The Mechanical Genome in Regulation and Infection

Thesis by

Yi-Ju Chen

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 5, 2015)

To my family.

Acknowledgements

Abstract

Biological information storage and retrieval is a dynamic process that requires the genome to undergo dramatic structural rearrangements. Recent advances in single-molecule techniques have allowed precise quantification of the nano-mechanical properties of DNA [1, 2], and direct *in vivo* observation of molecules in action [3]. In this work, we will examine elasticity in protein-mediated DNA looping, whose structural rearrangement is essential for transcriptional regulation in both prokaryotes and eukaryotes. We will look at hydrodynamics in the process of viral DNA ejection, which mediates information transfer and exchange and has prominent implications in evolution. As in the case of Kepler's laws of planetary motion leading to Newton's gravitational theory, and the allometric scaling laws in biology revealing the organizing principles of complex networks [4], experimental data collapse in these biological phenomena has guided much of our studies and urged us to find the underlying physical principles.

Publications

This thesis is based on the following publications:

(i) A Single-Molecule Hershey-Chase Experiment. Van Valen D, Wu D, Chen YJ, Tuson H, Wiggins P, Phillips R, 2012 Curr Biol 22:133943.

(ii) Poly(dA:dT)-rich DNAs are Highly Flexible in the Context of DNA Looping. Johnson S*, Chen YJ*, Phillips R, 2013 PLoS One 8(10):e75799.

(iii) Modulation of DNA loop lifetimes by the free energy of loop formation. Chen YJ*, Johnson S*, Mulligan P*, Spakowitz A, Phillips R, 2014 Proc Natl Acad Sci USA 111(49):17396401.

(iv) Interplay of Protein Binding Interactions, DNA Mechanics, and Conformational Entropy in DNA Looping Kinetics. Mulligan P, Chen YJ, Phillips R, Spakowitz A, 2015 submitted.

(v) Effects of host-cell environment on bacteriophage DNA injection. Kegel W, Chen YJ, Wu D, Gelbart W, Phillips R, 2015 in preparation.

Chapter 2 and Chapter 3 are based on publications (ii), (iii), and (iv). Chapter 4 represents (i), (v), and preliminary results from a follow-up experiment to (i), designed and conducted in collaboration with Timur Zhiyentayev and Long Cai.

Contents

List of Figures

List of Tables

Chapter 1

Introduction: Scalings in the Genomes and the Microbiomes

Often, some of the most fundamental clues to how a given phenomenon works come from experimentally observed scaling behaviors. Examples range from the relation of period and size of planetary orbits to the universal critical exponents of continuous phase transitions. The regularities that exist across many scales strongly suggest the form of interactions and the structure of the organizations. This need not be common in biology, as divergent evolution, extreme environments, and biological diversity are much appreciated. On the other hand, biology is inevitably subject to physical constraints, and improvements in quantitative measurements spanning a diverse range of biological entities have helped reveal similar behaviors in their information capacity, feedback control, pattern formation, and transport phenomena. A prominent class of examples is the allometric scaling laws that emerge from the fractal-like distribution networks, which are fundamental to the body plan of many organisms [4–6].

In the world of the microbes (Fig. 1.1(a) shows a collection of microbes with different morphologies) and their genomes, which are subjects of our studies, both the growth rate and the information content (genome length) are found to be scaling with the body mass of the organisms (see Fig. 1.1(c) and (d)), revealing principles governing their construction and information processing ability [7, 8]. Zooming in to the level of individual bacteria (Fig. 1.1(b) shows bacterial viruses infecting a single bacterium), several recent studies [3, 9–11] observed interesting biophysical properties of the bacterial cytoplasm, that macromolecules diffuse according to a slower time-dependency than random walks (see Fig. 1.1(e) for an example). In our own works studying the physical genome both *in vitro* and in bacteria, scalings in the data have also been very useful. In chapter 3 and chapter 4, we will discuss how the scalings helped identify an unexplored aspect in DNA looping mechanism (Fig. 1.1(f)), and determine the control parameters of bacteriophage ejection (Fig. 1.1(g)).

The organization of this thesis is as follows: Chapter 2 presents our *in vitro* study of DNA mechanics, chapter 3 examines the kinetics of protein-mediated DNA looping, and chapter 4 aims to look at the bacteriophage genome during the infection process. Finally, in chapter 5 we conclude by listing several open

questions related to the systems we study.

Our quest started from an observation pointed out by a series of studies [12–14] regarding the regulation of gene expression in eukaryotic cells, that DNA sequence plays a key role in controlling the transcription factor or RNA polymerase accessibility to the nucleosomes where the DNA wraps around, therefore the DNA sequence may serve as a secondary "genomic code" that instructs nucleosome distribution along the genome, reducing the gene expression activity of the regions with high nucleosomal density. The hypothesis was that DNA deformability, which contributes to the free energy of a DNA-wrapped nucleosome, is determined by its sequence. We tested this idea by studying the same nucleosome positioning sequences in a bacterial protein-mediated DNA looping motif (the *E. coli lac* operon) which we could assay the deformability by the tethered particle motion experimental technique. This result forms the basis of chapter 2. From this data, we also carefully studied the dynamics of the looping and unlooping events in chapter 3. Previous studies on this subject have not systematically examined the possible factors affecting the dynamics [15, 16], and we aimed to develop a theoretical framework for the kinetics and to integrate the results from tuning the relevant knobs, such as protein concentration, protein-DNA binding potential, and DNA length and sequence, into our kinetics theory.

We then turned our attention from the DNA mechanics of bacterial gene regulation to the infection dynamics of bacterial virus. Specifically, we study the tailed bacteriophage with icosahedral head and linear double-stranded DNA as its genome (in the order Caudovirales), which represents the oldest and most abundant bacteriophage type. This type of phage is unique for its high DNA packing density in the head (about 50% volume ratio), that requires ATP-consuming motors during its assembly process in the cell, and has been estimated to exhibit 60 *atm* of pressure on the phage capsid [17–20]. It was then proposed that bacteriophage could use this stored free energy in its compressed DNA as a mechanism for injection into the bacterium, and several *in vitro* studies have confirmed this idea in bacteriophage Lambda [21–24], while other mechanisms have also been observed [25, 26] in different phage species. Chapter 4 represents our attempts to extend the previous studies to measure the *in vivo* bacteriophage Lambda ejection dynamics at the single virus level, as well as the post-infection growth of the viral genomes.
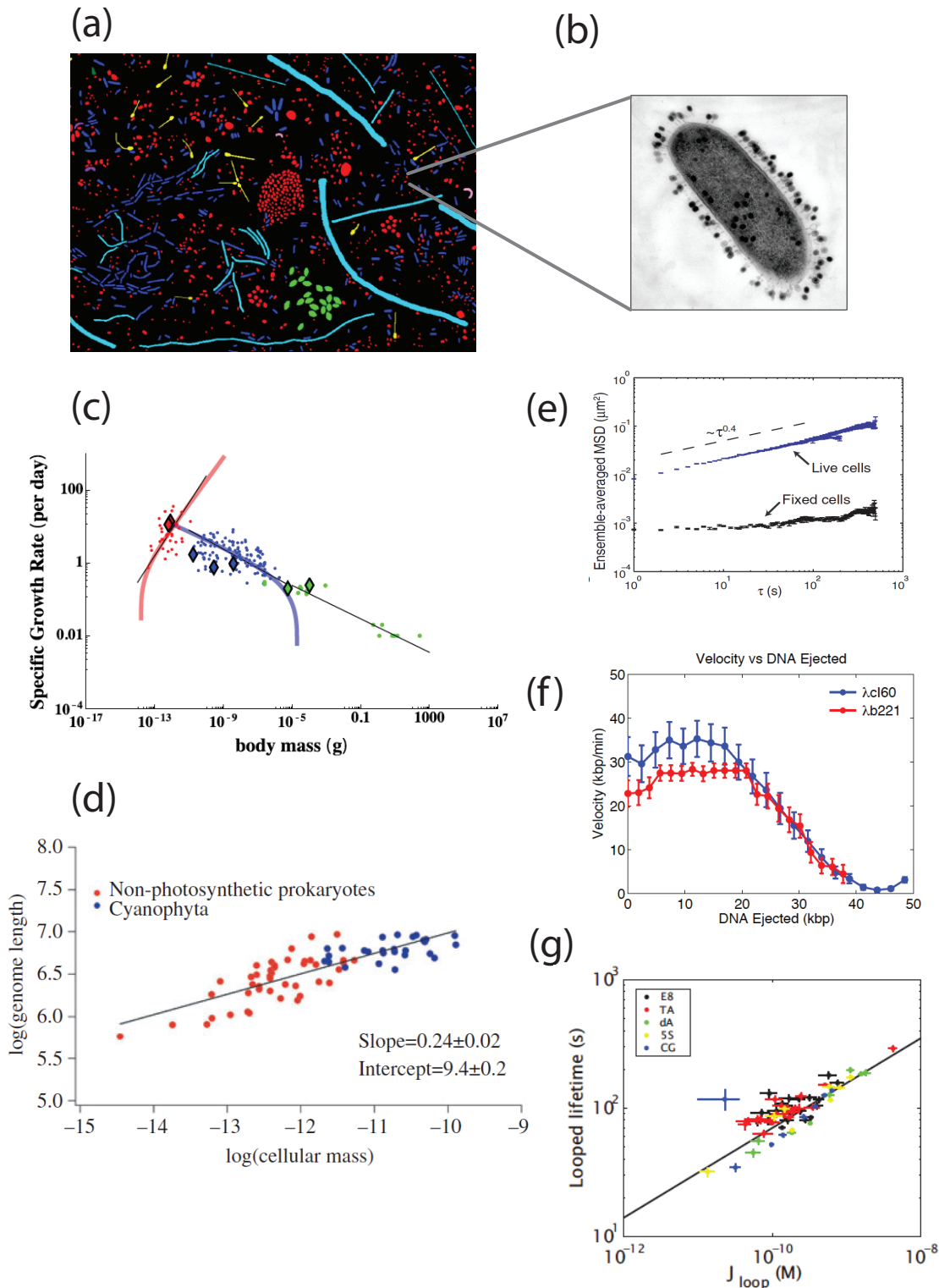
Figure 1.1: **Scalings in the microbes and their genomes. (a),** a microbial community, classified and pseudo-colored according to the cell morphologies. Adapted from reference [27]. **(b),** a transmission electron micrograph of an *E. coli* cell infected with bacteriophage T4 particles. Adapted from reference [28], courtesy of John Wertz. **(c),** growth rate of prokaryotes (colored red), eukaryotes (blue), and small metazoans (green), plotted with respect to body mass. Adapted from reference [8]. **(d),** genome length of microbes as a function of the cellular mass. Adapted from reference [7]. **(e),** mean-square-displacement of genomic loci in live *E. coli* cell scales sub-diffusively with time. Adapted from reference [10]. **(f),** bacteriophage Lambda *in vivo* ejection speed scales with the amount of DNA present in the bacterium. **(g),** lifetime of the protein-mediated DNA loop scales with the loop formation free energy.

Chapter 2

Mechanics in a Transcriptional Regulation Motif

This project is a collaboration with Stephanie Johnson, and with special thanks for David Wu for developing the acquisition and analysis codes.

A version of this chapter originally appeared as: "Poly(dA:dT)-rich DNAs are Highly Flexible in the Context of DNA Looping. Johnson S*, Chen YJ*, Phillips R, 2013 PLoS One 8(10):e75799".

2.1 Sequence-Dependent Mechanical Properties of DNA

Large-scale DNA deformation is ubiquitous in transcriptional regulation in prokaryotes and eukaryotes alike. Though much is known about how transcription factors and constellations of binding sites dictate where and how gene regulation will occur, less is known about the role played by the intervening DNA. In this work we explore the effect of sequence flexibility on transcription factor-mediated DNA looping, by drawing on sequences identified in nucleosome formation and ligase-mediated cyclization assays as being especially favorable for or resistant to large deformations. We examine a poly(dA:dT)-rich, nucleosome-repelling sequence that is often thought to belong to a class of highly inflexible DNAs; two strong nucleosome positioning sequences that share a set of particular sequence features common to nucleosome-preferring DNAs; and a CG-rich sequence representative of high G+C-content genomic regions that correlate with high nucleosome occupancy *in vivo*. To measure the flexibility of these sequences in the context of DNA looping, we combine the *in vitro* single-molecule tethered particle motion assay, a canonical looping protein, and a statistical mechanical model that allows us to quantitatively relate the looping probability to the looping free energy. We show that, in contrast to the case of nucleosome occupancy, G+C content does not positively correlate with looping probability, and that despite sharing sequence features that are thought to determine nucleosome affinity, the two strong nucleosome positioning sequences behave markedly dissimilarly in the context of looping. Most surprisingly, the poly(dA:dT)-rich DNA that is often characterized as highly inflexible in fact exhibits one of the highest propensities for looping that we have measured. These results argue for a need to revisit our understanding of the mechanical properties of DNA in a way that will provide a basis for

understanding DNA deformation over the entire range of biologically relevant scenarios that are impacted by DNA deformability.

Although it has been known since the work of Jacob and Monod that genomes encode special regulatory sequences in the form of binding sites for proteins that modulate transcription, only recently has it become clear that genomes encode other regulatory features in their sequences as well. Further, with the advent of modern sequencing methods, it is of great interest to have a base-pair resolution understanding of the significance of the entirety of genomes, not just specific coding regions and putative regulatory sites.

One well-known example of other information present in genomes is the different sequence preferences that confer nucleosome positioning [29–31], with similar ideas at least partially relevant in the context of architectural proteins in bacteria also [32]. It has been shown both from analyses of sequences isolated from natural sources and from *in vitro* nucleosome affinity studies with synthetic sequences that the DNA sequence can cause the relative affinity of nucleosomes for DNA to vary over several orders of magnitude, most likely due to the intrinsic flexibility, especially bendability, of the particular DNA sequence in question [31, 33–36]. The claim that intrinsic DNA sequence flexibility determines nucleosome affinity has led not only to many theoretical and experimental studies on the relationship between sequence and flexibility [37–43], but also to the elucidation of numerous sequence "rules" that can be used to predict the likelihood that a nucleosome will prefer certain sequences over others (summarized recently in [30, 35]). For example, AA/TT/AT/TA steps in phase with the helical repeat of the DNA, with GG/CC/CG/GC steps five base pairs out of phase with the AA/TT/AT/TA steps, are a common motif in both naturally occurring and synthetic nucleosome-preferring sequences [31, 35]. Similarly, the G+C content of a sequence and occurrence of poly(dA:dT) tracts have been very powerful parameters in predicting nucleosome occupancy *in vivo* [30, 44–47]. Our aim here is to explore the extent to which these sequences, when taken beyond the context of cyclization and nucleosome formation to another critical DNA deformation motif, exhibit similar effects on a distinct kind of deformation.

There has been an especially long history of the study of these intriguing sequence motifs known as poly(dA:dT) tracts, in the context of nucleosome occupancy as well as many other biological contexts. Such sequences, composed of 4 or more A bases in a row (A_n with $n \geq 4$) or two or more A bases followed by an equal number of T bases ($A_n T_n$ with $n \geq 2$), strongly disfavor nucleosome formation, both *in vivo* [48–51] and *in vitro* [12, 52–55], and are in fact thought to be one of the primary determinants of nucleosome positions *in vivo* [30, 49], with their presence upstream of promoters and in the downstream genes correlating with increased gene expression levels [48, 56, 57]. Poly(dA:dT) tracts show unique structural and dynamic properties in a variety of *in vitro* and *in vivo* assays (summarized recently in [49, 58]), with one of their hallmark characteristics being a marked intrinsic curvature [58]. There is evidence that poly(dA:dT) tracts may also be less flexible than other sequences [2, 59, 60], which is often given as the reason for their low affinity for nucleosomes, though there is some evidence that poly(dA:dT) tracts might actually be *more* flexible than other sequences [37]. It is clear, however, that some special property or properties of A-tracts

leads them to be especially resistant to the deformations that are required for DNA wrapped in a nucleosome [49, 58], and, indeed, to their important functions in several other biological contexts as well [58].

2.2 Nucleosome-Associated DNA Sequences with Unusual Deformability

In this work, we make use of sequences that, in the context of nucleosome formation and cyclization assays, appear to be associated with distinct flexibilities as a starting point for examining the question of what sequence rules control deformations induced by a DNA-loop-forming transcription factor, as opposed to those induced in nucleosomes. We have previously argued using two synthetic sequences that DNA looping does not necessarily follow the same sequence-dependent trends as do nucleosome formation and cyclization [61]. Here we expand our repertoire of sequences to specifically test the generalizability of three sequence features known to be important in nucleosome biology and cyclization. We focus in particular on the intriguing class of nucleosome-repelling, poly(dA:dT)-rich DNAs that are thought to be especially resistant to deformation, making use of a naturally occurring poly(dA:dT)-rich sequence that forms a nucleosome-free region at a yeast promoter [51]. We note that the poly(dA:dT)-rich DNA we use here differs from the *phased* A-tracts that have been extensively characterized in the context of DNA looping, both *in vivo* and *in vitro* [62–70]. Phased A-tracts contain short poly(dA:dT) tracts spaced by non-A-tract DNAs such that the poly(dA:dT) tracts are in phase with the helical period of the DNA, generating globally curved structures that are known to significantly enhance DNA looping [62–66]. The poly(dA:dT)-rich sequence we examine here contains *unphased* A-tracts that we do not anticipate to have a sustained, global curvature.

We compare the effects on looping of this poly(dA:dT)-rich DNA not only to the effects of two synthetic sequences we have previously studied, but also to those of two additional naturally occurring, genomic sequences: the well-known, strong nucleosome positioning sequence 5S from a sea urchin ribosomal subunit [71], which, along with the 601TA sequence we previously studied, contains the repeating AA/TT/TA/AT and offset GG/CC/CG/GC steps that are common in nucleosome-preferring sequences; and one of the GC-rich sequences that are abundant in the exons and regulatory regions (*e.g.* promoters) of human genes, and that correlate with high nucleosome occupancy *in vivo* [46, 47, 50]. The 5S sequence has been examined using both *in vitro* cyclization and *in vitro* nucleosome formation assays and, along with the two synthetic sequences E8 and 601TA [36, 72], can be used as a standard for comparison between our and other *in vitro* assays. The five sequences used in this work and their effects on nucleosomes are summarized in Table 2.1.

The poly(dA:dT)-rich sequence (from Fig. 4 of Ref. [51]), GC-rich sequence (from "Human 2" at http://genie.weizmann.ac.il/pubs/field08/field08_data.html), and 5S sequences (from Fig. 1 of [71]) were cloned into the pZS25 plasmid used in [61], with these eukaryotic sequences replacing the E8 or TA sequences in that plasmid. In cases where the loop lengths used in this study were shorter than the 147 bp that are wrapped in nucleosomes, the corresponding looping sequences used in TPM were taken from the middle of these sequences (relative to the nucleosomal dyad); in cases where the nucleosomal sequences were shorter than the desired loop length, they were padded at one end with the random E8 sequence [36, 61, 73]. See Figures 2.1 and 2.2 in File S1 for details. As in [61], "no-promoter" loops were flanked by the synthetic,

strongest known operator (repressor binding site) O_{id} and the strongest naturally occurring operator O_1; "with-promoter" loops were flanked by O_{id} and a weaker naturally occurring operator, O_2, because these with-promoter constructs are also used in *in vivo* studies of the effect of loop architecture on YFP expression, in which case O_2 is a more convenient choice of operator than O_1. Similarly, the motivation to include the *lac*UV5 promoter in the loop stems from parallel *in vivo* studies, in which the promoter is a natural part of the looping architecture. The promoter is included in the loop between the sequence of interest and the O_2 operator. Figures 2.1 and 2.2 in File S1 gives the exact sequences used in this work; Fig. 2.3(B) shows the TPM constructs schematically.

Table 2.1: **Nucleosome-positioning or nucleosome-repelling sequences.**

Sequence Name	Species	Genomic Position	Nucleosome Affinity
poly(dA:dT) ("dA")	Budding yeast (*S. cerevisiae*)	Chr III, 38745 – 39785 bp (Ref. [51])	∼3-fold *in vivo* nucleosome depletion relative to average genomic DNA (Fig. 2E of Ref. [50]); ∼ 2 k_BT increase in energy of nucleosome formation *in vitro* relative to 5S (Fig. 8D of Ref. [50]) *(estimates based on similar sequences)*
GC-rich ("CG")	Human	Chr Y, 4482107 – 4481956 bp (Ref. [50])	*(not determined)*
5S	Sea urchin (*L. variegatus*)	20 bp-165 bp from the *Mbo* II fragment containing 5S rRNA gene (Ref. [71])	1.6 k_BT decrease in energy of nucleosome formation compared to E8 *in vitro* (Ref. [36])
601TA ("TA")	synthetic, strong nucleosome positioning sequence (Refs. [36, 73, 74])	N/A	3 k_BT decrease in energy of nucleosome formation compared to E8 *in vitro* (Ref. [36])
E8	synthetic random (Refs. [36, 73])	N/A	*(used as a reference)*

The sequences described here were chosen because each has been found to have significant effects on *in vivo* nucleosome positions and/or *in vitro* nucleosome affinities, as shown in the rightmost column. The exception is the GC-rich sequence from humans: although its nucleosome affinity has not been directly determined either *in vivo* or *in vitro*, it is predicted to correlate with high nucleosome occupancy because of its high G+C content [45] and is occupied by a nucleosome(s) *in vivo* according to micrococcal nuclease digestion [50]. Two-letter abbreviations given in parentheses under each full sequence name will be used in figure legends in the rest of this work.

Cloning of the sequences of interest into the pZS25 plasmid was accomplished in either one or two steps. For the 5S sequences, oligomers were first ordered from Integrated DNA Technologies as single-stranded forward and reverse complements, consisting of 69 bp (for the "with-promoter" constructs) or 105 bp (for the "no-promoter" constructs) of the 5S sequence, plus the O_{id} and O_1/O_2 operators, and, where applicable, the lacUV5 promoter sequence. These oligomers were annealed and then ligated into the pZS25 plasmid at the AatII and EcoRI restriction sites that fall just outside the operators that flank the E8 or TA sequences in the original pZS25 plasmids [61]. Second, Quik-Change mutagenesis (Agilent Technologies) was performed to generate additional lengths (that is, to introduce insertions or deletions) of the 5S sequence from the initial 105 bp loop lengths. However, we found that this site-directed mutagenesis step generated distributions of products for the poly(dA:dT) constructs, possibly due to replication slipped mispairing over repetitive sequences [75]. Therefore all lengths of the poly(dA:dT) sequence, as well as of the GC-rich sequence, which also have the potential to contain such "slippery" regions, were created by ligation of synthesized oligomers into the pZS25 plasmid. All constructs were confirmed by sequencing (Laragen Inc.) to have clean sequence reads, and the approximately 450 bp digoxigenin- and biotin-labeled TPM constructs were created by PCR as described for the E8- and TA-containing constructs in [61, 76]. Sequences of TPM constructs were again confirmed by sequencing before use.

```
E8108: GGCCGGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACGC

TA108: GGCCGTTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCACGTGC

5S101: -------CATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S102: ------TCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S103: -----GTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S104: ----CGTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S105: ---ACGTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S106: --GACGTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S107: -TGACGTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA
5S108: ATGACGTCATAACATCCCTGACCCTTTAAATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCACCGAAGTCA

PolyA101: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC-gtgctgatcccct------
PolyA102: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC-gtgctgatcccctg-----
PolyA103: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC-gtgctgatcccctgt----
PolyA104: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC-gtgctgatcccctgtg---
PolyA105: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC-gtgctgatcccctgtgc--
PolyA106: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC-gtgctgatcccctgtgct-
PolyA107: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGACggtgctgatcccctgtgct-
PolyA108: ACCTTGTATTGTATTTCCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGACggtgctgatcccctgtgctg

CG101: -------GGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG102: ------AGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG103: -----GAGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG104: ----AGAGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG105: ---CAGAGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG106: --CCAGAGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG107: -TCCAGAGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
CG108: TTCCAGAGGGGCACCCACCAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAGGCACAGTTGTCAGGGACC
```

Figure 2.1: "No-promoter" looping sequences used in this work, compared to one length each of the E8 and TA sequences used in [61] (see Ref. [61] for additional lengths of the E8 and TA sequences). All sequences are listed 5′ to 3′. The O_{id} operator is immediately 5′ to these sequences, and has the sequence 5′-AATTGTGAGCGCTCACAATT-3′. O_1 is immediately 3′ and has the sequence 5′-AATTGTGAGCGGATAACAATT-3′. The 5S sequences shown here are the middle 101-108 bp of the full 5S sequence described by [71]; the CG sequences are the middle 101-108 bp from the Y-chromosome of "Human 2" at http://genie.weizmann.ac.il/pubs/field08/field08_data.html (see also Ref. [50]). The poly(dA:dT)-rich sequence from [51] is only 88 bp long and so was padded with E8 on the O_1-proximal end (as indicated by the lower-case letters). Poly(dA:dT) tracts, defined as stretches of 4 or more A bases in a row, are indicated in green; the TA/AA/AT/TT bases spaced ten bases apart that contribute to the nucleosome preferences of the TA and 5S sequences are indicated in red; and the CG/GG/CC/GC bases five bases out of phase with the TA/AA/AT/TT bases, which also contribute to nucleosome preference, are shown in blue. Note that the TA and GC bases on the 3′ end of the TA sequence, in boldface letters, are one base-pair out of phase with those on the 5′ end.

```
E8108(prom): --------TACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC---------------

TA108(prom): ----------CTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC-------------

5S101(prom): -----------------------------------TAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S102(prom): -----------------------------------TTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S103(prom): ----------------------------------CTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S104(prom): ---------------------------------GCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S105(prom): --------------------------------AGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S106(prom): -------------------------------TAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S107(prom): ------------------------------ATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------
5S108(prom): -----------------------------AATAGCTTAACTTTCATCAAGCAAGAGCCTACGACCATACCATGCTGAATATACCGGTTCTCGTCCGATCAC--------

PolyA101(prom): ----------------------GTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA102(prom): ---------------------CGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA103(prom): --------------------GCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA104(prom): -------------------TGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA105(prom): ------------------TTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA106(prom): -----------------TTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA107(prom): ----------------CTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------
PolyA108(prom): ---------------CCTTTGCGTGATGAAAAAAAAACTGAAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAAACAGGAC--------------------

CG101(prom): ------------------------CAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG102(prom): -----------------------CCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG103(prom): ----------------------GCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG104(prom): ---------------------TGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG105(prom): --------------------ATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG106(prom): -------------------GATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG107(prom): ------------------AGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
CG108(prom): -----------------CAGATGCCAGCTGGAGCTCTCCTGTATGAGGGATCTGTTGATTCCAGCTGGGAGGTGTCTGCTACTCAGGAG------------------
```

Figure 2.2: "With-promoter" looping sequences used in this work. Colors are the same as in Fig. 2.1. Note that these sequences are shorter versions of those in Fig. 2.1, so dashes indicate missing bases relative to the 108-bp version of each sequence given in Fig. 2.1. The O_{id} operator is immediately 5′ to these sequences, and has the sequence 5′-AATTGTGAGCGCTCACAATT-3′; the lacUV5 promoter, 5′-TTTACAATTAATGCTTCCGGCTCGTATAATGTGTGG-3′, is immediately 3′ to these sequences, followed immediately by the O_2 operator, 5′-GGTTGTTACTCGCTCACATTT-3′.

2.3 Probing Sequence-Dependent Deformability in the *lac* Operon

2.3.1 Assay DNA Deformability by Tethered Particle Motion

To measure the effect of these sequences on looping rather than nucleosome formation, we made use of a combination of an *in vitro* single-molecule assay for DNA looping, called tethered particle motion (TPM) [15, 77–79], with the canonical *E. coli* Lac repressor to induce looping, and a statistical mechanical model for looping that allows us to extract a quantitative measure of DNA flexibility, called the looping J-factor, for the DNA in the loop [61, 76]. We have recently demonstrated [61] that this combined method offers a powerful and complementary approach to established assays that have been used to probe the mechanical properties of DNA, particularly at short length scales, to great effect, such as ligase-mediated DNA cyclization [36, 43, 80–85] and measured DNA end-to-end distance by fluorescence resonance energy transfer [2, 86]. In particular, using the Lac repressor as a tool to probe the role of DNA deformability in loop formation allows us to examine the effect of sequence on the formation of shapes other than the roughly circular ones formed by cyclization and nucleosome formation, which we have argued may be an important caveat to discovering general flexibility rules from nucleosome formation and cyclization studies alone [61].

Tethered particle motion assays were performed as described in [61]. Briefly, linear DNAs, labeled on one end with digoxigenin and on the other end with biotin, were introduced into chambers created between a microscope slide and coverslip, with the coverslip coated nonspecifically with anti-digoxigenin. Streptavidin-coated beads (Bangs Laboratories, Inc) were then introduced into the chamber to complete the formation of tethered particles. The motion of the beads was tracked using custom Matlab code that calculated each bead's root-mean-squared (RMS) motion in the plane of the coverslip, and looping probabilities were extracted from these RMS-versus-time trajectories as the time spent in the looped state (reduced RMS), divided by total observation time. Similarly, the probabilities of the "bottom" versus "middle" states (see Results section) were defined as the time spent in a particular state, divided by the total observation time.

By measuring the looping probability of a construct at a particular repressor concentration, and using the repressor-operator dissociation constants for O_1, O_2 and O_{id} in [61], we can calculate the J-factor for that construct. All measurements in this work were carried out at 100 pM repressor, using repressor purified in-house. The relationship between the looping probabilities measured in TPM (p_{loop}), the repressor-operator dissociation constants for the two operators that flank the loop (K_1, K_2 and K_{id}), and the looping J-factor of the DNA in the loop (J_{loop}) can be described as

$$p_{\mathrm{loop}} = \frac{\frac{[R]J_{\mathrm{loop}}}{2K_AK_B}}{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_AK_B} + \frac{[R]J_{\mathrm{loop}}}{2K_AK_B}}, \tag{2.1}$$

where $[R]$ is the concentration of Lac repressor, and K_A and K_B are repressor-operator dissociation constants of the two operators flanking the loop (K_{id} and K_1 or K_2). A similar expression can be derived for the J-factors of the individual "bottom" and "middle" looped states and is given in [61].

2.3.2   Looping Probability Results

Our experimental approach to examining the effect of DNA sequence on looping combines an *in vitro* single-molecule assay for DNA looping, called tethered particle motion (TPM) [15, 77–79], with a statistical mechanical model that allows us to extract biological parameters from the single-molecule data [61, 76]. As shown schematically in Fig. 2.3(A), in TPM, a microscopic bead is tethered to a microscope coverslip by a linear piece of DNA, with the motion of the bead serving as a reporter of the state of the DNA tether: the formation of a protein-mediated DNA loop in the tether reduces the motion of the bead in a detectable fashion [15, 77–79]. We use the canonical Lac repressor from *E. coli* to induce DNA loops. Because more readily deformable sequences allow loops to form more easily, we can quantify sequence-dependent DNA flexibility by quantifying the looping probability, which we calculate as the time spent in the looped state divided by the total observation time (see Methods for details).

More precisely, our statistical mechanical model (described in the Methods section) allows us to extract a parameter called the looping J-factor from looping probabilities [61]. The J-factor is the effective concentration of one end of the loop in the vicinity of the other, analogous to the J-factor measured in ligase-mediated DNA cyclization assays [80, 87], and is mathematically related to the energy required to deform the DNA into a loop, ΔF_{loop}, according to the relationship:

$$J_{\mathrm{loop}} = 1 \text{ M } e^{-\beta \Delta F_{\mathrm{loop}}}, \tag{2.2}$$

where $\beta = 1/(k_B T)$ (k_B being Boltzmann's constant and T the temperature). A higher J-factor therefore corresponds to a lower free energy of loop formation. In the case of cyclization, where the boundary conditions of the ligated circular DNA are well understood, the J-factor can be expressed in terms of parameters describing the twisting and bending flexibility of the DNA, and its helical period [38, 43, 88, 89]. However, in the case of DNA looping by the Lac repressor, where the boundary conditions are not well known (summarized in Fig. 4 of [61]), an expression for the looping J-factor in terms of the twist and bend flexibility parameters of the loop DNA has not been described. Nevertheless, by measuring the J-factors for different sequences, we can comparatively assess the effect of sequence on the energy required to deform the DNA into a loop, and thereby gain insight into the sequence rules that control this deformation.

Given that 5S and TA share both sequence features and similar trends in apparent flexibility in the contexts of nucleosome formation and cyclization [36, 73, 74], we expected these two sequences to behave similarly to each other in the context of looping. On the other hand, since poly(dA:dT)-rich sequences are supposed to assume such unique structures as to strongly disfavor nucleosome formation [49, 58], while high GC content is one of the strongest predictors of high nucleosome occupancy [45, 50], we expected these two sequences to behave very differently from each other in the context of looping. Given the common assumption that poly(dA:dT)-rich DNAs are highly resistant to deformation, we especially did not expect to observe much, if any, loop formation with the poly(dA:dT)-rich, nucleosome-repelling sequence.

As shown in Fig. 2.3, none of these expectations were borne out. TA and 5S do not behave similarly, nor do CG and poly(dA:dT) behave especially dissimilarly, nor does poly(dA:dT) resist loop formation. Moreover, the behavior of these special nucleosome-preferring or nucleosome-repelling sequences is dependent on the larger DNA context, in that the addition of the 36-bp bacterial *lac*UV5 promoter sequence to these roughly 100-bp loops changes the relative looping probabilities of the five sequences (see Methods for the rationale behind the inclusion of this promoter). Without this promoter sequence (Fig. 2.3(C)), the two synthetic sequences, E8 and TA, exhibit comparable amounts of looping, while the three natural sequences, including both 5S and poly(dA:dT), all loop more than either E8 or TA. With the promoter (Fig 2.3(D)), however, TA loops more than E8, but 5S *less* than either E8 or TA. Both with and without the promoter the supposedly very different GC-rich and poly(dA:dT)-rich DNAs loop more than the random E8 sequence. The looping probabilities of the poly(dA:dT) sequence are especially surprising—instead of looping very little, as we expected, this sequence loops more than any other sequence without the promoter and a comparable amount to TA with the promoter.

These five sequences differ not only in looping probability, but also in the loop length at which that looping is maximal: the poly(dA:dT) sequence is maximized at 104 bp, the 5S and CG sequences at 105 bp, and the E8 and TA sequences at 106 bp. These different maxima could be explained by different helical periods for these five DNAs, though without more periods of data we cannot definitively quantify their helical periods. In the case of the poly(dA:dT) sequence, an altered helical period would not be unexpected, as pure poly(dA:dT) copolymers are known to have shorter helical periods (10.1 bp/turn) than random DNAs (10.6 bp/turn) [90, 91]. On the other hand, 5S exhibits the same helical period as E8 and TA in cyclization assays [73], so it is intriguing that its looping maximum occurs at a different length than that of E8 and TA, perhaps suggesting a different helical period in the context of looping than that of E8 and TA. The promoter does not appear to alter the maximum of looping for a given sequence. As noted above, it is difficult to use these looping data to comment further on other DNA elasticity parameters, in particular any sequence-dependent differences in torsional stiffness, but in Fig. 2.4 in File S1 we provide evidence that these sequences may share the same twisting flexibility, even if they differ in helical period.

The effect of the promoter on loop formation can be more clearly seen when looping J-factors are compared across sequences, instead of the looping probabilities. Because the no-promoter and with-promoter loops are flanked by different combinations of operators (Fig. 2.3(B); see also Methods), their looping probabilities cannot be directly compared. However, as described above and in the Methods section, we can use the statistical mechanical model that we have described for this system to extract J-factors from each looping probability [61]. These J-factors are shown in Fig. 2.5. Loop sequence can modulate the looping J-factor by at least an order of magnitude (compare the poly(dA:dT) J-factors to those of 5S with promoter or E8 and TA, no-promoter). The *lac*UV5 promoter has the largest effect on the TA and 5S sequences (though of opposite sign), but appears to have little effect on poly(dA:dT)-containing and E8-containing loops, and moderate effect on CG-containing loops. It is intriguing how large and diverse an effect the 36-bp *lac*UV5

promoter has on the roughly 100 bp loops we examine here; but one possible explanation for its minimal effect on the poly(dA:dT)-rich sequence, at least, compared to the others, is that the properties of A-tract structures tend to dominate over the properties of surrounding sequences [58]. We note that our results in [61] comparing the effect of sequence versus flanking operators on measured J-factors preclude the possibility that the differences between the no-promoter and with-promoter constructs are due to the difference in flanking operators. We also note that it is possible that the effect of the promoter stems not from the promoter sequence itself, but from the fact that the sequences of interest that form the rest of the loop are shorter when 36 bp of the loop are replaced by the promoter sequence. However, we consider this explanation to be less likely, because as shown in the left-hand panels of Fig. 2.3(C) and (D) above, we have measured the looping probabilities (and J-factors; see [61]) of more than two periods of E8- and TA-containing DNAs, allowing a direct comparison of loops that contain the same amount of E8 and TA both with and without the promoter (compare, for example, no-promoter loop lengths of 90 bp to with-promoter lengths of 120 bp). In this case we still find that without the promoter the J-factors of the E8- and TA-containing loops are indistinguishable, but with the promoter the TA sequence loops more than the E8 sequence, indicating that it is the promoter and not a shortening of some unique element(s) of the E8 or TA sequences that cause the difference in J-factors with versus without the promoter for these two sequences.
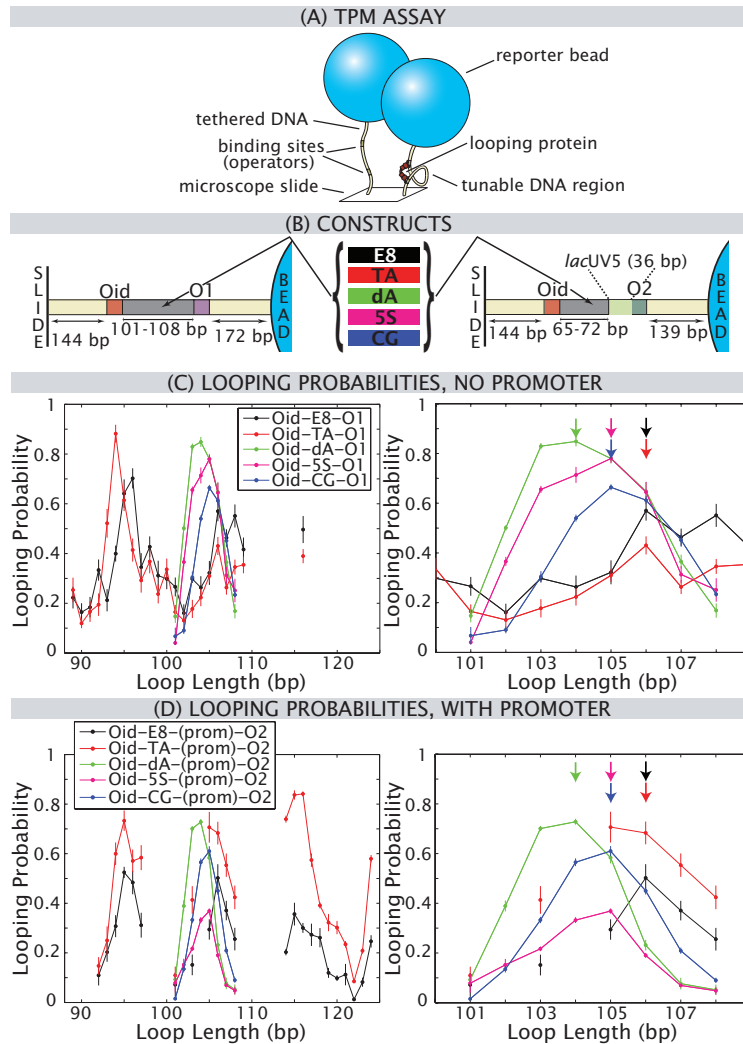
Figure 2.3: **Looping probability as a function of loop length and sequence. (A)** Schematic of the tethered particle motion (TPM) assay for measuring looping. In TPM, a bead is tethered to the surface of a microscope coverslip by a linear DNA. The motion of the bead serves as a readout for the state of the tether: if the DNA tether contains two binding sites for a looping protein such as the Lac repressor, and the looping protein is present and binds both sites simultaneously, forming a loop, the motion of the bead is reduced in a detectable fashion [15, 77–79]. The motion of the bead is observed over time, and the looping probability for a particular DNA is defined as the time spent in the looped (reduced motion) state, divided by the total observation time. **(B)** Schematic of the "no promoter" (left) and "with-promoter" (right) constructs used in this work. "Loop length" is defined as the inner edge-to-edge distance between operators (excluding the operators themselves, but including the promoter, if present). **(C)** Looping probabilities for the five sequences described in Table 2.1, without the bacterial *lac*UV5 promoter sequence as part of the loop. **(D)** Looping probabilities for the same five sequences but with the promoter sequence in the loop. Righthand panels in (C) and (D) show the same data as lefthand panels, except magnified around loop lengths 100-110 bp. The five sequences do not all share the same maxima of looping (colored arrows), not even the TA and 5S sequences that share similar sequence features (see Figures 2.1 and 2.2), though each sequence has the same maximum with and without the promoter (as far as can be determined with the current data; note that the with-promoter maximum for the TA sequence could be at 105 or 106, as those points are within error). All E8 and TA data (in particular, those outside of the 101-108 bp range) were previously described in [61].
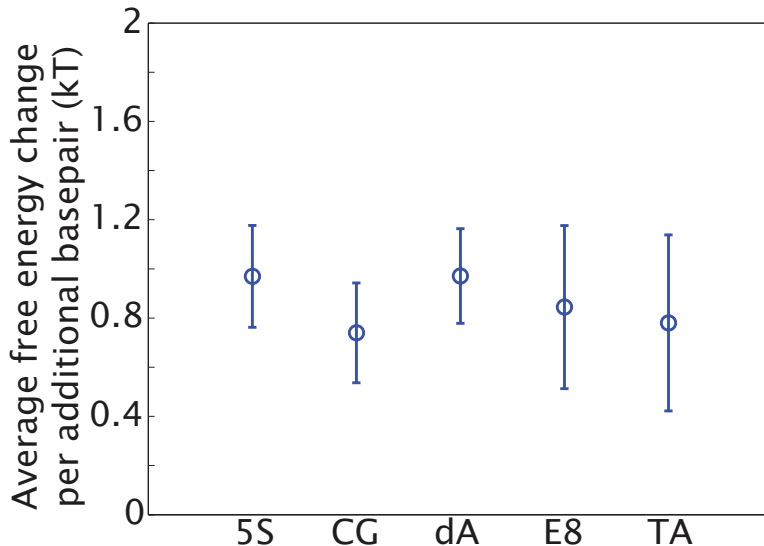
Figure 2.4: Sequence-dependent twist stiffness. In our data, differences in torsional stiffness between sequences would manifest as different amplitudes in the loop-length-dependent oscillations—that is, larger differences between peaks and troughs, or equivalently, different steepnesses of the slopes of the oscillations between peaks and troughs. One possible method for asking if we observe any such sequence-dependent changes in torsional stiffness is to fit the data of J-factors versus loop length in Fig. 2 in the main text to the functional form that has been derived for cyclization J-factors as a function of length (see Ref. [38]), even though the boundary conditions of looping and cyclization are very different, and discuss an "apparent" DNA stiffness. However, due to the lack of sufficient data in the troughs of the oscillations for all but the E8 and TA sequences, the errors on such a fitting attempt were too large for us to comment on the apparent DNA stiffness using this method. A second possible method for investigating a potential sequence-dependent twist stiffness is to calculate the average change in looping free energy (related to the looping J-factor through Eq. 1 in the main text) between loops of n basepairs and loops of $n-1$ basepairs, for each sequence. This gives a measure of the amplitude of the length-dependent oscillations for each sequence. If the five sequences we examine differ in twist stiffness, we would expect the average change in looping free energy per basepair added to the loop to be different for the different sequences, in that stiffer sequences would have larger oscillations (or steeper slopes). We show such a calculation here, that is, the average change in free energy per basepair added to the loop. The five sequences show the same change in free energy as a function of additional basepairs, suggesting that they may share the same torsional stiffnesses, though again with limited data it is difficult to make conclusive statements. Data shown here are for no-promoter loops only, and the E8 and TA data are for 101-108 bp loop lengths only, for consistency with the other sequences.
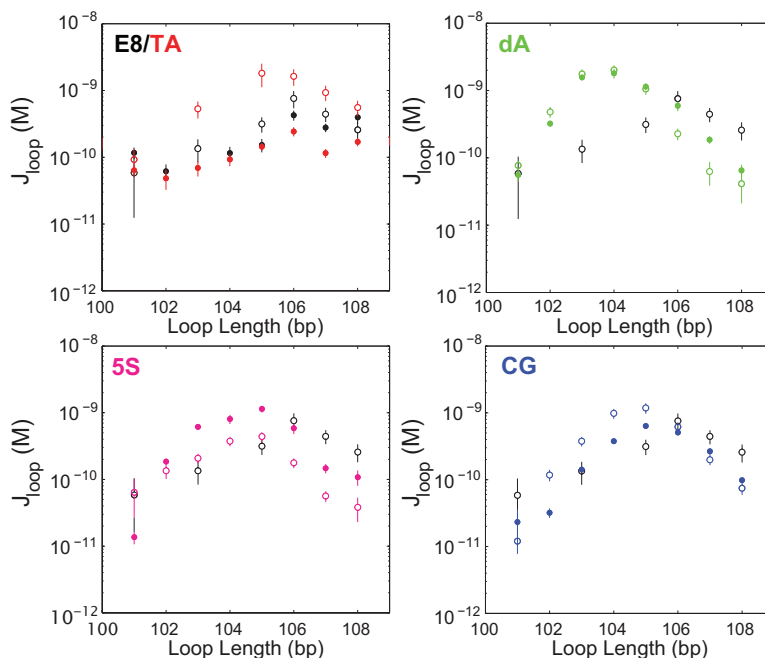
Figure 2.5: **Looping J-factors as a function of loop length and sequence.** J-factors for sequences without (closed circles) and with (open circles) the *lac*UV5 promoter were extracted from the data in Fig. 2.3 as described in the Methods section. The J-factor is a measure of the free energy of loop formation (and is related to the bending and twisting flexibility of the DNA in the loop): the higher the J-factor, the lower the free energy required to deform the loop region DNA into a loop. As described in [61], the addition of the promoter to the E8 loop sequence does not significantly affect its J-factor, so the J-factor for E8 with the promoter is shown as a reference in all panels (black open points). In contrast to E8, the addition of the promoter does change the J-factors for three of the four other sequences, making the TA-containing loops more flexible, but the 5S and, to a lesser extent, the CG sequences less flexible. Interestingly, the poly(dA:dT) sequence, like the E8 sequence, is unchanged with the inclusion of the promoter. We note that because the no-promoter versus with-promoter constructs contain different combinations of repressor binding sites, we can only use J-factors, not looping probabilities, to quantitatively examine the effect of the promoter; the statistical mechanical model of Eqn. 3.15 allows us to make this comparison.

2.4   Discussions

2.4.1   Indications for Loop Structure.

TPM trajectories not only provide information about the free energy of loop formation, captured by the J-factors discussed in the previous section, but also contain some information about the preferred loop conformation as a function of sequence, through the observed length of the TPM tether when a loop has formed. In fact, previous work from our group and others has shown that the Lac repressor can support at least two observable loop conformations for any pair of operators, with any sequence, because these conformations lead to distinct tether lengths in TPM [61, 63–65, 65, 66, 76, 92–94]. Although the underlying molecular details of these two looped states, which we label the "middle" ("M") and "bottom" ("B") states according to their tether lengths relative to the unlooped state, are as yet unknown, they must differ in repressor and/or DNA conformation in a way that alters the boundary conditions of the loop, since they are distinguishable in TPM. It has been proposed that the two states arise from the four distinct DNA binding topologies allowed by a V-shaped Lac repressor similar to that shown in the Lac repressor crystal structure [95, 96], and/or two repressor conformations, the V-shape seen in the crystal structure and a more extended "E" shape [65, 66, 92, 94, 97–99]. It is likely, in fact, that the two observed looped states are each composed of more than one microstate (that is, some combination of V-shaped and E-shaped repressor conformation(s) and associated binding topologies [95]). Even without knowing the details of the underlying molecular conformation(s) of these two states, however, we can use them to provide a window into the effect of sequence on preferred loop conformation.

In particular, by examining the relative probability of the two looped states as a function of both loop length and loop sequence, we can assess the contributions of sequence to the energy required to form the associated loop conformation(s). As shown in Figure 2.8, which of the two looped states predominates depends in a complicated way upon the loop sequence, the presence versus absence of the lacUV5 promoter, and the loop length. In [61], we showed that having E8 or TA in the loop region, over two to three helical periods, leads to alternating preferences for the middle versus the bottom looped state, with the middle state predominating when the operators are in-phase and looping is maximal, but the bottom state predominating when the operators are out-of-phase. The inclusion of the promoter in the loop increases the preference for middle state for out-of-phase operators. These trends are captured in the top left panel of Fig. 2.8.

These trends do not hold for the three genomically sourced loop sequences (CG, dA, and 5S). For the poly(dA:dT)-rich sequence, as with E8 and TA, the promoter increases the preference for the middle looped state for out-of-phase operators; for 5S, however, the presence of the promoter decreases the preference for the middle state. The preferred state of the CG sequence is mostly insensitive to the presence versus absence of the promoter. Both with and without the promoter, though, the middle state is generally preferred ($J_{\mathrm{loop,M}}/J_{\mathrm{loop,tot}} \geq 0.5$) at more loop lengths for the genomically sourced DNAs than for the synthetic sequences, insofar as we are able to determine from the lengths shown in Fig. 2.8. These results demonstrate

a complicated dependence of preferred loop state on sequence that does not always follow overall trends in looping free energy: for example, 5S and TA are the two sequences that show the largest change in J-factor with the inclusion of the promoter, but E8 and TA are the sequences that show the largest change in preferred looped state with the promoter. However, the trend seen in the preceding section with CG and poly(dA:dT) having more in common than 5S and TA holds true for preferred loop conformation as well.

A different measure of loop conformation can be derived from the TPM tether lengths themselves—that is, from the measured root-mean-squared motion of the bead, $\langle R \rangle$, as in the example trajectory shown in Fig. 2.10(A), which exhibits three clear states, the two looped states and the unlooped state. Because of variability in initial tether length, even in the absence of Lac repressor, we calculate a *relative* measure of tether length for the unlooped and looped states, where the motion of each bead is normalized to its motion in the absence of repressor. We might expect, then, that in the presence of repressor, the unlooped state would fall at a relative $\langle R \rangle$ of zero, and the looped states at negative values. However, as can be seen in the sample trace in Fig. 2.10(A) and in the lefthand panels of Fig. 2.10(B), the unlooped state in the presence of repressor is actually shorter than the tether in the absence of repressor (*i.e.,* the horizontal black dashed line in Fig. 2.10(A) lies above the mean of the unlooped state in the blue data). In [61] we present evidence for this shortening of the unlooped state in the presence of repressor being due to the bending of the operators induced by the Lac repressor protein that is observed in the crystal structure of the Lac repressor complexed with DNA [96]. (We note that this is a Lac repressor-specific result; compare, for example, the recent results from Manzo and coworkers with the lambda repressor [100], where a similar shortening of the unlooped state is shown to be due to nonspecific binding. For example, the Lac repressor does not exhibit the dependence of the looped tether length on repressor concentration that is seen with the lambda repressor [61, 100]).

As shown in Fig. 2.10(B), the length of the TPM tether in both the unlooped and looped states is similar but not identical for the five sequences and eight lengths that we examine here. The most obvious modulation of tether length correlates with loop length, with the shortest unlooped- and looped-state tether lengths occurring near the maxima of the looping probability. We believe this modulation with length is due to the phasing of the bends of the DNA tether as it exits the repressor-bound operators in the looped state, or the phasing of the bent operators in the unlooped state. At the repressor concentration we use here, the unlooped state should be primarily composed of the doubly-bound state [61], meaning that the two operators are both bent by bound repressor. As shown schematically in Fig. 2.10(C), when these bends are in-phase, the tether length should be shortest (and also the looping probability is highest, because the operators are in-phase). A similar argument can be made for the modulation of the looped state, regarding the relative phases of the tangents of the DNA exiting the loop.

It is interesting to consider how the sequence of the loop might influence the length of the tether in the unlooped state, when no loop has formed; see, for example, the CG with-promoter versus 5S with-promoter sequences, where the latter is consistently longer than the former (Fig. 2.10(B)). We do not see a sequence dependence to tether length in the absence of repressor, ruling out the possibility of a detectable intrinsic

curvature to the CG sequence. We speculate instead that CG alters the trajectory of the DNA as it exits the bend in the operators in the unlooped state, compared to the trajectory when the sequence next to the operators is 5S, leading to a consistent difference in unlooped tether lengths.

Interestingly, in contrast to its influence on preferred looped state (middle versus bottom), the promoter does not alter the length of the tether for a given sequence at a given loop length (see also the bottom left panel of Fig. 2.6 and Fig. 2.7 in File S1). On the other hand, as shown in Fig. 2.10(D), the poly(dA:dT)-rich sequence, noticeably more so than the other sequences, stands out as a sequence that does strongly affect the tether length of the loop, in that it mandates a very narrow range of tether lengths as a function of looping J-factor (related, for a particular sequence, to the loop length or equivalently the operator spacing). A similar but less pronounced trend can be observed for the unlooped state with the GC-rich sequence (Fig. 2.10(D)). The other sequences allow much more variability in tether length as a function of J-factor/operator spacing (see Figure 2.7 in File S1). This strong trend in tether length as a function of J-factor could be evidence of the formation of special, defined loop structures with the GC-rich and poly(dA:dT)-rich sequences that constrain the allowed loop conformations as a function of operator spacing more than the other sequences do.

Further computational and modeling efforts will be required to relate these data on tether lengths and preferred loop length to loop structure, similarly to how Towles and coworkers have used TPM tether lengths to show that different DNA loop topologies can explain the observed tethered lengths of the two looped states [95]. However, even without currently knowing the underlying molecular details causing these sequence-specific trends in tether length and preferred loop state, and therefore in loop conformation, it is clear that it is the loop sequence, and not the Lac repressor itself, that determines the loop conformation to a large degree. It has been shown recently that the Lac repressor is capable of accommodating many different loop conformations [66], which is consistent with the results we present here. We hope that computational and modeling efforts with these data, as well as continued efforts to use assays such as FRET to directly probe loop conformation [63–66], will shed light on this complex interplay between sequence and loop conformation.
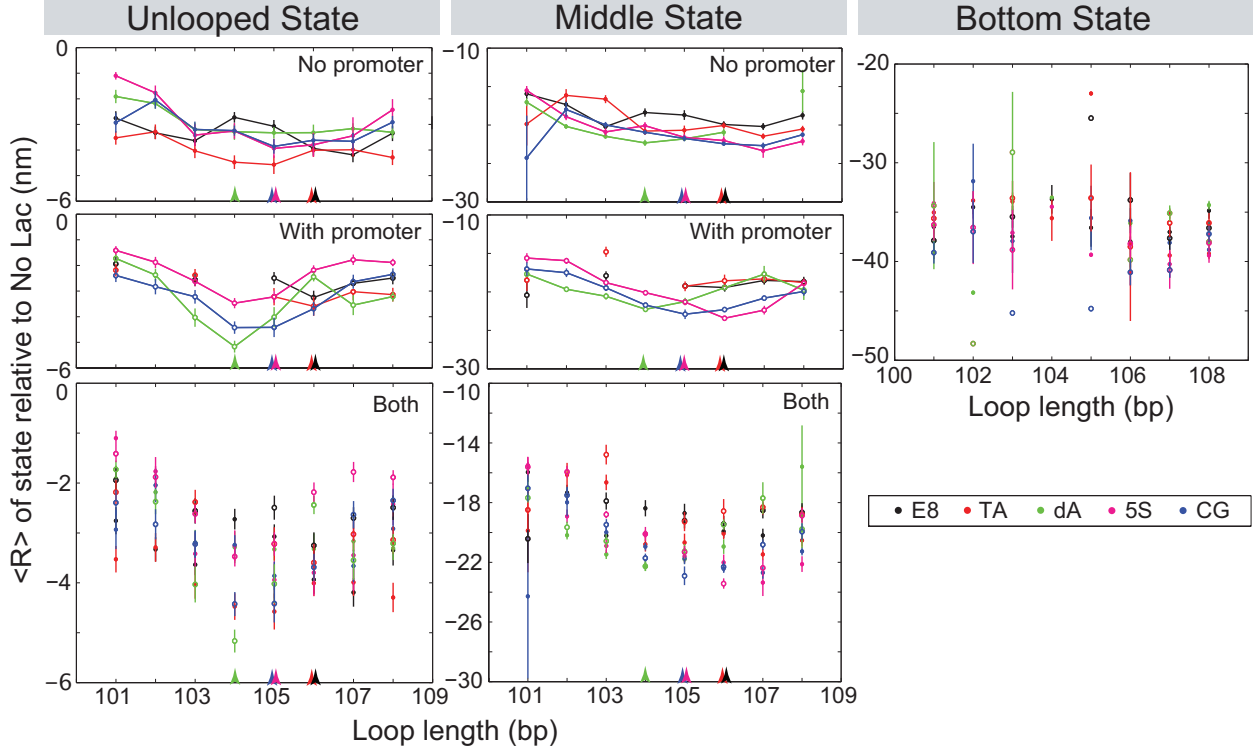
Figure 2.6: Tether lengths of looped and unlooped states as a function of loop length and sequence. The y-axes here and in the next figure are population averages of the difference in RMS between each bead's tether in the absence of protein, and the indicated state in the presence of protein (as described in the Supplementary Material of [61], there is sufficient tether-to-tether variability in the absence of protein, which we attribute to the variability in the diameters of the beads we use, that trends in tether lengths in the presence of protein are only observable when normalized to the length of each tether in the absence of protein). The bottom state has very poor statistics for most sequences and so should be considered indicative only of the ballparks of tether lengths we observe for that state. For the unlooped and middle states, however, we observe a modulation of tether length with loop length, with the shortest tether lengths for both states occurring near the maximum of looping (indicated for each sequence by the colored arrows at the bottoms of the plots). As argued more extensively in [61], we believe the reduction in tether length in the unlooped state in the presence of protein, compared to the tether length in the absence of protein, is due to the bending of the operators induced by the Lac repressor protein. The Lac repressor is known to bend the DNA of the O_{id} operator by 45 degrees [96]; our previous work suggests that a bound Lac repressor also bends the other operators but to a lesser degree, with the extent of bending directly proportional to the strength of the operator. At the repressor concentration we use here, the unlooped state should be primarily composed of the doubly-bound state [61], meaning that the two operators are both bent by bound repressor. When these bends are in-phase, the tether length is shortest (and also the looping probability is highest, because the operators are in-phase). We believe this explains the modulation of tether length in the unlooped state. A similar argument can be made for the modulation of the middle looped state, regarding the relative phases of the tangents of the DNA exiting the loop.
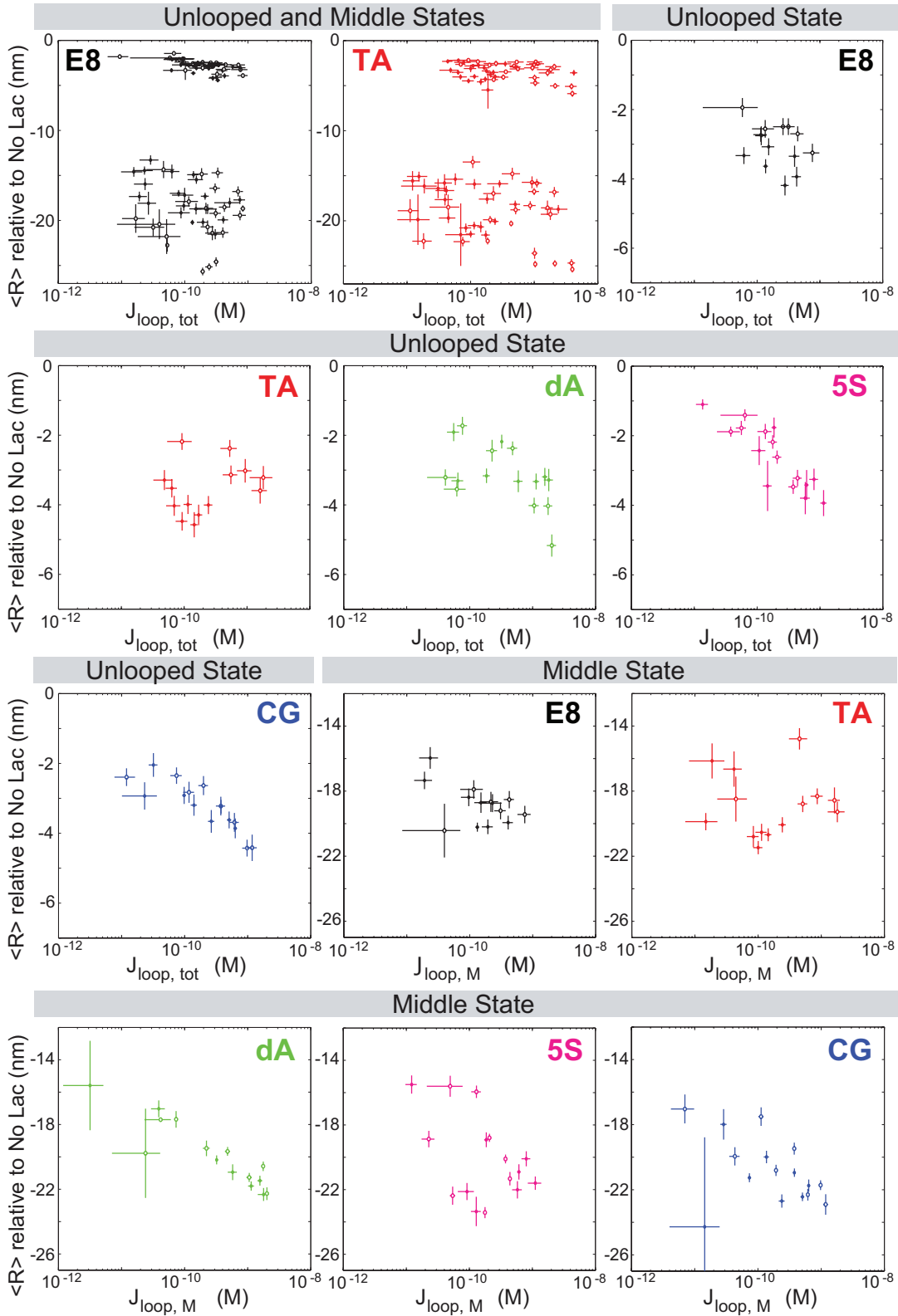
Figure 2.7: Tether length as a function of looping J-factor. Here as in the previous figure we show tether lengths for the indicated state normalized by each bead's RMS in the absence of protein, but as a function of J-factor, rather than loop length. Unlooped state tether lengths are always plotted versus the total J-factor, whereas middle state tether lengths
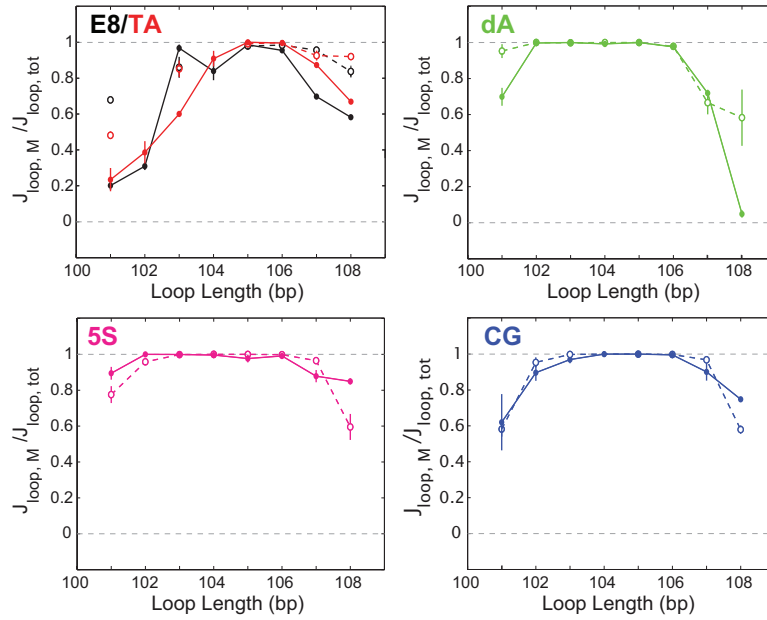
Figure 2.8: **Comparison of the likelihood of the "middle" (longer) versus "bottom" (shorter) looped states.** The y-axes indicate the fraction of the total J-factor that is contributed by the middle state (as in Fig. 2.5, since the with- and without-promoter constructs have different operators, J-factors and not looping probabilities must be compared). That is, when the ratio $J_{\mathrm{loop,M}}/J_{\mathrm{loop,tot}}$ is unity, indicated by a horizontal black dashed line, only the middle state is observed; when this ratio is zero, again indicated by a horizontal black dashed line, only the bottom state is observed. Closed circles are no-promoter constructs; open circles are with-promoter. E8 and TA data are a subset of those in [61]. Figure 2.9 shows the looping probabilities and J-factors for the two states instead of the relative measures shown here.
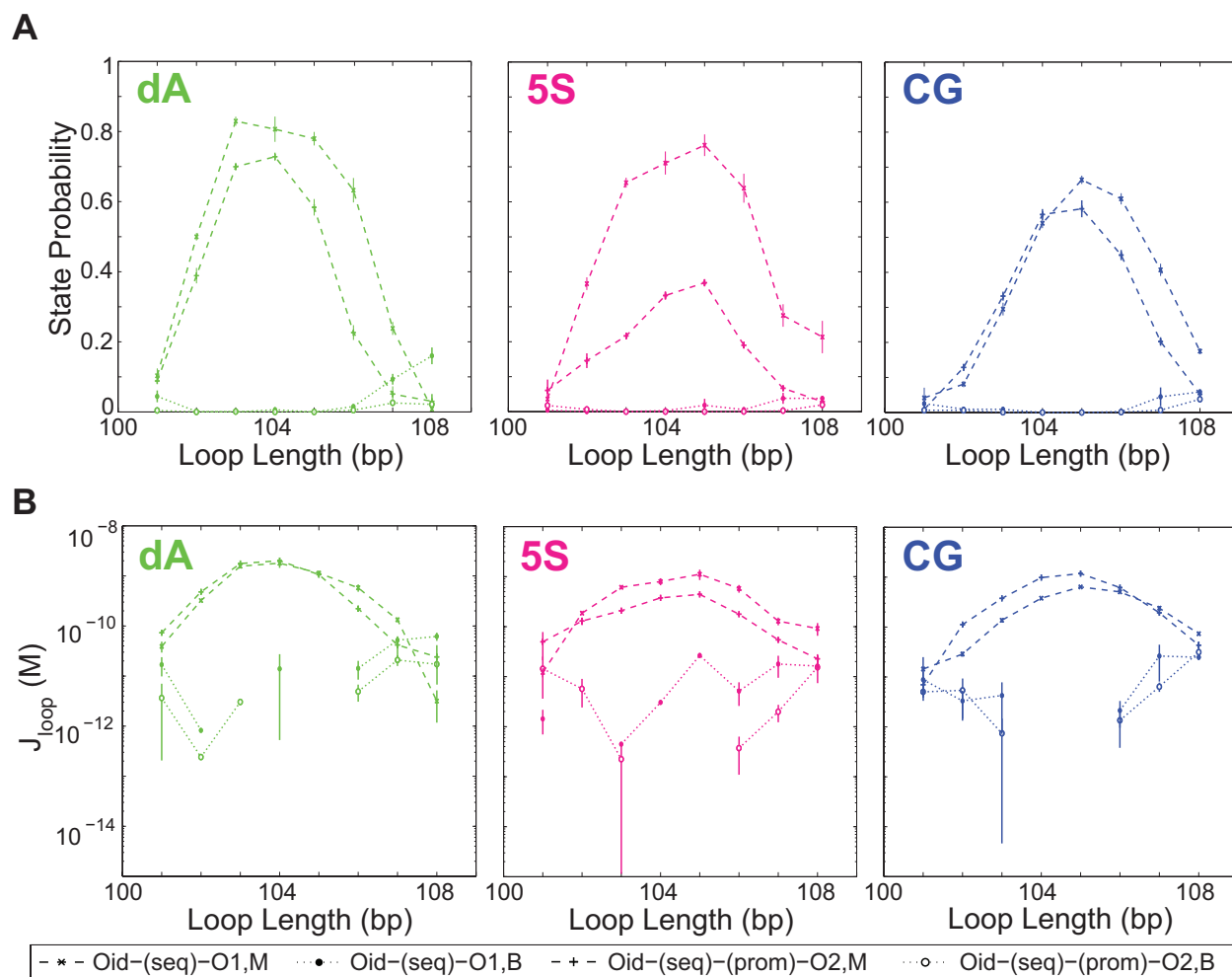
Figure 2.9: Looping probabilities **(A)** and J-factors **(B)** for the "middle" (M) and "bottom" (B) states separately. See Fig. 3 and the Supplementary Material of [61] for corresponding data for the E8 and TA sequences. Note that because of the different operators in the no-promoter versus with-promoter data, in some cases the no-promoter data has a lower looping probability than the with-promoter data, but yet a larger J-factor.
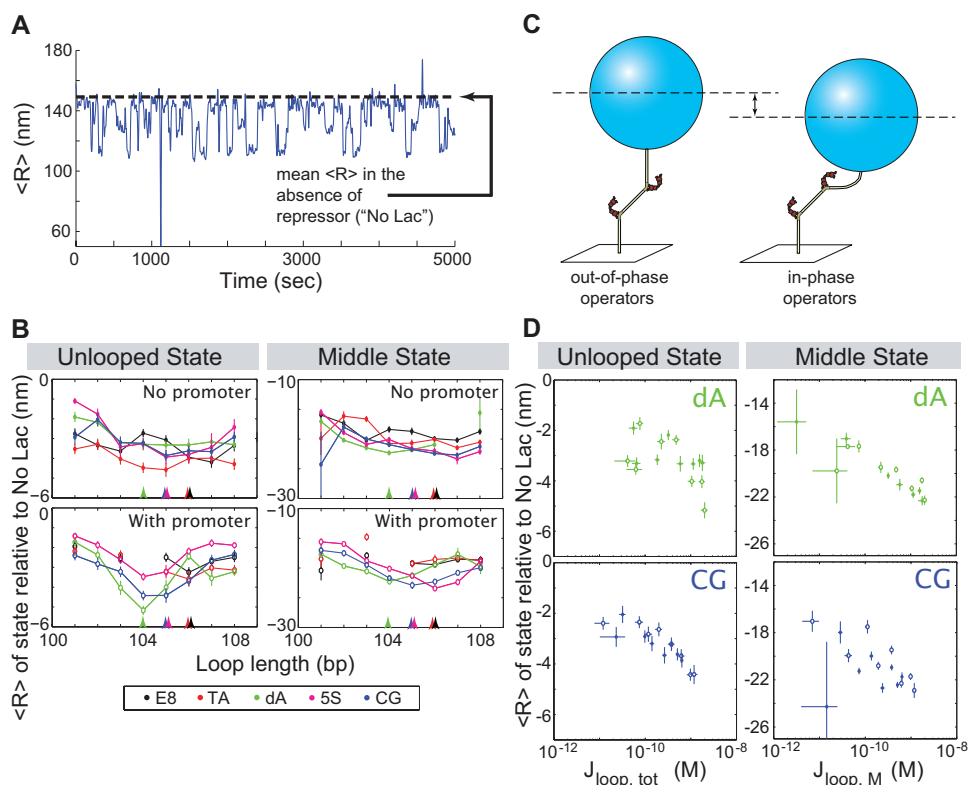
Figure 2.10: **Tether length as a function of loop length, sequence and J-factor. (A)** Sample TPM time trajectory showing the smoothed (*i.e.* Gaussian-filtered) root-mean-squared motion, $\langle R \rangle$, of a single bead. This construct shows an unlooped state and two looped states, the "middle" state around 130 nm, and the "bottom" state around 110 nm. Black horizontal dashed line indicates the average $\langle R \rangle$ for this particular tether in the absence of repressor. Due to variability in tether length even in the absence of repressor [61], on the y-axes in (B) and (D) we plot a relative measure of tether length, by normalizing the mean $\langle R \rangle$ value for a particular state to the mean $\langle R \rangle$ of each tether in the absence of repressor, and then taking the population average of this difference. **(B)** Tether length as a function of loop length. We observe a modulation of tether length with loop length, with the shortest tether lengths for both the unlooped and looped states occurring near the maximum of looping (indicated for each sequence by the colored arrows at the bottoms of the plots). See Fig. 2.6 in the Supporting Information for bottom state lengths. **(C)** Schematic of our proposed model for the observed variations in unlooped tether length as a function of loop length, which we attribute to the phasing of the bends that the repressor creates upon binding the operators. A similar argument can be made for the looped states. Note that to emphasize the effect of bending from the operators, here we have for the most part represented the DNA as straight segments. **(D)** Tether length as a function of J-factor. Unlooped state tether lengths are plotted versus the total J-factor, whereas middle state tether lengths are plotted versus the J-factor for the middle state. As in (A), in general the length of the tether in both the unlooped and middle looped states is shorter at larger J-factors for a particular sequence. However, this trend is sharper for some sequences than others (see Fig. 2.7 in the Supporting Information for the other sequences, which generally have more scatter than either the dA or CG sequences).

2.4.2 Comparing Looping and Nucleosome Formation

In [61] we showed that the synthetic E8 and TA sequences show no sequence dependence to looping in the absence of the *lac*UV5 promoter but a nucleosome-like sequence dependence in the presence of the promoter. We hypothesized that perhaps the promoter alters the preferred state of the loop to one whose shape is more similar to that of DNA in a nucleosome or DNA minicircle formed by cyclization, leading to similar sequence trends with the promoter as with nucleosomes. We still attribute the difference in the patterns of sequence dependence that we observe between looping and nucleosome formation to the role of the shape of the deformation in determining the observed deformability of a particular sequence. However, we have shown here with a broader range of sequences that the role of the promoter in controlling loopability is more complicated than we had previously hypothesized. Neither with nor without the promoter does loop formation follow the sequence trends of nucleosome formation. As shown in Figure 2.11, if looping J-factors did follow the same patterns of sequence preference as do cyclization J-factors and nucleosome formation free energies, a plot of the looping J-factors versus cyclization J-factors for the various sequences we have studied here would fall on a line with a positive slope. We find that this is not the case; in fact, without the promoter there is perhaps a slight anticorrelation between looping J-factors and cyclization J-factors (and no discernible correlation with the promoter).

The J-factors plotted in Figure 2.11 are the maximum looping or cyclization J-factors over a particular period. Specifically, the looping J-factors used are those at 104 bp for dA, 105 for 5S and CG, and 106 for E8 and TA; the cyclization J-factors are for 94 bp of the E8, 5S or TA sequences and are taken from [73]. Although we are not directly comparing identical lengths between cyclization and looping, the general trends hold regardless of lengths chosen. In fact, identifying the loop length that corresponds to a particular cyclization length is difficult, given that the flanking operators for looping must be taken into account in some fashion. That is, for cyclization, DNA length is easy to compute—it is simply the length of the oligomer used in the ligation reactions. However, in the case of looping, it is unclear if the appropriate length for comparison is just the DNA in the loop (excluding the operators), or the length between the midpoints of the operators, or including all of the operators. Similarly, we are not comparing identical loop lengths across sequences; we chose to compare loop flexibility at the looping maximum for each sequence in an attempt to compare lengths at which the operators are most likely to be in phase, such that we are comparing only bending and not twisting flexibility. Finally, we note that here we are interested in the same kind of comparison that Cloutier and Widom were in Ref. [36], which was the inspiration for this figure; in [36], Cloutier and Widom compared cyclization and nucleosome formation free energies, even though the cyclization experiments were performed with roughly 100 bp DNAs and the nucleosome formation assays with roughly 150 bp DNAs. Likewise, we do not expect that the fragments of nucleosome-preferring or nucleosome-repelling sequences that we examine here in the context of looping will necessarily have exactly the same characteristics as the full-length nucleosomal sequences from which they were derived; but we are interested in comparing general trends in observed flexibility of these roughly 110 bp loops with those of roughly 100 bp ligated minicircles

and of roughly 150 bp nucleosomal DNAs.

The strong correlation between a sequence's ease of cyclization and of nucleosome formation, as shown in Fig. 2.11(A), has been used to argue that nucleosome sequence preferences depend largely on the intrinsic mechanical properties of a DNA, particularly its bendability [36], though other mechanisms have also been proposed, such as that described by Rohs and coworkers, which depends not on sequence-dependent DNA flexibility but on sequence-dependent minor groove shape [101]. We have shown here that three sequence features that commonly determine nucleosome preferences, either through their effect on DNA flexibility or on other structural aspects recognized by the nucleosome, do not likewise determine looping, arguing for the need to identify a different set of sequence features that determine loopability. The most striking contrast between previously established sequence "rules" derived from nucleosome studies and the trends in looping J-factors that we observe here is that of the nucleosome-repelling, poly(dA:dT) sequence, which has the lowest looping free energy that we have quantified so far. Other *in vitro* assays predominantly show poly(dA:dT) copolymers to be highly resistant to deformations; for example, Vafabakhsh and coworkers recently used a FRET-based cyclization assay, analogous to traditional ligase-mediated cyclization assays, to show that poly(dA:dT)-rich sequences have cyclization rates significantly smaller than other sequences such as E8 and TA [2]. Although ease of cyclization is often equated with bendability, it appears that such observed bendability is more context-dependent than has been previously appreciated: that is, the simplest model that one would write down to describe the energetics of these different deformed DNAs would feature the persistence length as the governing parameter that is used to characterize bendability, and yet, the distinct responses seen in looping, nucleosomes and cyclization belie that simplest model. It will be informative to extend this study of an unphased poly(dA:dT) tract in DNA loops to include more sequences containing both pure poly(dA:dT) copolymers and naturally-occuring poly(dA:dT)-rich DNAs that exclude nucleosomes *in vivo*, in order to elucidate the precise role of poly(dA:dT)-tracts in determining looping. It is clear, however, that poly(dA:dT)-rich DNAs should not be exclusively thought of as stiff or resistant to bending in all biological contexts.
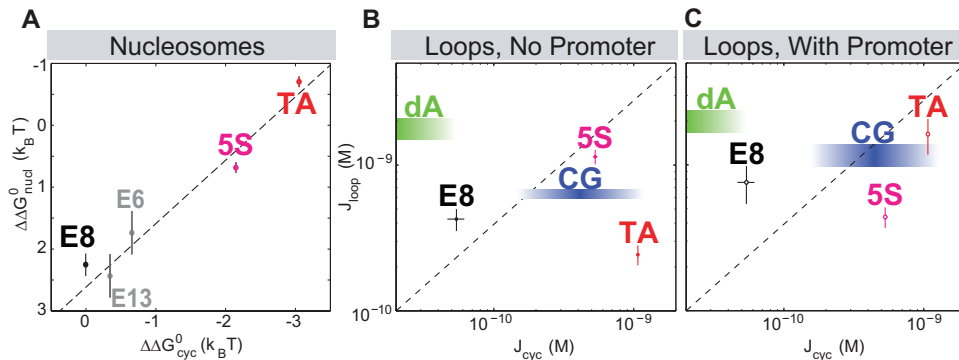
Figure 2.11: **Comparing trends in sequence flexibility for looping versus cyclization and nucleosome formation.** **(A)** Nucleosome formation and cyclization share trends in sequence flexibility, with sequences that have lower energies of nucleosome formation ($\Delta\Delta G^0_{nucl}$) also having lower energies of cyclization ($\Delta\Delta G^0_{cyc}$). Cloutier and Widom used this correlation to argue that the same mechanical properties, particularly the bendability, of the DNA contributed to nucleosome formation as to cyclization [36]. The energy of cyclization, ΔG^0_{cyc}, is related to the cyclization J-factor for a particular DNA, J_i, through the relationship $\Delta G^0_{cyc} = -RT \ln(J_i/J_{ref})$, where T is the temperature, R is the gas constant and J_{ref} is an arbitrary reference molecule (see Ref. [36] for details). Adapted from Refs. [36, 102]. **(B)** Looping J-factors for the no-promoter data do not show the same trends in sequence dependence as do cyclization and nucleosome formation: if anything, a higher cyclization J-factor correlates with a lower looping J-factor. **(C)** Same as (B) but for with-promoter DNAs. The cyclization J-factors of the poly(dA:dT)-rich and GC-rich sequences that we use here have not been reported, so they are shown as shaded regions whose height reflects the uncertainty in the looping J-factors we measure, and whose width reflect our estimates about what their cyclization behavior should be. In particular, the poly(dA:dT)-rich sequence exhibits very low nucleosome occupancy *in vivo* [50, 51], and similar sequences have high energies of nucleosome formation *in vitro* [12, 50], which, according to the logic of (A), should correspond to a low cyclization J-factor, probably lower than that of E8. Some poly(dA:dT)-rich DNAs were in fact recently directly shown to cyclize less readily than random sequences [2]. In contrast, the GC-rich sequence should be a good nucleosome former (though the nucleosome affinity of this particular sequence has not been tested either *in vivo* or *in vitro*), and so its cyclization J-factor is probably comparable to that of 5S and TA, the other strong nucleosome-preferring sequences on this plot. Additional details of how this plot was generated can be found in the Methods section.

2.4.3   Is G+C content a good parameter for DNA deformability?

A second striking contrast between our results here and previously established rules for nucleosome formation concerns the role of G+C content in determining loop formation. The G+C content of a DNA is one of the most powerful parameters for predicting nucleosome occupancy *in vivo* [45, 47], with higher G+C content correlating with higher occupancy. However, as shown in Fig. 2.12, G+C content offers little predictive power for loopability, or is anticorrelated with looping. We note that a recent, systematic DNA cyclization study demonstrated a quadratic dependence of DNA bending stiffness on G+C content [43]. In our case of protein-mediated DNA looping, the looping J-factor contains contributions from protein elasticity in addition to those from DNA elasticity, and our DNA sequences contain A-tracts and GGGCCC motifs that were excluded in [43], making a direct comparison between our results and theirs difficult; but it is possible that the looping J-factor is neither correlated or anticorrelated with G+C content but instead depends quadratically on G+C content, as do cyclization J-factors. More data will be necessary to make a strong statistical statement about the anticorrelation or lack of correlation between the looping J-factor and G+C content, and to determine the form of the relationship between the looping J-factor and G+C content (*e.g.* quadratic versus linear), but we propose low G+C content as the starting point of a potential new sequence "rule" for predicting looping J-factors, and a fertile realm of further investigation. Finally, we have shown that the repeating AA/TT/TA/AT and GG/CC/GC/CG steps that characterize the 5S and TA sequences, as well as many nucleosome-preferring sequences, do not likewise determine looping J-factors, as these two sequences behave very differently from each other in the context of transcription factor-mediated DNA looping.

In summary, we find that the poly(dA:dT)-rich sequence that strongly excludes nucleosomes *in vivo* [51] and that belongs to a class of sequences usually thought of as highly resistant to deformation is in fact the strongest looping sequence we have studied so far. Moreover, the 5S and TA sequences, which share sequence features important to nucleosome formation (see Figures 2.1 and 2.2 in File S1 and Ref. [35]) as well as trends in apparent flexibility in *in vitro* cyclization and nucleosome formation assays [36, 73, 74], behave very differently from each other in the context of looping. We also find that G+C content, a good predictor of nucleosome occupancy, is not likewise positively correlated with looping, and in fact our data suggest the G+C content and looping may be *anti*correlated. Taken together, these results strongly suggest that very different sequence rules determine DNA looping versus cyclization and nucleosome formation, possibly because of the protein-mediated boundary conditions that differ between looping geometries and nucleosomal geometries, and that the biophysical characteristics of poly(dA:dT)-rich DNAs and their biological functions may be more diverse and context-dependent than has been previously appreciated.

Our work in no way undermines previous claims of the sequence dependence to nucleosome formation and/or occupancy either *in vivo* or *in vitro*; rather, it demonstrates that the "rules" of sequence flexibility derived from cyclization and nucleosome formation studies are inapplicable to DNA looping, possibly due to the difference in the boundary conditions and therefore DNA conformations involved in forming a protein-mediated loop versus a DNA minicircle or a nucleosome. It will be interesting to extend these studies of the
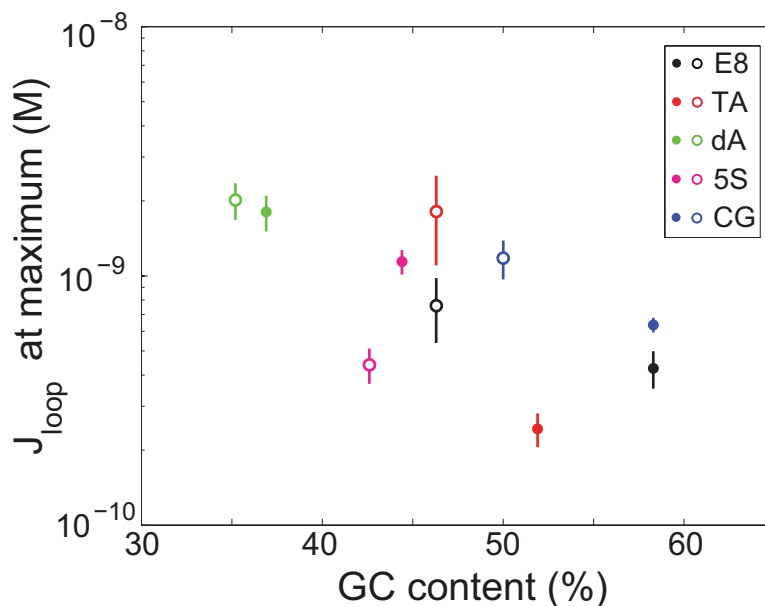
Figure 2.12: **Maximum looping J-factor as a function of loop G+C content.** Maximum J-factors for each of the five sequences, with (closed circles) and without (open circles) promoter, are plotted with respect to each sequence's G+C content. For nucleosomes, G+C content strongly correlates with nucleosome occupancy [45]. In contrast, it appears that G+C content and loopability are anti-correlated. Loop lengths plotted here are the same as in Fig. 2.11.

role of sequence in loop formation to other DNA looping proteins besides the Lac repressor. As noted above, it has been shown recently that the Lac repressor can accommodate many different loop conformations [66]. The variety in tether lengths and preferred looped states that we observe are consistent with a forgiving Lac repressor protein. Nucleosomes, on the other hand, have a more fixed structure that should not be as accommodating to a range of helical periods and DNA polymer conformations (hence the hypothesis that poly(dA:dT)-rich DNAs disfavor nucleosome formation because they adopt geometry that is incompatible with the structure of the DNA in a nucleosome [49]). It would be informative to measure the looping J-factors of these same sequences with a more rigid looping protein. It will also be interesting to see if other bacterial promoter sequences have similar effect of altering the looping boundary condition as the very strong and synthetic *lac*UV5 promoter. In fact, the *lac*UV5 promoter should be a key starting point for identifying sequences that have a strong effect on looping, since it can have significant effects on the behavior of a loop, even when it comprises only one-third of the loop length.

Chapter 3

DNA Looping Kinetics

This project is a collaboration with Peter Mulligan, Prof. Andrew Spakowitz (Stanford University) and Stephanie Johnson.

A version of this chapter originally appeared in: "Modulation of DNA loop lifetimes by the free energy of loop formation. Chen YJ*, Johnson S*, Mulligan P*, Spakowitz A, Phillips R, 2014 Proc Natl Acad Sci USA 111(49):17396401", and "Interplay of Protein Binding Interactions, DNA Mechanics, and Conformational Entropy in DNA Looping Kinetics. Mulligan P, Chen YJ, Phillips R, Spakowitz A, 2015 submitted".

3.1 Lifetimes of Short DNA Loops Scale with Looping Free Energy

Many key cellular processes, from gene regulation to metabolism, require the coordinated physical interaction of biological macromolecules. A classic example is the coordination of DNA and proteins in DNA loop formation, which is a recurring design principle from viruses [103] to animals [104]. Despite their prevalence, many questions remain about how these loops form and function *in vitro* and *in vivo* [38, 105]. Here, we ask how the mechanics of the protein and DNA in a protein-mediated loop govern looping and unlooping dynamics, an issue which has only recently begun to be explored [106].

Recent advances in single-molecule techniques have allowed precise quantification and deeper understanding of the physical properties of DNA [1, 2], and we have taken advantage of one such technique here, in combination with a classic bacterial looping protein. The polymer physics implications of this work, however, are more general than the particular system we focus on. The looping protein we examine here, the Lac repressor, so named because of its role in repressing transcription at the *lac* promoter in the bacterium *Escherichia coli*, was one of the first described examples of a genetic regulator that acts through specific DNA-protein interactions. Looping is accomplished by means of two DNA binding domains per Lac repressor molecule, allowing it to bind two of its specific DNA sites ("operators") simultaneously, with the intervening DNA adopting a looped conformation [96], as shown schematically in Fig. 3.1**a**. To measure loop formation and breakdown, we make use of the *in vitro* single-molecule tethered particle motion (TPM)

assay [15, 61, 77, 79], a simple but powerful technique for investigating DNA-protein interactions. In TPM, a micron-sized bead is tethered to one end of a linear DNA, with the other end attached to a microscope coverslip (Fig. 3.1**b**). The motion of the bead gives a readout of the effective length of the bead's DNA "tether", such that loop formation induced by the binding of the Lac repressor to its two operators results in a quantifiable reduction of the bead's motion. We record the trajectory of looping and unlooping for each DNA molecule as a function of time in thermal equilibrium (Fig. 3.1**c**).

These trajectories contain a wealth of information about the DNA-protein interactions in our system. One such quantity that we will focus on here is the looping J-factor, $J_{loop} = (1\,\text{M})e^{-\beta \Delta F}$, that encapsulates the thermodynamic cost ΔF to deform the DNA (and possibly the protein) into the looped conformation, related to the cyclization J-factor often used to measure the flexibility of DNA *in vitro* [81]. We previously measured J_{loop} for looping constructs generated by a library of DNAs with different loop lengths and sequences (Fig. 3.1**c** and Fig. 3.18 in Section 3.4.1), to examine how DNA mechanics affect the energetics of loop formation [61, 107]. We showed that this library of constructs allowed us to tune J_{loop} over two orders of magnitude.

Here, we will instead focus on the looping and unlooping "lifetimes" (durations; see Fig. 3.1**c**) for this same library of DNAs. We find that the loop breakdown process at the DNA-protein interface is sensitive to the whole loop's deformation, with both looping and unlooping kinetics exhibiting rather simple forms of scaling with the looping free energy. Such a dependence has not, to our knowledge, been previously reported experimentally [15, 16] or considered in standard physical models for DNA looping [108–112], and suggests DNA looping as a member of a broader class of phenomena where applied force [113–117] or internal stress stored in polymers [2, 106, 118, 119] modulates biochemical reaction rates. Moreover, this result implies possible influence of DNA mechanics on evolution, since both the speeds of turning gene regulation *on* and *off* may be critical for fitness. We provide an explanation for the molecular origins of this dependence and develop a novel theory of looping kinetics, allowing us to probe experimentally inaccessible details of the looping pathway and the looping transition state.

In this work, we use the common single-molecule analysis technique of half-amplitude thresholding (see details in Sec. 3.4.1) to obtain distributions of the amount of time each TPM tether spends in the looped or unlooped state, called looped or unlooped lifetimes.

We begin by developing a simple kinetic framework for understanding what the measured state lifetimes tell us about the underlying physics of the system, with the basic elements given in Fig. 3.2. As discussed in more detail in Sec. 3.2, using standard kinetic analyses we can express the mean unlooped state lifetime, $\langle \tau_{unlooped} \rangle$, in terms of the repressor concentration $[R]$ and the rate constants diagrammed in Fig. 3.2, as

$$\langle \tau_{unlooped} \rangle = \frac{\left(1 + \frac{k_{off}^A}{[R]k_{on}^A}\right)\left(1 + \frac{k_{off}^B}{[R]k_{on}^B}\right)}{k_{on}^{\alpha}\frac{k_{off}^A}{[R]k_{on}^A} + k_{on}^{\beta}\frac{k_{off}^B}{[R]k_{on}^B}}. \tag{3.1}$$

Note that $\langle \tau_{unlooped} \rangle$ contains two different kinds of rates: $\{k^A_{on/off}, k^B_{on/off}\}$, for the binding/unbinding of the first repressor head to the DNA (We are distinguishing between binding and unbinding to operator A versus operator B, because several operators with different affinities for repressor have been described), as well as $\{k^\alpha_{on}, k^\beta_{on}\}$ for the binding of a second operator when the repressor has already bound the first one, which we here allow to differ from the rates for the initial binding event (see Fig. 3.2). On the other hand, $\langle \tau_{looped} \rangle = \frac{1}{k^\alpha_{off}+k^\beta_{off}}$ contains only the loop-affected dissociation rates $\{k^\alpha_{off}, k^\beta_{off}\}$, which we made no *a priori* assumptions and allow to differ from the simple unbinding events $\{k^A_{off}, k^B_{off}\}$. Experimental measurements of the unlooped and looped lifetimes then tell us how looping affects $k^{\alpha/\beta}_{on}$ and $k^{\alpha/\beta}_{off}$, respectively. Regulation of association rates by flexible linkers and polymer ring closure rates have been discussed in the framework of confined diffusion [108–112], and the effect of confined diffusion from the elastic DNA-repressor loop is likely to dominate $\{k^\alpha_{on}, k^\beta_{on}\}$ in our case as well. However, dissociation rates are usually thought of as local phenomena and dependent only on the interaction strength at the molecular interface, a hypothesis implicitly used in previous works on DNA looping kinetics [15, 16, 106, 120]. In contrast, in force spectroscopy experiments, an applied force changes a reaction free energy profile by adding a linear term to it. As a result, the equilibrium constant of the reaction, as well as both the "on" and "off" rates (e.g., association and dissociation of chemical bonds, folding and unfolding of RNA or nucleosomes), depend on the pulling force [113–117]. With our kinetic measurements, we can address the question of to what degree the dissociation process (i.e. the looped lifetime) is simply a local interaction and to what degree it is affected by the elastic deformation energies of the protein and DNA chain.

As exemplified by Fig. 3.3**a** for sequence dA (see Fig. 3.1**b**), the unlooped and looped state lifetimes extracted from our TPM data show a modulation with loop length, just as the J-factor (equivalently, the deformation free energy of the system) does. The other sequences are similarly plotted in Fig. 3.21 in Sec. 3.4.2 and exhibit more complex behavior when loop length varies more than one helical repeat. While the unlooped and looped state lifetimes are complicated functions of the loop length and DNA sequence, they are approximately monotonic when plotted versus the J-factor, as shown in Fig. 3.3**b**. Moreover, this behavior is roughly independent of both loop sequence and, within the range of lengths examined here, loop length, as shown in Fig. 3.3**c,d**: unlooped and looped state lifetimes for five different sequences spanning one to two helical periods of DNA all follow the same trend with J. Since J is known to be a function of loop length and DNA sequence, it can be viewed as encompassing the effects of the polymer parameters (within the range examined here) on the looping dynamics.

In contrast to the common view that dissociation rates are local phenomena only, these data suggest that the loop dissociation and association kinetics are both regulated by J_{loop}. We note that the looping J-factor is sometimes interpreted as effective cohesive-end concentration or effective repressor concentration, and increasing effective concentration is thought to facilitate association kinetics [121, 122]. However, this concept does not explain how the dissociation kinetics are modulated by an effective concentration: according to the simple kinetic framework discussed above, the looped-state lifetime should not depend on repressor

concentration $[R]$. If we take the effective concentration interpretation of J_{loop} literally, the fact that the dissociation kinetics, *i.e.* the looped lifetime shown in Fig. 3.3, depend on J_{loop} is inconsistent with this framework. Explaining the dependence of the dissociation kinetics on J_{loop} requires a different interpretation of J_{loop} beyond effective concentration and more akin to how applied force distorts bonding free energy in force spectroscopy experiments [113–117]. We first apply transition state theory to obtain some intuition about how J_{loop} can modulate loop formation and breakdown rates, then turn to a more sophisticated framework that more explicitly models the polymer mechanics. We note that our analysis based on free-energy landscapes is theoretically equivalent to expressing the effect of polymer deformation in terms of force and torque acting on the bond [106, 118, 119], and we used the free-energy treatment because of its conceptual simplicity.

The magenta curve in Fig. 3.2**c** shows a pathway between one unlooped state, where operator A is bound, and the looped state with both operators bound. The transition state on this path has an unknown structure and a total free energy $F_{transition}$. The activation energies for the forward and reverse transitions are given by $\Delta F^{\ddagger}_{unloop} = (F_{transition} - F_{unlooped})$ and $\Delta F^{\ddagger}_{loop} = (F_{transition} - F_{looped}) = (\Delta F^{\ddagger}_{unloop} - \Delta F + E_B)$, where ΔF is the free energy of deforming the DNA into the looped state and E_B is the favorable free energy of binding operator B. The loop formation and breakdown rates are given by transition state theory to be

$$k^{\beta}_{on} = k^{\beta}_0 e^{-\beta \Delta F^{\ddagger}_{unloop}} \tag{3.2}$$

and

$$k^{\beta}_{off} = k^{\beta}_0 e^{-\beta \Delta F^{\ddagger}_{loop}} = k^{\beta}_0 e^{-\beta(\Delta F^{\ddagger}_{unloop} - \Delta F + E_B)}, \tag{3.3}$$

where k^{β}_0 absorbs contributions from diffusivity and the shape of the free energy pathway. Similar equations can be derived for k^{α}_{on} and k^{α}_{off} (Sec. 3.2). Combining these with the equations given above for $\langle \tau_{unlooped} \rangle$ and $\langle \tau_{looped} \rangle$, we find that the mean unlooped and looped lifetimes will scale as

$$\langle \tau_{unlooped} \rangle \propto e^{\beta \Delta F^{\ddagger}_{unloop}} \tag{3.4}$$

and

$$\langle \tau_{looped} \rangle \propto e^{\beta(\Delta F^{\ddagger}_{unloop} - \Delta F + E_B)}. \tag{3.5}$$

Given the data in Fig. 3.3, we must conclude that not only $\Delta F^{\ddagger}_{unloop}$, but also $\Delta F^{\ddagger}_{loop}$, are determined by the DNA deformation energy, encapsulated in ΔF. We note that the looped and unlooped state lifetimes in Fig. 3.3**c,d** scale roughly linearly with J when plotted in log-log scale, and so we can obtain an approximate form of the relationship between τ and J (equivalently, ΔF). We find that $\langle \tau_{unlooped} \rangle \propto J^{-0.48 \pm 0.03}_{loop}$, and $\langle \tau_{looped} \rangle \propto J^{0.35 \pm 0.02}_{loop}$. This provides some intuition into how far the unknown transition state is, in terms of the elastic deformation ΔF, from the looped and unlooped states. Moreover, if the DNA chain were bent and twisted almost all the way into the needed shape before binding of the second operator, the unlooped

lifetime would scale as J_{loop}^{-1} and the looped lifetime would be independent of J_{loop} and chain length, a very different kinetic behavior from what we find here. This change in the J_{loop} dependence could shift the kinetic lifetimes an order of magnitude over just one decade in loop length.

It is clear that this is not the case for our system; that is, our finding that both the looped and unlooped lifetimes are dependent upon J_{loop} reveal that the process of going from the unlooped state to the transition state encompasses some, but not all, of the elastic deformation of the chain. Instead, the scaling exponents for τ versus J give an indication of the degree of release in elastic strain when moving from the looped state to the transition state. This model stands in contrast to previous models of ring closure that describe the DNA looping process as the first-passage time of the two ends coming within an approximately zero distance of separation [108–110].

To explore which elastic energies from the DNA and protein may be contributing to the reaction landscape of Fig. 3.2c, we have developed a molecular-level model for the looping and unlooping processes. We reason that the free energy landscape has contributions from bending and twisting deformations of the DNA and a binding energy (between the free repressor binding domain and the empty operator) of finite interaction length, as described in Fig. 3.1a. This results in a distance from the looped state to the transition state, similar to the ideas of finite-scale chemical bonding in Refs. [113, 114].

The DNA loop region is modeled as a wormlike chain, which describes the polymer as an elastic thread that is subjected to thermal fluctuations [123]. While there is still considerable debate about the elastic nature of DNA at short lengths [43, 124], the wormlike chain model has a clear physical basis and its application resulted in reasonable value for the persistence length. The bending energy for a specific conformation is given by

$$\beta E_{bend} = \frac{L_p}{2} \int_0^L ds \kappa(s)^2, \tag{3.6}$$

which depends on the square of the local curvature $\kappa(s)$. For a specific polymer conformation, the local curvature is equivalent to the inverse of the radius of a circle that is tangent to the curve (*e.g.* a straight chain segment has zero curvature and a tangent circle with infinite radius). This quadratic bending energy is consistent with linear elasticity theory of a thin elastic beam with a bending modulus $k_B T L_p$ (where L_p is the persistence length). The polymer conformational free energy $F_{conf}(r)$ gives the free energy for fixing the end-to-end distance of the two operators to be r, incorporating both the bending deformation energy and the entropy of different DNA conformations. We show some example configurations from a Monte Carlo simulation in Fig. 3.4a. To determine $F_{conf}(r)$, we find the Green function $G(r)$, which gives the probability of the two ends being a distance r apart, by summing over all possible paths and weighting each by $e^{-\beta E_{bend}}$. We have previously derived the exact result for the Green function [125] and use this result to calculate the conformational free energy $\beta F_{conf}(r) = -\log\left[r^2 G(r)\right]$.

The binding of the operators to the Lac repressor requires proper orientational alignment between the binding face of the operators and the Lac binding domains. The intervening DNA length determines the undeformed orientation of the DNA helix at the operator, and proper alignment for binding incurs energetically

costly twist deformation upon rotating the DNA into its proper orientation (which is in fact the origin of the modulation of J with loop length, noted in the text accompanying Fig. 3.3a). We define the twist angle θ to be the angle of rotation about the DNA axis at the unbound operator away from the ground-state untwisted angle (*i.e.* $\theta = 0$ is untwisted). We consider a simple model for the twist free energy $\beta F_{twist}(\theta) = \frac{L_t}{2L}\theta^2$, which is quadratic in the local twist deformation and evenly distributes the twist deformation over the length of the DNA between the operators. The twist persistence length L_t represents the resistance to twist deformation. This model neglects the geometric coupling between twist and writhe of the chain, which becomes more relevant at longer chain lengths where out-of-plane conformations are not prohibited by the bending deformation energy [125, 126].

The binding free energy, which drives the formation of the looped state, is modeled as a potential well with depth ϵ_0 and an interaction length scale δ. The separation of the two operators at the surface of the DNA strands is given by $r_a(r, \theta) = \sqrt{(r-a)^2 + a^2 - 2(r-a)a\cos(\theta - \theta_{op})}$, where r is the end-to-end distance of the DNA strands and θ is the twist angle at the unbound operator (see Fig. 3.9 for an illustration). The preferred twist angle $\theta_{op} = 2\pi(L/L_{turn}) + \theta_0$ gives the twist angle that orients the empty operator to face the Lac repressor binding domain, where θ_0 defines the twist angle necessary for docking DNA into Lac repressor even at integer DNA helical repeats. The representative images in Fig. 3.4a show how the DNA twists to orient the Lac repressor with the unbound operator as the ends are brought together. The DNA structure dictates the cross-sectional radius a (assumed to be $a = 1\,\text{nm}$) [127] and the helical length L_{turn} (assumed to be $L_{turn} = 10.46\,\text{bp} \times 0.34\,\text{nm}$) [125]. The binding free energy F_{bind} is then given by

$$\beta F_{bind}(r, \theta) = \begin{cases} \frac{-2\epsilon_0}{1+\exp\left[\frac{r_a(r,\theta)}{\delta}\right]}, & r > 2a, \\ \infty, & r \leq 2a, \end{cases} \tag{3.7}$$

which includes a steric cutoff at $r = 2a$ to account for the overlap of DNA backbone segments. This simple binding model aims to model the basic interaction between the DNA operator and Lac repressor by introducing only the binding affinity ϵ_0 and the interaction length δ to capture the physical interaction. More detailed models of interaction could include more molecular detail, but our goal is to give the simplest representation of binding without introducing additional parameters that do not have well-defined values.

The three thermodynamic contributions $F_{conf}(r)$, $F_{twist}(\theta)$, and $F_{bind}(r, \theta)$ combine to give the total free energy landscape $F_{total}(r, \theta)$, as shown in Fig. 3.4a for $L = 101\,\text{bp}$ and parameters $\epsilon_0 = 23.5$ (in $k_B T$ units), $\delta = 1.3\,\text{nm}$, $L_p = 48\,\text{nm}$, $L_t = 15\,\text{nm}$, and $\theta_0 = 0.003\,\pi$. We find the minimum free energy path from the looped state X, over the transition state Y, to the unlooped state Z, and plot each of the three free energy contributions, as well as the total free energy along this path, in Fig. 3.4b. From transition state theory, the looped lifetime is simply proportional to $e^{-\beta\Delta F_{loop}^{\ddagger}}$, and the unlooped lifetime is proportional to $e^{-\beta\Delta F_{unloop}^{\ddagger}}$, as given in Eqns. 3.58 and 3.59 above. We use the more sophisticated Fokker-Planck formalism, and treat the reaction from the looped to unlooped state (and vice versa) as diffusion on a one-dimensional potential energy landscape, given by $F_{total}(r)$ along the minimum free energy path shown in Fig. 3.4b. Twist

angle relaxation is much faster than changes in the end-to-end distance [128], and we avoid introducing an additional, poorly characterized parameter for the twist angle diffusivity by reducing the problem down to one-dimension. We calculate the mean looped lifetime as the average first passage time from anywhere in $r < r_Y$ to leaving the transition state at $r = r_Y$, and similarly for the mean unlooped lifetime. The looping J-factor is calculated from the polymer free energy difference,

$$J_{loop} = (1\,\mathrm{M})e^{-\beta \Delta F} = (1\,\mathrm{M}) \exp[-\beta(F_{poly}^{loop} - F_{poly}^{unloop})]. \tag{3.8}$$

Here, F_{poly}^{loop} and F_{poly}^{unloop} are calculated by averaging only the polymer elastic energies ($F_{conf} + F_{twist}$) (*i.e.* excluding the binding energy) over the end-to-end distance r smaller and larger than $r = r_Y$, respectively, with a Boltzmann weight given by $e^{-\beta F_{total}(r)}$. We refer the reader to Sec. 3.3 for more details.

To compare these theoretical results with the experimental results of Fig. 3.3, we find model parameters for the elastic parameters L_p, L_t, which could vary with DNA sequence, and the binding parameters ϵ_0, δ, and θ_0, which should be consistent across all sequences with the same operators. We obtain $\theta_0 = 0.003\,\pi$ by looking at the peaks in the J_{loop} data (see Fig. 2 in Ref. [107]), which occur when the twist-free orientation $2\pi L/L_{turn}$ is aligned with θ_0. The model is able to reproduce the basic qualitative features of the data across a range of parameters, and we chose values of $\epsilon_0 = 23.5$ (in $k_B T$ units), and $\delta = 1.3\,\mathrm{nm}$ as representative of a good fit to the data across all five sequences for the set of operators used here. These parameters are within the expected range, given the size of the Lac repressor arm (around $3 - 4\,\mathrm{nm}$ from the crystal structure) [96] and the binding energy of the repressor to DNA of approximately $16\,k_B T$ [129]. We then varied the elastic parameters to find the best fit for each sequence, obtaining values of the persistence length L_p ranging from 48 to 51 nm and the twist persistence length L_t ranging from 10 to 70 nm. The values for L_p are close to the canonical value for dsDNA of $L_p = 53\,\mathrm{nm}$ [130]. While our twist persistence differs from the canonical value of $L_t = 110\,\mathrm{nm}$ [128], we note that our twist model is much simplified. We do not include the details of the end orientations, twist angle entropic effects, and twist-writhe coupling, all of which could lead to the lower value of L_t that we determined.

Theoretical predictions for the unlooped and looped mean lifetimes are shown in Figs. 3.4c and d, using the same values as in Fig. 3.4a. The theoretical lifetimes (black curves) exhibit an approximate power-law trend with J_{loop} for the lengths ranging from 89 to 115 bp. Notably, the value of J_{loop} exhibits both oscillations and an average increase as the length is increased from 89 to 115 bp. In this regard, the quantity J_{loop} serves as a critical determinant of the looped and unlooped lifetimes.

The major novel feature introduced in our molecular model for looping is the treatment of the protein binding energy that has a well depth of ϵ_0 and an interaction radius of δ. These parameters are specifically dependent upon the properties of the protein and the operator binding interface. In addition to the data shown in Fig. 3.3, we also analyzed TPM trajectories with a different set of operators, specifically with the O_1 operator replaced by the slightly weaker O_2. These data are plotted as blue dots in Fig. 3.5. We have previously determined the energetic difference between these two operators to be $1.5\,k_B T$ [129]. Since only

one operator's affinity was changed, we would expect the resulting value of ϵ_0 to be reduced by $0.75\,k_BT$. (We note that in these operator-swap experiments, the sequence of the loop was somewhat altered as well, but not its total length, and as such we expect most or all of the change to be due to the difference in the binding well depth ϵ_0). Our model prediction, given in black in Fig. 3.5, clearly agrees well with the experimental results when ϵ_0 is reduced by $0.75k_BT$ and all other parameters are kept the same as in Fig. 3.4.

Consistent with our theory, only the looped lifetimes are affected by the change of operator. In the free energy plot in Fig. 3.4b, we see that the binding energy (blue) only begins to affect the total free energy curve (black) once the configuration is to the left of the transition state; *i.e.* it is in the looped state. Likewise, we see that the twist energy (red) only begins at end-to-end distances less than the transition state end-to-end distance r_Y. Thus, our molecular level model has given us clear insight into the elastic deformation present at the transition state, and this agrees well with our experimental measurements.

The other major parameter introduced for the binding energy was a finite length-scale for the DNA-protein interaction. This parameter is critical to explain our findings that both the looped and unlooped lifetimes depend upon the J-factor. The finite length-scale of interaction δ affects the transition state Y by changing the end-to-end distance and twist angle at which this state occurs. For large δ, the transition state would occur at a farther end-to end distance and thus exhibit a notable release of deformation energy as compared to the looped state X, leading to a dependence of the looped lifetime on J_{loop}. This parameter δ phenomenologically models both the size [96] and flexibility [131] of the protein mediating the loop, and could also account for other effects such as electrostatic interactions or nonspecific binding leading to sliding along the DNA chain [132]. Thus, in experiments with proteins of smaller size or flexibility than Lac repressor, we would expect a decreased scaling exponent and hence decreased dependence of the looped lifetime on J_{loop}. We will explore the effect of this interaction distance further in an upcoming manuscript.

Using the single-molecule technique of tethered particle motion to examine looping and unlooping lifetimes by the classic Lac repressor looping protein, we have shown here that both the looped and unlooped lifetimes depend upon the J-factor, indicating that the dissociation rate is dependent upon the DNA and protein elasticity. These findings are unexpected based on the common treatment of the J-factor as an effective protein concentration, and have been ignored by previous studies of DNA looping. We also note that the J-factor-modulated state lifetimes, having a 1 to 10-minute dynamical range, are comparable to *E. coli*'s cell division time. The state lifetimes are sensitive to how they scale with J, and within a decade of loop length variation the response times can change an order of magnitude. It is therefore interesting to explore how DNA mechanics modulates the *in vivo* looping and unlooping rates and assess its influence on how individual cells respond to nutrient fluctuations. To explain our experimental results, we have developed a molecular-level model that accounts for the role of both the polymer and protein deformation in DNA looping and unlooping kinetics. This model includes a simple but straight-forward and physically-derived picture for the three energies necessary to explain looping in short, stiff chains: bending, twisting, and binding. The binding energy used allows us to incorporate the protein elasticity through the introduction of a finite

length-scale of interaction that modulates the degree of favorable binding depending upon the end-to-end distance. We find the lifetimes calculated from this model to be in good agreement with our experimental results with realistic physical parameters, and that the model provides additional insights into the properties of the transition state and how the elastic energy changes during the course of the looping reaction. Finally, we note that long-range ordering of opening and closing kinetics by the system's free energy landscape should be a general framework that goes beyond the DNA or repressor-specific variables, and can be applied to other elastic systems such as ligand-receptor reaction [133] or protein assembly [134], where flexible tethers are important for the biological functions.

Note: TPM experiments were performed as previously described [61, 107]. Details of our implementation of the half-amplitude thresholding procedure used to quantify looped and unlooped lifetimes are given in Sec. 3.4.1, and a comparison of our results to those in previous studies using TPM to measure Lac repressor looping and unlooping rates, showing good agreement between our results and these previous studies, is given in Sec. 3.4.1. Experimental errors are reported as standard errors on the means, calculated according to the bootstrapping method described in Sec. 3.4.1.

a Molecular picture of protein-mediated DNA looping

b Schematic of the tethered particle motion (TPM) assay

c Looping trajectory and dwell time measurement
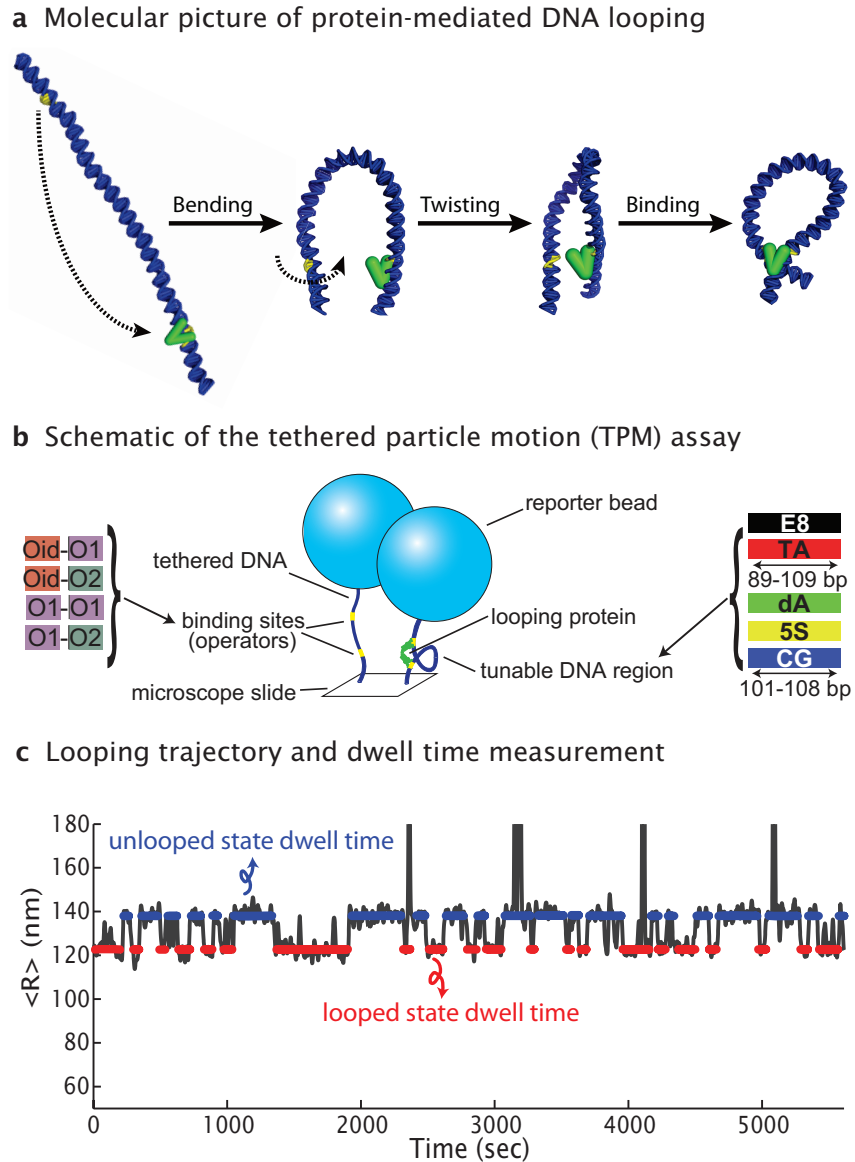
Figure 3.1: **DNA looping dynamics measured by tethered particle motion. a,** Loop formation requires the DNA chain to bend and twist to bring the binding sites together and properly orient them.**b,** The TPM setup, in which single DNA molecules tether microscopic beads to a slide. Looping due to the Lac repressor binding the two operators on the DNA reduces the bead's motion. **c,** Sample TPM trajectory, $\langle R \rangle$ versus time, recorded from a single tether and segmented into unlooped (blue) and looped (red) states. The lifetime of a state is how long a trajectory remains in that state before transitioning to a different one.
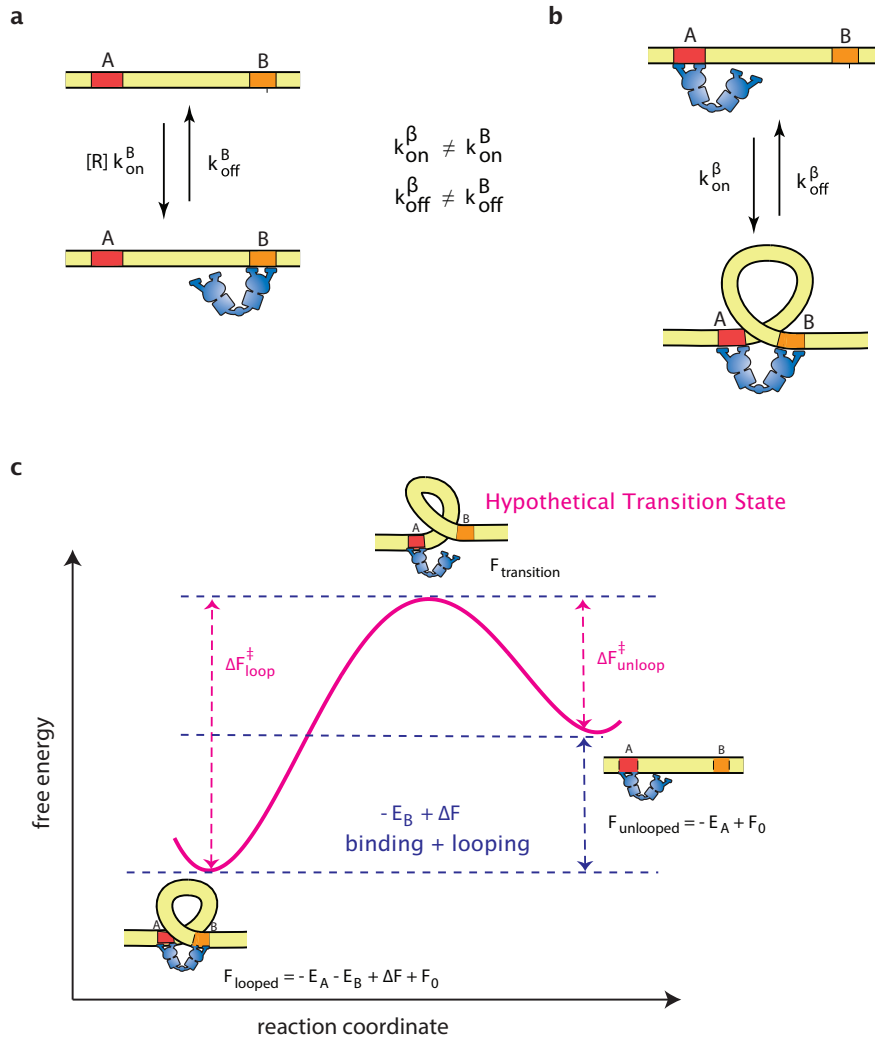
Figure 3.2: **Kinetic framework of protein-mediated looping. a,** Repressor association with operator B is a pseudo-first-order reaction with rate $[R]k_{on}^B$. Dissociation is a zeroth-order reaction with rate k_{off}^B. **b,** Association of the repressor bound at operator A with the unbound operator B has a rate k_{on}^β, and dissociation from operator B has a rate k_{off}^β. Note that here we are distinguishing between the binding/unbinding of the first repressor head (rate constants $k_{on}^A, k_{off}^A, k_{on}^B, k_{off}^B$) and the binding/unbinding of the second repressor head to the same DNA (rate constants $k_{on}^\alpha, k_{off}^\alpha, k_{on}^\beta, k_{off}^\beta$). **c,** The reaction curve (magenta) shows a hypothetical transition state, with unknown free energy, controls both the forward the reverse reactions. What is known from equilibrium measurements is the difference between the unlooped and looped states, $-E_B + \Delta F$ (binding energy plus the penalty for looping deformation).
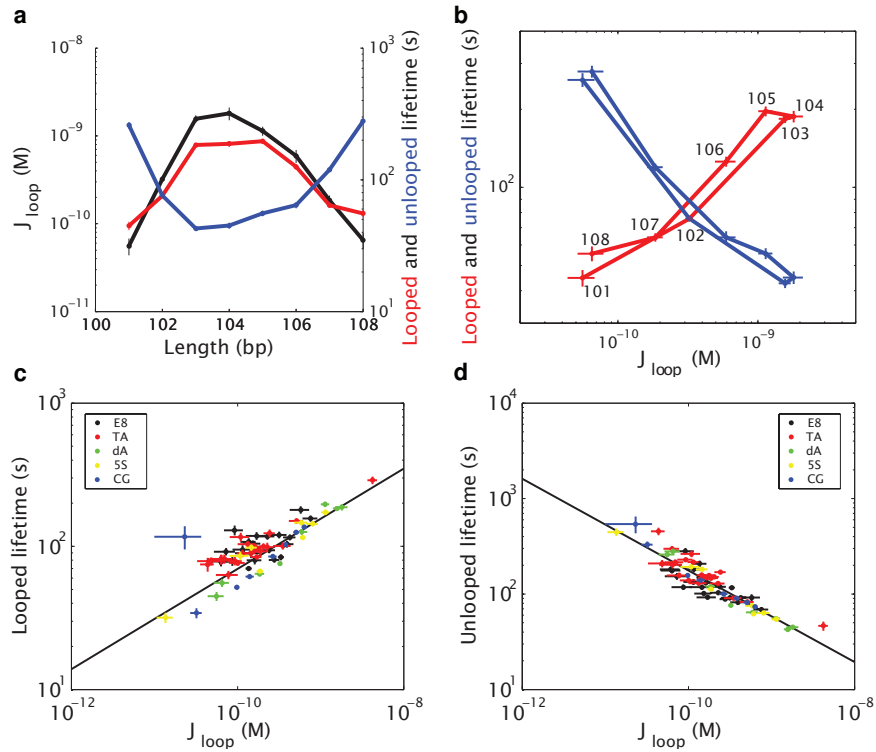
Figure 3.3: **Experimentally measured state lifetimes**. **a,** Looping J-factor (black), mean unlooped lifetime (blue), and mean looped lifetime (red), for one helical period of the "dA" sequence. **b,** The same data as in (a), but with lifetimes plotted versus the J-factor and the loop length (in bp) marked for the looped lifetime curve. **c,** Mean looped state lifetime and **d,** mean unlooped state lifetime versus J_{loop}, for one helical repeat each for three sequences ("dA", "5S", "CG"), and two helical repeats for two sequences ("E8","TA").
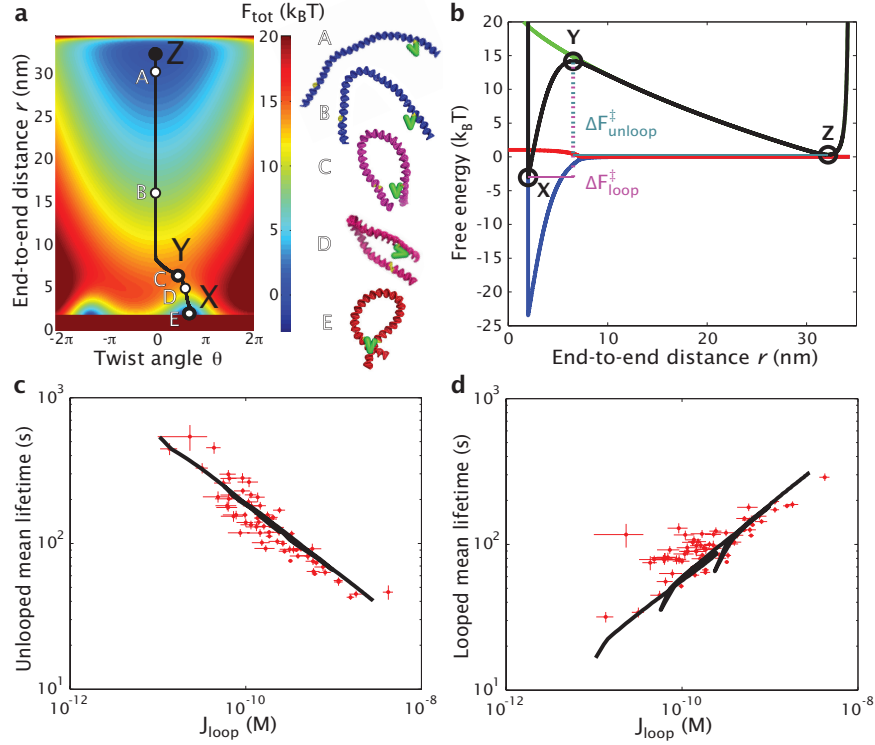
Figure 3.4: **Molecular model for DNA looping. a,** Total free energy surface versus end-to-end distance r and twist angle θ. In this plot, $L = 101$ bp and the parameters are $\epsilon_0 = 23.5$ (in $k_B T$ units), $\delta = 1.3$ nm, $L_p = 48$ nm, $L_t = 15$ nm, and $\theta_0 = 0.003\,\pi$. The black curve indicates the minimum free energy path between the looped state (X) and the unlooped state (Z), passing through the transition state (Y). Representative DNA conformations (as predicted by Monte Carlo simulation) at five different end-to-end separations are shown to the right of the free energy surface, where the degree of twisting is indicated by the DNA coloration ranging from blue for $\theta = 0$ to red for $\theta = 0.678\,\pi$. **b,** Free energy along the minimum free energy path. The total free energy (black) is a combination of the polymer free energy (green), the twisting free energy (red), and the binding free energy (blue). The free energy barriers to leave the looped and unlooped states are $\Delta F^{\ddagger}_{loop}$ and $\Delta F^{\ddagger}_{unloop}$, respectively. **c,** Unlooped lifetime behavior. The experimentally determined unlooped lifetimes (red dots) are plotted versus J_{loop}, with the black line corresponding to the theoretical prediction as L is varied from 89 to 115 bp. **d,** Looped lifetime behavior. The looped lifetimes from the experiments (red dots) and theory (black line) are plotted using the same parameters as in panel **c**.
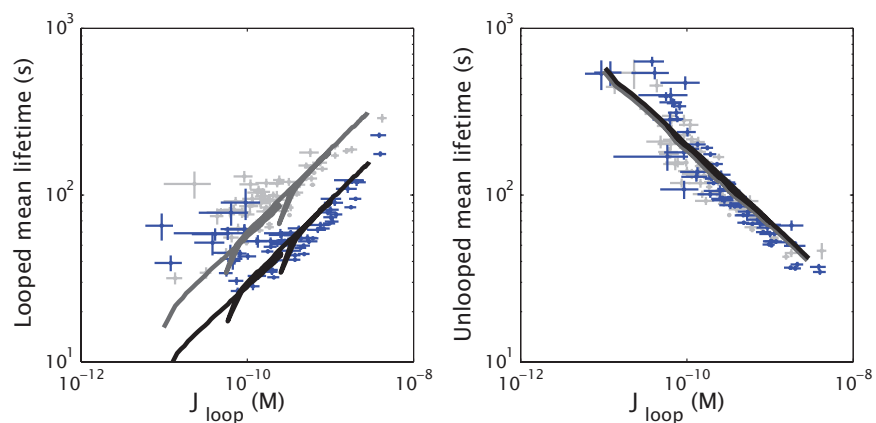
Figure 3.5: **Changing operator affinity shifts looped lifetimes. a,** Looped lifetime behavior. The experimentally determined looped lifetimes (blue) for all five sequences with O_{id} and O_2 operators, instead of O_{id} and O_1 as in Fig. 3.3, are plotted versus J_{loop}, with the black line corresponding to the theoretical prediction as L is varied from 89 to 115 bp. All the model parameters are the same as in Fig. 3.4 except ϵ_0 has been reduced to $22.75\,k_BT$. For reference, the light gray dots are the data from the O_{id} and O_1 operators and the dark gray line is the theory curve from Fig. 3.4. **b,** Unlooped lifetime behavior. The unlooped lifetimes for the O_{id} and O_2 operators (blue) and theory (black line) are plotted, the same as in panel **a**.

3.2 Theoretical Analysis of Protein-Mediated DNA Looping

In these sections we describe in detail the kinetics scheme we introduce in 3.1, and give a fuller account of its reconciliation with our experimental results.

3.2.1 Kinetics scheme for protein-mediated looping.

States of the looping system and assumptions in our derivation.
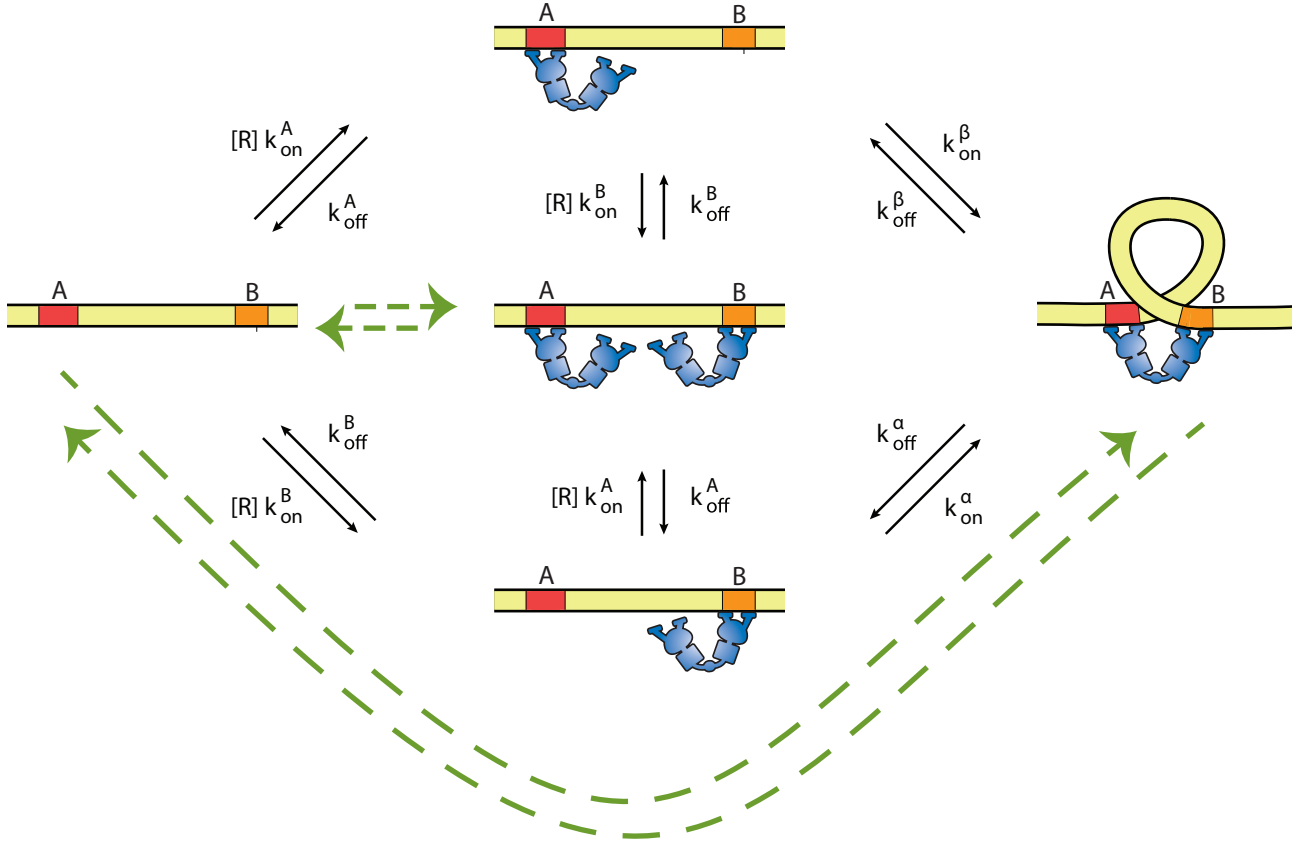
Figure 3.6: **Kinetic scheme for looping.** In thermal equilibrium, for a DNA (yellow line) with two Lac repressor binding sites A and B (orange and red boxes) exposed to a solution of Lac repressor molecules (blue shapes), the system transitions between the four unlooped states and the looped state discussed in the text. Kinetic rates governing each transition are labeled above the arrows indicating that transition. Each transition is either first-order (characterized by one rate constant k with unit $\frac{1}{time}$) or pseudo-first-order (characterized by a rate $[R]k$ with units $\frac{1}{time}$). We assume that the concentration of repressor is much larger than the concentrations of its binding sites, such that the concentration of free repressor in solution is effectively constant over the course of the experiment (see [61] for details and a discussion of when this assumption breaks down). Therefore the second-order binding steps of the form $R + A \rightarrow RA$ can be simplified to pseudo-first-order. Examples for the forbidden pathways discussed in (iv) are shown in green arrows.

We start by setting up some basic parameters of the system and stating some assumptions required for

our derivation. As described in our previous work in which we derived a statistical mechanical model for the probability of loop formation by the Lac repressor [61, 76], our system can adopt five coarse-grained states, illustrated in Fig. 3.6: no protein bound on DNA, either operator (i.e. binding site) A or B being bound, both operators bound, and the looped state. Their relative statistical weights are given by

$$w_0 = 1, \tag{3.9}$$

$$w_{A-bound} = \frac{[R]}{K_A}, \tag{3.10}$$

$$w_{B-bound} = \frac{[R]}{K_B}, \tag{3.11}$$

$$w_{AB} = \frac{[R]^2}{K_A K_B}, \tag{3.12}$$

$$w_{loop} = \frac{[R] J_{loop}}{2 K_A K_B}. \tag{3.13}$$

The relationship between the looping probabilities measured in TPM, p_{loop}, the repressor-operator dissociation constants for the two operators that flank the loop, K_A and K_B, and the looping J-factor of the DNA in the loop, J_{loop}, can be described as

$$p_{loop} = \frac{w_{loop}}{w_0 + w_{A-bound} + w_{B-bound} + w_{AB} + w_{loop}} \tag{3.14}$$

$$= \frac{\frac{[R] J_{loop}}{2 K_A K_B}}{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B} + \frac{[R] J_{loop}}{2 K_A K_B}}. \tag{3.15}$$

where $[R]$ is the concentration of Lac repressor. J_{loop} is related to the looping free energy by $J_{loop} = 1 \text{ M } e^{-\beta \Delta F}$. "1 M" is the standard state calculated by discretizing the solution volume to molecular size lattice and the exponential factor is the occupation number given by Boltzmann distribution. K_A, K_B and $[R]$ are also in molar. Generally, the "looping free energy" contains contributions from the bending, twisting energy change and entropic change from both the DNA and the protein. Figure 3.6 summarizes our kinetic model for the transition state theory. We use several simplifications as follows.

(i) We assume that two LacI binding to operator A and B are independent, although we know the binding of LacI does bend the operators and has a small effect on the tether RMS (as reported in [61]), and in principle binding of two LacI could couple. (Some researchers discussed protein-protein coupling through both binding to DNA.)

(ii) We assume at all the LacI concentrations we work with, there is negligible nonspecific binding of LacI to DNA. In fact, the dissociation constant of the LacI for its specific binding sites (operators) is about 10^{-9} times smaller than non-specific binding [135]. The dissociation constant for non-specific binding is in the $10^{-5} M$ range, which is much higher than the highest repressor concentration we use ($10^{-8} M$). Non-specific

binding, if it occurred, might shorten the DNA tethers and result in higher local concentration of adsorbed LacI available to bind to DNA. Ref. [100] reported that nonspecific binding of phage Lambda cI protein to DNA shortens loop length and broadens its distribution in a concentration-dependent manner. This also changes and broadens the observed k_{on} and k_{off} of cI to an operator (which are supposed to be independent of cI concentration). However, in our system the shortening of tether by protein binding is small and is not concentration dependent, as shown in [61]. Our LacI looping system therefore does not suffer from non-specific binding at the concentrations we work with, and we do not need concentration-dependent correction terms for the rate constants.

(iii) No depletion of LacI from the solution. In our experiments, LacI is present in excess over the DNA, so the LacI concentration in solution is essentially constant. Hence, the second-order reaction of a repressor binding to an operator (for example, $R + A \rightarrow RA$) becomes a pseudo-first-order reaction. All the reaction steps in Figure 3.6 are thus first-order or pseudo-first-order reactions.

The above assumptions were also used in the statistical mechanical model for looping probability (see further discussion of these issues in [61]). The observed looping probabilities fit well to this statistical mechanical model (see Ref. [61]), and gives us confidence that assumptions (i), (ii) and (iii) are indeed valid for our experimental conditions. We next consider the kinetics that govern the transition among these states and adopt a few assumptions commonly used in chemical kinetics analysis.

(iv) We assume transitions occur at one molecular junction at a time (with the probability of order $k\delta t$, where k is a reaction rate and δt is an infinitesimal time interval). Simultaneous actions at two junctions would be very unlikely (because for two events to happen at the same time the probability would be of order $kk'(\delta t)^2$). For example, the probability of both heads of LacI grabbing the operators simultaneously while the loop forms is negligible. One of its ends always has to bind first and wait for the polymer fluctuation to bring the other end close to the other operator. Similarly, two LacI binding to two operators simultaneously is a process of negligible significance. The same is true for simultaneous dissociation at two junctions or any other high-order transitions. The two example forbidden pathways are shown in Figure 3.6 in green arrows.

(v) As discussed in Sec. 3.4.1 and Fig. 3.18, with TPM we actually observe two different looped states, with distinct RMS values, that we call the "bottom" and "middle" looped states. They are likely looped states with different geometries and they may interconvert (see Ref. [136]). However, for the most part here we consider only the coarse-grained total looped state, because, as described in Ref. [107], the bottom state is rare, and we do not have much data on bottom state lifetimes for most constructs. Hence, our kinetic analysis focuses on the switching rates between the unlooped state and the dominant looped state.

Master equations.

From the kinetics pathways shown in Figure 3.6, the master equations of the system are given by

$$\frac{dp_0}{dt} = -([R]k_{on}^A)p_0 + k_{off}^A p_{A-bound} - ([R]k_{on}^B)p_0 + k_{off}^B p_{B-bound}, \tag{3.16}$$

$$\frac{dp_{A-bound}}{dt} = -(k_{off}^A + [R]k_{on}^B + k_{on}^\beta)p_{A-bound} + [R]k_{on}^A p_0 + k_{off}^B p_{AB} + k_{off}^\beta p_{loop}, \tag{3.17}$$

$$\frac{dp_{B-bound}}{dt} = -(k_{off}^B + [R]k_{on}^A + k_{on}^\alpha)p_{B-bound} + [R]k_{on}^B p_0 + k_{off}^A p_{AB} + k_{off}^\alpha p_{loop}, \tag{3.18}$$

$$\frac{dp_{AB}}{dt} = -(k_{off}^A + k_{off}^B)p_{AB} + [R]k_{on}^A p_{A-bound} + [R]k_{on}^B p_{B-bound}, \tag{3.19}$$

$$\tag{3.20}$$

and

$$\frac{dp_{loop}}{dt} = -k_{off}^\beta p_{loop} + k_{on}^\beta p_{A-bound} - k_{off}^\alpha p_{loop} + k_{on}^\alpha p_{B-bound}. \tag{3.21}$$

These probabilities must add up to one, that is,

$$p_0 + p_{loop} + p_{AB} + p_{A-bound} + p_{B-bound} = 1. \tag{3.22}$$

Application of detailed balance.

Before turning to a derivation of expressions for average lifetimes in terms of the rate constants in Fig. 3.6 and Fig. 3.6, we next establish some basic relationships between the statistical weights introduced in the previous section and the kinetic parameters in Fig. 3.6, which will be used in the next section.

In Fig. 3.6, k_{on}^α, k_{off}^α represent the on and off rate constants for LacI to bind to and dissociate from operator A, when its other end is bound to operator B. Similarly, k_{on}^β, k_{off}^β represent the on and off rate constants for operator B when the Lac repressor already binds to A with one end. As discussed above, we assume all the reaction steps are first-order or pseudo-first-order, so the units of k_{on}^α, k_{off}^α, k_{on}^β, k_{off}^β, k_{off}^A, k_{off}^B and $[R]k_{on}^A$, $[R]k_{on}^B$ are $\frac{1}{[time]}$. Because of DNA looping, we expect the Lac repressor on rates to operator A and B, k_{on}^α and k_{on}^β, to differ from the original on rates $[R]k_{on}^A$ and $[R]k_{on}^B$. In earlier work, J_{loop} was used interchangeably with the *effective* local concentration of LacI [121]. If we consider the looping reaction at operator A when the LacI has already bound to other site B, in the usual sense of intermolecular binding using an effective protein concentration $[I]$, we can write the reaction as $I + B - bound \rightarrow looping$, while the usual binding reaction of LacI onto site A is $R + A \rightarrow RA$. Making use of the statistical weights given in Eq. 3.9, 3.10, 3.11 and 3.13, the relative amount of the different species in these reactions are

$$\frac{[looping]}{[B-bound]} = \frac{w_{loop}}{w_{B-bound}} = \frac{J_{loop}}{2K_A}, \tag{3.23}$$

and

$$\frac{[RA]}{[A]} = \frac{w_{A-bound}}{w_0} = \frac{[R]}{K_A}. \tag{3.24}$$

Hence, we can identify $[I]$ with $\frac{J_{loop}}{2}$. This is why the looping J-factor is usually considered an effective repressor concentration.

J_{loop} affects how easily the empty site could find a repressor molecule in its close proximity. A large J_{loop} means that when one hand of the repressor is bound to one site, it is more easily brought to the other site because it is coupled to the DNA. Hence, the on rates k_{on}^{α} and k_{on}^{β}, which depend upon the chance for the operator to see a LacI when the other end of of the LacI has already bound to the other operator, should be regulated by the dynamics and strain of DNA loop formation. Indeed, loop formation kinetics have been modeled by diffusion in an effective energy landscape which is governed by the polymer stiffness and size [108–111]. On the other hand, the off rates k_{off}^{α} and k_{off}^{β} are in general difficult to model despite various efforts in the literature, because the detailed dynamical atomic position, orientation and electrostatic interactions (coupled by solvent) are not known. Here we do not exclude the possibility that the off rates k_{off}^{α} and k_{off}^{β} are also dependent on J_{loop} and are different from k_{off}^{A} and k_{off}^{B} as well.

If we look at the transitions between operator B being bound and the looped state, we can write down detailed balance as

$$k_{on}^{\alpha} \cdot p_{B-bound} = k_{off}^{\alpha} \cdot p_{loop}. \tag{3.25}$$

Similarly, between operator A being bound and the looped state, detailed balance says

$$k_{on}^{\beta} \cdot p_{A-bound} = k_{off}^{\beta} \cdot p_{loop}. \tag{3.26}$$

Inserting the statistical weights from Eq. 3.10, 3.11, and 3.13 to the above detailed balance relations leads to

$$\frac{k_{on}^{\alpha}}{k_{off}^{\alpha}} = \frac{p_{loop}}{p_{B-bound}} = \frac{w_{loop}}{w_{B-bound}} = \frac{\frac{[R]J_{loop}}{2K_A K_B}}{\frac{[R]}{K_B}} = \frac{J_{loop}}{2K_A}, \tag{3.27}$$

and

$$\frac{k_{on}^{\beta}}{k_{off}^{\beta}} = \frac{p_{loop}}{p_{A-bound}} = \frac{w_{loop}}{w_{A-bound}} = \frac{\frac{[R]J_{loop}}{2K_A K_B}}{\frac{[R]}{K_A}} = \frac{J_{loop}}{2K_B}. \tag{3.28}$$

On the other hand, we can also relate the dissociation constants K_A and K_B to the kinetic rate constants. The detailed balance between the state with nothing bound on DNA and operator A bound by one Lac repressor is

$$([R]k_{on}^{A}) \cdot p_0 = k_{off}^{A} \cdot p_{A-bound}. \tag{3.29}$$

The dissociation constant K_A by definition is the ratio of concentrations $\frac{[R][A]}{[RA]}$ for the reaction $R + A \rightleftharpoons RA$.

Since $\frac{[A]}{[RA]}$ is just $\frac{p_0}{p_{A-bound}}$, K_A is given by

$$K_A = \frac{[R]p_0}{p_{A-bound}} = \frac{k_{off}^A}{k_{on}^A}. \tag{3.30}$$

Similarly, the detailed balance between the state with nothing bound on DNA and operator B bound by one Lac repressor is

$$([R]k_{on}^B) \cdot p_0 = k_{off}^B \cdot p_{B-bound}. \tag{3.31}$$

Thus K_B has the form

$$K_B = \frac{[R]p_0}{p_{B-bound}} = \frac{k_{off}^B}{k_{on}^B}. \tag{3.32}$$

Associate the average lifetimes with rate constants.

Experimentally we cannot distinguish the four unlooped states because they yield similar RMS values. What are experimentally observable are the lifetimes of the looped state and the unlooped state. For each construct (specified by loop sequence, length and operators), measured at a particular LacI concentration, we therefore obtain distributions of the *Looped* lifetimes $p(\tau_{looped})$ and the *Unlooped* lifetimes $p(\tau_{unlooped})$, which give average lifetimes $< \tau_{looped} >$ and $< \tau_{unlooped} >$ for each construct, as well as looping probabilities, p_{loop}, which are the total time spent in the looped state divided by the total observation time. Because the system is either in the looped or unlooped state, the total time spent in each state, as well as the average lifetime spent in each state, are by definition associated with the probability in each state and are given by

$$\frac{< \tau_{looped} >}{< \tau_{unlooped} >} = \frac{p_{loop}}{1 - p_{loop}} = \frac{w_{loop}}{w_0 + w_{A-bound} + w_{B-bound} + w_{AB}} \tag{3.33}$$

$$= \frac{\frac{[R]J_{loop}}{2K_AK_B}}{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_AK_B}} \propto J_{loop}. \tag{3.34}$$

We will try to connect the rate constants in the model (because the rate constants are useful when we talk about transition state theory later) to the experimentally observed lifetimes. From the looped state to the unlooped states there are two paths, determined by rate constants k_{off}^α and k_{off}^β. The lifetime distribution assuming a one-step Poisson process (see [28] for an ion channel example) is

$$p(\tau_{looped}) = (k_{off}^\alpha + k_{off}^\beta)e^{-(k_{off}^\alpha + k_{off}^\beta)\tau_{looped}}. \tag{3.35}$$

Hence the average looped lifetime is

$$< \tau_{looped} >= \frac{1}{(k_{off}^\alpha + k_{off}^\beta)}. \tag{3.36}$$

Now, if the unlooped state is truly one single state, the average unlooped lifetime would be

$$< \tau_{unlooped} >^{single\ state} = \frac{1}{(k_{on}^\alpha + k_{on}^\beta)}. \tag{3.37}$$

The experimentally observed average lifetime has an inverse relationship with the rate constants. Moreover, this inverse relationship actually still holds, up to some proportionality constants, even when the state is composed of multiple microstates. To show this is the case, we return to the statement above that the unlooped state involves four indistinguishable states in TPM. Looking at Figure 3.6 we know that if the system starts within any of the four unlooped states, before it reaches the looped state, it can only follow the arrows with the these rate constants: k_{on}^α, k_{on}^β, $[R]k_{on}^A$, $[R]k_{on}^B$, k_{off}^A, k_{off}^B. Hence, we know only these rate constants will appear in the functional form of the unlooped lifetime distribution, even though we do not know the analytical solution. In other words,

$$p(\tau_{unlooped}) = \text{Function}(k_{on}^\alpha, k_{on}^\beta, [R]k_{on}^A, [R]k_{on}^B, k_{off}^A, k_{off}^B). \tag{3.38}$$

We next make use of the fact that lifetimes are associated with the looping probabilities to derive an expression for the average unlooped lifetime, by combining Eqs. 3.34, 3.45, 3.27, and 3.28 as follows. Starting from Eq. 3.34, $< \tau_{unlooped} >$ can be written as

$$< \tau_{unlooped} > = < \tau_{looped} > \frac{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B}}{\frac{[R]J_{loop}}{2K_A K_B}}. \tag{3.39}$$

We then plug in the expression of $< \tau_{looped} >$ given by Eq. 3.45 to the expression of $< \tau_{unlooped} >$ above and obtain

$$< \tau_{unlooped} > = \frac{1}{(k_{off}^\alpha + k_{off}^\beta)} \frac{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B}}{\frac{[R]J_{loop}}{2K_A K_B}} \tag{3.40}$$

$$= \frac{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B}}{(\frac{J_{loop}k_{off}^\alpha}{2K_A})(\frac{[R]}{K_B}) + (\frac{J_{loop}k_{off}^\beta}{2K_B})(\frac{[R]}{K_A})} \tag{3.41}$$

Next, we can use use the detailed balance relations for the α and β rate constants given by Eq. 3.27 and Eq. 3.28 and arrive at

$$< \tau_{unlooped} > = \frac{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B}}{(k_{on}^\alpha)(\frac{[R]}{K_B}) + (k_{on}^\beta)(\frac{[R]}{K_A})}. \tag{3.42}$$

Then we multiply both dividend and divisor with the factor $\frac{K_A K_B}{[R]^2}$ and arrive at the form

$$< \tau_{unlooped} > = \frac{\frac{K_A K_B}{[R]^2} + \frac{K_B}{[R]} + \frac{K_A}{[R]} + 1}{k_{on}^\alpha \frac{K_A}{[R]} + k_{on}^\beta \frac{K_B}{[R]}} = \frac{(1 + \frac{K_A}{[R]})(1 + \frac{K_B}{[R]})}{k_{on}^\alpha \frac{K_A}{[R]} + k_{on}^\beta \frac{K_B}{[R]}}. \tag{3.43}$$

We can also replace K_A and K_B with kinetic rate constants given in Eq. 3.30 and Eq. 3.32 and obtain

$$< \tau_{unlooped} > = \frac{(1 + \frac{k_{off}^A}{[R]k_{on}^A})(1 + \frac{k_{off}^B}{[R]k_{on}^B})}{k_{on}^\alpha(\frac{k_{off}^A}{[R]k_{on}^A}) + k_{on}^\beta(\frac{k_{off}^B}{[R]k_{on}^B})}. \tag{3.44}$$

We see that indeed $< \tau_{unlooped} >$ can be expressed in terms of only the set of parameters k_{on}^α, k_{on}^β, $[R]k_{on}^A$, $[R]k_{on}^B$, k_{off}^A, k_{off}^B. Also, even though we do not have an analytical solution for the probability distribution of the unlooped lifetime, and this expression we have for the average unlooped lifetime $< \tau_{unlooped} >$ is redundant, we confirm that it is inversely proportional to k_{on}^α and k_{on}^β. This inverse proportionality between average lifetime and the rate constants is all we need when we apply transition state theory (which predicts the rate constants' dependence on the looping J-factor) to explain the power-law-like dependence of lifetimes on J-factor.

3.2.2 The Concentration Dependence of Kinetics.

Having established a basic framework for the relationships between average lifetimes, repressor concentration, looping probabilities, and rate constants, we next turn to a comparison between our experimental results and our theoretical framework, starting with experimental observations of the dependence of the average lifetime on repressor concentration.

The effects of repressor concentration and flanking operator affinity on looped and unlooped lifetimes are shown in Fig. 3.7. Based on the kinetic scheme in Fig. 3.6 and the definition of p_{loop}, the loop breakdown rates k_{off}^α and k_{off}^β are related to the experimentally observed mean looped lifetime $\langle \tau_{looped} \rangle$ by

$$\langle \tau_{looped} \rangle = \frac{1}{\left(k_{off}^\alpha + k_{off}^\beta \right)}. \tag{3.45}$$

The loop formation rates k_{on}^α and k_{on}^β are related to the mean unlooped lifetime $\langle \tau_{unlooped} \rangle$ by

$$\langle \tau_{unlooped} \rangle = \frac{\left(1 + \frac{k_{off}^A}{[R]k_{on}^A}\right) \left(1 + \frac{k_{off}^B}{[R]k_{on}^B}\right)}{k_{on}^\alpha \frac{k_{off}^A}{[R]k_{on}^A} + k_{on}^\beta \frac{k_{off}^B}{[R]k_{on}^B}} \tag{3.46}$$

$$= \langle \tau_{looped} \rangle \left(\frac{1}{p_{loop}([R])} - 1 \right). \tag{3.47}$$

Looped state lifetimes are independent of repressor concentration.

According to Eq. 3.45, $< \tau_{looped} > = \frac{1}{(k_{off}^\alpha + k_{off}^\beta)}$. Also, both rate constants k_{off}^α and k_{off}^β are determined by DNA-protein interaction and are independent of $[R]$. As a result, the looped state lifetimes should exhibit

no dependence on repressor concentration $[R]$. As shown in Fig. 3.7b, our data are reasonably consistent with this prediction.

The value of $< \tau_{looped} >$ does depend on operator binding site identity, however: the four constructs have one common binding site (say they have the same k_{off}^{β}), so the values of $< \tau_{looped} >$ will depend on the different k_{off}^{α}. We see that from Figure 3.7b, constructs with stronger binding sites have longer $< \tau_{looped} >$ and hence smaller k_{off}^{α}. (The order of LacI-operator binding affinities is $O_{id} > O_1 > O_2$.) As we will see in section 3.2.3 in Eq. 3.57, there is a $e^{-\beta E_A}$ term in k_{off}^{α}. This could possibly qualitatively explain why construct with stronger binding site (bigger E_A, here we use positive values of binding energies) would have smaller k_{off}^{α}. However, we do not know the pre-factors for different constructs, and we do not have a definitive model to predict the operator dependence of rate constants, so we cannot conclude definitively that this $e^{-\beta E_A}$ term is the source of the variation in average lifetime between constructs.

The unlooped lifetime depends on repressor concentration via a shifted equilibrium.

On the other hand, the unlooped lifetime given by Eq. (3.46) does depend on $[R]$, and in this case the data shown in Fig. 3.7a is consistent with the predicted $[R]$ dependence. Briefly, the shape of this dependence on R is consistent with the prediction of Eq. 3.47: since $\langle \tau_{looped} \rangle$ is $[R]$-independent, the $[R]$-dependence of $\langle \tau_{unlooped} \rangle$ comes from the $[R]$-dependence in the looping probability and is captured by the ratio of unlooped and looped probabilities, or the "relative unlooped frequency" $\frac{1}{p_{loop}([R])} - 1 \equiv \omega$ in Eq. (3.47) as explained below. That is, $\langle \tau_{unlooped} \rangle$ should scale linearly with ω, as shown in Fig. 3.7c.

The unlooped lifetime depends on $[R]$ as shown in Figure 3.7a. We can explain this dependency by Eq. 3.46,

$$< \tau_{unlooped} >= \frac{(1 + \frac{k_{off}^A}{[R]k_{on}^A})(1 + \frac{k_{off}^B}{[R]k_{on}^B})}{k_{on}^{\alpha}(\frac{k_{off}^A}{[R]k_{on}^A}) + k_{on}^{\beta}(\frac{k_{off}^B}{[R]k_{on}^B})}. \tag{3.48}$$

We expect all the rate constants k's to be independent of repressor concentration $[R]$ (because they are determined by local interactions between molecules and will not be affected by other repressor molecules present in the system). However, $< \tau_{unlooped} >$ depends on $[R]$ explicitly. $[R]$ shifts the equilibrium among the internal unlooped states—if we look at the looping probability as a function of repressor concentration, given in Figure 3.8, we see some similarity between it and Figure 3.7a.

We do not know the rate constants in Eq. 3.48, but it is identical to Eq. 3.39 which says

$$< \tau_{unlooped} >=< \tau_{looped} > \frac{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B}}{\frac{[R]J_{loop}}{2K_A K_B}}. \tag{3.49}$$

Since the values of K_A, K_B and J_{loop} are constants for each construct, and also we know that $< \tau_{looped} >$ is independent of $[R]$, this tells us how $< \tau_{unlooped} >$ depends on $[R]$.

Compare Eq. 3.49 with looping probability given in Eq. 3.15, we see that the way $< \tau_{unlooped} >$ depend

on $[R]$ is entirely governed by how p_{loop} depend on $[R]$. The repressor concentration $[R]$ shifts the equilibrium and hence controls the looping probability and the unlooped lifetime. We can rewrite Eq. 3.49 in the form

$$< \tau_{unlooped} >=< \tau_{looped} > \frac{1 + \frac{[R]}{K_A} + \frac{[R]}{K_B} + \frac{[R]^2}{K_A K_B}}{\frac{[R] J_{loop}}{2 K_A K_B}} =< \tau_{looped} > (\frac{1}{p_{loop}([R])} - 1). \qquad (3.50)$$

Since $< \tau_{looped} >$ is constant over $[R]$, $< \tau_{unlooped} >$ should scale as $\frac{1}{p_{loop}([R])} - 1 = \frac{p_{unlooped}([R])}{p_{loop}([R])} \equiv \omega$. We compute this relative unlooped frequency ω, *i.e.* the odds of being in the unlooped state, from the measured looping probabilities given in Figure 3.8, with the errors on ω calculated using our standard bootstrapping procedure (see [61] and Sec. 3.4.1 above), and show in Figure 3.7c that $< \tau_{unlooped} >$ and ω indeed are correlated linearly.

3.2.3 The J-factor Dependence of Kinetics.

Looped and unlooped lifetimes scale with looping J-factor.

Figure 3.3 in 3.1 show the lifetimes plotted as a function of looping J-factors. We see that $< \tau_{looped} >$ and $< \tau_{unlooped} >$ both scale with J_{loop}, without loop *sequence* or *helical period* dependence, despite the fact that the looping boundary conditions are different for the sequences we measured. (When looping probabilities are close to 0 or when J-factor is small, the state transition events are rare. Less lifetime data is available from the same amount of total observation time and the average lifetimes will have bigger error bars.) Figure 3.3a,b show opposite trends in looped lifetime and unlooped lifetime with respect to the looping J-factor. Comparing the trends in Figure 3.3 with those in Figure 3.7a,b, we see that the interpretation of J_{loop} as effective repressor concentration fails to capture the opposite dependencies of looped and unlooped lifetimes with J_{loop}.

It is especially intriguing that not only loop formation kinetics, but also loop breakdown kinetics, are controlled by the looping J-factor, which is a function of DNA and protein bending and torsional flexibilities and the geometry of these polymers. The kinetics of loop formation are usually modeled by diffusion in a one-dimensional energy landscape governed by the polymer stiffness [108–111], from which the first passage time from unlooped to looped state is calculated. Hence, it is not surprising that the loop formation process is dependent on the J-factor, which is a function of DNA flexibility. On the other hand, the loop breakdown process is less addressed and the off rate constants of chemical reactions are usually considered to be only dependent on *local* bimolecular interaction strength, which can be tuned by ionic concentration and temperature in the environment. It is surprising that our data reveals that the *long-distance* DNA looping plays a role in the loop breakdown process. The DNA looping system gives an novel example where both on and off rate constants are modulated by the same, action-at-a-distance mechanical quantity. We will try to explain this scaling in the following arguments.

Transition state theory.

We use transition state theory to phenomenologically model the fact that both looped and unlooped lifetimes are dependent on J_{loop}, and that $< \tau_{looped} >$ and $< \tau_{unlooped} >$ have opposite trends with respect to J_{loop}. Transition state theory in chemical kinetics assumes a transition state with a higher energy between the initial and final states of the reaction. The transition state is short-lived and its detailed conformation and energy are usually unknown. (First-principle simulations can be helpful in obtaining information about the structure and free energy of the transition state, and have been attempted for protein-mediated looping. However, in the case of repressor-mediated DNA looping, the boundary conditions of the loop are not yet well-established, so simulations of the looping J-factor can yield quite diverse results, depending on the assumptions made about these boundary conditions [61]. On the other hand, in the case of ligase-mediated DNA cyclization, there is an exact solution for the free energy of the cyclized DNA because the boundary conditions are well-defined, and we expect the transition state for cyclization reactions to be easier to model than repressor-mediated looping. The same lifetime scaling with free energy should hold true for the opening and closure dynamics in cyclization, and could serve as a way to check the simulated transition state free energy.)

If the activation energy of a path is E_{act}, the rate constant would have the form $k = k_0 e^{-\beta E_{act}}$. k_0 is determined by the detailed atomic orientations and the short-ranged electrostatic interactions of the molecules. In principle, the forward and reverse paths and the transition states for them do not need to be the same, but in this case of looping and unlooping, assuming the same forward and reverse paths would be the first order approximation. In fact, as will be seen below, the assumption that the forward and reverse reactions follow the same path naturally explains both dependencies of the looped and unlooped lifetimes to the looping J-factors.

We summarize our hypothesized free energy landscape in Figure 3.2 in 3.1. From the dissociation constants K_A and K_B we can obtain E_A and E_B, the binding energy of the LacI molecule to the operators A and B. In Figure 3.2, we consider the transition between the state in which only binding site A is bound by LacI, and the looped state. We denote $\Delta F_{loop}^{\ddagger}$ as the configurational free energy change to the highly distorted DNA and possibly also bending of the Lac repressor in the transition state. (Ref. [92] suggests that the Lac repressor itself has configurational change during looping). The free energy of the unlooped, looped and the transition states are thus given by

$$F_{unlooped} = -E_A + F_0. \tag{3.51}$$

$$F_{looped} = -E_A - E_B + \Delta F + F_0. \tag{3.52}$$

$$F_{transition} = -E_A + F_0 + \Delta F^{\ddagger}_{unloop}. \tag{3.53}$$

Here, F_0 is the reference state configurational free energy of the unlooped state. Hence, the activation energy of the forward path is $(F_{transition} - F_{looped}) = \Delta F^{\ddagger}_{loop}$. The activation energy of the reverse path is $(F_{transition} - F_{unlooped}) = \Delta F^{\ddagger}_{unloop}$. As a result, the on and off rate constants for operator B are given by

$$k^{\beta}_{on} = k^{\beta}_0 e^{-\beta \Delta F^{\ddagger}_{unloop}}. \tag{3.54}$$

$$k^{\beta}_{off} = k^{\beta}_0 e^{-\beta \Delta F^{\ddagger}_{loop}} \propto e^{-\beta(\Delta F^{\ddagger}_{unloop} - \Delta F - E_B)} \propto e^{-\beta(\Delta F^{\ddagger}_{unloop} - \Delta F)}. \tag{3.55}$$

Similarly, if we consider the transition between the state that only binding site B is bound and the looped state, we can obtain the rate constants for operator A as

$$k^{\alpha}_{on} = k^{\alpha}_0 e^{-\beta \Delta F^{\ddagger}_{unloop}}. \tag{3.56}$$

$$k^{\alpha}_{off} = k^{\alpha}_0 e^{-\beta \Delta F^{\ddagger}_{loop}} \propto e^{-\beta(\Delta F^{\ddagger}_{unloop} - \Delta F - E_B)} \propto e^{-\beta(\Delta F^{\ddagger}_{unloop} - \Delta F)}. \tag{3.57}$$

The above relationships automatically satisfy the detailed balance constraints given in Eq. 3.27 and 3.28 that $\frac{k^{\alpha}_{on}}{k^{\alpha}_{off}} \propto e^{-\beta \Delta F} \propto J_{loop}$ and $\frac{k^{\beta}_{on}}{k^{\beta}_{off}} \propto J_{loop}$.

Comparing the above on-rates in Eq. 3.54 and Eq. 3.56 with the relationship between unlooped lifetime and kinetic rates given in Eq. 3.46, we see that the experimentally measured unlooped lifetimes are governed by

$$< \tau_{unlooped} > = \frac{(1 + \frac{k^A_{off}}{[R]k^A_{on}})(1 + \frac{k^B_{off}}{[R]k^B_{on}})}{k^{\alpha}_{on}(\frac{k^A_{off}}{[R]k^A_{on}}) + k^{\beta}_{on}(\frac{k^B_{off}}{[R]k^B_{on}})} = \frac{(1 + \frac{k^A_{off}}{[R]k^A_{on}})(1 + \frac{k^B_{off}}{[R]k^B_{on}})}{k^{\alpha}_0 e^{-\beta \Delta F^{\ddagger}_{unloop}}(\frac{k^A_{off}}{[R]k^A_{on}}) + k^{\beta}_0 e^{-\beta \Delta F^{\ddagger}_{unloop}}(\frac{k^B_{off}}{[R]k^B_{on}})} \propto e^{\beta \Delta F^{\ddagger}_{unloop}}. \tag{3.58}$$

On the other hand, comparing the above off-rates in Eq. 3.55 and Eq. 3.57 with the relation between looped lifetimes and kinetic rates given in Eq. 3.45, we see that the experimentally measured looped lifetimes are governed by

$$< \tau_{looped} > = \frac{1}{(k^{\alpha}_{off} + k^{\beta}_{off})} = \frac{1}{(k^{\alpha}_0 e^{-\beta(E_A + \Delta F^{\ddagger}_{unloop} - \Delta F)} + k^{\beta}_0 e^{-\beta(E_B + \Delta F^{\ddagger}_{unloop} - \Delta F)})} \propto e^{\beta(\Delta F^{\ddagger}_{unloop} - \Delta F)}. \tag{3.59}$$

The lifetimes also satisfy the constraint from the looping probability given in Eq. 3.34 that $\frac{<\tau_{looped}>}{<\tau_{unlooped}>} \propto e^{-\beta \Delta F} \propto J_{loop}$.

The transition state depends on the polymer construct.

For a polymer or a DNA-protein polymer system, the free energy of the closed or looped conformation involves an increase in elastic energy and a change in entropy due to constrained polymer fluctuations. The looped conformation and its elastic energy depend on the construct—specifically, the size, geometry and rigidities of the polymer system. The entropy change is also governed by the looped conformation and hence depends on the construct. As a result, the looping free energy and its structure are determined by the polymer system construct, even if the exact solution is not available for systems like protein-mediated DNA looping. On the other hand, despite the unknown structure of the transition state, we have some hints about what it should be. In computing the first passage time to polymer ring closure, the landscape of configurational free energy is usually calculated based on the polymer rigidities and size, and the system is modeled as diffusion in the free energy landscape (or effective potential) [108]. The time scale to looping is thus given by the rate to cross a barrier whose height is determined by the polymer construct, applying some variations of the transition state theory (such as Kramer's theory). This scheme is identical to using transition state theory in our looping system, and the barrier height bears the same concept as what would be the transition state in our system. Hence, we argue that in our system the transition state free energy is also governed by the polymer construct. Since both the transition state free energy and the looping free energy are governed by the polymer construct, their values are related.

In other words, we argue that the transition state configurational free energy change $\Delta F^{\ddagger}_{unloop}$ is related to the looping free energy change ΔF, since they are both determined by the polymer construct. For a given construct whose ΔF is known, the value of $\Delta F^{\ddagger}_{unloop}$ is uniquely determined. There is a mapping from ΔF to $\Delta F^{\ddagger}_{unloop}$,

$$\Delta F^{\ddagger}_{unloop} = \Delta F^{\ddagger}_{unloop}(\Delta F). \tag{3.60}$$

Then the unlooped and looped lifetimes given in Eq. 3.58 and 3.59 become

$$< \tau_{unlooped} > \propto e^{\beta \Delta F^{\ddagger}_{unloop}(\Delta F)}, \tag{3.61}$$

and

$$< \tau_{looped} > \propto e^{\beta(\Delta F^{\ddagger}_{unloop}(\Delta F) - \Delta F)}. \tag{3.62}$$

This means that the kinetic rates as well as the looped and unlooped lifetimes have to scale with $J_{loop} \propto e^{-\beta \Delta F}$. In particular, since $\Delta F \propto \frac{\ln J_{loop}}{-\beta}$,

$$< \tau_{unlooped} > \propto e^{\beta \Delta F^{\ddagger}_{unloop}(\Delta F)} \propto e^{\beta \Delta F^{\ddagger}_{unloop}[\frac{\ln J_{loop}}{-\beta}]} \tag{3.63}$$

and

$$< \tau_{looped} > \propto e^{\beta(\Delta F^{\ddagger}_{unloop}(\Delta F) - \Delta F)} \propto e^{\beta(\Delta F^{\ddagger}_{unloop}(\Delta F)} \cdot e^{-\beta \Delta F} \propto e^{\beta(\Delta F^{\ddagger}_{unloop}[\frac{\ln J_{loop}}{-\beta}])} \cdot J_{loop}, \tag{3.64}$$

showing that the lifetimes are functions of J_{loop} and hence the data collapse we observe in Figure 3.3.

Approximate power-law-like scaling.

Without knowing the exact form of $\Delta F_{unloop}^{\ddagger} = \Delta F_{unloop}^{\ddagger}(\Delta F)$, we can still estimate how $\Delta F_{unloop}^{\ddagger}$ and ΔF are related numerically. We ask what their relative numerical values are, quantified by the ratio n such that

$$\Delta F_{unloop}^{\ddagger}(\Delta F) = n\Delta F. \tag{3.65}$$

Within a small perturbation of parameter space (in our case, the DNA constructs share a similar range of loop lengths and have similar helical repeats; their measured looping J-factors also fall within similar ranges), we can assume n is approximately constant over this range.

If we then insert Eq. 3.65 into the lifetimes given by Eq. 3.61 and Eq. 3.62, we get

$$< \tau_{unlooped} > \propto e^{\beta \Delta F_{unloop}^{\ddagger}(\Delta F)} = e^{\beta(n\Delta F)} \propto J_{loop}^{-n}, \tag{3.66}$$

and

$$< \tau_{looped} > \propto e^{\beta(\Delta F_{unloop}^{\ddagger}(\Delta F) - \Delta F)} = e^{\beta(n-1)\Delta F} \propto J_{loop}^{-(n-1)}. \tag{3.67}$$

Thus, we expect to see approximate power-law-like scaling of the lifetimes with respect to the looping J-factor, given by

$$\log\left(< \tau_{unlooped} >\right) = (-n)\log\left(J_{loop}\right) + constant, \tag{3.68}$$

and

$$\log\left(< \tau_{looped} >\right) = (1-n)\log\left(J_{loop}\right) + constant. \tag{3.69}$$

As noted above, protein-mediated DNA looping is complicated because of as-yet poorly-established boundary conditions. We propose this approximate power-law-like scaling observation as a way to probe looping transition state free energies, and that it may be useful in future simulation efforts. The development of the transition state theory has laid the groundwork to understand the dependence of both the loop formation and breakdown kinetics on the polymer properties encapsulated within the looping J-factor.
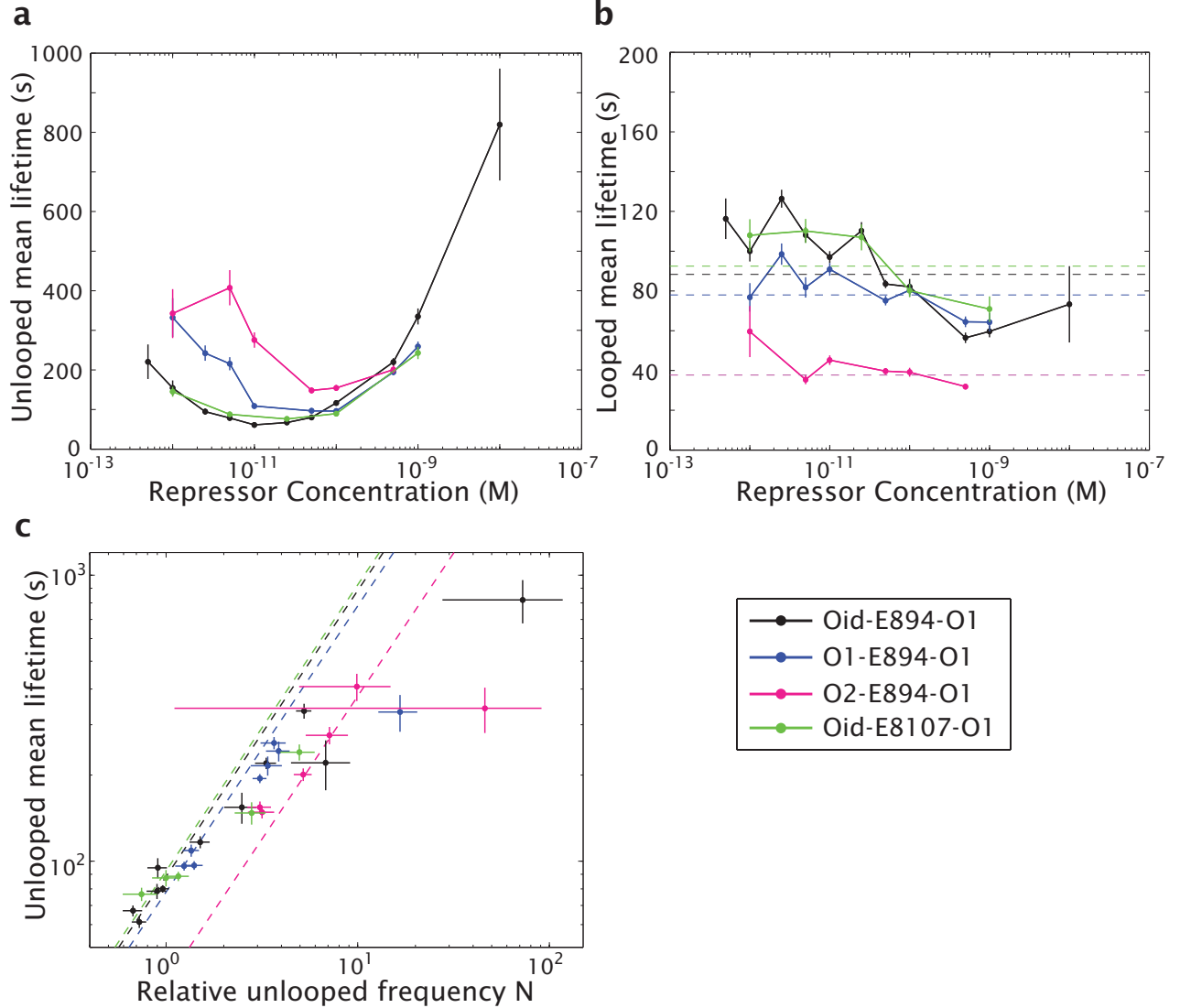
Figure 3.7: **Mean state lifetime as a function of repressor concentration and flanking operators**. All of these constructs contain the E8 sequence in the loop, but two (black and green) have the same flanking operators but different loop lengths, while the rest have the same loop length as the black data but different flanking operators. Naming convention details are given in the Methods. **a,** Mean unlooped state lifetimes versus repressor concentration $[R]$. **b,** Mean looped state lifetimes versus repressor concentration. Horizontal dashed lines represent the average looped state lifetime over all concentrations for a particular construct, weighted by the error at each concentration. As discussed in the text, the mean lifetime of the looped state should be invariant with concentration; see also Fig. 3.8 and Sec. 3.4.2 below. **c,** Mean unlooped state lifetimes versus the relative unlooped frequency ω, defined as the term that multiplies $\langle \tau_{looped} \rangle$ in Eq. (3.47). The minimum unlooped mean lifetime for each construct in panel **a** corresponds to the minimal relative unlooped frequency and maximal looping probability. Dashed lines show $\langle \tau_{unlooped} \rangle = b * \omega$, with slope b for each construct corresponding to the weighted-average looped state lifetimes shown as horizontal dashed lines in panel **b**. Given the relationship between $\langle \tau_{unlooped} \rangle$ and ω derived in Eq. (3.47), which predicts a linear relationship with a slope corresponding to the mean looped state lifetime, the dashed lines in this panel are therefore a parameter-free "fit" demonstrating that the relationship between $\langle \tau_{unlooped} \rangle$, $\langle \tau_{looped} \rangle$ and ω are in good agreement with the predictions of our model.
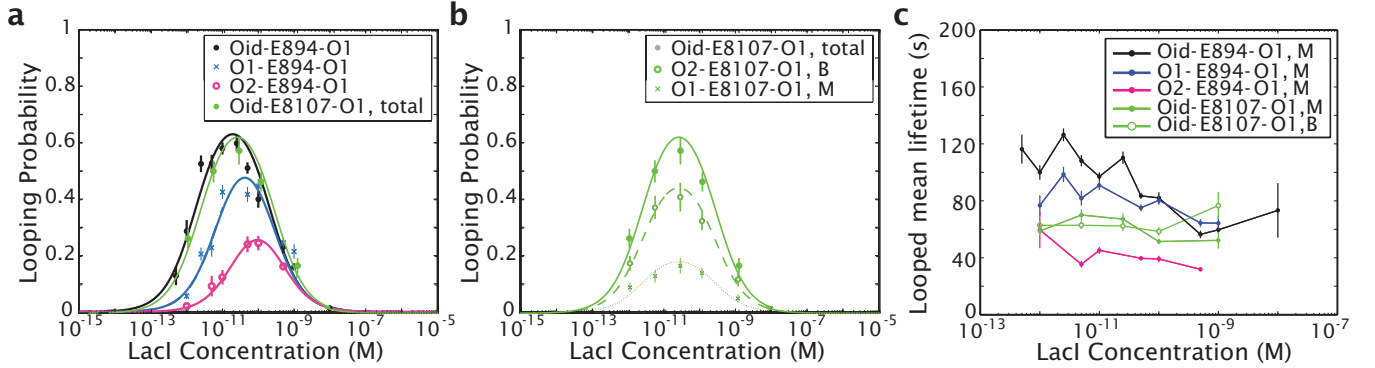
Figure 3.8: **Looping probability (a,b) and looping dwell time (c) as a function of repressor concentration.** Data in **a, b** are mean looping probabilities and are adapted from [61]; errors are standard errors on the mean. Curves are fits to our statistical mechanical model (see Eq. 3.15) with J_{loop}, K_A and K_B for each sequence as the fit parameters. Note that the two constructs represented by filled circles have the same flanking operators but different loop lengths; the other constructs have the same loop length and sequence as the black data but have different flanking operators. The difference between **a** and **b** is that **b** shows the looping probabilities for the two looped states, the "middle" (M) and "bottom" (B) states described in Sec. 3.4.1 above, for the one of these four constructs that has both states. The other three constructs shown in **a** have only the "middle" state. In **c**, the mean dwell times of the looped states whose probabilities are shown in **a, b** are plotted. Data are the same as in Fig. 3.7 above, except here we have plotted the dwell times of the two looped states of the Oid-E8107-O1 construct separately. Note that when the dwell times of the two states are plotted separately, the invariance with repressor concentration that our theoretical framework predicts is more clear. It is possible, then, that the looped state in the Oid-E894-O1 (and possibly that of O1-E894-O1 as well), which appears as a single tether length in our data, is in fact composed of different microstates, which, if plotted separately, would then appear invariant with concentration. This hypothesis is supported by the fact that the distributions of the looped state dwell times whose mean values are plotted here are not singly exponentially distributed (see Fig. 3.20 below), and by the conclusions of the vb-HMM analysis mentioned in Sec. 3.4.1 above (see Ref. [136]).

3.3 The Transition State: Comparing the Scales of DNA Deformation and Protein Binding

The information needed for life is maintained and inherited by successive generations within deoxyribonucleic acid or DNA. It is not simply a repository of information through the base pair sequence. Rather, the physical properties of DNA play a key role in regulating gene expression and ultimately many key cellular and organismal processes. One of the main ways the physics of the DNA molecule enters is through looping, which can allow different genomic regions to act in concert. Examples of this looping range from mating type switching in yeast [137], to the spread of histone marks in eukaryotes [138], to the supercoiled organization of the bacterial genome [139]. In all these cases, the elastic properties of the DNA molecule itself affect both the thermodynamic probability to form a loop and the kinetic lifetimes of looped configurations.

The dynamics of loop formation for general polymers has been well-studied, and most theoretical work focuses on understanding the looping time [108, 110, 140] or the average time to form a loop. Because of the importance of looping in the context of DNA, much work has been done to understand how the elastic properties of semiflexible polymers like DNA [141] affect the looping kinetics [109, 142]. Experimental studies of DNA looping have looked at the cyclization rate for two ends of a chain to ligate [36, 80, 83], and from this work the Jacobson-Stockmeyer factor, or J-factor, is first applied to DNA looping [80, 143]. This important concept captured the effect of the polymer chain on the looping rate, acting as the ratio of the equilibrium rate constants for looping and unlooping. Further experimental work expanded the study of DNA looping dynamics to cases where proteins, such as Lac repressor [15, 16, 120] and Fok1 [106], helped to mediate the looping. Several theoretical studies have examined the role of the protein on the kinetics [97, 144, 145]. Our work builds upon the polymer physics of DNA looping and extends this to include a simple physical model for the protein that can be used to study both the looped and unlooped lifetimes.

In a previous paper, we examined experimental observations of both the looped and unlooped lifetimes via tethered particle motion (TPM) [146]. The DNA loop is mediated by the Lac repressor protein, which stabilizes the looped conformation by binding to two locations (called operators) on the DNA. We show a dependence of the looped lifetime on the polymer energetics that is not predicted from the existing physical models of DNA looping. We explain these findings by recognizing that the loop formation process is not the same as two DNA ends being brought together, as in cyclization experiments. The Lac repressor protein that holds the loop together has a larger range of distances, with varying interaction strength, at which it can form a loop. This interaction range has important effects on the transition state of the looping reaction and consequently the kinetics of looping and unlooping.

In this work, we use our theoretical model for DNA looping to address the kinetic behavior that results from the physical properties of the DNA as well as those of the Lac repressor protein. Using this approach, we explore the range of behaviors that arise from varying DNA lengths and protein interaction distance. The next section introduces our model, which treats the looping reaction as a diffusion on an free energy

surface. Our solution allows us to calculate not only the lifetimes of the looped and unlooped states, but also the looping J-factor. In Sec. 3.3.2, we analyze the role of the interaction distance on the looping dynamics, as well as the interplay of twist and bend with varying polymer length. We also discuss how the J-factor does not follow from a local concentration of the ends of the polymer chain. Despite the fact that our model has been built to discuss the dynamics of protein-mediated DNA looping, the theoretical framework and underlying physical parameters are sufficiently general so as to allow our analysis to be applied more broadly to polymer binding dynamics in other contexts.

3.3.1   A Molecular Model

We develop a molecular-level model for DNA looping and unlooping processes associated with protein binding. Our model was first introduced to model looping kinetics in TPM experiments [146], and this manuscript provides a detailed development and extensive analysis of the range of its behaviors. Our approach is sufficiently general to incorporate the details of different DNA-binding protein systems, as well as other DNA and non-DNA polymer looping. However, the model is sufficiently simple to render predictions of looping kinetics with minimal model parameters and without resorting to major computational simulations. A schematic of our model is provided in Fig. 3.9.

The model includes the energetic and entropic penalties from bending, as well as the elastic deformation energy from the twisting rigidity of the DNA. Additionally, we capture the physical deformation and orientation of the protein through the interaction parameters in our protein binding energy. These free-energy contributions combine to give the free energy landscape over which the looping and unlooping process occur. Our analytical model predicts the looping and unlooping kinetics as a solution to the Fokker-Planck equation, which treats these processes as diffusive on the free energy landscape. By solving the dynamics of this diffusion process, we calculate the average lifetime to stay in either the looped or unlooped state and determine how these kinetic rates are connected to the free energies of the DNA and protein.

Free energy landscape of DNA looping

Double-stranded DNA behaves as a semiflexible polymer. For short lengths below the persistence length, the DNA behaves as a rigid rod, while at longer lengths the DNA tends a flexible random-walk in space. The wormlike chain model, which describes the polymer as an elastic strand subjected to thermal fluctuations [123], is a suitable model for DNA for short (tens of base pairs) to long lengths (thousands of base pairs), and thus spans the length scale for most transcription factor-mediated looping motifs. Notably, experimental measurements of short DNA strands (tens of base pairs) have shown varying agreement with the wormlike chain, particularly when the DNA is highly deformed [43, 124]. Our goal here is to provide a simple picture of the looping kinetics, and our model can be easily modified to include a non-linear elastic model that has been proposed for short DNA strands [141].

Deviations from a straight polymer conformation results a quadratic bending energy penalty. The bending
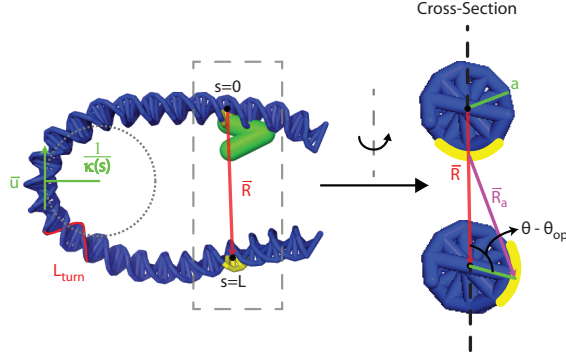
Figure 3.9: Schematic representation of our model of DNA (blue) and Lac repressor (green). The DNA chain, parametrized by s, has a tangent vector $\vec{u}$ labeled in green, and the curvature κ is the inverse of the tangent radius at that point. The site-to-site distance R is measured from the center of the DNA chain at the two binding sites, $s = 0$ and $s = L$. Because of the natural DNA helicity, the two binding sites differ in their phases on the DNA surface by an angle θ_{op}. The DNA chain has a radius a and twist angle θ, resulting in a spatial distance R_a between the two binding sites (yellow).

energy for the chain is given by

$$\beta E_{\text{bend}} = \frac{l_p}{2} \int_0^L ds \left| \frac{d\vec{u}}{ds} \right|^2 = \frac{l_p}{2} \int_0^L ds \kappa(s)^2, \tag{3.70}$$

where l_p is the persistence length, $\beta = 1/(k_B T)$, and L is the contour length of the chain between binding sites. Figure 3.9 provides definitions of the geometric quantities used in the wormlike chain model. The arclength parameter s gives the position along the chain, where $s = 0$ is the position of the first binding site, and $s = L$ is the position of the second site. The curvature $\kappa(s)$ represents the change in the tangent vector as you move along the chain, such that a larger $\kappa(s)$ indicates a more locally bent configuration at position s. The tangent vector $\vec{u}(s)$ is related to the local curvature as $\kappa(s) = \left| \frac{d\vec{u}}{ds} \right|$. Note, we use $0.34\,\text{nm}$ per bp to convert between length units [127]. In this treatment, the twist of the chain (discussed below) is decoupled from the bending deformation. Although the coupling between twist and bend can be incorporated into the theory [125], the resulting treatment would require more details of the protein geometry and the rotational dynamics than are incorporated into our current theory.

The free energy of the wormlike chain model, which includes both elastic and entropic contributions, is calculated from the Green function

$$G(\vec{R}; L) = \int \mathcal{D}[\vec{u}(s)] \, \delta \left[\vec{R} - \int_0^L ds \vec{u}(s) \right] \exp(-\beta E_{\text{bend}}). \tag{3.71}$$

This gives the probability of the two binding sites on the DNA being separated by $\vec{R}$. Due to rotational invariance, the Green function is only a function of the separation distance $R = |\vec{R}|$ [$i.e.$ $G(\vec{R}; L) = G(R; L)$]. The probability of each conformation has a Boltzmann weighting of $\exp(-\beta E_{\text{bend}})$, and the total probability is integrated over all possible chain conformations that have the same site-to-site vector $\vec{R}$. We use our

exact solution to the Green function for a wormlike chain [125, 147, 148] and use the methods described in Ref. [149] to calculate the Green function here. The conformational free energy is given by

$$\beta F_{\text{conf}}(R) = -\log\left[R^2 G(R; L)\right],\tag{3.72}$$

which gives the free energy of fixing the site-to-site distance to R.

The other contribution to the polymer energy arises from the twisting rigidity of the DNA. The formation of a looped conformation, whether for a ligation reaction between two ends of DNA or a looped structure in a DNA-protein complex, requires proper orientational alignment between the two ends. Thus, the intervening DNA of the loop may need to be twisted, with the free energy penalty

$$\beta F_{\text{twist}}(\theta) = \frac{l_t}{2L}\theta^2,\tag{3.73}$$

given to be quadratic in the local twist deformation. The angle θ gives the angle of rotation about the DNA axis away from the ground-state untwisted angle, such that $\theta = 0$ is untwisted. This model assumes the twist deformation is evenly distributed along the length of the DNA, with a twist persistence length l_t.

We note that this model for twist is the simplest model that captures the elastic orientational penalty for deforming the chain, and thus the potential value of l_t could differ substantially from the intrinsic value, which has been measured as $110\,\text{nm}$ [128]. As such, it does not include the three dimensional orientation that the two ends would need to meet, entropic contributions from the twist angle degree of freedom, nor the geometric coupling between twist and writhe of the chain. Our theory is used to model looping kinetics in TPM [146], resulting in best-fit values of the twist persistence lengths ranging from $10\,\text{nm}$ to $70\,\text{nm}$. In this work, we will focus on the behavior of our model in this range of values.

In Fig. 3.9, we show the geometric orientations for the chain in the untwisted and bound states. The undeformed orientation of the DNA is $2\pi(L/L_{\text{turn}})$, which depends upon the length of the segment to be looped due to the natural helicity of the DNA chain. The length of one complete rotation around the DNA axis $L_{\text{turn}} = 10.46\,\text{bp}$ is marked in red in Fig. 3.9. The preferred twist angle $\tilde{\theta}_{\text{op}} = 2\pi(L/L_{\text{turn}}) + \theta_0$ gives the twist angle that orients one end to face the other for binding, where θ_0 is an intrinsic angle for the orientation needed for the two ends to loop and may vary for each looping system. In the effective model that incorporates the Lac repressor, $\tilde{\theta}_{\text{op}}$ represents the angle to twist a given loop length to optimally align the unbound operator to bind the free binding domain of the Lac repressor protein. This is shown in Fig. 3.9. We note that alignment of the orientation θ to the preferred twist angle $\tilde{\theta}_{\text{op}}$ occurs at $\theta = \tilde{\theta}_{\text{op}} - 2\pi n$, where n is any integer value. In this work, we assume n takes the value n_{min} that minimizes the twist deformation, *i.e.* the bound state only includes the least twisted conformation. Thus, the twist angle goes from $\theta = 0$ to $\theta = \tilde{\theta}_{\text{op}} - 2\pi n_{\text{min}}$ as the two ends are brought closer together. We define the optimal angle $\theta_{\text{op}} = \tilde{\theta}_{\text{op}} - 2\pi n_{\text{min}}$ to define the preferred twist angle based on the minimal twist.

The separation of the two binding sites on the surface of the DNA molecule depends on the orientation.

For a fixed site-to-site-distance R measured from the middle axis of the DNA strand, the actual separation R_a between the binding sites is shown in Fig. 3.9 and is given by

$$R_a(R,\theta) = \sqrt{(R-a)^2 + a^2 - 2a(R-a)\cos(\theta - \theta_{\mathrm{op}})}. \tag{3.74}$$

In general, we assume that the DNA steric radius $a = 1\,\mathrm{nm}$ [127]. When the strand is twisted such that $\theta = \theta_{\mathrm{op}}$, the two ends are oriented as close as possible. These angles are diagrammed in Fig. 3.9.

The binding free energy drives the formation of the looped state. We choose a simple form for the energy that can capture the general trend of attraction at long-range and repulsion at short-range, regardless of details at finer scales. We describe it with a potential well with depth ϵ_0 and an interaction length scale of δ. We also include a steric cutoff at $R = 2a$ to account for the overlap of DNA backbone segments, and thus the energy goes to infinity when R is less than this value. Our model does not rely on additional geometric parameters that may not be well-characterized and may differ for each system. Thus, our conclusions are applicable to a wide range of potential DNA looping systems.

Taking the Lac repressor-mediated DNA looping as an example, ϵ_0 represents the favorable interaction energy of the Lac repressor to the operator, while δ captures the distance at which this interaction begins to be felt. Given the flexibility of the protein, the size of the Lac repressor arms, and the ability to non-specifically bind the DNA, the value of δ should reasonably be a few nanometers [96, 131, 132]. The binding energy ϵ_0 should fall close to the measured binding energy between the DNA operator and Lac repressor of 15-18 $k_B T$ [129]. These values could vary considerably for other looping systems. The binding free energy is given by

$$\beta F_{\mathrm{bind}}(r,\theta) = \begin{cases} \dfrac{-2\epsilon_0}{1+\exp\left[\frac{R_a(R,\theta)}{\delta}\right]}, & R > 2a, \\[2ex] \infty, & R \le 2a, \end{cases} \tag{3.75}$$

which depends on the distance between the two backbone sites R_a. Thus, the binding free energy is minimized both when the two ends are brought together (small site-to-site distance R) as well as oriented appropriately $(\theta = \theta_{\mathrm{op}})$.

The total free energy landscape $F_{\mathrm{total}}(R,\theta)$ shown in Fig. 3.10a is the sum of $F_{\mathrm{conf}}(R)$, $F_{\mathrm{twist}}(\theta)$, and $F_{\mathrm{bind}}(R,\theta)$. The landscape is calculated for parameters $\epsilon_0 = 23.5 k_B T$, $\delta = 1.3\,\mathrm{nm}$, $l_p = 48\,\mathrm{nm}$, $l_t = 15\,\mathrm{nm}$, and $\theta_0 = 0.003\,\pi$. These parameters correspond to fits to experimental measurements for Lac-repressor mediated looping [146]. The looped-state minimum occurs at $R = 2a$, and the chain has twisted to align the bindings sites, such that $\theta = \theta_{\mathrm{op}}$. We see this progression in the twist angle in the images of DNA (taken from a Monte Carlo simulation), where the color shows the twist angle going from zero (blue) to $\theta_{\mathrm{op}} = 2.16$ (red). The two patches of low energy near $R = 2a$ arise from either clockwise or counter-clockwise rotation of the chain to align the sites, although twisting in one direction requires less total twist.

To reduce this energy surface to a single reaction coordinate (the site-to-site distance R), we must
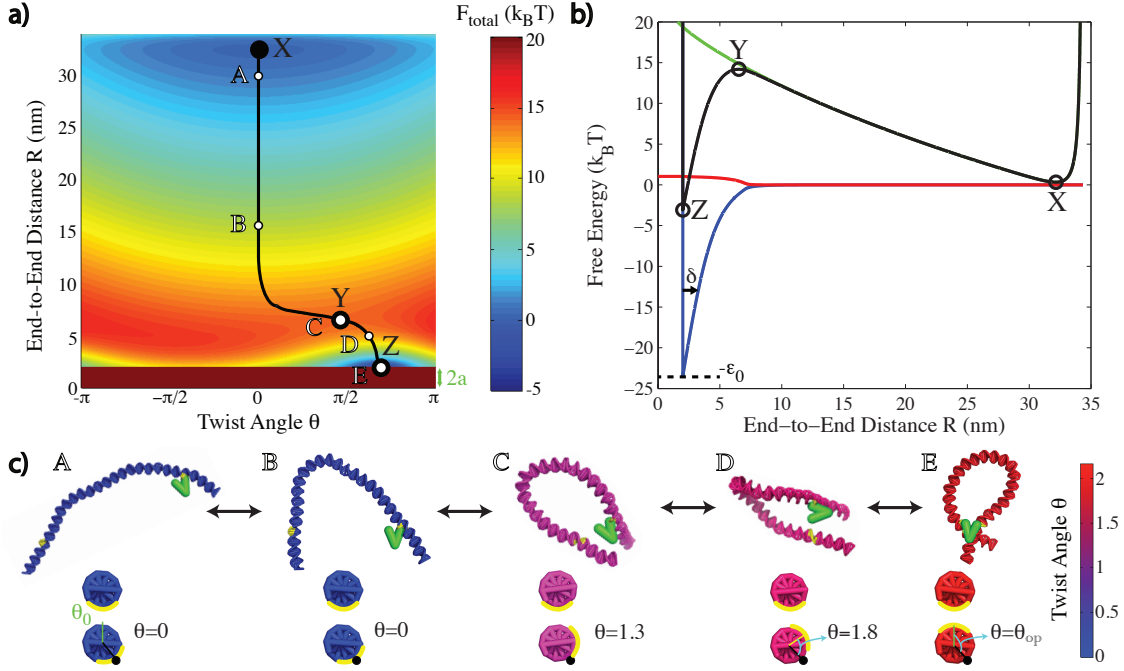
Figure 3.10: Free energy landscape for DNA looping and unlooping. **a)** Total free energy surface versus site-to-site distance R and twist angle θ. The black curve indicates the minimum free energy path between the unlooped state (X) and the looped state (Z), passing through the transition state (Y). **b)** Free energy along the minimum free energy path. The minimal-path free energy F_{min} (black) is a combination of the conformational free energy F_{conf} (green), the twisting energy F_{twist} (red), and the binding free energy F_{bind} (blue). **c)** Schematic of the looping reaction. The conformations are snapshots from a Monte Carlo simulation at different site-to-site distances. The coloration shows the variation in the twist angle, and the cross-sections views show θ throughout the binding process.

calculate the minimum free energy path from the unlooped state at point X, over the transition state Y, to the looped state Z. We calculate this minimum-path free energy $F_{\text{min}}(R)$ by determining the value of the twist angle θ that corresponds to the minimum of F_{total} for a given value of R. This procedure is numerically equivalent to finding the path along which the perpendicular gradient is a minimum, which we obtained by adapting code from Ren and Vanden-Eijnden for the zero temperature string method [150]. In Fig. 3.10b, we plot the minimal-path free energy F_{min} (black) along with the values of F_{conf} (green), F_{twist} (red), and F_{bind} (blue) along this path. The minimal-path free energy serves as the input to our determination of the looped and unlooped lifetimes.

Kinetic behavior from Fokker-Planck equation

Transition-state theory holds that the looped and unlooped lifetimes depend on the energy barriers to leave the looped and unlooped states, respectively [114]. These energy barriers are given by $\Delta F_{\text{loop}}^{\ddagger}$ and $\Delta F_{\text{unloop}}^{\ddagger}$, with the lifetimes proportional to $\exp(-\beta \Delta F_{\text{loop}}^{\ddagger})$ and $\exp(-\beta \Delta F_{\text{unloop}}^{\ddagger})$, respectively. A better approximation comes from using Kramers theory, which defines the proportionality based on the curvatures

at the minima (states X and Z) and maximum (state Y) of the free energy path [151, 152]. The complete Fokker-Planck treatment uses the full shape of the free energy landscape [153] to determine the transition rates. Within the Fokker-Planck formalism, we treat the reaction from the looped to unlooped state (and vice versa) as diffusion on a one-dimensional potential energy landscape, given by $F_{\min}(R)$ along the minimum free energy path shown in Fig. 3.10b. The major assumption of this model, which should hold well for shorter and stiffer chains, is that the polymer chain is in a local equilibrium. Our reaction coordinate is the site-to-site distance R, and each of the model parameters for the energy enters through its effect on $F_{\min}(R)$.

To start, we write the Fokker-Planck equation as

$$\left(\frac{\partial}{\partial t} - \Gamma^R\right) G^R(R', t | R, 0) = 0, \tag{3.76}$$

where

$$\Gamma^R = D\left(\frac{\partial^2}{\partial R^2} + \frac{\partial \beta F_{\min}}{\partial R}\frac{\partial}{\partial R}\right). \tag{3.77}$$

The Green function $G^R(R', t | R, 0)$ for the process is defined as the probability of going from an site-to-site distance R at time $t = 0$ to R' at time t. The governing Fokker-Planck equation is written using a backward representation (*i.e.* the Feynman-Kac formula), since we use the initial position R in this time-evolution equation. Note, we have already found the optimal twist angle θ for each site-to-site distance R, so θ does not appear explicitly in the Fokker-Planck equation. The first term in Eq. 3.77 represents the entropic driving force for moving along the free energy landscape, while the second term captures the potential energy opposing conformational changes that result in higher free energies. The diffusion coefficient D is a free parameter in our model, which we use to rescale the magnitude of the looped and unlooped lifetimes. In our previous work [146], we choose $D = 1.2 \times 10^{-5}\,\mathrm{nm^2/s}$ to align the measured looped and unlooped lifetimes to the experimental values from Lac-repressor-mediated looping. By solving Eq. 3.76 for free energy $F_{\min}(R)$, we calculate the flux of particles that cross the transition state in either direction and the average time that is spent in either the looped or the unlooped state.

We define the survival probability $S(R, t)$ for remaining in the same state (looped or unlooped) after a given time t. When starting at an site-to-site distance R within that state, the survival probability is given by

$$S(R, t) = \int_{R_{\min}}^{R_{\max}} dR' R'^2 G^R(R', t | R, 0). \tag{3.78}$$

For the looped state, R is located between $R_{\min} = 2a$ and $R_{\max} = R_Y$, and for the unlooped state, the site-to-site distance R must be between $R_{\min} = R_Y$ and $R_{\max} = L$. The survival probability follows the same dynamic equation as the Green function G^R (Eq. 3.76), which can be obtained by performing the integration in Eq. 3.78. From $S(R, t)$, we calculate the first passage time $T(R) = \int_0^\infty dt S(R, t)$ as the average time to go from an site-to-site distance R to the transition state at $R = R_Y$. Following from Eq. 3.76, we write

$$\Gamma^R T = D\left(\frac{\partial^2 T}{\partial R^2} + \frac{\partial \beta F_{\min}}{\partial R}\frac{\partial T}{\partial R}\right) = -1, \tag{3.79}$$

with boundary conditions of $T(R = R_Y) = 0$ for both the looped and unlooped states, $\frac{\partial T}{\partial R}(R = 2a) = 0$ for the looped state, and $\frac{\partial T}{\partial R}(R = L) = 0$ for the looped state. By letting $\Phi = \frac{\partial T}{\partial R}$, we can solve the resultant first-order ODE with an integrating factor. Finally, we average $T(R)$ over all possible starting site-to-site distances to obtain the mean looped and unlooped lifetimes. The mean looped lifetime is given by

$$\langle \tau_{\text{loop}} \rangle = \frac{1}{Q_{\text{loop}}} \times \qquad\qquad\qquad (3.80)$$
$$\int_{2a}^{R_Y} dR \int_{R}^{R_Y} dR' \int_{2a}^{R'} dR'' e^{-\beta F_{\min}(R) + \beta F_{\min}(R') - \beta F_{\min}(R'')}$$

where the looped state ranges from $R = 2a$ to $R = R_Y$, and the mean unlooped lifetime is

$$\langle \tau_{\text{unloop}} \rangle = \frac{1}{Q_{\text{unloop}}} \times \qquad\qquad\qquad (3.81)$$
$$\int_{R_Y}^{L} dR \int_{R_Y}^{R} dR' \int_{R'}^{L} dR'' e^{-\beta F_{\min}(R) + \beta F_{\min}(R') - \beta F_{\min}(R'')}$$

where the unlooped state ranges from $R = R_Y$ to $R = L$. We define the partition function for the looped state as $Q_{\text{loop}} = \int_{2a}^{R_Y} dR \exp\left[-\beta F_{\min}(R)\right]$ and the unlooped-state partition function as $Q_{\text{unloop}} = \int_{R_Y}^{L} dR \exp\left[-\beta F_{\min}(R)\right]$.

From our model, we calculate the J-factor for the looping reaction. This quantity is defined as the exponential of the free energy difference between the looped and unlooped states for the polymer chain only, as discussed in previous work [61, 107, 146]. The J-factor is given by

$$J_{\text{loop}} = (1\,\text{M}) \exp\left[-\beta(F_{\text{poly}}^{\text{loop}} - F_{\text{poly}}^{\text{unloop}})\right], \qquad\qquad\qquad (3.82)$$

where we assume a standard state of $1\,\text{M}$, as in previous treatments [61, 107, 146]. The free energies $F_{\text{poly}}^{\text{loop}}$ and $F_{\text{poly}}^{\text{unloop}}$ only include the energetic contributions from the polymer chain configuration and do not include the binding interaction. We write

$$F_{\text{poly}}^{\text{loop}} = -k_B T \log Q_{\text{loop}} \qquad\qquad\qquad (3.83)$$
$$- \frac{\int_{2a}^{R_Y} dR F_{\text{bind}}(R) \exp\left[-\beta F_{\min}(R)\right]}{Q_{\text{loop}}},$$

where the first term captures the total free energy of the looped state and the second term removes the binding energy contribution from $F_{\text{bind}}(R)$. We similarly can write

$$F_{\text{poly}}^{\text{unloop}} = -k_B T \log Q_{\text{unloop}} \qquad\qquad\qquad (3.84)$$
$$- \frac{\int_{R_Y}^{L} dR F_{\text{bind}}(R) \exp\left[-\beta F_{\min}(R)\right]}{Q_{\text{unloop}}}.$$

These two terms, when put into Eq. 3.82, allow us to calculate J_{loop}.

3.3.2 Results and Discussions

The model developed above connects the kinetics of polymer looping to the free energy of the polymer chain and associated binding interactions. This accurately reflects the behavior of shorter chains, where the limiting rates for loop formation are governed by the high energetic costs for bending and twisting the chain. The results of these calculations reveal the strength of the coupling between the looping dynamics of these short chains and the physical forces associated with the polymer rigidity and the binding interactions. Our results build upon our previous experimental work on Lac repressor looping [146], but our findings here are applicable to a broad range of systems involving DNA looping.

Effect of the interaction range

Forming a loop in DNA (*e.g.* DNA cyclization [36, 80, 83]) requires one end to enter into the vicinity of the other. The probability of the polymer having a conformation that allows binding is described by the J-factor, which is first described by Jacobson and Stockmayer in their work on ring formation in polycondensation reactions [143]. Since then, the J-factor has been applied to the cyclization probability of DNA [36, 80, 83] as well as protein-mediated looping probabilities [61, 107]. These experiments measure the J-factor from the ratio of the association and dissociation rates for ring or loop formation, resulting in its measure as a concentration. We note that these would be the inverse of the unlooped and looped state lifetimes, respectively. Like a Boltzmann conformation weighting, this ratio is proportional to the exponential of the free energy difference for the polymer chain between the looped and unlooped states, and essentially accounts for the probability of the polymer adopting a potential binding conformation given the associated entropic and elastic costs. Theoretical work on the J-factor utilizes polymer chain models to determine the probability distribution for the site-to-site distance with specific orientational constraints [125]. In this manner, the J-factor can be calculated as the local concentration of one end in the vicinity of the other end [154].

We describe above in Eqs. 3.83 and 3.84 the calculation of the free energy difference for the polymer chain from the free energy landscape. We define each state as the site-to-site distances shorter than (for the looped state) or longer than (for the unlooped state) the transition-state site-to-site distance R_Y, and we take an appropriate thermodynamic average of the free energy within each state, removing the non-polymeric contributions (*i.e.* the binding energy). The J-factor calculated from our model is plotted in Fig. 3.11a. For now, we just consider the black curve, whose parameters are chosen based on our fits to the experimentally determined loop lifetimes [146], with $\delta = 1.3\,\text{nm}$, $\epsilon_0 = 23.5\,k_B T$, $\theta_0 = 0.03$, and $l_p = 48\,\text{nm}$. Throughout this section, we only discuss our model without twist (*i.e.* $l_t = 0$) and defer our discussion of twist to the following sections. We see that at short lengths, the J-factor is small because of the high elastic and entropic penalty to bring the DNA into the looped shape. It reaches a peak value where $L \approx 3.4 l_p$, matching results of other calculations [125]. At longer lengths, the J-factor begins decreasing again since the entropic penalty for keeping a short site-to-site distance is more thermodynamically significant than the reduced penalty from
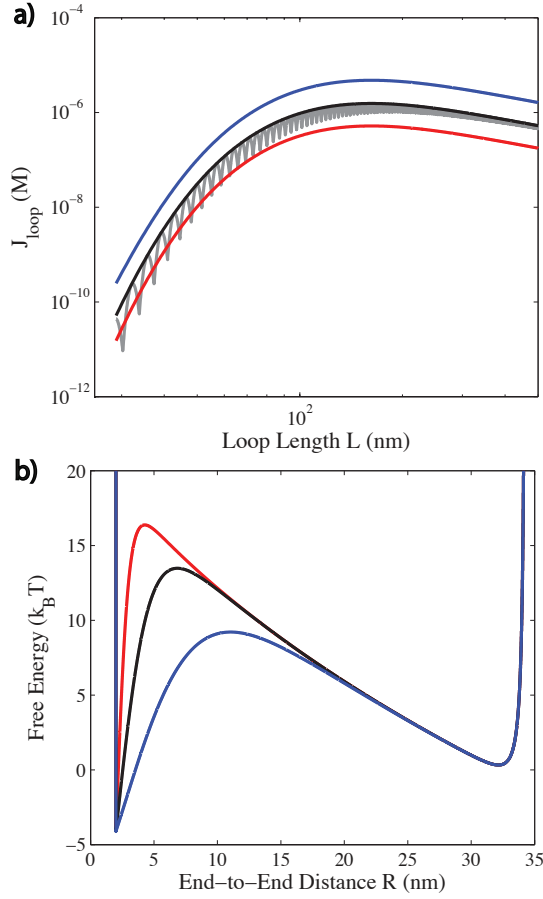
Figure 3.11: Interaction distance shifts the J-factor and the transition state. **a)** Looping J-factor plotted versus loop length, for a chain with $\epsilon_0 = 23.5\,k_BT$, $l_p = 48\,\text{nm}$, and $\theta_0 = 0.003\,\pi$. We vary δ from $0.5\,\text{nm}$ (blue) to $1.3\,\text{nm}$ (black) to $3\,\text{nm}$ (red) with the twist persistence length $l_t = 0\,\text{nm}$. In gray, we plot $\delta = 1.3\,\text{nm}$ but with $l_t = 15\,\text{nm}$. **b)** Free energy $F_{\min}$ versus site-to-site distance for a strand of length $L = 101\,\text{bp}$, with all other parameters the same as in **a**.

the bending energy with increasing chain length.

One of the major features of our model is the presence of an interaction distance δ that allows the binding energy to be felt from larger site-to-site distances. Decreasing this interaction distance shifts the transition-state site-to-site distance R_Y to the left, as is seen in Fig. 3.11b. Here, we plot the free energy along the minimal path $F_{\min}(R)$ at three different values of the interaction distance: $3\,\text{nm}$ (blue), $1.3\,\text{nm}$ (black), and $0.5\,\text{nm}$ (red). All other parameters are set to the same values as in the black J-factor curve in Fig. 3.11a. At small values of δ, the transition state occurs at a separation that is only marginally larger than the looped state conformation. With a smaller range of site-to-site distances and a larger barrier to clear the transition state, the polymer free energy $F_{\text{poly}}^{\text{loop}}$ of the looped state rises, and accordingly the J-factor decreases (as per Eq. 3.82).

In Fig. 3.11a, we also plot the J-factor for $\delta = 3\,\text{nm}$ (blue) and $\delta = 0.5\,\text{nm}$ (red), with the other parameters the same as in the black curve. The increased interaction range from $1.3\,\text{nm}$ to $3\,\text{nm}$ causes the J-factor to

shift up by a multiplicative factor of 3.08 (for lengths past the maximum J-factor). The shift from $0.5\,\mathrm{nm}$ to $1.3\,\mathrm{nm}$ is 2.99. We note that although the interaction distance is a property of the protein and its binding potential, it enters into the measured J-factors by its effect on the total free energy as described above for Fig. 3.11b. Since many looping systems could be mediated by different sized proteins with varying flexibility and other properties, the measured J-factors must account for these shifts or otherwise the predictions could be off by an order of magnitude or more.

In Fig. 3.12, we plot the looped and unlooped lifetimes versus length for the same three values of the interaction distance δ as for the J-factors in Fig. 3.11a. These curves show the looped lifetime increasing with increasing loop length, and conversely, the unlooped lifetimes decrease with loop length.

While length can be a useful measure of the polymer flexibility, it does not account for changes in other properties, such as the persistence length, which can influence the probability of the polymer forming a looped conformation. The J-factor captures all of the contributions of bending and twisting within our model, and we can plot the lifetimes versus the J-factor to understand how the polymer behavior affects the rates of looping and unlooping. The ratio of the looped and unlooped lifetimes is the J-factor (*i.e.* $\langle \tau_{\mathrm{loop}} \rangle / \langle \tau_{\mathrm{unloop}} \rangle = J_{\mathrm{loop}}$). Since the J-factor, as described above, is viewed as the local end concentration [80, 154] and appears as a reactant within the statistical weight for the looped complex in a model for Lac-repressor looping [95, 146], we would expect the J-factor to largely affect the unlooped lifetime (*i.e.* the association rate of the two ends).

As discussed in our previous work [146], experimental measurements of the looped lifetime showed a surprising dependence upon the looping J-factor. In Fig. 3.12a, we plot the looped lifetime $\langle \tau_{\mathrm{loop}} \rangle$ versus length, and in Fig. 3.12b, we show the looped lifetime versus the J-factor, over the same range of lengths ($85\,\mathrm{bp}$ to $250\,\mathrm{bp}$) as in Fig. 3.12a. The colors correspond to the same three values of the interaction distance δ as shown in Fig. 3.11. The inset to Fig. 3.12b shows a comparison between the looped lifetime from our theory (with $\delta = 1.3\,\mathrm{nm}$ and $l_t = 15\,\mathrm{nm}$) and the TPM experimental data found in Ref. [146]. We argue that the interaction length scale δ plays a key role in the looped lifetime dependence on the J-factor [146]. Larger values of the interaction distance δ shift the location of the transition state R_Y to larger site-to-site distances, as seen in Fig. 3.11b. When the transition-state distance R_Y is much larger than the looped state distance R_X, there is a significant release of elastic deformation of the chain going from the looped state to the transition state. We can see the shift in the dependence of the looped lifetime on the interaction distance in Fig. 3.12b. As we decrease δ from $3\,\mathrm{nm}$ (red) to $1.3\,\mathrm{nm}$ (black) to $0.5\,\mathrm{nm}$ (blue), the slope of the looped lifetime gradually goes away, as does the dependence of the looped lifetime on the elastic properties of the polymer.

The J-factor still plays a role in the kinetics of association (*i.e.* the unlooped lifetime). In Fig. 3.12c, we show the unlooped lifetimes versus length, and in Fig. 3.12d, we plot the unlooped lifetimes versus J_{loop} over the same length ranges as in Fig. 3.12c. The inset to Fig. 3.12d shows a comparison between the unlooped lifetime from our theory (with $\delta = 1.3\,\mathrm{nm}$ and $l_t = 15\,\mathrm{nm}$) and the TPM experimental data found
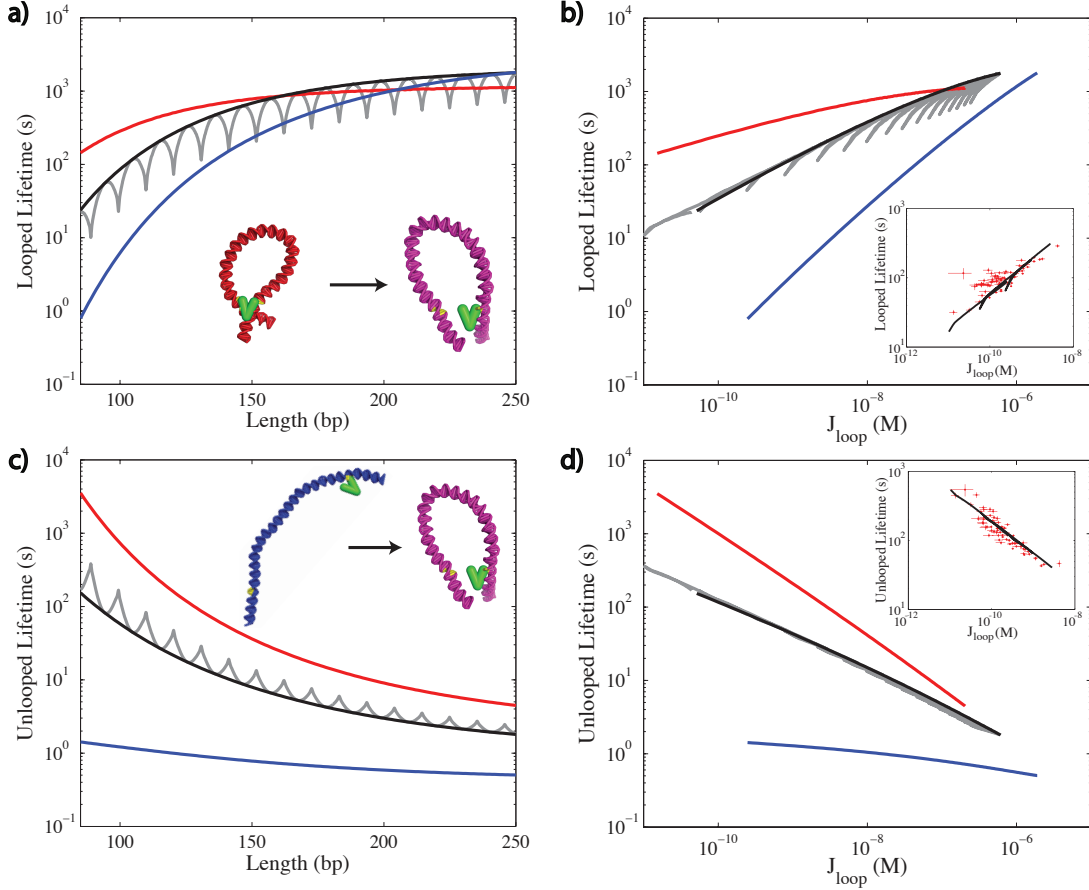
Figure 3.12: Looped and Unlooped Lifetimes. **a)** Looped lifetimes plotted versus loop length. We vary δ from 0.5 nm (blue) to 1.3 nm (black) to 3 nm (red) with the twist persistence length $l_t = 0$ nm (see text for other parameter values). In gray, we plot $\delta = 1.3$ nm but with $l_t = 15$ nm. **b)** Looped lifetimes versus J-factors for DNA chains running from lengths of 85 bp to 250 bp with the same parameters and colors as in **a**. The inset shows a comparison between our theory (with $\delta = 1.3$ nm and $l_t = 15$ nm) and TPM experiments from Ref. [146]. **c)** Unlooped lifetimes versus length, with the same parameters as in **a**. **d)** Unlooped lifetimes versus J-factors, with the same parameters as **a**. The inset shows a comparison between our theory (with $\delta = 1.3$ nm and $l_t = 15$ nm) and TPM experiments from Ref. [146].

in Ref. [146]. Since J_{loop} is part of the association reaction statistical weight, we would have assumed, treating J_{loop} as a concentration in a first-order reaction, that the reaction rate for leaving the unlooped state would be proportional to J_{loop}. Thus, the unlooped lifetime would be inversely proportional to the J-factor. Instead, we see that the unlooped lifetime does decline with increasing J-factor, but not as J_{loop}^{-1}. As the DNA chain is less deformed at the transition state R_Y, the unlooped lifetime exhibits a less significant dependence on the J-factor.

We quantify the dependence of the lifetimes on the J-factor by measuring the scaling exponents b and c for a fit of $\tau_{\mathrm{loop}} \propto J^b$ and $\tau_{\mathrm{unloop}} \propto J^c$. We perform the fit for lengths L varying from 185 bp to 225 bp to ensure we are within a length regime behaving as a power law. The scaling b of the looped lifetime ranges from 0.1032 (red, $\delta = 0.5$ nm) to 0.3072 (black, $\delta = 1.3$ nm) to 0.7147 (blue, $\delta = 3$ nm). Meanwhile, the

scaling of the unlooped lifetime with J also increases with δ, going from -0.7451 (red) to -0.5361 (black) to -0.1662 (blue). The difference between the looped and unlooped scaling exponents is approximately 1, as expected by the ratio of the looped to unlooped lifetimes being proportional to the J-factor.

Thus far, we only discuss the behavior for $l_p = 48\,\text{nm}$, corresponding to the approximate value of the DNA persistence length. It is now illustrative to explore the impact of bending rigidity on the looping kinetics, both for fundamental insight and for addressing the impact of changes in sequence in modulating the behavior. If we increase the persistence length, the increased resistance to bending results in the looped state experiencing greater elastic resistance, and the conformational free energy from Eq. 3.72 (plotted in green in Fig. 3.10b) starts increasing more sharply as the two ends come together. Consequently, the release of elastic stress from the looped state to the transition state is higher, and we therefore expect the scaling dependence to be inversely proportional to the persistence length. This results in a balance between the length scale of the elastic stiffness and the length scale of binding, affecting the transition state location and the kinetic lifetimes.
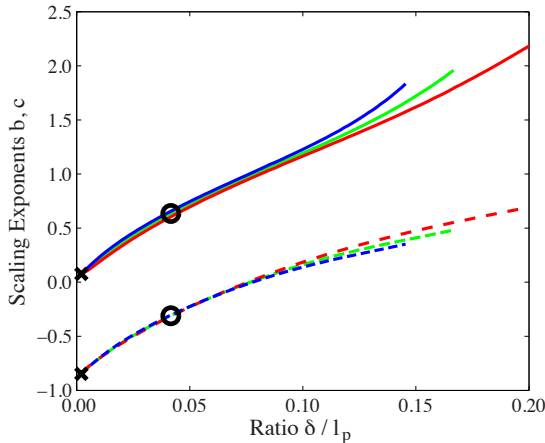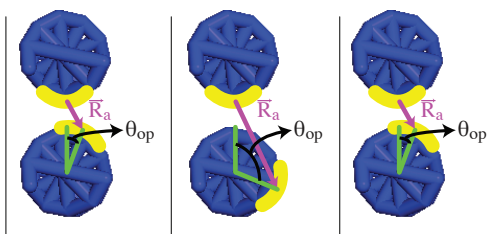
Figure 3.13: Scaling dependence versus J-factor. Scaling exponents b (solid curves) and c (dashed curves), where $\tau_{\text{loop}} \propto J^b$ and $\tau_{\text{unloop}} \propto J^c$, versus δ/l_p for three different persistence lengths: $40\,\text{nm}$ (blue), $48\,\text{nm}$ (green), and $55\,\text{nm}$ (red). Points identified by crosses and circles are described in the text.

We consider persistence lengths varying from $40\,\text{nm}$ (red) to $48\,\text{nm}$ (green) to $55\,\text{nm}$ (blue) and interaction radii varying from $0.1\,\text{nm}$ to $8\,\text{nm}$. The scaling exponents b for the looped (solid lines) and unlooped (dotted lines) lifetimes are plotted in Fig. 3.13 versus the ratio of the interaction range over the persistence length δ/l_p. We calculate the exponents from values of the DNA length ranging from $120\,\text{bp}$ to $150\,\text{bp}$ for a chain with no twist ($l_t = 0$), $\epsilon_0 = 23.5\,k_BT$, and $\theta_0 = 0.003\,\pi$. Except when δ is very large, the curves fall on the same curve, showing that the dependence on the J-factor is due to the balancing between the binding interaction length scale and the elastic deformation length scale. Deviations at large δ occur due to the diffusion time scale to leave the looped state exceeding the time scale for crossing the energetic barrier. Since the ratio of the looped to the unlooped lifetimes are proportional to J^1, the difference between the scaling laws remains approximately 1. For small values of δ/l_p, the scaling exponent for the looped lifetimes

goes to zero, and the scaling exponent for the unlooped lifetime approaches one, since the looped lifetime becomes independent of the J-factor.

L	95 bp	100 bp	105.46 bp
θ_{op}	0.114 π	0.624 π	0.114 π
F_{twist}	0.198 k_BT	5.658 k_BT	0.178 k_BT
F_{conf}	20.198 k_BT	19.535 k_BT	18.785 k_BT

Figure 3.14: Table showing the change in the relaxed orientation angle θ_{op} and bend and twist free energies at three different lengths.

The role of twist elasticity

Up to this point, we have not included the effect of twist in our discussion. The general dependence of the looping dynamics on the elastic energy within the polymer, over many helical repeat lengths, is determined by the bending energy, as is the behavior of the J-factor itself. However, the twist resistance enters into the behavior and is the dominant factor in modulating the J-factor over a short range of DNA lengths [61, 125]. Since the helical repeat length L_{turn} of DNA is only 10.46 bp, the addition of just 5 bp can shift the orientation almost 180 degrees. Figure 3.14 gives values of the relaxed orientation angle θ_{op} and the corresponding bend and twist free energies over a single helical repeat. Thus, over short lengths, the twist free energy needed to align the DNA ends for appropriate binding (Eq. 3.73 evaluated at $\theta = \theta_{op}$) can shift from a minimum of zero when the DNA does not need to twist from its natural orientation to orient the two ends properly, to a maximum when $\theta_{op} = \pi$.

The twist free energy needed to align the ends affect the J-factor, which measures the free energy change from the unlooped relaxed state to the looped state. Being in the looped state requires the polymer to not only bend into a close site-to-site distance, but also to twist to allow the ends to bind. The addition of twist free energy lowers the J-factor from the no twist state (black curve in Fig. 3.11), with greater deviation with increasing DNA twist at half-integer turns. We plot the J-factor with twist (for $l_t = 15$ nm) in gray. The oscillations in the J-factor have a periodicity of L_{turn} due to the orientation changing through an angle of 2π over that length. The deviation from the black curve is strongest at short lengths, since the twist deformation is assumed to spread equally over the full length of the chain. At long lengths, these oscillations disappear. The presence of twist deformation also affects the looped and unlooped lifetimes, as shown by the

gray curves in Fig. 3.12. We see deviations from the black curves as the twist angle needed for the preferred binding orientation oscillates away from zero. Additionally, the one-to-one dependence between J-factor and the looped lifetime no longer holds.

When there is no penalty for twisting ($l_t = 0$), the DNA chain is free to take any twist angle θ and thus assume $\theta = \theta_{\mathrm{op}}$ at all site-to-site distances R. With twist resistance, the angle changes from $\theta = 0$ to $\theta = \theta_{\mathrm{op}}$ as diagrammed in Fig. 3.10c. The twist free energy primarily affects the looped state, as the DNA chain only begins to twist as it feels the favorable binding energy associated with orienting toward the other end, as shown in the red curve in Fig. 3.10b. Because the twist free energy only starts to increase for site-to-site distances smaller than the transition state at R_Y, only the energy barrier to leave the looped state is affected. The scaling law no longer holds for the looped lifetimes because the energy barrier for the looped lifetime and the J-factor increase proportionally to twist, so the scaling over lengths less than one twist oscillation look more like J_{loop}^1.

We can see these effects clearly in Fig. 3.15a, where we have plotted the looped (red) and unlooped (blue) lifetimes for lengths ranging from 85 bp to 135 bp, with $\delta = 0.1$ nm, $l_t = 50$ nm, and all other parameters the same as previously defined. In this plot, we use the elevated value $l_t = 50$ nm in order to accentuate the oscillations arising from twist deformation. The interaction distance $\delta = 0.1$ nm is very small, and the scaling exponent b for the looped lifetime without twist should be nearly zero, as indicated by the red dashed line and the crosses in Fig. 3.13. The unlooped lifetime without twist should be nearly proportional to J_{loop}^{-1}, which is indicated by the blue dashed line. The free energy barrier of the looped state and the J-factor are both dependent upon the twist free energy, which changes sharply with length, and the bending energy changes only mildly with length over a short length range. Thus, the looped lifetime over a single decade shows a scaling law closer to J_{loop}^1 (black dotted line). Since both the twisting and bending free energies are slightly dampened as you increase length, the scaling exponent looks a bit less than one. On the other hand, since the energy barrier to leave the unlooped state is largely unaffected by the twist, the unlooped lifetime over these short lengths follow a scaling that looks more like J_{loop}^0 (black dashed line). Over many decades in length, twist energetic penalty continues oscillating to smaller and smaller values, while the elastic bending resistance steadily decreases and is the dominant effect. Thus, over long lengths, the lifetimes follow the scaling trends without twist that are given in Fig. 3.13a.

The same general behavior is shown in Fig. 3.15b, where we plot the same polymer system as in Fig. 3.15a except for an interaction distance $\delta = 2$ nm. The scaling laws for the no twist deformation case are given for the looped (red dashed line) and unlooped (blue dashed line) lifetimes, taken from the values circled in Fig. 3.13. The curves plotted here come much closer to a one-to-one correspondence between the J-factor and the kinetic lifetimes. This occurs because the looped and unlooped lifetime scalings due to the bending resistance alone (as shown in Fig. 3.13) nearly match up with the slope at short length scales that is due to the twist oscillations. Thus, while certain values of the interaction radius could make the relationship appear to be a universal trend, the overall behavior of the J-factor cannot capture in a one-to-one relationship the
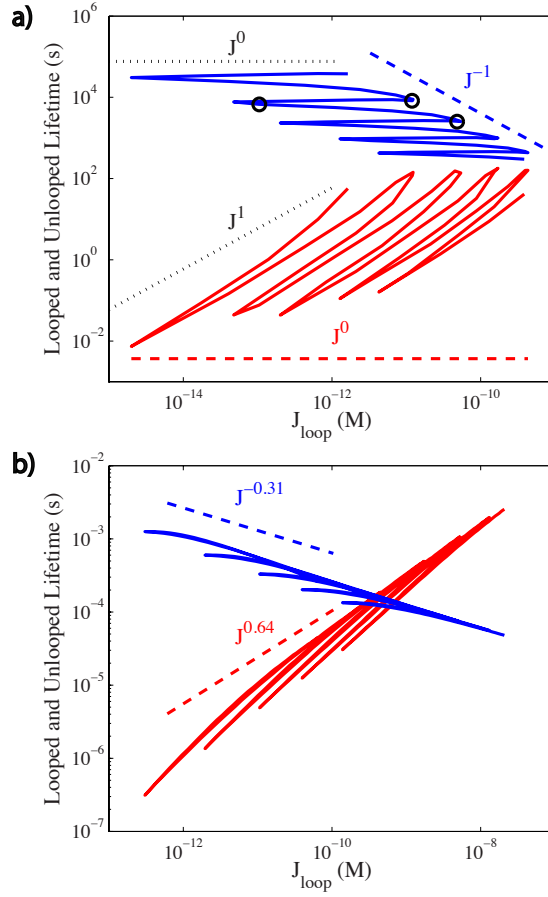
Figure 3.15: Twist influence on looped and unlooped lifetimes. **a)** Looped (red) and unlooped (blue) lifetimes for a DNA chain with persistence length of 48 nm, $\epsilon_0 = 23.5\,k_BT$, $\theta_0 = 0.003\,\pi$, $\delta = 0.1$ nm, and $l_t = 50$ nm. The loop lengths range from 85 bp to 135 bp. The three circles correspond to the DNA lengths in Fig. 3.14, and the scaling behavior for the dashed lines are from the crosses in Fig. 3.13. **b)** Looped (red) and unlooped (blue) lifetimes for DNA chain with the same parameters as in **a** except the interaction distance is $\delta = 2$ nm. The scaling behavior for the dashed lines are from the circles in Fig. 3.13.

lifetimes of the looped and unlooped states.

J-factor comparison

The effect of the protein properties on the J-factor challenges the conventional view of the J-factor as capturing the local concentration of one end of the polymer chain at the other end, which is calculated based on the site-to-site distribution of the polymer chain. We can write this local concentration as

$$
\begin{aligned}
J_{conc} &= \frac{10\,\mathrm{nm}^3\mathrm{M}}{6.02} \lim_{x \to 0} \frac{4\pi}{L^3} \frac{\int_0^x dR\, R^2 G(R)}{\frac{4}{3}\pi x^3} \\
&= \frac{10\,\mathrm{nm}^3\mathrm{M}}{6.02} \frac{G(R=0)}{L^3},
\end{aligned} \tag{3.85}
$$

where $G(R)$ is the Green function for the site-to-site distance defined in Eq. 3.71. The numerator captures the probability of the two ends being within x (nm) of each other, and the denominator is the volume encapsulated by the sphere of reaction radius x. Taking the limit gives us the local concentration at $R = 0$. The factor of $10\,\mathrm{nm^3M}/6.02$ comes from converting the units from 1 chain per $\mathrm{nm^3}$ (assuming L has units of nm) to M.

In Fig. 3.11a, we show that the J-factor is dependent on the value of the interaction radius δ, yet since the above equation has no dependence on δ, J_{loop} (from Eq. 3.82) does not have a simple correspondence to the local concentration at $R = 0$. Given that the looping reaction is not required to happen within a very small reaction window [96, 131, 132], a more appropriate view of the J-factor might be the concentration of the other end within a larger sphere. This approach is first used in Ref. [145] to show how the J-factor could be modulated by orders of magnitude given a large capture radius. In their work, they adjusted the capture radius to account for the effect of protein size on the J-factor of a wormlike chain polymer.
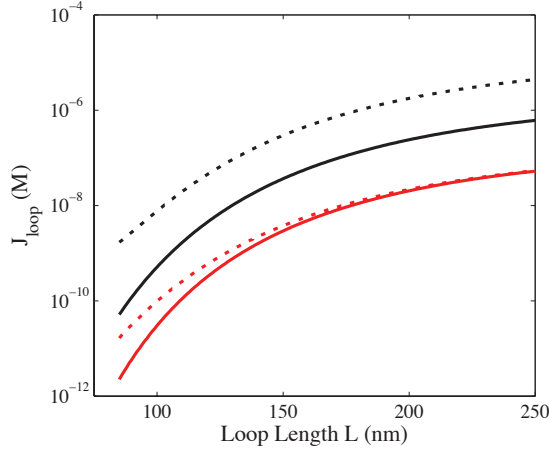
Figure 3.16: J-factor versus length. Black lines show plots of J_{loop}, as calculated in Eq. 3.82, for a DNA chain with the same properties as in Fig. 3.10 and $\delta = 1.3\,\mathrm{nm}$ (solid) or $\delta = 5\,\mathrm{nm}$ (dashed). The red curves show the J-factor without taking the limit of reaction radius x to zero, as defined in Eq. 3.85. The solid lines are for $x = 1.3\,\mathrm{nm}$ and the dashed for $x = 5\,\mathrm{nm}$.

To reproduce their results, we adjust Eq. 3.85 and not take the limit to zero, instead choosing some reaction radius x that it would need to be in to form a loop. Our calculation uses the full analytical solution to the Green function for the wormlike chain [125, 149], which was unavailable at the time that Douarche *et al.* did their work [145]. We see close quantitative agreement between our results, plotted in red in Fig. 3.16, and their calculations extending the saddle point approximation of Shimada and Yamakawa [126] in Fig. 4 of Ref. [145]. The solid line ($x = 1.3\,\mathrm{nm}$) and the dashed line ($x = 5\,\mathrm{nm}$) converge to a single curve at larger values of loop length. This convergence is qualitatively different from the behavior of J_{loop}, plotted in Fig. 3.11a. We plot J_{loop} as defined in Eq. 3.82 for $\delta = 1.3\,\mathrm{nm}$ (solid black) and $5\,\mathrm{nm}$ (dashed black). The impact of the binding energy and the protein properties on the J-factor persists at long lengths. Thus, the J-factor measured in experiments as a free energy difference between the looped and unlooped states have

qualitative distinctions from the local end concentration of the polymer chain.

Behavior at intermediate lengths

The DNA chain at these short loop lengths is naturally more rodlike in conformation, due to the high elastic penalty for binding. With increasing chain length to several persistence lengths, the DNA does not principally reside in a conformation close to full extension. Thus, the free energy minimum (at R_X) for the bending energy curve in green in Fig. 3.10b is not close to full extension as it is in the plot (full extension would be at $R = L = 34.34\,\mathrm{nm}$) for larger chain lengths. The J-factor, and thus the looping probability, peaks at chain lengths that are a few persistence lengths long. In the limit of long chain length, the wormlike chain model behaves as a Gaussian or flexible chain, with a conformations that behave as a random walk in space.
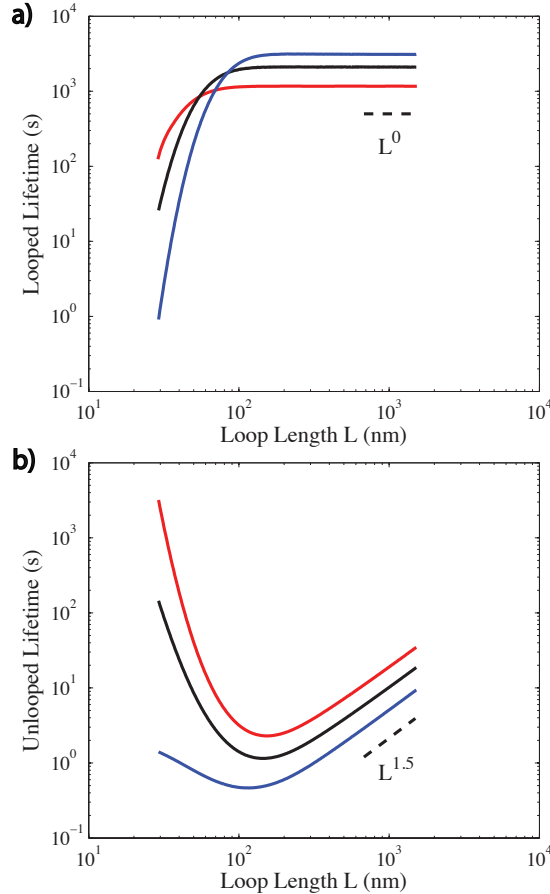
Figure 3.17: Behavior of looped and unlooped lifetimes at intermediate lengths. **a)** Looped lifetimes versus length for DNA chains running from lengths of 85 bp to 4500 bp. All curves are for a chain with $\epsilon_0 = 23.5\,k_B T$, $l_p = 48\,\mathrm{nm}$, and $\theta_0 = 0.003\,\pi$. The interaction distance varies with $\delta = 0.5\,\mathrm{nm}$ (blue), $\delta = 1.3\,\mathrm{nm}$ (black), and $\delta = 3\,\mathrm{nm}$ (red). The scaling law shows the behavior at larger values of length. **b)** Unlooped lifetimes versus length, with all the same parameters as **a**.

We plot the dependence of the looped and unlooped lifetimes for lengths spanning from 85 to 4500 bp

in Fig. 3.17, with the same parameters as in Fig. 3.11. As before, we see that the looped lifetimes increase with length for shorter length chains while the unlooped lifetimes decrease with length. However, at the DNA length at which the J-factor reaches its maximum and then begins to decrease, the unlooped lifetime begins to increase with length (with a scaling $b = 1$ from $\tau_{\text{unloop}} \propto J^b$), and the looped lifetime levels off and remains the same regardless of length. In the plots, we show that this behavior is true regardless of the interaction radius. Thus, the polymer goes to the same scaling at these longer lengths regardless of the short length-scaling which is shown to be dependent upon the value of δ.

At these intermediate length DNA chains, the elastic barrier to loop formation is no longer the dominant barrier and thus the release of elastic strain in going from the looped state to the transition state plays only a minimal role. As such, at lengths beyond the peak value of J-factor, the looped lifetime becomes independent of both J_{loop} and the loop length L. The unlooped lifetime reaches a scaling with length of $L^{1.5}$, as shown in Fig. 3.17b. This is consistent with predictions using the Gaussian chain model for the unloopend lifetime [110, 140]. At some larger length, however, the looped and unlooped lifetime behaviors exhibited in Fig. 3.17 break down, as the kinetics no longer are governed by the energetic barriers but depend upon the relaxation of the polymer chain. Other work has looked into this crossover and found that the unlooped lifetime scaling (often called the "looping time") shift from 1.5 to 2, which is characteristic of the relaxation time for a Rouse chain [110]. The length at which the Rouse relaxation time begins to dominate the looping kinetics by becoming slower than the local equilibration time has been calculated to be tens of persistence lengths (it is dependent upon the capture radius as well) [110]. Our model does not account for the effects of polymer relaxation, and so we restrict our discussion to short and intermediate chain lengths where the assumption of local equilibrium is valid.

Conclusions

One of the surprising results of the experimental measurements of the lifetimes of Lac repressor-mediated loops is the dependence of the looped lifetime on the J-factor. A view of the J-factor within the statistical mechanical framework of looping dynamics suggests that the effects of the polymer chain properties should be confined to influencing the association reaction alone, and most theoretical work on looping dynamics has focused on the unlooped lifetime dependence upon the polymer length and properties [108–112]. Dissociation rates have typically been thought of as only involving the separation of the binding surfaces at the interface between them and thus would not involve larger-scale events dependent on polymer conformation [15, 16, 106, 120]. With a simple physical model of a semiflexible polymer and a binding interaction with a finite lengthscale, we have shown how the J-factor can influence both the looped and unlooped lifetimes.

The major effect captured by our model is through the binding potential. The shape of this interaction energy influences the distribution of states that the DNA and protein adopt within the "looped state", which refers to all the site-to-site distances shorter than the transition state at R_Y. Because the strength of the binding energy varies with the site-to-site distance, the Boltzmann weight for each site-to-site distance within

the looped state varies, and through this, a dependence on the polymer configuration is introduced into the looped state free energy.

The main factor that influences the shape of the binding energy is the interaction distance δ. Physically, this parameter captures the range of the binding energy, *i.e.* how far away the two ends start to feel each other and consequently be influenced to approach and form a loop. At further distances, the two binding sites would have only a weak interaction, and the favorable binding energy decreases as the two ends approach.

Several factors contribute to why we model the binding energy this way and what should contribute to the magnitude of δ. On an atomic scale, the favorable electrostatic interactions that govern the binding increases as the two ends are brought together. For a protein-mediated loop process, such as that of Lac repressor, the protein size [96] and flexibility [131] plays an additional role. The size of a protein and the location of the binding domains affect the shape of the preferred looped conformations and the polymer site-to-site distances at which binding occurs. Internal flexibility of the protein can allow the binding sites to stretch with some energetic cost, resulting in the DNA feeling the binding interaction before it reaches the fully bound state. Finally, the binding domains on the protein may be able to non-specifically bind to the DNA and slide along the DNA into the binding site [132], effectively extending the range of lengths where the binding interactions occurs.

Our model, using δ to modulate the shape of the binding well, only captures these effects in the simplest manner possible. More detail could be added to properly account for the sterics and features of the protein within the looping reaction, as in Refs. [97, 99], as well as separate interaction energies for nonspecific binding of the DNA. While these would perhaps add to the quantitative accuracy for various systems, the qualitative features captured by simply recognizing the distance-dependent effects of the binding interaction provide an intuitive view of how the polymer physics impacts both the association and dissociation reactions. Future experimental work could seek to test the components of this interaction distance by probing other protein-mediated loop systems as well as by introducing modifications into the protein. For example, the hinge regions of Lac repressor could be lengthened (as in Ref. [16]), the salt concentrations altered, and the sequences near the binding site on DNA modified to increase nonspecific binding.

We emphasize that the interaction distance presented here plays a similar role to a reaction distance for the interaction, but due to the exponential shape of the binding energy, it is not possible to make them equivalent. The J-factor that is captured experimentally through the looping probabilities, as in Refs. [36, 61, 80, 83, 107], is a measure of the free energy difference of the polymer chain between the looped and unlooped states. While the J_{loop} term appears as if it is a local concentration within the kinetic framework, it cannot be reconciled as such because the shape of the binding energy affects the free energy of the looped state. In order to obtain the J-factors of previous works that have looked at protein size affecting the reaction radius [145, 155], we would need to assume a constant binding energy well within the looped state.

One of our major findings is that the scaling we observe for the looped and unlooped lifetimes versus

the looping J-factor is dependent upon the balance of the bending length scale (the persistence length) and the interaction length scale δ. As the length of the polymer chain increases, the impact of the bending free energy diminishes, leading to a decrease in the barrier to leave the unlooped state. The increased flexibility makes the looped state free energy lower, but because the strength of the binding energy is modulated by δ, the transition state is not lowered as much, and the looped lifetime consequently increases. Thus increasing δ pushes the slope higher, and similarly, decreasing the persistence length also affects this balancing between the looped state minimum Z and the transition state Y, resulting in the scaling dependence seen in Fig. 3.13.

The DNA mechanics are not solely a product of the bending energy, and over a short range of lengths, the twist resistance is the dominant energy that modulates the dynamic behavior. Small changes in length can have a large change on the total twist free energy by rotating the orientation out of phase by up to 180 degrees. The twist free energy, like the binding energy, primarily affects the looped state. Our model results in the twist free energy turning on as the two ends start to feel one another, which is captured through an energetic penalty to the binding energy. For example, at the transition state for $\delta = 1.3\,\mathrm{nm}$, we find that the DNA is roughly halfway to full twist, as shown in the schematic in Fig. 3.10c. If the transition state has very little twist relative to the most favorable binding orientation (i.e. $\theta_Y \ll |\theta_{\mathrm{op}}|$), going from the looped state to the transition state has a greater release in elastic energy, and the scaling versus J-factor over one repeat length should approach one. The nature of the transition state through which the looping reaction proceeds is largely unknown, but this model provides a framework through which to probe for understanding and see how the twistedness of the transition state can manifest itself through the observed scaling behavior.

The degree of transition state twist is largely controlled by the strength of the energetic binding penalty for misorientation, which is dictated by the difference between R_a and the site-to-site distance from the polymer mid-points R. The size of this penalty is controlled by the steric radius of the DNA. While we have taken the DNA radius to be $a = 1\,\mathrm{nm}$ [127], we note that it could reasonably be larger based on the ions around the DNA helix, possibly up to $1.4\,\mathrm{nm}$ [156]. The effect of a larger steric radius means a further distance that the protein would need to reach around for binding sites that were out of alignment. And thus, increased a depresses the slope by making the transition state more twisted, as discussed above. We chose our model for its simplicity in incorporating the effect of misalignment in orientation without introducing a parameter that we could not well characterized. However, more detail can be introduced to capture the nature of the twist and the orientation penalty for misalignment [99, 125, 131].

While we decouple the effects of twist from bending, we recognize that such decoupling is not possible within experiments. The persistence length of DNA is affected by DNA sequence, yet these same changes in sequence can also affect the helical repeat length, which we assumed to be constant at $10.46\,\mathrm{bp}$. The slight alterations in the locations of the peaks in the J-factor for different sequences in Ref. [61, 107] may be partially explained by changes in the helical repeat length. Despite this, the basic trend of the lifetime scaling over many decades should follow the scaling behavior without twist until the length is sufficiently long such that entropic effects become dominant over elastic contributions (i.e. past the maximum point in

the J-factor).

Further understanding of the interplay between the polymer physics and the kinetic looping events could come from Brownian dynamics simulations. Such simulations can vary the bending or twisting rigidity while using an exponential binding potential instead of a first-passage approach. Since the looping process is slow and thus computationally difficult to simulate, we intend to use simulations grounded in a proper chain discretization for a wormlike chain model [157, 158]. Such simulations could confirm some of our analytical findings as well as their limitations and potentially provide insight into a broader range of DNA looping problems that might not be as tractable analytically.

Finally, while our work has primarily been targeted to the looping of DNA, the physical insights are applicable to a much broader range of problems. Polymer looping is a general phenomenon important to a number of fields, and much work has been focused on looping within a biological context, as the creation of synthetically looped polymers has been challenging [159]. The dynamics in these looping processes is controlled by the balance between the elastic behaviors of the polymers involved and the shape of the interaction that is binding the two distal regions of the polymer together. Understanding the kinetics of DNA looping can answer some of the fundamental questions underlying many cellular processes, including viral infection [103], the transmission of epigenetic marks [138], and the conformational changes in DNA to regulate gene expression.

3.4 Supplementary Information

3.4.1 Data Analysis

Data acquisition and analysis overview.

In this work, we use the conventional half-amplitude thresholding method, described in the next section, to quantify looped and unlooped state durations ("lifetimes," also referred to in the literature as dwell times) in the tethered particle motion (TPM) data of [61, 107]. These data were obtained with Lac repressor purified in-house and used at a final concentration of 100 pM except as indicated in Fig. 3.7 below. DNA constructs consisted of loop lengths ranging from roughly 90 to 120 bp, composed of the five sequences described in Fig. 3.18: The synthetic, random E8 sequence [36, 73], the synthetic, strong nucleosome positioning sequence 601TA which we abbreviate "TA" [36, 73, 74], the strong naturally occurring 5S nucleosome positioning sequence [71], a poly(dA:dT)-rich DNA from a nucleosome-free region of a yeast promotor that we call "dA" [51], and a CG-rich sequence from humans that we call "CG" [50]. The loop sequence is flanked by various combinations of operators that are known to have different affinities for the Lac repressor: O_{id}, the strongest, O_1, roughly 4 times weaker, or O_2, about 5 times weaker than O_1. The affinities of the Lac repressor for these operators as measured by TPM are within those measured by traditional ensemble biochemical assays, as described in our previous work [61, 107]. In the main text we focus on the very strong, synthetic O_{id} operator and the weaker naturally occurring O_1 operator; data with additional operators are discussed briefly below. Roughly 150 bp of DNA separated the operators and the bead or the microscope slide, for total tether lengths of around 450 bp. Details of these constructs can be found in Fig. 3.18 and in Refs. [61, 107]; in figure legends constructs are referred to as <operator><loop sequence and length><operator>.

To summarize our previous work with these constructs, a DNA's sequence and length affect its bending and torsional rigidities and its helical repeat, which in turn affects the relative phasing of the two operators, and, it has been proposed, the conformation of the repressor protein in the looped state [38, 66, 92, 98, 99, 131]. Thus by fine-tuning DNA sequence and loop length systematically, we were able to generate a library of looping J-factors (J_{loop}) for this work. (For comparison with our theoretical framework, we focus here on simply the total looped state lifetimes, although we discuss briefly below subtleties arising from possible additional repressor conformations and/or loop structures.)

Details of our implementation of the half-amplitude thresholding procedure used to quantify looped and unlooped lifetimes are given in the next section. For each data point in the figures in this work (that is, data for a particular DNA construct at a particular repressor concentration), we obtain distributions of looped and unlooped lifetimes from roughly 30-50 tethers, observed from 10 to 100 minutes. Examples of these distributions are given in Sec. 3.4.2 below. We report the means of these distributions and the standard errors on the means. A comparison of our results to those in previous studies using TPM to measure Lac repressor looping and unlooping rates, showing good agreement between our results and these previous studies, is given in Sec. 3.4.2.

Kinetic analysis by half-amplitude thresholding.

In this work we used the conventional half-amplitude thresholding method [15, 16, 92, 106, 160–162] to obtain kinetic parameters from TPM data. An outline of this method is shown in Fig. 3.18. The basic concept of this method is to define thresholds, or root-mean-squared (RMS) values of the bead's motion, that delineate different states (for example, that define the difference between "unlooped" and "looped"), and then to define the lifetime of a particular state as the time during which a bead's RMS does not cross a threshold. The subtleties in this method arise from dealing with spurious events (for example, short-lived sticking events), and the limited time resolution of the Gaussian filter that is applied to the RMS trace before the lifetime analysis is done (see details in [61]). Our treatment of these subtleties is described below.

Trajectories were thresholded as described in [61], using the minima between Gaussians fit to the RMS histogram to define preliminary thresholds, which were manually adjusted as needed. The result of this thresholding allowed every time point in an RMS trace to be assigned a state: "U", unlooped; "M", middle looped state; "B", bottom looped state; or "Sp", spurious. Time points were labeled spurious where the RMS value exceeded the highest threshold (and was therefore most likely due to a tracking error, for example if a free bead in solution passed through the field of view), or where it fell beneath the bottom threshold (and was therefore most likely due to a sticking event). The "bottom" and "middle" looped states refer to the two different looped states, defined as two distinct tether lengths (differing in RMS by about 10 nm), that we and others have observed in looping studies with the Lac repressor [61, 63–66, 76, 92–94, 107]. However, for most of the constructs studied here the occurrence of the bottom state was too infrequent to allow reliable estimates of its mean lifetime. For this reason, and for consistency with the model derived in 3.1, the lifetimes shown in the figures in 3.1 are derived from an analysis in which the threshold between the middle and bottom looped states was ignored, such that trajectories were divided into only unlooped or looped states (and spurious states). Results for the middle state stemming from an analysis in which the bottom and middle looped states were separated out are given in Sec. 3.4.2 below.

Following the convention in the field [15, 16, 92, 160], we ignored any dwells shorter than twice the dead time of the filter (defined in the next section), treating them as follows: if a transition occurred to a state whose duration lasted shorter than twice the filter dead time, and the states just before and just after this too-short dwell were the same, we counted the flanking dwells plus the time in the too-short dwell as one long lifetime in the flanking state. If the states before and after the too-short dwell were different, however, we split the too-short dwell between the preceding and succeeding dwells. Excluding too-short dwells was performed before the removal of spurious states (so too-short spurious states, as well as too-short genuine states, were ignored).

We dealt with spurious states in a similar manner to too-short dwells: if a spurious state was preceded and followed by the same genuine state, then we assumed the underlying genuine state of the system did not change during the spurious event and considered the flanking dwells plus the time spent in the spurious state to be one long lifetime. If the flanking states were not the same, however, we counted half the spurious

event's duration towards the preceding event, and half towards the succeeding event. This approach is reasonable as long as most dwells in spurious states are significantly shorter than the average transition rate between genuine states. For example, if a sticking event occurs that lasts several minutes, and it is preceded and followed by dwells in the unlooped state, but in the rest of the trajectory the unlooped state transitions back and forth to a looped state(s) every few seconds, then it is unreasonable to assume that no transitions occurred during the minutes-long sticking event. We find the mean lifetime of spurious events in our assay to be 30 ± 3 seconds, independent of repressor concentration, loop sequence, loop length, and flanking operators. As can be seen in Fig. 3.7 and Figs. 3.21 and 3.22 below, although the mean lifetimes of genuine states for some constructs approach 30 seconds, the majority are longer. Therefore we consider our treatment of spurious events to be reasonable.[1]

The result of this thresholding procedure is a series of states and times spent in these states, for each trajectory in a data set. The mean and standard error of the lifetimes for a particular state were calculated over all trajectories in a data set. We note here one final subtlety to the calculation of these mean lifetimes, which is whether or not to include the first and last dwells in a given trajectory, whose observed duration is bounded not by transitions to new states, but on one side by the limitation of the observation time. If the dwells in each genuine state were exponentially distributed, then because of the property of memorylessness of single exponentials, we could include these first and last dwells. However, as noted below in Section 3.4.2, we find that almost none of the lifetime distributions are exponential. Therefore we excluded the first and last dwell of every trajectory from our analysis. In practice, most data sets contain so many dwells that the inclusion of these two extra dwells per trace did not change the mean lifetimes we calculated, with the exception of data sets for which the looping probability is either very high or very low, where our ability to obtain meaningful information about average lifetimes in the unlooped state (if looping is rare) or the looped state (if the looping probability is very close to 1, for example for the Oid-TA94-O1 construct at most repressor concentrations) is limited regardless of how we treat first and last dwells.

As an additional test of our algorithm for calculating lifetimes, we compared the mean lifetimes that we obtained for the Oid-E8107-O1 construct, at all five repressor concentrations, with mean lifetimes computed using a variational Bayes/hidden Markov model (vb-HMM) approach [136] similar to a method previously described for FRET [163] and single-particle tracking [164]. This vb-HMM approach is entirely orthogonal to the thresholding method described here, making use of maximum likelihood estimates of the true state at every point in a trajectory, yet results in comparable mean lifetimes to those shown in Figs. 3.7a and 3.7b below. We conclude that, while the vb-HMM approach is preferable for constructs with fast transitions

---

[1]We explored two additional approaches to dealing with spurious events that derive from an alternate assumption, that because we have no information about the true state of the system during a spurious event, we should excise the time spent in spurious states. Then the preceding and succeeding dwells can either be concatenated, as if the spurious event never occurred; or they can be counted as separate, regardless of whether they were the same or different states. Because most spurious states are relatively short-lived, the first approach, that of concatenating trajectories around excised spurious events, yields mean lifetimes that are comparable to those calculated by our chosen approach described above. However, the second approach, that of counting dwells before and after spurious events as entirely separate, leads to calculated lifetimes that are on average about half as long as those calculated from the other methods, because of the introduction of what we believe to be false transitions when no such transition actually occurred. We conclude that the method we followed gave the most reasonable estimate of the true lifetimes in genuine states.

and/or closely spaced states, or, as we argue in [136], for demonstrating that the looped states we observe are composed of multiple microstates, for our purposes here the half-amplitude approach sufficed.

Calculating the dead time of a filter.

The "dead time" of a filter refers to the duration of an event (looping or unlooping) that gives a half-amplitude response from the filter [165, 166]. The convention in the field is then to assume that the temporal resolution is twice the dead time [15, 16, 92, 160], that is, events shorter than twice the dead time cannot be resolved as true transitions between states instead of noise. In this section we will derive an expression for that dead time for the Gaussian filters that were discussed above.[2]

In this derivation we will consider the true signal from TPM to be a step function, and neglect the noise that is superimposed on this signal (though that noise also contributes to the temporal resolution of the experiment, it is ignored when calculating the filter dead time). For simplicity consider a two-state system, and let state 1 be at RMS = 0, and state 2 at RMS = A. For an event from state 1 to state 2 back to state 1, where the dwell in state 2 lasts time T and is centered at $t = 0$, we can write the corresponding raw, unfiltered TPM trace as

$$x(t) = A \cdot s_T(t), \tag{3.86}$$

where $s_T(t)$ is 1 between $t = -T/2$ and $t = +T/2$, and zero elsewhere. Then, by the definition above, the dead time of the filter will be an event duration T that produces a half-amplitude response in the filtered signal, *i.e.* such that the amplitude of the filtered signal becomes $A/2$, when the amplitude of the unfiltered signal is A.

If we apply a Gaussian filter $g(t)$ with some standard deviation σ_g to the step-function "trace", the sharp transitions from states 1 to 2 at $t = -T/2$ and from state 2 to 1 at $t = T/2$ will be smoothed, with the maximum of the filtered signal at $t = 0$, when the filter and underlying trace are aligned. Mathematically, we define "applying a filter" as convolving the filter $g(t)$ with the signal $x(t)$, such that the filtered signal filt_$x(t)$ can be written as

$$\text{filt}\_x(t) = \int_{-\infty}^{+\infty} g(t - \tau)x(\tau)d\tau. \tag{3.87}$$

Because a Gaussian is an even function, that is, symmetric about $t = 0$ such that $g(t) = g(-t)$, we know that at time $t = 0$,

$$\text{filt}\_x(0) = \int_{-\infty}^{+\infty} g(-\tau)x(\tau)d\tau \tag{3.88}$$

becomes

$$\text{filt}\_x(0) = \int_{-\infty}^{+\infty} g(\tau)x(\tau)d\tau \tag{3.89}$$

In order to find the dead time of the filter, we want to find the signal width T such that the maximum

---

[2]Thanks to Matthew Johnson at MIT for the outline of this derivation.

of filt_$x(t)$, which, as noted above, occurs in this example at $t = 0$, is equal to $A/2$. So the definition of the dead time of the filter, T_{dead}, becomes the condition that when the length of the dwell $T = T_{\mathrm{dead}}$,

$$\text{filt\_}x(0) = \int_{-\infty}^{+\infty} g(\tau)x(\tau)d\tau = \frac{A}{2}, \tag{3.90}$$

or

$$\text{filt\_}x(0) = \int_{-\infty}^{+\infty} g(\tau) \cdot A \cdot s_T(\tau)d\tau = \frac{A}{2}. \tag{3.91}$$

Note that A can be cancelled from both sides, so the dead time is independent of the signal's amplitude. That is, the dead time of the filter does not depend on the difference in RMS between states.

Since $s_\tau(\tau)$ is zero except between $-T_{\mathrm{dead}}/2$ and $T_{\mathrm{dead}}/2$, Eq. (3.91) becomes

$$\int_{-\frac{T_{\mathrm{dead}}}{2}}^{+\frac{T_{\mathrm{dead}}}{2}} g(\tau)d\tau = \frac{1}{2} \tag{3.92}$$

where we have already cancelled A from both sides.

Because $g(\tau)$ is a Gaussian, we can rewrite the integral on the left-hand side of Eq. (3.92) in terms of the cumulative distribution function of a Gaussian, usually given the variable Φ, where

$$\Phi(x) = \int_{-\infty}^{x} g(t)dt, \tag{3.93}$$

and whose solution is given by

$$\int_{-\infty}^{x} g(t)dt = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \tag{3.94}$$

Note that Φ is defined for a Gaussian whose standard deviation is 1; but we are considering a Gaussian with standard deviation σ_g. So when we write the integral in Eq. (3.92) in terms of $\Phi(x)$, we must write it as

$$\int_{-\infty}^{\frac{T_{\mathrm{dead}}}{2}} g(\tau)d\tau - \int_{-\infty}^{\frac{-T_{\mathrm{dead}}}{2}} g(\tau)d\tau = \Phi(T_{\mathrm{dead}}/(2\sigma_g)) - \Phi(-T_{\mathrm{dead}}/(2\sigma_g)), \tag{3.95}$$

expressing T_{dead} in terms of the σ_g of our filter. Given the solution to $\Phi(x)$ above, we have our final result for the condition on T_{dead},

$$\Phi\left(\frac{T_{\mathrm{dead}}}{2\sigma_g}\right) - \Phi\left(\frac{-T_{\mathrm{dead}}}{2\sigma_g}\right) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{T_{\mathrm{dead}}/(2\sigma_g)}{\sqrt{2}}\right)\right] - \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{-T_{\mathrm{dead}}/(2\sigma_g)}{\sqrt{2}}\right)\right] = \frac{1}{2}, \tag{3.96}$$

which simplifies to

$$\left[\mathrm{erf}\left(\frac{T_{\mathrm{dead}}/(2\sigma_g)}{\sqrt{2}}\right)\right] - \left[\mathrm{erf}\left(\frac{-T_{\mathrm{dead}}/(2\sigma_g)}{\sqrt{2}}\right)\right] = 1. \tag{3.97}$$

We can look up that the solution to this expression involving the error function ($\mathrm{erf}(x)$) occurs when $\frac{T_{\mathrm{dead}}}{2\sigma_g} \approx 0.67$, or that

$$T_{\mathrm{dead}} \approx 2 \cdot 0.67 \cdot \sigma_g. \tag{3.98}$$

Eq. (3.98) gives us an expression for the dead time in terms of the standard deviation of the Gaussian filter that we apply to our data. We apply the filter in Fourier space (such that the process becomes a multiplication between the filter and the Fourier-transformed data, instead of a convolution in time-space). In Fourier space the filter we use has the form

$$G(f) = e^{-0.3466(f/cf_G)^2}. \tag{3.99}$$

The factor of 0.3466 in the exponent is chosen to give 3 dB of attenuation at the cutoff frequency [165]. That is, when $f = cf_G$, the attenuation is half (3 dB corresponds to a change in power ratio of a factor of 2). For this to be the case, we must have a pre factor in the exponent of $\ln 2/2 = 0.3466$. cf_G is a rescaled cutoff frequency of the filter based on how we define our frequency axis. We choose to establish our frequency axis from $-$num. frames$/2$ to $+$num. frames$/2$, where "num. frames" is the number of image frames in a trajectory. If we want a cutoff frequency for the filter at f_{cG} Hz, then we must define

$$cf_G = \frac{\text{num. frames}}{fps \times f_{cG}^{-1}}, \tag{3.100}$$

where fps is the frame rate of the camera (30 Hz in our case). This is essentially a unit conversion, since our frequency axis is unitless but f_{cG} is in Hz. This conversion coincides with the convention of the Matlab fft command (by which we Fourier transform our data), which returns a vector the same length as the input vector, in frequencies from 0 to $fps/2$.

The Fourier transform of a Gaussian is a Gaussian, so in time space the Gaussian filter defined in Eq. (3.99) becomes

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma_g}e^{-\frac{t^2}{2\sigma_g^2}}, \tag{3.101}$$

where σ_g defines the width of the filter and is related to f_{cG} by [165]

$$\sigma_g = \frac{\sqrt{\ln 2}}{2\pi f_{cG}}. \tag{3.102}$$

We use $f_{cG} = 0.0326$ Hz, which corresponds to a Gaussian-shaped smoothing profile with a 4 second standard deviation in time space. Given this 0.0326 Hz cutoff frequency, according to Eq. (3.98) we calculate that the dead time of our filter is 5.5 seconds.

Fitting data of lifetimes versus J-factors.

Fitting was performed using the built-in lsqnonlin function in Matlab, with the standard errors on the mean lifetimes as the weights. Individual data sets (unlooped state vs. looped state, and data flanked by O_{id}-O_1 vs. O_{id}-O_2) were fit to a generic power law of the form $\langle \tau \rangle = a \times J_{loop}^b$, or all four data sets (the two state and the two pairs of flanking operators) were fit simultaneously to Eqs. 3.66 amd 3.67 in Sec. 3.2.3. That

is, for the global fit to all four data sets below, the m parameter was forced to be the same for all four data sets, but the proportionality constant was allowed to vary between data sets.

As in [61], we believe that the largest source of error in the data is the variability between tethers; therefore, as in [61], errors on the fit parameters were calculated according to a bootstrapping scheme, in which the trajectories that comprise each data set were resampled with replacement 10,000 times, and from these resampled sets, 10,000 new mean lifetimes, standard errors on the lifetimes, and J-factors were computed. The fits were then redone 10,000 times using these new mean lifetimes and J-factors, and the errors on the fit parameters taken to be the standard errors of the 10,000 new fit parameters. The fit parameters and their errors we obtain for the additional data are in reasonable agreement with that of Fig. 3.3 in 3.1.
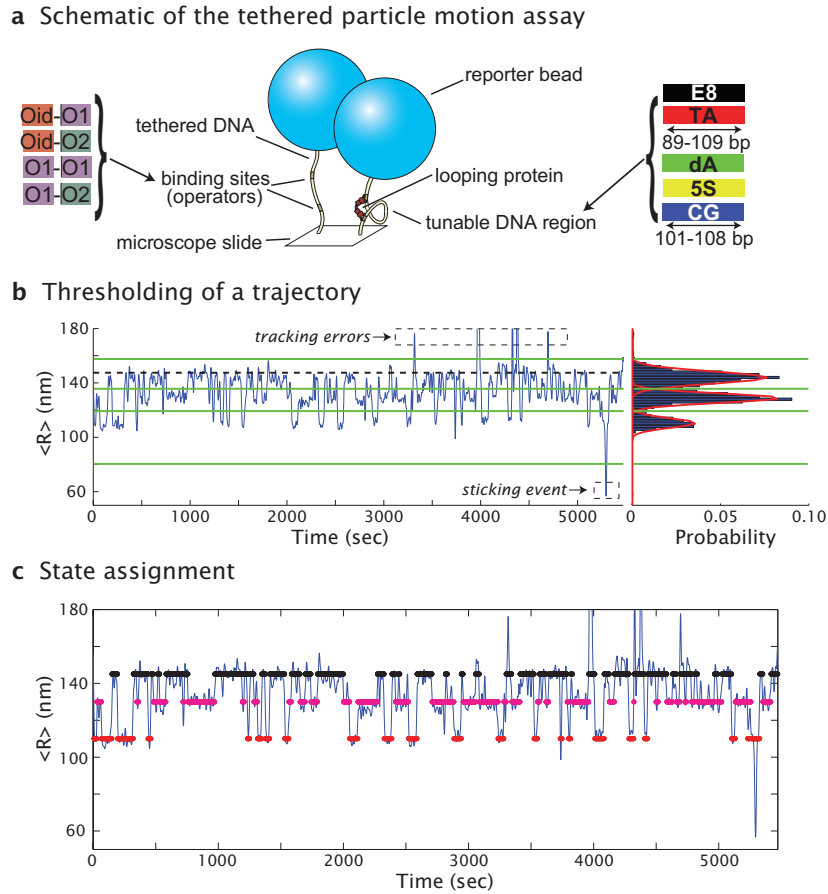
a Schematic of the tethered particle motion assay

b Thresholding of a trajectory

c State assignment

Figure 3.18: **Half-amplitude thresholding for obtaining kinetic information from TPM traces.** **a,** In tethered particle motion, single DNA molecules tether microscopic beads to the surface of a slide. The formation of a loop between two Lac repressor binding sites (operators) by the Lac repressor protein causes the motion of the bead to be reduced. By calculating the time-averaged root-mean-squared ("$\langle R \rangle$") motion of the bead over time, we can detect looping and unlooping events [15, 61, 77–79] (more precisely, $\langle R \rangle$ is our shorthand for the Gaussian-filtered root-mean-squared motion [sqrt($x_t^2 + y_t^2$)] of the bead, where x_t and y_t are a bead's (x, y) position at time t). The constructs used in this work contain varying lengths of DNA in the tunable DNA region, composed of five different sequences (the synthetic, random E8 sequence [36, 73], the synthetic, strong nucleosome positioning sequence 601TA which we abbreviate "TA" [36, 73, 74], the strong naturally occurring 5S nucleosome positioning sequence [71], a poly(dA:dT)-rich DNA from a nucleosome-free region of a yeast promotor that we call "dA" [51], and a CG-rich sequence from humans that we call "CG" [50]), and flanked by various combinations of operators that are known to have different affinities for the Lac repressor: O_{id}, the strongest, O_1, roughly 4 times weaker, or O_2, about 5 times weaker than O_1 [61, 107]. **b,** A sample TPM trajectory from a single bead, one that exhibits three genuine states (unlooped plus "middle" and "bottom" looped states) and several spurious states (both tracking errors and sticking events). The righthand panel shows a histogram of the $\langle R \rangle$ values for this trajectory. Red lines indicate a fit of three Gaussians to the histogram; thresholds (green lines) are chosen as the minima between Gaussians, except for the bottommost line, which is set at 80 nm for all traces (the shortest $\langle R \rangle$ that we can distinguish from sticking events), and the topmost line, which is set to the sum of the mean $\langle R \rangle$ in the absence of repressor (black dashed line) plus three standard deviations of the Gaussian fit to the $\langle R \rangle$ histogram in the absence of repressor. **c,** As described in the text, the thresholds defined as shown in (a) are used to assign a state (unlooped, black; middle loop, magenta; or bottom loop, red) to each point in the trajectory. Time spent in spurious states is reassigned to genuine states as described in the text. Note that this figure describes the analysis for calculating middle and bottom looped states separately; for the analyses used in 3.1, trajectories were thresholded into unlooped, looped and spurious only.

3.4.2 Additional Results

Characteristics of the lifetime distributions.

In the main text we plot the means and standard errors of the lifetimes that we obtain for various data sets. Here we ask what the full distributions of these lifetimes look like, and in particular, whether they are exponentially distributed.

A state will have exponentially distributed lifetimes if it is composed of only one microstate, such that transitions out of the state are governed by a single rate constant. As shown in Fig. 3.6, we know there are 4 microstates that contribute to what we observe as the unlooped state (no repressor bound, a repressor bound at one operator, a repressor bound to the second operator, or both repressors bound by different operators—all of which should have comparable tether lengths). We would therefore expect that the lifetimes of the unlooped state would not be exponentially distributed; and indeed a kinetic analysis by Wong and coworkers on similar constructs to those used here [92], as well as one by Revalee and coworkers with longer loops and some intrinsically curved sequences [162], found the unlooped lifetime distributions to be best fit by a mixture of two exponentials. On the other hand, we might expect the middle and/or bottom looped states that we observe to be singly exponentially distributed, if they contain only one microstate, which is in fact what Wong and coworkers found [92]. (Revalee and coworkers described only one looped state, which they found to be best fit by a biexponential distribution [162].)
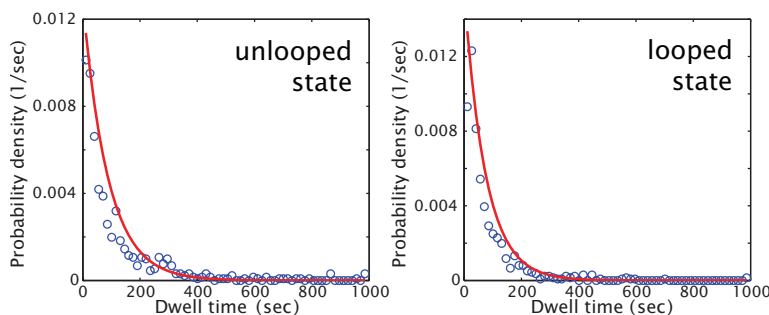
Figure 3.19: **Example lifetime histograms** for the construct Oid-E894-O1 in the presence of 100 pM repressor. Red curves are single exponential fits. This construct has only the middle looped state, whose lifetime distribution is shown in the right-hand panel. It is difficult to tell from these fits if a single exponential describes these distributions of lifetimes well, especially since we have found the results of the fits to vary significantly based on the size of the bins chosen. Here the bin size is 15 seconds. (Another way of avoiding binning is by plotting cumulative probability distributions, as in [106, 162], though we prefer the P-P plots described below for comparison to exponential distributions.)

A common method for determining whether lifetime distributions are exponentially distributed is to make histograms of the lifetimes, and fit exponentials to them (*e.g.*, [15, 92, 106]). An example of such histograms with a single-exponential fit is shown in Fig. 3.19. However, we have found the fidelity of this method to be subject to a significant amount of variability depending on the size of the bins chosen for the histogram, so instead here we use P–P ("percent–percent") plots to compare the empirical cumulative
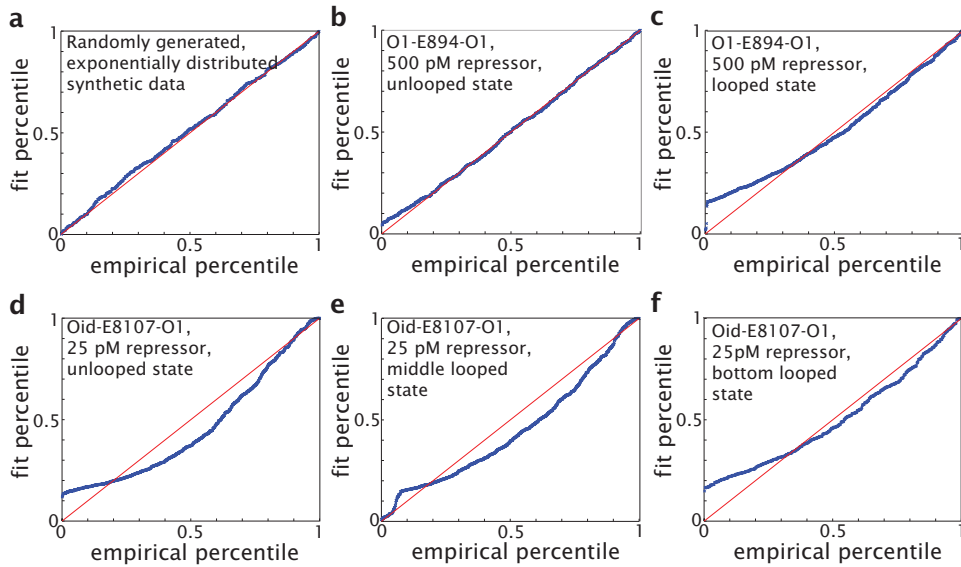
Figure 3.20: **P-P plots of the lifetimes that we obtain demonstrate that only one of our lifetime distributions is singly-exponentially distributed. a,** A P-P plot with randomly generated, exponentially distributed synthetic "data," with the same number of points as our lifetime distributions. When the empirical percentiles match the percentiles of the fit distribution, the plotted points (blue circles) will lie on the $y = x$ line (shown in red); the deviation of the points from the $y = x$ line measures the deviation of the empirical (measured) percentiles from the fit percentiles (and hence the deviation of the data from the fit model, in this case, an exponential). **b,** A P-P plot for the lifetimes obtained for the O1-E894-O1 construct's unlooped state at 500 pM repressor, compared to the cdf of a single exponential. In this case we conclude that a single exponential describes the experimental distribution well. **c,** A P-P plot for the same construct as in **b** but for the middle state. A single exponential does not describe these data well. **d-f,** P-P plots for a 107 bp loop of the E8 sequence flanked by the O_{id} and O_1 operators, at 25 pm repressor, a concentration at which both looped states are prevalent. None of these states is well-described by a single exponential, indicating that they most likely contain multiple microstates. The same holds for the other sequences, loop lengths, flanking operators and repressor concentrations lower than 500 pM that we have examined in this work. Note the common feature of blue points lying above the $y = x$ line at small percentiles, even in **b**, indicating that the measured lifetimes had too few counts at small values for the distribution to be well fit by an exponential; we suspect this feature is due at least in part to our limited temporal resolution which does not allow us to measure lifetimes shorter than 11 seconds.

distribution functions (CDF) of the measured lifetimes we obtain to the CDF of an exponential distribution fit according to maximum likelihood.[3]

As shown in Fig. 3.20, we find most of the states we observe are not exponentially distributed. The only lifetimes that we find to be exponentially distributed are those of the unlooped state at or above 500 pM repressor concentration. As discussed in [61], at high repressor concentrations (that is, above the concentration at which looping is maximal, which includes 500 pM) we expect the unlooped state to be dominated by the microstate in which both operators are bound by separate repressors; the construct shown in Fig. 3.20b has the added advantage that its two flanking operators are the same, which collapses the four possible unlooped state microstates into 3 microstates (the singly-bound-operator states being identical

---

[3]We are grateful to Matthew Johnson for the suggestion to use P–P plots and for the code to do the analysis, which can be freely downloaded from https://gist.github.com/mattjj/2356182 and https://gist.github.com/mattjj/5604903#file-qq-py.

in this case). The majority of our constructs, under the majority of repressor concentrations, do not show singly-exponentially distributed lifetime distributions for either their looped or unlooped states, indicating that all are composed of multiple microstates. We have in fact recently shown using a more sophisticated kinetic analysis of a subset of the data discussed in this work that the two looped states we observe are indeed composed of multiple microstates [136], as has been long supposed to be the case [95, 98, 99, 131], and which again is consistent with the characteristics of the lifetime distributions we obtain here.

Comparing looping rate constants for a loop flanked by identical operators to literature values.

For exponentially-distributed lifetimes, which characterize states whose exit rates are governed by a single rate constant, the rate constant for transitioning out of that state is simply the inverse of the mean lifetime of the state (see Sec. 3.2.1 above). As discussed in the previous section, we obtain exponentially distributed lifetimes only for the unlooped state at or above 500 pM. But at least for the unlooped state of those constructs that we have measured at high repressor concentration, we can compare the rate constants that we obtain to those recently reported by Wong and coworkers and Rutkauskas and coworkers on similar constructs, at the same salt concentrations that we use, and analyzed using half-amplitude threholding as we have done here [92, 94].

Wong and coworkers found their 133 bp and 138 bp loops, flanked by two O_{id} operators, to have unlooping rates (k_{off}^{α} in the language of Fig. 3.6, where in this case operators A and B are identical so $k_{off}^{\alpha} = k_{off}^{\beta}$) of 0.003 to 0.006 per second, and looping rates (k_{on}^{α}) of 0.005 to 0.03 per second, in the presence of 5.4 nM repressor. Rutkauskas and coworkers used a 285-bp loop flanked by two O_1 operators, and found unlooping rates from 0.023 to 0.046 per second. (As an aside, we note that the off-rate for O_1 in the absence of loop formation is 0.0047 per second as measured using nitrocellulose filter binding in 200 mM KCl [167], the same conditions as here and in [92, 94]).

If we assume exponentially-distributed lifetimes for our O1-E894-O1 construct at 500 pM repressor (shown in Fig. 3.20b and c above), we find a looping rate of roughly 0.005 per second, which is on the lower end of the range of values Wong and coworkers found with the stronger O_{id} operator. If we were to assume that the looped state is likewise also governed by only one rate constant (which is however unlikely, given the results of Sec. 3.4.2 above), we would calculate $k_{off}^{\alpha} \approx 0.02$ s^{-1}, which is faster than that of Wong and coworkers (which makes sense, given our weaker operator), and in good agreement with the values obtained by Rutkauskas and coworkers. Thus we are confident that our application of the half-amplitude thresholding method gives us reasonable values for approximate rate constants.

Approximate power-law-like scaling of lifetimes with J-factor as a function of flanking operators, and for the two states separately.

In the main text we show that both unlooped and looped state lifetimes have a power-law-like relationship to looping J-factors. In Fig. 3.21a and b we show that this relationship holds when one of the flanking

operators is changed (and the LacUV5 promoter sequence is added to the loop, because these sequences were originally designed for complementary *in vivo* and *in vitro* studies, with the promoter and the choice of operators being necessary for the *in vivo* work) [61, 107]. In Fig. 3.21a and b as well as in Fig. 3.3 in 3.1, the data are fit to a generic power law, as described in Sec. 3.4.1, with fit parameters given in Table 3.1. In Fig. 3.21c and d, we show the same data but fit to Eqs. 3.66 amd 3.67 in Sec. 3.2.3; these fit parameters are given in the bottom half of Table 3.1 (called "U/L, global").

Finally, in Fig. 3.22 we show the result of the lifetime analysis in which the middle and bottom looped states are thresholded separately (see Sec. 3.4.1), for both sets of flanking operators, again fit to generic power laws. Note that here we plot the middle looped state lifetimes against the looping J-factors of the *middle* state, rather than the total looping J-factor of both looped states combined. As noted above, for most constructs, especially the ones flanked by O_{id} and O_2 (and containing the promoter), the bottom state has such a low probability that $J_{loop,M}$ and $J_{loop,tot}$ are comparable. Note also that whether or not trajectories are thresholded according to unlooped-looped or unlooped-middle-bottom does not change the lifetimes of the unlooped state, as we would expect.

U/L	a (sec/M)	b (unitless)	U/M/B	a (sec/M)	b (unitless)
$\mathbf{U}_{Oid-O1}$	$28(\pm6) \times 10^{-4}$	-0.48 ± 0.03	$\mathbf{U}_{Oid-O1}$	$29(\pm7) \times 10^{-4}$	-0.48 ± 0.03
$\mathbf{U}_{Oid-O2}$	$6(\pm2) \times 10^{-4}$	-0.55 ± 0.02	$\mathbf{U}_{Oid-O2}$	$6(\pm2) \times 10^{-4}$	-0.55 ± 0.02
$\mathbf{L}_{Oid-O1}$	$2.2(\pm0.4) \times 10^5$	0.35 ± 0.02	$\mathbf{M}_{Oid-O1}$	$2.3(\pm0.3) \times 10^6$	0.5 ± 0.1
$\mathbf{L}_{Oid-O2}$	$1.9(\pm0.4) \times 10^5$	0.385 ± 0.02	$\mathbf{M}_{Oid-O2}$	$2.9(\pm0.7) \times 10^5$	0.4 ± 0.1
U/L, global	$\mathbf{C}_{U,Oid-O1}$ (sec/M)	$\mathbf{C}_{U,Oid-O2}$ (sec/M)	$\mathbf{C}_{L,Oid-O1}$ (sec/M)	$\mathbf{C}_{L,Oid-O2}$ (sec/M)	m (unitless)
	$2.9(\pm0.2) \times 10^{-4}$	$3.4(\pm0.2) \times 10^{-4}$	$1.00(\pm0.08) \times 10^6$	$4.6(\pm0.4) \times 10^5$	-0.422 ± 0.009

Table 3.1: Fit parameters for lifetimes as a function of J-factor. Two kinds of fits were performed on each of four data sets (unlooped vs. looped state, and O_{id}-O_1 vs O_{id}-O_2 as flanking operators): individual fits of each data set separately to a generic power law of the form $\langle \tau \rangle = a \times J_{loop}^b$, the parameters for which are given in the top left half of the table ("U/L"); or a global fit to all four data sets simultaneously to Eqs. 3.66 and 3.67 in Sec. 3.2.3, in which the m parameters was forced to be the same for all four data sets, but with the proportionality constants (here called C) allowed to be different for the looped vs. unlooped states and the two pairs of flanking operators. Fit parameters for this global fit are given in the bottom half of the table ("U/L, global"). In addition, we also performed fits to generic power laws for the analysis in which the middle and bottom looped states were thresholded separately, with parameters given in the top right half of the table ("U/M/B"). For most constructs the occurrence of the bottom state is too rare to allow a similar analysis on the bottom state.
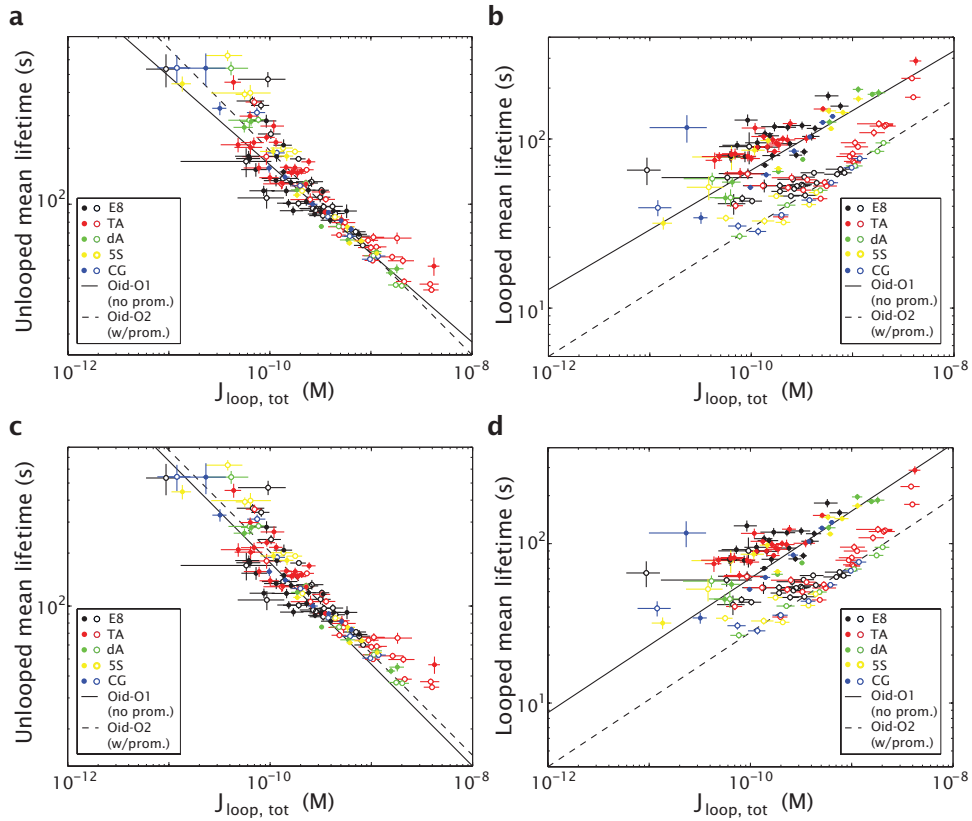
Figure 3.21: **Additional lifetimes data as a function of J-factor and flanking operators.** Unlooped state (**a, c**) and looped state (**b, d**) lifetimes as a function of J-factor and flanking operators. Closed circles are the same as those shown in 3.1, and represent constructs flanked by the O_{id} and O_1 operators; open circles are constructs in which 36 bp of the loop have been replaced by the LacUV5 promoter sequence, and the flanking operators are O_{id} and O_2, O_2 being about five times weaker than O_1. Lines in **a** and **b** represent fits to generic power laws, with fit parameters given in the top five rows of the three lefthand columns of Table 3.1; lines in **c** and **d** represent a global fit of all four data sets (looped and unlooped states with two combinations of operators) to Eqs. 3.66 amd 3.67 in Sec. 3.2.3, with fit parameters given in the bottom half of Table 3.1.
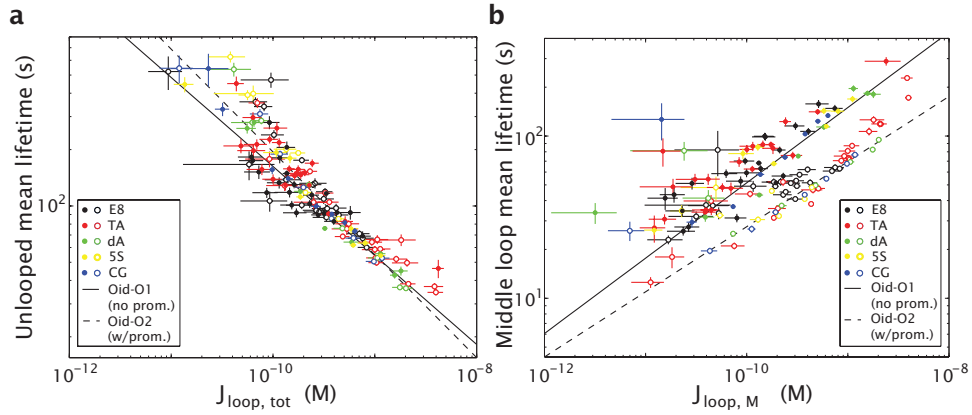
Figure 3.22: **Lifetimes vs. J-factors as a function of flanking operators, when the two looped states are considered separately.** Here the lifetimes for the middle state are plotted versus the looping J-factors for the middle state only, instead of the total looping J-factors used in the other plots here and in 3.1. Lines represent fits to generic power laws, with fit parameters given in the top five rows of the three righthand columns in Table. 3.1.
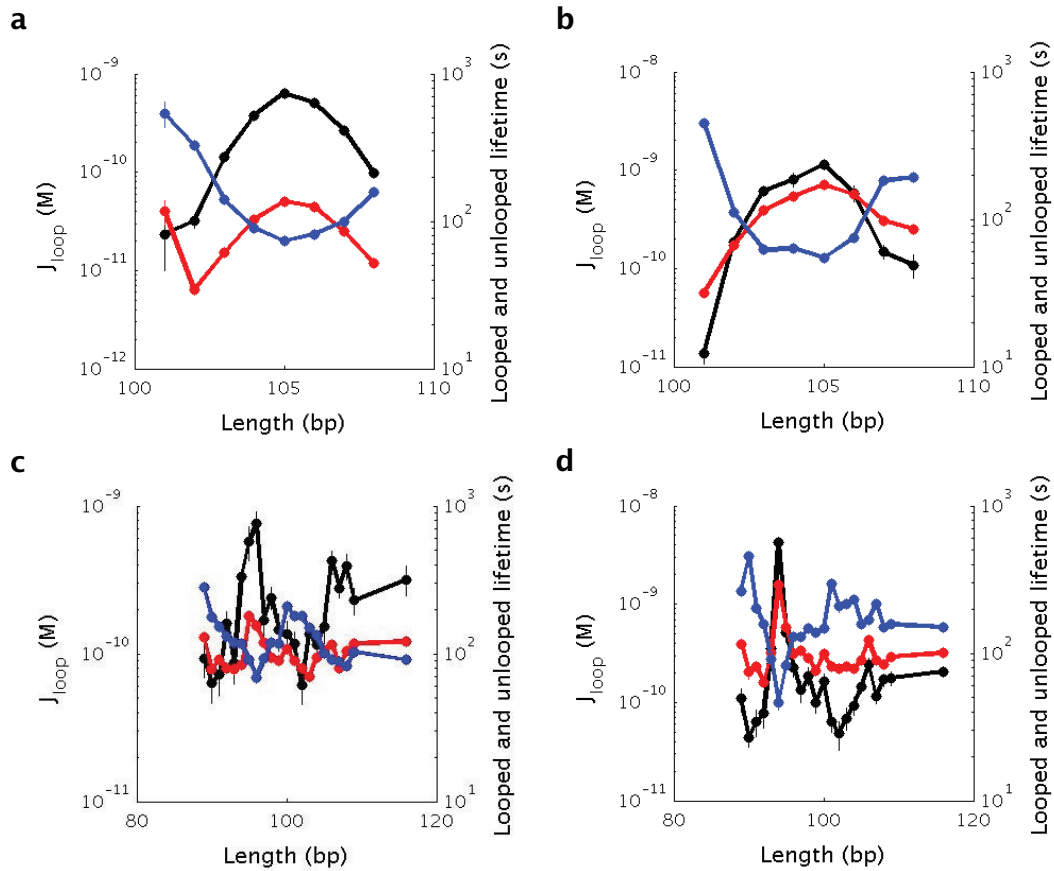
Figure 3.23: Unlooped (blue) lifetimes, looped (red) lifetimes, and looping J-factors (black) plotted as a function of loop length, for the DNA sequences **a,** CG, **b,** 5S, **c,** E8, **d,** TA.

3.5 Additional Discussions

3.5.1 The Cycling Time: Implications for Adaptation?
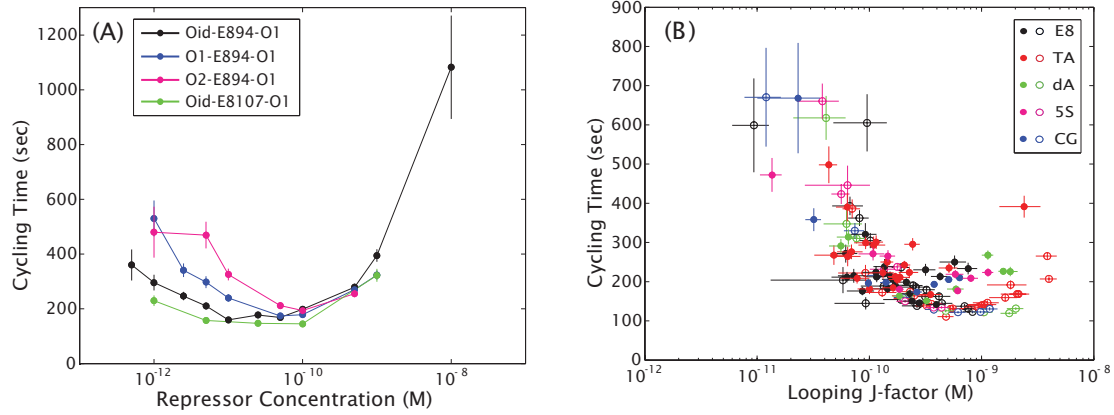
Figure 3.24: Cycling times (defined as the sum of averaged looped and unlooped dwell times) at various repressor concentrations and looping J-factors. **(A)** Cycling time as a function of repressor concentration. **(B)** Cycling time as a function looping J-factor. Open circles: with promoter constructs. Closed circles: no promoter constructs. The cycling time in our *lac*I mediated looping system changes for about 2-3 fold over two orders of magnitude of repressor concentration or looping J-factor.

We discuss the possible biological implication of the looping kinetics in this section. We define the cycling time T_C as $< \tau_{looped} > + < \tau_{unlooped} >$. This is the value representing how fast the system completes a looped-unlooped cycle and back to the looped state or vice versa. As pointed out in [100], it has a biological significance that it sets a limit for how fast the system could respond to environmental changes. [100] used λ immunity region and high λ repressor concentration resulting in non-specific binding. The cycling rate they got was nearly a constant in the one order of magnitude of repressor concentration ($10^{-8}M$ to $10^{-7}M$) they investigated, implying the system maintains fast response in this repressor concentration change. We plot the cycling time from our *lac* repressor system as follows.

Our results show that in our system, there is a range of conditions where the cycling time stays essentially constant within one order of magnitude ($10^{-11}M$ to $10^{-10}M$ repressor concentration in Figure 3.24(A) and $10^{-10}M$ to $10^{-9}M$ looping J factor range in Figure 3.24(B)). Our data covers a bigger range of conditions and reveals that the cycling time varies about 2 to 3 fold. The 2 to 3 fold change across the two orders of magnitude change in our conditions is still a pretty small variation. Our data supports the idea of an almost *conserved* cycling time, in a different looping system and without the nonspecific binding mechanism, still potentially serves to preserve the response time scale to environmental signals. It is interesting to note that the typical cycling time in our *lac*I looping system is 200 seconds at around $10^{-11}M$ to $10^{-10}M$ repressor concentration, while [100] reported typically 5 seconds cycling time for the λ repressor looping system at repressor concentration about $10^{-8}M$ to $10^{-7}M$. The *lac*I system is an endogenous sensor and switch for

nutrition in *E. coli*, while the λ repressor system is what a prophage in *E. coli* uses to make decision for lysogenic or lytic fate. Cycling times for both of the systems are short compared to the typical doubling time for *E. coli*, 1800 seconds, depending on the media, yet the viral decision making is almost 40-fold faster than the endogenous switch. The system-specific cycling time should be useful in considering biological circuit building blocks and may shed light on the optimization process during evolution.

3.5.2 A Toy Model for the Transition State

Consider a polymer of length L. The looping reaction coordinate has two degrees of freedom, r, the polymer end-to-end distance, and ϕ, the relative angle of the binding heads. From unlooped state to looped state, ϕ goes from Δ, the initial phasing given by where the two binding sites are along the DNA helix, to 0, when the mismatch is compensated. $\Delta = 2\pi \mod (\frac{L}{L_{helix}} - \Delta_0)$. The twist the polymer experiences is $(\Delta - \phi)$. Similarly from unlooped state to looped state, r goes from L, the polymer length, to 0. Because we are dealing with DNA loops shorter than one persistence length, so the end-to-end distance of the unlooped state is very close to L. Entropy plays a minor role in this case. This allows us to assume the polymer relaxation time scale is much shorter than the looping and unlooping times and the polymer is at thermal equilibrium anywhere along the free energy landscape.

Consider the binding energy has both r and ϕ dependence, so that

$$\beta E_{bind} = \frac{-2\epsilon_0}{1 + e^{\frac{r}{\delta}}} + \frac{-2\epsilon_0'}{1 + e^{\frac{\phi}{\delta'}}}. \tag{3.103}$$

At $r = 0, \phi = 0$, $\beta E_{bind} = -\epsilon_0 - \epsilon_0'$ should be about -15 to -20 kT.

The elastic energy consist of bending, twisting, and twist-bend coupling is written as

$$\beta E_{elastic} = \frac{L_p}{2L}\theta(r)^2 + \frac{L_t}{2L}(\Delta - \phi)^2 + \frac{G}{L}\theta(r)(\Delta - \phi). \tag{3.104}$$

The parameters L_p, $L_{p=t}$ and G are the bending persistence length, twist persistence length, and twist-bend coupling constant. For DNA at physiological condition they are about 50, 25 and 25 nm, respectively. $\theta(r)$ is the inverse of the polymer bending curvature and

$$r = L\frac{\sin\frac{\theta}{2}}{\frac{\theta}{2}}. \tag{3.105}$$

When r goes from L to 0, θ goes from 0 to 2π.

The total reaction energy landscape is thus

$$\beta E = \beta E_{bind} + \beta E_{elastic} = \frac{-2\epsilon_0}{1 + e^{\frac{r}{\delta}}} + \frac{-2\epsilon_0'}{1 + e^{\frac{\phi}{\delta'}}} + \frac{L_p}{2L}\theta(r)^2 + \frac{L_t}{2L}(\Delta - \phi)^2 + \frac{G}{L}\theta(r)(\Delta - \phi). \tag{3.106}$$

The energy of the unlooped state is (when there is no elastic energy)

$$\beta E_1 = \beta E(r = L, \phi = \Delta) = \beta E_{bind}(r = L, \phi = \Delta) + 0 \rightarrow 0. \tag{3.107}$$

The energy of the looped state is

$$\beta E_2 = \beta E(r = 0, \phi = 0) = -\epsilon_0 - \epsilon_0' + \frac{L_p}{2L}(2\pi)^2 + \frac{L_t}{2L}(\Delta)^2 + \frac{G}{L}(2\pi)\Delta. \tag{3.108}$$

The looping free energy is

$$\beta E_{loop} = \beta E_2 - \beta E_1 - (-\epsilon_0 - \epsilon_0') = \frac{L_p}{2L}(2\pi)^2 + \frac{L_t}{2L}(\Delta)^2 + \frac{G}{L}(2\pi)\Delta. \tag{3.109}$$

We define the transition state as the maximum of the energy landscape. To get the maximum of the landscape, we take partial derivative of βE with respect to r and ϕ,

$$\frac{\partial \beta E}{\partial r} = \frac{2\epsilon_0}{\delta}(1 + e^{r/\delta})^{-2}e^{r/\delta} + 0 + \frac{L_p}{L}\theta\frac{\partial\theta}{\partial r} + 0 + \frac{G}{L}(\Delta - \phi)\frac{\partial\theta}{\partial r}. \tag{3.110}$$

Since

$$\frac{\partial\theta}{\partial r} = \frac{\theta}{L\cos(\theta/2) - r}, \tag{3.111}$$

$$\frac{\partial \beta E}{\partial r} = \frac{2\epsilon_0}{\delta}\frac{e^{r/\delta}}{(1 + e^{r/\delta})^2} + [\frac{L_p}{L}\theta + \frac{G}{L}(\Delta - \phi)]\frac{\theta}{L\cos(\theta/2) - r}. \tag{3.112}$$

Similarly,

$$\frac{\partial \beta E}{\partial \phi} = 0 + \frac{2\epsilon_0'}{\delta'}(1 + e^{\phi/\delta'})^{-2}e^{\phi/\delta'} + 0 - \frac{L_t}{L}(\Delta - \phi) - \frac{G}{L}\theta. \tag{3.113}$$

We can use

$$\frac{\partial \beta E}{\partial r}(r^\star, \phi^\star) = 0 = \frac{2\epsilon_0}{\delta}\frac{1}{(e^{-r^\star/2\delta} + e^{r^\star/2\delta})^2} + [\frac{L_p}{L}\theta^\star + \frac{G}{L}(\Delta - \phi^\star)]\frac{\theta^\star}{L\cos(\theta^\star/2) - r^\star} \tag{3.114}$$

and

$$\frac{\partial \beta E}{\partial \phi}(r^\star, \phi^\star) = 0 = \frac{2\epsilon_0'}{\delta'}\frac{1}{(e^{-\phi^\star/2\delta'} + e^{\phi^\star/2\delta'})^2} - \frac{L_t}{L}(\Delta - \phi^\star) - \frac{G}{L}\theta^\star \tag{3.115}$$

to solve for $r^\star$ and $\phi^\star$. The notation $r^\star = L\frac{\sin\frac{\theta^\star}{2}}{\frac{\theta^\star}{2}}$. Note that if either $G = 0$ or $\epsilon_0' = 0$ (meaning no twist-bend coupling or relative angle of binding heads doesn't affect binding), Δ would not be able to control both $r^\star$ and $\phi^\star$.

The transition state energy is

$$\beta E^\star = \beta E(r = r^\star, \phi = \phi^\star) = \frac{-2\epsilon_0}{1 + e^{\frac{r^\star}{\delta}}} + \frac{-2\epsilon_0'}{1 + e^{\frac{\phi^\star}{\delta'}}} + \frac{L_p}{2L}\theta(r^\star)^2 + \frac{L_t}{2L}(\Delta - \phi^\star)^2 + \frac{G}{L}\theta(r^\star)(\Delta - \phi^\star). \tag{3.116}$$

At given $\epsilon_0, \epsilon_0', \delta, \delta', L_p, L_t, G$, and approximately small range of L, changing Δ (changing phasing) controls both $r^\star$ and $\phi^\star$. Hence $\beta E^\star$ should be determined by Δ.

The scaling exponent n would be

$$n = \frac{E^\star - E_1}{E_{loop}} \to \frac{E^\star}{E_{loop}} = \frac{\frac{-2\epsilon_0}{1+e^{\frac{r^\star}{\delta}}} + \frac{-2\epsilon_0'}{1+e^{\frac{\phi^\star}{\delta'}}} + \frac{L_p}{2L}\theta(r^\star)^2 + \frac{L_t}{2L}(\Delta - \phi^\star)^2 + \frac{G}{L}\theta(r^\star)(\Delta - \phi^\star)}{\frac{L_p}{2L}(2\pi)^2 + \frac{L_t}{2L}(\Delta)^2 + \frac{G}{L}(2\pi)\Delta}. \tag{3.117}$$

We hope to prove that for a set of given $\epsilon_0, \epsilon_0', \delta, \delta', L_p, L_t, G$ and L, n is approximately constant when changing Δ. We are also interested to know what n would be at a different polymer length L.

3.5.3 Could we obtain the waiting time distribution from solving the master equations?

Here we discuss possible future directions for solving the dwell time distribution for composite states. [168] looked at the single molecule enzymatic reaction and the turnover time. The turnover time was obtained by solving the $p(t)$ of the species from the master equation, given the initial condition is known (only pure substrate in the beginning). Following that idea, we could do the following: Assume the system is at thermal equilibrium. At any given moment we could call $t = 0$, and start to measure the waiting times from individual molecules. The waiting time distribution of the unlooped state from the measurement is $f_{unlooped}(t)$ (normalized). So on average, the probability that a transition is observed from unlooped to looped state occurs between t and $t + dt$ is $f_{unlooped}(t)dt$, but this is the exactly same as the probability increase or decrease of p_{loop}, which is dp_{loop}. On the other hand, the probability change of p_{loop} is governed by the master equation, so

$$f_{unlooped}(t)dt = dp_{loop}. \tag{3.118}$$

$$f_{unlooped}(t) = \frac{dp_{loop}}{dt} = -k_{off}^\beta p_{loop} + k_{on}^\beta p_{A-bound} - k_{off}^\alpha p_{loop} + k_{on}^\alpha p_{B-bound}. \tag{3.119}$$

In principle, we could solve the time evolution of $\frac{dp_{loop}}{dt}$ from the coupled set of master equations, since they are linear. To do so, we note that although we can not distinguish the various unlooped states, we do know that the system must be always at either $A-bound$ or $B-bound$ during the transition from looped to unlooped state or vice versa. So we could use the initial condition that $p_{A-bound} = 1$ or $p_{B-bound} = 1$ while all other p is 0 (Since we are observing single molecules and watch the time evolution of the probabilities, these p's are not the same as the statistical weights, so this doesn't violate the assumption that the system is at thermal equilibrium.) to solve the master equations, and obtain two solutions: $f_{unlooped}^A(t)$ and $f_{unlooped}^B(t)$ (normalized). Observing the system for a long enough time, we would see a mixture of both populations of dwell times. On the other hand, we know that the system is at thermal equilibrium, so the likelihood of the initial conditions $p_{A-bound} = 1$ and $p_{B-bound} = 1$ to occur is the statistical weights $w_{A-bound} = \frac{[R]}{K_A}$ and $w_{B-bound} = \frac{[R]}{K_B}$, respectively. The observed unlooped dwell time distribution should then be

$$f_{unlooped}(t) = \frac{w_{A-bound}}{w_{A-bound} + w_{B-bound}} f_{unlooped}^A(t) + \frac{w_{B-bound}}{w_{A-bound} + w_{B-bound}} f_{unlooped}^B(t). \tag{3.120}$$

Chapter 4

Bacteriophage Infection Dynamics

This project is a collaboration with David Van Valen, David Wu, Timur Zhiyentayev, Long Cai, and Willem Kegel.

A version of this paper originally appeared in: "A Single-Molecule Hershey-Chase Experiment. Van Valen D, Wu D, Chen YJ, Tuson H, Wiggins P, Phillips R, 2012 Curr Biol 22:133943", and "Effects of host-cell environment on bacteriophage DNA injection. Kegel W, Chen YJ, Wu D, Gelbart W, Phillips R, 2015 in preparation".

Viruses are ubiquitous in all the domains of life. One of the signature events of viral infections is the arrival of the viral genetic material within the host. As shown by Hershey and Chase more than half a century ago, many bacterial viruses carry out their infection process through a process of DNA translocation across the host cell membrane. However, since Hershey and Chase used bacterial viruses to definitively establish DNA as the molecular carrier of genetic information in 1952, the precise mechanisms of phage DNA translocation still remains a mystery. While bulk measurements have set a time scale for *in vivo* DNA translocation during bacteriophage infection, measurements of DNA ejection by single bacteriophages have only been made *in vitro*.

The advent of single-molecule biophysics resulted in a resurgence of interest in the life cycles of bacterial viruses, earlier central players in the development of molecular biology [169–171] and now providing a powerful testbed for quantitative analysis in biology [18, 172–176]. Much information has been gathered recently about the structure of both individual viral proteins and assembled viruses [20, 177]. These structural efforts are complemented by a variety of experiments that have elucidated the biophysics of binding of viruses to their host receptors [178–180], the mechanism of genome delivery [25, 181–185], and the mechanics of DNA packaging during viral assembly [18, 172, 173]. In addition, the long-standing interest in the developmental decision between lysis or lysogeny [186] has also been explored at the single-cell level [174–176].

One of the most exciting developments in this field in recent years has been the use of optical trapping to monitor the force build up during the packaging of viral DNA [18, 172, 173]. These seminal experiments watched individual viruses as their genomes were being packaged by the portal motor and measured the

force that builds up as a function of the amount of DNA that has been packed into the capsid. The insights resulting from these measurements led to an explosion of subsequent models and related experiments that have led to a largely self-consistent picture of the free energy cost to tightly compact the viral genome to a volume fraction of roughly 0.5 [21, 187–189].

Despite these advances in the study of genome mechanics in bacterial viruses, controversy surrounds the nature of the mechanisms responsible for driving DNA translocation from the viral capsid across the membrane into the cytoplasm [181]. Specifically, whether or not the energy stored in the compacted viral DNA plays an active role in the DNA translocation process *in vivo* is in dispute: one of the hypotheses that has been advanced as a result of the optical trapping packaging studies is that the stored free energy of packing is used as the driving force for ejection of the viral genome [18, 172, 173], a hypothesis supported by several different experimental approaches to the *in vitro* ejection problem [23, 182, 183, 190–195].

However, a pressure-driven mechanism alone is insufficient to account for the entire DNA translocation process. This is known as the *in vivo* pressure conundrum. Simply stated, it is thought that the turgor pressure [196, 197] 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 [198]. The driving force is thought to equalize with the cell pressure at around 50 percent of genome ejection [23, 199]. Similarly, the question of whether energy-consuming reactions must take place to chaperone the lambda DNA across the cell membrane in living cells has not been answered either, as was shown to be the case in T7 and phi29 [26, 184]; although lambda phage can invade energy-depleted cells, the velocity of ejection is unknown in this case [200]. T5 requires protein synthesis of phage proteins for infection but does not require metabolic energy, per se, for genome delivery [201, 202]. However, the bulk measurement of ejection dynamics does not account for non-synchronous ejection events, cell-to-cell variability in ejection dynamics, and have poor time resolution.

Several models have been proposed to explain how the remaining viral DNA is delivered; these models include active transport by motors, hydrodynamic drag, as well as stochastic mechanisms such as Brownian ratchets and Langmuir forces [22, 185, 187, 198, 203]. All of these models are intriguing and have been the subject of theoretical inquiry, but it is clear that determining which model best reflects the ejection process for phage lambda requires a new generation of experiments with single cell and single phage resolution in order to quantify the *in vivo* ejection process and provide precise facts concerning the mean rate of ejection and its variability as a function of key tunable parameters such as the viral genome length. The absence of such data makes further definitive modeling efforts difficult and the goal of the present work is to provide such data.

4.1 Dynamics and Force Measurements for *in vivo* Ejection

4.1.1 Real-time *in vivo* Ejection Assay

Building on the established *in vitro* DNA ejection assay

Our objective is to quantitatively image the DNA translocation process *in vivo* and to measure the ejection process in real time at the single cell level with the aim of obtaining experimental data capable of discriminating between the models proposed by the literature.

We first review the *in vitro* phage ejection protocol, which will serve as the basis of the corresponding experimental framework *in vivo*. The *in vitro* phage DNA ejection experiment, first developed by Mangenot *et al.* and later adapted to use with phage lambda [24, 191, 193], is as follows: 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. A solution of 10^8 - 10^{11} pfu/ml lambda phage was incubated in the assembled flow chambers for 10 minutes. Once focused, the chamber was washed with 200 μL of buffer + 1% oPOE. Buffers were either SM buffer for the SYTOX Orange measurement or 10 mM Tris, pH7.5, 2.5 mM $MgSO_4$. The solution to induce ejection consisted of buffer, 1% oPOE, 1% glucose oxidase/catalase, 1% LamB, 0.5% glucose, 1% betamercaptoethanol, and either 10^{-6} diluted SYBR gold or 500 nM SYTOX Orange. Calibration of lengths and data analysis was performed as in [193].

The concept of the experiment is in direct analogy to the Hershey-Chase experiment [204], where it was first directly shown that phage DNA is transferred directly into bacterial host and that DNA is the molecule of genetic inheritance. In the work described here, we adapt a similar strategy (see Figure 4.1a, b) to that exploited in the Hershey-Chase experiment, but with DNA labeled with fluorescent dyes as opposed to radioactive phosphorus to track the viral genome during an infection. By looking at this problem from a single molecule perspective, we are able to observe the cell-to-cell variability in the translocation process with high temporal resolution and gain insight into the mechanism of *in vivo* DNA ejection.

A schematic of our experimental design is shown in Figure 4.1a. To visualize DNA translocation, we use a fluorescent marker to stain the viral DNA while it still remains in the capsid [205]. The stained phages are then bound to a bacterial cell and then imaged with fluorescence microscopy over a sufficiently long time to monitor the infection process. The signature of an ejection event is a loss of fluorescence in the virus and a concomitant increase in the fluorescence within the bacterial cell (4.1b). The phage infection process is not compromised by fluorescent dyes. First, phages stained with cyanine dyes have previously been observed to infect cells [206]. Second, studies characterizing the interaction of cyanine dyes with DNA inside the phage capsid have demonstrated that for some dyes, stained phages remain intact [205]. Cyanine dyes have also been used to study the kinetics of viruses in living eukaryotic cells, demonstrating that some dyes have limited cytotoxicity [207]. 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, (iv) not cross the membrane of living cells, and (v) be sufficiently bright for quantitative analysis. Our screen identified SYTOX Orange as a dye with all of these properties.
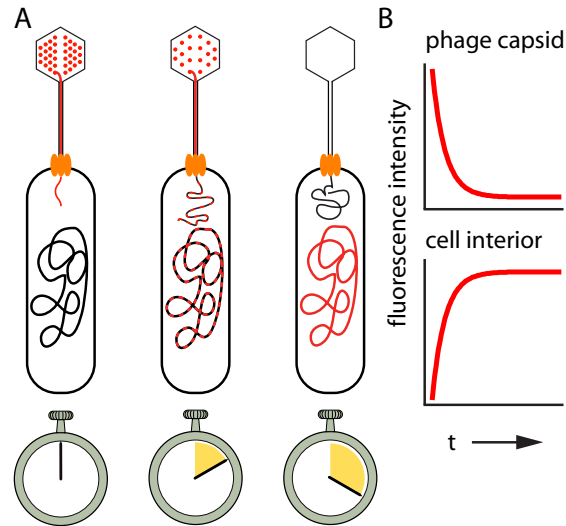
Figure 4.1: **A schematic for monitoring DNA translocation with pre-ejection labeling.** (a) The DNA is stained while still in the capsid. During ejection, the phage DNA carries its complement of cyanine dye with it, transferring fluorescence intensity from the capsid to the cellular interior. Eventually, the dye falls off the phage DNA and rebinds to the bacterium's genome. (b) The timing of ejection is determined by measuring the loss of fluorescence intensity from the capsid and the concomitant increase in intensity in the cellular interior.

Developing an *in vivo* version of ejection assay

The final experimental procedure for imaging DNA ejection *in vivo* is summarized as follows:

Glass coverslips were cleaned by sonication for 30 minutes in 1 M KOH followed by sonication in 100% ethanol with copious rinsing with purified water in between, and then dried on a hot plate. The coverslips were then briefly (5 seconds) immersed in a fresh solution of 1% polyethyleneimine, transferred into purified water, and finally dried with a stream of air. A microscope slide, double-sided tape, and the treated coverslip were then assembled into a flow chamber.

E. coli strain LE392 was grown up overnight in LB media at 37 ^{o}C. The saturated culture was then diluted 1:100 in M9 maltose-sup and grown for 3 hours at 37 ^{o}C until the culture reached OD600 $\sim$ 0.3. Debris and cellular contaminants were removed from the plate lysate by centrifuging the lysate for 5 minutes at 13,000 g. The supernatant was recovered and then stained with SYTOX Orange at a final concentration of 500 nM for 3 hours at room temperature. Prior to binding stained phages to *E. coli*, free dye was removed by diluting 100 μL of the phage suspension and then centrifuging the sample across a 100 KDa filter 4 times. Each round of centrifugation led to a 40-fold dilution of dye, reducing the final free concentration of dye to less than 200 pM. After the final round, the phages were brought up to the original volume of 100 μL with M9sup. Phages were then bound to cells either by mixing $\sim$ 50 μL of cells with phage at a multiplicity of infection (MOI) of $\sim$1 for 30 minute on ice. The cells were then flowed into the flow chamber and allowed to adhere to the surface for 2 minutes on an ice block. This method allows us to achieve desired MOI, as shown in Figure 4.2. Bulk measurements of phage binding and infection kinetics show that the initiation of phage ejection is slowed down considerably within this time period in this condition [176, 180, 208].

After the infected cells adhered to the flow chamber, the flow chamber was washed with 200 μL of M9sup with 1% GODCAT mixture, and 0.5% glucose . The chamber was then sealed with valap and imaged on a Nikon Ti-E Perfect Focus microscope. Samples were illuminated with mercury lamp excitation passed through a TRITC filter set. An environmental chamber was used to keep the sample at 37 ^{o}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 Hamamatsu C8484 camera, a Photometrics CoolSNAP ES2 camera, or an Andor iXON EMCCD camera. We observed better conservation of fluorescence between the phage and the cell with the Hamamatsu and Photometrics cameras as opposed to the Andor camera. The EM gain of the Andor camera allowed for shorter exposure times and higher time resolution.

Finally, for the 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 time point in the movie were registered with the first frame by cross-correlation; this registration was then applied to all fluorescence channels, removing spatial drift from the data set. Next, the registered phase images were all added together; this step greatly reduces difficulty of segmentation. Lastly, thresholding and minor morphological operations were

(a)

(b)

MOI=0.5

(c)
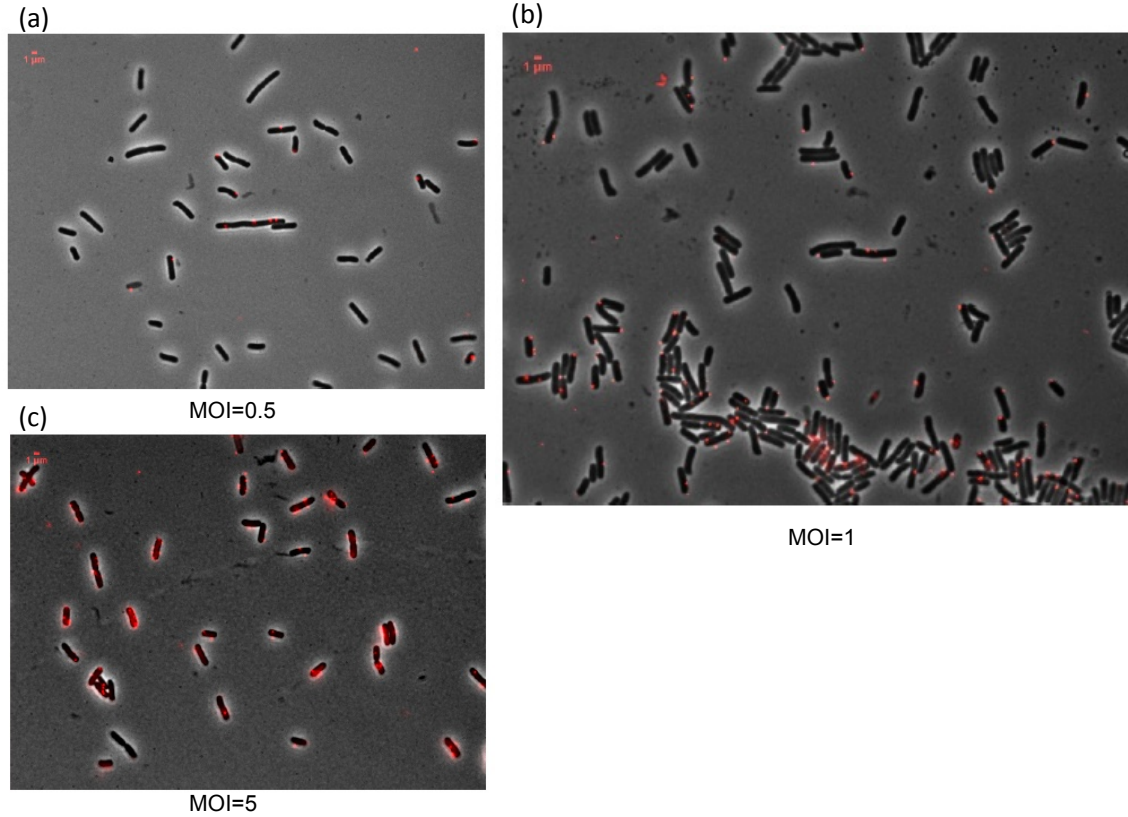
MOI=1

MOI=5

Figure 4.2: **Microscopy images of bacterial cells with attached phage Lambda particles** (stained with SYTOX Orange and pseudo-colored in red) shows control of the multiplicity of infection (MOI). (a), MOI= 0.5. (b), MOI=1. (c), MOI=5.

used to create a mask of the cell from the combined image. Cells were occasionally segmented manually when quality phase images were not available. Phage masks were created from the fluorescence channel by a similar process. Quantities of interest, including background levels, and fluorescence intensities inside the cell and phage in each frame were extracted for further analysis.
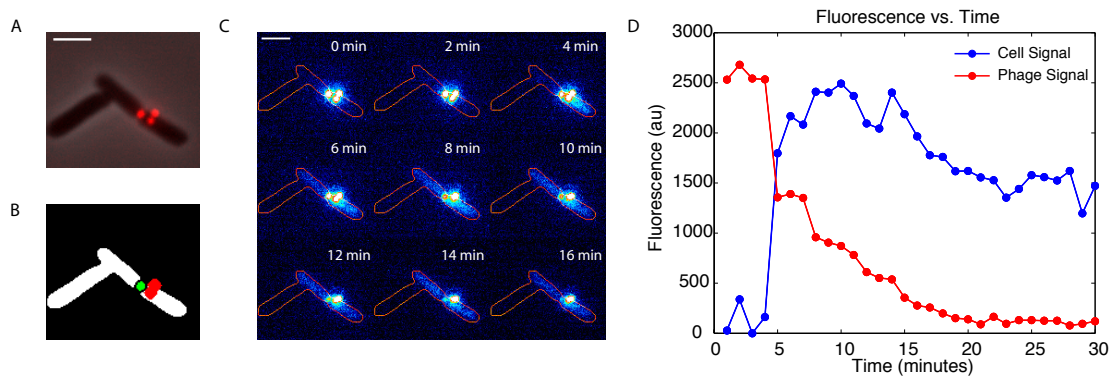
Figure 4.3: **Dynamics of DNA ejection.** (a) Viruses attached to the cell surface in this fluorescence image merged with its brightfield counterpart. (b) Segmentation masks of the cell (white), the phage which ejects its DNA (green), and the phages that do not eject their DNA (red). (c) Time sequence of the fluorescence in the cell. The edge of the cell is outlined for reference. (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 (a) and (c) is 2 microns.

Validating the experimental protocol: Potential issues with SYTOX Orange

We first show that SYTOX Orange does not alter the ejection process *in vitro* or *in vivo*. The first set of control experiments was aimed at examining how SYTOX Orange affects the infectivity of the phage. We first verified that SYTOX Orange could penetrate the phage capsid and stain phage DNA. This was done by simultaneously staining phage with SYTOX Orange and DAPI, a dye that readily penetrates the phage capsid [176, 209]. The results of this experiment are shown in Figure S1; the spatial co-localization of signal from both fluorescent dyes demonstrates that SYTOX Orange penetrates the phage capsid.

One important question concerning the presence of the SYTOX Orange dye is the extent to which it perturbs the dynamics of viral DNA ejection. This is an important control because whereas SYTOX Orange stains inside the capsid, other dyes allow visualization of DNA ejection but do not permit intra-capsid staining [193]. To explore this question, we used SYTOX Orange stained phages as the basis of an *in vitro* single-molecule experiment that has been developed for phages T5 and lambda [24, 191–193]. In these experiments, phages are adhered to the bottom of a flow chamber and coerced by a trigger protein to eject their genomes into free solution. Fluorescent DNA staining dyes enable the DNA to be visualized and a flow allows the length of the ejecting DNA to be measured; see the Materials and Methods for more details. Montages of characteristic ejections are shown in Figures 4.4a and 4.4b. We performed an *in vitro* single-molecule ejection assay using phages whose genomes were already stained with SYTOX Orange; these phages were ejected into a solution containing SYTOX Orange to ensure enough dye molecules were bound to enable visualization. The results of this experiment were compared to an experiment where phages were ejected into a solution containing SYBR Gold. The primary difference between the two measurements is that dye was present inside the capsid prior to ejection in the SYTOX Orange experiment whereas dye was only bound to the ejecting DNA in the SYBR Gold experiment [193]. The ejection trajectories for both sets of experiments were queried to find the average velocity of the ejecting DNA as a function of the amount of DNA remaining inside the capsid. This quantity is shown in Figure 4.4c. While there are some measurable differences between the two experiments (which are likely due to ionic differences between the two buffers), the key features of the dynamics are preserved. From this measurement, we conclude that the presence of dye inside the capsid during the ejection process likely has little impact on the dynamics.

We next examined the extent to which these dye molecules alter the macroscopic titers of infectious phage. In the bulk titer assay, phages are spread on a bacterial lawn and the viral plaques are counted as a way of measuring the number of viable phages in a given volume of solution. For this control experiment, we wanted to make sure that SYTOX Orange stained phages permit the life cycle of lytic phage and their host bacteria to continue. For this experiment, a sample of our phage λcI60 was stained with 5 μM of dye and then subjected to the appropriate dialysis. An unstained sample was also subjected to dialysis for comparison. The unstained phage had a titer of 5×10^{10} while the stained phage had a titer of 3.9×10^{10}, a drop of 20%. These results demonstrate that the presence of dye inside the capsid does not have a significant effect on the infectivity of lambda phage in bulk.

As a second control for infectivity, we imaged the lytic cycle of SYTOX Orange stained phages at the single cell level. Briefly, phages were stained with SYTOX Orange and unbound dye was removed by dialysis. The stained phages were then bound to cells, placed on an agar pad, and imaged. A single image was taken in fluorescence and the cells were subsequently followed with bright-field microscopy. Only one fluorescence image was taken to mitigate any possible photo-damage from the excitation of SYTOX Orange, as an oxygen scavenging system was not present. A montage of a characteristic cell is shown in Figure 4.4d. As seen in Figure 4.4d, cells with fluorescence were seen to lyse. We assume that fluorescence inside a cell signifies that the cell had previously been infected by phage. We demonstrate later that this must be the case by showing that SYTOX Orange does not cross intact bacterial membranes. All of the fluorescent cells that we observed went on to lyse (n=23). Cells with bound phages were also observed; of these, 80% went on to lyse (n=18). This measurement in conjunction with the bulk titering measurement demonstrates that SYTOX Orange does not interfere with the lytic pathway in any substantial way.

(a)

(b)

(c) Velocity vs DNA Ejected: SYTOX Orange vs. SYBR Gold

(d)

(e) No Phage Internal Control
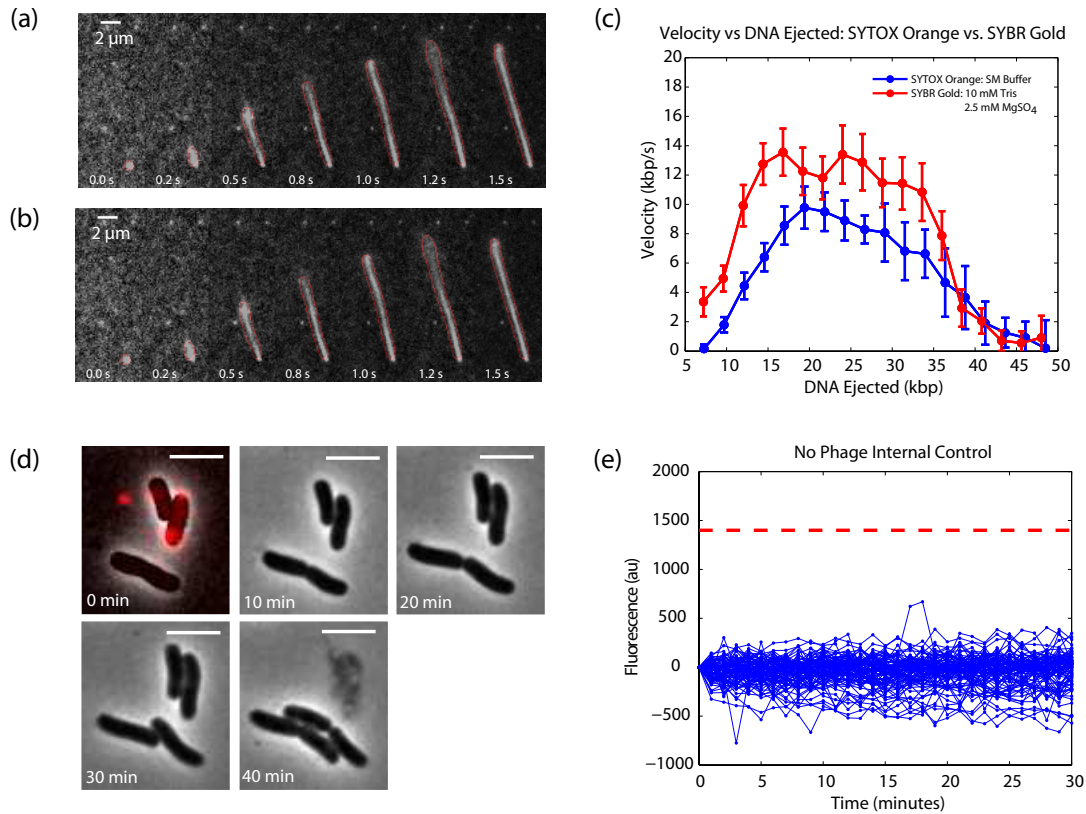
Figure 4.4: **Key controls for the single-molecule** *in vivo* **DNA ejection experiment.** (a,b) Data for the single-molecule *in vitro* DNA ejection experiment. An *in vitro* ejection experiment was performed on unstained phages (a) and pre-stained phages (b) to determine whether the presence of dye inside the capsid has any effect on the ejection dynamics. SYBR Gold was used in the experiment shown in (a) and SYTOX Orange was used in the experiment shown in (b). A MATLAB script was used to segment the DNA; the pixels identified as DNA are outlined. (c) Velocity vs DNA in the capsid for the single-molecule *in vitro* DNA ejection experiment. The presence of dye inside the capsid does not change the underlying dynamics of the ejection process. The difference between the two conditions can be attributed to subtle ionic differences in the buffers for the two experiments. (d) A montage demonstrating that cells that have been infected with stained phage will lyse. Scale bar: 4.8 microns. (e) No phage control. The fluorescence of 89 cells with no phage attached were monitored over the course of a data set. The red line indicates the average change in fluorescence of phages that ejected within the data set. No increase in fluorescence is seen, demonstrating that SYTOX Orange does not cross the membranes of living cells.

Validating the experimental protocol: Photobleaching or Cross-talk issues

We confirm that SYTOX Orange has low membrane permeability and suitable photobleaching properties. One concern about 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[210]. In such a scenario, spurious fluorescence would result in confounding effects which would interfere with our ability to identify and monitor putative ejection events. To mitigate the effects of oxidative damage during our high frame rate acquisitions (1-4 frames per minute), we used the GODCAT system (1% glucose oxidase-catalase mixture, 1%βME, and 0.5% glucose) to prevent the formation of oxygen free radicals. Glucose oxidase is also known to be protective to *E. coli* in the presence of oxygen radicals [210]. We also note that the GODCAT system is also known to mitigate photobleaching of fluorescent dyes, another important reason for its use. To investigate whether spurious increases in fluorescence occur, we examined phage-free cells in one of our *in vivo* ejection data sets. We segmented images in this data set for every cell without bound phage and then monitored the internal fluorescence of those cells over time. As can be seen in Figure 4.4e, there is no evidence from any of the 89 cells examined that there is an increase in fluorescence in the absence of phages. From this analysis, we conclude that spurious increases in fluorescence, if they do occur, are very rare events. Thus, when we do see an increase in fluorescence, we conclude the result must have been from phage ejecting its DNA into the cell interior.

Two other issues that might affect the quantitative interpretation of our *in vivo* ejection measurement are photobleaching and dye leak through the capsid: issues known to be concerns for fluorescent dyes. Photobleaching and dye leakage were quantitatively characterized by time-lapse imaging of stained phages bound to coverslips and analysis of non-infecting phages bound to cells. We assume in these cases that all loss of fluorescence is due to a combination of photobleaching and dye leakage. These measurements (shown in Figure S2-S5 and discussed in more detail in the Supplemental Information) show that these effects cannot account for the decrease in intensity observed in the trajectories shown (Figs. 4.3d, 4.15, and S6-S20). Furthermore, the infecting viruses show a rapid decay in fluorescence while the cellular cytoplasm shows a concomitant rise of fluorescence. This observation is consistent with the delivery of dye-bound DNA to the cell. The key result of these various controls is that the presence of SYTOX Orange in our phages does not materially alter the underlying dynamics of the viruses. More precisely, these experiments demonstrate that the cases where we see a simultaneous reduction in the fluorescence intensity within a phage and a corresponding increase in the fluorescence intensity of the bacterial host are actual infection events.

In order to properly characterize the ejection events, photobleaching and viral DNA ejection must be distinguishable. To explore the relative time scales of putative ejection events and photobleaching, we measured the rate of photobleaching and demonstrated that the time scales of ejection and photobleaching are well separated, as shown in Figure 4.5. Phages were prepared with SYTOX Orange as described in

the Materials and Methods. Briefly, λcI60 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 an observation 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, precisely the same conditions used in the *in vivo* experiment. The fluorescence images were processed using a custom MATLAB script.

This data (Figure 4.5) demonstrates that while significant photobleaching does occur over the course of 30 minutes, ejections, as shown in Figures 4.4 and 4.3 , range from 1 minute to 20 minutes. Thus for the typical ejection time scale of 15 minutes, photobleaching can account for at most a 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.
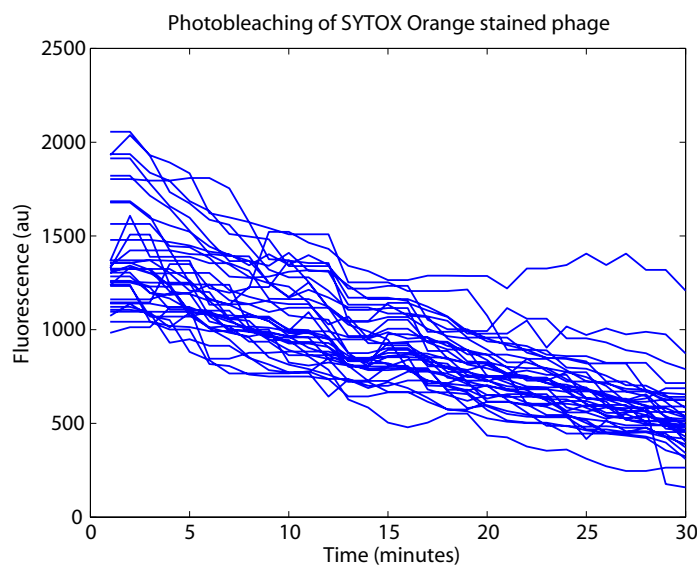
Figure 4.5: **Photobleaching in SYTOX Orange stained bacteriophage.** The trajectories of the total fluorescence above background for 36 phages are shown. The time scale for photobleaching is 30 minutes, and all the trajectories are monotonically decreasing.

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 4.6. This data is consistent with Figure 4.5 and demonstrates that the loss in fluorescence in ejecting phages

occurs on a faster time scale 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 4.5). The red lines, therefore, represent the fluorescence of non-ejecting 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.
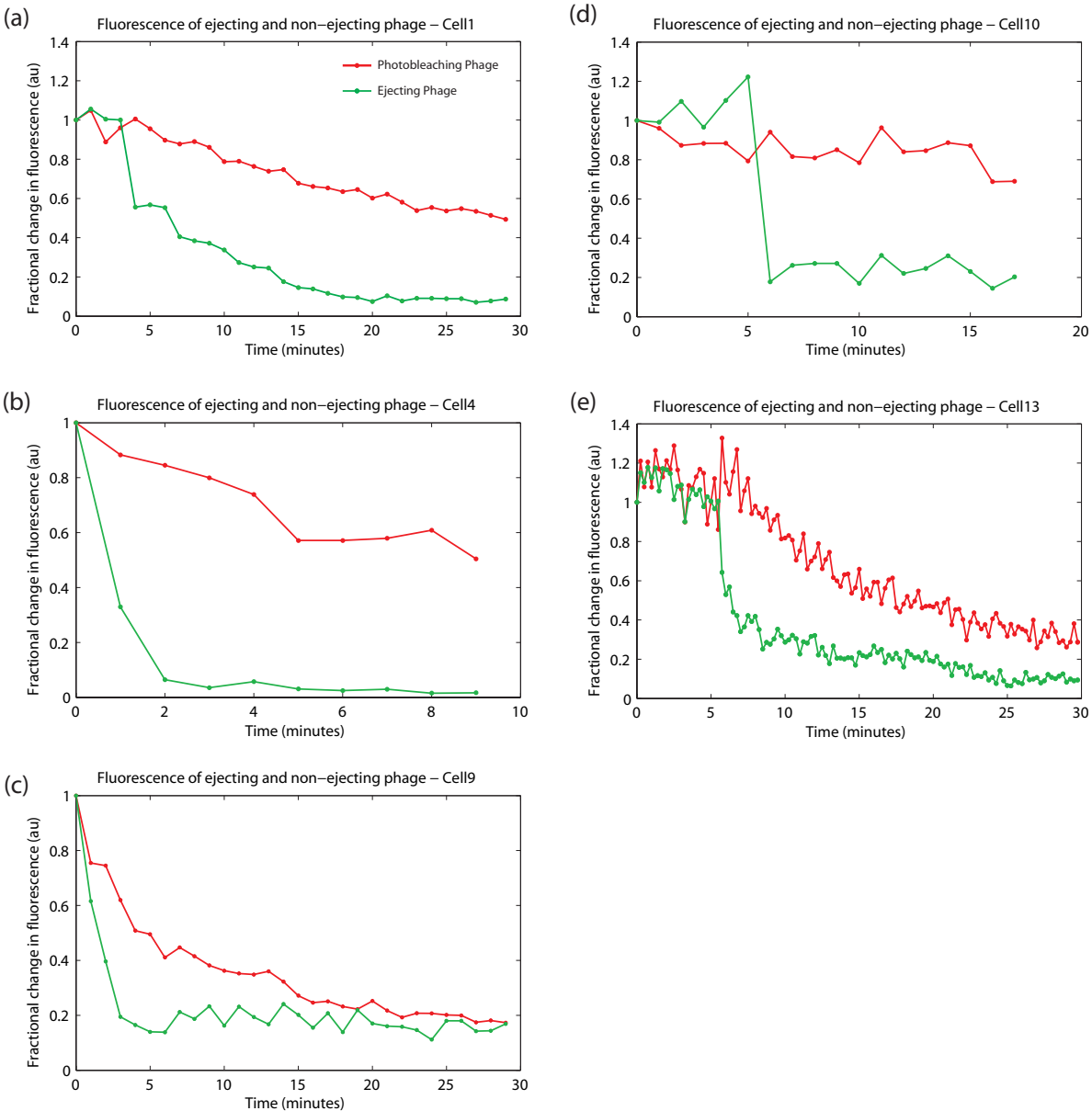
Figure 4.6: **Photobleaching in SYTOX Orange stained bacteriophage bound to cells.** Several cells had multiple stained phages that were bound simultaneously. Some of these phages did not display a significant decrease in fluorescence, indicating that they did not infect the cell. Here we plot the decrease 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 segmentation masks. 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.

Validating the experimental protocol: Accurate dynamics

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 [205]. We compared the SYTOX Orange stain with DAPI, which is known to be a quantitative indicator of DNA mass in phages [209]. As shown in Figure 4.7a, 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 4.7b simultaneous staining and co-localization of the viral DNA with both DAPI and STYOX Orange, confirming that SYTOX Orange is both a sensitive and specific indicator for the presence of phage lambda.

DAPI TRITC

Figure 4.7: **Simultaneous staining of phage λcI60 with SYTOX Orange and DAPI.** Here, we verify that SYTOX Orange will penetrate the capsid and stain ph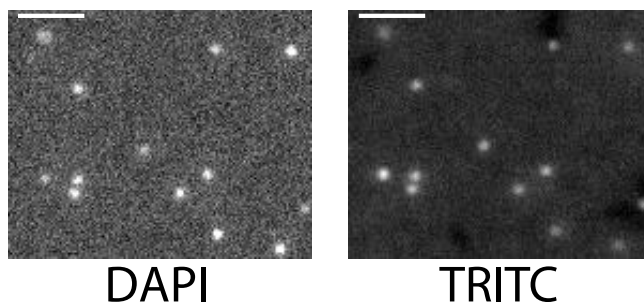age DNA. Phages were stained with a 0.5 μg/ml concentration of DAPI and 500 nM SYTOX Orange, then flowed into an observation chamber. It has been previously established that DAPI will readily penetrate the phage capsid and stain phage DNA [176, 209]. We observe perfect co-localization of the DAPI and SYTOX Orange signals, demonstrating that SYTOX Orange will enter the phage capsid stain the phage DNA. 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) [193]. The scale bar is 2 microns.

We then confirmed that SYTOX Orange unbinds from naked DNA on the timescale of seconds. As indicated schematically in Figure 1 in the main text, the loss in fluorescence from the virus should be accompanied by a concomitant increase in the fluorescence of the cell itself. This claim is predicated upon the idea that the fluorescent dye labels rapidly detach from the viral genome once they have left the confines of the viral capsid. This expectation is supported by earlier experiments which measured the kinetics of these dye molecules both for naked DNA and DNA within capsids [205]. These experiments demonstrate that for DNA within the capsid, the time scale of dye kinetics is of order hours while the dynamics of binding and unbinding of dyes on naked DNA takes seconds [211]. We confirmed this by performing our *in vitro* ejection assay on prestained phages, where they release their genome into a solution that has no free dye molecules. In this case, what is seen is a steady decrease in the fluorescence intensity of the viruses themselves, the signature of DNA exiting from the capsid, but no stretched DNA characteristic of the usual

in vitro assay because as soon as the DNA leaves the capsid, the dye falls off of the DNA, as shown in Figure 4.8. Furthermore, the kinetics of fluorescence loss in the phages matches the kinetics of *in vitro* ejection [24].

We also performed an *in vitro* ejection experiment with SYTOX Orange instead of SYBR gold following the method in [193], but exchanged buffer with dye to buffer without dye after completing DNA ejection; as is shown in Figure 4.9, the fluorescence signal drops by 70% within 7 seconds. From [211], the k_{off} for SYTOX Orange is 0.58 s^{-1}, and they achieve a 38% reduction in staining with a 100-fold dilution of 500 nM SYTOX Orange within a few seconds. In our view, these experiments support the interpretation of the *in vivo* ejection assay that when the viral genome enters the cell, the bound dyes fall off and they can then bind onto the much larger host cell genome, allowing us to visualize intracellular fluorescence.
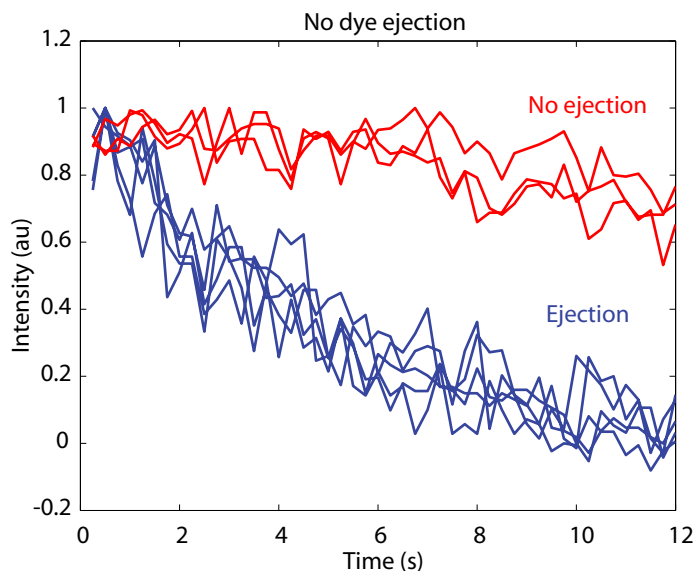
Figure 4.8: **Single-molecule *in vitro* DNA ejection of SYTOX Orange stained phages in the absence of external dye.** The phage capsids are rapidly destained as DNA is ejected out (blue) in the presence of lambda receptor, LamB, and dye falls off. The kinetics are consistent with *in vitro* ejections [24] in which the ejected DNA was stained, instead of the phage capsid DNA, as is shown here. Compare with phage capsids which did not eject (red) which demonstrates a drop in fluorescence intensity presumably due to photobleaching. The kinetics are markedly different. These ejection curves demonstrate that the *in vitro* kinetics of ejection are unaffected by staining the phage capsid DNA with SYTOX Orange.
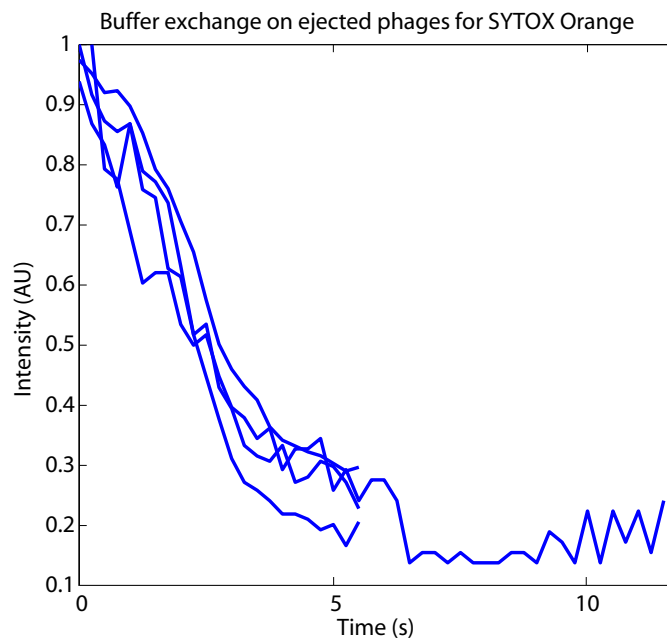
Figure 4.9: **Single-molecule** *in vitro* **DNA ejection in the presence of SYTOX Orange.** In order to probe dye unbinding kinetics, ejected phage, stained with SYTOX Orange, is destained and imaged over time. At early times, the buffer is exchanged for buffer without dye. The fluorescence signal on a strand of DNA decreases to background within 7 seconds. This is consistent with previous experiments [211] and also with our observation that dye coming off the phage DNA rebinds to cellular DNA.

Preparing phage Lambda sample

We use the plate lysis method to obtain high concentration phage sample (compared to liquid lysis). A typical plate of bacterial lawn with viral plaques is shown in Figure 4.10. We did not perform a CsCl density-gradient centrifugation purification, but since we can directly visualize the SYTOX Orange-stained phage particles infecting the cells, and observe cell bursts following the infection, this guarantees that these phage particles are infectious and have intact capsids and genomes. To prepare the plates, 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 ^{o}C water bath until needed. The host cell strain LE392 was grown up overnight in 5 mL of LB. The saturated cell culture was then centrifuged for 5 minutes at 5,000 g and the pellet was resuspended in 5 mL of SM buffer. A 100 μL aliquot of cells was then mixed with 1 μL of phage stock in a 14 mL culture tube and incubated at 37 ^{o}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 ^{o}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 ^{o}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 were then removed by centrifuging for 10 minutes at 5,000 g; the supernatant was recovered. Plate lysis typically yielded titers of $\sim 10^{10}$ pfu/ml. A similar procedure, but with more diluted phage, is used to determine the infectious phage concentration (titer).
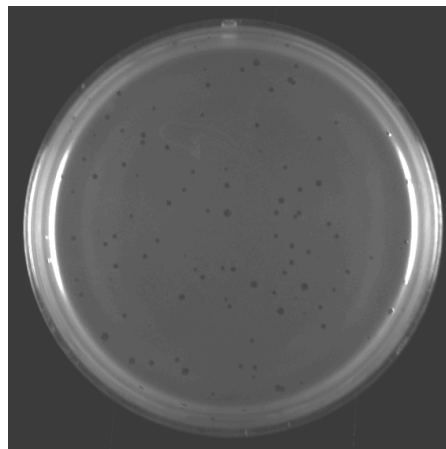

Figure 4.10: **A bacteriophage Lambda plate** used for plate lysis or titer. The agar plate is filled with a thin lawn of bacteria (opaque), and the individual plaques (clear) are regions where bacterial cells are digested, starting from a single infection event.

4.1.2   A Scheme to Verify Other Possible Ejection Driving Forces

A number of different hypotheses have been formulated for the actual translocation mechanism for phage λ, some of which are highlighted in Figure 4.11. 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 [203, 212–214]. The dynamics of viral ejection is governed by both the driving force associated with the packaged DNA and also the friction arising from any number of distinct sources. Hypotheses for the contributions to the driving force range from active cellular motors, free energy stored in the packaged phage DNA, to an osmotic pressure gradient between the cellular interior and the environment, and are likely different from one phage species to the next. On the other hand, the friction from packaged DNA-DNA interaction, DNA-phage tail tube and DNA-viscoelastic cellular environment interactions may all play important roles in the ejection dynamics, but are all difficult to model or measure at this point due to the rich and complicated nano-scale hydrodynamic and polymer physics present in this system. In fact, even in the case of *in vitro* ejection, the value of the friction was shown to be varying during the ejection process and its underlying physical origins are still under debate. Hence, in order to develop a mechanistic understanding of the ejection process, it is necessary to supplement the single-molecule Hershey-Chase experiment which records ejection dynamics with an assay not influenced by the friction to determine the ejection driving force separately. Information on the nature and magnitude of the ejection driving force from this static assay can then lead to a precise quantification of the friction and the viscoelastic cellular environment when combined with the single-molecule Hershey-Chase experiment. Therefore, we aim to develop an *in vivo* viral ejection method to probe the driving forces governing bacteriophage ejection dynamics.
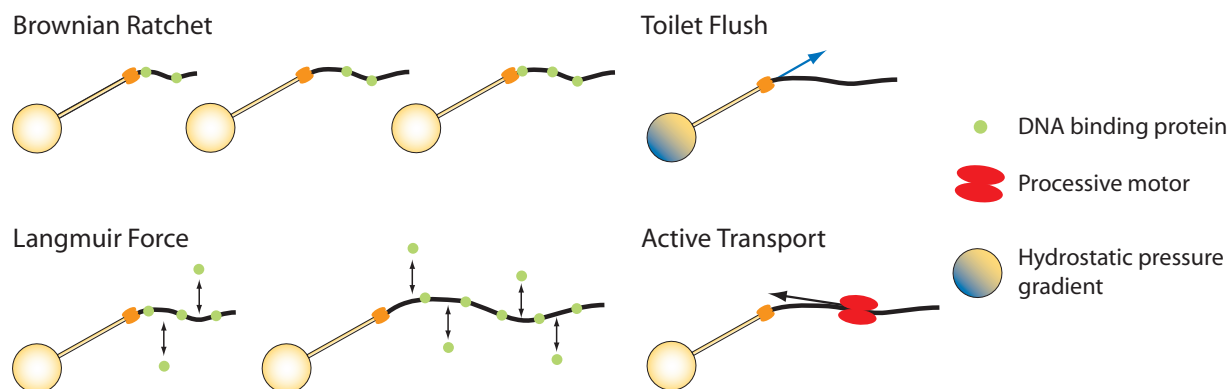
Figure 4.11: **Models for DNA ejection into living cells.** A number of hypotheses have been put forth for how DNA enters the highly crowded cellular interior. The Brownian ratchet and Langmuir models posit that DNA binding proteins attach to the DNA and prevent it from moving back out [212, 213]. The "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 [214]. The active transport mechanism refers to the translocation of molecular motors like RNA polymerase on the ejected DNA once it enters the cell [203].

In our experimental design, we use fluorescence in-situ hybridization (FISH) to examine the arrival of specific parts of the viral genome within the cellular interior. Such a static, single cell experiment is constructed as follows: fluorescence dye coupled short DNA oligos are designed according to the sequence of the bacteriophage genome at various positions (the resolution is about 0.5-1 kbp). Phage stock is mixed with cell culture to allow phage particles to adsorb to cells for 30 minutes on ice and achieve the desired multiplicity of infection. A temperature shift to 37 degrees Celsius starts the infection process. At an arbitrary time point, the reaction is fixated with formaldehyde. The cells are then slightly permeablized and heated to allow hybridization of the fluorescence probes to phage genome loci. Each locus is designed to be probed by a particular excitation/emission spectrum (in principle, more than 32 colors can be achieved by incorporating activator-photoswitchable reporter dye pairs used in super-resolution imaging). [215] Microscopic images distinguish successfully injected genomic loci from ones that are left outside of the cell at the fixation time point. Figure 4.12 shows a preliminary control experiment of this assay for wild type phage T7 infecting *E. coli* cells.
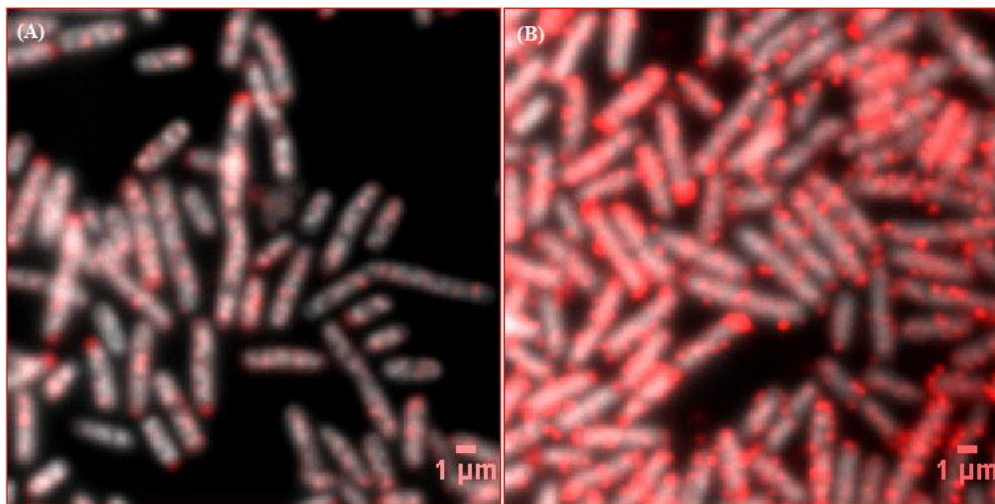
Figure 4.12: **Preliminary phage DNA fluorescence** *in situ* **hybridization control experiment.** A set of fluorescent probes is designed against the 1000 bp ejection start site of the wild type phage T7 DNA and is coupled with the dye Alexa Fluor 647. The cells are fixed at 10 minutes after the temperature shift, hybridized, and stained with DAPI (which traces the bacterial as well as viral DNA) before mounting to microscope. The Alexa Fluor 657 channel is represented in red and the DAPI channel is represented in grayscale. Thresholding is not applied to the images. (A) Null infection control. The cells are treated the same way except that they are not mixed with T7 phage. The low red signal in the cells is resulted from non-specific binding of probes to cellular structures or hydrophobic interaction between dye and lipid molecules that are not completely washed away. (B) The multiplicity of infection is roughly 10. The red phage signal and the intensity of each puncta in the cells is much higher compared to (A), providing preliminary evidence that the internalization of phage genome can indeed be captured by this assay and distinguished from non-specific background.

Since the fixation, permeablization and hybridization steps are optimized for *E. coli* cells, this assay is readily applied to all the different kinds of bacteriophage infecting *E. coli*. Under various perturbations of

the driving forces, the relevant ones for a particular phage would give rise to incomplete ejections, and the magnitude can be quantified by tracing out how much of the phage genome is successfully internalized with multi-color FISH and possibly super-resolution where at least 32 genomic loci are possible. The driving force given by active cellular motors can be knocked out by specific antibiotics, the pressure stored in packaged phage DNA can be tuned by the ionic strength in the surrounding solution, and the osmotic pressure in the environment can be modulated by the addition of polymer osmolytes (calibration of the latter two factors are available in literature and can be compared with previous *in vitro* experiments). These hypotheses regarding driving force will be tested among various double-stranded DNA *E. coli* phages differing in whether they have contractile tails or not, whether they encode their own DNA/RNA polymerases or not, and with very different tail/genome lengths despite the fact that all phages have approximately 0.5 packaging volume density.

The ultimate goal of this experiment is to test the hypotheses of ejection driving force (including active cellular motors, free energy stored in encapsidated viral DNA, and hydrodynamic imbalance), by using drug [216], salt concentration [214], and osmotic [217] perturbations during the ejection process, and determine which one can successfully stall the ejection. An illustration of this scheme is shown in Figure 4.13
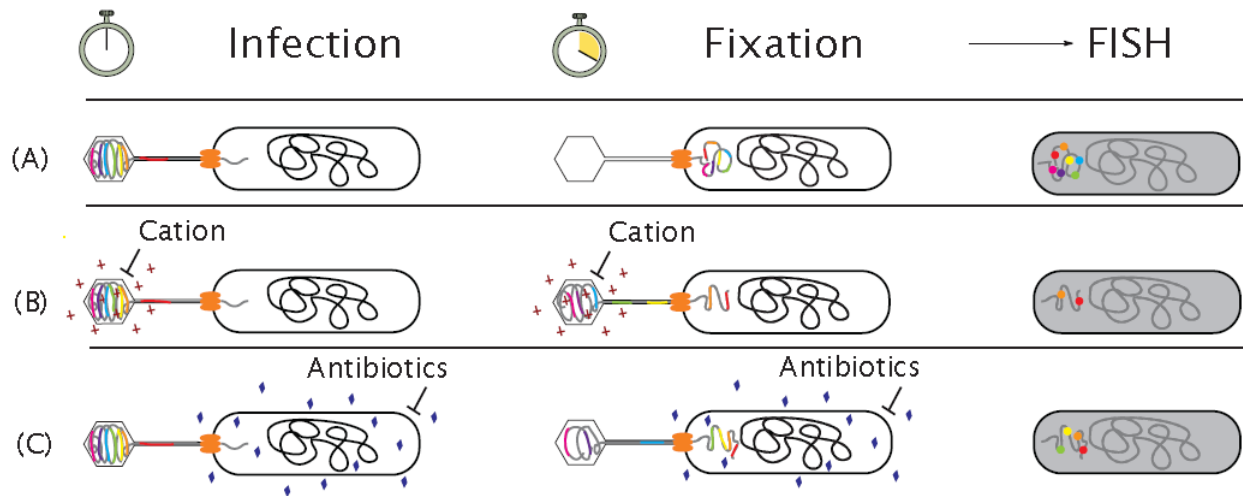
Figure 4.13: **A perturbation scheme to dissect ejection driving force.** Bacteriophage infection is initiated under different conditions, including (A) no perturbation, (B) phage genome condensed by cations of high ionic strength to reduce the store pressure in phage, (C) inhibiting host enzymes that may potentially pull in phage DNA by specific antibiotics. If the perturbations are relevant, ejection should halt, and only a portion of the phage genome would be present in the cell. DNA FISH is used to examine which part of the phage genome is in the cell, as compared to the "no perturbation" case where all the genome is in the cell.

To obtain a systematic solution to the question of bacteriophage *in vivo* ejection driving force, we have to push the FISH method forward to be able to simultaneously monitor two-color FISH.

The demonstration of the single-cell two-color FISH and our attempts to dissect the bacteriophage ejection process is shown in figure 4.14. We design a set of probes (40 nt DNA oligos) complementary to the known bacteriophage DNA sequence to cover about 0.5 -1 kbp of each desired position on the genome. Each set of probes is chemically coupled to a chosen color of fluorescence dyes. In the experiment shown in figure 4.14, the ejection start site of the bacteriophage genome is labeled with Alexa Fluor 647 (Life Technologies) and pseudo-colored in red; the ejection end site is labeled with Alexa Fluor 555 and pseudo-colored in green. When both colors co-localize within a diffraction-limited spot, they add up to produce yellow. We also stain the *E. coli* cell wall with Wheat Germ Agglutinin (WGA)-Alexa Fluor 488 Conjugate (Life Technologies) and pseudo-color the cell wall in white. At this point, the staining efficiency seems to be variable from cell to cell and the protocol requires further optimization. We use a temperature shift from 0 to 37 degrees Celsius to initiate the ejection process, after adsorbing bacteriophage T7 particles onto the *E. coli* cells. We chemically fix and stop the reaction (by formaldehyde) after 2 (figure 4.14(A)) and 10 (figure 4.14(B)) minutes of ejection. The permeablization, hybridization, and microscopic mounting protocols are similar to the standard procedure [215].
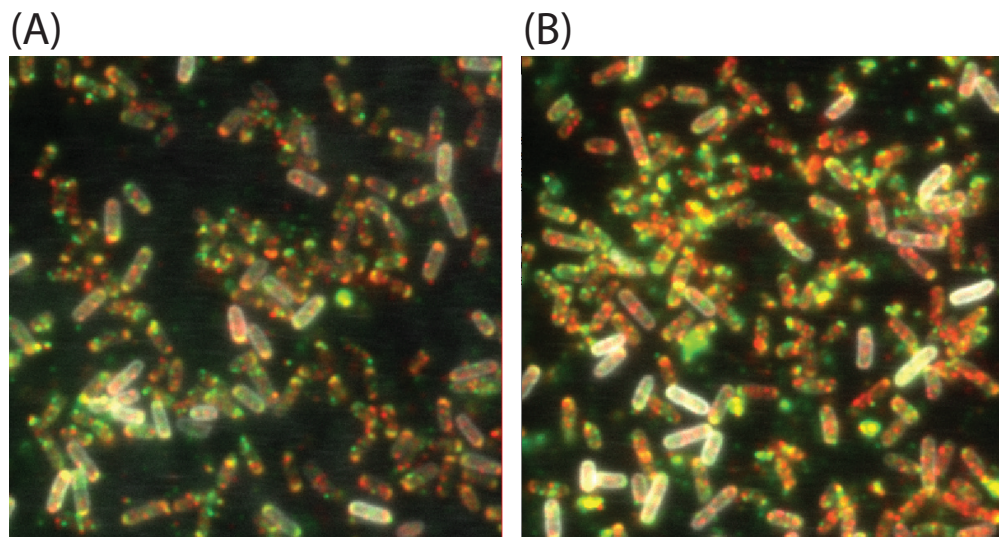
Figure 4.14: **Two-color bacteriophage DNA FISH experiment.** Bacteriophage T7 particles infecting *E. coli* cells are fixed and the DNA is visualized with fluorescence in-situ hybridization. Red: Alexa Fluor 647 probes labeling the ejection start site of the bacteriophage genome. Green: Alexa Fluor 555 probes labeling the ejection end site. Yellow: co-localization of red and green channels. White: *E. coli* cell wall stained with with Wheat Germ Agglutinin-Alexa Fluor 488 Conjugate. (A), the cells are fixed at 2 minutes after the ejection initiation. (B), the cells are fixed at 10 minutes after the ejection. (A) and (B) are at the same contrast. Note that scale bars do not apply to these fluorescence images.

According to previous bulk *in vivo* ejection studies in bacteriophage T7 [216], the ejection should be completed in 10 minutes at 37 degrees Celsius, so both the ejection start and end sites of the genome should

be internalized to the cell. On the contrary, at 2 minutes of the ejection, only the ejection start site should enter the cell, so we expected to see only the red color within the cell. However, comparing figure 4.14(A) to (B), both colors are present within the cell contour outlined by WGA. We conclude that three-dimensional resolution is required to dissect the difference between the two images. The reason is that in figure 4.14(A), the green color may actually be outside of the cell envelope, but our images do not have enough resolution to distinguish the z-positions, so the green and red color may appear to be on top of each other. Techniques such as cylindrical lens (which results in distorted diffraction profile when the spot is at different z-positions) [218] and super-resolution [215] should be adopted to resolve this issue. Alternatively, high-throughput sequencing techniques may serve as an independent confirmation of the microscopy results. We note that the total fluorescence intensity in each cell is on average higher in 4.14(B) than in (A). This is because T7 genome is known to be actively transcribed and replicated during the 10 minutes ejection process, hence more viral genetic molecules have been produces in (B).

4.2 Implications of the Ejection Dynamics

4.2.1 Time Scale and Speed Profile of Lambda *in vivo* Ejection

Briefly, our *in vivo* ejection method uses a fluorescent intercalating dye to monitor the bacteriophage DNA during its translocation from single phage Lambda particles to single *E. coli* cells. The choice of the dye and the micro-environment for the single cell infection are optimized such that frequency of infection and cell growth and lysis following the real-time imaging of ejections are compatible 'with bulk infection assays' with standard bulk infection assays, suggesting that the infection is at physiological conditions.

A typical ejection event for phage λcI60 obtained from this assay is shown in Figure 4.3. As seen in Figure 4.3a, the attachment of the viruses to the host is revealed by the presence of diffraction-limited spots on the cell surface. We identified pixels associated with either the virus or the cell (Fig. 4.3b) and queried the fluorescence intensity as a function of time. As shown in the montage of images (Fig. 4.3c), the ejection process is characterized by a loss of fluorescence intensity in the phage and a concomitant increase in fluorescence in the cellular interior (Fig. 4.3d). Note that the fluorescence inside the cell is diffuse; this reflects the dye molecules unbinding from the phage DNA and redistributing themselves along the host genome. This is expected from the residence time of SYTOX Orange on DNA, which is ~ 1 s [211]. A single-molecule experiment in which SYTOX Orange stained phages are ejected *in vitro* in the absence of external dye is consistent with this expectation; see the Supplemental Information, particularly Figures S2-S3, for more details. In the particular trajectory shown, the increase in cellular fluorescence is roughly equal to the decrease in phage fluorescence. The decrease in signal at the end of the trajectories in Figure 4.3d can be accounted for by photobleaching (Fig. S5). The two signals in Figure 4.3d are analogous to earlier *in vitro* measurements in which DNA ejection from viruses was inhibited through osmotic pressure 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 [183, 193]. Here, by analogy, the DNA ejected is measured through the cellular fluorescence and the DNA remaining in the capsid is similarly revealed by monitoring the fluorescence in those pixels identified as corresponding to the virus.

We performed our *in vivo* ejection assay for λcI60, a strain with a wild type genome length of 48.5 kbp. The results of a number of distinct ejection events are shown in Figure 4.15. 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* [24, 193]. We define the ejection time as the time required for 80% of the fluorescence intensity to leave the viral capsid. The mean time for ejection for λcI60 and the corresponding standard deviation was 5.2 ± 4.2 minutes ($n = 45$). In addition to the widespread variability in ejection time, there are a number of ejections which demonstrate pausing events. We define a pause as a non-decreasing fluorescence level greater than 2 minutes (our time resolution was typically 1 minute, so 2 minutes is the minimum time required to rule out spurious events); the mean

pause time for λcI60 was 5.4 ± 4.1 minutes ($n = 14$). Based on this observation, we have partitioned the trajectories into two classes: single-step and paused. These are shown in Figure 4.15a and Figure 4.15b, respectively.
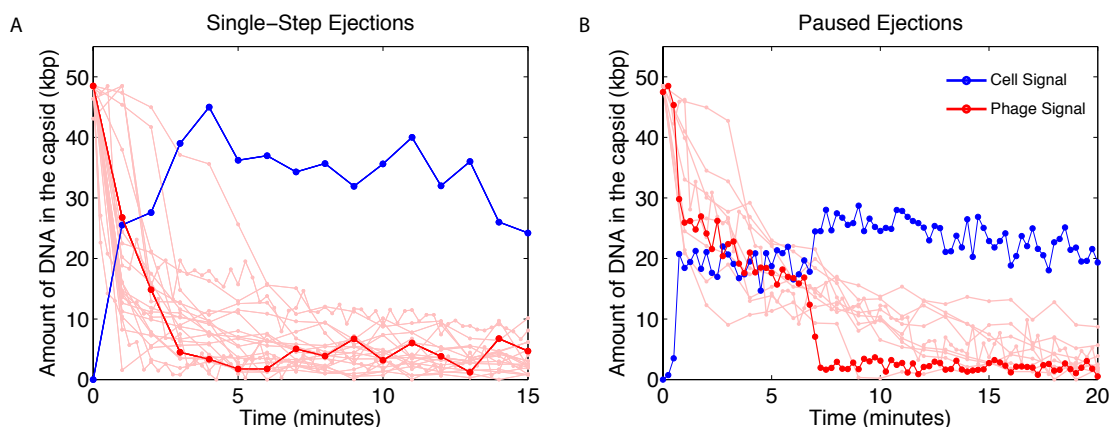

Figure 4.15: **Ejection trajectories from single-cell infections for λcI60.** The red trajectories show the time history of the DNA intensity within the virus and the blue trajectories show the concomitant increase in the fluorescence in the cellular interior. The solid red color highlights a characteristic ejection and the lighter red color displays other ejection events for reference. The conversion between arbitrary units and kbp was done by first subtracting each trace's minimum observed fluorescence from itself. Each trace was then normalized by the maximum drop in the phage fluorescence and then multiplied by the genome length, which is 48.5 kbp for λcI60. Only two representative traces for the intensity within the cell are shown, with the remaining trajectories presented in the SI. (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 SI.

We next asked if a reduction in driving force would produce significant differences in DNA translocation rates, an idea already used in our earlier *in vitro* measurements [23, 24]. In this earlier work, the ejection of phage λcI60, which has a genome length of 48.5 kbp was compared to the ejection of phage strain λb221, which has the b region of the genome removed and a shorter genome length of 37.7 kbp. Through bulk and single-molecule *in vitro experiments*, it was shown that the amount of DNA inside the viral capsid was a control parameter for *in vitro* DNA ejection. Once phage λcI60 has ejected 10.8 kbp of DNA, the ejection forces and dynamics are equivalent to that of phage λb221. To explore the effect of genome length changes on DNA translocation rates *in vivo*, we performed our *in vivo* ejection assay for phage λb221. The full set of trajectories for phage λb221 is shown in Figures S15-S20. The mean time for ejection was 2.58 ± 2.34 minutes ($n = 18$). One paused ejection was also observed, with a pause time of 5 minutes. For λb221, we also observed a number of ejections ($n = 10$) that did not finish during the course of the movie. The trajectories for these ejections, termed stalled ejections, are shown in Figures S19-S20. Stalled ejections were not observed for λcI60. One possibility is that stalling events are related to λb221's shorter genome and the consequent reduction in total driving force. Stalled ejections were not included in the averages given earlier.

Our measurements on both the wild-type and shortened genomes provide an opportunity to quantitatively examine the DNA translocation kinetics. One quantity of interest is the first-passage time for ejection, defined as the time at which a certain amount of DNA has entered the cell. For each trajectory, we extracted the first passage time for 20%, 50%, and 80% completion, as determined by the decrease in the starting phage fluorescence. The first-passage time distributions for λcI60 and λb221 are shown in Figure 4.16a. By taking the mean of this distribution, we obtain the mean first-passage time; this is plotted in Figure 4.16b. We note that the mean first-passage time is a quantity that is amenable to theoretical calculations [219]. Another way to view the mean first-passage time is as an "average" ejection trajectory. When viewed in this way, one interpretation of Figure 4.16b is that, within the error of the measurement, the "average" trajectories for λcI60 and λb221 have considerable overlap.

Another method of analyzing the dynamics is to determine the mean velocity at different amounts of DNA remaining in the capsid during an ejection. In this analysis, the velocity at each landmark is extracted from each trajectory. The mean velocity is then recorded and plotted as a function of the amount of DNA remaining in the capsid. The result of such an analysis for λcI60 and λb221 is shown in Figure 4.16c. This plot shows that there is little overlap between the two curves and that for lower amounts of DNA remaining in the capsid, the velocity for λb221 is higher than λcI60. This is to be contrasted with *in vitro* measurements where there is significant overlap between the two curves, with the "data collapse" in that case signifying that the dynamics are equivalent after λcI60 has ejected its first 10.8 kbp [24]. An alternative way to visualize this same data is to plot the mean velocity versus the amount of DNA ejected into the cell, as opposed to the amount of DNA remaining in the capsid. This is shown in Figure 4.16d. When plotted in this fashion, there is considerable overlap between the two curves, with a small difference observed for the first 20 kbp of ejection. This analysis is consistent with the mean first passage time analysis, which showed considerable overlap when plotted against the amount of DNA ejected.

For bacterial viruses, genome delivery is at the heart of the viral life cycle. And yet, this critical process of transfer of viral DNA from the virus to its host remains enigmatic as does the *in vivo* process of polymer translocation more generally. Beyond a purely intellectual understanding of this process, phage-mediated horizontal transfer of virulence factors is a noted cause of world-wide dysentery [220, 221]. Our objective was to design and perform an experiment with single-molecule resolution that would permit us to watch this process in real time, to measure the specific functional form of the speed of ejection as a function of the amount of DNA left in the capsid (and hence as a function of the amount out as well) and to characterize the cell-to-cell variability of that process. Using a fluorescent staining strategy in conjunction with video microscopy, we have performed such measurements for both wild type lambda phage (48.5 kbp) and a mutant with a shorter genome (37.7 kbp). The results of these experiments are featured in Figures 4.3, 4.15, 4.16 and S6-S20. 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. A number of ejections exhibited pauses with some never reaching completion during the course

of the experiment. We note that our single-molecule measurements are consistent with earlier estimates of a minute time scale for *in vivo* genome delivery of phage lambda from bulk experiments. This is because the bulk estimate should only be compared to the fastest observed single-molecule ejections, which were completed within one minute [25, 26].

A number of different hypotheses have been formulated for the actual translocation mechanism for phage λ, some of which are highlighted in Figure 4.11. 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. With respect to identifying the correct ejection mechanism for phage λ, our results provide both surprises and useful insights that constrain the space of possible models and will guide future modeling efforts. One key result is that the length of DNA remaining inside the capsid is not the sole control parameter that governs the ejection dynamics, as it is *in vitro*. In the *in vitro* experiments, the approximate collapse of the data from the different genome lengths on a single curve revealed that the DNA within the capsid is driving the kinetics of ejection [23, 24]. By way of contrast, in the *in vivo* ejection experiments reported here, an approximate data collapse is only revealed when the velocity is plotted with respect to how much DNA is *out* of the capsid and in the cell rather than how much DNA remains within the capsid. No collapse is seen when the velocity is plotted against the amount of DNA remaining inside the capsid.
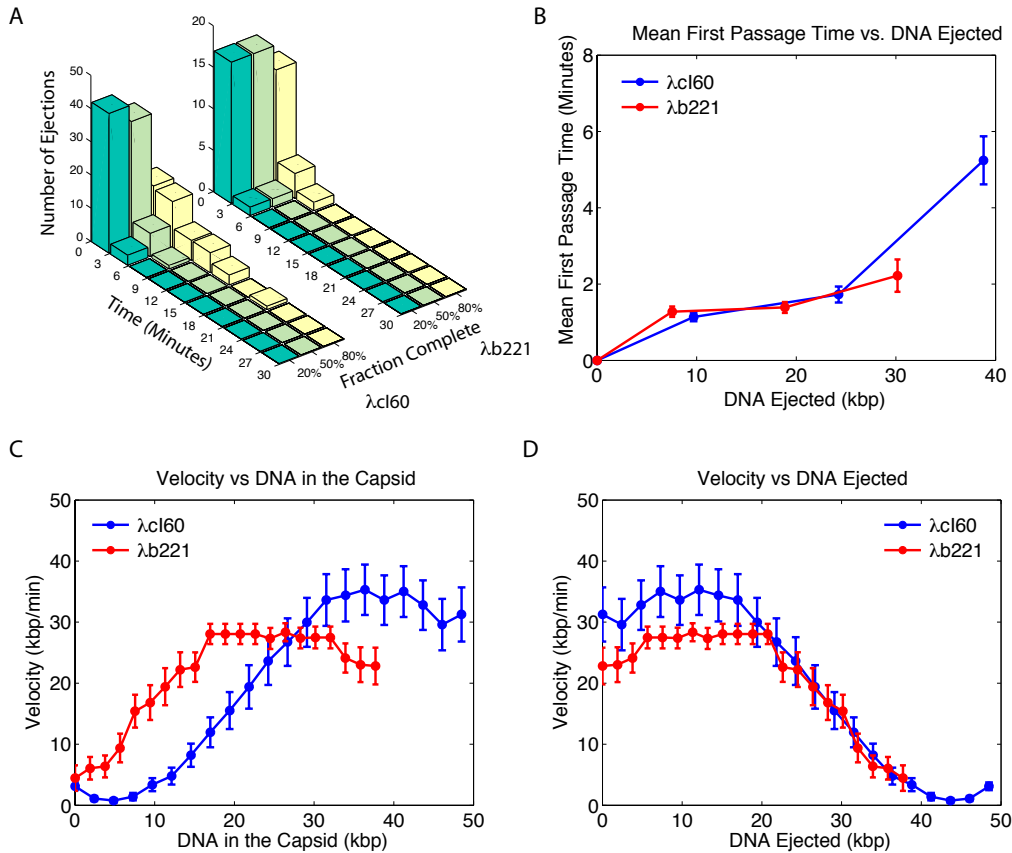
Figure 4.16: **An ensemble view of ejection times and dynamics for phage λ.** (a) Distributions of first-passage times for different fractions of completion of ejection. The histograms are determined by setting ejection thresholds of 20%, 50% and 80% complete as measured using the fluorescence. The distributions for both λcI60 (48.5 kbp) and λb221 (37.7 kbp) are shown. (b) Mean first-passage times for λcI60 and λb221. The means of the distributions shown in part (a) were used to calculate the mean first-passage time. There is little difference in the mean first-passage times of absolute amounts of DNA ejected between λcI60 and λb221. (c) Velocity of ejecting DNA plotted as a function of the amount of DNA remaining in the capsid. The initial portion of the ejection process is faster for λcI60 than λb221. There is no significant overlap between the two curves. (d) Velocity of ejecting DNA plotted as a function of the amount of DNA ejected. There is a small difference between the two curves for the first 20 kbp, and significant overlap after that. Only trajectories without pauses were used to generate (c) and (d).

4.2.2   Comparing with *in vitro* Ejection and Other Bacteriophages

In summary, we have examined the DNA ejection process for bacteriophage λ *in vivo* at the single-molecule level. We note that the techniques explored in this work may be generalizable to the study of other bacteriophages. It would be especially interesting to see a comparison between the bulk and single-molecule dynamics for bacteriophage T7 as bulk experiments have shown that the speed is constant throughout the ejection process *in vivo* [26, 181], as opposed to the variable rate reported here. We also note that the experimental platform presented here can be used to explore the effects of various genetic, chemical, and mechanical perturbations on the ejection process. Altogether, this work is an important step forward in understanding the mechanism behind genome delivery in bacteriophage λ and the more general problem of polymer translocation.

Our quantitative analysis of the ejection trajectories led to several new findings:

(1) Lambda ejection into cells takes several minutes, rather than the much shorter ≈ 10 second time scale previously reported for *in vitro* ejection [214].

(2) Using two Lambda mutants with different genome lengths (48.5 kbp and 37.8 kbp, respectively), it was shown that the *in vivo* ejection speed scales with the amount of DNA injected into the cell instead of the amount of DNA remaining in the capsid. By using these different length mutants, we can tune the mechanics of both the packaging and ejection process by controlling the amount of stored free energy of the DNA within the capsid. In the *in vitro* ejection case, the scaling with genome length remaining in the capsid indicates that the packaged DNA provides the ejection force [214].

(3) Also, the speed profile for phage Lambda injection is monotonically decreasing, which does not agree with an enzymatic, RNA polymerase catalyzed, constant speed profile previously reported for phage T7 [216].

(4) The *in vivo* ejection dynamics shows significant cell-to-cell variability and substantial pausing events, or even stalling events for the shorter genome mutant, which can only be resolved by single molecule assays like this. The single-virus-single-cell experiment not only eliminates ejection asynchrony and noise which can affect the data interpretation, but also reveals that the physical origins of the ejection dynamics *in vivo* are more subtle than their *in vitro* counterparts. In particular, besides active molecular motors, the cells also provide a viscoelastic environment through cytoplasmic forces and cellular proteins that could possibly explain the shapes and dependencies of the velocity curves.

The lack of data collapse with respect to DNA left in the capsid has significant implications for the role the energy stored in the compacted DNA plays during the *in vivo* ejection. If there were some portion of the ejection process that were governed solely by the energy in the compacted DNA, then during that portion we would expect the dynamics of λcI60 and λb221 to be identical when the amount of DNA remaining in the capsid is identical; this is the *in vitro* case as studied in [24]. Because the DNA-DNA repulsion inside the capsid is highest when the capsid contains more DNA, such a period would likely be at the beginning of the ejection process. As seen in Figure 4.16c, however, there exists no period of overlap between the velocity

curves for the two phage strains, and hence no period during the ejection process where the length of DNA in the capsid is the sole control parameter. From this observation, we can conclude that there is no period of time during which *in vivo* DNA ejection is governed solely by the electrostatic repulsion inside the capsid. Two-step models in which the the first half of the genome is delivered by the energy stored in the compacted DNA and the remainder is delivered by another mechanism are also not consistent with our data.

As mentioned earlier, another intriguing finding is the collapse of the velocity curves for λcI60 and λb221 (after 20 kbp) when plotted against the amount of DNA ejected into the cell. This raises the possibility that the amount of DNA ejected, as opposed to the amount of DNA in the capsid, is a key control parameter for this system. This picture would be consistent with models in which the mechanism is internal to the cell, as the only information such a mechanism would have available is the amount of DNA that has been brought inside the cell. The collapse of data on to a single curve is a powerful argument that has been used before to identify control parameters for *in vitro* DNA ejection as well as the lysis-lysogeny decision [23, 24, 176]. One limitation to applying this argument is that only two genome lengths have been tested here, and it is possible that this collapse does not hold for all viable genome lengths. Such reasoning also does not exclude a mixed picture, as mentioned above.

The origin of the apparent pauses might provide information about the ejection mechanism, as DNA-based motors acting against a load have been observed to pause [17, 222]. However, the pauses observed here are much longer than the pauses observed for motors and it is possible 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 [191, 192]. However, this is unlikely as pauses are not observed for phage lambda *in vitro* [24, 193]. We also cannot rule out the possibility that SYTOX Orange intermittently interferes with the ejection process.

We note that our results are contrary to what was shown in T7, in which a constant DNA ejection rate was seen with bulk measurements [26]. T7 has a capsid similar in size to lambda (60 nm vs. 58 nm, respectively) with a 40 kbp genome; however its tail is considerably shorter (23 nm vs. 150 nm, respectively) [181, 223]. It has been suggested that a constant velocity is suggestive of a purely enzyme driven model such as a molecular motor[26]. Such a feature is not seen in our data, as Figure 4.16d shows that once ~ 20 kbp of DNA has been ejected, there is a marked decrease in the ejection velocity. However, the anisotropic force-velocity relationship seen *in vitro* and the presence of pN level forces from the DNA-DNA repulsion inside the capsid make it unclear whether a constant ejection rate prediction would be true for lambda. Curiously, in references [224] and [225] it is noted that the transcription rate is between ~ 40 and 80 nt/sec which translates to roughly 10 minutes to pull DNA in, which is consistent with our observed time scale for λcI60.

The current data does not match previous calculations on mechanisms based on DNA binding proteins and thermal fluctuations [219]. Those calculations predict that after the first 10.8 kbp of DNA from λcI60 has been ejected, it should have the same dynamics as λb221 which, as discussed earlier, is not consistent with our

data. Also perplexing is the time scale of ejection. While there exists a discussion of polymer translocation in crowded environments in the literature that speaks on time scales, there are important differences between those analyses and our experimental system [226]. In particular, the origin of the friction exerted on the DNA that sets the time scale for ejection is poorly understood, both *in vitro* and now *in vivo*. A number of models assume a linear relationship between force and velocity, but it is now known that this assumption is not true *in vitro* [24, 219]. Whether this assumption is true *in vivo* is unclear, but we are skeptical of its applicability here. A better understanding of this friction would allow the current theories to speak quantitatively about the experimental observables in the single-molecule *in vitro* and *in vivo* DNA ejection experiments, and enable a richer interplay between theory and experiment.

In the late stages of viral infection, viral genomes and structural proteins are produced from different cellular machineries, then assemble into new viral particles. As opposed to self-assembly for many viruses, tailed bacteriophages, with double-stranded DNA genome, use energy-consuming molecular motors to package the genomes into the capsids and fill a large fraction (about 50%) of the volume [227–229]. Under physiological ionic conditions, the densely packed, negatively charged DNA is repulsive and could exert a pressure as high as 60 atmospheres to the capsid [227]. It is therefore hypothesized that the pressurized DNA could later facilitate the genome ejection process when infecting a new host.

This idea has been confirmed in the case of phage Lambda. Theoretical models based on the pressurized DNA mechanism [229, 230] correctly predicted the corresponding DNA configuration in the capsid [231], as well as the osmotic pressure in the surrounding solution required to balance out the driving force and halt the ejection [232, 233]. On a side note, other sources of driving force may also contribute significantly to the ejection process, depending on the designs of the virus. Some examples include active translocation by host motors found in phage T7 and ϕ29 [234, 235], Brownian ratchet resulting from cellular protein adsorption on the ejecting viral genome [236], and osmotic pressure gradient across the virus [237]. For simplicity, here we use bacteriophage Lambda, which has pressurized DNA as its dominating driving force, as a model system to study how genome translocation behaves under the physical constraints of the viral-host environment.

Two experimentally accessible tuning knobs that control the ejection force given by the pressurized DNA, are the ionic condition and the amount of DNA *remaining* in the capsid. During the course of ejection, the amount of DNA in the capsid decreases as a function of time. A previous *in vitro* experiment measuring the ejection dynamics of phage Lambda has shown, that the ejection speed scales with the instantaneous amount of DNA in the capsid, regardless of the total amount of genome the phage has to begin with [238]. This not only provides an additional evidence for the pressurized DNA idea, but also reveals notable aspects of the viscous resistance. Assuming linear drag, and estimating the drag constant from the measured ejection speed and the known ejection force as a function of genome length remaining in the capsid, it follows that the drag constant decreases almost exponentially with genome length in capsid, and implies that the part of DNA ejected into the solution contributes negligibly to the viscous resistance. A plausible explanation for the decaying drag constant uses the result of a unit-length polymer, moving through a thin tube and dragging

water along it, to calculate the shear stress. Since the number of unit-length DNA segments decreases, while the width of the tubes (approximately the separation between DNA segments in the capsid) increases, as more DNA is ejected, the total friction on the remaining DNA decreases during the course of ejection [239].

What provoked *in vivo* ejection measurements has to do with the fact that a Gram-negative bacterium, such as the *E. coli* infected by Lambda phage, has an osmotic pressure of $0.1 - 5$ atmospheres relative to the environment [240, 241]. The same magnitude of osmotic pressure was able to stop the ejection in the *in vitro* setting [232, 233]. This leads to a conundrum for *in vivo* ejection: How is DNA ejection almost always completed and leading to successful infection, despite the finite pressure in the cell?

In our previous work [242] on phage Lambda *in vivo* ejection, we found that the speed was about two orders of magnitude slower than that of *in vitro* ejection (typical timescales for *in vivo* and *in vitro* ejection are a few minutes and a few seconds, respectively). Moreover, the speed had a different characteristics from what measured *in vitro*. Roughly constant in the beginning of ejection, the speed then monotonically decreases and scales with the amount of DNA *ejected into the host-cell*. How does one reconcile the fact that DNA in the cell is the control parameter for *in vivo* ejection, with the prior understanding that the ejection force originated from the pressurized DNA mechanism scales with the amount of DNA in the phage capsid? How does one theorize the distinct and slower *in vivo* dynamics?

We argue that the dominant driving force for Lambda phage *in vivo* ejection remains the same, being the pressurized DNA in phage capsid and tuned by genome lengths and ionic conditions. For one thing, the most significant change from *in vitro* to *in vivo* ejection is the much reduced speed. We think that introducing an additional driving force is unnecessary, and would focus on how the viscous resistance elevates due to the presence of the cell. For another, there is no experimental evidence that Lambda phage ejection would be stopped by depleting the cell's metabolic activities. It is also known that transcription and translation from Lambda genome only start after the linear genome circularizes, therefore active cellular processes are unlikely to contribute to the DNA injection stage [243].

This leaves us with the cell's role as passively responding to the phage ejection with a resistive force. One clue might be that the bacterial cell interior is "crowded", with typical volume fraction of proteins and other macromolecules of ≈ 0.5. In these crowded, or 'glassy', cell interiors [244] anomalous diffusion of macromolecules has widely been observed, see e.g., [245, 246]. Anomalous diffusion is defined by $< r^2 > \propto Dt^\alpha$ with $0 < \alpha < 1$. In the case of bacterial chromosomal loci, $\alpha \approx 0.4$ [246], colloidal particles in dense worm-like micelles $\alpha \approx 0.3$ [247] and in case of RNA and several other macromolecules $\alpha \approx 0.7$ [245]. In Fig. 4.17 mean square displacements of several macromolecules in bacterial cytoplasm, including phage DNA, have been plotted. (These data have been extracted from references [244, 246] and [248].) On the relevant time windows these data consistently point to $\alpha \approx 0.5$. The injected phage DNA in this crowded cytoplasm may also diffuse anomalously, and give rise to a resistive force that scales with the amount of phage DNA in the cell, therefore the observed scaling in the ejection speed. Careful theoretical argument is still needed to associate anomalous diffusion with the phage ejection speed at late stage.
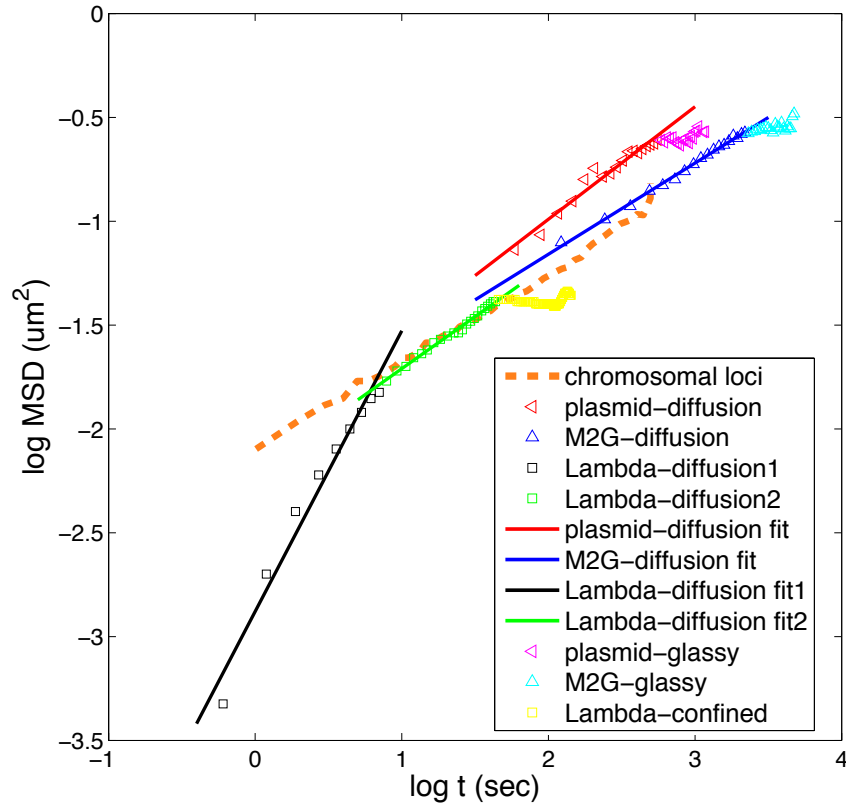
Figure 4.17: **Anomalous diffusion of macromolecules.** We replotted the diffusion behaviors measured in literature in this log-log plot of the mean squared displacement (μm^2) versus time (s) for several macromolecules in the bacterial cytoplasm. Except for the "glassy" and "confined" phases and "Lambda diffusion 1", the overall slope, which defines the coefficient for anomalous diffusion, is roughly $\alpha \approx 0.5$. The data representing "chromosomal loci" is from reference [246]. "Plasmid" (GFP-LacI-labeled mini-RK2 plasmid) and "M2G" (crescentin-GFP structure in metabolically active bacteria) are from reference [244]; the later stage with smaller slope is referred to as "glassy" in [244] and we only fit the first "diffusion" regime. The data "Lambda" is the bacteriophage lambda DNA monitored after injection and before genome integration, from reference [248], and we again identify regimes with different slopes as "diffusion 1", "diffusion 2", and "confined".

4.3 Towards Quantifying Post-Infection Dynamics

This project is a collaboration with Timur Zhiyentayev and David Wu.

Despite the well-understood genetics and biochemistry of bacteriophage Lambda, some open questions remain for the post-infection process, including the lytic-lysogenic decision making [249, 250], and the implication of burst size and timing to bacteriophage fitness [251]. One of the mysteries in the bacteriophage lytic pathway has to do with the chemical balance of the molecular species. It is often found that the subunits of molecular modules in the cell are synthesized to the exact ratio as in the assembled unit, which may play an important role in efficient assembly and may have been evolutionarily optimized, but the mechanisms are still being actively researched. The best-known example is the ribosome - there is stringent regulation of the amount of ribosomal RNA and ribosomal protein subunit expression in the cell [252–254]. To our knowledge, this question has not been addressed in the context of bacteriophage, which is known to have well-defined amounts of DNA and capsid protein subunits in the mature viral particles.

4.3.1 Measure Viral Lytic Production in Single Cells.
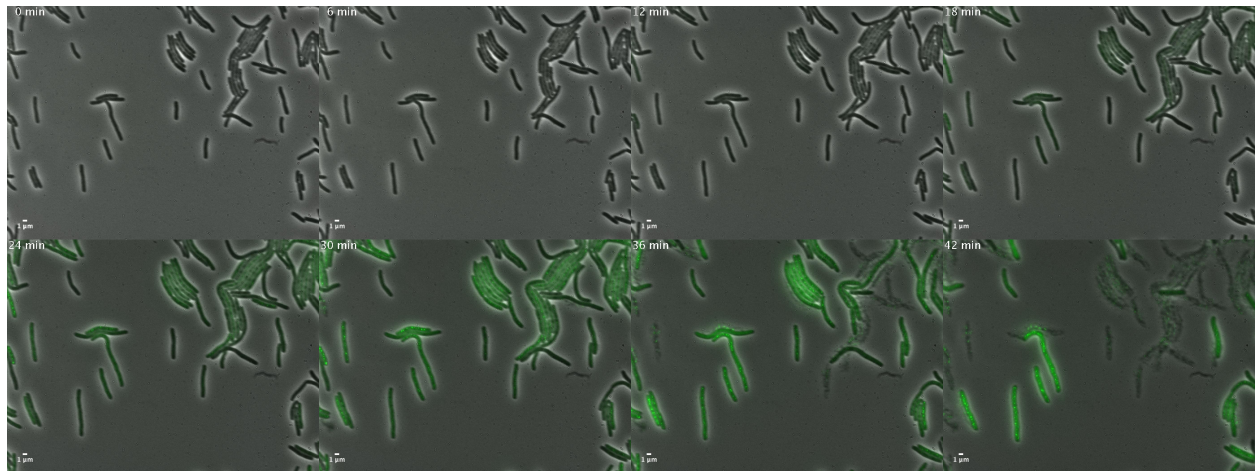
Figure 4.18: **The bacteriophage lambda lytic growth.** A time-lapse video microscopy of bacteriophage lysogen LZ2 growing on agar pad at 37 degrees Celsius. The LZ2 genome carries the capsid protein gpD-EYFP fusion and has a mutant lambda repressor, making the virus only lytic at this temperature. Phase-contrast images (gray-scale) and YFP channel images (green) are merged for each time points. Accompanying the production of viral machinery, the cellular metabolism is taken over and eventually leads to burst of cells. Scale bar: 1 micron. Time lapse: 6 minutes.

To investigate the stoichiometry problem, we turn to the bacteriophage lambda lysogen LZ2 [250]. It was constructed as a temperature-inducible lysogen (the viral genome stays integrated to the bacterial chromosome at 30 degrees Celsius and becomes lytic at 37 degrees Celcius), with the capsid protein gpD fused with EYFP. Figure 4.18 shows a typical time-lapse of cells expressing lambda LZ2. As more and more

viral products accumulate in the cells, the infection results in delayed cell growth and division, and finally leads to cell burst. Next, we will construct a library of fluorescence fusions to each capsid protein subunits to help us monitor the protein production in single cells as a function of time. The photo-bleaching technique will be used to calibrate the absolute copy number of the fluorescence proteins.

We also extend the FISH method to quantify the copy number of viral nucleic acids at each time point during the lytic growth. Our preliminary data for DNA FISH is shown in figure 4.19. We FISH the lambda gene A (the terminase) with oligos coupled to Alexa Fluor 594 dyes (pseudo-colored in green) and stain the total nucleic acid with DAPI (pseudo-colored in blue). Figure 4.19(A) is the lysogen cells, carrying one viral genome per bacterial chromosome, before heat induction. Figure 4.19(B) is cells after 40 minutes of heat induction. As predicted, we see that cells in (A) have one to two diffraction-limited puncta per cell, representing individual viral genomes. In (B) both the viral genome and total nucleic acid content are heavily ramped up after the lytic pathway is induced.

We think by extending these preliminary results, and combining both DNA and RNA FISH, it is possible to dissect the copy number profile during the infection process. The potential questions regarding the lytic production phase would be: How the growth curves of the viral products can be explained by using simple master equations describing the replication/transcription/translation of the well-characterized phage lambda genetic regulatory circuit? How wasteful the viral production is compared to the burst size (the average number of viral particles released per cell as measured in titer experiments)? Are the viral components (for example the capsid protein to genome ratio) are produced to the correct stoichiometry given by structural studies for mature viral particles? By quantifying the bacteriophage products as a function of time, we will be able to address fundamental issues regarding molecular balance in gene regulatory networks, and draw comparison to other actively-studied fields, such as the regulation of ribosomal RNA and protein expressions.
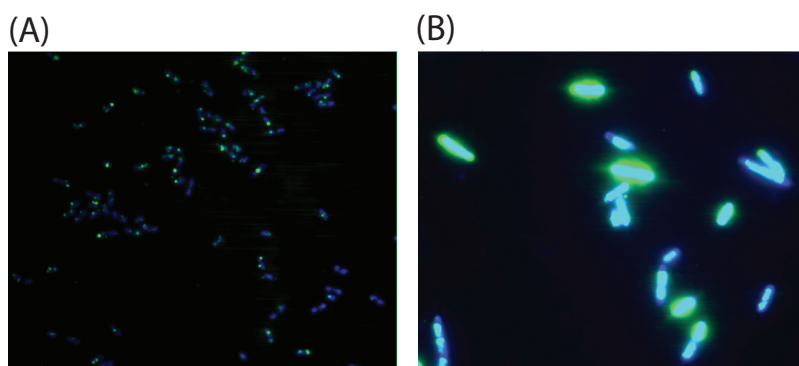
Figure 4.19: **Preliminary DNA FISH results for characterizing the viral lytic growth.** Green: Alexa Fluor 594, single-molecule DNA FISH of the lambda gene A. Blue: total nucleic acid with DAPI. (A), the LZ2 lysogen before induction. Each cell has 1-2 copies of chromosomes, hence 1-2 copies of viral genome. (B), the LZ2 lysogen after 40 minutes of induction. Both the viral DNA and total nucleic acid are heavily expressed compared to (A). (A) and (B) are at the same contrast. Note that scale bars do not apply to these fluorescence images.

4.3.2 Calibration of Protein Maturation Dynamics

Newly synthesized fusion protein needs finite time scale to mature (here we mean folding, reaction with oxygen, and all other relevant steps) properly to show fluorescence. The maturation is assumed to be a decay process with a time constant τ, which depends on the amino acid sequence and environmental conditions. The propagator $k(t - t')$ is thus

$$k(t - t') = 1 - e^{-(t-t')/\tau}, \tag{4.1}$$

which satisfies $k \to 1$ when $t \to \infty$.

For some protein ΔP synthesized at time t', the amount of protein that become fluorescent at time t is $k(t - t')\Delta P$.

Calibration when protein copy number is time-varying

In the case of viral capsid protein production in cells, the proteins are constantly being synthesized, and since we know both the capsid proteins and the fluorescent protein are very stable we can ignore their degradation. We denote the accumulated amount of protein up to time t as $P(t)$, and the rate of producing new protein as $r(t) = \frac{dP(t)}{dt}$. At $t = 0$, $P(0) = 0$.

The amount that are fluorescent at time t, $F(t)$, is given by

$$F(t) = \int_{t'=0}^{t'=t} k(t - t') \cdot dP(t') = \int_{t'=0}^{t'=t} k(t - t') \cdot r(t')dt' \tag{4.2}$$

$$= \int_{t'=0}^{t'=t} [1 - e^{-(t-t')/\tau}]r(t')dt' \tag{4.3}$$

$$= P(t) - \int_{t'=0}^{t'=t} e^{-(t-t')/\tau}r(t')dt', \tag{4.4}$$

because $\int_{t'=0}^{t'=t} r(t')dt' = P(t) - P(0) = P(t)$. Since the integration in the second term above only concerns t', we have

$$F(t) = P(t) - e^{-t/\tau} \cdot \int_{t'=0}^{t'=t} e^{t'/\tau}r(t')dt'. \tag{4.5}$$

We can use integration by parts, by identifying $u = e^{-(t-t')/\tau}$ and $dv = r(t')dt'$, so

$$F(t) = P(t) - e^{-t/\tau} \cdot [e^{t'/\tau}P(t')|_{t'=0}^{t'=t} - \int_{t'=0}^{t'=t} \frac{1}{\tau}e^{t'/\tau}P(t')dt'] \tag{4.6}$$

$$= e^{-t/\tau}\frac{1}{\tau} \int_{t'=0}^{t'=t} e^{t'/\tau}P(t')dt'. \tag{4.7}$$

Or equivalently, $\tau e^{t/\tau}F(t) = \int_{t'=0}^{t'=t} e^{t'/\tau}P(t')dt'$. We can differentiate both sides by dt, leading to

$$e^{t/\tau} P(t) = \frac{d}{dt} \int_{t'=0}^{t'=t} e^{t'/\tau} P(t') dt' = \tau \frac{d}{dt} [e^{t/\tau} F(t)] \tag{4.8}$$

$$= e^{t/\tau} F(t) + \tau e^{t/\tau} \frac{dF(t)}{dt} e^{t/\tau}. \tag{4.9}$$

Hence,

$$P(t) = F(t) + \tau \frac{dF(t)}{dt}. \tag{4.10}$$

Given that we know τ, we can use the measured 'apparent' amount of protein from fluorescence, $F(t)$, to obtain the actual copy number $P(t)$.

Measure the maturation decay time τ

(i) If it is possible to deposit some protein ΔP at time t', $P(t) = \Delta P \cdot \Theta(t - t')$, where $\Theta(t - t')$ is the step function, then the fluorescence measurement should give $F(t) = [1 - e^{-(t-t')/\tau}]\Delta P$ and we can fit $F(t)$ to get τ. This might be hard to implement experimentally.

(ii) In certain regulatory settings, there could be a regime where the protein production rate is constant. For example, we can assume the viral structural phage protein-fluorescent protein fusions do not degrade, and if we have a fixed number of mRNA, with constant translation rate we then have $r(t) = \alpha$, $P(t) = \alpha \cdot t$.

From equation (4.3), $F(t) = \int_{t'=0}^{t'=t} [1 - e^{-(t-t')/\tau}] \cdot \alpha \cdot dt' = \alpha(t - \tau + \tau e^{-t/\tau})$. This is also a simple functional form to fit τ.

(iii) If the protein production is linear with rate α from time 0 to T, and the production is stopped at T, then

$$\begin{aligned} r(t) &= \alpha, & 0 \leq t < T, \\ r(t) &= 0, & t \geq T, \\ P(t) &= \alpha t, & 0 \leq t < T, \\ P(t) &= \alpha T, & t \geq T. \end{aligned} \tag{4.11}$$

From equation (4.3),

$$F(t) = \int_{t'=0}^{t'=t} [1 - e^{-(t-t')/\tau}] \cdot \alpha \cdot dt',$$

$$= \alpha(t - \tau + \tau e^{-t/\tau}), \qquad\qquad\qquad\qquad 0 \le t < T,$$

$$F(t) = \int_{t'=0}^{t'=T} [1 - e^{-(t-t')/\tau}] \cdot \alpha \cdot dt' + \int_{t'=T}^{t'=t} [1 - e^{-(t-t')/\tau}] \cdot 0 \cdot dt',$$

$$= \alpha[T - \tau e^{-(t-T)/\tau} + \tau e^{-t/\tau}], \qquad\qquad\qquad t \ge T.$$

$$\text{(4.12)}$$

When $t \ge T$, $F(t) = \alpha[T - \tau e^{-(t-T)/\tau} + \tau e^{-t/\tau}] = \alpha T - \alpha \tau (e^{T/\tau} - 1)e^{-t/\tau} \equiv \alpha T - B(T, \tau)e^{-t/\tau}$ is probably a simpler functional form to fit τ than that in (ii).

(iv) Actually, the functional form of $r(t)$ doesn't matter that much. As long as we have

$$r(t) \ne 0, \quad 0 \le t < T,$$

$$r(t) = 0, \qquad t \ge T,$$

$$\text{(4.13)}$$

meaning that production is turned off at time T, we will end up with a simple $F(t)$ for $t \ge T$.

$$F(t) = \int_{t'=0}^{t'=T} [1 - e^{-(t-t')/\tau}] \cdot r(t') \cdot dt' + \int_{t'=T}^{t'=t} [1 - e^{-(t-t')/\tau}] \cdot 0 \cdot dt',$$

$$= P(T) - e^{-t/\tau} \int_{t'=0}^{t'=T} e^{t'/\tau} \cdot r(t') \cdot dt',$$

$$\equiv P(T) - B(T, \tau)e^{-t/\tau}, \qquad t \ge T.$$

$$\text{(4.14)}$$

The only t dependence is an exponential decay, and all other terms are constants. This allows us to fit τ easily.

Using kinetic model

The rate of producing non-fluorescent protein N from gene G is $r(t)$, and the non-fluorescent protein decay to fluorescent protein F with rate $\frac{1}{\tau}$ as the following,

$$G \xrightarrow{r(t)} N \xrightarrow{\frac{1}{\tau}} F. \qquad\qquad\qquad\qquad \text{(4.15)}$$

The differential equations are

$$\frac{dN}{dt} = r(t) - \frac{1}{\tau}N,$$

$$\frac{dF}{dt} = \frac{1}{\tau}N.$$

$$\text{(4.16)}$$

The total amount of protein at time t has contributions from the non-fluorescent and the fluorescent

ones, $P(t) = N(t) + F(t) = \tau\frac{dF}{dt} + F(t)$, which is equivalent to equation (4.10).

When $r(t)$ is constant, these differential equations can be easily solved and get the same forms as (ii) above.

The "'decay term" $\frac{1}{\tau}N$, depending on the current amount of N, is similar to the treatment of protein degradation (plus dilution) half-life in the bleach-chase experiments. However, I think there is a caveat in that this treatment means the decay process is memory-less. The molecules produced earlier would have the same decay probability as the molecules produced at a later time, so we could use $\frac{1}{\tau}N$. The fact that the system appears to be memory-less might be a property of the Poisson process being the source of the exponential decay. I think this memory-less property wouldn't hold in general for any propagator $k(t - t')$ in equation (4.2).

Chapter 5

Conclusions

Genomics, transcriptomics, and proteomics have recently been topics of much interest, with the overall goal of improving the predictive power between genotypes and phenotypes. The majority of works have focused on the information the genome encodes. Despite the fact that the field of genetics has existed for several decades, it remains challenging to accurately and precisely predict phenotypes from genotypes, particularly when dealing with relatively complex traits or describing the full developmental processes which require monitoring large networks of genes and proteins. More recently, advances have enabled high-throughput experimental techniques for DNA sequencing and gene expression profiles, as well as development of more rigorous computational algorithms in bioinformatics. However, these techniques and algorithms establish mainly correlations; they cannot probe deeper into the causal relationships and underlying mechanisms.

To quantitatively predict the statistical properties and dynamics of genetic circuits in response to changes in the environment, and to investigate the underlying mechanisms, we studied in this thesis the physical aspects of the DNA macromolecules and employed tools from statistical mechanics to describe the gene expression outcomes. We began by examining the mechanical properties of the DNA polymers, with the objective of understanding sequence-dependent deformability and its role in controlling nucleosome accessibility and corresponding gene expression in eukaryotic cells that had been observed [12–14]. We studied the different properties that those same sequences exhibit when taken in the context of a bacterial protein-mediated DNA looping motif. From this, we discovered that the biological contexts played a significant role in determining which combinations of the deformability degrees of freedom would show up. We attempted to assess whether simple parameters, including polyA or GC content of the DNA sequence, could be reliable predictors of the mechanical properties of those DNA polymers. This continues to be an active field of research that is receiving much attention from other researchers, applying high-throughput sequencing methods to collect more data on sequence-dependent mechanics of the genome.

In the same bacterial DNA looping motif mentioned previously, we also conducted an examination of its looping and unlooping dynamics and found a scaling relationship between the lifetimes and the loop formation free energy. Microscopically, the looping and unlooping are stochastic events, but their average durations must be consistent with the thermodynamic model previously devised [76, 255]. The finding that

the unlooping dynamics also depends on loop formation free energy was unexpected and went contrary to existing literature. Dissociation processes are generally considered as escaping the short-ranged binding potential, but we have presented evidence that the dissociation process is also influenced by the deformability of DNA. The mechanism in this case was the interaction distance of the binding potential is compatible with the characteristic scale of DNA bending, in the case of short loops where polymer relaxation can be ignored.

In the third part of this thesis, we explored the viral infection dynamics in a bacterial system. While most eukaryotic viruses are able to gain access to eukaryotic cells through membrane fusion, bacterial viruses attempting to penetrate bacteria must contend with the added complexity of a cell wall before gaining entry to the host. Consequently, viruses infecting a eukaryotic cell are able to carry their full structure into their host, while bacteriophages must instead inject their DNA into the host while the capsid remains outside. However, the bacteriophage DNA ejection process may still provide insights to how eukaryotic viral genomes enter the host nucleus. The consensus in existing literature is that bacteriophages use lytic domains on their protein shell to locally digest through the hosts cell wall. One issue in need of additional clarity, however, concerns how the cell wall, membranes, the periplasmic space, and the cytoplasm influence the phage DNA motion as it enters the cell. Our approach to answering this question was to directly measure the dynamics of DNA during ejection and compare these measurements with different models for the ejection driving forces and the resistive forces that the DNA may experience during the process. Previous *in vitro* experiments specific to phage Lambda have provided evidence that the main contributor to the injection driving force is the pressure resulting from confinement of DNA in the phage capsid [21–24]. After conducting an *in vivo* version of this phage Lambda injection experiment, we were able to quantitatively evaluate how the bacterial cellular environment impacts the dynamics of this process. The ejection was observed to slow by two orders of magnitude, and our results further indicated that the ejection velocity was dependent on the amount of phage DNA in the cell. We think the speed profiles we observed in phage Lambda ejection were a consequence of the viscoelasic, crowded bacterial cytoplasm, rather than an active mechanism similar to what has been observed in phage T7 (which would result in a constant ejection speed). We also developed a method based on *in situ* fluorescence hybridization to distinguish different possible ejection driving forces which should work in a broader range of phage species, and apply it to examined the post-infection phase of bacterial phage Lambda, where we were primarily concerned with the growth of various phage products over time.

Having reviewed both the previous work and the current achievements presented in this thesis, we now conclude by turning our attention to potential future directions and extensions of this work.

First, for the Lac repressor-mediated DNA looping system, the wild-type Lac operon actually consists of three protein binding sites, which resulted in three possible looping configurations. However, in our current experimental setup, we did not have sufficient resolution to distinguish these three distinct loop structures. Moreover, extending the model we have done here, using polymer physics to predict the dynamics of loop formation and breakdown, is made more challenging by the inherent complexities of the wild-type operon.

Second, even for a simplified case with only two binding sites, the corresponding *in vivo* looping probability measurements [256, 257] were not fully consistent with the *in vitro* results, which implies that the crowded cellular environment or the presence of nonspecific proteins may play a non-trivial role. Another potential future direction for this work would be to develop an *in vivo* experiment to measure looping dynamics in live cells. For much longer loops, such as the 2.3 kbp loop that occurs in the decision-making region on the phage Lambda genome, labeling techniques based on the fluorescent repressor-operator systems were used in visualizing the phage Lambda CI-mediated DNA looping in live *E. coli* cells, and a lower bound of 0.5 s state lifetimes was estimated [258]. However, it is unclear whether these labels (each contains 3 repairs of operator sites, spanning about 100 bp; together give 66 bp resolution reported in [258]) would be too big for the Lac repressor-mediated loop in the hundreds bp range, which makes experimental setup for further investigation into its *in vivo* dynamics more challenging.

Third, although the genetics of bacteriophages and their biochemical products are generally well understood and thoroughly documented, relatively little is known specifically about their diverse ejection processes. In the case of phage Lambda, it has been observed that both the *E. coli* outer membrane receptor LamB and inner membrane protein PtsM must be present for successful *in vivo* ejection. However, the reason for this requirement and the specific process by which this ejection occurs are not yet understood, and would warrant further study. The lack of understanding for the genome ejection dynamics process extends to other types of phages outside of the tailed dsDNA phage category, such as phage M13 that infects bacteria though the F pilus structure.

Finally, we acknowledge that much of our work discussed in this thesis, whether in the gene regulation motif or viral infection, focused on the dynamics, as this provides additional insight to an organisms fitness than the phenotypic traits alone. It is the temporal responses compared to an organism's lifespan and to the given rate of environmental changes that determines an organisms performance and survivability. Consequently, some trade-offs are expected. For instance, among bacteriophage species, experimental data has suggested a negative correlation between replication rate and several other properties, including host range, environmental structural stability, and lysogenic ability/frequency [259–261]. However, a more complete theory for these trade-offs has not been developed and would be another possible extension of the work discussed here. In addition, because viral infection, or horizontal gene transfer in general, plays a crucial role in shaping the characteristics of bacterial populations, there are also opportunities for integrating these effects into the current theoretical framework for evolutionary dynamics, which are currently focused largely on random mutations being the major driving force.

Bibliography

[1] P. Gross, N. Laurens, LB. Oddershede, U. Bockelmann, EJG. Peterman, and GJL. Wuite. Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.*, 7(9):731–736, 2011.

[2] R. Vafabakhsh and T. Ha. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science*, 337(6098):1097–1101, 2012.

[3] I Golding and EC Cox. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.*, 96(9):098102, 2006.

[4] GB. West, JH. Brown, and BJ. Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122–126, 1997.

[5] BJ. Enquist, JH. Brown, and GB. West. Allometric scaling of plant energetics and population density. *Nature*, 395(6698):163–165, 1998.

[6] BJ. Enquist, GB. West, EL. Charnov, and JH. Brown. Allometric scaling of production and life-history variation in vascular plants. *Nature*, 401(6756):907–911, 1999.

[7] Geoffrey B West and James H Brown. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. *Journal of Experimental Biology*, 208(9):1575–1592, 2005.

[8] Christopher P Kempes, Stephanie Dutkiewicz, and Michael J Follows. Growth, metabolic partitioning, and the size of microorganisms. *Proceedings of the National Academy of Sciences*, 109(2):495–500, 2012.

[9] IM. Tolic-Norrelykke, EL. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sorensen. Anomalous diffusion in living yeast cells. *Phys. Rev. Lett.*, 93(7):078102, 2004.

[10] Stephanie C Weber, Andrew J Spakowitz, and Julie A Theriot. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Physical review letters*, 104(23):238102, 2010.

[11] Stephanie C Weber, Julie A Theriot, and Andrew J Spakowitz. Subdiffusive motion of a polymer composed of subdiffusive monomers. *Physical Review E*, 82(1):011913, 2010.

[12] J. D. Anderson and J. Widom. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol Cell Biol*, 21(11):3830–3839, 2001.

[13] E. Segal, Y Fondufe-Mittendorf, L Chen, A Thastrom, Y Field, I Moore, J-P Wang, and J. Widom. A genomic code for nucleosome positioning. *Nature*, 442:772–778, 2006.

[14] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236):362–366, 2009.

[15] L. Finzi and J. Gelles. Measurement of Lactose repressor-mediated loop formation and breakdown in single DNA molecules. *Science*, 267:378–380, 1995.

[16] F. Vanzi, C. Broggio, L. Sacconi, and F. S. Pavone. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic Acids Res*, 34(12):3409–3420, 2006.

[17] 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.

[18] 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.

[19] 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.

[20] J. E. Johnson and W. Chiu. DNA packaging and delivery machines in tailed bacteriophages. *Curr Opin Struct Biol*, 17:237–243, 2007.

[21] P. K. Purohit, J. Kondev, and R. Phillips. Mechanics of DNA packaging in viruses. *Proc Natl Acad Sci USA*, 100:3173–3178, 2003.

[22] 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.

[23] 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.

[24] 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.

[25] L. Garcia and I. Molineux. Rate of translocation of bacteriophage T7 DNA across the membranes of *escherichia coli*. *J Bacteriol*, 14:4066–4076, 1995.

[26] P. Kemp, M. Gupta, and I.J. Molineux. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Mol M*, 53:1251–1265, 2004.

[27] Frank B Dazzo, Kevin J Klemmer, Ryan Chandler, and Youssef G Yanni. In situ ecophysiology of microbial biofilm communities analyzed by cmeias computer-assisted microscopy at single-cell resolution. *Diversity*, 5(3):426–460, 2013.

[28] Rob Phillips, Jane Kondev, Julie Theriot, and Hernan Garcia. *Physical biology of the cell*. Garland Science, 2012.

[29] E. Segal and J. Widom. What controls nucleosome positions? *Trends Genet*, 25:335–343, 2009.

[30] M. Radman-Livaja and O. J. Rando. Nucleosome positioning: How is it established, and why does it matter? *Dev Biol*, 339(2):258–266, 2010.

[31] V. Iyer. Nucleosome positioning: bringing order to the eukaryotic genome. *Trends Cell Biol*, 22:250–256, 2012.

[32] H. Nishida. Genome DNA sequence variation, evolution and function in bacteria and archea. *Curr Issues Mol Biol*, 15:19–24, 2012.

[33] H. R. Widlund, P. N. Kuduvalli, M. Bengtsson, H. Cao, T. D. Tullius, and M. Kubista. Nucleosome structural features and intrinsic properties of the TATAAACGCC repeat sequence. *J Biol Chem*, 274(45):31847–31852, 1999.

[34] M. Roychoudhury, A. Sitlani, J. Lapham, and D. M. Crothers. Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proc Natl Acad Sci USA*, 97(25):13608–13613, 2000.

[35] J. Widom. Role of DNA sequence in nucleosome stability and dynamics. *Quart Rev Biophys*, 34:1–56, 2001.

[36] T. E. Cloutier and J. Widom. Spontaneous sharp bending of double-stranded DNA. *Molec Cell*, 14(3):355–362, 2004.

[37] M. Hogan, J. LeGrange, and B. Austin. Dependence of DNA helix flexibility on base composition. *Nature*, 304(5928):752–754, 1983.

[38] J. P. Peters and L. J.. Maher 3rd. DNA curvature and flexibility *in vitro* and *in vivo*. *Quart Rev Biophys*, 43(1):22–63, 2010.

[39] W. K. Olson and V. B. Zhurkin. Working the kinks out of nucleosomal DNA. *Curr Opin Struct Biol*, 21(3):348–357, 2011.

[40] W. K. Olson, D. Swigon, and B. D. Coleman. Implications of the dependence of the elastic properties of DNA on nucleotide sequence. *Phil Trans R Soc Lond A*, 362:1403–1422, 2004.

[41] A. V. Morozov, K. Fortney, D. A. Gaykalov, V. M. Studitsky, J. Widom, and E. D. Siggia. Using DNA mechanics to predict *in vitro* nucleosome positions and formation energies. *Nuc Acids Res*, 37(14):4704–4722, 2009.

[42] S. Balasubramanian, F. Xu, and W. K. Olson. DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. *Biophys J*, 96(6):2245–2260, 2009.

[43] S. Geggier and A. Vologodskii. Sequence dependence of DNA bending rigidity. *Proc Natl Acad Sci USA*, 107(35):15421–15426, 2010.

[44] H.E. Peckham, R.E. Thurman, Y. Fu, J.A. Stamatoyannopoulos, W.S. Noble, K. Struhl, and Z. Weng. Nucleosome positioning signals in genomic DNA. *Genome Res*, 17:1170–1177, 2007.

[45] D. Tillo and T. R. Hughes. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics*, 10:442, 2009.

[46] S. Schwartz, E. Meshorer, and G. Ast. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol*, 16:990–995, 2009.

[47] D. Tillo, N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, Y. Field, J. D. Lieb, J. Widom, E. Segal, and T. R. Hughes. High nucleosome occupancy is encoded at human regulatory sequences. *PLOS ONE*, 5:e9129, 2010.

[48] V. Iyer and K. Struhl. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J*, 14(11):2570–2579, 1995.

[49] E. Segal and J. Widom. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol*, 19(1):65–71, 2009.

[50] Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol*, 4(11):e1000216, 2008.

[51] G.-C. Yuan, Y.-J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, 309:627–630, 2005.

[52] D. Rhodes. Nucleosome cores reconstituted from poly(dA-dT) and the octamer of histones. *Nucleic Acids Res*, 6(5):1805–1816, 1979.

[53] G.R. Kunkel and H.G. Martinson. Nucleosomes will not form on double-stranded RNA or over poly(dA)·poly(dT) tracts in recombinant DNA. *Nucleic Acids Res*, 9:6869–6888, 1981.

[54] A. Prunell. Nucleosome reconstitution on plasmid-inserted poly(dA)·poly(dT). *EMBO J*, 1:173–179, 1982.

[55] T. E. Shrader and D. M. Crothers. Effects of DNA sequence and histone-histone interactions on nucleosome placement. *J Mol Biol*, 216(1):69–84, 1990.

[56] K. Struhl. Naturally occurring poly(dA:dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc Natl Acad Sci USA*, 82(24):8419–8423, 1985.

[57] T. Raveh-Sadka, U. Levo, M.and Shabi, B. Shany, L. Keren, M. Lotan-Pompan, D. Zeevi, E. Sharon, A. Weinberger, and E. Segal. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat Genet*, 44:743–750, 2012.

[58] T. E. Haran and U. Mohanty. The unique structure of A-tracts and intrinsic DNA bending. *Quart Rev Biophys*, 42:41–81, 2009.

[59] H. C. M. Nelson, J. T. Finch, B. F. Luisi, and A. Klug. The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature*, 330(6145):221–226, 1987.

[60] B. Sutter, G. Schnappauf, and F. Thoma. Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters *in vivo*. *Nucleic Acids Res*, 28:4083–4089, 2000.

[61] S. Johnson, M. Lindén, and R. Phillips. Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res*, 40(16):7728–7738, 2012.

[62] D. M. Crothers, T.E. Haran, and J.G. Nadeau. Intrinsically bent DNA. *J Biol Chem*, 265:7093–7096, 1990.

[63] R. A. Mehta and J. D. Kahn. Designed hyperstable Lac repressor·DNA loop topologies suggest alternative loop geometries. *J Mol Biol*, 294:67–77, 1999.

[64] L. M. Edelman, R. Cheong, and J. D. Kahn. Fluorescence resonance energy transfer over 130 basepairs in hyperstable Lac repressor-DNA loops. *Biophys J*, 84(2):1131–1145, 2003.

[65] M. A. Morgan, K. Okamoto, J.D. Kahn, and D. S. English. Single-molecule spectroscopic determination of *lac* repressor-DNA loop conformation. *Biophys J*, 89(4):2588–2596, 2005.

[66] A. R. Haeusler, K. R. Goodson, T. D. Lillian, X. Wang, S. Goyal, N. C. Perkins, and J. D. Kahn. FRET studies of a landscape of Lac repressor-mediated DNA loops. *Nucleic Acids Res*, 2012.

[67] A. K. Cheema, N. R. Choudhury, and H. K. Das. A- and T-tract-mediated intrinsic curvature in native DNA between the binding site of upstream activator NtrC and the *nifLA* promoter of *Klebsiella pneumoniae* facilitates transcription. *J Bacteriol*, 181(17):5296–5302, 1999.

[68] A. Schulz, J. Langowski, and K. Rippe. The effect of the DNA conformation on the rate of NtrC activated transcription of *Escherichia coli* RNA Polymerase·σ^{54} holoenzyme. *J Mol Biol*, 300:709–725, 2004.

[69] M. Serrano, I. Barthelemy, and M. Salas. Transcription activation at a distance by phage Φ 29 protein p4. Effect of bent and non-bent intervening sequences. *J Mol Biol*, 219(3):403–414, 1991.

[70] A. E. Lilja, J. R. Jenssen, and J. D Kahn. Geometric and dynamic requirements for DNA looping, wrapping and unwrapping in the activation of *E. coli glnAp2* transcription by NtrC. *J Mol Biol*, 342:467–478, 2004.

[71] R. T. Simpson and D. W. Stafford. Structural features of a phased nucleosome core particle. *Proc Natl Acad Sci USA*, 80:51–55, 1983.

[72] A. Thastrom, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J Mol Biol*, 288:213–229, 1999.

[73] T. E. Cloutier and J. Widom. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc Natl Acad Sci USA*, 102(10):3645–3650, 2005.

[74] P. T. Lowary and J. Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, 276:19–42, 1998.

[75] ST Bzymek, M; Lovett. Evidence for Two Mechanisms of Palindrome-Stimulated Deletion in Escherichia coli: Single-Strand Annealing and Replication Slipped Mispairing. *GENETICS*, 158(2):527–540, 2001.

[76] 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(5):e5621, 2009.

[77] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature*, 352:444–448, 1991.

[78] H. Yin, R. Landick, and J. Gelles. Tethered particle motion method for studying transcript elongation by a single RNA polymerase molecule. *Biophys J*, 67(6):2468–2478, 1994.

[79] P. C. Nelson, C. Zurla, D. Brogioli, J. F. Beausang, L. Finzi, and D. Dunlap. Tethered particle motion as a diagnostic of DNA tether length. *J Phys Chem B*, 110(34):17260–17267, 2006.

[80] D. Shore, J. Langowski, and R. L. Baldwin. DNA flexibility studied by covalent closure of short fragments into circles. *Proc Natl Acad Sci USA*, 78(8):4833–4837, 1981.

[81] D. M. Crothers, J. Drak, J.D. Kahn, and S.D. Levene. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol*, 212:3–29, 1992.

[82] Y. Zhang and D. M. Crothers. High-throughput approach for detection of DNA bending and flexibility based on cyclization. *Proc Natl Acad Sci USA*, 100(6):31613166, 2003.

[83] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii. Cyclization of short DNAfragments and bending fluctuations of the double helix. *Proc Natl Acad Sci USA*, 102(15):5397–5402, 2005.

[84] R.A. Forties, R. Bundschuh, and M.G. Poirier. The flexibility of locally melted DNA. *Nucleic Acids Res*, 37:4580–4586, 2009.

[85] S. Geggier, A. Kotlyar, and A. Vologodskii. Temperature dependence of DNA persistence length. *Nucleic Acids Research*, 39:1419–1426, 2011.

[86] S. V. Kuznetsov, S. Sugimura, P. Vivas, D. M. Crothers, and A. Ansari. Direct observation of DNA bending/unbending kinetics in complex with DNAbending protein IHF. *Proc Natl Acad Sci USA*, 103(49):18515–18520, 2006.

[87] H. Jacobson and W. H. Stockmayer. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J Chem Phys*, 18(12):1600–1606, 1950.

[88] J. Shimada and H. Yamakawa. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules*, 17:689–698, 1984.

[89] David Shore and Robert L Baldwin. Energetics of dna twisting: I. relation between twist and cyclization probability. *Journal of molecular biology*, 170(4):957–981, 1983.

[90] L.J. Peck and J.C. Wang. Sequence dependence of the helical repeat of DNA in solution. *Nature*, 292(5821):375–378, 1981.

[91] F. Strauss, C. Gaillard, and A. Prunell. Helical periodicity of DNA, poly(dA)·poly(dT) and poly(dA-dT)·poly(dA-dT) in solution. *Eur J Biochem*, 118(2):215–222, 1981.

[92] O. K. Wong, M. Guthold, D. A. Erie, and J. Gelles. Interconvertible Lac repressor-DNA loops revealed by single-molecule experiments. *PLoS Biol*, 6(9):e232, 2008.

[93] D. Normanno, F. Vanzi, and F. S. Pavone. Single-molecule manipulation reveals supercoiling-dependent modulation of *lac* repressor-mediated DNA looping. *Nucleic Acids Res*, 36(8):2505–2513, 2008.

[94] D. Rutkauskas, H. Zhan, K. S. Matthews, F. S. Pavone, and F. Vanzi. Tetramer opening in LacI-mediated DNA looping. *Proc Natl Acad Sci USA*, 106(39):16627–16632, 2009.

[95] K. B. Towles, J. F. Beausang, H. G. Garcia, R. Phillips, and P. C. Nelson. First-principles calculation of DNA looping in tethered particle experiments. *Phys Biol*, 6(2):025001, 2009.

[96] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, 271(5253):1247–1254, 1996.

[97] E. Villa, A. Balaeff, and K. Schulten. Structural dynamics of the lac repressor–DNA complex revealed by a multiscale simulation. *Proc Natl Acad Sci*, 102(19):6783–6788, May 2005.

[98] D. Swigon, B. D. Coleman, and W. K. Olson. Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. *Proc Natl Acad Sci USA*, 103(26):9879–9884, 2006.

[99] Y. Zhang, A. E. McEwen, D. M. Crothers, S. D. Levene, and P. Fraser. Analysis of *in vivo* LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS One*, 1:e136, 2006.

[100] C. Manzo, C. Zurla, D.D. Dunlap, and L. Finzi. The effect of nonspecific binding of lambda repressor on DNA looping dynamics. *Biophys J*, 103:1753–1761, 2012.

[101] R. Rohs, S.M. West, A. Sosinsky, P. Liu, R.S. Mann, and B. Honig. The role of DNA shape in protein-DNA recognition. *Nature*, 461:1248–1253, 2009.

[102] H.G. Garcia, P.G. Grayson, L. Han, M. Inamdar, J. Kondev, P.C. Nelson, R. Phillips, J. Widom, and P.A. Wiggins. Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers*, 85:115–130, 2007.

[103] J. Griffith, A. Hochschild, and M. Ptashne. DNA Loops induced by cooperative binding of lambda-repressor. *Nature*, 322(6081):750–752, 1986.

[104] RW. Zeller, JD. Griffith, JG. Moore, CV. Kirchhamer, RJ. Britten, and EH. Davidson. A multimerizing transcription factor of sea urchin embryos capable of looping DNA. *Proc. Natl. Acad. Sci. USA*, 92(7):2989–2993, 1995.

[105] G.-W. Li, OG. Berg, and J. Elf. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat Phys*, 5(4):294–297, 2009.

[106] N. Laurens, D. A. Rusling, C. Pernstich, I. Brouwer, S. E. Halford, and G. J. L. Wuite. DNA looping by FokI: the impact of twisting and bending rigidity on protein-induced looping dynamics. *Nucleic Acids Res*, 40(11):4988–4997, 2012.

[107] S. Johnson, Y.-J. Chen, and R. Phillips. Poly(dA:dT)-rich DNAs are highly flexible in the context of DNA looping. *PLoS ONE*, 8(10):e75799, 2013.

[108] S. Jun, J. Bechhoefer, and B. Y. Ha. Diffusion-limited loop formation of semiflexible polymers: Kramers theory and the intertwined time scales of chain relaxation and closing. *Europhys Lett*, 64(3):420426, 2003.

[109] C. Hyeon and D. Thirumalai. Kinetics of interior loop formation in semiflexible chains. *J Chem Phys*, 124:104905, 2006.

[110] NM. Toan, G. Morrison, C. Hyeon, and D. Thirumalai. Kinetics of loop formation in polymer chains. *J Phys Chem B*, 112(19):6094–6106, 2008.

[111] D. Van Valen, M. Haataja, and R. Phillips. Biochemistry on a Leash: The Roles of Tether Length and Geometry in Signal Integration Proteins. *Biophys J*, 96:12751292, 2009.

[112] D. Reeves, K. Cheveralls, and J. Kondev. Regulation of biochemical reaction rates by flexible tethers. *Phys Rev E*, 84(2):021914, 2011.

[113] G. Bell. Models for the Specific Adhesion of Cells to Cells. *Science*, 200:618627, 1978.

[114] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–53, 1999.

[115] OK. Dudko, G. Hummer, and A. Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. USA*, 105(41):15755–15760, 2008.

[116] J. Liphardt, B. Onoa, SB. Smith, I. JR Tinoco, and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, 2001.

[117] S. Mihardja, A. Spakowitz, Y. Zhang, and C. Bustamante. Effect of force on mononucleosomal dynamics. *Proc. Natl. Acad. Sci. USA*, 103(43):15871–15876, 2006.

[118] Collin Joseph, Chiao-Yu Tseng, Giovanni Zocchi, and Tsvi Tlusty. Asymmetric effect of mechanical stress on the forward and reverse reaction catalyzed by an enzyme. *PloS one*, 9(7):e101442, 2014.

[119] Tung T Le and Harold D Kim. Probing the elastic limit of dna bending. *Nucleic Acids Res*, page gku735, 2014.

[120] YF. Chen, JN. Milstein, and JC. Meiners. Femtonewton Entropic Forces Can Control the Formation of Protein-Mediated DNA Loops. *Phys Rev Lett*, 104(4):048301, 2010.

[121] SM. Law, GR. Bellomy, PJ. Schlax, and MT. Record. *In vivo* thermodynamic analysis of repression with and without looping in lac constructs. Estimates of free and local lac repressor concentrations and

of physical properties of a region of supercoiled plasmid DNA *in vivo*. *J. Mol. Bio.*, 230(1):161–173, 1993.

[122] VM. Krishnamurthy, V. Semetey, PJ. Bracher, N. Shen, and GM. Whitesides. Dependence of Effective Molarity on Linker Length for an Intramolecular Protein-Ligand System. *J Am Chem Soc*, 129:1312–1320, 2007.

[123] Nobuhiko Saitô, Kunihiko Takahashi, and Yasuo Yunoki. The statistical mechanical theory of stiff chains. *Journal of the Physical Society of Japan*, 22(1):219–226, 1967.

[124] P. A. Wiggins, T. Van Der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, D. Dekker, and P. C. Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology*, 1:137–141, 2006.

[125] AJ. Spakowitz. Wormlike chain statistics with twist and fixed ends. *Europhys Lett*, 73(5):684–690, 2006.

[126] Jiro Shimada and Hiromi Yamakawa. Ring-closure probabilities for twisted wormlike chains. application to dna. *Macromolecules*, 17(4):689–698, 1984.

[127] Richard R Sinden. *DNA structure and function*. Gulf Professional Publishing, 1994.

[128] Zev Bryant, Michael D Stone, Jeff Gore, Steven B Smith, Nicholas R Cozzarelli, and Carlos Bustamante. Structural transitions and elasticity from torque measurements on dna. *Nature*, 424(6946):338–341, 2003.

[129] Hernan G Garcia and Rob Phillips. Quantitative dissection of the simple repression input–output function. *Proceedings of the National Academy of Sciences*, 108(29):12173–12178, 2011.

[130] Steven B Smith, Laura Finzi, and Carlos Bustamante. Direct mechanical measurements of the elasticity of single dna molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992.

[131] L. Czapla, M.A. Grosner, D. Swigon, and W. K. Olson. Interplay of Protein and DNA structure revealed in simulations of the *lac operon*. *PLOS ONE*, 8(2):e56548, 2013.

[132] Charalampos G Kalodimos, Nikolaos Biris, Alexandre MJJ Bonvin, Marc M Levandoski, Marc Guennuegues, Rolf Boelens, and Robert Kaptein. Structure and flexibility adaptation in nonspecific and specific protein-dna complexes. *Science*, 305(5682):386–389, 2004.

[133] C. Jeppesen, JY. Wong, TL. Kuhl, JN. Israelachvili, N. Mullah, S. Zalipsky, and CM. Marques. Impact of polymer tether length on multiple ligand-receptor bond formation. *Science*, 293(5529):465–468, 2001.

[134] DC. Zappulla and TR. Cech. Yeast telomerase RNA: A flexible scaffold for protein subunits. *Proc. Natl. Acad. Sci. USA*, 101(27):10024–10029, 2004.

[135] PHV. Hippel, Revzin A., CA Gross, and AC. Wang. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: 1. The lac operon: equilibrium aspects. *Proc. Natl. Acad. Sci. USA*, 71(12):4808–4812, 1974.

[136] S. Johnson, J.-W. van de Meent, R. Phillips, C. Wiggins, and M. Lindén. Multiple LacI-mediated loops revealed by Bayesian statistics and tethered particle motion. *under review at Nucleic Acids Res*, 2014.

[137] James E Haber. Mating-type genes and mat switching in saccharomyces cerevisiae. *Genetics*, 191(1):33–64, 2012.

[138] Shiv IS Grewal and Danesh Moazed. Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634):798–802, 2003.

[139] Lisa Postow, Christine D Hardy, Javier Arsuaga, and Nicholas R Cozzarelli. Topological domain structure of the escherichia coli chromosome. *Genes & development*, 18(14):1766–1779, 2004.

[140] Jeff ZY Chen, Heng-Kwong Tsao, and Yu-Jane Sheng. Diffusion-controlled first contact of the ends of a polymer: Crossover between two scaling regimes. *Physical Review E*, 72(3):031804, 2005.

[141] Paul A Wiggins and Philip C Nelson. Generalized theory of semiflexible polymers. *Physical Review E*, 73(3):031906, 2006.

[142] Reza Afra and Brian A Todd. Kinetics of loop formation in worm-like chain polymers. *The Journal of Chemical Physics*, 138(17):174908, 2013.

[143] W Stockmayer and H Jacobson. Intramolecular reaction in polycondensations. *J. Chem. Phys*, 18:1600–1606, 1950.

[144] Holger Merlitz, Karsten Rippe, Konstantin V Klenin, and Jörg Langowski. Looping dynamics of linear dna molecules and the effect of dna curvature: a study by brownian dynamics simulation. *Biophysical journal*, 74(2):773–779, 1998.

[145] Nicolas Douarche and Simona Cocco. Protein-mediated dna loops: Effects of protein bridge size and kinks. *Physical Review E*, 72(6):061902, 2005.

[146] Yi-Ju Chen, Stephanie Johnson, Peter Mulligan, Andrew Spakowitz, and Rob Phillips. Modulation of dna loop lifetimes by the free energy of loop formation. *Proc Natl Acad Sci USA*, 111(49):17396–401, 2014.

[147] Andrew J Spakowitz and Zhen-Gang Wang. Exact results for a semiflexible polymer chain in an aligning field. *Macromolecules*, 37(15):5814–5823, 2004.

[148] Andrew J Spakowitz and Zhen-Gang Wang. End-to-end distance vector distribution with fixed end orientations for the wormlike chain model. *Physical Review E*, 72(4):041802, 2005.

[149] Shafigh Mehraeen, Bariz Sudhanshu, Elena F Koslover, and Andrew J Spakowitz. End-to-end distribution for a wormlike chain in arbitrary dimensions. *Physical Review E*, 77(6):061803, 2008.

[150] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66(5):052301, 2002.

[151] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[152] Huan-Xiang Zhou. Rate theories for biologists. *Quarterly reviews of biophysics*, 43(02):219–293, 2010.

[153] Hannes Risken. *Fokker-Planck Equation*. Springer, 1984.

[154] J-F Allemand, S Cocco, N Douarche, and G Lia. Loops in dna: an overview of experimental and theoretical approaches. *The European Physical Journal E*, 19(3):293–302, 2006.

[155] Karsten Rippe, Peter H von Hippel, and Jörg Langowski. Action at a distance: Dna-looping and initiation of transcription. *Trends in Biochemical Sciences*, 20(12):500–506, 1995.

[156] Donald C Rau and V Adrian Parsegian. Direct measurement of the intermolecular forces between counterion-condensed dna double helices. evidence for long range attractive hydration forces. *Biophysical journal*, 61(1):246–259, 1992.

[157] Elena F Koslover and Andrew J Spakowitz. Discretizing elastic chains for coarse-grained polymer models. *Soft Matter*, 9(29):7016–7027, 2013.

[158] Elena F Koslover and Andrew J Spakowitz. Systematic coarse-graining of microscale polymer models as effective elastic chains. *Macromolecules*, 46(5):2003–2014, 2013.

[159] Hans R Kricheldorf. Cyclic polymers: Synthetic strategies and physical properties. *Journal of Polymer Science Part A: Polymer Chemistry*, 48(2):251–284, 2010.

[160] D. Colquhoun and B. Sakmann. Fluctuations in the microsecond time range of the current through single acetylcholine receptor ion channels. *Nature*, 294:464–466, 1981.

[161] N. Laurens, S. R. W. Bellamy, A. F. Harms, Y. S. Kovacheva, S. E. Halford, and G. J. L. Wuite. Dissecting protein-induced DNA looping dynamics in real time. *Nucleic Acids Res*, 37(16):5454–5464, 2009.

[162] J.D. Revalee, G.A. Blab, H.D. Wilson, J.D. Kahn, and J.-C. Meiners. Tethered particle motion reveals that LacI·DNA loops coexist with a competitor-resistant but apparently unlooped conformation. *Biophys J*, 106(3):705–715, 2014.

[163] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez Jr., and C. H. Wiggins. Learning rates and states from biophysical time series: A Bayesian approach to model selection and Single-Molecule FRET data. *Biophys. J.*, 97(12):3196–3205, 2009.

[164] F. Persson, M. Lindén, C. Unoson, and J. Elf. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nature Methods*, 10:265–269, 2013.

[165] D. Colquhoun and F. J. Sigworth. Fitting and statistical analysis of single-channel records. In B. Sakmann and E. Neher, editors, *Single Channel Recording*, pages 191–263. Plenum Press, 1983.

[166] C. Manzo and L. Finzi. Quantitative analysis of DNA-looping kinetics from Tethered Particle Motion experiments. *Methods Enzymol*, 475:199–220, 2010.

[167] R.B. Winter, O.G. Berg, and P.H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry*, 20(24):6961–6977, 1981.

[168] SC. Kou, BJ. Cherayil, W Min, BP. English, and XS. Xie. Single-molecule Michaelis-Menten equations. *J. Phys. Chem. B*, 109(41):19068–19081, 2005.

[169] M. Gottesman. Bacteriophage lambda: the untold story. *J. Mol. Biol.*, 293:177–180, 1999.

[170] 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.

[171] H. Echols and C. Gross. *Operators and promoters : The story of molecular biology and its creators.* University of California Press, New York, NY, 1st edition, 2001.

[172] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. The bacteriophage phi29 portal motor can package DNA against a large internal force. *Nature*, 413:748–752, 2001.

[173] 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.

[174] A. M. Oppenhein, O. Koniler, J. Stavans, D. L. Court, and S. Adhya. Switches in bacteriophage lambda development. *Ann Rev Gen*, 39:409–429, 2005.

[175] F. St-Pierre and D. Endy. Determination of cell fate selection during phage lambda infection. *Proc Natl Acad Sci USA*, 105:20705–20710, 2008.

[176] L. Zeng, S. O. 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.

[177] J. E. Johnson. Virus particle maturation: insights into elegantly programmed nanomachines. *Curr Opin Struct Biol*, 20:210–216, 2010.

[178] 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.

[179] 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.

[180] R. Moldovan, E. Chapman-McQuiston, and X. L. Wu. On kinetics of phage adsorption. *Biophys J*, 93:303–315, 2007.

[181] I. Molineux. No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Molec Microbiol*, 40:1–8, 2001.

[182] J. Bohm, O. Lambert, A. S. Frangakis, L. Letellier, W. Baumeister, and J. L. Rigaud. FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. *Curr Biol*, 11:1168–1175, 2001.

[183] 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.

[184] V. Gonzalez-Huici, M. Salas, and J. Hermoso. The push-pull mechanism of bacteriophage phi29 DNA injection. *Mol Microbiol*, 52(2):529–40, 2004.

[185] I.J. Molineux. Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? *Virology*, 344:221–229, 2006.

[186] M. Ptashne. *A Genetic Switch, 3rd ed.* Cold Spring Harbor Laboratory Press (Cold Spring Harbor), New York, 2004.

[187] S. Tzlil, 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.

[188] A. S. Petrov and S. C. Harvey. Structural and thermodynamic principles of viral packaging. *Structure*, 15:21–27, 2007.

[189] Z. Li, J. Wu, and Z. G. Wang. Osmotic pressure and packaging structure of caged DNA. *Biophys J*, 94:737–746, 2008.

[190] 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.

[191] 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.

[192] 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.

[193] 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.

[194] A. S. Petrov and S. C. Harvey. Packaging double-helical DNA into viral capsids: structures, forces, and energetics. *Biophys J*, 95:497–502, 2008.

[195] A. Siber, Dragar M., V. A. Parsegian, and R. Podgornik. Packing nanomechanics of viral genomes. *Eur Phys J E Soft Matter*, 26:317–325, 2008.

[196] J. B. Stock, B. Rauch, and S. Roseman. Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. *J Biol Chem*, 252:7850–7860, 1977.

[197] Yi Deng, Mingzhai Sun, and Joshua W. Shaevitz. Direct measurement of cell wall stress stiffening and turgor pressure in live bacterial cells. *Phys. Rev. Lett.*, 107:158101, Oct 2011.

[198] J. Kindt, S. Tzlil, A. Ben-Schaul, and W. M. Gelbart. DNA packaging and ejection forces in bacteriophage. *Proc Natl Acad Sci USA*, 98:13671–13674, 2001.

[199] 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.

[200] A.K. Filali-Maltouf and B. Labedan. The energetics of the injection process of bacteriophage lambda DNA and the role of the ptsM/pel-encoded protein. *Biochem Biophys Res Comm*, 130:1093–1101, 1985.

[201] P. Boulanger and L. Letellier. Ion channels are likely to be involved in the two steps of phage T5 dna penetration into escherichia coli cells. *J Biol Chem*, 267:3168–3172, 1992.

[202] A.K. Filali-Maltouf and B. Labedan. Host cell metabolic energy is not required for injection of bacteriophage T5 DNA. *J Bacteriol*, 153:124–133, 1983.

[203] M. M. Inamdar, W. M. Gelbart, and R. Phillips. Dynamics of DNA Ejection from Bacteriophage. *Biophys J*, 91:411–420, 2006.

[204] 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.

[205] 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.

[206] 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.

[207] 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.

[208] D. Mackay and V. Bode. Events in lambda injection between phage absorption and DNA entry. *Virology*, 72:154–166, 1976.

[209] 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.

[210] 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.

[211] 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.

[212] 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.

[213] 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.

[214] 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.

[215] E. Lubeck and L. Cai. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods*, 9(7):743–748, 2012.

[216] P. Kemp, M. Gupta, and IJ. Molineux. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Mol Microbiol*, 53(4):1251–1265, 2004.

[217] 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.

[218] B. Huang, W. Wang, M. Bates, and X. Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319(5864):810–813, 2008.

[219] M. M. Inamdar, W. M. Gelbart, and Phillips R. Dynamics of DNA ejection from bacteriophage. *Biophys J*, 91:411–420, 2006.

[220] M.K. Waldor and J.J Mekalanos. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*, 272:1910–1914, 1996.

[221] F. Hassan, M. Kamruzzaman, J.J. Mekalanos, and S.M. Faruque. Satellite phage TLCϕ enables toxigenic conversion by CTX phage through dif site alteration. *Nature*, 467:982–985, 2010.

[222] 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.

[223] I. Katsura and Hend. Length determination in bacteriophage lambda tails. *Cell*, 39:691–698, 1984.

[224] S. L. Gotta, O. L. Jr Miller, and S. L. French. rRNA transcription rate in *Escherichia coli. J Bacteriol*, 173:6647–6649, 1991.

[225] 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.

[226] Ajay Gopinathan and Yong Woon Kim. Polymer translocation in crowded environments. *Phys. Rev. Lett.*, 99:228106, Nov 2007.

[227] Douglas E Smith, Sander J Tans, Steven B Smith, Shelley Grimes, Dwight L Anderson, and Carlos Bustamante. The bacteriophage &phis; 29 portal motor can package dna against a large internal force. *Nature*, 413(6857):748–752, 2001.

[228] Yann R Chemla, K Aathavan, Jens Michaelis, Shelley Grimes, Paul J Jardine, Dwight L Anderson, and Carlos Bustamante. Mechanism of force generation of a viral dna packaging motor. *Cell*, 122(5):683–692, 2005.

[229] Prashant K Purohit, Mandar M Inamdar, Paul D Grayson, Todd M Squires, Jané Kondev, and Rob Phillips. Forces during bacteriophage dna packaging and ejection. *Biophysical journal*, 88(2):851–866, 2005.

[230] Prashant K Purohit, Jané Kondev, and Rob Phillips. Mechanics of dna packaging in viruses. *Proceedings of the National Academy of Sciences*, 100(6):3173–3178, 2003.

[231] Xiangyun Qiu, Donald C Rau, V Adrian Parsegian, Li Tai Fang, Charles M Knobler, and William M Gelbart. Salt-dependent dna-dna spacings in intact bacteriophage λ reflect relative importance of dna self-repulsion and bending energies. *Physical review letters*, 106(2):028102, 2011.

[232] Alex Evilevitch, Laurence Lavelle, Charles M Knobler, Eric Raspaud, and William M Gelbart. Osmotic pressure inhibition of dna ejection from phage. *Proceedings of the National Academy of Sciences*, 100(16):9292–9295, 2003.

[233] Paul Grayson, Alex Evilevitch, Mandar M Inamdar, Prashant K Purohit, William M Gelbart, Charles M Knobler, and Rob Phillips. The effect of genome length on ejection forces in bacteriophage lambda. *Virology*, 348(2):430–436, 2006.

[234] Ian J Molineux. No syringes please, ejection of phage t7 dna from the virion is enzyme driven. *Molecular microbiology*, 40(1):1–8, 2001.

[235] Víctor González-Huici, Margarita Salas, and José M Hermoso. The push–pull mechanism of bacteriophage ø29 dna injection. *Molecular microbiology*, 52(2):529–540, 2004.

[236] Charles S Peskin, Garrett M Odell, and George F Oster. Cellular motions and thermal fluctuations: the brownian ratchet. *Biophysical journal*, 65(1):316–324, 1993.

[237] Serge G Lemay, Debabrata Panja, and Ian J Molineux. Role of osmotic and hydrostatic pressures in bacteriophage genome ejection. *Physical Review E*, 87(2):022714, 2013.

[238] Paul Grayson, Lin Han, Tabita Winther, and Rob Phillips. Real-time observations of single bacteriophage λ dna ejections in vitro. *Proceedings of the National Academy of Sciences*, 104(37):14652–14657, 2007.

[239] Ji-Zeng Wang, Long Li, and Hua-Jian Gao. Compressed wormlike chain moving out of confined space: A model of dna ejection from bacteriophage. *Acta Mechanica Sinica*, 28(4):1219–1226, 2012.

[240] JB Stock, B Rauch, and S Roseman. Periplasmic space in salmonella typhimurium and escherichia coli. *Journal of Biological Chemistry*, 252(21):7850–7861, 1977.

[241] Yi Deng, Mingzhai Sun, and Joshua W Shaevitz. Direct measurement of cell wall stress stiffening and turgor pressure in live bacterial cells. *Physical review letters*, 107(15):158101, 2011.

[242] David Van Valen, David Wu, Yi-Ju Chen, Hannah Tuson, Paul Wiggins, and Rob Phillips. A single-molecule hershey-chase experiment. *Current Biology*, 22(14):1339–1343, 2012.

[243] Mark Ptashne. *A genetic switch: phage lambda revisited*, volume 3. Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY:, 2004.

[244] Bradley R Parry, Ivan V Surovtsev, Matthew T Cabeen, Corey S O'Hern, Eric R Dufresne, and Christine Jacobs-Wagner. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell*, 156(1):183–194, 2014.

[245] Felix Höfling and Thomas Franosch. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76(4):046602, 2013.

[246] Stephanie C Weber, Andrew J Spakowitz, and Julie A Theriot. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Physical review letters*, 104(23):238102, 2010.

[247] Jae-Hyung Jeon, Natascha Leijnse, Lene B Oddershede, and Ralf Metzler. Anomalous diffusion and power-law relaxation of the time averaged mean squared displacement in worm-like micellar solutions. *New Journal of Physics*, 15(4):045011, 2013.

[248] Asaf Tal, Rinat Arbel-Goren, Nina Costantino, Joel Stavans, et al. Location of the unique integration site on an escherichia coli chromosome by bacteriophage lambda dna in vivo. *Proceedings of the National Academy of Sciences*, 111(20):7308–7312, 2014.

[249] F. St-Pierre and D. Endy. Determination of cell fate selection during phage lambda infection. *Proc Natl Acad Sci USA*, 105:20705–20710, 2008.

[250] L. Zeng, S. O. 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.

[251] M. De Paepe and F. Taddei. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS Biol.*, 4(7):e193, 2006.

[252] PP. Deniss and M. Nomura. Stringent control of ribosomal-protein gene-expression in Escherichia coli. *Proc Natl Acad Sci USA*, 71(10):3819–3823, 1974.

[253] NJ. Pearson, HM. Fried, and JR Warner. Yeast use translational control to compensate for extra copies of a ribosomal-protein gene. *Cell*, 29(2):347–355, 1982.

[254] F. Bachand, DH. Lackner, J. Bahler, and PA. Silver. Autoregulation of ribosome biosynthesis by a translational response in fission yeast. *Mol. Cell. Biol.*, 26(5):1731–1742, 2006.

[255] 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.

[256] J. Q. Boedicker, H. G. Garcia, S. Johnson, and R. Phillips. DNA-bending protein HU masks the sequence-dependence of repressor-mediated loop formation *in vivo*. 2012.

[257] J. Boedicker, H. Garcia, and R. Phillips. Theoretical and experimental dissection of DNA loop-mediated repression. *PRL, in press*, 2011.

[258] Zach Hensel, Xiaoli Weng, Arvin Cesar Lagda, and Jie Xiao. Transcription-factor-mediated dna looping probed by high-resolution, single-molecule imaging in live e. coli cells. *PLoS biology*, 11(6):e1001591, 2013.

[259] Marianne De Paepe and François Taddei. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS biology*, 4(7):e193, 2006.

[260] Daniel H Goldhill and Paul E Turner. The evolution of life history trade-offs in viruses. *Current opinion in virology*, 8:79–84, 2014.

[261] Eric C Keen. Tradeoffs in bacteriophage life histories. *Bacteriophage*, 4(2), 2014.