Elucidating the structural mechanisms of capsid stability and assembly using a hyperthermophilic bacteriophage

ELUCIDATING THE STRUCTURAL MECHANISMS OF CAPSID STABILITY AND ASSEMBLY USING A HYPERTHERMOPHILIC BACTERIOPHAGE

A Dissertation Presented

By

NICHOLAS PAUL STONE

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

JULY 9TH, 2019

ELUCIDATING THE STRUCTURAL MECHANISMS OF CAPSID STABILITY AND ASSEMBLY USING A HYPERTHERMOPHILIC BACTERIOPHAGE

A Dissertation Presented

By

NICHOLAS PAUL STONE

This work was undertaken in the Graduate School of Biomedical Sciences

Biochemistry and Molecular Pharmacology

Under the mentorship of

_____

Brian Kelch, Ph.D., Thesis Advisor

_____

Kenan Murphy, Ph.D., Member of Committee

_____

William Royer, Ph.D., Member of Committee

_____

Chen Xu, Ph.D., Member of Committee

_____

Carolyn Teschke, Ph.D., External Member of Committee

_____

Celia Schiffer, Ph.D., Chair of Committee

_____

Mary-Ellen Lane, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

July 9th, 2019

ACKNOWLEDGEMENTS

I would like to take a brief moment to thank my all of my mentors and colleagues for their invaluable support and encouragement throughout my time in graduate school. I have been fortunate to have had Dr. Brian Kelch as my thesis mentor for the past six years. Brian has created an amazing lab environment, and I am grateful to have been a part of its early beginnings. Brian always pushed me to be the best scientist possible, and has always provided incredible support for both my thesis research and my professional development. I would also like to thank the members of the Kelch Lab, Dr. Christl Gaubitz, Janelle Hayes, Joe Magrino, Emily Agnello, Jacob Landeck, and Xingchen Liu for their continued advice, support, and friendship. I am also grateful to former lab members Dr. Brendan Hilbert and Dr. Caroline Duffy for their advice, guidance, and emotional support while traversing my early graduate school years. The advice and support provided additionally by the Schiffer and Royer lab members through our joint group meetings was an integral part of my development into an inquisitive scientist, and your impact on my development will not be forgotten.

Thank you to the faculty members who have served on my Qualifying Exam, TRAC, and Dissertation Defense committees: Dr. Duane Jenness, Dr. Andrei Korostelev, Dr. Kenan Murphy, Dr. William Royer, Dr. Celia Schiffer, Dr. Mohan Somasundaran, Dr. Chen Xu, and Dr. Carolyn Teschke. I thank you for

the significant amount of time you have invested into helping me accomplish my thesis goals, and for helping me evolve into the scientist I am today.

I must also recognize those who have directly impacted my research results by offering technical support. I owe a large thanks to Dr. Brendan Hilbert for teaching me every aspect of x-ray crystallography, and Dr. Gabriel Demo for spending all too much of his time teaching me how to process data and helping me troubleshoot 3D refinement of virus particles. Dr. Kevin Halloran and Daniel Hidalgo also provided immense support to the design and execution of several assays described in this thesis. I must also thank the members of the Cryo-EM Core facility, Dr. Chen Xu, Dr. Kangkang Song, and Dr. Kyounghwan Lee for donating their time and expertise to teaching me how to prepare samples and collecting the data presented in this work.

I would lastly like to thank the students, faculty, and staff in the Department of Biochemistry and Molecular Pharmacology for their continuous support, encouragement, and friendship throughout my tenure at UMass.

ABSTRACT

Nearly all viruses encapsulate their genomes in protective protein shells known as capsids. Capsids self-assemble from repeating protein subunits, which surround the viral genome. Many viruses use a powerful biomotor to pump DNA into preformed capsid shells. Therefore, not only does the capsid protect the genome from environmental stress, it additionally stabilizes against high internal pressure caused by the tightly-packaged genome inside. To understand how capsids remain stable despite extreme conditions, I use thermophilic bacteriophage P74-26 as a model to probe the structural mechanisms that govern capsid assembly and stability. P74-26 capsids have a similar architecture to capsids of mesophilic tailed bacteriophages, allowing direct comparison to elucidate the structural basis of enhanced thermostability. Here I determine the structure of the P74-26 capsid decoration protein, which contains a core beta-barrel domain termed the 'beta-tulip' domain. The beta-tulip domain is conserved in structural proteins from both Herpesviruses and phage, as well as a broad-spectrum Cas9 inhibitor, providing evidence of shared evolutionary ancestry. Additionally, my high-resolution structure of the P74-26 virion capsid reveals unique interdigitated architectural features that contribute to enhanced stability in the thermophile. P74-26 has a significantly larger capsid than related mesophiles yet retains the same icosahedral geometry, demonstrating a novel mechanism for increasing capsid capacity. Furthermore, my thesis work explores capsid assembly and maturation mechanisms *in vitro,* establishing P74-26 as a platform for future development of novel nanoparticles and therapeutic delivery systems. Taken together, this work illuminates the incredible stability of a thermophilic virus and illustrates its utility as a powerful tool for studying viral maturation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF THIRD PARTY COPYRIGHTED MATERIAL

The content of Figure 1.5 has been adapted from a published manuscript (Mannige and Brooks, 2010), and reproduced with permission from the publishing journal, PLoS One.

Citation:

> Mannige RV, Brooks CL. Periodic table of virus capsids: implications for natural selection and design. *PLoS ONE*. 2010.

LIST OF ABBREVIATIONS AND NOMENCLATURE

Double-stranded DNA	**dsDNA**
Major Capsid Protein	**MCP**
Phage P74-26 Decoration Protein	**Dec^{P74-26}; gp87**
Phage Lambda Decoration Protein	**gpD**
Large Terminase	**TerL**
Small Terminase	**TerS**
Herpes Simplex Virus	**HSV**
Human Cytomegalovirus	**HCMV**
Simian Virus 40	**SV40**
Guanidine Hydrochloride	**GdnHCl**
Polyethylene Glycol	**PEG**
2-Mercaptoethanol	**βME**
Triplex Forming Oligonucleotide	**TFO**
Triplex Binding Site	**TBS**

PREFACE

The work in Chapter II of this thesis has been previously published in the journal *Structure* as:

> Stone NP, Hilbert BJ, Hidalgo D, Halloran KT, Kelch BA. A Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex Proteins and an Anti-CRISPR Protein. *Structure*. 2018.

The work in Chapter III of this thesis has been formatted into a manuscript and submitted for peer review.

CHAPTER I

Introduction

Classification and morphology of double-stranded DNA viruses

Viruses are the most numerous biological particles on the planet, far outnumbering their hosts by at least one order of magnitude (Koonin and Dolja, 2013; Suttle, 2007; Wommack and Colwell, 2000). Most viruses use protein shells known as capsids to protect their genomes during transport from one host cell to another. Virus capsids vary widely in size, shape, construction, and encapsulation mechanism, but all capsids play a similar role in stabilizing the viral genome and protecting it from harsh extracellular environments (Mateu, 2013). This thesis focuses on the capsid stability mechanisms employed by double-stranded DNA viruses, in particular that of the tailed bacteriophages.

In the Baltimore virus classification system, group I viruses are characterized by the presence of a double-stranded DNA (dsDNA) genome (Baltimore, 1971). The dsDNA virus group features enveloped and non-enveloped viruses and includes human pathogens (i.e. Adenoviruses, Poxviruses, and Herpesviruses), and most bacterial viruses (known as bacteriophages). One family of bacteriophages is the Caudoviruses, or tailed bacteriophages, which are non-enveloped dsDNA viruses featuring an icosahedral capsid and a flexible tail structure used for host cellular attachment and infection (Ackermann, 1998). The order Caudoviridae currently includes more than 350 unique characterized species from a diverse array of hosts (International Committee on Taxonomy of Viruses: talk.ictvonline.org/taxonomy; Lefkowitz et al., 2018). Caudoviruses are classified into three viral families based

on their unique tail morphologies: Podoviridae, Myoviridae, and Siphoviridae (Ackermann, 1998). Podoviruses have short, non-contractile tails that typically assemble onto the capsid shell. The well-characterized Podovirus P22 features a short tail tube with tail spike proteins necessary for host attachment and DNA ejection (Hartwieg et al., 1986; Tang and Marion, 2005). Myoviruses such as bacteriophage T4 feature long contractile tails consisting of an inner tail tube, and outer contractile tail sheath (Aksyuk et al., 2011; Yap and Rossmann, 2014). Siphoviruses, which comprise more than half of all Caudoviruses, feature long non-contractile tails with significant flexibility (Davidson et al., 2012). Siphoviruses feature a tail tube structure similar to Myoviruses that is preassembled prior to capsid attachment, but lack an outer contractile tail sheath.

The dsDNA Herpesviruses, which are closely related to Caudoviruses, infect animal hosts and includes many human pathogens such as Herpes Simplex Virus 1 and 2, Human Cytomegalovirus, Epstein-Barr virus, and Kaposi's Sarcoma-associated Herpesvirus (Davison, 2007a). Herpesviruses feature longer genomes than most tailed phages and accordingly have capsids with a larger interior capacity. Herpesvirus genomes range from 125-250 kb in length, which are packaged into preformed capsid shells in a similar manner to Caudoviruses (Davison, 2007b; Honess, 1984). Herpesviruses and tailed phages have homologous icosahedral capsid assemblies, with significant structural conservation of core capsid proteins (Veesler and Cambillau, 2011; Veesler and Johnson, 2013). In contrast to tailed phage however, Herpesvirus virions feature

an additional outer tegument layer and envelope required for infection that surround the icosahedral capsid (Kelly et al., 2009; Mettenleiter, 2002; Yu et al., 2011).

All dsDNA viruses package their genomes into capsid shells, generating significant pressures within the capsid (Earnshaw and Casjens, 1980; Evilevitch et al., 2004). The internal pressure caused by the tightly-packed genome necessitates significant stabilization of dsDNA virus capsid shells. In this thesis, I seek to probe the structural mechanisms of capsid stability and assembly using a thermophilic Caudovirus to elucidate the molecular underpinnings of the viral maturation mechanism.

Viral maturation in dsDNA viruses

Maturation is a critical stage in the life cycle of all viruses that leads to the production of mature, infectious virus particles (Delgui and Rodríguez, 2013; Mateu, 2013). The viral maturation process in dsDNA viruses occurs in three main stages: 1) assembly of a preformed capsid shell known as the procapsid, 2) ATP-dependent genome packaging, and 3) capsid expansion and final virion assembly (Figure 1.1) (Steven et al., 2005; Veesler and Johnson, 2012). First, a small cohort of essential structural proteins undergo a complex self-assembly reaction to generate a relaxed, spherical protein shell known as the procapsid. Following assembly, a powerful ATPase motor packages the viral genome into the procapsid with high efficiency (Johnson and Chiu, 2007; Rao and Feiss, 2008). Finally, the pressure generated by the genome packaging reaction leads to capsid expansion from the spherical procapsid state into an expanded, isometric capsid. Once packaging is completed, final assembly of the mature infectious virion occurs, typically including the attachment of a tail structure required for host infection. The events of viral maturation are highly coordinated, and lead to the rapid assembly of a homogeneous population of infectious virions.

Viral maturation is an incredibly dynamic process. Capsid expansion causes significant rearrangement of the capsid subunits, requiring a large amount of flexibility to accommodate the incoming genome (Dokland and Murialdo, 1993; Gertsman et al., 2009; Prasad et al., 1993). Despite this inherent

flexibility the capsid subunits must also stabilize the capsid against high internal pressure caused by the packaging reaction, retaining rigidity in the mature virion. This thesis seeks to determine the structural mechanisms of capsid stability and assembly in a novel thermophilic bacteriophage model system. Using the hyperthermophilic bacteriophage P74-26 (Minakhin et al., 2008; Yu et al., 2006) – a close relative of well-characterized mesophilic Caudoviruses – we can directly compare thermophilic and mesophilic virions and dissect the structural and biochemical mechanisms of capsid thermostability.

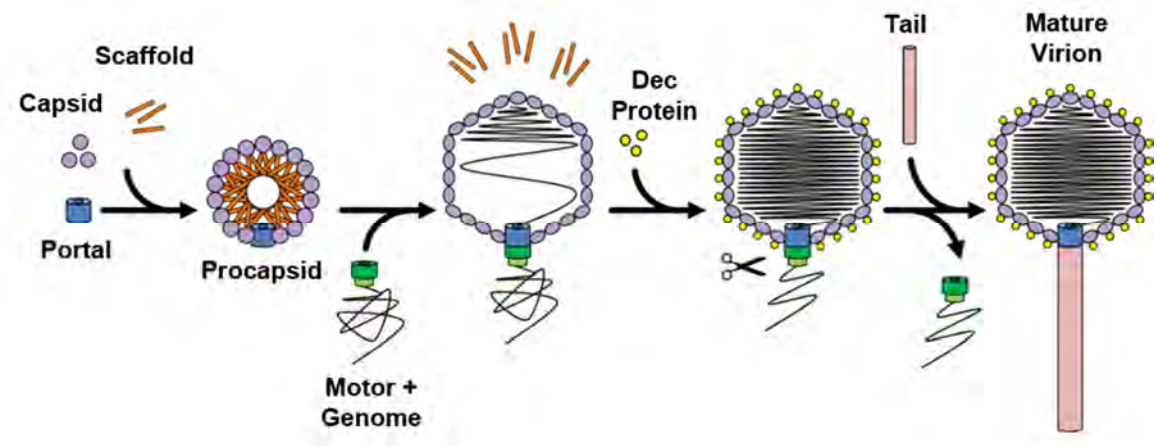

Figure 1.1. Viral maturation in dsDNA viruses. Maturation begins with the assembly of the procapsid from repeating subunits of one or more capsid coat proteins, a scaffolding protein, and portal protein. Then, the packaging motor complexed with genome binds to the capsid portal vertex and DNA is pumped into the capsid in an ATP-dependent manner. Loss of the scaffolding protein causes capsid expansion from a spherical procapsid to isometric capsid. In some dsDNA viruses, expansion exposes binding sites for decoration proteins or other auxiliary stabilizing proteins. Upon completion of packaging the packaging motor and genome dissociate from the capsid and final assembly of the mature, infectious virion occurs.

Procapsid assembly

In most tailed phages, procapsids are typically assembled from repeated copies of a few essential structural proteins; a Major Capsid Protein that forms the bulk of the capsid shell, a scaffolding protein or domain to aid in capsid assembly, and the portal protein, which forms a pore for genome entry and exit. These three protein subunits are the minimal requirements for assembly of dsDNA virus procapsids, though in certain families such as Herpesviruses additional components are typically required for proper assembly (Cardone et al., Yang and Baines, 2008).

The Major Capsid Protein

The Major Capsid Protein (MCP) or coat protein forms the majority of the capsid shell in dsDNA viruses. The viral capsid shell is constructed from pentameric and hexameric oligomers of MCP or equivalent coat proteins termed pentons and hexons, respectively – collectively known as capsomers (Caspar and Klug, 1962). In the isometric assembly, pentons are situated at 11 of 12 icosahedral vertices in the capsid, while the portal occupies the final icosahedral vertex (Hendrix, 1978; Parent et al., 2018). Hexons comprise the remaining surface area of the capsid shell (Figure 1.2A,B), and the number of hexons incorporated in the icosahedron typically varies in correlation with capsid size (Prasad and Schmid, 2012).

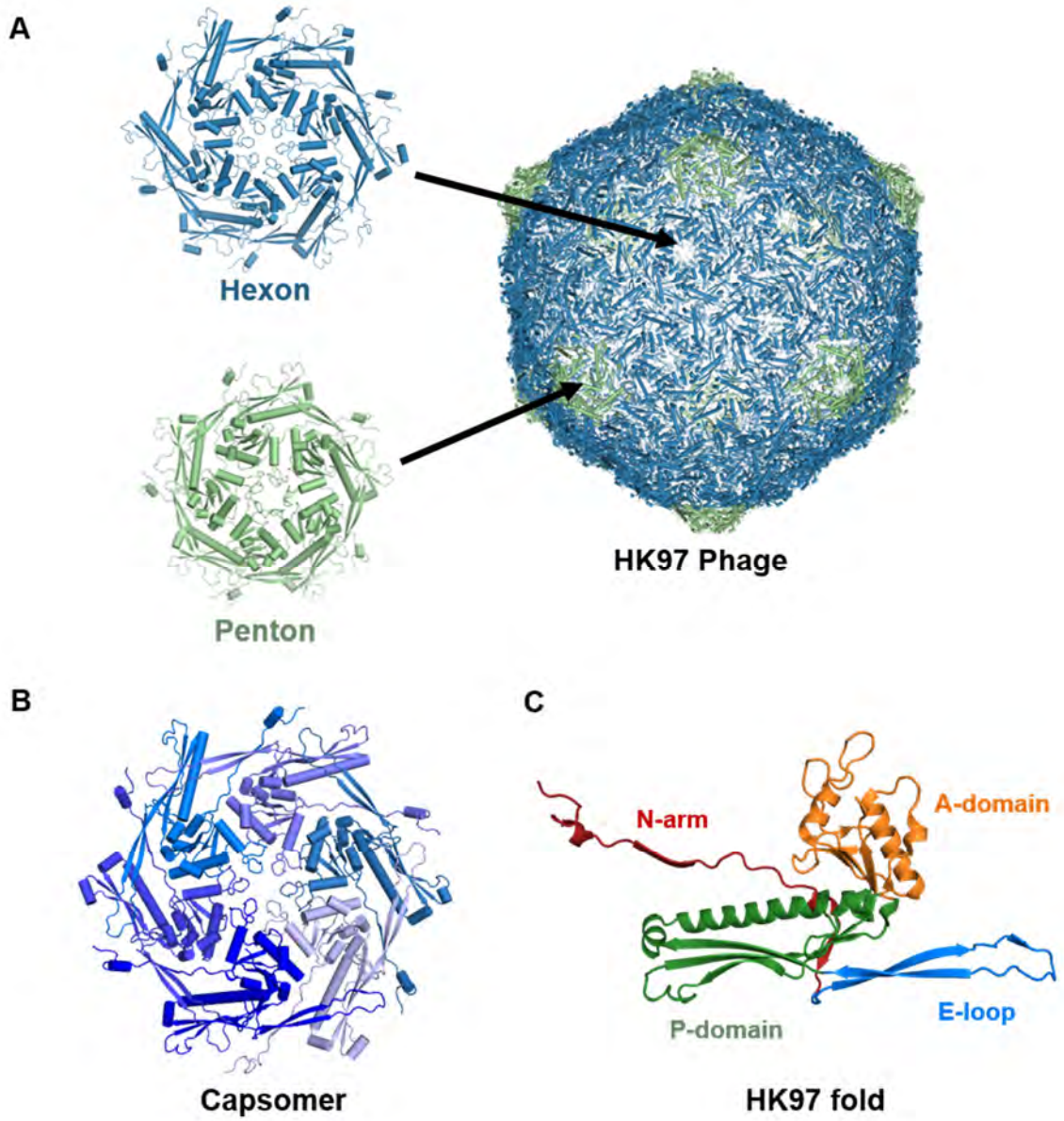

Figure 1.2. The HK97 fold is conserved in many dsDNA virus Major Capsid Proteins.
A) Orientation of HK97 hexons (blue) and pentons (green) within the icosahedral capsid assembly. Penton subunits form the fivefold icosahedral vertices, while the hexons coat the remaining surface of the capsid shell.
B) Capsomer schematic showing the organization of MCP monomers in a single hexon from phage HK97.
C) Schematic of the HK97 Major Capsid Protein colored by core fold domain, highlighting the common features of the HK97 fold conserved in most tailed phages (PDB: 1OHG).

Structure elucidation of dsDNA virus capsids reveals that MCPs from a diverse array of viruses often share common architectures, and these viruses can be sorted into defined ancestral lineages based on the overall fold of MCP (Krupovic and Bamford, 2011). Papillomavirus and Polyomavirus MCPs have a 'Jelly-roll fold' architecture formed by two adjacent 4-stranded β-sheets arranged in a barrel-like conformation (Chen et al., 2000; Stehle et al., 1996). While present in these two families of dsDNA viruses, the Jelly-roll fold is much more prevalent amongst viruses with RNA genomes (Krupovic and Bamford, 2011; Rossmann and Johnson, 1989). dsDNA viruses from nine different families including Adenoviruses and Poxviruses contain a 'double Jelly-roll fold' that consists of two β-barrel Jelly-roll domains joined by a linker region. Unlike the single Jelly-roll fold, the double Jelly-roll fold is found only in dsDNA viruses (Khayat and Johnson, 2011; Roberts et al., 1986; San martin and Van, 2018).

One of the most common MCP architectural lineages in dsDNA viruses is the 'HK97 fold'. The HK97 fold is conserved in viruses that infect all domains of life, and is found in both Caudoviruses and Herpesviruses (Krupovic and Bamford, 2011; Suhanovsky and Teschke, 2015). The HK97 fold is named after bacteriophage HK97, the first virus of this lineage to be structurally characterized (Gertsman et al., 2009; Wikoff et al., 2000). HK97 folds from these viruses share several conserved domains and architectural features, principal among them the P-domain, A-domain, N-arm, and E-loop (Figure 1.2C). These core domains are conserved in all HK97-like viruses (Figure 1.3).

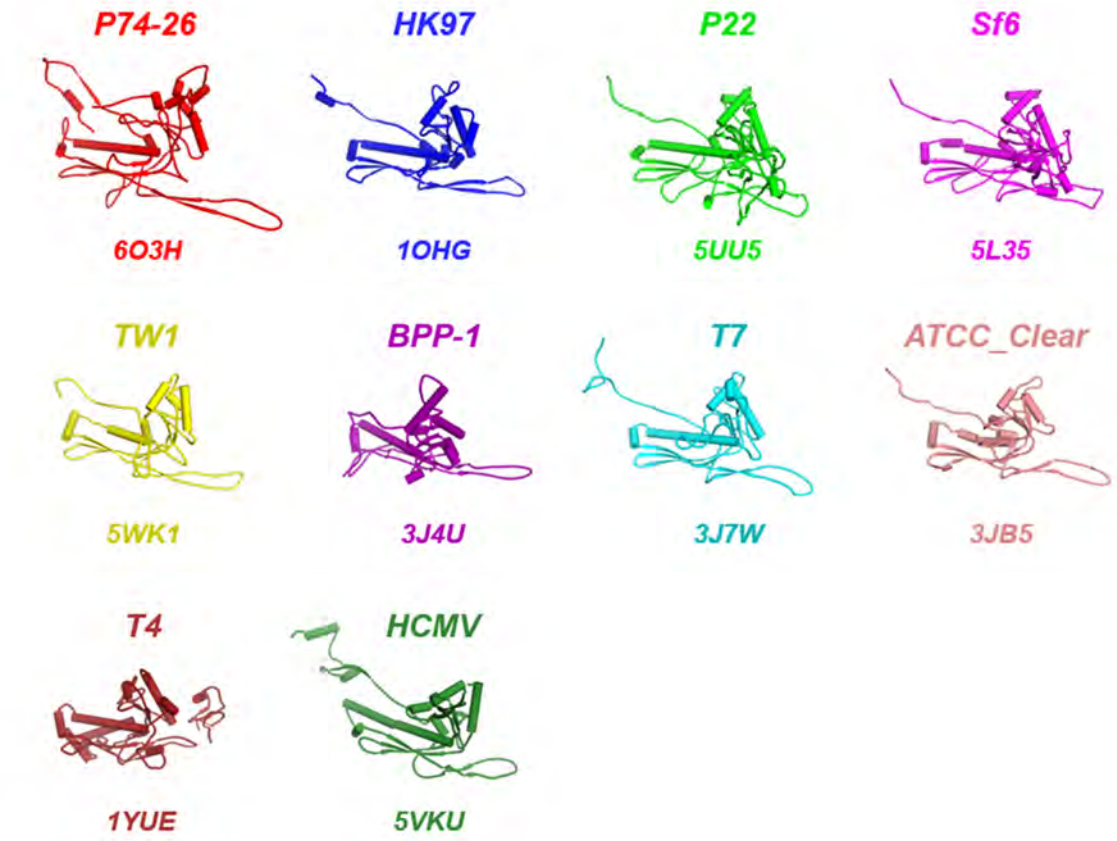

Figure 1.3. Conservation of Major Capsid Protein HK97 folds. HK97-folds of MCPs from several Caudoviruses, and the conserved HK97-fold domain of Human Cytomegalovirus (HCMV) MCP. PDB accession codes are listed below each model.

The P-domain forms one of two main globular regions of the HK97 fold. The P-domain consists of a long 3-4 stranded β-sheet and a long α-helix termed the ‘spine helix’ that spans the length of the domain. The A-domain forms the other core globular domain of the HK97 fold, and includes a 5-stranded β-sheet that forms a ‘β-hinge’ between the A- and P-domains (Gertsman et al., 2009; Teschke and Parent, 2010; Wikoff et al., 2000). Together, these two globular domains form the bulk of the HK97 fold surface and form the core characteristic ‘L’-shape. Within the capsomer assembly, the HK97 fold A-domains are situated toward the center, with the P-domains of the HK97 fold emanating radially outward toward the capsomer edges (Figure 1.2B).

The additional architectural features of the HK97 fold in general are loops and extensions that emanate from these core globular domains. The E-loop forms a long β-hairpin that extends outward from the core HK97 fold. The E-loop forms extensive contacts with an adjacent MCP subunits, forming critical interactions with the N-arm and/or P-domain from a neighboring MCP subunit stabilizing the capsid shell (Lander et al., 2008; Suhanovsky and Teschke, 2015; Wikoff et al., 2000). In many HK97-like viruses the N-arm of the HK97 fold plays a role in stabilizing the capsid shell. The N-arm is often oriented toward the capsid interior, but in some viruses adopts alternate conformations to form stabilizing interactions with neighboring intra-capsomer MCP subunits (Lander et al., 2008; Liu et al., 2010), adjacent capsomers (Bayfield et al., 2019; Guo et al.,

2014; Zhao et al., 2017), or decoration proteins on the capsid surface (Bayfield et al., 2019; Zhang et al., 2013).

Additional features found in several HK97-like viruses include the G-loop – a glycine rich extension from the P-domain that interacts with the E-loop to stabilize capsomers (Bayfield et al., 2019; Tso et al., 2014), and the F-loop – a small insertion loop between the N-arm and E-loop that further stabilizes the threefold capsid axes (Liu et al., 2010). A number of dsDNA viruses have additional insertion domains, often called I-domains, within the A-domain or E-loop that are essential for procapsid assembly and particle stability (Rizzo et al., 2014; Fokine et al., 2005; Zhao et al., 2017).

Herpesvirus MCPs also contain the conserved HK97 fold domain, though the architecture of MCP is significantly more complex than that seen in Caudoviruses. Herpesvirus MCPs contain both a ‘floor’ and ‘tower’ region; the floor region includes the conserved HK97 fold and three additional extension domains, while the tower region includes an additional three domains required for assembly that are unique to the Herpesviruses (Dai et al., 2018; Dai and Zhou, 2018; Yu et al., 2017).

Portal protein

The portal protein forms a dodecameric ring assembly with a central pore that is used for genome entry and exit in the viral capsid (Bazinet and King, 1988; Driedonks et al., 1981; Valpuesta and Carrascosa, 1994; Newcomb et al., 2001).

The portal ring occupies one of the 12 fivefold icosahedral vertices of the capsid and serves as the primary attachment site for the genome packaging motor and later the tail assembly, forming a symmetry mismatch between the capsid shell and the neck/tail region of the virion (Hendrix, 1978; Parent et al., 2018). The portal dodecamer plays a key role in multiple steps of viral maturation including nucleation of procapsid assembly, genome packaging and termination, and final assembly of the mature virion.

Portal nucleates procapsid assembly by recruiting scaffolding proteins and MCP to a single dodecameric ring, leading to only one point of entry and egress in each assembled procapsid (Motwani and Teschke, 2019; Newcomb et al., 2005). During genome packaging, the portal ring complexes with the genome packaging motor, forming a central pore through which genomic DNA enters the procapsid (Black, 1989; Earnshaw and Casjens, 1980). Previous studies in phage P22 suggest that portal also plays a role in termination of packaging, as portal undergoes conformational changes concurrent with conformational switching of capsid expansion. This conformational switching has been suggested to cause a change in binding affinity between the DNA packaging motor and portal, leading to motor dissociation upon packaging completion (Casjens et al., 1992; Lokareddy et al., 2017).

Scaffolding protein

Proper procapsid assembly often requires an interior or exterior scaffolding component that is not present in the mature virion (Casjens and King, 1974; Dokland, 1999; Prevelige and Fane 2012). Without an interior scaffolding protein, MCPs in some Caudoviruses assemble into aberrant structures (including 'polyhead' structures and capsid filaments) or smaller isometric heads constructed from fewer MCP subunits (Choi et al., 2006; Earnshaw and King, 1978; Howatson and Kemp, 1975; Thuman-Commike et al., 1998).

Internal Scaffolding proteins in dsDNA viruses occupy a large amount of the procapsid interior, necessitating removal to allow entry of the viral genome during maturation. Many Caudoviruses and Herpesviruses use a single interior scaffolding protein to direct capsid assembly, which is either degraded by a maturational protease (Liu and Roizman, 1991; Prevelige and Fane, 2012; Thomsen 1995) or ejected from the capsid shell during or after genome packaging (Prasad et al., 1993; Chen et al., 2011). Phages HK97 and T5 have a unique scaffolding domain (termed the delta-domain in HK97) within the MCP that facilitates procapsid assembly in a similar manner to independent scaffolding proteins, wherein proteolytic cleavage of the scaffolding domain leads to capsid expansion (Huet et al., 2010; Gertsman et al., 2009; Huang et al., 2011).

Structural characterization of the interior scaffolding protein from phage Φ29 showed it is primarily α-helical, consisting of a long coiled-coil formed by two

α-helices with adjacent helix-turn-helix motifs at one end that form an arrowhead-like capsid binding domain (Morais et al., 2003). Previous work with scaffolding proteins from phage P22 and 80α additionally determined that the capsid binding region of the Scaffolding protein forms electrostatic interactions with the MCP capsomer interior surface, thus facilitating recruitment of capsomers to the procapsid assembly (Cortines et al., 2011; Dearborn et al., 2017).

Genome packaging in dsDNA viruses

All viruses encapsulate their genomes in capsid protein shells. Many viruses with small genomes encapsulate their genomes during capsid assembly, but incorporation in such a manner is unsuitable for viruses with complex genomes. dsDNA viruses have large genomes that predominantly exceed 15 kb in length (Hatful, 2008). As such, most dsDNA viruses use a powerful ATPase motor known as the terminase to package the genome into preformed procapsids (Johnson and Chiu, 2007; Black, 1989; Rao and Feiss, 2008). The terminase motor complex consists of both a small and large terminase (TerS and TerL, respectively). TerS is responsible for recognition of genomic DNA (Al-zahrani et al., 2009; Catalano et al., 1995), while TerL houses both ATP-hydrolysis and nuclease functionality of the terminase motor required for translocation and termination of packaging (Rao and Feiss, 2008).

Genome packaging reaction

The genome packaging reaction begins with the recognition of replicated genomic DNA by the terminase motor (Black, 1989; Catalano, 1995). The viral genome is typically replicated as end-to-end concatemers as a byproduct of rolling-circle amplification, which are used as the packaging substrate (Casjens and Gilcrease, 2009). The terminase complex along with the genome concatemer then assembles onto the unique portal vertex of the procapsid to facilitate genome packaging and DNA is then pumped into the procapsid in an ATP-dependent manner (Catalano et al., 1995; Hsiao and Black, 1977; Fujisawa et al., 1991; Yeo and Feiss, 1995). In many dsDNA viruses including phage P22, the incoming DNA begins to generate interior capsid pressure and causes the ejection of the scaffolding protein through pores in the capsid shell. The loss of the scaffolding protein then leads to expansion into an expanded isometric particle (Prasad et al., 1993). The genome packaging reaction generates a significant amount of pressure at the capsid interior, compressing the DNA within the capsid to near-crystalline density upon completion of packaging (Earnshaw and Casjens, 1980; Evilevitch et al., 2004). The compression of genomic DNA causes a significant buildup of pressure inside the capsid that can reach 60 atmospheres (Purohit et al., 2005; Smith et al., 2001). To compete with the increasing capsid pressure, the powerful terminase motor is capable of packaging against forces nearing 60 pN (Fuller et al., 2007; Smith et al., 2001).

Packaging mechanisms and termination

In dsDNA viruses, there are two types of genome packaging: unit-length packaging and headful packaging. In ‘unit-length’ packaging viruses such as phage Lambda, the terminase motor cleaves at specific cut sites before and after genome packaging, resulting in exactly one complete genome from the concatemeric DNA packaging into the capsid (Catalano, 1995; Catalano 2005; Fujisawa and Morita, 1997). In contrast, terminases of ‘headful’ packaging viruses such as phage P22 package the concatemeric genome nonspecifically, completely filling the capsid shell (Black, 1989; Coren et al., 1995; Streisinger et al., 1967; Casjens et al., 1988). This leads to inclusion of >1 genome length of DNA packaged in the capsid. This redundancy of headful packaging creates a circularly permuted genome that is suitable for rolling-circle amplification in the host cell (Casjens and Gilcrease, 2009). Upon packaging completion, the large terminase cleaves the concatemeric DNA and dissociates from the portal vertex (Bhattacharyya and Rao, 1993). Dissociation of motor and genome leaves the portal vertex exposed for addition of tail proteins and final virion assembly.

Capsid expansion and mature virion assembly

The genome packaging reaction and loss of interior Scaffolding proteins causes significant rearrangement of the capsid subunits in dsDNA viruses. In some viruses such as HK97, proteolytic cleavage of the interior scaffold precedes the genome packaging reaction (Conway et al., 1995; Duda et al., 1995). In Caudoviruses and Herpesviruses with an independent Scaffolding protein, loss of the interior scaffold and capsid expansion occurs concurrent with the genome packaging reaction (Prevelige and Fane, 2012). Loss of the Scaffolding protein either from proteolytic cleavage or pressure from the incoming genome leads to an approximate 10% expansion in capsid size, transitioning from a spherical procapsid to an isometric virion (Dokland and Murialdo, 1993; Fokine and Rossmann, 2014; Prasad et al., 1993). Expansion of the capsid yields flatter MCP capsomers, forming an expanded rigid particle with an overall thinner capsid shell (Figure 1.4) (Chen et al., 2011; Gertsman et al., 2009; Preux et al., 2013; Wikoff et al., 2000). In several viruses, expansion of the capsid shell also reveals binding sites for additional auxiliary proteins, such as decoration or cementing proteins found in Caudoviruses, which bind preferentially to the outer surface of the expanded capsid shell (Lander et al., 2008; Tang et al., 2006; Yang et al., 2000).

After capsid expansion and completion of genome packaging in most Caudoviruses, a preassembled tail structure required for infection of the host cell is attached to the Portal vertex (Ackermann, 1998). The tail includes a baseplate

or tip necessary for cellular attachment in Myoviruses and Siphoviruses (Dowah and Clokie, 2018; Taylor et al., 2016) and a tail tube composed of repeated major tail protein subunits (Pell et al., 2009). Myovirus tails are contractile, and include an additional outer tail sheath that contracts during cell attachment and injection (Leiman et al., 2004; Taylor et al., 2016). Many Caudoviruses also possess a tail Tape Measure Protein, which dictates the overall length of the tails leading to uniform tail assembly (Xu et al., 2014). Rather than attachment of a preassembled tail fiber, Podoviruses such as P22 assemble a small injection complex at the portal vertex that includes the addition of tail spike proteins required for cellular attachment (Berget and Poteete, 1980; Tang et al., 2005). In dsDNA viruses, this final assembly and attachment of the tail structure to the DNA-filled capsid shell generates the mature, infectious virion.

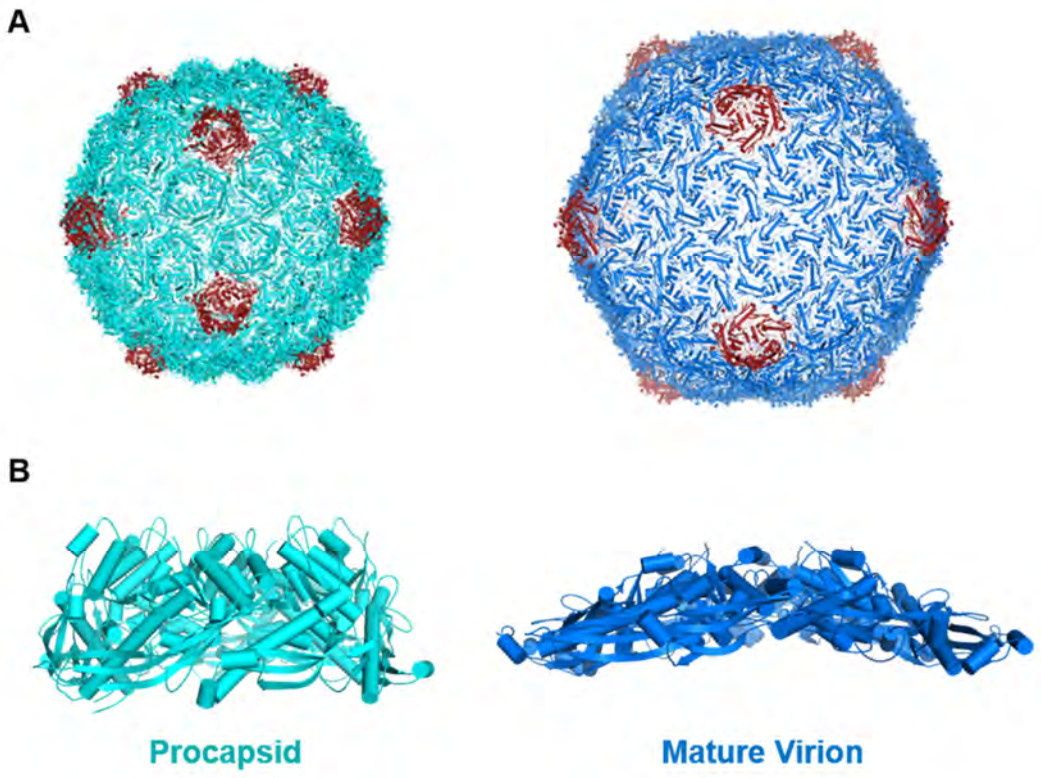

Figure 1.4. Capsid maturational expansion causes thinning of the capsid shell.
A) Morphological comparison of phage HK97 procapsid (cyan) and mature virion (blue) with pentons colored in red.
B) Capsid expansion causes thinning of the capsid wall by adopting a 'flattened' capsomers in the mature virion conformation. Procapsid and mature virion hexons are shown in cyan and blue, respectively.

Capsid architecture and assembly in icosahedral viruses

Icosahedral virus capsids vary widely in size, ranging from small viruses that house only a few essential genes, to the aptly-named giant viruses with genomes exceeding 1 Mb in length (Colson et al., 2013). Icosahedral capsids are formed from multiple copies of one or more coat proteins that assemble into capsomers – both pentons and hexons – that are assembled into the complete capsid shell. Icosahedral capsid size is typically modulated by the incorporation of an increasing amount of capsid subunits during assembly, leading to capsids with a larger interior volume (Llorente et al., 2014; Mannige and Brooks, 2009; 2010; Prasad and Schmid, 2012). Despite large differences in capsid size, a single set of core principles describes the geometric assembly of most icosahedral virus capsids.

Icosahedral capsids contain fivefold, threefold, and twofold rotational axes within the overall assembly. The fivefold axes comprise the 12 total icosahedral vertices of the capsid shell (including the portal vertex), while the two- and threefold axes are formed at the junctions between two or three capsomers (Prasad and Schmid, 2012). The capsomer interaction sites are considered to be quasi-twofold or threefold when situated directly adjacent to a penton. Between the 12 fivefold vertices of the capsid sit 20 total icosahedral faces that vary with the size of the capsid. Thus, most icosahedral viruses have 12 penton subunits that form the icosahedral vertices and a variable number of hexon subunits that dictate the overall size of the capsid shell (Llorente et al., 2014; Mannige and

Brooks, 2009; 2010). According to the principles of 'quasi-equivalence theory' first proposed by Caspar and Klug, icosahedral capsids of any size are constructed from 60 total asymmetric units by varying the arrangement of capsomers and total number of coat protein subunits incorporated in the assembly (Caspar and Klug, 1962). For the construction of a perfect icosahedron the total number of coat protein subunits required for assembly can be denoted as 60*T, where T is defined as the triangulation number.

Since all icosahedral viruses have 12 fivefold vertices (11 pentons and a portal) regardless of capsid size, the distance between two pentons – and the arrangement of hexons in the icosahedral faces – is used as a standardized measurement of capsid size. Starting from one penton, the number of hexon 'steps' required to reach the nearest neighboring penton can be expressed as steps in one direction, *h*, and a second direction 60˚ off-axis, *k,* where $h \geq 1$ and $k \geq 0$ (Figure 1.5A). Using these two stepwise measurements, capsid triangulation number is defined by the equation: $T = h^2 + h*k + k^2$ (Caspar and Klug, 1962). The simplest icosahedral viruses are constructed from 60 total coat protein subunits and have T=1 geometry. These capsids contain only the 12 pentons at the icosahedral vertices with no hexons on the capsid surface (Mannige and Brooks, 2009). Thus for most icosahedral viruses T>1, addition of hexons, rather than significant modulation of the MCP conformation, increases the capsid surface area and interior volume. This causes viruses of a given T-number to

adopt discrete capsid sizes (Mannige and Brooks, 2009). A notable exception to this mechanism of increasing capsid capacity is discussed in Chapter III.

In order to retain icosahedral geometry, capsomers within the capsid surface are often modulated into specific conformations. While pentons situated at the icosahedral vertices adopt only a single conformation, the subunit-subunit dihedral angles of hexons varies with capsomer position and particle size (Mannige and Brooks, 2009; 2010). Many Caudoviruses have T=7 symmetry, constructed from 11 pentons and 60 total hexons (Figure 1.5B). In this arrangement, each hexon contacts a single penton and all 60 hexons required for assembly are considered to be situated in near-equivalent chemical environments (Caspar and Klug, 1962). In viruses T=3, 4, and 7, only a single hexon conformation is required for assembly. In larger icosahedral viruses, proximity to the fivefold vertices and the edges of icosahedral faces leads to two or more hexon conformations within the capsid assembly (Mannige and Brooks, 2009; Zandi et al., 2004). For example, Herpesviruses have T=16 geometry with 150 total hexons (Figure 1.5C). Thus, the assembly includes two different hexon conformations; bent or ‘winged’ hexons situated at the edges of the icosahedral faces, and ‘flat’ hexons in the center of the icosahedral face (Mannige and Brooks 2010). In Chapter III, I propose that both capsid size (i.e. T-number) and hexon conformational heterogeneity have a significant impact on capsid stability by introducing increased complexity and thus more potential weak points to the capsid assembly.

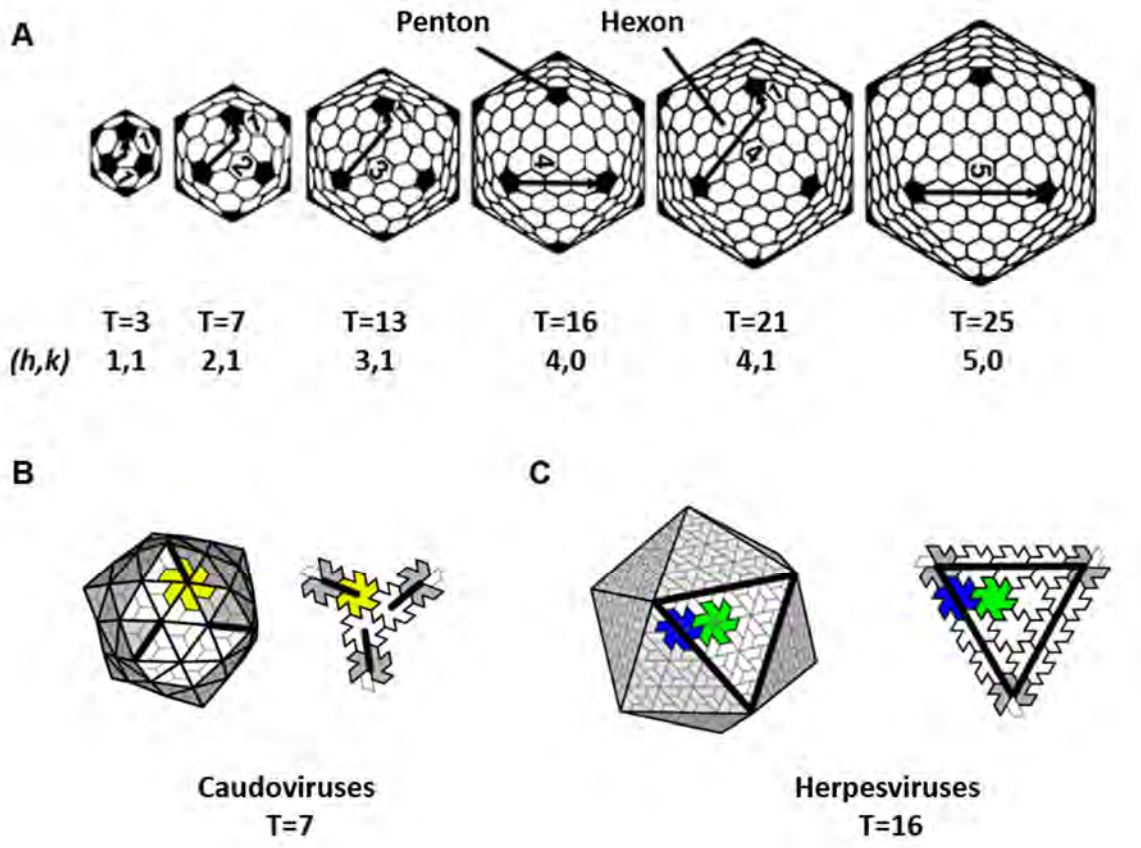

Figure 1.5. Capsid capacity and relationship to icosahedral geometry.
A) In most dsDNA viruses, capsid size is regulated by addition of hexon subunits to the icosahedral assembly. Icosahedral viruses are classified by Triangulation number (T), calculated by the overall distance in steps *h* and *k* between icosahedral vertices.
B) Schematic representation of the T=7 icosahedron commonly found in Caudoviruses. Right, icosahedral face highlighting a single hexon (yellow). All hexons in the T=7 assembly share the same conformation and are considered equivalent.
C) In T>7 viruses including T=16 Herpesviruses, multiple conformations of hexons (blue and green) are required to produce the complete assembly.
Figure adapted from Mannige RV, Brooks CL III (2010) Periodic Table of Virus Capsids: Implications for Natural Selection and Design. PLOS ONE.

Structure determination of virus capsids

There are two main methods that historically have been essential for the elucidation of virus and capsid structures: Electron microscopy (EM) and X-ray crystallography. Following the introduction and early adoption of EM in the 1930's and 1940's, biologists quickly realized the utility of EM for the characterization of virus morphology and diagnosis of viral pathogens (Nagler and Rake, 1948; Ruska, 1987). The development of negative-staining EM provided a rapid method for preparing microscopy grids using aqueous samples and heavy metal stain to generate high-contrast images. Negative-staining EM was useful for the early classification of viruses and phage, providing some of the earliest high-quality images of virions and phages (Ackermann and Prangishvili, 2010; Brenner et al., 1959a, 1959b; Hall, 1955; Huxley and Zubay, 1961; Tikhonenko, 1970).

The first atomic-resolution structures of virus capsids were determined by X-ray crystallography (Abad-Zapatero et al., 1980; Harrison et al., 1978). To date, more than 400 virus capsid structures have been determined by crystallography. It is particularly difficult to elucidate structures of large capsids and viruses with irregular symmetry elements. Thus, dsDNA viruses and tailed phages in particular are largely underrepresented by crystallographic structures (Jiang and Tang, 2017; http://viperdb.scripps.edu/xray.php). Wikoff and colleagues were able to elucidate the 3.4 Å structure of the phage HK97 capsid, a significant undertaking, only using virus-like particle assemblies of the Major

Capsid Protein (Wikoff et al., 2000). Several other virus structures determined using crystallography likewise used surrogate virus-like particles or mutagenesis to facilitate crystallization (Burmeister et al., 2015; Ford et al., 2013; Guu et al., 2009).

The development of cryo-electron microscopy (cryo-EM) was a significant leap forward for structure determination by eliminating the need for heavy-metal staining or fixing of specimens to prevent electron beam-induced damage of samples. Cryo-EM grids are flash frozen in liquid ethane to produce a thin layer of vitreous (non-crystalline) ice to immobilize particles in their native state (Dubochet 1988). Prepared as such, particles adopt multiple orientations within the vitreous ice, facilitating 3-dimensional computational reconstruction (Chang et al., 2015). Furthermore, sample preparation without fixative or the need for crystallization allows for the structural characterization of virions and capsid intermediate structures in their native states with minimal manipulation. Despite the success of cryo-EM in analyzing virus capsid structure, resolution was still largely limited preventing high resolution structure determination. Early cryo-EM reconstructions were often paired with atomic structures of capsid components solved by crystallography to generate pseudoatomic models of complete capsid assemblies (Baker et al., 1999).

High-resolution cryo-EM and virus structure

In recent years, the introduction of Direct Electron Detection Devices has made single-particle cryo-EM a powerful technique for determining structures of macromolecules at near-atomic resolution (Faruqi and McMullan, 2011; Jin et al., 2008; Milazzo et al., 2010; 2011; Mooney et al., 2011). The emergence of direct detectors offers significant advantages over previous data collection methods – traditional photographic film exposure and digital collection using Charge-Coupled Detectors (CCD). CCD detectors use an indirect scintillation counting method to convert electron dosage counts into a digital micrograph. The development of CCD cameras did introduce high-throughput digital data collection methods and on-the-fly image analysis (Carragher et al., 2000; Nakamura et al., 2010; Zhang et al., 2003), but photographic film was often preferred due to the higher Detective Quantum Efficiency (DQE) and thus a higher signal-to-noise ratio compared to CCD output (McMullan et al., 2009; Bammes et al., 2011). Direct electron detectors allow for direct exposure of the incident electron beam, eliminating the need for digital conversion as in the case of CCD detectors, while achieving a higher DQE than photographic film (Wu et al., 2016). Additionally, direct electron detectors advanced high-throughput data collection and the rapid collection of limited dose movie frames for beam-induced motion correction (Campbell et al., 2012; Li et al., 2013). Direct electron detectors paired with high-powered electron microscopes and the rapid

development of single-particle reconstruction software packages in recent years has led to the so-named ‘Resolution Revolution’ in cryo-EM (Kühlbrandt, 2014).

The resolution revolution unsurprisingly had a significant impact on capsid structure determination, improving both data collection and processing. Cryo-EM grid sample preparations can be significantly less cumbersome than crystallization and the time from sample preparation to 3D reconstruction is greatly reduced. To date nearly 400 capsid structures have been solved using cryo-EM methods. In contrast to crystallographic structures, cryo-EM reconstructions sample a wider variety of capsid sizes, expanding well beyond sizes practical for crystallographic methods (viperdb.scripps.edu/cryoem.php). Given the inherent highly-ordered symmetry of icosahedral virus capsids, dsDNA viruses are ideal candidates for structure determination by cryo-EM. During data processing and 3D reconstruction, 60-fold icosahedral symmetry can be imposed on the refinement procedure to significantly improve resolution (Guo and Jiang, 2015; Jiang and Tang, 2018). Icosahedral refinement methods paired with improvements in direct detector technology has led to the elucidation of several atomic resolution capsid structures of dsDNA viruses, which have continued to advance our understanding of capsid assembly and maturation (Table 1.1). With the rapid advancement of direct detectors and improved data processing techniques, capsid reconstructions continue to improve, with dsDNA virus capsid structures now exceeding 3 Å resolution (Chapter III; Zhao et al., 2017).

Table 1.1. Atomic resolution structures of Herpesviruses and Caudoviruses.

Virus Family	Species	Capsid Structure	Resol.* (Å)	Method	PDB code
Herpesviridae	HCMV	Mature virion	3.9	cryo-EM	5VKU
	HSV-2	Mature virion	3.1	cryo-EM	5ZAP
Myoviridae	T4	Isometric head	3.3	cryo-EM	5VF3
Podoviridae	BPP1	Mature virion	3.5	cryo-EM	3J4U
	P22	Procapsid†	3.8	cryo-EM	2XYY
		Mature virion†	4.0	cryo-EM	2XYZ
		Mature virion	3.3	cryo-EM	5UU5
	Sf6	Mature virion	2.9	cryo-EM	5L35
	T7	Head intermediate	3.5	cryo-EM	3J7W
		Mature virion	3.6	cryo-EM	3J7X
Siphoviridae	80α	Procapsid	3.7	cryo-EM	6B23
		Procapsid	3.8	cryo-EM	6B0X
	ATCC_Clear	Capsid	3.7	cryo-EM	3JB5
	HK97	Head intermediate II	3.6	x-ray	3E8K
		Head intermediate II	3.9	x-ray	2FT1
		Head intermediate IV	3.8	x-ray	2FSY
		Mature Virion	3.4	x-ray	1OHG
	P23-45	Expanded head	3.7	cryo-EM	6I9E
	P74-26	Mature virion	2.8	cryo-EM	6O3H
	TW1	Mature virion	3.6	cryo-EM	5WK1

***** Atomic resolution; defined as ≤ 4Å
† C_α backbone model

Capsid stability mechanisms in dsDNA viruses

The maturation of dsDNA virus capsids causes a substantial conformational switch in capsid architecture, flattening the capsid shell and increasing the interior volume (Figure 1.4). The capsid additionally accommodates the viral genome, which generates a large amount of pressure within the capsid. dsDNA viruses use a number of different mechanisms to stabilize the capsid shell against a high interior pressure. High-resolution structure determination of dsDNA virus capsids and capsid proteins has played an instrumental role in elucidating the structural mechanisms of capsid stability in dsDNA viruses.

The complex organization of MCP subunits as hexons and pentons within the capsid generates inherent weaknesses that require stabilization by extensive interactions between capsid subunits, and the addition of other structural proteins to the assembly. dsDNA viruses in particular have evolved a wide array of strategies to stabilize the capsid against high internal pressures. In phage HK97, capsid expansion leads to autocatalytic formation of covalent crosslinks between neighboring capsomers at the threefold axes, the junction point of three capsomers. This crosslinking creates an isopeptide bond between the E-loop and P-domain of two adjacent MCP subunits across the threefold axis, forming a protein ‘chain mail’ linking adjacent capsomers (Figure 1.6A) (Gertsman et al., 2009; Wikoff et al., 2000). This mechanism of generating covalent linkages

between capsid subunits is unique to HK97 and is not seen in any other dsDNA viruses.

Alternatively, some dsDNA viruses use auxiliary proteins such as decoration proteins for additional capsid stabilization, which bind to the expanded capsid as either dimers (Baker et al., 2013; Zhang et al., 2015) or trimers (Lander et al., 2008; Pietilä et al., 2013; Qin et al., 2010; Shen et al., 2012; Stone et al., 2018; Wang et al., 2018; Yang et al., 2000) at capsomer junction sites. In the well-characterized phage Lambda for example, the decoration protein gpD forms homotrimers that bind to the threefold and quasi-threefold axes, stabilizing each capsomer junction in the entire capsid assembly (Figure 1.6B) (Lander et al., 2008; Yang et al., 2008). The decoration protein is essential for encapsulation of the genome. Capsid assembly in the absence of Lambda gpD leads to an inability to contain the full-length genome (Sternberg and Weisberg, 1977). The decoration protein includes an extended N-terminal arm motif that forms extensive contact with a neighboring capsomer, creating a stabilizing 4-stranded β-sheet with the E-loop and N-arm of two adjacent MCP monomers (Lander et al., 2008). Phage T4 features a small outer capsid protein termed ‘Soc’ that adopts a similar homotrimeric arrangement at the threefold and quasi-threefold axes, despite having a prolate head morphology. Nevertheless, Soc performs a similar function to Lambda gpD in stabilizing the capsid shell (Qin et al., 2010). Dimeric decoration proteins differ in structure to their trimeric counterparts, but form similar stabilizing interactions albeit across the twofold axes between

capsomers (Zhang et al., 2015). Other auxiliary decoration proteins in some viruses such as the T4 highly-immunogenic outer capsid protein ‘Hoc’ and the SPP1 decoration protein bind to the center of hexons, but have little effect on stabilization of the capsid shell (Ross et al., 1985; Sathaliyawala et al., 2010; White et al., 2012).

Herpesviruses use Triplex proteins to stabilize the capsid shell in place of phage decoration proteins at the threefold axes in the capsid assembly. Rather than binding to the outer surface of the capsid after maturational expansion, Triplex proteins are an integral part of the assembly, filling pores in the capsid surface (Newcomb et al., 1993; Yu et al., 2017; Zhou et al., 1995). In contrast to the homo-trimeric decoration proteins found in some phages, Herpesvirus Triplex proteins form asymmetric trimers comprised of two different Triplex proteins, Tri1 and Tri2, in a 2:1 ratio (Figure 1.6C). Despite substantial differences in protein arrangement, Herpesvirus Triplex proteins share significant structural homology with Caudovirus decoration proteins within the core β-barrel domain, suggesting a common evolutionary lineage (Chapter II; Yu et al., 2017). In the Herpesvirus Human Cytomegalovirus (HCMV), additional interactions between HK97-fold domains in the MCP further stabilize capsomer junctions. Unique α-helical dimerization domains bind across the twofold axes, while an extended N-arm ‘lasso’ domain interacts with adjacent MCP E-loops, forming an additional ‘lasso triangle’ around the threefold axes that lies beneath the Triplex trimer (Yu et al., 2017).

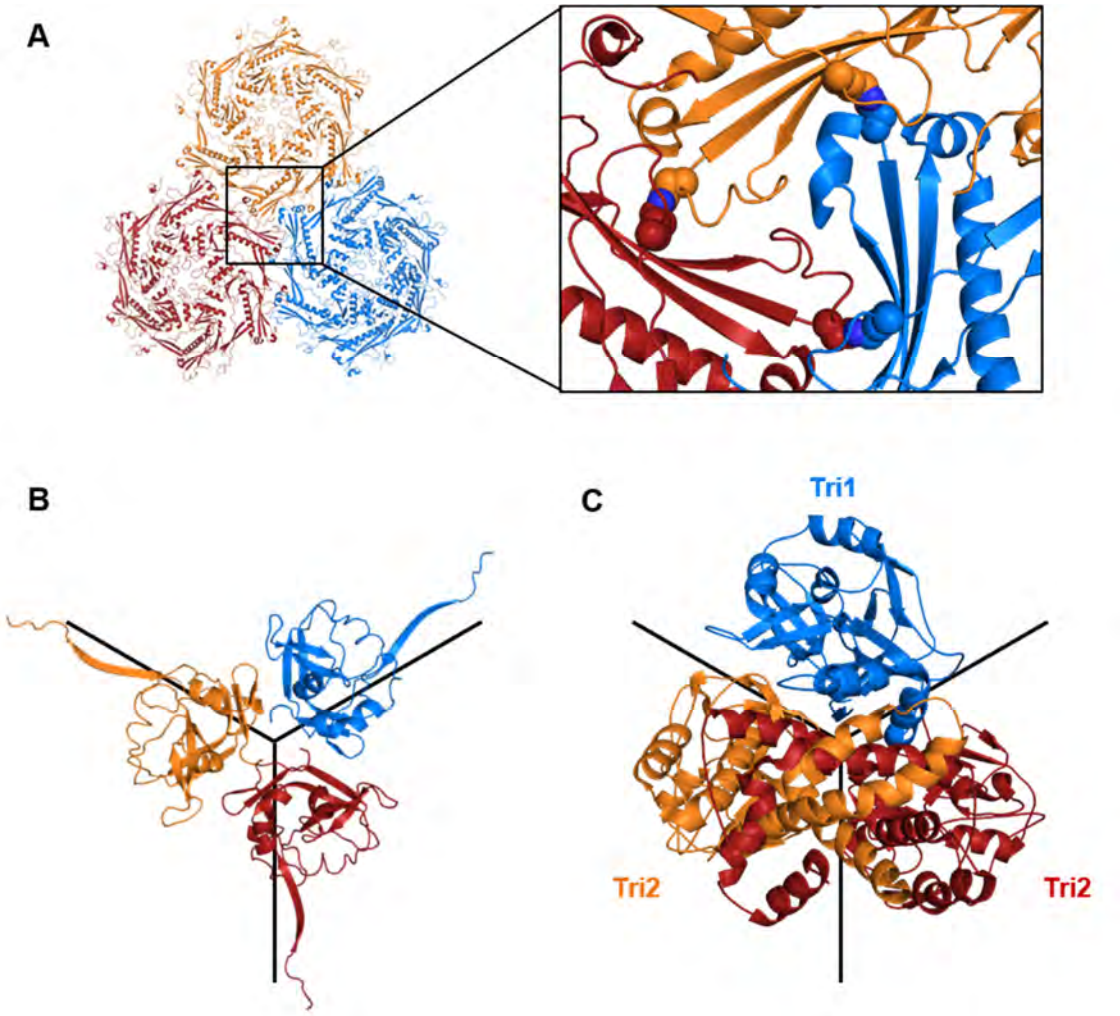

Figure 1.6. Mechanisms of capsid stabilization at the threefold capsid axes.
A) Many dsDNA viruses require additional stabilization at the threefold axes where capsomers meet. Upon capsid expansion, phage HK97 adopts a protein chain mail conformation, forming covalent crosslinks about the threefold axes (Inset; PDB: 1OHG).
B) dsDNA decoration proteins often form homotrimers that bind to the outer surface of the capsid shell (PDB: 6O3H).
C) Herpesvirus capsids stabilize threefold axes using Triplex heterotrimers formed from two copies of Tri2 and a single copy of Tri1 (PDB: 5VKU).

In other dsDNA viruses that lack decoration proteins or other auxiliary capsid proteins, capsids form extensive interactions between MCP subunits to stabilize the capsid shell. In phage T7 the extended E-loop of the mature virion MCP binds to the P-domain of an MCP monomer in an adjacent capsomer at the threefold axis, and forms extensive contact with the P-domain of an adjacent MCP subunit within the same capsomer (Agirrezabala et al., 2007; Guo et al., 2014). The interaction at the threefold axis functions similarly to the 'chain mail' linkages seen in phage HK97, though no covalent crosslinks are known to exist in the T7 assembly. In phage P22, the extended N-arm of MCP interacts with a protruding loop in the P-domain to stabilize the threefold axes throughout the capsid shell (Parent et al., 2010). Phage Sf6 MCP forms extensive hydrogen bond networks between the MCP E-loop and adjacent MCP subunit P-domains in a manner similar to T7. Additionally, the Sf6 N-arm is significantly extended compared to other HK97-like phages and faces the outer capsid surface, forming stabilizing interactions across the twofold axes of the icosahedral assembly (Zhao et al., 2017). Though these mechanisms of capsid stabilization vary widely in dsDNA viruses, the interactions described all contribute to stability of the 2-fold and 3-fold axes, suggesting that capsomer junctions are critical stabilization points within the capsid architecture.

Thermophilic bacteriophage P74-26 as a model for viral thermostability

For the work presented in this thesis, I use the hyperthermophilic bacteriophage P74-26 to elucidate the molecular mechanisms of capsid stability and assembly. P74-26 was initially isolated from alkaline hot springs (Yu et al., 2006) and infects the host bacterium *Thermus thermophilus,* which has an optimal growth temperature of 65-72 ˚C (Oshima and Imahori, 1974). P74-26 features an isometric capsid head and a tail structure approximately 850 nm in length (Figure 1.7A), the longest known virus tail discovered to date (Minakhin et al., 2008; Yu et al., 2006).

Similar to other tailed phages, P74-26 uses a powerful terminase motor to package its genome into preformed procapsid shells, generating immense pressure within the capsid interior. However, P74-26 capsids must also contend with the added stress of an extreme high-temperature natural environment. This provides a unique perspective from which to study viral maturation, and can further our understanding of maturation and stability in thermophilic viruses.

Another key defining feature of P74-26 is its close relation to mesophilic Caudoviruses. Thermophilic virus genomes often have little or no homology to other viruses annotated in online databases and sometimes represent unique or otherwise unclassified viral families (Rice et al., 2001; 2004). This is perhaps due to enhanced ‘evolvability’ of thermophilic proteins and complexes, as increased stability allows a higher tolerance for mutagenesis. Thus, thermophilic viruses

likely have a higher tolerance for sequence divergence compared to potential related ancestors, more frequently incorporating potentially destabilizing mutations (Bloom et al., 2006; Finch and Kim, 2018; Sterpone and Melchionna, 2012). Accordingly, thermophilic viruses often have unique capsid morphologies that bear little or no resemblance to mesophilic viruses (Rice et al., 2001).

In contrast to other thermophilic viruses, early proteomics characterization of P74-26 and its close relatives identified putative gene functions for 25% of the total open reading frames using publicly-available database searches (Minakhin et al., 2008). To date, several structural capsid proteins in P74-26 have been identified, including the Major Capsid Protein, decoration protein, scaffolding protein, portal, tail tube protein, and tail tape measure protein (Figure 1.7B) (Minakhin et al., 2008; Stone et al., 2018). Interestingly, several putative gene products in P74-26 share sequence homology with mesophilic viral proteins, suggesting that P74-26 has a closer relationship to mesophilic viruses than typically seen with other extremophiles (Minakhin et al., 2008).

The close relationship between P74-26 and several well-characterized mesophilic Caudoviruses and Herpesviruses poses a significant advantage in determining mechanisms of stability. Comparison of homologous thermophilic and mesophilic proteins has proven useful for elucidating stability mechanisms in a diverse array of proteins (Kumar et al., 2000; Razvi and Scholtz, 2006), but virus capsids have not been extensively studied in such a manner. For the first time, this model system allows for the direct comparison of thermophilic and

mesophilic virus capsids from which we can elucidate the core principles of capsid thermostability. Furthermore, extensive characterization of P74-26 can provide insight into evolutionary relationships and help to elucidate common ancestral lineages amongst Caudoviruses and other viral families.

Due to inherent stability, proteins from thermophilic systems are also widely amenable to *in vitro* studies. Previous work in our lab has demonstrated that P74-26 proteins recombinantly expressed in *E. coli* can be recovered in high yield using conventional purification methods (Hilbert et al., 2015; Hilbert et al., 2017; Stone et al., 2018), providing a robust model for *in vitro* study.

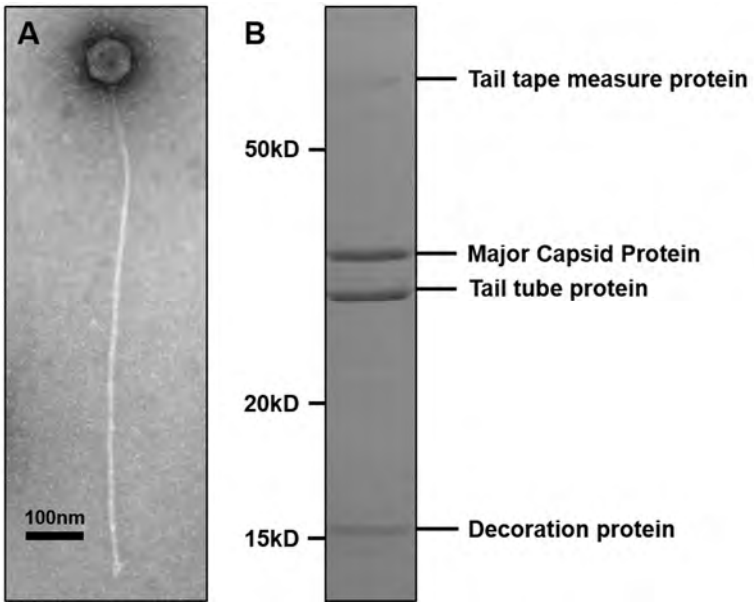

Figure 1.7. Thermophilic bacteriophage P74-26.
A) Negative-staining electron micrograph of the P74-26 virion. The mature virions feature an icosahedral capsid and the longest known virus tails.
B) SDS-PAGE analysis of purified mature virions reveals the predominant structural elements within the capsid assembly.

Scope of Thesis

Despite a long history of phage and virus structural studies spanning more than 70 years, there are only a small number of structures that reveal atomic-level details of the mechanisms of viral maturation. Furthermore, nearly all of these models are of temperate mesophilic viruses, and do little to explain the mechanisms driving stability and assembly in viruses that thrive in unique environments. This thesis aims to elucidate the molecular mechanisms that govern stability and assembly in the thermophilic bacteriophage P74-26 and derive a model of capsid thermostability and determine the capsid protein subunits and architectures that are critical for virion stabilization.

In Chapter II, I elucidate the structure of the P74-26 capsid decoration protein and characterize its enhanced stability compared to homologous mesophilic decoration proteins. This work additionally revealed a conserved ‘β-tulip’ domain found in several viral structural proteins in both Herpesviruses and phage, providing evidence for a common ancestry.

In Chapter III, I elucidate the 2.8-Å structure of the P74-26 phage capsid using single-particle cryo-EM. The atomic structure of the icosahedral capsid lattice reveals a complex network of interleaved protein-protein interactions and unique architectures that stabilize the mature virion in the thermophile. P74-26 capsid structure determination also led to the discovery of a novel mechanism for enhancing the interior capacity of a capsid shell while conserving a stable T=7

capsid geometry. Together, this work illuminates the incredible stability of an extremophilic virus and establishes a novel viral platform for future studies of viral maturation.

CHAPTER II

A hyperthermophilic phage Decoration protein suggests a common evolutionary origin with Herpesvirus Triplex proteins and an anti-CRISPR protein

Preface

The work presented in this chapter is adapted from a previous publication titled “A hyperthermophilic phage Decoration protein suggests common evolutionary origin with herpesvirus Triplex proteins and an anti-CRISPR protein” published in the journal *Structure* on July 3, 2018.

This manuscript was written by Brian Kelch and me, and together we designed the experimental plan. I performed equilibrium unfolding assays, crystallographic data processing and refinement, model building, and structural analysis. Contributing authors Dr. Brendan Hilbert and Daniel Hidalgo performed protein crystallization and collected x-ray diffraction data. Dr. Kevin Halloran assisted with experimental design and data analysis for equilibrium unfolding assays described in the text. Jooyoung Lee performed the Cas9 inhibition assay described in the text.

Introduction

Tailed bacteriophage (also known as Caudoviruses) are the most common biological entities on Earth (Suttle, 2007; Wommack and Colwell, 2000). Caudoviruses assemble viral particles by actively packaging the viral genome inside a preformed, self-assembled capsid structure (Catalano, 2005). Because the DNA fills the capsid to nearly crystalline density, the pressure inside the capsid head is estimated to be extremely high (>6 MPa) (Evilevitch et al., 2004). Thus, the capsid structure must withstand high internal strain, in addition to the challenges imposed by the fluctuating external environment.

The capsids of Caudoviruses are primarily comprised of a single major capsid protein (MCP) of the HK97 fold, which assembles into an icosahedral shell (Aksyuk and Rossmann, 2011; Suhanovsky and Teschke, 2015). The MCP subunits interact with each other to build an interdigitated structure with high structural integrity. Proper construction of the capsid is directed by the portal ring that nucleates self-assembly of MCP, as well as a scaffolding protein that chaperones the MCP subunits to ensure the proper size and shape of the shell (Lurz et al., 2001; Prevelige and Fane, 2012; Prevelige et al., 1993). The MCP protein forms units called capsomers of five or six subunits, termed pentons or hexons, respectively. The capsid typically assembles with 11 pentons and a variable number of hexons, depending on the triangulation number of the virus shell. The capsid structure is altered during genome packaging, causing expansion of the shell, loss of the scaffolding protein, and a morphological

change from nearly spherical to an icosahedral structure. The icosahedral shell is stable against the high internal pressure from the condensed DNA filling the capsid interior.

Many viruses have evolved mechanisms to stabilize the quasi and icosahedral three fold axes of the capsid, which are necessary for stabilization of the icosahedral shell. For example, the Siphovirus HK97 uses covalent linkages at the three-fold axes that form a ‘chain-mail’ to stabilize the capsid structure (Wikoff et al., 2000). On the other hand, many viruses instead use auxiliary capsid proteins to add structural integrity, in particular the trimeric decoration proteins that bind at the three-fold axes of the capsid icosahedron (Sternberg and Weisberg, 1977). Decoration proteins (also known as cementing proteins) play an important role in capsid stability and assembly. Cryo-electron microscopy (cryo-EM) structures reveal that decoration proteins specifically interact with the three-fold axis by inserting their N-terminal region into a groove formed by neighboring capsomers (Lander et al., 2008). Decoration proteins only bind to the expanded, icosahedral capsid because the sites for binding are occluded in the smaller, spherical procapsids (Tang et al., 2006). Studies of the Lambda phage decoration protein revealed that binding of decoration protein trimers strengthens the capsid assembly against both increased temperature and mechanical deformation (Gilcrease et al., 2005; Sae-Ueng et al., 2014; Sternberg and Weisberg, 1977). However, it is unclear how evolution modulates decoration

protein-mediated stabilization. Moreover, it is unclear how decoration proteins are related to other proteins found throughout viral and cellular lineages.

Herpesviruses have a similar mechanism of capsid assembly as tailed bacteriophage, which has implied that Herpesviruses evolved from Caudoviruses (Rixon and Schmid, 2014). Both classes of viruses use similar DNA packaging machinery, with a terminase-class motor and portal complex that acts as both an entrance and exit for the viral genome (Bazinet and King, 1985; Hendrix, 1978; Hsiao and Black, 1978). Herpesvirus and Caudovirus capsids are primarily comprised of a major capsid protein with the HK97 fold (Suhanovsky and Teschke, 2015) and direct self-assembly using a similar scaffolding protein (Fane and Prevelige, 2003). Herpesvirus particles have trimeric Triplex proteins at the three-fold axes of the capsid (Newcomb et al., 1993). Triplex is an attractive target for HCMV vaccine development (Choi et al., 2016). The Triplex complex is an integral part of both the immature procapsid and the final capsid shell, unlike decoration proteins of Caudoviruses (Heymann et al., 2003). Moreover, the Triplex proteins are substantially larger and share no significant sequence similarity with decoration proteins. Thus, the shared mechanisms of particle assembly and evolutionary relationships of these capsid proteins remain unclear.

Phage structural proteins can be evolutionary sources of proteins that perform other functions, particularly phage-host interactions (Basler et al., 2012; Fraser et al., 2006; Veesler and Cambillau, 2011). A critical class of proteins that modulate phage-host interactions are the Anti-CRISPR proteins, which are

encoded by phage and other mobile genetic elements to inhibit the CRISPR/Cas (Clustered Regularly Interspersed Short Palindromic Repeats, and CRISPR-associated system, respectively) adaptive immune system of prokaryotes (Borges et al., 2017). Beyond their critical role in the molecular arms race between prokaryotes and viruses, Anti-CRISPR (Acr) proteins are important for medicine and biotechnology. Anti-CRISPR proteins have been found in bacterial mobile elements and have been proposed to increase virulence of these bacterial strains (Maxwell, 2016). Acr proteins have been shown to improve CRISPR/Cas9-based genome editing by limiting off-target effects (Shin et al., 2017), as well as converting a CRISPR-associated nuclease into a transcriptional repressor (Bondy-Denomy et al., 2015). Despite their importance, the evolution of Acr proteins is still largely mysterious. As has been noted previously (Bondy-Denomy et al., 2013; Pawluk et al., 2016), Acr genes are often encoded near viral structural genes, which leads us to hypothesize that some Acr genes may have evolved from structural components of viruses.

Here we investigate phage decoration protein structure, function, and evolution using a hyperthermophilic phage. Phage P74-26 is found in hot springs and infects *Thermus thermophilus* bacteria (Yu et al., 2006). P74-26 has the longest tail of known viruses (~1 μm) and thrives at 70 ˚C (Minakhin et al., 2008; Yu et al., 2006), a temperature at which related mesophilic phage are disabled in minutes (Bauer and Evilevitch, 2015). We identify the protein gp87 as the decoration protein of P74-26 phage. Our 1.7-Å resolution structure of gp87

reveals it has the same fold as the phage Lambda decoration protein, despite a lack of sequence similarity. We show that the hyperthermophilic protein is substantially more stable than its mesophilic homologs. Furthermore, we identify a conserved β-barrel domain of the decoration protein that is found in Herpesvirus Triplex proteins and the Anti-CRISPR protein AcrIIC1, suggesting that these diverse proteins share a common evolutionary ancestor. Our work provides the groundwork for understanding the high stability of thermophilic viruses. Moreover, these studies illustrate a deep connection between the capsid machinery of tailed bacteriophage and Herpesviruses, and lead to a potential mechanism for the evolution of an anti-CRISPR protein.

Materials and Methods

Growth and purification of P74-26 virions

Phage stocks were prepared using fresh overnight cultures of *Thermus thermophilus* grown at 65 ˚C in Thermus Growth Medium (0.8% (w/v) Tryptone, 0.4% (w/v) Yeast Extract, 0.3% (w/v) NaCl, 1 mM $MgCl_2$, 0.5 M $CaCl_2$). For preparation of P74-26 phage stock, 6 mL of fresh *T. thermophilus* (OD_{600} = 1.0) was inoculated with 4 mL of purified phage stock at $1x10^6$ Plaque Forming Units per mL (PFU/mL) for adsorption. Adsorption reaction mixture was incubated for 10 minutes at 65 ˚C, then inoculated into 1 L Thermus Growth Medium. The culture was then incubated at 65 ˚C, shaking for 4-5 hours, yielding a high-titer phage lysate (>$1x10^9$ PFU/mL). Lysates were spun at 4,000 x g for 20 minutes at 4˚ C to remove cell debris, then the supernatant was treated with DNase I and RNase A to a final concentration of 2 Units/mL and 1 µg/mL, respectively and incubated at 30˚ C for one hour. Solid NaCl was added to the phage stock to a final concentration of 1 M while stirring, then culture was incubated on ice for one hour and spun at 11,000 x g for 20 minutes at 4 ˚C. To precipitate virions, solid PEG-8,000 was added to a final concentration of 10% (w/v) while stirring and phage stock was incubated on ice overnight.

To pellet virions, precipitated phage stock was spun at 11,000 x g for 20 minutes at 4 ˚C. The phage pellet was then resuspended in 2 mL of resuspension buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM $MgS0_4$). 0.4 g

solid CsCl was added to resuspension. Phage resuspension was then added to CsCl step gradients (five steps: 1.2 g/mL (2 mL), 1.3 g/mL (2 mL), 1.4 g/mL (2 mL), 1.5 g/mL (2 mL), and 1.7 g/mL (1 mL); made in 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM $MgS0_4$) prepared in 12 mL ultracentrifuge tubes (Seton Scientific). Gradients were spun in Beckman SW-40Ti swinging bucket rotor at 38,000 RPM for 18 hours at 4 ˚C. P74-26 virions, which sediment at ~1.5 g/mL CsCl, were isolated and dialyzed twice overnight at 4 ˚C into 2 L of 50 mM Tris-HCl pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$.

SDS-PAGE analysis

30 µL samples of virions (~$1x10^{11}$ PFU/mL) were run on 12% SDS-PAGE gels. Samples were incubated in SDS loading buffer for five minutes at 95 ˚C and run on gels at 180V for 45 minutes. Gels were fixed with 50% (v/v) ethanol, 10% (v/v) acetic acid and then stained with Coomassie Blue in 5% (v/v) ethanol, 7.5% (v/v) acetic acid. Gels were imaged on an Amersham Imager 600 (GE Healthcare). Densitometry was performed using ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997-2016).

Electron microscopy

CsCl gradient purified virions (~1x10^{11} PFU/mL) were applied to 400-mesh copper grids (Electron Microscopy Sciences) coated in carbon. 3.5 µL samples were applied to carbon surface of grid and incubated for 30 seconds, then excess sample was removed from grids. Following sample application grids were stained with 1% Uranyl Acetate (pH 4.5) and visualized in a Philips CM120 electron microscope (120kV) fitted with Gatan Orius SC1000 camera. Micrographs were collected at a magnification of 19,500X.

Cloning, expression, and purification of gp87

P74-26 gp87 was synthesized by the Genscript Corporation and subcloned into BamHI and XhoI sites of the pSMT3 vector with a cleavable N-terminal His6-SUMO tag (Yunus and Lima, 2009). Restriction enzymes were purchased from New England BioLabs, and oligonucleotide primers were obtained from Integrated DNA Technologies. P74-26 gp87 forward primer: GATCGGATCCATGGATAAAATTCAACTG; P74-26 gp87 reverse primer: GATCCTCGAGTCAGCGCGTGTAGTCAAAGAAATAG.

P74-26 gp87 was expressed in *E. coli* BLR-DE3 cells containing the pSMT3-gp87 plasmid. Cultures were grown in Terrific Broth supplemented with 30 µg/mL kanamycin at 37 ˚C to an OD_{600} of 0.7. Cultures were then incubated at 4 ˚C for 20 minutes, then overnight expression at 18 ˚C was induced with a final

concentration of 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG). Cells were then pelleted and resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM KCl, 20 mM Imidazole, 5 mM 2-mercaptoethanol (βME), 10% (v/v) Glycerol). All subsequent purification steps were performed at room temperature. Cells were lysed in a cell disruptor and pelleted. Cleared lysate was then applied to 2x5 mL His-Trap columns (GE Healthcare) pre-equilibrated in buffer A. P74-26 gp87 was then eluted with buffer B (50 mM Tris-HCl pH 7.5, 300 mM KCl, 500 mM Imidazole, 5 mM βME, 10% (v/v) Glycerol). Eluate was dialyzed into 50 mM Tris-HCl pH 7.5, 150 mM KCl, 2 mM DTT and the His-SUMO tag was cleaved with Ubiquitin-like Specific Protease 1 (ULP1) overnight. The dialyzed protein was passed over a 5mL His-Trap to remove cleaved His-SUMO tag. Cleaved eluate was then concentrated to 60 μM, aliquotted and flash frozen in liquid nitrogen, then stored at -80 ˚C.

Crystallization, data collection, and structure determination

P74-26 gp87 native crystals were formed by hanging-drop vapor diffusion at 25˚ C. 1mg/mL protein was mixed 1:1 with well solution containing 100 mM Tris-HCl pH 7.0, 19.5% (w/v) PEG 3350. Crystals were soaked in cryoprotectant (100 mM Tris-HCl pH 7.0, 21% (w/v) PEG 3350) and flash frozen in liquid nitrogen prior to data collection. Initial native dataset was collected to 1.9 Å using a MicroMax007-HF/Rigaku Saturn 944 CCD detector x-ray diffraction system at

wavelength 1.54 Å (home source). Following native dataset collection, the native crystal was thawed, soaked in well solution supplemented with 1 M KI for 60 seconds, and flash frozen again. A KI derivative dataset was then immediately collected to 2.3 Å, with anomalous signal extending to 2.8 Å according to phenix.xtriage. Using a separate crystal, an additional native dataset was later collected to 1.7 Å at Advanced Photon Source beamline 19-BM at wavelength 0.979 Å. Diffraction datasets were processed using HKL3000 (Otwinowski, 1997). P74-26 gp87 structure was solved by SAD phasing of the 2.3 Å derivative dataset using the iodide anomalous signal with PHENIX autosol (Dauter et al., 2000; Zwart et al., 2008). 11 iodides were found per monomeric subunit. Structure refinement and model building were performed with the programs PHENIX (Adams et al., 2010) and COOT (Emsley and Cowtan, 2004). Structure was deposited in the Protein Data Bank (www.rcsb.org), PDB code: 6BL5.

Folding and refolding analysis

P74-26 gp87 stocks at final concentrations of 2.5, 5, 10, or 15 μM were prepared in Tris-Buffered Saline, pH 7.5 (Boston BioProducts) and either 7.5M guanidine hydrochloride (GdnHCl) or no GdnHCl. These stocks were mixed using a Hamilton Microlab 500 series titrator to yield final samples with the desired concentration of GdnHCl. GdnHCl concentration of stocks was determined using an ABBE Mark II refractometer (Reichert). Fluorescence

emission spectra for GdnHCl titrations at 2.5, 5, and 10 μM concentrations of gp87 were collected at 25˚ C using a Fluoromax 4 Spectrofluorometer (Horiba Scientific). P74-26 gp87 was excited at 295 nm and emission spectra were collected from 310 to 400 nm for each sample. To check for hysteresis, we measured fluorescence after 24, 48, and 120 hours of incubation at 25 ˚C. No significant difference in fluorescence was seen between 24 hours and 120 hours, indicating that folding/unfolding reached equilibrium. Circular dichroism (CD) spectra of titrations of gp87 at 5 μM and 15 μM final concentrations were collected at 25 ˚C using a Jasco-810 spectropolarimeter (Jasco, Inc) equipped with a temperature control system. Equilibrium unfolding was monitored from 215-260 nm. The Trp fluorescence and CD data were globally fit using single-value decomposition analysis (Simler et al., 2004) to a trimer unfolding to three monomers model ($N_3 \rightleftharpoons 3U$) across all wavelengths for all spectra collected. Data was fit using an in-house least squares fitting program Savuka (Gualfetti et al., 1999).

SEC-MALS

P74-26 gp87 was run on tandem size exclusion chromatography – multi-angle light scattering (SEC-MALS) by injecting a 100 μL sample at a concentration of 1 mg/mL (~60 μM) in 50 mM Tris-HCl pH 7.5, 150 mM KCl, 2 mM DTT. Protein was filtered through a 0.2 μM syringe filter. Elution was

monitored using a Dawn Heleos-II MALS detector and Optilab T-rex differential refractive index detector (Wyatt Technology). Elution peaks were selected and molar mass was determined using the ASTRA6 analysis program (Wyatt Technology). The peak for the gp87 trimer eluted after 10 minutes at a concentration of ~20 μM as measured by refractive index and absorbance.

Structural analysis

Homology searches using both full-length P74-26 gp87 and the β-tulip domain alone were performed using NCBI BLAST, DALI protein structure comparison server, and NCBI Vector Alignment Search Tool (VAST) (Holm and Rosenström, 2010; Madej et al., 2014). Structure-based sequence alignment between P74-26 gp87 and Lambda gpD was performed using the PDBeFold program (EMBL-EBI - www.ebi.ac.uk/msd-srv/ssm/). Multiple sequence alignments with homologous proteins were performed using the Clustal Omega program (EMBL-EBI - www.ebi.ac.uk/Tools/msa/clustalo/). C_α RMSD and Z-score calculations for full-length protein and β-Tulip domain alignments were determined using PDBeFold. Surface area calculations for gp87, gpD, and SHP were performed using the program InterProSurf (http://curie.utmb.edu/prosurf). ILVF cluster analysis of hydrophobic networks gp87 and gpD was performed using the program BASiC Networks (http://biotools.umassmed.edu/ccss/ccssv2/basic.cgi).

NmeCas9 inhibition assay

Purification of NmeCas9 and anti-CRISPR proteins is as described in Pawluk *et al*. (Pawluk et al., 2016). NmeCas9 sgRNA targeting N-TS4B (Amrani *et al*., bioRxiv, 2017) was generated by *in vitro* T7 transcription (Epicentre). *In vitro* cleavage assay was performed as previously described (Pawluk et al., 2016). Briefly, NmeCas9 (50 nM) was pre-incubated with anti-CRISPR or gp87 protein in reaction buffer (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 1 mM DTT, and 5 mM $MgCl_2$) for 10 minutes at room temperature. Next, *in vitro* T7 transcribed sgRNA (75 nM) was added to the mixture and incubated for another 20 minutes. pUC19 plasmid containing the target site N-TS4B was linearized by Sca I digestion. The linearized plasmid was added at ~1 nM final concentration. The reactions were incubated at 37 ˚C for 1 hour, treated with proteinase K at 50 ˚C for 10 minutes, and visualized after electrophoresis in a 1% agarose/TAE gel.

Results

Major components of P74-26 virions

We purified P74-26 virions from infections of *Thermus thermophilus* strain HB8 using a combination of PEG precipitation and gradient centrifugation. The P74-26 virions consist of an icosahedral capsid ~80 nm in diameter and long, non-contractile tails exceeding 800 nm in length (Figure 2.1A).

To determine whether the P74-26 capsid is stabilized by covalent interactions (as in the phage HK97 (Duda, 1998; Wikoff et al., 2000)), we analyzed the proteome of P74-26 virions by SDS-PAGE. The major capsid protein migrates on SDS-PAGE at its expected molecular weight (46.6 kDa), similar to results obtained previously (Figure 2.1B)(Minakhin et al., 2008). This result is in contrast to phage HK97, whose MCPs are linked covalently and run on SDS-PAGE as a smear at very high molecular weight (Duda, 1998). Thus, we conclude that the P74-26 capsid is stabilized by non-covalent interactions.

We hypothesized that P74-26 capsids are stabilized by decoration proteins. Our SDS-PAGE analysis reveals other major components of the virions: the tail tube protein (gp93; 37.9 kDa; the tape measure protein (gp95; 550 kDa) and gp87 (16.3 kDa), whose function in the virion is unknown. Based on size, abundance, and gene location, we hypothesize that gp87 acts as a decoration protein. gp87 is similar in size to known decoration proteins gpD and SHP from phages Lambda and P21, respectively (~12 kDa for both). Moreover, the relative

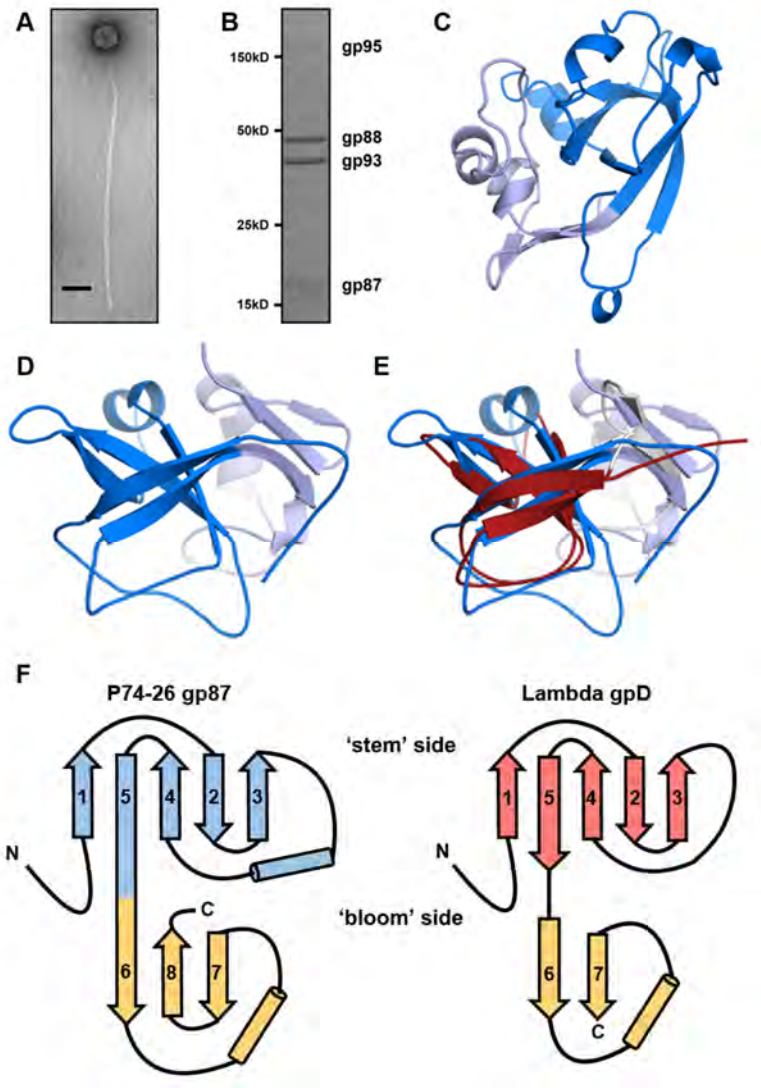

Figure 2.1. P74-26 gp87 is a thermophilic capsid decoration protein.
A) Negative-stain electron micrograph of purified P74-26 virion; scale bar, 100 nm.
B) SDS-PAGE analysis of P74-26 virions reveals major structural components including gp87, gp88 (major capsid protein), gp93 (tail protein), and gp95 (tape measure protein).
C & D) 1.7-Å resolution structure of P74-26 gp87 with the five-stranded β-tulip domain highlighted in dark blue.
E) Structure-based alignment of P74-26 gp87 (β-tulip domain in blue) and Lambda decoration protein gpD (grey, β-tulip domain in red, PDB: 1C5E) reveals significant structural homology despite high sequence variance.
F) Topology diagrams of P74-26 gp87 and Lambda gpD reveal conserved architecture of β-tulip domain flanked by a small mixed α/β domain.

Table 2.1. P74-26 gp87 data collection and refinement statistics

Data Collection	KI Derivative	Native (APS)
Space Group	P 63	P 63
Wavelength	1.54 (Home source)	0.979 (19-BM)
Resolution range	29.38 – 2.34	25.84 – 1.69
Unit cell angles	89.74, 89.74, 31.36	89.52, 89.52, 31.39
Unit cell dimensions	90, 90, 120	90, 90, 120
Total reflections	67505	119377
Unique reflections	6255	16355
Multiplicity	10.8 (9.7)	7.3 (7.2)
Completeness (%)	98.9 (95.6)	99.9 (100)
Mean I/sigma I	16.2 (3.2)	14.3 (2.3)
Wilson B factor	32.14	26.76
R-merge	0.125 (0.728)	0.057 (0.824)
R-meas	0.131 (0.768)	0.062 (0.888)
R-pim	0.039 (0.242)	0.023 (0.328)
CC ½	0.998 (0.847)	0.999 (0.810)
Refinement		
R-work %		17.87
R-free %		21.26
RMS bonds		0.01
RMS angles		1.05
Ramachandran favored %		96.1
Rotamer outliers %		0
Clashscore		1.01
Average B		36.8

abundance of gp87 is close to that expected for a decoration protein (318 ± 6 copies per virion by gel band densitometry). Finally, the genes encoding decoration proteins and MCPs are usually proximal to each other (Casjens, 2005; Hatfull, 2008); the P74-26 MCP is encoded by gene 88, adjacent to gene 87.

P74-26 gp87 structure reveals similarity to phage decoration proteins

To test whether gp87 is a decoration protein, we determined its atomic structure by x-ray crystallography. Purified recombinant P74-26 gp87 readily crystallized. We solved the structure to 2.3-Å resolution by single-wavelength anomalous dispersion (SAD) using iodide ions soaked into the crystals, and phases subsequently were extended to a 1.7-Å resolution native dataset (Table 1; Figure 2.1C,D). The structure reveals a core domain comprised of a five-stranded anti-parallel β-barrel, followed by a mixed α/β subdomain. According to the nomenclature for beta-strand topology (Richardson, 1981), the barrel has a topology of (+3, +1, -2, -1) (Figure 2.1F). An α-helical linker connects strands 3 and 4. One end of the β-barrel is flared open, while the other is capped with loops, such that the domain resembles a tulip. We refer to the domain as a β-tulip due to this resemblance (Figure 2.2A). The smaller C-terminal domain consists of a three-stranded anti-parallel β-sheet capped by an α-helix and extended loop structure. The N-terminal 16 amino acids are not visible in the structure, most

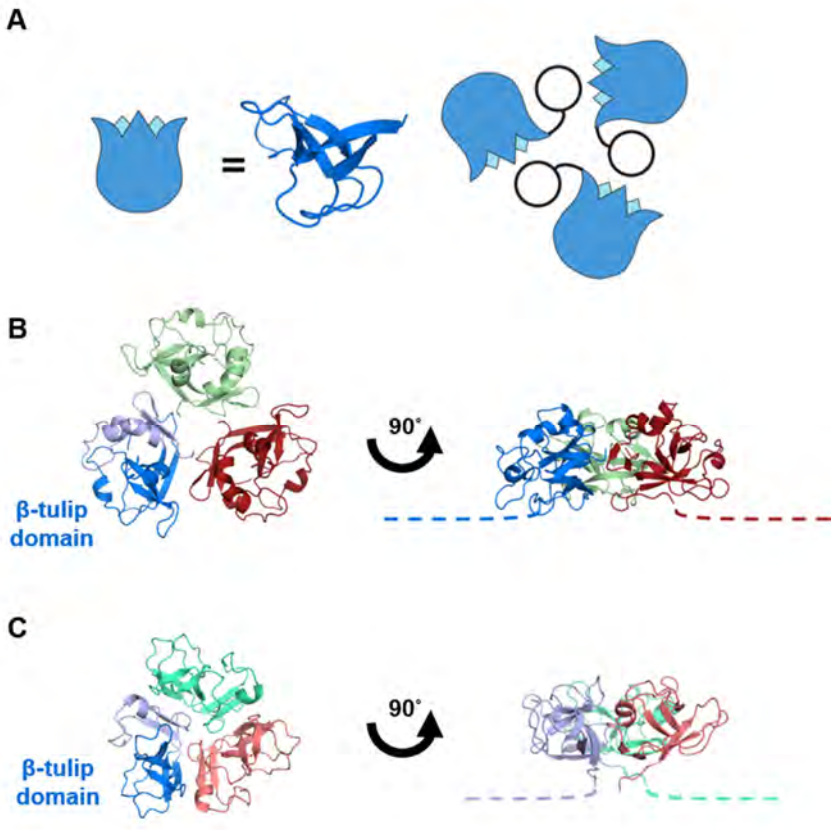

Figure 2.2. Interactions of the decoration protein for capsid stabilization.
A) Model of decoration protein trimer highlighting positions of the β-tulip domains (blue) within the assembly.
B) P74-26 gp87 trimer highlights difference in trimer assembly characterized by a ~20˚ outward rotation of each of the gp87 trimer subunits. The N-terminal capsid binding region of both crystal structures is disordered, and is drawn as proportional dotted lines in A and B.
C) Structure of the Lambda gpD trimer shows the orientation of gpD from the top of the capsid (left) and rotated 90˚ to the side (right).

likely due to disorder. The crystal packing reveals a trimer of gp87 proteins arranged in a ‘head-to-tail’ fashion (Figure 2.2B). The ‘bloom’ end of the tulip is interacting with the mixed α/β-domain of a neighbor, primarily through hydrophobic interactions. β-tulip residues F27, L29, F31, I80, and L99 form an intermolecular hydrophobic cluster with the C-terminal domain residues I135 and F140 in the neighboring subunit.

The structure and assembly of P74-26 gp87 is very similar to that observed for other well-known decoration proteins. The gp87 protein exhibits the same fold as the gpD protein from phage Lambda and the SHP protein from lamboid phage P21 (Forrer et al., 2004; Yang et al., 2000) (Figure 2.1E,F; Figure 2.3A,B). Both gpD and SHP contain an N-terminal β-tulip domain and a C-terminal mixed α/β subdomain (C_α RMSD compared to gp87 is 2.3 Å and 2.5 Å, respectively). The similarity between the β-tulip domains of gp87 and gpD is particularly high (C_α RMSD = 2.1 Å; Table 2). Moreover, all three proteins have an N-terminal region that is disordered in the crystal structures (16, 14, and 11 residues for gp87, gpD, and SHP, respectively). The C-terminal domains have an overall similar fold (strand-loop-helix-strand). In gpD and SHP, the two β-strands of the C-terminal subdomain form a parallel β-sheet, whereas gp87 has an insertion of an extra β-strand that results in the formation of an anti-parallel three-stranded β-sheet (Figure 2.3C,D). Thus, the fold of gp87 is nearly identical to that of gpD and SHP.

The decoration proteins gpD and SHP also crystallized as trimers, with a similar (but not identical) arrangement around the three-fold axis of the trimer as in gp87 (Figure 2.2B, 2.2C). The trimerization interface is comprised of the core β-tulip domain of one subunit contacting the C-terminal domain of an adjacent subunit. This arrangement places the unstructured N-terminal region of each subunit at the 'bottom' of the trimer such that it can attach to the capsid shell (Lander et al., 2008). Thus, the similarities in both tertiary and quaternary structure, and its high abundance in the virion lead us to conclude that gp87 is a decoration protein. The close structural and functional similarity is notable despite the low sequence homology between the thermophilic gp87 and its mesophilic cousins (~9% sequence identity) (Figure 2.3E).

Despite the lack of clear sequence similarity, we detect several residues that are conserved across decoration proteins. We built a structure-based sequence alignment of phage decoration proteins to identify any trends in conservation of residues throughout decoration proteins. We identify only 7 residues that are largely conserved across known homologs (Figure 2.3E). We note two conserved residues (Asp 66 and Glu 83 in Lambda gpD) poised near the capsid binding surface of the decoration protein that may play a role in stabilizing the capsid when gpD is bound to the virion. Future studies will determine whether these putative interactions are important for stabilizing the capsid assembly.

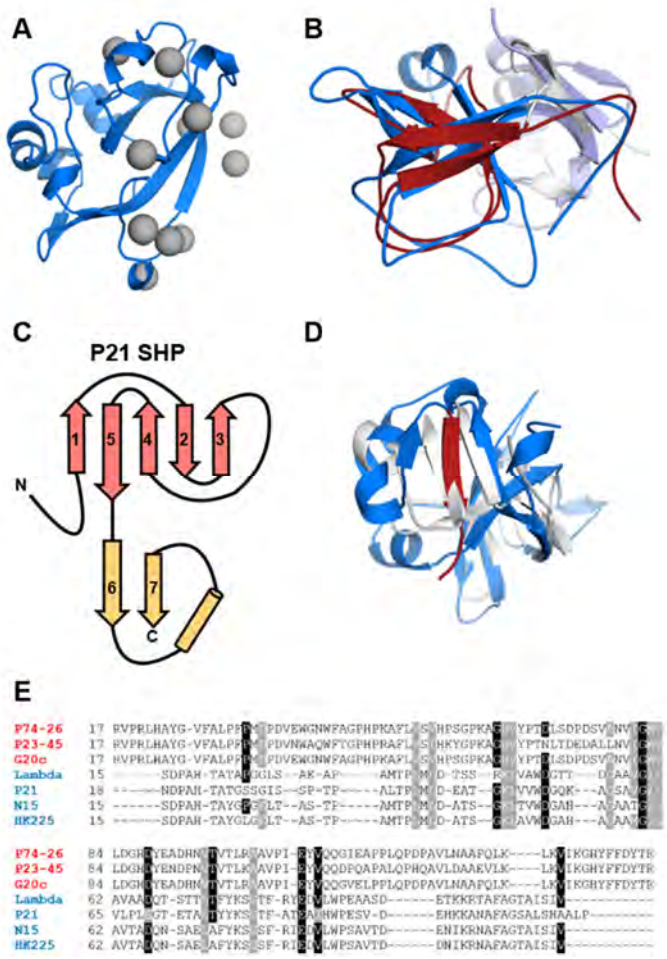

Figure 2.3. P74-26 gp87 has significant similarity to decoration protein SHP.
A) Iodide ion binding positions are shown mapped onto 1.7-Å structure of P74-26 gp87 (iodides in grey).
B) Structural alignment of gp87 (β-tulip in blue) with the decoration protein SHP from phage P21 (white, β-tulip in red, PDB: 1TD3).
C) Topology diagram of the P21 decoration protein SHP.
D) Structural alignment of P74-26 gp87 (blue) with Lambda gpD (grey) highlighting additional strand in gp87 (red) forming an antiparallel β-sheet in the C-terminal domain.
E) Structure-based multiple sequence alignment of thermophilic and mesophilic decoration proteins reveals conserved residues. Thermophilic and mesophilic decoration proteins are color coded red and blue, respectively.

Table 2.2. C_α RMSD comparison of β-tulip domains. RMSD values reported in Å; N_{aln} = number of residues aligned for each comparison.

	P74-26 gp87	Lambda gpD	P21 SHP	HCMV Tri1	HCMV Tri2	Anticrispr AcrIIC1	φ29 gp12 (N-term)	φ29 gp12 (C-term)	TW1 gp56 (N-term)	TW1 gp56 (C-term)
Lambda gpD	N_{aln}: 54 RMSD: 2.12 Z score: 8.4									
P21 SHP	N_{aln}: 55 RMSD: 1.96 Z score: 8.0	N_{aln}: 54 RMSD: 0.79 Z score: 9.7								
HCMV Tri1	N_{aln}: 53 RMSD: 2.40 Z score: 3.4	N_{aln}: 48 RMSD: 2.43 Z score: 4.2	N_{aln}: 46 RMSD: 2.38 Z score: 4.0							
HCMV Tri2	N_{aln}: 43 RMSD: 2.24 Z score: 4.9	N_{aln}: 44 RMSD: 2.24 Z score: 6.0	N_{aln}: 44 RMSD: 2.36 Z score: 6.2	N_{aln}: 43 RMSD: 1.64 Z score: 4.9						
Anticrispr AcrIIC1	N_{aln}: 45 RMSD: 2.61 Z score: 3.8	N_{aln}: 38 RMSD: 3.07 Z score: 3.8	N_{aln}: 37 RMSD: 2.94 Z score: 3.2	N_{aln}: 38 RMSD: 2.29 Z score: 2.7	N_{aln}: 39 RMSD: 2.63 Z score: 3.6					
φ29 gp12 (N-term)	N_{aln}: 39 RMSD: 1.70 Z score: 5.1	N_{aln}: 46 RMSD: 2.03 Z score: 5.3	N_{aln}: 47 RMSD: 1.92 Z score: 6.4	N_{aln}: 46 RMSD: 2.65 Z score: 3.7	N_{aln}: 46 RMSD: 2.16 Z score: 4.0	N_{aln}: 43 RMSD: 3.03 Z score: 2.5				
φ29 gp12 (C-term)	N_{aln}: 45 RMSD: 1.51 Z score: 6.6	N_{aln}: 42 RMSD: 2.08 Z score: 5.4	N_{aln}: 42 RMSD: 2.04 Z score: 5.0	N_{aln}: 45 RMSD: 2.38 Z score: 5.1	N_{aln}: 42 RMSD: 2.16 Z score: 5.6	N_{aln}: 40 RMSD: 2.27 Z score: 3.5	N_{aln}: 44 RMSD: 2.97 Z score: 2.9			
TW1 gp56 (N-term)	N_{aln}: 54 RMSD: 2.77 Z score: 4.5	N_{aln}: 49 RMSD: 2.27 Z score: 4.8	N_{aln}: 51 RMSD: 2.22 Z score: 4.2	N_{aln}: 46 RMSD: 2.81 Z score: 2.1	N_{aln}: 43 RMSD: 2.59 Z score: 3.9	N_{aln}: 40 RMSD: 2.98 Z score: 3.0	N_{aln}: 46 RMSD: 1.78 Z score: 5.7	N_{aln}: 48 RMSD: 2.76 Z score: 2.5		
TW1 gp56 (C-term)	N_{aln}: 53 RMSD: 2.29 Z score: 4.8	N_{aln}: 51 RMSD: 1.57 Z score: 6.7	N_{aln}: 52 RMSD: 1.47 Z score: 7.7	N_{aln}: 48 RMSD: 2.26 Z score: 4.0	N_{aln}: 42 RMSD: 1.84 Z score: 5.7	N_{aln}: 39 RMSD: 2.92 Z score: 2.9	N_{aln}: 50 RMSD: 1.92 Z score: 6.3	N_{aln}: 48 RMSD: 2.09 Z score: 6.7	N_{aln}: 52 RMSD: 2.46 Z score: 4.2	
MoeA	N_{aln}: 54 RMSD: 2.44 Z score: 4.7	N_{aln}: 48 RMSD: 2.53 Z score: 4.2	N_{aln}: 50 RMSD: 2.94 Z score: 4.3	N_{aln}: 47 RMSD: 3.13 Z score: 2.6	N_{aln}: 39 RMSD: 2.54 Z score: 4.6	N_{aln}: 44 RMSD: 2.67 Z score: 4.9	N_{aln}: 44 RMSD: 2.78 Z score: 4.6	N_{aln}: 45 RMSD: 2.45 Z score: 4.1	N_{aln}: 55 RMSD: 3.38 Z score: 2.2	N_{aln}: 54 RMSD: 2.50 Z score: 4.2

P74-26 gp87 is more stable than mesophilic homologs

We investigated whether the thermophilic decoration protein is more stable than its mesophilic counterparts. We first investigated whether gp87 forms a stable quaternary assembly in solution using size-exclusion chromatography/multi-angle light scattering (SEC-MALS). We find that gp87 is a stable trimer in solution at ~20 μM, with a measured molecular mass of 53 kDa (actual molecular mass of trimer is 49 kDa) (Figure 2.4A). This result contrasts with Lambda gpD, which is monomeric in solution even up to millimolar concentrations under similar solvent conditions (Imber et al., 1980; Yang et al., 2000). However, gpD crystallized as a trimer. Thus, P74-26 gp87 forms a more stable trimer than Lambda gpD.

We investigated the stability of the trimer by guanidine hydrochloride (GdnHCl)-induced unfolding, using both circular dichroism (CD) and tryptophan fluorescence spectroscopies (Figure 2.5). We measured unfolding across a range of protein concentrations (2.5 to 15 μM) to determine if trimerization is coupled with unfolding. Our CD and fluorescence data match each other quite well across all measured protein concentrations, with no hysteresis. We observe a single, sharp, and reversible transition between 3 M and 4 M GdnHCl with a midpoint of approximately 3.6 M GdnHCl (Figure 2.4B). In contrast, the midpoint of the Lambda gpD and P21 SHP unfolding curves are at 1.4 M and 1.1 M GdnHCl, respectively (Forrer et al., 2004), indicating that gp87 trimer is substantially more stable than its mesophilic homologs. The data from the CD

and fluorescence experiments were globally fit to a trimer to three unfolded monomers model ($N_3 \rightleftharpoons 3U$). Our attempts to fit the data to a three-state model failed to converge on a reasonable solution, indicating that trimer dissociation is coupled to unfolding. The estimated stability of the trimer was found to be 42.1 ± 1.3 kcal (mol trimer)$^{-1}$ with a denaturant dependence of 7.8 ± 0.4 kcal (mol trimer)$^{-1}$ M^{-1}. In contrast, the stability estimated for the gpD monomer is ~5.2 kcal (mol)$^{-1}$.

Similarities between Herpesvirus Triplex and phage decoration proteins

Our finding that the decoration protein fold is specified with very low sequence conservation led us to investigate whether this fold is found in other proteins, but has been overlooked by classical comparative genomics approaches. We first investigated whether this ancient fold may be found in the Herpesviruses, because Herpesviruses are thought to be direct descendants of tailed bacteriophage (Baines and Weller, 2005; Chen et al., 2011; Duda et al., 2013; Rixon and Schmid, 2014; Selvarajan Sigamani et al., 2013;). We focused on the Triplex complex of Herpesviruses, because both Triplex and phage decoration proteins are capsid components that bind and stabilize the three-fold axes of the capsid (Lander et al., 2008; Yu et al., 2017). Thus, we hypothesized that the Triplex proteins and phage decoration proteins may share an evolutionary origin.

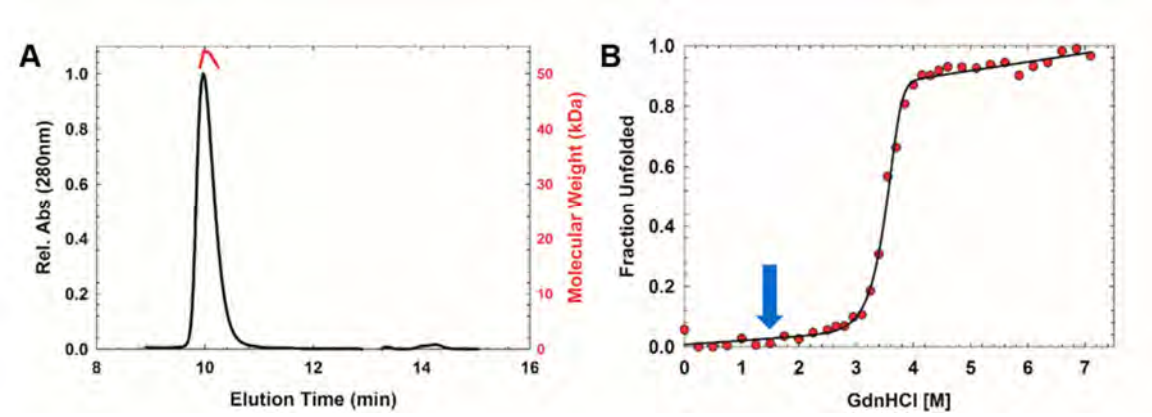

Figure 2.4. Thermophilic decoration protein has enhanced stability compared to mesophilic homologs.
A) P74-26 gp87 forms a stable trimer in solution as determined by size exclusion chromatography-multi angle light scattering (SEC-MALS). Predicted molecular mass: 49 kDa, measured molecular mass: 52 kDa.
B) Representative equilibrium fraction unfolding curve of P74-26 gp87 at 5 μM shows a steep unfolding transition from 3 M to 4 M GdnHCl; excitation = 295 nm, emission = 325 nm. The solid line represents the global fit to a model of trimer to three unfolded monomers. Blue arrow indicates the comparable transition midpoint of Lambda gpD unfolding.

We find that core β-tulip domain of phage decoration proteins is structurally similar to the core trimerization region of the Triplex proteins (Figure 2.6A,B). The Triplex complex is a heterotrimer consisting of one copy of Tri1 and two copies of the Tri2 protein. Both Tri1 and Tri2 contain a cryptic β-tulip domain centrally located in their structure. In both cases, the β-tulip domains are discontinuous, with large insertions in loops between strands of the tulip (Figure 2.6C). The β-tulip domains of Tri1 and Tri2 have the same basic size, shape, and topology as β-tulip domains of phage decoration proteins (C_α RMSD = 2.4 and 2.2, respectively compared to gp87; Figure 2.6A-C; Table 2). The β-tulip region is the only region of obvious structural similarity between Tri1, Tri2, and decoration proteins. In all three subunits of the Triplex trimer, the tulip domains participate in the primary trimerization interface (Figure 2.6D). Moreover, the Triplex β-tulip domains are arranged roughly parallel to the capsid shell, with the bloom end of the tulip as an interaction surface, similar to the phage decoration proteins. The N-terminal regions of Triplex proteins are pointed toward the capsid surface and make key interactions with the HK97 fold of MCP (Yu et al., 2017), in an analogous manner as phage decoration proteins. Both Tri1 and Tri2 contain a helical region inserted in between strands 3 and 4 of the β-tulip domain. This helical region is substantially larger in Tri2, which forms a ‘bracing arm’ that tightly dimerizes the two Tri2 proteins. Insertion of helical regions is common between strands 3 and 4 of the β-tulip domain, as seen in P74-26 gp87 and other β-tulip proteins (see below). These similarities in tertiary and quaternary

structure suggest a shared evolutionary history between the phage decoration proteins and the Triplex proteins of Herpesviruses.

β-tulip domains are general protein-protein interaction motifs enriched in viruses

We sought to determine if other proteins contain β-tulip domains. Unsurprisingly, sequence similarity search algorithms such as PSI-Blast (Altschul et al., 1997) found no significant homologs outside of decoration proteins or Triplex proteins. Because the β-tulip domain does not appear to be associated with a specific sequence motif, we searched for similar structures in the protein structure database using the programs DALI and VAST (Holm and Rosenström, 2010; Madej et al., 2014). Interestingly, β-tulip domains were identified in three seemingly unrelated proteins: 1) the anti-CRISPR protein AcrIIC1, 2) the tailspike protein gp12 of phage Φ29, and 3) the molybdenum metabolizing protein MoeA.

The anti-CRISPR protein AcrIIC1 consists of a single β-tulip with a small α-helical insertion (Figure 2.7A,B). This small protein is a broad spectrum Cas9 inhibitor and binds directly to the Cas9 HNH nuclease active site, preventing DNA cleavage (Harrington et al., 2017). Despite lacking any discernible sequence similarity, the β-tulip of AcrIIC1 is structurally very similar to that of P74-26 gp87, with a C_α RMSD of 2.6 Å (Figure 2.7A; Table 2). The β-tulip

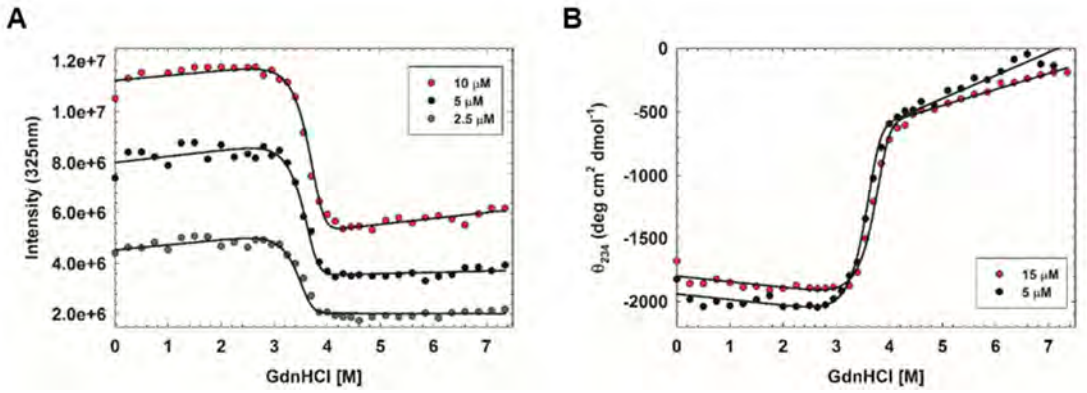

Figure 2.5. Equilibrium unfolding of P74-26 gp87.
A) Equilibrium unfolding of gp87 at monomer concentrations of 10 μM (red), 5 μM (black), and 2.5 μM (cyan) was monitored by tryptophan fluorescence; excitation 295 nm, emission 325 nm.
B) Equilibrium unfolding of gp87 at monomer concentrations of 5 μM (black) and 15 μM (red) monitored by circular dichroism at 234 nm. Solid lines represent the fit to a model of trimer to three unfolded monomers.

domains of both AcrIIC1 and gp87 contain a helical linker that connects strands 3 and 4 of the tulip. Unlike other β-tulip domains, AcrIIC1 is monomeric and interacts with its partner (Cas9) through the ‘stem’ end of the β-tulip (Figure 2.7C). Despite significant structure conservation between gp87 and AcrIIC1, gp87 does not inhibit NmeCas9 cleavage *in vitro* (Figure 2.8).

The Φ29 tailspike protein gp12 uses its tandem β-tulip domains (termed the D4 region) to direct assembly of the tailspike trimer (Xiang et al., 2009). Recently, Rossmann and colleagues described structural homology between the C-terminal portion of gp12 and the decoration protein gp56 from the marine siphovirus TW1 (Wang et al., 2017). The TW1 decoration protein contains N- and C-terminal β-tulip domains with significant structural similarity to other β-tulip containing proteins (Figure 2.9E-H, Table 2). Our structural comparison of gp12 with the decoration protein from P74-26 revealed that the D4 region of gp12 consists of two consecutive β-tulip domains forming intramolecular contacts in a bloom-to-stem fashion (Figure 2.9A-D). This arrangement is echoed in the quaternary structure, as the bloom end of the C-terminal β-tulip domain interacts with the stem end of the N-terminal tulip in the adjacent subunit of the trimer (Figure 2.9D). The N-terminal β-tulip domain has a small β-hairpin inserted between strands 4 and 5, which projects outward away from the ring of tulips. The molybdenum biosynthetic enzyme MoeA contains a C-terminal β-tulip domain that aids in oligomerization of MoeA (Schrag et al., 2001). As often found

in other β-tulip domains, the MoeA tulip contains a short α-helical linker between strands 3 and 4.

Discussion

Increased stability of a thermophilic decoration protein

P74-26 is a thermophilic virus found in hot springs between 60-75 ˚C. We sought to determine how the P74-26 virion maintains stability under these harsh conditions. We find that the P74-26 capsid is not stabilized by covalent cross-links, unlike HK97 phage that stabilizes the capsid through 'covalent chainmail' of the major capsid protein (Wikoff et al., 2000). Instead we hypothesize that gp87 acts as a decoration or 'cementing' protein to stabilize the capsid. gp87 is present in the virion with a stoichiometry expected for a decoration protein, and adopts similar tertiary and quaternary structure as mesophilic decoration proteins. Despite the similar structure and function, gp87 is widely divergent at the sequence level from other known decoration proteins. The low sequence homology may preclude identification of decoration proteins using classical comparative genomics approaches; structural and/or biochemical analyses may be necessary for proper annotation.

Our structural and biochemical characterization of gp87 provides insight into the mechanisms of capsid stabilization. Phage P74-26 thrives at 70˚ C, a temperature at which phage Lambda disassembles readily ($t_{1/2}$ ~5 min)(Bauer and Evilevitch, 2015). The P74-26 decoration protein is more stable than its mesophilic counterpart, gpD (Figure 2.4B; (Forrer et al., 2004)). Furthermore, isolated gp87 is a stable trimer in solution with a stability of 42.1 ± 1.3 kcal (mol

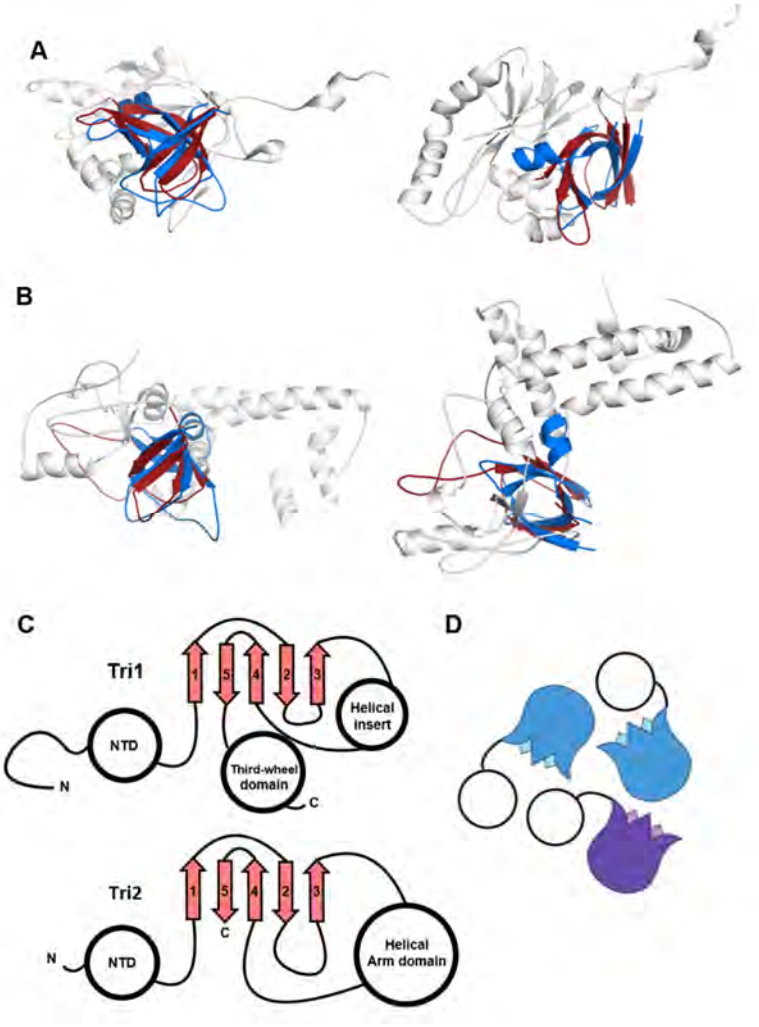

Figure 2.6. Structural similarity of phage decoration protein trimers and HCMV Triplex proteins.
A & B) The P74-26 gp87 β-tulip domain (blue) is similar to central domain of HCMV Tri 1 (A) and Tri 2 (B) proteins (grey β-Tulip domains in red, PDB: 5VKU).
C) Central domain of HCMV Tri 1 and Tri2 has conserved β-tulip topology.
D) HCMV Triplex proteins form an asymmetric trimer consisting of two molecules of tri2 (blue) and one molecule of tri1 (purple).

trimer)$^{-1}$, whereas gpD is monomeric in solution. Because our unfolding data show that gp87 unfolds in a single transition, we hypothesize that dissociation of the trimer and unfolding of monomers are tightly linked. Thus, we propose that gp87 has increased stability due to both tertiary and quaternary interactions.

Our analysis of the gp87 crystal structure reveals the likely mechanism for gp87 stabilization. Each subunit in the P74-26 decoration protein is larger, burying more hydrophobic surface within each protein's core as compared to the decoration proteins from phages Lambda and P21 (4470 Å^2, 3382 Å^2, and 3828 Å^2 for gp87, gpD, and SHP, respectively)) (Forrer et al., 2004). In particular, gp87 displays larger clusters of hydrophobic residues (I, L, V, and F) than gpD (Figure 2.10A-C). ILVF clusters have recently been hypothesized to be key determinants of folding and stability (Kathuria et al., 2016). The larger ILVF clusters in gp87 may increase the intrinsic stability of each subunit of the trimer. The ILVF clusters are more extensive in thermophilic gp87 than in mesophilic gpD. A gpD trimer contains six ILVF clusters, two per subunit. In contrast, a gp87 trimer has 9 ILVF clusters, three of which span from the β-tulip domain of one subunit to the C-terminal domain of the adjacent subunit (Figure 2.10A,B). Our structure also indicates that the interfacial interactions between decoration subunits increases stability of the complex. The interfaces of the trimeric assembly bury greater surface area in gp87 than SHP or gpD. The greater buried surface area is expected to afford greater stabilization of the trimer. Tighter interactions between

subunits will link trimerization to the unfolding of decoration protein. Indeed, our unfolding data exhibit a single transition, supporting this model.

The gp87 structure also reveals insights into how the decoration protein interacts with the capsid to stabilize the overall assembly. The N-terminal region of decoration proteins is known to interact directly with MCP, forming an extended beta-sheet across the HK97-fold E-loop (Lander et al., 2008). This structuring of the N-terminal region creates struts that stabilize the three-fold axes in the icosahedral shell. The N-terminal region of gp87 is substantially longer than those of mesophilic gpD and SHP (27, 19, and 21 residues long, respectively). The longer arm could provide a stronger foothold for the trimer into the capsid structure, stabilizing the three-fold axes. Moreover, the globular domains of gp87 are larger than those of its mesophilic cousins, which could allow more interaction surface between the core trimer and the MCP shell. Furthermore, we observe that the orientation of the gp87 subunit is different than observed for gpD and SHP, with each subunit rotated downward (toward the capsid shell) by ~20˚. The altered orientation could place the globular domains of gp87 in closer contact with the capsid shell to afford tighter contacts and greater stabilization. Finally, we identified several residues that are conserved across the known decoration proteins and we note that several of these residues appear to be poised to interact directly with the capsid shell. Interestingly, the two prime candidate residues for MCP interactions appear to interact directly with the capsid in cryo-EM structures of Lambda phage (Lander

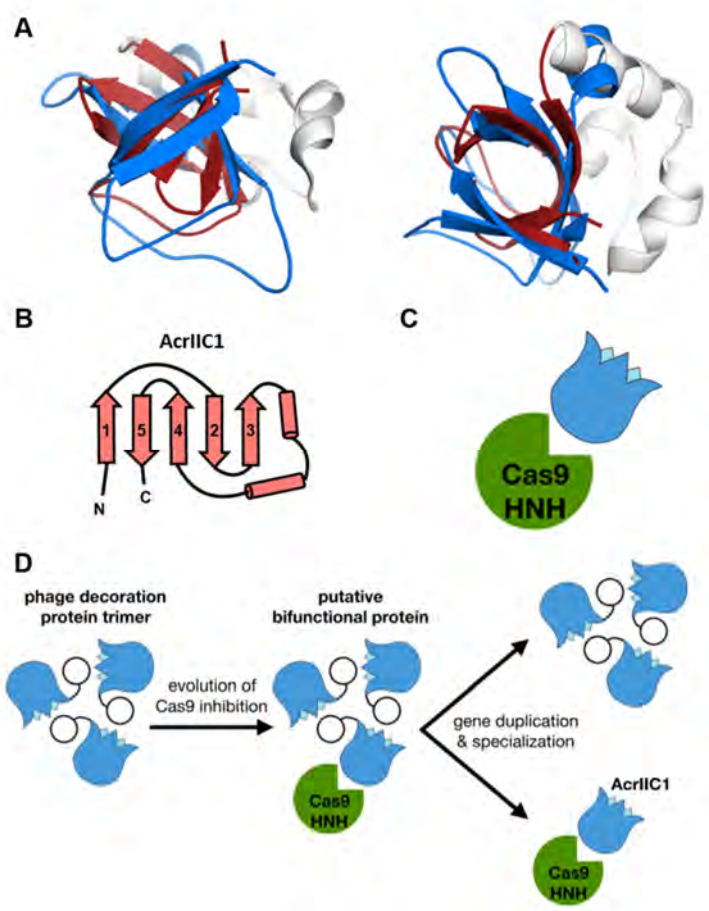

Figure 2.7. β-tulip domain suggests evolution of anti-CRISPR proteins from phage structural proteins.
A) Structural alignment of P74-26 gp87 β-tulip domain (blue) with the anti-CRISPR protein AcrIIC1 (β-tulip in red, PDB: 5VGB).
B) Topology diagram of AcrIIC1 reveals conserved β-tulip domain architecture.
C) Cas9 binds stem side of the AcrIIC1 β-tulip domain rather than the bloom side.
D) Decoration protein trimers form interactions with neighboring subunits through the 'bloom' end of the β-tulip domain, leaving the 'stem' end exposed. Anti-CRISPR AcrIIC1 binds to the Cas9 HNH domain through the stem end of the β-tulip domain and may have evolved from a bifunctional phage decoration protein.

et al., 2008). Further structural and biochemical analyses will be necessary to uncover whether these interactions are important for stabilization of the capsid shell.

Evolutionary relationship of Herpesvirus Triplex proteins and phage decoration proteins

Our analysis reveals that the β-tulip domain of decoration proteins is structurally and functionally conserved in Herpesvirus Triplex proteins. Although there is only 10% and 18% sequence identity between the β-tulips of P74-26 gp87 with HCMV Tri1 and Tri2, their similarities deepen our understanding of the evolutionary linkage between these two classes of viruses. The structures of the β-tulip domains are similar, with the same arrangement of β-strands in the tulip. Both decoration protein and Triplex β-tulip domains participate in trimerization, in each case with the 'bloom' side positioned at the interface. Additionally, the β-tulip domains arrange themselves in a similar manner. Both decoration proteins and Triplex complexes position themselves at the three-fold axes of the capsid icosahedron. In all cases, the C-terminal regions are pointed away from the capsid shell and the N-terminal region binds to the capsid surface, with direct interactions with the E-loop of the MCP HK97 fold (Lander et al., 2008; Yu et al., 2017). Moreover, even the variable regions appear to have significant similarities. For example, gp87, Tri1, and Tri2 have a helical region inserted between strands

3 and 4 of the β-tulip domain. In summary, the combined structural and organizational similarities illustrate the shared evolutionary history of Herpesvirus Triplex and phage decoration proteins.

For the past decade or more, evidence of the shared evolutionary history of Herpesviruses and Caudoviruses has accumulated (Rixon and Schmid, 2014). In particular, the MCPs share the HK97 fold, although those of Herpesviruses have complicated elaborations (Suhanovsky and Teschke, 2015; Wikoff et al., 2000). Both virus families also direct the assembly of the capsid using a scaffolding protein that is part of procapsid but not found in the infectious particle (Fane and Prevelige, 2003; Johnson, 2010). Furthermore, the genome packaging machinery and mechanism is largely conserved between the two classes of viruses (Bazinet and King, 1985; Hendrix, 1978; Hsiao and Black, 1978). Here we add to the similarities of these different classes of viruses, showing that the decoration proteins share an evolutionary history as well. Thus, the major genes involved in virus structure and assembly are conserved throughout Caudoviruses and Herpesviruses.

While a shared evolutionary history may imply mechanistic similarities between decoration and Triplex proteins, there are substantial differences. First, the Triplex proteins are not three-fold symmetric as in phage decoration proteins (Figure 2.2A; Figure 2.6D); the two Tri2 subunits grip each other in a tight embrace, while the Tri1 protein makes less extensive contact (Yu et al., 2017). The heterotrimeric nature of the Triplex complex has been hypothesized to allow

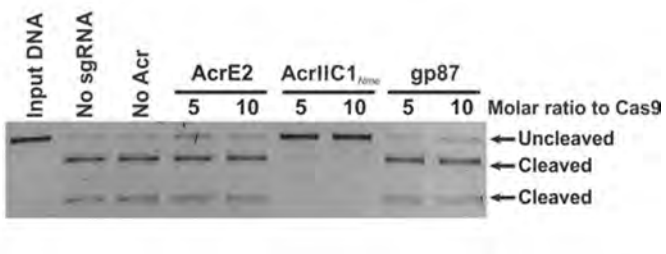

Figure 2.8. gp87 does not inhibit DNA cleavage by NmeCas9 *in vitro*.
Linearized plasmid DNA bearing a protospacer (target) next to a PAM sequence was subjected to *in vitro* cleavage by purified NmeCas9. NmeCas9 was pre-incubated with purified anti-CRISPR proteins AcrIIC1$_{Nme}$ and AcrE2 as a positive and negative control, respectively. Molar ratios of anti-CRISPR protein or gp87 relative to NmeCas9 are shown at the top of each lane. Input DNA and cleaved products are indicated with arrows on the right. The NmeCas9 cleavage assays shown are representative of three independent replicates.

each subunit in the complex to adopt a different role in capsid assembly (Saad et al., 1999). Decoration proteins attach to the capsid late in the assembly process after capsid expansion (Casjens and Hendrix, 1974; Dokland and Murialdo, 1993), while the Triplex proteins bind the procapsid early in assembly before DNA is packaged and stay bound throughout all stages of virus particle assembly (Trus et al., 1996). This difference in assembly may reflect the fact that the Triplex proteins embed between MCP capsomers and are an integral part of the capsid surface (Yu et al., 2017). This difference could be explained by the fact that the decoration proteins' interactions with capsid components are less extensive and the decoration protein binding site is only accessible after capsid expansion (Lander et al., 2008; Tang et al., 2006).

The shared evolutionary history suggests that Triplex proteins evolved from phage decoration proteins and, over the course of time, acquired a more integral role in capsid structure. As the Triplex formed the basis for the three-fold axes in Herpesvirus procapsids and capsids, additional domains were inserted into loops between strands of the β-tulip fold. We note that MCP also underwent domain insertions within loops of the HK97 fold during the course of evolution, suggesting a common mechanism for elaboration of the basic folds found in viral capsids. We also note that the Triplex's role as a necessary component of the procapsid required evolution of a separate mechanism for strengthening the capsid against the internal pressure of packaged DNA. In Herpes Simplex Virus 1, this is achieved by the protein UL25 (Sae-Ueng et al., 2014; Snijder et al.,

2017), which binds the penton vertices of the capsid. In the herpesvirus KSHV, the SCP protein cross-links MCP subunits within hexons (Dai et al., 2015).

Evolutionary origin of an anti-CRISPR protein

The anti-CRISPR protein AcrIIC1 is a broad-spectrum inhibitor of Cas9 nuclease activity. AcrIIC1 binds directly to the HNH nuclease domain of Cas9, preventing target cleavage (Harrington et al., 2017; Pawluk et al., 2016). Anti-CRISPR inhibitors of Cas9 have found use in biotechnology applications to decrease off-target DNA cleavage (Shin et al., 2017). Despite the importance of anti-CRISPR proteins, their evolution has remained mysterious (Pawluk et al., 2017).

Our discovery of the structural similarity of AcrIIC1 and phage decoration proteins suggests a shared evolutionary ancestor. AcrIIC1 consists primarily of a β-tulip fold with a strong structural similarity to the β-tulip domain of gp87 (C_α RMSD = 2.6 Å, Figure 2.7A; Table 2). Much like other β-tulips, AcrIIC1 contains a small helical insert between strands 3 and 4. However, we detect no significant sequence homology between AcrIIC1 and gp87 (6% sequence identity). This is not unexpected because the decoration proteins of Lambda and P74-26 phage share very little conservation, despite conserved function. Given the low sequence homology between β-tulip containing proteins, structure homology is much more useful than sequence similarity in identifying protein ancestry.

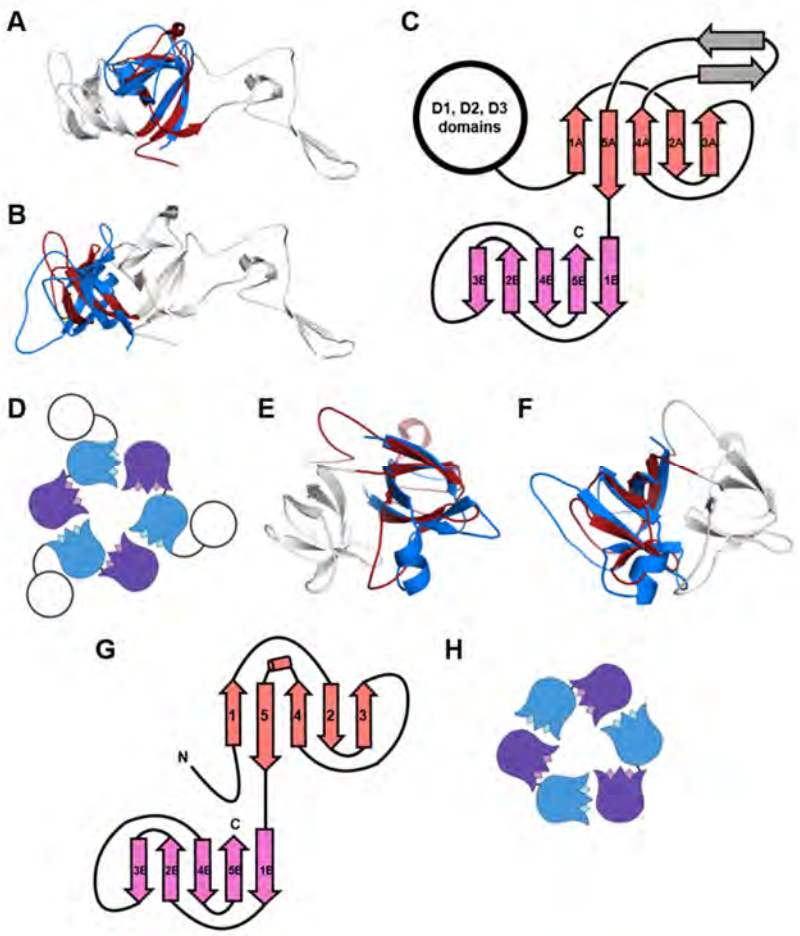

Figure 2.9. Conservation of the β-tulip domain in the decoration protein gp56 from the marine siphovirus TW1 and phage tail spike proteins.
A) Structure-based alignment of P74-26 gp87 β-tulip domain (blue) with the N-terminal β-tulip domain (red) of the D4 domain of Phi29 tail spike protein gp12.
B) Alignment of P74-26 gp87 β-tulip domain with the C-terminal β-tulip domain of gp12.
C) Topology diagram of Phi29 gp12 shows the orientation of the tandem β-tulip domains.
D) Model of the head-to-tail β-tulip domain orientation in gp12.
E) Structure-based alignment of gp87 β-tulip domain (blue) with the N-terminal β-tulip domain from TW1 gp56 (red).
F) Structure-based alignment of gp87 β-tulip domain (blue) with the C-terminal β-tulip domain from TW1 gp56 (red).
G) Topology diagram of TW1 gp56 shows the orientation of the tandem β-tulip domains.
H) Model of the β-tulip domain orientation in TW1 gp56.

Our findings suggest shared ancestry between AcrIIC1 and phage decoration proteins, but do not clarify which preceded the other. Although it is formally possible that AcrIIC1 is the evolutionary origin for phage decoration proteins, we find it more likely that phage decoration proteins predated the evolution of AcrIIC1. AcrIIC1 uses the stem side of its β-tulip domain to bind the Cas9 HNH domain (Harrington et al., 2017), in contrast to other β-tulip proteins that interact primarily through the 'bloom' side (Figure 2.7C). Thus, interaction with the Cas9 HNH nuclease could have evolved in a decoration protein without disrupting its central trimerization role. Gene duplication would then allow this putative bi-functional protein to specialize for each individual function (Figure 2.7D). One hurdle for this evolutionary pathway is that Acr proteins must act quickly upon phage infection to prevent CRISPR-mediated genome cleavage. However, decoration proteins are typically expressed late in infection (White et al., 2012) and typically remain outside the cell upon infection. How a decoration protein could be present during early stages of phage infection is not clear, although these proteins could be packaged with the genome and injected into the host cell, as is found for many phage proteins (Falco et al., 1977; Mullaney and Black, 2014; Strauss and King, 1984; Zhao et al., 2016). Alternatively, we propose that AcrIIC1 could have evolved in the context of a prophage, where the Cas9 inhibitory effect could have evolved for gene regulation or other functions (Bondy-Denomy et al., 2015).

The evolution of Anti-CRISPR proteins has been mysterious. There are currently over twenty unique families of Acr proteins that have been reported (Pawluk et al., 2017). These proteins are often small (50-150 residues) and have no clear sequence or structural features in common. Moreover, the mechanisms of Acr inhibition are diverse (Bondy-Denomy et al., 2015). Because of their remarkable diversity and small size, Acr proteins have been proposed to evolve *de novo* from small open reading frames (Pawluk et al., 2017). Here we show that the AcrIIC1 protein most likely evolved from the phage decoration protein fold. Future experiments (outside the scope of this paper) could investigate the evolutionary path between decoration and Anti-CRISPR proteins, perhaps by utilizing decoration proteins as phage display platforms (Yang et al., 2000).

Because β-tulip domains have extremely low or no sequence homology, we could only identify this relation by structural similarity rather than sequence motifs. We propose that some Acr proteins may have evolved from other phage proteins, but their structural homology has been masked by the rapid evolution and insertion of new structural elements. Regardless, *de novo* evolution remains a likely explanation for many Acr proteins. Recent work from the Cordes group has shown that proteins from young, ‘non-coding’ genes can have rudimentary ability to fold into compact, native-like states, suggesting a mechanism for evolution of new folds (Bungard et al., 2017).

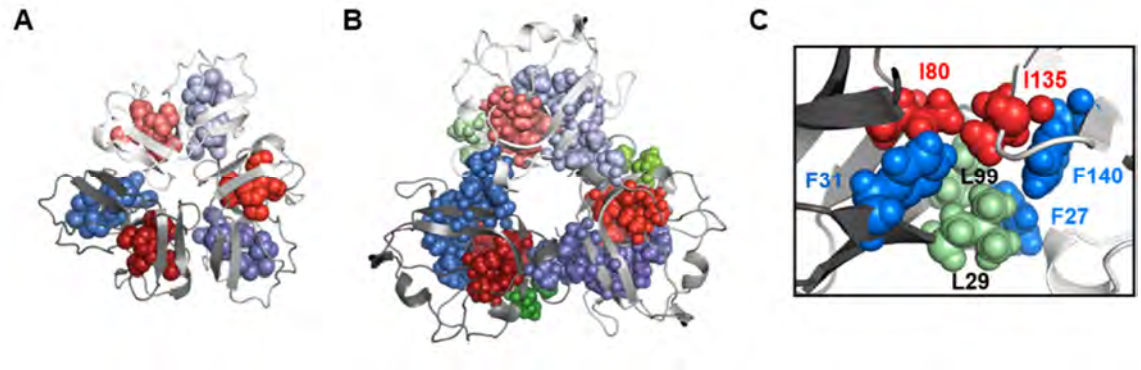

Figure 2.10. P74-26 gp87 forms a stable trimer through extensive hydrophobic network.
A) Hydrophobic ILVF clusters mapped onto the structure of Lambda gpD trimer show clusters within the β-tulip (blue) and C-terminal domains (red) that are unconnected to each other.
B) ILVF clusters mapped onto the P74-26 gp87 trimer show an extended hydrophobic network, with clusters forming intermolecular and intramolecular interactions (clusters shown in blue span multiple subunits).
C) gp87 β-tulip domains interact with neighboring subunits primarily through hydrophobic interactions, forming an intermolecular cluster consisting of Ile (red), Leu (green), and Phe (blue) residues.

The Beta-tulip fold as a widespread protein interaction motif

Although this structural motif is quite small, we could only identify β-tulip domains in a handful of proteins. There are other examples of 5-stranded β-barrels, (e.g. HIV gp120 (Pan et al., 2015)), but these domains use a different topology than found in β-tulip domains, suggesting that they evolved independently. The β-tulip fold is structurally conserved despite lacking strong sequence conservation. This is reminiscent of the HK97 fold (Suhanovsky and Teschke, 2015), as well as the well-studied Triose phosphate isomerase (TIM) Barrel fold (Nagano et al., 2002).

The β-tulip fold is enriched in viral proteins. Of the five classes of β-tulip proteins that we identified, four of them are viral in origin: phage decoration proteins, Herpesvirus Triplex proteins, phage tailspike proteins, and phage anti-CRISPR proteins. The only outlier is MoeA, a molybdenum cofactor biosynthesis enzyme that is found throughout life. Whether the MoeA β-tulip evolved through convergent or divergent evolution is still unknown. Regardless, we propose that the viral β-tulip proteins share a common ancestor.

Recently, Rossmann and colleagues have elucidated the structure of the capsid stabilizing protein gp56 from the marine siphovirus TW1, which shares structural homology with phage Lambda gpD and the tailspike protein from phage Φ29 (Wang et al., 2017). TW1 gp56 consists of two β-tulip domains arranged in bloom to stem orientation around the central axis of the trimer

(Figure 2.9H). The N-terminal β-tulip features a short α-helical linker between strands 4 and 5 (Figure 2.9G). This contrasts with the other β-tulip containing proteins, as α-helical linkers appear to be most common between strands 3 and 4. Given the similarity between gp56 and Φ29 tailspike, Rossmann and colleagues noted that in general the 'gp56-like fold' is capable of forming stable trimers in capsid stabilizing proteins, but also in proteins with differing functions (Wang et al., 2017).

One of the viral β-tulip proteins is the Φ29 phage tailspike protein, which forms a homotrimeric enzyme complex used for recognition and digestion of host cell wall structure (Moreno and Bluzat-Moreno, 1978). The D4 region of the tailspike protein contains a pair of β-tulip domains arrayed in a head-to-tail fashion around the trimerization interface. Both β-tulip domains have significant structure homology to P74-26 gp87 (C_α RMSD = 1.7 Å and 1.5 Å for N and C-terminal β-tulip domains of D4, respectively; Figure 2.9A-D; Table 2). The tandem β-tulip domains in D4 act as an 'autochaperone' to allow trimerization of the D1, D2 and D3 regions. Removal of D4 results in non-productive assembly of the tailspike trimer (Xiang et al., 2009), most likely due to kinetic traps in the folding of the highly interdigitated β-helix motifs found in the D1, D2, and D3 domains (Betts and King, 1999). (We note that an Adenovirus uses the tailspike β-helix fold as a decoration protein to stabilize the capsid three-fold axes (Menéndez-Conejero et al., 2017). Thus, both the β-tulip and β-helix regions of the tailspike have evolved to stabilize the capsid.) Interestingly, Rossmann and

colleagues have noted that the D4 region of gp12 shows significant sequence homology to the minor capsid protein VP260 of *Paramecium bursaria* chlorella virus 1 (PBCV1) and some bacterial adhesion proteins (Xiang et al., 2009). The major capsid protein of PBCV1 exhibits a jelly roll fold instead of the HK97 fold found in Caudoviruses and Herpesviruses (Nandhagopal et al., 2002). Therefore, we propose the β-tulip domain plays a role in capsid stabilization and molecular recognition across a broad swath of viral lineages, not just those of the HK97 fold viral lineage.

Our analysis reveals some general features of the β-tulip domain as a protein interaction module. The β-tulip domain tends to function in the context of a multimer, particularly trimers (gpD, gp87, Triplex, gp12, MoeA), although AcrIIC1 is an exception. β-tulip domains act as protein-protein interaction modules, with the primary interaction surface mediated by the ‘bloom’ side of the tulip, in particular strand 1. Again, AcrIIC1 is an exception. Finally, the β-tulip motif is quite malleable in terms of sequence as well as inserted structural elements. Insertions of auxiliary elements, in particular in between strands 3 and 4 of the tulip, are common. Insertion of extra domains into the fundamental fold appears to be a common feature of capsid coat proteins; for example, HCMV major capsid protein has multiple insertions in loops of its core HK97 fold (Yu et al., 2017). The alterations in sequence and structure of the β-tulip domain shown here provides an example of the flexibility of the fold, as well as highlights the challenges in uncovering this fold in other proteins using comparative genomics.

CHAPTER III

Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure

Preface

The work presented in Chapter III is adapted from a manuscript titled "Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure" that has been submitted for publication. A preprint version of the manuscript was published on bioRxiv on January 5th, 2019 (https://www.biorxiv.org/content/10.1101/473264v2).

Brian Kelch and I wrote the manuscript and developed the experimental design together. I performed phage purifications, cryo-EM sample preparation and optimization, data collection, data processing, 3D refinement, and structural analysis. Contributing author Dr. Gabriel Demo assisted with cryo-EM data processing and refinement and structural analysis. Emily Agnello performed the phage thermostability assay described in the text.

Introduction

Capsids are protein shells that surround and protect the viral genome. Capsid proteins often self-assemble into icosahedral structures with a quasi-equivalent arrangement of individual subunits. In quasi-equivalence theory (Caspar and Klug 1962), the capsid subunits use similar interactions throughout the assembly, but are arranged in slightly different ways in non-symmetry related environments.

Icosahedral capsid proteins form substructures called capsomers, with each capsomer comprised of either five or six subunits (pentons and hexons, respectively). Typical icosahedral capsids can be described by the triangulation number (T), which describes the complexity of the icosahedral symmetry. Capsids are composed of 60T protein subunits, resulting in an assembly of twelve pentons and a variable number of hexons (10 x (T-1)). Although capsid structure has been intensely studied for over a half century, the principles underlying capsid size and stability remain elusive.

Capsid size is typically controlled by the underlying capsid geometry: the triangulation number and whether the capsid is prolate or isometric. Most frequently, the triangulation number is altered to control capsid size. Higher triangulation number results in more hexons and, therefore, a larger capsid; lowering the triangulation number will result in a smaller capsid. Another common mechanism for increasing capsid capacity is to convert a spherical, isometric

capsid into an elongated icosahedron, referred to as a prolate capsid. A prolate capsid consists of a cylinder with two quasi-icosahedral caps at the ends, resulting in increased interior capacity. In many viruses, an interior scaffolding protein or domain dictates the assembly of homogeneous capsids with a defined geometry and size (Prevelige and Fane 2012; Dearborn et al., 2017). Small alterations to capsid protein primary sequence can change T number or can convert a capsid between isometric and prolate (Eiserling et al., 1970; Fiedler et al., 2012). Thus, the evolutionary barriers for altering capsid size through these mechanisms are quite low. It remains unknown if there are other mechanisms for capsid size control and whether these mechanisms have any ramifications for virus fitness, such as capsid stability.

Capsid stability is vital for virion survival in an external fluctuating environment until it finds a new host cell. Across the viral world, there is a wide spectrum of capsid stabilities (Mateu, 2013). Tailed phages and similar viruses utilize more stable capsids because of the high internal pressure of their tightly packaged genome (Evilevitch et al., 2003; 2004) and because the capsid never disassembles during the infection cycle (Casjens and Molineux, 2012; Davidson et al., 2012; Leiman and Shneider, 2012). Tailed bacteriophages, as well as herpesviruses, use Major Capsid Proteins of the conserved HK97 fold (Suhanovsky and Teschke, 2015). Capsids of these viruses first form as spherical procapsids, which convert to an icosahedral shape concomitant with genome packaging, release, or cleavage of auxiliary proteins. The conversion to

an icosahedral shape is mediated through conformational rearrangement of the HK97 folds (Hendrix and Johnson, 2011; Prevelige and Fane, 2012).

Thermophilic virus capsids are amongst the strongest because they survive in an especially harsh environment. Previous studies of thermophilic viruses have focused on capsids of various shapes including icosahedral (Veesler et al., 2013), filamentous (Dimaio et al., 2015; Kasson et al., 2017; Liu et al., 2018), helical (Ptchelkine et al., 2017), and lemon-shaped (Hochstein et al., 2018; Hong et al., 2015). However, for most of these viruses, close mesophilic homologs are not available, which makes it challenging to identify the structural mechanisms that underlie thermostability.

We use the thermophilic, tailed bacteriophage P74-26 to elucidate the structural mechanisms of thermostability (Hilbert et al., 2015; 2017; Stone et al., 2018). P74-26 is an especially long-tailed siphovirus that infects *T. thermophilus* (Minakhin et al., 2008; Yu et al., 2006) (Figure 3.1A). We use P74-26 to compare with similar lambdoid phages and other mesophilic Caudoviruses. Because stability mechanisms of lambdoid phages have been studied in great detail (Bauer and Evilevitch, 2015; Bauer et al., 2015; Hernando-Pérez et al., 2014; Lander et al., 2008; Qiu, 2012), this comparison presents a unique opportunity to identify mechanisms of capsid stability.

Here, we report the structure of the P74-26 capsid. A series of lassos topologically tether subunits together to stabilize the capsid through a novel

mechanism. The decoration protein forms a unique cage to lock the mature capsid in place. Finally, the structure reveals T=7 geometry despite a capsid capacity about twice that of typical T=7 Caudoviruses. P74-26 uses larger and flatter capsomers to achieve this larger capsid. Our work suggests that capsid geometry plays a critical role in virus stability.

Materials and Methods

Growth and Purification of P74-26 virions

Phage stock preparation and virion purification were performed as previously described (Minakhin et al., 2008; Stone et al., 2018). P74-26 phage was propagated in the host strain *Thermus thermophilus* HB8 using fresh overnight cultures grown at 65 °C in Thermus Medium (4 g L^{-1} yeast extract, 8g L^{-1} tryptone, 3g L^{-1} NaCl, 1 mM $MgCl_2$, 0.5 mM $CaCl_2$). 4 mL of P74-26 phage stock at $1x10^6$ plaque forming units per mL (PFU mL^{-1}) was combined with 6 mL of fresh *T. thermophilus* HB8 and incubated at 65 °C for 10 minutes. Reaction mixture was then inoculated into 1 L of Thermus Medium, and incubated at 65 °C while shaking for 4-5 hours. Lysate was then spun at 4,000 x g for 20 minutes, and supernatant was incubated with DNase I (2 Units mL^{-1}) and RNase A (1 µg mL^{-1}) for 1 hour at 30 °C. To precipitate virions, solid NaCl was added to 1M final concentration and Polyethylene Glycol MW 8,000 was added to a final concentration of 10% (w/v) while stirring. Phage stock was then incubated on ice overnight. The next day, precipitated phage stock was spun at 11,000 x g for 20 minutes at 4 °C. The resulting phage pellet was resuspended in 2 mL of 50 mM Tris pH 7.5, 100 mM NaCl, and 1 mM $MgSO_4$. Resuspension was supplemented with 0.4 g solid CsCl and added to a CsCl step gradient (2 mL steps each of 1.2, 1.3, 1.4, 1.5 g mL^{-1} CsCl and 1 mL cushion of 1.7 g mL^{-1} CsCl, in 50 mM Tris pH 7.5, 100 mM NaCl, 1 mM $MgSO_4$). Gradients were prepared in 12 mL ultracentrifuge tubes and spun in a Beckman SW40-Ti rotor at 38,000 RPM for

18 hours at 4 °C. Sedimented virions were isolated and dialyzed twice overnight into 2L of 50 mM Tris pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$ at 4 °C. P74-26 virions were then concentrated to $1x10^{12}$ PFU mL^{-1} for subsequent use in electron microscopy.

Thermal Stability Assay

Purified P74-26 virions were diluted in sample buffer (50 mM Tris pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$) to a final concentration of ~$1x10^8$ PFU mL^{-1} and incubated at 80º C for 0, 0.5, 1, 2, 4, or 6 hours in triplicate. Incubated virions were serially diluted and 100 µL aliquots of several dilutions were each mixed with 150 µL of fresh *T. thermophilus* HB8 (OD_{600} = 1.0). The adsorption reactions were then incubated at 65º C for 10 minutes and mixed with 2.5 mL molten Thermus Medium supplemented with 0.75% agar. The molten reaction mixture was then spread evenly on top of TR Gellan Gum plates (2g L^{-1} Yeast Extract, 4g L^{-1} Tryptone, 1g L^{-1} NaCl, 15g L^{-1} Gellan gum, 1.5 mM $MgCl_2$, 1.5 mM $CaCl_2$, pH 7.5). Once the soft agar solidified, plates were inverted and incubated overnight at 65 °C yielding a bacterial lawn with discernible phage plaques. Plaques were then counted from three plates for each incubation.

Negative-staining electron microscopy

Infections of *Thermus thermophilus* HB8 were performed as described above (growth/purification of P74-26 virions). Aliquots of the infection culture were diluted 1:1 in thermus growth medium and applied to carbon-coated 400-mesh grids one hour after infection. Excess sample was then blotted from the grid surface, and grids were finally stained with 1% uranyl acetate. Data was then collected on a 120kV Philips CM-120 microscope fitted with a Gatan Orius SC1000 detector.

Cryo-EM Specimen Preparation

400 copper mesh lacey carbon support film grids (Electron Microscopy Sciences) were glow discharged using a Pelco easiGlow (Pelco) for 45 seconds at 20 mA (negative polarity) prior to sample application. 3.5 µL of P74-26 virions at $1x10^{12}$ PFU mL^{-1} was applied to a cleaned grid at 22º C and 90% humidity in a Vitrobot Mark IV (FEI). Samples were then blotted for 8 seconds after a wait time of 15 seconds. Sample coated grids were then vitrified by plunging into liquid ethane. Prepared grids were then stored submerged in liquid nitrogen prior to data collection.

Data Collection

Micrographs were collected on a 300kV Titan Krios electron microscope (FEI) equipped with a K2 Summit direct electron detector (Gatan). Images were collected at a magnification of 130,000X in super-resolution mode at a pixel size of 0.529 Angstroms per pixel and a total dose of 48 e^{-} $Å^{-2}$ per micrograph. Micrographs were collected with a target defocus of -0.2 to -1.2 µm. For each micrograph, 32 frames were motion-corrected using the *Align Frames* module in IMOD (http://bio3d.colorado.edu/imod/betaDoc/man/alignframes.html) with 2x binning yielding a final pixel size of 1.059 Å/pixel. In total, 4,611 micrographs were obtained from a single 3-day collection.

Data Processing

CTFFind4 within the *cis*TEM software package (Grant et al., 2018) was used to determine defocus values for each resulting micrograph. An initial particle set of 37,046 particles was then picked using the *Find Particles* action in *cis*TEM with a specified maximum particle radius of 390 Å. This led to selection primarily of P74-26 capsids without selecting virion tail structures. The resulting particle stack was extracted using a box size of 1024 pixels, and then sorted into 20 classes using 20 cycles of 2D classification in *cis*TEM. Particles from 2D class averages containing complete icosahedral capsids were extracted, yielding a stack of 28,880 particles. The particle stack was then used to generate an *ab*

initio model for subsequent refinement using the *Ab-initio 3D* action in *cis*TEM. Finally, 4x binned, 2x binned, and unbinned particle stacks, and particle parameters were exported to the Frealign software package (Frealign version 9.11; (Grigorieff, 2007)) for subsequent 3D refinement and reconstruction with icosahedral symmetry imposed.

To speed up initial data processing, the 4x binned particle stack was used for 5 cycles of global alignment (mode 3) with a resolution range of 300-12 Å in Frealign using the *ab initio* generated model as a template. The 2x binned particle stack was then subjected to 12 rounds of refinement (mode 1) in Frealign, increasing the high-resolution limit each 3 cycles from 12 to 6 Å. The 2x binned particle stack was then subjected to 20 rounds of 3D classification using 4 total classes and a high-resolution limit of 6 Å. Two classes containing high quality particles were combined using *merge_classes*.exe in Frealign, yielding a particle substack containing 23,178 particles. The particle substack (1.059 Å/pixel) was then subjected to 8 cycles of refinement and reconstruction in Frealign mode 1, yielding a resolution of 3.3 Å (0.143 FSC cutoff). FSC curves were calculated in Frealign using odd and even particle half-sets. The map was sharpened using automatically calculated B-factors (-98 $Å^2$) using *bfactor.exe* in Frealign. The sharpened map was used for preliminary model building and structure refinement. The final cycle of refinement included Ewald sphere curvature corrected reconstructions of the same and opposite handedness by setting IEWALD in Frealign to either 1 or -1, respectively. The Ewald sphere

curvature corrected reconstruction of the same handedness yielded the final map with a resolution of 2.8 Å (0.143 FSC cutoff; 3.1 Å at 0.5 FSC cutoff). The final map was sharpened with a B-factor of -100 Å using *bfactor.exe* in Frealign. The final map was then used for final model building.

Model Building

For the P74-26 MCP atomic model, a poly-alanine chain was initially built into the density map and manually refined using Coot (Emsley and Cowtan, 2004). Then, side chains were assigned, and a complete model of MCP was built and refined manually. For the P74-26 decoration protein, the previously determined atomic model of the decoration protein trimer (PDB: 6BL5) was rigid-body fit into a corresponding portion of the density map. The N-terminal 16 residues of the decoration protein were then built into the density map, and the protein trimer was refined manually in Coot. The models of the MCP monomer and decoration protein trimer were then refined into the P74-26 density map using the Phenix real-space refine procedure (Adams et al., 2010). These models were then used to generate the P74-26 icosahedral asymmetric unit. The MCP P- and A-domains were first rigid-body fit into the density corresponding to each MCP subunit of the asymmetric unit (six hexon subunits and one penton subunit). Then, the N-arm, C-arm, and E-loop of each subunit were individually rigid-body fit into the corresponding density to account for variability in the orientation of these loops throughout the asymmetric unit. The refined decoration protein model was then fit into the corresponding density of the asymmetric unit,

comprising two decoration protein trimers and an additional monomer. The complete asymmetric unit was then manually refined in Coot, and final real-space refinement of the asymmetric model was performed using Phenix. The final real-space refinement consisted of 10 cycles of refinement using the 2.8-Å Ewald sphere corrected map as input. The real-space refinement statistics are listed in Table 3.1.

Database Accession Numbers

Model coordinates of the P74-26 capsid asymmetric unit were deposited to the Protein Data Bank (PDB: 6O3H). The density map of the icosahedral capsid reconstruction was deposited to the EM Data Bank (EMD-0618).

Table 3.1: Structure determination and refinement

Deposited structures	
PDB accession no.	6O3H
EMDB accession no.	EMD-0618
Data collection	
Microscope	FEI Titan Krios
Detector	Gatan K-2
Voltage (kV)	300
Magnification	130,000
Electron exposure ($e^-/Å^2$)	48
Defocus range (μm)	-0.2 to -1.2
Pixel size (Å)	0.529
Data processing	
Number of particles refined	28,880
Final number of particles	23,178
Imposed symmetry	Icosahedral
Map-sharpening B-factor ($Å^2$)	-100
Final resolution (Å)	2.8
Asymmetric unit refinement	
Map correlation (%)	80.1
R.M.S.D. (bonds)	0.01
R.M.S.D. (angles)	0.93
All-atom clashscore	5.79
Ramachandran favored (%)	92.2
Ramachandran allowed (%)	7.3
Ramachandran Outliers (%)	0.49
Rotamer outliers (%)	0.42
C-beta deviations	0

Results

Thermostability of P74-26 virions

We sought to elucidate the principles underlying the thermal stability of the P74-26 virus particle. Although it is clear that P74-26 is more stable than its mesophilic counterparts, the thermostability of P74-26 particles has never been directly measured. To address this, we heated samples of P74-26 virions to 80 °C and measured phage infectivity as a proxy of virus particle integrity (Figure 3.1B). This is a very stringent test for capsid stability, as disruption of many other structures within the virion (e.g. tail, neck, baseplate, etc.) can lead to loss of infectivity. Indeed, we find that P74-26 capsids remain intact after loss of tails and packaged DNA following incubation at 80 °C (Figure 3.1C). P74-26 virions remain stable and infectious when incubated at 80 °C for long periods ($t_{1/2}$ = 49.5 ± 0.7 minutes). In contrast, mesophilic Lambda phage rapidly loses infectivity when incubated at high temperatures in a similar buffer ($t_{1/2}$ = 48 sec at 75 °C and extrapolated $t_{1/2}$ ~ 5-7 seconds at 80 °C (Bauer and Evilevitch, 2015)). These experiments demonstrate that P74-26 is, to our knowledge, the most stable Caudovirus known.

Overall capsid structure

To determine the structural mechanism of P74-26 stability, we used single particle cryo-EM to determine the high-resolution structure of the P74-26 capsid.

In the raw images, the capsids are clearly isometric, icosahedral particles filled with DNA (Figure 3.2A,B). The ~0.9 μm long, flexible tails emanate from a five-fold vertex. We extracted capsid particles from the raw images and performed single particle reconstruction imposing icosahedral symmetry (Figure 3.3). Because the portal and tail complex sit at a unique vertex, icosahedral averaging removes features for these structures. The resolution of the reconstruction is 2.8 Å or 3.1 Å according to gold standard 0.143 or 0.5 Fourier Shell Correlation (FSC) criteria (Figure 3.2C; Figure 3.4A,B). (However, we note that the disordered DNA in the interior can adversely impact these calculations.) The P74-26 capsid is 824 Å in diameter from vertex to vertex, and 770 Å from face to face. The capsid exhibits T=7 (*laevo*) symmetry, which is surprising for a capsid of this size. We could easily identify density consistent with the major capsid protein (gene product 86, hereafter referred to as MCP) and the decoration protein (gene product 87, hereafter referred to as $Dec^{P74\text{-}26}$). We had previously determined the crystal structure of the trimeric globular regions of $Dec^{P74\text{-}26}$, which consists of a β-tulip domain followed by a mixed α/β subdomain (Figure 3.4C; Figure 3.5A) (Stone et al., 2018). The $Dec^{P74\text{-}26}$ crystal structure was easily placed into the cryo-EM maps with minor adjustments. Furthermore, we clearly resolve the entire ~23 residue N-terminal arm of $Dec^{P74\text{-}26}$ (hereafter called the Dec-arm), most of which was missing from the crystal structure. Likewise, the entire chain of MCP is clearly represented in the reconstruction maps. Thus, we have determined the complete structures, from N- to C-termini, of the two major

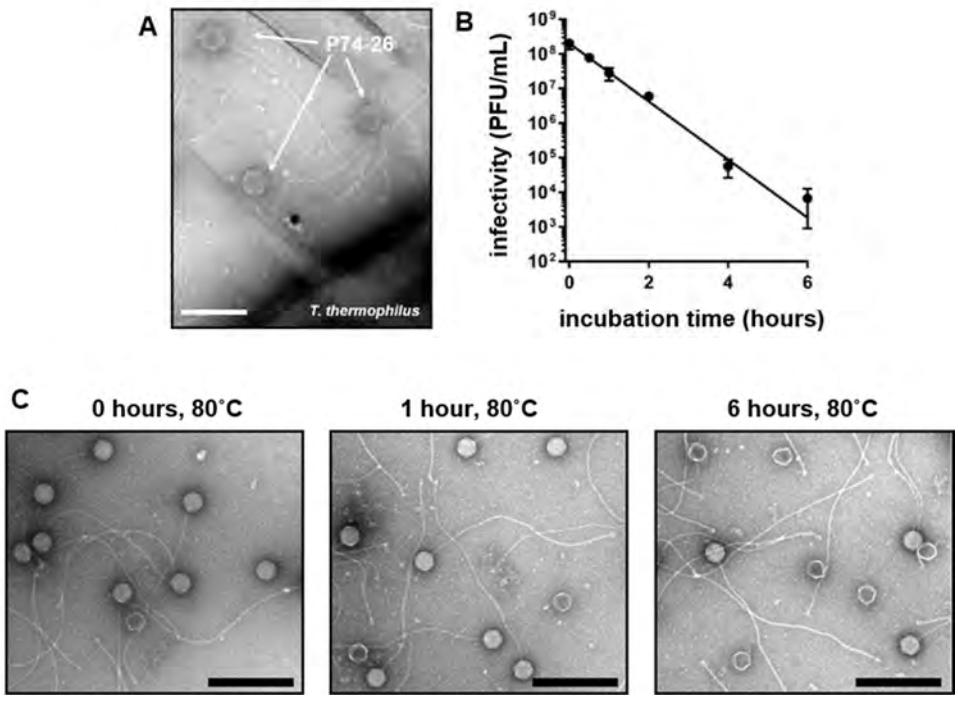

Figure 3.1. P74-26 is a thermostable virus.
A) Negative stain electron micrograph of phage P74-26 infecting *T. thermophilus* HB8; scale bar, 200 nm.
B) Purified P74-26 virions retain significant infectivity after 80˚C incubation.
C) Negative-staining electron micrographs of purified P74-26 virions following incubation at 80˚C, imaged after 0, 1, and 6 hour incubations (see main text figure 1); scale bars, 500 nm.

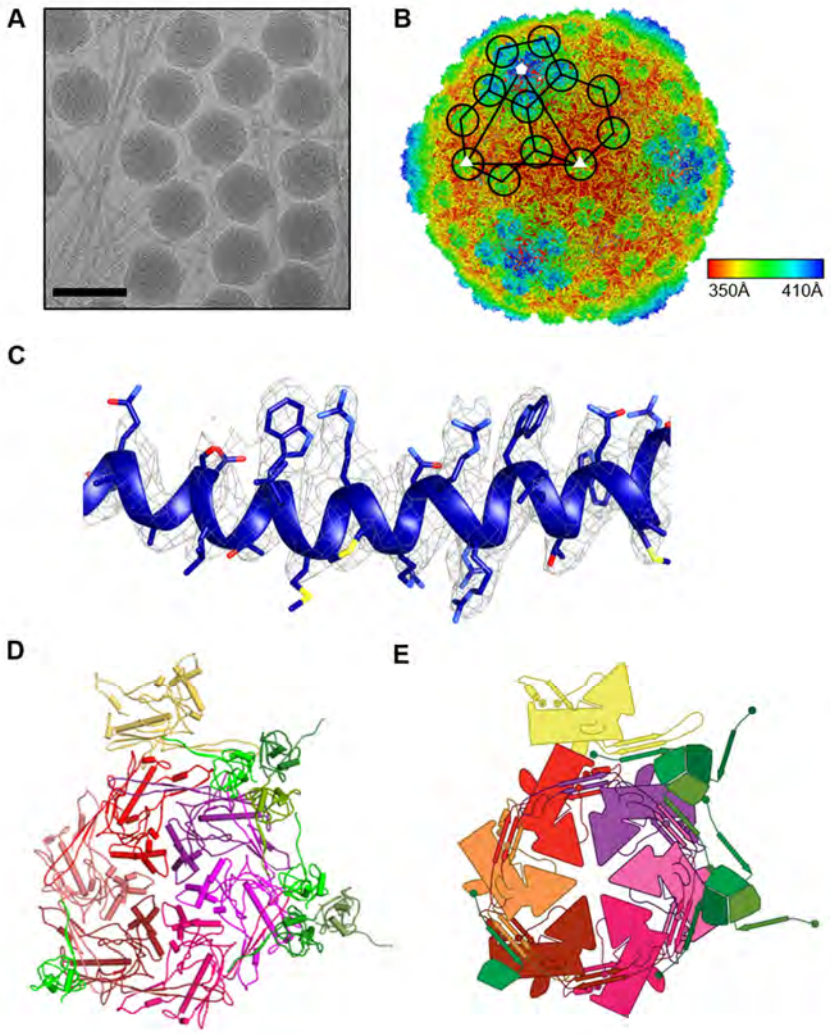

Figure 3.2. Determination of P74-26 capsid structure by cryo-EM.
A) Representative micrograph of purified P74-26 virions from high-resolution data collection; scale bar, 100 nm.
B) Icosahedral reconstruction of P74-26 mature capsid, colored radially from the center of the volume (red to blue). One penton and two hexons of MCP are outlined on the volume, and decoration protein trimers are circled. The black triangle indicates the icosahedral asymmetric unit, with three- and five-fold axes labeled as small white triangles and pentagons, respectively.
C) Representative section of the density map with fitted model of the MCP P-domain spine helix.
D and E) Atomic model of P74-26 icosahedral asymmetric unit (d) and corresponding cartoon schematic (e). In the asymmetric unit, the Major Capsid Protein (MCP) comprises one complete hexon (shown in shades of red) and a single penton subunit (yellow). The Decoration protein subunits (gp87) are shown in shades of green.

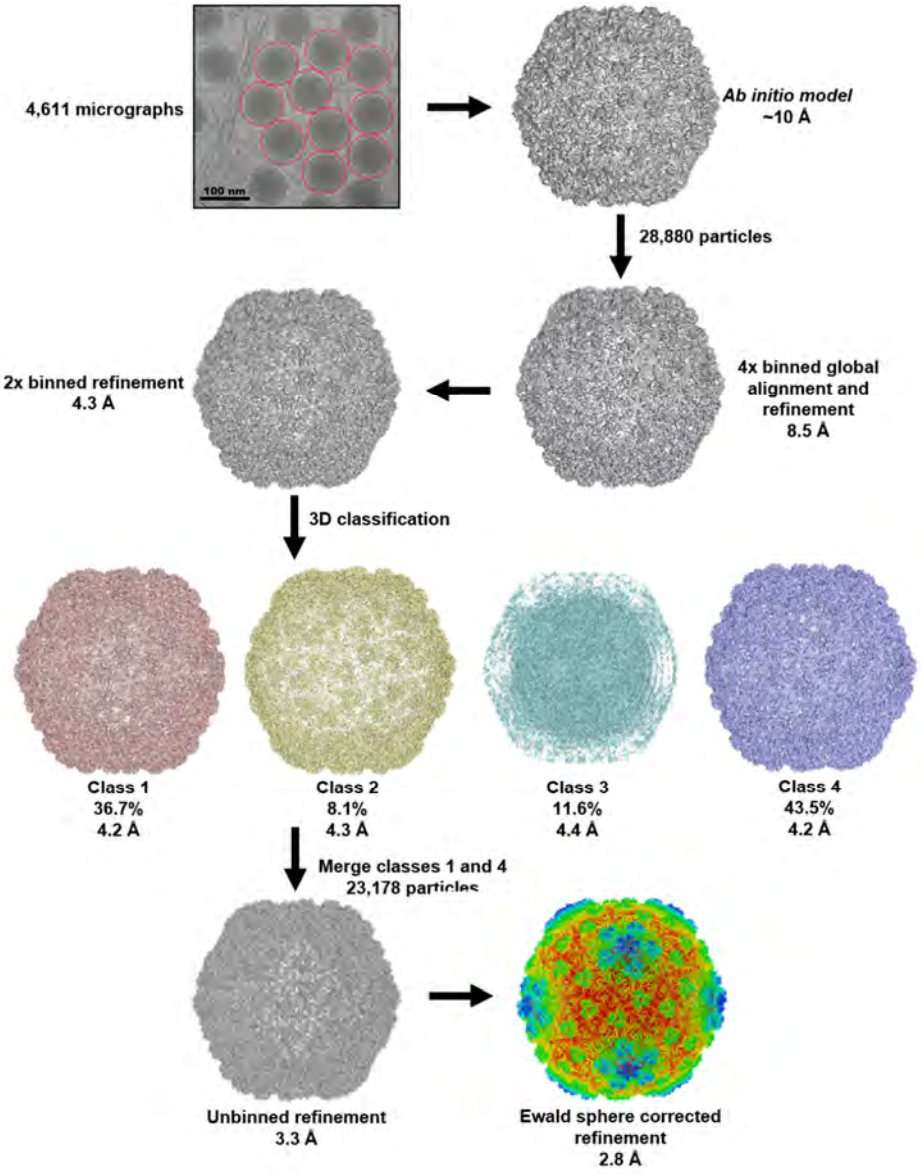

Figure 3.3. Schematic of P74-26 capsid cryo-EM refinement and classification. All particles were initially aligned to a single *ab initio* model generated from the data. 3D classification was performed with the 2x binned particle stack, and classes 1 and 4 were merged prior to final refinement of the capsid. Ewald sphere curvature correction was performed on the final round of refinement.

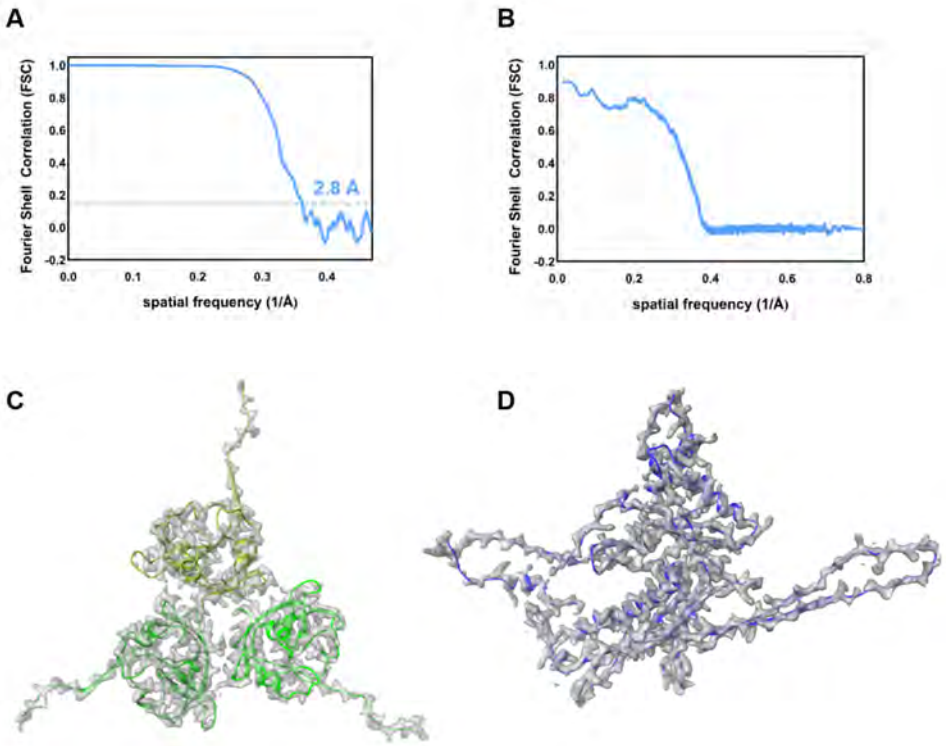

Figure 3.4. P74-26 icosahedral reconstruction structure validation.
A) Fourier shell correlation curve for the P74-26 icosahedral reconstruction; resolution is 2.8 Å at the gold standard FSC cutoff of 0.143 (dotted gray line).
B) Fourier shell correlation curve for the comparison of the cryo-EM map to the coordinate model of the icosahedral asymmetric unit.
C) Model of the $Dec^{p74\text{-}26}$ protein trimer fit into the corresponding map density.
D) Overall fit of the P74-26 MCP monomer model fitted to the corresponding map density.

components of the P74-26 capsid (see Table 3.1 for reconstruction and model statistics).

MCP adopts the expected HK97 fold (Figure 3.4D; Figure 3.5B;). The HK97 fold has two globular domains: the rectangular P-domain (peripheral domain; residues 107 through 186, and 326 through 373) and the triangle-shaped A-domain (axial domain; residues 192 through 320, and 384 through 391). Attached to these domains are a series of loops and extended arms that facilitate protein-protein interactions to stabilize the capsid. MCP contains an especially long N-terminal region (the N-arm; residues 1 through 52) and β-hairpin called the E-loop (extended loop; residues 60 through 106). MCP also contains four non-classical elements that are either unique to P74-26 or are not found in most HK97 folds: *1)* a G-loop region (residues 166-188) of the P-domain that forms a small beta-hairpin; *2)* an S-loop (residues 123-133) that lies at the N-terminal end of the P-domain spine helix; *3)* a T-loop (residues 330 through 346) that forms a flap off the bottom of the P-domain; and *4)* an extended C-terminal arm (the C-arm; residues 392 through 409) that extends away from the A-domain.

We observe conformational heterogeneity amongst the seven copies of MCP in the asymmetric unit that is restricted to the E-loop and N-arm (Figure 3.6A). The MCP conformer found in the penton exhibits larger conformational changes than the hexon subunits. The E-loop displays the largest variability in conformation with a ~27 Å movement of the tip (comparing the penton subunit to

the ‘D’ subunit of the hexon). The N-arm is the only other region that displays large heterogeneity across MCP conformers, with a ~19 Å movement of the N-arm tip. Both conformational changes can be described as rigid body rotations at hinge regions that lie at the bases of the E-loop and N-arm. There are minimal conformational differences throughout the rest of the HK97 fold. This is different than observed in other Caudoviruses such as phage Sf6, where there is conformational variability throughout the HK97 fold (Zhao et al., 2017).

The P74-26 capsid is exceptionally large for a T=7 virus

To our knowledge, the P74-26 capsid is far larger than observed in all other structures of T=7 viruses. Compared to other T=7 phage, the P74-26 capsid inner diameter is longer by ~140 Å than average, and 115 Å longer than the next biggest capsid (Figure 3.7A; Table 3.2). This results in a capsid capacity that is about twice as large as normal for T=7 Caudoviruses. The large capsid holds the 83 kb genome of P74-26, which is nearly twice as long as average for T=7 Caudoviruses. The coevolution of larger capacity and genome results in a typical packaging density (0.52 versus 0.54 bp nm^{-3} for P74-26 and other Caudoviruses, respectively).

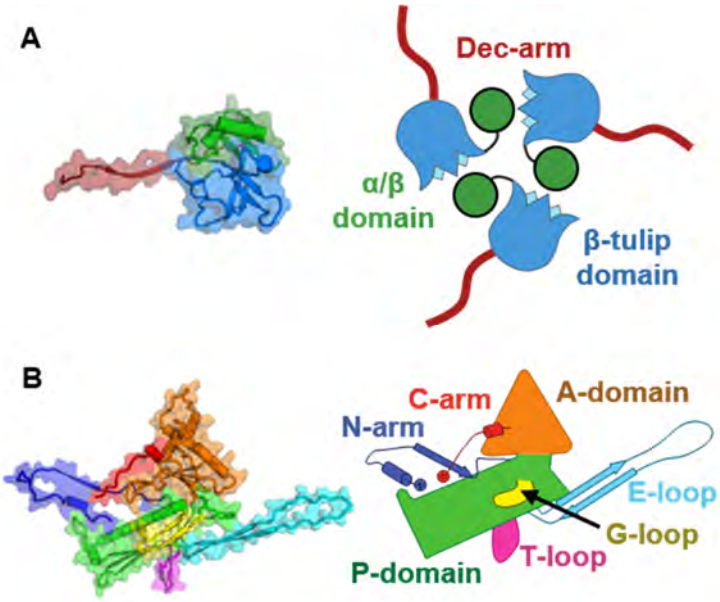

Figure 3.5. Structural models of P74-26 MCP and Dec.
A) Structure of P74-26 Dec protein monomer colored by domain (left) with corresponding schematic of the Dec trimer (right).
B) Structure of P74-26 MCP colored by domain (left) with corresponding labeled schematic (right).

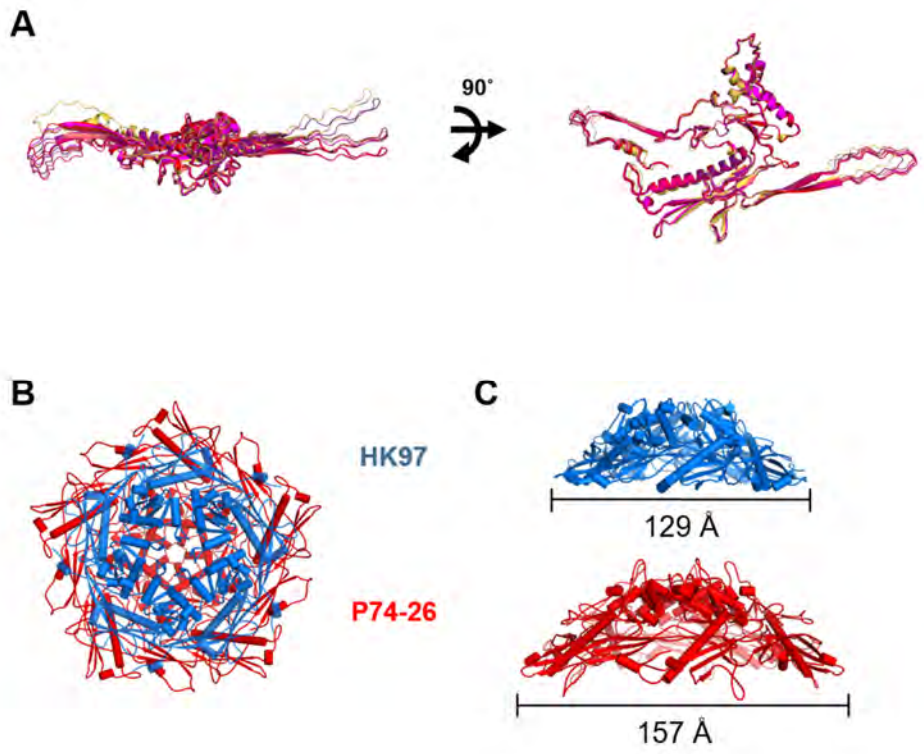

Figure 3.6. MCP subunit and capsomer orientation in P74-26.
A) Overlay of MCP subunits in the P74-26 asymmetric unit shows the conformational heterogeneity of the E-loop and N-arm lassos, and the conserved orientation of the A- and P-domains.
B) Overlay of P74-26 (red) and phage HK97 (blue) pentons displayed as a top-down view.
C) Side profile comparison of P74-26 and HK97 pentons, values indicate longest diameter of respective pentons.

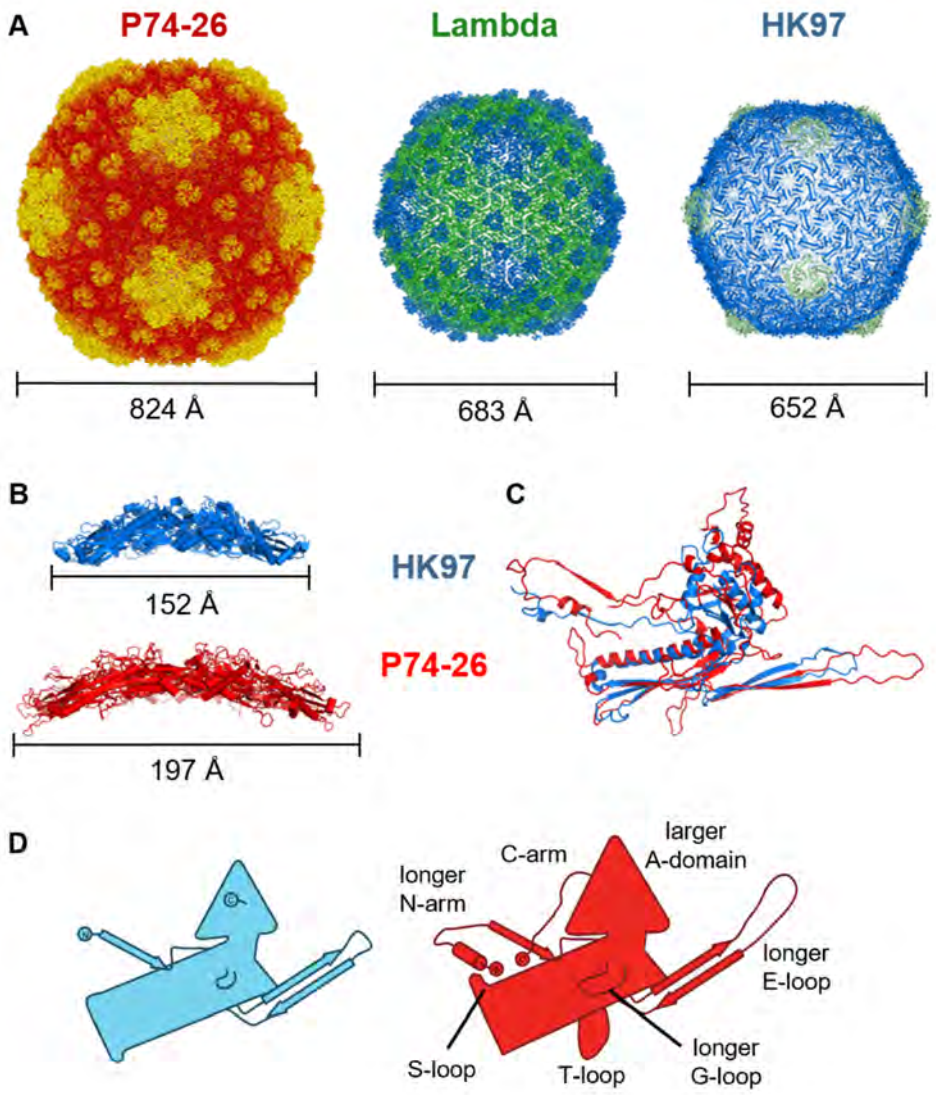

Figure 3.7. P74-26 has significantly increased capsid capacity compared to T=7 mesophilic viruses.
A) Size comparison of P74-26 and homologous mesophilic T=7 phages Lambda and HK97. Corresponding capsid diameters (longest outer diameter measured from 5-fold axes) are listed below each structure.
B) Size comparison of hexons from HK97 (blue) and P74-26 (red); longest hexon diameters are listed below each structure.
C) Overlay of HK97 and P74-26 MCP, aligned on the HK97 fold P-domain.
D) Schematic representations of MCP from HK97 (left, blue) and P74-26 (right, red). Labels indicate features unique to P74-26 MCP.

Table 3.2: Sizes of T=7 phage

Virus	Outer Diameter (Å) †	Inner Diameter (Å) ‡	Genome Size (kb)	Decoration Proteins	Packaging density (bp/nm^3)	PDB/EMDB accession code
80α	618	488	43.9	none	0.72	6C21
ATCC_Clear	620	518	29.7•	none	0.41	3JB5
BPP-1	676	526	42.5	2-fold	0.56	3J4U
CW02	670	550	49.4	3-fold	0.57	3J1A
HK97	652	520	39.7	none	0.54	1OHG
Lambda	683	547	48.5	3-fold	0.57	EM-5012
P22	686	534	41.7	none	0.52	5UU5
P-SSP7	654	508	45	none	0.66	2XD8
Sf6	690	532	39	none	0.49	5L35
SPP1	676	556	45.9	hexon center	0.51	4AN5
SYN5	664	538	46.2	none	0.57	4BML
T7	664	522	39.9	none	0.54	3J7X
TW-1	700	558	39.9	3-fold	0.44	5WK1
ε15	668	536	39.7	2-fold	0.49	3C5B
HSTV-2§	738	608	68.2	3-fold	0.58	EM-2235
P74-26	**824**	**673**	**83.3**	**3-fold**	**0.52**	**6O3H**

† Outer diameter is defined as twice the length of the longest radius from the map center to the capsid shell.

‡ Inner diameter is defined as twice the length of the shortest radius from the map center to the capsid shell.

• Genome length reported from homolog propionibacterium phage PA6.

§ HSTV-2 is an archaeal virus.

The larger size of the T=7 capsid is due to much larger and flatter capsomers than found in other Caudoviruses. The P74-26 pentons and hexons are 157 Å and 197 Å in diameter in the longest dimension, larger than normal capsomers (average of 129 Å and 158 Å, respectively). The P74-26 capsomers are also less curved than found in typical Caudoviruses, even though the spherical factor of the P74-26 capsid particle is higher than homologs (approximately 0.4 for P74-26 vs. 0.25 for HK97) (Figure 3.6B,C; Figure 3.7B).

P74-26 MCP occupies much more surface area than observed in other HK97 folds. The MCP of P74-26 contains a typical number of residues (Table 3.3), so this is not simply due to lengthening the MCP primary sequence. Instead, P74-26 efficiently uses extended loop architecture to create a structure with large surface area. There are eight structural changes that increase the surface area of MCP. The first change is the P74-26 E-loop β-hairpin, which is 44 residues in length, whereas the range of E-loops in other HK97 folds is 32 to 34 residues. The greater number of residues allows the E-loop to extend much further (~70 Å vs ~53 Å for P74-26 and HK97 phage, respectively). The second change is an elongated N-arm (52 residues extending ~44 Å, versus 36 residues and ~37 Å extension, averaged across other Caudoviruses). The N-arm of P74-26 also reverses direction to directly contact the P-domain. This unique N-arm architecture covers much more surface area than other N-arms. The third change is the P74-26 C-arm, which forms a unique extended structure that runs along the outside of the A-domain. The C-arm binds two adjacent A-domains within a

capsomer such that the axial region of the capsomer is enlarged (Figure 3.2D,E; Figure 3.8A). The A-domain surface area is increased by the fourth and fifth major structural changes: a unique helix at the tip of the A-domain that extends the axial dimension by ~6 Å, and an extended loop region (residues 295 through 311) that widens the A-domain (Figure 3.8C,D). The final three major structural changes are three loops that emanate from the P-domain (the G-loop, S-loop, and T-loop). Each of these three regions of the P-domain play critical roles in intra- or inter-capsomer interactions, which we delineate below.

MCP forms rings, 'lassos', and flaps to topologically link subunit-subunit interactions

The MCP-MCP interactions within a capsomer are much more extensive and intricate than in other Caudoviruses (Figure 3.9A). The interaction area between two adjacent MCP proteins is ~3150 $Å^2$, which is larger than most other T=7 HK97 folds. Only phage Sf6 buries more surface area (~3277 $Å^2$), much of which is contributed by the large insertion domain that makes intra-capsomer interactions in lieu of a decoration protein (Zhao et al., 2017). The hydrophobic contribution to interface stability is greater for P74-26 than other T=7 Caudoviruses. The estimated hydrophobic interaction free energy between two MCP subunits is -34.1 kcal mol^{-1} for P74-26 versus an average of -25.0 kcal mol^{-1} for other viruses (Table 3.3). However, the number of hydrogen bonds and salt

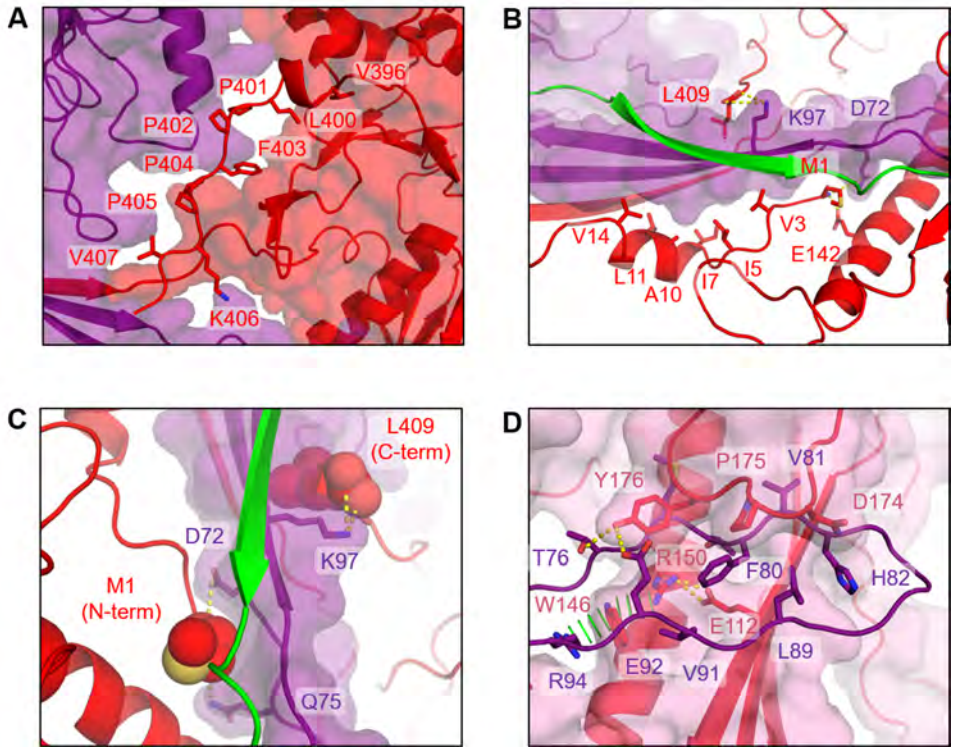

Figure 3.8. Intra-capsomer interactions in P74-26.
A) The C-terminal arm forms hydrophobic interactions with both the A-domain of the MCP subunit from which it emanates (red), as well as an adjacent subunit (purple) stabilizing inter-capsomer interactions.
B) The MCP N- and C-termini (red) both interact with the E-loop β-sheet consisting of the E-loop (purple), N-arm (red) and Dec-arm (green). The C-arm interacts on the top of the β-sheet, while the N-arm binds along the bottom. Thus, the two arms collaborate to form a 'pincer'-like interaction to hold the neighboring E-loop in place.
C) The MCP N- and C-termini (shown as spheres) use ionic interactions to act as 'pincers' clasping the E-loop β-sheet from either side.
D) Hydrophobic interactions, hydrogen bonds, and salt bridges tightly secure the E-loop (purple) to the P-domain and G-loop of a neighboring subunit (red).

Table 3.3: Analysis of inter-subunit interactions across T=7 phages

Phage	MCP length (residues)	Average interface area (Å²)†	$\Delta G_{hydrophobic}$ (kcal/mol) ‡	N_{HB}•	N_{SB}§	Structure resolution (Å)	PDB accession code
P74-26	409	3150	-34.1	34.3	3.8	2.8	6O3H
HK97	282	2346	-22.3	31.8	9.5	3.4	1OHG
Sf6	423	3277	-31.9	36.2	4.8	2.9	5L35
TW1	352	2449	-21.4	30.2	5.3	3.6	5WK1
P22	430	3230	-24.4	41.8	9.3	3.3	5UU5
BPP-1	331	2416	-18.4	27.0	12.5	3.5	3J4U
T7	398	2861	-29.9	25.3	3.8	3.5	3J7W
ATTC_Clear	315	2226	-26.6	22.8	1.5	3.7	3JB5

† Calculated as difference in total accessible surface areas of isolated and interfacing subunits within a hexon divided by two, averaged over all six hexon interfaces.

‡ Average estimated hydrophobic interaction energy between two subunits within a hexon, averaged over all six hexon interfaces. ΔGhydrophobic indicates the solvation free energy gain upon formation of the interface, in kcal/mol. The value is calculated as difference in total solvation energies of isolated and interfacing structures.

• Number of hydrogen bonds between two subunits within a hexon, averaged over all six hexon interfaces.

§ Number of salt bridges between two subunits within a hexon, averaged over all six hexon interfaces.

bridges stabilizing intra-capsomer interactions are typical (34 versus 31 hydrogen bonds, and 4 versus 7 salt bridges for P74-26 and others, respectively).

The architecture of the interfaces between MCP subunits is substantially different for P74-26, with loops and extended arms that are intertwined to provide stability against internal pressure. Like other HK97 folds, the E-loop forms important subunit-subunit interactions within the capsomer. However, there are two unique networks of interactions that form topological links to establish tight E-loop binding.

The Dec-arm, N-arm, and C-arms all collaborate to form a ring that completely surrounds the β-sheet region of the E-loop. The Dec-arm forms an anti-parallel β-sheet with the E-loop of a neighbor, forming the outer surface of the capsid (Figure 3.9A), similar to that of phage Lambda (Lander et al., 2008). The N-arm consists of four sections: 1) a ‘hand’ that comprises the N-terminal four residues, 2) a helical ‘forearm’, 3) a turn that forms the ‘elbow’ or ‘crook’, and 4) the ‘upper arm’ that forms a β-strand conformation. The hand, forearm, and upper arm all form intimate contacts with the E-loop of an adjacent MCP. The upper arm forms a parallel β-sheet with the E-loop along the inner surface of the capsid shell, similar to that observed in nearly all other HK97 folds (Suhanovsky and Teschke, 2015). However, the P74-26 N-arm is unique in that the elbow interacts with a neighboring Dec protein (see below for more details) and makes a sharp turn to orient the forearm and hand underneath the E-loop β-sheet. The forearm helix and hand interact with the E-loop using tightly packed hydrophobic

interactions, with the N-terminus amine group forming salt bridges with the P-domain (Glu142) and the E-loop of the neighbor (Asp72) (Figure 3.9C).

P74-26 MCP contains a C-arm that stabilizes inter-subunit interactions and, to our knowledge, is completely absent in other HK97 folds. The 18-residue C-arm extends down the length of the A-domain, ending at the E-loop of the neighboring MCP. The C-arm uses hydrophobic interactions to glue the A-domains of two adjacent MCP subunits together. The C-terminal four residues bind directly on top of the E-loop of a neighbor, interacting with all four strands in the sheet (Figure 3.9D). The four sidechains form a series of hydrophobic and salt bridge interactions to stabilize the C-arm over the sheet, with the C-terminal carboxylate forming salt bridges with Arg12 of the Dec-arm and Lys97 of the E-loop. Thus, both N- and C-arms grip the neighboring E-loop in a pincer architecture, with the C-arm binding from above, while the N-arm binds from below (Figure 3.8B,C). All three arms act to completely surround the E-loop of each neighbor in a ring of tight interactions, stabilizing the inter-subunit assembly (Figure 3.9E). To our knowledge, this type of architecture is unique.

The end of the E-loop attaches to the neighboring MCP. The E-loop is intercalated by several residues from the neighboring subunit (Glu112, Trp146, and Arg150 of the P-domain and Tyr176, Pro178, and Asn179 of the G-loop), preventing it from forming a canonical beta-hairpin structure (Figure 3.8D; Figure 3.9B). The long G-loop is critical for this interaction, as it allows these residues to interdigitate between the E-loop strands. The interaction between the P-domain

and G-loop of one subunit with the E-loop of its neighbor is mediated by numerous hydrophobic interactions, hydrogen bonds, and salt bridges (Figure 3.8D). In this manner, the E-loop acts as a lasso to topologically attach to the next subunit within the capsomer.

The N-arm also topologically links multiple proteins within the capsid assembly (Figure 3.10A). The elbow interacts tightly with the P-domain of a neighboring MCP, and the β-tulip domain and Dec-arm of a nearby $Dec^{P74\text{-}26}$ (Figure 3.10B). The elbow-to-Dec interaction is critical for stabilizing inter-capsomer interactions (described in more detail in the following section). The curve of the elbow directs the forearm and hand back toward the same HK97 fold from which the N-arm emanates (Figure 3.10A). The P-domain contains the S-loop, a small loop of 11 residues that inserts into a groove between the forearm helix and the hand residues. These structural elements are locked together using a complicated web of hydrogen bonds, salt bridges, and hydrophobic interactions that latches the forearm and hand in place (Figure 3.10C). In this manner, the S-loop effectively closes the N-arm into a long, closed loop that links different capsid subunits together. Taken together, we observe two separate lasso interactions emanating from each side of MCP: one consisting of the E-loop, and the other comprised of the N-arm/S-loop combination.

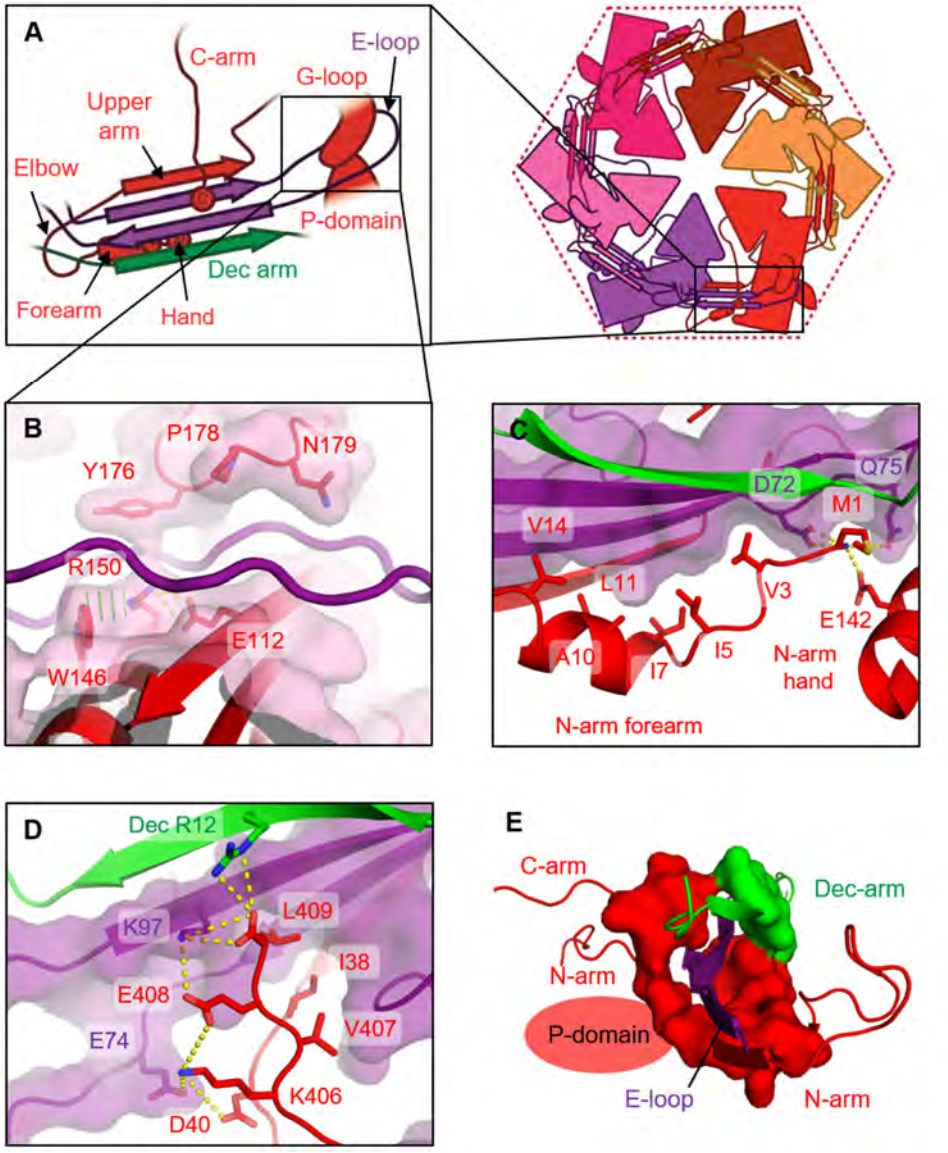

Figure 3.9. Lasso interactions stabilizing the E-loop.
A) Diagrams depicting interactions of the MCP E-loop (purple) with a neighboring MCP subunit (red) and Dec-arm (green).
B) The E-loop lasso (purple) is stabilized by the G-loop and P-domain of a neighboring subunit (red).
C) The N-arm 'forearm' and 'hand' bind underneath an adjacent E-loop.
D) The C-arm binds on top of all four strands of the E-loop β-sheet.
E) The β-hairpin of the E-loop (purple) is stabilized by a closed 'ring' consisting of the MCP N- and C- arms (red), and the Dec-arm (green).

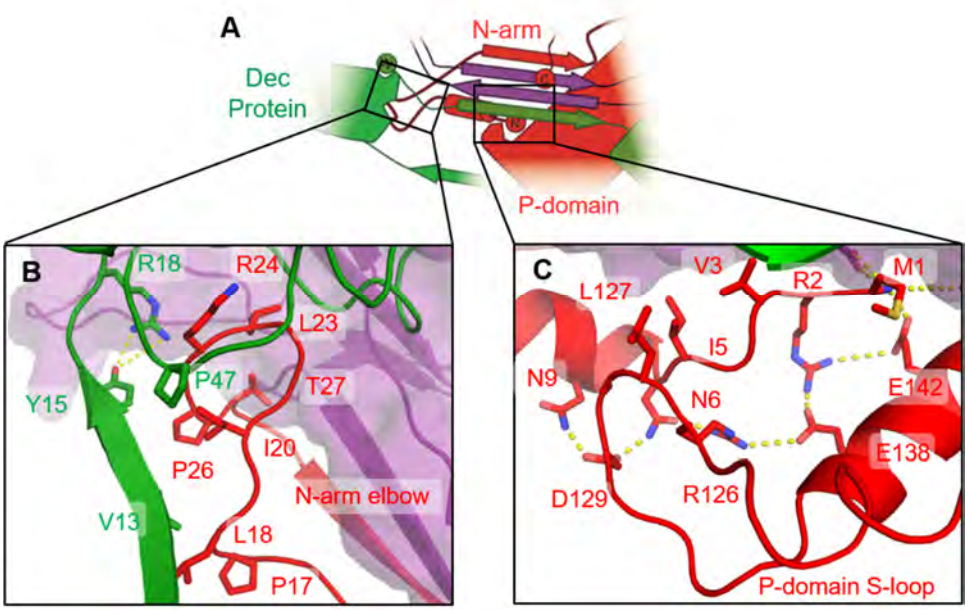

Figure 3.10. Lasso interactions stabilizing the N-arm.
A) Diagram depicting the N-arm lasso interactions with the P-domain (red) and adjacent Dec^{P74-26} protein (green).
B) The N-arm 'elbow' (red) extends to the β-tulip domain of a Dec^{P74-26} protein (green) at an adjacent three-fold axis.
C) The N-arm 'hand' binds to the P-domain S-loop, creating a closed loop lasso.

The N-arm forms overlapping flaps that stabilize the assembly at the two-fold and quasi-two-fold axes of symmetry between capsomers (Figure 3.11A,C,D). The N-arm elbow region of an MCP subunit in one capsomer interacts with the S-loop, forearm, and hand of the MCP subunit in the capsomer across the two-fold/quasi-two-fold axis. The interaction is primarily mediated through hydrophobic interactions. Importantly, the sidechain of Leu127 in the S-loop sits inside the opening of N-arm such that it acts as a 'hitching post' that the N-arm wraps around (3.13A). The elbow is placed on the outside face of the capsid, while the forearm and hand form the inside surface of a neighboring capsomer. Thus, both MCP proteins form two inter-locking flaps across the two-fold and quasi-two-fold axes.

MCP uses the T-loop to stabilize the interaction where three capsomers meet (i.e. the three-fold and quasi-three-fold axes) (Figure 3.11A). (We define these capsomers as α, β, and γ for clarity.) The T-loop forms the inside surface of the capsid, where it binds the P-domain of an MCP in a neighboring capsomer in a 'tongue-in-groove' fashion (Figure 3.11B,E). The interactions are reciprocated such that the T-loop of capsomer α binds the P-domain of capsomer β, and so on. Thus, the elbows form two-fold flaps on the capsid exterior to stabilize the two-fold and quasi-two-fold axes, while the T-loop forms tongue-in-groove interactions on the capsid interior to stabilize the three-fold and quasi-three-fold axes.

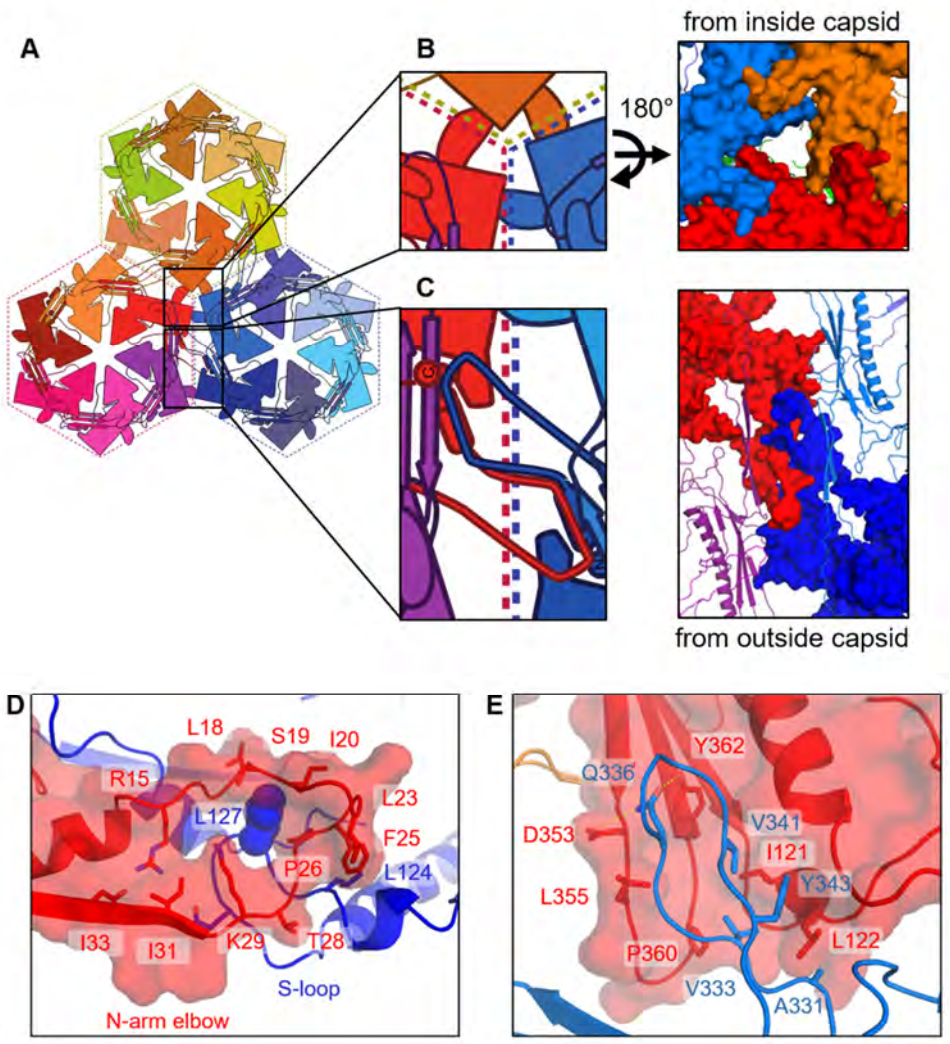

Figure 3.11. Inter-capsomer stabilization by extended flaps in P74-26 MCP.
A) Diagram of the three-fold axis where three capsomers (α-red, β-blue, and γ-yellow) meet.
B) The T-loop flap interacts with the P-domain of a neighboring subunit, forming a ring around the quasi-three-fold/three-fold axes. Right, tongue-in-groove interactions of T-loops and adjacent P-domains, as viewed from the capsid interior.
C) The N-arms of two adjacent subunits form interleaving flaps where the N-arm binds to the S-loop of the neighboring subunit across the quasi-two-fold/two-fold axes.
D) Leu 127 from the P-domain latch (dark blue) acts as a 'hitching post' for the N-arm lasso of a neighboring MCP subunit (red) across the quasi-two-fold/two-fold axes.
E) The 'tongue-in-groove' interaction of the MCP T-loop (light blue) and an adjacent MCP P-domain (red) at the quasi-three-fold/three-fold axes.

The decoration protein forms an inter-capsomer cage

The $Dec^{P74\text{-}26}$ trimer binds the capsid at the icosahedral three-fold and quasi-three-fold axes, with the Dec-arm pointing to the neighboring three-fold/quasi-three-fold axis. $Dec^{P74\text{-}26}$ forms numerous interactions throughout the capsid. Each $Dec^{P74\text{-}26}$ subunit contacts nine different proteins: six MCP subunits in three different capsomers, two other $Dec^{P74\text{-}26}$ subunits within the same trimer, and one $Dec^{P74\text{-}26}$ subunit at the neighboring three-fold/quasi-three-fold axis (Figure 3.12A). The total interaction area for a single 146-residue $Dec^{P74\text{-}26}$ subunit is ~4100 $Å^2$. This is vastly more extensive than seen for mesophilic decoration proteins. For example, phage TW1 has a decoration protein of similar size (149 residues), but each Dec^{TW1} subunit buries only ~60% of the equivalent surface area (~2670 $Å^2$) and interacts with seven other protein chains. Additionally, $Dec^{P74\text{-}26}$ uses much greater hydrophobic interactions than mesophilic homologs (estimated free energy of hydrophobic interactions is -38 versus -9 kcal mol^{-1} for P74-26 and TW1, respectively; Table 3.4).

The majority of $Dec^{P74\text{-}26}$ interactions are with capsomer α, where it contacts two separate MCP chains, termed α1 and α2. The $Dec^{P74\text{-}26}$ globular region primarily makes contact with the P-domain and G-loop of the α1 subunit. The $Dec^{P74\text{-}26}$ to P-domain interaction surface is comprised of a series of hydrophobic interactions and salt bridges (Figure 3.13A). The interface with the α2 subunit is the most substantial, accounting for ~1300 $Å^2$ of interaction area. Most of this interaction is due to the previously mentioned interaction of the Dec-

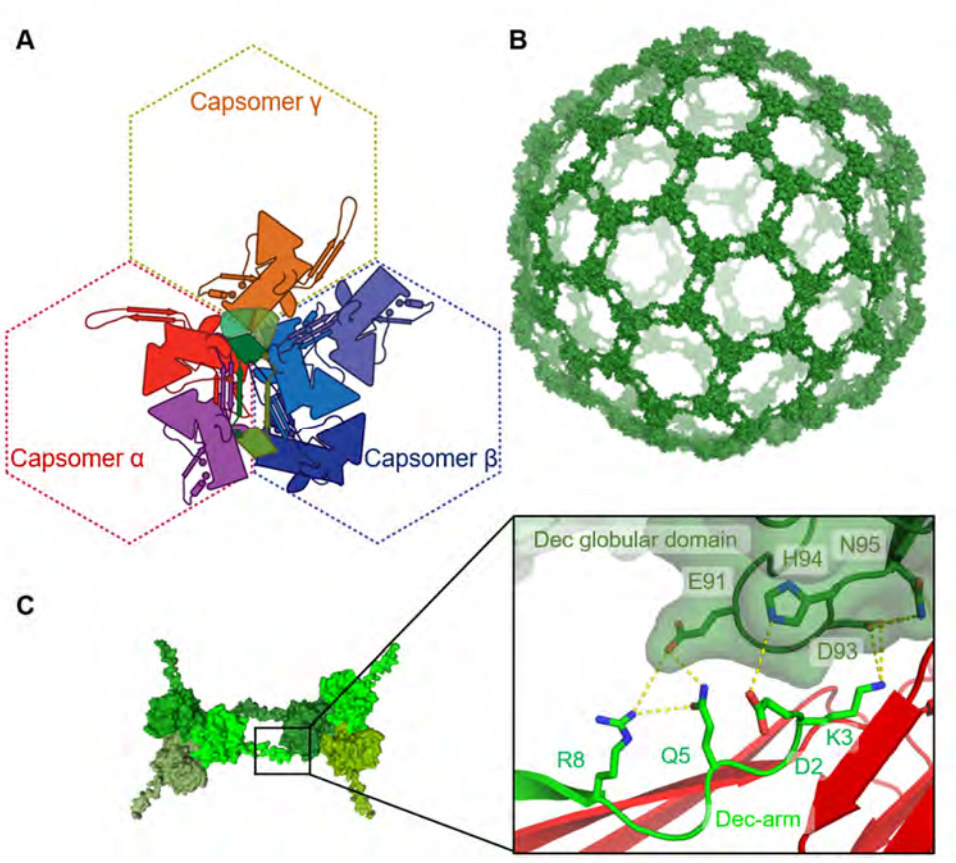

Figure 3.12. Stabilization of the P74-26 capsid by Dec^{P74-26}.
A) A single Dec^{P74-26} protein monomer (dark green) stabilizes capsomer interfaces by binding nine different proteins: three Dec^{P74-26} subunits (green) and six MCP subunits from three capsomers.
B) The Dec^{P74-26} proteins form an interconnected 'cage' that surrounds the capsid.
C) The Dec-arm from one subunit (light green) interacts with the β-tulip domain of a Dec^{P74-26} protein (dark green) positioned at an adjacent quasi-three-fold/three-fold axis.

Table 3.4: Analysis of decoration protein interactions in P74-26 and phage TW1

Phage	Dec length (residues)	Average interface area (Å²)†	$\Delta G_{hydrophobic}$ (kcal/mol) ‡	N_{HB}•	N_{SB}§	No. of proteins contacted	Structure resolution (Å)	PDB accession code
P74-26	146	4162	-38	41	24	9	2.8	6O3H
TW1	149	2676	-9	34	13	7	3.6	5WK1

† Calculated as the total interaction surface of a decoration protein monomer with adjacent MCP and Dec subunits.

‡ Total estimated hydrophobic interaction energy between Dec and binding partners, averaged from three Dec proteins within Dec trimer. ΔGhydrophobic indicates the solvation free energy gain upon formation of the interface, in kcal/mol. The value is calculated as difference in total solvation energies of isolated and interfacing structures.

• Total number of hydrogen bonds between Dec and interacting partners, averaged from three Dec proteins within trimer.

§ Number of salt bridges between Dec monomer and interacting proteins, averaged from three Dec proteins within trimer.

arm with the E-loop (Figure 3.10A). However, additional contacts are formed between the Dec-arm and the α2 G-loop (Figure 3.13B).

The Dec-arm reaches to the $Dec^{P74\text{-}26}$ trimer at the neighboring three-fold/quasi-three-fold axis. The N-terminal seven residues of the Dec-arm use hydrogen bonds and salt bridges to bind the neighboring $Dec^{P74\text{-}26}$ β-tulip domain in an induced-fit mechanism (Figure 3.12C). This same Dec subunit reciprocates this interaction using its Dec-arm, to create two proteins embracing each other 'at arm's length'. These interactions effectively link each of the $Dec^{P74\text{-}26}$ trimers throughout the capsid surface into one interlocked cage (Figure 3.12B). To our knowledge, this interconnected architecture of decoration proteins has not been observed in mesophilic viruses.

Each $Dec^{P74\text{-}26}$ globular region also interacts with the other two neighboring capsomers, contacting three MCP subunits in capsomer β and one in capsomer γ. The β1 MCP subunit uses its elbow region to bind $Dec^{P74\text{-}26}$, burying 449 $Å^2$ of area. The G-loop of the β2 MCP subunit binds the side of the $Dec^{P74\text{-}26}$ β-tulip domain, using hydrogen bonds and hydrophobic interactions, burying 257 $Å^2$ (Figure 3.14A). The β3 MCP subunit binds underneath the $Dec^{P74\text{-}26}$ globular region, using the tip of the E-loop to make hydrogen bonds (Figure 3.14B). Finally, γ1 MCP makes a few hydrogen bonds to $Dec^{P74\text{-}26}$ using its P-domain (Figure 3.14C).

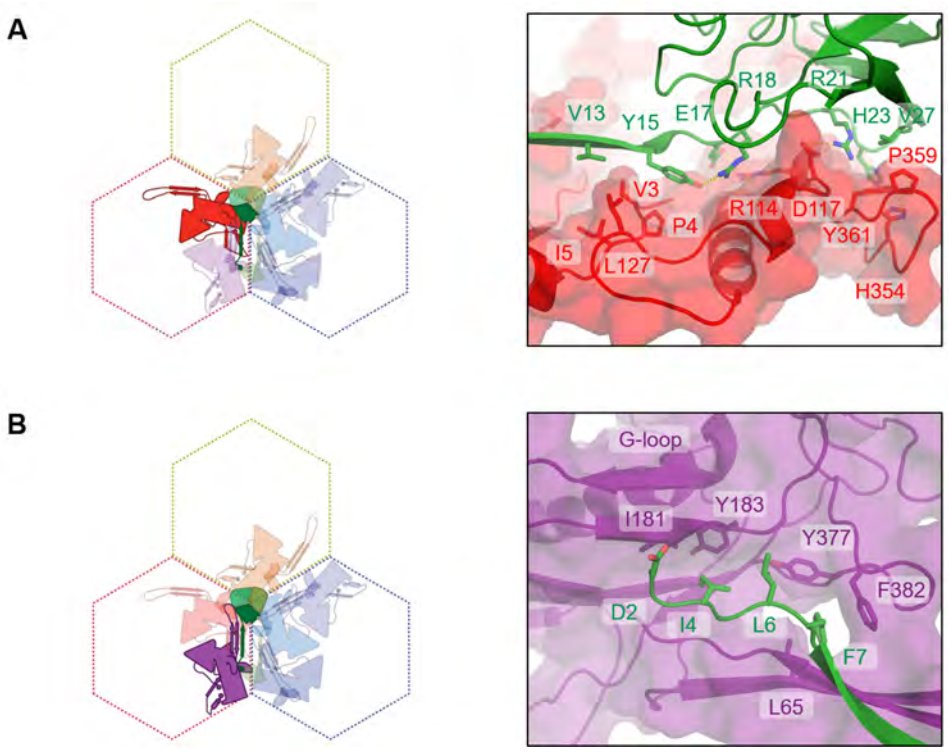

Figure 3.13. Inter-capsomer interactions of Dec with α capsomer in P74-26.
a) The β-tulip domain of Dec (green) interacts extensively with the P-domain of the α1 subunit (red).
b) The Dec-arm (green) extends beyond the E-loop β-sheet, interacting with the G-loop and A-domain of the α2 subunit (purple).

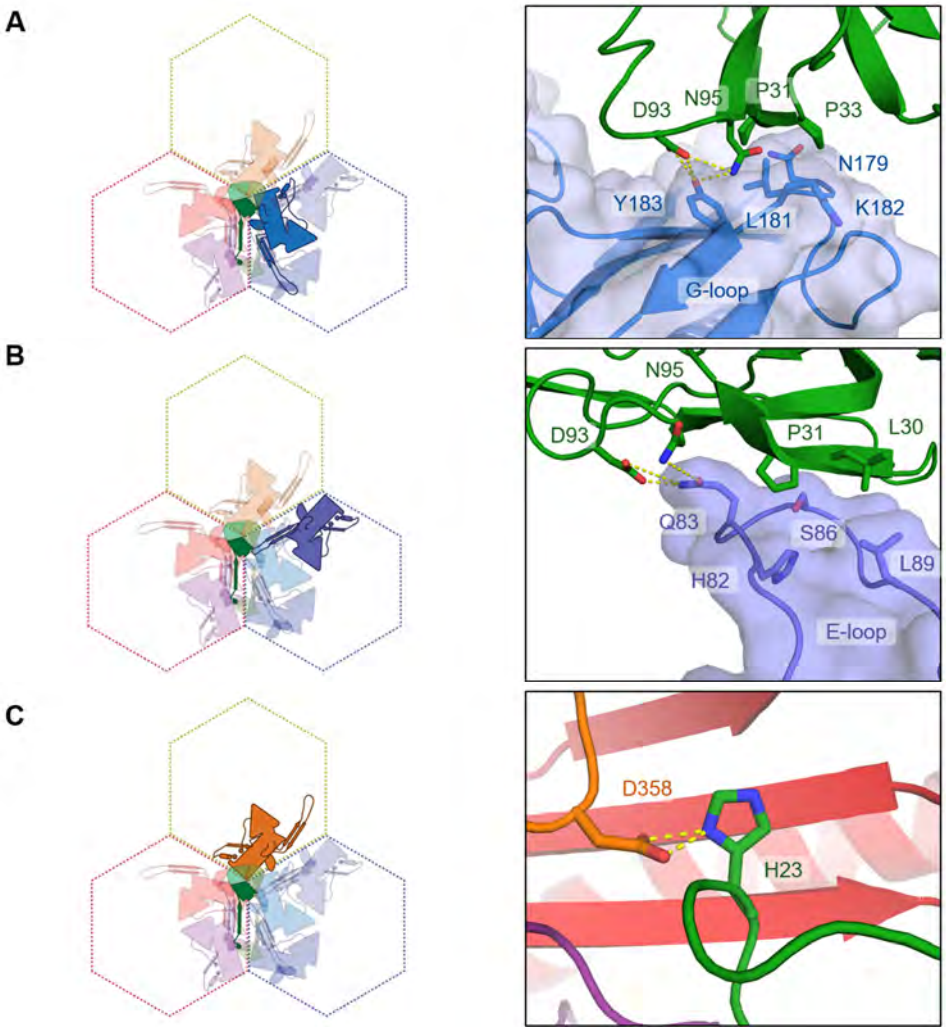

Figure 3.14. Inter-capsomer interactions of Dec with β and γ capsomers in P74-26.
A) The Dec β-tulip domain (green) interacts with the extended G-loop of the β2 MCP subunit (blue) across the quasi-two-fold/two-fold axis.
B) The β-tulip domain of Dec (green) interacts with the tip of the E-loop of the β3 subunit (lavender).
C) The β-tulip domain of Dec (green) extends across the quasi-three-fold/three-fold axis, hydrogen bonding with the P-domain of the γ1 subunit (orange).

Discussion

Architectures for enhancing stability: lassos, rings, and flaps

We observe enhanced hydrophobic interactions at the subunit-subunit interfaces within the capsid. Hydrophobic interactions are estimated to be >2 fold higher for P74-26 than for mesophilic homologs (Table 3.3; Table 3.4). This observation can partially explain the enhanced thermostability of the P74-26 capsid, as the hydrophobic effect increases in strength at high temperature (Huang and Chandler, 2000). In contrast, we observe no significant alteration in the number of hydrogen bonds or salt bridges (Table 3.3; Table 3.4), other interactions that have been seen to confer thermostability for some globular proteins (Petsko, 2001; Szilágyi and Závodszky, 2000; Vogt et al., 1997; Zhou and Dong, 2003).

We were not particularly surprised to find increased hydrophobic interactions in the P74-26 capsid. Numerous studies of thermophilic globular proteins show increased hydrophobic forces as a major contributor to thermal stability (Hennig et al., 1997; Knapp et al., 1999; Lim et al., 1997; Petsko, 2001; Voorhorst et al., 1997). However, there are two things that make the P74-26 capsid a unique model system: 1) high internal pressure from tightly packaged DNA induces mechanical stress in the capsid (Bauer et al., 2015; Evilevitch et al., 2003; 2004), and 2) it is a self-assembling structure in which inter-subunit architecture and topology plays an important role in overall stability. We can

derive these principles by comparing the P74-26 structure with those of numerous mesophilic homologs.

We find that the P74-26 capsid is stabilized by several loops and extensions that form topological linkages between subunits. The unique E-loop lasso attaches to the G-loop and P-domain of a neighboring MCP, which act as a 'hitching post' for topologically tethering the lasso. Furthermore, the N- and C-arms, together with the Dec-arm, completely ring the E-loop β-strands, another architectural element unique to P74-26 (Figure 3.9E). Thus, the E-loop is ringed toward the base and lassoed towards the tip.

A second lasso is formed by the N-arm, which forms both intra-capsomer and inter-capsomer interactions. The N-arm binds to the E-loop of a neighboring MCP within a capsomer through the upper-arm strand, the forearm helix, and hand region (Figure 3.9A,C). Furthermore, the elbow, forearm, and hand stabilize capsomer-capsomer interactions by binding with $Dec^{P74\text{-}26}$ and an MCP subunit that lies across the two-fold/quasi-two-fold axes (Figure 3.10A,B). While the N-arm is technically not a closed loop, P74-26 effectively closes the loop using the unique S-loop to fix the position of the forearm and hand regions (Figure 3.10C). We find no similar N-arm lasso architecture in mesophilic Caudoviruses (Chen et al., 2017; Guo et al., 2014; Hryc et al., 2017; Wang et al., 2017; Wikoff et al., 2000; Zhang et al., 2013; Zhao et al., 2017), which suggests that this architecture is important for enhancing capsid stability.

Although lassos are not found in other Caudoviruses, the distantly related herpesviruses contain analogous lasso architecture in the N-arm of the HK97 fold Dai and Zhou, 2018; Dai et al., 2018; Yu et al., 2017). Much like the N-arm of P74-26, the herpesvirus N-arm lassos are not true closed loops. Despite this seeming similarity, the herpesvirus lassos function differently. They exclusively stabilize inter-capsomer interactions, whereas the P74-26 N-arm lasso stabilizes both intra-capsomer and inter-capsomer interactions. Moreover, the pentons of herpesvirus capsids do not exhibit lasso interactions and the interactions are variable within hexon subunits, while P74-26 has nearly identical lasso interactions throughout both pentons and hexons. These observations indicate that the lasso architecture likely evolved independently and highlights the flexibility of topological stabilization mechanisms. We hypothesize that the extendable arm architecture facilitates evolution of stronger interactions within self-assembling systems such as capsids. These open-ended loops can be easily enhanced through serial single-residue extension. Perhaps this ease of evolution is why extended arm lassos are found in both P74-26 and herpesviruses. These extended lassos are similar to N- and C- terminal extensions that mediate assembly in other viruses (i.e. C-terminal extensions in SV40 capsid coat proteins (Stehle et al., 1996)). We anticipate that these types of extended lassos can be useful for engineering more stable capsids and other self-assembling particles.

Another advantage of the lasso architecture is that it can adopt a less extended conformation. P74-26 MCP has two lassos on either end of the HK97 fold, both of which are presumably present in the much smaller procapsid. By using these lasso structures, P74-26 can retain high stability, while providing conformational flexibility to expand during maturation. We hypothesize that the lassos are less extended in the procapsid; upon capsid expansion, the lassos reach their full extension, where they lock into place. The full extension observed in the mature capsid would provide tensional integrity, as we discuss below.

P74-26 uses interweaved flaps to topologically stabilize inter-capsomer contacts. The T-loop stabilizes inter-capsomer interactions by inserting into a groove on the P-domain of an MCP subunit in a neighboring capsomer. These T-loop interactions are found ringing the three-fold/quasi-three-fold axes along the inside face of the capsid (Figure 3.11B,E). Similarly, the inter-capsomer two-fold/quasi-two-fold interactions are stabilized on the outside face of the capsid by the interleaved arrangement of the N-arms (Figure 3.11C,D). These overlapping structures resemble the interleaved arrangement of flaps in the top of a moving box. In this manner, the outside and inside faces of the capsid are stabilized by two separate interleaving flap interactions. We propose the ‘moving box’ arrangements seen at symmetry and quasi-symmetry axes greatly strengthens the capsid against internal pressure because they are topologically challenging to disrupt. However, these arrangements are also presumably challenging to

assemble, which raises the important question of how the P74-26 capsid assembles with an interleaved architecture.

A cage of decoration proteins stabilizes the mature capsid

Decoration proteins increase capsid stability (Hernando-Pérez et al., 2014; Lander et al., 2008; Yang et al., 2008), although additional roles have been postulated (Newcomer et al., 2019). The three-fold/quasi-three-fold axes are stabilized by the $Dec^{P74\text{-}26}$ trimer. Compared to mesophilic Caudoviruses, the $Dec^{P74\text{-}26}$ trimer interacts with more subunits across a much larger interaction area (Figure 3.12A). The total interaction area per $Dec^{P74\text{-}26}$ subunit is quite remarkable: ~4100 $Å^2$ for a 146-residue protein. Our previous study showed that $Dec^{P74\text{-}26}$ is substantially more stable than its mesophilic homologs, and this stabilization is primarily through formation of an extraordinarily tight trimer (Chapter II; Stone et al., 2018). However, the trimerization interactions only account for a small fraction of the total $Dec^{P74\text{-}26}$ interaction area (~18% of the total interaction area per $Dec^{P74\text{-}26}$ subunit). This suggests that the $Dec^{P74\text{-}26}$ interaction with the capsid contributes a substantial amount of stability.

Interactions between $Dec^{P74\text{-}26}$ trimers forms a ‘cage’ holding the capsid together (Figure 3.12B). This arrangement is unique to P74-26. For example, phages Lambda and TW1 use a very similar Decoration Protein fold (Stone et al., 2018), but the interaction of their Dec-arm with other capsid proteins is much

more limited (Lander et al., 2008; Wang et al., 2017). Furthermore, the unrelated decoration protein of phage L does not connect with neighboring trimers, and in fact is missing at the quasi-three-fold axes (Newcomer et al., 2019). T4 phage is decorated with the Soc protein that interacts with neighboring Soc subunits at the three-fold and quasi-three-fold axes; however, Soc is present in relatively low occupancy (~50%), so the cage is incomplete (Chen et al., 2017). Because decoration proteins are usually absent in the procapsid (Lander et al., 2008), we expect that the $Dec^{P74\text{-}26}$ cage would assemble cooperatively upon capsid expansion to stabilize the capsid. Future experiments will interrogate the role of cooperativity in assembly and stability.

A tensegrity mechanism for enhancing capsid stability

How do forces from internal pressure act on the capsid, and how does the capsid architecture resist these stresses? If we assume that the pressure from encapsulated DNA is distributed uniformly around the isometric capsid, then all points on the capsid experience a vector of force that is perpendicular to the surface of the capsid. By analogy, the capsid is experiencing forces that are similar to those of a balloon. Thus, the internal pressure causes lateral stresses on capsid interactions. While the high internal pressure exhibited by the phage challenges capsid stability, it may allow for stability mechanisms that rely on tensegrity. It is possible that the internal pressure can be harnessed to produce a

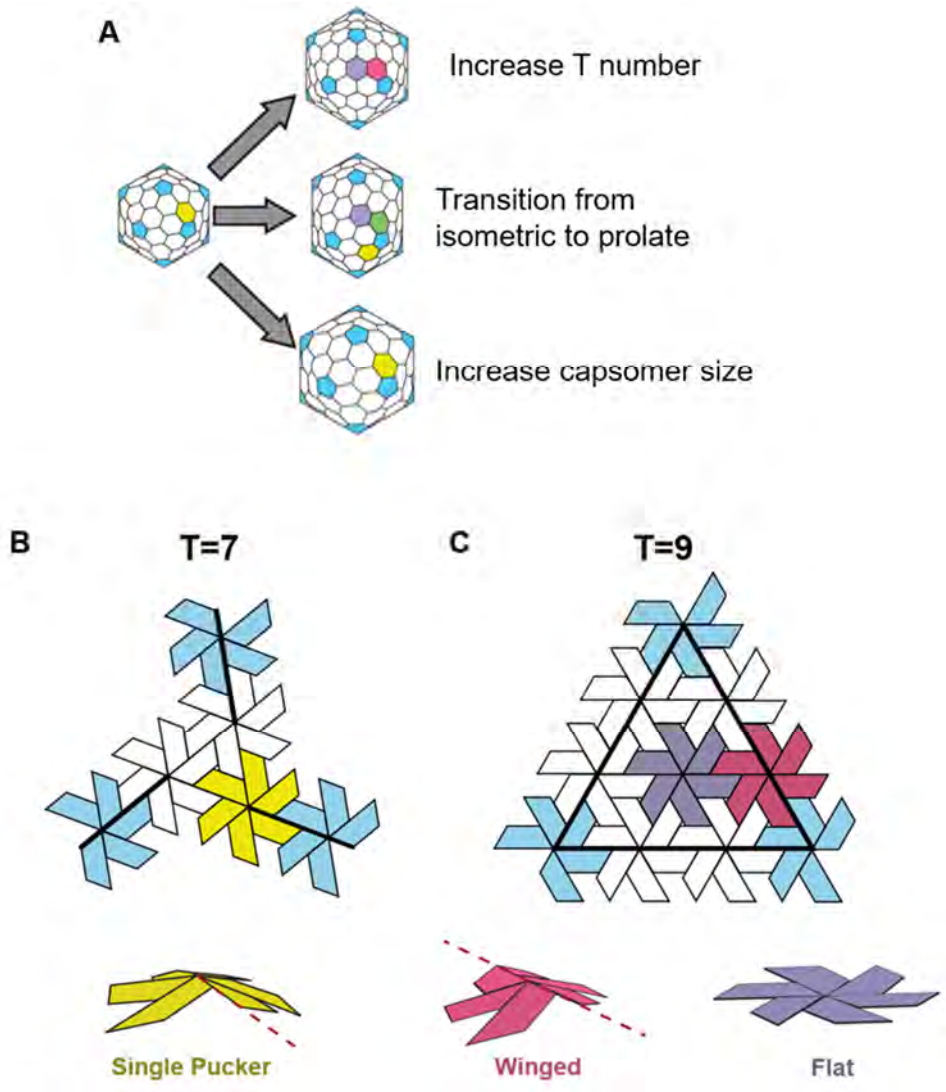

Figure 3.15. Mechanisms for increasing capsid capacity.
A) P74-26 adopts a novel mechanism for enhancing capsid capacity by increasing the size of capsomers, while retaining T=7 geometry.
B and C) Schematic representation of the faces of a T=7 (A) and T=9 (B) icosahedron. T=7 viruses consist of equivalent pentons as well as equivalent hexons, which adopt a 'single pucker' conformation. In T=9 viruses, two hexon conformations, 'winged' and 'flat' are required to form the larger capsid assembly. *Adapted from Mannige, et al, PLoS One, 2010.*

particle stabilized by the stress on the individual subunits. In support of this, picornavirus capsids, which contend with considerably less internal pressure, can be stabilized by minor modifications to the capsid (Wang et al., 2015).

The architecture of the P74-26 capsid is constructed to withstand lateral stress through tensional integrity. Tensional integrity, or tensegrity, is a generalized mechanism for architectural stability that involves structured regions held together by a network of flexible elements that are under continuous tension (Fuller, 1961). In the P74-26 capsid, the A- and P-domains are the structured regions, and the lassos and extended arms are the flexible elements that transmit tension. For example, the E-loop lasso will become taut against the 'hitching post' formed by the G-loop and the P-domain of the neighbor. Likewise, the N-arm forms a lasso whose end is held in place by the S-loop that locks into the groove between the forearm and hand (Figure 3.10C). Thus, we predict that the S-loop will exhibit the hallmarks of a 'catch bond', a non-covalent bond that becomes stronger under tension (Dembo et al., 1988). Furthermore, the P74-26 capsid contains several flaps that interleave with each other. These interactions would topologically resist the lateral and longitudinal stresses of internal pressure. As a whole, these lasso and flap elements use tension to resist structural failure of the capsid. The tensegrity mechanism observed here is simply a more elaborate example of capsid tensegrity suggested by Caspar many years ago (Caspar, 1980).

The lasso, flap, and arm interactions are positioned such that the internal pressure distributes the stress across multiple bonds. For example, the E-loop β-sheet experiences forces along the axis of the sheet. Thus, all bonds holding the sheet together are under stress rather than the orthogonal geometry in which the stress is just on the bonds at the end of the sheet. Capsid failure would require simultaneous disruption of many bonds (a 'shearing' geometry), rather than an unzipping geometry in which the bonds rupture one at a time (Rohs et al., 1999). Pioneering single molecule studies have shown that a shearing geometry requires much higher forces to disrupt than when forces act in an unzipping geometry (Guinn et al., 2015; Jagannathan et al., 2012; Wang et al., 2013). Thus, the P74-26 capsid is constructed such that lateral forces act in a shearing geometry, resulting in high tensegrity.

Mechanisms for altering capsid capacity

The capsid of P74-26 is larger than in most Caudoviruses, which correlates with its abnormally large genome. Most T=7 Caudoviruses have genome sizes between 30 and 50 kb (Table 3.2), while phage P74-26's genome is nearly twice as long at 83 kb (Minakhin et al., 2008). Based on genome size we had predicted that the capsid would be T=12 (average genome size ~80kb (Lander et al., 2012)), although T=9 or T=13 would have been possible (average genome size ~70kb or ~120kb, respectively). The P74-26 capsid attains this

larger size by significantly increasing the capsomer size rather than changing the icosahedral complexity. The capsomer is larger because the P74-26 MCP covers more surface area than normal, despite a typical length for MCP. Consequently, the capsomer is slightly thinner than normal (Figure 3.7B). Thus, the number of residues in MCP does not predict total area covered, and genome size does not predict triangulation number.

Recently, Bayfield et al. determined the structure of a closely related T=7 thermophilic phage, which similarly uses enlarged capsomers to increase capsid capacity (Bayfield et al., 2019). To our knowledge, this is a new mechanism for increasing capsid capacity. There are two classic mechanisms for enlarging a capsid: 1) increasing the triangulation number, and 2) conversion of an isometric to prolate head. In the first case, hexons are added across all faces of the capsid, whereas in the second, hexons are added across ten of the icosahedral faces such that the capsid is elongated in one dimension (Figure 3.15A). In both cases, the capsomers remain the same size. Here, we have identified a third mechanism for evolution of a larger capsid: increasing the size of the capsomer.

These three mechanisms have very different evolutionary barriers. The two classical mechanisms can be implemented through simple mutations and have been observed numerous times. In many viruses, simple point mutations modify the triangulation number (Butan et al., 2010; Ferreira et al., 2003; Fiedler et al., 2012). Furthermore, the triangulation number of some capsids can be altered without changing MCP sequence (Allison et al., 2003; Crowther et al.,

1994; Dearborn et al., 2017). Likewise, single point mutations in T4 phage convert the capsid from prolate to isometric or generate ‘giant’ heads in which the long axis of the prolate head is lengthened (Doermann et al., 1973; Eiserling et al., 1970). Therefore, the evolutionary barriers for altering capsid volume through the two classical mechanisms appear to be quite low. In contrast, the enlarged capsomer strategy identified here requires multiple, extensive alterations to the sequences of capsid proteins. The larger P74-26 capsid necessitates large changes to the eight separate modifications to the MCP structure, as well as the Dec-arm (Figure 3.5A,B; Figure 3.7C,D). This begs the question: why did phage P74-26 utilize this seemingly more challenging evolutionary strategy rather than the ‘easier’, classical strategies? What constraints prevented evolution of a larger capsid through the classical routes?

Our first hypothesis is that the lassos, flaps, and arms stabilizing the P74-26 capsid require a larger capsomer for function. It is possible that the lassos need extra space to open enough for a ‘hitching post’ structure to insert. Likewise, the flaps and arms may require a certain length to elicit their stabilizing activity. If this were the case, then the architectural elements stabilizing the capsid require larger than normal capsomers. In this scenario, the larger capsomer is the selected structural feature and the T=7 geometry is a spandrel: a biological structure that is a byproduct of evolution rather than a result of direct selection (Gould and Lewontin, 1979). However, we do not favor this hypothesis because lassos are found in herpesviruses, in which the HK97 fold is a typical

size. (Herpesvirus MCPs have several other ‘tower’ domains that increase size, but these domains are not part of the main capsid ‘floor’ and do not contribute to the HK97 fold (Yu et al., 2017)). Furthermore, other Caudovirus MCPs contain long N-arms (e.g. Sf6 phage (Zhao et al., 2017)) or E-loops that are opened nearly as wide as P74-26 (e.g. P22 phage (Hryc et al., 2017), but these proteins are of typical size. Nonetheless, this hypothesis remains untested.

A second hypothesis is that the genome size and capsid capacity coevolved through small concomitant increases. If the ancestral phage evolved a slightly larger genome than can be accommodated in the capsid, then there may be selective pressure for a slightly larger capsid. Increasing T number or transition to a prolate head substantially increases the capsid volume, resulting in large drops in internal pressure. These transitions may be discouraged because internal pressure needs to be maintained for infection (Bauer and Evilevitch, 2015). To avoid large changes in internal pressure, larger capsomers may slowly coevolve with a larger genome.

Our final hypothesis is that capsid geometry has a direct effect on overall capsid stability. We hypothesize that the T=7 geometry is inherently more stable than higher triangulation numbers due to variable conformations of hexons. All capsids that are T=9 or higher have more than one type of hexon present, while all capsids T≤7 have exactly one type of hexon (except T=1, which has no hexons; (Mannige and Brooks, 2009; 2010)). For example, T=7 has a ‘single-pucker’ hexon conformation, while T=9 has both winged and flat hexons (Figure

3.15B,C). We also note that prolate capsids have multiple types of hexons (generally three or more hexon conformations; Figure 3.15A). Thus, major capsid proteins in T>7 viruses must accommodate the hexon conformational heterogeneity, which may adversely affect stability.

We hypothesize that T=7 geometry is the highest complexity (i.e. largest size) that is inherently stable. More complex geometries would introduce instability through variation in hexon conformation. This inherent instability may require extra stabilization mechanisms to mitigate, such as decoration proteins to cement the structure in place. We envision two non-mutually exclusive disadvantages of T>7 geometry. First, each of the separate hexon conformations must remain functional and stable, which would constrain evolution of MCP proteins for greater stability. The second benefit is that a lower triangulation number results in fewer subunit-subunit interfaces, thus minimizing the number of weak points in the capsid. In support of this hypothesis, the extremophilic, archaeal virus HSTV-2 (*Halorubrum sodomense* tailed virus 2) packages its ~68 kb genome into a T=7 head (Pietilä et al., 2013). HSTV-2 utilizes a larger-than-normal capsid, and also has trimer decoration proteins that sit at the three-fold/quasi-three-fold axes. The fact that this mechanism for capsid enlargement has only been observed in extremophiles supports the idea that T=7 geometry has a beneficial effect on stability. In further support of our hypothesis, all known T>7 capsids use decoration proteins (to our knowledge), while many T=7 viruses lack decoration proteins.

If various triangulation numbers have different inherent stability, this would suggest that each geometry exhibits weak points at different regions of the capsid, as has been predicted from theoretical work (Wilson, 2016). We hypothesize that the three-fold/quasi-three-fold axes represent the weak points in a T=7 lattice. In support of this hypothesis, decoration proteins of T=7 Caudoviruses are commonly found at the three-fold/quasi-three-fold axes (Table 3.2) (Lander et al., 2008; Newcomer et al., 2019; Shen et al., 2012; Wang et al., 2017). Furthermore, these axes are stabilized by covalent cross-links in HK97 phage (Wikoff et al., 2000) and T-loop flaps in P74-26 (Figure 3.11B). To examine this idea further, we note that T=9 phage also use decoration proteins at the three-fold axes (Choi et al., 2008; Grose et al., 2014), while T=12 and T=13 phage use decoration proteins at the centers of capsomers (Effantin et al., 2006; Fokine et al., 2004; Lander et al., 2012).

We note that all of our analysis has primarily focused on Caudoviruses. These viruses do not generally break down their capsids as part of their lifecycle, so the capsid has no selective pressure to be labile. In fact, the high pressure of packaged DNA presents a high selective pressure to evolve stable capsids. It is likely that other types of viruses use different stability mechanisms, particularly viruses that disassemble their capsids as a necessary part of their lifecycle.

CHAPTER IV

Discussion

Principles of capsid assembly and stability

All viruses package their genomes into capsid shells, which protects the genetic material from the external environment and ensures robust pathogenicity. Capsids are marvels of biological engineering that contend with a wide variety of physical stresses, environmental conditions, and conformational changes throughout the viral life cycle. In dsDNA viruses, capsids must be flexible enough to allow significant conformational heterogeneity during maturation and rigid enough to stabilize against high interior pressures caused by the genome packaging reaction (Earnshaw and Casjens, 1980). In my thesis, I sought to elucidate the molecular mechanisms driving capsid maturation, assembly, and stability in dsDNA viruses. For this I used the hyperthermophilic bacteriophage P74-26 as a model for understanding capsid thermostability. My work characterizes P74-26 capsid subunit stability compared to related mesophilic viruses and identifies a number of unique architectural features of the thermophilic virion capsid that are critical for global stabilization. Furthermore, my work with P74-26 led to the development of novel methods and *in vitro* assays utilizing the capsid machinery, which will be valuable for future studies of virus structure and function.

In Chapter II, we structurally and biochemically characterize the capsid Decoration or 'Cementing' protein from P74-26 ($Dec^{P74\text{-}26}$). $Dec^{P74\text{-}26}$ is significantly more stable than mesophilic homologs largely due to an extensive

hydrophobic network that persists throughout the trimeric assembly, wherein hydrophobic clusters form rings that completely surround the 3-fold capsid axes.

The Decoration protein structure led to the discovery of the β-tulip domain, a five stranded β-barrel domain conserved in mesophilic phage structural elements. The β-tulip domain is additionally conserved in Herpesvirus Triplex capsid proteins, providing evidence of a shared evolutionary ancestry between bacteriophages and human pathogens. The β-tulip domain is additionally found in the anti-CRISPR protein AcrIIC1 – a broad-spectrum Cas9 inhibitor – and this discovery led to a proposed mechanism for divergent evolution of the inhibitor from a viral trimeric structural protein.

In Chapter III we elucidate the P74-26 capsid architecture to determine the structural mechanism of global capsid thermostability. Using single-particle cryo-EM with imposed icosahedral symmetry we solved the capsid structure to 2.8 Å, built a *de novo* model of the Major Capsid Protein (MCP), and refined the structure of the capsid Decoration protein. P74-26 MCP has the conserved HK97-fold found in related Caudoviruses, with several extended loops, ‘lassos’ and ‘flaps’ that form enhanced topological linkages between capsid subunits and stabilize the 2-fold and 3-fold axes formed by interfacing capsomers. Furthermore, we elucidate the stability mechanism of the Decoration protein trimer, which forms highly-cooperative interactions with MCP subunits surrounding the 3-fold/quasi 3-fold axes. These interactions are completely

unique to the thermophile and serve to further stabilize highly-pressurized virions against extreme temperatures.

Interestingly, P74-26 exhibits T=7 icosahedral capsid geometry despite having a genome nearly twice the length of typical T=7 viruses, enhancing the interior volume of the capsid instead by forming larger and flatter capsomers. This leads to the assembly of T=7 capsids with nearly twice the interior volume compared to related Caudoviruses. From this we form the hypothesis that limiting capsid geometry stabilizes P74-26 by limiting the overall complexity of capsid assembly. By subverting the incorporation of additional subunits during assembly, P74-26 limits the number of 3-fold axes that form potential weak points in the capsid shell. Thus, P74-26 enhances thermostability by both modulating canonical capsid subunit architecture and constraining icosahedral capsid geometry.

Taken together, this work establishes a robust platform for future studies of viral maturation and outlines core principles for increasing capsid stability and internal capacity. Phage P74-26 is an incredibly stable virus, and the molecular mechanisms I have characterized can be adapted to a broad array of future applications for virus particles. The discoveries made in my thesis will be valuable for engineering novel particles for *in vitro* encapsulation, developing viral-based nanomaterials, and designing improved viral vectors for therapeutics delivery.

Future Directions

There are still a great number of open questions regarding viral maturation in dsDNA viruses, and P74-26 in particular, that I was unable to address in the span of my thesis research. I outline below several future projects that would answer a number of open questions regarding capsid assembly, stability, viral maturation, and host-pathogen interactions. In particular, I focus on future structural studies and how this work can advance our understanding of capsid expansion, DNA ejection, and the genome packaging machinery. Future work will also uncover additional key structural features that contribute to virion stability in P74-26 and broaden our knowledge of how viruses cope with extreme environments. These projects build upon the work I have done to establish P74-26 as a model system for studying capsid stability, and explore new questions regarding a unique thermophilic bacteriophage and the underlying mechanisms of viral maturation in all dsDNA viruses.

Native capsid architecture

In order to reveal the structural mechanism of thermostability in the P74-26 mature virion capsid (Chapter III), particles were icosahedrally averaged during the refinement procedure. Using 60-fold icosahedral symmetry to process the P74-26 capsid dataset, I could refine the capsid asymmetric unit at high-resolution using a relatively small set of selected particles. This method is commonly used for determining structures of icosahedral viruses, and the high-order symmetry contributes significantly to the resolution of the capsid shell reconstruction (Guo and Jiang, 2014). However, there are non-icosahedral structural features of the virion that are averaged out during this procedure and are not represented in the final capsid reconstruction, primarily: 1) the unique Portal vertex through which DNA enters the capsid, and 2) the capsid interior, including the packaged genome housed inside the mature virion. Furthermore, the icosahedral reconstruction does not fully represent the true native state of the mature virion. Non-icosahedral features can, however be resolved using asymmetric or low-symmetry reconstructions of the P74-26 capsid.

Asymmetric and low-symmetry reconstructions have proven useful for determining structures of mature dsDNA virions (Liu et al., 2019; Parent et al., 2018; Wang et al., 2018), though final resolution is often significantly lower than in high-symmetry reconstruction. Symmetric reconstruction of full mature virions is often complicated due to inherent symmetry mismatches between the capsid head and the tail assembly. Icosahedral capsids, including P74-26, exhibit 5-fold

rotational symmetry. In contrast, the Portal vertex and tail assemblies typically exhibit 12-fold or 6-fold rotational symmetry. An asymmetric or low-symmetry reconstruction of the P74-26 capsid coupled with the Portal/neck region will not only reveal the architecture of the DNA ejection apparatus, but will also further our understanding of the role of capsid subunits in the context of the native virion. Symmetric reconstruction with 6-fold symmetry imposed is likely the best method for determining a high-resolution reconstruction of the P74-26 neck region. In this case, a focus mask including the features of the unique capsid vertex would be used to exclude capsid features. This would limit bias from the 5-fold symmetry of the head and yield a representative reconstruction of the neck region. Low-symmetry and masked reconstructions have been successful in both Caudoviruses and Herpesviruses, providing insight into the mechanisms of genome packaging and DNA ejection (Chang et al., 2005; McElwee et al., 2018; Tang et al., 2006)

One other intriguing prospect is elucidating the structure of the mature virion capsid interior. While the icosahedral reconstruction provides some evidence of interior genome organization, this density is difficult to interpret meaningfully due to the symmetry imposed during refinement. Furthermore, it remains unclear whether P74-26 contains any additional protein components within the capsid. However, preliminary cryo-EM micrographs suggest the presence of a packaged protein using a technique known as “bubblegrams”. Briefly, slight overexposure of cryo-EM samples (higher electron dose) causes

protein within the capsid to bubble from radiation damage, while the capsid protein shell and packaged genome are largely unaffected (Wu et al., 2012). In P74-26, we see this bubbling effect occur at the tail vertex, and the opposite polar vertex of the virion (i.e. the “top” of the virion; Figure 4.1). The density exposed at the tail vertex of the capsid likely corresponds to the interior portion of the Portal/neck assembly, but the density at the polar vertex remains unaccounted for. These preliminary results indicate that protein may be packaged along with genomic DNA into the capsid. This is not entirely unexpected, as some viruses such as Hepatitis B Virus package polymerase polypeptides along with the viral genome (Selzer and Zlotnick, 2015; Wang et al., 2014). It is also possible that the bubbling at the polar vertex is an artifact of genome organization within the capsid. Further experimentation will be necessary to characterize the capsid features causing the bubblegrams in P74-26.

For the capsid interior asymmetric reconstruction, the capsid outer shell would be excluded from refinement using a mask created from the virion capsid shell structure. Thus, the outer capsid features will not dominate particle orientation, and genomic DNA and any additional protein components inside the head can be resolved without significant bias. In dsDNA viruses, the genome typically adopts a single-spooled conformation within the capsid shell, coiling against the capsid interior surface upon entry (Black, 1989; Cerritelli et al., 1997; Jiang et al., 2006; Lepault et al., 1987). As DNA continuously enters the capsid,

layers form wherein the genome becomes progressively less ordered toward the center of the particle. Thus, genome organization can be resolved layer-by-layer moving toward the center of the virion capsid (Ilca et al., 2019; Koning et al., 2016; Liu et al., 2019), progressively masking layers of the viral genome. This structure will help to elucidate interactions between the genome and the capsid shell, and provide valuable insight into how dsDNA is compressed to near-crystalline density within the virion.

While still technically challenging these capsid structures can be elucidated using the sample preparation procedure described in Chapter III, subverting additional specimen optimization. It is also possible to reprocess preexisting datasets used previously for icosahedral refinement. The micrographs used for icosahedral refinement feature closely-packed P74-26 capsids which could be reprocessed as described above without imposed symmetry (Figure 4.2A). In order to achieve mid- to high-resolution structural information, it will likely be necessary to collect larger datasets with more particles suitable for refining capsid substructures. The other major limiting factor in obtaining these reconstructions is the computational power required to process virus particles without high-symmetry. While still feasible, asymmetric reconstructions of large datasets featuring large P74-26 particles will require allocation of significantly more computational resources and time compared to icosahedral reconstruction.

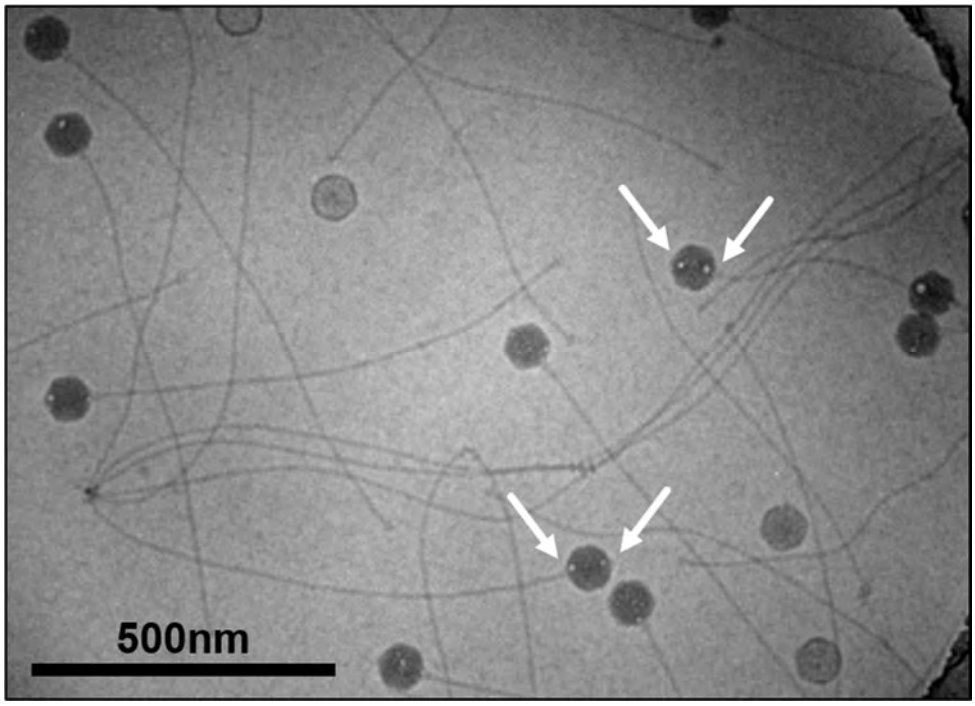

Figure 4.1. Overexposure of P74-26 virions suggests the presence of an interior capsid protein. Limited overexposure to the electron beam generates 'bubblegrams' in P74-26 samples frozen in vitreous ice. Bubblegrams reveal localized puncta of radiation damage at the tail vertex and opposite pole (white arrows) suggesting the mature virion capsid may include a packaged protein component.

The P74-26 tail assembly

Nearly all bacterial viruses possess a tail that is essential for recognition of the host cell surface and ejection of DNA from the capsid into the host cell cytoplasm (Molineux, 2006). The tail attaches to the Portal vertex of the capsid following genome packaging, generating an infectious virion. At the far end of the tail fiber sits the baseplate, a macromolecular claw-like protrusion primed for host cellular attachment. Attachment of baseplate to the host cell triggers a signal cascade that travels through the tail tube and to the capsid to signal ejection (Davidson, et al., 2011).

P74-26 is a Siphovirus with a flexible non-contractile tail, which is the longest known tail of any virus; nearly 1 μm long at full extension (Minakhin et al., 2008). A tail of such size is counterintuitive, as we predict the significant length would negatively impact efficiency of DNA ejection and virion stability. We also do not currently know why P74-26 tails are significantly longer than tails found in related viruses and thus, the evolutionary advantage of a long tail and the mechanism of DNA ejection through the tail tube remain poorly understood. Structural characterization using cryo-EM will lead to a better understanding of the molecular mechanisms of host cell infection, DNA ejection, and signal transduction in P74-26. Ideally, it would be beneficial to structurally characterize the tail assembly both pre- and post-ejection in order to map conformational changes caused by the infection mechanism. The samples used above for capsid reconstruction reflect a pre-ejection state, as the mature virions are

poised for host cellular attachment. Isolation of post-ejection tail complexes suitable for cryo-EM may prove difficult. While possible to optimize purification of viral particles in a post-ejection state, it is also possible to induce this conformational switch *in vitro.* In mesophilic phage, addition of lipopolysaccharide (LPS) and/or other cellular membrane proteins to purified virions induces DNA ejection (Andres et al., 2010; Jin et al., 2016). A similar approach could be used to induce ejection in P74-26. We do not currently know precisely what host factors are critical for attachment to *Thermus thermophilus*, as no LPS exists in the host strain HB8 (Suda et al., 2012). Previous characterization of glycolipids present in the cell membrane does, however, identify a number of putative host attachment factors (Suda et al., 2012). Nevertheless, structural characterization of the tail complex will help to elucidate the mechanisms of cell recognition and the signal cascade leading to DNA ejection.

The P74-26 tail primarily consists of the tail tube protein (TTP). The tail is non-contractile and likely comprised of stacked rings of TTP oligomers rather than forming a helical filament (Davidson, et al., 2011). In other Siphoviruses, the TTP typically forms hexameric rings (Fokine and Rossmann, 2014). As such, the best single-particle cryo-EM reconstruction approach would use rotational symmetry (i.e. 6-fold symmetry) to achieve the highest possible resolution. Reconstructions of the tail tube and baseplate structures pose unique challenges compared to the capsid reconstructions described above. First, sample preparations and previously collected datasets for capsid reconstructions are not

suitable for tail reconstruction. While the virions do have intact tails, these samples are optimized to produce highly concentrated close-packed virus capsids and as a result the tails aggregate, preventing clear selection of tail segments or baseplates for reconstruction. Future data collection for tail tube and baseplate structures may require significant dilution and further buffer optimization compared to datasets used for capsid reconstructions (Figure 4.2B). The tail also has significant flexibility, making reconstruction of the tail tube difficult. For this it will likely be necessary to reconstruct a small section of the tail tube using a focus mask consisting of only a few (optimally 3-5) layers of stacked oligomers. This would alleviate significant conformational heterogeneity caused by extreme flexibility throughout the tail structure.

Additionally, it will be difficult to amass a large dataset of baseplate particles given the significant length of the tail in P74-26. At full extension, the tail is nearly 1 μm long, which is wider than a recorded micrograph at a typical magnification used for data collection. Future experiments will be necessary to determine the most efficient method for specimen optimization. For these reconstructions it may also be beneficial to prepare samples for cryo-EM using purified P74-26 tails in the absence of capsid. We can isolate the tails from high-titer phage stocks grown in *Thermus thermophilus*. Tails are preassembled concurrent with procapsids during viral maturation, and we can separate the virion precursor components using sucrose gradient separation.

Alternatively, we can structurally characterize host cell infection using cryoelectron tomography (cryo-ET). In tomography, samples are imaged at progressive tilt angles to form a 3D representation of a single specimen rather than averaging a large number of 2D particle images (Ericus et al., 2015; Subramaniam et al., 2007). This technique could be applied to P74-26 infections in *Thermus thermophilus* to map conformational changes of the tail complex during infection, while avoiding the limitations of a single-particle approach outlined above. Recent work with the Podovirus P22 uses a similar cryo-ET method to elucidate the mechanism of host cell attachment and DNA injection (Wang et al., 2019). Furthermore, this emerging technique can offer valuable insight into virus maturation within the host cell with respect to viral protein interactions with cellular architecture and native assembly within the host cytoplasm (Cope et al., 2011; Iwasaki et al., 2010).

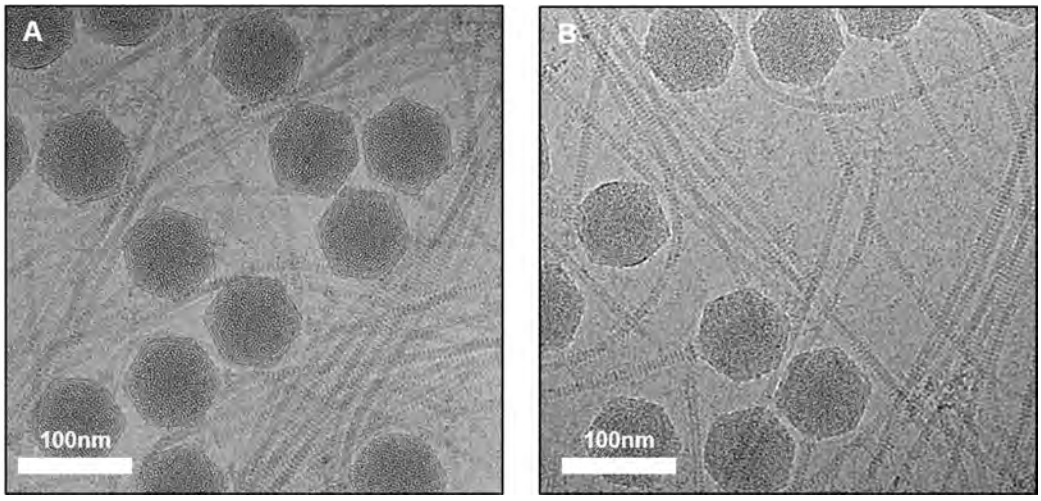

Figure 4.2. Cryo-EM sample preparation of P74-26 mature virions.
A) Representative micrograph used for icosahedral reconstruction in Chapter III. Virion tails aggregate at high concentrations, making the dataset unsuitable for tail tube and baseplate reconstructions.
B) At lower concentrations tail fibers are less prone to aggregation, which may be advantageous for selection of tail tube sections and tail baseplate structures. Tail fiber flexibility may also hinder 3D reconstruction efforts

Capsid maturation intermediates

The work presented in this thesis primarily focuses on capsid structure and stability of mature, DNA-filled virions. This provides only a static view at a single stage in the viral life cycle, but capsids are dynamic particles that undergo large conformational changes during maturation (Catalano, 2005). With our P74-26 particle purification protocol, we have the ability to further probe the capsid expansion mechanism by elucidating structures of capsid intermediates. Purification from high-yield phage stocks generates highly-pure P74-26 virions and procapsids suitable for cryo-EM. Isolated procapsids are native, fully-functional particles suitable for genome packaging (see Appendix II; Figure 2.2, 2.3) and viable candidates for single-particle cryo-EM. Recently, Bayfield et al. determined cryo-EM reconstructions of the procapsid and expanded head of a closely-related thermophilic phage (Bayfield et al., 2019). Using a similar approach, we can determine high-resolution structures of P74-26 procapsids, and can map the conformational changes between the procapsid and mature virion in extreme detail.

Additionally, we can also determine high-resolution structure of intermediate expanded, empty capsids. Procapsid particles in other Caudoviruses can be expanded *in vitro* using mild denaturants to generate a large population of expanded empty heads that lack the interior Scaffolding protein (Tuma et al., 1998). A similar approach could be used to generate expanded heads in P74-26 without causing significant damage to the capsid

assembly overall. Expanded heads have an icosahedral morphology, and resemble the mature virion capsid in both size and shape, but do not contain packaged DNA. Comparison of the expanded empty and mature virion capsid structures would allow for further exploration of the capsid stability mechanism outlined in Chapter III. We pose the hypothesis that the immense pressure caused by the tightly-packaged DNA within P74-26 capsids served to strengthen the overall stability of the capsid. Using high-resolution structures of expanded capsids with and without packaged DNA, it is possible to discern differences in conformation caused by interior capsid pressure and further examine the role tensegrity plays in stabilizing the capsid.

As with the mature virion cryo-EM reconstructions, the procapsid and expanded head reconstructions can be performed with or without imposed symmetry. In a similar manner, icosahedrally averaged reconstructions can provide near-atomic resolution details regarding subunit-subunit interactions across the capsid surface. Asymmetric or low-symmetry reconstructions will again be more technically challenging and computationally demanding, but could help to reveal details about how significant conformational changes during capsid expansion impact subunit interactions at the Portal vertex.

Preliminary experiments show P74-26 procapsids remain intact and retain their spherical unexpanded shape when frozen in vitreous ice. Procapsids and expanded heads will require further sample optimization to produce quality samples for high-throughput data collection. Particle density and dispersity is

lower than expected when compared to virion sample preparation. Different buffer conditions (i.e. detergents) and EM grid preparations will need to be screened and significantly improved to create monodisperse samples or close-packed procapsid and expanded head particles suitable for data collection.

Capsid packaging machinery

Viral genome packaging is critical for the generation of infectious mature virions. However, little is known about the precise interactions between the packaging machinery and genomic DNA during active packaging. Previous studies have proposed several different mechanisms for DNA translocation by the Large Terminase motor (TerL) in several model systems (Dixit et al., 2012; Hilbert et al., 2015; Ray et al., 2010; Sun et al., 2008). One such model is the so-called DNA 'crunching' model, wherein the translocation machinery compresses incoming DNA and abrogates the canonical B-form double helix during genome packaging (Harvey, 2015; Ray et al., 2010). Alternatively, our lab previously proposed a 'lever' mechanism in P74-26, in which conformational switching in the ATPase domain of TerL generates the force required for translocation (Hilbert et al., 2015). High-resolution structure determination of a stalled, actively-packaging procapsid will help to elucidate the precise mechanism of DNA translocation in P74-26 and identify any DNA conformational changes made during genome packaging.

The *in vitro* packaging assay I developed in P74-26 (Appendix II) lends itself to future structural studies of the genome packaging mechanism. For the *in vitro* packaging assay, purified procapsids and TerL are mixed with DNA substrate in the presence of ATP. This minimal complex is sufficient for packaging DNA *in vitro*, and can be used to generate viable targets for structural characterization.

As described above, asymmetric and low-symmetry cryoEM reconstructions would also be necessary for elucidating structures of the packaging machinery. In this particular case, native procapsids would be combined with TerL, DNA, and/or the regulatory small terminase subunit (TerS) to form an active packaging complex. It is possible to generate ‘Y-shaped’ or forked DNA constructs that will stall the packaging reaction, locking the complex in an active conformation (Black and Rao, 2012; Ray et al., 2010). A similar DNA substrate design could be used to stall P74-26 packaging. Stalled complexes could then be structurally characterized using a single-particle cryo-EM approach. Alternatively, incubation at low temperature where packaging occurs at a significantly slower rate (see Appendix II; Figure A2.3) would allow for trapping a packaging motor unperturbed by mechanical blocks. For this project, packaging reaction conditions, DNA construct design, and sample conditions for EM will require significant optimization.

In vitro assembly of P74-26

Capsid assembly is a highly-regulated process. Complex protein-protein interactions and assembly of capsid substructures must be properly mediated to generate homogeneous capsids with high fidelity. The work presented in Appendix I seeks to elucidate the capsid assembly mechanism using a fully-reconstituted *in vitro* assembly assay in P74-26. A functional *in vitro* reconstitution system would be a valuable tool with which to study capsid stability and maturation and can be exploited for a wide array of therapeutic and materials science applications.

P74-26 MCP recombinantly expressed and purified from *E. coli* has a propensity for spontaneous self-assembly into procapsid-like particles consistent in size with native procapsids. This occurs without an interior scaffolding component, which, to our knowledge, is unique amongst Caudoviruses as an additional scaffolding protein or domain is typically required for assembly of procapsid-like particles *in vitro* (Huet et al., 2010; Spilman et al., 2013). While spontaneous assembly of freshly-purified MCP was inefficient (~1% total mass assembled), incubation at high temperature favors assembly with nearly 70% efficiency.

Though assembly of MCP *in vitro* is an exciting prospect on its own, it will be beneficial in the future to reconstitute procapsid-like particles that incorporate both the Scaffolding protein and Portal. Portal nucleates assembly by

recruiting Scaffolding protein, leading to construction of the procapsid *in vivo*. Recent work using phage P22 demonstrated that Portal can be incorporated into procapsid-like particles assembled *in vitro*, similarly nucleating capsid construction and forming homogeneous particles (Motwani and Teschke, 2019). The major developmental limitation of the P74-26 assembly model is that we cannot currently express the Scaffolding protein by conventional methods in *E. coli*, as we can with other capsid structural components. Coexpression with MCP or additional capsid components may be necessary to sequester the Scaffolding protein into assembling capsids, avoiding potential cell toxicity or aggregation issues. It may also be beneficial to express the Scaffolding protein in the host strain by generating constructs in vectors designed for constitutive expression in *Thermus thermophilus.* Future experiments will also be necessary to optimize incorporation of additional capsid components into the particle assembly.

With these core capsid structural elements, it is possible to reconstitute packaging-competent procapsids solely using recombinant protein. With an *in vitro* assembly assay capable of generating functional procapsids it will be possible to further examine structure-function relationships in the capsid and identify interactions critical for assembly. *In vitro* reconstitution will provide a method to introduce mutations and modulations to capsid components without a prerequisite established genetic system in P74-26.

An *in vitro* assembly assay in P74-26 would additionally have a number of applications in the future development of therapeutic strategies and novel

engineered nanomaterials. Viruses and virus-like particles have been used extensively as nanocontainers, nanoreactors, and delivery vehicles with a diverse range of cargoes (de la Escosura et al., 2009; Minten et al., 2009; Patterson et al., 2014; Uchida et al., 2007). The thermostability of P74-26 would be a significant advantage in developing it for future uses in a laboratory setting. Regardless of successful incorporation of additional capsid components, optimized self-assembly of MCP alone still provides a powerful tool that can be adapted to encapsulate desired cargo in thermostable procapsid-like particles.

Novel *in vitro* assays to study terminase motor function

The work presented in Appendices II and III describes novel approaches to monitor P74-26 procapsid packaging and terminase motor translocation *in vitro*. Our lab has extensively characterized the P74-26 packaging ATPase motor, TerL (Hilbert et al., 2015; Hilbert et al., 2017). This work identified residues critical for ATP hydrolysis and DNA binding, and proposed a mechanism for DNA translocation. While we have studied ATP hydrolysis and DNA binding *in vitro,* we currently lack robust methods for probing the mechanistic details of translocation and genome packaging.

With a small amount of additional optimization, these novel *in vitro* assays I have established using P74-26 can be used to test hypotheses regarding the terminase motor mechanism and structure-function relationships using partially and fully assembled packaging complexes. Future adaptation of this work will lead to the development of powerful single-molecule assays that can parse mechanistic details of the packaging reaction on biologically relevant timescales.

The *in vitro* packaging assay described in Appendix II is fully functional, and can be used in its current state to probe TerL function in complex with the procapsid (Figure A2.1). Previously characterized mutants that inhibit ATP hydrolysis or DNA binding are easily substituted for wild-type TerL in the packaging reaction, providing a rapid method for characterizing critical residues in the context of active packaging. This assay can also be adapted to single-

molecule formats to determine how TerL converts ATP hydrolysis into mechanical motion, pumping DNA into the capsid shell. Fluorescently-labeled DNA fragments could be used to monitor individual packaging events to measure kinetics of initiation and packaging. The force generation mechanism of the terminase motor can likewise be studied on the single-molecule level using optical tweezer experiments. For these experiments, immunolabeled procapsid and DNA are attached to small beads captured in an optical laser trap. The beads are brought close together in a flowcell containing ATP to initiate packaging. Then, the beads are separated to create tension on the DNA strand. As packaging commences, DNA length and motor stalling force can then be measured over time to determine force generation and kinetics of packaging (Figure 4.3; Smith, 2011).

The novel triplex dissociation assay described in Appendix III will require further optimization prior to use in functionally characterizing TerL. In this assay, TerL translocation is monitored via dissociation of a fluorescently-labeled triplex oligonucleotide probe bound to a dsDNA template (Figure A3.1), similar to assays previously developed to study other DNA translocases (Firman and Szczelkun, 2000; Graham et al., 2010; McClelland et al., 2005). Dissociation of the triplex probe by TerL in the presence of ATP leads to an increase in fluorescence (Figure A3.3), but it remains unclear if TerL is actively translocating DNA. Preliminary experiments suggest partial recovery of probe fluorescence occurs in the presence of ADP and ATP analogs, suggesting DNA binding may

be sufficient for triplex dissociation. Future work will be necessary to determine whether this assay is a viable method for monitoring DNA translocation. Novel dsDNA templates that introduce inhibitory molecular roadblocks can additionally be designed to demonstrate directional translocation and determine if this assay is functional.

If proven successful, *in vitro* DNA translocation can also be adapted to single-molecule formats. It is possible to generate dsDNA templates with multiple successive triplex binding sites that can be affixed in a flow cell to monitor fluorescence changes upon addition of TerL. Unique fluorescent labels can be used for multiple triplex oligonucleotides, and thus individual dissociation events can be detected in rapid succession to monitor DNA translocation. This approach could be adapted similarly to DNA curtain analysis, a promising method for monitoring DNA translocation using TIRF microscopy (Figure 4.4; Collins et al., 2014). Using this approach, thousands of TerL translocation events can be monitored concurrently in real-time at the single-molecule level to elucidate DNA translocation kinetics. Together, these assays provide powerful tools that will further our understanding of the packaging reaction and the mechanism of DNA translocation in this unique thermophilic virus.

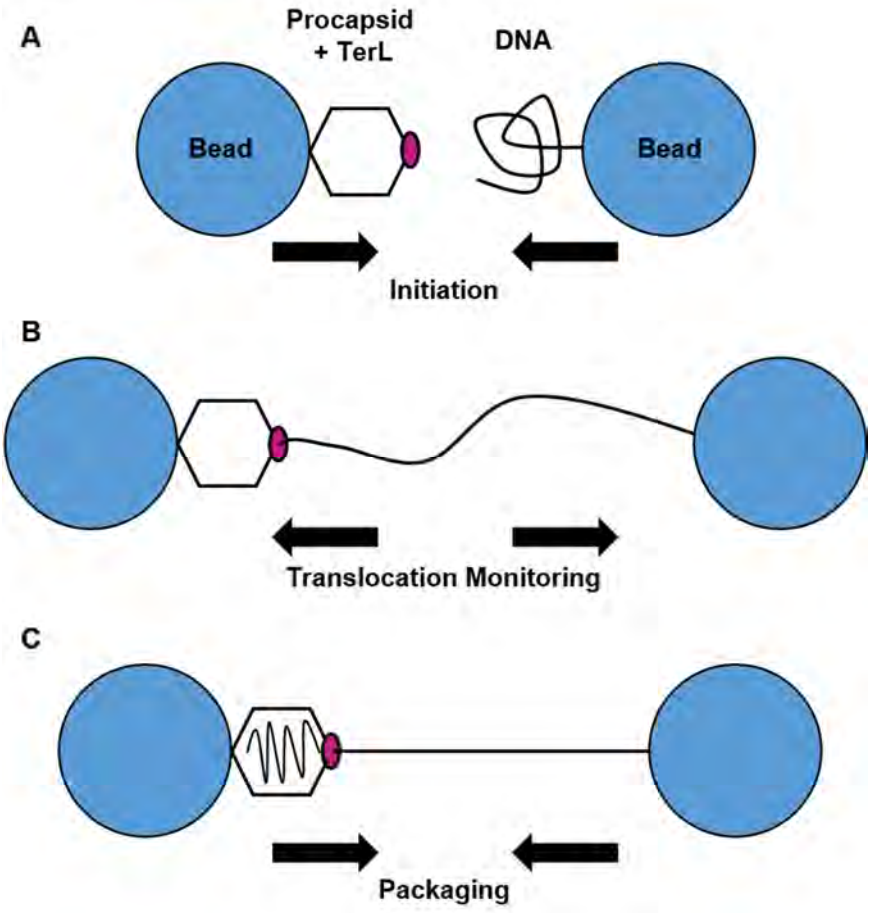

Figure 4.3. Optical trap assay to monitor genome packaging.
A) The procapsid:terminase complex, and dsDNA substrate are attached to two beads both immobilized in optical traps. The beads are brought close together to facilitate DNA:terminase motor attachment and initiate DNA packaging.
B) Following attachment, optical trap beads are then separated to fully extend the DNA strand to monitor DNA translocation.
C) ATP-dependent packaging pulls the beads closer together, which can be used to monitor translocation kinetics and the stalling force of the terminase motor.

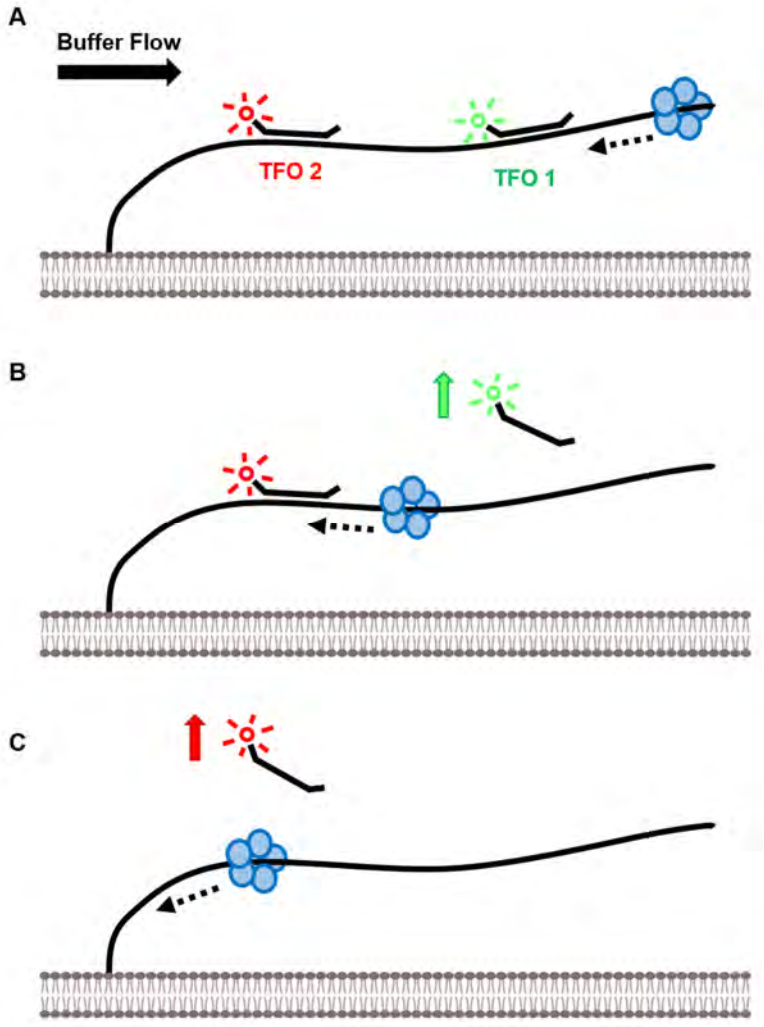

Figure 4.4. Single-molecule triplex dissociation assay.
A) Substrate DNA with multiple triplex binding sites and uniquely labeled Triplex Forming Oligonucleotides (TFOs) is affixed to a lipid bilayer or glass slide in a flow-cell chamber. Addition of P74-26 terminase motor and ATP initiates translocation along the dsDNA template in a similar manner to DNA curtain analysis.
B and C) Translocation of the terminase motor along dsDNA sequentially removes the fluorescently-labeled TFO triplex probes, causing an increase in fluorescence upon dissociation. Sequential dissociation events can be used to study kinetics of terminase motor translocation.

Concluding remarks

This thesis elucidates structural mechanisms of capsid stability in the hyperthermophilic bacteriophage P74-26, and establishes a foundation to develop novel *in vitro* tools and techniques for future studies of virus assembly. In addition, my work outlines a framework for structural and biochemical characterization of P74-26 that can be adapted to answer a number of open questions with regard to virion assembly and maturation. Overall, my thesis is an exciting beginning to many projects that will further our understanding of these mechanisms in dsDNA viruses at the molecular level and continue to inspire future researchers to explore the complex nature of viruses.

Appendix I

In vitro assembly of P74-26 capsids

Introduction

All viruses encapsulate their genomes in capsid shells comprised of a large number of repeating protein subunits, which undergo a complex self-assembly reaction to generate homogeneous particles with high fidelity (Perlmutter and Hagan, 2016; Zlotnick, 2005). This process relies on a small number of unique structural proteins to build procapsids suitable for genome packaging. Despite the critical role of procapsid assembly in viral maturation, little is known about the mechanistic details of the assembly process. Thus, understanding the molecular mechanisms driving capsid assembly *in vitro* is critical not only for the treatment of human infections, but also to the understanding of complex biological assemblies.

This appendix highlights my work to establish an *in vitro* assay with which to probe the P74-26 capsid assembly mechanism. P74-26 is a thermophilic virus with a marked increase in capsid stability compared to related mesophilic viruses, yet it shares a conserved capsid architecture. Furthermore, given the enhanced stability of the P74-26 virion demonstrated in Chapters II and III, particles assembled *in vitro* can be engineered and adapted for future use as robust nanoparticles and therapeutic delivery vehicles.

Materials and Methods

Cloning, Expression, and Purification of P74-26 capsid components

P74-26 Major Capsid Protein (MCP; P74-26 gp88) was synthesized by the Genscript Corporation and cloned into the pSMT3 vector using BamHI and XhoI (New England Biolabs) cut sites within the multiple cloning site. The pSMT3 vector construct contains a cleavable N-terminal 6X His-SUMO tag (Yunus and Lima 2009). Oligonucleotide primers for cloning were ordered from Integrated DNA Technologies:

MCP forward primer: 5'-GATCGGATCCATGCGTGTCCCGATTAAC-3'

MCP reverse primer: 5'-GATCCTCGAGTCACAGTTCGACTTTCGGCGG-3'.

P74-26 MCP was expressed in *E. coli* strain BLR-DE3 cells grown in the presence of 30 µg/mL kanamycin in Terrific Broth supplemented with phosphate buffer. Cells were grown at 37 ˚C to an OD_{600} of 0.6-0.8. Cultures were then incubated at 4 ˚C for 20 minutes, and expression was induced with the addition of 1mM isopropyl-β-d-thiogalactopyranoside (IPTG). Cultures were incubated overnight at 18 ˚C, with a final OD_{600} of 10-11. Following expression, cells were then pelleted and resuspended in buffer A (50 mM Tris-HCl pH 7.5, 300 mM KCl, 20 mM Imidazole, 5 mM 2-mercaptoethanol, 10% (v/v) Glycerol). Cells were then flash frozen and stored at -80 ˚C until use. Thawed cells were then lysed via two passes through a cell diruptor, and lysate was spun at 20,000 x g for 40 minutes to pellet cell debris. The supernatant was then filtered through a 5 µm prefilter

and then a 0.45 µm filter, and applied to 2x5mL His-Trap columns (GE) which were pre-equilibrated in buffer A. Following protein binding, the column was washed with 2 column volumes of buffer A, and then P74-26 MCP was eluted with 3 column volumes of buffer B (50 mM Tris-HCl pH 7.5, 300 mM KCl, 500 mM Imidazole, 5 mM 2-mercaptoethanol). Eluate was dialyzed into buffer A, with Ubiquitin-like Protease 1 (ULP1) to cleave the His-SUMO protein tag overnight at 4 ˚C. Dialyzed, cleaved protein was passed over a fresh pre-equilibrated His-Trap column to remove the cleaved tag. Cleaved protein was then dialyzed overnight into 50 mM Tris-HCl pH 7.5, 150 mM KCl. The dialyzed protein was concentrated to 3 mg/mL, flash frozen in liquid nitrogen, and stored at -80 ˚C. The P74-26 Decoration protein (P74-26 gp87) was purified using a similar protocol, as described previously (Stone et al., 2018; Chapter II).

SEC-MALS

All tandem Size Exclusion Chromatography – Multi-Angle Light Scattering was performed using 100 µL injections at a monomer concentration of 1 mg/mL in 50 mM Tris-HCl pH 7.5, 150 mM KCl. P74-26 MCP samples were filtered through a 0.2 µM syringe filter prior to sample injection. The sample elution was monitored using a Dawn Heleos-II multi-angle light scattering detector and an Optilab T-rex differential refractive index detector (Wyatt Technology). Elution peaks for

assembled and monomeric P74-26 MCP were selected and analyzed using the ASTRA6 software package (Wyatt Technology).

Electron Microscopy

Clean carbon-coated 400-mesh copper grids (Electron Microscopy Supply) were glow discharged for 1 minute at 60 V using a Pelco easi-glow (Pelco). 3.5 µL samples of P74-26 MCP were then applied to glow discharged grids and incubated for 30 seconds. Excess buffer was then removed using Whatman filter paper, and the grids were then washed three times using 1% Uranyl Acetate stain. On the final wash, excess stain was removed and the grids were allowed to dry at room temperature prior to imaging. Data was collected using a Philips CM120 microscope fitted with a Gatan Orius SC1000A detector. Micrographs were collected using the Gatan Digital Micrograph software; magnification information is denoted in respective figure legends.

Results and Discussion

Expression and purification of P74-26 capsid structural components

For this project, I sought to develop an *in vitro* assembly assay using the structural components of P74-26 procapsids and mature virion capsids: Major Capsid Protein (MCP), Decoration protein (Dec), Scaffolding protein, and Portal. His-tagged constructs of P74-26 MCP, Dec, and Portal express well in *E. coli* and are easily purified from cell lysates (Figure A1.1A). However, we were unable to express the Scaffolding protein despite varying expression protocols or using different expression constructs. Viral Scaffolding proteins can be difficult to recombinantly express, due to their propensity for aggregation when not sequestered by MCP for assembly. We also observe no Scaffolding protein expression when coexpressed with MCP, a technique which has been used in mesophilic virus systems to generate procapsid-like assemblies including the interior protein scaffold *in vitro* (Bazinet and King, 1988; Prevelige et al., 1988; Lee and Guo, 1995). The Scaffolding protein is essential for the assembly of procapsids suitable for genome packaging and future work will be necessary to optimize its expression and purification.

Recombinant P74-26 Major Capsid Protein has a propensity for self-assembly

Despite the lack of recombinantly expressed P74-26 Scaffolding protein, negative-staining electron microscopy reveals MCP is capable of self-assembly *in vitro*, which to our knowledge is unique to P74-26. Assembly in other dsDNA viruses is typically dependent upon interior or exterior scaffolding proteins Dokland, 1999; Prevelige and Fane, 2012). Certain phage coat proteins contain a Scaffolding domain (i.e. HK97 delta-domain; Oh et al., 2014), but the interior scaffold is still required for assembly. It is possible that other thermophilic virus coat proteins may share this propensity for self-assembly without an interior scaffold, though no such examples have yet been characterized.

P74-26 MCP primarily forms particles of two distinct sizes, both large (Figure A1.1B) and small (Figure A1.1C). The larger assembled particles are approximately 60 nm in diameter, which is consistent with the size of native P74-26 procapsids (~62 nm). Interestingly, the smaller particles are approximately 30 nm in diameter; significantly smaller than expected for P74-26. Though there is no interior scaffold guiding assembly, MCP likely still forms particles with inherent icosahedral geometry outlined by quasi-equivalence theory (Caspar and Klug, 1962). Icosahedral viruses have discrete geometric organizations based upon the incorporation of additional capsomers in the final assembly, denoted by the Triangulation number, T. Our hypothesis is that the large self-assembled MCP particles have T=7 geometry, equivalent to the native capsid structure, while the

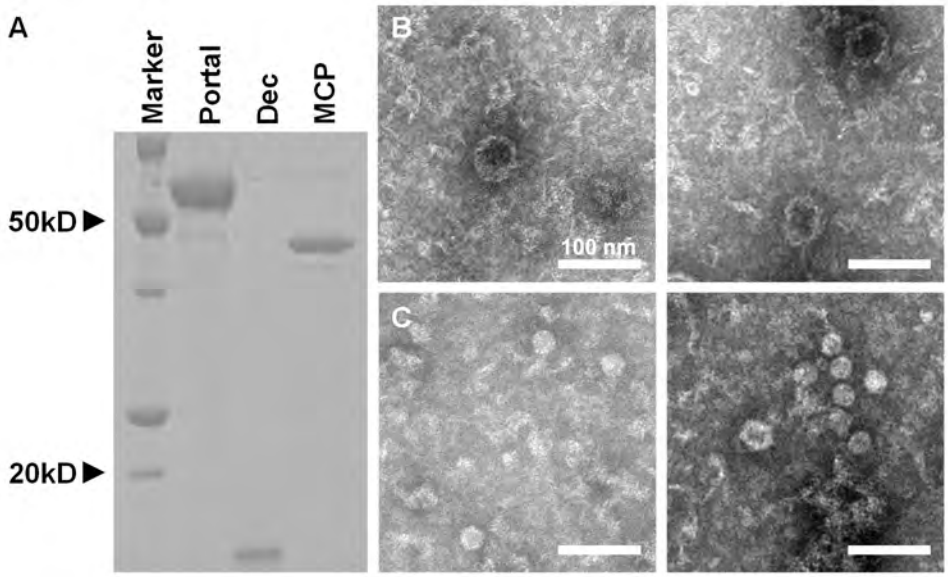

Figure A1.1. Purified P74-26 MCP has a propensity for self-assembly *in vitro*.
A) SDS-PAGE gel analysis of purified P74-26 capsid structural proteins; Portal protein, Decoration protein (Dec) and Major Capsid Protein (MCP).
B) Purified MCP stored at -80 ˚C forms procapsid-like particles approximately 60 nm in diameter, consistent in size with native T=7 procapsids.
C) P74-26 MCP also forms capsid-like particles with a smaller diameter, likely sampling a smaller icosahedral geometry (i.e. T=4); scale bars, B and C = 100 nm.

smaller particles have a lower T number (i.e. T=4) and are comprised of fewer MCP subunits. Future studies will be required to determine if self-assembled MCP particles have canonical icosahedral geometries.

We next wanted to characterize the procapsid-like particles and determine the efficiency of spontaneous MCP self-assembly using tandem Size-Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS; Figure A1.2). Purified MCP elutes as two major populations; a large early peak corresponding to procapsid-like particles and a smaller late peak of unassembled MCP. The procapsid-like particles have a hydrodynamic radius (Rh(Q)) of 30.3 nm, which is consistent with the previous electron microscopy results. The calculated molecular weight of the procapsid-like particle peak is 17.8 MDa roughly similar to the predicted molecular weight of a T=7 icosahedron constructed from 420 total MCP subunits (19.2 MDa). We see no evidence of a distinct peak representing the smaller (T=4) particles, as the heterogeneous assembled particles likely co-elute. The molecular weight of the unassembled MCP is consistent with monomeric P74-26 MCP (46.6 kDa), which accounts for the vast majority of MCP in the sample. The large scattering peak only contains ~1% of total protein mass, while ~99% of the MCP remains monomeric.

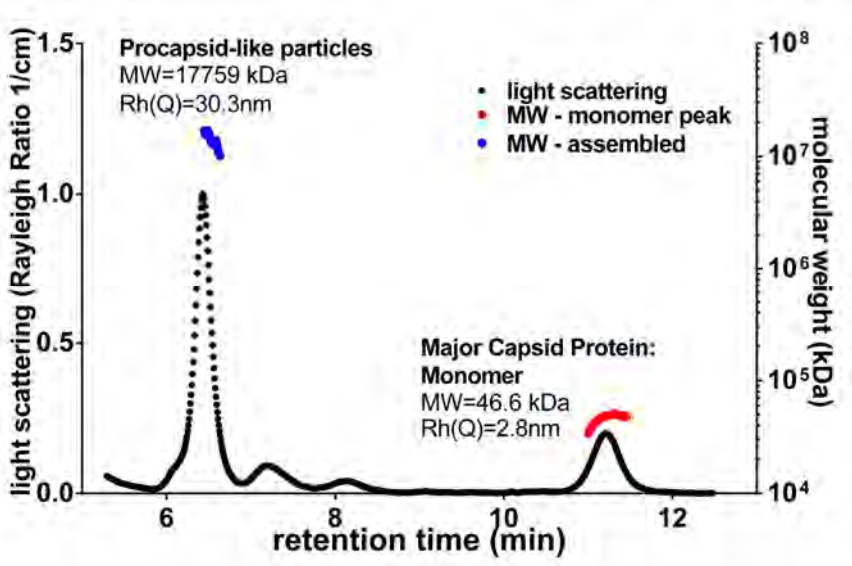

Figure A1.2. Spontaneous self-assembly of P74-26 MCP forms procapsid-like particles. SEC-MALS analysis of purified MCP shows two discrete populations: monomeric MCP (molecular weight denoted in red) and procapsid-like assembled MCP (molecular weight denoted in blue). Assembled MCP forms particles consistent in size with native P74-26 procapsids (~60 nm in diameter).

Optimization of procapsid-like particle assembly of MCP

Though P74-26 MCP does spontaneously form procapsid-like particles *in vitro*, assembly is unfavorable and MCP is largely monomeric. We next wanted to determine if we can induce favorable assembly of MCP. Addition of crowding agents commonly used to induce oligomerization of macromolecules (i.e. PEG and glycerol) did not have a significant effect on particle assembly. Addition of 5 or 10 mM $CaCl_2$ had a modest effect on MCP assembly, though procapsid-like particles are largely heterogeneous, often forming particles and aggregates larger than native T=7 P74-26 capsids. Incubation at high temperatures however, did have a significant positive effect on assembly. MCP incubated at 4 ˚C for 6 hours has a slight improvement on assembly (5% efficiency; Figure A1.3A), similarly forming procapsid-like particles of multiple sizes (Figure A1.3B). In contrast, 67% of MCP assembles into procapsid-like particles following incubation at 50 ˚C for 18 hours (Figure A1.3C). Moreover, heat treated MCP assembles into particles consistent in size with native T=7 capsids (Rh(Q): 36.3 nm) with less variation in particle size (Figure A1.3D). Negative-staining electron microscopy shows no evidence of T=4 procapsid-like particle assembly. Thus, incubation at high temperature not only significantly improves efficiency of MCP assembly, but also decreases particle heterogeneity.

The experimental work presented in this appendix centers around self-assembly of P74-26 MCP in the absence of other capsid proteins. While this project establishes P74-26 as a viable platform for building reconstituted particles

in vitro that can be used to develop thermostable nanoparticles, it will require significant future work to enhance efficiency of assembly. First, successful expression (or coexpression with MCP) and purification of the Scaffolding protein will be essential for assembly of homogenous particles. MCP assembles without scaffold, but even with improved efficiency yields a heterogeneous particle population. Incorporation of Portal within procapsid-like particles will likewise be necessary for generating packaging-competent particles *in vitro*, and will aid in the assembly of packaging-competent procapsid particles.

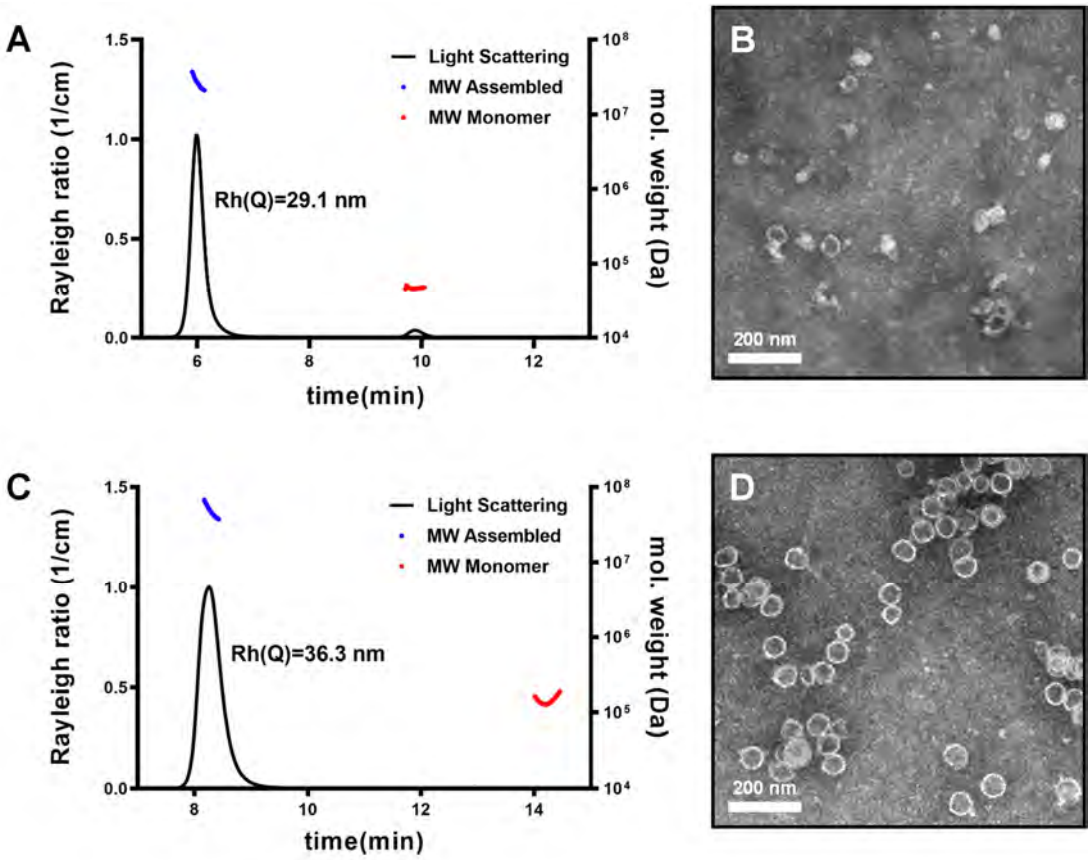

Figure A1.3. Incubation at high temperature induces favorable assembly of capsid-like particles from P74-26 MCP.
A) SEC-MALS of purified MCP incubated at 4 ˚C shows partial assembly of large, capsid-like particles. Molecular weights for assembled (blue) and monomeric (red) MCP are consistent with predicted values.
B) Negative stain electron micrograph showing assembled particles consistent in size with native procapsids.
C) SEC-MALS of purified MCP incubated for 18 hours at 50 ˚C. MCP favorably assembles into capsid-like particles.
D) Representative negative-stain electron micrograph of heat-induced capsid-like particle assembly.

Appendix II

Development of a P74-26 *in vitro* packaging assay

Introduction

In vitro packaging assays have been developed in several mesophilic bacteriophage systems to probe the genome packaging reaction and study terminase motor function (Gope and Serwer, 1983; Dale and Greenaway, 1985; Rao et al., 1992; Droge and Tavares, 2000). In general, these *in vitro* packaging assays employ purified procapsids, purified terminase motor or extract, and ATP to package linearized DNA substrate into the capsid shell. The *in vitro* packaging assay is a bulk DNase protection assay, wherein DNA packaged into the capsid is protected from subsequent digestion by DNase I. Following the packaging reaction, any remaining unpackaged DNA is digested via treatment with DNase, while the DNA protected in the capsid remains intact. The capsids are then digested, and packaged DNA is analyzed using gel electrophoresis (Figure A2.1). This section outlines my preliminary work developing an *in vitro* packaging assay using thermophilic phage P74-26.

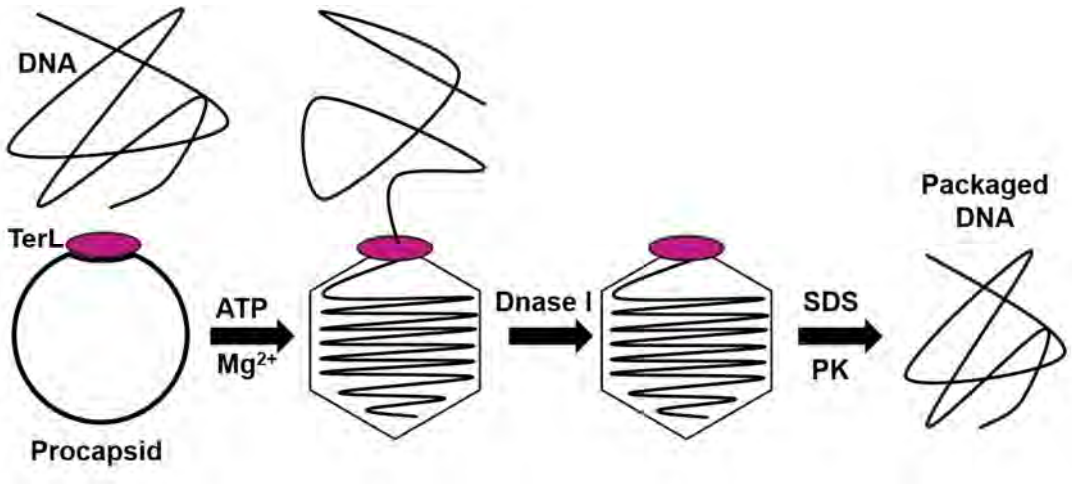

Figure A2.1. A novel *in vitro* packaging assay using the hyperthermophilic bacteriophage P74-26.
Purified procapsids, purified terminase motor (TerL), and linearized DNA substrate are combined, and the reaction is initiated upon addition of ATP. Reactions are then incubated at given temperatures for 2 hours. Following packaging, reactions are treated with DNase I to digest any unpackaged DNA. Procapsids are then digested with Proteinase K, and protected DNA can be assayed via gel electrophoresis.

Materials and Methods

Purification of P74-26 procapsids

P74-26 phage stock preparation and initial Cesium Chloride gradient purification was performed as previously described in Chapters II and III. For electron microscopy and *in vitro* packaging, procapsids isolated from CsCl gradients require further purification. ~1-2 mL of procapsids isolated from CsCl gradients (sedimented at 1.3 g/mL CsCl) were dialyzed into 2 L of dialysis buffer (50 mM Tris pH 8.0, 10 mM Nacl, 10 mM $MgCl_2$) overnight at 4 ˚C. Dialyzed procapsids were then concentrated to 2 mL total volume. Sucrose density gradients were assembled in 14 mL ultracentrifuge tubes (Beckman) through repeated flash freezing of 2mL steps of 30%, 25%, 20%, 15%, and 10% sucrose in 50 mM Tris pH 8.0, 10 mM NaCl, 10 mM $MgCl_2$. Frozen gradients were stored at -20˚ C prior to use. Sucrose gradients were allowed to thaw overnight at 4 ˚C prior to centrifugation to form continuous density gradients. Procapsids were then added and the gradients were spun in a Beckman SW40Ti rotor at 25,000 RPM at 4 ˚C for 2 hours. Gradients were divided into 500 µL fractions. Fraction composition was analyzed by SDS-PAGE gel analysis, and fractions containing procapsids were then pooled and dialyzed into dialysis buffer twice overnight at 4 ˚C. Procapsids were concentrated to ~10^{10} particles per µL (approximated by gel densitometry) for *in vitro* packaging experiments.

Negative staining electron microscopy

In order to ensure the quality of procapsid purification, particles were analyzed by negative staining electron microscopy. 3 µL of concentrated procapsids were added to carbon-coated 400-mesh copper grids (Electron Microscopy Sciences) and incubated for 30 seconds. Excess sample was then blotted off the grid using Whatman filter paper. The sample-coated side of the grids was then dipped into drops of 1% uranyl acetate stain solution, and excess stain was blotted off. This procedure was repeated twice, and then grids were allowed to dry for 1 minute. Grids were analyzed on a Philips CM120 120kV microscope equipped with an Orius SC1000 detector (Gatan). Images were then collected at 31,000 or 53,000 X magnification using the Gatan Digital Micrograph software suite.

In vitro packaging assay

Packaging reactions were set up using ~$2x10^{11}$ procapsids and 2 µM purified large terminase (TerL) or mutant TerL construct in 1X reaction buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 5% w/v polyethylene glycol MW 8000, 3 mM spermine, 3 mM spermidine, 3 mM putrescine, 3 mM beta-mercaptoethanol). The procapsid/TerL mixture was then incubated at 20 ˚C for 30 minutes. 200 ng of DNA substrate (pet28 expression plasmid linearized by Sma I restriction digestion) was added, and the reactions were initiated with addition of ATP to a final concentration of 10 mM. Reactions were then incubated at given

temperatures for 2 hours in a thermalcycler heat block. Following the packaging reaction, 5 mg of DNase I was added and reactions were incubated at 37 ˚C for 1 hour to digest unpackaged DNA substrate. 5 µg of Proteinase K in 0.25 M EDTA, 2% SDS was then added and packaging reactions were incubated at 65 ˚C for 1 hour to inactivate DNAse I and digest the procapsids. To assay for packaged DNA, reactions were finally run on 1% agarose gels in 1X Tris-Acetate EDTA buffer at 70V for approximately 1 hour.

Results and Discussion

Purification and characterization of P74-26 procapsids

The most critical component of the *in vitro* packaging assay is highly-pure procapsid particles suitable for packaging. Currently, there is no established P74-26 genetic system, which would allow us to generate packaging mutants to cause accumulation of procapsids during infection. Alternatively, I developed a two-stage density gradient purification protocol to isolate native P74-26 procapsids from wild-type phage infections. *In vivo*, both procapsids and phage tails are preassembled prior to genome packaging (Catalano, 2005). Thus in a live phage infection, packaging-competent procapsids are an abundant species that can be isolated from mature virions and other contaminants such as cell debris and phage tail structures. To generate purified procapsids, P74-26 phage particles precipitated from large-scale phage stocks (Stone et al., 2018) are first purified using Cesium Chloride (CsCl) step gradient density ultracentrifugation. The procapsids typically sediment at ~1.3 g/mL CsCl, while mature virions sediment at ~1.5 g/mL CsCl, allowing for the isolation of immature procapsids suitable for packaging. At this stage, the procapsid fractions from the CsCl gradient additionally contain *Thermus thermophilus* cell debris, damaged virions, and phage tails (data not shown). The procapsid-containing fractions are then subjected to separation via continuous 10-30% sucrose gradient density utracentrifugation. Sucrose gradient fractions contain highly-pure P74-26 procapsid particles, largely free from contaminants.

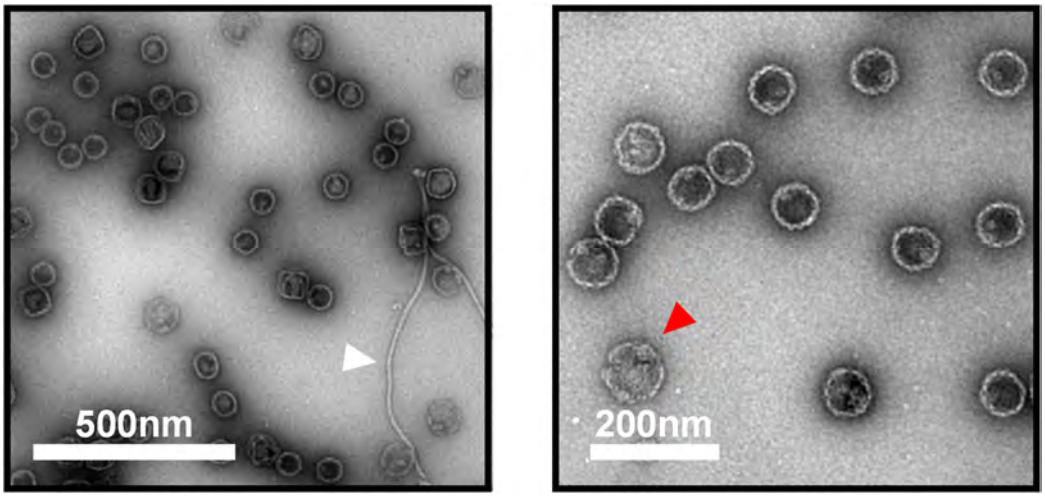

Figure A2.2. Purification of native P74-26 procapsids.
Negative-stain electron micrographs of sucrose gradient purified P74-26 procapsids. Gradient purification separates procapsids from residual cell debris and assembled phage tails, though some tail structures persist in pooled fractions (left, white arrow). The pooled fractions consist primarily of functional, spherical procapsids (damaged procapsid denoted by red arrow, right).

To assess the quality of procapsid preparations, concentrated particles to be used in the packaging assay were evaluated with negative-staining electron microscopy (Figure A2.2). P74-26 procapsids are spherical particles ~70 nm in diameter that are permeable to Uranyl Acetate stain. This stain permeability of the 'empty' procapsid particles contrasts the exclusion of stain seen in the case of genome-filled virions (see Chapter III). The procapsid preparations still include some contaminants (Figure A2.2), though contamination is significantly reduced compared to purification without sucrose gradient separation. Additionally, the mechanical stress caused by pipetting and concentrating the particles leads to a small population of damaged procapsids, which similarly do not significantly impact the efficiency of bulk *in vitro* packaging.

P74-26 procapsids package at high temperatures in vitro

We next sought to test whether highly-pure P74-26 procapsids can be packaged *in vitro*. Procapsids are packaged using purified TerL (Hilbert et al., 2015) in the presence of a linearized plasmid DNA substrate, and ATP. P74-26 uses a 'headful' packaging mechanism, and genome packaging is not dependent on recognition sequences or Cos site initiation (Catalano, 2005). Thus, we use a linearized pet28 plasmid backbone for subsequent packaging experiments. Packaging reactions are incubated at the given temperatures as described above for 2 hours. P74-26 procapsids package linearized plasmid DNA optimally at 60

˚C, showing significantly lower packaging efficiency at 50 ˚C and 70 ˚C (Figure A2.3A). At 60 ˚C packaging efficiency varies slightly with each experiment and procapsid preparation, though optimally we see nearly 100% efficiency of packaging using 100 ng of linearized plasmid DNA substrate (Figure A2.3A). In the absence of TerL or ATP we see no protected DNA, suggesting that the purified procapsid particles are free from significant contamination with genomic DNA or DNA-filled mature virions.

The TerL construct we used for DNA packaging contains a non-cleavable poly-histidine (His) tag at the C-terminal end (pet24-TerL). During maturation, the C-terminal domain of TerL binds to the portal protein assembly to pump DNA into the procapsid. However, given the efficiency of packaging, we do not suspect the His tag negatively impacts binding of TerL to portal during packaging. We additionally used a TerL construct with a cleaved N-terminal His tag (pet28-TerL; Figure A2.3B). The pet28-TerL construct also packages linearized plasmid DNA, though with lower efficiency than the pet24-TerL construct under the same experimental conditions. The difference in efficiency may be due to a difference in enzymatic activity between the two purified constructs rather than a difference in portal binding affinity, though further experimentation is necessary.

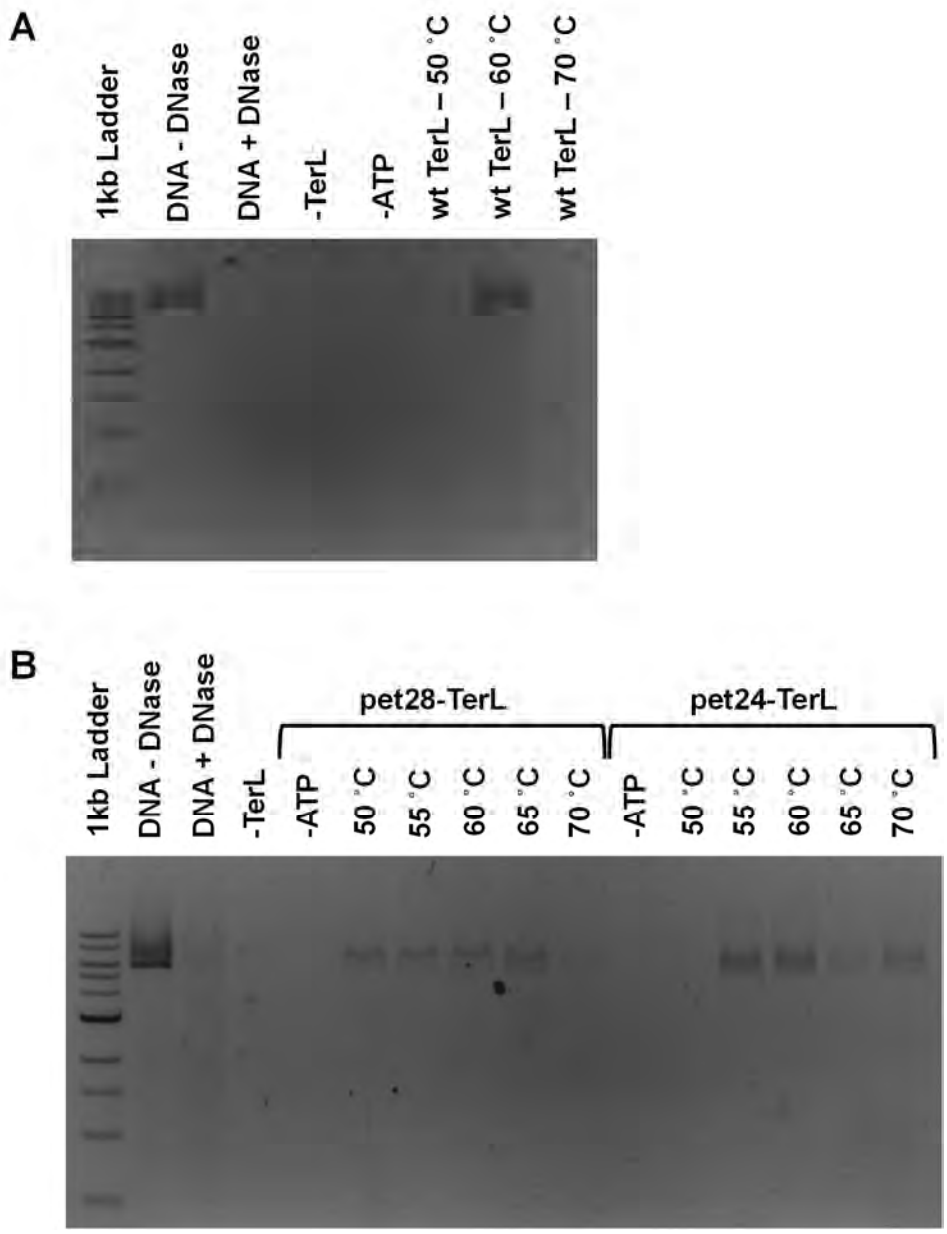

Figure A2.3. P74-26 *in vitro* packaging is efficient at high temperature.
A) Unprotected input DNA is digested by treatment with DNase I, and in the absence of TerL or ATP, no packaged DNA is protected from digestion. Procapsid packaging is optimally active at approximately 60 ˚C. Control lanes were similarly incubated at 60 ˚C
B) Procapsids are efficiently packaged by two different TerL constructs; TerL with a cleavable N-terminal His tag (pet28-TerL) and TerL with a non-cleavable C-terminal His tag (pet24-TerL).

P74-26 TerL catalytic mutant and C-terminal deletion inhibit DNA packaging

We finally sought to determine how previously-characterized TerL mutant constructs impact DNA packaging *in vitro*. Our lab previously generated a deletion construct that has a 32 residue deletion corresponding to a flexible 'tail' region at the C-terminus of TerL (TerL ΔC-term). Under identical packaging conditions the TerL ΔC-term construct inhibits DNA packaging, likely because the deletion abolishes Portal binding and thus formation of a functional packaging complex (Figure A2.4A). Additionally, our lab previously identified residue 150 as the catalytic glutamate required for hydrolysis of ATP. Accordingly, the TerL-E150A mutant construct completely abrogates ATP hydrolysis (Hilbert et al., 2015). TerL-E150A similarly exhibits no packaging activity *in vitro* (Figure A2.4A). Preliminary titration experiments with increasing concentrations of TerL-E150A added to the packaging reaction additionally show cooperative inhibition of packaging activity (Figure A2.4B). TerL must form the biologically active pentameric ring in order to bind Portal and translocate DNA. Thus, incorporation of a single mutant subunit into the pentameric assembly is sufficient to stall the packaging reaction, rendering the entire TerL ring inactive.

Perspective

The *in vitro* packaging assay described here is a powerful tool for probing the molecular underpinnings of the genome packaging reaction in a thermophilic system. Using purified native procapsids from phage P74-26, we have established a reproducible protocol for packaging linearized DNA substrate with high efficiency. Future work will need to be done to assess the efficiency of packaging a variety of DNA substrates, including full-length genomic DNA.

Overall, this work provides a framework to further investigate the DNA translocation mechanism of the Large Terminase. With the packaging assay, we can further investigate how previously-characterized TerL mutations that impact DNA binding affinity and enzymatic activity disrupt viral maturation using a biologically relevant *in vitro* reconstitution system. Additionally, the P74-26 packaging assay can plausibly be adapted in the future to single-molecule formats. Recent work with fluorescence-based assays and optical tweezer experiments in phages Lambda, T4, and Φ29 demonstrate that these techniques can parse motor translocation kinetics and measure force generation during packaging (Smith, 2011; Smith et al., 2001; Fuller et al., 2007). Performing single-molecule assays with the P74-26 packaging complex will help to further illuminate the packaging motor mechanism and structure-function relationships in a unique thermophilic model system.

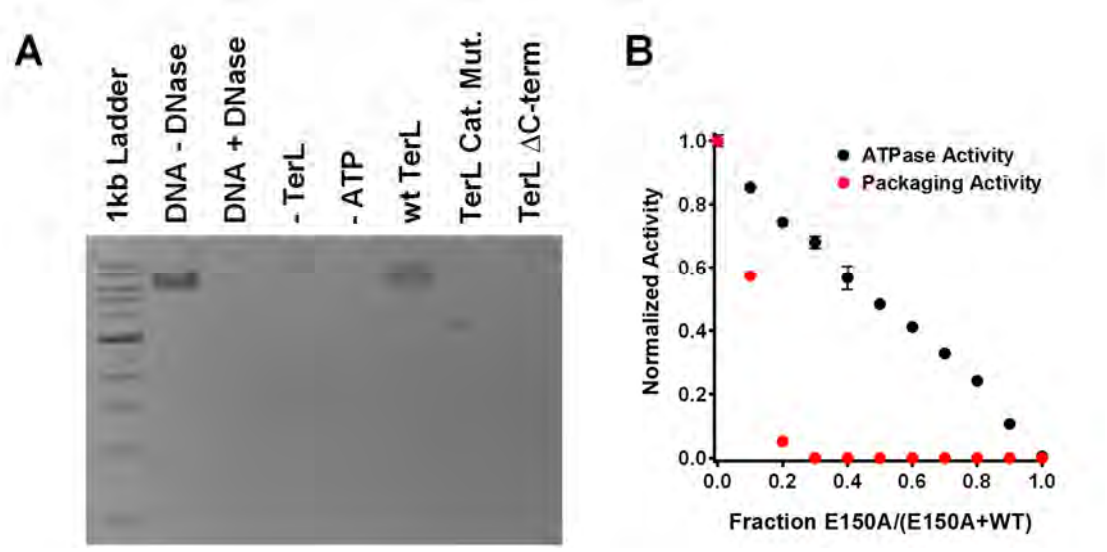

Figure A2.4. P74-26 TerL catalytic mutant and C-terminal tail deletion constructs inhibit procapsid packaging.
A) The wild-type pet24-TerL construct favorably packages the linearized DNA substrate, while the E150A mutation (Cat. Mut.) completely abrogates packaging activity. Truncation of the C-terminal tail of TerL required for procapsid binding likewise leads to loss of packaging ability.
B) Preliminary titration of TerL-E150A suggests cooperative inhibition of DNA packaging activity (red), compared to ATPase activity inhibition (black).

Appendix III

Development of a P74-26 *in vitro* translocation assay

Introduction

For this project, we sought to develop an *in vitro* translocation assay to probe the activity of the P74-26 Large Terminase packaging motor (TerL). To do this, I adapted a novel triplex dissociation assay previously used to investigate the activity of helicases and other DNA translocases (Firman and Szczelkun, 2000; McClelland et al., 2005; Seidel et al., 2005; Sivanathan et al., 2006). For this assay, a fluorescently-labeled triplex forming oligonucleotide probe (TFO) is bound to a double-stranded DNA template containing a complementary 'Triplex Binding Site' (TBS) inserted into a larger dsDNA template. At low pH, the dsDNA template and labeled probe form a pyrimidine-pyrimidine-purine DNA triplex formed via base pairing of an additional poly-pyrimidine strand from the TFO in the dsDNA major groove (Frank-Kamenetskii and Mirkin, 1995). The stable triplex substrate can then be brought to physiological pH for the assay. When bound to the dsDNA template the fluorescent probe exhibits a lower total fluorescence compared to unbound TFO probe. Upon addition of enzyme and ATP, the motor translocates across the dsDNA portion of the substrate, and dissociates the triplex probe from the DNA template (Figure A3.1). At physiological pH the TFO probe cannot rebind to the dsDNA template, and translocation of P74-26 TerL can therefore be monitored by changes in fluorescence caused by dissociation of the free triplex probe.

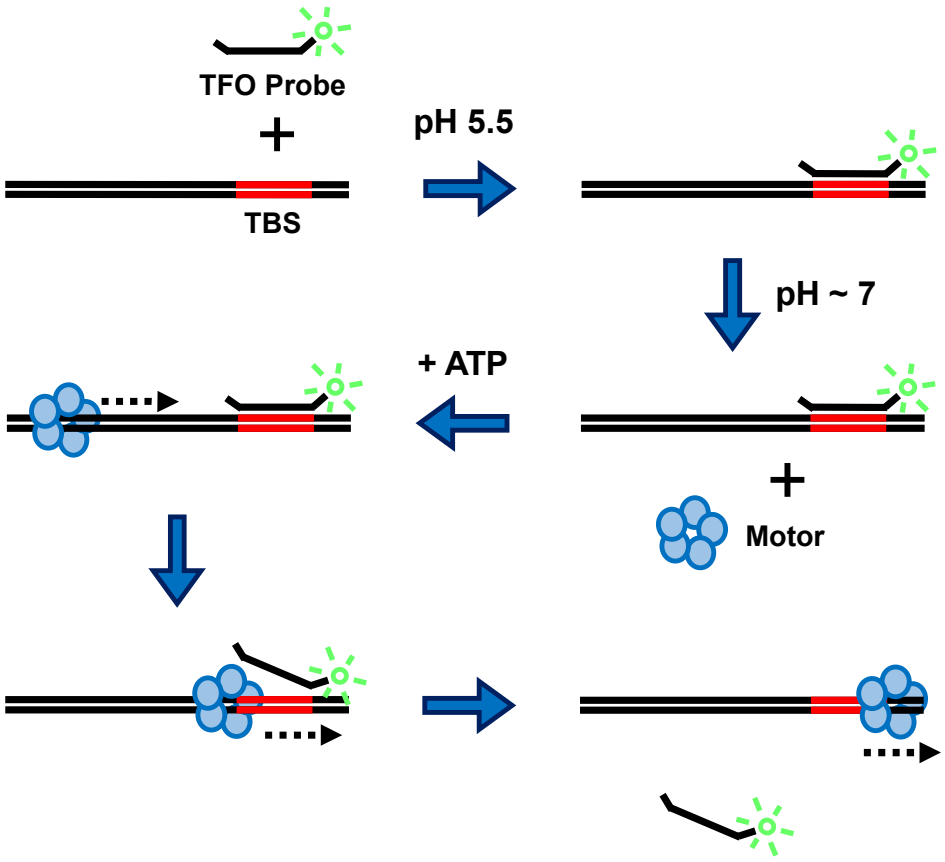

Figure A3.1. A novel triplex dissociation assay for TerL translocation.
Cartoon schematic of the triplex dissociation assay. Double-stranded DNA with an incorporated Triplex Binding Site (TBS, red) is combined with a poly-pyrimidine TAMRA-labeled Triplex Forming Oligonucleotide (TFO, green). At low pH, the TFO binds to the TBS, forming the stable triplex substrate, which exhibits lower fluorescence compared to free TFO probe. After formation at low pH, the triplex substrate is brought to physiological pH. Addition of the terminase motor (blue) and ATP initiates translocation along the DNA substrate. When challenged by the presence of triplex DNA, the motor dissociates the TFO probe, leading to an increase in the fluorescent signal.

Materials and Methods

Purification of Protein

Recombinant Large Terminase protein from phage P74-26 ($TerL^{P74-26}$) and TerL mutants were expressed and purified as previously described in Hilbert et al., 2015.

DNA templates and triplex formation

Oligonucleotides:

TBS – Forward
5'-phosphate/AATTCAAGAAAAGAAAGAAGAAAGAAAGGTAC-3'

TBS – Reverse
5'-phosphate/CTTTCTTTCTTCTTTCTTTTCTTG-3'

TAMRA Triplex Probe
5'-TAMRA/TTCTTTTCTTTCTTCTTTCTTTAACC-3'

A 1000bp fragment derived from the P74-26 genome, spanning across the origin of replication was synthesized and cloned into the pUC57 cloning vector (GenScript). A 959 bp DNA fragment was PCR amplified from the construct, adding either EcoRI or BamHI restriction endonuclease cut sites for downstream processing. The DNA substrate was digested with either EcoRI or BamHI endonucleases (NEB), and cut DNA fragments were purified using a PCR cleanup kit (Qiagen). Triplex Binding Site (TBS) forward and reverse oligonucleotides were combined at a final concentration of 1 μM. To anneal the binding site, the TBS oligos were heated to 95 ˚C for 5 minutes, and slowly

brought to room temperature overnight. The cleaved and purified DNA fragments and annealed triplex binding site were then ligated for 1 hour using T4 Ligase (NEB) to generate a double-stranded DNA template including the Triplex Binding Site. The assembled DNA then served as the core double-stranded template for triplex substrate formation.

Triplex Formation:

The triplex formation reaction was set up as previously described in McClelland et al., 2005. Briefly, 50 nM double-stranded DNA template and 25 nM TAMRA-labeled Triplex Probe were then mixed in triplex formation buffer (10 mM MES pH 5.5, 12.5 mM $MgCl_2$) in a tube wrapped in aluminum foil, and incubated at 20 ˚C overnight. Following the formation reaction, the triplex probe was stored at 4 ˚C prior to use.

Triplex Dissociation Assay

Prior to use, the triplex template was diluted 1:10 in reaction buffer (50 mM Tris pH 8.0, 10 mM MgCl2, 1mM DTT). Final reactions were set up with 2.5 nM DNA/Triplex template and 1 μM $TerL^{p74\text{-}26}$ protein in reaction buffer. ATP was then added to a final concentration of 5 mM to start the reaction, with a final reaction volume of 100 μL.

Triplex formation and dissociation were assayed using either gel electrophoresis or fluorimetric analysis. For gel electrophoresis, 6X Orange G loading dye was added to 20 µL triplex samples to a final 1X concentration. The samples were then run on 0.25X TAM gels (40mM Tris-acetate, 1 mM $MgCl_2$) with 5% polyacrylamide in 1X TAM running buffer (Whitehouse et al., 2003). The gels were run at 4 ˚C for approximately 45 minutes at 70V. The gels were then imaged using a Typhoon FLA 9000 fluorescence imager (GE) equipped with a 532 nm emission filter. Additionally, triplex dissociation reactions were analyzed using a Fluoromax 4 (Horiba). Measurements were made in a 125 µL quartz cuvette (Starna Cells) containing the complete triplex dissociation reaction, triplex template, or free TAMRA-triplex probe. Emission wavelength scans were collected from 565 to 700 nm using an excitation wavelength of 550 nm and a bandpass of 6 nm.

Results and Discussion

Formation of a stable, fluorescently-labeled triplex DNA substrate

For the assembly of the triplex template DNA, we use dsDNA derived from the genome of P74-26. A 959 bp segment of the P74-26 genome was then ligated to a Triplex Binding Site (TBS) DNA duplex fragment, forming an offset binding site for the TFO probe (Figure A3.1). The 5' TAMRA-labeled TFO probe consists of a 22-nucleotide poly-pyrimidine stretch complementary to the TBS, and a 4-nucleotide non-complementary 'foot' at the 3' end. Incubation of 50 nM dsDNA template (including the TBS binding site) with 25 nM labeled TFO probe at pH 5.5 leads to efficient assembly of the triplex substrate. Using an excess of the unlabeled dsDNA template ensures that the TFO is completely sequestered to form fluorescently-labeled triplex DNA substrate (see Figure A3.2).

In order to assess the thermal stability of template DNA, the triplex substrate is then diluted 1:10 into reaction buffer (at pH 7.5) and incubated for 1 hour at given temperatures. Following incubation, triplex templates were run on a non-denaturing Tris-Acetate gel to assess heat denaturation of the template. At temperatures above 40 ˚C, the TFO probe partially dissociates from the dsDNA template (Figure A3.2). Given the partial dissociation at higher temperatures, the preliminary experiments described below were performed at room temperature (25 ˚C).

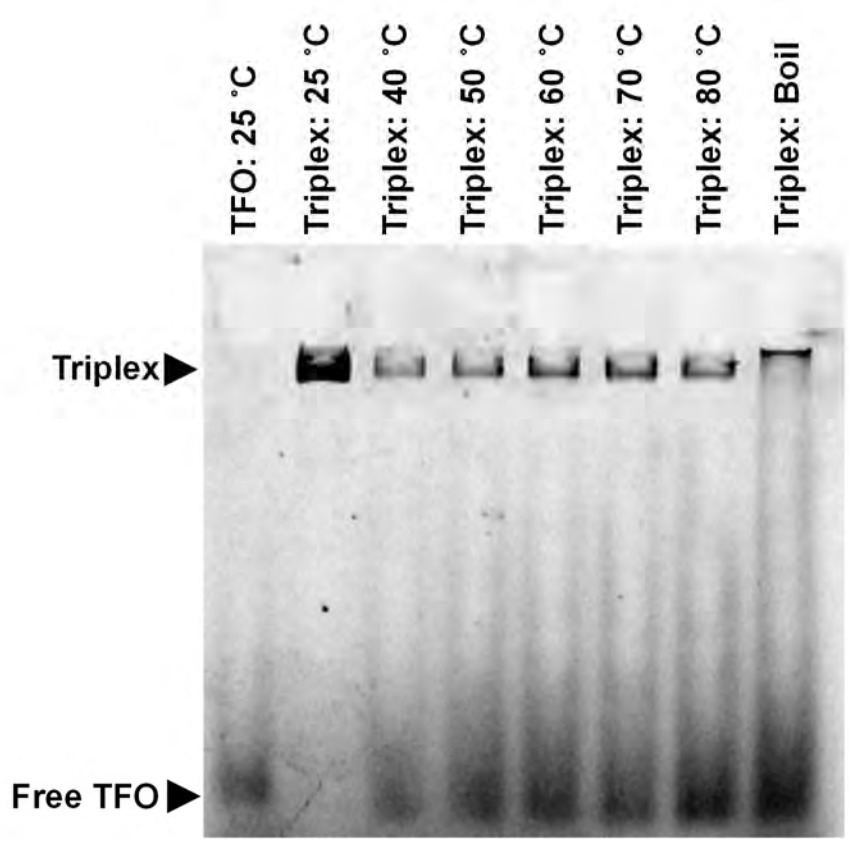

Figure A3.2. Formation of stable TAMRA-labeled triplex DNA substrate.
Duplex DNA substrate and Triplex Forming Oligonucleotide (TFO) form a stable triplex template at pH 5.5. The formed triplex substrate remains stable under physiological conditions (pH 7.5; upper gel band). When incubated at high temperatures, the TFO begins to dissociate from the duplex DNA template (lower gel band).

Monitoring P74-26 TerL translocation via triplex dissociation

We sought to determine whether triplex probe dissociation can be used to monitor TerL translocation *in vitro*. To monitor fluorescence of the TAMRA probe we used an excitation wavelength of 550 nm and monitored emission wavelengths of 565-700 nm to capture peak emission of the probe at ~580 nm. The stable triplex template exhibits significantly lower fluorescence compared to free TFO when diluted in reaction buffer. Addition of 1 μM TerL and 5 mM ATP to the triplex substrate leads to a 2-fold increase in TAMRA fluorescence, suggesting TerL dissociates TFO from the dsDNA template (Figure A3.3).

We also monitored triplex dissociation in the presence of non-hydrolyzable ATP analogs or ADP to inhibit DNA translocation. In this case, we would expect no changes in fluorescence when compared to the triplex template alone. Interestingly, in the presence of ATPγS, AMP-PNP or ADP, TAMRA fluorescence appears to modestly increase compared to the stable triplex substrate, but recovery is less efficient than in the presence of ATP (Figure A3.4). One possible explanation is that TerL can still bind to template DNA, and binding may partially dissociate the TFO. Previous work performed in our lab suggests that in the presence of ADP, TerL can weakly bind to a given DNA substrate (Hilbert et al.,

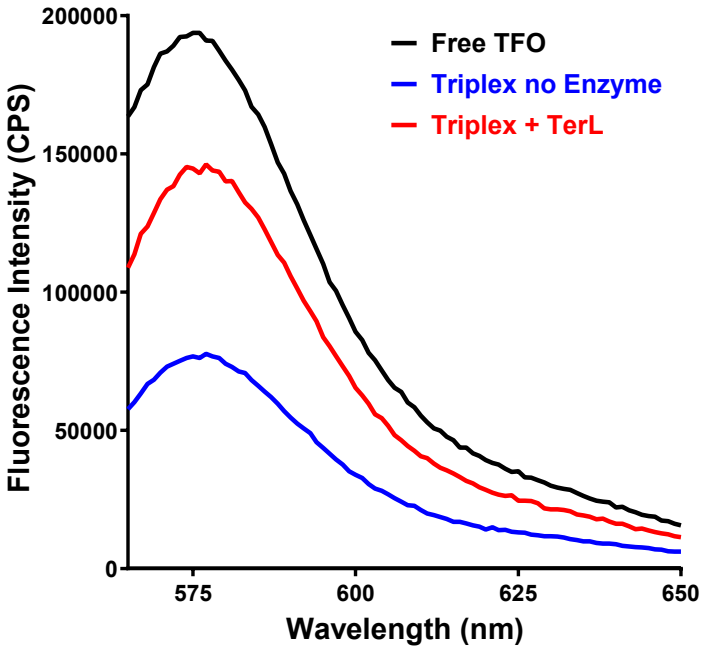

Figure A3.3. TerL translocation leads to partial rescue of triplex probe fluorescence.
The fluorescently-labeled Triplex Forming Oligonucleotide (TFO) exhibits higher fluorescence when not bound to a duplex DNA template (black), versus the assembled Triplex substrate (blue). Upon addition of the Large Terminase (TerL) and ATP (red), the TFO is displaced leading to an increase in fluorescence.

2017). In the presence of a non-hydrolyzable ATP analog, TerL can likewise bind to dsDNA, and may dissociate TFO without translocating along the DNA template. Further experimentation will be necessary to determine the mechanism of TFO dissociation, and establish this method as a viable assay for monitoring DNA translocation.

Future directions

In the future, a triplex dissociation assay optimized for stopped-flow or continuous flow methodologies will be beneficial for kinetic analysis of P74-26 TerL translocation, similar to experiments described previously using other translocases (Graham et al., 2010). An established triplex dissociation assay using TerL can also be adapted to a single-molecule format. Recently-developed 'DNA curtain' experiments allow for the direct monitoring of interactions between protein and DNA in real-time (Collins et al., 2014). In this method, strands of DNA including multiple triplex binding sites would be fixed in a flow cell. A solution of TerL and ATP would then be flowed in, then translocation and subsequent triplex probe dissociation can be monitored at the single-molecule level. Using TIRF microscopy to monitor fluorescence changes caused by TerL translocation, one can potentially monitor hundreds of individual translocation reactions in real-time to discern kinetics of P74-26 TerL locomotion.

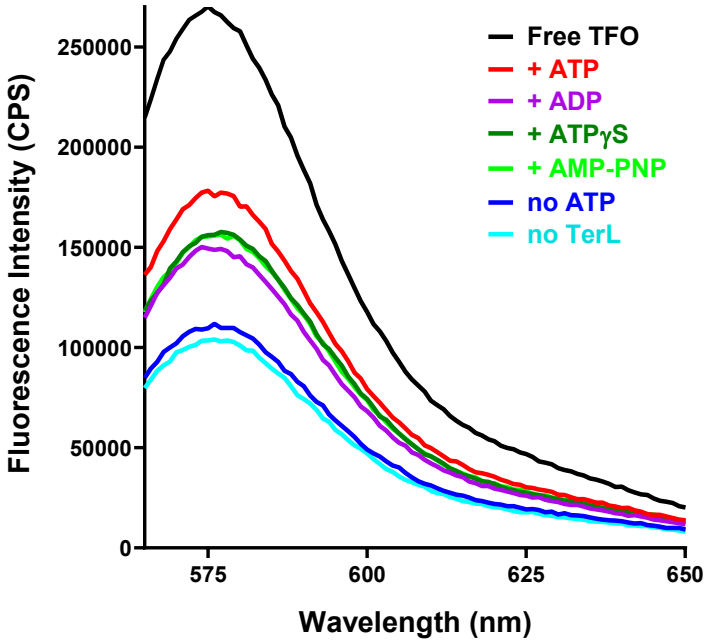

Figure A3.4. Triplex dissociation in the presence of ATP analogs.
The TFO probe is displaced by TerL in the presence of ATP, causing an increase in fluorescence (blue; red). In the presence of ADP (purple) or ATP analog (green), fluorescence is partially restored by probe dissociation, likely caused by DNA binding rather than translocation.

REFERENCES

Abad-Zapatero C, Abdel-meguid SS, Johnson JE, et al. Structure of southern bean mosaic virus at 2.8 A resolution. Nature. 1980;286(5768):33-9.

Ackermann HW, Prangishvili D. Prokaryote viruses studied by electron microscopy. Arch Virol. 2012;157(10):1843-9.

Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66(Pt 2):213-21.

Adrian M, Dubochet J, Lepault J, Mcdowall AW. Cryo-electron microscopy of viruses. Nature. 1984;308(5954):32-6.

Agirrezabala X, Velázquez-muriel JA, Gómez-puertas P, Scheres SH, Carazo JM, Carrascosa JL. Quasi-atomic model of bacteriophage t7 procapsid shell: insights into the structure and evolution of a basic fold. Structure. 2007;15(4):461-72.

Aksyuk AA, Kurochkina LP, Fokine A, et al. Structural conservation of the myoviridae phage tail sheath protein fold. Structure. 2011;19(12):1885-94.

Aksyuk AA, Rossmann MG. Bacteriophage assembly. Viruses. 2011;3(3):172-203.

Allison SL, Tao YJ, O'riordain G, Mandl CW, Harrison SC, Heinz FX. Two distinct size classes of immature and mature subviral particles from tick-borne encephalitis virus. J Virol. 2003;77(21):11357-66.

Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402.

Al-zahrani AS, Kondabagil K, Gao S, Kelly N, Ghosh-kumar M, Rao VB. The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. J Biol Chem. 2009;284(36):24490-500.

Andres D, Hanke C, Baxa U, Seul A, Barbirz S, Seckler R. Tailspike interactions with lipopolysaccharide effect DNA ejection from phage P22 particles in vitro. J Biol Chem. 2010;285(47):36768-75.

Baker ML, Hryc CF, Zhang Q, et al. Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. Proc Natl Acad Sci USA. 2013;110(30):12301-6.

Baker TS, Olson NH, Fuller SD. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol Mol Biol Rev. 1999;63(4):862-922, table of contents.

Bammes BE, Rochat RH, Jakana J, Chen DH, Chiu W. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. J Struct Biol. 2012;177(3):589-601.

Basler M, Pilhofer M, Henderson GP, Jensen GJ, Mekalanos JJ. Type VI secretion requires a dynamic contractile phage tail-like structure. Nature. 2012;483(7388):182-6.

Bauer DW, Evilevitch A. Influence of Internal DNA Pressure on Stability and Infectivity of Phage λ. J Mol Biol. 2015;427(20):3189-3200.

Bauer DW, Li D, Huffman J, et al. Exploring the Balance between DNA Pressure and Capsid Stability in Herpesviruses and Phages. J Virol. 2015;89(18):9288-98.

Bayfield OW, Klimuk E, Winkler DC, et al. Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. Proc Natl Acad Sci USA. 2019;116(9):3556-3561.

Bazinet C, King J. Initiation of P22 procapsid assembly in vivo. J Mol Biol. 1988;202(1):77-86.

Bazinet C, King J. The DNA translocating vertex of dsDNA bacteriophage. Annu Rev Microbiol. 1985;39:109-29.

Benson SD, Bamford JK, Bamford DH, Burnett RM. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell. 1999;98(6):825-33.

Berget PB, Poteete AR. Structure and functions of the bacteriophage P22 tail protein. J Virol. 1980;34(1):234-43.

Betts S, King J. There's a right way and a wrong way: in vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. Structure. 1999;7(6):R131-9.

Bhattacharyya SP, Rao VB. A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. Virology. 1993;196(1):34-44.

Black LW, Rao VB. Structure, assembly, and DNA packaging of the bacteriophage T4 head. Adv Virus Res. 2012;82:119-53.

Black LW. DNA packaging in dsDNA bacteriophages. Annu Rev Microbiol. 1989;43:267-92.

Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proc Natl Acad Sci USA. 2006;103(15):5869-74.

Bondy-denomy J, Garcia B, Strum S, et al. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature. 2015;526(7571):136-9.

Bondy-denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature. 2013;493(7432):429-32.

Borges AL, Davidson AR, Bondy-denomy J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. Annu Rev Virol. 2017;4(1):37-59.

Brenner S, Horne RW. A negative staining method for high resolution electron microscopy of viruses. Biochim Biophys Acta. 1959;34:103-10.

Brenner S, Streisinger G, Horne RW, Champe SP, Barnett L, Benzer S, Rees MW. Structural components of bacteriophage. J Mol Biol. 1959;1(3):281-92.

Bungard D, Copple JS, Yan J, et al. Foldability of a Natural De Novo Evolved Protein. Structure. 2017;25(11):1687-1696.e4.

Burmeister WP, Buisson M, Estrozi LF, et al. Structure determination of feline calicivirus virus-like particles in the context of a pseudo-octahedral arrangement. PLoS ONE. 2015;10(3):e0119289.

Butan C, Lokhandwala PM, Purdy JG, Cardone G, Craven RC, Steven AC. Suppression of a morphogenic mutant in Rous sarcoma virus capsid protein by a second-site mutation: a cryoelectron tomography study. J Virol. 2010;84(13):6377-86.

Campbell MG, Cheng A, Brilot AF, et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. Structure. 2012;20(11):1823-8.

Cardone G, Heymann JB, Cheng N, Trus BL, Steven AC. Procapsid assembly, maturation, nuclear exit: dynamic steps in the production of infectious herpesvirions. Adv Exp Med Biol. 2012;726:423-39.

Carragher B, Kisseberth N, Kriegman D, et al. Leginon: an automated system for acquisition of images from vitreous ice specimens. J Struct Biol. 2000;132(1):33-45.

Casjens S, Hayden M. Analysis in vivo of the bacteriophage P22 headful nuclease. J Mol Biol. 1988;199(3):467-74.

Casjens S, Wyckoff E, Hayden M, et al. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J Mol Biol. 1992;224(4):1055-74.

Casjens SR, Gilcrease EB. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Methods Mol Biol. 2009;502:91-111.

Casjens SR, Hendrix RW. Locations and amounts of major structural proteins in bacteriophage lambda. J Mol Biol. 1974;88(2):535-45.

Casjens SR, Molineux IJ. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol. 2012;726:143-79.

Casjens SR. Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol. 2005;8(4):451-8.

Caspar DL, Klug A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol. 1962;27:1-24.

Caspar DL. Movement and self-control in protein assemblies. Quasi-equivalence revisited. Biophys J. 1980;32(1):103-38.

Catalano CE, Cue D, Feiss M. Virus DNA packaging: the strategy used by phage lambda. Mol Microbiol. 1995;16(6):1075-86.

Catalano CE. Viral genome packaging: genetics, structure, and mechanism. Springer; 2005.

Cerritelli ME, Cheng N, Rosenberg AH, Mcpherson CE, Booy FP, Steven AC. Encapsidated conformation of bacteriophage T7 DNA. Cell. 1997;91(2):271-80.

Chang J, Weigele P, King J, Chiu W, Jiang W. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. Structure. 2006;14(6):1073-82.

Chen DH, Baker ML, Hryc CF, et al. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. Proc Natl Acad Sci USA. 2011;108(4):1355-60.

Chen XS, Garcea RL, Goldberg I, Casini G, Harrison SC. Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. Mol Cell. 2000;5(3):557-67.

Chen Z, Sun L, Zhang Z, et al. Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses. Proc Natl Acad Sci USA. 2017;114(39):E8184-E8193.

Cheng Y, Grigorieff N, Penczek PA, Walz T. A primer to single-particle cryo-electron microscopy. Cell. 2015;161(3):438-449.

Choi KH, McPartland J, Kaganman I, Bowman VD, Rothman-denes LB, Rossmann MG. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J Mol Biol. 2008;378(3):726-36.

Choi KH, Morais MC, Anderson DL, Rossmann MG. Determinants of bacteriophage phi29 head morphology. Structure. 2006;14(11):1723-7.

Choi KY, Root M, Mcgregor A. A Novel Non-Replication-Competent Cytomegalovirus Capsid Mutant Vaccine Strategy Is Effective in Reducing Congenital Infection. J Virol. 2016;90(17):7902-19.

Collins BE, Ye LF, Duzdevich D, Greene EC. DNA curtains: novel tools for imaging protein-nucleic acid interactions at the single-molecule level. Methods Cell Biol. 2014;123:217-34.

Colson P, De lamballerie X, Yutin N, et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch Virol. 2013;158(12):2517-21.

Conway JF, Duda RL, Cheng N, Hendrix RW, Steven AC. Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J Mol Biol. 1995;253(1):86-99.

Cope J, Heumann J, Hoenger A. Cryo-electron tomography for structural characterization of macromolecular complexes. Curr Protoc Protein Sci. 2011;Chapter 17:Unit17.13.

Coren JS, Pierce JC, Sternberg N. Headful packaging revisited: the packaging of more than one DNA molecule into a bacteriophage P1 head. J Mol Biol. 1995;249(1):176-84.

Cortines JR, Weigele PR, Gilcrease EB, Casjens SR, Teschke CM. Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. Virology. 2011;421(1):1-11.

Crowther RA, Kiselev NA, Böttcher B, et al. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell. 1994;77(6):943-50.

Dai X, Gong D, Lim H, et al. Structure and mutagenesis reveal essential capsid protein interactions for KSHV replication. Nature. 2018;553(7689):521-525.

Dai X, Gong D, Xiao Y, Wu TT, Sun R, Zhou ZH. CryoEM and mutagenesis reveal that the smallest capsid protein cements and stabilizes Kaposi's sarcoma-associated herpesvirus capsid. Proc Natl Acad Sci USA. 2015;112(7):E649-56.

Dai X, Zhou ZH. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science. 2018;360(6384)

Dale JW, Greenaway PJ. In vitro packaging of DNA. Methods Mol Biol. 1985;2:245-50.

Dauter Z, Dauter M, Rajashankar KR. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Acta Crystallogr D Biol Crystallogr. 2000;56(Pt 2):232-7.

Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL. Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol. 2012;726:115-42.

Davison AJ. Comparative analysis of the genomes. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge. Cambridge University Press. 2007.

Davison AJ. Overview of Classification. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge. Cambridge University Press. 2007.

Davison AJ. Overview of Classification. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge. Cambridge University Press. 2007.

de la Escosura A, Nolte R, Cornelissen J. Viruses and protein cages as nanocontainers and nanoreactors. J Mater Chem. 2009;(19):2274-78.

Dearborn AD, Wall EA, Kizziah JL, et al. Competing scaffolding proteins determine capsid size during mobilization of pathogenicity islands. Elife. 2017;6.

Delgui LR, Rodríguez JF. Virus Maturation. In: Mateu M. Structure and physics of viruses. Dordrecht. Springer. 2013;68:395-415.

Dembo M, Torney DC, Saxman K, Hammer D. The reaction-limited kinetics of membrane-to-surface adhesion and detachment. Proc R Soc Lond, B, Biol Sci. 1988;234(1274):55-83.

Dimaio F, Yu X, Rensen E, Krupovic M, Prangishvili D, Egelman EH. Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. Science. 2015;348(6237):914-7.

Doermann AH, Eiserling FA, Boehner L. Genetic control of capsid length in bacteriophage T4. I. Isolation and preliminary description of four new mutants. J Virol. 1973;12(2):374-85.

Dokland T, Murialdo H. Structural transitions during maturation of bacteriophage lambda capsids. J Mol Biol. 1993;233(4):682-94.

Dowah ASA, Clokie MRJ. Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. Biophys Rev. 2018;10(2):535-542.

Driedonks RA, Engel A, Tenheggeler B, Van driel . Gene 20 product of bacteriophage T4 its purification and structure. J Mol Biol. 1981;152(4):641-62.

Dröge A, Tavares P. In vitro packaging of DNA of the Bacillus subtilis bacteriophage SPP1. J Mol Biol. 2000;296(1):103-15.

Dubochet J, Adrian M, Chang JJ, et al. Cryo-electron microscopy of vitrified specimens. Q Rev Biophys. 1988;21(2):129-228.

Duda RL, Hempel J, Michel H, Shabanowitz J, Hunt D, Hendrix RW. Structural transitions during bacteriophage HK97 head assembly. J Mol Biol. 1995;247(4):618-35.

Duda RL, Oh B, Hendrix RW. Functional domains of the HK97 capsid maturation protease and the mechanisms of protein encapsidation. J Mol Biol. 2013;425(15):2765-81.

Duda RL. Protein chainmail: catenated protein in viral capsids. Cell. 1998;94(1):55-60.

Earnshaw WC, Casjens SR. DNA packaging by the double-stranded DNA bacteriophages. Cell. 1980;21(2):319-31.

Earnshaw WC, King J. Structure of phage P22 coat protein aggregates formed in the absence of the scaffolding protein. J Mol Biol. 1978;126(4):721-47.

Effantin G, Boulanger P, Neumann E, Letellier L, Conway JF. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. J Mol Biol. 2006;361(5):993-1002.

Eiserling FA, Geiduschek EP, Epstein RH, Metter EJ. Capsid size and deoxyribonucleic acid length: the petite variant of bacteriophage T4. J Virol. 1970;6(6):865-76.

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 12 Pt 1):2126-32.

Ercius P, Alaidi O, Rames MJ, Ren G. Electron Tomography: A Three-Dimensional Analytic Tool for Hard and Soft Materials Research. Adv Mater Weinheim. 2015;27(38):5638-63.

Evilevitch A, Castelnovo M, Knobler CM, Gelbart WM. Measuring the force ejecting DNA from phage. J Phys Chem. 2004;108(21):6838-43.

Evilevitch A, Lavelle L, Knobler CM, Raspaud E, Gelbart WM. Osmotic pressure inhibition of DNA from phage. Proc Natl Acad Sci USA. 2003;100(16):9292-5.

Falco SC, Laan KV, Rothman-denes LB. Virion-associated RNA polymerase required for bacteriophage N4 development. Proc Natl Acad Sci USA. 1977;74(2):520-3.

Fane BA, Prevelige PE. Mechanism of scaffolding-assisted viral assembly. Adv Protein Chem. 2003;64:259-99.

Faruqi AR, Mcmullan G. Electronic detectors for electron microscopy. Q Rev Biophys. 2011;44(3):357-90.

Ferreira D, Hernandez R, Horton M, Brown DT. Morphological variants of Sindbis virus produced by a mutation in the capsid protein. Virology. 2003;307(1):54-66.

Fiedler JD, Higginson C, Hovlid ML, et al. Engineered mutations change the structure and stability of a virus-like particle. Biomacromolecules. 2012;13(8):2339-48.

Finch AJ, Kim JR. Thermophilic Proteins as Versatile Scaffolds for Protein Engineering. Microorganisms. 2018;6(4)

Firman K, Szczelkun MD. Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. EMBO J. 2000;19(9):2094-102.

Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG. Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci USA. 2004;101(16):6003-8.

Fokine A, Leiman PG, Shneider MM, et al. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. Proc Natl Acad Sci USA. 2005;102(20):7163-8.

Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage. 2014;4(1):e28281.

Ford RJ, Barker AM, Bakker SE, et al. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. J Mol Biol. 2013;425(6):1050-64.

Forrer P, Chang C, Ott D, Wlodawer A, Plückthun A. Kinetic stability and crystal structure of the viral capsid protein SHP. J Mol Biol. 2004;344(1):179-93.

Frank-kamenetskii MD, Mirkin SM. Triplex DNA structures. Annu Rev Biochem. 1995;64:65-95.

Fraser JS, Yu Z, Maxwell KL, Davidson AR. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. J Mol Biol. 2006;359(2):496-507.

Fujisawa H, Morita M. Phage DNA packaging. Genes Cells. 1997;2(9):537-45.

Fujisawa H, Shibata H, Kato H. Analysis of interactions among factors involved in the bacteriophage T3 DNA packaging reaction in a defined in vitro system. Virology. 1991;185(2):788-94.

Fuller DN, Raymer DM, Kottadiel VI, Rao VB, Smith DE. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. Proc Natl Acad Sci USA. 2007;104(43):16868-73.

Fuller RB. Tensegrity. Portfolio and Art News Annual. 1961.

Gertsman I, Gan L, Guttman M, et al. An unexpected twist in viral capsid maturation. Nature. 2009;458(7238):646-50.

Gilcrease EB, Winn-stapley DA, Hewitt FC, Joss L, Casjens SR. Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. J Bacteriol. 2005;187(6):2050-7.

Gope R, Serwer P. Bacteriophage P22 in vitro DNA packaging monitored by agarose gel electrophoresis: rate of DNA entry into capsids. J Virol. 1983;47(1):96-105.

Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond, B, Biol Sci. 1979;205(1161):581-98.

Graham JE, Sherratt DJ, Szczelkun MD. Sequence-specific assembly of FtsK hexamers establishes directional translocation on DNA. Proc Natl Acad Sci USA. 2010;107(47):20263-8.

Grant T, Rohou A, Grigorieff N. cisTEM, user-friendly software for single-particle image processing. Elife. 2018;7.

Grigorieff N. FREALIGN: high-resolution refinement of single particle structures. J Struct Biol. 2007;157(1):117-25.

Grose JH, Belnap DM, Jensen JD, et al. The genomes, proteomes, and structures of three novel phages that infect the Bacillus cereus group and carry putative virulence factors. J Virol. 2014;88(20):11846-60.

Gualfetti PJ, Bilsel O, Matthews CR. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from Escherichia coli. Protein Sci. 1999;8(8):1623-35.

Guinn EJ, Jagannathan B, Marqusee S. Single-molecule chemo-mechanical unfolding reveals multiple transition state barriers in a small single-domain protein. Nat Commun. 2015;6:6861.

Guo F, Jiang W. Single particle cryo-electron microscopy and 3-D reconstruction of viruses. Methods Mol Biol. 2014;1117:401-43.

Guo F, Liu Z, Fang PA, et al. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. Proc Natl Acad Sci USA. 2014;111(43):E4606-14.

Guu TS, Liu Z, Ye Q, et al. Structure of the hepatitis E virus-like particle suggests mechanisms for virus assembly and receptor binding. Proc Natl Acad Sci USA. 2009;106(31):12992-7.

Hall CE. Electron densitometry of stained virus particles. J Biophys Biochem Cytol. 1955;1(1):1-12.

Harrington LB, Doxzen KW, Ma E, et al. A Broad-Spectrum Inhibitor of CRISPR-Cas9. Cell. 2017;170(6):1224-1233.e15.

Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. Tomato bushy stunt virus at 2.9 A resolution. Nature. 1978;276(5686):368-73.

Hartwieg E, Bazinet C, King J. The DNA injection apparatus of phage p22. Biophys J. 1986;49(1):24-6.

Harvey SC. The scrunchworm hypothesis: transitions between A-DNA and B-DNA provide the driving force for genome packaging in double-stranded DNA bacteriophages. J Struct Biol. 2015;189(1):1-8.

Hatfull GF. Bacteriophage genomics. Curr Opin Microbiol. 2008;11(5):447-53.

Hendrix RW and Johnson JE. Bacteriophage HK97 capsid assembly and maturation. In: Rossmann MG, Rao VB. Boston, MA. Springer US. 2011;351-63.

Hendrix RW. Symmetry mismatch and DNA packaging in large bacteriophages. Proc Natl Acad Sci USA. 1978;75(10):4779-83.

Hennig M, Sterner R, Kirschner K, Jansonius JN. Crystal structure at 2.0 A resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile Thermotoga maritima: possible determinants of protein stability. Biochemistry. 1997;36(20):6009-16.

Hernando-pérez M, Lambert S, Nakatani-webster E, Catalano CE, De pablo PJ. Cementing proteins provide extra mechanical stabilization to viral cages. Nat Commun. 2014;5:4520.

Hilbert BJ, Hayes JA, Stone NP, Duffy CM, Sankaran B, Kelch BA. Structure and mechanism of the ATPase that powers viral genome packaging. Proc Natl Acad Sci USA. 2015;112(29):E3792-9.

Hilbert BJ, Hayes JA, Stone NP, Xu RG, Kelch BA. The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. Nucleic Acids Res. 2017;45(6):3591-3605.

Hochstein R, Bollschweiler D, Dharmavaram S, et al. Structural studies of tailed spindle virus reveal a structural paradigm used in the assembly of spindle-shaped viruses. Proc Natl Acad Sci USA. 2018;115(9):2120-2125.

Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010;38(Web Server issue):W545-9.

Honess RW. Herpes simplex and 'the herpes complex': diverse observations and a unifying hypothesis. The eighth Fleming lecture. J Gen Virol. 1984;65:2077-107.

Hong C, Pietilä MK, Fu CJ, Schmid MF, Bamford DH, Chiu W. Lemon-shaped halo archaeal virus His1 with uniform tail but variable capsid structure. Proc Natl Acad Sci USA. 2015;112(8):2449-54.

Howatson AF, Kemp CL. The structure of tubular head forms of bacteriophage lambda; relation to the capsid structure of petit lambda and normal lambda heads. Virology. 1975;67(1):80-4.

Hryc CF, Chen DH, Afonine PV, et al. Accurate model annotation of a near-atomic resolution cryo-EM map. Proc Natl Acad Sci USA. 2017;114(12):3103-3108.

Hsiao CL, Black LW. DNA packaging and the pathway of bacteriophage T4 head assembly. Proc Natl Acad Sci USA. 1977;74(9):3652-6.

Hsiao CL, Black LW. Head morphogenesis of bacteriophage T4. III. The role of gene 20 in DNA packaging. Virology. 1978;91(1):26-38.

Huang DM, Chandler D. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. Proc Natl Acad Sci USA. 2000;97(15):8324-7.

Huang RK, Khayat R, Lee KK, et al. The Prohead-I structure of bacteriophage HK97: implications for scaffold-mediated control of particle assembly and maturation. J Mol Biol. 2011;408(3):541-54.

Huet A, Conway JF, Letellier L, Boulanger P. In vitro assembly of the T=13 procapsid of bacteriophage T5 with its scaffolding domain. J Virol. 2010;84(18):9350-8.

Huxley HE, Zubay G. Preferential staining of nucleic acid-containing structures for electron microscopy. J Biophys Biochem Cytol. 1961;11:273-96.

Ilca SL, Sun X, El omari K, et al. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. Nature. 2019;570:252-6.

Imber R, Tsugita A, Wurtz M, Hohn T. Outer surface protein of bacteriophage lambda. J Mol Biol. 1980;139(3):277-95.

Israel V. E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. J Virol. 1977;23(1):91-7.

Iwasaki K, Omura T. Electron tomography of the supramolecular structure of virus-infected cells. Curr Opin Struct Biol. 2010;20(5):632-9.

J.D. Baines, S.K. Weller, Cleavage and Packaging of Herpes Simplex Virus 1 DNA, in: Catalano CE, Viral Genome Packaging Machines: Genetics, Structure, and Mechanism, 1st ed., Springer, 2005:135–150.

Jagannathan B, Elms PJ, Bustamante C, Marqusee S. Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. Proc Natl Acad Sci USA. 2012;109(44):17820-5.

Jiang W, Chang J, Jakana J, Weigele P, King J, Chiu W. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature. 2006;439(7076):612-6.

Jiang W, Tang L. Atomic cryo-EM structures of viruses. Curr Opin Struct Biol. 2017;46:122-129.

Jin L, Milazzo AC, Kleinfelder S, et al. Applications of direct detection device in transmission electron microscopy. J Struct Biol. 2008;161(3):352-8.

Jin Y, Sdao SM, Dover JA, et al. Bacteriophage P22 ejects all of its internal proteins before its genome. Virology. 2015;485:128-34.

Johnson JE, Chiu W. DNA packaging and delivery machines in tailed bacteriophages. Curr Opin Struct Biol. 2007;17(2):237-43.
Johnson JE. Virus particle maturation: insights into elegantly programmed nanomachines. Curr Opin Struct Biol. 2010;20(2):210-6.

Kasson P, Dimaio F, Yu X, et al. Model for a novel membrane envelope in a filamentous hyperthermophilic virus. Elife. 2017;6

Kathuria SV, Chan YH, Nobrega RP, Özen A, Matthews CR. Clusters of isoleucine, leucine, and valine side chains define cores of stability in high-energy states of globular proteins: Sequence determinants of structure and stability. Protein Sci. 2016;25(3):662-75.

Kelly BJ, Fraefel C, Cunningham AL, Diefenbach RJ. Functional roles of the tegument proteins of herpes simplex virus type 1. Virus Res. 2009;145(2):173-86.

Khayat R, Johnson JE. Pass the jelly rolls. Structure. 2011;19(7):904-6.

Knapp S, Kardinahl S, Hellgren N, Tibbelin G, Schäfer G, Ladenstein R. Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon Sulfolobus acidocaldarius at 2.2 A resolution. J Mol Biol. 1999;285(2):689-702.

Koning RI, Gomez-blanco J, Akopjana I, et al. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. Nat Commun. 2016;7:12524.

Koonin EV, Dolja VV. A virocentric perspective on the evolution of life. Curr Opin Virol. 2013;3(5):546-57.

Krupovic M, Bamford DH. Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. Curr Opin Virol. 2011;1(2):118-24.

Kühlbrandt W. Biochemistry. The resolution revolution. Science. 2014;343(6178):1443-4.

Kumar S, Tsai C, Nussinov R. Factors enhancing protein thermostability. Protein Engineering, Design and Selection. 2000;13(3):179-91.

Lander GC, Baudoux AC, Azam F, Potter CS, Carragher B, Johnson JE. Capsomer dynamics and stabilization in the T = 12 marine bacteriophage SIO-2 and its procapsid studied by CryoEM. Structure. 2012;20(3):498-503.

Lander GC, Evilevitch A, Jeembaeva M, Potter CS, Carragher B, Johnson JE. Bacteriophage lambda stabilization by auxiliary protein gpD: timing, location, and mechanism of attachment determined by cryo-EM. Structure. 2008;16(9):1399-406.

Lee CS, Guo P. Sequential interactions of structural proteins in phage phi 29 procapsid assembly. J Virol. 1995;69(8):5024-32.

Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. 2018;46(D1):D708-D717.

Leiman PG, Chipman PR, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell. 2004;118(4):419-29.

Leiman PG, Shneider MM. Contractile tail machines of bacteriophages. Adv Exp Med Biol. 2012;726:93-114.

Lepault J, Dubochet J, Baschong W, Kellenberger E. Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples. EMBO J. 1987;6(5):1507-12.

Li X, Mooney P, Zheng S, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Methods. 2013;10(6):584-90.

Lim JH, Yu YG, Han YS, et al. The crystal structure of an Fe-superoxide dismutase from the hyperthermophile Aquifex pyrophilus at 1.9 A resolution: structural basis for thermostability. J Mol Biol. 1997;270(2):259-74.

Liu F and Zhou ZH. Comparative virion structures of human herpesviruses. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge. Cambridge University Press. 2007.

Liu F, Roizman B. The herpes simplex virus 1 gene encoding a protease also contains within its coding domain the gene encoding the more abundant substrate. J Virol. 1991;65(10):5149-56.

Liu X, Zhang Q, Murata K, et al. Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. Nat Struct Mol Biol. 2010;17(7):830-6.

Liu Y, Jih J, Dai X, Bi GQ, Zhou ZH. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. Nature. 2019;570:257-61.

Liu Y, Osinski T, Wang F, et al. Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. Nat Commun. 2018;9(1):3360.

Llorente JM, Hernández-rojas J, Bretón J. A minimal representation of the self-assembly of virus capsids. Soft Matter. 2014;10(20):3560-9.

Lokareddy RK, Sankhala RS, Roy A, et al. Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. Nat Commun. 2017;8:14310.

Lurz R, Orlova EV, Günther D, et al. Structural organisation of the head-to-tail interface of a bacterial virus. J Mol Biol. 2001;310(5):1027-37.

Madej T, Lanczycki CJ, Zhang D, et al. MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic Acids Res. 2014;42(Database issue):D297-303.

Mannige RV, Brooks CL. Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling. Proc Natl Acad Sci USA. 2009;106(21):8531-6.

Mannige RV, Brooks CL. Periodic table of virus capsids: implications for natural selection and design. PLoS ONE. 2010;5(3):e9423.

Mateu MG. Assembly, stability and dynamics of virus capsids. Arch Biochem Biophys. 2013;531(1-2):65-79.

Maxwell KL. Phages Fight Back: Inactivation of the CRISPR-Cas Bacterial Immune System by Anti-CRISPR Proteins. PLoS Pathog. 2016;12(1):e1005282.

Mcclelland SE, Dryden DT, Szczelkun MD. Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases. J Mol Biol. 2005;348(4):895-915.

McElwee M, Vijayakrishnan S, Rixon F, Bhella D. Structure of the herpes simplex virus portal-vertex. PLoS Biol. 2018;16(6):e2006191.

McMullan G, Chen S, Henderson R, Faruqi AR. Detective quantum efficiency of electron area detectors in electron microscopy. Ultramicroscopy. 2009;109(9):1126-43.

Menéndez-conejero R, Nguyen TH, Singh AK, et al. Structure of a Reptilian Adenovirus Reveals a Phage Tailspike Fold Stabilizing a Vertebrate Virus Capsid. Structure. 2017;25(10):1562-1573.e5.

Mettenleiter TC. Herpesvirus assembly and egress. J Virol. 2002;76(4):1537-47.

Milazzo AC, Cheng A, Moeller A, et al. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. J Struct Biol. 2011;176(3):404-8.

Milazzo AC, Moldovan G, Lanman J, et al. Characterization of a direct detection device imaging camera for transmission electron microscopy. Ultramicroscopy. 2010;110(7):744-7.

Minakhin L, Goel M, Berdygulova Z, et al. Genome comparison and proteomic characterization of Thermus thermophilus bacteriophages P23-45 and P74-26: siphoviruses with triplex-forming sequences and the longest known tails. J Mol Biol. 2008;378(2):468-80.

Minten IJ, Hendriks LJ, Nolte RJ, Cornelissen JJ. Controlled encapsulation of multiple proteins in virus capsids. J Am Chem Soc. 2009;131(49):17771-3.

Molineux IJ. Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it?. Virology. 2006;344(1):221-9.

Moody MF. Geometry of phage head construction. J Mol Biol. 1999;293(2):401-33.

Mooney P,Contarato D, Denes P, Gubbens A, Lee B, Lent M, Agard D. A High-Speed Electron-Counting Direct Detection Camera for TEM. Microsc and Microanal. 2011;(17):1004-5.

Morais MC, Kanamaru S, Badasso MO, et al. Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. Nat Struct Biol. 2003;10(7):572-6.

Moreno F, Bluzat-moreno FG. Evidence that the neck appendages are adsorption organelles in Bacillus subtilis bacteriophage phi29. J Virol. 1978;27(3):831-4.

Mullaney JM, Black LW. Bacteriophage T4 capsid packaging and unpackaging of DNA and proteins. Methods Mol Biol. 2014;1108:69-85.

Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol. 2002;321(5):741-65.

Nagler FP, Rake G. The Use of the Electron Microscope in Diagnosis of Variola, Vaccinia, and Varicella. J Bacteriol. 1948;55(1):45-51.

Nakamura N, Shimizu Y, Shinkawa T, et al. Automated specimen search in cryo-TEM observation with DIFF-defocus imaging. J Electron Microsc (Tokyo). 2010;59(4):299-310.

Nandhagopal N, Simpson AA, Gurnon JR, et al. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc Natl Acad Sci USA. 2002;99(23):14758-63.

Newcomb WW, Homa FL, Brown JC. Involvement of the portal at an early step in herpes simplex virus capsid assembly. J Virol. 2005;79(16):10540-6.

Newcomb WW, Juhas RM, Thomsen DR, et al. The UL6 gene product forms the portal for entry of DNA into the herpes simplex virus capsid. J Virol. 2001;75(22):10923-32.

Newcomer RL, Schrad JR, Gilcrease EB, et al. The phage L capsid decoration protein has a novel OB-fold and an unusual capsid binding strategy. Elife. 2019;8.

Oh B, Moyer CL, Hendrix RW, Duda RL. The delta domain of the HK97 major capsid protein is essential for assembly. Virology. 2014;456-457:171-8.

Olia AS, Prevelige PE, Johnson JE, Cingolani G. Three-dimensional structure of a viral genome-delivery portal vertex. Nat Struct Mol Biol. 2011;18(5):597-603.

Oshima T, Imahori K. Description of *Thermus thermophilus*, a nonsporulating thermophilic bacterium from a Japanese thermal spa. Int J Syst Bacteriol. 1974;24:102-12.

Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Meth Enzymol. 1997;276:307-26.

Pan R, Gorny MK, Zolla-pazner S, Kong XP. The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel. J Virol. 2015;89(15):8003-10.

Parent KN, Khayat R, Tu LH, et al. P22 coat protein structures reveal a novel mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. Structure. 2010;18(3):390-401.

Parent KN, Schrad JR, Cingolani G. Breaking Symmetry in Viral Icosahedral Capsids as Seen through the Lenses of X-ray Crystallography and Cryo-Electron Microscopy. Viruses. 2018;10(2).

Patterson DP, Schwarz B, Waters RS, Gedeon T, Douglas T. Encapsulation of an enzyme cascade within the bacteriophage P22 virus-like particle. ACS Chem Biol. 2014;9(2):359-65.

Pawluk A, Amrani N, Zhang Y, et al. Naturally Occurring Off-Switches for CRISPR-Cas9. Cell. 2016;167(7):1829-1838.e9.

Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol. 2018;16(1):12-17.

Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proc Natl Acad Sci USA. 2009;106(11):4160-5.

Perlmutter JD, Hagan MF. Mechanisms of virus assembly. Annu Rev Phys Chem. 2015;66:217-39.

Petsko GA. Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". Meth Enzymol. 2001;334:469-78.

Pietilä MK, Laurinmäki P, Russell DA, et al. Insights into head-tailed viruses infecting extremely halophilic archaea. J Virol. 2013;87(6):3248-60.

Prasad BV, Prevelige PE, Marietta E, et al. Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. J Mol Biol. 1993;231(1):65-74.

Prasad BV, Schmid MF. Principles of virus structural organization. Adv Exp Med Biol. 2012;726:17-47.

Preux O, Durand D, Huet A, et al. A two-state cooperative expansion converts the procapsid shell of bacteriophage T5 into a highly stable capsid isomorphous to the final virion head. J Mol Biol. 2013;425(11):1999-2014.

Prevelige PE, Fane BA. Building the machines: scaffolding protein functions during bacteriophage morphogenesis. Adv Exp Med Biol. 2012;726:325-50.

Prevelige PE, Thomas D, King J. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. Biophys J. 1993;64(3):824-35.

Prevelige PE, Thomas D, King J. Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. J Mol Biol. 1988;202(4):743-57.

Ptchelkine D, Gillum A, Mochizuki T, et al. Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. Nat Commun. 2017;8(1):1436.

Purohit PK, Inamdar MM, Grayson PD, Squires TM, Kondev J, Phillips R. Forces during bacteriophage DNA packaging and ejection. Biophys J. 2005;88(2):851-66.

Qin L, Fokine A, O'donnell E, Rao VB, Rossmann MG. Structure of the small outer capsid protein, Soc: a clamp for stabilizing capsids of T4-like phages. J Mol Biol. 2010;395(4):728-41.

Qiu X. Heat induced capsid disassembly and DNA release of bacteriophage λ. PLoS ONE. 2012;7(7):e39793.

Rao VB, Black LW. Structure and assembly of bacteriophage T4 head. Virol J. 2010;7:356.

Rao VB, Feiss M. The bacteriophage DNA packaging motor. Annu Rev Genet. 2008;42:647-81.

Rao VB, Thaker V, Black LW. A phage T4 in vitro packaging system for cloning long DNA molecules. Gene. 1992;113(1):25-33.

Ray K, Sabanayagam CR, Lakowicz JR, Black LW. DNA crunching by a viral packaging motor: Compression of a procapsid-portal stalled Y-DNA substrate. Virology. 2010;398(2):224-32.

Razvi A, Scholtz JM. Lessons in stability from thermophilic proteins. Protein Sci. 2006;15(7):1569-78.

Rice G, Stedman K, Snyder J, et al. Viruses from extreme thermal environments. Proc Natl Acad Sci USA. 2001;98(23):13341-5.

Rice G, Tang L, Stedman K, et al. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc Natl Acad Sci USA. 2004;101(20):7716-20.

Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem. 1981;34:167-339.

Rixon FJ, Schmid MF. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. Curr Opin Virol. 2014;5:105-10.

Rizzo AA, Suhanovsky MM, Baker ML, et al. Multiple functional roles of the accessory I-domain of bacteriophage P22 coat protein revealed by NMR structure and CryoEM modeling. Structure. 2014;22(6):830-41.

Roberts MM, White JL, Grütter MG, Burnett RM. Three-dimensional structure of the adenovirus major coat protein hexon. Science. 1986;232(4754):1148-51.

Rochat RH, Liu X, Murata K, Nagayama K, Rixon FJ, Chiu W. Seeing the portal in herpes simplex virus type 1 B capsids. J Virol. 2011;85(4):1871-4.

Rohs R, Etchebest C, Lavery R. Unraveling proteins: a molecular mechanics study. Biophys J. 1999;76(5):2760-8.

Ross PD, Black LW, Bisher ME, Steven AC. Assembly-dependent conformational changes in a viral capsid protein. Calorimetric comparison of successive conformational states of the gp23 surface lattice of bacteriophage T4. J Mol Biol. 1985;183(3):353-64.

Rossmann MG, Arnold E, Erickson JW, et al. Structure of a human common cold virus and functional relationship to other picornaviruses. Nature. 1985;317(6033):145-53.

Rossmann MG, Johnson JE. Icosahedral RNA virus structure. Annu Rev Biochem. 1989;58:533-73.

Ruska E 1987; Nobel lecture: the development of the electron microscope and of electron microscopy. Biosci Rep. 7: 607- 629

Saad A, Zhou ZH, Jakana J, Chiu W, Rixon FJ. Roles of triplex and scaffolding proteins in herpes simplex virus type 1 capsid formation suggested by structures of recombinant particles. J Virol. 1999;73(8):6821-30.

Sae-ueng U, Liu T, Catalano CE, Huffman JB, Homa FL, Evilevitch A. Major capsid reinforcement by a minor protein in herpesviruses and phage. Nucleic Acids Res. 2014;42(14):9096-107.

San martín C, Van raaij MJ. The so far farthest reaches of the double jelly roll capsid protein fold. Virol J. 2018;15(1):181.

Sathaliyawala T, Islam MZ, Li Q, Fokine A, Rossmann MG, Rao VB. Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. Mol Microbiol. 2010;77(2):444-55.

Schmid MF, Hecksel CW, Rochat RH, Bhella D, Chiu W, Rixon FJ. A tail-like assembly at the portal vertex in intact herpes simplex type-1 virions. PLoS Pathog. 2012;8(10):e1002961.

Schrag JD, Huang W, Sivaraman J, et al. The crystal structure of Escherichia coli MoeA, a protein from the molybdopterin synthesis pathway. J Mol Biol. 2001;310(2):419-31.

Seidel R, Bloom JG, Van noort J, et al. Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. EMBO J. 2005;24(23):4188-97.

Selvarajan sigamani S, Zhao H, Kamau YN, Baines JD, Tang L. The structure of the herpes simplex virus DNA-packaging terminase pUL15 nuclease domain suggests an evolutionary lineage among eukaryotic and prokaryotic viruses. J Virol. 2013;87(12):7140-8.

Shen PS, Domek MJ, Sanz-garcía E, et al. Sequence and structural characterization of great salt lake bacteriophage CW02, a member of the T7-like supergroup. J Virol. 2012;86(15):7907-17.

Shin J, Jiang F, Liu JJ, et al. Disabling Cas9 by an anti-CRISPR DNA mimic. Sci Adv. 2017;3(7):e1701620.

Simler BR, Doyle BL, Matthews CR. Zinc binding drives the folding and association of the homo-trimeric gamma-carbonic anhydrase from Methanosarcina thermophila. Protein Eng Des Sel. 2004;17(3):285-91.

Sivanathan V, Allen MD, De bekker C, et al. The FtsK gamma domain directs oriented DNA translocation by interacting with KOPS. Nat Struct Mol Biol. 2006;13(11):965-72.

Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. Nature. 2001;413(6857):748-52.

Smith DE. Single-molecule studies of viral DNA packaging. Curr Opin Virol. 2011;1(2):134-41.

Snijder J, Radtke K, Anderson F, et al. Vertex-Specific Proteins pUL17 and pUL25 Mechanically Reinforce Herpes Simplex Virus Capsids. J Virol. 2017;91(12)

Spilman MS, Damle PK, Dearborn AD, et al. Assembly of bacteriophage 80α capsids in a Staphylococcus aureus expression system. J Virol. 2013;434(2):242-50.

Stehle T, Gamblin SJ, Yan Y, Harrison SC. The structure of simian virus 40 refined at 3.1 A resolution. Structure. 1996;4(2):165-82.

Sternberg N, Weisberg R. Packaging of coliphage lambda DNA. II. The role of the gene D protein. J Mol Biol. 1977;117(3):733-59.

Sterpone F, Melchionna S. Thermophilic proteins: insight and perspective from in silico experiments. Chem Soc Rev. 2012;41(5):1665-76.

Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. Curr Opin Struct Biol. 2005;15(2):227-36.

Stone NP, Hilbert BJ, Hidalgo D, et al. A Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex Proteins and an Anti-CRISPR Protein. Structure. 2018;26(7):936-947.

Strauss H, King J. Steps in the stabilization of newly packaged DNA during phage P22 morphogenesis. J Mol Biol. 1984;172(4):523-43.

Subramaniam S, Bartesaghi A, Liu J, Bennett AE, Sougrat R. Electron tomography of viruses. Curr Opin Struct Biol. 2007;17(5):596-602.

Suda Y, Okazaki F, Hasegawa Y, et al. Structural characterization of neutral and acidic glycolipids from Thermus thermophilus HB8. PLoS ONE. 2012;7(7):e35067.

Suhanovsky MM, Teschke CM. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. Virology. 2015;479-480:487-97.

Suttle CA. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol. 2007;5(10):801-12.

Szilágyi A, Závodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure. 2000;8(5):493-504.

Tang L, Gilcrease EB, Casjens SR, Johnson JE. Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. Structure. 2006;14(5):837-45.

Tang L, Marion WR, Cingolani G, Prevelige PE, Johnson JE. Three-dimensional structure of the bacteriophage P22 tail machine. EMBO J. 2005;24(12):2087-95.

Taylor NM, Prokhorov NS, Guerrero-ferreira RC, et al. Structure of the T4 baseplate and its function in triggering sheath contraction. Nature. 2016;533(7603):346-52.

Thomsen DR, Newcomb WW, Brown JC, Homa FL. Assembly of the herpes simplex virus capsid: requirement for the carboxyl-terminal twenty-five amino acids of the proteins encoded by the UL26 and UL26.5 genes. J Virol. 1995;69(6):3690-703.

Thuman-commike PA, Greene B, Malinski JA, King J, Chiu W. Role of the scaffolding protein in P22 procapsid size determination suggested by T = 4 and T = 7 procapsid structures. Biophys J. 1998;74(1):559-68.

Tikhonenko AS. Electron-microscopic methods of investigation of bacteriophages. In: Ultrastructure of bacterial viruses. Plenum Press; New York, NY. 1970;294.

Trus BL, Booy FP, Newcomb WW, et al. The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins VP19c and VP23 in assembly. J Mol Biol. 1996;263(3):447-62.

Uchida M, Klem MT, Allen M, Suci P, Flenniken M, Gillitzer E, Varpness Z, Liepold LO, Young M, Douglas T. Biological Containers: protein cages as multifunctional nanoplatforms. Adv Materials. 2007;19(8):1025-42.

Valpuesta JM, Carrascosa JL. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Q Rev Biophys. 1994;27(2):107-155.

Veesler D, Cambillau C. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev. 2011;75(3):423-33.

Veesler D, Johnson JE. Virus maturation. Annu Rev Biophys. 2012;41:473-96.

Veesler D, Ng TS, Sendamarai AK, et al. Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. Proc Natl Acad Sci USA. 2013;110(14):5504-9.

Vogt G, Woell S, Argos P. Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol. 1997;269(4):631-43.

Voorhorst WG, Warner A, De vos WM, Siezen RJ. Homology modelling of two subtilisin-like proteases from the hyperthermophilic archaea Pyrococcus furiosus and Thermococcus stetteri. Protein Eng. 1997;10(8):905-14.

Wang C, Tu J, Liu J, Molineux IJ. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. Nat Microbiol. 2019;4(6):1049-1056.

Wang L, Xu X, Kumar R, et al. Probing DNA clamps with single-molecule force spectroscopy. Nucleic Acids Res. 2013;41(16):7804-14.

Wang X, Ren J, Gao Q, et al. Hepatitis A virus and the origins of picornaviruses. Nature. 2015;517(7532):85-88.

Wang Z, Hardies SC, Fokine A, et al. Structure of the Marine Siphovirus TW1: Evolution of Capsid-Stabilizing Proteins and Tail Spikes. Structure. 2018;26(2):238-248.

White HE, Sherman MB, Brasilès S, et al. Capsid structure and its stability at the late stages of bacteriophage SPP1 assembly. J Virol. 2012;86(12):6768-77.

Whitehouse I, Stockdale C, Flaus A, Szczelkun MD, Owen-hughes T. Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Mol Cell Biol. 2003;23(6):1935-45.

Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. Science. 2000;289(5487):2129-33.

Wilson DP. Protruding Features of Viral Capsids Are Clustered on Icosahedral Great Circles. PLoS ONE. 2016;11(4):e0152319.

Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev. 2000;64(1):69-114.

Wu S, Armache JP, Cheng Y. Single-particle cryo-EM data acquisition by using direct electron detection camera. Microscopy (Oxf). 2016;65(1):35-41.

Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, Rossmann MG. Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Mol Cell. 2009;34(3):375-86.

Xu J, Hendrix RW, Duda RL. Chaperone-protein interactions that mediate assembly of the bacteriophage lambda tail to the correct length. J Mol Biol. 2014;426(5):1004-18.

Yang F, Forrer P, Dauter Z, et al. Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat Struct Biol. 2000;7(3):230-7.

Yang K, Baines JD. Domain within herpes simplex virus 1 scaffold proteins required for interaction with portal protein in infected cells and incorporation of the portal vertex into capsids. J Virol. 2008;82(10):5021-30.

Yang Q, Maluf NK, Catalano CE. Packaging of a unit-length viral genome: the role of nucleotides and the gpD decoration protein in stable nucleocapsid assembly in bacteriophage lambda. J Mol Biol. 2008;383(5):1037-48.

Yap ML, Rossmann MG. Structure and function of bacteriophage T4. Future Microbiol. 2014;9(12):1319-27.

Yeo A, Feiss M. Specific interaction of terminase, the DNA packaging enzyme of bacteriophage lambda, with the portal protein of the prohead. J Mol Biol. 1995;245(2):141-50.

Yu MX, Slater MR, Ackermann HW. Isolation and characterization of Thermus bacteriophages. Arch Virol. 2006;151(4):663-79.

Yu X, Jih J, Jiang J, Zhou ZH. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. Science. 2017;356(6345)

Yu X, Shah S, Lee M, et al. Biochemical and structural characterization of the capsid-bound tegument proteins of human cytomegalovirus. J Struct Biol. 2011;174(3):451-60.

Yunus AA, Lima CD. Purification of SUMO conjugating enzymes and kinetic analysis of substrate conjugation. Methods Mol Biol. 2009;497:167-86.

Zandi R, Reguera D, Bruinsma RF, Gelbart WM, Rudnick J. Origin of icosahedral symmetry in viruses. Proc Natl Acad Sci USA. 2004;101(44):15556-60.

Zhang P, Borgnia MJ, Mooney P, et al. Automated image acquisition and processing using a new generation of 4K x 4K CCD cameras for cryo electron microscopic studies of macromolecular assemblies. J Struct Biol. 2003;143(2):135-44.

Zhang X, Guo H, Jin L, et al. A new topology of the HK97-like fold revealed in Bordetella bacteriophage by cryoEM at 3.5 A resolution. Elife. 2013;2:e01299.

Zhang X, Guo H, Jin L, et al. Correction: A new topology of the HK97-like fold revealed in Bordetella bacteriophage by cryoEM at 3.5 Å resolution. Elife. 2015;4

Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008;9:40.

Zhao H, Li K, Lynn AY, et al. Structure of a headful DNA-packaging bacterial virus at 2.9 Å resolution by electron cryo-microscopy. Proc Natl Acad Sci USA. 2017;114(14):3601-3606.

Zhao H, Speir JA, Matsui T, et al. Structure of a Bacterial Virus DNA-Injection Protein Complex Reveals a Decameric Assembly with a Constricted Molecular Channel. PLoS ONE. 2016;11(2):e0149337.

Zhao L, Kopylov M, Potter CS, Carragher B, Finn MG. Engineering the PP7 Virus Capsid as a Peptide Display Platform. ACS Nano. 2019;13(4):4443-4454.

Zhou HX, Dong F. Electrostatic contributions to the stability of a thermophilic cold shock protein. Biophys J. 2003;84(4):2216-22.

Zlotnick A. Theoretical aspects of virus capsid assembly. J Mol Recognit. 2005;18(6):479-90.

Zwart PH, Afonine PV, Grosse-kunstleve RW, et al. Automated structure solution with the PHENIX suite. Methods Mol Biol. 2008;426:419-35.

Teschke CM, Parent KN. 'Let the phage do the work': using the phage P22 coat protein structures as a framework to understand its folding and assembly mutants. Virology. 2010;401(2):119-30.