STRUCTURAL STUDIES OF THE PHAGE G CAPSID AND HELICAL TAIL SHEATH USING CRYO-EM

by

Brenda Gonzalez

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biological Sciences

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Leifu Chang
Department of Biological Sciences

Dr. Angeline Lyon
Department of Chemistry

Dr. Wen Jiang, Chair
Department of Biological Sciences

Dr. David Thompson
Department of Chemistry

Dr. Thomas Walter
Department of Biological Sciences

Approved by:
Dr. Janice P. Evans

Dedicated to my family and loved ones for their support

ACKNOWLEDGMENTS

I would like to acknowledge the various mentors, friends, and family that have supported me, and helped me achieve my goal of becoming a first generation PhD. I would like to thank Dr. Laszlo Csonka for accepting me in his lab as a high school student, which was a significant opportunity in my education that piqued my interest in academia. From then, I have met various mentors throughout my undergraduate studies and beyond that I have been able to learn from.

I would like to thank my research advisor, Dr. Wen Jiang, committee members, Dr. Leifu Chang, Dr. David Thompson, Dr. Tom Walter, and Dr. Angeline Lyon, and my lab mates for their patience and support through my graduate studies. The guidance and advice I received from these individuals about my project was critical in my success. I would also like to recognize all the professors that have helped guide me through various applications and courses that have helped build my foundation in science. My colleagues and friends in the Department of Biological Sciences were a major source of support and crucial in my success in graduate school. I would like to especially thank Frank Vago, Lyman Monroe, and Dr. Corey Moore for their help in designing and troubleshooting experiments, discussing ideas, writing and editing, and analyzing data.

During my time in graduate school, I received financial and moral support from the Purdue Office of Graduate Diversity Initiatives and SACNAS. I would like to thank the colleagues that participated in these programs with me and faculty especially Dr. Ignacio Camarillo, Dr. Morris Levy, and Dr. Maria Levy for their support and encouragement. Finally, I would like to thank my family for supporting me and believing in my goals throughout these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Phages, viruses that infect bacteria, have been used for many studies in understanding fundamentals of molecular biology and taking advantage of their natural antimicrobial properties (Harper 2021). They are often noted for their overwhelming abundance and are recognized as the most abundant biological entities in the world (Harper 2021). The field has grown since the early 20th century, and now, there are several classes of phages that have been observed and characterized (Ackermann 2009). Within this abundant class of biological organisms, the order called *Caudovirales*, is the most populated group of phages to date (Harper 2021). In this order of viruses, the dsDNA genome phages have 2 main components, the icosahedral capsid, and a tail (Harper 2021). Though many tailed phages have been studied for many decades, new information about phages is still being found. Important findings such as the CRISPR gene editing tool adapted from phages in 2007 (Barrangou, Fremaux et al. 2007) have contributed to new biotechnology that impacts human health. For this reason, studies on phages have proven to be valuable in understanding fundamental biological questions and advancing basic research.

In this dissertation, we investigated phage G, which has the largest capsid and genome of propagated phage studied to date (Donelli 1968, Sun and Serwer 1997, Pope 2011, Hua, Huet et al. 2017). By studying phage G, we may add to the knowledge of this relatively unexplored group of Jumbo phages with remarkably larger genomes (>200kbp) (Yuan and Gao 2017)to understand how their structure and function may be similar or different to the commonly studied, smaller bacteriophages, such as T4 and λ. For a majority of these studies, we outline how our structural biology insights of phage G using cryo-EM (cryo-Electron Microscopy) have shown it's icosahedral capsid of ~ 180 nm in diameter at the 5-fold icosahedral vertex is composed of hexamer and pentamer proteins similar to what's been discovered in other, smaller tailed phages (González, Monroe et al. 2020). Our observations from microscopy data also show unique mechanistic properties in phage G's tail that are inconsistent with current model of tail contraction within the *Myoviridae* family of tailed phages. Data suggest phage G's structure and organization of its helical tail are still similar to contractile phages such as T4 (Amos and Klug 1975, Abuladze, Gingery et al. 1994) and phi812 (Nováček, Šiborová et al. 2016), however, the mechanism of the tail sheath movement is inconsistent with the existing ideas of myophage function (Harper 2021).

CHAPTER 1. OVERVIEW OF CAUDOVIRALES

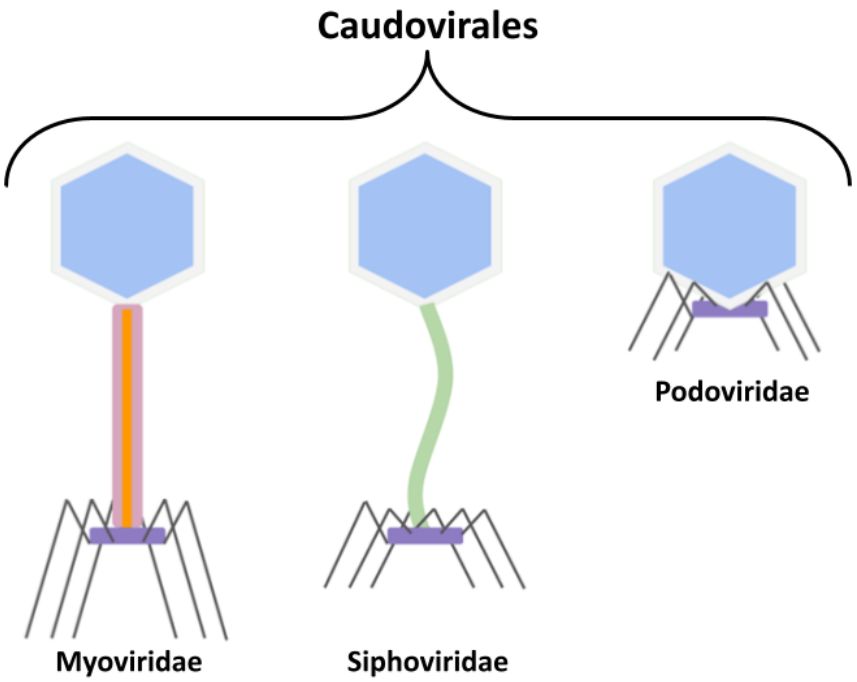


Figure 1.1. The *Caudovirales* order of tailed phages. (Left) *Myoviridae* family of dsDNA tailed phages that is recognizable by its contractile sheath tail. This is the prototypical depiction of a bacteriophage. (Middle) *Siphoviridae* family of tailed phages that lacks tail sheath proteins and is non-contractile. It is also generally more flexible. (Right) *Podoviridae* family which barely has a recognizable tail.

1.1 The three families of tailed phages

Caudovirales are the most identified group of viruses studied (Harper 2021). They have three common features: (i) an icosahedral or prolate protein capsid, (ii) a tail connected to one of the vertices, and (iii) a dsDNA genome. There are three families within the order of tailed *Caudovirales*, and they are distinguished by their tail morphology (Figure 1.1). All three families reported have differential characteristics in the tail structure. *Podoviridae* have an extensile tail that use capsid-packaged proteins that are extended and used for host penetration and genome ejection (Hrebík, Štveráková et al. 2019). *Podoviridae* are 15% of identified *Caudovirales* (Ackermann and Prangishvili 2012). *Siphoviridae*, which comprise more than half of all identified and characterized *Caudovirales,* possess noncontractile tails. Current ideas of *Siphoviridae* describe proteins at the end of its tail that contact specific complexes in the cell wall to facilitate

dsDNA ejection (Harper 2021). *Myoviridae* are described to have contractile tail sheath that drive tail tube movement to help penetrate the host cell (Hu, Margolin et al. 2015). About 25% of known *Caudovirales* are *Myoviridae* (Harper 2021).

1.2 Caudovirales Capsid Structure

Head Assembly

The assembly of phages within the *Caudovirales* order includes the presence of large protein multimers of varying symmetries and unique genome packaging mechanisms. Assembly of the capsid begins at the portal protein, which is a dodecamer. The scaffolding proteins assemble onto the connector (portal) protein, and the capsid proteins assemble around this scaffold (Casjens and King 1975) to form a capsid precursor. Once the capsid proteins have come together, the terminase subunits bind to the portal and form a packaging motor for the dsDNA. The motor transfers the dsDNA into the capsid, which helps with the expansion of the capsid and maturation (Conway, Duda et al. 1995, Fang, Wright et al. 2008, Guo, Liu et al. 2014). After spooling of the dsDNA, and/or during the DNA packing stage, scaffolding proteins used for the initial assembly of the capsid, are digested by host aminopeptidases, hydrolases, etc., (Suhanovsky and Teschke 2015). When the capsid is full (translated by the terminase signal), then the DNA is cleaved, portal shuts, the terminase dissociates, and the connector complex binds to the portal (Conway, Duda et al. 1995, Guo, Liu et al. 2014, Harper 2021). The dsDNA packed, mature capsid is attached to the completed tail structure to finish assembly (Harper 2021).

T4 Capsid Assembly

The capsid assembly of T4 has been more extensively studied and used to understand myophage formation. In the first stage of capsid assembly of T4, the dodecamer portal assembles on the cytoplasmic membrane of the host, with the aid of the gp40. The prohead core proteins, gp21, gp22, gp67, and gp68; the initiation proteins, IPI, IPII, and IPIII; and gpalt form what is to be an internal scaffold, on the portal complex (Fokine and Rossmann 2014). The main two capsid structural components, gp23 and gp24 assemble around the core in hexamer and pentameric symmetries, respectively, creating a procapsid. The maturation process in structural assembly includes the function of a prohead protease, gp21, that cleaves the N-terminus residues of gp23,

gp24, gp67 and gp68 and digests gp21, g22, the internal proteins, a gpalt into small fragments (Fokine and Rossmann 2014).

The function of capsid protein cleavage during maturation is mostly to trigger a rearrangement to flatten capsid walls and increase internal capsid volume (Conway, Duda et al. 1995, Fokine and Rossmann 2014, Harper 2021). In T4, at this stage, the capsid leaves the host membrane and a complex of terminase proteins and dsDNA bind to the portal protein to begin the genome-packaging. The terminase is made up 2 components, the small and large terminase subunits. During the packaging, free energy derived from ATP hydrolysis is thought to drive the terminase motor to mechanically push the DNA into the capsid (Yap and Rossmann 2014).

DNA packaging in *Bacillus subtilis* phage ϕ29 works differently in that it lacks a small terminase subunit, and instead uses a 174-nucleotide pRNA that helps TerL to package the genome. In ϕ29, it has been described that TerL, the terminase protein, has ATPase, endonuclease, and translocase activity, and oligomerizes into a pentamer at the portal vertex for packaging. In ϕ29, 4 of the 5 subunits in the motor sequentially hydrolyze ATP and can move 2.5 bp per hydrolysis of each subunit. The internal capsid pressure is a signal for termination and motor ejection (Rao and Feiss 2015).

The same signal for packaging termination is also used for T4 and Mu. However, P2 detects a conserved sequence at the beginning of a new genome copy which triggers termination. In both mechanisms, the large terminase subunit cuts the dsDNA to terminate packaging. The neck proteins associate with the portal after packaging is terminated. In T4, gp13 and gp14 serve as adapters and bind to the dodecamer portal proteins to the tail. In T4, there are fibritin proteins that decorate the neck – making up the collar and whisker. In T4, the full capsid is joined to the assembled tail, to then complete assembly (Yap and Rossmann 2014).

1.3 Capsid organization

The basic building blocks of *Caudoviral* capsids are relatively well conserved in architecture, however, the sequence similarity may be low (Suhanovsky and Teschke 2015). In other cases, some have cement proteins, also called decoration proteins for added stability (Prevelige Jr 2008,

Vernhes, Renouard et al. 2017, González, Monroe et al. 2020). *Caudoviral* capsid size also varies from phage G, the largest isolated and propagated capsid to date at 180 nm diameter measured from 5-fold to 5-fold vertex (Ageno, Donelli et al. 1973, Hua, Huet et al. 2017, González, Monroe et al. 2020) versus the smallest of 14 kb *Rhodococcus* phage RRH1 that encodes just 20 genes (Petrovski, Dyson et al. 2012). Furthermore, there is also variation in the icosahedral structure of *Caudoviral* capsids in whether they are elongated (prolate) such as ϕ29 (Morais, Choi et al. 2005) and T4 (Fokine, Chipman et al. 2004).

Capsid proteins

The most recognized feature in the capsid proteins of *Caudovirales* is the fundamental HK97 fold (Wikoff, Liljas et al. 2000). The capsid proteins are then assembled into hexamers and pentamers, in some cases the pentamer proteins are formed by a distinct protein as in the case with T4 with gp23 and gp24 (Fokine, Chipman et al. 2004), that associate into an icosahedron that can be described by T number symmetry.

The concept of T number was first described in the Caspar-Klug Theory described in their 1962 publication on "Physical principles in the construction of regular viruses" (Caspar and Klug 1962). The idea is based on the organization of a viral icosahedron being composed of 60 subunits that form 20 triangles that make the facets of the icosahedron (Caspar and Klug 1962). There are 3 different points of symmetry in the icosahedron: 5-fold, 2 -fold, and 3-fold. At the 12 vertices of an icosahedron, there is 5-fold symmetry, and at the edges of the triangle facets, there is 2-fold symmetry. In the very center of each triangular facet of the icosahedron, there is 3-fold symmetry. In Caspar & Klug's report, the T number (also called the triangulation number) is defined by the squared length of each triangular edge that makes up the facets in the icosahedron.

A simplified version of the T number from the Viral Zone website can be described as the number of hexamers (h) that can be traced in a straight line from one pentamer (k) to the next (h and k), which can then be applied to the Triangulation number equation: $T=h^2+hk+k^2$ (Caspar and Klug 1962, Hulo, De Castro et al. 2011). Typical triangulation numbers observed icosahedron viruses range from the T=3 prolate capsid of *Bacillus* phage phi29 (450 Å wide and 540 Å long) (Tao, Olson et al. 1998, Xu, Wang et al. 2019), to the largest capsid structure reported to date of phage

G at T=52 and 180 nm wide at the furthest pentamer-to-pentamer distance (Donelli 1968, Hua, Huet et al. 2017, González, Monroe et al. 2020).

A common phage building block

Wikoff, *et al.* reported the HK97 capsid protein structure using x-ray crystallography, which demonstrated the canonical *Caudoviral* major capsid protein architecture shared by tailed phages at 3.6 Å resolution (Wikoff, Liljas et al. 2000). Despite the low sequence identity observed among capsid proteins of *Caudovirales*, there is a shared protein fold architecture that has been observed ubiquitously among tailed phages (Suhanovsky and Teschke 2015). Curiously, this pattern of capsid protein fold is also reported in the coat protein of eukaryotic viruses, such as *Herpesviridae* (Baker, Jiang et al. 2005, Yuan, Wang et al. 2018). They are a family of viruses with linear DNA genomes and a T=16 icosahedral symmetry protein core surrounded by tegument, an envelope, and envelope proteins, further demonstrating the utility of the structural fold among different viral capsid symmetries. Additionally, naturally occurring biological structures, such as encapsulin nanocompartments – icosahedral protein shell delivery vehicles, which are natural to bacteria – share the common HK97 capsid protein fold observed in tailed phages (Sutter, Boehringer et al. 2008).

1.4 Tail organization

Podoviridae

The tails in *Podoviridae* are the shortest of all the three families of tailed phages and are non-contractile. Tail fibers and spikes are common in *Podoviridae*, such as T7, a well-studied *E. coli* phage with a tail of about 30 nm long (Cuervo, Pulido-Cid et al. 2013). Tail assembly in T7 does not happen until the capsid assembly and dsDNA packaging is completed. From then, the portal protein, which is a 12 subunit complex of gp8 in T7, gp11, and gp12, and fibers come together to form the noncontractile tail complex.

In the Podoviridae, *Salmonella* phage P22, the main component organization of its tail is like T7. In P22, the trimeric tail fibers of gp9 have host attachment properties. As in the case of other Podoviridae such as phi29, P22 has host lipopolysaccharide digestion, which has been proposed to be a strategy in helping with host adhesion and infection progression (Harper 2021).

Siphoviridae

In both families of tailed phages with long tails, *Myoviridae* and *Siphoviridae,* the assembly of the tail begins at the distal tip, towards the baseplate. *Siphoviridae* are characterized by their inability to contract. A common example used to illustrate this process can be explained with phage Lambda, which is a common coliphage that has been studied for a relatively long time. The steps described below are mostly from studies done by Katsura (1990), Davidson, Cardarelli et al. (2012).

The Lambda tail assembly starts with two simultaneous events. First, the tail tube made of gpH will be covered in tail chaperones of gpG and gpGT. This assembly will then be joined to the baseplate structure, which is a trimer of gpJ, and subunits of gpI, gpL, and gpK for phage Lambda. The baseplate, tail tube measure proteins, and the tail chaperone proteins form the preliminary tail structure. The next step involves the replacement of the tail chaperone proteins by gpV, Lambda's major tail protein, and following this, gpU caps off the structure. Finishing protein add-ons such as gpZ, the packaged Lambda capsid, and the tail fibers are then fixed onto the phage tail to finish assembly (Katsura 1990).

Myoviridae

Out of the three families of tailed phages, the *Myoviridae* have the only contractile tails. They are marked by their unique ability to rearrange their tail sheath proteins as a process during infection to facilitate genome ejection through unclear mechanisms (Harper 2021). A good example to learn about tailed phage assembly and function is the *E. coli* phage Mu (Büttner, Wu et al. 2016). Below, I explain more on Büttner's et al.'s findings of the sequence of Mu tail assembly. (Büttner, Wu et al. 2016)

The first step in beginning the tail assembly in Mu is the union of the proteins that make up the baseplate, which is generally larger, and more complex in *Myoviridae*. In Mu, a trimer of gp27, a trimer of gp5, and gp5.4 join as a starting point for the baseplate. From there, hexamers of Mup46, gp25, Mup47, Mup48, gp6, and gp7 complexes assemble around that structure to form the final baseplate foundation. From there, the preassembled tail tube and tape measure complex is then joined to the baseplate, that can then recruit the tail sheath proteins. After the 6 stranded, right-handed helix organization of tail sheath proteins is assembled onto the tube proteins, the tail

structure is joined to the dsDNA packaged, icosahedral capsid. Finally, proteins for the tail fibers are joined to the baseplate for the finishing touches.

Tail sheath contraction in Myoviridae with T4 as an example

T4 is arguably one of the most well studied *Myoviridae* model systems in phage biology. Numerous reports have been produced from the Rossmann lab at Purdue reporting the tail structure, morphology, and contraction mechanism of the T4 tail. In T4, the tail is assembled in an independent pathway from the prolate, icosahedral dsDNA, packaged capsid. As mentioned in the previous section, the tail assembly is initiated by the baseplate formation (Leiman, Kanamaru et al. 2003, Harper 2021). In the case of T4, a trimer of gp11 is first put together, which then serves as a starting point for the baseplate assembly (Leiman, Kanamaru et al. 2003). The baseplate is then attached to the tail tube, which can then fully assemble into the tail sheath decorated structure, completing the contractile tail machine of T4.

In T4, the mechanism of contraction has been proposed to begin at the long tail fibers (LTF) (Leiman, Kanamaru et al. 2003), which can then help anchor the short tail fibers (STF) onto the fuzzy, lipopolysaccharide-coated surface of the *E. coli* host. The tail fiber binding is then thought to induce a "hexagon-to-star" conformational change in the baseplate and an irreversible contraction in the tail sheath (Crowther, Lenk et al. 1977). Researchers have suggested that assembly of the tail sheath proteins, gp18, onto the gp19 tail tube structure is done in a high energy conformation (Leiman, Kanamaru et al. 2003) compared to a stretched out spring. Additionally, studies have found that gp25, a baseplate component of T4, is structurally alike to the sheath protein, gp18, and is proposed to initiate the tail contraction by altering the sheath protein organization, causing contraction in a domino effect. Reports have also shown that the contraction of the sheath proteins is a rigid body movement (Aksyuk, Leiman et al. 2009) that begins at the distal end of the tail and is propagated upwards towards the capsid in a wave-like motion (Moody 1973). However, it should be noted that contraction does not mean DNA release. Previous studies observed that T4 tail contraction could be artificially induced by urea (Leiman, Kanamaru et al. 2003).

Table 1.1. T4 tail genesis proteins adapted from (Leiman, Kanamaru et al. 2003)

Tail protein	Mass (kDa)	Copy Numbers	Location	Function
gp3	19.7	6	Added right before tail terminator, gp15	Maintain correct tail tube length (Vianelli, Wang et al. 2000)
gp53	23.0	6	At the inner portion of the baseplate interacting with the tail tube gp6-gp25	Proposed to help propagate the conformational change from host binding to the tail sheath to initiate contraction (Leiman, Chipman et al. 2004).
gp5	63.7	3	At the tip	lysozyme to help dissolve host surface for tube penetration (Leiman, Arisaka et al. 2010)
gp6	74.4	12	At the inner portion of the baseplate interacting with the tail tube in complex with gp25-gp53	Proposed to help propagate the conformational change from host binding to the tail sheath to initiate contraction (Leiman, Chipman et al. 2004).
gp7	119.2	6	Part of the baseplate assembly in contact with gp10	Transfers the host-binding signal received from gp10, another baseplate component, which eventually gets transferred to the tail sheath to initiate contraction (Leiman, Arisaka et al. 2010)
gp8	38.0	12	Intermediate baseplate protein	Involved in the hexagonal-to-star conformational change in the baseplate that happens during host attachment (Leiman, Arisaka et al. 2010)
gp9	31.0	18	Connects long tail fibers to baseplate	Helps trigger conformational change in baseplate upon tail fiber binding to host (Leiman, Arisaka et al. 2010)
gp10	66.2	18	Connects short tail fibers to baseplate	Extends the short tail fibers toward host during adsorption (Leiman, Arisaka et al. 2010)
gp11	23.7	18	Associated to short tail fibers in the inactive form	Rotates about 100° and makes contacts with the long tail fiber upon host attachment, helping trigger an overall baseplate conformational change (Leiman, Arisaka et al. 2010)
gp12	55.3	18	Short tail fibers	They are extended after the long tail fibers attach the host surface to bring the tail tube in contact with the membrane (Leiman, Arisaka et al. 2010)
gp15	31.4	6	Neck region	Thought to play a role in attachment of the packaged capsid to the tail (King 1968)
gp18	71.2	144	Tail sheath	Irreversibly rearranges from extended to contracted during infection, and is proposed to help protrude the inner tail tube to help penetrate through host membrane (De Rosier and Klug 1968, Kanamaru, Leiman et al. 2002, Leiman, Kanamaru et al. 2003)
gp19	18.5	144	Tail tube	Scaffold for tail sheath assembly and has 40 Å-diameter channel where dsDNA genome is transferred during infection (Leiman, Kanamaru et al. 2003)
gp25	15.1	6	Baseplate component	Proposed to initiate tail sheath assembly in the extended conformation, which is thought to be a high energy state. Also, it has been proposed to initiate tail contraction following host adhesion (Taylor, Prokhorov et al. 2016)
gp27	44.4	3	Central spike component at tip	Is in complex with gp5 and gp5.4 to make a complex for the spike at the tail tip (Taylor, Prokhorov et al. 2016)
gp29	64.4	6	Baseplate component	Controls the tail tube, gp19, assembly length by an unclear mechanism. Current theories include gp29's unfolded state influences the incorporation of gp19 as the tail tube assembly extends (Leiman, Arisaka et al. 2010)
gp48	39.7	6	Baseplate component	Associates with gp54 to serve as a base for the tail-tube junction and also a starting point for tail spike formation (Taylor, Prokhorov et al. 2016)
gp54	35.0	6	Baseplate component	Binds to gp48 to serve as a base for tail tube assembly and also is a base for tail spike assembly (Taylor, Prokhorov et al. 2016)

CHAPTER 2. PHAGE G CAPSID STUDIES BY CRYO-EM

This chapter is from a 2020 publication in collaboration with other scientists across different institutions (González, Monroe et al. 2020). Their contributions are noted and are outlined in the journal website.

Phage G has the largest capsid and genome of any known propagated phage. Many aspects of its structure, assembly, and replication have not been elucidated. Herein, we present the dsDNA-packed and empty phage G capsid at 6.1 and 9 Å resolution, respectively, using cryo-EM for structure determination and mass spectrometry for protein identification. The major capsid protein, gp27, is identified and found to share the HK97-fold universally conserved in all previously solved dsDNA phages. Trimers of the decoration protein, gp26, sit on the 3-fold axes and are thought to enhance the interactions of the hexameric capsomeres of gp27, for other phages encoding decoration proteins. Phage G's decoration protein is longer than what has been reported in other phages, and we suspect the extra interaction surface area helps stabilize the capsid. We identified several additional capsid proteins, including a candidate for the prohead protease responsible for processing gp27. Furthermore, cryo-EM reveals a range of partially full, condensed DNA densities that appear to have no contact with capsid shell. Three analyses confirm that the phage G host is a *Lysinibacillus,* and not *Bacillus megaterium*: identity of host proteins in our mass spectrometry analyses, genome sequence of the phage G host, and host range of phage G.

2.1 Introduction

Phage G is a bacteriophage with a remarkably large capsid (~180 nm in diameter at the 5-fold vertex) and packaged genome length (626 kbp) (Donelli 1968, Ageno, Donelli et al. 1973, Donelli, Dore et al. 1975, Donelli, Griso et al. 1976, Hua, Huet et al. 2017) compared to the median genome size of 52 kbp derived from available phage sequences (Al-Shayeb, Sachdeva et al. 2020). Phage G was originally isolated and characterized using a host bacterium classified as Bacillus megaterium (Donelli 1968). Since its discovery in 1968, structural studies of this phage have had limited resolution (Ageno, Donelli et al. 1973, Donelli, Dore et al. 1975, Donelli, Griso et al. 1976, Sun and Serwer 1997, Hua, Huet et al. 2017), eventually achieving cryo-EM resolution of 15 Å

(Hua, Huet et al. 2017). The reported limitation to resolution is the large size (Hua, Huet et al. 2017) and challenges in culturing and purifying the intact virion.

All tailed dsDNA phages share several structural features, including major capsid proteins (MCPs) that have a tertiary structure in common, the HK97 fold (Wikoff, Liljas et al. 2000, Suhanovsky and Teschke 2015). The HK97 fold is identifiable by its characteristic core domains, the N-terminal arm, E loop, A domain, and P domain (Wikoff, Liljas et al. 2000, Rader, Vlad et al. 2005). In addition to the MCP, many dsDNA, tailed phages encode decoration proteins located on the outer surface of the major capsid network (Sathaliyawala, Islam et al. 2010, Wang, Hardies et al. 2018, Newcomer, Schrad et al. 2019). These decoration proteins have been observed at different positions and in different oligomerization states, but are frequently observed as trimers (Sae-Ueng, Liu et al. 2014, Stone, Hilbert et al. 2018), as in phages Lambda and TW1 (Lander, Evilevitch et al. 2008, Wang, Hardies et al. 2018). The trimeric decoration proteins are thought to stabilize the 3-fold symmetry axes between three MCP hexamers, and therefore, the overall capsid organization (Wang, Hardies et al. 2018, Newcomer, Schrad et al. 2019), by various proposed mechanisms (Lander, Evilevitch et al. 2008, Sathaliyawala, Islam et al. 2010, Sae-Ueng, Liu et al. 2014, Stone, Hilbert et al. 2018, Wang, Hardies et al. 2018, Newcomer, Schrad et al. 2019).

The size of the G capsid raises many questions regarding (1) how such a large structure assembles, packages such a lengthy dsDNA genome, and maintains stability, and (2) what differences exist in relation to smaller ds DNA phages. In addressing these questions, we have characterized the phage G virion structure using cryo-EM and mass spectrometry. We have identified the MCPs and reconstructed empty capsid and dsDNA-packed virions to 9 and 6.1 Å resolution, respectively. Our reconstructions highlighted the lack of regular, concentric rings of packaged DNA observed in most phages (Serwer, Hayes et al. 1992, Chang, Weigele et al. 2006, Jiang, Chang et al. 2006, Lander, Evilevitch et al. 2008), suggesting a potentially novel phage G dsDNA packing mechanism. Additionally, our reconstruction shows the presence of a decoration capsid protein, gp26, which we propose is responsible for stabilizing the MCP, gp27, and hexamers. Consistent with this proposal, our SDS-PAGE results confirmed the phage G capsid is not stabilized by intermolecular covalent bonding interactions, as in the case with HK97 (Wikoff, Liljas et al.

2000).We found the first smaller-than-capsid, packaged DNA that appears to be detached from the inner capsid wall. We corrected the identity of the phage G host (PGH) to a *Lysinibacillus* species.

2.2 Results

2.2.1 Phage G empty and full capsid cryo-EM reconstruction

In our studies, we used single-particle cryo-EM (Cheng 2018) to image purified phage G particles, as shown in a representative micrograph (Figure 2.1). Data collection and reconstruction statistics are shown in Table 2.1. Phage particles in micrographs consist mostly of DNA-filled virions, with subpopulations of both empty phage particles and partially DNA-filled capsids. Our capsid structures of phage G's DNA-full (Figure 2.1, Supplementary Figure 2.S1) and empty (Supplementary Figure 2.S2) were reconstructed with jspr (Guo and Jiang 2014) to 6.1 and 9 Å resolution, respectively, using the "gold-standard" FSC criterion (Guo and Jiang 2014) (Supplementary Figure 2.S1). In total, 2564 and 243 particles were used to reconstruct the dsDNA full and empty capsid densities, respectively. More in-depth surface view depictions of both empty and full G capsid structures are outlined in Supplementary Figure 2.S2. The empty and full capsid reconstructions superimpose very well, and we therefore only further depict the higher-resolution full reconstruction in Figure 2.1B.

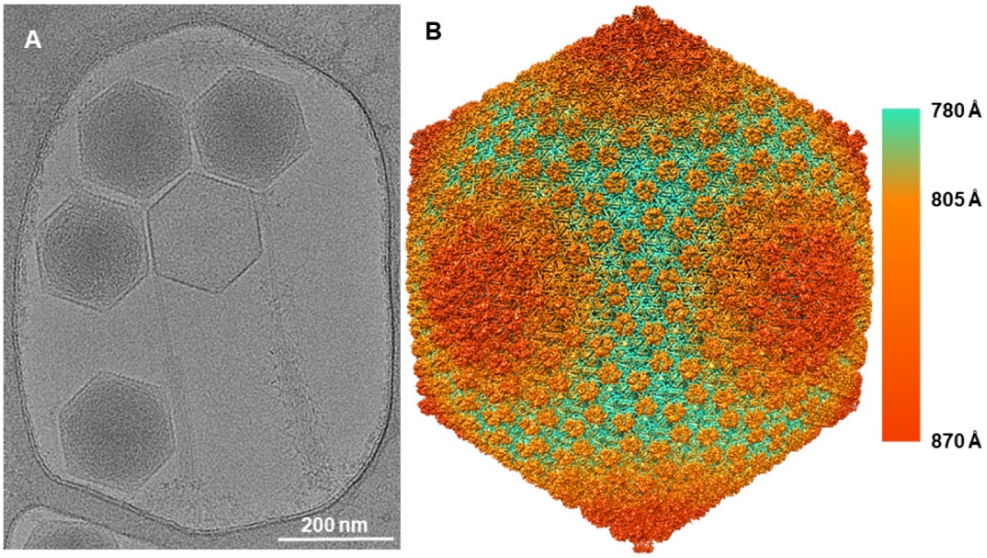


Figure 2.1. Phage G capsid cryo-EM structure. (a) A representative micrograph shows particles with various capsid states. (b) Phage G dsDNA-full capsid reconstruction at 6.1 Å resolution with 2564 particles. The structure displayed is surface colored radially using the color bar on the right.

Table 2.1. Phage G cryo-EM reconstruction and model statistics.

Parameters	Full capsid	Empty capsid
Nominal magnification	8700	
Voltage (KeV)	300	
Electron exposure (e –/Å^2)	14.5	
Defocus range (μm)	1–5	
Detector	K2	
Movies	375	
Beam size (μm)	3.35	
Symmetry imposed	Icosahedral	
Total particles for reconstruction	2564	243
Pixel size in reconstruction (Å)	1.742	2.900
Box size used for reconstruction (pixels)	1280	768
Map resolution (Å)	6.1	9.0
FSC threshold	0.143	
Image processing software used	jspr	
EMDB Accession	EMD-21695	EMD-21702

2.2.2 Phage G gp27 hexamer and gp26 trimer organization

Phage G has the largest bacteriophage capsid reported to date with an average diameter of ~ 180 nm along the 5-fold vertex (Hua, Huet et al. 2017). Its capsid is composed of gp27 hexamers and gp26 trimers organized into T = 52 *dextro* (h = 2, k = 6) icosahedral symmetry (Supplementary Figure 2.S3). The structure of gp27 (Figure 2.2B), the 282-amino-acid phage G MCP according to the phage G NCBI genome from 2018, was first modeled using Robetta (Kim, Chivian et al. 2004). This structure was then refined to the densities of gp27 in the virion 3D map using molecular dynamics flexible fitting (MDFF) method (McGreevy, Teo et al. 2016). The phage G gp27 model (Figure 2.2B) reported high similarity in structure to HK97's MCP, gp5 (Wikoff, Liljas et al. 2000). Despite the structural similarity, phage G's gp27 and HK97's mature gp5 have only 23% protein sequence identity, as determined by blastp alignment. Sequence alignment of phage G's gp27 and HK97's gp5 revealed 102 extra amino acids in the N terminus of HK97 gp5 (Conway, Duda et al. 1995). In HK97 gp5, the first 102 amino acids comprise the coiled-coil delta domain, which serves

as a scaffold during procapsid assembly, but is removed by cleavage during maturation (Conway, Duda et al. 1995, Oh, Moyer et al. 2014).

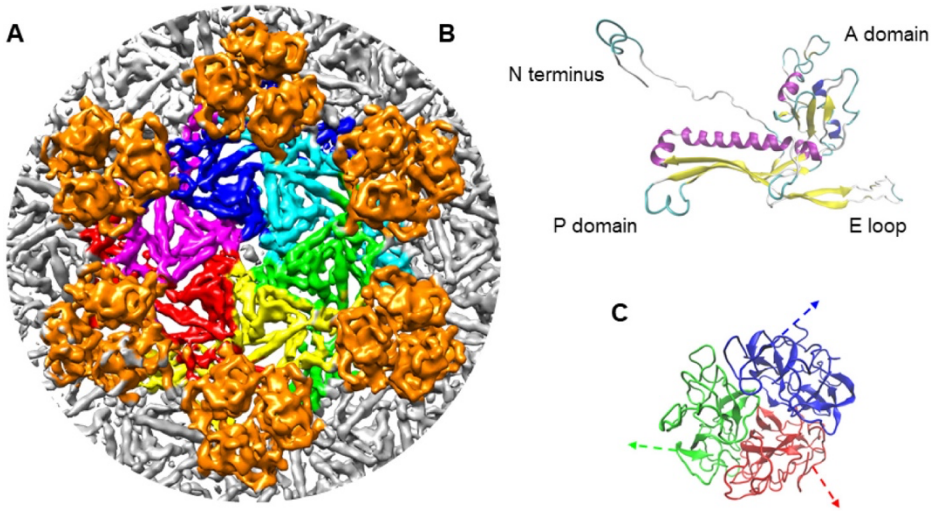


Figure 2.2. Phage G MCP hexamer and decoration protein trimer arrangement. (a) Phage G's MCP, gp27, and its decoration protein, gp26 arrangement. The structure shows gp27 hexamers and gp26 trimers positioned at the 3-fold axes around the hexamers. (b) An overview of phage G's gp27 homology model. All three domains of the structure are consistent with the domains described in HK97's MCP including the A domain, E loop, P domain. (c) The phage G decoration, gp26, protein oligomerizes into trimers. The first 15 amino acids were omitted in modeling because of their flexibility. Arrows indicate the direction the N terminus extends in the capsid density contacting neighboring gp27 subunits.

Phage G's capsid had a decoration protein, identified as gp26, organized into trimers on the outer surface of the capsid (Hua, Huet et al. 2017). The trimers were arranged around the 3-fold axes between three MCP hexamers (Figure 2.2C). The trimers are similar in structure and arrangement to what has been described for the decoration proteins of phages Lambda and TW1 (Lander, Evilevitch et al. 2008, Wang, Hardies et al. 2018). All three trimers contain a compact, beta sheet-rich core and an N-terminal extension. Phage G's gp26 is 55 amino acids longer than Lambda's gpD and 17 amino acids longer than gp56 of TW1 (Lander, Evilevitch et al. 2008, Wang, Hardies et al. 2018). Because of this limitation, we used TW1's gp56 (Wang, Hardies et al. 2018) as a reference for structure modeling of phage G's trimeric decoration protein, gp26, from the observed EM density (Figure 2A). Homology modeling, docking, and refinement of gp26 into our density

suggests a flexible N-terminal portion of approximately 15 amino acids, which was omitted from the model in Figure 2.2C.

The triangulation number (*T*-number) is represented by the following formula: $T = h^2 + hk + k^2$, where *h* and *k* represented the hexamer units counted toward the nearest pentamer vertex in a straight line (Caspar and Klug 1962).

In Lambda's 6.8 Å resolution capsid structure, a density bridges the N-terminal extension of the decoration protein, gpD, with the MCP E-loop region (Lander, Evilevitch et al. 2008), suggesting a stabilizing interaction between the two, and we observed a density in a similar position between phage G decoration protein gp26 and MCP gp27. The marine Siphoviridae, TW1, decoration protein structure and organization also closely resemble phage G's capsid structure (Wang, Hardies et al. 2018). All three decoration proteins have extended, flexible N-terminal segments that interact with MCP hexamers (Lander, Evilevitch et al. 2008, Prevelige Jr 2008, Wang, Hardies et al. 2018).

2.2.3 Lack of concentric DNA rings in the phage G virion structure

In all previously reported bacteriophage virion structures (Cerritelli, Cheng et al. 1997, Jiang, Chang et al. 2006, Lander, Evilevitch et al. 2008, Chen, Baker et al. 2011), a concentric ring pattern emerges from symmetrical reconstructions of the dsDNA genome in the capsid, even if less than half the phage DNA is packaged. However, we observed no concentric rings in our symmetric phage G capsid reconstruction (Figure 2.3A). Correspondingly, it is evident in the radial density plots that the signature equispaced intensities from phage DNA packaged in the capsid are absent in our analysis (Figure 2.3C). DNA-derived intensity, however, does exist in our virion reconstruction because the pixel intensity in the DNA-packed structure is higher than it is for the empty structure (Figure 2.3C), as expected.

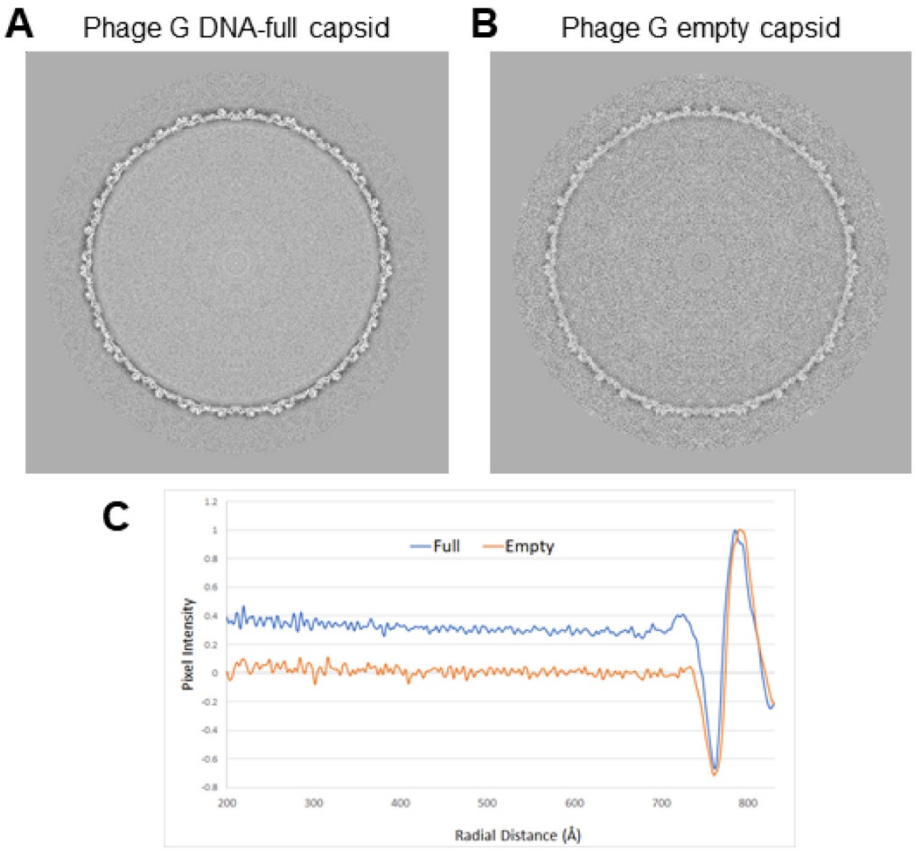


Figure 2.3. Lack of concentric dsDNA rings in the phage G DNA-full structure. (a) Middle volume slice of phage G's DNA-full capsid. (b) Middle volume slice of phage G's empty capsid. (c) Radial distance plots were generated for the dsDNA-full (blue) and empty (orange) phage G structures.

2.2.4 Partially full DNA condensates within the phage G capsid

To further explore the reason behind the lack of concentric rings in our virion reconstruction, we visually sorted the particles in our micrographs as summarized in Table 2.2. Among 3620 phage G capsid particles from our cryo-EM dataset, we sorted into four categories: DNA-full, partially full, capsid ruptured, and empty (no DNA visible); examples are in Figure 2.4. Overall, 88.6% of virus particles were classified as DNA-full, 8.0% empty, 2.3% DNA-partially full, and 1.0% ruptured with some condensed DNA retained. All particles categorized as partially full can be seen in Supplementary Figure 2.S4. All particles categorized as ruptured with associated condensed DNA are shown in Supplementary Figure 2.S5.

Table 2.2. Phage G capsid state counts (total count = 3620)

State	Full	Empty	Partially full	Ruptured
Count	3207	293	84	36
Percent	88.6%	8.0%	2.3%	1.0%

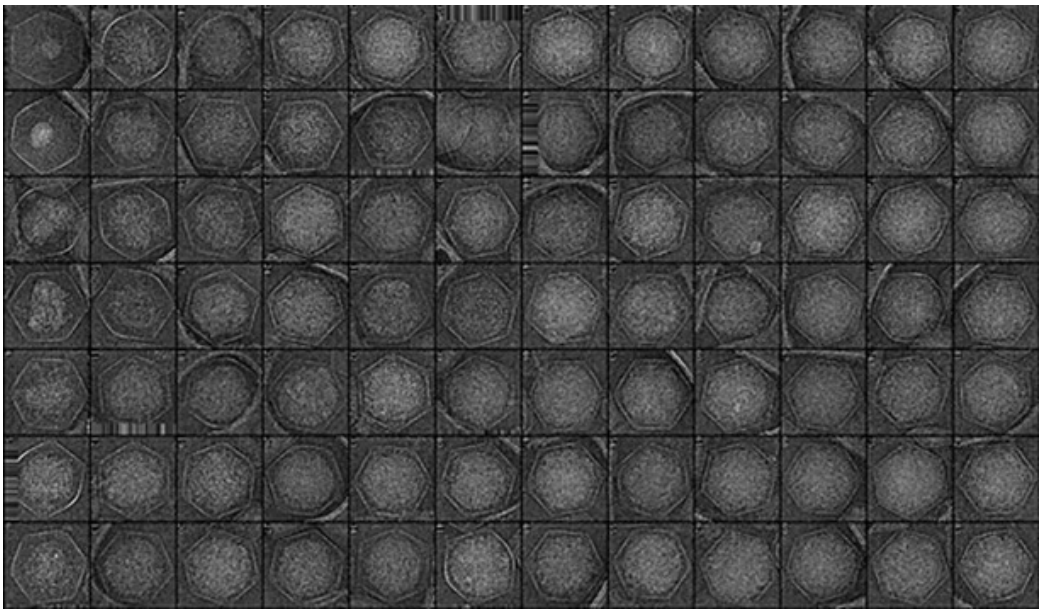


Figure 2.4. Partially full phage G particles gallery. In our cryo-EM micrographs, we observed particles that appeared to have DNA condensates that were seemingly detached from the capsid shell. The gallery above shows the particles that were used to quantify the apparent DNA condensate width (measured from the shortest dimension). Some of the particles in this gallery show toroid-like features. In total, 84 particles categorized as partially full are shown above. Striped edges in some of the particles are because of the micrograph edge.

We further characterized 84 DNA-partially full particles by measuring the apparent width of the condensed DNA in the smallest dimension (Figure 2.5). The values ranged from 495 to 1606 Å and have an average of 1341 Å, about 84% of the average diameter of the capsid (1590 Å from Figure 2.3C). A particle with ~ 816 Å DNA (Figure 2.5, leftmost inset) is particularly interesting as its shape clearly resembles a side view of a DNA toroid.

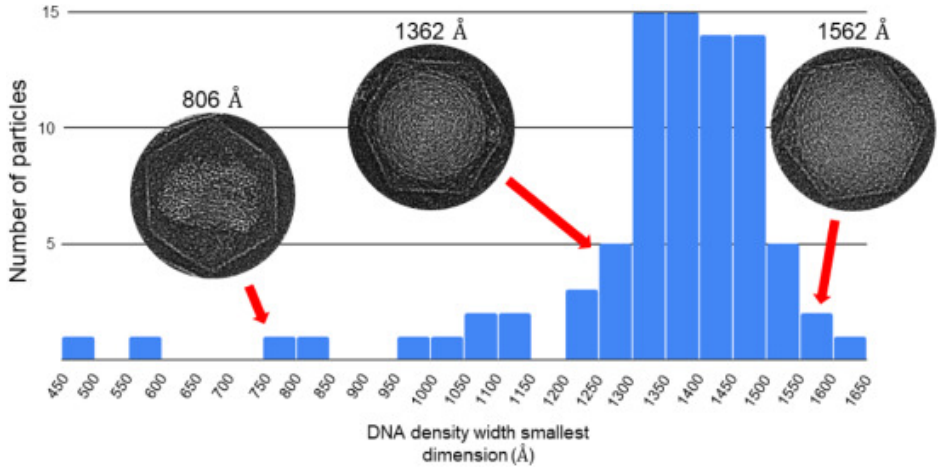


Figure 2.5. Histogram of partial DNA diameter. We further analyzed the group of 84 particles categorized as having partially full, compacted DNA density, by plotting a histogram of their DNA density width measured in the shortest dimension. The measured DNA density diameter values within the group of heterogeneous DNA density phage G capsids ranged from 495 to 1606 Å. Some particles appear to have toroid-like organizations from the side and top views as shown in the left two insets in this figure.

To visualize particles of this type in 3D, we used cryo-electron tomography on our phage G cryo-EM grid. From these data, we found one particle that appeared to have partially packaged DNA (Supplementary Figure 2.S6, bottom right), relative to neighboring full particles (Supplementary Figure 2.S6, top right). Also, it appeared to have a gap between the capsid shell and the DNA density, supporting our observations from the single-particle cryo-EM micrographs (Supplementary Figure 2.S4). The low abundance in our cryo-ET data collected reflected the low abundance of these partially full particles in our dataset (2.3%).

2.2.5 Identification of phage G proteins by mass spectrometry

Sixty-one phage G proteins were detected by mass spectrometry analyses with an FDR of 0% and a minimum number of two peptides (Table 2.3). An additional 22 phage G proteins were identified with a single peptide (the mass spectra of each were visually inspected prior to inclusion in Table 2.3). The number of individual proteins we detected in phage G is comparable to the numbers of different proteins identified in other large myovirus (Fokine, Chipman et al. 2004, Thomas, Rolando et al. 2008, Lecoutere, Ceyssens et al. 2009, Effantin, Hamasaki et al. 2013, Thomas, Quintana et al. 2016). The identified phage G proteins ranged in molecular weight from 5.7 kDa

for gp429, a protein of unknown function, to 313.2 kDa for gp183, the putative tape measure protein, i.e., the protein that determines the length of the tail (Mahony, Alqarni et al. 2016). The total number of mass spectra identified for the phage G proteins varied from 1 for 17 proteins to 427 for gp255 (Table 2.3). The range of mass spectral counts for the different proteins reflects in part, variable copy numbers. For instance, high numbers of total mass spectra were identified for gp26 and gp27, consistent with the structural determination that these proteins are present in high copy numbers in the head. Among the proteins for which low numbers of total mass spectra were identified (Table 2.3), were proteins that are candidates for lysins (gp23) and holins (gp24 and gp25). Identification of gps 23–25 was unexpected since lysis proteins are not present in the virion, suggesting they co-purified with the phage. Nonetheless, we believe the majority of phage G proteins identified by mass spectrometry are virion proteins because many non-virion proteins (e.g., the terminase protein, gp1, and DNA polymerase, gp329) were not detected.

The genes encoding mass spectrometry-identified capsid proteins are distributed throughout the phage G genome, highlighting the absence of major morphogenesis gene region, unlike many other tailed phages. The major implication is that for the many phage G virion genes whose functions have not been determined using sequence-based searches, their functions also cannot be hypothesized based on their gene locale. Nonetheless, we noted that the proteins known to be essential for the formation of a tailed phage icosahedral capsid all had their genes located within a relatively short region of the phage G genome between G_1 and G_27 (Figure 2.6). These genes included those for the large terminase subunit (gp1), portal protein (gp14), prohead protease (gp19, see below), head decoration protein (gp26), and MCP (gp27). These proteins were all detected by mass spectrometry (Table 2.3), with the exception of the large terminase protein, because of its transient role during morphogenesis(Casjens and King 1975).

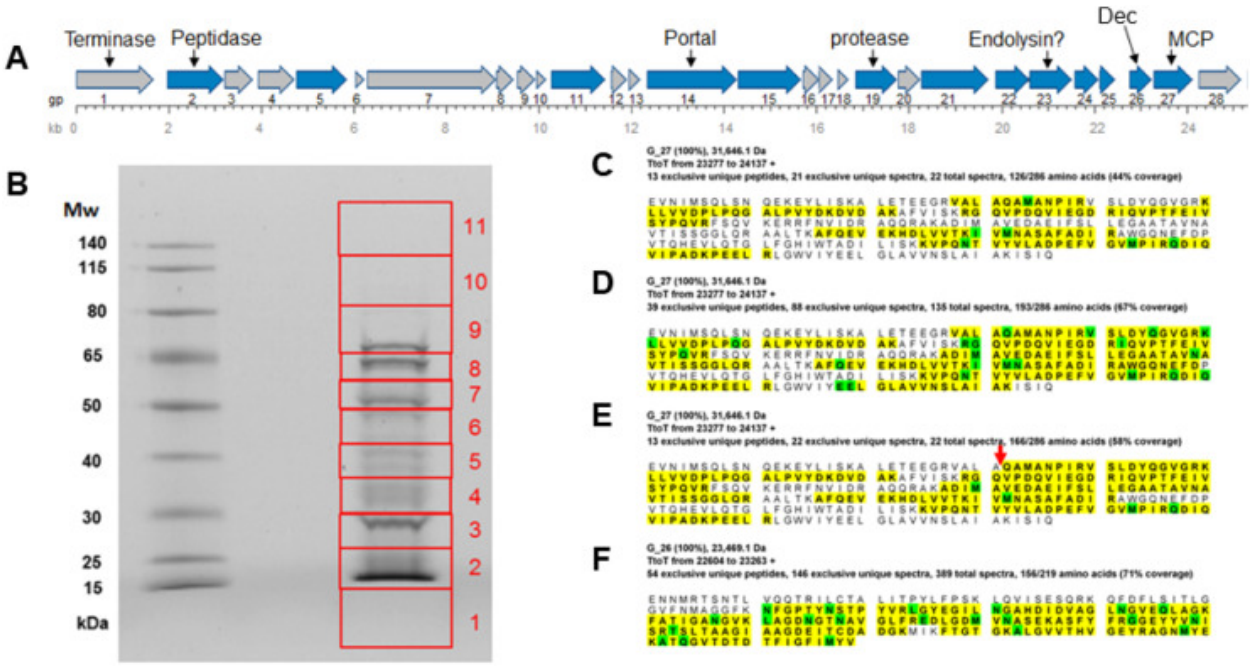


Figure 2.6. Identification of major Phage G head proteins by mass spectrometry. (a) Phage G genome region encoding the major head morphogenesis genes, including the terminase, portal, candidate prohead protease, MCP and head decoration protein (Dec). Genes whose products were identified by mass spectrometry are shaded blue. (b) SDS-PAGE gel of sucrose gradient purified Phage G virions. Individual gel slices that were subjected to mass spectrometric analyses are indicated. The mass spectral protein sequence coverage of the MCP gp27 from different gel slices are shown in (c) gel slice 4, (d) gel slice 3, and (e) gel slice 2. Red arrow indicates cleavage by the prohead protease, gp19. The mass spectral protein sequence coverage of the head decoration protein gp26 from gel slice 2 is shown in (f). Amino acids matched to a mass spectrum are shaded in yellow. Amino acids marked in green potentially have a post-translational modification (e.g., phosphorylation).

Table 2.3. Mass spectrometry analysis

Accession number[a]	MW (kDa)	Exclusive peptides	Total spectra	Sequence coverage (%)	Putative function/comment, best HHpred match (E-value)[b]
G_1	63.5		ND		Terminase large subunit, 3CPE_A DNA packaging protein gp17 T4 phage (1.4E − 33)
G_2	46.7	7	12	22.6	Predicted Zinc-dependent peptidase, 5EUF_B Protease (1.4E − 48)
G_3	25.2		ND		Transcriptional regulator, 4NQW_A ECF RNA polymerase sigma factor (7.1E − 22)
G_4	30.8		ND		
G_5	41.6	1	1	2.9	
G_6	8.1	3	4	29.9	
G_7	104.2		ND		
G_8	15.1		ND		
G_9	15.2		ND		
G_10	5.3		ND		
G_11	43.7	9	28	28.5	Candidate for prohead scaffold, coils prediction, pI 4.12
G_12	13.6		ND		
G_13	10.2		ND		
G_14	72.1	8	8	12.9	Portal, PF04860.12 (2.9E − 20)
G_15	53.6	3	4	7.1	Possible minor capsid protein, PF06152.11, Phage_min_cap2 (E-value: 0.18)
G_16	14.0		ND		
G_17	11.9	1	1	15.5	
G_18	9.4	1	2	10.5	
G_19	33.6	5	5	16.8	Putative prohead protease, PF04586.17; Peptidase_S78 (3.0E − 16)
G_20	18.9		ND		
G_21	55.4	6	11	17.0	Candidate for prohead scaffold, coils prediction, p*I* 4.83
G_22	26.9	4	4	17.6	Candidate endolysin, 1JWQ_A *N*-acetylmuramoyle-l-alanine amidase (3.3E − 30)
G_23	61.0	4	6	15.3	Predicted ribosomal slippage extending G_22 CDS
G_24	19.9	2	5	21.3	Three predicted transmembrane helices (residues 7–26; 41–60; 72–94)
G_25	13.3	2	6	24.4	PF09682.10; Phage_holin_6_1 (1.7E − 19). Predicted transmembrane helix (residues 5–27)
G_26	17.3	49	424	68.5	Head decoration
G_27	31.2	36	185	66.8	Major capsid, 3QPR_A MCP (4.1E − 30)

ND, not detected by mass spectrometry.

a Proteins detected by mass spectrometry, labeled as annotated in the GenBank: JN638751.1 deposited phage G genome.

b HHpred matches with an E-value > 0.18 were not included.

Mass spectral detection of proteolytic processing of the MCP

Inspection of the mass spectral sequence coverage of the MCP, gp27, provided further evidence that this protein is cleaved during head maturation. Two semi-tryptic peptides were identified for gp27, supporting that gp27 is cleaved *C*-terminal to A-27: 28-QAMANPIR-35 in gel slice 2 (Figure 2.6) and 28-QAMANPIRVSLDYQGVGRKLLVVDPLPQGALPVYDK-63 in gel slice 3. This site in gp27 was previously identified as the maturation cleavage site using Edman degradation (Hua, Huet et al. 2017). In addition, we observed that the peptide coverage of gp27 in higher molecular weight gel slices including the tryptic peptide (24-VALAQAMANPIR-35, at a 100% peptide identification probability). The identification of this peptide indicates that cleavage after A-27 did not occur in all copies of gp27 (Figure 2.6C and D) and suggests heterogeneous processing of the gp27 *N*-terminal propeptide, as observed for the MCPs of other large myoviral capsids (Donelli, Griso et al. 1976).

2.2.6 Identification of a candidate for the phage G prohead protease

The identification of proteolytic processing in phage G gp27 (Figure 2.6B–E) induced us to search for the protease responsible for its cleavage. Although the phage G genome encodes several protease candidates (e.g., gp2 and gp257, annotated as a "zinc-dependent peptidase" and a "ClpP-like protease," respectively from the Bacillus phage G genome), our bioinformatics analyses support gp19 as a stronger candidate for the phage G prohead protease. The reasons are the following.

We initially identified gp19 as a prohead protease candidate via a Psi-Blast search from gp19, which revealed matches to prohead proteases from other tailed phages. These include a match to gp77, the putative protease of Rhodococcus phage E3 (YP_008061118) in the first iteration (1e − 21), and the known prohead protease gp123 (which has a nested scaffold gp123*) of Bacillus phage 0305phi8-36 (YP_001429613.1) in the second iteration (2e − 12). A HHpred search from phage G gp19 also identified numbers of phage prohead proteases, the best match (3.0E − 16) being to Pfam PF04586.17 (Supplemental Figure S7). Pfam PF04586.17 represents the MEROPS family S78 of Caudoviral prohead serine proteases for which HK97 gp4 is the type peptidase. HK97 gp4 has an important role in head maturation, cleaving the 102-residue delta domain from

the N terminus of the MCP gp5 (Conway, Duda et al. 1995, Duda, Martincic et al. 1995, Rader, Vlad et al. 2005, Duda, Oh et al. 2013). Although the match between gp19 and Pfam PF04586.17 is highly diverged (13% identity), gp19 conserves the residues (H-129 and S-180) known to be essential for catalysis in HK97 gp4. These residues are conserved in all known phage prohead proteases that are serine proteases (Cheng, Shen et al. 2004, Liu and Mushegian 2004), including phage E3 gp77 and phage 0305phi8-36 gp123.

2.2.7 Determination that the PGH is a *Lysinibacillus* species

Since the first publications on phage G (Donelli 1968), its host has been referred to as *B. megaterium* (also sometimes *Bacillus megatherium*). During the current study, we made several observations that were not in agreement with the PGH being *B. megaterium*. We note that the taxonomic assignment of PGH was originally made decades before the development of molecular methods for bacterial identification. Initially, during TEM analyses of phage G adsorbed to its host (Figure 2.7), we observed that the bacterial cell diameter (~ 1 μm, similar to that of *Escherichia coli*) was not as large as expected for *B. megaterium*, one of the largest known bacteria with a cell volume ~ 100 × that of *E. coli*.

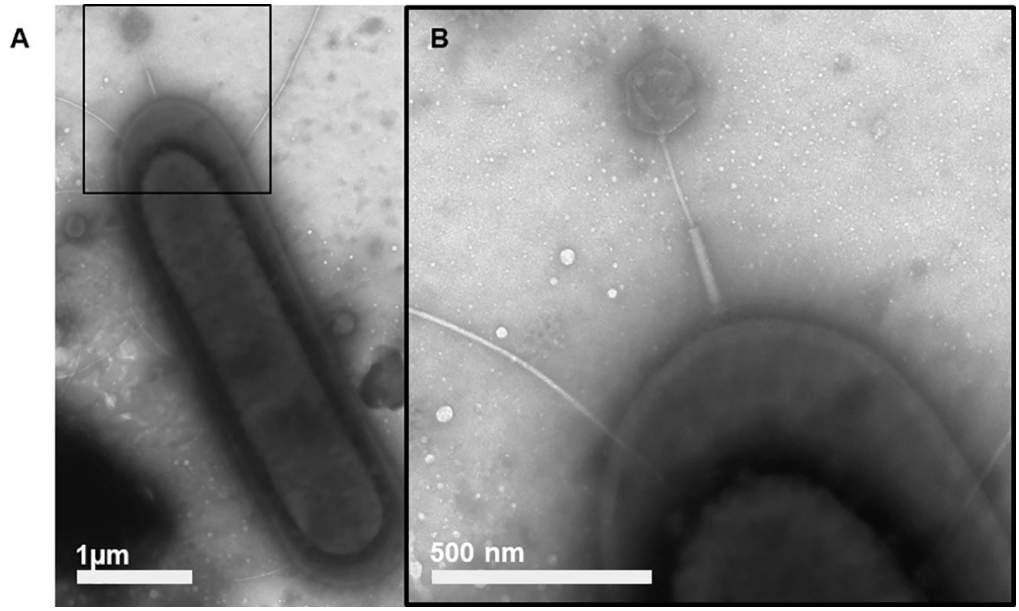


Figure 2.7. TEM micrographs of phage G adsorbed to *Lysinibacillus* species host. (A) A zoomed-out view of an infected host cell shows 3 phage G virion adhered. (B) A zoomed in view of phage G attached to its host.

Table 2.4. Bacterial host proteins identified by mass spectrometry in phage G samples.

Protein name	Organism	Accession number[a]	MW (kDa)	Exclusive peptides	Total mass spectra	Sequence coverage (%)
ATP synthase subunit beta, atpD	*L. sphaericus* (strain C3–41)	sp | B1HM56 | ATPB_LYSSC	51.3	17	80	53
Elongation factor Tu, tuf	*L. sphaericus* (strain C3–41)	sp | B1HMZ0 | EFTU_LYSSC	43.3	11	17	35
ATP synthase gamma chain, atpG	*L. sphaericus* (strain C3–41)	sp | B1HM55 | ATPG_LYSSC	31.4	6	14	29
Ribonuclease Y, rny	*L. sphaericus* (strain C3–41)	sp | B1HR37 | RNY_LYSSC	58.5	5	8	13
ATP synthase subunit b, atpF	*L. sphaericus* (strain C3–41)	sp | B1HM52 | ATPF_LYSSC	19.1	5	7	25
ATP synthase subunit delta, atpH	*L. sphaericus* (strain C3–41)	sp | B1HM53 | ATPD_LYSSC	19.4	3	5	22
60 kDa chaperonin, groL	*L. sphaericus* (strain C3–41)	sp | B1HT15 | CH60_LYSSC	57.3	4	4	11
ATP synthase subunit alpha, atpA	*L. sphaericus* (strain C3–41)	sp | B1HM54 | ATPA_LYSSC	54.7	2	3	6
Glutamate–tRNA ligase, gltX	*L. sphaericus* (strain C3–41)	sp | B1HNM5 | SYE_LYSSC	55.3	2	3	5
Serine hydroxymethyltransferase	*L. sphaericus* (strain C3–41)	sp | B1HM45 | GLYA_LYSSC	44.8	2	3	5
d-alanine aminotransferase, dat	*L. sphaericus* (strain C3–41)	sp | P54693 | DAAA_LYSSH	31.5	2	2	9
Meso-diaminopimelate d-dehydrogenase, dapdh	*L. sphaericus* (strain C3–41)	sp | Q9KWR0 | DAPDH_LYSSH	35.6	2	2	6
Phenylalanine–tRNA ligase alpha subunit, pheS	*L. sphaericus* (strain C3–41)	sp | B1HWA2 | SYFA_LYSSC	38.4	2	2	6
Probable malate:quinone oxidoreductase, mqo	*L. sphaericus* (strain C3–41)	sp | B1HNU2 | MQO_LYSSC	54.3	2	2	5

a Accession number based on NCBI deposited genome for txid444177 for *L. sphaericus* (strain C3–41).

In agreement with the TEM data, 14 *Lysinibacillus sphaericus* strain C3–41 proteins were identified in our phage G samples (Table 2.4), but no *B. megaterium* proteins. The detection of bacterial proteins that co-purified in the phage preparation was possible because the mass spectrometry searches were conducted against a database of phage G proteins (in this instance, all of the phage G proteins annotated in GenBank accession JN638751.1) concatenated to the 2012 Swiss-Prot database (2012_11_170320; version 51.6) which included *B. megaterium* and *L. sphaericus* sequences for homologous proteins.

After evaluating the mass spectrometry results, we sequenced the genome of PGH using Illumina technology (2X300 paired end reads) yielding 2,129,816 reads, 99.80% of which were determined to match the *Lysinibacillus* (strain W446) 12871 Branch (Taxon ID 400634) by Cosmid Blastn analyses using a library of 32,282 unique prokaryotic genomes. This finding, combined with our mass spectrometric identification of *Lysinibacillus* proteins in the phage G sample, strongly supports that PGH is a species of *Lysinibacillus*, not *B. megaterium*. Current efforts are focused on refining the sequences assembly/annotation of PGH to permit future database searching of the MS data.

To further investigate the host range of phage G, we performed a spot test of phage G on three strains of *L. sphaericus*, two strains of *B. megaterium*, and one strain of *L. fusiformis* compared to PGH. The efficiency of plating of phage G on all the *Lysinibacillus* and *B. megaterium* strains was < 0.01% with no individual plaques detected on any of the strains other than PGH (Table 2.5). These data are consistent with the PGH genome data showing PHG to be neither *B. megaterium* nor *L. sphaericus*, but rather a novel species of *Lysinibacillus*. Our data also suggest phage G may not have a broad host range, but further studies are required to confirm its full lytic spectrum.

Table 2.5. Sensitivity of different bacteria to phage G.

Bacteria tested Name	BGSCID[a]	Other identifier(s)	Phage G EOP[b] (%)
Lysinibacillus sp.[c]	–	PGH	100
B. megaterium (type strain)	7A36	ATCC 14581	< 0.01[d]
B. megaterium	7A1	899	< 0.01[d]
L. sphaericus (type strain)	13A6	ATCC 14577	No lysis
L. sphaericus	13A10	ATCC 12123, NRS T111	< 0.01[d]
L. sphaericus	13A18	ATCC 4525, NRS 348	< 0.01[d]
L. fusiformis (type strain)	19A1	ATCC 7055 T, NRS 350	< 0.014

a Bacillus Genetic Stock Center Identification Number.

b EOP, efficiency of plating.

c Assigned in this study.

d Lysis, including turbid lysis, was observed in several spots containing high numbers of particles (10^7 and 10^6 pfu), which we attribute to lysis from without as no single plaques were observed.

2.3 Discussion

2.3.1 Phage G capsid components and organization resemble those of other dsDNA tailed phages

As shown by our capsid reconstruction, phage G's overall capsid structure is consistent with what has been reported for other commonly studied phages. Phage G's overall capsid shell protein organization most resembles those of coliphage Lambda and marine Siphovirus TW1, despite dramatic differences in capsid and genome size. Both Lambda and TW1 capsids are much smaller than phage G in capsid shell diameter (~ 60 and ~ 68 nm in diameter, respectively) and have considerably shorter genomes (48.5 and 40 kbp, respectively), while phage G's capsid is 160 nm in diameter and has a 626-kbp packaged genome (Lander, Evilevitch et al. 2008, Hua, Huet et al. 2017, Wang, Hardies et al. 2018). The MCPs of all three phages (1) have the canonical HK97 gp5-like MCP fold, (2) are arranged with icosahedral symmetry with hexamer-pentamer clustering, and (3) have a decoration protein that sits on top of the major capsid icosahedron at 3-fold axes. Moreover, the decoration proteins for all three phages have a common architecture, a mostly beta-strand-rich core that oligomerizes into trimers and an N-terminal extension that interacts with the E-loop MCP hexamers (Lander, Evilevitch et al. 2008, Wang, Hardies et al. 2018). However, the length of the decoration protein in these three phages differs, with phage G's decoration protein being the longest, being 17 residues longer than TW1's gp56, and 55 amino acids longer than Lambda's gpD (Lander, Evilevitch et al. 2008, Wang, Hardies et al. 2018).

These findings suggest that G capsid assembly occurs by mechanisms similar to those of smaller phages. In agreement with what has been observed in other phages, the G decoration protein makes multiple contacts with the MCP, providing evidence that one function of this decoration protein is to keep the overall capsid structure stable (Prevelige Jr 2008). Phage G's decoration protein is significantly longer than TW1's and Lambda's decoration proteins, which might increase stabilization. Nonetheless, the basic strategy appears the same for G, TW1, and Lambda.(Lander, Evilevitch et al. 2008, Wang, Hardies et al. 2018).

2.3.2 Phage G DNA packaging

In contrast, we have observed two novel features of phage G DNA packaging: (1) compacted, smaller-than-capsid DNA densities with some toroid-like views and (2) lack of concentric DNA rings in the symmetric 3D reconstruction. In contrast, all previously reported phage virion 3D structures had concentric rings of encapsulated dsDNA genome (Cerritelli, Cheng et al. 1997, Chang, Weigele et al. 2006, Jiang, Chang et al. 2006, Lander, Evilevitch et al. 2008, Serwer, Wright et al. 2014). This is true not only for well-studied phages, such as T3, T4, T7, and P22, but also for jumbo phages, such as PhiKZ, N3, PBS1, 121Q, and ϕRSL1 (Effantin, Hamasaki et al. 2013, Hua, Huet et al. 2017). The rings are also observed in T3 capsids with subgenomic DNA content of less than 58% . Thus, the low packaging density of G DNA is not the cause of absence of the rings. Of note, in previous studies, the DNA-full capsid symmetric reconstruction of phage G also did not show any obvious concentric rings, although this was not discussed (Hua, Huet et al. 2017).

In our phage G cryo-EM micrographs, the envelope of packaged DNA was smaller than the capsid shell and variable in size, ranging from ~ 42 nm to a capsid-filling ~ 150 nm. At all stages of the partial packaging of other studied phages, the DNA extends to the inner surface of the capsid shell. The gap observed here, between the DNA density and the capsid shell, would not be expected (Serwer, Wright et al. 2014). In addition, in the case of phage T3, the number of concentric DNA rings increases with increasing amounts of packaged DNA.

These unique observations raise questions about how phage G's DNA packaging mechanisms differ from mechanisms of other well-studied phages. We hypothesize that the large G capsid size

and the stiffness of dsDNA, together are responsible for these observations. As dsDNA has persistence length around ~ 50 nm (Mitchell, Glowacki et al. 2017), the packaged dsDNA genome of the smaller phages is biased against condensing until it contacts either the capsid's shell or an already condensed DNA segment, as it seems for phage T3 (Serwer, Wright et al. 2014) . The result is DNA close packing that manifests as concentric rings in the 3D reconstructions, with the rings appearing first just underneath the shell and progressively increasing toward the center as more dsDNA is packaged. In contrast, phage G's capsid is roughly three times larger than a persistence length. Thus, the packaged dsDNA is less constrained to pack against the inner surface of the capsid shell. The reduced constraint on the packaged DNA results in various DNA conformations that do not produce the appearance of concentric rings in the 3D reconstruction. These conformations might include toroids.

The possibility of DNA packaging-generated DNA toroids within the phage G capsid is supported by several reports of polycation-induced DNA toroids in artificial systems including DNA containing liposomes (Manning 1978, Grønbech-Jensen, Mashl et al. 1997, Lambert, Letellier et al. 2000, Shen, Downing et al. 2000). Furthermore, phage T5 DNA toroids inside the virion have been induced in vitro by shielding the charges on DNA with polyamines (Leforestier, Šiber et al. 2011). DNA toroid formation has been reported at the collar region of phi29 (Tang, Peng et al. 2007). In Hua et al.'s report on phage G's genome electrophoresis, G DNA migrated slower than expected, hinting at modification of DNA charge (Hua, Huet et al. 2017). The modified charge may make a further contribution to formation of these DNA condensates.

2.3.3 Phage G capsid assembly

We note that, although we expect phage G to have a scaffold protein that is cleaved by the protease gp19 and removed from the head during head expansion, further analyses are required to identify the scaffold protein. The identification of scaffold proteins is complicated by the variability of scaffold proteins utilized by different phages. For instance, some phages have a dedicated scaffold protein (e.g., T4 gp22), whereas others do not (e.g., the delta domain of the HK97 gp5 acts as its scaffold (Huang, Khayat et al. 2011). In some phages, the scaffold is nested within the protease gene (e.g., gp123* in 0305phi8–36 (Huang, Khayat et al. 2011)). Since the phage G MCP, gp27, does not have the equivalent of the N-terminal delta domain of HK97 gp5, nor does protease gp19

have a nested scaffold gene, it seems likely that phage G would have a separate protein with primary function of scaffolding.

Although there are no known conserved residues found in all scaffold proteins, many tailed phage scaffold proteins have coiled-coil motifs and/or low pIs (Oh, Moyer et al. 2014). Based on having these two characteristics, there are two potential candidates for phage G scaffold/core proteins, gp11 and gp21, whose genes are located in the genome region that includes other major head proteins. However, rigorous identification requires further analyses, ideally of precursor head particles.

2.3.4 Phage G infects a *Lysinibacillus*

We also found that the PGH is a *Lysinibacillus*. We initially questioned the historic assignment of PGH as *B. megaterium* because the cells in our EM images were smaller than the known size of the *B. megaterium* cells (Figure 2.7). Our doubts were confirmed when mass spectrometry analysis of phage G identified proteins from a bacterium different to *B. megaterium* (Table 2.4). Through whole-genome sequencing, we determined that phage G's host is a *Lysinibacillus* species, a genus not known when phage G was isolated in 1969 (Donelli 1968). The findings from our proteomic and genomic analyses were additionally supported by host range tests, which showed phage G to be unable to replicate in several strains of *B. megaterium*. The previous misclassification of the PGH bacterium was likely because phage G was first isolated and described in 1968 (Donelli 1968). This was prior to the reclassification of numbers of *Bacillus* species into several new genera, including the genus *Lysinibacillus*. The subsequent establishment of the genus *Lysinibacillu*s was based on rRNA gene sequencing, DNA–DNA hybridization, and other diagnostic characteristics, such as the presence of L-lys–D-Asp in the peptidoglycan (Thomas, Hardies et al. 2007). We expect further characterization of phage G's host will help reveal the mechanisms by which this unusual virus replicates in this bacterium.

2.4 Methods

2.4.1 Phage G propagation and purification

Preparation of phage G was done as was previously described (Sun and Serwer 1997). In brief, phage G was grown in agarose overlays, concentrated by centrifugal pelleting, and subjected to rate zonal centrifugation in a sucrose gradient. After recovery from the gradient, phage G was stored in 0.01 M Tris–Cl (pH 7.4), 0.01 M MgSO4, 6% polyethylene glycol MW 3350.

2.4.2 PGH genome sequencing

Phage G's host was sent for whole-genome sequencing using Omega Bioservices (Norcross, GA, USA). Reported results identify phage G's host is a *Lysinibacillus* species. A more in-depth analysis of the PGH genome is underway (manuscript in progress).

2.4.3 Identification of phage G virion proteins by mass spectrometry

Purified phage G was obtained from zonal, sucrose gradient centrifugation as described above. The extracted fractions of pure virus were boiled in SDS-PAGE loading buffer for 5 min and separated by 1-D SDS-PAGE in a Criterion XT MOPS 12% SDS-PAGE reducing gel (Bio-Rad) and proteins visualized by staining with colloidal Coomassie blue. The gel lane was divided into 11 slices (Figure 2.6B). After destaining, proteins in the gel slices were reduced with TCEP (tris(2-carboxyethyl) phosphine hydrochloride) and then alkylated with iodoacetamide before digestion with trypsin (Promega). HPLC-electrospray ionization-tandem mass spectrometry was accomplished on a Thermo Fisher LTQ Orbitrap Velos Pro mass spectrometer fitted with a New Objective Digital PicoView 550 NanoESI source. On-line HPLC separation of the digests was accomplished with an Eksigent/AB Sciex NanoLC-Ultra 2-D HPLC system: column, PicoFrit™ (New Objective; 75 μm i.d.) packed to 15 cm with C18 adsorbent (Vydac; 218MS 5 μm, 300 Å). Precursor ions were acquired in the Orbitrap in centroid mode at 60,000 resolution (m/z 400); data-dependent collision-induced dissociation spectra of the six most intense ions in the precursor scan were acquired at the same time in the linear trap. Mascot (v2.6.0; Matrix Science; London, UK) was used to search the mass spectrometry files for each slice against a database containing phage G protein sequences concatenated with the SwissProt database (2012_11_170320; version 51.6). The phage G fasta file was exported from NCBI GenBank Accession JN638751.1 and adjusted to

include terminator (stop codon) to terminator ("TtoT") sequences that included the upstream sequences from each predicted protein start site to ensure that the predicted start sites were valid.

Subset searching of the Mascot output by X! Tandem (Craig and Beavis 2004), determination of probabilities of peptide assignments and protein identifications, and cross correlation of the Mascot and X! Tandem identifications were accomplished by Scaffold (v4.8.4; Proteome Software, Portland, Oregon), using either the files for the individual slices or the "MudPIT" option to combine the results from all slices in the lane. Quality filters in Scaffold resulting in an FDR of 0% at the protein level were as follows: peptide, 95%; protein, 99.0%; and minimum number of peptides, 2. The tandem mass spectra for each protein identified with only a single peptide were visually inspected to verify mass accuracy of < 5 ppm and detection of the majority of sequence informative fragments.

2.4.4 Testing of different bacteria for growth of phage G

Two strains of B. megaterium (7A36 and 7A1), three strains of L. sphaericus (13A6, 13A10 and 13A18), and one strain of L. fusiformis (19A1) were acquired from the Bacillus Genetic Stock Center (BGSC, http://www.bgsc.org/). The strains and PGH were propagated in sterile Luria Bertani broth supplemented with 2% nutrient broth (LB + N), 2 mM CaCl2, and 2 mM MgCl2 at 34–35 °C. Overnight cultures (100 µl) were inoculated into 4 ml of molten overlay containing 0.34% agar and poured over hard agar bottom plates. A 10-fold dilution series of a phage G stock (1.1 × 1010 pfu/ml) was spotted onto each overlay, incubated overnight at 28 °C, and examined the next morning for evidence of phage G growth.

2.4.5 Phage G cryo-EM data collection and image processing

Three microliters of phage G sample were deposited on a 400 mesh Ted Pella ultrathin lacey carbon grid and incubated for 30 min in a humid chamber on ice. The grid was then washed with 10 µl of 0.2 TM buffer (0.01 M Tris–Cl (pH 7.4), 0.01 M MgSO4, 6% polyethylene glycol MW 3350). Using a Gatan CP3 plunger, the grid was then blotted for 9 s with Whatman 1 filter paper at 65% humidity, then plunge-frozen in liquid ethane. The plunge-frozen, phage G grid was then imaged using a Titan Krios equipped with a Gatan K2 Summit direct electron detector in super-

resolution mode at the Purdue Cryo-EM Facility with a nominal magnification of 8700X resulting in 1.742 Å/super-resolution pixel. A total of 375 movies were collected.

The movies were motion corrected using motioncorr (Li, Mooney et al. 2013). The particles were manually selected using EMAN2 e2boxer.py program (Tang, Peng et al. 2007), extracted with batchboxer.py in jspr (Guo and Jiang 2014). Full capsid particles were sampled in a 12802 box without down sampling, while the empty capsid particles were down-sampled to 2.90 Å/pixel by FFT cropping to 7682 image size for further image processing. Subsequent CTF estimation de novo initial model, and icosahedral refinement/reconstruction were done using jspr (Guo and Jiang 2014). The density maps of the full and empty capsid structures are available from the Electron Microscopy Database (EMDB) under accession numbers EMD-21695 and EMD-21702, respectively.

2.4.6 Modeling and refinement of phage G capsid proteins to cryo-EM density

Sequence information for gp27 was input into Robetta for 3D structure prediction (Kim, Chivian et al. 2004). Results indicated that gp27 adopted the canonical HK97 gp5-like fold (Duda, Martincic et al. 1995, Rader, Vlad et al. 2005, Duda, Oh et al. 2013, Oh, Moyer et al. 2014). Based on the locations and shape of gp26, we noticed the similarity to TW1's decoration protein (Wang, Hardies et al. 2018). Using TW1's gp56 as a template, we generated an atomic model of gp26 used in subsequent docking and refinement steps. The first 15 flexible N-terminal amino acids of gp26 were omitted to facilitate refinement.

The gp26 and gp27 models were fit into the phage G virion density map using the collage program in Situs (Wriggers 2012). The top scoring docked models were saved individually. The cross-correlation coefficient was calculated for each set of docked models and their respective map segments. By varying the cross-correlation cutoff values, we were able to efficiently select the best docked model for each subunit. From this docking, the full gp27 hexamers, and partially modeled gp26 trimers of the phage G capsid were modeled in the density.

Further refinement of the gp26 and gp27 fully atomistic models to the density map was done using MDFF (McGreevy, Teo et al. 2016). To select an appropriate density matching weight relative to

the stereochemical energy for refinement, fitting was first performed using the even half-map densities (DiMaio, Zhang et al. 2013). An icosahedral facet density of the capsid was extracted from the model and used for refinement with even half-map reconstruction using weights of 0.1, 0.3, 0.5, and 0.7 over a 1-ns minimization and 1 ns MDFF simulation as a method to balance density and model energy during refinement as previously described (DiMaio, Zhang et al. 2013). The final frame of the resulting trajectory was used to calculate a cross-correlation against the odd half-map reconstruction. The weight value that resulted in the highest cross-correlation was then found to be 0.5 and was taken to be the optimal scale factor of those tested. This scale factor was then used for final MDFF fitting to the full reconstructed density. The gp27 and gp26 atomic models are available from the Protein Data Bank (PDB) under the accession ID 6WKK.

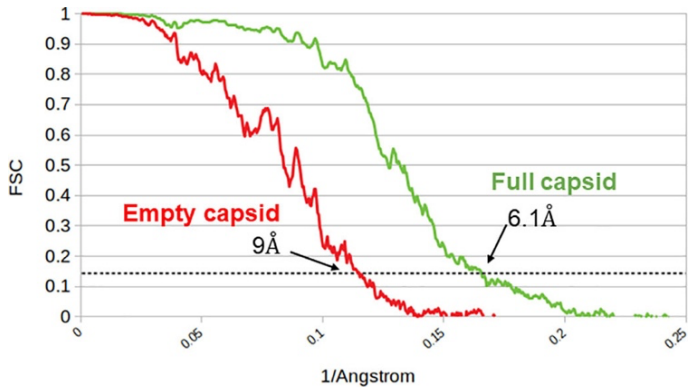
2.4.7 Cryo-electron tomography

Samples for tomography were the same used for single particle cryo-EM data collection. Thirty-two tilt series were collected using SerialEM (Mastronarde 2005) software with a tilt range of −60° to 60° in 2° increments and a nominal magnification of 8700 × on a Titan Krios with a Gatan K2 Summit direct electron detector at the Purdue Cryo-EM Facility. Tilt series were collected using a total dose of 60 e −/Å2 and a cosine distribution scheme of exposure dose. The images were binned for image processing at 1.3-nm/pixel sampling. The raw tilts were aligned using IMOD (Mastronarde and Held 2017) and reconstructed using the methods described in depth from Yan et al. using MBIR (Yan, Venkatakrishnan et al. 2019). From these tilt series, a total of 76 particles were observed. All particles were visually evaluated and one partially full particle was identified and shown in Supplementary Figure S6.

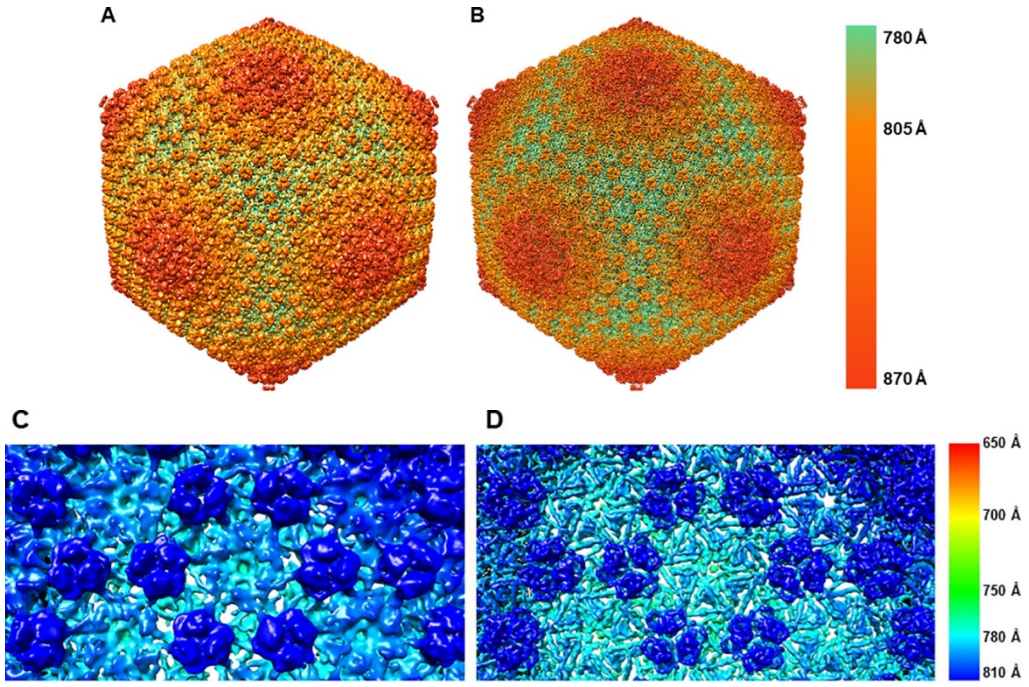
2.4.8 Negative stain analysis of host-phage G interactions

PGH bacteria were grown to mid-log phase in TB media (Sun and Serwer 1997). Cells were infected at a MOI of 1 and lysis was monitored by OD600 nm. The infected cells were then spun down at 5000g for 10 min and washed twice in PBS buffer, pH 7.0, and resuspended to a final OD600 = 1.0. The infected cells were then deposited onto lacey carbon grids coated with graphene oxide and stained with phosphotungstic acetate solution for negative stain screening. The grid was imaged on a Tecnai T20 200 kV Electron Microscope using a Gatan US1000 2Kx2K CCD camera.

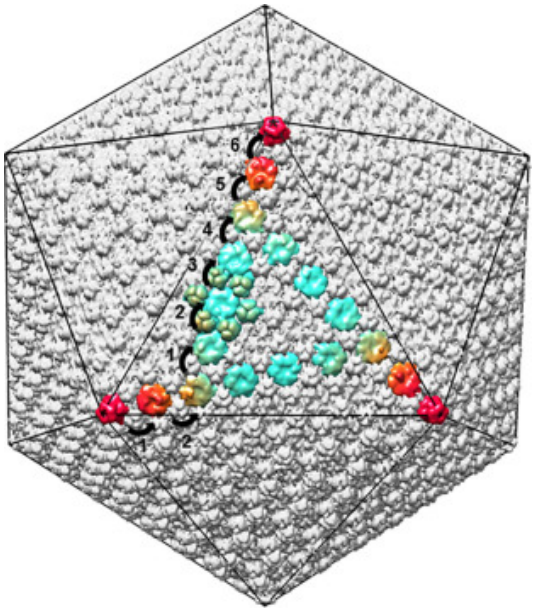
2.5 Supplemental Information



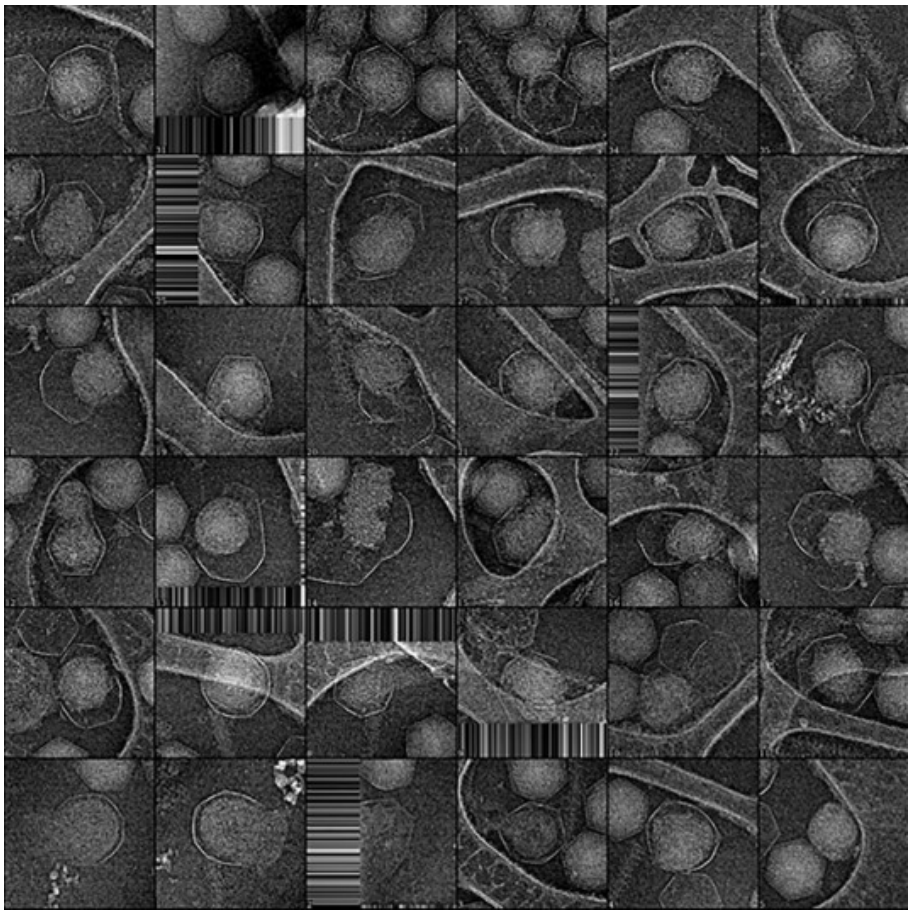
Supplemental Figure 2.S1. Phage G capsid FSC curves. Phage G's dsDNA-full and empty capsid "gold-standard" FSC curves show the structures at 6.1 Å (green) and 9 Å (red) resolution, respectively. The dsDNA-full capsid reconstruction was generated with 2564 particles. The empty capsid structure was obtained from 243 particles. The empty structure FSC curve was generated from down sampled images and therefore does not extend as far as the full structure FSC.



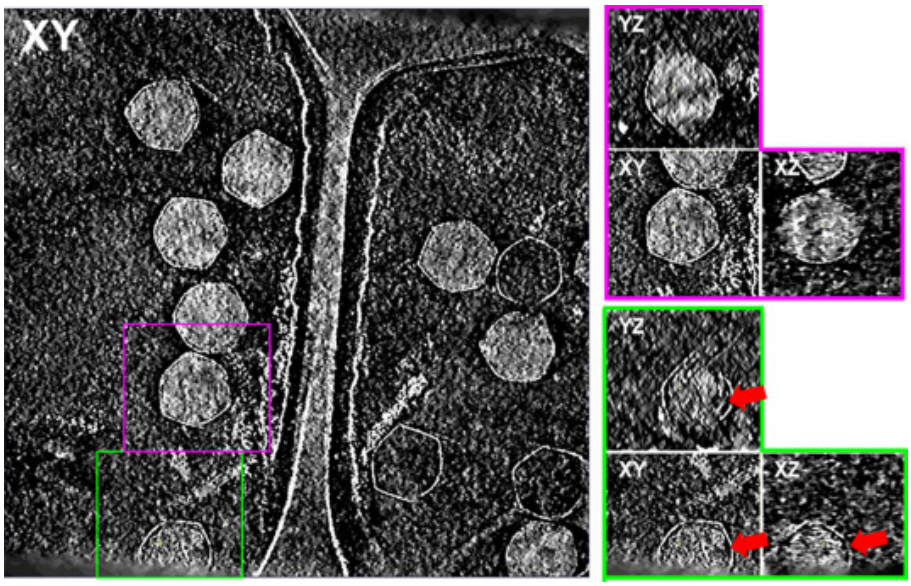
Supplemental Figure 2.S2. Phage G empty and full surface views. Phage G's dsDNA-empty and full capsid structures at 9 Å (A) and 6.1 Å (B) surface views, respectively, are shown above. The zoomed in views of hexamers and trimers at the 3-fold icosahedral axis are shown in panels C and D for the empty and full phage G capsid structures, respectively. Color keys are shown on the right side.



Supplemental Figure 2.S3. Phage G's $T = 52$ icosahedral symmetry. The figure represents phage G's $T = 52$ *dextro* ($h = 2$, $k = 6$) icosahedral symmetry, the largest of all known phage capsids to date. The triangulation number (T-number) is represented by the following formula: $T = h^2 + hk + k^2$, where h and k represented the hexamer units counted toward the nearest pentamer vertex in a straight line (Caspar and Klug 1962). Select hexamers, pentamers, and trimers are displayed in color. The black arrows indicate the hexamers counted to calculate the icosahedral T-number.



Supplemental Figure 2.S4. Ruptured phage G particles gallery. Ruptured particles in our cryo-EM micrographs of phage G showed associated DNA in a compacted state. In total, we found 36 ruptured particles with similar features pictured in the gallery above. Striped edges in some of the particles are from the edge of the micrograph.



Supplemental Figure 2.S5. Cryo-ET tomogram of phage G with particle showing DNA density detached from the capsid shell. The image on the left is a section from a cryo-ET tomogram showing several phage G capsids reconstructed using MBIR (Yan, Venkatakrishnan et al. 2019). Boxes in green and pink indicate particles that were further displayed in the X/Y/Z orthogonal views on the right. The partially full particle highlighted in the green box shows signs of a gap between the DNA density and the capsid shell (red arrow). The DNA density in the green boxed particle appears weaker than the other full particles in the tomogram. A neighboring particle with full DNA density boxed out in pink does not show this feature. Because of the low abundance of partially full particles in our dataset (2.3%), it was difficult to find more partially full particles in our tomography.

CHAPTER 3. STRUCTURAL STUDIES OF PHAGE G TAIL DEMONSTRATE A NON-CANONICAL TAIL CONTRACTION

This chapter is from a manuscript developed and submitted to MDPI Viruses in 2021 in collaboration with other scientists across different institutions. Their contributions were noted upon submission.

Abstract: Phage G is recognized as having the largest genome among isolated, propagated phages. In this study, negative stain electron microscopy of the host-phage interaction revealed a state where the tail sheath was contracted towards the distal tip and decoupled from the head-neck region. This is inconsistent with the canonical myophage tail contraction, where the sheath contracts upward, coupled to the head-neck. Our cryo-EM structures of the non-contracted and contracted tail sheath show that the protein fold of the sheath protein is consistent with smaller, contractile phages such as T4 and phi812 and phage G's sheath subunits move as rigid bodies during contraction, as described for T4. A comparison of the tail sheath helical symmetry, timetree of the sheath protein, and organization of the tail morphogenesis region of phage G genome also suggest similarity to other myophages. This observation of phage G's tail could be due to a missing anchor point at the upper end of tail sheath that allows the decoupling of the sheath from the head-neck region. Despite the genomic, sequence, and structural similarity to other myophages, the noncanonical tail contraction might indicate a unique evolutionary status of phage G among tailed phages that call for further investigation.

3.1 Introduction

Phage G's virion is remarkable because of its size — it is recognized as the largest phage that has been propagated in lab (Donelli 1968, Hua, Huet et al. 2017, Yuan and Gao 2017) – which raises many questions as to how such a large structure is assembled and maintains stability, as well as its role in infection. In our recent report on phage G, we analyzed head using cryo-electron microscopy (cryo-EM). From these data, we observed heterogeneity in the dsDNA density in the head, and we obtained the structure of its T=52, icosahedral capsid to 6.1 Å resolution, which highlighted the HK97-like major capsid proteins and a Lambda-like decoration trimers (González,

Monroe et al. 2020). Additionally, there was heterogeneity in the dsDNA density of its 498 kbp genome(González, Monroe et al. 2020).

In this study, we focused on the phage G tail structure. Our interest in analyzing phage G's tail stemmed from our initial negative stain EM analysis of phage G/host images, in which the tail sheath appeared decoupled, post-contraction, from the head/tail neck region, unlike those previously reported for the well-studied T4 and other myophages (Moody 1973, Amos and Klug 1975, Leiman, Chipman et al. 2004, Aksyuk, Leiman et al. 2009, Effantin, Hamasaki et al. 2013, Hu, Margolin et al. 2015, Nováček, Šiborová et al. 2016). In myophage, the canonical contraction of the tail sheath happens upwards toward the neck and the tail tube protrudes through the bottom of the baseplate into the host cell (or is visible if the particle is not adsorbed to a host cell) (Moody 1973, Amos and Klug 1975, Leiman, Chipman et al. 2004, Aksyuk, Leiman et al. 2009, Effantin, Hamasaki et al. 2013, Hu, Margolin et al. 2015, Nováček, Šiborová et al. 2016). Phage G's tail had previously been studied by negative stain EM over five decades ago by the original researchers that discovered phage G, but information since then about phage G's tail has been limited (Donelli 1968, Donelli, Guglielmi et al. 1972, Ageno, Donelli et al. 1973, Donelli, Dore et al. 1975, Donelli, Griso et al. 1976).

The early studies by Donelli, et. al in 1972 of phage G found that the tail was approximately 450 nm in length, and was composed of a 6-start, right-handed helix (Donelli, Guglielmi et al. 1972). Based on this analysis, it was hypothesized that both the non-contracted and contracted tail sheath had helical symmetry. The original structural studies from the 1970s also reported that upon contraction, the tail sheath underwent a compression of 22.6 Å along the axial direction with a change in twist around the axis by 6.6° per subunit (Donelli, Guglielmi et al. 1972). However, likely due to the challenges that come with helical reconstruction from electron micrographs (Egelman 2014), cryo-EM 3D reconstruction analysis had not yet been applied to the phage G tail. Here, we perform this analysis after presenting observations on the unusual contraction of the phage G tail. We also present studies to probe the evolutionary relationship of phage G relative to other tailed phages.

3.2 Materials and Methods

3.2.1 Phage G Propagation and Purification

Phage G was grown and isolated as previously described (Sun and Serwer 1997, González, Monroe et al. 2020). Briefly, phage G was amplified as a plate stock in agarose overlays, which were collected, and large debris cleared by centrifugation. Phage G was then purified by zonal centrifugation in a sucrose gradient and stored in 0.01 M Tris–Cl (pH 7.4), 0.01 M MgSO4, 6% polyethyleneglycol MW 3350.

3.2.2 Negative Stain EM of Phage G/host Interaction

We also used negative stain EM to look at phage G attached to its host strain (PGH) bacterium recently revised to a *Lysinibacillus* sp. (González, Monroe et al. 2020). PGH cells were grown to mid-log phase in TB media (10 g tryptone, 5 g NaCl in 1000 ml H_20) (Serwer, Estrada et al. 1995) and were infected at an MOI of 1 and monitored by OD600 nm until lysis occurred. The lysed and remaining cells were then spun down at 5000g for 10 min and washed twice in PBS buffer, pH 7.0, and resuspended to a final OD600 = 1. The resuspended content was then deposited onto lacey carbon grids coated with graphene oxide and stained with phospho-tungstic acetate (PTA) solution for negative stain EM. The grid was imaged on a Tecnai T20 200 kV Electron Microscope using a Gatan US1000 2Kx2K CCD camera.

3.2.3 Cryo-EM Data Collection

The cryo-EM data collection for this study was previously described in Gonzalez, et. al 2020 (González, Monroe et al. 2020). Briefly, a drop of 3µl of purified phage G was put onto a 400-mesh Ted Pella ultrathin lacey carbon grid and incubated for 30 min in a humid chamber on ice. The grid was then washed with 10µl of buffer (0.01 M Tris–Cl (pH 7.4), 0.01 M MgSO4, 6% polyethylene glycol MW 3350). The grid was frozen using a Gatan CP3 plunger, where it was blotted for 9 s with Whatman 1 filter paper at 65% humidity, then plunge-frozen in liquid ethane.

The frozen phage G grid was then imaged using the Titan Krios equipped with a Gatan K2 Summit direct electron detector in super-resolution mode at Purdue Cryo-EM Facility with 8700X nominal

magnification, and a sampling of 1.742 Å/super-resolution pixel. Overall, 375 movies were collected, which were then motion corrected using motioncorr (Li, Mooney et al. 2013). The movie averages were 2x binned to 3.484 Å/pixel for further image processing.

3.2.4 Categorization of Tail Contraction States from Cryo-EM Micrographs

In our cryo-EM dataset, it is apparent there are multiple phage G tail states even in the absence of the host. To further explore these states, we visually classified particles with complete tails into either contracted or non-contracted tail sheath conformations. Furthermore, we observed whether the contracted tail sheath was present in either near the head, in the middle section of the tail, or near the tail tip.

3.2.5 Helical Reconstruction of Non-contracted and Contracted Phage G Tail Sheath

Phage G tail was manually picked using the Relion helical picker (He and Scheres 2017). Helical segments were extracted with a 18.7 Å rise and specifying 5 asymmetric subunit using a 224 pixel box. In total, 22,755 segments were extracted. To obtain the initial helical twist and rise values of phage G tail components, the extracted segments were used in CryoSPARC for further analyses (Punjani, Rubinstein et al. 2017). The tail tube, non-contracted, and contracted tail sheath were visually separated based on 2D classification results and used for ab-initio modeling specifying C6 symmetry. From there, a relatively low-resolution 3D model was generated that could then be used to analyze the helical symmetry (e.g. twist and rise) parameters using the *helicalSym.py* program from the jspr package (Guo and Jiang 2014).

To further improve the 3D reconstructions, several additional rounds of 2D classification were done to remove heterogeneity. Helical refinement was then done using CryoSPARC to generate refined, helical models for the non-contracted and contracted tail sheath structures. All parameters used for the helical refinement were default, except for the following in the contracted phage G helical reconstruction: 27.13º twist, 18.89 Å rise, 15 maximum symmetry order, C6 symmetry. All parameters used for the helical refinement of the non-contracted were default, except for the following: 20.57º twist, 41.53 Å rise, 40 maximum symmetry order, C6 symmetry.

3.2.6 Bioinformatic Evolutionary Analysis of Phage G Tail Sheath

A protein sequence-based evolutionary analysis of the phage G sheath protein gp178 was performed as described previously (Hardies, Thomas et al. 2016) by Stephen C. Hardies (University of Texas Health Science Center San Antonio). Briefly, the T02 aligner (Hughey and Krogh 1996, Karplus, Barrett et al. 1998) obtained from the U.C.S.C bioinformatics group (https://compbio.soe.ucsc.edu/sam.html) was used to align a homologs set generated as the union of PsiBlast hits keyed with diverse known myoviral sheath protein sequences. A subset of sheath proteins covering a broad representation of the resulting tree was then selected for refinement. The quality of alignment across two of the most divergent lineages (T4 to Bxz1, and Bxz1 to phage G) was checked by HMM-HMM comparison using HHpred (Steinegger, Meier et al. 2019), resulting in limiting the final tree to the region corresponding to residues 357-655 of T4 gp18, and removal of sequences not contiguously aligned in that region. A further check to avoid perturbation of the tree by recombinant or partially aligned sequences was to divide that region in two and remove any sequences not producing a congruent tree in the two subsections. The final tree was calculated using MrBayes (Ronquist, Teslenko et al. 2012) with an independent gamma rate relaxed clock model (Bouckaert, Heled et al. 2014). The time scale was set by alignment of nodes in the SPO1, LP65, Bastille clade and the T4, Aeh1, KVP40, Syn1 clade with a scaled large terminase tree as described (Hardies, Thomas et al. 2016).

3.3 Results

3.3.1 Negative Stain EM of Phage G Host Attachment

The goal of these analyses was to gain a better understanding of the phage G tail structure and its role in infection. Our interest in phage G's sheath was initially piqued by preliminary TEM of negatively stained phage G particles adsorbed to the PGH host cell wall. In those micrographs, we observed numbers of particles whose tail sheath was contracted in a manner that was not consistent with previous descriptions of the tail behavior of other myoviruses, where the contracted tail sheath is always coupled to the neck region (Moody 1973, Amos and Klug 1975, Leiman, Chipman et al. 2004, Aksyuk, Leiman et al. 2009, Effantin, Hamasaki et al. 2013, Hu, Margolin et al. 2015, Nováček, Šiborová et al. 2016). Those initial results led us to perform a Cryo-electron microscopy

reconstruction of the phage G tail sheath to gain a better understanding of the structural basis for its curious behavior.

TEM of negatively stained phage G and its host revealed particles adsorbed to the PGH host surface (Figure. 3.1A). In various cases, the head did not have strong density – suggesting that this conformation could have been post genome ejection. All the phage G particles observed appeared to have a contracted tail conformation with the tail sheath located at the distal end of the tail in contact with the host surface via the tail fibers (Figure. 3.1). This is in stark contrast with typical observations of contracted myophages, such as T4, where the contracted tail sheaths are located at the opposite end next to the DNA-containing head (Moody 1973, Amos and Klug 1975, Leiman, Chipman et al. 2004, Aksyuk, Leiman et al. 2009, Effantin, Hamasaki et al. 2013, Hu, Margolin et al. 2015, Nováček, Šiborová et al. 2016). To find if these unusual observations of contracted tail sheath near the tail tip are unique to our data, we examined the literature and found multiple negative stain images of purified phage G particles in the 1973 report of the discovery and characterization of phage G, in which the contracted phage G tail sheath was seen at different positions along the tail (below to the head, at the middle of the tail, and near the tail tip) although the localization of contracted tail sheath near the tail tip was not discussed (Figure. 3.1B) (Ageno, Donelli et al. 1973).

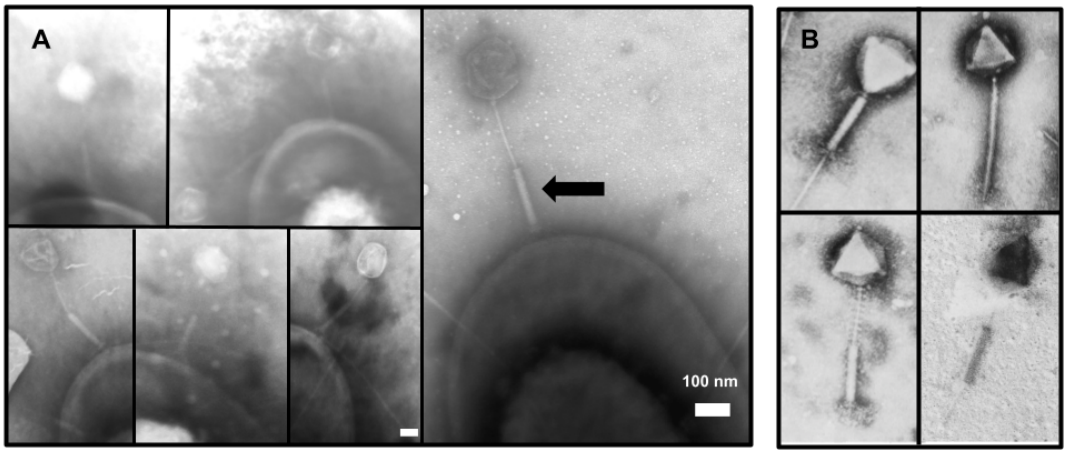


Figure 3.1. TEM of negatively stained phage G showing multiple states of the contractile sheath (A) Phage G particles attached to the host PGH cell and (B) Micrographs of Phage G virions reproduced from Ageno et al. 1973 (Ageno, Donelli et al. 1973). All phage G particles observed to be adsorbed to the host all had their tail's contracted toward the distal tip of phage G's tail which was in contact with host cell surface. The arrow indicates an example of one sheath contracted in a non-canonical manner. The white bar represents 100 nm in length. B. In the studies by Ageno, et al., the phage G tail sheath was observed to be contracted at different positions along the tail (next to the head, at the middle of the tail, and near the tail tip), demonstrating the odd behavior of the phage G sheath, we observed both by TEM and Cryo-EM (see below).

3.3.2 Tail Contraction States in Our Cryo-EM Data

To rule out the possibility of the above observations being negative stain artifacts, we collected single particle cryo-EM images of purified phage G particles. In these analyses we observed multiple states of the tail contraction among the phage G particles, similar to the earlier observations noted above (Figure. 3.1B). We manually categorized phage G particles from our cryo-EM micrographs based on the morphology of the tail sheath and its location along the tail. Phage particles with completely visible, easily distinguishable tails were counted, resulting in 364 distinct phage G particles identified.

All the tails were then categorized based on their tail i) contraction state, ii) the location of the contracted tail sheath relative to the head, and iii) head state, as summarized in Table 1. Examples of each state described in Table 3.1 are shown in Figure. 3.2. The two contraction states are non-

contracted (uniform thickness along the entire tail) and contracted (variable thickness along the tail). The locations of contracted tail sheath are grouped into 3 places, near head, middle, and near-tip. As we previously reported, phage G sometimes has a partially full head with the DNA appearing centralized and detached from the capsid protein shell (González, Monroe et al. 2020), and we refer to this state as the partial DNA state. Capsids with apparent DNA-full or DNA empty head categorizations we assigned visually.

Overall, there were 211 particles (58%) that were identified as non-contracted, with the remaining 153 (42%) phage G particles having contracted tail sheath states. The tail sheath contractile states did not appear significantly influenced by the presence or absence of DNA in the head. In DNA-full phage, 40% of tails exhibited contraction, whereas in partial DNA and DNA-empty phage, 58% exhibited contraction, although this percentage number is less reliable due to the small number of particles (21 out of 36) in these states.

The most unusual observation was that 123 of the 153 contracted-sheath particles (80%) had the sheath detached from the neck region and in contact with the baseplate at the distal tip of the tail. Only 18% of the contracted particles had the tail sheath located under the neck region. The remaining 2% had the tail sheath in the middle of the tail, making no contact with either the baseplate or neck. The contracted states with the tail sheath near the middle of the tail or at the distal tip of the tail are inconsistent with the observations and the proposed molecular mechanisms described for myophages, such as T4 (Leiman, Arisaka et al. 2010), in which the tail sheath must be mechanically coupled to the head at the neck region in all contractile states.

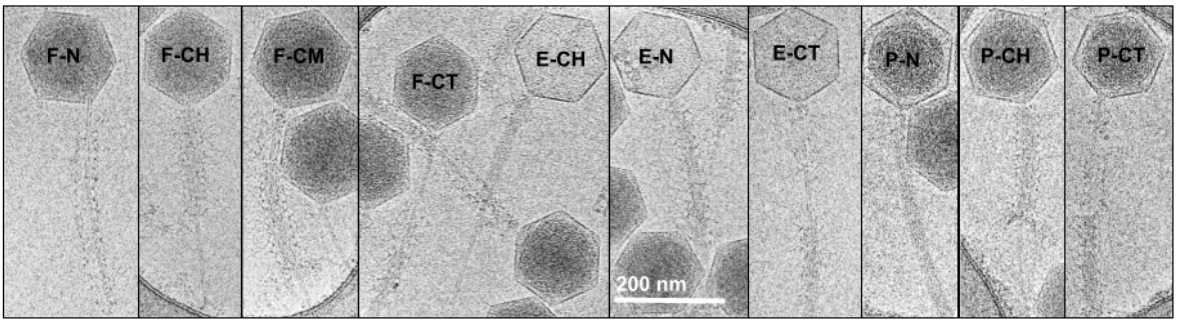


Figure 3.2. Categorized tail states from single-particle cryo-EM of purified phage G in different tail contraction states in the absence or presence of DNA. Particles are labeled as F: DNA full, P: Partial DNA, E: empty - N: Non-contracted, CH: Contracted near head, CM: Contracted Middle, CT: Contracted at Tip as shown by the representative images.

Table 3.1. Tail contraction and head states from cryo-EM micrographs.

Capsid state	**Non-contracted**	**Contracted Tail Location**		
		Near Head	Middle	Near Tip
DNA full	196	25	3	104
Partial DNA	3	1	0	7
Empty	12	1	0	12

3.3.3 Phage G Tail Components from 2D Classification

The various tail contraction states we observed from our cryo-EM micrographs (Figure. 3.2) raised questions about the mechanism of phage G tail contraction. To have better insight into this non-canonical tail contraction behavior, we continued our investigation by determining the structure of the tail sheath in phage G using single-particle cryo-EM analysis.

Focusing on the tail structural analysis, the phage G tail was picked from micrographs using the manual helical particle picking tool in Relion (He and Scheres 2017) from the head-tail junction down past the tail tip. From there, segments were extracted using the Relion helical segmenting tool (He and Scheres 2017), with a 93.5 Å interbox distance and a 224 pixel box size. In total, 22,755 particles were extracted. Using CryoSPARC (Punjani, Rubinstein et al. 2017), the particles were then subjected to multiple rounds of reference-free 2D classification. From the 2D classification results, we were able to detect different phage G tail components, including: the neck region (Figure. 3.3C), the non-contracted tail sheath (Figure. 3.3D), the tail tube (Figure. 3.3E),

the tail tube and contracted tail sheath junction (Figure. 3.3F), and the contracted tail sheath (Figure. 3.3G).

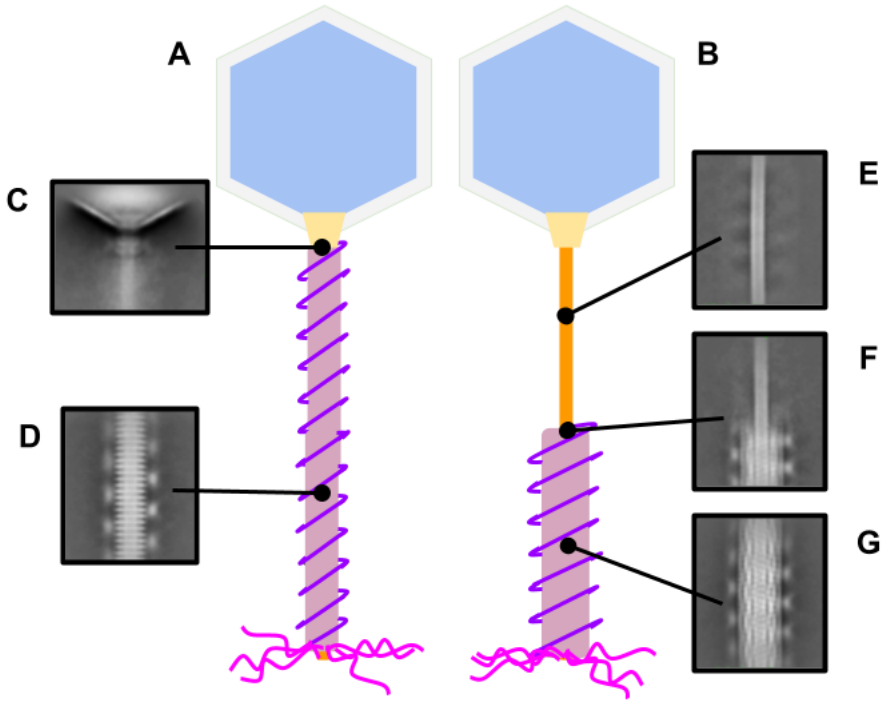


Figure 3.3. 2D classification of phage G tail components. The cartoons above represent the non-contracted (A) and contracted (B) tail sheath states of phage G. The small insets are 2D classifications from the cryo-EM data that point to the corresponding tail components in the cartoons as follows: (C) neck region, (D) non-contracted sheath, (E) tube, (F) contracted sheath and tube junction, (G) contracted sheath.

During the 2D classification analysis, we also noticed density on the side of the tail sheath that corresponds to the outer coil density that has also been described in Donelli, et. al's 1972 negative stain EM study (Donelli, Guglielmi et al. 1972). This is a unique feature, that to our knowledge, has not been observed in other phages. To take a closer look at this feature, we measured the distances between the peaks of density and assigned it to the outer coil diameter and helical pitch (Figure. 3.4).

In the non-contracted tail sheath, the measured distance of the outer coil diameter was 390 Å. In the contracted tail sheath conformation, the outer coil density diameter increases to 490 Å (Figure. 3.4). The axial distance between the peaks of outer coil density were the same (210 Å) in both tail sheath states (Figure. 3.4). It is interesting to note that the outer coil maintains its axial pitch while its diameter undergoes significant change. It suggests that neither end of the outer coil is coupled to the tail sheath and, instead, the upper end of outer coil is coupled to the neck region and the bottom end is coupled to the baseplate/tail tube to maintain the total length and helical pitch during sheath contraction while its diameter was forced to widen by the contracted sheath. The outer coil density was not well resolved in our cryo-EM reconstructions of the tail sheath and had a blurred appearance in the 2D classification results. We speculate these results are because the coil is dynamic in nature and/or has different helical symmetries than the tail sheath (Figure. 3.4).

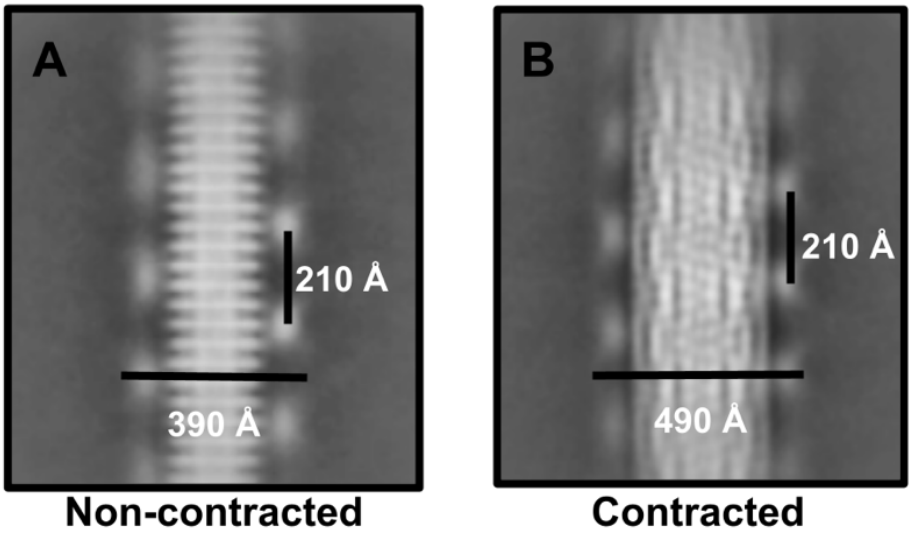


Figure 3.4. Phage G outer coil density in 2D classification. Density from the outer coil was observed in phage G sheaths that were (A) non-contracted, and (B) contracted. In both sheath states the outer coil had the same measured axial pitch distance (210 Å). However, the outer coil's diameter expanded approximately 100 Å in the transition from the non-contracted state (390 Å) to the contacted state (490 Å).

3.3.4 3D Cryo-EM Structure of the Non-contracted and Contracted Phage G Tail Sheath

The 2D classification results of phage G highlighted the heterogeneity and variety in its tail states. The 2D classes visually identified as non-contracted and contracted tail sheath were subjected to

3D helical reconstruction using CryoSPARC (Punjani, Rubinstein et al. 2017). The resolution of the non-contracted and contracted tail sheath cryo-EM density is approximately 7-8 Å, and 6-7 Å, respectively. The apparent secondary structure elements, such as alpha helices, that are clearly observed in the structures (Figs. 3.5, 3.6, 3.7), support the resolution claims. The outer and inner diameter of the non-contracted tail sheath of phage G is approximately 240 Å and 60 Å, respectively (Figure. 3.5 and Table 3.2). After contraction the outer and inner diameter of phage G's tail sheath expands in diameter to 320 Å and 120 Å in the outer an inner diameter, respectively (Figure. 3.5 and Table 3.2). The tail sheath proteins of phage G are arranged in a six strand, right-handed helix around the inner tail tube structure (shown in transparent grey Figure. 3.5), through which the dsDNA is transferred during infection. The tail sheath proteins in the non-contracted state, in each helical strand, are organized with a symmetry of 20.57° twist and a rise of 41.53 Å (Figure. 3.5 and Table 3.2).

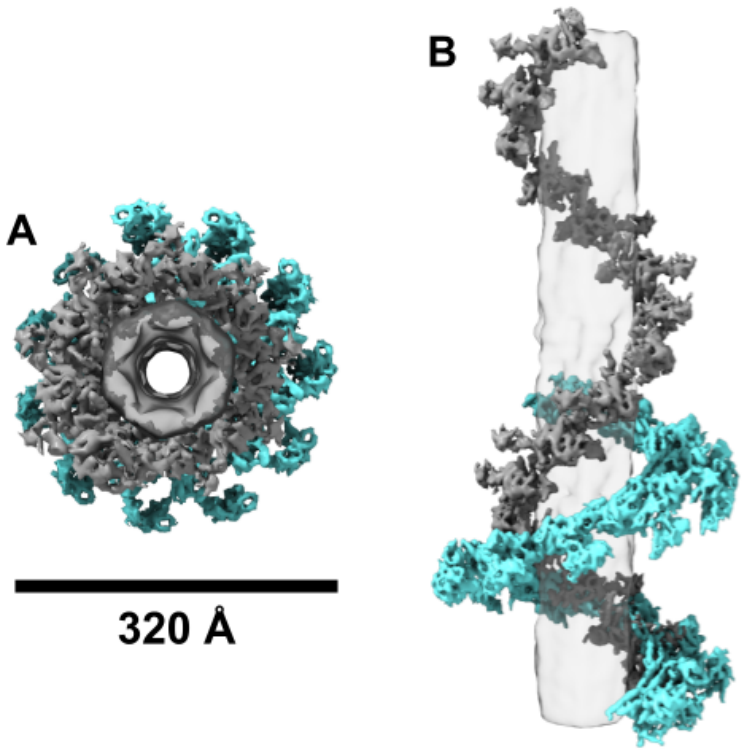


Figure 3.5. Non-contracted and contracted tail sheath organization. This is a visual representation of 16 tail sheath subunits in the non-contracted conformation (grey), and in the contracted conformation (cyan) cryo-EM structures in the top (A) and side (B) view. The tail tube density is shown in transparent grey.

The contracted tail sheath structure maintains the six stranded, right-handed helical arrangement, but the sheath proteins compact to form a helix with helical symmetry of 27.13º twist and 18.89 Å rise (Figure. 3.5 and Table 3.2). With the compacting from the tail sheath contraction, the outer diameter of the sheath widens by 80 Å and the inner diameter doubles in width. The total length of the phage G tail is 4500 Å, and the non-contracted tail sheath organization has a rise of 41.53 Å, therefore, ~648 sheath subunits are arranged on the tail in total. From this, the calculated length the completely contracted phage G tail sheath would be 108 subunits per strand multiplied by 18.89 Å rise per subunit, which is 2040 Å. This is consistent with our observations of the phage G contracted state from our negative stain and cryo-EM micrographs, where about half of the tail length is occupied by the contracted tail sheath (Figs. 3.1 and 3.2).

Table 3.2. Cryo-EM helical structure details of phage G tail sheath in non-contracted versus contracted states.

Helical structure features	Non-contracted	Contracted
Outer diameter (Å)	240	320
Inner diameter (Å)	60	120
Pitch (Å)	706.01	245.57
Rise (Å)	41.53	18.89
Twist (°)	20.57	27.13

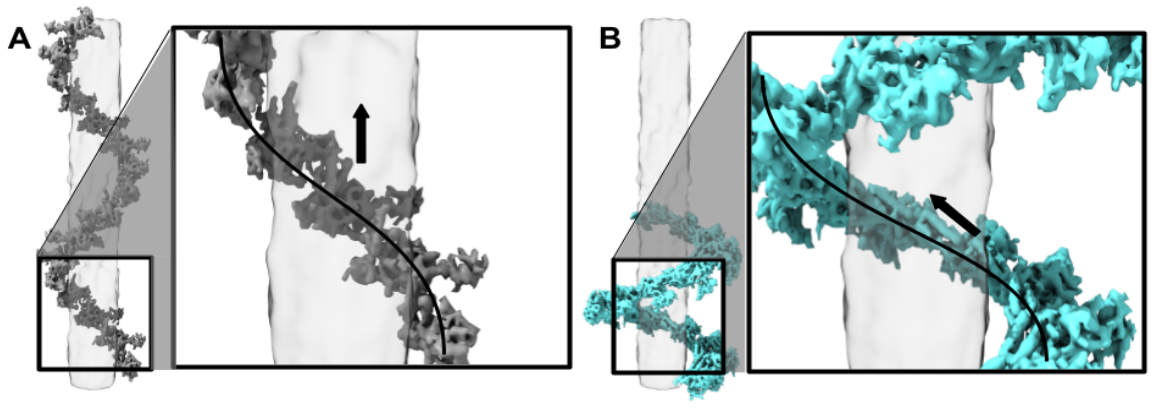


Figure 3.6. Rigid phage G tail sheath subunit twisting upon contraction. (A) non-contracted; (B) contracted. There are contacts with two long alpha helices in the inner core of the tail sheath (grey) in parallel with the tube (transparent grey). The red arrow indicates the direction of the two alpha helices. Upon contraction, sheath proteins go through a rigid movement (see Figure. 7A) where the subunits twist approximately 30º. In both (A) and (B) the phage head would be at the top of the figure.

3.3.5 Phage G Tail Sheath Subunit Structure and Arrangement

At the subunit level, phage G's tail sheath (gp178) is a 579 amino acids long protein that is 63 kDa (Pope 2011). The sheath proteins form a six-stranded, right-handed helix that wraps around the tail tube structure, and it is a dynamic feature of the tail. The phage G sheath protein has a larger core close to the tail tube, and a smaller, outward protruding density region (Figs. 3.6, 3.7). Within the core region of the phage G sheath protein, multiple rod-like densities of alpha helices are resolved (Figure. 3.7) and there are two main alpha helices that are closest in proximity to the tube structure (transparent grey Figure. 3.6). The outer most facing (away from the tail tube) density of phage G's sheath protein is less well resolved than the core region (Figure. 3.7A). From the density, there appears to be less alpha helical density in this region.

At the domain level, phage G's sheath protein appears to have highly similar overall organization as T4 (Aksyuk, Leiman et al. 2009), as labeled in Figure 7B. From the partial crystal structure of T4, the four domains I, II, III, and IV are organized from most exposed to most buried towards the tube, respectively (Aksyuk, Leiman et al. 2009). Domain I (residues 98-188) has a 6 stranded beta-barrel and an alpha helix (Aksyuk, Leiman et al. 2009). The domain II is then composed of a two-layer beta sandwich, surrounded by 4 alpha helices, and is defined by residues 88-97 and 189-345 (Aksyuk, Leiman et al. 2009). The domain III is then described as a beta sheet with 6 beta strands and 6 alpha helices and is residues 20–87 and 346–510 (Aksyuk, Leiman et al. 2009). Finally, domain IV, has not been resolved in the T4 crystal structure, but is composed of the termini (residues 1-20 and 510-659) (Aksyuk, Leiman et al. 2009). Cryo-EM structural studies on phi812 also found that its sheath protein, gp103, had a similar organization at the individual sheath protein subunit level to T4 (Nováček, Šiborová et al. 2016). In phi812, the cryo-EM density in the domain IV region has been described to contain 2 major helices (Nováček, Šiborová et al. 2016). The region has not been described from the T4 studies because of the limited structural information (Aksyuk, Leiman et al. 2009, Leiman, Arisaka et al. 2010, Aksyuk, Kurochkina et al. 2011), but we have also observed these two helices in our phage G gp178 cryo-EM density (Figs. 3.6, 3.7).

Like T4, in phage G's gp178 sheath structure, domain I is the outermost facing region of the sheath (Fig 3.7B). It is also the least resolved portion of the sheath protein density (Figure. 3.7A). The density for domain II is not obvious, as outlined for T4 (Arisaka, Nakako et al. 1988, Aksyuk,

Kurochkina et al. 2011), which could be due to phage G's sheath protein, gp178 (579 amino acids) being smaller than T4's gp18 (659 amino acids) (Arisaka, Nakako et al. 1988). In phage G, as for T4's gp18, domain III is mainly alpha helical (Figure. 3.7B) (Aksyuk, Leiman et al. 2009). A prominent bundle of three alpha helices in domain III of phage G's density matches well with T4's helices in the same region (Figure. 3.7B).

In phage G's gp178 sheath protein, domain I is the outermost facing region of the sheath (Fig 3.7B). It is also the least resolved portion of the sheath protein density (Figure. 3.7A). The density for domain II is not obvious, as outlined for T4 (Arisaka, Nakako et al. 1988, Aksyuk, Kurochkina et al. 2011), which could be due to phage G's sheath protein, gp178 (579 amino acids) being smaller than T4's gp18 (659 amino acids) (Arisaka, Nakako et al. 1988). In phage G, as for T4's gp18, domain III is mainly alpha helical (Figure. 3.7B) (Aksyuk, Leiman et al. 2009). A prominent bundle of three alpha helices in domain III of phage G's density matches well with T4's helices in the same region (Figure. 3.7B).

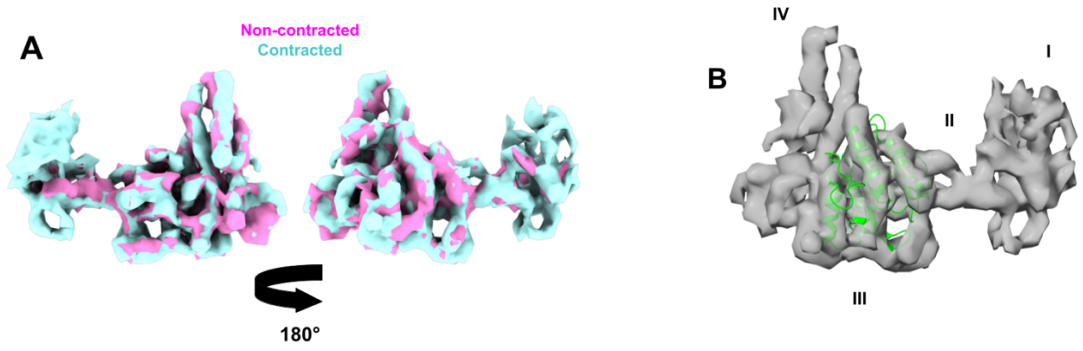


Figure 3.7. Phage G tail sheath subunit structure. A. The phage G tail sheath subunit from the non-contracted (pink) and contracted (cyan) structures were superimposed using ChimeraX fit-to-model feature (Pettersen, Goddard et al. 2021). The densities superimpose almost completely, except for the outer domain region, where the non-contracted structure (pink) has poorly defined density. B. The phage G sheath protein structure has several alpha helices as shown by the rod-like density of the contracted tail sheath subunit (transparent grey). The phage G sheath structure also follows the same domain organization described for T4 (Aksyuk, Leiman et al. 2009, Aksyuk, Kurochkina et al. 2011). The green alpha helices were fit from the T4 tail sheath crystal structure residues 21 – 201 (PDB: 3FOA) (Aksyuk, Leiman et al. 2009).

From our cryo-EM analysis, we obtained the structures of the non-contracted and the contracted phage sheath (Figure. 3.6). We compared the individual sheath protein density in both states and found the individual sheath proteins maintained the same subunit structure in both contraction states (Figure. 3.7A). As demonstrated in Figure. 3.6, the structural information suggests the sheath proteins move as rigid bodies by approximately 30º. The non-contracted phage G sheath state has a 41.53 Å rise and 20.27º twist (Table 3.2). Upon contraction, the rise decreases by 22.63 Å, and the twist increases by 6.56º (Table 3.2). The contraction causes a rigid rotation of the sheath protein subunits in the helical tail, which leads to a difference in subunit packing of the helix (Figure. 3.6).

3.3.6 Phage G Tail Sheath Helical Symmetry Compared to Other Known Phages

After obtaining the cryo-EM structure of the phage G non-contracted and contracted sheath, we then investigated how its helical symmetry parameters compared to existing structures of tailed phages. We collected available phage tail data from literature and plotted the helical symmetry of all the data, including our results for the phage G tail sheath (Figure. 3.8, Table 3.3).

In this comparison, we included information from 4 myophages (T4, Φ812, ΦKZ, and ΦRSL1) (Abuladze, Gingery et al. 1994, Lecoutere, Ceyssens et al. 2009, Effantin, Hamasaki et al. 2013, Nováček, Šiborová et al. 2016), and 7 siphophages (λ, YSD1, SPP1, Araucaria, T5, P2, and TP901-1) (Katsura 1990, Sassi, Bebeacua et al. 2013, Büttner, Wu et al. 2016, Mahony, Alqarni et al. 2016, Arnaud, Effantin et al. 2017, Hardy, Dunstan et al. 2020, Zinke, Sachowsky et al. 2020). The results show two relatively distinguishable clusters. The first cluster that coincides with the non-contracted phage tail sheath (grey squares and circle) ranges from 17–22° twist and 36–42.8 Å rise. All the non-contracted myophage tail sheath structures are found in this cluster plus the λ, YSD1, and SPP1 (shown as X's), which are also found nearby (Figure. 3.8). The second cluster coinciding with the contracted tail sheath is not as tightly grouped and can be found ranging from a 27.18 - 34.1° twist to 16.4 - 18.89 Å rise in Figure. 3.8. In this cluster, only the myophage structures were grouped. The 4 phage tail structures falling outside of these two clusters were all siphophages (shown as X's) in Figure. 3.8 (Araucaria, TP901-1, T5, and P2).

Overall, the T4 tail is 925 Å long, and contracts to 420 Å (Moody 1973, Aksyuk, Leiman et al. 2009, Taylor, Prokhorov et al. 2016), and the phi812 non-contracted tail is 2,020 Å long, and

contracts to 808 Å (Nováček, Šiborová et al. 2016). The phage G tail sheath is ~4500 Å long in the non-contracted state and compressing to 2040 Å upon contraction. Although the length of phage G's tail is more than 2X longer than the phi812 tail and almost 5X longer than the T4 tail, the tail sheath of all three phages are 6-stranded, right-handed helix tails with relatively similar helical symmetry in the non-contracted and contracted states (Figure. 3.8, Table 3.3).

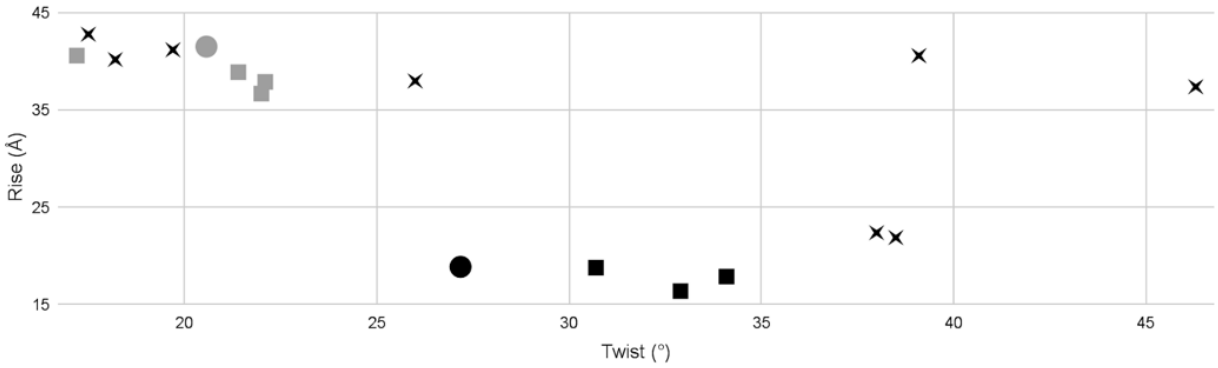


Figure 3.8. Helical symmetry parameters of tail structure across various phages. The grey boxes are the non-contracted tail sheath structures of myophage. The black boxes are the contracted myophage structures, the X's represent siphophages, and the circles are the phage G structures from this study. In this study, we included information from 4 myophages (T4, phi812K1-420, phiKZ, and phiRSL1) (Abuladze, Gingery et al. 1994, Lecoutere, Ceyssens et al. 2009, Effantin, Hamasaki et al. 2013, Nováček, Šiborová et al. 2016), and 7 siphophages (λ, YSD1, SPP1, Araucaria, T5, P2, and TP901-1) (Katsura 1990, Sassi, Bebeacua et al. 2013, Büttner, Wu et al. 2016, Mahony, Alqarni et al. 2016, Arnaud, Effantin et al. 2017, Hardy, Dunstan et al. 2020, Zinke, Sachowsky et al. 2020).

Table 3.3. Helical symmetry comparison among various tailed phage.

	Virus	Structure	Twist (°)	Rise (Å)	Reported Resolution (Å)	EMDB entry	Host Gram (-/+)	Citation
Myophage	**phage G**	non-contracted sheath	20.57	41.53	7-8	-	+	Current study
		contracted sheath	27.18	18.89	6-7	-		
	phi812K1-420	non-contracted sheath	21.4	38.9	6.2	4051	-	(Nováček, Šiborová et al. 2016)
		contracted sheath	30.7	18.8	4.2	4052		
	phiRSL1	non-contracted sheath	22.1	37.9	9.6	2244	-	(Effantin, Hamasaki et al. 2013)
	phiKZ	contracted poly sheath	34.1	17.9	19.0	5331	-	(Aksyuk, Kurochkina et al. 2011)
		non-contracted sheath	22	36.7	18.0	5332		
	T4	contracted sheath	32.9	16.4	N/A	N/A	-	(De Rosier and Klug 1968); (Leiman, Chipman et al. 2004)
		non-contracted	17.2	40.6	N/A	N/A		
		tube	18.2	40.2	3.4	8767		(Zheng, Wang et al. 2017)
Siphophage	**spp1**	tail	38.5	21.9	4.0	10792	+	(Zinke, Sachowsky et al. 2020)
	YSD1		19.7	41.2	3.5	22183	-	(Hardy, Dunstan et al. 2020)
	Lambda		17.5	42.8	6.4	20242	-	(Campbell, Duda et al. 2020)
	T5		39.1	40.6	6.0	3692	-	(Arnaud, Effantin et al. 2017)
	p2		46.3	37.4	22.0	2464	-	(Bebeacua, Tremblay et al. 2013)
	araucaria		26	38	24.0	2337	+	(Sassi, Bebeacua et al. 2013)
	TP901-1		38	22.4	20.0	2228	+	(Bebeacua, Lai et al. 2013)

3.3.7 Evolutionary Analysis of Phage G Tail Sheath Protein gp178

Since our structural analyses demonstrated that the overall structures and helical arrangements of the subunits of the phage G tail sheath had similarity to those of other myophages, we performed a protein sequence-based evolutionary analysis of the phage G sheath protein gp178 that were computed as described previously (Hardies, Thomas et al. 2016). To compare sheath proteins from an array of genetically diverged phages, the analyses were based on the more highly conserved *C*-terminal region of the sheath protein (residues 357-655 of T4 gp18). The resulting time tree was calibrated in time by congruence with a large terminase tree in the T4 and SPO1 clades.

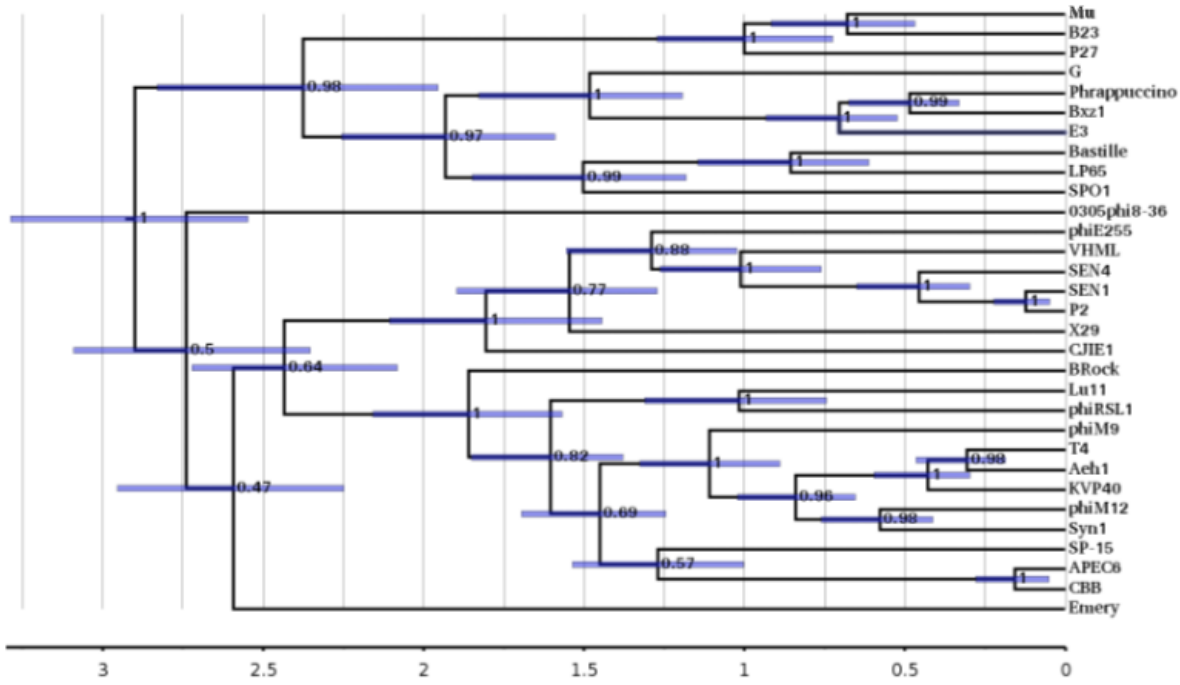


Figure 3.9. Phage G sheath timetree. A timetree bioinformatic analysis was done to analyze the evolutionary relationship among various phage tail sheath proteins. The timeline unit was 1.0 Gya. It covered residues 357-655 of T4 gp18 and was calibrated in time by congruence with a large terminase tree in the T4 and SPO1 clades. The error bars are shown in purple and are accompanied by their respective quantities.

These analyses found that, despite phage G's sheath being structurally similar to that of T4 in both its helical symmetry (Figure. 3.8) and at the individual sheath subunit level (Figure. 3.7), the sheath proteins of the two phages are diverged such that they likely derived from an ancestral split that occurred over 2.5 billion years ago (Figure. 3.9). In addition, even the sheath proteins of what are

currently the closest known relatives to the phage G sheath protein (in phages Phrappuccino, Bxz1, and E3) are predicted to have evolved from an ancestral split that occurred almost 1.5 Gya.

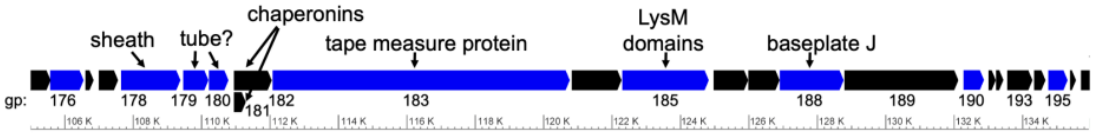


Figure 3.10. Scheme of the phage G genome region containing the major tail morphogenesis genes. Gene products (gp) identified by mass spectrometry (González, Monroe et al. 2020) in purified virions are shaded blue.

3.3.8 Phage G Genome Tail Morphogenesis Region

The intriguing dichotomy of phage G sheath's similarities and differences to those of other myoviruses led us to seek a better understanding of other components of the phage G contractile tail. The phage G sheath gene is located toward the 5' end of a module of genes in its genome (NCBI: NC_023719) (Pope 2011) whose synteny is reminiscent of that observed in many, but not all, myoviral genomes (e.g., Mu (Morgan, Hatfull et al. 2002), 0305phi8-36 (Hardies, Thomas et al. 2007), ARV1(Kaliniene, Šimoliūnas et al. 2017)). For example, downstream of the sheath gene is a pair of chaperonin genes, and immediately downstream of those is the tape measure protein (TMP) gene and a series of genes that likely encode baseplate components and tail fibers (Figure. 3.10). The tail chaperonins (gp181 and gp182) were annotated as such due to their having a signature translational frameshift that was first described for the G-T genes of Lambda (Levin, Hendrix et al. 1993). Gp181 and gp182 were not identified in our previous mass spectral analyses of phage G (González, Monroe et al. 2020), consistent with their predicted function.

However, numbers of proteins encoded in the phage G tail morphogenesis region were identified by mass spectrometry, and have other characteristics and/or sequence similarity (as determined by BlastP or HHpred) that strongly support their being components of the tail (González, Monroe et al. 2020). For instance, one of these proteins gp183 has the classic characteristics of a TMP. The TMP gene stands out in many phage genomes simply for its length (often the longest gene), which is as a consequence of the TMP's role as the tail length determinant (e.g., T4 (Abuladze, Gingery et al. 1994) and TP901-1 (Mahony, Alqarni et al. 2016)). To perform this role, a few copies (3-6)

of the TMP extend as an alpha helical structure from the baseplate to the neck region within the central core of tail tube. Considering its function and the length of the phage G tail (450 nm) it is not surprising that gp183 is 2893 residues long and has a predicted high overall content of a-helices (50 %) and coils (44 %) (Lupas, Van Dyke et al. 1991, Wang, Li et al. 2016, González, Monroe et al. 2020). The C-terminus of gp183 likely assists with genome ejection into the host cell by interacting with, likely degrading, the cell wall – an additional function of the TMP in some phages (Rodríguez-Rubio, Gutiérrez et al. 2012). This expectation is based on Blast and HHpred matches in the C-terminal region of gp183 to the LytD superfamily of Beta- N-acetylglucosaminidases (residues 2439-2595, 1.99e-29) and Peptidase family M23 (residues 2622-2716, 1.20e-35). Other proteins whose genes are downstream of the TMP gene, also had homology to domains and/or phage tail proteins that support their roles as baseplate and/or tail fiber proteins. For instance, gp188, has diverged similarity to the T4 baseplate wedge protein gp6 (4.1e-28) and P2 gpJ (2.6e-25).

The tube candidates are gp179 (242 residues) and gp180 (189 residues), the genes for which are located between the sheath gene and the "G-T" chaperonins which is the normal location for the tube gene in many myoviruses (Figure. 3.10). Both gp179 and gp180 have diverged homology to a series of tube proteins, including those from diffocins (e.g., xkdM of the Bacillus subtilis prophage PBSX), myoviruses (e.g., gp19 of T4) and even siphophages (e.g., gpV of Lambda) as determined by HHpred (all matches had probabilities >95 %). Whether phage G gp179 and gp180 both form part of the tube, or one of them forms the tube and the other has a different function (e.g., binding of the tube to the neck or baseplate), is unclear. This unexpected finding warrants further research to clarify their roles.

3.4 Discussion

3.4.1 Unusual tail sheath contraction in phage G

In our initial investigation with negatively stained phage G/host cells and purified phage G using cryo-EM, our images suggested the tail sheath contraction showed a differed behavior from the canonical depiction of myophage contraction. Although phage G's contracted tail sheath was seen at various locations along the tail in the cryo-EM images of purified phage G (Figure. 3.2), all of

the phage G observed in our negative stain analysis of phage G attached to host cell surface depict it as having the contracted tail sheath portion towards the tip of the tail, decoupled from the neck region (Figure. 3.1). This opposes the observations of T4 phage, the type phage of myophages, in which the contracted tail sheath was always coupled to the head/neck region.

This observation is in conflict with the current model proposed for DNA ejection through the protruded tail tube that is powered by the contraction of tail sheath for myophages (King 1968, Moody 1973, Aksyuk, Leiman et al. 2009). In the current model based on the structural observations with T4 (Moody 1973, Leiman, Chipman et al. 2004, Kostyuchenko, Chipman et al. 2005), the irreversible tail sheath contraction in myophages is proposed to be initiated by baseplate conformational change initiated by interactions from tail fiber attachment to the host cell surface. This contraction propagates from the baseplate in an upwards wave motion towards the head (Moody 1973, Aksyuk, Leiman et al. 2009). It is thought that there is stored potential energy within the uncontracted sheath, whereby upon contraction, the energy is transferred to the tail tube for puncturing the host cells membrane and cell wall to eject the dsDNA genome into host cell cytoplasm (Moody 1973, Leiman, Chipman et al. 2004, Kostyuchenko, Chipman et al. 2005, Leiman, Arisaka et al. 2010).

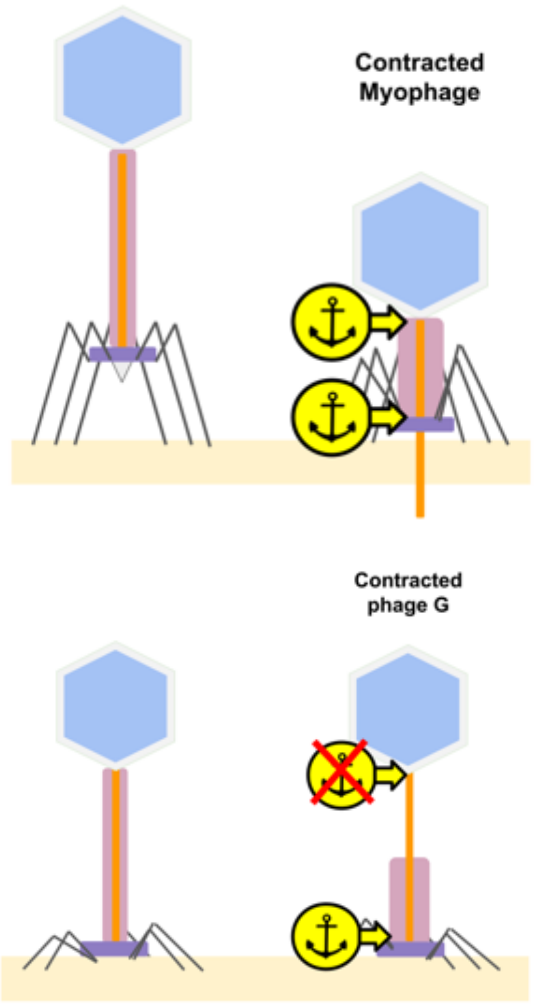


Figure 3.11. Tail sheath anchor points in a typical myophage and phage G during contraction. A. In the canonical tail sheath contraction described for myophage (Leiman, Chipman et al. 2004, Kostyuchenko, Chipman et al. 2005), where the tail sheath stays in contact with the neck region in all states by the interactions with the tail baseplate and head/neck via the two anchor points at the bottom and upper end of tail sheath, respectively. B. In phage G, the upper anchor is likely missing, which results in the non-canonical tail contraction where the contracted sheath is positioned at the tail tip as observed in our phage G/host cell images (Figure. 1A).

3.4.2 Missing Tail Sheath Anchor Point at the Neck Region in Phage G

As previously described the canonical tail sheath contraction described for myophages (Leiman, Chipman et al. 2004, Kostyuchenko, Chipman et al. 2005) is based on the tail sheath staying in contact with the head/neck region in all states. During contraction of a typical myophage, the tail sheath contracts upward toward the head, and this is proposed to help penetrate the inner tube

through the host cell membrane to proceed with infection (Moody 1973, Leiman, Chipman et al. 2004, Hu, Margolin et al. 2015) as shown in Figure. 3.11A.

Structurally, this process requires two anchor points, one at the bottom to tie the bottom end of tail sheath to the baseplate and the other at the top to tie the top of tail sheath to the head/neck of the phage. Upon contraction, the shortening of tail sheath and the widening of the inner diameter of the tail sheath will release the grip of inner tail tube, bring down the head/neck/tail tube, and force the inner tube slip through the shortened tail sheath and penetrate through the host cell membrane before phage DNA genome is ejected into host cytoplasm (Molineux 2001).

There have been anchor points described for T4 as illustrated in Figure. 3.11 with the yellow arrows. In T4, these proteins have been described as gp3 which contacts right before the tail terminator (Vianelli, Wang et al. 2000), gp15 which helps with the head to tail attachment (King 1968), gp25 which is hypothesized to initiate contraction with structural similarity to the sheath (Taylor, Prokhorov et al. 2016).

The non-canonical tail contraction with tail sheath staying at the tail tip of phage G (Figure. 3.1A) could be explained by the model as described above but with the upper anchor missing – while the functional bottom anchor still holds the contracted tail sheath at the host cell surface, the loss of the upper anchor could no longer help protrude the tail tube mechanically. Furthermore, in the isolated phage G as shown in our cryo-EM images (Figure. 3.2), the contracted sheath would no longer be forced to stay at the tail tip due to the lack of attachment of the bottom end of sheath/baseplate to the host cell surface, but could instead freely slide and stay at an arbitrary location along the tail tube.

Our analysis of phage G's tail morphogenesis gene region also gives us more information about the structural components of its tail in relation to other myophage (Figure. 3.11). Despite the similarity in the overall synteny of the phage G major tail gene module with those of other phages, and the functional assignment of a handful of genes, there remains more questions regarding the

other phage G tail components than what is known. For instance, what is the full complement of genes required to form its tail, baseplate and helical fibers? Based on the number of different proteins identified as part of the phage G virion, as well as precedents in other structurally complex myoviruses (e.g., SPO1, T4) there is likely >20 different proteins required to form the phage G tail. Similarly, what is the role of each of the phage G tail components? Research to address these questions is certain to generate novel findings as hinted at by the existence of two phage G proteins (gp179, gp180) that have similarity to known tail tube proteins (Figure. 3.10).

3.4.3 Evolutionary Implication of Phage Tail Mediated Infection Mechanisms and Future Directions

The observations of a non-canonical tail sheath contraction in phage G were intriguing. Further investigations of the organization of genes in tail morphogenesis region of the genome, sequence analyses of the sheath proteins, the tail sheath structures at multiple levels – helical arrangement of the entire sheath, protein fold of the sheath protein, and their changes upon contraction – have shown noticeable similarities and evolutionary relationship to other myophages.

Few reports have mentioned a tail contraction behavior similar to phage G, where the contracted sheath can be found decoupled from the neck region. Negative stain images of isolated phage in contractile states decoupled from the neck region were found for *Listeria monocytogenes* phage 0176 (Loessner and Calendar 2006), and three phages infecting *Burkholderia pseudomallei*, KS5, KS14 (Lynch, Stothard et al. 2010), and ST2 (Yordpratum, Tattawasart et al. 2011). In this report, we give the first structural insight of the contraction mechanism and protein structure of the sheath in phage G to probe the contraction states we have observed.

Thus, the results reported in this study provide an additional discovery of a seemingly outlier of myophage with its tail sheath contraction and DNA ejection mechanism that cannot be explained by current model for myophages. Many follow-up studies will be needed to further clarify how phage G uses its tail to attach to the host cell surface and inject its DNA genome into host cell cytoplasm, and if this process could still be considered the same process for all myophages, an

evolutionary intermediate between myophages and siphophages, or should be considered a new class of phages with a distinctive mechanism of infection.

We suggest the following structural studies: cryo-electron tomography studies of phage G/host cell interactions at different time points of post-infection; single particle cryo-EM studies of phage G at atomic-model resolutions (3.5 Å or better) with an emphasis on the tail sheath and its junction with head/neck; role of the outer coil in the infection process.

CHAPTER 4. CONCLUSIONS

In summary, from the results of all our studies with phage G using cryo-EM, we have discovered some notable points about its structure including:

- Heterogenous, partially-full states of seemingly centralized, compacted genome in our single-particle cryo-EM data
- The capsid structure of phage G is remarkably larger than other better studied *Myoviridae* such as T4, but its capsid building blocks are generally shaped, and assembled like smaller phages
- We suspected, and ultimately showed through genome sequencing and host specificity studies that phage G's host was a *Lysinibacillus*, not a *B. megaterium*, as mainly referenced in the past
- Even when purified, and in the absence of host, the phage G tail appears to adopt multiple contracted conformations along the length of the tail, which is inconsistent with current the current understanding of *Myoviridae*
- Our observations of phage G's tail contracting towards the distal tip and decoupled from the neck region are consistent with previous studies done by Donelli, Guglielmi et al. (1972)
- We successfully used helical reconstruction without the need for helical index in real space to obtain the contracted and extended structures of phage G's tail sheath to subnanometer resolution
- We found the non-contracted phage G tail sheath to have the helical symmetry of: 20.57° twist and a rise of 41.53 Å
- We found the contracted phage G tail sheath had a helical symmetry of: 27.13º and 18.89 Å rise
- From our preliminary host adsorption studies with negative stain TEM using phage G, we found an unusual tail contraction pattern where the tail sheath contracted towards the distal end of the tip, decoupled from the neck region

Overall, information on large phages, such as phage G (Hua, Huet et al. 2017) is severely limited in comparison to the amount of knowledge available for smaller phages. This may be due to difficulty in purification and preparation of the large phage (Hua, Huet et al. 2017, Yuan and Gao 2017). Historically, research on phage has led to fundamental developments in our understanding of molecular biology, which has helped us advance influential technologies such as CRISPR-Cas9 and phage therapy (Harper 2021). Furthermore, since phage are so abundant and ubiquitous, understanding their ecology and biological mechanisms is beneficial for us in advancing our basic understanding of biology and infection. By studying phage G, an outlier in the current group of knowledge for tailed phage, we have added a better understanding of its structural components—giving us insight on how such a large phage is assembled. Furthermore, these being our initial studies with phage G, we have noted its tail contraction behavior where the contracted sheath is decoupled from the neck region, which has only been reported a few times (*Listeria monocytogenes* phage 0176 (Loessner and Calendar 2006), and three phages infecting *Burkholderia pseudomallei*, KS5, KS14 (Lynch, Stothard et al. 2010), and ST2 (Yordpratum, Tattawasart et al. 2011)), and structural information on this observation was previously unavailable. Understanding how and why phage G and the other reported phage contract could help us understand the process of infection and the evolutionary significance of the tail sheath.

There are still many experiments that could be designed to further investigate its function and assembly. For example, genetic manipulation of phage G to knockout selective tail proteins could help us understand which components are necessary for contraction, and address the importance of this function in host infection. Furthermore, there are also still various questions left to explore on phage G assembly. Phage G's dsDNA-packaged genome is approximately 626kbp (Sun and Serwer 1997, Pope 2011, Hua, Huet et al. 2017, González, Monroe et al. 2020), and it's host genome is 2.1Mbp (González, Monroe et al. 2020), begging the question as to how many phage can be assembled per infection being that the genetic material is limited.

In conclusion, our work with phage G has uncovered information about its capsid structure, host identity, and unique tail behavior and organization. Together, this information is a great starting point for continuing studies with phage G in combination with its host. Based on our existing data, we expect phage G to have unique properties during host adsorption and we would expect to find

new details about its assembly with experiments such as cryo-ET. Although phage G has been described and recognized as a special virus giant since 1968 (Donelli 1968), there are still many opportunities to learn about basic phage biology using phage G as a model system.

REFERENCES

Abuladze, N. K., M. Gingery, J. Tsai and F. A. Eiserling (1994). "Tail length determination in bacteriophage T4." Virology **199**(2): 301-310.

Ackermann, H.-W. (2009). Phage Classification and Characterization. Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions. M. R. J. Clokie and A. M. Kropinski. Totowa, NJ, Humana Press**:** 127-140.

Ackermann, H.-W. and D. Prangishvili (2012). "Prokaryote viruses studied by electron microscopy." Archives of virology **157**(10): 1843-1849.

Ageno, M., G. Donelli and F. Guglielmi (1973). "Structure and physico-chemical properties of bacteriophage G. II, The shape and symmetry of the capsid." Micron (1969) **4**(4): 376-403.

Aksyuk, A. A., L. P. Kurochkina, A. Fokine, F. Forouhar, V. V. Mesyanzhinov, L. Tong and M. G. Rossmann (2011). "Structural conservation of the myoviridae phage tail sheath protein fold." Structure **19**(12): 1885-1894.

Aksyuk, A. A., P. G. Leiman, L. P. Kurochkina, M. M. Shneider, V. A. Kostyuchenko, V. V. Mesyanzhinov and M. G. Rossmann (2009). "The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria." EMBO J **28**(7): 821-829.

Al-Shayeb, B., R. Sachdeva, L.-X. Chen, F. Ward, P. Munk, A. Devoto, C. J. Castelle, M. R. Olm, K. Bouma-Gregson and Y. Amano (2020). "Clades of huge phages from across Earth's ecosystems." Nature **578**(7795): 425-431.

Amos, L. and A. Klug (1975). "Three-dimensional image reconstructions of the contractile tail of T4 bacteriophage." Journal of molecular biology **99**(1): 51-64.

Arisaka, F., T. Nakako, H. Takahashi and S.-l. Ishii (1988). "Nucleotide sequence of the tail sheath gene of bacteriophage T4 and amino acid sequence of its product." Journal of virology **62**(4): 1186-1193.

Arnaud, C.-A., G. Effantin, C. Vivès, S. Engilberge, M. Bacia, P. Boulanger, E. Girard, G. Schoehn and C. Breyton (2017). "Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection." Nature communications **8**(1): 1-9.

Baker, M. L., W. Jiang, F. J. Rixon and W. Chiu (2005). "Common ancestry of herpesviruses and tailed DNA bacteriophages." Journal of virology **79**(23): 14967-14970.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.

Bebeacua, C., L. Lai, C. S. Vegge, L. Brøndsted, M. van Heel, D. Veesler and C. Cambillau (2013). "Visualizing a complete Siphoviridae member by single-particle electron microscopy: the structure of lactococcal phage TP901-1." Journal of virology **87**(2): 1061-1068.

Bebeacua, C., D. Tremblay, C. Farenc, M.-P. Chapot-Chartier, I. Sadovskaya, M. van Heel, D. Veesler, S. Moineau and C. Cambillau (2013). "Structure, adsorption to host, and infection mechanism of virulent lactococcal phage p2." Journal of virology **87**(22): 12302-12312.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut and A. J. Drummond (2014). "BEAST 2: a software platform for Bayesian evolutionary analysis." PLoS computational biology **10**(4): e1003537.

Büttner, C. R., Y. Wu, K. L. Maxwell and A. R. Davidson (2016). "Baseplate assembly of phage Mu: Defining the conserved core components of contractile-tailed phages and related bacterial systems." Proc Natl Acad Sci U S A **113**(36): 10174-10179.

Campbell, P. L., R. L. Duda, J. Nassur, J. F. Conway and A. Huet (2020). "Mobile loops and electrostatic interactions maintain the flexible tail tube of bacteriophage lambda." Journal of molecular biology **432**(2): 384-395.

Casjens, S. and J. King (1975). "Virus assembly." Annual review of biochemistry **44**(1): 555-611.

Caspar, D. L. and A. Klug (1962). Physical principles in the construction of regular viruses. Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press.

Cerritelli, M. E., N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy and A. C. Steven (1997). "Encapsidated conformation of bacteriophage T7 DNA." Cell **91**(2): 271-280.

Chang, J., P. Weigele, J. King, W. Chiu and W. Jiang (2006). "Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery." Structure **14**(6): 1073-1082.

Chen, D.-H., M. L. Baker, C. F. Hryc, F. DiMaio, J. Jakana, W. Wu, M. Dougherty, C. Haase-Pettingell, M. F. Schmid and W. Jiang (2011). "Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus." Proceedings of the National Academy of Sciences **108**(4): 1355-1360.

Cheng, H., N. Shen, J. Pei and N. V. Grishin (2004). "Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease." Protein Science **13**(8): 2260-2269.

Cheng, Y. (2018). "Single-particle cryo-EM—How did it get here and where will it go." Science **361**(6405): 876-880.

Conway, J., R. Duda, N. Cheng, R. Hendrix and A. Steven (1995). "Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system." Journal of molecular biology **253**(1): 86-99.

Craig, R. and R. C. Beavis (2004). "TANDEM: matching proteins with tandem mass spectra." Bioinformatics **20**(9): 1466-1467.

Crowther, R., E. V. Lenk, Y. Kikuchi and J. King (1977). "Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4." Journal of molecular biology **116**(3): 489-523.

Cuervo, A., M. Pulido-Cid, M. Chagoyen, R. Arranz, V. A. González-García, C. Garcia-Doval, J. R. Castón, J. M. Valpuesta, M. J. van Raaij, J. Martín-Benito and J. L. Carrascosa (2013). "Structural characterization of the bacteriophage T7 tail machinery." J Biol Chem **288**(36): 26290-26299.

Davidson, A. R., L. Cardarelli, L. G. Pell, D. R. Radford and K. L. Maxwell (2012). "Long noncontractile tail machines of bacteriophages." Adv Exp Med Biol **726**: 115-142.

De Rosier, D. and A. Klug (1968). "Reconstruction of three dimensional structures from electron micrographs." Nature **217**(5124): 130-134.

DiMaio, F., J. Zhang, W. Chiu and D. Baker (2013). "Cryo-em model validation using independent map reconstructions." Protein Science **22**(6): 865-868.

Donelli, G. (1968). "ISOLATION OF A BACTERIOPHAGE OF EXCEPTIONAL DIMENSIONS ACTIVE IN R MEGATHERIUM." ATTI DELLA ACCADEMIA NAZIONALE DEI LINCEI RENDICONTI-CLASSE DI SCIENZE FISICHE-MATEMATICHE & NATURALI **44**(1): 95-&.

Donelli, G., E. Dore, C. Frontali and M. Grandolfo (1975). "Structure and physico-chemical properties of bacteriophage G: III. A homogeneous DNA of molecular weight 5× 108." Journal of molecular biology **94**(4): 555-565.

Donelli, G., G. Griso, L. Paoletti and S. Rebessi (1976). Capsomeric Arrangement in the Bacteriophage G Head. Sixth Eur. Reg. Conf. Electron Microsc.(Jerusalem).

Donelli, G., F. Guglielmi and L. Paoletti (1972). "Structure and physico-chemical properties of bacteriophage G. I. Arrangement of protein subunits and contraction process of tail sheath." J Mol Biol **71**(2): 113-125.

Duda, R. L., K. Martincic and R. W. Hendrix (1995). "Genetic basis of bacteriophage HK97 prohead assembly." Journal of molecular biology **247**(4): 636-647.

Duda, R. L., B. Oh and R. W. Hendrix (2013). "Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation." Journal of molecular biology **425**(15): 2765-2781.

Effantin, G., R. Hamasaki, T. Kawasaki, M. Bacia, C. Moriscot, W. Weissenhorn, T. Yamada and G. Schoehn (2013). "Cryo-electron microscopy three-dimensional structure of the jumbo phage ΦRSL1 infecting the phytopathogen Ralstonia solanacearum." Structure **21**(2): 298-305.

Egelman, E. H. (2014). "Ambiguities in helical reconstruction." Elife **3**: e04969.

Fang, P.-A., E. T. Wright, S. T. Weintraub, K. Hakala, W. Wu, P. Serwer and W. Jiang (2008). "Visualization of Bacteriophage T3 Capsids with DNA Incompletely Packaged In Vivo." J Mol Biol **384**(5): 1384-1399.

Fokine, A., P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao and M. G. Rossmann (2004). "Molecular architecture of the prolate head of bacteriophage T4." Proceedings of the National Academy of Sciences **101**(16): 6003-6008.

Fokine, A. and M. G. Rossmann (2014). "Molecular architecture of tailed double-stranded DNA phages." Bacteriophage **4**(2): e28281.

González, B., L. Monroe, K. Li, R. Yan, E. Wright, T. Walter, D. Kihara, S. T. Weintraub, J. A. Thomas and P. Serwer (2020). "Phage G structure at 6.1 AA resolution, condensed DNA, and host identity revision to a lysinibacillus." Journal of Molecular Biology **432**(14): 4139-4153.

Grønbech-Jensen, N., R. J. Mashl, R. F. Bruinsma and W. M. Gelbart (1997). "Counterion-induced attraction between rigid polyelectrolytes." Physical Review Letters **78**(12): 2477.

Guo, F. and W. Jiang (2014). Single particle cryo-electron microscopy and 3-D reconstruction of viruses. Electron Microscopy, Springer**:** 401-443.

Guo, F., Z. Liu, P.-A. Fang, Q. Zhang, E. T. Wright, W. Wu, C. Zhang, F. Vago, Y. Ren and J. Jakana (2014). "Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions." Proceedings of the National Academy of Sciences **111**(43): E4606-E4614.

Hardies, S. C., J. A. Thomas, L. Black, S. T. Weintraub, C. Y. Hwang and B. C. Cho (2016). "Identification of structural and morphogenesis genes of Pseudoalteromonas phage φRIO-1 and placement within the evolutionary history of Podoviridae." Virology **489**: 116-127.

Hardies, S. C., J. A. Thomas and P. Serwer (2007). "Comparative genomics of Bacillus thuringiensis phage 0305φ8-36: defining patterns of descent in a novel ancient phage lineage." Virology Journal **4**(1): 1-17.

Hardy, J. M., R. A. Dunstan, R. Grinter, M. J. Belousoff, J. Wang, D. Pickard, H. Venugopal, G. Dougan, T. Lithgow and F. Coulibaly (2020). "The architecture and stabilisation of flagellotropic tailed bacteriophages." Nature communications **11**(1): 1-11.

Harper, D. R. (2021). Bacteriophages : biology, technology, therapy.

He, S. and S. H. W. Scheres (2017). "Helical reconstruction in RELION." J Struct Biol **198**(3): 163-176.

Hrebík, D., D. Štveráková, K. Škubník, T. Füzik, R. Pantůček and P. Plevka (2019). "Structure and genome ejection mechanism of Staphylococcus aureus phage P68." Science advances **5**(10): eaaw7414.

Hu, B., W. Margolin, I. J. Molineux and J. Liu (2015). "Structural remodeling of bacteriophage T4 and host membranes during infection initiation." Proceedings of the National Academy of Sciences **112**(35): E4919-E4928.

Hua, J., A. Huet, C. A. Lopez, K. Toropova, W. H. Pope, R. L. Duda, R. W. Hendrix and J. F. Conway (2017). "Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the HK97 fold." MBio **8**(5).

Huang, R. K., R. Khayat, K. K. Lee, I. Gertsman, R. L. Duda, R. W. Hendrix and J. E. Johnson (2011). "The Prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation." Journal of molecular biology **408**(3): 541-554.

Hughey, R. and A. Krogh (1996). "Hidden Markov models for sequence analysis: extension and analysis of the basic method." Bioinformatics **12**(2): 95-107.

Hulo, C., E. De Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios and P. Le Mercier (2011). "ViralZone: a knowledge resource to understand virus diversity." Nucleic acids research **39**(suppl_1): D576-D582.

Jiang, W., J. Chang, J. Jakana, P. Weigele, J. King and W. Chiu (2006). "Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus." Nature **439**(7076): 612-616.

Kaliniene, L., E. Šimoliūnas, L. Truncaitė, A. Zajančkauskaitė, J. Nainys, A. Kaupinis, M. Valius and R. Meškys (2017). "Molecular analysis of Arthrobacter myovirus vB_ArtM-ArV1: we blame it on the tail." Journal of virology **91**(8): e00023-00017.

Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka and M. G. Rossmann (2002). "Structure of the cell-puncturing device of bacteriophage T4." Nature **415**(6871): 553-557.

Karplus, K., C. Barrett and R. Hughey (1998). "Hidden Markov models for detecting remote protein homologies." Bioinformatics (Oxford, England) **14**(10): 846-856.

Katsura, I. (1990). "Mechanism of length determination in bacteriophage lambda tails." Adv Biophys **26**: 1-18.

Kim, D. E., D. Chivian and D. Baker (2004). "Protein structure prediction and analysis using the Robetta server." Nucleic acids research **32**(suppl_2): W526-W531.

King, J. (1968). "Assembly of the tau of bacteriophage T4." Journal of molecular biology **32**(2): 231-262.

Kostyuchenko, V. A., P. R. Chipman, P. G. Leiman, F. Arisaka, V. V. Mesyanzhinov and M. G. Rossmann (2005). "The tail structure of bacteriophage T4 and its mechanism of contraction." Nature structural & molecular biology **12**(9): 810-813.

Lambert, O., L. Letellier, W. M. Gelbart and J.-L. Rigaud (2000). "DNA delivery by phage as a strategy for encapsulating toroidal condensates of arbitrary size into liposomes." Proceedings of the National Academy of Sciences **97**(13): 7248-7253.

Lander, G. C., A. Evilevitch, M. Jeembaeva, C. S. Potter, B. Carragher and J. E. Johnson (2008). "Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM." Structure **16**(9): 1399-1406.

Lecoutere, E., P. J. Ceyssens, K. A. Miroshnikov, V. V. Mesyanzhinov, V. N. Krylov, J. P. Noben, J. Robben, K. Hertveldt, G. Volckaert and R. Lavigne (2009). "Identification and comparative analysis of the structural proteomes of ϕKZ and EL, two giant Pseudomonas aeruginosa bacteriophages." Proteomics **9**(11): 3215-3219.

Leforestier, A., A. Šiber, F. Livolant and R. Podgornik (2011). "Protein-DNA interactions determine the shapes of DNA toroids condensed in virus capsids." Biophysical journal **100**(9): 2209-2216.

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

Leiman, P. G., F. Arisaka, M. J. van Raaij, V. A. Kostyuchenko, A. A. Aksyuk, S. Kanamaru and M. G. Rossmann (2010). "Morphogenesis of the T4 tail and tail fibers." Virol J **7**: 355.

Leiman, P. G., P. R. Chipman, V. A. Kostyuchenko, V. V. Mesyanzhinov and M. G. Rossmann (2004). "Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host." Cell **118**(4): 419-429.

Levin, M. E., R. W. Hendrix and S. R. Casjens (1993). "A programmed translational frameshift is required for the synthesis of a bacteriophage λ tail assembly protein." Journal of molecular biology **234**(1): 124-139.

Li, X., P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard and Y. Cheng (2013). "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM." Nature methods **10**(6): 584-590.

Liu, J. and A. Mushegian (2004). "Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages." Journal of bacteriology **186**(13): 4369-4375.

Loessner, M. J. and R. Calendar (2006). The Listeria bacteriophages. The bacteriophages, Oxford University Press**:** 593-601.

Lupas, A., M. Van Dyke and J. Stock (1991). "Predicting coiled coils from protein sequences." Science: 1162-1164.

Lynch, K. H., P. Stothard and J. J. Dennis (2010). "Genomic analysis and relatedness of P2-like phages of the Burkholderia cepacia complex." BMC genomics **11**(1): 1-26.

Mahony, J., M. Alqarni, S. Stockdale, S. Spinelli, M. Feyereisen, C. Cambillau and D. V. Sinderen (2016). "Functional and structural dissection of the tape measure protein of lactococcal phage TP901-1." Sci Rep **6**: 36667.

Manning, G. S. (1978). "The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides." Quarterly reviews of biophysics **11**(2): 179-246.

Mastronarde, D. N. (2005). "Automated electron microscope tomography using robust prediction of specimen movements." Journal of structural biology **152**(1): 36-51.

Mastronarde, D. N. and S. R. Held (2017). "Automated tilt series alignment and tomographic reconstruction in IMOD." Journal of structural biology **197**(2): 102-113.

McGreevy, R., I. Teo, A. Singharoy and K. Schulten (2016). "Advances in the molecular dynamics flexible fitting method for cryo-EM modeling." Methods **100**: 50-60.

Mitchell, J. S., J. Glowacki, A. E. Grandchamp, R. S. Manning and J. H. Maddocks (2017). "Sequence-dependent persistence lengths of DNA." Journal of chemical theory and computation **13**(4): 1539-1555.

Molineux, I. J. (2001). "No syringes please, ejection of phage T7 DNA from the virion is enzyme driven." Molecular microbiology **40**(1): 1-8.

Moody, M. (1973). "Sheath of bacteriophage T4: III. Contraction mechanism deduced from partially contracted sheaths." Journal of molecular biology **80**(4): 613-635.

Morais, M. C., K. H. Choi, J. S. Koti, P. R. Chipman, D. L. Anderson and M. G. Rossmann (2005). "Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of ϕ29." Molecular cell **18**(2): 149-159.

Morgan, G. J., G. F. Hatfull, S. Casjens and R. W. Hendrix (2002). "Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus." Journal of molecular biology **317**(3): 337-359.

Newcomer, R. L., J. R. Schrad, E. B. Gilcrease, S. R. Casjens, M. Feig, C. M. Teschke, A. T. Alexandrescu and K. N. Parent (2019). "The phage L capsid decoration protein has a novel OB-fold and an unusual capsid binding strategy." Elife **8**: e45345.

Nováček, J., M. Šiborová, M. Benešík, R. Pantůček, J. Doškař and P. Plevka (2016). "Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate." Proceedings of the National Academy of Sciences **113**(33): 9351-9356.

Oh, B., C. L. Moyer, R. W. Hendrix and R. L. Duda (2014). "The delta domain of the HK97 major capsid protein is essential for assembly." Virology **456**: 171-178.

Petrovski, S., Z. A. Dyson, R. J. Seviour and D. Tillett (2012). "Small but sufficient: the Rhodococcus phage RRH1 has the smallest known Siphoviridae genome at 14.2 kilobases." Journal of virology **86**(1): 358-363.

Pettersen, E. F., T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin (2021). "UCSF ChimeraX: Structure visualization for researchers, educators, and developers." Protein Science **30**(1): 70-82.

Pope, W. H., Pedulla,M.L., Ford,M.E., Peebles,C.L., Hatfull,G.H. and Hendrix,R.W (2011). Bacillus phage G, complete genome.

Prevelige Jr, P. E. (2008). "Send for reinforcements! Conserved binding of capsid decoration proteins." Structure **16**(9): 1292-1293.

Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nat Methods **14**(3): 290-296.

Rader, A., D. H. Vlad and I. Bahar (2005). "Maturation dynamics of bacteriophage HK97 capsid." Structure **13**(3): 413-421.

Rao, V. B. and M. Feiss (2015). "Mechanisms of DNA packaging by large double-stranded DNA viruses." Annual review of virology **2**: 351-378.

Rodríguez-Rubio, L., D. Gutiérrez, B. Martínez, A. Rodríguez, F. Götz and P. García (2012). "The tape measure protein of the Staphylococcus aureus bacteriophage vB_SauS-phiIPLA35 has an active muramidase domain." Applied and environmental microbiology **78**(17): 6369-6371.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck (2012). "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." Systematic biology **61**(3): 539-542.

Sae-Ueng, U., T. Liu, C. E. Catalano, J. B. Huffman, F. L. Homa and A. Evilevitch (2014). "Major capsid reinforcement by a minor protein in herpesviruses and phage." Nucleic acids research **42**(14): 9096-9107.

Sassi, M., C. Bebeacua, M. Drancourt and C. Cambillau (2013). "The first structure of a mycobacteriophage, the Mycobacterium abscessus subsp. bolletii phage Araucaria." Journal of virology **87**(14): 8099-8109.

Sathaliyawala, T., M. Z. Islam, Q. Li, A. Fokine, M. G. Rossmann and V. B. Rao (2010). "Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages." Molecular microbiology **77**(2): 444-455.

Serwer, P., A. Estrada and R. A. Harris (1995). "Video light microscopy of 670-kb DNA in a hanging drop: shape of the envelope of DNA." Biophysical journal **69**(6): 2649-2660.

Serwer, P., S. J. Hayes and R. H. Watson (1992). "Conformation of DNA packaged in bacteriophage T7: analysis by use of ultraviolet light-induced DNA-capsid cross-linking." Journal of molecular biology **223**(4): 999-1011.

Serwer, P., E. T. Wright, Z. Liu and W. Jiang (2014). "Length quantization of DNA partially expelled from heads of a bacteriophage T3 mutant." Virology **456**: 157-170.

Shen, M. R., K. H. Downing, R. Balhorn and N. V. Hud (2000). "Nucleation of DNA condensation by static loops: formation of DNA toroids with reduced dimensions." Journal Of The American Chemical Society **122**(19): 4833-4834.

Steinegger, M., M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger and J. Söding (2019). "HH-suite3 for fast remote homology detection and deep protein annotation." BMC bioinformatics **20**(1): 1-15.

Stone, N. P., B. J. Hilbert, D. Hidalgo, K. T. Halloran, J. Lee, E. J. Sontheimer and B. A. Kelch (2018). "A hyperthermophilic phage decoration protein suggests common evolutionary origin with herpesvirus triplex proteins and an anti-CRISPR protein." Structure **26**(7): 936-947. e933.

Suhanovsky, M. M. and C. M. Teschke (2015). "Nature′ s favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold." Virology **479**: 487-497.

Sun, M. and P. Serwer (1997). "The conformation of DNA packaged in bacteriophage G." Biophysical journal **72**(2): 958-963.

Sutter, M., D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban and N. Ban (2008). "Structural basis of enzyme encapsulation into a bacterial nanocompartment." Nature structural & molecular biology **15**(9): 939-947.

Tang, G., L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees and S. J. Ludtke (2007). "EMAN2: an extensible image processing suite for electron microscopy." Journal of structural biology **157**(1): 38-46.

Tao, Y., N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann and T. S. Baker (1998). "Assembly of a tailed bacterial virus and its genome release studied in three dimensions." Cell **95**(3): 431-437.

Taylor, N. M., N. S. Prokhorov, R. C. Guerrero-Ferreira, M. M. Shneider, C. Browning, K. N. Goldie, H. Stahlberg and P. G. Leiman (2016). "Structure of the T4 baseplate and its function in triggering sheath contraction." Nature **533**(7603): 346-352.

Thomas, J. A., S. C. Hardies, M. Rolando, S. J. Hayes, K. Lieman, C. A. Carroll, S. T. Weintraub and P. Serwer (2007). "Complete genomic sequence and mass spectrometric analysis of highly diverse, atypical Bacillus thuringiensis phage 0305ϕ8–36." Virology **368**(2): 405-421.

Thomas, J. A., A. D. B. Quintana, M. A. Bosch, A. C. De Peña, E. Aguilera, A. Coulibaly, W. Wu, M. V. Osier, A. O. Hudson and S. T. Weintraub (2016). "Identification of essential genes in the Salmonella phage SPN3US reveals novel insights into giant phage head structure and assembly." Journal of virology **90**(22): 10284-10298.

Thomas, J. A., M. R. Rolando, C. A. Carroll, P. S. Shen, D. M. Belnap, S. T. Weintraub, P. Serwer and S. C. Hardies (2008). "Characterization of Pseudomonas chlororaphis myovirus 201ϕ2-1 via genomic sequencing, mass spectrometry, and electron microscopy." Virology **376**(2): 330-338.

Vernhes, E., M. Renouard, B. Gilquin, P. Cuniasse, D. Durand, P. England, S. Hoos, A. Huet, J. F. Conway and A. Glukhov (2017). "High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid." Scientific reports **7**(1): 1-14.

Vianelli, A., G. Wang, M. Gingery, R. Duda, F. Eiserling and E. Goldberg (2000). "Bacteriophage T4 self-assembly: localization of gp3 and its role in determining tail length." Journal of bacteriology **182**(3): 680-688.

Wang, S., W. Li, S. Liu and J. Xu (2016). "RaptorX-Property: a web server for protein structure property prediction." Nucleic acids research **44**(W1): W430-W435.

Wang, Z., S. C. Hardies, A. Fokine, T. Klose, W. Jiang, B. C. Cho and M. G. Rossmann (2018). "Structure of the marine siphovirus TW1: Evolution of capsid-stabilizing proteins and tail spikes." Structure **26**(2): 238-248. e233.

Wikoff, W. R., L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix and J. E. Johnson (2000). "Topologically linked protein rings in the bacteriophage HK97 capsid." Science **289**(5487): 2129-2133.

Wriggers, W. (2012). "Conventions and workflows for using Situs." Acta Crystallographica Section D: Biological Crystallography **68**(4): 344-351.

Xu, J., D. Wang, M. Gui and Y. Xiang (2019). "Structural assembly of the tailed bacteriophage ϕ29." Nature communications **10**(1): 1-16.

Yan, R., S. V. Venkatakrishnan, J. Liu, C. A. Bouman and W. Jiang (2019). "MBIR: A cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information." Journal of structural biology **206**(2): 183-192.

Yap, M. L. and M. G. Rossmann (2014). "Structure and function of bacteriophage T4." Future microbiology **9**(12): 1319-1327.

Yordpratum, U., U. Tattawasart, S. Wongratanacheewin and R. W. Sermswan (2011). "Novel lytic bacteriophages from soil that lyse Burkholderia pseudomallei." FEMS microbiology letters **314**(1): 81-88.

Yuan, S., J. Wang, D. Zhu, N. Wang, Q. Gao, W. Chen, H. Tang, J. Wang, X. Zhang and H. Liu (2018). "Cryo-EM structure of a herpesvirus capsid at 3.1 Å." Science **360**(6384).

Yuan, Y. and M. Gao (2017). "Jumbo Bacteriophages: An Overview." Frontiers in Microbiology **8**(403).

Zheng, W., F. Wang, N. M. Taylor, R. C. Guerrero-Ferreira, P. G. Leiman and E. H. Egelman (2017). "Refined cryo-EM structure of the T4 tail tube: exploring the lowest dose limit." Structure **25**(9): 1436-1441. e1432.

Zinke, M., K. A. Sachowsky, C. Öster, S. Zinn-Justin, R. Ravelli, G. F. Schröder, M. Habeck and A. Lange (2020). "Architecture of the flexible tail tube of bacteriophage SPP1." Nature communications **11**(1): 1-9.