Title
Dissecting the Operating Mechanism of a Biological Motor One Molecule at a Time

Author
Chistol, Gheorghe

Publication Date
2013

Dissecting the Operating Mechanism of a Biological Motor One Molecule at a Time

by

Gheorghe Chistol

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor in Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos J. Bustamante, Chair
Professor Ahmet Yildiz
Professor Michael Botchan

Spring 2013

Dissecting the Operating Mechanism of a Biological Motor One Molecule at a Time

Abstract

Dissecting the Operating Mechanism of a Biological Motor One Molecule at a Time

by

Gheorghe Chistol

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Carlos J. Bustamante, Chair

Double-stranded DNA viruses, including most bacteriophages and mammalian herpesviruses, package their genomes into a pre-formed protein capsid during their self-assembly. DNA is compacted to near-crystalline densities at the end of packaging. This remarkable mechanical task is performed by a powerful ATP-driven molecular machine known as the packaging motor. Bacteriophage Phi29, a model system for studying DNA packaging, has a 19.3-kbp genome and its packaging motor is composed of a connector, an RNA scaffold, and a pentameric ring ATPase.

Ring ATPases of the ASCE superfamily perform a variety of cellular functions. An important question about the operation of these molecular machines is how the ring subunits coordinate their chemical and mechanical transitions. Here we present the first comprehensive mechanochemical characterization of a homomeric ring ATPase – Phi29 gp16 – which translocates dsDNA in cycles composed of alternating dwells and bursts. We use high-resolution optical tweezers to determine the effect of nucleotide analogs on the cycle. We find that ATP hydrolysis occurs sequentially during the burst and that ADP release is interlaced with ATP binding during the dwell, revealing a high degree of coordination among ring subunits. Moreover, we show that the motor displays an unexpected division of labor: although all subunits of the homo-pentamer bind and hydrolyze ATP during each cycle, only four participate in translocation whereas the remaining subunit plays an ATP-dependent regulatory role.

Several viral packaging motors have been shown to slow down as the capsid fills up with DNA, but it remains unclear how the packaging velocity is regulated. Here we use high-resolution optical tweezers to monitor the base-pair-scale packaging dynamics at various degrees of capsid filling. By comparing the burst duration at various degrees of capsid filling and different external forces, we estimate an internal force of ~20 pN at 100% filling, much lower than the motor's stall force. We find that the motor's step size is continuously modulated by capsid filling, in quantitative agreement with measurements of DNA rotation by the Phi29 packaging motor. In addition, we find the motor switches on and off at high filling by entering into long-lived pauses, which may allow DNA relaxation within the capsid. Together, our results reveal that the motor is not passively stalled by a large internal force at high filling as suggested by previous models. Instead, the motor is actively throttled down via several mechanisms in response to DNA encapsidation. The intricate crosstalk between the motor and the capsid plays a key role in orchestrating the molecular events leading to packaging termination and virus maturation, and may represent a general design principle shared by different viruses.

Dedication

To Constantin Ojog, Svetlana Ojog, Mihai Ojog, Alexandru Ojog,
Adela Dumitraşco, and Mihai Dumitraşco
for their love, support, and encouragement.

To my late mother, Valentina Chistol,
and my late grandmother, Ana Chistol,
who will be missed dearly.

Table of Contents

Chapter 1: Introduction

Chapter 2: Data Analysis

Chapter 3: Dissecting the Mechano-Chemical Cycle of a Homomeric Ring ATPase

Chapter 4: Dissecting the Throttle Control Mechanisms of a Viral Packaging Motor

Chapter 5: Additional Projects and Conclusions

List of Figures

Chapter 1: Introduction

Chapter 2: Data Analysis

Chapter 3: Dissecting the Mechano-Chemical Cycle of a Homomeric Ring ATPase

Chapter 4: Dissecting the Throttle Control Mechanisms of a Viral Packaging Motor

Chapter 5: Additional Projects and Conclusions

Acknowledgements

This thesis would not have been possible without the help and support of many talented and inspiring people. I am indebted to Shixin Liu who worked with me very closely on all the Phi29 projects. From Shixin I learned how to be rigorous and systematic, how to write in the most efficient possible manner, and how to tell a story in a paper. I also indebted to Jeffrey R. Moffitt, who trained me in the science of optical tweezers, and the art of single-molecule experiments. Jeff instilled in me a healthy sense of skepticism, a sense so vital for a scientist.

I would like to thank Christian Kaiser, Craig Hetherington, Rodrigo Maillard, Jae Yen Shin, Ninning Liu, Sara Tafoya, Maya Sen, Manchuta Dangkulwanich, Toyotaka Ishibashi, Lacramioara Bintu, Yara Meija, Adam Politzer, Shannon Yan, Daniel Goldman, Troy Lionberger, Sam Leachman, Cesar Dias Celiz, Marta Kopaczynska, Maurizio Righini, Antony Lee, Melania Strycharska, Yves Coello, Wei Cheng, Courtney Hodges, Ariel Kaplan, Aathavan Karunakaran, and all the other current and former members of the Bustamante and Tinoco labs for stimulating conversations and guidance, scientific as well as personal.

I wish to express a very special thanks to Liem Huynh for supporting me in the tough moments, inspiring me to explore alternatives, and brightening up my last year of graduate school. I want to thank Anne Takizawa, Natalie Thung, and Melina Winterton for helping me navigate the labyrinths of bureaucracy at UC Berkeley. I am indebted to Ignacio Tinoco, Andreas Martin, Shelley Grimes, Paul J. Jardine, Ahmet Yildiz, and Michael Botchan for the friendly guidance, thought provoking suggestions, and the general collegiality that each of them offered to me over the years.

Finally I would like to express my deep appreciation and gratitude to my advisor, Dr. Carlos J. Bustamante, for his mentorship, unwavering support, and for inspiring me to pursue a career in science.

Chapter 1
Introduction

1.1 The Phi29 dsDNA Packaging Motor

Phi29 is a dsDNA bacteriophage that pumps its 19.3 kb genome into a preformed capsid using a powerful packaging motor. Upon the completion of packaging, the packaging motor partially dissociates from the full prohead and a very elaborate tail structure is assembled onto the capsid portal. Phi29 infects its host bacterium, *Bacillus subtilis*, by injecting viral DNA through the tail assembly. DNA packaging is a remarkable process that must overcome large electrostatic, entropic, and bending energies required to compact DNA to near-crystalline densitiesities. A large class of virusses, including herpesviruses, adenovirusses, poxviruses, and tailed bacteriophages, employ portal packaging motors to translocate their linear genomes into protein shells (Black 1989, Bazinet and King, 1985).

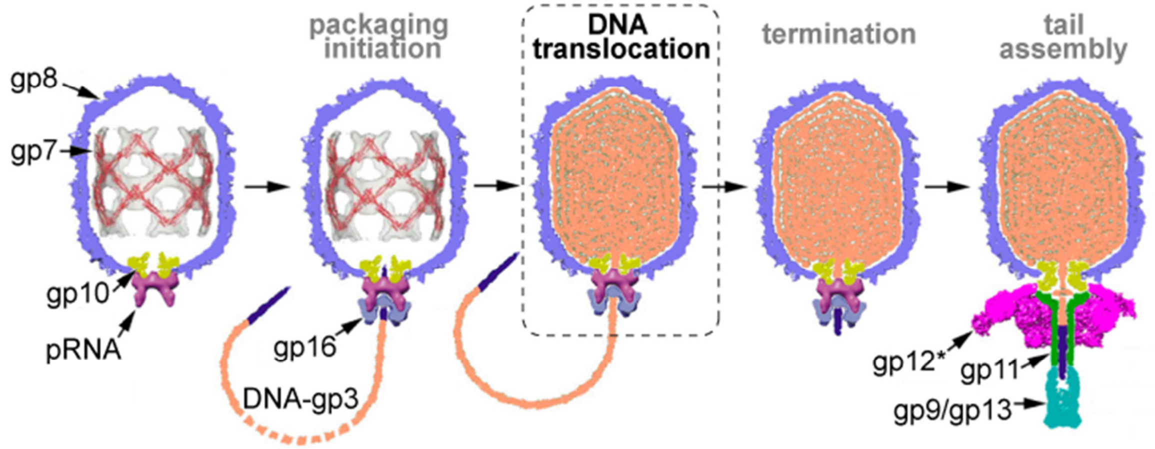


Figure 1.1.1: Phi29 Viral Assembly

The Phi29 capsid self-assembles from gp8 subunits around a gp7 protein scaffold. The connector is a 12-meric ring of gp10. A pentameric ring of pRNA assembles at the open vertex of the capsid. The Phi29 ATPase, gp16, then assembles onto the pRNA scaffold to form a pentameric ring. During initiation, the Phi29 packaging motor recognizes the gp3-labeled end of the viral genome and begins to translocate DNA into the prohead. After the entire 19.3-kb genome has been packaged, the pRNA and gp16 rings dissociate from the portal and the tail is assembled. Figure courtesy of Paul J. Jardine.

The Phi29 packaging motor consists of 3 components that assemble at one of the vertices of the icosahedral prohead (Morais et al., 2008): a protein connector ring mounted inside the capsid (Figure 1.1.2A, cyan), a prohead RNA ring (pRNA) that assembles outside the capsid (Figure 1.1.2A, magenta), and a ring of ATPase proteins that assemble onto the pRNA scaffold (Figure 1.1.2A, blue). The connector ring consists of 12 identical subunits (gp10) that form a cone-like structure ~7.5nm tall, and ~13.8nm in diameter at its widest, with a central channel large enough to accommodate dsDNA (Simpson et al., 2000). The pRNA ring consists of five identical subunits (Morais et al., 2008) that form a scaffold attaching the ATPase to the capsid. The wildtype pRNA contains 174 nucleotides, however the motor can also function with a truncated, 120-nt version of the pRNA (Zhao et al., 2008). Although pRNA is unique to Phi29 and its close relatives (the *Podoviridae* family of tailed bacteriophages), it performs the role of scaffold proteins in other tailed bacteriophages. The ATPase ring consists of five identical gp16 subunits that assemble onto the pRNA scaffold forming a central pore through which DNA is translocated into the capsid (Morais et al., 2008). Gp16 is the active motor component that couples the chemical energy stored in ATP to mechanical work needed to pump DNA into the capsid (Guo et al., 1987). Gp16 is related to a large superfamily of proteins that translocate a polymer substrate through a central pore, including helicases, protein unfoldases, and other DNA translocases (Iyer et al., 2004). Unlike most ring ATPases which are hexameric, the Phi29 ATPase is pentameric, presumably because the icosahedral capsid imposes its five-fold symmetry on the pRNA and ATPase rings (Morais et al., 2008).

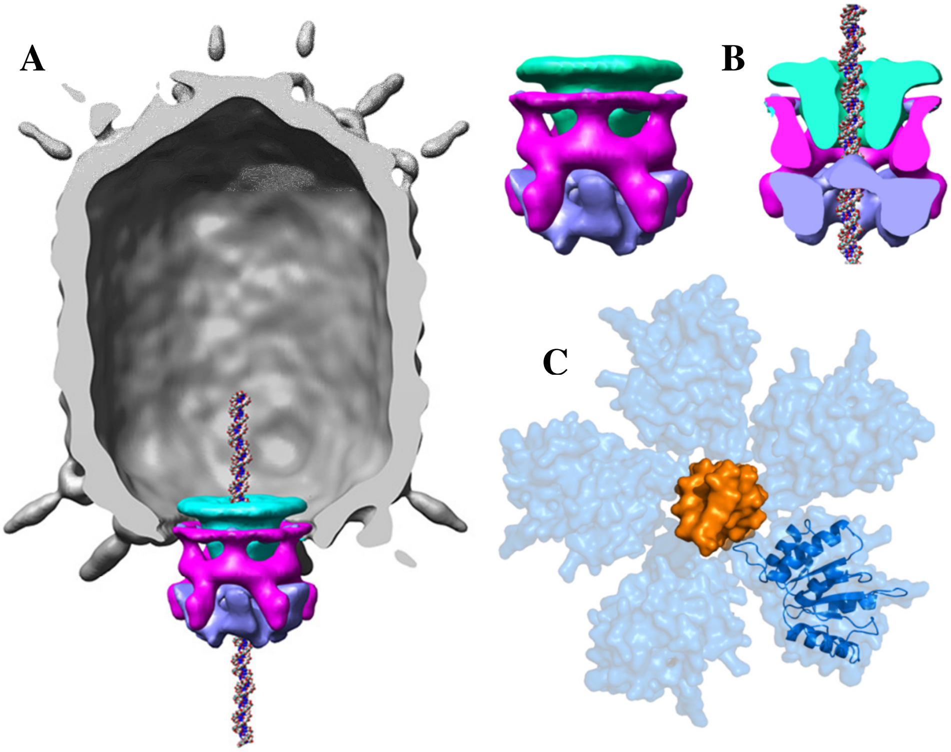


Figure 1.1.2: The Structure of the Phi29 Packaging Motor

(A) CryoEM reconstruction of the Phi29 capsid (gray) and packaging motor consisting of a connector ring (cyan), pRNA ring (magenta), and ATPase ring (blue). For scale, a model of dsDNA was inserted through the motor. (B) Cross section of the Phi29 packaging motor. (C) Structural model of the Phi29 ATPase ring (for details see Chapter 5.1). One gp16 subunit is highlighted with a ribbon diagram, dsDNA is shown in orange. CryoEM reconstruction images courtesy of Marc C. Morais.

The Phi29 packaging motor has been studied extensively using traditional biochemical, structural and single-molecule techniques, and it has become not only a model system for viral packaging motors, but also a model system for understanding ring ATPases. To investigate how the Phi29 motor operates, I used a dual trap high resolution optical tweezers assay illustrated in Figure 1.1.3A. Packaging motors, proheads, and DNA were first assembled in a test-tube, where packaging was then initiated by adding ATP. One end of the DNA is covalently linked to gp3, a small protein required to initiate packaging. After 30-45 seconds DNA packaging was halted by adding ATPgS, a non-hydrolyzable nucleotide analog. The resulting stalled complexes were then incubated with micron-sized steptavidin-coated beads which bound the biotinylated end of DNA outside the capsid. The single-molecule experiment was performed in a micro-fluidics chamber where a streptavidin-coated bead was held in one optical trap, and an anti-prohead antibody-coated bead was held in the second optical trap. Upon bringing the two beads in close proximity to each other, the prohead bound to the antibody bead and formed a tether as depicted in Figure 1.1.3A. When ATP buffer was flown into the chamber, the motor resumed packaging and gradually reduced the length of DNA tethered between the beads. In addition to measuring the contour length of DNA versus time, optical tweezers were used to apply pico-Newton (pN) forces against the packaging motor, enabling us to study the coupling between the chemical and mechanical cycles of the motor.

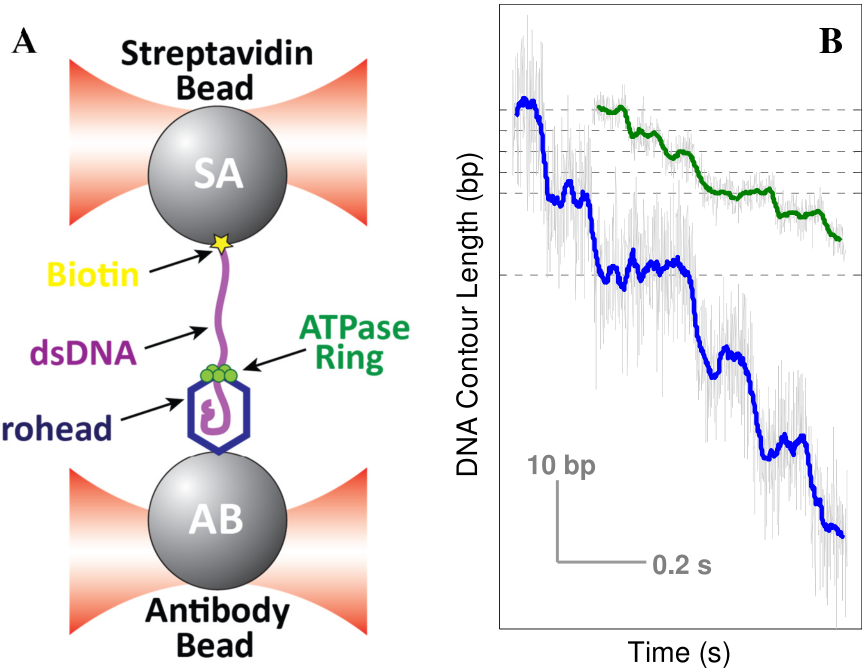


Figure 1.1.3: The Single Molecule DNA Packaging Assay
(A) The packaging complex consisting of a prohead, packaging motor and dsDNA tether is held between two micron-sized polystyrene beads by two optical traps. As the motor translocates DNA into the capsid, the DNA tether is shortened, pulling the two beads out of their respective traps. The force exterted by the trap on a bead is proportional to the bead displacement from the trap center. The distance between the two traps is intermittently adjusted to maintain the DNA tension within predefined limits. (B) Sample packaging traces at low force (7-10 pN, blue) and high force (30-40 pN, green). Raw data (light gray) was boxcar filtered. At low force Phi29 packages in nearly instantaneous 10-bp increments called bursts (blue curve). At high force, the burst breaks up into four 2.5-bp steps (green curve).

The first single-molecule Phi29 study discovered that the Phi29 packaging motor can generate mechanical forces in excess of 50 pN (Smith et al., 2001), more than ten times the force generated by molecular motors that power our muscles (Svoboda and Block, 1994). The same study found that packaging gradually slows down as DNA fills the prohead, leading to the hypothesis that a large internal pressure builds up within the viral capsid and that packaging stops when the internal force exceeds the stall force of the motor (Smith et al., 2001). When operating against small opposing forces (5-10 pN) the Phi29 motor translocates DNA in 10-bp bursts separated by ~100-ms dwells (Figure 1.1.3B) (Moffitt et al., 2009). Experiments at high external forces (30-50 pN) revealed that the burst phase consits of four 2.5-bp steps, indicating that four gp16 subunits translocate DNA during a single dwell-burst cycle (Moffitt et al., 2009). Past work showed that at least four ATPase subunits bind nucleotides during the dwell phase (Moffitt et al., 2009), and that inorganic phosphate release is the chemical transition which powers DNA translocation during the burst phase (Chemla et al., 2005). A series of very elegant experiments with modified substrates revealed that the motor makes crucial electrostatic contacts with the DNA backbone during the dwell phase (Figure 1.1.4), and that non-specific steric interactions propel DNA during the burst phase, helping to explain the non-integer nature of the 2.5 bp/subunit step-size (Aathavan et al., 2009).

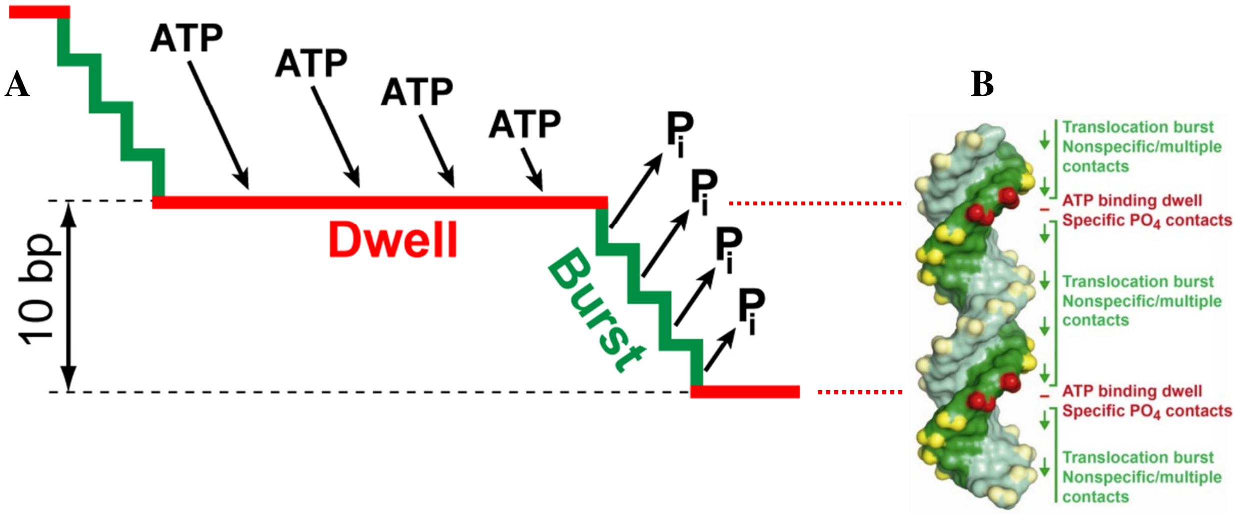


Figure 1.1.4: The Dwell-Burst Cycle of the Phi29 Packaging Motor
(A) The Phi29 motor translocates dsDNA in 10-bp bursts (green) separated by stationary dwells (red). The burst is comprised of four 2.5-bp steps, each powered by the release of Pi. At least four out of five ring subunits bind ATP sequentially during the dwell. (B) During the dwell phase, the motor makes a unique electrostatic contact with a pair of phosphates on the 5'-3' strand. During the burst phase, the motor pushes DNA via non-specific steric interactions.

Despite the wealth of knowledge accumulated over a decade of single-molecule experiments in our laboratory, there were several fundamental questions about the Phi29 packaging motor that remained unanswered. How are ATP binding, ATP hydrolysis and product release coordinated among gp16 subunits during the dwell-burst cycle? Why does this pentameric ATPase employ only four of its subunits to translocate DNA during the burst phase? What is the role of the special, non-translocating subunit? What are the factors that break the symmetry of the ATPase ring to designate such a special subunit? How does the motor function when the capsid is nearly full and how does it prevent accidental release of DNA due to the large internal pressure? This dissertation provides answers to these questions and more.

1.2 Ring NTPases Related to Phi29 Gp16

The Phi29 ATPase (gp16) belongs to the large superfamily of Additional Strand Conserved Glutamate (ASCE) proteins which use free energy derived from NTP hydrolysis to perform a wide array of cellular functions, ranging from replication initiation and transcription regulation, to cargo transport, chromosome segregation, protein degradation, and viral packaging (Lyubimov et al. 2011). Although proteins from this superfamily have the same basic fold, they may have unique accesory domains optimized for a particular task. For example many viral ATPases have a nuclease domain that can be activated to cut DNA when the capsid is filled (Black, 1989), transmembrane proteins such as FtsK have transmembrane domains (Massey et al., 2006), helicases such as Rho have secondary nucleic-acid binding region outside the central pore (Skordalakes et al., 2003).

Gp16 belongs to the PilT/FtsK branch within the RecA clade of the ASCE superfamily (Figure 1.2.1), together with other viral encapsidation proteins and dsDNA translocases. A large portion of ASCE proteins function as homomeric rings (Figure 1.2.1, red), with either five, six, or seven identical subunits, although hexamers are most common. These proteins share a common structural architecture consisting of five alpha helices (Figure 1.2.1, yellow) and four beta sheets (Figure 1.2.1, blue). The conserved Walker A (WA) and Walker B (WB) motifs are involved in nucleotide binding and hydrolysis, the arginine finger (R) is thought to stabilize the transition state of the nucleotide in the binding pocket and presumably participates in inter-subunit coordination (Figure 1.2.1). In the remainder of this section I will discuss a few ASCE ring ATPases in order to provide some context for understanding the Phi29 packaging motor.

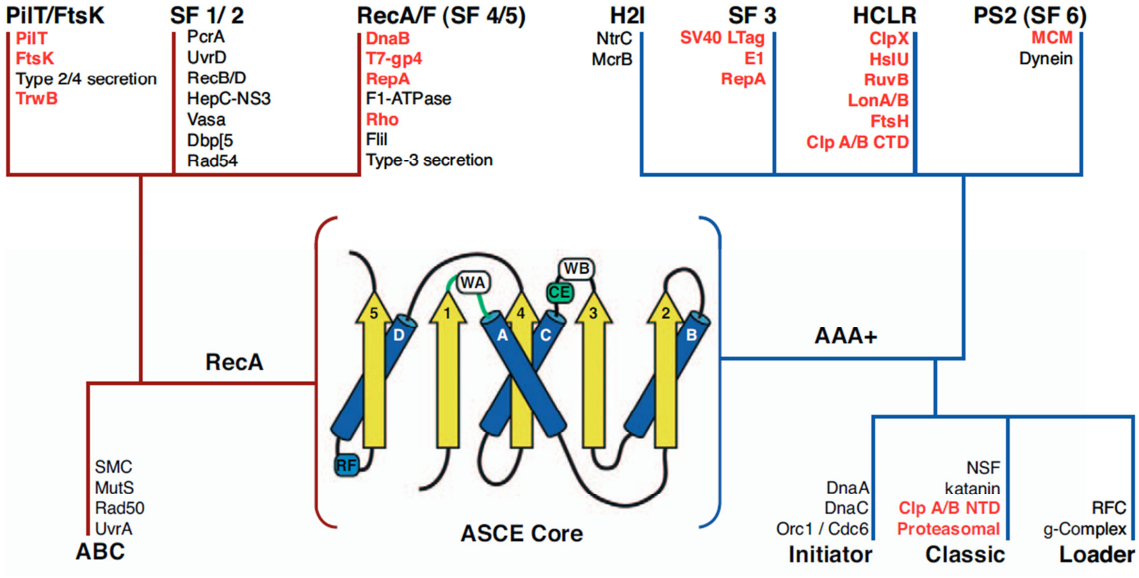


Figure 1.2.1: Additional Strand Conserved Glutamate (ASCE) Superfamily Tree
Many proteins from the ASCE superfamily are homomeric ring translocases (highlighted in red). The Phi29 gp16 ATPase belongs to the PilT/FtsK clade of the RecA-like branch of the ASCE superfamily.
Reprinted from Lyubimov et al., 2011 with permission from Elsevier © 2011.

E1, Rho, and DnaB helicases are the only homomeric ring NTPases that have been co-crystallized with their nucleic acid substrate and nucleotides bound to catalytic pockets (Enemark et al., 2006, Thomsen et al., 2009, Itsathitphaisarn et al., 2012). In all three cases the interface between two adjacent subunits forms a catalytic pocket where ATP binds to the Walker A/B motifs of one subunit, while the arginine finger from the other subunit stabilizes the transition-state of the nucleotide. In addition, all three structures contain single-stranded (ss) RNA or ssDNA bound inside the central channel of the ring. Pore loops extending from each ATPase subunit contain positively charged residues (usually lysines) that contact the negatively charged backbone of the nucleic acid. These crystal structures represent the state of the art when it comes to understanding the architecture of ring ATPases because all the molecules of interest are present in the “native” context. Unfortunately several efforts to co-crystallize dsDNA ring translocase with their substrates have not been successful.

The packaging motors of bacteriophages T4 and λ have been studied extensively using traditional biochemistry techniques (Rao and Black, 2000; Murialdo, 1991). Initial single-molecule experiments with T4 and λ packaging motors revealed that these molecular machines can generate very large forces (Cottadiel et al., 2012, Zhang et al., 2011, Tsai et al., 2009, Fuller et al., 2007) and exhibit large variations in packaging velocity from motor to motor, presumably hinting at a relatively uncoordinated ATPase operation. The T4 gp17 ATPase has been crystallized as a monomer, which was later fit into the electron density of a cryo-EM reconstruction of stalled complexes (Sun et al., 2008). The reconstructed T4 gp17 ring displays some unexpected and perhaps controversial features, such as an "internal arginine finger" that acts *in cis*, in contrast to most other ring ATPase crystal structures where the arginine finger acts *in trans*. The reconstructed T4 ATPase ring structure should be taken with a large grain of salt because the protein did not crystallize in its oligomeric form, and the electron density map used for fitting was of low resolution.

FtsK and SpoIIIE are fast dsDNA translocases from the same branch of the ASCE superfamily as Phi29 gp16. Their role is to pump DNA across a lipid membrane during bacterial sporulation. Single molecule experiments have revealed that FtsK/SpoIIIE can generate tens of pN of force, and are extremely fast, translocating DNA at speeds between 3000 and 6000 bp/s, about 30-60 times faster than the Phi29 packaging motor (Saleh et al., 2004, Crozat et al., 2010, Lee et al., 2012). Although FtsK has been crystallized as a hexameric ring (Massey et al., 2006), its structure does not shed light on motor-substrate interaction because dsDNA could not be co-crystallized. In addition, the FtsK structure is of limited use for understanding inter-subunit coordination because the catalytic pockets are widely open and the arginine finger is located too far from the nucleotide binding site (Massey et al., 2006). Nevertheless, both FtsK and SpoIIIE are fascinating molecular motors because of their incredibly high velocities and their ability to strip proteins bound to DNA (Chivers et al., 2010). In addition, both FtsK and SpoIIIE are highly amenable to single-molecule manipulations and measurements.

The ClpX protein unfoldase and the T7 gp4 DNA helicase are two intensely studied homomeric ring NTPases which share a lot of similarities, which is very surprising given that they operate on such different substrates. The crystal structures of these two ring motors exhibit clear structural asymmetries – only four out of the six catalytic pockets are occupied by a nucleotide, and the remaining two pockets are 180 degrees apart (Singleton et al., 2000; Massey et al., 2006). It has been proposed that this "dimer of trimers" asymmetry arises because of stress that builds up within the ring upon nucleotide binding. The atomic models of ClpX and T7 gp4 were initially thought to be consistent with a sequential rotary coordination of nucleotide binding and hydrolysis (Singleton et al., 2000; Massey et al., 2006). Single molecule experiments with the T7 gp4 helicase indicate a moderate degree of inter-subunit coordination in the NTPase ring (Sun et al., 2011), but the details of this mechanism remain unclear. Ensemble measurements of catalytic activity of single-chain ClpX hexamers were consistent with a stochastic inter-subunit coordination model, in which individual ClpX monomers would turnover ATP independently of each other (Martin et al., 2005). Preliminary single-molecule experiments with optical tweezers found that ClpX translocates the unfolded polypeptide in discrete steps (Aubin-Tam et al., 2011; Maillard et al., 2011). As part of the Maillard et al. team, I performed high-resolution measurements and found that ClpX translocated its substrate in increments of ~1 nm, ~2 nm, or ~3 nm, suggesting a partial degree of inter-subunit coordination and a fundamental step size of approximately 1 nm/subunit (Figure 1.2.3).

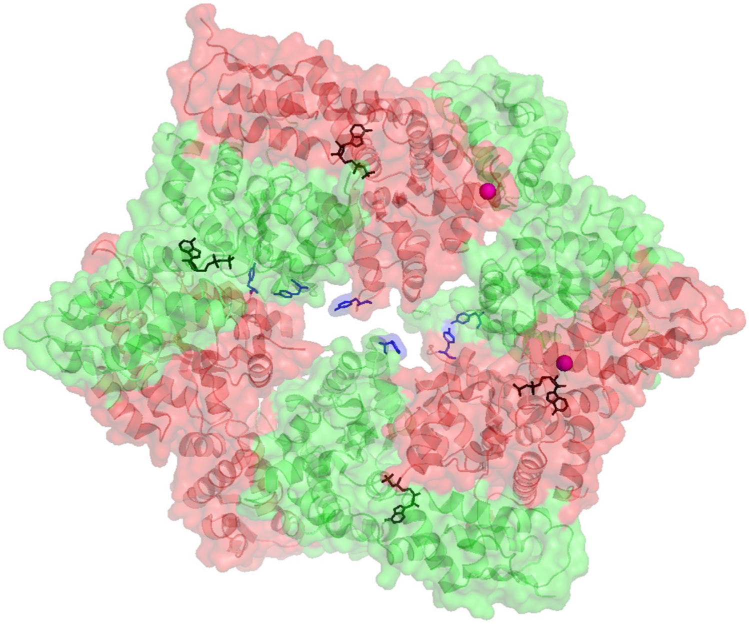


Figure 1.2.2: Asymmetric Structure of the ClpX Homo-Hexamer
Alternating ClpX subunits are shown in red and green, four ADP molecules are shown in black sticks, and two magnesium ions are shown in magenta. Tyrosine Y153 residues are highlighted in blue. It has been shown that Y153 is part of the substrate-binding loop and it has been proposed that Y153 acts as a "steric paddle" to push the unfolded polypeptide chain through the central pore of ClpX (Martin et al., 2009). PDB code 3HWS.

Any review of the ASCE superfamily, no matter how brief, would be incomplete without mentioning the F1 ATP-synthase and the replicative helicase Mcm2-7. F1 is a heteromeric hexamer composed of alternating α and β subunits and a central shaft (γ subunit) which can rotate inside the channel of the F1 ring (Abrahams et al., 1994). In nature, a proton gradient drives the rotation of the γ subunit, which in turn drives the synthesis of ATP within the three catalytic pockets at the interface of α and β subunits (Boyer et al., 1997). It has been shown that F1 can also operate in reverse, i.e. the hydrolysis of ATP within the catalytic pockets can power the rotation of the central shaft (Itoh et al., 2004). A series of very elegant single molecule imaging experiments revealed that the operation of F1 subunits is coordinated in a strict sequential rotary fashion (Adachi et al., 2007). F1 was the first ring ATPase to have its mechano-chemical cycle dissected in great detail (Adachi et al., 2007). Initially it was thought that the inter-subunit coordination in F1 was dictated by the orientation of the γ domain, however ultra-fast AFM experiments have recently revealed that the F1 ring maintains its coordination even in the absence of the central shaft (Uchihashi et al., 2011), suggesting that the subunit firing scheme is somehow hard-wired in the structure of the motor. Unlike F1 which only has two types of subunits, the helicase Mcm2-7 is a heteromeric ring with six unique subunits that assemble in a well-defined order (Costa et al., 2011). Although very similar to each other in sequence and structure, the Mcm subunits are functionally distinct (Bochman et al., 2008). Mcm2-7 requires ancillary factors for activation, which hindered the study of in-vitro helicase activity until recently (Ilves et al., 2010). Mcm2-7 is a very attractive system for studying inter-subunit coordination in single-molecule experiments precisely because its subunits can be arranged in a specific order without resorting to covalent linking.

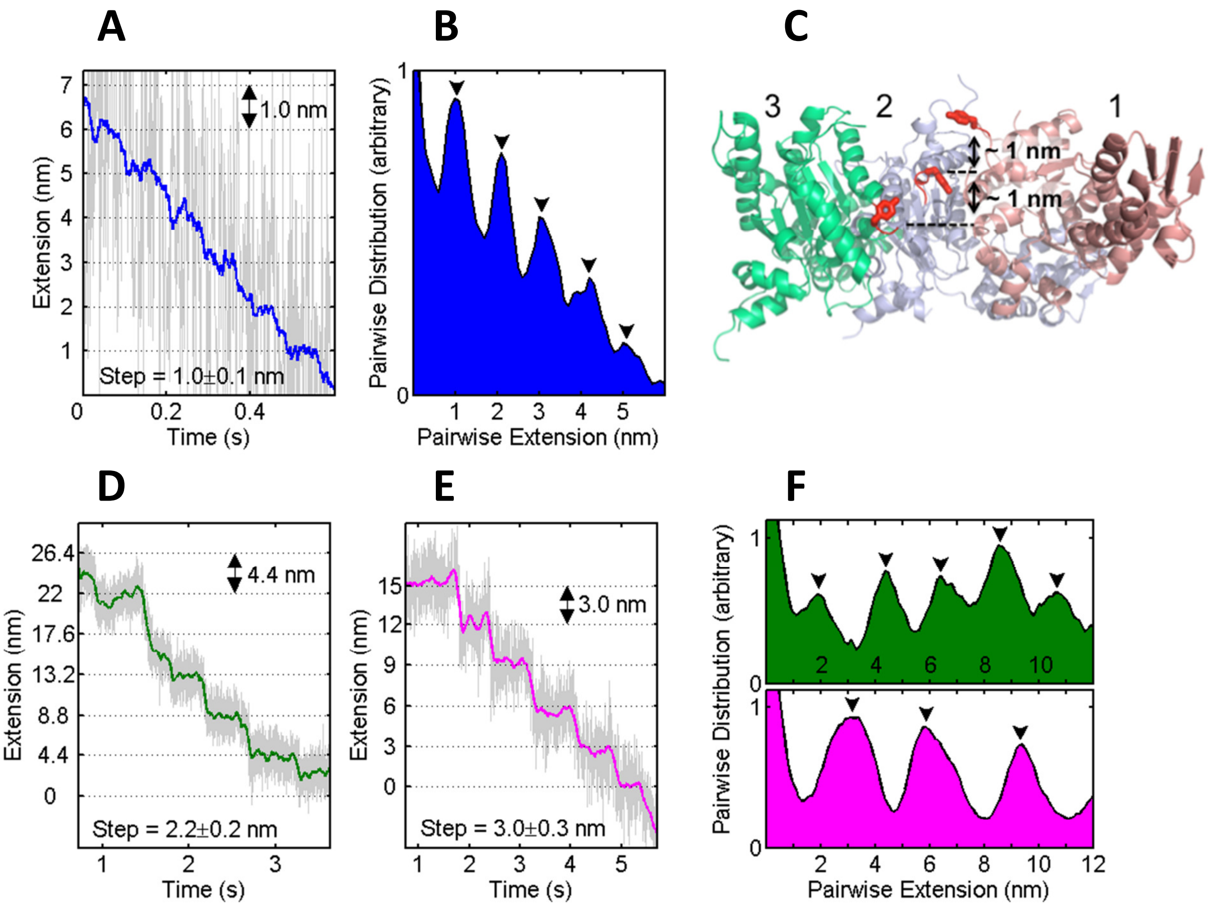


Figure 1.2.3: Translocation Steps Suggest Limited Coordination among ClpX Subunits
(A) Fragment of a ClpXP translocation region at ~10 pN displaying stepping periodicity of 1.0 nm. Raw data downsampled to 500 Hz are shown in gray, and data boxcar filtered to 20 Hz shown in blue. (B) The pairwise distance distribution for the trace in (A) shows a 1 nm periodicity (black arrows). (C) Side view of the nucleotide-bound ClpX hexamer (Glynn et al., 2009) with three subunits removed to allow visualization of the pore 1 loops (red) with the critical Tyr153 (stick representation) shows distinct staggering relative to the pore axis. The loop of the nucleotide-free subunit 1 is close to the top of the pore, whereas the loops of the nucleotide-bound subunits 2 and 3 are in an intermediate and bottom position, respectively. The distances between Tyr153 in each of these staggered positions are ~1 nm. (D and E) Fragment of a ClpXP translocation region at ~14 pN exhibiting clear steps. (F) The pairwise distance distributions for (D) and (E) indicate ~2-nm and ~3-nm periodicity.
Reprinted from Maillard et al., 2011 with permission from Elsevier © 2011.

1.3 High Resolution Optical Tweezers

The idea that a tightly focused beam of light can be used to trap microscopic particles was first proposed by Ashkin in 1970, but it wasn't until 1980's that optical traps could be built and used to levitate colloidal particles in liquids (Ashkin et al., 1986). What followed was a rapid proliferation of optical trapping in many fields of science and engineering, including biophysics (Ashkin, 2000). At first optical tweezers were used to investigate the elastic properties of DNA (Bustamante et al., 2003), and study the motion of molecular motors in vitro (Block et al., 1990; Kuo and Sheetz, 1993; Svoboda and Block, 1994). The use of optical trapping in biophysics quickly expanded after improvements in electronics and laser technology in the late 1990's. For a while the technique was limited by various sources of noise and drift (Neuman et al., 2004, Moffitt et al., 2008). The dual-trap optical tweezers that we currently operate (Figure 1.3.1) were designed to isolate the environmental noise that would otherwise couple into the system via cover-glass surface or micro-pipette attachments (Moffitt et al., 2006). This section provides an overview of the dual-trap optical tweezers instrument (Figures 1.3.1 and 1.3.2).

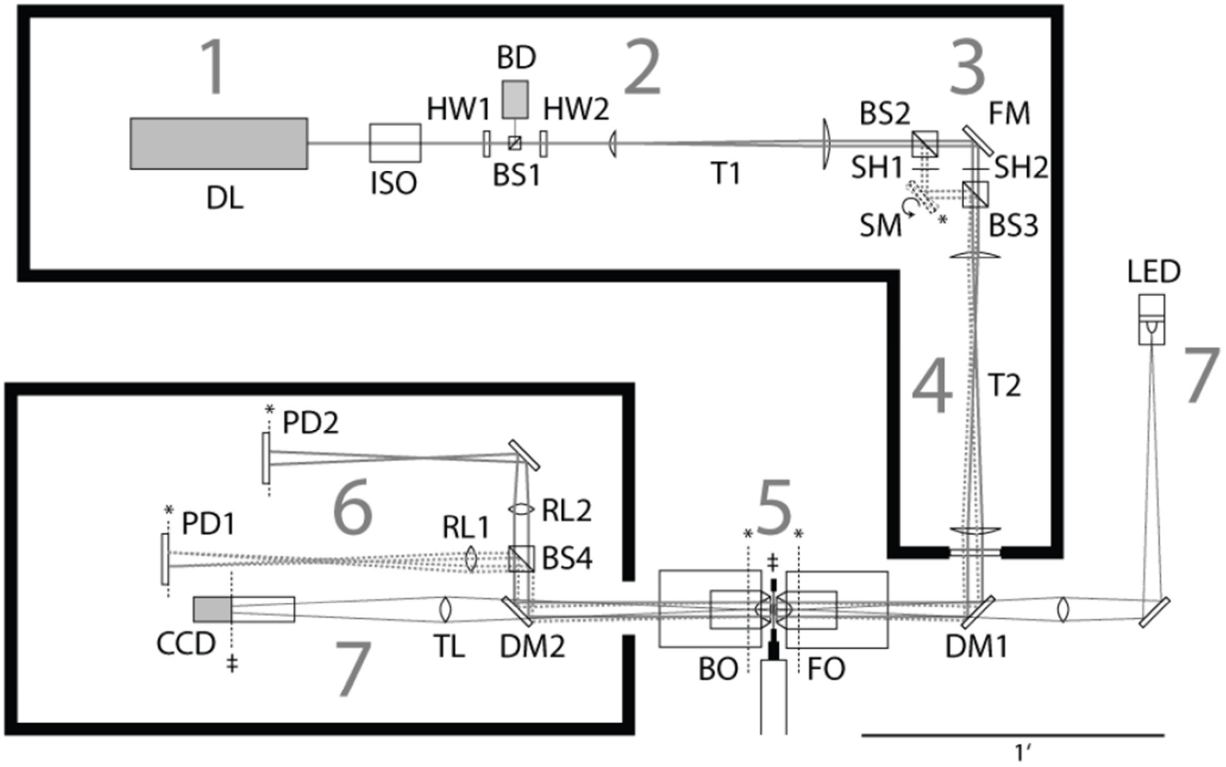


Figure 1.3.1: Diagram of the High Resolution Dual Trap Optical Tweezers

The instrument employs a1064-nm diode-pumped, solid-state, $YdVO_4$ laser (DL). The instrument layout is broken down in 7 sections which are discussed in detail in Chapter 1.3. Thick black lines denote the position of the Plexiglas enclosures that protect the components from stray air flow. Diagram courtesy of Jeff Moffitt.

Figure 1.3.1 illustrates the design of the dual trap optical tweezers consisting of seven stages. Stage 1 contains a diode-pumped 1064 nm solid state laser (DL) which outputs a TEM00 mode beam with a Gaussian profile, and an optical isolator (ISO) which effectively acts as a unidirectional light valve, preventing reflected light from destabilizing the laser feedback or causing damage to the laser head.

Stage 2 provides absolute and relative trap power control as well an initial expansion of the laser beam. The first motorized half-wave plate (HW1) is used to rotate the polarization axis of the laser beam before it reaches the polarizing beam-splitter (BS1), which reflects vertically polarized light into the beam dump (BD) and allows horizontally polarized light to pass through. The polarization axis of the light emerging from BS1 is then rotated by the second motorized half-wave plate (HW2). As a result, HW1 controls the total power of the light that will be used to form the optical traps, whereas HW2 controls the distribution of the total power between the two traps. The first beam-expanding telescope T1 consists of two plano-convex lenses with focal lengths of 38 mm and 175 mm and expands the laser beam-waist to ~4 mm. Note that the beam waist is the distance from the center of the beam to where the beam intensity drops to $1/e^2$ the maximum intensity.

Stage 3 provides trap steering control and expands the laser beam size to its final width. A polarized beam-splitter (BS2) reflects vertically polarized light to a steerable mirror (SM), and allows horizontally polarized light to reach a fixed mirror, FM. Light reflecting off the fixed and steerable mirrors is merged into a single beam by another polarizing beam-splitter (BS3). The vertically and horizontally polarized light beams will form the steerable and fixed traps respectively. The motion of the steerable trap is controlled by deflecting the steerable mirror SM. Electro-mechanical shutters SH1 and SH2 are used to block the light beams and therefore turn off the respective optical trap.

Stage 4 consists of telescope T2 containing two plano-convex lenses of focal lengths 100 mm and 200 mm. This telescope serves three purposes: (1) to expand the beam waist from ~4 mm to ~8 mm, in order to overfill the 8-mm aperture of the trapping objective (FO), (2) to de-magnify by 2x the angular deflection imparted by the steerable mirror (SM) on the steerable beam (dotted lines), and (3) to optically conjugate the plane of the steerable mirror with the back focal plane of the trapping objective (FO) (conjugated planes are denotes by stars). Two planes are optically conjugated if an angular deflection of a ray at the first plane results only in an angular deflection and no translational deflection of the same ray at the second plane.

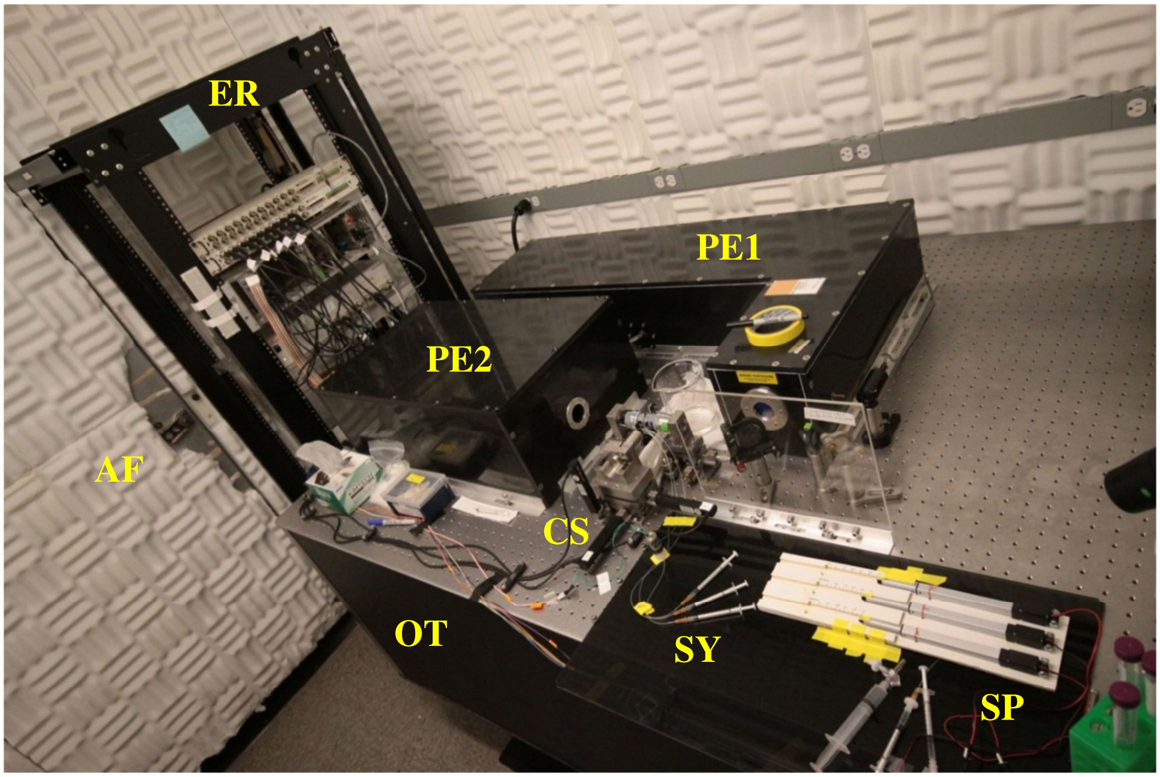


Figure 1.3.2: Overview of the Dual Trap Optical Tweezers Instrument Room
AF – acoustic foam, ER – electronics rack, OT – optical table, CS – chamber stage equipped with three linear motors for chamber travel, SY – syringes with beads and buffers, SP – home-made syringe pump, PE1 and PE2 – Plexiglas enclosures covering the optical elements.

Stage 5 is the portion of the instrument where trapping and manipulation takes place. The front objective (FO) forms the steerable and fixed traps by focusing collimated light emerging from telescope T2. The back objective (BO) collects the highly divergent light from the optical traps and collimates it. Both FO and BO are water immersion, 60x magnification, 1.2 numerical aperture, infinity corrected, differential-interference contrast (DIC) Nikon objectives (CFI Plan Apo VC 60X WI; Nikon, Melville, NY). The use of water immersion objectives allows us to trap deep within the microfluidics chamber (shown in Figure 1.3.1 sandwiched between the two objectives). Note that the back aperture plane of the trapping objective (FO) is optically conjugated to the plane of the steerable mirror via the telescope T2, which ensures that the steerable beam is always centered on the back aperture of the trapping objective and is not clipped asymmetrically upon its deflection by the steerable mirror.

Stage 6, i.e. the detection stage, consists of a polarizing beam-splitter (BS4) that separates the light from the two traps, and two photo-detectors (PD1 and PD2) which are optically conjugated to the back aperture of the back objective BO by two relay lenses (RL1 and RL2) of focal length +150 mm. A micron-sized dielectric bead held in an optical trap scatters light, which interferes with the un-scattered light to produce an interference pattern that is imaged onto the surface of the photo-detector. When the bead is exactly centered on the trap axis, the resulting interference pattern is centered onto the PD. If the bead is deflected from the trap center by Brownian motion or an external force, the centroid of the interference pattern shifts on the surface of the detector (Pralle et al., 1999). For relatively small bead displacements (up to ~100 nm in our case) the detector signal is linearly proportional to the bead displacement, which can be computed using the well-established Brownian calibration technique (Tolić-Nørrelykke et al., 2004).

Stage 7 provides real-time video of the trapped beads and their surroundings, which enables the instrument user to catch individual beads and navigate around the chamber. The LED provides blue light which is focused onto the back aperture of the trapping objective (FO) to provide uniform Kohler illumination of the sample plane. The dichroic mirror DM1 reflects IR light while allowing visible light to pass through, thus merging the laser and LED light before they enter the trapping stage. A second dichroic mirror (DM2) reflects IR light toward the detection stage, while allowing visible light to pass through a tube lens (TL) to a video camera (CCD). As a result, the plane containing the trapped beads inside the microfluidics chamber is imaged onto the surface of the CCD (both denoted by ǂ in Figure 1.3.1).

The alignment, calibration and operation of dual-trap optical tweezers are described in exhaustive detail in Jeff Moffitt's PhD dissertation (Moffitt, 2009) as well as several articles (Moffitt et al., 2006, Bustamante et al., 2008, Tolić-Nørrelykke et al., 2004).

Chapter 2
Data Analysis

2.1 Data Filtering

The raw data in our single-molecule experiments is recorded at 2.5 kHz bandwidth, has a low signal-to noise ratio (Figure 2.1.1, light gray), and its analysis is memory intensive. Usually a simple downsampling filter or a sliding boxcar filter is applied to the raw data prior to any analysis. The downsampling filter simply reduces the bandwidth of a dataset by an integer factor N. For example a 5x downsampling filter averages every 5 points of the raw data into a single filtered point, effectively reducing the size of the data set five-fold. Figure 2.1.1 illustrates downsampling filters ranging from 5x to 100x. Downsampling is the default filter for Phi29 single-molecule data, and is a simple way to improve the signal to noise ratio and to reduce dataset size prior to memory-intensive operations such as step-finding, pause detection, translocation velocity calculation, etc.

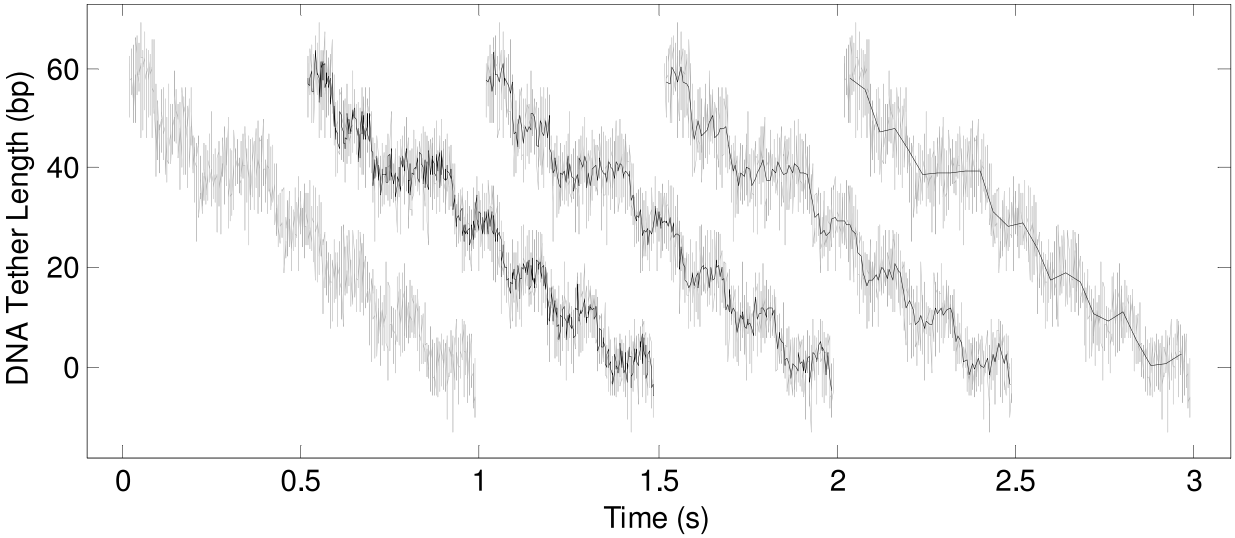


Figure 2.1.1 Downsampling Filters Applied to a High Resolution Phi29 Trace
Raw 2.5 kHz data shown in light gray. 5x, 10x, 20x, and 100x downsampled data shown in black from left to right.

The sliding boxcar filter operates on an N-point window of data, computes and records their average, then shifts the window by one point and repeats the procedure (Figure 2.1.2). The filtered data generated this way is correlated across the width of the N-point window. Note that unlike the downsampling filter, boxcar filtering does not reduce the number of data points. We use the sliding boxcar filter to condition data for computing pairwise distance distributions, and to generate visually appealing plots. It is worth remembering that boxcar filtered data is correlated across N points, which can lead to artifacts with some analysis techniques.

Filtering should be done cautiously, with special considerations for data properties and subsequent analysis. For instance under-filtering the data can lead to excessive computation time when finding steps with the Schwartz Information Criterion algorithm (Kalafut and Visscher, 2008). On the other hand, over-filtering the data may obscure any dynamics on a time-scale smaller than the filter window. How does one pick the optimal filter window? The answer depends on the type of subsequent analysis. If filtering is applied prior to finding translocation steps, it is crucial to first empirically estimate the average dwell duration (i.e. the time between successive steps) and ensure that there are at least 5-10 filtered data points per dwell.

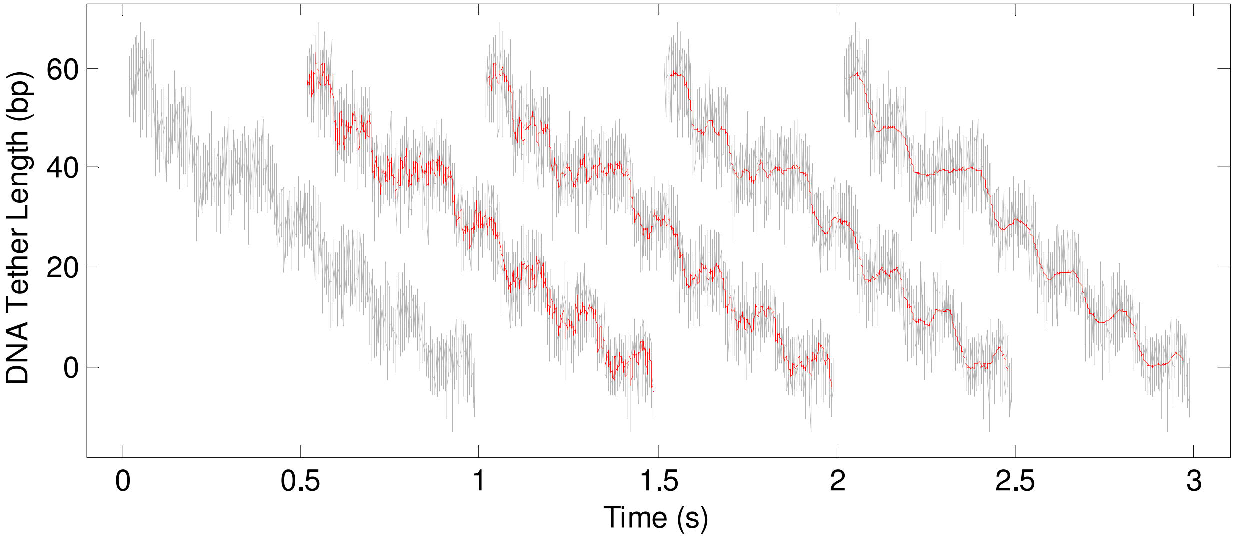


Figure 2.1.2 Sliding Boxcar Filters Applied to a High Resolution Phi29 Trace
Raw 2.5 kHz data shown in light gray. 10-pt, 20-pt, 50-pt, and 100-pt sliding boxcar filtered data is shown in red (from left to right).

Any downsampling or smoothing filter will improve the signal-to-noise ratio at the cost of reduced time-resolution, and filtering is always a compromise between the two. A common beginner's mistake is to pick a filter window size and hard-code it in the analysis pipeline, which often causes unforeseen issues when using very different datasets. It is strongly recommended to make the filter window size an adjustable parameter. In addition, one should verify that the final result (for example pause-free DNA translocation velocity) is insensitive to the filter parameters.

2.2 Pairwise Distance Distribution Analysis

Recent advances in instrumentation have enabled us to study the operation of molecular motors with optical tweezers at sub-nanometer spatial resolution. The Phi29 packaging motor is the poster-child of the basepair-resolution trend in single-molecule studies. High-resolution single-molecule trajectories allow us to follow motor operation one step at a time, and provide us with a wealth of information that was previously out of experimental reach. We can now measure step size distributions, dwell time distributions, identify micro-slips, and follow fast dynamic processes at the sub-nanometer scale. However, this extra information comes at a steep price: expensive and complex instrumentation, very high instrument stability standards, the associated low throughput of "high-resolution data", and increasingly more sophisticated analysis methods. Most of the Phi29 single molecule work now requires a high-resolution approach. The first stage in any high-resolution analysis is the preliminary identification of high-resolution data. The periodogram analysis (Block and Svoboda, 1995), sometimes called the pairwise distance distribution (PWD) analysis (Moffitt et al., 2009; Chistol et al., 2012) is a fast and convenient method to measure the intrinsic periodicity of a molecular motor recording, i.e. the motor step-size. The PWD can be calculated via either a slow brute-force method, or a fast method that relies on the discrete Fourier transform.

The brute-force method is a simple but computationally expensive implementation in which the raw data is first downsampled, then all possible $n(n-1)/2$ point-to-point distances are computed (here n is the number of downsampled points). The pairwise distance distribution is generated by plotting a histogram of the pairwise distances, as shown in the sample Matlab code below.

```
function [Count, Distance] = BruteForcePWD(RawContour,FilterFactor,BinSize)
    FilteredContour   = DownSample(RawContour,FilterFactor);
    PairwiseDistances = []; %initialize the list of point-to-point distances
    for i=1:length(FilteredContour)-1
        %compute distances between ith point and the ones after that
        temp              = FilteredContour((i+1):end)-FilteredContour(i);
        PairwiseDistances = [PairwiseDistances temp];
    end
    PairwiseDistance  = abs(PairwiseDistances); %distance magnitude
    HistogramBins     = (0+BinSize/2):BinSize:max(PairwiseDistances);
    [Count, Distance] = hist(PairwiseDistances,HistogramBins);
end
```

Note that the resulting PWD distribution will depend on the filter bandwidth (Figure 2.2.1) and the histogram bin size. As long as one does not over-filter or under-filter the data, the PWD will display local maxima at distances that are an integer multiple of the fundamental step size.

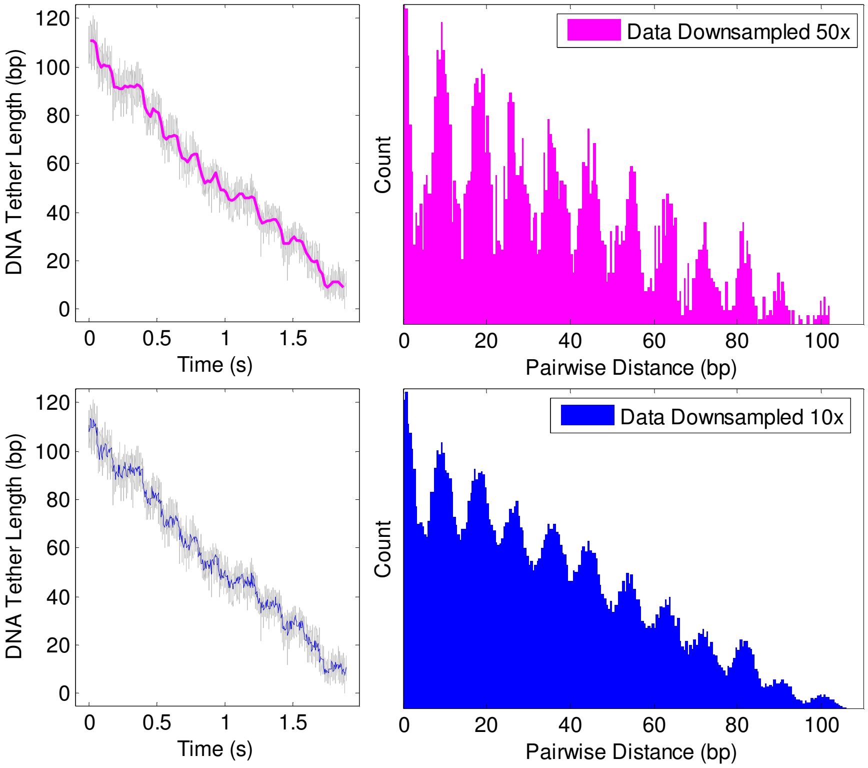


Figure 2.2.1 Computing the Pairwise Distance Distribution via the Brute-Force Method
Pairwise distance distributions were generated from the same Phi29 trace with 50x downsampling (magenta) and 10x downsampling (blue) using the brute-force method and a 0.5-bp histogram bin-size. The trace is shown on the left with 2.5 kHz data in light gray and filtered data in magenta/blue, and the respective PWD is shown on the right.

Although the brute-force PWD method is very simple, it is quite slow, and the computation time scales as n^2, where n is the number of data points. In addition, the number of point pairs drops at longer distances, resulting in a diminishing PWD intensity (Figure 2.2.1). These shortcomings are addressed by a more efficient, Fast Fourier Transform (FFT) based method, which measures the intrinsic spatial periodicity of the signal (Block and Svoboda, 1995). The Matlab code below is a compact implementation of the fast PWD method. Using a sliding boxcar filter is recommended for this method as it does not reduce the number of data points and produces a smooth distribution.

```
function [Amplitude, Distance] = FastFourierPWD(RawContour,FilterFactor,BinSize)
    FilteredContour = BoxcarFilter(RawContour,FilterFactor);
    HistogramBins   = min(FilteredContour):BinSize:max(FilteredContour);
    [n, d]          = hist(FilteredContour,HistogramBins); %bin filtered data
    PowerSpectrum   = abs(fft(n)).^2;
    Inverse         = real(ifft(PowerSpectrum)); %inverse FFT
    Amplitude       = Inverse(1:round(length(Inverse)/2));
    Distance        = (0:(length(Amplitude)-1))*BinSize;
end
```

Figure 2.2.2 illustrates the difference between PWDs computed using the brute-force and the FFT-based methods. Note the jagged appearance of the distribution when a downsampling filter is used.

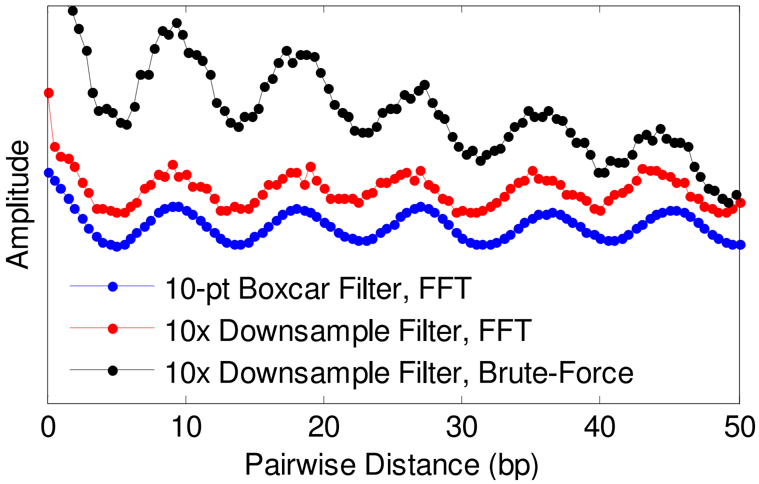


Figure 2.2.2: Comparing Brute-Force and FFT-based methods for Computing PWDs
On the bottom in blue: PWD generated by the FFT-based method using a 10-pt boxcar filter.
In the middle in red: PWD generated by the FFT-based method using a 10x downsampling filter.
On the top in black: PWD generated by the brute-force method using a 10x downsampling filter.
The red and black curves were offset vertically for clarity. PWDs were computed using the same Phi29 trace.

Figure 2.2.3 displays PWDs generated from the same Phi29 trace using the FFT method with boxcar filters of different window size. The PWD contrast, i.e. the ratio between the height of local maxima and local minima, depends on filtering but the peak location is essentially unchanged except when data is over-filtered (50-pt window), which leads to artifacts.

Pairwise distance distribution analysis is very powerful, fast, and convenient for analyzing molecular motor traces, but it has some limitations and its results must be interpreted with caution. (1) PWDs are difficult to interpret when dwell durations vary significantly, as the longer dwells contain a disproportionately high number of points. For example, if there are two very long dwells 30 bp apart, the PWD will display a very large local maximum at a pairwise distance of 30 bp. (2) PWDs are difficult to interpret if the molecular trajectory contains a distribution of step sizes. For example if a trace contains a mix of 10-bp and 7.5-bp steps, the first local maximum will be located at a pairwise distance somewhere between 7.5 and 10 bp, depending on the fraction of steps of a given size. (3) Frequent slipping in a molecular trajectory can have a significant impact on the shape of the PWD, and can lead to the long-range de-correlation of the periodic signal. It is recommended to break up the trace in slip-free regions prior to PWD analysis. (4) Filter type and filter bandwidth have some effect on the shape of the pairwise distribution, and it is recommended to vary filter settings and ensure that the final result is not an artifact of filtering. (5) The FFT-based PWD calculation may cause artifacts when analyzing very short segments of a trace (spanning only a few steps), so the brute-force method is recommended in that case.

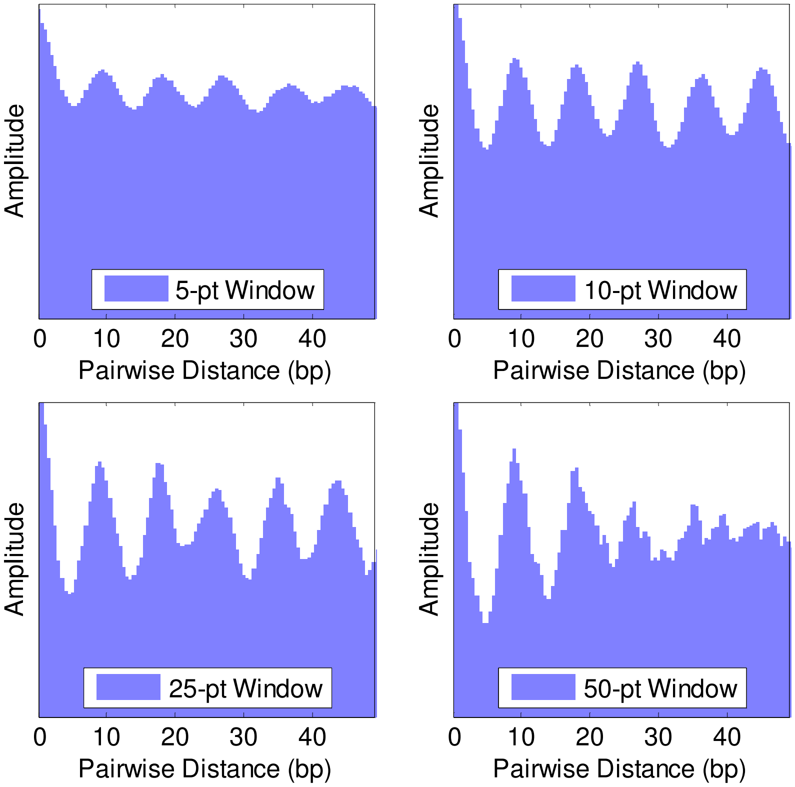


Figure 2.2.3: Illustrating the Effect of Boxcar Filter Window Size on PWDs
PWDs were computed using the same Phi29 trace, a 0.5-bp histogram bin width, and boxcar filters with 5-pt, 10-pt, 25-pt, and 50-pt windows. The first local maximum of all PWDs reveals a 10-bp periodicity corresponding to the 10-bp burst of Phi29. Note that higher-order maxima are affected by over-filtering (50-pt window).

2.3 Kernel Density Estimation

To overcome some of the limitations of the PWD analysis, I adapted the kernel density estimator (Rosenblatt, 1955; Parzen, 1961) for the analysis of high-resolution molecular motor trajectories. One can think of the kernel density as a generalized form of a histogram. In a histogram every single data point is represented as a bar of unit height. Representing a distribution with a histogram has several drawbacks: (1) it depends on the chosen histogram bin size, (2) any information about the uncertainty associated with each data-point is discarded, and (3) the histogram values are discrete and discontinuous. In a kernel density plot, every data point is represented as a Gaussian centered at the value of the data-point with a width (σ) given by the uncertainty of the data-point value.

The kernel density can be used to represent the Phi29 motor residence time on DNA. To construct the kernel density I passed the raw data through a downsampling filter (Figure 2.3.1 left panel). Every filtered point was represented as a Gaussian of unit area (Figure 2.3.1 center panel). The width of each Gaussian was given by the standard error associated with each filtered point. For illustrative purposes, every 30th filtered point and its Gaussian are shown in blue in Figure 2.3.1. The sum of all individual Gaussian contributions yields the total kernel density Figure 2.3.1, right panel). Since each Gaussian is normalized to unit area, every filtered data point contributes equally to the kernel density.

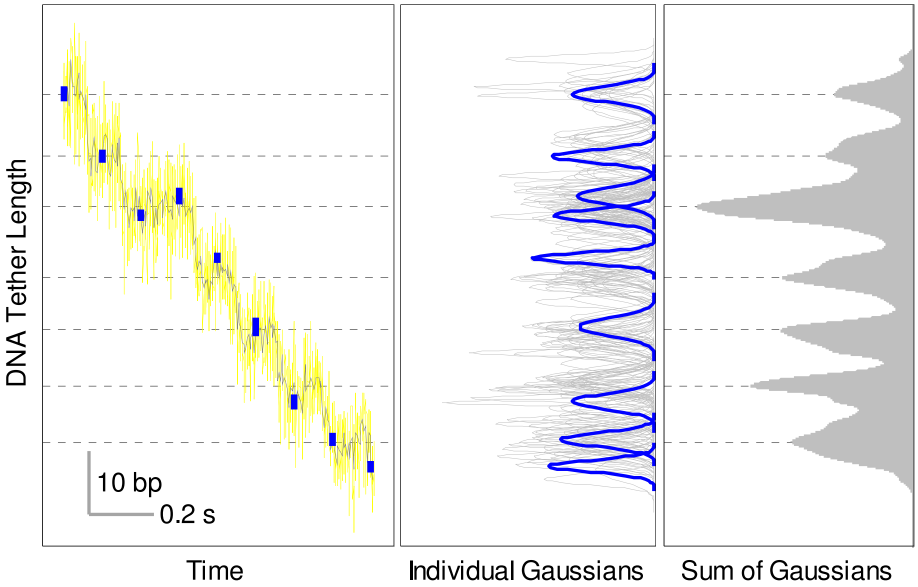


Figure 2.3.1: Constructing the Kernel Density Representation of a High Resolution Phi29 Trace
Left Panel: Fragment of a Phi29 translocation trace, raw data shown in yellow, 10x downsampled data shown in gray. Center Panel: Gaussian representations of each downsampled point are shown in gray. Right Panel: kernel density estimate obtained by summing the individual contributions of filtered data points. Every 30th downsampled point is highlighted in blue in the left and center panels. Dashed lines indicate the location of local maxima in the kernel density estimate.

The kernel density representation of a high resolution Phi29 trace exhibits pronounced local maxima and minima, corresponding to dwells and bursts respectively (Figure 2.3.1, right panel). I sometimes refer to the kernel density representation of a Phi29 trace as the “residence time histogram” because it represents the amount of time the molecular motor spends at a particular location on the DNA.

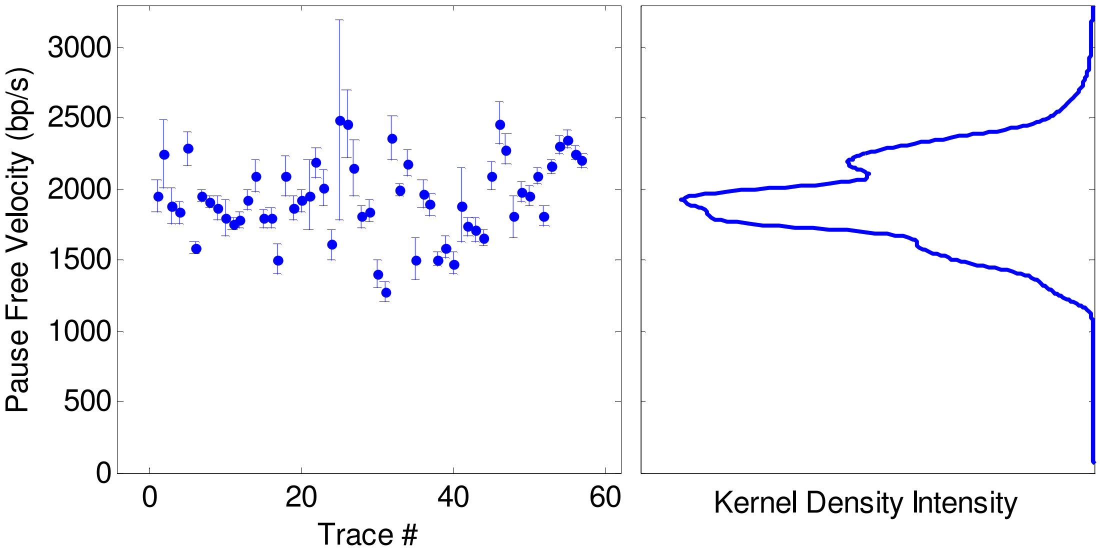


Figure 2.3.2: Kernel Density Estimate of the Mean Velocity Distribution of SpoIIIE
Left Panel: Each data point represents the mean pause-free translocation velocity computed for one molecule of SpoIIIE. Error bars indicate the 95% confidence interval. Right Panel: Mean velocity distribution that was obtained by summing up individual Gaussian contributions for each point in the left panel. Data courtesy of Ninning Liu.

The kernel density can also be used to construct distributions from several individual measurements as illustrated in Fig 2.3.2, which shows the pause-free velocity of SpoIIIE, a very fast dsDNA translocase. Each data-point represents the average translocation velocity for one molecule, and error-bars correspond to the 95% confidence interval. The individual Gaussian contribution of each SpoIIIE molecule was computed and summed up to obtain the kernel density estimate of the mean velocity distribution (Figure 2.3.2, right panel). The final distribution suggests that it may contain fast and slow sub-populations of SpoIIIE molecules.

As any other analysis method, the kernel density analysis has some limitations. (1) The kernel density representation is not informative for traces containing slips, as the motor traverses the same location on DNA again after a slip. It is best to break up the trace into slip-free regions and analyze each region separately. (2) The exact shape of the kernel density depends somewhat on the downsampling filter, although the location of local maxima should remain essentially unchanged. (3) The signature of small-scale or fast motion may be obscured on a kernel density plot, especially in the vicinity of very large local maxima. (4) Over-filtering or under-filtering the data may result in artifacts or poor contrast in the kernel density representation.

2.4 Finding Steps in Noisy Data

To extract the most information out of a high-resolution single-molecule trace, it is necessary to resolve individual motor steps and measure the distribution of dwell durations and step sizes. Dwell time distributions provide clues about the number of rate-limiting transitions taking place during the dwell, whereas step size distributions provide insight about the motor inter-subunit coordination and the nature of the motor-substrate interaction. Several step-finding algorithms have been used in single molecule data analysis, including the Kerssemaker's method (Kerssemakers et al., 2006), the Student t-test algorithm (Carter and Cross, 2005), the Schwartz Information Criterion method (Kalafut and Visscher, 2008), and others (Arunajadai and Cheng, 2013).

I relied mainly on two step-finding algorithms – the Student *t*-test method (Carter and Cross, 2005) and the Schwartz Information Criterion (SIC) method (Kalafut and Visscher, 2008). I prefer the SIC method since it requires no adjustable parameters in its original form and is generally more robust. The original SIC algorithm uses the formula $SIC(j_1, \ldots, j_k) = (k+2)log(n) + nlog\left(\hat{\sigma}^2_{j_1,\ldots,j_k}\right)$, where n is the number of data points, k is the number of steps, and $\hat{\sigma}^2_{j_1,\ldots,j_k}$is the maximum likelihood estimator of variance when k steps are fitted to the data. The SIC step detection is an iterative procedure, starting out with no steps assigned to the data. In the first stage of analysis the algorithm places a hypothetical step at a time-point, effectively breaking up the trace into two hypothetical dwells. The mean value and the data variance is computed for each hypothetical dwell, and are then used to calculate the SIC score for this hypothetical step assignment. This procedure is repeated, placing a hypothetical step at every time point and computing the respective SIC score. The step placement that minimizes the SIC value is accepted and the procedure advances to the next stage, where the algorithm attempts to find the best location for another step. SIC values are computed for all the possible locations of the new step, and the best placement for a new step is one that minimizes the overall SIC score. If adding a new step reduces the SIC score compared to the previous round, the new step is accepted and the procedure continues, until adding another step fails to further decrease the SIC score.

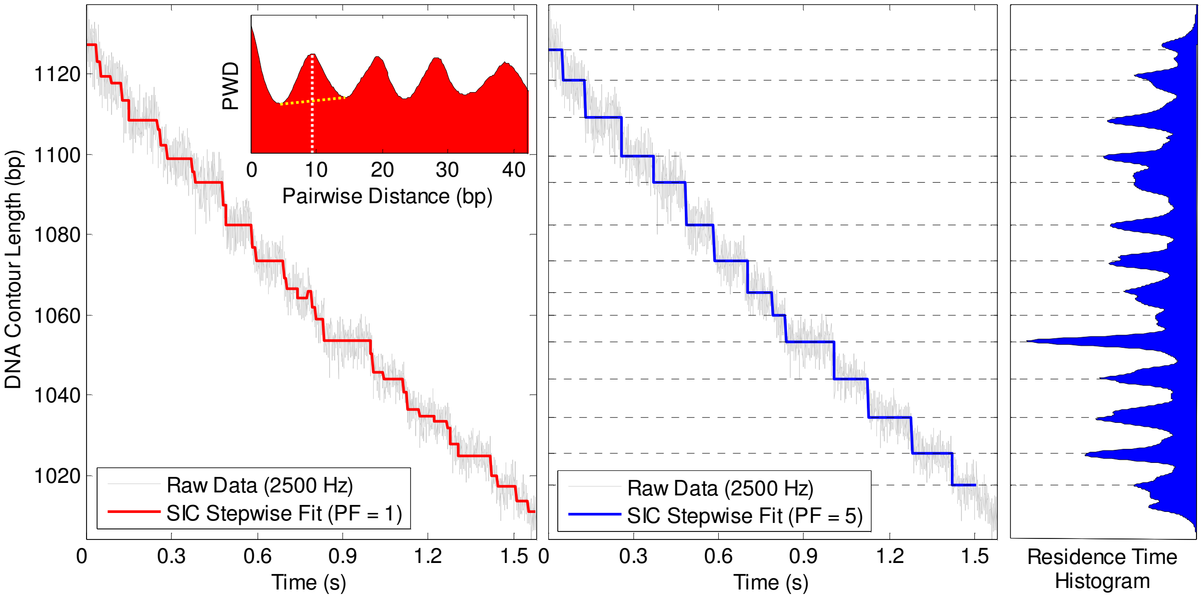


Figure 2.4.1: Stepwise Fits Generated by the Original and the Modified SIC Step-Finding Algorithms
Left Panel: The stepwise fit (red line) generated by the standard SIC method (penalty factor PF = 1) for a high-resolution Phi29 translocation trace (raw data in light gray). Inset: the pairwise distance distribution for this trajectory reveals a 10-bp signal periodicity. The dashed yellow line highlights the local minima around the first local maximum, indicated by the white dashed line. Center Panel: Blue line indicates the stepwise fit generated by the modified SIC method using penalty factor PF = 5 for the same raw data (light gray). Right Panel: The residence time histogram (i.e. the kernel density estimate) generated by downsampling the raw data by 10x. Dashed horizontal lines denote the position of local maxima in the residence time histogram.

While the original SIC algorithm form works well with simulated data that contains white noise, it tends to over-fit real-world experimental data, which contains colored noise. To prevent over-fitting, I introducing an additional penalty factor (PF) as follows: $SIC(j_1, \ldots, j_k) =$

$PF(k+2)log(n)+nlog\left(\hat{\sigma}^2_{j_1,\ldots,j_k}\right)$. The appropriate penalty factor can be estimated by analyzing a portion of the packaging trace where the motor is stalled (usually at the beginning of the trace). I found that a penalty factor of 3-5 worked best (Figure 2.4.1). The authors of the method recommend analyzing unfiltered data (2.5 kHz in our case), which results in very slow execution times in Matlab, even with optimized code. To speed-up the step-finding process I downsampled the data by 10-50x, depending on the conditions of the experiment. Kernel density analysis was used to validate the candidate dwells identified by the modified SIC algorithm. I selected packaging traces that displayed clear stepping based on their PWD (Figure 2.4.1, left panel, red inset). I defined the PWD contrast as the height of the local maximum divided by the height of the baseline set by the nearest local minima. Traces that exhibited a PWD contrast of 1.2 or higher usually contained very clear stepping and were used for further dwell-time analysis.

Every step-finding method has unique strengths and limitations, but no step-finding algorithm can be trusted blindly. Generally, the dwells and bursts identified by a step-finding algorithm must be treated as "candidates" and later verified and validated by an independent method, either a different step-finding algorithm, or a complementary residence time histogram analysis (Figure 2.4.1). It goes without saying that any sophisticated analysis method is powerless if the experimental data is of poor quality.

2.5 Estimating Confidence Intervals via Bootstrapping

When analyzing experimental data, one is always faced with estimating the confidence interval for some measured quantity, for example the mean dwell duration. It is often tempting to use the standard deviation or standard error as a measure of the confidence interval for the mean. The standard error of the mean is sometimes misleading, and in some cases even impossible to compute, for example if one is interested in the value of $n_{min}=\langle\tau\rangle^2/(\langle\tau^2\rangle-\langle\tau\rangle^2)$, for a dwell time distribution $\{\tau_i\}$. Bootstrapping provides a simple and robust means of estimating the confidence interval of any quantity that can be calculated from a set of experimental measurements (Efron, 1979). Let's imagine that we have a set of 1224 experimentally measured dwell durations (Figure 2.5.1, left panel) and we want to compute the mean dwell duration with 95% confidence. To this end, we draw at random 1224 dwell duration values from the actual data set (drawing the same value twice is allowed), effectively generating a simulated data set. We then compute and record the quantity of interest – mean dwell duration – for the simulated distribution. The sample code below illustrates how bootstrapping can be done in Matlab.

```
function MeanDurationList = BootstrapMeanDwellDuration(DwellDurations,BootstrapN)
    %index of all the values in the DwellDurations
    Indexes = 1:1:length(DwellDurations);
    %initialize the list of boostrapped Mean Duration values
    MeanDurationList = nan(1,BootstrapN);
    for i=1:BootstrapN
        %draw with replacement, drawing the same value twice is ok
        CurrInd             = randsample(Indexes,length(DwellDurations),1);
        MeanDurationList(i) = mean(DwellDurations(CurrInd));
    end
end
```

After repeating the "resample and compute" procedure several times, we can plot the distribution of mean dwell durations computed from the simulated data sets (Figure 2.5.1, right panel). The average dwell duration of the original distribution is given by the most likely value in the bootstrap histogram (dashed line) and the upper and lower confidence bounds can be computed by including 95% of the data around the mean (horizontal dotted line). If the original data set were smaller, for example only ~100 dwell durations, the bootstrap histogram would have been much wider, and therefore the confidence interval would have been much broader. For convenience, the kernel density estimator can be used to generate a smooth probability distribution for the bootstrapped values (Figure 2.5.1, right panel, dark blue curve). Note that this method enables the user to calculate an arbitrary confidence interval (68%, 95%, etc.) and takes into account the size of the original data set by drawing exactly 1224 dwell durations in each simulation round.

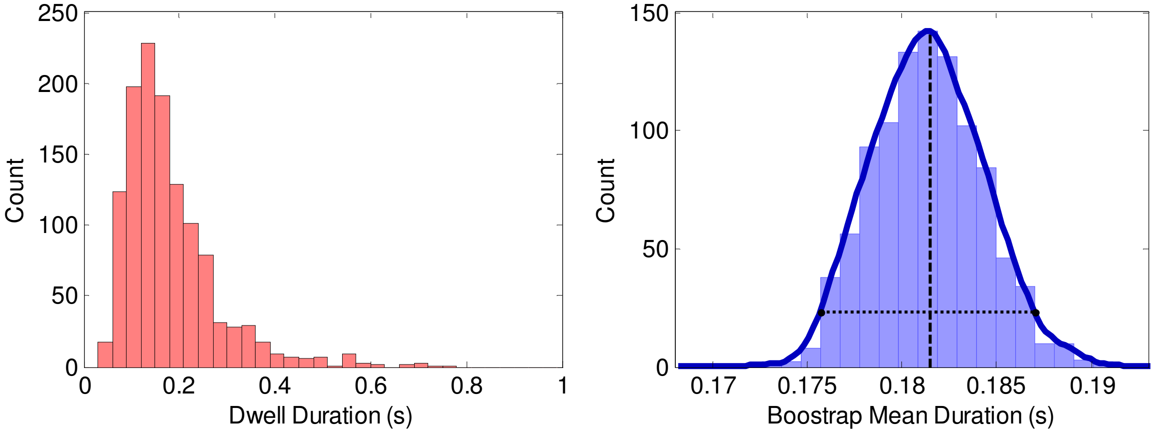


Figure 2.5.1 Using Bootstrapping to Compute the 95% Confidence Interval for the Mean Dwell Duration Left Panel: The dwell duration distribution at saturating [ATP] and 60-70% capsid filling (N = 1224). Right Panel: The distribution of mean dwell durations from 1000 rounds of bootstrapping using the original dwell duration distribution (light blue). The kernel density representation of the distribution is shown in dark blue. The dashed vertical line indicates the most likely mean dwell duration, and the dotted horizontal line denotes the 95% confidence interval. $\langle \tau \rangle = 0.182$ s, 95% Confidence interval: 0.175 s - 0.187 s.

Estimating confidence intervals via bootstrapping can answer the question "If I were to collect twice as much data, how much will it improve my error-bars?" It should be pointed out that the bootstrap method may fail when operating on very small data sets that under-sample the underlying distribution.

2.6 Conclusions and Data Analysis Recommendations

This chapter is only a brief overview of the most common methods I used for analyzing single molecule data, as it is impossible to summarize everything I learned during the past six years. Below are some lessons I have learned the hard way. (1) Never trust a single analysis method, especially if it yields the results you are looking for. (2) Always double-check for loopholes, failures in logic, silent bugs in the code. (3) Verify that your results do not change qualitatively when using different parameters or different analysis procedure. (4) Be aware of the strengths and limitations of a particular method, and use them to your advantage. (5) Do not attempt to write the "holy grail" of all analysis procedures, i.e. the analysis code that will do everything, as it is bound to fail. Always write small functions that are easy to read and will perform one procedure really well. If you need a slightly different procedure for a different purpose, copy and modify the original code. (6) Comment and organize your code religiously, as you may need to go back and modify it at a later time. Writing good code is slower, but will save you time in the future. (7) Organize your code in folders, with each folder dedicated to a particular task, for example "Brownian Calibration", "SIC Step Finding", "Velocity

Calculation”, etc. If you keep all your analysis functions in one folder, over the years your library will grow to unmanageable sizes. (8) Always balance the cost of doing some sophisticated analysis manually and writing new code that would automate that procedure. If you anticipate that you might have to go back and re-do the manual analysis, it is strongly recommended to invest some time to write an automated method. (9) When writing code, favor simplicity and clarity over sophistication and complexity. Simple code may run a bit slower, but it will be easy for you and others to edit later, and you are more likely to make mistakes or overlook loopholes in highly complex code. Remember: computers are always getting faster, and you can always leave your computer to crunch some data while you go out to get lunch or coffee. (10) Iterate on your analysis. Do not attempt to write the perfect analysis script in one stroke. Use the 80/20 principle – i.e. get 80% of the functionality working in 20% of the time, then repeat. Sometimes 80% is good enough for an initial assessment. As you learn more about you’re the data, you can improve the code to satisfy 90% of your needs. (11) Automatically generate and save plots that illustrate the major steps of your analysis. Review these plots after the analysis is complete and look for issues, inconsistencies, failures, and anything out of the ordinary. (12) Although automation is usually preferred, some analysis requires manual supervision, but can be partially automated by creating Graphical User Interface modules. (13) Get to know your data! It is not uncommon to use one or two “representative” traces to test a new analysis script then apply that analysis to all the traces. More often than not, “representative” traces are in fact outliers, i.e. ideal-case examples of experimental data. It is strongly recommended to familiarize yourself with all the data, and look for patterns, exceptions, common issues, potential pitfalls, etc. (14) Accept other people’s harsh criticism, don’t take it personally, and thank them for it. It is much better to receive harsh criticism from your lab mates than from the reviewers of your manuscript, or even worse – the readers of your paper.

Chapter 3
Dissecting the Mechanochemical Cycle of the Phi29 Homomeric Ring ATPase

Material for this chapter has been drawn from Chistol and Liu et al., 2012 with permission from Elsevier © 2012.

3.1 Background

Multi-subunit, ring-shaped NTPases (Nucleoside Triphosphatases) are a group of enzymes that drive translocation or rotation of their substrates, such as nucleic acids and polypeptides, by coupling nucleotide binding and hydrolysis to conformational changes in the substrate-binding motif (Bustamante et al., 2004). A large number of these oligomeric ring NTPases belong to the ASCE [Additional Strand Conserved E (glutamate)] superfamily, and are involved in a multitude of cellular tasks, including DNA replication, transcription regulation, protein degradation, cargo transport, chromosome segregation, and viral genome packaging (Erzberger and Berger, 2006; Iyer et al., 2004).

A central question about the operating principles of ring NTPases is how these motors couple and coordinate chemical events to processive mechanical movement (Lyubimov et al., 2011; Singleton et al., 2007). It is difficult to answer this question using traditional biochemical methods which necessarily involve ensemble averaging. Although crystal structures containing the motor, the substrate, and the bound nucleotides often possess sufficient resolution to relate chemical states to mechanical states of the motor, they do not provide dynamic information about the overall mechanochemical cycle. The advent of single-molecule techniques has made it possible to follow the trajectories of individual molecules in real time. By directly imaging the angular movement of a heteromeric rotary motor, F_1-ATPase, Kinosita and colleagues achieved the most comprehensive mechanochemical characterization of an ASCE ring NTPase to date (Adachi et al., 2007). However, a similar level of description has not yet been accomplished for any homomeric nucleic acid or polypeptide translocase, which comprises a major fraction of ASCE ring NTPases. This is largely due to the lack of robust and sensitive assays to monitor nanometer-scale substrate translocation. We have recently developed high-resolution optical tweezers that enable us to observe the discrete translocation steps of the DNA packaging motor of the *Bacillus subtilis* bacteriophage Phi29 (Moffitt et al., 2009). Thus this double-stranded DNA (dsDNA) translocase is an ideal model system to study the mechanochemical coupling and coordination mechanism of homomeric ring ATPases.

The functional core of the Phi29 packaging motor is a homo-pentameric ATPase, gene product 16 (gp16) (Morais et al., 2008), that belongs to the HerA/FtsK clade of the ASCE superfamily (Burroughs et al., 2007). The biochemical and biophysical properties of gp16 have been studied extensively by bulk and single-molecule experiments (Casjens, 2011; Hetherington et al., 2012; Rao and Feiss, 2008). Previously we reported that the Phi29 packaging motor translocates DNA by cycling through two phases: a stationary or “dwell” phase during which ATPs are loaded to the ring, and a translocation or “burst” phase during which 10 base pairs (bp) of DNA are packaged in four 2.5-bp steps (Figure 3.1.1) (Moffitt et al., 2009). Each 2.5-bp translocation step is powered by the release of inorganic phosphate (Pi) from one subunit (Figure 3.1.1D) (Chemla et al., 2005), indicating that only four of the five ring subunits are involved in DNA translocation in every 10-bp cycle. We have also shown that the motor makes a specific electrostatic contact with a pair of adjacent backbone phosphates on the 5'-3' DNA strand every

10 bp and that this phosphate contact plays an important regulatory role in the motor's cycle (Aathavan et al., 2009).

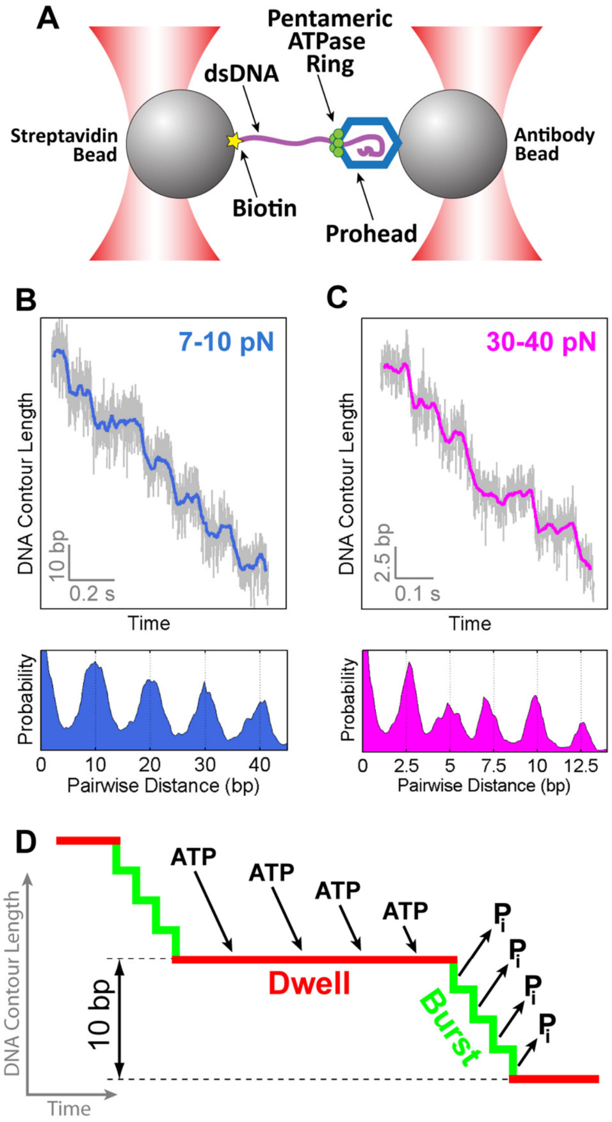


Figure 3.1.1: The Phi29 Motor Packages DNA via a Dwell-Burst Mechanism
(A) Dual-trap optical tweezers used to study the Phi29 packaging motor. (B) A sample packaging trace collected at an external force of 7-10 pN (top) and the corresponding pairwise distance distribution (PWD) (bottom). 2500 Hz (light gray) raw data were filtered and decimated to 250 Hz (blue). In this force regime, DNA is translocated in nearly-instantaneous 10-bp increments. (C) A sample packaging trace collected at an external force of 30-40 pN (top) and its corresponding PWD (bottom). In this force regime the 10-bp bursts break up into four 2.5-bp steps. (D) Diagram of the dwell-burst cycle of the Phi29 motor.

These results set the stage to pursue a full description of the mechanochemical coupling in the Phi29 packaging motor. Specifically, we asked: Where in the cycle, burst or dwell, does ATP hydrolysis occur? Where in the cycle does ADP release happen? How do individual ATPase subunits coordinate their catalytic cycles? Does the non-translocating subunit bind and hydrolyze nucleotide each cycle? Does the non-translocating subunit play any role in the motor operation?

To determine how chemical transitions are coupled to mechanical motion and how subunits are coordinated in the Phi29 ATPase, we use various nucleotide analogs to perturb specific chemical transitions of the catalytic cycle. We follow the translocation dynamics of single packaging motors with high-resolution optical tweezers (Figure 3.1.1A). By monitoring the mechanical response of the motor to these chemical perturbations we show that ATP hydrolysis occurs in the burst phase and that ADP release occurs in the dwell phase. Furthermore, we establish that ATP binding and ADP release are interlaced and strictly coordinated during the dwell phase. Moreover, our data indicate that the special non-translocating subunit binds and hydrolyzes ATP at a well-defined stage in the dwell-burst cycle, in order to play an essential regulatory role in the motor's operation. This finding reveals an unprecedented division of labor among homomeric ring ATPases. Finally, we present the first complete mechanochemical model for the operation of a homomeric ring ATPase based on the dynamic information obtained from single-molecule data.

3.2 Determining the Timing of ATP Hydrolysis

To determine where in the dwell-burst cycle ATP hydrolysis occurs and how the individual hydrolysis events are coordinated among subunits, we monitored DNA translocation by Phi29 packaging complexes in the presence of ATPγS, a non-hydrolyzable ATP analog. It was previously shown that ATPγS binds to the motor and induces long pauses (Chemla et al., 2005). We analyzed single-molecule trajectories at base-pair resolution to determine what phase of the cycle (dwell or burst) is affected by ATPγS. Figure 3.2.1A shows individual packaging traces at various ATPγS concentrations and a fixed, saturating ATP concentration. ATPγS-induced events appeared as pauses in DNA translocation that became more frequent with increasing nucleotide analog concentration (Figure 3.2.1A-B). The majority (~75%) of all ATPγS-induced pausing events consisted of a single pause (Figure 3.2.1C). Surprisingly, a significant portion (~25%) of all ATPγS-induced pausing events consisted of two or more consecutive pauses separated by 10 bp (Figure 3.2.1C). We refer to the latter fraction of pausing events as *pause clusters*.

We then asked whether the multiple pauses observed within a pause cluster were caused by the successive binding of multiple ATPγS molecules to the motor. At the highest [ATPγS] sampled (2.0 μM), we observed an average of one ATPγS-induced pausing event for every 170 bp of DNA packaged (Figure 3.2.1B). Therefore the probability of observing two consecutive pauses caused by two independent ATPγS binding events within a cluster should be less than 1/17 (~6%). This analysis indicates that each pause cluster is caused by the binding of a single ATPγS to the motor. This conclusion is reinforced by the following observations: (1) the number of ATPγS-induced pausing events per kilobase of DNA packaged – including both single pauses and pause clusters – increases linearly with [ATPγS] (Figure 3.2.1B), indicating that each pausing event is caused by the binding of a single nucleotide analog; (2) the distribution of pausing event durations is well fit by a single-exponential, indicating that a single stochastic process – most likely the dissociation of a single ATPγS – is responsible for the termination of the pausing event (Figure 3.2.1D); in contrast, the distribution of durations of individual pauses (including single pauses and pauses within a pause cluster) cannot be fit by a single-exponential, further suggesting that the dissociation of ATPγS from the motor determines the lifetime of the entire pausing event but not the duration of individual pauses that make up a cluster; (3) the average duration of pausing events is independent of [ATPγS]; and (4) the fraction of pause clusters among all pausing events is also independent of [ATPγS] (Figure 3.2.1E inset). Taken together, our results support a mechanism in which all pausing events are caused by a single non-hydrolyzable ATP analog binding to the ring. While the analog remains bound, the motor can stochastically take a few 10-bp bursts, resulting in a pause cluster. As will be discussed later, the observation of pause clusters provides insight about the division of labor among subunits and their coordination in the ring.

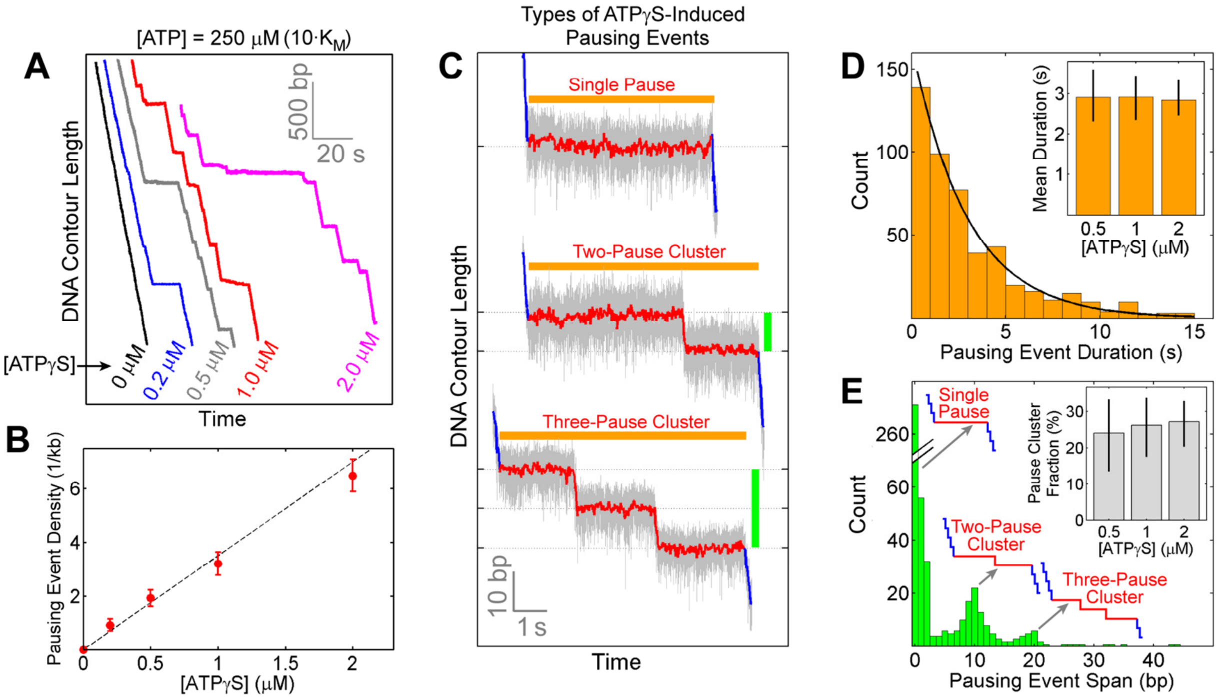


Figure 3.2.1: Non-Hydrolyzable ATP Analog Induces Pausing Events

(A) Packaging traces at saturating [ATP] (250 μM) and various amounts of ATPγS. (B) The density of ATPγS-induced pausing events (number of pausing events per kilobase of DNA translocated) is linearly proportional to [ATPγS]. (C) ATPγS-induced pausing events consist of one or more pauses separated by 10-bp bursts. Pausing events are characterized by their duration (orange bar) and span (the length of DNA translocated during an event, green bar). Pausing events consisting of two or more pauses are referred to as pause clusters. (D) The duration of pausing events pooled from all ATPγS concentrations is well fit by a single exponential ($N = 516$). (Inset) The mean pausing event duration is independent of [ATPγS]. (E) The span of pausing events pooled from all [ATPγS] comprises three distinct groups: single pauses with ~0-bp span, two-pause-clusters with ~10-bp span, and three-pause clusters with ~20-bp span. The finite width of the peaks is due to noise in single-molecule data. (Inset) The fraction of ATPγS-induced pausing events that are pause clusters is independent of [ATPγS].

Having established that all pausing events are caused by the binding of a single ATP analog to the motor, we can now determine where ATP hydrolysis occurs in the dwell-burst cycle. Previous work indicated that Pi release precedes or coincides with each 2.5-bp power stroke during the burst phase (Chemla et al., 2005; Moffitt et al., 2009). Therefore, within a single subunit, ATP hydrolysis should happen after nucleotide binding and before the power stroke. However, this constraint does not uniquely determine the timing or the coordination of hydrolysis with respect to other chemical events in the motor's cycle. ATP hydrolysis can occur during the dwell phase, either interlaced with nucleotide binding (Figure 3.2.2A, H-Scenario 1) or temporally segregated from binding (Figure 3.2.2A, H-Scenario 2). Alternatively, ATP hydrolysis could take place immediately before Pi release during the burst phase (Figure 3.2.2A, H-Scenario 3).

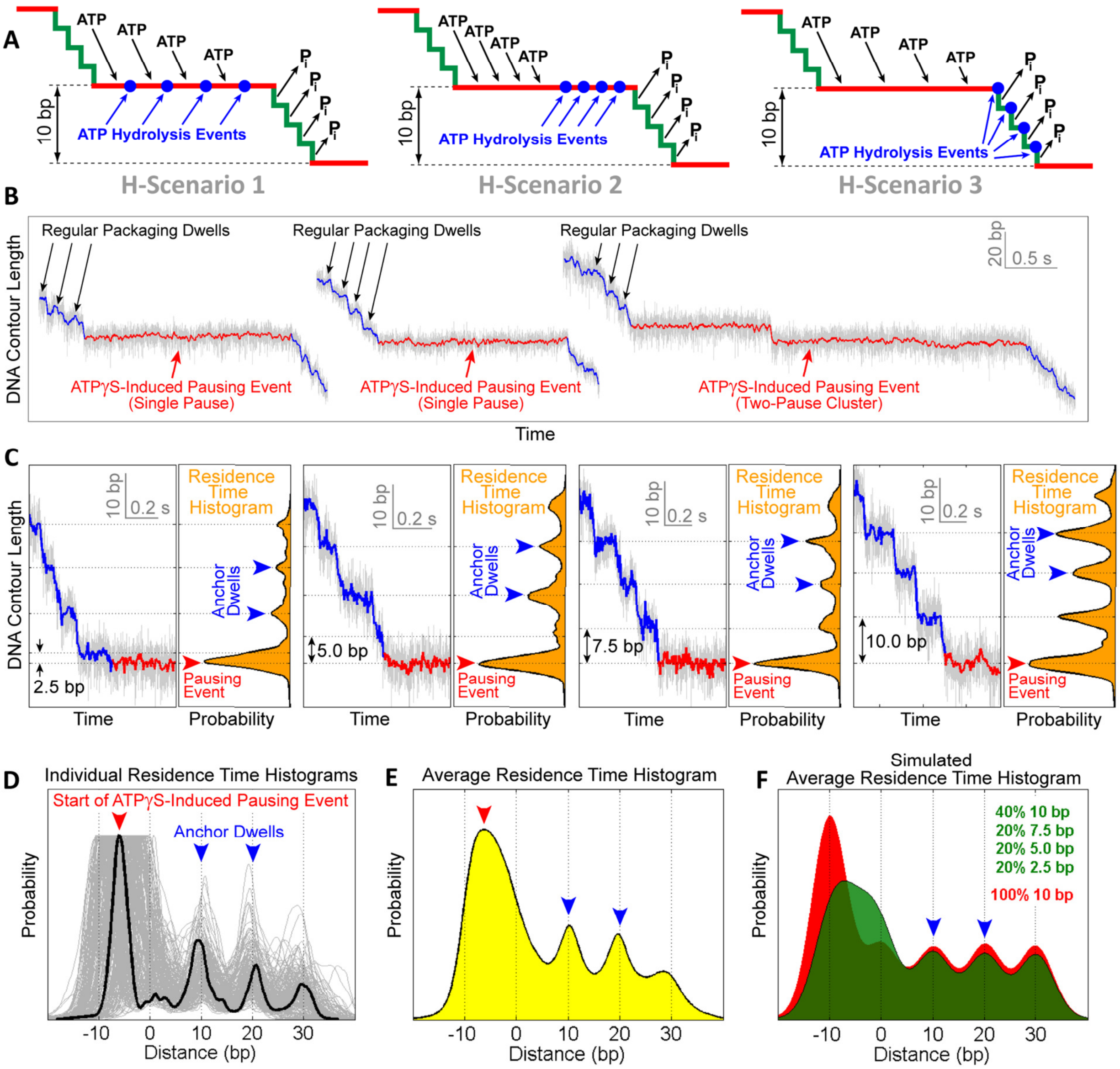


Figure 3.2.2: Determining the Timing of ATP Hydrolysis in the Dwell-Burst Cycle

(A) Possible ATP-hydrolysis scenarios. H-Scenario 1: ATP hydrolysis is interlaced with ATP binding during the dwell. H-Scenario 2: ATP hydrolysis occurs during the dwell after all subunits have bound ATP. H-Scenario 3: ATP hydrolysis is interlaced with Pi release during the burst. (B) Overview of ATPγS-induced pausing events highlighting the difference in duration between regular packaging dwells (black arrows) and pauses (red arrows). (C) Detailed view of four sample packaging traces containing an ATPγS-induced pausing event, from 40-bp upstream of the pausing event to 0.5-sec after the start of the pausing event. Regular packaging is shown in blue and the start of the pausing event is shown in red. The large peak (red arrow) in the residence time histogram corresponds to the start of the pausing event. Two regular dwells (blue arrows) were used as anchors for aligning different residence time histograms. (D) Residence time histograms superimposed and aligned using their two anchor dwells, which are located at 10-bp and 20-bp marks (blue arrows). The 0-bp mark denotes the position of the regular dwell immediately before a pausing event. The burst size before an ATPγS-induced pausing event is given by the distance between the 0-bp mark and the large peak in the histogram (red arrow). (E) The residence time histogram obtained by averaging all histograms from panel D ($N = 209$), using experimental data from all [ATPγS]. (F) The average residence time histogram constructed from simulated data for scenarios where hydrolysis occurs during the dwell (red) or during the burst (green).

Using high-resolution optical tweezers, we can distinguish among these alternative scenarios by determining what mechanical phase of the cycle is interrupted by the non-hydrolyzable analog. If all ATP hydrolysis events occur during the dwell (H-Scenarios 1 and 2), the analog-induced pausing event should always start during a dwell and should always be preceded by a complete 10-bp burst, regardless of which subunit binds the analog (Figure 3.2.3A). In contrast, if hydrolysis occurs during the burst (H-Scenario 3), we expect the motor to pause after it takes zero, one, two, or three 2.5-bp steps depending on the order of the analog-bound subunit in the hydrolysis sequence, resulting in a 10-bp, 2.5-bp, 5.0-bp, or 7.5-bp pre-pause burst, respectively (Figure 3.2.3B). Note that when the motor takes zero 2.5-bp steps before pausing, the pre-pause burst appears as 10 bp in size.

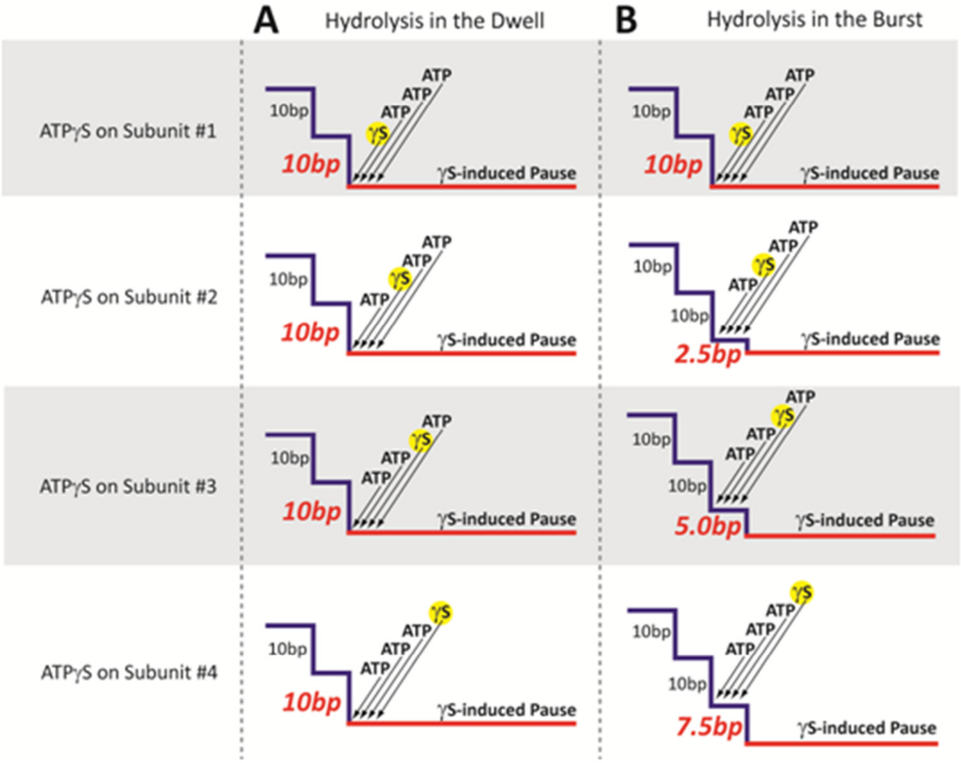


Figure 3.2.3: Predicted Burst Sizes before ATPγS-Induced Pausing Events
(A) If ATP hydrolysis occurs during the dwell, the burst immediately before an ATPγS-induced pausing event should always be 10 bp in size, regardless of the analog position within the ATPase ring. (B) If ATP hydrolysis occurs during the burst, the size of burst immediately before an ATPγS-induced event can be 2.5 bp, 5.0 bp, 7.5 bp, or 10 bp, depending on the position of the analog-bound subunit.

Analysis of individual molecular trajectories revealed a broad distribution containing both complete (10-bp) and incomplete (2.5-bp, 5.0-bp, and 7.5-bp) pre-pause bursts (Figure 3.2.2B-C). To objectively analyze the distribution of pre-pause burst sizes without biasing against smaller bursts, we constructed residence time histograms, which represent the amount of time the motor resides at a particular location along the DNA (Figure 3.2.2C, orange histograms). On these histograms, regular packaging dwells appeared as small peaks, while ATPγS-induced pauses appeared as large local maxima (Figure 3.2.2C, red arrows). Two regular packaging dwells (Figure 3.2.2C, blue arrows) upstream of the pause served as anchor points. The distance between the beginning of a pausing event and its anchor dwells provides a direct measurement of the pre-pause burst size. Furthermore, the anchor dwells enabled us to align all traces and construct an average residence time histogram (Figure 3.2.2D-E). Note that the large peak corresponding to the pause is broad (Figure 3.2.2E, red arrow) and centered away from the 10-bp lattice defined by the anchor points (Figure 3.2.2E, dotted vertical lines). For comparison, Figure 3.2.2F shows simulated average residence time histograms corresponding to different hydrolysis scenarios. If all pre-pause bursts are 10 bp in size (H-Scenarios 1 and 2), the pause peak should fall on the 10-bp lattice (Figure 3.2.2F, red histogram). If instead a significant fraction of pre-pause bursts is 2.5 bp, 5.0 bp, or 7.5 bp in size (H-Scenario 3), the pause peak should be broad and centered off the 10-bp lattice (Figure 3.2.2F, green histogram). The average residence time histogram constructed from experimental data is clearly consistent with the latter case (Figures 3.2.2E-F), indicating that ATP is hydrolyzed during the burst.

Our data suggest that ATPγS is able to temporarily stall the motor after it takes zero, one, two, or three 2.5-bp steps, depending on the position of the analog-bound subunit relative to the first subunit performing hydrolysis. This observation requires that hydrolysis events be strictly coordinated and sequential; otherwise the ATP-bound subunits would always hydrolyze first, resulting in the same burst size prior to each pausing event. To summarize, our results demonstrate that hydrolysis occurs during the burst phase in a strictly coordinated fashion, with each subunit hydrolyzing ATP immediately before taking a 2.5-bp step (Figure 3.2.2A, H-Scenario 3).

3.3 Determining the Coordination of ATP Binding and ADP Release among Ring Subunits

To determine the timing and coordination of ADP release within the dwell-burst cycle, we perturbed DNA packaging using orthovanadate (VO_4^{3-}), a Pi analog that forms stable complexes with ADP, delaying the dissociation of ADP from the binding pocket (Baird et al., 1999; Sharma and Davidson, 2000; Yang and Catalano, 2004). Within a single subunit, ADP release must occur after nucleotide hydrolysis and before a new ATP molecule docks. These requirements allow only three possible ADP release scenarios: all ADP release events are temporally segregated from ATP binding at the beginning of the dwell (Figure 3.3.1A, R-Scenario 1), ADP release is interlaced with ATP binding during the dwell (Figure 3.3.1A, R-Scenario 2), or ADP release is interlaced with Pi release during the burst (Figure 3.3.1A, R-Scenario 3).

Small amounts of Na_3VO_4 (50 nM) were added to an ADP-free packaging buffer, such that vanadate could form complexes only *in situ* with ADP that remained bound to the ATPase after the last round of hydrolysis. We found that ADP-vanadate also induces pausing events (Figure 3.3.1B). However, in contrast to the ATPγS results, the bursts preceding ADP-vanadate-induced pausing events are overwhelmingly 10-bp in size, and the pause peak in the average residence time histogram falls on the 10-bp lattice (Figure 3.3.1C). This result is only consistent with scenarios where ADP release occurs during the dwell phase (Figure 3.3.1A, R-Scenarios 1 and 2).

We have previously reported that all ATPs are loaded to the gp16 ring during the dwell phase and that at any given time during the dwell only one ATPase subunit is capable of binding ATP (Moffitt et al., 2009). As shown above, ADP release also occurs during the dwell phase. The question then arises: how do individual gp16 subunits coordinate their ADP release and ATP binding transitions? Are ADP release and ATP binding temporally segregated (Figure 3.3.1A, R-Scenario 1), or are they interlaced (Figure 3.3.1A, R-Scenario 2)? To discriminate between these two scenarios, we sought to probe the effect of ADP on the motor dynamics. High-resolution packaging traces were collected at saturating [ATP] and various [ADP]. Increasing [ADP] from 0 to 6×[ATP] gradually lengthens the duration of all dwells in a linear fashion (Figure 3.3.1D-E). The effect of ADP is clearly distinct from that of ATPγS or ADP-vanadate, which cause long-lived pausing events once bound. Thus, the dissociation rate of ADP must be sufficiently fast so that binding of one single ADP does not induce distinct pauses. The increase in the dwell duration at high [ADP] is most likely due to multiple rounds of ADP binding and release events from an ATPase subunit before that subunit finally docks ATP. These experiments also show that ADP does not affect the burst size or burst duration (Figure 3.3.1D-F).

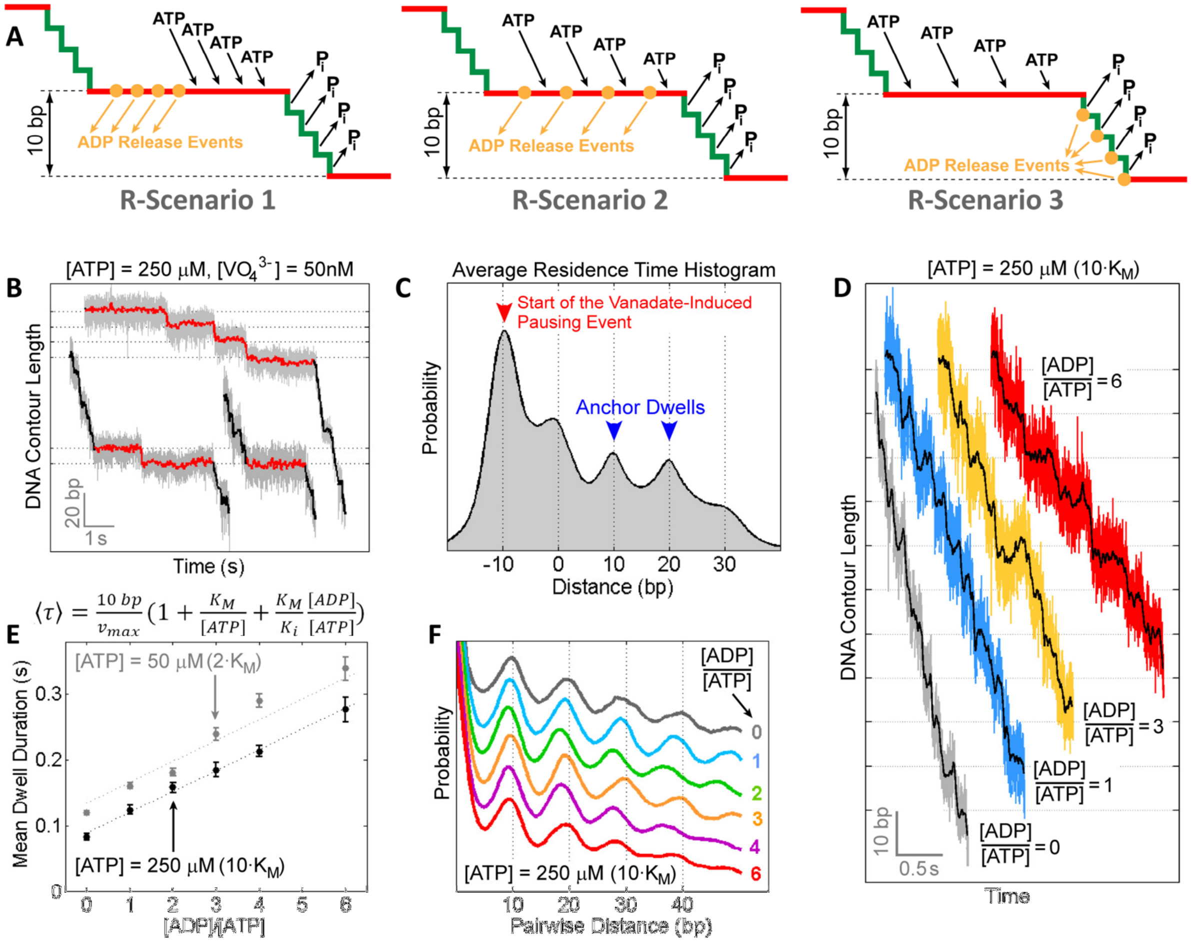


Figure 3.3.1: Determining the Timing and Coordination of ADP Release

(A) Possible ADP-release scenarios. R-Scenario 1: ADP release is temporally segregated from ATP binding during the dwell. R-Scenario 2: ADP release is interlaced with ATP binding during the dwell. R-Scenario 3: ADP release is interlaced with hydrolysis and Pi release during the burst. (B) Sample packaging traces in an ADP-free packaging buffer containing saturating [ATP] and 50 nM of sodium orthovanadate. ADP-vanadate-induced pausing events (red) consist of single pauses or pause clusters containing multiple pauses separated by 10-bp bursts. (C) The average residence time histogram for ADP-vanadate-induced pausing events (*N* = 37). The blue arrows denote the anchor dwells, and the red arrow indicates the start of the pausing event. (D) Sample packaging traces at saturating [ATP] (250 μM) and various ADP concentrations. (E) The mean dwell duration as a function of [ADP] at [ATP] = 250 μM (black) and 50 μM (gray). Data were fit to a competitive inhibition model. (F) PWD plots computed from translocation traces at [ATP] = 250 μM and various [ADP]. The peaks at integer multiples of 10 bp indicate that the burst size is 10 bp, and is not affected by [ADP].

The dependence of the mean dwell duration on [ADP] is well described by a competitive-inhibition model (Figure 3.3.1E), consistent with our previous results (Chemla et al., 2005). If ADP release and ATP binding occur in an alternate fashion (Figure 3.3.1A, R-Scenario 2), ADP and ATP always compete for the same state of the ATPase, making this scenario consistent with competitive inhibition. Moreover, in this case, the mean dwell duration should be linearly dependent on [ADP] since at any given time only one binding site is available. On the contrary, if ADP release and ATP binding were temporally segregated (Figure 3.3.1A, R-Scenario 1), ADP would act as a non-competitive inhibitor, and inhibition by ADP could not be offset with increasing [ATP], as observed (Chemla et al., 2005). Moreover, R-Scenario 1 predicts a non-linear dependence of the mean dwell duration on [ADP], because multiple binding sites would be simultaneously available for ADP binding (Segel, 1975). Therefore, our results and those from previous studies (Chemla et al., 2005; Moffitt et al., 2009) are consistent with a model in which ADP release events are interlaced with ATP binding during the dwell phase (Figure 3.3.1A, R-Scenario 2). An alternative model in which ADP inhibition occurs through an off-pathway state can be shown to be invalid (Chistol and Liu et al., 2012, Supplementary Discussion).

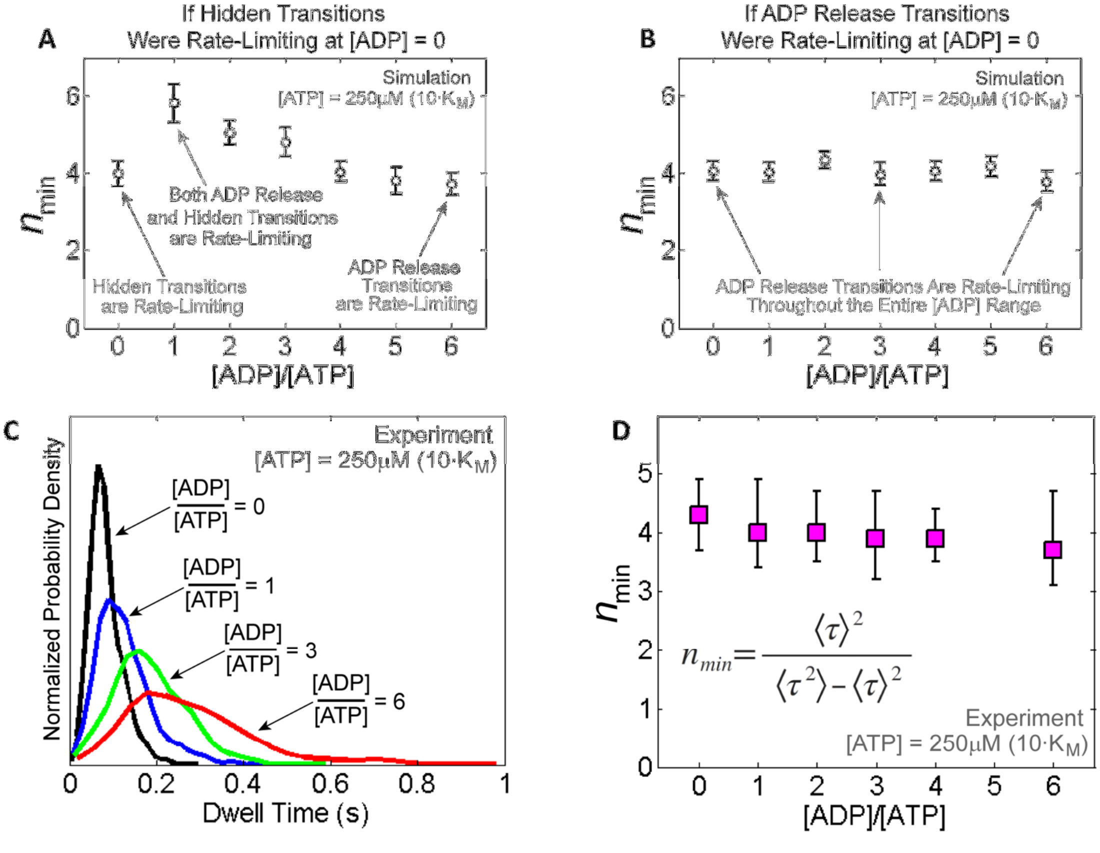


Figure 3.4.1: ADP Release Events Are Rate-Limiting in the Dwell at Saturating [ATP]
(A) Simulated n_{min} dependence on [ADP] if four hidden transitions were rate-limiting during the dwell at saturating [ATP] and zero [ADP]. n_{min} is computed from the dwell time distribution and represents a lower limit for the number of rate-limiting transitions during the dwell. (B) Simulated n_{min} dependence on [ADP] if four ADP release transitions were rate-limiting at saturating [ATP] and zero [ADP]. (C) Experimental dwell time distributions at saturating [ATP] (250 μM) and various [ADP]. (D) n_{min} computed from experimental dwell time distributions at various [ADP]. Error bars represent 95% confidence intervals estimated via bootstrapping.

3.4 ADP Release Events Are Rate-Limiting at Saturating [ATP]

As shown previously (Moffitt et al., 2010; Schnitzer and Block, 1995), the ratio of the squared mean of dwell times to their variance, $n_{min} = \langle\tau\rangle^2/(\langle\tau^2\rangle - \langle\tau\rangle^2)$, constitutes a lower bound for the number of rate-limiting events in a dwell: the larger the number of rate-limiting events the higher the $n_{\min}$ value. Using $n_{\min}$ analysis, we previously established that, at saturating [ATP], the dwell phase contains at least 4 rate-limiting transitions of unknown identity, which are distinct from ATP binding (Moffitt et al., 2009). By determining how $n_{\min}$ varies with [ADP], it is possible to establish the identity of those rate-limiting events.

As shown in Figure 3.3.1D-E, high concentrations of ADP prolong the amount of time the motor spends in the dwell phase, most likely by delaying the replacement of ADP with ATP in the catalytic pockets, and effectively rendering ADP release rate-limiting. Let us now imagine that the dwell phase contains at least four hidden transitions which are neither ATP binding nor ADP release, and that these hidden transitions are rate-limiting at saturating [ATP] and zero [ADP]. As [ADP] is raised, the ADP release transition effectively becomes slower. When the average duration of one ADP release event becomes comparable to the average duration of one hidden transition, the number of rate-limiting events within the dwell phase will necessarily increase ($n_{\min}$ should increase). As [ADP] is raised further, ADP release transitions become slower than the hidden transitions, and eventually ADP release events become the only rate-limiting steps ($n_{\min}$ should decrease). In other words, if processes other than ADP release were rate-limiting at zero [ADP], the identity of the rate-limiting transitions should change with increasing [ADP], and the value of $n_{\min}$ should first rise, reach a peak, and then decrease asymptotically. Monte Carlo simulations of different rate-limiting scenarios corroborate this prediction (Figure 3.4.1A-B).

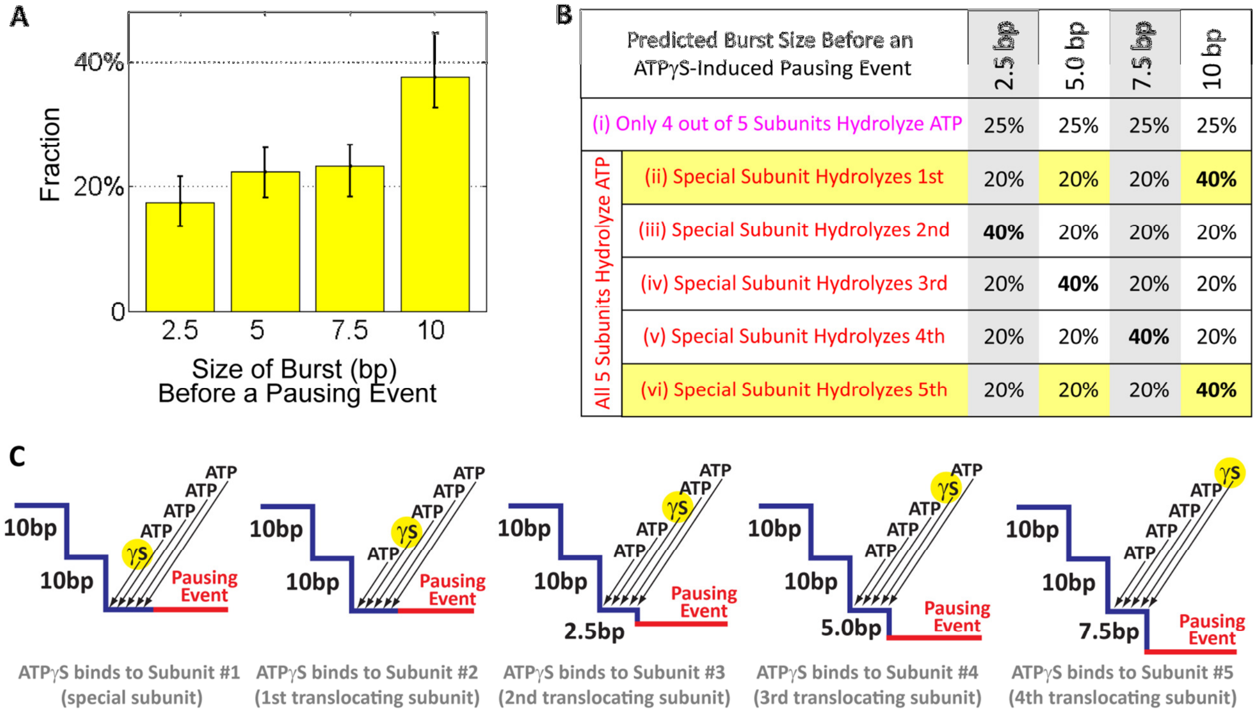
We compiled dwell time distributions for packaging traces in buffers containing saturating [ATP] (250 μM) and [ADP] ranging from 0 to 1500 μM (Figure 3.4.1C). Remarkably, $n_{\min}$ remained constant even though the mean dwell duration increased ~3-fold across the entire range of ADP concentrations tested (Figures 3.4.1D-E). This result can only be explained if, at saturating [ATP], ADP release events are already rate-limiting at zero [ADP] and remain so at all ADP concentrations. Therefore we conclude that under saturating [ATP] conditions at least four ADP release events rate-limit the duration of the dwell.

3.5 Determining the Role of the Special Non-Translocating Subunit

The results presented above reveal that the Phi29 packaging motor operates as a highly coordinated machine in which four of its five subunits release ADP, bind ATP, hydrolyze ATP, and translocate DNA in a precisely timed fashion throughout its mechanochemical cycle. Several questions then arise naturally: What is the role of the non-translocating special subunit? Does it bind ATP? Does it hydrolyze ATP? Three observations provide clues to answer these questions.

(1) Approximately 40% of all bursts preceding ATPγS-induced pausing events are 10-bp in size, whereas the remaining 60% are evenly split among 2.5-bp, 5.0-bp, and 7.5-bp bursts (Figure 3.5.1A). This observation can be most naturally explained by a mechanism in which all five motor subunits, including the special one, bind and hydrolyze nucleotide in every dwell-burst cycle. To demonstrate this point, let us consider the contrary, namely that only the four translocating subunits bind and hydrolyze ATP every dwell-burst cycle. If so the pre-pause burst size should be evenly distributed among 2.5 bp, 5.0 bp, 7.5 bp, and 10 bp (Figure 3.5.1B, Case (i)). In contrast, if instead all five gp16 subunits bind and hydrolyze ATP during every dwell-burst cycle, we should expect a non-uniform pre-pause burst size distribution (Figure 3.5.1B,

Cases (ii)-(vi)). Our observation that the pre-pause burst sizes are not uniformly distributed indicates that all five subunits bind and hydrolyze ATP in each cycle. Moreover, the shape of the pre-pause burst size distribution can be used to determine the temporal order of hydrolysis of the special subunit with respect to the other four translocating subunits. As shown in Figure 3.5.1B-C, the experimentally observed 2:1:1:1 ratio of 10-bp, 7.5-bp, 5.0-bp, and 2.5-bp pre-pause bursts, respectively, requires that the special subunit be either the first or the last to hydrolyze nucleotide.



	Predicted Burst Size Before an ATPγS-Induced Pausing Event	2.5 bp	5.0 bp	7.5 bp	10 bp
	(i) Only 4 out of 5 Subunits Hydrolyze ATP	25%	25%	25%	25%
All 5 Subunits Hydrolyze ATP	(ii) Special Subunit Hydrolyzes 1st	20%	20%	20%	**40%**
	(iii) Special Subunit Hydrolyzes 2nd	**40%**	20%	20%	20%
	(iv) Special Subunit Hydrolyzes 3rd	20%	**40%**	20%	20%
	(v) Special Subunit Hydrolyzes 4th	20%	20%	**40%**	20%
	(vi) Special Subunit Hydrolyzes 5th	20%	20%	20%	**40%**

Figure 3.5.1: The Nucleotide State of the Special Subunit

(A) Distribution of the burst size before an ATPγS-induced pausing event, measured from residence time histograms (Figure 3D). 10-bp bursts occur roughly twice as often as 2.5-bp, 5.0-bp, or 7.5-bp bursts (N = 209). Error bars represent 95% confidence intervals estimated via bootstrapping. (B) Predicted distribution of 2.5-bp, 5.0-bp, 7.5-bp, and 10-bp bursts before a pausing event for various cases of the nucleotide state of the special subunit. Cases (ii) and (vi) are consistent with experimental observations (panel A). (C) Diagrams for the pre-pause burst size distribution given Case (ii) from panel B. In this case, all five subunits, including the special one, bind and hydrolyze nucleotide during every cycle and the special subunit is always the first to undergo hydrolysis. Arrows indicate the temporal order of nucleotide binding events.

(2) The pausing event span histogram displays peaks at 0, 10, and 20 bp (Figure 3.2.1E), indicating that during an ATPγS-induced pause cluster, the motor predominantly takes 10-bp bursts. We obtained similar results when the experiments were repeated with a different non-hydrolyzable ATP analog, AMP-PNP. These observations show that, while one ATPase subunit is occupied by a non-hydrolyzable nucleotide analog, the motor is capable of packaging DNA in nearly-instantaneous 10-bp bursts separated by pauses that are much longer than regular dwells. This phenomenon requires that: (i) during the first pause of a pause cluster the motor resets, and the ATPγS-bound subunit adopts the identity of the special non-translocating subunit, whether or not it fulfilled this role in the previous cycle; and (ii) that while the special subunit remains

bound to ATPγS, the remaining four subunits bind ATP, hydrolyze ATP, and translocate DNA by 2.5 bp each. In this mechanism, the long pauses observed within a pause cluster reflect the response of the motor to the inability of the newly assigned special subunit to hydrolyze the analog nucleotide.

(3) ADP-vanadate can also induce pause clusters that contain 10-bp bursts (Figure 3.3.1B), indicating that, during an ADP-vanadate-induced pause cluster, only one subunit is bound to ADP-vanadate, and that this subunit assumes the special identity. This observation implies that, besides hydrolysis, timely product release by the special subunit is also required for the normal activity of the ring.

In conclusion, our observations support a coordination mechanism in which all five subunits, including the special one, bind and hydrolyze ATP every dwell-burst cycle. We have already shown that hydrolysis events must be strictly coordinated and sequential among the four translocating subunits. The pre-pause burst size distribution now indicates that hydrolysis of the special subunit is also strictly timed with respect to those of the other subunits – it either initiates or concludes the hydrolysis cascade. This mechanism naturally reconciles the discrepancy between the bulk measurement of five ATP molecules consumed for every 10 bp of DNA packaged (Chemla et al., 2005; Guo et al., 1987), and the single-molecule measurement of four 2.5-bp power strokes per dwell-burst cycle. The strict timing and coordination of the catalytic activity of the special subunit leads us to speculate that the hydrolysis by the special subunit is coupled to an essential regulatory process, such as triggering or resetting the motor's mechanochemical cycle.

Note that the role of the special subunit is inferred assuming that nucleotide analogs trap the ATPase at a certain on-pathway state without dramatically altering the kinetic mechanism of the motor, and that ATPγS is equally likely to bind to any one of the five catalytic sites. This is a reasonable assumption, and it has been widely used to dissect the reaction pathway of other NTPases. We acknowledge that alternative models for the operation of the motor can be envisioned; here we have chosen the most parsimonious one to explain the data.

3.6 Symmetry Breaking in a Homomeric Ring ATPase

It has been shown that functional asymmetry exists in heteromeric ring NTPases such as dynein (Carter et al., 2011), the bacterial clamp loader (Johnson and O'Donnell, 2003), and the eukaryotic MCM helicase (Bochman et al., 2008). In these enzymes, a subset of the subunits is responsible for the mechanical task, while the others are either inactive or play regulatory roles. On the other hand, subunits from homomeric ring NTPases are generally believed to undertake the same mechanical task, although they could be in different nucleotide states at a particular stage in the mechanochemical cycle (Lyubimov et al., 2011). Here we show, for the first time, that a homomeric ring ATPase can also display a division of labor: only four of the five gp16 subunits translocate the substrate, while the special one appears to play a critical regulatory function.

While structural differences between subunits in a heteromeric ring NTPase naturally provide the basis for functional asymmetry, it is less clear how such division of labor arises in a homomeric ring motor. It is possible that the functional asymmetry observed in gp16 is caused by the cracking of its ring into a lock-washer shape as has been seen in the crystal structures of Rho, DnaA, and DnaC (Duderstadt et al., 2011; Mott et al., 2008; Skordalakes and Berger, 2003), or by an asymmetric closed planar ring as seen for ClpX (Glynn et al., 2009). However, any asymmetric structure must always involve an underlying symmetry-breaking mechanism.

For example, in the case of F_1-ATPase, the three-fold symmetry of the trimer of $\alpha\beta$ dimers is most likely broken by the successive interaction of each β subunit with the central γ subunit that acts as an external coordination agent of the cycle (Abrahams et al., 1994). In the Phi29 packaging motor, it has been shown that the electrostatic contact made by the motor with a pair of adjacent DNA phosphates every 10 bp plays an important regulatory role in the motor's operation (Aathavan et al., 2009). Our results here suggest that one of the five ATPase subunits also plays a critical regulatory function. Hence, it is sensible to propose that the subunit making the electrostatic contact with DNA is the one that adopts the identity of the special, non-translocating subunit. In other words, the symmetry breaking in the Phi29 motor could arise during every dwell from the crucial electrostatic contact the motor makes every 10 bp with the DNA backbone.

What is the role of this non-translocating subunit? We have inferred that the special subunit either initiates or concludes the hydrolysis cascade (Figure 3.5.1A-B). Therefore, the role of this subunit could be to time the motor's cycle by starting or terminating the burst. Of these two possibilities, we favor the one in which the special subunit is the first to hydrolyze ATP, using this energy to break its strong electrostatic contact with the DNA backbone and to trigger the burst phase. When the special subunit is occupied by a nucleotide analog (ATPγS, AMP-PNP, ADP-vanadate, etc.), the breaking of the motor-DNA contact at the beginning of the burst and the release of the first ADP molecule at the beginning of the dwell must take place spontaneously. Hence the progression of the cycle slows down due to the absence of the timing signal arising from the ATP turnover by the special subunit.

Is the regulatory role always performed by the same subunit, or is that task passed around the ring in subsequent cycles? We have argued that, within a pause cluster, the analog-bound subunit assumes the special role and retains it throughout successive cycles in a cluster. One could argue that a model in which the identity of the special subunit does not change is only valid when one of the motor subunits is bound to ATPγS. However, this model is also favored by symmetry considerations. Since the helical pitch of dsDNA is 10.4 bp, the subunit best positioned to make the crucial regulatory phosphate interaction after a 10-bp burst is the same subunit that made that contact in the previous dwell. We note that this model does not prevent the motor from reassigning the identity of the special subunit under certain circumstances when regular packaging is disrupted, such as after slips or upon analog-induced pausing events, nor does it exclude the small probability of special subunit reassignment during normal packaging due to the intrinsic stochasticity of the system.

3.7 Summary and Conclusions

Based on our previous results and current findings, we can now propose a complete model of the mechanochemical cycle of the Phi29 packaging motor, as depicted in Figure 3.7.1. At the end of the burst phase, all five motor subunits contain ADPs from the previous hydrolysis cycle. At the beginning of the new dwell phase, the special subunit makes the crucial phosphate contact, causing the release of its ADP. The subsequent docking and tight binding of ATP triggers the release of ADP from the next subunit. Coordinated release and binding events then repeat around the ring until all five subunits have replaced ADP with ATP. Next, the special subunit is the first to hydrolyze ATP, breaking its electrostatic contact with the DNA and triggering the sequential hydrolysis, Pi release, and 2.5-bp DNA translocation by the remaining four subunits. The same model is also represented schematically in Figure 3.7.2.

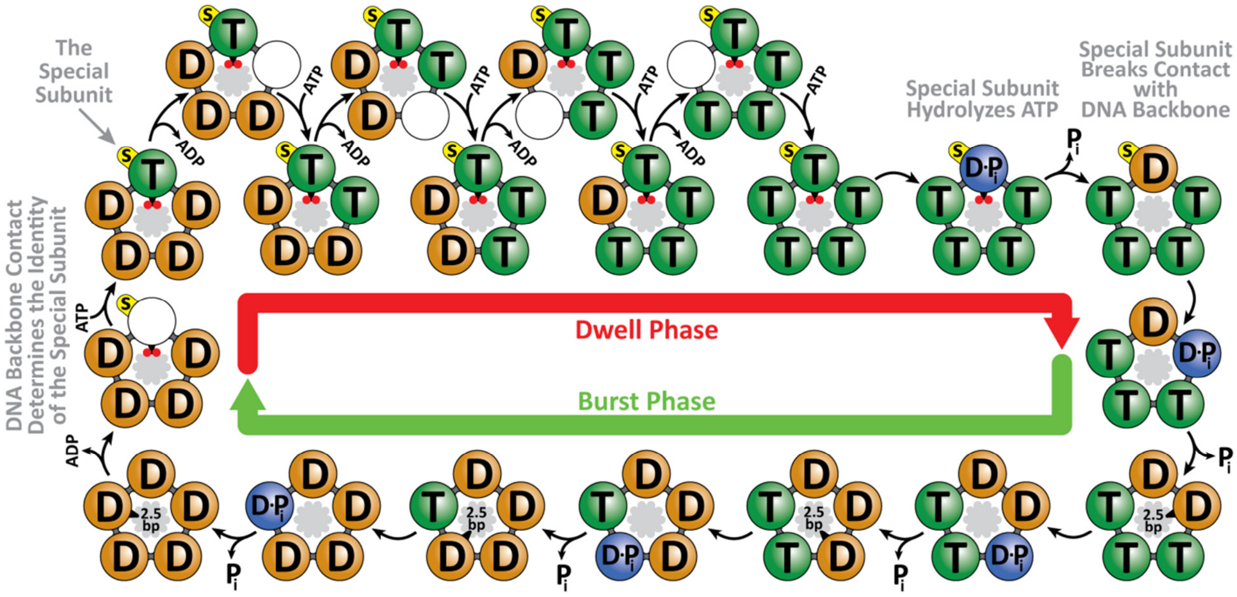


Figure 3.7.1: The Complete Mechanochemical Model of the Phi29 ATPase

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 Pi, 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.

The model above depicts a ring ATPase with multiple levels of coordination among the catalytic cycles of individual subunits. First, ADP release and ATP binding occur in an interlaced and sequential fashion. It is intuitive to see how binding and release events can be kinetically coupled between nearest neighbors in a ring structure: ATP tight binding in one subunit may provide the strain energy to open the binding pocket of the following subunit, facilitating ADP release (Wang and Oster, 1998; Yu et al., 2010). Significantly, our finding that ATP binding and ADP release are interlaced during the dwell explains why we experimentally determined an ATP-binding Hill coefficient of $n = 1$ (Chemla et al., 2005; Moffitt et al., 2009), in apparent contradiction with a highly coordinated motor operation, since only one site is available for ATP docking at any given time during the dwell and tight-binding represents a largely irreversible step (Moffitt et al., 2009).

Second, hydrolysis also occurs sequentially around the ring. In related ring NTPases such as the T7 gp4 helicase and the Phi12 P4 packaging motor, it has been suggested that hydrolysis of the preceding subunit drives the insertion of its arginine finger into the next catalytic site, accelerating the next hydrolysis event (Kainov et al., 2008; Singleton et al., 2000). It is conceivable that the same mechanism is employed by Phi29 gp16, as comparative genomic studies have identified a highly conserved arginine finger (R146) (Burroughs et al., 2007).

Finally, the Pi release events could either be strictly interlaced with hydrolysis or take place stochastically following hydrolysis and preceding ADP release. The great processivity of the motor and intricate coordination of other chemical events make the stochastic Pi release model less likely.

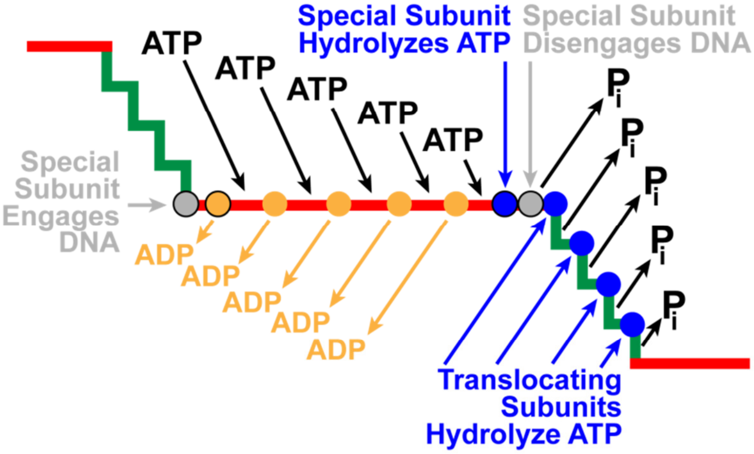


Figure 3.7.2: Summary of the Phi29 Mechano-Chemical Cycle Model
An alternative representation of the model depicted in Figure 3.7.1. The thick line represents an idealized translocation trace: dwells and bursts are shown in red and green respectively. Time flows from left to right. For simplicity only one dwell-burst cycle is illustrated. Note that the ADP molecules released during the dwell phase are left over from the previous mechano-chemical cycle.

The operation of ASCE ring NTPases has been proposed to occur through either a rotary (Adachi et al., 2007; Adelman et al., 2006; Crampton et al., 2006; Enemark and Joshua-Tor, 2006; Mancini et al., 2004; Massey et al., 2006; Singleton et al., 2000; Sun et al., 2008; Thomsen and Berger, 2009), a concerted (Gai et al., 2004), or a stochastic mechanism (Martin et al., 2005), which have been seen as mutually exclusive of each other. In the model we present here for the Phi29 motor, nucleotide binding, hydrolysis, and release events all proceed from one subunit to the next in an ordinal fashion, consistent with a canonical rotary mechanism. However, the level of detail attained in our study also reveals elements that are typically associated with either the concerted or the stochastic mechanism. First, observations of an “all-ATP-bound” and an “all-ADP-bound” state in the crystal structures of the SV40 LTag helicase have been used to support a concerted mechanism for this motor in which all subunits turnover nucleotides simultaneously (Gai et al., 2004). We note that in our model, the motor also adopts an “all-ATP-bound” and an “all-ADP-bound” state at certain points in the cycle, even though nucleotide turnover is sequential. Second, several motors, such as ClpX and MCM helicases, have been shown to tolerate multiple inactive subunits, a phenomenon that has been rationalized by complete or partial stochastic models (Ilves et al., 2010; Martin et al., 2005; Moreau et al., 2007). Here we show that the Phi29 motor, although highly coordinated, exhibits flexibility reminiscent of a stochastic mechanism, bypassing the usual coordination scheme and allowing alternative subunits to translocate DNA when one subunit is temporarily inactivated. A mechanism that allows such flexibility might be particularly useful for overcoming mechanical or chemical obstacles. The combination of coordination and adaptability may be a general design feature shared by ring NTPases, and may serve as a cautionary tale against strict canonical mechanisms which need not be mutually exclusive: motors may possess features from multiple mechanisms to better suit their unique biological functions.

Chapter 4
Dissecting the Throttle Control Mechanisms of a Viral Packaging Motor

4.1 Background

Double-stranded DNA viruses, including most bacteriophages and mammalian herpesviruses, package their genomes into a pre-formed protein capsid during their self-assembly (Gelbart and Knobler, 2009; Sun et al., 2010). DNA is compacted to near-crystalline densities at the end of packaging (Earnshaw and Harrison, 1977). This remarkable mechanical task is performed by a powerful ATP-driven molecular machine known as the packaging motor (Rao and Feiss, 2008; Casjens, 2011). Bacteriophage Phi29, a model system for studying DNA packaging, has a 19.3-kbp genome and its packaging motor is composed of a connector, an RNA scaffold, and a pentameric ring ATPase (Morais, 2012). Several viral packaging motors have been shown to slow down as the capsid fills up with DNA (Smith et al., 2001; Fuller at al., 2007). However, it remains unclear how the packaging velocity is regulated as DNA fills the capsid. Here we used high-resolution optical tweezers to monitor the base-pair-scale packaging dynamics at various levels of capsid filling. We found that the effect of internal filling on motor operation is unexpectedly complex. Not only is the force-sensitive burst phase of the packaging cycle slowed, the force-insensitive dwell phase also slows down, revealing that filling and force are not equivalent in affecting the motor dynamics. This finding prompted us to make an accurate measurement of the opposing force acting from within the capsid, which is critical for understanding the DNA organization inside the capsid and the ejection energetics. Unexpectedly, we found that the motor step size is continuously modulated by the internal filling. Moreover, a class of long-lived pauses increasingly punctuates normal translocation. The dynamics of these pauses are significantly different in a mutant motor lacking a DNA-contacting loop in its connector, which also showed elevated slipping upon the completion of packaging. The picture emerging from these results shows that, instead of being passively stalled by a high opposing force, the motor throttles down by actively changing its behavior in response to DNA encapsidation. This study not only presents a detailed dissection of the complete packaging process that is applicable to other viruses, but also underlines that many new insights could be gained by studying molecular motors in the presence of the full set of regulatory factors.

4.2 External Force and Internal Filling Affect Different Portions of the Motor's Cycle

To probe packaging at different levels of capsid filling, we created dsDNA constructs of various lengths ranging from 6 kbp to 21 kbp (0.3 − 1.1 times the length of the Phi29 genome). The packaging dynamics of individual capsid-motor-DNA complexes were monitored by a dual-trap optical tweezers instrument (Figure 4.2.1A). We first collected packaging traces at various ATP concentrations and at a low external load of 7-10 pN (Figure 4.2.1B). Packaging traces were divided into 1-kbp segments corresponding to ~5% filling level increments. The average packaging velocity at each filling segment follows Michaelis-Menten kinetics, $v = V_{\max}\,[\mathrm{ATP}] / (K_m + [\mathrm{ATP}])$. Both $V_{\max}$ and K_m decrease as DNA fills up the capsid (Figure 4.2.1C). Their ratio, $V_{\max}/K_m$, which describes the effective ATP binding rate ($v \approx (V_{\max}/K_m)\,[\mathrm{ATP}]$, when $[\mathrm{ATP}] << K_m$) (Visscher et al., 1999; Chemla et al., 2005), also drops as the filling degree increases (Figure 4.2.1D). In contrast, an external force only changes $V_{\max}$ and K_m, but does not affect $V_{\max}/K_m$, as reported by our earlier work (Figure 4.2.1D, inset) (Chemla et al., 2005). The dependence of $V_{\max}$, K_m, and $V_{\max}/K_m$ on a certain perturbation (force, filling, etc.) provides information about the location of the perturbation-sensitive transitions within the chemical cycle of the motor (Keller and Bustamante, 2000).

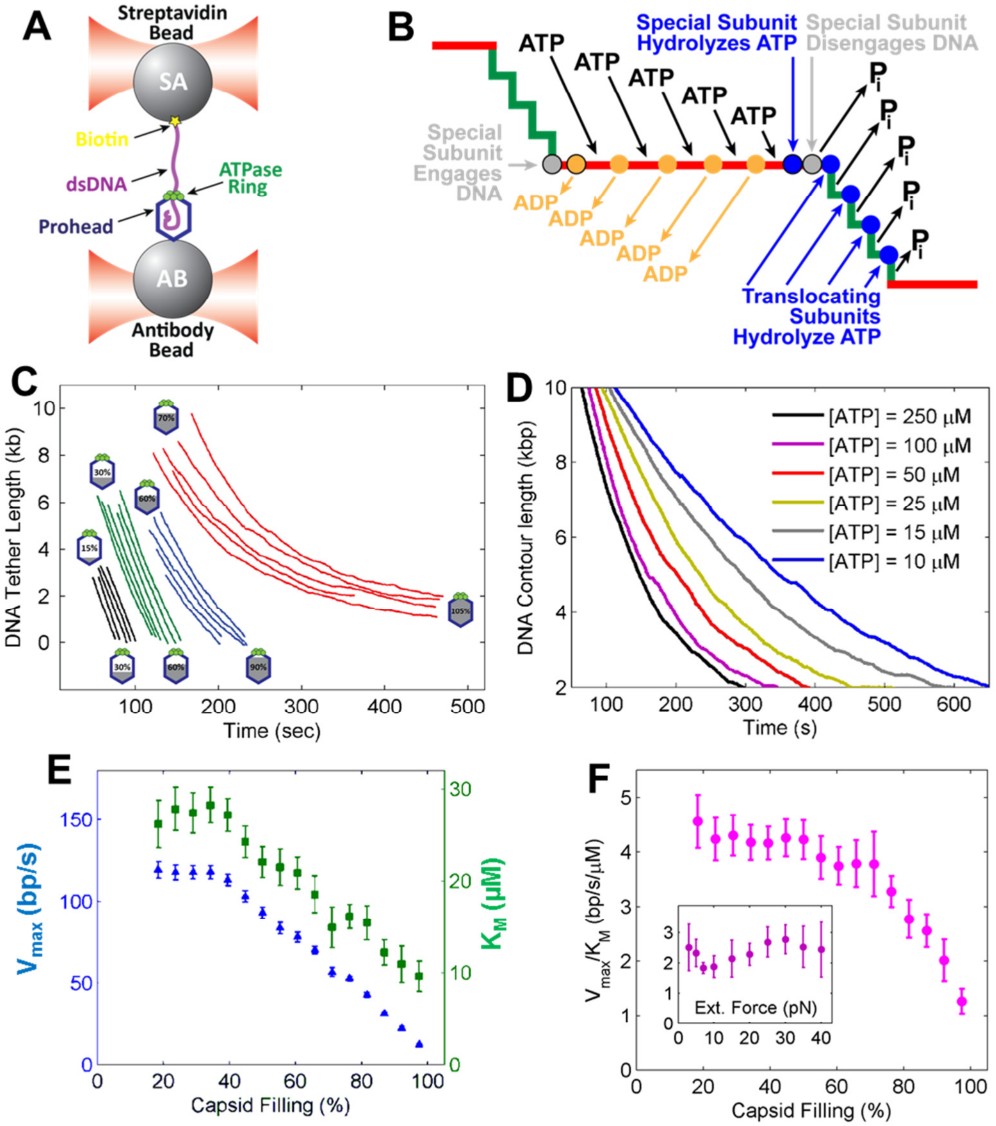


Figure 4.2.1: External Force and Internal Filling Affect Different Portions of the Motor's Cycle
(A) Single-molecule experiment geometry. DNA tether length was varied from 6kb to 21kb. (B) Mechano-chemical model of the Phi29 motor operation at low filling. (C) Sample DNA packaging traces at saturating [ATP] at capsid fillings ranging from 15% to 105%. (D) Sample packaging traces with 21kb DNA tethers at [ATP] ranging from 10µM (sub-K_M) to 250µM (saturating). (E) V_{max} (blue) and K_M (green) dependence on capsid filling. (F) V_{max}/K_M ratio, which is proportional to the effective ATP binding rate, drops as a function of filling. Inset: The V_{max}/K_M ratio is not affected by external force at low filling.

We have previously shown that ATP binds to the subunit in two steps: a reversible "docking" transition followed by an irreversible "tight-binding" transition (Chemla et al., 2005; Moffitt et al., 2009). The observed drop in $V_{\max}/K_m$ as DNA fills the capsid could be caused by a decrease in the ATP docking rate $k_{\text{ATP\_dock}}$, an increase in the ATP unbinding rate $k_{\text{ATP\_undock}}$, or a reduction in the ATP tight-binding rate $k_{\text{ATP\_tight\_bind}}$. However, a change in $k_{\text{ATP\_dock}}$ or $k_{\text{ATP\_undock}}$

should not affect $V_{\max}$; whereas decreasing $k_{\text{ATP\_tight\_bind}}$ causes $V_{\max}$, K_{m}, and $V_{\max}/K_{\mathrm{m}}$ to simultaneously drop, consistent with our results. This analysis suggests that the internal filling slows down the ATP tight-binding step, a transition during which the ATPase commits to hydrolysis by making stronger contacts with ATP (Oster and Wang, 2000).

Next, we examined the effect of ADP on packaging at different filling degrees. Packaging traces were collected at a fixed concentration of ATP (250 μM) and various concentrations of ADP (0 to 1500 μM) (Figure 4.2.2A). At all filling degrees, the inverse of packaging velocity is linearly dependent on the ADP concentration (Figure 4.2.2B), suggesting that the effect of ADP can be described by a competitive-inhibition model (Segel, 1975), $v = \frac{V_{max}[ATP]}{K_m\left(1+\frac{[ADP]}{K_{D,ADP}}\right)+[ATP]}$. Previously we have demonstrated that competitive inhibition by ADP is only consistent with a subunit coordination scheme in which ADP release events are interlaced with ATP binding events (Figure 4.2.1B) (Chistol et al., 2012). Here we show that this type of coordination still applies at high filling. The dissociation constant for ADP, $K_{\mathrm{D,ADP}}$, does not change significantly with DNA filling (Figure 4.2.2C).

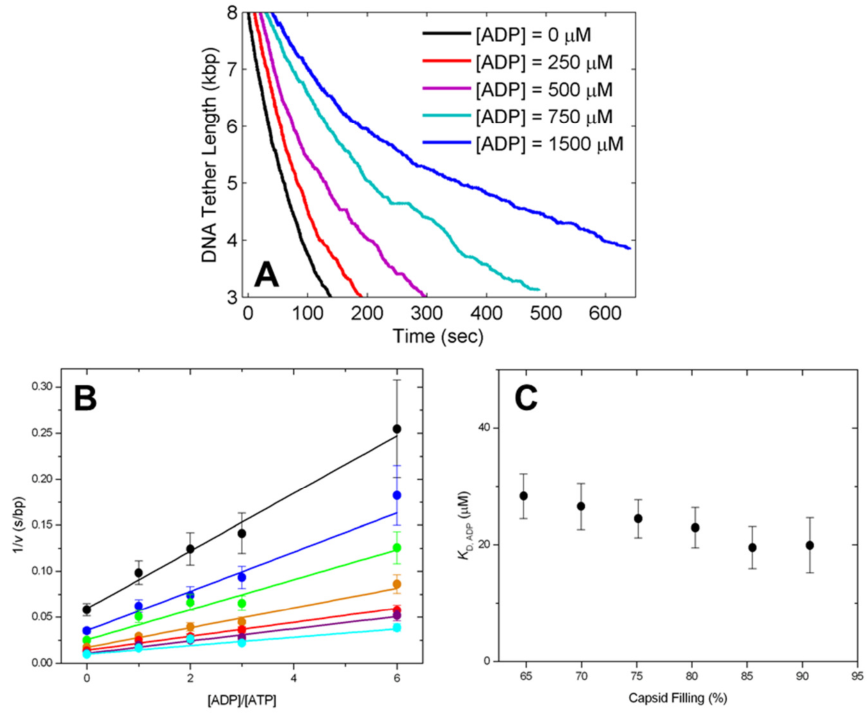


Figure 4.2.2: Internal Filling Does Not Affect the ADP Dissociation Constant

(A) Sample packaging traces with 21 kb DNA tethers at [ATP] = 250 μM and [ADP] ranging from 0 μM to 1500 μM. (B) Plots of inverse velocity versus [ADP]/[ATP] at capsid fillings ranging from 65% (light blue) to 90% (black) reveal a linear dependence consistent with simple competitive inhibition. (C) ADP dissociation constant values versus capsid filling computed from the linear fits in (B).

4.3 Measuring the Magnitude of the Internal Force within the Viral Capsid

As the stiff, highly-charged dsDNA is being compressed into the small capsid, the motor must overcome significant free energy costs caused by electrostatic self-repulsion, bending rigidity, and entropy loss of the confined DNA (Riemer and Bloomfield, 1978), which give rise to an internal resisting force. Estimating the amplitude of the internal force is critical for understanding DNA organization inside the capsid and the energetics of the subsequent ejection process (Molineux and Panja, 2013).

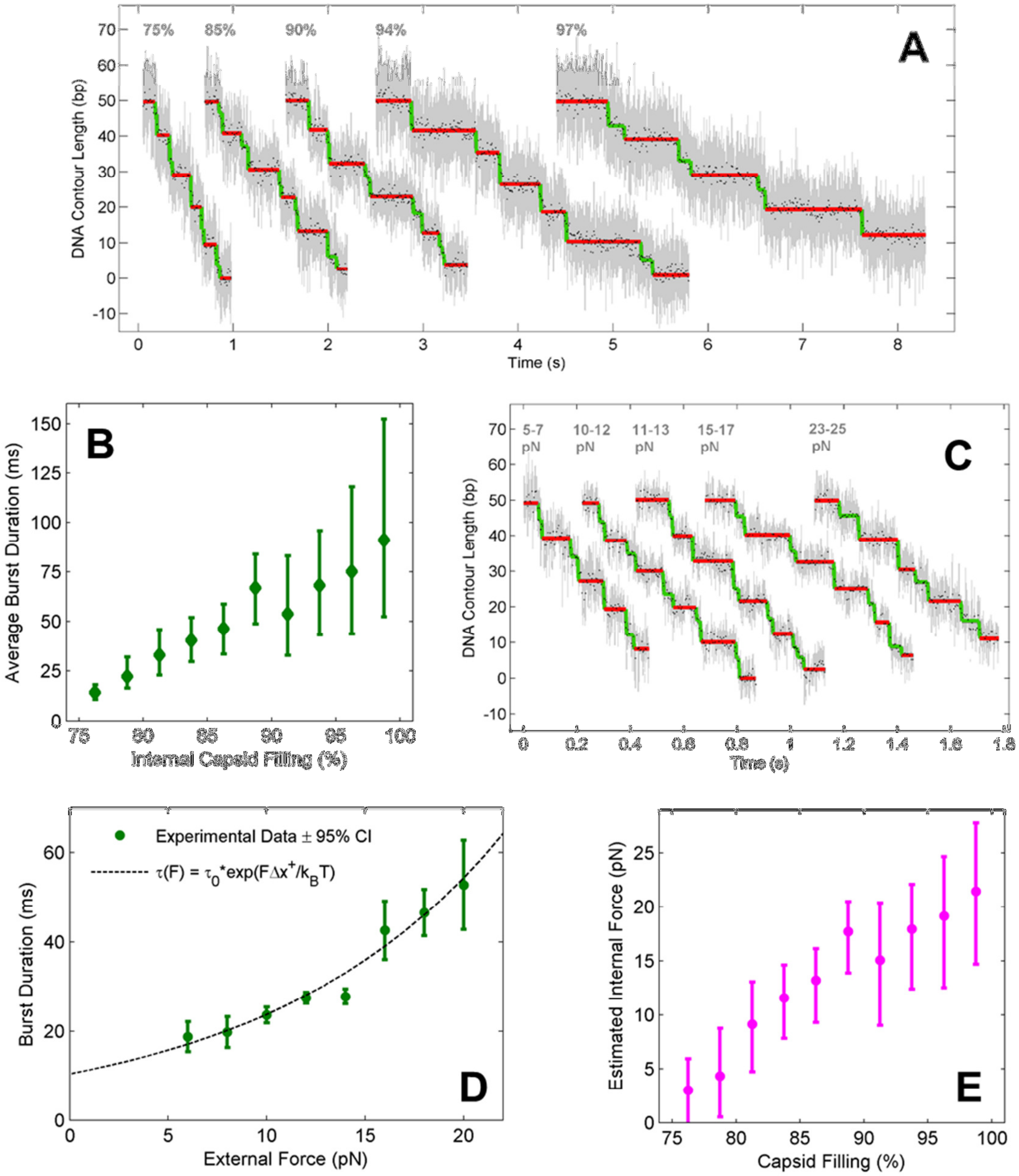


Figure 4.3.1: Measuring the Magnitude of the Internal Force within the Viral Capsid
(A) Sample packaging traces at capsid fillings of 75-100%. Raw 2.5kHz data shown in light gray, 20x downsampled data shown in black dots. Burst phase (green) and dwell phase (red) assignment was done using a modified SIC step-finding algorithm (see Chapter 2.4). (B) Average burst duration versus filling at saturating [ATP] and low external force (6-9pN). (C) Sample packaging traces at low filling (<30%) and external forces ranging from 5 pN to 25 pN. (D) Average burst duration versus external force at low capsid filling. (E) Estimate of the internal force acting via DNA on the packaging motor as a function of filling. This estimate was computed by comparing the average burst duration at various filling (B) and external force (D). Where applicable error-bars indicate the 95% confidence interval computed via bootstrapping.

Previously the internal force as a function of filling was inferred by relating the packaging velocities measured at different capsid filling levels to those measured at different externally applied forces, under the assumption that the effects of filling and external opposing force on motor dynamics are equivalent (Smith et al., 2001; Rickgauer et al., 2008). However, our results described above demonstrate that such an assumption is not valid: internal filling alters the value of $V_{\max}/K_{\mathrm{m}}$, while an opposing external force does not (Figure 4.2.1D). This new insight begs a more reliable method for estimating the internal force.

To that end we examined the dynamics of individual mechanochemical cycles from high-resolution packaging data collected at various opposing external forces and degrees of internal capsid filling (Figure 4.3.1A-B). We have previously shown that each cycle is composed of a "dwell" phase during which all five ring subunits release ADP and load ATP, and a "burst' phase during which four of the five subunits hydrolyze ATP, release P_i, and translocate DNA (Moffitt et al., 2009; Chistol et al., 2012) (Figure 4.2.1B). The burst phase contains the force-generating transition that drives DNA translocation and hence should be force-sensitive. Indeed, we observed that the burst phase is prolonged by applying an external force (Figure 4.3.1A, C). On the contrary, the duration of the stationary dwell phase is essentially insensitive to force. The packaging dynamics at different filling degrees revealed a drastically different picture (Figure 4.3.1A-B): both dwell and burst phases become longer as DNA fills the viral capsid, further confirming that internal filling and external force influence motor operation via different mechanisms. The lengthening of the burst duration caused by internal filling is equivalent to the effect of an external load, indicating that the packaging motor experiences an opposing internal force that is transmitted directly through DNA from inside the capsid. Importantly, our ability to measure the dwell and burst durations separately allowed us to estimate this internal force accurately, as opposed to previous studies that only measured the overall packaging velocity. By correlating the average burst duration at different capsid filling and different external forces, we estimated the internal force as a function of capsid filling (Figure 4.3.1E). Our results reveal an internal force of ~10 pN at 85% filling and ~20 pN at 100% filling, about 3-5 times lower than earlier internal force estimates (Smith et al., 2001; Rickgauer et al., 2008). It is worth noting that those earlier estimates were significantly higher than the resisting force predicted by several analytical modeling and numerical simulation studies (Evilevitch et al., 2004; Purohit et al., 2005; Spakowitz and Wang, 2005). Our new internal force estimate reconciles this discrepancy and may help further refine the current models for viral genome packaging and ejection (Purohit et al., 2005; Spakowitz and Wang, 2005; Tzlil et al., 2003; Forrey et al., 2006; Petrov et al, 2007).

4.4 Burst Size Dependence on Capsid Filling and Its Implications for Motor Operation

Having measured the burst durations as a function of filling, we next measured the burst sizes, i.e., the amount of DNA packaged in each motor cycle. At low filling (<30%), the average burst size is 9.8±0.3 bp, consistent with our previous measurement (Moffitt et al., 2009). Surprisingly, the average burst size monotonously decreases with filling, approaching 8.9±0.3 bp at 100% filling (Figure 4.4.1A). To our knowledge, this is the first example of a molecular motor's step size being continuously modulated by an external factor, in this case the capsid filling. We showed before that each burst consists of four steps with a non-integer step size (2.5 bp) (Moffitt et al., 2009), and that the motor-DNA interactions within a burst are mostly non-specific and steric (Aathavan et al., 2009). The past and current results together strongly suggest that the step size of the Phi29 motor is not dictated by the geometrical periodicity of the substrate track like the cytoskeleton motors kinesin (Howard et al., 1989) and myosin (Mehta et al., 1999), but is determined by the amplitude of the motor's internal conformational motions, which can be modulated externally.

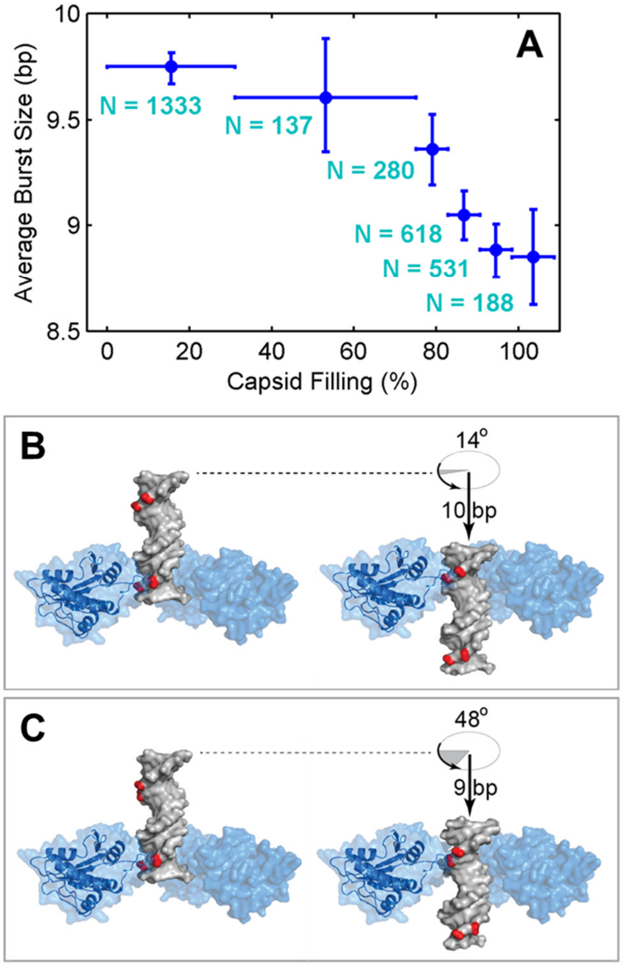


Figure 4.4.1: Burst Size Dependence on Capsid Filling and Its Implications for Motor Operation
(A) Average burst size versus capsid filling. N is the number of bursts measured at a given condition. Vertical error-bars indicate the 95% confidence interval computed via bootstrapping. (B) Predicted DNA rotation relative to the motor for a 10-bp burst size. For clarity only three gp16 subunits are shown, and a ribbon diagram is used to highlight the subunit that contacts the DNA backbone during the dwell. The pairs of backbone phosphates contacted by the motor during subsequent dwells are colored red. (C) Predicted DNA rotation relative to the motor for a 9-bp burst size.

Furthermore, the change in burst size at high filling has crucial implications for the motion of DNA through the packaging motor. We have previously proposed that one subunit in the Phi29 homo-pentameric ATPase ring is special, and that this special subunit is the one that makes the crucial contact with a pair of DNA backbone phosphates during each dwell (Chistol et al., 2012; Aathavan et et al., 2009). Because B-form dsDNA has an average helical pitch of 10.4 bp, at low filling the motor needs to rotate DNA by ~14° after a 10-bp burst in order to bring the nearest pair of backbone phosphates back into alignment with the special subunit (Figure 4.4.1B). Indeed, using a single-molecule rotor-bead assay, we recently measured a rotation density (average amount of DNA rotation per DNA length packaged) of ~1.4°/bp at low filling

(Craig Hetherington, private communication). As the burst size reduces from 10 bp to 9 bp, more rotation is needed (~5.4°/bp) to align the DNA phosphates with the special subunit (Figure 4.4.1C). This predicted rotation density change is in quantitative agreement with the experimental measurements made by Craig Hetherington (data not shown, private communication). Moreover, these results suggest that our previously proposed mechanochemical model (Chistol et al., 2012), in which the special subunit does not change its identity in successive cycles, is applicable at both low and high capsid filling.

4.5 Dwells and Long Lived Pauses at High Capsid Filling

The lengthening of the dwell duration is the main cause of the packaging velocity reduction. Interestingly, towards the late stage of packaging (>60% filling), a series of long lived pauses (LLP) punctuate regular DNA translocation (Figure 4.5.1A-B). High-resolution analysis revealed that LLPs predominantly occur in the dwell phase. The probability of a normal dwell with duration of several seconds is vanishingly small. Thus the LLP represents an off-pathway motor state caused by high filling. We then separately analyzed regular dwells and LLPs. The regular dwell duration increases with the filling level (Figure 4.5.1C), and remains [ATP] dependent (Figure 4.5.1C, inset). The Michaelis-Menten analysis above (Figure 4.2.1E-F) suggested that ATP tight-binding, which occurs in the dwell phase, is decelerated at high filling. At saturating [ATP], n_{min}, which reflects the number of rate-limiting steps in the dwell, decreases from ~5 at 60% filling to ~2 at 100% filling (Figure 4.5.1D). We have previously shown that ADP release events are rate-limiting during the dwell at low filling (Chistol et al., 2012). Our current analysis indicates that ATP tight-binding becomes rate-limiting at high filling (Figure 4.2.1C-D), and that not all ATPase subunits have the same tight-binding rate.

We then investigated the properties of LLPs. Both the frequency and the duration of LLPs increase as DNA filling proceeds (Figure 4.5.1E-F). Thus the motor spends most of its time in this inactive state, making the packaging velocity reach essentially zero at 100% filling. The frequency of LLP is independent of [ATP] (Figure 4.5.1E, inset), indicating that pause entry occurs in a different kinetic block from ATP binding. i.e., the motor enters the LLP state either before or after all subunits have bound ATP (Figure 4.2.1B). The application of large opposing external loads does not induce LLPs; thus it is unlikely that these pauses are caused by the buildup of the internal force. An alternative origin of LLP would be conformational changes of the motor induced by the highly compressed DNA. The allostery between the capsid and the motor has been found in the related bacteriophage P22, whose connector (portal) undergoes a significant conformational change when the capsid is tightly packed with DNA (Lander et al., 2006). Here we show that such structural rearrangement is likely further transmitted into the ATPase. Although the Phi29 ATPase packages a unit-length genome in nature and lacks an endonuclease domain, it is still tempting to speculate that LLPs may be part of the "head-full" mechanism exploited by many dsDNA bacteriophages to switch on their nuclease activity and to cleave their concatemeric genomes once the filling level reaches certain threshold (Casjens, 2011). The occurrence of LLP is heterogeneous among different viral particles, consistent with the variation of the packaged genome size observed in the head-full phages (Nurmemmedov et al., 2007).

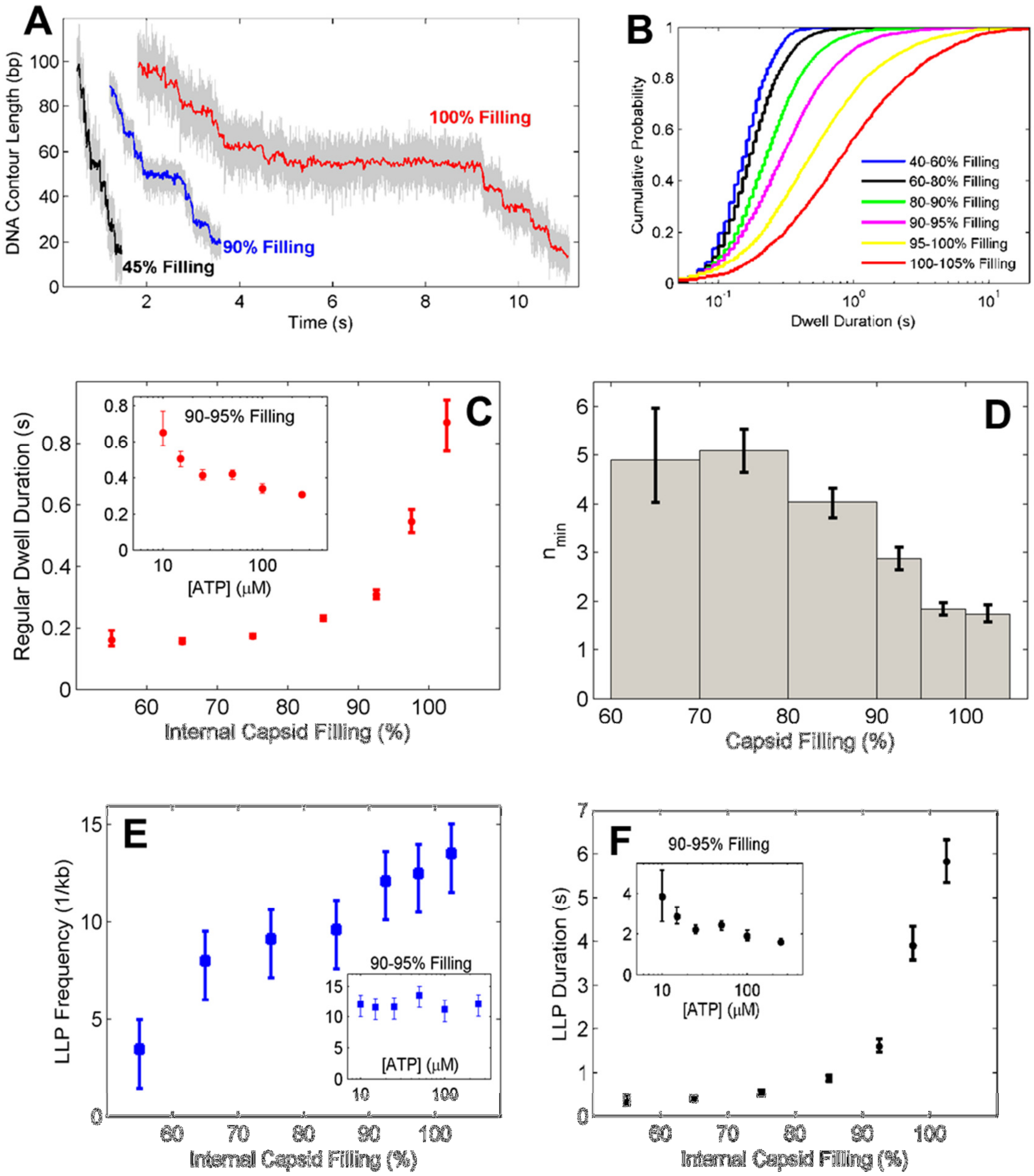


Figure 4.5.1: Dwells and Long Lived Pauses at High Capsid Filling

(A) Sample packaging traces at various fillings illustrate the existence of unusually long dwells. (B) Cumulative dwell time distribution at various capsid fillings. (C) Mean duration of regular dwells as a function of filling at saturating [ATP]. Inset: Mean dwell duration is [ATP] dependent at a fixed filling. (D) n_{min} dependence on capsid filling. n_{min} represents the lower-limit for the number of rate-limiting events during the dwell phase. (E) Frequency of long lived pauses (LLP) increases with capsid filling at saturating [ATP]. Inset: LLP frequency is independent of [ATP] for a given filling. (F) The average LLP duration increases with filling at saturating [ATP]. Inset: Average LLP duration versus [ATP]. Where applicable error-bars indicate the 95% confidence interval computed via bootstrapping.

4.6 Role of the Connector at High Filling

Upon packaging termination, the ATPase and pRNA rings detach from the prohead and the tail proteins assemble onto the connector, generating an infectious virion (Rao and Feiss, 2008). Therefore the connector likely plays a central role in orchestrating the transition from packaging to tail assembly. We created a connector mutant by deleting a DNA-contacting loop (channel loop) projecting into the central channel through which the DNA passes (Figure 4.6.1A, inset). The mutant motor showed drastically different packaging dynamics from the wild-type motor at high filling (Figure 4.6.1A).

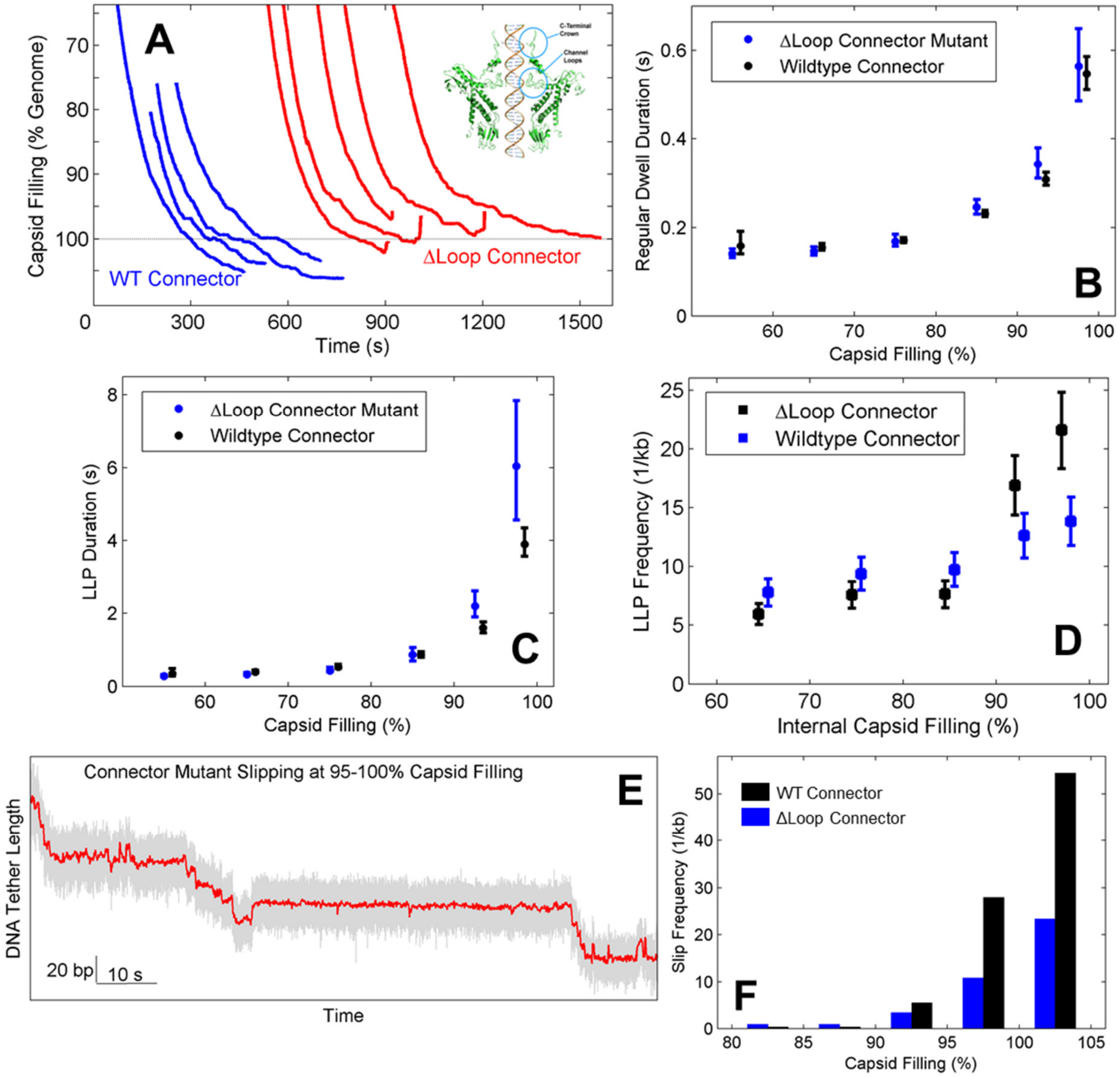


Figure 4.6.1: The Role of the Connector at High Filling

(A) Sample packaging traces with 21 kb DNA tethers of wild-type connectors (blue) and mutant connectors with channel loop deletions (red). (B) The mean duration of regular dwells is essentially the same for WT and ΔLoop mutant connectors. (C) LLP duration at very high filling is longer for the ΔLoop mutant compared to the WT connector. (D) LLP frequency at very high filling is greater for the ΔLoop mutant than for the WT connector. (E) Detail view of a ΔLoop mutant trace containing regular dwells, LLPs, and micro-slips. (F) At very high filling the ΔLoop mutant connector is more prone to slipping than the WT connector. Where applicable error-bars indicate the 95% confidence interval computed via bootstrapping.

Most of the wild-type motors managed to package 1.05 times the genome length, while most of the mutant motors underwent fatal slippage before 100% filling. This result suggests that the mutant fails to retain the packaged DNA within the capsid, confirming a recent bulk biochemical study (Grimes et al., 2011). Interestingly, high-resolution analysis revealed that the normal dwells are not affected by the mutation (Figure 4.6.1B), but the duration and frequency of LLPs are higher for the mutant (Figure 4.6.1C-D). In addition, the mutant showed elevated slipping frequency (Figure 4.6.1E-F). These observations indicate a crosstalk between the ATPase and the connector that is filling-dependent. At low filling, the connector channel loop is not essential for translocation. However, once LLP starts to occur, the interaction between the DNA and the channel loop is needed for the ATPase to exit from an LLP and to continue packaging to full-length. This interaction also appears to prevent the motor from slipping. After the ATPase dissociates, the connector alone serves as the valve to keep the DNA inside the capsid until tail assembly further stabilizes the fully-packaged virion (Fang et al., 2012; Tavares et al., 2012).

4.7 Summary and Conclusions

The decrease of packaging rate caused by internal filling has been reported in several dsDNA viruses including bacteriophages Phi29 and λ (Smith et al., 2001; Fuller et al., 2007; Rickgauer et al., 2008), and until now has been simply explained by an effective internal force that opposes packaging. Here we show that this view oversimplifies the profound influence that DNA filling confers on the packaging motor. The internal force even at the end of packaging is much lower than the motor's stall force (>60 pN) (Smith et al., 2001) and hence is not enough to halt DNA translocation.

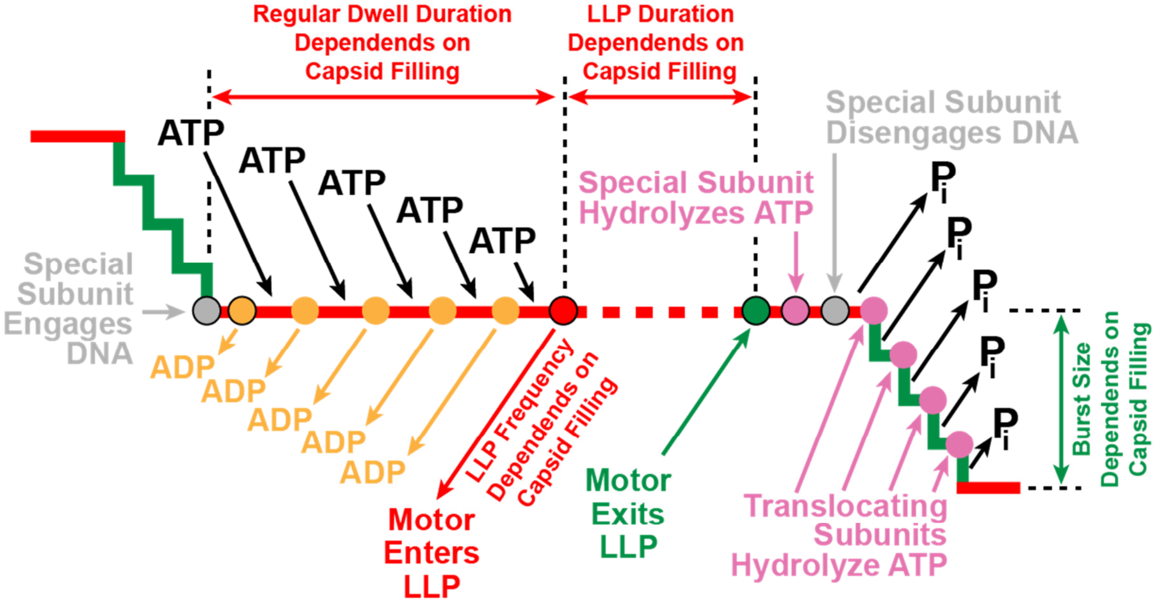


Figure 4.7.1: Revised Mechano-Chemical Model of the Phi29 Packaging Motor's Operation
The thick red and green lines represent the dwell ad burst phases respectively. Time flows from left to the right. ATP binding, ADP release, ATP hydrolysis and Pi release are indicated by arrows. In response to the capsid filling up with DNA, the motor is throttled down via several mechanisms: (1) the lengthening of the regular dwell duration, (2) the appearance of long lived pauses, and (3) the reduction in burst size.

The slower rate of nucleotide tight-binding, the reduced step size, and the more frequent transition to a prolonged paused state all point to a unique motor throttling mechanism in which the ATPase is continuously adjusting its operation in response to the changing environment inside the capsid (Figure 4.7.1). Given that the internal DNA organization differs among viral particles (Leforestier and Livolant, 2010; Widom and Baldwin, 1983) and thus the internal force is heterogeneous, the active allosteric mechanism may prove to be evolutionally advantageous in uniformly terminating packaging compared to the passive stalling mechanism. The internal force is even lower *in vivo* because of the high ionic strength (Fuller et al., 2007) and, in addition, is countered by the cytoplasmic osmotic pressure (Evilevitch et al., 2003). Therefore, it is unlikely to be the main driving force for genome ejection during viral infection as previously conjectured. Recent studies (Molineux and Panja, 2013; Gonzalez-Huici et al., 2004; Van Valen et al., 2012) have suggested that the viral genome is translocated into the host cell by other mechanisms such as water flow or enzymes such as RNA polymerase actively pulling on DNA. Finally, our work highlights that, with the physical and chemical properties of a growing number of molecular motors being elucidated, understanding how their operation is modulated by their physiological partners will shed new light on the mechanisms of the biological processes performed by these nano-machines.

Chapter 5
Additional Projects and Conclusions

5.1 Homology-Based Modelling and Structural Analysis of the Phi29 Gp16 ATPase

Our study of the Phi29 packaging motor has been impeded by the lack of a crystal structure for the gp16 ring ATPase. Here I used Phyre2 to generate a structural model of gp16. I located the canonical Walker A, Walker B, and arginine finger motifs in the Phi29 ATPase model. In addition, I identified a putative DNA-binding loop with two conserved residues that could explain the nature of the observed motor-substrate interaction (Aathavan et al. 2009).

I gathered distant homologs of Phi29 gp16 (332 aa, RefSeq code YP_002004545) in three iterations of Psi-Blast against the NR protein database with an E-value threshold of 1e-5 (Altschul et al. 1997). I used MAFFT with the E-INS-I option (Katoh et al. 2008) to obtain a multiple sequence alignment (MSA). I viewed and edited the MSA using Jalview (Waterhouse et al. 2009). The final MSA (Figure 5.1.1) contains 18 sequences from the Psi-Blast search (including the query), all of which are confirmed or predicted viral packaging proteins. For reference I also included five well-studied viral packaging ATPases (Guo et al. 1987) and ten extensively studied ATPases from the RecA-like branch of the ASCE superfamily. In the final MSA sequences are labeled either "Psi-Blast", "GuoPaper", or "RelATPase" to denote how they were selected (Figure 5.1.1).

I used the Phyre2 structure prediction web-server to model Phi29 gp16 in normal mode (Kelley et al. 2009). Human RNA helicase A (PDB code 3LLM) was automatically selected as the template for homology modelling. Phyre2 modeled 150 out of 332 residues in gp16 (45% coverage) with 97% confidence. I compared the predicted structure of gp16 with the solved crystal structure of FtsK (PDB code 2IUU), a distantly related dsDNA translocase from *Pseudomonas aeruginosa* (Massey et al. 2006). Since the RNA helicase (Schutz et al. 2010) that served as the template for gp16 does not form an oligomeric ring, I could not use it to analyze and interpret the gp16 model. I relied on PyMol to view and align the FtsK structure and the gp16 model. Using FtsK for comparative analysis with the gp16 model provides two major advantages: (i) like gp16, FtsK is a dsDNA translocase; and (ii) both FtsK and gp16 belong to the same branch (PilT/FtsK) of the ASCE superfamily. However, there are also drawbacks to using FtsK: (i) FtsK forms a homo-hexamer whereas gp16 forms a homo-pentamer; and (ii) FtsK was not co-crystallized with its dsDNA substrate (in fact the vast majority of ASCE ring translocases have not been co-crystallized with their substrate).

I used the final MSA to identify three structural motifs in gp16: Walker A, Walker B, and the arginine finger (Figure 5.1.1). The catalytic pocket in ring NTPases is located at the interface between two subunits. While one subunit provides the Walker A/B motifs for nucleotide binding, the other subunit inserts its arginine finger in the proximity of the bound nucleotide and stimulates hydrolysis. The arginine finger is also thought to play an role in coordinating the activity of individual subunits in the ring (Lyubimov et al. 2011). In Phi29 gp16 Walker A is located at G24-S31 and Walker B is found at V112-F120 (Figure 5.1.1). The multiple sequence alignment identified R146 as the putative arginine finger in gp16 (Figure 5.1.1).

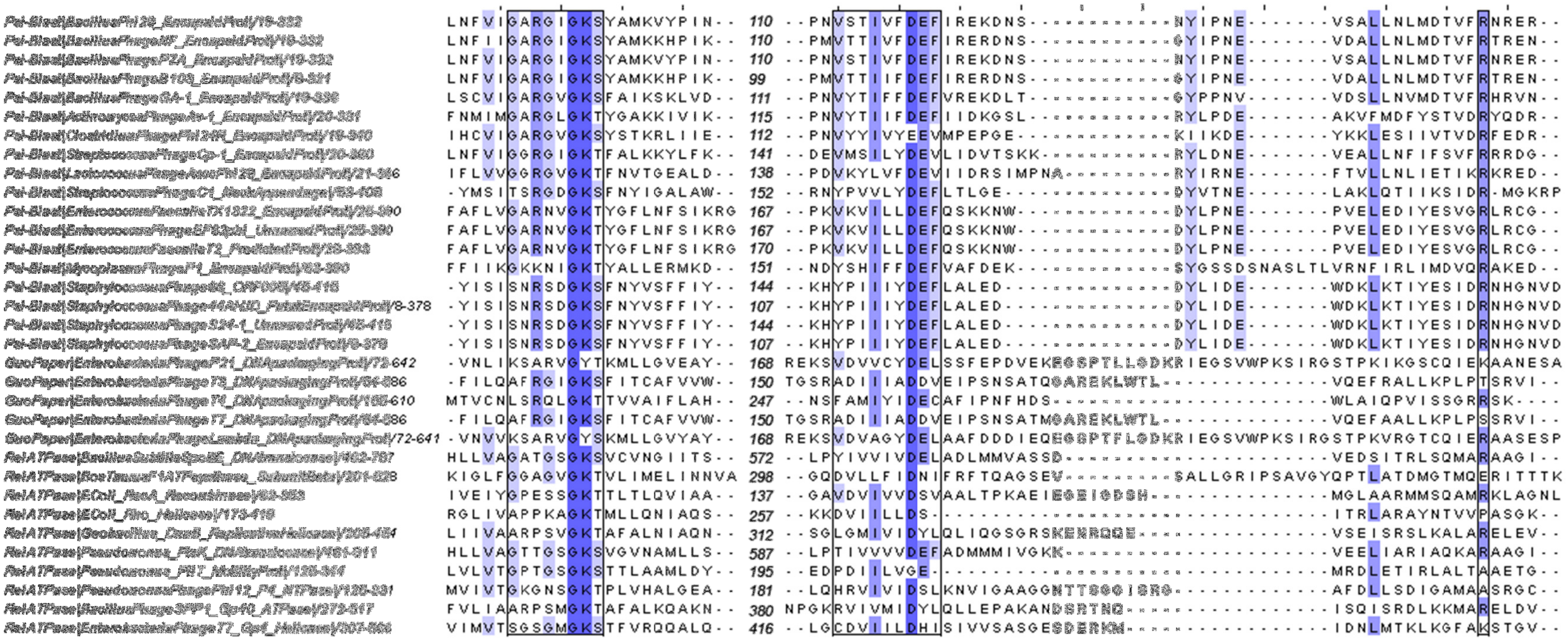


Figure 5.1.1: Multiple Sequence Alignment of Phi29 gp16 and Related ATPases

The MSA includes top gp16 homologues identified by Psi-Blast (labeled "PsiBlast"), five well-characterized viral packaging ATPases (labeled "GuoPaper"), and ten RecA-like ASCE NTPases (labeled "RelATPase"). The first box contains the Walker A motif, the second box contains the Walker B motif, and the third box highlights the arginine finger motif. In Phi29 gp16 Walker A spans G24-S31, Walker B spans V112-F120, and R146 is the putative Arginine finger.

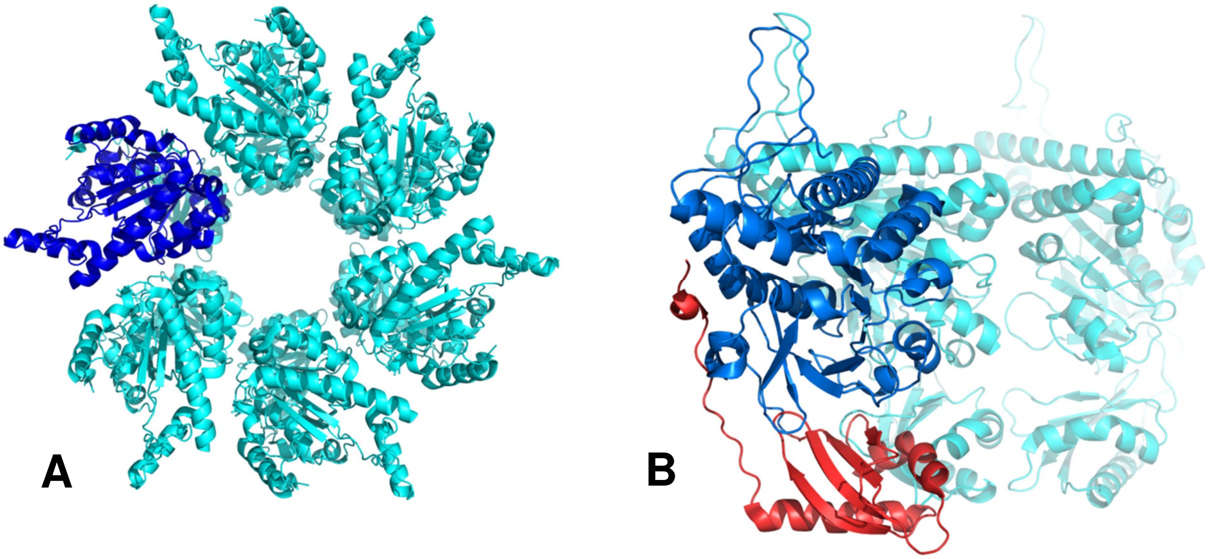


Figure 5.1.2: The Crystal Structure of FtsK – a Homo-Hexameric Ring ATPase Related to Phi29 Gp16
(A) Overview of the FtsK homo-hexameric ring ATPase (Massey et al. 2006). One subunit is highlighted in blue, the rest are colored cyan. (B) Each FtsK subunit consists of a β-domain (blue), and an α-domain (red). The β-domain is related to other RecA-like NTPases, the α-domain is unique to FtsK.

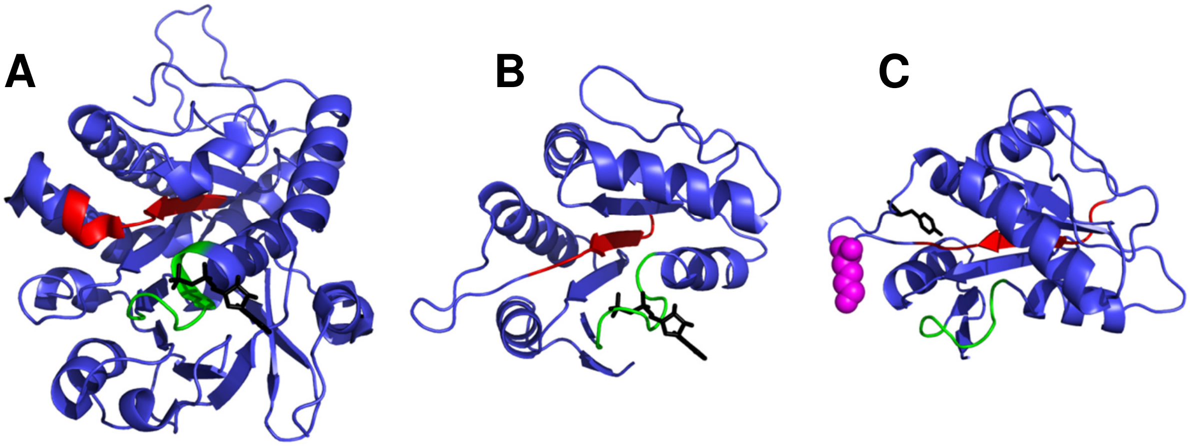


Figure 5.1.3: Comparing Walker A/B Motifs in FtsK and Phi29 Gp16
(A) FtsK motor domain with Walker A shown in green, Walker B in red, and ADP in black. (B) Phi29 gp16 model with Walker A in green, Walker B in red, and ADP in black. (C) Close-up view of the putative DNA-binding loop of gp16. A conserved Lysine (K124), shown with magenta spheres, is best positioned to form the unique electrostatic contact with the DNA backbone during the dwell. A conserved Tyrosine (Y129) is shown with black sticks. Y129 could be used to push DNA via steric interactions during the burst. Although Y129 faces toward the Walker B motif (red) in this model, it is possible that Y129 would be flipped out, and therefore best positioned to contact the dsDNA substrate during the burst.

I used the crystal structure of the FtsK to further analyze the gp16 model generated by Phyre2. FtsK is a dsDNA translocase that is responsible for chromosome segregation in dividing bacteria. FtsK assembles as a homo-hexameric ring (Figure 5.1.2A), and each subunit is composed of two domains, α and β, shown in red and blue in Figure 5.1.2B (Massey et al. 2006). While the α-domain is unique to FtsK, the β-domain is similar to other RecA-like NTPases. In the remaining figures and analysis, I show and make references only to the β-domain of FtsK. Figure 5.1.3A illustrates the location of Walker A (G466-S473) and Walker B (T589-F597) motifs in the structure of an FtsK subunit. Walker A is located at the junction of an alpha helix and a loop and interacts with the ADP bound in the catalytic pocket (Figure 5.1.3A, black sticks). Walker B is located in the hydrophobic core of the protein (Figure 5.1.3A, green) but also extends out to a loop and the beginning of an alpha-helix.

The gp16 homo-pentamer model was obtained by aligning five gp16 subunits with five subunits from the FtsK hexamer, then rotating each gp16 monomer by 12 degrees (Figure 5.1.4). The central pore of the resulting ATPase ring is large enough to accomodate dsDNA. Although FtsK and gp16 share a very low sequence identity, the main features of these two proteins align reasonably well. To identify a putative DNA-binding loop in gp16, I viewed the reconstructed gp16 homo-pentamer from the side as shown in Figure 5.1.4B, which illustrates gp16 orientation relative to DNA passing through the central pore of the ring. Only one loop from each subunit protrudes into the central channel, and this loop contains two conserved residues, K124 and Y129. Conserved positively charged residues have been identified in the substrate-binding loops of several ring translocases, and they have been shown to form specific electrostatic contacts with the nucleic acid backbone (Enemark et al. 2006, Thomsen et al. 2009, Itsathitphaisarn et al, 2012). It is therefore plausible that the K124 residue in gp16 is involved in making the unique contact between the Phi29 motor and dsDNA during the dwell phase (Aathavan et al. 2009). A similar conserved bulky hydrophobic residue has been identified in the substrate-binding loop of the protein unfoldase ClpX, which is thought to interact with the substrate in a non-specific fashion (Glynn et al. 2009). It is conceivable that the conserved Y129 in gp16 pushes the substrate during the burst phase, when the Phi29 motor relies on non-specific interactions to move DNA in 2.5-bp steps.

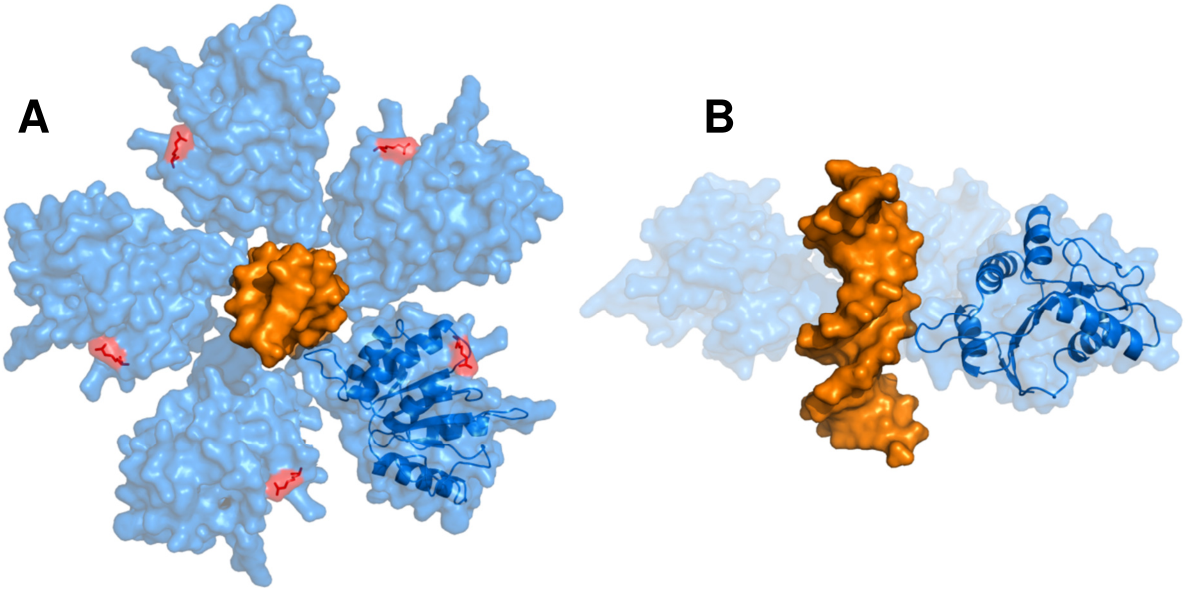


Figure 5.1.4: The Structural Model of The Phi29 ATPase Ring

(A) The gp16 homo-pentamer model was obtained by aligning five gp16 subunits with five subunits from the FtsK hexamer, then rotating each gp16 monomer by 12 degrees. The central pore of the resulting ATPase ring is large enough to accomodate dsDNA (orange). The putative arginine (R146) finger is shown in red. (B) Side view of the ATPase ring showing that the putative DNA-binding loop is well positioned to interact with the substrate. For clarity, only three gp16 subunits are shown.

In this study I used a homology-based structure prediction server to model the Phi29 gp16 ATPase, and compared the predicted structure with the solved structure of a distantly related dsDNA translocase. I located the canonical Walker A, Walker B, and arginine finger motifs in gp16, and identified a putative DNA-binding loop. Two conserved residues – K124 and Y129 – may help explain how the Phi29 motor makes a specific electrostatic contact with its substrate during one part of its cycle (dwell) and a non-specific contact with DNA during another part of the cycle (burst). While this gp16 model may be useful for a superficial characterization and could guide targeted mutagenesis, it has several shortcomings: (i) this model includes only about half of all amino-acids in the gp16 sequence; (ii) it does not reveal the complex network of interactions at various stages in the hydrolysis cycle; and (iii) it does not display the gp16 monomer in the native context of a pentameric ring.

5.2 Devising a Built-In Molecular Ruler with Biotin-Streptavidin Barriers

Dual trap optical tweezers enable us to resolve relative motion at the sub-nanometer scale, but it is relatively difficult to measure absolute distances very precisely. For example we know the position of the Phi29 motor on a DNA tether within ~100-200 bp. This uncertainty arises mainly for two reasons: (1) the usual Brownian power-spectrum calibration is accurate within ~5-10%, and (2) there is some variability in the bead diameter size (~5%). Such small calibration errors have not posed a major problem until recently, when we measured a decrease in burst size from ~10 bp at low capsid filling, to ~9 bp at very high capsid filling. These experiments require not only high precision (which can be achieved by repeating the experiments many times and accumulating a lot of measurements), but also high accuracy, which begs for a novel calibration approach. I came up with an internal molecular ruler concept, which requires at least two pausing sites engineered at specific locations on the DNA tether being packaged by the Phi29 motor (Figure 5.2.1). The motor would package normally until it encounters the first site, where it pauses for a while then continues packaging until the second site, where it pauses again. If the two pauses can be identified unambiguously, one can easily determine both an offset correction, and a scaling correction for the DNA tether length.

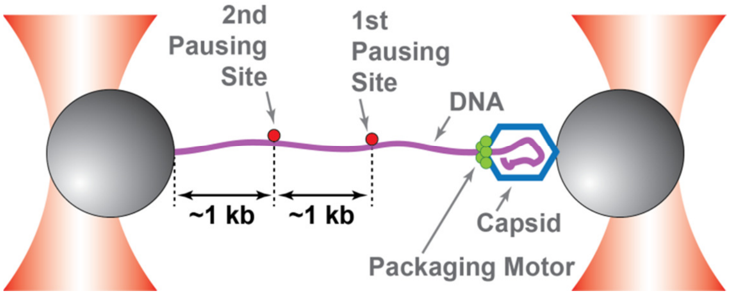


Figure 5.2.1: Built-In Molecular Ruler Concept for Phi29 Single Molecule Experiments
The Phi29 packaging complex is tethered between two polystyrene beads held in two optical traps. Two pausing sites are engineered at specific locations along the DNA substrate, and the motor must pause at each site long enough to unambiguously identify the pause. In addition, the motor should be able to successfully traverse the site after the pause.

A similar internal-calibration concept has been used for correcting calibration errors in RNA Polymerase II single molecule experiments, using a programmed stall site where PolII pauses if a certain nucleotide is omitted from the buffer (Hodges et al., 2009). I chose an internal biotin-streptavidin obstacle as the pausing site for Phi29 due to the strength of the interaction and the simplicity of engineering an internally biotinylated DNA substrate. It has been previously shown in biochemical assays that a ring motor closely related to gp16, FtsK, can break the biotin-streptavidin linkage (Chivers et al., 2010). As a proof of principle, I designed a DNA tether 6.7-kb long with a gp3-molecule on one end, and a dig moiety on the opposite end. I engineered a single internal biotin 1 kb away from the dig end. The gp3 molecule is necessary for packaging initiation in vitro. The dig moiety is used to attach the DNA tether to an anti-Dig antibody coated bead (figure 5.2.2, left panel). I used this construct to make stalled Phi29 packaging complexes as described previously (Smith et al., 2001; Chemla et al., 2005) that were restarted during the optical tweezers experiment. Streptavidin was added to the stall buffer to create a barrier that would be too bulky to pass through the central pore of the Phi29 packaging motor. The right panel in Figure 5.2.2 illustrates representative traces obtained in single-molecule packaging experiments with a single internal biotin-streptavidin barrier. The Phi29 motor paused for several seconds when it encountered the bulky obstacle. About 40% of the packaging complexes that reached the biotin-streptavidin barrier successfully crossed it and translocated the entire DNA tether, whereas the other 60% underwent terminal slips from the obstacle-induced pause.

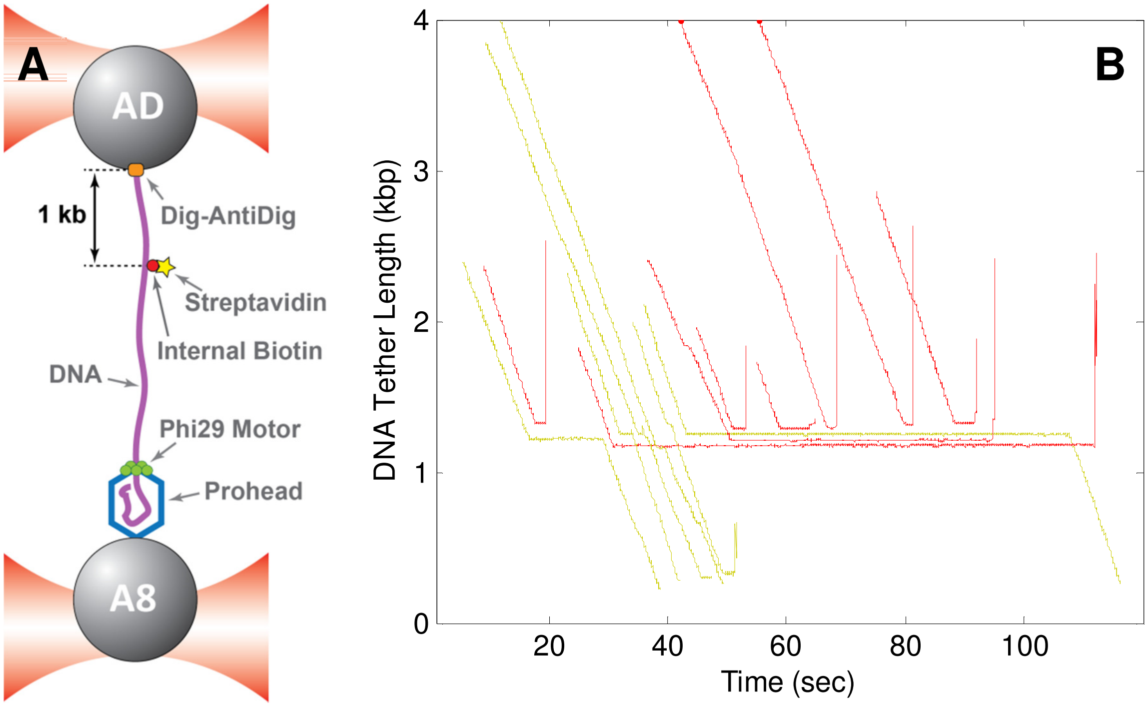


Figure 5.2.2: The Phi29 Packaging Motor Can Cross a Single Streptavidin-Biotin Obstacle
Left: The single molecule experiment geometry. DNA was attached to the top bead via a dig-antidig linkage. A biotin was covalently attached to the DNA substrate ~1kb away from the dig terminus, and streptavidin was later bound to the biotin. Note that the packaging buffer does not contain any streptavidin, to prevent re-binding by molecules from solution. Right: Representative Phi29 translocation traces on the DNA substrate with an internal biotin-streptavidin barrier. ~40% of all traces successfully cross the obstacle (green), and the remaining ~60% slip off the DNA substrate after attempting to cross the barrier (red).

A detailed look at the barrier-induced pauses reveals multiple futile crossing attempts (Fig 5.2.3), in which the motor translocates up to 10-bp worth of DNA and recoils back to its original position, suggesting that the motor does not fully disengage the DNA substrate during these attempts. In addition, it appears that the terminal slips take place after a futile barrier crossing attempt. I would like to point out that the Phi29 motor successfully crossed the biotin-streptavidin barrier only when the packaging buffer contained no free streptavidin, and the only streptavidin present in the experiment was the molecule bound to the internal biotin covalently attached to the DNA tether. The lifetime of a single biotin-streptavidin linkage is on the order of hundreds of hours, so it is highly unlikely that the successful barrier crossings observed in my experiments were due to the spontaneous dissociation of streptavidin. I would also like to point out that the position of the obstacle-induced pause varies among different single-molecule trajectories, and is mainly due to differences in bead size, ~100-200 bp, corresponding to ~30-60 nm, i.e. within the bounds specified by the bead manufacturer. Although having two engineered pausing sites is optimal for correcting any calibration errors, a single pausing site is also sufficient, since the packaging motor stops when it reaches the end of the DNA tether, effectively providing an internal control. In the initial experiments I collected a few traces that successfully crossed the barrier and exhibited a clear pause at the end of the DNA tether, which allowed me to conclude that our instrument can measure relative changes in contour length with a ~3% error.

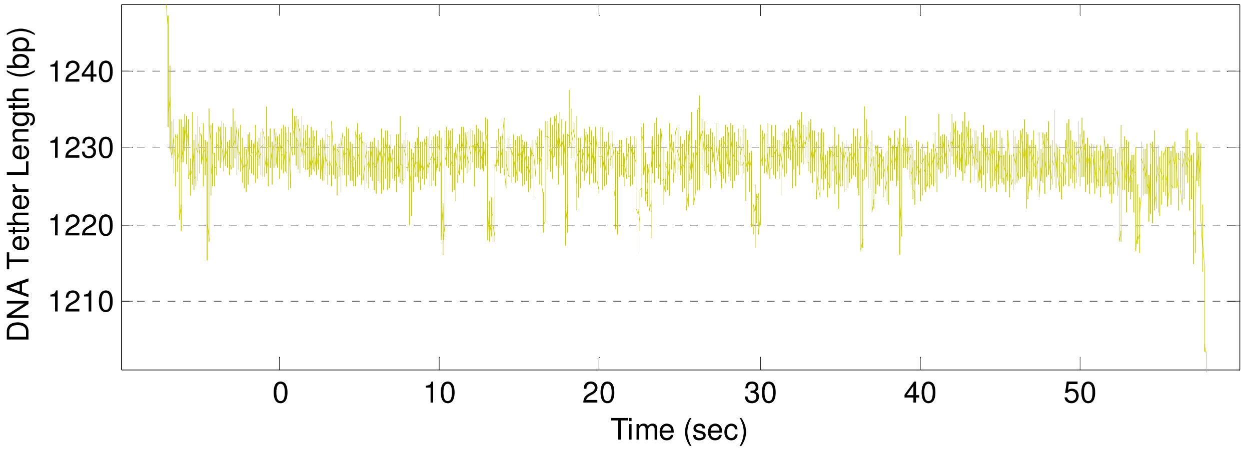


Figure 5.2.3: Phi29 Pauses in Front of a Biotin-Streptavidin Barrier Before Successfully Crossing It
The Phi29 packaging motor successfully crosses the obstacle by "bulldozing" the streptavidin from the biotin that was covalently attached to the DNA tether. Note that Phi29 requires several attempts before succeeding, and that each attempt is no larger than the 10-bp burst size of the motor.

My initial experiments show that an internal biotin-streptavidin barrier induces easily detectable pauses at a specific location on the DNA tether. In future experiments I will test a complete internal molecular ruler construct as illustrated in Figure 5.2.1, which will allow me to localize the pause within a 10-bp window (i.e. a single burst size), enabling us to accurately measure ~1 kb change in DNA contour length with an uncertainty of 10-20bp. In addition, an internal biotin-streptavidin obstacle provides a convenient marker for a variety of experiments in which knowing the absolute position of the motor on DNA is essential. For example, we intend to study how the Phi29 packaging motor translocates A-form RNA (~50-100 bp of dsRNA), and we can use an internal obstacle to indicate the location of the dsRNA patch in the single-molecule trajectory. The biotin-streptavidin obstacle also provides some insight about how a ring ATPase bulldozes proteins off a DNA duplex. It would be interesting to measure how the crossing probability and pause duration depends on various factors: ATP concentration, external force, the identity of the DNA strand containing the biotin moiety, etc.

5.3 Investigating the pRNA Role in Motor Operation with Single Molecule Experiments

Single molecule experiments allow us to focus solely on DNA packaging, which is not possible with ensemble packaging assays, where the effects of initiation and packaging cannot be disentangled. Mutations in the pRNA could potentially affect several chemical/mechanical transitions in the packaging motor: ATP binding rate, ATP Hydrolysis rate, Pi release rate, ADP release rate, various aspects of ATPase inter-subunit coordination, burst size, ATPase-DNA interaction, capsid-ATPase interaction. Previous studies have shown that ADP release and ATP binding occur during the dwell, while ATP hydrolysis and Pi release occur during the burst (Moffitt et al., 2009, Chistol et al., 2012). I used the single-molecule packaging assay to screen a variety of pRNA mutants produced by Shelley Grimes, and this section provides a summary of my findings.

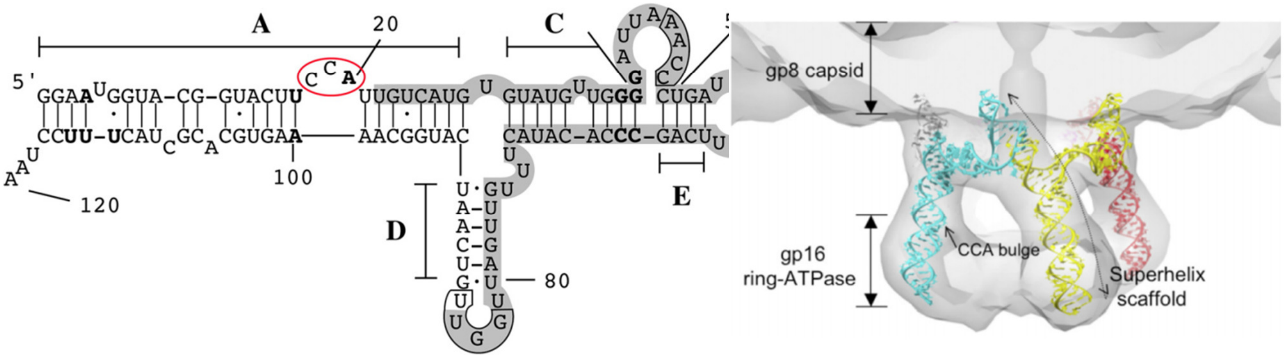
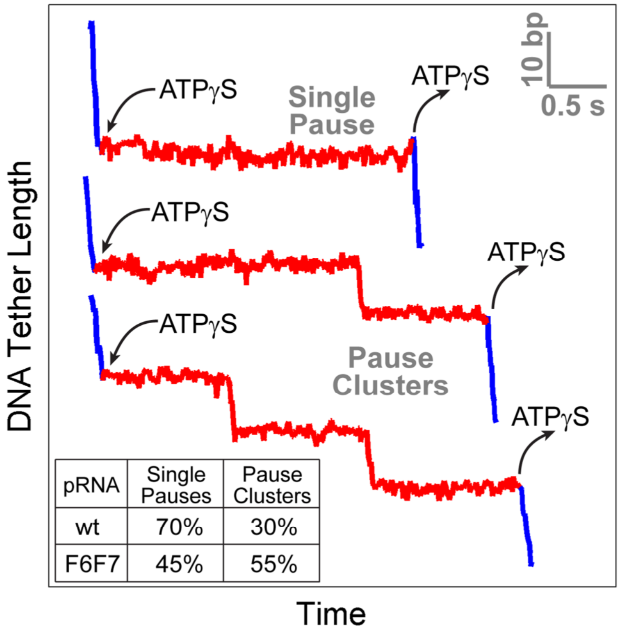


Figure 5.3.1: The Structure of Phi29 pRNA Monomer and Homo-Pentamer
Left: Predicted secondary structure of the 120 base pRNA. Bases that are conserved in Phi29 relatives are shown in bold. The bases involved in the intermolecular pseudoknot that forms the pRNA multimer are boxed. Bases 1–28 and 92–117 form the A-helix. The CCA bulge is circled in red; the adenine residue in bold is conserved among the Phi29 relatives. Reprinted from Zhao et al., 2008 with permission from Elsevier © 2008. Right: Monomeric pRNAs (yellow, cyan, red) fit within the cryo-EM electron density map of the prohead with pRNA. Reprinted from Ding et al., 2011 with permission from the National Academy of Sciences © 2011.

The F6+F7 pRNA ring consists of alternating F6 and F7 monomers, which are essentially identical to wildtype monomers, except at the molecular pseudo-knot (see Figure 5.3.1, left panel). The four bases that form the pseudoknot have been engineered to arrange pRNA monomers in an alternating fashion: either F6+F7+F6+F7+F6 or F7+F6+F7+F6+F7. Note that in each case there is a discontinuity within the pRNA ring because two neighboring F6 or F7 monomers cannot base-pair at the pseudoknot. Bulk packaging and ATPase assays indicate that F6+F7 pRNA rings behave just like the wildtype. Single molecule packaging experiments confirmed the results obtained in bulk experiments. I found that packaging complexes with F6+F7 pRNA translocate at 110+/-5 bp/s, essentially indistinguishable from wildtype, and that the F6+F7 pRNA did not affect the burst size. Since the pRNA ring symmetry is broken in F6+F7 assemblies, I hypothesized that it might affect how the packaging motor behaves during an ATP analog-induced pause cluster (see Chapter 3). During a pause cluster, one ATPase subunit remains bound to ATPγS, while the remaining four subunits remain capable of turning over ATP and translocating DNA in 10-bp bursts. I performed packaging experiments with F6+F7 mutants in a saturating [ATP] buffer containing trace amounts of ATPγS, and found that F6+F7 mutants are more likely to enter into pause clusters compared to WT pRNA (Figure

5.3.2). This difference is statistically significant and it suggests that the mismatched interface within the F6+F7 pRNA ring does affect how the gp16 ATPase subunits communicate.



pRNA	Single Pauses	Pause Clusters
wt	70%	30%
F6F7	45%	55%

Figure 5.3.2: Assessing the Effect of the pRNA Interface Mismatch Within the F6+F7 pRNA Ring
F6+F7 pRNA ring exhibits more ATPγS-induced pause clusters (55+/-5%) than the wildtype pRNA (30+/-5%).

AbBeEaΔCC is a pRNA construct in which three pRNA monomers (Ab, Be, and Ea) have been arranged in a specific order by engineering the intermolecular pseudoknot. For example the Ab monomer can interact with Ea and Be monomers to form the (Ea)(Ab)(Be) trimer. ΔCC is a pRNA mutation in which two C bases are deleted from the CCA bulge on the A helix (Figure 5.3.1, left panel). Homomeric pRNA rings with the ΔCC mutation support binding of gp16 to the prohead, but the resulting complexes are incapable of packaging detectable amounts of DNA (Zhao et al., 2008). This mutation has been shown to alter the angle of the pRNA super-helices that for a pentameric scaffold for the gp16 ATPase (Zhao et al., 2008). The AbBeEaΔCC pRNA was designed to test DNA packaging activity in the presence of one ΔCC mutation - (Ab)(Be)(EaΔCC)(Ab)(Be), or two ΔCC mutations - (Be)(EaΔCC)(Ab)(Be)(EaΔCC) (Shelley Grimes, private communication). Single-molecule screens revealed that packaging complexes with AbBeEaΔCC pRNA are 25-30% slower than wildtype at saturating ATP, but maintain the 10-bp burst size (Figure 5.3.3). Since translocation bursts are extremely fast, the observed decrease in velocity is entirely due to an increase in duration of the dwell phase. Most AbBeEaΔCC complexes never restart packaging in the single molecule assay, and are prone to releasing DNA from the capsid in a controlled manner (Figure 5.3.4, left panel), which is clearly distinct from slipping. This mutant has very low activity in the optical tweezers assay, and the amount of data currently available is insufficient to constrain the value of K_M with any reasonable degree of confidence.

The ΔU5 mutation removes the 5th pRNA base located on the A-helix (Figure 5.3.1, left panel), and is thought to affect the angle and flexibility at the tip the pRNA scaffold (Shelley Grimes, private communication). Bulk assays indicate that packaging complexes with homo-pentameric ΔU5 pRNA have reduced gp16 binding and DNA packaging efficiency compared to wildtype pRNA (Shelley Grimes, private communication).

pRNA	Vmax (bp/s)	Km (μM)	Vmax/Km (bp/s/μM)	Hill Coeff	Mean Length Packaged on 21kb DNA (bp)
Wildtype	115+/-5	25+/-3	4.7+/-0.6	1.2+/-0.2	19,900+/-300
ΔU5	103+/-4	20+/-3	5.2+/-0.8	1.1+/-0.3	19,300+/-200
U92ins	104+/-5	19+/-3	5.5+/-0.9	0.9+/-0.3	18,500+/-500
AbBeEaΔCC	85+/-5	na	na	na	na

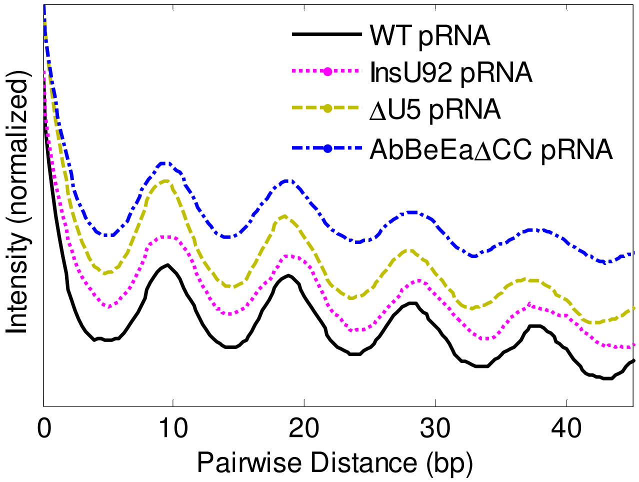


Figure 5.3.3 The Effect of pRNA Mutations on Phi29 Motor Operation
Top: Table summarizing the findings of my initial pRNA mutant screening. V_{max}, K_M, V_{max}/K_M and Hill Coefficient values refer to motor operation at low capsid filling (<30%). Bottom: Pairwise distance distributions for traces at low capsid filling (<30%) indicate that pRNA mutations do not affect the burst size of the Phi29 packaging motor.

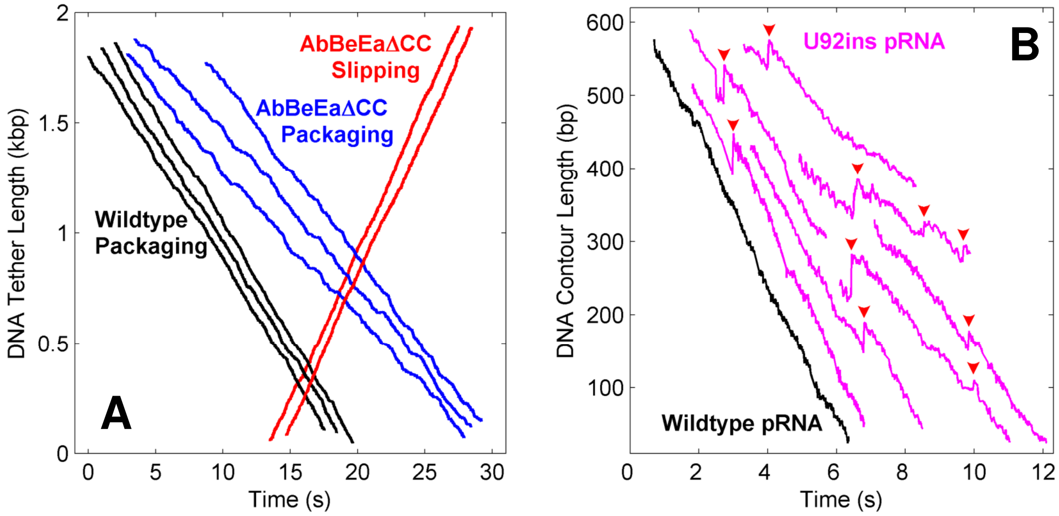


Figure 5.3.4: Anomalies in DNA Packaging Traces from pRNA Mutants
(A) AbBeEaΔCC sample translocation traces (blue) and "reverse translocation" traces (red) compared to wildtype translocation trajectories (black). (B) U92ins traces (magenta) illustrating slipping (red arrowheads) at [ATP] = 25μM (~K_M) at the beginning of the trace. Wildtype packaging complexes (black) do not slip at this ATP concentration.

Initial screens of the homo-pentameric ΔU5 pRNA mutant in single-molecule experiments revealed no significant difference between the mutant and the wildtype at low capsid filling (Figure 5.3.3). I found that packaging complexes with the ΔU5 homo-pentamer are unable to retain DNA within the capsid at very high filling (Figure 5.3.5), and undergo fatal slips when the capsid is nearly full. In these experiments the packaging motor is given a 21-kb DNA tether, which is longer than the 19.3-kb genome of Phi29. In contrast, wildtype packaging complexes package about 20 kb of DNA before reaching a complete stop, and are able to prevent the DNA already inside the capsid from slipping out (Figure 5.3.5).

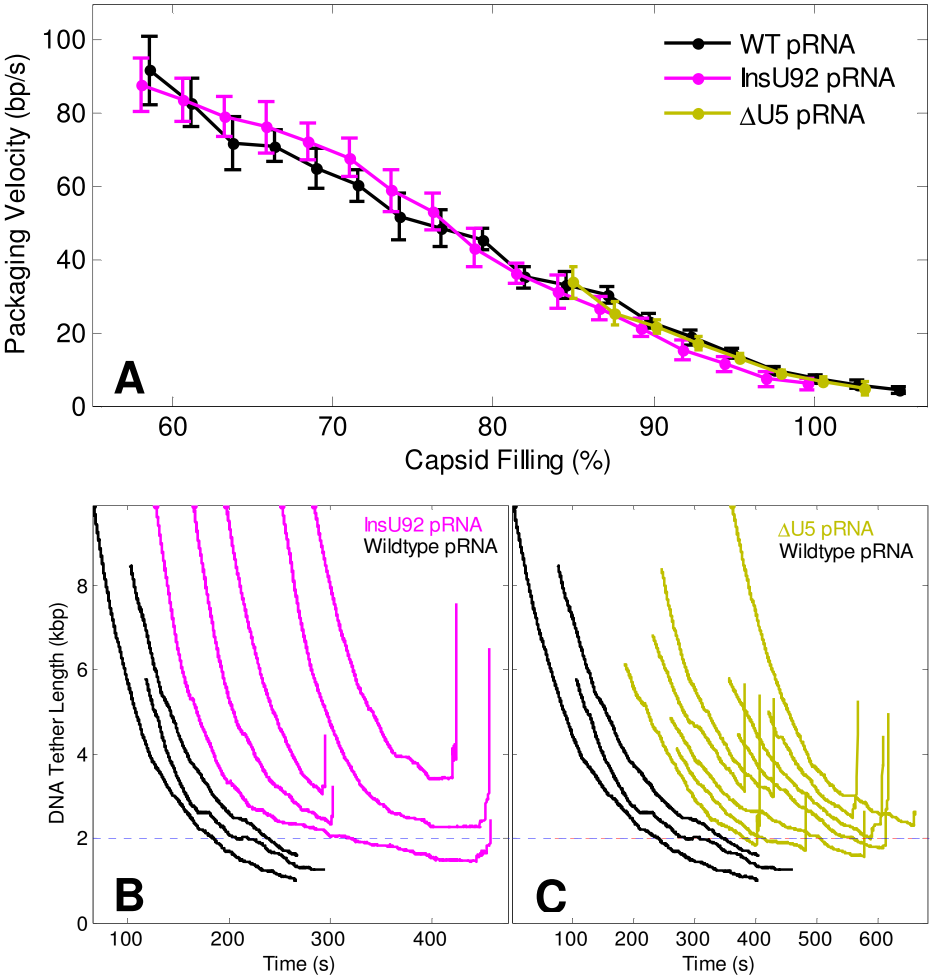


Figure 5.3.5: The Effect of pRNA Mutations on DNA Packaging Dynamics at High Filling
(A) Pause-free DNA packaging velocity versus capsid filling. Error-bars indicate the 95% confidence interval. (B) Sample packaging traces of U92ins homo-pentameric pRNA mutant (magenta) and wildtype (black). (C) Sample packaging traces of ΔU5 homo-pentameric pRNA mutant (green) and wildtype (black). In panels (B) and (C) the dashed horizontal line indicates the 100% capsid filling mark.

The U92ins mutation introduces an extra base at the hinge between the A-helix and the D-loop of the Phi29 pRNA (Figure 5.3.1, left panel). Ensemble experiments show that DNA packaging efficiency and gp16 binding to the prohead is reduced in complexes with homo-pentameric U92ins pRNA (Zhao et al, 2012). My single-molecule screens of this homo-pentameric mutant reveal no significant differences from wildtype pRNA at low capsid filling and high [ATP] (Figure 5.3.3). Surprisingly, at low ATP concentration (~K_M) the U92ins mutant exhibited significantly more slips than wildtype, especially in the initial ~500-bp portion of the packaging trace (Figure 5.3.4, right panel). Some of these initial slips lead to tether rupture, resulting in low experiment throughputs. However, the slipping observed in my optical tweezers experiments is unlikely to cause the low packaging efficiency observed in biochemical assays because there is no external force acting on the DNA in ensemble experiments. My results suggest that the low DNA packaging efficiency observed in bulk is due to an assembly and initiation defect caused by the U92ins mutation. Single-molecule packaging screens with a 21-kb DNA tether revealed that U92ins mutant particles are unable to contain the DNA within the capsid at near 100% filling, and undergo fatal slips (Figure 5.4.5B) just like the ΔU5 mutant. Note that the pause-free velocity of U92ins and ΔU5 mutants is essentially indistinguishable from that of wildtype packaging complexes (Figure 5.4.5A), indicating that the mutations do not affect the force-generation capacity of the motor. Future experiments with pRNA mutants will investigate the role of pRNA in throttling down the gp16 ATPase operation at high capsid filling.

5.4 Building a Low-Cost Syringe Pump

Single-molecule optical tweezers experiments require the use of microfluidics chambers with 2-3 channels, with a different buffer flowing through each channel. The traditional method of controlling buffer flow is to have the fluids in different 1-mL syringes and manually push ~50-100 μL of liquid at a time. Over the years various members of our lab designed a few different automated flow control systems. One system relied on Clippard air valves to control the pressure of gas acting on different syringes, but it was too complex and cumbersome for my experiments. Another system used 3 Hamilton valves (~$700/each) to control the gravity driven flow of buffers. In addition to being expensive and complex, this system was prone to bacterial contamination. A different flow-control system employed syringe pumps that cost several hundreds of dollars.

Figure 5.4.1 Low-Cost Syringe Pump Overview
The system consists of three syringes placed on a white foam-board platform. Each syringe is driven by one actuator (silver) mounted onto the platform. The assembly is placed on a Plexiglas sheet to prevent buffer spills onto the optical table. A solenoid pinch valve (upper left corner) is used to block any residual flow from the middle syringe to the central channel of the microfluidics chamber.

I set out to build a custom syringe pump system using affordable linear actuators from Firgelli. These actuators come in a variety of models, but I chose the L12-S model (~$80/each) powered by 12 VDC, with a gear ratio of 210:1 (the slowest model), and 100 mm of travel (sufficient for 1 mL syringes). Other actuator models can be controlled from a computer or a dedicated controller, but they cost extra. The L12-S model has a built in min/max limiter switch that prevents damage to the motor when it reaches the extremities of its travel path. The actuator extends when a positive voltage is applied, and it retracts when a negative voltage is applied. The motor speed can be controlled by varying the voltage between 0 and 12V.

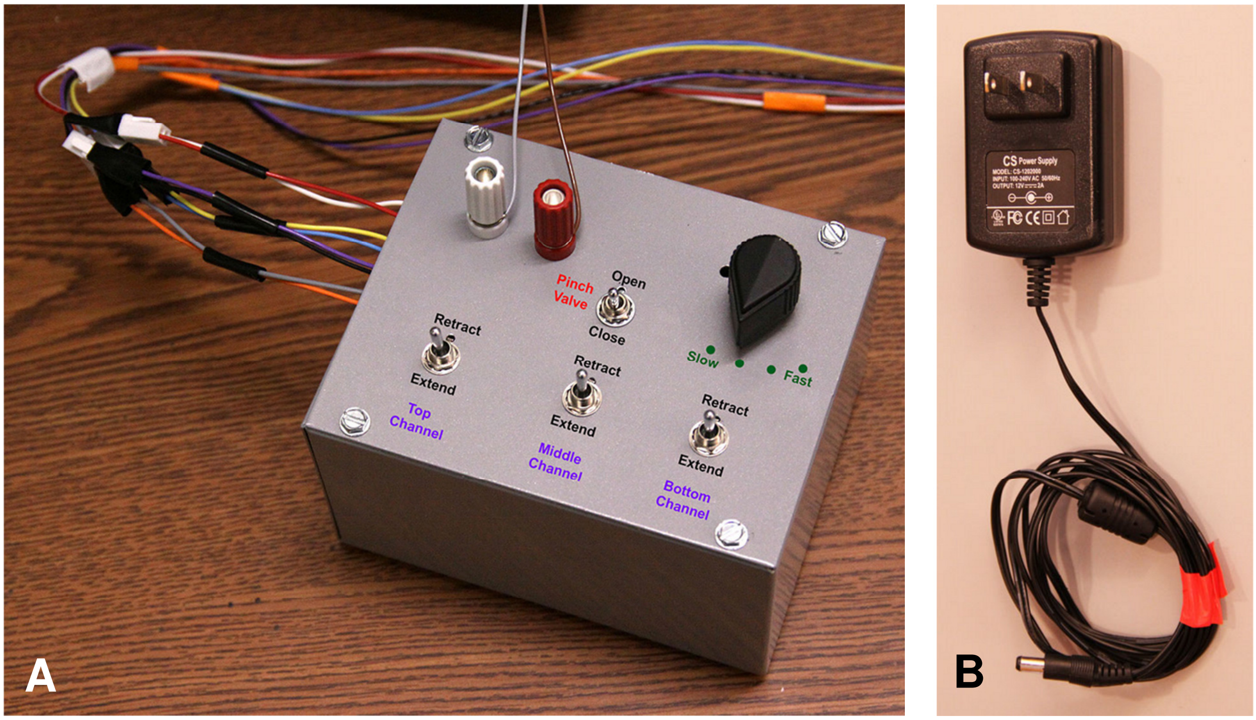


Figure 5.4.2: Analog Controller for the Low-Cost Syringe Pump System
Left: Three dual polarity – dual position switches control the movement direction for each actuator (blue labels). The speed of all motors is controlled by a potentiometer (black knob, green labels). The state of the pinch valve is controlled by a dual position toggle switch (red label). Both the pinch valve and actuators require 12V to operate, so a single power source is needed (white and red banana plugs). Right: A cheap AC-DC converter can be used to power the syringe pump system, with a voltage output of 12V and a current up to 0.5 A.

The low-cost syringe pump assembly consists of 3 linear actuators mounted on a foam-board platform for syringes (Figure 5.4.1). The motors are powered and operated via an analog controller box (Figure 5.4.2A). The extension/retraction speed of all three actuators is controlled by a 50 Ω potentiometer (Figure 5.4.2A, black knob) that reduces the voltage supplied to all actuators to a fraction of the original 12V. At their highest setting, the motors extend by ~5-10 mm/s, but they can be operated at speeds as low as ~1 mm/s. Each actuator is operated using a dual polarity - dual position toggle switch (Figure 5.4.2A, blue labels). Power can be provided by a cheap AC-DC converter (Figure 5.4.2B). Note that this system does not allow fine control over buffer flow-rates; instead it is used to flush the microfluidics chamber with fresh buffer at 10-15 minute intervals. I added a solenoid pinch valve (Cole-Palmer, YO-98302-00, ~$70/each) that prevents any residual flow in the central channel, where optical trapping takes place (Figure

5.4.1 upper left corner). The pinch valve opens when a 12 VDC voltage is applied to its terminals, but is otherwise closed in the absence of any voltage. The valve is operated by a dual-position toggle switch (Figure 5.4.2A, red label). In conclusion, this system can be built quickly and cheaply, does not require a computer to run, and is suitable for simple buffer-exchange applications, but lacks fine control over buffer flow-rates.

5.5 Building a Low-Cost Gel Documentation System

Gel documentation (Gel Doc) systems are used to image agarose or PAGE gels on UV/VIS trans-illuminators. Gel Doc systems consist of a camera and lens for image acquisition; a computer and display for control and processing; and a software suite for quantifying or adjusting the images. Some commercial Gel Docs have a built-in thermal printer. The heart of any commercial gel doc system is the camera, usually with a small CCD or CMOS sensor and an imaging lens. The major advantage of a commercial Gel Doc is the integration of image acquisition with image processing and quantification. I built a custom Gel Documentation system for our lab and published the design online (Chistol, 2011). Several biology labs around the world have successfully built their own Gel Doc using this design. The system consists of: (1) Camera and lens, (2) Camera enclosure, (3) Computer, (4) Software.

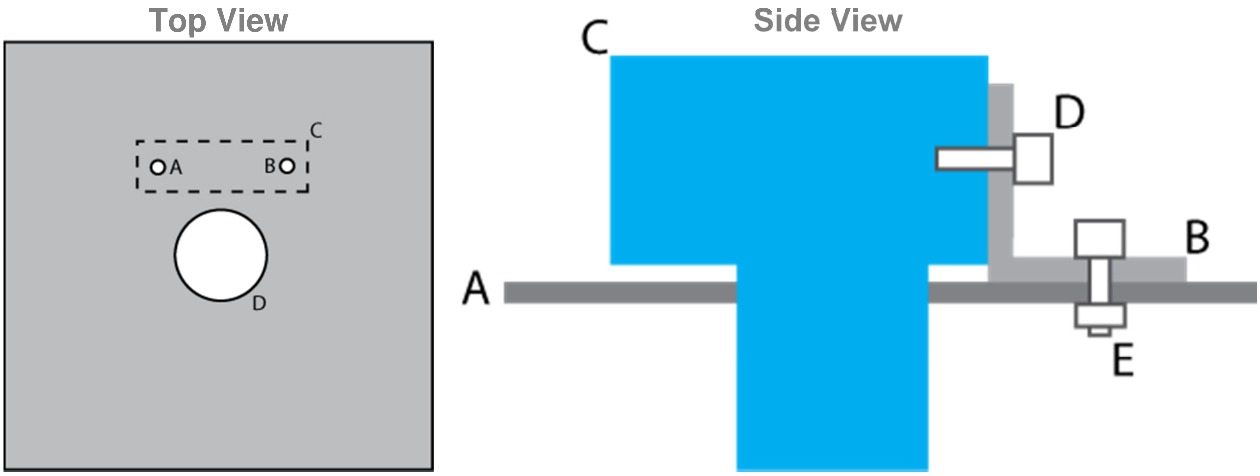


Figure 5.5.1: Camera Mounting Diagram for the Low-Cost Gel Doc
(Top View) A and B are holes drilled for screws that hold the L-bracket, C is the footprint of the L-bracket, and D is the cutout for the lens (73 mm in diameter). (Side View) A is the top portion of the Gel Doc enclosure, B is L-bracket (10-15cm long), C represents the dSLR camera, D is the screw through the camera tripod socket, and E represents the two screws that attach the L-bracket to the enclosure (through holes A and B in the Top View diagram).

(1) Camera and Lens. All modern dSLR cameras have sensors that are several times larger than point-and-shoot cameras and are therefore well suited for a Gel Doc. As of 2013, only Canon ships their dSLRs with free remote capture software that allows the user to control the camera from a computer via a USB cable. The Gel Doc camera must have a live-view feature which allows the user to adjust focus/exposure from the computer while viewing a real-time video stream from the camera (Canon Rebel XS, XSi, T1i, T2i, T3i, T3, etc.). The cheapest live-view enabled dSLRs costs ~$500 new with a kit lens, but used or refurbished models may cost as little as ~$200. The 18-55 mm kit lens is sufficient for all but the most demanding users. Given that ethidium bromide and SYBR safe absorb deep blue/UV light and emit orange/red light, a

color filter is needed to block short-wavelength light. The most convenient commercially available filters are Tiffen Orange 16 or Tiffen Orange 21 glass filters (~$20 new for a 58 mm-diameter filter that fits the 18-55 mm Canon kit lens). In addition, an AC-DC power supply (~$60) is needed to power the camera continuously, and long USB cable (~$10) is needed to connect the camera to the computer.

(2) The camera enclosure/platform can be as simple as a thick cardboard box which supports the camera. A metal, wood or Plexiglas enclosure is recommended for durability. The Chemistry department machine shop manufactured our enclosure using 5 mm-thick black Plexiglas. The enclosure was designed to fit on top of the SYBR Safe transilluminator, UV transilluminator, and a visible light table (for imaging Coomassie and Silver stained gels). The enclosure is a cube measuring 30 cm on each side. The top surface has a circular cutout for the lens measuring 73 mm in diameter (Figure 5.5.1).

(3) The computer can be either a PC or Mac with very basic specifications, so a low cost machine should suffice. It is recommended to always wear gloves when touching the keyboard and mouse as they will be contaminated with trace amounts of ethidium bromide – a known carcinogen.

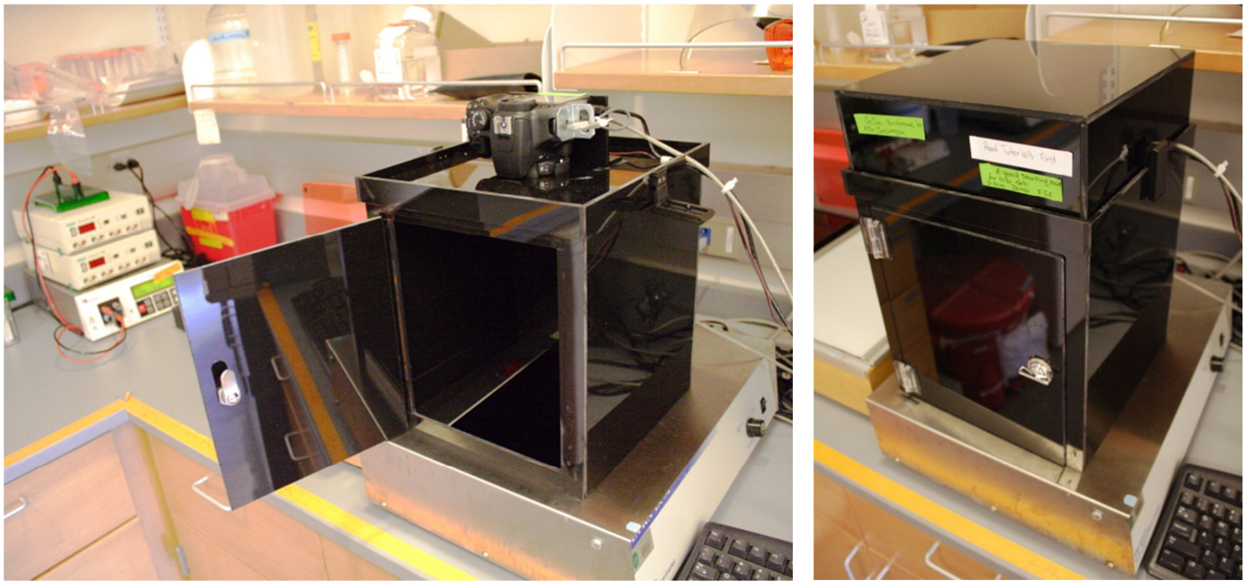


Figure 5.5.2: Overview of the Low-Cost Gel Doc

Left: the Gel Doc unit placed on a UV transilluminator with its top cover removed and the front door open for easy access to the gel. Right: the Gel Doc with its top cover on. The cover protects the camera against accidental damage and prevents light leaks from through the camera viewfinder and the cutout for the lens. The top cover is removed very rarely, only to reset the camera in case of a software error.

(4) The software for the Gel Doc consists of Canon EOS Utilities for image capture and ImageJ for image processing and gel band quantification. EOS Utilities, which comes bundled with the camera, provides a live-view feed and allows the user to control exposure and focus. The camera should be set in Manual mode, black and white color mode, with no sharpening or contrast enhancement. The focus mode should be set to “Manual” on the lens barrel, which then enables the user to adjust focus from the computer alone. Optimal focusing is achieved by ensuring that the gel edges are sharp in the live-view feed. Before imaging a gel, the Gel Doc door should be closed and the UV transilluminator turned on. Optimal exposure settings may

vary depending on the gel, but the following parameters can be used as a starting point: shutter speed of ~1 sec, lens aperture of ~f/5.6, camera ISO of ~400. If the gel bands are faint, the camera sensitivity (ISO) can be increased to 1600 or 3200, and the shutter speed can then be increased to several seconds. Overexposure should be avoided because it leads to the loss of information about the brightest bands. The low-cost Gel Doc does not have a strictly linear response, but its images can be used for quantification if known quantities of DNA are loaded on the gel as controls. Any image manipulation software such as ImageJ, Adobe Photoshop, Adobe Photoshop Elements, GIMP, Paint.NET can be used to invert gel images and adjust grayscale levels. ImageJ can be used to select the bands and quantify their intensity.

In conclusion, this Gel Doc system is convenient, relatively cheap, simple to construct, and should last a long time. The dSLR camera has a life expectancy of 50-100k shutter actuations, and a well-made enclosure will outlast the camera. Commercially available Gel Doc systems are more convenient due to software integration, but the homebuilt Gel Doc produces high-quality gel images (Figure 5.3.3) for a lot less money.

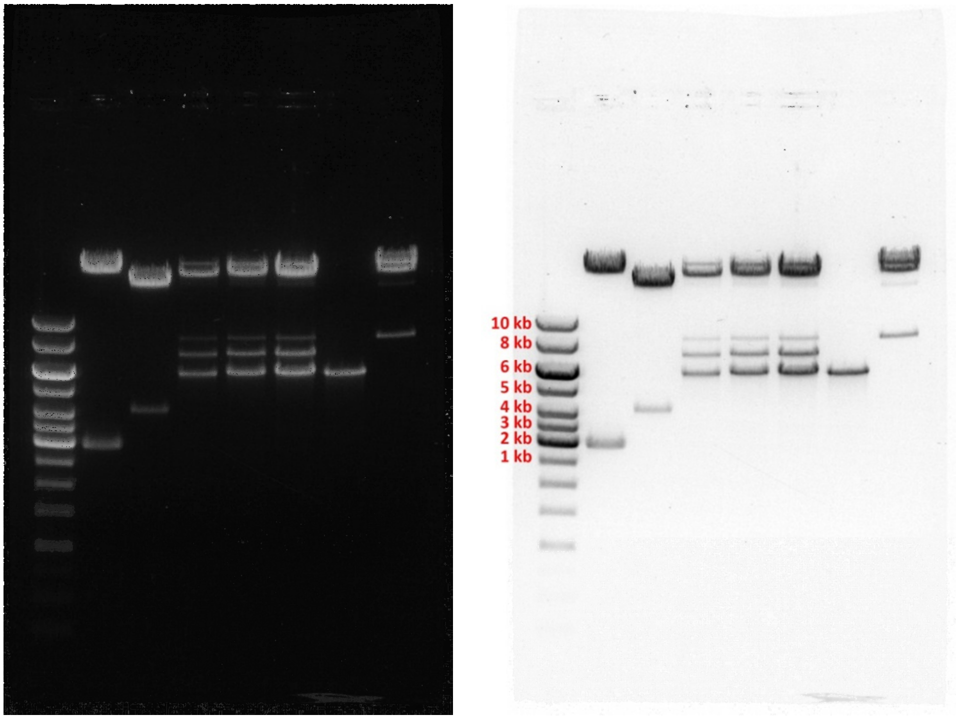


Figure 5.5.3: Sample Gel Image Acquired on the Low-Cost Gel Doc
0.5% Agarose Gel and 1xTAE imaged with the low-cost Gel Doc. The original image (left) was inverted (right) in Adobe Photoshop.

5.6 Concluding Thoughts

It is remarkable how much we have learned about the operation of the Phi29 packaging motor from single-molecule optical trapping experiments. Imagine studying the operation of an internal combustion engine simply by watching a truck drive in various conditions without ever looking under the hood. To test the power of the truck's engine you may drive the truck on a progressively steeper hill until it no longer advances. This is exactly what was done in the earliest single-molecule experiments with Phi29 – the packaging motor was given DNA to translocate under an increasing external opposing force until the motor stalled (Smith et al., 2001). To understand how the internal combustion engine generates power from its fuel, you may drive a truck fueled with low-grade gasoline, high-grade gasoline, alcohol-gasoline mixtures, diesel, or perhaps more exotic types of fuel. In the case of the Phi29 motor, its operation was monitored at different fuel concentration ([ATP]), and exhaust product concentrations ([ADP] and [Pi]) (Chemla et al., 2005). To study the factors that influence the traction between the truck's wheels and the ground, you may test-drive the truck on mud, gravel, sand, pavement, wet road, snow, etc. To understand how the Phi29 packaging motor interacts with its track (DNA), the motor was given a variety of substrates to translocate: double-stranded DNA, DNA with a neutral backbone, single-stranded DNA, nicked DNA, and even artificial polymers that the motor would never encounter in nature (Aathavan et al., 2009).

It is probably becoming very obvious that I am stretching my analogy a bit too far. In fact the internal combustion engine and truck analogy is essentially useless if I were to describe the projects that I worked on. By the time I started studying the Phi29 packaging motor, a great deal was already known about this system. It was thought that a huge pressure built up inside the viral capsid during packaging, which eventually prevented any more DNA from being stuffed into the prohead (Smith et al., 2001). It had been shown that each motor cycle consisted of a stationary dwell phase and a translocation burst phase (Moffitt et al., 2009). A previous study had determined that the motor "fueled up" with ATP during the dwell phase, and that only four of its five subunits actively translocated DNA during the burst phase (Moffitt et al., 2009). It was shown that the burst was 10-bp in size and that the motor made specific interactions with the negative charges on the DNA backbone every 10 bp (Moffitt et al., 2009, Aathavan et al., 2009). Together with previous work, the findings presented in this dissertation provide the most detailed characterization of a homomeric ring motor to date.

In our first study, described in Chapter 3, we investigated the timing and coordination of ADP release and ATP hydrolysis within the operational cycle of the Phi29 ring ATPase. We discovered that ADP release is interlaced with ATP binding during the dwell phase, and that ADP release is rate limiting at saturating ATP and low external force (Chistol et al., 2012). We found that the four DNA-translocating subunits hydrolyze ATP during the burst phase, immediately before the power-stroke. We provided evidence suggesting that the non-translocating subunit plays a critical regulatory role in each dwell-burst cycle. Given the geometry of the dsDNA helix, and the burst size of the motor, the identity of this special subunit is most likely preserved several cycles in a row. This is the first known example of a homomeric ring ATPase where a division of labor arises among subunits.

In our second study, described in Chapter 4, we investigated how the Phi29 motor packages DNA into the viral capsid at very high filling. We discovered that Phi29 employs several mechanisms to throttle down its packaging motor as the viral capsid fills up with DNA. First, the dwell phase is gradually lengthened, owing to lower ATP tight-binding rates at higher filling. Second, the motor entered into long lived pauses at capsid fillings above ~80%, and these pauses became longer and more frequent at higher filling, indicating that capsid filling can regulate packaging via an on/off switch. Third, the burst duration became longer at higher filling, indicating that an opposing force acted though DNA on the packaging motor from within the capsid. Finally, the burst size was gradually reduced from 10 bp at low filling to approximately 9 bp at very high filling – the first observation of step-size modulation by an external factor in a ring motor. Our burst duration measurements indicate that the opposing internal force at 100% capsid filling does not exceed ~20 pN, a value 3-5x lower than previously estimated (Smith et al., 2001; Rickgauer et al., 2008). The existence of a filling-dependent on/off switch could explain how viruses that package concatenated genomes can activate a nuclease domain that cuts the DNA only when the capsid is full (Lander et al., 2006). The signal that forces the motor to enter a long lived pause could also double as the signal that controls the nuclease domain. The surprising finding that burst size decreases slightly at high filling, together with DNA rotation measurements (Hetherington, private communication) suggest that the Phi29 ATPase symmetry is broken by a unique set of motor-DNA contacts, and that these contacts determine the identity of the special subunit.

Although we have achieved a remarkably detailed characterization of the Phi29 motor operation, our study of this system is far from over. Several questions remain to be addressed in future research. What are the structural elements responsible for inter-subunit coordination? Why is the Phi29 ATPase so tightly coordinated? How does this compare to other ring ATPases? What is the exact role of the special subunit? How is information about capsid filling communicated to the ATPase through the connector and pRNA?

Having spent most of the last five years worrying about all the little things that make the experiments “just work” I now find it very refreshing to look back at our findings and the path we took to get there. More than ever I am amazed by the complexity and elegance of the tiny molecular machinery that makes life possible.

References

- Aathavan, K., Politzer, A.T., Kaplan, A., Moffitt, J.R., Chemla, Y.R., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2009). Substrate interactions and promiscuity in a viral DNA packaging motor. Nature 461, 669-673
- Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 A of F1-ATPase from bovine heart mitochondria. Nature 370, 621-628
- Adachi, K., Oiwa, K., Nishizaka, T., Furuike, S., Noji, H., Itoh, H., Yoshida, M., and Kinosita, K., Jr. (2007). Coupling of rotation and catalysis in F1-ATPase revealed by single-molecule imaging and manipulation. Cell 130, 309-321
- Adelman, J.L., Jeong, Y.J., Liao, J.C., Patel, G., Kim, D.E., Oster, G., and Patel, S.S. (2006). Mechanochemistry of transcription termination factor Rho. Mol Cell 22, 611-621
- Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Research, (1997) 25 (17)
- Arunajadai, S.G., and Cheng, W., Step Detection in Single-Molecule Real Time Trajectories embedded in Correlated Noise, PLOS One (2013), Vol. 8, Issue 3
- Ashkin A., and Dziedzic, J. M., Observation of light scattering from nonspherical particles using optical levitation, Applied Optics (1980), Vol. 19, No. 5
- Ashkin, A., et al., Observation of a Single-Beam Gradient Force Optical Trap for Dielectric Particles, Optics Letters (1986), Vol. 11, No. 5
- Ashkin, A., et al., History of Optical Trapping and Manipulation of Small-Neutral Particle, Atoms, and Molecules, IEEE Journal on Selected Topics in Quantum Electronics (2000), Vol. 6, No. 6
- Aubin-Tam M., et al., Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine, Cell (2011), Vol. 145, 257-267
- Baird, C.L., Harkins, T.T., Morris, S.K., and Lindsley, J.E. (1999). Topoisomerase II drives DNA transport by hydrolyzing one ATP. Proc Natl Acad Sci USA 96, 13685-13690
- Barry, E.R., et al., Intersubunit Allosteric Communication Mediated by a Conserved Loop in the MCM Helicase, PNAS (2008), Vol. 106, No. 4, 1051-1056
- Bazinet, C., and King, J., The DNA Translocating Vertex of dsDNA Bacteriophage; Ann. Rev. Microbiol. (1985), 39:109-29
- Berg-Sorensen, K., and Flyvbjerg, H. (2004). Power spectrum analysis for optical tweezers. Rev Sci Instrum 75, 594-612
- Black, L. W., DNA Packaging in dsDNA Bacteriophages, Ann. Rev. Microbiol. (1989), 43:267-92
- Block, S.M., et al., Bead Movement by Single Kinesin Molecules Studies with Optical Tweezers, Nature (1990), Vol. 348
- Block, S. M. and Svoboda, K., Analysis of High Resolution Recordings of Motor Movement, Biophysical Journal (1995), Vol. 68, 230s-241s
- Bochman, M.L., Bell, S.P., and Schwacha, A. (2008). Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. Mol Cell Biol 28, 5865-5873
- Boyer, P. D., The ATP Synthase – a Splendid Molecular Machine, Annu. Rev. Biochem. (1997), 66:717-49

- 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
- Bustamante C., et al., Ten Years of Tension: Single-Molecule DNA Mechanics, Nature (2003), Vol. 421
- Bustamante, C., Chemla, Y.R., Forde, N.R., and Izhaky, D. (2004). Mechanical processes in biochemistry. Annu Rev Biochem 73, 705-748
- Bustamante, C., Chemla, Y.R., and Moffitt, J.R. (2008). High-resolution dual-trap optical tweezers with differential detection. In Single-Molecule Techniques: A Laboratory Manual, P.R. Selvin, and T. Ha, eds. (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press), pp. 297-324
- Carter, A.P., Cho, C., Jin, L., and Vale, R.D. (2011). Crystal structure of the dynein motor domain. Science 331, 1159-1165
- Casjens, S.R. (2011). The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol 9, 647-657
- Chemla, Y.R., Aathavan, K., Michaelis, J., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2005). Mechanism of force generation of a viral DNA packaging motor. Cell 122, 683-692
- Chistol, G. et al. High degree of coordination and division of labor among subunits in a homomeric ring ATPase. Cell 151, 1017-1028 (2012)
- Chistol, G., DIY Gel Imaging System, Lab Times (2011), 64-65
- Chivers, C.E., et al., A Streptavidin Variant with Slower Biotin Dissociation and Increased Mechanostability, Nature Methods (2010), Vol. 7, No. 5
- Crampton, D.J., Mukherjee, S., and Richardson, C.C. (2006). DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. Mol Cell 21, 165-174
- Costa, A., et al., The Structural Basis for MCM2-7 Helicase Activation by GINS and Cdc45, Nature Structural and Molecular Biology (2011), Vol., No. 4
- Crozat, E., et al., Separating Speed and Ability to Displace Roadblocks During DNA Translocation by FtsK, The EMBO Journal (2010), Vol. 29, 1423-1433
- Ding, F., et al., Structure and Assembly of the Essential RNA Ring Component of a Viral DNA Packaging Motor, PNAS (2010), Vol 108
- Duderstadt, K.E., Chuang, K., and Berger, J.M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. Nature 478, 209-213
- Earnshaw, W.C. & Harrison, S.C. DNA arrangement in isometric phage heads. Nature 268, 598-602 (1977)
- Efron, B., Bootstrap Methods: Another Look at the Jackknife, The 1977 Rietz Lecture, The Annals of Statistics (1979), Vol. 7, No. 1, 1-26
- Enemark, E.J., and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. Nature 442, 270-275
- Erzberger, J.P., and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct 35, 93-114
- Evilevitch, A., Lavelle, L., Knobler, C.M., Raspaud, E. & Gelbart, W.M. Osmotic pressure inhibition of DNA ejection from phage. Proc. Natl. Acad. Sci. USA 100, 9292-9295 (2003)

- Evilevitch, A., Castelnovo, M., Knobler, C.M. & Gelbart, W.M. Measuring the force ejecting DNA from phage. J. Phys. Chem. B 108, 6838-6843 (2004)
- Fang, H., Jing, P., Hague, 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)
- Forrey, C. & Muthukumar, M. Langevin dynamics simulations of genome packing in bacteriophage. Biophys. J. 91, 25-41 (2006)
- Fuller, D.N. et al. Measurements of single DNA molecule packaging dynamics in bacteriophage lamda reveal high forces, high motor processivity, and capsid transformations. J. Mol. Biol. 373, 1113-1122 (2007)
- Fuller, D.N. et al. Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi29. Proc. Natl. Acad. Sci. USA 104, 11245-11250 (2007)
- Gai, D., Zhao, R., Li, D., Finkielstein, C.V., and Chen, X.S. (2004). Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. Cell 119, 47-60
- Gelbart, W.M. & Knobler, C.M. Pressurized viruses. Science 323, 1682-1683 (2009)
- Glynn, S.E., Martin, A., Nager, A.R., Baker, T.A., and Sauer, R.T. (2009). Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. Cell 139, 744-756
- Gonzalez-Huici, V., Salas, M. & Hermoso, J.M. The push-pull mechanism of bacteriophage phi29 DNA injection. Mol. Microbiol. 52, 529-540 (2004)
- Goodno, C.C. (1982). Myosin active-site trapping with vanadate ion. Methods Enzymol 85, 116-123
- Grimes, S., Ma, S., Gao, J., Atz, R. & Jardine, P.J. Role of Phi29 connector channel loops in late-stage DNA packaging. J. Mol. Biol. 410, 50-59 (2011)
- Guo, P., Peterson, C., and Anderson, D. (1987). Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage phi29. J Mol Biol 197, 229-236
- Harjes, E., et al., Structure of the RNA Claw of the DNA Packaging Motor of Bacteriophage Phi29, Nucleic Acids Research (2012), Vol. 40, No. 19, 9953-9963
- Hetherington, C.L., Moffitt, J.R., Jardine, P.J., and Bustamante, C. (2012). Viral DNA packaging motors. In Comprehensive Biophysics, E. Egelman, ed. (Academic Press), pp. 420-446
- Hodges, C., et al., Nucleosomal Fluctuations Govern the Transcription Dynamics of RNA Polymerase II, Science (2009), Vol. 325, No. 5940, 626-628
- Howard, J., Hudspeth, A.J. & Vale, R.D. Movement of microtubules by single kinesin molecules. Nature 342, 154-158 (1989)
- Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol Cell 37, 247-258
- Itoh, H., et al., Mechanically Driven ATP Synthesis by F1-ATPase, Nature (2004), Vol. 427
- Iyer, L.M., Leipe, D.D., Koonin, E.V., and Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 146, 11-31
- Iyer, L.M., et al., 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 Research (2004), Vol. 32, No. 17, 5260-5279

- Johnson, A., and O'Donnell, M. (2003). Ordered ATP hydrolysis in the gamma complex clamp loader AAA+ machine. J Biol Chem 278, 14406-14413
- Kainov, D.E., Mancini, E.J., Telenius, J., Lisal, J., Grimes, J.M., Bamford, D.H., Stuart, D.I., and Tuma, R. (2008). Structural basis of mechanochemical coupling in a hexameric molecular motor. J Biol Chem 283, 3607-3617
- Kalafut, B., and Visscher, K. (2008). An objective, model-independent method for detection of non-uniform steps in noisy signals. Comput Phys Commun 179, 716-723
- Katoh K., Toh H., Recent developments in the MAFFT multiple sequence alignment program, Briefings in Bioinformatics (2008), 9 (4)
- Keller, D. & Bustamante, C. The mechanochemistry of molecular motors. Biophys. J. 78, 541-556 (2000)
- Kelley L.A., Sternberg M.J.E., Protein structure prediction on the web: a case study using the Phyre server, Nature Protocols (2009), 4
- Kerssemakers, J.W.J., et al., Assembly Dynamics of Microtubules at Molecular Resolution, Nature (2006), Vol. 442
- Kottadiel, V.I., et al., The Dynamic Pause-Unpackaging State, an Off-Translocation Recovery State of a DNA Packaging Motor from Bacteriophage T4, PNAS (2012), Vol. 109, No. 49, 20000-20005
- Lander, G.C. et al. The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science 312, 1791-1795 (2006)
- Lee, J.Y., et al., Single-Molecule Imaging of DNA Curtains Reveals Mechanisms of KOPS Sequence Targeting by the DNA Translocase FtsK, PNAS (2012), Vol. 109, No. 17
- Leforestier, A. & Livolant, F. The bacteriophage genome undergoes a succession of intracapsid phase transitions upon DNA ejection. J. Mol. Biol. 396, 384-395 (2010)
- Lyubimov, A.Y., Strycharska, M., and Berger, J.M. (2011). The nuts and bolts of ring-translocase structure and mechanism. Curr Opin Struct Biol 21, 240-248
- Maillard, R. A. et al., ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates, Cell (2011), Vol. 145, 459-469
- Mancini, E.J., Kainov, D.E., Grimes, J.M., Tuma, R., Bamford, D.H., and Stuart, D.I. (2004). Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. Cell 118, 743-755
- Martin, A., Baker, T.A., and Sauer, R.T. (2005). Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. Nature 437, 1115-1120
- Martin, A., et al., Pore Loops of the AAA+ ClpX Machine Grip Substrates to Drive Translocation and Unfolding, Nature Structural & Molecular Biology (2008), Vol. 15, No. 11
- Massey, T.H., Mercogliano, C.P., Yates, J., Sherratt, D.J., and Lowe, J. (2006). Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. Mol Cell 23, 457-469
- Mehta, A.D. et al. Myosin-V is a processive actin-based motor. Nature 400, 590-593 (1999)
- Moffitt, J.R., et al., Differential Detection of Dual Traps Improves the Spatial Resolution of Optical Tweezers, PNAS (2006), Vol. 103, No. 24, 9006-9011
- Moffitt, J.R., Chemla, Y.R., Aathavan, K., Grimes, S., Jardine, P.J., Anderson, D.L., and Bustamante, C. (2009). Intersubunit coordination in a homomeric ring ATPase. Nature 457, 446-450

- Moffitt, J.R., et al., Viral DNA Packaging at Base Pair Resolution, University of California Berkeley (2009), PhD Dissertation
- Moffitt, J.R., Chemla, Y.R., and Bustamante, C. (2010). Methods in statistical kinetics. Methods Enzymol 475, 221-257
- Moffitt, J.R., Chemla, Y.R., and Bustamante, C. (2010). Mechanistic constraints from the substrate concentration dependence of enzymatic fluctuations. Proc Natl Acad Sci USA 107, 15739-15744
- Molineux, I.J. & Panja, D. Popping the cork: mechanisms of phage genome ejection. Nat. Rev. Microbiol. 11, 194-204 (2013)
- 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
- Morais, M.C. The dsDNA packaging motor in bacteriophage Phi29. in Viral Molecular Machines (eds. Rossmann, M.G. & Rao, V.B.) 511-547 (Springer, 2012)
- Moreau, M.J., McGeoch, A.T., Lowe, A.R., Itzhaki, L.S., and Bell, S.D. (2007). ATPase site architecture and helicase mechanism of an archaeal MCM. Mol Cell 28, 304-314
- Mott, M.L., Erzberger, J.P., Coons, M.M., and Berger, J.M. (2008). Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. Cell 135, 623-634
- Murialdo, H., Bacteriophage Lambda DNA Maturation and Packaging, Annu. Rev. Biochem. (1991), 60:125-53
- Neuman, K. C. and Block, S. M., Optical Trapping, Review of Scientific Instruments (2004), Vol. 75, No. 9
- Nurmemmedov, E., Castelnovo, M., Catalano, C.E. & Evilevitch, A. Biophysics of viral infectivity: matching genome length with capsid size. Q. Rev. Biophys. 40, 327-356 (2007)
- Oster, G. & Wang, H. Reverse engineering a protein: the Mechano-Chemistry of ATP synthase. Biochim. Biophys. Acta 1458, 482-510 (2000)
- Parzen, E., On Estimation of a Probability Density Function and Mode, The Annals of Mathematical Statistics (1961), 1065-1076
- Petrov, A.S. & Harvey, S.C. Structural and thermodynamic principles of viral packaging. Structure 15, 21-27 (2007)
- Pralle, A., et al., Three-Dimensional High-Resolution Particle Tracking for Optical Tweezers by Forward Scattered Light, Microscopy Research and Technique (1999), 44:378-386
- Purohit, P.K. et al. Forces during bacteriophage DNA packaging and ejection. Biophys. J. 88, 851-866 (2005)
- Rao, V.B., and Feiss, M. (2008). The bacteriophage DNA packaging motor. Annu Rev Genet 42, 647-681
- Rao, V.B., and Black, L.W., DNA Packaging in Bacteriophage T4, Madam Curie Bioscience Database, NCBI Bookshelf
- Rickgauer, J.P. et al. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage phi29. Biophys. J. 94, 159-167 (2008)
- Riemer, S.C. & Bloomfield, V.A. Packaging of DNA in bacteriophage heads: some considerations on energetics. Biopolymers 17, 785-794 (1978)
- Rosenblatt, M., Remarks on Some Nonparametric Estimates of a Density Function, The Annals of Mathematical Statistics (1955), 832-837

- Saleh, O. A., et al., Fast, DNA-Sequence Independent Translocation by FtsK in a Single-Molecule Experiment, The EMBO Journal (2004), 2430-2439
- Satapathy, A.K., et al., Coupling dTTP Hydrolysis with DNA Unwinding by the DNA Helicase of Bacteriophage T7, Journal of Biological Chemistry (2011), Vol. 286, No. 39, 34468-34478
- Schnitzer, M.J., and Block, S.M. (1995). Statistical kinetics of processive enzymes. Cold Spring Harb Symp Quant Biol 60, 793-802
- Segel, I.H. (1975). Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems (New York, John Wiley & Sons, Inc.)
- Sharma, S., and Davidson, A.L. (2000). Vanadate-induced trapping of nucleotides by purified maltose transport complex requires ATP hydrolysis. J Bacteriol 182, 6570-6576
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem 76, 23-50
- Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell 101, 589-600
- Schutz P., Wahlberg E., Karlberg T., Hammarstrom M., Collins R., Flores A., Schuler H.; Crystal structure of human RNA helicase A (DHX9): structural basis for unselective nucleotide base binding in a DEAD-box variant protein, Journal of Molecular Biology (2010), 400
- Skordalakes, E., and Berger, J.M. (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell 114, 135-146
- Smith, D.E. et al. The bacteriophage phi29 portal motor can package DNA against a large internal force. Nature 413, 748-752 (2001)
- Spakowitz, A.J. & Wang, Z.G. DNA packaging in bacteriophage: is twist important? Biophys. J. 88, 3912-3923 (2005)
- Spudich, J. A., et al., Optical Traps to Study Properties of Molecular Motors, Cold Spring Harbor Protoc. (2011), doi:10.1101/pdb.top066662
- 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
- Sun, S., Rao, V.B. & Rossmann, M.G. Genome packaging in viruses. Curr. Opin. Struct. Biol. 20, 114-120 (2010)
- Svoboda, K., and Block, S. M., Force and Velocity Measured for Single Kinesin Molecules, Cell (1994), Vol 77, 773-784
- Tavares, P., Zinn-Justin, S. & Orlova, E.V. Genome gating in tailed bacteriophage capsids. in Viral Molecular Machines (eds. Rossmann, M.G. & Rao, V.B.) 585-600 (Springer, 2012)
- Thomsen, N.D., and Berger, J.M. (2009). Running in reverse: the structural basis for translocation polarity in hexameric helicases. Cell 139, 523-534
- Tolić-Nørrelykke, I. M., MatLab Program for Precision Calibration of Optical Tweezers, Computer Physics Communications (2004), Vol. 159, 225-240
- Tsay, J.M., et al., The Q Motif of a Viral Packaging Motor Governs its Force Generation and Communicates ATP Recognition to DNA Interaction, PNAS (2009), Vol. 106, No. 34, 14355-14360

- Tzlil, S., Kindt, J.T., Gelbart, W.M. & Ben-Shaul, A. Forces and pressures in DNA packaging and release from viral capsids. Biophys. J. 84, 1616-1627 (2003)
- Uchihashi, T., et al., High-Speed Atomic Force Microscopy Reveals Rotary Catalysis of Rotorless F1-ATPase, Science (2011), Vol. 333
- Van Valen, D. et al. A single-molecule Hershey-Chase experiment. Curr. Biol. 22, 1339-1343 (2012)
- Visscher, K., Schnitzer, M.J. & Block, S.M. Single kinesin molecules studied with a molecular force clamp. Nature 400, 184-189 (1999)
- Xiang, Y., et al., Structural Changes of Bacteriophage Phi29 upon DNA Packaging and Release, The EMBO Journal (2006), Vol. 25, 5229-5239
- Yodh, J. G., et al., Insight into Helicase Mechanism and Function Revealed through Single-Molecule Approaches, Quarterly Reviews of Biophysics (2010)
- Wang, H., and Oster, G. (1998). Energy transduction in the F1 motor of ATP synthase. Nature 396, 279-282
- Waterhouse A.M., Procter J.B., Martin D.M.A., Clamp M., Barton G.J., Jalview Version 2 – a multiple sequence alignment editor and analysis workbench, Bioinformatics (2009), 25 (9)
- Widom, J. & Baldwin, R.L. Tests of spool models for DNA packaging in phage lambda. J. Mol. Biol. 171, 419-437 (1983)
- Yang, Q., and Catalano, C.E. (2004). A minimal kinetic model for a viral DNA packaging machine. Biochemistry 43, 289-299
- Yardimci, H., et al., Bypass of a Protein Barrier by a Replicative DNA Helicase, Nature (2012), Vol. 492
- Yu, J., Moffitt, J., Hetherington, C.L., Bustamante, C., and Oster, G. (2010). Mechanochemistry of a viral DNA packaging motor. J Mol Biol 400, 186-203
- Zhang, Z. et al., A Promiscuous DNA Packaging Machine from Bacteriophage T4, PLOS Biology (2011), Vol. 9, Issue 2
- Zhao, W., Morais, M.C., Anderson, D.L., Jardine, P.J., and Grimes, S. (2008). Role of the CCA bulge of prohead RNA of bacteriophage Phi29 in DNA packaging. J Mol Biol 383, 520-528
- Zhao, W., et al., A Three-Helix Junction is the Interface between Two Functional Domains of Prohead RNA in Phi29 DNA Packaging, Journal of Virology (2012), Vol. 86, No. 21