 8 h. A summary of all the conditioned used is provided in Table 5.1.

Table 5.1 Conditions of *E. coli* 0157:H7/ phage CBA120 turbidity assay.

Condition	*E.coli*	Phage	Tailspike protein (μg/mg)
E.coli control	+	-	-
Phage control	+	+	-
TSP1control	+	-	50
TSP1	+	+	5,25,50, 100
TSP2 control	+	-	50
TSP2	+	+	5,25,50, 100
TSP3 control	+	-	24
TSP3	+	+	5,10,15, 24
TSP1 and TSP2 control	+	-	50
TSP1 and TSP2	+	+	5,25,50, 100

5.2.5   Halo assay

Halo assays were performed by the Nelson Lab. A 50 mL culture *E. coli* O157:H7 (ATCC 700728) was grown overnight at 37 °C to an $OD_{600} = 1.6$. The cells were centrifuged at 4,150 rpm, and at 4 °C for 10 min and the supernatant was removed. The cell pellet was washed three times with 1x PBS buffer and resuspended in 1 mL PBS. 500 μL of the washed cells were mixed with 10 mL 0.7% agarose solution and poured into a 100 mm plate. After solidifying, 0.5 cm wells were made using sterile plastic dropper tips. Initially, 10 μL of 6 mg/mL TSP 1-3 in PBS was added to the wells, and for TSP3 mutants halo assay 45 μg protein wild-type TSP3 and TSP3 mutants were added to the wells. The plates were incubated at 37 °C for up to 48 h. This protocol was adapted from K. A. Hughes et al.[13]. The formation of clearing zones, corresponding to glycosidase activity, was observed

and photographed.

5.3　*Results*
5.3.1　TEA mutants and phage CBA120

The LPS of *E. coli* O157:H7 contains a modified galactose residue as one-fourth of a repeating polysaccharide motif of the O-antigen. Without the gal operon GalE, GalT, GalK, and GalU proteins, *E.coli* EDL933 is unable to synthesize N-acetyl-D- galactose of the O157 antigen[91]. Because of the exposed LPS core *E. coli* O157:H7 gal mutants are 70% more susceptible to P1 transduction phage than the wild-type *E. coli* O157:H7[91].

A plaque forming assay was used to investigate the phage CBA120 ability to infect the wild-type *E. coli* O157:H7 and the O157 gal mutants. When phage CBA120 is inoculated with *E. coli* O157:H7 (ATCC 700728) that produces intact O-antigen, phage plaques form - confirming reports that CBA120 infects *E. coli* O157: H7[27, 91]. Non-O157:H7 *E. coli* (ATCC 35218) has a different O-antigen　O6[92], and no plaque formation was seen after inoculation with phage CBA120, as expected from its specificity (Table 5.2). It is believed that interactions between the O-antigen and phage tailspike proteins are essential to the adsorption process and without the correct TSP O-antigen receptor,　phage CBA120 tailspikes are unable to degrade the LPS and reach the primary receptor on the cell surface[21]. The idenity of CBA120's primary receptor is not known: however in phage infection irreversable binding to a primary receptor on the cell surface is essentail to the infection process[22]. It is also important to note that *E. coli* strain 35218 may not possess the necessary primary receptor to facilitate infection. Nonetheless, phage CBA120 ability to infect *E. coli* strain 700728 but not strain 35218 is consistent with the previous study of Kutter et. al[27], which showed the selectivity of phage CBA120 to　*E. coli* O157: H7.

To probe the importance of the O157 O-antigen to phage CBA120 infection, O157 gal mutants (TEA 023, TEA 026 and TEA 028) were inoculated with the phage. None of the gal mutants showed plaque formation, confirming that the O157:H7 O-antigen is required for phage CBA120 infection (Table 5.2). Previously, studies of phage P22 adsorption demonstrated that phage P22's TSP must interact with several *S.typhimurium* O-antigen chains for phage adsorption to proceed[86]. The authors hypothesized that phage P22 must interact with three O-antigen chains and orient itself appropriately with respect to the primary receptor for infectivity [86]. It is tempting to speculate that a similar process occurs when phage CBA120 infects *E. coli* O157:H7.

Table 5.2 Phage CBA120 plaque formation on *E. coli* O157:H7 gal mutants.

E. coli strain	35218	700728	TEA023	TEA026	TEA028
Plaque formation	-	+	-	-	-

- no plaque, + plaque

5.3.2 Adsorption inhibition Assay

The phage adsorption assay tests how many phages are adsorbed onto the bacterial surface during an 8-minute incubation. By adding exogenous TSPs to the phage titer, their effect on phage CBA120's ability to infect *E. coli* O157:H7 (ATCC 700728) can be assessed. If a TSP is involved in phage attachment, it would compete with the phage for the receptors on the bacteria and inhibit phage ability to bind to the bacteria, resulting in higher free phage titer in the supernatant compared with a reference assay performed without TSP. Phage supernatant without bacteria provides the maximum phage titer control for the

subsequent plaque formation. Table 5.3 summarizes the combination of components used in these experiments.

Phage titer and the subsequent plaque formation is limited by the number of phages that can traverse the LPS and infect the bacteria. Negative control proteins, BSA and inactive PlyC, in the presence of bacteria and phage have no significant effect on phage CBA120 adsorption and should yield the same phage titer as the reference mixture of phage and bacteria. Indeed, they produced a lower phage titer than the control comprising phage alone without bacteria. TSPs 1-3 do not form any plaques without the addition of phage CBA120, which confirms that it is the phage CBA120 that creates the plaques and not the TSPs alone. Of the three TSPs tested, plaque formation following the addition of TSP1 was similar to that of the BSA and inactive PlyC negative controls. Lower plaque formation indicates that TSP1 is not competing with the phage CBA120 receptors as it does not lower free phage titer by preventing adsorption.

Unexpectedly, both TSP2 and TSP3 increased plaque number dramatically, and thus infection efficiency, when compared to the reference of maximum phage titer in the absence of bacteria (Figure 5.2). Higher phage titers can be explained by the enzymatic activity of TSP2 and TSP3 which increases phage access to the primary receptors of *E. coli* O157:H7 and thus enhances phage CBA120's infectivity. Such activity must not prevent phage binding to the bacterial cells and could include non-specific cleavage of surface polysaccharides other than the LPS, or LPS modification that does not cleave the

saccharide backbone. Nonetheless, the adsorption inhibition confirmed TSP2 and TSP3

activities, although it did not reveal a TSP1 activity.

Table 5.3 Conditions of *E. coli* O157:H7/CBA120 Adsorption Inhibition Assay

Proteins	EC 700728	Phage	Protein amount (ug)
None (free phage control)	-	+	-
BSA (Control)	+	+	100
Inactive PlyC (Control)	+	+	23
TSP1	+	+	100
TSP2	+	+	100
TSP3	+	+	100
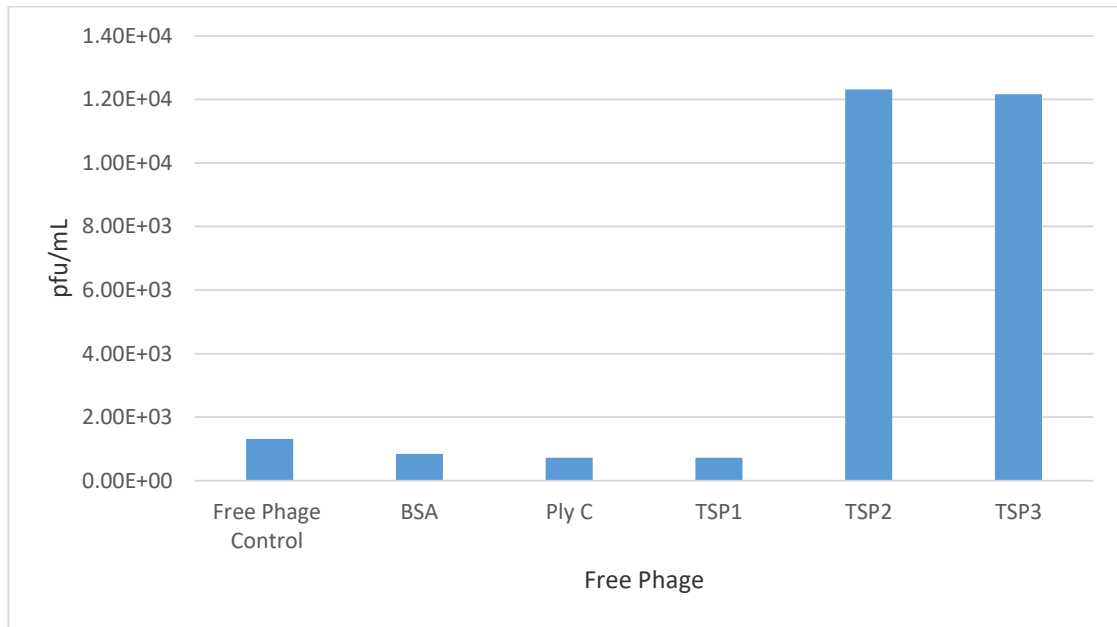
Figure 5.2 Phage CBA120 adsorption inhibition assay.
The histogram bars are as follows: The control free phage titer (first); phage titers in the presence of BSA and inactive PlyC after incubation with *E. coli* O157:H7 (2 and 3). Phage titers in the presence of TSP1-3 in (4-6).

5.3.2 Turbidity assay

Phage CBA120 replication has been documented by Kutter et. Al[27] in 2011. Under

anaerobic conditions, the gradual bacterial lysis began after 90 minutes. In an attempt to

investigate the effect TSPs have on host bacterial infections, *E. coli* O157:H7 and varying

concentrations of TSPs and phage were incubated together at 37 °C and growth was monitored at 600 nm. Controls included *E. coli* O157:H7 cells without phage or TSP to monitor normal anaerobic growth, *E. coli* cells and phage without TSP as the reference phage lysis level. Moreover, cultures include TSPs and *E. coli* 700728 without phage to confirm that TSPs alone does not affect cell growth and not plaques were formed (data not included).

TSP1 and TSP3 showed no effect on phage CBA120's ability to infect *E. coli* O157:H7. The cell density of *E. coli* O157:H7 increased over the first 100 minutes for all conditions after which differences were apparent. The control cells without phage CBA120, either with the *E. coli* O157:H7 alone or in the presence of a TSP displayed logarithmic growth until they reached a plateau after 500 minutes. In the presence of phage CBA120, bacterial growth reached a maximum between 100 and 200 minutes and then the cell density declined sharply as phage CBA120 lysed them. The presence of varying amounts of TSP1 and TSP3 had minimal effect on the *E. coli* cell growth.

Figure 5.3 turbidity assay of *E. coli* O157:H7 and TSP1.
Turbidity assay displaying the growth of E. coli O157:H7 as a function of time (minutes). the TSPl_control contains 50mg/mL TSP1without CBA120 phage, TSPl_100 contains 100mg/mL TSP1with phage CBA120, TSPl_50 contains 50mg/mL TSP1 with phage CBA120, TSPl_25 contains 25mg/mL TSP1with phage CBA120, TSPl_5 contains 5mg/mL TSP1with phage CBA120, and Cells Control contains only *E. coli*.
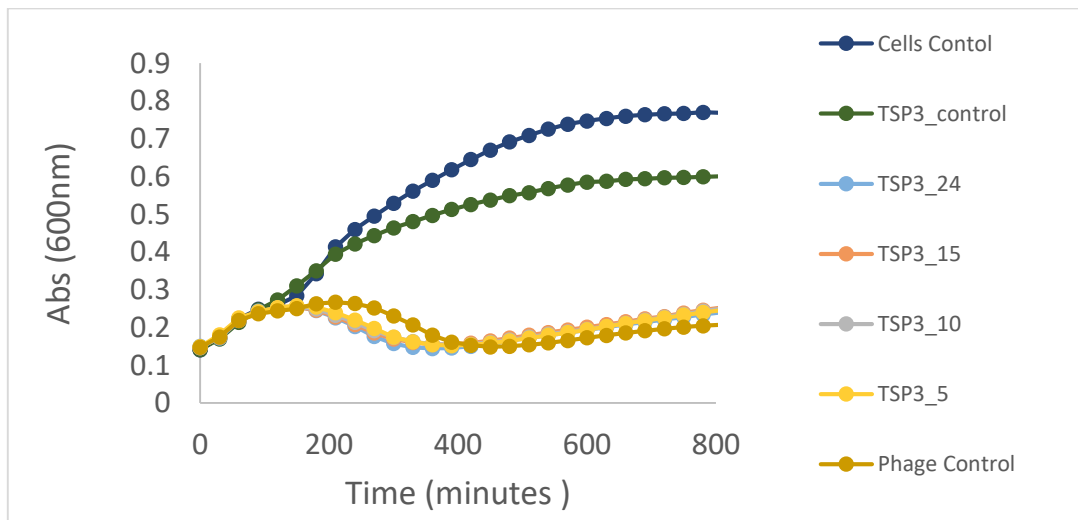
Figure 5.4 Turbidity assay of *E. coli* O157:H7 and TSP3.
Turbidity assay displaying the growth of *E. coli* O157:H7 as a function of time (minutes). the TSP3_control contains 24mg/mL TSP3 without CBA120 phage, TSP3_24 contains 24mg/mL TSP3 with phage CBA120, TSP3_15 contains 15mg/mL TSP3 with phage CBA120, TSP3_10 contains 10mg/mL TSP3 with phage CBA120, TSP3_5 contains 5mg/mL TSP3with phage CBA120, and Cells Control contains only *E. coli*.

In contrast to TSP1 and TSP3, which did not affect bacterial cell infection by phage CBA120 TSP2 exhibited concentration dependent inhibition of phage infectivity. The highest concentration of TSP2, 100 ug/mL of abolished phage CBA120 infectivity, and this culture growth curve is similar to the cells controls with and without TSP2 but lacking the phage. Decreasing the amount of exogenous TSP2 increased the cell infection by phage CBA120 resulting in decreased *E. coli* growth. These results support the conclusion that TSP2 competes with the phage for the secondary *E. coli* O17:H7 receptor and in the highest concentration completely prevents phage attachment to the LPS. Turbidity assays with the gal mutants, TEA023, TEA026, and TEA028, which lack the O-antigen, revealed that phage CBA120 was unable to infect and lyse *E. coli* O157:H7 (data not shown). By analogy, we hypothesize that the phage had no access to the bacterial LPS receptor in the presence of TSP2 either because TSP2 occupied all receptor sites or because TSP2 degraded the LPS within the first 100-200 min.
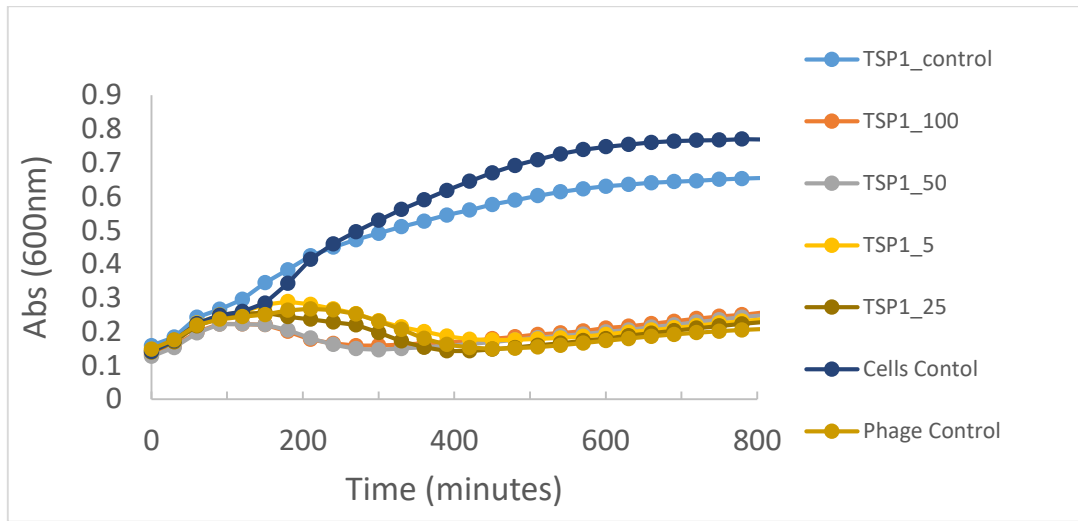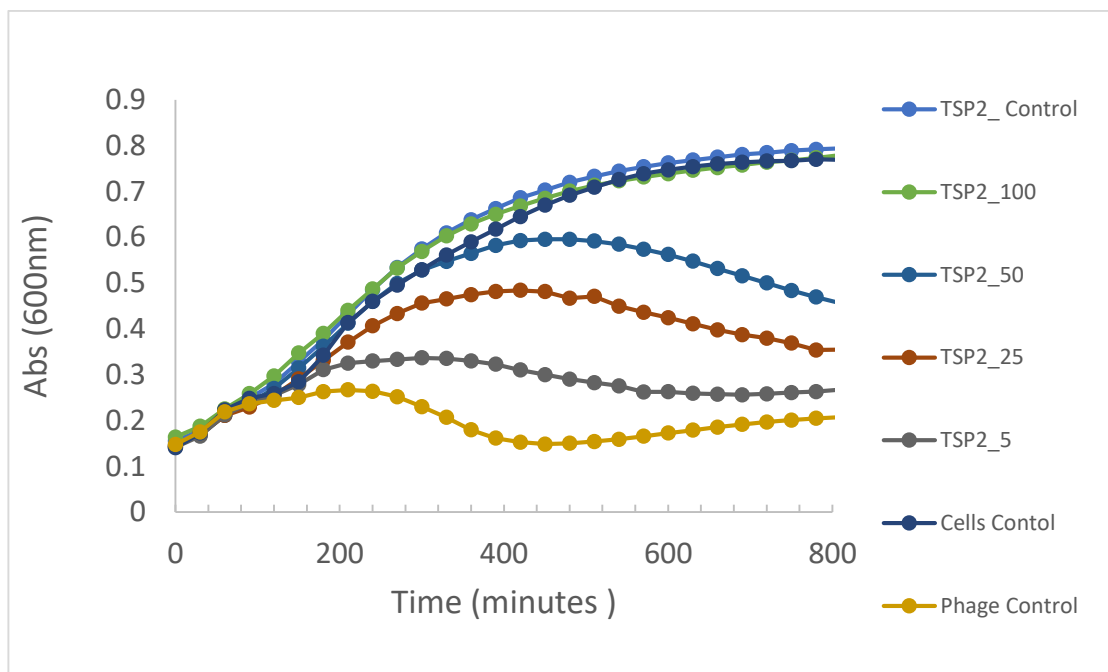
Figure 5.5 Turbidity assay of *E. coli* O157:H7 and TSP2.
Turbidity assay displaying the growth of *E. coli* O157:H7 as a function of time (minutes). the TSP2_control contains 50mg/mL TSP2 without phage CBA120, TSP2_100 contains 100mg/mL TSP2 with phage CBA120, TSP2_50 contains 50mg/mL TSP2 with phage CBA120, TSP2_25 contains 25mg/mL TSP2 with phage CBA120, TSP2_5 contains 5mg/mL TSP1with phage CBA120, and Cells Control contains only *E. coli*.

5.3.3   Halo assay and Identification of TSP3 Catalytic Residues

Although tailspikes do not lyse bacterial cells, the thinning of the LPS layer can be viewed

as a halo of less density or less opacity on an agar dish embedded with sensitive bacteria.

Whereas this does not address the specific glycosidic cleavage site, it does allow for semi-

quantitative demonstration and comparison of activity.

Purified TSP1-3 applied to *E. coli* O157: H7 embedded in agarose caused the formation of

halos extending from the well site. The halos indicate that *E. coli* Ol57:H7 O-antigen or

other surface polysaccharides are substrates of TSPs 1-3. Out of the three TSPs tested TSP1

101

showed the smaller less clear halo. Both TSP2 and TSP3 formed more evident halos, by comparison, consistent with a more active depolymerase activity. While these findings are indicative of TSP enzymatic activity, they do not divulge information about catalytic rates or substrate specificity. While the 3D structures of these TSPs suggest that all have glycosidase activity, it is possible to form halos by degrading non-specific polysaccharides, particularly in the case of the low activity of TSP1. This possibility was tested further by expanding the halo assay to examine other bacterial strains and species. It was found that the TSPs form halos on both *E. coli* O157:H7 strains tested (ATCC 7000728 and ATCC 43894). In contrast, none of the TSPs formed halos when applied to either a non-O157:H7 strain of *E. coli* (ATCC 35218) or other Gram-negative bacterial species (*K. pneumoniae*, *P. aeruginosa*, *A. baumannii*). Although the data initially suggested that TSP1-3 display specificity for the O157:H7 O-antigen, these TSPs were also capable of inducing halos when using the *E. coli* mutants TEA023 (knockout of the *galU* gene) and TEA026 (knockout of the *galETKM* operon), which lack the O157 O-antigen. Taken together, TSP1 and TSP3 glycosidase activity does not appear to require the presence of the O157 antigen. Whether they act on the outer core or inner core moieties of LPS or other bacterial surface polysaccharides remains to be determined.
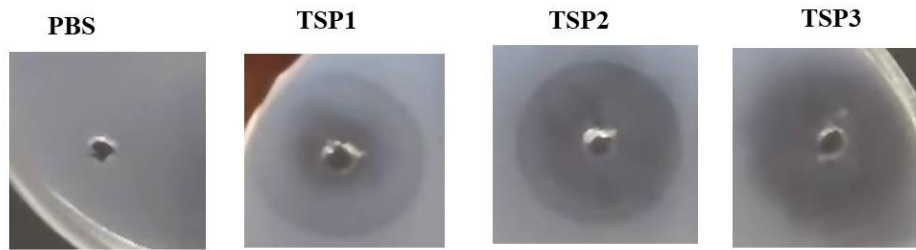
Figure 5.6 Halo formation by phage CBA120 TSPs 1-3.
TSP halo formation on *E. coli* O157:H7 agar plates. PBS negative control shows no halo. TSP1 forms the least distinct halo while TSP2 and TSP3 form similar size halos.

The ability of the TSPs to induce halos provides an assay to probe residues proposed to comprise the catalytic machinery by site-directed mutagenesis. This has been determined for TSP3, and now studies of the TSP2 and TSP1 catalytic residues are in progress.

Using the halo assay, equal weighted amounts of purified wild-type and mutant TSP3 proteins were analyzed (Figure 5.7,Table 5.4) in order to determine which amino acid residues comprise the catalytic machinery. As expected, a halo was observed when spotting wild-type TSP3 (Figure 5.7, well 1), while the buffer only negative control (Figure 5.7, well 9) was devoid of activity. Similar to the negative control, the TSP3 E362Q:D383N:D426N triple mutant (Figure 5.7, well 8) was incapable of generating a halo, implying glycosidase activity was abolished and thus confirming the location of the active site. Individual TSP3 point mutants were assayed next in order to identify which specific residues are required for catalysis. When compared to wild-type, the TSP3 D383N mutant (Figure 5.7, well 3) displayed comparable activity, indicating D383 is dispensable for activity, whereas the E362Q (Figure 5.7, well 2) and D426N (Figure 5.7, well 4)

mutants exhibited activity defects. Contrary to E362Q:D383N (Figure 5.7, well 5) and D383N:D426N (Figure 5.7, well 7), the activity of the TSP3 E362Q:D426N double mutant (Figure 5.7, well 6) was abolished. This result suggests the two essential catalytic residues of TSP3 are E362 and D426, located on two adjacent subunits.

Table 5.4 TSP3 active site mutant halo assay results.

Well	Mutation	Halo Formation
1	N/A	++
2	E362Q	+
3	D383N	++
4	D426N	+
5	E362Q and D383N	+
6	E362Q and D426N	-
7	D383N and D426N	+
8	E362Q, D383N, and D426N	-
9	PBS control	-

Well numbering corresponds to Figure 5.7
Symbols are as follows: -, No observable halo; +, 0-1.5 cm halo diameter; ++, >1.5 cm halo diameter.
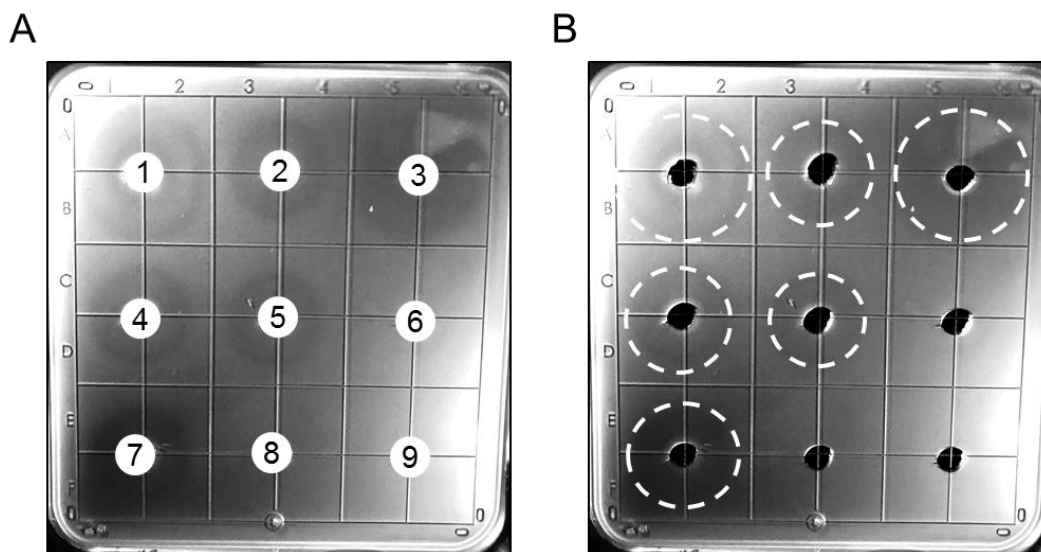
Figure 5.7 TSP3 wild-type and TSP3 active site mutants halo assays.
E. coli strain ATCC 700728 was embedded in agarose. Wells (3 mm) were cut out of the agarose and loaded with 10 µL (6 mg/mL) wild-type TSP3 or active site mutants and incubated overnight to visualize tailspike activity. (A) (1) wild-type TSP3; (2) TSP3 E362Q; (3) TSP D383N; (4) TSP3 D426N; (5) TSP3 E362Q:D383N double mutant; (6) TSP3 E362Q:D426N double mutant; (7) TSP3 D383N:D426N double mutant; (8) TSP3 E362Q:D383N: D426N triple mutant; (9) PBS control. (B) Identical image to (A) with a dashed white line to indicate the halo edge to aid in visualization. Quantitation of Figure 5.7 is shown in Table 5.4.

5.3.4   Conclusions

Several conclusions may be derived from the activity studies. First, with the employed methods, TSP1 has the least observable activity compared with the other tailspikes. Nevertheless, the halo assay confirms that O157:H7 LPS or another polysaccharide is a substrate for the enzyme even though the halo it induces is not as pronounced as the halos from the other TSPs. The turbidity assay and the phage adsorption assay indicate that within the time range of these experiments (minutes), TSP1 lacks enzymatic activity toward the phages receptor. However, the halo assay enables the ongoing site-directed mutagenesis studies that will define the active site residues.

Of the three TSPs, only TSP2 exhibits receptor binding activity, as it was able to inhibit phage CBA120 infection of *E. coli* O157:H7. Thus, it appears the TSP2 is the glycosidase that degrades the *E. coli* O157:H7 LPS. Once the catalytic residues are determined by site-directed mutagenesis, the TSP2 mutants could also be evaluated by the turbidity assay to determine whether the elimination of the enzyme activity also eliminates TSP2's ability to inhibit Phage CBA120 infection.

TSP3's enzymatic activity appears to be similar to that of TSP2 in several aspects. Both TSP2 and TSP3 have more robust halo producing activity than TSP1. Similarly, to TSP2, TSP3 also increases the efficacy of phage adsorption, which may be caused not only competing with the phage for binding to the secondary receptor during the 8 min incubation period but also by enzymatic activity that these TSPs have towards *E. coli* O157:H7's LPS or other polysaccharide during the day period required for plaque formation. However, in contrast to TSP2, TSP3 does not inhibit phage CBA120 infection, and therefore its substrate is probably not the O-antigen. The halo assay with TSP3 mutants confirm that Glu362 and Asp426 comprise that catalytic machinery even though the TSP3 substrate remains elusive.

Chapter 6: Conclusions and Future Prospects

Phage CBA120 specially infects the food pathogen E. coli O157:H7[27]. Exploring the interactions between phage CBA120 and *E. coli* O157:H7 could lead to discovering new biocontrol agents to fight or monitor *E. coli* O157:H7 infections. Through a collaboration between the Herzberg and Nelson Labs, we have been able to reach several conclusions about the structure and function of phage CBA120's tailspike proteins. Electron microscopy images of phage CBA120 showed its unusual tail structure. Bioinformatics studies and sequence analysis allows us to predict structural domains and their function. Finally, the structure determination of TSP1-3 has revealed putative active sites and catalytic residues. The catalytic residues of TSP3 were confirmed by site-directed mutagenesis, and those of TSP1 and TSP2 are currently being investigated.

To better understand CBA120's tailspike proteins in phage infectivity both functional and structural studies will continue to be pursued. Determining the structures of the N-terminal domains of TSP2 and TSP4 will provide valuable insight into the assembly of the tailspike complex. Analytical ultracentrifugation experiments will continue to probe the interaction between the TSPs. Future studies will also focus on determining the substrates of TSP1, TSP3, and TSP4 through biochemical assays. Ultimately these studies will not only provide insight into the infectivity mechanism of phage CBA120 but they will also serve as a paradigm to understand how *Ackermannviridae* infect their host.

Appendix

Appendix 1: ***Protein Sequence and DNA sequences of TSP2-4***

	gene product (gp)	Open Reading Frame	Accession
TSP2	gp163	orf211	YP_004957865.1
TSP3	gp164	orf212	YP_004957866.1
TSP4	gp165	orf213	YP_004957867.1

TSP2

Protein Sequence
```
MTRNVEELFGGVITAPHQIPFTYKSNVGGETFLSLPFYPVTGVVTINGGMQVPLDNFEIEGNTLNLGRALS
KGDVVYCLFDKILSPEDTAKGIRIYKFQAVGGETEFTPDFTSYGVQSLYIGGEYKTPEIEYSYDSTTGKVS
LQTALSAGVWVVAEMSVKQPNISPAFDRSIQEIARSANVKDSEVIVSTDTISLLDGKKVVYDIATQTSYGL
PTIPDGSVISSVSAGKLNYNPGDVQVDLLPLEDSFINVINTLGRNDGAKYIGECHSVADLRNTEPTMDGQR
IILKQHTAGTLLGGGVFRALIDGTGKTDNNGTVIKTVGGAAWLRVNADRVNPFMFGALGGSNDDTIPVQSC
VDSGKATQLTDAHYVSNIQLKYNTSSIYGSGLHYSRLHQLPSATGNCITIKDTCSLIVLDAFGVYGTGAQQ
GTSFTAGTTGIYVETPSGLSADYPFHTTADPRRDLCISKVHIAGFDEYGLNIDSGNFSVTTDSLLVNHINQ
VGVRCATTDWTWTNIQVNTCGKQCLVLDGCGNGRIIGGKFIWANWQPYGTVGQFPGITINNSQNMVINGIE
VQDCGGNGIEISESYSISMNGLNTNRNGINANNTFYNIVFNKSDAVINGFVGLNYAANSGSGANSSAGNFQ
FLSNDCSVTINGVVETGYMGINFIGDNNIINPTNSDLSINGLVNYSKTGLQTMNETPTFDGVSTTPVYVSV
PSSVGQVNGLRLSQANKDKLLYSRTAGPEGITMAAVIVPTISGAEVFNFMAIGSGFSDTSNSLHLQLVIDA
SGKQTIALLLGGDGTTQILSGDLPNDLKLQSGVPYHIAIGAKPGYFWWSILNIQTGKRIRRSFRGAYLAVP
FNSIFGLTSSLTFFSDSNAGGDACSGVGAKVYVGMFSSENDYVASRYYNLINPVDPTKLISYRILDSSI
```

Codon Optimized Gene
```
ATGACCCGCAATGTGGAAGAACTGTTTGGTGGTGTTATTACCGCACCGCATCAGATTCCGTTTACCTATAA
AAGCAATGTTGGTGGCGAAACCTTTCTGAGCCTGCCGTTTTATCCGGTTACCGGTGTTGTTACAATTAATG
GTGGTATGCAGGTTCCGCTGGATAACTTTGAAATTGAAGGTAATACCCTGAATCTGGGTCGTGCACTGAGC
AAAGGTGATGTTGTTTATTGCCTGTTCGACAAAATTCTGAGTCCGGAAGATACCGCAAAAGGTATTCGCAT
TTACAAATTTCAGGCAGTGGGTGGTGAAACCGAATTTACACCGGATTTTACCAGCTATGGTGTTCAGAGCC
TGTATATTGGTGGTGAGTATAAAACACCGGAAATCGAGTATAGCTATGATAGCACCACCGGTAAAGTTAGC
CTGCAGACCGCACTGAGTGCCGGTGTTTGGGTTGTTGCAGAAATGAGCGTTAAACAGCCGAATATTAGTCC
GGCATTTGATCGTAGCATTCAAGAAATTGCACGTAGCGCCAATGTTAAAGATAGCGAAGTTATTGTTAGCA
CCGATACCATTAGCCTGCTGGATGGTAAAAAAGTGGTTTATGATATTGCAACCCAGACCAGTTATGGTCTG
CCGACCATTCCGGATGGTAGCGTTATTAGCAGCGTTAGCGCAGGTAAACTGAACTATAATCCGGGTGATGT
TCAGGTTGATCTGCTGCCGCTGGAAGATAGCTTTATTAACGTGATTAATACGCTGGGTCGTAACGATGGTG
CCAAATATATCGGTGAATGTCATAGCGTTGCCGATCTGCGTAATACCGAACCGACCATGGATGGTCAGCGC
ATTATTCTGAAACAGCATACCGCAGGCACCCTGCTGGGTGGTGGTGTGTTTCGTGCCCTGATTGATGGTAC
AGGTAAAACCGATAATAATGGCACCGTGATTAAAACCGTTGGTGGTGCAGCATGGCTGCGTGTTAATGCAG
ATCGTGTTAATCCGTTTATGTTTGGTGCACTGGGTGGTAGCAATGATGATACAATTCCGGTTCAGAGCTGT
GTTGATAGCGGTAAAGCAACCCAGCTGACCGATGCACATTATGTTAGCAATATCCAGCTGAAATATAACAC
```

```
CAGCAGCATTTATGGTAGTGGCCTGCATTATAGCCGTCTGCATCAGCTGCCGAGCGCAACCGGTAATTGCA
TTACCATTAAAGATACCTGCAGCCTGATTGTTCTGGATGCATTTGGTGTTTATGGTACGGGTGCACAGCAG
GGCACCAGCTTTACAGCAGGTACAACCGGTATTTATGTGGAAACCCCGAGCGGTCTGAGCGCAGATTATCC
GTTTCATACCACCGCAGATCCGCGTCGTGATCTGTGTATTAGCAAAGTTCATATTGCCGGTTTTGATGAGT
ATGGCCTGAATATTGATTCCGGCAATTTTAGCGTTACCACCGATAGTCTGCTGGTGAATCATATTAATCAG
GTTGGTGTTCGTTGTGCAACCACCGATTGGACCTGGACCAACATTCAGGTTAATACCTGTGGTAAACAGTG
TCTGGTACTGGATGGTTGTGGTAATGGTCGTATTATTGGTGGCAAATTTATCTGGGCAAATTGGCAGCCGT
ATGGTACAGTTGGTCAGTTTCCGGGTATTACGATTAATAACAGTCAGAACATGGTGATCAACGGCATTGAA
GTTCAGGATTGTGGTGGCAATGGTATTGAAATTAGCGAAAGCTATAGCATCAGCATGAATGGTCTGAATAC
CAATCGCAACGGTATTAATGCCAACAACACCTTCTATAACATCGTGTTCAATAAAAGCGACGCGGTGATTA
ATGGTTTTGTGGGTCTGAATTATGCAGCAAATAGCGGTAGCGGTGCAAATAGCAGTGCAGGTAATTTTCAG
TTTCTGAGCAATGATTGTAGCGTGACCATTAACGGTGTGGTTGAAACCGGTTATATGGGCATTAATTTCAT
CGGCGACAACAACATTATTAACCCGACCAATAGCGATCTGAGCATTAATGGCCTGGTGAATTATAGCAAAA
CCGGACTGCAGACCATGAATGAAACACCGACCTTTGATGGTGTGAGCACCACACCGGTTTATGTGAGCGTT
CCGAGCAGCGTTGGTCAGGTGAATGGCCTGCGTCTGAGCCAGGCAAATAAAGATAAACTGCTGTATAGCCG
TACCGCAGGTCCGGAAGGTATTACCATGGCAGCGGTTATTGTTCCGACCATTAGTGGTGCCGAAGTGTTTA
ACTTTATGGCAATTGGTAGCGGTTTTAGCGATACCTCAAATAGCCTGCATCTGCAGCTGGTGATTGATGCA
AGCGGTAAACAGACCATTGCGCTGCTGCTGGGAGGTGATGGCACCACACAGATTCTGAGCGGTGATCTGCC
GAATGATCTGAAACTGCAGAGCGGTGTTCCGTATCATATTGCAATTGGTGCTAAACCGGGTTATTTTGGT
GGTCCATTCTGAACATTCAGACCGGTAAACGTATTCGTCGTAGCTTTCGTGGTGCATATCTGGCAGTTCCG
TTTAATAGCATTTTTGGTCTGACCAGCAGCCTGACCTTTTTTAGTGATAGCAATGCCGGTGGTGATGCATG
TAGCGGTGTTGGTGCAAAAGTTTATGTTGGTATGTTTAGCAGCGAGAACGATTATGTTGCAAGCCGCTATT
ATAACCTGATCAATCCGGTGGATCCGACCAAACTGATTAGCTATCGTATTCTGGATAGCAGCATTCATCAT
CACCATCACCATTAA
```

TSP3
Protein Sequence

```
MISQFNQPRGSTSIEVNKQSIARNFGVKEDEVIYFTAGIDLSGFKVIYDESTQRAYSLPFGIVSGTTAISL
DERAILTHSAGSVDLGELAVSREEYVTLPGSFNFGHTINVKNELLVHDDKKYRWDGSLPKVVAAGSTPDSS
GGVGLGAWLSVGDAALRAELNTKVSDGTFPATIKYKYGLPSVIDGAIYRTVQDKLDDFVFLEDFGGKDDAG
STDNSIAFRKAFASGARKIRLRGSGVYGMATRDIELPAKYEIIGNAKNPEIKYLGTDTSFTMFTLTGSGPA
SNQWKQGGMFRDLIISSDVKINWMLGRHVQNLDYDRVFFYNSATVLNNYHYVNFTRCERWGSAFIGRADLN
TIQFISESPKFHLCFSSGSPIDVWDTADLAITKCTMFAGDYAVRTRVTQKQVTAPDLFAGYPVLITCSVFD
AVRGHAWDLEGSVYSTITGNLVSAGRDTNSHGAYIKGGRSLSLTGNVFTYCGNYGLVLEDVQQSGFVGNVF
NGNKTGGLGTLACKDLSIVGGSMGTTYVRGGYYTQPVGYSDISSNSTGILLSGVAFDEALTTKVYLDTSIT
TRNKVINCSGVPDTIARGSTANRPANPQASYQYYDTTLGIPIWWNSVSGTWKNAAGADV
```

Codon Optimized Gene
```
ATGATCAGCCAGTTTAATCAGCCTCGTGGTAGCACCAGCATTGAAGTTAATAAACAGAGCATTGCCCGTAA
CTTTGGCGTGAAAGAAGATGAGGTTATCTATTTTACCGCAGGCATTGATCTGAGCGGCTTTAAAGTTATTT
ATGATGAAAGCACCCAGCGTGCATATAGCCTGCCGTTTGGTATTGTTAGCGGCACCACCGCAATTAGCCTG
GATGAACGTGCAATTCTGACCCATAGCGCAGGTAGCGTTGATCTGGGTGAACTGGCAGTTAGCCGTGAAGA
ATATGTTACCCTGCCTGGTAGCTTTAACTTTGGTCATACCATTAACGTGAAAAACGAACTGCTGGTGCACG
ACGACAAAAAATACCGTTGGGATGGTAGTCTGCCGAAAGTTGTTGCAGCAGGTAGCACACCGGATAGCAGC
GGTGGTGTTGGTCTGGGTGCATGGCTGAGCGTTGGTGATGCAGCACTGCGTGCAGAACTGAATACCAAAGT
TTCAGATGGCACCTTTCCGGCAACCATCAAATACAAATATGGTCTGCCGAGCGTTATTGATGGTGCAATTT
ATCGTACCGTTCAGGATAAACTGGATGATTTTGTGTTCCTGGAAGATTTCGGTGGTAAAGATGATGCCGGT
AGCACCGATAATAGCATTGCATTTCGTAAAGCATTTGCAAGCGGTGCACGTAAAATTCGTCTGCGTGGTAG
CGGTGTTTATGGTATGGCAACCCGTGATATTGAACTGCCTGCAAAATATGAAATTATCGGCAATGCCAAAA
ACCCGGAAATCAAATATCTGGGCACCGATACCAGCTTTACCATGTTTACCCTGACCGGTAGTGGTCCGGCA
AGCAATCAGTGGAAACAGGGTGGTATGTTTCGTGATCTGATTATTAGCTCCGACGTGAAAATCAATTGGAT
GCTGGGTCGTCATGTTCAGAATCTGGATTATGATCGCGTGTTCTTTTATAACAGCGCAACCGTTCTGAACA
ACTATCACTATGTGAATTTTACCCGTTGTGAACGTTGGGGTAGCGCATTTATTGGTCGTGCCGATCTGAAT
ACGATTCAGTTTATTAGCGAAAGCCCGAAATTTCACCTGTGTTTTAGCAGTGGTAGCCCGATTGATGTTTG
GGATACCGCAGATCTGGCAATTACCAAATGCACCATGTTTGCCGGTGATTATGCAGTTCGTACCCGTGTTA
```

```
CCCAGAAACAGGTTACCGCACCGGACCTGTTTGCAGGTTATCCGGTTCTGATTACCTGTAGCGTTTTTGAT
GCCGTTCGTGGTCATGCATGGGATCTGGAAGGTAGCGTGTATAGCACCATTACCGGCAATCTGGTGAGCGC
AGGTCGTGATACCAATAGCCATGGTGCCTATATCAAAGGTGGTCGTAGCCTGAGCCTGACCGGCAATGTTT
TTACCTATTGTGGTAATTATGGTCTGGTGCTGGAAGATGTTCAGCAGAGCGGTTTTGTTGGTAATGTGTTT
AATGGCAATAAAACCGGTGGCCTGGGCACCCTGGCATGTAAAGATCTGAGCATTGTTGGTGGTAGCATGGG
CACCACCTATGTTCGTGGTGGTTATTATACCCAGCCGGTTGGTTATAGCGATATTAGCAGCAATAGCACCG
GTATTCTGCTGAGCGGTGTTGCATTTGATGAAGCACTGACCACCAAAGTGTATCTGGATACCAGTATTACC
ACCCGCAATAAAGTGATTAATTGCAGCGGTGTTCCGGATACCATTGCACGTGGTAGTACCGCAAATCGTCC
GGCAAATCCGCAGGCCAGCTATCAGTATTATGATACCACACTGGGTATTCCGATTTGGTGGAATAGCGTTA
GCGGTACATGGAAAAATGCAGCCGGTGCAGATGTTCATCATCATCACCATCATTAA
```


TSP4

Protein Sequence
```
MANKPTQPLFPLGLETSESSNIKGFNNSGTIEHSPGAVMTFPEDTEVTGLPSSVRYNPDSDEFEGYYENGG
WLSLGGGGIRWETLPHAPSSNLLEGRGYLINNTTGTSTVVLPSPTRIGDSVTICDAYGKFATYPLTVSPSG
NNLYGSTEDMAITTDNVSATFTWSGPEQGWVITSGVGLGQGRVYSREIFTQILASETSAVTLNTPPTIVDV
YADGKRLAESKYSLDGNVITFSPSLPASTELQVIEYTPIQLGNGGGSGSSTITWVYNGGSAIGGETEITLD
IVVDDVPAIDINGSRQYKNLGFTFDPLTSKITLAQELDAEDEVVVIINGTPNIYNQIDYTLREVARVTNVK
DTEVIYFSVGAVLSGYKVIYDKVTQRSYFIPELPTGTTAVSLSSSAILVHSAGSVDLGALAVSREEYVTLS
GTFDSGAVINTKNELLTHTDGKYRWDGTLPKTVAAGSTPATTGGVGSGAWLSVGDASLKSNLNKPNGLSYI
GTVSSVSELSSIAGLIGDSIILDSYVDGFNLGGGVMVAVNSDTVVDNIVTFQGNGVVWKRKLFNGVADVYE
AGYTGTGDLAIFINKINAVGFDCIVPVSGEITTPIIFDIAKGALIGKNKCTLIESASATGDYYLTIVNTDT
DYTNRDVINATALMTGVSFVGKGTRKLAIGGSTSGEVSELRISNCGFISTAGIEFLDNAYRILFDKCALSR
SFTNSVIFNSPANSGEVIKFNHCWMVDNGGPFTFKNGQFIFDSCSLPAGKKSGYFDPVVALSDNATTVFTN
GNIEYQPGQSFVGFTVDGSSRLSISDSTILLPNDYSTVPIVNNGDGVVSLNNCSLPLYGSTTIATGFATRQ
LIGGLSKKIMSRGCYPRAGFITSNWNLGCIVSPYINSVSNGSGQFENISNWTLSQTGTDVVTVTTGNDVPN
DLMFSTSFVLSVPTVGAAANFTQTIIDCEPGRYFQLGFWAKNTTTTLASIRFLDQQGNAVADSIGYNIPVG
NTFNFYALVDCVPPGAYKAEINFNVSSIVGGIAIHNVIYGLI
```

Codon Optimized Gene
```
ATGGCCAATAAACCGACCCAGCCGCTGTTTCCGCTGGGTCTGGAAACCAGCGAAAGCAGCAACATTAAAGG
TTTTAACAATAGCGGCACCATCGAACATAGTCCGGGTGCAGTTATGACCTTTCCGGAAGATACCGAAGTTA
CCGGTCTGCCGAGCAGCGTTCGTTATAATCCGGATAGTGATGAATTTGAGGGCTATTATGAAAATGGTGGT
TGGCTGAGCCTGGGTGGTGGTGGTATTCGTTGGGAAACCCTGCCGCATGCACCGAGCAGCAATCTGCTGGA
AGGTCGTGGTTATCTGATTAATAACACCACCGGCACCAGCACCGTTGTGCTGCCGAGCCCGACCCGTATTG
GTGATAGCGTTACCATTTGTGATGCCTATGGAAAATTTGCAACCTATCCGCTGACCGTTAGCCCGAGCGGT
AATAATCTGTATGGTAGCACCGAAGATATGGCAATTACCACCGATAATGTTAGCGCAACCTTTACCTGGTC
AGGTCCGGAACAGGGTTGGGTTATTACCAGCGGTGTTGGTCTGGGTCAGGGTCGTGTTTATAGCCGTGAAA
TCTTTACCCAGATTCTGGCAAGCGAAACCAGTGCAGTTACCCTGAATACCCCTCCGACCATTGTTGATGTT
TATGCAGATGGTAAACGTCTGGCCGAAAGCAAATATTCACTGGATGGTAATGTGATTACCTTTAGCCCGTC
ACTGCCTGCAAGCACCGAACTGCAGGTTATTGAGTATACCCCGATTCAGCTGGGTAATGGTGGTGGCAGCG
GTAGCAGTACCATTACCTGGGTTTATAATGGCGGTAGCGCAATTGGTGGTGAAACCGAAATTACCCTGGAT
ATTGTTGTTGATGATGTTCCGGCAATTGATATTAATGGTAGCCGTCAGTATAAAAACCTGGGCTTTACCTT
TGATCCTCTGACCAGCAAAATTACACTGGCACAAGAACTGGATGCAGAAGATGAAGTTGTTGTTATTATCA
ATGGCACCCCGAACATCTATAACCAGATTGATTATACCCTGCGTGAAGTTGCACGTGTTACCAATGTTAAA
GATACGGAAGTGATCTATTTTAGCGTTGGTGCAGTTCTGAGCGGTTACAAAGTGATTTATGATAAAGTTAC
CCAGCGCAGCTATTTTATCCCGGAACTGCCGACCGGTACAACCGCAGTTAGCCTGAGCAGCAGCGCAATTC
TGGTTCATAGCGCAGGTAGCGTTGATCTGGGTGCACTGGCAGTGAGCCGTGAAGAATATGTGACCCTGAGC
GGCACCTTTGATAGCGGTGCAGTGATTAATACCAAAAATGAACTGCTGACCCATACCGATGGTAAATATCG
TTGGGATGGCACCCTGCCGAAAACCGTTGCAGCAGGTAGCACACCGGCAACCACAGGTGGTGTTGGTAGCG
GTGCCTGGCTGAGTGTTGGTGATGCAAGCCTGAAAAGCAATCTGAACAAACCGAATGGTCTGAGCTATATT
GGCACCGTTAGCAGCGTTAGCGAACTGAGCAGTATTGCAGGTCTGATTGGTGATTCAATTATCCTGGATTC
ATATGTGGATGGCTTTAATCTGGGTGGCGGTGTTATGGTTGCAGTTAATAGCGATACCGTGGTGGATAACA
TTGTTACCTTTCAGGGCAATGGTGTTGTGTGGAAACGTAAACTGTTTAATGGTGTGGCCGATGTGTATGAA
```

110

```
GCAGGTTATACCGGCACCGGTGATCTGGCAATTTTCATTAACAAAATTAACGCCGTGGGCTTCGATTGTAT
TGTTCCGGTTAGCGGTGAAATTACCACACCGATTATCTTTGATATTGCAAAAGGTGCCCTGATCGGCAAAA
ACAAATGTACCCTGATTGAAAGCGCAAGCGCAACCGGTGATTATTATCTGACCATTGTGAATACCGATACC
GATTATACCAACCGCGACGTTATTAATGCCACCGCACTGATGACCGGTGTTAGCTTTGTTGGTAAAGGCAC
CCGTAAACTGGCCATTGGTGGTAGCACAAGCGGTGAAGTGAGCGAACTGCGTATTAGCAATTGTGGTTTTA
TTAGCACCGCAGGCATTGAGTTTCTGGATAATGCATATCGCATCCTGTTTGATAAATGTGCACTGAGCCGT
AGCTTTACCAATAGCGTTATCTTTAATAGTCCGGCAAATAGTGGCGAGGTGATCAAATTTAACCATTGCTG
GATGGTTGATAACGGTGGTCCGTTTACCTTTAAAAACGGTCAGTTTATCTTCGATAGCTGTAGCCTGCCTG
CAGGTAAAAAAAGCGGTTATTTTGATCCGGTTGTTGCCCTGAGCGATAATGCAACCACCGTGTTTACGAAT
GGCAACATTGAATATCAGCCTGGCCAGAGCTTTGTGGGTTTTACCGTTGATGGTAGCAGCCGTCTGAGCAT
TAGCGATAGCACAATTCTGCTGCCGAATGATTATTCAACCGTTCCGATTGTTAATAACGGTGATGGTGTGG
TGAGCCTGAATAATTGCAGCCTGCCGCTGTATGGCAGTACCACCATTGCAACCGGTTTTGCAACCCGTCAG
CTGATTGGCGGTCTGAGCAAAAAAATCATGAGCCGTGGTTGTTATCCGCGTGCAGGCTTTATTACCTCAAA
TTGGAATCTGGGCTGTATTGTTAGCCCGTATATTAACAGCGTGAGCAATGGTAGTGGTCAGTTTGAAAATA
TCAGCAATTGGACACTGAGCCAGACCGGCACAGATGTTGTTACCGTGACCACCGGTAATGATGTGCCGAAC
GATCTGATGTTTAGCACCAGCTTTGTGCTGAGCGTTCCGACCGTTGGTGCCGCAGCCAATTTCACCCAGAC
AATTATTGATTGTGAACCGGGTCGTTATTTCCAGCTGGGTTTTTGGGCAAAAAATACCACAACAACCCTGG
CAAGCATTCGTTTTCTGGATCAGCAGGGTAACGCAGTTGCAGATAGCATTGGTTATAACATTCCGGTTGGC
AACACCTTTAACTTTTATGCCCTGGTTGATTGTGTTCCGCCTGGTGCATATAAAGCCGAAATTAACTTTAA
CGTGAGCAGCATTGTGGGTGGTATTGCGATTCATAATGTTATCTATGGCCTGATCCATCATCACCACCATC
ATTAA
```

Appendix 2: *Primers*

Appendix Table 1 Primers for TSP3 Site-Directed Mutagenesis

Mutation	Template	Sequence (5' > 3')*
E326Q	pBAD24::*TSP3*	[Phos]- ATTCAGTTTATTAGC**CAG**AGCCCGAAATTTCAC
D383N	pBAD24::*TSP3*	[Phos]- GTTTGGGATACCGCA**AAC**CTGGCAATTACCAAA
D426N	pBAD24::*TSP3*	[Phos]- ACCTGTAGCGTTTTT**AAC**GCCGTTCGTGGTCAT
E326Q D383N	pBAD24::*TSP3* *E362Q*	[Phos]- GTTTGGGATACCGCA**AAC**CTGGCAATTACCAAA
E326Q D426N	pBAD24::*TSP3* *E362Q*	[Phos]- ACCTGTAGCGTTTTT**AAC**GCCGTTCGTGGTCAT
E326Q D383N D426N	pBAD24::*TSP3* *E362Q D383N*	[Phos]- ACCTGTAGCGTTTTT**AAC**GCCGTTCGTGGTCAT

* Point mutations bold

Appendix Table 2 Primers for TSP4 Constructs

TSP4 Construct	Template	Sequence (5' > 3')	Restriction enzyme
OP1-F	pBAD24::*TSP4*	[Phos]-CGCGAATTCATTATGGCCAATAAACC	EcoR1
OP2-F	pBAD24::*TSP4*	[Phos]-CGCGAATTCATGACCTTTCCGGAAGATACCGAAG	EcoR1
OP3-F	pBAD24::*TSP4*	[Phos]-CGCGAATTCATGAACATCTATAACCAGATTGATTATAC	EcoR1
OP4-F	pBAD24::*TSP4*	[Phos]-CGCGAATTCATGGTTGGTGATGCAAGCCTGAAAAGC	EcoR1
OP5-F	pBAD24::*TSP4*	[Phos]-CGCGAATTCATGGGCACCGGTGATCTGGCAATTTTC	EcoR1
OP7-F	pBAD24::*TSP4*	[Phos]-CGCGAATTCATGGCAATTACCACCGATAATGTTAGC	EcoR1
OP12-R	pBAD24::*TSP4*	[Phos]-CGCTCTAGATTAATGATGGTGGTGATGATGCGGGGTGCCATTGATAATAACAAC	Xba1
END-R	pBAD24::*TSP4*	[Phos]-CGCTCTAGATTAATGATGGTGGTGATGATGGATCAGGCCATAGATAACATTATG	Xba1

Bibliography

1. Pires, D. P.; Oliveira, H.; Melo, L. D. R.; Sillankorva, S.; Azeredo, J., Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Applied Microbiology and Biotechnology* **2016,** *100* (5), 2141-2151.

2. Wells, J. G.; Davis, B. R.; Wachsmuth, I. K.; Riley, L. W.; Remis, R. S.; Sokolow, R.; Morris, G. K., Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare Escherichia coli serotype. *Journal of Clinical Microbiology* **1983,** *18* (3), 512-520.

3. Bavaro, M. F., E. coli O157:H7 and Other Toxigenic Strains: The Curse of Global Food Distribution. *Current Gastroenterology Reports* **2012,** *14* (4), 317-323.

4. Goldwater, P. N.; Bettelheim, K. A., Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Medicine* **2012,** *10* (1).

5. Matthews, K. R.; Kniel, K. E.; Montville, T. J., *Food Microbiology: An Introduction, Fourth Edition*. American Society of Microbiology: 2017.

6. Scallan, E.; Hoekstra, R. M.; Angulo, F. J.; Tauxe, R. V.; Widdowson, M.-A.; Roy, S. L.; Jones, J. L.; Griffin, P. M., Foodborne Illness Acquired in the United States—Major Pathogens. *Emerging Infectious Diseases* **2011,** *17* (1), 7-15.

7. Summary of Recall Cases in Calendar Year 2017. *United States Department of Agriculture* **2018**.

8. *Summary of Recall Cases in Calendar Year 2013*; Recall Summaries 2017; United States Department of Agriculture: 2015//03/24, 2015.

9.   Important Bankruptcy Notice. https://www.soynutbutter.com/.

10.   Scharff, R. L., State Estimates for the Annual Cost of Foodborne Illness. *Journal of Food Protection* **2015,** *78* (6), 1064-1071.

11.   Soon, J. M.; Chadd, S. A.; Baines, R. N., Escherichia coli O157:H7 in beef cattle: on farm contamination and pre-slaughter control methods. *Animal Health Research Reviews* **2011,** *12* (02), 197-211.

12.   Rangel, J. M.; Sparling, P. H.; Crowe, C.; Griffin, P. M.; Swerdlow, D. L., Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. *Emerging Infectious Diseases* **2005,** *11* (4), 603-609.

13.   Hughes, K. A.; Sutherland, I. W.; Jones, M. V., Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase. *Microbiology (Reading, England)* **1998,** *144 (Pt 11)*, 3039-3047.

14.   Silagyi, K. S. Biofilm Formation By Escherichia Coli O157:H7. pdf, University of Maryland, 2007.

15.   Costerton, J. W., Bacterial Biofilms: A Common Cause of Persistent Infections. *Science* **1999,** *284* (5418), 1318-1322.

16.   Raetz, C. R. H.; Whitfield, C., Lipopolysaccharide Endotoxins. *Annual Review of Biochemistry* **2002,** *71* (1), 635-700.

17.   Wang, L.; Reeves, P. R., Organization of Escherichia coli O157 O Antigen Gene Cluster and Identification of Its Specific Genes. *Infection and Immunity* **1998,** *66* (8), 3545-3551.

18.   Samuel, G.; Hogbin, J. P.; Wang, L.; Reeves, P. R., Relationships of the Escherichia coli O157, O111, and O55 O-Antigen Gene Clusters with Those of

Salmonella enterica and Citrobacter freundii, Which Express Identical O Antigens. *Journal of Bacteriology* **2004,** *186* (19), 6536-6543.

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

20. Casjens, S. R., Comparative genomics and evolution of the tailed-bacteriophages. *Current Opinion in Microbiology* **2005,** *8* (4), 451-458.

21. Hughes, K. A.; Sutherland, I. W.; Clark, J.; Jones, M. V., Bacteriophage and associated polysaccharide depolymerases--novel tools for study of bacterial biofilms. *Journal of Applied Microbiology* **1998,** *85* (3), 583-590.

22. Casjens, S.; J Molineux, I., *Short Noncontractile Tail Machines: Adsorption and DNA Delivery by Podoviruses*. 2012; Vol. 726, p 143-79.

23. Orlova, E. V., Bacteriophages and Their Structural Organisation. In *Bacteriophages*, Kurtbke, p., Ed. InTech: 2012.

24. Adriaenssens, E. M.; Ackermann, H.-W.; Anany, H.; Blasdel, B.; Connerton, I. F.; Goulding, D.; Griffiths, M. W.; Hooton, S. P.; Kutter, E. M.; Kropinski, A. M.; Lee, J.-H.; Maes, M.; Pickard, D.; Ryu, S.; Sepehrizadeh, Z.; Shahrbabak, S. S.; Toribio, A. L.; Lavigne, R., A suggested new bacteriophage genus: "Viunalikevirus". *Archives of Virology* **2012,** *157* (10), 2035-2046.

25. Park, M.; Lee, J.-H.; Shin, H.; Kim, M.; Choi, J.; Kang, D.-H.; Heu, S.; Ryu, S., Characterization and Comparative Genomic Analysis of a Novel Bacteriophage, SFP10, Simultaneously Inhibiting both Salmonella enterica and

Escherichia coli O157:H7. *Applied and Environmental Microbiology* **2012,** *78* (1), 58-69.

26.   Lim, J. Y.; Yoon, J. W.; Hovde, C. J., A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157. *Journal of microbiology and biotechnology* **2010,** *20* (1), 5-14.

27.   Kutter, E. M.; Skutt-Kakaria, K.; Blasdel, B.; El-Shibiny, A.; Castano, A.; Bryan, D.; Kropinski, A. M.; Villegas, A.; Ackermann, H.-W.; Toribio, A. L.; Pickard, D.; Anany, H.; Callaway, T.; Brabban, A. D., Characterization of a ViI-like Phage Specific to Escherichia coli O157:H7. *Virology Journal* **2011,** *8* (1), 430.

28.   Adriaenssens, E. M.; Van Vaerenbergh, J.; Vandenheuvel, D.; Dunon, V.; Ceyssens, P.-J.; De Proft, M.; Kropinski, A. M.; Noben, J.-P.; Maes, M.; Lavigne, R., T4-Related Bacteriophage LIMEstone Isolates for the Control of Soft Rot on Potato Caused by 'Dickeya solani'. *PLoS ONE* **2012,** *7* (3), e33227.

29.   Shahrbabak, S. S.; Khodabandehlou, Z.; Shahverdi, A. R.; Skurnik, M.; Ackermann, H. W.; Varjosalo, M.; Yazdi, M. T.; Sepehrizadeh, Z., Isolation, characterization and complete genome sequence of PhaxI: a phage of Escherichia coli O157 : H7. *Microbiology* **2013,** *159* (Pt_8), 1629-1638.

30.   Hooton, S. P. T.; Timms, A. R.; Rowsell, J.; Wilson, R.; Connerton, I. F., Salmonella Typhimurium-specific bacteriophage ΦSH19 and the origins of species specificity in the Vi01-like phage family. *Virology Journal* **2011,** *8* (1), 498.

31. Bradley, D. E.; Kay, D., The Fine Structure of Bacteriophages. *Journal of General Microbiology* **1960,** *23* (3), 553-563.

32. Anany, H.; Lingohr, E. J.; Villegas, A.; Ackermann, H.-W.; She, Y.-M.; Griffiths, M. W.; Kropinski, A. M., A Shigella boydii bacteriophage which resembles Salmonella phage ViI. *Virology Journal* **2011,** *8* (1), 242.

33. Turner, D.; Reynolds, D.; Seto, D.; Mahadevan, P., CoreGenes3.5: a webserver for the determination of core genes from sets of viral and small bacterial genomes. *BMC Research Notes* **2013,** *6* (1), 140.

34. Alikhan, N.-F.; Petty, N. K.; Ben Zakour, N. L.; Beatson, S. A., BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **2011,** *12* (1).

35. Kropinski, A. M.; Anany, H.; Kuhn, J. H.; Tolstoy, I.; Kutter, E.; Adriaenssens, E. M. Proposal 2017.001B.A.v3.Ackermannviridae. http:// ictv.globa l/propo sals-17/2017.001B.A.v3.Acker mannv irida e.zip.

36. King, A. M. Q.; Lefkowitz, E. J.; Mushegian, A. R.; Adams, M. J.; Dutilh, B. E.; Gorbalenya, A. E.; Harrach, B.; Harrison, R. L.; Junglen, S.; Knowles, N. J.; Kropinski, A. M.; Krupovic, M.; Kuhn, J. H.; Nibert, M. L.; Rubino, L.; Sabanadzovic, S.; Sanfaçon, H.; Siddell, S. G.; Simmonds, P.; Varsani, A.; Zerbini, F. M.; Davison, A. J. J. A. o. V., Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). **2018,** *163* (9), 2601-2631.

37.   Adriaenssens, E. M.; Wittmann, J.; Kuhn, J. H.; Turner, D.; Sullivan, M. B.; Dutilh, B. E.; Jang, H. B.; van Zyl, L. J.; Klumpp, J.; Lobocka, M.; Moreno Switt, A. I.; Rumnieks, J.; Edwards, R. A.; Uchiyama, J.; Alfenas-Zerbini, P.; Petty, N. K.; Kropinski, A. M.; Barylski, J.; Gillis, A.; Clokie, M. R. C.; Prangishvili, D.; Lavigne, R.; Aziz, R. K.; Duffy, S.; Krupovic, M.; Poranen, M. M.; Knezevic, P.; Enault, F.; Tong, Y.; Oksanen, H. M.; Rodney Brister, J. J. A. o. V., Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. **2018,** *163* (4), 1125-1129.

38.   Chen, C.; Bales, P.; Greenfield, J.; Heselpoth, R. D.; Nelson, D. C.; Herzberg, O., Crystal structure of ORF210 from E. coli O157:H1 phage CBA120 (TSP1), a putative tailspike protein. *PloS One* **2014,** *9* (3), e93156.

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

40.   Sturtevant, J. M.; Yu, M. H.; Haase-Pettingell, C.; King, J., Thermostability of temperature-sensitive folding mutants of the P22 tailspike protein. *The Journal of Biological Chemistry* **1989,** *264* (18), 10693-10698.

41.   Prokhorov, N. S.; Riccio, C.; Zdorovenko, E. L.; Shneider, M. M.; Browning, C.; Knirel, Y. A.; Leiman, P. G.; Letarov, A. V., Function of bacteriophage G7C esterase tailspike in host cell adsorption: Function of bacteriophage G7C esterase tailspike. *Molecular Microbiology* **2017,** *105* (3), 385-398.

42.   Scholl, D.; Merril, C., The Genome of Bacteriophage K1F, a T7-Like Phage That Has Acquired the Ability To Replicate on K1 Strains of <em>Escherichia coli</em&gt. *Journal of Bacteriology* **2005,** *187* (24), 8499.

43.   Steinbacher, S.; Baxa, U.; Miller, S.; Weintraub, A.; Seckler, R.; Huber, R., Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors. *Proceedings of the National Academy of Sciences* **1996,** *93* (20), 10584.

44.   Chang, J.; Weigele, P.; King, J.; Chiu, W.; Jiang, W., Cryo-EM Asymmetric Reconstruction of Bacteriophage P22 Reveals Organization of its DNA Packaging and Infecting Machinery. *Structure* **2006,** *14* (6), 1073-1082.

45.   Pintilie, G.; Chen, D.-H.; Haase-Pettingell, Cameron A.; King, Jonathan A.; Chiu, W., Resolution and Probabilistic Models of Components in CryoEM Maps of Mature P22 Bacteriophage. *Biophysical Journal* **2016,** *110* (4), 827-839.

46.   Steinbacher, S.; Seckler, R.; Miller, S.; Steipe, B.; Huber, R.; Reinemer, P., Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science (New York, N.Y.)* **1994,** *265* (5170), 383-386.

47.   Baxa, U.; Steinbacher, S.; Miller, S.; Weintraub, A.; Huber, R.; Seckler, R., Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide. *Biophysical Journal* **1996,** *71* (4), 2040-2048.

48.   Andres, D.; Gohlke, U.; Broeker, N. K.; Schulze, S.; Rabsch, W.; Heinemann, U.; Barbirz, S.; Seckler, R., 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), 486-494.

49. Kulikov, E.; Kropinski, A. M.; Golomidova, A.; Lingohr, E.; Govorun, V.; Serebryakova, M.; Prokhorov, N.; Letarova, M.; Manykin, A.; Strotskaya, A.; Letarov, A., Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB_EcoP_G7C. *Virology* **2012,** *426* (2), 93-99.

50. Choi, K. H.; McPartland, J.; Kaganman, I.; Bowman, V. D.; Rothman-Denes, L. B.; Rossmann, M. G., Insight into DNA and Protein Transport in Double-Stranded DNA Viruses: The Structure of Bacteriophage N4. *Journal of Molecular Biology* **2008,** *378* (3), 726-736.

51. Notredame, C.; Higgins, D. G.; Heringa, J., T-coffee: a novel method for fast and accurate multiple sequence alignment11Edited by J. Thornton. *Journal of Molecular Biology* **2000,** *302* (1), 205-217.

52. Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A. N.; Alva, V., A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **2018,** *430* (15), 2237-2243.

53. Leiman, P. G.; Shneider, M. M.; Mesyanzhinov, V. V.; Rossmann, M. G., Evolution of Bacteriophage Tails: Structure of T4 Gene Product 10. *Journal of Molecular Biology* **2006,** *358* (3), 912-921.

54. Leiman, P. G.; Arisaka, F.; van Raaij, M. J.; Kostyuchenko, V. A.; Aksyuk, A. A.; Kanamaru, S.; Rossmann, M. G., Morphogenesis of the T4 tail and tail fibers. *Virology Journal* **2010,** *7* (1), 355.

55.   Yap, M. L.; Klose, T.; Arisaka, F.; Speir, J. A.; Veesler, D.; Fokine, A.; Rossmann, M. G., Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proceedings of the National Academy of Sciences* **2016,** *113* (10), 2654.

56.   Leiman, P. G.; Kanamaru, S.; Mesyanzhinov, V. V.; Arisaka, F.; Rossmann, M. G., Structure and morphogenesis of bacteriophage T4. *Cellular and Molecular Life Sciences (CMLS)* **2003,** *60* (11), 2356-2370.

57.   Pickard, D.; Toribio, A. L.; Petty, N. K.; van Tonder, A.; Yu, L.; Goulding, D.; Barrell, B.; Rance, R.; Harris, D.; Wetter, M.; Wain, J.; Choudhary, J.; Thomson, N.; Dougan, G., A Conserved Acetyl Esterase Domain Targets Diverse Bacteriophages to the Vi Capsular Receptor of *Salmonella* enterica Serovar Typhi. *Journal of Bacteriology* **2010,** *192* (21), 5746.

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

59.   Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T., Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research* **2013,** *41* (Web Server issue), W349-357.

60.   Barbirz, S.; Müller, J. J.; Uetrecht, C.; Clark, A. J.; Heinemann, U.; Seckler, R., Crystal structure of Escherichia coli phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. *Molecular Microbiology* **2008,** *69* (2), 303-316.

61.   Kabsch, W., Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D* **2010,** *66* (2), 133-144.

62. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallographica Section D* **2013,** *69* (7), 1204-1214.

63. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **2010,** *66* (2), 213-221.

64. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004,** *60* (12 Part 1), 2126-2132.

65. Echols, N.; Moriarty, N. W.; Klei, H. E.; Afonine, P. V.; Bunkoczi, G.; Headd, J. J.; McCoy, A. J.; Oeffner, R. D.; Read, R. J.; Terwilliger, T. C.; Adams, P. D., Automating crystallographic structure solution and refinement of protein-ligand complexes. *Acta Crystallographica Section D* **2014,** *70* (1), 144-154.

66. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **1997,** *53* (3), 240-255.

67. Schneider, T. R.; Sheldrick, G. M., Substructure solution with SHELXD. *Acta Crystallographica Section D Biological Crystallography* **2002,** *58* (10), 1772-1779.

68. Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C.,

MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **2010,** *66* (1), 12-21.

69.   Laakso, L. M.; Holm, L., Dali server update. *Nucleic Acids Research* **2016,** *44* (W1), W351-W355.

70.   Simpson, P. J.; Xie, H.; Bolam, D. N.; Gilbert, H. J.; Williamson, M. P., The Structural Basis for the Ligand Specificity of Family 2 Carbohydrate-binding Modules. **2000,** *275* (52), 41137-41142.

71.   von Schantz, L.; Håkansson, M.; Logan, D. T.; Walse, B.; Österlin, J.; Nordberg-Karlsson, E.; Ohlin, M., Structural basis for carbohydrate-binding specificity—A comparative assessment of two engineered carbohydrate-binding modules. *Glycobiology* **2012,** *22* (7), 948-961.

72.   Davies, G.; Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure* **1995,** *3* (9), 853-859.

73.   KOSHLAND Jr., D. E., Stereochemistry and The Mechanism of Enzymatic Reactions. **1953,** *28* (4), 416-436.

74.   Plattner, M.; Shneider, M. M.; Arbatsky, N. P.; Shashkov, A. S.; Nazarov, S.; Prokhorov, N. S.; Taylor, N. M. I.; Knirel, Y. A.; Leiman, P. G., Structure and function of the branched receptor-binding complex of bacteriophage CBA120. University of Texas Medical Branch, Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, 301 University Blvd, Galveston, TX 77555-0647: 2019.

75.   Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W., iMOSFLM: a new graphical interface for diffraction-image processing with

MOSFLM. *Acta Crystallographica. Section D, Biological Crystallography* **2011,** *67* (Pt 4), 271-281.

76.   Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography* **2006,** *62* (8), 859-866.

77.   Sheldrick, G. M., Macromolecular phasing with SHELXE. *Zeitschrift für Kristallographie - Crystalline Materials* **2002,** *217* (12).

78.   Terwilliger, T. C.; Adams, P. D.; Read, R. J.; McCoy, A. J.; Moriarty, N. W.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Zwart, P. H.; Hung, L.-W., Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D Biological Crystallography* **2009,** *65* (6), 582-601.

79.   Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX : building new software for automated crystallographic structure determination. *Acta Crystallographica Section D Biological Crystallography* **2002,** *58* (11), 1948-1954.

80.   Müller, J. J.; Barbirz, S.; Heinle, K.; Freiberg, A.; Seckler, R.; Heinemann, U., An Intersubunit Active Site between Supercoiled Parallel β Helices in the Trimeric Tailspike Endorhamnosidase of Shigella flexneri Phage Sf6. *Structure* **2008,** *16* (5), 766-775.

81.   Mayans, O.; Scott, M.; Connerton, I.; Gravesen, T.; Benen, J.; Visser, J.; Pickersgill, R.; Jenkins, J., Two crystal structures of pectin lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. *Structure* **1997,** *5* (5), 677-89.

82.   Beck, T.; Krasauskas, A.; Gruene, T.; Sheldrick, G. M., A magic triangle for experimental phasing of macromolecules. *Acta Crystallographica Section D* **2008,** *64* (11), 1179-1182.

83.   Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A., Comparative Protein Structure Modeling Using Modeller. **2006,** *15* (1), 5.6.1-5.6.30.

84.   Zhang, Y. J. B. B., I-TASSER server for protein 3D structure prediction. **2008,** *9* (1), 40.

85.   Seul, A.; Müller, J. J.; Andres, D.; Stettner, E.; Heinemann, U.; Seckler, R., Bacteriophage P22 tailspike: structure of the complete protein and function of the interdomain linker. *Acta Crystallographica Section D Biological Crystallography* **2014,** *70* (5), 1336-1345.

86.   Israel, V., A Model for the Adsorption of Phage P22 to Salmonella typhimurium. *Journal of General Virology* **1978,** *40* (3), 669-673.

87.   Kreisberg, J. F.; Betts, S. D.; King, J., β-Helix core packing within the triple-stranded oligomerization domain of the P22 tailspike. *Protein Science* **2000,** *9* (12), 2338-2343.

88.   Benton, C. B.; King, J.; Clark, P. L., Characterization of the Protrimer

        Intermediate in the Folding Pathway of the Interdigitated β-Helix Tailspike

        Protein. *Biochemistry* **2002,** *41* (16), 5093-5103.

89.   Eriksson, U.; Svenson, S. B.; Lönngren, J.; Lindberg, A. A., Salmonella Phage

        Glycanases: Substrate Specificity of the Phage P22 Endo-rhamnosidase. **1979,** *43*

        (3), 503-511.

90.   Xiang, Y.; Leiman, P. G.; Li, L.; Grimes, S.; Anderson, D. L.; Rossmann, M.

        G., Crystallographic Insights into the Autocatalytic Assembly Mechanism of a

        Bacteriophage Tail Spike. *Molecular Cell* **2009,** *34* (3), 375-386.

91.   Ho, T. D.; Waldor, M. K., Enterohemorrhagic Escherichia coli O157:H7 gal

        Mutants Are Sensitive to Bacteriophage P1 and Defective in Intestinal

        Colonization. *Infection and Immunity* **2007,** *75* (4), 1661.

92.   Vollmer, T.; Hinse, D.; Kleesiek, K.; Dreier, J., The Pan Genera Detection

        Immunoassay: a Novel Point-of-Issue Method for Detection of Bacterial

        Contamination in Platelet Concentrates. *Journal of Clinical Microbiology* **2010,**

        48 (10), 3475.