The Dissertation Committee for Huzhang Mao Certifies that this is the approved version of the following dissertation:

STRUCTURE OF THE GP16 ATPase FROM THE φ29 dsDNA PACKAGING MOTOR

Committee:

Marc C. Morais, Ph.D. (Supervisor) _____

Kyung H. Choi, Ph.D. _____

Jose M. Barral, Ph.D. _____

Dean, Graduate School _____

Structure of the gp16 ATPase from the φ29 dsDNA Packaging Motor

by

Huzhang Mao

Dissertation

Presented to the Faculty of the Graduate School of The University of Texas Medical Branch in Partial Fulfillment of the Requirements for the Degree of

Master in Biochemistry and Molecular Biology

The University of Texas Medical Branch

December, 2011

Structure of the gp16 ATPase from the φ29 dsDNA Packaging Motor

Publication No.

Huzhang Mao, M.S.

The University of Texas Medical Branch, 2011

Supervisor: Marc C. Morais

Bacillus subtilis bacteriophage φ29 packages double-stranded DNA (dsDNA) into a preformed viral shell, or procapsid, and serves as a model system for studying genome packaging in eukaryotic dsDNA viruses such as poxviruses, herpesviruses and adenoviruses. Encapsidation of bacteriophage φ29 DNA is driven by a phage-encoded molecular motor. This motor is powered by an oligomeric ATPase, gp16, or gene product 16, that converts energy obtained from ATP hydrolysis into translocation of dsDNA. In this study, we solved the gp16 ATPase structure via X-ray crystallography. The resultant monomeric structure indicated that gp16 ATPase adopts a modified Rossmann fold (Rossmann *et al.*, 1974), in which six conserved β-strands form a central β-sheet, and adjacent β-strands are linked by intervening α-helices, and that the protein is a member of the ancient P-loop additional strand catalytic E (ASCE) NTPase superfamily. The active site responsible for ATP hydrolysis is located on one side of the central β-sheet. Superposition of our structure on related ring-forming members of the ASCE superfamily

indicated that residue Arg148 protrudes from the other side of the β-sheet and is well-positioned to insert its side chain into the active site of a neighboring ATPase to trigger sequential ATP hydrolysis events around an oligomeric ATPase ring.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ASCE……..……Additional strand catalytic E

AUC…….Analytical Ultracentrifugation

Cat E……....Catalytic Glutamate

dsDNA……..double-strand DNA

gp10……gene product 10

OB……oligonucleotide/oligosaccharide binding

pRNA…….……prohead RNA

RMSD……..…..Root Mean Square Deviation

Sen……....Phosphate Sensor

S.V……..……Sedimentation Velocity

WA…….……Walker A Motif

WB……..……Walker B Motif

Chapter 1: Introduction

1.1 Bacteriophage φ29

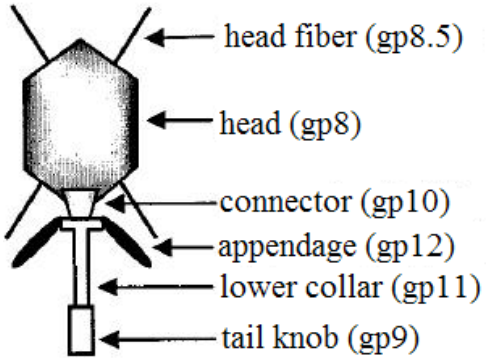


Figure 1: Schematic structure of phage φ29 (Modified from Meijer *et al.*, 2001).

Bacteriophage φ29 (Riley et al., 1965) is a member of the Podoviridae family and infects *Bacillus subtilis*. Phage φ29 (Figure 1) contains a 19.3 kbp dsDNA genome that encodes about 20 proteins encapsidated in an icosahedral head with quasi T=3, Q=5 symmetry that measures about 540 Å long and 450 Å wide. The head consists of 30 hexameric and 11 pentameric capsomers, with a dodecameric head-tail connector head protein replacing a pentamer at one of the twelve icosahedral vertices. Attached to the head is a pseudo 6-fold symmetric 440 Å long neck/tail complex (Tao *et al.*, 1998) that recognizes and attaches to the host cell. Phage φ29 is one of the smallest dsDNA tailed phages, and its relatively small genome (about 40% the size of phage λ and 10% the size of phage T4) affords the opportunity to uncover the function of every gene in the virus (Grimes *et al.*, 1997). The morphogenesis pathway of phage φ29 begins with the assembly of the prolate prohead that consists of 12 copies of head-tail connector protein gp10 (Guo *et al.*, 1987), about 130 copies of the scaffolding protein (gp7), 235 copies of the major capsid protein (gp8) and 55 copies the trimeric head fiber protein (gp8.5). A pentameric φ29-encoded RNA (pRNA) binds to the connector-vertex of the prohead to complete prohead assembly. The viral genome, covalently attached at both ends to gene product3 (gp3), is then packaged into the prohead in a process driven by the phage-encoded ATPase, gp16. The scaffolding protein gp7 exits during packaging (Morais *et al.*, 2003), although the mechanism and route of scaffold release is poorly understood. Upon completion of DNA packaging, the lower collar (gp11), tail (gp9) and appendages (gp12) attach to the DNA-filled prohead to form the mature phage (Grimes *et al.*, 2002).

1.2 Phage φ29 DNA Packaging Motor

The φ29 packaging motor is composed of three components: a dodecameric connector protein (gp10), an oligomeric pRNA and a multimeric ATPase (Figure 2). Hence, the connector, pRNA and ATPase form three concentric rings, and the DNA-gp3 is believed to be threaded through a continuous channel along their shared central axis and into the phage capsid. This overall picture of the motor not only provides a general context for understanding the structure and function of individual macromolecular motor components, but also provides a framework for understanding how different motor components interact with each other to coordinate the mechanochemical cycle of the motor. The structures and functions of each of these three components are summarized below.

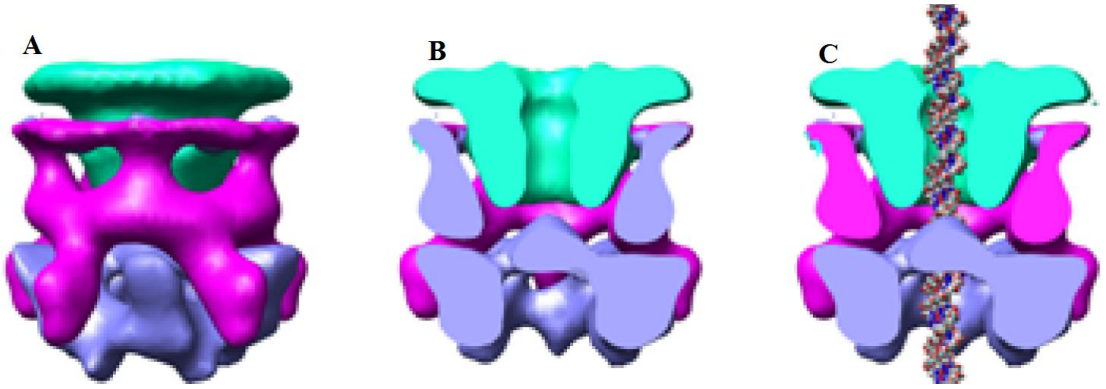


Figure 2: Reconstruction of phage φ29 DNA packaging motor: connector (green), pRNA multimer (magenta) and pentameric gp16 (blue). A. Side view. B. Cross-section of side view. C. Cross-sectional side view with DNA modeled in the central channel of the motor. (Morais *et al.*, 2008).

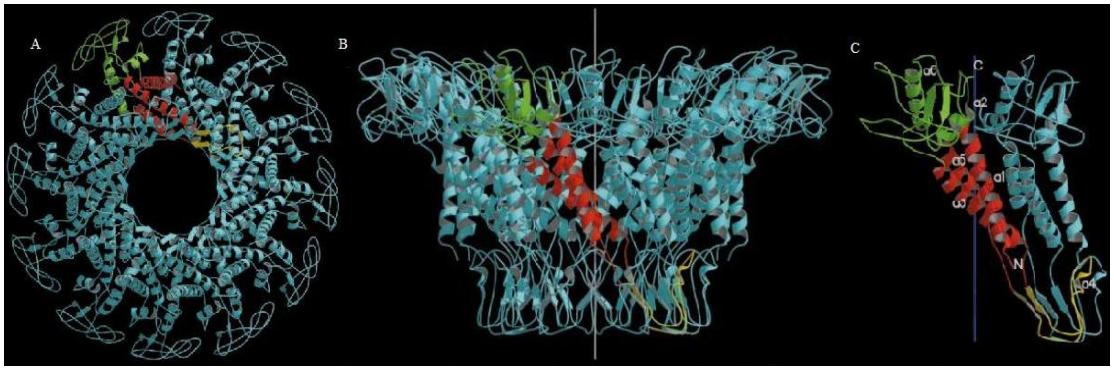


Figure 3: Phage φ29 dodecameric connector structure ribbon diagram. A. Top view. B. Side view. C. Side-view of two monomers (Simposn, *et al.*, 2000). In all three panels a single

monomer is colored such that wide-end is shown in green, the central region in red, and the narrow region in yellow; all other monomers are colored blue.

The connector (Figure 3) is 75 Å long and can be divided into three regions: the narrow end, the central part, and the wide end which have external radii of 33, 47 and 69 Å, respectively. The internal diameter of the connector is 36 Å at the narrow end and 60 Å at the wide end (Simpson, *et al.*, 2000).

The pRNA consists of three-helices organized around a U-rich 3-way junction (Figure 4), and has two domains: domain I consists of 117 bases at the 5' end, and domain II consists of 44 bases at the 3' end, with a single strand of 13 bases connecting the two domains (Figure 4). The A helix in domain I is proposed to bind to gp16 ATPase, and other parts within this domain are responsible for prohead binding, connector binding, as well as pRNA oligomerization (Morais *et al.*, 2008; Ding *et al.*, 2011).

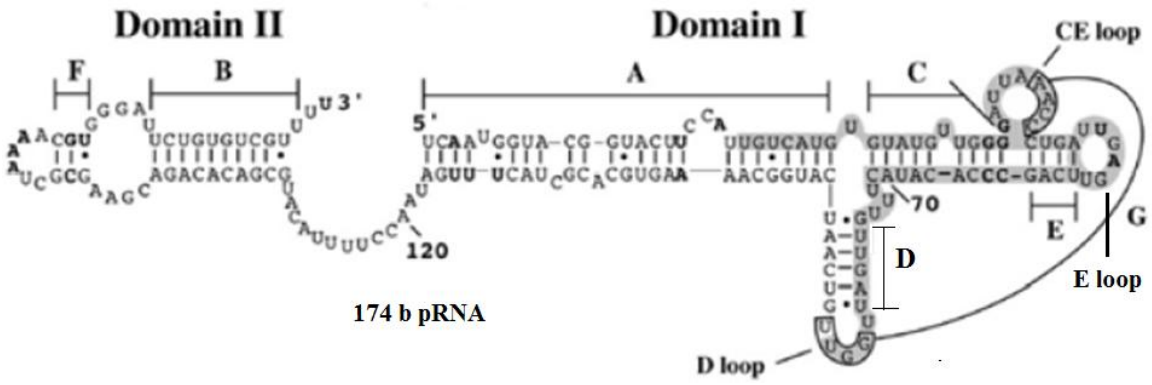


Figure 4: Predicted secondary structure of phage φ29 pRNA (A-G: helix A-G). (Grimes *et al.*, 2002).

pRNA is predicted to form multimeric ring around the connector via intermolecular interactions between the CE & E loop within domain I, and this had been confirmed by X-ray crystallography and cryoEM reconstruction (Morais *et al.*, 2001; Morais *et al.*, 2005; Morais *et al.*, 2008; Ding *et al.*, 2011; Grimes and Morais, unpublished data) (Figure 5).

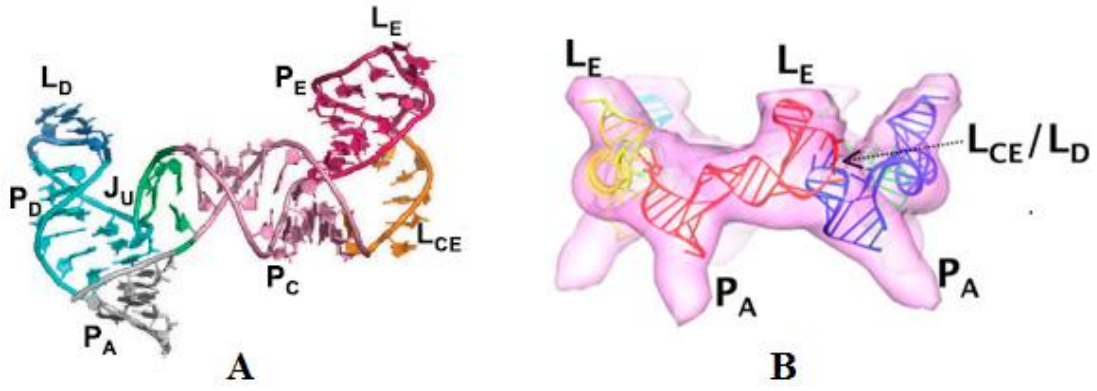


Figure 5. Structure of pRNA prohead-binding domain. (A) pRNA protomer. (B) Fitting of pRNA ring structure into its cryoEM envelope; intermolecular base-pairing between the CE & D loop is indicated by a black arrow (Ding *et al.*, 2011).

1.3 The packaging gp16 ATPase

In order to compact DNA into proheads, considerable energy is required to pay the enthalpic and entropic costs resulting from bending DNA, charge-charge repulsions, and the reduction of conformational space. The gp16 couples the chemical energy released during ATP hydrolysis to the mechanical translocation of the DNA. The 39-kDa 332-residue gp16 consists of two domains (Koti *et al.*, 2008): the first ~200 N-terminal amino acids correspond to the ASCE (Additional Strand Catalytic E (glutamate)) P-loop NTPase domain, and the last ~130 residues comprise the C-terminal domain (Figure 6). Sequence analysis shows that C-terminal domain likely adopts a five-stranded oligonucleotide/oligosaccharide binding (OB) fold, indicating a possible role in pRNA binding (Koti *et al.*, 2008) or DNA translocation (Sun, *et al.*, 2007, Sun *et al.*, 2008).

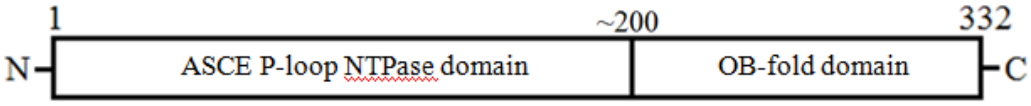


Figure 6. Schematic diagram of φ29 gp16 showing its two domain structure: Additional Strand Catalytic E (glutamate)) (ASCE) P-loop NTPase domain (amino acid 1 ~ 200), and oligonucleotide/oligosaccharide binding (OB) fold domain (amino acid 201 ~ 332) are indicated.

The N-terminal domain, or ATPase domain, contains several motifs characteristic of the ASCE P-loop superfamily, including an adenine binding motif, Walker A/B motifs, a catalytic glutamate, a phosphate sensor motif, and an arginine finger (Figure 7A). It's believed that the Walker A/B motifs participate in ATP hydrolysis by binding and positioning ATP (Walker A and B) and coordinating a critical catalytic Mg^{2+} ion that activates the catalytic water nucleophile (Walker B). The phosphate sensor monitors the hydrolysis state of ATP, and thus likely plays a role in coordinating ATP hydrolysis with the sequential conformational changes that constitute the mechano-chemical cycle of the motor**.** The arginine finger is typically inserted from one monomer into the active site of a neighboring monomer in an oligomeric ring, where the positively charged guanidinium group stabilizes both the pentavalent transition state of ATP hydrolysis and the phosphate leaving group. Since the arginine finger belongs to a neighboring subunit, it may also play a role in coordinating the sequential hydrolysis of ATP around the oligomeric ring of ATPases (Figure 7B).

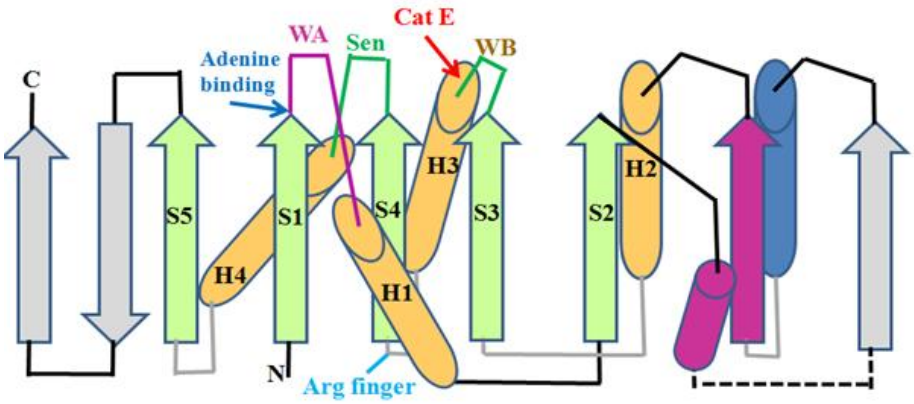


Figure 7. Secondary structural prediction of gp16 ATPase. Five parallel beta strands (S1, S2, S3, S4, and S5, colored light green) form the central β-sheet, and each strand is separated by an intervening α-helix (H1, H2, H3, and H4, colored light orange). The motifs involved in ATP hydrolysis (Adenine binding motif, Walker A/B motifs, catalytic glutamate) are predicted to reside on one side of the sheet, while the residue responsible for coordination of ATP hydrolysis within the multimeric ATPase ring is predicted to lie on the other side (Burroughs *et al.*, 2007; Morais 2011 chapter).

Interestingly, gp16 by itself has weak ATPase activity, which is stimulated by RNA or DNA. Furthermore, maximum ATPase activity is reached only when gp16 binds to the prohead-pRNA complex as a ring, and DNA is present (Reid *et al.*, 1994b; Reid *et al.*, 1994c; Zhang *et al.*, 1994; Zhang *et al.*, 1995), suggesting that efficient ATP hydrolysis requires coordination of neighboring subunits around the ring. Thus, high resolution structural information about the entire motor complex is needed to fully understand the role of g16 in DNA packaging.

1.4 DNA Packaging Model

As one of the most powerful biological motors known, the φ29 packaging motor can generate forces up to 70 pN (Rickgauer *et al.*, 2008). Several models have been proposed to explain the fundamental mechanism of DNA translocation. These models generally fall into two basic types: rotary motors where rotation of the connector is coupled to DNA translocation, and linear models where different domains within a protein alternately grab and release DNA (Figure 8). Rotary models are largely based on the ideas of Hendrix (1978), who proposed that the symmetry mismatch between the 5-fold prohead shell and 12-fold connector would facilitate rotation of the connector. Rotation of the connector around the helical DNA would thus "screw" the DNA into the capsid, much as a nut moves on a bolt. In contrast to rotary-based mechanisms, the φ29 motor was also suggested to function more like a linear motor, in which phage DNA-packaging could be better explained by an "inchworm-type mechanism", in which opening and closing of the ATP binding cleft causes movement of a DNA-binding domain in the ATPase that is coupled to DNA translocation (Draper *et al.*, 2007).

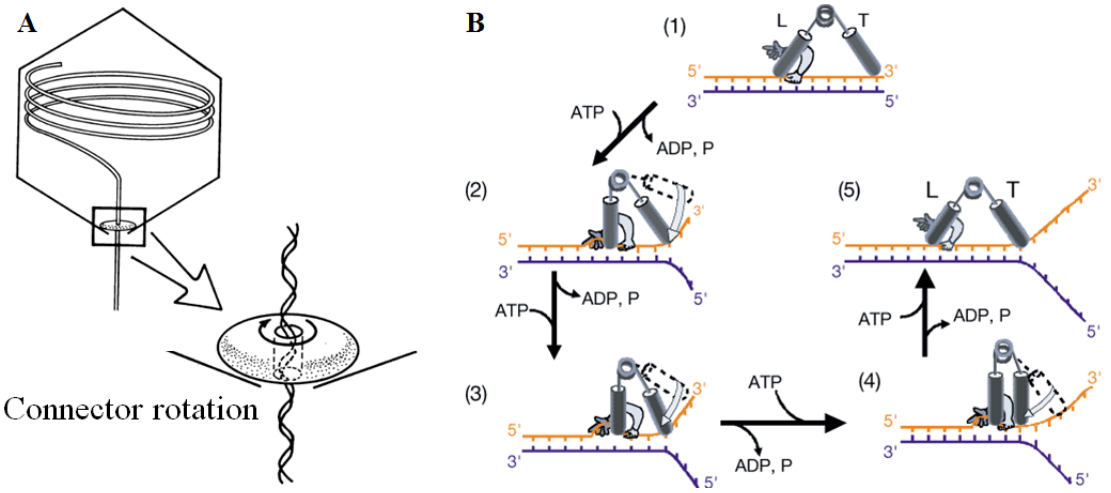


Figure 8. Cartoons showing two proposed DNA packaging mechanisms utilized by φ29: A. Rotary model in which rotation of the connector screws DNA into the prohead (Hendrix 1978); B. Linear model in which opening and closing of DNA binding domains induced by ATP hydrolysis is associated with DNA translocation. Individual strands of DNA duplex is colored in orange and blue. The protein contains two domains: a leading domain (L, which binds to the DNA duplex) and an ATP binding domain (T, which binds ATP and introduces conformational changes in the leading domain upon ATP hydrolysis) (Bianco *et al.*, 2000).

More recently, a "push and roll" model derived from a homology model of gp16 made using the φ12 ssRNA packaging ATPase as a template was suggested based on observation from single particle experiments including: 1) that 10bp DNA is packaged per four ATP hydrolysis events within the pentameric ring (Moffitt *et al.*, 2009); and 2) binding affinity for ATPase to DNA is high in ATP-bound state, and low in nucleotide-free state (Chemla *et al.*, 2005). In this model, specific interactions between the motor and the DNA are not required, and the force powering genome encapsidation is generated by non-specific steric contacts between motor and the major groove of DNA (Yu *et al.*, 2010).

However, without an atomic structure of gp16, and hence an atomic model of the whole motor, the mechanism of DNA translocation utilized by phage φ29 cannot be fully understood. Here, we describe the X-ray crystal structure of N-terminal ASCE-NTPase domain of the gp16 ATPase complexed with AMP-PNP to ~1.9Å resolution. Structural comparison with other ATPase motors suggests that gp16

ATPase is more related to oligomeric ATPases, and superposition of gp16 onto various ring-forming ATPases suggest mechansims by which dsDNA is encapsidated in the procapsid.

Chapter 2 Methods

2.1 Protein Expression and Purification

Expression and purification of gp16: Clones of full-length gp16 and its N- and C-terminal domains (identified from limited trypsin treatment; see below) were cloned into pET30a at NdeI/XhoI sites and transformed to *E. coli* for expression. The transformed cells were grown in LB medium in presence of kanamycin to an optimal OD_{600} of 0.5, and induced with 0.6 mM isopropyl-2-thiogalactopyranoside (IPTG) overnight at 18°C. The cells were spun down and resuspended in a buffer containing 20mM Tris-HCl, 300mM NaCl, and pH7.5, followed by sonication. The gp16-containing supernatant was incubated with Talon (TM) resin and eluted with a 5mM to 300mM imidazole gradient. The fractions containing the protein were collected, concentrated, and loaded onto a gel filtration column. Fractions containing pure protein were collected and concentrated to desired concentrations for crystallization experiments.

Limited trypsin treatment: Purified gp16 was concentrated to 1mg/ml, followed by incubation with 2% trypsin (mass ratio of trypsin to protein) for 5~10 minutes at room temperature. Stable protein fragments were separated by SDS-PAGE gel and stable bands were either extracted for Mass Spectrometry analysis, or the whole gel was transferred to PVDF membrane for protein N-terminal sequencing to identify the trypsin cleavage site; in this way, both N- and C-terminal boundaries were determined for the two major bands resulting from limited trypsin proteolysis.

Clones of individual domains of gp16: Genes that encode individual N- and C-terminal domains of gp16 indentified from limited trypsin treatment were cloned from synthesized full-length gene *gp16* via polymerase chain reaction (PCR). Gene fragments were then inserted to plasmid pET30a at NdeI and XhoI sites. The

combinational plasmid containing the inserted gene was transformed to *E. coli* JM101 competent cells for plasmid extraction and gene sequencing at UTMB Molecular Genomics Core. After confirmation of the desired insertion, the combinational plasmid was transformed to *E. coli* BL21(DE3) competent cells for protein expression and purification as described above.

2.2 X-ray Crystallography

Protein crystallization: Extensive screening was performed to find conditions suitable for gp16 crystallization (full-length and individual domain constructs). In addition, a novel in-drop trypsinization protocol was applied to crystallization trials, in which 2% trypsin was added to protein solution and allowed to incubate for 5 -10 minutes at room temperature prior to subsequent crystallographic condition screening. Crystallization trials were greatly expedited by the integrated, robotic crystallization system from Rigaku available at UTMB Structural Core Facility.

X-ray crystallography: X-ray data was collected at the UTMB Structural Core Facility using an X-ray area detector systems that contain an ultra-fine-focus high brilliancy X-ray generator and focusing multilayer optics: these initially included a MacScience DIP2030H-VLM dual 30cm diameter imaging plate detector and Bruker-AXS SMART 2k CCD. X-ray data was also collected using new diffraction equipment purchased by the X-ray core facility in the summer of 2011 with unsurpassed high-brilliance FR-E++DW Superbright X-ray generator and RAXIS-IV++ crystallography system with both Cu and Cr optics from Rigaku, USA. In order to solve the phase problem via anomalous dispersion phasing methods, SAD data on a selenomethionine derivitized protein were collected at the GCPCC (Gulf Coast Protein Crystallography Consortium) protein crystallography beam line at the CAMD (Center for Advanced Microstructures and Devices) synchrotron at Baton Rouge, Louisiana.

Chapter 3 Structure Determination of φ29 gp16 ATPase

3.1 Protein Crystallization

Attempts to crystallize full length gp16 were challenging due to the full-length protein's poor solubility and tendency to self-aggregate upon purification (Figure S1). Analytical ultracentrifugation experiments showed that purified full-length gp16 trends to form multimers in solution rapidly (Figure 9, Figure S2).

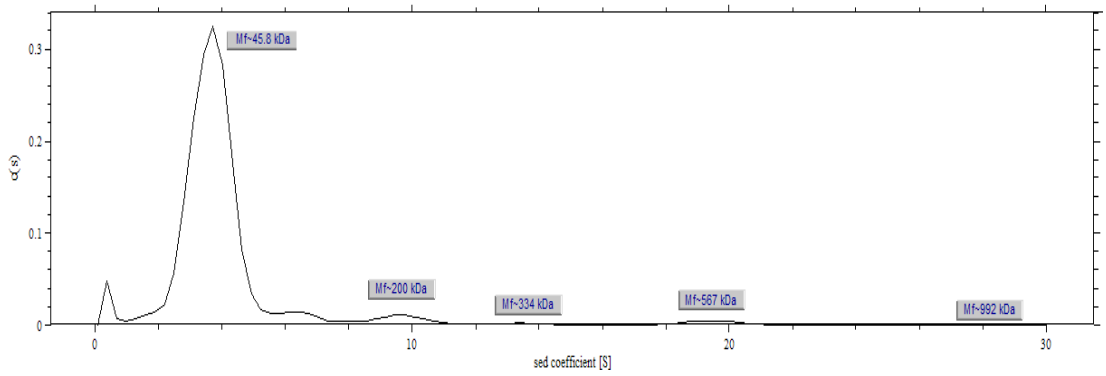


Figure 9. Analytical ultracentrifugation analysis of full-length gp16 (0.54mg/ml) in solution. Though most protein (>85%) stays monomer in solution, higher oligomerization states present. Peaks labeled with 45.8 kDa, 200 kDa, and 334 kDa correspond to gp16 monomer, pentamer, and octamer, respectively.

Despite the low solubility and self-aggregation of full-length gp16, we were able to obtain diffracting crystals from full-length protein using the novel in-drop trypsinization protocol (see Methods) in conditions consisting of 2M ammonium sulfate, 100mM citric acid, pH5.5 (Figure 10). Mass spectrometry analysis of the resultant crystals revealed that first 122 amino acids of full-length gp16 were crystallized, which embodies part of the predicted catalytic ATPase core.

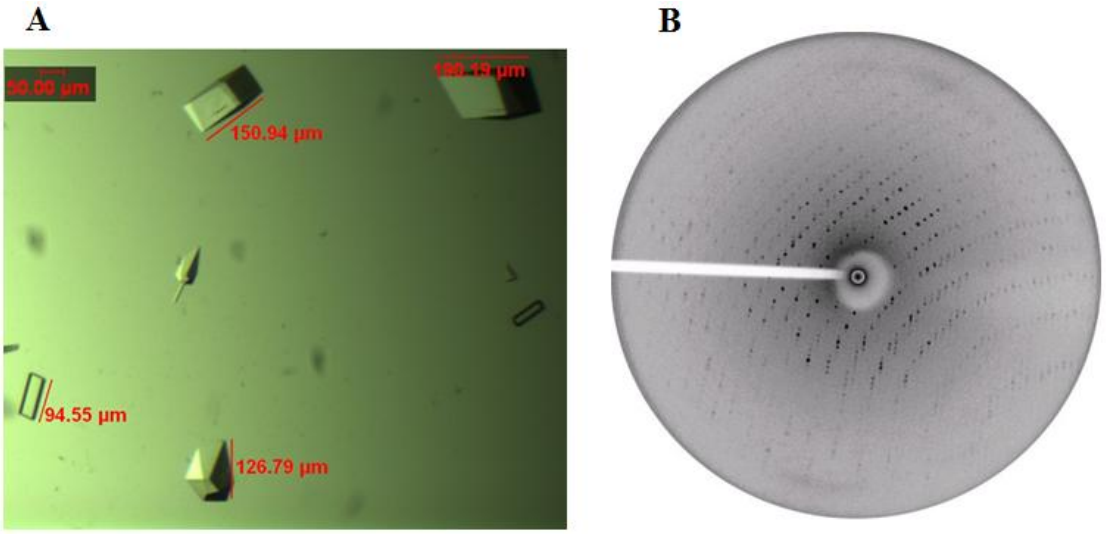


Figure 10. Crystal of gp16 N-terminal domain resulted from in-drop trysinization. (A) Crystals of gp16 N-terminal domain from in-drop trypsinization. Crystals adopted multiple geometric forms, such as cuboid, diamond, and other polyhedron. (B) Diffraction pattern of crystals shown in (A), which fell into space group P1, a=35Å, b=37Å, c=40Å, α=97 °, β=101 °, γ=92 °.

The fact that crystals from in-drop trypsinization contained only first ~122 amino acids of N-terminal domain of gp16 was quite unexpected, as positions around 122 correspond to the ATPase catalytic core, and do not correspond to the cleavage site between N- and C-terminal domains identified as R207 in previous experiments. In order to detect domain boundaries within full-length gp16, gp16 was subjected to limited trypsinization (Methods). N-terminal sequencing and mass spectrometry analysis revealed that limited trypsin treatment cleaved full-length gp16 at a position between Arg207 and Leu208 **(**Figure S1**)**. Individual fragments of both N- and C-terminal domains were expressed and purified, followed by crystallization condition screening (Figure S1). We thus suspect that an additional cleavage reaction occurred during in-drop proteolysis. Although we were only able to map the 1st 122 amino acids using mass spectromety, we suspect the crystals contain the first ~ 200 amino acids of N-terminal domain. Cleavage at residue 122 would mean we cleaved only part of the fold, which seems unlikely since this ASCE fold is the basic structural unit. Furthermore, the dimensions of the asymmetric unit are similar to that of a

crystal of untreated ATPase domain complexed with AMP-PNP (see below). Hence, the peptide containing residues beyond 122 were not visible in mass spectromety, or trypsin cleaved a loop between secondary structural elements, yet the overall fold remained intact. We are in the processing of solving the structure of data obtained from these crystals, and the resulting electron density is expected to confirm the overall fold remains intact and possibly identify an additional cleavage site in one of the connecting loops in the structure. Attempts to crystallize the N- and C- terminal domains, even the N-terminal domain identified via limited proteolysis, were unsuccessful. Interestingly, addition of trypsin to the N-terminal construct resulted in crystals identical to those obtained for the full-length protein using in-drop proteolysis. These results suggest that an additional cleavage was necessary to generate a readily crystallizable fragment. As discussed above, we are working to solve the structures obtained from these crystals, which should reveal the necessary additional cleavage.

3.2 Structure of gp16 ATPase and AMP-PNP Complex

Gp16 N-terminal domain is an ATPase that participates in multiple steps of ATP hydrolysis, including ATP binding, ATP hydrolyzing, ADP stabilization, and phosphate leaving. The ability to observe sequential conformational changes that occur among those steps would greatly facilitate understanding of ATP hydrolysis mechanism utilized by φ29 ATPase and its relatives. In order to trap the ATP binding state, we tried co-crystallization of the N-terminal ATPase domain of gp16 ATPase (identified via limited proteolysis) with the ATP analog AMP-PNP. Fortunately, we obtained well-diffracting crystals in conditions containing either 2.8M sodium acetate trihydrate, pH7.0 or 3.5M sodium formate, pH7.0 **(**Figure S3**)**. It is worth noting that these constructs were never treated with trypsin, and thus contained the entirety of the N-terminal ATPase domain. To solve the phase problem and determine the structure, a selenomethionine derivative was prepared and used for single wavelength anomalous dispersion (SAD) data collection at the protein crystallography beam line at CAMD synchrotron at Baton Rouge, Louisiana. Selenomethionine sites were located in

Patterson maps from which initial phases could be determined and then improved via solvent flattening. Data indexing, integration, space group determination, and scaling were carried out using HKL3000 (Minor et al., 2006); SAD structure solution and initial model building was carried out using the program PHENIX (Adams et al., 2010). Further model building was improved using the program ARP/wARP (Langer *et al.*, 2008) and manual intervention.

3.2.1 Structural Overview

We solved the structure using single-wavelength anomalous dispersion on a selenomethionine substituted protein a few weeks ago. So far, we cannot see the His-tag, the first four amino acids, or the last 8 amino acids, and density for the AMP-PNP molecule is not entirely clear; overall, additional work is needed to finish refining the structure.

The gp16 ATPase structure (Figure 11A) was determined to ~1.9Å resolution. The structure is approximately an oval ball measuring about 45 × 33 × 33Å. As expected, the protein adopts a modified Rossmann fold (Rossmann *et al.*, 1974) that is characteristic of ASCE superfamily members. In this case, there is a nine-stranded β-sheet that is sandwiched between α-helices on either side, and five parallel β-strands connected by intervening α-helices compose the central β-sheet. Functional motifs responsible for ATP binding and hydrolysis were mapped onto the structure and are all located on one side of the central β-sheet, except for the potential communicating arginine finger, either Arg146 or Arg148, which sits on the other side of the central β-sheet (Figure 11B).

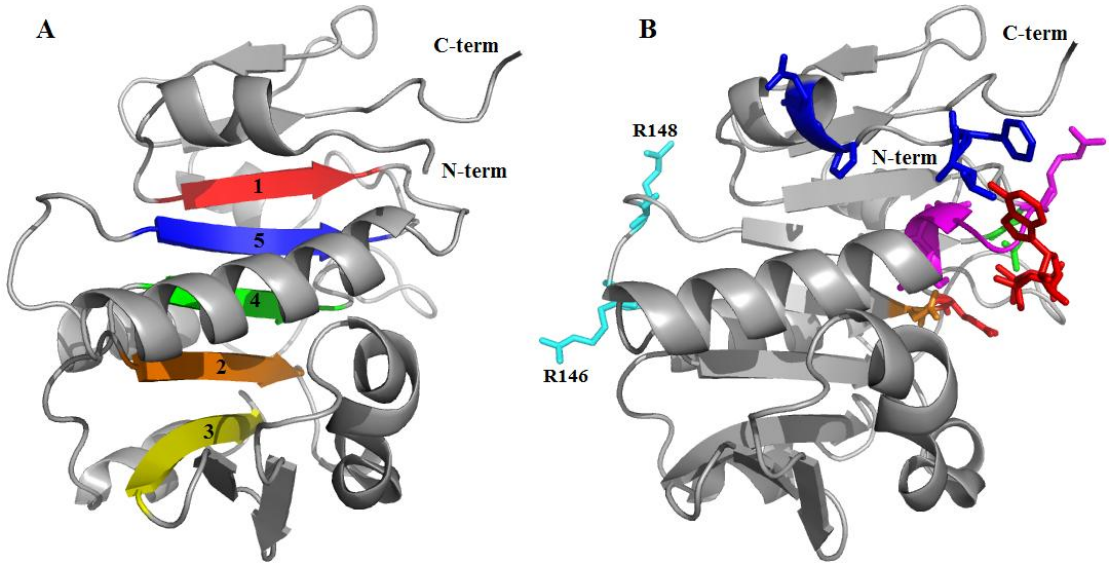


Figure 11. Structure of gp16 ATPase. (A) A ribbon representation of gp16 ATPase. The parallel five β-strands in the central β-sheet (colored in red, blue, green, orange, and yellow, respectively) follow a 15423 topology. (B) Residues associated with nucleotide binding and hydrolysis. Residues associated with adenine binding, Walker A/B motifs, Cat E, and the phosphate sensor are shown in sticks and colored in blue, purple, yellow, orange, and green, respectively. An ATP molecule (colored in red) is docked into the active site based on the observed AMP-PNP density within the binding cleft in the T4 gp16-ATP complex (Sun *et al.*, 2007).

Table 1. Data Collection and Refinement Statistics

	ATP Se-Met
Data Collection	
X-ray source	Protein Crystallization Beamline at CAMD*
Wavelength (Å)	0.97941
Resolution (Å)	50-1.93
Space group	$P2_12_12_1$
Unit cell (Å)	a= 33.1
	b = 36.8
	c = 139.1
Unique reflections	13,022

R_{sym} (%)	6.9 (23.1)
I/σ	31.5 (5.4)
Completeness (%)	96.6 (72.6)
Average redundancy	9.1

* CAMD: Center for Advanced Microstructures & Devices, Baton Rouge, Louisiana.

Refinement	
Resolution (Å)	50-1.93
R_{work}/R_{free} (%)	20.3/25.9
Average B factor ($Å^2$)	35.6
RMSD bonds (Å)	0.008
RMSD angles (°)	1.2

3.2.1 Structural and Sequence Comparison

Despite low sequence identity, gp16 ATPase superimposed well (Table 2) onto various ATPases with distinct cellular functions, such as bacteriophage T4 gp17 ATPase (Sun *et al.*, 2007), α- and β- subunits of bovine heart mitochondroia F1-ATPase (Abrahams *et al.*, 1994), and Pcr DNA helicase (Velankar *et al.*, 1999). Structural alignment of those ATPases results in an approximately identical position around the tip of the central β-sheet where nucleotide or its analog is bound (Table 2).

Table 2: Structural Superpositions of gp16 ATPase with Other Motor ATPases

Protein	Function	PDB Entry	Number of Equivalent Cα Atoms	RMSD(Å)
T4 gp17	dsDNA phage packaging motor	2O0H	160	4.2
φ12 P4	ssRNA phage packaging motor	1W44	96	6.9
T7 gp4	hexameric helicase	1E0J	144	5.9
F1-ATPase α	protein pump	1H8H	144	5.2
F1-ATPase β	protein pump	1H8H	144	6.2
RecA	Recombination	1XMS	128	5.6
RecG	monomeric helicase	1GM5	136	5.6

PcrA+DNA	monomeric helicase	2PJR	80	4.3
FtsH	hexameric metalloprotease	1IY0	120	6.4
FtsK	hexameric translocase	2IUT	80	5.2

In addition to structure based comparison, previous sequence alignment also detected similar amino acids arranged around the spatially conserved motifs among gp16 ATPase and other motor ATPases (Figure 12). G/A-XXXXGK(T/S) (where X can be any amino acid) is present in virtually all Walker A motifs; a conserved Asp residue at the tip of Walker B motif (ZZZZD, Z represents hydrophobic amino acid) is directly followed by a carboxylate group, usually Glu; a tripeptide sequence T/S-G/A-T/S 20-30 amino acids downstream of the Walker B motif consists the phosphate sensor; as well as ZQ in the adenine binding motif, is located ~15 amino acids upstream of Walker A motif. Extensive mutational analysis has demonstrated that mutations within conserved motifs could generate mutants defect in either ATP hydrolysis or DNA translocation, further illustrating their essential role in motor function.

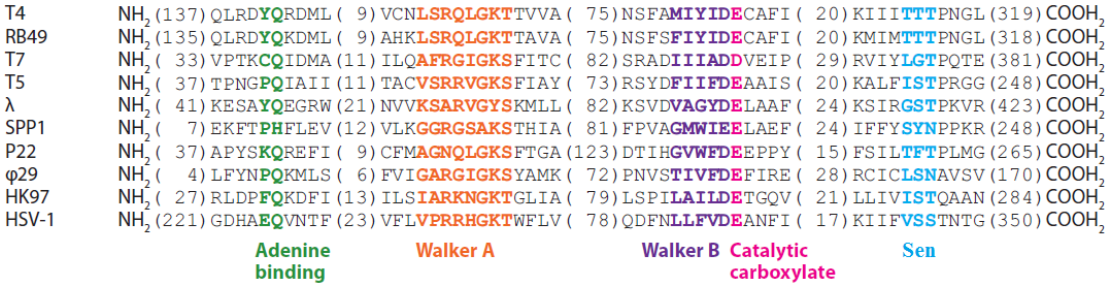


Figure 12. Sequence alignment of ATPases showing functional motifs: Adenine binding, Walker A, Walker B, Catalytic carboxylate, and Sen, colored in green, orange, purple, pink, and cyan, respectively. These ATPase motifs are conserved among phage φ29, T4 and other phages. Numbers in parentheses represent the number of amino acids (Rao, *et al.*, 2008).

3.2.3 ATPase Activity Center

As mentioned above, a putative arginine finger of gp16 ATPase is inserted into the active site of a neighboring ATPase site. Two arginine finger candidates, Arg146 and Arg148, are observed in the crystal structure. Both arginine residues are at the

base of β-4 strand, and thus we tried to distinguish arginine finger between Arg146 and Arg148 by superimposing our ATPase structure onto another arginine finger containing ATPase, the FtsK translocase hexamer. Though we believe that gp16 forms pentamer on the motor, fitting of gp16 oligomers onto other ring structures would still be informative. For example, pRNA forms different oligomers in crystal and on the motor using same residues (Morais, *et al.*, 2008; Ding *et al.*, 2011). In this case, two Arg fingers can be distinguished by superposition of a gp16 hexamer onto FtsK translocase hexamer, indicating that Arg148 is within close proximity of active ATPase site from a neighboring subunit, while Arg146 protrudes outwards the ring (Figure 13). Thus, Arg148 is more likely to function as an Arg finger and trigger sequential ATP hydrolysis within a ring structure (Morais chapter, 2011).

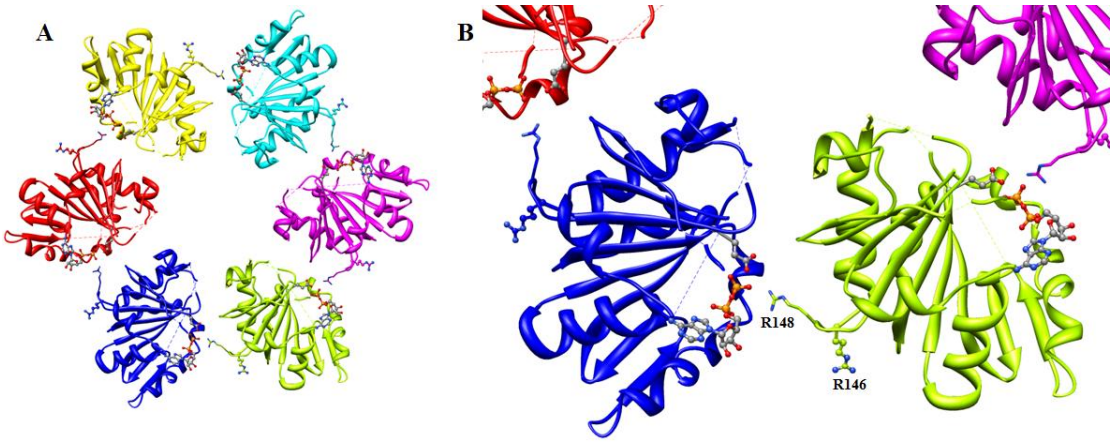
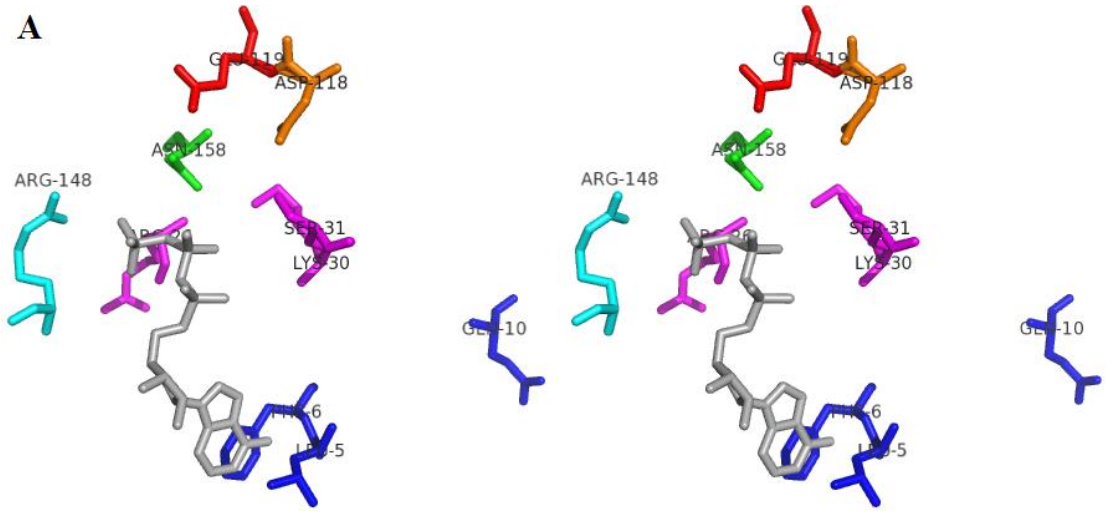


Figure 13. Arg finger. (A) Superposition of six gp16 ATPases onto the FtsK translocase hexamer results in a gp16 ATPase hexameric ring. In this model, Arg148 is inserted into active cleft of a neighboring subunit in clockwise pattern. Six ATPase subunits are colored in blue, red, yellow, cyan, purple, and green, respectively. Two Arg finger candidates, Arg146 and Arg148 are shown as sticks. An ATP is also shown in sticks, and colored according to atoms: gray, blue, orange, and red for C, N, P and O, respectively. Although this ATP is from the FtsK structure, it superimposes well with the putative AMP-PNP molecule from our structure. (B) Enlarged view of Arg finger between blue and green subunits from (A). R148 from green ATPase inserts its side chain into the hydrolysis site of the blue ATPase, while R146 protrudes outwards the ring structure and loses contact with the blue subunit.

3.2.3 ATP Binding Site

Based on observed structural features in gp16 ATPase and sequence- and structure-based alignment information, potential protein-substrate interaction events within gp16 ATPase active center can be suggested for φ29 gp16. As shown in Figure 14, an incoming ATP molecule is stabilized by 1) hydrophobic stacking interaction between adenine base and Phe-Leu dipepetide, and 2) hydrogen bonding between the heterocyclic ring and Gln10. Besides, ATP is further stabilized by electrostatic interaction between α-phosphate and Arg26 within Walker A motif. Residues Lys30 and Ser31 are part of Walker A phosphate binding P-loop and correctly orient ATP phosphate chain into the active site cleft. Ser31, together with Asp118 from Walker B motif, coordinates an Mg^{2+} ion to precisely position γ-phosphate into the catalytic sites. The Glu119 activates a water molecule for nucleophilic attack on the pre-positioned γ-phosphate. The arginine finger, Arg148, from a neighboring monomer stabilizes the pentavalent transition state and facilitates the leaving of Pi after hydrolysis. Asn158 monitors ATP hydrolysis events and possibly relays this information to other parts within the molecule via conformational changes (Story *et al.*, 1992).



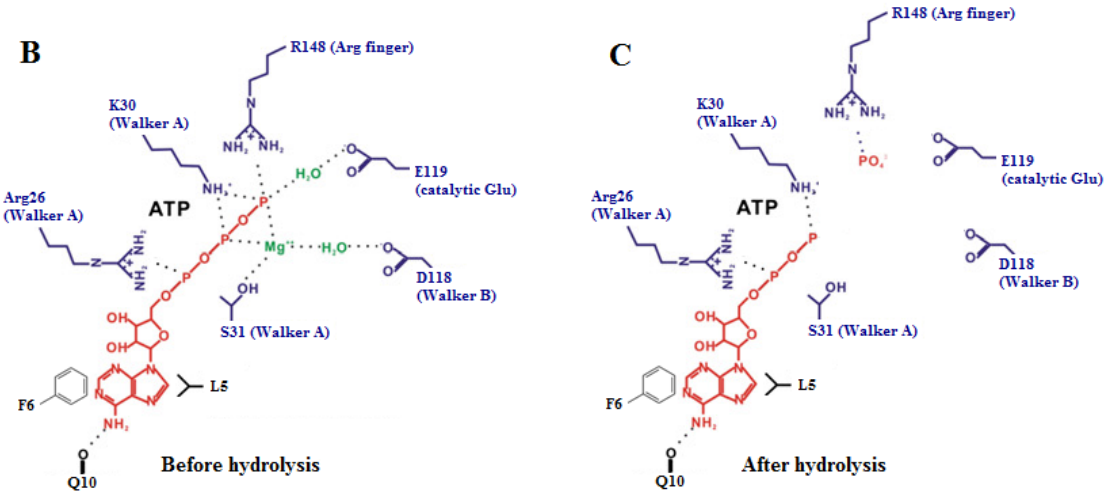


Figure 14. gp16 ATPase active center. (A) Stereodiagram representation of residues in the ATP-binding cleft of gp16 ATPase. ATP is colored in gray and residues from adenine binding motif, Walker A/B motifs, Cat E, Sen and Arg finger are colored as in Figure 10B. The Arg148 is inserted into the active site from an adjacent monomer. (B) Schematic view of molecular interaction between the ATP molecule and contacting amino acids. Residues interacting with adenine bases are colored in black, others are colored in blue. One Mg^{2+} ion and two water molecules are colored in green. As the native structure is crystallized with non-hydrolysable AMPPNP; at this stage of our refinement, we cannot be sure that the proposed Mg^{2+} is not a water molecule, and further refinement and analysis is necessary to confirm this metal site. Before hydrolysis, β- and γ-phosphates are stabilized by residues from Walker A/B motifs, after hydrolysis triggered by Glu119, the leaving of γ-phosphate is facilitated by Arg148 from a neighboring monomer. Asn158 captures conformational changes in the process.

Chapter 4 Discussion

4.1 Oligomerization of gp16

CryoEM reconstructions have shown that gp16 ATPases form a pentameric ring on the prohead by interacting with the distal ends of pRNA A-helix (Koti *et al.,* 2008; Morais *et al.,* 2008). Since ATPases from φ29 relatives could form higher order oligomers, or even decamers in solution, it's possible that gp16 ATPases first form pentamer in solution and then assemble the resultant pentamer onto prohead. However, data from analytical ultracentrifugation and gel filtration show that both full-length gp16 protein and individual N-/C-terminal domains sare monomers in solution **(**Figure 9**,** Figure S2**)**. Also, reported formation of dimers and trimers of recombinantly expressed gp16 ATPase in solution (Lee & Guo 2006; Koti *et al.,* 2008) suggests that it is unlikely that the gp16 ATPase pentamer is pre-assembled. Instead, gp16 forms a pentamer by virtue of binding to the pentameric pRNA at the motor vertex. It's been shown that gp16 ATPase can either assemble directly on the pRNA and then bind the DNA-gp3 substrate, or first interact with DNA-gp3 and then attach to the pRNA (Koti *et al.,* 2008). Consequently, it's still not clear which assembly pathway is utilized in infected bacteria.

4.2 Arg finger and gp16 ATPase Classification

As mentioned above, φ29 ATPase is classified as an ASCE P-loop NTPase based on conserved WalkerA/B motifs, the Cat E that immediately follows Walker B motif, and an arginine finger. Though the spatial position of arginine finger is highly conserved among different branches of ASCE superfamily, its sequence position varies, which has caused a controversy as to which branch of ASCE superfamily φ29 ATPase should be assigned (Burroughs *et al.,* 2007). The Arginine finger within the HerA/FtsK superfamily is located at the base of strand β-4 and inserts its side chain into the active cleft of an adjacent monomer in an oligomeric ring structure. While in the Helicase superfamily, the characteristic arginine finger is also at the N-terminal end of strand β-4 in one of the two nucleotide binding domains. Here, arginine finger inserts its side chain into a neighboring active site in an inter-domain rather than

inter-subunit fashion. In contrast, for members from Terminase Large Subunit (TLS) family, the arginine finger is in the middle Walker B motif downstream of strand β-3 (Burrough *et al*, 2007). Unexpectedly, crystal structure of T4 ATPase revealed that arginine finger here functions *in cis* rather than *in trans* (Sun *et al.*, 2007).

Before solving gp16 ATPase structure, arginine residues from two positions were proposed to serve as arginie finger candidates, Arg26 in the Walker A motif and Arg14 at the base of strand β-4, which suggest different classifications of the φ29 ATPase within ASCE superfamily. Our gp16 ATPase structure demonstrates that like T4, a putative arginine finger, Arg26, also exists. Although it is in a slightly different position and conformation than in T4, minor movements or rotamer shifts would allow it interact with the ATP molecule within the ATPase monomer, as in T4. In contrast, based on superposition with the FtsK translocase hexamer, Arg148 is also pointed into the active site of its neighbor, and is in good position to trigger ATP hydrolysis there. Hence, we cannot definitively use the position of the Arginine finger to discriminate between the two possible branches of the ASCE superfamily. However, regardless of which branch our structure should occupy, based on mechanistic arguments presented below, we believe the φ29 ATPase functions more like the ring ATPases such as the Ftsk/herA and RecA superfamily.

4.3 DNA Packaging Mechanisms based on superposition with other ASCE superfamily members

Though extensive genetic, biochemical, biophysical, structural and single particle information emerged during past years regarding genome encapsidation in φ29, the fundamental mechanism of DNA translocation remains elusive. Up to now, rotary models, linear models, and models with non-integer step size have been described. Currently, the most prevalent possible packaging models are based on structural homology with motor ATPases from T4, the FtsK translocase, and the ssRNA virus φ12. Each of these mechanisms can now be evaluated in light of our structrure.

1) Mechanism based on phage T4 gp17 (Terminase Large Subunit (TLS) branch of

the ASCE superfamily)

ATP hydrolysis in gp17 triggers movement of its C-terminal nuclease domain relative to its N-terminal ATPase domain, which in turn pushes DNA, bound to the nuclease domain, into the procapsid (Sun *et al.*, 2008). Although the ATPase domains of φ29 gp16 and T4 gp17 are quite similar, their C-terminal domain differs (Figure 15). T4 gp17 C-terminus is a nuclease domain responsible for cutting and packaging DNA, while φ29 packages a unit length genome, and a nuclease domain is not necessary. It's possible that gp16 C-terminal OB domain plays a similar role, however we suspect that they cannot use similar packaging strategies since 1) the relative arrangement of gp16 C-terminal OB domain relative to the N-terminal ATPase domain in φ29 is quite different from the relative arrangement of the analogous domains in T4, which would result in a different domain coordination manner compared to T4; 2) the position of Arg finger in T4 gp17 favors an *in-cis* activation mechanism within a multimeric ring structure, while spatial position of Arg finger in φ29 gp16 prefers an *in-trans* activation fashion within a ring-like structure (Sun *et al.*, 2007; Burroughs *et al.*, 2007).

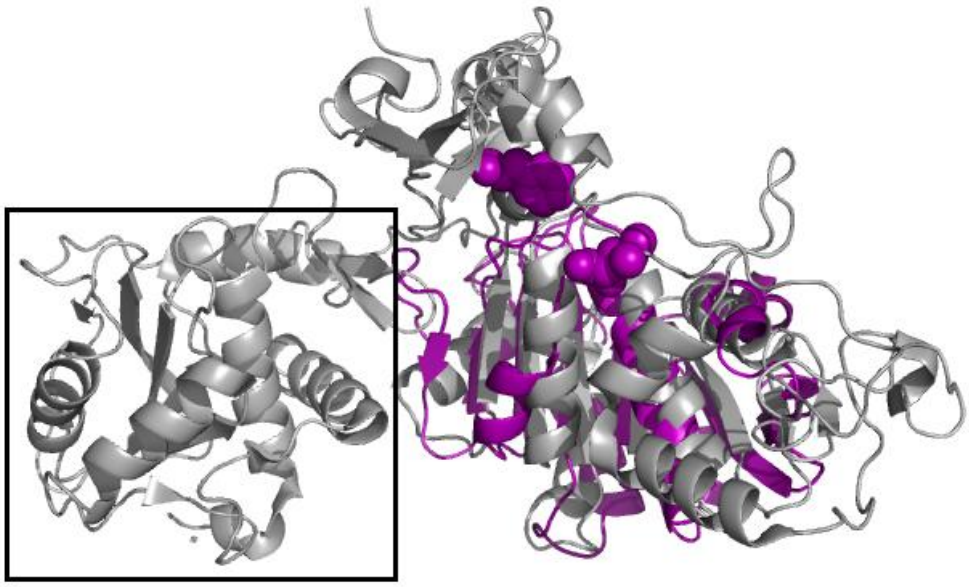


Figure 15. Structural comparison of φ29 gp16 ATPase and T4 gp17. gp16 ATPase and gp17 are colored in purple and gray, respectively. The N- and C-termini of gp16 ATPase are shown in spheres, and the C-terminus of gp17 is enclosed in a black box. It's not likely that the

C-terminus of gp16 will be present at a similar position as that of gp17 C-teminal nuclease domain.

2) Mechanism in FtsK translocase (FtsK/HerA branch of the ASCE superfamily)

Similar to T4, the proposed DNA translocation mechanism used by ftsK invokes the relative movement of two domains upon ATP hydrolysis in the ATPase domain; ATP hydrolysis induced conformational changes result in binding and release of DNA with respect to α- and β- domains of FtsK alternatively. Thus, ~1.6 bp DNA is translocated per subunit within the hexameric FtsK ring (Massey *et al.*, 2006). Like T4, the proposed mechanism relies on movement of an additional domain not present in φ29 to pump DNA into the head. However, in this case, superposition of the ATPase domain of φ29 gp16 onto the Ftsk β-domain would place the C-terminal OB domain in approximately the same position as the FtsK α-domain (Figure 16). Hence, we cannot rule out a similar mechansim for φ29 where OB domain functions analogously to the α-domain in FtsK.

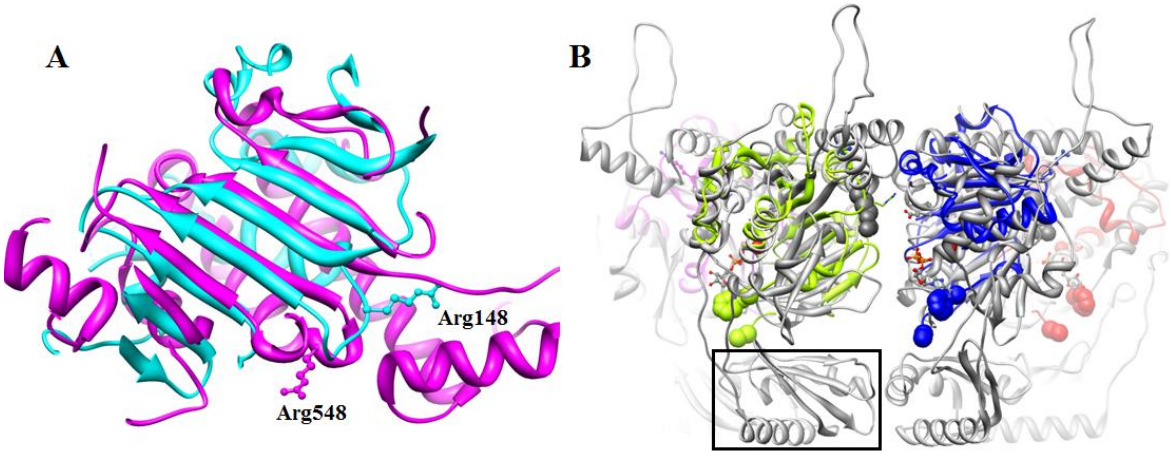


Figure 16. Structural comparison of φ29 gp16 ATPase and the bacterial FtsK translocase. (A) Superposition φ29 gp16 ATPase (cyan) onto FtsK β-domain (purple) demonstrates the conserved central β-sheet, and arginine fingers arranged at relatively the same spatial position are shown in sticks. (B) Side view of superposition of two gp16 ATPases (yellow and blue) onto Ftsk motor domain hexamer (gray). The N- and C-termini of gp16 ATPase are shown in spheres, and the α-domain of FtsK is enclosed in a black box. It's possible that C-terminal domain of gp16 ATPase would be in a similar location as the α-domain of FtsK and thus

function similarly during DNA translocation.

3) Mechanism in phage φ12 P4 (RecA branch of the ASCE superfamily)

Rather than relying on movement of an additional domain to drive DNA translocations, φ12 uses a Lys241 containing luminal loop within its hexamer channel to package DNA. Hydrolysis of ATP causes this luminal loop to change position, which exerts a force on the ssRNA genome that moves it into procapsid (Mancini *et al.*, 2004). Interestingly, φ29 has an analogous lysine, Lys124, at relative position as Lys241 in φ12. Furthermore, both gp16 ATPase and the φ12 ATPase have similarly positioned arginine fingers capable of *in-trans* activation of ATP hydrolysis in neighboring subunits (Figure 17). Thus, it's possible that φ29 uses DNA packaging mechanism similar to the φ12 P4 ATPase.

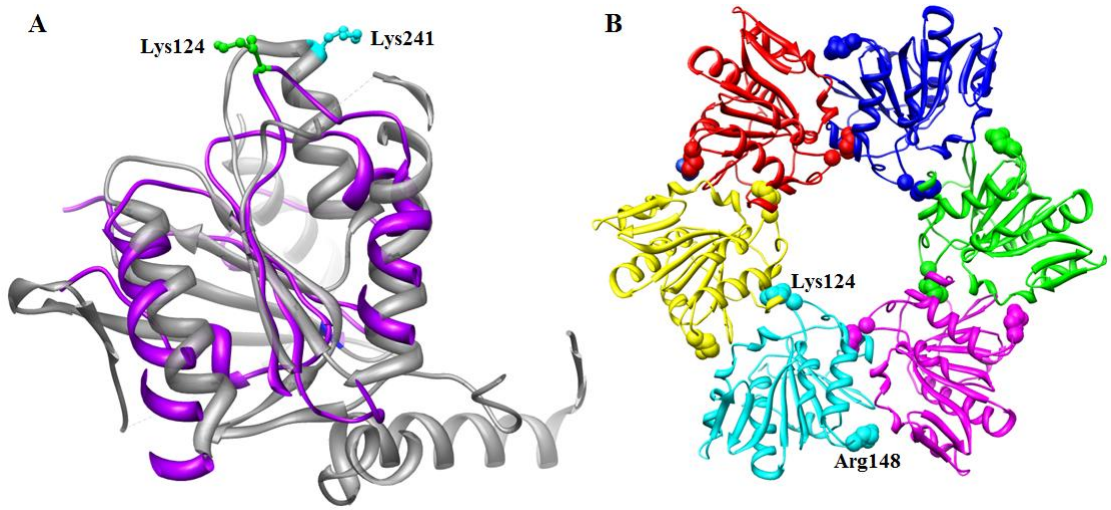


Figure 17. Structural comparison of φ29 gp16 ATPase to φ12 P4. (A) Superposition of φ29 gp16 ATPase (purple) onto φ12 P4 (gray), showing the conservation of central β-sheet between the two proteins. Lys241 that is responsible for DNA translocation in φ12 P4 is shown in sticks and colored in cyan, and a spatially equivalent Lysine, Lys124 that may play a similar role in DNA packaging for φ29 gp16 ATPase is also shown in sticks and colored in green. (B) Result of superimposing six gp16 ATPases (colored in cyan, yellow, red, blue, green, and purple in order) onto the P4 hexamer. Lys124 that is likely to contact DNA is located inside the ring, while Arg148 inserts into the neighboring ATPase site. Both Arg and Lys residues are shown in van der Waals spheres and two residues with the cyan monomer are

labeled. Though gp16 ATPase ring is modeled as a hexamer, a similar mode of operation is entirely consistent with a pentameric ring

4.4 Possible Packaging Mechanism

Several mechanistic generalizations can be drawn from the above analyses of the characteristic functional motifs in ASCE ATPases and superpositions with other ASCE superfamily members. When applied to φ29 packaging ATPase, a picture of its role in genome encapsidation begins to emerge. From superposition with related ring-forming ATPases, it seems likely that ATP and Mg^{2+} are bound at an inter-subunit interface such that adjacent monomers both contribute residues to the active site. Hence, ATP binding and hydrolysis directly affects adjacent subunits and the motor can thus respond to allosteric interactions between multiple ATP and DNA binding sites. In a ring-like structure, this arrangement can facilitate coordination of conformational changes required for DNA translocation. The glycine-rich Walker A P-loop and the adenine binding motif within the active site cleft recognize and bind the adenine base in ATP and thus help correctly orient the nucleotide substrate for hydrolysis. Walker A and Walker B loops both contribute to binding the catalytic Mg^{2+} ion. The catalytic glutamate immediately following the Walker B aspartate helps to activate the Mg^{2+}-bound water molecule for nucleophilic attack on the γ-phosphate of ATP. In order to couple coordinated conformational changes to particular nucleotide states, the φ29 ATPase likely uses the phosphate sensor at the tip of strand β-4 to distinguish ATP from its hydrolysis products. The arginine finger at the base of strand β-4 stabilizes the pentavalent transition state and phosphate leaving group during hydrolysis and, because the arginine finger belongs to a neighboring subunit, this *in trans* activation might also be important in initiating the next ATP hydrolysis event in a neighboring subunit in the oligomeric ring.

During each power stroke, the positively charged lumenal Lys124 from one subunit is electrostatically steered to the DNA backbone as in P4 (Fig. 13A). Upon engagement with the DNA, the lever then pushes sterically, orthogonal to the backbone axis; assuming B-form DNA geometry, this push will cause the

right-handed DNA helix not only to be translated in the packaging direction, but also rotated in a left-handed direction, an prediction that has now been confirmed by single particle experiments (Hetherington and Bustamante, personal communication). It is not entirely clear how ATP hydrolysis is coupled to force generation by the lumenal lysines, but it was proposed that, similar to the F_1-ATPase, movement of loops in the active site is coupled to twisting of the central β-sheet, and that this deformation is stored as elastic energy that can be used by the lumenal lysines to apply a force on the DNA (Abrahams *et al.,* 1994; Kabaleeswaran *et al.,* 2006; Yu *et al.,* 2010).

The proposed reaction scheme begins with one subunit bound to DNA. Assuming that the subunit initially interacting with DNA is in an ATP-bound state, then the electrostatic coupling between the lysine lever and the DNA will be tight (Chemla *et al.,* 2005; Moffitt *et al.,* 2009), thus holding the DNA in place and preventing slippage. It has been shown that affinity for DNA decreases in the ADP bound state (Chemla *et al.,* 2005). Hence the lever affinity for DNA would decrease upon ATP hydrolysis at end of the power stroke that accompanies Pi release (Chemla *et al.,* 2005) causing the DNA to roll to the lever on the adjacent subunit. If the next subunit in the ring is in the ATP-bound state, it has a high affinity for DNA and can thus accept the incoming DNA. The proposed rolling motion thus carries the DNA from one subunit to the next. Similar to P4 and FtsK, each power stroke also causes the arginine finger in the active subunit to be inserted into the active site of its neighbor, thus promoting hydrolysis in the next subunit around the ring (Lisal & Tuma 2005; Kainov *et al.,* 2008). However, because ADP release is slow, when it comes time for the fifth subunit to fire, subunit 1, the next subunit in a pentameric ring is still bound to ADP, thus having a low affinity for DNA and rendering it incapable of accepting the rolling DNA substrate. As a result, at the end of every four power strokes, the cycle pauses, constituting a dwell phase before the fifth subunit can fire, during which time four ATPs are loaded into the catalytic sites. This prediction is consistent with single particle experiments showing that DNA moves in 10 base pair bursts (consisting of 4 2.5 base pair substeps) followed by long pauses (Smith *et al*

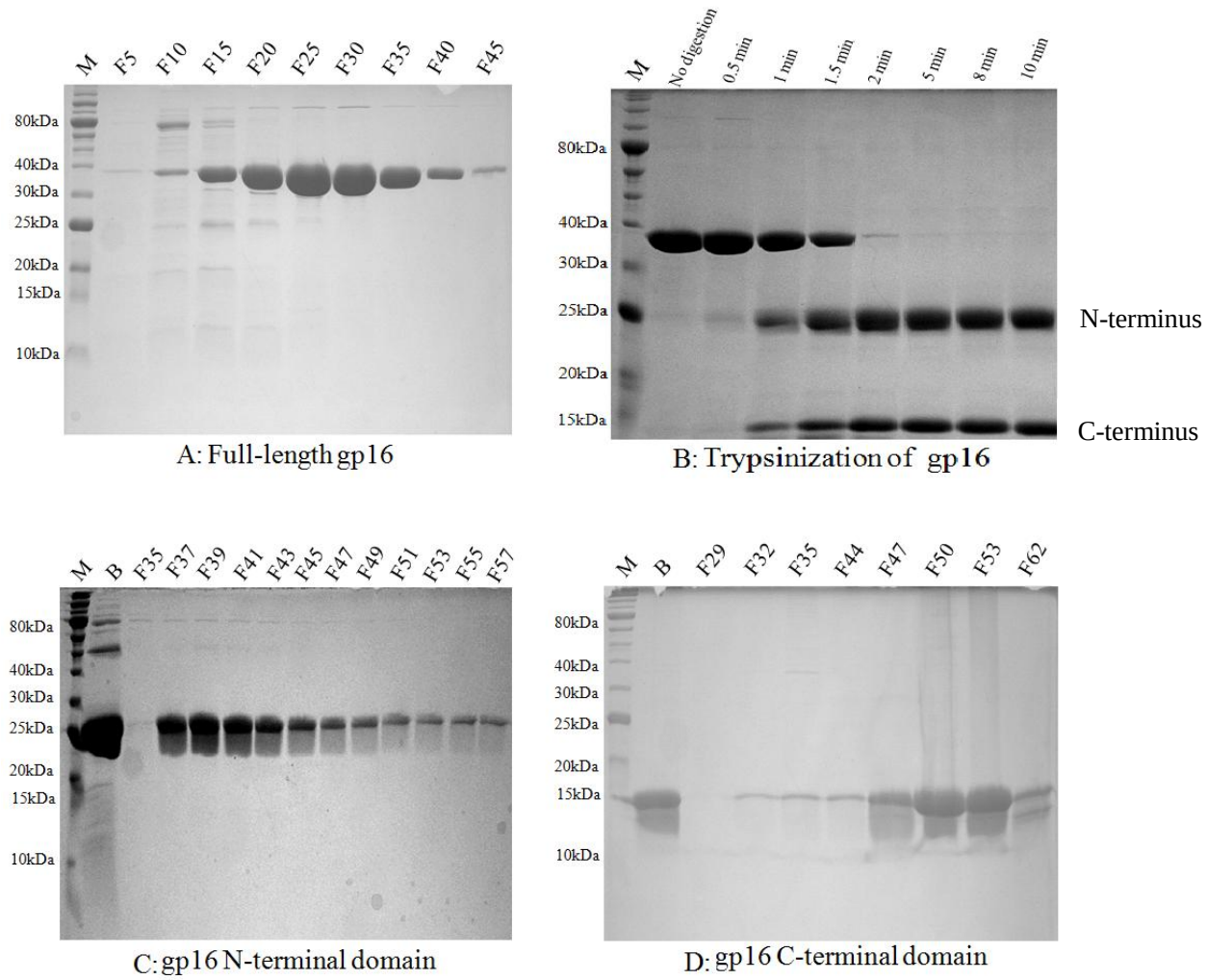
2000; Moffit *et al*. 2009). The next burst phase of four power strokes starts once spontaneous ATP hydrolysis takes place in the fifth site without insertion of an arginine finger. A necessary consequence of this mechanism is that the order of subunit firing changes each cycle. For example, if the subunit firing order of a particular cycle is 1,2,3,4 then the firing order in the subsequent cycle will be 5, 1, 2, 3 and so on. In summary, the push-and-roll model naturally accounts for the observed frequency and duration of the dwell phase, explains how a four 2.5 base pair step cycle arises from a planar pentameric motor, and provides a new perspective on how a multimeric ATPase might transport DNA.

Appendix

Supplementary Figures

Supplementary Figure 1: Purification and limited trypsinization of gp16 and purification of resultant N- and C-terminal domains.

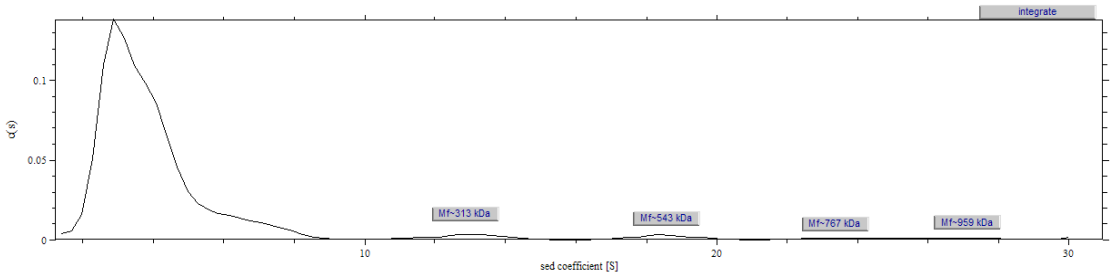
Full length gp16 was subjected to 2% trypsin treatment which was terminated at different time points by addition of SDS sample buffer. Two major bands resulted from this experiment that were sent for mass spectrometry (MALDI-TOF/TOF) analyis and protein N-terminal sequencing after transfer to a PVDF membrane. Individual clones of indentified gp16 N- and C-terminal domains were then made for protein expression and purification. Due to instability and self-aggregation of full-length gp16, 100mM imidazole was included in protein buffer. A: SDS-PAGE gel of full-length gp16 after Talon resin; B: SDS-PAGE gel of limited trypsinization of full-length gp16; C: SDS-PAGE gel of gp16 N-terminal domain (gp16 ATPase) after gel filtration; D: SDS-PAGE gel of gp16 C-terminal domain after gel filtration. (M: protein marker; B: before gel filtation, F5: 5th fraction; F10: 10th fraction; 1 min: 1 min of trypsinization.)



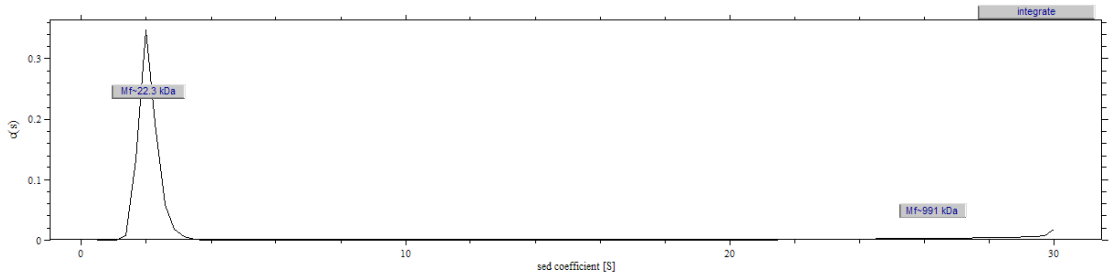
Supplementary Figure 1: Protein purification and limited trypsinization.

Supplementary Figure 2: Analytical ultracentrifugation (AUC) analysis of full length gp16 and C-terminal domain.

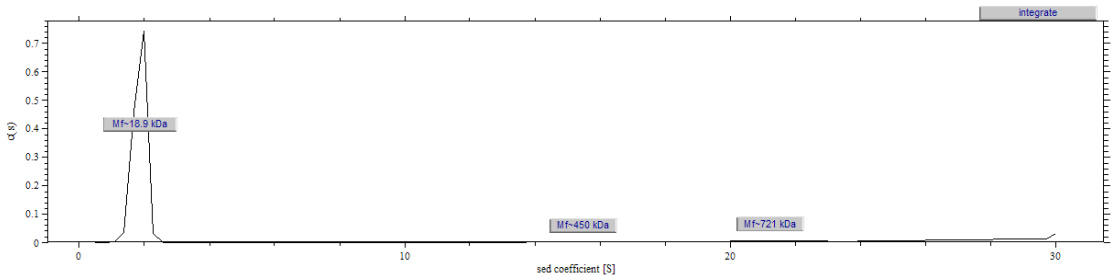
Oligomerization states of full-length gp16 and the gp16 C-terminal domain were identified via sedimentation velocity at 18℃. Protein was concentrated to desired concentrations before AUC. Full-length gp16 was concentrated to 0.38mg/ml (A), 0.54mg/ml (Figure 9), and AUC data were measured at A_{280}. The gp16 C-terminal domain was concentrated to 3.12mg/ml (B), 5.43mg/ml (C), and 7.92mg/ml (D), and AUC data were collected at A_{299}. For full-length gp16, 100mM imidazole was removed by overnight dialysis. Though part of gp16 C-terminal domain aggregated during AUC, the aggregates sedimented so fast that it didn't interfere with subsequent data fitting.



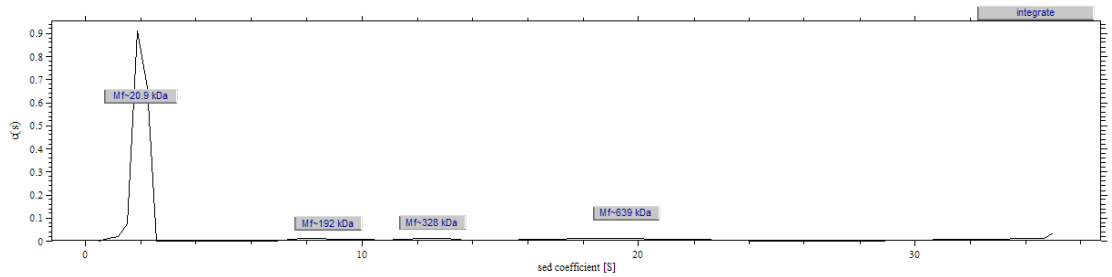
A: Full-length gp16 (0.38mg/ml)



B: gp16 C-terminal domain (3.12mg/ml)



C: gp16 C-terminal domain (5.43mg/ml)

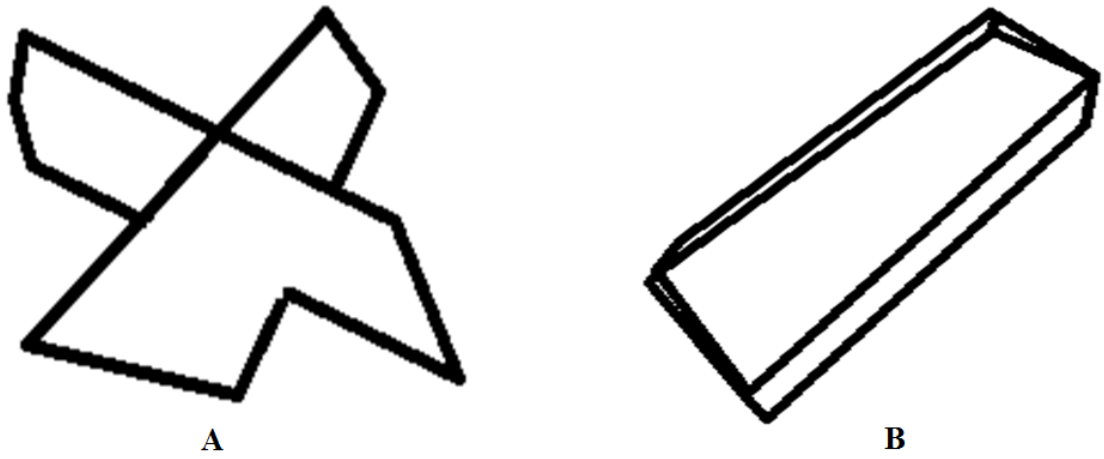


D: gp16 C-terminal domain (7.92mg/ml)

Supplementary Figure 2: AUC analysis of oligomerization states gp16 and its C-terminal domain.

Supplementary Figure 3: Crystals of gp16 N-terminal domain and AMP-PNP complex.

N-terminal domain of gp16 was mixed with 0.5mM ANP-PNP followed by crystal screening. Two conditions were found that gave crystals with different shapes, but with exactly the same unit cell parameters. Two flat crystals within condition A (2.8M sodium acetate trihydrate, pH7.0) crossed each other, while a single cuboid crystal was presented in condition B (3.5M sodium formate, pH7.0).



Supplementary Figure 3: Crystal shapes of N-terminal gp16 complexed with AMP-PNP found in two conditions A (A) and B (B). Original crystals were used in diffraction experiments before photographs could be taken.

5. References

Abrahams, J. P., A. G. Leslie, R. Lutter and J. E. Walker (1994). Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. *Nature.* 370(6491): 621-628.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart (2010). PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Cryst* D66: 213-221.

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

Chemla, Y. R., K. Aathavan, J. Michaelis, S. Grimes, P. J. Jardine, D. L. Anderson and C. Bustamante (2005). Mechanism of force generation of a viral DNA packaging motor. *Cell.* 122(5): 683-692.

Ding, F., C. Lu, W. Zhao, K. R. Rajashankar, D. L. Anderson, P. J. Jardine, S. Grimes, A. Ke, (2011). Sturcutre and assembly of the essential RNA ring component of a viral DNA packaging motor. *Proc Natl Acad Sci U S A.* 108(18): 7357-7362.

Draper., B, V. B. Rao. (2007). An ATP hydrolysis sensor in the DNA packaging motor from bacteriophage T4 suggests an inchworm-type translocation mechanism. *J. Mol. Biol.* 369:79-94.

Grimes, S., D. L. Anderson. (1997). The bacteriophage ϕ29 packaging proteins supercoil the DNA ends. *J. Mol. Biol.* 266:901-914.

Grimes, S., P. J. Jardine and D. Anderson (2002). Bacteriophage phi 29 DNA packaging. *Adv Virus Res.* 58: 255-294.

Guo, P., S. Erickson, D. Anderson. (1987). A small viral RNA is required for *in vitro* packaging of bacteriophage Phi29 DNA. *Science.* 236:690-694.

Guo, P., C. Peterson, D. Anderson (1987a). Initiation events in in-vitro packaging of bacteriophage phi 29 DNA-gp3. *J Mol Biol.* 197(2): 219-228.

Hendrix, R. W. (1978). Symmetry mismatch and DNA packaging in large bacteriophages. *Proc Natl Acad Sci U S A.* 75(10): 4779-4783.

Iyer, L. M., K. S. Makarova, E. V. Koonin and L. Aravind (2004b). 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(17): 5260-5279.

Kainov, D. E., E. J. Mancini, J. Telenius, J. Lisal, J. M. Grimes, D. H. Bamford, D. I. Stuart and R. Tuma (2008). Structural basis of mechanochemical coupling in a hexameric molecular motor. *J Biol Chem.* 283(6): 3607-3617.

Lee, T. J. and P. Guo (2006). Interaction of gp16 with pRNA and DNA for genome packaging by the motor of bacterial virus phi29. *J Mol Biol* 356(3): 589-599.

Langer, G., S. S. Cohen, V. S. Lamzin and A. Perrakis (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc.. 3(7): 1171-1179.

Lisal, J. and R. Tuma (2005). Cooperative mechanism of RNA packaging motor. *J Biol Chem.* 280(24): 23157-23164.

Mancini, E. J., D. E. Kainov, J. M. Grimes, R. Tuma, D. H. Bamford, D. I. Stuart. (2004) . Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell.* 118:743-755.

Massey, H. M., C.P. Mercogliano, J. Yates, D. J. Sherratt, J. Löwe. (2006). Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol Cell*. 23: 457-469.

Minor, M., M. Cymborowski, Z. Otwinowski, M. Chruszcz. (2006). HKL3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Cryst.* D62: 859-866.

Mitchell, M. S., S. Matsuzaki, S. Imai and V. B. Rao (2002). Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. *Nucleic Acids Res.* 30(18): 4009-4021.

Moffitt, J. R., Y. R. Chemla, K. Aathavan, S. Grimes, P. J. Jardine, D. L. Anderson, C. Bustamante (2009). Intersubunit coordination in a homomeric ring ATPase. *Nature* 457(7228): 446-450.

Morais, M. C. 2011. "Structure and Assembly of the Bacteriophage ϕ29 dsDNA Packaging Motor", *XXIInd Biennial Conference on Phage/Virus Assembly*, Port Aransas, TX.

Morais, M. C., S. Kanamaru, M. O. Badasso, J. S. Koti, B. A. L. Owen, C. T. McMurray, D. L. Anderson, and M. G. Rossmann. (2003). Bacteriophage φ29 scaffolding protein gp7 before and after prohead assembly. *Nat. Struct. Biol.*, 10:572-576.

Morais, M. C., J. S. Koti, V. D. Bowman, E. Reyes-Aldrete1, D. L. Anderson, M. G. Rossmann. (2008). Defining molecular and domain boundaries in the bacteriophage ϕ29 DNA packaging motor. *Structure*. 16(8):1267-1274.

Rao, V. B. and M. Feisis, (2008). The bacteriophage DNA packaging motor. *Annu. Rev. Genet*. 42:746-81.

Reid, R. J., J. W. Bodley and D. Anderson (1994b). Identification of bacteriophage phi 29 prohead RNA domains necessary for in vitro DNA-gp3 packaging. *J Biol Chem.* 269(12): 9084-9089.

Reid, R. J., F. Zhang, S. Benson and D. Anderson (1994c). Probing the structure of bacteriophage phi 29 prohead RNA with specific mutations. *J Biol Chem*. 269(28): 18656-18661.

Reilly, B. E. (1965). A study of the baceteriophage of Bacillus subtilis and their infectious nucleic adics. Ph.D thesis. Case Western Reserve University, Detroit, Michigan.

Rickgauer, J. P., D. N. Fuller, S. Grimes, P. J. Jardine, D. L. Anderson and D. E. Smith (2008). Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage phi29. *Biophys J.* 94(1): 159-167.

Rossmann, M. G., D. Moras and K. W. Olsen, (1974). Chemical and biological evolution of nucleotide-binding protein. *Nature* **250**(463): 194-199.

Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. (2000) Structure of the bacteriophage Φ29 DNA packing motor. *Nature (London)*, 408: 745-750.

Story, R. M., T.A. Steitz. (1992). Structure of the RecA protein ADP complex. *Nature*. 355: 374-376.

Sun, S., K. Kondabagil, B. Draper, T. I. Alam, V. D. Bowman, Z. Zhang, S. Hegde, A. Fokine, M. G. Rossmann and V. B. Rao (2008). "The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell*, **135**(7): 1251-1262.

Sun, S., K. Kondabagil, P. M. Gentz, M. G. Rossmann and V. B. Rao (2007). The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. *Mol Cell*, **25**(6): 943-949.

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

Velankar, S. S., P. Soultanas, M. S. Dillingham, H. S. Subramanya, D. B. Wigley. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, 97(1):75-84.

Yu, J., J. Moffitt, C. L. Hetherington, C. Bustamante, G. Oster, (2010). Mechanochemistry of a viral DNA packaging motor. *J Mol Biol.* 400(2): 186-203.

Zhang, C., C. S. Lee and P. Guo (1994). The proximate 5' and 3' ends of the 120-base viral RNA (pRNA) are crucial for the packaging of bacteriophage phi 29 DNA.

Virology **201**(1): 77-85.

Zhang, C., T. Tellinghuisen and P. Guo (1995). Confirmation of the helical structure of the 5'/3' termini of the essential DNA packaging pRNA of phage phi 29. *RNA* **1**(10): 1041-1050.

Vita

Huzhang Mao is the youngest of four children born to Mr. Jianfa Mao and wife Mrs. Limei Jiang. He was born on the 1st of October 1986 in the county of Yujiang of Jingxi Province in China. He spent twelve years in Ou'tang Elementary School, Huangzhuang Middle and Yujiang NO.1 High School. He graduated with a Bachelor's of Scienc in Biological Technology from Nankai University in the city of Tianjin in 2009, soon after which he joined UTMB for his graduate studies.

Permanent address:

Mao Group, Ou'tang Village, Huzhangzhuang Township, Yujiang County, Jiangxi Province, China. 33204

This dissertation was typed by Huzhang Mao.