

# ***In vitro* DNA mechanics in Gene Regulation: One Molecule at a Time**

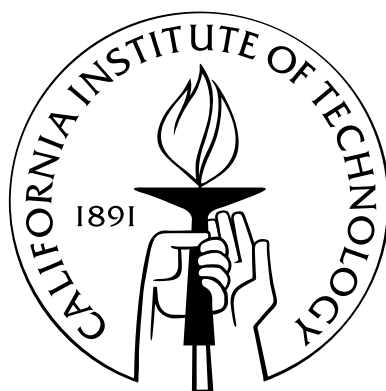
Thesis by

Lin Han

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2008

(Defended 10 July 2007)

© 2008

Lin Han

All Rights Reserved

# Acknowledgements

First of all I would like to thank my advisor, Rob Phillips, for his incredible support in the past six years. Doing experiments in his lab has been a rich and rewarding experience for me. From the bottom of my heart, I am extremely grateful for his wisdom, encouragement, and immense patience with my writing. I would like to extend my thanks to Amy Phillips for hosting lovely parties to let the group relax. Thanks also to Katie Miller for her administrative assistance. I would never have made it here without help from my labmates, Frosso Seitaridou, Seth Blumberg, Dave Wu, Hernan Garcia, Darren Segall, Paul Wiggins, Heun Jin Lee, Eric Peterson and Tristan Ursell. I would especially like to express my thanks to Paul Grayson for being my answers.com. It has been a pleasure working with all of them. I also want to say thank you to Yan Poon and Corey Wilson for their kind help with purification of lac repressor protein.

When I joined Caltech I never imagined that I would have the chance to travel to other parts of the USA and Europe for research. I am glad to have met many wonderful scientists along the way and would like to express my gratitude to all of them. In particular, I thank Laura Finzi and David Dunlap for teaching me the tethered particle motion method and for Laura's gracious hospitality in Milan. I am grateful for the hospitality of Cees Dekker and Chris Meiners during a visit to their labs. I am greatly indebted to my collaborators, Phil Nelson, Jon Widom, Jeff Gelles, and Bob Schleif for their advice and many valuable discussions. I would also like to thank Kathleen Matthews and Jason Kahn for providing the Lac repressor for the looping experiments.

Much love to my family, Ru, and Sotiris, whose patient love enabled me to complete this work.

# Abstract

The biological significance of DNA is primarily attributed to its sequence information. On the other hand, the mechanical properties of DNA can play a critical role in a wide variety of biological processes. One prime example is DNA looping in the context of transcriptional regulation. The emergence of single molecule tracking techniques in the last two decades presents an unprecedented opportunity for studying looping kinetics. One such powerful technique, tethered particle motion (TPM), harnesses the Brownian motion of a microsphere as a means of reporting on the excursion of its tethered molecule, such as DNA. The present work focuses on a looping system found in *Escherichia coli*, which is mediated by the Lac repressor (LacI) protein. TPM is used to measure individual, real-time looping/unlooping events in DNA of various length and sequence characteristics. By monitoring the magnitude, frequency, and time interval of these features while tuning different parameters, such as LacI concentration, DNA length and DNA sequence, one can survey a host of important information about looping kinetics. A measurement of the LacI concentration dependence of looping probability was found to be in quantitative agreement with a simple thermodynamic model, which also led to the measurement of free energy of LacI-mediated looping, the first such measurement in a single molecule, *in vitro* setting. A quantitative characterization of free energy was obtained under conditions of different inter-operator spacing, systematically varied from 300 to 310 base pairs in one-base-pair increments. An important conclusion from this study is that free energy is modulated by DNA's helical structure, yet the energy difference between the aligned and unaligned operator configurations is small compared to expectation from simple polymer physics. TPM measurements also revealed an additional looped state, lending support to the hypothesis that two distinct conformations of LacI, the closed and open forms, can coexist. This study also confirmed that the presence of certain DNA sequences, particularly TA pairs in the minor groove of the nucleosomal positioning sequence, makes DNA substantially softer than a corresponding random sequence. This provides direct support for the notion of sequence-dependent DNA elasticity. Finally, a surprising result is that loops as short as 100 base pairs-only two-thirds

the persistence length of DNA-can form by LacI-DNA binding. Classical elasticity theory almost forbids this, suggesting that LacI itself plays a more direct role in the bending process, or classical understanding of DNA elasticity breaks down at length scales comparable to its persistence length.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>List of Figures</b>	<b>xxii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Tightly Bent DNA is a Fact of Life . . . . .	1
1.2 DNA Looping Studies <i>In vivo</i> and <i>In vitro</i> . . . . .	5
1.3 Sequence Dependence of DNA Flexibility . . . . .	8
1.4 Structure of the Thesis . . . . .	9
<b>2 Calibration of Tethered Particle Motion method</b>	<b>15</b>
2.1 Introduction . . . . .	16
2.2 Results and Discussion . . . . .	17
2.2.1 Data Selection Criteria . . . . .	18
2.2.2 Acquisition Time . . . . .	20
2.2.3 Calibration of Motion . . . . .	24
2.2.4 Time Constant as a Function of Tether Length and Bead Size . . . . .	28
2.2.5 Ionic Effect . . . . .	28
2.2.6 Smearing Effect . . . . .	30
2.3 Applications to DNA Looping . . . . .	35
2.4 Conclusions . . . . .	38
<b>3 Data Analysis</b>	<b>39</b>
3.1 Particle Tracking . . . . .	39
3.2 Drift Correction . . . . .	40

3.2.0.1	Introduction to Digital Filters . . . . .	44
3.2.0.2	Butterworth Filter: Drift Correction . . . . .	44
3.3	Thresholding . . . . .	46
3.3.0.3	Gaussian Filter: Reducing Noise . . . . .	48
3.3.0.4	The Threshold Detector . . . . .	52
3.3.0.5	Effect of Window Size . . . . .	53
3.4	Diffusion-Hidden Markov Model Method . . . . .	60
3.4.1	Basic Concept of Hidden Markov Model . . . . .	62
3.4.1.1	Diffusion of a Tethered Particle . . . . .	63
3.4.2	Diffusive-HMM . . . . .	67
3.4.2.1	Apply DHMM to TPM Data . . . . .	72
<b>4</b>	<b>Lac Repressor Concentration Dependence of DNA Looping</b>	<b>75</b>
4.1	Introduction . . . . .	75
4.2	Results . . . . .	77
4.3	Conclusions . . . . .	89
<b>5</b>	<b>Length Dependence of DNA looping</b>	<b>91</b>
5.1	Length Dependence: 1 bp Resolution for a Whole Helical Turn . . . . .	91
5.2	Kinetic Analysis . . . . .	98
<b>6</b>	<b>Sequence Dependence of DNA Looping</b>	<b>101</b>
6.1	Introduction . . . . .	101
6.2	Results . . . . .	104
6.2.1	Formation of the Wild-type Loop . . . . .	104
6.2.2	Sequence Dependence . . . . .	106
6.2.3	Theoretical Analysis . . . . .	108
<b>7</b>	<b>Conclusion and Future Work</b>	<b>113</b>
7.1	Conclusions . . . . .	113
7.2	Future Directions . . . . .	115

<b>A</b>	<b>Materials and Methods</b>	<b>117</b>
A.1	Construction of Plasmids with Two LacI Binding Sites for Length Dependence Measurements . . . . .	117
A.2	Construction of Plasmids with Two LacI Binding Sites for Sequence Dependence Measurements . . . . .	118
A.3	Construction of Labeled DNAs . . . . .	118
A.4	TPM Sample Preparation . . . . .	120
<b>B</b>	<b>Sequences</b>	<b>127</b>
B.1	Sequences Between the Two Operators for Phasing Experiments (300-310 bp) . .	127
B.2	Sequences Between the Two Operators for Sequence-Dependence Experiments . .	129
	<b>Bibliography</b>	<b>131</b>



# List of Figures

1.1	Biological examples of tightly bent DNA. (a) Transcription factor mediated DNA looping, (b) DNA packing in the nucleosome, (b) DNA packing in bacterial viruses. Adopted from Garcia <i>et al.</i> (2007). . . . .	2
1.2	Gene regulation in <i>lac</i> Operon. (a) Three structural genes, <i>lac Z</i> (which codes for $\beta$ -galactosidase), <i>lac Y</i> (which codes for lactose permease), and <i>lac A</i> (which codes for transacetylase), are under the control of the same promoter where RNA polymerase binds and starts transcription. The gene encoded for the regulatory protein, Lac repressor, is located Upstream of the <i>lac</i> operon. The operators O1, O2 and O3 are the binding sites for Lac repressor. Locations of the primary operator O1 and the two auxiliary operators O2 and O3 with respect to the CAP binding site and promoter region for RNA polymerase. The ends of the <i>LacI</i> and <i>lacZ</i> genes are also provided as reference. (b)The operon is " off ". In the absence of lactose, <i>LacI</i> binds to the operator and prevents RNA polymerase initiating the gene transcription. (c) The operon is " on". In the presence of lactose, the inducer, <i>LacI</i> no longer binds to the operator, and transcription starts. Adopted online from <a href="http://www.blc.arizona.edu/Marty/411/Lectures/Figures/">http://www.blc.arizona.edu/Marty/411/Lectures/Figures/</a> . . . . .	3
1.3	<i>In vivo</i> DNA looping by Lac repressor. (a) Data from Müller-Hill <i>et al.</i> (Muller <i>et al.</i> 1996) showing repression as a function of distance between operators. (b) Extracted looping free energy based on equilibrium statistical mechanics (Garcia <i>et al.</i> 2007; Bintu <i>et al.</i> 2005a). Adopted from Garcia <i>et al.</i> (2007). . . . .	6
1.4	Correlation between the flexibility of DNA and the stability of the DNA-histon complex. The free energy of DNA wrapping on the histone H32H42 tetramer Is strongly correlated with the free energy of cyclization. . . . .	9

- 1.5 (a) Illustration of TPM method. Schematics of both the unlooped and looped states which show how the effective tether length is a reporter of the state of looping. Left: Motion of the tethered bead in the absence of the DNA binding protein. Right: Reduced motion of the tethered bead due to the change in the effective tether length by DNA binding protein. Typical tethers have a length of 1000 bp and typical bead sizes are 200 to 1000 nm in diameters. (b) Calibration curves of Brownian motion versus DNA tethered length with different bead sizes. . . . . 10
- 1.6 Two methods of TPM data analysis to obtain the kinetic information. (a) Basic concept of thresholding method. 1 Windowing data to reduce noise. 2 Setting up a threshold through which transition events are identified and durations of such events are quantified. 3 Histogram of all the durations and fit it to obtain the life time  $\tau$ . (b) Basic concept of Diffusive-HMM method. 1 Choose an observable variable suitable for HMM application. 2 Set up probability distribution of the observable variable  $\vec{r}_i$  at time  $i$  for each hidden state  $q_i$  with unknown parameter: rate constants  $ks$ . 3 Calculate the total probability of all possible trajectories. 4 Maximize the likelihood and extract the unknown parameters: rate constants. . . . . 11
- 1.7 RMS motion distribution averaged over 4s at different concentration of Lac repressor. Experimental data are in diamonds (blue: Di-operator DNA; red: Mono-operator DNA) and three-Gaussian fit is in green. Looping is optimized at a certain repressor concentration, above which it is less favored as the repressor concentration goes up. Two gray dashed lines represent the expected motion, based on our calibration measurements (data not shown), for 901bp DNA and the same DNA when 306 bp (the interoperator spacing) are subtracted off of the full length 901bp tether. . . . . 12
- 1.8 Histogram of the Brownian motion for DNAs with two LacI binding sites spaced from 300 to 310 bp. Experimental data are in blue diamonds and three-Gaussian fit is in red. Two gray lines represent the expected motion, based on our calibration measurements (data not shown), for 901bp DNA and the same DNA when 305 bp (the interoperator spacing) are subtracted off of the full length 901bp tether. . . . . 13
- 1.9 Histogram of Brownian motion distribution of different sequences, including when binding sites are in and out of phase. Gray dash lines are expected motion from calibration data. . . . . 14

- 2.1 Illustration of the tethered particle method. (a) Motion of the tethered bead in the absence of the DNA binding protein. (b) Reduced motion of the tethered bead due to the change in the effective tether length induced by a DNA binding protein that loops or bends the DNA. (c) Simulated looping data with 4s variance filter. The green point is the 4s averaged RMS value at time  $t = 800s$ . The inset shows the x-y positional data corresponding to the RMS trajectory. . . . . 17
- 2.2 Selection of qualified tethers. Displays of good and bad data, which inspire the selection criteria. (a) Trajectory associated with a good data set. The green line corresponds to the 4-second averaged RMS motion in a control experiment with a stuck bead. (b) Trajectory for a bead that failed the motion symmetry test. (c) Trajectory associated with nonuniform motion caused by nonspecific binding, seen as a downward spike between 0 and 50 seconds. (d) Scatter plot of in-plane motion. This case reveals symmetric motion. (e) Scatter plot of in-plane motion. This case is asymmetric and corresponds to data that are rejected. In all figures, the dots correspond to raw data and the blue line to data after averaging over a four-second time window. The DNA used here is 1206 bp long and the bead size is 490 nm in diameter. (f) Distribution of bead excursions with selection rules. Three filters are applied progressively. Original data: red; after application of minimal motion filter: cyan; after application of symmetry filter: blue; and after application of final filter: pink. The DNA used here is 901bp long and bead size is 490nm in diameter. . . . . 21
- 2.3 Amplitude of Brownian motion and the average standard deviation as a function of the averaging time for different length DNA tether lengths. The length of the DNA varies from 199 to 2625 bp. The top row shows the averaged RMS motion as a function of time interval, and the lower row shows the standard deviation as a function of time interval. Black dotted lines indicate the time over which the system reaches equilibrium. Red lines made by hand suggest the motion reaches saturation. . . . . 23

2.4	Brownian motion of bead as a function of the tether length for different sized microspheres. Each red point in the figure is the average of equilibrium amplitude of RMS motion over 20 to 200 qualified beads, which is calculated using equation 2.1 with $t = 5$ seconds for $R = 100$ nm (bottom data set), $t = 10$ seconds for $R = 245$ nm (middle data set) and $t = 20$ seconds for $R = 485$ nm (top data set). Points in green represents the RMS motion averaged over 4s for the same data sets. Points: experimental data for different sized beads. The blue and black dashed curves are second-order polynomial fits to the datasets obtained by using the different averaging times. . . .	25
2.5	Experimental values for RMS motion of bead center, projected to the plane of the surface, averaged over 5, 10, and 20 seconds of observation for bead size in radius 100, 245 and 485nm, respectively. Each dot represents the average of approximately 20 to 200 different observed beads with the given tether length. Curve: theoretically predicted RMS motion, corrected for the blurring effect of our long shutter time. The curve is a <i>zero parameter</i> prediction based on the known value of DNA persistence length in solvent conditions like ours, $\xi = 45$ nm Strick <i>et al.</i> (1998), and the manufacturer's specification for the bead radius. . . . .	27
2.6	Autocorrelation function for a single measurement. Top: bead radius 100 nm and 1338 base pairs. Middle: bead radius of 245 nm and 1323 base pairs. Bottom: bead radius of 485 nm and 1338 basepairs . . . . .	29
2.7	Histogram of time constants calculated. Left: bead radius 100 nm and 1338 base pairs. Center: bead radius of 245 nm and 1323 base pairs. Right: bead radius of 485 nm and 1338 basepairs . . . . .	29
2.8	Time constant as a function of contour length and bead radius. . . . .	30
2.9	The $RMS_{4s}$ motion of 901bp DNA tethered beads in different concentrations of KCl and $MgCl_2$ . . . . .	31
2.10	The $RMS_{4s}$ motion as a function of camera exposure time. The effect of camera exposure time was studied by examining the bead's RMS motion for different choices of exposure time. Each data point corresponds to a different camera setting. . . . .	32
2.11	Comparison of typical trajectories averaged over 4 seconds in the presence of Lac repressor showing loop formation and breakdown between DNAs with different tether lengths. (a) Total length of DNA is 901 bp. (b) Total length of DNA is 1128 bp. Operator center to center distance is 325.5 bp, which is the same for both DNAs. . . .	36

2.12	Calibration curve for 490nm size bead and its application in looping experiments. Blue: calibration curve for $\text{RMS}_{4s}$ motion as a function of tethered DNA length. Red: $\text{RMS}_{4s}$ motion in the presence of 100pM LacI. DNA is 901bp long with operator center to center distance of 325.5 bp. Green: $\text{RMS}_{4s}$ motion in the presence of 100 pM LacI. DNA is 450 bp long with operator center to center distance of 114.5 bp. Each pair of black dashed lines represent the expected motion for actual total length and the length subtracts off the operator center to center distance, respectively. . . .	37
3.1	Cross-correlation algorithm used in particle tracking. (a) A typical image of the bead observed using DIC microscope and its corresponding intensity plot on the right. (b) Cross-correlation between two images with the same bead. (c) Calculation of the centroid of the peak in the cross-correlation. . . . .	41
3.2	The drift in the raw position data and the data after drift correction. (a) The raw $x$ position, the position after drift correction and the drift estimation. (b) The $x - y$ scatter plot showing that the motion in $x - y$ plane is radial symmetric. . . . .	42
3.3	A diagram with one simple example explaining how to use a filter to recover a signal corrupted with random noise. (a) Noise corrupted signal (in orange) and the true signal (in blue). (b) A power spectrum showing different frequency components. By identifying and locating the dominant frequencies, a reasonable filter shown in (c) is selected to recover the true signal. (c) The selected band-pass filter $B(f)$ , in which only the dominant frequencies get to pass and other frequencies due to noise are blocked. (d) Comparison between the recovered signal (in red) and the true signal (in blue). . . . .	43
3.4	The frequency response of the Butterworth filter with different orders. The higher the order, the steeper the attenuation. . . . .	45
3.5	Power spectrum of both tethered and stuck bead. Three beads' power spectrum are plotted here, including two tethered beads and one stuck bead selected from the same field of view. . . . .	45
3.6	Drift correction for a struck bead using first order Butterworth filter with cutoff frequency 0.1 Hz . (a) The raw $x$ position of the stuck bead (in blue) and the drift detected using Butterworth filter (in red). (b) The drift-corrected $x$ position due to vibration. . . . .	47

3.7	Two methods of TPM data analysis to obtain the kinetic information. (a) Basic concept of thresholding method. 1 Windowing data to reduce noise. 2 Setting up a threshold through which transition events are identified and durations of such events are quantified. 3 Histogram of all the durations and fit it to obtain the lifetime $\tau$ . (b) Basic concept of Diffusive-HMM method. 1 Choose an observable variable suitable for HMM application. 2 Set up probability distribution of the observable variable $\vec{r}_i$ at time $i$ for each hidden state $q_i$ with unknown parameter: rate constants $k$ s. 3 Calculate the total probability of all possible trajectories. 4 Maximize the likelihood and extract the unknown parameters: rate constants. . . . .	49
3.8	Raw data and filtered data. Typical data set filtered at 0.0326 Hz (pink) and 0.1304 Hz (green) resulting in different signal-to-noise ratios. The blue dots represent the unfiltered data before. . . . .	51
3.9	Typical two state data set filtered at 0.0326 Hz showing Brownian motion as function of time. The red curve is a double Gaussian fit to the two peaks and the midpoint between them is taken as the dividing line between the two states. The two-state's system was observed in the presence of both LacI and IPTG when the bead was tethered with 901 bp long DNA (Later on, no such data can be repeated). . . . .	51
3.10	Transitions detected using thresholding methods. The green lines represent the states the system is in. (a) The Gaussian filtered bead's motion. (b) Transitions detected by single-threshold method. . . . .	52
3.11	The step response of the Gaussian filter. The original step function is plotted using the blue dashed curve and the filtered signal is given by the red curve. . . . .	53
3.12	Pulses with different width and their corresponding filtered output. (a) Original pulses (in blue), the output of the filter (in red) and the half-amplitude threshold (in green). (b) Transitions detected with duration measured. . . . .	56
3.13	Looped and unlooped time distribution from data filtered with different cutoff frequencies. . . . .	58

3.14 Simplified kinetic scheme in the three-state-system. (0): The state in which No LacI binds. (1): The state where LacI only binds to one of the two binding sites. The state in the shadow is ignored for simplification. (2): Open-looped configuration where LacI is extended when binds to the two binding sites at the same time. (3): Closed-looped configuration where LacI is in a V-shape when binding to the two binding sites at the same time. (4): Two LacI present resulting in occupancy of both binding sites. . . . . 59

3.15 State transitions in a hidden Markov model. Circles represent the hidden states  $q$  in the system, transitions among which are governed by a set of probabilities called transition probabilities. In a particular state an observable output, presented as hexagons, can be generated, according to the associated probability distribution  $P$ . It is only the output, not the state visible to an external observer and therefore states are "hidden" to the outside. . . . . 60

3.16 A 2-state urn and ball model which illustrates the general case of a discrete symbol HMM. Two steps are involved in finding the best Markov model that describes the behavior of the system. (a) Specification of the general topology of the model including five basic elements: 1) Number of hidden states. 2) The number of distinct observation symbols per state. 3) Output probability distribution. 4) State transition probability. 5) The initial state distribution. (b) Optimization of Markov model parameters. Knowing the probability distribution of each observable output in each hidden state and the transition probability containing the transition rates we are interested in, the total probability of the observation can be calculated and then optimized to get the hidden parameters of the model using maximum-likelihood techniques. . . 61

3.17 Autocorrelation of experimental positional data. (a) Autocorrelation showing exponential decay. Green line: exponential fit. (b) Autocorrelation in logarithm. Red line represents the linear fit from which the spring constant  $k_{\text{eff}}$  and drag coefficient  $\gamma$  are extracted. . . . . 66

3.18 Construction of probability distribution for the next position. (a) Histogram displacements  $\delta x$  and  $\delta y$  at  $t+1$  when the distance to anchor point is  $\rho_t$ . (b) Mean of  $\delta x$  and  $\delta y$  as function of  $\rho$  and its corresponding second-order polynomial fit. (c) Standard deviations of  $\delta x$  and  $\delta y$  as function of  $\rho$  and its corresponding first-order polynomial fit. Markers with different colors in plot (b) and (c) represent different data sets. . . 67

3.19	Transition detected using 3-state DHMM method. (a) Transitions detected (in green) with Gaussian filtered $\sqrt{\langle R^2 \rangle_{4s}}$ motion superposed. (b) Corresponding histogram of $\sqrt{\langle R^2 \rangle_{4s}}$ . (c) An enlarged plot showing the region circled in red in (a). Orange represents the threshold value and pink arrows point at the corresponding positional data points in the pulse regions. . . . .	71
4.1	Different representations of TPM data. (a) Scatter plot of drift-corrected positional data. Each dot corresponds to the instantaneous projected position of the bead at a given instant in time. (b) Running average of RMS motion over 4s. $R$ is the distance from the bead center (dots in panel (a)) to the tether attachment point (centroid of all dots in panel (a)). Red and green lines represent the expected motion, based on our calibration measurements (data not shown), for 901bp DNA and the same DNA when 305 bp (the interoperator spacing) are subtracted off of the full length 901bp tether. (c) RMS distribution. DNA used here is pUC305L1 with 100pM Lac repressor. . . . .	78
4.2	Typical trajectories with mono-operator DNA. (a) $O_{id}$ is removed. (b) $O_1$ is removed.	79
4.3	Typical Brownian motion traces of the DNA tethered beads in the presence of different concentrations of Lac repressor varying from 1pM to 100nM. DNA is 901bp in total length with 306bp inter-operator spacing. . . . .	80



- 4.4 Concentration dependence of looping probability. The histograms show the distribution of RMS motions averaged over 4s at different concentrations of Lac repressor. Experimental data are represented by diamonds (blue: di-operator DNA; red: mssssssssssssso-operator DNA) and a fit to the data using a sum of three Gaussians is shown in green. The fit parameters are displayed in table 4.1. The two gray dashed lines represent the expected motion, based on our calibration measurements (data not shown), for 901 bp DNA and the same DNA when 306 bp (the interoperator spacing) are subtracted off of the full length 901 bp tether. The short tether loop has smaller motion than expected by simple subtraction of the inter-operator spacing, which is expected due to the imperfect exiting angle of DNA from the loop structure. The bottom and top curves show control experiments: bottom, no Lac repressor; top, no second operator. The widths of the distributions for the two control experiments are slightly different. One possible explanation is that there are two possible configurations for the DNA construct without the second operator, empty or bound by Lac repressor. Crystal studies shows that a sharp bend in operators occurs upon Lac repressor binding (Lewis *et al.* 1996), which would result in a slightly reduced effective tether length. . . . . 82
- 4.5 Averaging effect among the data set. Data obtained with pUC306L1 DNA in the presence of 10 pM Lac repressor. . . . . 83
- 4.6 Looping probability at different concentration of Lac repressor. The DNA used in these experiments is 901 bp long and the inter-operator spacing is 306 bp. The vertical axis gives looping probability (fraction of time spent in either of the two looped states). The points were obtained using two different methods: The blue points were obtained by calculating the fraction of time that the system spends in one of the two looped states, using a transition detection algorithm based on a Hidden Markov Model (HMM) (Beausang *et al.* 2007; Beausang and Nelson 2007). The yellow points are derived from the histogram distribution by integrating the area under the two looped states over total area. The red curve is a fit to the experimental data using the statistical mechanics model described in the text. . . . . 84

- 4.7 States and weights for the Lac repressor-DNA system (Bintu *et al.* 2005b). Each of the five state classes shown in the left column has a corresponding statistical weight. All of the weights have been normalized by the state in which the DNA is unoccupied. State (v) is treated as a single looped state, even though there are multiple distinct looped configurations. . . . . 86
- 4.8 Measured and computed probabilities for different states of Lac repressor and operator DNA. The curves are all the results from the statistical mechanical model using parameters obtained from the measured data points. Blue curve: the probability of empty DNA with no Lac repressor bound. Red curve: the probability of the looped state. Green curve: the probability of the state with two Lac repressor molecules bound, one associated with each of the operators. Pink curve: the probability of the state where one Lac repressor is bound to the Oid site. Cyan curve: the probability of the state where one Lac repressor is bound to the O1 site. The blue circles are the data points for the probability of the looped state obtained using the Hidden Markov Model based transition detection method (Beausang *et al.* 2007) and computing the fraction of the overall trajectory when the DNA is looped, with no reference to which looped state. The orange squares show the fractional equilibrium population obtained by calculating the area under the Gaussian curve in the looping region in the histogram. The black dashed line indicates the Lac repressor concentration used in the phasing experiments. . . . . 87
- 4.9 Computed probabilities for different states of Lac repressor and operator DNA for DNA with two identical operators O1. The curves are all the results from the statistical mechanical model using parameters  $\Delta\epsilon_1$  and  $\Delta G_{loop}$  obtained from our experiments. Blue curve: the probability of empty DNA with no Lac repressor bound. Red curve: the probability of the looped state. Green curve: the probability of the state with two Lac repressor molecules bound, one associated with each of the operators. Pink curve: the probability of the state where one Lac repressor is bound to the Oid site. Cyan curve: the probability of the state where one Lac repressor is bound to the O1 site. The black d lines indicate the Lac repressor concentration used in the previous experiments. . . . . 88

5.1	Typical Brownian motion traces of the DNA tethered beads with inter-operator spacing from 300 to 310 bp in 1 bp increment. Lac repressor concentration is 100 pM. . . . .	92
5.2	Histogram of the Brownian motion for DNAs with two Lac repressor binding sites spaced from 300 to 310 bp. Experimental data are in blue diamonds and three-Gaussian fit is in red. The fit parameters are displayed in tab. 5.1. The two gray dashed lines represent the expected motion, based on our calibration measurements (data not shown), for 901 bp DNA and the same DNA when 305 bp (the interoperator spacing) are subtracted off of the full length 901 bp tether. . . . .	94
5.3	Total looping free energies with the corresponding looping probability for each of the inter-operator spacings from 300 to 310 bp. A periodic behavior over the entire helical repeat is observed for looping probability, and its maximum occurs at interoperator spacing of 306 bp. . . . .	95
5.4	Stability of the two looped states as a function of inter-operator spacing. . . . .	97
5.5	Effective tether length as a function of inter-operator spacing. Blue: unlooped state. Green: middle state. Cyan: bottom state. Red: actual tether DNA length. Pink: DNA length subtracts inter-operator spacing. . . . .	97
5.6	Rate constants as a function of inter-operator spacing. (a) Looping rates. (b) Inter-conversion rates between two looped states. (c) Unlooping rates. . . . .	99
6.1	Typical TPM recording for wild type interoperator spacing found in the <i>lac</i> operon. (a) In the presence of Lac repressor, looping occurs as indicated both by the trajectories and the histograms. (b) Control experiment in which there is no Lac repressor present. (c) Control experiment using both IPTG and Lac repressor. The red lines correspond to the expected excursions for the unlooped state and the green line corresponds to the “expected” excursion for a DNA fragment from which the DNA fragment between the two operators has been removed. The DNA construct is 351 bp in length with wild type inter-operator DNA between binding sites $O_{id}$ and $O_1$ . . . . .	105
6.2	Histogram of Brownian motion distribution for wild-type interoperator spacing. The histogram is fit to a sum of two Gaussians. The looping free energy can be calculated based on the thermodynamic model. . . . .	106

6.3	Typical TPM recordings for different sequences and different lengths. The sequences labeled E8 are random sequences and those labeled TA are nucleosomal positioning sequences. The integer after the label refers to the length of the fragment between the two binding sites. For each sequence and length, two recordings are shown. . . . .	107
6.4	Histogram of Brownian motion distribution of different sequences, as well as when binding sites are in and out of phase. Grey dash lines are expected motion from calibration data. . . . .	109
6.5	Looping free energy as a function of inter-operator spacing. Distance between the two binding sites is referred as inter-operator spacing rather than the operator center to center distance. . . . .	112
A.1	Synthesis of DNA construct. (a) The procedure of construction of the plasmid with two LacI binding sites. (b) The basic idea of producing labeled DNA using PCR reaction. . . . .	119
A.2	Promoter regions of the different constructs. (a) Promoter region of pZS25-YFP which has a variant of the lacUV5 promoter and an $O_1$ operator upstream overlapping the -35 region. (b) Final construct that allows to insert arbitrary DNA sequences between a $O_{id}$ and $O_1$ operators. . . . .	120
A.3	Illustration of TPM sample preparation. (a) Sketch of the flow cell. (b) Two schemes of making DNA tethers. . . . .	121

# Chapter 1

## Introduction

### 1.1 Tightly Bent DNA is a Fact of Life

Since DNA's double helical structure was revealed by Watson and Crick (1953), most studies of this molecule center on its genetic information content. On the other hand, many of the mechanisms by which genetic information is stored and used and regulated involve mechanical deformation of DNA. Indeed, the tight bending of DNA has farreaching consequences for a multitude of biological processes. Figure 1.1 shows three distinct examples of the way in which genomic DNA is subject to tight bending. Figure 1.1 (a) shows DNA looping by Lac repressor: a ubiquitous motif involved in transcriptional regulation. Figure 1.1 (b) and (c) respectively illustrate how DNA packs in the nucleosome and bacterial viruses. Remarkably, genomic packing often involves bending DNA molecules on scales that are small in comparison with their persistence length. Just as any deformable objects, such as the trunk of a palm tree, has a characteristic length over which it is hard to bend, so do polymers like DNA. This length scale, which for microscopic polymers is defined as

$$\xi_p = \kappa/k_B T,$$

is characterized by a competition between the material's intrinsic flexural rigidity  $\kappa$  (Boal 2002), and thermal fluctuations. For a polymer in solution, the persistence length is the scale over which the thermal energy bends DNA. For DNA, this value is  $\approx 50$  nm, which is equivalent to  $\approx 150$  base pairs (bp) (Rivetti *et al.* 1996). When DNA is bent on a scale shorter than  $\xi_p$ , we refer to it as highly or tightly bent, implying a significant energy cost required to induce such bending. Interestingly, in many of its most important roles, DNA adopts exactly these forms of tightly bent configurations.

In the setting of transcriptional regulation, there are a host of regulatory architectures both in

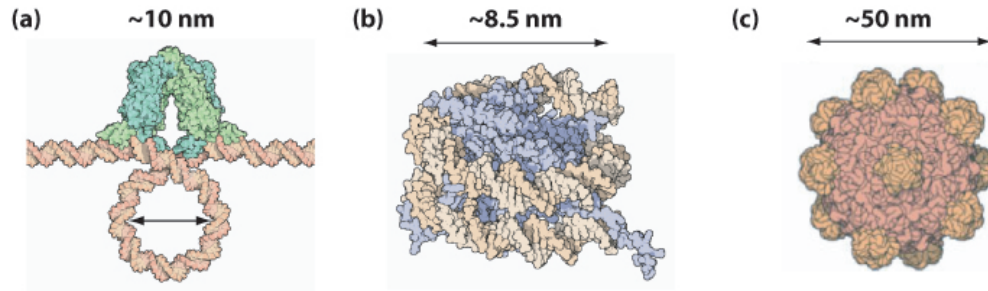


Figure 1.1: Biological examples of tightly bent DNA. (a) Transcription factor mediated DNA looping, (b) DNA packing in the nucleosome, (b) DNA packing in bacterial viruses. Adopted from Garcia *et al.* (2007).

prokaryotes and eukaryotes that require the interaction of sequences on the DNA which are not adjacent (Adhya 1989; Schleif 1992; Matthews 1992). These interactions are mediated by DNA-binding proteins that deform the DNA. Recently, it has become possible to perform genome-wide surveys to determine the entirety of looped configurations induced by a given protein (Loh *et al.* 2006; Wei *et al.* 2006). One of the most transparent examples of DNA looping is in bacteria where some repressors and activators can bind at two sites simultaneously resulting in a DNA loop. This effect was first elucidated in the context of the arabinose operon (Dunn *et al.* 1984). It is an amusing twist of history that the two regulatory motifs considered by Jacob and Monod, namely, the switch that makes the decision between the lytic and lysogenic pathways after phage infection (Ptashne 2004) and the decision making apparatus associated with lactose digestion in bacteria (Schleif 1992), both involve DNA looping as well. Table 1.1 shows some examples of regulatory processes that involve looping. It is worth remarking that in the majority of cases, the length of the resulting loop is comparable to or smaller than the persistence length ( $\approx 150\text{bp}$ ).

The *lac* system has been a classic model for investigation both *in vivo* and *in vitro* to understand the role of tightly bent DNA in gene regulation. The *lac* operon refers to the genes responsible for lactose transport and metabolism in bacteria (Müller-Hill 1996). As shown in figure ??(a), it consists of three adjacent structural genes, *LacZ*, *LacY* and *LacA*, which are under the control of the same promoter. This promoter is the site where RNA polymerase binds and initiates transcription. The *lac* operon is regulated by several factors, including the presence of glucose and lactose. The regulatory response to lactose requires the regulatory protein Lac repressor (*LacI*) which is encoded by *LacI* nearby. In particular, in the absence of lactose, as illustrated in figure 1.2 (b), *LacI* binds to the operator site which overlaps with the promoter site. The bound *LacI* prevents RNA poly-

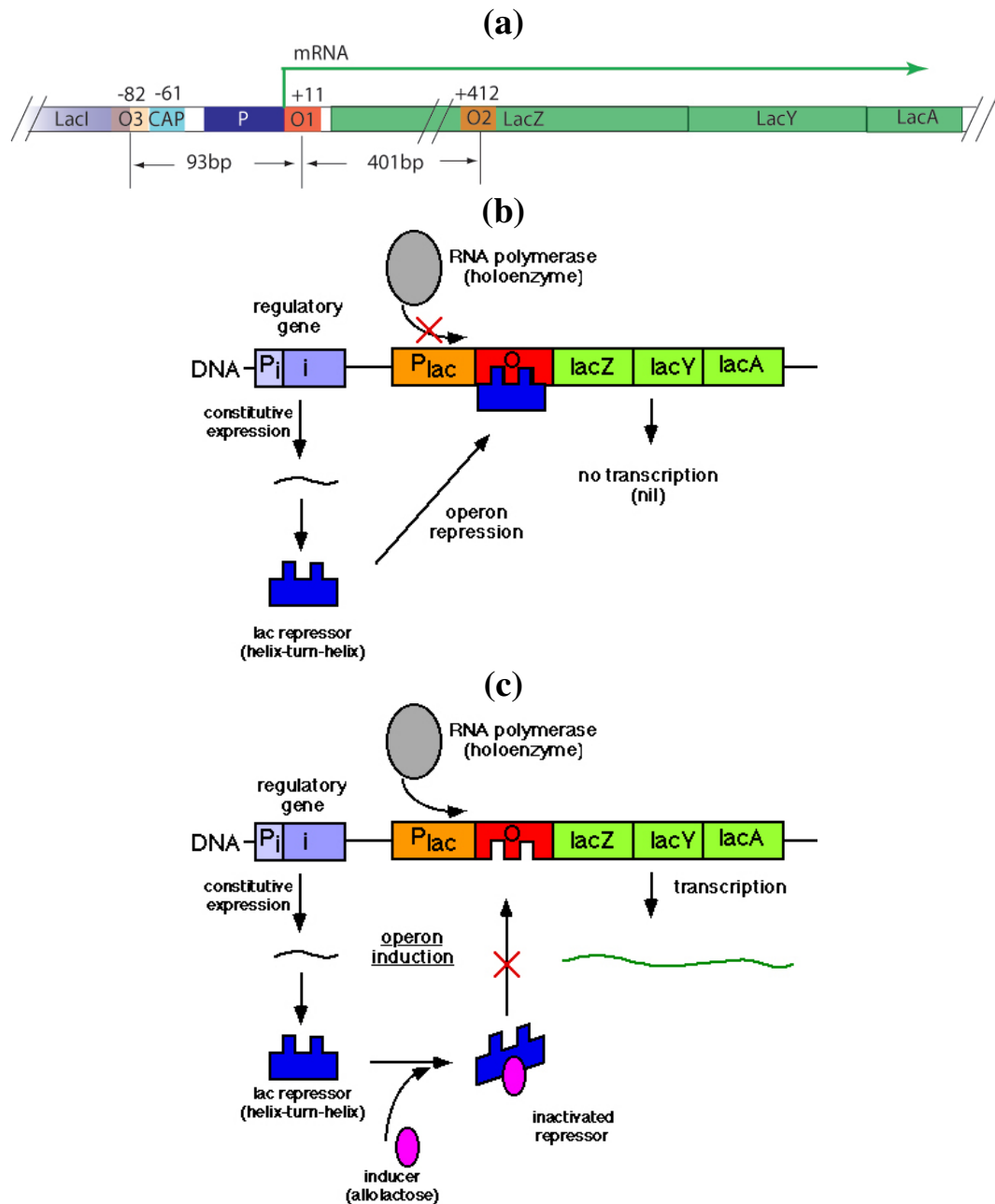


Figure 1.2: Gene regulation in *lac* Operon. (a) Three structural genes, *lac Z* (which codes for  $\beta$ -galactosidase), *lac Y* (which codes for lactose permease), and *lac A* (which codes for transacetylase), are under the control of the same promoter where RNA polymerase binds and starts transcription. The gene encoded for the regulatory protein, Lac repressor, is located Upstream of the *lac* operon. The operators O1, O2 and O3 are the binding sites for Lac repressor. Locations of the primary operator O1 and the two auxiliary operators O2 and O3 with respect to the CAP binding site and promoter region for RNA polymerase. The ends of the *LacI* and *lacZ* genes are also provided as reference. (b) The operon is "off". In the absence of lactose, *LacI* binds to the operator and prevents RNA polymerase initiating the gene transcription. (c) The operon is "on". In the presence of lactose, the inducer, *LacI* no longer binds to the operator, and transcription starts. Adopted online from <http://www.blc.arizona.edu/Marty/411/Lectures/figures/>

Table 1.1: DNA looping in prokaryotic and eukaryotic transcriptional regulation. Loop lengths and mechanisms of action of some of the best known looping systems in bacteria and eukaryotes. Note that these loop lengths suggest tightly bent configurations since the *in vitro* measured persistence length is 150 bp. Adopted from Garcia *et al.* (2007)

Molecule or locus	Mode of action	Wild type loop lengths (bp)	References
Lac repressor	Repression	92, 401	Müller-Hill (1996)
AraC	Repression, activation	210	Schleif (1992)
Gal repressor	Repression	115	Schleif (1992)
Deo repressor	Repression	270, 599, 869	Schleif (1992)
Nag repressor	Repression	93	Plumbridge and Kolb (1998)
NtrC	Activation	110 ~ 140	Mathews (1992)
$\lambda$ repressor	Repression, activation	~ 2, 400	Ptashne (2004); Dodd <i>et al.</i> (2004)
XylR	Activation	~ 150	Inouye <i>et al.</i> (1990)
PapI	Activation	~ 100	Mathews (1992); Nou <i>et al.</i> (1995)
RXR	Activation	30 – 500	Yasmin <i>et al.</i> (2004)

merase from transcribing, either by inhibiting the binding of RNA polymerase (Majors 1975; Nick and Gilbert 1985; Friedman *et al.* 1995) or stopping its entry into the processive elongation phase (Chen *et al.* 1971; Straney and Crothers 1987). Therefore, under lactose-free conditions, the operon is "off". When faced with an absence of glucose and the presence of lactose, the LacI undergoes a conformational change which in turn dramatically decreases the affinity of LacI to operator. As long as no repressor binds to the operators, RNA polymerase that recognizes the promoter can transcribe the operon's structural genes into mRNA and the operon is "on" (Figure 1.2 (b)). The efficiency of repression of the lactose operon in *Escherichia coli* is modulated by the Lac repressor's loop-forming effect on DNA. The *lac* operon consists of three recognition sites for LacI in a stretch of 500 bp, as illustrated in figure 1.2 (a). The primary operator O1 is located at position 11 relative to the promoter. Two auxiliary operators, O2 and O3, lie within the LacI and lac Z gene, respectively. O2 is 93 bp upstream of O1 while O3 is 401 bp downstream of O1. The affinity of LacI to these three operators in decreasing order is: O1 > O2 > O3. Maximally efficient repression is achieved via DNA looping by the simultaneous binding of LacI protein to two operators (Muller *et al.* 1996; Mehta and Kahn 1999). This is because DNA looping can significantly enhance protein association to the lower-affinity site due to the tethering effect of DNA looping (Schleif 1992; Zhang *et al.* 2006).



## 1.2 DNA Looping Studies *In vivo* and *In vitro*

To understand the physical mechanism of the biological action at a distance revealed by DNA looping it is necessary to bring both *in vitro* and *in vivo* experiments as well as theoretical analyses to bear on this important problem. Over the last few decades there have been a series of impressive and beautiful experiments from many quarters. In the *in vivo* context, it is especially the work of Müller-Hill and coworkers that demonstrates the intriguing quantitative implications of DNA looping for regulation (Muller *et al.* 1996). In their experiments, they tuned the length of the DNA loop in one-base-pair increments and measured the resulting repression of a reporter gene. More recently, these experiments have been performed with mutant bacterial strains which were deficient in architectural proteins such as HU (Becker *et al.* 2005). The observation of such experiments is that the amplitude of gene repression oscillates with length of the loop with a periodicity of  $\approx 11$  bp, which is consistent with the DNA helical repeat (Dunn *et al.* 1984; Lee and Schleif 1989; Becker *et al.* 2005; Muller *et al.* 1996). The characteristic oscillations in gene repression amplitude in *lac* operon is shown in figure 1.3(a). This experiment underscores quantitatively the importance of DNA mechanics in gene regulation, even at the single base-pair level. An important parameter that can be extracted from this type of study, using thermodynamic models (Buchler *et al.* 2003; Bintu *et al.* 2005c), is the free energy of looping, which measures the energy involved in forming the looped configuration as a function of the inter-operator distance (Law *et al.* 1993; Bintu *et al.* 2005a; Vilar and Saiz 2005; Saiz *et al.* 2005). As illustrated in figure 1.3(b), the minimum in looping free energy is located at around 70bp, which is about half the persistence length of DNA. This rather unexpected result has been suggested to arise from the greater effective flexibility of *in vivo* versus *in vitro* DNA. The mechanics of DNA inside of living cells is considerably more complicated than the *in vitro* description. In particular, besides the free energy of DNA looping itself, it is also necessary to consider a number of other factors including (i) the geometry and flexibility of the looping protein (Friedman *et al.* 1995; Allemand *et al.* 2006), (ii) the fact that DNA is negatively supercoiled inside the cell (Vologodskii and Cozzarelli 1996) and (iii) the presence of structural proteins such as HU, IHF and H-NS (Becker *et al.* 2005). The contributions and importance of each factor needs to be decoupled in order to analyze the role of tightly bent DNA in transcriptional regulation. In order to tackle each factor separately, *in vitro* approaches using the tools of solution biochemistry and single molecule biophysics, have been employed.

Bulk binding assays, such as filter binding assays and electrophoretic mobility shift assays, are

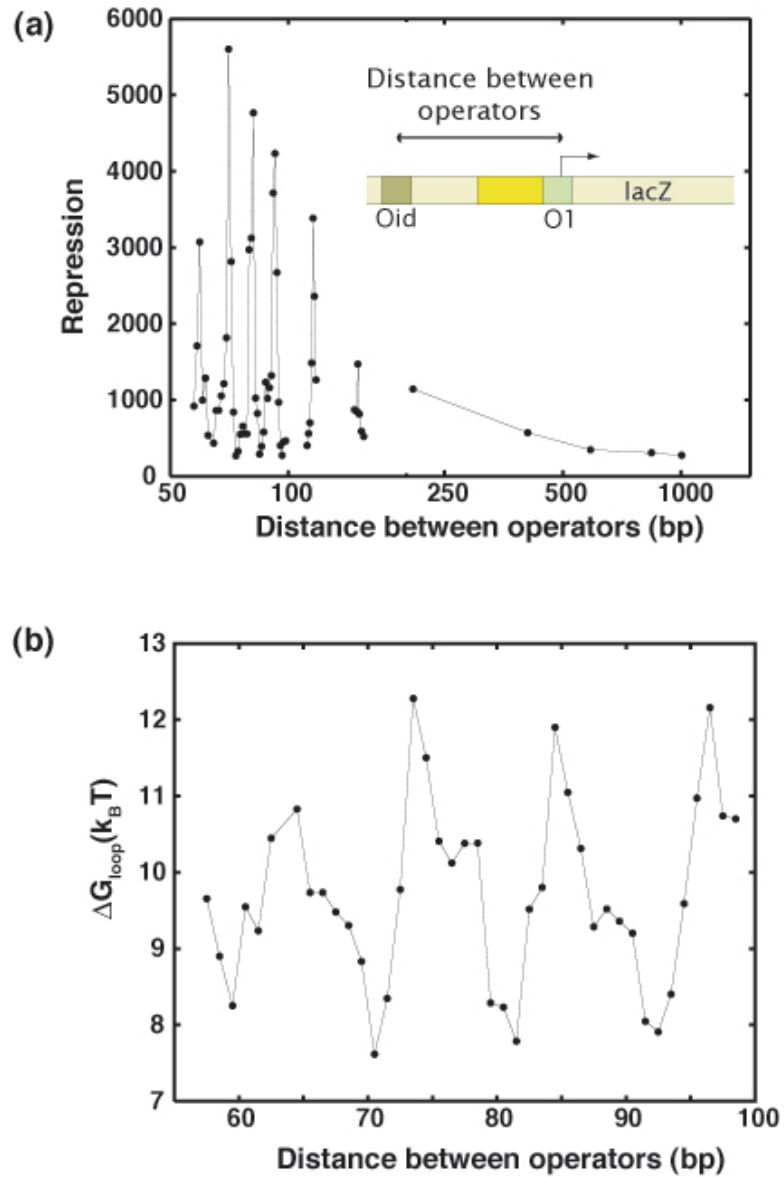


Figure 1.3: *In vivo* DNA looping by Lac repressor. (a) Data from Müller-Hill *et al.* (Muller *et al.* 1996) showing repression as a function of distance between operators. (b) Extracted looping free energy based on equilibrium statistical mechanics (Garcia *et al.* 2007; Bintu *et al.* 2005a). Adopted from Garcia *et al.* (2007).

used to measure the affinity of proteins to their DNA targets. These techniques offer the opportunity for systematically varying parameters, such as DNA length, the degree of supercoiling, or presence of structural proteins, thereby decoupling the various effects contributing to looping processes *in vivo*. Using the electrophoretic mobility shift assay, Müller-Hill *et al.* found that the looping probability qualitatively decreases as the inter-operator spacing is reduced from 210bp to 60bp (Krämer *et al.* 1987), which agrees with the quantitative observations of Matthews *et al.* using the filter binding assay (Hsieh *et al.* 1987). Hence there is an obvious disparity between the preferred length of DNA looping *in vivo* and *in vitro*. The same concepts have been applied to supercoiled plasmids to examine the role of supercoiling (Krämer *et al.* 1988; Whitson *et al.* 1987; Borowiec *et al.* 1987), and demonstrated an increase in the affinity of LacI binding to a single operator site. Their results suggest that negative supercoiling favors the association of LacI and its target DNA sites, resulting in an increase in looping probability and looping stability. Furthermore, this probability is found to remain relatively constant over the distance between 100 and 500 bp (Whitson *et al.* 1987). It is also found that supercoiling could shift the optimal spacing for loop stabilization depending on the degree of supercoiling, suggesting the helical repeat is changing (Krämer *et al.* 1988). Although such studies demonstrate the role of supercoiling and inter-operator spacing in DNA looping, they lack either quantitative kinetic information or systematic investigation.

Another important class of experiments that have shed light on the mechanics of DNA looping *in vitro* are single-molecule measurements using the Tethered Particle Motion (TPM) method (Schafer *et al.* 1991). Finzi and Gelles were the first to apply TPM to detect Lac repressor mediated loop formation and breakdown, and to elucidate the kinetic information of such processes (Finzi and Gelles 1995). In this method, a DNA molecule is tethered between a microscope slide and a microsphere which is large enough to be imaged with conventional optical microscopy. The Brownian motion of the bead serves as a reporter of the underlying DNA dynamics. In particular, when the molecule is unlooped, the tether has its full length, and the excursions of the bead are large. When the DNA is looped, the tether is shortened and the excursions are reduced (Yin *et al.* 1994; Tolic-Norrelykke *et al.* 2004a; Blumberg *et al.* 2005; Pouget *et al.* 2004; Segall *et al.* 2006a; Nelson *et al.* 2006a; Zurla *et al.* 2006; Broek *et al.* 2006; Zurla *et al.* 2007). Thus, modulations in motion reflect conformational changes in the tethered molecule. This method has recently revealed (Wong *et al.* 2007) the presence of two-looped states which is consistent with the presence of multiple configurations observed using FRET (Edelman *et al.* 2003; Morgan *et al.* 2005), electron microscopy studies (Ruben and Roos 1997) and suggested by x-ray crystallography studies (Friedman *et al.*

1995). All of these experiments suggest an important role for protein flexibility.

Structural proteins, such as HU and IHF, have been known to alter the effective flexibility of DNA (Noort *et al.* 2004). Recent TPM and AFM studies have demonstrated the influence of specifically-bound IHF protein on Lac-mediated DNA looping (Zurla *et al.* 2007). These results suggest that a sharp bend on DNA occurs upon IHF protein binding. Moreover, looping could either be enhanced or attenuated depending on where this bend is located as well as the relative phasing of the two LacI binding sites. Though this study successfully investigated the effect of specific, sharp bending or non-specific compaction by IHF affected on DNA looping, no systematic data are available to draw conclusions on the influence of such proteins on DNA looping with inter-operator distance varying in one base pair increments.

As stated above, significant effort has been devoted in the past two decades to understanding the role of tightly bent DNA in gene regulation. *In vitro* investigations have qualitatively shown that a variety of different players *in vivo*, such as DNA supercoiling, DNA length, or the presence of structural proteins, significantly alter looping efficiency and stability. However, such studies lack the single-base pair resolution that would enable the distinction of whether individual contributions are sufficient to explain the observed dependence of repression on inter-operator spacing in living biological systems.

### 1.3 Sequence Dependence of DNA Flexibility

So far we have reviewed how length is a key parameter in the study of DNA mechanics. Nucleotide sequence is another important way in which the energetics of tightly bent DNA can be tuned. For example, sequence dependence has been shown to influence nucleosome formation, where  $\approx 147$ bp long DNA has to be wrapped into loops of only  $\approx 80$  bp per superhelical turn. Remarkably, some sequences wrapping onto nucleosomes are favored over others by more than a factor of 5000 (Cloutier and Widom 2004, 2005). This preferred affinity is not due to any particularly favorable site-specific interaction between DNA and nucleosomal proteins (Luger *et al.* 1997). Instead it is a consequence of indirect readout (Gromiha *et al.* 2004), where different DNA sequences have varying abilities to adopt particular conformations required by nucleosomes. As illustrated in figure 1.4, the direct correlation between the free energy cost of DNA cyclization and the free energy of nucleosome formation demonstrated the important role of flexibility of DNA sequences in nucleosomes formation (Roychoudhury *et al.* 2000; Cloutier and Widom 2005). Unlike the DNA-nucleosome wrapping

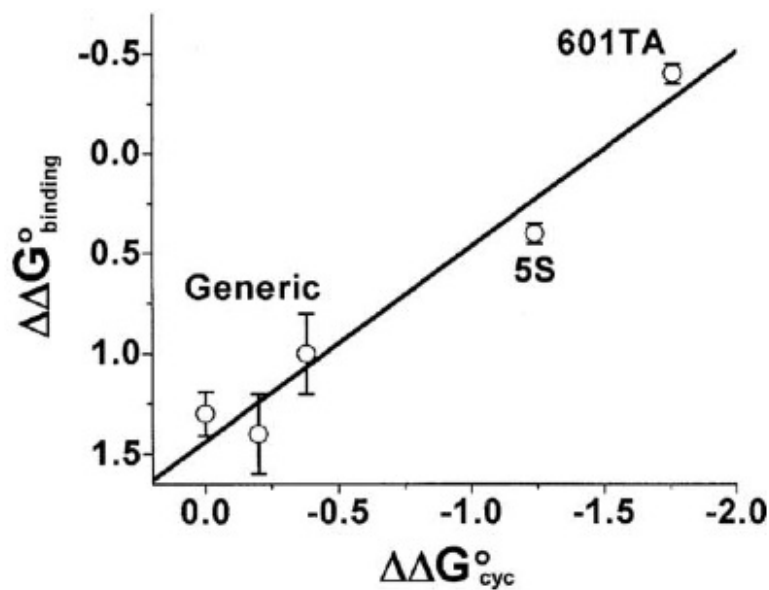


Figure 1.4: Correlation between the flexibility of DNA and the stability of the DNA-histon complex. The free energy of DNA wrapping on the histone H32H42 tetramer is strongly correlated with the free energy of cyclization.

system, sequence dependence has not yet been demonstrated in DNA looping systems, such as the ones listed in table 1.1.

## 1.4 Structure of the Thesis

The goal of this thesis is to exam the role of DNA mechanics in transcriptional regulation. In particular, we use DNA looping in the *lac* operon as a window on the elastic properties of DNA fragments. Single-molecule biophysics makes it possible to perform systematic and quantitative studies on how the DNA looping process responds to parameter changes, such as operator distance, DNA sequences and the concentration of repressor proteins.

Chapter 2 introduces the basic concept of the Tethered Particle Motion (TPM) method, followed by establishment of a well-characterized calibration of TPM for a ranges of parameters, such as bead size and the tethered length. TPM experiments often employ a measure of the particle's Brownian motion in order to probe biological dynamics associated with conformational change of the tether. Often a relative change in the Brownian motion is deemed sufficient for TPM analysis. However, a complete understanding of how the motion is dependent on the physical properties of the tethered

particle complex would permit more quantitative and accurate evaluation of TPM data. In order to better characterize the measurement capabilities of TPM experiments involving DNA tethers, we have carried out a detailed calibration of TPM on magnitude of Brownian motion as a function of DNA length, particle size, and ionic strength of buffer. We also explore how experimental parameters such as acquisition time and exposure time affect the apparent motion of the tethered particle. We vary the DNA length from 200 bp to 2.6 kbp and consider particle diameters of 200, 490, and 970 nm.

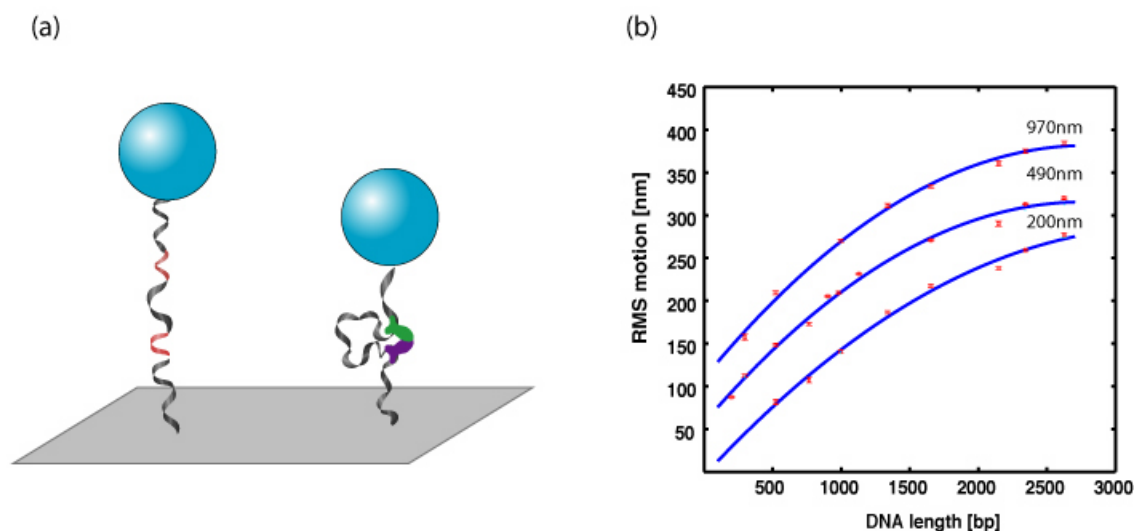


Figure 1.5: (a) Illustration of TPM method. Schematics of both the unlooped and looped states which show how the effective tether length is a reporter of the state of looping. Left: Motion of the tethered bead in the absence of the DNA binding protein. Right: Reduced motion of the tethered bead due to the change in the effective tether length by DNA binding protein. Typical tethers have a length of 1000 bp and typical bead sizes are 200 to 1000 nm in diameters. (b) Calibration curves of Brownian motion versus DNA tethered length with different bead sizes.

Chapter 3 demonstrates how to extract kinetic information of looping processes from the direct experimental outcome, i.e. ,the record of movements of the beads. Positions can be tracked in each frame, from which important rates can be evaluated. Two methods shown in figure 1.6 can be used to obtain such kinetic information. One is the conventional thresholding method, and the other one is based on the Hidden Markov Model.

Chapter 4 quantitatively characterize the dependence of looping on the concentration of Lac repressor. At very low concentrations, we expect that there will be negligible looping. At intermediate

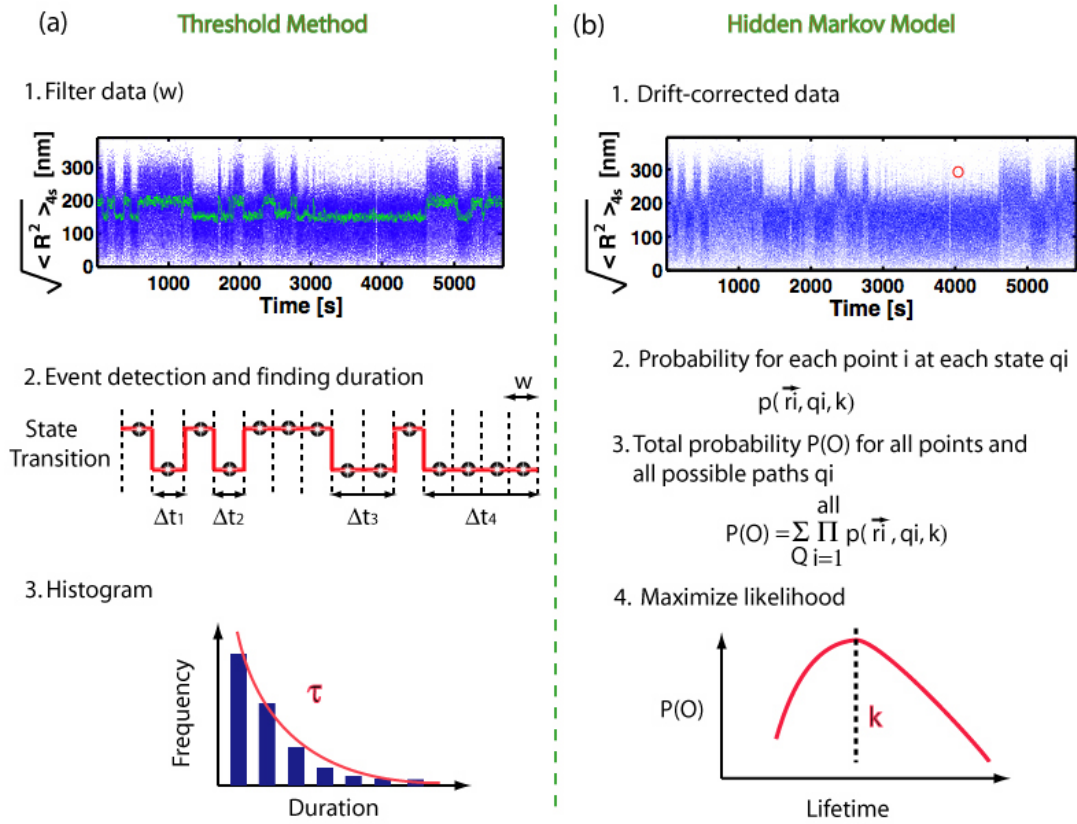


Figure 1.6: Two methods of TPM data analysis to obtain the kinetic information. (a) Basic concept of thresholding method. 1 Windowing data to reduce noise. 2 Setting up a threshold through which transition events are identified and durations of such events are quantified. 3 Histogram of all the durations and fit it to obtain the life time  $\tau$ . (b) Basic concept of Diffusive-HMM method. 1 Choose an observable variable suitable for HMM application. 2 Set up probability distribution of the observable variable  $\vec{r}_i$  at time  $i$  for each hidden state  $q_i$  with unknown parameter: rate constants  $k$ s. 3 Calculate the total probability of all possible trajectories. 4 Maximize the likelihood and extract the unknown parameters: rate constants.

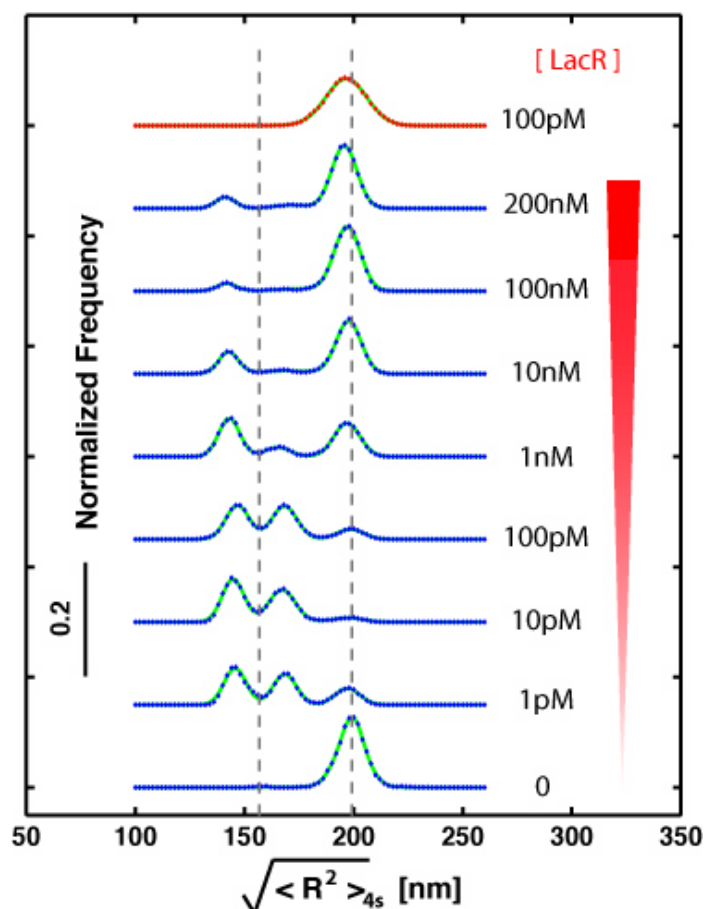


Figure 1.7: RMS motion distribution averaged over 4s at different concentration of Lac repressor. Experimental data are in diamonds (blue: Di-operator DNA; red: Mono-operator DNA) and three-Gaussian fit is in green. Looping is optimized at a certain repressor concentration, above which it is less favored as the repressor concentration goes up. Two gray dashed lines represent the expected motion, based on our calibration measurements (data not shown), for 901bp DNA and the same DNA when 306 bp (the interoperator spacing) are subtracted off of the full length 901bp tether.

concentrations, the equilibrium situation will be dominated by states in which a single repressor molecule is bound to the DNA at the strong operator, punctuated by transient looping events. In the very high concentration limit, each operator will be occupied by a repressor making the formation of a loop nearly impossible. This progression of qualitative behavior revealed in figure 1.7 is then further modeled using thermodynamic model to obtain the important parameters, such as binding energy and looping free energy.

Chapter 5 systematically examine how one single base pair change can alter the looping process. Oscillation in looping probability is observed as the distance between two binding sites changes



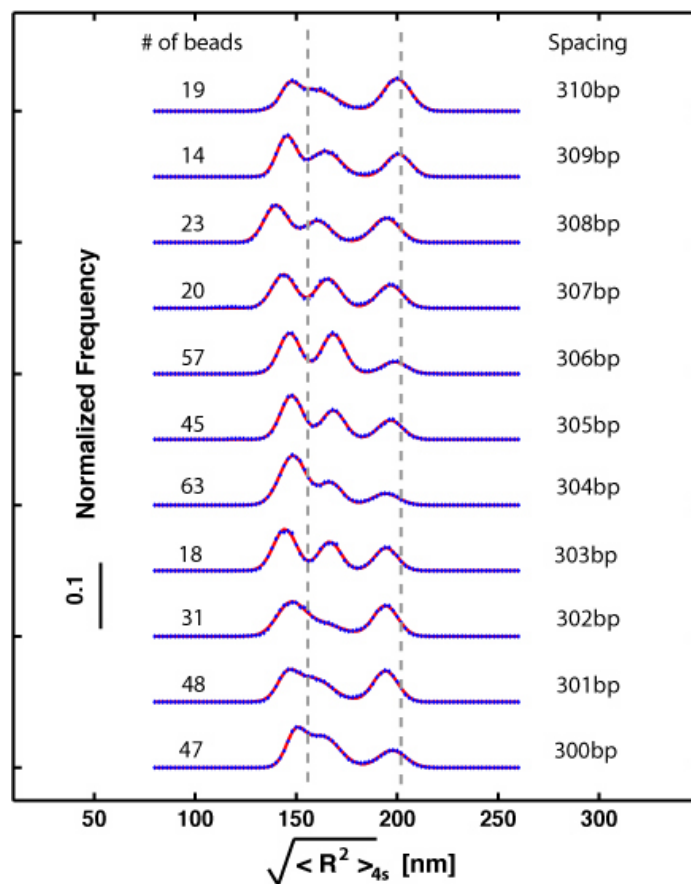


Figure 1.8: Histogram of the Brownian motion for DNAs with two LacI binding sites spaced from 300 to 310 bp. Experimental data are in blue diamonds and three-Gaussian fit is in red. Two gray lines represent the expected motion, based on our calibration measurements (data not shown), for 901bp DNA and the same DNA when 305 bp (the interoperator spacing) are subtracted off of the full length 901bp tether.

from 300 to 310 bp in one-base pair increments. Over the distances used here, looping is rather robust, which is consistent with early *in vivo* studies of the gene repression on the similar distances (Muller *et al.* 1996). Moreover, two different looped configurations are detected with different response to the twist effect. When two binding sites are on the same side of DNA, i.e., in-phase, two looped states experience nearly equally probability; however, when they are on the opposite side of DNA, i.e., out-of-phase, one looped configuration is favored.

Chapter 6 illustrates how sequences with different bendability can enhance and stabilize DNA looping probability. Special nucleosome sequences that are thought to be highly bendable, when placed between the operators of the *lac* operon, form much more stable in-phase loops than random

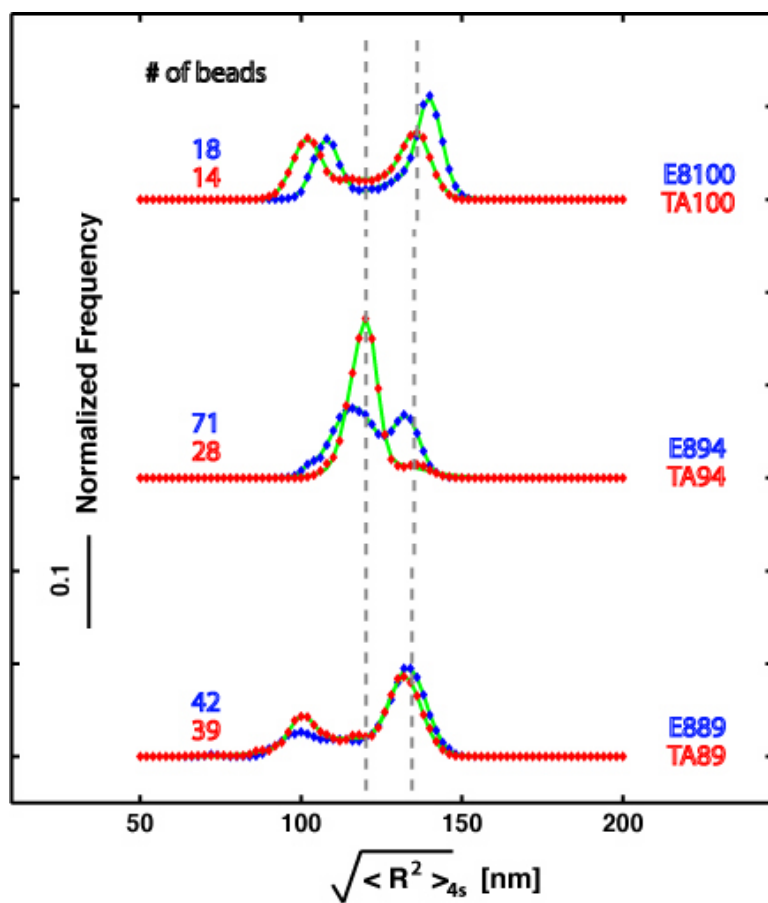


Figure 1.9: Histogram of Brownian motion distribution of different sequences, including when binding sites are in and out of phase. Gray dash lines are expected motion from calibration data.

sequences of the same length, However, for out-of-phase loop formation, the difference between them are relatively small. Moreover, the in-phase and out-of-loops seem to have different configurations.

Appendices are included with detailed protocols that may be followed to replicate the experiments and the DNA sequences used in the experiments.

## Chapter 2

# Calibration of Tethered Particle Motion method

The Tethered Particle Motion (TPM) method has been used to observe and characterize a variety of protein-DNA interactions including DNA looping and transcription. TPM experiments exploit the Brownian motion of a DNA-tethered bead to probe biological dynamics associated with conformational change of the tether. In these experiments, a relative change in the excursion of the bead is used as a reporter of the underlying macromolecular dynamics and is often deemed sufficient for TPM analysis. However, a complete understanding of how the motion depends on the physical properties of the tethered particle complex would permit more quantitative and accurate evaluation of TPM data. For instance, such understanding can help extract details about a looped complex geometry (or multiple coexisting geometries) from TPM data. In order to better characterize the measurement capabilities of TPM experiments involving DNA tethers, we have carried out a detailed calibration of TPM magnitude and intrinsic time constant as a function of DNA length, particle size and ionic strength of the buffer. We also explore how experimental parameters such as acquisition time, exposure time and tracking algorithm affect the apparent motion of the tethered particle. We vary the DNA length from 200 bp to 2.6 kbp and consider particle diameters of 200, 490 and 970 nm. We also present a systematic comparison between measured particle excursions and theoretical expectations, which helps clarify both the experiments and models of DNA dynamics.

## 2.1 Introduction

Single molecule studies are enriching our understanding of biological processes by providing a unique window on the microtrajectories of individual molecules rather than on their ensemble-averaged behavior. A large portion of these studies is devoted to exploring the intricacies of protein-DNA interactions that are central to gene regulation, DNA replication and DNA repair. The resolution of nm-scale distances involved in such interactions poses a significant challenge. The emergence of the tethered particle motion (TPM) method offers a practical and relatively simple solution. In this method, a biopolymer is tethered between a stationary substrate and a small particle which is large enough to be imaged with conventional optical microscopy as shown in figure 2.1. The Brownian motion of the bead serves as a reporter of the underlying macromolecular dynamics. Changes in motion reflect conformational transformations of the tethered molecule. Such changes may be caused by processive walking of RNA polymerase (Schafer *et al.* 1991; Yin *et al.* 1994), DNA looping (Finzi and Gelles 1995; van den Broek *et al.* 2006; Zurla *et al.* 2006, 2007; Wong *et al.* 2007; Vanzi *et al.* 2006; Beausang *et al.* 2007), DNA hybridization (Singh-Zocchi *et al.* 2003), DNA bending (Toli-Nrrelykke *et al.* 2006), Holliday junction formation (Pouget *et al.* 2004) or RNA translation (Vanzi *et al.* 2003).

While TPM is simple in principle, its implementation suffers from a variety of challenges. Sample preparation can be compromised by multiple tethered microspheres, non-specific absorption, transient sticking events and dissociation of the tether joints (Pouget *et al.* 2004; Vanzi *et al.* 2006; van den Broek *et al.* 2006; Blumberg *et al.* 2005; Nelson *et al.* 2006a). In addition, image analysis of TPM data is complicated by instrumental drift and the stochastic nature of the tethered particle's motion. The importance of these complications depends on the relationship between the net observation time, exposure time for a single position measurement and the intrinsic relaxation time of the tethered particle (which depends upon the bead size). Thus, proper quantification of the spatial and temporal resolution of TPM measurements requires an understanding of how particle motion depends on tether length, particle size and other experimental parameters.

To facilitate quantitative TPM measurements we have conducted an extensive characterization of the dynamics of DNA tethered microspheres. We focus exclusively on TPM behavior in the absence of externally applied force (as might be applied via magnetic or optical tweezers). Our aims are threefold. First, we would like to understand what features of sample preparation, data acquisition and data analysis have a marked effect on the reproducibility and reliability of TPM

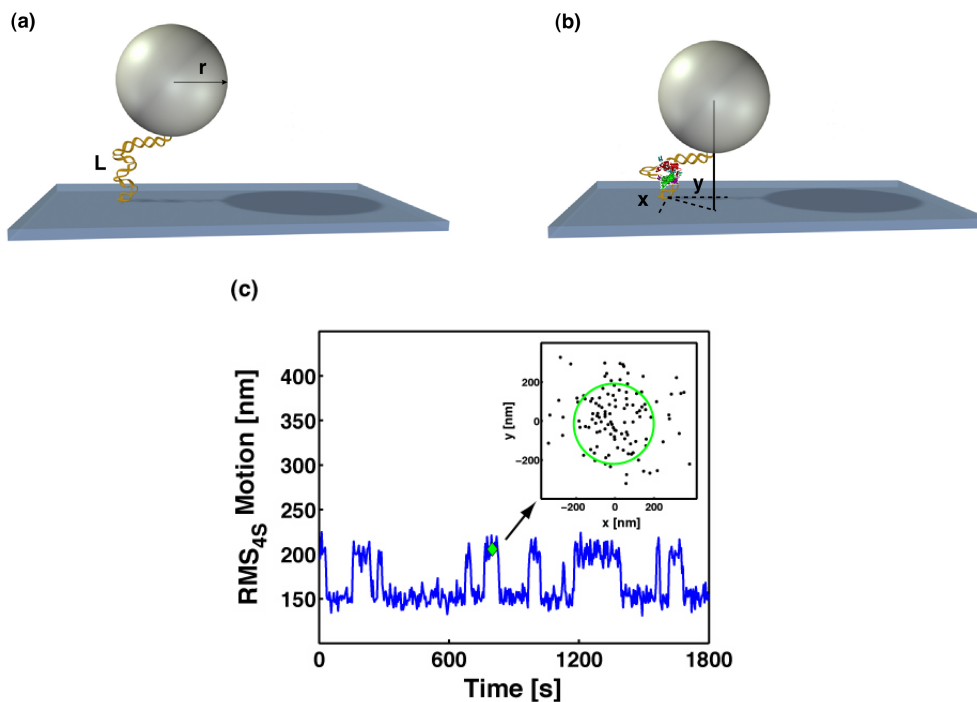


Figure 2.1: Illustration of the tethered particle method. (a) Motion of the tethered bead in the absence of the DNA binding protein. (b) Reduced motion of the tethered bead due to the change in the effective tether length induced by a DNA binding protein that loops or bends the DNA. (c) Simulated looping data with 4s variance filter. The green point is the 4s averaged RMS value at time  $t = 800$ s. The inset shows the x-y positional data corresponding to the RMS trajectory.

measurements. Second, we want to calibrate the relationship between particle motion, tether length and observation time so that subsequent TPM experiments can be quantitatively interpreted. Proper calibration allows more precise predictions concerning the way a particular conformational change of the tether affects TPM and these calibrations have been central to our attempts to understand the complex dynamics of Lac repressor induced looping of DNA. Third, we want to understand the physical processes that govern TPM so that we can assess the significance of our measurements. This is especially timely in light of recent theoretical activity aimed at uncovering the role of bead size, tether length and attachment flexibility in altering observed TPM trajectories.

## 2.2 Results and Discussion

For typical TPM experiments, once video images of the projected position of the bead have been acquired, the first step in the analysis is to determine position traces for all tethered particles. In

the standard DIC microscopy setup that we use, the particles are tracked in the image plane using a cross-correlation method (Gelles *et al.* 1988), which amounts to a two-dimensional projection of the full three-dimensional motion. Tracking of motion in all three dimensions has been accomplished using evanescent fields or diffraction rings (Blumberg *et al.* 2005), but this involves additional calibration and technical challenges. Thus we limit our analysis to two-dimensional tracking. The positions of the bead are subject to a slow drift due to vibrations of the experimental apparatus, which has been removed using a first order Butterworth filter at 0.05Hz cutoff frequency (Vanzi *et al.* 2006). For our Brownian statistic, we choose the square root of the sum of the variances of the drift corrected particle position ( $x, y$ ) along two orthogonal image plane axes as defined by

$$\text{RMS}_t = \sqrt{\langle (x - \bar{x})^2 + (y - \bar{y})^2 \rangle_t}, \quad (2.1)$$

where  $t$  is the time interval over which the RMS motion is measured. equation 2.1 permits us to capture the tether dynamics using a single scalar quantity as is shown in figure 2.1(c).

### 2.2.1 Data Selection Criteria

Though single-particle data often reveal detailed features of the dynamics of protein-DNA interactions, in fact, often such data are plagued with artifacts. In the context of TPM experiments, for example, a fair amount of the collected data is contaminated. In particular, we suspect that artifacts arise mainly due to non-specific binding of the bead and DNA to each other and the surface and multiple DNA attachments on the same bead. To get good calibration data, we want to get our best possible handle on aberrant TPM behavior. For that reason, we utilize several selection criteria to ensure that the particles we observe are "typical" of those that are not sticking to the surface or multiple DNA molecules. In order to root out the unusable trajectories, we have implemented a series of objective criteria, such as "minimum motion", "motion symmetry" and "uniformity" to be described below. Only  $\text{RMS}_{4s}$  data of a tether is used to check whether tethers are qualified or not. The first criterion, "minimum motion" is applied to discard beads which cannot be differentiated from beads stuck to the glass substrate. Due to imperfections of the glass surface and bead (Singh-Zocchi *et al.* 2003), laboratory vibrations, as well as CCD and tracking algorithm resolution uncertainties, a residual motion will persist even for beads stuck to the glass surface. To exemplify this point and quantify the residual motion, a control experiment was carried out with beads but no DNA present. An example of these kinds of trajectories is shown in figure 2.2. For stuck beads,

the motions as characterized by  $\text{RMS}_{4s}$  (in green in the figure) is substantially less than that for the tethered bead (blue curve in the figure), but it is still non-zero. Based on these control experiments, data sets exhibiting average excursions,  $\text{RMS}_{4s}$ , lower than 30 nm cannot be differentiated from stuck beads and are therefore rejected.

The second criterion we impose, "motion symmetry", requires that the motion of the bead must reflect in-plane symmetry consistent with a single DNA tether. In particular, ideally, a tethered particle should exhibit symmetric in-plane motion about its anchor point. The motional symmetry is determined from the covariance matrix (Blumberg *et al.* 2005; Nelson *et al.* 2006a), which is given by

$$C = \begin{pmatrix} \sigma_{x_1x_1} & \sigma_{x_1x_2} \\ \sigma_{x_2x_1} & \sigma_{x_2x_2} \end{pmatrix},$$

where

$$\sigma_{x_i x_j} = \frac{1}{N} \sum_{k=1}^N x_i^k x_j^k - \frac{1}{N^2} \left( \sum_{k=1}^N x_i^k \right) \left( \sum_{k=1}^N x_j^k \right)$$

are the second moments of the bead's position. Here  $N$  is the number of frames for a given run and  $x_1^k, x_2^k$  are the in-plane coordinates (i.e., the position  $x, y$ ) of the microsphere for frame  $k$  as obtained from the raw data. The eigenvalues ( $\lambda_1, \lambda_2$ ) for the covariance matrix are proportional to the square of the major and minor axis corresponding to the in-plane displacement of the bead. Given the radial symmetry of the problem a necessary condition for a "good" tether is one in which the major and minor axes are nearly equal. A measure of equality between the two axes is determined through the symmetry factor  $s = \sqrt{\frac{\lambda_1}{\lambda_2}}$ . In practice, symmetry factors with  $s \leq 1.1$  are deemed acceptable, as they generate reproducible data. (Here we assume  $\lambda_1 > \lambda_2$  without loss of generality.) Fig. 2.2 displays scatter plots for the in-plane motion of several beads to illustrate the distinction between "good" and "bad" tethers. The first plot passes the symmetry test and would serve as a qualified tether while the second would be rejected on the grounds that it fails the symmetry test. A rationale for the asymmetric scatter is that this bead may be attached to multiple DNA molecules (Pouget *et al.* 2004). Such a geometry would be consistent with the observed asymmetric motion.

A third criterion, "uniformity", qualifies tethers on the basis of the consistency of their motion over time and is designed to eliminate unwanted non-specific binding events, such as non-specific binding of DNA to the bead or glass surface for short periods of time. In order to rule out these events, we first divide the whole time series into 10 subwindows and then calculate the average

RMS<sub>4s</sub> motion,  $A_i$ , within each time window  $i$ , followed by:

$$u = \sigma\left(\frac{A_1}{A_{tot}}, \frac{A_2}{A_{tot}}, \dots, \frac{A_i}{A_{tot}}, \dots\right),$$

where  $A_{tot}$  is the overall average RMS<sub>4s</sub> motion. Only data sets with standard deviations between 0 and 0.2 are accepted. The bottom figure of 2.2(c) displays what appears to be a non-specific binding event, even though the motion of the bead meets the symmetry criterion.

The first two selection criteria discard tethers which are corrupt from the start, while the third selection criterion eliminates unwanted interactions that occur during some part of the run. However, if the purpose of the experiment is to identify interesting molecular binding events such as those leading to DNA looping or bending, then the last criterion cannot be applied since these binding events have the same signature as those rejected by the uniformity criterion. On the other hand, it should be enforced in cases like ours, which are aimed at characterizing simple DNA tethers. Filtering the data that fail the last criterion leaves tethers which experience a consistency in their Brownian motion. Figure 2.2(f) displays a broad distribution in the measured RMS<sub>4s</sub> before selection (in red). Upon applying our three selection criteria, around 50% of the data are qualified, displaying a well-defined measure for the Brownian motion (in pink). This figure suggests that the main source of the corruption comes from the asymmetric in-plane motion. The simplest fix for suppressing these events is reducing the concentration of the DNA used in the experiments. Our reason for showing the data is to illustrate some of the difficulties that arise in making careful TPM measurements.

## 2.2.2 Acquisition Time

As noted above, the positional data are converted into the actual statistic we use to characterize the motion by measuring the RMS motion over a particular time window (usually four seconds). The choice of an optimal time interval is one of the questions that should concern anyone using the tethered particle method. An excessively short time interval may not capture the full phase space available to the tethers, while long-term measurements result in a loss of temporal resolution. An additional consideration is that there is statistical error due to the stochastic nature of Brownian motion. This means that even if there were no drift and our particle tracking algorithm was infinitely precise, there would still be error in measuring the motion of a tethered particle over a defined acquisition time.



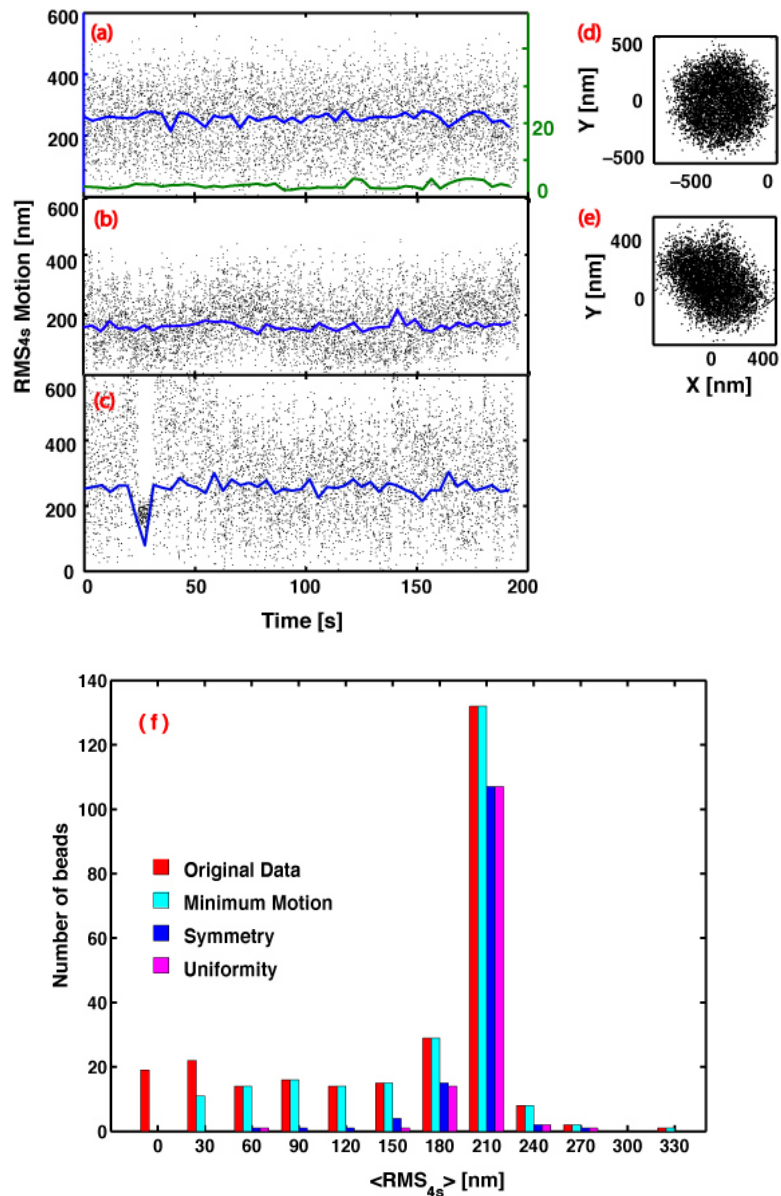


Figure 2.2: Selection of qualified tethers. Displays of good and bad data, which inspire the selection criteria. (a) Trajectory associated with a good data set. The green line corresponds to the 4-second averaged RMS motion in a control experiment with a stuck bead. (b) Trajectory for a bead that failed the motion symmetry test. (c) Trajectory associated with nonuniform motion caused by nonspecific binding, seen as a downward spike between 0 and 50 seconds. (d) Scatter plot of in-plane motion. This case reveals symmetric motion. (e) Scatter plot of in-plane motion. This case is asymmetric and corresponds to data that are rejected. In all figures, the dots correspond to raw data and the blue line to data after averaging over a four-second time window. The DNA used here is 1206 bp long and the bead size is 490 nm in diameter. (f) Distribution of bead excursions with selection rules. Three filters are applied progressively. Original data: red; after application of minimal motion filter: cyan; after application of symmetry filter: blue; and after application of final filter: pink. The DNA used here is 901bp long and bead size is 490nm in diameter.

To obtain an intuitive feel for how acquisition time effects both the average value and averaged standard deviation in motion measurements, we examined the statistics of bead motions for different choices of such time interval as shown in figure 2.3. The average value in motion is calculated as  $\langle \text{RMS}_{\text{time}} \rangle$ , which is averaged over entire population. The averaged standard deviation relates the error in motion measurements, which is referred as  $\text{Std}_{\text{time}}$ , where

$$\text{Std}_{\text{time}} = \sqrt{\langle \text{RMS}_{\text{time}}^2 \rangle - \langle \text{RMS}_{\text{time}} \rangle^2},$$

We consider several different statistics just to see how different statistical measures respond to choice of the averaging time interval. Each point in the figure corresponds to quantities measured over a defined time interval. The trajectories used to generate this data were 200 seconds long. The graphs in the top row illustrate how the averaged RMS motion depends upon acquisition time. DNA lengths range from 200 to almost 3000 bp and we report data for beads with three different diameters: 200, 490 and 970 nm.

We find that the amplitude of the RMS motion first increases with the number of successive images used for its calculation, before saturating at the equilibrium value. To understand this better, imagine a bead in a harmonic well. There is a characteristic timescale over which the bead remembers its previous positions. This timescale is given by  $\tau_0 = \frac{\gamma}{k_{\text{eff}}}$ , where  $\gamma$  is the drag coefficient and  $k_{\text{eff}}$  is the spring constant. If the time interval is smaller than  $\tau_0$ , the measured motion will have a big variation. If the time interval is a lot larger than  $\tau_0$ , the measured motion will represent the equilibrium motion. When bead size is fixed, as the tether length increases, both the motion and the time required for the bead to explore sufficient configurations increases. This is expected due to the increased excursion range of the tether, and longer tethers tend to have smaller effective spring constants. When the length of DNA is fixed, as bead size increases, both the equilibrium motion and equilibration time increase. An intuitive explanation of the bead size effect is that the radius of the bead is added on to the tether length itself. However, it is not quite so simple because the particle does not have a persistence length. In addition, the effect of the bead's presence is a little more complicated because the particle can rotate around its attachment point and because of excluded volume effects between the particle and the surface (Segall *et al.* 2006b). Both of these latter effects reduce the measured motion. The increase in equilibration time results from the viscous drag of larger beads since the drag,  $\gamma$ , is linearly related to the bead's radius, and the time constant is linearly related to the viscous drag (for a quadratic potential). The graphs in the bottom

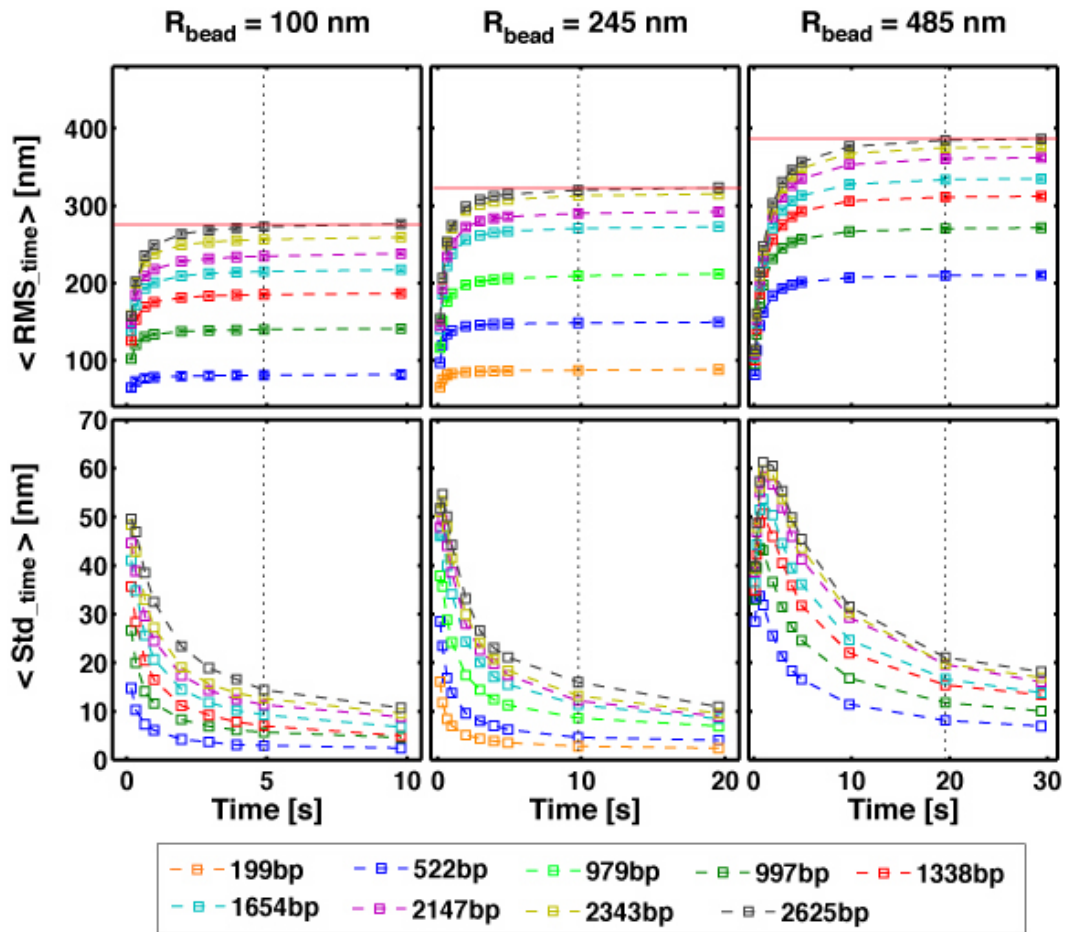


Figure 2.3: Amplitude of Brownian motion and the average standard deviation as a function of the averaging time for different length DNA tether lengths. The length of the DNA varies from 199 to 2625 bp. The top row shows the averaged RMS motion as a function of time interval, and the lower row shows the standard deviation as a function of time interval. Black dotted lines indicate the time over which the system reaches equilibrium. Red lines made by hand suggest the motion reaches saturation.

row demonstrate how the statistical error,  $\text{Std}_{\text{time}}$ , is related to such time interval. As the number of the successive images used for the RMS calculation increases, the error decreases. Larger beads can introduce larger fluctuation, because of their slow diffusion.

### 2.2.3 Calibration of Motion

TPM experiments are generally designed for detection of a discrete conformational change of biopolymers. It is necessary to see how the Brownian motion changes as a function of tethered DNA length. A further benefit of calibration data is that we can get a sense of how big of a change we will be able to measure. For example, can the TPM technique resolve 100 bp looping? What about 90-degree bending? Can we distinguish two looped configurations that only differ by the angle the DNA makes upon leaving the loop? Lastly, we can compare our calibration data to theoretical models of TPM motion. This helps to identify relevant physical principles that might affect the rate of biomolecular interactions and increases our confidence in both the experimental protocol, and in the mathematical models themselves.

As discussed before, the measured amplitude of RMS motion depends on the time interval over which it is calculated (figure 2.3). To unequivocally determine the relation between the measured excursion and the tethered DNA length, we choose the timescale that guarantees that we will attain saturation motion for the longest DNA molecules. As indicated in figure 2.3, we systematically used 5, 10 and 20 seconds for beads with diameters of 200, 490 and 970 nm, respectively.

The concept of the experiments is to carry out a series of TPM measurements at each tether length and for each bead size. The point is to quantify the expected motion for different choices of the tether length and the bead size since both of these geometric parameters are reflected in the measured bead excursions. These data are shown in figure 2.4. Each point in that figure is the average of RMS motion of 20 to 200 qualified beads. To characterize the relationship of equilibrium RMS motion to the actual tethered DNA length, a second-order polynomial was fitted to the experimental data. The fitting parameters are listed in Table 2.1. The fits are valid only over a limited range of tether lengths, but serve as a useful scheme for interpolating when performing real DNA looping measurements, for example. For very short and long tether lengths, these fits are no longer valid and result in spurious artifacts. Black dashed curves represent motion averaged over 4 seconds for the three bead sizes covered, demonstrating the measured motion of larger beads or longer DNA doesn't reach the saturation for very short acquisition times.

Intuitively, the slope of these curves can be interpreted as the sensitivity of the TPM, i.e., the

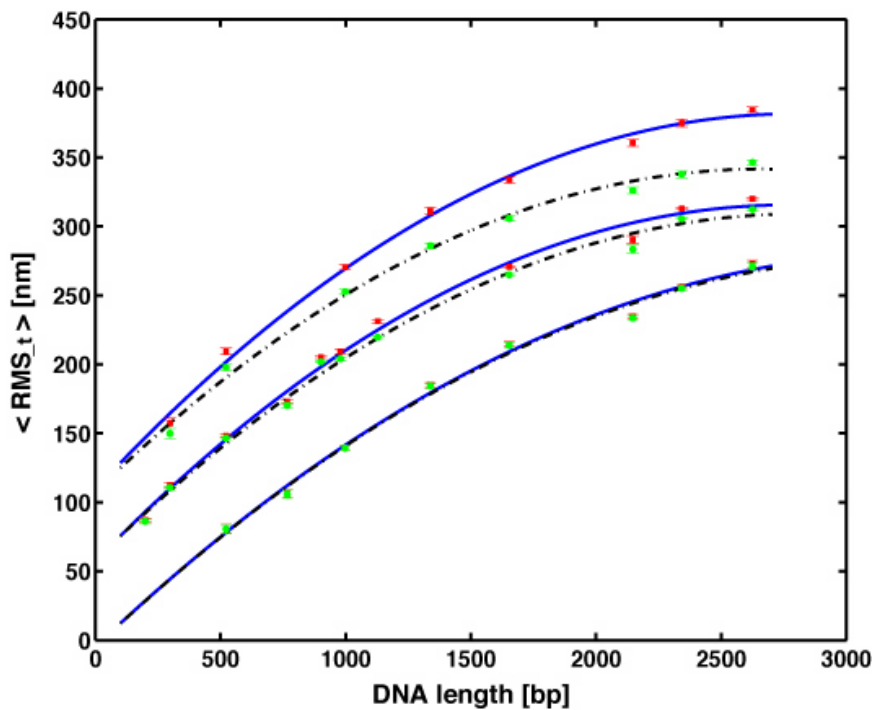


Figure 2.4: Brownian motion of bead as a function of the tether length for different sized microspheres. Each red point in the figure is the average of equilibrium amplitude of RMS motion over 20 to 200 qualified beads, which is calculated using equation 2.1 with  $t = 5$  seconds for  $R = 100$  nm (bottom data set),  $t = 10$  seconds for  $R = 245$  nm (middle data set) and  $t = 20$  seconds for  $R = 485$  nm (top data set). Points in green represents the RMS motion averaged over 4s for the same data sets. Points: experimental data for different sized beads. The blue and black dashed curves are second-order polynomial fits to the datasets obtained by using the different averaging times.

Table 2.1: Parameters of quadratic function  $ax^2 + bx + c$  obtained for fitting both the equilibrium motion data and 4 seconds interval data

Time [s]	Diameter	$a \times 10^{-5}$	b	c
5	200 nm	$-2.58 \pm 0.68$	$0.17 \pm 0.02$	$-4.5 \pm 14.8$
10	490 nm	$-3.37 \pm 0.47$	$0.19 \pm 0.01$	$57.3 \pm 7.2$
20	970 nm	$-3.49 \pm 0.46$	$0.20 \pm 0.01$	$109.5 \pm 8.7$
4	200 nm	$-2.60 \pm 0.69$	$0.17 \pm 0.02$	$-4.75 \pm 14.7$
4	490 nm	$-3.17 \pm 0.41$	$0.18 \pm 0.01$	$58.05 \pm 6.6$
4	970 nm	$-3.31 \pm 0.48$	$0.18 \pm 0.01$	$107.7 \pm 8.7$

ability of TPM to resolve different tether lengths. However caution must be used in this analysis, because the accompanying statistical noise is not constant. More explicitly, a comparison of the slopes reveals that larger beads provide better spatial resolution, however, their slower diffusion also implies more statistical noise in an a measurement. Thus in each experiment, a compromise must be sought between good spatial (large beads) and temporal (small beads) resolution.

To gain more confidence in our understanding of the apparatus and analysis, we compared our experimental calibration curve to one we generated from a mathematical model of the system, following Nelson *et al.* (2006b). The root-mean-square excursion of the projected bead location away from its tether point is controlled by various competing effects: (i) In the absence of any bead or wall, the chain’s endpoint would execute 3D Brownian motion subject to a restoring force from the tether. (ii) But the bead’s rotatory Brownian motion implies that its center, which is what we observe, lies a considerable distance away from the chain endpoint. (iii) The bead–wall exclusion pushes the bead upward, reducing its transverse excursions.

To account for all these effects, we modified the Gaussian sampling method previously used in (Segall *et al.* 2006b; Nelson *et al.* 2006b; Czaplá *et al.* 2006) (see smearing section). Our code generated many simulated DNA chains and bead orientations, applied the steric constraints (Segall *et al.* 2006b), and reported the RMS deviation of the bead center from the attachment point. We applied a correction to this theoretical result, to account for the bead’s motion during the rather long shutter time (about 31 msec, see smearing section). Figure 2.5 shows that an *a priori* calculation of the expected motion matches the data fairly well with no fitting at all. The remaining discrepancy with our data may reflect unremoved instrumental drift, for example high-frequency motion that our Butterworth filter cannot distinguish from true Brownian motion. The variance of such noise would add in quadrature with the true Brownian motion, shifting the entire curve up by a constant.

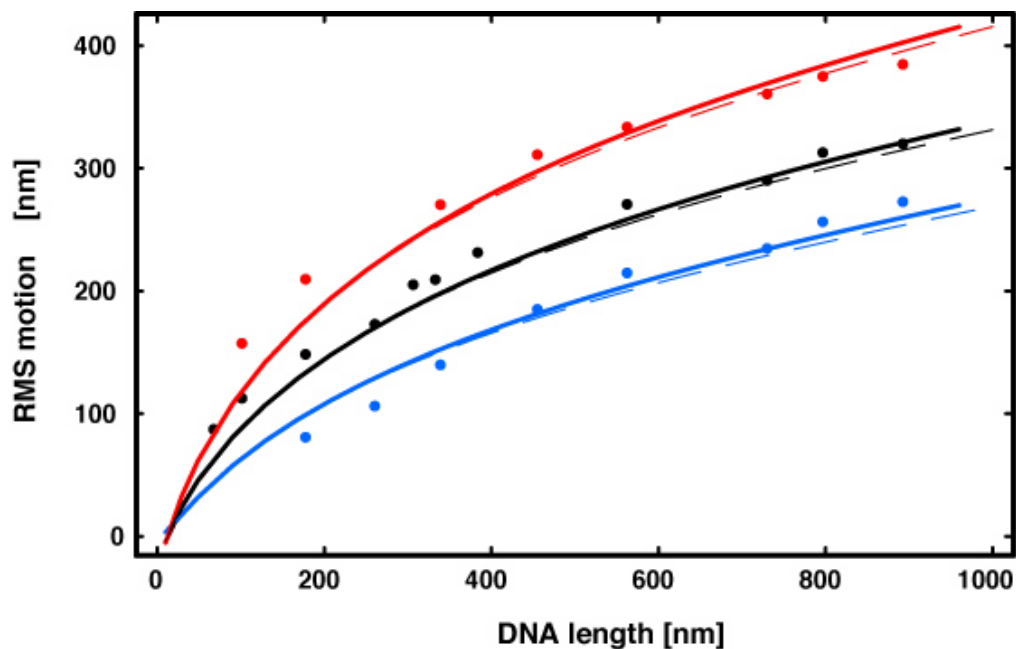


Figure 2.5: Experimental values for RMS motion of bead center, projected to the plane of the surface, averaged over 5, 10, and 20 seconds of observation for bead size in radius 100, 245 and 485nm, respectively. Each dot represents the average of approximately 20 to 200 different observed beads with the given tether length. Curve: theoretically predicted RMS motion, corrected for the blurring effect of our long shutter time. The curve is a *zero parameter* prediction based on the known value of DNA persistence length in solvent conditions like ours,  $\xi = 45$  nm Strick *et al.* (1998), and the manufacturer's specification for the bead radius.

### 2.2.4 Time Constant as a Function of Tether Length and Bead Size

Having explored the dependency of the spatial resolution of the tethered particle method on parameters such as bead size and contour length, we now explore the dependency of these parameters on temporal resolution. A measure of the temporal resolution in the TPM is the time constant  $\tau$  of the bead associated with its motion. The time constant of the bead gives a measure of how long it takes the bead to return to its equilibrium position and hence equilibrate. The larger the time constant, the longer the bead takes to equilibrate and the longer it takes to reduce thermal noise. Ideally, time constants should be "short" compared to the time scale governing the dynamics of the biomolecular interactions of interest.

To determine the time constants from the data we compute the auto-correlation function

$$\langle \delta \mathbf{r}_{\perp}(n\Delta t) \cdot \delta \mathbf{r}_{\perp}(0) \rangle_{\text{exp}} = \frac{1}{N} \sum_i^N \mathbf{r}_{\perp}((i+n)\Delta t) \cdot \mathbf{r}_{\perp}(i\Delta t) - \left( \frac{1}{N} \sum_i^N \mathbf{r}_{\perp}^2(i\Delta t) \right), \quad (2.2)$$

where  $\Delta t$  is the inverse of the frame rate and  $N$  is the number of frames taken. Given the experimental auto-correlation function we make the assumption that it decays as a single exponential

$$\langle \delta \mathbf{r}_{\perp}(n\Delta t) \cdot \delta \mathbf{r}_{\perp}(0) \rangle_{\text{exp}} \approx Ae^{-t/\tau}. \quad (2.3)$$

Applying a least-squares fit of Eq. 2.3 to the experimental results we obtain the associated time constant for in-plane motion.

Figure 2.6 displays a sample of the experimentally computed autocorrelation function and the single exponential fit for a given tether length and three-bead radii. Figure 2.7 displays a histogram of all of the time constants computed for that given tether length.

In figure 2.8 we display the experimentally determined time constants as a function of tether length for various bead sizes. The time constants increase as a function of tether length and bead size. The increase in time constants with respect to both parameters is expected. Larger beads have a stronger drag force and therefore take a longer time to return to equilibrium; longer tethers are floppier molecules which take a longer time to equilibrate.

### 2.2.5 Ionic Effect

TPM is often used to detect the conformational change of DNA due to protein binding. Most of these binding interactions depend strongly on ionic conditions. Here we investigate whether ion



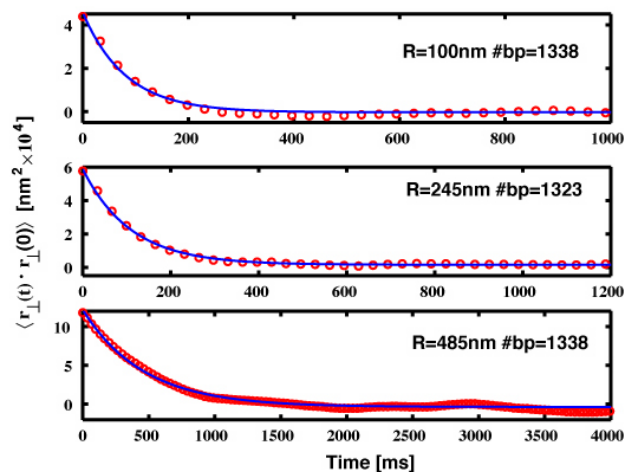


Figure 2.6: Autocorrelation function for a single measurement. Top: bead radius 100 nm and 1338 base pairs. Middle: bead radius of 245 nm and 1323 base pairs. Bottom: bead radius of 485 nm and 1338 basepairs

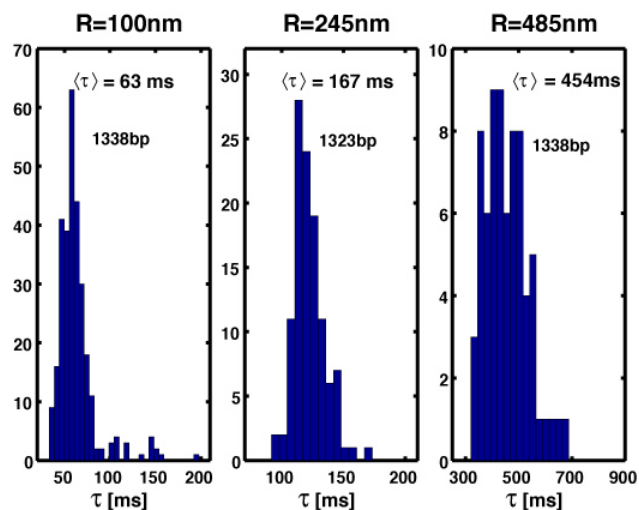


Figure 2.7: Histogram of time constants calculated. Left: bead radius 100 nm and 1338 base pairs. Center: bead radius of 245 nm and 1323 base pairs. Right: bead radius of 485 nm and 1338 basepairs

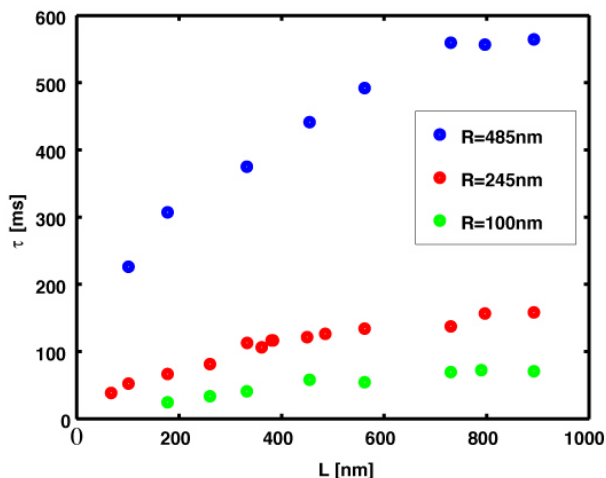


Figure 2.8: Time constant as a function of contour length and bead radius.

conditions also play a role in simple tethered particle motion. Figure 2.9 plot the Brownian motion at different KCl and  $\text{MgCl}_2$  concentration, respectively. We find that the motion appears to increase by 20 nm for KCl concentration larger 100 mM, but then levels off. This is a surprising result because prior research has shown that the persistence length of DNA does not change much over this range of ionic strength (Baumann *et al.* 1997) and that the change is such that the persistence length increases at low salt (due to removal of ionic screening). One possible explanation is that there are some interaction between the glass surface and the beads. At relative low salt concentration, such attractive interaction is not all screened off.

With  $\text{MgCl}_2$ , the motion barely changes. This is also surprising because DNA is thought to compact at this concentration of Mg (Baumann *et al.* 1997). It could be that our buffer ingredients have a shielding effect that Bustamante *et al.* did not see. Although we do not understand the trends with ionic concentration, for our current purposes it is useful to note that around 130 mM KCl and 4 mM Mg, there is little effect of ion concentration. Thus, it is not a significant source of error in our calibration curves.

### 2.2.6 Smearing Effect

The smearing effect refers to the time averaging due to the finite exposure time of the camera detector. In order to study such an effect on the RMS motion of the tethered particle, an experiment was performed with 901bp DNA and a 490 nm diameter bead. Signals of such tethered beads were detected at different camera exposure times. The movies were captured at 1, 5, 10, 30, 50, 100, 200,

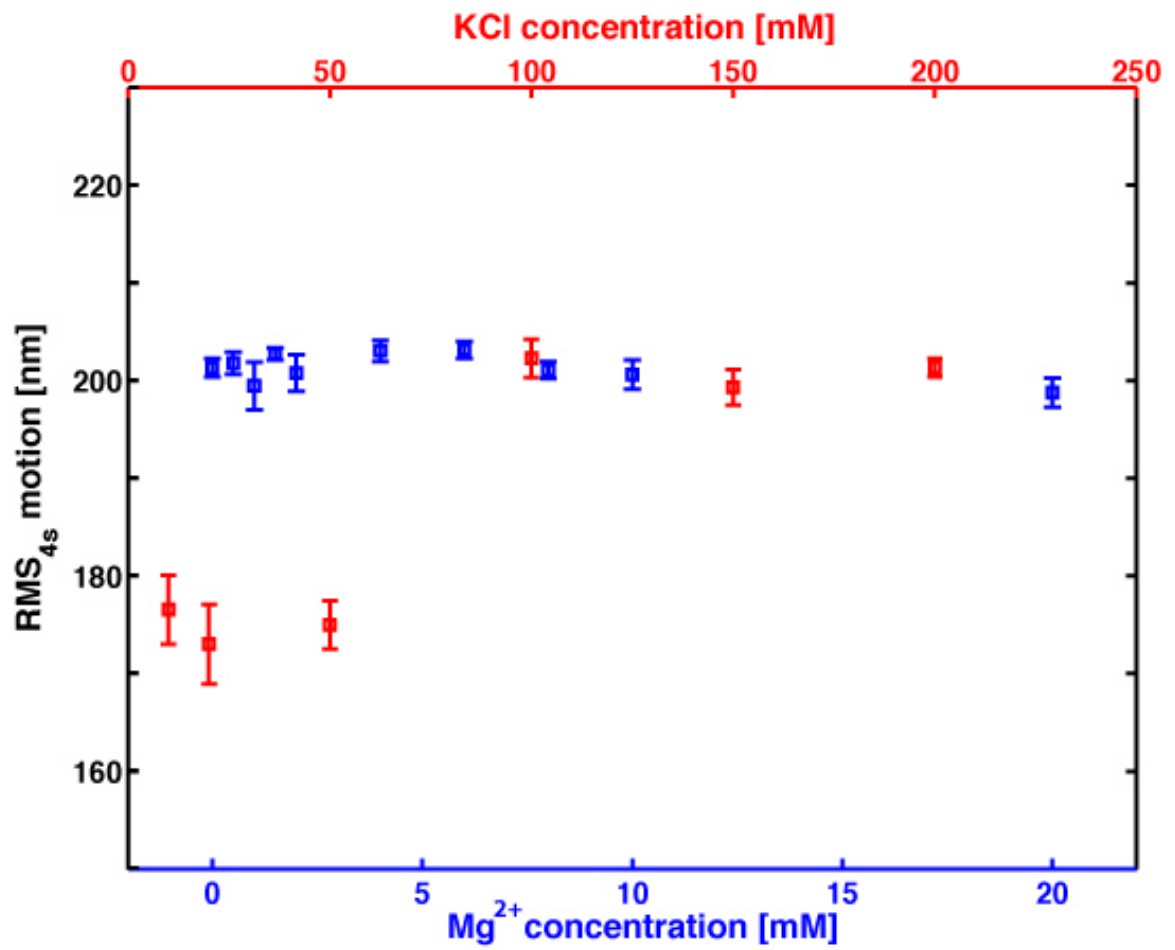


Figure 2.9: The RMS<sub>4s</sub> motion of 901bp DNA tethered beads in different concentrations of KCl and MgCl<sub>2</sub>.

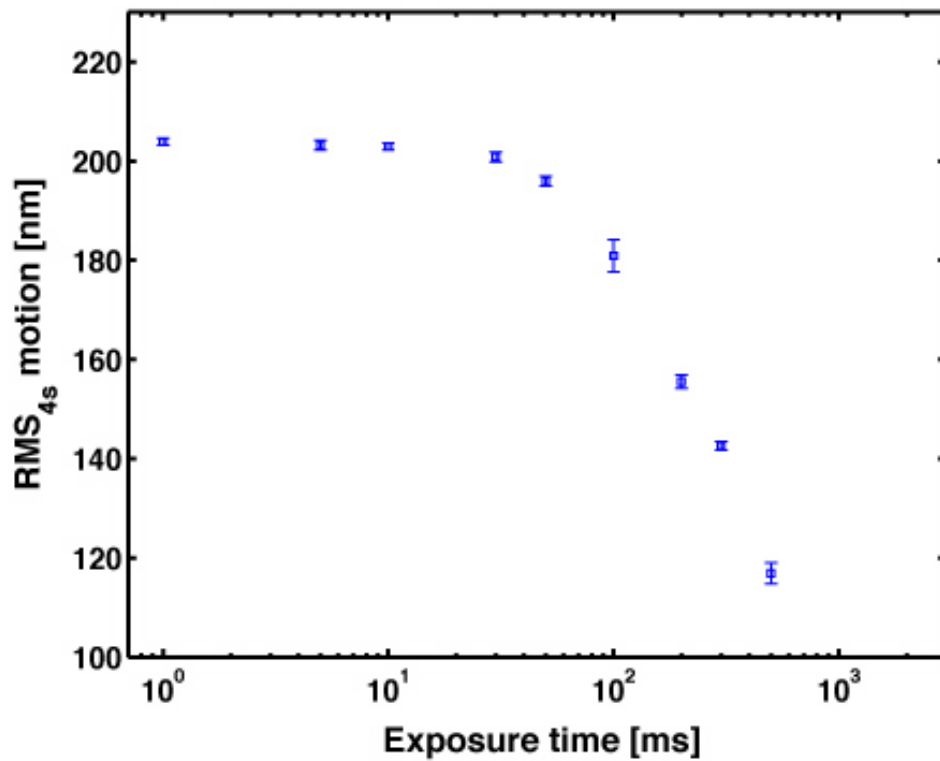


Figure 2.10: The RMS<sub>4s</sub> motion as a function of camera exposure time. The effect of camera exposure time was studied by examining the bead's RMS motion for different choices of exposure time. Each data point corresponds to a different camera setting.

300 and 500 ms exposure durations. The effects of averaging over the camera's exposure time were apparent. Figure 2.10 shows the RMS motion at different exposure times. It reaches saturation when the exposure time is smaller than 30ms, and decreases sharply above this value.

These effects can be considered from a theoretical perspective. Here we describe the correction to the theoretical calibration curve in figure 2.5 to account for bead motion during a single frame exposure. The image of a static bead is a 2D distribution of intensity,  $I_s(\mathbf{r} - \mathbf{r}_0)$ , where  $\mathbf{r}$  is the projected position in the focal plane. It reflects the "actual" bead image, the microscope pointspread function, uncertainties from finite pixel size, etc.

Finite shutter time blurs the image of a diffusing particle. As an extreme example, suppose the shutter were open for a time much longer than the bead's time to diffuse through its range of motion; then we would observe a blurred image centered on zero, and larger than the static image of the bead. In fact, some TPM implementations study this enlarged apparent bead image (Finzi and Gelles 1995). The bead-tracking method, which we use, discards the apparent image size and instead studies the apparent bead center position as a function of time. We now ask, how is this apparent bead center related to the true instantaneous bead position?

If we leave the shutter open only for a very short time, say 1 msec out of the total video frame time, then we may expect there would be very little blurring (Nelson *et al.* 2006b). But in the present work, the shutter is open for almost the entire 33 msec video frame; we need a correction to account for this fact.

Suppose we knew that at some time  $t$  the bead's true position is  $\mathbf{r}_0$ . This is the quantity we want but can not observe directly. At a later time  $t + \delta\tau$ , we only know the probability distribution function (pdf) of the bead's possible positions: It is centered on a new point  $\mathbf{r}_{\delta\tau}$ . For tethered 2D Brownian motion, and infinitesimal  $\delta\tau$ , the new distribution  $P(\mathbf{r}; \delta\tau)$  is a Gaussian of width  $\sqrt{2D\delta\tau}$  centered on  $\mathbf{r}_{\delta\tau} = \mathbf{r}_0 + (\mathbf{f}/\zeta)\delta\tau$ , where  $\mathbf{f}$  is the restoring force of the tether,  $\zeta = k_B T/D = 6\pi\eta$  is the Stokes drag constant, and  $\eta$  is the viscosity of water. We can estimate the force by the Gaussian chain approximation,  $\mathbf{f} = -k_B T \mathbf{r}_0 / (L\xi)$  where  $\xi$  is the persistence length. The average expected image at time  $t + \delta\tau$  is then the convolution of the static image  $I_s$  with  $P$ . This intensity distribution is centered at  $\mathbf{r}_{\delta\tau}$ .

We can find the average blurred image by dividing the finite shutter time  $\delta t$  into small slices  $\delta\tau$ , finding the expected average image at each  $\delta\tau$ , and adding them all together. The average blurred image will be stretched relative to the static image, and its center will just be the average of the various  $\mathbf{r}_{\delta\tau}$ . This center will be shifted radially inward relative to the initial  $\mathbf{r}_0$ , so call it  $S(\rho_0)\mathbf{r}_0$ , where  $\rho_0 = |\mathbf{r}_0|$ .  $S(\rho_0) < 1$  is a scale factor function that we wish to find.

In the framework of the above approximations, the center  $\rho_{\delta\tau}$  obeys

$$\frac{d\rho}{d\delta\tau} = v(\rho) = -\frac{1}{6\pi\eta R_b} \frac{k_B T}{L\xi} \rho, \quad (2.4)$$

Let  $T_s = 6\pi\eta R_b L\xi/k_B T$ . So  $\rho(\delta\tau) = \rho_0 e^{-\delta\tau/T_s}$ . The average of this center position over a finite shutter time  $\delta t$  is  $S(\rho_0)\rho_0$  where

$$S(\rho_0) = \frac{T_s}{\delta t} [1 - e^{-\delta t/T_s}], \quad (2.5)$$

Notice  $S$  is actually independent of  $\rho_0$ . For very small  $\delta t$  we get  $S \rightarrow 1 - \frac{1}{2}(\delta t/T_s)$ . For large  $\delta t$ , we have  $S \rightarrow 0$ .

We conclude that every report of  $\mathbf{r}$  is systematically too small by a factor of  $S$ , which depends on the shutter time  $\delta t = 31$  msec and the tether length  $L$  (and other fixed quantities). If we want to predict the experimental data we should take the theoretical prediction, e.g., for  $\sqrt{\langle \rho^2 \rangle}$ , and correct it, here by a factor of  $S$ . This correction is trivial to apply (comes out of the averaging sign), because  $S$  is independent of  $\rho_0$ .

The preceding discussion made some poor approximations. For example the drag constant is much bigger than the Stokes-law formula used above, due to wall effects. Nor is the tether end-end distance equal to  $\rho$  (there is also the distance from bead attachment to bead center, plus foreshortening due to projection to  $x - y$  plane). Nor is the tether's entropic elasticity well represented by the Gaussian-chain formula. For all these reasons, we replaced Eqs. 2.4–2.5 by a phenomenological formula obtained from our data. We computed the average shift in the apparent  $\mathbf{r}$  from video frame  $N$  to  $N + 2$ , separated by  $\Delta t = 66$  msec, and noted it was of the radially symmetric form  $\langle \rho' \rangle_\rho / \rho \approx 0.55$ . We then replaced Eq. 2.4 by

$$\frac{d\rho}{d\delta\tau} = v(\rho) = -\frac{V_*}{L} \rho, \quad (2.6)$$

where  $V_*$  is a constant determined from the observed frame-to-frame shifts via Eq. 2.6, that is,  $e^{-(V_*/L)\Delta t} = \langle \rho' \rangle_\rho / \rho$ , or  $V_*\delta t = 87$  nm/msec. Then Eq. 2.5 becomes

$$S = \frac{L}{V_*\delta t} [1 - e^{-\delta t V_*/L}], \quad (2.7)$$

The corrected curve in figure 2.5 was obtained from the Monte Carlo simulation described in (Segall

*et al.* 2006b; Nelson *et al.* 2006b) by multiplying by  $S$ .

### 2.3 Applications to DNA Looping

One of the signal achievements of the tethered particle method has been its use in studying DNA looping. Many transcriptional regulatory motifs involve the binding of transcription factors that bind at more than one site simultaneously forming a loop of the intervening DNA. The TPM technique has been used to explore these problems. The calibration analysis performed here can serve as the basis of a more careful evaluation of DNA looping and bending by DNA-binding proteins.

The tethered particle motion (TPM) has proved to be a valuable technique for studying conformational changes of DNA molecules via looping and bending processes. Going one step further, a proper, systematic calibration of TPM measurements provides information relating the bead's observed motion to the actual tethered length of DNA. This can be used to optimize the design of subsequent DNA looping experiments; for example, what is the optimal total DNA length and bead size needed to detect a 305 bp loop with a good signal-to-noise ratio? TPM sensitivity is influenced by the reporter bead size: smaller beads give better temporal resolution, but at the expense of poorer spatial resolution. Thus if the aim of the experiment is just to obtain information on conformational change, a large bead ought to be used, because for a given length of DNA, a larger bead translates to more change in its motion. Conversely, if time resolution is the most important factor, smaller beads should be used. Once the bead size has been determined by the experimental objectives, our aim is to find a good length of DNA that will maximize the measurable response to conformational changes. In principle, one should use tethers such that the difference between the two states of interest corresponds to a region over which the calibration curve has the largest slope since the motion is not in a linear relationship with actual tethered DNA length. For example, suppose that the bead's diameter is fixed at 490 nm and the loop that we would like to detect is 325 bp (distance between two binding sites). Choosing DNA length around 2500 bp would be a bad idea, because the relative change will be too small to be differentiated from fluctuation. How about 1100 and 900 bp? The calibration curve tells us a total length of 1100 bp will be worse than 900 bp for detecting a 325 bp loop, because the change in motion is smaller. As illustrated in figure 2.11, 900 bp long DNA indeed offers a better sensitivity, resulting in differentiating all possible states nicely. In comparison, when DNA is 1128 bp in total length, the looping signal is not as well resolved as the one in 900 bp.

Another benefit derived from the calibration curve, as illustrated in figure 2.12, is a better un-

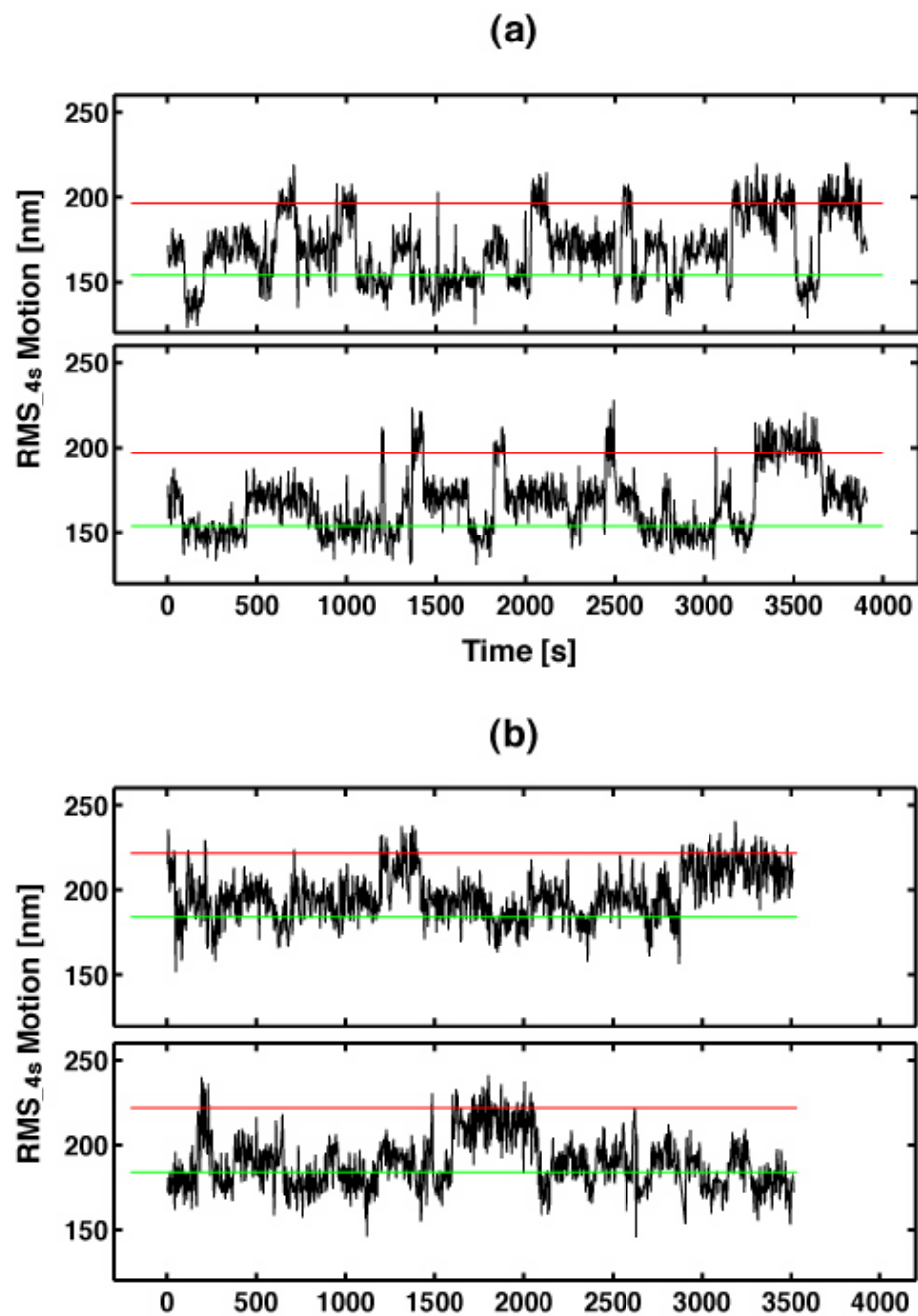


Figure 2.11: Comparison of typical trajectories averaged over 4 seconds in the presence of Lac repressor showing loop formation and breakdown between DNAs with different tether lengths. (a) Total length of DNA is 901 bp. (b) Total length of DNA is 1128 bp. Operator center to center distance is 325.5 bp, which is the same for both DNAs.



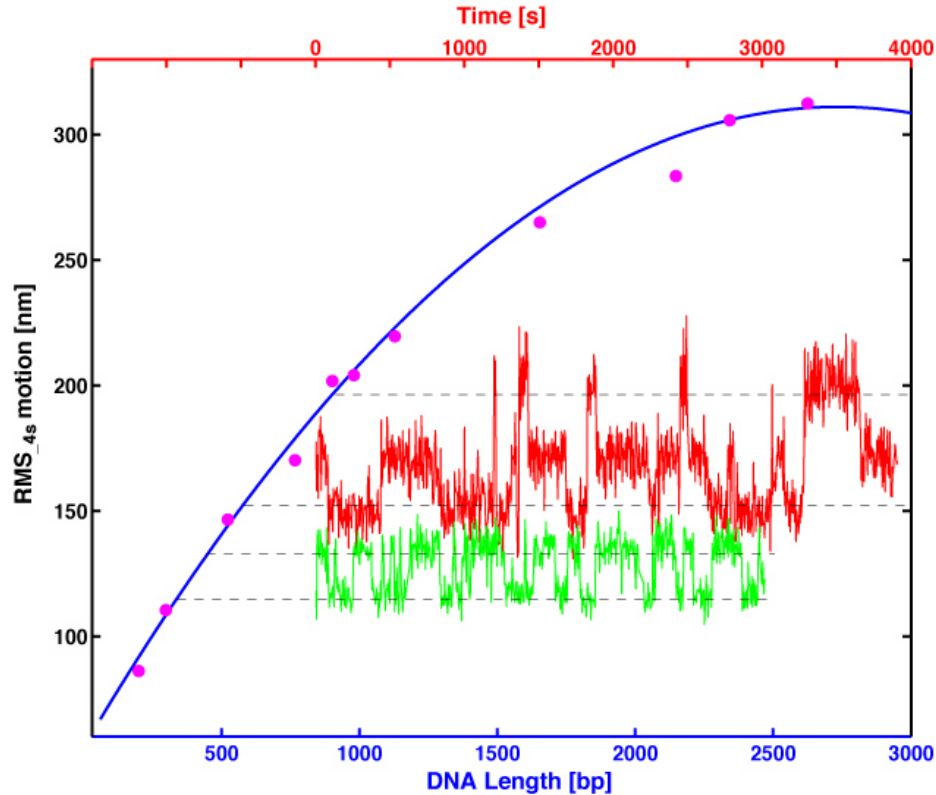


Figure 2.12: Calibration curve for 490nm size bead and its application in looping experiments. Blue: calibration curve for  $\text{RMS}_{4s}$  motion as a function of tethered DNA length. Red:  $\text{RMS}_{4s}$  motion in the presence of 100pM LacI. DNA is 901bp long with operator center to center distance of 325.5 bp. Green:  $\text{RMS}_{4s}$  motion in the presence of 100 pM LacI. DNA is 450 bp long with operator center to center distance of 114.5 bp. Each pair of black dashed lines represent the expected motion for actual total length and the length subtracts off the operator center to center distance, respectively.

Understanding of the geometry of the conformational changes we have studied. For example, there are three clearly separated states shown in the red line in figure 2.12, where the total length of DNA is 901bp and the loop size is about 325 bp. The first two black dashed lines from the top represent the expected motion for 901bp long DNA, and 901 minus 325 bp, respectively. By comparing the effective tether length upon loop formation to them, we might gain insight into what the two looped species are.

## 2.4 Conclusions

The tethered particle motion method is one of the simplest tools for performing single-molecule experiments on DNA-protein complexes. The central idea is to use the Brownian motion of a small particle tethered to a DNA molecule as a reporter of the underlying macromolecular dynamics of the DNA in its complexes with DNA-binding proteins. The point of this chapter has been to examine the challenges that are inherent in making useful quantitative measurements using this method. One of the main outcomes of that effort has been the development of calibration curves which illustrate how tethered-particle excursions depend upon both bead size and tether length. Our own recent calculations argued that the presence of a bead at the end of a tethered DNA molecule will lead to excluded-volume forces that will tend to elongate the DNA. This effect depends upon the bead size. The work presented here largely corroborates this picture.

## Chapter 3

# Data Analysis

In our Tethered Particle Motion (TPM) experiments, DNA is anchored on a coverslide with a microsphere attached to the other end. By measuring the range of diffusion of a tethered bead, we can detect the conformational change of single DNA molecules due to Lac repressor (LacI) mediated DNA looping at the single molecule level. Furthermore, The kinetics of the looping and unlooping processes can also be obtained. The goal of this chapter is to show the progressive of ideas uading from raw images of beads to looping trajectories.

### 3.1 Particle Tracking

The motion of the bead is recorded through a Differential Interference Contrast (DIC) microscope at 30 frames per second. Figure 3.1 (a) shows an image of the bead and its corresponding intensity plot. Note that under DIC observing conditions the bead appears to contain an equal amount of bright and dark shaded regions. The position of the bead is tracked in the  $x - y$  plane and recorded as raw data for further analysis. We take advantage of the good contrast of the bright/dark shades on the bead to track its position using a cross-correlation method (Gelles *et al.* 1988). The cross-correlation between two images is calculated as

$$\mathbf{C}_0(x, y) = \sum_{i=-a/2}^{a/2} \sum_{j=-b/2}^{b/2} \mathbf{I}_1(x + i, y + j) \mathbf{I}_2(i, j),$$

which can be normalized to give

$$\mathbf{C}(x, y) = \frac{\mathbf{C}_0(x, y)}{\max(\mathbf{C}_0(x, y))},$$

where  $a$  and  $b$  are the dimensions in the  $x$  and  $y$  directions, respectively.  $\mathbf{I}_1$  and  $\mathbf{I}_2$  are the intensity of two different images.  $\mathbf{I}_2$  represents the template intensity. The effect of the above calculation is to find the position in one frame for which the surrounding intensity distribution most closely matches the template frame, whose center is taken to be the point  $(0, 0)$ . The cross-correlation calculation yields sharp peaks in regions where the intensity distribution in the frame closely matches the template intensity distribution, as shown in figure 3.1 (b). To get rid of the background matching, a threshold value  $T$  is subtracted from each point in the cross-correlation, and points where the difference is negative are discarded, as illustrated in figure 3.1 (c). From the remaining points the coordinates of the centroid  $(x_c, y_c)$  are computed as:

$$x_c = \frac{\sum_x x[\mathbf{C}(x, y) - T]}{\sum_{x,y} [\mathbf{C}(x, y) - T]},$$

$$y_c = \frac{\sum_y y[\mathbf{C}(x, y) - T]}{\sum_{x,y} [\mathbf{C}(x, y) - T]}.$$

The point  $(x_c, y_c)$  is taken as the position of the bead in the image  $\mathbf{I}_1$ . The set of calculated coordinates  $(x_c, y_c)$  reflect the translational movement of the bead with sub-pixel resolution. Depending on the signal-to-noise characteristics, this cross-correlation tracking algorithm can at times yield spatial resolutions better than a tenth of a pixel (Huang *et al.* 1997). When a sequence of consecutive video frames is analyzed, a single template derived from the first frame is used to analyze all subsequent frames. An alternative is to compare images to the preceding frame's image, however, it will lead to a small but artificial drift in the positional data due to accumulation of the tracking errors. Therefore, one single template matching scheme is selected for all subsequent analysis.

## 3.2 Drift Correction

The positions of the bead are corrupted by a slow drift due to vibrations of the experimental apparatus, as shown in a raw data plot in figure 3.2 (a). Note that the time scale of this drift is slower than the LacI-mediated dynamics of the bead, making the signal amenable to filtering. The following sections include the basic concept of digital filtering and the details of the Butherworth filter used for drift correction.

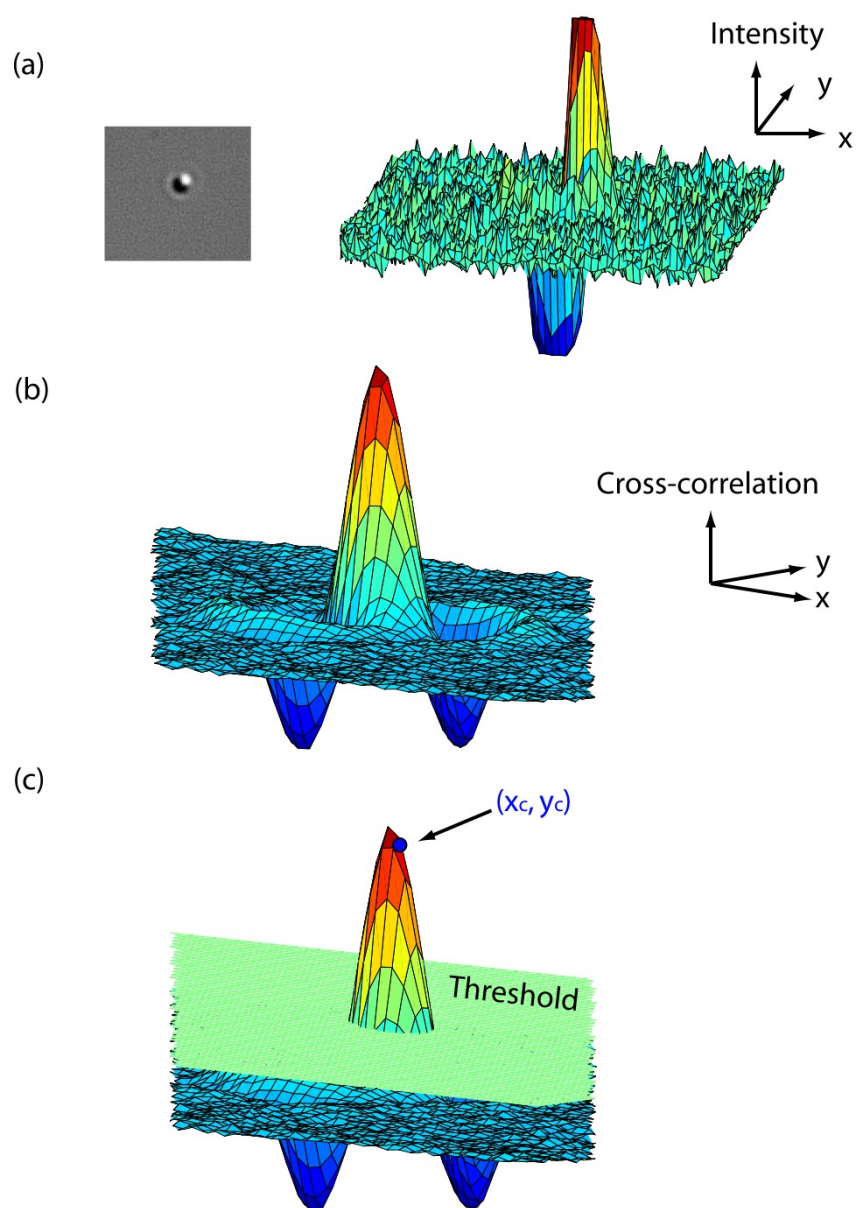


Figure 3.1: Cross-correlation algorithm used in particle tracking. (a) A typical image of the bead observed using DIC microscope and its corresponding intensity plot on the right. (b) Cross-correlation between two images with the same bead. (c) Calculation of the centroid of the peak in the cross-correlation.

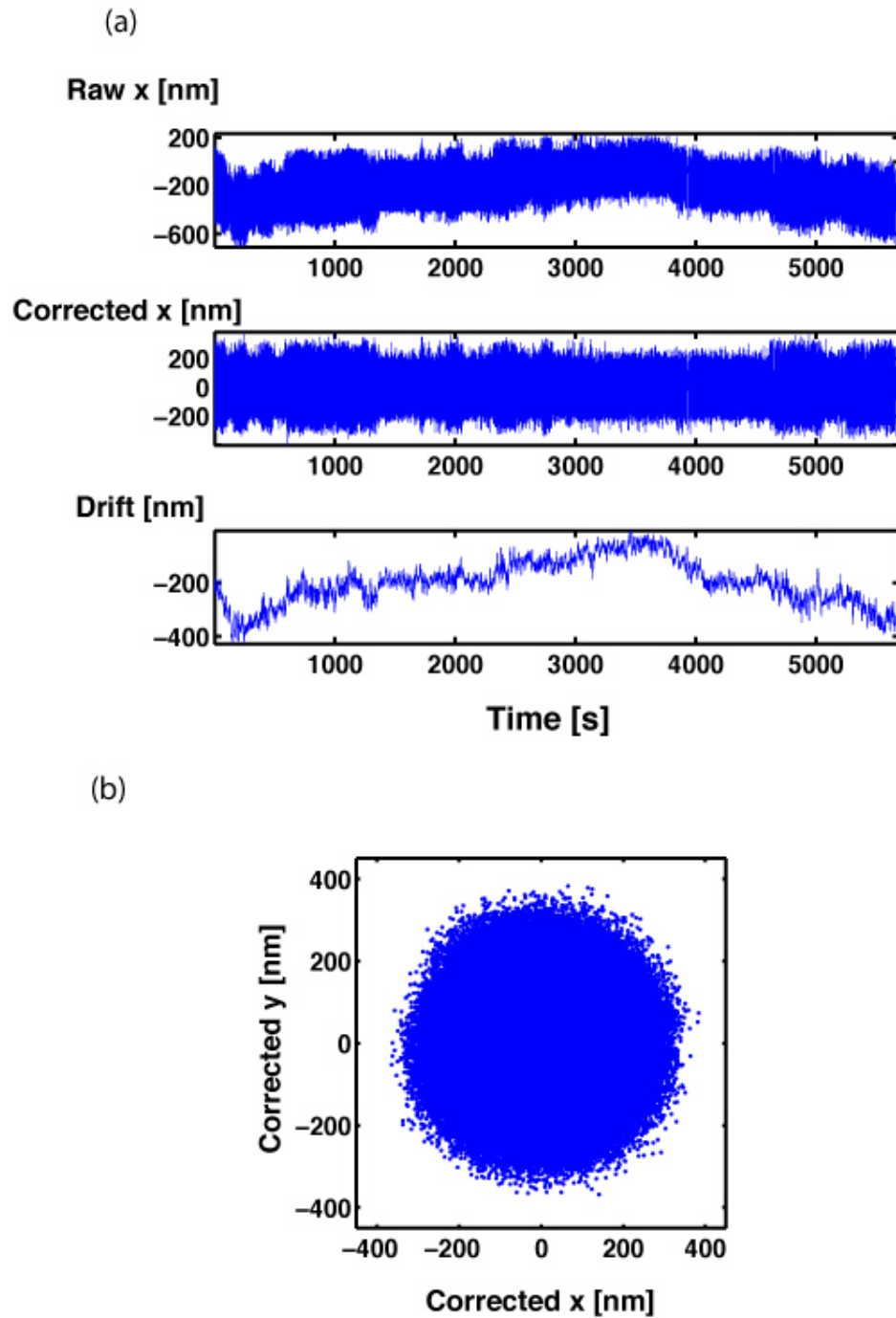


Figure 3.2: The drift in the raw position data and the data after drift correction. (a) The raw  $x$  position, the position after drift correction and the drift estimation. (b) The  $x - y$  scatter plot showing that the motion in  $x - y$  plane is radial symmetric.

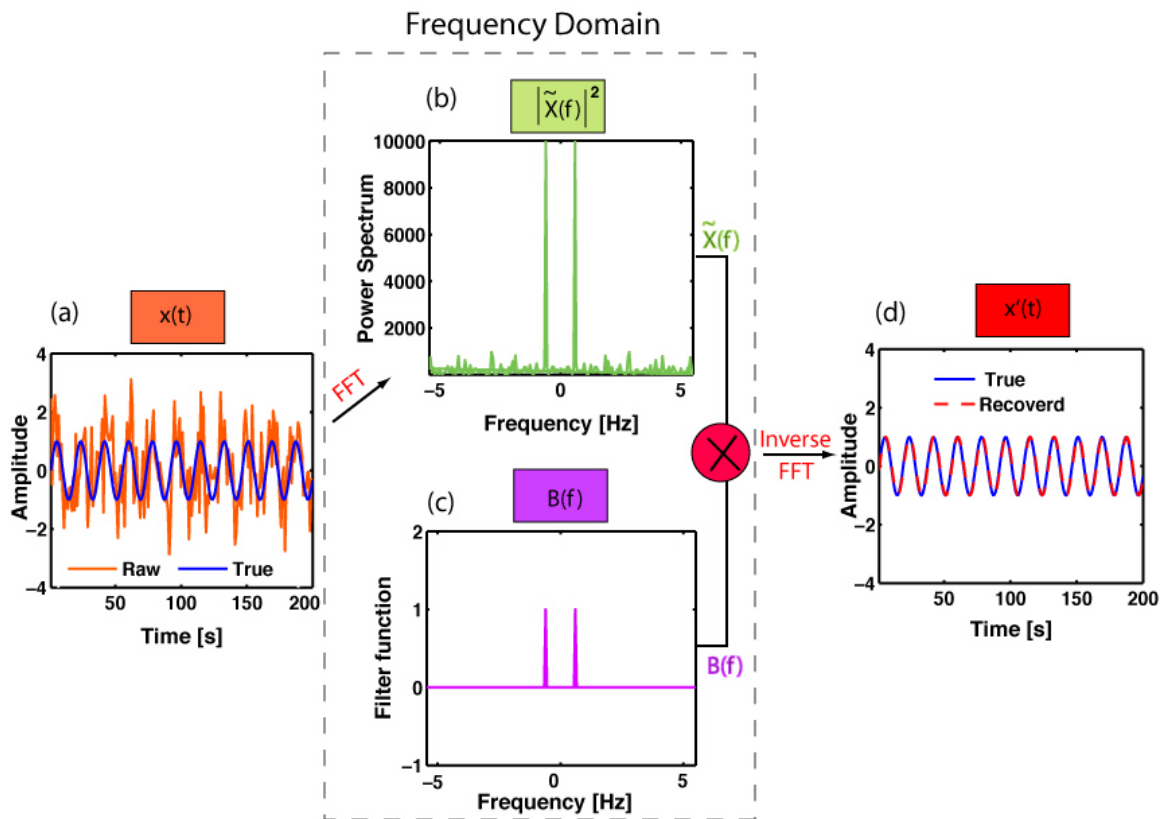


Figure 3.3: A diagram with one simple example explaining how to use a filter to recover a signal corrupted with random noise. (a) Noise corrupted signal (in orange) and the true signal (in blue). (b) A power spectrum showing different frequency components. By identifying and locating the dominant frequencies, a reasonable filter shown in (c) is selected to recover the true signal. (c) The selected band-pass filter  $B(f)$ , in which only the dominant frequencies get to pass and other frequencies due to noise are blocked. (d) Comparison between the recovered signal (in red) and the true signal (in blue).

### 3.2.0.1 Introduction to Digital Filters

A data signal normally contains a mixture of different frequency components. Signal filtering is often used to eliminate unwanted frequencies from the data signal. While the correct filter settings can significantly improve the signal to noise (S/N), incorrect settings can distort the signal presentation and even eliminate the signal completely. Therefore, it is important to understand the concept of signal filtering. Filtering can be performed either in the time domain by the operation of convolution, or in the frequency domain by multiplication of the frequency response. Frequency filtering is based on the Fourier Transform, through which frequency spectral contents of the signal can be obtained. A filter can emphasize some frequency components and attenuate some others and it can also operate on the phases of the spectral components, by delaying them by different amounts. To apply a filter, as illustrated in a simple example in figure 3.3, the original signal  $x(t)$  (in this case a noise corrupted sine signal generated by superposing sine signal and Gaussian noise together) has to be first transformed into the frequency domain  $\tilde{X}(f)$ , then multiplied with the filter function  $B(f)$  in a frequency-by-frequency fashion. The product has to be retransformed using the Inverse Fourier Transform to obtain the filtered signal  $x'(t)$  in the time domain. To recover the true signal, a band-pass filter is employed in this case with cutoff frequencies selected based on the power spectrum of the raw signal shown in figure 3.3 (b). Such filter allows the main signal to pass while blocking the other frequencies, as shown in figure 3.3 (c).

### 3.2.0.2 Butterworth Filter: Drift Correction

This section is focused on the details of how the Butterworth filter is used to correct positional data corrupted by slow drift. The Butterworth filter is a modified step function with a smooth bandpass response and more gradual out-of-band attenuation and is defined as:

$$B(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_c}\right)^{2n}}},$$

where  $f_c$  is the cutoff frequency and  $n$  is the order of the filter indicating the sharpness of the filter around the cutoff. It can be seen that as  $n$  approaches infinity, it becomes a step function and frequencies below cutoff will be fully passed, while frequencies above cutoff will be suppressed. For smaller values of  $n$ , the cutoff will be less sharp. Figure 3.4 shows the frequency response of different orders of Butterworth filter. Any motion with frequencies larger than the cutoff frequency



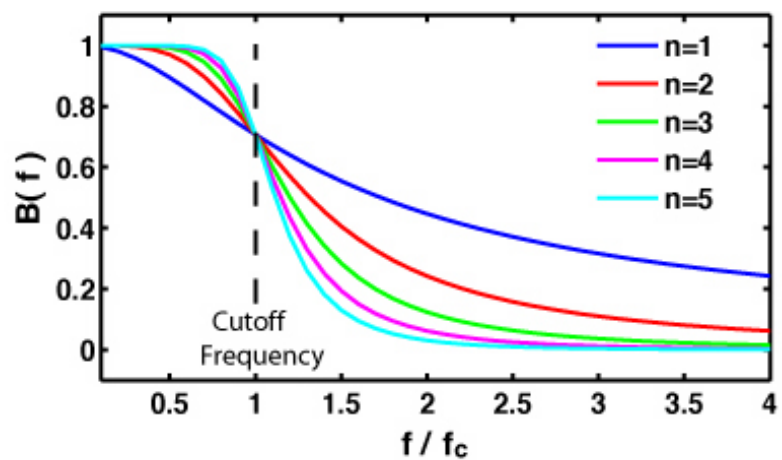


Figure 3.4: The frequency response of the Butterworth filter with different orders. The higher the order, the steeper the attenuation.

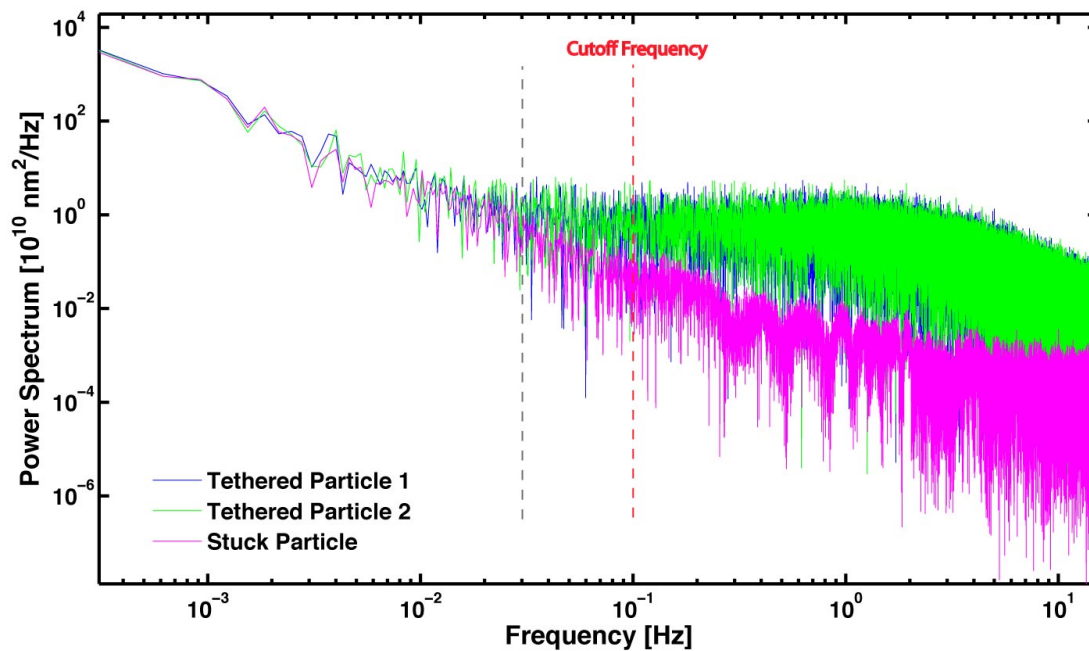


Figure 3.5: Power spectrum of both tethered and stuck bead. Three beads' power spectrum are plotted here, including two tethered beads and one stuck bead selected from the same field of view.

is gradually attenuated. In the TPM sample preparation, there are always some particles that are stuck to the surface nonspecifically without a DNA linked in between, which we refer to as stuck particles. As a results, motions of the stuck particles can be used to explore the slow drift in our experiment. To find a reasonable cutoff frequency for correcting the drift, the power spectra of both tethered particles and the stuck particle are inspected in figure 3.5. These three particles are selected from the same field of view, which suggests they should experience similar drift in their positional data. As shown in figure 3.5, by comparing the three power spectra, the frequencies below 0.02Hz are believed to be due to drift in this case. Because drift is different from sample to sample or day to day, a cutoff frequency of 0.1Hz (a little higher than 0.02) is chosen from here on to tolerate the variations. To determine the efficiency of our drift-correcting filter, we first test it on the stuck bead position data. Figure 3.6 (a) superposes the raw position data and the drift characterized using Butterworth filter together, showing indeed the drift in our TPM experiments can be estimated by such filter. Figure 3.6 (b) illustrates that the drift-corrected position data have fluctuations in the range of 30 nm, which is reasonable. Figure 3.2 shows how effective this drift correction is when we apply it to the tethered bead positions. In figure 3.2 (a) we show raw position data, drift-corrected position data and drift estimation using the first order low pass Butterworth filter at  $f_c=0.1$  Hz. Figure 3.2 (b) shows that drift corrected positions have a good radial symmetry.

So far, we accomplished the first step in the data analysis, which is obtaining the drift-free positions of tethered beads. The ultimate aim of the analysis is to obtain kinetic information for both looping and unlooping processes. Two strategies are used here to get the rate constants. One is the conventional thresholding method, which uses the variance of the bead position as the experimental variable reflecting the effective tethered DNA length. In this method, a threshold is chosen to determine individual loop formation and breakdown events, and then quantitative analyses of the durations of such events are performed and histogrammed to obtain the rates. An alternative method based on the Hidden Markov Model is developed to obtain the rate constants directly from the positional data. In the following sections, these methods are described in detail, and the disadvantages and advantages associated with each method are discussed.

### 3.3 Thresholding

The thresholding method is used to identify individual loop formation and breakdown events mediated by LacI. The microsphere exhibits a restricted Brownian diffusion due to the DNA tethers with

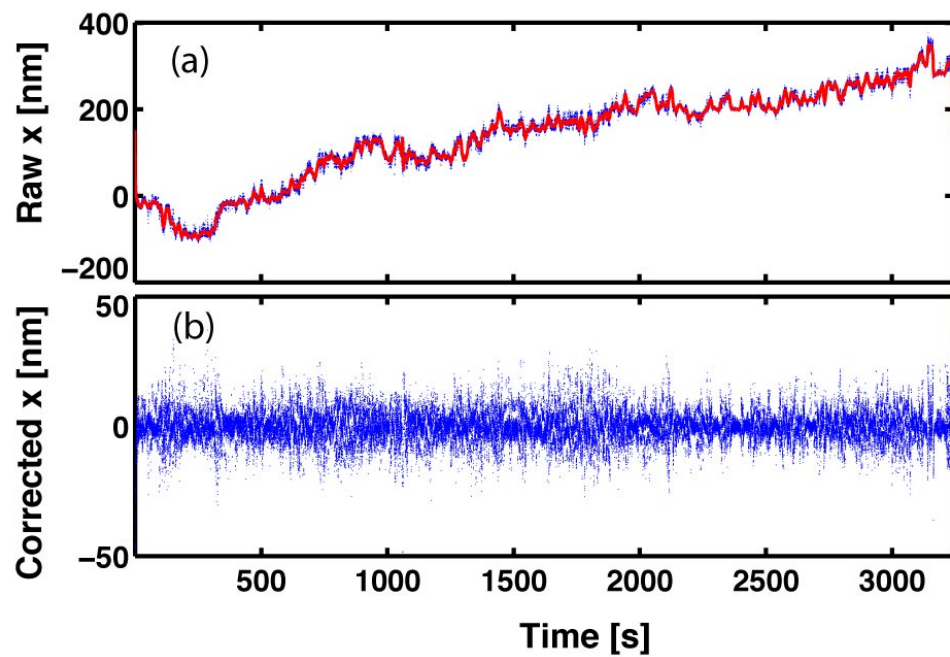


Figure 3.6: Drift correction for a struck bead using first order Butterworth filter with cutoff frequency 0.1 Hz . (a) The raw  $x$  position of the struck bead (in blue) and the drift detected using Butterworth filter (in red). (b) The drift-corrected  $x$  position due to vibration.

an amplitude reflecting the effective tethered DNA length itself. The whole procedure is illustrated in figure 3.7 (a). We first have to find an experimental output to represent the motion of the bead. The radius of mobility of the bead,  $R_t$  at time  $t$ , is calculated as:

$$R_t = \sqrt{x_t^2 + y_t^2},$$

where  $x$  and  $y$  are drift-corrected positions. Such data, as shown in blue dots in figure 3.7(a-1) and (b-1), are so noisy that it is impossible to determine the loop formation and breakdown events directly though there is a discernable pattern in the data. Therefore, the first step in the thresholding method requires smoothing these data to increase the signal-to-noise ratio to acceptable levels. A threshold in the second step is then applied to such filtered data, as shown in green in figure 3.7(a-1), to determine individual loop formation and breakdown events. Quantitative analyses of the durations of such looping and unlooping events are performed. In the final step, the kinetic information is derived from population statistics of the lifetimes.

### 3.3.0.3 Gaussian Filter: Reducing Noise

In the thresholding method, the signal of transition events is buried in the thermal fluctuations, as shown in blue dots in figure 3.7. To reduce the background noise sufficiently to allow looping events to be detected and characterized, some filtering on the data is required. From the radius of mobility of the bead, the RMS of the mobility of the bead in a given interval of time is obtained by applying a Gaussian filter to  $\sqrt{\langle R_t^2 \rangle}$ . A Gaussian filter is a kind of moving average filter and it can be used to reduce random noise while retaining a sharp step response and increasing the signal-to-noise ratio (SNR). This step requires a choice of cutoff frequency. Basically, we want to find a good frequency cutoff to reduce statistical fluctuations enough in the running average, so that when a threshold is applied, we rarely mistake a fluctuation for a looping event. Different cutoff frequencies result in different time resolution and signal-to-noise ratio. These two factors require a compromise. Filtering with a low frequency cutoff makes higher signal-to-noise ratio and increases the confidence in the discrimination between different states, but decreases the time resolution. On the other hand, filtering with a high frequency cutoff maintains higher time resolution but makes lower signal-to-noise ratio, resulting in increasing the number of false events. Understanding the concept of the Gaussian filter will provide us useful insights into how to choose the cutoff. The

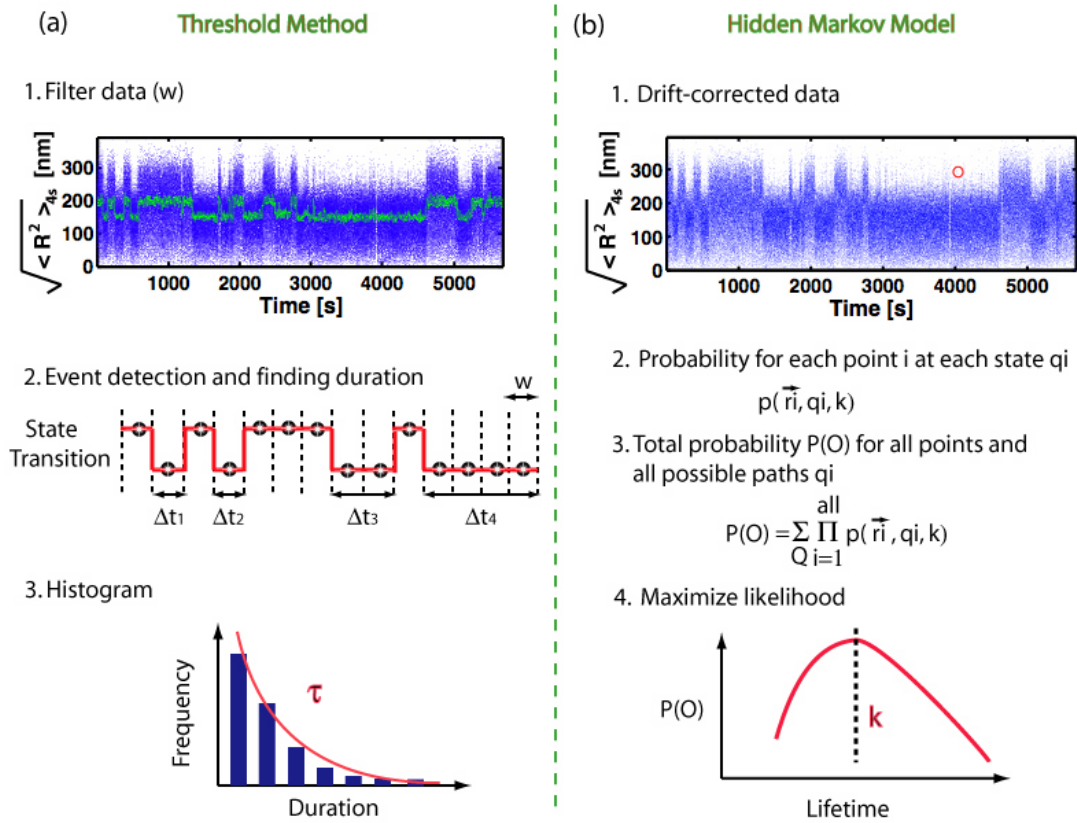


Figure 3.7: Two methods of TPM data analysis to obtain the kinetic information. (a) Basic concept of thresholding method. 1 Windowing data to reduce noise. 2 Setting up a threshold through which transition events are identified and durations of such events are quantified. 3 Histogram of all the durations and fit it to obtain the lifetime  $\tau$ . (b) Basic concept of Diffusive-HMM method. 1 Choose an observable variable suitable for HMM application. 2 Set up probability distribution of the observable variable  $\vec{r}_i$  at time  $i$  for each hidden state  $q_i$  with unknown parameter: rate constants  $k$ s. 3 Calculate the total probability of all possible trajectories. 4 Maximize the likelihood and extract the unknown parameters: rate constants.

Gaussian filter has a frequency response function defined as (Colquhoun and Sigworth 1995):

$$G(f) = \exp\left(-0.3466 \frac{f^2}{f_c^2}\right), \quad (3.1)$$

where the constant 0.3466 is chosen to give 3dB of attenuation at the cutoff frequency  $f_c$ . The impulse response of the Gaussian filter in the time domain can be obtained through Inverse Fourier transform of equation 3.1 and can be written in the same form as a Gaussian probability distribution:

$$g(t) = \frac{1}{(2\pi)^{1/2}\sigma_g} \exp\left(-\frac{t^2}{2\sigma_g^2}\right) \quad (3.2)$$

where the standard deviation of the impulse is characterized by  $\sigma_g$ . It corresponds to the window size of the filter with value inversely proportional to  $f_c$ :

$$\sigma_g = \frac{(\ln 2)^{1/2}}{2\pi f_c}. \quad (3.3)$$

To see how this window size relates to the moving average, we have to understand how this filter works in the time domain. As mentioned before, applying a filter to a signal  $s$  can also be obtained by a convolution operation in the time domain and it can be expressed mathematically as:

$$y(i) = s * g = \sum_j g(j) \cdot s(i - j),$$

where  $*$  denotes the convolution operation and  $y(i)$  is the final convolution product. From the above expression and considering  $g(j)$  being a Gaussian distribution, the number of points contributing to the final convolution product for each point  $i$  is about  $\sigma_g$ , since Gaussian distribution is around zero at other points. In another word, for any  $y(i)$ , a Gaussian filter takes about a window size  $\sigma_g$  of points with the  $i$ th point in the middle and "blend" them all together according to the filter function and generates an output at  $i$ . The bigger this window size, the more averaging effect, which results in a better signal-to-noise ratio, but poorer time resolution.

Figure 3.8 illustrates the effect of different window size resulting in different signal-to-noise ratio. The pink and green represent data filtered with cutoff frequencies 0.0326 and 0.1304 Hz, which correspond to window size of 1 and 4 seconds, respectively. When the window size is small, filtering is limited resulting in residual noise which is indistinguishable from the short looping events. When the window size is big enough, the filtered data shows a clear discrimination between different

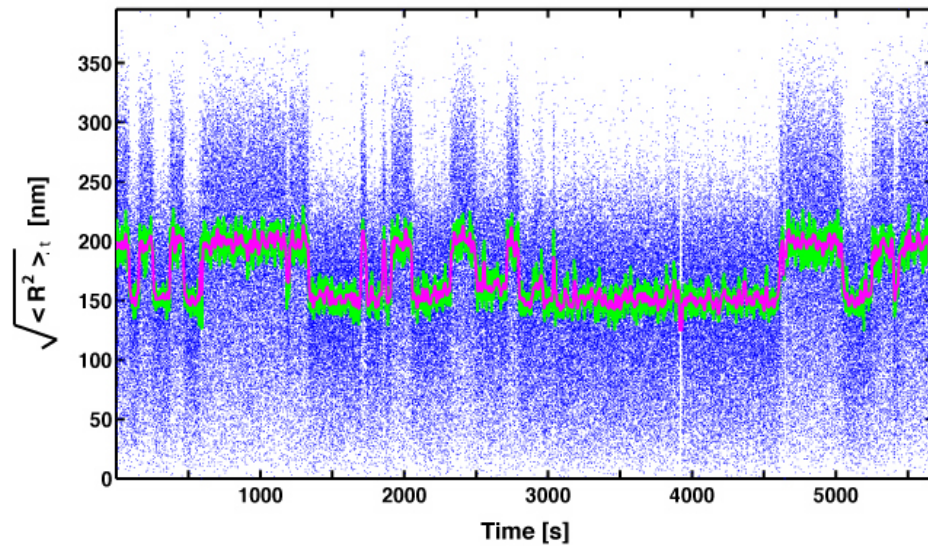


Figure 3.8: Raw data and filtered data. Typical data set filtered at 0.0326 Hz (pink) and 0.1304 Hz (green) resulting in different signal-to-noise ratios. The blue dots represent the unfiltered data before.

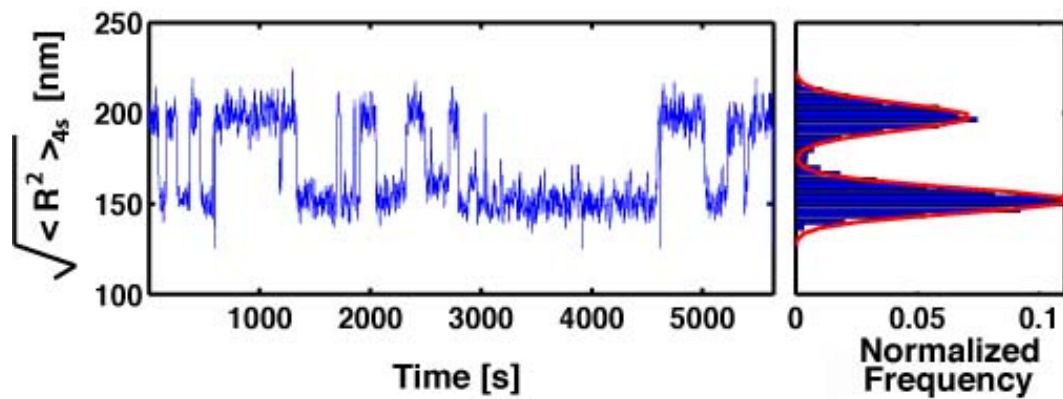


Figure 3.9: Typical two state data set filtered at 0.0326 Hz showing Brownian motion as function of time. The red curve is a double Gaussian fit to the two peaks and the midpoint between them is taken as the dividing line between the two states. The two-state's system was observed in the presence of both LacI and IPTG when the bead was tethered with 901 bp long DNA (Later on, no such data can be repeated).

states. The cutoff frequency of 0.0326 Hz is finally chosen to obtain not only a good signal-to-noise ratio but also a reasonable time resolution. A typical data set is shown in figure 3.9. The next step is to set a threshold and to determine when transitions between different states take place and to obtain other quantitative parameters, such as rate constants.

### 3.3.0.4 The Threshold Detector

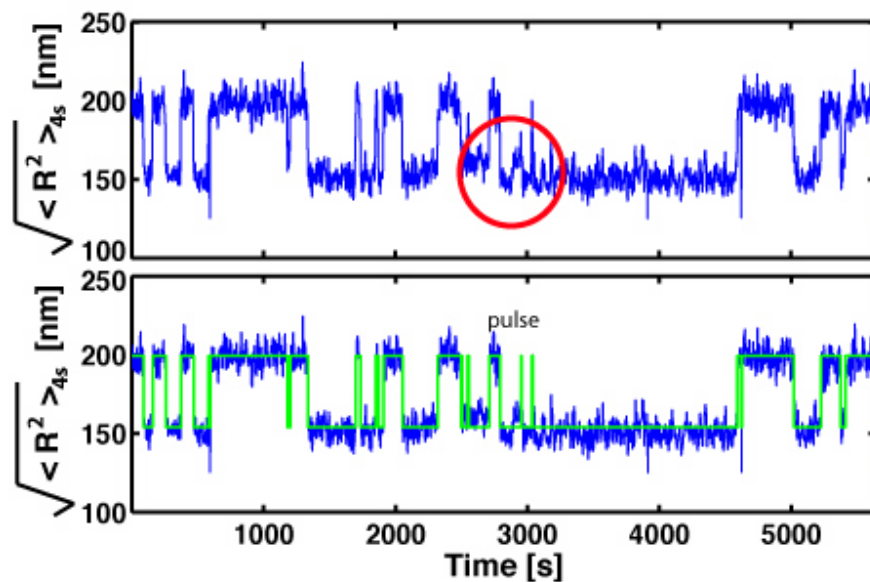


Figure 3.10: Transitions detected using thresholding methods. The green lines represents the states the system is in. (a) The Gaussian filtered bead's motion. (b) Transitions detected by single-threshold method.

The threshold is a boundary between different states and is used to differentiate between different states. Every crossing of the threshold is interpreted as a transition between different states. This method is the most widely used method for TPM data analysis (Finzi and Gelles 1995; Vanzi *et al.* 2006). The threshold needs to be set high enough to avoid including a significant number of spurious peaks caused by noise but low enough to capture as many true events as possible. In order to find this critical threshold, a double Gaussian was fitted to the  $\sqrt{\langle R_t^2 \rangle}$  distribution shown in the right plot of Figure 3.9 and the middle point between the two Gaussian peaks is then taken as the threshold. Figure 3.10 shows the transitions revealed by the thresholding method. Transitions



detected in the region enclosed by the red circle remained uncertain, and it is hard to distinguish short pulses from random noise fluctuations. That raises another question: What is the detection limit for a given cutoff frequency and threshold?

### 3.3.0.5 Effect of Window Size

During the convolution process, points within the size  $\sigma_g$  will be averaged according to the the filter function. In our case, it is a Gaussian distribution. Short pulses will be smeared due to this averaging and will become indistinguishable from fluctuations. To find the shortest pulse after filtering can be detected, we have to understand how a Gaussian filter responds to pulses of different width. First, let us consider the response for a step function  $f(t)$  :

$$f(t) = \begin{cases} 1 & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

The response is given as a convolution product  $F(t)$  of the signal  $f(t)$  and the Gaussian filter

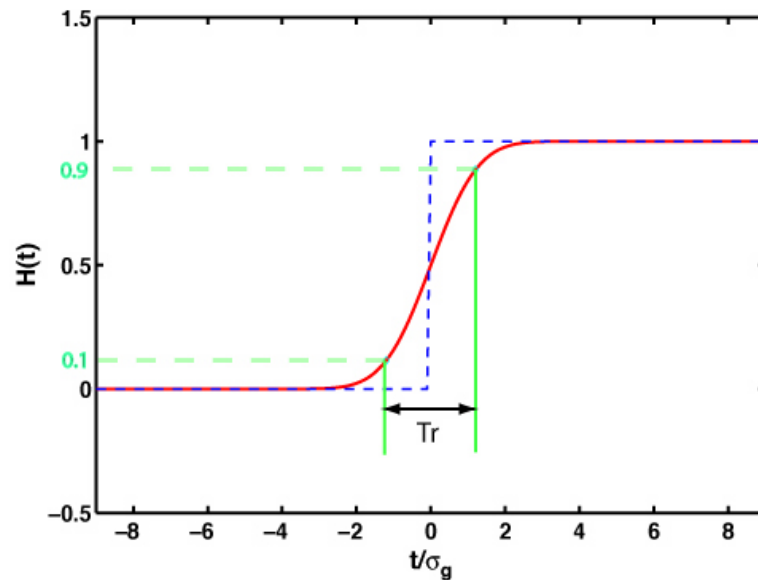


Figure 3.11: The step response of the Gaussian filter. The original step function is plotted using the blue dashed curve and the filtered signal is given by the red curve.

response  $g(t)$ , and based on the definition of convolution, we have:

$$F(t) = f(t) * g(t) = \int_{-\infty}^{\infty} f(s)g(t-s)ds.$$

We then split the integral into two parts:

$$F(t) = \int_{-\infty}^t g(s)f(t-s)ds + \int_t^{\infty} g(s)f(t-s)ds.$$

When  $s$  is greater than  $t$ , we have  $f(t-s) = 0$ . Only one part is left to contribute to the final convolution product:

$$F(t) = \int_{-\infty}^t g(s)f(t-s)ds.$$

Now plug in the expression for Gaussian response  $g(t)$  in equation 3.2 and  $f(t-s) = 1$  when  $s < t$ , and the result of this analysis is:

$$\begin{aligned} F(t) &= \int_{-\infty}^t 1 \cdot \frac{1}{(2\pi)^{1/2}\sigma_g} \exp\left(-\frac{s^2}{2\sigma_g^2}\right) ds \\ &= \frac{1}{(2\pi)^{1/2}\sigma_g} \int_{-\infty}^0 \exp\left(-\frac{s^2}{2\sigma_g^2}\right) ds + \frac{1}{(2\pi)^{1/2}\sigma_g} \int_0^t \exp\left(-\frac{s^2}{2\sigma_g^2}\right) ds. \end{aligned}$$

The first term is half of the Gaussian integral and the second one can be written in terms of an error function:

$$F(t) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{t}{\sqrt{2}\sigma_g}\right).$$

This is our final convolution product of the step response. This step response, important in time domain applications, shows how quickly the filter responds to an abrupt change in the input signal, which is illustrated as the red curve in figure 3.11. To represent the amount of averaging, a parameter  $T_r$  is defined to correspond to the minimum length of a pulse to which the filter gives a nearly full-amplitude response. One common definition for a risetime is the time between the 10 % and 90 % amplitude points of the transition in the output of the filter, as indicated in figure 3.11. To find out if a pulse after filtering can be detected, we have to know what the peak value of the output signal is and then compare it to the threshold. Now consider a rectangular pulse  $b(t)$  of length  $w$  and unit

amplitude:

$$b(t) = \begin{cases} 1 & \text{for } 0 \leq t \leq w, \\ 0 & \text{otherwise.} \end{cases}$$

This rectangular signal  $b(t)$  can be written as a linear combination of the step function  $f(t)$ :

$$b(t) = f(t) - f(t - w).$$

The distributivity property of convolution gives the pulse response  $B(t)$ :

$$B(t) = F(t) - F(t - w)$$

Substituting the step response  $F(t)$  into the equation above, we get:

$$B(t) = \frac{1}{2} \operatorname{erf} \left( \frac{t}{\sqrt{2}\sigma_g} \right) - \frac{1}{2} \operatorname{erf} \left( \frac{t - w}{\sqrt{2}\sigma_g} \right).$$

Next, we need to find the maximum of the pulse response  $B(t)$ , which tells us the magnitude of the peak of Gaussian filtered rectangular pulse. It is worthy noting that the function  $B(t)$  is symmetric about  $t = \frac{w}{2}$ , which is where the peak value should be located. Then the maximum of  $B(t)$  can be simply obtained by using the value of  $B(\frac{w}{2})$ :

$$\begin{aligned} B_{\max} &= B\left(\frac{w}{2}\right) \\ &= \operatorname{erf} \left( \frac{w}{2\sqrt{2}\sigma_g} \right). \end{aligned} \quad (3.4)$$

If  $\sigma_g$  is plugged in using equation 3.3, we have the relationship between peak value and cutoff frequency:

$$B_{\max} = \operatorname{erf}(2.668f_c w). \quad (3.5)$$

The risetime is then calculated as:

$$T_r = T_{10-90} = 2^{3/2}\sigma_g \operatorname{erf}^{-1}(0.8) = 0.3396/f_c. \quad (3.6)$$

The expression of  $T_r$  shows how fast the filter responds to a step function for a given cutoff frequency. Figure 3.12 demonstrates the relationship between filtered output and the input noise free pulse width with the unit of  $T_r$ . The short pulses narrower than the dead time  $T_d$  are missed altogether, because they never reach the threshold after filtering. We can calculate the exact value of  $T_d$  by finding the pulse width that gives a half-amplitude response. According to equation 3.4, we have

$$\operatorname{erf}\left(\frac{T_d}{2^{3/2}\sigma_g}\right) = 0.5. \quad (3.7)$$

If we solve equation 3.7, we find the dead time:

$$T_d = 2^{3/2}\operatorname{erf}^{-1}(0.5)\sigma_g = 1.349\sigma_g. \quad (3.8)$$

Using equation 3.6 and Equation 3.3, we have the relationship between  $T_d$  and  $T_r$  or  $f_c$ :

$$T_d = 0.538 T_r = 0.179/f_c.$$

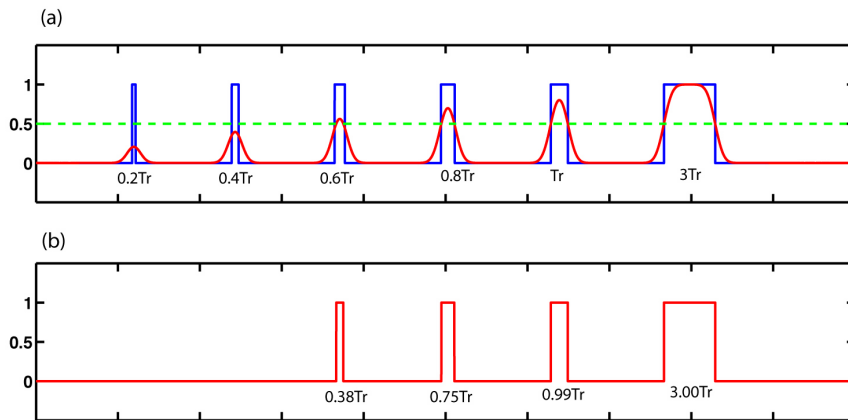


Figure 3.12: Pulses with different width and their corresponding filtered output. (a) Original pulses (in blue), the output of the filter (in red) and the half-amplitude threshold (in green). (b) Transitions detected with duration measured.

Equation 3.8 illustrates that a wide window has a longer dead time, which would smear out some genuine short-lived events and mistakenly lump them together and report them as false long lived events. To see how the choice of window size affects our system, a few cutoff frequencies: 0.1304, 0.0652, 0.04346, 0.0326 and 0.0163 Hz are used, corresponding to the window size of 1,

1.3, 2, 4 and 8 seconds. Figure 3.13 shows when the window size is small, the signal is noisier and it is more likely to mistake the fluctuation as a transition resulting in lots of short-lived events. When the window size is very big, the number of short-lived events is small due to the substantial dead time of the filter. In conclusion, the thresholding methods described here require windowing the data to smear out the fluctuations in the system, however, in the meantime, not only is the time resolution limited, but also some true events get mistakenly destroyed or ignored. Therefore, the rate constants obtained by histogramming the time distribution are seriously biased by the choice of the filter size. Some further correction for missed and false events could be done based on the properties of the Gaussian filter (Vanzi *et al.* 2006). However, it requires that we understand the complete kinetics of the system very well. Unlike the simple looping and unlooping condition shown in (Vanzi *et al.* 2006), our system is much more complex. It does not just have one looped state, it contains two looped states. A simplified kinetic scheme is shown in figure 3.14. To simplify the whole process, one of the two states in which only one LacI binds to one of the two binding sites shown in (2) is considered and the other state shown in the shadow is ignored. To understand this complex system is hard and to correct for missed and false events for this system is even more challenging. Therefore, the thresholding method is not very reliable for analyzing our data to obtain the unbiased rate constants.

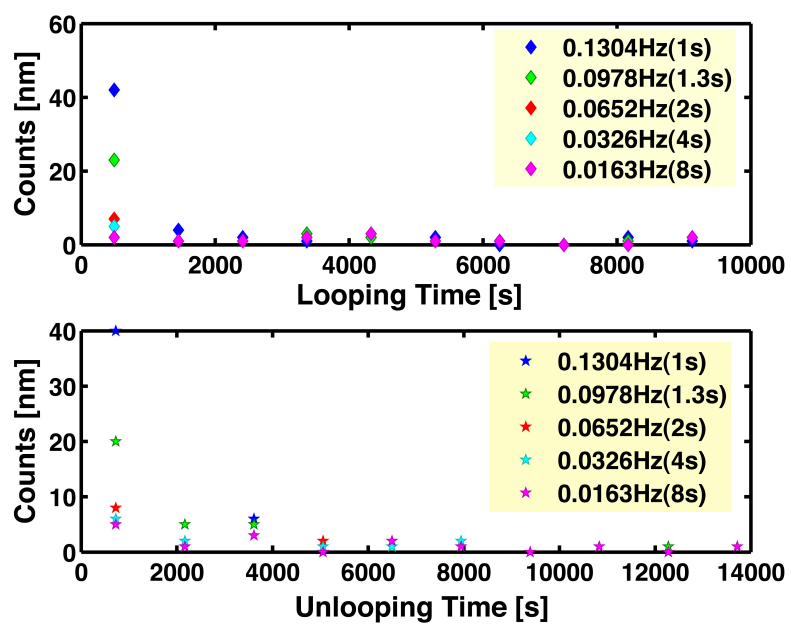


Figure 3.13: Looped and unlooped time distribution from data filtered with different cutoff frequencies.

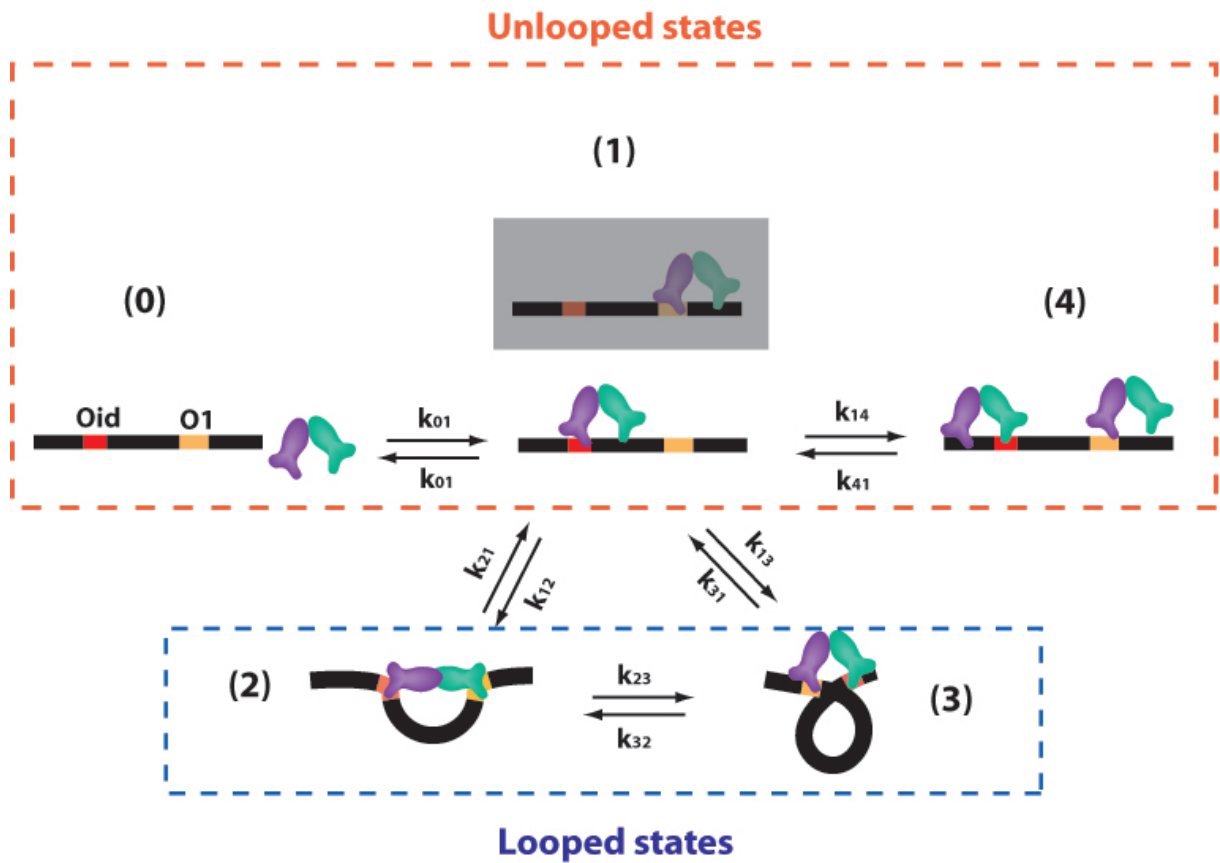


Figure 3.14: Simplified kinetic scheme in the three-state-system. (0): The state in which No LacI binds. (1): The state where LacI only binds to one of the two binding sites. The state in the shadow is ignored for simplification. (2): Open-looped configuration where LacI is extended when binds to the two binding sites at the same time. (3): Closed-looped configuration where LacI is in a V-shape when binding to the two binding sites at the same time. (4): Two LacI present resulting in occupancy of both binding sites.

### 3.4 Diffusion-Hidden Markov Model Method

The Hidden Markov Model (HMM) is widely used in analyzing patch-clamp measurements of ion channels. In a typical patch-clamp experiment, the opening and closing of the channel is directly monitored by measuring the alternation between a higher and lower current signal flowing through the patch. This gating behavior of ion channels can be modeled as a Markov chain, in which each state of the model represents a physical conformation of the channel. Similar to ion channel measurements, a typical TPM experiment gives the position of the tethered bead in time, as shown in figure 3.2, which suggests the range of the diffusion alternates due to the conformational change of the tether. We are therefore tempted to apply the HMM method to our data to obtain quantitative information such as looping and unlooping rates. Compared to conventional thresholding method, the HMM method does not require data to be windowed, which is an advantage in terms of temporal resolution.

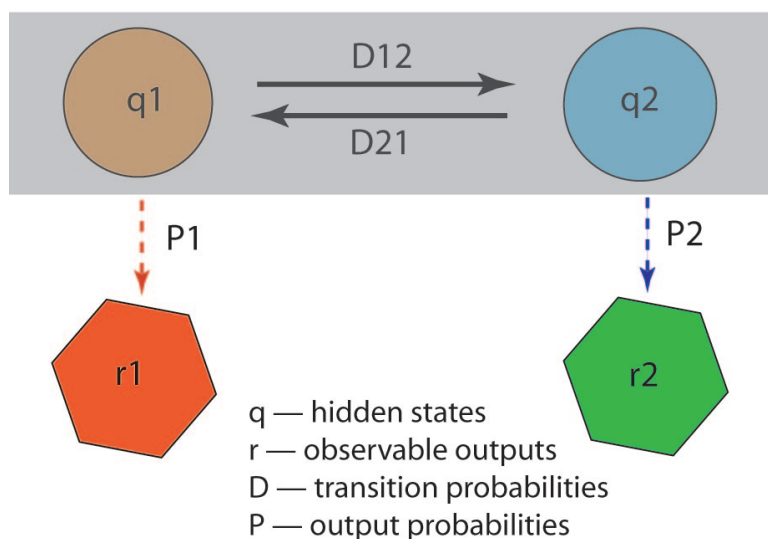


Figure 3.15: State transitions in a hidden Markov model. Circles represent the hidden states  $q$  in the system, transitions among which are governed by a set of probabilities called transition probabilities. In a particular state an observable output, presented as hexagons, can be generated, according to the associated probability distribution  $P$ . It is only the output, not the state visible to an external observer and therefore states are "hidden" to the outside.



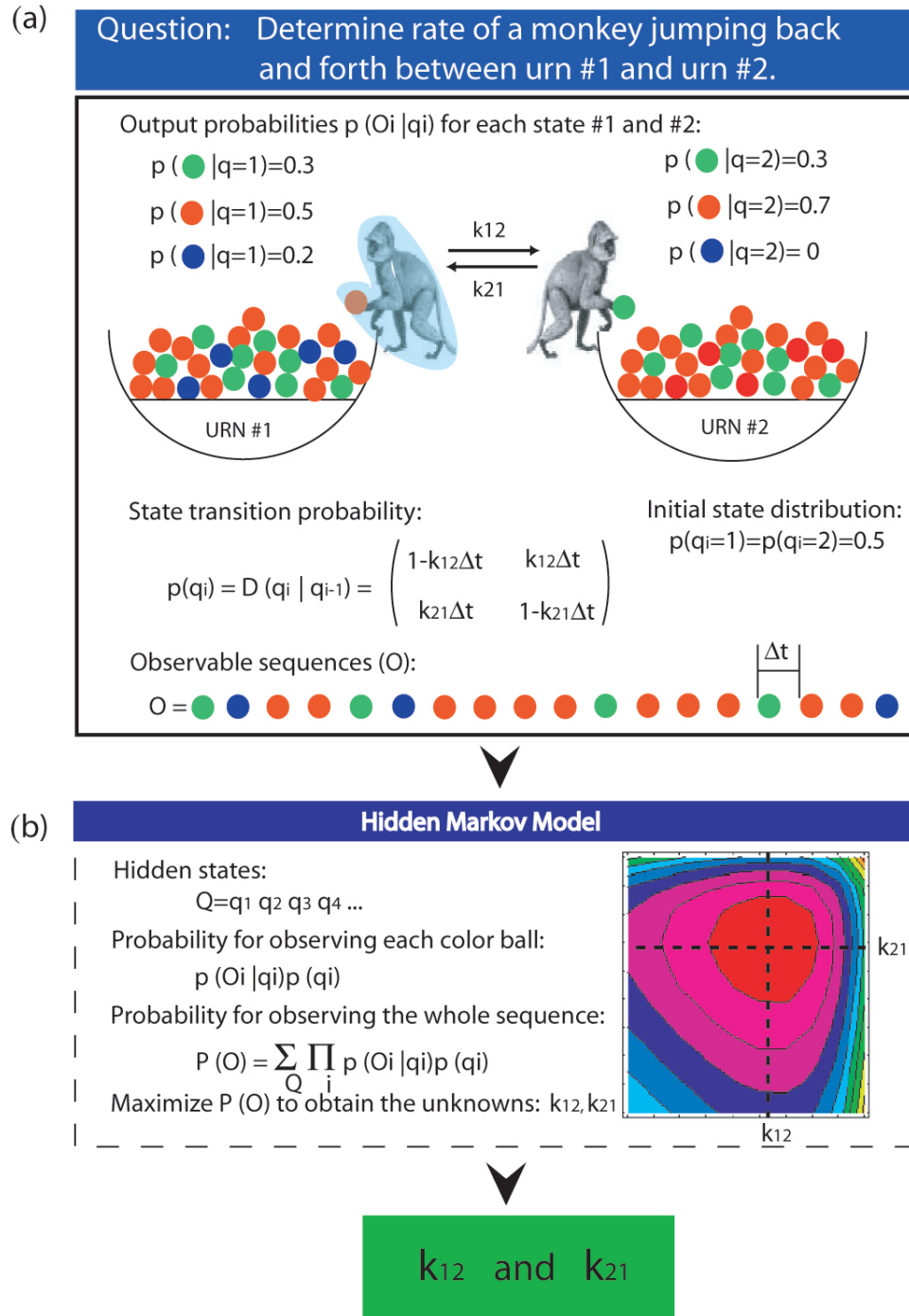


Figure 3.16: A 2-state urn and ball model which illustrates the general case of a discrete symbol HMM. Two steps are involved in finding the best Markov model that describes the behavior of the system. (a) Specification of the general topology of the model including five basic elements: 1) Number of hidden states. 2) The number of distinct observation symbols per state. 3) Output probability distribution. 4) State transition probability. 5) The initial state distribution. (b) Optimization of Markov model parameters. Knowing the probability distribution of each observable output in each hidden state and the transition probability containing the transition rates we are interested in, the total probability of the observation can be calculated and then optimized to get the hidden parameters of the model using maximum-likelihood techniques.

### 3.4.1 Basic Concept of Hidden Markov Model

In the hidden Markov model, the state is not directly visible, but variables influenced by the state are visible. Each state has a probability distribution over the possible observable outputs. Therefore the sequence of outputs generated by an HMM gives some information about the sequence of states. More precisely, standard HMM supposes that an observed signal reflects two processes (Rabiner 1989): a hidden process that generates a time series using variable  $q_t$  according to a Markov process with some transition probability distribution  $D(q_{t+1}|q_t)$ , and an observed signal  $r_t$  at each instant  $t$ , which is drawn from a probability distribution  $P(r_t|q_t)$  as shown in figure 3.15. In a Markov process, the value of the hidden variable  $q_{t+1}$  (at time  $t+1$ ) only depends on the value of the hidden variable  $q_t$  (at time  $t$ ). Similarly, the value of the observed variable  $r_t$  only depends on the value of the hidden variable  $q_t$  (both at time  $t$ ). Two steps are involved in finding the best Markov model that describes the behavior of the system, which is illustrated in figure 3.16. First, the general topology of the model must be specified with the number of states and the connectivity allowing transitions among states. In the case of urn-and-ball model in figure 3.16 (a), two hidden states are urn No. 1 and urn No. 2 and the output variables are colored balls. The state transition probability is governed by the matrix of transition rates, which are what we are interested in. The second step is the optimization of Markov model parameters, as shown in figure 3.16 (b). Once we know the probability distribution of each observable output in each hidden state and the transition matrix among different states, the total probability of the observation can be calculated over all possible trajectories and then optimized to get the hidden parameters of the model using maximum-likelihood techniques. In the ion channel analysis, the hidden states could be the open or closed channel with different current amplitude output as the observable. Channel opening and closing are random and the current amplitude flowing through the channel only depends on which state the channel is in. These two properties of the ion channel reach the standard qualifications of HMM application. We might be tempted to apply it to our system. Let  $q_t$  denote the state that our DNA tether is in and output observable  $r_t = (x_t, y_t)$  to be the position of the bead. However, our TPM output is not a pure Markov process as explained next. The ability to form a loop depends on where the bead is: for example, if the bead is too far from the tethering point, then loop formation is forbidden until the bead has diffused closer and bring the two binding sites closer to each other. This is against the assumption made in standard HMM. Moreover, the bead location  $r_{t+1}$  at time  $t$  depends not only on the state  $q_t$  at time  $t$ , but also on the bead location  $r_t$  at time  $t$  under the acquisition condition here, as explained

in detail in the following section. Both of these reasons prevent applying HMM directly to our TPM data, modification of the usual formulation is required.

### 3.4.1.1 Diffusion of a Tethered Particle

As mentioned above, the standard HMM requires that the output variable only depends on the hidden state, not on anything else. When we try to apply HMM to our TPM data with the bead's position as observable output, we need to find out if it satisfies this criterion. The goal of this section is to verify the second point made above that the bead's location is coupled from one frame to the next through the autocorrelation function of the bead position. The tethered particle can be thought of as a particle in a harmonic well with spring constant  $k_{\text{eff}}$ . The motion of a tethered particle subject to Brownian motion as well as viscous and the spring force due to the tethered DNA is governed by a momentum balance. In one dimension, this can be expressed as:

$$m \frac{d^2 x}{dt^2} = \gamma \frac{dx}{dt} + k_{\text{eff}} x - f(t), \quad (3.9)$$

where  $x(t)$  is the position of sphere center,  $\gamma$  is the drag coefficient and  $m$  is the mass of the tethered bead. In our case,  $m = 6.2 \times 10^{-23}$ g, which is so small that we can ignore the acceleration term. The simplified equation is:

$$\gamma \frac{dx}{dt} + k_{\text{eff}} x = f(t) \quad (3.10)$$

The condition

$$x(0) = 0. \quad (3.11)$$

fixes the starting point for the trajectory. And the condition

$$\frac{dx}{dt} = 0 \quad \text{at} \quad t = -\infty$$

assures sufficient time to reach thermodynamic equilibrium by  $t=0$  and therefore eliminates any effect of the initial velocity.

To solve equation 3.10, multiply it on both sides by  $e^{\frac{k_{\text{eff}}}{\gamma} t}$ , and organize in the way below:

$$\frac{d}{dt} \left( x \cdot e^{\frac{k_{\text{eff}}}{\gamma} t} \right) = \frac{1}{\gamma} e^{\frac{k_{\text{eff}}}{\gamma} t} f(t).$$

Then integrate from  $-\infty$  to  $t$  to obtain the position of the bead at time  $t$  given by

$$x(t) = \frac{1}{\gamma} \int_{-\infty}^t e^{-\frac{k_{\text{eff}}}{\gamma}(t-s)} f(s) ds. \quad (3.12)$$

Two assumptions are used here to characterize the Brownian force. First, these forces are taken to be random both in direction and magnitude, and they are uncorrelated on the time scale of particle motion. These conditions can be expressed mathematically through the ensemble average: for random in directions and magnitude, we have

$$\langle f(t) \rangle = 0.$$

The fact that Brownian force is uncorrected from time  $t$  to time  $t + 1$  can be expressed mathematically by

$$\langle f(t)f(t + \tau) \rangle = F\delta(\tau), \quad (3.13)$$

where  $F$  can be determined from the second assumption: at equilibrium, the kinetic energy is partitioned equally among the three translational modes of the particle. From energy conservation, we have:

$$\frac{1}{2}k_{\text{eff}}x^2 = \frac{1}{2}k_{\text{B}}T. \quad (3.14)$$

Using the expression of  $x(t)$  in equation 3.12, the position autocorrelation function can be expressed as:

$$\begin{aligned} R(\tau) &= \langle x(t)x(t + \tau) \rangle \\ &= \left\langle \frac{1}{\gamma} \int_{-\infty}^t e^{-\frac{k_{\text{eff}}}{\gamma}(t-s)} f(s) ds \cdot \frac{1}{\gamma} \int_{-\infty}^{t+\tau} e^{-\frac{k_{\text{eff}}}{\gamma}(t+\tau-q)} f(q) dq \right\rangle \\ &= \frac{1}{\gamma^2} e^{-\frac{k_{\text{eff}}}{\gamma}\tau} \left\langle \int_{-\infty}^t e^{-\frac{k_{\text{eff}}}{\gamma}(t-s)} f(s) ds \cdot \int_{-\infty}^{t+\tau} e^{-\frac{k_{\text{eff}}}{\gamma}(t-q)} f(q) dq \right\rangle \\ &= \frac{1}{\gamma^2} e^{-\frac{k_{\text{eff}}}{\gamma}\tau} \left\langle \int_{-\infty}^t e^{-\frac{k_{\text{eff}}}{\gamma}(t-s)} f(s) ds \cdot \left[ \int_{-\infty}^t e^{-\frac{k_{\text{eff}}}{\gamma}(t-q)} f(q) dq + \int_0^{\tau} e^{-\frac{k_{\text{eff}}}{\gamma}(t-q)} f(q) dq \right] \right\rangle. \end{aligned}$$

According to equation 3.13, the above equation can be simplified as:

$$\begin{aligned} R(\tau) &= \frac{1}{\gamma^2} e^{-\frac{k_{\text{eff}}}{\gamma}\tau} \int_{-\infty}^t e^{-\frac{k_{\text{eff}}}{\gamma}2(t-s)} f(s) ds, \\ R(\tau) &= \frac{F}{2k_{\text{eff}}\gamma} e^{-\frac{k_{\text{eff}}}{\gamma}\tau}. \end{aligned} \quad (3.15)$$

From equation 3.14 and equation 3.15, we can obtain  $F = 2k_{\text{B}}T\gamma$ . The autocorrelation can be written as:

$$R(\tau) = \frac{k_{\text{B}}T}{k_{\text{eff}}} e^{-\frac{k_{\text{eff}}}{\gamma}\tau}.$$

The logarithm of the autocorrelation of positions from experimental data is shown in figure 3.17. The spring constant  $k_{\text{eff}}$  and the drag coefficient  $\gamma$  can be extracted from the linear fit of the experimental data. We find  $k_{\text{eff}} = 2.8 \times 10^{-4}$  pN/nm, diffusion constant  $D = 3.0 \times 10^5$  nm<sup>2</sup>/s and the characteristic decay time constant  $\tau = 50$  ms. The diffusion coefficient for the bead used here, exhibiting free Brownian motion (i.e., no tethering effects) can be calculated as:

$$D_0 = \frac{k_{\text{B}}T}{6\pi\eta a} = 8.9 \times 10^5 \text{ nm}^2/\text{s},$$

where  $a$  is the radius of the bead and  $\eta$  is the fluid viscosity. This value is more than a factor of four larger than the one obtained from experimental data. This discrepancy may be due to the fact that the bead is very close to the surface when it is tethered by DNA. If we consider this effect, according to Dufresne *et al.* (2000), the first-order-corrected diffusion coefficient near the wall  $D_{\text{w}}$  can be written as:

$$D_{\text{w}} = \left(1 - \frac{9}{16} \frac{a}{h}\right) D_0 = 4.9 \times 10^5 \text{ nm}^2/\text{s}.$$

In the equation above, we use the DNA contour length as the distance  $h$  from particle to the wall, which underestimates the wall effect and results in a larger  $D_{\text{w}}$ . By considering this wall effect, we bring closer the diffusion constant estimated from the simple calculation above and the one extracted from the experimental data.

The characteristic decay time  $\tau$  obtained from fitting the experimental data supports the observation that the Brownian motion of the tethered bead has an intrinsic time scale of 50 ms. This time scale is slow compared with the sampling frequency (30 Hz), which prevents the direct implementation

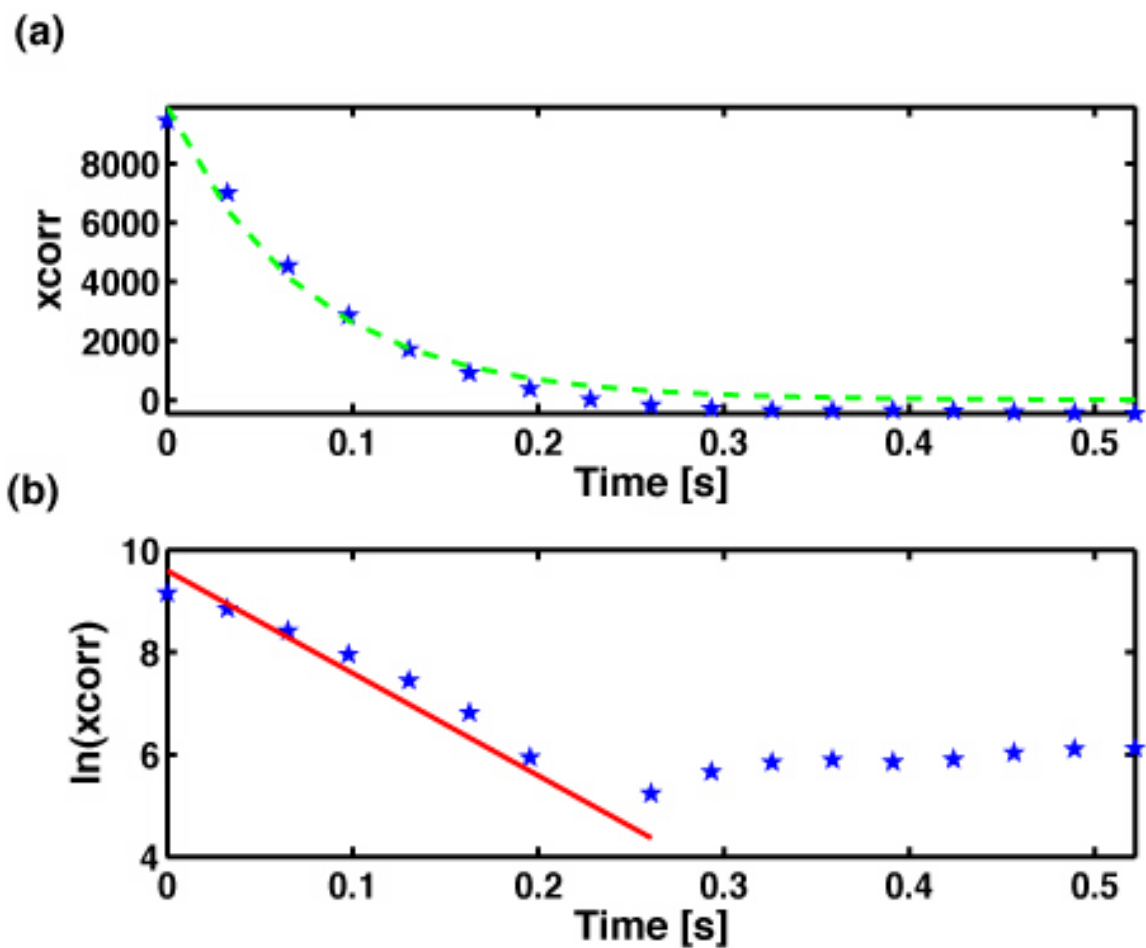


Figure 3.17: Autocorrelation of experimental positional data. (a) Autocorrelation showing exponential decay. Green line: exponential fit. (b) Autocorrelation in logarithm. Red line represents the linear fit from which the spring constant  $k_{\text{eff}}$  and drag coefficient  $\gamma$  are extracted.

of the traditional hidden Markov model to TPM data.

### 3.4.2 Diffusive-HMM

Beausang. *et al.* developed a method, namely diffusive HMM, for analyzing TPM data (Beausang *et al.* 2007). The goal of this method is trying to employ the concept of HMM and customize it for TPM data, through which the rate constants are directly derived from the positional data obtained in the TPM experiments. Based on standard HMM, we first have to choose an observable variable, which is only dependent on the hidden state that the system is in. In the absence of LacI protein, the

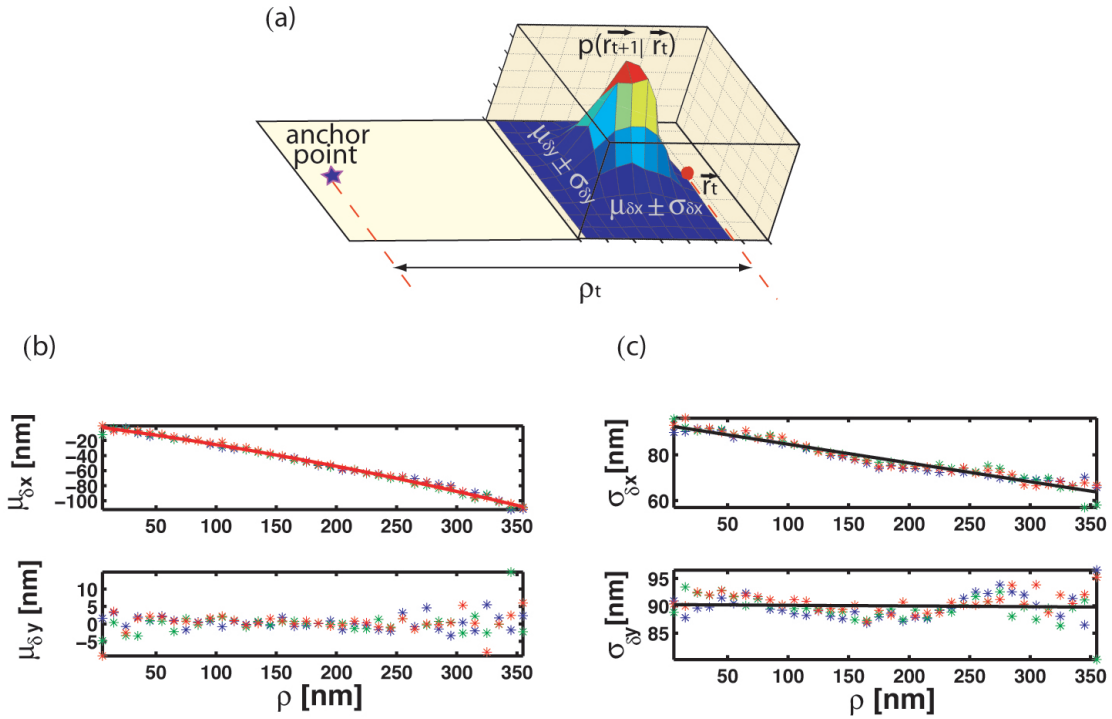


Figure 3.18: Construction of probability distribution for the next position. (a) Histogram displacements  $\delta x$  and  $\delta y$  at  $t+1$  when the distance to anchor point is  $\rho_t$ . (b) Mean of  $\delta x$  and  $\delta y$  as function of  $\rho$  and its corresponding second-order polynomial fit. (c) Standard deviations of  $\delta x$  and  $\delta y$  as function of  $\rho$  and its corresponding first-order polynomial fit. Markers with different colors in plot (b) and (c) represent different data sets.

bead executes tethered Brownian motion, and this motion is itself a Markov process, i.e., the beads displacement  $\delta \vec{r}_{t+1}$  depends only on  $\vec{r}_t$ , not on earlier positions. Therefore, this displacement could be used as the observed variable in the HMM. The next step, as shown in figure 3.7, is to obtain the

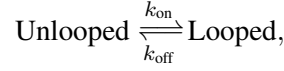
probability for a given point at time  $t$  to be in each state  $p(\vec{r}_t, q_t)$ . To find the required modification for applying HMM to our TPM data, we first need to get the probability distribution for the next position  $p(\vec{r}_{t+1}|\vec{r}_t)$  for each state using the observed variable  $\delta\vec{r}_{t+1}$ . To avoid modeling the complicated dynamical details of tethered particle Brownian motion near the surface, this information for each state is extracted phenomenologically from the control experimental data. Such control experiments carrying similar dynamical information of the beads, are performed in the following ways: (i). To obtain the information for the unlooped state, the bead's motion is observed in the absence of the DNA looping protein LacI. (ii). For the looped state, we monitor the bead's motion in the presence of a LacI mutant V52C instead of LacI itself. This mutant is designed to permit disulfide bond formation, which makes important contacts that are critical to DNA binding. As a result, V52C has increased affinity for DNA operators (Falcon *et al.* 1997), leading to a measurement of primarily looped states. Such data containing only one type of looped state are selected to obtain the information that serves as input to the HMM model. An alternative way to obtain the information for looped state is using a shorter tethered DNA which has similar effective tethered length to the one when the loop forms. However, to obtain such similar effective tethered length is tricky because the resolution in our motion-vs-length calibration data is poor. From data obtained in the control experiments, we first calculate the displacements between two adjacent frames throughout the entire time-series data, and then histogram  $\delta x$  and  $\delta y$  as function of distance from the anchor point  $\rho_t$ , where  $\rho_t = \sqrt{x_t^2 + y_t^2}$ . Figure 3.18 (a) shows that the probability distributions for the next point  $\vec{r}_{t+1}$  are both roughly 2D Gaussians centered around a point that depends on  $\vec{r}_t$ , with widths  $\sigma$  reflecting the random excursions of Brownian motion between two adjacent frames. Mathematically, it can be expressed as:

$$P(\vec{r}_{t+1}|\vec{r}_t) = \frac{1}{\sqrt{2\pi\sigma_{\delta x}^2(\rho_t)}} \exp[-(\delta_x - \mu_{\delta x}(\rho_t))^2/2\sigma_{\delta x}^2(\rho_t)] \times \frac{1}{\sqrt{2\pi\sigma_{\delta y}^2(\rho_t)}} \exp[-\delta_y^2/2\sigma_{\delta y}^2(\rho_t)] \quad (3.16)$$

where  $\mu_{\delta x}$ ,  $\sigma_{\delta x}$  and  $\sigma_{\delta y}$  denote the mean displacement in the  $x$  direction, and the standard deviation in the  $x$  and  $y$  directions, respectively. These three values are functions of  $r$ , which are shown in figure 3.18 (b) and (c). The lines are from the polynomial fit to the control experimental data: the red line represents the fitting for mean displacement in the  $x$  direction and black lines represent the fitting for standard deviation of the displacement in either  $x$  or  $y$  directions. This expression above gives the distribution of the next point location  $\vec{r}_{t+1}$  given the current location  $\vec{r}_t$ . In order to incorporate the hidden state dependence, we construct a joint distribution function  $D(q_{t+1}, \vec{r}_{t+1}|q_t, \vec{r}_t)$ ,



the probability of observing  $(q_{t+1}, r_{t+1})$  given  $(q_t, r_t)$  as follows. Let us first consider the simple two state (i.e., looped and unlooped) system,



and let "1" and "2" denote the unlooped and looped states, respectively.

$$\begin{aligned} D(1, \vec{r}_{t+1}|1, \vec{r}_t) &= (1 - k_{\text{on}}T) p_1(\vec{r}_{t+1}|\vec{r}_t) \text{ if } \rho_{t+1} \text{ allows looping,} \\ &= p_1(\vec{r}_{t+1}|\vec{r}_t) \text{ otherwise,} \\ D(1, \vec{r}_{t+1}|2, \vec{r}_t) &= (k_{\text{off}}T) p_2(\vec{r}_{t+1}|\vec{r}_t) \text{ if } \rho_{t+1} \text{ allows looping,} \\ &= 0 \text{ otherwise,} \\ D(2, \vec{r}_{t+1}|2, \vec{r}_t) &= (1 - k_{\text{off}}T) p_2(\vec{r}_{t+1}|\vec{r}_t), \\ D(2, \vec{r}_{t+1}|1, \vec{r}_t) &= k_{\text{off}}T p_2(\vec{r}_{t+1}|\vec{r}_t), \end{aligned}$$

where  $k_{\text{on}}$  and  $k_{\text{off}}$  are the loop-formation and loop-breakdown rate constants, and  $T$  is the time interval between two adjacent frames. If the system is initially in the unlooped state ( $q_t = 1$ ) and  $\rho_t$  is too large to permit loop formation, then at time  $t+1$ , the system must remain in the unlooped state ( $q_{t+1} = 1$ ). However, if  $q_t = 1$  and  $\rho_t$  is less than the maximum excursion observed for beads with a permanently looped tether in the control experiments, then both final states are allowed. When the initial state is looped, the state of the system at  $t+1$  could be at either of the two states without restriction on the location of the bead. To simplify the expressions with matrix notation, let the rows denote the state at time  $t$  and columns denote the state at  $t+1$ :

$$D(q_{t+1}, \vec{r}_{t+1}|q_t, \vec{r}_t) = \left[ \begin{pmatrix} 1 - \theta k_{\text{on}}T & \theta k_{\text{on}}T \\ k_{\text{off}}T & 1 - k_{\text{off}}T \end{pmatrix} \begin{pmatrix} p_1(\vec{r}_{t+1}|\vec{r}_t) & 0 \\ 0 & p_2(\vec{r}_{t+1}|\vec{r}_t) \end{pmatrix} \right]$$

, where  $\theta = 1$  if looping is allowed and 0 otherwise. For a three state system, as illustrated in figure 3.14, we have:

$$D(q_{t+1}, \vec{r}_{t+1} | q_t, \vec{r}_t) = \begin{pmatrix} 1 - \theta_1 k_{12} T - \theta_3 k_{13} T & \theta_1 k_{12} T & \theta_3 k_{13} T \\ k_{21} T & 1 - k_{21} T - \theta_2 k_{23} T & \theta_2 k_{23} T \\ k_{31} T & k_{32} T & 1 - k_{31} T - k_{32} T \end{pmatrix} \times \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{pmatrix}$$

where  $\theta_1, \theta_2$  and  $\theta_3 = 1$  if certain looping configurations are permitted, and zero otherwise. From this expression, we can obtain the probability of the system being at each state at time  $t+1$  if we know the bead's displacement from time  $t$  to  $t+1$ . So far, we achieve the second step of the HMM application by following the above procedure, which is outlined in figure 3.7 (b). To obtain the total probability for the observed sequence  $P(O)$  for the next step, the above calculation is repeated for all pairs of points in the data and summed over all possible sets of the hidden states  $q_t$ , which results in:

$$\begin{aligned} P(O) &= P(\vec{r}_N, \vec{r}_{N-1}, \dots, \vec{r}_t, \dots, \vec{r}_1) \\ &= \sum_{\{q_i\}} D(q_N, \vec{r}_N | q_{N-1}, \vec{r}_{N-1}) \times D(q_{N-1}, \vec{r}_{N-1} | q_{N-2}, \vec{r}_{N-2}) \times \\ &\quad \dots \times D(q_{t+1}, \vec{r}_{t+1} | q_t, \vec{r}_t) \times \dots \times D(q_2, \vec{r}_2 | q_1, \vec{r}_1). \end{aligned}$$

In this expression, the rate constants  $k$ 's are unknown. The number of parameters depends on the number of the states the system has. The rate constants can be determined by maximizing  $P(O)$ . The best single trajectory for the observed sequence can also be obtained once we have all the parameters in the model. Figure 3.19 shows the transitions obtained with the DHMM method. In plot(a), the detected transition is superposed on the Gaussian-filtered bead motion used in the thresholding method. In a large number of cases the thresholding method appears to be as efficient as the DHMM method at capturing transitions. However, narrow transitions are not captured by the thresholding method because of window averaging, as explained in the previous section. Plot (c) zooms to one of these narrow regions and clearly shows the different results resolved by these

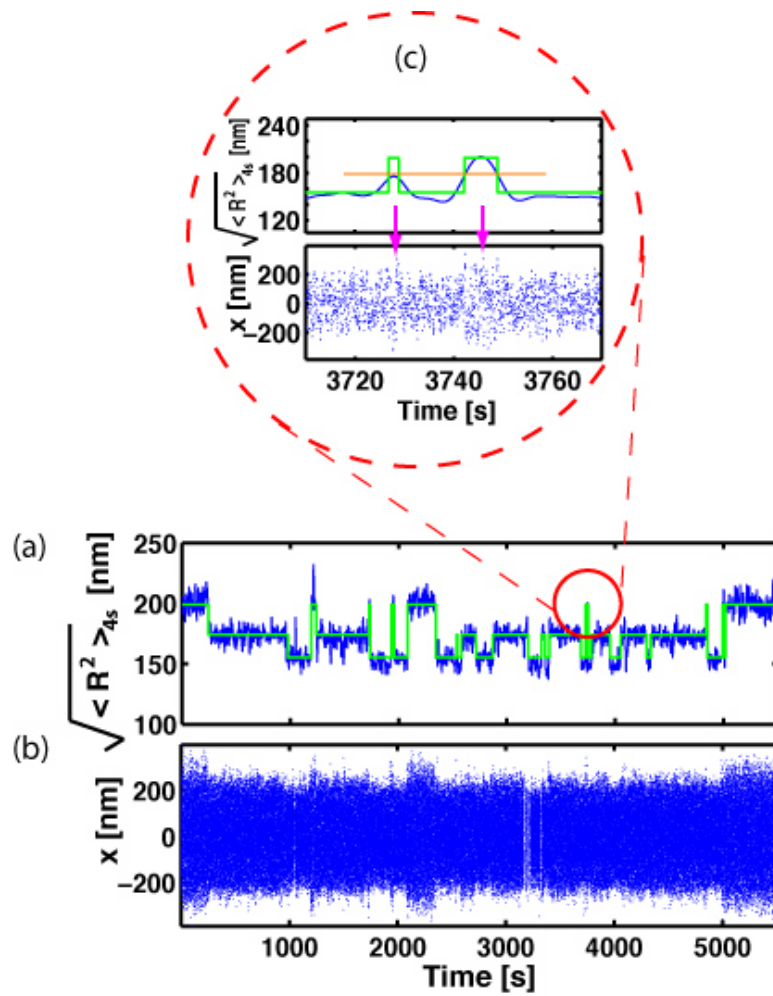


Figure 3.19: Transition detected using 3-state DHMM method. (a) Transitions detected (in green) with Gaussian filtered  $\sqrt{\langle R^2 \rangle_{4s}}$  motion superposed. (b) Corresponding histogram of  $\sqrt{\langle R^2 \rangle_{4s}}$ . (c) An enlarged plot showing the region circled in red in (a). Orange represents the threshold value and pink arrows point at the corresponding positional data points in the pulse regions.

methods. To gain further insight about what happens during these short pulses, drift-free positional data is inspected. In the bottom plot in (c), as the pink arrows point out, a few spikes take place during that period, which demonstrates the higher temporal resolution of the DHMM compared to threshold method.

### 3.4.2.1 Apply DHMM to TPM Data

Our measurements pose some additional challenges when we try to apply this method to our TPM data. First, besides long term drift, which is effectively removed by filtering, the positional data can be further corrupted by transient sticking events or transient erroneous tracking of position due to free-floating particles. Dealing with these issues is a great challenge for data analysis. One obvious solution is to remove points believed to be the corrupted. However, this creates another bias by artificially merging two adjacent events, or omitting real events. To overcome this problem we assign the corrupted points an equal probability of occupying each state; therefore, what might be happening during the corrupted sequence is determined by the maximum likelihood technique. The second broad challenge involves dealing with variability of the calculated rate constants across beads. One possible solution is inspired by the conventional thresholding method. For each data set from one bead, the best possible trajectory can be obtained from the DHMM method, and durations for each state can be determined. We might be tempted to simply collect all the individual durations for each state resolved by the DHMM method and obtain the time constant from the combined histogram of all durations, as was done for the thresholding method. However, if we use such parameters to the DHMM model, it might not result in the most likelihood rate constants for all the trajectories. An alternative, more elegant solution is to obtain the rate constants by maximizing the probability for entire data sets. The advantage of this approach over the previous one is that ambiguous events in one bead might be resolved because of the additional information provided by other beads.

We have described two different methods for extracting the rate constants describing the LacI-DNA interaction. First, the conventional thresholding method uses an amplitude threshold to distinguish the different looped and unlooped states. This approach requires window averaging of data to boost signal-to-noise, which inevitably leads to loss of some information. Moreover, the final result depends strongly on the window size. The second method, based on the Diffusive Hidden Markov Model, does not rely on window averaging, thus temporal resolution is not sacrificed for the sake of better signal-to-noise. Furthermore, it is easy to extend this method to a higher number of states

system. For these reasons, we ultimately decide to exclusively employ the DHMM method in our analysis.

## Chapter 4

# Lac Repressor Concentration Dependence of DNA Looping

### 4.1 Introduction

The biological significance of DNA is primarily attributed to its sequence information. On the other hand, there are a wide range of processes for which DNA's physical basis as a stiff polymer also matters (Garcia *et al.* 2007). For example, the packaging of DNA into nucleosomes appears to select for sequence motifs that are particularly flexible (Segal *et al.* 2006; Cloutier and Widom 2004). In the setting of transcriptional regulation, there are a host of regulatory architectures both in prokaryotes and eukaryotes which require the interaction of sequences on the DNA which are not adjacent (Adhya 1989; Schleif 1992; Matthews 1992). These interactions are mediated by DNA-binding proteins which have to deform the DNA. Of late, it has become possible to perform genome-wide surveys to determine the entirety of looped configurations induced by a given protein (Loh *et al.* 2006; Wei *et al.* 2006). One of the most transparent examples of DNA looping is in bacteria where some repressors and activators can bind at two sites simultaneously resulting in a DNA loop. This effect was first elucidated in the context of the arabinose operon (Dunn *et al.* 1984). It is an amusing twist of history that the two regulatory motifs considered by Jacob and Monod, namely, the switch that makes the decision between the lytic and lysogenic pathways after phage infection (Ptashne 2004) and the decision making apparatus associated with lactose digestion in bacteria (Schleif 1992; Müller-Hill 1996), both involve DNA looping as well.

To understand the physical mechanism of the biological action at a distance revealed by DNA looping it is necessary to bring both *in vitro* and *in vivo* experiments as well as theoretical analyses to bear on this important problem. Over the last few decades there have been a series of impressive

and beautiful experiments from many quarters that inspired our own work. In the *in vivo* context, it is especially the work of Muller-Hill and coworkers that demonstrates the intriguing quantitative implications of DNA looping for regulation (Muller *et al.* 1996). In their experiments, they tuned the length of the DNA loop in one base pair increments and measured the resulting repression. More recently, these experiments have been performed with mutant bacterial strains which were deficient in architectural proteins such as HU (Becker *et al.* 2005). On the *in vitro* side, single molecule experiments using the tethered-particle method have also contributed significantly (Finzi and Gelles 1995; van den Broek *et al.* 2006; Vanzi *et al.* 2006; Zurla *et al.* 2006). The idea of these experiments is to tether a piece of DNA to a microscope cover slip with a bead attached to the end. The DNA construct has the relevant binding sites (operators) for the protein of interest along the DNA and when one of these proteins binds, it shortens the length of the tether. As a result of the shorter tether, the Brownian motion of the bead is reduced. Hence, the size of the random excursions of the bead serves as a reporter for the status of the DNA molecule (i.e., looped or unlooped, DNA-binding protein present or not).

In addition to single-molecule studies, *in vitro* biochemical assays have also shed important light on the interactions between transcription factors and their DNA targets. Both filter binding assays and electrophoretic mobility shift assays have been widely used to study how variables dictating DNA mechanics such as length and degree of supercoiling, alter the looping process (Krämer *et al.* 1987; Hsieh *et al.* 1987; Krämer *et al.* 1988; Whitson *et al.* 1987; Borowiec *et al.* 1987). One of the missing links in the experimental elucidation of these problems is systematic, single-molecule experiments which probe both the length and sequence dependence of DNA looping. Our view is that such systematic experiments which complement the corresponding systematic *in vivo* studies will help clarify the way in which both length and sequence contribute to the probability of DNA looping. To that end, we have carried out experiments which probe the DNA looping process over a range of concentrations of repressor protein and for a series of different loop lengths. In addition, intrigued by the sequence preferences observed in nucleosomal DNA, we have made looping constructs in which these nucleosomal sequences are taken out of their natural eukaryotic context and are inserted between the operators that serve as binding sites for the Lac repressor. The point of this exercise is to see how the looping dynamics depends upon these tunable parameters, namely, repressor concentration (Chapter 4), length (Chapter 5), and sequence (Chapter 6).

## 4.2 Results

A typical experimental trace resulting from these experiments is shown in figure 4.1. As seen in the figure, as with recent work from Gelles and co-workers (Wong *et al.* 2007), there are clearly two distinct looped states as seen both in the trajectory and the histogram. Our working hypothesis is that these two looped states correspond to two different configurations of the Lac repressor molecule and its attendant DNA which we call the "open" and "closed" configurations. Control experiments (figure 4.2) with one of the two binding sites removed show no looping events, which further supports the idea that the two lower states are indeed looped configurations. Direct inter-conversion between the two looped species suggested these two looped configurations are due to the different conformations of Lac repressor protein (Wong *et al.* 2007) rather than two different DNA topologies (Friedman *et al.* 1995; Mehta and Kahn 1999; Semsey *et al.* 2004).

In order to extract quantities such as the free energy of looping associated with repressor binding, we require looping data at a number of different concentrations. The experiments described here use DNA constructs harboring two different operators,  $O_{id}$  and  $O_1$  as Lac repressor binding sites and the inter-operator spacing is 306 bp, as shown in figure 4.3(a). Representative single-molecule trajectories at each concentration tested here are presented in figure 4.3 (b). At very low concentrations, we expect that there will be negligible looping. At intermediate concentrations, the equilibrium situation will be dominated by states in which a single repressor molecule is bound to the DNA at the strong operator,  $O_{id}$ , punctuated by transient looping events. In the very high concentration limit, each operator will be occupied by a repressor making the formation of a loop nearly impossible. This progression of qualitative behavior is revealed in figure 4.4 which shows data from eight distinct concentrations of Lac repressor, as well as a single operator control in which the DNA lacks a secondary operator for Lac repressor to bind to. These curves are generated by summing the histograms from *all* of the individual trajectories for each concentration. One of the intriguing features of this data is the way in which the two looped states are turned off as a function of the concentration of Lac repressor is increased to very high levels. This phenomenon is expected since the Lac repressor exists always as tetramers under the conditions used here (Levandoski *et al.* 1996; Barry and Matthews 1999) and competition for binding at the second operator between loose Lac repressor and Lac repressor bound to the other operator is stronger as the concentration of Lac repressor increases. However, the two different looped species have slightly different response at high repressor concentrations. For example, at 1 nM concentrations, the intermediate looped state



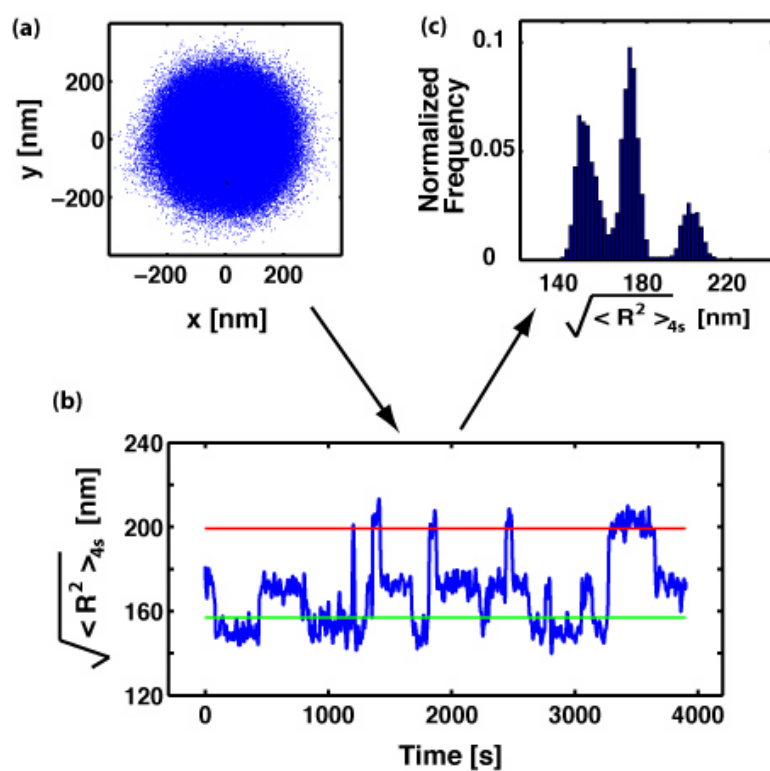


Figure 4.1: Different representations of TPM data. (a) Scatter plot of drift-corrected positional data. Each dot corresponds to the instantaneous projected position of the bead at a given instant in time. (b) Running average of RMS motion over 4s.  $R$  is the distance from the bead center (dots in panel (a)) to the tether attachment point (centroid of all dots in panel (a)). Red and green lines represent the expected motion, based on our calibration measurements (data not shown), for 901bp DNA and the same DNA when 305 bp (the interoperator spacing) are subtracted off of the full length 901bp tether. (c) RMS distribution. DNA used here is pUC305L1 with 100pM Lac repressor.

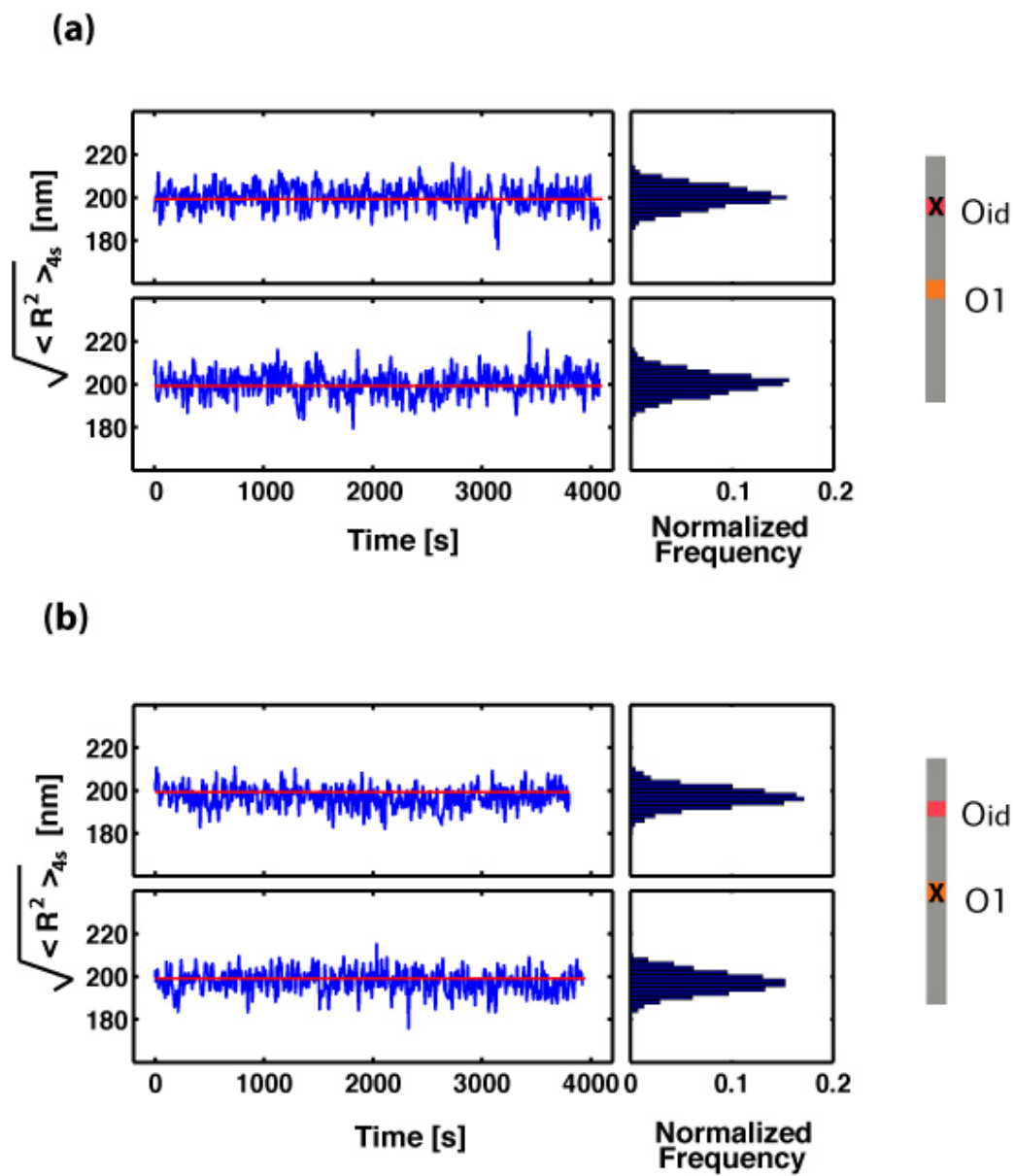


Figure 4.2: Typical trajectories with mono-operator DNA. (a)  $O_{id}$  is removed. (b)  $O_1$  is removed.

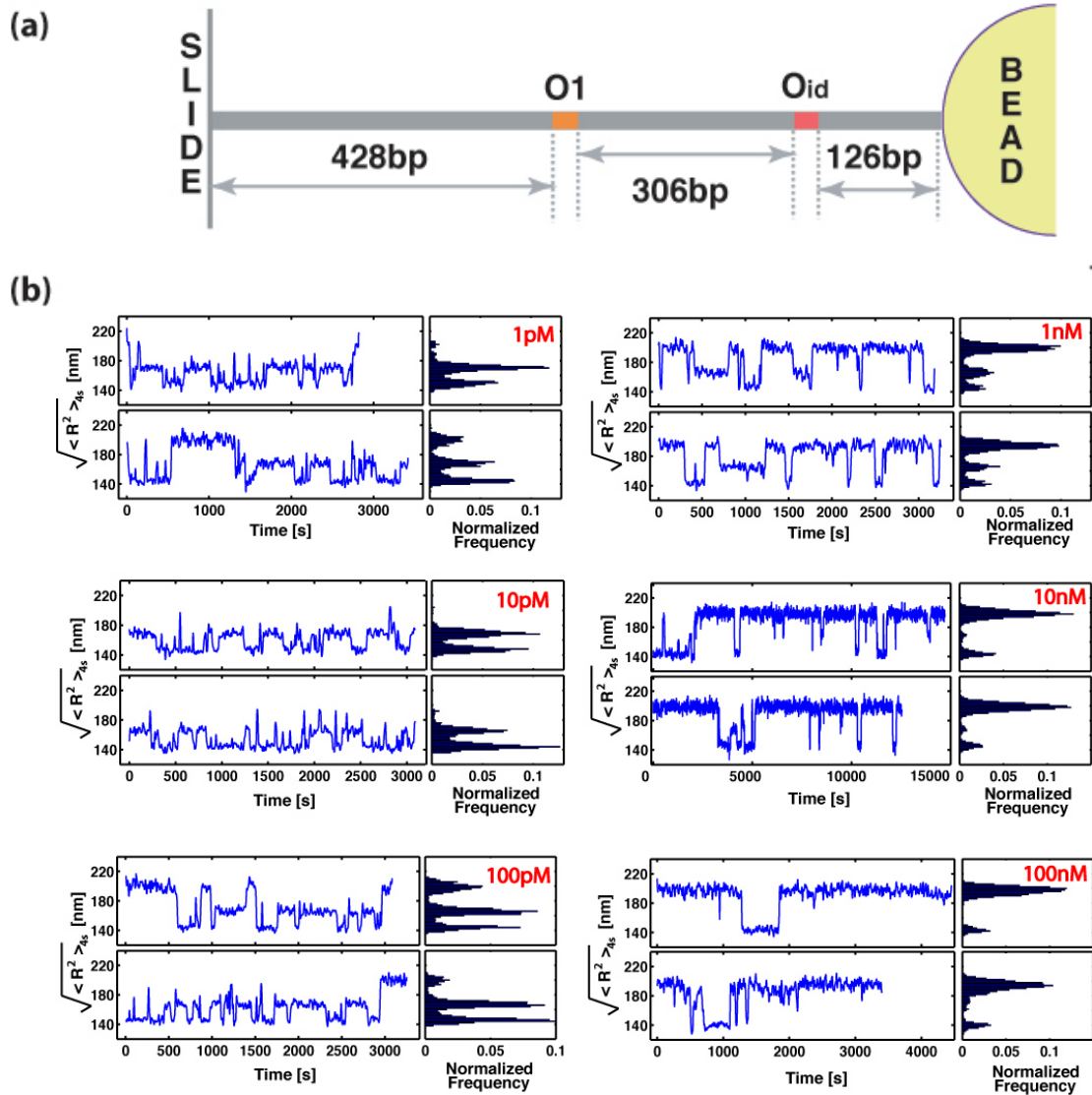


Figure 4.3: Typical Brownian motion traces of the DNA tethered beads in the presence of different concentrations of Lac repressor varying from 1pM to 100nM. DNA is 901bp in total length with 306bp inter-operator spacing.

Table 4.1: Gaussian fit parameters in the Lac repressor concentration titration experiment. Bin size is 2 nm. N is total number of beads.

[LacI]	a1	b1	c1	a2	b2	c2	a3	b3	c3	N
1pM	0.067	145.72	7.05	0.056	168.44	7.51	0.030	197.15	7.96	13
10pM	0.079	144.80	7.03	0.059	167.20	8.55	0.0075	197.03	10.41	26
100pM	0.062	146.79	7.50	0.061	168.28	8.16	0.018	198.57	8.63	57
1nM	0.070	143.03	6.81	0.0170	165.22	8.21	0.061	196.96	8.35	12
10nM	0.040	142.64	6.25	0.0055	169.34	19.83	0.096	197.74	8.04	17
100nM	0.014	141.79	6.10	0.0042	178.65	19.38	0.12	197.15	8.31	16
200nM	0.020	141.23	6.27	0.0061	171.41	11.52	0.11	195.73	8.10	8

has become very infrequent while the shortest looped state remains competitive. No concentrations below 1 pM are tested for technical reasons, such as non-specific adsorption of the surface and tetramer dissociation into dimers. Similar concentration dependence on LacI mediated DNA looping was studied before (Vanzi *et al.* 2006). However, only three concentrations, 4, 20 and 100 pM, were examined. Their result showed looping is suppressed as the concentration goes up, but not when the concentration is very low.

As mentioned above, figure 4.4 is a result of summing all the trajectories of multiple DNA molecules to gain a better statistic. If the system is ergodic, it should ideally be equivalent to accumulation of a long trajectory of a single DNA molecule. We accumulated 1 hour or more trajectories of 10 to 70 different beads for DNA molecules under the same condition. Figure 4.5 revealed the improvement of the statistic as the number of beads for accumulating goes up.

One way to characterize the looping probability as a function of concentration is shown in figure 4.6. There are several different ways to obtain such a plot. First, by examining the trajectories, we can simply compute the fraction of time that the DNA spends in each of the different states, with the looping probability given by the ratio of the time spent in either of the looped states to the total elapsed time (blue points in figure 4.6). A second way of obtaining this same quantity is to use figure 4.4 and to compute the areas under the different peaks and to use the ratios of areas as a measure of looping probability (brown points in figure 4.6).

These results can also be explored from a theoretical perspective using the tools of statistical mechanics (Bintu *et al.* 2005b; Saiz and Vilar 2006a). The goal of such a model is to compute the probability of the various microstates available to the repressor-DNA system as shown in figure 4.7. The simplest model posits 5 distinct states (Wong *et al.* 2007; Finzi and Gelles 1995; Vanzi *et al.* 2006): both operators empty,  $O_{id}$  occupied by repressor without looping,  $O_1$  occupied by repressor

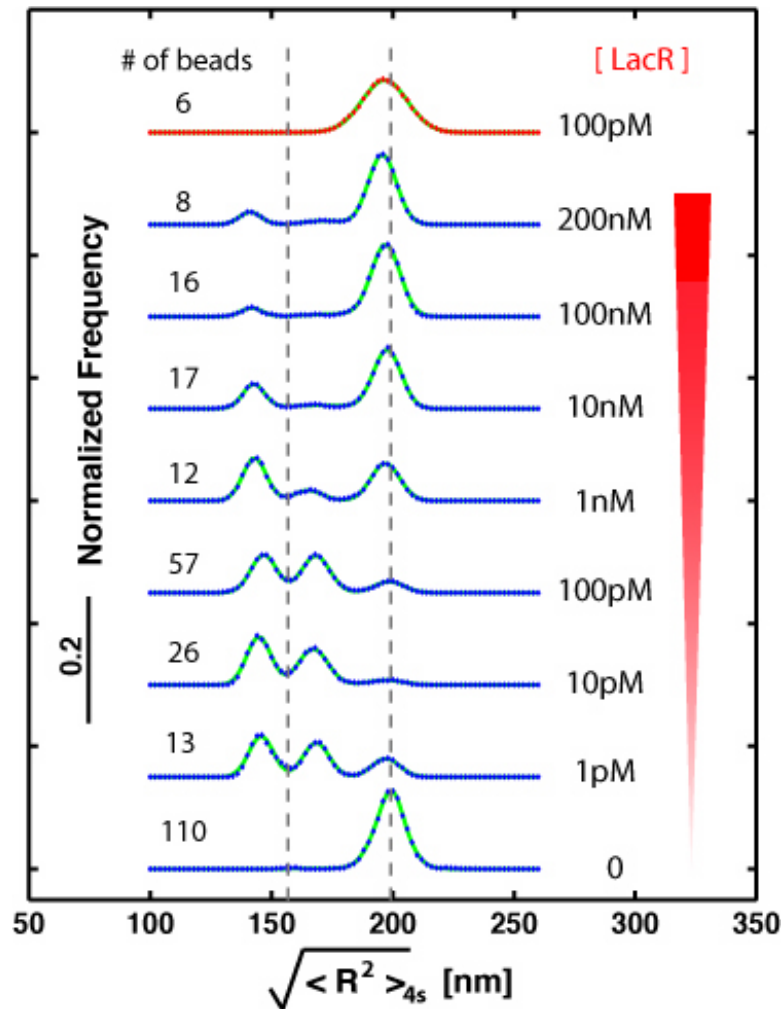


Figure 4.4: Concentration dependence of looping probability. The histograms show the distribution of RMS motions averaged over 4s at different concentrations of Lac repressor. Experimental data are represented by diamonds (blue: di-operator DNA; red: mssssssssssso-operator DNA) and a fit to the data using a sum of three Gaussians is shown in green. The fit parameters are displayed in table 4.1. The two gray dashed lines represent the expected motion, based on our calibration measurements (data not shown), for 901 bp DNA and the same DNA when 306 bp (the interoperator spacing) are subtracted off of the full length 901 bp tether. The short tether loop has smaller motion than expected by simple subtraction of the inter-operator spacing, which is expected due to the imperfect exiting angle of DNA from the loop structure. The bottom and top curves show control experiments: bottom, no Lac repressor; top, no second operator. The widths of the distributions for the two control experiments are slightly different. One possible explanation is that there are two possible configurations for the DNA construct without the second operator, empty or bound by Lac repressor. Crystal studies shows that a sharp bend in operators occurs upon Lac repressor binding (Lewis *et al.* 1996), which would result in a slightly reduced effective tether length.

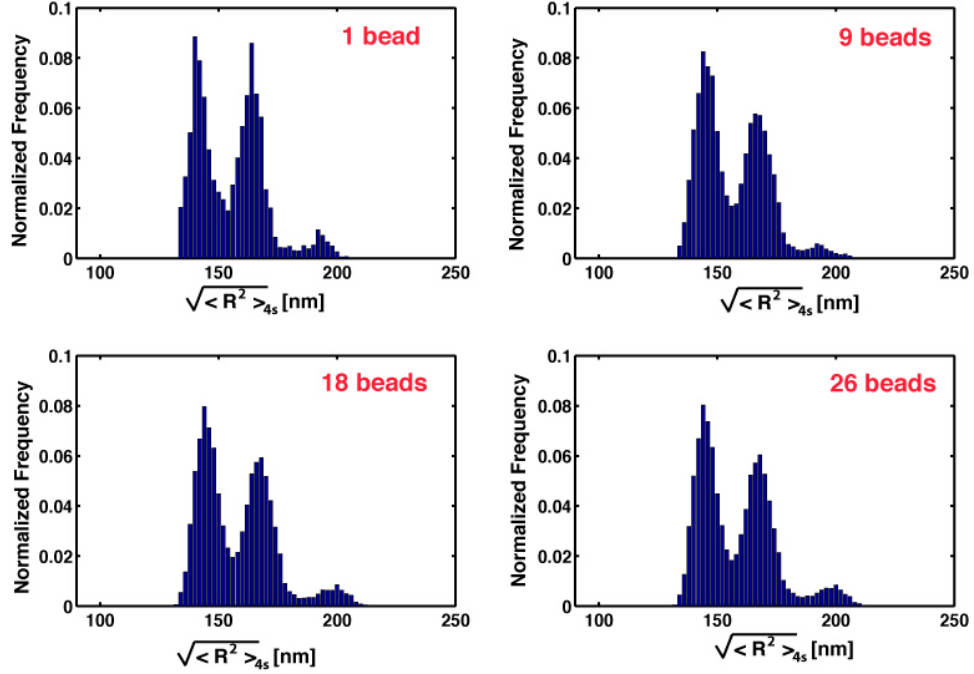


Figure 4.5: Averaging effect among the data set. Data obtained with pUC306L1 DNA in the presence of 10 pM Lac repressor.

without looping,  $O_{id}$  and  $O_1$  separately occupied by single repressors and the looped state (which we treat in the simplest model as having only one looped configuration). The model does not take into account of the effect of nonspecific binding of LacI to non-operator DNA. Such effect could be negligible because the equilibrium association constant of LacI to non-operator DNA at conditions similar to ours is around  $10^4 \text{ M}^{-1}$  (Revzin and von Hippel 1977), which is around 7 magnitudes less than specific binding. Given such association constant, the ratio between nonspecific bound LacI and the free LacI in solution can be calculated as:

$$\begin{aligned} \frac{[RD]}{[R]} &= K_a^{non} \times [D] \\ &\approx 1.8 \times 10^{-5}, \end{aligned}$$

where  $[RD]$  is the concentration of non-specifically bound LacI,  $[R]$  is the concentration of LacI in solution, and  $[D]$  is the DNA concentration, which is around 2 pM in our experiment. For  $[R] = 200 \text{ nM}$ , we have  $[RD] \approx 3.6 \text{ pM}$ , which is much less than the free LacI in solution.

The key parameters that show up in this model are the binding energies of repressor for the two operators,  $\Delta\epsilon_{id}$  and  $\Delta\epsilon_1$ , the looping free energy  $\Delta G_{loop}$  and the concentration of repressor

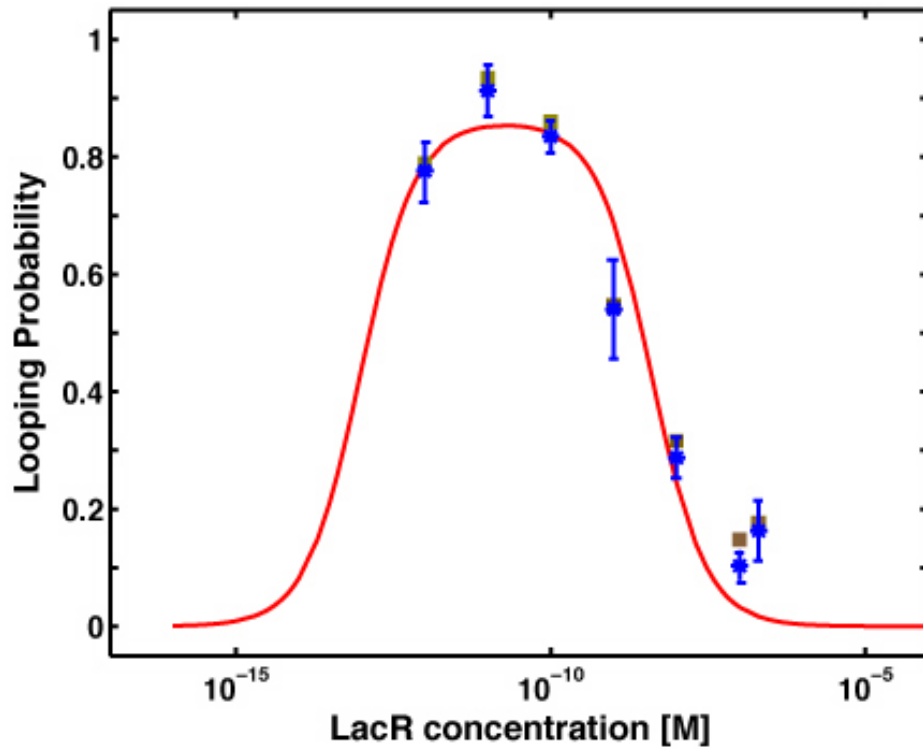


Figure 4.6: Looping probability at different concentration of Lac repressor. The DNA used in these experiments is 901 bp long and the inter-operator spacing is 306 bp. The vertical axis gives looping probability (fraction of time spent in either of the two looped states). The points were obtained using two different methods: The blue points were obtained by calculating the fraction of time that the system spends in one of the two looped states, using a transition detection algorithm based on a Hidden Markov Model (HMM) (Beausang *et al.* 2007; Beausang and Nelson 2007). The yellow points are derived from the histogram distribution by integrating the area under the two looped states over total area. The red curve is a fit to the experimental data using the statistical mechanics model described in the text.

[R]. The binding energy here contains two components. One is the standard positional free energy required for bringing one Lac repressor molecule to its DNA binding site at 1 M concentration of Lac repressor. The other is the interaction free energy due to the physical contact upon protein binding (Saiz and Vilar 2006b). The associated free energy with each configuration gives the statistical weights of the equilibrium probability (listed in the right column of figure 4.7). For example, to obtain the probability of the looped state, we construct the ratio of state (v) in the figure to the sum over all five states, as shown in equation 4.1:

$$P_{\text{loop}} = \frac{[R] \exp [-(\Delta\epsilon_{\text{id}} + \Delta\epsilon_1 + \Delta G_{\text{loop}})/k_{\text{B}}T]}{Z_{\text{tot}}}, \quad (4.1)$$

where  $Z_{\text{tot}}$  is the normalization factor (the partition function) obtained by summing over all the states available to the system and given by

$$\begin{aligned} Z_{\text{tot}} = & 1 + [R] \exp [-(\Delta\epsilon_{\text{id}})/k_{\text{B}}T] + [R] \exp [-(\Delta\epsilon_1)/k_{\text{B}}T] \\ & + [R] \exp [-(\Delta\epsilon_{\text{id}} + \Delta\epsilon_1 + \Delta G_{\text{loop}})/k_{\text{B}}T] + [R]^2 \exp [-(\Delta\epsilon_{\text{id}} + \Delta\epsilon_1)/k_{\text{B}}T]. \end{aligned} \quad (4.2)$$

Using the data of figure 4.6, we can actually obtain the looping free energy, as well as the binding energies by fitting the data to equation 4.1 with the result:  $\Delta\epsilon_{\text{id}} = -28.0 \pm 7.3 k_{\text{B}}T$ ,  $\Delta\epsilon_1 = -21.3 \pm 2.8 k_{\text{B}}T$  and  $\Delta G_{\text{loop}} = 19.5 \pm 1.4 k_{\text{B}}T$ . The two binding energies are in good agreement with previous studies, where they measured either the equilibrium association or dissociation constant of Lac repressor binding to symmetric operator sequence  $O_{\text{id}}$  or wild type operator  $O_1$ . Conversion of the results of previous studies into binding energies gives:  $\Delta\epsilon_{\text{id}} = -28.5 \pm 0.2 k_{\text{B}}T$  in 200 mM KCl (the same salt conditions used in our experiments) (Levandoski *et al.* 1996) and  $\Delta\epsilon_1 = -24.8 \pm 0.1 k_{\text{B}}T$  in 150 mM KCl (Zhang and Gottlieb 1993). Though similar in magnitude, our higher energy for  $\Delta\epsilon_1$  likely arises because we used higher ionic strength conditions and higher ionic strength increases the binding energy, i.e. binding is less favored.

Once the parameters that characterize the model are in hand, we can plot the probability of all five possible states as a function of the Lac repressor concentration as shown in figure 4.8. What this figure reveals is that at the concentrations we normally work, the system is dominated by the looped state and the state with single occupancy of  $O_{\text{id}}$ . However, previous studies (Finzi and Gelles 1995; Vanzi *et al.* 2006) have demonstrated that more than one species are dominated in the unlooped states using a similar setup, where two identical operators  $O_1$  were used, and the inter-







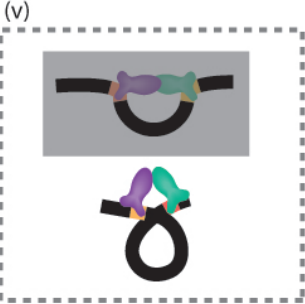
Representative States	Weights
(i) 	1
(ii) 	$[R]\exp(-\Delta\epsilon_{id}/k_B T)$
(iii) 	$[R]\exp(-\Delta\epsilon_1/k_B T)$
(iv) 	$[R]^2\exp[-(\Delta\epsilon_1+\Delta\epsilon_{id})/k_B T]$
(v) 	$[R]\exp[-(\Delta\epsilon_1+\Delta\epsilon_{id}+\Delta G_{loop})/k_B T]$

Figure 4.7: States and weights for the Lac repressor-DNA system ( Bintu *et al.* 2005b). Each of the five state classes shown in the left column has a corresponding statistical weight. All of the weights have been normalized by the state in which the DNA is unoccupied. State (v) is treated as a single looped state, even though there are multiple distinct looped configurations.

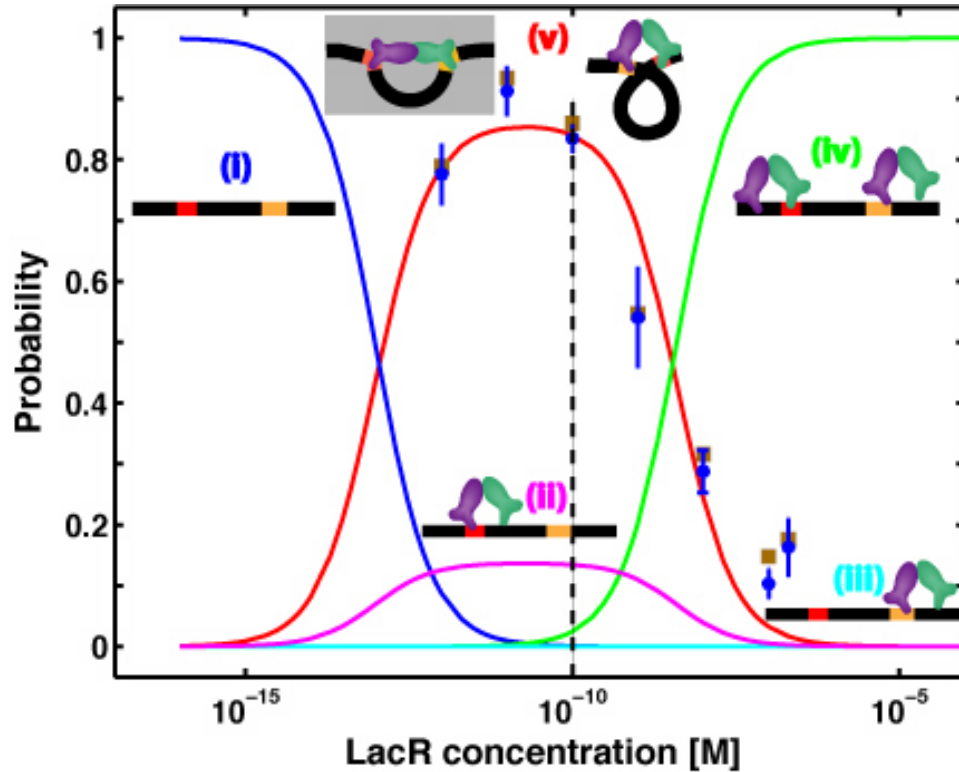


Figure 4.8: Measured and computed probabilities for different states of Lac repressor and operator DNA. The curves are all the results from the statistical mechanical model using parameters obtained from the measured data points. Blue curve: the probability of empty DNA with no Lac repressor bound. Red curve: the probability of the looped state. Green curve: the probability of the state with two Lac repressor molecules bound, one associated with each of the operators. Pink curve: the probability of the state where one Lac repressor is bound to the Oid site. Cyan curve: the probability of the state where one Lac repressor is bound to the O1 site. The blue circles are the data points for the probability of the looped state obtained using the Hidden Markov Model based transition detection method (Beausang *et al.* 2007) and computing the fraction of the overall trajectory when the DNA is looped, with no reference to which looped state. The orange squares show the fractional equilibrium population obtained by calculating the area under the Gaussian curve in the looping region in the histogram. The black dashed line indicates the Lac repressor concentration used in the phasing experiments.

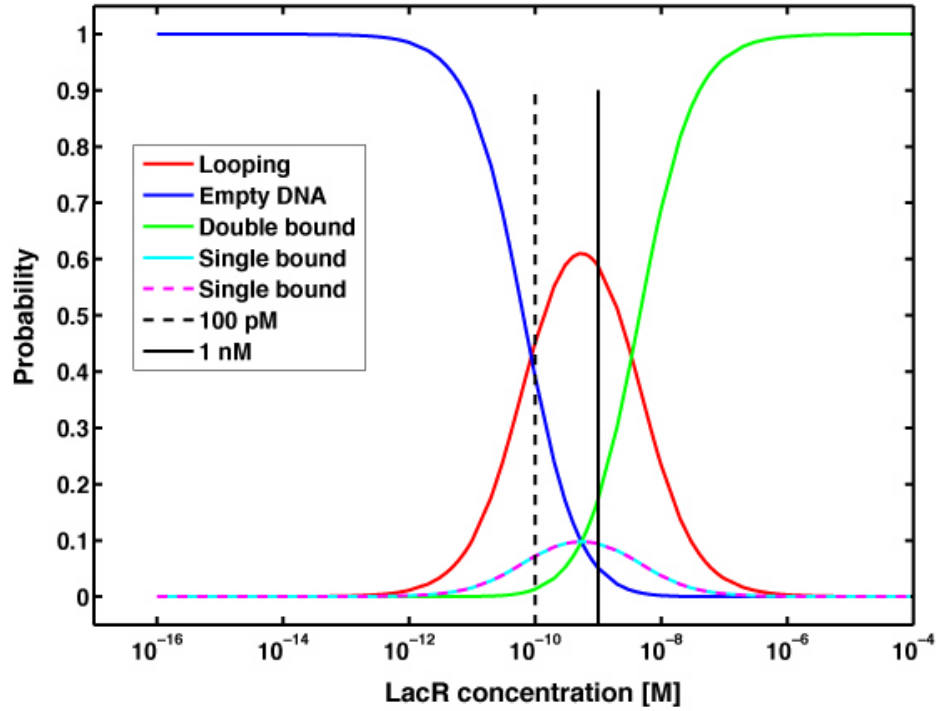


Figure 4.9: Computed probabilities for different states of Lac repressor and operator DNA for DNA with two identical operators O1. The curves are all the results from the statistical mechanical model using parameters  $\Delta\epsilon_1$  and  $\Delta G_{loop}$  obtained from our experiments. Blue curve: the probability of empty DNA with no Lac repressor bound. Red curve: the probability of the looped state. Green curve: the probability of the state with two Lac repressor molecules bound, one associated with each of the operators. Pink curve: the probability of the state where one Lac repressor is bound to the Oid site. Cyan curve: the probability of the state where one Lac repressor is bound to the O1 site. The black d lines indicate the Lac repressor concentration used in the previous experiments.

operator spacing is 21 bp less. In order to see if this little difference in the setup might lead to such big variations, we plot the probability of each state for their system using the binding energy  $\Delta\epsilon_1 = -21.3 \pm 2.8$  and  $\Delta G_{loop} = 19.5 \pm 1.4 k_B T$  obtained from our model. As illustrated in figure 4.9, when the [LacI] is 100pM (used by Vanzi *et al.* (2006)), there are three main species in the unlooped state: two types of single LacI bound DNAs and empty DNAs. When the [LacI] is 1 nM (used by Finzi and Gelles (1995)), four species coexist in the unlooped state and they are: two types of single LacI bound DNAs, empty DNAs and double-bound DNAs. The results revealed by our model could explain the phenomenon observed in the previous work.

### 4.3 Conclusions

Using tethered particle motion method, we are able to observe the Lac repressor concentration dependence of DNA looping. Lac repressor is a tetramer capable of simultaneously binding to two sites, thereby looping the intervening DNA. At very low concentrations, we expect that there will be negligible looping. At intermediate concentrations, the equilibrium situation will be dominated by states in which a single repressor molecule is bound to the DNA at the higher affinity operator,  $O_{id}$  in our case, punctuated by transient looping events. In the very high concentration limit, both operators will be occupied by a separate repressor, making the formation of a loop nearly impossible. We observe this type of concentration dependence in our experiments, and find it to be in quantitative agreement with a simple statistical mechanics model of looping probability. We were able to calculate the free energy associated with looping, having first subtracted the contribution of LacI-DNA binding, obtaining its first experimental measurement in a single molecule, *in vitro*, setting. Further more, we obtained unambiguous experimental evidence for the existence of an additional looped state, which is in agreement with theoretical predictions (Zhang *et al.* 2006; Swigon *et al.* 2006; Geanakopoulos *et al.* 2001; Balaeff *et al.* 2006) and other recent experiments (Gelles *et al.*, unpublished). The two looped configurations are expected to be the consequence of two distinct conformations of LacI, namely, the closed (V-shaped), and open form.

## Chapter 5

# Length Dependence of DNA looping

In the last chapter we characterized the concentration dependence of DNA looping dynamics. Here we study the effect of varying another parameter, DNA length. To make it possible to carry out a thorough analysis of the interplay between theories of transcriptional regulation (and DNA looping) and experiment, we have carried out a series of DNA looping experiments using the tethered-particle method for lengths ranging from 300 to 310 bp in one bp increments. The experiments described here use DNA constructs harboring two different operators,  $O_{id}$  and  $O_1$  as Lac repressor binding sites.

### 5.1 Length Dependence: 1 bp Resolution for a Whole Helical Turn

The beautiful *in vivo* repression experiments of Muller *et al.* (1996) demonstrate that the length of the DNA loop formed by Lac repressor strongly affects the probability of loop formation (especially for loop lengths less than 150 bp). In particular, they (and others) have observed phasing in which the relative orientations of the two operators changes the ease with which repressor can bind. Similar phasing effects have been observed in *in vitro* cyclization assays (Cloutier and Widom 2004, 2005). What has not been clear is how to concretely relate these seemingly disparate results on DNA mechanics from the *in vivo* and *in vitro* settings. Our idea was to systematically examine the same progression of DNA lengths that have been observed *in vivo* using TPM experiments. To that end, we have measured TPM trajectories for a series of interoperator spacings measured in 1 bp increments. Representative single-molecule trajectories at each inter-operator distance tested here are presented in figure 5.1.

The results of this systematic series of measurements are shown in figure 5.2. Each plot shows the probability of the three states for a particular interoperator spacing. What is particularly striking

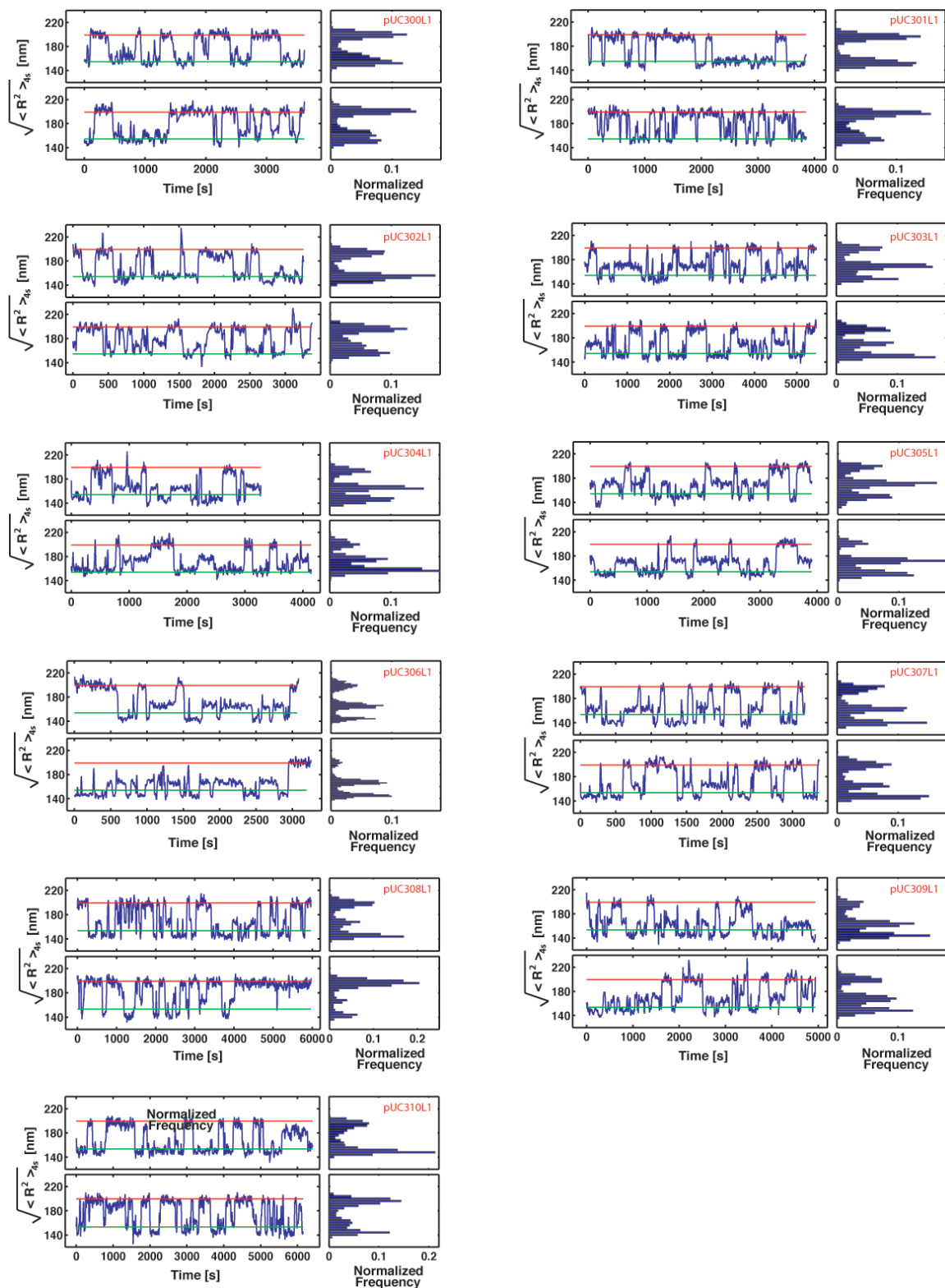


Figure 5.1: Typical Brownian motion traces of the DNA tethered beads with inter-operator spacing from 300 to 310 bp in 1 bp increment. Lac repressor concentration is 100 pM.

Table 5.1: Gaussian fit parameters in the phasing experiment. Bin size is 2 nm. N is total number of beads.

Spacing	a1	b1	c1	a2	b2	c2	a3	b3	c3	N
300 bp	0.046	149.54	6.46	0.048	162.70	12.12	0.026	197.48	9.45	47
301 bp	0.032	145.44	6.69	0.037	157.62	13.79	0.048	194.18	8.65	48
302 bp	0.043	146.89	9.39	0.020	161.03	15.74	0.047	194.25	8.61	31
303 bp	0.063	144.45	8.07	0.044	166.60	7.68	0.035	194.47	7.99	18
304 bp	0.076	148.38	9.17	0.034	167.12	7.92	0.018	194.17	9.24	63
305 bp	0.067	147.78	7.69	0.044	168.26	8.05	0.029	196.59	8.45	45
306 bp	0.062	146.79	7.50	0.061	168.28	8.16	0.018	198.55	8.66	57
307 bp	0.052	143.77	8.13	0.045	165.76	8.45	0.035	196.46	8.87	20
308 bp	0.057	140.01	8.23	0.033	160.45	9.47	0.037	194.53	9.40	23
309 bp	0.062	145.30	6.80	0.039	164.37	10.57	0.035	200.53	8.52	14
310 bp	0.027	146.50	5.79	0.033	158.57	15.01	0.050	199.89	9.53	19

about the data is that the intermediate looped state has a stronger dependence on the phasing than does the shorter looped state.

To obtain the probability of the system being in each state, we can examine the trajectories and detect any possible transitions between each state using DHMM (Beausang *et al.* 2007; Beausang and Nelson 2007). We then simply compute the fraction of time that the DNA spends in each of the different states, with the probability in one state is given by the ratio of the time spent in that states to the total elapsed time. The results are listed in table 5.2. Figure 5.3 shows a series of plots of the relative likelihood of the looped and unlooped states as a function of the DNA length between the two operators. Our result on looping probability (the insert figure in Figure 5.3) shows a periodic behavior with helical phasing. The maximum looping is achieved when two binding sites are 306 bp apart suggesting that at this distance, the two sites are on the same side of DNA. i.e., in-phase. The ability to form stable out-of-phase (two binding sites are on the opposite side of the DNA) loops with only a small reduction in stability is consistent with previous studies (Krämer *et al.* 1987; Wong *et al.* 2007). The relatively stable looping through the entire helical repeat is also consistent with relative constant repression level *in vivo* for similar inter-operator spacing (Muller *et al.* 1996). Looping probability (the inset figure in Figure 5.3) can be converted into the looping free energy based on the thermodynamic model described in chapter 4. One might expect when two operators on the opposite sites of the DNA, more energy is required to make the DNA in good registry in order for Lac repressor binding.

The shorter looped configuration turns out to be more robust than the middle one in terms of

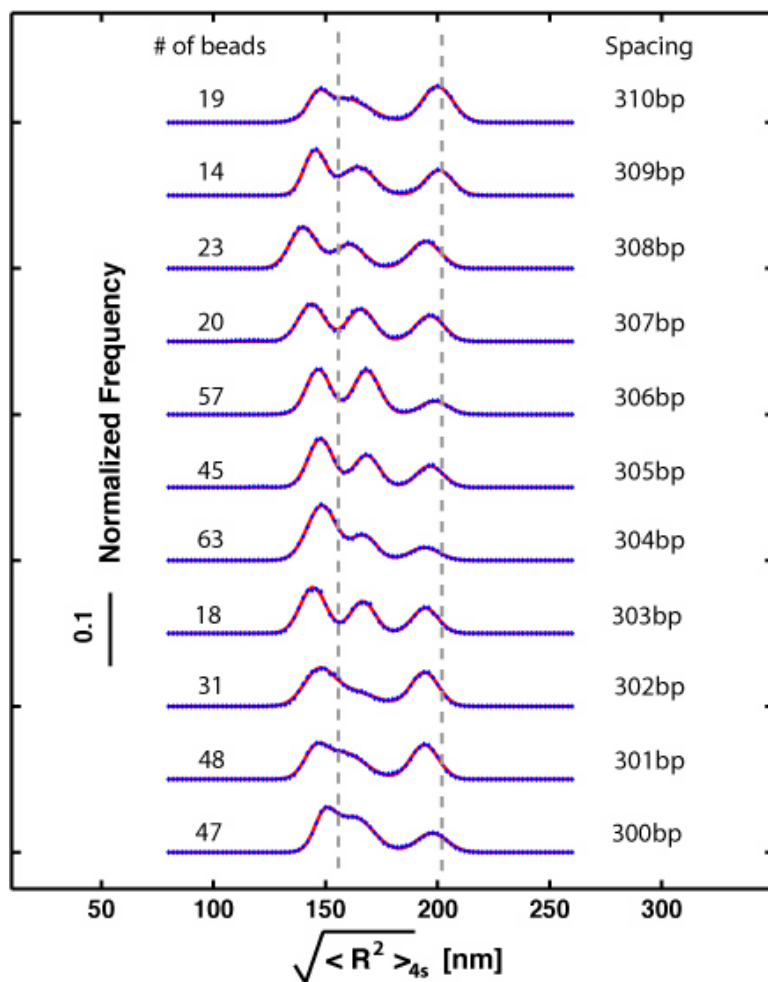


Figure 5.2: Histogram of the Brownian motion for DNAs with two Lac repressor binding sites spaced from 300 to 310 bp. Experimental data are in blue diamonds and three-Gaussian fit is in red. The fit parameters are displayed in tab. 5.1. The two gray dashed lines represent the expected motion, based on our calibration measurements (data not shown), for 901 bp DNA and the same DNA when 305 bp (the interoperator spacing) are subtracted off of the full length 901 bp tether.



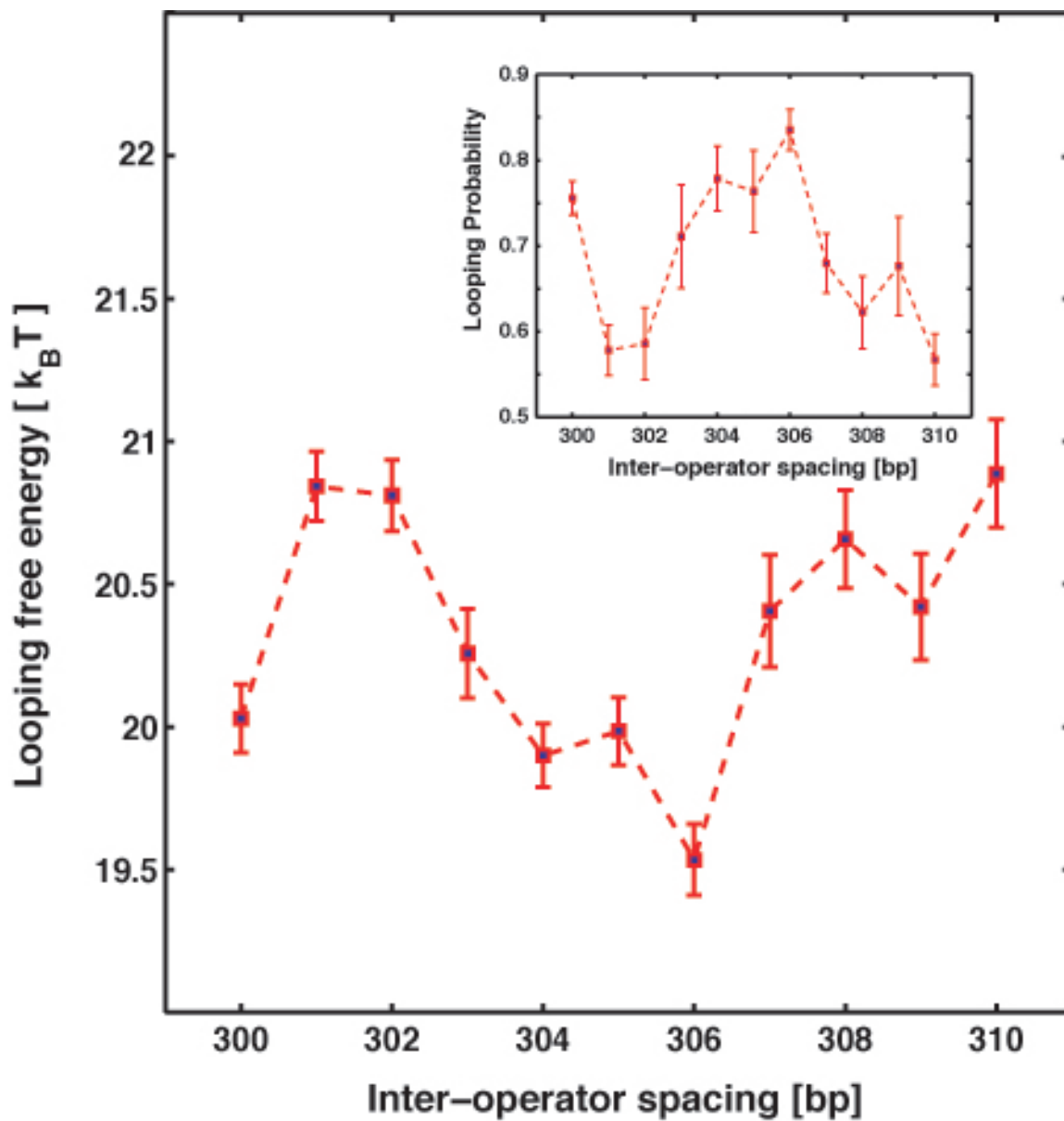


Figure 5.3: Total looping free energies with the corresponding looping probability for each of the inter-operator spacings from 300 to 310 bp. A periodic behavior over the entire helical repeat is observed for looping probability, and its maximum occurs at inter-operator spacing of 306 bp.

Table 5.2: Fraction of time spent in each state and the looping free energy. U: unlooped state. M: long tether looped state. B: short tether looped state. Lac repressor concentration is 100 pM

Spacing	U	M	B	$\Delta G_{loop} (k_B T)$
300 bp	0.24	0.33	0.43	$20.03 \pm 0.12$
301 bp	0.42	0.16	0.42	$20.84 \pm 0.12$
302 bp	0.41	0.14	0.45	$20.81 \pm 0.12$
303 bp	0.29	0.30	0.41	$20.26 \pm 0.16$
304 bp	0.22	0.29	0.49	$19.90 \pm 0.11$
305 bp	0.24	0.30	0.46	$19.99 \pm 0.12$
306 bp	0.16	0.40	0.44	$19.53 \pm 0.13$
307 bp	0.32	0.31	0.37	$20.42 \pm 0.20$
308 bp	0.38	0.16	0.46	$20.66 \pm 0.17$
309 bp	0.32	0.22	0.46	$20.42 \pm 0.19$
310 bp	0.43	0.22	0.35	$20.89 \pm 0.19$

responding to the twist effect. The relative stabilities of the two looped states are compared in Figure 5.4, where the ratio of time spent on each looped state and time in unlooped state is calculated. As shown here, when it is in phase, two looped configurations have nearly equally probability. When it is out of phase, the bottom looped state is dominant.

The two looped structures, as explained before, are suspected to be due to the two different conformations of Lac repressor protein (Wong *et al.* 2007). However, the question of which looped state corresponds to which conformations of Lac repressor is still unclear. Based on our experiments of tuning both Lac repressor concentration and distance between binding sites, we see that the middle looped state is more sensitive to both concentration and twist change. To gain more insight into this problem, we examined the effective tether length upon loop formation. As illustrated in Figure 5.5, the effective tether length of the unlooped state (blue line) is in good agreement with the actual tether length (red line). For the looped state, the effective tether length of the bottom looped state is close to the length one would expect when the DNA between the two binding sites has been removed (Pink line), surprisingly however, the one for middle looped state is around 100 bp longer than the length represented by the pink line. It is hard to imagine what sort of looped configuration will reduce the tether length by an increment smaller than the actual distance between the two binding sites. Even if we consider the protein size, which is about 17.2 nm long when it is in open conformation (Ruben and Roos 1997), it does not account for this 100 bp difference.

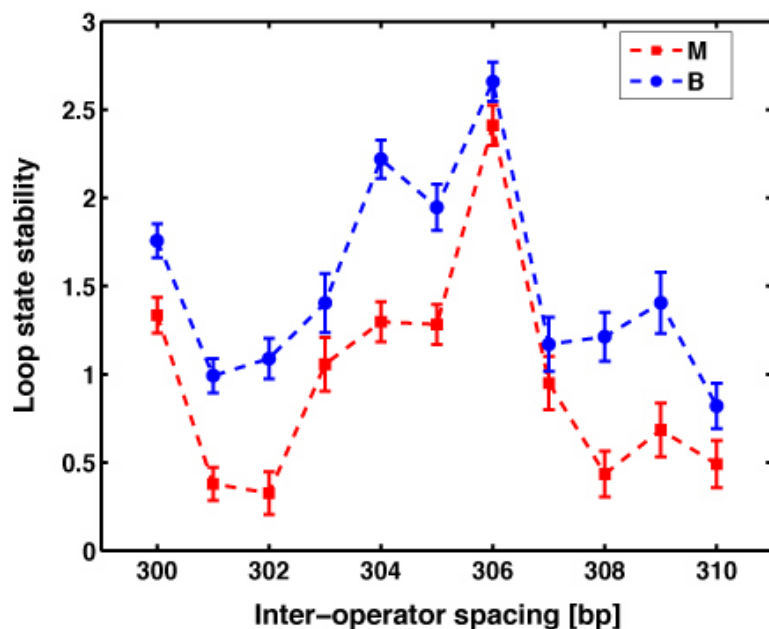


Figure 5.4: Stability of the two looped states as a function of inter-operator spacing.

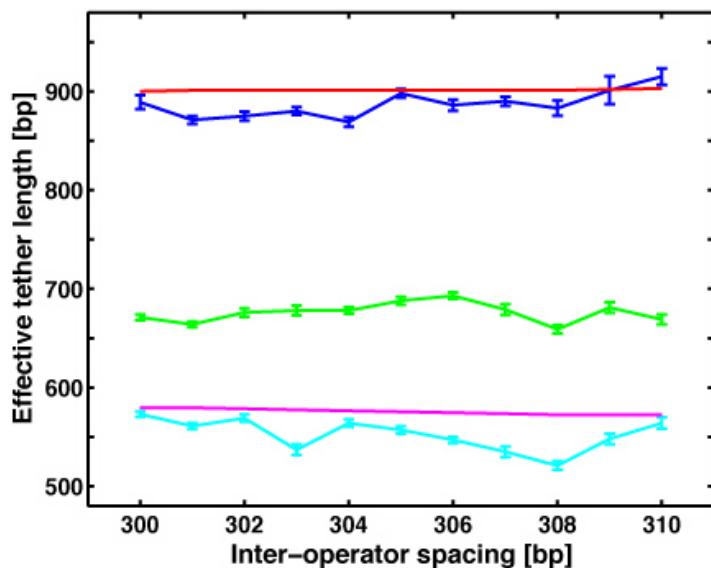


Figure 5.5: Effective tether length as a function of inter-operator spacing. Blue: unlooped state. Green: middle state. Cyan: bottom state. Red: actual tether DNA length. Pink: DNA length subtracts inter-operator spacing.

## 5.2 Kinetic Analysis

Our thermodynamic model in chapter 4 demonstrates that at LacI concentration of 100 pM, which is we used in this experiment, the unlooped state is only dominated by one species (single bound LacI at  $O_{id}$ ). Under this condition, it is reasonable to assume our system is a three-state system, i.e., unlooped state, short tethered looped state, long tethered loop state. We then apply three-state DHMM (Beausang *et al.* 2007; Beausang and Nelson 2007) to extract the rate constants (listed in table 5.3) and most likely trajectories. Figure 5.6(a) shows formation of one type of looped configuration is more sensitive to phasing than the other one and when two operators are 306 bp apart, looping is most likely to occur. Figure 5.6(b) displays the inter-conversion rate between two looped states and shows loop conversion from long tethered loop state to short tethered loop state is easier than the other way around.

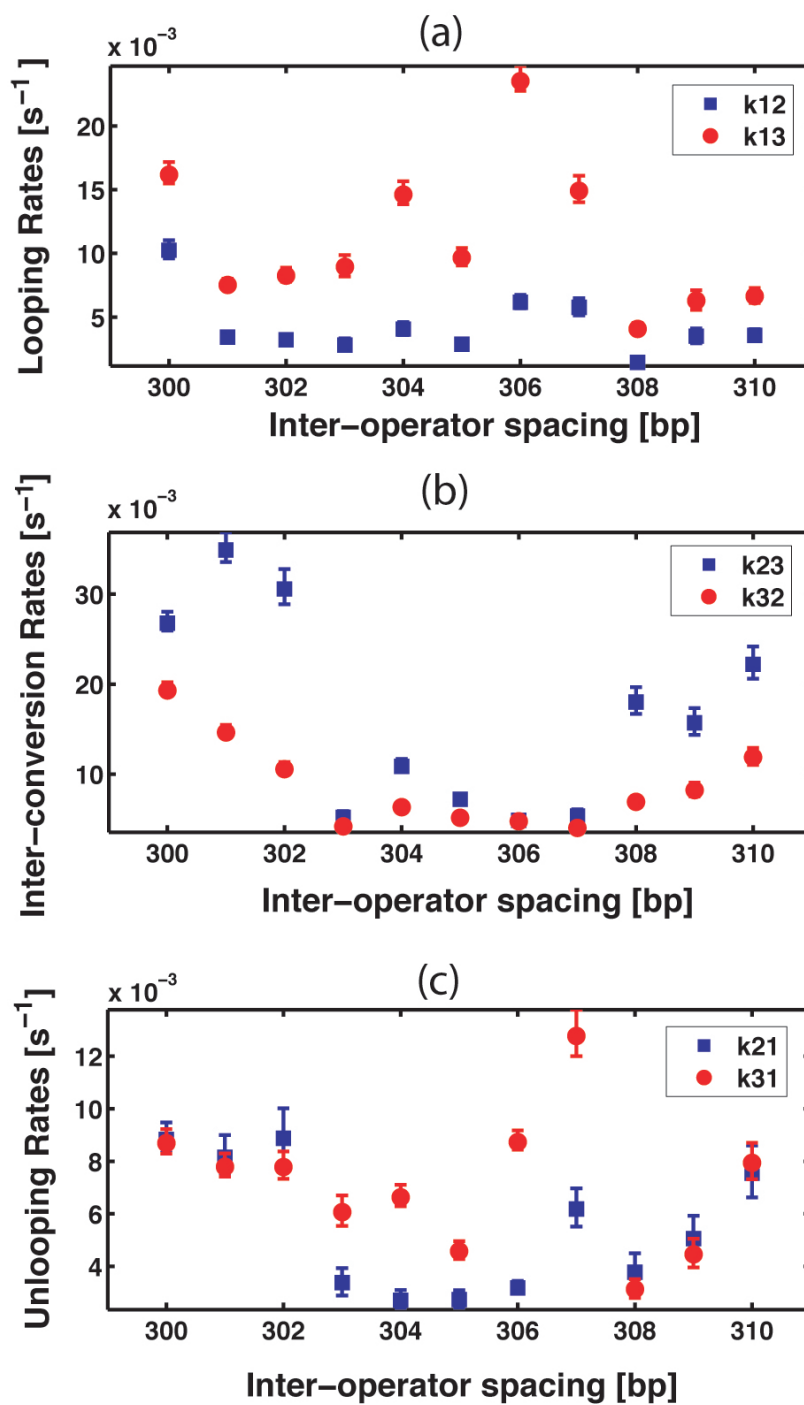


Figure 5.6: Rate constants as a function of inter-operator spacing. (a) Looping rates. (b) Inter-conversion rates between two looped states. (c) Unlooping rates.

Table 5.3: Rate constants. "1" denotes the unlooped state, "2" denotes the middle looped state and "3" denotes the bottom looped state. "kij" is the rate constant from state i to state j.

(bp)	( $10^{-3}/s$ )	( $10^{-3}/s$ )	( $10^{-3}/s$ )	( $10^{-3}/s$ )	( $10^{-3}/s$ )	( $10^{-3}/s$ )
Spacing	k12	k12 low	k12 high	k13	k13 low	k13 high
300	10.2412	0.6287	0.7880	16.1840	0.6981	0.9758
301	3.4270	0.3024	0.3571	7.5260	0.3708	0.4994
302	3.2154	0.3647	0.4190	8.2534	0.4851	0.6292
303	2.8115	0.4641	0.5172	8.9411	0.7504	0.9167
304	4.0633	0.4738	0.5520	14.6099	0.7600	1.0379
305	2.8659	0.3620	0.4120	9.6457	0.5913	0.7653
306	6.1641	0.4639	0.5740	23.4976	0.7464	1.1896
307	5.7658	0.6177	0.7260	14.9100	0.8925	1.1795
308	1.4481	0.2848	0.3091	4.0669	0.4087	0.4844
309	3.5038	0.5570	0.6197	6.2936	0.7137	0.8289
310	3.5675	0.4239	0.4857	6.6399	0.5334	0.6463
Spacing	k23	k23 low	k23 high	k32	k32 low	k32 high
300	26.7661	0.8235	1.2852	19.2905	0.5781	0.9144
301	34.8989	1.3365	1.9508	14.6422	0.5591	0.8180
302	30.5626	1.6879	2.2266	10.5507	0.6040	0.7782
303	5.1984	0.6034	0.7058	4.1944	0.4680	0.5415
304	10.8594	0.6384	0.8300	6.3154	0.3896	0.4904
305	7.2086	0.4843	0.6081	5.1586	0.3323	0.4271
306	4.8728	0.2557	0.3446	4.7786	0.2350	0.3201
307	5.3517	0.6543	0.7493	4.0464	0.5107	0.5785
308	18.0152	1.3277	1.6547	6.9280	0.5045	0.6312
309	15.7047	1.3449	1.6321	8.2158	0.7007	0.8473
310	22.1994	1.5958	1.9873	11.8713	0.8446	1.0413
Spacing	k21	k21 low	k21 high	k31	k31 low	k31 high
300	8.8315	0.5084	0.6457	8.6915	0.3947	0.5317
301	8.1579	0.7172	0.8402	7.7957	0.3758	0.5081
302	8.8724	1.0078	1.1393	7.7828	0.4498	0.5933
303	3.3829	0.4882	0.5481	6.0632	0.5229	0.6360
304	2.7096	0.3519	0.3937	6.6328	0.3453	0.4689
305	2.7370	0.3089	0.3565	4.5805	0.2986	0.3806
306	3.1763	0.2168	0.2694	8.7296	0.2882	0.4464
307	6.1880	0.6731	0.7814	12.7686	0.7621	1.0034
308	3.7811	0.6552	0.7173	3.1261	0.3272	0.3835
309	5.0592	0.7725	0.8714	4.4649	0.5040	0.5860
310	7.5472	0.9245	1.0532	7.9434	0.6262	0.7607

## Chapter 6

# Sequence Dependence of DNA Looping

Complexes of DNA and associated DNA-binding proteins are central to the mechanisms of transcriptional regulation in all organisms. Interestingly, the mechanical properties of DNA can play a critical role in dictating how such complexes are formed and on what time scale. One prime example is DNA looping in the context of transcriptional regulation. Recently, there has been a flurry of activity aimed at determining whether or not the short-length scale (i.e., smaller than the persistence length) mechanical properties of DNA are different than demanded by the wormlike chain model. To explore the connection between looping, transcriptional regulation and DNA mechanics, we have carried out measurements and calculations of the looping of short (<100 bp) DNA fragments in the *lac* operon. In this case, the Lac repressor molecule binds at two sites on the DNA and loops the intervening DNA. Using single-molecule techniques (tethered-particle motion) we have considered several different sequences with loop lengths less than 100 bp and find that these constructs are easily looped by Lac repressor and spend a much larger fraction of their time in the looped state than would be expected on the basis of the wormlike chain-type models of DNA elasticity. In particular, we find that no auxiliary DNA-bending proteins, nor supercoiling, are required to assist loop formation. In addition, we show that special nucleosome sequences that are thought to be highly bendable, loop more easily than random sequences of the same length when they are placed between the Lac repressor operators.

### 6.1 Introduction

There has been a surge of recent interest from many quarters concerning the length- and sequence-dependence of DNA mechanics and how it affects the ways in which organisms manage their genomes during DNA packing, repair and regulation. From a biological perspective, the mechan-

ics of DNA is one of the factors that governs the mechanism whereby complexes of transcription factors assemble on DNA and on what time scales. Indeed, assembly of complexes of activators, repressors and other helper molecules on DNA is one of the primary ways that cells “decide” when, where and how much to express specific genes.

DNA mechanics arises in a diverse array of important biological problems such as the looping of DNA in transcriptional regulation and the packing of DNA into nucleosomes (Garcia *et al.* 2007). One of the time-honored ideas in the physics of DNA is that of the persistence length. The persistence length is the length scale over which DNA is mechanically stiff and is roughly 150 bp in length (Rivetti *et al.* 1996). Stated differently, for DNA fragments shorter than 150 bp, there is a large energetic cost associated with bending the DNA. To the extent that this physical picture is correct, it raises interesting biological questions about how DNA-binding proteins such as transcription factors and repair enzymes are able to bring regions on the DNA with small genomic distances into close physical proximity.

Many biological processes involve bending DNA precisely at these short scales. For example, DNA looping by transcription factors in the *lac* operon involves loop formation with a loop length of 92 bp (Schleif 1992; Müller-Hill 1996). Similarly, the macromolecular assembly associated with packing DNA in the eukaryotic nucleus, the nucleosome, involves bending 147 bp  $1\frac{3}{4}$  times around an octamer of proteins known as the histone octamer. In both of these cases, the typical sizes of the resulting structures are sufficiently small that the conventional view of DNA elasticity holds that there is a high elastic energy cost associated with their formation. In this chapter we unite two seemingly disparate biological problems as seen through the prism of their similar *physical* features. In particular, we use DNA looping in the *lac* operon as a window onto the mechanics of short DNA fragments of different sequences. Single-molecule biophysics makes it possible to perform systematic and controlled studies of the response of short DNAs which are induced to loop by DNA-binding proteins such as the Lac repressor used here (Yin *et al.* 1994; Tolic-Norrelykke *et al.* 2004b; Blumberg *et al.* 2005; Pouget *et al.* 2004; Segall *et al.* 2006b; Nelson *et al.* 2006a; Zurla *et al.* 2006; Broek *et al.* 2006; Zurla *et al.* 2007).

The idea of the experiment is that a single molecule of DNA is anchored to a microscope cover slip with a large (i.e.,  $\approx 0.25 \mu\text{m}$ ) bead attached to the opposite end. The bead serves as a reporter of the dynamics of the underlying DNA molecule. In particular, when the DNA is looped through the binding of the DNA-binding protein, the tether length is reduced and the excursion of the bead is correspondingly reduced. By examining the trajectories of the bead over extended periods (i.e.,  $>$



1000 s), it is possible to monitor the dynamics of the looping process. The goal of our experiments was to examine two intriguing questions concerning the interplay between the formation of repression complexes and DNA mechanics: (i) are short DNA fragments more easily bent than expected on the basis of conventional wormlike chain-type models of DNA elasticity physics and does this aid in the formation of repression complexes and (ii) does DNA sequence strongly influence the likelihood of DNA looping, e.g., by reducing the mechanical stiffness of the DNA molecule? The answers to these questions matter both physically and biologically. For example, though there is a growing understanding of the combinatorial logic of cis-regulatory regions in transcription, the mechanisms by which transcription factors bind and bend DNA to form regulatory complexes remain largely unknown.

From a physical perspective, recent work on both experimental and theoretical fronts has suggested that short DNA is more flexible than expected on the basis of the wormlike chain model (Cloutier and Widom 2004; Yan and Marko 2004; Cloutier and Widom 2005; Yan *et al.* 2005; Wiggins and Nelson 2006; Shroff *et al.* 2005; Wiggins *et al.* 2006). From a biological perspective, the question of mechanical stiffness of DNA arises from studies of nucleosomes which demonstrate sequence preferences for which parts of a given DNA molecule will wrap around the histone octamer. Further, *in vitro* cyclization measurements (Cloutier and Widom 2004, 2005), *in vitro* nucleosome assembly experiments and systematic studies of sequence preferences in nucleosomes in yeast and other organisms suggest that these preferences are a reflection of the enhanced bendability of certain special sequences. Nevertheless, some of these ideas remain controversial, in part because of experimental complexities inherent to cyclization measurements. The aim of our experiments and associated calculations is to employ a completely different approach from those used so far (in particular, a single-molecule experiment) to tackle the question of the short-length scale mechanical properties of DNA and how they relate to protein-DNA complexes.

In this chapter we employ the tethered particle motion technique to systematically explore the differences between the mechanical properties of nucleosomal and random DNAs. We also show that Lac repressor can loop 92 bp worth of DNA from the wild type *lac* operon without any other proteins involved (e.g., without the assistance of CAP or another DNA-bending protein). We then demonstrate that nucleosomal DNAs are even more readily looped than their random counterparts.

## 6.2 Results

### 6.2.1 Formation of the Wild-type Loop

There are three binding sites for Lac repressor in the wild type *lac* operon:  $O_1$ ,  $O_2$  and  $O_3$ . These three binding sites give rise to two different loops that are relevant for transcriptional regulation, both of them including  $O_1$ . Operators  $O_1$  and  $O_2$  are separated by 401 bp whereas the operators  $O_1$  and  $O_3$  are only separated by 92 bp (Müller-Hill 1996), much shorter than the DNA persistence length. As part of the suite of experiments we have done to probe the length- and sequence-dependence of DNA mechanics, we have examined looping of DNA by Lac repressor for the wild-type interoperator spacing of 92 bp. Our rationale for examining this case was to see if *in vitro* looping experiments might shed some light on the question of whether the maximum in looping probability observed in *in vivo* studies (i.e.,  $< 80$  bp) is surprising or not. It has been speculated, for example, that the ability to form short-length loops *in vivo* is dependent on the presence of accessory proteins or supercoiling (Garcia *et al.* 2007). The results of our TPM experiments on the wild-type interoperator spacing are shown in figure 6.1. In particular, figure 6.1(a) shows several TPM trajectories for this construct in the presence of Lac repressor. This figure reveals trajectories characterized by a number of transitions between the unlooped and looped states. Several controls were run as part of an attempt to make sure that these putative looping states are indeed due to looping. For example, figure 6.1(b) shows the results of a no-protein control in which the excursions of the bead were monitored in the absence of Lac repressor protein. A second reassuring control involves the inducer IPTG which has the effect of significantly reducing the affinity of Lac repressor for DNA as shown in figure 6.1(c).

To probe the meaning of these results, it is important to extract their quantitative implications. Fig. 6.2 shows a histogram from the result of many trajectories with the wild-type interoperator spacing. By fitting this histogram to two Gaussians corresponding to the excursions of the bead when in the looped and unlooped states, it is possible to reckon the looping probability itself, and from that, the looping energy. The basis of the calculation of the looping free energy is to use a statistical mechanics model in which the various allowed states of the system are each assigned their appropriate Boltzmann weight. These weights involve binding energies and, for the looped state, the looping free energy as well.

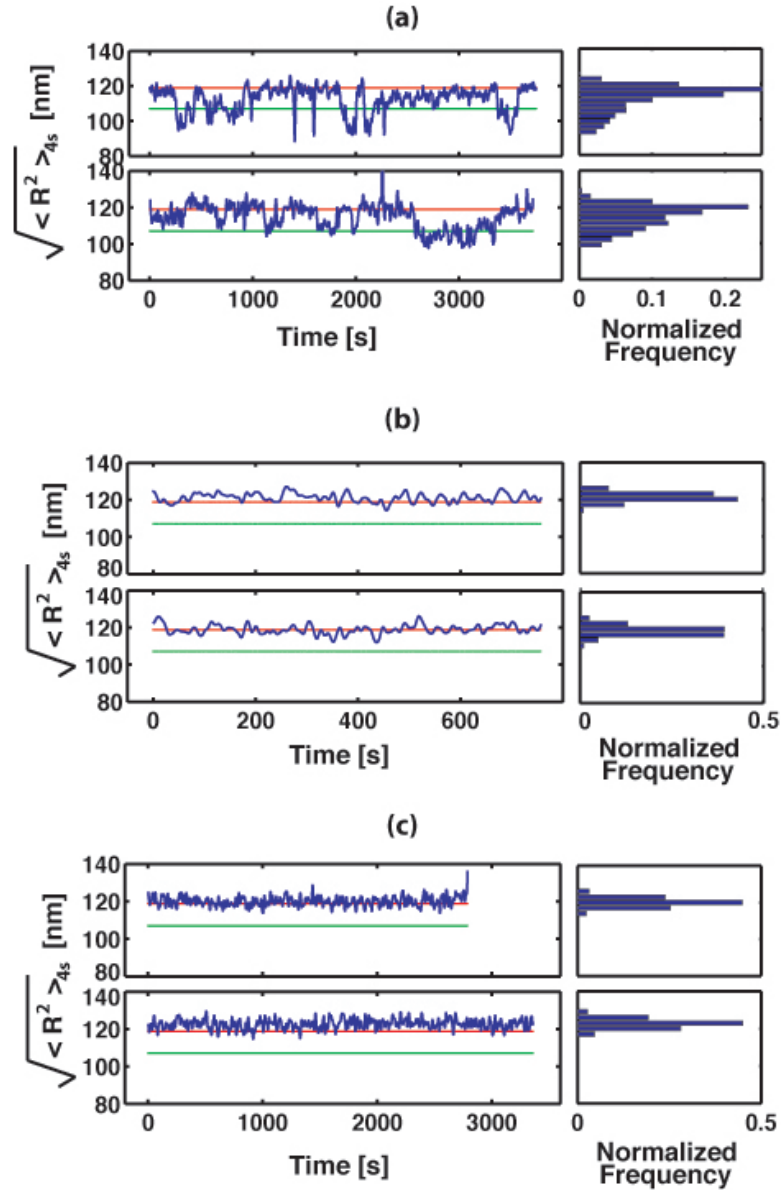


Figure 6.1: Typical TPM recording for wild type interoperator spacing found in the *lac* operon. (a) In the presence of Lac repressor, looping occurs as indicated both by the trajectories and the histograms. (b) Control experiment in which there is no Lac repressor present. (c) Control experiment using both IPTG and Lac repressor. The red lines correspond to the expected excursions for the unlooped state and the green line corresponds to the “expected” excursion for a DNA fragment from which the DNA fragment between the two operators has been removed. The DNA construct is 351 bp in length with wild type inter-operator DNA between binding sites  $O_{id}$  and  $O_1$ .

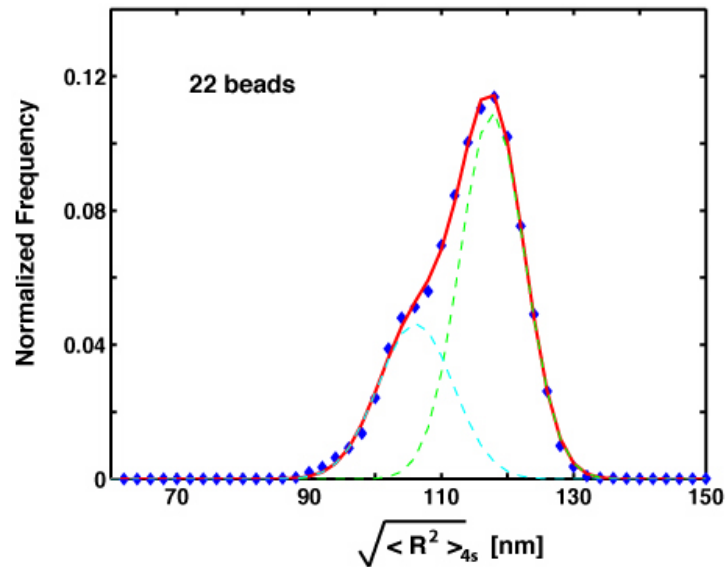


Figure 6.2: Histogram of Brownian motion distribution for wild-type interoperator spacing. The histogram is fit to a sum of two Gaussians. The looping free energy can be calculated based on the thermodynamic model.

### 6.2.2 Sequence Dependence

There is evidence from many different fronts that DNA sequence plays a key role in determining the stability of nucleosomes. The favorability of some sequences relative to others has been ascribed to the fact that the more favorable sequences are elastically softer, making the energy cost to bend the DNA around the histone octamer less severe (Cloutier and Widom 2004). To test the flexibility of such sequences in a totally different context, we have taken nucleosomal positioning sequences and installed them in the *lac* operon, i.e. between the two operator sites. Our reasoning was that if these special sequences are more easily bent, this should be revealed by looping of these sequences by the Lac repressor.

The random sequences E8 and nucleosomal positioning sequences TA used here are the same in both sequence and length as those used in previous DNA cyclization experiments (Cloutier and Widom 2004, 2005). In the cyclization experiment, the  $J$  factor was measured for DNA with length varying from 89 bp to 105 bp. These results showed 94 bp gives the local optimal length for both random and special sequences. In addition, 89 bp and 100 bp corresponds to the local minimum in the cyclization free energy. As a result, we postulated that when such DNAs are inserted between the two binding sites for Lac repressor, the two operators are either in-phase (94 bp) or out-of-phase

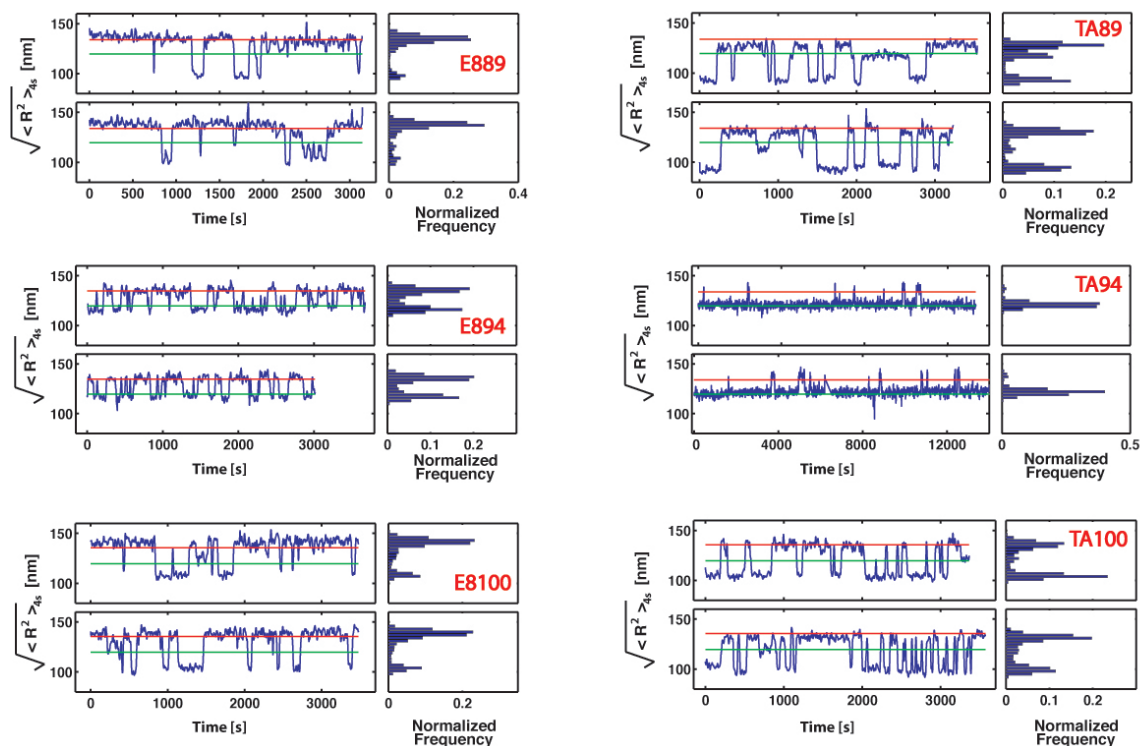


Figure 6.3: Typical TPM recordings for different sequences and different lengths. The sequences labeled E8 are random sequences and those labeled TA are nucleosomal positioning sequences. The integer after the label refers to the length of the fragment between the two binding sites. For each sequence and length, two recordings are shown.

(89 bp, 100 bp). An example of the kind of data that we have collected for a random and nucleosomal sequence for several different interoperator spacings is shown in figure 6.3. As is clear both from the trajectories and the histograms, the TA sequence (i.e., the nucleosomal sequence) loops more easily, at least for the case when the binding sites are in-phase,

A second way to examine the character of these short DNA loops is to compare the histograms from the different lengths and sequences as shown in figure 6.4. The three sets of histograms correspond to three different interoperator spacings which permit an exploration of the phasing effect as the operators go in and out of helical registry. Interestingly, the 94 bp loop reveals that the nucleosomal positioning sequences are much more inclined to loop than random sequences of the same length. Another interesting observation that results from this figure is that the favored loop state is different depending upon the phasing. The presence of two different looped configurations jibes with earlier experiments. Two different looped configurations induced by Lac repressor have been observed in TPM experiments like those described here (Wong *et al.* 2007; Han *et al.* 2007).

Further, multiple looped configurations have been argued to arise as the consequence of the large scale conformational change of the Lac repressor protein, i.e., V shaped to Extended form (Ruben and Roos 1997).

### 6.2.3 Theoretical Analysis

Both the observed length and sequence dependence of the formation of a repression complex are intriguing from the perspective of DNA mechanics. To better understand these results, we have appealed to two classes of models: i) statistical mechanics models of the probability of DNA-repressor complexes which depend upon the looping free energy and ii) Monte Carlo simulations of the TPM experiment itself which include the energetics of the bent DNA and excluded volume interactions of the bead with the cover slip. The statistical mechanics models are described briefly in the Supplementary Material and have been worked out in detail elsewhere. The exciting outcome of these models in conjunction with the experimental results presented here is that they permit a direct calculation of the length and sequence dependence of the looping free energies.

The looping free energy difference,  $\Delta\Delta G_{loop}$ , for a 94 bp loop between a random sequence and a nucleosomal positioning sequence is comparable to that found in cyclization experiments (Cloutier and Widom 2004, 2005) (table 6.1). The decrease of this quantity by  $1 k_B T$  compared to the cyclization results can be attributed to the size of the protein bridge (Douarche and Cocco 2005). Additionally, the magnitude of the motion upon looping corresponds to the effective tether length. This suggests that the favored in-phase loop is a closed loop where the Lac repressor protein adopts a V-shape.

When half a helical repeat is added to the system, the difference in looping energy between the two sequences becomes much smaller than the difference found in cyclization experiments. Given the same loop geometry as in the in-phase case a similar difference between the two sequences would be expected. However, this does not have to hold true if there is a different favored configuration. Both figure 6.4 and figure 6.3 clearly show that the favored out-of-phase looped configuration is different from the in-phase looped configuration. Together with the implication that random sequence and nucleosomal sequence have similar twist flexibilities (Cloutier and Widom 2005), it suggests that bendability does not play a significant role in the favored out-phase loop. We then conclude that this favored loop is induced by an extended conformation of Lac repressor where DNA is much less bent than the one induced by V-shaped Lac repressor. Although favored looped configurations are dependent on phasing, the unfavored type of looped configurations is also

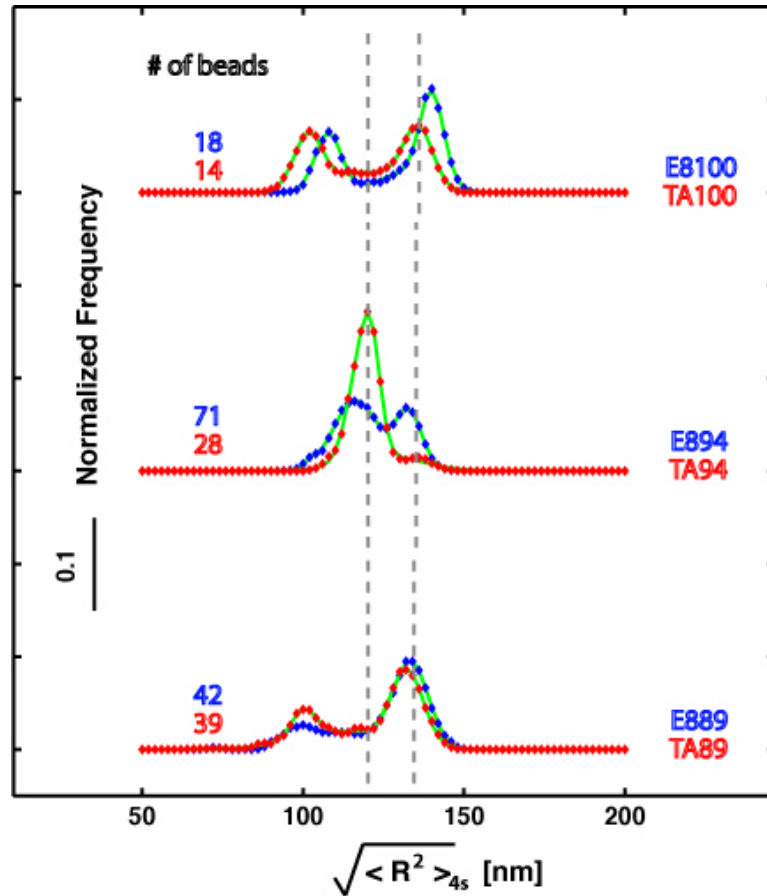


Figure 6.4: Histogram of Brownian motion distribution of different sequences, as well as when binding sites are in and out of phase. Grey dash lines are expected motion from calibration data.

occasionally observed. The nucleosomal positioning sequence DNAs spend relatively more time on the unfavored looped state than the random sequence when it is out of phase. On the other side, the random sequence DNAs spend relatively more time on the unfavored looped state when it is in phase. This result suggests the flexibility of Lac repressor protein turn to adopt the low-energy cost looped structure by a large conformational change.

One of the puzzles that has so far been unresolved concerning DNA mechanics at these short scales is whether or not *in vivo* and *in vitro* experiments tell a different story. In particular, *in vivo* DNA mechanics experiments in which repression of a given gene is measured as a function of the interoperator spacing (Muller *et al.* 1996; Becker *et al.* 2005) have the provocative feature that the maximum in repression (corresponding to the minimum in looping free energy) correspond to interoperator spacings that are shorter than the persistence length. Some speculate that this *in vivo* behavior results from the binding of helper proteins such as the architectural protein HU (Becker

*et al.* 2005; Garcia *et al.* 2007) or the control of DNA topology through the accumulation of twist. In the TPM measurements reported here, there are neither architectural proteins nor proteins that control the twist of the DNA. As a result, these experimental results serve as a jumping off point for a quantitative investigation of whether DNA at length scales shorter than the persistence length behaves more flexibly than expected on the basis of the wormlike chain model. To address this question, we performed a series of simulations of the probability of DNA looping for short, tethered DNAs like those described here using a variant of the wormlike chain model to investigate the looping probability.

The fraction of time spent in the looped configuration is controlled by several competing effects. For example, suppose that a repressor tetramer is bound to the stronger operator,  $O_{id}$ . Shortening the interoperator spacing reduces the volume over which the other operator ( $O_1$ ) wanders relative to the second binding site on the repressor, increases the apparent local "concentration" of free operator in the neighborhood of that binding site, and hence enhances looping. But decreasing the interoperator spacing also has the opposite effect of discouraging looping, due to the larger elastic energy cost of forming a shorter loop. Moreover, a shorter overall DNA construct increases the entropic force exerted by bead-wall avoidance, again discouraging looping (Segall *et al.* 2006b). To see what our measurement of this looping equilibrium tells us, we therefore needed to calculate in some detail the expected local concentration of operator (the "looping  $J$  factor") based on a particular mathematical model of DNA elasticity. We chose a harmonic-elasticity model (a generalization of the traditional wormlike chain model), to see if it could adequately explain our results, or if, on the contrary, some non-harmonic model (for example the one proposed in (Yan and Marko 2004; Wiggins *et al.* 2005)) might be indicated.

To perform the required calculation, we modified the Gaussian sampling method previously used in (Segall *et al.* 2006b; Nelson *et al.* 2006a; Czapla *et al.* 2006). Our code generated many simulated DNA chains, applied steric constraints (Segall *et al.* 2006b), and reported what fraction of accepted chain/bead configurations had the two operator sites separated by 7 nm, the distance between operator centers as .jpgn in PDB structure 1LBG (Lewis *et al.* 1996). The standard elastic model as an isotropic rod is inadequate for the description of DNA loops only a few helical repeats in length (see, for instance Czapla *et al.* (2006)), so we modified the elasticity to account for bend anisotropy and bend-roll coupling. We did not account for sequence dependence, however, so we can only make comparisons to our experimental results with random-sequence DNA. We adjusted the overall magnitude of our DNA elasticity matrix to yield a value of overall persistence length



$\xi = 45$  nm appropriate for our experiment's buffer conditions (Strick *et al.* 1998). The chain generation accounted for bead-wall, bead-chain, and wall-chain avoidance, but not chain-chain; nor did we consider any interactions involving the repressor tetramer other than binding.

The result of the simulation was that the looping  $J$  factor for short loops was 0.015 times as great for the constructs with interoperator spacing around 100 bp as for those with spacing around 300 bp; this ratio was about a hundred times smaller than the experimentally determined ratio of  $J(100 \text{ bp})/J(300 \text{ bp}) \approx 1.7$  at optimum helical phasing (Han *et al.* 2007).

We conclude that the hypotheses of harmonic elasticity, plus a rigid V-shaped protein coupler, cannot explain our experimental results. One possible explanation, for which other support has been growing, is the hypothesis of DNA elastic breakdown at high curvature (Yan and Marko 2004; Wiggins *et al.* 2005). Alternatively, the LacI tetramer may spend an appreciable fraction of its time in an alternate ("open") conformation, where the operator binding sites are separated by more than 7 nm. To determine which of these mechanisms (or both) is at work would require a detailed understanding of the free energy change of LacI opening, or new experiments using a coupler known to be inflexible. We note, however, that in our measurements we can often discern a separation of the looped-state histogram peak into two distinct subpeaks, as we and others have already found with longer interoperator spacings (Wong *et al.* 2007; Han *et al.* 2007). This preliminary observation suggests that both the "open" and the "closed" conformations are appreciably populated, in contradiction with our theoretical calculation based on harmonic DNA elasticity.

Table 6.1: Looping free energy.  $\Delta\Delta G_{loop}$  is the looping free energy difference between nucleosomal positioning sequence and random sequence. Looping free energy  $\Delta G_{loop}$  is calculated from equation 6.1, where  $Z_{tot}$  is the normalization factor (the partition function) and given by equation 6.3, binding energy  $\Delta\epsilon_{id} = -28.0 \pm 7.3 k_B T$ ,  $\Delta\epsilon_1 = -21.3 \pm 2.8 k_B T$  (Han *et al.* 2007). The minimum between the two Gaussian peaks is chosen to be the threshold between unlooped and unlooped states.  $P_{loop}$  is then calculated by dividing the area belonging to looped states by the total area under the Gaussian curve. ”\*” indicates the numbers are obtained from Widom’s experiments.

Molecule	Looping free energy $\Delta G_{loop} k_B T$			
	Spacing (bp)	89	94	100
E8		21.69	20.45	20.83
TA		21.55	18.57	20.67
$\Delta\Delta G_{loop}$		0.15	1.90	0.16
$\Delta\Delta G_{loop}^*$		2.30	2.99	1.97

$$P_{loop} = \frac{[R] \exp [-(\Delta\epsilon_{id} + \Delta\epsilon_1 + \Delta G_{loop})/k_B T]}{Z_{tot}}, \quad (6.1)$$

$$Z_{tot} = 1 + [R] \exp [-(\Delta\epsilon_{id})/k_B T] + [R] \exp [-(\Delta\epsilon_1)/k_B T] + [R] \exp [-(\Delta\epsilon_{id} + \Delta\epsilon_1 + \Delta G_{loop})/k_B T] + [R]^2 \exp [-(\Delta\epsilon_{id} + \Delta\epsilon_1)/k_B T]. \quad (6.2)$$

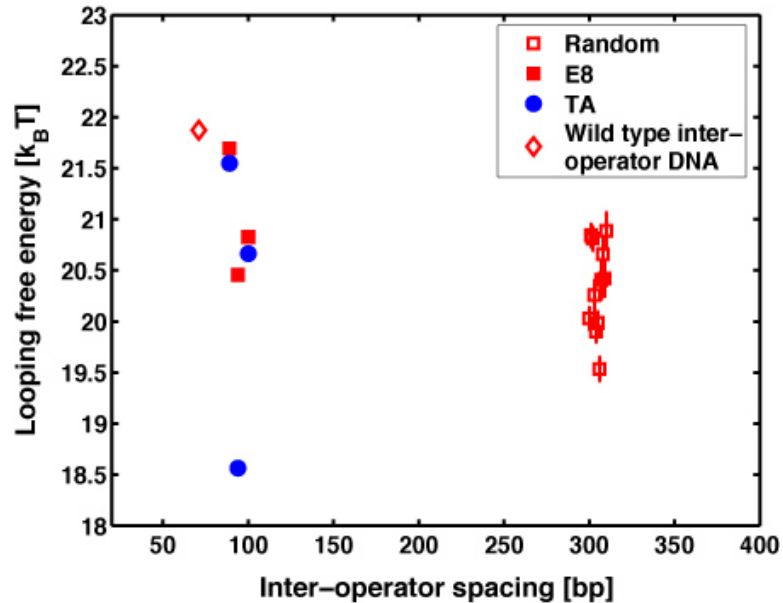


Figure 6.5: Looping free energy as a function of inter-operator spacing. Distance between the two binding sites is referred as inter-operator spacing rather than the operator center to center distance.

## Chapter 7

# Conclusion and Future Work

### 7.1 Conclusions

The motivation for this thesis has been to systematically explore the role of sharply bent DNA in gene regulation. We choose as our model system, Lac repressor (LacI)-mediated loop formation. LacI is a widely studied protein found in *Escherichia coli*, although looping is a ubiquitous motif in gene regulation. The workhorse of our experimental effort has been a single DNA molecule tracking method called tethered particle motion (TPM), in which one end is attached to a glass coverslip and the other to a polystyrene microsphere reporter particle. Brownian motion of the particle reveals the underlying conformational changes of DNA. This technique is sensitive enough to detect nanometer-scale displacement changes accompanying looping/unlooping events, which register as "jumps" in Brownian motion. By monitoring parameters such as the amplitude, frequency, and time interval of these features one can survey a host of important information about looping kinetics.

The main accomplishments and conclusions of this work are summarized as follows.

(i) In our constant struggle against unwanted measurement noise, we have obtained a calibration curve to optimize our ability to detect different sized loops. Contrary to what one might naively expect, the Brownian motion magnitude is not linearly proportional to DNA length. Thus by referring to the calibration curve we can reliably translate between the looping-mediated jumps in observed motion and corresponding change in effective tether length. Moreover, we have gone one step further and repeated the calibration for a total of three different sized polystyrene beads. Ultimately we used the intermediate (490 nm) beads, as these were a reasonable compromise between spatial and temporal resolution.

(ii) LacI is a tetramer capable of simultaneously binding to two sites, thereby looping the intervening DNA. At very low concentrations, we expect that there will be negligible looping. At

intermediate concentrations, the equilibrium situation will be dominated by states in which a single repressor molecule is bound to the DNA at the higher affinity operator,  $O_{id}$  in our case, punctuated by transient looping events. In the very high concentration limit, both operators will be occupied by a separate repressor, making the formation of a loop nearly impossible. We observe this type of concentration dependence in our experiments, and find it to be in quantitative agreement with a simple statistical mechanics model of looping probability. We were able to calculate the free energy associated with looping, having first subtracted the contribution of LacI-DNA binding, obtaining its first experimental measurement in a single molecule, *in vitro*, setting.

(iii) We carried out the first systematic and quantitative characterization of the free energy of looping in a single molecule *in vitro* environment. The interoperator spacing was varied from 300 to 310 bp in 1 bp steps. Our results show that free energy is modulated by DNA's helical structure, and are consistent with analogous *in vitro* work on cyclization (Cloutier and Widom 2004, 2005), and *in vivo* work on repression (Muller *et al.* 1996; Becker *et al.* 2005). In a striking departure from classical elasticity theory, we find that the looping energy difference between the case of aligned operators (i.e. they face the same side of the helical structure) and unaligned operators is quite small.

(iv) DNA looping is usually considered to be a two step (i.e., "on/off") process. However, we obtained unambiguous experimental evidence for the existence of an additional looped state, which is in agreement with theoretical predictions (Zhang *et al.* 2006; Swigon *et al.* 2006; Geanacopoulos *et al.* 2001; Balaeff *et al.* 2006) and other recent experiments (Wong *et al.* 2007). The two looped configurations are expected to be the consequence of two distinct conformations of LacI, namely, the closed (V-shaped), and open form.

(v) Certain DNA sequences are more flexible than others, directly supporting the increasing amount of evidence that DNA elasticity is sequence-dependent. In particular, we confirmed that the periodic presence of TA pairs in the minor groove of the nucleosome positioning sequence makes this substantially softer than random sequences: we measured a difference of  $\approx 2k_B T$  in the free energy of looping.

(vi) DNA *in vitro* spontaneously forms loops as short as 100 base pairs, mediated by LacI. This value is remarkable for two reasons: first, it is only two thirds the persistence length of DNA, which is almost forbidden by classical elasticity theory; second, it contradicts previous speculation that such sized loops can only form *in vivo* with the aid of structural proteins or supercoiling properties. This result allows us to hypothesize that either LacI can undergo large conformational changes to

adopt different DNA topologies, or that our classical understanding of DNA elasticity breaks down at length scales comparable to its persistence length.

## 7.2 Future Directions

I would like to close with my thoughts on what are the next important questions we can ask, after having obtained some interesting results about Lac repressor mediated DNA looping. First, the LacI looping system definitely deserves to be studied further. Our results have hit upon several fascinating effects that I would love to spend another five years of my life investigating (but I am generous so I will leave the opportunity to others). I am especially curious about the two looped structures we observed, the closed and open forms, and feel that it would be worth probing further. For instance, one can imagine performing a mutation that will strengthen the interaction between the two LacI dimers. This should lock the LacI into the V-shaped form. If this prediction is correct, then we can further hypothesize that a TPM experiment will reveal only one looped state. Furthermore, by systematically increasing the rigidity of LacI we can investigate the role of protein flexibility in loop formation, particularly small loops comparable to the persistence length. Another possibility for future research that follows up on what we have already done is to explore the effect of adding structural proteins such as HU, IHF to the TPM measurement, and observing changes in looping kinetics. It would also be interesting to consider supercoiled DNA in the context of looping. Furthermore, we can measure looping in the presence of another popular protein, AraC, which only forms loops with supercoiled DNA. A systematic study of these factors might help us gain a better understanding of DNA mechanics *in vivo*.

Perhaps the ultimate mystery of looping lies in its operation in eukaryotic organisms. This would involve studying a more complex set of interactions between DNA with multiple binding sites and transcription factors. To give a sense of this complexity, consider that a 2000 bp DNA contains more than 10 binding sites for a single protein that mediates looping (?). All sites do not need to be occupied at the same time, and in fact gene expression is regulated by loop formation between a particular pair of these sites. The sheer number of conformational states makes experiments on this system a daunting challenge.

## Appendix A

# Materials and Methods

### A.1 Construction of Plasmids with Two LacI Binding Sites for Length Dependence Measurements

DNA used in the looping project requires two Lac repressor (LacI) binding sites spaced a certain distance apart. To do so, point mutations are applied using QuickChange site-direct mutagenesis kits (Stratagene). The basic procedure utilizes a supercoiled dsDNA template plasmid with an insert of interest and two synthetic oligonucleotide primers containing desired mutations. The two primers, complementary to opposite strands of the vector plasmid, are extended in a PCR similar reaction. The primers are designed in such way that mutations are located in the middle of the primers and around 15bp complementary to the template plasmid at each side. Incorporation of the primers generates a mutated plasmid containing staggered nicks. Then DpnI restriction reaction is followed to digest the parent template DNA which is methyolated and select for mutation-containing synthesized DNA. Plasmid pUC19 was chosen here as a starting template because it is not only a high copy plasmid but also contains two LacI binding sites:  $O_1$  and  $O_3$ . We first mutate a few basepairs in the  $O_3$  site such that it will eliminate the binding affinity for this site. The resulting plasmid is called pUC19O1 indicating it only has one single  $O_1$  site. To construct another binding site on the pUC19O1 plasmid, we replace 20bp with the LacI binding sequence  $O_{id}$  at a series of locations separated by 1bp with desired distance from  $O_1$  using the mutagenesis method again. For some of the secondary site construction, we have to use either deletion or addition from already made plasmids with two designed binding sites. The details on primers and templates used in this process are listed in figure A.2. The final product contains two binding sites  $O_1$  and  $O_{id}$  spaced at designed distance. figure A.1 lists all the operator sequences involved in the experiments. Figure

Table A.1: Operator sequences.  $O_{3*}$  is the mutation from  $O_3$ , who doesn't have any affinity to LacI.

Operator	Sequence
$O_{id}$	AATTGTGAGCGCTCACAATT
$O_1$	AATTGTTATCCGCTCACAATT
$O_3$	AATTGCGTTGCGCTCACTGCC
$O_{3*}$	AATTGCGTTGAGCTCGAGGTT

A.1 (a) illustrates the procedure for creating two binding sites separated by the desired spacing.

## A.2 Construction of Plasmids with Two LacI Binding Sites for Sequence Dependence Measurements

Plasmid pZS22-YFP was kindly provided by Michael Elowitz. The main features of the pZ plasmids are located between unique restriction sites (Lutz and Bujard 1997). The YFP gene comes from plasmid pDH5 (University of Washington Yeast Resource Center (Rosenfeld *et al.* 2005)).

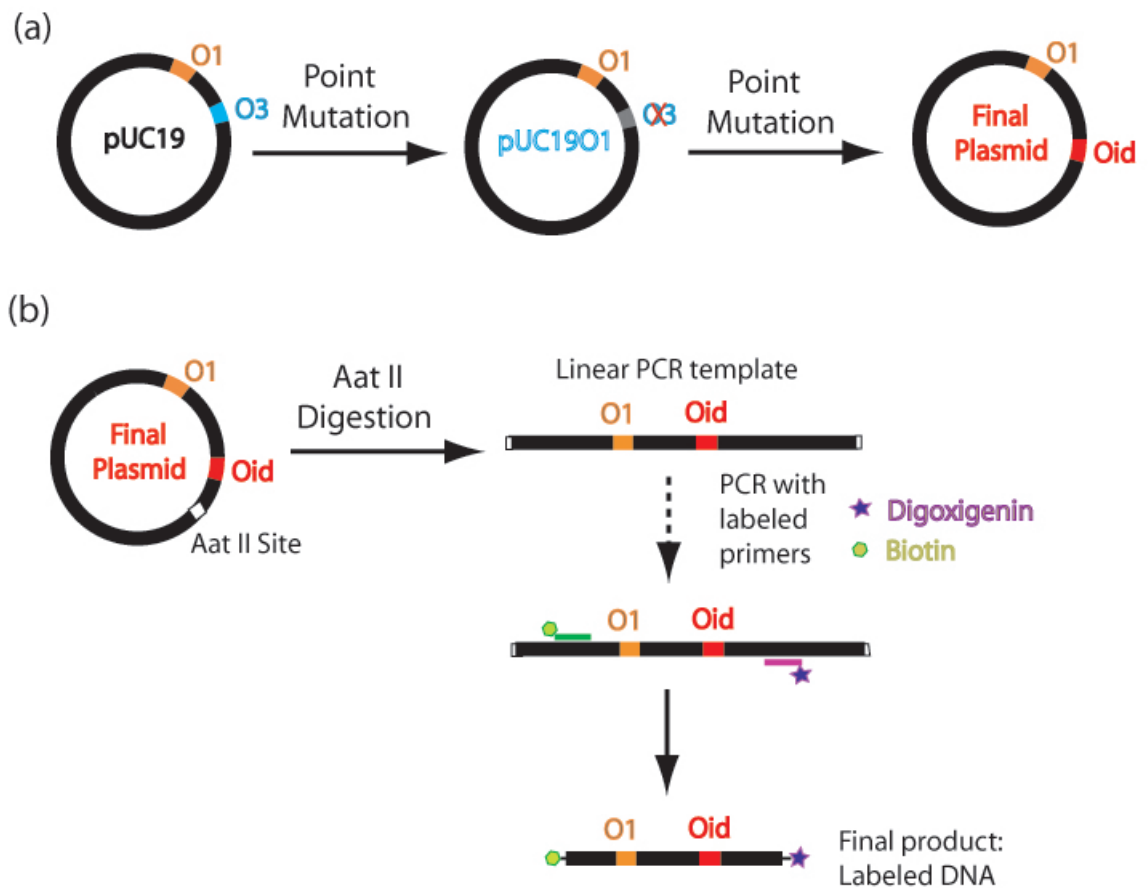
A variant of the lacUV5 promoter (Müller-Hill 1996) was created by annealing complimentary oligos and placing the resulting molecule between the EcoRI and XhoI sites of pZS22-YFP in order to create pZS25-YFP. This promoter included the -35 and -10 regions of the lacUV5 promoter, an AseI site between the two signals and an  $O_1$  operator at position -45 as shown in figure A.2a.

Sequences E8-89 and TA-89 (Cloutier and Widom 2004, 2005) were obtained by PCR using primers with a flanking AatII site and  $O_{id}$  operator upstream and with a flanking  $O_1$  operator, -35 region and AseI site downstream. A plasmid where any DNA sequence can be integrated between a  $O_{id}$  and  $O_1$  operators was then created by replacing the AatII-AseI region of pZS25-YFP with the PCR product as seen in figure A.2b. For the two 89 base pair sequences this gave rise to pZS25.E89-YFP and pZS25.T89-YFP.

Finally, the different lengths used by Cloutier and Widom were generated from these templates using site directed mutagenesis (Quikchange II, Stratagene).

## A.3 Construction of Labeled DNAs

In TPM experiments, DNA is linked between the substrate and a bead. Two pairs of linkers: biotin-streptavidin and digoxigenine-anti-digoxigenine, are chosen to permit specific linkage of the DNA





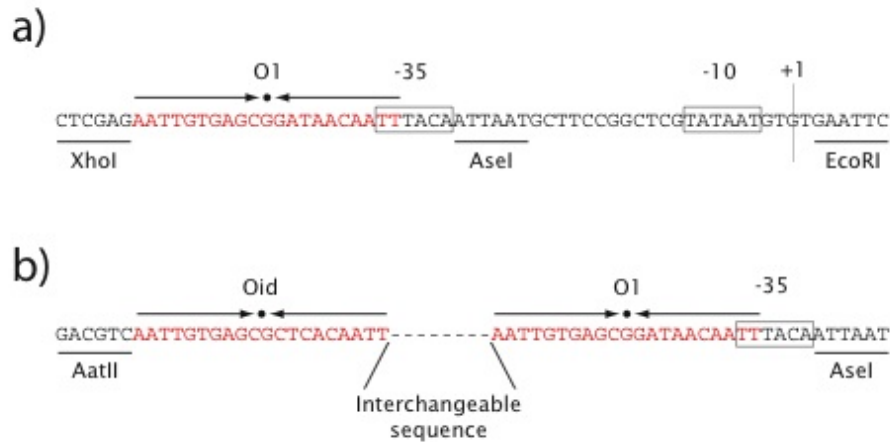


Figure A.2: Promoter regions of the different constructs. (a) Promoter region of pZS25-YFP which has a variant of the lacUV5 promoter and an  $O_1$  operator upstream overlapping the -35 region. (b) Final construct that allows to insert arbitrary DNA sequences between a  $O_{id}$  and  $O_1$  operators.

to a polystyrene microsphere and glass coverslip, respectively. Polymerase Chain Reaction (PCR) was used to amplify such labeled DNA with two modified primers. Each primer is designed to be about 20bp in length and linked with either biotin or digoxigenine at 5' end (MWG). To optimize the PCR reaction, the linearized plasmids with Aat II cut are used as the templates. Detailed information of PCR is listed in table A.3. The PCR products were then purified by gel extraction (QIAquick Gel Extraction Kit, QIAGEN) and the concentration of the DNA were measured using quantitative DNA electrophoresis. Construction of labeled DNA is shown in figure A.1 (b).

#### A.4 TPM Sample Preparation

TPM sample preparation involves the construction of the relevant DNA tethers with their associated reporter beads. Streptavidin (Bangs lab) or neutravidin (Molecular Probes) coated microspheres of diameter 200, 490 and 970 nm served as our tethered particle. In contrast to the 490 and 970 nm microspheres, the 200 nm microspheres were fluorescent. Prior to the incubation with DNA, the beads were sonicated for about 5 minutes to avoid agglomeration. Then a buffer exchange on the beads was performed by three cycles of centrifugation and resuspension in TPB buffer. TPB consisted of 20 mM Tris-acetate, pH=8.0, 130 mM KCl, 4 mM  $MgCl_2$ , 0.1 mM DTT, 0.1 mM EDTA, 20  $\mu$ g/ml acetylated BSA (Sigma-Aldrich), 80  $\mu$ g/ml heparin(Sigma-Aldrich)) and 3 mg/ml casein (Sigma), filtered with 300 kD MWCO polysulfone membrane (Milipore). This combination of reagents

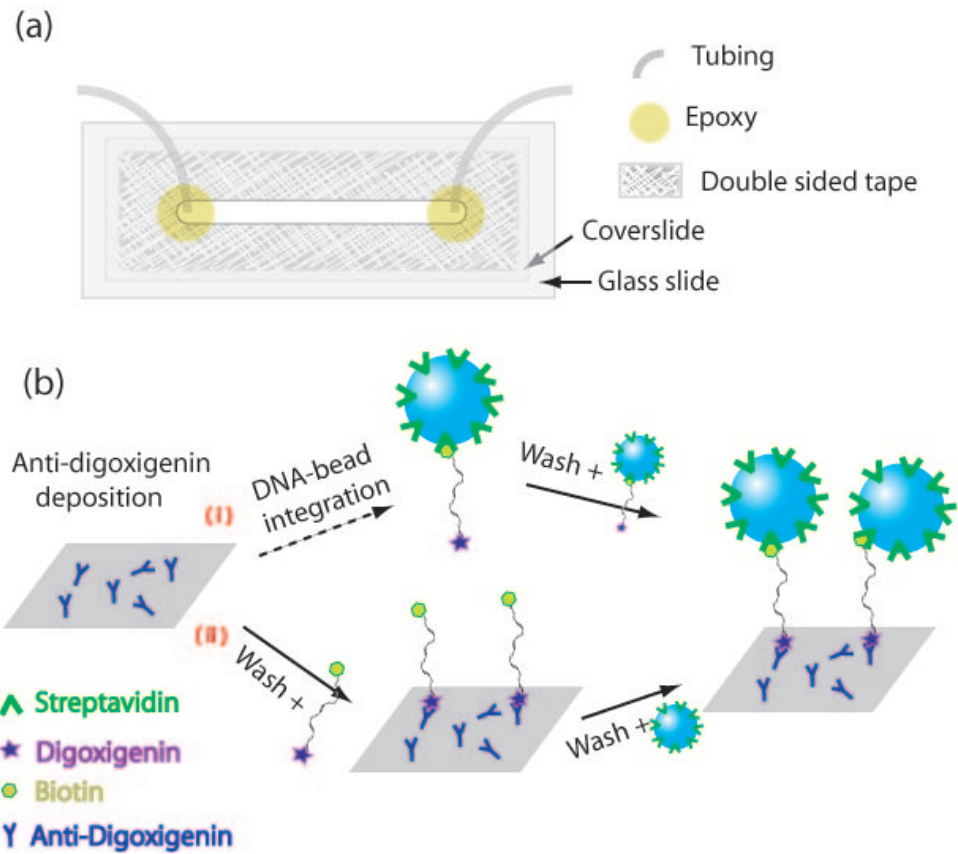


Figure A.3: Illustration of TPM sample preparation. (a) Sketch of the flow cell. (b) Two schemes of making DNA tethers.

was chosen in an attempt to maximize sample yield and longevity, while minimizing non-specific adsorption of DNA and microspheres onto the coverslip.

Tethered particle samples were created inside a 20-30  $\mu\text{l}$  flow cell made out of a glass slide with one hole near each end, glass coverslip, double-sided tape and tygon tubing. The coverslip and glass slide were cleaned either with 4N HCl for 24 hours or with plasma cleaning for 5 minutes and then the flow cell was constructed in a way shown in the figure A.3 (a). Two tygon tubings served as an input and output were inserted into the holes on the glass slide and sealed with epoxy. A reaction chamber was created by cutting a channel on the double sided tape, which glues the coverslip and glassslide together. Making the end of the channel round and as close to the holes of the glass slide as possible is important to avoid generating bubbles. The flow cell was then heated for about 20 seconds to seal securely.

We used two different schemes for DNA tether assembly. One requires to attach the bead to the DNA first, then anchors into the glass substrate. The other one fixes the DNA on the glass slide first, then attaches bead onto it. The experimental results showed there is no significant difference in terms of tether dynamics. The whole procedure of this DNA tether assembly processes is shown in the figure A.3 (b). First the flow chamber was incubated with 20  $\mu\text{g}/\text{mg}$  polyclonal anti-digoxigenin (Roche) in PBS buffer for about 25 minutes, and then rinsed with 400  $\mu\text{l}$  wash buffer (TPB buffer with no casein) followed by 400  $\mu\text{l}$  of TPB buffer. For the first scheme, microsphere-DNA complexes were created by incubating approximately 100 pM microspheres with 10 pM labeled DNA in TPB buffer for at least an hour. The 10:1 ratio of beads to DNA was designed to minimize the occurrence of multiple DNA strands attached to a single microsphere. The tethering procedure was completed by introducing 50  $\mu\text{l}$  of the microsphere-DNA complexes into the flow cell for four to ten minutes. Additional tethering yield could be accomplished by another round of incubation with fresh microsphere-DNA complexes. For the second scheme, 500  $\mu\text{l}$  of labeled DNA in TPB buffer with about 2pM concentration was flushed into the chamber and incubated for around 1 hour. Then about 10 pM of beads was introduced into the chamber and incubate for 20 minutes after washing with 750  $\mu\text{l}$  TPB buffer to remove the unbound DNAs. Finally, for both schemes, unbound microspheres were removed by flushing the chamber with 1 mL TPB buffer. Once microspheres were introduced into the flow cell, tether integrity was improved by taking care to minimize flow rates within the sample chamber.

For experiments that investigated the effect of KCl, TPB was replaced with 10 mM Tris-HCl, pH 7.4, 10 mM KCl, 0.1 mM EDTA, 0.2 mM DTT, 5% DMSO and 1 mg/mL casein. A similar TPB

Table A.2: Materials used in the mutagenesis process for creating plasmids with two LacI binding sites. The capital letters in the primer sequences indicate the mutations. '()' indicates the deletion and '[' ]' indicates addition.

Molecule	Primer	Template	Action	Resulting
pUC1901	Mut0	pUC19	Replace	pUC19 with $O_1$ only
pUC300	Mut1	pUC1901	Replace	$O_1$ -300bp- $O_{id}$
pUC301	Mut2	pUC301	Delete 1bp	$O_1$ -301bp- $O_{id}$
pUC302	Mut3	pUC1901	Replace	$O_1$ -302bp- $O_{id}$
pUC303	Mut4	pUC1901	Replace	$O_1$ -303bp- $O_{id}$
pUC304	Mut5	pUC1901	Replace	$O_1$ -304bp- $O_{id}$
pUC305	Mut6	pUC1901	Replace	$O_1$ -305bp- $O_{id}$
pUC306	Mut7	pUC1901	Replace	$O_1$ -306bp- $O_{id}$
pUC307	Mut8	pUC1901	Replace	$O_1$ -307bp- $O_{id}$
pUC308	Mut9	pUC1901	Replace	$O_1$ -308bp- $O_{id}$
pUC309	Mut10	pUC308	Add 1bp	$O_1$ -309bp- $O_{id}$
pUC310	Mut11	pUC308	Add 2bp	$O_1$ -310bp- $O_{id}$
pUC71	Mut12	pUC1901	Replace	$O_1$ -71bp- $O_{id}$

Primer sequences(5' – >3'):

Mut0:ctaactcacattaattgcgttg AgctcGAGgTT cgctttccagtc Mut1: catacagccggaag (G) cataaagtgtaaagc

Mut2: ctcgaaagaaca AATTGTGAGCGCTCACAATT aagccaggaacc

Mut3: ctcgaaagaacaat AATTGTGAGCGCTCACAATT aggccaggaaccg

Mut4: cgaaagaacatg AATTGTGAGCGCTCACAATT gccaggaaccgt

Mut5: ggaaagaacatgt AATTGTGAGCGCTCACAATT gccaggaaccgta

Mut6: gaaagaacatgtg AATTGTGAGCGCTCACAATT ccaggaaccgtaa

Mut7: cgaaagaacatgtga AATTGTGAGCGCTCACAATT caggaaccgtaaaaag

Mut8: ggaaagaacatgtgag AATTGTGAGCGCTCACAATT aggaaccgtaaaaagg

Mut9: gaaagaacatgtgagc AATTGTGAGCGCTCACAATT ggaaccgtaaaaaggc

Mut10: catacagccggaag [C] cataaagtgtaaagc

Mut11: catacagccggaag [CG] cataaagtgtaaagc

Mut12: ctaactcacattaattg TgAG(G)cgctcacAATT cgctttccagtc

buffer was used to make sample for our  $Mg_{2+}$  dependence experiments, except the KCl concentration was 200 mM KCl. The experiment on exploring how the exposure time of the camera affects the magnitude of the tethered particle motion, samples were made identically to the calibration samples.

Table A.3: Materials used in amplifying labeled DNA for looping project using PCR.  
**For Looping Project: Length Dependence:**

Molecule	Template	Length(bp)	Resulting
pUC300L1	pUC300	900	Dig - 427bp- $O_1$ -300bp- $O_{id}$ -132bp - Bio
pUC301L1	pUC301	901	Dig - 427bp- $O_1$ -301bp- $O_{id}$ -132bp - Bio
pUC302L1	pUC302	901	Dig - 427bp- $O_1$ -302bp- $O_{id}$ -131bp - Bio
pUC303L1	pUC303	901	Dig - 427bp- $O_1$ -303bp- $O_{id}$ -130bp - Bio
pUC304L1	pUC304	901	Dig - 427bp- $O_1$ -304bp- $O_{id}$ -129bp - Bio
pUC305L1	pUC305	901	Dig - 427bp- $O_1$ -305bp- $O_{id}$ -128bp - Bio
pUC306L1	pUC306	901	Dig - 427bp- $O_1$ -306bp- $O_{id}$ -127bp - Bio
pUC307L1	pUC307	901	Dig - 427bp- $O_1$ -307bp- $O_{id}$ -126bp - Bio
pUC308L1	pUC308	901	Dig - 427bp- $O_1$ -308bp- $O_{id}$ -125bp - Bio
pUC309L1	pUC309	902	Dig - 427bp- $O_1$ -309bp- $O_{id}$ -125bp - Bio
pUC310L1	pUC310	903	Dig - 427bp- $O_1$ -310bp- $O_{id}$ -125bp - Bio
pUC71L1 *	pUC71	351	Dig - 427bp- $O_1$ -310bp- $O_{id}$ -125bp - Bio

Primer sequences(5' – >3'):

Plen901F: Dig - ACAGCTTGTCTGTAAGCGGATG

Plen901R: Bio - CGCCTGGTATCTTTATAGTCCTGTC

71F: DIG-AAAACGACGGCCAGTGAATTCG

71R: BIO-AGCCGAACGACCGAGCGCAG

**For Looping Project: Sequence dependence:**

Molecule	Template	Length(bp)	Resulting
pTA89L1	pTA89	445	Dig - 144bp- $O_{id}$ -TA89bp- $O_1$ -171bp - Bio
pTA94L1	pTA94	450	Dig - 144bp- $O_{id}$ -TA94bp- $O_1$ -171bp - Bio
pTA100L1	pTA100	456	Dig - 144bp- $O_{id}$ -TA100bp- $O_1$ -171bp - Bio
pE889L1	pE889	445	Dig - 144bp- $O_{id}$ -E889bp- $O_1$ -171bp - Bio
pE894L1	pE894	450	Dig - 144bp- $O_{id}$ -E894bp- $O_1$ -171bp - Bio
pE8100L1	pE8100	456	Dig - 144bp- $O_{id}$ -E8100bp- $O_1$ -171bp - Bio

Primer sequences(5' – > 3'):

Pseq450F: Dig - ATGCGAAACGATCCTCATCC

Pseq450R: Bio - GCATCACCTTCACCCTCTCC

Table A.4: Materials used in amplifying labeled DNA for calibration using PCR.  
**For TPM calibration:**

Molecule	Primers(5' – > 3')	Length(bp)
$\lambda$ 82	CAGTAAACGTCTGTTGAGCA GGGAATAACACCATGAAAAA	82
$\lambda$ 199	GGTTATCGAAATCAGCCACAGCGCC GGATACGTCTGAACTGGTCAC	199
$\lambda$ 298	ATTTGGTTGATCGTGGTG TCGTATTCGTCAAAGGGA	298
$\lambda$ 522	GCCACCTGTTACTGGTCGAT ACGGTGGAAACGATACTTGC	522
$\lambda$ 766	GGTGAACGATGCGTAATGTG TCAGCTGAATGGTGCAGTTC	766
$\lambda$ 997	CAAAGGCGGTTAAGGTGGTA GGCTGTACCGGACAATGAGT	997

# Appendix B

## Sequences

### B.1 Sequences Between the Two Operators for Phasing Experiments (300-310 bp)

300:

ccacacaacatacagaccggaacataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgagctcgaggtcgc  
 ttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttct  
 cgtcactgactcgtcgcctcggctggtcggctcggcgagcggatcagctcactcaaaggcggttaatacggttatccacagaatcagggga  
 taacgcaggaaagaaca

301:

ccacacaacatacagaccggaagcataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgagctcgaggtc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 ctgctcactgactcgtcgcctcggctggtcggctcggcgagcggatcagctcactcaaaggcggttaatacggttatccacagaatcaggg  
 gataacgcaggaaagaaca

302:

ccacacaacatacagaccggaagcataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgagctcgaggtc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 ctgctcactgactcgtcgcctcggctggtcggctcggcgagcggatcagctcactcaaaggcggttaatacggttatccacagaatcaggg  
 gataacgcaggaaagaacat

303:

ccacacaacatacagaccggaagcataaagtgtaaagcctgggggcctaataagtgagtgagctaactcacattaattgcgttgagctcgaggttc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 tcgctcactgactcgtcgcctcggtcgttcggctcggcgagcggatcagctcactcaaaggcgtaatacggttatccacagaatcagggg  
 ataacgcaggaaagaacatg

304:

ccacacaacatacagaccggaagcataaagtgtaaagcctgggggcctaataagtgagtgagctaactcacattaattgcgttgagctcgaggttc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 ctcgctcactgactcgtcgcctcggtcgttcggctcggcgagcggatcagctcactcaaaggcgtaatacggttatccacagaatcagg  
 ggataacgcaggaaagaacatgt

305:

ccacacaacatacagaccggaagcataaagtgtaaagcctgggggcctaataagtgagtgagctaactcacattaattgcgttgagctcgaggttc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 ctcgctcactgactcgtcgcctcggtcgttcggctcggcgagcggatcagctcactcaaaggcgtaatacggttatccacagaatcaggg  
 gataacgcaggaaagaacatgtg

306:

ccacacaacatacagaccggaagcataaagtgtaaagcctgggggcctaataagtgagtgagctaactcacattaattgcgttgagctcgaggttc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 ctcgctcactgactcgtcgcctcggtcgttcggctcggcgagcggatcagctcactcaaaggcgtaatacggttatccacagaatcaggg  
 gataacgcaggaaagaacatgtga

307:

ccacacaacatacagaccggaagcataaagtgtaaagcctgggggcctaataagtgagtgagctaactcacattaattgcgttgagctcgaggttc  
 gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttc  
 tcgctcactgactcgtcgcctcggtcgttcggctcggcgagcggatcagctcactcaaaggcgtaatacggttatccacanaatcagggg  
 ataacgcangaaagaacatgtgag



308:

ccacacaacatacagaccggaagcataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgagctcgaggt  
 cgtttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagagggcgtttgcgtattggcgctcttccgctt  
 cctcgtcactgactcgtcgcctcggctcgttcggctgcggcgagcgggtatcagctcactcaaaggcggtaatacggttatccacanaatcag  
 gggataacgcangaagaacatgtgagc

309:

ccacacaacatacagaccggaagccataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgagctcgagg  
 ttcgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagagggcgtttgcgtattggcgctcttccgctt  
 ttctcgtcactgactcgtcgcctcggctcgttcggctgcggcgagcgggtatcagctcactcaaaggcggtaatacggttatccacagaatcag  
 gggataacgcaggaaagaacatgtgagc

310:

ccacacaacatacagaccggaagccataaagtgtaaagcctggggtgcctaatagtgagtaactcacattaattgcgttgagctcgaggt  
 tcgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagagggcgtttgcgtattggcgctcttccgctt  
 cctcgtcactgactcgtcgcctcggctcgttcggctgcggcgagcgggtatcagctcactcaaaggcggtaatacggttatccacagaatcag  
 gggataacgcaggaaagaacatgtgagc

## **B.2 Sequences Between the Two Operators for Sequence-Dependence Experiments**

E8100:

ggccggaggctgctgcgtagaactacttttatttatcgctccacggctgatccc ctgtgctgttgccgtgttatctcaggttagtacgacgtc-  
 cgc

E894:

ggcctgctgtagaactacttttatttatcgctccacggctgatcccctg tgctgttgccgtgttatctcaggttagtacgacgtccc

E889:

ggccggctgctgtagaactacttttatttatcgctccacggctg ctgatcccctgtgctgttgccgtgttatctcaggttagtacgacc

TA100:

ggccgtaattggcgtagcaagctctagcaccgcttaaacgca cgtacgcgctgtctaccgcgtttaaccgccaataggattactactagtctc-  
tagc

TA94:

ggccgggctgtagcaagctctagcaccgcttaaacgcacgta cgcgctgtctaccgcgtttaaccgccaataggattactactagtctctac

TA89:

ggccgggctgtagcaagctctagcaccgcttaaacgcacgtacgcg ctgtctaccgcgtttaaccgccaataggattactactagtc

# Bibliography

- S. Adhya. Multipartite genetic control elements: communication by DNA loop. *Annual review of genetics*, 23:227–50, 1989.
- J. F. Allemand, S. Cocco, N. Douarche, and G. Lia. Loops in DNA: An overview of experimental and theoretical approaches. *European Physical Journal E*, 19(3):293–302, 2006.
- A. Balaeff, L. Mahadevan, and K. Schulten. Modeling DNA loops using the theory of elasticity. *Physical review*, 73:031919, Mar. 2006.
- J. K. Barry and K. S. Matthews. Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. *Biochemistry*, 38(20):6520–8, May 1999.
- C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 94(12):6185–90, June 1997.
- J. F. Beausang and P. C. Nelson. Diffusive hidden markov model characterization of DNA looping dynamics in tethered particle experiments. Submitted, 2007.
- J. F. Beausang, C. Zurla, C. Manzo, D. Dunlap, L. Finzi, and P. Nelson. DNA Looping Kinetics Analyzed Using Diffusive Hidden Markov Model. *Biophysical Journal*, Feb. 2007.
- N. A. Becker, J. D. Kahn, and L. J. Maher. Bacterial repression loops require enhanced DNA flexibility. *Journal of Molecular Biology*, 349(4):716–30, June 2005.
- L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips. Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev*, 15(2):125–35, 2005a.

- L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: models. *Current opinion in genetics and development*, 15(2):116–24, Apr. 2005b.
- L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*, 15(2):116–24, 2005c.
- S. Blumberg, A. Gajraj, M. W. Pennington, and J. Meiners. 3-d characterization of Tethered Microspheres by Total Internal Reflection Fluorescence Microscopy. *Biophys. J.*, page biophysj.105.061242, 2005.
- D. H. Boal. *Mechanics of the cell*. Cambridge University Press, Cambridge [England] ; New York, 2002.
- J. A. Borowiec, L. Zhang, S. Sasse-Dwight, and J. D. Gralla. Dna supercoiling promotes formation of a bent repression loop in lac dna. *J Mol Biol*, 196(1):101–11, 1987.
- B. V. D. Broek, F. Vanzi, D. Normanno, F. S. Pavone, and G. J. Wuite. Real-time observation of dna looping dynamics of type iie restriction enzymes naei and nari. *Nucleic Acids Res*, 34(1):167–74, 2006.
- N. E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*, 100(9):5136–41, 2003.
- B. Chen, B. de Crombrughe, W. B. Anderson, M. E. Gottesman, I. Pastan, and R. L. Perlman. On the mechanism of action of lac repressor. *Nature: New biology*, 233(37):67–70, Sept. 1971.
- T. E. Cloutier and J. Widom. Spontaneous sharp bending of double-stranded DNA. *Mol Cell*, 14(3):355–62, 2004.
- T. E. Cloutier and J. Widom. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc Natl Acad Sci USA*, 102(10):3645–50, 2005.
- D. Colquhoun and F. J. Sigworth. Fitting and statistical analysis of single-channel records. In B. Sakmann and E. Neher, editors, *Single-Channel Recording*, pages 483–587. Plenum Press, New York, second edition, 1995.
- L. Czapla, D. Swigon, and W. K. Olson. Sequence-dependent effects in the cyclization of short DNA. *Journal of Chemical Theory and Computation*, 2(3):685–695, 2006.

- I. B. Dodd, K. E. Shearwin, A. J. Perkins, T. Burr, A. Hochschild, and J. B. Egan. Cooperativity in long-range gene regulation by the lambda ci repressor. *Genes Dev*, 18(3):344–54, 2004.
- N. Douarche and S. Cocco. Protein-mediated DNA loops: Effects of protein bridge size and kinks. *Physical Review E*, 72(6):061902–(1–10), 2005.
- E. R. Dufresne, T. M. Squires, M. P. Brenner, and D. G. Grier. Hydrodynamic coupling of two brownian spheres to a planar surface. *Physical review letters*, 85(15):3317–20, Oct. 2000.
- T. M. Dunn, S. Hahn, S. Ogden, and R. F. An operator at -280 base pairs that is required for repression of arabid operon promoter: addition of dna helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci U S A*, 81(16):5017–20, 1984.
- L. M. Edelman, R. Cheong, and J. D. Kahn. Fluorescence resonance energy transfer over approximately 130 basepairs in hyperstable lac repressor-dna loops. *Biophys J*, 84(2 Pt 1):1131–45, 2003.
- C. M. Falcon, L. Swint-Kruse, and K. S. Matthews. Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. *The Journal of biological chemistry*, 272(43):26818–21, Oct. 1997.
- L. Finzi and J. Gelles. Measurement of lacose repressor-mediated loop formation and breakdown in single dna molecules. *Science*, 267:378–380, 1995.
- A. M. Friedman, T. O. Fischmann, and T. A. Steitz. Crystal structure of lac repressor core tetramer and its implications for dna looping. *Science*, 268(5218):1721–7, 1995.
- H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins. Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers*, 85(2):115–30, Feb. 2007.
- M. Geanacopoulos, G. Vasmatzis, V. B. Zhurkin, and S. Adhya. Gal repressosome contains an antiparallel DNA loop. *Nature structural biology*, 8(5):432–6, May 2001.
- J. Gelles, B. Schnapp, and M. Sheetz. Tracking kinesin-driven movements with nanometre-scale precision. *Nature*, 331(4):450–453, 1988.

- M. M. Gromiha, J. G. Siebers, S. Selvaraj, H. Kono, and A. Sarai. Intermolecular and intramolecular readout mechanisms in protein-dna recognition. *Journal of Molecular Biology*, 337(2):285–294, 2004.
- L. Han, P. Grayson, S. Blumberg, J. Beausang, P. Nelson, and R. Phillips. Length and concentration dependence of DNA looping, in preparation, 2007.
- W. T. Hsieh, P. A. Whitson, K. S. Matthews, and R. D. Wells. Influence of sequence and distance between two operators on interaction with the lac repressor. *J Biol Chem*, 262(30):14583–91, 1987.
- H. Huang, D. Dabiri, and M. Gharib. On errors of digital particle image velocimetry. *Measurement Science and Technology*, 8(12):1427–1440, Dec. 1997.
- S. Inouye, M. Gomada, U. M. Sangodkar, A. Nakazawa, and T. Nakazawa. Upstream regulatory sequence for transcriptional activator xylr in the first operon of xylene metabolism on the tol plasmid. *J Mol Biol*, 216(2):251–60, 1990.
- H. Krämer, M. Niemoller, M. Amouyal, B. Revet, B. V. Wilcken-Bergmann, and B. Muller-Hill. lac repressor forms loops with linear dna carrying two suitably spaced lac operators. *Embo J*, 6(5):1481–91, 1987.
- H. Krämer, M. Amouyal, A. Nordheim, and B. Muller-Hill. Dna supercoiling changes the spacing requirement of two lac operators for dna loop formation with lac repressor. *Embo J*, 7(2):547–56, 1988.
- S. M. Law, G. R. Bellomy, P. J. Schlax, and M. T. Record Jr. In vivo thermodynamic analysis of repression with and without looping in lac constructs. estimates of free and local lac repressor concentrations and of physical properties of a region of supercoiled plasmid dna in vivo. *J Mol Biol*, 230(1):161–73, 1993.
- D. H. Lee and R. F. Schleif. In vivo DNA loops in araCBAD: size limits and helical repeat. *Proc Natl Acad Sci U S A*, 86(2):476–80, 1989.
- M. M. Levandoski, O. V. Tsodikov, D. E. Frank, S. E. Melcher, R. M. Saecker, and M. T. Record. Cooperative and anticooperative effects in binding of the first and second plasmid Osym operators to a LacI tetramer: evidence for contributions of non-operator DNA binding by wrapping and looping. *Journal of molecular biology*, 260(5):697–717, Aug. 1996.

- M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. Crystal structure of the lactose operon repressor and its complexes with dna and inducer. *Science*, 271(5253):1247–54, 1996.
- Y. H. Loh, Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim, and H. H. Ng. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics*, 38(4):431–40, Apr. 2006.
- K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648):251–60, 1997.
- R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*, 25(6):1203–10, 1997.
- J. Majors. Initiation of in vitro mRNA synthesis from the wild-type lac promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 72(11):4394–8, Nov. 1975.
- K. S. Matthews. Dna looping. *Microbiol Rev*, 56(1):123–36, 1992.
- R. A. Mehta and J. D. Kahn. Designed hyperstable Lac repressor.DNA loop topologies suggest alternative loop geometries. *Journal of molecular biology*, 294(1):67–77, Nov. 1999.
- M. A. Morgan, K. Okamoto, J. D. Kahn, and D. S. English. Single-molecule spectroscopic determination of lac repressor-dna loop conformation. *Biophys J*, 89(4):2588–96, 2005.
- J. Muller, S. Oehler, and B. Muller-Hill. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol*, 257(1):21–9, 1996.
- B. Müller-Hill. *The lac Operon : a short history of a genetic paradigm*. Walter de Gruyter, Berlin ; New York, 1996.
- P. C. Nelson, C. Zurla, D. Brogioli, J. F. Beausang, L. Finzi, and D. Dunlap. Tethered particle motion as a diagnostic of dna tether length. *J Phys Chem B Condens Matter Mater Surf Interfaces Biophys*, 110(34):17260–7, 2006a.

- P. C. Nelson, C. Zurla, D. Brogioli, J. F. Beausang, L. Finzi, and D. Dunlap. Tethered particle motion as a diagnostic of DNA tether length. *Journal of Physical Chemistry B*, 110(34):17260–17267, 2006b.
- H. Nick and W. Gilbert. Detection in vivo of protein-DNA interactions within the lac operon of *Escherichia coli*. *Nature*, 313(6005):795–8, 1985.
- J. V. Noort, S. Verbrugge, N. Goosen, C. Dekker, and R. T. Dame. Dual architectural roles of huf formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci U S A*, 101(18):6969–74, 2004.
- X. Nou, B. Braaten, L. Kaltenbach, and D. A. Low. Differential binding of lrp to two sets of pap dna binding sites mediated by pap i regulates pap phase variation in *Escherichia coli*. *Embo J*, 14(23):5785–97, 1995.
- J. Plumbridge and A. Kolb. Dna bending and expression of the divergent nage-b operons. *Nucleic Acids Res*, 26(5):1254–60, 1998.
- N. Pouget, C. Dennis, C. Turlan, M. Grigoriev, M. Chandler, and L. Salome. Single-particle tracking for DNA tether length monitoring. *Nucleic Acids Research*, 32(9), 2004.
- M. Ptashne. *A genetic switch : phage lambda revisited*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 3rd edition, 2004.
- L. Rabiner. A tutorial on hidden markov models and selected applications inspeech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb. 1989.
- A. Revzin and P. H. von Hippel. Direct measurement of association constants for the binding of *Escherichia coli* lac repressor to non-operator DNA. *Biochemistry*, 16(22):4769–76, Nov. 1977.
- C. Rivetti, M. Guthold, and C. Bustamante. Scanning force microscopy of dna deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J Mol Biol*, 264(5):919–32, Dec 1996.
- N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717):1962–5, 2005.



- M. Roychoudhury, A. Sitlani, J. Lapham, and D. M. Crothers. Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proc Natl Acad Sci U S A*, 97(25):13608–13, 2000.
- G. C. Ruben and T. B. Roos. Conformation of lac repressor tetramer in solution, bound and unbound to operator dna. *Microsc Res Tech*, 36(5):400–16, 1997.
- L. Saiz and J. M. Vilar. In vivo evidence of alternative loop geometries in dna-protein complexes. *q-bio.BM/0602012*, 2006a.
- L. Saiz and J. M. Vilar. DNA looping: the consequences and its control. *Current opinion in structural biology*, 16(3):344–50, June 2006b.
- L. Saiz, J. M. Rubi, and J. M. Vilar. Inferring the in vivo looping properties of dna. *Proc Natl Acad Sci U S A*, 2005.
- D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature*, 352(6334):444–8, Aug. 1991.
- R. Schleif. DNA looping. *Annu Rev Biochem*, 61:199–223, 1992.
- E. Segal, Y. Fondufe-Mittendorf, . Chen, A. Thastrom, Y. Field, I. K. Moore, J. P. Wang, and J. Widom. A genomic code for nucleosome positioning. *Nature*, 442(7104):772–8, 2006.
- D. E. Segall, P. C. Nelson, and R. Phillips. Volume-exclusion effects in tethered-particle experiments: bead size matters. *Phys Rev Lett*, 96(8):088306, 2006a.
- D. E. Segall, P. C. Nelson, and R. Phillips. Volume-exclusion effects in tethered-particle experiments: Bead size matters. *Physical Review Letters*, 96(8):088306–(1–4), 2006b.
- S. Semsey, M. Y. Tolstorukov, K. Virnik, V. B. Zhurkin, and S. Adhya. DNA trajectory in the Gal repressosome. *Genes and development*, 18(15):1898–907, Aug. 2004.
- H. Shroff, B. M. Reinhard, M. Siu, H. Agarwal, A. Spakowitz, and J. Liphardt. Biocompatible force sensor with optical readout and dimensions of 6 nm<sup>3</sup>. *Nano Lett*, 5(7):1509–14, 2005.
- M. Singh-Zocchi, S. Dixit, V. Ivanov, and G. Zocchi. Single-molecule detection of DNA hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7605–10, June 2003.

- S. B. Straney and D. M. Crothers. Lac repressor is a transient gene-activating protein. *Cell*, 51(5): 699–707, Dec. 1987.
- T. R. Strick, V. Croquette, and D. Bensimon. Homologous pairing in stretched supercoiled DNA. *Proc. Natl. Acad. Sci. USA*, 95:10579–10583, 1998.
- D. Swigon, B. D. Coleman, and W. K. Olson. Modeling the lac repressor-operator assembly: The influence of dna looping on lac repressor conformation. *Proc Natl Acad Sci U S A*, 103(26): 9879–84, 2006.
- S. Toli-Norrelykke, M. Rasmussen, F. Pavone, K. Berg-Srensen, and L. Oddershede. Stepwise bending of DNA by a single TATA-box binding protein. *Biophysical journal*, 90(10):3694–703, May 2006.
- S. F. Tolic-Norrelykke, A. M. Engh, R. Landick, and J. Gelles. Diversity in the rates of transcript elongation by single rna polymerase molecules. *J Biol Chem*, 279(5):3292–9, 2004a.
- S. F. Tolic-Norrelykke, A. M. Engh, R. Landick, and J. Gelles. Diversity in the rates of transcript elongation by single RNA polymerase molecules. *J Biol Chem*, 279(5):3292–9, 2004b.
- B. van den Broek, F. Vanzi, D. Normanno, F. S. Pavone, and G. J. Wuite. Real-time observation of DNA looping dynamics of Type IIE restriction enzymes NaeI and NarI. *Nucleic acids research*, 34(1):167–74, 2006.
- F. Vanzi, S. Vladimirov, C. R. Knudsen, Y. E. Goldman, and B. S. Cooperman. Protein synthesis by single ribosomes. *RNA (New York, N.Y)*, 9(10):1174–9, Oct. 2003.
- F. Vanzi, C. Broggio, L. Sacconi, and F. S. Pavone. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic acids research*, 34(12):3409–20, 2006.
- J. M. Vilar and L. Saiz. Dna looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev*, 15(2):136–44, 2005.
- A. Vologodskii and N. R. Cozzarelli. Effect of supercoiling on the juxtaposition and relative orientation of dna sites. *Biophys J*, 70(6):2548–56, 1996.

- J. D. Watson and F. H. Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in Nature, number 4356 April 25, 1953. *Nature*, 248(5451):765, Apr. 1953.
- C. L. Wei, Q. Wu, V. B. Vega, K. P. Chiu, P. Ng, T. Zhang, A. Shahab, H. C. Yong, Y. Fu, Z. Weng, J. Liu, X. D. Zhao, J. L. Chew, Y. L. Lee, V. A. Kuznetsov, W. K. Sung, L. D. Miller, B. Lim, E. T. Liu, Q. Yu, H. H. Ng, and Y. Ruan. A global map of p53 transcription-factor binding sites in the human genome. *Cell*, 124(1):207–19, Jan. 2006.
- P. A. Whitson, W. T. Hsieh, R. D. Wells, and K. S. Matthews. Influence of supercoiling and sequence context on operator dna binding with lac repressor. *J Biol Chem*, 262(30):14592–9, 1987.
- P. A. Wiggins and P. C. Nelson. Generalized theory of semiflexible polymers. *Phys Rev E Stat Nonlin Soft Matter Phys*, 73(3 Pt 1):031906, 2006.
- P. A. Wiggins, P. C. Nelson, and R. Phillips. Exact theory of kinkable elastic polymers. *Phys. Rev. E*, 71:021909–(1–19), 2005.
- P. A. Wiggins, T. V. D. Heijde, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology*, 1(2):137–141, 2006.
- O. K. Wong, M. Guthold, D. A. Erie, and J. Gelles. Interconvertable lactose repressor-DNA looped complexes revealed by single-molecule experiments, submitted, 2007.
- J. Yan and J. F. Marko. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.*, 93:108108–(1–4), 2004.
- J. Yan, R. Kawamura, and J. F. Marko. Statistics of loop formation along double helix DNAs. *Physical review E*, 71:061905, June 2005.
- R. Yasmin, K. T. Yeung, R. H. Chung, M. E. Gaczynska, P. A. Osmulski, and N. Noy. Dna-looping by rxr tetramers permits transcriptional regulation "at a distance". *J Mol Biol*, 343(2):327–38, 2004.
- H. Yin, R. Landick, and J. Gelles. Tethered particle motion method for studying transcript elongation by a single RNA polymerase molecule. *Biophys. J.*, 67(6):2468–2478, 1994.

- X. Zhang and P. A. Gottlieb. Thermodynamic and alkylation interference analysis of the lac repressor-operator substituted with the analogue 7-deazaguanine. *Biochemistry*, 32(42):11374–84, Oct. 1993.
- Y. Zhang, A. E. McEwen, D. M. Crothers, and S. D. Levene. Statistical-mechanical theory of DNA looping. *Biophysical journal*, 90(6):1903–12, Mar. 2006.
- C. Zurla, A. Franzini, G. Galli, D. Dunlap, D. E. A. Lewis, S. Adhya, and L. Finzi. Novel tethered particle motion analysis of CI protein-mediated DNA looping in the regulation of bacteriophage lambda. *J. Phys.: Condens. Matter*, 18(14):S225–S234, 2006.
- C. Zurla, T. Samuely, G. Bertoni, F. Valle, G. Dietler, L. Finzi, and D. D. Dunlap. Integration host factor alters LacI-induced DNA looping. *Biophysical chemistry*, 128(2-3):245–52, July 2007.