Torturing Molecular Motors:
Single-Molecule Studies of Viral Packaging and Flagellar Switching

by

Adam Terrall Politzer

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jan T. Liphardt, Chair
Professor Carlos Bustamante
Professor Adam Arkin
Professor Kathleen Ryan

Fall 2013
Abstract

Torturing Molecular Motors: Single-Molecule Studies of Viral Packaging and Flagellar Switching

by

Adam Terrall Politzer

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Jan T. Liphardt, Chair

Molecular motors play a central role in all known living systems. Cells use them to fight the never ending battle against entropy. Motors allow cells to organize their components in space and time to perform complex functions. In this thesis I focus on two molecular motors: the packaging motor of φ29, and the flagellar motor of Escherichia coli. These are both well-studied motors with decades of research poured into them. By leveraging past knowledge, I was able to undertake subtle studies of their function. I tortured these motors by putting them in sub-optimal operating conditions to test their limits.

For the phage motor we were primarily interested in what sort of contacts it uses to engage and exert force on its DNA substrate. We queried the nature of these contacts by making a series of modified DNA substrates, with a focus on methylphosphonate substitutions that maintain the structure of DNA while removing the negative charge. Interestingly, we found that the negative charge of the phosphates was not required for the motor to translocate DNA, but that the charge stabilizes the interaction and makes the motor less likely to slip. We also discovered that the motor interacts more strongly with one of the DNA strands, and we used these modifications to identify a set of periodic contacts.

For the flagellar motor I was interested in the role of proton motive force (PMF) in changing the rotational direction of the motor, which is controlled by a complex on its cytoplasmic side called the flagellar switch. The motor is powered by PMF, but does the PMF also affect the switch? We found that in fact the switch is affected by changes in PMF. This has implications for the switching mechanism. The switch is not only affected by the chemical
binding signal, but is also affected by its associations with the motor and potentially its proximity to charges on the cell membrane.
Acknowledgments

This dissertation work was made possible by the contributions and support of many people. Carlos Bustamante started me off on the phage project and provided perceptive suggestions. Aathavan Karunakaran shared with me the wonders of hands-on single molecule experiments and was an excellent resource in the lab throughout our collaboration on the phage project. Jeff Moffitt and Craig Hetherington provided many fruitful discussions about the phage motor and beyond. Ariel Kaplan contributed the key insight of the logistic regression found in Appendix I that helped tie the phage work together. I had positive interactions with many others in the Bustamante Lab, all of whom I cannot list, but include, Arthur Edelstein, Courtney Hodges, Yann Chemla, Melania Strycharska, and Andy Spakowitz.

I would like to thank Jan Liphardt for providing the freedom and resources to pursue my interests. I would like to acknowledge all past and present members of the Liphardt Lab, and especially Derek Greenfield, Jessica Walter, Will Draper, Tom Yuzvinsky, Ann McEvoy, Raja Ghosh, Jake Siegel, and Alan Lowe for making the lab a lighter and more interesting place. I also thank Jessica for helping me get started with proteorhodopsin, and Will for writing and maintaining the microscope control and acquisition software.

Outside the lab I thank Yuhai Tu, Richard Berry, and George Oster for discussions of my results, and Ruchira Datta for copy-editing.

I am grateful to Howard Berg, who hosted me in his lab at Harvard when I was starting the flagellar switch work in 2010. Junhua Yuan was generous with his time as he helped me get started down the right path. Richard Branch has provided helpful discussions of interpreting the flagellar switch data. I also am indebted to Larry Smith and Roberta Towner who let me stay at their house during my Harvard sojourn.

I would also like to thank my family. My parents’ contributions to my life are too numerous to list here. They are so consistently good to me that sometimes I forget how lucky I am to have them.

Finally, Saheli Rupa Datta has been a steady source of much support and joy for which I am most grateful.
Chapter 1

Introduction
Molecular Motors

Molecular motors are ubiquitous and necessary for life as we know it. Barring advances in synthetic nanotechnology, and for the purpose of this thesis, a molecular motor is a nanoscale protein structure that converts stored chemical energy into mechanical force and motion. Motors allow cells to organize components and move in ways that would be impossible if cells relied on diffusion alone. Without motors life would be limited to simple sacks of molecules floating around. Indeed, even the simplest bacteria and archaea are full of molecular motors. All of the fundamental processes of the Central Dogma of biology (DNA replication, transcription, and translation) require motors.

Motors are made of intricately folded polypeptide chains composed of hundreds of amino acids, comprising thousands of atoms. They have length scales of nanometers and can associate in complexes to form structures of tens of nanometers. Concepts familiar from macroscopic mechanics such as work and torque extend down to these tiny machines. However a major difference is that at this scale thermal fluctuations play a significant role. The forces the motors exert typically range from 1-100 pN, and the individual steps are nanometer sized, so units of pN·nm are useful when measuring their work and energy. At room temperature, thermal fluctuations, \( k_b T \), are approximately 4 pN·nm.

We are motivated to study molecular motors primarily to try to understand the world around us. How do living systems convert chemical energy into mechanical movements? How do motors interact with their substrates? How do motors control and modulate their activity? One day it may be possible to design our own molecular motors. Biological protein machines are ideally suited to exert piconewton forces on the nanometer scale. A simple DNA sequence can specify complex structure and function for a protein, so once a useful protein is identified biological tools make it possible to quickly and easily produce trillions of copies. However, before we can build our own molecular motors, we must first understand how nature’s motors work.

Over the past few decades new experimental techniques have made it possible to observe the function of individual motors as they move, one step at a time (Howard, 2001, Schliwa, 2003). The fundamentals of a wide range of molecular motors’ operation have been established, but many details remain unknown. Some motors are driven by energy stored in chemical bonds in molecules such as ATP, while others are driven by electrochemical gradients. In this thesis I examine one of each. The bacteriophage \( \varphi 29 \) DNA packaging motor is driven by ATP hydrolysis, while the \textit{Escherichia coli} flagellar motor is...
driven by protons that move due to the proton motive force. In both cases we put the motors into non-standard operating conditions to elucidate further details of their mechanisms.

There are also similarities in our experimental approach. The motors are too small to see directly. We want to observe them in action, so techniques like electron microscopy that require vacuum are not applicable. To study these nanomachines we attach them to micron-sized polystyrene beads that are visible in a light microscope. Single motors are capable of displacing the micron-sized beads such that their movement can be tracked. Tracking single motors allows us to discern details that would not be present in ensemble measurements of many motors.

To illustrate the value of single measurements as compared to bulk measurements, Steve Chu likes to point out that if you were to measure a random sample of humans as a group you would find that each individual has one testicle and one ovary. We of course know that this is false, and to get to the truth we would need to examine the humans individually. Similarly, studying single molecules individually allows us to get closer to the truth of their inner workings. For example, in the phage packaging motor, single molecule measurements revealed that on the microscale the motor does not move at a constant rate; rather, it moves in bursts of coordinated steps (Moffitt et al., 2009). In the flagellar switch, individual measurements revealed that it has a much higher signal amplification than previously thought (Cluzel et al., 2000, Yuan and Berg, 2013).

In addition to offering a means to identify and separate heterogeneous behavior, single molecule measurements also offer opportunities to apply and measure forces on these molecules. This is extremely useful and important to the study of motors, since their primary activity involves generating forces. By attaching a bead to a motor, it is possible to record the forces that the motor exerts. One of the best tools for manipulating these types of beads and measuring forces of this scale is the optical tweezers. Originally invented to trap particles (Ashkin, 1970), it later became apparent that optical tweezers can be used to measure movements and forces. The underlying principle is that a dielectric, such as the polystyrene beads that we use, is attracted to a focused spot of light. The attractive force is typically proportional to the bead’s distance from the focus, so the optical trap behaves like a Hookean spring. Thus, force can be determined by measuring the bead’s displacement from the center of the trap.

By tethering such a bead to DNA this ability to determine force can be used to probe the mechanics of DNA-protein complexes. The robustness of these techniques is grounded in the detailed understanding of the mechanical
properties of DNA, which were actually first measured using a related single-molecule method that uses ferromagnetic beads and magnetic fields (Smith et al., 1992). Once the force-extension properties of DNA were determined, it became possible to study proteins that act on DNA (Davenport et al., 2000, Wuite et al., 2000).

Another similarity between the two motors in this thesis is that they both come with natural handles that we can attach beads to: the phage motor has the prohead, and the flagellar motor has the flagellum. This is important because attaching a large bead to a molecular motor directly can inhibit the activity of the motor. There are ways to make such an attachment by using an intermediate tether, but one always has to make sure that it does not impede the motor. The motors in this thesis do not have this challenge, and that is part of the reason that researchers were able to develop robust single-molecule assays for them.

These experiments have a certain simple appeal. While technically challenging—and there are certainly a range of complicated chemical reactions involved—ultimately we are observing single motors execute simple movements. There is a certain satisfaction and sense of wonder that comes from stripping away the biochemical abstractions, and watching single molecules move. Even more satisfying is having predictions about their movement confirmed in real time. For example, the first time we ran the modified DNA experiments with the phage motor it was extremely exciting to see the motor pause at the location where we made the modification. In the case of the flagellar motor, we were able to turn a knob to control the PMF, and it was satisfying to see the motor immediately respond. To come to the results presented in this thesis many repeated trials were required, which dragged on at times, but the goal of revealing non-trivial aspects of these motors carried the experiments along.

**Phage Packaging Motor**

There are two main areas of context in which the phage motor is interesting (beyond the pure question of how this molecular machine functions): as a central component of phage biology, and also as a model system of other DNA translocases.

Phages are extremely abundant. The human body contains 10 times as many bacterial cells as human cells, and for every bacterium there are about 10
phages (Letarov and Kulikov, 2009). Beyond the human body, anywhere there are bacteria, there are phages. A liter of seawater can contain $10^{10}$ phages, and it's estimated that there are $10^{31}$ phages on earth (Breitbart et al., 2002). These phages are of enormous ecological significance because they kill approximately half of the earth’s bacteria every two days (Gravitz, 2012). Phages also mediate transfer of DNA between bacteria, greatly expediting evolution (Paul, 1999). On a medical note, as more and more bacteria develop antibiotic resistance, phage therapy to kill pathogenic bacteria may prove a useful and vital tool to fight infection (Harper et al., 2011).

The $\phi$29 phage packaging motor is also extremely interesting as a DNA translocase. Its ATPase shares sequence homology with other motors that perform a range of essential tasks, including DNA recombination, repair, and segregation, and as such it is a member of the Additional Strand Conserved E (ASCE) family of ATP-hydrolyzing proteins (Singleton et al., 2007, Iyer et al., 2004, Burroughs et al., 2007). Like many of these motors it contains a multimeric ring of identical subunits that encircle the DNA substrate. The phage motor is an appealing model system because it is straightforward to produce large quantities of it, and because it comes with a convenient handle already attached: the prohead. Beads coated with antibodies raised to the capsid proteins can be attached without impeding the motor. The lessons we learn about the phage motor may be applicable to the other similar motors. It’s unlikely that they work in exactly the same way, but there are probably similar underlying principles.

The packaging motor plays a key role in the lifecycle of tailed bacteriophages, which are one of the most abundant families of phages. These phages pack their capsid tightly full of DNA. This allows them to perform a useful trick: they store mechanical energy when they are in the richly energetic environment (high [ATP]) of their host cell, as the DNA is packed to nearly crystalline density, as shown in Figure 1-1 (Comolli et al., 2008). Energy is stored when the DNA is confined due to factors of electrostatic auto-repulsion of the charged DNA, mechanical bending of the DNA, and reduction of entropy. After the phage lyses its host cell, it finds itself in a much less energetic environment. When it encounters a new bacterium to infect it uses the stored mechanical energy to inject its DNA payload. In this way the phages can use the energy provided by the previous host cell to infect the next host cell, even when the cells are nowhere near each other in space or time.

This work focuses on the $\phi$29 packaging motor. $\phi$29 infects the common soil bacterium Bacillus subtilis, and is one of the most well-studied phages, with much of its biochemistry and structure established. Importantly for this work, researchers have developed a robust in vitro DNA packaging assay, which has
been extended to the single-molecule level (Smith et al., 2001, Chemla et al., 2005, Rickgauer et al., 2008).

φ29 has a 19.3 kb double-stranded DNA genome that encodes 19 genes (Grimes et al., 2002). It is densely packed into a 50x44 nm icosahedral protein capsid by the packaging motor (Figure 1-2). The packaging motor contains a pentameric ring of ATPase proteins. Between it and the capsid is a pentameric ring of “packaging” pRNA that links it to a dodecameric protein structure called the connector, which is connected to the opening of the capsid (Simpson et al., 2000, Zhao et al., 2008, Morais et al., 2008). The connector and RNA are essential for packaging, so it is not known if they are involved in exerting force on the DNA, or if the ATPase directly engages the substrate. Once the DNA is completely inside, the ATPase and RNA rings disassemble and the tail protein structure forms. The tail stabilizes the full capsid, and later recognizes surface markers on the target cell, attaches the phage to the cell, penetrates the cell, and then signals to the phage portal to eject its DNA.

The motor is powered by ATP hydrolysis. Bulk packaging measurements found that the motor packages ~2 basepairs of DNA per molecule of hydrolyzed ATP (Chemla et al., 2005). High resolution measurements of single motors revealed that the fundamental step size is actually 2.5 basepairs (Moffitt et al., 2009). This step size is consistent with the stall force of the motor, which was measured to be ~110 pN (Rickgauer et al., 2008). The work performed just below the stall force is the product of the stall force and step size, yielding 94 pN·nm, which is quite close to the free energy of ATP hydrolysis, ~80 pN·nm at standard conditions (Stryer, 1995). Since 110 pN is a lower bound for the stall force, the motor has an efficiency of at least 85%.

Interestingly the motor shows two phases: a burst of four 2.5 basepair steps followed by a dwell phase (Moffitt et al., 2009). Apparently the subunits are coordinated as each subunit waits for the other subunits to bind ATP before they hydrolyze it. Strain is thought to communicate the state of the subunits to their neighbors. The non-integer step size was unexpected, and it implies that the motor cannot rely on a specific periodic feature of DNA to engage. This suggests that the motor at least partially uses non-specific steric interactions to exert force on its DNA substrate.

The work in Chapter 2 of this thesis was motivated by wanting to understand how the motor engages DNA. We performed a range of experiments with a set of substrates, each lacking different elements of standard DNA structure. We found that the motor uses both specific and relatively nonspecific contacts to transmit force to the DNA.
Chemotaxis and the Flagellar Motor

Bacterial chemotaxis is a fascinating field because it studies a complex network that has been well-characterized. The fact that it is well-characterized does not mean that all the questions are solved however. Rather, every new discovery leads to more questions, and when tackling these new questions we can stand on the vast mountain of existing knowledge.

Chemotaxis refers to the movement toward or away from certain chemicals. A large part of the early success in studying the chemotaxis network stemmed from the fact that its mutants are non-lethal, so they are much more practical to study than mutations that lead to cell inviability. Chemotaxis also has clear and accessible experimental read-outs of activity because the output of the network is a physical response. Early read-outs included assays that counted the number of cells that swam into capillaries containing chemical gradients (Adler and Dahl, 1967). Another convenient assay that has persisted is the growth and movement of cells on a petri dish of nutrient-containing semi-solid agar gel: as the cells metabolize nutrients they form chemical gradients, and the cells move to form “swarm rings” that are visible to the naked eye (Adler, 1966). The accessibility of these types of experiments means that all of the chemotaxis proteins in Escherichia coli are known, and their function is generally understood, due to a wide range of studies of mutants. The chemotaxis network is one of the best characterized signaling pathways.

It is also possible to access another level of detail by studying the response of individual cells. Howard Berg famously followed single swimming cells with a tracking microscope to learn that they alternate periods of straight swimming (“runs”) with periods of reorientation (“tumbles”), and they suppress tumbling when the environment is improving (Berg and Brown, 1972). Having established that basic phenomenon, a more practical technique is to observe rotation of the flagellar motors themselves. This was first done by tethering flagella to a surface and watching the cell bodies rotate (Silverman and Simon, 1974). A more precise method is to attach a polystyrene bead to a flagellar stub on an immobilized cell, and follow its rotation (Ryu et al., 2000). Beads are less prone to random intermittent surface interactions and becoming stuck, so their velocity is a better read-out of motor activity than the previously used cell body rotation.

The main components of the chemotaxis pathway are conserved across the majority of motile bacteria (Wadhams and Armitage, 2004). We use E. coli as a model organism because of the ease of culturing it, and because of the extensive body of work that has previously characterized it. E. coli also happens
to have one of the simplest chemotaxis networks, as many other bacteria have additional accessory proteins (Rao et al., 2004). But since the basic components are similar, the things we learn about *E. coli* are broadly applicable to other species of bacteria. Understanding chemotaxis network motifs is also of general interest to understanding other protein networks.

Motile bacteria execute chemotaxis by undergoing a directed random walk. They swim 20-30 µm/s by rotating helical flagella that extend 5-15 µm outside the cell (*E. coli* have 6 on average). In between the flagella and the motors that rotate them is a universal joint, or “hook,” that allows the flagella to bend and form a bundle behind the cell due to hydrodynamic forces (Macnab, 2003). When the motors rotate counterclockwise (as seen looking at the motor from the outside of the cell) the cell is propelled forward (Larsen et al., 1974). As shown in Figure 1-3, when one or more motors reverse direction to clockwise, those flagella move out of the bundle and the cell stops moving forward and tumbles in place (Turner et al., 2000). By changing the frequency of tumbles, the cell is able to undergo a directed random walk.

The range of run times is governed by physical constraints of diffusion (Berg and Purcell, 1977). On the low end runs are rarely shorter than 1 s because in less than that time the cell does not outpace diffusion, so the cell cannot determine if conditions have significantly improved or worsened. The upper range of runs is about 10 s, because beyond that time the rotational diffusion of the cell means its direction is no longer correlated. That is, because of Brownian motion, on average a cell the size of *E. coli* cannot swim straight for more than 10 s. Because of these constraints wildtype cells tune their response to fall within this useful range.

Cells sense their environment with chemoreceptors. Chemoreceptors are transmembrane proteins that contain a binding site to which chemicals of interest bind. The attractant chemicals they sense are typically fuel sources for the cell, and are often amino acids or sugars. Repellents include weak acids and metal ions. Binding of a ligand causes a conformation change in the chemoreceptor, which is sensed by downstream chemotaxis proteins.

The chemotaxis network is a prototypical example of a two-component signal transduction system, of which there are hundreds in bacteria. It has been extensively studied and reviewed (Bren and Eisenbach, 2000, Baker et al., 2006a, Baker et al., 2006b, Parkinson et al., 2005, Sourjik, 2004, Bourret and Stock, 2002). See Figure 1-4 for an overview. The canonical system contains a histidine kinase that is autophosphorylated with a phosphate group from adenosine triphosphate. The phosphate is then transferred to an aspartate residue on a second, typically soluble, protein, called a response regulator. When the response regulator is phosphorylated it generally changes
conformation, which alters the binding affinity for its target. Many response regulators regulate transcription of genes, but in the chemotaxis network the response regulator directly binds to the flagellar switch.

The histidine kinase of the chemotaxis network is CheA, which is bound to chemoreceptors. CheA phosphorylates the response regulator CheY, and the phosphorylation rate depends on the binding status of the chemoreceptor. The rate is high when no ligand is bound and low when an attractant ligand binds. Phosphorylated CheY (CheY-P) then diffuses throughout the cell, and when it binds to the flagellar switch it increases the probability of clockwise rotation. Meanwhile, a third protein, CheZ, dephosphorylates CheY at a constant rate, which helps keep the CheY pool from becoming saturated.

Also essential are two more proteins that affect CheA activity: CheR and CheB. These proteins can modulate the kinase activity of CheA by methylating (CheR) and demethylating (CheB) specific glutamic acid residues on the chemoreceptor. In addition to phosphorylating CheY, CheA also phosphorylates CheB. When phosphorylated, CheB demethylates chemoreceptors, which lowers CheA’s kinase activity. This introduces a feedback loop that allows the cell to sense changes in chemical concentration, rather than absolute chemical concentration. This allows the cell to maintain its run lengths within the useful regime, and constantly optimize its position in its environment, which is full of chemical gradients. Somewhat astoundingly, this adaptation allows *E. coli* to sense chemical gradients over five orders of magnitude of chemical concentration.

There is some structural diversity among bacterial flagellar motors (Chen et al., 2011), but they all contain a similar set of parts. Again we focus on the *E. coli* motor, as it is the best studied and the focus of this work. The *E. coli* motor is 45 nm in diameter and spans the inner membrane, cell wall, and outer membrane. The complex is 11 MDa, and contains 12 different core proteins, many of which are multimeric rings made of ~26-110 copies of each protein. There are an additional ~30 proteins involved in assembly and gene regulation of this complex structure. Because of the complexity and energetic cost of this structure, synthesis is highly regulated and proceeds in a hierarchical order with multiple checkpoints (Berg, 2003, Macnab, 2003).

The motor contains a rotor made of four distinct rings and a central rod (Figure 1-5). The rotor is surrounded by the stator, which is anchored to the peptidoglycan cell wall. The stator is made of the proteins MotA and MotB, which form units consisting of 4 MotA and 2 MotB, with a total of up to at least 11 units (Reid et al., 2006). The stator units contain a proton channel, which is the motor’s means of harnessing energy. The motor is powered by the proton motive force (PMF), which is the electrochemical potential of protons.
across the inner cell membrane, and is normally generated by cellular respiration. At a standard PMF of -150 mV each proton provides \( \sim 6 \, k_B T \) of energy, and \( \sim 1200 \) protons per revolution move through the stator, turning the motor at speeds of up to \( \sim 150 \) Hz. There is an aspartate residue in MotB that is protonated, which causes a conformation change in MotA, which then exerts a force on the rotor protein FliG via an electrostatic interaction (Sowa and Berry, 2008).

The direction of rotation is mediated by the switch complex, which contains FliG, FliN, and FliM. At room temperature in the absence of CheY-P the motor rotates counterclockwise, which corresponds to straight swimming. When CheY is phosphorylated, CheY-P binds to FliM, which leads to a conformation change in FliG and ultimately clockwise rotation.

Motor composition is surprisingly dynamic, even as it is rotating. Experiments in cells lacking MotAB that then started expressing these stator proteins showed that the units can associate individually and function independently (Blair and Berg, 1988). Velocity traces in these “resurrection” experiments show a staircase pattern as consecutive stator units engage with the rotor. Fluorescence recovery after photobleaching (FRAP) experiments with GFP-MotB showed that in functional motors the stator units exchange on average every \( \sim 25 \) seconds (Leake et al., 2006). Additional FRAP studies showed that FliM and FliN exchange on a time scale of \( \sim 10 \) min, but FliG does not exchange (Fukuoka et al., 2010, Delalez et al., 2010).

One final particularly interesting property of the chemotaxis system is its high gain. Gain is defined as the ratio of fractional change in output to fractional change in input. In the chemotaxis network the output is the rotation direction bias of the motor, the input is binding of ligand to chemoreceptors, and it has a gain of \( \sim 100 \) (Segall et al., 1986). Much of the gain takes place at the site of the chemoreceptors, which have a gain of \( \sim 35 \) (Sourjik and Berg, 2002). They accomplish this by being closely clustered and interacting with their neighbors; one chemoreceptor that binds ligand can activate 35 of its neighbors. The remaining gain occurs at the flagellar switch.

The activity of the switch displays a steep sigmoidal dependence on [CheY-P]. The steepness of the sigmoid can be quantified with the Hill coefficient, \( N \) in the equation:

\[
A = \frac{[L]^N}{K_d + [L]^N}
\]

where \( A \) is fractional activity, \([L]\) is ligand concentration, and \( K_d \) is the dissociation constant (Stryer, 1995). The switch was first measured to have a Hill coefficient of \( \sim 4 \) (Scharf et al., 1998), however that measurement was
flawed because it measured [CheY-P] in populations of cells, which have a
distribution of different actual concentrations, smearing out the response curve.
Cluzel et al. (2000) performed a better measurement by quantifying CheY-P in
single cells using a fluorescent protein tag, and found a Hill coefficient of ~10.
Most recently, Yuan and Berg (2013) found that the switch itself adapts to
different concentrations of [CheY-P] by changing the number of FliM units, so
Cluzel et al.’s measurement was actually a measurement of motors with
different numbers of FliM. Correcting for the constant number of FliM, the
Hill coefficient is actually as high as 21.

With these results in mind, in addition to studying the switch to figure
out its general mechanism, we are also motivated to try to figure out how it
achieves such high cooperativity. In Chapter 3 I examine if PMF may play a
role in maintaining this high cooperativity.
Figure 1-1. Cryo-EM reconstruction of packed φ29 heads. The DNA is packed so tightly that it is nearly crystalline. a) Top-down cross-section and radially integrated intensity b) Side view cross-section and linearly integrated intensity. c) 3D reconstructions. Individual layers of DNA are apparent. The periodicity of the peaks of DNA density is approximately 2.2 nm, quite close to the diameter of double-stranded DNA of ~2 nm. Figure adapted from Comolli et al., 2008.
Figure 1-2. Structure of the φ29 packaging complex. The dodecameric connector (cyan) extends within the capsid (grey) and is bound to the pRNA (magenta), which in turn is bound to the ATPase (blue). Figure from Morais et al., 2008.
Figure 1-3. Frames of a movie that captures a tumble of a swimming *E. coli*. Flagella were labeled with a fluorescent dye. The movie was recorded at 60 Hz, but only every other frame is shown. The cell enters the frame swimming down and to the left, and then multiple motors reverse direction and the flagellar bundle breaks apart. The cell reorients and then the bundle is reformed in the last two frames, as the cell swims off in a new direction. Figure from Turner et al., 2000.
Figure 1-4. Chemotaxis network schematic. Two dimeric chemoreceptors—methyl-accepting chemotaxis proteins (MCPs)—are shown. In addition, two chemotaxis protein CheW monomers and a CheA dimer are shown interacting with the highly conserved signaling domain of the MCPs in the cytoplasm. A decrease in attractant concentration induces trans-auto phosphorylation of the CheA dimer, which phosphorylates the response regulator CheY. Phosphorylated CheY then binds to the flagellar motor to bias clockwise rotation (tumbles). Phosphorylated CheA also phosphorylates another response regulator—the methylesterase CheB. Phosphorylated CheB competes with a constitutive methyltransferase, CheR, to control the degree of methylation of specific glutamates in the MCPs. This resets the signaling state of the receptors and allows them to adapt to the present concentration of attractant and to sense subsequent changes. The dephosphorylation of phosphorylated CheY is accelerated by the phosphatase CheZ. Figure and caption adapted from Wadhams and Armitage, 2004.
Figure 1-5. Flagellar motor schematic and electron micrograph. MotA, MotB, and components of the transport apparatus (dashed ellipse) do not survive extraction with detergent and, therefore, are not shown on the right. This reconstruction is derived from rotationally averaged images of about 100 hook–basal body complexes of Salmonella polyhook strain SJW880 embedded in vitreous ice. The radial densities have been projected from front to back along the line of view, so this is what would be seen if one were able to look through the spinning structure. Figure and caption adapted from Berg, 2003.
References


Chapter 2

Substrate Interactions and Promiscuity in a Viral DNA Packaging Motor

The work presented in this chapter was adapted with permission from

Summary

Many processes in biology, including DNA recombination, prokaryotic chromosome segregation, gene transposition, and viral DNA packaging, involve the translocation of DNA or RNA by ATP-driven ring motors belonging to the ASCE/AAA superfamily (Iyer et al., 2004, Erzberger and Berger, 2006). While their mechanism of mechanochemical conversion is beginning to be understood (Hopfner and Michaelis, 2007), little is known about how these motors engage their nucleic acid substrates. Here we explore translocase–DNA interactions of the bacteriophage φ29 DNA packaging motor using a single-molecule optical tweezers assay (Grimes et al., 2002). To investigate the role of the phosphate charge, we test the ability of the motor to package DNA constructs with regions of uncharged methylphosphonate (MeP) modifications. The motor actively traverses these inserts, though with reduced probability compared to native DNA, indicating that phosphate charges are not essential, though important for translocation. By changing the length of the MeP inserts and restoring the charge to one or the other DNA strand selectively, we establish that important contacts are made with phosphate charges on the 5’-3’ strand every 10 base pairs. High-resolution measurements of the motor dynamics as it packages neutral DNA reveal that phosphate contacts serve both a load-bearing role and a regulatory role in the mechanochemical cycle. In addition, the motor is able to package inserts lacking bases and sugars, inserts with single-stranded gaps or unpaired bulges, and even a non-biological linker, suggesting that the motor makes promiscuous contacts with a wide variety of chemical moieties over a range of geometries. Such diverse contacts may be the basis for the large substrate diversity observed within the ASCE superfamily.
Introduction

The ASCE superfamily of proteins consists of structurally similar ATPases associated with diverse cellular activities involving metabolism and transport of proteins and nucleic acids in all forms of life (Iyer et al., 2004, Erzberger and Berger, 2006). A subset of these enzymes are multimeric ringed pumps responsible for dsDNA transport in processes including genome packaging in adenoviruses, herpesviruses, poxviruses, and tailed bacteriophages (Burroughs et al., 2007). The *Bacillus subtilis* phage φ29 packages a 6.6-µm (19.3 kb) dsDNA genome into a 42x54 nm prolate icosahedral protein capsid, overcoming the large elastic, electrostatic, and entropic barriers associated with compacting DNA to near crystalline densities (Grimes et al., 2002). This remarkable feat is accomplished by a motor complex composed of a dodecameric head-tail connector protein (Simpson et al., 2000), a pentameric ring of prohead-RNA (Morais et al., 2001, Morais et al., 2008), and a pentameric ring of the ASCE ATPase gene product 16 (gp16) (Morais et al., 2008, Burroughs et al., 2007).

The mechanochemistry of this motor is relatively well understood. Single-molecule studies have shown that it processively translocates against forces as high as 60 pN (Smith et al., 2001). The hydrolysis cycles of the individual ATPase subunits are highly coordinated: the ATPase ring loads multiple ATP molecules during the first portion of its mechanochemical cycle, the *dwell phase*, and then during the second portion of the cycle, the *burst phase*, the motor translocates the DNA in a 10-bp burst, composed of four 2.5-bp steps (Moffitt et al., 2009), triggered by the release of hydrolysis products (Chemla et al., 2005). In contrast, very little is known about the motor-DNA interaction. Does the motor contact a single DNA element, such as a phosphate or a base, or are its contacts distributed over multiple parts of the DNA? What is the role of the DNA geometry in this interaction? Finally, are these contacts only responsible for transmitting force, or do they also regulate the kinetic cycle?

Results and Discussion

To address these questions we used optical tweezers to study the dynamics of single prohead-motor complexes (Smith et al., 2001, Rickgauer et al., 2008) as they package DNA containing a region of specifically modified substrate (*Figure 2-1a*). To test the role of the phosphate backbone charge in motor-DNA interactions we inserted a 10 bp region of ds methylphosphonate DNA (ds-MeP). In this modification the charged oxygen is replaced with an
uncharged isosteric methyl group, conserving the B-form structure of DNA (Thiviyanathan et al., 2002, Strauss and Maher, 1994) (Figure 2-1b, inset). Thus, it is possible to infer the role of this chemical modification in a native geometric context. Figure 2-1b shows sample packaging traces under 5 pN of constant tension and saturating [ATP] (1 mM). Packaging proceeds normally until the motor encounters the inserted modification, where it pauses and then either successfully traverses the insert or completely dissociates. In stark contrast to related helicases where disruption of a single charge-backbone interaction via substrate modification (Eoff et al., 2005, Kawaoka et al., 2004, SenGupta and Borowiec, 1992) or site-directed mutagenesis (Mancini et al., 2004, Dillingham et al., 1999) completely abolishes translocation, the packaging motor traverses 10 bp of neutral DNA with a probability of ~80% (Figure 2-1c).

How do we rationalize such a high traversal probability? One possibility is that phosphate charges are essential for active translocation, but the motor is capable of crossing the insert, without contacting it, by exploiting diffusive fluctuations. A second possibility is that phosphate interactions are not essential, and the motor can exert force on the insert via non-charge contacts. We discriminate between these mechanisms by exploiting the strong force dependence of diffusive traversal time (Howard, 2001). Figure 2-1c shows that there is only a two-fold increase in pause duration with a 15 pN increase in force, much less than the $10^5$-fold increase predicted for diffusion across a 10 bp distance. Figure 2-1c also shows that lowering [ATP] increases the pause duration and decreases the traversal probability, providing further support for an active, [ATP]-dependent traversal mechanism. These results indicate that the motor actively traverses the insert by making contacts with elements other than the phosphate charge. However, the fact that packaging is affected by the absence of these charges indicates that native packaging involves both charge and non-charge contacts.

To determine if phosphate charges from both strands are equally important, we created hybrid inserts in which one strand has a MeP backbone and the other has an unmodified, charged backbone. We used a 30 bp insert to accentuate the effect of the chargeless section (the traversal probability of packaging a 30 bp dsMeP insert at 5 pN is ~4%, see Fig. 1d). Figure 2-1d shows that at 5 pN the traversal probability of the hybrid insert with MeP on the strand packaged from 3'-5' is almost 90%, whereas the traversal probability of the hybrid insert with MeP in the 5'-3' strand is 10%. This result clearly indicates that the most important phosphate interactions are made with the 5'-3' strand in the direction of packaging. Such preferential interaction with a single DNA strand has been shown for the monomeric dsDNA translocase
EcoR124I (Stanley et al., 2006), though it is more surprising in the case of a ring-ATPase in which multiple subunits of the ring are simultaneously in close proximity to both strands.

Next, we asked if a critical length is involved in the interaction of the motor and its DNA substrate. Figure 2-1e shows that as we increase the length of the neutral insert from 5 to 10 bp there is no statistically significant change in the traversal probability until the 1 bp increase from 10 to 11 bp which results in a two-fold reduction. However, further increasing the length to 15 bp does not produce a similar change. The location of this discrete change in traversal probability at first appears inconsistent with the 10 bp of DNA packaged by the motor each full mechanochemical cycle (Moffitt et al., 2009), but these results are easily reconciled if the motor makes contact with two adjacent phosphates, with either contact sufficient for packaging (See Figure 2-1f). The crystal structure of the related BPV helicase E1 with its ssDNA substrate reveals simultaneous contact with adjacent phosphate charges (Enemark and Joshua-Tor, 2006), lending support to this interpretation. Thus, the most important phosphate interactions are with adjacent phosphates every 10 bp on the 5’-3’ strand. However, the fact that the traversal probability does not drop to zero above 10 bp again indicates that the motor makes additional non-charge contacts. More complicated models, such as redundant translocating contact points in the motor that span 10 bp, are also possible but are neither necessitated by this data nor suggested by the crystal structures of related motors (Massey et al., 2006, Enemark and Joshua-Tor, 2006).

We next investigated the specific role of phosphates in the mechanochemical cycle by probing the base-pair-scale dynamics of the motor at an uncharged insert. Phosphate contacts may serve two possible roles in the mechanochemical cycle. Either they provide the long-lived contacts required for keeping the enzyme attached to the DNA, increasing its processivity, or they serve a sensory role, accelerating a chemical rate, such as ATP hydrolysis, upon detecting that the DNA is bound and properly oriented. These two roles of the phosphate charge predict different behaviors of the motor as it traverses the modified insert. If the phosphate provides load-bearing contacts, its absence will increase the dissociation rate of the motor, and the insert-induced pause will consist of a series of attempts to package followed by small slips. Alternatively, if the role of the phosphate charge is sensory, translocation will proceed normally with no slips, but the time between packaging steps will be greatly lengthened, due to the decreased rate of catalytic turnover.

To reveal the dynamics of the motor as it crosses a neutral insert (10-bp ds-MeP), we followed packaging using a dual-trap optical tweezers with higher spatial and temporal resolution (Moffitt et al., 2009). Figure 2-2a shows that
the pauses observed at low resolution are actually remarkably dynamic events, containing two types of sub-pauses at distinct locations along the modified DNA insert. The first type of sub-pause, which we term an *upstream pause* because it occurs at longer DNA tether lengths, occurs in a uniform position, ± 1 bp (s.d.), and has longer average durations, 1.00 ± 0.08 s (s.e.m.; Figures 2-2b&d). These sub-pauses are followed by either brief disengagement (slips) of the motor or packaging attempts. These attempts are themselves followed by a second class of sub-pause, which we term a *downstream* pause. These sub-pauses occur over a larger range of positions, ± 3 bp (s.d.), have shorter durations, 80 ± 10 ms (s.e.m.), and are terminated either with slips or with the successful packaging of the insert (Figures 2-2c&e). After slips from either the upstream or downstream pauses the motor typically recovers and repackages the DNA to the position of the upstream pause. Occasionally the motor does not recover from a slip, resulting in a terminal slip. Figure 2-2f contains the branching probabilities of these events.

The upstream- and downstream-pause time distributions are both well-described by single exponential decays (Figure 2-2b&c), suggesting that i) these processes are governed by single rate-limiting kinetic events and that ii) each class of sub-pause corresponds to an individual kinetic event, as opposed to a weighted sum of different kinetic processes. In combination with the distinct mean durations, these observations suggest that each class of sub-pause corresponds to a unique mechanochemical phase of the motor: ATP-loading dwell or stepping burst. To identify the kinetic state of the motor associated with each class of sub-pause, we measured the dynamics of motors traversing 5 bp of neutral DNA, an insert shorter than the 10-bp span of the full mechanochemical cycle. Since contacts during the burst phase occur every 2.5-bp while contacts during the dwell phase occur only every 10 bp, all motors must make burst phase contacts with the neutral DNA while a fraction of the motors will burst across the short insert, never making dwell phase contacts with the neutral DNA. If the upstream pause occurs during the dwell phase, then only a fraction of the motors will display such a pause. Figure 2-6 shows that this is indeed the case; only ~50% of the motors packaging 5 bp of ds-MeP display a long pause, strongly suggesting that the upstream pause corresponds to the dwell phase. This conclusion is further supported by the average duration of the downstream pauses and the exponential nature of their distribution, which are both similar to the micro-dwells observed during the burst phase on charged DNA under similar conditions (Moffitt et al., 2009). Finally, the highly uniform position of the upstream pause indicates that this single-base-pair position is unique; suggesting that after slipping and
repositioning, the motor resides at the boundary of the charged and neutral DNA during its dwell phase.

The multiple slip/attempt phenotype observed in Figure 2-2a indicates that removing the phosphate charge reduces the processivity of the motor. Since the motor slips both from upstream and downstream pauses, load-bearing phosphate contacts are made during both the dwell and burst phase of the cycle. Additionally, upstream pauses are on average 10-fold longer than dwells on normal DNA under the same experimental conditions, ~100 ms (Moffitt et al., 2009); thus, the absence of the phosphate charge slows the dwell phase, revealing an additional sensory role for this contact. The fact that a single kinetic event dominates the upstream pause duration, in sharp contrast to the ~4 kinetic events that are known to be rate-limiting for the dwell phase on charged DNA (Moffitt et al., 2009), indicates that the mechanochemical cycle of the packaging motor contains a single kinetic checkpoint—a process that halts the chemical cycle until the DNA is correctly positioned. In contrast, the average duration of the contacts during the burst phase is not significantly modified on neutral DNA, hence these contacts likely do not involve a significant sensory role.

These dynamics suggest a mechanism for the motor to cross neutral inserts. Successful traversal is a kinetic competition between the increased off-rate of the motor on neutral DNA and the time it takes to successfully complete the mechanochemical events necessary to traverse this modified DNA. There is a small probability of loading the necessary ATPs and starting the burst, out-competing the increased slipping rate on the neutral DNA; however, as illustrated by the probabilities in Figure 2-2f, the large probability of recovering from a slip allows the motor to attempt to package the insert many times, amplifying this small probability to the large traversal probabilities observed in Figure 1. This kinetic competition also explains the dramatic decrease in the ability of the motor to package the 30 bp of neutral DNA since this would require multiple consecutive mechanochemical cycles to out-compete slipping. Finally, the branching probabilities observed in Figure 2-2f predict traversal probabilities consistent with the values in Figure 2-1, providing further support for this model.

The fact that the motor resides for a finite time in the neutral DNA (Figure 2-2) provides further evidence for additional, non-charge contacts. To test the role of sugars and bases in these interactions, we created an insert with these elements removed (Figure 2-3a). These chemical modifications also disrupt the helical geometry of the DNA; thus, to isolate the role of DNA geometry, we created inserts with no chemical modification yet with large disruptions to the helix; namely, single-stranded gaps (Figure 2-3b) and
unpaired bulges (Figure 2-3c). Remarkably, the φ29 motor is capable of packaging all of these modifications (Figure 2-3b&c) and the force dependence of the mean pause durations is inconsistent with a purely diffusive model. Notably, the motor shows lower traversal probability and higher pause durations for gaps on the 5’-3’ strand than those on the 3’-5’ strand, with a similar trend seen in pause durations of the bulges, consistent with the important 5’-3’ contacts discussed above. Our findings are consistent with the successful packaging of short unpaired bulges in phage lambda (Pearson and Fox, 1988), though they differ from reported results in φ29 (Moll and Guo, 2005) and T4 (Oram et al., 2008) where single-stranded gaps were not packaged. However, disparate experimental methods likely account for these apparent differences.

Many of the modifications we have probed (Figures 2-1 and 2-3) change multiple features of the DNA simultaneously, e.g. the 20 nt single-stranded insert removes bases and sugars and disrupts the B-form geometry. Thus, in order to extract the relative importance of the different chemical moieties while controlling for other changes, we performed a set of multivariate logistic regressions (Agresti, 2003) on different sub-sets of our data. This analysis provides a quantitative ranking of the importance of the different contacts and their force dependence, see Figure 2-4 and Figure 2-5. These regressions confirm the importance of the phosphates every 10 bp on the 5’-3’ strand and reveal that these contacts are only important under the application of force. In addition, the regression analysis reveals less important contacts with phosphates on both strands on a smaller distance scale, which, remarkably, remain important at zero applied load. This residual importance may arise from the force the motor generates on the DNA, as opposed to the external load applied optically, supporting our conclusion that these contacts occur during the burst phase.

The regression analyses also provide information on the nature of the non-charge contacts, revealing that important, but minor, contacts are made with bases or sugars (Figure 2-5a). However, these regressions also predict that removing all of the characteristic features of nucleic acid—phosphates, bases, and sugars—will result in a reduced but finite traversal probability. This observation suggests that a component of these additional, non-charge interactions is not DNA-specific. We tested this surprising result by challenging the motor with a polymer lacking any resemblance to nucleic acid. Remarkably, the motor packages this insert as well (Figure 2-3d), revealing a non-specific component to the motor-DNA interaction. Such interactions may correspond to well-defined chemical contacts, i.e. hydrogen bonds or hydrophobic interactions, which do not require a specific nucleic acid moiety. Alternatively,
it is possible that these interactions are steric in nature. A prediction of such a steric drive mechanism is that important amino acids in DNA translocation can be replaced by any residue with a bulky side chain, a behaviour suggested by mutational analysis of the DNA binding loop of the related hexameric helicase SV40 (Shen et al., 2005).

Conclusion

Taken together, our data indicate that the motor-DNA interaction involves a wide variety of contacts with the full complement of nucleic acid moieties, as well as contacts not specific to DNA. In addition, our data suggest that the type of motor-DNA interaction changes during the course of the mechanochemical cycle. During the dwell phase, when ATP is loaded, the motor maintains strong, load-bearing contact with the DNA via ionic interactions with adjacent phosphates on the 5′-3′ strand. These contacts play a sensory role, coupling the mechanical and chemical cycles. In the burst phase, when the DNA is translocated, the motor makes contact with a variety of chemical moieties, charges on both strands, bases or sugars, and additional non-DNA-specific contacts. Given the relatively long duration of the contacts during the dwell phase, specific, strong, ionic interactions may be preferable. In contrast, contacts during active translocation in the burst phase must be made and broken more quickly; thus, it would be advantageous to make less specific, more promiscuous, weaker contacts. Figure 2-4 summarizes the important motor-DNA interactions and the specific point in the mechanochemical cycle at which these contacts are made.

Our findings have broad implications for the mechanism of the packaging motor. In particular, non-specific contacts may serve an important role in generating a step size that is a non-integer repeat of the chemical periodicity of DNA. Moreover, the observation that important contacts are made below distances of ~10 bp provides evidence against an open ring model of translocation in which only two subunits contact the DNA, limiting the mechanism by which a non-integer step size can be generated (Moffitt et al., 2009). In parallel, our findings have broader implications for the family of ASCE ring-ATPases. The use of alternate contacts during different portions of the mechanochemical cycle as well as the use of sensory contacts may serve as a general mechanism by which multimeric motors synchronize and coordinate the hydrolysis cycles of their individual subunits with the position of their substrate. The use of non-specific, perhaps steric, contacts may reflect the existence of a conserved translocation mechanism shared by nucleic acid and
polypeptide translocases of the ASCE superfamily; such non-specific contacts may have facilitated the evolution of peptide translocases from nucleic acid translocases (Mulkidjianian et al., 2007).

Materials and Methods

DNA substrates. For each construct, two (or in the case of the gaps, three) modified or unmodified DNA oligonucleotides were purchased from Fidelity Systems or IDT. The oligos hybridize to form a DNA segment with three nucleotide overhangs. 4187 and 4008 bp DNA fragments with distinct non-palindromic 3-nucleotide overhangs were prepared by PCR, restriction digest, and agarose gel purification. The primer for the 4008 bp PCR fragment contains a terminal biotin. These fragments were ligated to the insert oligos and the 8.2 kb product was gel purified.

Single-molecule packaging. Single-molecule packaging assays were performed as previously described (Rickgauer et al., 2008, Smith et al., 2001, Chemla et al., 2005, Fuller et al., 2007) in packaging buffer containing 50 mM Tris-HCl pH 7.8, 50 mM NaCl, 5 mM MgCl$_2$, and 1mM ATP unless mentioned otherwise. Briefly, the biotinylated DNA is bound to a 2.1 μm streptavidin-coated polystyrene bead held by a micropipette and the prohead-gp16-ATPase is attached to a second bead held in an optical trap. Packaging is initiated in situ by bringing the beads together (Rickgauer et al., 2008, Fuller et al., 2007). The instrument (Smith et al., 2003) is run in force-feedback mode, where the separation between the beads is adjusted to maintain constant tension in the DNA molecule. High resolution packaging measurements were conducted on a dual trap instrument described previously (Moffitt et al., 2006) with materials prepared in the same fashion as above, except 860 nm beads were used.

Analysis. The modified DNA insert is expected to reach the motor after packaging of 4.2 kb. However, variability in attachment geometry of the DNA to the beads and other systematic errors introduce uncertainty in the measurement of tether length, so the insert was considered to be between the 2.5 and 4.5 kb position. The longest pause in this interval was scored as the insert-induced pause. In traces in which a pause could not be identified a pause of 0.1 seconds, the temporal resolution, was assigned. The rates of pausing (0.017 kb$^{-1}$) and slipping (0.02 kb$^{-1}$) during normal packaging (Chemla et al., 2005) are small enough that these events do not significantly bias our measurements. All probabilities were calculated with the Laplace estimator (Lewis and Sauro, 2006) which is given by $p_{Laplace}=(x+1)/(n+2)$ where $x$ is the number of successes and $n$ is the number of trials. This is a better estimator
than the maximum-likelihood $\chi^2/n$, especially when $n$ is small or the probabilities being estimated are near zero or one (Lewis and Sauro, 2006). $n_{\text{min}}$ values are a measure of the degree to which a distribution is exponential and provide a strict lower limit on the number of rate-limiting kinetic events (Moffitt et al., 2009). These values were calculated by taking the ratio of the mean dwell time squared over the variance in the dwell times. $n_{\text{min}}$ error bars are standard deviations and were calculated via a bootstrap method.
Figure 2-1. Packaging of neutral DNA analogs. 

**a.** A prohead-motor complex bound to a microsphere is held in an optical trap while a micropipette holds a second microsphere bound to DNA containing a modified insert. A tether is formed and packaging is initiated when the beads are briefly brought into close proximity in the presence of ATP. 

**b.** Representative traces of DNA containing 10 bp ds-MeP packaged by the motor against a constant force of 5pN. Blue traces show traversal following a pause, and the red trace shows a terminal dissociation event following a pause. Inset are a schematic of the insert, with MeP nucleotides in red and unmodified nucleotides in blue, and the chemical structure of a MeP nucleotide.

**c.** Force and ATP dependence of traversal probability and pause duration of 10 bp ds-MeP inserts. 

**d.** Motor traversal probability of 30 bp ds-MeP and DNA-MeP hybrid inserts at 5 pN. 

**e.** Traversal probability of ds-MeP inserts at 5 pN force as a function of insert length. P-values (calculated using the two-tailed Fisher exact test) between 9 and 10 bp, and 10 and 11 bp are indicated. 

**f.** Translocation cycle length and footprint size limits from MeP length dependence. This scheme shows the position of a subunit that contacts the DNA before and after a full mechanochemical cycle, i.e. 10 bp. Contact with a single phosphate would produce a drop in traversal probability between 9 bp ds-MeP and 10 bp ds-MeP whereas contact with two phosphates would produce the observed drop between 10 bp ds-MeP and 11 bp ds-MeP. Regular phosphates are in blue, MeP in red, and the motor subunit contact loop in green. In c, d, and e the traversal probability is plotted using the Laplace best estimator (Lewis and Sauro, 2006), with 95% confidence intervals from the adjusted Wald method (Agresti and Coull, 1998), and the error bars of the pause durations are the SEM.
Figure 2-2. High resolution dynamics at neutral-DNA insert. a. Top: representative high resolution trace of the base-pair-scale dynamics at an insert composed of 10 bp of ds-Mep. Bottom: close up of a segment of this pause which shows that the dynamics consist of two classes of sub-pauses, upstream and downstream pauses, separated by attempts and punctuated by slips and repackaging events. b. Histogram of upstream pause durations. c. Histogram of downstream pause durations. d. Histograms of position changes in the upstream pause. The distributions in (c) and (d) have $n_{\text{min}}$ values—the ratio of the mean squared to the variance—of $1.1 \pm 0.1$ (s.d.) and $1.3 \pm 0.4$ (s.d.), respectively. $n_{\text{min}}$ provide a method for determining if a distribution is well-described by a single-exponential decay (Moffitt et al., 2009, Schnitzer and Block, 1995)—a single exponential decay has an $n_{\text{min}}$ value of 1—thus, these two distributions are well-described by single exponential decays. e. Histogram of distance between upstream and downstream pauses. f. A cartoon model of the dynamics of these pauses with average lifetimes and inter-conversion probabilities.
Figure 2-3. Substrate promiscuity. The force dependence of traversal probabilities and pause durations of different modified DNA substrates. Schematic illustrations of the inserts and the chemical structural modifications are shown. a. 10 nt ds-abasic phosphate backbone b. 20 nt ssDNA (poly-AC) on 5’-3’ and 3’-5’ strands c. 10bp bulge (poly-AC) on 5’-3’ and 3’-5’ strands d. ds-linker. In a-d the traversal probability is plotted using the Laplace best estimator (Lewis and Sauro, 2006), with 95% confidence intervals from the adjusted Wald method (Agresti and Coull, 1998), and the error bars of the pause durations are the SEM.
Figure 2-4. The Motor-DNA Contacts. The relative importance of the motor-DNA contacts as inferred from the traversal probabilities at 5 pN of all of the measured modifications (Figures 2-1&2-3) using a logistic regression (Appendix 1; Figure 2-5). The units of “contact importance” correspond to the inverse of the distance over which the removal of the specific moiety would reduce the traversal probability to 50%. The mechanochemical phase of the motor as it moves along the DNA is indicated to the right.
Figure 2-5. Logistic regression of the importance of contacts with different DNA moieties. See Appendix 1 for details of the regression analysis. SP—contacts with phosphates every 10 bp on the 5’-3’ strand. P35—contacts with phosphates in the 5’-3’ strand. P35—contacts with phosphates in the 3’-5’ strand. B35—contacts with bases and sugars in the 3’-5’ strand. a. Logistic regression using all data collected at 5 pN. Red bars — values for the coefficients of the model parameters (error bars are 95% c.i.). Blue symbols — Mean coefficients averaging the best 18 models, totalling 95% of the Akaike Weights. These results provide a ranking of the importance of different contacts during packaging. b. Experimental measurements (red) and model predictions (green) for the experiments at 5pN. The experimental result of the ds-linker was not included in the regression calculation. c. Small sample Akaike Information Criteria (AICc) values for regressions using P53, P35, B35 and a control parameter instead of SP. The control parameters represent phosphates removed with periodicities from 6 to 18 bp instead of SP. d. Regression using MeP inserts, at a variety of forces. The special phosphates have only a force-dependent weight, consistent with those contacts occurring during the dwell. The rest of the phosphates in both strands have also a zero-force component, likely a result of being contacted during the burst, when the motor generates force.
Figure 2-6. High Resolution Studies of Traversal of 5 bp of dsMeP. To probe the identity of the long, upstream pauses observed at high resolution in the crossing of 10 bp of dsMeP, we package shorter regions of neutral DNA, 5 bp of dsMeP. a. If the motor is situated on the DNA such that one dwell occurs upstream of the modified DNA by more than 5 bp from the start of the neutral insert, the 10-bp burst (Moffitt et al., 2009) will position the dwell within the neutral region. If the upstream pause corresponds to a dwell phase interaction with the neutral DNA, we expect to observe a pause of a long duration. b. However, if the motor is situated such that this dwell occurs less than 5 bp from the start of the neutral DNA, the 10-bp burst will carry the motor across the neutral DNA and the next dwell will be positioned on charged DNA. If the upstream pause corresponds to the dwell phase, then we would expect to observe no discernable pause in packaging. Since there is a slight variability in the 10-bp burst size (Moffitt et al., 2009), we expect that the motors will randomize their position on the DNA across the 4 kb of charged DNA which must be packaged before the insert. Thus, if the upstream pause corresponds to a dwell phase interaction with the neutral DNA, we expect to see only ~50% of the motors pause at the insert. c. Example packaging traces of 5 bp dsMeP using the high resolution optical tweezers under 1 mM [ATP] and ~15 pN of opposing load. Note that some traces show no long pause distinguishable from the normal distributions of dwell phase. d. The mean pause duration, the traversal probability, and the probability of observing a pause (N_{phage}=11; N_{pause}=5). Error bars for the probabilities are 95% confidence intervals calculated via the adjusted Wald method and the standard error of the mean for the pause duration. The observed pauses occurred at a mean position of 3840 ± 20 bp (s.d.); and a region of DNA 3 times the standard deviation in the position of the pauses was used as the candidate position of the insert. To account for the variability in dwell times on the DNA during normal packaging, we scored any event with durations of 300 ms or longer as a pause, a duration 3-fold larger than the typical dwell time on charged DNA. Because the probability of observing a pause is statistically inconsistent with 100%, we conclude the upstream pause corresponds to the dwell phase.
References


LEWIS, J. R. & SAURO, J. 2006. When 100% really isn’t 100%: improving the accuracy of small-sample estimates of completion rates. Journal of Usability studies, 1, 136-150.


Internal Force Resisting Viral DNA Packaging in Bacteriophage phi29.

Biophys J, 94, 159-67.


Chapter 3

The Bacterial Flagellar Switch is Sensitive to Proton Motive Force

This chapter contains previously unpublished research performed in collaboration with Richard Branch, Junhua Yuan, Howard C Berg, Carlos Bustamante, and Jan Liphardt.
Summary

The *Escherichia coli* flagellar switch controls the direction of flagellar motor rotation and allows cells to execute a biased random walk towards better environments. Traditionally, the motor’s ability to convert chemical information in the form of [CheY-P] into distinct rotational states has been considered separately from the motor’s ability to convert an electrochemical potential (the cellular proton motive force, PMF) into rotary motion. Here, we directly quantify the role of PMF in switching. We have used *E. coli* cells whose PMF can be reversibly controlled via the light-powered proton pump proteorhodopsin. Using a constitutively active CheYD13K strain to avoid confounding cellular PMF-dependent processes, we found that the motor switches less frequently at reduced PMF, in part due to dissociation of stators at low PMF. The switch frequency with three or fewer stators is extremely low; above that number switching is more sensitive to changes of PMF level than to changes of stator number. Therefore, the switch rates are not simply proportional to velocity, but rather are differentially sensitive to the structural dynamics of the motor and to the external driving potential. Moreover, when PMF is reduced, cooperativity—as measured by the Hill coefficient—falls approximately two-fold: the switch becomes less sensitive to extreme CheY-P levels and moves to more equal occupancy of the CCW and CW states. Therefore, while the bacterial flagellar motor primarily converts energy in the form of a chemical potential into mechanical work, it may also use the PMF to boost its sensitivity to chemical signal inputs.

Introduction

*Escherichia coli* bacteria propel themselves by spinning helical flagella with rotary molecular motors (Berg, 2003). When motors rotate counterclockwise (CCW) cells swim straight, and when a motor switches direction to clockwise (CW) the cell tumbles. By modulating their tumbling frequency cells are able to execute a biased random walk and move to more favorable environments. Concentration changes of small molecules are detected by chemoreceptors and a signal is transmitted to motors via the phosphorylated response regulator CheY-P. The network of chemotaxis proteins that controls [CheY-P] has been well characterized, but the motor’s switch complex remains an active area of research.

The switch is a particularly interesting system because it has an extremely high gain; that is, a small incremental change in [CheY-P] causes a large change
in the output of the switch. The steepness, or cooperativity, of this response can be quantified with the Hill coefficient, which Cluzel et al measured to be 10 (Cluzel et al., 2000). More recently it has become apparent that the switch adapts by modulating the number of binding sites for CheY-P (FliM units), and when that is taken into account, the Hill coefficient for a fixed number of FliM units is as high as 21, which is over five-fold greater than that of the next largest allosteric protein complex (Yuan and Berg, 2013). Moreover, the largest measured value of binding cooperativity of CheY-P to FliM is only 2 (Sourjik and Berg, 2002). The existing model of conformational spread would require a CheY-P binding cooperativity over 10 to result in a Hill coefficient of 21, so the source of the high Hill coefficient is not known.

A powerful technique to better understand the mechanism of the switch is to study how the switch activity responds to a range of physical parameters. Researchers have been doing this for many years, and interestingly, in addition to the canonical sensitivity to [CheY-P], switching of the motor is affected by changes in temperature (Turner et al., 1996), load (Fahrner et al., 2003), and proton-motive-force (PMF) (Khan and Macnab, 1980). However, the previous measurements of switching versus PMF were limited by being slow and irreversible, and they were also done in cells with intact chemotaxis signalling networks that could potentially respond to PMF, which means that their results did not provide a clear picture of pure switch activity.

The development of the light-powered proton pump proteorhodopsin as a tool to manipulate PMF in E. coli (Walter et al., 2007) makes it possible to study in vivo cellular responses to PMF in a controlled and reversible manner. The flagellar motor is a particularly appealing system because previous measurements established that its speed of rotation is proportional to PMF (Gabel and Berg, 2003), so motor rotation speed is a convenient reporter of PMF. Past experiments that depleted PMF in a variety of ways showed that switching of the flagellar motor is sensitive to PMF: at low PMF, motors in wildtype cells spin nearly exclusively counterclockwise (CCW), corresponding to straight-swimming (Khan and Macnab, 1980). We decided to use proteorhodopsin to achieve finer and reversible control of PMF to better understand this sensitivity. To isolate motor effects from fluctuations in activity of the chemotaxis network, we used a set of chemotaxis mutants that were not known when the first PMF switching experiments were conducted. These mutants have fixed concentrations of activated CheY, so the switch complex receives a constant chemical signal and any observed changes in switch rates should be due to inherent properties of the motor.
Results and Discussion

To visualize the motor’s action we tethered 1 µm polystyrene beads to sheared “sticky” (Scharf et al., 1998) flagellar stubs on cells stuck to a glass coverslip (Figure 3-1a) and imaged the beads with darkfield illumination and a ~700 Hz EMCCD camera (Figure 3-1b, c). We first reduced PMF by sealing cells in an airtight chamber of lactate-containing chemotaxis buffer so that they would exhaust the available oxygen and not be able to maintain PMF via respiration. Once the PMF started to drop (as measured by speed of rotation of the beads), we induced proteorhodopsin proton-pumping with a green excitation light and were able to restore PMF. (Beads are imaged with red light to avoid excitation of proteorhodopsin.) The intensity of the green light can be adjusted to reversibly control PMF (Figure 3-1d). Under these conditions the PMF change upon increasing light intensity is fast (<0.1 s), whereas the decrease of the PMF is considerably slower (~30-60 s).

While conducting these experiments we observed a previously reported phenomenon: at reduced PMF the torque-generating stators are less stable on the motor (Fung and Berg, 1995, Tipping et al., 2013). When a cell is taken from high to low back to high PMF by modulating the excitation light intensity, the velocity does not always fully recover to its initial value (Figure 3-1e). If the light is then left at high intensity the velocity sometimes recovers in discrete steps of the same magnitude, consistent with stepwise re-engagement of stators. (With the load of a 1 µm bead the additional velocity per stator at normal PMF is a clearly discernible ~5.6 Hz). To isolate PMF effects from stator effects and to sample a range of PMFs we used light sequences consisting of repeated on-off light intervals that were fast enough (30-60 s) that most of the time stators did not disengage.

To isolate motor switching effects due to changes in PMF from concomitant alterations in chemotaxis network activity, we used two parallel approaches. The first used a mutant strain with ΔcheZΔcheB (lacking the CheY phosphatase and the methylesterase that competes with CheY for phosphate and inactivates the receptor kinase respectively), and CheY under control of a trc promoter, which was previously estimated to keep >99% of the CheY phosphorylated (Alon et al., 1998). As a result the signal to the motor is solely controlled by the amount of expressed CheY. In the second approach we used a strain that expresses cheYD13K (CheY*), which contains a point mutation that locks CheY into the activated state, even in the absence of phosphorylation (Korobkova et al., 2004). We expressed CheY* in cells lacking the chemotaxis
operon (ΔcheA-Z) to further ensure that there was no cryptic interference from the chemotaxis network.

Interestingly, these two “motor-isolated” strains displayed different behavior: the ΔcheZΔcheB strain behaved the same as previously reported wildtype cells: motors spin almost exclusively CCW at low PMF (Figure 2a), whereas the CheY* strain did not have a strong CCW bias at reduced PMF (Figure 3-2b). We suspect that this is because the ΔcheZ ΔcheB strain was not truly motor-isolated at low PMF. Even lacking the CheZ phosphatase, CheY has auto-dephosphorylation activity with a half-life of less than 30 s (Lukat et al., 1990, Hess et al., 1988). Since PMF depletion happens on the same timescale, [CheY-P] likely dwindles as it auto-dephosphorylates at lowered PMF. At reduced PMF, the FₒF₁ ATP synthase runs in reverse, consuming ATP to pump protons out of the cell in an attempt to maintain PMF. ATP is the source of the phosphate on CheY-P, so we hypothesized that ATP depletion reduces the rate of CheY phosphorylation (Figure 3-2d). We also suspected that this consumption of ATP is the reason that the PMF decreases much more slowly than it increases. We tested this hypothesis by constructing a strain identical to the ΔcheZΔcheB strain, with the addition of a ΔatpB-C deletion, so that it lacked ATP synthase. Confirming our hypothesis, the ΔatpB-C cells no longer have the strong CCW bias at low PMF, and they also show a much faster PMF decrease than cells with intact ATP synthase (Figure 3-2c). Since the [CheY-P] in the ΔcheZΔcheB strain dwindles at low PMF, altered switching at low PMF in this strain is not purely due to the motor. For this reason we focus on the CheY* strain for the rest of this work.

Next, we studied the low PMF switching behavior in more detail. As the PMF decreases, the motor switches less frequently, reaching run times of many tens of seconds between switch events (Figure 3-3a). As previously mentioned, at low PMF stators dissociate, so we sought to separate effects of reduced PMF and stator number. We first examined switching as a function of PMF by repeatedly turning the excitation light on for 30 s and off for 60 s. In these experiments cells swept out a range of ~50-100% relative PMF, and the time spent at low PMF was short enough that stator dissociation events were rare, as evidenced by immediate full recovery of velocity. As the motor's velocity decreased, switch events became less frequent, as switching rates for both directions decreased (CCW->CW and CW->CCW) (Figure 3-3b). Over the range examined with these ‘mild’ on-off light sequences the switching rates in both directions were roughly proportional to PMF (Figure 3-3c). The two switching rates changed in tandem, so that the bias did not change significantly with PMF.
We also observed switching as a function of stator number by allowing cells to fall below \(~30\%) of their normal PMF for a few minutes and then suddenly increasing excitation light to maximum intensity. The PMF increases in \(< 0.1 \text{ s}\), whereas the stators re-associate in tens of seconds, resulting in staircase-shaped velocity vs. time plots (Figure 3-3d). Each flat interval corresponds to the motor running at a constant velocity with a constant number of stators; the velocity jumps in multiples of about \(5.6 \text{ Hz}\) when additional stators incorporate into the motor. (The motor can accommodate up to at least 11 stators (Reid et al., 2006).) These experiments allowed us to observe switching at a range of different numbers of engaged stators at constant PMF. Notably, switching was rare at stator numbers below four, with runs of tens of seconds in both directions. (Bias is poorly defined in this extremely low switch rate regime, though we found that motors were more likely to rotate CW.) Once the fourth stator entered the motor, it immediately began to switch, and switching rate was only mildly influenced by increasing stator number (Figure 3-3e). Because stator dissociation and re-association is stochastic and not always reversible, it was more difficult to collect data for switching vs. stator number than for switching vs. PMF, so the exact dependence of switching on stator number is unclear beyond the distinct rare switching at 1 to 3 stators.

A possible explanation for the sensitivity of switch rates to PMF and stator number is one previously suggested by Fahrner et al. (Fahrner et al., 2003): the switch may be sensitive to proton flux. (The motor proton flux is likely tightly coupled to rotation velocity (Meister et al., 1987).) This concept would entail a simple mechanical coupling between rotation and the switch. We are able to at least partially test this idea by comparing switching in regimes where the velocity is reduced, either by PMF or stator number, to a similar value. To minimize uncertainty from cell-to-cell variability we made comparisons within traces of single cells (Figure 3-4a). The medium PMF, high stator region (region 2) has lower switching rates than the medium stator, high PMF region (region 3), which can be seen in the dwell time histograms (Figure 3-4b). A two-sample Kolmogorov–Smirnov test rejects the null hypothesis that the dwell time distributions from these two regions of this trace are from the same underlying distribution with \(p < 10^{-4}\) for CCW and \(p < 10^{-5}\) for CW. The same result of lower switch rates at reduced PMF (region 2) than at reduced stator number (region 3) is present in 22 out of 23 cells where such comparisons were possible. Kolmogorov–Smirnov tests comparing the distributions from regions 2 and 3 of the 23 cells yields \(p < 0.01\) in 21 of 23 for CCW and 22 of 23 for CW. Combined histograms are shown in Figure 3-4d and e, and as expected the Kolmogorov–Smirnov \(p\) values of the combined
distributions are extremely low: $p < 10^{-169}$ for CCW and $p < 10^{-234}$ for CW. We conclude that the switch is more sensitive to PMF than it is to stator number, and that these changes in switching rate cannot simply be explained by a single mechanical coupling that depends on rotation velocity.

So far we have focused on the intermediate bias regime where switching rates are more easily measured, as excessively long rotation-state dwell times are rare. However, we were also curious about the switch’s sensitivity to PMF in cells with more extreme bias (cells with high or low [CheY-P]), based on Tu’s (Tu, 2008) suggestion that the ultrasensitivity of the switch (Hill coefficient of $\sim 10$) (Cluzel et al., 2000) is due to a non-equilibrium process that consumes energy provided by the PMF. To test this notion, we examined cells with either high or low CCW biases at normal PMF, due to low or high [CheY*], respectively. We subjected these CheY* cells to the repeated on-off light sequences that minimize stator dissociation to isolate the PMF sensitivity. When the PMF was decreased, the bias of cells with high starting bias decreased to more intermediate values, while the bias of cells with low starting bias increased to more intermediate values (Figure 3-5a). These observations show that the motor becomes less sensitive to [CheY-P] at reduced PMF, and therefore, the motor’s bias relaxes to more equal occupancy of the CW and CCW states. At $\sim 70\%$ relative PMF the decrease in sensitivity corresponds to a reduced Hill coefficient of $\sim 6$ (Figure 3-5b), which is still highly cooperative, but nearly a two-fold reduction compared to the value at full PMF. It is likely that the cooperativity drops further when the PMF is reduced to below 70%, but we have been unable to collect sufficient data in that operating regime due to stator dissociation, which as discussed, results in a sudden but reversible decrease in switching rates, precluding reliable measurement of the motor’s bias. Ultimately, while we do not have a definitive demonstration that the switch uses energy from the PMF to increase the steepness of its response to [CheY-P], our results are suggestive.

Another, non-mutually exclusive, explanation of these results uses the framework of the classic allosteric Monod-Wyman-Changeux (MWC) model (Monod et al., 1965), which has previously been used to model the flagellar switch (Alon et al., 1998):

$$B = \frac{(1 + \frac{Y}{K})^N}{L \left(1 + \frac{Y}{CK}\right)^N + \left(1 + \frac{Y}{K}\right)^N}$$

Where $B$ is CCW bias, $Y$ is [CheY-P], $N$ is the number of switch subunits, $K$ is the CCW CheY-P dissociation constant, $C$ is the ratio of CW and CCW
CheY-P dissociation constants, and $L$ is the CCW to CW equilibrium constant in the absence of CheY-P. $L$, $K$, and $C$ can be expressed in terms of the relative free energy difference, $E_a$, between CW and CCW states for a single subunit (Duke et al., 2001):

$$L = e^{NE_a/k_BT}$$
$$K = Y_{1/2}e^{-E_a/k_BT}$$
$$C = e^{2E_a/k_BT}$$

Where $Y_{1/2}$ is the [CheY-P] that yields $B = 0.5$. ($Y_{1/2} \propto e^{-E_L/k_BT}$ where $E_L$ is the [CheY-P] binding energy in the absence of conformational change.)

In this model a change in cooperativity can arise from a change in $E_a$ (Figure 3-5b, inset). Since the switch is in close proximity to the inner membrane (Francis et al., 1992), it is possible that electric field from the PMF could influence the energy landscape of the switch proteins. In addition to the relative free energy of the conformations, the electric field could also affect the relative free energy of the transition state, which would lead to the observed reduction in both switching rates at low PMF.

Another possible factor within the MWC model that could change cooperativity is the number of proteins in the switch ring, as the ring has been shown to be dynamic and the number of subunits can change (Yuan et al., 2012). However, this possibility is less likely to explain our results than the electric field because the changes in switching rates are instantaneous, whereas the previously observed FliM exchange rate is of the order of 10 minutes (Delalez et al., 2010, Fukuoka et al., 2010), and the signal dependent FliM occupancy change required many 10s of seconds (Yuan and Berg, 2013). One could suggest that there is a different, much faster, type of FliM dissociation and reassociation when PMF is changed, but that seems less likely. Moreover, in the recent experiments that observed switching versus changes in FliM number, the response curve shifted to the left in addition to changing slope, whereas in the PMF results we only see the change in slope. An interesting side note about that result is that in order for $N$ to shift the curve left, $L$ must not actually be a function of $N$. This can be interpreted with the above equations if the main free energy conformation change does not occur in FliM, but rather in the more stable protein of the switch, FliG. In that case $L$ would be a function of the number of FliG units instead of the number of FliM units. Because of the fast response to PMF increases and because the response curve changes differently from previously measured response to changing FliM occupancy, we
do not believe that the change in switching rates can be explained by a simple change in motor composition.

Conclusion

These measurements of switching of the flagellar motor at reduced PMF show that the electrochemical proton gradient that drives motor rotation also changes the switching frequency and the sensitivity of the motor to CheY-P. The changes in switching as the motor slows down due to PMF are similar to Fahrner et al.’s measurement of switching versus load (Fahrner et al., 2003), where switching rates decreased at high load (low velocity), and cells with a strong CCW bias shifted to a more moderate CCW bias. Therefore it seemed possible that the same mechanism is responsible for the PMF and load dependence. However, in this work we show that the dependence of switching on stator number is not as strong as that of PMF for similar velocities, so there must be some other contributing factor besides rotation speed. What this factor may be, as well as the mechanism by which the switch uses PMF to increase its sensitivity and frequency awaits further theoretical and experimental study.

Materials and Methods

Strains and Plasmids All strains were descended from AW405. Deletions were generated with the PCR fragments and the λ-Red system (Datsenko and Wanner, 2000), as well as P1vir phage transduction from Keio collection knockouts (Baba et al., 2006). Plasmids were constructed via PCR and sequence and ligase independent cloning (SLIC) (Li and Elledge, 2007).

Plasmids:
pBAD-PR: pBAD-Proteorhodopsin, amp-R pBR322 origin.
pATP1: fliC<sup>x</sup>, pBAD-PR, ampicillin-R, pBR322 origin.
pATP5: lac-cheYD13K, fliC<sup>x</sup>, lacI, chloramphenicol-R, p15A origin.

Strains:
VSJ85: gift from Karen Fahrner; AW405 ΔcheZ ΔcheB and trc-cheY.
ATP25: ATP18 ΔcheA-Z + pATP5 + pBAD-PR.

Cell Culture Cells were grown from a 1:100 dilution of overnight culture shaking in T-broth at 33 °C for ~3 hours to an OD<sub>600</sub> of 0.3 and then 0.1%
Arabinose, 20-200μM IPTG, and 10μM retinal were added, and cells were returned to the incubator for another 1 hour.

**Sample Preparation** Cells were harvested by centrifugation at 2,500g and washed with chemotaxis buffer (10 mM potassium phosphate, 0.1 mM EDTA, 10mM lactic acid, pH 7.0). Flagella were sheared by passing solution through two 25-gauge needles 100 times. Coverslips were prepared by coating with 0.01% poly-L-lysine, and a chamber was formed with two strips of double-stick tape and a second coverslip. Cells were added to the chamber, incubated for 2 minutes, rinsed, then 1.0 μm polystyrene beads (Bangs Labs) were added, incubated 2 minutes, and rinsed. The chamber was sealed with 5-minute epoxy.

**PMF Control and Imaging** PMF started to decline after 15-30 minutes, at which point it was restored with excitation light. Green excitation light came from a 175 W Xenon bulb passed through a 540/25 filter. Imaging light came from a Zeiss 1.2/1.4 N.A. darkfield condenser passing red light through a 700/30 filter. Imaging light was separated from excitation light with a 565 longpass dichroic filter. Excitation light was controlled with a combination of neutral density filters and a variable aperture Sutter shutter. Maximum light intensity on the sample was approximately 100mW/cm². The sample was imaged with a Nikon 60x .5-1.25 variable NA objective onto an Andor iXon camera run at ~700 Hz.

**Analysis** Movies of beads were converted to angular velocity traces and switching events were detected with custom Matlab software using a two-threshold crossing algorithm (Yuan et al., 2009).
Figure 3-1. Light control of PMF and measurement of flagellar motor rotation. a. *E. coli* expressing proteorhodopsin can be powered by light. Excited proteorhodopsin pumps protons out of the cell, and protons re-enter via the flagellar motor, causing it to rotate. Rotation of the motor is monitored by tethering 1-µm beads to sheared flagellar stubs. b. Beads are imaged with dark field microscopy and an EMCCD running at ~700 Hz, and the centroid is found (red circle). The cell scatters much less light than the bead, so only the bead is visible in this image. c. Track of centroid positions for 10 seconds of rotation. d. Angular speed as proteorhodopsin excitation light is turned on (green background) and off (white background) repeatedly. e. Angular speed after excitation light is decreased to a low level (faint green background) and then increased. In this trace stator subunits dissociated at low PMF and can be seen re-associating at high PMF in a stepwise manner in the second half of the trace. (The dissociation events are partially masked by the rapid PMF change.)
Figure 3-2. Responses of different strains to changes in PMF. 

a. \( \Delta \text{cheZ} \Delta \text{cheB} \). Nearly exclusive CCW rotation at reduced PMF (similar result found in n=16 cells). 

b. \( \text{cheYD13K} \Delta \text{cheA-}Z \). No large change in CCW bias at reduced PMF (similar result found in n=16 cells). 

c. \( \Delta \text{cheZ} \Delta \text{cheB} \Delta \text{atpB-}C \). Much faster velocity decrease, but no large change in CCW bias (similar result found in n=10 cells). 

d. Proposed explanation of strong CCW bias in \( \Delta \text{cheZ} \Delta \text{cheB} \) and wildtype strains: low PMF leads to low [ATP], which leads to low [CheY-P], and ultimately extreme CCW bias.
Figure 3-3. Switching at a range of PMF values and stator numbers. 

a. Velocity trace as PMF decreases; in this case, PMF and stator number are both decreasing. Dwell times shown in red: CCW filled circles, CW open circles. 

b. Repeated on-off excitation light leads to repeatable sweep of PMF values, with unchanging stator number, [without spending too much time at low PMF, which leads to stator dissociation.] In blue (CW) and light blue (CCW) are switching rates calculated for 5 second bins, and CCW bias is shown in red. 

c. Switching rates and bias calculated for the range of PMF values for 9 cells with the same excitation light sequence as shown in (b). Open circles are means across cells, weighted by the standard errors of the mean shown as error bars. Relative PMF values are displayed as velocity-equivalent stator units (using 5.6 Hz/stator). Both switching rates are roughly proportional to PMF, and the bias does not change significantly as a function of PMF. 

d. Excitation light increase following a few minutes at low PMF sweeps out different stator numbers at full PMF, as the stators reengage. 

e. Switching rates and bias calculated for the range of stator values for 10 cells with a similar excitation light sequence as shown in d. (Because stator re-association is stochastic and takes longer than PMF changes, the data is more sparse than the switching versus PMF data shown in (b) and (c).)
Figure 3-4. Long time course of cheYD13K at different PMF values and stator numbers. 

a. Excitation light intensity is indicated by intensity of green background. At medium light intensity the relative PMF falls to 60%, as the stator number remains unchanged (indicated by the immediate full recovery of velocity in the second high excitation light interval). At low light intensity the relative PMF falls to 20%, and the engaged stator number falls from 8 to 2. In the final high excitation light interval the PMF immediately recovers, and the stators re-engage in a stepwise manner. Inset: zoom of region of light increase. Velocity increase occurs in < 0.1 s, and the switch rate increases. 

b. Dwell time histograms of the region bounded by dotted lines in part a., CCW in blue, CW in light blue, with mean and standard error of the mean indicated in red. 

c. Switch rates for 20 second bins of the time course shown in (a). 

d. Combined CCW dwell-time distributions for 23 cells for max PMF/max stator number (panel (a) region 1) compared to medium PMF (region 2) and medium stator number (region 3). Cells were only included if they contained a region of at least 20 seconds at max-PMF medium-stator-number (region 3) that had velocity less than or equal to the medium-PMF max-stator-number region (region 2). 

e. Combined CW dwell time distributions for same cells as (d).
Figure 3-5. CCW bias versus PMF. a. Bias is plotted as a function of relative PMF using CheY* cells and the on-off light sequences used in Figure 3-2 and Figure 3-3b. The regions with no excitation light and decreasing PMF were divided into 5 equal sections, for which the bias and relative PMF were calculated. Each light line corresponds to a single cell, and they are colored red, green, and blue according to their high (10 cells) medium (16 cells) and low (20 cells) CCW bias at maximum PMF (which corresponds to low, medium, and high [CheY*] respectively). The bold lines are the mean of each group of cells. Cells with such high or low [CheY*] that they never switched did not begin to switch at reduced PMF, and were not included in the averages. b. The dark black line is the fit to the switch sensitivity measured in Cluzel et al., 2000, with a hill coefficient of 10. The dashed line is an estimate of the sensitivity of the motor at a relative PMF of 70%, based on the data shown in (a), and it has a hill coefficient of 6. The inset colored free energy diagrams show a way the relative free energies of the two rotational states and the transition state may change at reduced PMF (dashed line) that would lead to the observed slopes depicted in (a).
References


YUAN, J., BRANCH, R. W., HOSU, B. G. & BERG, H. C. 2012. Adaptation 
Appendix I

Tables and Logistic Regression of Phage Data

The material presented in this appendix was adapted with permission from

Table 1: List of Insert Oligos

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Strand</th>
<th>Sequence (from 5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 bp ds-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGCCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>9 bp ds-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCCTCAGAACTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>CCATGGTCAGTGAGCCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>10 bp ds-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>11 bp ds-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>CCATGGTCAGTTCTCCTCCTC</td>
</tr>
<tr>
<td>15 bp ds-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGCCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>30 bp ds-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGCCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>5’-3’ 15 bp Hybrid-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>5’-3’ 30 bp Hybrid-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>3’-5’ 30 bp Hybrid-MeP</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>Linker</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>10 PO4 Abasic</td>
<td>Top</td>
<td>GAGAACGTGACTCACACTGCCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>5’-3’ 10 nt ss-gap</td>
<td>Top-Left</td>
<td>CGCGACTGACGTGACTGTC</td>
</tr>
<tr>
<td></td>
<td>Top-Right</td>
<td>CAGTTGTGACGTGACTGTC</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>5’-3’ 20 nt ss-gap</td>
<td>Top-Left</td>
<td>CGCGACTGACGTGACTGTC</td>
</tr>
<tr>
<td></td>
<td>Top-Right</td>
<td>CAGTTGTGACGTGACTGTC</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCGACTGTGAGTATCAGTTCTCCTC</td>
</tr>
<tr>
<td>3’-5’ 20 nt ss-gap</td>
<td>Bottom-Left</td>
<td>GCACAGTCACTGTCAGTTGTC</td>
</tr>
<tr>
<td></td>
<td>Bottom-Right</td>
<td>GCACAGTCACTGTCAGTTGTC</td>
</tr>
<tr>
<td>5’-3’ 10 nt bulge</td>
<td>Top</td>
<td>GAGAACATGGCAGCTGACGTGCACATCGCAGCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCTTATGTGACTGAGCCATGTTCTCCTC</td>
</tr>
<tr>
<td>5’-3’ 20 nt bulge</td>
<td>Top</td>
<td>GAGAACATGGCAGCTGACGTGCACATCGCAGCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCTTATGTGACTGAGCCATGTTCTCCTC</td>
</tr>
<tr>
<td>3’-5’ 10 nt bulge</td>
<td>Top</td>
<td>GAGAACATGGCAGCTGACGTGCACATCGCAGCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCTTATGTGACTGAGCCATGTTCTCCTC</td>
</tr>
<tr>
<td>3’-5’ 20 nt bulge</td>
<td>Top</td>
<td>GAGAACATGGCAGCTGACGTGCACATCGCAGCTCAGTGATAGT</td>
</tr>
<tr>
<td></td>
<td>Bottom</td>
<td>ATCAGTGAGGGCTTATGTGACTGAGCCATGTTCTCCTC</td>
</tr>
</tbody>
</table>

Modifications are in red, MeP in red/bold. The top strand is defined as the 5’-3’ strand in the direction of packaging.
Table 2: Summary of results from all inserts

<table>
<thead>
<tr>
<th>Modification</th>
<th>Substrate Name</th>
<th>Force (pN)</th>
<th>Total # phages</th>
<th>Successful Traversal</th>
<th>Traversal Probability,*</th>
<th>95% c.i.**</th>
<th>Mean Pause (s)</th>
<th>SEM (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Charges</td>
<td>5 bp ds-Mep</td>
<td>5</td>
<td>12</td>
<td>12</td>
<td>0.93</td>
<td>0.79 - 1.00</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>5 bp ds-Mep</td>
<td>16 ± 2*</td>
<td>11</td>
<td>11</td>
<td>0.92</td>
<td>0.77 - 1.00</td>
<td>1.5</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>5 bp ds-Mep</td>
<td>20</td>
<td>11</td>
<td>11</td>
<td>0.92</td>
<td>0.77 - 1.00</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>5 bp ds-Mep</td>
<td>5</td>
<td>20</td>
<td>16</td>
<td>0.77</td>
<td>0.58 - 0.93</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>10 bp ds-Mep</td>
<td>5</td>
<td>103</td>
<td>84</td>
<td>0.81</td>
<td>0.73 - 0.88</td>
<td>3.3</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>10 bp ds-Mep</td>
<td>11 ± 4*</td>
<td>9</td>
<td>5</td>
<td>0.55</td>
<td>0.27 - 0.81</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>10 bp ds-Mep</td>
<td>20</td>
<td>27</td>
<td>7</td>
<td>0.28</td>
<td>0.13 - 0.45</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>11 bp ds-Mep</td>
<td>5</td>
<td>26</td>
<td>12</td>
<td>0.46</td>
<td>0.29 - 0.65</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15 bp ds-Mep</td>
<td>2.5</td>
<td>11</td>
<td>4</td>
<td>0.38</td>
<td>0.15 - 0.65</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>15 bp ds-Mep</td>
<td>5</td>
<td>33</td>
<td>11</td>
<td>0.34</td>
<td>0.20 - 0.50</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>15 bp ds-Mep</td>
<td>10</td>
<td>21</td>
<td>5</td>
<td>0.26</td>
<td>0.10 - 0.45</td>
<td>31</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>30 bp ds-Mep</td>
<td>5</td>
<td>22</td>
<td>0</td>
<td>0.04</td>
<td>0.00 - 0.13</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>30 bp ds-Mep</td>
<td>0.5</td>
<td>19</td>
<td>2</td>
<td>0.14</td>
<td>0.02 - 0.35</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>30 bp ds-Mep</td>
<td>1</td>
<td>44</td>
<td>3</td>
<td>0.09</td>
<td>0.02 - 0.19</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>10 bp ds-Mep, 25μM ATP</td>
<td>5</td>
<td>43</td>
<td>22</td>
<td>0.51</td>
<td>0.37 - 0.65</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5' 3' 30 bp hybrid MeP</td>
<td>5</td>
<td>38</td>
<td>3</td>
<td>0.1</td>
<td>0.02 - 0.22</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>5' 3' 30 bp hybrid MeP</td>
<td>1</td>
<td>23</td>
<td>12</td>
<td>0.52</td>
<td>0.33 - 0.71</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3' 5' 30 bp hybrid MeP</td>
<td>5</td>
<td>25</td>
<td>22</td>
<td>0.85</td>
<td>0.69 - 0.97</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>3' 5' 30 bp hybrid MeP</td>
<td>20</td>
<td>17</td>
<td>9</td>
<td>0.53</td>
<td>0.31 - 0.74</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>5' 3' 15 bp hybrid MeP</td>
<td>5</td>
<td>18</td>
<td>8</td>
<td>0.45</td>
<td>0.25 - 0.66</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>No bases, no sugars</td>
<td>10 Phosphate-Abasic</td>
<td>5</td>
<td>27</td>
<td>25</td>
<td>0.9</td>
<td>0.76 - 0.99</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>10 Phosphate-Abasic</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>0.77</td>
<td>0.51 - 0.96</td>
<td>74</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>10 Phosphate-Abasic</td>
<td>20</td>
<td>12</td>
<td>1</td>
<td>0.14</td>
<td>0.00 - 0.38</td>
<td>112</td>
<td>N/A</td>
</tr>
<tr>
<td>Gaps</td>
<td>5' 3' 20 nt ss-gap</td>
<td>5</td>
<td>17</td>
<td>8</td>
<td>0.47</td>
<td>0.26 - 0.69</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>5' 3' 20 nt ss-gap</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>0.47</td>
<td>0.23 - 0.71</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>3' 5' 20nt ss-gap</td>
<td>5</td>
<td>27</td>
<td>21</td>
<td>0.76</td>
<td>0.59 - 0.90</td>
<td>1.3</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>3' 5' 20nt ss-gap</td>
<td>10</td>
<td>25</td>
<td>21</td>
<td>0.81</td>
<td>0.65 - 0.94</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5' 3' 10 nt ss-gap</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>0.73</td>
<td>0.44 - 0.95</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>Bulges</td>
<td>5' 3' 10 nt bulge</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>0.8</td>
<td>0.51 - 1.00</td>
<td>1.7</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>5' 3' 20 nt bulge</td>
<td>5</td>
<td>11</td>
<td>7</td>
<td>0.62</td>
<td>0.35 - 0.85</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5' 3' 10 nt bulge</td>
<td>20</td>
<td>11</td>
<td>9</td>
<td>0.77</td>
<td>0.51 - 0.96</td>
<td>2.6</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>5' 3' 20 nt bulge</td>
<td>20</td>
<td>13</td>
<td>6</td>
<td>0.47</td>
<td>0.23 - 0.71</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>3' 5' 10 nt bulge</td>
<td>5</td>
<td>14</td>
<td>11</td>
<td>0.75</td>
<td>0.52 - 0.93</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>3' 5' 20 nt bulge</td>
<td>5</td>
<td>25</td>
<td>12</td>
<td>0.48</td>
<td>0.30 - 0.67</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>3' 5' 10 nt bulge</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>0.83</td>
<td>0.54 - 1.00</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>3' 5' 20 nt bulge</td>
<td>20</td>
<td>12</td>
<td>7</td>
<td>0.57</td>
<td>0.32 - 0.81</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Non biological</td>
<td>Linker</td>
<td>5</td>
<td>33</td>
<td>32</td>
<td>0.94</td>
<td>0.83 - 1.00</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Linker</td>
<td>20</td>
<td>24</td>
<td>23</td>
<td>0.92</td>
<td>0.78 - 1.00</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Linker</td>
<td>30</td>
<td>27</td>
<td>19</td>
<td>0.69</td>
<td>0.54 - 0.84</td>
<td>1.5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

*Traversal probability calculating using the Laplace estimator (Lewis)  
**95% confidence interval calculated using the adjusted Wald method (Agresti and Coull, 1998)  
# These measurements were made using a high resolution dual trap instrument without force feedback. The error bar on the force is the standard deviation in the average force at the insert (Moffitt et al., 2009)
Table 3: Predicted force dependence of pauses

<table>
<thead>
<tr>
<th>Modification</th>
<th>Contour length (nm)</th>
<th>Force-1 (pN)</th>
<th>Extension-1 (nm)*</th>
<th>Pause-1 (s)</th>
<th>Force-2 (pN)</th>
<th>Extension-2 (nm)*</th>
<th>Pause-2 (s)</th>
<th>Measured pause ratio</th>
<th>Predicted Pause ratio**</th>
</tr>
</thead>
<tbody>
<tr>
<td>10bp ds-Mep</td>
<td>3.4</td>
<td>5</td>
<td>3.1</td>
<td>3.3</td>
<td>20</td>
<td>3.3</td>
<td>7.4</td>
<td>2.1±1</td>
<td>16,000</td>
</tr>
<tr>
<td>15bp ds-Mep</td>
<td>5.1</td>
<td>5</td>
<td>4.7</td>
<td>34.8</td>
<td>10</td>
<td>5.0</td>
<td>30.6</td>
<td>0.9±0.6</td>
<td>160</td>
</tr>
<tr>
<td>5'-3' 20nt ssDNA</td>
<td>14.7</td>
<td>5</td>
<td>6.3</td>
<td>19.8</td>
<td>10</td>
<td>8.9</td>
<td>36.1</td>
<td>2±2</td>
<td>290,000</td>
</tr>
<tr>
<td>3'-5' 20nt ssDNA</td>
<td>14.7</td>
<td>5</td>
<td>1.0</td>
<td>12.4</td>
<td>10</td>
<td>1.9</td>
<td>74.0</td>
<td>6±5</td>
<td>19</td>
</tr>
<tr>
<td>ds-10-PO4-Abasic</td>
<td>7.7</td>
<td>5</td>
<td>0.5</td>
<td>30</td>
<td>2.1</td>
<td>1.5</td>
<td>3.4±0.8</td>
<td>568,650</td>
<td></td>
</tr>
</tbody>
</table>

* An estimate of the minimum value of the extensions of inserts was calculated from the extensible worm-like chain model (Wang et al., 1997). The persistence length and stretch modulus of DNA neutralized with spermine, 40 nm and 1200pN (Baumann et al., 2000), were used as an estimate of the properties of ds-Mep DNA; the values for polyethylene glycol, 0.35nm and 1560pN (Kienberger et al., 2000), were used as an estimate for the abasic phosphate and the abasic linker inserts. Finally, the ssDNA values of 0.75nm and 800pN (Smith et al., 1996) were used for the ssDNA gaps.

**Predicted using \( t = 2\left(\frac{x^2}{2D}\right)\left(\frac{kT}{Fx}\right)^2 \left\{\exp\left(-\frac{Fx}{kT}\right) - 1 + \frac{Fx}{kT}\right\} \) (Howard, 2001)

Table 4: Summary of high resolution dynamics

<table>
<thead>
<tr>
<th></th>
<th>Mean (s)</th>
<th>STD (s)</th>
<th>SEM (s)</th>
<th>N</th>
<th>Nstop</th>
<th>Nattempt</th>
<th>Npass</th>
<th>Nrecover</th>
<th>Nterm</th>
<th>pstop*</th>
<th>pattempt*</th>
<th>pPass*</th>
<th>pRecover*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Pause</td>
<td>1.00</td>
<td>0.96</td>
<td>0.08</td>
<td>129</td>
<td>101</td>
<td>28</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.78</td>
<td>(0.70,0.85)</td>
<td>0.22</td>
<td>(0.15,0.30)</td>
</tr>
<tr>
<td>Attempt Pause</td>
<td>0.079</td>
<td>0.069</td>
<td>0.013</td>
<td>28</td>
<td>23</td>
<td>5</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.80</td>
<td>(0.64,0.93)</td>
<td>0.20</td>
<td>(0.07,0.36)</td>
</tr>
<tr>
<td>Slips</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>124</td>
<td>120</td>
<td>4</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.96</td>
<td>(0.92,0.99)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Traversal probability calculating using the Laplace estimator (Lewis); in parentheses are 95% confidence intervals calculated using the adjusted Wald method (Agresti and Coull, 1998)

Table 5: Encoding of selected inserts for logistic regression analysis

<table>
<thead>
<tr>
<th>Modification</th>
<th>Extension (nm)</th>
<th>Phosphates 5'-3'</th>
<th>Phosphates 3'-5'</th>
<th>Sugars/Bases 5'-3'</th>
<th>Sugars/Bases 3'-5'</th>
<th>Special Phosphates</th>
<th>Native Structure</th>
<th>Molecules Traversed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 bp ds-Mep</td>
<td>1.6</td>
<td>-5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>10 bp ds-Mep</td>
<td>3.2</td>
<td>-10</td>
<td>-10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>84</td>
<td>103</td>
</tr>
<tr>
<td>11 bp ds-Mep</td>
<td>3.5</td>
<td>-11</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>5'-3' MeP30 hybrid</td>
<td>9.6</td>
<td>-30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>ds-10-PO4-Abasic</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>-10</td>
<td>-10</td>
<td>0</td>
<td>-1</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>ds-linker</td>
<td>0.5</td>
<td>-10</td>
<td>-10</td>
<td>-10</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>32</td>
<td>33</td>
</tr>
</tbody>
</table>
Table 6: Logistic Regression Models of Data collected at 5pN

<table>
<thead>
<tr>
<th>Function</th>
<th>Model</th>
<th>K</th>
<th>-2 log(L)</th>
<th>AICc</th>
<th>Δ</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Log-Log</td>
<td>P53,P35,B35,SP</td>
<td>4</td>
<td>45.99</td>
<td>55.80</td>
<td>0.00</td>
<td>0.21</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>S,P53,P35,SP</td>
<td>4</td>
<td>46.81</td>
<td>56.63</td>
<td>0.82</td>
<td>0.14</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>E,S,P53,P35,SP</td>
<td>5</td>
<td>44.35</td>
<td>57.21</td>
<td>1.40</td>
<td>0.11</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>E,P53,P35,B35,SP</td>
<td>5</td>
<td>44.53</td>
<td>57.39</td>
<td>1.59</td>
<td>0.10</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>S,P53,P35,B35,SP</td>
<td>5</td>
<td>44.68</td>
<td>57.54</td>
<td>1.73</td>
<td>0.09</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>S,P53,P35,B35,SP</td>
<td>5</td>
<td>45.22</td>
<td>58.08</td>
<td>2.28</td>
<td>0.07</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>P53,P35,B35,SP</td>
<td>5</td>
<td>45.67</td>
<td>58.53</td>
<td>2.72</td>
<td>0.05</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>E,S,P53,P35,B35,SP</td>
<td>5</td>
<td>42.74</td>
<td>58.94</td>
<td>3.14</td>
<td>0.04</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>E,S,P53,P35,B53,SP</td>
<td>6</td>
<td>43.41</td>
<td>59.61</td>
<td>3.80</td>
<td>0.03</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>E,P53,P35,B35,SP</td>
<td>6</td>
<td>44.05</td>
<td>60.25</td>
<td>4.45</td>
<td>0.02</td>
</tr>
<tr>
<td>Probit</td>
<td>E,P53,P35,B35,SP</td>
<td>5</td>
<td>47.49</td>
<td>60.35</td>
<td>4.54</td>
<td>0.02</td>
</tr>
<tr>
<td>Probit</td>
<td>E,S,P53,P35,SP</td>
<td>5</td>
<td>47.99</td>
<td>60.85</td>
<td>5.04</td>
<td>0.02</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>S,P53,P35,B53,SP</td>
<td>6</td>
<td>43.68</td>
<td>60.88</td>
<td>5.08</td>
<td>0.02</td>
</tr>
<tr>
<td>Probit</td>
<td>E,S,P53,P35,B35,SP</td>
<td>6</td>
<td>45.46</td>
<td>61.66</td>
<td>5.85</td>
<td>0.01</td>
</tr>
<tr>
<td>C. Log-Log</td>
<td>E,S,P53,P35,B35,SP</td>
<td>7</td>
<td>42.69</td>
<td>62.56</td>
<td>6.78</td>
<td>0.01</td>
</tr>
<tr>
<td>Probit</td>
<td>E,S,P53,P35,B35,SP</td>
<td>6</td>
<td>46.39</td>
<td>62.59</td>
<td>6.79</td>
<td>0.01</td>
</tr>
<tr>
<td>Probit</td>
<td>E,P53,P35,B35,SP</td>
<td>6</td>
<td>46.98</td>
<td>63.18</td>
<td>7.38</td>
<td>0.01</td>
</tr>
<tr>
<td>Probit</td>
<td>P53,P35,B35,SP</td>
<td>4</td>
<td>53.67</td>
<td>63.49</td>
<td>7.68</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Shown are the best 18 models, ranked and sorted by their AICc (Small sample Akaike Information Criteria). Function – Link function for the generalized linear model. K – Number of parameters in the model. L – Likelihood function, Δ – Difference in AICc with respect to the best model. The Akaike Weights (Burnham and Anderson, 2002), w = exp(-Δ/2)/Σ exp(-Δ/2), represent the likelihood of the specific model, given the data set.
Modeling of modification traversal probability with a logistic regression analysis

The data set of traversal probabilities collected at different forces and from inserts lacking variable numbers and types of DNA chemical moieties, in one or both strands, allows us to address the relative importance of contacts with these moieties in a native packaging context, i.e. when all other moieties are present. Logistic regression (LR) analysis is a statistical method aimed at modeling the relationship between a dichotomous response variable and a set of explanatory variables (predictors). With a data set composed of multiple observations (e.g. molecules tested), each corresponding to a certain combination of the predictors (e.g. removal of bases and structure), LR provides a convenient and rigorous way to make inferences about the effect of each of the predictors separately (e.g. the effect of removing the bases without perturbing the structure). The first step in LR is a transformation of the probability (defined in [0,1]) into a variable defined in [-∞, +∞] using one of several sigmoidal functions. We use the term “logistic regression”, somewhat liberally, to denote not only a regression using the logit transformation but also regressions with other link functions appropriate for a binomial response variable, i.e. Probit and Complementary LogLog. The transformed probability is then modeled as a linear combination of the explanatory variables. For example, the classical logit function given by

\[ \text{Logit}(p) = \log\left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \]

where p is the probability that the response variable equals one of its possible values, e.g. “success”, X_1, X_2,... X_k are the explanatory variables and \( \beta_0, \beta_1, \beta_2, ..., \beta_k \) their coefficients. A regression procedure based on a maximum likelihood criterion is used then to find these coefficients. While the LR model does not necessarily represent the true physical model, it provides information on the effect of each of the predictors on the response variable. Notably, if different predictors have the same units, then the values for their coefficients show the relative magnitude of their effect on the response variable.

We performed a logistic regression analysis of our traversal probability data, using as potential predictors for the model: 1) The number of phosphate charges removed in each strand (denoted by the variable P53 for the phosphates in the 5’-3’ strand, and P35 for the phosphates in the 3’-5’ strand), 2) The number of bases and sugars removed in each strand (B53 and B35), 3) The number of special phosphates removed every 10 bp (SP), 4) The presence of native double-strand structure (S equals 0 for native DNA or Mep inserts, -1 for the rest of the inserts), and 5) The end-to-end extension of the insert under
the applied force (E, a continuous variable calculated using the contour and persistence lengths as in Table 3). For example, for 5 bp ds-MeP, where 5 phosphates are “removed” from both strands, the predictors are P53 = -5 bp, P35 = -5 bp, B53 = 0, B35 = 0, SP = 0 and E = 1.58 nm. Table 5 shows additional examples for this encoding method. The unpaired bulge modifications do not “remove” elements from the DNA and therefore were not included in this analysis. The analysis was performed using the statistical package XLSTAT 2009 (Adinsoft) and the statistical toolbox of Matlab (Mathworks).

Our a priori set of explanatory variables includes all the chemical moieties of DNA and additional parameters that vary in the different experiments. Does each one of these parameters, when controlling for the others, really affect translocation? The answer to this question can be found by a methodical process of model selection. A variety of models can be constructed using different subsets of the explanatory variables defined above, and each is characterized by a different quality of fit to the data (as represented by the likelihood function, $L$). However, despite the improvement in $L$ achieved by adding additional parameters, it is important to avoid over-fitting the data, which will lead to wrong conclusions. Information theory provides a convenient and rigorous way to rank the different models that can be constructed, in order to select the best model that includes only those variables that have an effect on the calculated probability. The (small-sample) Akaike's information criterion, $\text{AIC}_c = -2 \log(L) + 2K \left(1 + \frac{n}{K-1}\right)$ where $K$ is the number of parameters and $n$ is the number of samples, provides a balance between bias and parsimony. In order to be included in the model, a certain variable needs to improve the likelihood of the model in a way that outweighs the additional complexity of a model with a larger number of variables.

We first performed a regression using all of our data collected against a 5 pN tension. With 7 variables and 3 link functions (Logit, Complementary LogLog given by $\log(-\log(1-p))$, and probit, equal to the inverse of the normal cumulative distribution function), it is possible to construct 381 different models, ranging from those that include only a single explanatory variable to the complete set. As seen in Table 6, the model with the lowest $\text{AIC}_c$ (i.e. the best model) is a Complementary Log-Log model that includes the phosphates in both strands (P53, P35), the bases in the 3’-5’ strand (B35), and contacts with phosphates separated 10 bp on the 5’-3’ strand (SP). The results for the predictors’ coefficients are summarized in Figure 2-5a. Figure 2-5b shows the predicted traversal probabilities calculated for all our measurements, alongside the measured probabilities.
As an additional control for the validity of the parameter SP, we defined additional variables representing the removal of phosphates with periodicities ranging from 6 to 18 bp, and for each of these variables we performed a regression in which the control variable replaced SP. Figure 2-5c shows the calculated AICc’s from these regressions. The dashed line in the figure indicates the AICc for a model that includes only P53, P35 and B35. Only SP produces a significant improvement in the quality of the model.

The results of the regression performed on the 5pN data set confirm our previous analysis of the neutral ds-MeP experiments, indicating the importance of phosphate contacts every 10 bp on the 5’-3’ strand. They also confirm the existence of contacts with the rest of the phosphates on both strands. Interestingly, the regression also reveals non-charge contacts in the form of contacts with bases or sugars, a subtle feature that may not be obvious from direct inspection of the traversal probabilities. Moreover, since all the perturbations are parameterized with the same units, i.e. the number of bp over which the change is made, the regression provides a ranking of the relative importance of the different contacts at 5 pN. Clearly, contacts every 10 bp on the 5’-3’ strand are more important than other contacts on the same strand, and those in turn are more important than the phosphates, bases and sugars on the 3’-5’ strand. Using the above results, we are now able to predict the traversal probability for a particular perturbation (e.g. removing a phosphate in the 3’-5’ strand), while all the other contacts are not perturbed. We define the “importance” of each contact as the inverse of the length (bp) that need to be perturbed in order to reduce the probability to 50%. These results are summarized in Figure 4 as a “heat map” on DNA.

Table 6 also shows the AICc differences, $\Delta_i = \text{AICc}_i - \text{AICc}_{\text{best}}$, over all the candidate models. Using these differences we calculate the Akaike Weights (Burnham and Anderson, 2002) given by

$$w_i = \frac{\exp\left[-\frac{1}{2} \Delta_i\right]}{\sum_1^R \exp\left[-\frac{1}{2} \Delta_i\right]},$$

where R is the number of models being considered. These weights (shown in Table 6) quantify the plausibility of this model being the actual best model in the set of candidate models, i.e. the likelihood of the model, given the data. Since the weight of evidence for our selected model being the best model is not highly conclusive (e.g. the weight for the best model is only 1.5 times larger than the weights of the second-best), it is reasonable to explore the predictions of additional models in our set using multi-model inference techniques (Burnham and Anderson, 2002). The 18 models with the highest $w$’s comprise...
95% of the model likelihood (i.e. \( \sum_{i=1}^{18} w_i > 0.95 \)), and we use this subset for our multi-model inference. We use the information from the different models in two different ways. First, we compute the model-averaged coefficients by averaging the predicted values for the coefficients from the different models, weighting these values with the corresponding \( w_i \). While the set of average coefficients includes also an effect for the insert extension and structure (data not shown), the values for the coefficients corresponding to the phosphates and bases, including the special phosphates, are in good agreement with our prediction based on one selected model only (Figure 2-5a). A second way of accounting for the information content of the different models is to sum, for each parameter, the \( w \) values of all the models in the set that include this particular parameter. The result can be viewed as the likelihood of a parameter (Burnham and Anderson, 2002), given the data. These results also support our previous results: All the models in the top 95% likelihood contain the phosphates in both strands and the phosphates separated by 10 bp, while most of them contain also the bases and sugars on the 3'-5' strand. The likelihood of P53, P35, B35 and SP is 0.95, 0.95, 0.55, and 0.96 respectively.

The regression results indicate that the motor makes contact with different chemical moieties of DNA, and that the traversal probability can be explained in terms of the number of these moieties that are removed. Surprisingly, they also reveal that removing all of these moieties—phosphates, bases, and sugars—will result in a reduced but finite traversal probability, indicating the existence of additional contacts, not included in our models. Notably, this is true not only for the best model, but for all of the models in our candidate set.

Next, we turned our attention to the effect of force on the traversal probability. We concentrate on the effect of force on the phosphate contacts, which the previous regression revealed as the most important specific contacts, and performed a regression on the data collected for ds-MeP and ss-MeP, against different forces. Since different contacts may have a different force dependence, we introduced the force dependence by using as predictors the products of each one of the chemical predictors with the force (e.g. P53*F). This means that for each of these predictors there will now be two terms in the linear function, and the regression will find two coefficients: a “force-independent” coefficient (e.g. the coefficient of P53) and a “force-dependent” one, (the coefficient of P53*F). The results of the model are shown in Figure 2-5d. Remarkably, the entire effect of the special phosphates is a force-dependent one (i.e the coefficient for SP * F is significant, but the one for SP is not),
predicting that without an external force there will be no effect of removing these phosphates. This surprising result is an additional indication of the fact that these contacts belong to the dwell phase, where the only load on the motor is the externally applied force. The rest of the phosphate contacts have both a force-dependent and a force-independent weight, supporting the fact that these phosphates are contacted during the burst phase, when the motor generates force. Comparing the force-dependent weights of the various phosphates shows that the contacts during the dwell exhibit much higher force sensitivity than the contacts during the burst, an indication of the higher “strength” of the special contacts (as the force increases, removing this contact will result in a much larger decrease in probability. This means that without SP the motor cannot sustain a high force, indicating that these are strong contacts). Note that force-independent and force-dependent weights should not be compared directly since the absolute value of the latter depends on the choice of force units.
References


LEWIS, J. 2006. When 100% Really Isn’t 100%: Improving the Accuracy of Small-Sample Estimates of Completion Rates. *Journal of Usability Studies*, 1, 136-150.


Appendix II

Protocols
SLIC Protocol

Sequence and Ligase Independent Cloning (SLIC) (Li and Elledge, 2007, Li and Elledge, 2012) allows for powerful and easy DNA manipulation. Thanks go to Will Draper for promoting it in our lab. Gibson Assembly uses similar principles and has become quite popular, although we have not found it to be obviously better than SLIC.

The general idea is to generate pieces of DNA with homologous single-stranded DNA overhangs (“sticky ends”), combine them, and then transform them into *E. coli*, to repair the gaps/nicks and amplify the DNA. The combined DNA must assemble into a plasmid vector with an origin of replication and antibiotic selection marker. The sticky ends are generated by using the 3’ exonuclease activity of T4 polymerase in the absence of dNTPs. This “chew-back” reaction is quenched by adding a dNTP.

The SLIC reaction is extremely robust, and we have had routine success with combining up to five fragments. Success with up to ten fragments (with reduced efficiency) has been reported in the literature. I used homologous sequences of 30 basepairs, although it is possible to use overlaps of as low as 20 basepairs with reduced efficiency.

The most critical and error-prone step is designing the sequences of the primers, so it pays to double and triple check them. In the time since I did my assemblies New England BioLabs started selling a Gibson Assembly kit, which seems a bit superfluous, but they do have on their website a useful tool to design primers: http://nebuilder.neb.com.

To design primers manually I found forward and reverse 20 nucleotide oligos for each fragment (the free/donation-ware program ApE http://biologylabs.utah.edu/jorgensen/wayned/ape/ is a useful tool for this task). Then I fused the reverse complement of each to the 5’ end of the adjoining fragment. Our oligo supplier had a 35 nucleotide cutoff for oligos to be ready the next day, so I then removed the first 5 nucleotides. This can be confusing, and sketching it can be helpful.

The general schematic for a single junction looks something like this:
Here is an example with over-simplified sequences for clarity. Say we want to join the following pieces:

End of fragment 1: 
AAAAAAAAAAAAAAAAATAA
TTTTTTTTTTTTTTTTATT

Beginning of fragment 2: 
ATGGGGGGGGGGGGGGGGG
TACCCCCCCCCCCCCCCCC

The fragment 1 reverse primer is TTATTTTTTTTTTTTTTTTT, and the fragment 2 forward primer is ATGGGGGGGGGGGGGGGGG. The full combined oligos will then be aaaaaaaaaaaaataaATGGGGGGGGGGGGGGGGG and ccccccoccccocccccccatTTATTTTTTTTTTTTTTTTT. Trimming 5 nucleotides of the 5’ end yields aaaaaaaaaaaaataaATGGGGGGGGGGGGGGGGG and ccccccoccccocccccccatTTATTTTTTTTTTTTTTTTT. These 35 nucleotide oligos would result in a 30 basepair overlap between the two fragments.

Note that these oligos are for separate PCR reactions, and each of their primer pairs would be determined by carrying out the same steps on the other side of each fragment.

These example sequences of course would not actually work due to mispriming of the long runs of the same nucleotide. This is the one slight weakness of this technique: you need to find usable PCR primer sites, so it’s not 100% sequence independent. However, modern polymerase enzymes (such as PfuUltra II) are quite robust. If you have flexibility in choosing your junction sites, it’s best to follow PCR primer best practices: 40-60% GC content, Tm = 50-60 °C, length = 15-25 nt, avoid secondary structure, avoid over 4 di-nucleotide repeats, and avoid contiguous repeated nucleotide runs over 4 nucleotides. These are not strict guidelines, but keep an eye out for trouble with PCR products if you violate them.

The plasmid backbone fragment can be generated by restriction digest, but I always generated it with PCR. Generating it with PCR is the most convenient way, but there is the slight possibility of developing mutations, which might be overlooked since the full backbone is not always sequenced.

Short fragments less than 100 basepairs do not work well in the chew-back reaction. However, it is possible to order a pair of long primers that leave the desired sticky ends when annealed, and add them after the chew-back reaction.

The completed PCR reactions should always be run on a gel, since it is important to have a distinct PCR product. If the band is smeared, the SLIC reaction probably will not work. The few times that I did not get a distinct PCR band I increased annealing temperature or switched PCR enzymes (from
PfuUltra II to Phusion) and had success. Alternatively you can order different primers.

DpnI digestion is used to get rid of any contaminating PCR template DNA, as it cleaves methylated DNA. Assuming your DNA comes from a DAM+ strain, it will be methylated. PCR product DNA is not methylated, and will not be affected by DpnI. To save time and effort, the DpnI digestion can be combined with the T4 chew-back reaction.

It is also worth ordering primers to check the SLIC product via PCR and sequencing. Checking PCRs should span the fragment junctions.

Protocol

1) Perform PCR as described in high-fidelity polymerase enzyme (such as PfuUltra II) manual. Of a 50 µL PCR, run 2 µL on a gel to confirm clean bands.

2) Perform a Qiagen PCR cleanup with MinElute columns and elute into 10 µL.

3) Quantify DNA concentration with Nanodrop using 1.5 µL solution.

4) DpnI digest and T4 polymerase chew-back: 5 µL reaction containing 150 ng of backbone and roughly equimolar quantities of DNA fragments (to avoid unnecessary dilution, up to 4:1 insert to backbone ratio is fine). Add 0.5 µL 10x NEB2 and BSA, plus 0.5 µL Fermentas DpnI and 0.2 µL NEB T4 polymerase. Mix and incubate at room temperature for 30 min.

5) Quench by adding 0.5 µL 10 mM dATP (or could use any other dNTP).

6) Anneal by incubating at 37 °C for 30 min.

7) Transform ~3 µL of reaction into Top10 chemically competent cells.

Next Day

Colony PCR is a useful technique to screen colonies. Make a PCR master mix with all components except for template and split it into PCR tubes (strip tubes work well for this). To pick cells from colonies, first number a set of clean
small Eppendorf tubes (or PCR strip tubes) and add 50 µL dH₂O to each. Number or otherwise label the tubes. Then touch a pipette tip to a colony (don’t worry about scraping up the entire colony), and then swish around/pipet up-down in the 50 µL of dH₂O. Repeat with other colonies, and then add 0.5 µL of each to PCR.

If the PCR looks good, then you can use the appropriate tube of cells in water to inoculate a 3 ml LB culture for miniprep of plasmid DNA and check via sequencing.

References

**E. coli chemical competence protocol**

Will Draper adapted this from a Wendel Lim lab protocol.

This is a robust chemical competence protocol which should give good results even in inexperienced hands. The original protocol was for making freezer stocks of chemically competent cells, a la Invitrogen's TOP10, but I use this procedure for routine subcloning, and in fact have never frozen these cells. I find it easy enough just to grow them up the day of. Using the TOP10 strain, I routinely get transformation efficiencies on the order of 0.05 x 10^9 cfu/µg DNA, which is roughly 20 fold less efficient than Invitrogen TOP10, and more than good enough for the things I do. With optimization, I am sure that number could be pushed much closer to TOP10's, but I simply have no need to spend the time doing that.

**Step 1: Prepare buffers**

First, you need to prepare the two competence buffers, TSS and KCM. TSS requires PEG 3350, which is fairly slow to dissolve, so give it some time, and use shaking. I make a 50% solution, dissolving it on a shaker in the warm room.

1x TSS (volumes are for 10 mL):
- 10% PEG 3350 (2 mL of a 50% solution)
- 5% DMSO (0.5 mL)
- 20 mM MgCl2 (0.2 mL 1 M MgCl2)
- LB-broth to 10 mL

1x KCM (volumes are for 10 mL):
- 0.1 M KCl (0.5 mL of 2 M KCl)
- 0.03 M CaCl2 (0.3 mL of 1 M CaCl2)
- 0.05 M MgCl2 (0.5 mL of 1 M MgCl2)
- H2O to 10 mL

**Step 2: Prepare cells**

This protocol works on cells in either exponential or stationary phase, though efficiencies are dramatically reduced the further a culture is in to stationary phase. Efficiencies from an overnight culture, freshly in stationary phase, are good enough for routine subcloning, and is a big time-saver. To prepare cells in exponential phase, start a day culture using a 1:100 dilution of fresh overnight culture in LB, allowing ~0.5–1 mL of day culture per
transformation. Grow to an OD600 of roughly 0.5–0.8, which should take roughly 3 hours at 37 C.

**Step 3: Harvest**

Chill the cells briefly on ice, and then harvest by centrifugation. If you only have a couple transformations to do, it is easiest to use the microcentrifuge (16.1 kRCF for 0.5 minutes), but if you're doing more than four or so, the refrigerated desktop centrifuge is easier (5 kRCF for 5 minutes). Discard the supernatant, and resuspend cells in 50 µL ice-cold TSS per transformation (i.e. concentrate the liquid culture roughly 10–20 fold). From now on, keep the cells chilled on ice until the heat shock. If you're making freezer stocks, aliquot and freeze the competent cells at this point. When you're ready to perform the transformation, thaw cells on ice, and proceed to the next step.

**Step 4: Transform and Recover**

Add your plasmid DNA to 50 µL of competent cell suspension. Mix the DNA with a gentle swirling motion of the pipette tip, do not mix by pipetting up and down. Then, add an equal volume of ice cold 1x KCM, and, again, mix with a gentle swirling motion of the pipette tip, do not mix by pipetting up and down. Allow the cells to rest for between 1 and 30 minutes (longer is supposed to increase efficiency, but I've never tested that), and then heat shock at 42 C for 90 seconds. If you are using a heat block for the heat shock, add some deionized water to the wells to ensure good thermal contact with your tubes. Place the cells back on ice immediately after the heat shock. Add ~500 mL LB to the cell suspension, and shake at 37 C.

Depending on the plasmid's antibiotic resistance, shake for about an hour before plating. You can go a little shorter for ampicillin resistance, and will need a little longer for chloramphenicol resistance. Plate between 100 and 200 µL on the appropriate antibiotic plate, and throw in a warm place overnight (or on your bench over the weekend). Always save your transformation tubes so that you can re-plate the suspension if your plates don't yield any transformants.

**Optional**

It's often useful to perform a positive control for transformation, particularly when you're using competent cells for cloning and vector construction. A positive control will eliminate ambiguities as to where such a protocol failed. Was the ligase bad? Was there even any DNA in there? Were the competent cells bad? (No! I did the positive control!)
To do this, grow enough cells for one extra transformation. Add 1 µL of Invitrogen pUC19 plasmid DNA (it should be at 10 pg/µL) as the DNA, and heat shock as normal. Plate 100 µL of the cells, and the next day count the number of colonies on the plate. Multiply that number by $6 \times 10^5$ to calculate the efficiency in cfu/µg DNA. An average number of colonies is something on the order of 100.
**P1 Transduction Protocol**

Adapted for the Liphardt Lab by Derek Greenfield.

This is a powerful technique for moving gene/operon knockouts and mutations between strains of *E. coli*. The phage will package ~90 kb pieces of the genome and inject into recipient cells where the DNA can homologously recombine with the new genome. The one caveat is that the phage can carry neighboring genes along with your gene of interest. This is usually not a problem, but it is something to look out for. For example, I transduced a chemotaxis knockout (ΔcheA-Z) from a strain with flagella (*fliC*) to a strain without *fliC*, and approximately half of the resulting colonies had *fliC*. This makes sense, since *cheA* is approximately 27 kb away from *fliC*. In this case screening was easy because of the clear phenotype of being non-motile.

Once a phage infects a cell in the presence of Ca+2, the cell will die. But, if phage infects a cell and then the Ca+2 is removed/chelated, the phage will not produce any additional infectious phage. Therefore, transductants will only be isolated when a single phage (carrying *E. coli* DNA) infects a cell—if multiple phages infect the same cell, the cell dies. That is the importance of doing multiple dilutions of phage, there is a sweet spot—too few phage per cell and not enough transductants; too many, and all cells perish.

**P1 broth = LB + 5 mM CaCl₂ + 0.2% glucose (11.1 mM)**

Glucose should be filter sterilized, not autoclaved; store at room temperature. 1 M sodium citrate stock should be pH’d to 5.5-6 using HCl, filter sterilized; store at room temperature.

**Lysate preparation**

Use donor strain with selectable marker; P1 will transfer 90kb pieces of the genome.

1. Grow donor strain overnight in 1 mL L.B.
2. Dilute overnight culture 1:100 in 2 mL fresh P1 broth. Grow with aeration at 37 °C for 60-90 minutes, until slightly turbid.
3. Add 10⁸-10⁹ (100 µL) P1 phage stock lysate to the culture, continue growing at 37 °C. Monitor for 1-3 hr until the culture has lysed completely. If culture does not clear, let tube sit for a 2-4 hours and check for cell debris.
(cloudy wide pellet at bottom—more obvious if compared to a control). If there is debris, there’s probably enough phage for a successful transduction.

4. Transfer 1–2mL to eppendorf. Important: chloroform will react with certain plastics, including many culture tubes, so only add chloroform to eppendorfs. Add several drops of chloroform to the lysate and vortex. Wait a few minutes. Centrifuge away the debris and chloroform (16,100 rpm, 2 min) and transfer the supernatant to a fresh tube. Store at 4 °C for up to 1 year.

**Transduction**

Use recipient strain that lacks selectable marker.

1. Grow recipient strain overnight in 2 mL LB
2. On the next day, harvest the cells by centrifugation (4500 rpm, 5 min) and resuspend in half the volume (1 mL) fresh P1 broth.
3. Set up three reactions:
   a) 100 µL P1 donor lysate + 100 µL recipient cells
   b) 100 µL 1:20 diluted P1 donor lysate (diluted in P1 broth) + 100 µL recipient cells – this is important, don’t skip it! Sometimes the diluted phage will be the only plate that has colonies.
   c) 100 µL P1 broth + 100 µL recipient cells
4. Mix tubes well and incubate at 37 °C for 30 min without shaking. This allows phage to infect, but not lyse, recipient cells.
5. Add 200uL of 1 M sodium citrate (pH 5.5-6) then 1 mL LB to each tube, and incubate at 37 °C with shaking for at least 1 hr to allow recombination and expression of the antibiotic resistance marker.
6. Plate 200 µL of each reaction on an appropriate antibiotic-containing plate. Also plate 50uL of your P1 donor lysate as a control to check for contaminating donor cells remaining in your donor lysate.
7. You should get 10 to 2000 colonies on plates A and B, and no colonies on the two control plates. Your colonies are on a plate that is covered with P1 phage. To isolate cells from remaining phage, colony purify cells:
   a) Pick single colonies with a pipette tip and transfer to 500 µL LB + 100 µL sodium citrate. Vortex for 5-10 sec.
   b) Streak onto a normal LB + antibiotic plate to obtain single colonies.
   c) Pick a single colony, grow in normal LB + antibiotic to make stocks and restreak onto LB + antibiotic plate for use.
8. Test your purified colony for the presence of the mutant gene you intended to transduce using colony PCR or by looking for a phenotype.
Troubleshooting

If you see no cells on your transduction plates, try plating all of your reaction by spinning 16100g x 2min and removing all but 50-200 µL of liquid. If no colonies are obtained, try the transduction again but with different dilutions of phage.

If your donor cells don’t clear, check the concentration of viable phage:
1. Grow 2mL overnight strain that can be infected by your phage in LB
2. Resuspend overnight in 1mL P1 broth.
3. Dilute phage to $10^{-6}$ and $10^{-8}$ in P1 broth in 100 µL volumes.
4. Mix 100 µL cells + 100 µL each phage dilution and incubate at 37 °C for 30 minutes without shaking to allow phage to adsorb and infect cells.
5. While phage is incubating, make 0.7% top agar by weighing 0.28 g of low-melt agar in 50 mL falcon tube, then add 40 mL LB. Microwave for 10 seconds at a time, vortexing in between until agar is melted. Alternatively, you can autoclave regular agar and LB powder.
6. Transfer agar to 45 C water bath and allow to cool. When cool, add CaCl$_2$ to 5mM (2mL of 100 mM) and MgSO$_4$ to 100 mM (0.4 mL of 1 M, skip this?).
7. When phage and cells are ready: Aliquot 3mL of warm top agar in culture tubes (or other large tubes) and quickly add cells plus phage, vortex well, and pour onto LB plates before top agar solidifies.
8. Allow top agar to solidify, then flip plates and incubate at 37 C overnight.
9. Count clear plaques and determine phage concentration (pfu).
Proteorhodopsin and Tethered Bead Protocol

This is a tricky experiment for a number of reasons. It is low throughput, as only one motor is recorded at a time, but in order to make meaningful observations, many motors need to be documented. Each motor must be recorded for a minimum of a few minutes, but the greater limiting factor is finding usable motors.

The proteorhodopsin is surprisingly robust, as I never had it fail (the one time light did not restore PMF in proteorhodopsin cells I had forgotten to add the retinal). On the other hand, the efficiency of obtaining “good” spinning beads from day to day a bit fickle. A “good” bead spins quickly (~60 Hz for fully energized cells spinning 1 µm beads), smoothly, about a small diameter (significantly less than the diameter of the bead itself), and in a round or ellipsoid path. Looking at it by eye through the eyepieces of the microscope the bead looks slightly blurred. If you can discern individual rotations of a bead on a fully energized cell, then it is spinning too slowly, either because the flagellar stub is too long, or because it is partially stuck. Quickly scanning by eye is useful to find good bead candidates.

I also found it useful to track trajectory and rotation speed in real time. To help determine if a bead was worth recording. In addition to finding beads that are spinning sufficiently fast, looking at the trajectory on a computer screen helps avoid wasting time on beads with non-canonical tethering geometries. Before becoming more selective with bead selection I found that I had up to ~25% of beads rotate the “wrong” direction. That is, in the absence of CheY, motors spin exclusively CCW at room temperature, but I would sometimes find beads spinning CW.

How could a bead spin the wrong way? The canonical bead is sitting above the cell with the flagellar stub below it (assuming the objective is pointed down and the cell is sitting on the bottom surface of the chamber). However, other geometries are possible. For example, a bead may be on the side of the cell and rotate in a vertical plane with respect to the objective. If that plane tilts out slightly the rotation direction will appear reversed. Ultimately, if the flagella stub is connected to the top hemisphere of the bead, rotation will appear reversed.

At Harvard the bead trajectory is tracked using two slit-covered photomultiplier tubes, and the real-time trajectory is displayed with LabView. At Berkeley we tracked the beads with a fast EMCCD camera using a small (10x10-16x16) ROI, and we used custom acquisition software to display the bead trajectory track and frequency in real time. By using these tools to identify good beads, I was able to reduce the number of wrong way spinners to under
~5%. This of course does not matter for non-switching motors, but it is relevant for when you are observing a switching bead and there is no other reference for which direction is which.

Finding good beads is also dependent upon the cells being properly stuck to the surface. If there is not enough poly-L-lysine, then the cells will not stick well. They can detach while recording, and it also seems that they are harder to find in the first place. If the density of poly-L-lysine, is too high, the cells remain well-stuck, but it becomes harder to find good beads. So there is a sweet-spot of poly-L-lysine density. Unfortunately the effectiveness of poly-L-lysine varies from batch-to-batch, and within the same bottle it can decrease on a month timescale. I always used the same concentration of poly-L-lysine, but I varied the incubation time to adjust the amount on the slide. Unfortunately I did not have a straightforward quantitative assay of poly-L-lysine effectiveness, so it was a bit of trial and error.

Mid-way through my experiments I heard that the Shaevitz Lab uses polyethylenimine instead of poly-L-lysine to stick cells to slides. I never used it, but it might be worth trying, because of the problems with poly-L-lysine.

I did try sticking cells to the cover slip via centrifugation, but it was not consistent enough to be useful. Cells were well-stuck, but I did not have a way of ensuring predictable coverage density, so I did not end up using the centrifuge for cell adhesion.

**Protocol**

1) Start overnight culture of cell strain containing proteorhodopsin, lacking wildtype flagella, and containing sticky flagella in T-broth (1% tryptone, 0.5% NaCl) at 33 °C shaking.

2) Grow 1:100 dilution in 2 mL T-broth from overnight 3 hr at 33 °C shaking.

3) Induce proteorhodopsin and CheY*, and add retinal:
   - 0.1 % arabinose (1:200 from stock)
   - 20-200uM IPTG
   - 10uM retinal (1:1000 from ethanol stock)
   - Return to 33 °C shaking for 1 hr

4) Measure OD_{600} (typically 0.3-0.4)
5) Centrifuge 2 min at 2,500 RCF, remove supernatant, and resuspend in 1 mL chemotaxis buffer + lactate (10 mM potassium phosphate, 0.1 mM EDTA, 10 mM lactic acid, 10mM Lactate, pH 7.0).

6) Pipette 0.35 mL poly-L-lysine onto a 24x60 mm coverslip. Incubate at room temperature for 2-10 minutes, and then rinse well with dH$_2$O. You can proceed with the following steps during the incubation. After rinsing, shake water drops off, and gently blot any remaining with a Kimwipe.

7) Remove half (0.5 mL) of cell suspension and put remainder at 4 °C for later the same day if you want to make a second chamber. Shear flagella by passing solution through two 25-gauge needles 100 times. (Tubing is epoxied to two needles so that the solution can be passed back and forth between two 3 mL syringes.)

8) Centrifuge 2 min at 2,500 RCF, remove supernatant, and resuspend in enough chemotaxis buffer + lactate to put the cells at OD$_{600}$ = ~2-4 (refer to OD$_{600}$ from step 4)

9) Form a chamber by sticking two short pieces of double-stick tape (~2x16 mm) to the sides of a 22x22 mm coverslip, and then placing on the middle of the poly-L-lysine-coated coverslip.

10) Flow in ~38 µL of sheared and concentrated cell solution by pipetting onto one edge of chamber and incubate 2 min at room temperature.

11) Flow in 50 µL of beads (1:20 dilution of Bangs 1um) by pipetting on one edge and wicking the other edge with the corner of a Kimwipe and incubate 2 min at room temperature.

12) Wash with 225 µL chemotaxis buffer + lactate by flowing through the chamber.

13) Seal all four sides of chamber with 5-minute epoxy

14) Wait a few minutes for epoxy to set, and then put on microscope. PMF will drop in about 20 minutes, depending on cell density and temperature.

15) Find a good bead and execute your desired excitation light sequence.