2019-09-13

The Mechanism and Regulation of Bacteriophage DNA Packaging Motors

Janelle A. Hayes
University of Massachusetts Medical School

THE MECHANISM AND REGULATION OF BACTERIOPHAGE DNA PACKAGING MOTORS

A Dissertation Presented

By

JANELLE ANNA HAYES

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

SEPTEMBER 13TH, 2019

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

THE MECHANISM AND REGULATION OF BACTERIOPHAGE DNA

PACKAGING MOTORS

A Dissertation Presented
By

JANELLE ANNA HAYES

This work was undertaken in the Graduate School of Biomedical Sciences
Biochemistry and Molecular Pharmacology Program

Under the mentorship of

Brian A. Kelch, PhD., Thesis Advisor

Celia A. Schiffer, PhD., Member of Committee

Daniel N. Bolon, PhD., Member of Committee

Andrei A. Korostelev, PhD., Member of Committee

Paul J. Jardine, PhD., External Member of Committee

William E. Royer, PhD., Chair of Committee

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

September 13th, 2019

Dedication

This work is dedicated to my grandparents, both living and passed. Thanks for loving and believing in me.

Acknowledgments

I'd like to thank my advisor, Dr. Brian Kelch, for his incredible insight and support for this work. Coming into his lab, I knew nothing about structural biology and even the word 'biochemistry' terrified me. Now, hopefully this document proves that is no longer the case. Much of that is due to Brian's thoughtful mentorship and guidance over the past six years, for which I am incredibly grateful.

I'd also like to acknowledge my fellow lab members, both past and present. Dr. Brendan Hilbert, the co-first author on the publication in Chapter II, was an invaluable resource when I first started in the Kelch lab. Thank you Brendan for teaching me crystallography and protein purification. Dr. Caroline Duffy, thank you for showing me how to do an ATPase assay (that was important!), passing on your Excel skills, and for being an all-around A+ friend. To Dr. Nicholas Stone, thank you for being a huge support with the cryoEM work. Learning EM is a difficult task, and your guidance in all aspects of my project has made an incalculable difference. Dr. Christl Gaubitz, thank you for your help with the cryoEM sample preparation and processing suggestions, you are an endless source of knowledge. To Joseph Magrino, thanks for always talking science with me and giving feedback on ideas. To Brendan Page, thanks for the support with TerL purifications, your help was greatly appreciated. To Emily Agnello, Jacob Landeck, and Xingchen Liu, thanks for joining the lab, and I know you all will do great!

To the members of the Schiffer and Royer labs, your advice in group meeting has been vital for moving my projects forward. Thanks for your feedback, enthusiasm, encouragement, reagents, and equipment. I'd like to especially acknowledge Ellen Nalivaika, who was my go-to person for many questions over the years, and Ashley Matthew, who has been a great friend and cheerleader through all of grad school.

Of course, I also have to acknowledge my committees, both thesis and dissertation. Dr. William Royer, thank you for chairing all of my committees, from qualifying exam, to TRAC, and now DEC. Your help in moving forward through grad school has been tremendous. Dr. Celia Schiffer, thank you for all of your insight over the years, and for co-sponsoring my F31. I fully believe your additions to my fellowship made all of the difference in getting funded. Dr. Daniel Bolon, thank you for your guidance with the linked-construct, I learned so much from you contributions to that project. Dr. Osman Bilsel, thank you for serving on my committee for almost 5 years, I was very sad to see you go, but am grateful for all your help. To my new DEC committee members Dr. Andrei Korostelev and Dr. Paul Jardine from the University of Minnesota, I'm grateful for your service on my committee, and for your input on my dissertation.

In the BMP department there are too many to thank individually. In my opinion, we are the greatest department and I am very happy with my choice to join. Specifically, I would like to thank the Ryder and Bolon labs for allowing us to use their plate readers. The ATPase experiments in this work would never have

been possible without your help. To the Munson lab: thank you for letting me use your workstation to process EM data, it really sped things along. To the Matthews lab, thank you for letting me perform the CD experiments in your lab and for sharing reagents when needed.

To both EM platforms I owe a lot of gratitude. In EM core, Dr. Greg Hendricks and Dr. Lara Strittmatter taught me how to prepare grids and use the microscopes in the core. At the cryoEM facility, Dr. Chen Xu provided enormous amounts of support through training, project advice, and strategy, all of which were valuable contributions to the work performed in Chapter III. Dr. Kangkang Song and Dr. Kyounghwan Lee were also extremely helpful with sample preparation, screening, and data collection, all while teaching me the ins and outs of data collection. I thank beamline scientists at ALS sector 5 (Lawrence Berkeley National Laboratory), namely Dr. Banumathi Sankaran, APS 23-ID-B (Argonne National Laboratory), and NSLS X25 (Brookhaven National Laboratory) for technical support with X-ray diffraction data collection.

I would also like to acknowledge our collaborators at the University of Maryland in Dr. Lindsey Black's lab. Dr. Bingxue Yan and Dr. Black provided the T4 procasid and terminase reagents, along with advice on the *in vitro* DNA packaging project described in Appendix 2.

To all of the those in the BMP office: thanks for always taking care of the little things we take for granted every day. Special thanks for Karen Welch for always being there to chat, and for keeping the candy jar full! Additionally, I'm

grateful for the support from Luca, Maria, Irene, and Ivery, who kept the labs running through my entire time in graduate school.

I'd also like to acknowledge my mom, dad, sister, and extended family. Your support and encouragement means the world. Last but not least, I also thank my fiancé, Dr. Noah Cohen. I could not have done this without you. It's been a long road, but I'm glad to have met you along the way.

Abstract

Many double-stranded DNA viruses use a packaging motor during maturation to recognize and transport genetic material into the capsid. In terminase motors, the TerS complex recognizes DNA, while the TerL motor packages the DNA into the capsid shell. Although there are several models for DNA recognition and translocation, how the motor components assemble and power DNA translocation is unknown.

Using the thermophilic P74-26 bacteriophage model system, we discover that TerL uses a *trans*-activated ATP hydrolysis mechanism. Additionally, we identify critical residues for TerL ATP hydrolysis and DNA binding. With a combination of x-ray crystallography, SAXS, and molecular docking, we build a structural model for TerL pentamer assembly. Apo and ATP analog-bound TerL ATPase domain crystal structures show ligand-dependent conformational changes, which we propose power DNA translocation. Together, we assimilate these findings to build models for both motor assembly and DNA translocation.

Additionally, with the P76-26 system, we identify the TerS protein as gp83. I find that P74-26 TerS is a nonameric ring that stimulates TerL ATPase activity while inhibiting TerL nuclease activity. Using cryoEM, I solve 3.8 Å and 4.8 Å resolution symmetric and asymmetric reconstructions of the TerS ring. I observe in P74-26 TerS, the conserved C-terminal beta-barrel is absent, and instead the region is flexible or unstructured. Furthermore, the helix-turn-helix motifs of P74-

26 TerS are positioned differently than those of known TerS structures, suggesting P74-26 uses an alternative mechanism to recognize DNA.

Table of Contents

List of tables

List of figures

List of copyrighted materials produced by author

Figure number	Publisher	License number
Figure 1.8	National Academy of Sciences	None required
Chapter II	National Academy of Sciences	None required
Appendix III	Oxford University Press	4643680444438

List of third party copyrighted material

Figure number	Publisher	License number
Figure 1.2	Elsevier	4643680676512
Figure 1.5	Elsevier	4643680814405
Figure 1.7	Elsevier	4643680989720
Figure 1.9	Elsevier	4643681217210

Abbreviations

ASCE	Additional strand conserved glutamate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BME	2-mercaptoethanol
CryoEM	Cryo-electron microscopy
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ELP	Empty large prohead
ESP	Empty small prohead
HOC	Highly immunogenic outer capsid protein
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
MBP	Maltose binding protein
NADH	Nicotinamide adenine dinucleotide hydrogen
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
PP	Prescission protease
SAD	Single-wavelength anomalous dispersion
SAXS	Small angle x-ray scattering
SDS	Sodium dodecyl sulfate
SEC-MALS	Size-exclusion multi-angle light scattering
TAE	Tris-acetate-EDTA
TerL	Large terminase
TerS	Small terminase
WT	Wild-type

Preface

Portions of the work in this dissertation were performed or assisted by other individuals. Segments of the introduction will be published as a review chapter in the 4th Edition of the Encyclopedia of Virology (Elsevier). This review was written primarily by myself, with input and suggestions from my advisor, Dr. Brian Kelch.

The work in Chapter II is a co-first author publication with our previous postdoctoral researcher, Dr. Brendan Hilbert. In this chapter, I optimized the crystallization conditions for the apo TerL ATPase domain in the I23 space group. Dr. Hilbert performed the phasing, model building, and refinement for all structures, as well as the crystallization for the SeMet, ADP•BeF_3, R139A variant, and $P3_221$ apo space group structures. I cloned, expressed, and purified all wildtype and TerL variants, with the exception of the R39A, R102A, and R139A/E150A double variant, which were cloned by Dr. Caroline Duffy. I also performed all ATPase and DNA binding assays, with the exception of those in Figures 2.2D and E, which were done by Dr. Nicholas Stone. Synchrotron data collection was done at ALS sector 5 (Lawrence Berkeley National Laboratory) for SetMet and I23 space group apo crystals by Dr. Banumathi Sankaran. Dr. Hilbert and I made samples for SAXS, and the data was analyzed by Dr. Kelch. Dr. Kelch also performed the molecular docking and modeling in the manuscript. The manuscript was written by Dr. Kelch, Dr. Hilbert, and myself as a joint effort. All

data analysis in the publication are the result of numerous conversations between the three of us.

The work in Chapter III was performed largely by myself, under the advising of Dr. Kelch. I created and optimized the TerS expression and purification protocols, and performed the SEC-MALS, ATPase, and DNA binding assays. Dr. Hilbert performed the nuclease assay, and Dr. Stone performed the negative stain and subsequent 2D classification of the TerS particles. I optimized the cryoEM sample preparations, data processing, model building, and refinement. When needed, I sought advice from other lab members such as Dr. Stone and Dr. Christl Gaubitz. Datasets were collected with the help of Dr. Chen Xu, Dr. Kangkang Song, and Dr. Kyounghwan Lee in the cryoEM core facility. Circular dichroism studies were assisted by Noah Cohen. All analysis, discussion, and writing was performed by myself with insight and suggestions from Dr. Kelch.

The discussion section was written by myself, with limited input from my advisor. All work performed in Appendix 1 was undertaken by myself, with the exception of the E150A/WT doping experiment, which was done by Dr. Stone. In Appendix 2, this work was done in collaboration with Dr. Lindsey Black's lab at the University of Maryland. Dr. Bingxue Yan from the Black lab purified and supplied us with the T4 proheads and TerL samples. Dr. Black and Dr. Yan also provided helpful insight for *in vitro* packaging reaction optimization. All experiments in Appendix 2 were carried out by myself. In appendix 3, my

contribution to this publication is portions of the DNA binding data, which was recapitulated by Dr. Hilbert. Additionally, I also played a role in optimizing the nuclease assay that is featured throughout the publication.

CHAPTER I

Introduction

Introduction

Bacteriophage are the world's most abundant entities, with studies estimating their numbers to reach over 10^{31} particles on Earth[1]. In order for their populations to achieve this astounding level of prevalence, bacteriophage infect bacteria and replicate at rates which produce dozens of phage in minutes[2]. Because of their rapid reproduction speed and overwhelming numbers, bacteriophage are used as model systems to study viral replication and maturation. From them, we have discovered the fundamentals of biology, such as DNA as hereditary material[3], the genetic code[4], and mRNA[5].

One task all viruses must accomplish is transporting their genome into a capsid to create infectious progeny. While viruses with small genomes use an enveloping process led primarily by capsid proteins[6,7], viruses with long genomes are not capable of this mechanism. For this task, these viruses have developed ingenious DNA recognition systems and powerful motors. Motors are found in viruses infecting organisms from all three domains of life. This includes viruses as large as the giant mimiviruses to small viruses such as adeno-associated viruses[8–10]. For many of these motors, the molecular mechanisms and even the identities of motor components are unknown. However, in all cases, these viruses use an ATPase motor to actively pump RNA or DNA into the small volume of the viral particle, strictly confining the genome. In this manner, the motor overcomes the entropic and enthalpic penalties for encapsulating a large amount of genetic material.

In many motor-driven viruses, the genome is packaged to very high density, often approaching a quasi-crystalline state[11]. For example, the ~48 kb genome of phage lambda is ~16 μm end-to-end, but is packaged into a head with a diameter of only ~60 nm[12]. The motor needs to be extremely powerful in order to pump the genome against the internal pressure that builds during packaging, and also highly regulated to ensure complete and efficient encapsulation. Because of their ability to work against high pressure, viral genome packaging motors are a particularly exciting model for understanding motor force-generation and regulation. In the following chapters, my work explores the genome recognition and translocation mechanisms of double-stranded DNA viral motors.

In this introduction, I will focus on the best-understood classes of viral genome packaging motors: the terminases and phi29-type motors. These motors catalyze double-stranded DNA packaging in tailed bacteriophage, as well as herpesviruses. Because phage are the most numerous biological entities, these motors likely perform the bulk of viral genome packaging on Earth[13,14]. Figure 1.1 depicts a typical packaging cycle for a terminase motor with discrete DNA recognition, motor assembly, DNA translocation, DNA cleavage, and final maturation stages. While my work mainly focuses on terminase motors, the phi29 system provides important insight into the chemomechanical steps of the

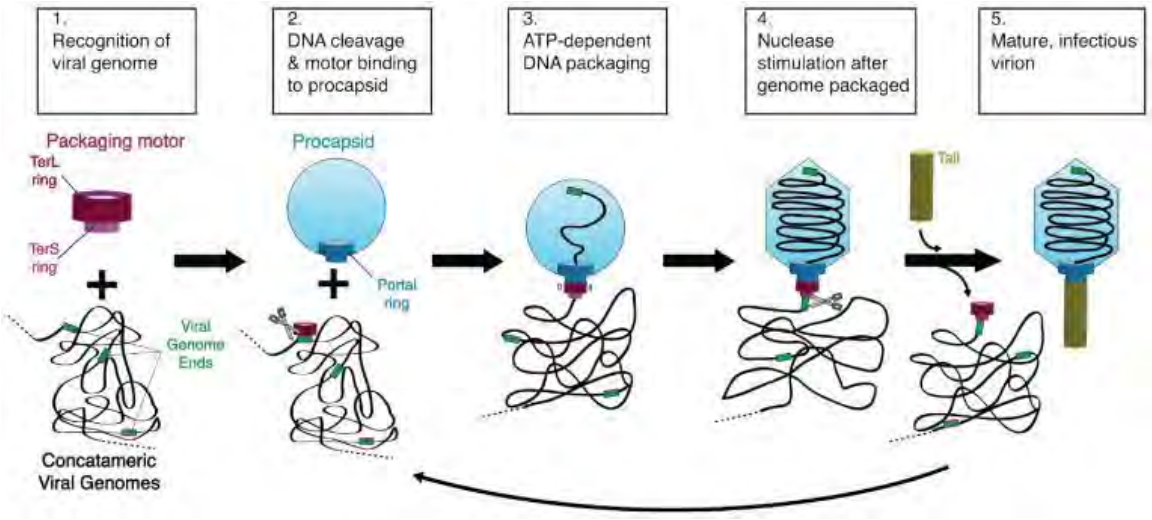


Figure 1.1. Schematic of a terminase motor packaging cycle. Genome packaging in bacteriophage and TerL constructs. Maturation of dsDNA viruses is a five-step process: (1) the small terminase (TerS) subunit of the motor recognizes the concatemeric viral genome; (2) the motor binds to the portal complex on the capsid; (3) TerL hydrolyzes ATP to translocate DNA into the capsid; (4) after the genome is translocated into the capsid, TerL switches its enzymatic activity from translocation to DNA cleavage; and (5) the motor is finally released for maturation of another virus while portal binds to the tail proteins to complete a mature, infectious virion. This figure and legend is reprinted with permission from: Hilbert et al Nucleic Acids Res 2017;45(6):3591-3605.

packaging reaction, and thus is included here. In the following sections, I describe the major components of the packaging motor, as well as their role in DNA encapsulation.

The Portal

Both terminase and phi29-type motors pump DNA through a dodecameric ring complex embedded in one of the five-fold vertices of the icosahedral shell. In terminase motors, this complex is called the 'portal', whereas it is called 'connector' in phi29-type motors. Not only does DNA enter through the portal ring during packaging, but it exits through the same channel. The portal also serves as the binding site for the neck and tail proteins that assemble upon maturation of the virus particle[15–17]. Numerous crystal structures of portal complexes from diverse phages reveal that these proteins share a common core structure[18–23]. This indicates that all portals are evolutionarily related, despite low sequence homology in their primary sequence. The portal ring has a central channel just wide enough for double-stranded DNA to pass through. This finding has led to the theory that the portal strips the incoming genome of any proteins bound to the DNA[24]. Moreover, this tight channel is suggested to play a critical role in the packaging process itself, perhaps by acting as a one-way valve[25,26] or even as a primary component of the force-generation step[27,28]. In all examined cases, the dodecameric portal complex creates a symmetry mismatch with the capsid by replacing a pentameric coat protein complex at the five-fold vertex of the

icosahedron. Whether this symmetry mismatch plays a functional role in packaging remains unknown, although recent work suggests a possible role in genome packaging termination[19].

The Motor: An ASCE ATPase

Phi29 motors primarily consist of the connector assembly, ATPase motor, and a specialized prohead RNA molecule, known as ‘pRNA ’(Figure 1.2). During packaging, the connector protein binds to the ATPase motor through interactions with the pRNA, and the motor translocates the DNA through the connector into the capsid. In terminase motors, the portal assembly binds to the ATPase subunit (the large terminase, or TerL) directly without pRNA. Terminase motors have an additional endonuclease domain, which is used to initiate and terminate packaging (discussed below). Despite the differences in motor assembly, both phi29 and terminase motors rely on a similar catalytic core to power DNA translocation: the ASCE ATPase fold.

The ATPase subdomain of viral genome packaging motors is derived from the ASCE (additional strand conserved glutamate) division of ATPases. The ASCE division contains numerous superfamilies of ATPases, including widely studied AAA+ ATPases, ABC transporters, RecA, and SF1/SF2 helicases (Figure 1.3)[29]. ASCE ATPases are ancient machines that function across all cellular life,

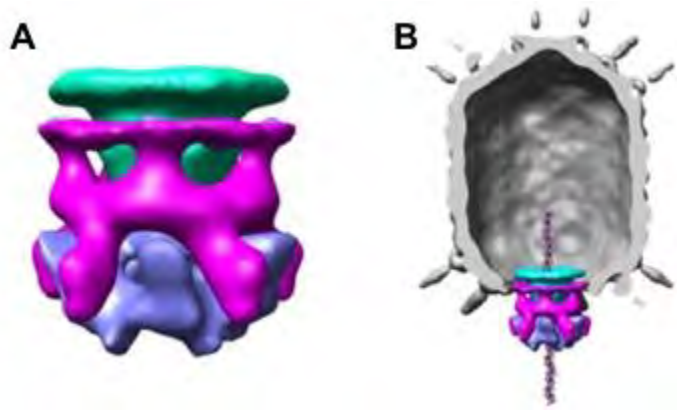


Figure 1.2. CryoEM structure of the phi29 motor. Density from difference maps isolating the connector (green), pRNA (magenta) and ATPase (blue) were combined to visualize the DNA packaging motor in ϕ29. (A) Close-up view of the phi29 motor. (B) The motor is shown in the context of the entire prohead. The front half of the prohead density has been removed so all the motor components can be seen. This figure and legend has been adapted with permission from: Morais et al Structure 2008;16(8):1267-74.

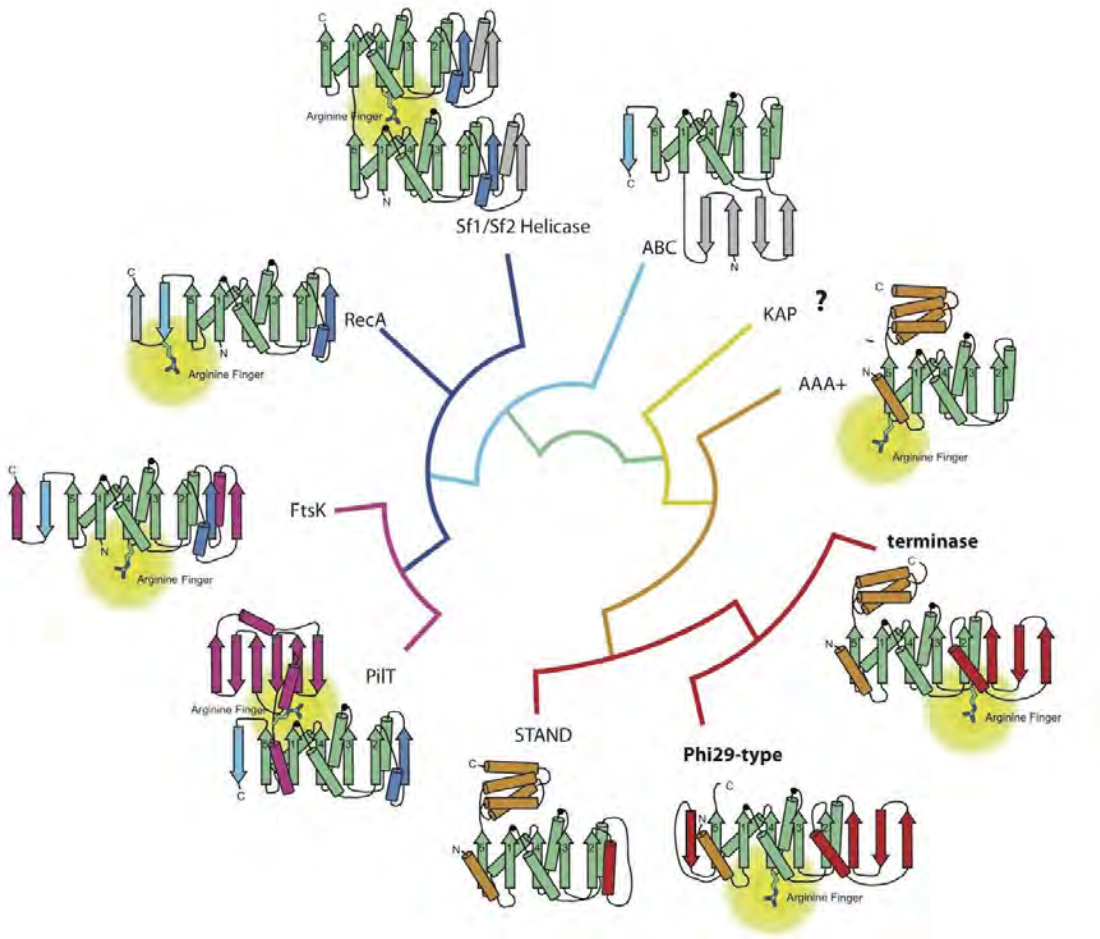


Figure 1.3. Phylogenetic tree of ASCE ATPases. The core Rossmann fold is shown in green, while branch-specific structural features are shown in various colors. The terminases and phi29-type ATPase families likely evolved from AAA+ ATPases.

as well as many viruses. Although their molecular mechanisms and roles are remarkably diverse, ASCE ATPases usually function as multimers that convert the chemical energy of ATP into mechanical work. ASCE ATPases also generally use a similar ATP hydrolysis mechanism. Thus, studies of the packaging motor structure and mechanism have broad relevance towards understanding the general principles of molecular machines.

The ASCE nucleotide binding motif uses a Rossmann 'βαβ' fold[30,31] which is comprised of a β-sheet sandwiched between multiple α-helices. Different clades within the ASCE tree may add variable elements to this base structure, however there are four features that are common to all ASCE ATPases: the Walker A and B motifs, a catalytic base that is usually a glutamate residue, and a *trans*-acting motif (Figure 1.4). The Walker A motif (also known as the P-loop; consensus sequence [G/A]xxxxGK[T/S], where 'x' indicates any amino acid), is a nucleotide binding motif that forms hydrogen bonds with the ATP phosphates in the active site. The threonine/serine residue in the Walker A motif is critical, as it coordinates a metal ion, often a magnesium, for ATP hydrolysis. Downstream in the protein sequence, the Walker B motif (consensus sequence φφφφ[D/E], where 'φ' indicates a hydrophobic amino acid), further assists in binding this metal, which coordinates the β and γ phosphates of ATP. Immediately downstream of the Walker B motif, a glutamate residue acts as an essential catalytic base, activating water for nucleophilic attack on the γ-phosphate of

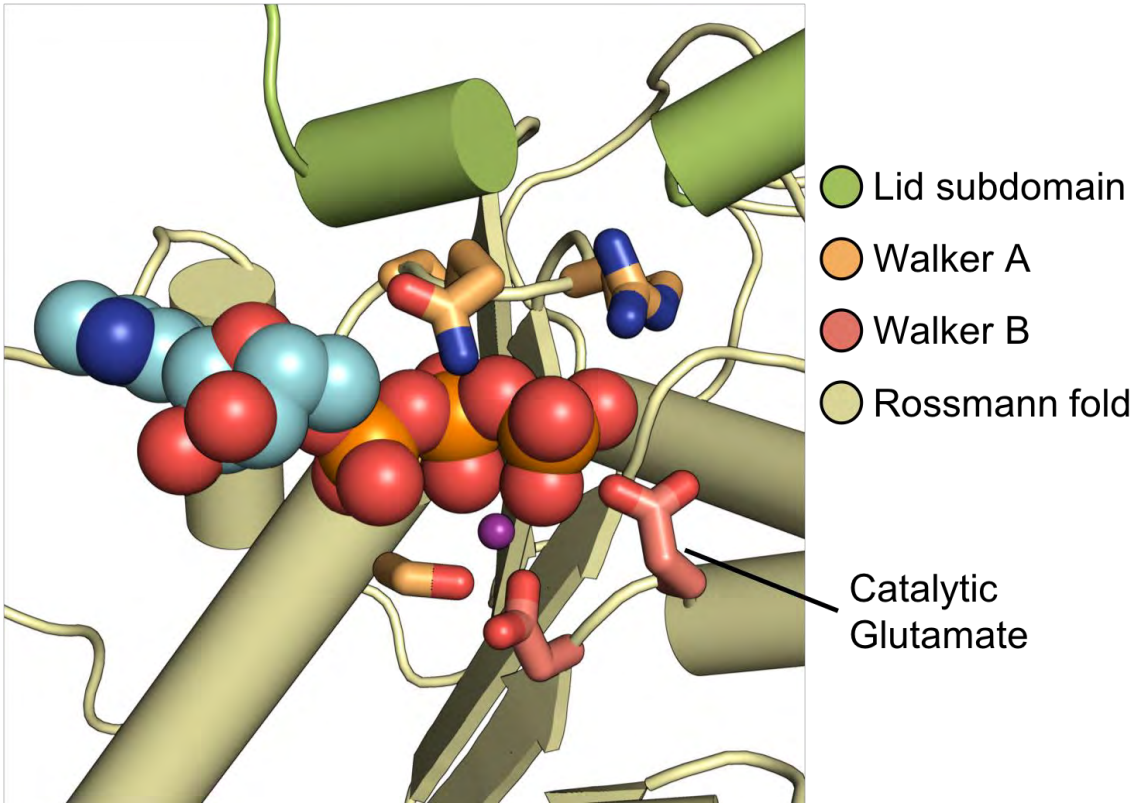


Figure 1.4. The terminase ATPase active site. The P74-26 bacteriophage ATPase domain is shown. ADP•BeF_3, a non-hydrolyzable ATP analog, is bound in the active site. The Walker A motif hydrogen bonds with the ATP phosphates, and the conserved serine residue coordinates a magnesium ion (purple sphere). The Walker B motif further assists in binding the magnesium, with the catalytic glutamate coordinating a water (not shown) for nucleophilic attack.

ATP[29,32,33]. (Often this ‘catalytic glutamate’ is included in the Walker B motif, which yields a consensus sequence φφφφ[D/E]E.) The fourth feature, a positively charged *trans*-acting element from a neighboring subunit (known as the arginine finger) interacts with the γ phosphate of ATP to stabilize the hydrolysis transition state. This *trans*-acting element also links nucleotide sensing and ATP hydrolysis between subunits, a necessity for regulating activity within oligomeric ATPases.

Terminase and phi29-like motors have numerous similar features, suggesting a relatively close phylogenetic relationship. Previous phylogenetic analysis classified the terminases and the phi29-type motors as disparate families within the ASCE division of ATPases. The phi29-type motors were placed within the HerA/FtsK family of dsDNA translocases and the terminases as an offshoot of RecA ATPases[34,35]. This classification was based purely on sequence analysis and secondary structure prediction. However, recent structural and mechanistic discoveries led to the conclusion that terminase and phi29-type motors are much more closely related and are two distinct clades within the same family (Figure 1.3)[36–40]. The ATPase subunits of phi29 and terminase motors have a core Rossmann fold with additional elements appended. For example, there is an additional α-helix and three antiparallel β-strands inserted between strand 2 and helix B of the core Rossmann fold, differentiating them from other ASCE ATPases[36–41]. An N-terminal helix has also been added, where it packs atop helices A, B, and C of the core Rossmann fold. The phi29 ATPase domain also contains a short antiparallel β-strand inserted

between helix D and β-strand 5[36]. Additionally, terminase motors have a C-terminal nuclease domain, while phi29 motor proteins lack nuclease activity. This is not surprising, because the phi29 motor does not cleave DNA during packaging (see below for more details). Finally, the terminase motor has a three-helix bundle subdomain in between the core ATPase domain and the nuclease domain. We refer to this subdomain as the 'lid' because it caps the ATPase active site (others have referred to this substructure as a 'linker' domain[39] or ATPase subdomain II[41]). The three-helix bundle lid subdomain is distinguished from the four-helix bundle lid subdomains of AAA+ and STAND ATPases. It is unclear whether phi29 has a similar lid subdomain, as the protein used for crystallization only contained the core Rossmann-fold of the ATPase[36].

In addition to similar ATPase domain structure, both terminase and phi29 proteins form oligomers when bound to the portal/connector. In the absence of the portal/connector proteins, the subunits of terminase and phi29 motors are generally monomeric[36,39,40,42,43]. CryoEM and biochemical studies of the terminase motor show a pentameric assembly of subunits[37,41]. However the oligomeric state of the phi29 motor remains contentious, as some biochemical studies suggest a hexameric assembly[44].

Assembly of the oligomeric motor onto the portal/connector facilitates ATPase and presumably packaging activity. Both terminase and phi29 motors use a *trans*-activated ATPase mechanism through an arginine finger motif[36,37,45], similar to most other ASCE ATPases[46,47].The arginine finger is positioned within

the interface between two subunits, directly contacting the ATP γ-phosphate to stabilize the developing negative charge during the ATP hydrolysis transition state. In phi29, the arginine finger also causes ejection of ADP from the post-hydrolysis state[48]. Thus, the arginine finger residue not only plays a catalytic role, but also 'communicates' nucleotide status between subunits within the ring. Because the arginine finger is central to motor function, its position on the protein surface dictates the relative orientation of neighboring subunits within the ring. In terminase and phi29 motors, the arginine finger is located in different regions of the protein, suggesting that the overall orientation of ATPase subunits may be altered between the two classes.

Phi29 and terminase motors are further distinguished by the accessory subunits required for efficient packaging. Terminase motors use the small terminase protein (TerS), which specifically recognizes the viral genome and modulates large terminase ATPase and nuclease activity (see 'Terminases: Packaging Initiation' section below)[42,43,49,50]. Phi29 lacks a clear homolog of TerS, and instead uses the gp3 protein which is covalently bound to the 5' ends of the viral genome[51]. The gp3-DNA complex is suspected to be the recognition element for packaging initiation[52,53]. Additionally, phi29 uses the 174 base pair long pRNA for binding the motor to the connector. The pRNA forms a pentameric assembly on the connector, binding to the phi29 motor through a spoke-like RNA secondary structure[23,36,54]. Terminase motors do not require a pRNA to bind portal, and bind through direct protein-protein interactions[41].

Phi29 motors

The viral genome packaging machinery in phage phi29 is arguably the best understood system. The phi29 packaging motor was discovered in the 1970s, and an *in vitro* packaging system was established in 1986[55]. The ability to reconstitute genome packaging from purified components opened the door for years of incisive study. Furthermore, the phi29 motor chemomechanical cycle has been elucidated to high detail using an elegant single-molecule optical trapping assay for motor activity[56]. Early optical trap experiments established phi29 as an exceptionally powerful motor, working against forces averaging 57 pN[56], a force close to when dsDNA fundamentally alters its conformation through over-stretching[57]. Subsequent experiments identified individual subunit stepping, which provided incredible detail into the mechanisms of force generation[58]. Finally, recent structural analysis has visualized a functioning holoenzyme at low resolution and various individual components at high resolution[36].

Phi29 genome replication and packaging begins with the gp3 protein. At the beginning of viral genome replication, gp3 covalently binds to the 3' end of the genome, where it acts as a primer for replicating the 5' strand[59]. The gp3 priming domain dimensions and charge mimic DNA, allowing the polymerase to initiate replication by covalently linking a nucleotide to the hydroxyl of a serine within gp3[59]. After DNA replication is complete, both ends of the DNA, known as left and right, remain bound to gp3. DNA-gp3 forms lariat loops[60], with the left end junction preferentially interacting with the packaging motor bound to the

prohead[61,62]. Binding to the motor causes DNA to supercoil, with some evidence suggesting that the supercoil wraps around the connector protein outside of the capsid[63]. In an unknown mechanism, the motor begins packaging the gp3-DNA, possibly through packaging the loop in its entirety, or through gyrase activity that has been speculated but never directly observed, allowing the loop to resolve.

The translocation of the genome into the empty phi29 capsid occurs in a complicated, ATP-dependent mechanism. Each translocation cycle can be separated into two phases: a ‘burst’ phase in which DNA physically pushes into the capsid, and a ‘dwell’ phase in which the motor resets (Figure 1.5). At the beginning of the ATP hydrolysis cycle, each subunit of the phi29 pentameric motor binds ATP[64]. Once the ring is fully loaded with ATP, ATP hydrolysis is stimulated in a sequential order around the ring[48,58,64]. During the burst phase, DNA is translocated by a total of 10 bp[58]. By slowing the rate of translocation, Bustamante and colleagues showed that the 10bp burst actually consists of four discrete 2.5bp substeps[58]. Thus, the pentameric phi29 motor uses four translocation events per cycle, which suggests that not every subunit directly translocates DNA per cycle.

One of the five motor subunits hydrolyzes ATP, but does not translocate DNA, and is designated the ‘special’ subunit. The exact role of the special subunit is still unclear; however, it is critical for regulating the motor during the dwell phase, which occurs in between each of the 10 bp translocation increments. It is thought that the special subunit grips DNA during the dwell

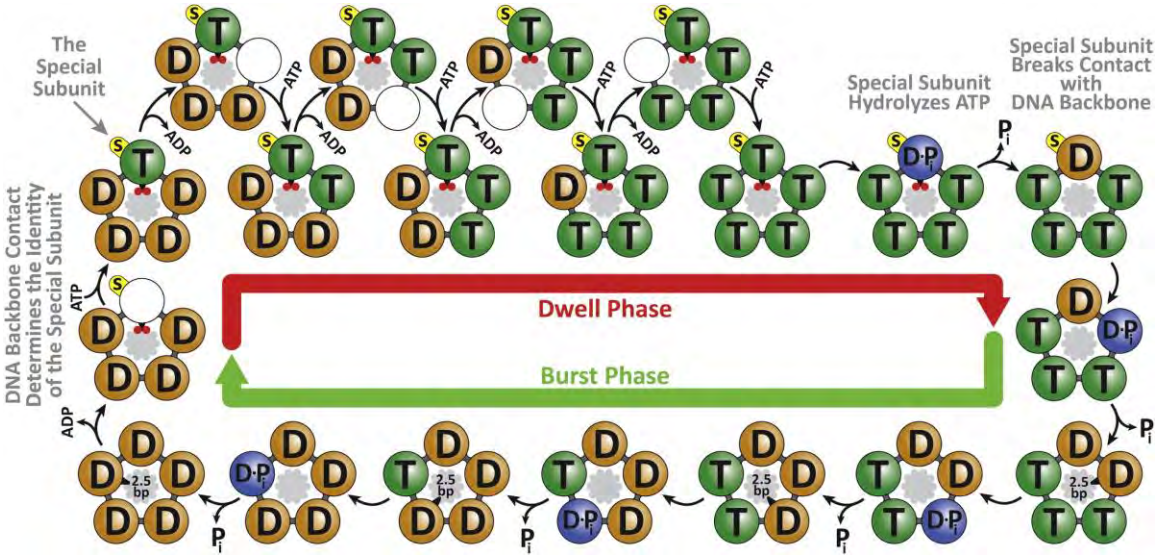


Figure 1.5. The phi29 chemomechanical ATPase cycle. At the end of the burst, all subunits are ADP-bound ("D" label). At the beginning of the dwell, the motor makes an electrostatic contact with two backbone phosphates (small red circles) on the dsDNA substrate (inside the ATPase ring). This unique contact determines the identity of the special subunit (yellow label "s"). The formation of the electrostatic contact facilitates ADP release by the special subunit. Subsequent ATP ("T" label) binding and ADP release events are interlaced, with ATP binding to one subunit enabling ADP release from its neighbor. After all five subunits have bound ATP, the special subunit hydrolyzes ATP ("D•Pi" label), releases P_i, and uses the hydrolysis free-energy to break the electrostatic contact with DNA, triggering the burst phase. During the burst, the remaining four ATP-bound subunits sequentially hydrolyze ATP, release Pi, and translocate DNA by 2.5 bp. The motor-DNA geometry (10.0 bp burst size versus 10.4 bp dsDNA helical pitch) favors a mechanism in which the same subunit is special in consecutive cycles. This figure and legend is reprinted with permission from: Chistol et al Cell 2012;151(5):1017-28.

phase, and releases ADP from its active site, allowing ATP to bind. The arginine finger of the special subunit and each subsequent subunit triggers ADP exchange for ATP sequentially within the ring, resulting in a fully ATP-loaded motor[48,58]. The DNA-bound special subunit then initiates ATP hydrolysis, breaking its contact with DNA, and triggering the burst translocation phase. During the four subsequent ATP hydrolysis events, ATP is hydrolyzed and P_i released. P_i release induces the conformational change that drives the translocation step, packaging 2.5 bp of DNA before handing it to a neighboring subunit. Upon the power stroke conformational change, the arginine finger of the currently active subunit positions into the neighboring subunit's active site, priming the next subunit for an ATP hydrolysis event. In this way, hydrolysis in the next subunit is triggered, and the DNA is processively packaged[48].

During the various packaging steps, DNA needs to undergo multiple 'grip-and-release' cycles. DNA is held by each subunit tightly during the translocation sub-steps as well as during the dwell; however, each subunit must relinquish its hold on DNA to easily pass off the DNA to a neighboring subunit within the ring. DNA is gripped mostly through the 5'-3' strand, with less contact made with the 3'-5' strand and minor contacts with sugars and nucleotide bases[65]. Upon passing to the next subunit during translocation, the DNA undergoes rotation[66]. As sequential hydrolysis and DNA translocation occurs during the packaging reaction, translocation step size and DNA rotation per cycle starts to vary. As more DNA is encapsulated, the step decreases from 2.5 to 2.3 bp, and the

rotation of DNA increases from -14° to -48° per increment. This is in accord with increased packaging density within the capsid and subsequent elevated backpressure, which the motor must work harder against in order to continue the packaging reaction. Recent work has shown that this backpressure can be relieved by a temporary pause in packaging, which allows the DNA to relax in the capsid[67]. Because this relaxation time is far longer than the packaging time, this result suggests that the fast speed of packaging causes DNA to adopt a metastable conformation. These findings illustrate that that motor speed has an important role in determining capsid internal pressure and genome dynamics.

The structure of the phi29 motor is beginning to come into focus. The connector assembly structure was first determined to high resolution by x-ray crystallography, along with a low-resolution cryoEM structure of the motor actively packaging[23]. Subsequent cryoEM studies show that the portal undergoes large conformational changes in structure upon packaging[68] and suggest a pentameric arrangement of the pRNA and ATPase components of the motor[69]. The pentameric arrangement has been challenged, with the more commonly seen hexameric ATPase arrangement proposed[44,45]. However, asymmetric reconstructions of the phi29 motor structure exhibit a clear five-fold or pseudo-five-fold arrangement of subunits[36,48,54]. Over the past decade, high-resolution structures of most of the motor components have been determined: a majority of the pRNA structure was determined by crystallography and NMR[70,71], as well as the ATPase domain of the gp16 motor protein[36]. By docking the high-resolution

structures into the cryoEM density, we are nearing a complete picture of the phi29 motor[36]. High-resolution structures of the motor in action are still needed to understand the motor's mechanism in detail. However, the rapidly developing field of cryoEM is nearly certain to profoundly impact our understanding of the phi29 motor in the future.

Terminase motors

In comparison to phi29 motors, less is known about terminase motor mechanism. Because terminases motors have been studied from multiple phage, here I synthesize the results from many model systems, such as phages T4, lambda, SPP1, P22, Sf6, and P74-26.

Packaging Initiation

Terminases initiate genome packaging using a small terminase protein. The small terminase (TerS) binds to the viral genome[50,72–77] and modifies the enzymatic activities of the large terminase protein[42,43,49,50,78]. During the initial stages of DNA packaging, TerS recognizes the concatemeric DNA and presents it to TerL, which cuts the DNA to form the termini that is inserted into the capsid[79]. In unit-length genome packaging phage with cohesive ends, this cleavage sequence is known as '*cos*', but in headful packaging terminases, it is termed '*pac*' (see 'Termination' section below for description on unit-length and headful packaging bacteriophage). *Cos* sequences are specifically cleaved by

the terminase to form complementary ends that ligate and circularize the DNA[80]. *Pac* sequences, on the other hand, are more variable, and cuts occur loosely within a general region of the genome[80]. In some phages, such as P22, this region is quite short (around 22 bases)[81], whereas in phages such as Sf6, the *pac* site is up to ~1800 bases long[82].

DNA recognition occurs in the region of the *cos* or *pac* signal. In lambda phage, a *cos* site bacteriophage, the DNA recognition component (gpNu1) binds to a site termed *cosB*, which is downstream of the *cosN* cleavage site [83]. In SPP1 phage, a *pac* site-containing bacteriophage, the TerS protein has been shown to bind two flanking areas of the cleavage site within the *pac* sequence [77]. The *pac* region in the SPP1 genome is intrinsically bent[77], which may serve as a feature for DNA recognition[84].

Structural studies of TerS proteins have furthered models for binding to *pac* and *cos* sites. In lambda phage, a NMR structure of the gpNu1 N-terminal region presents a dimer with externally facing winged helix-turn-helix motifs, a common DNA binding motif[73]. This was unexpected, as it was unclear how opposite-facing motifs would bind DNA. However, with the inclusion of the DNA-bending host-factor protein IHF, a model was developed in which the lambda *cos* site bends into a hairpin, bringing two regions within the *cosB* site close together, promoting gpNu1 binding[73]. Further studies of *pac* site TerS proteins show more complicated assemblies[49,50,85–87] (Figure 1.6). These complexes are made of 8 to 11 subunits, all of which form rings with central pores. From first impression, it

appears that DNA threads through the center of the TerS pore into the TerL motor. However, similar to lambda phage, the N-terminal region of most of these proteins forms a DNA binding motif (in this case a helix-turn-helix) that faces externally around the ring. In *Shigella flexneri* phage Sf6, mutations in this putative DNA binding domain abrogate or weaken DNA affinity[85]. Additionally, as is the case with *Shigella* phage Sf6 and *Bacillus subtilis* phage SF6, the pore of the TerS ring is too narrow for DNA binding. These findings, paired with the results that certain *pac* sequences are intrinsically bent, suggest DNA wraps around TerS, similar to the way nucleosomes interact with DNA.

While the wrapping mechanism is the current preferred model for TerS DNA binding, there is evidence that it may not be universal. For example, the P22 TerS assembly does not exhibit external helix-turn-helix motifs around the perimeter of the ring[50]. In fact, the P22 protein is folded differently from other TerS proteins. In P22 TerS, the N-terminal and central regions of the sequence form the ring appendages, whereas in other systems the N-terminal region alone forms this domain[50]. Additionally, the pore of the P22 ring is wide enough to accommodate DNA binding, and therefore is hypothesized to grip DNA[50]. While it is possible that P22 has evolved divergent mechanisms for binding DNA, it is unclear why the overall morphology of the protein is conserved. Thus, there is much to learn about the structural features of TerS proteins.

Beyond the N-terminal region, TerS assemblies generally have a central domain, known as an oligomerization domain, and a C-terminal β-barrel domain

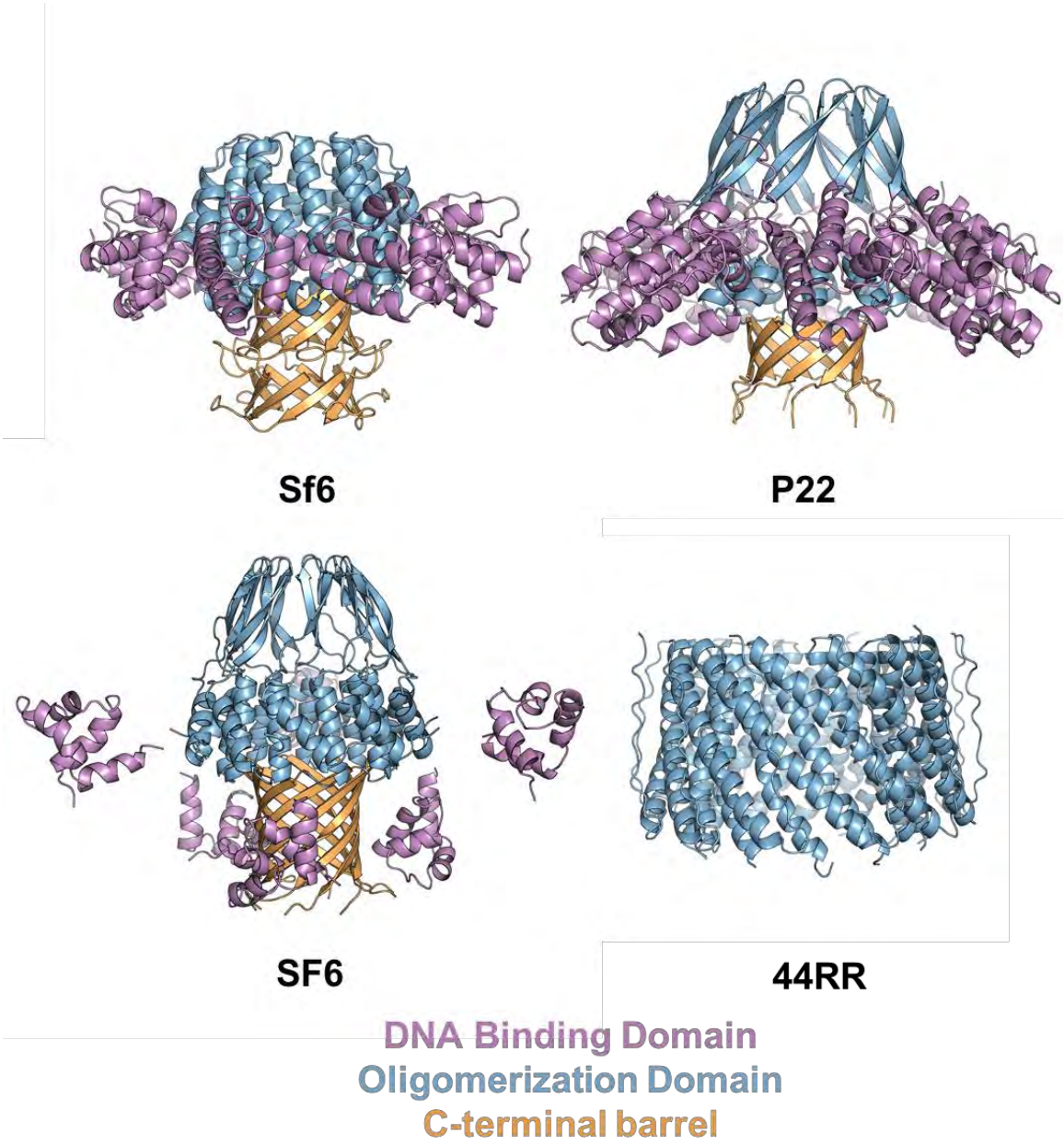


Figure 1.6. X-ray crystal structures of TerS assemblies. The crystal structures of Sf6, P22, SF6, and 44RR (PBD codes: 3HEF, 3P9A, 3ZQQ, and 3TXS). In general, TerS proteins have a putative DNA binding domain, also known as a helix-turn-helix motif (purple), a central oligomerization (blue), and a C-terminal barrel (orange). The 44RR TerS structure is of an N and C terminal truncation construct.

(Figure 1.6). The oligomerization domain forms the pore of the TerS assembly, and as the namesake suggests, contacts neighboring subunits to join the ring together. In most phage, this region is α-helical, with some systems exhibiting an anti-parallel β-hairpin insertion within the domain[50,87]. Conversely, the role of the C-terminal β-barrel is less clear. Some studies show removing the barrel introduces flexible stoichiometry, as formation of the β-barrel may strictly constrain the oligomeric state of the ring[49,87]. Functionally, the C-terminal region of the TerS is implicated in binding and regulating large terminase activity[49,50,88–90]. It has been seen across multiple systems that TerS affects TerL through increasing ATPase activity and inhibiting nuclease activity[42,43,49,50,78,91].

Outside these limited findings, very little is known about the interaction between TerS and TerL. The interaction mode of the two proteins is unclear, and no atomic resolution structures of the co-complex exist. Additionally, DNA-binding motifs on the perimeter of the ring obscure the binding mechanism between the two complexes, as coaxial stacking of the rings is unlikely. Furthermore, not much is known about the extent of TerS' role in initiation. *In vivo*, TerS is essential for DNA packaging, but *in vitro* it is dispensable, and in some cases inhibitory[92–94]. Whether TerS plays a transient role in packaging initiation by handing DNA to TerL or if it remains bound to the motor throughout packaging is unknown.

DNA translocation

After initiation, DNA packaging commences. Terminase motors translocate DNA at a faster rate than their phi29 counterpart, with the T4 motor taking 5 to 10 minutes to package its 171 kb genome[95]. While terminases are faster translocases, they exhibit more variable speed. *In vitro*, the T4 motor translocates DNA at an average rate of 700bp/s, peaking at 2,000bp/s with an ATP turnover rate of 300/s[95]. The terminase of phage lambda peaks at ~600bp/s at low capsid density, but this rate drops to ~200bp/s towards the end of packaging[96]. Meanwhile, phi29 packaging peaks at around 165bp/s *in vitro*, making it significantly slower[97]. Therefore, the motor speed seems to be correlated with genome size: faster motors are found in phages with larger genomes such that the total packaging time is similar across multiple phages[98]. In comparison to other ASCE ATPases, the T4 large terminase moves along DNA considerably quicker than the fastest known helicase RecBCD, which unwinds DNA at a rate of ~1 kb/s[99,100]. However, it is slower than the Ftsk and SpoIIIE dsDNA translocases, which move at a rate of 5-17 kb/s and ~4 kb/s respectively[101–104].

The speed of the motor coupled with increased forces inside the capsid during packaging requires an incredibly powerful ATPase to complete the packaging reaction. Fuller et al 2007[95] estimates that the power density of the T4 motor is approximately 5,000 kW/m^3, which is twice that of a typical automobile engine. How terminases generate this tremendous power is one of the central questions that has been debated within the field for decades, and several models

for terminase mechanism have been proposed to explain this phenomenon (see below).

Terminases are generally composed of an N-terminal ATPase domain and a C-terminal nuclease domain. The two domains are connected by a short linker, allowing for some degree of flexibility between the two halves of the protein. In addition to the ASCE ATPase family features listed in the above section, the large terminase ATPase domain contains a subdomain known as the 'lid subdomain' immediately upstream of the linker. The lid subdomain forms the upper portion of the nucleotide binding pocket, and is essential for ATPase activity (Figure 1.4)[37]. In related AAA+ proteins, the lid plays an important role in propagating nucleotide-dependent conformational changes to adjacent subunits within the ring[105,106].

Although the mechanism of translocation in terminases remains unknown, many models have been proposed. Some of these models have been proven incorrect. For example, an early hypothesis that the portal ring rotates within the capsid to drive translocation[107] has been disproven by observations that the portal does not rotate during packaging or in the mature virion[108,109]. Here I will examine several of the current leading models in detail.

The "Inchworm" translocation model

One of the primary models for DNA translocation by terminases is an "inchworm" model in which DNA is pulled into the capsid via a spring-like motion

(Figure 1.7). This model was originally derived from biochemical and structural analysis of the T4 phage large terminase protein, including high-resolution crystal structures of the T4-TerL protein, and a low-resolution cryoEM structure of the T4 procapsid bound to TerL[38,41]. A modified version of the inchworm hypothesis was proposed later from high-resolution crystal structures of Sf6 phage large terminase[39]. In this model, the ATPase domains bind to the capsid and portal while the nuclease domains grip DNA in the center of the pore. In the inchworm mechanism, a key catalytic residue (Arg162) is repositioned into the active conformation from both DNA binding to the nuclease domain and ATP binding to the active site in the core ATPase domain (Figures 1.7A and B). This induces ATP hydrolysis, which in turn is predicted to rotate the lid (also known as subdomain II) by ~6° and move the short flexible linker region between the ATPase and nuclease domains by about 3 Å (Figure 1.7C). This conformational change is proposed to align a set of ion pairs between the ATPase and the nuclease domains, resulting in a stronger electrostatic attractive force between the two domains. The electrostatic interactions pull the domains towards each other, and this ‘relaxed’ to ‘tensed’ conformational change translocates 2 base pairs of DNA upward into the capsid. During the reset phase of translocation, ADP and/or P_i release from the active site reorganizes the terminase subunit back into the relaxed state via the loss of negative charges from ADP and P_i release (Figure 1.7D). ATPase subdomain II rotates back, attenuating the electrostatic force between the ATPase and nuclease domains. The nuclease

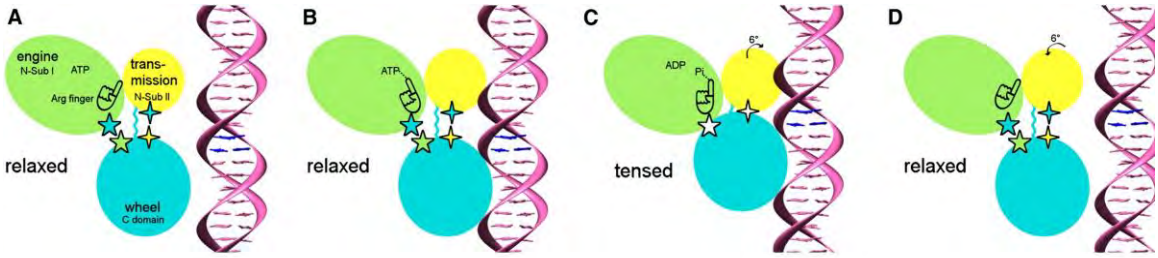


Figure 1.7. Proposed 'inchworm' model for the mechanism of DNA translocation by the large terminase. Panels (A)–(D) relate to the sequence of events that occur in a single large terminase molecule. The large terminase N-terminal subdomain I, subdomain II, and C-terminal domain are represented as green, yellow, and cyan ovals, respectively. The five-pointed stars show the charge interactions between the N-terminal subdomain I and the C-terminal domain. The four-pointed stars show the charge interaction between the N-terminal subdomain II and the C-terminal domain. The flexible linker between N- and C-terminal domain is represented by a wiggly cyan line. (A) The large terminase C-terminal domain is ready to bind DNA.(B) The C-terminal domain, when bound to the DNA, brings the DNA closer to the N-terminal domain of the same subunit. Conformational change in the N-terminal domain causes Arg162 to be placed into the ATPase active center in preparation for hydrolysis.(C) Hydrolysis of ATP has rotated the N-terminal subdomain II by about 6°, thereby aligning the charge pairs resulting in an electrostatic attraction that moves the C-terminal domain and the DNA 6.8 Å (equivalent to the distance between two base pairs) closer to the N-terminal domain and into the capsid. (D) ADP and P_i are released and the C-terminal domain returns to its original position. DNA is released and is aligned to bind the C-terminal domain of the neighboring subunit. This figure and legend is adapted with permission from: Sun et al Cell 2008;135(7):1251-62.

domain releases its grip on DNA, which is presumably passed to a neighboring subunit preceding or during the release of ADP and/or P_i.

While this model paints a comprehensive picture of the terminase packaging reaction, the proposed mechanism does not explain several observations. First, previous studies found the motor binds the portal in the reverse orientation, with the nuclease domains contacting the portal rather than the ATPase domains[110–113]. Second, with the ATPase domains splayed radially from each other and bound to portal, the *trans*-activation mechanism required for ringed ASCE ATPase activity cannot occur. Third, the regions predicted to mediate interactions with DNA show particularly low conservation and are only found in T4-like phage. Finally, the observed conformational changes do not fully explain DNA translocation. The small changes in conformation do not match a ~2 bp step size that has been predicted for terminases[114]. No significant conformational changes were observed in the structures of T4-TerL. This may be due to crystal contacts within the active site or the mutation of Walker B motif in the crystallization constructs. Moreover, there is a discrepancy in the Sf6 structures wherein the ATPγS structure, which should be locked in a 'ATP-like state' is identical to the ADP-bound and apo states[39]. This may be due to the fact that the Sf6 TerL was crystallized in the apo state, and then nucleotide ligands were soaked into the pre-assembled crystals. In many cases, the crystal lattice in the apo state does not allow for a ligand-induced conformational change in soaked crystals[115].

The “Lever” translocation model

Another model for viral motor DNA translocation is the lever model (Figure 1.8). My work with ATPase motors has contributed to developing this model, which is further described in Chapter II. In this model, a lever-like conformational change in the ATPase domain driven by ATP hydrolysis forms the power stroke of DNA translocation. In contrast to the aforementioned “inchworm” model, the “lever” model positions the ATPase domains as the central hub of the ring, while the nuclease domains protrude radially and attach to portal and capsid. Evidence of this assembly model comes from several crystal structures, molecular docking, and various biochemical analysis to determine the critical residues for motor function[37].

The identification of a *trans*-acting arginine finger was a critical step in developing the lever model. Identifying this residue led to reinterpretation of the “inchworm" structural model in which the ATPase domains do not contact neighboring subunits. The “lever” model places the ATPase domains into a ring to facilitate *trans*-activated ATP hydrolysis, placing the nuclease domains in contact with portal. Further unbiased pentamer docking experiments of a monomeric crystal structure of the P74-26 ATPase domain positions the predicted arginine fingers precisely for *trans*-activated ATP hydrolysis, providing a second layer of evidence for the inverted structural model. Interestingly, the pentameric model also fits well into the T4 packaging motor electron density map, indicating its potential biological relevance.

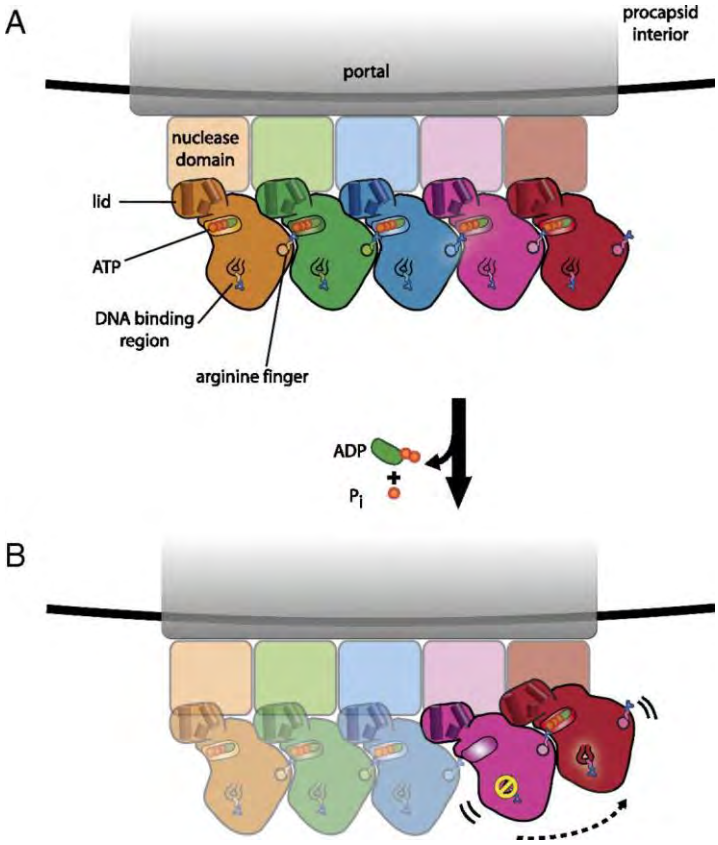


Figure 1.8. Proposed 'lever' model for the mechanism of DNA translocation by the large terminase. (A) Large terminase ring is shown. The nuclease domains interact with the portal complex (translucent gray rectangle) and possibly the procapsid (black curve). DNA is not shown but interacts with the large terminase through the DNA interaction motif. Each subunit's lid is bound tightly to the Rossmann fold of the adjacent subunit. (B) Upon ATP hydrolysis and release by the magenta subunit, the lid stays bound to the blue subunit and the Rossmann fold rotates 13° upward. To allow for this movement, the adjacent red subunit must also move in concert with the magenta subunit. To represent that the second site of symmetry breaking is unknown, the other three ATPase domains are faded. After hydrolyzing ATP, the magenta subunit releases DNA to the red subunit to translocate DNA upward through the pore; into the pore of the portal complex; and, ultimately, inside the procapsid. The release of DNA at each cycle by the ATP-hydrolyzing subunit allows for unidirectional DNA translocation. This figure and legend is reprinted with permission from: Hilbert et al Proc Natl Acad Sci USA 2015;112(29):E3792-9.

Furthermore, the pore of the pentameric docked model is lined with several conserved basic residues, indicating a role in DNA binding. Mutating these residues abrogates TerL binding to DNA[116]. Interestingly, mutation of residues in the nuclease domain has no substantial effect on DNA binding affinity. In fact, the TerL ATPase domain can bind DNA with nearly the same affinity as full-length TerL[116,117]. Assimilating these findings with the structural model yields a mechanistic model where the ATPase domains grip and translocate DNA in the central pore of the ring, while the nuclease domains bind to the portal and do not interact with the DNA during packaging (Figure 1.8A).

The lever model predicts that DNA is translocated by a lever-like motion of the ATPase domain while gripping to DNA[37]. These motions were identified by comparing crystal structures of P74-26 phage TerL, which shows significant conformational changes in the apo state or bound to a non-hydrolyzable ATP analog. Comparing the apo and ATP analog bound states, the lid subdomain undergoes a 13° rigid-body rotation. Because modeling and mutational analysis suggests that the lid subdomain forms the primary interaction surface between adjacent ATPase domains, this rotation has a substantial effect on the conformation of the ring. With the lid subdomain stabilized through neighboring contacts, the 13° rotation is transmitted to the Rossmann fold, which is the large domain of the ATPase module. This results in a lever-like movement of the ATPase domain (Figure 1.8B). This movement is calculated to shift the DNA binding region 8 Å upward towards the capsid and rotate DNA an estimated 2.3°.

This is very similar to the 2.5 bp DNA translocation step measured in phi29 motors[58,66]. It is predicted that upon either P_i or ADP release, the gripping subunit loses affinity for DNA and passes it off to the adjacent subunit. The adjacent subunit will now grip DNA tightly because it is in an ATP-bound state, which has been shown to have the highest affinity for DNA[37,116]. This handoff would then allow the subunit that hydrolyzed ATP to release ADP and P_i so that it can bind ATP, thereby resetting for the next cycle. By acting in concert with two subunits simultaneously, the DNA can be translocated efficiently. Moreover, the motor produces very high force because the lever arm (the entire Rossmann fold) is quite long.

The DNA "crunching" or "compression" translocation model

A third DNA packaging model is the DNA "crunching" or "compression" model (Figure 1.9). In this mechanistic model, the motor uses 'torsion' caused by DNA compression inside the terminase assembly to propel DNA into the capsid[118]. In other words, the DNA itself plays an active role in the force generation process. Specifically, with the motor attached to portal and DNA bound in the center, both portal and the terminase subunits grip the DNA and hydrolyze ATP (Figure 1.9A). ATP catalysis leads to a brief DNA compression which disrupts the B-form DNA structure, transiently adopting A-form DNA in what is referred to as a "crunched" state (Figure 1.9B)[28,112,118–120]. The return of this compressed state into the B-form state is assumed to cause the DNA

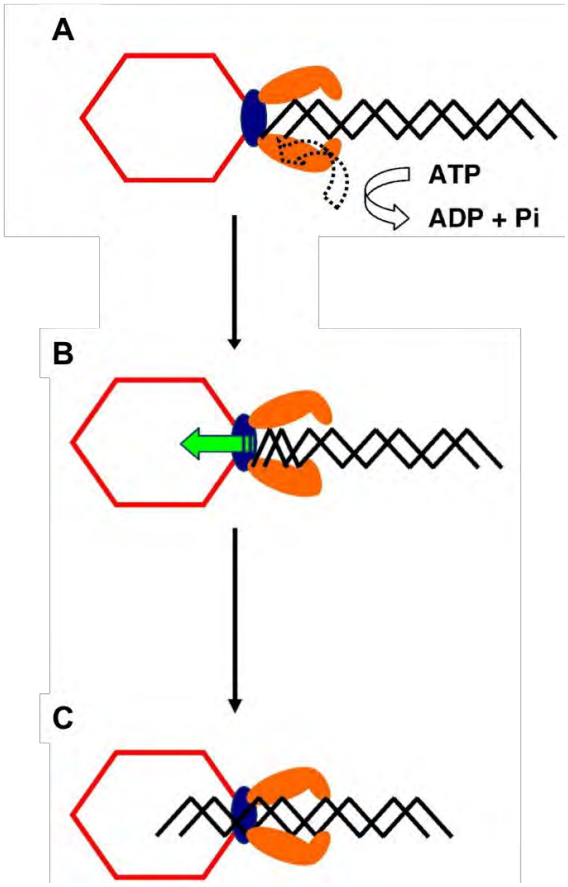


Figure 1.9. Phage DNA packaging via a torsional compression mechanism. (A) A fully assembled packaging complex is shown, with the empty prohead (red) carrying the portal ring (blue) shown at one vertex, and the DNA duplex drawn in black. Only two subunits of the large terminase (orange) are shown for clarity. The terminase subunits are sketched with a minor lobe, representing a flexible region that undergoes a conformational change (black dashed arrow) coupled with ATP binding and hydrolysis (white solid arrow). The exact steps and temporal order by which the binding and hydrolysis of ATP is coupled to movement of the large terminase during the reaction cycle is not specified. (B) Directed linear motion of the flexible arm of the large terminase subunit engages the DNA substrate and translocates this towards the prohead. This movement coupled with interaction of the DNA with the portal region causes induced changes to the helical pitch and temporarily stores energy. (C) The stored energy is released by translocation into the prohead (green arrow), restoring the B-form helical repeat. This figure and legend is adapted with permission from: Oram et al J Mol Biol. 2008; 381(1): 61–72.

translocation power stroke during the packaging cycle (Figure 1.9C). Evidence for this mechanism lies in several observations. A FRET experiment using a double-labeled fluorescent Y-DNA designed to stall the T4 packaging motor estimated 22-24% DNA compression within the terminase motor when packaging was stalled[118]. Additionally, subsequent packaging experiments using an intercalating dye, YOYO-1, showed dye release from the DNA during the packaging reaction. This observation is attributed to DNA compression inside the motor, as covalently attached DNA-binding labels are more readily packaged[112,118]. Observations that the motor does not readily package RNA-DNA hybrids also suggest this mechanism, as RNA-DNA hybrid complexes assume a mostly A-form structure and therefore are not compressible, preventing the scrunched to relaxed transition from driving the translocation power-stroke[121]. Two slightly different variants on the scrunching model have been proposed, with one model hypothesizing that the scrunching occurs within the terminase central pore[112], while the other model hypothesizing that the compression occurs within the portal channel[28,120]. I note that the compression model is not mutually exclusive with either of the two other models discussed here; it is possible that force generation is through a combination of the compression model and either the inchworm or lever models.

Packaging Termination

Phi29 and related bacteriophages replicate their genome as monomers rather than concatemers as most phages do. Thus, packaging is terminated

when the entirety of the DNA strand is packaged and no DNA cleavage is required. A 16 Å resolution cryoEM structure of a mature phi29 viral particle shows the DNA packaged inside the prohead with density in the channel of the connector and within the tailtube, both of which were assigned to the right-end gp3[68]. A second cryoEM structure of fiberless phi29 particles at 7.8 Å resolution confirmed the presence of gp3 within the tailtube, indicating that the right-end gp3 does not get fully packaged, while the connector-tailtube channel contains DNA bent in a toroid-like structure[122]. This novel 60 Å-in-diameter highly-bent DNA structure appears to be the result of DNA under extreme pressure. While the function of the toroid structure is unknown, the authors speculate it plays a role in holding the DNA inside the tail during infection. Further studies are necessary to explore this idea.

Terminases package concatemeric genomes and therefore must cut the DNA at the end of packaging. For viruses in this class, the genomes are typically synthesized by rolling circle DNA replication, resulting in multiple copies of the genome linked in a series (i.e. concatemers). Packaging is accomplished in one of two ways. In the first mechanism, typified by the lambda phage, the terminase cuts a specific site in the genome, resulting in exactly one genome length being packaged into the head (referred to here as ‘unit-length packaging’). In the second mechanism (known as ‘headful packaging’), the volume of packaged DNA triggers the nuclease cleavage mechanism, rather than a specific DNA sequence. The result of a headful packaging mechanism is that each capsid is

filled with slightly more than one genome of DNA, creating terminally redundant ends that can be circularly permuted[123,124].

For ‘unit-length’ bacteriophage such as lambda phage, the motor nuclease domain cuts at a specific sequence, known as *cosN* in lambda phage. Nicks are made on either side of the DNA strand, leaving a 5’ twelve nucleotide overhang[125]. For the initial cleavage event, a complex with the small terminase protein (gpNu1) and the *E. coli* protein Integration Host Factor (IHF) increase specificity and affinity for the large terminase to the *cosN* site. IHF bends the DNA nearly 180°[126,127], positioning two gpNu1 binding sites adjacent to each and presumably facilitating gpNu1 DNA binding, which in turn promotes large terminase nicking at *cosN*[73,83,128]. During termination, an upstream site, *cosQ*, and a downstream site, *I2*, are necessary for accurate nicking. Mutations in *cosQ* lead to loss of cleavage at the *cosN* site, causing aberrant termination[129]. It is proposed that *cosQ* plays a role in properly positioning the terminase/nuclease domains for antiparallel DNA cleavage at the *cosN* site, and that the volume of packaged DNA may trigger its recognition[130].

In comparison to ‘unit-length’ packaging bacteriophage, ‘headful’ bacteriophage do not terminate packaging at specific DNA sequences; instead, termination is determined by the volume of the head[131–133]. Because headful phages generally package slightly more than one genome into the capsid, the genome is terminally redundant for circular permutation[80,123]. This process can result in the phenomenon of generalized transduction, in which phages can carry

segments of bacterial genome from one cell to another[134]. There are several strategies for headful packaging. For example, phage P22 cleaves after ~2-10% of the next genome in the concatemer, which results in the next packaging cycle starting with a region downstream from the initial packaging recognition sequence[124,131]. Other cleavage strategies are seen across different phage, suggesting that termination (and initiation) are controlled idiosyncratically by different phage[80].

Headful terminases must trigger nuclease activity through a mechanism that senses how full the capsid is[124]. One hypothesis for how this is accomplished relies on the conformational state of the portal protein. In a recent study of the P22 portal ring, the authors found that during packaging the portal adopts an asymmetric assembly that binds tightly to the terminase motor (it is of note that the discovery of this asymmetric state also addressed how a pentameric motor could bind a dodecameric portal, a question that has eluded the field for several decades)[19]. However, once the head nears complete fullness, DNA binds around the capsid-enclosed portion of portal[19,135]. DNA binding in this region triggers the asymmetric portal assembly to adopt a symmetric ring, thus losing affinity for the pentameric motor and presumably discharging it from the capsid head[19]. After portal releases the motor, the motor nuclease domains cleave DNA, severing the packaged DNA from the rest of the concatemer, and terminating the packaging reaction.

Headful packaging mechanisms necessitate strict control to prevent premature cleavage of DNA. The small terminase subunit may play a regulatory role in this process, as it suppresses large terminase nuclease activity (while increasing ATPase activity)[43,49]. How this mechanism works is unclear, and the role of the small terminase during DNA translocation has yet to be fully resolved.

One model of DNA cleavage relies on separate DNA 'binding' and 'cleavage' interfaces of the nuclease domain. Terminases have nuclease domains of the RuvC/RNase-H family that use three acidic active site residues to coordinate two divalent metals (either Mg^{2+} or Mn^{2+}) to catalyze the cleavage reaction[39–41,116,136–139]. Crystal structures of the T4 nuclease domain show a patch of basic residues far from the active site that are presumed to make electrostatic contacts with the DNA backbone during 'translocation mode'[41]. Once endonuclease activity is triggered, the DNA is transferred through an unknown mechanism to the nuclease active site for cleavage. This putative DNA translocation interface is proposed to regulate the endonuclease activity by holding DNA away from the active site during translocation. However, it was later noted that no positively charged residues from this region are conserved throughout other bacteriophage nuclease domains[41,137], and that the nuclease domain has very poor affinity for DNA[116].

Further structural studies attempt to explain nuclease regulation through nuclease domain conformational changes. Analysis of the SPP1 bacteriophage nuclease domain crystal structure predicts that a conserved extended β-hairpin

clashes with DNA bound in an RNase-H-like conformation, and that the loop must reorient upon DNA binding[137]. Heterogeneity of loop positioning in different bacteriophage species and normal mode analysis calculations support this idea, suggesting that loop flexibility may limit active site accessibility for DNA, and thus acts as a nuclease regulation mechanism[137]. Later studies using thermophilic terminases suggest that movement of this loop is not required if the terminase uses a different DNA binding mode[116,136]. It was suggested that the nuclease domain cleaves DNA in a fashion similar to the thermophilic RuvC resolvase rather than RNaseH[140]. In this model, the β-hairpin and residues surrounding the active site cradle DNA within a groove, positioning it in a desirable conformation for cleavage[116,136].

Another suggested mechanism of cleavage regulation is that DNA packaging occurs too quickly for nucleolytic activity to take place[117,130]. While this is possible, extended motor stalls for extended time periods (seconds to hours) do not result in DNA cleavage, indicating this ‘kinetic competition’ model is not a primary mechanism for nuclease regulation[95,96,111,118,141].

Recently, the ‘steric hindrance model’ was proposed to explain regulation of terminase nucleolytic activity (Figure 1.10). In this mechanism, the nuclease domains bind to portal and capsid during packaging, preventing the nuclease active site from engaging with DNA[116]. Evidence suggests that the ATPase domains are solely responsible for gripping DNA, while the nuclease domains have no measurable affinity, supporting this arrangement. After ejection from

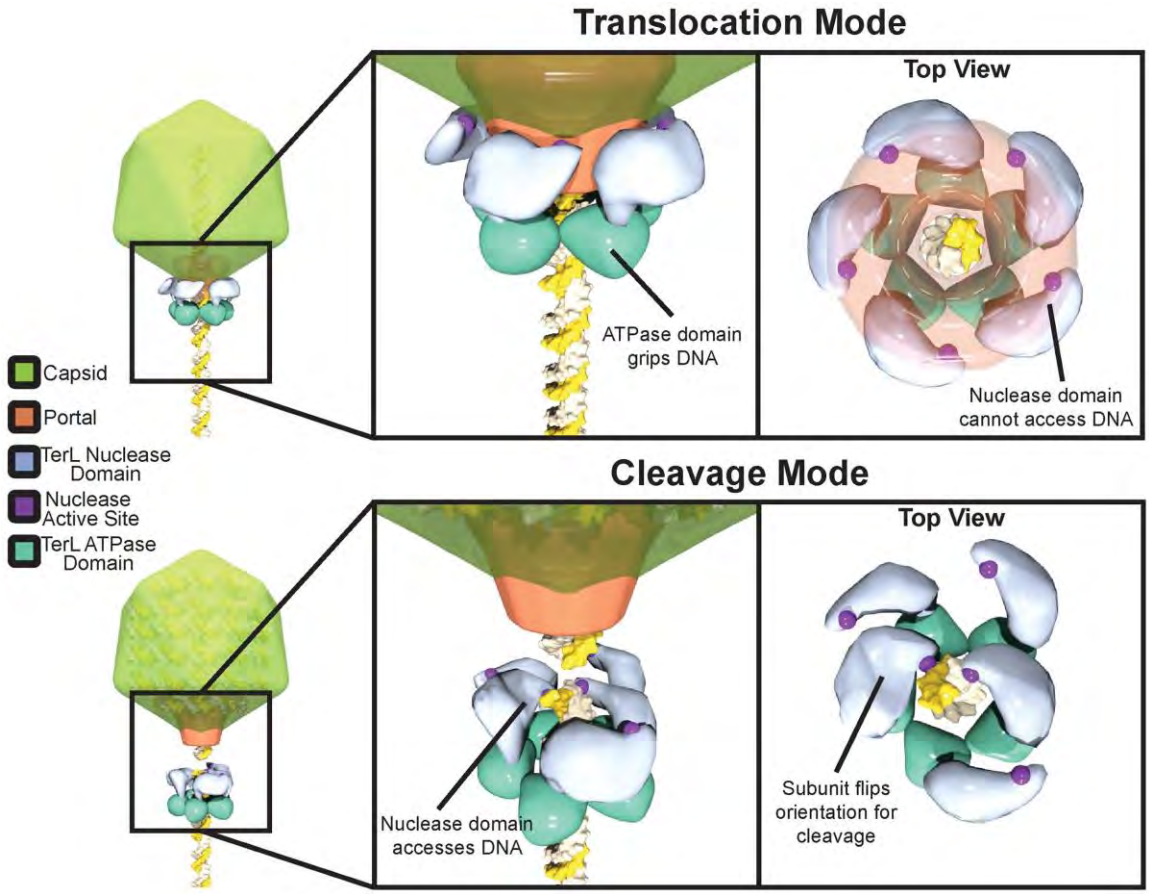


Figure 1.10. Proposed 'steric hindrance' model for nuclease regulation. During 'translocation mode' the nuclease domain active site is sequestered from DNA by interactions of the large terminase with portal and capsid, preventing premature cleavage. The ATPase domain serves as the sole surface for gripping DNA during packaging. Upon completion of packaging the large terminase enters 'cleavage mode'. The large terminase loosens its attachment to the portal and capsid, releasing the inhibition of the nuclease domains. The ATPase domains remain tightly bound to DNA. The nuclease domains rearrange to cleave each of the antiparallel DNA strands. Although depicted as a blunt cut, cleavage could also leave overhangs depending on how both nuclease domains engage DNA. This figure and legend is reprinted with permission from: Hilbert et al Nucleic Acids Res 2012;45(6):3591-3605.

portal, two nuclease domains bind and cleave DNA in an antiparallel arrangement, forming a double-stranded break. This rearrangement of the nuclease domains is made possible by a flexible linker connecting to the ATPase domain, allowing the nuclease domains to adopt the correct orientation for DNA cleavage[37,39,41,116,138,142].

Once packaging is finished, the motor is replaced with the virus neck and tail. In most phage, the neck proteins are either hexameric or dodecameric, matching the 12-fold symmetry of the portal[17,143,144]. The neck proteins assemble on portal, connecting portal and the capsid to the pre-assembled bacteriophage tail[145]. Bacteriophage tails are diverse and used for phage classification. In tailed-bacteriophage (otherwise known as *caudovirales*) there are three distinct classes: *myoviridae*, *podoviridae*, and *siphoviridae*[146]. *Myoviridae* have long contractile tails, while *podoviridae* have short non-contractile tails, and *siphoviridae* have long non-contractile tails. All three of these tail-types are hollow, and DNA is ejected through them during infection. Once tail and appendage assembly is complete, the viruses reach full maturity .

It is not clear how the DNA does not get prematurely released during the exchange of motor for tail. Various mechanisms for maintaining the pressurized DNA have been proposed. For example, the portal could act as a one-way valve[25,26]. Additionally, the conformation of the DNA itself may help maintain packaged genome. Cryo-electron microscopy studies show the DNA tightly spooled in the capsid, typically with the central axis of the spool coincident with

the long axis of the tail[17,122,135,147–149]. The arrangement of DNA appears to be most homogenous at the periphery of the shell where the genome contacts the capsid directly, with the 'layers' of DNA becoming progressively more heterogeneous towards the center. DNA is often found to be ordered in the portal complex and even within the tail, as if it were locked down but poised for ejection. Typically, the DNA sequence that enters the head first is the last to leave the head upon infection, suggesting a simple unidirectional model for spooling of DNA in the head[150] (T4-like phage are an exception; the first sequence in during packaging is the first sequence out during infection[151]). Moreover, there are often proteins that also reside in or nearby this channel that are ejected soon after infection[152,153]; these 'ejection proteins' presumably play an important role in early stages of infection, including the process of crossing the periplasm[154]. It remains a mystery whether these proteins hold DNA in the capsid until infection or the motor plays a role in packaging these ejection proteins inside the capsid.

Scope of thesis research

During my dissertation, my focus has been the mechanisms of DNA translocation and recognition in viral packaging motors. In Chapter II, I present my work with the large terminase protein of the thermophilic bacteriophage P74-26. With the help of a previous postdoctoral researcher, I crystallized the TerL ATPase domain, which guided numerous motor characterization experiments. I identified critical residues for ATP hydrolysis, assembly, and DNA binding, all of

which was synthesized with molecular docking and SAXS experiments to form a model for pentameric motor assembly. With this structural model and experimental data, we formed the “lever” mechanistic model for how the motor grips and translocates DNA. These findings were novel, as they contributed new models for motor mechanism and regulation to the field.

In Chapter III, I describe my work establishing the TerS protein of P74-26 as a system for studying small terminases. For this project, we identified a putative TerS gene in the P74-26 genome. I developed the purification protocol for the putative TerS protein and characterized the complex. I showed that the complex forms a nonameric assembly, binds DNA, and regulates TerL activity, thus verifying the protein as a small terminase. I also determined symmetric and asymmetric cryoEM reconstructions of the TerS ring to 3.8 Å and 4.8 Å resolution respectively. With these structures, I hypothesize that the C-terminal β-barrel is not conserved throughout all TerS proteins, and thus the binding and regulation of TerL is not dependent on it in P74-26 TerS. Additionally, I observe that the proposed ‘DNA binding’ helix-turn-helix motifs are arraigned in a different conformation than known TerS structures, indicating that P74-26 TerS likely uses a different DNA binding mechanism.

Portions of this chapter are under review as a section in the 4th Edition of the Encyclopedia of Virology (Elsevier).

CHAPTER II

Structure and mechanism of the ATPase that powers viral genome packaging

Abstract

Many viruses package their genomes into procapsids using an ATPase machine that is among the most powerful known biological motors. However, how this motor couples ATP hydrolysis to DNA translocation is still unknown. Here, we introduce a model system with unique properties for studying motor structure and mechanism. We describe crystal structures of the packaging motor ATPase domain that exhibit nucleotide-dependent conformational changes involving a large rotation of an entire subdomain. We also identify the arginine finger residue that catalyzes ATP hydrolysis in a neighboring motor subunit, illustrating that previous models for motor structure need revision. Our findings allow us to derive a structural model for the motor ring, which we validate using small-angle X-ray scattering and comparisons with previously published data. We illustrate the model's predictive power by identifying the motor's DNA-binding and assembly motifs. Finally, we integrate our results to propose a mechanistic model for DNA translocation by this molecular machine.

Introduction

Double-stranded DNA (dsDNA) viruses ranging from bacteriophages to the human pathogens of the herpesvirus family form infectious virions by packaging their genomes into preformed procapsids using a powerful ATPase machine[155]. The viral genome packaging motor is a multicomponent molecular machine that must complete several tasks in sequential order, the foremost of which is the ATP-dependent pumping of viral DNA into the procapsid (Figure

2.1A). Because the DNA progresses from a flexible state to a semi-crystalline state as it fills the capsid interior, the motor pumps against a tremendous force. The pressures inside the filled capsid are estimated to reach 60–70 atm[56,156], equivalent to 10-fold higher than a bottle of champagne. Thus, the viral packaging motor represents one of the most powerful biological motors known[56,95].

The central component of the packaging motor is the ATPase subunit, which drives DNA translocation. The ATPase subunit is a member of the additional strand, conserved glutamate (ASCE) superfamily of ATPases[34,35]. In herpesviruses, as well as many bacteriophages, this ATPase is from a specific clade of the ASCE family called the terminase family[35,155]. In viruses that use a terminase-type motor for genome packaging, the motor consists of several proteins that assemble into homomeric rings[79] (Figure 2.1A). The large terminase (TerL) protein harbors the motor's two enzymatic activities[79]: the ATPase activity that pumps DNA into the capsid and an endonuclease domain that cleaves packaged DNA from the remaining concatemeric DNA when the capsid is full. Cryo-electron microscopy (cryo-EM) studies indicate that a pentamer of TerL subunits attaches to the capsid by binding to a dodecameric assembly called portal[41]. However, there are conflicting reports as to the orientation of TerL relative to portal during packaging [41,113,157].

A structural model for the bacteriophage T4 TerL ring has been previously proposed[41] with these salient features: (i) the nuclease domains assemble to

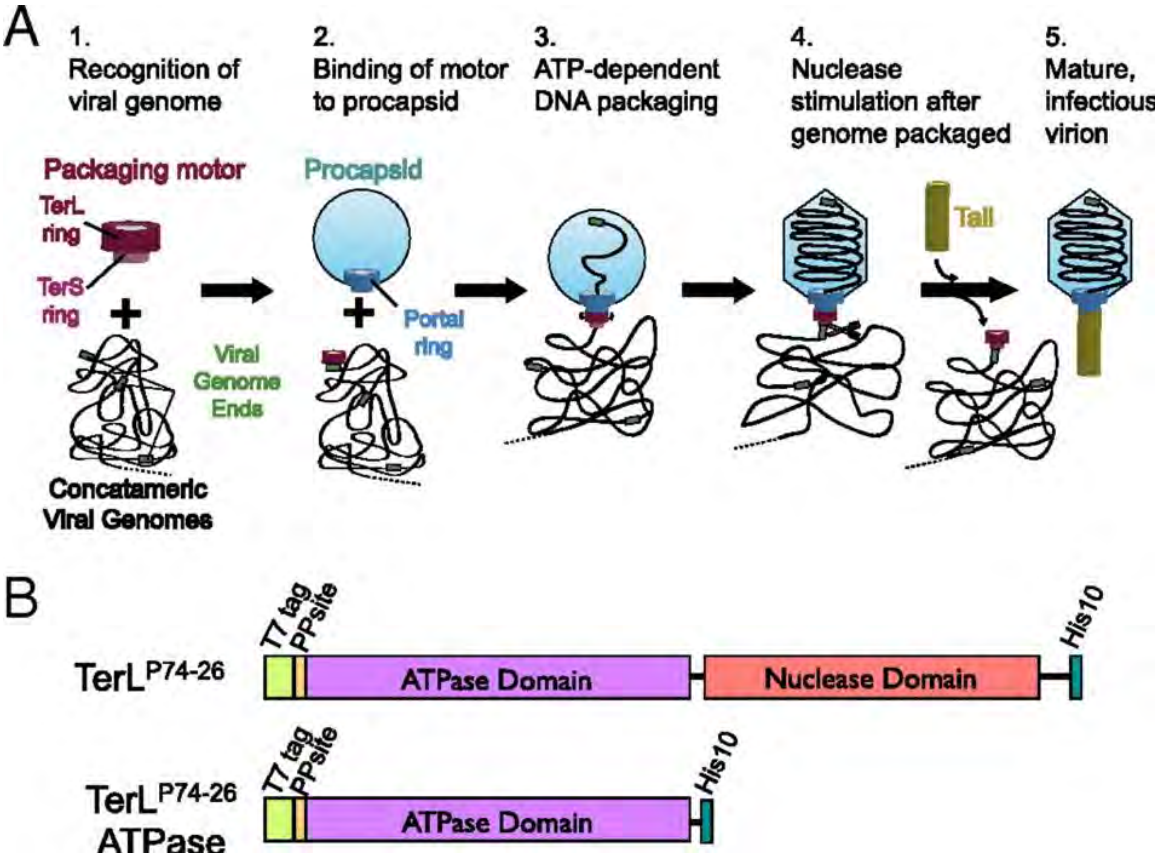


Figure 2.1. Genome packaging in bacteriophage and TerL constructs. (*A*) Maturation of dsDNA viruses is a five-step process: (1) the small terminase (TerS) subunit of the motor recognizes the concatemeric viral genome; (2) the motor binds to the portal complex on the capsid; (3) TerL hydrolyzes ATP to translocate DNA into the capsid; (4) after the genome is translocated into the capsid, TerL switches its enzymatic activity from translocation to DNA cleavage; and (5) the motor is finally released for maturation of another virus while portal binds to the tail proteins to complete a mature, infectious virion. (*B*) TerL constructs used in this study. Full-length P74-26 TerL is produced with an N-terminal T7-gp10 expression tag that is removed by prescission protease (PP) cleavage. The C-terminal His10 tag is noncleavable.

constitute the ring distal to the capsid, with the ATPase domains protruding radially to interact with portal; (ii) consequently, it was proposed that the active site residues that catalyze ATP hydrolysis are derived from one subunit exclusively; (iii) the pore of the ring has a net negative charge and is largely unconserved; and (iv) DNA was proposed to be bound primarily by the nuclease domain. This model is substantially different from models observed for other nucleic acid translocases of the ASCE family, in which the pore is largely conserved and positively charged, and ATP hydrolysis is catalyzed *in trans*[105,158–161]. Resolving these disparities between the proposed assembly and conserved features of the ASCE family is necessary to understand the molecular mechanisms of motor action.

Several issues have prevented elucidation of the terminase mechanism. First, the TerL proteins that have been studied *in vitro* thus far fail to form rings in isolation[42,43,78,162]. Second, crystal structures of the TerL proteins from bacteriophages Sf6 and T4[38,39] (gp2 and gp17, respectively; hereafter known as Sf6 TerL and T4 TerL) do not show ATP-dependent conformational changes sufficient to power DNA translocation. Consequently, the structural mechanism coupling ATP hydrolysis to DNA translocation remains nebulous. Structural analysis of the WT TerL ATPase domain in different nucleotide states will be necessary to address this issue.

To answer these critical questions, we used a novel model system for understanding motor function. Because packaging motor ATPases are often

insoluble[55,163,164], we first sought a packaging motor system that is highly soluble and forms rings in isolation to enable mechanistic analyses in a simplified system. We anticipated that TerL from a thermophilic phage would have higher stability, solubility, and propensity to form a pentamer. Moreover, we expected that the features necessary for DNA translocation would be more exaggerated in a thermophilic motor because packaging is expected to be more difficult at high temperature due to the increased entropic penalty of ordering DNA within the capsid[156].

Here, we describe the TerL protein from the phage P74-26 (hereafter known as P74-26 TerL), a thermophilic siphovirus that infects *Thermus thermophilus*[165]. We show that the P74-26 TerL assembles into a pentamer that has ATPase and DNA-binding activity. We report the structure of the P74-26 TerL ATPase domain in both the apo and ADP•BeF_3-bound states, and observe a large conformational change in a subdomain in response to ATP hydrolysis and release. We also show that ATP hydrolysis is catalyzed by a conserved arginine finger residue that is provided *in trans* by a neighboring TerL subunit. Finally, a combination of biochemical, biophysical, structural, and computational data is used to build a structural model for the TerL ring and a mechanistic model for how ATP hydrolysis results in translocation of DNA.

Materials and Methods

Cloning

The gene for P74-26 TerL was synthesized with codon optimization for expression in E. coli by Genscript Corporation. This gene was cloned into the BamHI and XhoI sites of a modified pET24a vector with an N-terminal T7-gp10 expression tag and prescission protease cut site, and a C-terminal noncleavable His10 tag[106]. Site-directed mutagenesis was performed using a mutagenesis protocol similar to QuikChange[166]. Enzymes were purchased from New England BioLabs. Oligonucleotides were obtained from IDT. The ATPase domain was constructed by deleting the codons encoding the C-terminal 229 residues using the aforementioned mutagenesis protocol[166]. This deletion site was chosen based on limited proteolysis and bioinformatic analysis (data not shown).

Expression and Purification of P74-26 TerL

All proteins were expressed in E. coli BLR-DE3 cells containing the pET24a-P74-26-TerL plasmid. Bacterial cultures were grown in Terrific Broth (Research Products International) supplemented with 30 μg/μL kanamycin at 37 °C until an OD600 of 0.7 was reached. Cells were then placed at 4 °C for 20 min, after which overnight expression at 18 °C was induced by the addition of isopropyl-β-D-thiogalactopyranoside to 1 mM. Cells were pelleted by centrifugation and resuspended in buffer A [500 mM NaCl, 20 mM imidazole, 50 mM Tris (pH 8.5), 5 mM β-Mercaptoethanol (βME), 10% (vol/vol) glycerol] before being flash-frozen in liquid nitrogen for storage. BLR-DE3 cells with SeMet-labeled ATPase domain (residues 1–256) were grown in minimal media. When cultures reached an

OD600 of 0.5, they were supplemented with amino acids, except for Met, and grown to an OD600 of 0.8. Cultures were then supplemented with SeMet. Induction and harvesting were performed as above.

For full-length constructs, thawed cells were lysed in a cell disruptor and pelleted to clear the lysate. Lysate was loaded onto a 10-mL centrifuge spin-column (Thermo Scientific) containing Ni-affinity beads (Thermo Scientific) pre-equilibrated with buffer A. Protein was eluted with buffer B [500 mM NaCl, 500 mM imidazole, 50 mM Tris (pH 8.5), 5 mM βME, 10% (vol/vol) glycerol] and dialyzed overnight at room temperature into buffer QA [25 mM Tris (pH 8.5), 125 mM NaCl, 2 mM DTT, 10% (vol/vol) glycerol]. Before dialysis, prescission protease was mixed with eluate to cleave the T7 expression tag. The sample was then loaded onto spin-columns containing Q resin (GE Healthcare) pre-equilibrated with buffer QA. After protein loading, the Q resin was washed with buffer QA and contaminants were eluted with buffer QA1 [25 mM Tris (pH 8.5), 200 mM NaCl, 2 mM DTT, 10% (vol/vol) glycerol]. Protein was eluted with buffer QB [25 mM Tris (pH 8.5), 450 mM NaCl, 2 mM DTT, 10% (vol/vol) glycerol] and dialyzed overnight at room temperature into buffer QA before concentrating.

For purification of the TerL-ATPase domain, cells were lysed and pelleted as above. After pelleting, all purification steps proceeded at room temperature. Lysate was loaded onto a pre-equilibrated 5-mL His-Trap column (GE Healthcare). The column was washed with buffer A and eluted with buffer B. Eluate was diluted four-fold with buffer A, and prescission protease was added.

The protein was dialyzed overnight in buffer QA. The protein was then filtered and flowed over a 5-mL HiTrap Q-HP column (GE Healthcare) pre-equilibrated in buffer QA. The flow-through was collected, concentrated, and flash-frozen.

Coupled Enzyme ATPase Assays

Coupled enzyme assays were performed at room temperature [167] using the following concentrations (unless otherwise noted): 6 U/mL pyruvate kinase, 6 U/mL lactate dehydrogenase, 1 mM phosphoenolpyruvate, 340 μM NADH, 50 mM Tris (pH 7.5), 500 μM Tris(2-carboxyethyl)phosphine, 100 mM potassium chloride, 5 mM magnesium chloride, 1 mM ATP, and 0.5 μM TerL. Absorbance was measured in a 96-well format with a Perkin–Elmer Victor3 1420 multichannel counter using an excitation filter centered at 355 nm, with a bandpass of 40 nm. In this assay, every NADH oxidized to NAD+ corresponds to one ATP hydrolyzed. To convert the measured absorbance directly to ATP concentration, we performed a standard curve so as to measure the NADH extinction coefficient directly under our experimental conditions.

To understand how TerL concentration affects ATPase activity, we measured ATPase activity over a wide concentration range of TerL. The rise in ATPase rate is slow and nonlinear at low TerL concentrations, but it increases linearly at higher concentrations (Figure 2.2D). The nonlinear activity at low TerL concentrations is not due to the limits of our assay, because we are able to measure ATPase rates reliably below the rate observed here using a different

ATPase. The slow increase in ATPase rate at low concentration is most likely due to assembly of a TerL pentamer. It is noteworthy that the gp16 ATPase from the packaging motor of the phi29 phage shows a similar nonlinear increase in ATPase rate[168], suggesting that an assembly of the ATPase subunits may be a necessary step in full ATPase activity for other viral motors. Regardless, the experiments that follow were performed at a concentration of TerL (0.5–1 μM) that is well within the linear range. Data comparing the rate of ATP hydrolysis to [ATP•$MgCl_2$] were fit using the Hill equation: $v = V_{max} * [ATP•MgCl_2]^n/(K_{half\text{-}maximal}^n + [ATP•MgCl_2]^n)$, where V_{max} is the maximum rate, n is the Hill coefficient, and $K_{half\text{-}maximal}$ is the substrate concentration that yields half-maximal activity.

DNA-Binding Assays

Full-length WT (0.13 nmol) or mutant P74-26 TerL (0.13 nmol) was incubated with 125 ng of 100-bp ladder (New England BioLabs) for 30 min at room temperature in apo, ADP, ADP•BeF_3, or ATP buffer as indicated. Apo buffer was composed of 50 mM Tris (pH 7.5), 150 mM KCl, and 1 mM DTT. ADP buffer was composed of 50 mM Tris (pH 7.5), 150 mM KCl, 1 mM DTT, 1 mM ADP, and 10 mM $MgCl_2$. ADP•BeF_3 buffer was composed of 50 mM Tris (pH 7.5), 150 mM KCl, 1 mM DTT, 1 mM ADP, 10 mM NaF, 4 mM $BeCl_2$, and 10 mM $MgCl_2$. ATP buffer was composed of 50 mM Tris (pH 7.5), 150 mM KCl, 1 mM DTT, 1 mM ATP, and 10 mM $MgCl_2$. Samples were then loaded onto a 1.5% (wt/vol) agarose

gel with a 1:14,000 dilution of GelRed dye (Phenix Research) and run at 80 V for ~2 h.

Crystallization

SeMet-labeled, apo, and ADP•BeF_3 -bound P74-26 TerL ATPase domain were crystallized with the hanging drop vapor diffusion method. Drops were set with final protein concentrations of 3–8 mg/mL apo, SeMet-labeled, and ADP•BeF_3-bound samples were premixed with 10 mM magnesium chloride and 10 mM DTT before crystallization. The R139A variant was premixed only with DTT. ADP•BeF_3 crystals were premixed with 2 mM ADP, 10 mM sodium fluoride, and 4 mM beryllium fluoride. ADP (2 mM) was premixed with SeMet-labeled protein. SeMet-labeled crystals formed in 0.9 M ammonium sulfate and 0.1 M trisodium citrate (pH 5.0). Cubic apo crystals formed in 0.7 M ammonium sulfate and 0.1 M trisodium citrate (pH 5.0). Apo R139A cubic crystals formed in 0.5 M ammonium sulfate and 0.1 M trisodium citrate (pH 4.5). Hexagonal apo crystals formed in 11% (wt/vol) PEG 2000, 0.15 M ammonium sulfate, and 0.1 M trisodium citrate (pH 5.0). ADP•BeF_3-bound crystals formed in 19% (wt/vol) PEG 3350, 0.05 M ammonium sulfate, and 0.1 M trisodium citrate (pH 5.0). Single crystals formed only after microseeding. Cryogenic buffer conditions for each crystal type consisted of elevated respective precipitant concentrations supplemented with 30% (vol/vol) ethylene glycol. Cryogenic conditions also contained all reagents that were premixed with each respective protein sample before crystallization.

Data Collection and Structure Solution

SeMet and apo ATPase domain datasets were collected at ALS beamline 5.0.1 at a wavelength of 0.9774 Å. SeMet structure experimental phases were calculated by SAD[169] using the PHENIX Autosol pipeline[170]. The SeMet model was used for molecular replacement for the apo structure. The second apo ATPase crystal data were collected at Advanced Photon Source beamline 23-ID-B at a wavelength of 1.033 Å. ADP•BeF_3 crystal data were collected at Brookhaven National Laboratory beamline X25 at a wavelength of 1.100 Å. ADP•BeF_3 data were truncated due to anisotropy in the diffraction. After scaling, data were processed using the University of California, Los Angeles Molecular Biology Institute Diffraction Anisotropy Server[171]. The apo R139A mutant crystal data were collected on a home source MicroMax007-HF/Saturn 944 CCD detector X-ray diffraction system at a wavelength of 1.54 Å. A C-terminal–truncated apo structure was used for molecular replacement. All datasets were processed with HKL3000[172]. Model building and refinement were performed with Coot[173] and PHENIX[174].

SAXS Data Collection and Analysis

Samples were dialyzed into buffer [25 mM Tris (pH 8.5), 125 mM NaCl, 4 mM DTT, 2% (vol/vol) glycerol] using dialysis buttons (Hampton Research) and filtered through a 0.45-μm membrane (Millipore). WT P74-26 TerL exhibited low levels of aggregation, which hampered interpretation of data. We hypothesized that the basic patch (Figure 2.4) contributes to the aggregation of P74-26 TerL.

To avoid aggregation, we made the R104E mutant. Indeed, this mutant shows no signs of aggregation (Figure 2.11A-C). Samples were shipped at 4 °C to the ALS, Lawrence Berkeley National Laboratory, for data collection at the SIBYLS beamline[175]. Data were collected at 10 °C with exposures of 0.5 s, 1 s, and 6 s and at several concentrations of TerL, and buffer blanks were subtracted to give protein scattering. Due to radiation sensitivity at longer exposures, only the 1-s exposure was used for analysis. Data were processed using the ATSAS software package[176]. The SAXS reconstruction was calculated using GASBOR[191] in real-space mode using fivefold symmetry. The docked model was optimally fit to the SAXS envelope by converting the GASBOR dummy atoms to a map using the "molmap" and "fit to map" commands in UCSF Chimera[177].

Molecular Docking and Modeling of the P74-26 TerL Ring

The model of the P74-26 TerL ATPase ring was calculated with the ADP•BeF_3-bound structure of the P74-26 TerL ATPase domain using M-ZDOCK[178]. Fivefold symmetry was applied, but no other restraints were imposed during the docking. To model the P74-26 TerL nuclease domain, we used I-TASSER[179] to generate a homology model with the nuclease domains of Sf6 TerL and T4 TerL as structural templates. To model the full-length ring, we aligned the full-length T4 structure[41] [Protein Data Bank (PDB) ID code 3CPE] onto the ATPase domains of our model and then superposed the P74-26 TerL nuclease homology model onto the T4 nuclease domains.

To model the potential conformational changes that drive DNA translocation within the ring upon ATP hydrolysis, we first positioned the apo ATPase structure onto a subunit within the modeled ring, using only the lid subdomain for alignment. To model the effect of this change on downstream subunits in the ring assembly, a second ATPase domain (bound to ADP•BeF_3) was placed adjacent to the apo subunit in an orientation identical to the orientation in the ring model. To calculate the approximate movement of individual subunits that may be associated with ATP hydrolysis, we took advantage of the convenient output format of M-ZDOCK, in which all symmetry-related atoms are in the same x–y plane. Therefore, overall translation of a DNA-binding residue upon ATP hydrolysis and release (in this case, C_α of Arg101) can be simply calculated from the change in the z coordinate. To calculate the in-plane rotation angle per step, we translated the C_α of Arg101 of the "translocating" subunit back into the original plane. We then calculated the in-plane angle swept out by the motion of this subunit by measuring the angle between these three points: the original position of Arg101 C_α within the ring, the center of mass of all Arg101 C_α atoms, and the new (back-projected along the z axis) Arg101 C_α position.

Modeling of the T4 TerL Ring and Fitting to Cryo-EM Reconstruction

To make the model of the T4 TerL ring, we aligned the full-length T4 TerL structure (PDB ID code 3CPE)[41] onto the ATPase domains of our docked

pentameric model using PyMOL (DeLano Scientific). The resulting model was then fit into the cryo-EM reconstruction of the T4 capsid with the TerL ring bound (EMDB-1572)[41] in UCSF Chimera using the fit to map command[177].

Results

Characterization of P74-26 TerL

We recombinantly expressed and purified P74-26 TerL in Escherichia coli (Figure 2.2A). P74-26 TerL predominantly elutes in size exclusion chromatography (SEC) at a volume consistent with a pentamer [molecular mass by SEC (MM_{SEC}) of ~270 kD, $MM_{pentamer}$ of 285 kD] with a small monomer peak (MM_{SEC} of ~57.3 kD, $MM_{monomer}$ of 57.1 kD) (Figure 2.2B). This observation is notable, because other TerL proteins are predominantly monomeric in isolation[42,43,78,162], only forming a pentameric ring when bound to the procapsid[41]. P74-26 TerLis an active ATPase that exhibits apparent cooperativity in ATP hydrolysis, suggesting that the active sites are coupled (Figure 2.2C–E). P74-26 TerL tightly binds DNA when locked into an ATP-bound state by the ATP mimic ADP•beryllium trifluoride (BeF_3) (Figure 2.2F). Thus, because the isolated P74-26 TerL protein can assume a pentamer consistent with its functional form, we argue that P74-26 TerL is an excellent minimal model for understanding packaging motor structure.

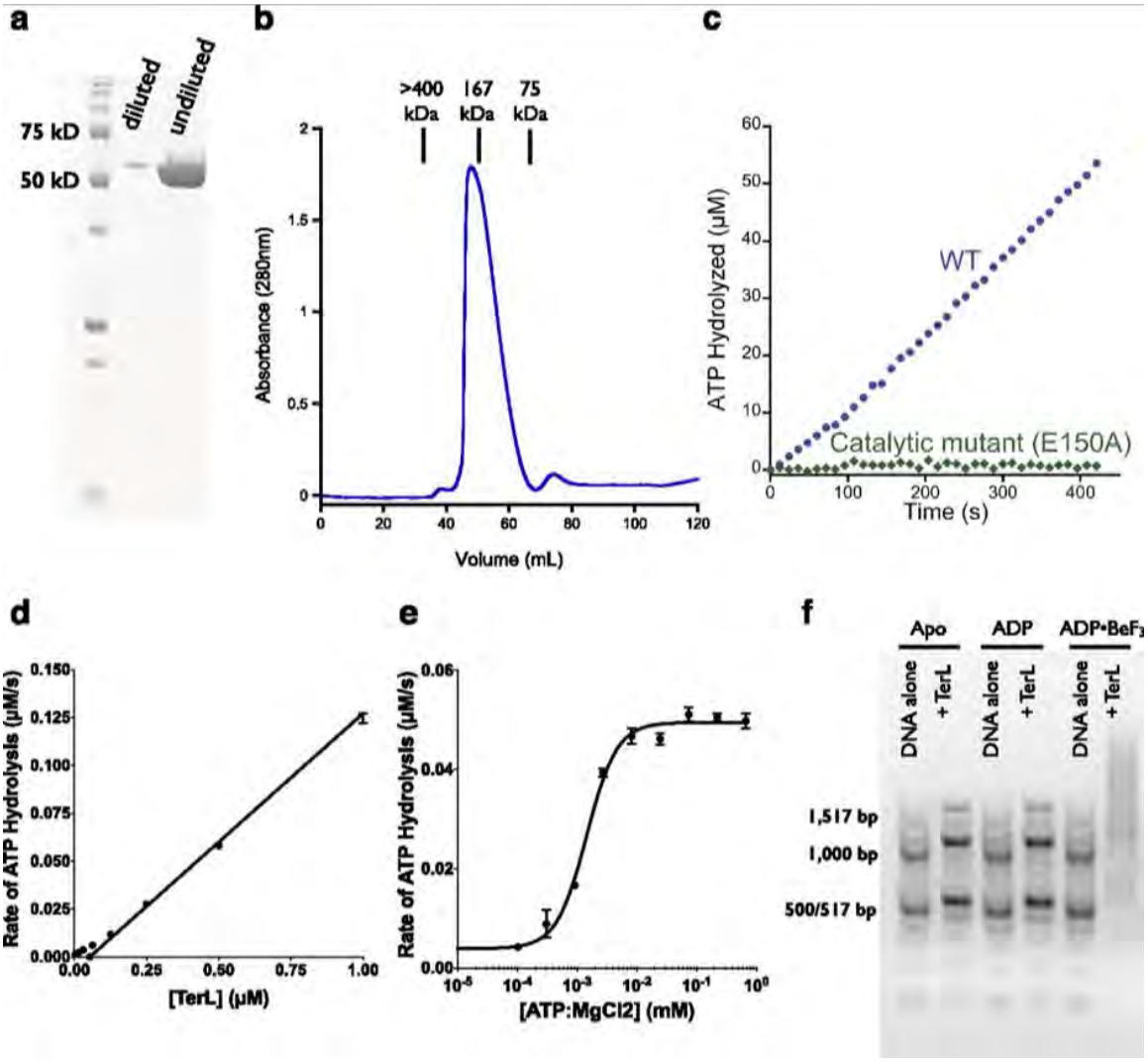

a
75 kD
50 kD
diluted
undiluted
b
>400 kDa
167 kDa
75 kDa
Absorbance (280nm)
Volume (mL)
c
ATP Hydrolyzed (μM)
WT
Catalytic mutant (E150A)
Time (s)
d
Rate of ATP Hydrolysis (μM/s)
[TerL] (μM)
e
Rate of ATP Hydrolysis (μM/s)
[ATP:MgCl2] (mM)
f
Apo
ADP
ADP•BeF3
DNA alone
+ TerL
1,517 bp
1,000 bp
500/517 bp

Figure 2.2. Characterization of P74-26 TerL. (*A*) SDS/PAGE analysis of P74-26 TerL purity. Diluted P74-26 TerL shows that the protein runs as a single band and that the undiluted protein shows no significant protein contaminants. Precision Plus Protein Unstained Standards (BioRad) are used as molecular mass markers. (*B*) P74-26 TerL elutes in SEC (S200) at a volume consistent with a pentamer. (*C*) P74-26 TerL shows significant ATPase activity as measured by a coupled-enzyme assay (50), whereas a variant that removes the catalytic glutamate (E150A P74-26 TerL) is completely inactive, suggesting that the observed ATPase activity in WT TerL is not due to contaminants. The E150A variant's active site is disabled *in cis* but can still donate an intact arginine finger to an adjacent subunit in the assembly. (*D*) Steady-state ATPase activity is shown for different concentrations of TerL. Error bars are the SD from at least three replicates. (*E*) Michaelis–Menten kinetic analysis of P74-26 TerL. Titration of ATP indicates that P74-26 TerL has high affinity for ATP•$MgCl_2$ (concentration for half-maximal activity of ~1.5 ± 0.2 μM) and robust turnover [apparent catalytic rate constant ($k_{cat,app}$) of ~0.091 ± 0.003 s^{-1}]. ATP hydrolysis is sigmoidal with respect to the concentration of ATP•$MgCl_2$, with an observed Hill coefficient of 1.7 ± 0.3. This apparent cooperativity suggests that there is coupling between ATPase active sites within the TerL assembly. (*F*) P74-26 TerL binds DNA when incubated with an ATP mimic. One hundred twenty-five nanograms of 100-bp ladder (New England BioLabs) was incubated with either $TerL_{P74-26}$ or a buffer control for 30 min at room temperature and run on a 1.5% (wt/vol) agarose gel at 80 V. Tight binding to DNA is only observed for P74-26 TerL that is preincubated with ADP•BeF_3. Slight shifts in the DNA migration are observed for P74-26 TerL in the apo state or incubated with ADP or ATP, which may indicate weak binding.

Structure of the TerL ATPase Domain

We next sought to elucidate the structural mechanism of genome packaging by P74-26 TerL. Because DNA translocation is ATP-dependent, considerable insight can be obtained from the structure of the isolated ATPase domain. Therefore, we expressed, purified, and crystallized a construct that corresponds to the N-terminal ATPase domain of P74-26 TerL (residues 1–256; Figure 2.1B). The structure in the absence of nucleotide was determined by single-wavelength anomalous dispersion (SAD)[169] using selenomethionine (SeMet)-labeled protein, and phases were extended to 2.1 Å using diffraction data from native protein (Table 2.1). A second crystal form yielded another structure of the apo P74-26 TerL ATPase domain to 1.9 Å. The structures from each crystal form are very similar (C_α rmsd of 0.8Å; Figure 2.3A and Table 2.2). Unlike the full-length P74-26 TerL, the isolated ATPase domain does not form a pentameric arrangement in either crystal form.

We note three prominent features of the P74-26 TerL ATPase domain (Figure 2.4). First, P74-26 TerL, as in other terminase ATPase domains[38,39], contains a C-terminal subdomain (residues 221–251 in P74-26 TerL) that sits above the ATPase active site. We call this region the “lid” subdomain with reference to the structurally unrelated but analogous lid subdomains found in other ASCE ATPases[46]. Lid subdomains are often used in contacts between adjacent subunits in ASCE oligomers[105,180], and their conformation can be

Table 2.1. Crystallographic statistics

Parameter	Apo-1	Apo-1 R139A	ADP•BeF3	Apo-2
Data collection				
Wavelength, Å	0.9774	1.54	1.100	1.033
Resolution range, Å	45.81–2.097 (2.171–2.097)	28.98–2.532 (2.622–2.532)	38.46–2.068 (2.142–2.068)	34.4–1.931 (2–1.931)
Space group	I 2 3	I 2 3	P 32	P 32 2 1
Unit cell a, b, c	129.578 129.578 129.578	129.578 129.578 129.578	76.92 76.92 131.12	62.89 62.89 133.1
Unit cell α, β, γ	90 90 90	90 90 90	90 90 120	90 90 120
Total reflections	298,214 (29,811)	527,226 (52,149)	247,951 (24,350)	113,291 (10,868)
Unique reflections	21,390 (2,138)	11,395 (1,232)	51,948 (5,228)	23,556 (2,255)
Multiplicity	13.9 (13.9)	43.1 (42.3)	4.8 (4.7)	4.8 (4.8)
Completeness, %	99.99 (100.00)	100.00 (100.00)	98.02 (98.03)	99.68 (97.75)
Mean I/σ(I)	15.80 (2.48)	45.99 (2.59)	14.24 (6.39)	10.71 (2.96)
Wilson B-factor	46.87	62.63	23.19	24.32
R-merge	0.087 (1.275)	0.074 (2.214)	0.058 (0.2278)	0.090 (0.504)
R-meas	0.090 (1.323)	0.075 (2.240)	0.066 (0.258)	0.102 (0.569)
R-pim	0.024 (0.353)	0.011 (0.343)	0.031 (0.120)	0.045 (0.256)
CC1/2	0.999 (0.759)	1.000 (0.952)	0.997 (0.988)	0.997 (0.851)
CC*	1.000 (0.929)	1.000 (0.988)	0.999 (0.997)	0.999 (0.959)
Refinement				
R-work	0.182 (0.263)	0.201 (0.386)	0.201 (0.246)	0.163 (0.214)

Parameter	Apo-1	Apo-1 R139A	ADP•BeF3	Apo-2
R-free	0.211 (0.292)	0.244 (0.440)	0.243 (0.301)	0.206 (0.263)
No. of nonhydrogen atoms	2,221	2,126	6,872	2,351
Macromolecules	2,085	2,044	6,273	2,122
Ligands	35	30	111	20
Water	101	52	490	209
Protein residues	251	250	764	258
rms, bonds	0.009	0.007	0.007	0.009
rms, angles	1.13	0.92	0.99	1.17
Ramachandran favored, %	99	96	98	98
Ramachandran outliers, %	0	0	0.13	0
Clash score	2.88	2.21	3.19	0.47
Average B-factor	57.8	71.4	34.9	29.4
Macromolecules	57.3	70.6	34.1	28.2
Solvent	56.2	68.4	43.7	39.5

- CC, correlation coefficient; I/σ(I), intensity divided by error of intensity; R, residual factor; R-meas, redundancy independent merging R factor; R-pim, precision indicating merging R factor.

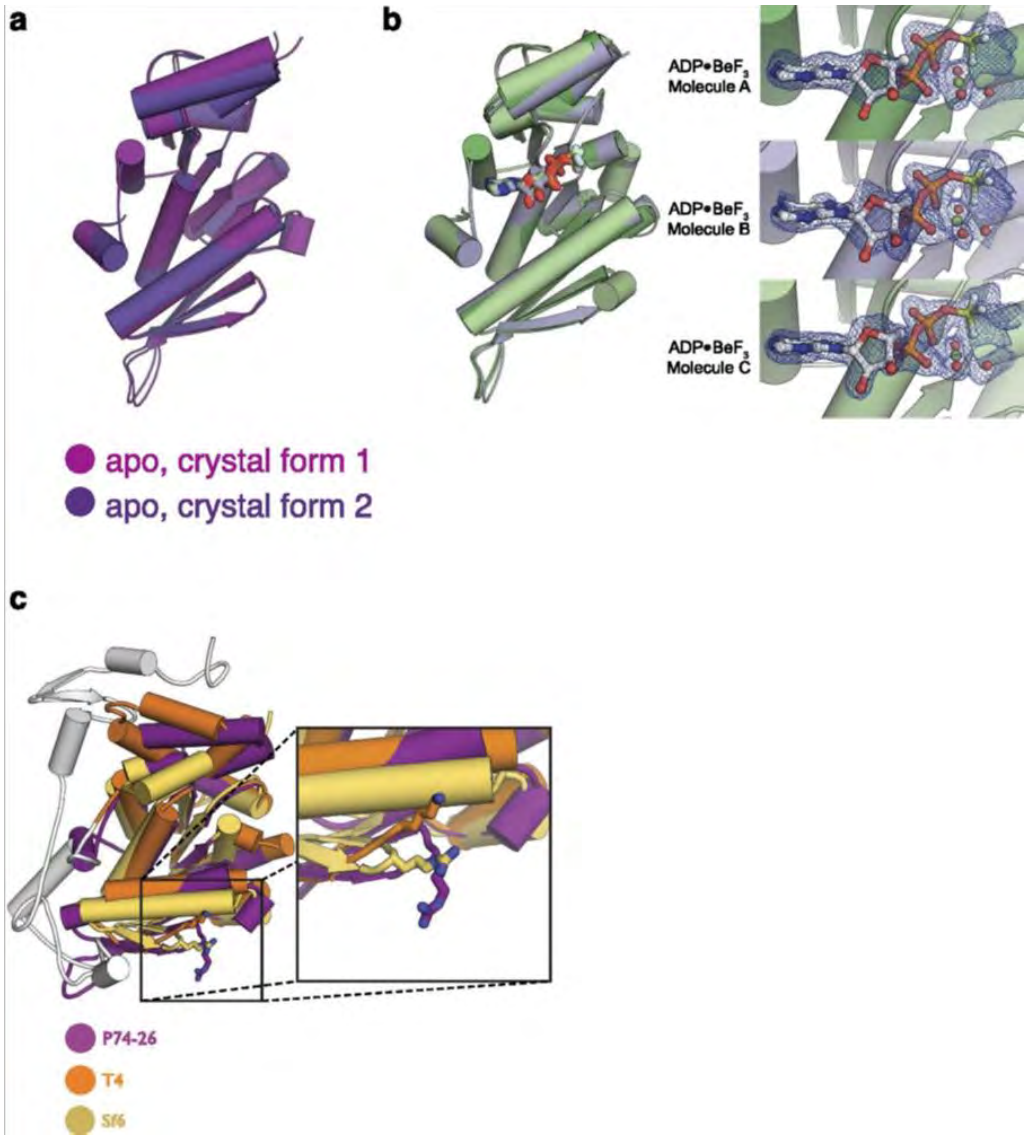

a
b
ADP•BeF3 Molecule A
ADP•BeF3 Molecule B
ADP•BeF3 Molecule C
apo, crystal form 1
apo, crystal form 2
c
P74-26
T4
Sf6

Figure 2.3. Comparison of P74-26 TerL structures. (*A*) Similarity between the two apo structures of the P74-26 TerL ATPase domain (C_α rmsd of 0.8 Å). The structures from both crystal forms are superposed with small deviations at the top of the lid and beta-hairpin loop. These regions make different lattice contacts in the two crystal forms. (*B*, *Left*) Superposition of the three individual ADP•BeF_3-bound ATPase domains shows that there are only minor differences between each monomer. The loop linking beta-strands 4 and 5 displays significant (>1 Å) differences between all three conformers. This loop is longer than seen in the structures of the Sf6 TerL and T4 TerL proteins[38,39], and it is in different packing environments in each monomer in the asymmetric unit. Thus, we conclude that this loop is relatively flexible in P74-26 TerL. (*B*, *Right*) Omit map density for the ADP•BeF_3 ligands, magnesium ions, and waters in each active site. Omit maps contoured at 3σ (0.34 $e^-/Å^2$) are shown for the active sites of all three ATPase domains in the asymmetric unit. Although molecule A has a slightly different electron density in the active site, its overall conformation is essentially the same as in molecules B and C, with a C_α rmsd within active site residues of 0.17 Å and 0.16 Å, respectively. (*C*) Comparison of the ATPase domains of TerL orthologs. TerL ATPase domains from Sf6 (tan) and T4 (orange) are superposed onto P74-26 (purple), with C_α rmsd values of ~2.1 Å and 1.7 Å, respectively. Note that the gray region in T4 TerL is unique to T4 and closely related phages, and was not used for superposition. (*Inset*) Conservation of the P74-26 TerL basic patch. Lys223 in T4 TerL is at the same position as DNA-binding residue Arg101 of P74-26 TerL. Arg82 in Sf6 TerL is at the same position as Arg101 of P74-26 TerL.

Table 2.2. Cα rmsd

	Apo-1	Apo-2	BeF_3-A	BeF_3-B	BeF_3-C	Apo-1 R139A
Apo-1		0.778	1.225	1.197	1.056	0.165
Apo-2			1.055	1.127	0.897	0.784
BeF_3-A				0.526	0.461	1.231
BeF_3-B					0.641	1.119
BeF_3-C						1.053
Apo-1 R139A						

Values are measured in angstroms.

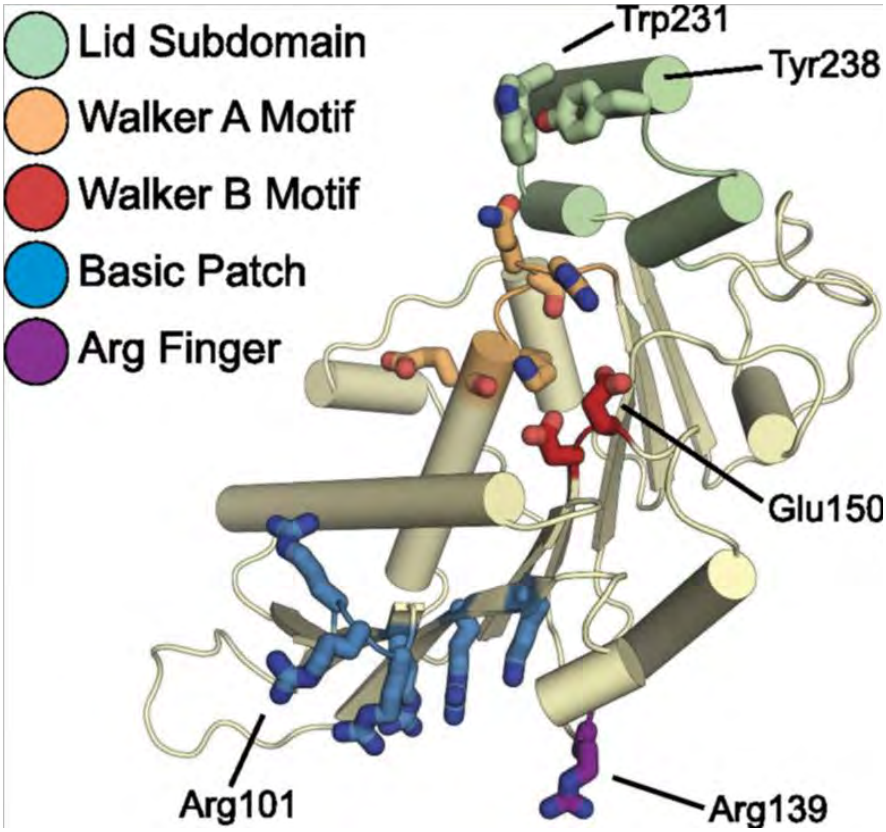


Figure 2.4. Structure of the P74-26 TerL ATPase domain. Characteristics of the TerL ATPase domain are color-coded. The Walker A and Walker B motifs form the active site. The lid subdomain (residues 221–251) is adjacent to the ATP-binding site and contains large hydrophobic residues that are positioned away from the hydrophobic core of the Rossmann fold. The hydrophobic residues Trp231 and Tyr238 are solvent-exposed, suggesting that they could participate in protein–protein interactions. Arg100, Arg101, Arg102, Arg104, Arg128, and Lys130 form a continuous, solvent-exposed basic patch. Arg101 is identified as critical for DNA binding. Arg139, positioned ~25 Å from the active site, is established as necessary for catalysis *in trans*.

modulated by ATP hydrolysis[105,106]. Second, we observe a large patch of six basic residues (R100, R101, R102, R104, R128, and K130) in the turns and strands of the antiparallel section of the beta-sheet, a region that is specific to the terminase ASCE subfamily (6). A basic residue is conserved at the Arg101 position in the other large terminase structures [Lys223 and Arg82 in T4[41] and Sf6[39], respectively] (Figure 2.3C). We hypothesize that this region binds to the DNA backbone during packaging. In support of this hypothesis, we observe a sulfate ion bound to Arg101 in all of our crystal forms. The third region of note is a conserved arginine that is on the opposite face from the active site. We propose that this residue is used to activate ATP hydrolysis in a neighboring subunit in the assembly (discussed below).

To determine the conformational changes induced by ATP binding, we solved a 2.0-Å crystal structure of the P74-26 TerL ATPase domain co-crystallized with the ATP mimic, ADP•BeF_3. The packing of the ADP•BeF_3 cocrystal is unrelated to the apo crystals, with three ADP•BeF_3-bound P74-26 TerL ATPase domains in the asymmetric unit (Figure 2.3B). The residues of the Walker A and B motifs are in the active conformation (Figure 2.5A). The largest conformational change between the apo and ADP•BeF3 states occurs in the lid subdomain. Upon binding the ATP analog, the lid rotates ~13° in a rigid body motion (Figure 2.5B). Upon binding the ATP mimic, the P-loop residues Arg39 and Gln40 rotate to contact the gamma-phosphate mimic directly. The P-loop's

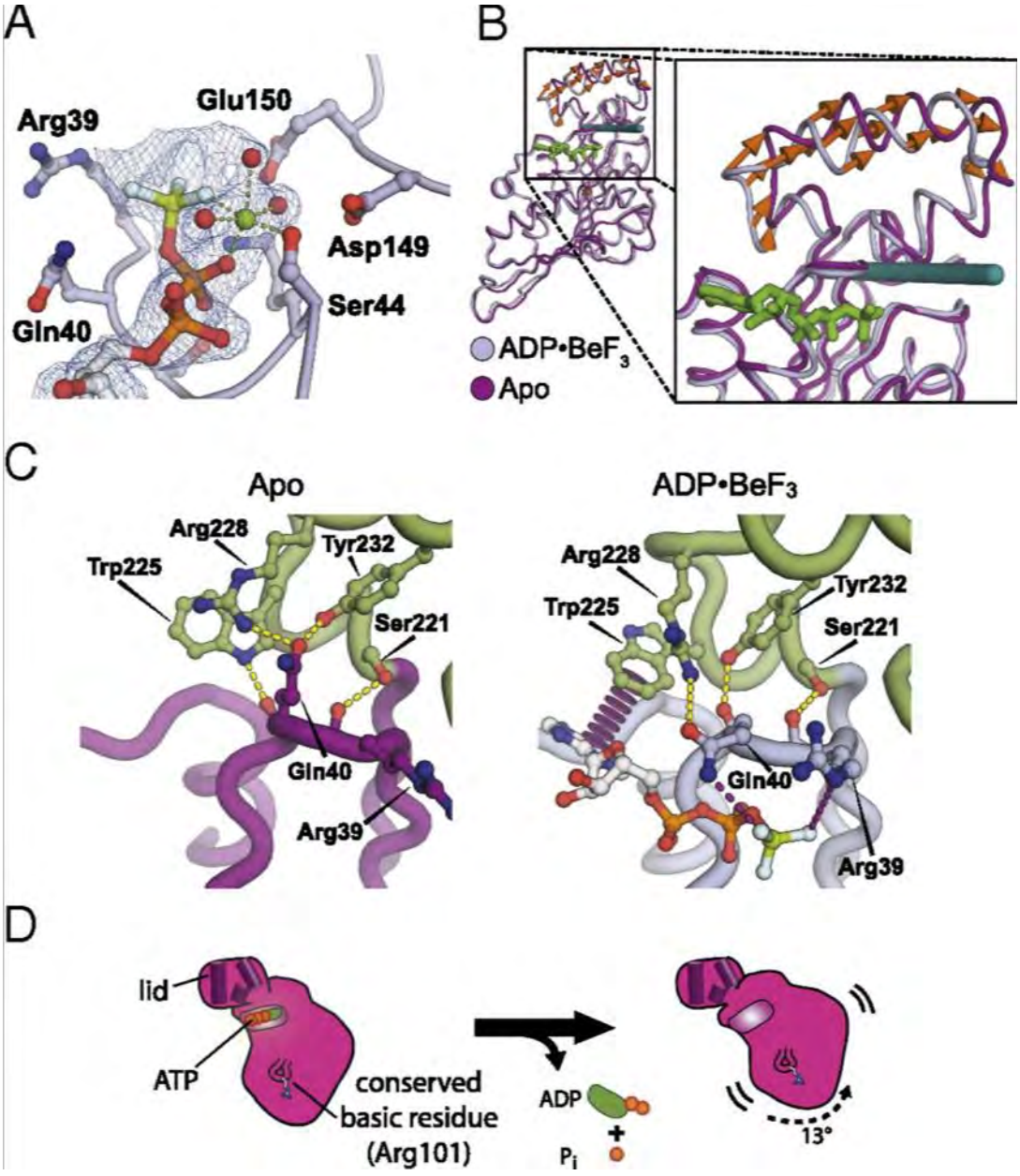

A
Arg39
Glu150
Asp149
Ser44
Gln40
B
ADP•BeF3
Apo
C
Apo
ADP•BeF3
Arg228
Tyr232
Trp225
Ser221
Gln40
Arg39
D
lid
ATP
conserved
basic residue
(Arg101)
ADP
+
Pi
13°

Figure 2.5. Structure of the P74-26 TerL ATPase domain in complex with an ATP analog. (*A*) ADP•BeF_3 is bound in the TerL active site. An omit map density (blue mesh) contoured to 3σ (0.34 $e^{-}/Å^2$) shows ADP•BeF_3, a magnesium ion (green), and waters (red) bound in the TerL active site. Ser44, an ADP beta-phosphate oxygen, fluoride, and three waters coordinate the magnesium ion. Glu150 is pointed toward the BeF_3 moiety to catalyze ATP hydrolysis. (*B*) Conformational changes upon ATP hydrolysis and release. The ATPase domain of the ADP•BeF_3-bound structure (light purple, with ADP•BeF_3 in green) was superposed onto the Rossmann fold of the apo structure (dark purple). The lid was not used for superposition. Vectors (orange arrows) show alpha-carbon position differences of 2.5 Å or greater between the two structures. The teal bar marks the lid's axis of rotation, which lies just off the ATP gamma-phosphate mimic. (*C*) P-loop interactions with the lid (green) appear to drive conformational changes. Hydrogen bonds are shown as yellow dashes. (*Left*) Backbone carbonyls of P-loop residues Arg39 and Gln40 (dark purple) interact with side chains of lid residues Ser221 and Trp225. (*Right*) Arg39 and Gln40 change position to bind ADP•BeF_3, pulling at the base of the lid. The new Gln40 position binds Tyr232, stabilizing the new lid conformation. The Trp225 indole rearranges to stack with the adenine base (purple disks), and Arg39 and Gln40 side chains engage the beryllium fluoride (purple dashes). (*D*) Schematic illustration depicting the conformational change of the lid subdomain in response to nucleotide hydrolysis and release. The lid is held fixed such that the Rossmann fold rotates by ∼13° outward.

interactions with several lid residues (Ser221, Trp225, Arg228, and Tyr232) result in a commensurate rotation of the lid toward the ATP adenine ring (Figure 2.5C). Thus, the motion appears to be induced by the P-loop, which bridges the gamma-phosphate mimic and residues at the base of the lid. The rotation of the lid upon ATP binding results in a reorientation of the lid's hydrophobic residues such that they are now less solvent-exposed, potentially altering TerL's interactions within the motor (Figure 2.5D).

Identification of the Trans-acting Arginine Finger

The structures of TerL ATPases reported here and elsewhere[38,39] display incomplete active sites, presumably because all were crystallized without formation of an active ring. These monomeric TerL structures have only two positively charged residues within the P-loop that contact the gamma-phosphate of ATP to stabilize the transition state for hydrolysis. However, other ASCE family members have at least three positive residues that contact the gamma-phosphate[105,106,158–161,181]. Indeed, our ATPase domain construct lacks ATPase activity (Figure 2.6A), confirming that the active site in our structure is incomplete. In ASCE-type ATPases, the third positive charge is provided *in trans* by a residue known as the "arginine finger," which is most often arginine, although it can also be lysine[46]. The location of the arginine finger within the Rossmann fold varies widely across the ASCE family, which means that ASCE subfamilies often have different relative orientations of their ATPase domains

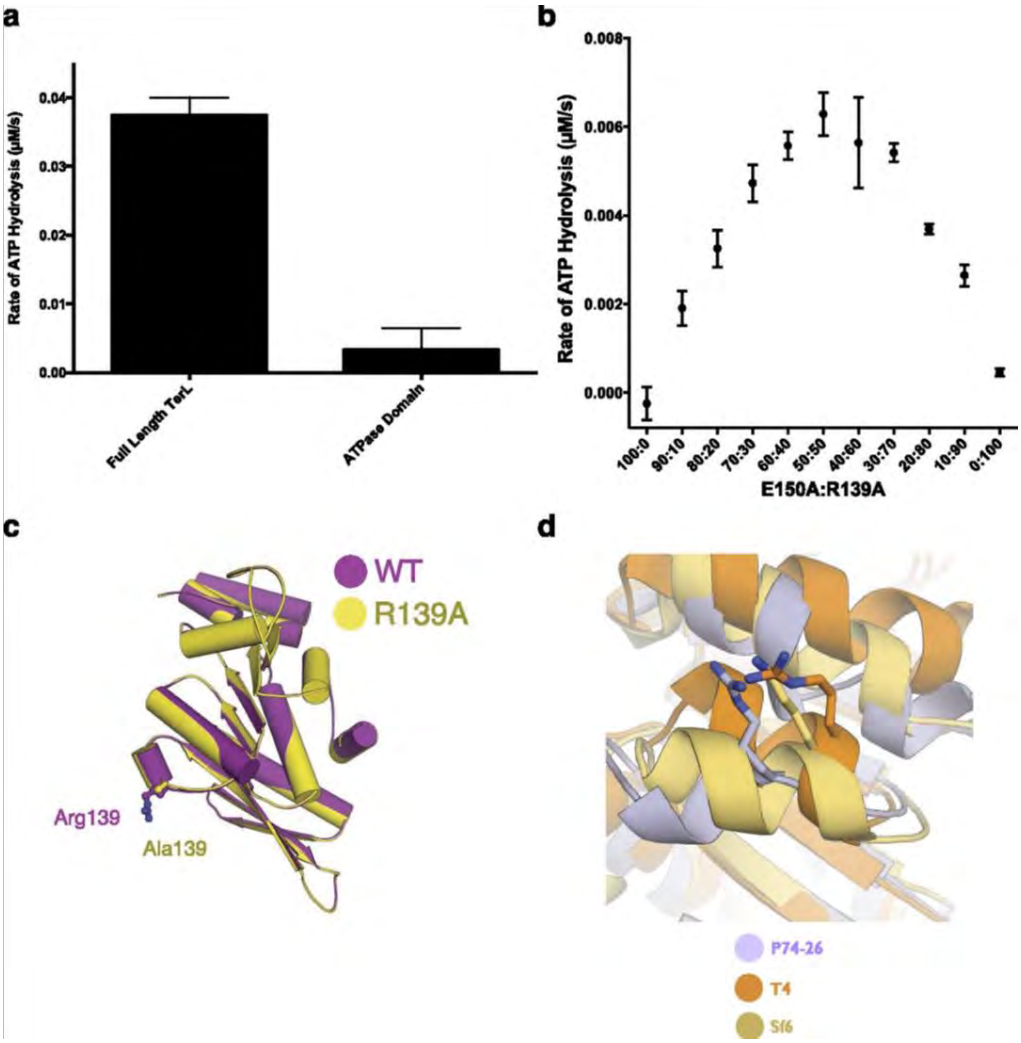


Figure 2.6. Identification of the *trans*-acting arginine finger. (*A*) P74-26 TerL ATPase domain is not catalytically active. Steady-state ATPase activity is shown for 0.5 μM TerL. Error bars are the SD from at least three replicates. (*B*) Mixture of the R139A and E150A single mutants has optimal ATP hydrolysis at a 1:1 ratio. TerL-R139A and TerL-E150A are mixed at the indicated ratios, with a final concentration of 1 μM total TerL, and steady-state ATPase activity is measured. (*C*) R139A mutation does not cause conformational changes in the ATPase domain. The 2.5-Å crystal structure of the R139A P74-26 TerL ATPase domain (yellow) is superposed onto the 2.0-Å structure of the WT structure (dark purple). Position 139 is shown in a ball-and-stick representation for both variants. (*D*) Arginine finger residue is conserved in other TerL ortholog structures. T4 TerL and Sf6 TerL (PDB ID codes 3CPE and 4IEE, respectively) were superposed onto P74-26 TerL, with the arginine finger residue shown in stick representation.

within the oligomer[182]. Thus, the identification of the *trans*-acting arginine finger in TerL is necessary both for elucidation of the ATPase mechanism and for determining the architecture of the TerL ring assembly.

We used a two-step strategy to identify the arginine finger of P74-26 TerL. First, we screened mutants of surface-exposed arginines in full-length P74-26 TerL for ATPase activity. For mutants that showed no ATPase activity in this initial scan, we tested whether an inactive P74-26 TerL protein (mutated Walker B motif) could restore activity by donating its intact arginine finger.

We measured steady-state ATPase activity for 11 different arginine mutants in full-length P74-26 TerL (Figure 2.7A). The R39A, R139A, and R235A variants completely abrogate ATPase activity, making them candidates for the arginine finger. Of these candidate residues, Arg39 and Arg235 reside within or near the active site, suggesting that they may affect catalysis *in cis*. Conversely, Arg139 is located ~25 Å away from the gamma-phosphate of ATP (Figure 2.4).

To assess which of our candidates is the arginine finger, we developed a biochemical complementation assay. We mixed each of our three candidate arginine finger mutants with the E150A variant whose active site is disabled *in cis* (Figure 2.7B) but can still donate an intact arginine finger to an adjacent subunit in the assembly. Thus, E150A should restore activity to an arginine finger mutant, but not to a mutant whose ATPase activity is disabled *in cis*. Mixing E150A with R139A results in restoration of ATPase activity, whereas mixtures of E150A with R39A or R235A exhibit no significant ATP hydrolysis (Figure 2.7B). ATPase

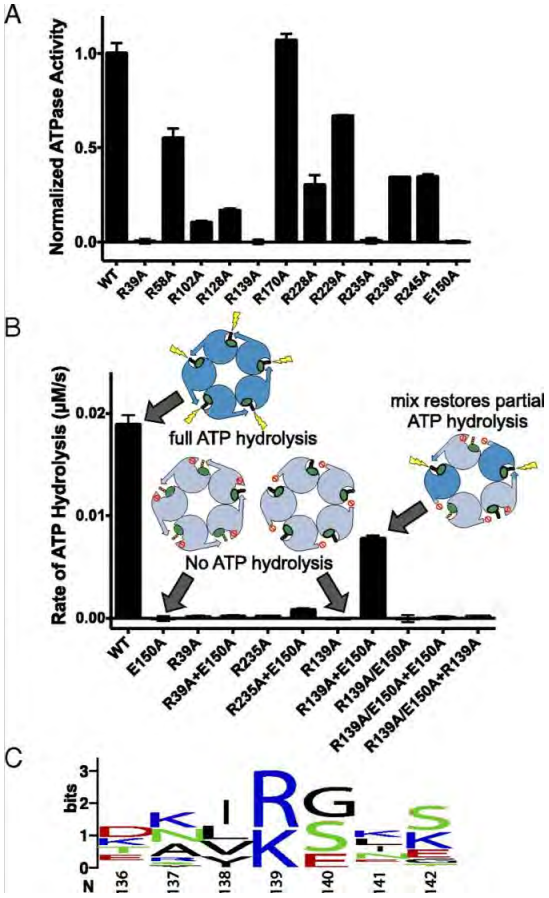


Figure 2.7. Identification of the TerL arginine finger residue. (*A*) ATP hydrolysis rates of P74-26 TerL arginine mutants. The only arginine mutants that abolish ATPase activity are R39A, R139A, and R235A (compare activity with activity of the inactive E150A mutant). The final TerL concentration is 0.5 μM. For all relevant panels in this figure, error bars are the SD from at least three replicates. R39 is a conserved P-loop residue proposed by Sun et al.[41] to be a *cis*-acting arginine finger. (*B*) Arginine finger complementation assay. Different TerL variants at 0.5 μM each are mixed either with an alternate TerL mutant in a 1:1 ratio (as indicated by the "+" sign) or with buffer, and their steady-state ATPase activity is measured. Note that the R139A/E150A double mutant is a single variant and not a mixture. Only a mixture of the R139A and E150A mutants significantly restores activity. Schematic illustrations illustrate that the Walker B mutant (E150A) has a debilitated active site but can donate its intact arginine finger (shown as an arrow), whereas the arginine finger mutant has an intact active site but lacks ATPase activity due to loss of the *trans*-acting residue. (*C*) Arginine finger residue is conserved in other TerL ortholog sequences. A logo diagram[183] was made from a sequence alignment of 70 TerL proteins. The residue numbering is shown for P74-26 TerL.

activity is maximal at a 1:1 ratio of E150A to R139A (Figure 2.6B), as expected for complementation of an arginine finger mutant. In contrast, a double mutant combining both the R139A and E150A mutations in the same protein has no activity, and cannot be complemented by either the E150A or R139A single mutant. Our complementation results mirror the results of similar experiments from Cox et al.[184] definitively establishing that the ASCE family member RecA uses trans-interactions to catalyze ATP hydrolysis.

We verified that the R139A mutation does not perturb ATPase domain structure by solving the 2.5-Å resolution structure of the isolated ATPase domain with the R139A mutation. This structure is essentially identical to the WT structure (Cα rmsd = 0.2 Å; Figure 2.6C), suggesting the observed deficit is not due to structural changes *in cis*. In support of this finding, we observe that arginine or lysine is conserved at this position in the large terminase family (Figures 2.7C and 2.6D). Hence, we expect that the arginine finger is playing a similar role throughout the family of large terminases. Taken together, our results indicate that Arg139 is the arginine finger, which mediates ATP hydrolysis in TerL *in trans*. Our identification of the arginine finger illustrates that (i) ATP hydrolysis is catalyzed in trans and (ii) the previously proposed structural model[41] must be substantially refined to account for these trans-interactions within the TerL ring.

Modeling of the TerL Ring

Our two key results, the P74-26 TerL ATPase domain structure and identification of the arginine finger residue, allow us to create a model for the TerL ring assembly. The structure provides a basis for molecular docking, and the arginine finger provides a spatial restraint that must be satisfied for the TerL assembly. We used the program M-ZDOCK[178,185] to model the TerL ATPase ring. Fivefold symmetry was the only constraint applied. We used the *trans*-arginine finger interaction to validate the docking results independently. Although many ATPase rings are asymmetrical during function[158,186], we assume fivefold symmetry for simplicity. Regardless, most ringed ATPase structures were first modeled as symmetrical assemblies[181,187], and these models were refined later to show asymmetry during function[113,157,158,186]. After docking of the ATPase ring, we used the full-length crystal structure of T4 TerL[41] to orient a homology model of the P74-26 nuclease domain, which results in minimal steric clashes between the two domains. The soft energetic potential used in M-ZDOCK results in slightly shorter interatomic distances[178], and therefore a constricted ring; thus, we treat the molecular docking results in a qualitative fashion.

Molecular docking produced a TerL assembly that is consistent with our identification of the arginine finger. Arg139 contacts the ATP gamma-phosphate in a neighboring subunit, despite no restraint for this interaction imposed during the calculation (Figure 2.8A and B). Thus, these models satisfy a critical spatial interaction identified by our biochemical results. The lid subdomains interact with an adjacent subunit, with the nuclease domains arranged radially (Figure 2.8C).

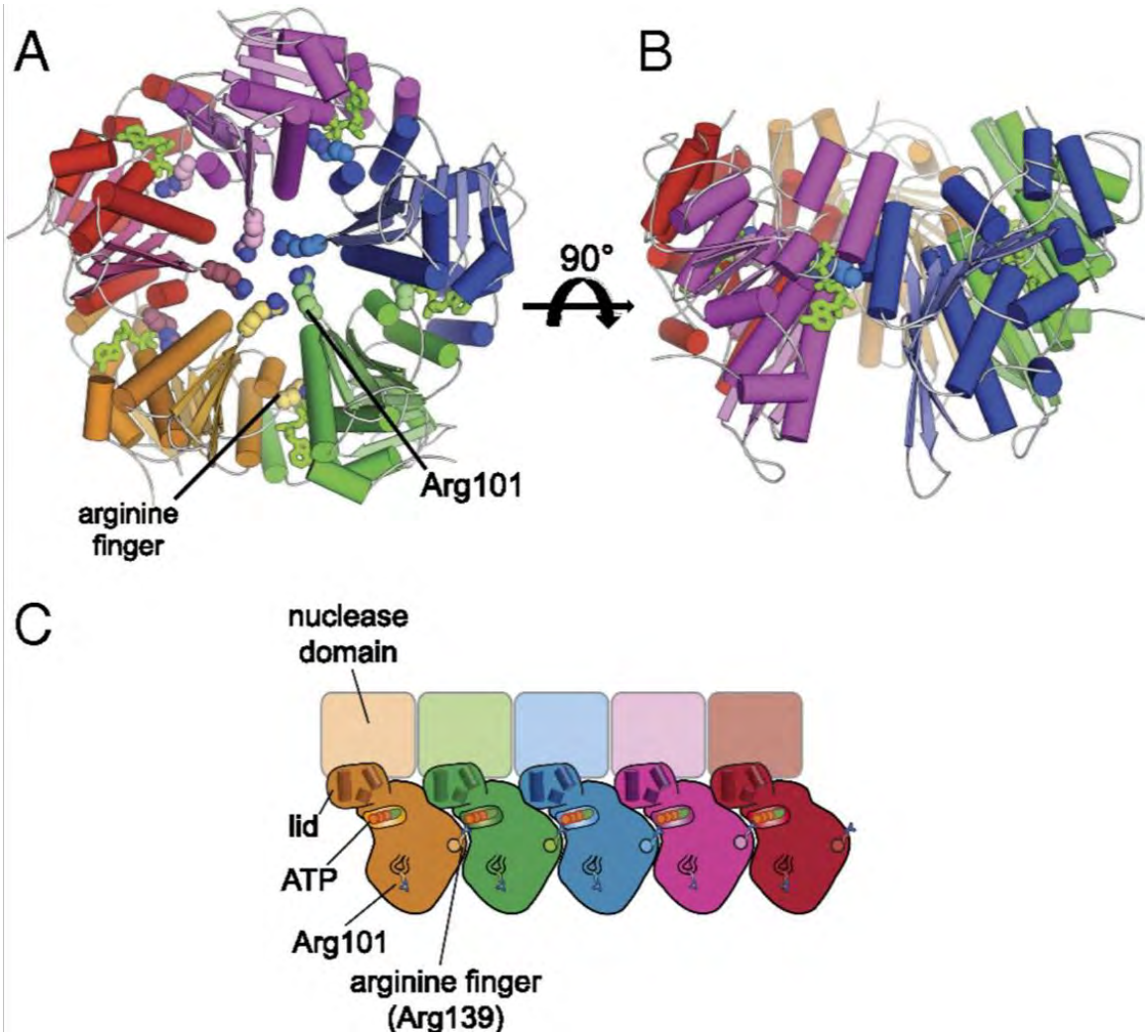


Figure 2.8. Derivation of the P74-26 TerL ring assembly structure. (*A*) Model of the P74-26 TerL ATPase ring derived from symmetry-constrained docking. A bottom-up view shows each of the five subunits in different colors, with the ATP ligand in green. The arginine finger and the conserved basic site (Arg101) are shown as colored spheres. (*B*) Side view of the P74-26 TerL ATPase ring derived from symmetry-constrained docking. Colors and the image are produced as in *A*. (*C*) Schematic depiction of the TerL ring. The TerL ring is shown from the perspective of the pore, with the ring artificially "broken" and flattened on the page so that all subunits can be viewed simultaneously. Each subunit is colored as shown in other panels in this figure. The nuclease domains are depicted as translucent squares to represent the ambiguity of the placement of the nuclease domain within the TerL ring. The loops protruding from the sides of each subunit represent the conserved basic patch (Arg101 in P74-26 TerL). The gamma-phosphate of ATP is contacted by the arginine finger from a neighboring subunit.

We tested our model's plausibility by comparing our TerL assembly with other TerL crystal structures and a cryo-EM reconstruction. First, orthologous TerL ATPase crystal structures[39,41] can be identically positioned without significant steric clashes of the ATPase domains (Figures 2.9A and C). In fact, the N-terminal extension that is unique to T4 is placed on the outside of the ring rather than lining the pore of the ring as in the previous proposal[41]. Moreover, the pores of our Sf6 TerL and T4 TerL ATPase models are positively charged (Figures 2.9B and D). Second, our structural model predicts that for other TerL proteins that are monomeric in isolation, ring assembly would be enhanced by ATP mediating cross-subunit interactions. Indeed, this dependency has been observed for T4[188], T3[189], and lambda[190]. Third, our T4 TerL ring model is a bowl-like assembly that reasonably fits the cryo-EM map[41] for an actively packaging T4 phage (Figure 2.9E and F). However, this fitting positions the ATPase domain ring distal to the capsid, whereas the nuclease domains interact with portal (Figure 2.9E S4E), an interaction that has been observed recently[113]. Thus, our revised TerL ring model is a different arrangement than that previously proposed.

We examined whether our structural model is globally consistent with a molecular envelope, as calculated from small-angle X-ray scattering (SAXS) data. Whereas WT P74-26 TerL exhibits slight aggregation, we obtained high-quality SAXS data by exploiting the R104E P74-26 TerL mutant, a more soluble variant than the WT protein (Figure 2.10A and B). R104E P74-26 TerL is well folded and has a radius of gyration of 48.0 ± 2.0 Å (Figure 2.10B and C),

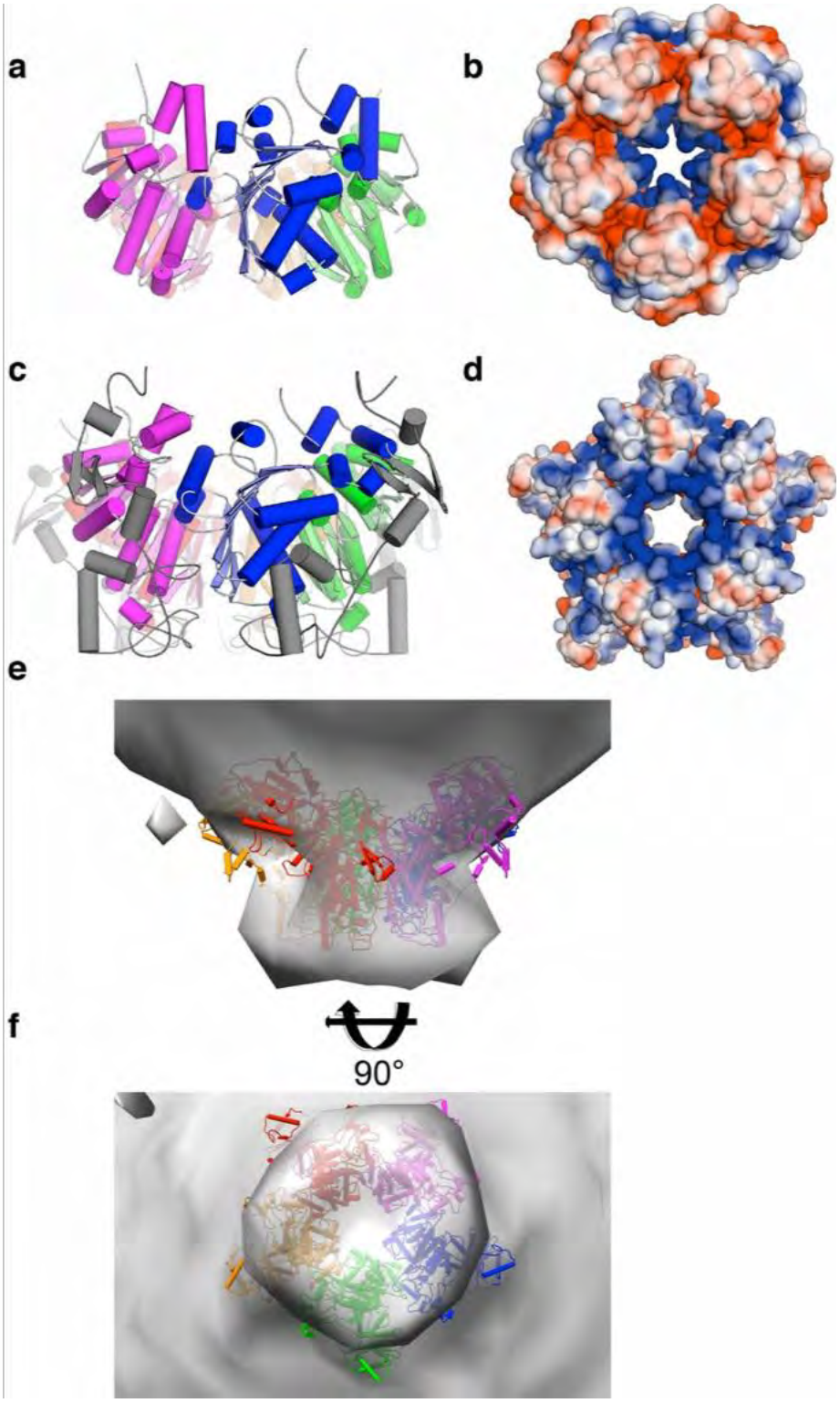

a
b
c
d
e
f
90°

Figure 2.9. Superposition of TerL orthologs onto the P74-26 TerL ATPase ring model. (*A*) Model of the Sf6 TerL ATPase ring. The TerL ATPase domain from Sf6 was superposed onto the docked model of the P74-26 TerL ATPase ring. (*Left*) Each subunit is shown in a different shade, and the ring is viewed from the side. There are no significant clashes in this model. (*B*) View of the electrostatic surface potential of the Sf6 TerL ring from the bottom shows a net positively charged pore as would be expected from the DNA-binding region. (*C*) Model of the T4 TerL ATPase ring. The TerL ATPase domain from T4 was superposed onto the docked model of the P74-26 TerL ATPase ring. The region specific to the T4 phages is shown in gray, the main ATPase domain is shown in various colors, and the ring is viewed from the side proximal to the nuclease domains. There are no significant clashes in this model, and the T4-specific region is present on the outside of the ring, away from the core machinery. (*D*) Electrostatic surface potential of the T4 TerL ring pore, as viewed from the bottom of the ring. Again, the residues within the pore have a net positive charge, as would be expected for a region that binds DNA. (*E* and *F*) Model of the T4 TerL ring fits a cryo-EM structure of the intact motor (*E* is viewed from the side and *F* from the bottom). The revised model of the T4 TerL ring is shown as a colored cartoon, with the cryo-EM envelope of the T4 capsid and packaging motor, which was calculated without fivefold averaging (gray surface).

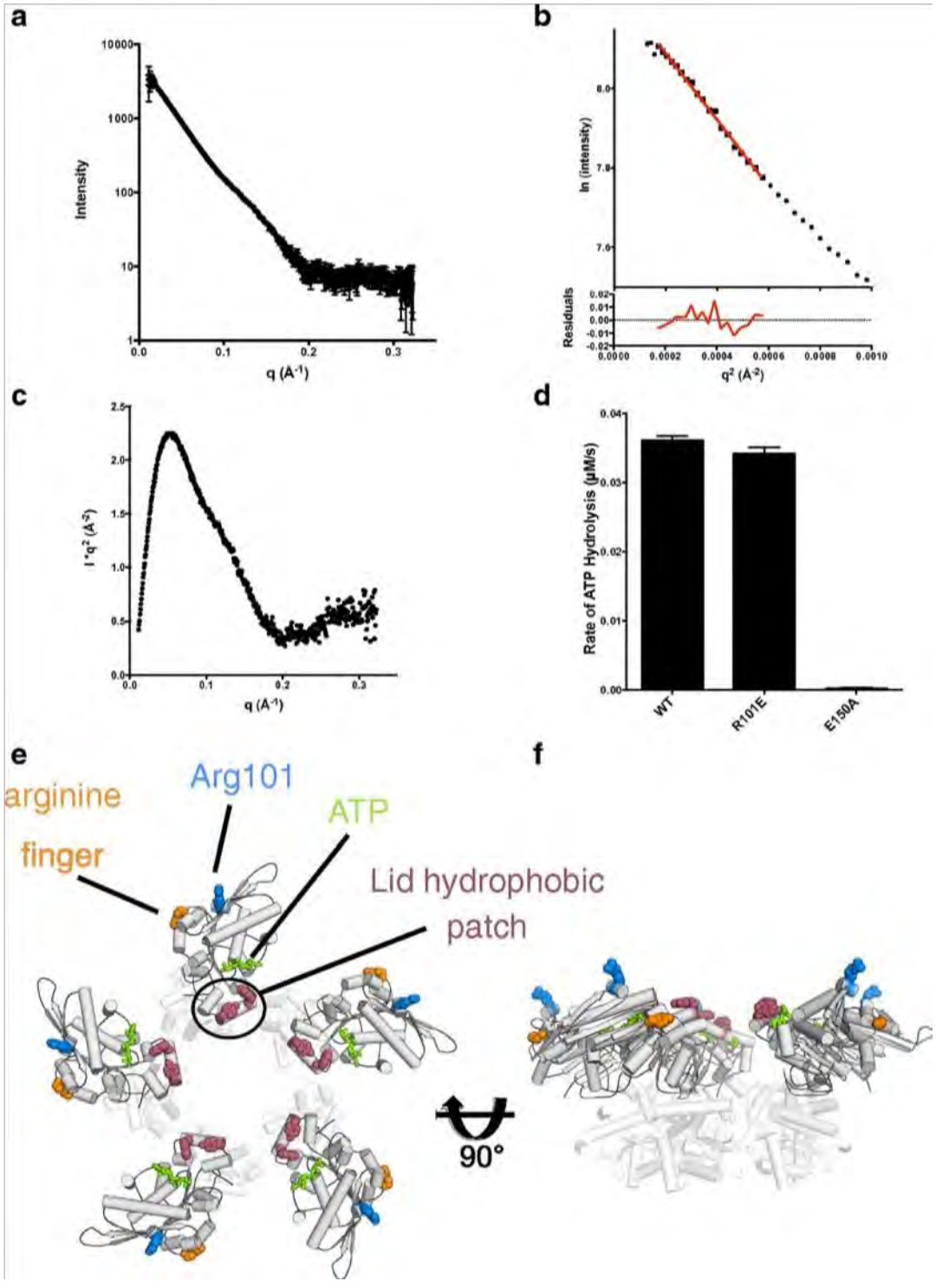

a
Intensity
q (Å⁻¹)
b
ln (intensity)
Residuals
q² (Å⁻²)
c
I*q² (Å⁻²)
q (Å⁻¹)
d
Rate of ATP Hydrolysis (μM/s)
WT
R101E
E150A
e
arginine finger
Arg101
ATP
Lid hydrophobic patch
90°
f

Figure 2.10. Functional and biophysical analysis of the P74-26 TerL pentameric assembly. (*A*) SAXS profile of a highly soluble variant of P74-26 TerL (R104E). Scattering intensity is shown as a semilog plot vs. the scattering angle (q). (*B*) Guinier plot of low scattering angles shows no aggregation, and a radius of gyration of 48.0 ± 2.0 Å. (*C*) P74-26 TerL construct is well-folded. A Kratky plot of the SAXS data is shown. The peak at low scattering angles and the relatively flat and low profile at high scattering angles are hallmarks of well-folded protein. (*D*) P74-26 TerL R101E variant is catalytically active. Steady-state ATPase activity is shown for 0.5 μM TerL. E150A and WT are shown as points of comparison. Error bars are the SD from at least three replicates. (*E*) Key functional residues mapped onto the previous model for the TerL assembly. The view is as seen from the perspective of the capsid/portal. The P74-26 TerL ATPase domain crystal structure and nuclease domain homology model were superposed onto the previous model for the T4 TerL ring (PDB ID code 3EZK). Key residues for testing each model are shown as colored spheres: The arginine finger (Arg139) is orange, the DNA-binding residue (Arg101) is blue, and the lid hydrophobic patch residues (W231 and Y238) are shown in rusty red. The nuclease domain homology model is semitransparent. (*F*) Same as in *E*, but viewed from the side. The procapsid and portal would be above the TerL ring.

consistent with a pentameric assembly. We performed an *ab initio* calculation of the P74-26 TerL molecular envelope using our SAXS data and imposing fivefold symmetry. The calculated molecular envelope fits the scattering data well ($\chi 2$ = 0.79) and is roughly shaped like a bowl, with a pore in the center and a V-shaped cross-section. The overall bowl-like shape of the P74-26 TerL ring model fits the SAXS reconstruction well; the ATPase domains form the V-shaped base, and the nuclease domains form the bowl's lip (Figure 2.11A). We note that extra density is present between the nuclease domains, possibly indicating conformational heterogeneity in the bowl's lip or a poor fit of our nuclease domain homology model. Regardless, our model is broadly consistent with reconstructions from both SAXS and cryo-EM.

We tested various aspects of our structural model to determine whether it has predictive power. Our model predicts that the lid's hydrophobic patch (Trp231 and Tyr238) contacts an adjacent subunit for arginine finger positioning (Figure 2.11B). To test the hypothesis that these residues are important for ATPase activity, we individually mutated Trp231 and Tyr238 to alanine. Mutation at each of these residues results in severe loss of ATPase activity (Figure 2.11C). Thus, the lid's hydrophobic patch is critical for TerL activity.

As an additional test of our model's predictive power, we used our model to identify the DNA-binding motif. Basic residues line the pore of the TerL ring, with a central pore residue (Arg101) that is conserved (Figure 2.11D and Figure 2.3C). A positive electrostatic environment in the pore is similar to other ASCE

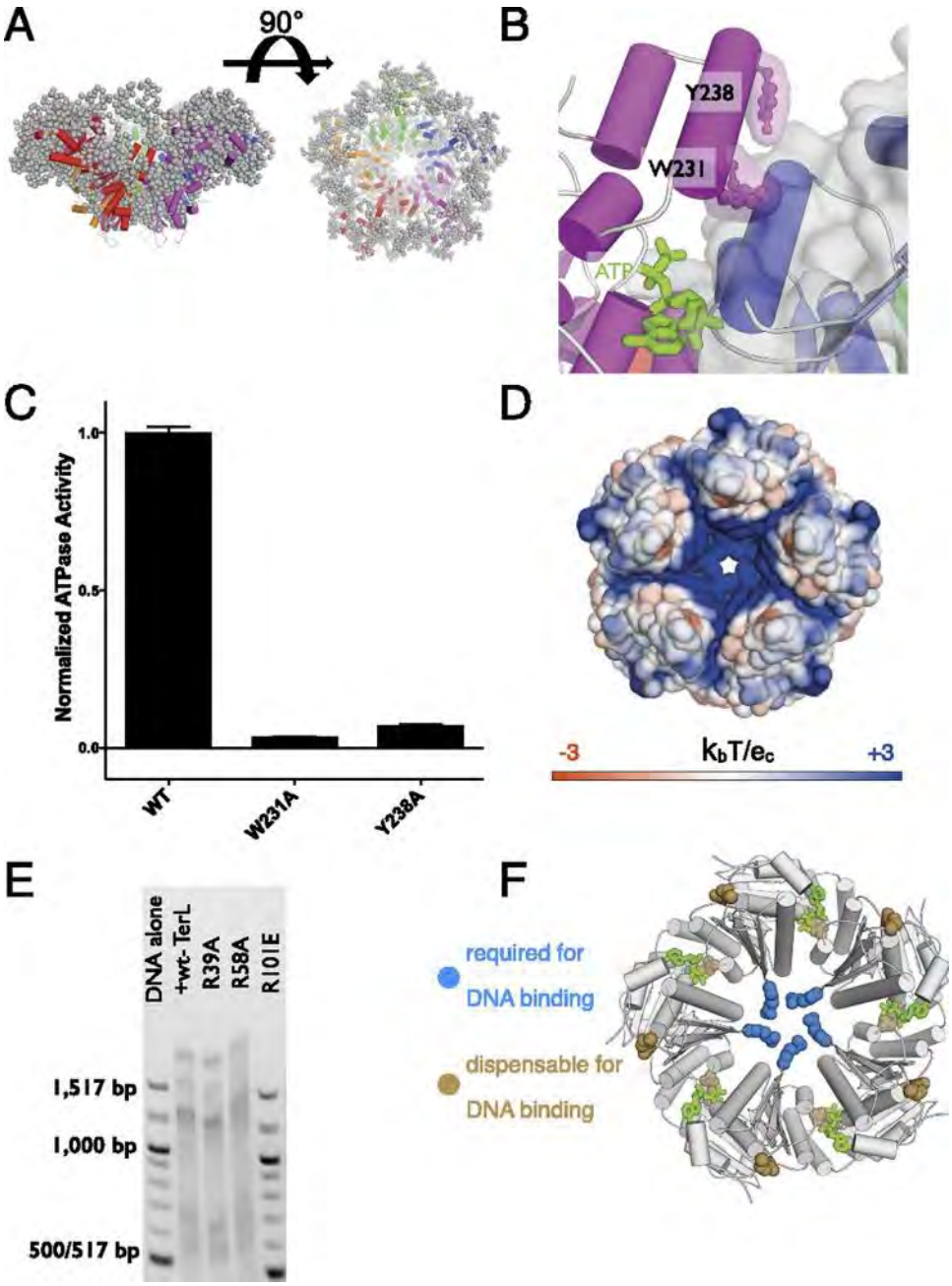

A
90°
B
Y238
W231
ATP
C
Normalized ATPase Activity
1.0
0.5
0.0
WT
W231A
Y238A
D
-3
kbT/ec
+3
E
DNA alone
+wt- TerL
R39A
R58A
R101E
1,517 bp
1,000 bp
500/517 bp
F
required for
DNA binding
dispensable for
DNA binding

Figure 2.11. Validation of the TerL ring assembly model. (*A*) Model of P74-26 TerL fits the *ab initio* SAXS envelope. The modelled P74-26 TerL ring structure (colored cartoon) is superposed onto the SAXS envelope calculated using GASBOR[191] (imposing fivefold symmetry). Dummy atoms for the SAXS envelope are shown as gray spheres. (*B*) Hydrophobic patch on the lid subdomain mediates critical intersubunit interactions. The structural model suggests that Trp231 and Tyr238 of the lid hydrophobic patch (ball-and-stick representation) mediate critical interactions with a neighboring subunit for positioning the arginine finger. (*C*) P74-26 TerL lid hydrophobic patch is critical for TerL function. Steady-state ATPase activity is shown for 0.5 μM TerL. Mutation of either W231 or Y238 results in a large decrease in ATPase activity. Error bars are the SD from at least three replicates. (*D*) Electrostatic map of the docked P74-26 TerL ATPase ring, with positive and negative surface potentials shown in blue and red, respectively. Note the positive charge lining the pore of the P74-26 TerL ATPase ring. The electrostatic surface was calculated using the APBS plug-in for PyMOL (DeLano Scientific)[192]. Figure is colored by electrostatic potential (k_bT/e_c) as indicated. (*E*) DNA binding for multiple arginine mutants. EMSA was carried out as in Fig. S1*F*, including several arginine mutants. Arg101 is required for DNA binding, whereas Arg39 and Arg58 are dispensable. (*F*) Mutation of the conserved pore arginine abrogates DNA binding. The ability of three different arginine mutants to bind DNA tightly in the EMSA assay was mapped onto the model of the P74-26 TerL ring. As predicted by the structural model, Arg101 is necessary for DNA binding and mutations distal to the pore (R39A and R58A) have no measurable effect on DNA binding. A 100-bp DNA ladder (New England BioLabs) is used as a standard.

nucleic acid translocases[158,159,161]. We hypothesize that Arg101 binds DNA during packaging. To test this hypothesis, we mutated several surface arginines in full-length P74-26 TerL and measured the ability of these mutants to interact with DNA in the presence of ADP•BeF_3 (Figure 2.11E). Mutation of Arg101 causes loss of DNA binding, whereas mutation of residues distal to the pore (Arg39 and Arg58) has no measurable effect (Figure 2.11F). Although a defect in binding ATP could cause loss of DNA binding, the R101E mutant retains WT levels of ATPase activity (Figure 2.10D), indicating that the DNA-binding defect is not due to loss of ATP binding. Thus, we have identified the critical DNA-binding motif in TerL, supporting our structural model for the TerL ring.

Discussion

Our results suggest a substantially different organization of the TerL ring than the initial model that was proposed previously in a ground-breaking study of the T4 terminase. The previous model, which is based on fitting the T4 TerL crystal structure to a cryo-EM reconstruction with a resolution of 34 Å, proposed that the TerL ATPase domains form a ring that contacts portal through the lid subdomains[41]. The pore of the ATPase ring has a net negative charge and is lined by a portion of TerL that is unique to T4. The nuclease domain was proposed to form a ring distal to portal that grips and translocates DNA through the pore. Because the ATPase active sites do not contact a neighboring subunit,

Sun et al.[41] proposed that an invariant arginine in the P-loop (Arg162 in T4 TerL, Arg39 in P74-26 TerL) acts as the arginine finger to catalyze ATP hydrolysis.

Our identification of the conserved *trans*-acting arginine finger indicates that the previous model requires significant modification. Although the P-loop arginine previously proposed as the *cis*-acting arginine finger is necessary for packaging[193] and ATPase activity[41] (Figure 2.7A), our data do not support the hypothesis that it is the arginine finger. Instead, we propose that this residue, Arg39 in P74-26 TerL, is conceptually analogous to the sensor II arginine found in the family of ATPases known as the ATPases associated with diverse cellular activities (AAA+), which aids ATP hydrolysis *in cis* and confers movement of the AAA+ lid subdomain upon ATP hydrolysis[105,106]. In order for the previous model to accommodate the *trans*-acting arginine finger, a rotation of the ATPase domain of ~110° is necessary. Although formally possible, this rotation of the ATPase domain does not fit the cryo-EM density for the actively translocating motor[41] and would disrupt portal interactions. Therefore, we disfavor this model.

Our docked model for the TerL ring satisfies the distance constraint imposed by the arginine finger and revises the orientation of TerL relative to portal. We propose that the nuclease domains form a radially arranged ring that is proximal to portal, whereas the ATPase ring is distal. Our model positions the lid subdomains at the interface between adjacent ATPase subunits, where they assist in positioning the arginine finger, as supported by our mutagenesis data (Figure 2.11C). In the previously proposed model[41], the lid residues interact with

portal (Figure 2.10E and F). However, because our sample lacks portal, the observed ATPase defects in lid mutants are not due to a disruption of portal interactions, but are consistent with our proposed role in TerL assembly.

Our updated model accounts for functional conservation across the TerL family. First, our model contains a largely basic patch lining the pore. Within this patch, we identify Arg101 as a critical component of the DNA-binding motif. With conservation of a basic residue at this position, our model is congruent with the terminase family. Notably, in the previously proposed structural model, the equivalent residue in T4 TerL is positioned distal to the pore and the nuclease domain was hypothesized to bind DNA[41] (Figure 2.10E and F), both of which are inconsistent with our observations. Second, our identification of the arginine finger in TerL brings the mechanism of ATP hydrolysis into accord not only with the terminase family but also with the rest of the ASCE family. Conservation of the DNA-binding and catalytic mechanism therefore supports our updated structural model.

Although our model allows accurate predictions for regions of function, the overall model is qualitative in nature. As previously mentioned, the ring dimensions are slightly constricted due to the docking algorithm, with the pore's smallest inner diameter measuring ~16 Å, as calculated from the C_β positions of surface-exposed arginines lining the pore (we use the C_β position because it is rigidly fixed). Thus, the modeled pore is too small to accommodate dsDNA. We propose that the TerL ring is expanded relative to the docked model because the

docked model exhibits contacts that are clearly too close with several overlapping atoms, which would artificially constrict the ring. Furthermore, our model positions the nuclease domain based on the full-length T4 TerL structure, which crystallized as a monomer[41]. The nuclease domain position may be ortholog-dependent or may be altered upon ring formation and/or upon portal binding. Nuclease position is important because the nuclease domain may be playing a key role in pentamer assembly, considering that the isolated P74-26 TerL ATPase domain crystallizes as a monomer. Although ongoing and future investigations will refine the details of our structural model, we have established that it captures essential aspects of the TerL ring through its consistency with prior data, as well as its predictive power. Our structural model, combined with our nucleotide-dependent structural changes, allows us to derive a preliminary mechanistic model for DNA translocation in this family.

Based on our TerL ring model and the observed conformational changes upon ATP hydrolysis and release, we propose a mechanistic model for DNA translocation. During genome packaging, DNA is gripped in the center of the TerL ring. In this model, it is assumed that the TerL ring hydrolyzes ATP one subunit at a time, and not in a concerted all-or-none mechanism. This assumption is consistent with the mechanism of ATP hydrolysis in other ringed ASCE ATPases[58,158,159,161]. When one subunit hydrolyzes ATP, it undergoes a conformational rearrangement such that the lid pivots 13° around the Rossmann fold (Figure 2.5B). If we further assume that the lid remains bound to the

neighboring subunit throughout packaging, as seen in homologs[194], the result would be the Rossmann fold pivoting outward and upward toward the capsid (Figure 2.12).

We propose that the conformational change from one ATP hydrolysis event propagates to an adjacent subunit, sterically exerting force on the adjacent subunit such that the two subunits move in concert. Because TerL only binds DNA tightly in the ATP-bound form (Figure 2.2F), ATP hydrolysis at one subunit will lead to that individual subunit releasing DNA. During a hydrolysis event, the ATP-hydrolyzing subunit loses its grip on DNA but initiates motion that is propagated to the adjacent ATP-bound subunit, which is still gripping DNA through interaction with Arg101. The conformational change of the ATP-hydrolyzing subunit results in an upward translocation motion of DNA at the adjacent subunit and resets the motor for unidirectional translocation. We estimate that this motion would translocate DNA perpendicular to the plane of the ring by ~8 Å per hydrolysis event, or about 2.4 bp. In addition, we predict that DNA would rotate in the plane of the ring by ~2.3° for each step. Because our model is qualitative, our estimated step rotation and size should be viewed with reservation. However, these values compare favorably with the 2.5 bp translocation[58] and ~3.5° rotation per step[66] measured for the phi29 motor at low packaging force. Our proposed “lever-like” mechanism for force generation is in contrast to the previous model, wherein a “spring-like” motion of DNA-bound nuclease domains translocates DNA through the pore[41]. Because several other

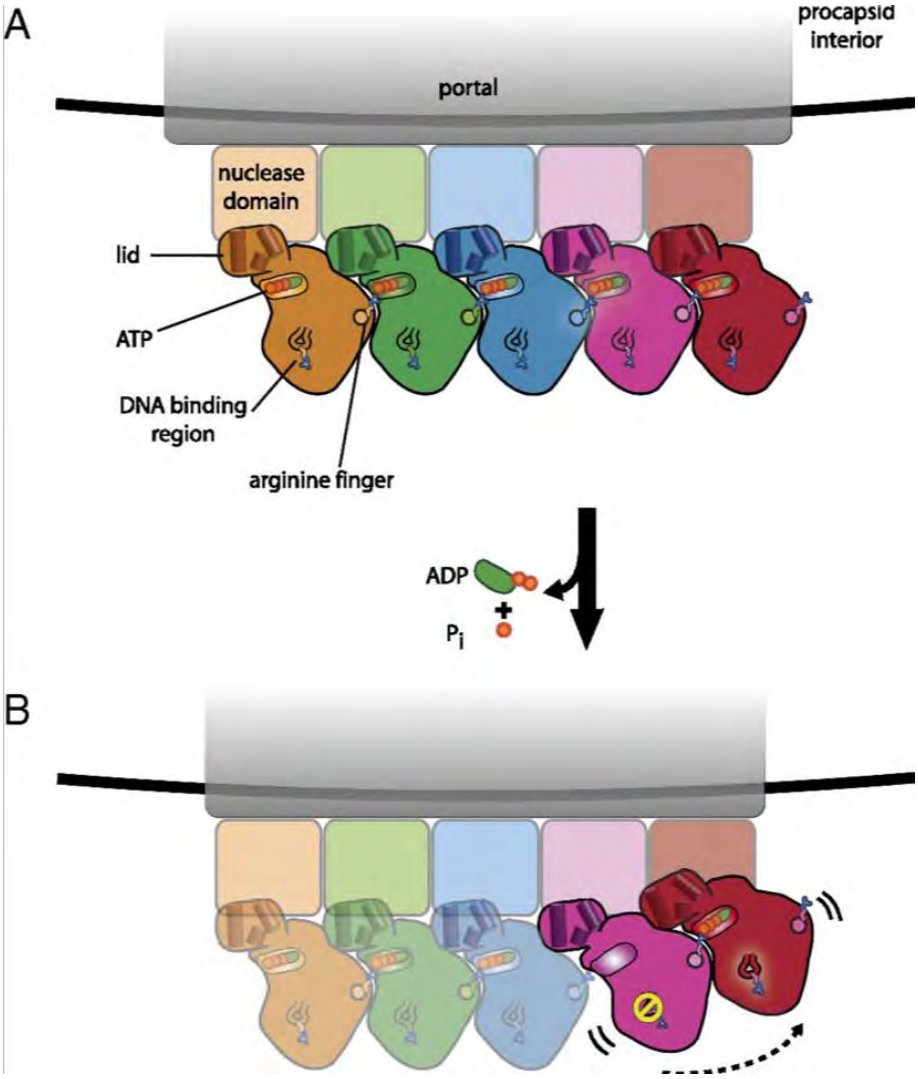


Figure 2.12. Proposed model for the mechanism of DNA translocation by TerL. (*A*) TerL ring is shown as in Fig. 6*F*. The nuclease domains interact with the portal complex (translucent gray rectangle) and possibly the procapsid (black curve). DNA is not shown but interacts with TerL through the DNA interaction motif (Arg101). Each subunit's lid is bound tightly to the Rossmann fold of the adjacent subunit. (*B*) Upon ATP hydrolysis and release by the magenta subunit, the lid stays bound to the blue subunit and the Rossmann fold rotates 13° upward. To allow for this movement, the adjacent red subunit must also move in concert with the magenta subunit. To represent that the second site of symmetry breaking is unknown, the other three ATPase domains are faded. After hydrolyzing ATP, the magenta subunit releases DNA to the red subunit to translocate DNA upward through the pore; into the pore of the portal complex; and, ultimately, inside the procapsid. The release of DNA at each cycle by the ATP-hydrolyzing subunit allows for unidirectional DNA translocation.

studies have evaluated DNA packaging in the context of the previous TerL structural and mechanistic models, our work illustrates that these studies should be reinterpreted within the context of the updated TerL model.

It is unclear how this conformational change will affect subunits further downstream than the two moving subunits. The breakage of symmetry at the ATP hydrolysis site necessitates at least one other site of symmetry breaking to maintain a closed ring structure. Alternatively, the TerL assembly may form an open lock washer shape that dynamically alternates which subunits cap the ends of the lock washer, as has been proposed for the ASCE helicase DnaB[158]. Although our current model is fivefold symmetrical, recent studies of ASCE family members illustrate that significant asymmetry exists during motor function[158,159,186]. Future refinement of our structural model will identify how TerL asymmetry drives DNA translocation.

This work was previously published in PNAS[37].

CHAPTER III

The cryoEM structure of a thermophilic small terminase protein

Abstract

Many tailed bacteriophage use a DNA packaging motor to encapsulate their genetic material during viral maturation. An important step of the packaging cycle is genome recognition, which is performed by the ‘small terminase’ subunit (TerS). TerS binds viral DNA and transfers it to the large terminase (TerL) for packaging. While there are several crystal structures of TerS proteins, the DNA and TerL binding mechanisms of the protein are still unclear. Here, we present a novel model system for studying TerS DNA and TerL binding. We identify the TerS protein of the thermophilic bacteriophage P74-26 as gp83. We determine P76-26 TerS oligomerizes into a stable nonamer (9-mer), which binds DNA, stimulates TerL ATPase activity, and inhibits TerL nuclease activity. Using cryoEM, we solve 3.8 Å and 4.8 Å resolution reconstructions of the symmetric and asymmetric TerS complexes. The cryoEM structures show P74-26 TerS forms a nonameric ring with a central pore wide enough for DNA accommodation. Interestingly, P74-26 TerS also contains an inflexible helix-turn-helix motif, and appears to lack the conserved C-terminal β-barrel domain in the TerL binding region. Together, our findings suggest the putative TerL binding region in P74-26 TerS is unstructured, and further studies will determine which TerS regions bind DNA.

Introduction

During their life cycle, viruses undergo a multi-step reproduction process. Viruses infecting all domains of life, from bacteria to eukaryotes, must replicate and encapsulate their genetic material to create infectious particles. For viruses with large genomes, transporting genetic material into the capsid is an energetic challenge, and many viruses have evolved motor systems to accomplish this task. In viruses with concatemeric double-stranded DNA genomes, a motor known as a 'terminase motor' is used. Terminase motors are composed of three components: a 'portal' channel, a 'small terminase' DNA recognition protein, and a 'large terminase' enzymatic motor[195]. During maturation, the portal, which is embedded within the capsid wall, acts as an adaptor to connect the capsid to the large terminase. The large terminase (TerL) binds portal and pumps DNA through its pore into the capsid. In order for this packaging step to occur, the motor must first recognize and cleave the viral genome concatemer to form a free end of the DNA to thread into portal. This DNA-recognition task is performed by the small terminase (TerS), which binds a recognition sequence known as '*pac'* and transfers the DNA to TerL for subsequent cleavage and packaging. TerS has an important role in initiation, as aberrant *pac* recognition impedes faithful genome packaging [196,197].

Despite being studied for several decades, how TerS binds to *pac* is still unclear. In many viral genomes, the *pac* site is located within the gene for TerS itself[50,78,81,82,198,199]. In some systems, such as SPP1, the *pac* site appears to

have intrinsic flexibility, suggesting a role for DNA bending in TerS recognition[199]. Further clues for the DNA binding mechanism come from structures of TerS proteins. In all currently known *pac* recognizing TerS proteins, the protein multimerizes to form a ring with a central pore[49,50,86,87]. In some of these assemblies, such as *Shigella flexneri* phage Sf6 and *Bacillus subtilis* phage SF6, the pore is too narrow to accommodate double-stranded DNA binding[86,87]. Interestingly in these structures, the outward-facing N-terminal domain is a helix-turn-helix motif, a common DNA-binding motif. Studies of Sf6 TerS indicate that mutation of this region of the protein abrogates DNA binding, suggesting a nucleosome-like wrapping mechanism[85]. The exception to this model is the TerS structure of phage P22. In P22, the perimeter of the ring lacks the helix-turn-helix motif, and the pore is wide enough to accommodate DNA[50]. This finding led to the second hypothesis that DNA binds in the center and is threaded through the pore.

Despite the differences in the DNA binding region of the protein, TerS rings retain the same mushroom-like shape with a C-terminal β-barrel. The β-barrel is of interest, as it is conserved in all TerS structures to date, and interacts with TerL[50,88]. TerS binding to TerL increases TerL's ATPase activity while inhibiting nuclease activity[42,43,49,50,78], indicating TerS has a regulatory effect on DNA packaging. Additionally, the β-barrel can control TerS assembly, as removing it causes polydisperse ring formation[49,87]. Therefore, the C-terminal β-

barrel has been hypothesized to be important for both TerS oligomerization and regulation of TerL activity.

In past studies, we have used the thermophilic phage model system P74-26 to probe the mechanisms behind different stages of the viral life cycle[37,116,200]. Here, we identify the small terminase gene of phage P74-26 and develop it as a model system. We characterize the complex as a nonameric ring that binds DNA, activates P74-26 TerL ATPase activity, and inhibits TerL nuclease activity. We report symmetric and asymmetric cryo-EM reconstructions of P74-26 TerS to overall resolutions of 3.8 Å and 4.8 Å resolution, respectively. Our structures show that P74-26 TerS retains the N-terminal helix-turn-helix motif, while also having a wide enough pore for DNA binding. In comparison to other phages, the helix-turn-helix motif is in an alternate conformation, suggesting P74-26 TerS binds DNA in a different manner compared to other TerS proteins. Most interestingly, the C-terminal region of P74-26 TerS is unstructured, indicating that the β-barrel seen in other TerS structures is not strictly conserved, nor is a β-barrel essential for regulating P74-26 TerL and stabilizing P74-26 TerS ring stoichiometry.

Materials and methods

Cloning

The TerS P74-26 gene was synthesized with codon optimization for expression in *E. coli* by Genscript Corporation. The gene was cloned into the BamHI and

NdeI sites of a modified pET28a vector with an N-terminal His6-T7-gp10 expression tag and a Prescission protease cut site. Enzymes were purchased from New England BioLabs. Oligonucleotides were purchased from IDT.

Expression and purification of proteins

Protein was expressed in BL21-DE3 cells containing the pET28a-TerS plasmid. Bacterial cultures were grown at 37°C in Terrific Broth supplemented with 30 μg/ml kanamycin until an OD_{600} of 0.7 was reached. Cells were moved to 4°C for 20 minutes, after which expression was induced by addition of IPTG (isopropyl-b-D-thiogalactopyranoside) to 1 mM. Cells were then returned to an 18°C incubator to shake overnight. Cells were pelleted and resuspended in 'Buffer A' (500 mM NaCl, 20 mM Tris pH 7.5, 20 mM imidazole and Roche cOmplete EDTA-free Protease Inhibitor Cocktail dissolved to a final concentration of 1x). Resuspended cells were flash frozen in liquid nitrogen for long-term storage at -80°C. Thawed cells were lysed using a cell disrupter, and lysis was pelleted via centrifugation. Cleared lysate was filtered using a 0.45 μM filter. All subsequent steps occurred at room temperature unless noted. Lysate was loaded and recirculated over Ni-affinity beads (Thermo-Scientific) for 2.5 hours, which had been pre-equilibrated with Buffer A. Beads were subsequently washed with 5 column volumes of Buffer A without protease inhibitors. The protein-bound beads were transferred to a 50 mL conical containing 1.25 mg of purified prescission protease, which was incubated overnight on a nutator. The following day, the resin was transferred to

a gravity flow column, and the flow-through was collected, alongside a 1 column volume wash of the resin with Buffer A. The flow-through was then concentrated and injected onto a HiPrep 26/60 Sephacryl S200-HR gel filtration column that had been pre-equilibrated with gel filtration buffer (250 mM NaCl, 20 mM Tris pH 7.5) at 4°C. Fractions corresponding to the TerS peak were pooled, concentrated to 17 mg/mL, and flash frozen in liquid nitrogen for storage at -80°C. P74-26 TerL was expressed and purified as previously described in Chapter II[37].

Size Exclusion Chromatography Multi-Angle Light Scattering (SEC-MALS)

SEC-MALS was performed at room temperature using a 1260 Infinity HPLC system (Agilent), a Dawn Helios-II multi-angle light scattering detector (Wyatt Technology), and an Optilab T-rEX differential refractive index detector (Wyatt Technology). Detectors were aligned, corrected for band broadening, and photodiodes were normalized using a BSA standard. Samples were diluted to 1 mg/mL with Gel Filtration buffer and filtered through a 0.22 µM filter. 50 µL of sample was injected onto a WTC-030S5 size exclusion column with a guard (Wyatt Technology) that had been pre-equilibrated overnight with Gel Filtration buffer. Data analysis was performed with Astra 6 software (Wyatt Technology).

DNA binding

TerS DNA binding was performed using the P74-26 gp83 DNA sequence that was PCR amplified from the P74-26 phage genome. P74-26 forward primer:

ATGAGCGTGAGTTTTAGGGACAGGG; P74-26 reverse primer: CTAGGTCTTAGGCGTTTCATCCGCC. Oligonucleotides were purchased from IDT. To assess DNA binding, TerS was dialyzed into a buffer containing 25 mM potassium glutamate and 10 mM Tris pH 7.5. TerS was then incubated for 30 minutes with 50 ng of the P74-26 gp83 gene in an 8 µL volume sample. After incubation, 2 µL of 5x Orange G loading dye was added to the samples, yielding the final protein concentration indicated on the gel. Samples were run on a 1% (wt/vol) TAE-agarose gel with a 1:10,000 dilution of GelRed dye (Phenix Research) for 90 minutes at 80 volts.

Coupled-enzyme ATPase assays

ATPase experiments were performed as previously described in Chapter II[37].

Nuclease assays

Nuclease experiments were performed as described in Appendix 1[116].

Negative stain electron microscopy

3.5 µL of 900 nM TerS (monomer) was applied to a glow-discharged carbon-coated 400 mesh copper EM grid and incubated for 30 seconds. Sample was blotted off, and the grid was washed with water and blotted two times. Grid was stained with 1% uranyl acetate and imaged using a 120kV Philips CM-120

electron microscope with a Gatan Orius SC1000 detector. Relion 2.0 was used for 2D classification[201].

CryoEM specimen preparation

For dataset one, 400 mesh 2/2 Holey Carbon C-Flat grids (Protochips) were incubated with ethyl acetate until dry. Grids were glow discharged for 60 seconds at 20 mA (negative polarity) with a Pelco easiGlow glow discharge system (Pelco). Samples were prepared to yield a final concentration of 19.5 μM TerS (nonamer), 150 mM NaCl, 20 mM Tris (pH 7.5), and 0.015% amphipol A8-35. For dataset two, the same sample was applied to a 200 mesh 2/2 UltrAuFoil Holey Gold grid (Quantifoil) that was glow-discharged for 120 seconds at 20 mA. For both datasets, 3 μL of sample was applied to grid at 10°C and 95% humidity in a Vitrobot Mark IV (FEI). Samples were blotted for 4 seconds with a blot force of 5 after a 10 second wait time. Samples were then vitrified by plunging into liquid ethane and were stored in liquid nitrogen until data collection.

CryoEM data collection

Micrographs were collected on a Titan Krios electron microscope (FEI) at 300 kV fitted with a K2 Summit direct electron detector (Gatan). Images were collected at 130,000x in super-resolution mode with a pixel size of 0.529 Å/pixel and a total dose of 50 e-/Å2 per micrograph. Micrographs were collected with a target defocus range of -1.4 to -2.6 for both datasets one and two. Dataset one was

collected with one shot focused on the center of the hole. For dataset two, the first 549 images were collected with four shots per hole at 0° tilt, and the remaining 1,077 images were collected at a 30° tilt with two shots per hole. In total, 2,822 micrographs were collected.

CryoEM data processing

Micrograph frames were aligned using the *Align Frames* module in IMOD with 2x binning, resulting in a final pixel size of 1.059 Å/pixel. Initial CTF estimation was performed using CTFFIND[202] within the cisTEM suite. Particles were picked with a characteristic radius of 40 Å using *Find Particles* in the cisTEM software package[203]. Particles were then extracted with a largest dimension of 120 Å and a box size of 256 pixels. Selected particles were subjected to 7 rounds of 2D classification using cisTEM. Each round of 2D classification consisted of 20 iterative cycles with 50 to 100 classes. After each round, the classes were examined and noisy classes were excluded before subjection to the next round of classification. The final round of 2D classification yielded 295,395 particles, which were exported into Relion format. *Ab-initio* 3D reconstruction was performed with cisTEM using a particle subset selected for an even distribution of views from the 2D classification images. *Ab-initio* 3D reconstruction was performed using 2 starts with 40 cycles per start. CTF correction was re-estimated using gCFT[204] and the particles were re-extracted in Relion 3.0[205]. Classification was done in Relion 3.0 using C1 symmetry into 6 classes for 60

iterations with a mask diameter of 140 Å. For the first asymmetric structure, classes 1 and 6 were combined (152,315 particles) for 3D refinement in Relion 3.0 using C1 symmetry. For the symmetric reconstruction, class 1 (84,860 particles) was sub-selected for 3D refinement in Relion 3.0 using C9 symmetry. For the second asymmetric structure, class 5 (86,969 particles) was sub-selected for asymmetric refinement using C1 symmetry. CTF refinement and subsequent post-processing were performed after 3D refinement for all symmetric and asymmetric reconstructions in Relion 3.0. Resolution was calculated using gold-standard FSC curve calculation and a cutoff of 0.143.

CryoEM model building

To build the atomic models of the TerS structure, the helix-turn-helix motifs and oligomerization domains of the g20c crystal structure (PDB code 4XVN) were rigid body fit into the cryoEM density for each subunit separately using the Chimera "*Fit to map*" command[177]. Each chain in the symmetric and asymmetric models consisted of residues 1 to 137. For the symmetric structure, one chain was manually refined in Coot[206], and 9-fold symmetry was repopulated using PyMol. For the class 5 asymmetric structure, the symmetric model was fit into the density and each helix-turn-helix motif and oligomerization domain were separately fit in Coot using the "rigid body refine" tool. Model refinement was performed in Phenix using the real-space refinement tool with three cycles of refinement per round. Rotamer restraints, Ramachandran restraints, and NCS

restraints were used during refinement. Group ADP values were calculated on a per residue basis.

Multibody refinement

Multibody refinement was performed on the C1 refined structure of classes 1 and 6. Separate masks were generated around a single helix-turn-helix domain and the remainder of the protein. The largest body was designated as body 1, and the helix-turn-helix motif was designated as body 2. Body 1 was set to refine with a rotational value of 0° and a translational value of 0 pixels in relation to body 2. Body 2 was set to refine with a rotational value of 20° and a translational value of 5 pixels in relation to body 1.

Focused classification

For focused classification, the C1 refined structure of classes 1 and 6 were used. Symmetry expansion was performed using the *relion_particle_symmetry_expand* program in Relion 3.0[205] to rotate each particle 40° nine times to align all monomers within the ring into the same position. Separate masks were created for a single TerS subunit and the remaining eight subunits. Particle subtraction was implemented using *Particle Subtraction* in Relion 3.0 to remove the eight remaining subunits, and focused classification was performed with the masked single subunit into 4 classes.

Local resolution and electrostatic potential calculation

Local resolution calculation was performed using the “Local resolution” function of Relion 3.0. Electrostatic maps were generated using the PyMol APBS plugin[192].

Secondary structure prediction

Secondary structure of TerS monomers was predicted using JPred4[207].

Circular Dichroism

TerS and TerL samples were dialyzed overnight at 4°C into a buffer containing 125 mM NaCl and 25 mM Tris (pH 8.5). Each protein was diluted to 1 μM in an identical buffer containing 1 mM 2-mercaptoethanol. Samples were placed in a double-chamber cuvette in which the pathlength of each chamber was 0.438 cm. Samples were read from 215 to 260 nm with a scan rate of 50 nm per min and a response time of 8 seconds with a Jasco J-810 circular dichroism spectrophotometer (Jasco, Inc). Each sample was read in triplicate, and an averaged spectra was generated. Samples were then mixed and read again. The spectra were converted to mean residue ellipticity (MRE) with the equation

$$\frac{(\Theta \times 100{,}000)}{[conc] \times (l) \times (n)}$$

Results

Characterization of P74-26 gp83

We sought to identify and characterize the P74-26 phage TerS to illuminate how thermophilic small terminase proteins recognize the viral genome. In related phage genomes, the small terminase gene directly precedes the large terminase gene. The P74-26 large terminase was previously identified as gp84[208], leading to the prediction that gene 83 encodes the small terminase protein. The gp83 protein is 171 amino acids (a.a) in length, and is thus similar in size to known TerS proteins, such as P22 (162 a.a), T4 (164 a.a), SF6 (145 a.a) and Sf6 (140 a.a). To further verify its identity, the putative TerS protein was recombinantly expressed and purified to near homogeneity (Figure 3.1A). Size-exclusion multi-angle light scattering (SEC-MALS) shows gp83 assembles into a stable 9-mer complex, with a measured molecular mass of 170 kDa (compared to 171 kDa calculated by sequence) and a polydispersity index of 1.000, indicating a monodisperse assembly (Figure 3.1B). The oligomerization state of gp83 is consistent with that of mesophilic TerS proteins, which assemble into 8 to 11 subunit oligomers[49,50,86,87].

To determine if gp83 binds DNA like other TerS proteins, we performed electromobility shift assays. Because many other TerS oligomers recognize a sequence within their own gene[50,78,81,82,198,199], we used the P74-26 gp83 DNA sequence to evaluate DNA binding. The gp83 complex binds DNA weakly, as indicated by smearing within the gel (Figure 3.1C). Although the K_d for DNA binding cannot be accurately quantified, low DNA binding affinities were

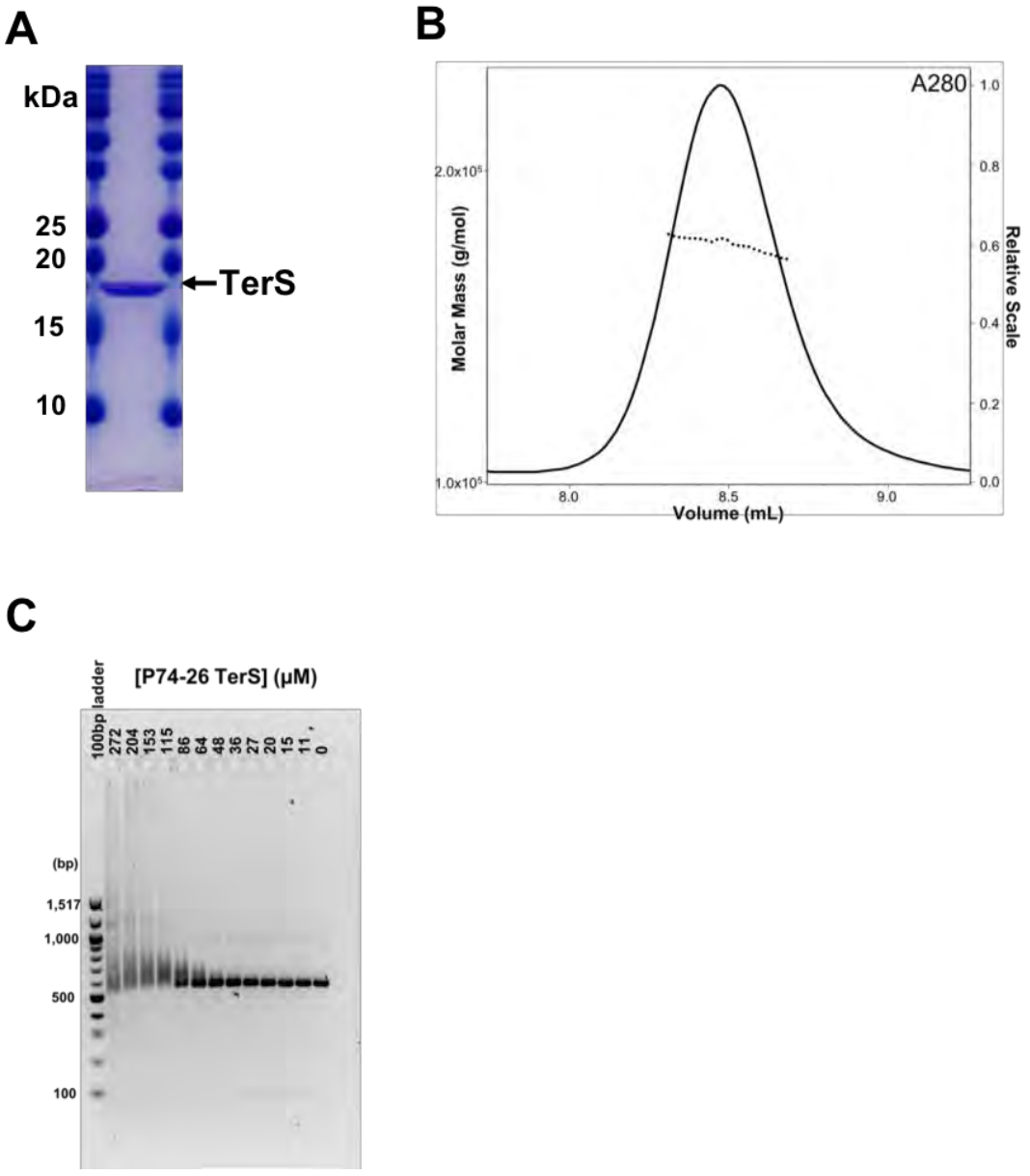


Figure 3.1. Characterization of P74-26 TerS. (A) SDS-PAGE gel of purified P74-26 TerS. (B) SEC-MALS of P74-26 TerS. The UV absorbance at 280 nm wavelength is shown. The measured molecular mass of the complex is 170 kDa, compared to 171 kDa calculated from sequence of a 9-mer. The polydispersity index is 1.000. (C) P74-26 TerS binds DNA with weak affinity. Titrating P74-26 TerS from 0 to 272 μM (monomer) with 50 ng of the P74-26 TerS gene shows TerS has a low affinity for DNA.

observed for other TerS proteins[84,85], indicating gp83 displays typical affinity for DNA.

We also find that gp83 regulates the enzymatic activities of TerL [42,43,49,50,78]. Upon mixing gp83 with P74-26 TerL, TerL ATPase activity increases 4.4-fold (Figure 3.2A). This suggests a direct interaction between TerL and gp83, as no DNA is present in the experiment. Previous studies also established this, with T4 TerS reportedly increasing T4 TerL ATPase activity up to 50-fold[42]. As seen with other systems, gp83 also inhibits TerL nuclease activity 3.3-fold[43,49,91] (Figure 3.2B). Taken together with the genetic, SEC-MALS, DNA binding, and ATPase activity, these results verify gp83 is the TerS of P74-26.

The cryo-electron microscopy structure of P74-26 TerS

We next used electron microscopy (EM) to determine the structure of P74-26 TerS. Negative stain EM shows homogenous TerS particles with even distributions of top and side views (Figure 3.3). From 2D classification, we observe that TerS forms a donut-shaped assembly with a central pore. To further elucidate the structural features of P74-26 TerS, we prepared samples of the complex for cryoEM. Unlike negative stain samples, cryoEM samples show preferred orientation for the top and bottom views of the ring and slight aggregation (Figure 3.4A). The lack of side views severely hampers initial structure determination, and the middle portion of the ring cannot be resolved (Figure 3.4B).

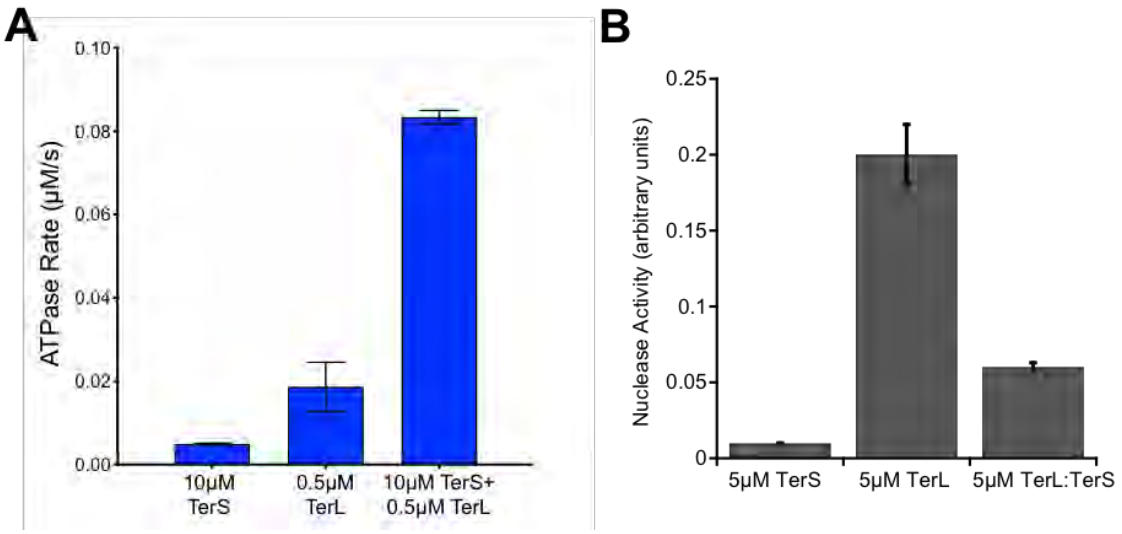


Figure 3.2. P74-26 TerS regulates the enzymatic activity of P74-26 TerL. (A) P74-26 TerS increases ATPase activity of P74-26 TerL 4.4-fold, (B) P74-26 TerS decreases P74-26 TerL nuclease activity 3.3-fold.

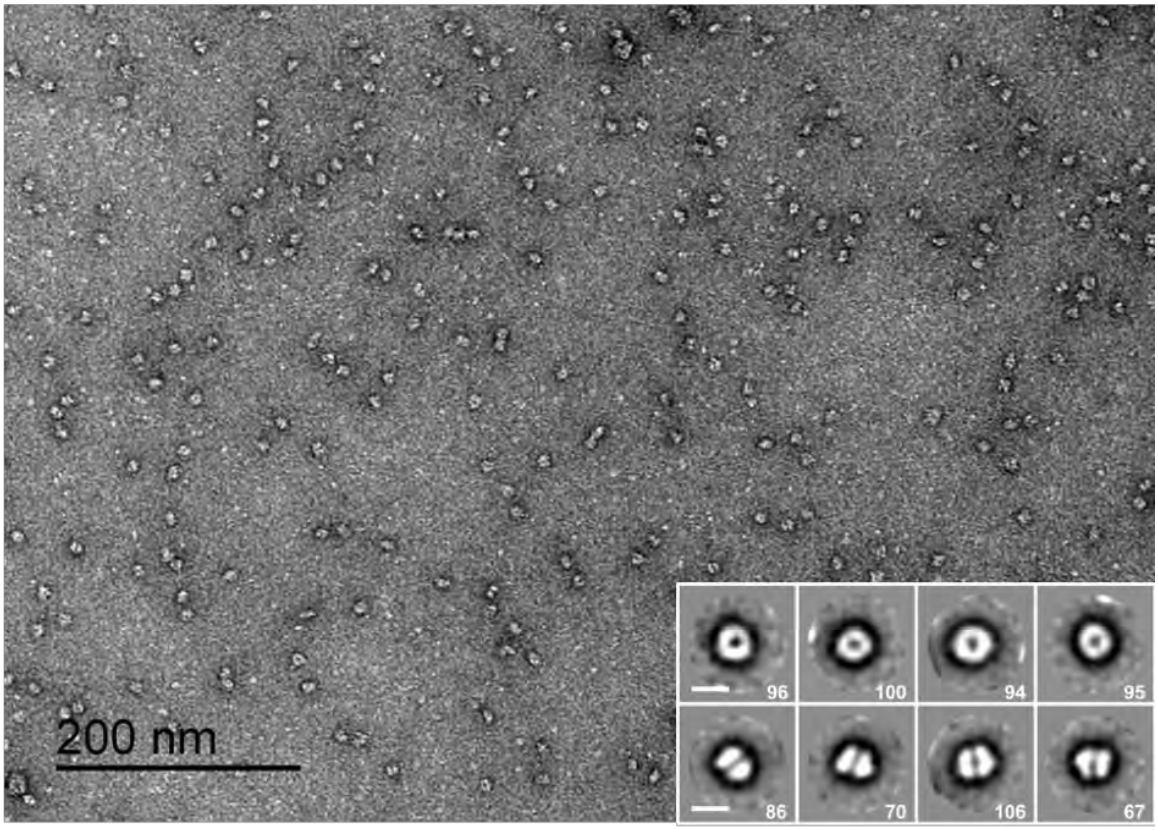


Figure 3.3. Negative stain of P74-26 TerS. Negative stain of P74-26 TerS shows that particles are homogenous and evenly dispersed. 2D classification (*inset*) shows an even distribution of top and side views.

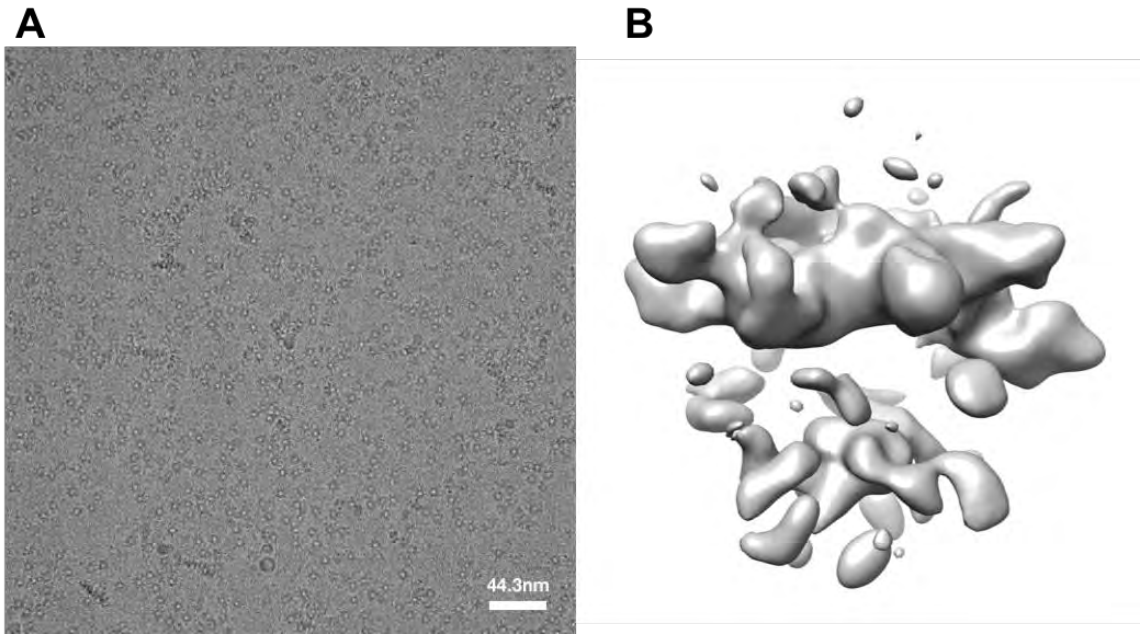


Figure 3.4. cryoEM samples of P74-26 TerS have preferred orientation. (A) CryoEM micrograph shows primarily top and bottom views of the TerS ring. Image taken on FEI Talos Acrtica at 200 kV with a Gatan K2 Direct Detector. (B) 3D reconstruction lacks density in center of ring (side view) due to preferred orientation.

To increase particle side views, we used a combination of sample additives and tilted data collection. Out of the many additives tested, such as CHAPS, NP40, LMNG and octyl-β-glucoside, amphipol A8-35 appeared to have the greatest effect on particle view distribution. After collecting a set of un-tilted images, we used a 30° tilt to obtain additional particle views (Figures 3.5A-C). Initial 3D classification of the combined datasets produces six different classes, several of which are of particular interest (Figure 3.6A). Classes 1 and 6, which account for over 50% of all particles, show apparent 9-fold symmetry. Asymmetric refinement of these combined classes generates a reconstruction with an overall resolution of 4.5 Å (Figures 3.7A-C). The features of this reconstruction remain 9-fold symmetric. Therefore, we refined class 1, the best resolved class containing 84,460 particles, with C9 symmetry to further improve the resolution. 3D refinement of class 1 with imposed symmetry results in a reconstruction of the TerS ring to an overall resolution of 3.8 Å (Figure 3.8A-C).

Using the symmetric reconstruction we built an atomic model of P74-26 TerS (Figures 3.9A-C). The model was constructed using the g20c TerS crystal structure as a starting point (PDB: 4XVN), which has 98.2% amino acid sequence identity to P74-26 TerS. In P74-26, each TerS monomer has a N-terminal helix-turn-helix motif and an oligomerization domain with two antiparallel helices. These helices stack against the oligomerization domain helices of the neighboring subunit, forming a ring with a central pore. The arrangement of helices in the pore is highly reminiscent of the central oligomerization domains of

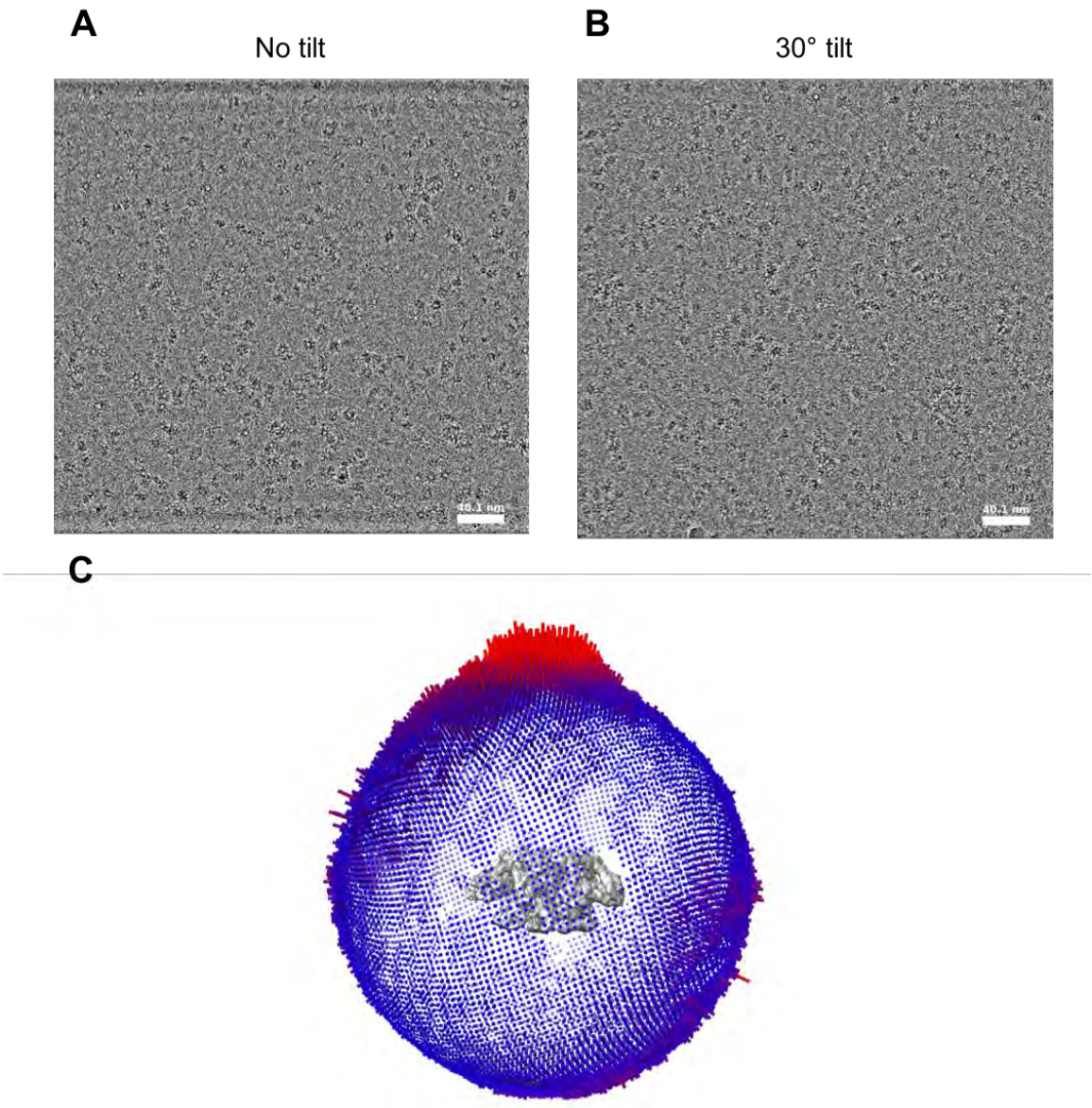


Figure 3.5. Tilted and un-tilted data collection on gold grids. Image taken on FEI Talos Acrtica at 200 kV with a Gatan K2 Direct Detector. (A) un-tilted and (B) 30° tilted micrographs were used to provide side views of the P74-26 complex. (C) Angular distribution of particles for asymmetric reconstruction of classes 1 and 6. The TerS particles still show preferred orientation, but enough side views are captured for 3D reconstruction.

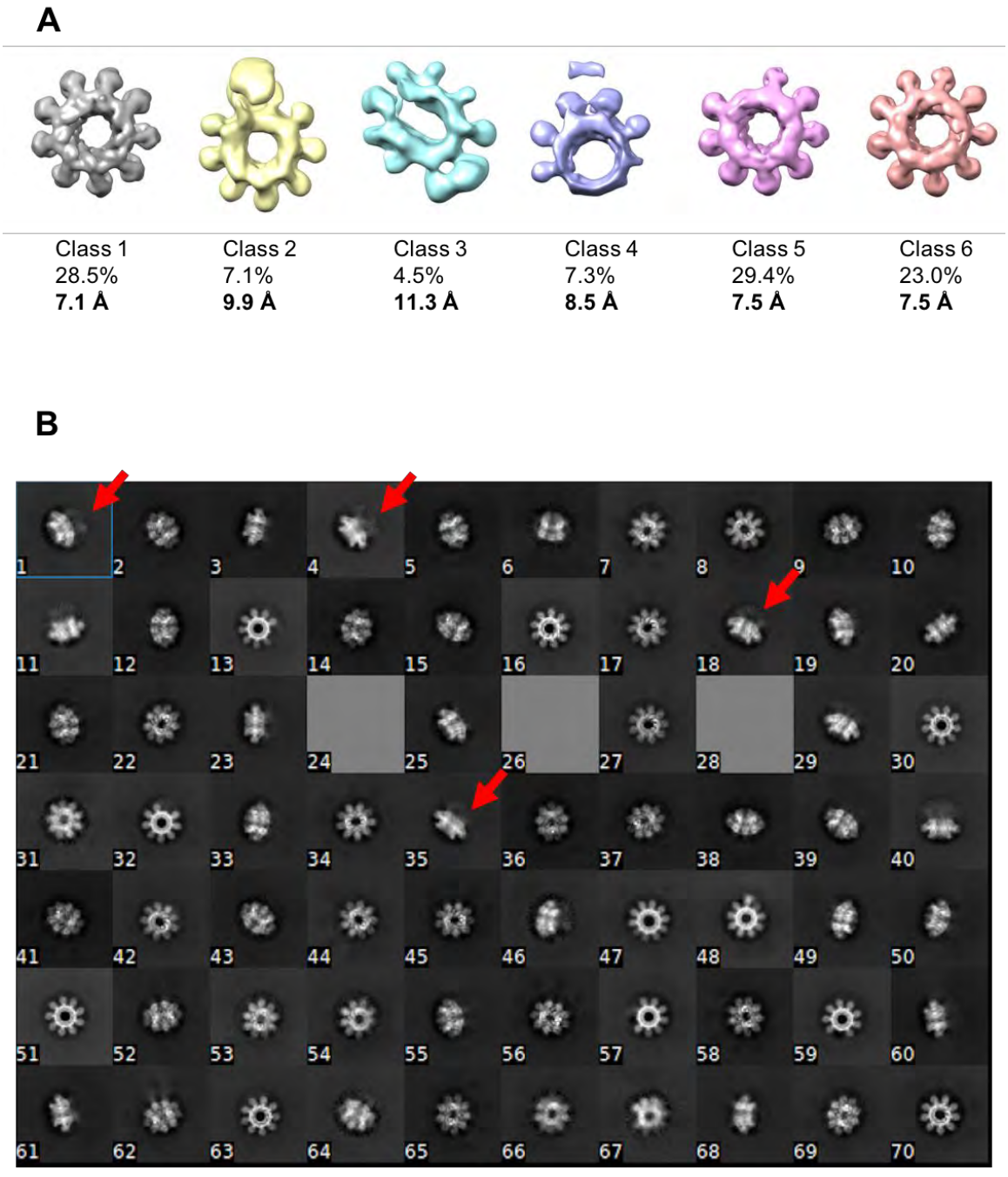


Figure 3.6. P74-26 TerS classification. (A) Asymmetric 3D classification shows 9-fold symmetry in the TerS ring. (B) P74-26 TerS 2D classes show blurred density where C-terminal region should be (red arrows).

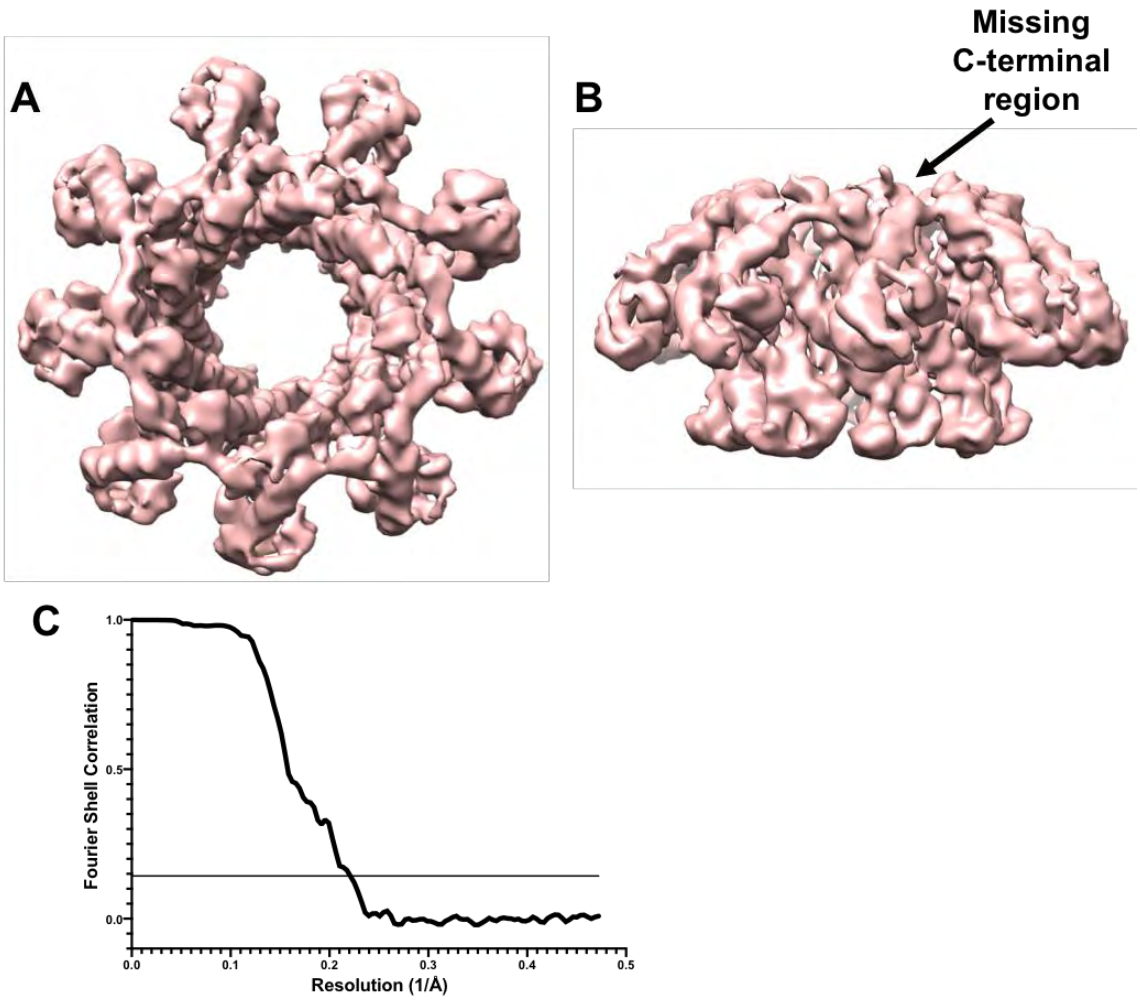


Figure 3.7. Asymmetric refinement of classes 1 and 6. (A) 4.5 Å resolution asymmetric 3D reconstruction of the P74-26 TerS ring shows 9-fold symmetry (top). (B) Side view of TerS reconstruction. (C) Gold-standard FSC curve of class 1 and 6 asymmetric 3D reconstruction. Flat line represents the FSC 0.143 cutoff.

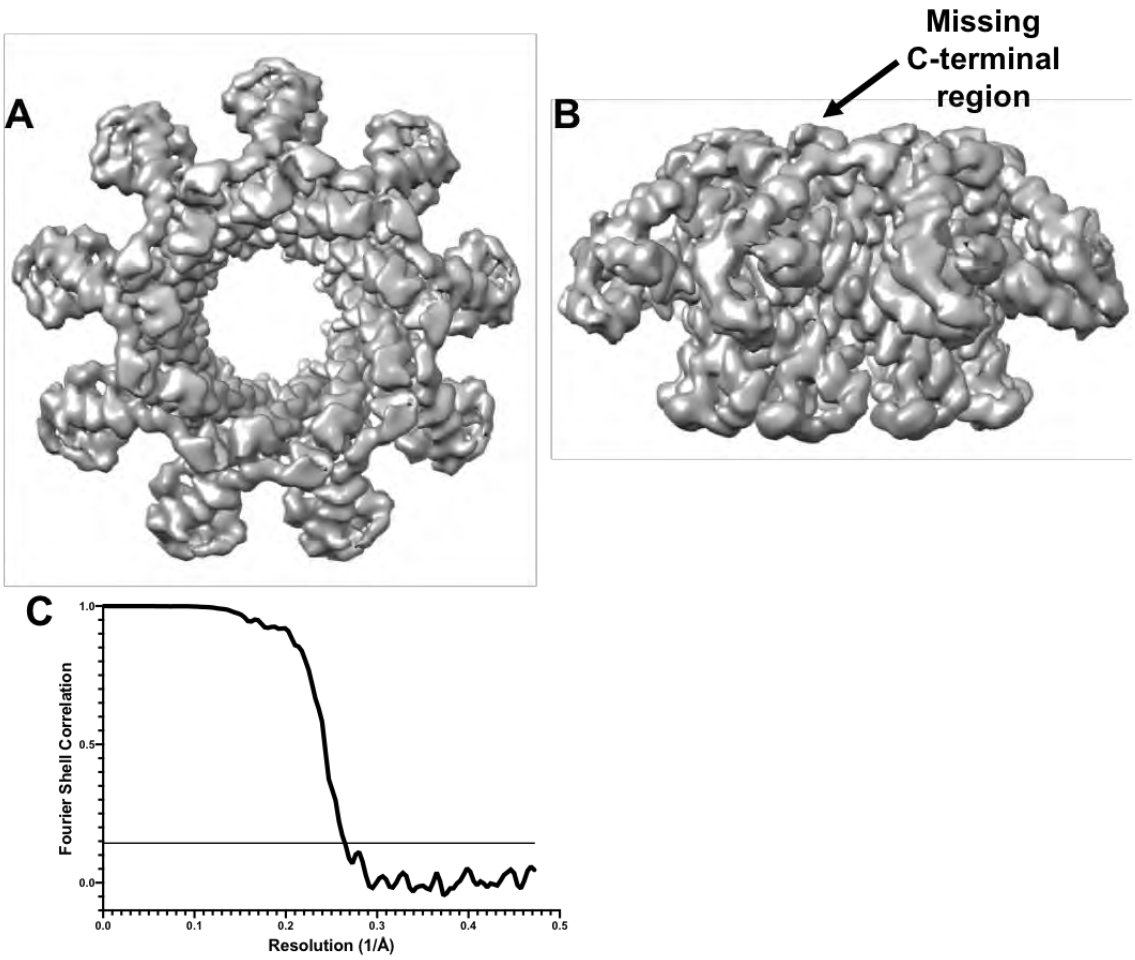


Figure 3.8. Symmetric refinement of class 1. (A) 3.8 Å resolution C9 symmetric 3D reconstruction of the P74-26 TerS ring (top). (B) Side view of TerS reconstruction. (C) Gold-standard FSC curve of symmetric 3D reconstruction. Flat line represents the FSC 0.143 cutoff.

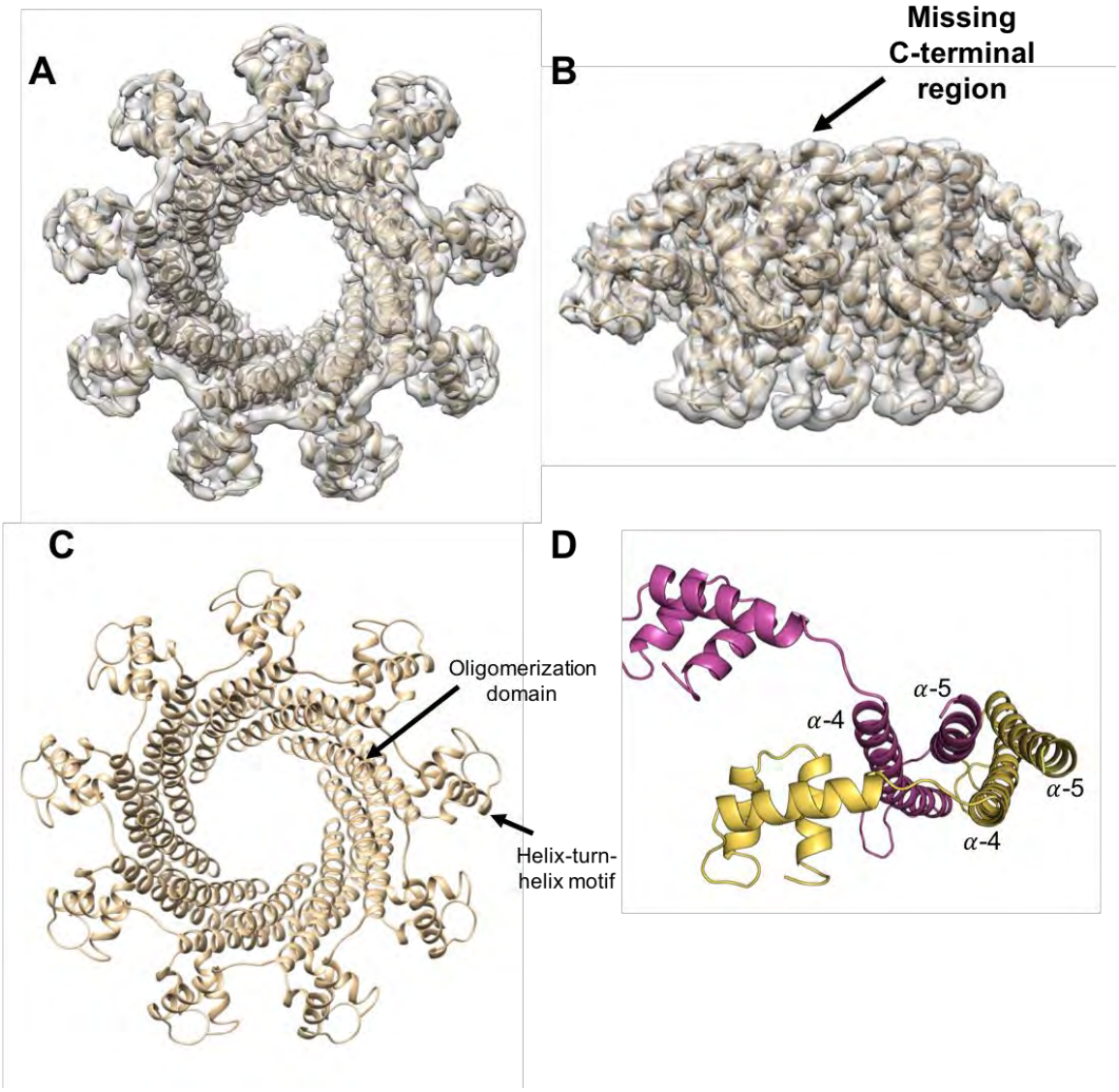


Figure 3.9. Model of P74-26 TerS. (A) Built atomic model in 3.8 Å resolution P74-26 TerS symmetric reconstruction (top). (B) Side view of atomic model in TerS reconstruction. (C) Top view of atomic model, with helix-turn-helix domains and oligomerization domains labeled. (D) In each subunit, alpha-helix 5 packs into the crevice formed by alpha-helices 4 and 5 in the counter-clockwise subunit. For simplicity, only two subunits (pink and yellow) are shown.

SF6 and 44RR TerS, with α-helix 5 of the oligomerization domain positioned in the crevice between α-helices 4 and 5 of the counter-clockwise subunit[49,87] (Figure 3.9D). The central oligomerization domains appear to be well-ordered, as local resolution of the 3D reconstruction shows the center of the pore has the highest resolution at 3.6 Å (Figure 3.10A and B). The poorest resolution, as low as 4.5 Å, is found around the perimeter of the ring in the helix-turn-helix motifs, which extend off the N-termini of the oligomerization domains like the spokes of a wheel (Figure 3.10A and B).

Surprisingly, the last 35 C-terminal residues of the protein are missing in the reconstruction. In mesophilic TerS proteins, this region forms a β-barrel with neighboring subunits and is responsible for TerL binding[50,86–88]. Both asymmetric and symmetric TerS reconstructions lack density for this region (Figures 3.7B, 3.8B, 3.12B). In 2D classification, side views of the protein are blurred where the C-terminal region should be, indicating the region is present, but not resolvable due to high flexibility or lack of defined structure (Figure 3.6B). Interestingly, secondary structure prediction designates this region as an α-helix that is connected to the oligomerization domain by a loop (Figure 3.11). This is a surprising finding, as β-barrels are found in all TerS structures to date[50,86,87], and are predicted with high confidence for *Shigella* phage Sf6 and *Bacillus* phage SF6 (Figure 3.11). For P22 phage, the secondary structure prediction omits the β-barrel present in the crystal structure from residues 129 to 135, designating this region as looped/unstructured. Interestingly, the secondary structure of the most

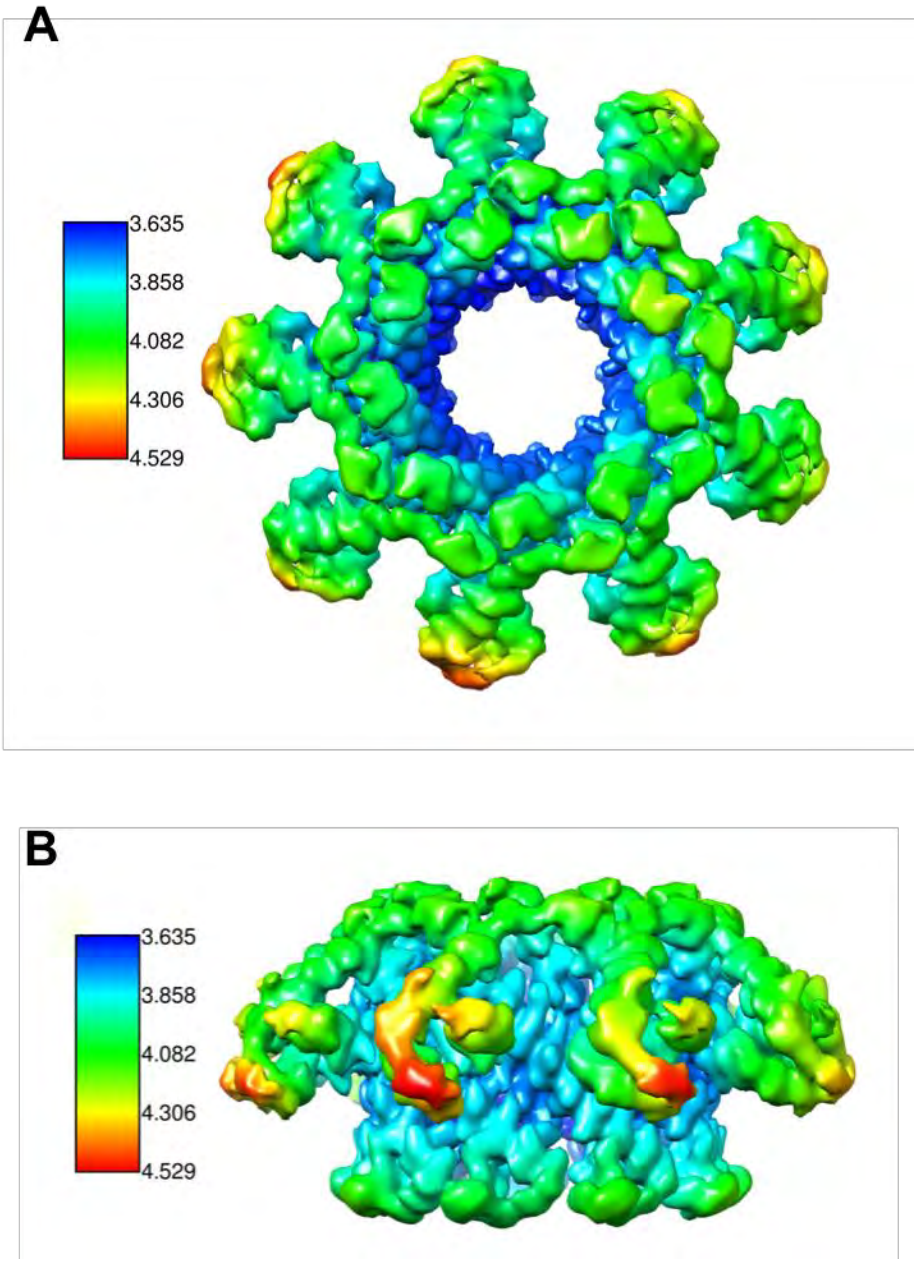


Figure 3.10. Local resolution of the P74-26 ring. Highest resolution of the P74-26 TerS ring is in the center of the pore, while the lowest resolution is in the helix-turn-helix domain. (A) Top view of the ring (B) Side view of the ring.

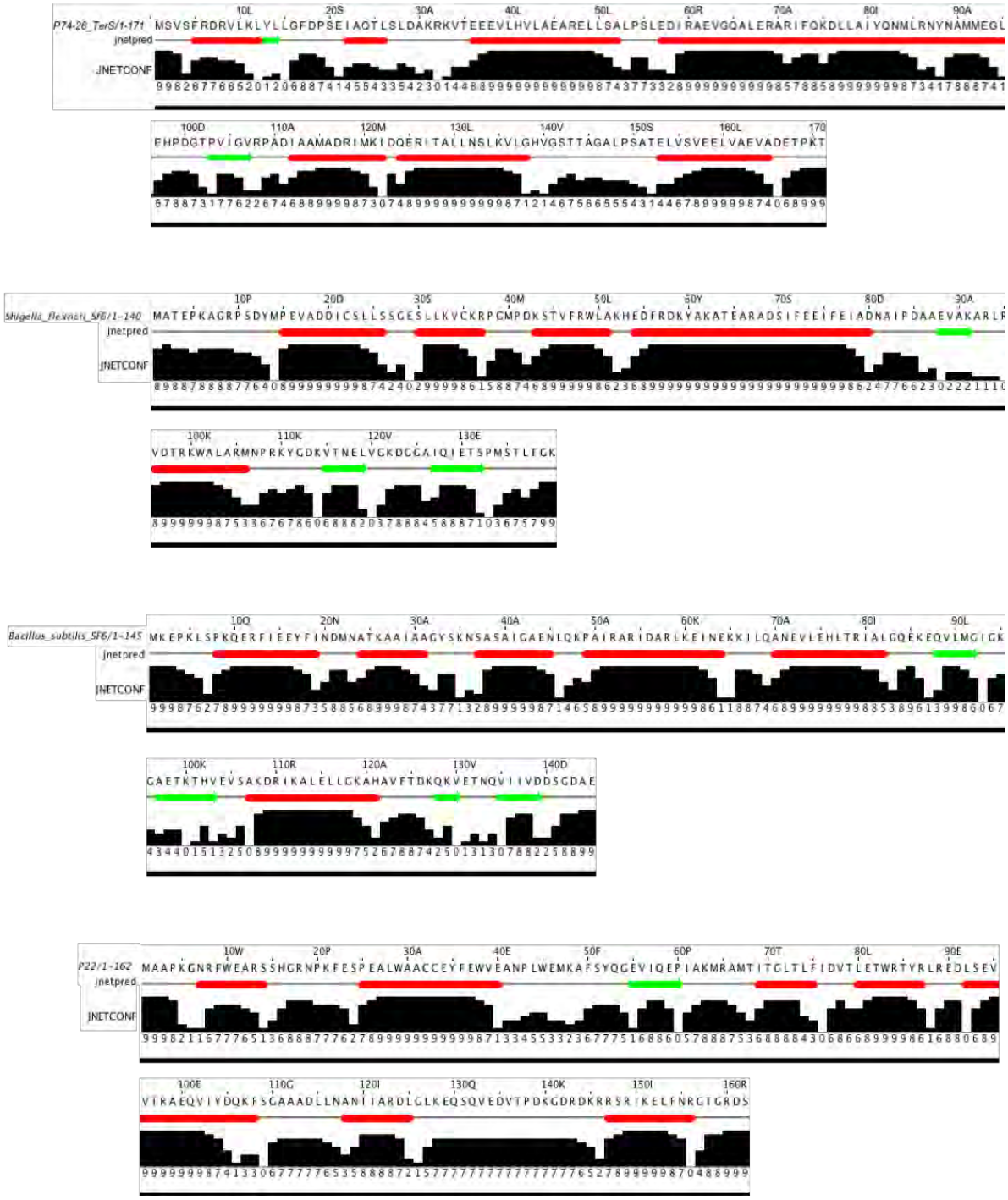
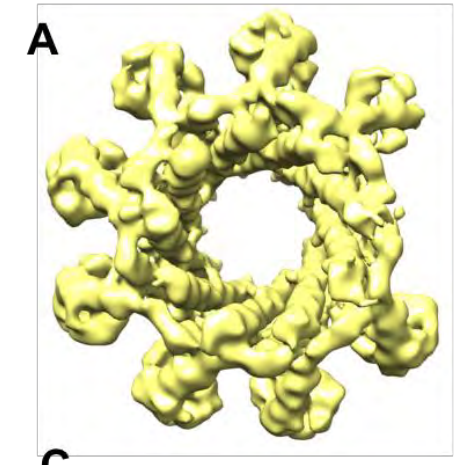
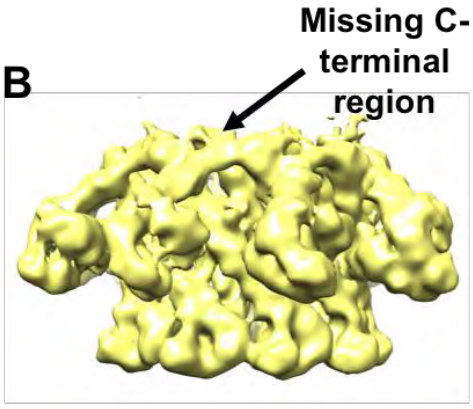


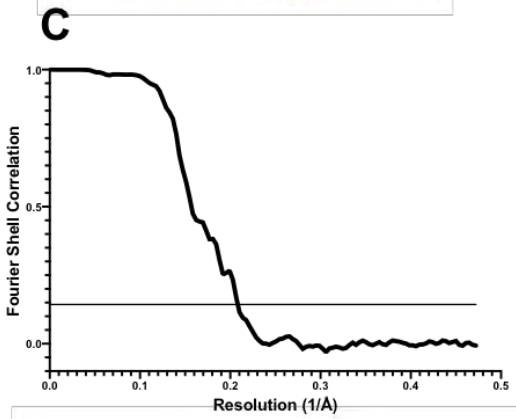
Figure 3.11. Secondary Structure Prediction of TerS proteins. Secondary structure was predicted using the JPred4 algorithm[207]. 'jetnetpred' indicates the consensus prediction (red tubes for helical regions, green arrows for beta regions). 'JNETCONF' indicates the confidence estimate for the prediction (0-9). High values indicate high confidence, while low values indicate low confidence.

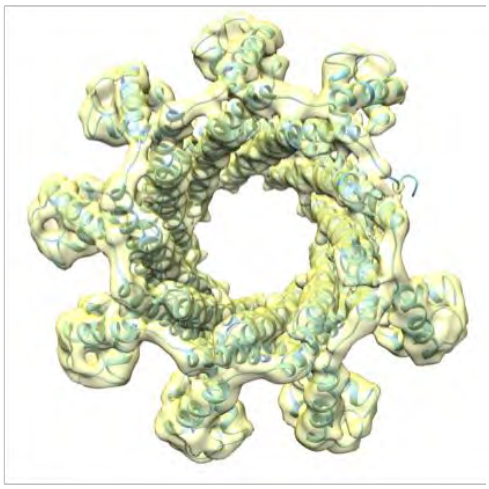
C-terminal region of P22, which is not resolved in the crystal structure, is also predicted to be α-helical. Secondary structure prediction algorithms suggest this region is α-helical with high confidence, a finding previously suggested by other groups as well[50]. From these findings, we speculate both P22 and P74-26 TerS proteins use an α-helical motif to bind their respective TerL components.

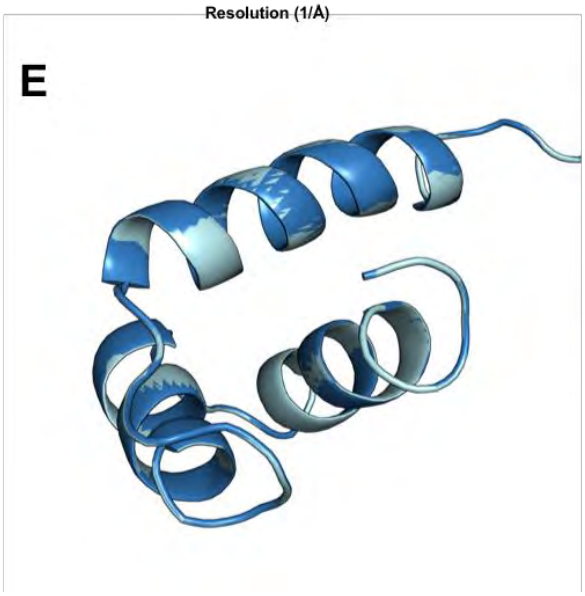
In other TerS structures, such as SF6 and Sf6, the helix-turn-helix motifs are flexible in regards to the central oligomerization domain[85–87]. This flexibility permits the helix-turn-helix motifs, termed ‘DNA binding domains’ in these phages, to stagger during DNA wrapping, allowing DNA to adopt a less constrained conformation. To investigate if the same conformational changes occur between the helix-turn-helix motifs of P74-26 TerS, we used several methods. First, we refined the asymmetric structure of class 5, the most asymmetric class with 86,969 particles. As other TerS structures show movement in the N-terminus[85–87], it is possible the missing helix-turn-helix domain in this class is due to the inherent flexibility of this region. 3D refinement produces a reconstruction to an overall resolution of 4.8 Å (Figure 3.12A-C). The reconstruction was used to create an atomic model of the class 5 structure by rigid body fitting each domain of the symmetrical model into the density (Figure 3.12D). Upon comparing each chain of the class 5 asymmetric model to all other chains within the model, no differences in helix-turn-helix motif orientation relative to the oligomerization domains were observed (Figure 3.12E). To determine if the missing helix-turn-helix motif is the result of proteolysis rather than protein


A


B
Missing C-terminal region


C
Fourier Shell Correlation
1.0
0.5
0.0
0.0
0.1
0.2
0.3
0.4
0.5
Resolution (1/Å)


D


E

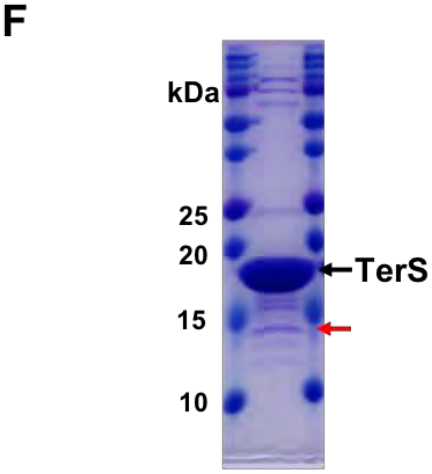

F
kDa
25
20
15
10
TerS

Figure 3.12. Asymmetric class 5 reconstruction of P74-26 TerS. (A) 4.8 Å resolution C1 symmetric 3D reconstruction of the class 5 P74-26 TerS ring (top) shows a missing DNA binding domain. (B) Side view of TerS reconstruction. (C) Gold-standard FSC curve of symmetric 3D reconstruction. Flat line represents the FSC 0.143 cutoff. (D) Class 5 atomic model in 3D reconstruction. (E) Superposition of chains A (dark blue) and C (light blue) of class 5 asymmetric model showing no conformational changes. (F) SDS-PAGE gel of P74-26 showing proteolysis (red arrow), the likely cause of the missing DNA binding domain.

flexibility, we ran concentrated purified protein on an SDS-PAGE gel. The gel shows minor proteolysis of TerS, with a band at the approximate size of a subunit missing a helix-turn-helix motif (Figure 3.12F). Using gel densitometry, we estimate that approximately 4.5% of the protein is proteolysed at this size, which is comparable to the 3% estimated by cryoEM. This suggests that the missing helix-turn-helix motif in class 5 is due to proteolysis, rather than conformational heterogeneity within the TerS ring.

We further probed for heterogeneity between the helix-turn-helix domains within classes 1 and 6. First, multibody refinement was used to examine potential movement between a single helix-turn-helix motif and the remainder of the TerS complex. Separate masks were generated between the helix-turn-helix motif and the remainder of the protein. Multibody refinement shows movement between the two bodies, however the small size of the helix-turn-helix motif (~6 kDa) prevents accurate refinement of the body (Figure 3.13A). Therefore, it is inconclusive whether the movement seen is a result of different conformations of the helix-turn-helix, or poor alignment due to insufficient signal from this region. To look at a larger part of the protein, focused classifications were done on a TerS monomer. First, symmetry expansion was implemented to align all monomers within the particle set. Particle subtraction was then used to remove the signal from the remaining eight subunits within the ring, and a focus mask was generated around the remaining subunit. 3D classification was performed on the expanded particle set to search for different conformations within the monomer.

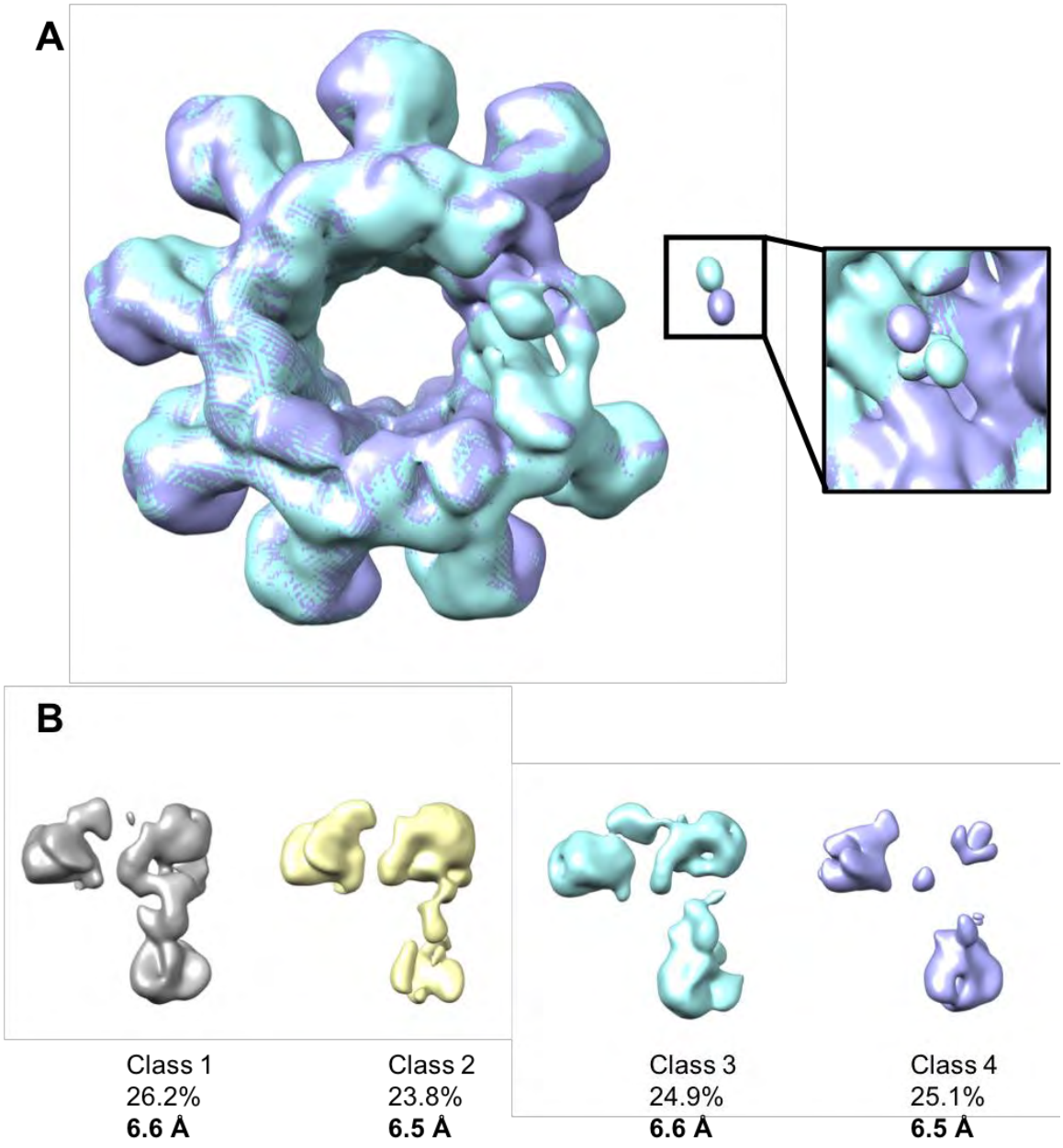


Figure 3.13. Multibody refinement and focused classification. (A) Multibody refinement of a single helix-turn-helix domain shows very small movements between initial (blue) and final (purple states) states, however the DNA binding domain itself cannot be refined (inset). (B) Focused classification of a single TerS subunit into four classes shows poor alignment due to the small size of the refined region.

Despite the increased signal from using a larger portion of the protein (~19k Da), the resulting four classes split evenly into reconstructions with patchy density, indicating poor alignment of the domain, again likely due to insufficient signal from the region (Figure 3.13B).

Comparison of the P74-26 TerS helix-turn-helix domain to mesophilic TerS structures

The presence of helix-turn-helix domains on the perimeter of the P74-24 TerS ring suggests DNA wraps around the assembled complex. This DNA-wrapping mechanism has also been hypothesized for Sf6, SPP1, and T4-like phages [85–87,209]. In comparison to *Shigella* phage Sf6 TerS crystal structures, the P74-26 helix-turn-helix motifs extend outward and rotate 56° counter-clockwise with respect to the central oligomerization domains (Figure 3.14A). This rotation positions α-helix 3, a helix predicted to bind the major groove, nearly perpendicular to the central oligomerization domains, whereas in Sf6 this helix is at a 70° angle relative to the oligomerization domains. With *Bacillus* SF6, the helix-turn-helix motifs are tethered to the oligomerization domains by highly flexible linkers, and the visible helix-turn-helix motifs are positioned in dramatically different conformations within the same crystal structure[87]. In relation to the ‘down’ pointing domain, the most similarly positioned SF6 helix-turn-helix motif, the P74-26 helix-turn-helix motif is rotated 53° clockwise with respect to the oligomerization domain (Figure 3.14B). Therefore, in comparison

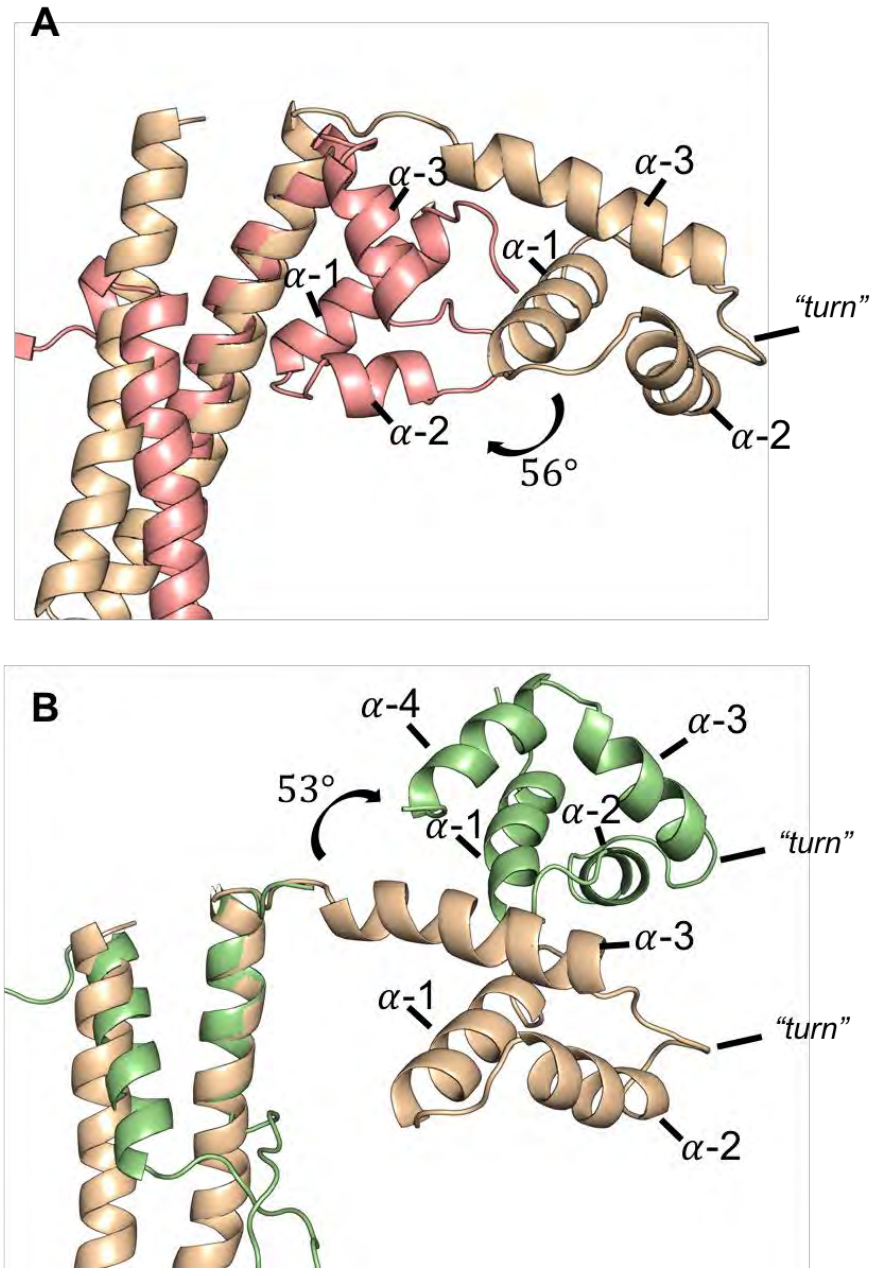


Figure 3.14. Structural alignments of P74-26 and mesophilic TerS proteins. The oligomerization domains of single subunits were aligned. (A) Alignment of the symmetric P74-26 TerS model (tan) with Sf6 phage TerS (pink; PDB 3HEF) shows the Sf6 TerS DNA binding domain is rotated 56° in relation to the P74-26 helix-turn-helix motif (B) Alignment of the symmetric P74-26 TerS model (tan) with SF6 phage TerS (green; PDB 3ZQQ) shows the SF6 DNA binding domain rotates 53° related to P74-26 helix-turn-helix domain.

to Sf6 and SF6 TerS proteins, the helix-turn-helix domains of the P74-26 TerS model are oriented differently in relation to the oligomerization domains, suggesting there are mechanistic distinctions in how the three TerS proteins bind DNA.

Interestingly, the ‘turn’ of the helix-turn-helix domain in P74-26 TerS contains basic and polar residues. These residues, specifically Lys31, Arg32, Lys33, and Thr35, may potentially bind the DNA phosphate backbone. In SF6, it was confirmed that residues in this ‘turn’ region conferred a non-specific effect on DNA recognition[84]. From this, we predict that the ‘turn’ region of the P74-26 helix-turn-helix domain primarily binds DNA phosphates through non-specific interactions with basic and polar residues.

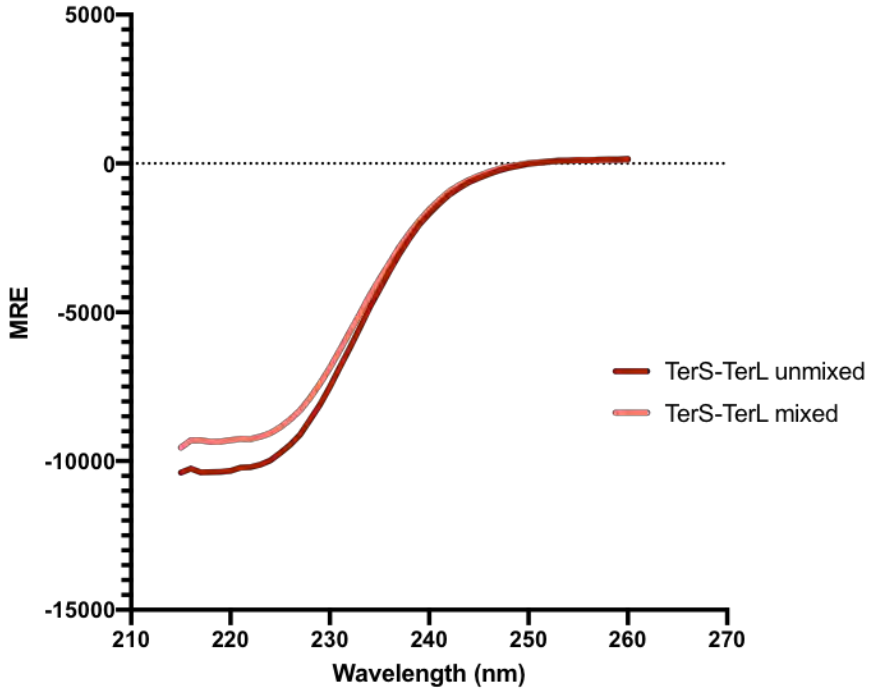
Discussion

Our discovery and characterization of the P74-26 small terminase has led to several observations regarding viral motor assembly. Previously, it seemed that the β-barrel feature of the C-terminal region was conserved throughout phage, as the presence of a β-barrel has been shown to restrain the oligomerization state of the complex[49,87]. In other systems, removing the barrel results in polydisperse oligomers[49,87], which is thought to occur because the formation of the barrel requires strict interactions between β-strands of neighboring subunits. However, in our extensive analysis of the cryoEM data, we find no evidence of β-barrel formation, yet our TerS assemblies remain

completely monodisperse according to SEC-MALS (Figure 3.1B). Therefore, in phage such as P74-26 TerS, a β-barrel is not critical for retaining correct stoichiometry.

Additionally, it is known that the TerS C-terminal region makes critical contacts with the large terminase for packaging[50,88]. This raises the question of how the small terminase of this thermophilic phage binds TerL, and what the nature of this interaction is. It is possible that P74-26 TerS requires a stimulus, such as DNA binding, TerL, or a different protein to order the C-terminal region. Circular dichroism studies of TerS and TerL do not show significant differences in secondary structure upon mixing the two samples, indicating that secondary structure does significantly change in either protein (Figure 3.15). It is possible in these experiments that the difference may be too subtle to detect, as the putative β-strands would contribute less to signal than α-helices in circular dichroism. Due to the chirality of DNA, the experiment was not attempted with DNA. Future studies will assess whether DNA binding orders the TerS C-terminal region, along with exploring the nature of the TerS/TerL interaction.

Currently, there are several hypotheses for how P74-26 TerS binds DNA. First, P74-26 TerS likely uses the ‘turn’ of the helix-turn-helix motif to interact with the DNA backbone. The concentration of basic residues in this region (Lys31, Arg32, Lys33) creates a positively-charged surface (Figure 3.16A and B), which could potentially interact with negatively-charged DNA phosphates. In SF6 TerS DNA binding studies, it was shown that the ‘turn’ and N-terminal region of α-helix



Sample	222:218 signal
TerS/TerL unmixed	0.984
TerS/TerL mixed	0.990

Figure 3.15. Circular dichroism of P74-26 TerS and TerL. Circular dichroism of P74-26 TerS and TerL unmixed (red) and mixed (pink) samples do not show significant changes in protein secondary structure (alpha:beta signal in table) when the two proteins are mixed.

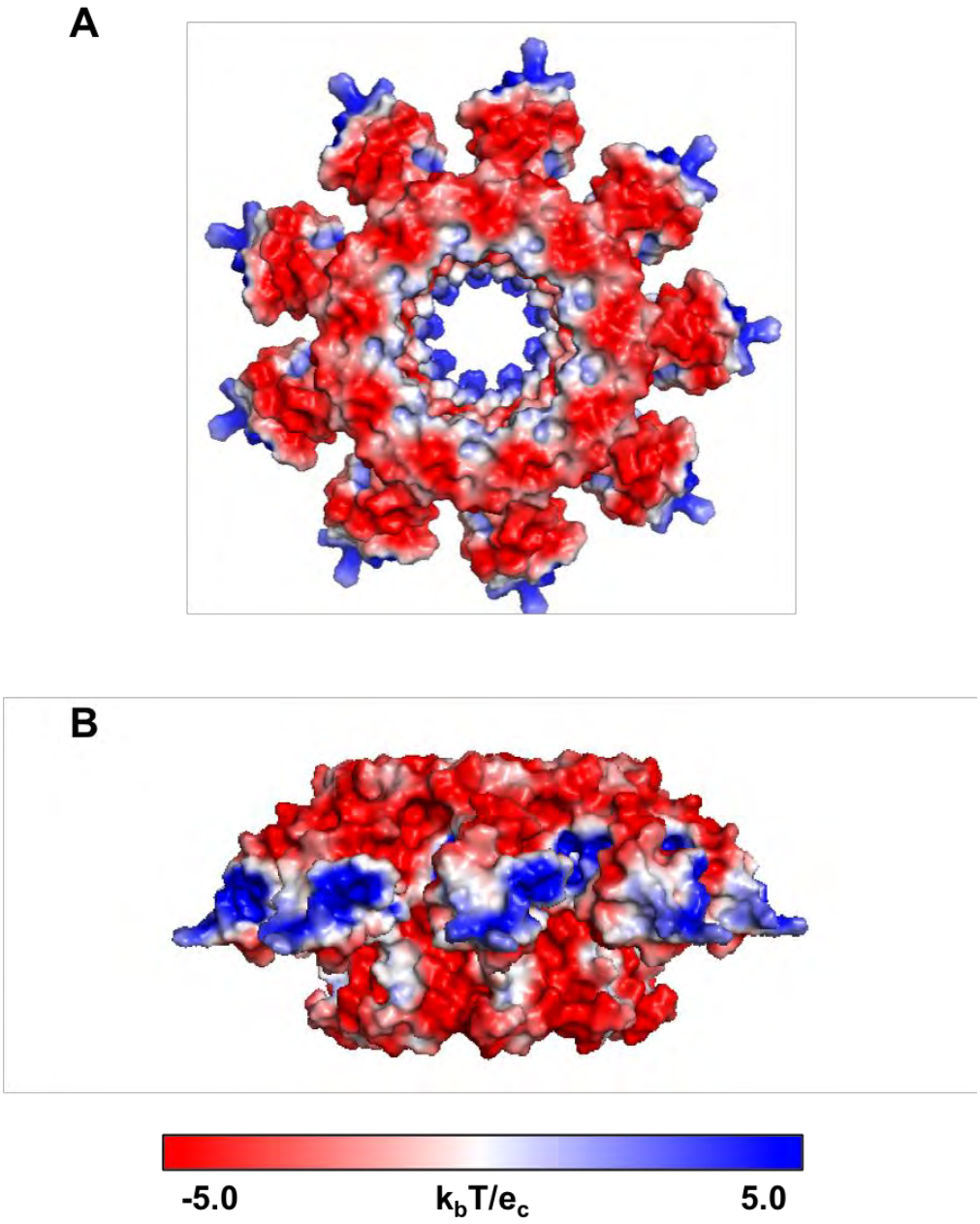


Figure 3.16. Electrostatics of P74-26 TerS ring. (A) Top and (B) side views of P74-26 TerS electrostatics using the APBS Pymol Plugin (Delano Scientific)[192]. Blue coloring indicates a net positive charge, while red coloring indicates a net negative charge. Positive charges are concentrated in the helix-turn-helix domains and the center of the TerS pore.

3 contributed to non-specific DNA binding[84], demonstrating the importance of this region for recognizing DNA. In P74-26, the distance between these ‘turn’ regions is approximately 11 bases of DNA, which compares favorably with the known 10.5 bases per turn configuration of double-stranded B-form DNA. This indicates the DNA phosphate backbone may properly align with the charged loop region, and thus the loop is positioned ideally for DNA binding. From local resolution estimation, the loop also appears to be flexible, as indicated by the lower resolution of this region (Figure 3.10). This flexibility may allow the loop adopt different conformations to further accommodate DNA binding.

Second, the different orientation of the helix-turn-helix domains in P74-26 TerS suggests there could be a functional distinction between how the thermophilic small terminase binds DNA compared to its mesophilic relatives. In all cases, the predicted major groove binding helix (α-helix 3) is positioned differently between the three phage types. Since phage must be able to recognize their own DNA in a sea of other bacterial and phage DNA, the differences in motif orientation may provide additional DNA recognition specificity. Future mutational studies of the helix-turn-helix motifs and surrounding regions will determine their roles in sequence recognition.

Third, although asymmetric reconstructions and local resolution estimation of the ring do not indicate flexibility between the helix-turn-helix motifs in P74-26 TerS, it is possible the motifs rearrange in the presence of DNA or higher temperature. Under different conditions, P74-26 TerS may adopt similar

conformations seen in Sf6 and SF6. Additionally, DNA-induced conformational changes could promote sequence-specific interactions of DNA with other regions of TerS. Future studies will determine the pliability of the helix-turn-helix motifs of P74-26 TerS in relation to outside factors and how it relates to sequence recognition.

Finally, despite the presence of helix-turn-helix motifs on the P74-26 TerS perimeter, we cannot exclude DNA binding in the TerS pore. The narrowest part of the P74-26 TerS pore has a 29 Å diameter, which is large enough to accommodate double-stranded DNA (~20 Å in diameter). In certain related mesophilic phages, such as *Shigella* phage Sf6, the pore is too narrow for double-stranded DNA to bind, eliminating this region as a potential DNA binding surface. In contrast, for P22 bacteriophage TerS, the hypothesized DNA binding mechanism is within the pore region[50], and it is possible P74-26 TerS binds DNA in a similar manner. Viewing the electrostatic surface potential of P74-26 TerS, there are several positively charged rings within the pore alongside negatively charged areas (Figure 3.16A and B). This area may potentially form a DNA binding interface which simultaneously attracts and repels the DNA phosphate backbone, propelling it into the motor. Interestingly, it's predicted that P22 has an α-helical C-terminus following the β-barrel, similar to the secondary structure prediction of P74-26 TerS[50]. However, it is of note that P22 TerS structures lack helix-turn-helix motifs on the perimeter of the ring, differentiating this structure from P74-26 and other TerS proteins.

As we gather more information about TerS proteins, we can assemble a clearer picture of how TerS proteins bind viral DNA and TerL, and how secondary, tertiary, and quaternary structure dictates the recognition mechanism (Table 3.1). For example, in TerS proteins with a α-helical C-terminal regions, does DNA thread through the pore? Or does every TerS with a helix-turn-helix motif wrap DNA around the ring? Perhaps both? Future studies will determine if there is a modular nature to TerS domains, which may be useful to engineer motors to specifically recognize and transport nucleic acid-based therapeutics.

Together, our work presents a novel thermophilic system for studying small terminase proteins and their role in viral maturation. To our knowledge, this is the first atomic resolution cryoEM structure of a small terminase protein, and also the first with an alternate conformation of the C-terminal region. Future studies of the P74-26 motor will elucidate the conformation of the C-terminal region and its role in TerL binding and enzymatic regulation, as well as the DNA binding mechanism of the TerS assembly.

Phage TerS	**Pore diameter (Å)**	**HTH motif?**	***Pac* length (nt)**	**C-terminus 2° structure**	**Favored DNA binding mode**
P74-26	29	Yes	?	Predicted alpha	?
Sf6	14	Yes	1800	Beta	Wrapping
SF6/SPP1	18	Yes	242	Beta	Wrapping
P22	20	No	22	Beta/ predicted alpha	Threading
44rr	28	?	?	?	Wrapping

Table 3.1. Comparison of ringed TerS proteins with known structures
PDB Codes: Sf6 (3HEF) SF6 (3ZQQ) P22 (3P9A) 44rr (3TXQ)

CHAPTER IV

Discussion

Summary of Dissertation Work

For my dissertation studies, my work focused on the mechanisms of viral motor DNA translocation and DNA recognition. In Chapter II, I detail my studies on the TerL ATPase, the enzymatic workhorse of the motor. In Chapter III, I present my work on the motor TerS subunit, which recognizes and binds viral DNA for packaging. Together, my studies have broadened our understanding of viral maturation by identifying critical motor features for assembly and mechanism.

In Chapter II, I show TerL motors use a *trans*-activated ATPase mechanism to regulate ATP hydrolysis. Using apo P74-26 TerL ATPase domain crystal structures, I identify critical residues for ATP hydrolysis and DNA binding. With molecular docking and SAXS, we create a ring model for TerL ATPase assembly. We also solve a second TerL ATPase domain crystal structure bound to ADP•BeF3, a non-hydrolyzable ATP analog. This structure shows a 13° rigid body rotation of the lid subdomain in relation to the apo structure, indicating a lever-like conformational change occurs during the ATPase cycle.

With these results, we generate models for both TerL motor assembly and DNA translocation. In our structural model, the ATPase domains form the central ‘hub’ of the motor, with the nuclease domains attached to portal. The motor ring grips DNA in the center of the ring through interactions of basic residues with DNA phosphates. Within the ring, the lid subdomain contacts a neighboring subunit, anchoring it in place. Upon ATP hydrolysis and release, the lid

subdomain rotates 13°. Since the lid subdomain is held in place by the neighboring subunit, the rotation transmits through the remainder of the ATPase domain. This conformational change pushes DNA upward into the capsid 8 Å in a ‘lever’-like movement. After the translocation step, the motor subunit releases DNA to a neighboring subunit, continuing the translocation cycle.

In Chapter III, I present my work on the P74-26 thermophilic TerS subunit of the motor. For this project, I develop and characterize the P76-26 TerS system for studying viral motor DNA recognition. I identify gp83 as the P74-26 TerS protein, and show the gp83 subunits oligomerize into a stable nonameric assembly. I demonstrate P74-26 TerS exhibits the characteristics of other TerS proteins, such as DNA binding, TerL ATPase activity stimulation, and TerL nuclease activity inhibition. I then solve asymmetric and symmetric cryoEM reconstructions of the TerS complex. With the reconstructions, I observe the C-terminal β-barrel of the complex, conserved in all other TerS structures[39,49,50,86,87], is absent, and thus the region appears to be unstructured or highly flexible. Secondary structure prediction indicates the C-terminal region may not form a β-barrel, and instead is α-helical. In addition, we find the helix-turn-helix motifs in P74-26 TerS are quite rigid in comparison to other mesophilic systems, and the pore is wide enough to accommodate dsDNA.

Together, our findings suggest a β-barrel may not be essential for P74-26 TerS assembly and TerL regulation. P74-26 TerS may be able to regulate its own assembly and TerL activity using a yet-to-be-discovered mechanism. In terms of

the DNA binding mechanism, P74-26 TerS appears competent to either wrap or thread DNA, warranting additional exploration of the P74-26 TerS DNA binding mechanism. These studies establish P74-26 TerS as a system for studying viral DNA recognition, with the cryoEM structure serving as a platform for future DNA binding studies.

Future directions

Structures of the TerL motor

Although research in the past several decades has elucidated important viral motor packaging steps, there are still many unanswered questions. First, there are no high resolution structures of assembled motors. While numerous models exist for motor oligomerization and DNA translocation, a full motor structure is needed to complement these studies. Visualizing the conformational transitions, intersubunit interactions, and DNA gripping contacts will shed light on how the motor generates the power to package DNA against the immense back pressure inside the capsid.

Assembled viral motors are difficult to crystallize, and all attempts to date produce monomeric TerL structures[37–39,41,210]. This may be because many ASCE ATPase rings are asymmetric, possibly impeding crystal lattice formation. In the past, many crystal structures of ASCE rings used nucleic acid to stabilize the complex by scaffolding the assembly[158–161]. DNA in crystallography can be difficult to optimize, as its length must match the complex footprint closely to

encourage crystal lattice formation. To date, there are no published viral motor crystal structures with DNA.

To overcome this challenge, cryoEM is an ideal technique. Single particle cryoEM has the ability to deal with heterogeneity and asymmetry because particles are treated individually. Thus far, cryoEM on isolated motor components is problematic, as the majority of motors do not assemble in the absence of portal and capsid. For those capable of assembling in isolation, aggregation prohibits proper structure determination. Our system, P74-26 TerL, is no exception. While P74-26 TerL assembles *in vitro* at high concentrations, it has a propensity to disassemble and aggregate in both cryo and negative stain conditions, making it a difficult sample to optimize.

Despite these challenges, a cryoEM structure of an actively packaging motor is feasible, and has already been accomplished at low resolution by other groups[41]. By using an *in vitro* packaging system, the TerL proteins assemble on portal and capsid, preventing aggregation. Furthermore, DNA may facilitate assembly. Forked DNA substrates for stalling the motor were previously designed[111,118], and these can be used to lock the motor in place for high-resolution structure determination (Appendix 2). Additionally, non-hydrolyzable or slowly-hydrolyzing ATP analogs (such as ADP•BeF_3 or ATPγS) can be used to trap the motor in a translocating state. With these reagents, a cryoEM structure is achievable.

The largest barrier to determining the structure of an actively packaging motor is creating a cryo-compatible packaging system. In many systems, such as P22, *in vitro* packaging is not possible despite decades of attempts. In the systems with established *in vitro* packaging reactions, the components are often not cryo-compatible. In the majority of cases, the packaging reactions need high molecular weight PEGs or active protein cannot be purified without glycerol. For P74-26, we developed an *in vitro* packaging reaction in which TerL is active without glycerol, but still requires PEG. In the future, we will further optimize this reaction to use lower amounts of PEG, or employ dialysis to remove the PEG post-packaging reaction.

Single-molecule studies of the TerL motor

Even without a structure of the TerL motor, there are numerous avenues to pursue regarding viral motor mechanism. A major unexplored area is single-molecule studies with the thermophilic motor, as were performed with the phi29 motor[48,56,58,64,65,211]. While phi29 genome packaging is elucidated to fine detail, the field currently understands the terminase motor DNA translocation cycle to a lesser degree. For instance, do terminases also hydrolyze ATP in a sequential order? What is their step size? Is there a similar ‘special’ subunit? Do they follow a similar chemomechanical cycle? Given there are considerable differences in DNA translocation speeds of phi29 and terminases, it is possible the two types of motors regulate packaging differently. Single-molecule optical trap experiments

are a powerful tool to assess DNA translocation step size, speed, and chemomechanics. In the future, our lab would like to explore how terminases differ from phi29 motors on a deeper level using single-molecule work. Furthermore, these studies will assess how thermophilic motors differ from mesophilic motors. From this, we will learn how thermophilic motors perform their roles within the extreme conditions they inhabit.

TerL linked-construct

Little is known about the order of ATP hydrolysis in terminase motors. Phi29 motors use a sequential ATP hydrolysis order, which is untested in terminase systems. To address this question, I created the single-chain polypeptide construct (linked-construct) described in Appendix 1. Linking the pentameric motor into a single polypeptide chain allows for absolute control over motor assembly and mutation placement, which batch experiments cannot guarantee. With this tool, we can mutate specific ATPase sites within the ring and assess how the mutations affect assembled ATPase activity, similar to the work performed with ClpX[180,194,212–217]. Outside determining the order of ATP hydrolysis, the linked-construct has numerous applications. We can tag the construct for fluorescent studies to observe motor dynamics[194] and also use it in conjunction with single molecule studies for DNA translocation. Another major application of the linked-construct is cryoEM. Linking the subunits prevents motor dissociation, and lowers protein concentrations for grid preparation, decreasing

the chances of sample aggregation. With the linked-construct, it may be feasible to determine the structure of the TerL motor.

TerS and DNA binding

The work described in Chapter III of this dissertation positions this project for further exploration of the TerS DNA binding mechanism. With a structure, we can pinpoint areas within the pore and/or helix-turn-helix motifs to probe for DNA affinity. To start, the residues in the 'turn' region (Lys31, Arg32, Lys33, Thr35) will be mutated, and DNA binding will be assessed by EMSA, as shown in Chapter III. Additional residues to test in the helix-turn-helix motif are those in helix 3, which in other helix-turn-helix proteins binds the DNA major groove[218]. Within the pore, there are several residues to probe for DNA binding as well. These include Arg108, Arg118, and Lys121. Determining whether these mutations affect DNA binding will show which region of the protein binds DNA.

If residues in the helix-turn-helix motif appear involved in DNA binding, we will further explore the wrapping mechanism using FRET. For this, we will use DNA that is approximately the same length as the ring circumference (for P74-26, this would be a 100 nucleotide long fragment of dsDNA) and conjugate appropriate FRET pairs to either end of the DNA. If the DNA wraps around the TerS ring, we expect to see high FRET efficiency, whereas low efficiency indicates a non-wrapping mechanism. This experiment will give insight into whether TerS truly wraps DNA or binds DNA in a non-wrapping fashion.

CryoEM of P74-26 TerS and DNA

Future directions of the TerS cryoEM project include a structure of TerS bound to DNA. Initial attempts at this structure were unsuccessful due to the low affinity of TerS for DNA *in vitro*. The DNA concentration needs to be very high to ensure efficient DNA binding, likely in the millimolar range. DNA shows high contrast in cryoEM, and unbound DNA causes excessive background in the images. In order to obtain a TerS with DNA structure, we will crosslink DNA to the complex. Ideally, crosslinked TerS:DNA will be removed from excess DNA though immunoprecipitation or affinity tag purification. With this sample, we will directly visualize how TerS binds DNA to gain insight into viral genome recognition.

P74-26 TerS genome binding site identification

For all DNA binding experiments, we can use gp83 DNA, as P74-26 TerS has some affinity for this region (Chapter III). However, in the future we will identify the true binding site for P74-26. A recent study of the highly similar phage P23-45 identified a *pac* cleavage site between genes 77 and 78[18]. In P74-26, an identical site is found within an analogous region between genes 76 and 77. The *pac* cleavage site location in this region suggests TerS binds nearby, but no true TerS binding sequence is known.

To identify the binding site for P74-26 TerS, we can PCR amplify this region from the P74-26 genome and evaluate discrete sections for TerS DNA

binding with EMSA. In this manner, we will identify the minimal TerS recognition sequence. If we are unable to locate a discrete binding site in this region, we will look outside the area using a ChIP-seq-like experiment. In this experiment, we crosslink TerS to P74-26 genomic DNA. We then shear the DNA, and a TerS-specific antibody immunoprecipitates TerS bound to DNA fragments. The DNA fragments are sequenced, and the data is analyzed to search for TerS enrichment sites. We will then perform *in vitro* binding experiments to verify putative TerS binding sites. Identifying these sites will provide us with an optimal sequence for DNA binding experiments.

Exploring the modular nature of TerS

In Chapter III, my work explores the idea that some regions of the TerS protein are dispensable for DNA recognition and TerL enzymatic regulation. I find that P74-26 does not have the conserved C-terminal β-barrel domain, which all other known TerS structures contain[49,50,85–87]. Secondary structure predictions suggest the P74-26 C-terminal region is α -helical rather than β-stranded. Interestingly, in P22 TerS, secondary structure prediction suggests the C-terminus ends in an α-helix[50]. In P22 TerS, there are no helix-turn-helix motifs, raising the hypothesis that DNA binds in the center of the pore. Since the C-terminus of TerS binds to TerL, the conformation of the protein in this region is important for dictating this interaction. Perhaps an α-helical C-terminus promotes a coaxially stacked interaction, whereas something that interacts in a

perpendicular fashion may use a different motif. Future studies will address the structure of the TerS C-terminal region and how it interacts with TerL.

To explore the nature of the TerS C-terminus, we will alter the C-terminal region to replicate other TerS protein C-termini. Similar chimeric proteins have already been made; in one study the domains of Sf6 and P22 were interchanged to show that the N-terminal Sf6 domains are specific to the Sf6 *pac* site[82]. Further chimeric constructs will elucidate the core principles of how phage DNA is recognized and transferred to TerL. With a deep understanding of these principles, these concepts can be applied to optimize nucleic-acid based nanotechnology to bind and transfer DNA in the most efficient manner.

Characterizing the TerS and TerL interaction

Another future direction is characterizing the TerS and TerL interaction. We know from the cryoEM structure that the P74-26 TerS C-terminal region is unstructured or highly flexible. Given these findings, it is unclear how TerS and TerL interact. A large portion of this interaction will be dictated by how DNA binds TerS. If DNA binds in the center, a coaxially stacked interaction is the likely configuration of the two proteins (Figure 4.1B). In this model, DNA threads through the center of TerS into the TerL pore. In a second model, where DNA wraps around the TerS ring, TerS and TerL may bind in a perpendicular fashion (Figure 4.1A). This allows for DNA to both wrap around TerS and thread into TerL.

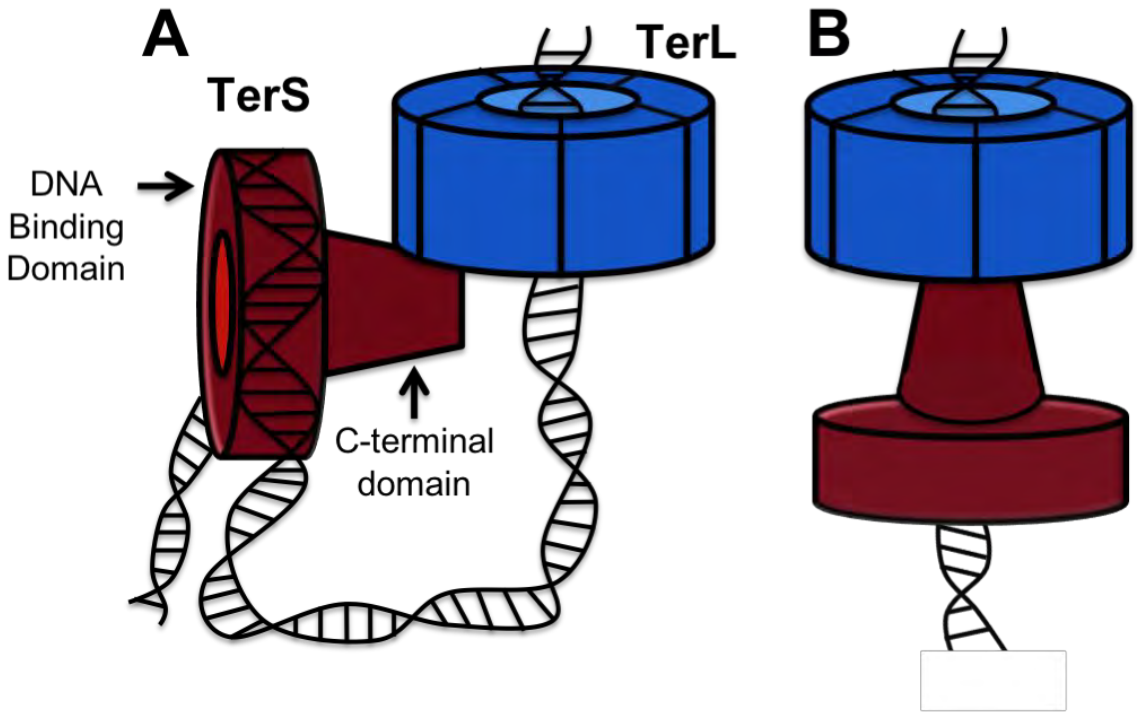


Figure 4.1. Potential TerS:TerL binding interactions. TerS (red) may interact with TerL (blue) in a (A) perpendicular manner, where DNA wraps around the TerS ring and threads into TerL or (B) a coaxially stacked manner, where DNA threads through the TerS and TerL pores.

Currently there are no high resolution structures of the TerS:TerL complex. Negative stain attempts with mesophilic systems produce difficult to interpret low resolution structures[219]. Future directions for this project include using cryoEM to obtain a reconstruction of the complex. As mentioned, TerL is challenging to use in EM because of its propensity to disassemble and aggregate on grids. It is possible to use crosslinking to prevent this and promote assembly. Additionally, DNA may encourage complex formation, and could ensure production of a biologically relevant complex.

If TerL assembly and aggregation prevents structure determination, the linked-construct may assist in overcoming these barriers. A caveat to this construct is that the stoichiometry of TerS to TerL in the biological complex is unknown. It is possible that only a few TerL subunits bind the TerS ring, which has been suggested in other models[49,219]. In these models, TerS hands DNA off to one or more TerL subunits, which bind DNA and then fully assemble on portal. Since the linked-construct is fully pentameric, this may obscure the true stoichiometry of the complex. However, the interactions between TerS and TerL may still be preserved, and the structure of the two together is worth pursuing for this reason.

Even without a structure of TerS and TerL, we can still analyze the interactions between the two proteins. Mutational analysis of the P74-26 TerS C-terminal region will determine which residues are important for TerL binding. ATPase assays are an appropriate readout strategy, as mutations that abrogate

the TerS:TerL interaction will decrease TerL ATPase stimulation. Additionally, crosslinking in conjunction with mass spectrometry may also show which TerS regions interact with TerL. It is also worthwhile to repeat the circular dichroism studies under diverse conditions. P74-26 is thermophilic, and heat may reveal structure changes within the complex that are otherwise unseen at room temperature. Together, these studies will elucidate how TerS binds to TerL to transfer DNA and regulate TerL activity.

Perspectives

The viral motor and ASCE ATPase field

The work I completed in my dissertation has uncovered important findings in both the viral motor and ASCE ATPase fields. My work on viral motor ATPases (Chapter II) was the first to introduce a *trans*-activated ATPase mechanism to the viral packaging field. Prior to this work, it was believed ATPase activity was controlled *in cis* because available models positioned the ATPase domains in isolation from eachother[41]. Introducing the *trans* model reconciled viral motor ATPase mechanism with other findings in the ASCE ATPase field, unifying viral motors with other ASCE ATPases through conserved mechanism.

Since its introduction, the *trans* ATPase mechanism was widely adopted by other bacteriophage systems. Multiple systems have since identified *trans*-acting arginine fingers in the TerL ATPase, strengthening its place in the viral motor literature[36,45,210]. Interestingly, in phi29 motors, the arginine finger resides

in a different position of the ATPase fold[27,36,37]. As the arginine finger acts *in trans*, alteration in position indicates the phi29 and terminase motors assemble differently. This finding begs the question as to how ATPases with different topologies perform the same role. Future studies will address this concept and add further insight into structure/function relationships of ASCE ATPases.

In terms of mechanism, the 'lever' model introduced in Chapter II is novel to the viral motor packaging field. With the lever model, we presented a different way for DNA to be translocated based off of ASCE ATPase knowledge and observations with the P74-26 system. With our model, we suggested a mechanism for force generation, assembly, and intersubunit coordination, which may apply to other ASCE ATPase proteins as well. Hopefully, our work will be used by others in the ASCE ATPase field to understand the mechanisms governing their own systems.

My work with the small terminase (Chapter III) introduced a new structure and model system for studying TerS DNA recognition. This structure showed not all TerS proteins have a C-terminal β-barrel, and provided a new platform to understand TerS assembly and TerL recognition. To my knowledge, it is also the first atomic resolution cryoEM structure of a TerS protein. CryoEM has great potential for use in studying other viral motor proteins, and I hope that this work will encourage others to consider EM with their own systems. Many mesophilic systems express protein poorly, and EM is a way to progress structural studies despite this issue.

One important note about the cryoEM dataset is that I used a tilted-data collection strategy to overcome severe particle preferred orientation[220]. While this is not a novel technique, for smaller particles many electron microscopists refrain from collecting tilted data due to thicker ice and subsequent particle contrast loss. While these are certainly barriers in tilted data collection, this work shows sub-4 Å structures are achievable with the tilted collection strategy. Preferred orientation is a common problem in cryoEM, and although initial optimization should be made at the level of sample preparation, tilted data collections are a valuable tool for overcoming this issue.

Prospects for drug design and delivery

Viral motors can be targeted with small molecule inhibitors for therapeutic use. Herpesviruses are among the human pathogens that use a terminase motor for maturation, and exposure to herpesviruses has been linked to birth defects[221], cancer[222] and meningoencephalitis[223]. Developing inhibitors against the motor will prevent virus propagation, making infections easier to control. Currently available terminase inhibitors against herpesviruses quickly evolve resistance mutations that make them less effective[224–227]. These inhibitors require optimization, and studies of viral motor structure and mechanism will contribute to their development. Future prospects involve engineering better inhibitors to prevent the maturation and replication of these herpesvirus for the purpose of preventing virus-related complications.

Finally, the study of motor structure and mechanism will lead to new biotechnological applications. It is easy to imagine roles for the motor in nanomaterials for targeted delivery of biological cargo. For example, the pRNA of phi29 has already been used as a therapeutic delivery device[228–231], and the portal was recently engineered as a nanopore for sensing various biomolecules[232]. It is possible that the motor itself can be engineered to package materials into nanoparticles to deliver nucleic acid based therapeutics. Additionally, understanding how the small terminase recognizes DNA will improve the efficiency of viral delivery systems by allowing researchers to better tailor recognition proteins to their targets. In the future, our work on viral motor structure and mechanism will translate to better technology in the nanoparticle and drug delivery realms of research.

Appendix I

The TerL linked-construct

Introduction

As previously mentioned, ASCE ATPases and viral motors use a *trans*-activated ATPase mechanism (Chapter II). In *trans*-activated ATP hydrolysis, neighboring subunits contribute to hydrolyzing ATP bound in a single subunit. Because the order of ATP hydrolysis in the ring determines how subunits undergo conformational changes during DNA translocation, correct *trans*-activated ATP hydrolysis regulation is important for ringed ASCE ATPase function.

In ringed ASCE ATPases, there are several mechanisms for the order of *trans*-activated ATP hydrolysis: sequential, probabilistic, concerted, and random. In the sequential mechanism, a subunit only hydrolyzes ATP if the subunit immediate upstream it has hydrolyzed ATP first. This is a commonly used mechanism for many ASCE ATPases, such as T7 and E1 helicases, and Rho1 transcription termination factor[159,161,233]. In the probabilistic mechanism, ATP hydrolysis can occur in any subunit, however the probability that ATP hydrolysis will occur in a specific subunit is contingent on other factors, such as substrate binding, ligand binding, or the states of neighboring subunits. ClpX, a very well-studied ASCE ATPase, uses a probabilistic mechanism to catalyze protein unfolding[212]. In the concerted ATP hydrolysis mechanism, all subunits cycle through the ATPase cycle simultaneously, leading to all subunits firing at once. This mechanism was predicted for the SV40 largeT antigen[234]. Finally, the

random mechanism, where there is no order for ATP hydrolysis, while theoretically possible, has never been observed in nature.

In phi29 motors, the order of ATPase activity within the ring has long been established as sequential through extensive single molecule studies[48,64]. However, in the related terminase motors, this mechanism has never been tested. Preliminary ATPase studies where a catalytically dead P74-26 TerL variant (E150A) was mixed at varying ratios to wildtype TerL showed a linear decrease in activity as the E150A variant ratio increased, indicating a probabilistic or random order of ATP hydrolysis (Figure A1.1A). Determining between these two mechanisms is not possible in a bulk system because it does not allow for specific mutation placement within the fully formed ring, which is required to uncover a probabilistic mechanism[212]. Additionally, incomplete or transient ring oligomerization may obscure a sequential mechanism and make it appear linear. Therefore, we needed additional studies to address the question of ATP hydrolysis order.

To determine the order of TerL ATP hydrolysis, I engineered covalently-linked constructs in which multiple TerL subunits were joined together in a single polypeptide chain (Figure A1.1B). With this construct, we could precisely position specific mutations within the ring in order to control the activity of individual subunits, which will unveil the order of ATP hydrolysis. This idea was inspired largely by the work done in the ClpX field, where a similar construct uncovered the ClpX probabilistic ATPase mechanism[212].

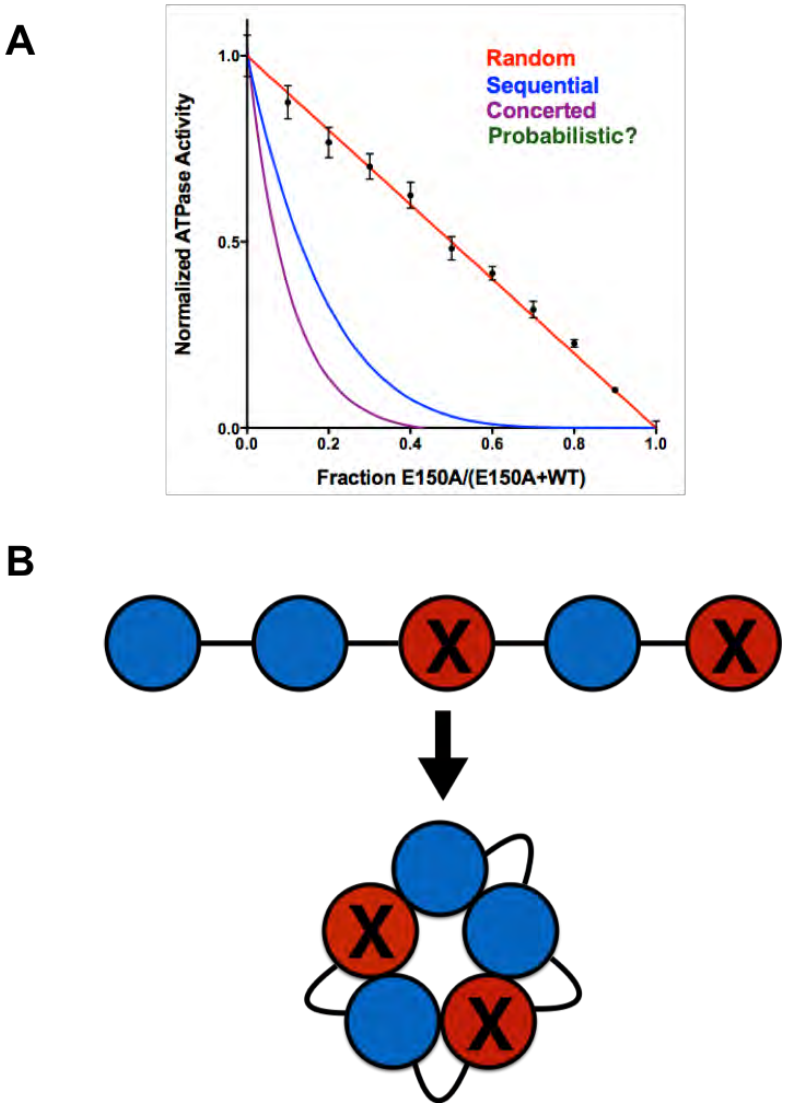


Figure A1.1. The linked-construct is a tool for studying ATP hydrolysis. (A) Simulated data of different models of ATP hydrolysis order overlaid with experimental results. A 'random' mechanism (red line) would show a linear decrease in activity as the ratio of a catalytically dead P74-26 variant (E150A) increases relative to wildtype TerL. Sequential and concerted mechanisms (blue and purple lines respectively) would sharply decrease in activity as E150A concentration increases. A probabilistic mechanism cannot be predicted to behave in a specific way because the mechanism is highly dependent on specific placement of mutations within the ring, and therefore would not be unveiled in this experiment. Black dots indicate experimental results, which overlay with the random or possibly probabilistic models. (B) Example of a linked pentameric construct. Red subunits indicate mutation subunits in the construct, which retain their positioning upon ring formation.

Materials and Methods

Cloning

The linked hexameric construct fused to maltose binding protein (MBP) was synthesized and subcloned into a pET28a vector by Genscript. The pentameric construct was made using the hexameric construct and deleting subunit 1 from the hexamer with the Gibson Assembly cloning method (New England BioLabs).

Protein Expression and Purification

All proteins were expressed in E. coli BL21 cells. Bacterial cultures were grown in Terrific Broth (Research Products International) supplemented with 30 μg/μL kanamycin at 37 °C until an OD600 of 0.7 was reached. Cells were then placed at 4 °C for 20 min, after which expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside to 1 mM. Cells were returned to the incubator to shake overnight at 18°C. Cells were pelleted by centrifugation and resuspended in buffer A [500 mM NaCl, 20 mM imidazole, 50 mM Tris (pH 8.5), 5 mM β-Mercaptoethanol (βME), 10% (vol/vol) glycerol] supplemented with 1x final concentration of protease inhibitor cocktail (Roche cOmplete EDTA-free tablets) before being flash-frozen in liquid nitrogen for storage. Thawed cells were lysed in a cell disruptor and pelleted to clear the lysate. From this point forward, all steps were performed at 4°C. Lysate was loaded onto 5 mL (per 2 liters of culture) Ni-affinity beads (Thermo Scientific) pre-equilibrated with buffer A with protease inhibitors. Protein was run over the resin at a rate of 5 mL/min and

immediately washed with 10 column volumes of buffer A + protease inhibitors to rid the sample from contaminating proteases. Protein was then eluted with 10 column volumes buffer B [500 mM NaCl, 500 mM imidazole, 50 mM Tris (pH 8.5), 5 mM βME, 10% (vol/vol) glycerol]. The eluted protein was collected and run slowly (1 mL/min) over 3 mL of amylose resin (New England BioLabs) pre-equilibrated with 5 column volumes of buffer CB [125 mM NaCl, 25 mM Tris (pH 8.5), 2mM DTT, 10% (vol/vol) glycerol]. After loading, the resin was washed with 12 column volumes of buffer CB. Protein was eluted with buffer CB supplemented with 10 mM maltose, and twenty fractions of 1 mL volume were collected. The protein eluted in the first six fractions, and these fractions were pooled and concentrated to 1.5 mg/mL.

Coupled-enzyme ATPase Assays

Coupled enzyme ATPase assays were performed as described in Chapter II at concentrations of 0.05 μM hexamer, 0.06 μM pentamer, and 0.3 μM unlinked TerL to normalize for active site number per construct.

Nuclease Assays

Nuclease assays were performed as described in Xu *et al.*[210] with 10 mM Manganese Chloride. 120 ng of pUC57 plasmid was used as the DNA substrate, and final protein concentrations are indicated in the experimental results.

Negative Stain Electron Microscopy

3.5 µL of 260 nM pentameric linked-construct was applied to a glow-discharged carbon-coated 400 mesh copper EM grid and incubated for 30 seconds. Sample was blotted off, and the grid was washed with water and blotted two times. Grid was stained with 1% uranyl acetate and imaged using a 120kV Philips CM-120 electron microscope with a Gatan Orius SC1000 detector.

Results

As P74-26 TerL has a long (36 amino acid) unstructured tail[37,116], a short linker was used to connect adjacent subunits. I selected a (Gly-Gly-Ser-Gly) x 2 linker for optimal flexibility and solubility (iGem Linker Repository). We determined in preliminary studies that this linker is highly susceptible to proteolysis, and therefore a tandem-affinity purification construct was designed, with a N-terminal thrombin-cleavable His6 tag and a C-terminal prescission-protease cleavable maltose-binding protein (MBP) tag (Figure A1.2A). This assured only full-length protein was used in subsequent experiments

To make the construct amenable to future mutagenesis, the construct nucleotide sequence was diversified, so unique primers could be engineered to mutate specific sites in the DNA sequence. At the time, it was unclear whether the motor was hexameric or pentameric (it is now strongly believed to be pentameric[19]), so a hexameric construct was synthesized with the intention of deleting a subunit for the pentameric constructs. The hexameric construct is

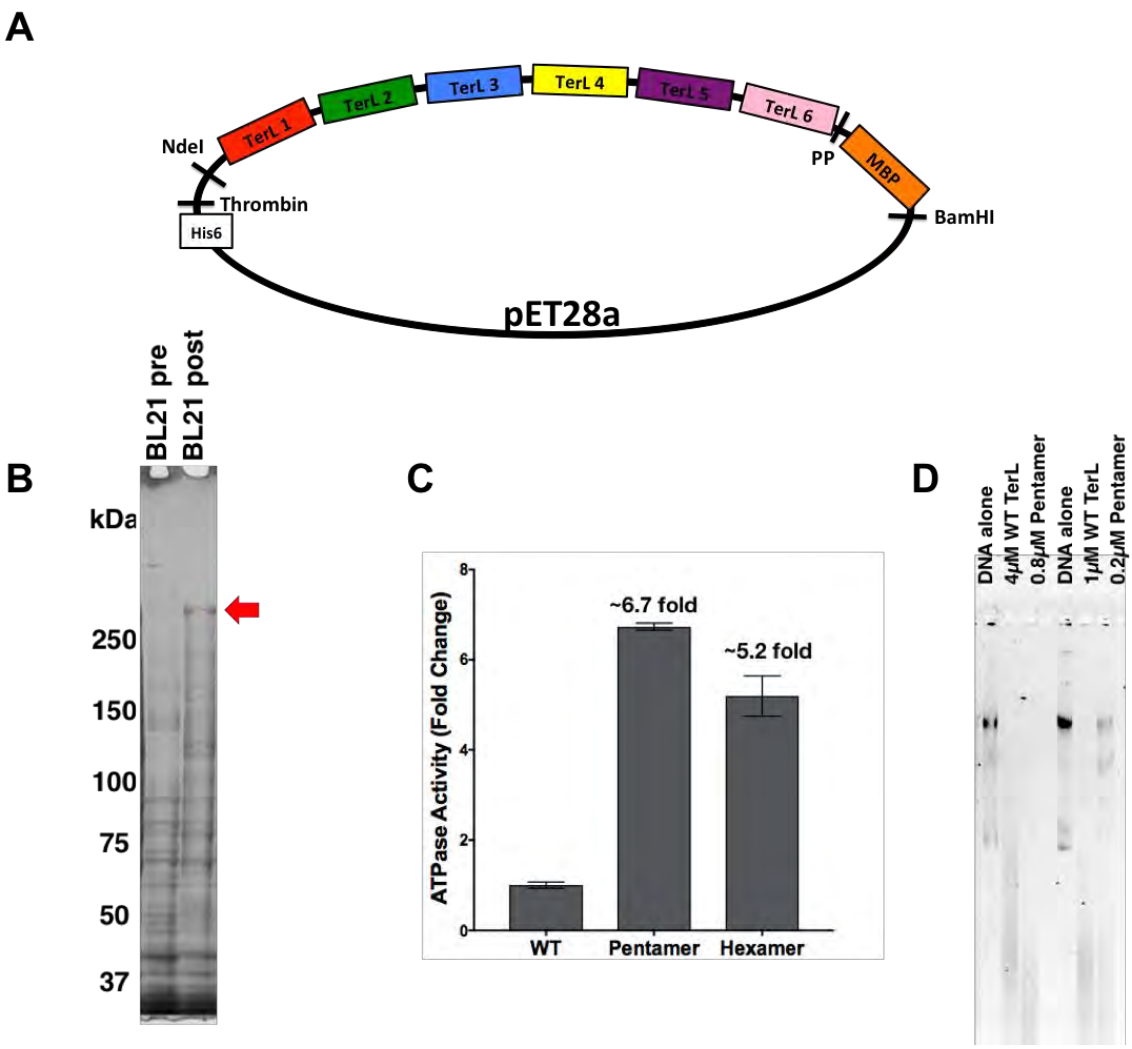


Figure A1.2. Characteristics of the P74-26 TerL construct. (A) Schematic of the hexameric TerL construct in a pET28a vector. Six TerL constructs are connected by short intersubunit linkers. There is a thrombin-cleavable His6 tag on the N-terminus and a prescission-protease cleavable maltose binding protein (MBP) tag on the C-terminus for tandem affinity purification. (B) Expression of linked hexamer pre- and post-induction in BL21 cells. Red arrow indicates expressed construct (C) ATPase activity of the linked constructs show that the pentamer and hexameric constructs are more active than the unlinked wildtype protein. (D) The pentameric linked construct shows less nuclease activity at 0.2 μM concentration compared to wildtype protein, but is more active at 0.8 μM than a comparable amount of wildtype P74-26 TerL.

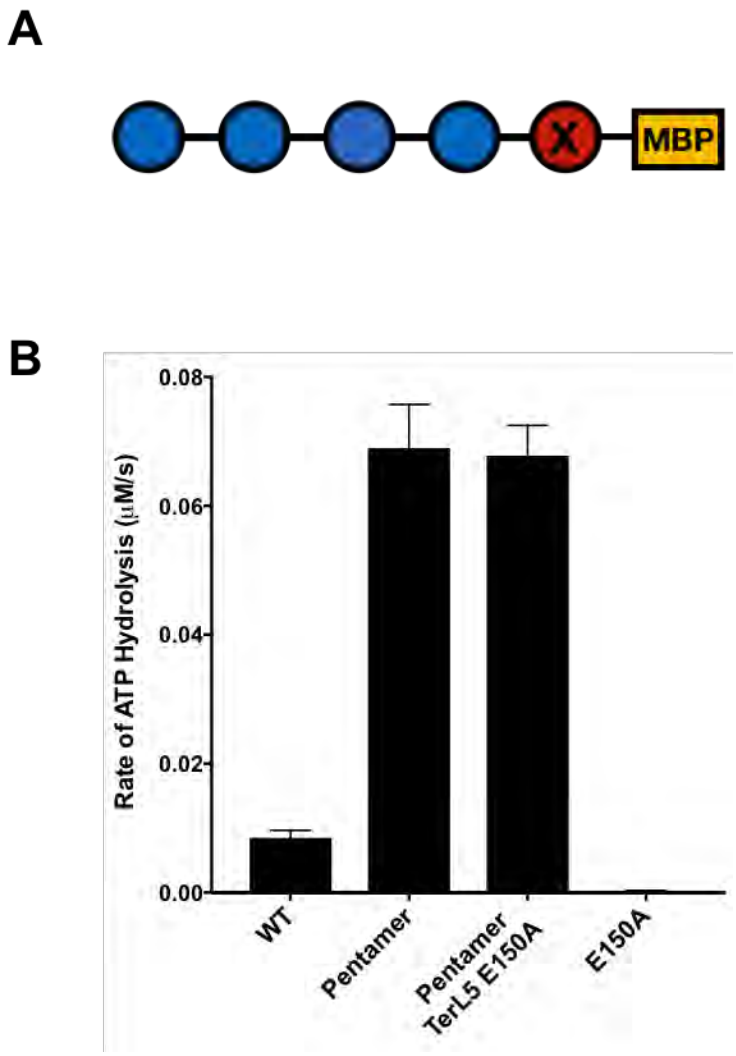
10,338 nucleotides long, yielding a 377 kDa protein. For the pentameric construct, the first subunit was deleted using Gibson assembly cloning, yielding a 8,559 nucleotide construct that is 321 kDa in size.

Protein expression and purification was challenging and arduous. The constructs express impressively for a protein of this size (Figure A1.2B), but undergo severe proteolysis despite use of protease inhibitors and cold temperatures. At the end of purification, less than 300 µg of construct is recovered per liter of expression. No conditions where the protein can be cleaved from the MBP tag without precipitation were found, and thus all experiments were done with the His6 and MBP tags intact.

Surprisingly, both hexameric and pentameric constructs show incredible activity (Figure A1.2C). The pentamer shows 6.7-fold ATPase activity over wild-type, and the hexamer shows a 5.2-fold activity increase (Figure A1.2C). The increase in activity is most likely due to increased *trans*-interactions between subunits, whereas the wildtype protein assemblies are not stable. Additionally, the pentameric construct is slightly more active than the hexameric construct. This is likely because the true stoichiometry in phage is pentameric, and the hexameric construct forces a less ideal ring assembly, slightly hindering ATP hydrolysis. Around the time of this experiment, strong evidence was published showing that the terminase motor is pentameric[19], so further experiments were performed with the pentameric construct only.

Characterization of the linked pentamer nuclease activity showed that at 0.2 μM linked-construct, the protein was less active than a comparable amount of wildtype protein (1 μm) (Figure A1.2D). However, increasing the protein concentrations 4-fold to 0.8 μM results in greater DNA degradation compared to wildtype protein. It is unclear why the lower concentration of linked-construct is less endoucleolyticaly active, however it is possible the linker off of the nuclease domain constrains the domain slightly, resulting in less catalytic efficiency.

The first mutation made in the pentameric construct was the catalytic glutamate in the last subunit (TerL5 E150A) (Figure A1.3A). If the order of ATP hydrolysis is concerted, we would expect to see abolition of ATPase activity when a single mutation is introduced. If the mechanism is sequential, ATPase activity would drastically decrease, but the construct would still retain some activity (Figure A1.1). If the order is probabilistic or random, the activity would be $4/5t^{h}$ of the fully-active linked-construct. However, when I performed this experiment, I observed no change in ATPase activity between the fully active linked-construct and the TerL 5 E150A variant (Figure A1.3B). There are two explanations for these observations. (1) The motor subunits are capable of compensating for a ‘dead’ subunit by hydrolyzing ATP at a faster rate, or (2) the protein concentration is inaccurate by about 20%. The second theory is the likely explanation, because due to the low yield and high extinction coefficient of the protein (491,225 M^{-1} g^{-1}), small errors in A280 reading are magnified in the



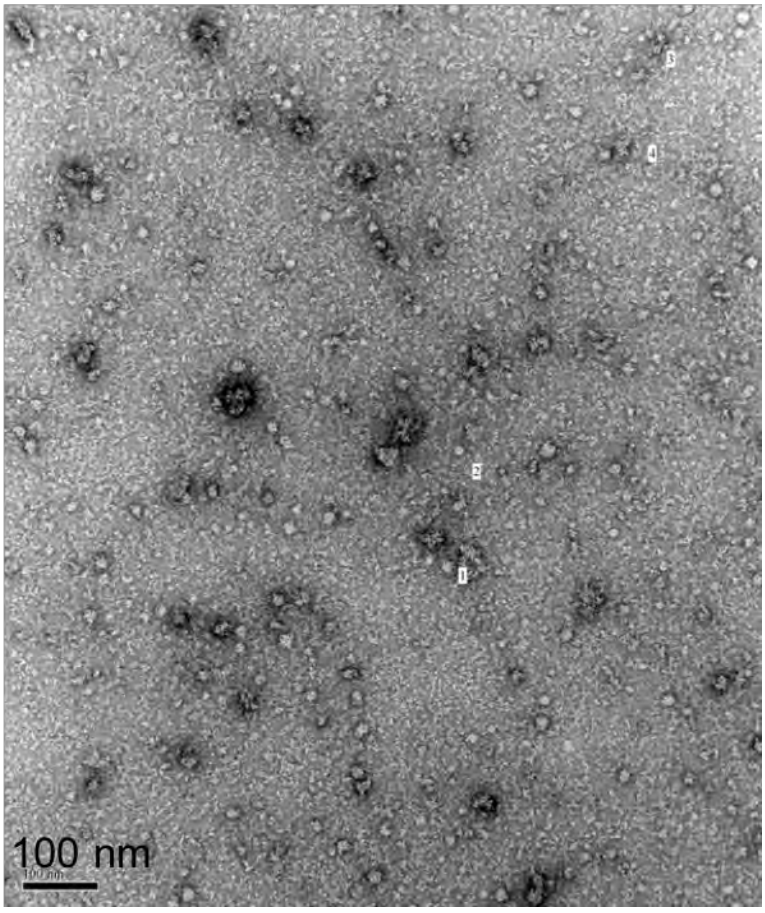
A1.3. Linked construct mutations show ambiguous results. (A) Schematic of the pentameric linked construct with a mutation of the catalytic glutamate of subunit 5 (TerL5 E150A). (B) ATPase activity of the pentameric linked construct and the pentameric TerL5 E150A are similar.

protein concentration calculation. Because of this and other reasons, the linked-construct project was put on hold, and other mutations have not been tested.

Discussion

My work on the linked-construct is the first of its kind in the viral motor field. While the process of determining the order of ATP hydrolysis using the construct is tedious due to low protein yield, it is still a valuable tool for studying the motor. For instance, the construct is fully assembled and therefore may be useful for cryoEM structure determination. The wildtype TerL protein behaves very poorly in cryo and negative stain EM, but initial negative stain attempts with the linked-construct are promising (Figure A1.4). Single particles are visible, and the sample appears to be optimizable for cryoEM. Currently, the protein is stored in maltose-containing buffer, which may cause issues with cryo grids and therefore should be removed before proceeding.

Other applications for the linked-construct include fluorescence and single molecule studies. The ClpX linked-construct has been used with fluorescence to determine ClpX ring dynamics[194], and the TerL linked-construct could potentially be used in a similar way. Additionally, the TerL linked-construct should be tested in the P74-26 *in vitro* packaging assay as well. If the construct displays translocation activity, it could give unparalleled insight into the packaging cycle. Therefore, the linked-construct has many potential uses in viral motor studies.



A1.4. Negative stain EM shows individual linked pentamer particles.

Appendix II

The cryoEM structure of an actively packaging motor

Introduction

For decades, the viral motor field has attempted to obtain a structure of a functional motor to understand how double-stranded DNA viruses package DNA. To date, there are few high-resolution structures of the TerL ATPase, and of those that exist, the protein structures are monomeric[37–39,210]. It is seemingly difficult to crystalize the motors in a fully assembled form, likely because of inherent asymmetry within the TerL ring. Fortunately, there are other techniques available to study viral motor structure, the most prominent of which is cryoEM. Recently, cryoEM has become a first-line technique for many large protein complexes due to the advent of direct electron detectors[235,236]. In terms of viral motors, cryoEM is a valuable tool for obtaining motor structures, as asymmetry in the complex does not necessarily deter users from obtaining high-resolution reconstructions.

In 2008, Cell published a cryoEM structure of an actively packaging T4 bacteriophage motor attached to a capsid[41]. In this structure, the resolution of the motor component was very low (~34 Å), making it difficult to interpret. Since publication of this T4 reconstruction, new technology has made obtaining high-resolution structures incredibly more feasible. Additionally, we began a collaboration with Dr. Lindsey Black's lab at the University of Maryland, who is one of the world's experts on the T4 system. For these reasons, we decided to revisit the T4 bacteriophage capsid and motor cryoEM work with hopes of obtaining a high-resolution reconstruction of an actively packaging motor. With

this structure, we expect to elucidate the mechanism of terminase motor packaging to unprecedented detail.

For this project, I attempted to optimize well established *in vitro* packaging assays for T4 bacteriophage to make them cryo-compatible. This proved challenging, as the T4 TerL protein requires high amount of glycerol in the storage buffer (~30%) to retain packaging activity and PEG in the reaction to package, both of which are not cryo-compatible. I also faced challenges with constructing the Y-DNA that was designed to stall the motor, and found optimizing the cryo grids to be difficult with the low concentrations of proheads in the reaction. In conclusion, to make this project feasible, non-hydrolyzable ATP analogs, crosslinking reagents, and more concentrated bacteriophage components will be needed.

Materials and Methods

Prohead and TerL purification

Proheads and TerL were purified by Lindsey Black's lab at the University of Maryland[78,237].

Negative stain electron microscopy of ESPs and ELPs

3.5 µL of 1.1x 10^{10} ph/ml ESPs or 6.8x10^{9} ph/ml ELPs (stored in a buffer containing 50 mM Tris pH 8.0, 5 mM $MgCl_2$ 0.5 mM EDTA) were applied to a glow-discharged carbon-coated 400 mesh copper EM grid and incubated for 30

seconds. Sample was blotted off, and the grid was washed with water and blotted two times. Grid was stained with 1% uranyl acetate and imaged using a 120kV Philips CM-120 electron microscope with a Gatan Orius SC1000 detector.

Cryo-electron microscopy of ESPs and ELPs

400 mesh 2/2 Holey Carbon C-Flat grids (Protochips) were incubated with chloroform until dry. Grids were glow discharged for 60 seconds at 20 mA (negative polarity) with a Pelco easiGlow glow discharge system (Pelco). 3 µL of $1.1x\ 10^{10}$ ph/ml ESPs or $6.8x10^{9}$ ph/ml ELPs (stored in a buffer containing 50 mM Tris pH 8.0, 5 mM $MgCl_2$ 0.5 mM EDTA) were applied to the grid at 22°C and 95% humidity in a Vitrobot Mark IV (FEI). Samples were blotted for 2 seconds with a blot force of 2 after a 15 second wait time. After a 0.5 second drain time, samples were vitrified by plunging into liquid ethane and were stored in liquid nitrogen until data collection. Grids were imaged using a 120kV Philips CM-120 electron microscope with a Gatan Orius SC1000 detector.

T4 in vitro *packaging*

The T4 *in vitro* packaging reaction protocol was adapted from Black *et al.* 2006[94]. In a 16 µL reaction, 2.0×10^{9} ph/mL (storage buffer contains 50 mM Tris pH 8.0, 5 mM $MgCl_2$, 0.5 mM EDTA) 530 nM T4 TerL (storage buffer contains 50 mM Tris pH 8.0, 0.1 mM EDTA, 5 mM $MgCl_2$, 0.2 mM $ZnCl_2$, 30% glycerol), 460 ng of dsDNA, 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 6 mM $MgCl_2$, 2 mM

spermidine, 1 mM putrescine, 5% PEG 20,000 (or substitute indicated), 0.5 mM DTT, and 5 mM ATP were incubated for 1 hour at room temperature to encourage DNA packaging. Next, 2 µg of DNaseI was added and incubated for 30 mins at 37°C. After the incubation, 3 µL of 1 part proteinase K (20 mg/mL), 1 part EDTA (0.5 M, pH 8), 1 part SDS (10%) was added and incubated at 65°C for 30 minutes. To make the gel sample, 10.5 µL of the reaction and 2 µL sample buffer were run on a 1% (wt/vol) TAE-agarose gel with a 1:10,000 dilution of GelRed dye (Phenix Research) for 75 minutes at 85 volts.

T4 centrifugation buffer exchange protocol

This protocol was adapted from Hegde *et al.* 2012[157]. In a 50 µL reaction, $2x10^{11}$ ELP proheads and a 10:1 ratio of TerL to portal monomer (12 portal molecules per prohead) were used in a packaging reaction (see above) and incubated for 1 hour at room temperature. The digestion steps of the packaging reaction were foregone, and samples were spun at 17,000 x g at 4°C for 45 minutes. 35 µL of supernatant was removed, and the pellet was washed with 500 µL of prohead buffer (50 mM Tris pH 8.0, 5 mM $MgCl_2$, 0.5 mM EDTA) with 5 mM ATPγS. Samples were spun at 17,000 x g for 45 minutes at 4°C. As much supernatant as possible was removed from the washed pellet, and the pellet was resuspended in 20 µL prohead buffer plus 5 µL of 6x SDS protein loading buffer. Sample was run on a 12% acrylamide SDS-PAGE gel, and stained with Coomassie blue for imaging.

Y-DNA design and creation

Oligonucelotide Y-DNA primers were ordered from IDT. Primers were designed to have a built-in EagI restriction site upon annealing. Y-DNA-PreCut-1 primer sequence: GGCCGATTCATAGGTAGCATACGTGCCCGGGCCACTATGGCAGAT CGAATTTTTTTCGATCTGCCATAGTGGCCCGGGC with 5' phosphorylation. Y-DNA-PreCut-2 primer sequence: CGGGCCCGGTGATACCGTCTAGCTTT TTTTAAGCTAGACGGTATCACCGGGCCCGACGTATGCTACCTATGAATC. Primers were resuspended and annealed in a buffer of 20 mM Tris pH 8.0, 1 mM $MgCl_2$. Leader sequences were PCR amplified from a ScaI digested pUC19 plasmid with a primer that adds an EagI sequence onto one end. 2:1 ligations were performed of Y:leader overnight with T4 ligase (New England BioLabs) at 16°C.

Polylysine and polyglutamine treated grids

400 mesh 2/2 Holey Carbon C-Flat grids (Protochips) were incubated with chloroform until dry. Grids were glow discharged for 60 seconds at 20 mA (negative polarity for polylysine, positive polarity for polyglutamine) with a Pelco easiGlow glow discharge system (Pelco). 5 µl of 0.1% polylysine or polyglutamine was incubated on the grid for 90 seconds. Additive was wicked off

with Whatman paper, and grids were washed two times with water. Samples were prepared as explained above, with the substitution of 1 second blot time.

Results

For this work, the proposed approach was to use the T4 *in vitro* packaging system alongside a forked 'Y-DNA' substrate that was previously designed and tested by our collaborator to stall the T4 motor during packaging[111,118] (Figure A2.1). Using this Y-DNA eliminates the need for non-hydrolyzable ATP analogs, which may hydrolyze over time. We predicted that locking the motor in place would assist in cryoEM reconstruction by stabilizing the motor in a fixed position prior to freezing, which also ensures DNA would be captured in the TerL channel.

Electron microscopy of T4 proheads

For T4 bacteriophage, there are two types of packaging-capable proheads: Empty Small Proheads (ESPs) and Empty Large Proheads (ELPs). As their namesake suggests, ESPs are the unexpanded, smaller precursors to ELPs[237,238]. What causes the transition from ESP to ELP between the two proheads is not known, however both can be used for *in vitro* packaging[94,237]. Initial negative stain evaluation was discouraging for both samples of empty proheads (Figure A2.2A and B). Both ESP and ELP samples looked shriveled, heterogenous, and appeared to clump together on the grid. Nevertheless, we attempted cryoEM on the proheads to see if the stain deformed the capsids.

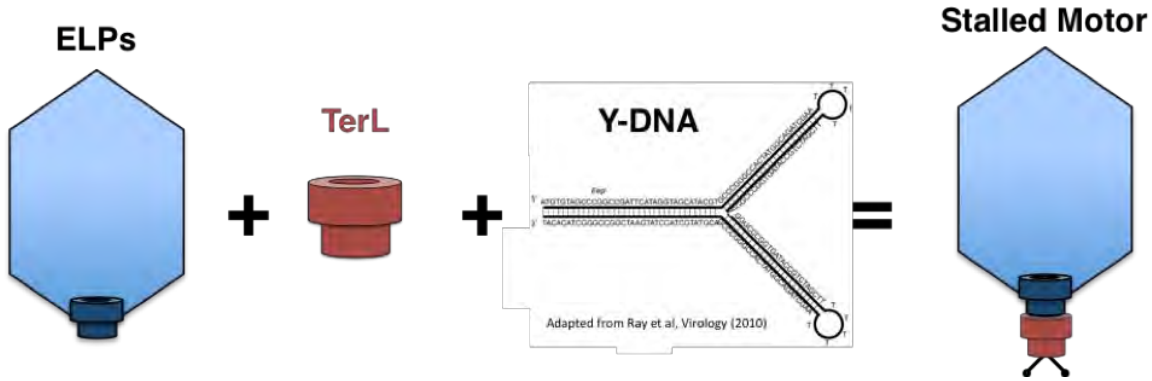


Figure A2.1. Proposed approach for stalling T4 packaging motors for cryoEM. In this approach, we combine the established *in vitro* packaging system for T4 bacteriophage[157] with Y-DNA[111,118] to stall the motor for cryoEM imaging.

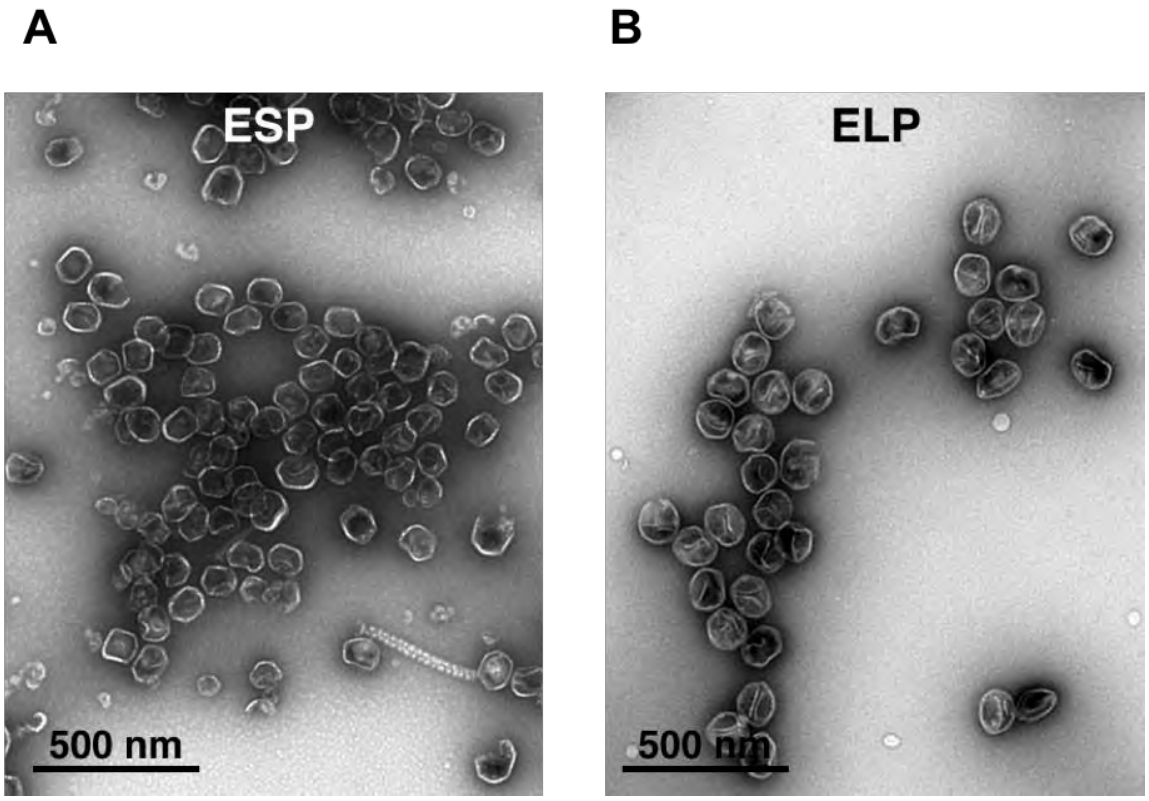


Figure A2.2. Negative-stain electron microscopy of ESP and ELP particles. Both Empty Small Proheads (ESP; A) and Empty Large Proheads (ELP; B) are heterogenous, deformed, and aggregated in negative stain conditions. These images were taken at 40,000x magnification.

Fortunately, this turned out to be the case. Whereas ESPs still remained heterogenous and unstable, the ELPs samples significantly improved (Figure A2.3A and B). In the ELP sample, the ELP particles appear highly stable, homogenous in shape, and show a multitude of orientations (Figure A2.3B). Since the ELP sample appeared to be highly amenable to cryoEM, we decided to proceed with ELPs for future optimization.

On the Talos Arctica, the K2 direct detector unveiled more details of the T4 prohead (Figure A2.4). It is possible to see an area of greater contrast at one end of each prohead (red arrows) that likely corresponds to the portal ring. Additionally, hair-like protections extend off the capsid surface. These are likely the 'HOC' proteins, which are 40 kDa flexible proteins that reach up to 60 Å away from the capsid and number 155 proteins per capsid[239]. These findings were incredibly encouraging for high-resolution structure determination using the T4 system.

Cryo-compatible T4 in vitro *packaging optimization*

Next, I went on to evaluate the T4 reagents for *in vitro* packaging. In the *in vitro* packaging assay, proheads, TerL, DNA, ATP, and reaction buffer are mixed and incubated to facilitate DNA packaging (Figure A2.5A). DNase is then added to degrade extra unpackaged DNA, and afterwards proteinase K and SDS are used to digest the capsid and expose the packaged DNA. Because the T4 TerL protein is usually stored in a non-cryo-compatible buffer with 30% glycerol, I

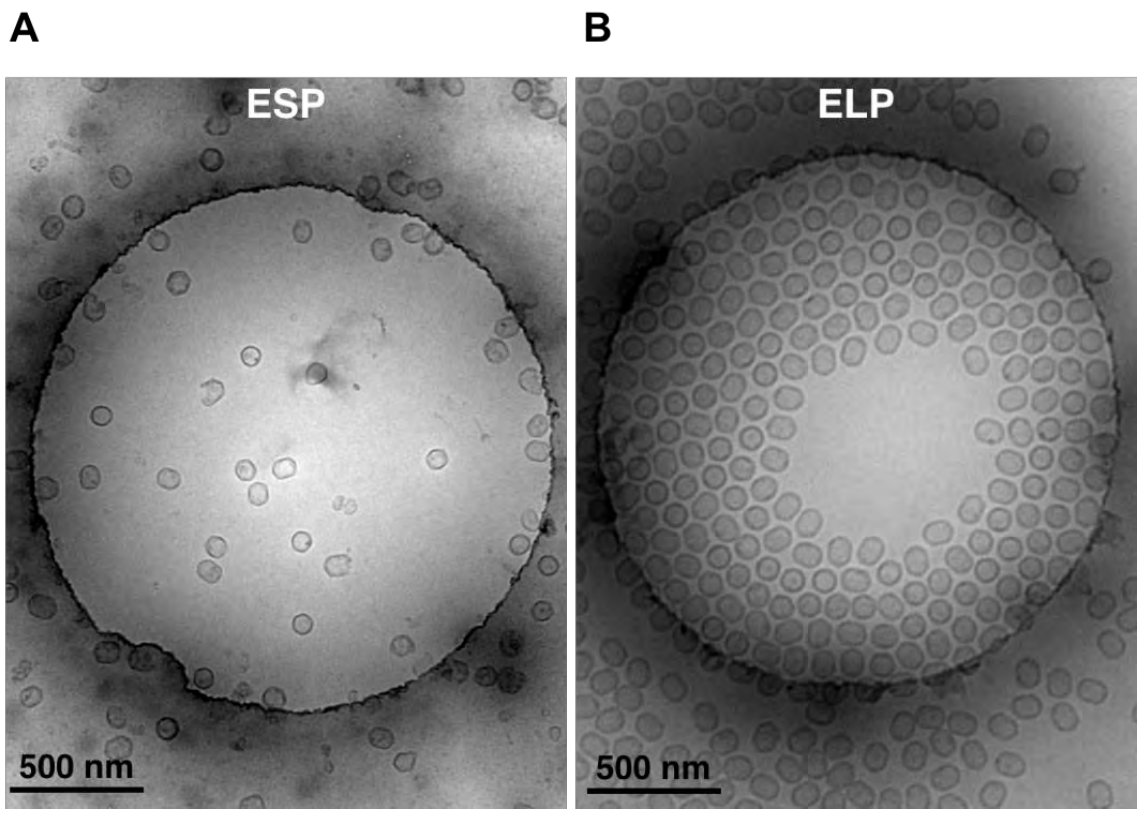


Figure A2.3. Cryo-electron microscopy of ESP and ELP particles. (A) Empty Small Proheads (ESPs) are deformed and heterogeneous in cryoEM conditions. (B) In contrast, Empty Large Proheads (ELP) are homogenous, stable, and adopt multiple orientations in cryoEM conditions. These images were taken at 31,000x magnification.

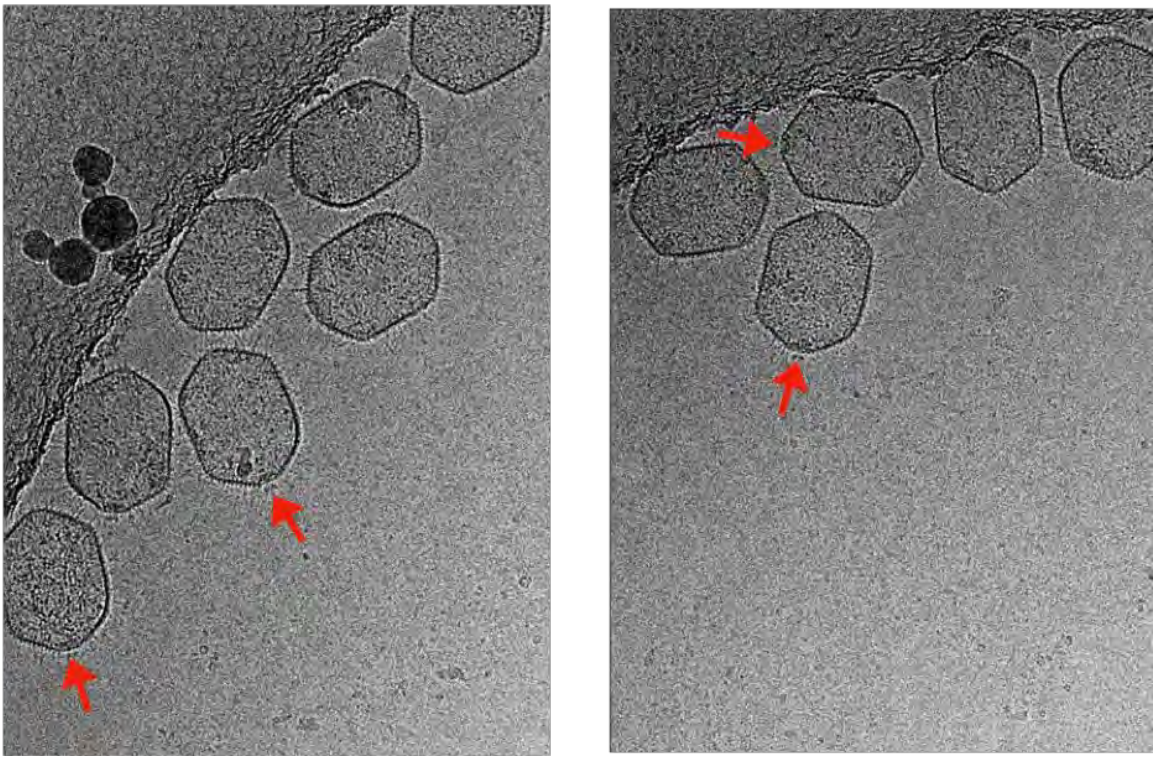
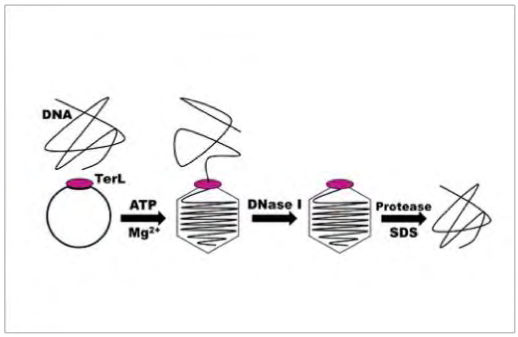
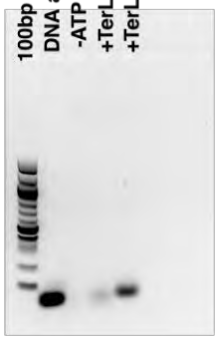
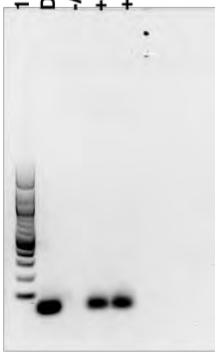
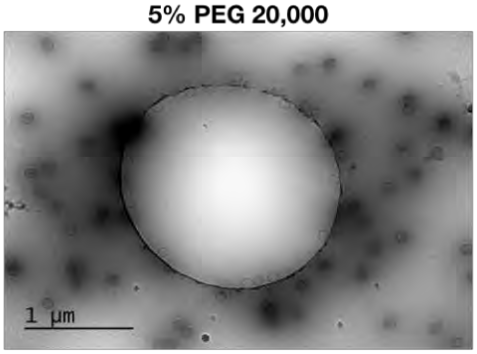


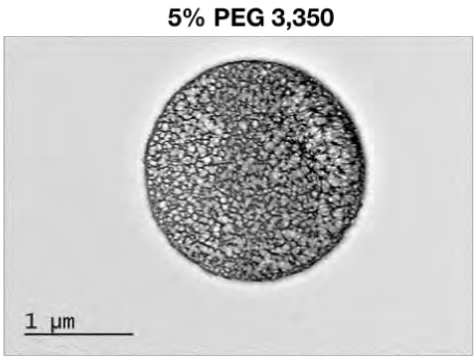
Figure A2.4. High resolution cryo-electron microscopy of ELP particles show details of proheads. Using the Talos Arctica and K2 direct electron detector, we can see density of the portal region (red arrow) on the ELP proheads. The hair-like projections are most likely the T4 HOC protein, a capsid decoration protein.


A
DNA
TerL
ATP
Mg2+
DNase I
Protease
SDS


B
100bp ladder
DNA alone
-ATP
+TerL/ no glycerol
+TerL/ + glycerol


C
100bp ladder
DNA alone
-ATP
+5% PEG 3350
+ 5% PEG 20K


D
5% PEG 20,000
1 µm


E
5% PEG 3,350
1 µm


F
100bp ladder
DNA alone
-ATP
+5% Ficoll
+ 5% PEG 20K

Figure A2.5. Optimization of a cryo-compatible *in vitro* packaging reaction is challenging. (A) *In vitro* packaging assay. Proheads, TerL, DNA, and ATP are incubated to package DNA. DNase I is added to digest unpackaged DNA, and proteinase K and SDS are used to degrade the procapsids and release packaged DNA. (B) *In vitro* packaging assay shows T4 TerL purified without glycerol is significantly less active than TerL with glycerol. (C) PEG 3,350 can be substituted for PEG 20,000 without significant changes in packaging activity. (D) 5% PEG 20,000 pushes procapsids out of the cryo grid holes. (E) 5% PEG 3,350 results in 'dry' grids. (F) Ficoll substitution for PEG 20,000 abrogates packaging activity.

tested whether the T4 TerL protein remained active without glycerol in the purification. Unfortunately, during the short amount of time between purification to experiment (overnight), T4-TerL almost completely loses activity (Figure A2.5B). In addition, we were informed by our collaborator that PEG is necessary to have in the packaging reaction. I tested cryoEM samples with the recommended 5% PEG 20,000 and found that the PEG pushed the particles out of the holes (Figure A2.5D). I was able to obtain equivalent amounts of packaging with PEG 3,350 (Figure A2.5C), but in cryoEM the grids were dry (Figure A2.5E). Substituting PEG for a different crowding reagent, 5% Ficoll, led to no packaging (Figure A2.5F). From these results, I concluded that the packaging reaction has to take place with glycerol and PEG in the buffer, which we would later remove through buffer exchange or dialysis prior to freezing.

T4 packaging buffer exchange assay

Previously published work showed T4 TerL co-pellets with T4 proheads with centrifugation[157]. With this protocol, I attempted to buffer exchange the proheads bound to TerL into a cryo-compatible buffer. I was unable to confirm that TerL co-pellets with T4 capsid because there is a contaminant in the prohead sample that runs at the same molecular weight as T4 TerL on SDS-PAGE (Figure A2.6A). Additionally, is appears that T4 TerL pellets on its own (Figure A2.6B), making it ambiguous whether it actually binds to the capsid in this assay.

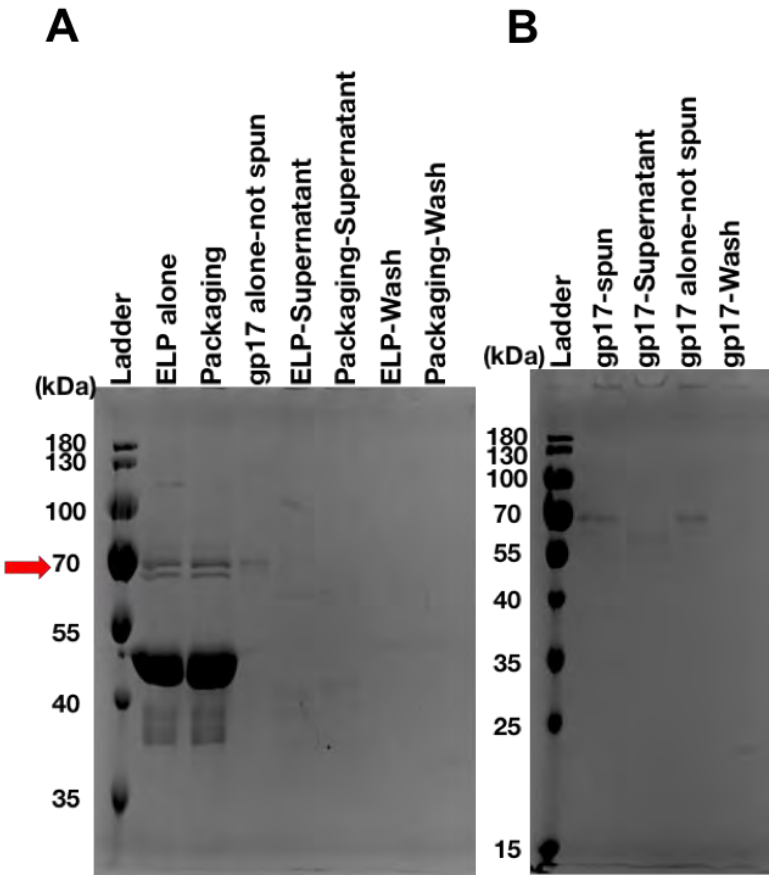


Figure A2.6. Centrifugation buffer exchange protocol test for T4 proheads and T4 TerL (gp17). An *in vitro* packaging assay was performed, and subjected to the centrifugation protocol outlined in Materials and Methods. It is unclear if T4 TerL co-pellets with ELP proheads because (A) there is a significant ELP prohead preparation contaminant than runs at the same size as T4 TerL (red arrow) and (B) T4 TerL sediments on its own.

Y-DNA formulation

It was incredibly challenging to create Y-DNA. The initial construct design consisted of two 90mer oligos, each of which form half of the ‘Y’. In the ‘Y’ stem, there is an EagI cut site, which when digested with EagI restriction enzyme forms an overhang to be annealed and ligated to a DNA ‘leader sequence’ containing a complementary overhang. I decided to order the Y-DNA primers with the EagI cleavage site pre-engineered to eliminate the need for restriction enzyme digestion. To facilitate ligation, I ordered the 5’ overhanging-end primer for the Y-DNA with 5’ phosphorylation, which is not standard for primers ordered from IDT. Regardless of these measures and uncountable attempts, I could not obtain Y ligated to the leader sequence. The leader sequence appears to self-ligate (Figure A2.7A), so I swapped the EagI site for a BbsI restriction site, which is non-palindromic and cannot self-ligate. With the BbsI Y-DNA and leaders, no annealing was observed (Figure A2.7B). Interestingly, with the EagI cut site the ‘Y’ does not self-ligate (Figure A2.7A), leading to the conclusion that the issue stems from Y-DNA formation. Our collaborator had some success with creating the Y-DNA, however the yield of the ligated product after gel recovery was so low that it could not be used in any experiments. From these attempts, we concluded that Y-DNA was not a straight-forward method for stalling the motor, and a non-hydrolyzable analog should be considered more seriously.

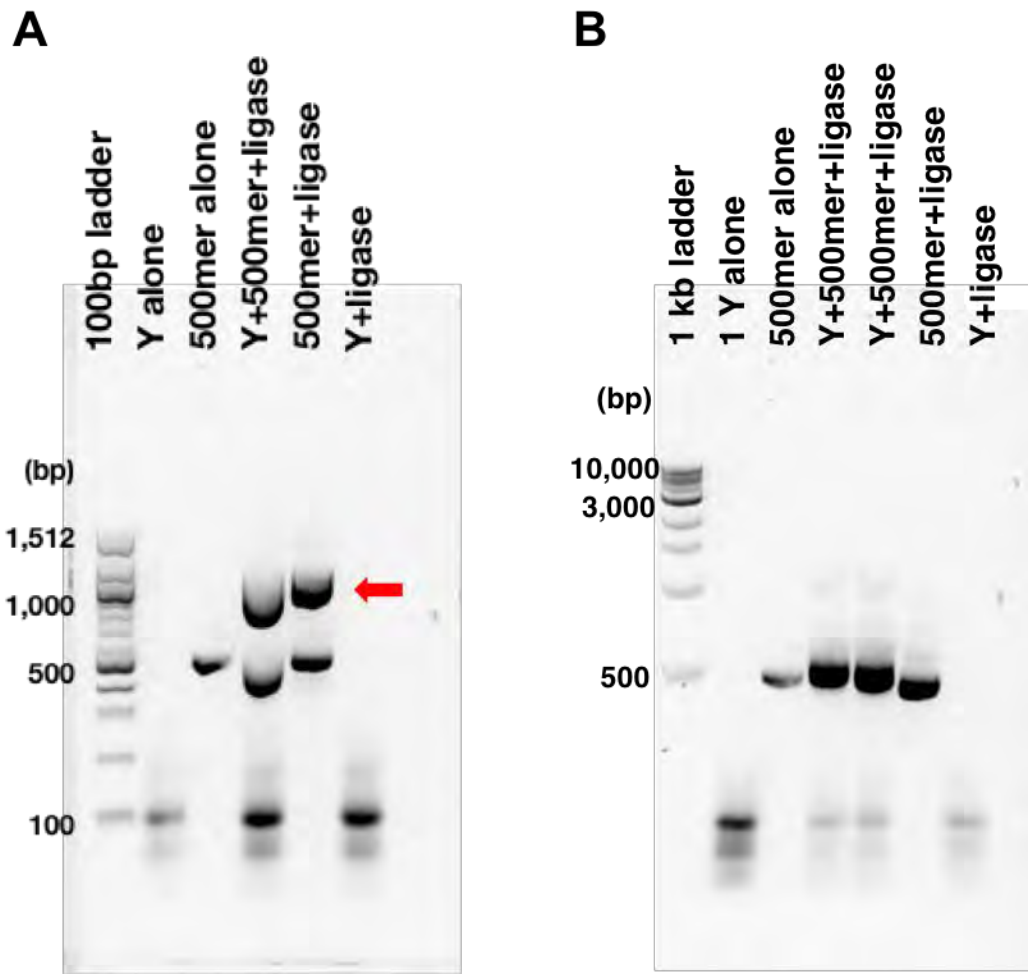


Figure A2.7. Creating Y-DNA proves challenging. (A) EagI digested leader sequence tends to self-ligate (red arrow) but Y-DNA does not. (B) Swapping the EagI site for the non-palindromic BbsI site shows almost no ligation occurs between the Y-DNA and the leader sequence.

Grid optimization for ELP proheads

In many of my cryoEM images I noticed the proheads tend to stick to carbon when capsid concentration is lowered. Since we must dilute the capsids with other reagents to package DNA, we need a way to push the capsids into the grid holes. To do this, I tried coating the carbon on the cryo grids with polylysine with hopes that the capsids would be repelled into the holes. The proheads appeared to be attracted to the polylysine even more, so this reagent was not usable (Figure A2.8A). I next tried polyglutamine, which has the opposite charge of lysine. Unfortunately, polyglutamine denatured the protein and disrupted ice formation (Figure A2.8B). From this work, it appears the samples must be highly concentrated in order to go into the holes, as seen in (Figure A2.3B).

Discussion

Optimizing the T4 samples for cryoEM proved to be a highly challenging project. T4 TerL purifications yield small amounts of protein, and T4 capsid preps are tedious, forcing optimization to be time-consuming and difficult. The main challenges behind this project are (1) removing glycerol and PEG from the packaging reaction to make the samples cryo-compatible (2) creating a reagent to stabilize the motor, such as Y-DNA (3) concentrating the samples enough to have particles in the grid holes. For these obstacles, there are several alternative options that have yet to be tried.

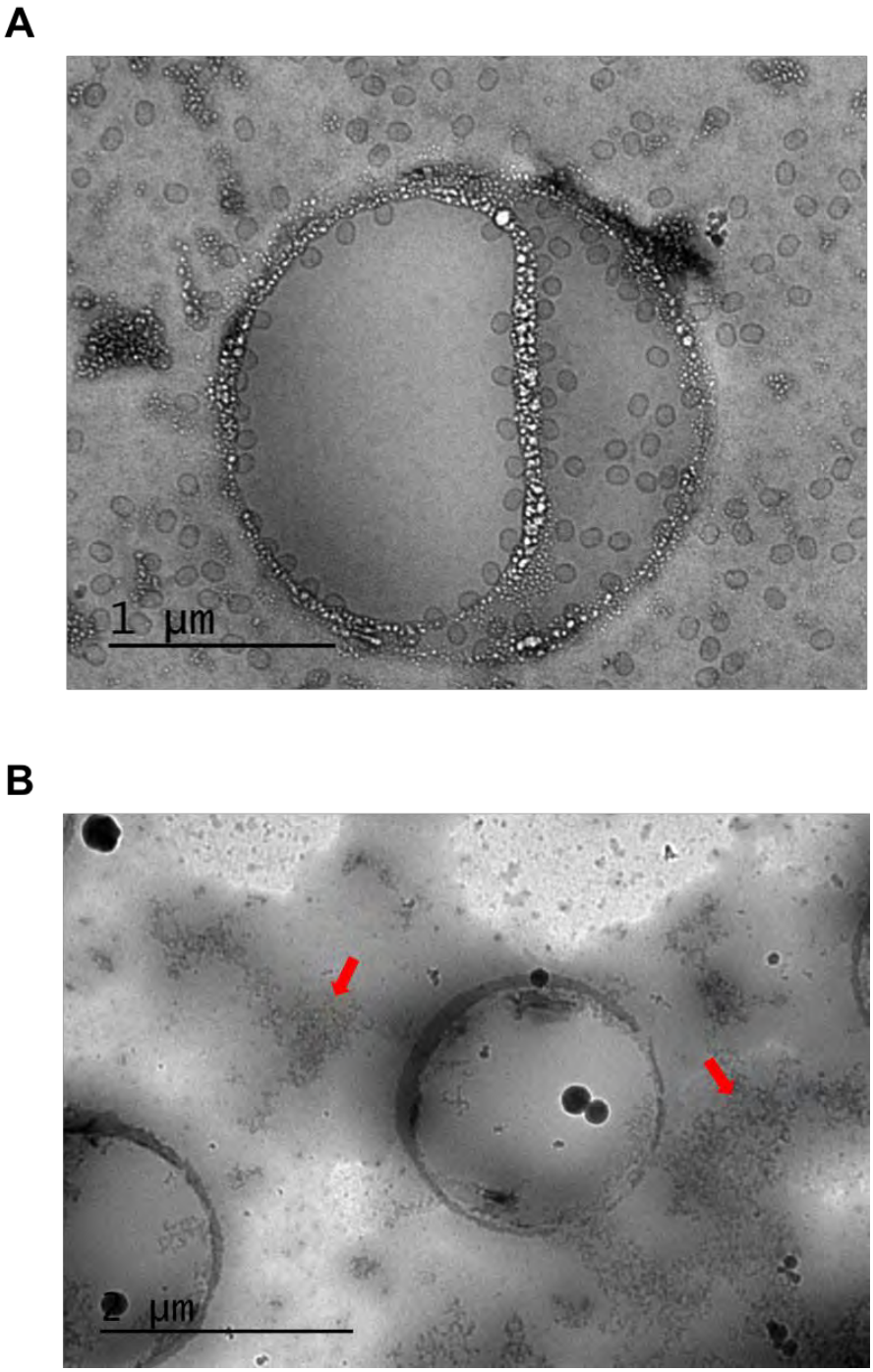


Figure A2.8. Treating cryo-grids to force proheads into the holes. (A) Polylysine treatment appears to attract prohead particles, as seen in this partially polylysine covered hole. (B) Polyglutamine treatment leads to protein denaturation (red arrows) on grids.

In regards to the T4 TerL purification, our collaborators informed us if the protein is used immediately after purification it retains activity without glycerol. It appears that the hours of overnight shipping are enough for T4 TerL to lose activity, making it a highly time sensitive reaction. Our collaborators have offered to travel to UMass to purify the protein onsite in the future, which may solve the issue of having glycerol in the reaction.

Additionally, our collaborators have reported the motor packages DNA with lower amounts of PEG (1-2%), albeit at a lower efficiency. We have not tested these low PEG amounts in cryoEM, and therefore do not know how these conditions behave when vitrified. Therefore, there is potential to further optimize this condition.

In terms of stabilizing the motor, the Y-DNA approach requires optimization to improve DNA yields. However, there are unexplored techniques for locking the motor in place, including slowly-hydrolyzable and non-hydrolyzable ATP analogs (ATPγS, AMP-PNP, ADP•BeF_3), and crosslinking. Crosslinking is a valuable tool if the packaging reaction cannot be optimized without glycerol and PEG. In this case, we will crosslink the packaging complex and use dialysis to remove the glycerol and PEG before imaging. Prior to this, the protein and proheads must be purified in buffers without Tris, which they currently contain, as this is a crosslink quenching reagent for the commonly used crosslinkers. In my opinion, a combination of crosslinker and non-hydrolyzable

ATP analog will be necessary to stall the motor complex and keep it intact for imaging.

For forcing viral particles into the holes for cryoEM imaging, there are other techniques we can try as well. If we use crosslinker and dialysis to remove glycerol and PEG from the packaging reaction, the next step could be simply to concentrate the particles using typical concentration methods. If the particles do not remain intact during concentration, other grids are available for testing. Gold foil grids do not have a carbon coated surface for the particles to stick to, so they may push the particles into the holes. Graphene oxide or carbon-coated grids may also attract particles to the holes. Additionally, for other virus projects in our lab, lacey carbon grids have worked well in the past[240] and may work for our purposes also.

In summary, although there are many challenges with obtaining a cryoEM reconstruction of an actively packaging T4 motor, there are several untested approaches to solve these issues. Future work on this project will test the aforementioned methods with hopes of solving the issues faced in this work.

Appendix III

The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain

Published online 13 January 2017

Nucleic Acids Research, 2017, Vol. 45, No. 6 ***3591–3605***
doi: 10.1093/nar/gkw1356

The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain

Brendan J. Hilbert[1], Janelle A. Hayes[1], Nicholas P. Stone[1], Rui-Gang Xu[2] and Brian A. Kelch[1,*]

[1]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA and [2]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK

Received October 13, 2016; Revised December 21, 2016; Editorial Decision December 22, 2016; Accepted January 05, 2017

ABSTRACT

Many viruses use a powerful terminase motor to pump their genome inside an empty procapsid shell during virus maturation. The large terminase (TerL) protein contains both enzymatic activities necessary for packaging in such viruses: the adenosine triphosphatase (ATPase) that powers DNA translocation and an endonuclease that cleaves the concatemeric genome at both initiation and completion of genome packaging. However, how TerL binds DNA during translocation and cleavage remains mysterious. Here we investigate DNA binding and cleavage using TerL from the thermophilic phage P74-26. We report the structure of the P74-26 TerL nuclease domain, which allows us to model DNA binding in the nuclease active site. We screened a large panel of TerL variants for defects in binding and DNA cleavage, revealing that the ATPase domain is the primary site for DNA binding, and is required for nuclease activity. The nuclease domain is dispensable for DNA binding but residues lining the active site guide DNA for cleavage. Kinetic analysis of DNA cleavage suggests flexible tethering of the nuclease domains during DNA cleavage. We propose that interactions with the procapsid during DNA translocation conformationally restrict the nuclease domain, inhibiting cleavage; TerL release from the capsid upon completion of packaging unlocks the nuclease domains to cleave DNA.

INTRODUCTION

Most double-stranded DNA viruses package their genomes using an adenosine triphosphate (ATP)-dependent motor to pump DNA into an empty capsid protein shell. As DNA fills the shell, internal pressure builds due to confinement of the highly charged DNA. Therefore, these motors have evolved to become some of the most powerful bio-motors known (1,2). For this reason, there is much interest in engineering packaging motors for delivery of nucleic acid therapeutics and as functionalized nano-devices. Moreover, genome packaging motors from herpes viruses are the targets of various Food and Drug Administration approved anti-viral drugs (3–9).

There are two distinct families of packaging motors for membrane-free dsDNA viruses: the terminase family and the Phi29-family motors (10). Here we focus on the more common terminase packaging apparatus, which has been studied in many viral systems (11–18). Terminase motors consist of the portal, large terminase (TerL) and small terminase (TerS) proteins, each of which assembles into a homomeric ring (19,20). Genome packaging by terminases can be broadly summarized as a five-step process (21) (Figure 1). (Step 1) First, the motor recognizes the concatemeric viral genome, primarily through TerS binding (22–26). (Step 2) Next, TerL cleaves the DNA at a specific site and binds to the portal complex. (Step 3) TerL uses ATP hydrolysis to translocate DNA (27–29) through the portal ring into the capsid (30). (Step 4) Upon completing the translocation of at least one genome-length of DNA, TerL switches its enzymatic activity from translocation to cleavage (31). This cleavage occurs either after encapsidating exactly one genome length (termed 'unit-length packaging') or after the capsid is completely filled with DNA, resulting in slightly more than one-genome length being packaged (termed 'headful packaging'). (Step 5) Finally the terminase subunits are released from the capsid for maturation of another virus, while portal binds to the tail proteins to complete a mature, infectious virion (13). Although the sequence of these events has been well studied, the structural mechanism for each step is largely unknown. In particular,

*To whom correspondence should be addressed. Tel: +1 508 856 8322; Fax: +1 508 856 6464; Email: brian.kelch@umassmed.edu

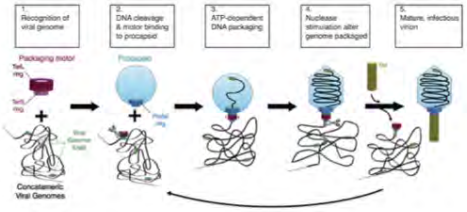


Figure 1. Schematic of a generic genome packaging reaction catalyzed by a terminase enzyme.

how the motor holds DNA during either translocation or cleavage remains obscure.

The TerL protein is the catalytic engine of the packaging apparatus, harboring the two enzymatic activities of the motor: the adenosine triphosphatase (ATPase) that drives DNA translocation, and the endonuclease that cleaves genome concatemers at both initiation and termination of packaging (32,33). The terminase motor is capable of generating high force (stall force up to ~60 pN) and high speeds (up to ~2000 bp/s) (2,34). However, several distinct structural mechanisms have been proposed for both the force generation reaction and DNA cleavage (7,18,35–41). Even more mysterious is the mechanism of nuclease activity regulation; current models for nuclease regulation include TerL auto-inhibition (38,42), catalytic regulation via competition between fast DNA translocation and slow DNA cleavage (37,43), and inhibition mediated by TerS (37,39,42,44). Careful dissection of TerL structure and mechanism is necessary to discern between competing models for TerL activity and regulation.

TerL contains two domains: a C-terminal nuclease domain of a RNaseH fold (7,18,35,38–39,41) and an N-terminal ATPase domain of the ASCE (additional strand, conserved glutamate) superfamily (45). The TerL protein forms an oligomeric ring (46) that binds to the procapsid (35). The oligomeric state for large terminases have been shown to be either tetrameric for lambda (47–49), while the distinct family of Phi29 motors are more controversial as to their oligomeric state (50–53). The C-terminal tail of TerL is thought to interact with the portal and/or the procapsid (32,46,54–59), although an alternate arrangement in which the N-terminal ATPase domain contacts the portal has also been proposed (35,60). Our group recently proposed a model for the TerL ring in which the ATPase domains form a ring with intersubunit contacts contributing to ATP hydrolysis and a small inner pore for binding DNA (46). In this model, the nuclease domains are positioned on the periphery of the TerL ring so that they can use their C-terminal tails to interact with the portal ring. Because the ATPase domains line the inner pore of the ring, this model suggests that the ATPase domains are the primary point of contact for DNA during DNA translocation. However, this model did not explain how DNA cleavage occurs. How does the nuclease domain contribute to DNA binding? Does the interaction surface change during the cleavage reaction?

Here we report enzymatic and structural characterization of TerL DNA binding and nuclease function. We use TerL from the thermophilic phage P74-26 (TerL$^{P74\text{-}26}$) (61) due to its high expression, solubility and stability. Employing a thermophilic terminase affords us a unique opportunity to separately evaluate DNA binding and cleavage, as the latter function only occurs at elevated temperatures. We show that both tight DNA binding and cleavage are nucleotide dependent. Our analysis of cleavage kinetics reveals that dual strand cleavage is fast, suggesting that multiple nuclease domains collaborate to cut both strands of the double helix. We also report the structure of the P74-26 nuclease domain, which we use to map the contributions of individual residues to both DNA binding and cleavage. Our data indicate that the ATPase domain is the primary determinant of DNA binding and that the nuclease domain is dispensable for DNA binding. We integrate our results to propose a mechanism for how TerL switches between DNA translocase and nuclease modes.

MATERIALS AND METHODS

Protein expression and purification

Both the isolated ATPase domain (1–256) and full-length P74-26 TerL mutants were expressed and purified as previously described (46). The nuclease domain (residues 256–485) was subcloned from our previously described pET24a full-length construct and was overexpressed identically to the above constructs (46). Cells were lysed in a cell disruptor and pelleted. The lysate was applied to a 10-ml His-Trap column (GE Healthcare) preequilibrated in buffer A (500 mM NaCl, 20 mM Imidazole, 50 mM Tris pH 8.5, 5 mM βME, 10% (v/v) glycerol). The column was washed with buffers A, followed by buffer A' (150 mM NaCl, 20 mM Imidazole, 50 mM Tris pH 7.5, 5 mM βME, 10% (v/v) glycerol). Protein was eluted with buffer B (150 mM NaCl, 250 mM Imidazole, 50 mM Tris pH 8.5, 5 mM βME, 10% (v/v) glycerol). Eluate was dialyzed into buffer QA (125 mM NaCl, 25 mM Tris pH 7.5, 2 mM DTT, 10% (v/v) glycerol) and the tag was cleaved with prescission protease overnight. Dialysate was loaded onto a 10-ml Q column (GE Healthcare) preequilibrated with buffer QA. The column was then washed with buffer QA. Protein was eluted by applying a 0–100% (v/v) gradient of buffer QA to buffer QB (1 M NaCl, 25 mM Tris pH 7.5, 2 mM DTT, 10% glycerol (v/v)). Eluate was injected onto an S200 HR26/60 (GE Healthcare) column preequilibrated with gel filtration buffer (125 mM NaCl, 25 mM Tris pH 7.5, 4 mM DTT), and eluted in overlapping peaks consistent with dimer (58.2 kDa from gel filtration, 57.4 kDa calculated mass) and monomer (~30 kDa from gel filtration, 28.7 kDa calculated mass). Eluted protein was concentrated to ~20 mg/ml and flash frozen in liquid nitrogen.

Crystallization, structure determination and refinement

Native crystals formed in hanging drops containing 20 mg/ml TerL Nuclease domain mixed 2:1 with buffer containing 0.23 M sodium phosphate monobasic/potassium phosphate dibasic pH 6.2 and 2.5 M sodium chloride and 4 mM dTMP. Crystals were plunged into cryoprotectant containing 0.28 M sodium phosphate monobasic/potassium

phosphate dibasic pH 6.2, 4 M sodium chloride and 2.5 mM dTMP before being flash frozen in liquid nitrogen. Data were collected at the Advanced Light Source at SIBYLS beamline 12.3.1 at wavelength 1.000 Å. Heavy atom derivative crystals were obtained by incubating native crystals with 3 mM potassium hexachloroplatinate 24 h prior to flash freezing. Cryoprotectant for heavy atom derivative crystals contained 3.4 mM potassium hexachloroplatinate. Derivative crystal data were collected at the Advanced Photon Source GM/CA CAT beamline 23ID-B at wavelength 0.855 Å in inverse beam mode. All diffraction data were processed with HKL3000 (62). Platinum bound to Met 265 allowed SAD phasing (63) of the 2.7 Å derivative crystal dataset using the PHENIX autosol pipeline (64). Native dataset anisotropic diffraction data were corrected with the UCLA Diffraction Anisotropy Server (65) and phases were extended to 2.6 Å resolution. Model building and structure refinement were performed with COOT (66) and PHENIX (67). The structure was deposited in the RCSB (PDB code 5TGE).

DNA binding and nuclease digestion

ADP-Beryllium Fluoride (ADP-BeF_3) was formed by incubating 50 mM Tris pH 8.5, 150 mM potassium chloride, 1 mM DTT, 1 mM ADP, 10 mM sodium fluoride, 4 mM beryllium chloride and 10 mM magnesium chloride for 2 h prior to usage. TerL and ADP-BeF_3 were mixed and incubated for 5 min prior to addition of 150 ng of plasmid pET28a (final concentration of 30.3 μM base pairs). Upon DNA addition, samples were incubated at room temperature (DNA binding) or 60°C (nuclease digestion) for 30 min unless otherwise indicated. During kinetics experiments, addition of cold 25 mM ethylenediaminetetraacetic-acid (EDTA) (final) and rapid cooling in an ice bath quenched cleavage. Unless otherwise noted, all DNA cleavage samples were quenched with 1.5% (w/v) sodium dodecyl sulphate (SDS) (final) to prevent TerL's DNA binding activity from perturbing DNA migration through the gel. Standard 1.5% (w/v) agarose Tris-Acetate EDTA pH 8.0 gels were used, with the exception of isolated ATPase domain DNA-binding assay. The pH of the running buffer and gel were raised to pH 8.5 to account for the ATPase domain PI of 8.1. Gels were imaged on an LAS 3000. Gel densitometry was performed using ImageJ (68).

To analyze and rank TerL variant DNA binding, the lanes of all variants were first aligned relative to the DNA ladder on each gel. Migration distance of the most intense DNA bands relative to wild-type migration distance was assessed to rank TerL variants into three general categories. Bands that migrated roughly equivalent to wild-type were designated as unaffected. Likewise bands that migrate equivalent to free plasmid were designated 'severe defect'. Intermediate migration distances were designated 'moderate defect'.

For monitoring DNA cleavage, variants were considered to have wild-type levels of digestion if there was no longer any relaxed or linear plasmid bands remaining. Variants were ranked as 'moderate defect' if the supercoiled band had been cut, but relaxed or linear bands remained. Variants were ranked as 'severe defect' if a supercoiled band remains and there was no significant smearing of degraded, lower molecular weight fragments. Each variant was tested in at least duplicate.

Fitting of the kinetic data was performed using the GraphPad Prism software (GraphPad Inc.) using equations described in Freifelder *et al.* (69) and Cowan *et al.* (70). Our Freifelder-Trumbo analysis utilized the relative fraction, based on densitometry analysis, of supercoiled, relaxed, and linearized plasmid bands to calculate the ratio of single strand nicks to linearization events per molecule during early time points. The fraction of linearized plasmid DNA, f_L, was used to calculate linearization events (n_{linear}) with Equation (1):

$$f_L = n_{\text{linear}} \times e^{(-n_{\text{linear}})} \quad (1)$$

The fraction of supercoiled DNA, f_{Sc}, was used to calculate the single strand nicks (n_{nick}) with Equation (2):

$$f_{Sc} = e^{-(n_{\text{nick}}+n_{\text{linear}})} \quad (2)$$

Simulated data for purely single strand cleavage was calculated from Equation (3) where h represents the maximum base pair distance between two nicks on opposing DNA strands that results in linearization, and L is the total DNA base pairs in the plasmid:

$$n_{\text{linear}} = \frac{\left[(n_{\text{nick}})^2 \times (2h+1)\right]}{4L} \quad (3)$$

If a plasmid has been nicked once, we calculated the probability, R_n, of no additional nicks occurring on the opposing DNA strand within a taboo zone with $2h$ width, that would result in linearization. The taboo zone, b, is expressed as a fraction of the total plasmid, $2h/L$. This probability, Equation (4), yields the percentage of plasmid molecules that have been relaxed by a single nick, but have not yet been linearized given the number of nicks, n and taboo zone fraction size, b.

$$Rn = 2^{1-n} \sum_{k=0}^{n/2} \binom{n}{2k} * (1-kb)^{n-1} \quad (4)$$

The full time course of DNA cleavage was globally fit to the nuclease model described in Cowan *et al.* (70). Supercoiled, relaxed and linearized plasmid band intensities were fit to the following three equations, respectively:

$$Sc\ (t) = Sc_0 \times e^{(-(k_{\text{nicking}}t + k_{\text{cleavage}}t))} \quad (5)$$

$$R\ (t) = \left(1 - e^{(-(R_0+k_{\text{nicking}}t))}\right) \times e^{(-(L_0+k_{\text{cleavage}}t))} \quad (6)$$

$$L\ (t) = \left(L_0 + k_{\text{cleavage}}t\right) \times e^{(-(L_0+k_{\text{cleavage}}t))} \quad (7)$$

Where Sc, R and L are the fractions of supercoiled, relaxed, and linearized plasmid and Sc_0, R_0 and L_0 are the initial fractions of those species. k_{nicking} and k_{cleavage} are the nicking and dual-strand cleavage rates, respectively.

Isolation of P23-45 Phage genomic DNA

An initial stock of phage P23-45 was kindly provided by the Severinov laboratory. A fresh culture of *Thermus ther-*

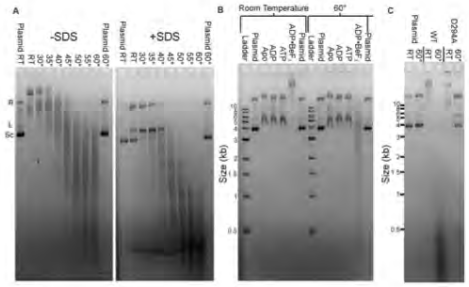


Figure 2. Characterization of TerL$^{P74\text{-}26}$ DNA-binding and nuclease activity. (**A**) Elevated temperature enhances TerL$^{P74\text{-}26}$ nuclease activity. Plasmid DNA migrates slowly when mixed with 15 μM TerL$^{P74\text{-}26}$ (left panel, no SDS) at low temperatures, indicating TerL$^{P74\text{-}26}$ primarily binds DNA at these temperatures. DNA cleavage occurs at higher temperatures indicated by the low molecular weight smearing. Addition of SDS (right panel) reveals minimal cleavage at room temperature with increased intensity of the relaxed (R) and linear (L) plasmid bands, and concomitant decrease in the supercoiled (Sc) band. At temperatures ≥40°C, we observe robust cleavage that increases as the temperature is raised. (**B**) At room temperature, DNA weakly binds to TerL$^{P74\text{-}26}$ in the apo state or when incubated with ADP or ATP. Locking TerL into an 'ATP-bound' state with the non-hydrolyzable analog ADP•BeF_3 results in tight DNA binding. At 60°C TerL cleaves DNA, but only in the presence of ADP•BeF_3. Buffer control samples containing ADP•BeF_3 do not exhibit perturbed plasmid migration (final lane at each temperature). (**C**) Mutation of D294, the conserved nuclease active site residue necessary for metal coordination, to alanine results in a severe loss of nuclease activity (10 μM protein) without affecting TerL$^{P74\text{-}26}$'s affinity for DNA.

mophilus HB8 was grown to an OD_{600} of ∼1.0 in growth medium (0.8% (w/v) Tryptone, 0.4% (w/v) Yeast Extract, 0.3% (w/v) NaCl, 1 mM $MgCl_2$, and 0.5 mM $CaCl_2$) (71). A total of 150 μl of fresh culture was combined with 100 μl P23-45 phage stock at a concentration of 10^6 Plaque Forming Units per ml (PFU/ml) and incubated at 65°C for 10 min. This mixture was then inoculated into 20 ml of fresh growth medium and incubated for 4–6 h at 65°C. The culture was spun at 4000 × *g* for 20 min to remove cell debris. Supernatant (>10^9 PFU/ml) was then treated with DNase I (final concentration, 2 Units/ml) and incubated at 30°C for 30 min. Genomic DNA was extracted from P23-45 phage stocks using the Phage DNA Isolation Kit (Norgen Biotek Corp) according to the manufacturer's protocol.

RESULTS

TerL$^{P74\text{-}26}$ displays robust nuclease activity

We previously established that TerL$^{P74\text{-}26}$ binds DNA at room temperature in the presence of ADP•BeF_3 using an electrophoretic mobility shift assay (EMSA) (46). Because P74-26 phage is a thermophile, we hypothesized that the relatively low temperature (∼20°C) of our previous DNA binding experiments 'masked' the underlying endonuclease activity. Indeed, raising the temperature stimulates TerL$^{P74\text{-}26}$'s nuclease activity, with robust cleavage of plasmid DNA at ∼40°C that accelerates at 60°C (Figure 2A). TerL$^{P74\text{-}26}$ must also be locked in an 'ATP-bound' state by a non-hydrolyzable ATP analog to efficiently cleave DNA; we only observe DNA cleavage with TerL$^{P74\text{-}26}$ preloaded with ADP•BeF_3 and no cleavage in the apo or ADP-loaded states (Figure 2B). Incubation with ATP does not result in DNA cleavage because robust TerL ATPase activity (46) rapidly converts all available ATP to ADP. Therefore, TerL$^{P74\text{-}26}$ needs to be locked in an 'ATP-bound' state for productive cleavage, indicating a strong linkage between the ATPase and nuclease activities.

To exclude the possibility of a co-purified contaminant being responsible for the observed nuclease activity, we mutated an absolutely conserved metal-coordinating active site residue (D294) to alanine. TerL$^{P74\text{-}26}$ is expected to have a two-metal coordinated active site as is found in related nucleases (7,37–39,42,72–73). D294 is the best candidate for mutagenesis because this residue coordinates one of the metal ions in the active site (7,35,37–39). The D294A variant displays an almost complete loss of cleavage activity, while displaying EMSA band shifts characteristic of wild-type DNA binding (Figure 2C). These results illustrate that the nuclease activity is due to TerL$^{P74\text{-}26}$ and not a contaminant. Therefore, isolated TerL$^{P74\text{-}26}$ retains the three critical activities necessary for terminase function: ATPase activity, DNA-binding activity (46) and DNA cleavage.

To our knowledge, TerL$^{P74\text{-}26}$ is the only known large terminase that can both bind and cleave DNA as a full-length protein. The isolated ATPase domain of TerL^{T4} is competent to bind DNA, whereas the full-length protein shows no significant affinity for DNA (74). Other full-length TerL proteins exhibit significant *in vitro* nuclease activity (7,37,39,41,75), although DNA-binding activity is undetectable. Therefore both DNA binding and cleavage can be separately dissected with TerL$^{P74\text{-}26}$.

We next investigated specificity of the DNA binding and cleavage activities of TerL$^{P74\text{-}26}$. As we have shown previously, TerL$^{P74\text{-}26}$ binds DNA with no sequence specificity, as all bands in a 1 kb-ladder are shifted (46). When incubated at 60°C, TerL$^{P74\text{-}26}$ degrades all bands in the 1kb-ladder, as well as negatively supercoiled plasmid, linearized plasmid and other linear fragments (Supplementary Figure S1). Thus, TerL$^{P74\text{-}26}$ both binds and cleaves DNA with no discernible sequence specificity. TerL proteins from other headful packaging phages cleave DNA with no sequence specificity, indicating that TerL$^{P74\text{-}26}$ is similar to most other TerL proteins in this respect.

To investigate the linkage between DNA binding and cleavage, we measured the TerL$^{P74\text{-}26}$ concentration dependence for both activities. DNA binding and cleavage reactions were performed at different temperatures, but otherwise with identical reaction conditions and an identical plasmid substrate. Strikingly, we observe that DNA binding and DNA cleavage exhibit nearly identical dependence on TerL concentration (Figure 3A and B). At TerL$^{P74\text{-}26}$ concentrations ≤2 μM (one TerL monomer for every 15.2 DNA base pairs), no DNA binding or cleavage is observed within the 30 min of the reaction. However, at >2 μM, TerL$^{P74\text{-}26}$ exhibits significant binding at room temperature, with higher TerL$^{P74\text{-}26}$ concentrations resulting in slower DNA migration. Likewise, at 60°C TerL$^{P74\text{-}26}$ concentrations higher than 2 μM result in substantial fragmentation of the plasmid over the 30-min time course. Increasing incu-

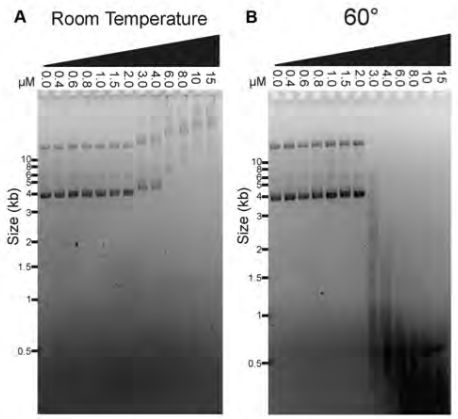


Figure 3. DNA binding and cleavage exhibit identical protein concentration dependence. (**A**) At room temperature, full length TerL$^{P74\text{-}26}$ binds DNA at concentrations greater than 2 μM (no SDS added to the samples). DNA migrates slower, indicating greater binding, as the TerL$^{P74\text{-}26}$ concentration is raised. (**B**) At 60°C, TerL$^{P74\text{-}26}$ cleaves DNA proportional to the degree of DNA binding observed at room temperature. Tighter binding (slower migration) at room temperature corresponds with more complete DNA digestion at 60°C.

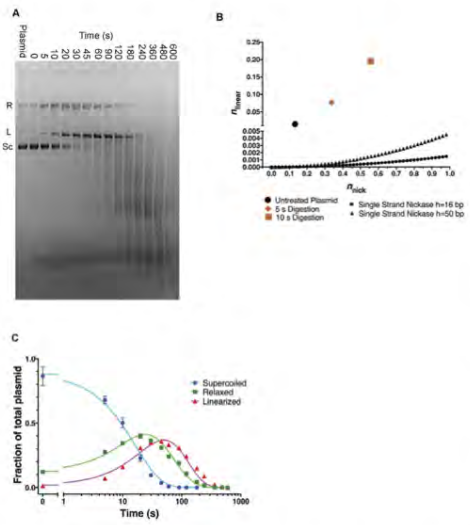


Figure 4. Kinetics of TerL$^{P74\text{-}26}$ nuclease activity. (**A**) A representative gel with 5 μM TerL$^{P74\text{-}26}$ incubated with plasmid at 60°C for different durations. Band intensities for supercoiled (Sc), relaxed (R) and linear (L) species were measured by densitometry of four replicates for Freifelder-Trumbo analysis and kinetic fitting. (SDS present in loading buffer.). (**B**) Freifelder-Trumbo (69) analysis of the untreated plasmid (black circle), 5 s timepoint (orange diamond) and 10 s timepoint (orange square) calculated with Equations (1-2). The simulated single strand nickase curves were calculated with Equation (3) for $h = 16$ bp (black hexagons) or 50 bp (black triangles). (**C**) Nuclease data were globally fit by a kinetic model (Equations 5-7) in which TerL can catalyze both nicking and dual strand cleavage (70). See 'Materials and Methods' section for more details. Residuals to the fit are shown in Supplementary Figure S3C.

bation time to 16 h at 60°C results in substantial, but incomplete DNA cleavage at 2 μM TerL$^{P74\text{-}26}$, with minor cleavage at 1.5 μM (Supplementary Figure S2). The coincident TerL dependencies of DNA binding and cleavage activities indicate that these two functions are tightly linked. Moreover, the steepness of the activity transition suggests that a cooperative process drives both DNA binding and cleavage. Band smearing and multiple TerL$^{P74\text{-}26}$ binding sites per plasmid prevent accurate measurement and fitting to calculate a K_d. However, binding occurs roughly at concentrations consistent with a K_d in the low μM range. The affinity of TerL$^{P74\text{-}26}$ for DNA is therefore similar to that measured for the Lambda phage TerL protein (K_d ~3–4 μM) (49), indicating that TerL$^{P74\text{-}26}$ is consistent with known terminase enzyme function.

Kinetic analysis of TerL$^{P74\text{-}26}$ cleavage

To further investigate the mechanism of TerL nuclease activity, we followed the kinetics of plasmid cleavage. We chose a TerL$^{P74\text{-}26}$ concentration of 5 μM due to the potent activity observed during the 30-min reaction. By measuring the intensities of the supercoiled, relaxed, and linear plasmid bands over a 10-min time course, we can distinguish between single-strand versus dual-strand cleavage (see below). To accurately quantify each band, we added SDS to our gel-loading buffer to prevent TerL$^{P74\text{-}26}$'s DNA binding activity from perturbing DNA migration. We observe a rapid loss of supercoiled DNA ($t_{1/2}$~20 s), whereas the relaxed and linearized plasmid bands increase and then decrease in intensity (Figure 4A). The relaxed plasmid increases in intensity (peak ~30 s), and is then degraded until it is undetectable at ~480 s. Linearized DNA increases with no observable lag until ~60 s, and then is degraded to form a smear, indicating more substantial fragmentation.

The rapid rise in the linearized fraction of DNA suggests TerL$^{P74\text{-}26}$ employs a dual-strand cleavage mechanism. The plasmid banding pattern closely resembles a mixed, dual-strand and single-strand cutting mechanism of DNaseI (76). In the presence of magnesium, DNaseI can only nick DNA. However, in the presence of manganese DNaseI can also cleave both strands, as well as produce single nicks. Campbell *et al.* used a similar plasmid digestion assay to characterize DNaseI activity (76). They observed a rapid rise in the fraction of linearized DNA in presence of manganese, as DNaseI's dual-strand cleavage activity quickly linearized the DNA. However, in the presence of magnesium, there is a significant lag in the appearance of linearized plasmid, because double-strand breaks only occur when DNaseI produces two nicks nearby each other on opposing strands. Because our plasmid size (5372 bp) is similar to that of Campbell *et al.* (5224 bp), the density of nicks and thus the cleavage patterns are comparable between the two experiments. We observe qualitatively similar cleavage pat-

terns for TerL$^{P74\text{-}26}$ plasmid digestion as for DNaseI•Mn^{2+} (Figure 4A), with no lag observed for the appearance of linearized plasmid. We therefore hypothesized that TerL$^{P74\text{-}26}$ both nicks and directly cleaves both DNA strands.

Quantifying the fractions of supercoiled, relaxed, and linearized DNA supports a dual strand cleavage mechanism. Freifelder-Trumbo analysis uses the proportion of supercoiled, relaxed, and linearized plasmid species to distinguish between dual strand cleavage versus the accumulation of a sufficient number of random nicks to result in eventual linearization (69). This analysis has previously determined the cleavage mechanism for other nucleases or DNA degrading small molecules (77–81). Because Freifelder-Trumbo analysis is limited to time points with low levels of fragmentation, we used only early time points (up to 10 s) for our Freifelder-Trumbo analysis where plasmid DNA has not yet been measurably fragmented (Supplementary Figure S3A). The number of DNA linearization events (n_{linear}) and nicks (n_{nick}) and were calculated from Equations (1) and (2) respectively. We simulated curves from a purely single-stranded cutting enzyme with Equation (3), with L base pairs in the plasmid (5372 bp) and h, the maximum base pair distance between two nicks on opposing DNA strands that results in linearization. The analyzed time points exhibit a proportion of double strand breaks that is >400 fold higher than that predicted for a random, purely single strand cutting mechanism for $h = 16$ bp (69) (Figure 4B). Therefore, we conclude that TerL$^{P74\text{-}26}$ has significant dual strand cleavage activity.

Because plasmid fragmentation limited us to a few data points for Freifelder-Trumbo analysis, we sought to verify our conclusions by investigating possible sources of error in our calculations. First, intercalation of ethidium bromide into supercoiled DNA differs from that of relaxed and linear DNA, often requiring a correction factor to adjust for a diminished fluorescent signal (82,83). Although we could not calculate the correction factor for our plasmid substrate, increasing the supercoiled signal intensity by a factor of 1.4, similar to other observations, does not perturb the results or affect our conclusion of dual strand cleavage (Supplementary Figure S3B and Table S1). Second, we adjusted the distance between two single strand nicks on opposing strands (h in Equation (3)) required for a double-strand break. Although linearization due to nicks 50 bp apart is physically unreasonable, increasing the maximum distance between opposing strand nicks from the standard 16 (69) to 50 bp does not change the result that TerL$^{P74\text{-}26}$ linearizes plasmid DNA two-orders of magnitude faster than predicted for a pure nicking mechanism (Figure 4B and Supplementary Table S1).

Because Freifelder-Trumbo analysis indicates TerL can directly cleave both DNA strands, we further investigated the nuclease mechanism by fitting the full TerL$^{P74\text{-}26}$ experimental data to a mechanism of mixed single-strand and dual-strand cleavage. We use a set of equations derived by Cowan *et al.* to describe this situation (70). The data were fit to Equations (5-7) (70). This model assumes dual strand cleavage is rapid enough that single strands cannot sufficiently accumulate as to cause substantial linearization. The Freifelder-Trumbo analysis supports this assumption as the linear to relaxed proportions of DNA are well above a simulated pure single strand cleaving mechanism. The fitted parameters include the nicking rate ($k_{nicking}$) and dual-strand cleavage rate ($k_{cleavage}$), as well as the initial fractions of the supercoiled (S_0), relaxed (R_0) and linearized (L_0) plasmid bands.

The high quality fit of the data (R^2 for global fit = 0.986; Figure 4C; Supplementary Figure S3C and Table S2) to the dual-strand linearization model confirms that TerL$^{P74\text{-}26}$ can catalyze dual strand cleavage. The nicking rate (0.0408 $\pm$ 0.0025 s^{-1}) is roughly twice that of the dual strand cleavage rate (0.0197 $\pm$ 0.0007 s^{-1}). These values suggest a plasmid would only accumulate two nicks on average for every linearization event. Importantly, after making an initial nick to relax a plasmid, Equation (4) allows us to calculate how many random single strand breaks a nicking enzyme would have to make for a 50% population of linearized plasmid (70). Given a taboo zone of 16 bp (69) on either side of the initial nick, plasmids would need to accumulate 23 nicks to form a 50% mixture of both relaxed and linearized plasmid molecules. Furthermore, if we increase the taboo zone to 50 bp on either side of the initial nick, as we did in the Freifelder-Trumbo analysis above, 50% of the plasmid would be linearized only after 13–14 nicks. Thus, the amount of nicks required for a single strand cutter to linearize 50% of a relaxed plasmid is more than an order of magnitude greater than our observed rate of nicks per linearization event. Only a dual-strand cleavage mechanism can account for the observed rapid linearization.

Because we observe rapid plasmid cleavage, we sought to determine whether a larger and physiologically relevant substrate would exhibit substantially different kinetics. A digestion time course with the genome of related phage P23-45 (92% identical between nucleotide sequence of the P23-45 and P74-26 genomes; 99.8% identity between amino acid sequence of P23-45 and P74-26 TerL proteins (71)) showed a similar rate of fragmentation (Supplementary Figure S3D). Only dual strand cleavage explains the rapid digestion of the phage genome. A pure nicking mechanism would result in significantly slower fragmentation of the 85 kb genome substrate relative to the 5.4 kb plasmid. Dual-strand cleavage however would rapidly fragment the genome as observed, further supporting our conclusion of dual-strand cleavage activity. Moreover, this result suggests that the DNA sequence alone cannot alter TerL$^{P74\text{-}26}$ mediated cleavage. Altogether, the evidence of dual-strand cleavage has important impacts on how the nuclease functions in the TerL protein (see 'Discussion' section).

Although our analysis of cleavage kinetics indicates that TerL performs both dual strand cleavage and single-strand nicking, this simple model may not perfectly describe the mechanism. The Freifelder-Trumbo analysis and fitting to the model described by Cowan *et al.* assume that binding and cleavage is random. We observe a weak banding pattern in the fragmented plasmid (Figure 4A and Supplementary Figure S2) suggesting TerL may have a mild preference for binding and/or cleaving at specific points in the plasmid. However, TerL$^{P74\text{-}26}$ binds (46) and cleaves numerous different substrates with no observable sequence specificity (Supplementary Figure S1). Given that TerL therefore binds and cleaves all tested DNA sequences, we assume that plasmid binding and cleavage remains mostly random. Given the ra-

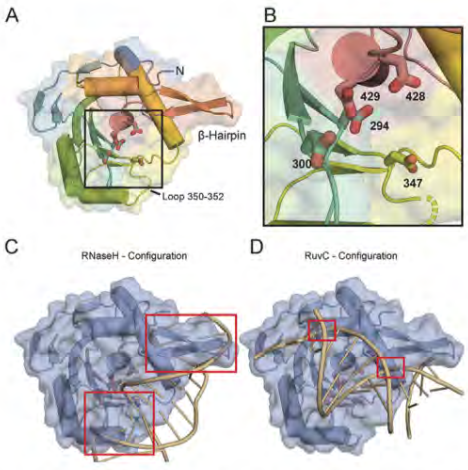


Figure 5. The TerL^{P74-26} nuclease domain structure. (**A**) Overall features of the TerL^{P74-26} nuclease domain structure. The electron density is missing for both a flexible loop at residues 350–352 (Gly-Val-Gly; dotted lines) and for C-terminal residues 450–485. Potential metal-coordinating active site residues (294, 300, 347, 428 and 429) are represented with sticks. A β-hairpin unique to the terminase family extends away from the nuclease domain. (**B**) Zoomed view of the nuclease active site shows the acidic residues for metal coordination. (**C**) The RNase H configuration of DNA bound the TerL^{P74-26}-ND was created by aligning an RNase H structure bound to a RNA/DNA hybrid duplex (85) to the TerL^{P74-26}-ND using default parameters in Chimera (87). The RNA:DNA duplex clashes (red boxes) with TerL^{P74-26}-ND. There are severe clashes in the regions around metal-coordinating residue 347 and the β-hairpin. (**D**) The RuvC configuration of DNA bound to TerL^{P74-26}-ND was generated using the structure of RuvC resolvase bound to a Holliday junction (86). Clashing is minimal, occurring at G424 and the side chains of R421 and R425. The flexibility of these residues suggests they may change conformation to accommodate DNA binding.

tio of linearized to relaxed molecules is >400 fold higher than a pure single strand cutter, and the goodness of the fit to the kinetic model, if there is a minor sequence preference, its impact is minimal.

Structure of TerL^{P74-26} nuclease domain

To gain insight into the structural mechanism of DNA cleavage by TerL^{P74-26}, we solved the structure of the TerL^{P74-26} nuclease domain (hereafter, TerL^{P74-26}-ND) to 2.6 Å resolution (Figure 5A). We obtained experimental phases from single-wavelength anomalous diffraction of a platinum derivative (see 'Materials and Methods' section) (Table 1). The overall fold of TerL^{P74-26} -ND is similar to those of other terminase nuclease domains (7,35,38–41), with an average C_α RMSD of 2.0 Å (for individual C_α RMSDs, see Supplementary Table S3).

Several high-resolution structures of the large terminase nuclease domain for highly related phage, G20C, are solved in the accompanying article from Xu *et al.* The protein sequences for TerL^{G20C}-ND and TerL^{P74-26}-ND are nearly identical, and differ only at residue 315 (G20C A315 versus P74-26 V315). TerL^{G20C}-ND crystallized in three crystal forms that are distinct from TerL^{P74-26}-ND. Overall, the structures of TerL^{G20C}-ND and TerL^{P74-26}-ND complement one another to provide key insight into TerL nuclease structure and function.

Table 1. Crystallographic data and refinement statistics

Data Collection	
Space group	P 43 21 2
Wavelength	1
Resolution range	47.51-2.60
Unit cell angles (°)	71.37 71.37 127.32
Unit cell dimensions (Å)	90 90 90
Total reflections	299 519
Unique reflections	10 480 (862)
Multiplicity	28 (28.8)
Completeness %	98 (98)
Mean I/sigma I	23.3 (6.8)
Wilson B factor	31.9
R-merge	0.089 (0.638)
R-meas	0.090 (0.650)
R-pim	0.017 (0.121)
CC1/2	0.999 (0.999)
CC*	1.000 (0.999)
Refinement	
R-work %	21.8
R-free %	24.7
RMS bonds	0.003
RMS angles	0.52
Ramachandran favored %	94
Ramachandran outliers %	0
Rotomer outliers %	0.62
Clashscore	4.05
Average B	38

The active site of TerL^{P74-26} -ND contains several metal-coordinating residues that are conserved across the terminase family. Because no divalent cations were added during purification and crystallization, we do not observe any metal coordination in the active site. As expected from structures of other TerL nuclease domains (7,35,38–41), we observe D294 in the heart of the active site accompanying several other metal-coordinating residues (D294, D300, D347, D428, D429) in TerL^{P74-26} (Figure 5A and B). The accompanying article by Xu *et al.* discusses specific residue interactions during metal coordination in detail for TerL^{G20C}-ND. The positions of acidic residues in TerL^{P74-26} most closely resemble the arrangement observed in phage T4/RB49 (Supplementary Figure S4), which binds metal with residues equivalent to TerL^{P74-26} D294, D347 and D429 (35,37).

Beyond the active site residues, there is remarkably little sequence conservation within the nuclease domain across the TerL family. The relative lack of conservation may reflect the fact that different viruses use varied strategies for cleaving DNA (21,84). Because of this lack of sequence conservation, identifying how DNA accesses the nuclease active site has been particularly challenging.

To address how DNA is positioned in the TerL nuclease active site, we compared the TerL nuclease domain to distantly related nuclease structures for which there are structures of substrates bound. In particular, the structure of human RNaseH bound to an RNA:DNA hybrid (85) and a more recent structure of *T. thermophilus* RuvC resolvase

bound to a Holliday junction (86) provide two different possibilities for the DNA orientation in the nuclease active site. By superposing (87) these two structures with that of TerL$^{P74\text{-}26}$-ND, we can model potential DNA interaction modes. Superposition of the RNaseH:RNA–DNA structure positions the DNA helix along a surface that extends from a flexible loop at residues 350–352, across the active site toward the N-terminus of the nuclease domain (Figure 5C). However, as has been noted previously (38–39,41), this positioning clashes with a β-hairpin that is present in all TerL proteins but is absent in other known members of the RNaseH superfamily of nucleases (38). Therefore, the β-hairpin must considerably flexible in order to accommodate DNA in the RNaseH configuration. Because of this substantial clash, the β-hairpin has been proposed to play an auto-regulatory role in controlling nuclease activity (38). In contrast, superposition of RuvC suggests an orthogonal DNA orientation (Figure 5D). Importantly, the β-hairpin does not produce a significant clash with modeled DNA but would instead provide a surface for cradling the DNA as it crosses the active site. Thus, the two RNaseH and RuvC models predict different roles for the β-hairpin: the RuvC model predicts that the β-hairpin assists in DNA cleavage while the RNaseH model predicts that the β-hairpin inhibits cleavage.

To further investigate the role of the nuclease domain in terminase function, we used the TerL$^{P74\text{-}26}$ nuclease structure to identify residues that may be important for DNA binding and DNA cleavage. We selected conserved or semi-conserved basic residues that are predicted to contact DNA, based on previous predictions of DNA binding surfaces (7,18,35,38–39,41) and our comparisons with RNaseH and RuvC. Combined with variants in the ATPase domain that we previously generated to study ATP hydrolysis and DNA binding (46), our panel includes 23 point mutations across both domains (Supplementary Table S4). We also used our isolated TerL$^{P74\text{-}26}$ ATPase domain (TerL$^{P74\text{-}26}$-AD) and TerL$^{P74\text{-}26}$-ND constructs to examine the overall role of each domain in TerL function. By separately measuring DNA-binding and nuclease cleavage for each variant, we provide critical insight into how DNA is bound and cleaved during viral genome packaging.

The ATPase domain is the primary DNA binding region

To assess how DNA binds to TerL$^{P74\text{-}26}$ we first focused on the ATPase domain, as we previously observed a complete loss of DNA binding from the R101E mutation (46). R101 is in a patch of basic residues along one surface of the ATPase domain that we predict forms the DNA binding surface within the pore of the assembled TerL ring (46). Here we extend this analysis to other residues across the surface of the ATPase domain, including other residues in the 'basic patch'. Three of the mutations in the basic patch (R102A, R104E and R128A) also display a complete loss of DNA binding, while the final basic patch variant (R121E) does not significantly affect binding (Figure 6A; Supplementary Figures S5 and S6A). To verify that these DNA-binding effects are due to specific disruption of the DNA binding interface, we tested variants with mutations predicted to be outside of the pore of the TerL ring (R58A and R170A). Neither of these mutations severely affects DNA binding affinity (Figure 6A). These results support our previous conclusion that this basic patch is critical for gripping DNA.

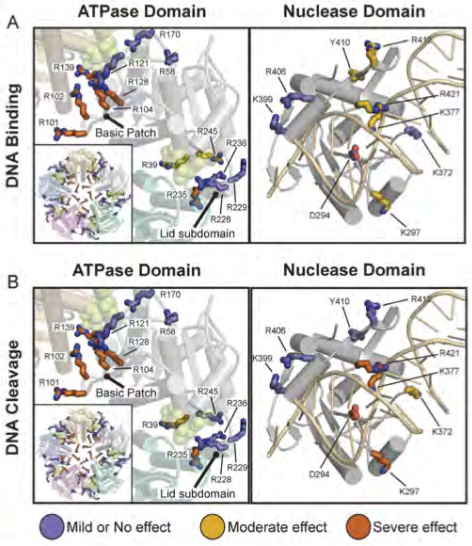


Figure 6. Mapping of residues important for DNA binding and cleavage. (**A**) Mutation of ATPase domain (left panel) basic patch residues severely inhibits DNA binding, whereas nuclease domain variants do not severely impact binding (right panel). DNA binding was assessed as described in the 'Materials and Methods' section. Variants were separated into three categories, as indicated by color. The ATPase domain is shown in the context of our ATPase ring model (46) to illustrate how mutational effects match the ring topology (green spheres indicate ADP•BeF_3 in the ATPase active site). Mutations that severely inhibit binding (orange) tend to be in the pore of the ATPase model (left panel inset) including the basic patch (R101E, R102A, R104E or R128A). Interfacial residues (R139A and R235A) also inhibit binding. (Mutational mapping of the ATPase ring model can be viewed in greater detail in Supplementary Figure S6.) The nuclease structure is shown with DNA in the RuvC configuration. Nuclease domain variants fail to significantly inhibit DNA binding. (**B**) Mapping of residues important for DNA cleavage. Variants were ranked as described in 'Materials and Methods' section. Residues in the ATPase that are important for binding are likewise critical for DNA cleavage (left panel). DNA binding is therefore a prerequisite for effective cleavage. Nuclease domain residues predicted to interact with DNA in the RuvC configuration (K297, K377 and R421) are critical for nuclease activity, suggesting that DNA binds in a similar orientation. Nuclease metal coordinating variant D294A serves as a negative control. Ranking of each residue's contribution to DNA binding and cleavage is shown in Supplementary Table S4.

Because TerL$^{P74\text{-}26}$ needs to be locked into an ATP-bound state to tightly grip DNA (Figure 2B), we next investigated how mutations in or near the ATPase active site affect DNA binding. R39 is a conserved residue in the P-loop of the active site and directly contacts the γ-phosphate group of ATP (46). R139 is the *trans*-acting arginine finger that is critical for ATP hydrolysis (46). We also tested several residues (R228, R229, R235, R236, R245) in the Lid subdo-

main, a region that caps the active site and changes conformation upon ATP hydrolysis and release (Figure 6A) (46). The R228A, R229A and R236A variants have no apparent effect on DNA binding, while the R39A and R245A variants exhibit a moderate decrease in DNA binding. Only the R139A and R235A variants display severe defects in DNA binding. Because the *trans*-acting arginine finger R139 and Lid subdomain residue R235 are important for both ATP hydrolysis and interactions between adjacent ATPase subunits (46), we hypothesize that the DNA binding defects observed with these variants is due to the severe loss of both ATP binding and/or ring assembly.

We next focused on the role of the nuclease domain in DNA binding. We mutated residues in the active site D294A and K377A), the β-hairpin (Y410A, R412A and R421E), and other regions that have been predicted in other structural studies of phages T4 (35) and Sf6 (18) to bind DNA (K297A, K372A, K399A and R406A). Interestingly, none of the mutations in the nuclease domain severely impact DNA binding (Figure 6A). Variants K297A and K377A display a moderate loss of affinity for DNA. Similarly, the mutations in the β-hairpin display modest defects. Neither R412 nor R421 are conserved in the β-hairpin, but structures of terminase nuclease domains often exhibit basic residues in similar locations (35,38–40), suggesting that basic residues may play some role in function. A third β-hairpin mutation (Y410A), designed to disrupt the β-hairpin structure, only moderately affected DNA binding. Overall these results indicate that the nuclease domain is not a primary determinant for high affinity DNA binding.

Because our panel of point mutants highlights the importance of the ATPase domain in binding DNA, we next investigated whether isolated domains bind DNA. TerL$^{P74\text{-}26}$-AD binds DNA at similar concentrations as full-length TerL$^{P74\text{-}26}$ (Figure 7A and Supplementary Figure S7). Interestingly, TerL$^{P74\text{-}26}$-AD binds DNA independent of nucleotide, indicating that the nuclease domain is important for the ATP-dependent regulation of DNA binding. In contrast, TerL$^{P74\text{-}26}$-ND does not detectably bind DNA, even at concentrations >30-fold higher than the approximate K_d for full-length TerL$^{P74\text{-}26}$ binding (Figure 7B). Therefore, the ATPase domain is necessary and sufficient for TerL to bind DNA, an event that is a prerequisite for nuclease activity.

Identifying the requirements for nuclease activity

We next examined our panel of variants to determine the role of individual residues on the DNA cleavage reaction. We find a strong correlation between DNA binding and nuclease activity across all ATPase mutants. Mutations that abrogate DNA binding likewise inhibit DNA cleavage, while mutations in the ATPase domain that do not disrupt DNA binding have no effect on nuclease activity. Specifically, mutations in the ATPase domain's basic patch (R101E, R102A, R104E and R128A) or the active site (R39A, R139A, R235A) show a severe loss of nuclease function (Figure 6B; Supplementary Figures S5 and S6B). These results support our finding that the isolated TerL$^{P74\text{-}26}$-ND fails to bind and cleave DNA. Therefore binding and cleavage hinge on the ATPase domain's ability to bind DNA.

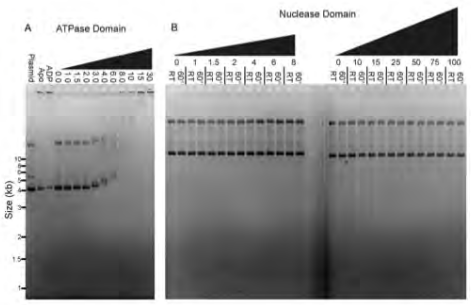


Figure 7. The ATPase domain is necessary and sufficient for DNA binding. (**A**) The isolated TerL$^{P74\text{-}26}$-AD binds DNA with the same affinity as full length TerL$^{P74\text{-}26}$ independent of ADP•BeF_3. A total of 15 μM TerL$^{P74\text{-}26}$-AD slows migration of a significant portion of plasmid DNA in the apo form or the presence of ADP. In the presence of ADP•BeF_3, DNA migration slows with TerL$^{P74\text{-}26}$-AD concentrations above 2 μM with decreasing migration proportional to the rise in protein concentration. Coomassie staining confirms the presence of TerL$^{P74\text{-}26}$-AD co-migrating with DNA (Supplementary Figure S7). (**B**) The isolated TerL$^{P74\text{-}26}$-ND neither binds nor cleaves DNA, even at concentrations 30-fold higher than where we observe binding and cleavage for full-length TerL$^{P74\text{-}26}$. DNA binding appears necessary for effective cleavage.

Interestingly, a subset of nuclease domain mutants disrupts DNA cleavage without severely impairing DNA binding. We observe a severe loss of nuclease activity in the K297A, K377A and R421E variants, and a moderate decrease in activity in the K372A variant. K297A, K377A and R421E are the only variants in full-length TerL$^{P74\text{-}26}$ where DNA binding remains relatively unperturbed yet nuclease activity is severely impacted. As mentioned previously, the RuvC and RNaseH binding modes predict very different behavior for several of the variants, particularly those with mutation in the β-hairpin. The RuvC binding mode predicts a favorable role for the β-hairpin in DNA cleavage, while the RNaseH mode predicts that the β-hairpin plays an auto-inhibitory role. Overall, cleavage defects in these variants are consistent with the DNA contacts predicted by the RuvC-like model of DNA binding (Figure 5D). In particular, R421 is on the face of the β-hairpin predicted to interact favorably with DNA in the RuvC binding mode; this residue is necessary for DNA cleavage activity (Figure 6B). In contrast, Y410 and R412 are on the opposite face of the β-hairpin and both are dispensable for nuclease activity. Regardless of whether the RuvC or RNaseH binding modes are correct for TerL nuclease engagement, our results demonstrate that cleavage depends on two factors, (i) the ability of TerL to bind DNA as dictated primarily by the ATPase domain, and, (ii) a specific set of nuclease domain residues predicted to position DNA for cleavage.

DISCUSSION

Proper terminase function in viral genome packaging requires precise spatiotemporal coordination and regulation of ATP hydrolysis, DNA binding and nucleolytic cleavage. Each of these functions must be individually examined in

order to piece together a packaging mechanism. We previously built a low-resolution structural model of a pentameric TerL ring that accurately predicted the position of the arginine finger (R139 in TerL^{P74-26}), as well as a key DNA-binding residue (R101 in TerL^{P74-26}) (46). We observed ATP-dependent conformational changes in the ATPase domain, suggesting that the Lid subdomain generates the force for DNA translocation through a lever-like motion. In this study we provide insights into terminase function that can then be applied to improve the existing models of genome packaging and further our understanding of one of nature's most powerful bio-motors.

The TerL ATPase domain tightly grips DNA

The TerL ATPase domain is indispensable for DNA binding. This conclusion is based on two major observations. First, we observe strong DNA binding with both full-length TerL^{P74-26} and the isolated ATPase domain. In contrast, the isolated nuclease domain of TerL^{P74-26} does not detectably bind or cleave DNA. Thus, the ATPase domain is both necessary and sufficient for DNA binding. Second, mutation of basic patch residues (R101, R102, R104 or R128) in full-length TerL^{P74-26} abrogates both DNA binding and cleavage. Conversely, none of the mutations located in the nuclease domain severely impact DNA binding, despite the fact that several residues are critical for DNA cleavage. These results suggest that the bulk of TerL affinity for DNA derives from the ATPase domain, as predicted by our previous model (46).

Several models for terminase:DNA binding predict a larger role of the nuclease domain in gripping DNA during translocation (18,35). Because we can separately measure DNA binding and cleavage with TerL^{P74-26}, we are able to directly test these predictions. Surprisingly we find that residues within the nuclease domain only make a small contribution to DNA affinity (Figure 6A and Supplementary Figure S5) and that the entire domain is dispensable for tight binding (Figure 7A and Supplementary Figure S7). Moreover, two semi-conserved residues (K399 and R406) in a region predicted to bind DNA (35) show no role in binding or cleaving DNA (Figure 6A). Therefore, we favor a model in which the ATPase domain is the primary DNA grip during both translocation and cleavage modes, and that the nuclease domain only engages DNA during genome cleavage (Figure 8). Although unlikely, it is possible that free TerL^{P74-26} is locked into 'cleavage mode' and uses the nuclease domain for gripping DNA when in 'translocation mode'. However, we do not favor this model because we observe no measurable affinity between the nuclease domain and DNA (Figure 7B), similar to results seen with T4-TerL (74). Moreover, the isolated ATPase domain, which is unlikely to be locked into 'DNA cleavage mode', displays tight DNA binding, as shown here (Figure 7A) and elsewhere for TerL^{T4} (74).

By separating the primary DNA gripping region from the nuclease active site, terminases have evolved an efficient means for regulating nuclease activity. First, the nuclease domain's low intrinsic DNA-binding affinity appears to be important for proper nuclease regulation. Although the nuclease active site must bind DNA with at least weak affinity

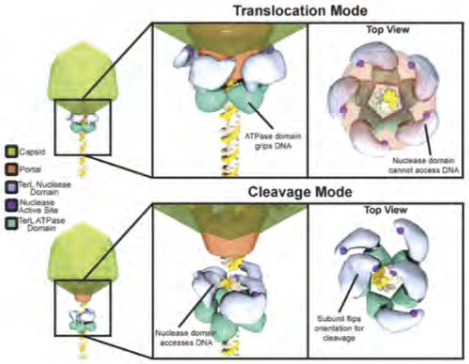


Figure 8. Proposed model for nuclease regulation. During 'translocation mode' the nuclease domain active site is sequestered from DNA by interactions of the TerL with portal and capsid, preventing premature cleavage. The ATPase domain serves as the sole surface for gripping DNA during packaging. Upon completion of packaging TerL enters 'cleavage mode'. TerL dissociates from the portal and capsid, releasing the inhibition of the nuclease domains. The ATPase domains remains tightly bound to DNA. The nuclease domains rearrange to cleave each of the antiparallel DNA strands. Although depicted as a blunt cut, cleavage could also leave overhangs depending on how both nuclease domains engage DNA.

in order to cleave, this affinity must be carefully balanced; TerL would catalyze spurious cleavage if the affinity were too strong, but would cleave inefficiently if the affinity were too weak. There appears to be a spectrum of intrinsic DNA binding affinities for TerL nuclease domains. Isolated T4- (37,74), Sf6- (18,40), CMV- (7) and HSV-TerL (41) nuclease domains cleave DNA, implying a modest affinity. On the other hand, isolated SPP1- (38), P22- (39) and P74-26-TerL nuclease domains fail to cleave DNA. Secondly, the flexible nature between the ATPase and nuclease domains allows the allosteric regulation of the nuclease. The ATPase domain places the nuclease domain in high local concentration with DNA to overcome the nuclease domain's intrinsically weak affinity for DNA. Moreover, by altering the position of the nuclease active site relative to DNA, terminase enzymes can easily regulate DNA cleavage (see below).

Kinetic analysis reveals details of TerL nuclease activity

Kinetic analysis of the TerL^{P74-26} nuclease activity reveals a mechanism of TerL cleavage consistent across the terminase family. TerL^{P74-26} rapidly cleaves supercoiled plasmid, with concomitant increases in both relaxed and linearized plasmid, followed by complete fragmentation. This pattern of cleavage qualitatively resembles a mechanism involving dual strand cleavage observed in DNaseI (76). Regarding the terminase family, our results also mirror those observed for T4 phage and CMV TerL-catalyzed nuclease activity (7,42). Similarly, increasing concentrations of HSV-1 TerL nuclease domain results in an initial increase in relaxed and linearized plasmid, followed by near complete digestion at high nuclease concentrations (41). These results were in-

terpreted as single strand nicking that eventually results in linearization and complete degradation (7,41). In contrast, our quantitative approach using Freifelder-Trumbo analysis and kinetic fitting of the nuclease reaction (Figure 4B and C) reveals significant dual strand cleavage by $TerL^{P74\text{-}26}$. This result places large constraints on the arrangement of the TerL nuclease domains during DNA cleavage. Because the qualitative cleavage data is consistent across the family, we propose that these constraints are universal to all terminases.

From a mechanistic perspective, dual strand cleavage requires flexibility of the nuclease domain relative to the ATPase domain. Endonucleases require two active sites arranged in an anti-parallel fashion for simultaneous dual strand cleavage (88,89). As previously suggested (37,42), the TerL nuclease domains would assume conformations roughly 180° relative to one another for their active sites to align with each of the antiparallel DNA strands. Therefore, there is significant flexibility between the ATPase and nuclease domains of TerL to allow this rearrangement. Indeed, there is much previous data to support this assertion. Limited proteolysis of the TerL proteins from P74-26 (Supplementary Figure S8), T4 (32) and P22 (39) indicates that the linker connecting the ATPase Lid subdomain to the nuclease domain is highly flexible. Additionally, crystal structures of TerL proteins from T4 and Sf6 show very different orientations of the nuclease domain relative to the ATPase domain (18,35). Therefore, we propose that the TerL ATPase domain ring tightly grips DNA while the TerL nuclease domain is flexibly tethered to adopt the necessary orientation for DNA cleavage.

How does dual strand cleavage occur? We envision two possibilities for cleavage of both strands: (i) a monomer of TerL cleaves both strands in rapid succession, or (ii) two subunits within a TerL oligomer cleave each strand contemporaneously. In the first mechanism, after cleaving the Watson strand, the nuclease domain of the TerL monomer must rapidly reorient by ~180° to cleave the Crick strand. In the second mechanism, two separate nuclease domains within a TerL oligomer can adopt orthogonal orientations to efficiently cleave both strands. Both mechanisms require a large degree of flexibility between the tightly bound ATPase domain and the nuclease domain. We do not observe a dimer oriented for dual strand cleavage within the nuclease domain crystal lattice (Supplementary Figure S9). Although we cannot decisively rule out dual strand cleavage by a monomer, we favor cleavage by a TerL oligomer for two reasons. First, the steep dependence on TerL concentration for both DNA binding and cleavage implies cooperative assembly of a TerL oligomer on DNA (Figure 3). Second, TerL requires ATP for both DNA binding and cleavage (Figure 2B), which implies that the interfacial contacts afforded by ATP binding (46) promote oligomerization on DNA. It is formally possible that a second TerL ring binds DNA in the opposite orientation (head to head) to the first ring during the cleavage reaction. This would require forming a large interface between rings as well as regulating the second ring binding to ensure it occurs only after capsid filling. We disfavor this model as there is no evidence for a head-to-head interaction or binding of a second ring at the end of packaging.

Regulation of TerL nuclease activity

During translocation, TerL nuclease activity must be inhibited to prevent premature cleavage. Two non-mutually exclusive possibilities could explain how TerL nuclease activity is regulated: (i) the 'kinetic competition model', wherein the rate of TerL ATP hydrolysis and DNA translocation significantly outpaces the rate of DNA cleavage until translocation slows upon maximal packaging, allowing cleavage to occur, and (ii) the 'steric block model', in which portal and/or TerS regulate the accessibility of the nuclease active site for DNA. We discuss these two possibilities below.

In the kinetic competition model the relative rates of ATP hydrolysis and DNA cleavage self-regulate TerL cleavage (37,43). If DNA translocation is much faster than the rate of cleavage, then the nuclease active site cannot stably engage DNA long enough for cleavage. Packaging would progress until the rate of DNA translocation sufficiently slows near the end of packaging. As the rates of translocation and cleavage become similar, the nuclease domain has enough time to engage a segment of DNA for successful cleavage. In the kinetic competition model, the rates of ATP hydrolysis and nuclease activity must be precisely balanced to prevent premature genome cleavage. However, two lines of evidence suggest that kinetic competition is not the regulatory mechanism. First, motor stalling events regularly occur during packaging in phages T4 (2) and Lambda (34,90) for periods of time up to ~5 s with no reported cleavage of DNA. Second, long-term motor stalls can been artificially induced in phage T4 with no significant cleavage occurring over the time scale of hours (57,91).

In the steric block model, inhibition is achieved by restricting the accessibility of the nuclease active site for DNA. Multiple lines of evidence indicate that the TerL C-terminal tail binds to the portal (32,46,54–59). We propose that this interaction with portal locks the nuclease domain in an orientation that prevents the nuclease active site from accessing DNA, thereby inhibiting premature genome cleavage during packaging. Upon completion of genome packaging, portal transmits a 'headful signal' that is thought to trigger DNA cleavage (92). We hypothesize that the headful signal facilitates dissociation of TerL from portal, releasing the nuclease domain from this restricted conformation, allowing the flexibly tethered nuclease domains to reorient and doubly cut DNA. Similarly, TerS can inhibit TerL-nuclease activity (37,39,42,44); therefore, similar contacts with TerS could sterically block TerL nuclease activity.

How does DNA engage the nuclease active site? We use both the structure of RNaseH bound to an RNA/DNA hybrid (85) or RuvC bound to a Holliday junction (86) to model how DNA accesses the active site. DNA binds in orthogonal orientations in these structures, leading to distinct predictions for behavior of some of the variants tested here. We favor a model of DNA binding similar to RuvC. TerL is more similar to RuvC than RNaseH in terms of structure and the surface of TerL is a better steric fit to the RuvC DNA orientation rather than that of RNaseH. As noted here and in other studies (38,39), the β-hairpin would clash with DNA in the RNaseH-like configuration (Figure 5C). These results have raised the question of whether the

β-hairpin adjusts its conformation to allow for productive access to the active site or if DNA is bound in a different orientation to the RNaseH model. In contrast, the model based on RuvC does not result in any substantial clash between the DNA and the protein (Figure 5D). Instead, the β-hairpin is positioned such that it can make favorable interactions along the DNA backbone. Importantly, our mutagenesis results are most consistent with the orientation of DNA predicted from the RuvC model. A β-hairpin residue predicted to directly form a salt bridge with the DNA backbone of the scissile strand (R421) is critical for nuclease activity, but two residues on the opposite face of the hairpin (Y410 and R412) are dispensable. In addition, three other residues in our panel of mutations are also predicted to interact with the scissile strand in the RuvC-like model (K297, K377 and R412); all three are critical for nuclease activity (Figure 6B). Furthermore, K372 is predicted to interact with the non-scissile strand, and the K372A variant shows a modest defect in nuclease activity. The phenotypes of the D294A, K297A and K377A (important for cleavage) and K399A and R406A (no role in cleavage) variants match both the RuvC and RNaseH orientations and do not effectively discriminate between the two models. Future studies will map the interactions with DNA in greater detail.

Our findings do not agree with previously proposed models of TerL DNA binding and cleavage. Translocation models for T4 (35) and Sf6 (18) both propose the nuclease domain as a primary site of DNA binding during packaging. Conversely, our TerL$^{P74\text{-}26}$ variant and domain analyses clearly demonstrate that the ATPase domain is the primary site of DNA binding. First, the TerL^{T4} model posits that residues R517, and R524 bind the DNA backbone during translocation (35). However, the equivalent residues in TerL$^{P74\text{-}26}$ (K399 and R406) are dispensable for binding. Second, the Sf6 model contends that the nuclease domain collaborates with the Lid subdomain (also referred to as the 'Linker subdomain') to grip DNA (18). However, key residues in the Sf6 structure that were predicted to bind DNA are either not conserved (R194, R305, R306 and K328 in Sf6) or show no binding defect when mutated in TerL$^{P74\text{-}26}$ (K192/R193, and R200 in Sf6; R236 and R245 in P74-26). Other residues predicted by the Sf6 model only have a slight effect; K372 in TerL$^{P74\text{-}26}$ (K323 in Sf6) is dispensable for DNA binding and only plays a minor role in DNA cleavage. A residue not predicted by the Sf6 model to bind DNA (K360 or R361 in Sf6; K377 in P74-26) is absolutely critical for DNA cleavage and has a moderate effect on DNA binding. Therefore, our data is not consistent with these two models, because we find that the ATPase domain is the primary site of DNA binding. However, it is possible that isolated TerL$^{P74\text{-}26}$ is locked in 'cleavage' mode, whereas these residues would have a measurable role in 'translocation' mode.

DNA must bind to the nuclease domain for the DNA cleavage reaction to occur, even if this binding is weak and/or transient. Our data are consistent with two of the residues that were predicted to be important for DNA cleavage in the TerL^{T4} and TerL^{Sf6} models. Residues flanking the active site are not necessary for full DNA binding affinity but are critical for DNA cleavage (R406 in T4 (35), which is K297 in P74-26; K416 in Sf6 (18), which is R412 in P74-26). The models therefore share some similarity for the TerL cleavage mechanism.

Terminases are conserved across many different families of dsDNA viruses, including human pathogens of Herpesviridae (11). There are several FDA-approved drugs on the market that target the terminase motor, and it is thought that these drugs' mode-of-action is through inhibition of the terminase's nuclease activity (3–9). Our studies reveal important aspects of the TerL nuclease mechanism and regulation, and provide a blueprint for future mode-of-action studies of small molecule inhibitors of terminase enzymes from human pathogens.

ACCESSION NUMBER

PDB Code: 5TGE.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank the Severinov laboratory (Rutgers University) for providing a sample of phage P23-45. We thank the Schiffer, Royer and Ryder labs for use of instrumentation and for helpful discussions. We thank members of the Antson Lab (University of York) for sharing data before publication and for thoughtful discussions. We thank beamline scientists at SIBYLS 12.3.1 (Lawrence Berkeley National Laboratory) and APS 23-ID-B (Argonne National Laboratory, award number GUP-39936) for technical support with x-ray diffraction data collection.

FUNDING

Pew Charitable Trusts Scholarship (to B.A.K.); Lawrence Berkeley; Argonne [GUP-39936]. Funding for open access charge: Pew Charitable Trust.
Conflict of interest statement. None declared.

REFERENCES

1. Smith,D.E., Tans,S.J., Smith,S.B., Grimes,S., Anderson,D.L. and Bustamante,C. (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature*, **413**, 748–752.
2. Fuller,D.N., Raymer,D.M., Kottadiel,V.I., Rao,V.B. and Smith,D.E. (2007) Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 16868–16873.
3. Buerger,I., Reefschlaeger,J., Bender,W., Eckenberg,P., Popp,A., Weber,O., Graeper,S., Klenk,H.D., Ruebsamen-Waigmann,H. and Hallenberger,S. (2001) A novel nonnucleoside inhibitor specifically targets cytomegalovirus DNA maturation via the UL89 and UL56 gene products. *J. Virol.*, **75**, 9077–9086.
4. Bogner,E. (2002) Human cytomegalovirus terminase as a target for antiviral chemotherapy. *Rev. Med. Virol.*, **12**, 115–127.
5. Dittmer,A., Drach,J.C., Townsend,L.B., Fischer,A. and Bogner,E. (2005) Interaction of the putative human cytomegalovirus portal protein pUL104 with the large terminase subunit pUL56 and its inhibition by benzimidazole-D-ribonucleosides. *J. Virol.*, **79**, 14660–14667.

6. Hwang,J.S., Kregler,O., Schilf,R., Bannert,N., Drach,J.C., Townsend,L.B. and Bogner,E. (2007) Identification of acetylated, tetrahalogenated benzimidazole D-ribonucleosides with enhanced activity against human cytomegalovirus. *J. Virol.*, **81**, 11604–11611.
7. Nadal,M., Mas,P.J., Blanco,A.G., Arnan,C., Sola,M., Hart,D.J. and Coll,M. (2010) Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 16078–16083.
8. Goldner,T., Hewlett,G., Ettischer,N., Ruebsamen-Schaeff,H., Zimmermann,H. and Lischka,P. (2011) The novel anticytomegalovirus compound AIC246 (Letermovir) inhibits human cytomegalovirus replication through a specific antiviral mechanism that involves the viral terminase. *J. Virol.*, **85**, 10884–10893.
9. Melendez,D.P. and Razonable,R.R. (2015) Letermovir and inhibitors of the terminase complex: a promising new class of investigational antiviral drugs against human cytomegalovirus. *Infect. Drug Resist.*, **8**, 269–277.
10. Burroughs,A.M., Iyer,L.M. and Aravind,L. (2007) Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn.*, **3**, 48–65.
11. Baines,J.D. and Weller,S.K. (2005) Cleavage and Packaging of Herpes Simplex Virus 1 DNA. In: Catalano,CE (ed.), *Viral Genome Packaging: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA, pp. 135–150.
12. Feiss,M. and Catalano,C.E. (2005) Bacteriophage Lambda Terminase and the Mechanisms of Viral DNA Packaging. In: Catalano,CE (ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA, pp. 5–39.
13. Rao,V.B. and Black,L.W. (2005) DNA Packaging in Bacteriophage T4. In: Catalano,CE (ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA, pp. 40–58.
14. Casjens,S. and Weigele,P. (2005) DNA Packaging by Bacteriophage P22. In: Catalano,CE (ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA, pp. 80–88.
15. Serwer,P. (2005) T3/T7 DNA Packaging. In: Catalano,CE (ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA, pp. 59–79.
16. Droge,A. and Tavares,P. and (2005) Bacteriophage SPP1 DNA Packaging. In: Catalano,CE (ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA, pp. 89–101.
17. Zhao,H., Finch,C.J., Sequeira,R.D., Johnson,B.A., Johnson,J.E., Casjens,S.R. and Tang,L. (2010) Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 1971–1976.
18. Zhao,H., Christensen,T.E., Kamau,Y.N. and Tang,L. (2013) Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 8075–8080.
19. Rao,V.B. and Feiss,M. (2008) The bacteriophage DNA packaging motor. *Annu. Rev. Genet.*, **42**, 647–681.
20. Feiss,M. and Rao,V.B. (2012) The bacteriophage DNA packaging machine. *Adv. Exp. Med. Biol.*, **726**, 489–509.
21. Catalano,C.E. (2005) *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. 1st edn., Springer, New York City, USA.
22. Shinder,G. and Gold,M. (1988) The Nul subunit of bacteriophage lambda terminase binds to specific sites in cos DNA. *J. Virol.*, **62**, 387–392.
23. Casjens,S., Sampson,L., Randall,S., Eppler,K., Wu,H., Petri,J.B. and Schmieger,H. (1992) Molecular genetic analysis of bacteriophage P22 gene 3 product, a protein involved in the initiation of headful DNA packaging. *J. Mol. Biol.*, **227**, 1086–1099.
24. Chai,S., Lurz,R. and Alonso,J.C. (1995) The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J. Mol. Biol.*, **252**, 386–398.
25. Lin,H. and Black,L.W. (1998) DNA requirements in vivo for phage T4 packaging. *Virology*, **242**, 118–127.
26. Wu,H., Sampson,L., Parr,R. and Casjens,S. (2002) The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. *Mol. Microbiol.*, **45**, 1631–1646.
27. Morita,M., Tasaka,M. and Fujisawa,H. (1993) DNA packaging ATPase of bacteriophage T3. *Virology*, **193**, 748–752.
28. Rao,V.B. and Mitchell,M.S. (2001) The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4. *J. Mol. Biol.*, **314**, 401–411.
29. Goetzinger,K.R. and Rao,V.B. (2003) Defining the ATPase center of bacteriophage T4 DNA packaging machine: requirement for a catalytic glutamate residue in the large terminase protein gp17. *J. Mol. Biol.*, **331**, 139–154.
30. Valpuesta,J.M. and Carrascosa,J.L. (1994) Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Q. Rev. Biophys.*, **27**, 107–155.
31. Bhattacharyya,S.P. and Rao,V.B. (1993) A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. *Virology*, **196**, 34–44.
32. Kanamaru,S., Kondabagil,K., Rossmann,M.G. and Rao,V.B. (2004) The functional domains of bacteriophage t4 terminase. *J. Biol. Chem.*, **279**, 40795–40801.
33. Duffy,C. and Feiss,M. (2002) The large subunit of bacteriophage lambda's terminase plays a role in DNA translocation and packaging termination. *J. Mol. Biol.*, **316**, 547–561.
34. Fuller,D.N., Raymer,D.M., Rickgauer,J.P., Robertson,R.M., Catalano,C.E., Anderson,D.L., Grimes,S. and Smith,D.E. (2007) Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.*, **373**, 1113–1122.
35. Sun,S., Kondabagil,K., Draper,B., Alam,T.I., Bowman,V.D., Zhang,Z., Hegde,S., Fokine,A., Rossmann,M.G. and Rao,V.B. (2008) The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell*, **135**, 1251–1262.
36. Black,L.W. (2015) Old, new, and widely true: the bacteriophage T4 DNA packaging mechanism. *Virology*, **479-480**, 650–656.
37. Alam,T.I., Draper,B., Kondabagil,K., Rentas,F.J., Ghosh-Kumar,M., Sun,S., Rossmann,M.G. and Rao,V.B. (2008) The headful packaging nuclease of bacteriophage T4. *Mol. Microbiol.*, **69**, 1180–1190.
38. Smits,C., Chechik,M., Kovalevskiy,O.V., Shevtsov,M.B., Foster,A.W., Alonso,J.C. and Antson,A.A. (2009) Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep.*, **10**, 592–598.
39. Roy,A. and Cingolani,G. (2012) Structure of p22 headful packaging nuclease. *J. Biol. Chem.*, **287**, 28196–28205.
40. Zhao,H., Lin,Z., Lynn,A.Y., Varnado,B., Beutler,J.A., Murelli,R.P., Le Grice,S.F. and Tang,L. (2015) Two distinct modes of metal ion binding in the nuclease active site of a viral DNA-packaging terminase: insight into the two-metal-ion catalytic mechanism. *Nucleic Acids Res.*, **43**, 11003–11016.
41. Selvarajan Sigamani,S., Zhao,H., Kamau,Y.N., Baines,J.D. and Tang,L. (2013) The structure of the herpes simplex virus DNA-packaging terminase pUL15 nuclease domain suggests an evolutionary lineage among eukaryotic and prokaryotic viruses. *J. Virol.*, **87**, 7140–7148.
42. Ghosh-Kumar,M., Alam,T.I., Draper,B., Stack,J.D. and Rao,V.B. (2011) Regulation by interdomain communication of a headful packaging nuclease from bacteriophage T4. *Nucleic Acids Res.*, **39**, 2742–2755.
43. Cue,D. and Feiss,M. (1997) Genetic evidence that recognition of cosQ, the signal for termination of phage lambda DNA packaging, depends on the extent of head filling. *Genetics*, **147**, 7–17.
44. Sun,S., Gao,S., Kondabagil,K., Xiang,Y., Rossmann,M.G. and Rao,V.B. (2012) Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 817–822.
45. Leipe,D.D., Koonin,E.V. and Aravind,L. (2003) Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.*, **333**, 781–815.
46. Hilbert,B.J., Hayes,J.A., Stone,N.P., Duffy,C.M., Sankaran,B. and Kelch,B.A. (2015) Structure and mechanism of the ATPase that powers viral genome packaging. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E3792–E3799.
47. Maluf,N.K., Gaussier,H., Bogner,E., Feiss,M. and Catalano,C.E. (2006) Assembly of bacteriophage lambda terminase into a viral

DNA maturation and packaging machine. *Biochemistry*, **45**, 15259–15268.
48. Andrews,B.T. and Catalano,C.E. (2012) The enzymology of a viral genome packaging motor is influenced by the assembly state of the motor subunits. *Biochemistry*, **51**, 9342–9353.
49. Yang,T.C., Ortiz,D., Nosaka,L., Lander,G.C. and Catalano,C.E. (2015) Thermodynamic Interrogation of the Assembly of a Viral Genome Packaging Motor Complex. *Biophys. J.*, **109**, 1663–1675.
50. Morais,M.C., Koti,J.S., Bowman,V.D., Reyes-Aldrete,E., Anderson,D.L. and Rossmann,M.G. (2008) Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor. *Structure*, **16**, 1267–1274.
51. Mao,H., Saha,M., Reyes-Aldrete,E., Sherman,M.B., Woodson,M., Atz,R., Grimes,S., Jardine,P.J. and Morais,M.C. (2016) Structural and molecular basis for coordination in a viral DNA packaging motor. *Cell Rep.*, **14**, 2017–2029.
52. Cao,S., Saha,M., Zhao,W., Jardine,P.J., Zhang,W., Grimes,S. and Morais,M.C. (2014) Insights into the structure and assembly of the bacteriophage 29 double-stranded DNA packaging motor. *J. Virol.*, **88**, 3986–3996.
53. Zhao,Z., De-Donatis,G.M., Schwartz,C., Fang,H., Li,J. and Guo,P. (2016) An arginine finger regulates the sequential action of asymmetrical hexameric ATPase in the double-stranded DNA translocation motor. *Mol. Cell. Biol.*, **36**, 2514–2523.
54. Yeo,A. and Feiss,M. (1995) Specific interaction of terminase, the DNA packaging enzyme of bacteriophage lambda, with the portal protein of the prohead. *J. Mol. Biol.*, **245**, 141–150.
55. Morita,M., Tasaka,M. and Fujisawa,H. (1995) Analysis of the fine structure of the prohead binding domain of the packaging protein of bacteriophage T3 using a hexapeptide, an analog of a prohead binding site. *Virology*, **211**, 516–524.
56. Lin,H., Rao,V.B. and Black,L.W. (1999) Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. *J. Mol. Biol.*, **289**, 249–260.
57. Dixit,A., Ray,K., Lakowicz,J.R. and Black,L.W. (2011) Dynamics of the T4 bacteriophage DNA packasome motor: endonuclease VII resolvase release of arrested Y-DNA substrates. *J. Biol. Chem.*, **286**, 18878–18889.
58. Dixit,A.B., Ray,K. and Black,L.W. (2012) Compression of the DNA substrate by a viral packaging motor is supported by removal of intercalating dye during translocation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 20419–20424.
59. Dixit,A.B., Ray,K., Thomas,J.A. and Black,L.W. (2013) The C-terminal domain of the bacteriophage T4 terminase docks on the prohead portal clip region during DNA packaging. *Virology*, **446**, 293–302.
60. Hegde,S., Padilla-Sanchez,V., Draper,B. and Rao,V.B. (2012) Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. *J. Virol.*, **86**, 4046–4057.
61. Yu,M.X., Slater,M.R. and Ackermann,H.W. (2006) Isolation and characterization of Thermus bacteriophages. *Arch. Virol.*, **151**, 663–679.
62. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. In: Carter,CW Jr and Sweet,RM (eds.), *Methods in Enzymology*. Academic Press, New York, USA, Vol. **276**, pp. 307–326.
63. Hendrickson,W.A. and Teeter,M.M. (1981) Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature*, **290**, 107–113.
64. Zwart,P.H., Afonine,P.V., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., McKee,E., Moriarty,N.W., Read,R.J., Sacchettini,J.C. *et al.* (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.*, **426**, 419–435.
65. Strong,M., Sawaya,M.R., Wang,S., Phillips,M., Cascio,D. and Eisenberg,D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8060–8065.
66. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
67. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
68. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
69. Freifelder,D. and Trumbo,B. (1969) Matching of single-strand breaks to form double-strand breaks in DNA. *Biopolymers*, **7**, 681–693.
70. Cowan,R., Collis,C.M. and Grigg,G.W. (1987) Breakage of double-stranded DNA due to single-stranded nicking. *J. Theor. Biol.*, **127**, 229–245.
71. Minakhin,L., Goel,M., Berdygulova,Z., Ramanculov,E., Florens,L., Glazko,G., Karamychev,V.N., Slesarev,A.I., Kozyavkin,S.A., Khromov,I. *et al.* (2008) 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.*, **378**, 468–480.
72. Nowotny,M. and Yang,W. (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.*, **25**, 1924–1933.
73. Yang,W., Lee,J.Y. and Nowotny,M. (2006) Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol. Cell*, **22**, 5–13.
74. Alam,T.I. and Rao,V.B. (2008) The ATPase domain of the large terminase protein, gp17, from bacteriophage T4 binds DNA: implications to the DNA packaging mechanism. *J. Mol. Biol.*, **376**, 1272–1281.
75. Gual,A., Camacho,A.G. and Alonso,J.C. (2000) Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1. *J. Biol. Chem.*, **275**, 35311–35319.
76. Campbell,V.W. and Jackson,D.A. (1980) The Effect of Divalent Cations on the Mode of Action of DNase I. *J. Biol. Chem.*, **255**, 3726–3735.
77. Povirk,L.F., Wubter,W., Kohnlein,W. and Hutchinson,F. (1977) DNA double-strand breaks and alkali-labile bonds produced by bleomycin. *Nucleic Acids Res.*, **4**, 3573–3580.
78. Colis,L.C., Woo,C.M., Hegan,D.C., Li,Z., Glazer,P.M. and Herzon,S.B. (2014) The cytotoxicity of (-)-lomaiviticin A arises from induction of double-strand breaks in DNA. *Nat. Chem.*, **6**, 504–510.
79. Jin,Y. and Cowan,J.A. (2005) DNA cleavage by copper-ATCUN complexes. Factors influencing cleavage mechanism and linearization of dsDNA. *J. Am. Chem. Soc.*, **127**, 8408–8415.
80. Zhang,Q., Xiang,Y., Liang,D., Peng,Y. and Guo,H. (2012) Efficient double-strand scission of plasmid DNA by quaternized-chitosan zinc complex. *Bioorg. Med. Chem. Lett.*, **22**, 1814–1817.
81. Branum,M.E., Tipton,A.K., Zhu,S. and Que,L. Jr (2001) Double-strand hydrolysis of plasmid DNA by dicerium complexes at 37 degrees C. *J. Am. Chem. Soc.*, **123**, 1898–1904.
82. Mirabelli,C.K., Huang,C.H. and Crooke,S.T. (1980) Comparison of DNA damage and single- and double-strand breakage activities on PM-2 DNA by talisomycin and bleomycin analogs. *Cancer Res.*, **40**, 4173–4177.
83. Shubsda,M.F., Goodisman,J. and Dabrowiak,J.C. (1997) Quantitation of ethidium-stained closed circular DNA in agarose gels. *J. Biochem. Biophys. Methods*, **34**, 73–79.
84. Li,S., Fan,H., An,X., Fan,H., Jiang,H., Chen,Y. and Tong,Y. (2014) Scrutinizing virus genome termini by high-throughput sequencing. *PLoS One*, **9**, e85806.
85. Nowotny,M., Gaidamakov,S.A., Ghirlando,R., Cerritelli,S.M., Crouch,R.J. and Yang,W. (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell*, **28**, 264–276.
86. Gorecka,K.M., Komorowska,W. and Nowotny,M. (2013) Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Res.*, **41**, 9945–9955.
87. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
88. Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.
89. Kim,Y.C., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.

90. delToro,D., Ortiz,D., Ordyan,M., Sippy,J., Oh,C.S., Keller,N., Feiss,M., Catalano,C.E. and Smith,D.E. (2016) Walker-A motif acts to coordinate ATP hydrolysis with motor output in viral DNA packaging. *J. Mol. Biol.*, **428**, 2709–2729.
91. Ray,K., Oram,M., Ma,J. and Black,L.W. (2009) Portal control of viral prohead expansion and DNA packaging. *Virology*, **391**, 44–50.
92. Casjens,S., Wyckoff,E., Hayden,M., Sampson,L., Eppler,K., Randall,S., Moreno,E.T. and Serwer,P. (1992) Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. *J. Mol. Biol.*, **224**, 1055–1074.

References

1. Comeau, A. M. *et al.* Exploring the prokaryotic virosphere. *Res. Microbiol.* **159**, 306–313 (2008).

2. Keen, E. C. A century of phage research: bacteriophages and the shaping of modern biology. *Bioessays* **37**, 6–9 (2015).

3. Hershey, A. D. & Chase, M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* **36**, 39–56 (1952).

4. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General Nature of the Genetic Code for Proteins. *Nature* **192**, 1227–1232 (1961).

5. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576–581 (1961).

6. Shi, G. & Suzuki, T. Molecular Basis of Encapsidation of Hepatitis C Virus Genome. *Front. Microbiol.* **9**, 396 (2018).

7. Green, T. J. *et al.* Common mechanism for RNA encapsidation by negative-strand RNA viruses. *J. Virol.* **88**, 3766–3775 (2014).

8. James, J. A. *et al.* Crystal structure of the SF3 helicase from adeno-associated virus type 2. *Structure* **11**, 1025–1035 (2003).

9. Chelikani, V., Ranjan, T. & Kondabagil, K. Revisiting the genome packaging in viruses with lessons from the ‘Giants’. *Virology* **466-467**, 15–26 (2014).

10. King, J. A., Dubielzig, R., Grimm, D. & Kleinschmidt, J. A. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *EMBO J.* **20**, 3282–3291 (2001).

11. Earnshaw, W. C., Honda, B. M., Laskey, R. A. & Thomas, J. O. Assembly of nucleosomes: the reaction involving X. laevis nucleoplasmin. *Cell* **21**, 373–383 (1980).

12. Bayer, M. E. & Bocharov, A. F. The capsid structure of bacteriophage lambda. *Virology* **54**, 465–475 (1973).

13. Bergh, O., Børsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).

14. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).

15. Lurz, R. *et al.* Structural organisation of the head-to-tail interface of a bacterial virus. *J. Mol. Biol.* **310**, 1027–1037 (2001).

16. Orlova, E. V. *et al.* Structure of a viral DNA gatekeeper at 10 A resolution by cryo-electron microscopy. *EMBO J.* **22**, 1255–1262 (2003).

17. Fokine, A. *et al.* The molecular architecture of the bacteriophage T4 neck. *J. Mol. Biol.* **425**, 1731–1744 (2013).

18. Bayfield, O. W. *et al.* Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 3556–3561 (2019).

19. Lokareddy, R. K. *et al.* Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nat. Commun.* **8**, 14310 (2017).

20. Sun, L. *et al.* Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat. Commun.* **6**, 7548 (2015).

21. Lebedev, A. A. *et al.* Structural framework for DNA translocation via the viral portal protein. *EMBO J.* **26**, 1984–1994 (2007).

22. Guasch, A. *et al.* Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle. *J. Mol. Biol.* **315**, 663–676 (2002).

23. Simpson, A. A. *et al.* Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408**, 745–750 (2000).

24. Casjens, S. R. The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol.* **9**, 647–657 (2011).

25. Grimes, S., Ma, S., Gao, J., Atz, R. & Jardine, P. J. Role of φ29 connector channel loops in late-stage DNA packaging. *J. Mol. Biol.* **410**, 50–59 (2011).

26. Fang, H., Jing, P., Haque, F. & Guo, P. Role of channel lysines and the ‘push through a one-way valve’ mechanism of the viral DNA packaging motor. *Biophys. J.* **102**, 127–135 (2012).

27. Guo, P., Noji, H., Yengo, C. M., Zhao, Z. & Grainge, I. Biological Nanomotors with a Revolution, Linear, or Rotation Motion Mechanism. *Microbiol. Mol. Biol. Rev.* **80**, 161–186 (2016).

28. Waters, J. T., Kim, H. D., Gumbart, J. C., Lu, X.-J. & Harvey, S. C. DNA Scrunching in the Packaging of Viral Genomes. *J. Phys. Chem. B* **120**, 6200–6207 (2016).

29. Leipe, D. D., Koonin, E. V. & Aravind, L. Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **333**, 781–815 (2003).

30. Rao, S. T. & Rossmann, M. G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**, 241–256 (1973).

31. Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem. Mol. Biol. Educ.* **43**, 206–209 (2015).

32. Geourjon, C. *et al.* A common mechanism for ATP hydrolysis in ABC transporter and helicase superfamilies. *Trends Biochem. Sci.* **26**, 539–544 (2001).

33. Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317**, 41–72 (2002).

34. Iyer, L. M., Makarova, K. S., Koonin, E. V. & Aravind, L. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res.* **32**, 5260–5279 (2004).

35. Burroughs, A. M., Iyer, L. M. & Aravind, L. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn.* **3**, 48–65 (2007).

36. Mao, H. *et al.* Structural and Molecular Basis for Coordination in a Viral DNA Packaging Motor. *Cell Rep.* **14**, 2017–2029 (2016).

37. Hilbert, B. J. *et al.* Structure and mechanism of the ATPase that powers viral genome packaging. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3792–9 (2015).

38. Sun, S., Kondabagil, K., Gentz, P. M., Rossmann, M. G. & Rao, V. B. The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. *Mol. Cell* **25**, 943–949 (2007).

39. Zhao, H., Christensen, T. E., Kamau, Y. N. & Tang, L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 8075–8080 (2013).

40. Xu, R.-G., Jenkins, H. T., Antson, A. A. & Greive, S. J. Structure of the large terminase from a hyperthermophilic virus reveals a unique mechanism for oligomerization and ATP hydrolysis. *Nucleic Acids Res.* **45**, 13029–13042 (2017).

41. Sun, S. *et al.* The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell* **135**, 1251–1262 (2008).

42. Leffers, G. & Rao, V. B. Biochemical characterization of an ATPase activity associated with the large packaging subunit gp17 from bacteriophage T4. *J. Biol. Chem.* **275**, 37127–37136 (2000).

43. Gual, A., Camacho, A. G. & Alonso, J. C. Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1. *J. Biol. Chem.* **275**, 35311–35319 (2000).

44. Schwartz, C., De Donatis, G. M., Fang, H. & Guo, P. The ATPase of the phi29 DNA packaging motor is a member of the hexameric AAA+ superfamily. *Virology* **443**, 20–27 (2013).

45. Zhao, Z. *et al.* An Arginine Finger Regulates the Sequential Action of Asymmetrical Hexameric ATPase in the Double-Stranded DNA Translocation Motor. *Mol. Cell. Biol.* **36**, 2514–2523 (2016).

46. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).

47. Thomsen, N. D. & Berger, J. M. Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases. *Mol. Microbiol.* **69**, 1071–1090 (2008).

48. Tafoya, S. *et al.* Molecular switch-like regulation enables global subunit coordination in a viral ring ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7961–7966 (2018).

49. Sun, S. *et al.* Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 817–822 (2012).

50. Roy, A., Bhardwaj, A., Datta, P., Lander, G. C. & Cingolani, G. Small terminase couples viral DNA binding to genome-packaging ATPase activity. *Structure* **20**, 1403–1413 (2012).

51. Mellado, R. P., Peñalva, M. A., Inciarte, M. R. & Salas, M. The protein covalently linked to the 5' termini of the DNA of Bacillus subtilis phage phi 29 is involved in the initiation of DNA replication. *Virology* **104**, 84–96 (1980).

52. Bjornsti, M. A., Reilly, B. E. & Anderson, D. L. Morphogenesis of bacteriophage phi 29 of Bacillus subtilis: oriented and quantized in vitro packaging of DNA protein gp3. *J. Virol.* **45**, 383–396 (1983).

53. Grimes, S. & Anderson, D. In vitro packaging of bacteriophage phi 29 DNA restriction fragments and the role of the terminal protein gp3. *J. Mol. Biol.* **209**, 91–100 (1989).

54. Cao, S. *et al.* Insights into the structure and assembly of the bacteriophage 29 double-stranded DNA packaging motor. *J. Virol.* **88**, 3986–3996 (2014).

55. Guo, P., Grimes, S. & Anderson, D. A defined system for in vitro packaging of DNA-gp3 of the Bacillus subtilis bacteriophage phi 29. *Proceedings of the National Academy of Sciences* **83**, 3505–3509 (1986).

56. Smith, D. E. *et al.* The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* **413**, 748–752 (2001).

57. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).

58. Moffitt, J. R. *et al.* Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446–450 (2009).

59. Kamtekar, S. *et al.* The phi29 DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition. *EMBO J.* **25**, 1335–1343 (2006).

60. Grimes, S. & Anderson, D. The bacteriophage phi29 packaging proteins supercoil the DNA ends. *J. Mol. Biol.* **266**, 901–914 (1997).

61. Guo, P., Peterson, C. & Anderson, D. Initiation events in in-vitro packaging of bacteriophage phi 29 DNA-gp3. *J. Mol. Biol.* **197**, 219–228 (1987).

62. Koti, J. S. *et al.* DNA packaging motor assembly intermediate of bacteriophage phi29. *J. Mol. Biol.* **381**, 1114–1132 (2008).

63. Turnquist, S., Simon, M., Egelman, E. & Anderson, D. Supercoiled DNA wraps around the bacteriophage phi 29 head-tail connector. *Proceedings of the National Academy of Sciences* **89**, 10479–10483 (1992).

64. Chistol, G. *et al.* High degree of coordination and division of labor among subunits in a homomeric ring ATPase. *Cell* **151**, 1017–1028 (2012).

65. Aathavan, K. *et al.* Substrate interactions and promiscuity in a viral DNA packaging motor. *Nature* **461**, 669–673 (2009).

66. Liu, S. *et al.* A viral packaging motor varies its DNA rotation and step size to preserve subunit coordination as the capsid fills. *Cell* **157**, 702–713 (2014).

67. Berndsen, Z. T., Keller, N., Grimes, S., Jardine, P. J. & Smith, D. E. Nonequilibrium dynamics and ultraslow relaxation of confined DNA during viral packaging. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8345–8350 (2014).

68. Xiang, Y. *et al.* Structural changes of bacteriophage phi29 upon DNA packaging and release. *EMBO J.* **25**, 5229–5239 (2006).

69. Morais, M. C. *et al.* Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor. *Structure* **16**, 1267–1274 (2008).

70. Ding, F. *et al.* Structure and assembly of the essential RNA ring component of a viral DNA packaging motor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7357–7362 (2011).

71. Harjes, E. *et al.* Structure of the RNA claw of the DNA packaging motor of bacteriophage Φ29. *Nucleic Acids Res.* **40**, 9953–9963 (2012).

72. Kypr, J. & Mrázek, J. Lambda phage protein Nu 1 contains the conserved DNA binding fold of repressors. *J. Mol. Biol.* **191**, 139–140 (1986).

73. de Beer, T. *et al.* Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Mol. Cell* **9**, 981–991 (2002).

74. Lin, H., Simon, M. N. & Black, L. W. Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging. *J. Biol. Chem.* **272**, 3495–3501 (1997).

75. Jackson, E. N., Jackson, D. A. & Deans, R. J. EcoRI analysis of bacteriophage P22 DNA packaging. *J. Mol. Biol.* **118**, 365–388 (1978).

76. Chai, S., Kruft, V. & Alonso, J. C. Analysis of the Bacillus subtilis bacteriophages SPP1 and SF6 gene 1 product: a protein involved in the initiation of headful packaging. *Virology* **202**, 930–939 (1994).

77. Chai, S., Lurz, R. & Alonso, J. C. The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J. Mol. Biol.* **252**, 386–398 (1995).

78. Baumann, R. G. & Black, L. W. Isolation and characterization of T4 bacteriophage gp17 terminase, a large subunit multimer with enhanced ATPase activity. *J. Biol. Chem.* **278**, 4618–4627 (2003).

79. Feiss, M. & Rao, V. B. The Bacteriophage DNA Packaging Machine. *Viral Molecular Machines* 489–509 (2012). doi:10.1007/978-1-4614-0980-9_22

80. Casjens, S. R. & Gilcrease, E. B. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol. Biol.* **502**, 91–111 (2009).

81. Wu, H., Sampson, L., Parr, R. & Casjens, S. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. *Mol. Microbiol.* **45**, 1631–1646 (2002).

82. Leavitt, J. C., Gilcrease, E. B., Wilson, K. & Casjens, S. R. Function and horizontal transfer of the small terminase subunit of the tailed bacteriophage Sf6 DNA packaging nanomotor. *Virology* **440**, 117–133 (2013).

83. Shinder, G. & Gold, M. The Nul subunit of bacteriophage lambda terminase binds to specific sites in cos DNA. *J. Virol.* **62**, 387–392 (1988).

84. Greive, S. J. *et al.* DNA recognition for virus assembly through multiple sequence-independent interactions with a helix-turn-helix motif. *Nucleic Acids Res.* **44**, 776–789 (2016).

85. Zhao, H., Kamau, Y. N., Christensen, T. E. & Tang, L. Structural and functional studies of the phage Sf6 terminase small subunit reveal a DNA-spooling device facilitated by structural plasticity. *J. Mol. Biol.* **423**, 413–426 (2012).

86. Zhao, H. *et al.* Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1971–1976 (2010).

87. Buttner, C. R. *et al.* Structural basis for DNA recognition and loading into a viral packaging motor. *Proceedings of the National Academy of Sciences* **109**, 811–816 (2012).

88. Gao, S. & Rao, V. B. Specificity of interactions among the DNA-packaging machine components of T4-related bacteriophages. *J. Biol. Chem.* **286**, 3944–3956 (2011).

89. Frackman, S., Siegele, D. A. & Feiss, M. The terminase of bacteriophage lambda. Functional domains for cosB binding and multimer assembly. *J. Mol. Biol.* **183**, 225–238 (1985).

90. Yang, Q., Berton, N., Manning, M. C. & Catalano, C. E. Domain structure of gpNu1, a phage lambda DNA packaging protein. *Biochemistry* **38**, 14238–14247 (1999).

91. Alam, T. I. *et al.* The headful packaging nuclease of bacteriophage T4. *Mol. Microbiol.* **69**, 1180–1190 (2008).

92. Al-Zahrani, A. S. *et al.* The small terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. *J. Biol. Chem.* **284**, 24490–24500 (2009).

93. Kondabagil, K. R., Zhang, Z. & Rao, V. B. The DNA translocating ATPase of bacteriophage T4 packaging motor. *J. Mol. Biol.* **363**, 786–799 (2006).

94. Black, L. W. & Peng, G. Mechanistic coupling of bacteriophage T4 DNA packaging to components of the replication-dependent late transcription machinery. *J. Biol. Chem.* **281**, 25635–25643 (2006).

95. Fuller, D. N., Raymer, D. M., Kottadiel, V. I., Rao, V. B. & Smith, D. E. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proceedings of the National Academy of Sciences* **104**, 16868–16873 (2007).

96. Fuller, D. N. *et al.* Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* **373**, 1113–1122 (2007).

97. Fuller, D. N. *et al.* Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi 29. *Proceedings of the National Academy of Sciences* **104**, 11245–11250 (2007).

98. Jardine, P. J. Slow and steady wins the race: physical limits on the rate of viral DNA packaging. *Curr. Opin. Virol.* **36**, 32–37 (2019).

99. Finkelstein, I. J., Visnapuu, M.-L. & Greene, E. C. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* **468**, 983–987 (2010).

100. Bianco, P. R. *et al.* Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409**, 374–378 (2001).

101. Saleh, O. A., Pérals, C., Barre, F.-X. & Allemand, J.-F. Fast, DNA-sequence independent translocation by FtsK in a single-molecule experiment. *EMBO J.* **23**, 2430–2439 (2004).

102. Pease, P. J. *et al.* Sequence-directed DNA translocation by purified FtsK. *Science* **307**, 586–590 (2005).

103. Lee, J. Y., Finkelstein, I. J., Crozat, E., Sherratt, D. J. & Greene, E. C. Single-molecule imaging of DNA curtains reveals mechanisms of KOPS sequence targeting by the DNA translocase FtsK. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6531–6536 (2012).

104. Ptacin, J. L. *et al.* Sequence-directed DNA export guides chromosome translocation during sporulation in Bacillus subtilis. *Nat. Struct. Mol. Biol.* **15**, 485–493 (2008).

105. Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* **13**, 676–683 (2006).

106. Kelch, B. A., Makino, D. L., O'Donnell, M. & Kuriyan, J. How a DNA polymerase clamp loader opens a sliding clamp. *Science* **334**, 1675–1680 (2011).

107. Hendrix, R. W. Symmetry mismatch and DNA packaging in large bacteriophages. *Proceedings of the National Academy of Sciences* **75**, 4779–4783 (1978).

108. Baumann, R. G., Mullaney, J. & Black, L. W. Portal fusion protein constraints on function in DNA packaging of bacteriophage T4. *Mol. Microbiol.* **61**, 16–32 (2006).

109. Hugel, T. *et al.* Experimental test of connector rotation during DNA packaging into bacteriophage phi29 capsids. *PLoS Biol.* **5**, e59 (2007).

110. Lin, H., Rao, V. B. & Black, L. W. Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. *J. Mol. Biol.* **289**, 249–260 (1999).

111. Dixit, A., Ray, K., Lakowicz, J. R. & Black, L. W. Dynamics of the T4 bacteriophage DNA packasome motor: endonuclease VII resolvase release of arrested Y-DNA substrates. *J. Biol. Chem.* **286**, 18878–18889 (2011).

112. Dixit, A. B., Ray, K. & Black, L. W. Compression of the DNA substrate by a viral packaging motor is supported by removal of intercalating dye during translocation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20419–20424 (2012).

113. Dixit, A. B., Ray, K., Thomas, J. A. & Black, L. W. The C-terminal domain of the bacteriophage T4 terminase docks on the prohead portal clip region during DNA packaging. *Virology* **446**, 293–302 (2013).

114. Morita, M., Tasaka, M. & Fujisawa, H. DNA packaging ATPase of bacteriophage T3. *Virology* **193**, 748–752 (1993).

115. Müller, I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta crystallographica. Section D, Structural biology* **73**, 79–92 (2017).

116. Hilbert, B. J., Hayes, J. A., Stone, N. P., Xu, R.-G. & Kelch, B. A. The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. *Nucleic Acids Res.* **45**, 3591–3605 (2017).

117. Alam, T. I. & Rao, V. B. The ATPase domain of the large terminase protein, gp17, from bacteriophage T4 binds DNA: implications to the DNA packaging mechanism. *J. Mol. Biol.* **376**, 1272–1281 (2008).

118. Ray, K., Sabanayagam, C. R., Lakowicz, J. R. & Black, L. W. DNA crunching by a viral packaging motor: Compression of a procapsid-portal stalled Y-DNA substrate. *Virology* **398**, 224–232 (2010).

119. Black, L. W. Old, new, and widely true: The bacteriophage T4 DNA packaging mechanism. *Virology* (2015).

120. Harvey, S. C. 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.* **189**, 1–8 (2015).

121. Oram, M., Sabanayagam, C. & Black, L. W. Modulation of the packaging reaction of bacteriophage t4 terminase by DNA structure. *J. Mol. Biol.* **381**, 61–72 (2008).

122. Tang, J. *et al.* DNA poised for release in bacteriophage phi29. *Structure* **16**, 935–943 (2008).

123. Streisinger, G., Emrich, J. & Stahl, M. M. Chromosome structure in phage t4, iii. Terminal redundancy and length determination. *Proceedings of the National Academy of Sciences* **57**, 292–295 (1967).

124. Casjens, S. & Hayden, M. Analysis in vivo of the bacteriophage P22 headful nuclease. *J. Mol. Biol.* **199**, 467–474 (1988).

125. Wu, R. & Taylor, E. Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage lambda DNA. *J. Mol. Biol.* **57**, 491–511 (1971).

126. Kosturko, L. D., Daub, E. & Murialdo, H. The interaction of E. coli integration host factor and lambda cos DNA: multiple complex formation and protein-induced bending. *Nucleic Acids Res.* **17**, 317–334 (1989).

127. Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295–1306 (1996).

128. Higgins, R. R. & Becker, A. Chromosome end formation in phage lambda, catalyzed by terminase, is controlled by two DNA elements of cos, cosN and R3, and by ATP. *EMBO J.* **13**, 6152–6161 (1994).

129. Cue, D. & Feiss, M. Termination of packaging of the bacteriophage lambda chromosome: cosQ is required for nicking the bottom strand of cosN. *J. Mol. Biol.* **280**, 11–29 (1998).

130. Cue, D. & Feiss, M. Genetic evidence that recognition of cosQ, the signal for termination of phage lambda DNA packaging, depends on the extent of head filling. *Genetics* **147**, 7–17 (1997).

131. Tye, B. K., Huberman, J. A. & Botstein, D. Non-random circular permutation of phage P22 DNA. *J. Mol. Biol.* **85**, 501–528 (1974).

132. Moore, S. D. & Prevelige, P. E. Bacteriophage p22 portal vertex formation in vivo. *J. Mol. Biol.* **315**, 975–994 (2002).

133. Weigele, P. R., Sampson, L., Winn-Stapley, D. & Casjens, S. R. Molecular genetics of bacteriophage P22 scaffolding protein's functional domains. *J. Mol. Biol.* **348**, 831–844 (2005).

134. Griffiths, A. J. F. *et al. An Introduction to Genetic Analysis*. (Macmillan, 2005).

135. Lander, G. C. *et al.* The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **312**, 1791–1795 (2006).

136. Xu, R.-G. *et al.* Viral genome packaging terminase cleaves DNA using the canonical RuvC-like two-metal catalysis mechanism. *Nucleic Acids Res.* **45**, 3580–3590 (2017).

137. Smits, C. *et al.* Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep.* **10**, 592–598 (2009).

138. Roy, A. & Cingolani, G. Structure of p22 headful packaging nuclease. *J. Biol. Chem.* **287**, 28196–28205 (2012).

139. Zhao, H. *et al.* Two distinct modes of metal ion binding in the nuclease active site of a viral DNA-packaging terminase: insight into the two-metal-ion catalytic mechanism. *Nucleic Acids Res.* **43**, 11003–11016 (2015).

140. Górecka, K. M., Komorowska, W. & Nowotny, M. Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Res.* **41**, 9945–9955 (2013).

141. delToro, D. *et al.* Walker-A Motif Acts to Coordinate ATP Hydrolysis with Motor Output in Viral DNA Packaging. *J. Mol. Biol.* **428**, 2709–2729 (2016).

142. Kanamaru, S., Kondabagil, K., Rossmann, M. G. & Rao, V. B. The functional domains of bacteriophage t4 terminase. *J. Biol. Chem.* **279**, 40795–40801 (2004).

143. Trottier, M. & Guo, P. Approaches to determine stoichiometry of viral assembly components. *J. Virol.* **71**, 487–494 (1997).

144. Lee, T. J., Schwartz, C. & Guo, P. Construction of bacteriophage phi29 DNA packaging motor and its applications in nanotechnology and therapy. *Ann. Biomed. Eng.* **37**, 2064–2081 (2009).

145. Tavares, P., Zinn-Justin, S. & Orlova, E. V. Genome gating in tailed bacteriophage capsids. *Adv. Exp. Med. Biol.* **726**, 585–600 (2012).

146. Fokine, A. & Rossmann, M. G. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**, e28281 (2014).

147. Agirrezabala, X. *et al.* Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J.* **24**, 3820–3829 (2005).

148. Jiang, W. *et al.* Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* **439**, 612–616 (2006).

149. Liu, X. *et al.* Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. *Nat. Struct. Mol. Biol.* **17**, 830–836 (2010).

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

151. Black, L. W. & Silverman, D. J. Model for DNA packaging into bacteriophage T4 heads. *J. Virol.* **28**, 643–655 (1978).

152. Jin, Y. *et al.* Bacteriophage P22 ejects all of its internal proteins before its genome. *Virology* **485**, 128–134 (2015).

153. Wu, W. *et al.* Localization of the Houdinisome (Ejection Proteins) inside the Bacteriophage P22 Virion by Bubblegram Imaging. *MBio* **7**, e01152–16 (2016).

154. Wang, C., Tu, J., Liu, J. & Molineux, I. J. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature microbiology* 1 (2019).

155. Rao, V. B. & Feiss, M. The bacteriophage DNA packaging motor. *Annu. Rev. Genet.* **42**, 647–681 (2008).

156. Alex Evilevitch, Martin Castelnovo, Charles M Knobler, A. & Gelbart, W. M. *Measuring the Force Ejecting DNA from Phage†*. **108**, (American Chemical Society, 2004).

157. Hegde, S., Padilla-Sanchez, V., Draper, B. & Rao, V. B. Portal-large terminase interactions of the bacteriophage T4 DNA packaging machine implicate a molecular lever mechanism for coupling ATPase to DNA translocation. *J. Virol.* **86**, 4046–4057 (2012).

158. Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J. & Steitz, T. A. The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell* **151**, 267–277 (2012).

159. Thomsen, N. D. & Berger, J. M. Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. *Cell* **139**, 523–534 (2009).

160. Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489–484 (2008).

161. Enemark, E. J. & Joshua-Tor, L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**, 270–275 (2006).

162. White, J. H. & Richardson, C. C. Gene 19 of bacteriophage T7. Overexpression, purification, and characterization of its product. *J. Biol. Chem.* **263**, 2469–2476 (1988).

163. Ibarra, B., Valpuesta, J. M. & Carrascosa, J. L. Purification and functional characterization of p16, the ATPase of the bacteriophage Phi29 packaging machinery. *Nucleic Acids Res.* **29**, 4264–4273 (2001).

164. Nadal, M. *et al.* Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16078–16083 (2010).

165. Yu, M. X., Slater, M. R. & Ackermann, H.-W. Isolation and characterization of Thermus bacteriophages. *Arch. Virol.* **151**, 663–679 (2006).

166. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).

167. Kornberg, A. & Pricer, W. E., Jr. Enzymatic phosphorylation of adenosine and 2,6-diaminopurine riboside. *J. Biol. Chem.* **193**, 481–495 (1951).

168. Lee, T. J., Zhang, H., Liang, D. & Guo, P. Strand and nucleotide-dependent ATPase activity of gp16 of bacterial virus phi29 DNA packaging motor. *Virology* **380**, 69–74 (2008).

169. Hendrickson, W. A. & Teeter, M. M. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290**, 107–113 (1981).

170. Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).

171. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8060–8065 (2006).

172. Otwinowski, Z. & Minor, W. [20] Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology* **276**, 307–326 (Academic Press, 1997).

173. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

174. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

175. Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612 (2009).

176. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1--towards automated and web-supported small-angle scattering data analysis. *Applied crystallography* **40**, s223–s228 (2007).

177. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

178. Pierce, B., Tong, W. & Weng, Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. *Bioinformatics* **21**, 1472–1478 (2005).

179. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).

180. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A. & Sauer, R. T. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**, 744–756 (2009).

181. DeLaBarre, B. & Brunger, A. T. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nat. Struct. Biol.* **10**, 856–863 (2003).

182. Wang, J. Nucleotide-dependent domain motions within rings of the RecA/AAA+ superfamily. *J. Struct. Biol.* **148**, 259–267 (2004).

183. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).

184. Cox, J. M., Abbott, S. N., Chitteni-Pattu, S., Inman, R. B. & Cox, M. M. Complementation of one RecA protein point mutation by another. Evidence for trans catalysis of ATP hydrolysis. *J. Biol. Chem.* **281**, 12968–12975 (2006).

185. Pierce, B. G. *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773 (2014).

186. Zhao, M. *et al.* Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**, 61–67 (2015).

187. Bailey, S., Eliason, W. K. & Steitz, T. A. Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* **318**, 459–463 (2007).

188. Vafabakhsh, R. *et al.* Single-molecule packaging initiation in real time by a viral DNA packaging machine from bacteriophage T4. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15096–15101 (2014).

189. 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* **185**, 788–794 (1991).

190. Yang, Q., Catalano, C. E. & Maluf, N. K. Kinetic analysis of the genome packaging reaction in bacteriophage lambda. *Biochemistry* **48**, 10705–10715 (2009).

191. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).

192. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041 (2001).

193. Rao, V. B. & Mitchell, M. S. The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4. *J. Mol. Biol.* **314**, 401–411 (2001).

194. Stinson, B. M. *et al.* Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine. *Cell* **153**, 628–639 (2013).

195. Feiss, M. & Rao, V. B. The Bacteriophage DNA Packaging Machine. in *Viral Molecular Machines* **726**, 489–509 (Springer US, 2011).

196. Schmieger, H. Phage P22-mutants with increased or decreased transduction abilities. *Mol. Gen. Genet.* **119**, 75–88 (1972).

197. Casjens, S. *et al.* Molecular genetic analysis of bacteriophage P22 gene 3 product, a protein involved in the initiation of headful DNA packaging. *J. Mol. Biol.* **227**, 1086–1099 (1992).

198. Casjens, S., Huang, W. M., Hayden, M. & Parr, R. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. *J. Mol. Biol.* **194**, 411–422 (1987).

199. Chai, S., Lurz, R. & Alonso, J. C. The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J. Mol. Biol.* **252**, 386–398 (1995).

200. Stone, N. P. *et al.* A Hyperthermophilic Phage Decoration Protein Suggests Common Evolutionary Origin with Herpesvirus Triplex Proteins and an Anti-CRISPR Protein. *Structure* **26**, 936–947.e3 (2018).

201. Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, (2016).

202. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

203. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *Elife* **7**, (2018).

204. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

205. Zivanov, J. *et al.* RELION-3: new tools for automated high-resolution cryo-EM structure determination. doi:10.1101/421123

206. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

207. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–94 (2015).

208. Minakhin, L. *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.* **378**, 468–480 (2008).

209. Gao, S., Zhang, L. & Rao, V. B. Exclusion of small terminase mediated DNA threading models for genome packaging in bacteriophage T4. *Nucleic Acids Res.* **44**, 4425–4439 (2016).

210. Xu, R.-G., Jenkins, H. T., Antson, A. A. & Greive, S. J. Structure of the large terminase from a hyperthermophilic virus reveals a unique mechanism for

oligomerization and ATP hydrolysis. *Nucleic Acids Res.* **45**, 13029–13042 (2017).

211. Chemla, Y. R. & Smith, D. E. Single-Molecule Studies of Viral DNA Packaging. in *Viral Molecular Machines* **726**, 549–584 (Springer US, 2011).

212. Martin, A., Baker, T. A. & Sauer, R. T. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**, 1115–1120 (2005).

213. Stinson, B. M., Baytshtok, V., Schmitz, K. R., Baker, T. A. & Sauer, R. T. Subunit asymmetry and roles of conformational switching in the hexameric AAA+ ring of ClpX. *Nat. Struct. Mol. Biol.* **22**, 411–416 (2015).

214. Martin, A., Baker, T. A. & Sauer, R. T. Distinct Static and Dynamic Interactions Control ATPase-Peptidase Communication in a AAA+ Protease. **27**, 41–52 (2007).

215. Shin, Y. *et al.* Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19340–19345 (2009).

216. Glynn, S. E., Nager, A. R., Baker, T. A. & Sauer, R. T. Dynamic and static components power unfolding in topologically closed rings of a AAA+ proteolytic machine. *Nat. Struct. Mol. Biol.* **19**, 616–622 (2012).

217. Cordova, J. C. *et al.* Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647–658 (2014).

218. Alberts, B. *et al. DNA-Binding Motifs in Gene Regulatory Proteins*. (Garland Science, 2002).

219. McNulty, R. *et al.* Architecture of the Complex Formed by Large and Small Terminase Subunits from Bacteriophage P22. *J. Mol. Biol.* **427**, 3285–3299 (2015).

220. Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

221. Buxmann, H., Hamprecht, K., Meyer-Wittkopf, M. & Friese, K. Primary Human Cytomegalovirus (HCMV) Infection in Pregnancy. *Dtsch. Arztebl. Int.* **114**, 45–52 (2017).

222. Jha, H. C., Banerjee, S. & Robertson, E. S. The Role of Gammaherpesviruses in Cancer Pathogenesis. *Pathogens* **5**, (2016).

223. Steiner, I. & Benninger, F. Manifestations of Herpes Virus Infections in the Nervous System. *Neurol. Clin.* **36**, 725–738 (2018).

224. Buerger, I. *et al.* A novel nonnucleoside inhibitor specifically targets cytomegalovirus DNA maturation via the UL89 and UL56 gene products. *J. Virol.* **75**, 9077–9086 (2001).

225. Melendez, D. P. & Razonable, R. R. Letermovir and inhibitors of the terminase complex: a promising new class of investigational antiviral drugs against human cytomegalovirus. *Infect. Drug Resist.* **8**, 269–277 (2015).

226. Krosky, P. M. *et al.* Resistance of human cytomegalovirus to benzimidazole ribonucleosides maps to two open reading frames: UL89 and UL56. *J. Virol.* **72**, 4721–4728 (1998).

227. van Zeijl, M. *et al.* Novel class of thiourea compounds that inhibit herpes simplex virus type 1 DNA cleavage and encapsidation: resistance maps to the UL6 gene. *J. Virol.* **74**, 9054–9061 (2000).

228. Guo, S., Tschammer, N., Mohammed, S. & Guo, P. Specific delivery of therapeutic RNAs to cancer cells via the dimerization mechanism of phi29 motor pRNA. *Hum. Gene Ther.* **16**, 1097–1109 (2005).

229. Shu, D., Shu, Y., Haque, F., Abdelmawla, S. & Guo, P. Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nat. Nanotechnol.* **6**, 658–667 (2011).

230. Guo, S., Piao, X., Li, H. & Guo, P. Methods for construction and characterization of simple or special multifunctional RNA nanoparticles based on the 3WJ of phi29 DNA packaging motor. *Methods* **143**, 121–133 (2018).

231. Jasinski, D., Haque, F., Binzel, D. W. & Guo, P. Advancement of the Emerging Field of RNA Nanotechnology. *ACS Nano* **11**, 1142–1164 (2017).

232. Cressiot, B., Greive, S. J., Mojtabavi, M., Antson, A. A. & Wanunu, M. Thermostable virus portal proteins as reprogrammable adapters for solid-state nanopore sensors. *Nat. Commun.* **9**, 4652 (2018).

233. Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).

234. Gai, D., Zhao, R., Li, D., Finkielstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47–60 (2004).

235. Subramaniam, S., Earl, L. A., Falconieri, V., Milne, J. L. & Egelman, E. H. Resolution advances in cryo-EM enable application to drug discovery. *Curr. Opin. Struct. Biol.* **41**, 194–202 (2016).

236. Grigorieff, N. Direct detection pays off for electron cryo-microscopy. *Elife* **2**, e00573 (2013).

237. Rao, V. B. & Black, L. W. DNA packaging of bacteriophage T4 proheads in vitro. Evidence that prohead expansion is not coupled to DNA packaging. *J. Mol. Biol.* **185**, 565–578 (1985).

238. Steven, A. C., Bauer, A. C., Bisher, M. E., Robey, F. A. & Black, L. W. The maturation-dependent conformational change of phage T4 capsid involves the translocation of specific epitopes between the inner and the outer capsid surfaces. *J. Struct. Biol.* **106**, 221–236 (1991).

239. Sathaliyawala, T. *et al.* Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. *Molecular Microbiology* **77**, 444–455 (2010).

240. Stone, N. P., Demo, G., Agnello, E. & Kelch, B. A. Principles for enhancing virus capsid capacity and stability from a thermophilic virus capsid structure. *bioRxiv* 473264 (2019). doi:10.1101/473264