

Case studies in quantitative biology: Biochemistry on a leash and a single-molecule Hershey-Chase experiment

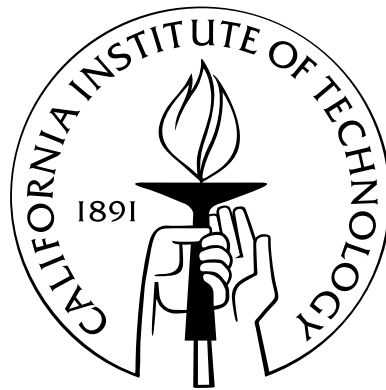
Thesis by

David Van Valen

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2011

(Defended April 25, 2011)

© 2011

David Van Valen

All Rights Reserved

Acknowledgements

Science is not a solitary endeavor. As such, there exist a number of people who have helped me along the path to a PhD, and they all deserve my thanks. I would like to thank my advisor, Rob Phillips and the members of the Phillips lab for graciously hosting me in their laboratory these past six years. I am particularly indebted to David Wu, who spent countless hours as a scientific mentor during my PhD. I also thank Paul Wiggins and Heun Jin Lee for the mentorship they provided during my stay at Caltech.

The research projects I worked on could not have been completed without the help of others. Mikko Haataja was instrumental in completing the theoretical analysis of signaling proteins. It was a pleasure working with Martin Linden, Jane Kondev, and Dan Reeves as well trying to extend these ideas to a number of different biological systems. The *in vitro* DNA ejection project could not have succeeded without the help of David Wu and Qicong Hu. Lastly my efforts in constructing a single-molecule Hershey-Chase experiment would not have met any success without the help of David Wu, Paul Wiggins, and Hannah Tuson. Several generations of students in the physiology course at the Marine Biological Laboratory in Woods Hole have worked on this project too, and I thank them as well. I would also like to thank the instructors at the Woods Hole physiology course, in particular Ron Vale and Dyche Mullins. It was, and still is, inspiring to see science conducted with such passion. I hope their passion will stay with me through the years.

I am deeply indebted to the Hertz Foundation for funding the majority of my stay at Caltech. Without their support, I would not have had the intellectual freedom necessary to develop and grow as a scientist. In particular, I thank Ray Sidney, whose generous support provided me this opportunity. I would also like to thank the community of Hertz fellows for their friendship and support over the years.

I would also like to thank my friends, particularly Adrian Elicegui, Daniella Escobar, and Kahlil Moreland. I thank you for your support and for providing an escape from academia.

Lastly, I would like to thank my family for enduring my absence during these past six years and for providing me with the love and support necessary as I've walked down this path. To my mother, father, and brother — from the bottom of my heart, I thank you for all of your love and support. This would not have been possible without you.

Abstract

The last 50 years of biological research has seen a marked increase in the amount of quantitative data that describes living systems. This wealth of data provides a unique opportunity to recast the pictorial level descriptions of biological processes in the language of mathematics, with the hope that such an undertaking will lead to deeper insights into the behavior of living systems. To achieve this end, we have undertaken three case studies in physical biology. In the first case study, we used statistical mechanics and polymer physics to construct a simple model that describes how flexible chains of amino acids, referred to as tethers, influence the information processing properties of signaling proteins. In the second case study, we studied the DNA ejection process of phage lambda *in vitro*. In particular, we used bulk and single-molecule methods to study the control parameters that govern the force and kinematics of the ejection process *in vitro*. In the last case study, we studied the DNA ejection process of phage lambda *in vivo*. We developed an assay that allows real-time monitoring of DNA ejection *in vivo* at the single-molecule level. We also developed a parallel system that allows the simultaneous visualization of both phage capsids and phage DNA at the single-cell level, constituting a true single-molecule Hershey-Chase experiment. The work described in this thesis outlines new tools, both in theory and experiment, that can be used to study biological systems as well as a paradigm that can be employed to mathematicize the cartoons of biology.

Contents

Acknowledgements	iii
Abstract	iv
1 Introduction	1
2 Biochemistry on a leash: Information processing in biological systems	4
2.1 Introduction	4
2.2 Statistical mechanical model of tethered receptor-ligand pairs	9
2.3 Polymer physics of protein tethers	13
2.4 Signaling proteins: Roles of tether geometry and length	15
2.4.1 Simple switches	16
2.4.2 Complex switches	19
2.4.2.1 Class I switches	20
2.4.2.2 Class II switches	23
2.4.2.3 Applications to reprogrammed N-WASP	23
2.4.2.4 SH3/cis-ligand interactions	26
2.5 Discussion	27
2.6 Appendix: A kinetic model for the simple switch	28
2.7 Appendix: Polymer models	36
2.7.1 The wormlike chain	36
2.7.2 The freely-jointed chain	40
2.7.3 Computing probability distributions with delta functions	43
2.8 Acknowledgements	45
3 Ion-dependent dynamics of <i>in vitro</i> DNA ejection for bacteriophage lambda	46
3.1 Introduction	46

3.2	Experimental design	48
3.3	Materials and methods	49
3.4	Results	52
3.5	Conclusion	58
3.6	Appendix	60
3.7	Acknowledgements	61
4	A single-molecule Hershey-Chase experiment	67
4.1	Introduction	67
4.2	Experimental design	68
4.3	Materials and methods	70
4.4	Results	73
4.5	Conclusion	77
4.6	Appendix	79
4.6.1	Dye-bound phage ejection controls	79
4.6.1.1	SYTOX Orange staining of phage is equivalent to DAPI	79
4.6.1.2	SYTOX-Orange-stained phage eject their DNA <i>in vitro</i>	80
4.6.1.3	SYTOX-Orange-stained phage behave like wild-type phage in terms of bulk infection	81
4.6.1.4	Cells in the absence of phage do not increase in fluorescence	82
4.6.1.5	Photobleaching and ejection time scales are separated	82
4.6.1.6	Single-phage ejection trajectories	84
4.6.2	Strains, media, and controls for constructing λ DVV1	89
4.7	Acknowledgments	90
	Bibliography	94

Chapter 1

Introduction

I begin with the following quote about Stokes' theorem from Michael Spivak's *Calculus on Manifolds* [1]:

"The reader probably suspects that the modern Stokes' Theorem is . . . difficult . . . On the contrary . . . the proof of this theorem is, in the mathematician's sense, an utter triviality . . . [However,] the statement of this triviality cannot be understood without a horde of difficult definitions . . . There are good reasons why the theorems should all be easy and the definitions hard."

While Spivak speaks specifically about Stokes' theorem, the essence of this quote is true in a more general sense. If one examines a system with enough tools, and couches the results in the appropriate language, then the underlying principles that govern it will become transparent. The lessons gleaned from this quote are particularly important for modern biology. The efforts of the scientific community over the last 50 years have brought about an increase in our knowledge about the information content of genomes, as well as careful biophysical characterizations of a number of important biological systems [2]. Suffice to say, there has never been more quantitative information available about living systems than there is now. The question that remains is — what should be done about it?

It is my belief that the quantitative nature of the information collected about living systems necessitates a shift in the language used to describe that information. The predominant method of capturing a biological "story" is with a picture; a brief reading of the classic textbook "*Essential Cell Biology*" is all that is necessary to convince the reader that this is true [3]. While much can be learned from these pictures, experience in other fields shows there is much to be gained by couching the science in the language of mathematics. Indeed, language is particularly important as it shapes both how we think, how we interpret, and even the very nature of the questions we can ask. This, in essence, is the theme of this thesis — mathematizing the cartoons of biology. Can we take these pictorial-level descriptions and recast them in the language of mathematics? Furthermore, what insights can we gain from such an undertaking?

Quantifying biology is not easy. However, the task becomes feasible if one initially limits the systems

under study. The ideal experimental system would have two features. The first feature is the presence of knobs in the experiment that can be tuned and are directly connected to experimental observables. The second feature is the presence of a transparent mathematical theory that predicts what happens to the system when those knobs are tuned. The presence of both features allows for a rich interplay between theory and experiment that deepens our understanding and enables us to fully harness the insights buried within the data. To this end, I undertook three separate case studies; each is described in full in the following chapters. These projects encompass all the methodologies required for this grander scheme — constructing a quantitative model from a cartoon-level description, using a previously existing experimental system with tunable knobs to test theoretical predictions, and building a new experimental system to take novel, quantitative measurements. With the broader picture in mind, I will briefly outline the three case studies and their findings below.

In the first chapter, we use statistical mechanics and simple ideas from polymer physics to develop a quantitative model of proteins whose activity is controlled by flexibly tethered ligands and receptors. We predict how the properties of tethers influence the function of these proteins and demonstrate how their tether length dependence can be exploited to construct proteins whose integration of multiple signals can be tuned. One case study to which we apply these ideas is that of the Wiskott-Aldrich Syndrome proteins as activators of actin polymerization. More generally, tethered ligands competing with those free in solution are a common phenomenon in biology, making this an important specific example of a widespread biological idea.

In the second chapter, we explore the control parameters that govern the dynamics of *in vitro* DNA ejection in bacteriophage lambda. Previous work demonstrated that bacteriophage DNA is highly pressurized, and this pressure has been hypothesized to help drive DNA ejection. Ions influence this process by screening charges on DNA; however, a systematic variation of salt concentrations to explore these effects has not been undertaken. To study the nature of the forces driving DNA ejection, we performed *in vitro* measurements of DNA ejection in bulk and at the single-phage level. We present measurements on the dynamics of ejection and on the self-repulsion force driving ejection. We examine the role of ion concentration and identity in both measurements, and show that the charge of counterions is an important control parameter. These measurements show that the mobility of ejecting DNA is independent of ionic concentrations for a given amount of DNA in the capsid. We also present evidence that phage DNA forms loops during ejection, and confirm this effect using optical tweezers.

In the third chapter, we use a case study built around phage lambda to study the transfer of biopolymers across membranes with specific reference to the transfer of DNA across bacterial membranes. Ever since Hershey and Chase used bacterial viruses to establish DNA as the molecular carrier of genetic information in

1952, the mechanism of DNA translocation for phage has been a mystery. While bounds have been established on a minimum time of *in vivo* DNA translocation, previous measurements of DNA translocation by single bacteriophages have only been made *in vitro*. Here, we present the first visualization of single bacteriophages infecting individual *Escherichia coli* cells by using the fluorescence of DNA-bound cyanine dyes. These experiments reveal that *in vivo* DNA translocation occurs on the time scale of 1 to 20 minutes, significantly longer than single-molecule ejections observed *in vitro*. We also measure the cell-to-cell variability in the translocation process, and show that a number of trajectories have significant pauses. In addition, we present a system for simultaneously monitoring the growth of both phage DNA and phage protein from infection to lysis at the single-cell level. This work provides a picture of the early stages of the infection process and sheds light on both the biology of viruses and more generally, the problem of polymer translocation.

While each of these chapters can be presented as an individual story, they are connected in spirit by goal of "mathematizing the cartoons of biology". It is my hope that the insights learned from this thesis stem not just from the science described within, but also in the approach it employs for understanding and studying living systems.

Chapter 2

Biochemistry on a leash: Information processing in biological systems

2.1 Introduction

Flexible chains of amino acids, referred to here as tethers, are important for the function of natural and synthetic proteins. Four examples of natural proteins in which such tether motifs are thought to play an active role are actin polymerization by formin proteins, the autoregulation of Src family kinases, the methylation of bacterial chemotaxis receptors, and the inactivation of potassium channels. Several of these examples are highlighted schematically in figure 2.1.

Actin is an important biopolymer; it is important for cell division, motility, and other tasks like phagocytosis. Given its vital role in the life of a cell, it is no surprise that the formation of actin filaments is tightly controlled by the cell. Formins are one class of proteins that participate in this process by promoting the growth of actin filaments. Formins have been shown to increase the rate of polymerization by as much as 5- or 15-fold; how they achieve this speed up remains an open question [4,5]. Recently, a physical model for this speed up has been proposed that treats the activation process like a ball in a cup [6]. In this model, depicted in figure 2.1a, flexible tethers connect binding sites for profilin-bound actin to an FH2 domain, the portion of formins known to bind to the barbed end of actin filaments. The flexible tethers serve to deliver actin monomers to the barbed end, enabling the polymerization process to proceed more rapidly than permitted by the diffusion limit. While this model has yet to be validated, it is an intriguing hypothesis that such tethers play an active role in formin-mediated actin polymerization

Src kinases are another family of proteins in which tethers are thought to play a prominent role [7, 8]. Src kinases are signaling proteins that control a number of cellular processes including adhesion and cell division. In particular, Src kinases are thought to possess both an inactive and active conformation. They also possess SH2 and SH3 domains, protein domains that recognize and bind to small peptide sequences.

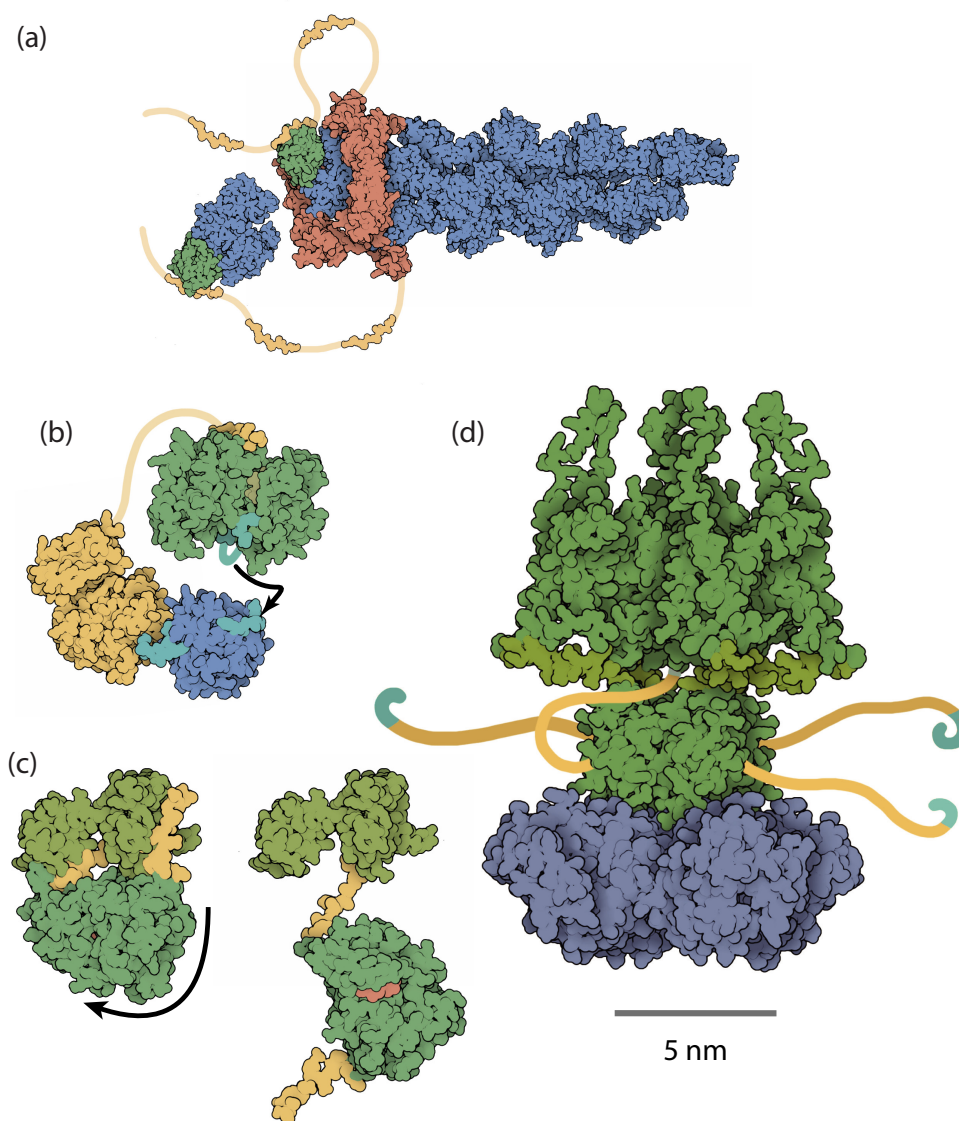


Figure 2.1: Biological examples of biochemistry on a leash. Receptors and ligands connected by a flexible tether are a common motif in biological systems. Images are courtesy of David Goodsell, Scripps Research Institute, La Jolla, CA. (a) Formin-mediated actin polymerization. In one model, formin proteins can use flexible tethers to grab actin monomers in solution and deliver them to the end of a growing actin filament. (b) A Ste5 scaffold protein (blue) is in complex with the kinase Ste7 (yellow), which binds the protein Fus3 (green) via a flexible linker. The tether-like linker helps recruit Fus3 to Ste5, where conversion to the active state is catalyzed. (c) Inactivation of Src tyrosine kinase. A flexible linker connects the SH2 domain and the tail portion of Src tyrosine kinase. It is thought that the binding of the phosphorylated tail to SH2 locks the protein in an inactive conformation. Theoretical studies suggest that the properties of the tether can influence the protein's function. (d) A Shaker-related potassium channel with flexible "chains" on the α subunit N-terminal ends, that terminate in a "ball". A few residues of one "ball" bind inside the pore and block the passage of ions.

In Src kinases, these domains are connected to their respective ligands by flexible tethers. In the current model of Src regulation, the SH2 and SH3 domains bind to their tethered ligands to stabilize the inactive conformation; this model is shown in figure 2.1b. The receptor-ligand interaction is suboptimal, enabling ligands free in solution to compete for binding with the tethered domain and activate the Src kinase [7, 8]. Experimental and theoretical studies have shown that the regulation is dependent on the composition of the tethers connecting the multiple subunits [9, 10]. In this system and others, the role tethers play can be thought of as biochemistry on a leash. In these cases, there is a competition between the tethered ligands and those that are free in solution. Simple physical models and clever experiments are starting to reveal how this competition plays out [11]. A quantitative picture will deepen our understanding of how signaling proteins work.

The inactivation of the voltage-gated Shaker potassium channel is another instance in which tethers are thought to play an important role in the biological function of a protein. The Shaker potassium channel adopts an open conformation when a large transmembrane voltage is applied, leading to a flux of potassium ions across the membrane. Once open, the channel can be spontaneously inactivated in a voltage-independent fashion; mutational studies by Hoshi et al. identified an N-terminal cytoplasmic portion of the ion channel as the portion responsible for this rapid inactivation [12–15]. The current model for inactivation posits that the cytoplasmic domain consists of a ball connected to a leash [14, 15]. The ball can bind to the pore and block the flux of potassium ions. In this model, the tether serves to effectively increase the concentration of balls seen by the open pore. Interestingly, the degree of inactivation has been shown to be dependent on the length of the tether. Mutants with longer tethers have less inactivation while mutants with shorter tethers exhibited more inactivation. Work by Timpe et al. demonstrated that this behavior was consistent with a model in which the tethers were modeled as a random walk [16]. As exhibited by these examples, simple polymer models have much to tell us about how tethers can influence the behavior of natural proteins.

Tethers are also believed to play an important role in the methylation of the chemotaxis receptors Tar and Tsr. The flexible C-terminus of these receptors is thought to bind to the methyl-transferases CheR and CheB, linking them to the receptor with a flexible tether. While tethered, these proteins can methylate the chemotaxis receptor, an event crucial for adaptation. The flexible tethers linking the transmembrane receptor with the methyl-transferase has been the focus of recent theoretical work [17]. Through use of a dynamic description of the tethers, Windisch et al. have demonstrated how tethering can lead to a more efficient search for substrates as well as control the orientation of the tethered CheR and CheB [17].

Another example of tethering is given by the scaffold Ste5, which is depicted in figure 2.1b, a scaffold that organizes the mating response pathway in budding yeast. Scaffolds are thought to be important organizers of signaling pathways; it has been postulated that they shape the input-output properties of signaling networks

by co-localizing enzymes and increasing their effective concentrations [18]. The case of Ste5 is intriguing because it is a concrete example of how a scaffold can direct the flow of information from stimulus to response inside the cell. The kinase Ste7 is a member of both the mating and starvation response pathways in yeast. When exposed to a mating signal, Ste7 activates the protein Fus3; the scaffold Ste5 is necessary to direct Ste7 down this pathway. In its absence, the starvation pathway is activated instead [18–20]. Recent work has shown how this change in information flow occurs [21]. Fus3 has both an active and inactive state and Ste5 is capable of catalyzing Fus3’s conversion to its active conformation. Curiously, there also exists a tethering interaction mediated by Ste7 that keeps the scaffold and Fus3 at a high effective concentration. In the absence of this interaction, a 100-fold higher concentration of scaffold is required to reach wild-type levels of Fus3 activation.

In addition to their role in natural systems, tethers have also been used to engineer proteins with novel and improved functions. Several studies have demonstrated that tethers can be used to enhance the stability of peptide homo-dimers. The Arc repressor is a protein made by bacteriophage P22 that is involved in the lysis-lysogeny decision. Wild-type Arc exists as a homodimer, as the monomers are unable to fold into a stable structure by themselves. Work by Robinson et al. has demonstrated that by tethering two Arc repressors together with a peptide linker, one can make a hybrid protein where the folded state is the preferred equilibrium configuration [22, 23]. The protein Rop, a homodimer that regulates plasmid replication, is another instance where the relationship between tethering and protein stability has been explored [24]. Interestingly, in both studies, synthetic biology was used as a platform to study the physical ramifications of biochemistry on a leash. The role the tether played in these systems was queried by systematically changing either the tether composition or tether length. Robinson et al. demonstrated that the changes in tether composition could influence both the folding and unfolding rate, leading to a six-order-of-magnitude change in the effective concentration each monomer saw of its tethered partner [23]. The work of Nagi et al. showed that in the case of Rop, the tether length dependence of the protein stability is consistent with a simple random walk model of the tether that accounts for excluded volume interactions [24]. In these instances, attempts to understand design principles for protein engineering have led to insights into biochemistry on a leash.

Recent work by Krishnamurthy et al. has demonstrated how tethers might be used to construct ligands that bind to multiple sites on a protein [11]. In their work, they connected human carbonic anhydrase and its ligand using a flexible polyethylene glycol tether. In this system, they were able to quantify the competition between free and tethered ligands and measure how this competition varied when the length of the tether was varied. From this, they were able to extract simple design rules for multivalent ligands. While this work has consequences for the ligand design, the results also speak to how biochemistry on a leash plays out in

natural biological systems.

Synthetic biologists have also used the ideas of biochemistry on a leash to engineer entirely synthetic scaffolds to optimize the yield of metabolic pathways. Metabolic engineering aims to turn living cells into factories capable of producing chemicals of interest [25]. When subverting a cell's molecular machinery, it is sometimes necessary to move an entire metabolic pathway from one organism into the more suitable host. This can present an engineering challenge, as the heterologous metabolic pathway can suffer from low efficiency when compared to its native host. One approach to solving this problem has been to use the idea of scaffolding [26, 27]. Dueber et al. have recently used scaffolds to improve the product yield when the melvalonate pathway of yeast is brought into *E. coli* [26]. Interestingly, melvalonate is a chemical precursor to artemisinin, an antimalarial drug. In their work, they created entirely synthetic scaffolds that assembled the key enzymes of this pathway into a complex, ensuring that all of the enzymes involved were spatially localized. By controlling the stoichiometries of each complex member, they were able to show that scaffolding can produce a 77-fold increase in product titer. The idea of using synthetic scaffolds has also been extended to improve the synthesis of glucaric acid from a novel metabolic pathway [27]. Each of these cases highlights how tethering motifs can be taken out of their natural context or designed from scratch to perform novel functions.

Another instance in which tethers have been used to design new proteins is the rewiring of the WASP family of signaling proteins [28]. Wiskott Aldrich Syndrome proteins, or WASPs, are a family of proteins that regulate the creation of new actin filaments by the Arp2/3 complex [29–34]. One common feature of these proteins is that activation of Arp2/3 is carried out by a C-terminal VCA domain. The VCA domain binds both the Arp2/3 complex and actin, and has been shown to deliver the first monomer in a new actin filament [31, 35]. As with Src kinases, regulation occurs through auto-inhibitory interactions that stabilize an inactive conformation [29, 34, 36]. Chemical signals can bind and competitively relieve this auto-inhibition; the binding of Cdc42 and PIP₂ to the family member N-WASP is a classic example [33, 36].

As with regulatory networks, reprogramming is often a metric for understanding [37–42]. Recent work in modular recombination has produced synthetic N-WASPs by tethering a VCA domain to a receptor and its associated ligand [28]. The interaction between the receptor and its tethered ligand serve to mimic the auto-inhibitory interactions in native N-WASP. Some of the richest results of this recombination technique involve constructs comprised of multiple receptor-ligand pairs on the same tether (see figure 2.6). In these cases, a binding partnership between one receptor-ligand pair will physically shorten the tether connecting the remaining pair, resulting in a different free energy of closure and novel signal integration behavior.

In conjunction with reprogramming, WASPs have also been the focus of quantitative modeling [28, 33, 43, 44]. Engineering and theoretical approaches have proven useful, as they serve to challenge and sharpen

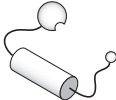
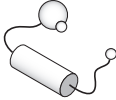
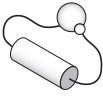
our understanding. In particular, one role of theory is to test models by having dials that can be tuned to produce falsifiable predictions. The cartoons that depict these proteins suggest the geometry of the tethers can be tuned to elicit different biological functions. Experiments have shown that changing tether lengths can affect the behavior of synthetic proteins; theoretical studies have reported similar findings [10, 23, 24, 28, 45]. The goal of this section is to examine how the general features of tethers influence the behavior of natural and synthetic proteins. In particular, we develop a quantitative model for tethered ligand-receptor pairs and show through a case study on N-WASP how the features of tethers can control the behavior of signaling proteins.

The outline of the remainder of this section is as follows. In section 2 we present our model for the interplay between free and tethered ligands. In section 3 we outline a simple random walk model used to describe flexible tethers. In section 4 we apply the model to predict the tether length dependence of a one-input synthetic N-WASP construct. We also extend our statistical mechanical model to study the input/output properties of multiple input constructs and illustrate how tethers can influence the signal integration properties of these proteins. We finish our discussion in section 5 by discussing the role tether geometry plays in protein-mediated signal integration and discussing how this model can be used to examine other cases of biochemistry on a leash.

2.2 Statistical mechanical model of tethered receptor-ligand pairs

In this work, we restrict ourselves to the case where tethered receptor-ligand binding inhibits protein activity. Our motivating example for selecting this motif are the synthetic WASP proteins that regulate actin polymerization by Arp2/3. A schematic for one of these constructs, which we call the simple switch, is outlined in figure 2.2. The receptor and tethered ligand, also called cis-ligand, are each connected to a tether. The two tethers originate from distinct, fixed locations in space. This is likely the case if there are some intervening secondary structural elements between the origins of the two tethers. In this motif, the switch is considered to be off when the receptor and cis-ligand are bound and on otherwise.

The starting point for our model is the assumption that the switches are effectively in thermal equilibrium. Because life is an inherently non-equilibrium process, care must be exercised when attempting to apply an equilibrium model to a biological system. Our model of the simple switch is one such case. We imagine that the active states of the switch are coupled to some productive reaction, as envisioned in figure 2.11. The intuition behind our equilibrium assumption is a separation of time scales between ligand binding and this productive reaction. Binding of ligands and the spatial fluctuations of unstructured amino acids are fast, occurring on a time scale of milliseconds or less [46–48]. For our motivating example, this process is Arp2/3-

State	Output	Boltzmann Factors	Thermodynamic Weight
	On	$\Omega_{on}^{tet} \Omega_L^{sol}$	1
	On	$\Omega_{on}^{tet} \Omega_{L-1}^{sol} e^{-\beta \epsilon_f}$	$\frac{[L]}{K_d}$
	Off	$\Omega_{off}^{tet} \Omega_L^{sol} e^{-\beta \epsilon_t}$	$\frac{J}{K_d}$

$$p_{on} = \frac{\text{On state 1} + \text{On state 2}}{\text{On state 1} + \text{On state 2} + \text{Off state}}$$

Figure 2.2: States and weights for the simple switch. The tethered ligand functions as an inhibitor and locks the protein into an inactive state when bound to the receptor. Ligands in solution can compete with cis-ligand to activate the switch. Statistical mechanics assigns statistical weights to each state and allows us to compute the probability of the protein to be in an active state, as shown in the ratio at the bottom.

mediated actin polymerization, a process that occurs on the timescale of seconds or longer. Because of this separation of timescales, we expect the productive reaction to only care about the equilibrium behavior of the switches. We give this argument a more concrete mathematical treatment in the appendix.

We are interested in finding the probability that a simple switch occupies an active state, a quantity we label p_{on} . In our model, this quantity is our metric for the activity of the simple switch. To proceed with the calculation, we first invoke the standard statistical mechanics toolkit. Statistical mechanics tells us that the probability of the system being in a given collection of microscopic states is given by

$$p_{state} = \frac{ge^{-\beta\epsilon}}{Z}, \quad (2.1)$$

where g is the degeneracy of the collection, $\beta \equiv \frac{1}{k_B T}$, ϵ is the energy of the collection, and Z is a normalization factor called the partition function. For our problem, there are three factors that determine the statistical weight of a collection — the strength of the interaction between the receptor and its ligands, the configurations of the ligands in solution, and the configurations of the tethers. Simple arguments lead to expressions for all three factors.

We treat the interaction between the receptor and its ligands by letting free ligand bind to the receptor with a binding energy of ϵ_f and cis-ligand bind with energy ϵ_t . As shown in figure 2.2, this identifies three different energy levels; one in which no ligand is bound, one in which free ligand is bound, and one in which cis-ligand is bound. Applying equation 2.1, we can write the probability the switch is active as

$$p_{on} = \frac{\sum_{i \in \text{active states}} g_i e^{-\beta\epsilon_i}}{\sum_{i \in \text{all states}} g_i e^{-\beta\epsilon_i}}. \quad (2.2)$$

The configurations of the ligands in solution and the configurations of the tethers contribute to the degeneracy of each energy level. Assuming independence of the contributions, we can write the degeneracy of the i^{th} energy level as $g_i = \Omega_i^{sol} \Omega_i^{tet}$, where Ω_i^{sol} and Ω_i^{tet} are the number of configurations available to the ligands in solution and tethers, respectively. In what follows, we will assume that the ligands are in excess of the switches; that is, the switches do not compete with each other for ligands in the solution. If there is one ligand in solution, then the number of configurations for that ligand scales with the volume of the solution (here we invoke ideas appropriate to a dilute solution). With L indistinguishable ligands in a solution of volume V , the number of configurations available to them scales as $\Omega_L^{sol} \propto V^L/L!$. When a ligand in solution binds to the receptor, its available volume is restricted to a small volume v_s . The number of configurations then scales as $\Omega_{L-1}^{sol} \propto V^{L-1}v_s/(L-1)!$. We can account for the configurations of the tethers by appealing

to basic polymer physics. The physical intuition is that the tethers can explore more configurations when the receptor is free than when it is bound to its cis-ligand, making it entropically favorable for the switch to be on. Models from polymer physics allow us to quantify this effect. What we wish to know are Ω_{on}^{tet} and Ω_{off}^{tet} , the number of configurations available to the tether in the on and off states, respectively. In the active state, we assume the tethers are unconstrained. In the inactive state, the configurations available to the tethers are restricted because their binding of the receptor with its cis-ligand forces the tether ends to lie near each other. The ratio between Ω_{off}^{tet} and the total number of possible tether configurations is the probability that the tether ends are close enough for binding to occur. In other words, if the probability of the two tether ends being close is p_{loop} , then

$$\frac{\Omega_{off}^{tet}}{\Omega_{on}^{tet}} = p_{loop}. \quad (2.3)$$

With these expressions in hand we compute the statistical weight for each state. Further, as shown in figure 2.2, we can write the probability that the switch is active as

$$p_{on} = \frac{\Omega_{on}^{tet}\Omega_L^{sol} + \Omega_{on}^{tet}\Omega_{L-1}^{sol}e^{-\beta\epsilon_f}}{\Omega_{on}^{tet}\Omega_L^{sol} + \Omega_{on}^{tet}\Omega_{L-1}^{sol}e^{-\beta\epsilon_f} + \Omega_{off}^{tet}\Omega_L^{sol}e^{-\beta\epsilon_t}}. \quad (2.4)$$

By dividing the top and bottom of this expression by $\Omega_{on}^{tet}\Omega_L^{sol}$, this expression reduces to

$$p_{on} = \frac{1 + \frac{\Omega_{L-1}^{sol}}{\Omega_L^{sol}}e^{-\beta\epsilon_f}}{1 + \frac{\Omega_{L-1}^{sol}}{\Omega_L^{sol}}e^{-\beta\epsilon_f} + \frac{\Omega_{off}^{tet}}{\Omega_{on}^{tet}}e^{-\beta\epsilon_t}}. \quad (2.5)$$

By definition, $\Omega_{L-1}^{sol}/\Omega_L^{sol} = v_s[L]$, where $[L]$ denotes the concentration of ligand. Thus, the probability p_{on} can be expressed as

$$p_{on} = \frac{1 + \frac{[L]}{v_s}e^{-\beta\epsilon_f}}{1 + \frac{[L]}{v_s}e^{-\beta\epsilon_f} + \frac{1}{v_s}p_{loop}e^{-\beta\epsilon_t}}. \quad (2.6)$$

At this point, it is convenient to define a probability density

$$J = \frac{p_{loop}}{v_s}, \quad (2.7)$$

characterizing the concentration of the tethered ligand in the vicinity of the tethered receptor. In this simple model, the receptor and tethered ligand are treated as point objects and are free to bind when the tether ends are within a small volume of each other. Naturally, the looping probability will depend on what this volume is; however, the probability density will not. Furthermore, we can formulate our entire problem in

terms of probability densities and calculate them directly from different polymer models. Readers familiar with DNA looping will find this probability density very similar to the J factor first used by Jacobson and Stockmayer to describe DNA cyclization [49]. Using the probability density, we can rewrite p_{on} as

$$p_{on} = \frac{1 + \frac{[L]}{\frac{1}{v_s} e^{\beta\epsilon_f}}}{1 + \frac{[L]}{\frac{1}{v_s} e^{\beta\epsilon_f}} + \frac{J}{\frac{1}{v_s} e^{\beta\epsilon_t}}} = \frac{1 + \frac{[L]}{K_d}}{1 + \frac{[L]}{K_d} + \frac{J}{K'_d}}, \quad (2.8)$$

where $K_d = \frac{1}{v_s} e^{\beta\epsilon_f}$ and $K'_d = \frac{1}{v_s} e^{\beta\epsilon_t}$ denote dissociation constants. Grouping terms into dissociation constants is useful because they provide a scale for concentrations and because they can be measured experimentally.

The physical interpretation of equation 2.8 is straightforward. The denominator reflects a competition between the ligands in solution and ligands that are tethered. Binding of free ligand to the receptor has a favorable energy but reduces the entropy of the ligands in solution. Binding of cis-ligand to the receptor is also energetically favorable but reduces the configurations available to the tethers, thus reducing their entropy. When $\frac{[L]}{K_d} \gg \frac{J}{K'_d}$, the free ligand wins the competition and $p_{on} \approx 1$. When $\frac{J}{K'_d} \gg \max[\frac{[L]}{K_d}, 1]$, on the other hand, the cis-ligand wins and $p_{on} \approx 0$. Also note that the probability density J has the units of concentration, and hence can be thought of as an effective concentration. In other words, it is the concentration of cis-ligand seen by the receptor. To quantify the interplay between the entropy of the solution and the entropy of the tethers, we need a physical model of the tethers. One such model, the random walk model, is outlined in the following section.

2.3 Polymer physics of protein tethers

In order to have a predictive framework for thinking about the tether-length dependence of signaling protein function, we must have a quantitative description of the competition between the tethered ligand and its untethered counterparts in solution. Specifically, we wish to employ a polymer model to describe the tethers and calculate the probability density of the tether ends. Our ultimate goal is to compute the effective concentration of cis-ligand seen by the receptor. We note that the meeting of two tethers with distinct origins is a problem that has been considered previously in the literature [50]. There are a variety of possible approaches to modeling tethers and in this paper we employ the simplest of them, a random walk model. The limitations of this model are briefly discussed at the end of the section.

In the random walk model, the tether is coarse-grained into a collection of independent segments, as depicted in figure 2.3. Each segment is allowed to rotate freely in space. The tether is composed of N

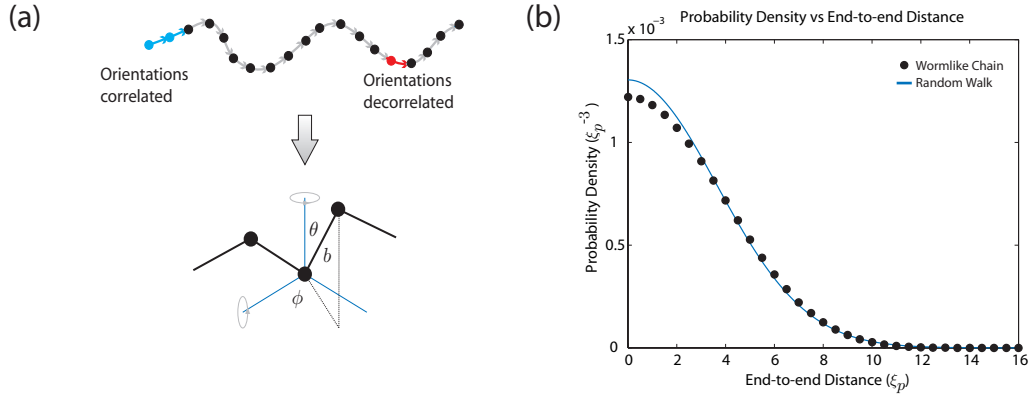


Figure 2.3: A random walk model for flexible polypeptide chains. (a) In this model, the polypeptide chain is separated into statistically independent segments, called Kuhn segments, of length b . The orientation of two monomers is perfectly correlated if they are in the same segment and completely uncorrelated if they are in different segments. (b) The probability distribution for the end-to-end distance vector for a chain of length $L = 20\xi_p$ computed with both the wormlike chain and random walk models. The computation for the wormlike chain was implemented using the method of Samuel et al. and is described in the appendix [51]. A comparison between the two polymer models shows that the random walk model is an acceptable approximation of the wormlike chain model for long chains.

statistically independent Kuhn segments of length b . The i^{th} segment has orientation \vec{r}_i and the end-to-end distance vector is given by $\vec{R} = \sum_{i=1}^N \vec{r}_i$. It follows from the central limit theorem that the probability distribution for the end-to-end distance vector, $G(\vec{R})$, takes the form of a Gaussian. The spatial components of \vec{R} are independent for long tethers, so we can write $G(\vec{R}) = G_x(x)G_y(y)G_z(z)$. The central limit theorem demands that G_x, G_y , and G_z be Gaussian as well. Gaussian distributions are characterized by their mean and variance. Symmetry demands the means be zero; the variance is given by $\langle \vec{R}^2 \rangle = Nb^2$. Independence implies the variances are given by $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = \frac{Nb^2}{3}$. Using these relations, we see that the probability distribution of \vec{R} is given by

$$G(\vec{R}; \xi, L) = \left(\frac{3}{4\pi\xi L} \right)^{\frac{3}{2}} \exp \left[-\frac{3\vec{R} \cdot \vec{R}}{4\xi L} \right], \quad (2.9)$$

where $L = Nb$ is the total length of the random walk and $\xi = \frac{b}{2}$ is its persistence length.

Because the receptor and cis-ligand are each connected to a tether, this simple model allows us to find their probability densities at each point in space and compute the effective concentration of cis-ligand seen by the receptor. We start by assuming the the receptor's tether has length L_1 and begins at some point \vec{r}_1' while the cis-ligand's tether has length L_2 and begins at \vec{r}_2' . If we denote the end of the receptor's tether as \vec{r}_1 and the end of the cis-ligand's tether as \vec{r}_2 , then the probability distribution for the tether ends are given by $G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1)$ and $G(\vec{r}_2 - \vec{r}_2'; \xi_T, L_2)$, respectively. To find the probability density of the cis-ligand as seen by the receptor, we must count the tether conformations where the receptor and cis-ligand are in close proximity to each other. This is accomplished by integrating the joint probability distribution of the

two tether ends, $G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1) G(\vec{r}_2 - \vec{r}_2'; \xi_T, L_2)$, over all possible positions of the two tether ends while employing a delta function to ensure that we only count the conformations where the tether ends are in the same location. The effective concentration is then

$$\begin{aligned} J &= \int d^3r_1 d^3r_2 \delta^{(3)}(\vec{r}_1 - \vec{r}_2) G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1) G(\vec{r}_2 - \vec{r}_2'; \xi_T, L_2) \\ &= G(\vec{D}; \xi_T, L_T), \end{aligned} \tag{2.10}$$

where $L_T = L_1 + L_2$ is the combined length of both tethers, ξ_T is their persistence length, and $\vec{D} = \vec{r}_1' - \vec{r}_2'$ is the vector between the origin of the two tethers. Note that in the above calculation, we have made use of the fact that the convolution of two Gaussians is itself a Gaussian.

The random walk model provides a simple way to compute the effective concentration of cis-ligand. However, this model is not without its flaws. For example, the wormlike chain (WLC) model, which attributes a bending energy for different tether conformations, has been shown to fit force-extension measurements and distributions of length sizes extracted from protein databases [52–55]. Other models, such as the rotational isomeric state model or the three-bead rotating chain are justified on experimental and structural grounds [56, 57]. More complicated models like the WLC are necessary to examine short tethers. On the other hand, the WLC and RW models provide the same quantitative description for long tethers (unpublished data). Further, none of the above models address the influence of self avoidance and nearby macromolecules on tether conformation. These excluded volume effects, particularly the role of nearby macromolecules, have been the subject of theoretical modeling [17, 58]. Recent experimental work suggests that they play an important role in tethered receptor-ligand systems [11]. While these concerns are important, we have chosen this model because our goal was to develop a simple model to explore tethering in proteins and to illustrate the kinds of effects such tethering can give rise to. Furthermore, we expect that our *qualitative* conclusions will remain valid even when self avoidance becomes important. Further consideration of excluded volume effects is beyond the scope of this work and should be the focus of future studies.

2.4 Signaling proteins: Roles of tether geometry and length

At this point, we are in a position to apply the tools from sections 2.2 and 2.3 to investigate the roles of tether length and geometry on the behavior of signaling proteins. First, we will discuss the case of simple switches, which contain a single ligand-receptor pair. Then, we will describe the cooperative integration of signals in more complex switches which contain multiple ligand-receptor pairs.

2.4.1 Simple switches

Our motivating example for the statistical mechanical model are the reprogrammed WASPs that link Arp2/3-dependent actin polymerization to chemical signals. In particular, we are interested in predicting how the properties of the tethers that connect the different domains influence the function of these proteins. To obtain such a prediction requires a number of model parameters. Fortunately, a number of studies exist on the structural biology of WASPs and the physical nature of amino acid chains that provide us with physical parameters that can be used to render the model predictive. Using the rule of thumb that one amino acid contributes ~ 0.4 nm to a protein's contour length, we estimate the flexible portions of the VCA domain to be ~ 21.6 nm in length. These portions are the black segments shown in figure 2.4a. The receptor, which in the case of the simple switch is a PDZ domain, is also thought to have unstructured residues at the C-terminus; we estimate these amino acids to be ~ 2.8 nm in length. For the simple switch studied by Dueber et al., we assume these are the only contributions to the tether [28]. Experimental and theoretical evidence suggests unstructured amino acids have a persistence length of ~ 0.4 nm [52–55, 59].

The remaining unknown parameter is the size of the rigid portion of the VCA domain, D as indicated schematically in figure 2.4. Recall that D arises in the model because the tethers for the receptor and ligand originate from different points. There are a number of reasons to expect this to be the case for the VCA domain. First, experiments have shown that N-WASP will bind Arp2/3 whether or not it is in an active conformation [33]. This data leads to the working assumption that the simple switch is always bound to Arp2/3. While this may not always be the case, only the subpopulation of Arp2/3s in solution that are bound to an N-WASP will contribute significantly to the creation of new actin filaments. Hence, we are justified in limiting the discussion to this case. Because the VCA domain of N-WASP is mostly unstructured, it is reasonable to ask where on Arp2/3 the VCA domain binds. Yeast two-hybrid experiments have shown that the acidic residues (i.e., the A domain in VCA) bind the protein p21, a member of the Arp2/3 complex [30]. The verprolin domain, or V domain, is known to bind actin monomers. Because the conserved domain, or C domain, has homology to the V domain, it has been proposed that it binds to the actin homolog Arp2 [59]. The fivefold decrease in the dissociation constant for VCA binding to Arp2/3 when the C domain is present supports this hypothesis [31]. If these two interactions exist, then they likely serve to hold the two tether origins fixed in space at separate points. This is illustrated in figure 2.4. Hence, the justification for applying this simple model to WASPs rests on the existence of these two interactions.

To find D , we turn to the titration experiments of the simple switch by Dueber et al. [28]. In this experiment, the relative activity, a measure of the speed up in actin polymerization caused by activated WASPs, was measured as a function of ligand concentration. We fit this experiment with fixed tether

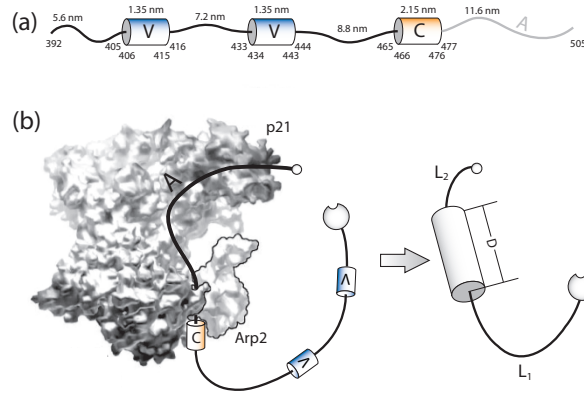


Figure 2.4: A physical model for the output domain of N-WASP. The VCA domain is thought to be a mixture of alpha helices and unstructured amino acids. (a) Separation of structure into residues assumed to have secondary structure and residues assumed to be unstructured. Here, cylinders represent stable alpha helices. (b) Model for binding of the VCA domain to Arp2/3 [60]. In our model, binding of VCA to Arp2/3 fixes the location of the C and A domains. This allows us to treat the C and A domains as a rigid cylinder and the rest of the VCA domain as a flexible tether.

lengths by assuming that p_{on} is equal to the relative activity and using least-squares to find the effective concentration that best matched their data. With this fit, we estimate D to be ~ 7.94 nm. This value is close to the distance between Arp2 and p21 in the Arp2/3 complex, the two members of the Arp2/3 complex that have been proposed as binding sites for the C and A domains, respectively. Finally, the dissociation constants for the receptor's interaction with both free and cis-ligand is taken to be $\sim 8 \mu\text{M}$ [28]. With these numbers, we can predict how the effective concentration (and hence the activity) changes with tether length.

The prediction for the tether length dependence of the simple switch is shown in figure 2.5. Lengthening short tethers increases effective concentrations while the opposite effect is seen for very long tethers. The crossover between the two regimes occurs at $L_T = \frac{D^2}{2\xi_T}$. This scaling behavior arises because a portion of the output domain is treated as a rigid rod. To enter a looped state, the tethers must cross this rod, a task that is difficult when the tether length is comparable to the rod size D . In this regime, extending the tethers across the rod becomes easier with longer tethers. When the tethers are very long, the opposite is seen. Because the rod is short compared to tethers, it has a small effect on how often the two tether ends meet. If the rod is ignored, then our simple polymer model treats the whole construct as a random walk. Longer walks decrease the chance the walk ends at the beginning, implying that longer tethers will decrease the likelihood of PDZ seeing its cis-ligand. In particular, in this regime the effective concentration obeys the scaling law $J \propto L_T^{-3/2}$.

There are several points about this model that are worth noting. First, we are assuming that p_{on} is equivalent to the relative activity that is measured in experiments. While the relative activity is likely a

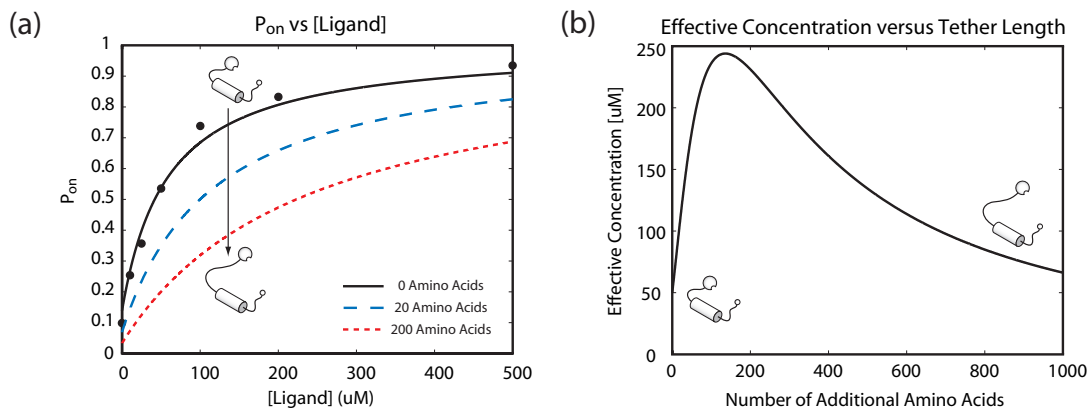


Figure 2.5: The tether length dependence of the simple switch. (a) The probability of the switch being active as a function of ligand concentration and tether length. The parameters used for the case study on reprogrammed N-WASP are described in the main text. Zero additional amino acids corresponds to the experimental construct with a tether length of ~ 24.4 nm. The additional amino acids used by Dueber et al. consist of serine-glycine repeats. Assuming the relative activity is equivalent to p_{on} , we can use experimental data from Dueber et al. to fit for unknown parameters in our model [28]. (b) The tether length dependence of the effective concentration of cis-ligand. We find that for short tethers, an increase in tether length leads to higher effective concentrations and hence a lower probability of occupying an active state. The reverse effect is seen in the model for very long tethers.

monotonically increasing function of p_{on} , one cannot say the two are equivalent until this has been proven experimentally. The use of this assumption may undermine a quantitative interpretation of our results, though there are other cases such as transcriptional regulation where this kind of thinking is appropriate [61]. This assumption is a reasonable first step and allows us to use our model to interpret experimental results and generate predictions. Second, the short tether regime may be more relevant for experiments. Looping likely renders the construct inactive through a conformational change in the VCA domain. Very long tethers may decouple the connection between looping and switch inactivation. Indeed, some constructs with long tethers are constitutively active [28]. Short tethers avoid this issue and hence might be more useful experimentally. Third, our results depend on our assumptions about the physical properties of the VCA domain while bound to Arp2/3. If one of the putative interactions does not exist, then the VCA domain would remain unstructured while bound to Arp2/3, and the effective concentration would decrease monotonically with increasing tether length with a $-\frac{3}{2}$ scaling exponent. Finally, we note that the tether length prediction here will have interesting consequences for how constructs with multiple tethered receptor-ligand pairs behave in our model. Indeed, the rich behavior afforded by the case of multiple receptor-ligand pairs associated with the same overall tether will serve as the basis of our discussion below.

2.4.2 Complex switches

In this section, we extend the statistical mechanical model described above to examine constructs with multiple receptor-ligand pairs, or complex switches. Such motifs might be one way in which evolution has expanded the signaling inventory by permitting multiple inputs to a given protein that activates other downstream components. The topologies of the complex switches under discussion are shown in figure 2.6. Complex switches accept multiple chemical inputs and exhibit interesting signal integration behavior. Some of the most interesting cases are reprogrammed N-WASP constructs that display antagonistic integration; one input is stimulatory while the other is inhibitory. Our goal is to demonstrate how tethers can play an active role in this decision making and propose physical mechanisms for how signaling proteins integrate multiple signals.

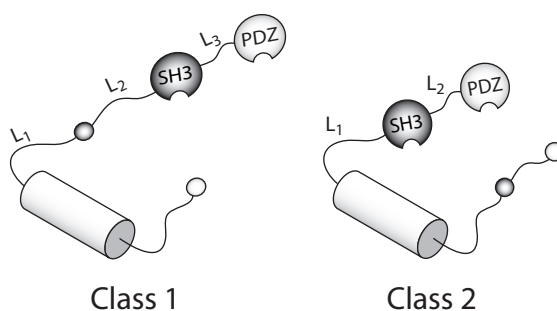


Figure 2.6: The topologies of complex switches. This figure outlines the topologies of the complex switches discussed in this work and highlights the relevant tether lengths. The statistical mechanical model used to examine the simple switch can be extended to examine constructs with multiple tethered receptor-ligand pairs. Note that there are two distinct mechanisms for signal integration. For class one switches, the interaction between the black receptor-ligand pair serves to control the length of the white receptor’s tether. The tether length dependence is then crucial for this construct’s signal integration behavior. For class two switches, the interaction between the black receptor-ligand pair serves to bring the white receptor-ligand pair into closer proximity. This likely increases the effective concentration of white cis-ligand and leads to cooperativity for the binding of the two inputs.

As with the simple switch, we can use statistical mechanics to construct a states and weights diagram for the two topologies in consideration; this is shown in figure 2.7. We restrict ourselves to the case where receptors only interact with ligands of the same color and where looping is the mechanism of inhibition. We point the reader’s attention to two important features of this extension. First, the topology of the complex switch, that is the arrangement of receptors and ligands on the tether, is crucial for determining what role tethers play in the switch’s signal integration properties. In class I switches, the black receptor-ligand pair controls the length of the white receptor’s tether. Because the white receptor-ligand pair controls the switch’s activation, the behavior with respect to tether length is crucial for determining how the switch responds to

white and black inputs. Conversely, class II switches do not share this topology. Interactions between the black receptor-ligand pair merely bring the white receptor-ligand pair into closer proximity (assuming the black receptor is smaller than the size of the cylinder). Hence, in the model, the switch's behavior is determined more by the origins of the tethers, a parameter governed by the secondary structure of the receptors, not the tether's physical properties.

Second, the model now contains three effective concentrations that must be calculated. These are

- The effective concentration of cis-ligand the black receptor sees when the white receptor/cis-ligand pair is unbound.
- The effective concentration of cis-ligand the white receptor sees when the black receptor/cis-ligand pair is unbound.
- The effective concentration of cis-ligand the white receptor sees when the black receptor/cis-ligand pair is bound.

Calculating these effective concentrations is not straightforward, as we can use structures from the PDB in some instances to abandon our point particle description of the receptors and ligands. Because of the physical size of the receptor, the tether ends for the receptor and cis-ligand are held at a fixed distance when the two species are bound; physical structures allow us to estimate this distance. In what follows, we return to our case study of reprogrammed N-WASP to demonstrate how this might be done.

2.4.2.1 Class I switches

As with the simple switch, we can use statistical mechanics to find the statistical weight of each state. The states and weights are shown in figure 2.7. Using these results, we can rewrite the pictorial representation of p_{on} algebraically as

$$p_{on} = \frac{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B][L_W]}{K_B K_W} + \frac{J_1}{K'_B} + \frac{J_1}{K'_B} \frac{[L_W]}{K_W}}{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B][L_W]}{K_B K_W} + \frac{J_1}{K'_B} + \frac{J_1}{K'_B} \frac{[L_W]}{K_W} + \frac{J_1}{K'_B} \frac{J_3}{K'_W} + \frac{J_2}{K'_W} + \frac{[L_B]}{K_B} \frac{J_2}{K'_W}}. \quad (2.11)$$

Here J_1 is the effective concentration of cis-ligand seen by the black receptor when the white pair is unbound, J_2 is the effective concentration of cis-ligand seen by the white receptor when the black pair is unbound, and J_3 is the effective concentration of cis-ligand seen by the white receptor when the black pair is bound. Further, K_B is the dissociation constant for free black ligand, K'_B is the dissociation constant for black cis-ligand, K_W is the dissociation constant for free white ligand, and K'_W is the dissociation constant for white cis-ligand.

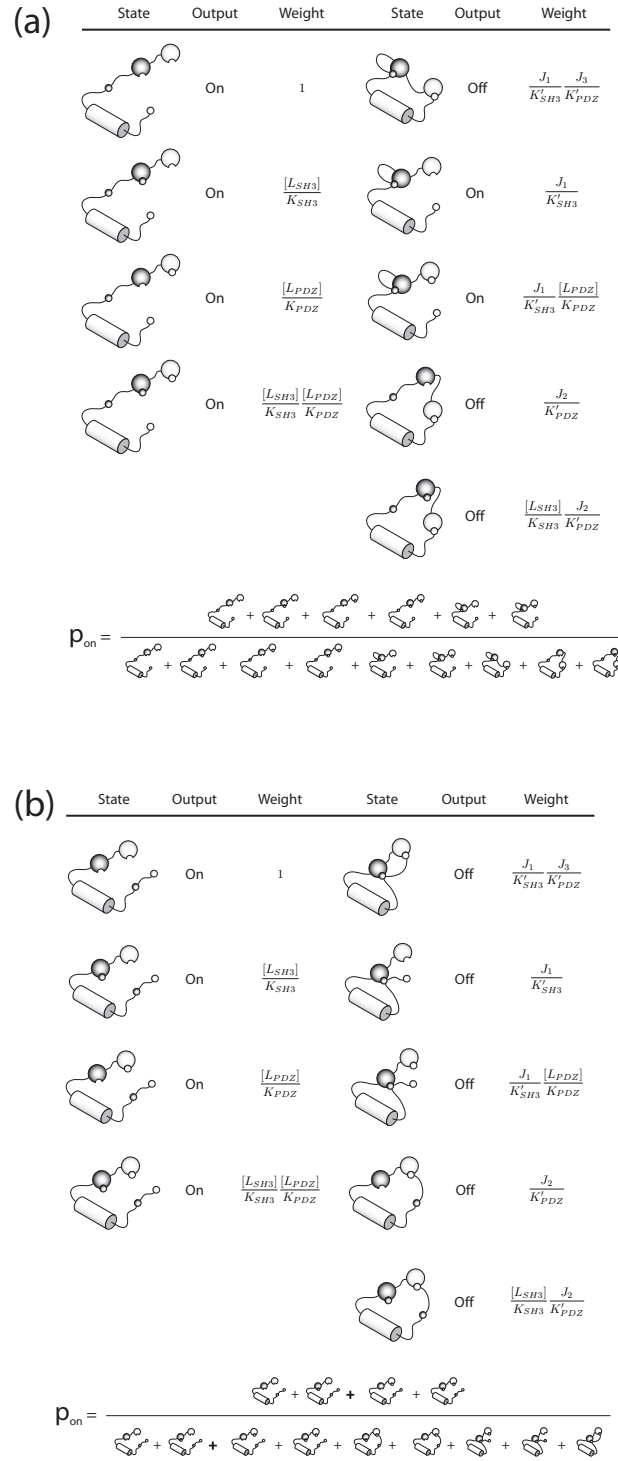


Figure 2.7: The states and weights for class I and class II complex switches. Statistical mechanics can be used to compute the thermodynamic weight of each state and the probability the switch occupies an active state. A complete loop must be formed to inactivate the switch. (a) Class I switches have been observed to display antagonistic gating in experiments. (b) Class II switches display cooperative integration in experiments.

In the reprogrammed N-WASP constructs designed by Dueber et al., an SH3 domain was used as the black receptor and a PDZ domain was used as the white receptor [28]. Fortunately, a crystal structure of SH3 allows us to relax our point particle approximation for the black receptor during our computation of J_1 . The crystal structure of an SH3 domain bound to its ligand is shown in figure 2.8.

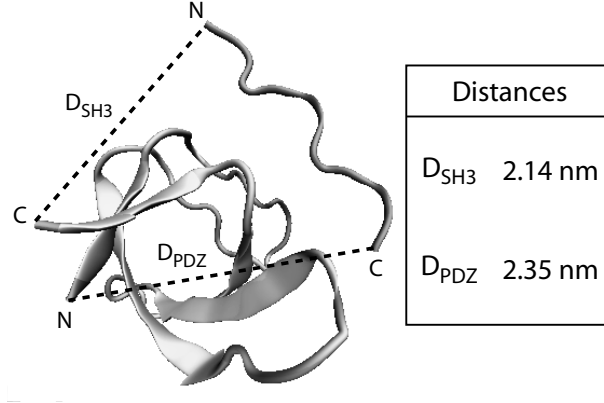


Figure 2.8: NMR structure of the SH3 domain. We can estimate the effective concentration of cis-ligand seen by an SH3 domain in class I switches by appealing to a structure of the bound complex, identification number 1PRM in the Protein Databank [62]. This structure can also be used to compute the effective concentration of cis-ligand seen by a PDZ domain in class II switches.

As seen in the figure, a tether connecting the C-terminus of an SH3 domain to the N-terminus of a ligand would have its beginning and end separated by a distance $D_{SH3} \sim 2.14$ nm in the bound state. Hence, J_1 is given by the expression

$$J_1 = \left(\frac{3}{4\pi\xi_T L_2} \right)^{\frac{3}{2}} \exp \left[-\frac{3D_{SH3}^2}{4\xi_T L_2} \right]. \quad (2.12)$$

The calculation of J_2 and J_3 is more straightforward. When the black receptor/cis-ligand pair is unbound, the tether connecting the two contributes to the length of the white receptor's tether. The total length of the white receptor's tether is then $L_1 + L_2 + L_3$, and J_2 is given by

$$J_2 = \left(\frac{3}{4\pi\xi_T (L_1 + L_2 + L_3)} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T (L_1 + L_2 + L_3)} \right]. \quad (2.13)$$

When the black receptor/cis-ligand pair is bound, the tether connecting the two is looped out and does not contribute to the length of the white receptor's tether. The total length of the white receptor's tether is then

$L_1 + L_3$, and J_3 is given by

$$J_3 = \left(\frac{3}{4\pi\xi_T(L_1 + L_3)} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T(L_1 + L_3)} \right]. \quad (2.14)$$

2.4.2.2 Class II switches

As with class I switches, we can use the pictorial representation of class II switches to write p_{on} algebraically as

$$p_{on} = \frac{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B][L_W]}{K_B K_W}}{1 + \frac{[L_B]}{K_B} + \frac{[L_W]}{K_W} + \frac{[L_B][L_W]}{K_B K_W} + \frac{[L_B]J_2}{K_B K'_W} + \frac{J_2}{K'_W} + \frac{J_1}{K'_B} \frac{[L_W]}{K_W} + \frac{J_1}{K'_B} + \frac{J_1 J_3}{K'_B K'_W}}. \quad (2.15)$$

The effective concentrations and dissociation constants have the same meaning for class II switches as they had for class I switches.

The calculation of J_1 and J_2 is straightforward, as we can use the results from the simple switch with tether lengths L_1 and $L_1 + L_2$, respectively. These effective concentrations are given by

$$J_1 = \left(\frac{3}{4\pi\xi_T L_1} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T L_1} \right], \text{ and} \quad (2.16)$$

$$J_2 = \left(\frac{3}{4\pi\xi_T (L_1 + L_2)} \right)^{\frac{3}{2}} \exp \left[-\frac{3D^2}{4\xi_T (L_1 + L_2)} \right]. \quad (2.17)$$

To find J_3 , we can again appeal to a crystal structure of an SH3 domain bound to its ligand. When SH3 is bound to its cis-ligand, the tether for the white receptor, in this case PDZ, originates from the N-terminus of the SH3 domain and the tether for the white ligand, in this case PDZ ligand, originates from the C terminus of the SH3 domain's cis-ligand. The tether origins in this case are separated by a distance D_{PDZ} , and J_3 is consequently given by the expression

$$J_3 = \left(\frac{3}{4\pi\xi_T L_2} \right)^{\frac{3}{2}} \exp \left[-\frac{3D_{PDZ}^2}{4\xi_T L_2} \right]. \quad (2.18)$$

The model parameters used for the case study on reprogrammed N-WASPs are in table 2.1.

2.4.2.3 Applications to reprogrammed N-WASP

To illustrate how tethers can play an active role in signal integration, we used our model to analyze two classes of reprogrammed N-WASP constructs engineered by Dueber et al. [28]. Curiously, the constructs corresponding to class I switches display antagonistic signal integration in experiments; the results for the application of our model to these switches are shown in figure 2.9. As seen in the figure, the simple switch's scaling behavior with tether length provides a physical mechanism for the antagonistic gating seen in these

Table 2.1: Model parameters for class I and class II complex switches. These parameters were taken from experiments performed by Dueber et al. [28] and correspond to constructs H1–H4 for class I switches and H14–H15 for class II switches.

Class I		Class II	
Parameter	Value	Parameter	Value
D	7.94 nm	D	7.94 nm
D_{SH3}	2.14 nm	D_{PDZ}	2.35 nm
L_1	21.6 nm	L_1	21.6 nm
L_2	7.2 nm	L_2	6.8 nm
L_3	6.8 nm		
K_B	0.1 μM	K_B	0.1 μM
K_W	8 μM	K_W	8 μM

constructs. Because we assume looping is the mechanism for inhibition, only binding between the white receptor/cis-ligand pair will be inhibitory. The role of the interaction between the black receptor/cis-ligand pair is to tune the length of the tether for the white receptor. As shown in the case study for the simple switch, lengthening the white receptor’s tether will increase its local concentration of cis-ligand. In the absence of free ligand, the black receptor is bound to its cis-ligand with high probability. The addition of black ligand to the solution competitively removes this interaction, allowing the tether connecting the black receptor/cis-ligand pair to be added to the length of the white receptor’s tether. As a result, changes in tether length are already a part of the natural repertoire of these signaling motifs. The effect of this change in tether length on the binding of the black receptor/cis-ligand pair is shown in figure 2.9b. When the interaction between the black receptor and its cis-ligand is removed, the effective concentration the white receptor sees of its cis-ligand slides up the curve to a higher point for the parameters corresponding to the constructs engineered by Dueber et al. [28]. The white receptor’s cis-ligand functions as an inhibitor, so removing the black receptor/cis-ligand interaction represses the switch. Because the role of free black ligand is to competitively remove that interaction, inputs through the black receptor are antagonistic.

The topology of class I switches afford tethers the opportunity to play an active role in how the two signals are processed. As seen in figure 2.5, there is a hump in the curve of effective concentration versus tether length. Where the system sits on the curve determines how class I switches will integrate the white and black signals. For the constructs engineered by Dueber et al., the parameters placed the system on the increasing part of the curve, where longer tethers lead to higher effective concentrations. This model predicts that for these constructs, lengthening the tether connecting the black receptor and its cis-ligand (L_2) will lead to stronger antagonistic integration. Finally, we note that if the system were on the decreasing part of the curve (e.g., by employing a large L_2), then the model predicts cooperative integration. In other words, both black and white ligands activate the switch.

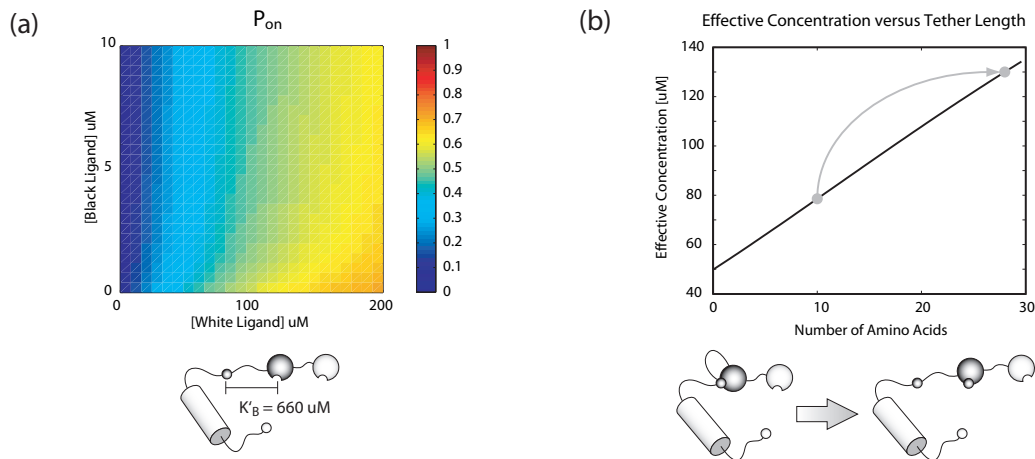


Figure 2.9: An activation profile for an antagonistic class I switch. For this profile, $K'_B = 660 \mu\text{M}$ and $K'_W = 8 \mu\text{M}$. The color represents p_{on} , with low values being blue and high values being black. When [white ligand] is increased the switch is activated, reflected by a color shift of blue to yellow or orange. When [black ligand] is increased, the switch is deactivated. This effect is best seen at high values of [white ligand]. Because the role of the black receptor/cis-ligand pair is to determine the length of white receptor's tether, the simple scaling properties of the simple switch provide a physical mechanism for antagonistic integration. The removal of SH3's interaction with its cis-ligand lengthens the white receptor's tether. The longer tether length for L_2 increases the effective concentration of cis-ligand seen by the white receptor and represses the switch.

The results for the application of our model to reprogrammed N-WASP constructs corresponding to class II switches is shown in figure 2.10. These constructs have been shown to exhibit cooperative integration in experiments, and can behave as either AND or OR gates. As seen in figure 2.10, the model displays the cooperative integration seen in experiments. Both black and white ligands activate the switch, with maximum activation achieved when both ligands are present. Mechanistically, the tether length dependence plays a less prominent role in the behavior of switches with this topology. Recall that for this topology, both binding of either black or white cis-ligand to the appropriate receptor is inhibitory. The cooperativity in the model arises because the black receptor/cis-ligand interaction controls the location of the tether origins for the white receptor and its cis-ligand. Binding of the black receptor/cis-ligand pair brings the tether origins for the white receptor/cis-ligand pair closer together, causing the white receptor to see a higher effective concentration of its cis-ligand. Binding of free black ligand not only removes an inhibitor but reduces the effective concentration the white receptor sees of its cis-ligand, making it easier for free white ligand to bind and activate the switch.

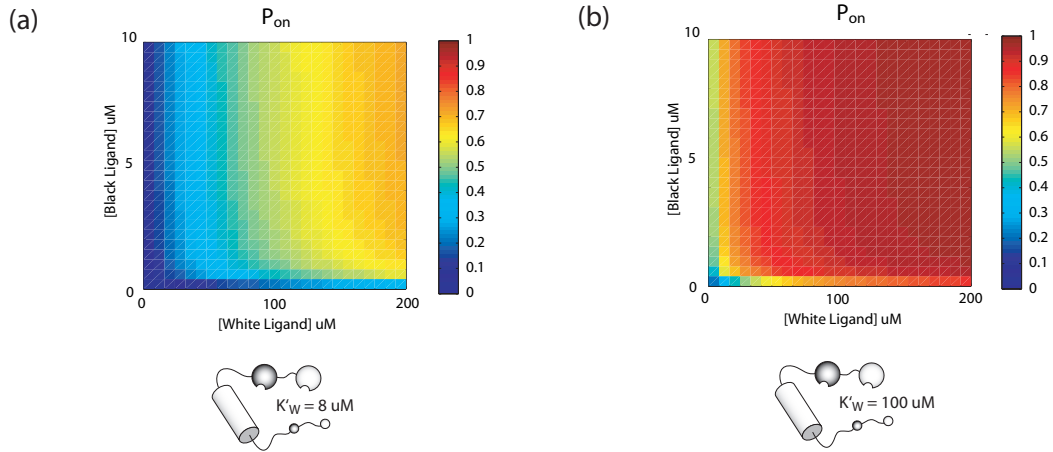


Figure 2.10: An activation profile for class II switches. The color represents p_{on} , with low values being blue and high values being black. (a) This construct exhibits the same behavior as an AND gate; both inputs are necessary to achieve maximal activation. Dissociation constants used were $K'_B = 1000 \mu\text{M}$ and $K'_W = 8 \mu\text{M}$. (b) This construct has behavior consistent with an OR gate; either input is sufficient to achieve maximum activation. Dissociation constants used were $K'_B = 1000 \mu\text{M}$ and $K'_W = 100 \mu\text{M}$.

2.4.2.4 SH3/cis-ligand interactions

The case study on reprogrammed N-WASP is useful for demonstrating how tethers can be involved in signal integration. However, during our analysis, we noted a discrepancy between theory and experiment for the interaction strength between SH3 and its cis-ligand. Experimentally, class I switches with $K'_B = 10 \mu\text{M}$ display antagonistic integration and switches with $K'_B = 660 \mu\text{M}$ only respond to inputs through the white receptor; the reverse is seen in our model. Two possible sources for this discrepancy are the SH3/cis-ligand interaction strength or the effective concentration SH3 sees of its cis-ligand. If either are reduced by two orders of magnitude, then our model qualitatively agrees with the experimental data.

Deciphering which term is the root of the discrepancy is non trivial. Concentrations are scaled by dissociation constants in the theory, making the two cases indistinguishable with the current data. Further, we found the same issue in each construct we analyzed that contained an SH3 domain. A simple switch made from an output domain and SH3 recently constructed by Dueber et al. is best described by an effective concentration of $\sim 0.2 \mu\text{M}$ for SH3's cis-ligand [63]. This is to be compared to $\sim 50 \mu\text{M}$ for the simple PDZ switch. It is difficult to explain a two-order-of-magnitude difference by a larger distance between the two tether origins, as such a distance is larger than the physical dimensions of the Arp2/3 complex. The behavior of the SH3 domain remains a mystery; one possible explanation is that the SH3/cis-ligand interaction strength is context dependent. The dissociation constants for SH3's interaction with free ligand has been measured;

the strength of this interaction may differ if the ligand is part of a larger construct [64–66].

Alternatively, the excluded volume effects alluded to in section 3 may play a significant role. The presence of large macromolecules like Arp2/3 may influence the conformations available to the tether, leading to a decrease in the effective concentration of cis-ligand. Recent work by Krishnamurthy et al. has suggested that such an effect is responsible for an order-of-magnitude decrease in the effective concentration of cis-ligand seen by human carbonic anhydrase in a tethered receptor-ligand system [11]. Whether this phenomenon could cause a two-order-of-magnitude effect is unclear. A new round of experiments, similar to the work of Krishnamurthy et al., that explores the tether length dependence of SH3/cis-ligand binding would be useful in exploring this issue [11]. While much stands to be learned when theory and experiment disagree, the resolution of this discrepancy is beyond the scope of our work. The key qualitative point to emerge from these calculations which transcends the details of our choice of parameters is the synergy between the different inputs that can exist for a single tether.

2.5 Discussion

We have used statistical mechanics and a random walk model to quantitatively explore the general theme of proteins with tethered receptor-ligand pairs. Through a case study on synthetic WASPs, we illustrated how tethers can influence protein-mediated decision making. We applied our model to constructs where multiple receptor-ligand pairs are connected by the same tether. In this case, we saw that the expected tether length dependence when the tether origins of the receptor and cis-ligand are separated provides a physical mechanism for antagonistic integration. These results demonstrate how tethers can play an active role in signal integration

There are a number of directions for future work. With regards to the case study, our assumptions about the binding of the VCA domain to Arp2/3 should be tested. The flexibility of regions assumed to be unstructured in our work can be ascertained with force-extension experiments. We have generated several predictions relevant to synthetic WASPs, namely that

- **Simple Switch:** For short tethers, an increase in tether length leads to higher effective concentrations.
- **Antagonistic Switch:** Lengthening the tether between SH3 and its cis-ligand will lead to stronger antagonistic integration.

Both of these predictions can be tested experimentally. We suggest using FRET measurements or fluorescence quenching as a more direct assay of probabilities [47, 48, 67]. Third, the role excluded volume plays in

this system, particularly with regards to the SH3 domain, should be measured with tether lengthening experiments akin to the work by Krishnamurthy et al. [11].

We end by noting that biochemistry on a leash is a thread that occurs throughout biology. Flexible tethers have been proposed to play a role in numerous processes, as mentioned above. A combination of simple physical models with experiments that vary tether lengths may lead to a clearer picture of the role tethers play in these systems.

2.6 Appendix: A kinetic model for the simple switch

In section 2.2 we developed a statistical mechanical model to compute the probability that a switch is active by invoking the assumption of thermodynamic equilibrium. Before applying this model to a real biological system, it is important to understand the limits of this assumption. The argument was built around the separation between the time scales of ligand binding and a productive subsequent activation reaction. Here, we give this argument a proper quantitative underpinning by examining a kinetic model of the simple switch.

Our kinetic model is summarized pictorially in figure 2.11. Ligands free in solution bind and unbind to the receptor with rate constants k_+ and k_- . The cis-ligand binds and unbinds to the receptor with rate constants k_c and k_- . In this model, we assume that ligand binding is diffusion limited. That is, the on rate is determined by how long it takes a ligand to come into contact with the receptor; this time is dependent on whether the ligand is free or if it is connected to a tether. The off rate is determined by the strength of the receptor-ligand interaction. Note that we treat the free ligand and cis-ligand as if they were identical and assign them the same off rate. As argued earlier, the switches are coupled to a slow productive reaction; in the case of WASPs, this is actin polymerization. This coupling is modeled by the final reaction; switches in the active state are depleted at a rate R to generate product.

It is straightforward to use chemical kinetics to convert our pictorial representation in figure 2.11 into a set of rate equations. The full linear set of kinetic equations that governs the system is

$$\frac{dS}{dt} = k_-S_L + k_-S_T - k_cS - k_+SL - RS, \quad (2.19)$$

$$\frac{dS_L}{dt} = k_+SL - k_-S_L - RS_L, \quad (2.20)$$

$$\frac{dS_T}{dt} = k_cS - k_-S_T, \quad (2.21)$$

$$\frac{dP}{dt} = RS + RS_L, \quad (2.22)$$

where S is the concentration of free switches, S_L is the concentration of switches bound to free ligand,

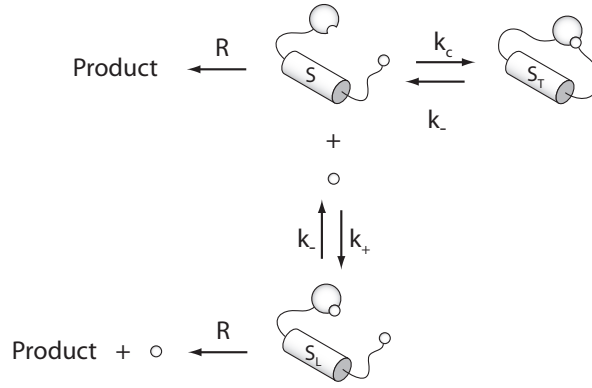


Figure 2.11: A kinetic model of the simple switch. This toy model assumes simple chemical kinetics for all reactions. Ligands free in solution bind and unbind to the receptor with rate constants k_+ and k_- while the tethered ligand binds and unbinds with rate constants k_c and k_- . Switches in an active state are consumed at a rate R to generate product.

S_T is the concentration of switches bound to tethered ligand, L is the concentration of free ligand, and P is the concentration of product. Note that we assume the concentration of free ligand is high and is essentially constant. Also note that in the above equations, we have neglected the spatial variations of the concentrations. This is a reasonable assumption if the system is initially well mixed.

The goal of our analysis is to compare a full treatment of the dynamics of the kinetic equations and determine when, if at all, an effective equilibrium approximation might hold. To this end, consider first the case $R = 0$. Starting from an initial condition of S_0 free switches, the characteristic relaxation time $\tau_{rel} \sim 1/\min[k_-, k_+L, k_c]$ governs the relaxation rate of the system towards thermal equilibrium. Once in equilibrium, the concentrations satisfy

$$S = p_S S_{tot}, \quad (2.23)$$

$$S_L = p_{S_L} S_{tot}, \quad (2.24)$$

$$S_T = p_{S_T} S_{tot}, \quad (2.25)$$

where $S_{tot} = S + S_L + S_T$ is the concentration of all switch species present in the system, and the probabilities

p_S , p_{S_L} , and p_{S_T} are given by

$$p_S = \frac{1}{1 + \frac{L}{K_d} + \frac{k_c}{k_+}}, \quad (2.26)$$

$$p_{S_L} = \frac{\frac{L}{K_d}}{1 + \frac{L}{K_d} + \frac{k_c}{k_+}}, \text{ and} \quad (2.27)$$

$$p_{S_T} = \frac{\frac{k_c}{k_+}}{1 + \frac{L}{K_d} + \frac{k_c}{k_+}}. \quad (2.28)$$

Now, when $R > 0$, we would like to know under which condition(s) we may regard the switches as being effectively in equilibrium, that is, $S_i(t) \approx p_i S_{tot}(t)$ where $i = [S, S_L, S_T]$. The approximate kinetic equations are given by

$$\frac{dS}{dt} = p_S \frac{dS_{tot}}{dt}, \quad (2.29)$$

$$\frac{dS_L}{dt} = p_{S_L} \frac{dS_{tot}}{dt}, \quad (2.30)$$

$$\frac{dS_T}{dt} = p_{S_T} \frac{dS_{tot}}{dt}, \quad (2.31)$$

$$\frac{dS_{tot}}{dt} = -R(p_S + p_{S_L}) S_{tot}, \text{ and} \quad (2.32)$$

$$\frac{dP}{dt} = R(p_S + p_{S_L}) S_{tot}. \quad (2.33)$$

That is, the relative ratios of S , S_L , and S_T remain fixed during the time evolution of the system. For brevity, we refer to the full kinetic model as FKM and the kinetic model with effective equilibrium as EKM.

Our strategy to find the limits of the equilibrium assumption is to compare the exact description with our expectation from equilibrium statistical mechanics. To do so, we numerically integrate the differential equations subject to a particular choice of initial conditions and compare the resulting concentrations of the various species. This integration requires values for the k_+ , k_- , k_c , and R . To find k_+ and k_- , we turned to the experiments of Gianni et al. and Dueber et al. [28, 46]. For PDZ domains, Gianni et al. have measured k_+ and find $k_+ \approx 8 \mu\text{M}^{-1}\text{s}^{-1}$ [46]. Their work also suggests that differences in K_d 's among different PDZ's and ligands can be attributed primarily to the off rate, k_- . As a result, we can use the measured value of k_+ in conjunction with the measured K_d to back out the value of k_- . Using the equation

$$k_- = K_d k_+, \quad (2.34)$$

we find that $k_- = 64 \text{ s}^{-1}$ to match $K_d = 8 \mu\text{M}$ measured by Dueber et al. [28].

Determining the value of k_c is more involved. Indeed, the details of the loop formation kinetics are described in the next section and here we spell out the conceptual basis for determining this rate constant. This parameter has two components. The first component is a collision rate, k_{col} , that describes how often the cis-ligand and receptor collide. The second component is a reaction probability, ρ , that describes how often binding occurs given a collision. This component is likely independent of the properties of the tether. If the two components are independent, then we can approximate the reaction rate as

$$k_c = \rho k_{col}. \quad (2.35)$$

k_{col} is calculated in the following section; for a tether length of 24.4 nm, we find that $k_{col} = 1.79 \times 10^6 \text{ s}^{-1}$. To estimate ρ , we again turn to the experiments of Gianni et al. [46]. Consider one receptor in a solution of ligands. The idea is to use the tether-free experiment to estimate parameters that we will later use to describe the kinetics of the tethered ligand. In particular, we need to determine the unknown constant ρ . Their data shows the reaction rate, k_+ , is $8 \mu\text{M}^{-1}\text{s}^{-1}$. Again, we can separate this into two components, implying

$$k_+ = \rho k_{col}^+. \quad (2.36)$$

We can compute the collision rate, k_{col}^+ for this reaction using the equation

$$k_{col}^+ = 4\pi D_L b, \quad (2.37)$$

where D_L is PDZ-ligand's diffusion coefficient in water and b is its approximate radius. This is a well known result from Smoluchowski [68]. We set $b = 1 \text{ nm}$ to estimate the size of a PDZ domain. We can estimate D_L using the Stokes-Einstein relation,

$$D_L = \frac{k_B T}{6\pi\eta r_L}, \quad (2.38)$$

where η is the viscosity of water and r_L is the radius of the ligand. We take $\eta = 10^{-3} \text{ kg m}^{-1}\text{s}^{-1}$ and $r_L = .565 \text{ nm}$ to obtain $D_L = 3.86 \times 10^{-10} \text{ m}^2\text{s}^{-1}$. These values give $k_{col}^+ = 2.92 \times 10^3 \mu\text{M}^{-1}\text{s}^{-1}$, implying that $\rho = 2.74 \times 10^{-3}$. Using this value for ρ , we estimate $k_c = 4900 \text{ s}^{-1}$.

The parameter R is the nucleation rate for new actin filaments. To estimate R , we turned to experiments performed by Zalvsky et al. [69]. In their work, they measured the creation rate for new filaments and fit their work to a four-step kinetic model. In this model Arp2/3, N-WASP, and actin come together to form a complex, which is then activated at a given rate to produce new actin filaments. This activation rate is identical to our definition of R ; for N-WASP, they found $R = 0.034 \text{ s}^{-1}$ [69]. Alternatively, the ATP

hydrolysis rate on Arp2 can serve as a surrogate for the nucleation rate. Work by Dayel et al. has shown that the VCA domain feeds Arp2/3 its first actin monomer and triggers ATP hydrolysis on Arp2 [35]. They have also demonstrated that the creation of new filaments proceeds at a rate close to the rate of ATP hydrolysis. Assuming that ATP hydrolysis and the creation of a new filament go hand-in-hand, the ATP hydrolysis rate can be a surrogate for the nucleation rate. Dayel et al. estimate that a lower bound for the ATP hydrolysis rate is $R \sim 0.1 \text{ s}^{-1}$. While the use of either value is reasonable, precision is not important. As long as R captures the appropriate time scale, then it will suffice to determine whether we can treat the system with an equilibrium model. The parameters used in the simulation are summarized in table 2.2.

Table 2.2: Numerical values for the rate constants used in the kinetic model.

Parameter	Value
k_+	$8 \mu\text{M}^{-1} \text{ s}^{-1}$
k_-	64 s^{-1}
k_c	$4.9 \times 10^3 \text{ s}^{-1}$
R	$.034 \text{ s}^{-1}$

Computing the collision rate. To compute k_{col} , we turn to analytic methods that have been previously developed [70–72]. The outline of this calculation is as follows. First, we find the probability distribution of the distance between the receptor and its cis-ligand. This distribution is then used as an effective entropic potential that governs the dynamics of the cis-ligand. Next, we treat the cis-ligand as a point particle that undergoes one-dimensional diffusion inside this potential. Its behavior is described by a one-dimensional Fokker-Planck equation. We then find the mean first passage time, which is defined as the average time for the ligand to first come within a short distance of the receptor provided it starts at the bottom of the potential well. The inverse of the mean first passage time is our estimate for k_{col} .

To compute the probability distribution for the end-to-end distance between the receptor and its cis-ligand, we return to the calculation of the effective concentration presented in the main text. The effective concentration is the probability density of the cis-ligand in the vicinity of the receptor. To compute the new quantity of interest, we want the probability density of the cis-ligand at some vector \vec{R} away from the receptor. Let receptor’s tether begin at \vec{r}_1' and end at \vec{r}_1 and let the cis-ligand’s tether begin at \vec{r}_2' and end at \vec{r}_2 . Let us continue to describe the tethers with the random walk model. The probability density of

interest is given by

$$G(\vec{R}) = \int d^3 r_1 d^3 r_2 G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1) G(\vec{r}_2 - \vec{r}_2'; \xi_T, L_2) \delta^3(\vec{R} - (\vec{r}_2 - \vec{r}_1)), \quad (2.39)$$

$$= \int d^3 x_1 G(\vec{r}_1 - \vec{r}_1'; \xi_T, L_1) G(\vec{R} + \vec{r}_1 - \vec{r}_2'; \xi_T, L_2), \quad (2.40)$$

$$= \left(\frac{3}{4\pi L_T \xi_T}\right)^{\frac{3}{2}} \exp\left[-\frac{3(\vec{R} - \vec{D})^2}{4L_T \xi_T}\right], \quad (2.41)$$

where $\vec{D} = \vec{r}_2' - \vec{r}_1'$ and $L_T = L_1 + L_2$. The probability distribution of the end-to-end distance is then

$$G(R) = R^2 \int d\Omega G(\vec{R}), \quad (2.42)$$

$$= R^2 \int_0^{2\pi} d\phi \int_{-1}^1 d(\cos\theta) \left(\frac{3}{4\pi L_T \xi_T}\right)^{\frac{3}{2}} \exp\left[-\frac{3}{4L_T \xi_T} (R^2 + D^2 - 2RD \cos\theta)\right], \quad (2.43)$$

$$= \left(\frac{3}{4\pi L_T \xi_T}\right)^{\frac{1}{2}} \frac{2R}{D} \sinh\left(\frac{3RD}{2L_T \xi_T}\right) \exp\left[-\frac{3}{4L_T \xi_T} (R^2 + D^2)\right]. \quad (2.44)$$

This distribution can be converted into an effective entropic potential using the equation

$$U_{eff}(R) = -k_B T \log G(R) + U_0, \quad (2.45)$$

where U_0 is a constant defined such that $\min U_{eff}(R) = 0$. The shape of this potential using $D = 7.94$ nm, $\xi_T = 0.4$ nm, and $L_T = 24.4$ nm is shown in figure 2.12. It is important to note that the tethered ligand will behave like it moves in this entropic potential only if we look at time scales that are long compared to the time scale of the tether's mechanical relaxation. A comparison of the contact rate for flexible amino acid chains obtained through fluorescence quenching measurements and the theory described by Szabo et al. and Hyeon et al. suggest that this is indeed the case and this approach is valid [47, 48, 70, 72].

Because we assume the ligand undergoes one-dimensional diffusion in a potential, the probability distribution of its location is governed by the Fokker-Planck equation,

$$\frac{\partial p(R, t)}{\partial t} = D_0 \frac{\partial}{\partial R} \left(\frac{\partial}{\partial R} p + \beta \frac{\partial U_{eff}}{\partial R} p \right), \quad (2.46)$$

where p is the probability distribution of the ligand and D_0 is the cis-ligand's apparent diffusion constant. Because we assume the receptor's location is fixed at the origin, D_0 is actually the sum of both the cis-ligand's

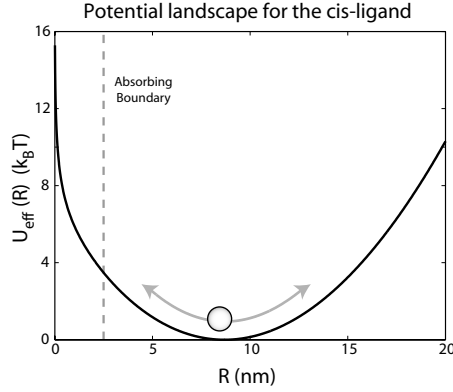


Figure 2.12: The potential landscape for the cis-ligand. The probability distribution of the distance between the receptor and its tethered ligand can be used to define an entropic potential. The cis-ligand is then assumed to undergo one-dimensional diffusion inside this potential well. We then use the Fokker-Planck equation to find the average time needed for the cis-ligand to encounter the absorbing boundary assuming it starts at the bottom of the potential well. The inverse of this mean first passage time is an estimate for the collision rate.

and the receptor's diffusion constant. The mean first passage time is given by the equation

$$\tau = \frac{1}{D_0} \int_a^r dR' \int_{R'}^{L_T} dR e^{\frac{U_{eff}(R') - U_{eff}(R)}{k_B T}}, \quad (2.47)$$

where a is the end-to-end distance of the tether at which the receptor and its cis-ligand are considered bound and r is the location of the potential's minimum [72]. We take a to be 2 nm, the approximate diameter of a PDZ domain. We take D_0 to be $6.04 \times 10^8 \text{ nm}^2\text{s}^{-1}$, the sum of D_L and the diffusion constant of a sphere of radius 1 nm in water. We computed this integral numerically using MATLAB to find $k_{col} = 1.79 \times 10^6 \text{ s}^{-1}$.

We numerically integrated the differential equations for the FKM and EKM using the MATLAB function `ode15s`. For the full model (FKM), we chose initial conditions that matched the concentrations used in the work of Dueber et al. [28]. These were $S = 50 \text{ nM}$, $S_L = 0 \text{ nM}$, $S_T = 0 \text{ nM}$, $L = 200 \text{ }\mu\text{M}$, and $P = 0 \text{ nM}$. The results are shown in figure 2.13a, where we plot the probability for the switch to be in either of its three states as a function of time. As shown in figure 2.13a, using these initial conditions results in an initial transient during which the concentrations of S_L and S_T rapidly increase. Following the initial transient, the probabilities remain constant for the duration of the simulation and are equal to the equilibrium probabilities set by equilibrium statistical mechanics. This behavior is independent of the initial conditions, as the system will display the same behavior provided the total concentration of all switch species, concentration of ligand, and rate constants remain the same.

We conclude that after the initial transients, the system enters a dynamic regime that is well characterized

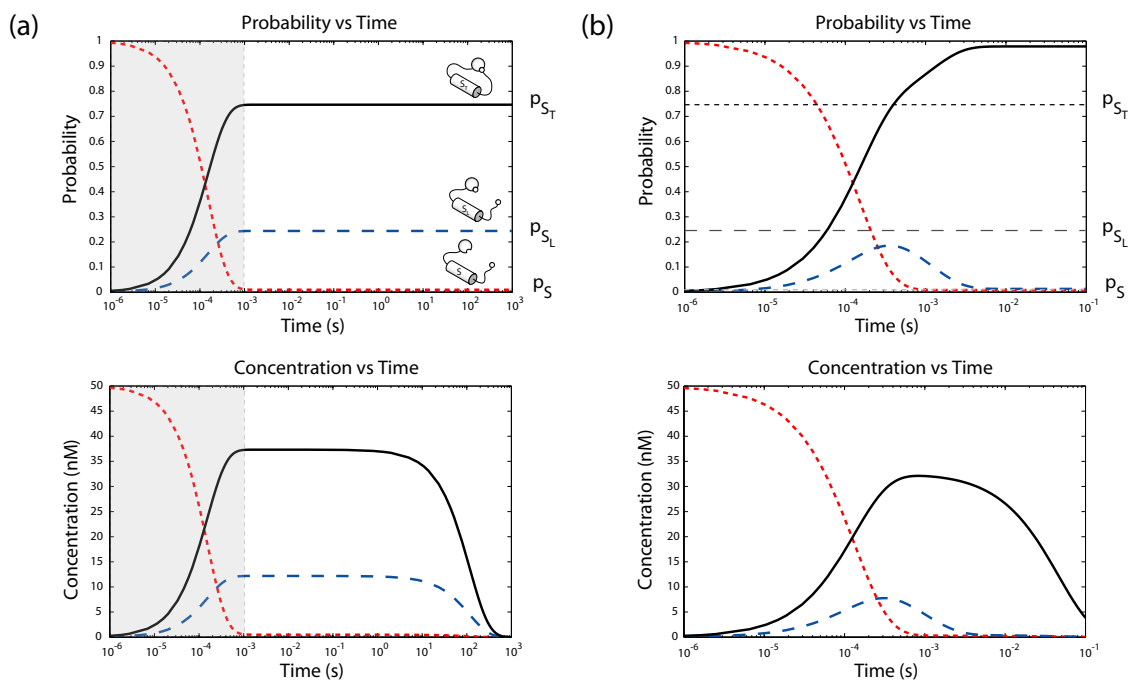


Figure 2.13: A simulation of the full kinetic model. (a) The probabilities and concentrations for S , S_L , and S_T as a function of time for $R = .034 \text{ s}^{-1}$. In the full kinetic model, the switches achieve a rapid pre-equilibria between the three states S , S_L , and S_T . After the initial transient, the probabilities for each of the three states is constant and equal to the probability predicted by equilibrium statistical mechanics. (b) The probabilities and concentrations for S , S_L , and S_T as a function of time for $R = 1000 \text{ s}^{-1}$. Because the timescale of the productive reaction is the same as that for ligand unbinding, the system never reaches an effective between the states of the simple switch. For this choice of parameters, equilibrium statistical mechanics considerations provide a poor description for the behavior of this system.

by the model invoking effective equilibrium. Hence, for our case study, the separation of timescales allows us to apply an equilibrium statistical mechanical model to a non-equilibrium system. The kinetic model also provides a vehicle to examine when this is not the case. During the initial transients, the three states are not in effective equilibrium. If actin polymerization occurred on a time scale comparable to or faster than the other rate constants, as shown in figure 2.13b, then the behavior of the system differs from the behavior based on equilibrium statistical mechanics considerations.

2.7 Appendix: Polymer models

Throughout the case study on N-WASP, we made reference to several different models for ideal chains. Each of these models has a degree of sophistication beyond the random walk model, and will be useful in exploring other examples of biochemistry on a leash. In this appendix, we briefly outline derivations for the probability distribution of the end-to-end distance vector for two such models, the wormlike chain model and the freely-jointed chain model. This appendix also adds some mathematical rigor to our use of delta functions to compute effective concentrations.

2.7.1 The wormlike chain

The wormlike chain model assigns a bending energy to each chain conformation; the probability of a chain adopting that conformation is then given by the requisite Boltzmann factor. We are interested in finding the probability distribution of the end-to-end distance vector provided that we have no constraints on the orientation of the two ends. To compute this distribution, we follow the work of Samuel et al. [51]. We express the probability distribution as an appropriate path integral and equate the distribution with the Green's function of a particular diffusion equation. We then express the Green's function as a matrix exponential and evaluate it numerically to find the appropriate distribution.

Let the parameter s be the arclength of a point on the chain; s is allowed to vary from 0 to L . The conformation of the chain is described by the unit tangent vectors $\hat{u}(s) = \langle u_x(s), u_y(s), u_z(s) \rangle$. The energy of a conformation is given by

$$E[\hat{u}(s)] = \frac{\kappa}{2} \int_0^L ds \left| \frac{\partial \hat{u}(s)}{\partial s} \right|^2, \quad (2.48)$$

where $\kappa = k_B T \xi_p$. With the energy of a conformation given, we are free to apply the canonical ensemble to calculate statistical averages. The polymer is allowed to achieve all possible conformations with each one being weighted by $e^{-\beta E}$. If the orientations of the ends are held fixed at \hat{u}_i and \hat{u}_f , then the probability

distribution of the end-to-end distance vector is

$$G\left(\vec{R}|\hat{u}_f, \hat{u}_i, L\right) = \int_{\hat{u}(0)=\hat{u}_i}^{\hat{u}(L)=\hat{u}_f} \mathcal{D}[\hat{u}(s)] \delta^{(3)}\left(\vec{R} - \int_0^L ds \hat{u}(s)\right) \exp\left[-\frac{\beta\kappa}{2} \int_0^L ds \left(\frac{\partial \hat{u}}{\partial s}\right)^2\right]. \quad (2.49)$$

To find $G\left(\vec{R}\right)$, the probability distribution with free ends, we must integrate $G\left(\vec{R}, \hat{u}_f|\hat{u}_i; L\right)$ over all possible orientations of \hat{u}_i and \hat{u}_f .

We are interested in the probability of the z-component of the end-to-end distance vector taking on a particular value. The probability distribution for z-component is given by

$$\begin{aligned} Q\left(\hat{u}_f, L; \hat{u}_i, 0|r_z\right) &= \int d^3R \delta^{(3)}\left(R_z - r_z\right) G\left(\vec{R}, \hat{u}_f|\hat{u}_i; L\right), \\ &= \int_{\hat{u}(0)=\hat{u}_i}^{\hat{u}(L)=\hat{u}_f} \mathcal{D}[\hat{u}(s)] \delta^{(3)}\left(\int_0^L ds u_z(s) - z\right) \exp\left[-\frac{\xi_p}{2} \int_0^L ds \left(\frac{\partial \hat{u}}{\partial s}\right)^2\right]. \end{aligned} \quad (2.50)$$

It is convenient to continue the calculation in Fourier space. The Fourier transform of Q is given by

$$\begin{aligned} \tilde{Q}\left(\hat{u}_f, L; \hat{u}_i, 0|k\right) &= \int_{-\infty}^{\infty} dr_z e^{ikr_z} Q\left(r_z, \hat{u}_f|\hat{u}_i\right), \\ &= \int_{\hat{u}(0)=\hat{u}_i}^{\hat{u}(L)=\hat{u}_f} \mathcal{D}[\hat{u}(s)] \exp\left[-\xi_p \int_0^L ds \left(\frac{1}{2} \left(\frac{\partial \hat{u}}{\partial s}\right)^2 - \frac{ik}{\xi_p} u_z(s)\right)\right]. \end{aligned} \quad (2.51)$$

This path integral is the green's function for the Schrodinger equation

$$\left(-\frac{1}{2\xi_p} \nabla_{\hat{u}}^2 + ik u_z + \frac{\partial}{\partial L}\right) I\left(\hat{u}_f, L; \hat{u}_i, 0|k\right) = 0, \quad (2.52)$$

where $\nabla_{\hat{u}}^2$ is the Laplacian on a unit sphere. Note that we are confined to the unit sphere because \hat{u} has unit magnitude. We move to spherical coordinates, where θ is the angle \hat{u}_f makes with the z-axis and ϕ is the angle the projection of \hat{u}_f in the x-y plane makes with the x-axis. Defining $z = \cos \theta$, this equation becomes

$$\left(-\frac{1}{2\xi_p} \nabla_{\hat{u}}^2 + ikz + \frac{\partial}{\partial L}\right) I\left(\hat{u}_f, L; \hat{u}_i, 0|k\right) = 0, \quad (2.53)$$

where

$$\nabla_{\hat{u}}^2 = (1 - z^2) \frac{\partial^2}{\partial z^2} - 2z \frac{\partial}{\partial z} + \frac{1}{1 - z^2} \frac{\partial^2}{\partial \phi^2}. \quad (2.54)$$

Because we are interested in the case of free ends, we will eventually integrate over the orientations of \hat{u}_i . The resulting function will have no ϕ dependence because ϕ dependence can only enter through \hat{u}_i . Hence,

we can restrict ourselves to studying the equation

$$\left(-\frac{1}{2}(1-z^2)\frac{\partial^2}{\partial z^2} + z\frac{\partial}{\partial z} + ikz + \frac{\partial}{\partial L}\right)I(k, \hat{u}_f; L) = 0. \quad (2.55)$$

We can find the Green's function by using separation of variables on z and L and then using an eigenfunction expansion. If the normalized eigenfunctions for the operator $\mathcal{L} = -\frac{1}{2}(1-z^2)\frac{\partial^2}{\partial z^2} + z\frac{\partial}{\partial z} + ikz$ are ψ_n with eigenvalues λ_n , then the Green's function can be written as

$$\tilde{Q}(z, L; z', 0|k) = \sum_{n=0}^{\infty} \psi_n(z)\psi_n(z')e^{\lambda_n L}. \quad (2.56)$$

Using the bra and ket notation, we can rewrite this as

$$\tilde{Q}(z, L; z', 0|k) = \sum_{n=0}^{\infty} |\lambda_n\rangle \langle \lambda_n| e^{\lambda_n L}. \quad (2.57)$$

In this notation, we see that \tilde{Q} is simply the matrix exponential of \mathcal{L} . This matrix exponential is easily evaluated when we use the eigenfunctions of \mathcal{L} as a basis because \mathcal{L} is a diagonal matrix in this basis. However, finding the eigenfunctions for \mathcal{L} is difficult. This problem can be approached numerically when we use the eigenfunctions of the operator $\mathcal{L}_0 = -\frac{1}{2}(1-z^2)\frac{\partial^2}{\partial z^2} + z\frac{\partial}{\partial z}$ as a basis. We denote this as the \mathcal{L}_0 . The eigenfunctions of \mathcal{L}_0 are the normalized Legendre polynomials

$$P_l(z) = \sqrt{\frac{2l+1}{2}} (2^l l!)^{-1} \frac{\partial^l}{\partial z^l} (z^2 - 1)^l. \quad (2.58)$$

To find the matrix for the matrix exponential of \mathcal{L} in this basis, we must first find the matrix for \mathcal{L} in the \mathcal{L}_0 basis. Using the following equations

$$\int_{-1}^1 dz P_n(z) \mathcal{L}_0 P_m(z) = \delta_{mn} \frac{n(n+1)}{2}, \quad (2.59)$$

$$\int_{-1}^1 dz P_n(z) ikz P_m(z) = \delta_{m,n+1} ik \frac{n+1}{\sqrt{(2n+1)(2n+3)}}, \quad (2.60)$$

we see that \mathcal{L} is a symmetric tri-diagonal matrix with matrix elements

$$\mathcal{L}_{n,n} = \frac{n(n+1)}{2}, \quad (2.61)$$

$$\mathcal{L}_{n,n+1} = \mathcal{L}_{n+1,n} = ik \frac{n+1}{\sqrt{(2n+1)(2n+3)}} \quad (2.62)$$

where n goes from 0 to ∞ . We can then obtain the matrix for $\tilde{Q}(z, L; z', 0|k)$ in the \mathcal{L}_0 basis by numerically

evaluating the matrix exponential in MATLAB. While the matrix of \mathcal{L} is actually infinite, we found that using the first 30X30 submatrix of \mathcal{L} was sufficient to accurately find the matrix exponential.

We are interested in the case of free ends, so we can integrate over z and z' , which represent the orientations of the final and initial tangent vectors. Because only the first Legendre polynomial is symmetric, we only need to consider the upper rightmost entry of the matrix for \tilde{Q} . In other words,

$$\tilde{Q}(z; L) = \int_{-1}^1 dz \int_{-1}^1 dz' \tilde{Q}(z, L; z', 0|k) = [e^{\mathcal{L}}]_{0,0} \quad (2.63)$$

where $e^{\mathcal{L}}$ is expressed in the \mathcal{L}_0 basis.

With a method to numerically evaluate $\tilde{Q}(k; L)$, we can find $Q(r_z; L)$ by performing a numerical integral using the quad function in MATLAB to obtain the inverse Fourier transform,

$$Q(r_z; L) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{-ikr_z} \tilde{Q}(k; L). \quad (2.64)$$

We can find the probability distribution for the end-to-end distance vector by noting that it is related to the probability distribution of the z-component of the end-to-end distance vector through the expression

$$Q(r_z; L) = \int_0^{\infty} dr_{xy} 2\pi r_{xy} G\left(\sqrt{r_{xy}^2 + r_z^2}\right), \quad (2.65)$$

$$= 2\pi \int_{r_z}^{\infty} dR R G(R), \quad (2.66)$$

where $R = \sqrt{r_{xy}^2 + r_z^2}$. Defining $S(R) = 4\pi R^2 G(R)$ as the probability distribution for the end-to-end distance, we get the expression

$$Q(r_z; L) = - \int_{\infty}^{r_z} dR \frac{1}{2R} S(R). \quad (2.67)$$

From this it follows that

$$S(R) = -2R \frac{\partial Q(R; L)}{\partial R}, \quad (2.68)$$

and hence

$$G(R) = -\frac{1}{2\pi R} \frac{\partial Q(R; L)}{\partial R}. \quad (2.69)$$

With the ability to evaluate Q numerically, we can use a four-point central difference approximation to perform the numerical derivative and obtain the probability distribution for the end-to-end distance vector.

2.7.2 The freely-jointed chain

Let us suppose we have a polymer composed of N segments each of length b_i , with each segment allowed to rotate freely in 3 dimensions. If \vec{r}_i is a vector of length b_i with the same orientation of the i 'th segment, then the end-to-end distance vector is given by

$$\vec{R} = \sum_{i=1}^N \vec{r}_i. \quad (2.70)$$

To calculate the probability distribution of \vec{R} , let us first write out the distribution for one segment. For the i 'th segment, its probability distribution is

$$G_i(\vec{r}_i) = \frac{1}{4\pi b_i^2} \delta^2(|\vec{r}_i| - b). \quad (2.71)$$

The probability distribution of \vec{R} is then

$$G_N(\vec{R}) = \prod_{i=1}^N \left[\int d^3 r_i G_i(\vec{r}_i) \right] \delta^3 \left(\vec{R} - \sum_{j=1}^N \vec{r}_j \right). \quad (2.72)$$

Using the Fourier representation of the delta function,

$$\delta^3 \left(\vec{R} - \sum_{j=1}^N \vec{r}_j \right) = \int \frac{d^3 k}{(2\pi)^3} \exp \left[i\vec{k} \cdot \vec{R} - i\vec{k} \cdot \sum_{j=1}^N \vec{r}_j \right], \quad (2.73)$$

we can write $G_N(\vec{R})$ as

$$G_N(\vec{R}) = \int \frac{d^3 k}{(2\pi)^3} e^{-i\vec{k} \cdot \vec{R}} \prod_{m=1}^N \left[\int d^3 r_m \frac{1}{4\pi b_m^2} \delta^3(|\vec{r}_m| - b_m) e^{-i\vec{k} \cdot \vec{r}_m} \right] \quad (2.74)$$

We can simplify this by performing the following integrals

$$\begin{aligned} \int d^3 r_m \frac{1}{4\pi b_m^2} \delta^3(|\vec{r}_m| - b_m) e^{-i\vec{k} \cdot \vec{r}_m} &= \frac{1}{2} \int_{-1}^1 d(\cos \theta) e^{-ikb_m \cos \theta}, \\ &= \frac{1}{2} \left[\frac{e^{-ikb_m \cos \theta}}{ikb_m} \right]_{-1}^1, \\ &= \frac{\sin(kb_m)}{kb_m} \end{aligned}$$

and

$$\begin{aligned}\int d\Omega e^{i\vec{k}\cdot\vec{R}} &= 2\pi \int_{-1}^1 d(\cos\theta) e^{ikR \cos\theta}, \\ &= \frac{4\pi \sin(kR)}{kR}.\end{aligned}$$

These two expressions combined yield

$$G_N(\vec{R}) = \frac{1}{2\pi^2 R} \int_0^\infty dk k \sin(kR) \prod_{m=1}^N \left(\frac{\sin(kb_m)}{kb_m} \right). \quad (2.75)$$

At this point, we can compute the integral numerically. If we were interested in the case where all the segments are of the same size, then we can obtain an analytic solution using a contour integral. Making the substitution $k = \frac{\eta}{b}$, this becomes

$$G_N(\vec{R}) = \frac{1}{2\pi^2 R b^2} \int_0^\infty d\eta \eta \sin\left(\frac{\eta R}{b}\right) \left(\frac{\sin\eta}{\eta}\right)^N. \quad (2.76)$$

Simplification yields

$$\begin{aligned}G_N(\vec{R}) &= \frac{1}{2\pi^2 R b^2} \int_0^\infty d\eta \eta \left(\frac{e^{\frac{i\eta R}{b}} - e^{-\frac{i\eta R}{b}}}{2i} \right) \left(\frac{\sin\eta}{\eta}\right)^N, \\ &= \frac{i}{4\pi^2 R b^2} \int_{-\infty}^\infty d\eta \eta e^{-\frac{i\eta R}{b}} \left(\frac{\sin\eta}{\eta}\right)^N, \\ &= \frac{i}{4\pi^2 R b^2} \int_{-\infty}^\infty d\eta \eta e^{-\frac{i\eta R}{b}} \frac{(e^{i\eta} - e^{-i\eta})^N}{(2i\eta)^N}, \\ &= \frac{1}{2^{N+2} i^{N-1} \pi^2 b^2 R} \sum_{j=0}^N (-1)^j \binom{N}{j} \int_{-\infty}^\infty d\eta \frac{\exp[i\eta(N-2j-\frac{R}{b})]}{\eta^{N-1}}.\end{aligned}$$

We need the integral $I_N(x) = \int_{-\infty}^\infty d\eta \frac{e^{i\eta x}}{\eta^{N-1}}$. There is a singularity at $\eta = 0$, but we can regularize it by letting $\eta \rightarrow \eta - i\epsilon$. This procedure gives

$$I_N(x) = \int_{-\infty}^\infty d\eta \frac{e^{ix(\eta-i\epsilon)}}{(\eta-i\epsilon)^{N-1}}. \quad (2.77)$$

This integral can be done by closing the contour with an infinite semicircle, provided the argument of the integral vanishes as the radius of the semicircle goes to infinity. Hence, if $x < 0$ we close the contour on the bottom and if $x > 0$ we close it on the top. There are no poles in the bottom semicircle, so there is no

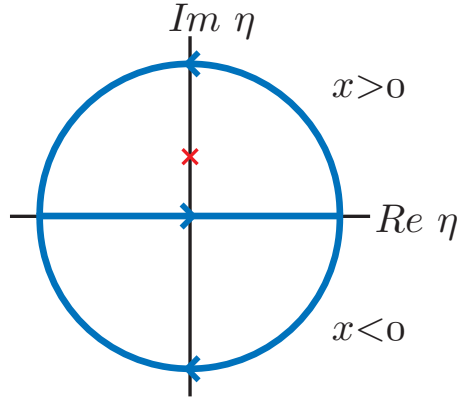


Figure 2.14: Contour integral for $I_N(x)$.

contribution. For the top semicircle, the residue theorem gives

$$\begin{aligned}
 I_N(x) &= 2\pi i \text{Res} \left[\frac{e^{ix(\eta-i\epsilon)}}{(\eta-i\epsilon)^{N-1}} \right], \\
 &= 2\pi i \text{Res} \left[\sum_{j=0}^{\infty} \frac{(ix)^j (\eta-i\epsilon)^{j-N+1}}{j!} \right], \\
 &= \frac{2\pi i^{N-1} x^{N-2}}{(N-2)!}.
 \end{aligned} \tag{2.78}$$

Our contour requires $x \geq 0$, implying $N - 2j - \frac{R}{b} \geq 0$. This gives the analytic form for $G_N(\vec{R})$,

$$G_N(\vec{R}) = \frac{1}{2^{N+1} \pi b^2 R} \sum_{0 \leq j \leq \frac{N - \frac{R}{b}}{2}} (-1)^j \binom{N}{j} \left(N - 2j - \frac{R}{b} \right)^{N-2}. \tag{2.79}$$

A cursory inspection of figure 2.15 reveals the behavior of $G_N(\vec{R})$. First, as N grows larger, the probability distribution can be approximated by a Gaussian. This is a consequence of the central limit theorem as demonstrated in the following section. Second, the Gaussian approximation overestimates the probability density when the magnitude of \vec{R} is small. Third, the derivation of the Gaussian approximation always invokes the separability of $G_N(\vec{R})$, i.e. $G_N(\vec{R}) = G_N(x)G_N(y)G_N(z)$. Staring at the analytic form of $G_N(\vec{R})$ for the freely jointed chain shows that separability is not true for small N . This is not surprising because the finite segment length imposes dependence among the x , y , and z components of \vec{R} . For large N , the constraint only becomes relevant when the chain is near maximum extension, an event that occurs with small probability [73].

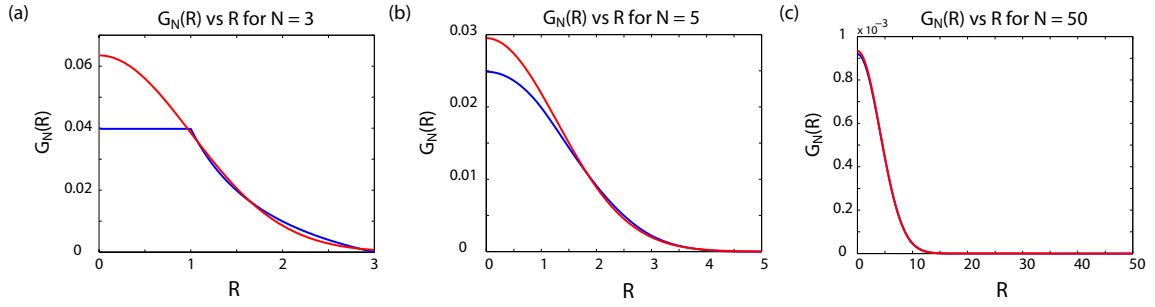


Figure 2.15: $G_N(\vec{R})$ for varying N . The blue curves are the analytic result and the red are a gaussian approximation. As N grows larger, the probability distribution approaches a gaussian.

2.7.3 Computing probability distributions with delta functions

In the case study on N-WASP, we explored the toy problem of a receptor and a ligand living on the end of two distinct tethers. We were interested in the notion of an effective concentration — the probability density of the ligand in the vicinity of the receptor. Our calculation went as follows. Suppose the probability density for the receptor being at location x_1 was $p_1(\vec{x}_1; L_1)$ and the probability density for the ligand being at location x_2 was $p_2(\vec{x}_2; L_2)$. Because the two random variables (the location of the receptor and ligand) are independent, the joint distribution is simply $p_1(\vec{x}_1; L_1)p_2(\vec{x}_2; L_2)$. We computed the effective concentration by integrating over all locations of both the receptor and ligand and employing a delta function that fires when the two are in the same location. In other words,

$$J = \int \int p_1(\vec{x}_1; L_1)p_2(\vec{x}_2; L_2)\delta(\vec{x}_1 - \vec{x}_2)d^3\vec{x}_1d^3\vec{x}_2 \quad (2.80)$$

where J is the effective concentration. In this appendix, we provide a brief derivation demonstrating why this method of calculation is sound.

Let us turn to the more general topic of how to deal with a change in variables when dealing with random variables. Let us suppose we have independent random variables X_1, \dots, X_n with each X_i having the probability distribution function $f_{X_i}(x_i)$. Now let us suppose that we have another random variable Y defined by

$$Y = G(X_1, \dots, X_n), \quad (2.81)$$

where G is a reasonably well-behaved function (continuous, bounded, etc.). The question we ask is: What

is the probability distribution for Y ? I claim that it is given by

$$f_Y(y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) \cdots f_{X_n}(x_n) \delta(y - G(x_1, \dots, x_n)) dx_1 \cdots dx_n. \quad (2.82)$$

To demonstrate this, let us use indicator functions. An indicator function $\Theta_A(a)$ is 1 when the input a is within the set A and 0 otherwise. Let us define a useful indicator function

$$\begin{aligned} \Theta(x, \epsilon) &= 1 \text{ if } -\epsilon < x < \epsilon, \\ &= 0 \text{ otherwise.} \end{aligned}$$

With this function in hand, let us compute the probability of the random variable Y being within ϵ of the value y . This is given by

$$P(Y \in [y - \epsilon, y + \epsilon]) = \int \cdots \int_{\Omega} f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1 \cdots dx_n, \quad (2.83)$$

where the domain Ω is restricted to $G^{-1}([y - \epsilon, y + \epsilon])$. This can be rewritten using the indicator function as

$$P(Y \in [y - \epsilon, y + \epsilon]) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) \cdots f_{X_n}(x_n) \Theta(y - G(x_1, \dots, x_n), \epsilon) dx_1 \cdots dx_n, \quad (2.84)$$

Now, we ask the question: What is the probability density of the random variable Y at position y ? We can approximate this density by taking the probability of Y being within the window $[y - \epsilon, y + \epsilon]$ and dividing by the width of the window, 2ϵ . In other words,

$$f_Y(y) \approx \frac{1}{2\epsilon} P(Y \in [y - \epsilon, y + \epsilon]), \quad (2.85)$$

$$\approx \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) \cdots f_{X_n}(x_n) \frac{1}{2\epsilon} \Theta(y - G(x_1, \dots, x_n), \epsilon) dx_1 \cdots dx_n. \quad (2.86)$$

To make the estimate more precise, we take the limit as ϵ goes to zero. This gives us

$$\begin{aligned} f_Y(y) &= \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) \cdots f_{X_n}(x_n) \frac{1}{2\epsilon} \Theta(y - G(x_1, \dots, x_n), \epsilon) dx_1 \cdots dx_n, \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) \cdots f_{X_n}(x_n) \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \Theta(y - G(x_1, \dots, x_n), \epsilon) dx_1 \cdots dx_n, \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) \cdots f_{X_n}(x_n) \delta(y - G(x_1, \dots, x_n)) dx_1 \cdots dx_n, \end{aligned}$$

where we have made use of the definition of the delta function during the final manipulation. While bringing

the limit inside the integral could use some closer scrutiny by a mathematician, the reasoning remains sound and can be extended to vector-valued functions of X_1, \dots, X_n as well. Returning to the case of the tethered ligand and receptor, we had the probability density functions $p_1(\vec{x}_1; L_1)$ and $p_2(\vec{x}_2; L_2)$. The function we are now interested in is $\vec{y} = \vec{x}_1 - \vec{x}_2$, the distance between the two objects. By the preceding argument, its probability distribution is given by

$$f_Y(y) = \int \int p_1(\vec{x}_1; L_1) p_2(\vec{x}_2; L_2) \delta(\vec{y} - (\vec{x}_1 - \vec{x}_2)) d^3 x_1 d^3 x_2. \quad (2.87)$$

The effective concentration is when \vec{x}_1 and \vec{x}_2 are in the same location — i.e., $\vec{y} = \vec{0}$. So we have

$$J = \int \int p_1(\vec{x}_1; L_1) p_2(\vec{x}_2; L_2) \delta(\vec{x}_1 - \vec{x}_2) d^3 \vec{x}_1 d^3 \vec{x}_2. \quad (2.88)$$

2.8 Acknowledgements

We thank Lacramioara Bintu, John Dueber, Paul Grayson, Alexander Grosberg, Jané Kondev, Wendell Lim, Tharathorn Rimchala, Boo Tseng, Paul Wiggins, and Brian Yeh for helpful discussions. David Van Valen and Rob Phillips acknowledge the support of a Fannie and John Hertz Foundation Yaser Abu-Mostafa Fellowship, a UCLA/Caltech MSTP Fellowship, and an NIH Pioneer Award. Mikko Haataja acknowledges support from the National Science Foundation through NSF-DMR Grant No. 0449184.

Chapter 3

Ion-dependent dynamics of *in vitro* DNA ejection for bacteriophage lambda

3.1 Introduction

Bacteriophages have played a key role in the emergence of both molecular and physical biology. They were an essential component of the Hershey-Chase experiment [74], which established that DNA was the molecule of inheritance, and have since provided an important technology for cloning and protein expression [75]. Phages have also been important model organisms for the study of macromolecular self-assembly; recent work has demonstrated assembly of viruses from their constituent components *in vitro* [76]. Bacteriophages have also been instrumental in the field of systems biology, as evidenced by the dissection of the transcriptional program responsible for the decision between lysis and lysogeny [77–81]. Finally, bacteriophages have played an important role in the field of single-molecule biophysics, as evidenced by the single-phage packaging experiments which revealed the large (many pico-Newton) forces generated by motors that package phage DNA into protein capsids [2]. The study of the bacteriophage life cycle has yielded much insight in the physical biology setting and we anticipate that it will continue to serve as a useful model system.

Here, we highlight four recent experimental techniques that examine how phage DNA is packaged into and released from protein capsids. These four techniques include single-molecule packaging of DNA, single-phage ejection measurements, osmotic suppression of DNA ejection, and cryo-electron tomography of partially packed phages. Single-molecule DNA packaging experiments have sought to elucidate the mechanics of how DNA gets packaged into the capsid; optical tweezers have been used to make systematic measurements of the packaging force for phage lambda and phi29 in a variety of different salt conditions [82]. What these experiments demonstrated was that by increasing the charge of the salts, the genome is packaged at a higher

velocity. Additionally, the internal force of the capsid DNA that resists packaging increases monotonically with a reduction in the charge of the salts.

These packaging experiments are complemented by single-phage ejection experiments, wherein phages are coerced into ejecting their DNA by addition of a trigger protein and then observed under a microscope with the aid of flow and DNA staining dyes [83,84]. These experiments illustrated the dynamics of the ejection process and have provided some insight into the frictional forces experienced by the DNA as it exits the capsid [84–86]. The principal quantity measured in these experiments is the velocity of ejection as a function of the quantity of DNA inside the capsid. The observations garnered from these ejection experiments reveal (broadly speaking) a picture in which the forces responsible for the ejection process decay as a function of packaged length in the same way as the forces that build up during the packaging process depend upon the packaged length: a reduction in charge increases the speed of ejections. However, the available ejection data only reflect measurements on a few ionic conditions.

It is also possible to measure the force exerted by the DNA inside phage capsids during ejections by inducing phages to eject their DNA into a series of increasingly osmotically-resistive solutions [87]. By varying the concentration of an osmolyte like polyethylene glycol (PEG), the resistive pressure is also varied. One can therefore measure the amount of DNA inside the phage capsid as a function of external pressure, and therefore how much force is generated by the DNA inside the capsid as a function of genome length [88–90]. It has been proposed that the DNA self-repulsion force can help push the DNA into the bacterial cytoplasm [91], but a consensus has yet to be reached. Alternative models propose that highly osmotic environments serve as a condensing agent for DNA exiting phage and actually promote ejection [90,92]. Others posit that the osmotic difference [93] between the intra- and extracellular environment is sufficient to "flush" DNA from the phage into the host [94]. However, when certain experimental conditions are met, by increasing the external pressure of the solution, phage ejections are inhibited [92]. To our knowledge, a systematic study of how different salt conditions influence ejections at different resistive pressures and velocities has yet to be undertaken.

Another way to study ejection and packaging is by looking at the structure of the DNA itself inside the capsid. Cryo-electron tomography experiments in which the packaging process of phage phi29 was paused by freezing have provided information on the structure of viral DNA as a function of length of DNA inside the capsid [95]. Others have investigated T5 [96,97], lambda [98,99], T3 [100], and P22 (see [101] and references therein). Recently, the interaxial spacing of condensed DNA was measured as a function of salts in order to corroborate viral DNA packaging theory [102].

The four experimental techniques described above have shed light on the forces, dynamics, and structure of DNA entering and leaving the viral capsid, and have been complemented by a vigorous modeling effort.

These include simple structural models taking into account bending energies and *in vitro* measurements of electrostatic repulsion [91, 103–106], density functional theory with few structural constraints that takes into account electrostatic repulsion/correlations and van der Waals forces [107], and molecular mechanics [95, 108, 109]. This body of work deepens our physical understanding of the packaging and ejection process.

In this work, we focus on the measurement of velocity and force of DNA ejection in bacteriophage lambda as a function of charge at fixed ionic strength. In previous experiments, it was shown that ejections are slower in the presence of magnesium as opposed to sodium [85]; however, these experiments were not performed at fixed ionic strength, making it difficult to discern the role of other ion specific factors, such as charge. Our results demonstrate that the ionic composition controls the observed velocity and force, rather than ionic strength. We thus propose that the charge of the counter-ion species is an important control parameter for this system. We also examined "looped" ejections, a type of ejection that was not analyzed in previous experiments and was assumed to be an artifact [85]. In these ejections, the first portion of the DNA exiting the capsid is stuck at the site of origin, leading to a very bright piece of DNA. This attachment is eventually broken by the flow, fully unfolding the DNA. We demonstrate that looped ejections exhibit the same dynamics as unlooped ejections and that they are another ejection class, and not an artifact. We confirm this using optical traps. Finally, we provide technical improvements on the single-phage ejection assay, in particular showing that low dye concentrations are necessary to avoid spurious (non-LamB induced) ejections.

3.2 Experimental design

Our study of *in vitro* DNA ejection in phage lambda centers around two assays — one that measures the kinematics of ejection from single phages and another that measures the force driving DNA ejection in bulk. Both of these methods have been used previously in the literature; herein we briefly describe conceptually how each assay works.

To observe the kinematics of single phages ejecting their genomes into real solution, we use an assay first pioneered by Mangenot et al. for phage T5, and later adapted to phage lambda by Grayson et al. [84, 85]. A schematic of the experiment is shown in figure 3.1a. A coverslip is passively coated with lambda phage by incubating a flow chamber with a solution of viruses; the viruses non-specifically adhere to the surface. Next, a solution is flown into the chamber that has three components — the protein LamB, the DNA staining dye SYBR gold, and the glucose oxidase/catalase oxygen scavenging system. The presence of the shigella variant of the protein LamB, the receptor for phage lambda *in vivo*, serves to initiate the DNA ejection process. Once the DNA starts to leave the capsid, the DNA staining dye SYBR gold is able to bind to the

DNA, enabling visualization by fluorescence microscopy. The presence of the glucose oxidase/catalase oxygen scavenging system prevents photodamage caused by oxygen radicals created by the constant excitation and emission of the dyes. Lastly, the entire experiment is done in the presence of a small shear flow. This shear flow stretches out the DNA, enabling us to quantify how much DNA has ejected at each time point.

To measure the force of DNA ejection, we use an assay first developed by Evilevitch et al. [87]. In this assay, lambda phages are incubated in a solution of LamB, DNase, and polyethylene glycol. The presence of polyethylene glycol serves as a crowding agent. The PEG molecules desire to occupy as much volume as possible and the ejection of DNA into solution is volume the PEG molecules can't occupy. This leads to a force that resists DNA ejection, and its magnitude is well calibrated [110]. In our modified protocol, we let the reaction reach equilibrium and then purify the DNA remaining in the viral capsids using proteinase K treatment to degrade both the DNase and the viral capsid. The DNA is then concentrated and its length measured using a home-built pulsed-field gel electrophoresis apparatus.

3.3 Materials and methods

Phage Purification. Lambda phage strain λ CI60 was purified according to Sambrook [111]. Briefly, stock phages were amplified by one round of plate lysis and one round of liquid phase lysis, followed by precipitation in PEG. Samples were then purified on a CsCl gradient followed by isopycnic centrifugation. Phages were dialyzed into the appropriate salt condition before use.

LamB purification. The membrane protein LamB was purified using a modified version of a protocol of Keller et al. [112]. The E. coli strain pop154 was grown overnight in LB + 0.2% maltose at 37 °C. The cells were pelleted at 3000 g at 4 °C for 30 minutes and re-suspended in 50 mM sodium phosphate pH 7.5, 100 mM NaCl, 2 mM EDTA, 5% sucrose. The re-suspended cells were subsequently lysed in a French press. The lysate was again centrifuged at 5000 g at room temperature for 10 minutes. The supernatant was collected and the outer membrane fraction was pelleted by ultracentrifugation at 30000 rpm in a Beckman Ti-60 rotor, 18 °C for 40 minutes. The pellet containing the membranes was resuspended in 20 mM sodium phosphate, pH 7.5, and 0.5% octylpolyoxyethylene (oPOE) in order to extract nonspecific membrane proteins. The solution was incubated in a heat bath at 40 °C for 50 minutes and again pelleted by ultracentrifugation. LamB was extracted by re-suspending the pellet in 20 mM sodium phosphate, pH 7.5, and 3% oPOE, incubating at 37 °C for 40 minutes, and pelleting the resulting mixture by ultracentrifugation. The supernatant was dialyzed to 20 mM sodium phosphate, pH 7.5, and 1% oPOE and then loaded onto a GE MBPTrap column and eluted with 20 mM sodium phosphate, pH 7.5, 1% oPOE, and 20 mM maltose after washing twice with the same solution without maltose. The LamB solution was then dialyzed with 50 mM Tris, pH 7.5, 10

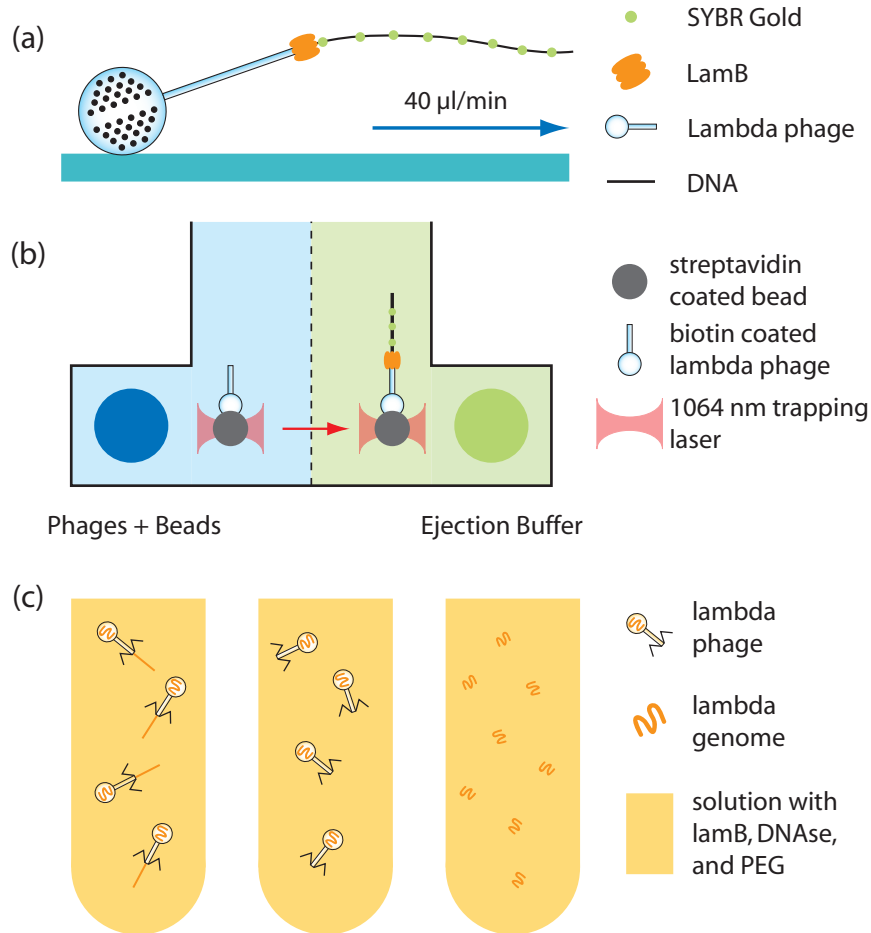


Figure 3.1: Schematics for observing DNA ejection by single-phage lambda virions *in vitro*. DNA was visualized using fluorescence microscopy. (a) Ejection at a surface. A flow chamber was incubated with lambda phage and the viruses were allowed to nonspecifically settle onto the surface. DNA ejection was triggered by flowing in LamB; the ejected DNA was visualized using SYBR gold. The flow field extends the ejected DNA, allowing quantification of the amount of DNA ejected by either measuring the length in pixels or measuring the total intensity. (b) Ejection on a bead. A parallel ejection assay using an optical trap was developed. Biotinylated phages were bound to streptavidin-coated polystyrene beads. A dual-input flow chamber was used; beads coated with lambda phage entered the chamber through one input and an ejection buffer consisting of LamB and SYBR gold entered through the other. In the presence of flow, the two fluids are well separated. An optical trap was used to move a phage-coated bead into the ejection buffer to initiate DNA ejection. (c) Measuring the force driving DNA ejection using an osmotic suppression assay. Phages are incubated in a solution containing LamB, DNase, and polyethylene glycol. The presence of PEG acts as a crowding agent and provides a resistive force against DNA ejection. Once the reaction has come to equilibrium, the DNA inside the capsid is purified and its length is quantified using pulsed field gel electrophoresis.

mM MgSO₄ (TM), and 1% oPOE. The resulting solution was diluted to 2 mg/mL. The concentration was assayed by measuring the absorbance at 280 nm and the purity was verified by SDS-PAGE.

Single-phage ejection assay. See figure 3.1a. We follow a protocol developed by Mangenot et al. [84,85]. Microscope coverslips were cleaned by sonication in 1M KOH for 10 minutes followed by sonication in water for 10 minutes and dried on a hot plate. Glass slides were drilled using a diamond covered drill bit and 5 inches of tubing was attached to the glass slide using epoxy. The flow chamber was assembled using laser-cut double-sided adhesive tape (41). A solution of $10^8 - 10^{11}$ pfu/ml lambda phage and 20 nm fluorescent polystyrene beads (focusing aids) was incubated in the assembled flow chambers for 10 minutes. Once focused, the chamber was washed with 200 L of buffer + 1% oPOE. Buffers had the following ionic compositions: 0, 2.5, 5, 7.5, or 10 mM NaCl, and 2.5, 1.875, 1.25, 0.625, or 0 mM MgSO₄, respectively, in addition to 10 mM Tris, pH 7.5. The solution to induce ejection consisted of buffer, 1% oPOE, 1% glucose oxidase/catalase, 10^{-6} diluted SYBR gold, and 1% LamB. Occasionally, 20 μ g/mL acetylated bovine serum albumin, 3 mg/mL casein, and 80 μ g/mL heparin were added to the buffer to block the glass surface to check if looping ejections disappeared. Calibration of lengths was performed as in [85].

Microscopy. Samples were imaged at 60X and 100X using a Nikon perfect focus system and either a SYBR gold or FITC filter set. Images were acquired with a Photometrics CoolSnap ES2 camera; the exposure time was 75 ms and the frame rate was 4 fps.

Image analysis. A custom segmentation algorithm was developed using MATLAB. Ejection events were identified in each movie by local adaptive thresholding on the fluorescence intensity. DNA was identified in each ejection by using background subtraction, a Wiener filter for de-noising, and a custom Canny filter for edge finding. The long axis length, total intensity, and background (the mean intensity of non-DNA related pixels) were extracted for each ejection. Trajectories were subjected to a 3 bin median filter for de-noising.

Force-dependence assay. We follow a protocol adapted from Evilevitch et al. [87]. Phages were quickly swirled in a solution of 1% oPOE, 1% LamB, 1% DNase, and appropriate buffer containing PEG at 37 °C and incubated for 1 hour at 37 °C [89]. The solution was then incubated in a 65 °C water bath for 50 minutes to break open capsids and inactivate DNase. DNA in capsids was then extracted by a 1:1 phenol:chloroform extraction followed by chloroform extraction and finally ethanol precipitation. DNA lengths were assayed using field inversion agarose gel electrophoresis (100 V forward 0.8 sec, 60 volts backwards 0.8 sec, for 7–10 hours). After staining with ethidium bromide, migration distance was quantified using MATLAB.

Optical trapping. A schematic of the experiment is outlined in figure 3.1b. Phages were biotinylated after a standard plate lysis by incubation with 10 M of sulfo-LC-biotin, after first dialyzing for 48 hours with 100 mM sodium bicarbonate, pH 8.2. The phages were then purified on a CsCl gradient. Biotinylated phages were incubated with 2 μ m streptavidin-coated polystyrene spheres for a few hours at room temperature on

a rotating rack. The streptavidin spheres were washed by resuspension in TM 3 times and diluted tenfold into TM during use. The phage ejection solution (see single-phage ejection assay above) was injected into one port, and the phages conjugated to streptavidin spheres were injected into the other port at 10 L/min. A 1064 nm laser at 100 mW was used to trap phage-bound spheres, which were moved across the boundary layer of the two solutions. The microscope was home built around a 60X IR corrected water immersion objective adapted with a 200 mm focal length tube lens, and imaged with an Andor Ixon EMCCD camera. The schematic of the chamber in figure 3.1b, was adapted from [113].

3.4 Results

Our first set of experiments focused on optimizing the single-phage ejection assay with regards to SYBR gold concentration, since we noticed that phages spontaneously eject their DNA in high concentrations of SYBR gold. SYBR gold is thought to affect the mechanical properties of DNA since other DNA stains have been shown to affect DNA's persistence length [114]. To investigate the origin of these ejections and to test the hypothesis that they are triggered by SYBR gold, we incubated phages in different salts and different amounts of SYBR gold and subsequently measured the number of active phages by titering. The results are given in table 3.1; errors in the titer values follow (counting) statistics. For high amounts of SYBR gold (1:10,000), we see a steep drop in the number of active phages, between two and tenfold for the buffers containing magnesium. The drop is particularly drastic in pure sodium (1,000-fold), a condition thought to increase the DNA pressure inside the capsid. One explanation is that sodium in concert with SYBR gold increases the DNA pressure enough to destabilize the phage. Alternatively, magnesium and SYBR gold could compete with each other for binding sites on the DNA, providing a protective effect. Another possibility is that SYBR gold acts by a mechanism similar to LamB. See Supplemental Information for more details. Regardless, the drop in titer is consistent with the hypothesis that SYBR gold induces ejections. When the SYBR gold concentration is reduced (1:1,000,000), we recover almost the entirety of the original titer value. We concluded that it was preferable to perform the experiment at this lower concentration to minimize the effect of SYBR gold.

When performing the single-phage ejection assay (figure 3.1a), we always noticed two types of ejections — ejections that looked similar to those previously reported, and those that revealed a different type of dynamics that have the appearance of a "partial ejection" and were previously viewed as artifactual (figure 3.3a) [85]. Furthermore, we always saw spurious ejections of both types, as indicated in table 3.1, last column. In our investigations into the origins of the spurious ejections, we found that, under certain salt conditions (table 3.1), we saw more "artifact" ejections after addition of LamB (up to threefold increase) in

Table 3.1: SYBR gold triggers DNA ejection independent of LamB; the presence of LamB influences the character of ejection. (a) Columns 1–3: Titers of phage lambda solutions as a function of buffer composition and SYBR gold concentration. CsCl-purified lambda phage were dialyzed in the appropriate buffer by washing and spin filtering. SYBR gold was added to bring the solution to the right concentration (none, high (1:10,000), and low (1:1,000,000)). The number of viable phage remaining was measured by titering. High concentrations of SYBR gold reduce the titer by more than an order of magnitude, especially in the absence of magnesium. This reduction in titer is because SYBR gold induces the bacteriophage to eject their DNA. Lowering the concentration of dye recovers the original titer levels. (b) Columns 4–5: Character of ejection in the presence and absence of LamB. A single-phage ejection assay was run for each condition with and without LamB. The number of ejections, both continuous and looped, was counted manually and is reported above. The fraction of ejections that are looped is higher in the presence of LamB.

Ionic composition	Titer ($\times 10^{11}$)	Titer in high SYBR ($\times 10^{11}$)	Titer in low SYBR ($\times 10^{11}$)	% looped in low SYBR with LamB	% looped in low SYBR with no LamB
50 mM Tris 10 mM MgSO ₄ 0 mM NaCl	1	0.41	0.55	88	35
10 mM Tris 2.5 mM MgSO ₄ 0 mM NaCl	0.76	0.07	0.59	58.8	13.2
10 mM Tris 1.875 mM MgSO ₄ 2.5 mM NaCl	0.90	0.10	0.86	57.6	25
10 mM Tris 1.25 mM MgSO ₄ 5 mM NaCl	0.79	0.03	0.97	50	48
10 mM Tris 0.625 mM MgSO ₄ 7.5 mM NaCl	0.93	0.08	1.10	44.9	20
10 mM Tris 0 mM MgSO ₄ 10 mM NaCl	0.69	0.001	0.55	55.3	6.9

the single-phage ejection assay. This led us to question their designation as "artifact".

However, it is well known that DNA will stick to glass beads [115], and the previous experiment does not rule out DNA sticking to the microscope coverslip, although this seems unlikely since the addition of LamB (in excess) would tend to block sites on the glass for the ejecting DNA to bind to. The stickiness of glass was confirmed by incubating lambda phage DNA on freshly cleaned coverslips; upon observation, the glass was covered in DNA (data not shown). Thus as another test, we performed the ejection assay in the presence of the surface blocking agents BSA, casein, and heparin. These blocking agents significantly reduced DNA's affinity to the surface [116] — when bare lambda DNA was again incubated in a flow chamber with these blocking agents, only a single strand was found after extensive searching efforts (data not shown). However, even with these blocking agents, "artifact" ejections were seen in the phage ejection assay. Fortunately, the

content of the "artifact" ejections could be deduced from serendipitous ejections: occasionally, it could be seen that the DNA was in fact in a looped state. This was quantified by an intensity histogram and is shown in figure 3.2a. We therefore term this class "loop" ejections in contrast to the ejections previously characterized, which we call "continuous" ejections. The addition of LamB tends to increase the amount of looping ejections (table 3.1) which suggests that looping ejections are physiological.

We further demonstrated that the looping ejections were not surface mediated by ejecting biotinylated phages off of optically trapped 2 micron (Movie S1) and 500 nm (Movie S2) streptavidin-coated polystyrene spheres. A dual-syringe pump pushes fluid into the chamber (figure 3.1b). Using an optical trap, we grabbed hold of a microsphere, and moved it across the laminar flow boundary layer into the ejection buffer. The continuous flow system limits diffusive mixing at the meeting juncture. After a LamB trimer finds its target, a phage ejects and we can monitor the ejection (figure 3.2b). We found that the looped ejections were as common as continuous ejections. These experiments support the hypothesis that loops occur when the exiting DNA sticks to the LamB or the phage capsid, as may be the case in phage T5 [117].

We next characterized the dynamics of both the looped and continuous ejections under different ionic compositions, again using low concentrations of SYBR gold as described above (figure 3.1a). We first flowed in the dye front to visualize any spontaneous or SYBR-gold-induced ejections. Even at a tenfold lower concentration of SYBR than previously reported [85], there were still some spurious ejections, although much fewer than would occur with SYBR gold concentrations used previously [85]. Any lower concentration of dye would have precluded accurate segmentation of the ejecting DNA. After a period of time upon which the rate of spurious ejections was observed to be steady, we added LamB to the solution. A sudden large increase in the rate of visualized ejections indicated that LamB was inducing a large fraction of subsequent ejections. However, it should be noted that the dynamics of ejections induced by SYBR gold and those induced by LamB are indistinguishable. The experiment was repeated for 5 different salt conditions, in which we systematically varied the amount of sodium and magnesium ions while keeping the ionic strength constant.

We analyzed as many ejections within a field of view as possible — reasons for excluding an ejection from analysis included overlapping with another ejecting phage, photo-damage during an ejection, or a looped ejection that did not completely unfold. Ejections extracted from the field of view were partitioned into two categories — "continuous" and "looped". These are highlighted in figure 3.3a and b, respectively. To quantify the dynamics, we designed a custom segmentation algorithm for the DNA strands. The relation between length of DNA in basepairs and in pixels was determined by measuring, under flow, different known lengths of lambda DNA cut with restriction enzymes, and fitting the result to a quadratic function (figure 3.5 in the appendix). This calibration disentangles polymer stretch and shear flow from actual DNA length [110]. The

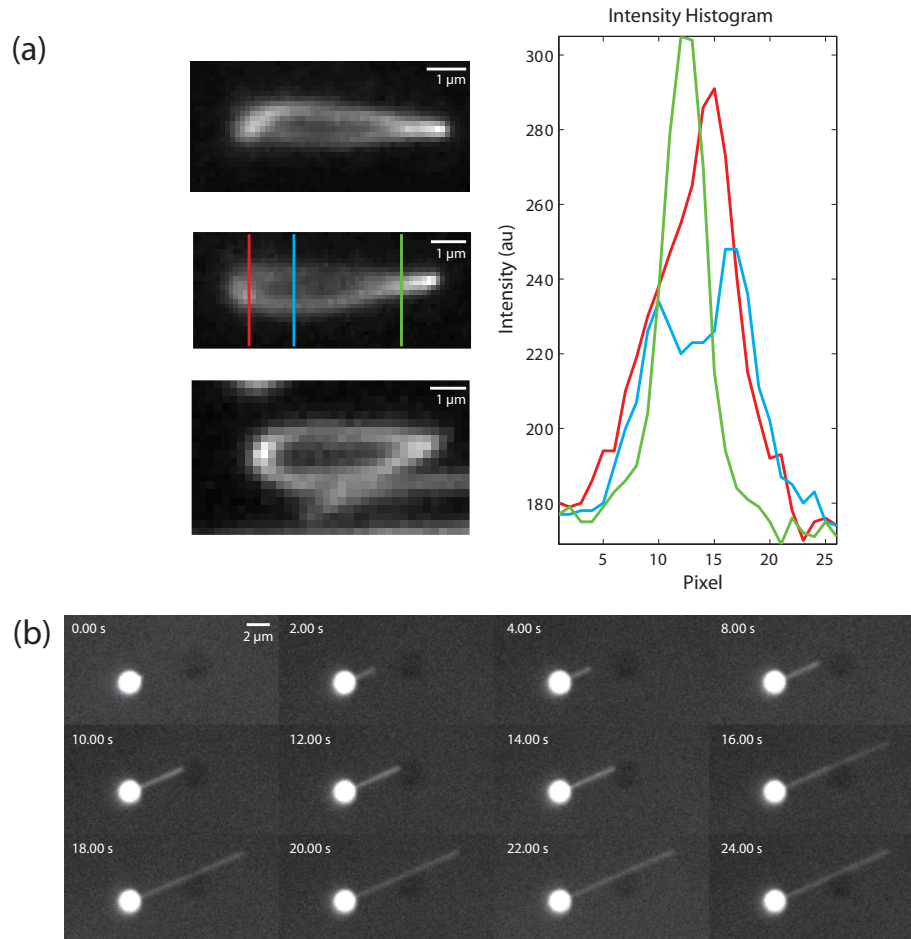


Figure 3.2: Looping during DNA ejection. (a) A montage of loops observed during the single-phage ejection assay. An intensity histogram constructed from one of the examples demonstrates the presence of loops during *in vitro* ejections. (b) Montage observation of a looped ejection during an optical trapping experiment. Phages were ejected off of optically trapped beads to suppress effects caused by the surface. Looped ejections are still observed, suggesting that they are not the byproduct of surface effects. The bead is 2 microns in diameter.

DNA's long axis length for each time point in a trajectory was then interpolated according to the quadratic function and normalized by 48.5 kbp, the full length of lambda phage DNA. This calibration is only accurate for continuous ejections; for looped ejections it is necessary to use intensity to discern the amount of DNA ejected. To do so, we take the total intensity of the DNA above background and normalize by the maximum intensity observed in a trajectory. This was done for both continuous and looped ejections (figure 3.3c and d, figure 3.6a–e).

A representative example of the trajectories is shown in figure 3.3c and d — trajectories for each ionic condition are in figure 3.6a–e. The fraction of DNA ejected is plotted versus time — the green curve uses fluorescence intensity to quantify total DNA ejected while the blue curve uses the calibrated length. It is clear in "looped trajectories" that the DNA is fact pinned at one end, as the trajectories exhibit a discontinuity in length at half-maximal ejection. A comparison between using calibrated DNA lengths and intensity as a measure of total ejected DNA is shown in figure 3.7a–e. Intensity measures give a larger estimate for the velocity of DNA ejected at early times because our segmentation algorithm misses the first bits of DNA ejection, and so changes in DNA length appear relatively large at small lengths; whereas at late stages in the ejection, length and intensity measures are equivalent metrics for amount of DNA ejected.

For each condition, we binned the trajectories to compute the average velocity as a function of DNA remaining in the capsid. This is plotted in figure 3.3e and f; figure 3.3e was computed using calibrated length while figure 3.3f was computed using intensities. Velocity plots for individual salt conditions are in figure 3.8a–e. From our analysis of the dynamics, we observed that individual trajectories were very similar to each other. As with previous work [85], this leads us to conclude that we are observing the same intrinsic dynamical process and that differences between trajectories are likely the result of measurement error. Secondly, our systematic perturbation of the sodium and magnesium concentrations — at fixed ionic strength — showed a consistent trend. As the amount of magnesium was lowered and the amount of sodium was increased, the speed of the ejections increased (figure 3.3e and f). From this, we conclude that it is the type of ion, not simply how much is present, that is an important control parameter that governs DNA ejection dynamics. We posit that the role of the positive ions is to screen the negative charge of the DNA backbone, which suggests that the charge of the counter-ions is the property of interest. Additionally, we observe that the continuous and looped ejections have velocity curves that are within error of each other (figure 3.3e and f). This suggests that the ejection mechanism for these two classes of ejections are very similar, if not identical, and that what happens to the first segment of DNA that exits the capsid determines whether an ejection will be continuous or looped.

To continue the physical characterization of the ejection process, we performed osmotic suppression experiments to determine the forces driving ejection in each of the different salt conditions. Osmotic pressure

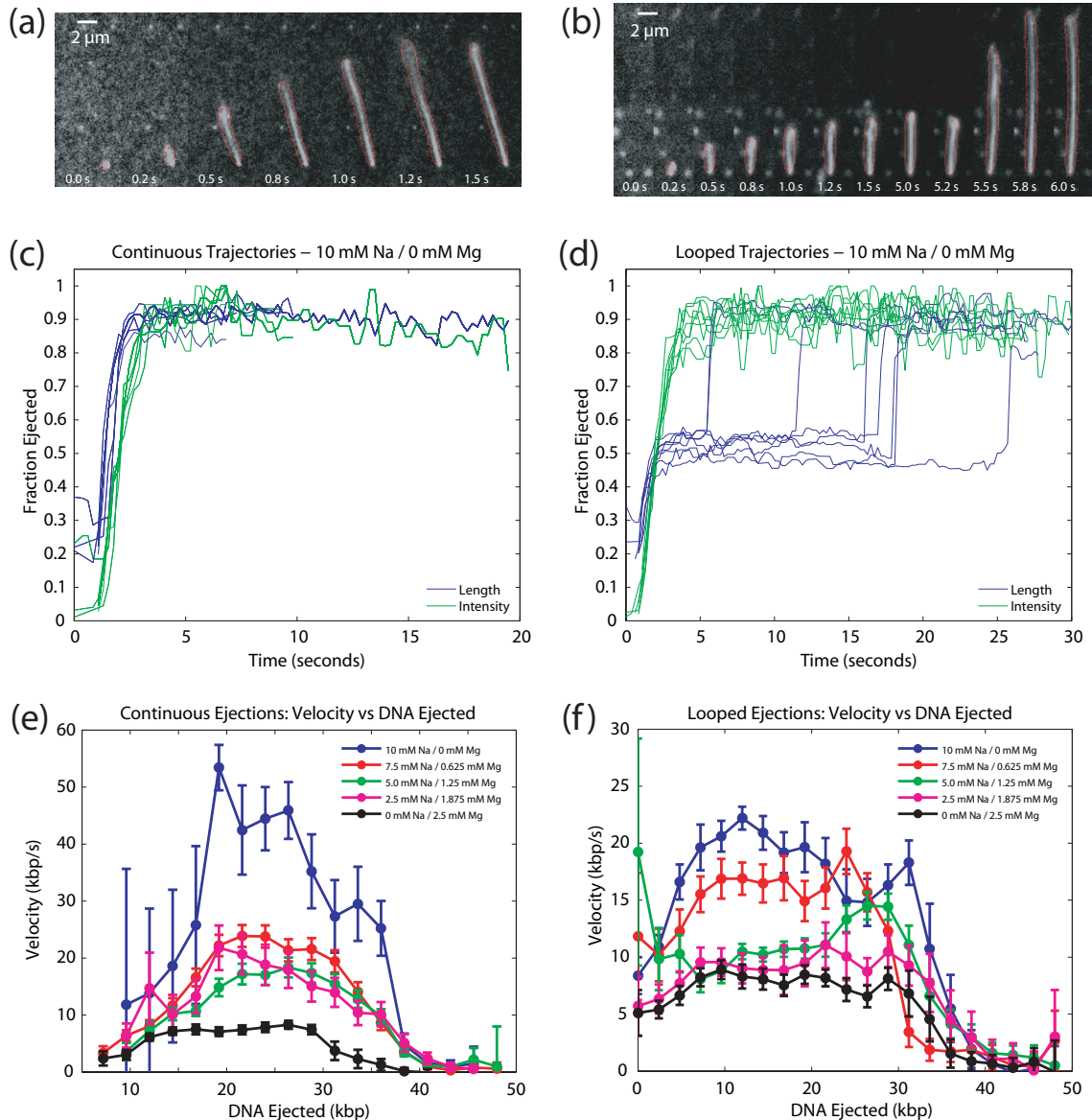


Figure 3.3: Trajectories of single-phage ejections and the ion-concentration dependence of the ejection dynamics. (a) Montage of continuous and (b) looped ejections. Movies of individual phages ejecting their DNA. The identified pieces of DNA are outlined in red. (c, d) Sample trajectories for continuous and looped ejections. The amount of DNA ejected was quantified by measuring either the calibrated long axis length or the total intensity above background of the segmented region. Lengths were determined by calibrating against restriction enzyme digested lambda DNA at known lengths, as described in the text. Individual trajectories are normalized by the maximum length/intensity observed during an individual trajectory to obtain the fraction ejected. Two qualitatively different ejections are observed — continuous ejections and looped ejections. The looped ejections reach half-maximal length before unfolding. The total intensity provides a way to quantify the total amount of DNA outside the capsid. (e) Velocity of ejection as a function of buffer composition for continuous trajectories and (f) looped. For each trajectory, the velocity of the DNA (measured using total intensity) at different landmark lengths was recorded. The mean velocity and the standard error of the DNA at each landmark length are plotted for each buffer condition. There is a clear trend in the data — buffers with more magnesium ions have a longer timescale for DNA ejection. Continuous and looped ejections have similar ejection dynamics in all buffer conditions.

varies with PEG 8000 concentration as given in [110]. Previous experiments have measured the amount of DNA ejected by pelleting phage capsids and measuring the absorbance at 260 nm. This measures the amount of DNA that has been chopped up by the DNase, which is assumed to be entirely ejected DNA. While the mass of DNA is thus accurately measured, it is not clear if the mass of ejected DNA can be quantitatively related to the length of DNA remaining in the phage capsids (see appendix, [118]). We used an alternative approach — purifying the DNA inside the capsids by standard extraction techniques and measuring its length directly by using field inversion gel electrophoresis (figure 3.9) [87]. This method is advantageous because we no longer need to assume that a constant fraction of phages eject for each PEG concentration [118]. Care was taken to avoid centrifugation, as it tends to fragment DNA (data not shown). The pressure driving ejection is shown in figure 3.4a. The general trend is that decreasing magnesium and increasing sodium increases the driving force, although the 5, 7.5, and 10 mM Na samples appear to be within error of each other. We observed that up to 65 atmospheres (~ 30 pN) of pressure was not enough to stall ejections in some salt conditions, which is consistent with theoretical expectations [85]. Atmospheres can be converted to force by multiplying the pressure by the end-on area of DNA, assuming a radius of 1.2 nm [89]. On top of the data are plotted theoretical predictions of the fraction ejected [89] as a function of pressure for 10 mM Na, 0 mM Mg, and 0 mM Na, 2.5 mM Mg, given the parameters determined in Rau et al. [119]. The theory, which assumes crystalline packing of DNA inside the capsid with intrastrand repulsive forces determined by the character of the counterions, fits the data quite well.

A simple model posits a linear relation between net forces and velocity, the coefficient of which is friction, or its inverse, mobility. We can determine how the mobility, denoted by $\sigma(l)$, of DNA inside the capsid depends on the salts (figure 3.4b) such that, $F(l) = v(l)/\sigma(l)$, where l is the DNA length in the capsid, F is the force as measured in figure 3.4a (multiplied by the area of end-on DNA), and v is the velocity as measured in figure 3.3e and f. This calculation assumes that the ejection is quasistatic; we assume that the DNA inside the capsid during the single-phage ejection experiments has relaxed to the equilibrium configuration in the osmotic suppression experiments. As previously observed [85], the mobility depends strongly on the amount of DNA remaining in the capsid. As the two extreme salt conditions overlap in figure 3.4b at increased DNA inside the capsid, we are unable to conclude that mobility is dependent on ion character at constant ionic strength.

3.5 Conclusion

We investigated the dynamics of DNA ejection in a variety of different salts, and also performed measurements on the self-repulsive force driving this process. We demonstrated that the counter-ion charge is an important

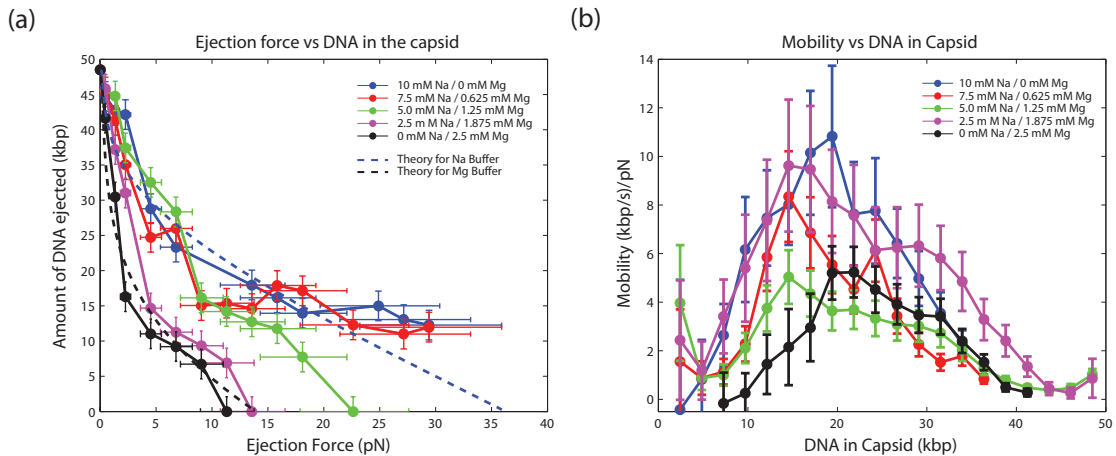


Figure 3.4: The force driving DNA ejection. Osmotic suppression experiments were performed in different buffer conditions to measure the force driving ejection. The fraction of DNA ejected was determined by measuring the amount of capsid DNA remaining using field inversion gel electrophoresis. (a) Increasing monovalent ion concentration raises the stall force of ejection at each length of DNA. Dotted lines are theoretical predictions according to [85]. (b) The mobility is measured as $v(l)/F(l)$. The trend is less clear regarding mobility and ion composition; it appears as though mobility is independent of ion composition for constant ionic strength. Vertical error bars were determined by the electrophoresis resolution, which was ± 2.1 kbp; horizontal error bars were determined by the pipetting accuracy of the PEG solutions, which had up to 5% variability.

control parameter for this class of experiments. These measurements should be of use to theorists working in physical virology. Of particular interest would be models that predict the force driving ejection in different salts, in addition to a better understanding of the time scale of the ejection process. We also investigated the origin of looped ejections. From our data, we can develop several possible models for these processes. It is possible that the end of the DNA gets stuck on the tail of the phage or LamB as it comes out, and that the rest of the genome subsequently comes out of the tail, accounting for the DNA loops occasionally seen (figure 3.2a). This may make it possible for phage to circularize even faster as the genome rushes to get inside its host, *E. coli*, as it is theorized that lambda delivers its genome directly into the cytoplasm without going through the periplasmic space [120]. Although the observation of looped ejections was first made in 2007 [85], this is the first attempt to explain this phenomenon. Of particular note is the relation of the current work to [82], which is an investigation of the packaging process as a function of salts; they note that by increasing the amount of multivalent cations in their packaging assay, the virus packages faster, and requires less force at an equivalent fraction of DNA packaged. This suggests that increasing magnesium reduces the amount of intrastrand DNA repulsion. In this current study, we see a similar effect; namely, an increased force is required to stall ejecting DNA at the same length, with increased monovalency. However, the mobility during DNA ejection appears to be independent of salt composition at our detection resolution. Notably,

the speed of ejection is faster during ejection in the presence of larger fractions of monovalent cations, which is consistent with the packaging data. Altogether, this work is a step forward in understanding the forces at play when bacteriophage lambda infects bacteria.

3.6 Appendix

Calibration of DNA length under flow. Figure 3.5 is the calibration of measured length of DNA in the microscope to actual length of DNA in kbp as determined by restriction enzymes [85]. The result is that length measured in the microscope is roughly quadratic with actual length in base pairs. We use this calibration to determine the actual length of ejected DNA in figure 3.3c and e, and figure 3.6.

Trajectory data. In this appendix we plot all trajectories (3.6) used to calculate velocities in figure 3.3e and f. In figure 3.7 and 3.8 we describe how the two measures of DNA ejected, length and intensity, are related to each other. In brief, fluorescence intensity consistently overestimates DNA at small lengths, and underestimates at long lengths (see figure 3.7 legend for details). Figure 3.6 demonstrates characteristic "continuous" and "looped" trajectories for different salt conditions. Looped trajectories clearly show a discontinuous change in length as a function of time, while the intensity measurements suggest a continuous exit of DNA from the capsid as a function of time and that all of the DNA has exited the capsid even though the length is much shorter than the fully stretched DNA. This suggests that for the "looped" trajectories the DNA is pinned to the capsid (figure 3.2b) and ultimately unfolds due to the presence of the flow, resulting in a discontinuous change in length as shown in the figure.

The discrepancy between measured length and intensity during the onset of phage ejection is because of our use of highly diluted SYBR gold dye (1:100,000-fold), which substantially reduces the signal/noise ratio. Our segmentation algorithm finds ejections by checking for fluorescence intensity. However, due to the low S/N, one strand of DNA, which to our eyes obviously appears as a single connected strand, looks like multiple disconnected strands to the computer. Thus, a median filter is used to smooth images so that those disconnected strands can become connected. At early times in a phage ejection, a small spot of DNA which is very dim but nonetheless detected has close to zero intensity count. However, after this small spot of DNA is smoothed by the median filter, the spot becomes elongated, resulting in the high length reported for near-zero DNA intensity.

Osmotic suppression assay. Evilevitch et al., 2003, developed a method to measure the amount of DNA remaining in phage capsids after incubation with LamB in the presence of osmolytes using spectrophotometry (absorbance at 260 nm) and by field inversion gel electrophoresis (Evilevitch et al., 2005); in the former measurement, the fraction of phages that did not eject was assumed to be constant at different concentrations

of PEG. On the basis of this assumption, the A260 measurement can be directly converted from mass to length; however, if the fraction of ejected phages changes from preparation to preparation, spectrophotometry will not produce an accurate measure of length. Moreover, sample-to-sample differences in preparation (DNA precipitation yield, especially) make it difficult to conclude that there is a constant unejected fraction — this is shown in figure 5 of Evilevitch et al., 2005, where there is up to a 30% variability in A260 measurements. To circumvent this issue, we chose to visualize the DNA length directly by performing agarose gel electrophoresis (figure 3.9) only. We see clearly that increasing the external pressure inhibits DNA ejection. By measuring the migration distance in each lane, we can relate the pressure exerted on the phage to the amount of DNA that was not remaining in the capsid, and in this fashion derive a pressure corresponding to the DNA left in the capsid as a function of amount of DNA ejected (figure 3.4a).

Above each lane in figure 3.9 is indicated the amount of external pressure induced by the presence of the PEG as measured in atmospheres. The relation between PEG concentration and pressure is described in Michel, 1983. Lanes labeled "L" correspond to the DNA ladder, with the marker lengths noted in (figure 3.9a). Increasing amounts of external pressure results in increasing amounts of DNA left inside the capsid after ejection. Migration distance in each lane was measured by looking for the maximum fluorescence peak relative to the 48.5 kbp peak, and comparing that to the corresponding ladder. Since electrophoresis was performed in many different batches, we present the results in such a way that the relevant pressure experiment in the appropriate salt condition is to the right of its corresponding ladder. Electrophoresis parameters: 100 V forward, 0.8 sec; 60 V backwards, 0.8 sec; 7–10 hours.

3.7 Acknowledgements

We thank the members of the physiology course at the Marine Biology Laboratory, Nicolas Chiaruttini, Paul Grayson, Zenan Chang, Alexander Grosberg, Ian Molineux, Michael Rubinstein, Virgile Viasnoff, William Gelbart, Charles Knobler, and members of the Phillips laboratory. We also gratefully acknowledge financial support from several sources, including a NIH Medical Scientist Training Program Fellowship, a Yaser Abu-Mostafa Hertz Fellowship, and a NIH Director's Pioneer Award. We also acknowledge the support of NSF grant number 0758343.

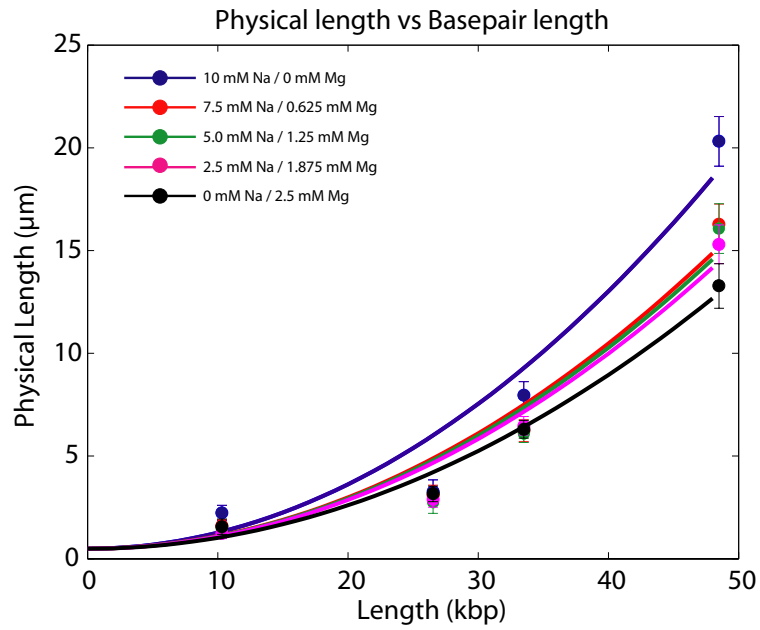


Figure 3.5: Calibration of tethered DNA length. To convert the physical length of DNA to the length in kbp, we performed a calibration as done by Grayson et al., 2007. Briefly, biotinylated lambda genomes were either treated with restriction enzymes (XhoI, NsiI, and PmlI) or kept untreated. The DNA was attached to a glass coverslip through a biotin-streptavidin bond. The DNA was extended with a $40 \mu\text{l}/\text{min}$ flow and the lengths of different pieces of DNA were measured manually using ImageJ. The mean length was fit to a quadratic function; error bars represent the standard deviation of the measured length.

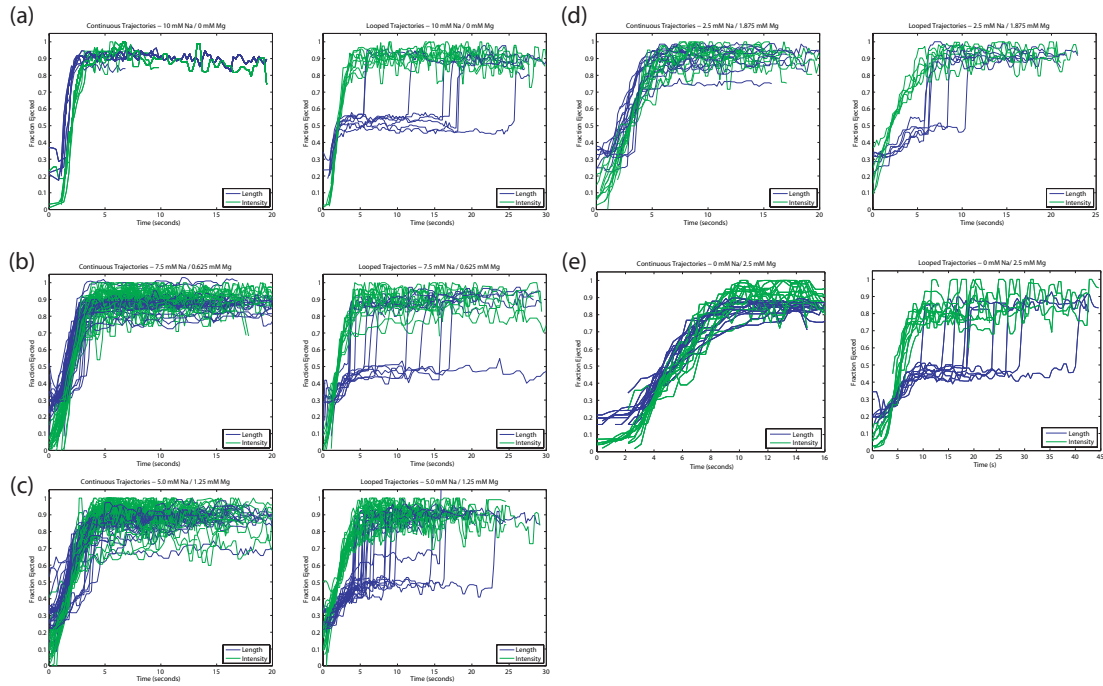


Figure 3.6: Trajectories used in the paper for all salt conditions. Green: intensity, blue: length. Intensity is the summed intensity of the DNA labeled by SYBR gold. The length is calibrated by measuring defined-length lambda DNA, via restriction digests, under the same flow conditions. Length is then calculated by interpolation. Looping trajectories uniformly unfold at half the maximum length. (a) Salt condition: 10 mM Tris, pH 7.4, 10 mM NaCl, 0 mM MgSO_4 . Number of trajectories: 5 continuous and 8 looped. (b) Salt condition: 10 mM Tris, pH 7.4, 7.5 mM NaCl, 0.625 mM MgSO_4 . Number of trajectories: 36 continuous and 13 looped. (c) Salt condition: 10 mM Tris, pH 7.4, 5.0 mM NaCl, 1.25 mM MgSO_4 . Number of trajectories: 43 continuous and 25 looped. (d) Salt condition: 10 mM Tris, pH 7.4, 2.5 mM NaCl, 1.875 mM MgSO_4 . Number of trajectories: 14 continuous and 6 looped. (e) Salt condition: 10 mM Tris, pH 7.4, 0 mM NaCl, 2.5 mM MgSO_4 . Number of trajectories: 15 continuous and 10 looped.

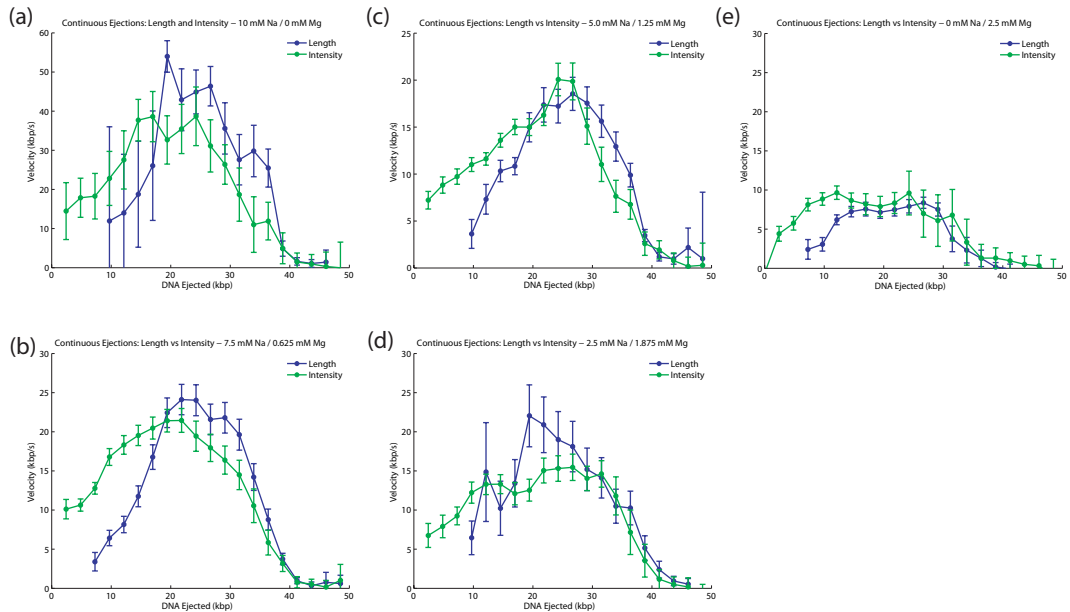


Figure 3.7: Velocities from the single-phage ejection experiment. Each panel has two classes — green lines denote velocities measured by intensity; blue lines denote velocities measured by length. Intensity measures give a larger estimate for the velocity of DNA ejected at early times because our segmentation algorithm misses the first bits of DNA ejection, and so changes in DNA length appear relatively large at small lengths. Since DNA lengths are hard to directly measure via restriction digest calibration at early times in ejections, we don't measure lengths less than around 5 kbp. The number of trajectories used in each velocity plot is listed in figure 3.5 caption. The errors were determined by standard error. (a) Salt condition: 10 mM Tris, pH 7.4, 10 mM NaCl, 0 mM MgSO_4 . (b) Salt condition: 10 mM Tris, pH 7.4, 7.5 mM NaCl, 0.625 mM MgSO_4 . (c) Salt condition: 10 mM Tris, pH 7.4, 5.0 mM NaCl, 1.25 mM MgSO_4 . (d) Salt condition: 10 mM Tris, pH 7.4, 2.5 mM NaCl, 1.875 mM MgSO_4 . (e) Salt condition: 10 mM Tris, pH 7.4, 0 mM NaCl, 2.5 mM MgSO_4 .

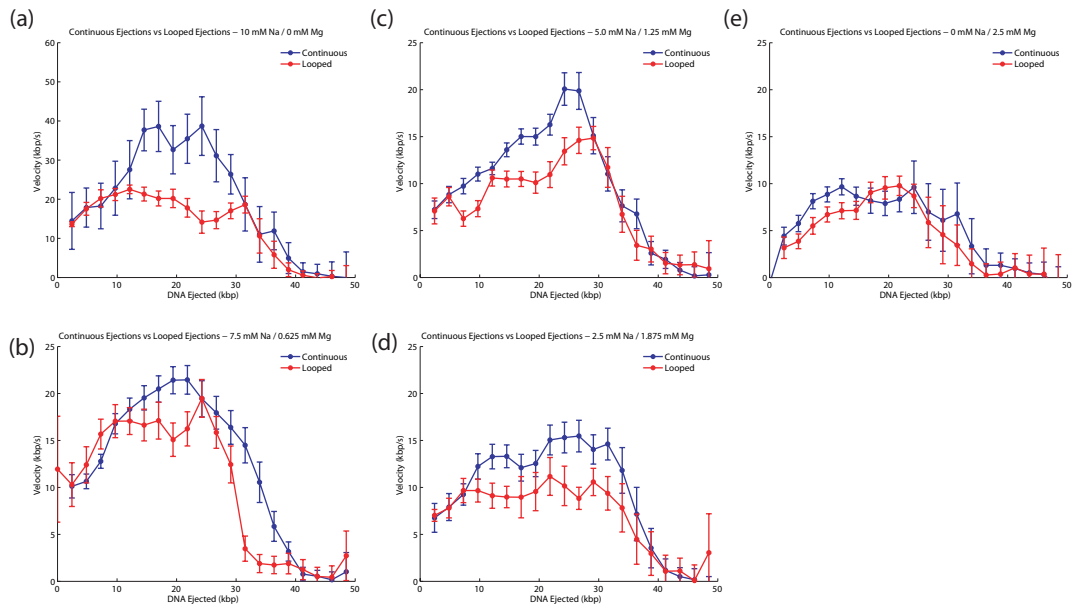


Figure 3.8: Velocity of DNA at different salt conditions. Red: velocity based on looped trajectories. Blue: velocities based on continuous trajectories. All velocities were determined using intensities; both classes of ejections show similar dynamics. The number of trajectories used in each velocity plot is listed in figure 3.5 caption. The errors were determined by standard error. (a) Salt condition: 10 mM Tris, pH 7.4, 10 mM NaCl, 0 mM MgSO_4 . (b) Salt condition: 10 mM Tris, pH 7.4, 7.5 mM NaCl, 0.625 mM MgSO_4 . (c) Salt condition: 10 mM Tris, pH 7.4, 5.0 mM NaCl, 1.25 mM MgSO_4 . (d) Salt condition: 10 mM Tris, pH 7.4, 2.5 mM NaCl, 1.875 mM MgSO_4 . (e) Salt condition: 10 mM Tris, pH 7.4, 0 mM NaCl, 2.5 mM MgSO_4 .

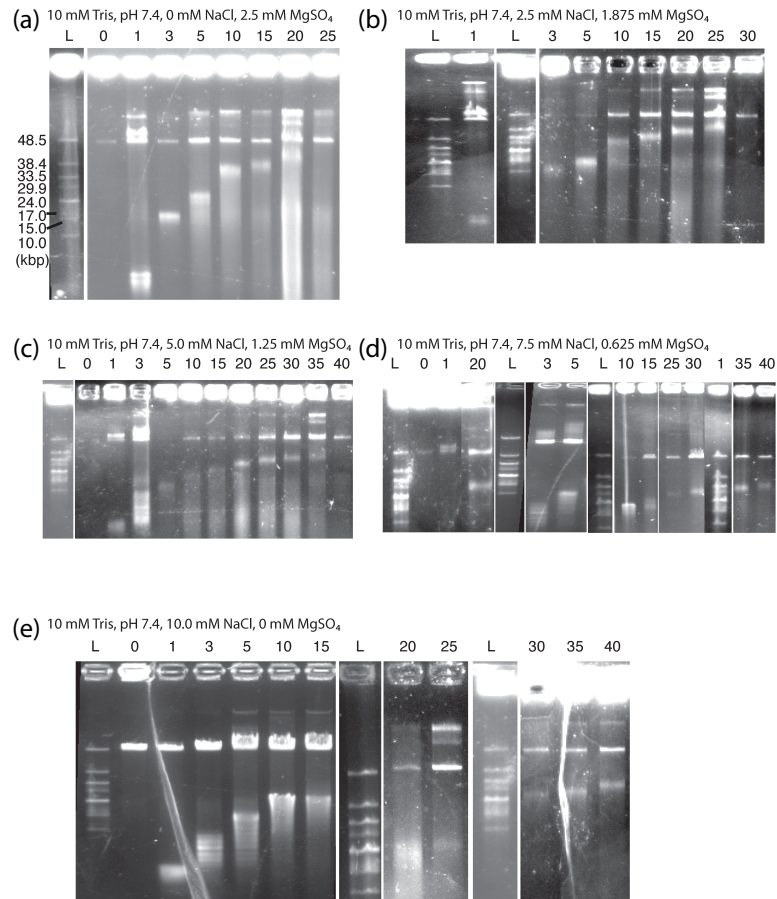


Figure 3.9: Field-inversion gel electrophoresis of force-dependence assay. A trend of DNA retention in the capsid with increasing external osmotic pressure is clearly visible. (a) Salt condition: 10 mM Tris, pH 7.4, 0 mM NaCl, 2.5 mM MgSO₄. (b) Salt condition: 10 mM Tris, pH 7.4, 2.5 mM NaCl, 1.875 mM MgSO₄. (c) Salt condition: 10 mM Tris, pH 7.4, 5.0 mM NaCl, 1.25 mM MgSO₄. (d) Salt condition: 10 mM Tris, pH 7.4, 7.5 mM NaCl, 0.625 mM MgSO₄. (e) Salt condition: 10 mM Tris, pH 7.4, 10 mM NaCl, 0 mM MgSO₄.

Chapter 4

A single-molecule Hershey-Chase experiment

4.1 Introduction

The advent of single-molecule biophysics resulted in a resurgence of interest in the life cycle of bacterial viruses, which were originally central players in the development of molecular biology [121–123]. Much information has been gathered recently about the structure of both individual viral proteins and assembled viruses (for examples, see references in [101, 124]). These structural efforts are complemented by a variety of experiments that have elucidated the biophysics of binding of viruses to their host receptors [125–127], the mechanism of genome delivery [87, 94, 128–131], and the mechanics of DNA packaging during viral assembly [2, 132, 133]. In addition, the long-standing interest in the developmental decision between lysis or lysogeny [134] has also been explored at the single-cell level [77, 79, 80]. The study of bacterial viruses remains a rich source of examples of fundamental biological problems: the self-assembly of proteins into higher-order structures, nucleic acid motors, and genetic decision making.

Controversy surrounds the nature of the mechanism responsible for driving DNA translocation from the viral capsid across the membrane into the cytoplasm [129]. Even the basic consideration of whether the energy stored in the compacted viral DNA plays an active role in the DNA translocation process continues to be disputed. One approach to exploring these questions has been through *in vitro* ejection measurements using both bulk and single-molecule approaches [83, 84, 86, 87, 89, 130, 135]. In the case of phage lambda, these packaging forces are sufficient to drive DNA ejection *in vitro*. *In vivo* bulk measurements of the ejection time for phages T7 and lambda [128] provide a lower bound on ejection times. Although these studies have contributed greatly to our understanding of phage infection, there has not been a single-molecule approach to measuring ejections either *in vivo* or against a resistive force such as that offered by the bacterial cytoplasm.

The current interpretation of these experiments can be summarized as the *in vivo* pressure conundrum.

Simply stated, it is thought that the turgor pressure [93] of the bacterial cell exerts a force on the incoming viral DNA. Once some fraction of the viral genome has been delivered, the driving force — solely derived from DNA compaction — is reduced and at some point is no longer sufficient to deliver the remaining DNA into the cell. The driving force is thought to equalize with the cell pressure at around 50 percent of genome ejection [88, 89]. Several models have been proposed to explain how the remaining viral DNA is delivered; these are schematized in figure 4.5. All of these models are intriguing and have been the subject of theoretical inquiry [91, 94, 105, 106, 136], but it is clear that determining which model best reflects the ejection process for phage lambda requires a new generation of experiments that quantify the *in vivo* ejection process.

Our objective, therefore, has been to quantitatively image the DNA translocation process *in vivo*. This is in direct analogy to the Hershey-Chase experiment [74], where it was first directly shown that phage DNA is transferred directly into bacterial hosts, and moreover, that DNA is the molecule of genetic inheritance. In the work described here, we adapt a similar strategy to that exploited in the Hershey-Chase experiment, but with the use of fluorescent labels rather than radioactivity. By looking at this problem from a single molecule perspective, we are able to observe the cell-to-cell variability in the translocation process and gain insight into the mechanism of *in vivo* DNA ejection. We also present a system that enables the simultaneous visualization of phage DNA and phage proteins *in vivo* that builds on previous work [80, 137, 138]. We use this system to examine the timescale of DNA replication and capsid assembly at the single-cell level.

The remainder of the chapter is organized as follows: We begin with a discussion of the experimental design and discuss our strategy for observing phage DNA *in vivo*. Next we discuss the protocols and analysis methods employed for this experiment. In the results section, we present our observations on the *in vivo* DNA translocation process. We demonstrate that the time scale of DNA translocation is 1 to 20 minutes and that there is marked cell-to-cell variability in the dynamics. We also image DNA replication and capsid assembly *in vivo* and comment on the the dynamics of both processes. We conclude by discussing our observations and what mechanistic insights we can gain from them.

4.2 Experimental design

We have considered two different scenarios for tracking the dynamics of the ejection process, pre-ejection labeling and post-ejection labeling. These are highlighted in figure 4.1a and b, respectively. The first scenario is to use a fluorescent marker to stain the viral DNA while it still remains in the capsid [139] and then to watch the ejection process as the fluorescence is transferred from the capsid to the cellular interior (figure 4.1a). There is much evidence that suggests this first approach is feasible. Phages stained with cyanine dyes have previously been observed to infect cells [140]. There have also been studies characterizing the

interaction of cyanine dyes with DNA inside the phage capsid and the effect their presence has on phage stability [139]. Further, cyanine dyes have also been used to study the kinetics of viruses in living eukaryotic cells, demonstrating that some dyes have limited cytotoxicity [141]. To identify a suitable dye, we screened a number of candidate dyes for their ability to (i) penetrate the phage capsid, (ii) maintain phage stability and infectivity while stained, (iii) preserve *in vitro* ejection kinetics while stained, and (iv) not cross the membrane of living cells. Our screen identified SYTOX Orange as a dye with all of these properties; protocols for its use and the results of the screen are further detailed in both the methods section and in the appendix.

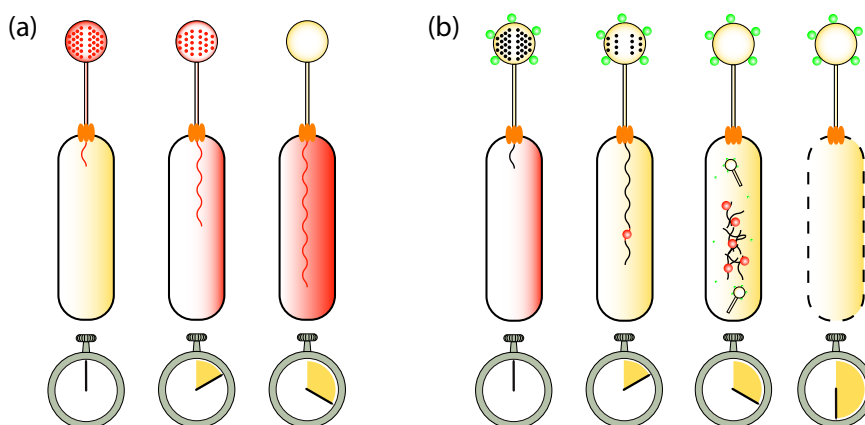


Figure 4.1: Schematic of experimental procedures for measuring *in vivo* phage DNA ejections. (a) A schematic for monitoring DNA translocation with pre-ejection labeling. The DNA is stained while still in the capsid. During ejection, the phage DNA carries its complement of cyanine dye with it transferring fluorescence intensity from the capsid to the cellular interior. (b) A schematic for post-ejection labeling of phage DNA. The infecting virus carries a YFP label on its capsid and a *parS* sequence in its genome. Polymerization of mCherry-ParB at the *parS* site fluorescently marks the phage genome when it is inside the cytoplasm. Dual color labeling allows both viral capsid protein and DNA to be monitored over the course of an infection.

We also employ a complementary strategy of adding ectopic binding sites which are labeled upon the entry of the viral DNA into the cell; this is schematized in figure 4.1b. When the *parS* site arrives in the cytoplasm and becomes accessible to the mCherry-ParB proteins, mCherry-ParB binds cooperatively at *parS*, producing a fluorescent focus [142]. This method builds upon earlier work in which fluorescently labeled capsids and fluorescently labeled genomes were separately monitored [79,80,137,138]. By supplementing the dye-labeled phages with this orthogonal protocol, we independently monitor the arrival of viral DNA within the cell in the absence of fluorescent dye molecules. To do this, we engineered a gpD-YFP phage containing a *parS* binding site and infected a strain of *E. coli* expressing mCherry-ParB. This allows us to monitor the fate of the viral DNA after its arrival in the bacterial cytoplasm and watch the implementation of certain key parts of its genetic program all the way to completion.

4.3 Materials and methods

Engineering λ DVV1 with recombineering. We used recombineering to engineer a phage that carries both a YFP label fused to the gpD protein and a *parS* sequence in the b region of the viral genome. A schematic of the engineered genome is shown in figure 4.4a; the full sequence is available on request. Our protocol is adapted from methods used by Court et al. and St. Pierre et al. [79, 138, 143]. We began by transforming the strain λ eyfp [137], which carries a lysogenized lambda phage with a YFP tag on the gpD protein, with the plasmid pKD46 [144]. The plasmid pKD46 carries an arabinose inducible λ Red system, which enables *E. coli* to undergo homologous recombination. Next, λ eyfp + pKD46 was grown to an OD600 of ~ 0.4 – 0.6 in LB + 50 μ g/mL ampicillin + 10mM L-arabinose, made electro-competent by washing the cells 5 times with 10% ice cold glycerol, and was stored at -80 °C for future use. We then amplified the *parS* sequence from the plasmid pALA1073 by PCR using primers that contained homology to both pALA1073 and to the b region of the lambda genome. The exact sequence of the primers are given in the appendix (table 4.3). The remaining plasmid DNA was then degraded by DpnI digestion. The electro-competent λ eyfp + pKD46 cells were then transformed with the insert by electroporation. The cells were given 2 hours to recover, plated on LB + 30 μ g/mL kanamycin plates, and grown overnight at 30 °C. The resulting colonies were screened for the presence of the *parS* site using colony PCR; colonies that tested positive were sent for sequencing (see 4.16). The results of the colony PCR screen and sequencing can be found in the appendix. The phage was then recovered from the lysogen using liquid lysis. Next, we transformed the strain LE392 with the plasmid pPlate*D and infected LE392+pPlate*D with the engineered phage and screened for lysogens using the kanamycin selection marker. This plasmid is necessary to correct the defect in capsid assembly (i.e. reduced burst sizes and the presence of empty capsids as revealed by electron microscopy) seen in phages expressing gpD fused to YFP [80, 137]. The final lysogen and the resulting phage were labeled λ DVV1.

Plate lysis. NZYM top agarose (NZYM + 0.7% agarose) and NZYM plates (NZYM + 1.5 % agar) were prepared prior to plate lysis. NZYM top agarose was melted on a hot plate and then stored in a 45 °C water bath until needed. The host cell strain, either LE392 or LE392 + pPlateD* for YFP phages, was grown up overnight in 5 mL of LB or LB + ampicillin. The saturated cell culture was then centrifuged for 5 minutes at 5000 g and the pellet was resuspended in 5 mL of SM buffer. A 100 μ L aliquot of cells was then mixed with 1 μ L of phage stock in a 14 mL culture tube and incubated at 37 °C for 20 minutes. Next, 3 mL of NZYM top agarose was added to the culture tube, gently mixed, and poured onto a NZYM plate. The plates were incubated for 12–16 hours at 37 °C or until lysis was visually apparent. After incubation, phages were recovered by pouring 5 mL of SM buffer onto the plate and placing the plate on a rocking station at 4 °C. After 5 hours, the SM buffer was recovered. The lysate was sterilized by adding chloroform to a

concentration of 1% and gently vortexing. Bacterial debris and chloroform was then removed by centrifuging for 10 minutes at 5000 g and recovering the supernatant. Plate lysis typically yielded titers of $\approx 10^{10}$ pfu/ml.

Liquid lysis. The protocol for liquid lysis follows a variant of Sambrook *et al.* [111]. An overnight of the desired lysogen (λ DVV1 for instance) was grown up overnight at 30 °C to saturation in LB. Four separate 1 L flasks were each filled with 250 mL of LBM supplemented with the appropriate antibiotics. Each flask was then inoculated with 1 mL of the saturated overnight culture and placed in a 30 °C shaker until the culture reached an OD600 of ~ 0.6 – 0.8 . Each flask was then removed from the shaker and then heat shocked in a 45 °C shaking water bath for 15 minutes to induce the lysogens to enter the lytic cycle. The flasks were then placed in a 37 °C shaker for 90 minutes, enough time for the lytic cycle to reach completion. This was verified both by monitoring the OD600 before and after the 37 °C incubation and by visually inspecting the culture for clarity and floating bacterial debris. The liquid lysate was sterilized by adding 10 mL of chloroform to each flask, mixing gently, and then incubating at room temperature for 1 hour. The aqueous portion of the mixture was then recovered and passed through a 5 μ m filter to remove any remaining bacterial debris. Bacterial DNA was degraded by treatment with DNase for 1 hour and NaCl was added to a concentration of 1 M. The phages were then concentrated by PEG precipitation. PEG 8000 was added to a concentration of 10% and the phages were allowed to precipitate overnight. The bacteriophage were pelleted by centrifugation at 11000 g for 10 minutes at 4 °C and gently resuspended in SM buffer. Liquid lysis typically yielded titers of $\approx 10^{10}$ pfu/ml. Titering of phages containing a YFP label on the capsid was performed with the strain LE392+pPlate*D to recover wild-type plaque morphology.

Real-time single-molecule ejection assay. Strain LE392 was grown up overnight in LB at 37 °C. The saturated culture was then diluted 1:100 in M9maltose-sup and grown for 3 hours at 37 °C until the culture reached OD600 ~ 0.3 . Phages from plate lysates were centrifuged for 5 minutes at 13,000 g to remove cellular contaminants. The phage-containing supernatant was recovered and then stained with SYTOX Orange at a final concentration of 500 nM for 3 hours at room temperature. Glass coverslips were cleaned by sonication for 30 minutes in 1 M KOH and 100% ethanol, with copious rinsing with milliQ water in between, and then dried on a hot plate while wrapped in aluminum foil. Flow chambers were assembled by first briefly (5 seconds) immersing the cleaned coverslip in a fresh solution of 2% polyethyleneimine (PEI), briefly immersing the coverslip in milliQ water, and then drying with a gentle stream of air. A microscope slide, double-sided tape, and the treated coverslip were then used to assemble a flow chamber. Prior to infection, free dye was removed from the phage suspension by putting 100 μ L of phage through 4–5 rounds of dilution and concentration by spin filtering with M9sup and an Amicon Ultra 4 filter with a 100 KDa. Each round led to a 40-fold dilution of dye, reducing the concentration of dye to less than 200 pM. After the final round, the phages were brought up to the original volume of 100 μ L with M9sup. Phages were then

bound to cells by mixing 57.5 μL of cells with 2.5 μL of phage (MOI 7) for 1 minute at room temperature. The cells were then flowed into the flow chamber and allowed to adhere to the surface for 2 minutes at room temperature. Measurements of phage binding and infection kinetics show that phages will not eject their DNA within this time period at room temperature [127, 145]. The flow chamber was then washed with 200 μL of M9sup supplemented with 1% GODCAT, 1% beta-mercaptoethanol, and 0.5% glucose. The chamber was then sealed with valip and imaged on a Nikon Ti-E Perfect Focus microscope after quickly focusing on the fluorescence channel. Samples were illuminated with mercury lamp excitation passed through a TRITC filter set. An environmental chamber was used to keep the sample at 37 °C. Snapshots of both the phase and fluorescence channels were taken either 1 or 4 times a minute, with a fluorescence excitation time of 500 ms or 300 ms, respectively. Images were collected using either a Hammamatsu C8484 camera or an Andor iXON EMCCD camera. We observed better conservation of fluorescence between the phage and the cell with the Hammamatsu camera as opposed to the Andor camera; see the appendix for more information. The EM gain of the Andor camera allowed for shorter exposure times and higher time resolution.

Imaging viral infection using λDVV1 . Strain LE392 + pAW61, which carries a plasmid harboring mCherry-ParB controlled by a pBAD promoter, was grown up overnight in LB + 50 $\mu\text{g}/\text{mL}$ ampicillin at 37 °C. The saturated culture was then diluted 1:100 into M9sup + 50 $\mu\text{g}/\text{mL}$ ampicillin and grown for 2 hours until the culture reached OD600 \sim 0.2. The cells were then induced to express mCherry-parB by adding L-arabinose to a concentration of 1 mM and incubating them for another hour at 37 °C (OD600 \sim 0.3). Prior to imaging, agar pads were made by placing a 24 mm x 60 mm coverslip on a sheet of parafilm, pipetting 2.5 mL of warm M9sup + 1.5% agarose on the coverslip, and then placing another 24 mm x 60 mm coverslip on top. After the solution solidified, small 5 mm x 5 mm squares were cut out using a razor blade. Cells were infected at an MOI of 10 (\approx 10 μL of cells and 10 μL of phage) at room temperature for 10 minutes. 2 μL of the cell/phage mixture was then spotted onto the agar pad and allowed to dry for 5 minutes at room temperature. The agar pad was then inverted and placed on a Willco dish with a glass coverslip bottom for imaging. Samples were imaged with a Nikon Ti-E perfect focus microscope. An environmental chamber was used to keep the sample at 37 °C. The samples were quickly focused using the FITC filter set and images were taken in phase and fluorescence every minute for 90 minutes. The samples were excited using mercury lamp fluorescence passed through a FITC (YFP) and TRITC (mCherry) filter set; exposure times were 1000 ms and 500 ms, respectively.

Image analysis. Cells of interest were manually identified in each movie and cropped from the field of view using ImageJ. The phase images for each cell were segmented using custom image analysis software created with MATLAB. Briefly, the phase images from each frame in the movie were registered with the first frame by finding the peak in the cross-correlation function. This registration was then applied to all

fluorescence channels as well, removing spatial drift from the data set. Next, the registered phase images were all added together; this step combines the information each frame contains about the location of the cell and greatly reduces difficulty of segmentation. Lastly, thresholding and minor morphological operations were used to create a mask of the cell from the combined image. This method also has the advantage of creating a mask for every frame simultaneously, rather than segmenting each frame individually. Cells were occasionally segmented manually when quality phase images were not available. Quantities of interest, including background levels, intensities, and the location of phages in each frame were extracted for further analysis.

4.4 Results

We first verified that our chosen dye, SYTOX Orange, could stain phages appropriately (figure 4.6) and allow for *in vitro* ejection events (figure 4.7). By binding these stained phages to cells, we were able to visualize ejection events. As seen in figure 4.2a, the attachment of the viruses to the host is revealed by the presence of diffraction-limited spots on the cell surface. Image segmentation is used to identify these spots by creating a mask that identifies the pixels associated with either the virus or the cell (figure 4.2b). With these masks, it is possible to query the total fluorescence intensity at each time step. As shown in the montage of images (figure 4.2c), the ejection process is characterized by a loss of fluorescence intensity in the phage and a concomitant increase in fluorescence in the cellular interior (figure 4.2d). Note that the fluorescence inside the cell is diffuse; we believe this reflects the dye molecules unbinding from the phage DNA and redistributing themselves along the host genome. This is to be expected, as the residence time of SYTOX Orange on DNA is roughly ~ 1 s [146]. These two signals are analogous to earlier *in vitro* measurements in which DNA ejection from viruses was inhibited through osmotic inhibition resulting in partial ejections. In that case, absorption measurements yielded the quantity of DNA that had escaped from the capsid while agarose gel electrophoresis measured the DNA remaining in the capsid [87,135].

Figure 4.3 shows the results of a number of distinct ejection events. For the measurements shown here, the fluorescence signal associated with the virus decreases on a time scale of minutes, a factor of ten or more longer than the corresponding dynamics observed *in vitro* [85,135]. The mean time for ejection, when greater than 95% of fluorescence intensity has left the virus capsid, is 11.9 ± 8.0 min ($n = 15$). In addition to the widespread variability in ejection time, there are a number of ejections which demonstrate what appear to be pausing events, which we define as non-decreasing fluorescence levels greater than 2 minutes (our time resolution was either 30 sec. or 1 min., so 2 min is the minimum time required to see a trend, according to Nyquist criterion); the mean pause time is 4.2 ± 1.0 min ($n = 11$). Within our criterion for defining pauses,

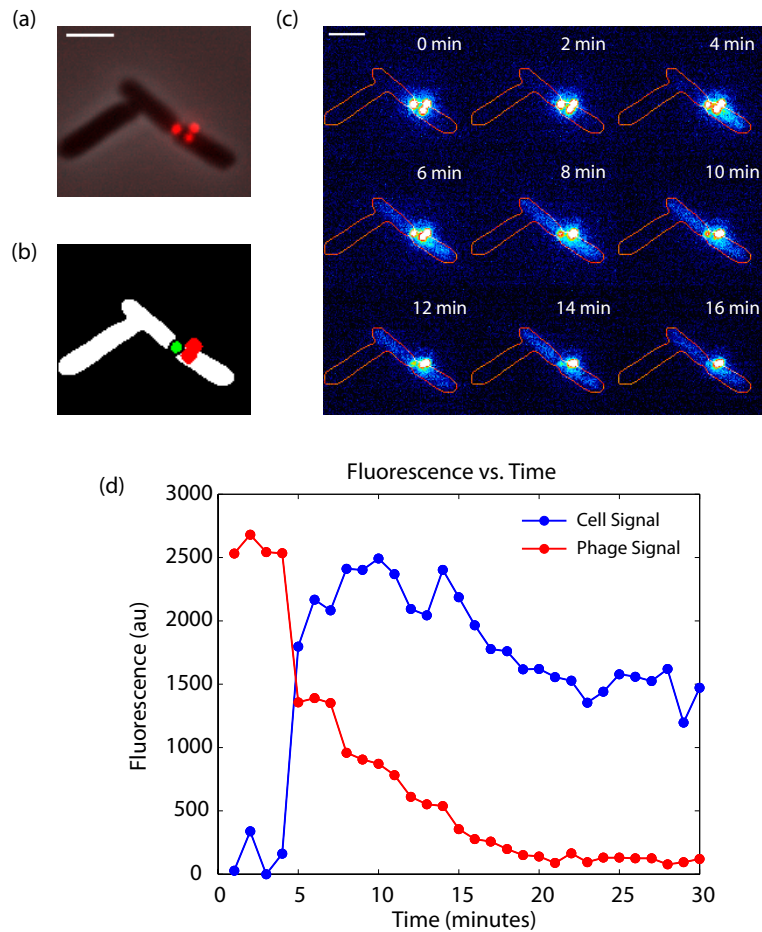


Figure 4.2: Dynamics of DNA ejection. (a) Viruses attached to the cell surface. (b) Segmentation masks of the cell (white), the phage which ejects its DNA (green), and the phages that do not eject their DNA (red). (c) Time sequence of the fluorescence in the cell. The edge of the cell is outlined. (d) Fluorescence intensity as a function of time. The intensity of the phage segmented region and the cell segmented region are each plotted separately. Note that in this ejection there appear to be steps and pauses. The scale bar in figures (a) and (b) is 2 microns.

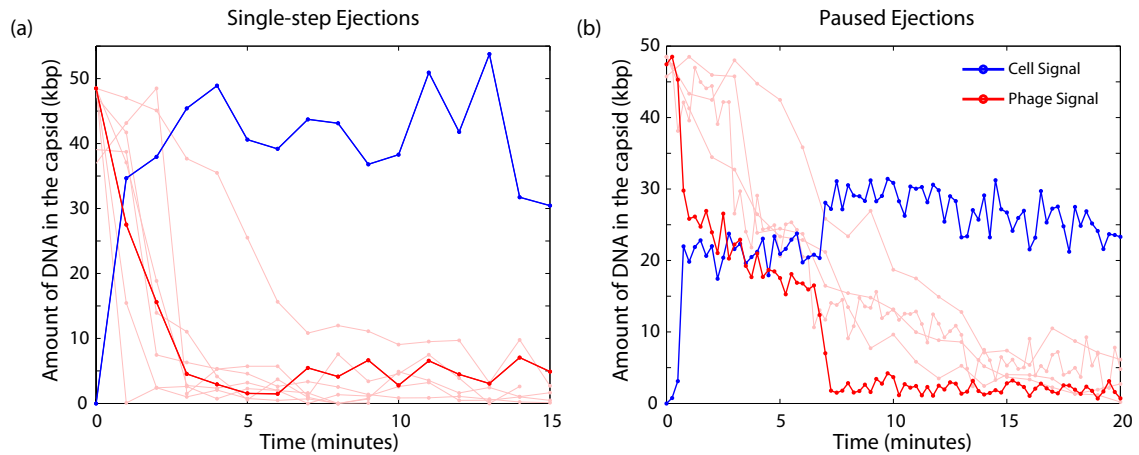


Figure 4.3: Ejection trajectories from single-cell infections. The red trajectories show the time history of the DNA intensity within the virus and the blue trajectories show the increase in the fluorescence in the cellular interior. Only two representative traces for the intensity within the cell are shown, with the remaining trajectories presented in the appendix. (a) Trajectories displaying a rapid and continuous ejection. (b) Trajectories that exhibit pausing events. The images associated with all of these time traces are shown in the appendix.

they all occurred within the first half of an ejection event. We have categorized the trajectories into two classes: single step and paused. These are shown in figure 4.3a and figure 4.3b, respectively. The full set of trajectories is shown in figures 4.11–4.13. Note that in the absence of attached phage, cells do not show any increase in fluorescence (figure 4.8), which may be a concern from oxidative damage to cell membranes.

Two issues that might affect the quantitative interpretation of this measurement are photobleaching and dye leak through the capsid. Although there is some degree of both effects over the time scale of the experiment, as discussed in more detail in the appendix (see figures 4.9 and 4.8), these effects cannot account for the decrease in intensity observed in the trajectories shown in figures 4.2d, 4.3, and in the appendix (figures 4.11–4.13). Closer analysis of these trajectories reveals that the rise in fluorescence in the cell plateaus at values comparable to the starting fluorescence in the phage. This supports our argument that contributions of photobleaching and dye leak to our observed signal are not significant, and that we are measuring the translocation of viral DNA into the cellular interior.

We were interested in a second method of observing the putative ejection events as a way to independently confirm their identity as DNA translocation processes; this would also allow us to better understand the fate of the viral genome once it enters the host. As indicated schematically in figure 4.1b, we simultaneously monitor the phage binding to the cell through a capsid protein gpD-YFP fusion, and the fate of the ejected DNA via mCherry-ParB in the host. A binding site on the phage genome serves as a seed for the localization of the mCherry-ParB fusion (figure 4.4a). An example of a time history using these read outs is shown in

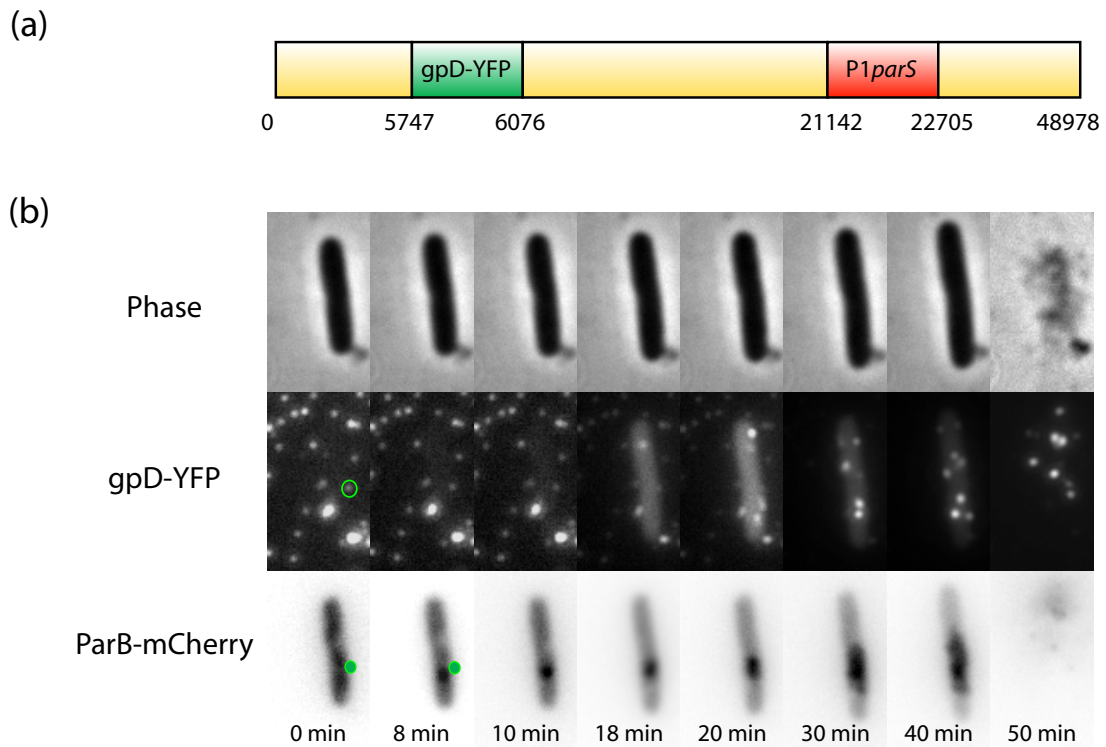


Figure 4.4: Monitoring ejection dynamics by binding of fluorescent proteins onto the viral genome. (a) Schematic of the genome of lambda phage showing the loci responsible for the fluorescence of the capsid (gpD-YFP) and for the binding of the ParB proteins (*P1parS*). The numbers beneath the schematic indicate where on the lambda genome the loci of interest are found. (b) Microscopy images showing the time course during and after the viral infection. The phage that are the possible sources of the initial ejection are highlighted in green. Over time, the presence of phage DNA and the formation of new viral capsids is revealed by the appearance of puncta in the both the mCherry-ParB channel and gpD-YFP channel, respectively.

figure 4.4. This enables us to visualize viruses that are bound to the cell surface, monitor the phage DNA as it first appears in the cellular interior, and observe the creation of new viral capsids. This is evident from both of the fluorescence channels which show an increase in the mCherry-ParB signal signifying the replication of viral DNA and the appearance of green fluorescent protein, first diffusely and then in the form of puncta (i.e., newly assembled phage capsids) (figure 4.4b). However, beyond its significance as an independent verification of the results shown in figures 4.2 and 4.3, this experiment also helps us develop a sense of the viral life cycle after the initial DNA ejection event.

Because we are monitoring fluorescence that reports on the state of both the viral genome and one of its protein products, by finding suitable ways to query the fluorescence, we can time key parts of the viral life cycle on a cell-by-cell basis. In particular, note that in the early times, the YFP signal is diffuse. We interpret this as the synthesis of viral capsid proteins. However, at later times, there is the clear appearance of distinct puncta, which we believe reflects the formation of intact viral capsids. Using data like that shown

in figures 4.2 and 4.4, our aim was to attempt to watch the viral infection process directly and to get a sense for the timescales of the replication of the viral genome, the synthesis of viral capsid proteins, the assembly of these capsids, and even the lysis of the host. The dynamics of the infection process was first spelled out carefully in experiments by Ellis and Delbruck revealing the one-step growth curve [147]. By monitoring the formation of viral plaques on a lawn of bacteria as a function of time after the infection process, they were able to infer the kinetics of viral binding, growth of phages within the infected host, and cell lysis. In the time since, a much higher resolution picture of these various processes has been assembled by pulling together insights from genetics, structural biology, electron microscopy, biochemistry, biophysical investigations, and many other directions as well. Here, we are able to directly observe each of these processes occurring in real-time underneath a microscope.

To get a sense for the timescales of translation and capsid assembly, we measured the initial time for each of these processes, using the appearance of viral DNA as the start time for infection. We found that the mean time to translation is $\sim 5 \pm 0.9$ minutes. This measurement is confounded by the maturation time of YFP, which suggests that translation might begin shortly after infection [148]. We also observe that DNA replication typically occurs on the timescale of ~ 10 minutes, capsids are first formed roughly $\sim 11 \pm 1.3$ minutes after infection, and that the time to lysis is $\sim 38 \pm 2$ minutes. Caution should be taken in strictly interpreting these numbers. The presence of mCherry-ParB could cause viral genomes to stick together, as has been observed in other systems, perturbing the natural time course of DNA replication [149]. Furthermore, it is also known that the assembly of viral capsids is defective when every gpD protein is tagged with YFP [80, 137]. However, even with these caveats in mind, we can still use these single-cell observations to get a sense for the timescale of each of these processes.

4.5 Conclusion

Bacterial viruses were once referred to by Max Delbruck as “the hydrogen atom of biology”. These words conjure up images of a model system that is at once tractable and yet capable of delivering lessons of wide biological significance in the same way that the hydrogen atom has served as a paradigm for quantum mechanics [150]. In the end, these viruses turned out to be much more than the simple model system that Delbruck envisioned, and along with one of their primary hosts, namely *E. coli*, have served as the primary workhorses for the development of modern molecular biology. Because of the deep knowledge about these viruses that was gained during this time period, they have now become an ideal tool for revisiting some of the very same questions that fueled that revolution, but now with an eye to precisely and quantitatively dissecting the physical processes that make up their life cycles.

For many bacterial viruses, one of the key early processes in the viral life cycle is the delivery of the viral genome into the host. A number of different hypotheses have been formulated for the actual translocation mechanism, some of which are highlighted in figure 4.5. In addition to the driving force due to the packaged DNA, these models propose that thermal fluctuations, hydrodynamic drag, and active molecular motors might each play a role in bringing the viral DNA into the bacterial cell. Naturally, the model that best reflects the actual physical mechanism should be consistent with experimental data. Based on our experience with previous observations of *in vitro* DNA ejection, we would not expect to see marked variability in the ejection trajectories as seen here, *in vivo*, with a purely mechanical mechanism [85, 86, 135]. The origin of the apparent pauses seen might provide a hint towards the ejection mechanism, as DNA-based motors acting against a load have been observed to pause [82, 151]. However, the pauses observed here are much longer than the pauses observed for motors and it is quite likely they could simply be a reflection of the cell-to-cell variability in turgor pressure. Another possibility is that the pauses observed here might also be related to mechanisms proposed for pauses observed *in vitro* for phage T5 [84, 86]. However, we feel this is unlikely as pauses are not observed for phage lambda *in vitro* [85, 135]. Curiously, [152] and [153] note that the transcription rate is between ~ 40 and 80 nt/sec which translates to roughly 10 minutes to pull DNA in, which matches our observed time scale. The current data does not match previous calculations on mechanisms based on DNA binding proteins and thermal fluctuations [154]. Also perplexing is the time-scale of ejection; the origin of the friction exerted on the DNA that sets the timescale for ejection is poorly understood, both *in vitro* and *in vivo*. It is possible that a deeper understanding of the friction will further the insights gained from our data.

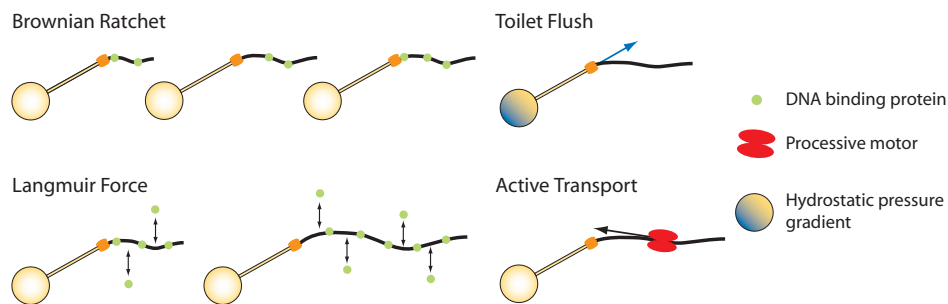


Figure 4.5: Models for DNA ejection into living cells. A number of hypotheses have been put forth for how DNA enters the highly crowded cellular interior. The Brownian ratchet and Langmuir models posit that DNA binding proteins attach to the DNA and prevent it from moving back out [155, 156]. The toilet flush model is based upon the idea of an osmotic imbalance between the cellular exterior and interior that is short circuited by viral binding [157]. The active transport mechanism refers to the translocation of molecular motors like RNA polymerase on the ejected DNA once it enters the cell [136].

As a step towards building a mechanistic picture of the ejection process *in vivo* in the case of phage

lambda, we have performed two different classes of *in vivo* single-molecule ejection experiments. The results of these experiments are featured in figures 4.2, 4.3, 4.4, and the appendix. These experiments reveal that the DNA translocation process, like many cellular processes, is subject to strong cell-to-cell variability with the ejection times exhibiting a wide range from approximately 1 to 20 minutes; furthermore, the character of the ejections are also highly variable, with some some ejection events exhibiting pauses. By simultaneously labeling the capsid proteins and the genomic DNA, it is possible to follow the viral life cycle in real time after the arrival of the viral DNA in the cytoplasm. In particular, by monitoring the fluorescence associated with ParB proteins, we have a real-time readout of the replication process of the viral DNA. By monitoring the fluorescence of the capsids, not only can we see phages bound to the cell surface, but we can also see the synthesis of the capsid proteins within the cell and their subsequent assembly into virus particles.

4.6 Appendix

The main purpose of this appendix is to provide detailed controls to demonstrate that our microscopy observations correspond to *bona fide* ejections of viral DNA into living cells. In particular, we first verified that the SYTOX Orange stain had appropriate specificity and sensitivity to visualize phage lambda. We next verified that SYTOX-Orange-stained phages eject their DNA in a comparable *in vitro* experiment, in contrast to many other dyes which allow visualization of DNA ejection but do not permit prior staining. After this, we made sure that SYTOX-Orange-stained phages does not impede the lifecycle of lytic phage and bacteria. We furthermore show that the host cells, in the presence of stained phage, do not have a spurious increase in fluorescence. That is, the presence of dyes and fluorescence excitation is known to degrade cell membranes. In order to rule out photobleaching artifacts, we measured the rate of photobleaching in our system and demonstrated that the time scales of ejection and photobleaching are well separated. Lastly, we provide all our ejection trajectories. We also supply the media and strains used in this work, as well as verification that phage λ DVV1 has the proper genotype.

4.6.1 Dye-bound phage ejection controls

4.6.1.1 SYTOX Orange staining of phage is equivalent to DAPI

We wanted to ensure that the phage DNA was indeed being labeled by SYTOX Orange. Though earlier work has shown the binding properties of similar dyes, it is still of interest to make comparisons in the case of phage lambda [139]. We compared the SYTOX Orange stain with DAPI, which is known to be a quantitative indicator of DNA mass in phages [158]. As shown in figure 4.6a, colocalization of the DAPI stain and the

YFP labeled capsid proteins demonstrates that the nucleic acid of each such virus can be identified with DAPI. We then show in figure 4.6b simultaneous staining and colocalization of the viral DNA with both DAPI and SYTOX Orange, confirming that SYTOX Orange is both a sensitive and specific indicator for the presence of phage lambda.

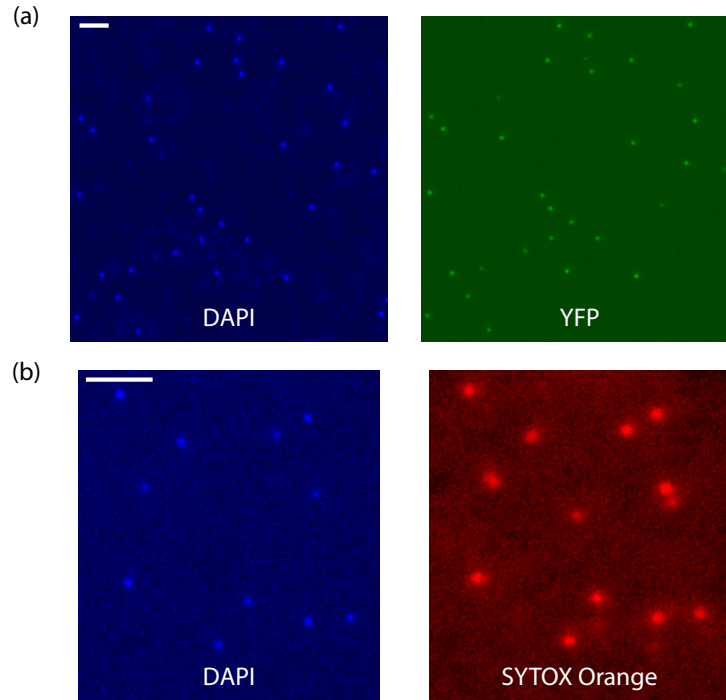


Figure 4.6: Simultaneous staining of λ DVV1 with SYTOX Orange and DAPI. Here, we verify that our engineered phage λ DVV1 express a YFP-tagged capsid protein and that SYTOX Orange will penetrate the capsid and stain phage DNA. In (a) and (b), phages were stained with a given concentration of dye (see caption for each image below), flowed into a chamber made of a KOH cleaned coverslip and double-sided tape, and then imaged. (a) Staining of λ DVV1 with a final DAPI concentration of $0.5 \mu\text{g}/\text{ml}$. It has been previously established that DAPI will readily penetrate the phage capsid [80,158]. We observe perfect co-localization of the DAPI and YFP signals, demonstrating that every phage has a YFP capsid label. (b) Co-staining of λ I60 with $0.5 \mu\text{g}/\text{ml}$ concentration of DAPI and $5 \mu\text{M}$ SYTOX Orange. Phages were stained with SYTOX Orange for 3 hours and then stained with DAPI for a few minutes. We observe perfect co-localization of the DAPI and SYTOX Orange signals, demonstrating that SYTOX Orange will enter the phage capsid. The presence of fluorescent puncta also provides evidence that SYTOX Orange will not adversely affect phage stability (in contrast to other dyes like SYBR Gold) [135]. An extra spot appears in (b) which is due to the presence of excess dye that was not dialyzed out. The scale bar in each image is 2 microns.

4.6.1.2 SYTOX-Orange-stained phage eject their DNA *in vitro*

Another important question concerning the presence of the SYTOX Orange dye is the extent to which it might perturb the dynamics of the viral DNA ejection. To explore this question, we use SYTOX-Orange-stained phages as the basis of an *in vitro* single-molecule experiment already performed in our laboratory

and that of others [84–86, 135]. This is an especially important control since the properties of SYTOX Orange are different from many other dyes which allow visualization of DNA ejection but do not permit prior staining [135]. As can be seen from Figure 4.7, the ejection profiles of SYTOX-Orange–stained lambda phage display fundamentally the same dynamics as phage ejecting into a solution of SYBR Gold [135] which “read out” the phage DNA only after it has exited the capsid.

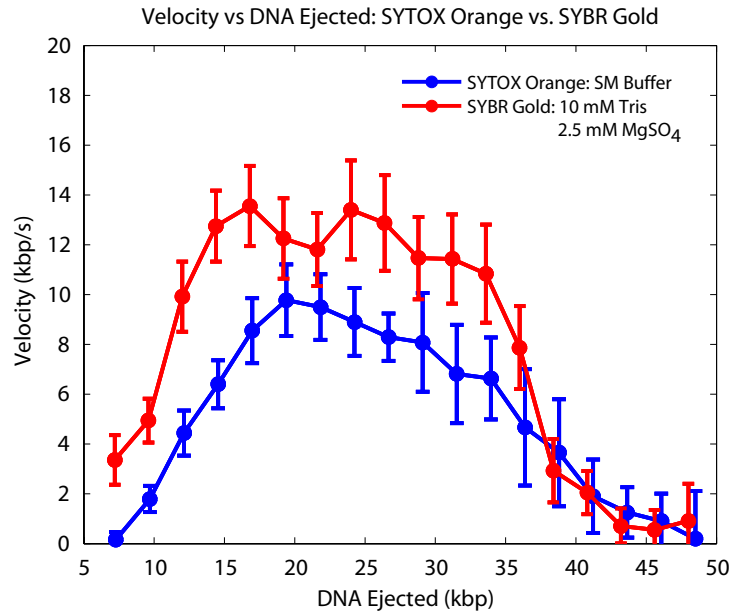


Figure 4.7: *In vitro* single-molecule ejection experiment using SYTOX Orange. Phages were first pre-stained with 5 μM SYTOX orange for three hours; the kinetics of *in vitro* DNA ejection from the pre-stained phage were then observed in SM buffer using previously established protocols, with 5 μM SYTOX Orange replacing SYBR Gold as the dye of choice [135]. The kinematics of DNA ejection in 10 mM Tris/ 2.5 mM MgSO₄, a solution that has a similar magnesium content as SM buffer, is also plotted for comparison. We see no significant difference in the dynamics of ejection for phages that were pre-stained as opposed to phages that were not.

4.6.1.3 SYTOX-Orange–stained phage behave like wild-type phage in terms of bulk infection

Another question that can be asked about the possible perturbations induced by the presence of the dye molecule is the extent to which these dye molecules alter the macroscopic titers of infectious phage. In particular, we wanted to make sure that SYTOX-Orange–stained phages permit the life cycle of lytic phage and their host bacteria to continue. In other words, after staining the phages with SYTOX Orange, by how much does the the titer of the phage decrease? The starting titer of our phages was $\sim 1 \times 10^{11}$. After staining with 5 μM of dye, and appropriate dialysis (see Materials and Methods), the resultant titer was measured to be $\sim 1 \times 10^{10}$ to $\sim 3 \times 10^{10}$ indicating between a three and tenfold drop in titer. This result confirms the results in [135]; that adding dye will result in a drop in titer but will still permit most phages to

still be active. From our microscopy observations (figure 4.6) and our titering experiments, we conclude that SYTOX Orange appropriately stains every phage, and that a significant fraction of those stained phages are actively infectious, and are able to produce phage progeny; thus, the activity of phage stained by SYTOX Orange is not decreased sufficiently to render the experiment impractical, unlike the effects of SYBR Gold staining [135].

4.6.1.4 Cells in the absence of phage do not increase in fluorescence

One of the consequences of the presence of nucleic acid dyes is the spurious increase in fluorescence of cells in the absence of phage attachment. In particular, the concomitant presence of dyes and fluorescence excitation is known, in some cases, to degrade the cell membranes via oxygen radicals and hydrogen peroxide generation. This results in cell death, which could cause an increase in intracellular fluorescence as dye leaks into the cells and binds to the nucleoid [159]. In these cases, spurious fluorescence could result in confounding effects which would interfere with our ability to identify and monitor putative ejection events. We have no evidence that such perturbations occur with SYTOX Orange at the concentrations used in these experiments.

We demonstrate here that cells, in the presence of stained phage elsewhere in the same chamber, do not have a spurious increase in fluorescence. We segmented images for every cell without bound phage and then monitored the internal fluorescence of those cells over time. As can be seen in figure 4.8, there is no evidence from any 89 cells inquired that there is an increase in fluorescence in the absence of phages. We moreover used the GODCAT system (see table 4.2) which is known to be protective to *E. coli* in the presence of oxygen radicals [159]. Thus, whenever we do see an increase in fluorescence, the result must have been from phage ejecting its DNA into the cell interior.

4.6.1.5 Photobleaching and ejection time scales are separated

In order to properly characterize the ejection events, photobleaching and viral DNA ejection must be distinguishable. To do this, we measured the rate of photobleaching and demonstrate that the time scales of ejection and photobleaching are well separated, as in Figure 4.9. Phages were prepared with SYTOX Orange as described in the Materials and Methods. Briefly, λ cI60 lysate was first centrifuged for 5 minutes at 13,000 g and the supernatant was stained with 500 nM SYTOX Orange for 3 hours. Free dye was then removed by 4 rounds of spin filtering through a 100 kDa cutoff Amicon filter at a 1:40 ratio of sample:buffer. A dilution of phage containing $\sim 10^9$ pfu/ml was then flowed onto KOH cleaned chamber and incubated for 2 minutes. The chamber was then washed with a solution of M9 medium supplemented with 1% GODCAT, 1% beta-mercaptoethanol, and 0.5% glucose. The sample was illuminated with mercury lamp excitation passed through a TRITC filter set; images were taken every minute with 500 ms exposure for 30 minutes,

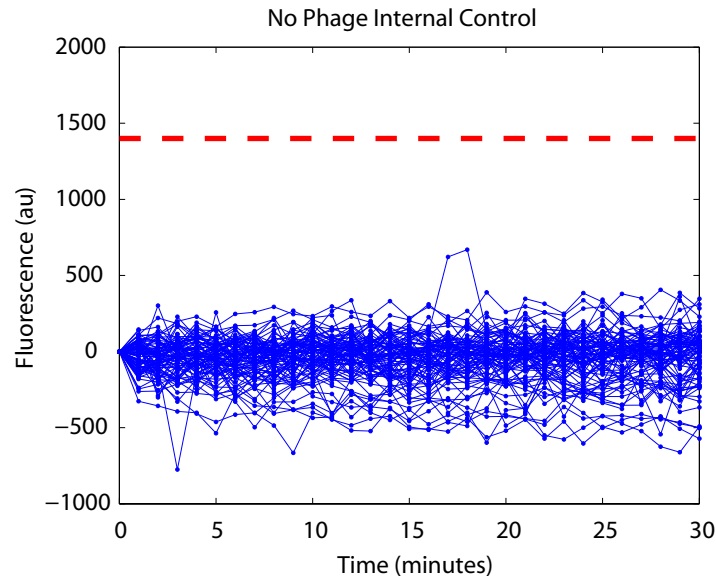


Figure 4.8: No phage internal control. 89 cells which had no discernible attached phages were analyzed for their level of fluorescence as a function of time (blue lines). The red line is the average total fluorescence change in an ejecting phage. The change in fluorescence for cells without phages is negligible compared to those with phages.

precisely the same conditions used in the *in vivo* experiment. The fluorescence images were processed using a custom MATLAB script.

This data (figure 4.9) demonstrates that while significant photobleaching does occur over the course of 30 minutes, ejections, as shown in figures 4.2 and 4.3 in the main text and 4.11 below, range from 1 minute to 20 minutes. Thus for the typical ejection timescale of 15 minutes, photobleaching can account for at most a $\sim 30\%$ loss of signal. On the other hand, the fluorescence losses seen for the putative ejection events were much more stereotyped and typically faster than the monotonic decreases seen to result from explicit photobleaching. This lends credence to our use of the rapid decrease in phage signal as a marker for putative *in vivo* ejection events.

A second way to explore the same question was to observe the cells in which several phages were bound. For those cases in which only one of the bound phages underwent a putative ejection event, by simultaneously monitoring the level of fluorescence in the other, non-ejecting phage, we could directly compare an ejecting and non-ejecting phage, both subject to photobleaching. Examples of these events are shown in figure 4.10. This data is consistent with figure 4.9 and demonstrates that the loss in fluorescence in ejecting phages occurs on a faster timescale than photobleaching and dye loss. The green lines were determined to be ejecting phages, because their fluorescence decrease mirrors the increase in fluorescence inside the cell, and is much faster than photobleaching (figure 4.9). The red lines, therefore, represent the fluorescence of non-ejecting

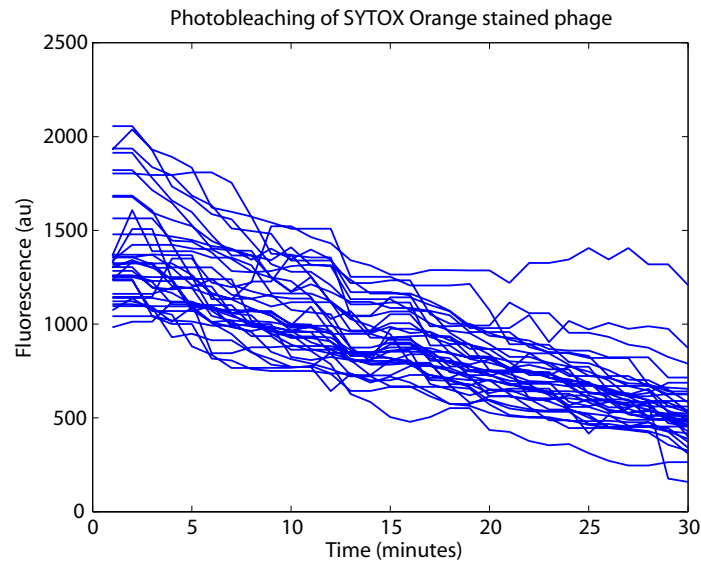


Figure 4.9: Photobleaching in SYTOX-Orange-stained bacteriophage. The trajectories of the total fluorescence above background for 36 phages are shown. The time scale for photobleaching is 30 minutes, and all the trajectories are decreasing monotonically.

phages. As can be seen, the time scale of photobleaching is much longer than that of ejection, which allows us to distinguish ejection from photobleaching.

4.6.1.6 Single-phage ejection trajectories

Here we provide all of our ejection trajectories. As can be seen from the trajectories below, the decrement of fluorescence in the phage is accompanied by a rise in fluorescence inside the cell, indicating that DNA is being transferred from the phage into the cell. These trajectories are the first direct *in vivo* measurements of DNA translocation, a single-molecule analog to the experiments by Hershey and Chase in 1952 [74]. As can be seen from the trajectories, the distribution of time to complete ejection is highly variable. In figures 4.11–4.13, we see that for the most part there is fluorescence conservation between ejecting phages (red lines) and increase in fluorescence in the cells (blue lines). The masks for ejecting phages are displayed in green, the masks for non-ejecting phages are displayed in red, and the masks for the cell are displayed in white. Often, trajectories contain steps and sometimes the fluorescence is not conserved. The lack of conservation could be due to dye being stripped off the phage as it translocates its DNA. It could also be due to nonlinearities introduced by the EM gain in the Andor iXon camera.

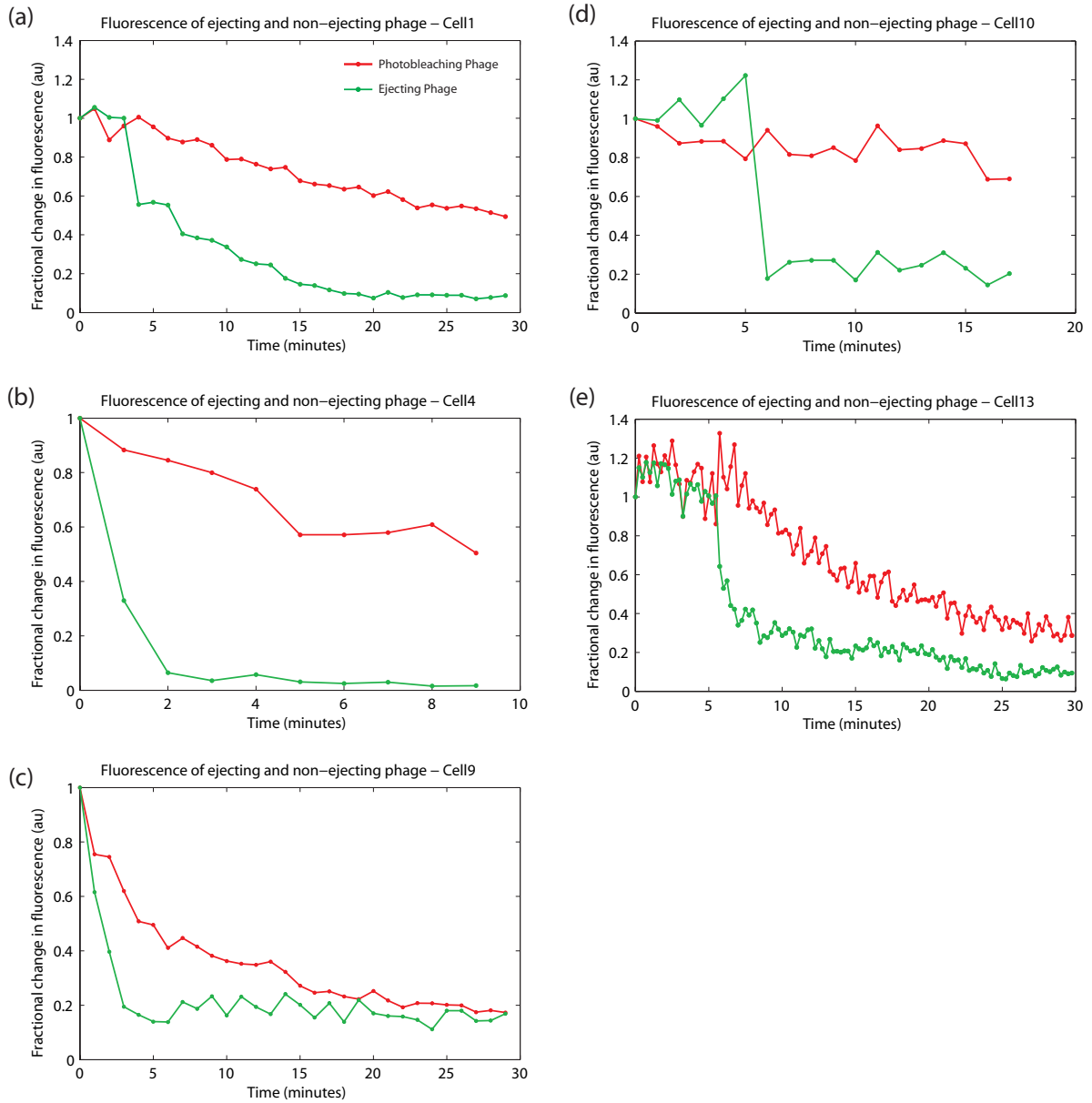


Figure 4.10: Photobleaching in SYTOX-Orange-stained bacteriophage bound to cells. Several cells had multiple stained phages that were bound simultaneously. Some of these phages did not display a significant decrease in fluorescence, indicating that they did not infect the cell. Here we plot the decrease in fluorescence in an infecting phage and a non-infecting phage in five different cells (a–e). The total fluorescence above background for each phage was determined using the segmentation masks displayed in figure 4.11. These values are normalized by the maximum observed fluorescence in each phage to allow for side-by-side comparison of the drop in fluorescence levels. Photobleaching occurs on a much longer time scale than ejections.

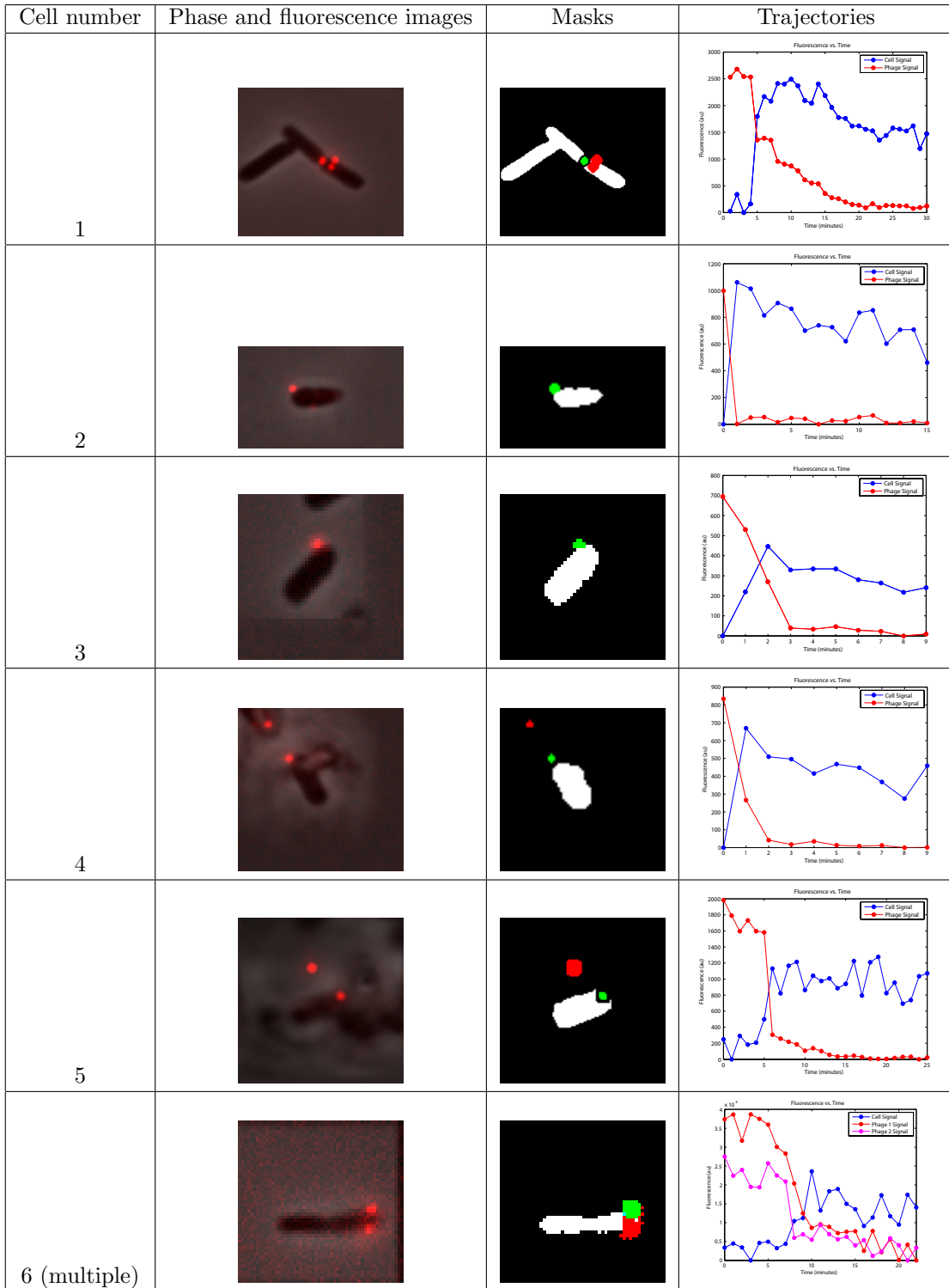


Figure 4.11: Single-molecule trajectories: Group 1. Each one of the rows in this series of images shows the phase and fluorescence image of a particular cell that has been segmented, the corresponding masks and the resulting fluorescence both from the virus-segmented region and from the whole cell. Cells with multiple ejecting phages are marked.

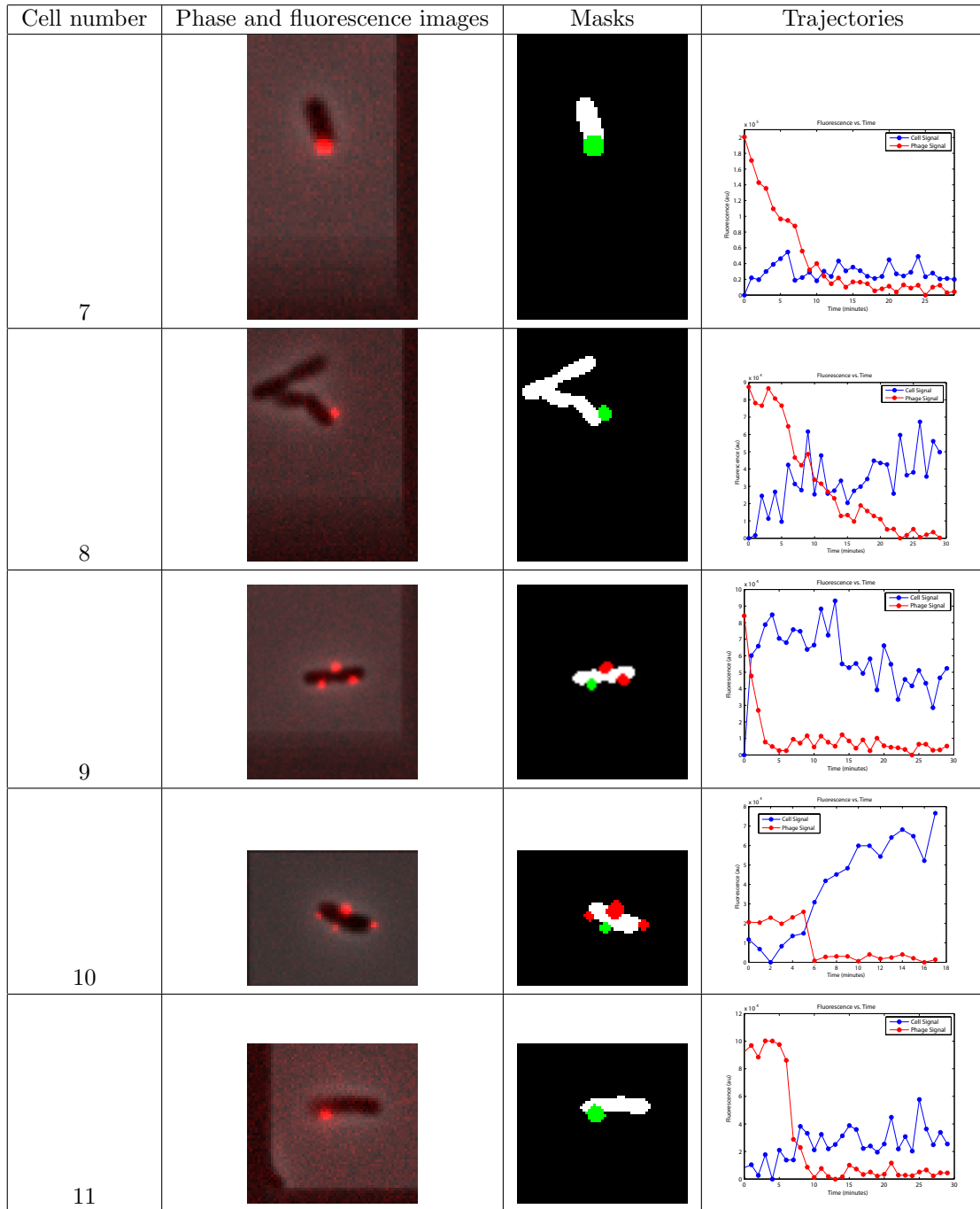


Figure 4.12: Single-molecule trajectories: Group 2. Each one of the rows in this series of images shows the phase and fluorescence image of a particular cell that has been segmented, the corresponding masks and the resulting fluorescence both from the virus-segmented region and from the whole cell. Cells with multiple ejecting phages are marked.

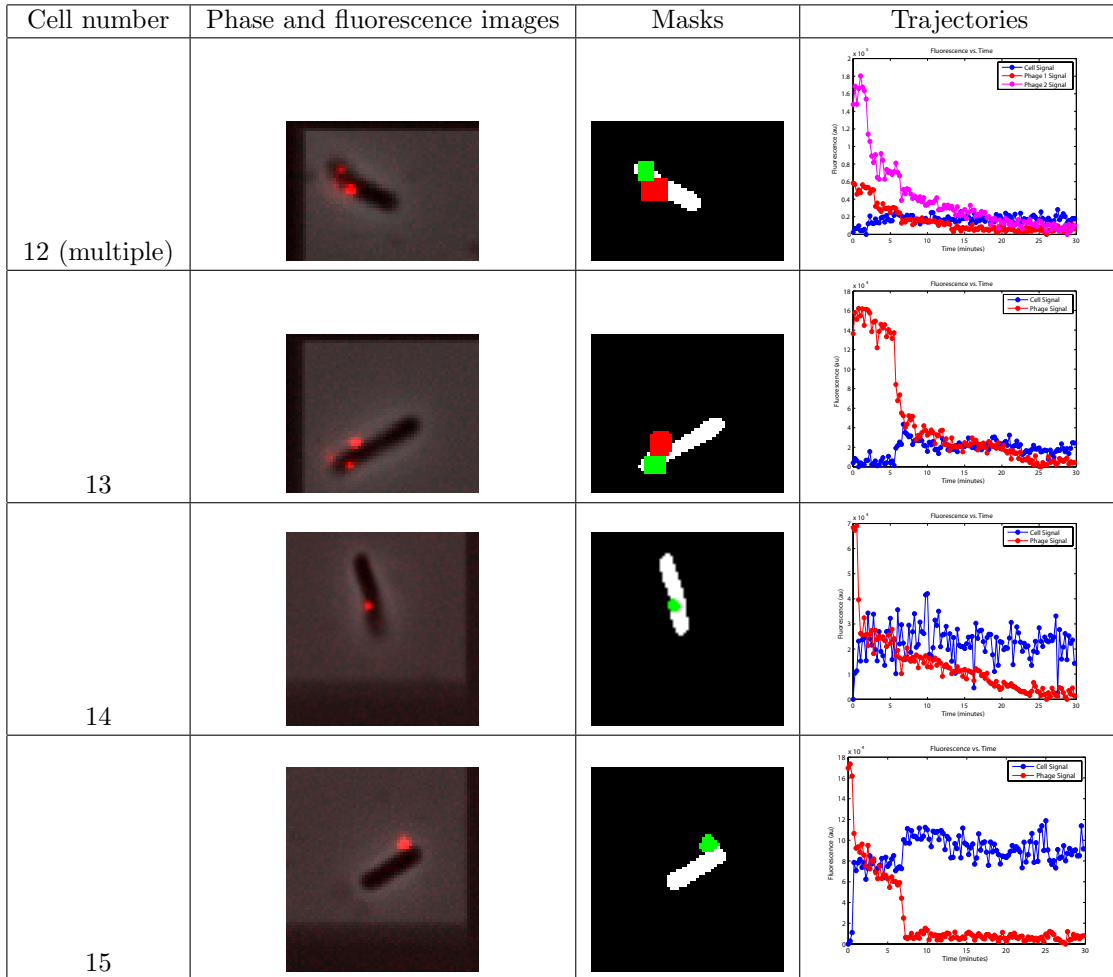


Figure 4.13: Single-molecule trajectories: Group 3. Each one of the rows in this series of images shows the phase and fluorescence image of a particular cell that has been segmented, the corresponding masks and the resulting fluorescence both from the virus-segmented region and from the whole cell. Cells with multiple ejecting phages are marked.

4.6.2 Strains, media, and controls for constructing λ DVV1

A description of the construction of phage λ DVV1 is outlined in the primary text. Here, we detail the strains that were used or constructed during this work (table 4.1), recipes for media and buffers (table 4.2), as well as controls that verify the genotype of phage λ DVV1 (table 4.3 and figures 4.14 and 4.15).

Table 4.1: Strain and plasmids that were used or constructed during the course of our study.

Strain, Phages, or Plasmid	Description
LE392	"Wild-type" <i>E. coli</i> that carries the supF amber suppression phenotype, allowing Sam7 phages to undergo lysis
λ cI60	A lambda phage with a clear plaque morphology mutation in the cI gene
λ eyfp	A lysogen with YFP fused to gpD [137]
λ eyfp + pKD46	Strain λ eyfp harboring a plasmid that expresses the λ Red system that allows the strain to undergo homologous recombination
LE392 + pAW61	Strain LE392 harboring a plasmid that expresses mCherry-ParB under the control of a pBAD promotor
LE392 + pPlate*D	Strain LE392 harboring a plasmid that produces the wild-type gpD protein. This plasmid overcomes the defects in capsid assembly seen in λ eyfp [80]. This strain was used to both propagate and titer YFP labeled phage.
λ DVV1	A phage with YFP fused to gpD and a P1 $parS$ sequence inserted into the b region. The strain was lysogenized into LE392+pPlate*D.
pALA1073	A plasmid harboring the P1 $parS$ sequence

Table 4.2: Recipes for various media and buffers used in our study.

Media	Description
NZYM	10g NZ amine, 5g NaCl, 5g Bacto-yeast extract, 2g MgSO ₄ ·7H ₂ O, 1 L milliQ water, autoclave
LB	10g Bacto-tryptone, 5g yeast extract, 10g NaCl, 1 L milliQ water, autoclave
LBM	LB + 10 mM MgSO ₄ ·7H ₂ O
M9sup	1X M9 salts, 1 mM thiamine hydrochloride, 0.4% glycerol, 0.2% casamino acids, 2 mM MgSO ₄ , 0.1 mM CaCl ₂
M9maltose-sup	1X M9 salts, 1 mM thiamine hydrochloride, 0.4% maltose, 0.2% casamino acids, 2 mM MgSO ₄ , 0.1 mM CaCl ₂
SM	5.8 g NaCl, 2 g MgSO ₄ ·7H ₂ O, 50 mL 1 M Tris-Cl, pH 7.5, 2 % w/v gelatin, milliQ water to 1 L, autoclave
TM	50 mM Tris-HCl (pH 7.4), 10 mM MgSO ₄
GODCAT	50 mg glucose oxidase, 111 μ L catalase, 500 μ L TM

Table 4.3: Primers used for engineering λ DVV1 by recombineering. Portions of the recombineering primers that carry homology to the lambda genome are in lower case and highlighted in red. The primers used for our colony PCR screen for successful insertions are also given.

Primer	Sequence
parS left	5'- aggcagcaaaatcatcagaaacgaacgcatcatcaagtgccggtcgtgca CACAGCTTTAGAGCGTTTTGCGAT -3'
parS right	5'- ccgtatccttcaccagcgtgtgccgttccaacttctgatattccctcc CGATAAAAAGCCGAAGCCTTAAAC -3'
parS integration check left	5' - AGATGCGGTGGCCTCAAAAAG - 3'
parS integration check right	5' - CCGGAACAGTTTTTCTGCTT - 3'

4.7 Acknowledgments

We are grateful to a number of people for help with experiments, advice, and critical commentary on the manuscript including: Heun Jin Lee, Maja Bialecka, Phillips lab, Talia Weiss, Vilawain Fernandes, Kari Barlan, Paul Grayson, Ido Golding, Lanying Zeng, Bill Gelbart, Chuck Knobler, Francois St. Pierre, and Drew Endy. We are also grateful to Ron Vale, Tim Mitchison, Dyche Mullins, and Clare Waterman as well as several generations of students from the MBL Physiology Course where this work has been developed over several summers. We also gratefully acknowledge financial support from several sources, including a NIH Medical Scientist Training Program Fellowship, a Yaser Abu-Mostafa Hertz Fellowship, and a NIH Director's Pioneer Award. We also acknowledge the support of NSF grant number 0758343.

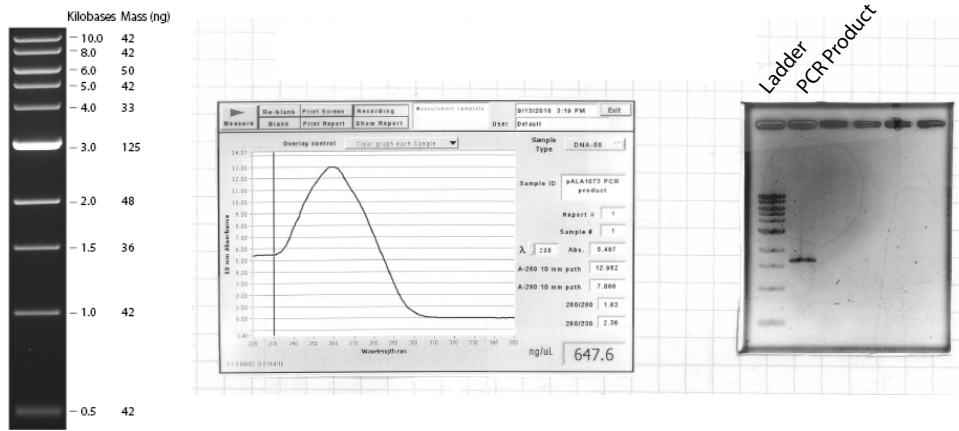


Figure 4.14: PCR amplification of the *P1parS* site on plasmid pALA1073 with primers containing homology to the b region of the lambda genome. The PCR product was purified from the host plasmid by DpnI digestion, and then used for recombineering. The expected insert size is ~ 1500 bp, which is confirmed by the gel.

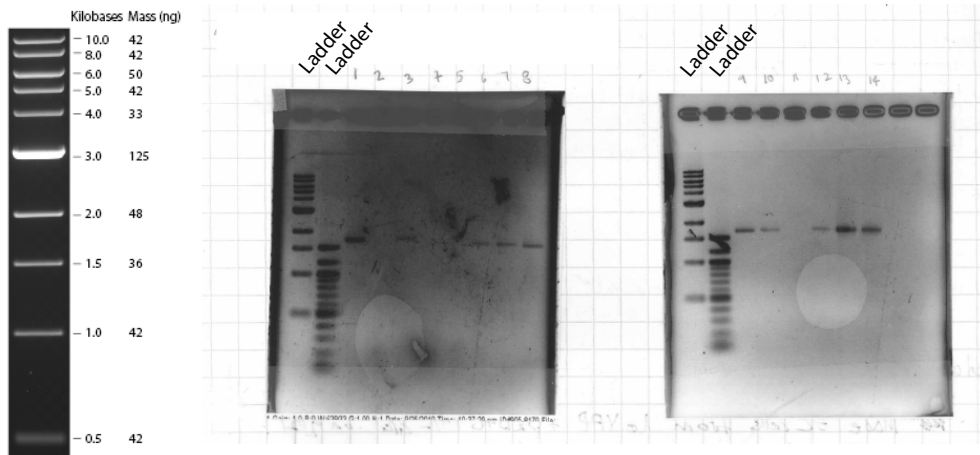


Figure 4.15: Colony PCR screen for the insertion of a *parS* site into the b region of the lambda genome. Single colonies from the selection step in the recombineering protocol were screened for the presence of a *P1parS* site using colony PCR. The PCR product is ~ 1702 bp with the insert and ~ 1988 bp without. Every colony that survived the Kanamycin selection step gave a positive result with the colony PCR screen.

```

Alignments
-----
>lcl|55993
Length=8668

Score = 1434 bits (776), Expect = 0.0
Identities = 868/929 (94%), Gaps = 9/929 (0%)
Strand=Plus/Plus

Query  41  CGATAAAAAGCCGAAG-CTTAAACTTTCGCCATTCAAATTTCACTATTAAGTACTGACTGTTT  99
      |||
Sbjct 2453 CGATAAAAAGCCGAAGCCTTAAACTTTCGCCATTCAAATTTCACTATTAAGTACTGACTGTTT 2512

Query 100  TTAAAGTAAATTACTCTAAAATTTCAAGGCGAAATCGCCACGATTCACCTTGGATTTTA  159
      |||
Sbjct 2513 TTAAAGTAAATTACTCTAAAATTTCAAGGTGAAATCGCCACGATTCACCTTGGATTTTA 2572

Query 160  CCTTCCTCCCCTCCTCCCgaaaaataaaaaaTTGCTTGTACGAGAAAGTCAACAAG  219
      |||
Sbjct 2573 CCTTCCTCCCCTCCTCCCgaaaaataaaaaaTTGCTTGTACGAGAAAGTCAACAAG 2632

Query 220  TGACTTTCAATAAAATCTCTCCGAAAAGGGATTACACAAAGTGCC-TTGTGTTAAGGA  278
      |||
Sbjct 2633 TGACTTTCAATAAAATCTCTCCGAAAAGGGATTACACAAAGTGCCCTTGTGTTAAGGA 2692

Query 279  AGAGTAAATTGAGTAACTTACGCGAATACCAGAATCGTATTGCAGATGACCTGCAGggggg  338
      |||
Sbjct 2693 AGAGTAAATTGAGTAACTTACGCGAATACCAGAATCGTATTGCAGATGACCTGCAGGGGG 2752

```

Figure 4.16: Sequencing results for the insertion of a parS site into the b region of the lambda genome. DNA generated from the colony PCR screen were sent for sequencing. Every colony came back positive. The results displayed here show that for this colony, there is 94% sequence identity between the region amplified from the lambda genome by colony PCR and the sequence of plasmid pALA1073.

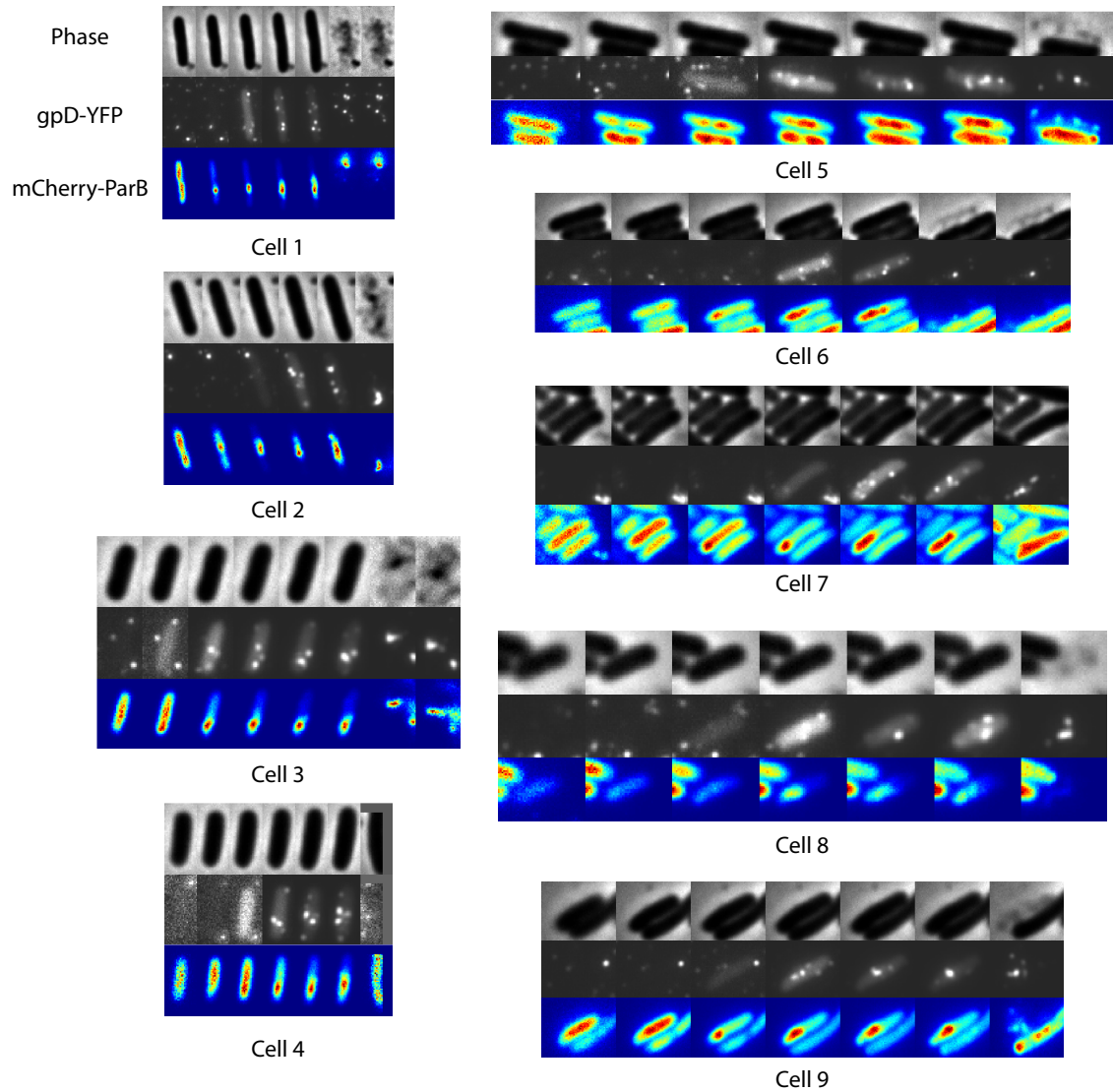


Figure 4.17: Montages of the phage lifecycle monitored with λ DVV1. Montages of phase images, gpD-YFP, and mCherry-ParB for 9 different cells are displayed.

Bibliography

- [1] M. Spivak. *Calculus On Manifolds*. Westview Press, 1971.
- [2] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. The bacteriophage straight phi29 portal motor can pack DNA against a large internal force. *Nature*, 413:748–752, 2001.
- [3] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Essential Cell Biology*. Garland Science, 2009.
- [4] S. Romero, C. Le Clainche, D. Didry, C. Egile, D. Pantaloni, and M. F. Carlier. Formin is a processive motor that requires profilin to accelerate actin assembly and associated atp hydrolysis. *Cell*, 119:419–429, 2004.
- [5] D. R. Kovar, E. S. Harris, R. Mahaffy, H. Higgs, and T. Pollard. Control of the assembly of ATP- and ADP-actin by formins and profilin. *Cell*, 124:423–435, 2006.
- [6] D. Vavylonis, D. Kovar, B. O’Shaughnessy, and T. Pollard. Model of formin-associated actin filament elongation. *Mol Cell*, 21:455–466, 2006.
- [7] J. Nguyen and W. Lim. How Src exercises self-restraint. *Nat Struct Biol*, 4:256–260, 1997.
- [8] S. Hubbard. Src autoinhibition: let us count the ways. *Nature Structural Biology*, 6(8):711–714, 1999.
- [9] S. Gonfloni, J. Williams, K. Hattula, A. Weijland, R. K. Wierenga, and G. Superti-Furga. The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *The EMBO Journal*, 16:7261–7271, 1997.
- [10] M. Young, S. Gonfloni, G. Superti-Furga, B. Roux, and J. Kuriyan. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell*, 105:115–126, 2001.
- [11] V. Krishnamurthy, V. Semetey, P. Bracher, N. Shen, and G. Whitesides. Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J Am Chem Soc*, 129:1312–1320, 2007.

- [12] W. Zagotta, T. Hoshi, and W. Aldrich. Gating of single Shaker potassium channels in *Drosophila* muscle and in *xenopus* oocytes injected with Shaker mRNA. *Proc Natl Acad Sci USA*, 86:7243–7247, 1989.
- [13] W. Zagotta and R. Aldrich. Voltage-dependent gating of Shaker A-type potassium channels in *Drosophila* muscle. *J Gen Physiol*, 95:29–60, 1990.
- [14] T. Hoshi, W. Zagotta, and W. Aldrich. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science*, 250:533–538, 1990.
- [15] W. Zagotta, T. Hoshi, and W. Aldrich. Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB. *Science*, 250:568–571, 1990.
- [16] L. Timpe and L. Peller. A random flight chain model for the tether of the Shaker K⁺ channel inactivation domain. *Biophys J*, 69:2415–2418, 1995.
- [17] B. Windisch, D. Bray, and T. Duke. Balls and chains — a mesoscopic approach to tethered protein domains. *Biophys J*, 91:2383–2392, 2006.
- [18] R. P. Bhattacharyya, A. Remenyi, B. J. Yeh, and W. A. Lim. Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual Review of Biochemistry*, 75:655–680, 2006.
- [19] K. Y. Choi, B. Satterberg, D. M. Lyons, and E. A. Elion. Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in *s. cerevisiae*. *Cell*, 78:499–512, 1994.
- [20] R. P. Bhattacharyya, A. Remenyi, M. C. Good, C. J. Bashor, A. M. Falick, and W. A. Lim. The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science*, 311:822–826, 2006.
- [21] M. Good, G. Tang, J. Singleton, A. Remenyi, and W. A. Lim. The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation. *Cell*, 136:1085–1097, 2009.
- [22] C. R. Robinson and R. T. Sauer. Equilibrium stability and sub-millisecond refolding of a designed single-chain Arc repressor. *Biochemistry*, 35:13878–13884, 1996.
- [23] C. Robinson and R. Sauer. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *Proc Natl Acad Sci USA*, 95:5929–5934, 1998.
- [24] A. Nagi and L. Regan. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des*, 2:67–75, 1997.

- [25] D. Na, T. Y. Kim, and S. Y. Lee. Construction and optimization of synthetic pathways in metabolic engineering. *Curr Opin Microbiol*, 13:363–370, 2010.
- [26] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, Prather K. L. J., and J. D. Keasling. Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology*, 27(8):753–759, 2009.
- [27] T. S. Moon, J. E. Dueber, E. Shiue, and K. Prather. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metabolic Engineering*, 12:298–305, 2010.
- [28] J. Dueber, B. Yeh, K. Chak, and W. Lim. Reprogramming control of an allosteric signaling switch through modular recombination. *Science*, 30:1904–1908, 2003.
- [29] H. Higgs, L. Blanchoin, and T. Pollard. Influence of the C-terminus of Wiskott-Aldrich syndrome protein (WASP) and the Arp2/3 complex on actin polymerization. *Biochemistry*, 38:15212–15222, 1999.
- [30] L. Machesky and R. Insall. Scar1 and the related Wiskott-Aldrich syndrome protein, WASP, regulate the actin cytoskeleton through the Arp2/3 complex. *Curr Biol*, 8:1347–1356, 1998.
- [31] J. B. Marchand, D. Kaiser, T. Pollard, and H. Higgs. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nat Cell Biol*, 3:76–82, 2001.
- [32] R. D. Mullins and L. Machesky. Actin assembly mediated by Arp2/3 complex and WASP family proteins. *Methods Enzymol*, 325:214–237, 2000.
- [33] K. Prehoda, J. Scott, R. D. Mullins, and W. Lim. Integration of multiple signals through cooperative regulation of the N-WASP–Arp2/3 complex. *Science*, 290:801–806, 2000.
- [34] R. Rohatgi, L. Ma, H. Miki, M. Lopez, T. Kirchhausen, T. Takenawa, and M. Kirschner. The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly. *Cell*, 97:221–231, 1999.
- [35] M. Dayel and R. D. Mullins. Activation of Arp2/3 complex: addition of the first subunit of the new filament by a WASP protein triggers rapid ATP hydrolysis on Arp2. *PLoS Biol*, 2:476–485, 2004.
- [36] A. Kim, L. Kakalis, N. Abdul-Manan, G. Liu, and M. Rosen. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature*, 404:151–158, 2000.
- [37] M. Ptashne and A. Gann. *Genes & Signals*. Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York), 2004.

- [38] A. Hochschild and M. Ptashne. Interaction at a distance between λ repressors disrupts gene activation. *Nature*, 336:353–357, 1988.
- [39] J. K. Joung, L. U. Le, and A. Hochschild. Synergistic activation of transcription by *Escherichia coli* cAMP receptor protein. *Proc Natl Acad Sci USA*, 90:3083–3087, 1993.
- [40] J. K. Joung, D. M. Koepp, and A. Hochschild. Synergistic activation of transcription by bacteriophage λ cI protein and *E. coli* cAMP receptor protein. *Science*, 265:1863–1866, 1994.
- [41] I. B. Dodd, K. E. Shearwin, A. J. Perkins, T. Burr, A. Hochschild, and J. B. Egan. Cooperativity in the long-range gene regulation by the λ ci repressor. *Genes Dev*, 18:344–354, 2004.
- [42] G. Ackers, A. Johnson, and M. Shea. Quantitative model for gene regulation by λ phage repressor. *Proc Natl Acad Sci USA*, 79:1129–1133, 1982.
- [43] I. Graham and T. Duke. The logical repertoire of ligand-binding proteins. *Phys Biol*, 2:159–165, 2005.
- [44] M. Buck, W. Xu, and M. Rosen. A two-state allosteric model for autoinhibition rationalizes WASP signal integration and targeting. *J Mol Biol*, 338:271–285, 2004.
- [45] H. X. Zhou. Quantitative relation between intermolecular and intramolecular binding of pro-rich peptides to SH3 domains. *Biophys J*, 91:3170–3181, 2006.
- [46] S. Gianni, A. Engstrom, M. Larsson, N. Calosci, F. Malatesta, L. Eklund, C. Ngang, C. Travaglini-Allocatelli, and P. Jemth. The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. *J Biol Chem*, 280:34805–34812, 2005.
- [47] R. Hudgins, F. Huang, G. Gramlich, and W. Nau. A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides. *J Am Chem Soc*, 124:556–564, 2002.
- [48] X. Wang, E. Bodunov, and W. Nau. Fluorescence quenching kinetics in short polymer chains: Dependence on chain length. *Opt Spectrosc*, 95:560–570, 2003.
- [49] H. Jacobson and W. Stockmayer. Intramolecular reaction in polycondensations. *I. The Theory of Linear Systems*, 18:1600–1606, 1950.
- [50] K. Ito, J. Chuang, C. Alvarez-Lorenzo, T. Watanabe, N. Ando, and A. Grosberg. Multiple point adsorption in a heteropolymer gel and the tanaka approach to imprinting: experiment and theory. *Prog Polym Sci*, 28:1490–1515, 2003.

- [51] J. Samuel and S. Sinha. Elasticity of semiflexible polymers. *Phys Rev E*, 66, 2002.
- [52] M. Rief, M. Gautel, F. Oesterhelt, J. Fernandez, and H. Gaug. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276:1109–1112, 1997.
- [53] M. Rief, J. Fernandez, and H. Gaub. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys Rev Lett*, 81:4764–4767, 1998.
- [54] R. Lim, N. P. Huang, J. Koser, J. Deng, K. H. A. Lau, K. Schwarz-Herion, B. Fahrenkrog, and U. Aebi. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc Natl Acad Sci USA*, 103:9512–9517, 2006.
- [55] H. X. Zhou. Loops in proteins can be modeled as worm-like chains. *J Phys Chem B*, 105:6763–6766, 2001.
- [56] P. Flory. *Statistical Mechanics of Chain Molecules*. John Wiley and Sons, Inc. (New York, New York), 1969.
- [57] H. Seyed-allaei. Three-bead rotating chain model shows universality in the stretching of proteins. *Phys Rev E*, 72:041908, 2005.
- [58] D. Segall, P. Nelson, and R. Phillips. Excluded-volume effects in tethered particle experiments: Bead size matters. *Phys Rev Lett*, 96:088306, 2006.
- [59] D. Chereau, F. Kerff, P. Graceffa, Z. Grabarek, K. Langsetmo, and R. Dominguez. Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc Natl Acad Sci USA*, 102, 2005.
- [60] R. Robinson, K. Turbedsky, D. Kaiser, J. B. Marchand, H. Higgs, S. Choe, and T. Pollard. Crystal structure of arp2/3 complex. *Science*, 294:1679–1684, 2001.
- [61] L. Bintu, N. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*, 15:116–124, 2005.
- [62] X. Wu, B. Knudsen, S. Feller, J. Zheng, A. Sali, D. Cowburn, H. Hanafusa, and J. Kuriyan. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure*, 3:215–226, 1995.
- [63] J. Dueber, E. Mirsky, and W. Lim. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol*, 25:660–662, 2007.

- [64] G. Posern et al. Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G. *Oncogene*, 16:1903–1912, 1998.
- [65] B. Knudsen, J. Zheng, S. Feller, J. Mayer, S. Burrell, D. Cowburn, and H. Hanafusa. Affinity and specificity requirements for the first Src homology 3 domain of the Crk proteins. *EMBO J*, 14:2191–2198, 1995.
- [66] J. Nguyen, C. Turck, F. Cohen, R. Zuckerman, and W. Lim. Exploiting the basis of proline recognition by SH3 and WW domains: Design of N-substituted inhibitors. *Nature*, 282:2088–2092, 1998.
- [67] M. Ward, J. Wu, and Y. Rao. Visualization of spatially and temporally regulated N-WASP activity during cytoskeletal reorganization in living cells. *Proc Natl Acad Sci USA*, 101:970–974, 2004.
- [68] M. Smoluchowski. *Phys Z*, 17:557, 1916.
- [69] J. Zalvsky, L. Lempert, H. Kranitz, and R. D. Mullins. Different WASP family proteins stimulate different Arp2/3 complex-dependent actin-nucleating activities. *Curr Biol*, 11:1903–1913, 2001.
- [70] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *J Chem Phys*, 72:4350–4357, 1980.
- [71] S. Jun, J. Bechhoefer, and B. Ha. Diffusion-limited loop formation of semiflexible polymers: Kramers theory and the intertwined time scales of chain relaxation and closing. *Europhys Lett*, 64:420–426, 2003.
- [72] C. Hyeon and D. Thirumalai. Kinetics of interior loop formation in semiflexible chains. *J Chem Phys*, 124, 2006.
- [73] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*. World Scientific (River Edge, NJ), 2004.
- [74] A. D. Hershey and M. M. Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Phys*, 36:39–56, 1952.
- [75] G. P. Smith. Surface presentation of protein epitopes using bacteriophage expression systems. *Curr Opin Biotechnol*, 2:668–673, 1991.
- [76] C. E. Catalano, D. Cue, and M. Feiss. Virus DNA packaging: the strategy used by phage lambda. *Curr Opin Biotechnol*, 16:1075–1086, 1995.

- [77] A. M. Oppenheim, O. Koniler, J. Stavans, D. L. Court, and S. Adhya. Switches in bacteriophage lambda development. *Ann Rev Gen*, 39:409–429, 2005.
- [78] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Ann Rev Gen*, 149:1633–1648, 1998.
- [79] F. St-Pierre and D. Endy. Determination of cell fate selection during phage lambda infection. *Proc Natl Acad Sci USA*, 105:20705–20710, 2008.
- [80] S. O. Zeng, L. Skinner, C. Zong, J. Sippy, M. Feiss, and I. Golding. Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell*, 141:682–691, 2010.
- [81] I. B. Dodd, K. E. Shearwin, and J. B. Egan. Revisited gene regulation in bacteriophage lambda. *Curr Opin Gen Dev*, 15:145–152, 2005.
- [82] D. N. Fuller, J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson, and D. E. Smith. Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi29. *Proc Natl Acad Sci USA*, 104:11245–11250, 2007.
- [83] S. L. Novick and J. D. Baldeschwieler. Fluorescence measurement of the kinetics of DNA injection by bacteriophage lambda into liposomes. *Biochemistry*, 27:7919–7924, 1988.
- [84] S. Mangenot, M. Hochrein, J. Radler, and L. Letellier. Real time imaging of DNA ejection from single phage particles. *Curr Biol*, 15:430–435, 2005.
- [85] P. Grayson, L. Han, T. Winther, and R. Phillips. Real-time observations of single bacteriophage lambda DNA ejections in vitro. *Proc Natl Acad Sci USA*, 104:13652–13657, 2007.
- [86] N. Chiaruttini, M. de Frutos, E. Augarde, P. Boulanger, L. Letellier, and V. Viasnoff. Is the in vitro ejection of bacteriophage DNA quasi-static? A bulk to single virus study. *Biophys J*, 99:447–455, 2010.
- [87] A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud, and W. M. Gelbart. Osmotic inhibition of DNA ejection from phage. *Proc Natl Acad Sci USA*, 100:9292–9295, 2003.
- [88] A. Evilevitch, J. W. Gober, M. Phillips, C. M. Knobler, and W. M. Gelbart. Measurements of DNA lengths remaining in a viral capsid after osmotically suppressed partial ejection. *Biophys J*, 88:851–866, 2005.
- [89] P. Grayson, A. Evilevitch, M. M. Inamdar, P. K. Purohit, W. M. Gelbart, and R. Phillips. The effect of genome length on ejection forces in bacteriophage lambda. *Virology*, 348:430–436, 2006.

- [90] A. Leforestier, S. Brasils, M. de Frutos, E. Raspaud, L. Letellier, P. Tavares, and F. Livolant. Bacteriophage T5 DNA ejection under pressure. *J Mol Biol*, 384:730–739, 2008.
- [91] J. Kindt, S. Tzllil, A. Ben-Schaul, and W. M. Gelbart. DNA packaging and ejection forces in bacteriophage. *Proc Natl Acad Sci USA*, 98:13671–13674, 2001.
- [92] M. Jeembaeva, M. Castelnovo, F. Larsson, and A. Evilevitch. Osmotic pressure: resisting or promoting DNA ejection from phage? *J Mol Biol*, 381:310–323, 2008.
- [93] J. B. Stock, B. Rauch, and S. Roseman. Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. *J Biol Chem*, 252:7850–7860, 1977.
- [94] I. J. Molineux. Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? *Virology*, 344:221–229, 2006.
- [95] L. R. Comolli, A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, C. Bustamante, and K. Downing. Three-dimensional architecture of the bacteriophage phi29 packaged genome and elucidation of packaging process. *Virology*, 371:267–277, 2008.
- [96] A. Leforestier and F. Livolant. Structure of toroidal DNA collapsed inside the phage capsid. *Proc Natl Acad Sci USA*, 106:9157–9162, 2009.
- [97] A. Leforestier and F. Livolant. The bacteriophage genome undergoes a succession of intracapsid phase transitions upon DNA ejection. *J Mol Biol*, 396:384–395, 2010.
- [98] A. Evilevitch. Effects of condensing agent and nuclease on the extent of ejection from phage lambda. *J Phys Chem B*, 110:22261–22265, 2006.
- [99] N. V. Hud and K. H. Downing. Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: the fine structure of DNA toroids. *Proc Natl Acad Sci USA*, 98:14925–14930, 2001.
- [100] P. A. Fang, E. T. Wright, S. T. Weintraub, K. Hakala, and W. Wu. Visualization of bacteriophage T3 capsids with DNA incompletely packaged in vivo. *J Mol Biol*, 384:1384–99, 2008.
- [101] J. E. Johnson and W. Chiu. DNA packaging and delivery machines in tailed bacteriophages. *Curr Opin Struct Biol*, 17:237–243, 2007.
- [102] A. Evilevitch, L. T. Fang, A. M. Yoffe, M. Castelnovo, D. C. Rau, V. A. Parsegian, W. M. Gelbart, and C. M. Knobler. Effects of salt concentrations and bending energy of the extent of ejection of phage genomes. *Biophys J*, 94:1110–1120, 2008.

- [103] T. Odijk. Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. *Biophys J*, 75:1223–1227, 1998.
- [104] P. K. Purohit, J. Kondev, and R. Phillips. Mechanics of DNA packaging in viruses. *Proc Natl Acad Sci USA*, 100:3173–3178, 2003.
- [105] S. Tzllil, J. T. Kindt, W. M. Gelbart, and A. Ben-Schaul. Forces and pressures in DNA packaging and release from viral capsids. *Biophys J*, 84:851–866, 2003.
- [106] P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev, and R. Phillips. Forces during bacteriophage DNA packaging and ejection. *Biophys J*, 88:1616–1627, 2005.
- [107] Z. Li, J. Wu, and Z. G. Wang. Osmotic pressure and packaging structure of caged DNA. *Biophys J*, 94:737–746, 2008.
- [108] A. S. Petrov and S. C. Harvey. Structural and thermodynamic principles of viral packaging. *Structure*, 15:21–27, 2007.
- [109] A. S. Petrov and S. C. Harvey. Packaging double-helical DNA into viral capsids: structures, forces, and energetics. *Biophys J*, 95:497–502, 2008.
- [110] B. E. Michel. Evaluation of the water potentials of solutions of polyethylene glycol 8000 both in the absence and presence of other solutes. *Plant Physiol*, 72:66–70, 1983.
- [111] J. Sambrook and D. W. Russell. *Molecular Cloning: A Laboratory Manual*, volume 95. Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York), 2001.
- [112] T. A. Keller, T. Ferenci, A. Prilipov, and J. P. Rosenbusch. Crystallization of monodisperse maltoporin from wild-type and mutant strains of various enterobacteriaceae. *Biochem Biophys Res Comm*, 199:767–771, 1994.
- [113] L. R. Brewer and P. R. Bianco. Laminar flow cells for single-molecule studies of DNA-protein interactions. *Nat Meth*, 5:517–525, 2008.
- [114] D. E. Smith, T. Perkins, and S. Chu. Laminar flow cells for single-molecule studies of DNA-protein interactions. *Macromolecules*, 29:1372–1373, 1996.
- [115] E. Joly. Purification of DNA fragments from agarose gels using glass beads. *Methods in Molecular Biology: Basic DNA and RNA Protocols*, 58, 1996.

- [116] L. Han, H. G. Garcia, S. Blumberg, K. B. Towles, J. F. Beausang, P. C. Nelson, and R. Phillips. Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One*, 4:e5621, 2009.
- [117] V. Viasnoff. personal communication. *The City of Paris Industrial Physics and Chemistry Higher Educational Institution*, 2010.
- [118] C. Sao-Jose, M. de Frutos, E. Raspaud, M. A. Santos, and P. Tavares. Pressure built by DNA packaging inside virions: enough to drive DNA ejection in vitro, largely insufficient for delivery into the bacterial cytoplasm. *J Mol Biol*, 374:346–355, 2007.
- [119] D. C. Rau, B. Lee, and V. A. Parsegian. Measurement of the repulsive force between polyelectrolyte molecules in ionic solution: hydration forces between parallel DNA double helices. *Proc Natl Acad Sci USA*, 81:2621–2625, 1984.
- [120] M. Esquinas-Rychen and B. Erni. Facilitation of bacteriophage lambda DNA injection by inner membrane proteins of the bacterial phosphoenol-pyruvate:carbohydrate phosphotransferase system (pts). *J Mol Microbiol Biotechnol*, 3:361–70, 2001.
- [121] M. Gottesman. Bacteriophage lambda: the untold story. *J Mol Biol*, 293:177–180, 1999.
- [122] J. D. Watson, G. S. Stent, and J. Cairns. *Phage and the Origins of Molecular Biology*. Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York), 2000.
- [123] H. Echols and C. Gross. *Operators and promoters: The story of molecular biology and its creators*. University of California Press (New York, New York), 2001.
- [124] J. E. Johnson. Virus particle maturation: insights into elegantly programmed nanomachines. *Curr Opin Struct Biol*, 20:210–216, 2010.
- [125] R. Edgar, A. Rokney, M. Feeney, S. Semsey, M. Kessel, M. B. Goldberg, S. Adhya, and A. B. Oppenheim. Bacteriophage infection is targeted to cellular poles. *Mol Microbiol*, 68:1107–1116, 2008.
- [126] L. Oddershede, J. K. Dreyer, S. Grego, S. Brown, and K. Berg-Sorensen. The motion of a single molecule, the lambda-receptor, in the bacterial outer membrane. *Biophys J*, 83:3152–3161, 2002.
- [127] R. Moldovan, E. Chapman-McQuiston, and X. L. Wu. On kinetics of phage adsorption. *Biophys J*, 93:303–315, 2007.
- [128] L. Garcia and I. Molineux. Rate of Translocation of Bacteriophage T7 DNA across the Membranes of *Escherichia coli*. *J Bacteriol*, 14:4066–4076, 1995.

- [129] I. Molineux. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molec Microbiol*, 40:1–8, 2001.
- [130] J. Bohm, O. Lambert, A. S. Frangakis, L. Letellier, W. Baumeister, and J. L. Rigaud. Fhva-mediated phage genome transfer into liposomes: a cryo-electron tomography study. *Curr Biol*, 11:1168–1175, 2001.
- [131] V. Gonzalez-Huici, M. Salas, and J. Hermoso. The push-pull mechanism of bacteriophage phi29 DNA injection. *Mol Microbiol*, 52(2):529–40, 2004.
- [132] D. N. Fuller, D. M. Raymer, J. P. Rickgauer, R. M. Robertson, C. E. Catalano, D. L. Anderson, S. Grimes, and D. E. Smith. Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. *J Mol Biol*, 373:1113–1122, 2007.
- [133] Y. R. Chemla, K. Aathavan, J. Michaelis, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante. Mechanism of force generation of a viral DNA packaging motor. *Cell*, 122:683–692, 2005.
- [134] M. Ptashne. *A Genetic Switch, 3rd ed.* Cold Spring Harbor Laboratory Press (Cold Spring Harbor, New York), 2004.
- [135] D. Wu, D. Van Valen, Q. Hu, and R. Phillips. Ion-dependent dynamics of DNA ejections for bacteriophage lambda. *Biophys J*, 99(4):1101–1109, 2010.
- [136] M. M. Inamdar, W. M. Gelbart, and R. Phillips. Dynamics of DNA Ejection from Bacteriophage. *Biophys J*, 91:411–420, 2006.
- [137] L. Alvarez, P. Thomen, T. Makushok, and D. Chatenay. Propagation of fluorescent viruses in growing plaques. *Biotech and Bioeng*, 96:615–621, 2007.
- [138] F. St. Pierre. *Determination of Cell Fate Selection During Phage Lambda Infection.* MIT Thesis, 2009.
- [139] M. Eriksson, M. Hardelin, A. Larsson, J. Bergenholtz, and B. Akerman. Binding of intercalating and groove-binding cyanine dyes to bacteriophage T5. *J Phys Chem B*, 111:1139–1148, 2007.
- [140] P. Mosier-Boss, S. Lieberman, J. Andrews, F. Rohwer, L. Wegley, and M. Breitbart. Use of fluorescently labeled phage in the detection and identification of bacterial species. *Appl Spec*, 57:1138–1144, 2003.
- [141] B. Brandenburg, L. Lee, M. Lakadamyli, M. Rust, X. Zhuang, and J. Hogle. Imaging poliovirus entry in live cells. *PLoS Biol*, 5:1543–1555, 2007.

- [142] Y. Li, K. Serguee, and S. Austin. The segregation of the *Escherichia coli* origin and terminus of replication. *Mol Microbiol*, 46:985–996, 2002.
- [143] R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg. *Lambda II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1983.
- [144] K. A. Datsenko and B. L. Wanner. One-step inactivation of chromosomal genes in *escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA*, 97:6640–6645, 2000.
- [145] D. Mackay and V. Bode. Events in lambda injection between phage absorption and DNA entry. *Virology*, 72:154–166, 1976.
- [146] X. Yan, R. Habbersett, T. Yoshida, J. Nolan, J. Jett, and B. Marrone. Probing the kinetics of SYTOX orange stain binding to double-stranded DNA with implications for DNA analysis. *Anal Chem*, 77:3554–3562, 2005.
- [147] E. L. Ellis and M. Delbruck. The growth of bacteriophage. *J Gen Physiol*, 22:365–384, 1939.
- [148] T. Nagai, K. Ibata, E. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol*, 1:87–90, 2002.
- [149] H. Nielsen, Y. Li, B. Youngren, F. Hansen, and S. Austin. Progressive segregation of the *escherichia coli* chromosome. *Molec Microbiol*, 61:383–393, 2006.
- [150] J. S. Rigden. *Hydrogen: The essential element*. Harvard University Press (Cambridge, MA), 2002.
- [151] R. J. Davenport, G. J. Wuite, R. Landick, and C. Bustamante. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science*, 287:2497–2500, 2000.
- [152] S. L. Gotta, O. L. Miller, and S. L. French. rRNA transcription rate in *Escherichia coli*. *J Bacteriol*, 173:6647–6649, 1991.
- [153] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer. *In vivo* dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol*, 14:796–806, 2007.
- [154] M. M. Inamdar, W. M. Gelbart, and R. Phillips. Dynamics of DNA ejection from bacteriophage. *Biophys J*, 91:411–420, 2006.
- [155] S. M. Simon, C. S. Peskin, and G. F. Oster. What drives the translocation of proteins? *Proc Natl Acad Sci USA*, 89(9):3770–3774, 1992.

- [156] C. S. Peskin, G. M. Odell, and G. F. Oster. Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys J*, 65(1):316–324, 1993.
- [157] P. Grayson and I. J. Molineux. Is phage DNA "injected" into cells — biologists and physicists can agree. *Curr Opin Microbiol*, 10(4):401–409, 2007.
- [158] A. W. Coleman, M. J. Maguire, and J. R. Coleman. Mithramycin- and 4'-6-diamidino-2-phenylindole (DAPI)-DNA staining for fluorescence microspectrophotometric measurement of DNA in nuclei, plastids, and virus particles. *J Histochem Cytochem*, 29(8):959–968, 1981.
- [159] J. P. Martin and N. Logsdon. Oxygen radicals mediate cell inactivation by acridine dyes, fluorescein, and lucifer yellow CH. *Photochem Photobiol*, 46(1):45–53, 1987.