Changing the DNA Sequence Alters the Kinetics of Protein-Mediated DNA Looping

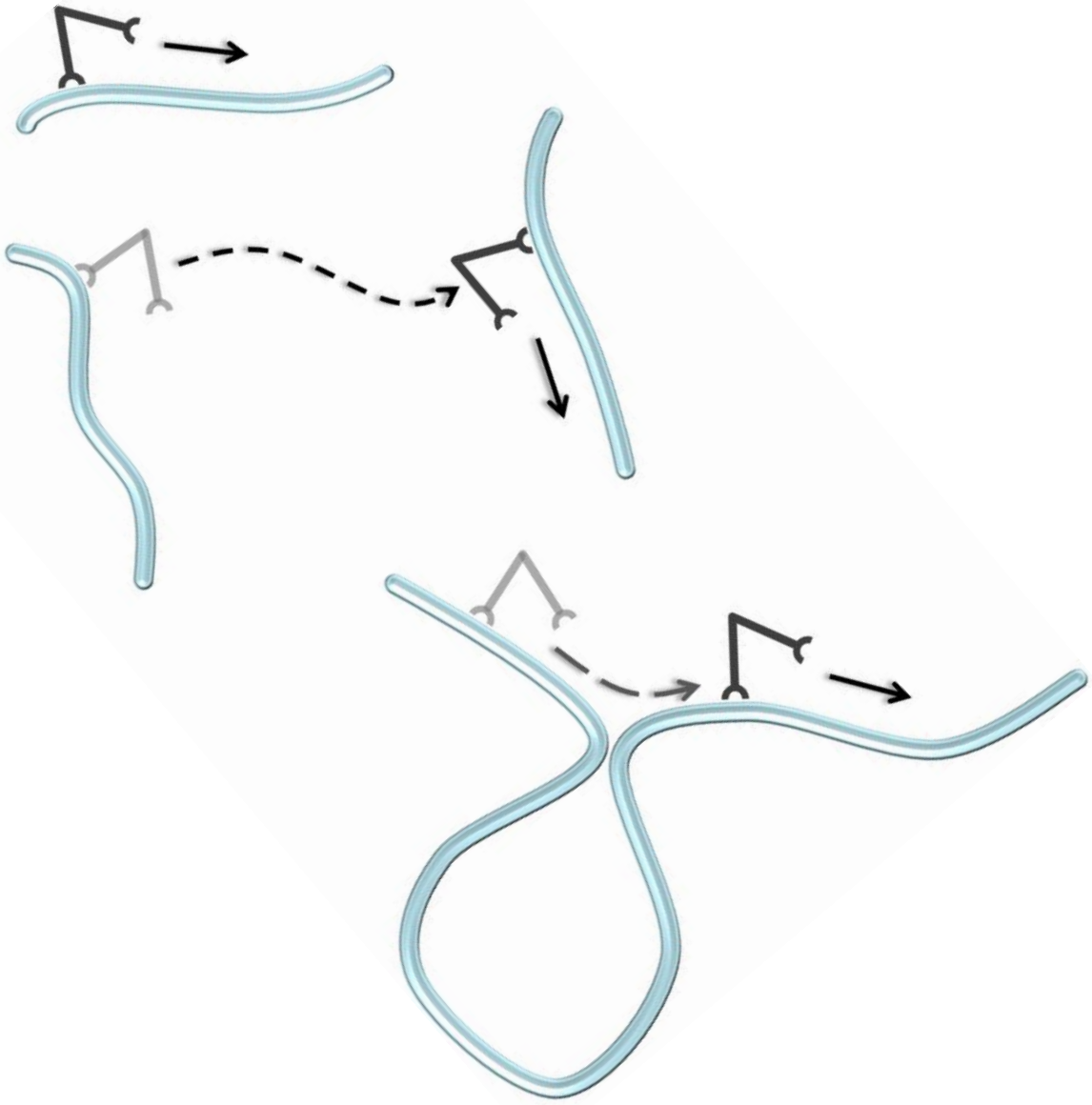
By

Joel David Revalee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in the University of Michigan
2013

Doctoral Committee:

Professor Jens-Christian Meiners, Chair
Professor Duncan Steel
Professor Nils G. Walter
Associate Professor Michal Zochowski



© Joel D. Revalee 2013
All Rights Reserved

DEDICATION

For my parents, Jerry and Peggy Revalee, without whose love and support this would not have been possible.

ACKOWLEDGMENTS

I would first like to acknowledge my advisor and mentor in experimental Biophysics, Dr. Jens-Christian Meiners. I am thankful for his patience, advice and most importantly his support over my time in graduate school.

I appreciate the help and advice that I have received from all members of the Meiners lab past and present. I'm grateful for the help and friendship of Dr. Yih-Fan Chen, who showed me how to make the various buffers I used in my experiments. I also appreciate the kindness and friendship of Dr. Krishnan Raghunathan, who taught me almost every biochemical technique that I've used. Many students have passed through over the years. I am glad to know Alan Kandinov, Justin Blaty and Ben Juliar, who all have bright futures ahead of them.

I'm thankful for the mentoring of Dr. Gerhard Blab, who taught me the experimental technique of Tethered Particle Motion, and who also taught me the basics of experimental biophysics.

I am especially grateful to Dr. Jason D. Kahn for helping this project in its early stages and for the generous gifts of DNA plasmids and lac Repressor protein (LacI). I am thankful for Dr. Kahn's students, Kathy Goodson, Aaron Haeusler and Larry Edelman for purifying the LacI.

I am thankful for my friends and lab mates, Julia Bourg and Dr. Rudra Kafle, who are both good scientists and great people. I enjoy our discussions about science, and I have grown particularly fond of Nepali cuisine.

I appreciate helpful discussions with Andy Hirsh and Todd Lillian about the theoretical models of DNA looping.

I have been blessed to work with many bright and motivated undergraduates. Henry Wilson, Mike Chu and Brian Lewis have bright futures ahead of them.

I was supported during my time in graduate school both by a National Science Foundation (NSF) Graduate Research Fellowship and by a Regents Fellowship from the University of Michigan.

I owe a great deal to Dr. Mohamed Laradji, who was my research advisor at the University of Memphis. He taught me how to use the theoretical tools of physics to perform research in biology. Without his support, I would not be a researcher in the field of biophysics. I am glad to have been a part of the Laradji research group, and I'm grateful for the friendship of all of the lab members there over the years. I also appreciate the friendship and help in membrane modeling of Dr. Michael Hore, who is an incredible scientist and good human being.

I am grateful for my friend Andres Millan, who has taught me the value of thinking outside of the box. Of all my friends, he is the only one who knows the answer to the question, "how is this even possible?"

I have benefitted from a loving and supporting family. My brother, Jerry, has always been an encouragement. My uncle, Larry Revalee, is a good man and a great uncle who has shared many good times with me in his visits to Ann Arbor. My grandmother, Fern, has always encouraged me.

I could not have asked for better parents than mine, Jerry and Peggy Revalee, to whom this work is dedicated.

Lastly, and most importantly, I acknowledge the guidance and love of my Lord and my God, Jesus Christ, whose strength has made me strong, and through Him I have done this work. It is His wonderful creation that I endeavor to understand.

> "Before the mountains were brought forth, or ever thou hadst formed the earth and the world, even from everlasting to everlasting, thou art God."
>
> Psalm 90:2

> "He revealeth the deep and secret things: he knoweth what is in the darkness, and the light dwelleth with him"
>
> Daniel 2:22

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

LIST OF APPENDICES

APPENDIX

ABSTRACT

DNA is more than just a blueprint – it is a dynamic molecule with physical and mechanical properties that are significant for its biological functions. At any given time, various enzymes bind to and release from a multitude of sites on DNA *in vivo*. These enzymes not only bind to DNA, but many of them also mediate the formation of DNA loops. In the present work, loop formation and breakdown in single DNA molecules is explored for intrinsically curved DNA substrates. It is found that while loops form and break down on the order of minutes in these DNAs, the protein cannot be competed off of an intrinsically curved DNA substrate with a large excess of competitor DNA for on the order of a day. This is posited to be due to an additional interaction between LacI and the nonspecific DNA near its binding sites. Furthermore, effects are observed on DNA loops in single molecule experiments when DNA outside the looping region is altered. To explain these observations, another interaction is hypothesized in which flanking DNA associates with LacI while in the looped state. Thus, the experimental results of this work reveal two additional protein-DNA interactions in a repressor protein (i.e. LacI) which is a bacterial paradigm of gene regulation. The modified kinetic schemes which follow from these newly hypothesized interactions reveal the robust nature of protein-DNA interactions and motivate future experiments with LacI. Furthermore, it is found that altering a mechanical property (i.e. intrinsic curvature) on a DNA substrate is observed to significantly alter the kinetics of biomolecular interactions between DNA and protein.

Chapter I

Introduction

1.1 DNA looping is a Physics research question!

From the construction of the first microscopes centuries ago and the electrical *in vitro* experiments with muscles in the 1800s to the modern elucidation of the structure of deoxyribonucleic acid (DNA) and the building of the first rigorous theoretical models of neuron action potentials, physicists have worked passionately to reach a fundamental understanding of the nature and principles at work in the micro-scale world of biological life.

The modern experimental and theoretical tools that I have acquired in my training as a physics researcher allow me to pursue scientific questions in molecular biology which both complement and enhance the current body of knowledge of the cellular micro-world. Furthermore, the kinds of questions that a physicist may ask about a biological system are qualitatively different from those in other disciplines. Whereas a biochemist may pursue the participating enzymes in an elaborate metabolic pathway, a physicist would attempt to characterize that system through simplifying assumptions which reproduce the thrust of its experimental behavior; these approaches are not mutually exclusive, and both yield fruitful

knowledge of biological processes. A particular example of this duality of research strategies is evident in the modern understanding of the dynamics of DNA in the cell.

Within the last few decades, much effort has been spent in elucidating the nanoscale properties of DNA both *in vitro* and *in vivo* (Ref Levens). Robust theoretical models of linear polymers have been applied to a variety of biological molecules (Ref Bustamante, Ref Marko). These models have advanced the current understanding of DNA as both an information-rich code for all of the proteins in a cell as well as a dynamic and interactive polymer (Ref Milstein 2011).

One particular process which is ubiquitous in the cell is that of protein-mediated DNA looping. Such an interaction is widely utilized at the molecular level to regulate gene expression (Ref Wilson 2007), excise segments of DNA (Ref Catto) and modulate viral phage infections (Ref Manzo). Because DNA looping involves the protein-mediated stabilization of a bent segment of DNA, it is a process that is sensitive to changes in DNA physical properties.

Moreover, many such physical properties of DNA have been examined, including the effects of line tension (Ref Blumberg 2005 (2), Ref Chen 2010 (1), Ref Chen 2010 (2)), bending stiffness (Ref Kandinov, Ref Raghunathan), persistence length (Ref Chen 2009), and twist and supercoiling (Ref Dorman, Ref Koslover). These properties and lines of inquiry are grounded in a physical picture of biological systems.

To fully characterize the function of DNA *in vivo*, one must understand the degree to which its physical properties can influence and perhaps even dictate the timescale and strength of its interactions with various cellular components. Thus, quantitative *in vitro* studies of important biological entities are an integral niche of contemporary biological experimentation that help to illuminate the wealth of dynamical and biochemical factors which underlie the molecular processes of life.

Chapter II

Background

2.1 Introduction

Years of biochemical research have revealed a range of protein-DNA interactions that affect transcription and, in fact, evince spatiotemporal variations which are necessary for the function of genetic networks (Ref Hasty). Theoretical models of even the most basic systems can quite easily invoke dozens of free parameters spread over numerous coupled differential equations to characterize the various components of the gene set in question (Ref Santillán).

2.2 The lac Repressor protein – the fundamental model

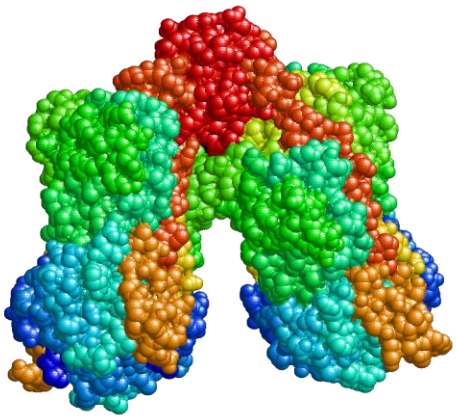


Figure 2.1 – Crystal structure of the lac Repressor Protein obtained from the Protein Data Bank (Ref PDB 1LBI).

It thus behooves any researcher to begin with the simplest possible system to quantitatively characterize the physical properties and dynamics of a protein-DNA interaction. For these reasons, experiments in the present work were performed on the lac Repressor protein (LacI, figure 2.1). Used universally in molecular biology, LacI is found in the bacterium *Escherichia Coli* (*E. coli*).

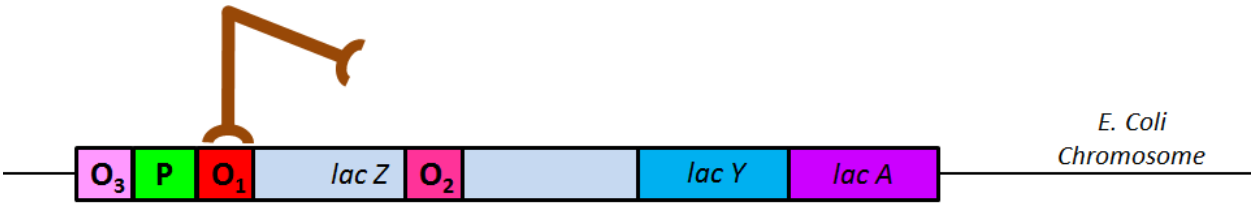


Figure 2.2 – Schematic of the lac operon *in vivo*. LacI (brown) binds simultaneously to two out of the three possible operators (O1, O2 & O3) in order to form a loop on the DNA and enhance repression of the three genes of the lac operon – *lacZ, lacY* & *lacA*. The promoter binding site (P) is shown in green.

LacI was one of the first transcription factors to be experimentally characterized (Ref Jacob). It represses transcription of the lac operon by binding simultaneously to two distant operator sites on the bacterial chromosome and stabilizing the intervening DNA, composed of the operon genes, in a loop (Ref Mossing, Ref Oehler 1990, Ref Oehler 1994, Ref Schleif); a schematic of the operon is shown in figure 2.2. In vivo, the lac operon contains three binding sites for the repressor protein (figure 2.2 – O_1, O_2 & O_3). These sites, called operators, vary in their binding affinities for LacI with O_1 having the strongest binding and O_3 having the weakest. These operators are variants of a palindromic ideal sequence ('AATTGTGAGCGCTCACAATT'); termed O_{SYM}, this sequence binds the DNA-binding domains (DBD) of LacI tighter than any of the *in vivo* operators.

LacI has been crystallized (Ref Lewis) and its well-resolved three-dimensional structure both while bound to DNA and to inducer is known (figure 2.1). The repressor protein is a tetramer composed of identical subunits; it has a plane of symmetry, and may be viewed as a dimer of dimers. Its N-terminal segments contain DNA-binding domains (DBD) which bind tightly to specific operator sites *in vivo*. The C-terminal domains of LacI comprise a 4-helix bundle (Ref Wilson 2007) which stabilizes a pair of LacI dimers into its native, tetrameric conformation; this section is non-covalently bound, and thus flexible. Evidence from *in vitro* bulk (Ref Taraban) and single-molecule experiments (Ref Rutkauskas) reveal that the tetramer

may change conformations from its V-shaped crystal structure to a more open conformation (figure 2.3).

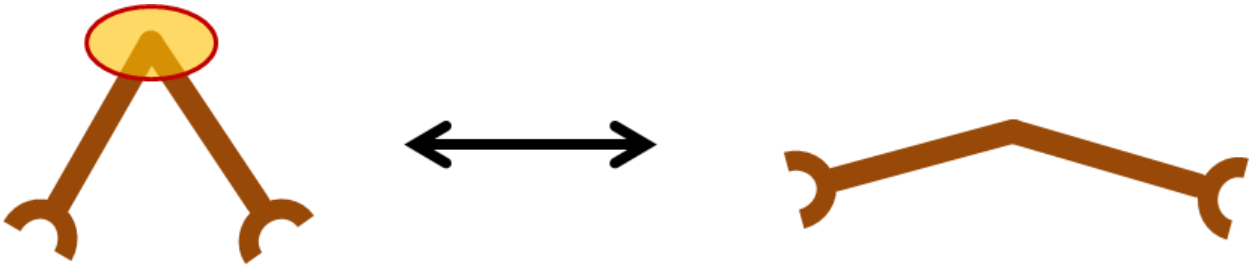


Figure 2.3 – A representation of LacI (brown) changing conformations from its V-shaped crystal structure (left) to a more open form (right). The 4-helix bundle of LacI, located inside the yellow ellipse (top left), is a non-covalently bound structure and is thought to act as a hinge and allow the repressor protein to adopt other conformations to relieve enthalpic strain in DNA looping (Ref Taraban, Ref Goyal, Ref Hirsh, Ref Haeusler).

2.3 Other proteins and biomolecules involved in the lac operon

LacI is not the only protein that binds to DNA in the vicinity of the lac operon. In fact, specific binding sites for catabolite activator protein (CAP) and RNA polymerase (RNAP) are immediately upstream from O_1 on the bacterial DNA (figure 2.2). When glucose levels in the cell are low, the level of cyclic adenosine monophosphate (cAMP) in the cytosol rises and cAMP binds to CAP as a ligand – increasing the affinity of CAP for binding of DNA (Ref Nelson). However, even if CAP binds to its specific DNA site, the operon is still repressed by LacI. For repression of the operon to cease, lactose levels in the cytosol must rise to the point that allolactose, a byproduct of β-galactosidase (which is itself the product of the *lacZ* gene – gray rectangle in figure 2.2) and the natural inducer of LacI, is present in high enough levels to bind LacI as a ligand and induce the operon. Upon binding of CAP to its specific site upstream of O_1, RNAP binds to the promoter site (the green rectangle in figure 2.2), interacts (protein-protein) with CAP and initiates transcription of the operon genes. All three genes of the operon are

transcribed by RNAP into a single mRNA transcript, which is then translated into three individual enzymes by the ribosome.

2.4 The operon genes and their enzymatic products

The three genes of the lac operon are translated into three separate enzyme products. The individual genes of the lac operon, *lacZ*, *lacY* and *lacA* (figure 2.2), are expressed, respectively, as the enzymes β-galactosidase, β-galactoside permease and β-galactoside transacetylase. β-galactosidase catalyzes the hydrolysis of lactose, a disaccharide, into glucose and galactose, both monosaccharides. If glucose and galactose are present in high enough cellular concentration, then it is possible for β-galactosidase to work in reverse and generate a significant population of disaccharides; one such disaccharide product is allolactose, the natural inducer of LacI. Allolactose is a chemical isomer of lactose and is composed of glucose and galactose bound with a different glycosidic linkage than in lactose. Thus, by having allolactose be an inducer of the lac operon, the cell has an elaborate concentration sensor of cytosolic lactose.

The product of the *lacY* gene, β-galactoside permease, is a transmembrane protein that pumps lactose into the *E. coli* bacterium. If both lactose and glucose are present in the environment of the bacterium, *E. coli* will preferably metabolize glucose until it is no longer present in high concentrations. Only then will levels of cAMP *in vivo* rise enough to bind CAP as a ligand, which then binds to its specific site near the operon to aid in transcription initiation. This preference of energy sources, glucose versus lactose, is termed catabolite repression; it is present in numerous organisms (Ref Deutscher), and serves as a mechanism for the *E. coli* bacterium to preferably metabolize the simplest sugars first.

Lastly, the product of the *lacA* gene is a cytosolic protein designated β-galactoside transacetylase. While it is known that this enzyme adds an acetyl group to galactosides, the purpose of this chemical reaction and any possible function (or lack thereof) of this enzyme in lactose metabolism is a mystery (Ref Roderick). This is the final component of the lac operon DNA.

While the description of all of the operon genes, as well as their various control elements, is complete, there remain other proteins that affect the transcription, translation and repression of the lac operon. An artificial inducer of LacI, isopropyl β-D-1-thiogalactopyranoside (IPTG), exists and is used widely in molecular biology. This chemical mimics allolactose and binds to LacI as a ligand (Ref Goodson).

In vitro experiments with LacI usually involve a "naked" DNA substrate, where it is "naked" in the sense that no other DNA-associated proteins are present. However, the cellular environment is quite crowded and a variety of proteins associate with the nucleoid DNA in *E. coli*. Proteins like HU and histone-like nucleoid-structuring protein (H-NS) bind DNA and alter its large-scale structure (Ref Dillon). These DNA-bound proteins, which are present all along the *E. coli* chromosome, function through processes such as inducing curvature, wrapping the DNA or even uniting distant segments (Ref Dillon, Ref Nir). While many of their structures are known (Ref Swinger), their effects on gene expression are still being experimentally investigated.

2.5 Other DNA looping proteins

Though the present work exclusively discusses DNA loops mediated by LacI, this repressor is not the only protein that facilitates DNA loop formation. In fact, the lac repressor

protein belongs to a class, called the LacI/GalR family, of similar repressors (Ref Swint-Kruse). The gal repressor (GalR), for which (along with LacI) this protein family is named, is also a DNA looping protein found in *E. coli* (Ref Qian). Whereas LacI is tetrameric in solution, GalR is dimeric and forms tetramers upon a dimer pair binding the two operators of the gal operon and mediating formation of a DNA loop (Ref Weickert). Such similar mechanisms are found across this family of proteins (Ref Meinhardt).

There are other DNA looping proteins present in the *E. coli* bacterium which are completely unrelated to those of the LacI/GalR family. One such example is the λ repressor, which is native to the phage *Enterobacteria phage λ* (abbreviated λ phage from this point forward) – a bacterial virus that infects *E. coli* (Ref Manzo). This viral protein is responsible for the environmentally-controlled switching in infected *E. coli* cells between states of lysogeny (non-virulence caused by integration of the viral DNA into the bacterial genome) and lysis (proliferation of new viruses caused by induction of the λ DNA). The λ repressor protein, called cI, loops DNA over several thousand base pairs and is responsible for maintaining the *E. coli* in a lysogenic (non-virulent) state (Ref Manzo). The cI protein is a dimer natively and may form DNA-bound tetramers and octamers to mediate formation of short- or long-range DNA loops (Ref Zurla).

All of these proteins serve important biological functions in the *E. coli* bacterium. While the mechanisms of DNA looping mediated by these enzymes as well as LacI have been detailed in this chapter, it remains to show how further studies of a classic prokaryotic repressor like LacI could yield new information. In fact, there are several fruitful questions about LacI and its DNA substrates that are open for experimental and theoretical exploration. These are the subjects of the chapters that follow.

Chapter III

DNA Loop Formation and Breakdown at the Single-Molecule Level

3.1 Introduction

Over the last few decades, research in Biophysics has progressed at an ever-increasing pace. New experimental (Ref Deniz) and theoretical (Ref Thirumalai) methods allow scientists to ask different kinds of questions about the nanoscale mechanisms at play in protein-DNA interactions (Ref Greenleaf). While years of thorough biochemical research have elucidated a range of such enzyme-nucleic acid associations, detailed descriptions of the underlying mechanics and dynamics of many these processes are still open for discovery. Moreover, in attempting to give a basic scientific understanding of many biomolecular interactions, most published descriptions resort to generic renderings of pairs of amorphous blobs (Ref Phillips). A major goal of biophysics research is thus to enhance the molecular understanding of cellular processes by transforming nanoscale cartoon-like descriptions in molecular biology into fully characterized physical pictures with well understood biochemical, thermodynamic and mechanical properties. Many researchers previously trained in physics experimental and theoretical techniques are choosing to give their talents to pursue this goal (Ref Oullette).

In order to dissect the dynamics of nucleic acids and proteins *in vivo*, biophysicists have developed a range of experimental techniques to isolate and probe the mechanics of individual pairs of biomolecules *in vitro*. The invention of an assortment of techniques, from force spectroscopy (i.e. optical trapping or tweezing, Ref Bustamante, Ref Marko, Ref Smith), atomic force microscopy (Ref Santos) and single-molecule Fluorescence Resonance Energy Transfer (smFRET, Ref Ha) to patch (voltage) clamping (Ref Zhao) and single-dye tracing in biomembranes (Ref Schutz), has opened up a diversity of new avenues of investigation in cellular processes.

3.2 Tethered Particle Motion (TPM) – Watching single DNA molecules move with visible light

To peer into the dynamics of DNA loop formation and breakdown mediated by lac repressor protein (LacI), a modern, single-molecule technique called Tethered Particle Motion (TPM, Ref Blumberg 2005 (1), Ref Finzi) is utilized; a simple schematic of the principle of TPM is shown in figure 3.1 (below). This technique was originally developed to observe the kinetics of transcription by RNA polymerase (Ref Schafer); it has since been applied to numerous protein-DNA interactions in which binding of an enzyme to a specific sites or sites on a substrate effectively shortens the range of Brownian motion of the DNA (Ref Fan, Ref Liu, Ref Plénat), which is approximated as a constrained, entropically elastic polymer (Ref Segall). The effective "shortening" of the DNA tether upon loop formation (figure 3.1, "L") is visualized as a drop in the root mean-squared (rms) motion of a streptavidin-coated polystyrene bead (540 nanometers, Spherotech Inc.) that is chemically attached to DNA at one end via a streptavidin-biotin bond (one of the DNA primers is biotinylated); the opposite end of the DNA molecule is attached to

the microscope slide via a digoxigenin chemical modification (to the other DNA primer) that binds to an antibody coating on the slide's coverslip.

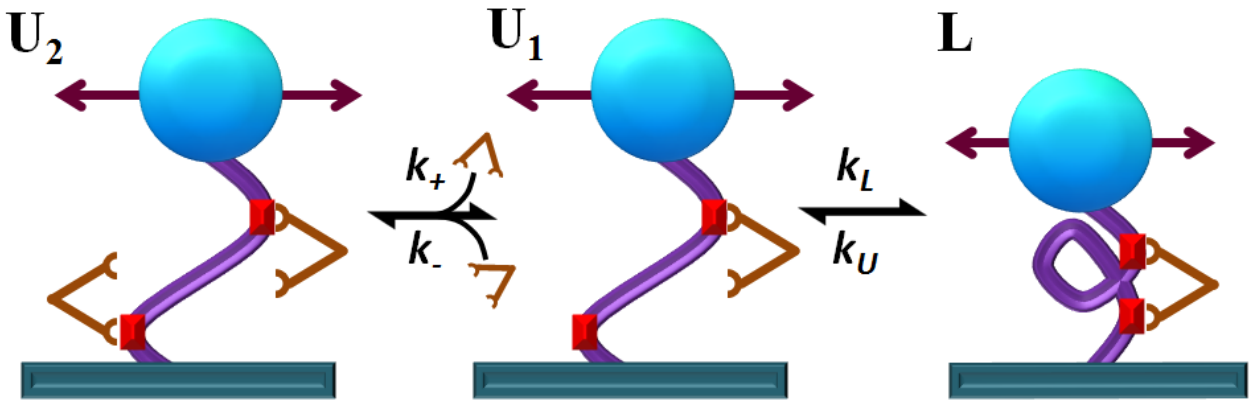


Figure 3.1 – A simple representation of the principle of loop detection by Tethered Particle Motion (TPM). In this work, the authors propose that the unlooped state, a combination of (U_2) doubly- and (U_1) singly-occupied DNA (purple, with operator binding sites in red) with their corresponding inter-conversions (at rates k_+ and k_-, respectively), reversibly (L) loops (at rate k_L) and unloops (at rate k_U) in the presence of lac Repressor protein (LacI, brown). The beads, tethered to DNA, are shown in light blue, while the arrows bracketing them demonstrate bead planar, root mean-squared (rms) Brownian motion.

3.3 Sample Preparation

3.3.1 Making flow cells

For the tethered particle assay, the DNA be tethered is attached to the bottom of a flow cell on one end, and to a polystyrene microsphere on the other end, following the protocols of Chen et al (Ref 2010 (1)). Custom made flow chambers were constructed by epoxy gluing Polyethylene Tygon tubing into holes drilled into a microscope cover slide (3"×1"×1 mm, Corning® 2947-3x1); a #1 cover slip was then heat-attached to the slide by means of parafilm (2"×250' roll, Parafilm M, PM-992), into which a thin channel had been cut. The finished chambers had typical dimensions of 20×2×0.1 mm^3. Finished chambers were incubated with anti-digoxigenin antibody (20 μg/mL, Roche, polyclonal from sheep) and the surface blocked

using 1% w/w casein in PTC buffer (20mM Tris-acetate, pH 8.0, 130 mM KCl, 4mM $MgCl_2$, 0.1 mM EDTA, 20 μg/mL bovine serum albumin (BSA), and 80 μg/mL heparin; additionally, 1 or 3 mg/mL α-casein was normally added) to reduce nonspecific surface interactions. Doubly labeled DNA was incubated with an excess of streptavidin-coated polystyrene microspheres (diameter 0.44 or 0.54 μm, Spherotech, IL) for 30-60 min, on a rotating rack to prevent sedimentation of the microspheres. Then, this mixture was introduced into the antibody-coated chamber and allowed to attach to the surface for 5-10 minutes. Unbound beads and DNA were then flushed out with 200-400 μL PTC buffer.

3.3.2 Acquisition of looping data

Loop formation and breakdown assays were conducted by filling the sample cell with LBB^+ buffer (LBB buffer contains 10 mM Tris-HCl, 200 mM KCl, 0.1 mM EDTA, pH 7.4; when using LBB+, the following were added: 0.2 mM DTT, and 5% DMSO) with LacI (generously donated by Dr. Jason D. Kahn) at concentrations ranging from 110-160 pM, and in some cases added 5 mM $MgCl_2$. The motion of tethered microspheres was observed with a custom-built microscope using bright-field illumination. The images were collected by a 60× objective (Olympus PlanApo 60x /1.45 Oil, Olympus, Center Valley, PA) and imaged onto a Peltier-cooled Cascade 650 (Photometrics) or Pixelink PL-B741 (PixeLINK ®) CCD camera. The images were then run through a band-pass filter (regular FITC band-pass filter) to avoid chromatic imaging artifacts. Video data is acquired at a frame rate ranging from 20 to 40 frames per second for typically 15 minutes to 4 hours per measurement (an average of about two hours per data acquisition); a time series of example images of a single bead from video data are given in figure 3.2. Further details about image data processing are found in appendix A.

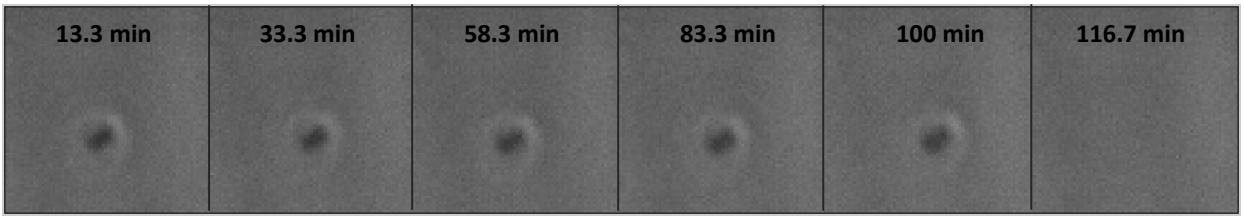


Figure 3.2 – Image video data acquired in Tethered Particle Motion (TPM) experiments. The six image snapshots in this figure are all from the same individual region of interest (ROI), at different time points, via light microscopy for a TPM experiment. Time stamps, in minutes, are given in black at the top of each snapshot. The dark spots surrounded by white rings are the (dark-centered) diffraction patterns at different times of a single bead tethered to the microscope coverslip. The time traces of x- and y- positions of this bead from all frames acquired are plotted in figure 3.3.

3.4 Analysis of bead positions

This section contains a series of figures from the full video data set of figure 3.2. While the bead positional tracking algorithm is described in appendix A, figures 3.3, 3.4, 3.5, 3.6 & 3.9 (as well as figure 3.2, above) describe in detail this bead's positional analysis, filtering and processing for TPM data.

Every bead from video data is isolated and, through a process described in appendix A, x- and y- positions are obtained. Once these are computed, the raw xy- coordinates as well as the individual images themselves are perused to check for bead ejection (defined below); explicitly, the chemical process for generating a bead-DNA-cover slip tether on a microscope slide is not permanent. Over the course of even a few minutes, the antibody-digoxigenin bond may release and allow the DNA/bead pair to unbind from its position on the microscope cover slip and diffuse freely in the solution of the flow cell; from this point on, this process will be referred to as bead "ejection". Ejection can be easily detected by a sudden apparent increase in rms motion due to the particle tracking algorithm no longer being able to localize the bead diffraction spot. In practice, bead image data must be closely monitored for ejection or other anomalies.

Once the "ejected" data is excised (figure 3.3), the remaining xy-positions are plotted and a rectangular area is defined outside which the data points are considered eccentric (figure 3.4). Only raw x- and y- positions within these constraints are used in the following processing steps.

The raw x- and y- positions are then run through a fifth-order, high-pass Butterworth filter of cutoff frequency 0.05 Hz (figure 3.5); this cutoff frequency is chosen in order remove long-timescale instrumental drift, which occurs at a slower pace than the abrupt changes in bead root-mean square (rms) displacement that accompany loop formation and breakdown events. The final processing step of positional data is performed on these filtered coordinates (figure 3.6).

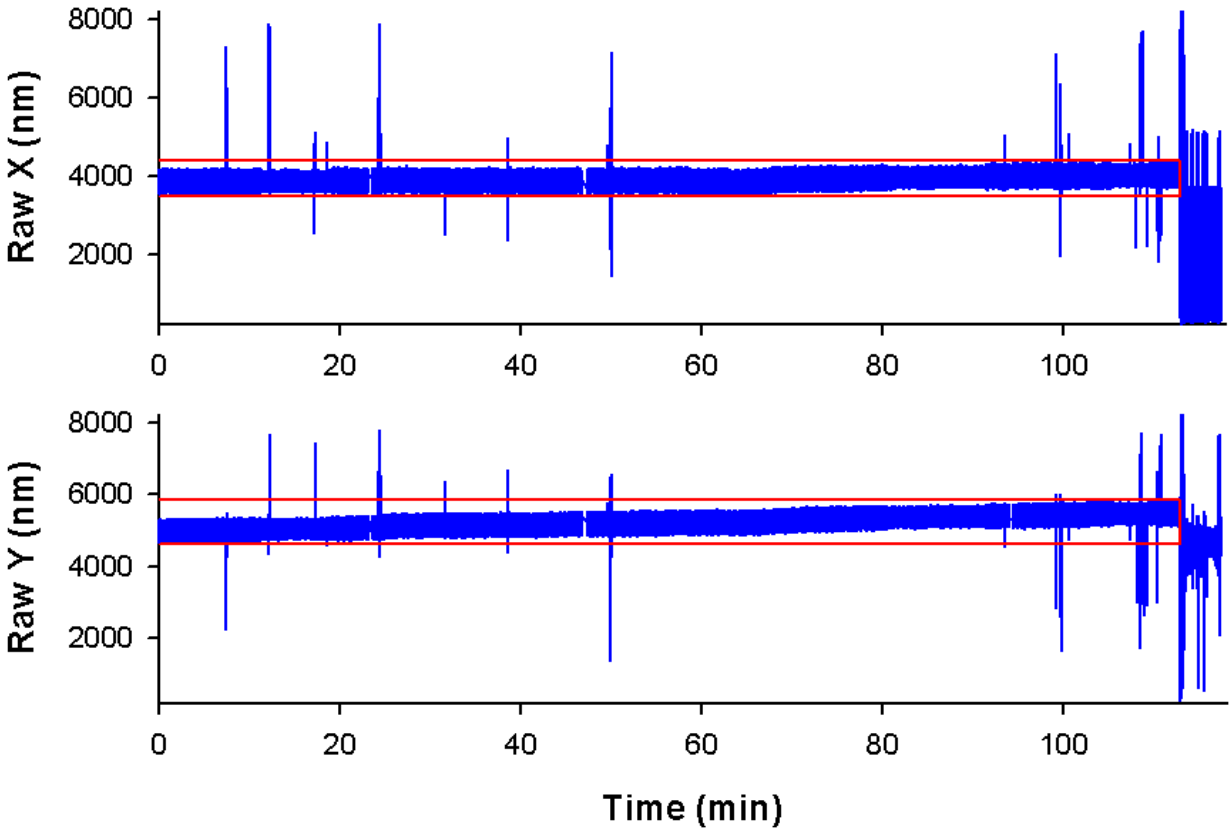


Figure 3.3 – Analyzing raw time trace data in Tethered Particle Motion (TPM) experiments. The two plots above show, respectively, the time courses (blue lines) of the x- and y- positions (in nanometers) of the bead (the same bead from figure 3.2). At 113 minutes in, the bead ejected, which is seen both by the absence of a bead in the 6th snapshot in figure 3.2 as well as the inability of the computer particle tracking algorithm to plot a well-defined bead course after this point in time. The red rectangles bounding a section of x- and y- data each show the portion of the unprocessed data that was used to infer bead motion. The red rectangles shown above bound the same positional area in figure 3.4.

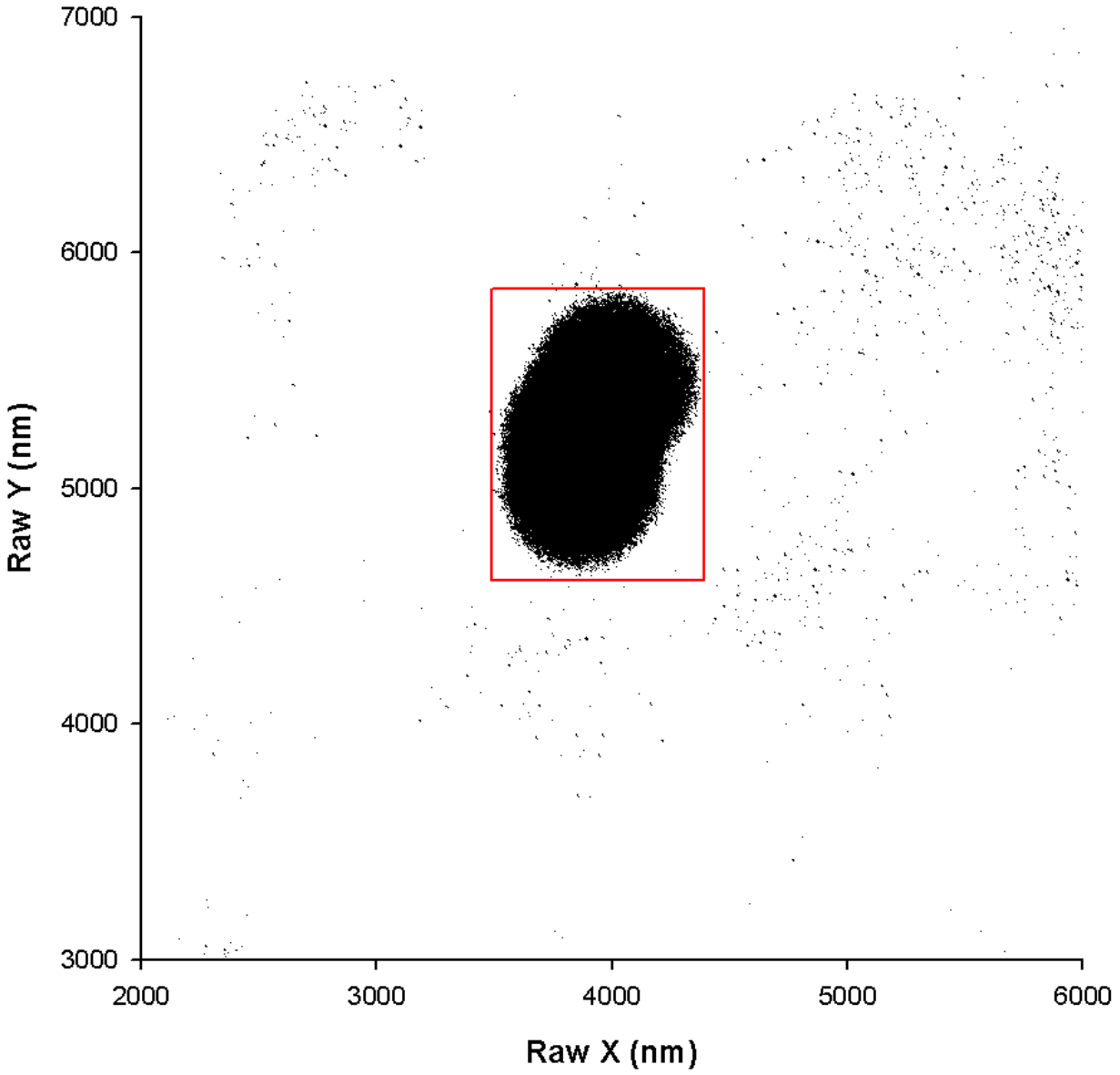


Figure 3.4 – Scatter plot of all of the raw x- and y- positional data (from figure 3.3) prior to bead ejection at 113 minutes. Acquired through Tethered Particle Motion (TPM) experiments, eccentric xy- coordinates are excised from the data by simply defining a rectangular boundary (red lines above) within the region of interest (ROI), outside which all points are excluded from data analysis. These points, though few in number, are the result of (rare) failures in the positional tracking algorithm of the bead to find the bead center. These two-dimensional bead positions have not been adjusted for slow-time drift, so the scatter plot is not circular. Further processing is necessary before this data is ready for use in positional root mean-square (rms) traces; the description of data processing steps continues in figures 3.5, 3.6 & 3.9. Figures 3.5 & 3.6 detail filtering of the xy-data that removes long-time drift, and figure 3.9 describes computation of root mean-square (rms) motion. The bead positions are given in nanometers. In data acquisition, all fitting is originally done in units of pixels; after all processing is finished, a multiplicative value of one hundred ten nanometers per camera pixel spatially normalizes the positional data.

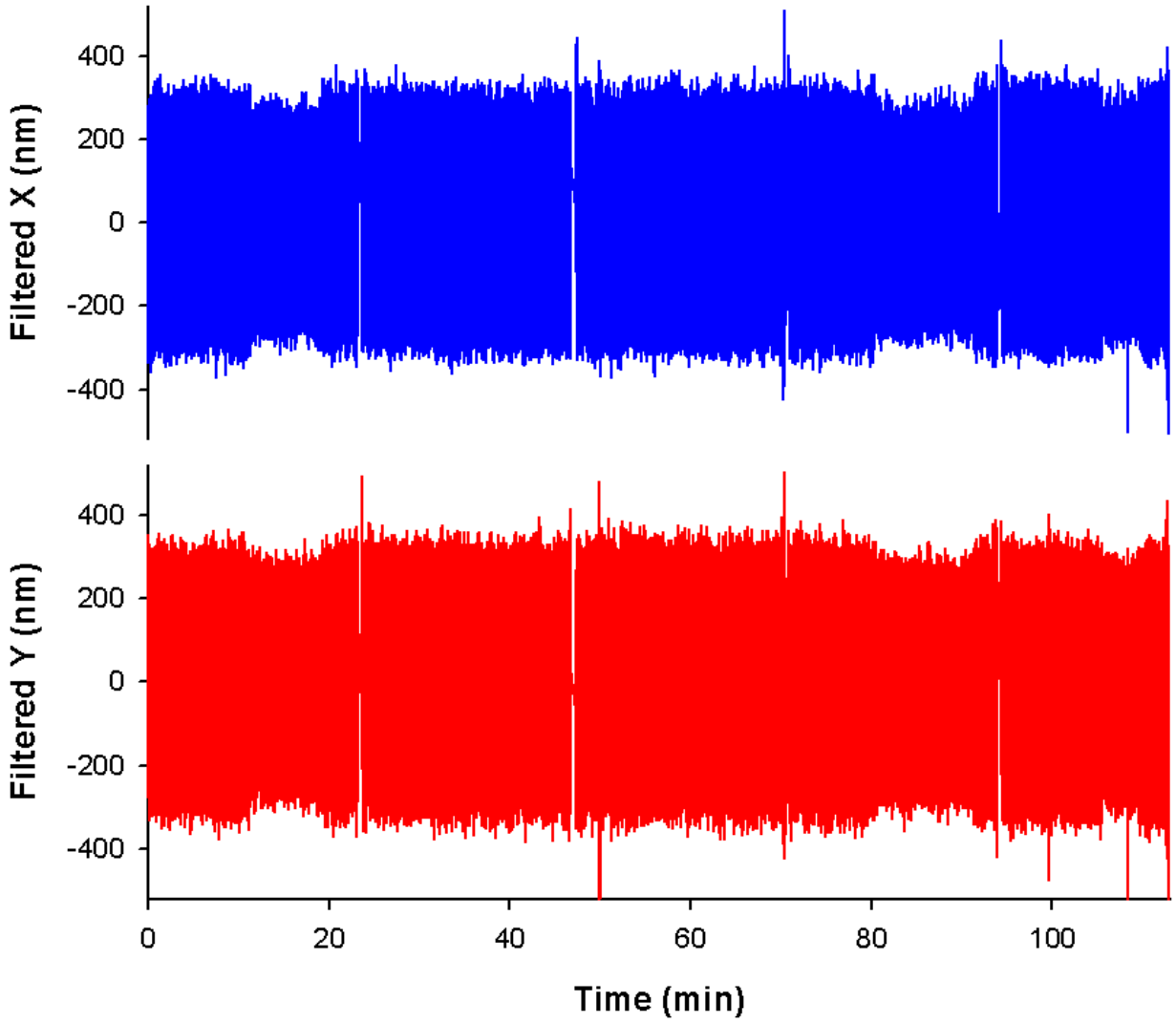


Figure 3.5 – Time courses of the filtered x- (blue) and y- (red) positional data. The raw x- and y- data from figure 3.4, after the processing steps depicted there, has been run through a fifth-order high-pass Butterworth filter with cutoff frequency 0.05 Hz; thus, any event with a characteristic timescale on the order of 20 seconds or more has been removed from the Fourier spectrum of the x- and y- positions. Positions have now been normalized by subtraction of the average positional vector, $\langle \vec{\rho} \rangle = \langle (x, y) \rangle$. Very few eccentric points now remain. From these filtered positions, I may now compute the bead radial length, $r = \sqrt{X^2_{FILTERED} + Y^2_{FILTERED}}$, at each time point. At this stage, the filtered x- and y- positions have minimal eccentricity and are circular to within the experimental error, as shown in figure 3.6.

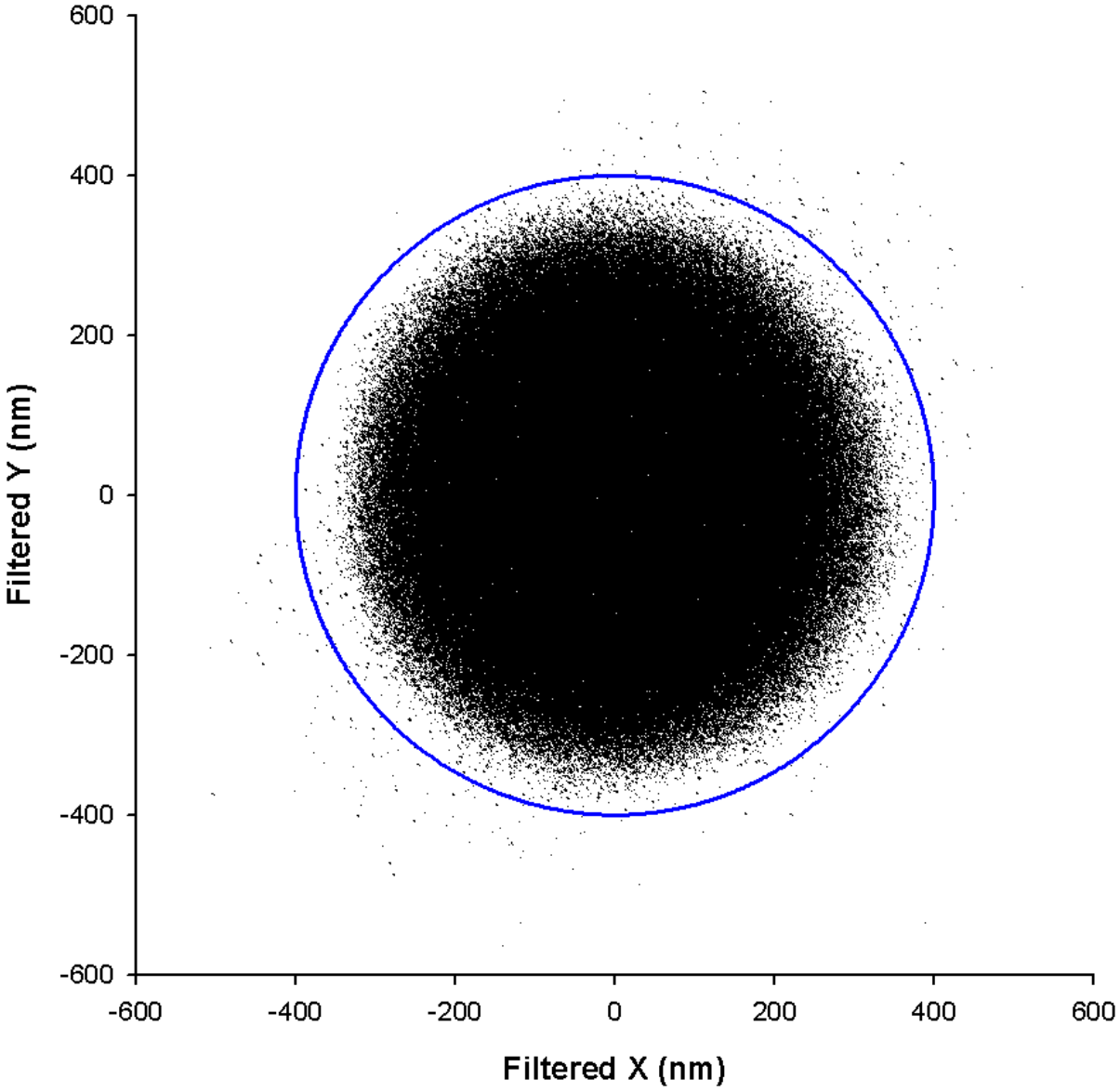


Figure 3.6 – Scatter plot of filtered x- and y- coordinates. The final stage of processing is performed on these values; a circle of appropriate radius is defined (blue circle) to exclude the optimal number of eccentric points. It is readily observed that these filtered coordinates are approximately circular. In fact, their precise circularity is measured as described in the text. Further processing of the x- & y- positional data is usually unnecessary. From this point, the only corrections or data artifacts that are removed are due to bead sticking on the coverslip (figure 3.7); bead sticking is easily distinguished by rms motion (figure 3.9) decreasing to less than 10% of the bead radius.

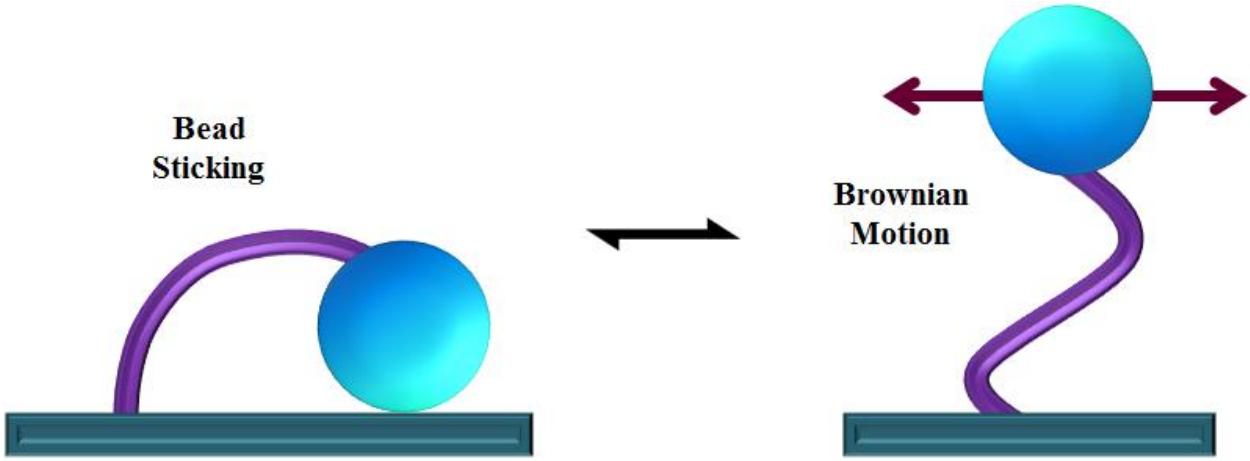


Figure 3.7 – A schematic of bead sticking. The bead sticking (left) is easily observed by a decrease in overall bead rms from its normal radius of motion (right) to less than 10% of the bead radius (in the experiments performed, the bead has diameter 540 nanometers). The DNA (purple) is tethered to a microscope coverslip (cyan) on one end and a polystyrene microsphere (light blue) on the other end. Bead sticking is reversible, and most sticking events last less than one minute.

3.5 Differentiating the "good" data from the "bad"

3.5.1 Multiple Tethers (MT)

Figures 3.2, 3.3, 3.4, 3.5 & 3.6 describe how, after data acquisition has finished and particle position tracking algorithms have been applied, the time courses of bead x- and y-coordinates are processed and filtered to exclude eccentric points and to remove instrument drift. At this stage, after these steps have been performed, a number of entire bead traces are excluded due to additional issues detailed below.

For each bead, a covariance matrix is computed of the filtered and processed x- and y-positions, $M_{COVARIANCE} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix}$. The eigenvalues of the covariance matrix, $(\lambda_{max}, \lambda_{min})$, are found. These eigenvalues can be used to measure the elliptical eccentricity of the bead's orbit around its mean position. The circularity is computed by taking the square root of the ratio of the

larger eigenvalue to the smaller one, $\sqrt{\lambda_{max}/\lambda_{min}}$, and the elliptical (orbital) eccentricity, ϵ, is computed as, $\epsilon = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} - 1$.

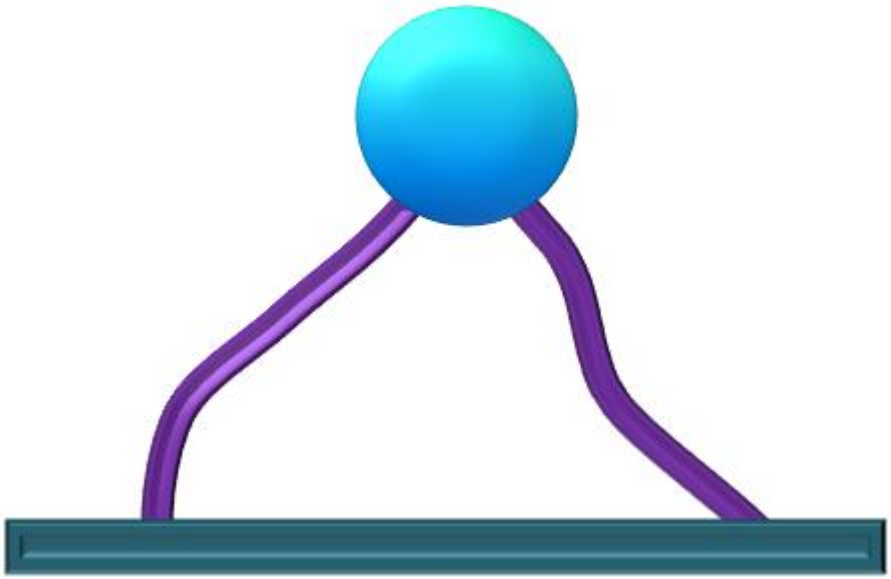


Figure 3.8 – A schematic of a multiple (i.e. double) tether. A bead (light blue) is tethered by more than one DNA tether (purple) to a microscope coverslip (cyan). A multiple tether is readily distinguishable by its anomalous orbital eccentricity. Such a multiple tether would have a "preferred" direction of motion perpendicular to the axis of tethering (i.e. out of the page).

The reason for computing the circularity of each bead is drawn schematically in figure 3.8. Although DNA tethers are incubated with an approximately four-fold excess of beads, there is a significant probability of multiple tethers on a single bead. These multiply tethered beads are identified by the fact that their multiple tethering points give them a preferred direction of motion (i.e. perpendicular to axis containing the tethering points). A singly-tethered bead has no preferred direction of motion, and thus should have eccentricity equal to zero, while a multiple tether should have nonzero eccentricity.

3.5.2 Multiple Beads (MB)

Because an excess of beads is incubated with DNA tethers, there is also the possibility that multiple beads associate into a glob; however, these are easily distinguished visually (they appear as an amalgamation of diffraction spots in a region of interest) or by computation of the

autocorrelation of the normalized, filtered positional vector, $\hat{r}(t) = (x_{FILT}(t), y_{FILT}(t))/r(t)$, where $r(t) = \sqrt{x_{FILT}^2(t) + y_{FILT}^2(t)}$. The autocorrelation function, $g(\tau) = \langle \hat{r}(t) \cdot \hat{r}(t+\tau) \rangle_t$, should follow the functional form of an exponential decay, $g(\tau; T_{CORR}) = e^{-\tau/T_{CORR}}$. For normal, singly-tethered beads, relaxation times (T_{CORR}) fall in the range of 60-70 milliseconds. Multi-bead conglomerates typically show larger bead relaxation times in the range of 100 milliseconds or above (even up to 300 milliseconds) in comparison to single-tethers. Bead relaxation times of less than 45 milliseconds were viewed as multiple tethers (confirmed by checking the bead circularity and also by the presence of decreased average rms) and were excluded from consideration.

All the ranges of bead autocorrelation times described above (multiple tether, normal bead, multiple bead) are bead size-dependent and the acceptable ranges of bead correlation times must be re-normalized if a different bead size is used. Throughout the data displayed below, only two bead sizes were used. For all 9C14 and 11C12 single-molecule data, polystyrene beads of diameter 440 nanometers (from Spherotech, Inc.) were employed. For all other DNAs used in tethered particle motion experiments, beads of diameter 540 nanometers were used. The autocorrelation time ranges described above are for the latter bead size (diameter 540 nanometers).

3.6 Computing the bead root mean-squared (rms) motion

After all processing and filtering have been performed on the x- and y- positional data, and after all spurious beads have been excluded, the root mean-square (rms) motion of the bead is computed as diagrammed in figure 3.9. From the processed, filtered x- and y- positions, a time series of bead radial values (rms), $r(t)$, is computed (figure 3.9 (a)); this bead radial value time

series is identical to the un-normalized $r(t)$ utilized in the positional autocorrelation calculations above.

In order to clearly distinguish looped or unlooped states in each data set, a smoothed, averaged version of the bead rms, $R_W(t) = \sqrt{\langle r^2(t) \rangle_W}$, is computed over a running window of W seconds; in all data sets, 2 second or 10 second running averages are found by taking the video frame rate, F ($71\ Hz$ for the data in figure 3.9), and using that to compute the number of frames corresponding to the running (time) window, W. The effect of smoothing the radial data is shown in figure 3.9 (b) & (c). The smoothed rms traces show clear state transitions, with looped states lasting several minutes at a time.

For the remainder of this work, all averaged rms traces (unless specified otherwise) are shown after being smoothed over a running window of ten seconds. This window size, W, is chosen for two reasons. First, any events less than ten seconds in duration, without further investigation of the associated rms and raw x- & y- positional data, could quite possibly be the result of anything from a spurious, single frame that failed in the position tracking algorithm to a momentary diffusion of an ejected bead through the focusing volume of the microscope setup. Second, increasing the window size decreases the noise of rms traces; this inverse relationship between the number of frames, N, in the filtering window and the noise, σ_{Rms}, follows $\sigma_{Rms} \sim 1/\sqrt{N}$. Moreover, the size of σ_{Rms} places a limitation on the number of rms levels or states that may be resolved in rms traces (more on that later).

Now that the rms traces are computed, it remains to be shown how looped and unlooped states are distinguished without relying upon the experimenter to "eyeball" the individual states. That is, it is possible forgo further manual data manipulation and to automate the higher-level calculations of looped and unlooped state lifetimes. This is covered in the next chapter (IV).

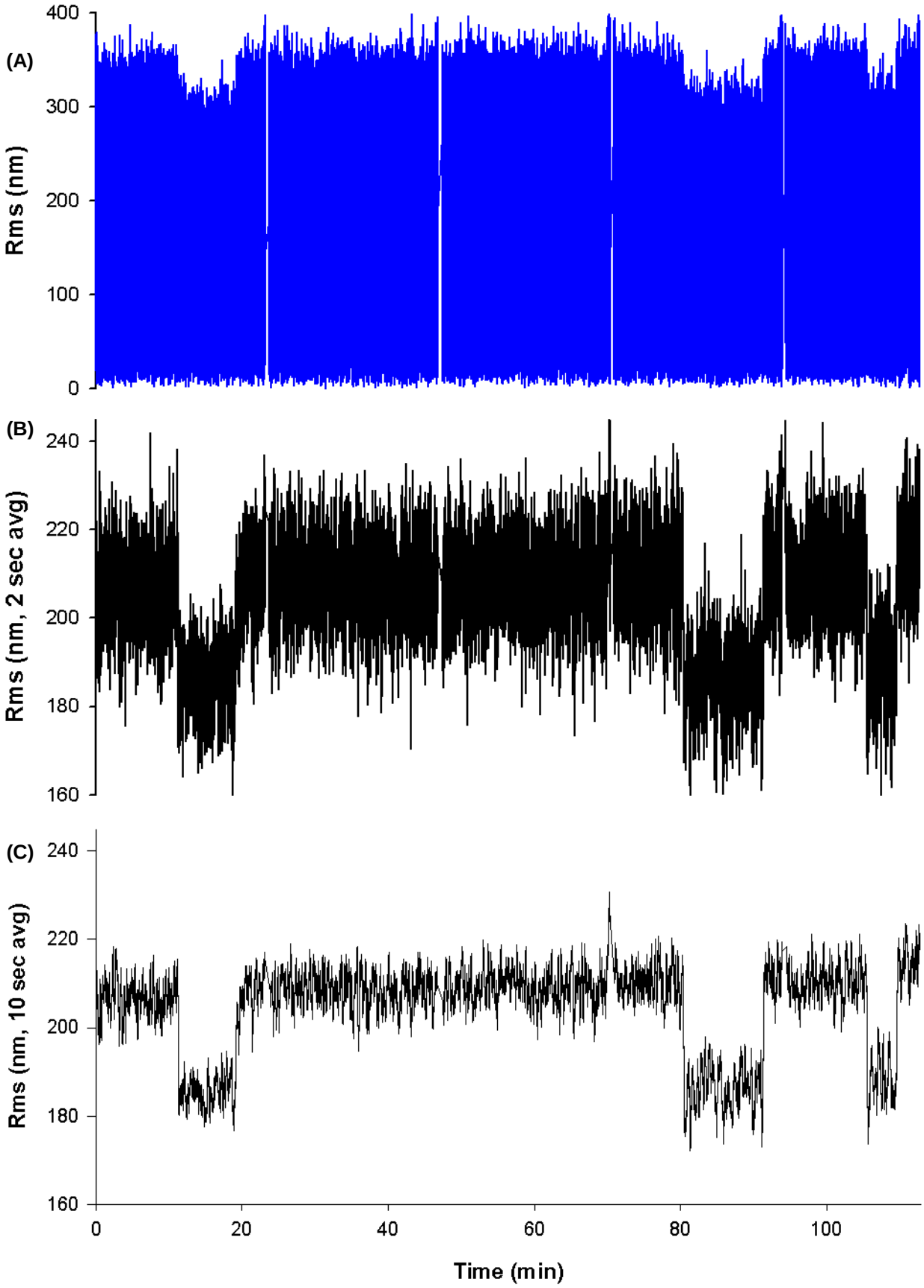

(A)
Rms (nm)
400
300
200
100
0
(B)
Rms (nm, 2 sec avg)
240
220
200
180
160
(C)
Rms (nm, 10 sec avg)
240
220
200
180
160
0
20
40
60
80
100
Time (min)

Figure 3.9 – The effect of averaging on bead positional root mean-square (rms) motion. The top plot (a) shows the time series of unaveraged bead rms (blue lines); this is the same data used in figures 3.2-3.6. When a (b) 2 second or (c) 10 second running average is applied to the data, not only does the rms motion look smoother visually, but looped and unlooped states become immediately distinguishable as well-defined rms levels.

Chapter IV

The Effects of Curvature on DNA Loops

4.1 Introduction

The characterization of the physical properties of DNA is a collaborative, ongoing effort in biophysical research that dates back many decades (Ref Hagerman, Ref Lankaš 2000) and continues into the present day (Ref Chen 2009, Ref Chen 2010 (1), Ref Chen 2010 (2), Ref Kandinov, Ref Raghunathan). Much effort has been spent in creating a full theoretical framework (Ref Goyal, Ref Kratky) that accounts for all of the empirical findings of DNA elasticity, chiefly the computations of average DNA bending stiffness and persistence length (Ref Bustamante, Ref Marko).

In the present work, a single molecule technique (Tethered Particle Motion (TPM)) is used to approach a different physical property; specifically, the question that is explored in the following sections is that of how the intrinsic curvature in a DNA substrate affects the kinetics of DNA loop formation and breakdown mediated by the lac Repressor protein (LacI). The various DNA constructs studied are described in some detail in appendix B at the end of this document. Although all studies reported here are novel in their results, the research findings presented were motivated by earlier biochemical work (Ref Mehta). This earlier work is described in the section that follows.

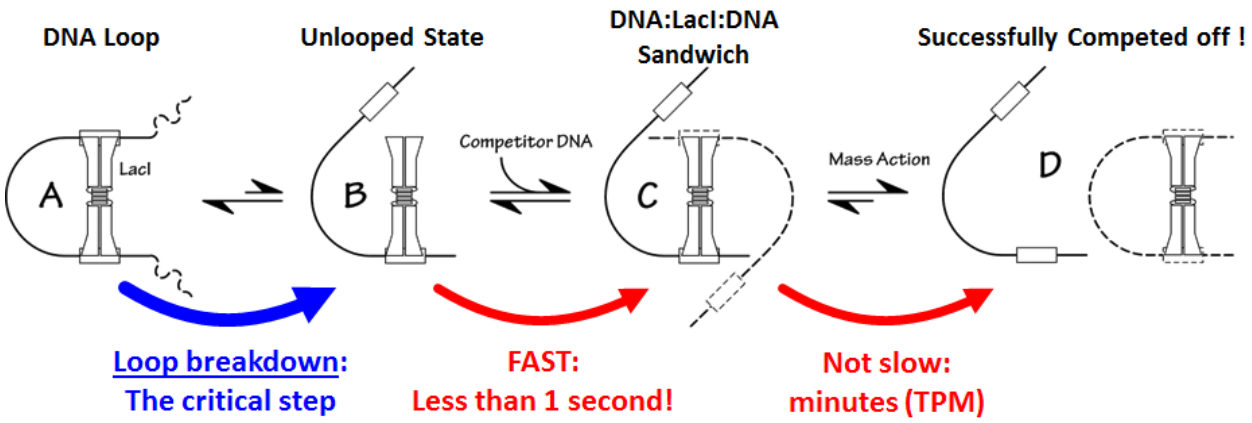


Figure 4.1 – A schematic of the bulk DNA competition assay of Mehta and Kahn (Ref). The (A) DNA substrate (solid lines) looped by LacI (B) breaks down and one DNA-binding domain (DBD) of LacI is free to (C) bind competitor DNA (dashed lines) and form a sandwich. At this point, LacI may either (B) let go of the competitor or (D) be competed off; there is equal chance of both. Section 4.2.3 details the calculation of the kinetic rate of transition from (B) to (C); it is found to be less than a second. Fitting of TPM unlooped state lifetime data to a kinetic model discussed later gives the rate of transition from (C) to (D); it is less than two minutes on average. Thus, the bulk DNA competition assay was thought to measure the loop breakdown rate, (A) to (B).

4.2 Motivation

4.2.1 Bulk DNA Competition Assays

The wealth of modern, *in vitro* experimental approaches in molecular biology enables researchers to ask new kinds of questions about the loop formation process. It is now feasible to compare quantitative models of this process to experiments probing underlying mechanics. For the lac repressor loop, the goal of testing sequence-dependent loop stability led to the design of DNA constructs with symmetric operators (abbreviated O_{SYM}) known to bind LacI more tightly (Ref. Sadler) bracketing phased A_6-tracts (Ref Barbič, Ref Lankaš 2010, Ref Strahs) that bend the DNA. The resulting LacI-DNA looped complexes were shown to be hyperstable *in vitro* using enzyme mobility shift (EMSA) DNA competition assays performed by Mehta & Kahn (Ref). The principle of this bulk competition assay is drawn above in figure 4.1. Three of Mehta and Kahn's constructs, designated 9C14, 11C12 and Unbent Control (UBC), are utilized in

single molecule experiments detailed in later sections of this chapter; 9C14 and 11C12 are hyperstable constructs, while UBC lacks any A-tract bends and is thus a straight control sequence for bulk and single molecule measurements.

In the bulk DNA competition assays by Mehta and Kahn (Ref), a one nanomolar concentration of the hyperstable DNA-LacI complex (after 0.5 hour initial incubation for LacI to find its DNA substrate) is incubated *in vitro* with up to a fifty-fold excess of free competitor DNA; the relative populations of free and LacI-bound substrate DNA are then measured after 0.5 and 24 hours incubation time with competitor. Relative population data after this time is summarized in table 4.1. Approximating DNA competition as a Poisson process, relative populations, P, of free DNA in the bulk competition assays may be converted into competition times, τ, by solving the relation $P(t) = exp(-t/\tau)$ at a given time point. Using these relations, lifetimes of hyperstable protein-DNA complexes are computed to be on the order of a day (table 4.1). However, one correction must be noted; although the UBC competition half-time, $T_{1/2}$, is computed to be about five hours (table 4.1), examination of the relative population data at 0.5 hours (Figure 5 in Ref Mehta) competition show approximately equal populations, and thus UBC's true competition half-time is on the order of 30 minutes.

DNA Construct	Relative Population After 24 hours	Contains A-Tract bends?	τ (hours)	$T_{1/2}$ (hours)
UBC	3.8 %	No	7.34	5.09
9C14	46.5 %	Yes	31.3	21.7
11C12	63.3 %	Yes	52.5	36.4

Table 4.1 – Bulk DNA competition data from Mehta & Kahn (Ref). Column 1 gives whether or not these DNA constructs contain intrinsic curvature in their looping regions. Column 2 lists the relative populations of LacI-bound substrate DNA after 24 hours incubation with a fifty-fold excess of competitor DNA. Columns three and four give, assuming an exponential decay process, the competition lifetimes or half-times, $T_{1/2} = \tau ln(2)$.

From the competition measurements in bulk assays, it was inferred that loop lifetimes in hyperstable DNAs were on the order of days. Also, lifetimes varied with the intrinsic curvature of the substrate DNA. This was expected based upon the design of the constructs, which were engineered to bend in different directions, thus giving rise to different predicted loop topologies.

Further experimentation using Fluorescence Resonance Energy Transfer (FRET) directly showed that hyperstable curved DNA constructs favored different pairs of operator-LacI binding orientations, or topologies, based upon their respective directions of curvature (Ref Edelman, Ref Morgan). These engineered hyperstable DNA templates were even used to generate a landscape of constructs upon which DNA competition could be performed (Ref Haeusler).

4.2.2 Computing the competition time from first-principles

In the bulk assay described above, how fast should free DNA competitor molecules at 50 nM concentration kinetically encounter bound LacI-DNA substrate at 1 nM concentration? If it is assumed that this process occurs as a diffusion-limited second-order reaction between two small molecules diffusing in water, then the kinetics of this system are well understood in the literature (Ref Alberty, Ref Zhou) and an approximate value for the second-order reaction rate, k_{COMP}, may be taken as $10^8 M^{-1} sec^{-1}$.

With this value for the reaction rate and the concentrations of DNA competitor, c_{COMP} (50 nM), one may write the rate of competition once the LacI-DNA is unlooped and thus LacI has a free DNA-binding domain (DBD) available to bind to competitor in solution as,

$$T_{1/2} = \frac{ln(2)}{k_{REACTION}} = \frac{ln(2)}{k_{COMP} C_{COMP}} \cong \frac{0.693}{10^8 M^{-1} sec^{-1} \times 5 \times 10^{-8} M} \tag{4.1}$$

which gives, $T_{1/2} = 0.14\ sec$. Equation 4.1 gives yields that in less than one second on average, once LacI-DNA unloops, competitor DNA binds to LacI! Thus, the bulk DNA competition assay was thought to be a sensitive diagnostic of DNA loop lifetimes in hyperstable constructs.

4.2.3 Theoretical predictions of DNA binding topologies for curved constructs

Elasticity models of DNA looping on the hyperstable constructs above predict that such intrinsic curvature can affect the bending energy of the DNA loop by as much as 37.5% (Ref Goyal (Figure 7G)); the corresponding absolute (elastic) energetic curvature benefit of 11.1 kJ/mol ($4.5k_BT$, Ref Goyal (Figure 7G)), when compared to the empirically measured DNA-LacI binding interaction energy of 64.9-66.1 kJ/mol (Ref Frank (Table 1)), is likely to greatly affect loop formation and stability. Furthermore, theoretical studies of energy landscapes for bent DNA constructs reveal that the LacI protein itself likely adopts different conformations in different binding topologies in order to reduce loop elastic strain (Ref Hirsh, Ref Lillian).

4.3 Single molecule measurements of looping kinetics

4.3.1 Measuring looped and unlooped state lifetimes in TPM data

In order to peer into the dynamics of hyperstable loop formation and breakdown, single molecule experiments were performed on two of Mehta & Kahn's intrinsically curved constructs, 9C14 and 11C12, as well as one control sequence, designated UBC, that lacked any A-tract bends. For each DNA construct used in tethered particle motion (TPM) experiments, several tethers in microscope slide region of interest would be observed in parallel for hours at a time. Details about image acquisition and data processing algorithms, including particle tracking, are explained in appendix A. Example traces from the TPM experiments of all three of the Kahn DNA constructs (Ref Mehta), after the processing steps of chapter III have been performed, are plotted in figure 4.2. DNA loop formation and breakdown are visible in these traces as abrupt transitions to well-defined lower or higher smoothed bead root mean-square (rms) levels.

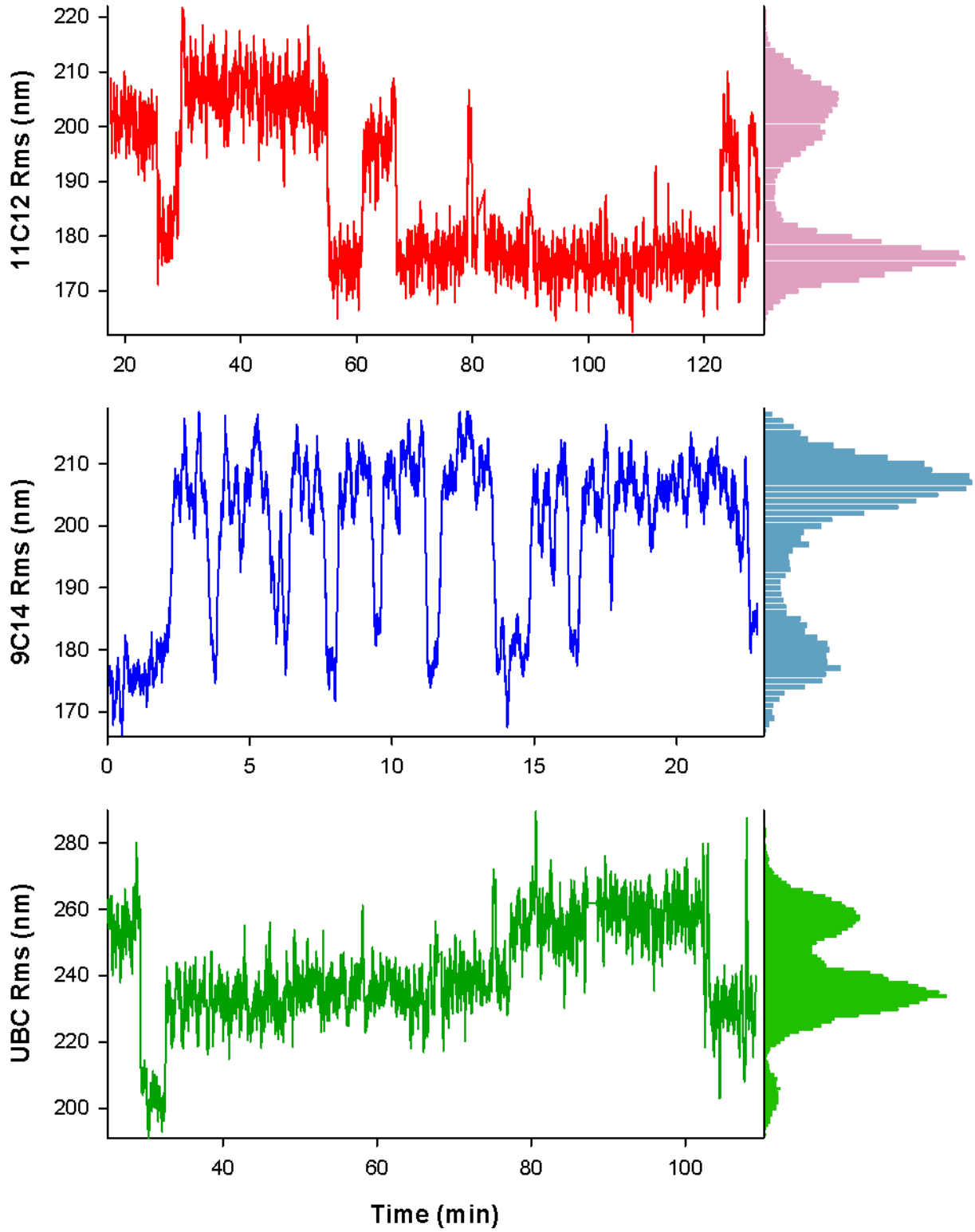


Figure 4.2 – Tethered Particle Motion (TPM) on the three Kahn DNAs (Ref). Above, rms time traces are given for 11C12 (red), 9C14 (blue) and Unbent Control (green) in the presence of LacI. The corresponding histograms are plotted to the right of each trace. It is apparent that transitions occur between well-defined rms levels for all DNAs shown above. When LacI is not present, such multi-level behavior is absent (data not shown). Transitions between rms levels for all DNAs measured occur on a timescale of minutes.

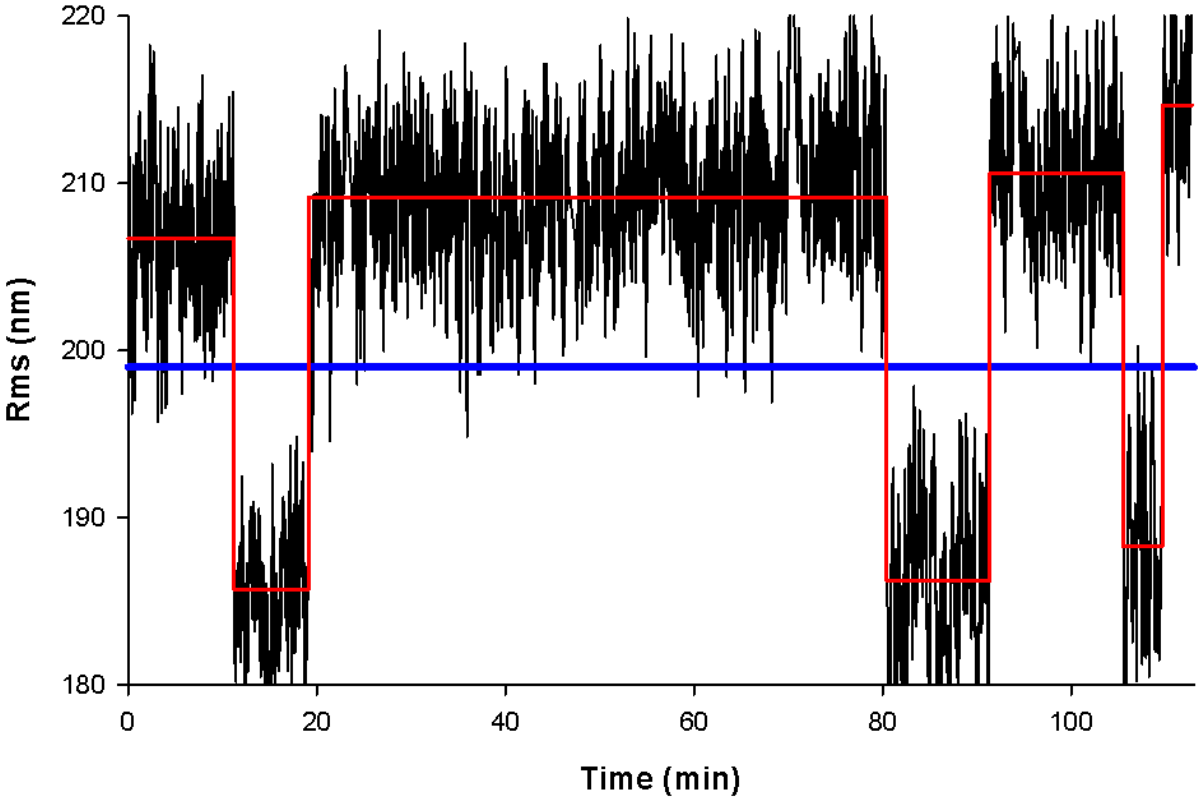


Figure 4.3 – Computing individual looped and unlooped lifetimes in single bead traces. The rms trace above (black lines) is for the same bead as was used in figure 3.9 in chapter III. The DNA construct used in the corresponding TPM experiment and analyzed in the above trace is 9C14, a curved sequence. It is readily seen that well-defined upper and lower levels of rms exist in this data. The higher level is taken to be the unlooped state, while the lower is taken as looped DNA-LacI complex. Any events less than half the filtering window of 10 seconds are excluded. In order to automate the state lifetime determinations, a threshold for looped vs. unlooped states (thick blue line) is determined for a given rms trace. The automatically determined levels are overlaid above the rms trace in red, with the red lines being drawn at the average rms values in each level.

For all three DNA traces it is readily visible that loops form and break down on a timescale of minutes. This is in marked contrast to results predicted from bulk assays. For instance, 11C12 was predicted at have a looped state half-life, $T_{1/2}$, of 36.4 hours. Taking the ratio of this timescale to the 0.14 seconds time predicted for competitor to find substrate DNA, one reaches an approximately one million-fold difference in looped and unlooped state dwell

times in bulk assays. However, in single-molecule experiments this ratio of times is computed to be on the order of unity. The exact looped and unlooped lifetime distributions are given later in this chapter.

The previous chapter detailed the procedures of processing and filtering bead positional data to obtain smoothed rms traces. After this point, it remains to precisely measure the looped and unlooped state dwell times. To accomplish this, a simple threshold-crossing method was utilized (Ref Finzi); this method is detailed in figure 4.3. Briefly, a level is determined above which the rms is assumed to be unlooped and below which it is assumed to be looped. The thresholding level is based on the minimum between the two (or more) Gaussian peaks in rms histograms corresponding to looped and unlooped DNA.

While other existing loop-calling methodologies rely upon the experimenter to distinguish looped states from unlooped states in the "raw" (i.e. unsmoothed) bead rms traces (Ref Chen 2010 (1)) or to ignore the time traces completely and determine relative state populations from binned rms histograms (Ref Han), the strength of the current method is that thresholding smoothed rms both forgoes observer bias in loop-calling while also revealing the fine structure of rms levels in TPM traces.

4.3.2 The spatial resolving power of TPM

As mentioned in the previous chapter, there is a theoretical limit for distinguishing two states in rms histogram populations. For two Gaussian-distributed populations in rms to be distinguishable, the sum of their half-widths at half maximum (HWHM) must be less than the relative distance between their distribution peaks. Thus, $\langle R_2 \rangle - \langle R_1 \rangle \geq HWHM(1) + HWHM(2) = \sqrt{2ln(2)}(\sigma_1 + \sigma_2) = 1.1774(\sigma_1 + \sigma_2)$. For TPM experiments, σ_{RMS} was found to vary inversely with the filtering window. For the 10 second filtering window utilized throughout

these experiments, σ_{RMS} falls in the range of 5-7 nm, giving an approximate resolution in rms of 15 nm.

4.3.3 The temporal resolving power of TPM

There are four characteristic timescales that are present in a TPM measurement. All of them are important. The first is the observation time. Test simulations (data not given) were run in which the looping process was approximated as a Poisson decay with characteristic time τ. It was found that if the observation time was brought down to a low integer multiple of the decay time, the data fitting would yield erroneous results.

The second characteristic time is the filtering window, W, which is held constant at 10 seconds over the course of the single molecule experiments on curved DNAs. State lifetimes (looped or unlooped) found to be less than half the value of the filtering window (5 seconds) were discarded. In the above-mentioned test simulations, it was found that if the ratio of the characteristic time, τ, to the filtering window, W, was less than a factor of two that the fits to lifetime data again yielded erroneous results.

Thirdly, the Butterworth filter applied to the x- and y- positional data in order to remove drift has a high-pass cutoff frequency of 0.05 Hz. This gives a characteristic allowable event time of 20 sec, and any process occurring slower than this is filtered out.

Finally, the microscope camera's frame rate at which images are acquired, a frequency that falls in the range of 20-40 Hz, gives the lower limit of observation. Although these short times of 25-50 milliseconds do not manifest in lifetime data, they are important in the computation of the autocorrelation time of the normalized x- y- positional vector (described in the previous chapter). To accurately measure this autocorrelation time, which normally is falls in the range of 50-75 msec, requires that bead image data must be acquired at a minimum of 20 Hz.

4.4 Obtaining statistical state dwell times from TPM data

4.4.1 Measured distributions of looped and unlooped states

After data thresholding is performed for all beads of a given DNA construct, the corresponding looped or unlooped state lifetimes are sorted temporally and plotted as a cumulative probability distribution. The distributions of looped and unlooped lifetimes for all three Kahn sequences are plotted in figure 4.4. Similar timescales are observed, as mentioned above, for looped and unlooped state distributions.

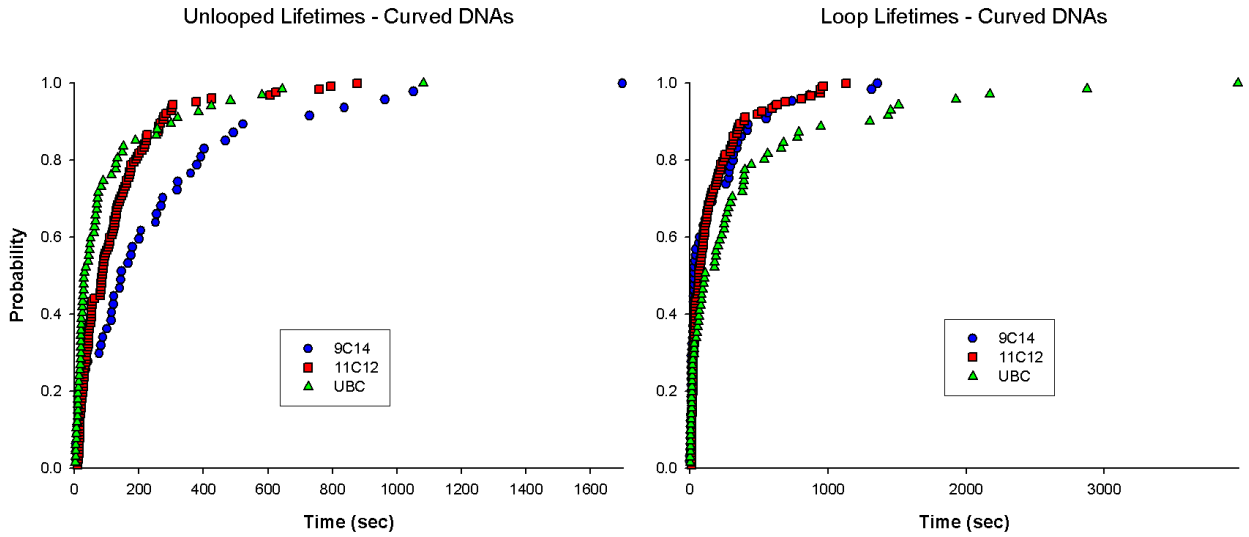


Figure 4.4 – Cumulative distributions of unlooped (left pane) and looped (right pane) state lifetimes. In both unlooped and looped distributions, 9C14 data is given as blue circles, 11C12 as red squares and UBC as green triangles (matching the color scheme from figure 4.2). Fitting of these data to single and double exponential functions is detailed below.

Before fitting any of this data to any functional forms, several observations can be made. First, 9C14 was predicted (Ref Lillian, Ref Mehta) to form a different LacI-DNA binding topoisomer than that for 11C12. This distinct binding orientation for 9C14 was predicted from the bulk DNA competitive resistances to increase the loop formation rate, compared to 11C12, by a factor of, $T_{1/2}(11C12)/T_{1/2}(9C14) = 36.4/21.7 = 1.67$. In single molecule experiments, the ratio of the average of unlooped state lifetimes, in seconds, for 9C14 in comparison to 11C12

was calculated to be, $\langle t_{unlooped}\rangle_{9C14}/\langle t_{unlooped}\rangle_{11C12} = 263.55/160.17 = 1.65$. This is in very good agreement with the bulk assay predictions. The longer unlooped time of 9C14 in relation to 11C12 is readily observable by inspection of the above unlooped lifetime cumulative distributions as a lower relative probability of loop formation for 9C14 at late times (left pane in figure 4.4).

Second, the looped state distributions for 9C14 and 11C12 (right pane in figure 4.4) overlay each other, and do not show very significant differences in their respective lifetimes ($\langle t_{looped}\rangle_{9C14}/\langle t_{looped}\rangle_{11C12} = 177.74/160.17 = 1.1$, or roughly a 10% difference in average looped states lifetime values, given in seconds).

Lastly, all averages and standard deviations of lifetime data, for all DNAs measured, are found to be on the order of minutes. Therefore, the mean observation time of several hours is sufficient to acquire looped and unlooped state lifetime data with minimal negligible probability of missing long-time rms events.

4.4.2 Fitting the looped and unlooped lifetime distributions to exponential functions

The looped and unlooped state distributions are well described as a Poisson decay process or sum of different Poisson decay processes. These obey single or double exponential distributions, given in equations 4.2 and 4.3 below as,

$$P_1(t;\tau) = 1 - exp\left(-t/\tau\right) \tag{4.2}$$

$$P_2(t;C,\tau_1,\tau_2) = C\left(1 - exp\left(-t/\tau_1\right)\right) + (1-C)\left(1 - exp\left(-t/\tau_2\right)\right) \tag{4.3}$$

where τ, C, τ_1 and τ_2 are fitting parameters. Because of the temporal constraints imposed by running window average, $W = 10\ seconds$, lifetimes shorter than half the window size have been excluded. A correction is thus necessary for events that are missed. This is carried out by

returning to the probability density function for a single exponential, $\mathcal{P}_1(t;\tau) = \tau^{-1}exp(t/\tau)$, and instead of integrating from zero to infinity, the integral is performed for a set of limits defining the minimal and maximal observed event times, t_{Short} and t_{Long}.

For almost all experiments performed, t_{Long} was much larger (several hours) than the looped and unlooped lifetimes (which lasted on the order of minutes), so it is taken to be infinite. However, the minimal time, t_{Short}, was observed to be significant for short-timescale loop formation events. Therefore, the single-exponential probability distribution was integrated and normalized to obtain,

$$P_1'(t;\tau,t_{Short}) = 1 - exp(-(t - t_{Short})/\tau). \quad (4.4)$$

A similar process of integration and normalization can be performed for the double-exponential probability density function, $\mathcal{P}_2(t;C,\tau_1,\tau_2) = C\mathcal{P}_1(t;\tau_1) + (1-C)\mathcal{P}_1(t;\tau_2)$, to obtain,

$$P_2'(t;C',\tau_1,\tau_2,t_{Short}) = \left(C\left(e^{t_{Short}/\tau_1} - e^{-t/\tau_1}\right) + (1-C)\left(e^{t_{Short}/\tau_2} - e^{-t/\tau_2}\right)\right) \times \left(Ce^{t_{Short}/\tau_1} + (1-C)e^{t_{Short}/\tau_2}\right)^{-1}. \quad (4.5)$$

Equation 4.5 can be simplified to,

$$P_2'(t;C',\tau_1,\tau_2,t_{Short}) = C'\left(1 - e^{-(t-t_{Short})/\tau_1}\right) + (1-C')\left(1 - e^{-(t-t_{Short})/\tau_2}\right) \quad (4.6)$$

where C' and C are related by,

$$\frac{C'}{1-C'} = \left(\frac{C}{1-C}\right)\exp\left(\left(\frac{1}{\tau_1} - \frac{1}{\tau_2}\right)t_{Short}\right). \quad (4.7)$$

From this point onward, equations 4.4 and 4.6 are referred to as "modified" exponential distributions, while equations 4.2 and 4.3 are called the "unmodified" distributions.

It is important to note that t_{Short} is not a fitting parameter. In each data set, the minimum measured lifetime, $min(t_{State})$, of each rms level was taken to be the value of t_{Short} for data

fitting in those state lifetimes; thus, different values of t_{Short} were used for fitting looped and unlooped distributions in a single DNA construct.

4.4.3 Results of fits to looped and unlooped lifetime data

With the unlooped and looped state TPM data plotted in distributions, it is now possible to fit single and double exponential distributions, both modified and unmodified, to the lifetimes. For all of the DNA constructs used in TPM, it is found that modified double-exponential functions give the best fits, as measured by minimization of the norm of the residuals, to both the unlooped as well as the looped state lifetime data. The fits of modified double exponential functions to both the looped and unlooped lifetimes for all DNAs are summarized in table 4.2.

DNA Construct	Unlooped (U) State Lifetimes (sec, τ_1, τ_2), with populations, C, in parentheses	Looped (L) State Lifetimes (sec, τ_1, τ_2) , with populations in parentheses	Number of data points for U & L
11C12	9.4 ± 1.2 (0.310 ± 0.045), 126.4 ± 1.8 (0.690)	13.99 ± 0.61 (0.535 ± 0.019), 199.2 ± 4.6 (0.465)	127 U, 124 L
9C14	4.3 ± 1.3 (0.279 ± 0.074), 266.0 ± 5.7 (0.721)	18.41 ± 0.87 (0.560 ± 0.016), 344 ± 18 (0.440)	47 U, 65 L
UBC	24.8 ± 1.2 (0.657 ± 0.025), 222 ± 21 (0.343)	17.0 ± 1.3 (0.328 ± 0.012), 429 ± 15 (0.672)	67 U, 71 L

Table 4.2 – Results of fitting the modified double exponential distribution function (equation 4.7) to unlooped (U) and looped (L) state lifetime data. Column 1 gives the DNA constructs measured, while columns three and four give, respectively, the best fit parameters for U and L data from figure 4.4 plus or minus the standard fitting errors. The fourth column gives the number of data points used to obtain the U & L fits for each DNA.

4.4.4 Interpreting fits to lifetime data

While these cumulative probability distribution functions give good fits to the looped and unlooped lifetime data, it remains to show why a double exponential should be required for both

types of data. For the looped state lifetimes, the fits give two time constants, τ_1 and τ_2, that differ by more than an order of magnitude for all DNAs observed. There are several possible explanations for the multiple-exponential behavior in the looped lifetime data.

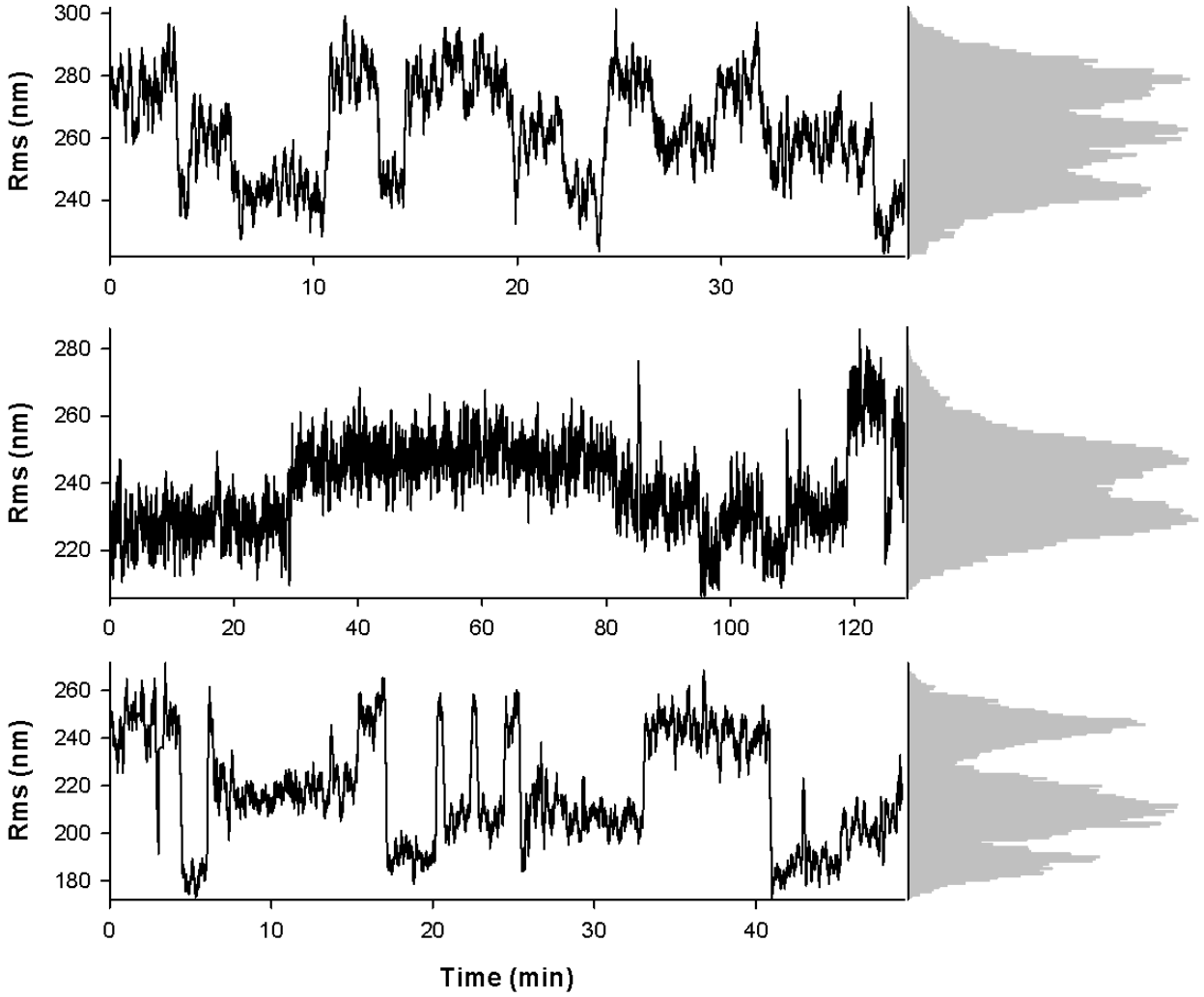


Figure 4.5 – Example TPM traces of UBC rms. All three rms traces (black lines) were from different beads of UBC data. Their corresponding rms histograms are plotted horizontally to the right of each trace (gray bars). It is clear that multiple rms levels exist in the UBC data plotted; these are also visible in the UBC trace of figure 4.4. The histograms for these rms traces show more than two peaks present, and transitions between looped rms levels are observed.

First it is possible that different topologies of binding for the looped operator-LacI-operator complex can give rise to states with differing energetic stabilities. Many experimental

(Ref Haeusler, Ref Mehta) and theoretical (Ref Goyal, Ref Lillian) studies have sought to observe and characterize the energetic effects of these different DNA binding topologies. While it cannot be ruled out that different topoisomers of DNA loops exist in the DNAs measured, it is unlikely that these could give rise to significantly different rms levels observed in the UBC construct. In figure 4.5, several example traces from UBC are given showing multiple rms levels below the upper, unlooped level that interconvert between each other; this interconversion is also visible in the UBC trace with its corresponding 3-peak histogram in figure 4.4 (green lines). If different looped rms levels corresponded to different binding topoisomers, then there should be no interconversion between these levels. The fact that the different levels observed interconvert between each other suggests that the multiple states in the Kahn DNAs are not due to binding topologies.

It should be noted that for 9C14 and 11C12, only two rms levels, corresponding to looped and unlooped LacI:DNA complex, were observed. This is not the case for other DNAs both used in the Meiners lab and elsewhere (Ref Han, Ref Rutkauskas). In fact, there is literature evidence (Ref Wong) that by simply varying the helical phasing between the two operator binding sites, two-level or three-level rms traces can be observed.

In addition to binding topoisomers, a second candidate to explain the multiple-level behavior of rms traces is by conformational changes in the protein. The "hinge" region of LacI at its C-terminus is a 4-helix bundle of non-covalently bound α-helices. Thus, in principle LacI could change conformation to a more open form. As discussed in chapter II and drawn in figure 2.3, there is good experimental evidence of this protein conformational change (Ref Rutkauskas, Ref Taraban). Experiments by Rutkauskas and coauthors (Ref) found that while "locking" the LacI protein in its canonical V-shaped form, which should prevent LacI conformational shifts,

did alter the population distributions in rms histogram in TPM experiments, this did not destroy the multiple-level (1 unlooped level plus multiple looped levels) nature of the rms traces in single molecule experiments. Other possible explanations for this multiple-level behavior are discussed in detail in a later chapter.

4.5 A Kinetic Model of Loop Formation

4.5.1 Keeping things simple

While the bi-exponential nature of the probability distribution of looped state lifetimes is easily understood as arising from different species of DNA loops, with several possible candidates, it is not immediately clear why two exponential terms are required to fit the unlooped state lifetimes. The simplest possible model with three kinetic parameters, analogous to the three fit parameters in the bi-exponential distribution, is given below in figure 4.6, and explained further in figure 4.7. This simple model was generated by Chen and coauthors (Ref 2010 (1)) to describe the loop formation process. A schematic of the model applied to TPM experiments is given in chapter III, figure 3.1. There is good literature evidence (Ref Han) that at the *in vitro* LacI concentrations used in the present experiments of 100-200 picomolar, there should be significant double-occupancy of the operators on a given DNA substrate. This double-occupancy would register as an additional, "dark" unlooped state in TPM experiments, designated as U_2.

$$U_2 \underset{k_-}{\overset{k_+}{\rightleftarrows}} U_1 \xrightarrow{k_L} L$$

Figure 4.6 – A simple model of unlooped kinetics. Here, two unlooped states are proposed. The first state, U_1, is a loop-ready state in which loop (L) formation, at rate k_L, can proceed. The second, U_2, is a "dark" state in which loop formation cannot proceed. The two unlooped states are interconvertible at rates, k_+, k_-.

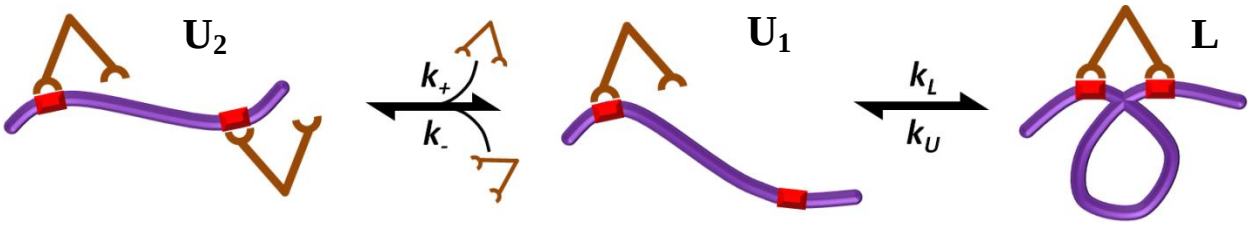


Figure 4.7 – A schematic of unlooped kinetics described by the kinetic model. LacI (brown) in the singly-occupied unlooped state, U_1, is bound to an operator site (red rectangle) along a DNA substrate (purple). In this stage LacI may form loops, L, at rate kL. However, before a loop can be formed in U_1 a second protein may bind to the other operator site (red) at rate k_- and send the DNA into the "dark" unlooped state, U_2, in there is no loop formation. Once one of the LacI proteins leaves its operator site, at rate k_+, the DNA returns to the loop-ready state and may form loops.

4.5.2 Solving the kinetic model

Although this kinetic model and its numerical solution method have been described previously (Ref Chen 2010 (1), Ref Chen 2010 (2)), here a novel analytic solution is presented in which the kinetic rates (k_+, k_-, k_L) are given exactly in terms of the bi-exponential fit parameters (C, τ_1, τ_2). This solution is derived below.

In order to obtain a probability distribution from the kinetic model, the one must start with the first-order kinetic decay rates of the populations of the two unlooped states. These kinetic equations obey

$$\dot{U}_1(t) = -(k_L + k_-)U_1(t) + k_+U_2(t) \tag{4.8}$$

$$\dot{U}_2(t) = k_-U_1(t) - k_+U_2(t). \tag{4.9}$$

It is noted that one term is neglected from equation 4.8. This term, which is equal to $k_L L(t)$, is ignored because of the manner in which unlooped lifetimes are measured; namely unlooped populations are distinguishable from looped populations and the breakdown of loops back into the unlooped state population does not register in considering the decay of individual unlooped

events. On the other hand, different species of unlooped populations are indistinguishable, so terms with U_2 appear in the decay equation for U_1 population and vice versa. Equations 4.8 and 4.9 can be compiled into a matrix of first-order coupled differential equations,

$$\vec{U}'(t) = \begin{pmatrix} \dot{U}_1 \\ \dot{U}_2 \end{pmatrix} = \begin{bmatrix} -(k_- + k_L) & k_+ \\ k_- & -k_+ \end{bmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \overline{\overline{K}}\vec{U}(t). \tag{4.10}$$

The corresponding eigenvalues of the matrix, $\overline{\overline{K}}$, are found to be,

$$\lambda_{\pm}[K] = -\frac{k_+ + k_- + k_L}{2} \pm \sqrt{\left(\frac{k_+ + k_- + k_L}{2}\right)^2 - k_+ k_L}. \tag{4.11}$$

From these eigenvalues, the state lifetimes of U_1 and U_2 may be obtained as $\tau_1 = -1/\lambda_+$ and $\tau_2 = -1/\lambda_-$. The assignment of τ's to λ's is later found to be interchangeable. Starting from the assumption that at time zero, all of the unlooped population is in the loop-ready state, U_1,

$$\vec{U}(0) = \begin{pmatrix} U_1(0) \\ U_2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{4.12}$$

a solution for $\vec{U}(t)$ can be obtained. For this solution, the total population in unlooped states, $P_U(t) = U_1(t) + U_2(t)$, may be written in a similar functional form as equation 4.3 (above),

$$P_U(t) = Ce^{\lambda_+ t} + (1 - C)e^{\lambda_- t} = Ce^{-t/\tau_1} + (1 - C)e^{-t/\tau_2}, \tag{4.13}$$

for the distribution of unlooped state lifetimes. The relative population in the first state above, C, is found to be,

$$C = \frac{k_L + \lambda_-}{\lambda_- - \lambda_+} = \frac{k_L - 1/\tau_2}{1/\tau_1 - 1/\tau_2}. \tag{4.14}$$

From this formula for C, the kinetic rate of loop formation, k_L, may be solved for as

$$k_L = \frac{C}{\tau_1} + \frac{(1 - C)}{\tau_2}. \tag{4.15}$$

Using the relations from the $\overline{\overline{K}}$ matrix eigenvalue solutions, $1/(\tau_1\tau_2) = \lambda_+\lambda_- = k_L k_+$, and, $1/\tau_1 + 1/\tau_2 = -(\lambda_+ + \lambda_-) = k_+ + k_- + k_L$, the other two kinetic rates are solved for as,

$$k_+ = \frac{1}{k_L\tau_1\tau_2}, \tag{4.16}$$

$$k_- = \frac{(1-C)}{\tau_1} + \frac{C}{\tau_2} - k_+. \tag{4.17}$$

4.5.3 Loop formation kinetics in the Kahn DNAs

Equations 4.15, 4.16 and 4.17 comprise an exact solution to the kinetic model of loop formation given in figure 4.6. The double-exponential fit parameters to unlooped lifetimes for 9C14, 11C12 & UBC may be plugged into these equations to obtain kinetic rates of transition for different unlooped state populations. The results of these computations are given in table 4.3.

Kinetic Rates	**11C12**	**9C14**	**UBC**
$\mathbf{1000 \times k_L}$ **(sec^{-1})**	38.4 ± 8.4	68 ± 37	28.0 ± 2.3
$\mathbf{1000 \times k_+}$ **(sec^{-1})**	21.9 ± 3.3	13.0 ± 4.0	6.48 ± 0.54
$\mathbf{1000 \times k_-}$ **(sec^{-1})**	54 ± 11	157 ± 64	10.3 ± 1.6
$\boldsymbol{\tau_L = 1/k_L}$**, U→L (sec)**	26.0 ± 5.7	14.7 ± 7.8	35.7 ± 2.9
$\boldsymbol{\tau_+ = 1/k_+}$**, 2→1 (sec)**	45.6 ± 6.8	77 ± 24	154 ± 13
$\boldsymbol{\tau_- = 1/k_-}$**, 1→2 (sec)**	18.5 ± 3.6	6.4 ± 2.6	97 ± 15

Table 4.3 – Kinetic rates in the loop formation model. The solutions of the kinetic model for the three Kahn DNAs (11C12, 9C14 & UBC) are given above in two forms. The first three rows of data give the kinetic rates in sec^{-1}. The last three lines give the inverse of the kinetic rates, which can be thought of as state lifetimes or average times for transition from a given state to a different one.

In table 4.3, it is immediately visible that the loop formation rate has a minimum for unbent control DNA. This agrees well with the design principles of the pre-bent DNAs, 9C14 and 11C12; chiefly, that the A-Tract bends reduce the thermodynamic barrier cost of forming loops. Furthermore, the characteristic loop formation time for all DNAs measured is in the range 14-36 seconds. This unlooped state lifetime range for curved DNAs 9C14 and 11C12 is several orders of magnitude larger than the expected time from the theoretical calculation performed in equation 4.1 above.

Kinetic Rates	**UBC**	**158 O_{SYM}**	**153 O_{SYM}**
C (relative population of 1st exponential)	0.657 ± 0.025	0.81 ± 0.05	0.38 ± 0.02
1st unlooped exp. time, τ_1 (sec)	24.8 ± 1.2	27 ± 1	16 ± 2
2nd unlooped exp. time, τ_2 (sec)	222 ± 21	97 ± 21	193 ± 19
$\tau_U = 1/k_L$, U→L (sec)	**35.7 ± 2.9**	**31.3 ± 2.9**	**37.1 ± 6.1**
$\tau_+ = 1/k_+$, 2→1 (sec)	154 ± 13	83.7 ± 8.3	83.3 ± 16.1
$\tau_- = 1/k_-$, 1→2 (sec)	97 ± 15	291 ± 143	34.8 ± 7.5

Table 4.4 – Kinetic rates for literature TPM data. Wong and coauthors (Ref Wong) used DNAs, designated 158 O_{SYM} and 153 O_{SYM} (so named because of their respective inter-operator distances), with inter-operator distances and operator binding strengths comparable to those found in the UBC construct (column 1). A table in the supporting information in this reference contained the bi-exponential fits to the unlooped lifetimes for these DNAs; they are listed in the first three rows of columns 2 & 3. For comparison purposes, their fit parameters were fed into the kinetic model solution and the corresponding loop formation rates compared to UBC were within the experimental errors of each other (bold). The solutions for kinetic rates are given in rows 4, 5 & 6.

To test the validity of the kinetic model, a literature search was performed and a series of TPM experiments with LacI by Wong and coauthors (Ref) was found in which the substrate

DNAs used were comparable to UBC both in their operator binding strengths (O_{SYM}'s) and in their inter-operator distances (158 bp and 153 bp for the pair of DNA constructs). These two literature DNA constructs, designated 158 O_{SYM} and 153 O_{SYM} (named for their respective inter-operator distances), were found to require double-exponential fits, the fit results of which Wong and coauthors listed in a table in their supporting information. These fit results for C, τ_1 and τ_2 were input into the kinetic model solutions (equation 4.15, 4.16 & 4.17) and kinetic rates were obtained; these results are listed in table 4.4. It is readily seen that the loop formation rates for all three DNAs are comparable. Thus, the kinetic model is validated independently through the literature.

4.6 Possible explanations

4.6.1 Does bead size matter?

There is a vast difference between the timescales observed in bulk and single molecule experiments with LacI. Several possible explanations exist for these curious results, and they must be addressed.

First it is possible that the presence of the large (diameter 0.54 micrometers) polystyrene bead affects the thermodynamics of the LacI-DNA system in a way that destabilizes the DNA loops. Chen and coauthors (Ref 2010 (1)) found that forces as low as 60 femtonewtons affected the loop formation rate. Given that the entropic force from the microscope coverslip on the bead for the contour lengths used is approximately 50 femtonewtons (Ref Chen 2009), falling in the range of loop-formation affecting forces, it was possible that the presence of the bead itself put an extra, entropic tension on the DNA substrate. To test this possibility, Milstein and coauthors (Ref 2010) performed TPM on a DNA construct attached, alternatively, to either a large

polystyrene microsphere of diameter 0.8 micrometers or a small gold nanoparticle of approximate diameter 30 nanometers. The resulting loop formation changed only by a factor of order unity. Thus, the experimental setup for TPM was found not to significantly affect looping kinetics.

4.6.2 What about the J-factors?

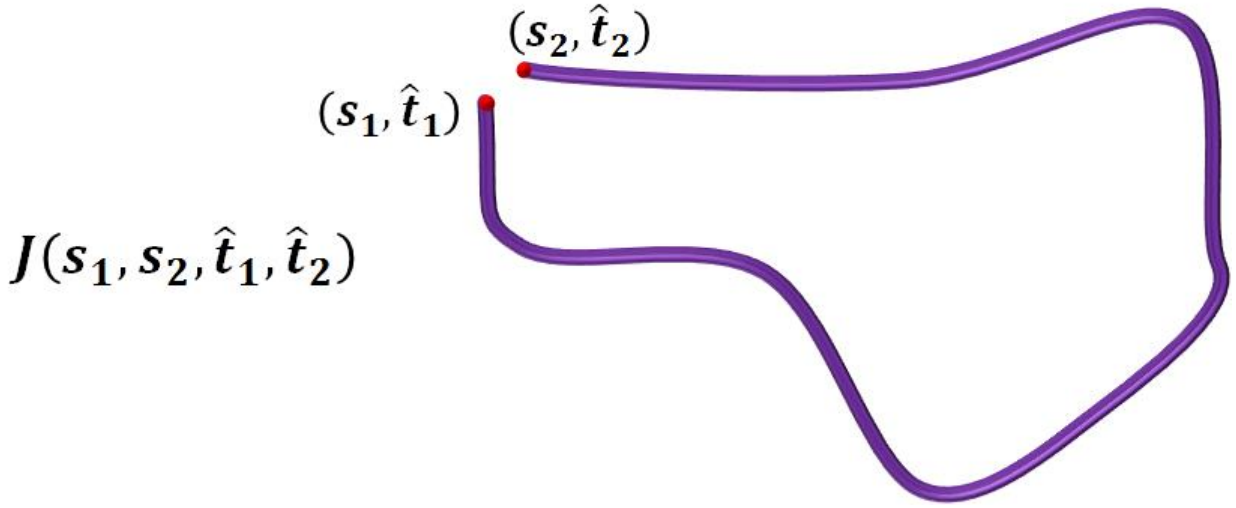


Figure 4.8 – A schematic drawing of the idea behind J factors. At a given position along a long polymer, the probability density that, through Brownian motion of the polymer, two points, s_1 and s_2, will meet is given by the function $J(s_1, s_2, \hat{t}_1, \hat{t}_2)$, where $\hat{t}_1$and $\hat{t}_2$are the positional tangential vectors of the polymer at points s_1 and s_2, respectively.

It is also possible that the probability density of one operator of a DNA substrate at the second operator site, called the looping J-factor, is higher for the curved sequences (figure 4.8). The J-factor may be thought of as an effective concentration, and it is usually given in molar units. However, it must be noted that the J-factor also accounts for orientational constraints (Ref Shimada), which in this case would be imposed on the DNA by LacI for a given binding topoisomer. Theoretical computations (Ref Wilson 2010) predict that the intrinsically bent

sequence 9C14 should have a J-factor that is approximately 10^{-5} molar, while random DNA of similar inter-operator length (i.e. UBC) should have a J-factor on the order of 10^{-9} molar. Part of the rationale of the bulk competition assays was that the competitor concentration of 50 nanomolar should dwarf the looping J-factor of straight DNA.

However, the ratio of J-factors for two constructs should be equal to the inverse of the ratio of the loop formation rates. These computations would predict a four-order-of-magnitude difference between the loop formation rates for 9C14 and UBC, which is not observed in TPM experiments.

4.6.3 The next step

The last possible explanation for the observed differences in bulk and single molecule data is that the bulk experimental data is somehow erroneous. In order to proceed with any further experimentation on curved DNA constructs, the bulk DNA competition assays must be confirmed. Specifically, confirmation of results must be seen at the single molecule level to understand the dynamics in play. This is the subject of the next chapter.

Chapter V

Single Molecule Competition Assays I

5.1 Introduction

Biochemical assays in bulk solutions allow a variety of useful chemical, thermodynamic and physical properties to be measured for molecules of interest. However, such methods average over the underlying dynamics of the molecules in question. Single particle paths and interactions necessitate different techniques. To observe individual events and to pry into the intricate molecular kinetics occurring on the nanoscale, a methodology is needed to observe single molecules of the desired type and to have a reporting system that yields such properties as interaction times and possible state transitions.

Single molecule experimental techniques allow important questions about interactions of individual molecules to be explored. For the *in vitro* system studied in the present work, the single molecule technique Tethered Particle Motion (TPM) makes possible the observation of individual loop formation and breakdown events. In this chapter, the TPM technique will be amended to include an extra influx of solution after LacI protein has been added. Specifically, the addition of a large excess concentration of competitor DNA transforms the simple TPM experiment into a Single Molecule Competition Assay (SMCA). With this experimental tool, the bulk results of LacI-DNA competitive resistance may be verified at the single molecule level.

5.2 Single molecule competition with short oligonucleotides

5.2.1 Experimental Methods

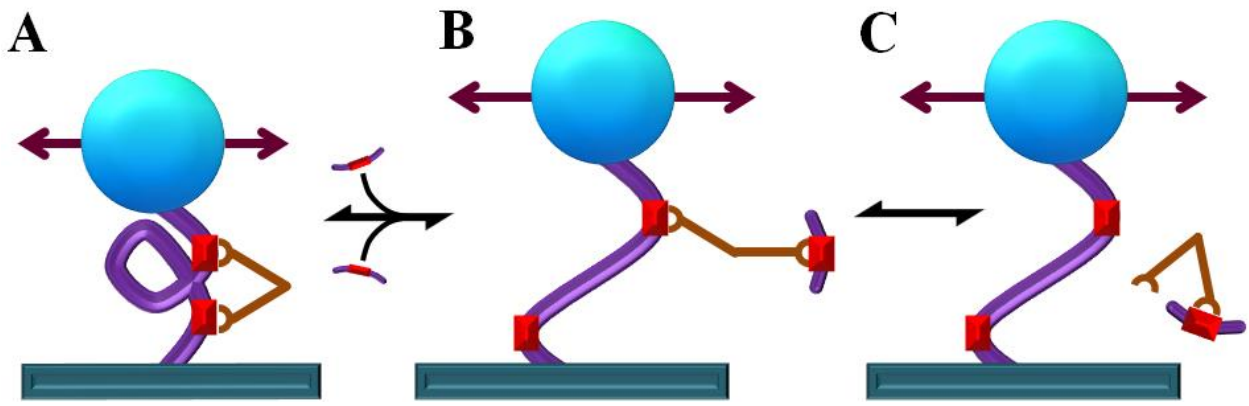


Figure 5.1 – A schematic of the single molecule competition assay (SMCA) with 40 bp oligonucleotides. Here, LacI (brown) is incubated with DNA (purple), during which time loops reversibly form (A) and breakdown on the substrate DNA, with operators given in red. After 30 minutes' incubation time, a 100 nanomolar solution of competitor DNA oligos (40 bp, short purple lines with red rectangles for the operators) is flowed into the cell. Upon loop breakdown (B), LacI was thought to quickly form a sandwich complex. At this juncture there is equal probability for LacI to unbind from its competitor and (A) reform loops on its tethered substrate or (C) for it to be competed off.

To perform a single-molecule competition assay (SMCA), the experimenter would begin by setting up a TPM experiment as described in chapter III. 9C14 was the substrate DNA utilized in these measurements. Following a modified version of the protocol of Mehta and Kahn (Ref), the flow cell would be incubated with LBB+ solution with LacI, at a concentration of around 100 picomolar, for approximately 30 minutes. After this time, a 100 nanomolar (note that Kahn and coauthors used 50 nM of double operator DNA for competitor; this is the rationale behind the current, and later, final concentrations of competitor used inside the flow cells in SMCA experiments) concentration of competitor DNA (unlabeled – that is, no chemical modifications were present in the competitor DNA) in LBB+ solution was flowed into the chamber. Afterward, flow cell tubes (through which solutions are flowed into the cell) were taped shut with Parafilm

to prevent drying of the chamber contents. Image data on the camera microscope was continuously acquired throughout this process. Normally, data acquisition would last for at least 2-3 hours after competitor was added to the flow cell. The principle of the single molecule competition experiments is diagrammed in figure 5.1.

In order to have a simplified test of the bulk enzyme mobility shift assay (EMSA) that measured competitive resistance time, a short oligonucleotide was utilized (abbreviated 'oligo' from this point onward) as the competitor DNA. This competitor DNA oligo was ordered as a pair of 40 base pair (bp) primer strands from Invitrogen^{TM}. The oligo contains the 20 bp symmetric/ideal operator flanking on either side by 10 bp of random DNA. The exact sequence of the 40 bp oligonucleotide is given in appendix B.

5.2.2 Initial results

Examples rms traces of 9C14 in the presence of 100 nM of 40 bp oligonucleotide competitor molecules are shown in figure 5.2. It is immediately apparent that LacI continues to form and breakdown loops on its 9C14 DNA substrate even in the presence of a large excess of competitor. The supposedly unbound LacI DNA-binding domain (DBD) of LacI in the unlooped state apparently does not interact with the competitor molecules in solution.

Thus, the Single molecule test of the bulk DNA competition EMSA has very perplexing results. First, it is confirmed that the 9C14-LacI complex indeed resist competition for many hours. However, during competition the unlooped state apparently remains unavailable for competitive binding of 40 bp oligonucleotides containing the ideal operator binding sites.

Further analysis is necessary for the kinetics underlying these single molecule DNA competition data to be properly understood. Particularly, a dissection of why the free DBD of

LacI does not “see” the competitor in solution motivates the experimenter to return the rough kinetic calculations of equation 4.1 in chapter IV (diagrammed in figure 5.3). Those approximations predicted that competitor molecules would find a free LacI binding site in less than a second. Here, LacI loops and re-loops on its substrate for many hours of observation.

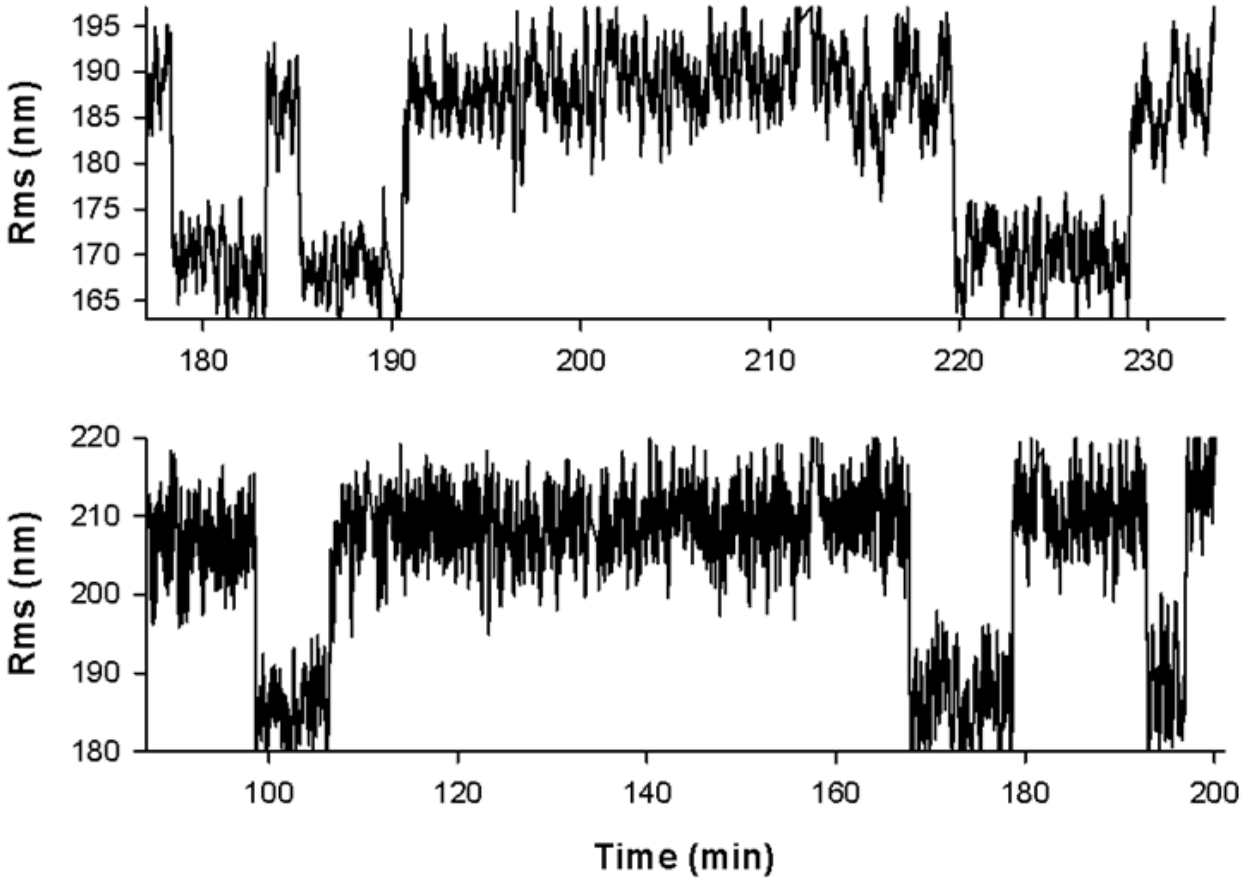


Figure 5.2 – Example rms traces of 9C14 single molecule competition with 40 bp oligonucleotides. Both these rms traces (black lines) originate from different beads in separate 9C14 SMCA data sets. At time zero in each example trace, 100 nM of competitor was flowed into the sample chamber. After almost four hours for the upper trace, and after over three hours for the lower trace, loop formation and breakdown proceeds on its tethered substrate.

It is also observed that the loop formation and breakdown rates do not change significantly in the competition assay. In the above traces, it is observed that looped and unlooped rms states last on the order of minutes. In fact, in the upper trace from around 190-220 minutes there is an unlooped state lasting around 30 minutes, while in the lower trace from

around 110-165 minutes there is an unlooped state that is over 45 minutes long! Such long unlooped states in the presence of competitor defy the interpretations put forth in the bulk assay, and warrant further theoretical consideration.

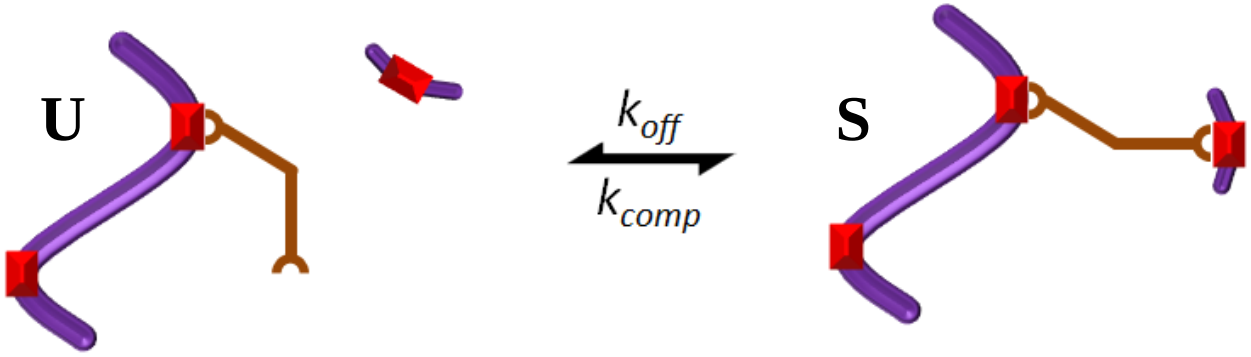


Figure 5.3 – Kinetic scheme of single molecule competition with 40 bp oligonucleotides. The DNA substrate (long purple line) is complexed with LacI (brown) in the unlooped state above (U) at one of its two operator sites (red rectangles). At kinetic rate k_{comp}, a 40 bp competitor DNA oligonucleotide (short purple line, with symmetric operator in red) binds to the free DNA-binding domain of LacI and forms a sandwich (S). In the sandwich complex, LacI may either re-form the LacI-DNA unlooped complex on its substrate (U), at rate k_{off}, or it may be competed off (not shown).

5.3 How much competitor is present in a flow cell?

The kinetic scheme in figure 5.3 must be re-visited. The bulk DNA competition utilized a competitor-to-substrate DNA stoichiometry of 50:1, with 1 nM of substrate and 50 nM of competitor. To fully understand the relative amount of competitor DNA as compared to tethered DNA in a single molecule experiment, an order-of-magnitude calculation is performed as follows.

To approximate the number of DNA tethers present in a flow cell, it is useful to know that the average microscope observation area of a flow cell has at least 10 beads. In a previous chapter, the lateral dimensions of the flow cell were given as $2 \times 20\ mm^2$, or $4 \times 10^{13} nm^2$. The microscope camera has an observable chip area of $1024 \times 1280\ pxl^2$, with each pixel

corresponding to 6725 nm. Accounting for the microscope's magnification of 60 ×, the approximate number of DNA tethers in a chamber may be computed as,

$$N_{Tethers} = \frac{10}{1310720\ pxl^2} \times \left(\frac{1\ pxl}{6725\ nm/60}\right)^2 \times 4 \times 10^{13}\ nm^2 = 2.4 \times 10^4. \quad (5.1)$$

Knowing that there are approximately 24000 tethers in a flow cell, it can now be computed how many competitor molecules are present in solution for every tethered LacI-DNA substrate. In a given experiment, the flow cell volume ranges from 30-70 µL. For an average of 50 µL of competitor DNA solution at a concentration of 100 nM, the number of competitor molecules in an enclosed flow cell is

$$N_{Comp} = 5 \times 10^{-5}\ L \times \frac{10^{-7}\ moles}{1\ L} \times \frac{6.022 \times 10^{23}}{1\ mole} = 3 \times 10^{12}. \quad (5.2)$$

Thus, there are approximately 3 trillion competitor molecules in a flow cell. Now, taking the ratio of the number of competitor DNA molecules to the number of substrate DNA tether molecules, it is found that for every DNA tethered to the microscope slide cover slip, there are approximately 10^8 competitor molecules floating free in solution. This represents a one hundred million fold excess of competitor DNA in a single molecule competition experiment!

5.4 Does the tetrameric repressor protein dimerize?

5.4.1 A hypothesized model of dimerized loop breakdown

Given the curious competition results at the single molecule level, it is prudent to reconsider the canonical model of loop formation and breakdown mediated by LacI. The above results suggest the existence of an unlooped state in which LacI is not available for competition. According to a series of classic biochemical assays (Ref Brenowitz, Ref Fickert, Ref Levandoski, Ref Royer), LacI should be predominantly tetrameric at the picomolar concentrations used in

TPM. However, concerns in the literature about the experimental methods employed in these measurements (Ref Oehler 1999) as well as implied evidence of long dimer-operator dissociation times of approximately one hour (Ref Fickert) motivated the hypothesis of a new model of looping kinetics, including the DNA competition assay, incorporating the possibility of LacI dimerization (figure 5.4). This model is described in 5.5.

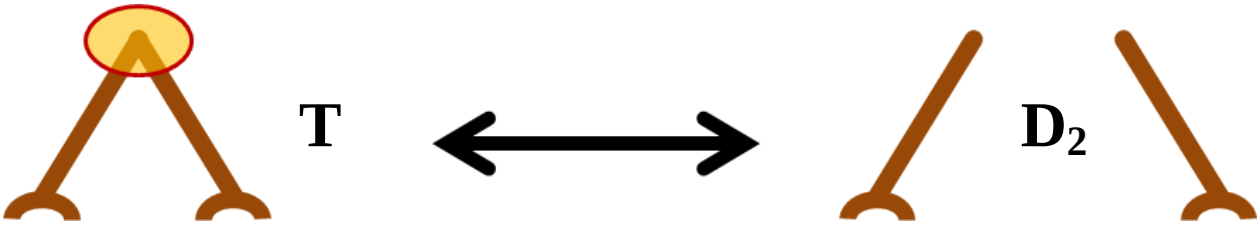


Figure 5.4 – The proposed tetramer-to-dimer breakdown of the lac Repressor protein. LacI (brown) is tetrameric in its native form; specifically, it is a dimer of dimers, with a 4-helix bundle that is thought of as a hinge region where the two dimers non-covalently join together (yellow ellipse). It is hypothesized that at certain concentration levels, the LacI tetramer (T) would reversibly dissociate into a pair of dimers (D_2). The associated loop formation and breakdown kinetics that incorporates these hypothesized interactions are detailed in figure 5.5.

If the predominant pathway for loop breakdown is dissociation of the tetramer into a DNA-bound dimer pair (figure 5.5, A → E), then there would exist a "silent" unlooped state in which neither of the DNA-binding domains (DBD) of LacI are free to bind to competitor. Furthermore, given that the dissociation time dimeric lac from its operator binding site is proposed to be on the order of an hour (Ref Fickert), the competition time would thus depend on the hypothesized small population of tetramer bound to DNA (figure 5.5, B).

These results would explain certain abnormalities observed in single molecule DNA competition assay rms traces. A "silent", long-lived unlooped state corresponding to looped-complex LacI breaking down into a DNA-bound dimer pair would be a plausible hypothesis for the 30- and 45- minute long unlooped states that are plotted in figure 5.2, and which are apparently unavailable for DNA competitive binding.

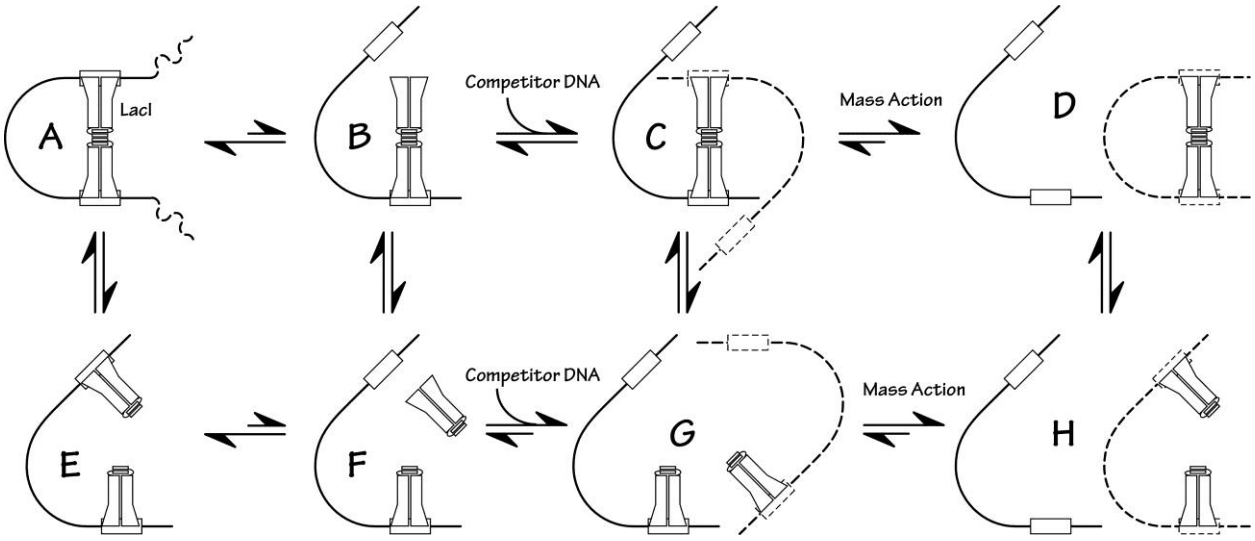


Figure 5.5 – A dimerized model of DNA competition and looping kinetics. This is a modified form of the competition assay schematic in figure 4.1 in the previous chapter. If lacI can dissociate into a dimer pair, then DNA loops (A) can break down into two DNA-bound dimers on its substrate (E). However, there is also a kinetic probability per unit time that one of LacI's DNA-binding domains (DBD) releases its bound operator (B) and is open for competition. Only then can competitor DNA bind substrate DNA in a sandwich complex (C), and from there be competed off (D). States E-H are hypothetical.

5.4.2 Testing the hypothesis

Given the radical claim of this hypothesis, an experimental test must be performed to validate whether or not tetrameric LacI indeed breakdowns into a dimer pair while bound to DNA. Such an experiment is possible with use of a mutant version of the lac Repressor protein. Designated Q231C, this protein was designed and purified by Matthews and Vanzi (Ref Rutkauskas) to replace a glutamine residue near the "top" of the "V" in canonical, tetrameric LacI with a cysteine residue. Two cysteine residues on LacI dimers can then be crosslinked with chemical linkers of variable lengths to "lock" the repressor in the tetrameric, V-shaped conformation (figure 5.6). This cysteine-crosslinked mutant protein was originally synthesized to demonstrate the presence of LacI tetramer opening (the protein conformational change drawn in figure 2.3) in single molecule DNA looping measurements (Ref Rutkauskas).

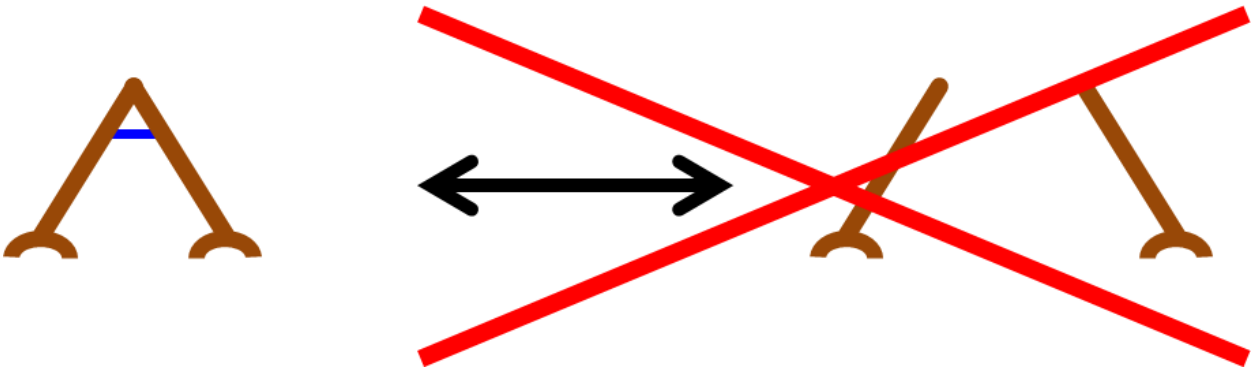


Figure 5.6 – A crosslinked mutant of lac Repressor protein. LacI (brown) is mutated to have a cysteine in place of a glutamine at residue 231 on the monomeric amino acid chain. Upon the chemical addition of a molecular crosslinker (blue line), tetrameric LacI can no longer (red lines) form a dimer pair (borwn lines to the right, above) like the drawing of figure 5.4. Thus, chemical crosslinking allows the exploration of protein-mediated DNA looping where only tetrameric LacI is present.

Single molecule TPM experiments were performed with Q231C, crosslinked with $BM[PEO]_2$, on the intrinsically bent construct 9C14. The TPM samples were prepared according to protocols described in chapter II, with the substitution of crosslinked Q231C (obtained by the generous donation of Francesco Vanzi). Representative rms traces of these TPM experiments are plotted in figure 5.7. The measured lifetime distributions are given in figure 5.8.

It is seen in the example rms traces in figure 5.7 that the rms level behavior and dwells do not change qualitatively, as hypothesized, with the "locking" of the tetramer by chemical crosslinking. In fact, long- and short- lived unlooped and looped states analogous to those measured in the fits to exponential distribution functions are observed in these traces. Overall, there is no significant (i.e. orders-of-magnitude) difference in 9C14 TPM data taken with wild-type LacI or chemically crosslinked mutant Q231C. Although the crosslinked LacI is known to slightly alter rms histograms (figure 5 in Ref Rukauskas), this effect is not large enough to explain 9C14's anomalous DNA competitive resistance. Thus, the hypothesis that loop breakdown must predominantly occur by LacI tetramer-to-dimers dissociation is disproven.

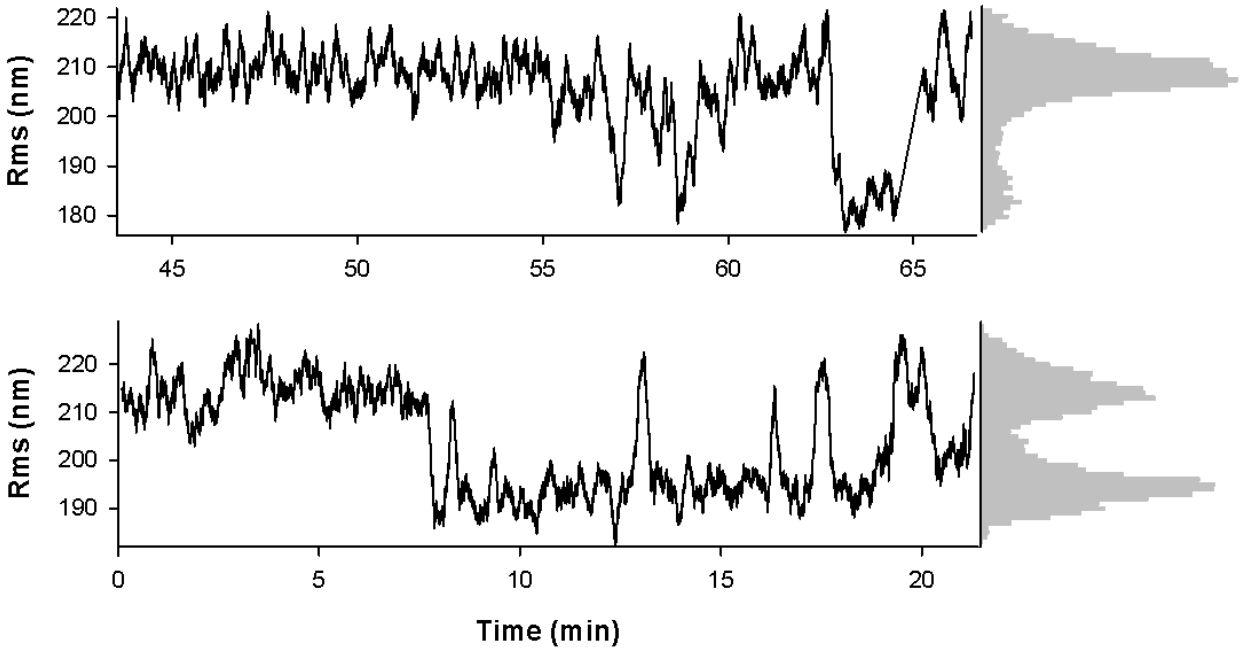


Figure 5.7 – Example rms traces from TPM experiments with the crosslinked mutant of lac Repressor protein (Q231C). Two sample rms traces from 9C14 (black lines), with their associated histograms (gray bars to the right of each trace), reveal that crosslinking of the looping protein does not significantly change the qualitative behavior of loop formation and breakdown kinetics. There are still long and short-lived looped and unlooped states, as seen in the above traces. It is apparent from the TPM data with the crosslinked mutant that LacI does not breakdown into a dimer while bound to DNA.

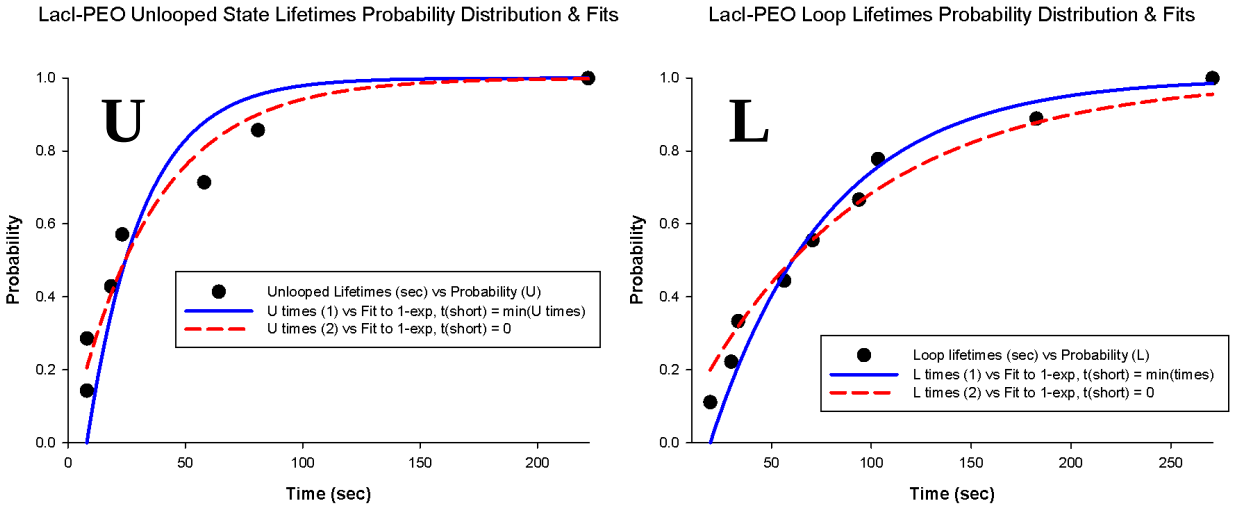


Figure 5.8 – Unlooped (U) and looped (L) lifetime distributions from TPM experiments with mutant LacI crosslinked with BM[PEO]$_2$ (abbreviated PEO). Graph legends are in boxes. Lifetime data points for U & L are black circles. Fits to modified single-exponential distributions are plotted as solid blue lines, while unmodified single-exponential distributions are the dashed red lines. The unmodified functions were better fits for the low number of data points acquired. For U, the fitted τ was 86.8 ± 5.4 sec, while the looped τ was 35.1 ± 3.8 sec.

5.5 A revised competition assay for 9C14

5.5.1 Back to the drawing board

To summarize the results so far, it has been found in bulk assays that intrinsically curved DNAs exhibit a strong resistance to competitive DNA binding, as compared to unbent control. Single molecule experiments both confirm these results and add that LacI reforms loops on its substrate without binding competitor for many hours. There appears to be a "silent" unlooped state in which one LacI DBD is bound to a specific operator site and the other, while not bound to specific DNA, is predominantly unavailable to bind competitor DNA. Finally, the TPM data above revealed that the "silent" unlooped state in not dimeric breakdown of the LacI tetramer.

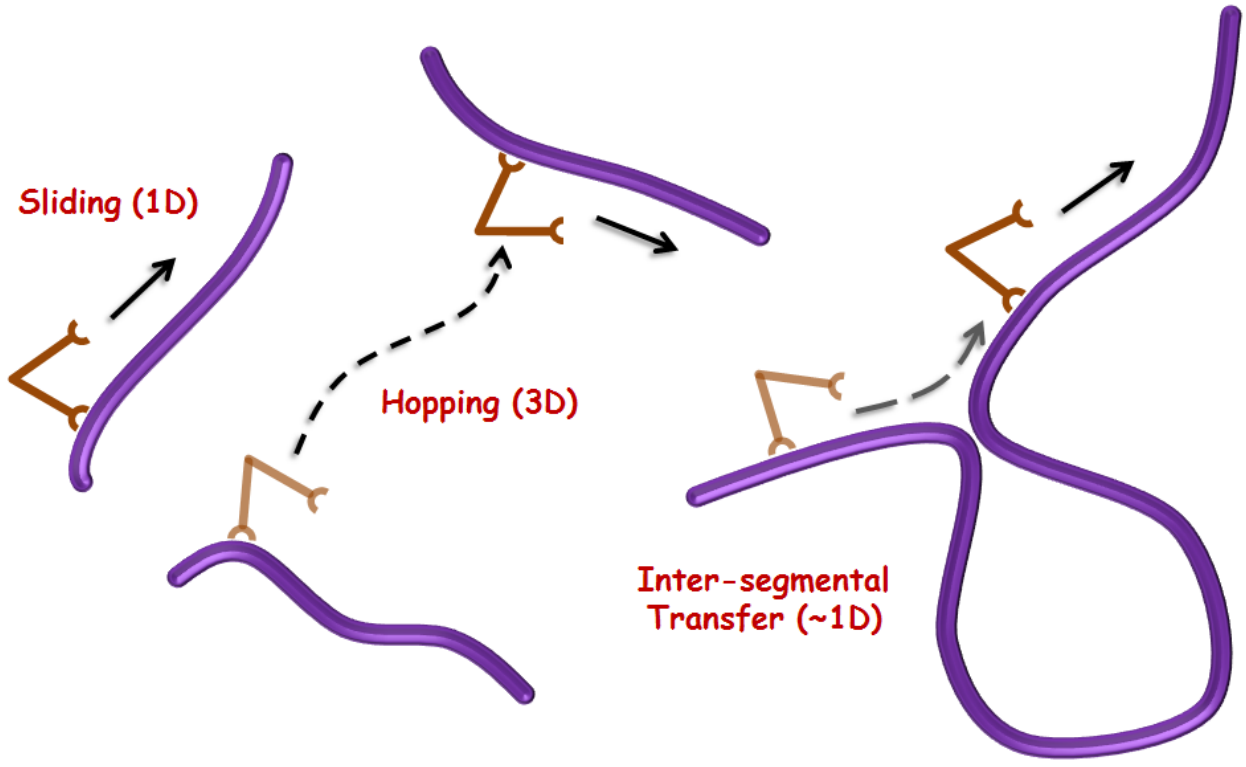


Figure 5.9 – Theoretical methods of target-search by a DNA-binding protein for its binding site on substrate DNA. These methods, proposed for the study of LacI (colored brown above, Ref Berg) searching along its substrate DNA (purple), include 1-dimensional sliding, 3-dimensional micro- and macro- hopping and quasi-1-dimensional inter-segmental transfer. The latter mechanism is facilitated by enzymes such as LacI (ref Hammar), GalR (Ref Qian), or the lambda repressor (Ref Manzo) that bring distant sites along a DNA contour close to each through a DNA loop. These 1-dimensional processes allow LacI to search faster than simple 3-dimensional diffusion.

It is appropriate to reconsider the kinetics of the DNA competition experiment in light of the above developments. The competitor DNA used was only 40 base pairs long, containing a single ideal *lac* operator (O_{SYM}) with 10 bp of random sequence DNA flanking either side. Perhaps this competitor DNA is insufficient relative to its substrate for the lac Repressor protein to appropriately explore it. Specifically, LacI is able to search the entire E. Coli genome rapidly via a variety of molecular mechanisms (Ref Berg, Ref Winter 1981 (1), Ref Winter 1981 (2)), which are diagrammed in figure 5.9. These mechanisms include sliding along the DNA double-helix, micro-hops, macro-hops and inter-segmental transfer (Ref Díaz). This multiplicity of search mechanisms allows LacI to (theoretically) search the entire bacterial genome of approximately 4 million bp on a timescale on the order a one second.

Because short 40 bp oligonucleotides were used in single molecule competition assays, the competitor sequence was probably not long enough for LacI to find the competitor O_{SYM} site by sliding along it with the free LacI DNA-binding domain. And since the off-rate for LacI from its operator site is apparently on the order of a day, as measured in the bulk competition assays (Ref Mehta), micro and macro hops are not significant while LacI is bound to a strong operator site like O_{SYM}. Thus, the dominant mechanism for LacI to leave its substrate in single molecule competition with 40 bp oligonucleotides is inter-segmental transfer.

5.5.2 Competition with full length competitors

To correct for this exclusion of LacI interactions, full end-to-end unlabeled DNAs (of identical sequence to the tethered substrates) were used as competitor. The long flanking ends of 9C14 correct for the omission of one-dimensional sliding of a free DNA-binding domain of LacI along competitor molecules and give a better approximation of the LacI-DNA complex *in vivo*.

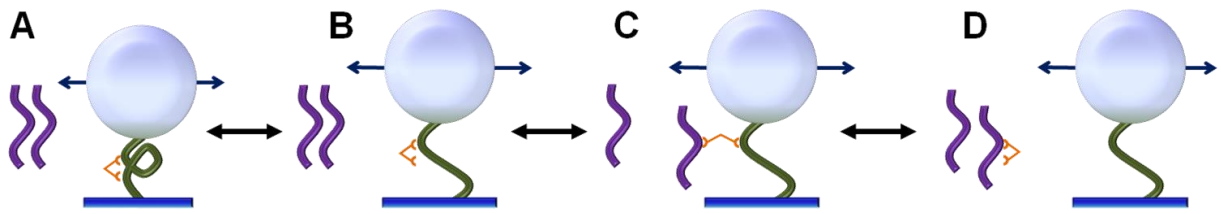


Figure 5.10 – Single molecule DNA competition assay with full sequence competitor. In the schematic above, DNA (purple lines) loops (A) mediated by LacI (small orange V-shape) and tethered to microscope slide coverslip (blue rectangle) and a polystyrene bead (white circle) may (B) break down, at which time a DNA-binding domain (DBD) of LacI should be free to (C) bind to competitor DNA to form a sandwich and (D) be competed off.

The principle of the adjusted single molecule competition assays is drawn above in figure 5.10, and representative rms traces of the full sequence SMCA are shown in figure 5.11. 9C14 is the only DNA substrate studied in the experiments discussed. At time zero, after a thirty minute pre-incubation with LacI in the flow cell, 50-75 microliters (the approximate flow cell volume) of 50 nanomolar full sequence competitor is added to the sample.

It is immediately observed that DNA loop formation and breakdown continues even in the presence of full sequence competitor DNA. For the 9C14 DNA construct measured presently, data observation and acquisition was carried out for many hours after competitor was added. The results in figure 5.11 show that LacI continues to reform on its intrinsically bent 9C14 substrate even after four hours (upper trace). In fact, enough data was acquired for looped and unlooped lifetime distributions (not shown) to be compiled and analyzed. Because of the lower number of data points (13 unlooped lifetimes), only an unmodified single-exponential fit could be made to unlooped state lifetimes, with a corresponding τ of 209 ± 32 seconds. While there were a low number of looped state lifetimes (17 in total), there were enough to attempt an unmodified double-exponential fit (equation 4.3); the measured parameters were, $C = 0.25 \pm 0.05$, $\tau_1 = 7.1 \pm 3.3\ seconds$, and $\tau_2 = 400 \pm 110\ seconds$. These values are very close to those measured without the presence of competitor and summarized in table 4.2. The loop formation

and breakdown kinetics of LacI bound to 9C14 is measured to be relatively invariant after several hours to DNA competition. Thus, LacI complexed with 9C14 (i.e. an intrinsically bent DNA substrate) mostly ignores competitor.

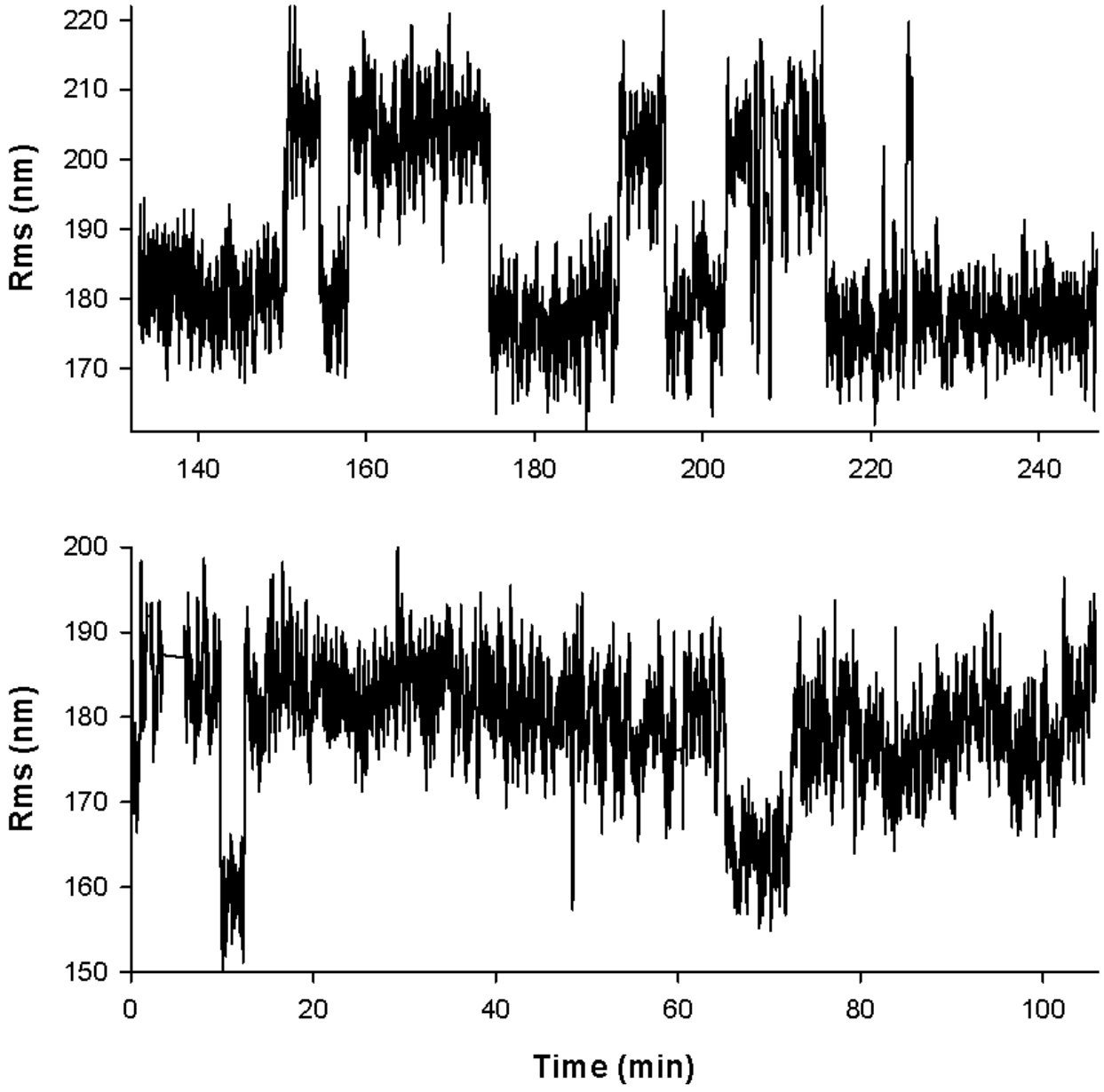


Figure 5.11 – Representative rms traces from the single molecule DNA competition assays with full sequence competitor. In both traces, 50 nanomolar of competitor DNA was added at time zero. It is observed that after several hours, LacI still reforms loops on its tethered substrate. These results dovetail nicely with the SMCAs that used 40 bp oligos, and confirm that LacI continues to mediate loop formation and breakdown on the intrinsically curved DNA (9C14) to which it is bound.

5.5.3 Where to go from here

To date, all single molecule competition experiments have been performed with the intrinsically-curved, competitively-resistant 9C14 DNA construct as substrate. To fully explore the question of DNA competitive resistance, it is necessary to explore the single molecule competition with other DNA substrates. In the next chapter, SMCAs results on a variety of DNA substrates will be detailed.

Chapter VI

Single Molecule Competition Assays II

6.1 Introduction

In the DNA constructs covered so far, single molecule experiments have dovetailed nicely with the results of bulk assays. While bulk DNA competition assays for LacI bound to curved constructs found that LacI could not be competed off for about a day, single molecule experiments both confirmed the bulk assay results and also revealed that loop formation and breakdown proceeds on an intrinsically curved DNA substrate even in the presence of a large excess of competitor. These results warrant further exploration of the DNA competition experiment. In the following sections, a variety of different DNA constructs are observed at the single molecule level in the presence of DNA competitor.

6.2 Unbent Control

While the single molecule competition assays (SMCA) covered so far have yielded surprising results, they have all been performed on the same DNA sequence (9C14). Other DNAs must be utilized to obtain a sense of loop formation breakdown dynamics in the presence of competitor. Thus, a control sequence is needed for use in SMCAs.

The most obvious choice for a DNA control in SMCAs is the same control sequence used in the bulk DNA competition assays (Ref Mehta). Designated Unbent Control (UBC), this sequence has comparable inter-operator distance to and shares identical operators with the 9C14 intrinsically bent sequence 9C14. However, UBC contains no phased A-tract bends, and is therefore a straight DNA sequence. TPM experiments measuring loop formation and breakdown events for single UBC molecules have been performed, and the results were summarized in tables 4.2 & 4.3 in chapter IV.

SMCAs with UBC were performed as full sequence SMCAs. In fact, for the remainder of this work, all SMCAs discussed were performed with full sequence competitor DNA that is identical in sequence to the tethered substrate but lacking any digoxigenin or biotin chemical labels. Examples traces from the Unbent Control competition experiments are plotted below in figure 6.1.

Plot A in figure 6.1 shows a representative trace of a bead in which, upon influx of competitor DNA solution, LacI was blown off. A sizeable portion of data was excluded because of this artifact. In plot B, it is seen that UBC continues to loop a few times after competitor DNA is added, but loop formation soon ceases and no looping events are observed after twenty minutes.

UBC is competed off on the order of 20-45 minutes. This competition time dovetails nicely with the competition half-time for UBC observed in bulk assays (Ref Mehta) of 30 minutes. Thus, for the analog of 9C14 lacking intrinsic curvature (UBC), DNA competition is successful over the experimental observation window and is measured to be less than one hour. With this data, it can now be said that DNA competition results at the bulk and single molecule levels do not contrast in their respective findings, but rather in their initial interpretations.

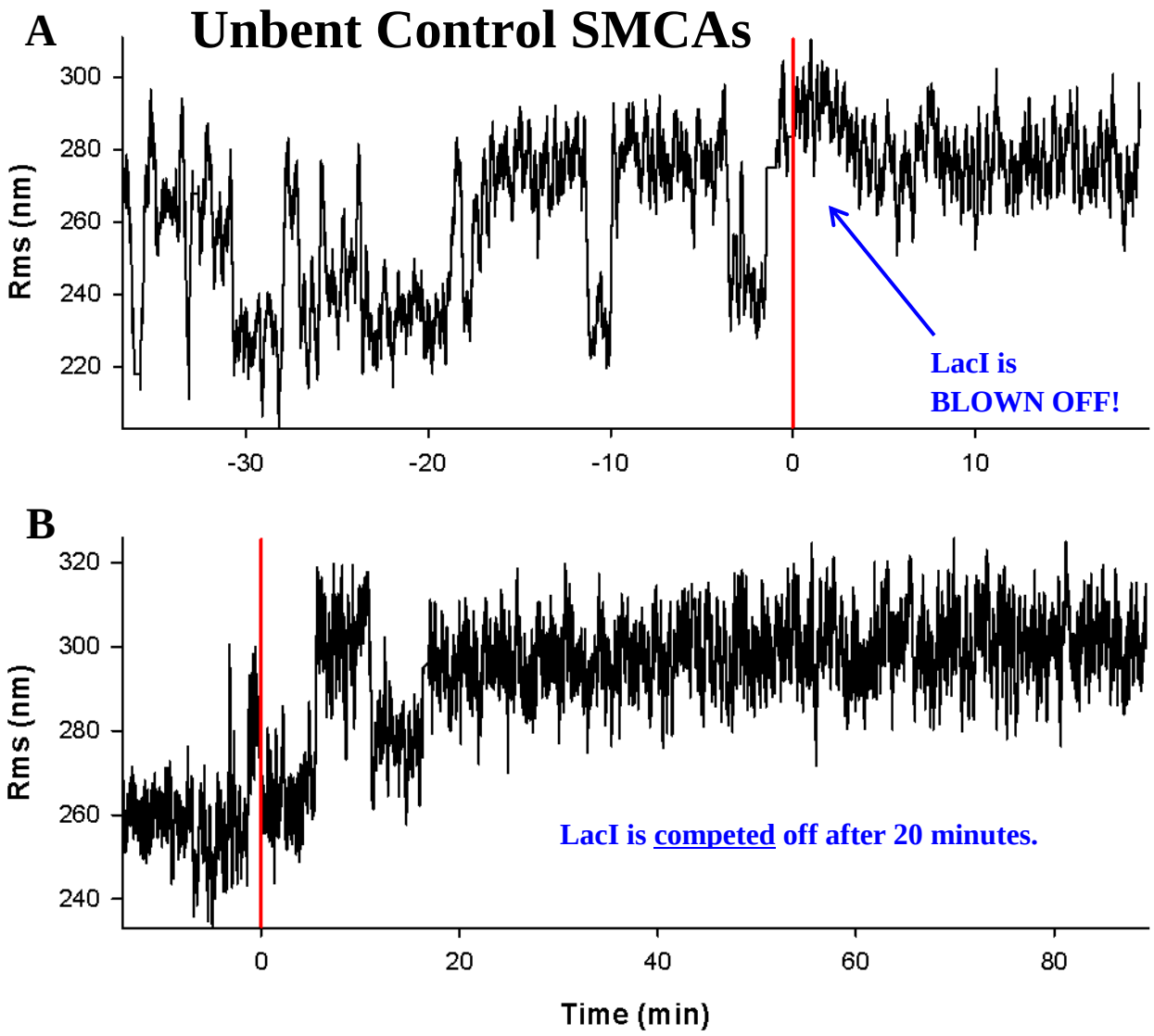


Figure 6.1 – Example rms traces from single molecule competition experiments with Unbent Control. At time zero in both traces, 50-75 µL (the approximate flow cell volume) of a 50 nanomolar solution of unlabeled competitor DNA in LBB+ buffer is flowed into the microscope chamber. In the upper trace (A), LacI is blown off upon addition of competitor DNA. A sizeable percentage of data points fall prey to this pathology of trace (A), and it is a limiting, complicating factor in data acquisition in the single molecule competition experiments. In the lower trace (B), LacI is competed off within 20 minutes. The red lines give the respective points in time at which competitor DNA solution is flowed into the slide chamber.

DNA competition is now shown to be a function of inter-operator intrinsic curvature. It is possible that the resistance to DNA competitor of a LacI-DNA complex is a function of other variables. This is explored in the following section.

6.3 Other non-curved DNAs

6.3.1 Motivation

DNA competition assays were performed on two additional, non-curved sequences in which the operator binding strength and inter-operator distance were varied. Designated 702 Unbent and 700 O_{SYM}, The sequences and design principles of these constructs are detailed in appendix B. 702 Unbent and 700 O_{SYM} contain operator center-to-center distances of 304 and 303 base pairs (bp), respectively, which is approximately twice the inter-operator distance for UBC of 158 bp. Thus, both DNAs will test the dependence of DNA competitive resistance on operator spacing relative to UBC. Furthermore, the sequences of 702 unbent and 700 O*SYM* are identical except in one respect – 702 unbent contains two O_1's (the strongest *in vivo* operator), while 700 O_{SYM}, like its name foretells, contains two symmetric ideal operator binding sites. Thus, the differences in single molecule measurements between these two DNAs are a measure of the effects of operator strength.

6.3.2 TPM with 702 Unbent and 700 O_{SYM}

Before SMCAs could be performed on these sequences, it was necessary to first measure loop formation and breakdown rates in a simple TPM experiment (i.e. without the presence of DNA competitor). Example TPM traces of 702 Unbent and 700 O_{SYM} and plotted in figure 6.2. Since the inter-operator distances for 702 Unbent and 700 O_{SYM} are about a factor of two larger than those of the Kahn DNAs (9C14, 11C12 & UBC), the relative distance between the looped and unlooped rms levels (ΔR_{States}) in these two new DNAs is considerably greater. For the new DNAs, ΔR_{States} typically falls in the range 65-80 nanometers, while for the Kahn DNAs ΔR_{States} was typically in the range of 25-40 nanometers. This larger difference between looped and

unlooped rms levels allows a smaller window size, W, to be used in computing the running-averaged rms.

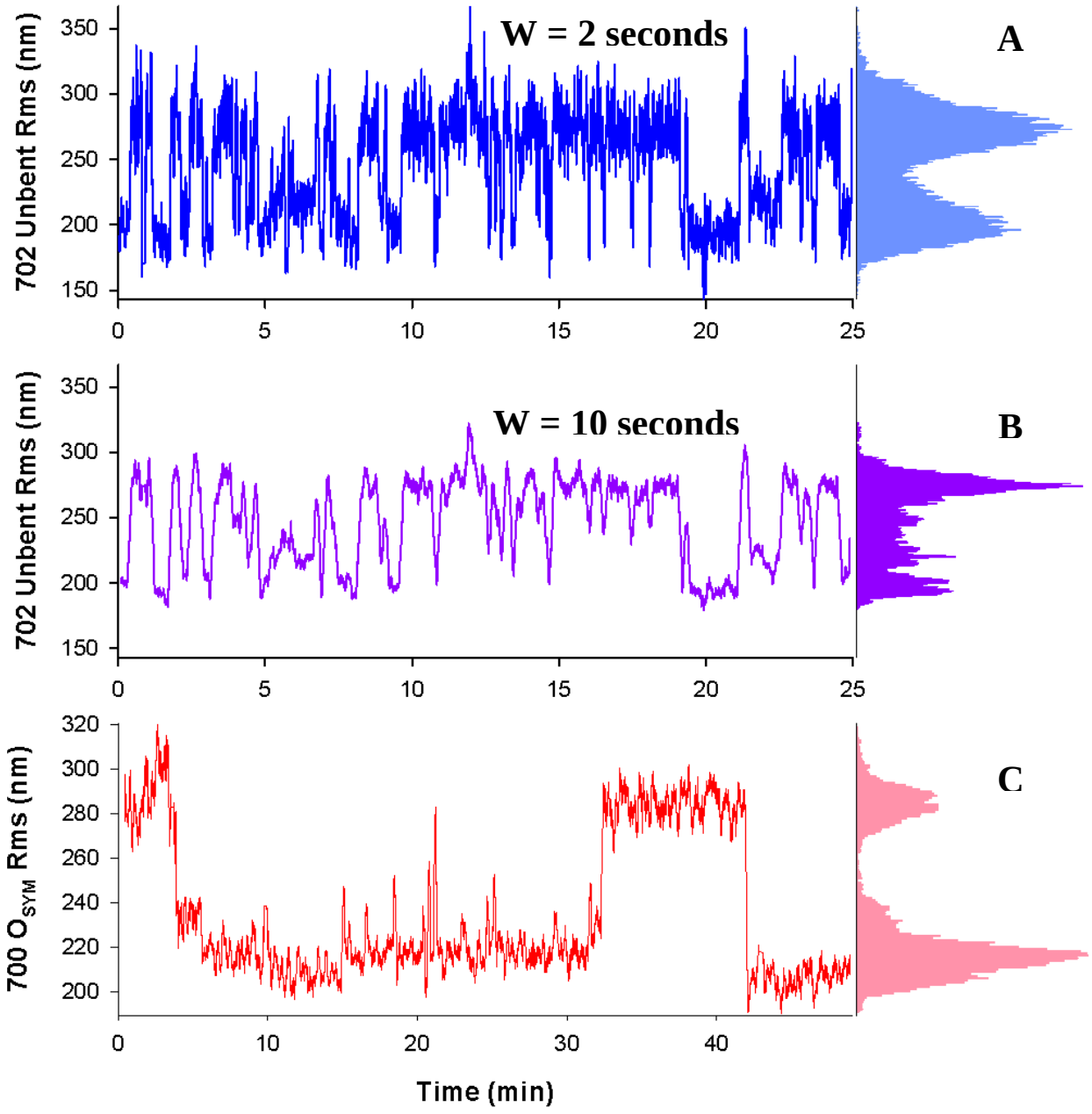


Figure 6.2 – Example rms traces and histograms from TPM with non-curved DNAs. Histograms are plotted to the right of each time trace. Trace A (top, blue lines) is the 702 Unbent DNA with a running window size, W, of 2 seconds. Trace B (middle, purple lines) is the same rms data with a running window on the rms of 10 seconds. While shorter temporal events are visible with the smaller window size, resolution of rms level spacing increases with increasing window size. In Trace C (lower, red lines), an rms time trace with $W = 10$ seconds is plotted for TPM experiments performed on 700 O_{SYM}. In both traces B and C, fine structure in the rms levels is visible both in the rms histograms (right) as well as in the time traces (left).

Traces A and B in figure 6.2 show the same 702 Unbent representative rms data with different window sizes, W. With the smaller window size of two seconds used in trace A, all events longer than two seconds are measurable. However, decreasing the window size decreases the spatial resolution as seen by the broader Gaussian peaks in the rms histograms plotted beside the time traces ($\sigma_{Peak} \sim 1/\sqrt{W}$). This smaller window size of two seconds is used for all lifetime data acquisition on 702 Unbent and 700 O_{SYM}, allowing many more lifetimes to be measured in thresholding and thus better fits to the modified double-exponential distribution functions.

The larger window size used in trace B shows that, although lesser temporal resolution is obtained, long window-size rms averaging reveals fine structure in the looped rms, visible in the histogram next to time trace B. As many as four Gaussian peaks are present in this histogram. Looped state fine structure is also visible the 700 O_{SYM} rms in trace C in the figure above. Thus, there is an inverse relationship between spatial resolution and temporal resolution in TPM.

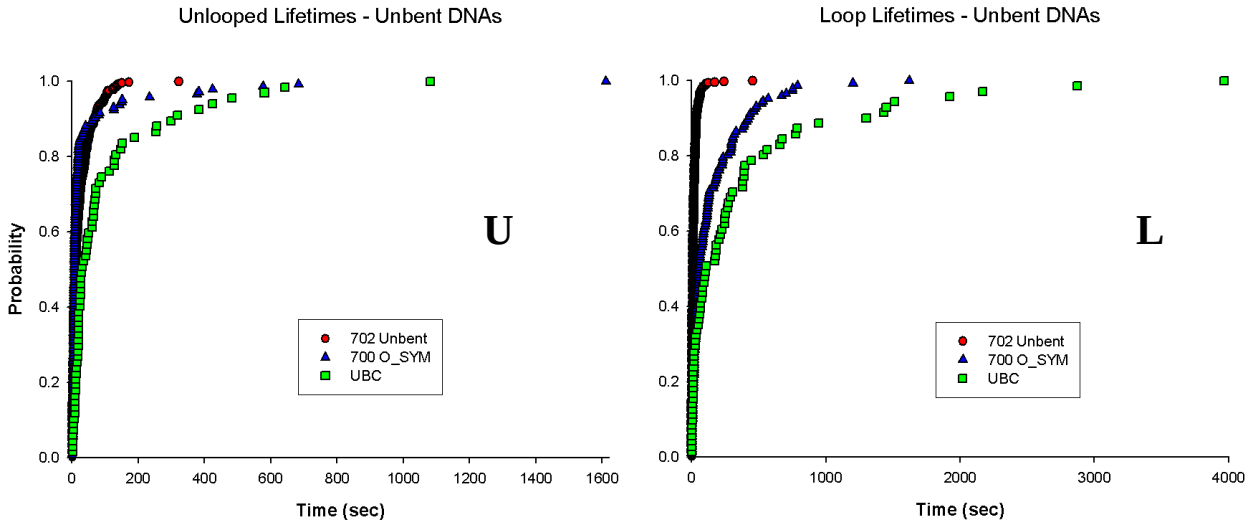


Figure 6.3 – Cumulative probability distributions of unlooped and looped lifetimes for unbent DNAs. In these probability distributions, unlooped (U) and looped (L) lifetimes are plotted for 702 Unbent (red circles), 700 O_{SYM} (blue triangles) and UBC (green squares). It is readily observed that, for short events, the probability distribution of 702 Unbent and 700 O_{SYM} unlooped lifetimes lay on top of each other. UBC shows the longest-lived unlooped states on average, visible as a gradual, exponential rise to maximum. The distribution of looped lifetimes reveals a steep rise to maximum for 702 Unbent, which in turn evinces the shortest looped states. Curiously, UBC shows both the longest-lived loop states and the shortest unlooped states.

DNA Construct	C	τ_1 (sec)	τ_2 (sec)	Number of looping events measured
700 O_{SYM}	0.435 ± 0.014	24.2 ± 1.1	227.3 ± 6.6	147
702 Unbent	0.210 ± 0.013	3.61 ± 0.15	17.14 ± 0.12	578

Table 6.1 – Fits to looped lifetime distributions of non-curved DNA constructs. These fits demonstrate the increased stability of the LacI-DNA complex when O_1 operators are replaced by symmetric, ideal operators (O_{SYM}), as observed by the large increased in both looped state lifetimes for 700 O_{SYM} relative to 702 Unbent. The first three columns give the fits of looped state lifetime cumulative distributions to modified double-exponential functions. The last column on the right gives the number of data points in each distribution (i.e. the number of looping events measured for each DNA listed).

Fit parameters & Kinetic Rates	UBC	700 O_{SYM}	702 Unbent	1151 O_1
C (relative population of 1st exponential)	0.657 ± 0.025	0.871 ± 0.007	0.525 ± 0.007	0.7 ± 0.2
1st unlooped exp. time, τ_1 (sec)	24.8 ± 1.2	5.43 ± 0.09	5.28 ± 0.08	5.2 ± 1.6
2nd unlooped exp. time, τ_2 (sec)	222 ± 21	136 ± 17	37.81 ± 0.43	77 ± 6
$\tau_U = 1/k_L$, U→L (sec)	**35.7 ± 2.9**	**6.20 ± 0.15**	**8.92 ± 0.22**	**7.2 ± 4.1**
$\tau_+ = 1/k_+$, 2→1 (sec)	154 ± 13	118.9 ± 3.2	22.36 ± 0.47	55 ± 18
$\tau_- = 1/k_-$, 1→2 (sec)	97 ± 15	46.0 ± 3.8	16.91 ± 0.63	20.5 ± 20.8
Number of unlooped lifetimes measured	67	144	580	247*

Table 6.2 – Fits to unlooped lifetimes of non-curved DNA constructs and kinetic model solutions. The UBC data is reproduced in column 1 from table 4.4 in chapter IV. Columns 2 and 3 give, respectively, the 700 O_{SYM} and 702 Unbent TPM analyses on the unlooping kinetics of these DNAs. The last column on the right contains data from the literature. Specifically, TPM measurements by Finzi and Gelles (Ref) on a DNA construct, designated here as 1151 O_1, are listed in the first three rows of column four. The DNA sequence of 702 unbent is originally derived from this construct of Finzi and Gelles. Using the loop formation kinetic model solutions given in equations 4.15-4.17, the loop formation time, $\tau_{U,\,as}$ well as the other kinetic parameters, τ_+ and τ_-, are obtained. The loop formation times for 702 Unbent and 1151 O_1 are comparable, giving a further proof of principle of the kinetic model utilized. The last row of the table gives the number of data points in each unlooped lifetime distribution. The asterisk next the data number of data points for 1151 O1 denotes that this value is effectively smaller than this because the analyzed unlooped lifetimes were binned into a probability histogram; such binning reduces the effective number of data points and thus increases the fitting errors relative to cumulative probability fits.

Looped and unlooped lifetime cumulative distributions are plotted in figure 6.3, and modified double-exponential function fit parameters to looped and unlooped lifetimes are summarized in tables 6.1 and 6.2, respectively. It is noted that the loop lifetimes of 702 Unbent compared to 700 O_{SYM} are about an order of magnitude smaller. This is also evident in the smaller number of looping events in figure 6.2 for 700 O_{SYM} as compared to 702 Unbent.

The unlooped state lifetimes are fit to the kinetic model solutions given in equations 4.15-4.17 (table 6.2). In the fourth column of this table, a DNA construct designated 1151_O1 is listed. This DNA's double-exponential fit parameters were obtained from literature measurements by Finzi and Gelles (Ref). In fact, the 702 Unbent sequence is a derivative construct of 1151_O1. Thus, the similarity of loop formation rates for the non-curved DNAs to 1151_O1 represents additional confirmation of both the loop formation kinetic model as well as the experimental techniques utilized in the present report.

It is evident that two unlooped states with vastly different stabilities exist. While the loop formation rate remains constant (bold row in the table) for the Finzi-derived DNAs (Ref), an additional, "dark" unlooped state (denoted earlier as U_2) exists which is unable to form loops. The DNA with stronger, symmetric ideal binding sites is found to dwell for longer times, τ_+, in this "dark" state, thus confirming that U2 is most likely doubly-occupied DNA substrate (i.e. one LacI protein bound to each DNA operator).

6.3.3 SMCAs with the non-curved DNAs

Single molecule competition experiments were performed on the non-curved DNAs 702 Unbent and 700 O_{SYM}. Example traces of both SMCAs are plotted in figure 6.4. The average competition time for 702 unbent is less than ten minutes, while for 700 O_{SYM} the competition

time falls in the range of 20-45 minutes. This represents a significant difference in competitive resistance which is attributable to the different operators, and their associated binding strengths to LacI, present in these two constructs. Both timescales are visible in the rms traces below.

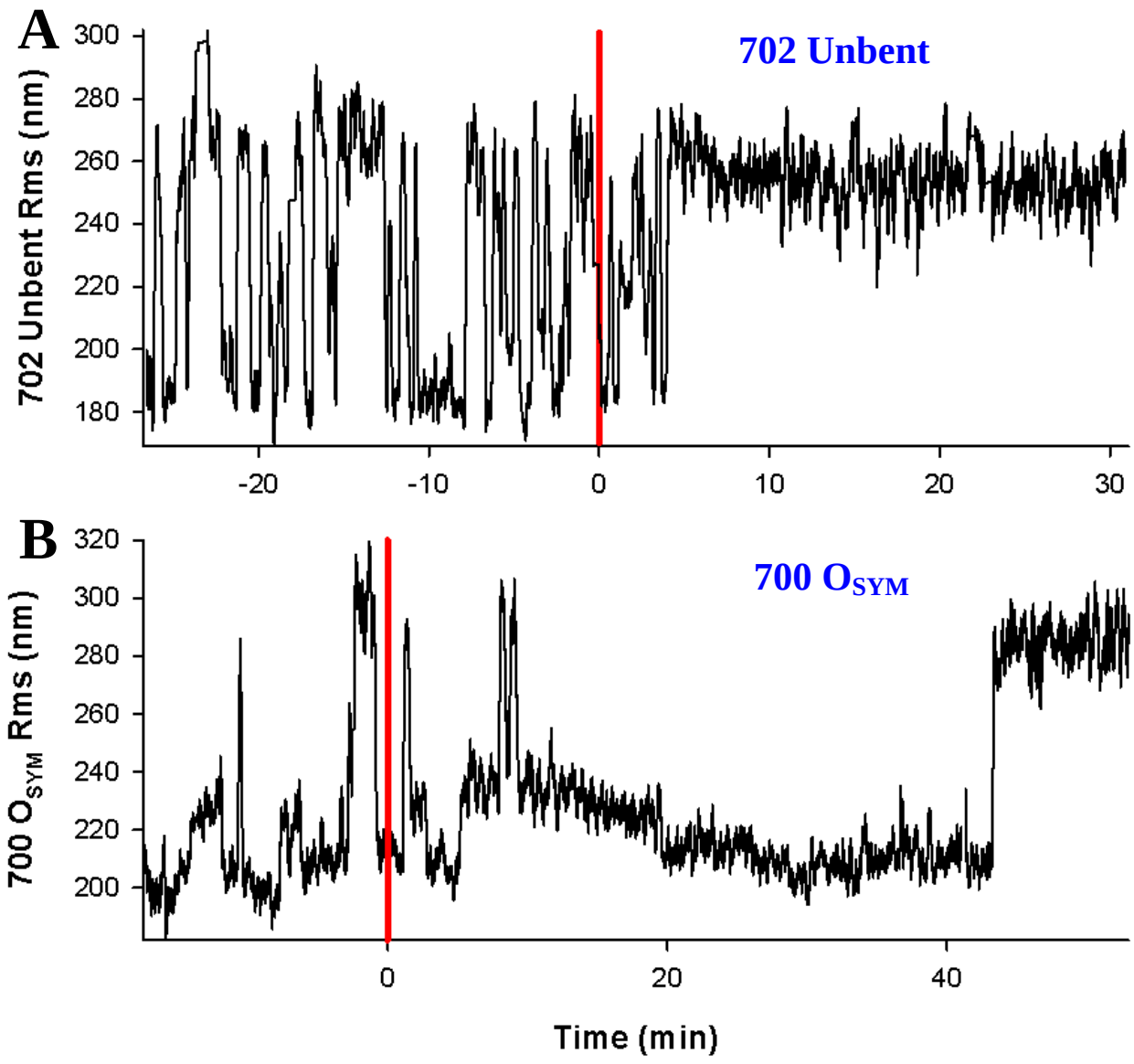


Figure 6.4 – Single molecule competition assays (SMCA) with 702 Unbent and 700 O_{SYM}. For each trace, 50-75 µL of a 50 nanomolar competitor DNA solution was added at time zero (vertical red lines in both traces). In trace A, a representative SMCA for 702 Unbent is shown. About 5 minutes after addition of competitor, loop formation ceases on 702 Unbent. For the SMCA with 700 O_{SYM} plotted in trace B, loop formation ceases after about 45 minutes. This large ratio reflects the stronger, symmetric ideal operator sites on 700 O_{SYM} as compared to the *in vivo* O1 operators on 702 Unbent.

6.4 A modified kinetic diagram of DNA competition

To explain and consolidate the results so far, a new kinetic model of the DNA competition process is proposed below in figure 6.5. Some form of "silent" unlooped state exists which is predominantly unavailable for competition. Given that a free DNA-binding domain (DBD) should have competitor DNA diffusively bind it and form a sandwich (figure 6.5, state D) in less than a second, this "silent" state must be one in which the DBD is occluded.

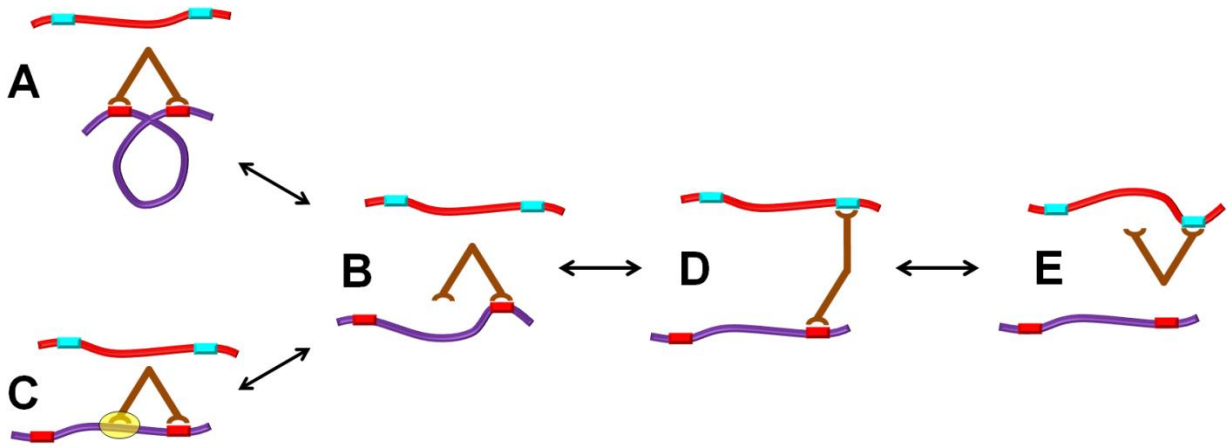


Figure 6.5 – A hypothesized model of the DNA competition experiment. The looped LacI-DNA complex (A) breaks down (B) at which point LacI (C) interacts with the nonspecific DNA near the operator site (yellow ellipse). Depending on LacI's affinity for the nonspecific DNA, LacI could only rarely interconvert to state B with one free DNA-bind domain (DBD). Once LacI has a free DBD (unbound to DNA), it may (D) bind to competitor and (E) be competed off.

Therefore, it is hypothesized that the LacI DBD that is unbound from operator DNA in the unlooped state is actually bound to nonspecific DNA in the "silent" state (figure 6.5, C). This would account for long total unlooped times (in the presence of a large excess of competitor DNA) that were observed in competition experiments with intrinsically curved DNA substrates. Furthermore, the presence of the intrinsic bends themselves could predominantly favor the nonspecifically-bound LacI-operator complex (the "silent" state) over a DBD unbound from DNA (figure 6.5, state B).

6.5 If LacI interacts nonspecifically with adjacent DNA …

If the nonspecific DNA, including DNA outside the looping region, is important in LacI-mediated DNA looping, then this must be directly tested. Experiments must be designed in which DNA outside of the looping region is altered in a variety of ways. Any effects on looped or unlooped stability would thus lend evidence to the newly hypothesized kinetic model of DNA competition (figure 6.5). Such tests are described in detail in the next chapter.

Chapter VII

The Flanking Sequence Matters

7.1 Introduction

The previous chapter detailed the startling hypothesis that LacI's interactions with nonspecific DNA are important in the kinetic diagram of DNA competition. If DNA outside of the looping region is truly important, then it is a straightforward to process to test this idea at the single molecule level. Experiments can be designed with varying flanking regions sequences, and the effects monitored by observing in changes in loop formation or breakdown kinetics in individual bead time traces. In the following sections, such a procedure is performed utilizing the Tethered Particle Motion (TPM) technique on variants of a curved DNA sequence (Ref Mehta).

7.2 TPM on DNAs with AT-Rich and CG-Rich flanking regions

The sequence of 9C14 was used to generate DNA constructs of varying flanking region composition. Specifically, 132 base pairs toward the 5' and 3' sides of the looping region were mutated to contain identical, "special" flanking regions with, alternatively, AT-Rich or CG-Rich regions. The CG contents of the AT-Rich and CG-Rich special regions are, respectively, 32.5 and 62.5 %. Un-mutated 9C14 has 52.3 % CG content in these regions.

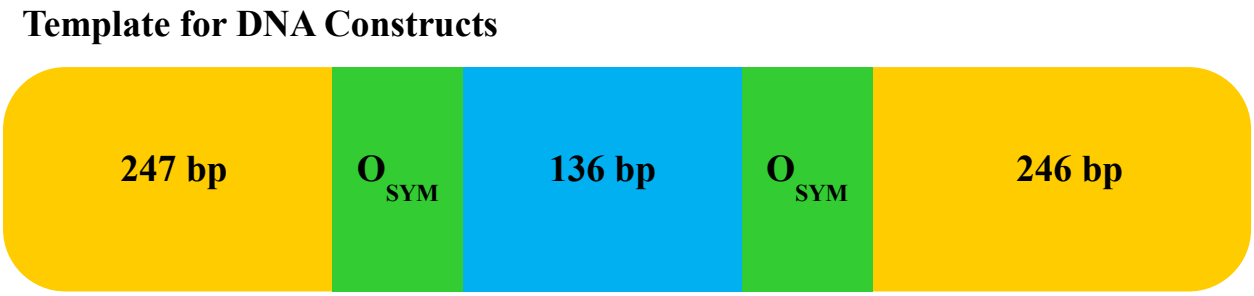


Figure 7.1 – The template for flanking sequence variants. The 9C14 DNA construct (sequence given in appendix B), is drawn schematically above. The inter-operator region (blue) is unchanged, as are the symmetric ideal operators (O_{SYM}'s, green). The flanking regions (yellow), however, are made alternatively AT-Rich or CG-Rich versus the original, random 5' and 3' flanking sequences in 9C14.

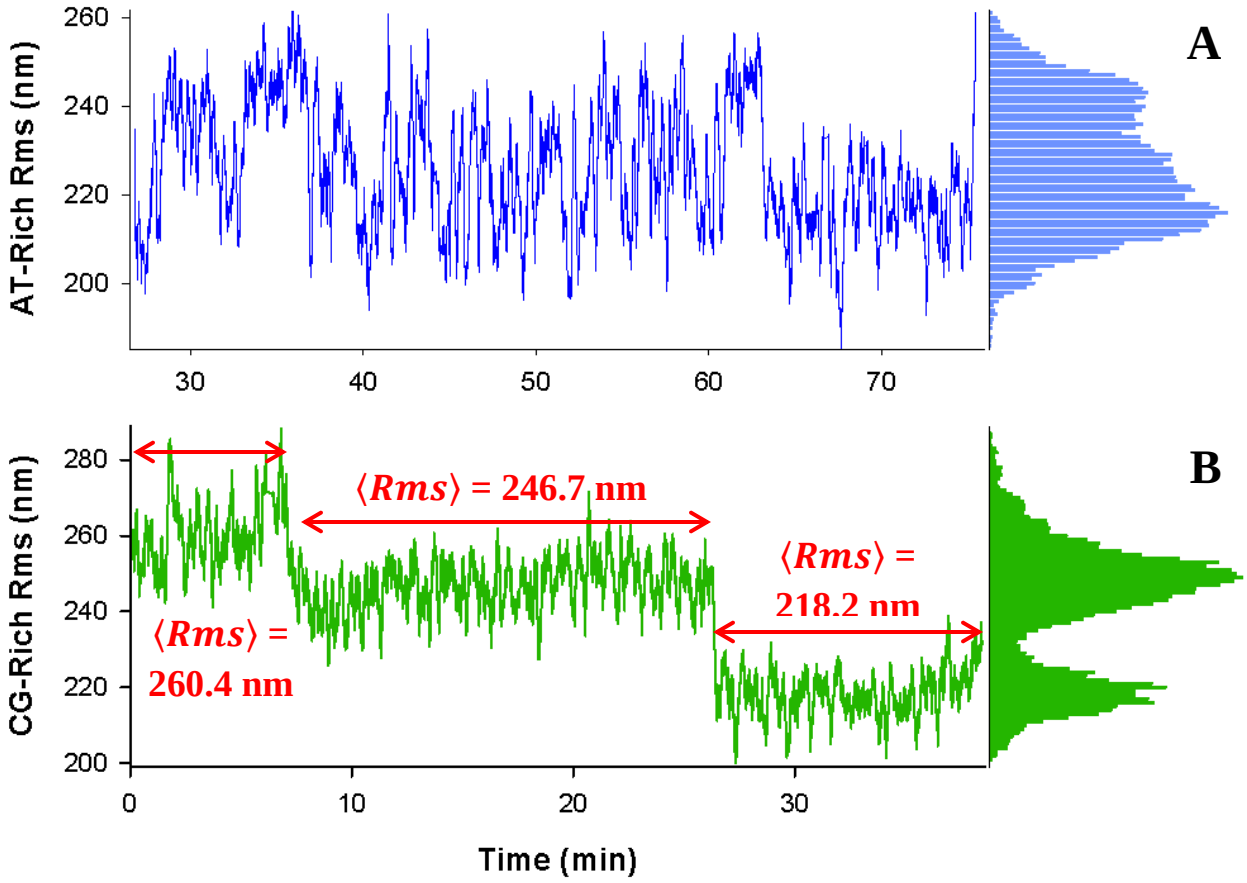


Figure 7.2 – Example rms traces & histograms of AT-Rich & CG-Rich TPM. For AT-Rich DNA (trace A, blue lines), a multiplicity of states is visible in both the anomalous time trace as well as the histogram, which contains a large smear in place of well-defined rms Gaussian peaks. For CG-Rich (trace B, green lines), multiple long-lived rms levels are observed. The average bead rms in each level is given in red, with red arrows demarcating each level's temporal length. The difference in rms levels for the CG-Rich trace is 13.7 and 28.5 nanometers, respectively. This data is in marked contrast to the simple, 2-level structure of un-mutated 9C14 TPM traces; comparison traces of un-mutated 9C14 and CG-Rich are given in figure 7.3.

The design principles of the new, special versions of 9C14 are illustrated above in figure 7.1, and full sequences for these and other DNAs is given in appendix B. These DNA sequences were obtained from new plasmids for these constructs (DNA 2.0, Inc.), and purified by polymerase chain reaction (PCR). As described in chapter III, end-labeled DNAs for use in TPM experiments were generated by PCR with primers containing digoxigenin and biotin chemical labels. TPM data was acquired for many hours at a time for each new sequence.

Examples traces from AT-Rich and CG-Rich TPM data are plotted above in figure 7.2. It is seen that AT-Rich has a multiplicity of short-lived rms levels (upper trace), while CG-Rich has several long-lived rms levels (lower trace). The TPM experimental data on sequences with special flanking regions is contrasted with that for unmodified 9C14 in figure 7.3. The AT-Rich DNA (the upper trace in figure 7.2) demonstrates anomalous looping kinetics. Specifically, rapid interconversion between closely spaced levels is observed.

In contrast, the CG-Rich DNA (lower traces in figures 7.2 and 7.3, respectively) show multiple closely-spaced levels that only interconvert very slowly. Moreover, the spacing observed between levels is on the order of ten nanometers, which is difficult to accurately resolve both spatially (because of the large rms running-average window size, W, that would be required to see them clearly) and temporally (due to such a large window size, W, obscuring short-time events) in TPM experiments.

Normal traces for unmodified 9C14 evince a simple two-peak Gaussian structure, corresponding to looped and unlooped states, in their rms histograms. For example, trace A in figure 7.3 contains two peaks centered at 203.7 and 178.3 nanometers (nm), giving a value for ΔRms of 25.4 nanometers. While the absolute value of unlooped rms changes depending on the bead size, the observed difference in rms levels for 9C14 of around 20-30 nm remains constant.

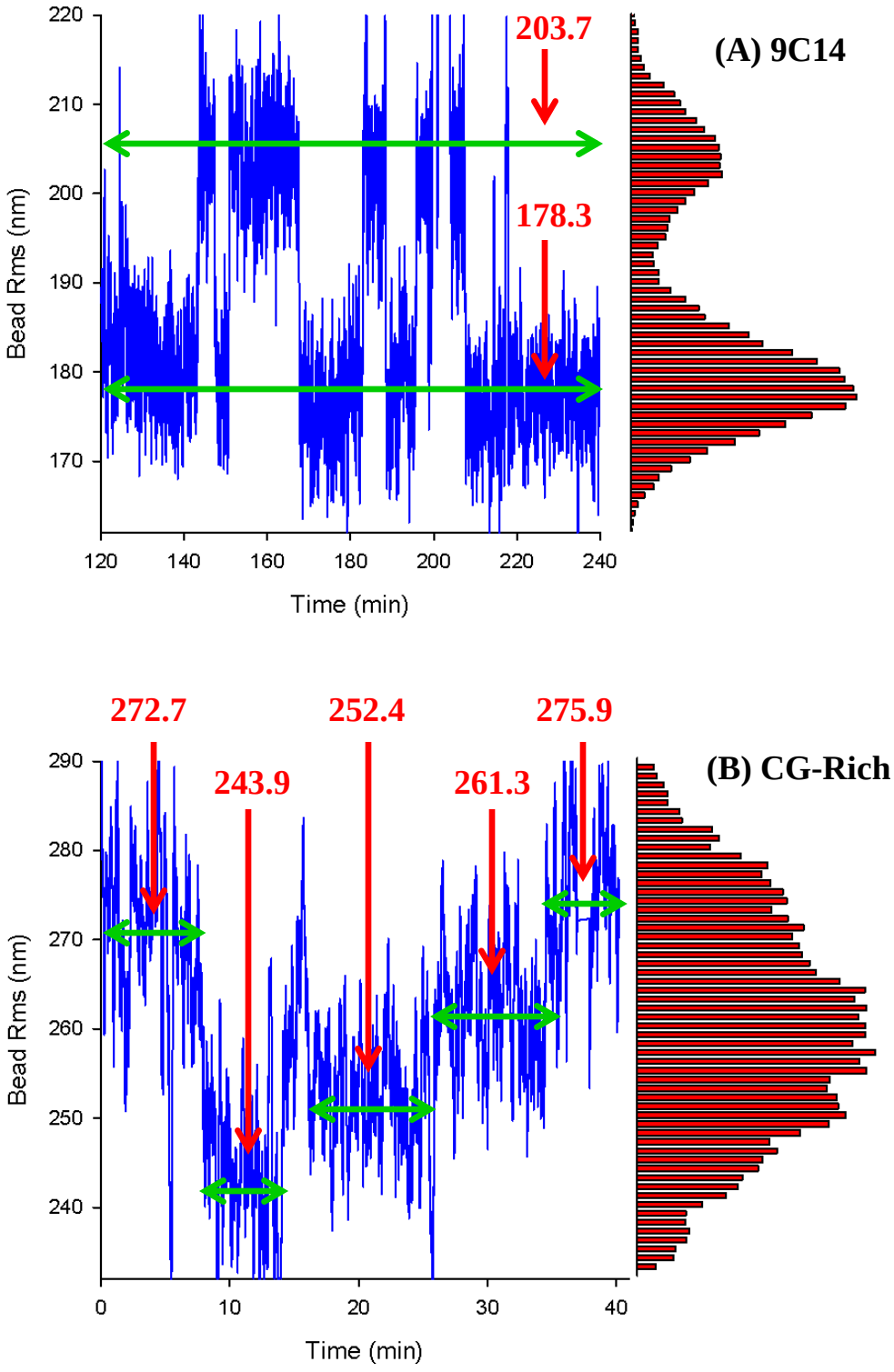


Figure 7.3 – Example rms traces & histograms of 9C14 & CG-Rich TPM. In trace A, a 9C14 rms trace from a full sequence SMCA is plotted; its corresponding rms histogram is plotted to the right of time the series. Two levels are visible in the time series, corresponding to two Gaussian peaks in the rms histogram. The Gaussian peak center lines are plotted as green arrows, and center positions are given in red in units of nanometers. In contrast, the data for the CG-Rich DNA (B) evince four different rms levels, which show as a smear in the rms histogram.

It is interesting to note that the ΔRms between the top and bottom level peaks of 32 nanometers for the CG-Rich trace above (B in figure 7.3, 275.9 and 243.9 nanometers) is comparable in magnitude to the range of rms level differences for unmodified 9C14. This consistency in top-to-bottom absolute rms level differences for holds across 9C14, AT-Rich and CG-Rich DNAs. This suggests that a form of symmetry among the looped rms levels is broken by altering the flanking regions in 9C14.

7.3 The meaning of it all

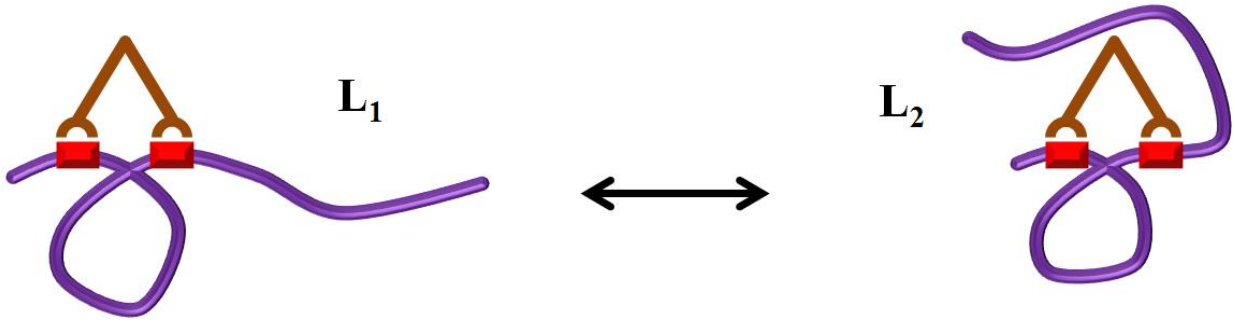


Figure 7.4 – A new LacI-Flanking sequence interaction model. In the above drawing, LacI (brown) bound to substrate DNA (purple, with operator binding sites drawn as red rectangles) in the looped state (L_1) has an observable interconversion into a second looped state (L_2) in which flanking region DNA is interacting with LacI in the looped state.

To interpret the TPM measurements with AT-Rich and CG-Rich DNAs, a new LacI-flanking DNA interaction model is proposed. In this model, drawn schematically in figure 7.4, LacI would reversibly interact with DNA in the flanking regions while in the looped state. Test measurements with a single-operator version of un-mutated 9C14 (data not shown) found that in the absence of a second symmetric operator site (in the presence of LacI), rms traces levels do not show multiple-level behavior. Thus, flanking DNA interactions with LacI in the unlooped state would most likely not register as rms changes in TPM experiments.

In summary, LacI is proposed to reversibly interact with flanking DNA in the looped state. Given the magnitude of rms changes observed, such interactions could be due to wrapping of this DNA region around LacI in the looped complex. There is *in vitro* evidence from thorough biochemical thermodynamic assays performed on single-operator constructs (Ref Tsodikov) that LacI interacts with non-operator DNA. However, those interactions were interpreted in those experiments as evidence for wrapping of the inter-looping region around LacI. While the current experiments do not rule out those interactions as possibly occurring *in vivo or in vitro*, effects were observed when the flanking region was altered. This suggests robust interactions between LacI and the entire DNA sequence in the neighborhood around the operator site (Ref Goodson).

Chapter VIII

Conclusions and future outlook

8.1 An optimized system

In conclusion, the lactose Repressor protein (LacI) is an archetype for transcription factor proteins specifically and protein-DNA interactions in general. The experimental results described in this work add new dimensions to the canonical understanding of protein-mediated DNA looping. Whereas at this juncture the simplified model of DNA looping was based on simple binding and unbinding of LacI to two distant operator sites, the results of the variety of single molecule performed have demonstrated the importance of additional interactions with LacI and its substrate DNA.

While it is well known that changes to the operator site sequence (Ref Oehler 1990) and inter-operator distance (Ref Müller) affect the lac Repressor system, two corollaries must now be added. First the DNA sequence of the inter-looping region is important. The addition of A-tract bends to make this region intrinsically curved modestly affects the loop formation and breakdown kinetics and drastically affects the resistance of the protein-DNA complex to DNA competition. Other systems like the nucleosomal positioning sequence (Ref Johnson) are known to have no effect on looping kinetics in the inter-operator region; however, if A-tract bends can

affect LacI's DNA competitive resistance, perhaps other DNA sequences also have a similar effect that is not observed by simple studies on loop formation and breakdown kinetics alone.

Second, the DNA sequence of the flanking region is important. Changes to the CG content of DNA outside of the looping region by less than 20% in the positive and negative directions visibly alters the Brownian motion of polystyrene beads attached to intrinsically curved DNAs. LacI thus interacts with DNA regions that were not recognized to be important for DNA looping dynamics.

8.2 Future directions

8.2.1 TPM with the *in vivo* sequence

In summary, the three regions of any DNA construct used in measurements of protein-mediated DNA looping (inter-operator, flanking and operator binding sites) are all important and they all play a role in the interaction kinetics of LacI with DNA. These studies further suggest that the entire neighborhood *in vivo* near a binding site for a given transcription factor protein could be an optimized system that takes advantage of all of these interactions. Future research directions with LacI could address many of these questions using the *in vivo* lac Operon DNA sequence as an *in vitro* diagnostic.

8.2.2 Footprinting and FRET as tests of possible interactions between LacI and nonspecific DNA

Also, if LacI truly interacts with nonspecific DNA in the neighborhood of an operator site, the nature of this interaction can be explored through two in vitro tests. First, DNase I footprinting experiments can be performed on the DNA constructs of variable flanking region sequence described in chapter VII. Such footprinting assays performed in the presence and

absence of LacI can reveal the nature and extent of any interactions between LacI and the flanking regions of a DNA construct.

Second, the modified version of LacI designated Q231C (discussed in chapter V) has cysteine residues that have been mutated into its amino acid chain. These residues were used to attach a linking molecule via sulfide bonds and thus produce a crosslinked LacI protein that could not dimerize or, in theory, change its conformation. However, these sulfide residues may also be used to attach other molecules of interest. There exist fluorophores with reduced sulfides in their chemical structures that could, in principle, be attached to Q231C and thus produce a LacI protein with an attached fluorophore. As long as the size of the fluorophore was very small compared to that of LacI, the steric effects of the fluorphore's presence could be neglected and this fluorophore-labeled LacI could be utilized in a FRET assay. In such a FRET experiment, a complementary fluorophore can be placed on substrate DNA in either the looping region or in the flanking region of various constructs. Observation and quantitation of the presence or absence of energy transfer can be interpreted as a diagnostic of the interactions of LacI with non-operator DNA. The precise location of the DNA fluorophore can also be varied, and the degree of relative energy transfer would thus give the LacI-DNA interaction strength at each position. Such an assay could be performed in both bulk FRET measurements and as a series of single-molecule FRET experiments.

8.2.3 A-tracts are not the only special sequences

As mentioned above, there are other "special" DNA sequences (e.g. the nucleosomal positioning sequence, which is abbreviated NPS from this point on) that are important for various protein-DNA interactions. While such special sequences may not affect the DNA looping process, it is possible that they interact with a looping protein in a way that is not

observed in simple measurements of looping kinetics. Thus, future single molecule DNA competition experiments could utilize different DNA looping proteins, perhaps those found in eukaryotes, in conjunction with DNA containing an NPS segment in the looping region.

8.2.4 Do other proteins matter in DNA looping?

While bacteria do not have nucleosomes, they do contain histone-like, nucleoid proteins that associate with the bacterial DNA (Ref Dillon). Given that these nucleoid-associated proteins like HU and H-NS bind to and interact with the bacterial DNA throughout its genome, it is possible that the presence of these proteins on a DNA substrate could affect the physical properties of the looping region for LacI *in vivo*. Thus, a straightforward experimental test of this idea would be to perform TPM measurements of loop formation and breakdown mediated by LacI in the presence and absence of HU or other nucleoid-associated proteins.

8.3 Final thoughts

In studying the lac Repressor protein, many experimental and theoretical techniques have been utilized. New methods of approaching biomolecular interactions needed to be developed, and new ways of thinking about protein-DNA interactions were engendered. Furthermore, the molecules in question in these experiments are not even directly observable with visible light and clever nanoscale reporting methods were exploited. Therefore, to study the biophysics of protein-mediated DNA looping has been to use the unseen to measure the unknown and thus to find the undiscovered.

APPENDICES

Appendix A

Image Processing and Positional Tracking

A.1 Image Processing

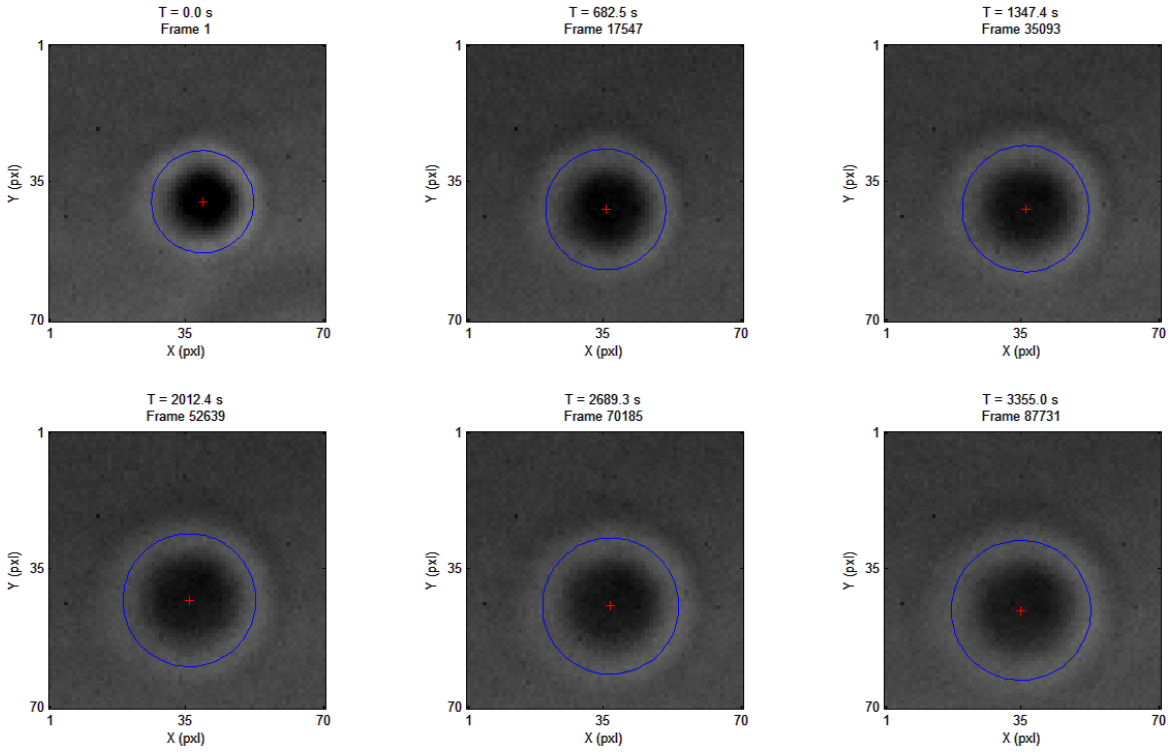


Figure A.1 – Sample time series of a single bead in a 70 pxl by 70 pxl region of interest (ROI) in the camera images. Red plus signs mark the computed x and y coordinates of the bead positions, while the blue circles denote the calculated bead radius. These six ROI snapshots span about 56 minutes of TPM data acquired for the Unbent Control (UBC) sequence. This time series proceeds begins with the first frame acquired in the top left corner with images from equally spaced time increments proceeding left-to-right, top-to-bottom; the ROI snapshot in the bottom right corner is from the last frame acquired.

Tethered particle motion (TPM) data (example bead images plotted in figure A.1) is acquired on a custom-built visible light microscope and imaged on a PixeLINK ® Pl-B741 camera. Images are acquired at video rate with a data acquisition script written in MATLAB ®. Although the total chip area of the camera is 1280 by 1024 pixel^2, only small portions of each frame are saved (figure A.2).

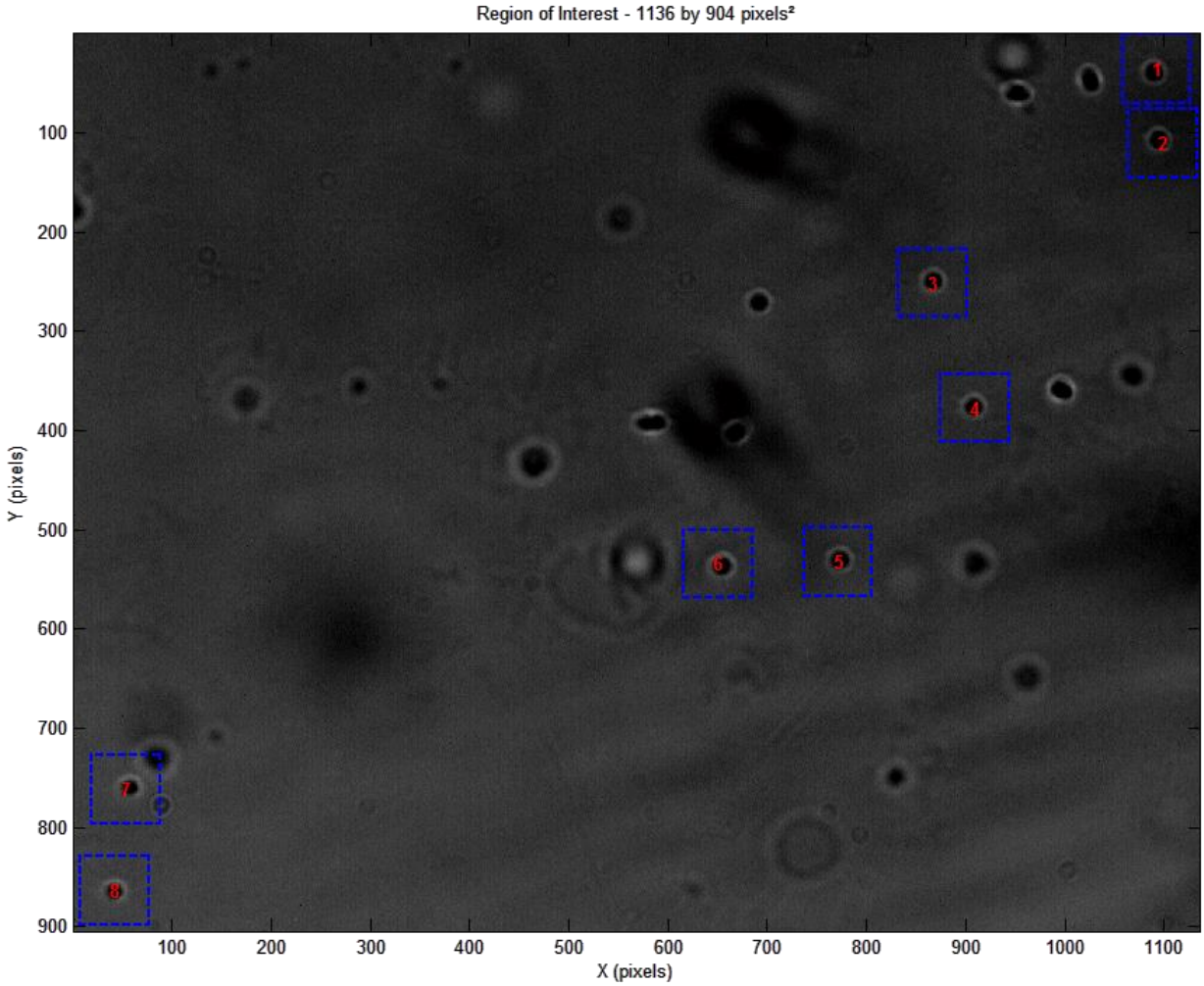


Figure A.2 – Full image saved from TPM data. In the example image above from UBC TPM, 8 regions of interest (ROIs, dashed blue squares) corresponding to 8 tethered beads (numbered 1-8 in red) are acquired for data analysis. A time series of chopped ROIs is given in figure A.1. The total area above is 1136 pxl by 904 pxl.

For the data in figure A.2, the total frame size acquired by the camera and digitally sent to the computer for processing is minimized to contain as little total area as possible while still

acquiring all bead data desired. As each frame is acquired, it is quickly chopped so that only the individual ROIs (dashed blue squares) are saved. While the total image area above is 1136 by 904 pxl^2, the actual area saved is the 70 pxl by 70 pxl area demarcated for each ROI (i.e. 39200 pxl^2), which is slightly less than four percent of each frame.

A.2 Mirrored Autocorrelations

After image acquisition, each ROI is analyzed as follows. Each ROI is correlated with the inverted mirror image of itself. Termed a "mirrored autocorrelation", the correlation maximum can be converted to bead x and y positions. The two-dimensional correlation function calculation is carried out in Fourier space by a Fast Fourier Transform (FFT). The 4900 pxl^2 ROI is padded (with pixel values equal to the mean intensity of the image) to optimize the FFT calculation, and thus the bead position is computed quickly for each ROI.

To achieve sub-pixel resolution, a two-dimensional quadratic maximum function,

$$z(x,y) = z_{max} - \beta_x(x - x_{max})^2 - \beta_y(y - y_{max})^2, \tag{A.1}$$

is fit to a neighborhood of points around the bead position computed from correlation, x_{Corr}, y_{Corr}. Using the relations below,

$$\beta_x = z(x_{Corr}, y_{Corr}) - 0.5\big(z(x_{Corr} + 1, y_{Corr}) + z(x_{Corr} - 1, y_{Corr})\big) \tag{A.2}$$

$$\beta_y = z(x_{Corr}, y_{Corr}) - 0.5\big(z(x_{Corr}, y_{Corr} + 1) + z(x_{Corr}, y_{Corr} - 1)\big) \tag{A.3}$$

$$\beta_x = \big(z(x_{Corr} + 1, y_{Corr}) - z(x_{Corr} - 1, y_{Corr})\big)/\big(4(x_{max} - x_{Corr})\big) \tag{A.4}$$

$$\beta_y = \big(z(x_{Corr}, y_{Corr} + 1) - z(x_{Corr}, y_{Corr} - 1)\big)/\big(4(y_{max} - y_{Corr})\big) \tag{A.5}$$

the bead center (x_{max}, y_{max}) can be more accurately solved for as,

$$x_{max} = x_{Corr} + \big(z(x_{Corr} + 1, y_{Corr}) - z(x_{Corr} - 1, y_{Corr})\big)/(4\beta_x) \tag{A.6}$$

$$y_{max} = y_{Corr} + \left(z(x_{Corr}, y_{Corr} + 1) - z(x_{Corr}, y_{Corr} - 1)\right)/\left(4\beta_y\right) \quad \text{(A.7)}$$

where the values for β_x and β_y are calculated from equations A.2 and A.3. This completes the positional tracking algorithm utilized in TPM.

A.3 Further Computations

Three additional calculations are made from the ROI image data for each bead. These are plotted below in figure A.3. First, a circularly averaged pixel intensity proceeding radially outward from the bead position (computed from equations A.6 and A.7), with 0.5-pixel bins, is computed (top plot in figure A.3). The radius that maximizes this function corresponds to the first maximum of the bead diffraction ring (white rings around the dark centers in the ROIs in figures 7.1 and 7.2). This is taken to be an effective bead size. Because of slow-time z-drift bringing the bead slightly out of focus (due to the oil on the oil-immersion microscope lens), the radius increases over time. To achieve sub-bin resolution, a quadratic fit is performed with the computed radius and two adjacent points around the maximum. Both raw bead radius (gray lines) and running-averaged intensity (solid blue lines) are shown.

Next, the mean intensity of the pixels around the edges of every ROI is taken as an image background intensity (figure A.3, middle trace). Background intensity is checked in each data set for abnormalities. The Fourier transform of the background intensity is also computed to make sure that no "preferred" frequencies appear (which could arise from the frequency of the illumination source).

Lastly, since an image autocorrelation function is computed to find the bead position, the value at the correlation maximum can be taken as signal intensity in TPM (middle plot in figure A.3).

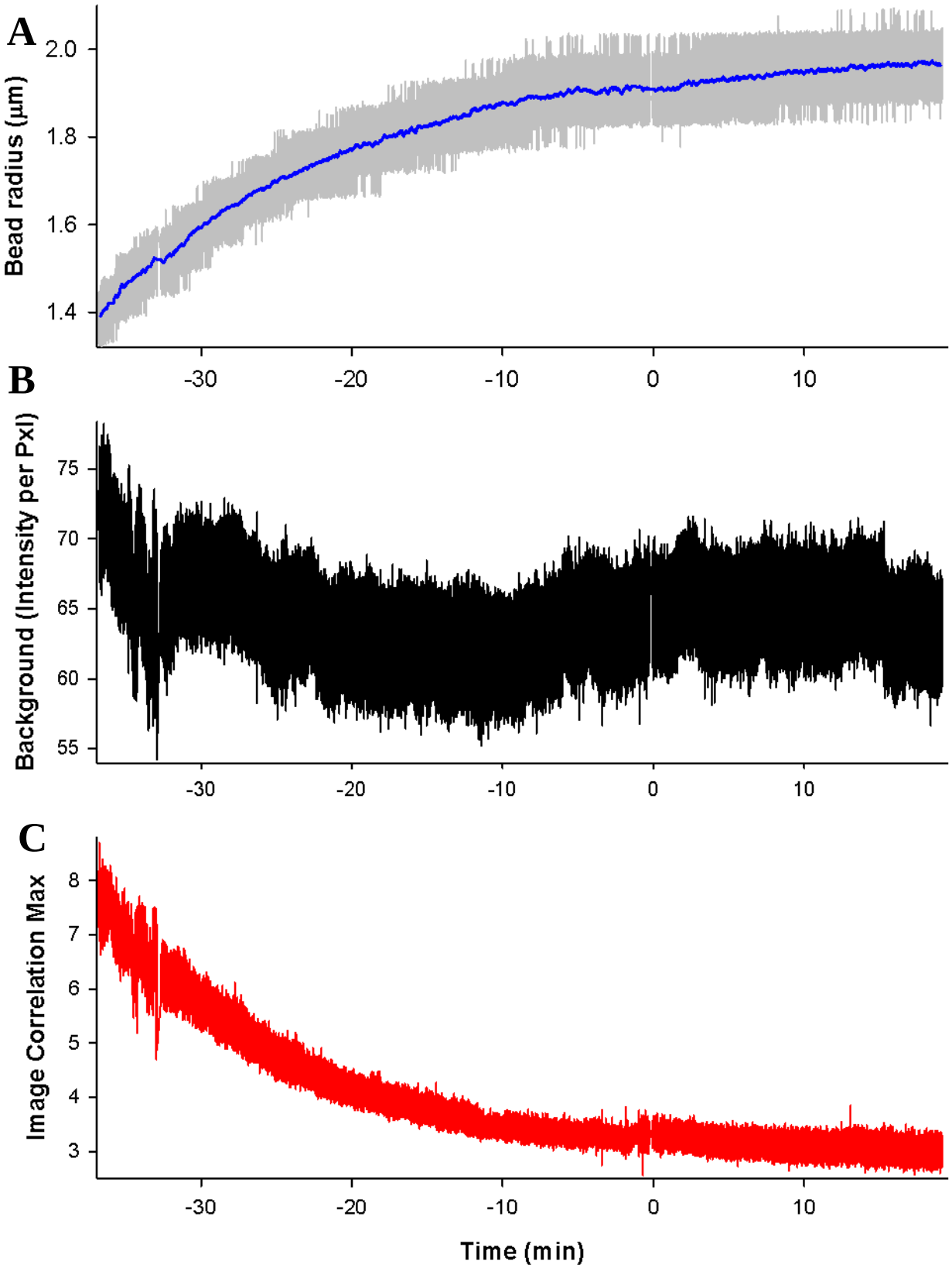

A
Bead radius (µm)
2.0
1.8
1.6
1.4
-30
-20
-10
0
10
B
Background (Intensity per Pxl)
75
70
65
60
55
-30
-20
-10
0
10
C
Image Correlation Max
8
7
6
5
4
3
-30
-20
-10
0
10
Time (min)

Figure A.3 – Quality control measures in TPM data. The top trace (A) gives the raw (gray lines) and running-averaged (blue lines) bead radius versus time. The middle trace (B) gives the image background (black lines), and the lower trace (C) shows the mirrored autocorrelation maximum (red lines), which is taken as a signal intensity, versus time.

The computations described in this section have been dubbed the "quality control" measures. Calculating the bead radius, image background and signal intensity allows exclusion of data with artifacts or abnormalities in these measurements. If no irregularities are found in these plots, then the TPM data is not thrown out. Now the analysis of bead positions as described in chapter III may be performed on this data set.

Appendix B

DNA Sequences

In this appendix, sequences of all DNAs used are listed with operators and other special sequence regions highlighted. The sequences of 9C14 and 11C12 are given below (figure B.1) with operators (symmetric ideal binding sites, abbreviated O_{SYM}, 20 base pairs (bp) each) in red, inter-looping region in green, flanking sequence in gray and primer binding regions in blue. The structure of the DNA sequence regions (color-coded) in these two constructs is described in tables B.1 & B.2.

The core sequences of 11C12, 9C14 and Unbent Control (UBC) are based sequences of the same names from the bulk competition DNA experiments of Mehta and Kahn (Ref). In this work, the flanking sequences of all DNAs used differ slightly on both sides of the inter-operator region. Mehta and Kahn's 9C14 has a total length of 352 bp, while the single-molecule 9C14 construct has end-to-end length of 669 bp. Particularly, the single-molecule 9C14 utilized in the present experiments has 181 extra bp toward the 5' end of flanking region #1; 6 bp were removed from the 3' end of flanking region #2 and 142 bp extra were added to the final construct. These additions were made for convenience of root mean-squared (rms) time series data (i.e. larger average rms values in $R(t)$ plots) and also because of concerns that the presence of a large polystyrene bead in tethered particle motion (TPM) experiments could sterically hinder

9C14 Sequence

TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC
GCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGC
AGGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCAC
TAAAGGGAACAAAAGCTGGGTACCGATATCTGCAGGTCAGTCTAGGTAAT
TGTGAGCGCTCACAATTAGATCTCAATTCGTACGGATCCGGTTTTTTGCCC
GTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCC
GTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGAATCGAAGCTAGCTA
ATTGTGAGCGCTCACAATTCGTTGTGGTAAAGCTTTGATATCAAGCTTATC
GATACCGTCGACCTCGAGGGGGGGCCGCCACCGCGGTGGAGCTCCAATTC
GCCCTATAGTGAGTCGTATTACTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATG
CGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGC
ACTCTCAGTACAA

11C12 Sequence

TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC
GCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGC
AGGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCAC
TAAAGGGAACAAAAGCTGGGTACCGATATCTGCAGGTCAGTCTAGGTAAT
TGTGAGCGCTCACAATTAGATCTCAATTCCTGTACGGATCCGCAAAAAACG
GGCAAAAAACGGCAAAAACGGGCAAAAAACGGCAAAAAACGGGCAAAAAA
CGGCAAAAAACGGGCAAAAAACCGCTACGCGTCCTAGAATCGAAGCTAGC
TAATTGTGAGCGCTCACAATTCGTTGTGGTAAAGCTTTGATATCAAGCTTA
TCGATACCGTCGACCTCGAGGGGGGGCCGCCACCGCGGTGGAGCTCCAAT
TCGCCCTATAGTGAGTCGTATTACTGGCGTAATAGCGAAGAGGCCCGCAC
CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGA
TGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGT
GCACTCTCAGTACAA

Figure B.1 – The sequences of curved DNA constructs 9C14 & 11C12. They are color-coded to highlight the different regions that are present in the DNA sequences, with primer binding sites in blue, flanking regions in gray, symmetric operators in red and the inter-looping regions in green. The sequences length for 9C14 and 11C12 are 669 and 668 base pairs (bp), respectively. In later DNAs, the template of 9C14 is used to generate sequences of identical inter-looping region and operator sequence but different flanking region sequence; the DNA total length is kept constant and the same primer binding sites are maintained. The structures of the different regions of the DNAs in this figure are detailed in tables B.1 & B.2.

the lac Repressor protein (LacI) from simultaneously binding to two operators and mediating DNA loop formation. Thus, additions were made to the flanking regions of all of the DNA constructs which originated from Mehta & Kahn (i.e. 9C14, 11C12 & Unbent Control).

DNA Sequence Region (9C14)	Region Length (bp)
Primer #1	**24**
Flanking #1	**223**
O_{SYM} #1	**20**
Inter-loop	**136**
O_{SYM} #2	**20**
Flanking #2	**221**
Primer #2	**25**
TOTAL	**669**

Table B.1 – The structure of DNA construct 9C14.The first column lists the different DNA segments (color-coded with the same scheme as figure 1) that comprise this construct, while the second gives the length of each segment in base pairs (bp). The inter-loop region length is given from end-to-end; in other sections of this work, it given from operator-center-to-operator-center.

DNA Sequence Region (11C12)	Region Length (bp)
Primer #1	**24**
Flanking #1	**223**
O_{SYM} #1	**20**
Inter-loop	**135**
O_{SYM} #2	**20**
Flanking #2	**221**
Primer #2	**25**
TOTAL	**668**

Table B.2 – The structure of DNA construct 11C12.This table is organized the same, including color-coding schemes, as table B.1. The same primer sequences are used for 9C14 and 11C12; there is only 1 bp difference in the total end-to-end lengths of these two DNA constructs.

Similar to 9C14, Mehta and Kahn's 11C12 has a total length of 351 bp, while the single-molecule 11C12 construct contains 668 bp. The single-molecule 11C12 also has 181 extra bp toward the 5' end of flanking region #1; 6 bp were also removed from the 3' end of 11C12 flanking region #2 and 142 bp extra were added to the final construct. The additional base pairs on the single-molecule versions of 9C14 and 11C12 originated from their respective plasmids by simple polymerase chain reaction (PCR) with primers that included plasmid DNA from farther toward each (5' and 3') end in the final constructs; after appropriate enzymatic cleavage of unwanted DNA in the products, the end-functionalized pre-bent DNA sequences were ready for experimental usage.

Unbent Control Sequence

GCGAGCCCATTTATACCTGAATATGGCTCATAACACCCCTTGTTTGCCTGG
CGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGA
AACGCCGTAGCGCCGATGGTAGTGTGGGGACTCCCCATGCGAGAGTAGGG
AACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCT
TTCGCCCGGGCTAATTAGGGGGTGTCGCCCTTCACACGTACTTAGTCGCT
GAAAATTGTGAGCGCTCACAATTAGATCTCAATTCCTGTACGGATCCACTG
AATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAA
CAGGCCAGCCATTACGCTCGTCATCAAAATCACTACGCGTCCTAGAATCGA
AGCTAGCTAATTGTGAGCGCTCACAATTCGTCAATCGAGTTCGTACCTAAG
GGCGACACAAAATTTATTCTAAATGCATAATAAATACTGATAACATCTTAT
AGTTTGTATTATATTTTGTATTATCGTTGACATGTATAATTTTGATATCAAA
AACTGATTTTCCCTTTATTATTTTCGAGATTTATTTTCTTAATTCTCTTTAA
CAAACTAGAAATATTGTATATACAAAAAATCATAAATAATAGATGAATAGT
TTAATTATAGGTGTTCATCAATCGAAAAAGCAACG

Figure B.2 – The sequence of DNA construct Unbent Control (UBC). Like the sequences listed in figure B.1 for 9C14 and 11C12, UBC's sequence is color-coded with the same coloring scheme to highlight the different regions that are present. In contrast to 11C12 and 9C14, this DNA lacks intrinsic curvature, and has an end-to-end length of 696 bp. The sizes of the different regions of UBC are detailed in table B.3.

Whereas the flanking regions of 9C14 and 11C12 were slightly modified, those of Unbent Control (UBC) were replaced entirely (this is due to the experimenter having to re-

order the plasmid). The flanking regions of UBC in single-molecule experiments were synthesized in PCR by simple inclusion of downstream sequences from the plasmid, while making sure to avoid any additional lac operator sites which could be present in the plasmid. The full end-to-end sequence of UBC (given in figure B.2 and detailed further in table B.3) contained 696 bp.

DNA Sequence Region (UBC)	Region Length (bp)
Primer #1	**20**
Flanking #1	**234**
O_{SYM} #1	**20**
Inter-loop	**138**
O_{SYM} #2	**20**
Flanking #2	**244**
Primer #2	**20**
TOTAL	**696**

Table B.3 – The structure of DNA construct UBC. This table is organized the same, including color-coding schemes, as tables B.1 & B.2. The sequence length of UBC is 696 bp – longer than that of either 9C14 or 11C12 by a few helical turns.

Two other DNA constructs, 9C14* AT-Rich and 9C14* CG-Rich, were based on the single-molecule sequence of 9C14. The sequences of both of these DNAs are given in figure B.3 and detailed further in tables B.4 & B.5. 9C14, 9C14* AT-Rich and 9C14* CG-Rich all have the same end-to-end length (669 bp) as 9C14. These three sequences have the same O_{SYM} operators, inter-loop region, primer binding sites and generic flanking regions. They differ only in the sequence of the "special" flanking regions (light blue in figure B.3 as well as tables B.4 & B.5), which are random, AT-Rich or CG-Rich for, respectively, 9C14, 9C14* AT-Rich and 9C14* CG-Rich. In 9C14* AT-Rich, the "special" flanking regions on the 5'

9C14* AT-Rich Sequence

TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC
GCCCAATACGCAAACTCTAGAATTTAGGCTATGAATAAATCTACGCATACA
TATAAATTGTGCAGATAAGTATCAGACCGAAATCAATACTTATATGTCTCT
AGCACTATAGATAAGCTCGGGCTAATATGAATAGGCTATTCTAGAAATTGT
GAGCGCTCACAATTAGATCTCAATTCGTACGGATCCGGTTTTTTGCCCGTT
TTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTT
TTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGAATCGAAGCTAGCTAATT
GTGAGCGCTCACAATTTCTAGAATTTAGGCTATGAATAAATCTACGCATAC
ATATAAATTGTGCAGATAAGTATCAGACCGAAATCAATACTTATATGTCTC
TAGCACTATAGATAAGCTCGGGCTAATATGAATAGGCTATTCTAGAATCGC
CCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTA
TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCT
CAGTACAA

9C14* CG-Rich Sequence

TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC
GCCCAATACGCAAACTCTAGAATTTAGGCCGCGAGCGAATCTACGCCGCC
AGCGAAATTGTGCAGCGCAGCGCCAGACCGAAATCAGCGCTCGCATGTCT
CTAGCACGCGAGCGCAGCTCGGGCTAACGCGAGCGGGCTATTCTAGAAAT
TGTGAGCGCTCACAATTAGATCTCAATTCGTACGGATCCGGTTTTTTGCCC
GTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCC
GTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGAATCGAAGCTAGCTA
ATTGTGAGCGCTCACAATTTCTAGAATTTAGGCCGCGAGCGAATCTACGCC
GCCAGCGAAATTGTGCAGCGCAGCGCCAGACCGAAATCAGCGCTCGCATG
TCTCTAGCACGCGAGCGCAGCTCGGGCTAACGCGAGCGGGCTATTCTAGA
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATG
CGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGC
ACTCTCAGTACAA

Figure B.3 – The sequences of DNA constructs 9C14* AT-Rich and 9C14* CG-Rich. Like the constructs given in figures B.1 & B.2, the two sequences above are color-coded with the same coloring scheme – with one exception; the flanking regions on the 5' and 3' sides of the inter-looping region are now divided into "generic" flanking sections which are still colored in gray and "special" regions which are colored in light blue. In the 9C14* AT-Rich sequence, both of the "special" regions that flank the operators outside the looping region are identical; this is also true for 9C14* CG-Rich. The sizes of the different regions of 9C14* AT-Rich and 9C14* CG-Rich are detailed in tables B.4 & B.5.

DNA Sequence Region (9C14* AT-Rich)	Region Length (bp)
Primer #1	**24**
Generic Flank #1	**91**
Special (AT-Rich) Flank #1	**132**
O_{SYM} #1	**20**
Inter-loop	**136**
O_{SYM} #2	**20**
Special (AT-Rich) Flank #2	**132**
Generic Flank #2	**89**
Primer #2	**25**
TOTAL	**669**

Table B.4 – The structure of DNA construct 9C14* AT-Rich. This table is organized the same, including color-coding schemes, as tables B.1, B.2 & B.3 – with one exception; the flanking region is now split into "generic" (still in gray) and "special" (light blue) regions. Including 9C14* CG-Rich, four of the seven DNAs utilized in single-molecule experiments shared the same primer binding sites.

DNA Sequence Region (9C14* CG-Rich)	Region Length (bp)
Primer #1	**24**
Generic Flank #1	**91**
Special (CG-Rich) Flank #1	**132**
O_{SYM} #1	**20**
Inter-loop	**136**
O_{SYM} #2	**20**
Special (CG-Rich) Flank #2	**132**
Generic Flank #2	**89**
Primer #2	**25**
TOTAL	**669**

Table B.5 – The structure of DNA construct 9C14* CG-Rich. This table is organized the same, including color-coding schemes, as table B.4. The CG content of the "special" flanking regions here is 62.5 %, whereas that in 9C14* AT-Rich is 32.5 %; these percentages do not include either of the two XbaI restriction sites (6 bp long each) at the beginning and ending of the "special" regions.

and 3' sides of the inter-loop region are identical; the same statement is true for the two special flanking regions of the 9C14* CG-Rich construct, but not for the original, single-molecule 9C14.

The other DNAs used in single molecule, TPM experiments are designated 702 unbent and 700 O_{SYM}; their sequences are detailed in figure B.4 and tables B.6 & B.7. 700 O_{SYM} and 702 unbent have identical flanking and inter-looping regions and differ only in their operator sequences. These constructs contain an inter-looping region that is more than twice the size of that found in Mehta and Kahn's constructs – 11C12, 9C14 or UBC (Ref). 702 unbent is a variant of a sequence used in a variety of TPM experiments with LacI performed in the Meiners lab (Ref Chen 2010 (1), Ref Chen 2010 (2), Ref Milstein 2010) and elsewhere (Ref Finzi).

702 unbent contains two identical operator binding sites with the *in vivo*, O_1 sequence ('**AATTGTGAGCGGATAACAATT**'), while 700 O_{SYM} contains two symmetric ideal operators like those found in all of the other DNAs used. The ideal operator is palindromic (sequence = '**AATTGTGAGCGCTCACAATT**'), and there is good experimental *in vitro* evidence that the dissociation constant of LacI from O_{SYM} relative to that from O_1 could be a factor of eight smaller (Ref Han, Ref Oehler 1994, Ref Sadler). No single-molecule experiments were performed on DNAs with the other two *in vivo* operator binding site sequences; designated O_2 ('**AAATGTGAGCGAGTAACAACC**') and O3 ('**GGCAGTGAGCGCAACGCAATT**'), these operators are auxiliary and only weakly bind to LacI (Ref Winter 1981 (1)). Although O_2 and O_3 enhance repression of the lac operon, repression is still possible even without these additional weak binding sites, but not without O_1 (Ref Oehler 1994).

702 Unbent Sequence

CCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGC
AGCGAGTCAGTGAGCGAGGAAGCGGAAGGCGAGAGTAGGGAACTGCCAG
GCATCAAACTAAGCAGAAGGCCCCTGACGGATGGCCTTTTTGCGTTTCTAC
AAACTCTTTCTGTGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCTT
TT**AATTGTTATCCGCTCACAATT**CCACACAACATACGAGCCGGAAGCATAA
AGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTATCAGCAT
CTAGCATGCAACCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAG
GGTGGTTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCAC
CGCCTGGAATTCAGTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGTG
A**AATTGTTATCCGCTCACAATT**CCCCGACGAGCTTCATGCCGTTAGTCGCA
CTGCAAGGGGTGTTATGAGCCATATTCAGGTATAAATGGGCTCGCGATAAT
GTTCAGAATTGGTTAATTGGTTGTAACACTGACCCCTATTTGTTTATTTTTC
TAAATACATTCAAATATGTAT**CCGCTCATGAGACAATAACCC**

700 O_{SYM} Sequence

CCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGC
AGCGAGTCAGTGAGCGAGGAAGCGGAAGGCGAGAGTAGGGAACTGCCAG
GCATCAAACTAAGCAGAAGGCCCCTGACGGATGGCCTTTTTGCGTTTCTAC
AAACTCTTTCTGTGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCTT
TT**AATTGTGAGCGCTCACAATT**CCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTATCAGCAT
CTAGCATGCAACCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCA
GGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCA
CCGCCTGGAATTCAGTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGT
GA**AATTGTGAGCGCTCACAATT**CCCCGACGAGCTTCATGCCGTTAGTCGC
ACTGCAAGGGGTGTTATGAGCCATATTCAGGTATAAATGGGCTCGCGATA
ATGTTCAGAATTGGTTAATTGGTTGTAACACTGACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTAT**CCGCTCATGAGACAATAACCC**

Figure B.4 – The sequences of DNA constructs 702 Unbent and 700 O_{sym}. The two sequences above are color-coded with the same coloring scheme as was utilized in figures B.1 & B.2. 702 unbent and 700 O_{sym} differ only in their respective operator sequences.

DNA Sequence Region (702 Unbent)	Region Length (bp)
Primer #1	**20**
Flanking #1	**183**
O_1 #1	**21**
Inter-loop	**283**
O_1 #2	**21**
Flanking #2	**153**
Primer #2	**21**
TOTAL	**702**

Table B.6 – The structure of DNA construct 702 Unbent. This table is organized the same, including color-coding schemes, as tables 1, 2 & 3. DNA constructs 702 Unbent and 700 O_{sym} (table 7) differ only in their operator sequences (O_1 vs. O_{sym}).

DNA Sequence Region (700 O_{SYM})	Region Length (bp)
Primer #1	**20**
Flanking #1	**183**
O_{SYM} #1	**20**
Inter-loop	**283**
O_{SYM} #2	**20**
Flanking #2	**153**
Primer #2	**21**
TOTAL	**700**

Table B.7 – The structure of DNA construct 700 O_{sym}. This table is organized the same, including color-coding schemes, as tables B.1, B.2, B.3 & B.6. The data obtained for DNA construct 700 O_{SYM} was the last to be measured for the present work. Both 702 Unbent and 700 O_{sym} were based on a DNA sequence from Finzi & Gelles (Ref).

Finally, five series of single-molecule competition assays (SMCA) were performed on, respectively, 9C14 (twice), Unbent Control, 702 unbent and 700 O_{SYM}. In the latter four SMCAs, the full end-to-end unlabeled sequences identical to the tethering DNA (i.e. DNAs

were synthesized with primers that lacked chemical labeling with digoxigenin or biotin molecules). However the first competition assay was performed on 9C14 with a 100 nanomolar concentration of a 40 bp oligonucleotide (abbreviated 'oligo' from this point onward); the sequence and organization of this short competitor oligo are detailed in figure B.5. This initial SMCA performed on 9C14 with 40 bp oligos was later re-done with the full end-to-end competitor sequence identical to the tethering DNA (i.e. 9C14).

Competitor Oligo Sequence

GAAGCTAGCTAATTGTGAGCGCTCACAATTCGTTGTGGTA

Figure B.5 – The structure of the 40 bp DNA oligonucleotide used in initial, single molecule DNA competition assays. This oligo, plus its complementary strand, contains two flanking regions of non-identical sequence (colored in gray), but identical length (10 bp), bracketing a single symmetric ideal operator, O_{SYM}. The reason for using an oligonucleotide at least 40 bp long is because of experimental *in vitro* evidence that LacI needs at least this minimal flanking DNA to properly bind to an operator binding site (Ref Frank).

Appendix C

Acquisition of Image Data

C.1 Introduction

In the present work, TPM – a single molecule technique – was used to explore the formation and breakdown of DNA loops. Thus far, it has been described in detail how, once image data has been acquired, the individual beads positions are computed (the particle tracking algorithm of appendix A), the positional data is filtered (chapter III, section 3.4) and from this the smoothed rms data (figure 3.9) is used to obtain a time trace of the LacI-DNA rms states. However, all of these steps began at the stage in which images of bead motion had already been acquired at video rate (i.e. approximately 30 Hz). The process of acquiring microscope images in a TPM experiment has not yet been described. The purpose of this appendix is thus to give a full description of the acquisition of TPM data from a visual light microscope.

C.2 The image acquisition GUI

Images of TPM data were acquired at video rate with a custom-made software program written in the MATLAB ® programming language. This script, which is given the function name *local_pxl_GUI_JDR_new,* uses C-language function libraries from the PixeLINK ® Software

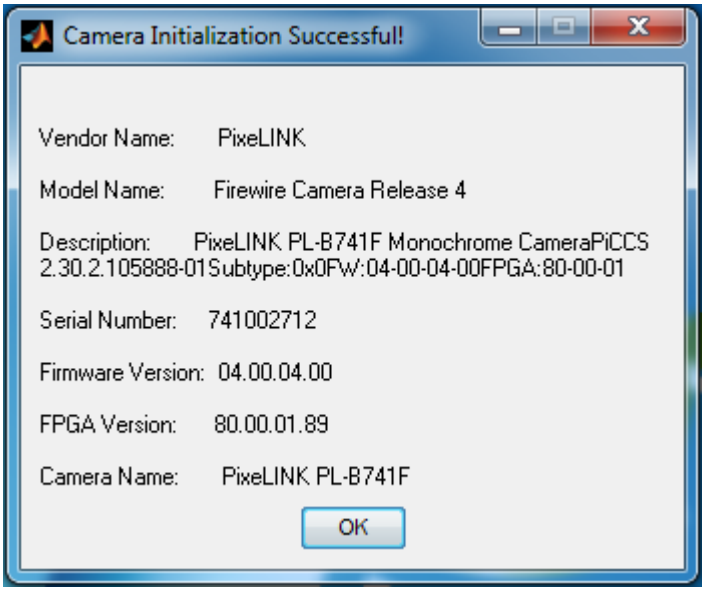


Figure C.1 – The camera info window. Upon opening the data acquisition program, this window appears – giving the description of the digital camera with which data will be acquired. The user must click "OK" before the acquisition graphical user interface (GUI) will open.

Developers Kit (SDK) to interface with the CCD camera. Before one uses this program, it is advisable to check whether the directory in which this MATLAB ® script is located is in the MATLAB ® path. Also, the C-language PixeLINK ® interface is formatted only for 32-bit versions of MATLAB ®, although future forms of this SDK software could be released for 64-bit versions of MATLAB ®.

Also, before using the custom-written image acquisition script, it is advantageous to acquaint oneself with the acquisition software, called "Capture OEM", which accompanies the PixeLINK ® camera. Specifically, it is advisable to check the various camera settings, some of which cannot be edited from the MATLAB ® script, and save the camera settings in a configuration (i.e. '.pcc') file; these settings can then be loaded again if the camera is unplugged or the various camera parameters are changed.

Lastly, the Lookup Table (LUT) settings can only be edited (at the time of this writing) from Capture OEM. The LUT maps the intensity values recorded by the camera to values saved in the image data. The LUT thus gives a precise control for image contrast and brightness; this is important for the particle tracking algorithm discussed in appendix A because proper bead positional tracking requires well-defined dark centers for each bead. Therefore, it is appropriate to adjust the LUT so that the beads have very dark centers surrounded by very bright rings.

When the user is ready to begin a TPM experiment, they must simply type in the program name, exactly as it is written above, in the MATLAB ® prompt. When this is done, a window will appear that gives a description of the digital CCD camera that will be used by this script for image acquisition (figure C.1). If the camera described in this window does not match the camera that is attached to the microscope, then all wires and cables must be checked to ensure that the microscope camera is attached to the personal computer (PC) in which the acquisition program is running. Once the user has checked the camera information, they must click the "OK" button before data acquisition can begin.

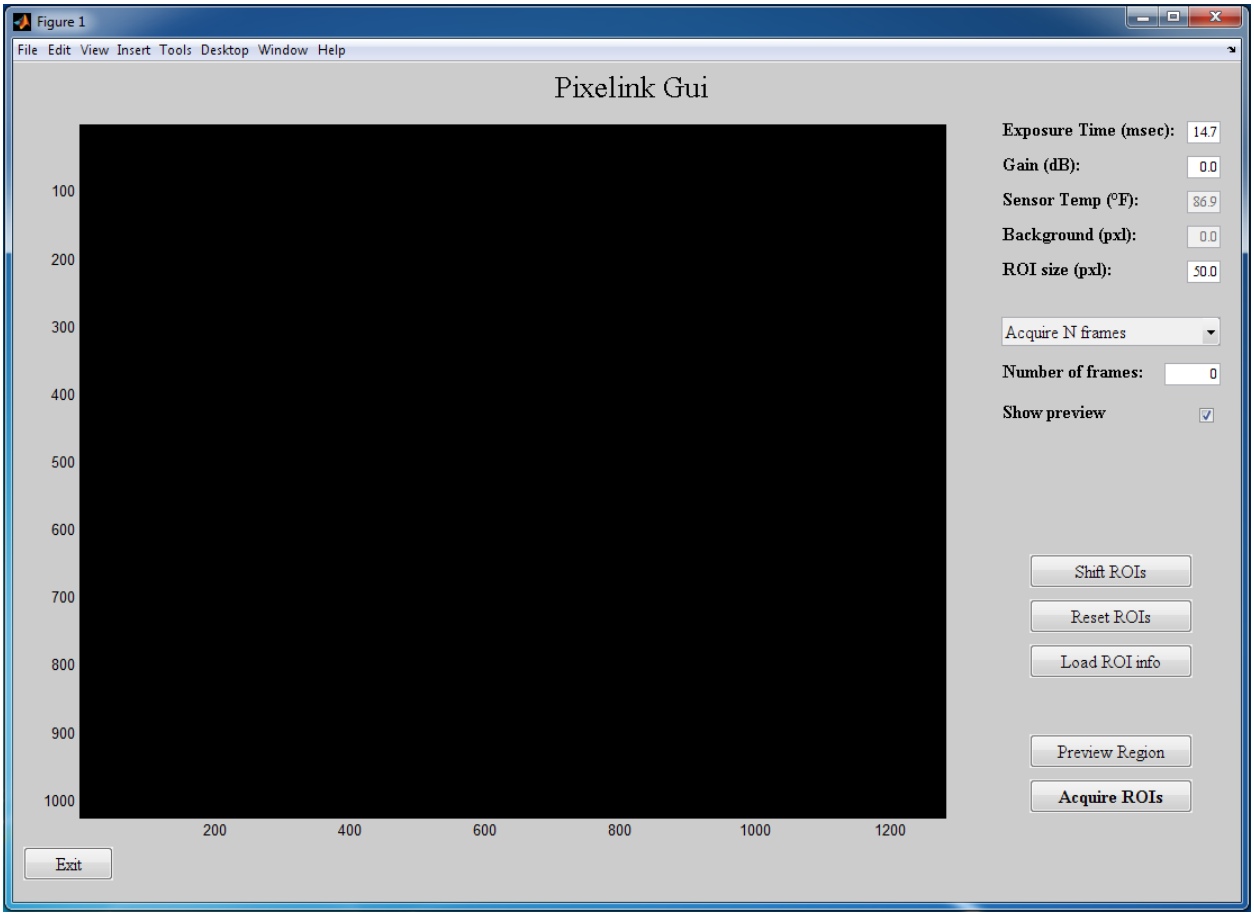


Figure C.2 – The acquisition GUI. This is the graphical user interface (GUI) from which all of the necessary camera settings and acquisition parameters can be adjusted. A variety of controls are present on the right-hand side of the GUI. The large black rectangle above is the preview window, from which a real-time preview of the flow cell observational area can be imaged via the "Preview Region" button. When the user is finished acquiring data, they must close this window by clicking the "Exit" button in the lower left-hand corner.

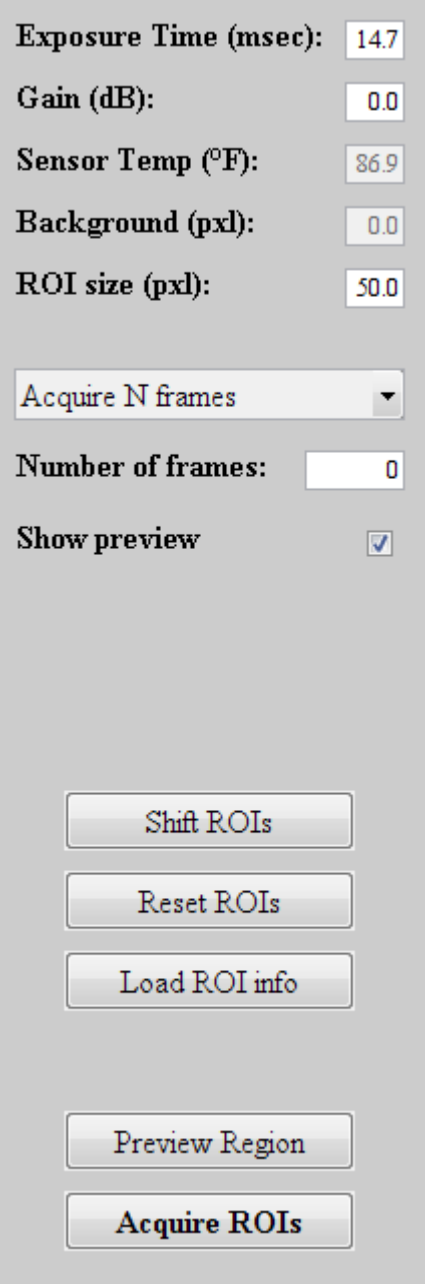


Figure C.3 – Control panel of the acquisition GUI. This enlarged view of the panel from the right-hand side of figure C.2 shows the readouts and controls of the acquisition GUI.

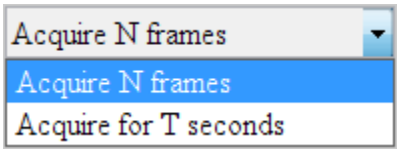


Figure C.4 – Drop-down menu from GUI control panel. This menu controls where data is acquired as a set total number of frames or as all frames in a given time interval.

Once the camera information window is closed, a new window will appear containing the graphical user interface (GUI) for data acquisition (figure C.2). The large black rectangle in the GUI is a preview pane which may be used to obtain a series of real-time images of the observational area in the flow cell; the axes of this preview pane are in units of pixels, with the full chip area being 1024 by 1280 pxl^2. In the bottom left-hand corner of the GUI is an "Exit" button, which is the only means to properly shutdown the acquisition program. On the right-hand side of the GUI is a panel of controls and readouts; this panel is displayed in a larger view in figure C.3.

The first two boxes in the GUI panel control the image exposure time, in milliseconds (msec), and the camera circuit gain, in decibels (dB). Both of these parameters can be altered by entering new values in their corresponding text boxes and pressing the "Enter" key on the PC keyboard. The third and fourth boxes of the GUI panel are readouts of the camera sensor temperature, in degrees Fahrenheit, and the average intensity value in all the pixels along the edges of a preview image, which is taken as the image background. The latter quantity varies with the illumination strength of the microscope light source. It is noted that the third and fourth

boxes are purely readouts and cannot be edited. The fifth box of the GUI panel gives the lateral size of each regions of interest (ROI), in pixels, for each bead to be acquired. This can be modified in the same way as for the first two boxes. It is noted that ROIs to be saved are square-sized regions, and thus the value of ROI size is both the x- as well as the y- dimension.

Below the top five control boxes on the right-hand panel, there is a drop-down menu (figure C.4) that controls one important aspect of TPM data acquisition. When acquiring data, this GUI script allows either (1) a set number of total frames to be saved or (2) a time interval, in seconds, to be defined during which all frames that fall in this interval should be saved. This drop-down menu manages this aspect of image acquisition and both options are interchangeable; in fact, if the second option is used then the MATLAB script converts a time interval into a set number of frames by simply multiplying the time interval (in seconds) by the camera frame rate. The exact value for the number of frames or the time interval can be entered in the box below the drop-down menu.

Underneath the drop-down menu and its corresponding box, there is a check-box that controls whether, once image data is actively being acquired, a data previewing window will appear. If checked, then every one hundred frames of acquisition the ROIs to be saved out of each frame are shown in this window (the window on the left-hand side of figure C.6). Though it is possible to preview every tenth frame or even every single frame to be saved, such previewing was found to cause many frames of image data (in some cases greater than fifty percent) to be missed (i.e. not acquired fast enough to be saved) and thus only judicious previewing of the image ROIs was found to be appropriate.

On the bottom half of the GUI panel there are five button controls. The first button, “Shift ROIs”, allows all data ROIs selected (ROI selection is described in the next section) to be

simultaneously shifted vertically or horizontally. The second button, “Reset ROIs”, clears all saved ROI positions as well as all information regarding whether the current data being acquired is part of a series of data acquisitions (a process that is discussed in the following sections). The third button, “Load ROI info”, allows the loading of all data saved in the “ROI.mat” file detailed in the following sections; this includes the ROI positions, ROI size (in pixels) as well as the above-mentioned information regarding whether or not the current acquisition is part of a series. This function is useful if data acquisition is paused or interrupted or if the GUI shutdowns for any reason. In those events, ROI info can simply be re-loaded using this button.

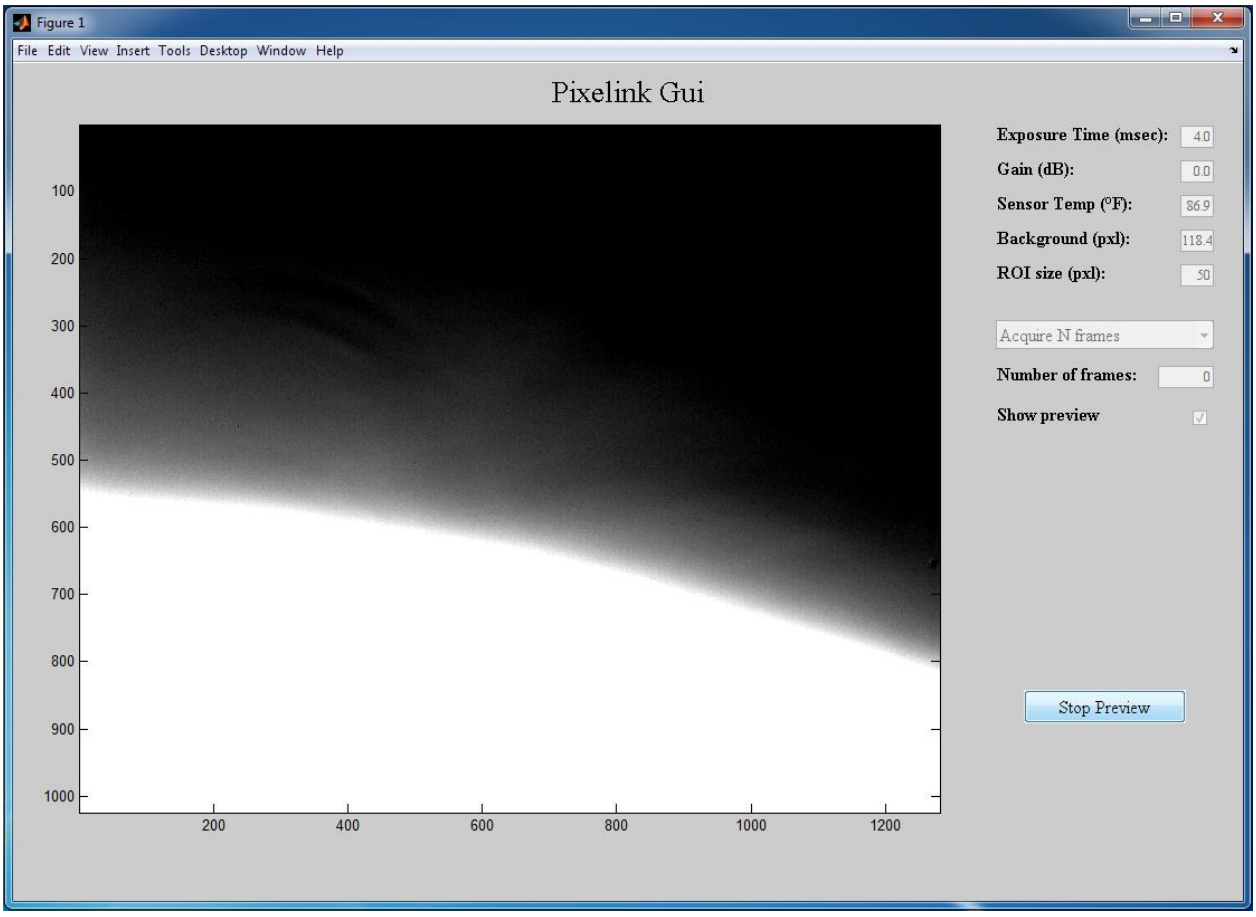


Figure C.5 – The GUI in preview mode. An image of the microscope (with nothing on it) is shown in the preview axes of the GUI window above. All of the control input boxes have been disabled and the readouts are now giving real-time information of the image background and the camera sensor temperature. The buttons in the lower half of the GUI controls on the right-hand side of the window have all disappeared, and the only remaining button, “Stop preview”, allows the user to make the GUI leave its “preview” mode.

The two bottom push-button controls are very important and will be used in every TPM experiment. The first is the “Preview Region” button, which, as described above, allows the user to put the GUI into a ‘preview’ mode (figure C.5) where images are continuously acquired and displayed on the preview axes. In this preview mode, all control input boxes are disabled and all push-button controls are hidden except for one – the “Stop Preview” button, which allows the user to exit preview mode; the “Stop Preview” button is only present while the GUI is in preview mode.

The preview mode is useful for several reasons. First, it allows adjustment of the microscope illumination level. The microscope illumination is proportional to the average intensity per pixel around the image edges (i.e. the background readout), which is displayed in the fourth box from the top. This “background” readout is activated during preview mode and gives a continuous time series of readouts for each preview frame displayed. Thus, fine tuning of the background illumination is possible in order to ensure a similar illumination level for every TPM experiment. It should be noted that in almost every TPM experiment, the exposure time was set at 4.0 milliseconds, the camera gain at 0.0 dB, the ROI size at 70 pixels and the background level at around 50-90 (which usually entailed that the illumination source was set at its maximum allowable setting). These camera settings were chosen to minimize the exposure time, allowing for a higher frame rate and lower image blur, and to minimize the gain (done because a higher gain value increases the camera noise).

The second purpose of preview mode is in the fine tuning of the lateral and horizontal positions of the observation area. Manual actuators attached to an adjustable xy-stage, to which the flow cell is attached, on the microscope setup are used to adjust this two-dimensional observational position. During adjustment of this position, it is helpful to work in the preview

mode of the acquisition GUI in order to watch as the xy-stage is moved and to precisely tune the (2D) position. A third actuator present on the microscope focusing lens allows adjustment of the camera focus (i.e. the z-position).

The last button of the GUI control panel is the "Acquire ROIs" button. This button allows the user to place the GUI in image acquisition mode. The process of defining ROIs in a preview frame and beginning data acquisition is detailed in the following section. However, before moving on to describe the image acquisition mode of the GUI, there is one final word of caution that must be given. It is strongly advised that the user only close the acquisition GUI by clicking the "Exit" button in the lower left-hand corner (figure C.2). Failure to do this can cause multiple errors in this MATLAB ® script and could necessitate either restarting the PC at which data is being acquired or unplugging the cable which connects the PixeLINK ® camera to the setup. The most common error caused by improper acquisition GUI shutdown is that the camera fails to uninitialize. The first step of the acquisition script is to initialize the camera for usage by the MATLAB ® script on the microscope PC to which the camera is attached (hence, the message in the window title of figure C.1, "Camera Initialization Successful!"). Improper shutdown thus usually means that the camera has not been uninitialized, and the simplest fix entails the following three steps. First, the user must close MATLAB ®. Second, the cable connecting the PixeLINK ® camera must be unplugged and (after a few seconds) plugged in again. This restores the camera to its default settings, including being uninitialized. Lastly, since the camera has been returned to its default settings, the user must load the camera configuration file ('.pcc') that they saved in the preparation steps detailed above. This process should restore all camera settings, including the LUT, and thus the user should be ready to restart the acquisition GUI in MATLAB ® and continue with any unfinished TPM experiment.

C.3 Acquiring TPM data using the GUI

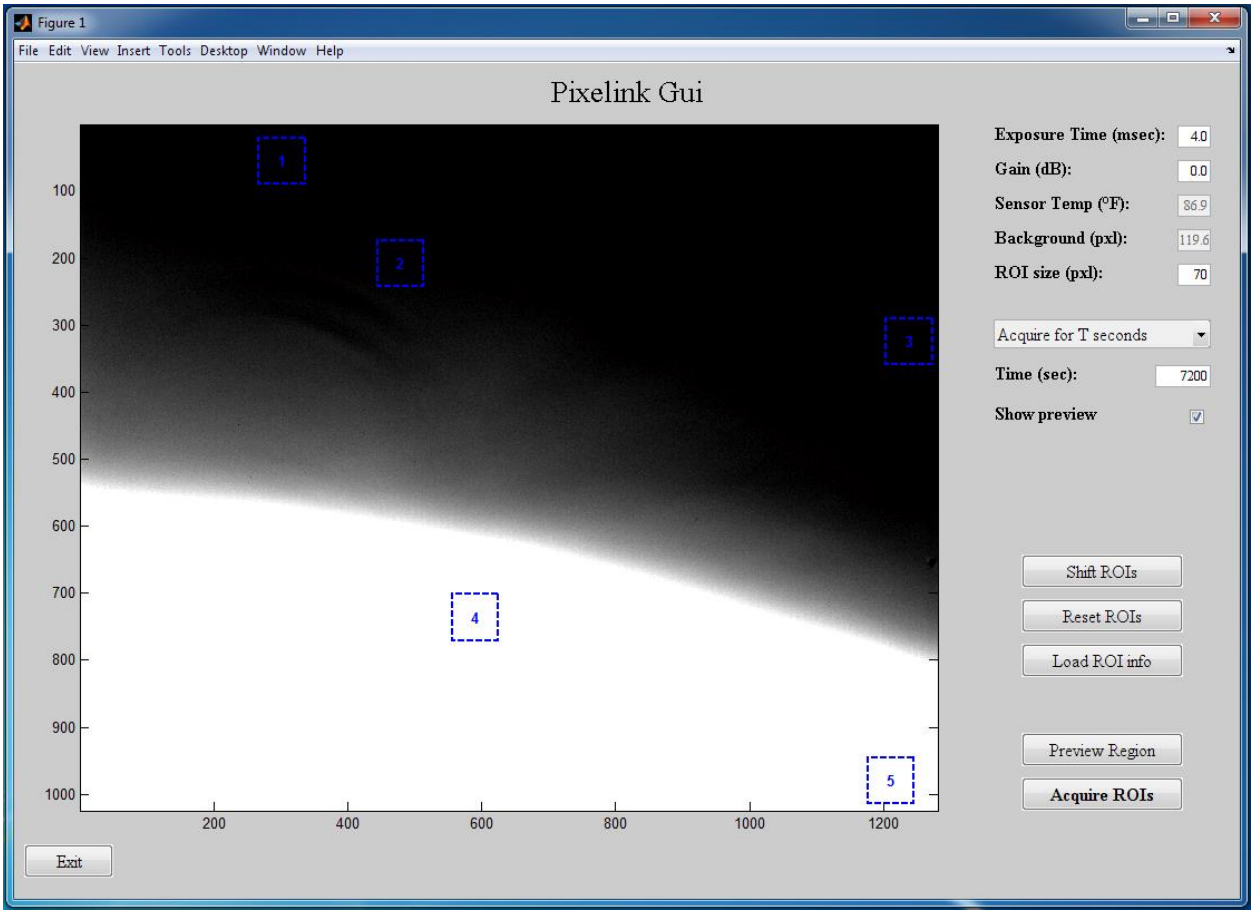


Figure C.6 – Selection of ROIs. Each dashed blue square represents a ROI to be saved. They are automatically numbered by the GUI script in order of ascending vertical position on the preview axes. Each ROI is selected by simple left-click of the PC mouse on the desired position; a ROI is then automatically created of desired size (given in the control panel on the right-hand side) centered on the position defined by mouse-click. A ROI can be deselected (i.e. removed from the list) by simply clicking inside its corresponding box (dashed blue lines).

Before any data can be acquired, ROIs must first be defined (figure C.6). In practice, each ROI corresponds to a single bead. Thus, the ROI size should be appropriately tuned so that it encompasses a minimal area for the bead's bright ring but is large enough so that it allows for the (tethered) Brownian motion of beads as well as any possible long-time positional drift or de-focusing. For 0.54 micron polystyrene beads, this value was found to be about 70 pixels (it is noted that the longer the user records data, two or more hours in the present work, the greater the possible drift and de-focusing will be).

Once the preview mode has been utilized and a frame is shown in the preview axes, appropriate ROIs can be then be selected. Although the preview frame in figure C.6 shows an empty microscope observational area, a preview frame for a TPM experiment would show many beads like that in figure A.2 in appendix A. ROIs (dashed blue lines) are selected by simply using the PC mouse to left-click on the desired position. The GUI script then automatically generates a ROI centered at this position of size given in the "ROI size" input box; a ROI can be removed by simply clicking inside its defined box (dashed blue lines). All ROIs are automatically (by the GUI script) numbered in order of increasing vertical position on the preview image.

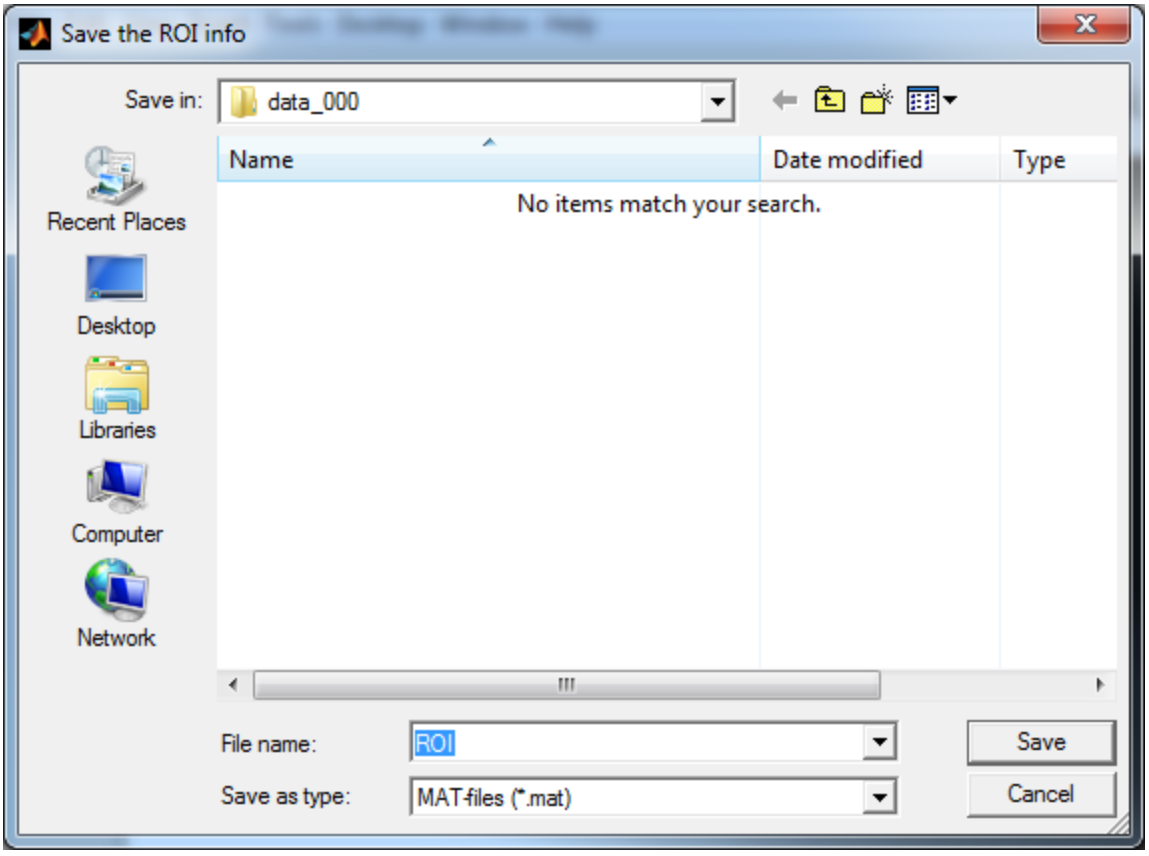


Figure C.7 – Where to save the data. When the user presses the "Acquire ROIs" button, this window appears. Here, the user chooses where to save the raw data (ROI images). It is important that the user saves the ROI file (above) with the name exactly as written in the prompt (i.e. "ROI"). Failure to do this could cause issues with the corresponding image analysis software (the algorithm of this image analysis software is detailed in appendix A).

Once the user is ready to begin image acquisition, they must click the “Acquire ROIs” button in the lower right-hand corner of the acquisition GUI. When this is done, a window will appear that prompts the user to choose the directory in which the ROI image data will be saved (figure C.7). It is important that the filename in the window prompt (“ROI”) remained unchanged. While image processing software could be altered (the algorithms used in image processing in this work are described in appendix A), its current format as of this writing requires the ROI information to be saved in a file named “ROI.mat” (this file and its contents are expanded upon in the next section).

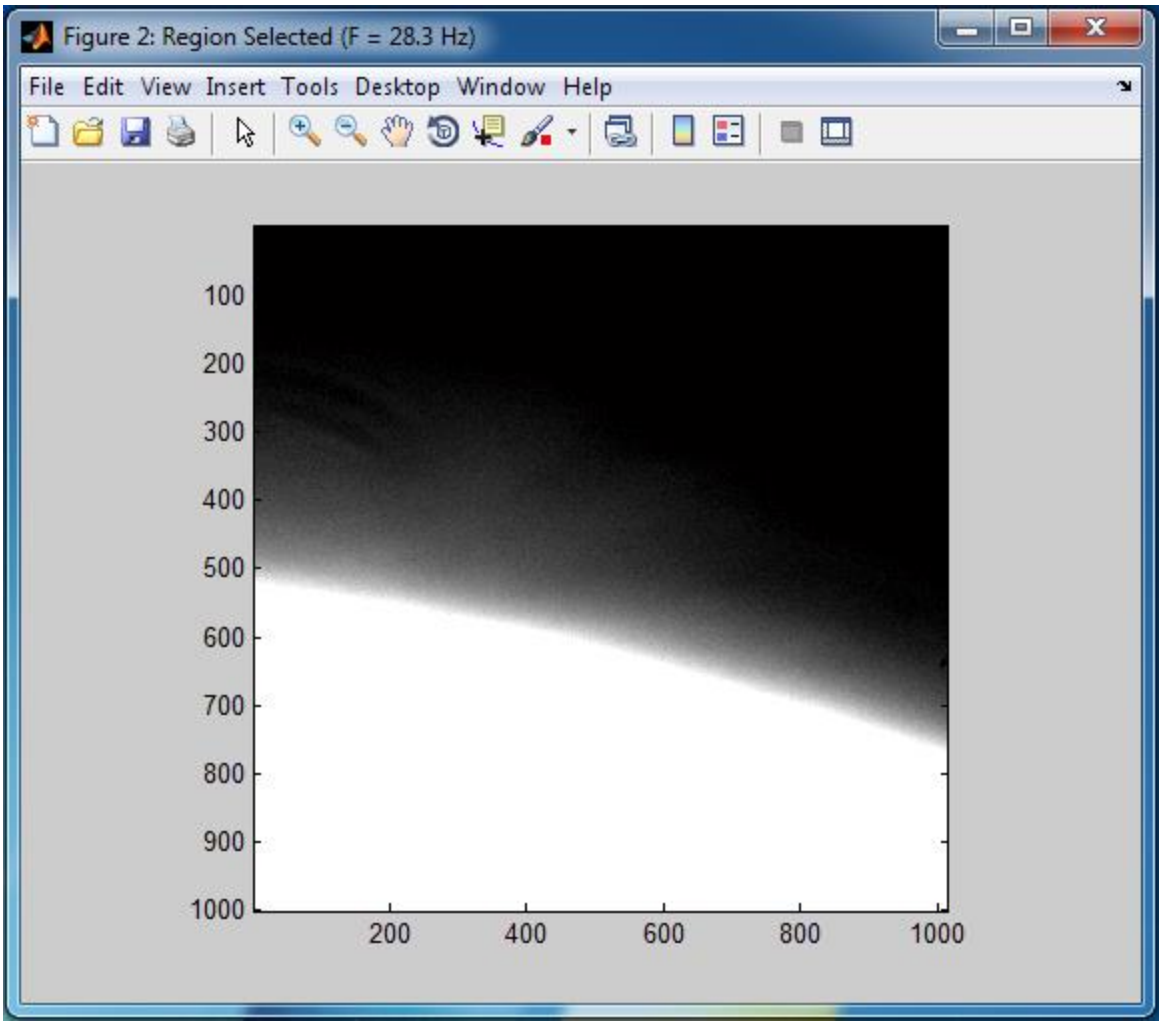


Figure C.8 – A preview of the region to be acquired. In order to acquire images at as high a frame rate (called “F” in the window title) as possible, the GUI script acquires the minimum sensor area from the CCD camera that includes all ROIs desired. That is the area shown in the axes of the above window.

Once the saving directory is chosen, and once the user has clicked the “save” button”, the acquisition GUI will disappear. Then, a series of two new windows will appear (figures C.8 & C.9). First, a window previewing the camera sensor area to be acquired is shown (figure C.8). This area is automatically chosen by the GUI script to minimize the sensor area for each frame and thus to maximize the camera’s frame rate of image acquisition. After inspection of the previewed sensor region, the user must close this window. Then, a second window will appear (figure C.9). This window gives a preview of the ROIs to be saved. This window must also be closed before image acquisition can proceed.

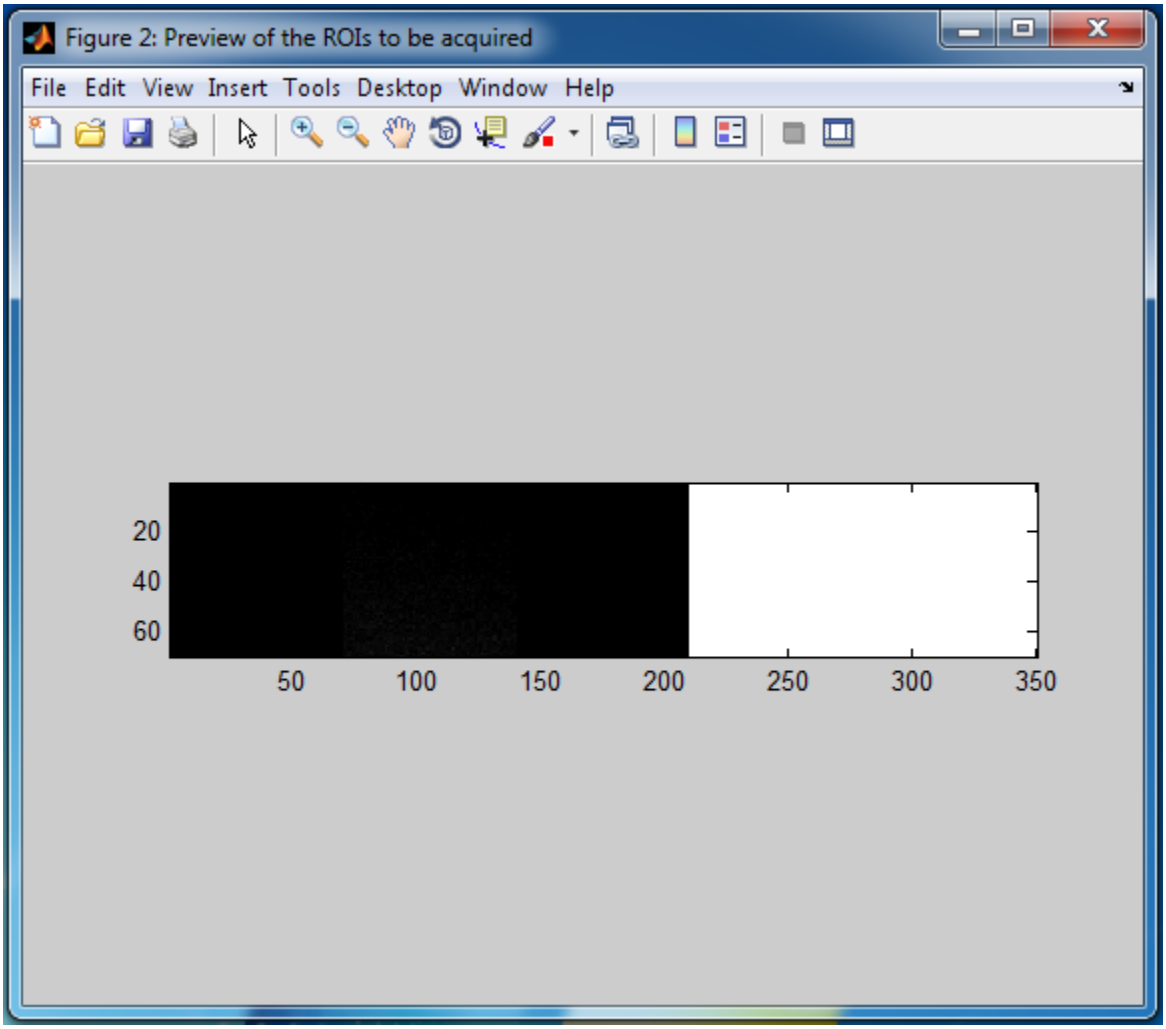


Figure C.9 – A preview of the ROIs to be acquired. Full frames of each image are not saved. Only the defined ROIs in each frame are saved. The above window is a preview of these ROIs which were defined in figure C.6.

Once the two preview windows shown in figures C.8 and C.9 have both been closed, image acquisition mode commences and ROI data in continuously saved. In this image acquisition mode, two windows are present (figures C.10 & C.11). Figure C.10 shows the previewing window for the ROIs to be saved. During image acquisition, such a previewing window serves several purposes. First, even though the ROIs from 100th frame are shown, this is sufficient to observe whether a bead has ejected for a given ROI. Second, the de-focusing and drift that occur over the long timescales of a TPM experiment can be closely monitored. Lastly, any further anomalies or microscope obstructions during image acquisition can be seen and corrected.

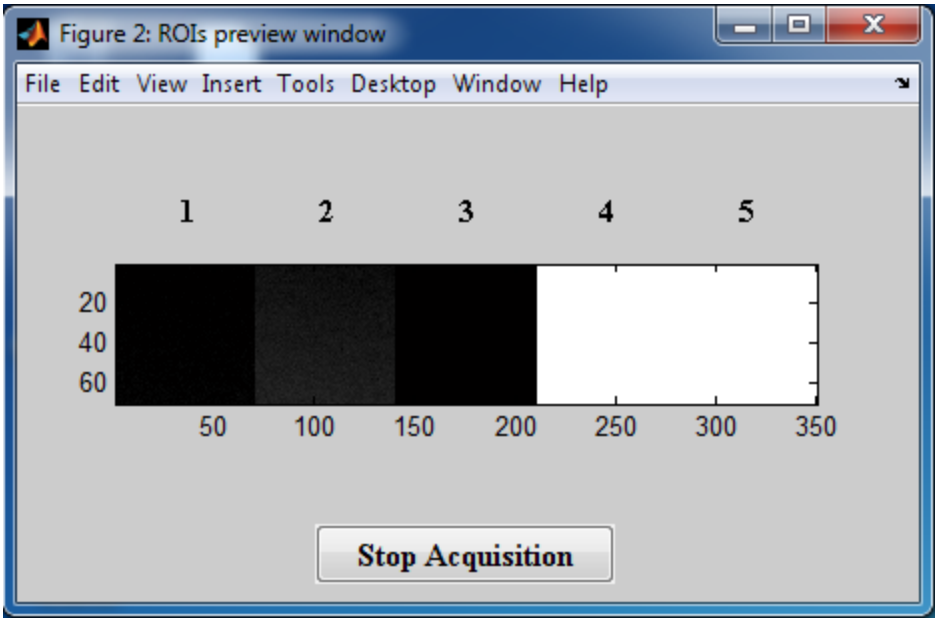


Figure C.10 – Preview of the ROIs during acquisition. During image acquisition, every 100th frame of data is used to make a preview image of the ROIs to be saved. In this window, a preview image is given of the 5 ROIs (numbered above) from figure C.6.

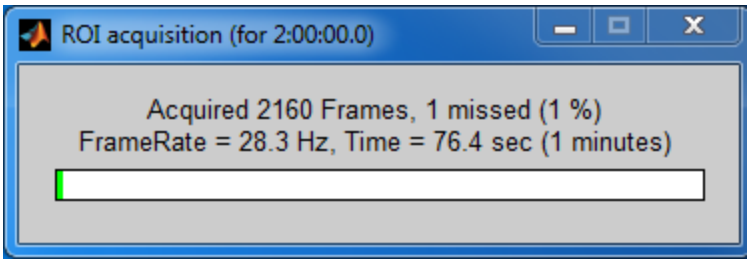


Figure C.11 – Progress bar for image acquisition. This window gives a progress bar for how much of the acquisition has been completed (plotted as a green horizontal fill in the lower empty bar at the bottom of the window).

The previewing window also contains a push-button to stop the image acquisition before it has acquired the number of frames desired (or has acquired over the full time interval inn

question). It must be noted that the GUI script only checks whether this button has been pressed once every hundred frames, so upon pressing the "Stop Acquisition" button it could take up to 5-10 seconds for the acquisition to actually cease.

The second window present in acquisition mode is a progress bar for the acquisition (figure C.11). This window contains several important pieces of information, which are updated after every ten frames of data acquired. The title of this window gives the length of acquisition (whether a time interval or a number of frames) in parentheses, which in this case was two hours. The first line of text inside this window gives the number of frames acquired so far as well as the overall acquisition progress (as a percentage). This line also gives the number of frames missed so far. The second line gives the camera frame rate, in Hertz (Hz), as well as the camera time during acquisition, in seconds. This second line also gives the experimental time in minutes, in parentheses, since the acquisition began (note this is distinct from the first time, given in seconds, in that the former is time measured by the camera and the latter is measured by the PC clock); this latter quantity is only updated after every one thousand frames of data acquired.

After acquisition is complete, both acquisition windows will close and the GUI window (figure C.6) will reappear. At this point, the user may exit the GUI program or, if desired, perform a second acquisition. If the user chooses to do the latter, then two options are available. First, the user may press the "Reset ROIs" button and choose new beads to acquire. However, if the user is performing a series of data acquisitions, they may simply use the ROIs already present. It is common for many beads to eject over the course of even a short acquisition, and there is a simple process to remove these ejected beads from the next series of acquisitions (and thus conserve hard drive space during a long experiment). After the previous acquisition, the user simply needs to left-click on each bead that has ejected. If the GUI accepts this input, the dashed

blue square around the ejected ROI should change color and become red; these ejected ROIs will not appear in future post-acquisition plots in the preview axes of the GUI window. This process can be repeated for a series of acquisitions as many times as desired.

```
>> clear all; close all; [fig_h, gui_handles, hCamera, f_alias] = local_pxl_GUI_JDR_new;

Camera Handle: 2.147484e+009
Functions alias: PixeLINK


Show preview in figure (handle = 2)


Frame 1000 MISSED (1 so far)!


Acquired 2901 frames over 102.620 sec at a framerate of 28.27 Hz, while missing 1 frames.
>>
```

Figure C.12 – Acquisition readouts in the command window. This figure shows the inputs from start to finish at the MATLAB ® prompt after data acquisition has finished. The GUI script is shown to have four output variables. The first, "fig_h", gives the figure handle of the GUI. The second, "gui_handles", gives the object handles for all graphical objects in the GUI window as a structure (this variable was quite useful for troubleshooting various design problems when this program was being written). The third variable, "hCamera", gives the handle for the CCD camera. Finally, the variable "f_alias" gives the alias for the function library of the C-language camera interface used by the acquisition GUI.

Once the GUI has been exited, the user will notice several readouts in the MATLAB ® prompt (figure C.12). It is seen that both the CCD camera handle and C-language functions alias are printed to the command window in separate lines. This occurred upon camera initialization when the GUI program first started. Also, when the GUI entered data acquisition mode, the figure handle of the previewing window was printed to the screen. During acquisition, every time a frame was missed a new line was printed to the screen showing which frame was missed as well as how many total frames had been missed by the acquisition script so far. Once acquisition mode was exited, a final line was printed showing the number of frames acquired, the acquisition time in seconds, the camera frame rate and the total number of missed frames. If this acquisition were part of a series, then the readout lines in figure C.12 would repeat for each new acquisition.

C.4 File formats and variable names

In this final section, the format of the files saved during each data acquisition is given along with a variable list with descriptions for each file that contains multiple variables. The most important file during acquisition, and by far the file with the largest byte size in the average TPM experiment, is the ROI raw images file, "ROI_images.Jdat". The format of this file is as follows. For each frame, a 32-bit unsigned integer that gives the frame number is saved; this value, divided by the camera's frame rate, corresponds to the image time used in TPM time series data plots. Following this 32-bit frame number, a matrix corresponding to the ROI images (like that in the previewing axes of figure C.10) of that frame is saved as a series of 16-bit unsigned integers. However, for the raw data the image matrix has been converted to a column vector by taking the columns of the image, from left-to-right, and saving each column. The length of each column (i.e. the image matrix vertical size) is equal to the "ROI_size" variable saved in the "ROI.mat" file, while the number of columns (i.e. the image matrix horizontal size) is equal to the number of ROIs times the "ROI_size". This format is repeated according to the number of frames, the variable "N_frames" saved in the "Framerate.mat" file, acquired in a TPM experiment. It is noted that the file "ROI_images.Jdat" is usually many gigabytes in size.

The next largest file should be called "saved_frames.Jdat". The format of this file is the same as that above, with one difference. After every ten thousand frames of acquisition, a full region is saved from the camera sensor in this file. As opposed to former method of saving the ROI data column-by-column, full regions are saved row-by-row from the top of the image to the bottom.

Finally the text file is saved containing timestamps from the PC clock after every 1000^{th} frame of data is acquired. The format of the data is N by 6, where is N is the total number of

timestamps. The six columns correspond to the year, month, day, hour (24-hour format), minute and second (accurate to three decimal places, and thus to milliseconds) given by the PC time.

Name of variable stored in "ROI.mat"	Variable description
ROI_edge_mat	A matrix of size N_{ROIs} by 4 which defines the edges of ROI in units of pixels; the 4 columns correspond to (left-to-right) ROI left edge position, ROI right edge position, ROI top edge position and ROI bottom edge position.
ROI_size	As its name suggests, this gives the ROI edge size in units of pixels.
check_prev_ROI	This is a logical (i.e. true/false) value telling whether or not the data from the current acquisition was part of a series of acquisitions that were connected to the previously acquired data.
Region_position	A four element row vector giving (from left-to-right) the x- and y- positions of the top left corner of the acquisition region as well as the horizontal and vertical dimensions of this region.
n_pxl	The total number of pixels in the region of acquisition.
n_ROI_more	This tells the number of ROIs added during the current data acquisition; this variable is necessary for an acquisition series in which ROIs are added and excised at each new stage of acquisition.
xy_ROI	A matrix of size N_{ROIs} by 2 which gives the ROI positions as defined by the user's clicks on the preview axes; the left and right columns correspond to the x- and y- positions for each ROI.
ejected_elems	This vector is a simple list of all ejected beads from the previous data acquisition, assuming that this data is part of an acquisition series; if it is not, then this variable should be empty.

Table C.1 – Description of "ROI.mat". All variables saved in this MATLAB ® format file are listed and detailed.

Name of variable stored in "Framerate.mat"	Variable description
Region_position	Same variable as that described in table C.1.
n_pxl	Same variable as that described in table C.1.
exposure_time	Camera exposure time in milliseconds.
camera_gain	Camera circuit gain in decibels (dB).
camera_temp	Camera sensor temperature in degrees Celsius.
N_frames	Number of frames acquired in this data set.
n_missed	Number of frames missed in during data acquisition.
FrameRate	Frame rate as computed by the camera firmware.
FrameRate_computed	Frame rate as computed by dividing N_frames by the acquisition time (according to the PC internal clock).
adjust_framerate	A logical (i.e. true/false) value that gives a value of "true" when the frame rates calculated by the above two methods have a percent difference greater than 0.5 %.

Table C.2 – Description of "Framerate.mat". All variables saved in this file are listed and detailed.

The final two files that are saved are named "ROI.mat" and 'Framerate.mat". These two MATLAB ® files both contain multiple saved variables corresponding to different camera and data acquisition parameters. The variables contained in the files "ROI.mat" and "Framerate.mat" are described, respectively, in tables C.1 and C.2.

References

Robert A. Alberty & Gordon G. Hammes, "Application of the theory of diffusion-controlled reactions to enzyme kinetics", *Journal of Physical Chemistry* **62** (1958), 154-159.

Anderj Barbič, Daniel P. Zimmer & Donald M. Crothers, "Structural origins of adenine-tract bending", *Proceedings of the National Academy of Sciences USA* **100** (2003), 2369-2373.

Otto G. Berg, Robert B. Winter & Peter H. von Hippel, "Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory", *Biochemistry* **20** (1981), 6929-6948.

Seth Blumberg, Arivalagan Gajraj, Matthew W. Pennington & Jens-Christian Meiners, "Three-Dimensional Characterization of Tethered Microspheres by Total Internal Reflection Fluorescence Microscopy", *Biophysical Journal* **89** (2005), 1272-1281.

Seth Blumberg, Alexei V. Tkachenko & Jens-Christian Meiners, "Disruption of Protein-Mediated DNA Looping by Tension in the Substrate DNA", *Biophysical Journal* **88** (2005), 1692-1701.

Michael Brenowitz, Amy Pickar & Elizabeth Jamison, "Stability of a Lac Repressor Mediated "Looped Complex"", *Biochemistry* **30** (1991), 5986-5998.

Carlos Bustamante, John F. Marko, Eric D. Siggia & S. Smith, "Entropic elasticity of lambda-phage DNA", *Science* **265** (September 9, 1994), 1599-1600.

Lucy E. Catto, Sumita Ganguly, Susan E. Milsom, Abigail J. Welsh & Stephen E. Halford, "Protein assembly and DNA looping by the FokI restriction endocnuclease", *Nucleic Acids Research* **34** (2006), 1711-1720.

Yih-Fan Chen, David P. Wilson, Krishnan Raghunathan & Jens-Christian Meiners, "Entropic boundary effects on the elasticity of short DNA molecules", *Physical Review E* **80** (2009), 020903.

Yih-Fan Chen, J. N. Milstein & Jens-Christian Meiners, "Femtonewton Entropic Forces Can Control the Formation of Protein-Mediated DNA Loops", *Physical Review Letters* **104** (2010), 048301.

Yih-Fan Chen, J. N. Milstein & Jens-Christian Meiners, "Protein-Mediated DNA Loop Formation and Breakdown in a Fluctuating Environment", *Physical Review Letters* **104** (2010), 258103.

Ashok A Deniz, Samrat Mukhopadhyay & Edward A Lemke, "Single-molecule biophysics: at the interface of biology, physics and chemistry", *Journal of the Royal Society Interface* **5** (2008),

Josef Deutscher, "The mechanisms of carbon catabolite repression in bacteria", *Current Opinion in Microbiology* **11** (2008), 87-93.

Mario A. Díaz de la Rosa, Elena F. Koslover, Peter J. Mulligan & Andrew J. Spakowitz, "Dynamic Strategies for Target-Site Search by DNA-Binding Proteins", *Biophysical Journal* **98** (2010), 2943-2953.

Shane C. Dillon & Charles J. Dorman, "Bacterial nucleoid-associated proteins, nucleoid structure and gene expression", *Nature Reviews Microbiology* **8** (2010), 185-195.

Charles J. Dorman, "DNA Supercoiling and Bacterial Gene Expression", *Science Progress* **89** (2006), 151-66.

Laurence M. Edelman, Raymond Cheong & Jason D. Kahn, "Fluorescence Resonance Energy Transfer over ~130 Basepairs in Hyperstable Lac Repressor-DNA Loops", *Biophysical Journal* **84** (2003), 1131-1145.

Hsiu-Fang Fan, "Real-time single-molecule tethered particle motion experiments reveal yje kinetics and mechanism of Cre-mediated site-specific recombination", *Nucleic Acids Research* **40** (2012), 6208-6222.

Reimund Fickert & Benno Müller-Hill, "How Lac Repressor finds *lac* Operator *in vitro*", *Journal of Molecular Biology* **226** (1992), 59-68.

Laura Finzi & Jeff Gelles, "Measurement of Lactose Repressor-Mediated Loop Formation and Breakdown in Single DNA Molecules", *Science* **267** (20 January, 1995), 378-380.

Diane E. Frank, Ruth M. Saecker, Jeffrey P. Bond, Michael W. Capp, Oleg V. Tsodikov, Sonya E. Melcher, Mark M. Levandoski & M. Thomas Record, Jr., "Thermodynamics of the Interactions of *Lac* Repressor with Variants of the Symmetric *Lac* Operator: Effects of Converting a Consensus Site to a Non-specific Site", *Journal of Molecular Biology* **267** (1997), 1186-1206.

Kathy A. Goodson, Zifan Wang, Aaron R. Haeusler, Jason D. Kahn & Douglas S. English, "LacI-DNA-IPTG Loops: Equilibria among Conformations by Single-Molecule FRET", *The Journal of Physical Chemistry B*, in press.

Sachin Goyal, Todd Lillian, Seth Blumberg, Jens-Christian Meiners, Edgar Meyhöfer & N. C. Perkins, "Intrinsic Curvature of DNA Influences LacR-Mediated Looping", *Biophysical Journal* **93** (2007), 4342-4359.

William J. Greenleaf, Michael T. Woodside & Steven M. Block, "High-Resolution, Single-Molecule Measurements of Biomolecular Motion", *Annual Review of Biophysics and Biomolecular Structure* **36** (2007), 171-190.

T. Ha, T. Enderle, D.F. Ogletree, D.S. Chemla, P.R. Selvin & S. Weiss, "Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor", *Proceedings of the National Academy of Sciences USA* **93** (1996), 6264-6268.

Aaron R. Haeusler, Kathy A. Goodson, Todd D. Lillian, Xiaoyu Wang, Sachin Goyal, Noel C. Perkins & Jason D. Kahn, "FRET Studies of a landscape of Lac repressor-mediated DNA loops", *Nucleic Acids Research* **40** (2012), 4432-4445.

Paul J. Hagerman, "Flexibility of DNA", *Annual Review of Biophysics and Biophysical Chemistry* **17** (1988), 265-286.

Petter Hammar, Prune Leroy, Anel Mahmutovic, Erik G. Marklund, Otto G. Berg & Johan Elf, "The *lac* Repressor Displays Facilitated Diffusion in Living Cells", *Science* **336** (2012), 1595-1598.

Lin Han, Hernan G. Garcia, Seth Blumberg, Kevin B. Towles, John F. Beausang, Philip C. Nelson & Rob Phillips, "Concentration and Length Dependence of DNA Looping in Transcriptional Regulation", *PLoS ONE* **4** (2009), e5621.

Jeff Hasty, David McMillen & J.J. Collins, "Engineered gene circuits", *Nature* **420**, 224-230 (14 November 2002).

Andrew D. Hirsh, Todd D. Lillian, Troy A. Lionberger & N. C. Perkins, "DNA Modeling Reveals an Extended Lac Repressor Conformation in Classic In Vitro Binding Assays", *Biophysical Journal* **101** (2011), 718-726.

Francois Jacob & Jacques Monod, "Genetic regulatory mechanisms in the synthesis of proteins", *Journal of Molecular Biology* **3** (June 1961), 318-356.

Stephanie Johnson, Martin Lindén & Rob Phillips, "Sequence dependence of transcription factor-mediated DNA looping", *Nucleic Acids Research* **40** (2012), 7728-7738.

Alan Kandinov, Krishnan Raghunathan & Jens-Christian Meiners, "Using DNA Looping to Measure Sequence Dependent DNA Elasticity", *Proceedings of SPIE* **8458** (2012), 845816.

Elena F. Koslover & Andrew J. Spakowitz, "Twist and Tension-Mediated Elastic Coupling between DNA-Binding Proteins", *Physical Review Letters* **102** (2009), 178102.

O. Kratky & G. Porod, "Röntgenuntersuchung Gelöster Fadenmoleküle", *Recueil des Travaux Chimiques des Pays-Bas* **686** (1949), 1106-1122.

F. Lankaš, N. Špačková, M. Moakher, P. Enkhbayar & J. Šponer, "A measure of DNA bending in nucleic acids structures applied to A-tract DNA", *Nucleic Acids Research* **38** (2010), 3414-22.

Filip Lankaš, Jiří Šponer, Pavel Hobza & Jörg Langowski, "Sequence-dependent Elastic Properties of DNA", *Journal of Molecular Biology* **299** (2000), 695-709.

Mark M. Levandoski, Oleg V. Tsodikov, Diane E. Frank, Sonya E. Melcher, Ruth M. Saecker & M. Thomas Record Jr., "Cooperative and Anticooperative Effects in Binding of the First and Second Plasmid O_{SYM} Operators to a LacI Tetramer: Evidence for Contributions of Non-operator DNA Binding by Wrapping and Looping", *Journal of Molecular Biology* **260** (1996), 697-717.

David Levens & Craig J Benham, "DNA stress and strain, *in silico*, *in vitro* and *in vivo*", Physical Biology **8** (2011), 035011.

Image from the RCSB PDB (www.rcsb.org) of PDB ID 1LBI (M. Lewis et. al., *Science* 1996).

M. Lewis, G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan & P. Lu, "Crystal Structure of the lactose operon repressor and its complexes with DNA and inducer", *Science* **271** (March 1, 1996), 1247-1254

Todd D. Lillian, Sachin Goyal, Jason D. Kahn, Edgar Meyhöfer & N. C. Perkins, "Computational Analysis of Looping of a Large Family of Highly Bent DNA by LacI", *Biophysical Journal* **95** (2008), 5832-5842.

Shih-Wei Liu, Jen-Fei Chu, Cheng-Ting Tsai, Hung-Chih Fang, Ta-Chau Chang & Hung-Wen Li, "Assaying the binding strength of G-quadruplex ligands using single-molecule TPM experiments", *Analytical Biochemistry*, in press.

Carlo Manzo, Chiara Zurla, David D. Dunlap & Laura Finzi, "The Effect of Nonspecific Binding of Lambda Repressor on DNA Looping Dynamics", *Biophysical Journal* **103** (2012), 1753-1761.

John F. Marko & Eric D. Siggia, "Stretching DNA", *Macromolecules* **28** (1995), 8759-8770.

Ruchi A. Mehta & Jason D. Kahn, "Designed Hyperstable Lac Repressor-DNA Loop Topologies Suggest Alternative Loop Geometries", *Journal of Molecular Biology* **294** (1999), 67-77.

Sarah Meinhadt, Michael W. Manley Jr, Nicole A. Becker, Jacob A. Hessman, L. James Maher III & Liskin Swint-Kruse, "Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression", *Nucleic Acid Research* **40** (2012), 11139-11154.

J. N. Milstein, Y. F. Chen & J.-C. Meiners, "Bead Size Effects on Protein-Mediated DNA Looping in Tethered-Particle Motion Experiments", *Biopolymers* **95** (2010), 144-150.

J. N. Milstein and J.-C. Meiners, "On the role of DNA biomechanics in the regulation of gene expression", *Journal of the Royal Society Interface* **8** (2011), 1673-1681.

Michael A. Morgan, Kenji Okamoto, Jason D. Kahn & Douglas S. English, "Single-Molecule Spectroscopic Determination of Lac Repressor-DNA Loop Conformation", *Biophysical Journal* **89** (2005), 2588-2596.

M.C. Mossing & M.T. Record, Jr., "Upstream Operators Enhance Repression of the Promoter", *Science* **233** (August 22, 1986), 889-892.

Johannes Müller, Stefan Oehler & Benno Müller-Hill, "Repression of *lac* Promoter as a Function of Distance, Phase and Quality of an Auxiliary *lac* Operator", *Journal of Molecular Biology* **257** (1996), 21-29.

David L. Nelson & Michael M. Cox, *Lehninger Principles of Biochemistry*, 4th edition, New York, NY: W. H. Freeman and Company (2005), 1093-1094.

Guy Nir, Moshe Lindner, Heidelinde R. C. Dietrich, Olga Girshevitz, Constantinos E. Vorgias & Yuval Garini, "HU Protein Induces Incoherent DNA Persistence Length", *Biophysical Journal* **100** (2011), 784-790.

Davide Normanno, Francesco Vanzi & Francesco Saverio Pavone, "Single-molecule manipulation reveals supercoiling-dependent modulation of *lac* repressor-mediated DNA looping", *Nucleic Acids Research* **36** (2008), 2505-2513.

Stefan Oehler, Regina Alex & Andrew Barker, "Is Nitrocellulose Filter Binding Really a Universal Assay for Protein-DNA Interactions?", *Analytical Biochemistry* **268** (1999), 330-336.

Stefan Oehler, Michele Amouyal, Peter Kolkhof, Brigette von Wilcken-Bergmann & Benno Müller-Hill, "Quality and position of the three *lac* operators of *E. coli* define efficiency of repression", *The EMBO Journal* **13** (1994), 3348-3355.

Stefan Oehler, Elizabeth R. Eismann, Helmut Krämer & Benno Müller-Hill, "The three operators of the *lac* operon cooperate in repression", *The EMBO Journal* **9** (1990), 973-979.

Jennifer Oullette, "Switching from Physics to Biology", *The Industrial Physicist* **9** (2003), 20-23.

Zhong Qian, Emilios K. Dimitriadis, Rotem Edgar, Prahathees Eswaramoorthy & Sankar Adhya, "Galactose repressor mediated intersegmental chromosomal connections in *Escherichia coli*", *Proceedings of the National Academy of Sciences USA* **109** (2012), 11336-11341.

Rob Phillips & Stephen R. Quake, "The Biological Frontier of Physics", *Physics Today* **59** (2006), 38-43.

Thomas Plénat, Catherine Tardin, Philippe Rousseau & Laurence Salomé, "High-throughput single-molecule analysis of DNA-protein interactions by tethered particle motion", *Nucleic Acids Research* **40** (2012), e89.

Krishnan Raghunathan, Alan Kandinov & Jens-Christian Meiners, *unpublished*

Steven L. Roderick, "The *lac* operon galactoside acetyltransferase", *Comptes Rendus Biologies* **328** (2005), 568-575.

Catherine A. Royer, Artemis E. Chakerian & Mathleen S. Matthews, "Macromolecular Binding Equilibria in the *lac* Repressor System: Studies Using High-pressure Fluorescence Spectroscopy", *Biochemistry* **29** (1990), 4959-4966.

Danielis Rutkauskas, Hongli Zhan, Kathleen S. Matthews, Francesco S. Pavone & Francesco Vanzi, "Tetramer opening in LacI-mediated DNA looping", *Proceedings of the National Academy of Sciences USA* **106** (2009), 16627-16632.

John R. Sadler, Henri Sasmor & Joan L. Betz, "A perfectly symmetric *lac* operator binds the *lac* repressor very tightly", *Proceedings of the National Academy of Sciences USA* **80** (1983), 6785-6789.

M. Santillán, "Bistable Behavior in a Model of the lac Operon in Escherichia Coli with Variable Growth Rate", *Biophysical Journal* **94** (2008), 2065-2081.

Nuno C. Santos & Miguel A.R.B. Castanho, "An overview of the biophysical applications of atomic force microscopy", *Biophysical Chemistry* **107** (2004), 133-149.

Dorothy A. Schafer, Jeff Gelles, Michael P. Sheetz & Robert Landick, "Transcription by single molecules of RNA polymerase observed by light microscopy", *Nature* **352** (August 1, 1991), 444-448.

R. Schleif, "DNA looping", *Annual Review of Biochemistry* **61** (1992), 199-223.

Gerhard J. Schutz, Max Sonnleitner, Peter Hinterdorfer & Hansgeorg Schindler, "Single molecule microscopy of biomembranes", *Molecular Membrane Biology* **17** (2000), 17-29.

Darren E. Segall, Philip C. Nelson & Rob Phillips, "Volume-Exclusion Effects in Tethered-Particle Experiments: Bead Size Matters", *Physical Review Letters* **96** (2006), 088306.

Jiro Shimada & Hiromi Yamakawa, "Ring-Closure Probabilities for Twisted Wormlike Chains: Application to DNA", *Macromolecules* **17** (1984), 689-698.

S.B. Smith, L. Finzi & C. Bustamante, "Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads", *Science* **258** (November 13, 1992), 1122-1126.

Daniel Strahs & Tamar Schlick, "A-Tract Bending: Insights into Experimental Structures by Computational Methods", *Journal of Molecular Biology* **301** (2000), 643-663.

Kerren K. Swinger, Kathryn M. Lemberg, Ying Zhang & Phoebe A. Rice, "Flexible DNA bending in HU-DNA cocrystal structures", The EMBO Journal **22** (July 15, 2003), 3749-3760.

Liskin Swint-Kruse & Kathleen S Matthews, "Allostery in the LacI/GalR family: variations on a theme", *Current Opinion in Microbiology* **12** (2009), 129-137.

Marc Taraban, Hongli Zhan, Andrew E. Whitten, David B. Langley, Kathleen S. Matthews, Liskin Swint-Kruse & Jill Trewhella, "Ligand-induced Conformational Changes and Conformational Dynamics in the Solution Structure of the Lactose Repressor Protein", *Journal of Molecular Biology* **376** (2008), 466-481.

D. Thirumalai, Edward P. O'Brien, Greg Morrison & Changbong Hyeon, "Theoretical Perspectives on Protein Folding", *Annual Review of Biophysics* **39** (2010), 159-183.

Oleg V. Tsodikov, Ruth M. Saecker, Sonya E. Melcher, Mark M. Levandoski, Diane E. Frank, Michael W. Capp & M. Thomas Record Jr., "Wrapping of Flanking Non-Operator DNA in *lac* Repressor-Operator Complexes: Implications for DNA looping", *Journal of Molecular Biology* **294** (1999), 639-655.

Michael J. Weickert & Sankhar Adhya, "The galactose regulon of *Escherichia coli*", *Molecular Microbiology* **10** (1993), 245-251.

C.J. Wilson, H. Zhan, L. Swint-Kruse & K.S. Matthews, "The lactose repressor system: paradigms for regulation, allosteric behavior, and protein folding", *Cellular and Molecular Life Sciences* **64** (2007), 3-16.

David P. Wilson, "A Generalized Theory of DNA Loop Formation: Considering the effects of Binding Topology and Sequence-Dependent Curvature", University of Michigan (2010), PhD Thesis.

Robert B. Winter & Peter H. von Hippel, "Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The *Escherichia Coli* Repressor-Operator Interaction: Equilibrium Measurements", *Biochemistry* **20** (1981), 6948-6960.

Robert B. Winter, Otto G. Berg & Peter 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** (1981), 6961-6977.

Oi Kwan Wong, Martin Guthold, Dorothy A. Erie & Jeff Gelles, "Interconvertible Lac Repressor-DNA Loops Revealed by Single-Molecule Experiments", *PLoS Biology* **6** (2008), 2028-2042.

Y Zhao, S Inayat, D A Dikin, J H Singer, R S Ruoff & J B Troy, "Patch clamp technique: review of the current state of the art and potential contributions from nanoengineering", *Proceedings of the Institution of Mechanical Engineers, Part N: Journal of Nanoengineering and Nanosystems* **222** (2008), 1-11.

Guozhi Zhou, Ming-Tat Wong & Guo-Qiang Zhou, "Diffusion-controlled reactions of enzymes: An approximate analytic solution of Chou's model", *Biophysical Chemistry* **18** (1983), 125-132.

Chiara Zurla, Calro Manzo, David Dunlap, Dale E. A. Lewis, Sankar Adhya & Laura Finzi, "Direct demonstration and quantification of long-range DNA looping by the λ bacteriophage repressor", *Nucleic Acids Research* **37** (2009), 2789-2795.