SINGLE-MOLECULE BIOPHYSICS OF DNA BENDING: LOOPING AND UNLOOPING

A Dissertation
Presented to
The Academic Faculty

by

Tung T. Le

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology
August 2015

SINGLE-MOLECULE BIOPHYSICS OF DNA BENDING: LOOPING AND UNLOOPING

Approved by:

Professor Harold Kim, Advisor
School of Physics
Georgia Institute of Technology

Professor James Gumbart
School of Physics
Georgia Institute of Technology

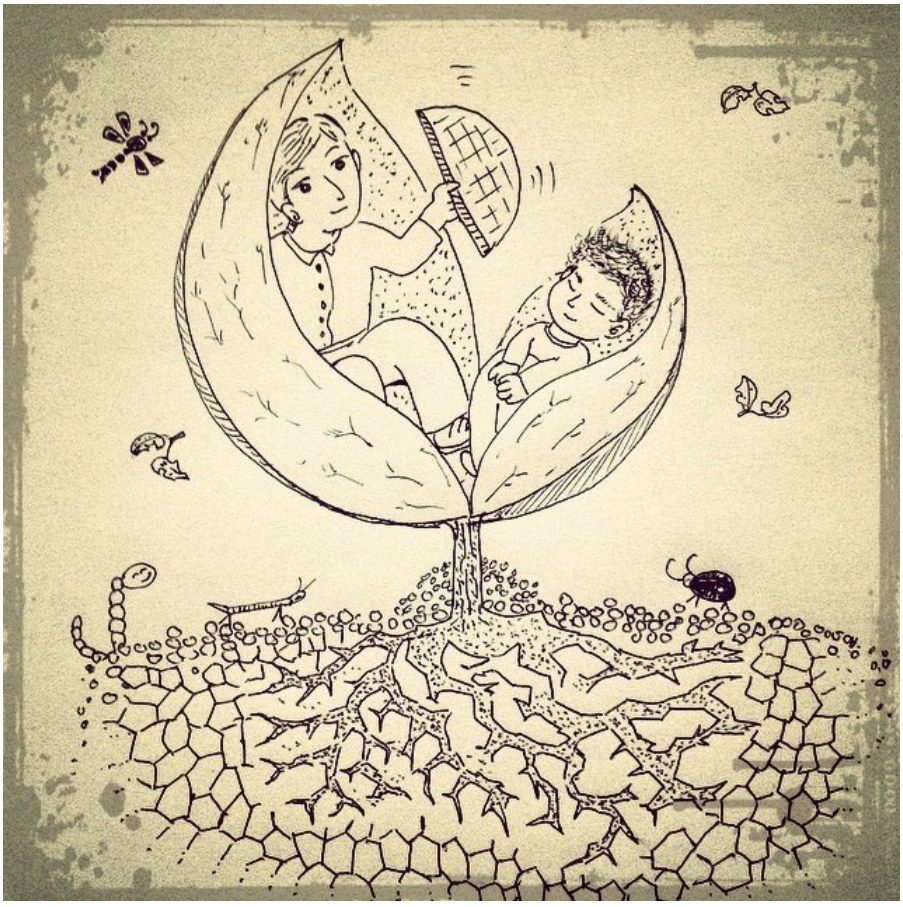
Professor Daniel Goldman
School of Physics
Georgia Institute of Technology

Professor Loren Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Professor Jennifer Curtis
School of Physics
Georgia Institute of Technology

Date Approved: April 09, 2015

A dedication to my father, with love.



ACKNOWLEDGEMENTS

First, I would like to give my highest appreciation to my advisor, Dr. Harold D. Kim, for his enduring guidance and generous support to my research during the graduate years. He has given me a unique opportunity to work independently and judge things in a scientific manner.

I would like to thank my previous advisor, Dr. Toan Nguyen, for his great mentor and support during my first year at Georgia Tech. I would also like to extend my appreciation to my committee members: Dr. Daniel Goldman, Dr. Jennifer Curtis, Dr. James Gumbart, and Dr. Loren Williams, for reading my thesis and giving me valuable comments. I want to thank Dr. Rasesh Parikh and other members of the Kim's lab for their kindness.

Outside the lab, I would like to give thanks to many people, my dear friends (Phuong Doan, Tien Hoang, Phuong Hoang) and many others in the Vietnamese Student Associations (VSA) in Atlanta, and my close friends at home (Vuong Linh Tran, Duy Nguyen) for their love and friendship.

Finally, I would like to express my deepest gratitude to my beloved parents and my two sisters. I want to honor my father, who has recently passed away, for having strong beliefs in me and in my choice. I really want him to feel so proud of me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis investigates bending mechanics of short double-stranded DNA below 200 base-pairs. DNA is usually portrayed as a straight duplex with an even helical pitch of almost 10.5 base pairs. However, DNA can adopt non-canonical structures due to thermal fluctuations, and often engages in numerous cellular activities such as transcription, viral packaging, and nucleosome formation through strong bending. Surprisingly, our understanding on the mechanical properties of DNA is still limited and fragmented. For example, we still do not clearly understand the origin of DNA stiffness, sequence-dependence of DNA curvature and flexibility, and DNA structure under extreme deformations.

My thesis is devoted to addressing a controversial topic of strong DNA bending. DNA flexibility at large-length scales can be well explained by an elastic polymer model known as the worm-like chain model, but the validity of the model at small length scales has been challenged over the last decade. We consider several subtle features that could potentially lead to the break-down of the worm-like chain model, such as local bendedness of the sequence and large bending angles. We used single-molecule Fluorescence Resonance Energy Transfer to track looping and unlooping of single DNA molecules in real time. We compared the measured looping and unlooping rates with theoretical predictions of the worm-like chain model.

We found that the intrinsic curvature of the sequence affects the looping propensity of short DNA and an extended worm-like chain model including the helical parameters of individual base pairs could adequately explain our measurements. For DNA with random sequence and negligible curvature, we discovered that the worm-like chain model could explain the stability of small DNA loops only down to a critical loop size. Below the critical loop size, the bending stress stored in the DNA loop became less sensitive to loop size, indicative of softened double-stranded DNA. The critical loop size is sensitive to salt condition, especially to magnesium at mM concentrations. This finding enabled us to

explain several contrasting results in the past and shed new light on the energetics of DNA bending.

CHAPTER I

INTRODUCTION

1.1 DNA bends

Double-stranded DNA (dsDNA) is arguably one of the most important molecules in science[1]. Besides its biological importance as a carrier of the genetic code, it has become a versatile molecular building block for self-assembly of sub-micron structures. Although the structure of dsDNA was solved nearly 60 years ago[2], how its structure affects biology still remains a central question. This thesis aims to investigate the mechanical properties of DNA on small length scales.

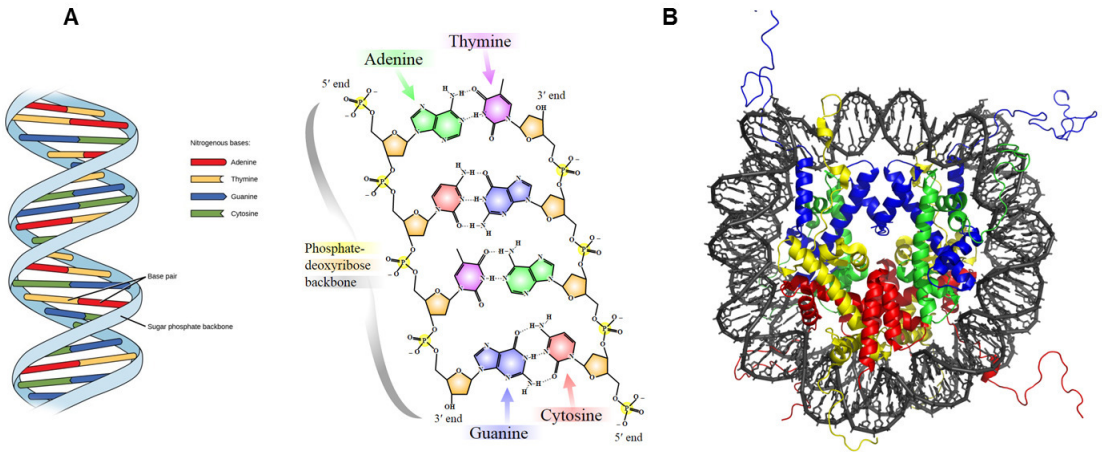


Figure 1.1: The dsDNA structure and DNA bending in the nucleosome. (A) A schematic representation of the double helix structure and the molecular structures of DNA base-pairs and backbone. (B) The crystal structure of the nucleosome core particle consisting of H2A (yellow) , H2B (red) , H3 (blue) and H4 (green) core histones, and DNA (gray). The view is from the top through the superhelical axis. Images adapted from `http://en.wikipedia.org/wiki/DNA`.

Duplex DNA is formed by two ribose-phosphate polymer chains (DNA strands) locked by hydrogen bonding between their complementary bases: guanine G (or adenine A) on one strand with cytosine C (or thymine T) on the other strand (Figure 1.1A). The electrostatic

repulsion of the phosphate groups in the backbone and the stacking of the bases on top of each other result in an unusually large rigidity of the molecule. This special structure helps preserving the genetic code but also poses a challenge to mechanical deformations of the DNA such as bending, twisting, and stretching.

Despite the large DNA rigidity, strongly deformed DNA is commonly found inside the cells. Some typical examples are packaging of viral DNA[3], transcription factor mediated DNA looping[4], and winding of DNA in nucleosomes[5]. Of particular interest is the nucleosome, the basic unit of DNA packaging in eukaryotes, which plays an important role in regulating transcriptional activities in the cells. The nucleosome particle consists of 147 base pairs (bp) of DNA wrapped in roughly 1.7 left-handed superhelical turns[6] around a histone octamer consisting of 2 copies each of the core histones H2A, H2B, H3, and H4 (Figure 1.1B), which corresponds to an 80-bp circle with a 4.5-nm radius. This requires a bending energy of $35k_BT$ if one assumes a purely elastic model for DNA bending[7]. This huge energy barrier suggests that formation of such tiny DNA loops driven by thermal excitations would be rare in the absence of the protein. Furthermore, it is also observed that dsDNA in the nucleosomes exhibits local structural disruptions such as large bending kinks at the contacts between DNA and histone octamer's surface[8]. Understanding the structure of DNA loop and the mechanics of DNA bending at small length scales is therefore of great importance since such local deviations could be averaged out over large length scales.

In this thesis, I will attempt to provide a quantitative model of DNA bending *in vitro* that may shed light on energetics of sharply bent DNA found *in vivo*. I will explain how DNA exploits its local sequence curvature or spontaneous structural disruptions to enhance its bending flexibility. I will also show how ionic environment around the molecule could affect its overall bendability, especially in the sub-persistence length regime.

1.2 The worm-like chain model of DNA

dsDNA is a semi-flexible polymer, in which successive segments display orientational cooperativity. The worm-like chain (WLC) model[9] can describe such behavior at the coarse-grained level while ignoring the structural details of dsDNA. In this model, the DNA polymer is simply considered as a trajectory that coincides with the helical axis of the DNA molecule.

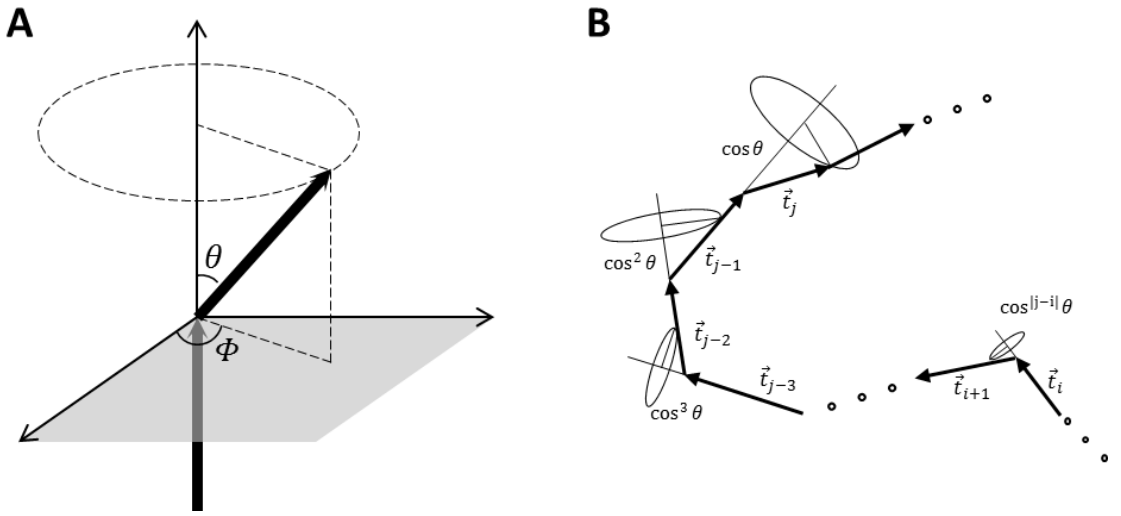


Figure 1.2: Freely rotating chain model. (A) Bond angle θ and torsion angle ϕ. (B) Schematic representation to calculate the correlation between bond vectors $\mathbf{l}_i$ and $\mathbf{l}_j$. Figure redrawn from Figure 2.5 in reference [10].

Consider a chain of N segments, $\mathbf{l}_1, \cdots, \mathbf{l}_N$, each segment with length l. All segments are connected successively, and each segment can rotate freely with respect to the next one. This is called the *freely jointed chain*. The relative orientation between two successive segments can be completely described by two angles, the bond angle θ and the torsion angle ϕ (see Figure 1.2A). If one constrains the bond angles θ between two segments but put no restrictions on the torsional angles, it becomes the *freely rotating chain*. Let's assume the bond angles are fixed. Because of the close interference between segments, the chain possesses some rigidity and two segments across the chain can exhibit angular correlation. The average angular correlation between segment i and segment j is:

$$\langle \mathbf{l}_i \cdot \mathbf{l}_j \rangle = l^2 \langle \mathbf{t}_i \cdot \mathbf{t}_j \rangle, \tag{1.1}$$

where $\mathbf{t}_i$ is the unit vector along segment i. One can calculate this factor by considering the correlation transmitted from segment j down to segment i across the chain (Figure 1.2B). For a freely rotating chain, the component of vector $\mathbf{t}_j$ normal to vector $\mathbf{t}_{j-1}$ averages out to zero due to the free rotations of the torsion angle ϕ_j. Only the longitudinal component of vector $\mathbf{t}_j$ along vector $\mathbf{t}_{j-1}$ can be transmitted, which is $\cos\theta$. If one keeps passing the correlation from segment $j-1$ to segment $j-2$, down to segment i, the overall correlation factor $\langle \mathbf{l}_i \cdot \mathbf{l}_j \rangle$ finally takes the form

$$\langle \mathbf{l}_i \cdot \mathbf{l}_j \rangle = l^2 \cos^{|j-i|}\theta. \tag{1.2}$$

Similarly, the projection of the end-to-end vector $\mathbf{R} = \sum_{i=1}^{N} \mathbf{l_i}$ onto the direction of the first segment, which represents the overall correlation of the chain,

$$\langle \mathbf{R} \cdot \mathbf{t}_1 \rangle = l(1 + \cos\theta + \cos^2\theta + \cdots + \cos^{N-1}\theta) = l\frac{1-\cos^N\theta}{1-\cos\theta}. \tag{1.3}$$

In the limit $N \to \infty$, this quantity is called the *persistence length*, L_p, which represents the memory of the initial direction of the chain, $L_p = l/(1-\cos\theta)$. The persistence length set a length scale above which the chain directional correlation becomes negligible[11]. The correlation between segment i and segment j can be rewritten in term of the persistence length as

$$\langle \mathbf{t}_i \cdot \mathbf{t}_j \rangle = \left(1 - \frac{l}{L_p}\right)^{|j-i|} \approx 1 - |j-i|\frac{l}{L_p} \approx \exp\left(-\frac{|j-i|l}{L_p}\right). \tag{1.4}$$

In the discrete Kratky-Porod WLC model, the angle θ is not fixed. Instead, it follows a Boltzmann distribution with the energy dictated by the bending angle between two consecutive segments

$$\epsilon_{i,i+1} = k_B T \frac{L_p}{2l}(\mathbf{t}_{i+1} - \mathbf{t}_i)^2 = k_B T \frac{L_p}{l}(1 - \cos\theta_{i,i+1}) \approx k_B T \frac{L_p}{2l}\theta_{i,i+1}^2, \tag{1.5}$$

where $\theta_{i,i+1}$ is the deflection angle between segment i and segment $i+1$ and $k_B T$ is the thermal energy. According to this model, directional change in the chain contour costs energy quadratically dependent on bending angle. As a result of the cooperativity between successive segments, the WLC model describes the polymer behavior between a rigid rod and a random coil. At room temperature, the polymer adopts a smoothly curved conformation, whereas at $T = 0K$, the polymer adopts a rigid rod conformation.

In the limit of infinitesimally small unit segments, we have the continuous WLC model that describes an isotropic rod with continuous flexibility[9]. By letting $N \to \infty$ and $l \to 0$ under the restriction that the total contour length is constant ($Nl = L$, the total contour length), we can express the end-to-end vector $\mathbf{R}$ and other quantities in term of a path integral

$$\mathbf{R} = \int_0^L \hat{t}(s)\mathrm{d}s, \tag{1.6}$$

where $\hat{t}(s) \equiv \frac{\partial \mathbf{r}(s)}{\partial s}$ is the tangential vector at position s across the contour of the polymer and $\mathbf{r}(s)$ is a radius vector that points to the position s. The correlation factor still follows Equation 1.4 as $\langle \hat{t}(0) \cdot \hat{t}(s) \rangle = \exp(-s/L_p)$. The total bending energy of the polymer can be expressed as

$$E = \frac{k_B T L_p}{2} \lim_{l \to 0} \left[\sum_{i=1}^{N-1} \left(\frac{\mathbf{t}_{i+1} - \mathbf{t}_i}{l} \right)^2 l \right] = \frac{k_B T L_p}{2} \int_0^L \left(\frac{\partial \hat{t}(s)}{\partial s} \right)^2 \mathrm{d}s. \tag{1.7}$$

Another useful quantity is the mean square end-to-end distance $\langle R^2 \rangle$ which characterizes the size of the polymer

$$\begin{aligned} \langle R^2 \rangle = \langle \mathbf{R} \cdot \mathbf{R} \rangle &= \int_0^L \mathrm{d}s \int_0^L \mathrm{d}s' \langle \hat{t}(s) \cdot \hat{t}(s') \rangle \\ &= \int_0^L \mathrm{d}s \int_0^L \mathrm{d}s' e^{-|s-s'|/L_p} = 2L_p L \left[1 - \frac{L_p}{L}(1 - e^{-L/L_p}) \right]. \end{aligned} \tag{1.8}$$

Under physiological condition, the typical value of L_p for dsDNA has been measured to be $45 - 53$ nm by various methods[12, 13, 14, 15, 16, 17]. For contour length much larger than L_p ($L \gg L_p$), DNA behaves like a flexible polymer $\langle R^2 \rangle = 2L_p L$, while when $L \leqslant L_p$, DNA can be described as a rod-like polymer with $\langle R^2 \rangle \approx L^2$. Despite its simplicity, the WLC model was shown to be able to explain DNA stretching experiments[18, 19, 14, 20] for DNA of a few kilobase pairs. For shorter DNA of a few hundreds of base pairs used in the cyclization experiments[21, 17, 22], the agreement is also remarkable. For even shorter DNA of ∼100 bp, the agreement is less certain and this will be the main theme of my thesis. More discussion on the applicability of the WLC model is summarized in Section 1.4.

1.3 J factor

To quantify the flexibility of dsDNA, we can determine the ring-closure probability of the DNA, which is also known as the J factor. The J factor of a linear DNA molecule is a quantity that measures the relative concentration of one end of the DNA in the vicinity of the other end. Consider a DNA polymer with contour length L in thermal equilibrium, one can define $P(r)$ as the probability distribution of the end-to-end distance r. $P(r)$ indicates the fraction of DNA conformations with end-to-end distance r and it should fulfill the normalization condition, $\int_{r=0}^{L} P(r)\mathrm{d}r = 1$. The relative density of the one end at a distance r from the other end, $j(r)$, can be calculated from $P(r)$ as

$$j(r) = \frac{P(r)}{4\pi r^2}. \tag{1.9}$$

This quantity, when expressed in term of molar unit, is the J factor.

The J factor depends on the contour length of the DNA and also on the relative geometry of the two ends. Shown in Figure 1.3 are some examples of how different geometric constraints affect the J factor according to the WLC model. The most extreme case is when the two ends are in direct contact and require torsional alignment for a complete matching of helical phases of the two termini (black curve in Figure 1.3). In this case, the J factor oscillates with a period of ∼10 bp (helical period of DNA) due to torsional rigidity. In general, the J factor becomes smaller and has a stronger length dependence as more constraints on end geometry are imposed. Analytical calculations of the J factor for the WLC model are available at all length scales[23, 24, 25], making it possible to experimentally check the validity of this classical polymer model.

Experimentally, the J factor can be measured by a technique called the bulk cyclization assay[16, 21, 17]. DNA is modified by biochemical methods to have two short single-stranded overhangs. The two overhangs are complementary to each other so that they form a pair of cohesive ends. The two cohesive ends can interact when they are brought to close proximity and trap the DNA in the looped state. The key idea is that by normalizing the equilibrium constant for cyclization ($K_{a,circular}$) by the equilibrium constant for bimolecular association between separate DNA molecules with identical cohesive ends ($K_{a,bimolecular}$),

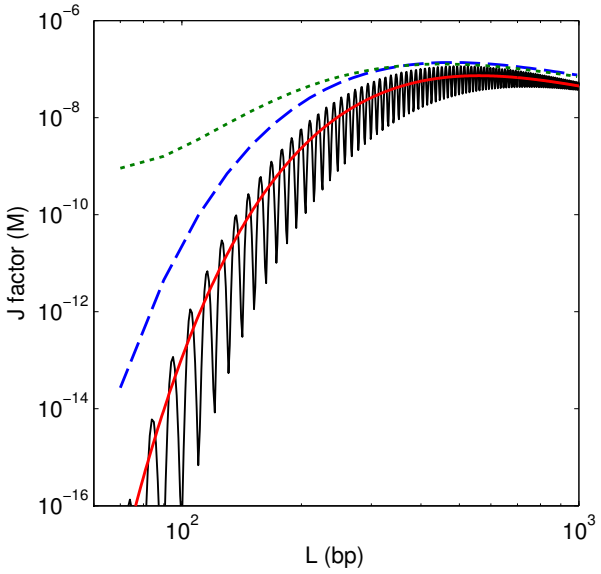


Figure 1.3: J factors predicted from the WLC model as a function of DNA length and the nature of the end geometry. For all calculations presented in this figure, the persistence length was chosen to be 48 nm. For the curve that requires torsional alignment, the torsional constant was set at 2.9×10^{-28} J.m[26] and the helical repeat was 10.5 bp. (Dotted green curve) Two ends require an end-to-end distance of 10 nm. (Dashed blue curve) Two ends require direct contact. (Red curve) Two ends require direct contact and share the same helical axis. (Black curve) Two ends require direct contact, same helical axis, and same helical face through torsional alignment. The approximation for J factor in the last case was derived by Shimada and Yamakawa[23, 27]. Otherwise, we used a semi-analytical estimation of the J factor with an end-to-end-distance constraint derived by Douarche and Cocco[24].

one can get the J factor (Figure 1.4A). In bulk cyclization, $K_{a,circular}$ or $K_{a,bimolecular}$ is hardly accessible. Therefore, the assay employs DNA ligase, which is an enzyme that can ligate the cohesive ends when the DNA is in the looped state (Figure 1.4B). The assay provides the accumulation rates of the products, k_C for ligase-catalyzed ring, and k_D for the ligase-catalyzed dimer formation. It is very important to note that only when the pre-equilibrium between the two states (marked by red boxes in Figure 1.4B) is very fast compared to ligation, the accumulation rates reflect the J factor (j). In such condition, we have

$$j = \frac{K_{a,circular}}{K_{a,bimolecule}} \approx \frac{k_C}{k_D}. \tag{1.10}$$

This approximation is valid when $k_{12} \gg k_{23}[E]$ and $k_{12}^* \gg k_{23}[E]$, where $[E]$ is the effective enzyme concentration, the rates k_{12}, k_{23}, and k_{12}^* are defined in Figure 1.4B. In other words, it requires the rate of substrate decay (k_{21} or k_{21}^*) to be much higher than the

rate of ligase binding to the sites ($k_{23} \cdot [E]$). While the cyclization technique has a good

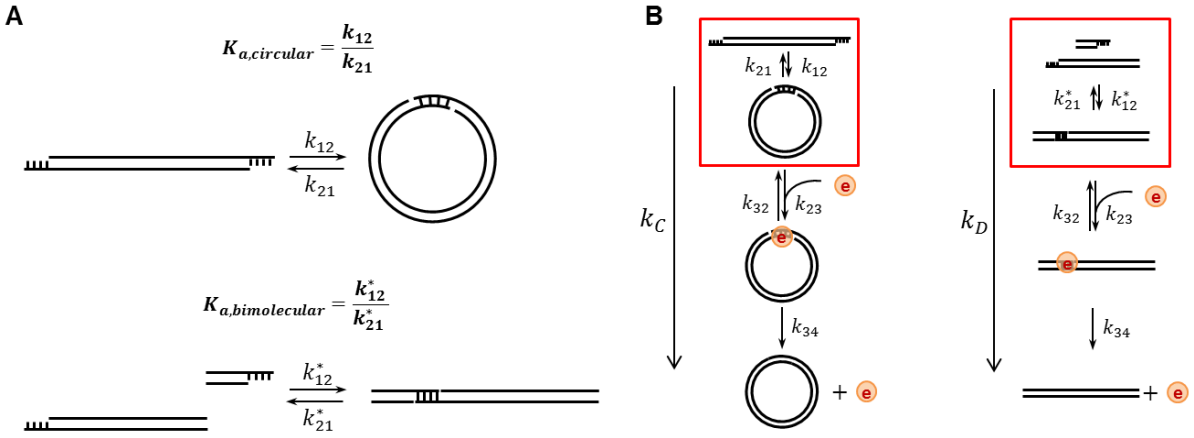


Figure 1.4: Schematic representation of cyclization. (A) (Upper) The equilibrium between the looped and unlooped states of a linear DNA with two sticky ends. Equilibrium constant $K_{a,circular}$ represents the fraction of molecules in the looped state over the unlooped state. (Lower) The association and dissociation of the two halves of the DNA molecule with similar cohesive ends. The equilibrium constants $K_{a,circular}$ and $K_{a,biomolecular}$ define the ring-closure probability or the J factor of the DNA ($j = K_{a,circular}/K_{a,biomolecular}$). (B) A schematic that represents the steps in the process of cyclization and ligation of a linear DNA molecule (left) and in the process of dimerization and ligation of two DNA halves (right). First step is the annealing of the cohesive ends that results in the DNA substates suitable for ligase binding. Second step represents the kinetics of ligase binding, and the last step is an irreversible ligation that produces DNA circles or dimers from the two halves.

sensitivity (Figure 1.3), it relies on a few assumptions as mentioned to correctly extract the J factor. Measurement of the J factor for DNA shorter than 150 bp turns out to be extremely sensitive to the ligase concentration. In fact, this subtlety became the center of controversy in the field, which will be highlighted in the next section.

1.4 Battles for and against the WLC model

Since proposed 60 years ago, the WLC model[9] has been constantly tested and scrutinized for its validity to describe the mechanics of DNA deformation at a wide range of length scales. Ligase-catalyzed cyclization has been used in a number of experiments to measure the ring closure probability, and it was found that the WLC model can faithfully explain the data for dsDNA at 200 bp or above[21, 17, 22] with a canonical value of the persistence length $L_p \approx 50$nm. In 2004, Cloutier and Widom reported a surprising result that the cyclization probabilities of several 94-bp DNA fragments were $3-5$ orders of magnitude

higher than the expected values predicted by the WLC model[28], which suggests that DNA has a superflexible bendability at 100-bp length scale. Their subsequent study also suggested a much more flexible torsional rigidity for short DNA around 100 bp[29]. These unexpected observations prompted the idea of DNA kinking as a molecular mechanism to explain the measured enhanced flexibility of DNA. DNA kinking is a local structural distortion in the double helix that can reduce the energetic cost for bending compared to a smooth, worm-like bending[30]. Shortly after the observations by the Widom group, two extended versions of the WLC model were proposed. The first model is the kinkable worm-like chain (KWLC) model[31] that considers kinks with disrupted base-stacking but intact base-pairing (kink type I). The second model is the meltable worm-like chain model[32] which assumes transiently melted base pairs (kink type II). In both models, the unstacked or melted base pairs can act as flexible hinges to facilitate sharp bending. These disruptions are thermally excited over a certain energy barrier. Strong bending, which imposes large stress on the helical structure, might facilitate the transition from a spring-like bending to kink-like bending. Careful analysis shows that to explain the data from the Widom group, the free energy of kink formation should be as low as $7k_BT$ (reference [33]).

These findings are in accordance with a different study using AFM imaging to analyze the conformations of DNA equilibrated on a surface[34]. By investigating the distribution of the bending angle across many DNA trajectories, Wiggin *et al.* found that highly bent conformations were observed at a higher frequency than the prediction from the WLC model. At a length scale of $\sim$ 10nm, the bending energy was found to depend linearly on the bending angle, $E = k_BT\alpha|\theta|$. The model was called the linear sub-elastic chain (LSEC) model, which could recapture the abnormally high cyclization J factor from the Widom group[35]. It is interesting to note that at larger length scales, LSEC becomes indistinguishable from WLC with a 50-nm persistence length[36]. Furthermore, all-atom molecular dynamics (MD) simulations[37] were performed to study the structures of the 94-bp 601 DNA minicircle[28] with or without torsional stress. The study reported that kinks (in the form of base-pair unstacking) indeed developed even in the torsionally relaxed circles during the simulation time of 80 nanoseconds. Therefore, kinkings appear to be an

inherent response of DNA to strong bending and twisting strain.

However, the WLC model still found support from other experiments. In 2005, Du *et al.*[38] reported a technical flaw in the cyclization experiments from the Widom group. They argued that the ligation condition used by the Widom group could not justify the J factor measurement. Since the cyclization of the $\sim$ 100-bp DNA is very inefficient, one is tempted to increase the ligase concentration to increase the accumulation rate of circular DNA. However, high ligase concentration used by the Widom group violated the condition $k_{12}^* \gg k_{23}[E]$ for dimerization. As a result, the second-order dimerization rate k_{12}^* was underestimated, and therefore, the J factor was severely overestimated. By adjusting the experimental condition, Du *et al.* found that the WLC model could explain cyclization data for DNA as short as 105 bp. Du *et al.* then proposed to use an enzymatic assay with single-stranded DNA nucleases to detect kinks or bubbles in small DNA minicircles[39]. While DNA minicircles at $\sim$90 bp were observed to resist enzymatic digestion, minicircles at $60-70$ bp could be digested. Based on these results, they set the lower limit for the WLC model at 70bp. Furthermore, a cryo-EM study[40] reconstructed the 3D conformations of one of the 94-bp DNA minicircles from Cloutier and Widom study[28] and revealed that the 94-bp DNA minicircle only bends smoothly without sharp kinks. Last but not the least, an MD simulation of small DNA minicircles[41] found results consistent with the DNA minicircle digestion data from the Vologodskii group. These controversies highlighted the importance to understand the mechanics of DNA deformation in the sub-persistence-length regime and called for new experimental strategies to accurately quantify DNA flexibility across a wide range of DNA lengths.

1.5 Reaching a unified view of strong DNA bending

In 2012, Vafabakhsh and Ha used single-molecule Fluorescence Resonance Energy Transfer (smFRET) to measure the J factor of DNA between 67 and 106 bp[42]. This smFRET looping assay did not require ligase since it employed long complementary ssDNA ends ($\sim$ 10 nucleotides) to stabilize the loops through strong base-pairing interaction. High salt conditions (1 M NaCl or 10 mM $MgCl_2$) were used to enhance the annealing of the cohesive

ends, making it possible to track the looping dynamics of extremely short DNA molecules. The authors reported that the J factor of those DNAs were a few orders of magnitude higher than the WLC prediction. Since there was no ligase involved, the results were considered valid compared to previous ligase-dependent studies[43]. However, shortly after the work was published, it was criticized for the possible impurity of the DNA samples, and poorly defined boundary condition for J factor comparison[44, 36].

We have also applied a similar smFRET looping assay[45, 46] in a number of DNA looping studies in various conditions. We first devised a strategy to prepare DNA samples with high purity. Since the intrinsic sequence curvature may enhance the apparent flexibility of short DNAs[47], we constructed several DNA fragments with same size but different intrinsic curvature and measured their J factors. We found that a small but persistent sequence intrinsic curvature could enhance the looping probability. Furthermore, a dinucleotide WLC model that accounted for such bendedness could quantitatively explain the length dependence of the measured looping J factor for several sequences taken from the yeast genome [45].

Next, we developed a new experiment to estimate the mechanical stress stored in DNA loops with size comparable to the other study[42] by measuring loop breakage kinetics[48]. Our method can be applied to conditions with varying ionic strength and type, including low salt concentration, which is inaccessible to other enzyme-based cyclization and digestion experiments. Our results support a KWLC model with salt-dependent kinkability, in which the elastic limit strongly depends on the presence of magnesium. Furthermore, in a condition comparable to other experiments, our model can faithfully explain previous data including bulk and single-molecule cyclization[38, 42] and DNA minicircle digestion[39] experiments, which clarifies the recent controversy in the field of DNA bending mechanics at small length scales.

It is worthwhile to mention that several recent experiments[49, 50] did not support the LSEC model for DNA bending proposed in the other study[34]. It was suggested that an artifact in the tracking algorithm might produce such sub-elastic behavior in the distribution of the bending angles. As will be explained below, our results also eliminated the LSEC

model.

1.6 Structure of the thesis

In Chapter 2, I will introduce single-molecule Fluorescent Resonance Energy Transfer (smFRET) and how we measure it with Total Internal Reflection Fluorescence Microscopy (TIRFM). I will also describe the experimental setup, sample preparation, data acquisition, and data analysis.

In Chapter 3, I will describe one application of the smFRET looping assay to study the looping dynamics dsDNA shorter than 200bp. We focus on the effect of intrinsic sequence curvature and contour length on DNA looping propensity. Furthermore, we perform a computational study to quantitatively explain the experimental data.

In Chapter 4, I will describe a new smFRET assay to study the elastic limit of dsDNA bending. The idea is to estimate the bending stress within a small DNA loop under strong bending via measuring the breakage dynamics of the loop. The relationship between the breakage dynamics and bending stress reveals a transition from the elastic regime to the nonelastic regime with enhanced flexibility. We will show that this limit is salt-dependent and is highly sensitive to the presence of magnesium. We will also show that our data support the KWLC model and is consistent with previous experimental data from cyclization to DNA minicircles digestion.

In the last chapter, I will present some preliminary data on sequence-dependent DNA flexibility in the length scales below 100bp. This extremely strong bending regime will potentially reveal interesting effects on dsDNA bending. We discovered that some DNA patterns with repeating units can exhibit high flexibility, which has some relevance to nucleosome positioning sequences.

CHAPTER II

SINGLE-MOLECULE FLUORESCENCE RESONANCE ENERGY TRANSFER AND TOTAL INTERNAL REFLECTION FLUORESCENCE MICROSCOPY

2.1 Fluorescence resonance energy transfer (FRET)

Förster (or Fluorescence) Resonance Energy Transfer (FRET) is a mechanism that governs the energy transfer between two fluorescent molecules[51, 52]. A donor in its electronic excited state may transfer energy to an acceptor through nonradiative dipole-dipole coupling[53]. FRET is a distance-dependent interaction and the FRET efficiency E_{FRET}, which quantifies the fraction of energy transfer, is inversely proportional to the sixth power of the distance between the donor and the acceptor:

$$E_{\mathrm{FRET}} = \frac{1}{1 + (R/R_0)^6}. \tag{2.1}$$

In this equation, R is the inter-fluorophore distance and R_0 is the Förster distance at which $E_{\mathrm{FRET}} = 0.5$. FRET is extremely sensitive to small change in the distance around R_0 (Figure 2.1A), making it a powerful tool to probe the inter-dye distance between $1 - 8$ nm[54]. Since FRET arises from a nonradiative dipole-dipole interaction between the donor and acceptor fluorescence, it depends on the overlap between the donor emission spectrum and the acceptor absorption spectrum, as well as the relative orientation of the two dipole moments. The Förster distance can be calculated as[52]:

$$R_0 = \left(\frac{9000\, Q_0\, (\ln 10) \kappa^2\, J}{128 \pi^5 n^4 N_{\mathrm{A}}} \right)^{1/6}, \tag{2.2}$$

where Q_0 is the fluorescence quantum yield of the donor in the absence of the acceptor, κ^2 is the dipole orientation factor, n is the refractive index of the medium, and N_{A} is Avogadro's number. The spectral overlap integral J is calculated as

$$J = \int f_{\mathrm{D}}(\lambda)\, \epsilon_{\mathrm{A}}(\lambda)\, \lambda^4\, \mathrm{d}\lambda, \tag{2.3}$$

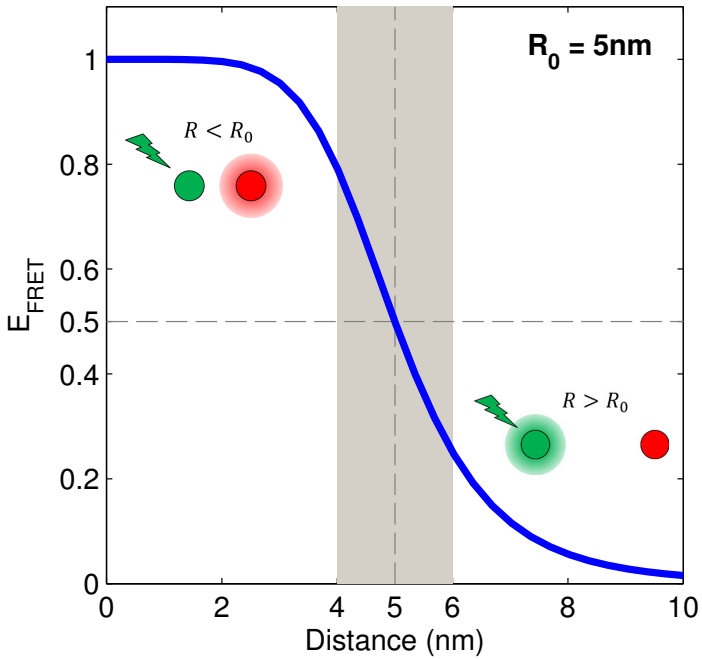


Figure 2.1: FRET efficiency as a function of the inter-dye distance R for the Förster radius R_0 of 5 nm. Shaded area represents the FRET values in the distance range $4-6$ nm which indicates the sensitivity of FRET around R_0.

where f_{D} is the normalized donor emission spectrum, and ϵ_{A} is the acceptor molar extinction coefficient. If the donor and acceptor are allowed to freely rotate in all directions during the excited state lifetime, $\kappa^2 = 2/3$. Otherwise, $0 < \kappa^2 < 4$ always hold. Here, we used Cyanine 3 (Cy3) and Cyanine 5 (Cy5), two common organic dyes with a Förster distance of $\sim$ 5nm[55] as the FRET pair in most of our studies. Having small size, appropriate spectra, comparable brightness, and high photostability in an oxygen-free medium, Cy3 and Cy5 have been widely employed in a number of FRET-based assays to study nucleic acids and protein conformational fluctuations or to study the interaction in several nucleic acids-protein complexes.

The strong distance dependence of FRET makes it a powerful "spectroscopic ruler" in detecting structural changes of biological molecules or the relative motion and interaction between two different macromolecules. In 1996, single-molecule FRET was first demonstrated by Taekjip Ha and coworkers[56]. By collecting the signals from both the donor and the acceptor in a FRET pair, the FRET efficiency can be obtained. If I_{D} is the intensity

of the donor and I_{A} is the intensity of the acceptor in one FRET pair, the FRET efficiency can be calculated as:

$$E_{FRET} = \frac{I_{\mathrm{A}}}{I_{\mathrm{D}} + I_{\mathrm{A}}}. \tag{2.4}$$

Furthermore, a FRET histogram obtained from hundreds of molecules not only gives an average number of E_{FRET} as in a steady-state ensemble measurement, but also provides more information on the distribution of the available conformations. Thus smFRET can facilitate detection of rare or short-lived states that can be easily hidden in a bulk measurement. More importantly, single-molecule FRET adds a new dimension to the problem by providing information on the dynamics of conformational fluctuations of a system in equilibrium[57].

As an example, we consider a double-stranded DNA that can freely deform under thermal excitation. One Cy3 and one Cy5 dye molecules are attached to the two ends of the DNA. The two ends of the DNA can temporarily stick with each other by a specific interaction when they come near each other. In the looped state, the end-to-end distance is almost zero resulting in a high FRET value while in the unlooped state, large end-to-end distance results in low FRET value. The system can be considered as a two-state system in which DNA molecule experiences consecutive looping and unlooping events. An ensemble measurement in steady state would yield an average FRET value and from this number, the relative population of the looped and unlooped states can be estimated only if one can determine the FRET efficiency of DNA samples in each state, which is not always practical in general. In an smFRET measurement, the time-dependent FRET traces from hundreds of DNA molecules can be easily obtained (some representative traces are shown in Figure 2.2A, one trace represents one molecule). One can extract the FRET values from multiple traces at one particular moment and construct a FRET histogram (Figure 2.2B). The FRET histogram shows two peaks which correspond to two DNA conformations, high FRET for the looped state and low FRET for the unlooped state. The dashed gray line in Figure 2.2B represents the average FRET value, which is normally reported by an ensemble measurement.

FRET distribution from several single traces recorded during the whole acquisition time

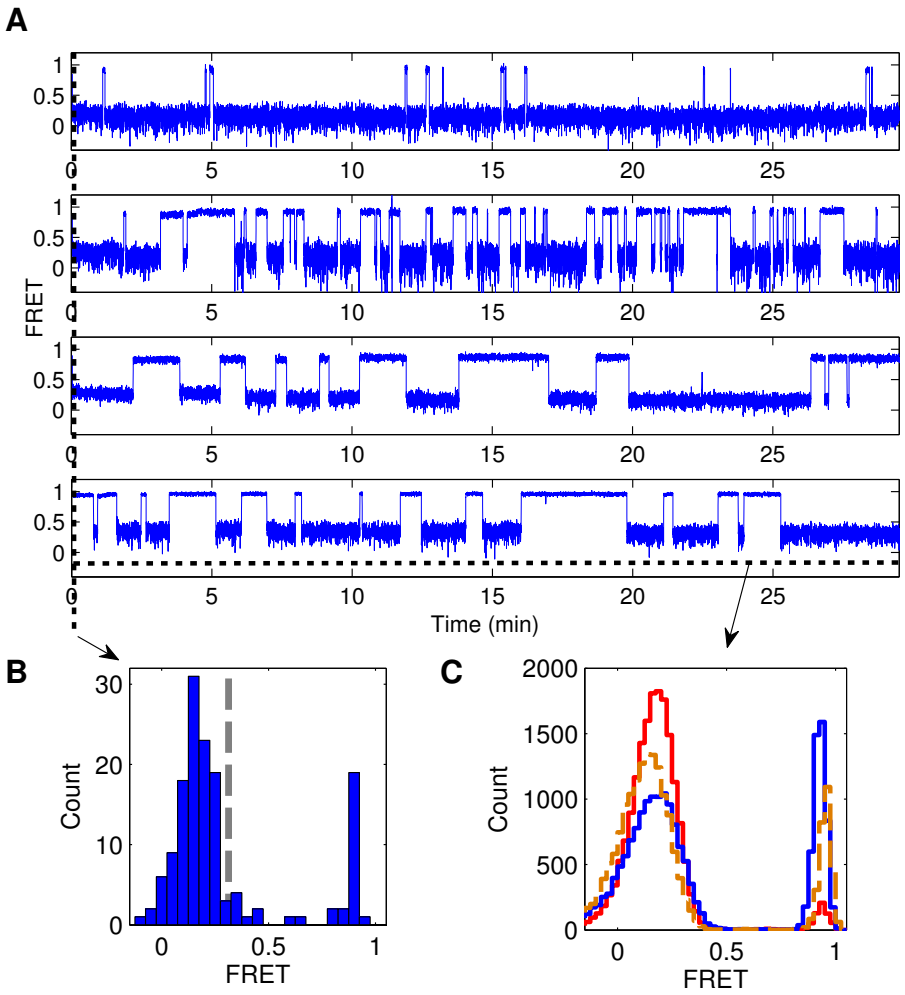


Figure 2.2: (A) Several FRET traces of dsDNA with two ends labeled with Cy3 and Cy5 dyes. Looped and unlooped conformations correspond to high FRET and low FRET states, respectively. The vertical dashed line indicates the moment that the FRET values from all traces were recorded to construct the histogram in (B). The horizontal dashed line indicates the time interval for collecting FRET values from several single molecules to construct the FRET histograms represented in (C). (B) The bulk FRET histogram built from 150 DNA molecules. For each trace, we recorded the mean FRET value of the first ten time frames ($\sim$ 1s). The gray dashed line indicates the mean FRET value, which is ~ 0.3. (C) Several FRET histograms, each curve represents data from one single-molecule trace. The time-lagged FRET values from each trace were binned together to construct this type of FRET histogram.

(Figure 2.2C) are also consistent with the ensemble histogram (Figure 2.2B). Importantly, the kinetics of the looping and unlooping processes can be directly accessed by recording the dwell-times that the DNA stays in the unlooped and looped states, respectively. Therefore, smFRET is unique in the sense that it can provide information on the dynamics of a system under equilibrium. Note that these smFRET traces reveals some inhomogeneity in the looping and unlooping dynamics of several chemically identical molecules (Figures 2.2A). The variation among several molecules becomes more apparent when comparing their single-molecule FRET histograms (Figures 2.2C).

2.2 Experimental design

2.2.1 Single-molecule fluorescence dyes

A common choice of donor and acceptor for the FRET pair is Cy3 as the donor and Cy5 as the acceptor (Figures 2.3A and 2.3B). They are organic dyes with high brightness, small size ($\sim$ 1nm), and have high water solubility. Cy3's emission spectra is significantly overlapped with the excitation spectra of Cy5 (Figure 2.3C), which makes FRET possible between the two dyes. The absorption and emission maximum are at 548 nm and 562 nm for Cy3, and 646 nm and 664 nm for Cy5, respectively. For FRET measurement, Cy3 (donor) was excited with a 532 nm laser (NT66-968, B&Wtek). Cy5 (acceptor) was directly excited with a 640 nm laser (CUBE 640-30FP, Coherent). The large separation between their absorbance spectra minimizes direct excitation of the Cy5 fluorophore from the 532-nm laser. Since the emissions from Cy3 and Cy5 have different wavelengths (562 nm and 664 nm), they can be separated by an appropriate bandpass filter or dichroic mirror (see Section 2.3).

To incorporate Cy3 and Cy5 into nucleic acids, one can use their ester derivatives, which are commercially available. The Cy3 NHS ester and the Cy5 NHS ester (Figures 2.3A and 2.3B) can readily react with the amino groups via acylation reaction. An amine-modified base that contains an amino group can be employed for this purpose. Some popular choice are the amino deoxythymidine dT C6 and amino deoxycytidine dC C6 (C6 is a 6-carbon spacer, which reduces the interaction between the biomolecule and the conjugated dye molecule), see Figure 2.3D. The advantage of these modified nucleotides is that they can

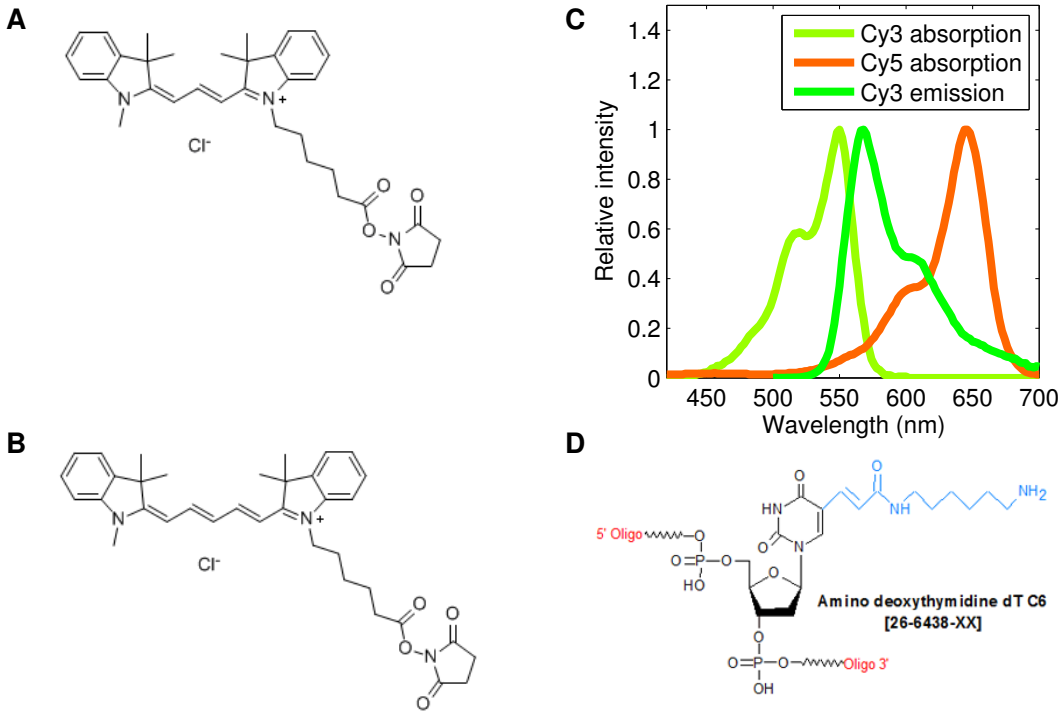


Figure 2.3: (A) Molecular structure of Cy3 NHS. (B) Molecular structure of Cy5 NHS. Images of the molecular structures of Cy3 NHS and Cy5 NHS were taken from `www.lumiprobe.com`. (C) Absorbance spectra of a Cy3-labeled DNA oligo and a Cy5-labeled DNA oligo. The emission spectrum of the Cy3-labeled oligo is also shown. These spectra were measured for the labelled DNA oligos with a concentration of 25 μM using a spectrofluorometer Synergy H1 Multi-Mode Reader from BioTek. The intensities were normalized by the maximum values. (D) Molecular structure of Amino Deoxythymidine (dT) C6 (Picture taken from `http://www.genelink.com/`).

preserve the base-pairing interaction and therefore, the dsDNA backbone is not disrupted by these modifications. Alternatively, Cy3 and Cy5 can be internally inserted into the DNA backbond via phosphoramidite chemistry but the intercalated dye disrupts one base-pairing interaction[58]. In this study, we prepared the labeled DNA using a biochemical technique called Polymerase Chain Reaction (PCR) which amplifies dsDNA from a DNA template and incorporates Cy3 or Cy5 from Cy3 or Cy5-labeled ssDNA primer. The primers were ordered from IDT but their preparation follows the two methods mentioned above. More information on DNA preparation procedure is described in Chapter 3.

2.2.2 Enhancing photophysical stability

Fluorescence dyes are not ever-glowing light bulbs. In fact, organic dyes commonly used in single-molecule studies of biological systems usually display photobleaching within a couple of seconds under excitation (Figure 2.4, top panel). Fluorophores like Cy5 also show some fast intensity fluctuations on the millisecond timescale, which is called photoblinking. These factors could complicate the interpretation of the data and limit the observation time. For example, one can mistakenly interpret photoblinking of Cy5 resulting in a zero FRET value as indicative of large distance between the donor and the acceptor. Short dye lifetime also hinders the detection of rare events and reduces the statistical accuracy of the measurement due to smaller data set. Therefore, a condition that could extend and stabilize the fluorescence emission is highly desirable.

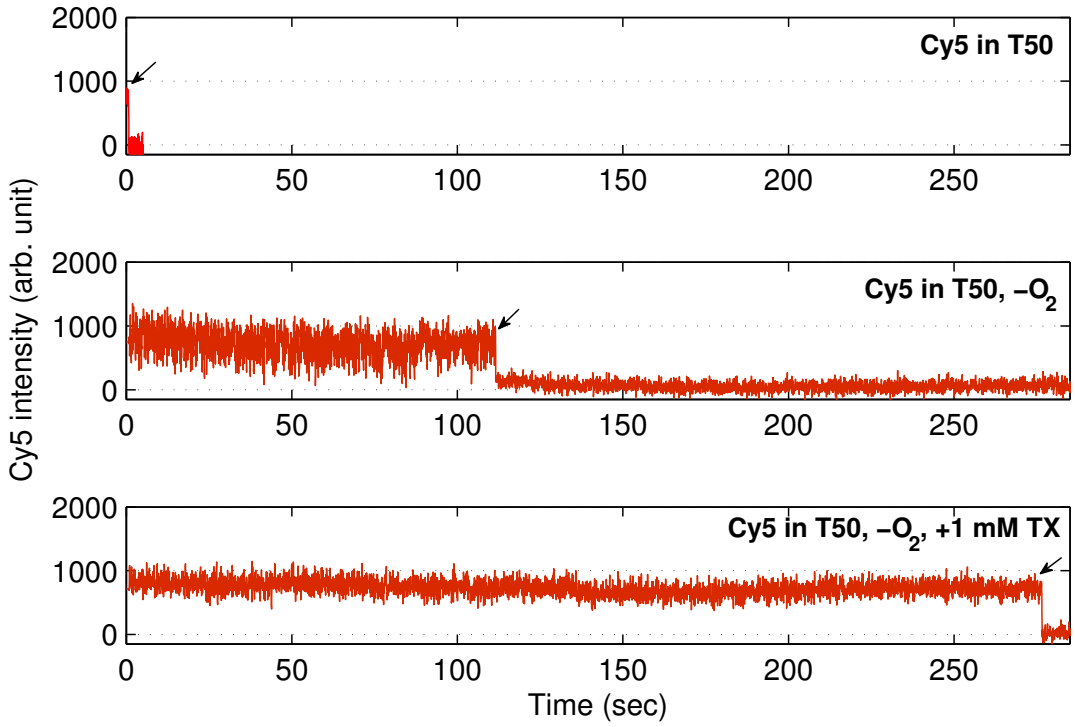


Figure 2.4: Effects of the oxygen scavenging system and Trolox on the photophysics of Cy5-conjugated DNA molecules. Fluorescence intensity traces taken with 40-ms resolution with direct excitation (640 nm) in T50 buffer only (top panel), in T50 buffer with oxygen scavenging system but no Trolox (middle panel), and in T50 buffer with the oxygen scavenging system and 1mM Trolox (lower panel). Arrows indicate the photobleaching of the fluorophore.

Fluorophores photobleach when oxygen oxydizes exposed chemical groups of organic dyes, making it no longer fluorescent. Therefore, an enzymatic oxygen scavenging system is added to the buffer to prolong the dye's lifetime (Figure 2.4, middle panel). The oxygen scavenging system employed in our study contains 5mM PCA (3,4-Dihydroxybenzoic acid) and 100nM PCD (Protocatechuate 3,4-Dioxygenase), which was shown to lower the oxygen level to only $\sim 3\mu$M[59]. PCD catalyzes the reaction of PCA with oxygen and produces additional proton H^+. But, additional protons only induce a negligible decrease in the pH of the imaging buffer over long period of incubation (up to an hour) in an air-tight chamber[60].

While the removal of oxygen extends the fluorophore's lifetime, it also causes frequent fast blinking in the dye's fluorescence (Figure 2.4, middle panel). This is because oxygen is a quencher of the triplet states. Trolox, which is the commercial name for 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid, is a water-soluble analog of vitamin E, which eliminates the photoblinking due to triplet state transition as well as blinking occurring on longer timescales[61]. Introducing Trolox at 1mM, along with an oxygen scavenging system like PCA/PCD, dramatically increases the photophysical stability of Cy3 and Cy5 and extends their lifetimes by several orders of magnitude (Figure 2.4, lower panel). Since Trolox hardly dissolves in H_2O, I first dissolved Trolox powder in 1M Tris pH 10 and adjusted the volume to make a 10mM (10X) solution at pH 7.0. The solution should be stirred by magnetic beads in the presence of open air and room light for ~ 24 hours since a partially degraded Trolox solution was found to have a much better quenching effect compared to a freshly dissolved Trolox solution[62].

2.3 Single molecule detection

2.3.1 Total internal reflection and the objective-type total internal reflection microscopy

Single molecule FRET is often combined with a Total Internal Reflection Fluorescence Microscopy (TIRFM) setup to observe hundreds of molecules at a time with a high signal-to-noise ratio and low background intensity. TIRFM uses an induced evanescent wave to selectively illuminate and excite fluorophores in a narrow region of the specimen immediately

adjacent to the glass-buffer interface. The evanescent wave is generated only when the incident light is totally internally reflected at the interface. The evanescent electromagnetic field decays exponentially from the interface and can penetrate to a depth of only a few hundreds of nanometer into the sample medium.

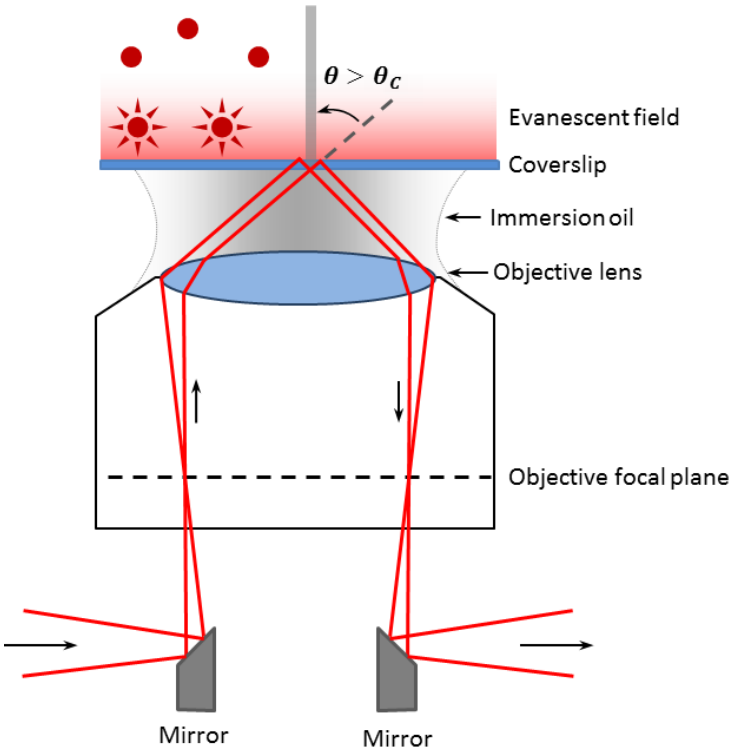


Figure 2.5: An objective-type fluorescence microscope with two small mirrors to reflect the incoming and outgoing excitation laser. After reflection from the first mirror, the incoming laser beam focuses at the back focal plane of the objective lens before reaching its periphery. With the incident angle larger than a critical angle, the beam gets totally internally reflected at the coverslip-sample medium interface and induces an evanescent field that can excite only molecules in the vicinity of that interface. The reflected beam passes through the other side of the objective and gets reflected away by the second mirror.

In the objective-type TIRFM, the laser beam is focused off-axis at the back focal plane of the objective lens. After exiting the objective lens, the collimated beam passes through the immersion oil and into the cover slip (Figure 2.5). After propagating through the glass cover slip, the excitation light beam encounters an interface with the aqueous sample of different refractive index and its direction changes according to the Snell's law. If the angle of incidence of the excitation light on the interface, θ, is greater than the *critical angle*, the

light beam undergoes total internal reflection and does not propagate into the sample but gets reflected back. The critical angle, θ_c, is given by:

$$\theta_c = \sin^{-1}(n_1/n_2), \tag{2.5}$$

where n_1 and n_2 are the refractive indices of the sample and the cover slip, respectively. To achieve TIR, the refractive index of the sample must be less than that of the cover slip. Typically, $n_1 = 1.33$ and $n_2 = 1.5$, so $\theta_c = 62 - 63°$. When $\theta > \theta_c$, the excitation light gets reflected off the coverslip-sample interface and induces a narrow electromagnetic field in the aqueous medium which is called the evanescent field. Since the intensity of the evanescent field decays exponentially with distance from the interface, only fluorophores close to the interface will be excited (Figure 2.5). This eliminates the background fluorescence from molecules in the bulk solution and dramatically enhances signal-to-noise. The penetration

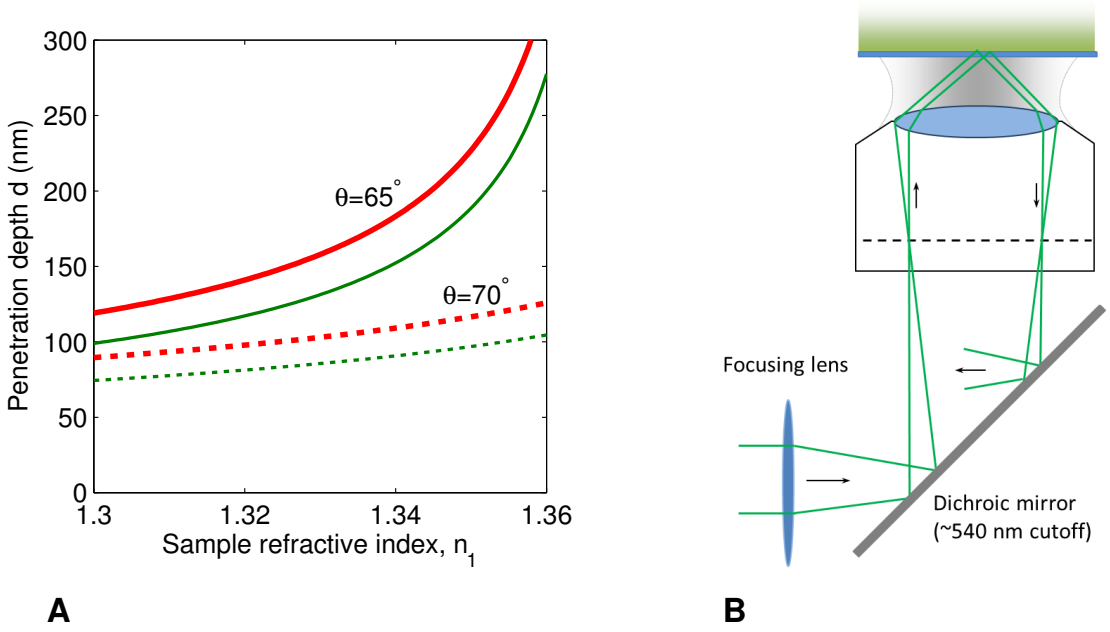


Figure 2.6: A The penetration depth of the evanescent field as a function the refractive index of the sample. Two wavelength values were used here: the green 532 nm and the red 640 nm. Incident angles were 65° and 70° as indicated in the figure. (B) A conventional objective-type TIRFM setup with an additional dichroic mirror. The dichroic mirror is used to reflect the incoming and outgoing excitation light and transmit the emission signal.

depth of the evanescent field, d, is given by:

$$d = \frac{\lambda_0}{4\pi\sqrt{n_2^2 \sin^2\theta - n_1^2}}. \tag{2.6}$$

where λ_0 is the wavelength of the excitation light in vacuum. Typical values for d are in the range $100 - 200$ nm (Figure 2.6A).

2.3.2 Objective-type TIRFM setup

A conventional setup of the objective-type TIRFM[58] requires an additional dichroic mirror mounted right under the objective to reflect the incoming and outgoing excitation laser (usually at 532 nm) and transmit the emission (Figure 2.6B). In the need of incorporating additional excitation wavelengths, the dichroic should have multiple bandpass regions, which is not always commercially available. Here we implemented a different setup[63] to avoid the dichroic mirror. In this setup, two tiny elliptical mirrors were placed under the objective back aperture. The first one is to direct the excitation laser into the objective. The second one reflects away the outgoing excitation laser to prevent it from entering into the emission detection pathway (Figure 2.5). The mirrors were mounted on two miniature translational stages that can move the mirrors horizontally in the direction that passes through center of the objective lens. By translating the beam sideways, the angle of refraction can be adjusted until total internal reflection happens. Figure 2.7A illustrates the excitation pathways of the 532-m and the 640-nm lasers which includes a telescope to create a collimating beam at an appropriate size and a combining dichroic mirror to merge the two laser pathways. The combined beam is then directed toward the miniature mirror by a focusing lens. Using an achromat at this step is critical since it will minimize the chromatic aberration at the back focal plane of the objective lens. The size of the miniature mirrors should be small enough so as not to block the emission signal from fluorescent molecules and still prevent the excitation laser from leaking into the detection optics.

2.3.3 Emission detection for dual-view imaging

For FRET experiments, the fluorescence emission from donor and acceptor molecules have to be separated by their wavelengths. The donor and the acceptor emissions were split into

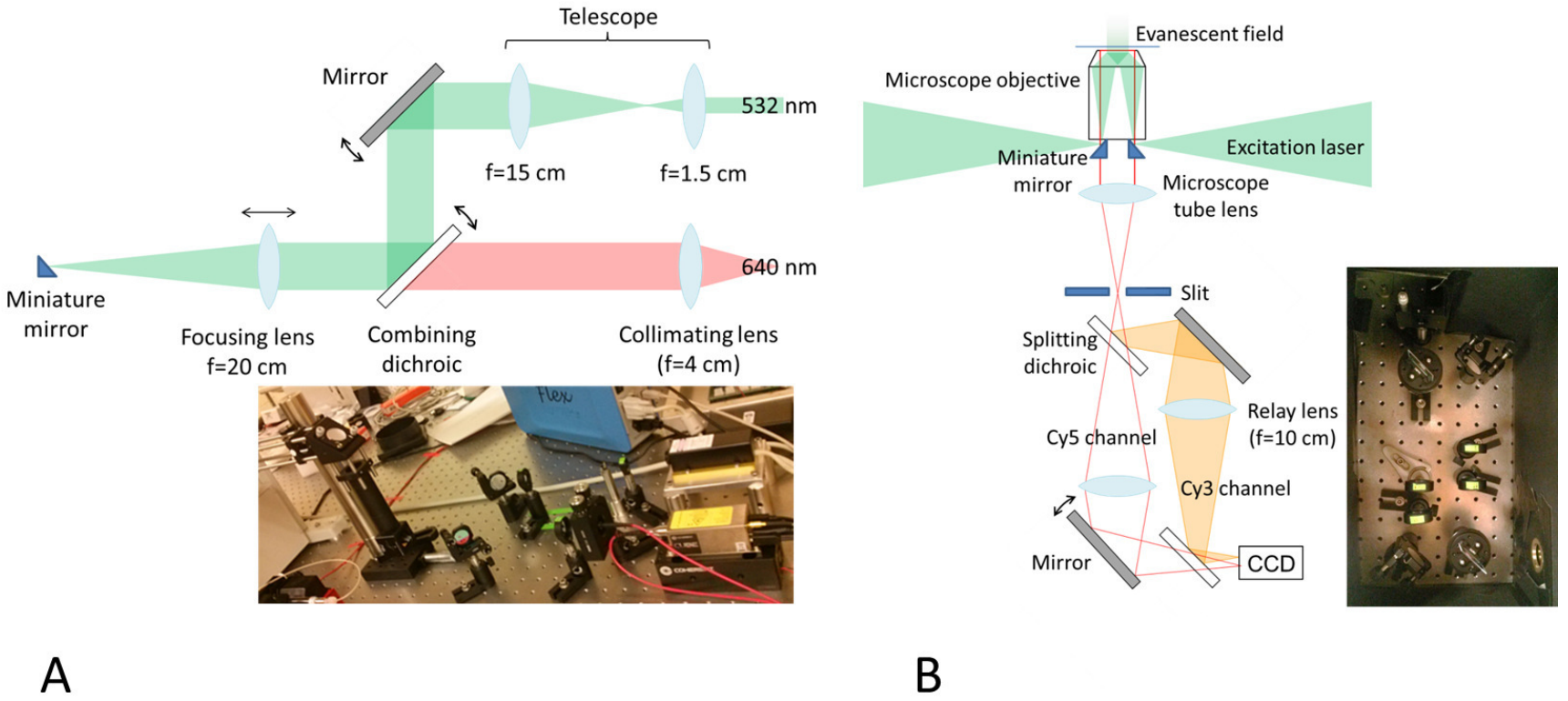


Figure 2.7: Objective-type TIRFM. (A) Excitation optics. The 532-nm and 640-nm lasers are collimated separately to a similar diameter, merged into the same path by a dichroic mirror, and focused onto a tiny elliptical mirror placed underneath the objective back aperture. It is important to use an achromatic lens for focusing to minimize chromatic aberration. (B) Microscope optics. The lateral position of the miniature mirror has to be finely adjusted so that it can reflect the laser beam through the objective while minimally blocking fluorescence emission. Therefore, it must be mounted on a small translation stage. Another mirror on the opposite side reflects the outgoing laser to prevent it from entering the detection optics. At the image plane outside the microscope body, an adjustable slit is placed to crop the image to half the size of the detection area of the CCD. The fluorescence emission is split by a dichroic mirror into two channels, and the relay lenses form the second image on the CCD with 1:1 magnification. One of the reflecting mirrors is slightly rotated to offset the donor and acceptor images. A bandpass and a longpass filters are used to reduce the background in the donor and acceptor channels, respectively.

two separate paths by a beam splitting dichroic mirror with a 650-nm edge (FF545/650-Di01-25x36, Semrock), see Figure 2.7B. The two channels use relay lenses to form the second image on the CCD with 1:1 magnification. To capture the donor and acceptor images by a single CCD, an adjustable slit is placed to crop the image to half the size of the detection area of the CCD and a reflecting mirror is slightly rotated to offset the donor and acceptor images along the horizontal direction. The two beams were merged by the second dichroic mirror and projected on two halves of an EMCCD (DU-897ECS0-# BV, Andor). Additional long-pass emission filters (LM01-552 for donor emission, FF650-Di01 for acceptor emission, Semrock) are used to reduce background signal. For alignment of the optics, it is helpful to illuminate a microscope slide with a scale etched on the surface and visualize the image coming out of the side port. It is important to make sure that all optics are centered along the light path and they are large enough not to clip the image. It is also critical to make both images focus at the exact same plane. This was done by slightly moving the lenses around until the images of the scale bar appear to be focused on both channels simultaneously.

2.3.4 Temperature control of the sample via the objective

Temperature control is essential for acquiring reproducible kinetic data. For temperature control, the objective is separated from the nosepiece of the microscope body to minimize heat transfer, and water from a temperature controlled chiller/heater is circulated through a brass collar that tightly fits around the interior metal beneath the objective jacket. This setup is able to achieve robust temperature control at the coverslip surface between 15 and 50°C (Figure 2.8). In this work, the sample temperature was maintained at 23° C.

2.4 Sample preparation to data acquisition

2.4.1 Surface immobilization

Surface immobilization is critical for smFRET experiments since the molecule has to be immobilized specifically onto the coverslip's surface (in objective-type TIRM) and its biological function should not be perturbed. Molecules such as DNA and protein can easily adsorb to an untreated glass surface without any passivation. A common surface passivation

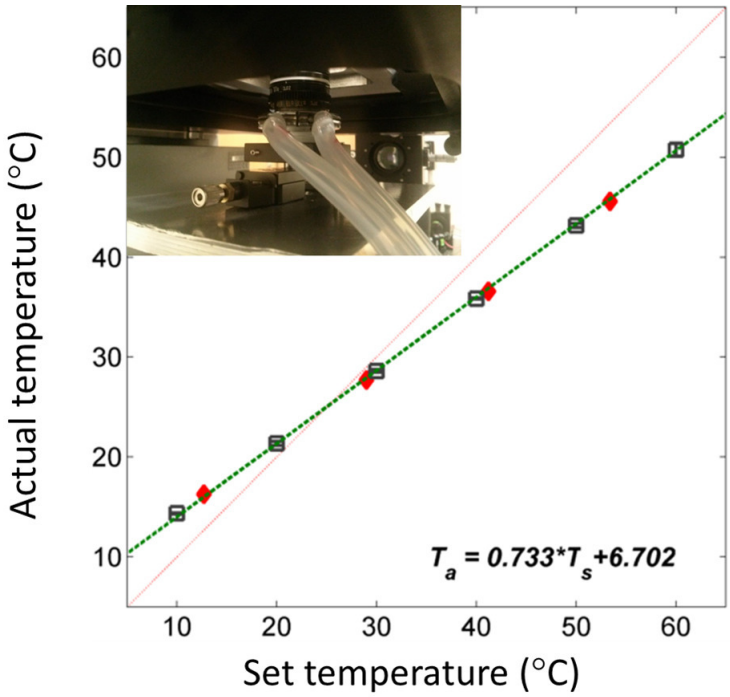


Figure 2.8: Temperature control for single-molecule experiments. The set temperature (T_s) is the reading on the water chiller/heater. The actual temperature (T_a) is measured by a thermocouple contacting the top surface of the coverslip on the microscope. The plot of the actual temperatures vs. six different set temperatures (black squares) shows excellent linearity (green dashed line). The robustness of the controller is shown by random measurements made at later times (red diamonds). The inset is the picture of the temperature controlling unit and the objective in their final assembly. The red line represents a slope of unity.

method for studies involving dsDNA is to use biotinylated bovine serum albumin (biotin-BSA)[57]. Biotin-BSA molecules adsorb to the glass surfaces and bind the biotinylated molecules through the multivalent NeutrAvidin protein (Figure 2.9A). They also repress additional nonspecific binding to the glass surface. For studies involving *stickier* molecules such as ssDNA and protein, for which the nonspecific binding is more severe, polyethylene glycol (PEG) can be used. When adsorbed to the surface, the PEG molecules create a dense brush-like polymer lawn (Figure 2.9B), which efficiently protects the surface from nonspecific binding[64].

In the original method to prepare the PEG surface[64], a pre-cleaned slide is aminosilanized by 2-aminopropyltriethoxysilane to connect primary amines to the glass surface (step (1), Figure 2.9C) and reacted with the N-hydroxysuccinimide (NHS) estermodified

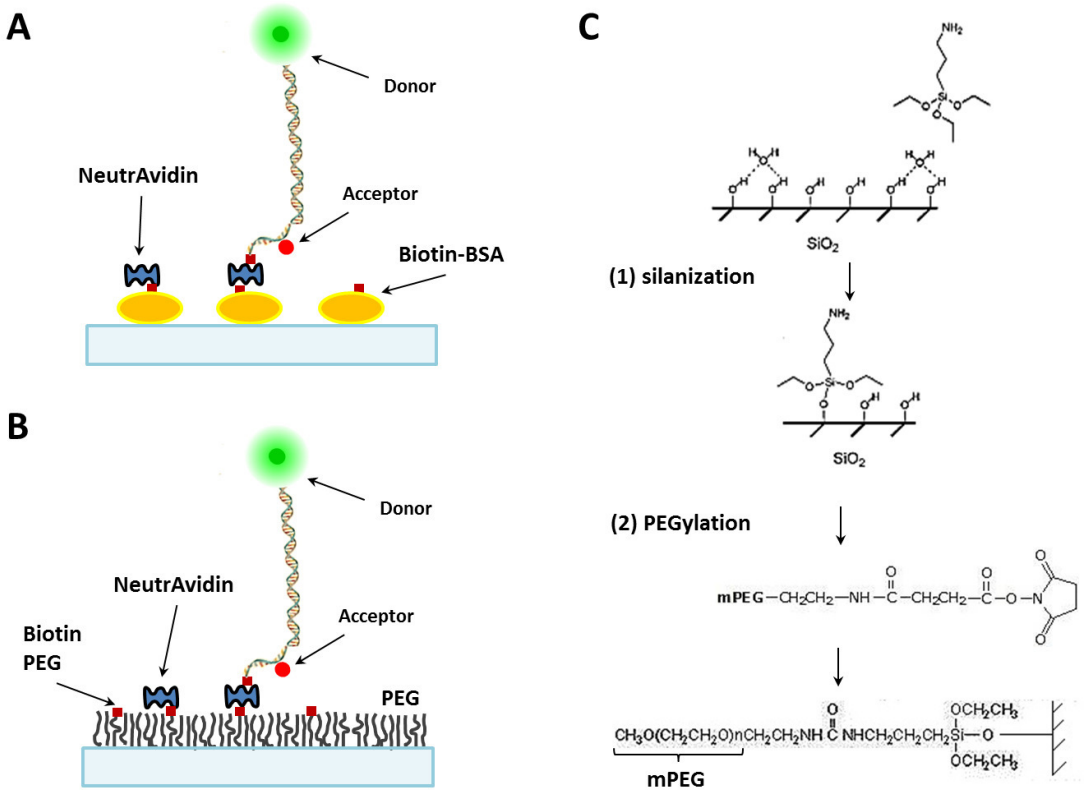


Figure 2.9: (A) Biotin-BSA surface. (B) PEG surface. (C) 2-step PEGylation protocol.

PEG (step (2) in Figure 2.9C), which also includes a small fraction of biotin-PEG-NHS ester for specific tethering. Since the second reaction processes very slowly, long incubation (several hours to overnight) is needed. An incomplete reaction will lead to a partially PEGylated surface which reduces the quality of the passivation. Therefore, a second row of PEGylation reaction may be needed[65]. Instead, we directly used a mixture of mPEG-silane and biotin-PEG-silane to process the PEGylation[66]. This reduces the time of surface treatment since the hydrolysis reaction during the aminosilanization (step (1) in Figure 2.9) is much less time-consuming. We also tried a new surface passivation method[67] using a DDS/Tween-20 surface and confirmed that the new method is comparable or better than PEG surface in terms of suppressing non-specific binding and preserving biomolecule activities.

2.4.2 Flow cell preparation

The sample chamber was prepared following a standard protocol[68, 58]. In short, the flow cells were constructed by sandwiching double-stick tape between a precleaned glass slide and a coverslip and applying epoxy at the edges of the channels. Note that on an objective-type TIR microscope, molecules on the coverslip are observed. Please refer to Figure 2.10A and the following steps for more details on the preparation of a functioning imaging chamber.

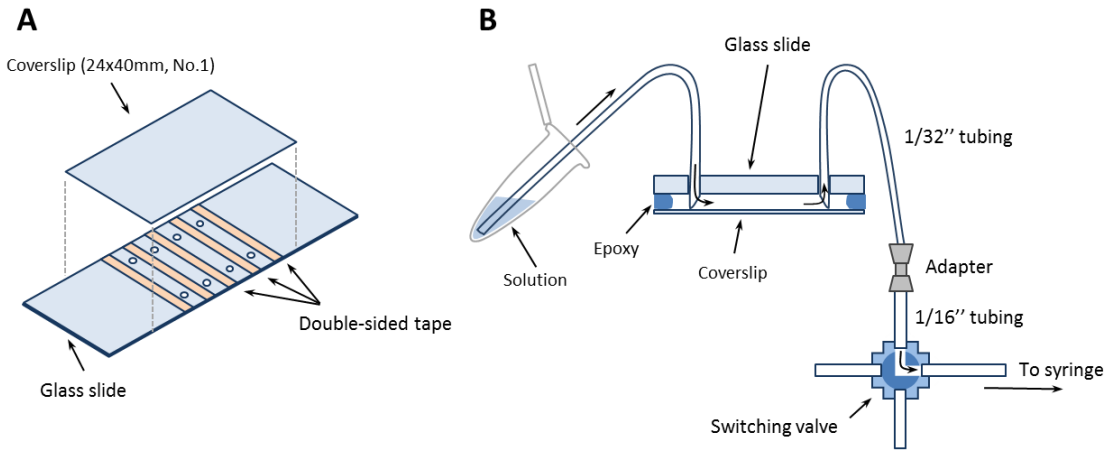


Figure 2.10: (A) Flow cell assembly. Holes are made with a 0.75mm-diameter drill bit (Starlite Industries). A 200 μL pipette tip can fit snugly into the hole for solution injection. (B) Assembly for automated solution injection. Two pieces of 1/32" tubing cut at an angle are fit into the inlet and outlet holes and affixed by 5-minute epoxy. A 1/32-to-1/16 microtight adapters (Upchurch Scientific) is used to connect the 1/32" tubing to the 1/16" tubing. The flow, which can be veered by a right angle flow switching valve (Upchurch Scientific), is pulled into the chamber by an air-tight glass syringe (Hamilton) via an automated pump. The syringe and pump are not shown in this figure.

1. Drill 6-7 pairs of holes along two opposite edges of a glass slide (3"×1") using a drill press and diamond drill bits. After drilling, rub the slide in flowing water to remove visible glass powder. Note: The holes serve as perfusion inlets and outlets. During drilling, cooling the slide with water is important to prevent cracking.

2. Place the slides upright in a glass jar and fill it with water. Sonicate for 15 min and transfer them into another glass jar dedicated for acetone cleaning. Fill it with acetone and sonicate for 15 minutes. Rinse the slides with ethanol using a spray bottle and then with water. Place them in a polypropylene jar, fill it with 5 M potassium

hydroxide, and sonicate for 15 minutes. Finally, sonicate the slides in water for 15 min. Clean the coverslips (No. 1, 24×40 mm) using the same protocol. Note: Cleaned slides and coverslips can be stored in distilled H_2O for long term use.

3. Mix 1 mg of biotin-PEG-silane (MW 3400) with 80mg of mPEG-silane (MW 2000) in 340μL of 0.1 M sodium bicarbonate solution. Mix well and centrifuge the mixture briefly to get rid of bubbles.

4. Put 80μL of the PEG solution on each slide and gently lower a coverslip over it. Wait for 45 min. Separate the coverslip from the slide with tweezers, rinse them with copious amount of dH_2O and let them dry in open air.

5. Place thin strips of double-stick tape across the slide to form channels. Carefully identify the passivated side of the coverslip and align it over tape. Firmly press the coverslip against the slide to form liquid-tight channels. Use 5-minute epoxy to seal the edges of the channels.

To exchange the imaging buffer, we connect the chamber to a syringe by using two pieces of tubing as the perfusion inlet and outlet (1/32-inch OD, Fisher). The tubing, which was cut at an angle ($\sim 30^\circ$) on one end, was fit into a hole on the glass slide. Small amount of epoxy was applied to affix the tubing. Note that the cut end allows a smooth flow across the channel (Figure 2.10B). A right angle flow switching valve was used to have additional control on the flow. New buffer can be introduced into the chamber by pulling a syringe via an automated pump.

2.5 Data processing and analysis

2.5.1 Data processing

Data acquisition is controlled by a win32 application program written and compiled in Microsoft Visual Studio. The program updates display in real time and saves each frame continuously. The program interacts with the camera (iXon, Andor) by using several dynamic link library files and header files provided by the manufacturer. To automate the shutter control and buffer exchange, a multifunction input/output board (VM140/K8061)

is used. The digital signal from the computer (for example, the on/off keying) is converted into the analog signal such as the low/high applied voltage on the devices with Transistor-Transistor Logic (TTL) gating input, which in turn turns the device off/on, respectively. The images are saved as a binary stream of numbers in order to minimize the size of the output movie file. To increase the signal-to-noise ratio, we use a 2×2 pixel binning.

2.5.2 Data analysis

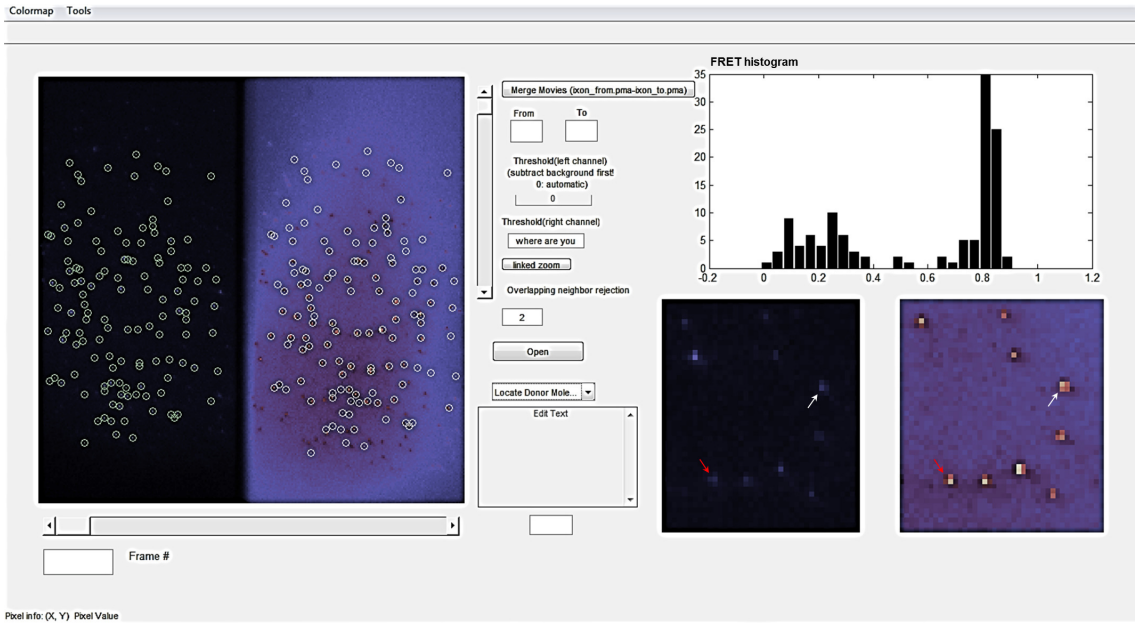


Figure 2.11: A GUI MATLAB program for data analysis. (Left) Two channels with donors (left) and acceptors (right). Small circles indicate single molecule localized by a peak searching algorithm. An affine transformation is used to select molecules with both donor and acceptor. (Upper right) FRET histogram of all molecules selected from an average image of the first 20 frames. (Lower right) Zoom-in pictures of an area showing donor and acceptor of several molecules. Arrows in same color indicate donor and acceptor of the same molecule.

Data analysis is done using MATLAB. A MATLAB GUI (graphical user interface) originally written by Harold Kim is used to handle all data processing post acquisition (Figure 2.11). First, we calculate an average image of donor and acceptor channel from the first twenty frames. We then subtract the local background to obtain an image with zero background level and identify spots that represent single molecules. Single molecule spots of donor and acceptor are localized by using an algorithm that can search for local intensity

peaks, which are well isolated and randomly distributed across the image. To colocalize donor and acceptor fluorescence emissions from the same molecule, one has to map the coordinates on the left-half side to those on the right-half side. This can be done by performing a point mapping via a spatial transformation (Figure 2.11). To generate the spatial transformation, the user manually localizes a few pairs of control points which obviously belong to the same molecule (Figure 2.11, lower right) and record their coordinates. These coordinates are used to generate an 'affine' transformation, which corrects for shearing effect, by the cp2tform function in MATLAB. Finally, the intensities from 3×3 pixels surrounding the peak are integrated to recover donor and acceptor intensities for each single molecule at every time frame.

Table 2.1: List of optics for single-molecule FRET setup using TIRFM and reagents for the DNA looping and unlooping assays

Name	Company	Catalog Number	Comments
Small DNA FRAG Extract Kit-100PR	VWR	97060-558	
Acrylamide 40% solution 500 mL	VWR	97064-522	
Bis-acrylamide 2% (w/v) solution 500 mL	VWR	97063-948	
GeneRuler 100 bp DNA Ladder, 100-1,000 bp	Fermentas	SM0241	
Mini Vertical PAGE System	VWR	89032-300	
Syringe filter 0.2 μm CS50	VWR	A2666	
Trolox	Sigma-Aldrich	238813-1G	triplet state quencher
Protocatechuic acid (PCA)	Sigma-Aldrich	08992-50MG	oxygen scavenging system
Protocatechuate 3,4-Dioxygenase (PCD)	Sigma-Aldrich	P8279-25UN	oxygen scavenging system

Table 2.1 (continued)

mPEG-silane, MW 2,000 1 g	Laysan Bio	MPEG-SIL-2000-1g	
Biotin-PEG-Silane, MW 3,400	Laysan Bio	Biotin-PEG-SIL-3400-1g	
Avidin, NeutrAvidin Biotin-binding Protein	Invitrogen	A2666	
Phusion Hot Start High-Fidelity DNA Polymerase	New England Biolabs	F-540L	
Gel/PCR DNA Fragments Extraction Kit	IBI Scientific	IB47020	
Premium plain glass microscope slides	Fisher Scientific	12-544-1	
VWR micro cover glass, rectangular, no. 1	VWR	48404-456	
Fisher Scientific Isotemp 1006S Recirculating Chiller/Heater	Fisher Scientific		temperature control
Objective Cooling Collar	Bioptechs	150303	temperature control
KMI53 Biological Micrometer Measuring Stage	Semprex	KMI53	
High Performance DPSS Laser 532 nm 50 mW	Edmund optics	NT66-968	Cy3 excitation
CUBE Fiber Pigtailed 640 nm, 30 mW, Fiber, FC/APC Connector	Coherent	1139604	Cy5 excitation
650 nm BrightLine Dichroic Beamsplitter	Semrock	FF650-Di01-25x36	splitting dichroic

Table 2.1 (continued)

LaserMUX Beam Combiner, reflects 514.5, 532, 543.5 nm lasers, 25 mm	Semrock	LM01-552-25	combining dichroic
Brightline Fluorescence Filter 593/40	Semrock	FF01-593/40-25	Cy3 emission filter
635 nm EdgeBasic LWP longpass Filter, 25 mm	Semrock	BLP01-635R-25	Cy5 emission filter
EMCCD iXon+	Andor Technology	DU-897E-CS0-#BV	
IX51 inverted microscope frame	Olympus		
Objective UApo N 100X/1.49 Oil TIRF	Olympus		
Immersion oil type-F for fluorescence microscopy	Olympus	IMMOIL-F30CC	
2 mm Diameter 45○ Rod Lens Aluminum Coated	Edmund optics	54-092	miniature mirror
1/4" Travel Single-Axis Translation Stage	Thorlabs	MS-1	translation of miniature mirror
⌀1" Achromatic Doublet, ARC: 400-700 nm, f=200 mm	Thorlabs	AC254-200-A	focusing lens
Adjustable Mechanical Slit	Thorlabs	VA100	
Dielectric Mirror	Thorlabs	BB1-E02	
⌀1" Achromatic Doublet, f = 100 mm	Thorlabs	AC254-100-A	relay lens
Lens Mount for ⌀1" Optics	Thorlabs	LMR1	
Dichroic Filter Mount	Thorlabs	FFM1	
Fixed Cage Cube Platform	Thorlabs	B3C	
Kinematic Mount for ⌀1" Optics	Thorlabs	KM100	

Table 2.1 (continued)

N-BK7 Plano-Convex Lens, ⌀1", f = 40 mm	Thorlabs	LA1422-A	collimating lens
N-BK7 Plano-Convex Lense, ⌀6.0 mm, f = 15 mm	Thorlabs	LA1222-A	telescope lens
N-BK7 Plano-Convex Lense, ⌀6.0 mm, f = 150 mm	Thorlabs	LA1433-A	telescope lens
1/32-inch OD tubing	Fisher Scientific	50-121-771	
GASTIGHT Syringe luer lock 5 mL	Hamilton	60375-555	
4-Way Valve PEEK L Flow	Upchurch Scientific	V-100L	
1/16 to 1/32 adapter, MicroTight, PEEK	Upchurch Scientific	P-881	

CHAPTER III

SINGLE-MOLECULE LOOPING ASSAY OF SHORT DOUBLE-STRANDED DNA TO STUDY THE EFFECT OF SEQUENCE CURVATURE ON DNA LOOPING

Recently, several studies have shown that short dsDNA loops more readily than the WLC model predicts. In most of these experiments, the intrinsic bendedness of dsDNA, which in theory can dramatically influence looping dynamics, was either avoided or unaccounted for. To investigate the effect of the shape of dsDNA on looping dynamics, we characterized the shapes of several synthetic dsDNA molecules of equal length but different sequences using gel electrophoresis. We then measured their looping rates using a FRET-based assay, and extracted the looping probability density or the J factor (j_M). We also used several dinucleotide angular parameter sets derived from the observed electrophoretic mobility to compute the j_M predicted by the WLC model for comparison. We found a strong correlation between curvature and j_M, although the measured j_M was higher than most dinucleotide model predictions. This result underscores the importance of determining the intrinsic shape of dsDNA to properly quantify the bending flexibility of short dsDNA.

3.1 Introduction

We start this chapter by briefly reminding the readers about the current controversy in quantifying DNA flexibility at short length scales. A more detailed discussion can be found in Chapter 1. Spontaneous looping dynamics of dsDNA has been traditionally measured by ligase-dependent cyclization[16]. In this assay, dsDNA molecules with 'sticky' (cohesive) ends are circularized or dimerized by DNA ligase. By comparing the rates of circle and dimer formation, the J factor (j_M) can be quantified in terms of an effective molar concentration of one end of the DNA in the vicinity of the other end. Widom *et al.* showed that j_M of dsDNA shorter than the persistence length was much higher than predicted by the WLC model[28]. Although this conclusion was questionable due to the high ligase concentration used[38], several other groups also reported that dsDNA loops more frequently than a worm-like chain using different methods[29, 34, 69, 70, 71]. However, most of these kinetics studies involved proteins which could affect mechanical properties of dsDNA due to their nonspecific binding[72, 73]. To overcome this issue, Vafabakhsh and Ha used a FRET-based single-molecule assay to measure looping dynamics of short dsDNA in the absence of proteins[42]. They measured j_M to be a few orders of magnitude higher than the WLC model prediction for less than 100-bp dsDNAs.

In most dsDNA looping studies, j_M vs. length relationship was used as the litmus test for the WLC model. The intrinsic bendedness of dsDNA was usually avoided or ignored because permanent bendedness of dsDNA can in principle lead to abnormally high j_M at short length scales[47]. Moreover, it is difficult to experimentally determine the exact shape of dsDNA in the ground state to use in j_M calculation. This poses the question: to what extent does curvature affect the looping dynamics of dsDNA? Here we designed several dsDNAs of equal length but different shapes and used a similar FRET-based assay to Vafabakhsh and Ha[42] to measure their j_M. To compute j_M predicted by the WLC model, we used the dinucleotide chain model with angular parameters optimized by the gel electrophoretic mobility of these dsDNAs. We showed that measured j_M values were in qualitative agreement with the dinucleotide WLC model although their absolute values were higher than *in silico* predictions in most cases. Our results suggest that mechanics of

large-angle dsDNA bending may not be compatible with that of small-angle dsDNA bending at least in the dinucleotide chain framework.

3.2 Experimental Design

We used two kinds of sequences: repeating artificial sequences and non-repeating genomic sequences. The repeating sequences were variations of sequences chosen from a large-scale nucleosome occupancy study by Kaplan *et al.*[74]. To design globally curved DNAs, we

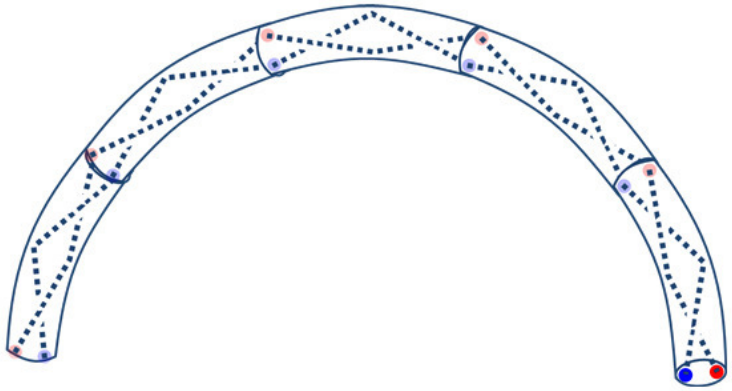


Figure 3.1: Curved dsDNA. A 10-mer dsDNA is represented as a curved tube segment, and the two single strands as dashed lines. The helical axis of any 10-mer DNA is not perfectly straight, and thus each concatenation of such a 10-mer will lead to incremental tilt of the helical axis in the same direction. This effect can be exploited to generate a planar, superhelical molecule.

repeat a 10-mer sequence. Since the helical repeat of dsDNA is close to 10 bp, any net deflection of the helical axis of the 10-mer will accumulate to produce a shape like a circular arc (Figure 3.1). For example, 5'-$(TTTATCATCCTTTATCATCCX)_7$-3' is a 147-bp curved DNA where X is a random extra base. Since the helical period is closer to 10.5 bp, an extra base is inserted after every two repeats (21 base pairs) to keep the curved structure as planar as possible. On the other hand, the 150bp genomic DNA is a generic DNA taken from the yeast genome. We also vary the length of the genomic DNA by shortening it from the 3' end.

All sequences, both artificial and genomic, were terminated at each end with a common 18-bp long adapter sequence for subsequent PCRs. The repeating sequences were commercially synthesized by GenScript. The detailed sequences are given below with the bold

parts as the adapter regions and the part in the parentheses as the repeating monomer. For each sequence, we only show one strand in the $5' \rightarrow 3'$ direction, the other strand is complementary to this strand.

3.2.1 186-bp long synthetic DNA sequences with varying shape (for each sequence, one repeating unit is put in the parentheses)

S1: **GTGCCAGCAACAGATAGC**(TTTATCAGTC)TTTATCAGTCGTTTATCAG
TCTTTATCAGTCGTTTATCAGTCTTTATCAGTCCTTTATCAGTCTTTATCAGT
CATTTATCAGTCTTTATCAGTCATTTATCAGTCTTTATCAGTCTTTTATCAGT
CTTTATCAGTCGTTT**CATTCGAGCTCGTTGTTG**

S2: **GTGCCAGCAACAGATAGC**(TTTATCATCC)TTTATCATCCGTTTATC
ATCCTTTATCATCCGTTTATCATCCTTTATCATCCCTTTATCATCCTTTATCAT
CCATTTATCATCCTTTATCATCCATTTATCATCCTTTATCATCCTTTTATCATC
CTTTATCATCCGTTT**CATTCGAGCTCGTTGTTG**

S3: **GTGCCAGCAACAGATAGC**(TTGATCATCC)TTGATCATCCGTTGATC
ATCCTTGATCATCCGTTGATCATCCTTGATCATCCCTTGATCATCCTTGATCA
TCCATTGATCATCCTTGATCATCCATTGATCATCCTTGATCATCCTTTGATCA
TCCTTGATCATCCGTTG**CATTCGAGCTCGTTGTTG**

S4: **GTGCCAGCAACAGATAGC**(TGACAGCAAC)TGACAGCAACTTGACAG
CAACTGACAGCAACGTGACAGCAACTGACAGCAACGTGACAGCAACTGACAGC
AACCTGACAGCAACTGACAGCAACTTGACAGCAACTGACAGCAACCTGACAGC
AACTGACAGCAACTTGA**CATTCGAGCTCGTTGTTG**

We also designed several sequences with nonuniform curvature from the two sequences S2 and S3: bent at the center (sequence S3-1), bent at the ends (sequence S3-2), and S-shaped (sequence S2-1). See Section 3.3.3 for more details on the sequence features.

S3-1: **GTGCCAGCAACAGATAGC**TTTATCATCCTTGATCATCCGTTGATC
ATCCTTGATCATCCGTTGATCATCCTTGATCATCCCTTGATCATCCTTGATCA
TCCATTGATCATCCTTGATCATCCATTGATCATCCTTGATCATCCTTTGATCA
TCCTTTATCATCCGTTG**CATTCGAGCTCGTTGTTG**

S3-2: **GTGCCAGCAACAGATAGC**TTGATCATCCTTGATCATCCGTTGATC

ATCCTTGATCATCCGTTGATCATCCTTGATCATCCCTTTATCATCCTTTATCAT
CCATTGATCATCCTTGATCATCCATTGATCATCCTTGATCATCCTTTGATCAT
CCTTGATCATCCGTTG**CATTCGAGCTCGTTGTTG**

S2-1: **GTGCCAGCAACAGATAGC**TTTATCATCCTTTATCATCCGTTTATC
ATCCTTTATCATCCGTTTATCATCCTTTATCATCCCTTTATCATCCGATGCTTT
ATCATCCATTTATCATCCTTTATCATCCATTTATCATCCTTTATCATCCTTTTA
TCATCCTTTATCATC**CATTCGAGCTCGTTGTTG**

3.2.2 Genomic DNA sequences with varying length

We varied the length of the genomic DNA by shortening the 3' end of the central 150-bp block without changing the two adapter regions. The range of DNA length is $102 - 186$ bp. We show here only two genomic sequences with 102bp and 186 bp.

186 bp: **GTGCCAGCAACAGATAGC**GCAATTTTAAAATATACAAAAATTAT
ATGTAGTATTTATAATTAGACATTTGTAAAGTGCGTTAAACTAATGATCTAGT
TGTCGTATTCTTCCTTAATTAGTTCTCTCTCCATATCGATCCATTTCTGTTCTG
CGAAGTTTTTGTCCTTA**CATTCGAGCTCGTTGTTG**

102 bp: **GTGCCAGCAACAGATAGC**GCAATTTTAAAATATACAAAAATTAT
ATGTAGTATTTATAATTAGACATTTGTAAAGTGCGTTAAA**CATTCGAGCTCG**
TTGTTG

3.2.3 Making the dsDNA with fluorophore and biotin label using PCR

These sequences were amplified using the following sets of primers:

Primer 1: GTGCCAGCAACAGATAGC,

Primer 2: /5Cy3/**GGTAAATTC**<u>A</u>**C**CAACAACGAGCTCGAATG,

Primer 2': /5Cy3/**TAAATTC**CTACAACAACGAGCTCGAATG,

Primer 3: /5BioTEG/GAAACATA**G**/iCy5/**GAATTTACC**GTGCCAGCAACAGATAGC.

Primer 4: CAACAACGAGCTCGAATG,

We performed two separate Polymerase Chain Reactions (PCRs) on each sequence, one with primer 1 and primer 2 (or primer 2') to integrate Cy3, and the other with primer 3 and primer 4 to integrate Cy5 and biotin. The 10-base long complementary regions in primer

2 and primer 3 stabilize DNA in the looped state (shown in bold). The underlined 'A' in primer 2 is a spacer that compensates for the internally linked Cy5 (iCy5) on primer 3. We also used primer 2' with a shorter complementary region to increase the number of looping events. To construct the dsDNA carrying the FRET pair and biotin, we mixed the two PCR products with an excess amount of Cy3-carrying strand (molar ratio 4:1) and allowed strand exchange by heating and cooling the sample (incubation at 98.5°C for 1.5 minutes, gradual cooling to 5°C with ramp rate of 0.1°C/s, and incubation at 5°C for 2 hours). See a scheme of the procedure in Figure 3.2. More than 90% of surface-immobilized molecules carried Cy5 as expected, and $\geq$ 50% of them carried a Cy3 partner as a result of efficient strand exchange.

3.2.4 DNA electrophoresis

To measure curvature of dsDNA, we implemented PAGE (Polyacrylamide gel electrophoresis). Gel electrophoresis is a method for separation and analysis of macromolecules such as DNA, RNA, and proteins, based on their size and charge. DNA molecules are separated by applying an electric field to move the negatively charged molecules through a gel matrix towards the positive electrode. Shorter molecules move faster and migrate farther than longer ones because shorter molecules migrate more easily through the pores of the gel. Similar to the length effect, curvature can also affect the sequence mobility as it requires more energy to straighten a bent molecule in passing through the gel pores, which causes larger retardation compared to a straight molecule at same size.

We used a similar protocol compared to previously published ones (29.2:0.8 acryl:bis, 5% in TBE buffer, PH 8.4 run at 5-8V/cm in 4°C)[75, 76]. We compared the mobility of a DNA against the bands of a 1 kb DNA ladder (Fermentas) and interpolated its apparent length. The apparent size of the DNA (R) was defined as the ratio of its apparent length to its real length. An electrophoresis image of the synthetic DNAs is shown in Figure 3.9A with their apparent sizes listed in Table 3.1.

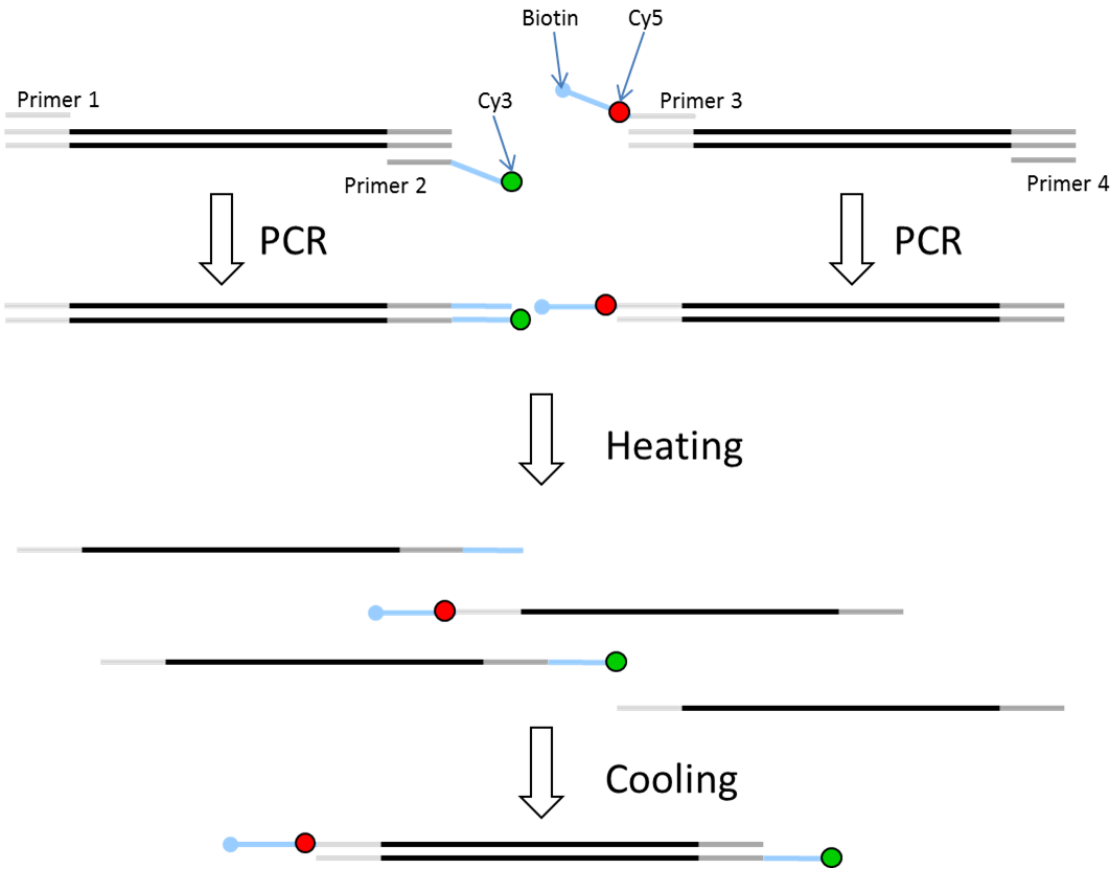


Figure 3.2: PCR-based preparation of dsDNA. Single-stranded DNAs are shown as a horizontal line. Their actual curvature is not depicted here. The central segment in black represents the unique part of the DNA fragment. Shown in light grey are the adapter sequences common to all DNAs used in this study. Primer 1 and 4 anneal to the front and back adapters, respectively. Primer 2 is labeled with Cy3 at the 5'-end. Primer 3 has a biotin tag at the 5'-end and an internally linked Cy5. The light blue regions of Primer 2 and 3 contain complementary sequences that function as sticky ends. Either a plasmid or genomic DNA that contains the sequence of interest is used as the template in two separate PCR reactions. Cy3-labeled dsDNA is produced using primer 1 and 2, and biotinylated Cy5-labeled DNA is produced using primer 3 and 4. These two PCR products are heated and cooled together for strand exchange. Four different dsDNAs are formed, only one of which contains Cy3, Cy5, and biotin.

3.2.5 Image acquisition

The laser power for Cy3 excitation was ~60 μW coming out of the objective. 2×2-binned Images were acquired at 25 frames per second with 40-millisecond exposure using our in-house software. Immediately before each experiment, we injected 10 μl of Neutravidin solution (100 μg/ml) into the channel and waited for two minutes before washing with T50

buffer (10mM Tris-HCl, 50mM NaCl, PH 7.0). DNA molecules were introduced at 50-100 pM for specific binding to the PEG-coated surface through Neutravidin-biotin interaction. Based on our design, the presence of Cy3 on the surface indicates the presence of Cy5 in the same molecule. The channel was finally flushed with the imaging buffer, which contained an oxygen scavenging system (100 mM PCD, 5 mM PCA, 1 mM Trolox slightly modified from [59]) and varying concentrations of NaCl (100-500 mM). All measurements took place at room temperature (22°C).

3.2.6 Data analysis

The movies captured as a series of 256×256 images were processed by a MATLAB code to generate single-molecule FRET traces and dwell-time histograms. The FRET traces were filtered using a 2-point sliding average. A typical FRET histogram showed two peaks which corresponded to the looped state (high FRET value) and the unlooped state (low FRET value). Each histogram was fitted using a double Gaussian function, and the intersection between the two Gaussian curves was used as a hard threshold between the two states. Dwell times in each state were collected from traces showing switching behaviors. For each DNA sequence, ∼500 traces from multiple viewing fields were used to generate dwell time histograms.

In this study, we either reported the lifetimes of the unlooped and the looped states, which are the normal mean dwell times measured in most single-molecule studies, or the mean first looping time. The first looping time is defined as the first passage time for an unlooped molecule to loop for the first time based on the observed FRET signal (Figure 3.3B). Thus, it can be measured from molecules that exist in the unlooped state at the beginning of observation. For this measurement, we used molecules with a shorter 7-bp complementary region so that the majority of the molecules were prepared in the unlooped state (∼99%). Molecules that were either missing one of the two dyes or were in the looped state (high FRET) at the beginning were discarded from analysis. Molecules that underwent Cy3 or Cy5 photobleaching before the first looping event were also discarded from analysis. Since the high FRET state with the 7-bp complementary region was very short-lived (<1s),

Cy5 photobleaching was extremely rare. The typical photobleaching lifetime of Cy3 in these experiments was ∼30-40 minutes, which was substantially longer than all looping times measured in our study.

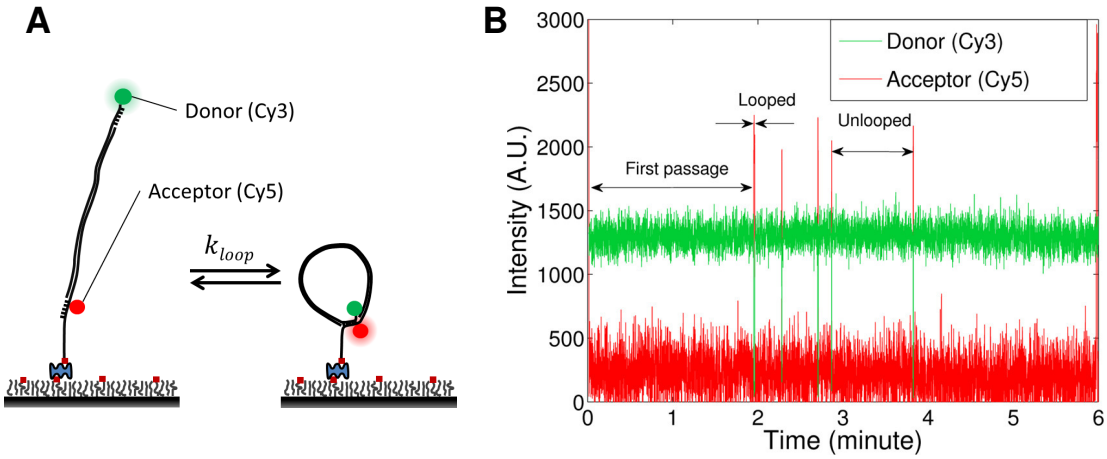


Figure 3.3: (A) Single-molecule FRET experiment. dsDNA molecules labeled with Cy3 (green) on one side and Cy5 (red) on the other are immobilized on the PEG (polyethylene glycol)-coated surface. Reversible looping and unlooping of dsDNAs result in high FRET (looped) and low FRET (unlooped) states. (B) A typical single-molecule time trace of donor (green) and acceptor (red) fluorescence intensities. The high and low FRET states are identified as the looped and unlooped states. Also shown are the first looping time and dwell times in the looped and unlooped states. See main text for their definitions. At the beginning and at the end of the acquisition, 532 and 640 nm lasers were switched on briefly to examine the presence of both Cy3 and Cy5 in each molecule.

To obtain the lifetime of the unlooped state or the mean first looping time, we plotted the number of molecules with both Cy3 and Cy5 that survived in the unlooped state as a function of time. We then fitted this decay curve using a double exponential function with a base line:

$$N(t) = N_\infty + N_f \exp(-k_f t) + N_s \exp(-k_s t). \tag{3.1}$$

The reason for this biphasic decay is not known. In this equation, N_∞ accounts for the number of dysfunctional dsDNA molecules that are not able to loop for unknown reasons. The mean first looping time was calculated as:

$$\langle t \rangle = N_f/(k_f(N_f + N_s)) + N_s/(k_s(N_f + N_s)). \tag{3.2}$$

The only difference between the two was whether the dwells in the unlooped state were

synchronized on their unlooping transitions or not. The looping rate is defined as the inverse of the mean first looping time.

3.3 Measuring DNA looping dynamics

3.3.1 FRET fluctuations reflect looping dynamics of dsDNA

To detect dsDNA looping, we designed DNA molecules with two regions: the looping region and the capturing region. The looping region is a 186-bp long duplex (unless mentioned otherwise), and the capturing regions are 7 or 10 base long single-stranded overhangs at both ends which can anneal to each other upon looping. One of the capturing regions also contains a biotin for surface immobilization (Figure 3.3A). The length of the capturing regions was chosen so that the looped state was stable enough to be detected by FRET, and unstable enough for the DNA to undergo reversible looping and unlooping.

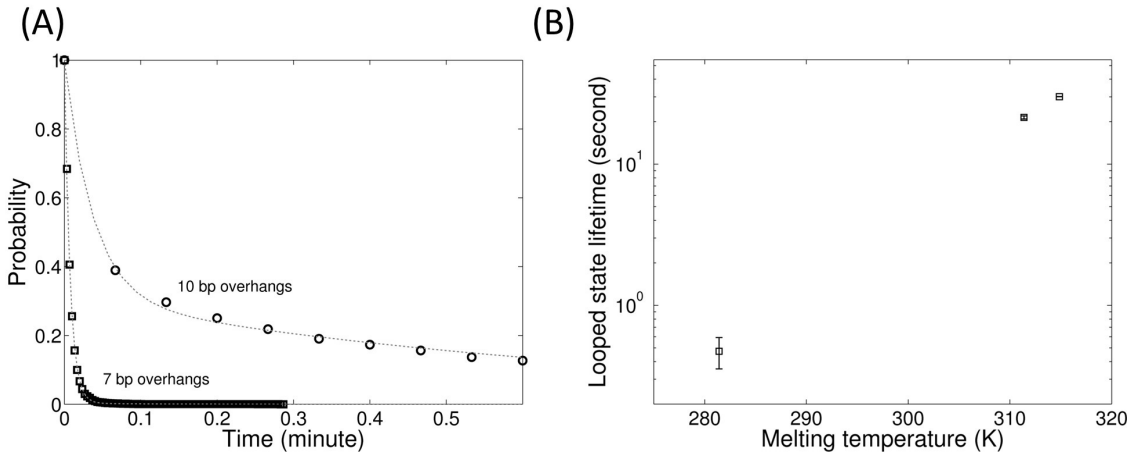


Figure 3.4: (A) The looped state dwell time distribution depends strongly on the strength of the overhangs. With shorter overhangs, not only did we observe a 60-fold difference in the looped state lifetime, but also a change in the dwell time distribution from double exponential to single exponential. (B) The relationship between the melting temperature of the capturing region and the looped state lifetime. The melting temperature of the capturing region was estimated using modified Breslauers thermodynamics at salt concentration of 500 mM (`http://www.finnzymes.fi/tm_determination.html`). The annealing overhangs are (from left to right): TAAATTC (7 bp, no internal dye), GGTAAATTCA (10 bp, no internal dye), and GGTAAATTCAC (10 bp, A is the non-complementary nucleotide for the internal Cy3). The looped state lifetime increases exponentially with the melting temperature, suggesting that the internal dye does not significantly affect base pairing between the overhangs.

We observed anti-correlated fluctuations of Cy3 (donor) and Cy5 (acceptor) from surface-immobilized dsDNA molecules (Figure 3.3B) due to reversible looping and unlooping of the DNA. We rule out blinking of the acceptor or switching of the acceptor dipole orientation as the cause of the fluctuations because:

(1) Cy5 was fluorescently active in the low FRET state as evidenced from the unintentional weak excitation of Cy5 by the 532 nm laser,

(2) the observed intensity changes of the dyes in the capturing regions depended on the DNA sequence of the looping region, and

(3) the mean dwell time (lifetime) of the looped state correlated with the number of complementary bases in the capturing region (Figure 3.4).

The lifetimes of the looped and unlooped states also changed with salt concentration (Figure 3.5). The looped state lifetime increased with salt, presumably because salt stabilizes base pairing. The unlooped state lifetime decreased with increasing salt concentration, consistent with salt promoting annealing between the capturing overhangs. Salt can also accelerate looping by increasing dsDNA flexibility, but this effect is likely small because the negatively charged phosphate groups are almost completely screened in the range of concentrations tested (145-505mM NaCl)[77]. We used 505 mM NaCl as the standard concentration for all other experiments of this study.

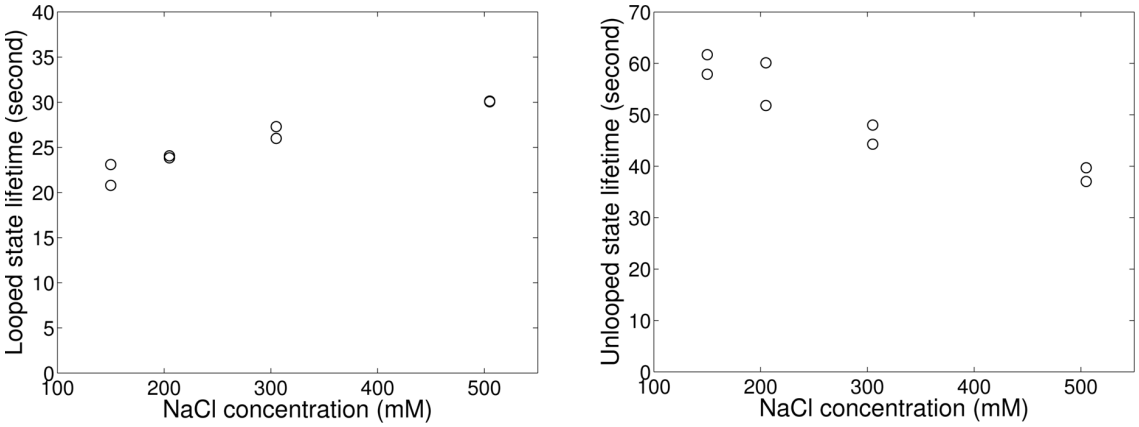


Figure 3.5: Salt dependence of looped and unlooped state lifetimes for sequence S3. Data from two repeated measurements are shown as black circles.

We decreased the length of a dsDNA whose sequence was derived from the yeast genome

from 186 to 102 bp, and observed a 20-fold increase in the unlooped state lifetime (Figure 3.6A). We emphasize that even if we had assumed that all molecules could loop eventually by fitting the decay curves with a zero-baseline exponential function, the change in the looping rate between 102bp and 186bp would still be less than 20-fold. To confirm that the measured looping kinetics was not affected by the acquisition time, we increased the acquisition time from 15 minutes to 45 minutes and measured the unlooped state lifetime of the 102bp dsDNA. As evident from Figure 3.7, the decay curves are nearly indistinguishable.

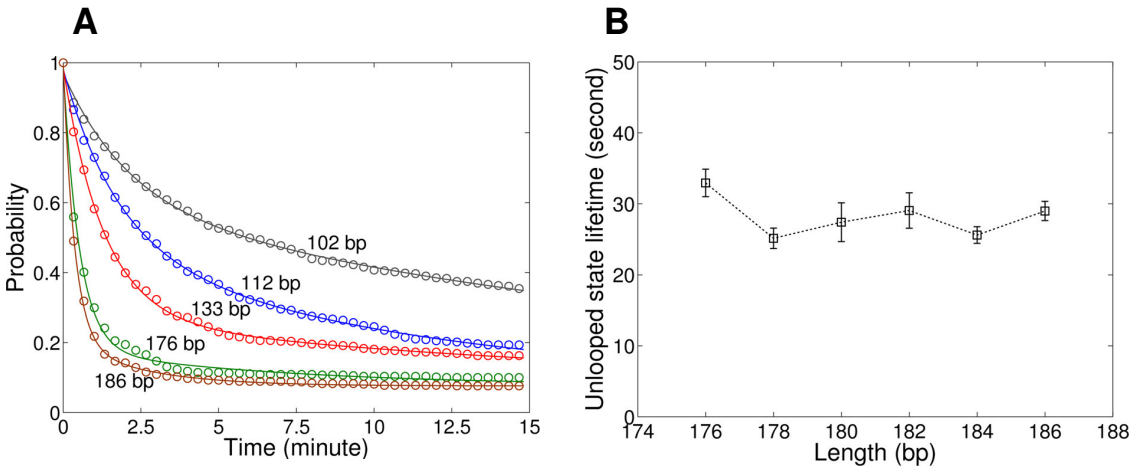


Figure 3.6: (A) Decay curves of the unlooped state population as a function of length (102 bp: gray, 112 bp: blue, 133 bp: red, 176 bp: green, 186 bp: brown). The decay curves were fitted using a double exponential function with a fixed baseline extracted from the 186-bp data. (B) The unlooped state lifetime of the genomic sequence DNA measured over one helical period. The small variation in the lifetime implies weak torsional stress for the dsDNA in the looped state.

In a ligation-based DNA cyclization assay that covalently links the 5'-end and the 3'-end of a dsDNA, j_M depends on the probability that the helical phases of the two ends match each other. This dependence results in oscillation of j_M as a function of length, whose peak-to-peak change can be as much as 100-fold at 180 bp according to the twisted WLC model (reference [23], also see Figure 1.3). To investigate whether our measured looping rate showed similar helical phase dependence, we measured looping rates of dsDNAs varying in length from 176 bp to 186 bp. We found that the looping rate changed no more than 1.3 times over this one helical period (Figure 3.6B). This result suggests that annealing between the two overhangs might not require as strict a helical phase match as ligation does in a

cyclization assay.

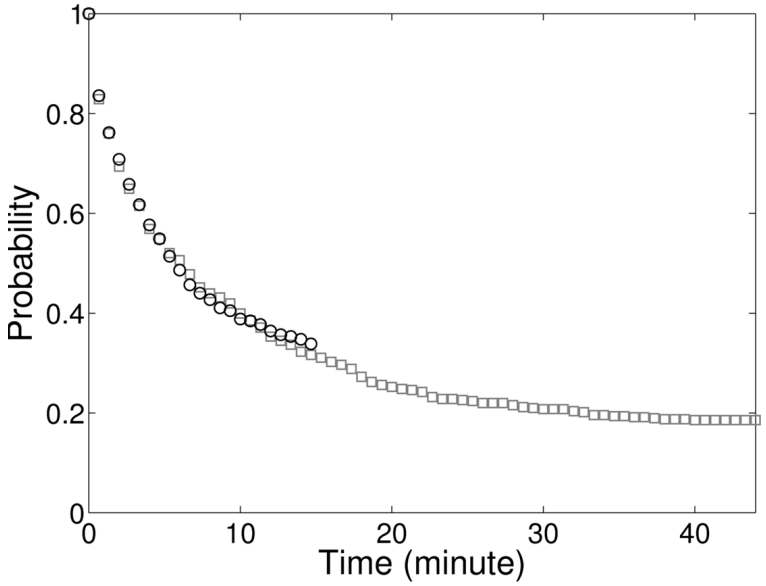


Figure 3.7: Dependence of the apparent looping kinetics on the acquisition time. Decay curves of the unlooped state population are compared between 15-minute (circles) and 45-minute (squares) acquisition. The dsDNA is a 102-bp long genomic sequence DNA. The lifetimes of the unlooped state are 842 seconds for 45 minute acquisition and 817 seconds for 15 minute acquisition.

3.3.2 Conversion of j_M from looping rate

The relationship between the equilibrium probability distribution of chain conformations and looping kinetics was given by Szabo and Schulten[78]. According to this relationship, the mean first passage time between the two reactive terminal sites of a polymer chain is given by:

$$\tau = \int_a^L dx \left[D(x)P_{eq}(x)\right]^{-1} \left(\int_x^L dy P_{eq}(y)\right)^2 + \frac{1}{\kappa_a P_{eq}(a)}, \tag{3.3}$$

where $P_{eq}(r)$ is the equilibrium probability density of the end-to-end distance being equal to r, L is the contour length of the polymer, D is the diffusion coefficient, and a is the inner boundary radius. κ_a reflects the trapping reaction between the two ends when $r = a$. In the case of a perfectly absorbing boundary condition ($\kappa_a \to \infty$), the second term vanishes.

We estimated the upper limit of this mean first passage time using an analytical expression of $P_{eq}(r)$ derived for a worm-like chain[79]. The diffusion coefficient is estimated to be $\sim 2.0 \times 10^8$ nm^2/s by treating the overhangs as 10-base long rigid rods in water[80]. The

boundary radius is assumed to be $\sim$ 5 nm. Using these values, the first term of Equation 3.3 is $\sim$ 5.5 ms, which is $10^2 - 10^3$-fold shorter than our measured looping time. Other studies using molecular dynamics simulations or an analytical calculation also yielded a much shorter first passage time with efficient trapping[81, 82, 83]. Therefore, the measured looping time likely reflects the second term in Equation 3.3, in which case the inverse of the mean first passage time is directly proportional to $P_{eq}(r)$. Moreover, if the measured looping time reflects the first term in the equation, we expect it to increase linearly with viscosity. However, the looping time varied only 1.2-fold between 4% and 30% glycerol, which indicates that the apparent looping is not diffusion limited (data not shown).

j_M is thus related to the mean first looping time (τ_{loop}) by:

$$j_M(a) = \frac{P_{eq}(a)}{4\pi a^2} = \frac{1}{\kappa_a(4\pi a^2)\tau_{loop}}. \tag{3.4}$$

To extract j_M, κ_a must be known. This parameter can be measured from free diffusion of one sticky end with respect to the other in a volume V, where $P_{eq}(a) = 4\pi a^2/V$. Substituting this in Equation 3.3 and assuming that annealing is not diffusion limited, we obtain the annealing time (τ_{anneal})

$$\tau_{anneal} \approx \frac{1}{N\kappa_a\,(4\pi a^2/V)}. \tag{3.5}$$

Here we used the fact that in the presence of N sticky ends, the annealing rate increases N-fold. Substituting κ_a from Equation 3.5 into Equation 3.4, we obtain

$$j_M(a) = \frac{N}{V}\frac{\tau_{anneal}}{\tau_{loop}} = c\frac{\tau_{anneal}}{\tau_{loop}}. \tag{3.6}$$

Therefore, j_M can be determined from three experimental measurables: concentration of sticky ends (c), the looping time τ_{loop}, and the annealing time τ_{anneal}. We realized this experiment by immobilizing the Cy5 containing single strand to the surface and introducing the Cy3-containing single strand at 50 nM (Figure 3.8A). Annealing of the Cy5 strand to the Cy3 strand resulted in detectable FRET bursts, and the first order rate constant was measured to be $0.45\pm0.04\times10^6$ $M^{-1}s^{-1}$ in 505 mM NaCl from several annealing experiments with various concentrations of the Cy3-containing single strand (Figure 3.8). This value is comparable to the values reported by the Ha group ($0.78 \pm 0.07 \times 10^6$ $M^{-1}s^{-1}$ in 1 M

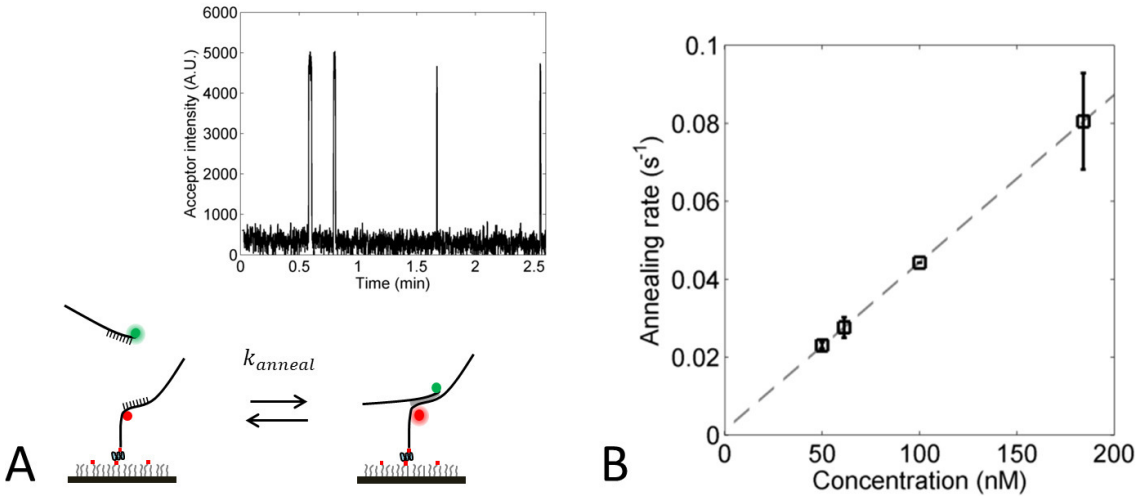


Figure 3.8: (A) Schematic view of the annealing rate measurement. Cy5 primer was immobilized on the surface, and Cy3 primer was introduced. Binding and unbinding of the Cy3 primer occurs reversibly and leads to detectable acceptor bursts, which are shown in the inset. (B) The annealing rate depends linearly on the concentration of the mobile primer. The slope of the line represents the first-order annealing rate constant.

NaCl and $0.26 \pm 0.04 \times 10^6$ $M^{-1}s^{-1}$ in 10 mM [Mg^{2+}]) that were obtained from dimerization kinetics between dsDNAs[42].

3.3.3 More curved molecules loop faster

Although measuring the length dependence of j_M has been used as the litmus test for the worm-like behavior, the interpretation of the result is nontrivial for a couple of reasons. First, changing the length not only changes the bending energy required for looping, but also the twisting energy. Second, curvature can be introduced locally when base pairs are added. Therefore, we kept the length of dsDNAs the same while varying the curvature in a global, predictable manner. To build curved DNA molecules, we concatenated a 10-mer sequence multiple times, with a random nucleotide inserted every 20 bases. As a result, the 10-mer sequence repeats itself every helical repeat (10.5 bp), and the static curvature or bendedness of the 10-mer, if any, will add up constructively. We designed four such repeating sequences, termed sequence S1 through S4. Sequence S1, S2, and S3 are more similar to each other than sequence S4. Sequence S2 and S4 were chosen from a previous

Table 3.1: The apparent size of several 186bp DNA sequences determined from a 100bp DNA marker.

Seq. name	S1	S2	S3	S4	S2-1	S3-1	S3-2
Apparent size	1.211	1.1491	0.9393	0.9623	1.086	0.9546	0.9469

study by Kaplan et al[74], and are supposed to possess opposite nucleosome affinity. We also designed three extra sequences with locally perturbed curvature: (1) we changed G to T in the two central 10-mers of sequence 3 (sequence S3-1), (2) we changed G to T in the terminal 10-mers of sequence S3 (sequence S3-2), and (3) we truncated the central 10-mer of sequence S2 in half to make an S-shaped molecule with an inflection point (sequence S2-1).

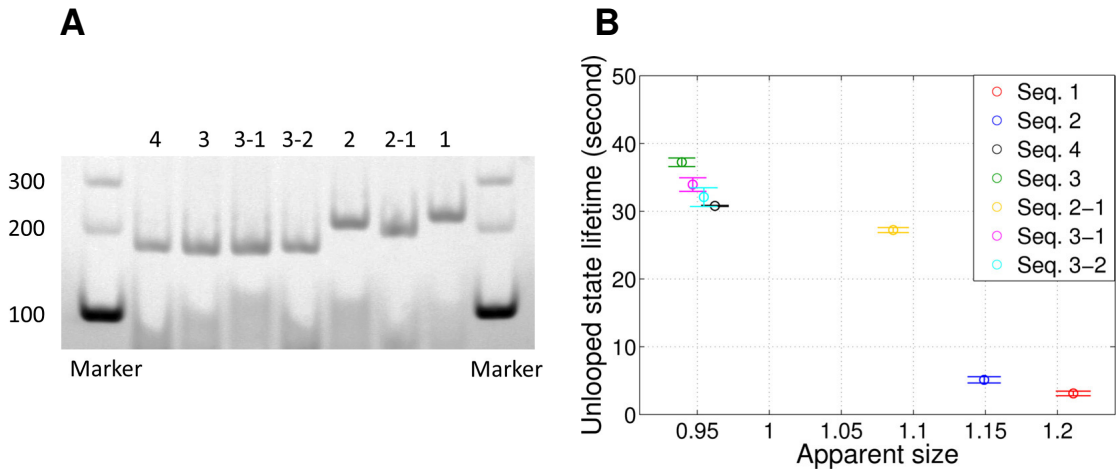


Figure 3.9: (A) Polyacrylamide gel electrophoresis pattern of synthetic DNAs. From left to right, the lanes contain 1 kb marker, sequence S4, sequence S3, sequence S3-1, sequence S3-2, sequence S2, sequence S2-1, sequence S1, and 1 kb marker. The exact sequences of these dsDNAs are given in the Section 3.2. (B) The relationship between the unlooped state lifetime and the apparent size.

We checked overall curvature of these dsDNAs by PAGE electrophoresis. The order of apparent size (R) of sequence S1 through S4 was $S1 > S2 > S4 > S3$ (Figure 3.9A). Among the sequence 3 variants, the apparent size was S3-2 > S3-1 > S3. The S-shaped molecule exhibited an apparent size between S2 and S4. Values for the apparent size of these seven sequences are listed in Table 3.1. Empirically, a curved dsDNA migrates more slowly than a straight dsDNA[75, 84]. Therefore, we could conclude that sequence S1 is most curved,

and sequence S3 is straightest. We measured looping kinetics of these dsDNAs, and found an anti-correlation between the apparent size and the unlooped state lifetime (Figure 3.9B). The strong anti-correlation among the repeat sequences suggests that the deflection angle increases monotonically in the order of S3 $\rightarrow$ S4 $\rightarrow$ S2 $\rightarrow$ S1. Despite locally perturbed curvature, other sequences also followed the anti-correlation between apparent size and looping time, although the S-shaped molecule (sequence S2-1) deviated notably from the overall trend.

3.4 Coarse-grained simulations

In this section, we attempt to use computer simulation to quantitatively describe of the shape-dependent looping kinetics of dsDNA. First, I want to introduce the DNA model and the simulation algorithm to simulate DNA fluctuations in thermal equilibrium.

3.4.1 The wedge-angle model for dsDNA deformation

We start with a discrete model for DNA deformation at single base-pair level, which is also known as the wedge-angle model[85]. The mutual orientation between two base pairs can be completely described by three deflection angles which are roll (ρ), tilt (τ), and twist (Ω). To define these angles, we follow the Cambridge convention[86] which is illustrated in Figure 3.10. On one base pair plane, we impose three mutually perpendicular unit vectors that define a local basis ($\mathbf{d}_1^{(i)}$, $\mathbf{d}_2^{(i)}$, $\mathbf{d}_3^{(i)}$). $\mathbf{d}_3^{(i)}$ is perpendicular to the base pair plane and points toward the center of the next base-pair, $\mathbf{d}_1^{(i)}$ points to the center of the major groove, and $\mathbf{d}_2^{(i)}$ is determined by the mutual perpendicularity of the three unit vectors ($\mathbf{d}_2^{(i)} = \mathbf{d}_3^{(i)} \times \mathbf{d}_1^{(i)}$). The next base pair can be totally identified by three rotations about the three local axes with the angles of roll, tilt, and twist, relative to its predecessor. In particular, twisting represents a rotation about $\mathbf{d}_3^{(i)}$, rolling represents a rotation about $\mathbf{d}_2^{(i)}$, and tilting represents a rotation about $\mathbf{d}_1^{(i)}$. Since there are four types of nucleotide (A, T, G, and C), there are $4^2 = 16$ dinucleotide steps in total. However, they are not totally independent since several steps share similar orientational parameters (for example, AG and CT, AC and GT, TC and GA, TG and CA, AA and TT, GG and CC). Therefore, there are 10 independent types of dinucleotide steps which are chosen to be AA, AC, AG, AT,

CA, CC, CG, GA, GC, and TA. Note that the roll and twist angles are the same for the complementary pairs such as AG and CT, AC and GT, etc., while the tilt angle reverses sign for such pairs[87].

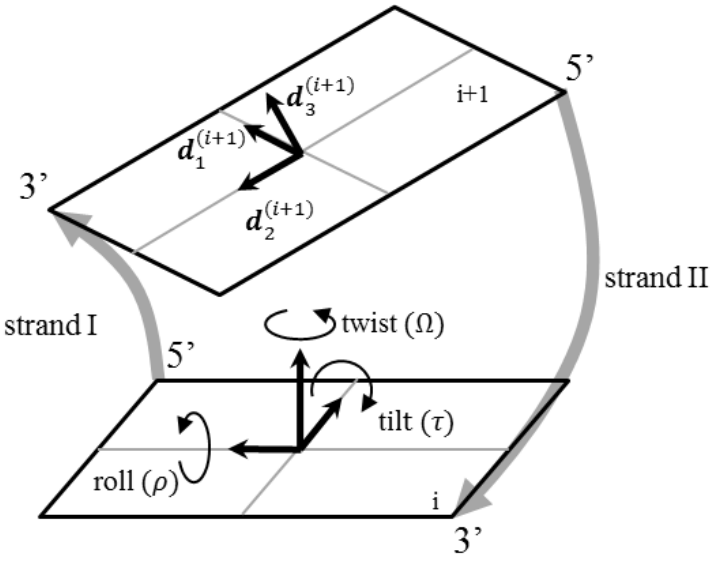


Figure 3.10: Orientational parameters of a base pair plane in a dinucleotide step (redrawn from Figure 1 in [84]). The superscript represents the base-pair index.

B-DNA has a property of having its base pairs very nearly perpendicular to the helix axis which is manifested by small deviations of roll and tilt from zero. However, the accumulation of such small local deformations at individual steps between base pairs could produce large-scale measurable curvature of the DNA. In the literature, there are a number of dinucleotide models aiming to capture the intrinsic curvature of DNA at the macroscopic scales although they often have widely different values for rolls, tilts, and twists. For example, Table 3.2 lists the angle values of some dinucleotide steps (in degree) from the Bolshoy-Trifonov model[88], which shows various deviations from their canonical values for B-DNA structure ($\rho_0 = 0$, $\tau_0 = 0$, $\Omega_0 = 34.3°$).

Given a set of helical parameters, one can readily reconstruct the helical structure using matrix manipulation method. In the wedge-angle model, the base pairs are treated as rigid units, which can be completely described by an orthonormal frame with an origin located at the center of the base pair and a set of three mutually perpendicular unit vectors. For

Table 3.2: The helical parameters for some dinucleotide steps from the Bolshoy-Trifonov model[88].

	Roll ρ_0 (°)	Tilt τ_0 (°)	Twist Ω_0 (°)
AA	-7.6141	2.6891	35.6200
AC	-1.3780	3.8267	34.4000
AG	7.2242	-3.0377	27.7000
AT	3.0655	5.1829	31.5000

base pair i, we set the center of the base pair at $\mathbf{r}^{(i)}$ and the three local axes to be ($\mathbf{d}_1^{(i)}$, $\mathbf{d}_2^{(i)}$, $\mathbf{d}_3^{(i)}$) (Figure 3.11A). To determine frame $i+1$, we perform a three-dimensional rotation whose representation in the basis of frame i is given by an orthogonal 3-by-3 matrix $\mathbf{R}^{(i)}$. For example, if the twist angle is $+\Omega$, we have $\mathbf{R}^{(i)} = \begin{pmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{pmatrix}$. New basis ($\mathbf{d}_j^{(i+1)}$, $j = 1, 2, 3$), which describes frame $i+1$, is obtained by rotating $\mathbf{d}_j^{(i)}$ by an angle Ω about $\mathbf{d}_3^{(i)}$, or in other words, $\mathbf{d}_j^{(i+1)} = \mathbf{R}^{(i)}\mathbf{d}_j^{(i)}$.

When considering all deformation angles such as roll, tilt, and twist, the rotation matrix $\mathbf{R}^{(i)}$ involves successive single rotations about $\mathbf{d}_2^{(i)}$ by an amount ρ, $\mathbf{d}_1^{(i)}$ by an amount τ, and $\mathbf{d}_3^{(i)}$ by an amount Ω. Therefore, the overall rotation matrix $\mathbf{R}^{(i)}$ is

$$\begin{aligned}\mathbf{R}^{(i)}(\rho,\tau,\Omega) &= \mathbf{R}_{\mathbf{d}_2^{(i)}}(\rho)\mathbf{R}_{\mathbf{d}_1^{(i)}}(\tau)\mathbf{R}_{\mathbf{d}_3^{(i)}}(\Omega) \\ &= \begin{pmatrix} \cos\rho\cos\Omega - \sin\tau\sin\rho\sin\Omega & -\cos\tau\sin\Omega & \sin\rho\cos\Omega + \sin\tau\cos\rho\sin\Omega \\ \cos\rho\sin\Omega + \sin\tau\sin\rho\cos\Omega & \cos\tau\cos\Omega & \sin\rho\sin\Omega - \sin\tau\cos\rho\cos\Omega \\ -\cos\tau\sin\rho & \sin\tau & \cos\tau\cos\rho \end{pmatrix}\end{aligned} \tag{3.7}$$

To remove degrees of freedom of the global rotation and the global translation that are irrelevant to DNA shape, the first base pair is fixed with the center at the origin of the laboratory frame ($\mathbf{r}^{(1)} = 0$) and the three unit vectors ($\mathbf{d}_1^{(1)}$, $\mathbf{d}_2^{(1)}$, $\mathbf{d}_3^{(1)}$) are set to align along the external coordinates ($\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$). Following that choice, the location and orientation of all other base pairs can be readily determined by performing matrix multiplication (refer to Figure 3.11A for more clarity)

$$\mathbf{d}_j^{(i)} = \mathbf{R}^{(1)}\mathbf{R}^{(2)}\ldots\mathbf{R}^{(i-1)}\mathbf{e}_j, \quad j = 1, 2, 3, \quad i = 2, 3, \cdots, N, \tag{3.8}$$

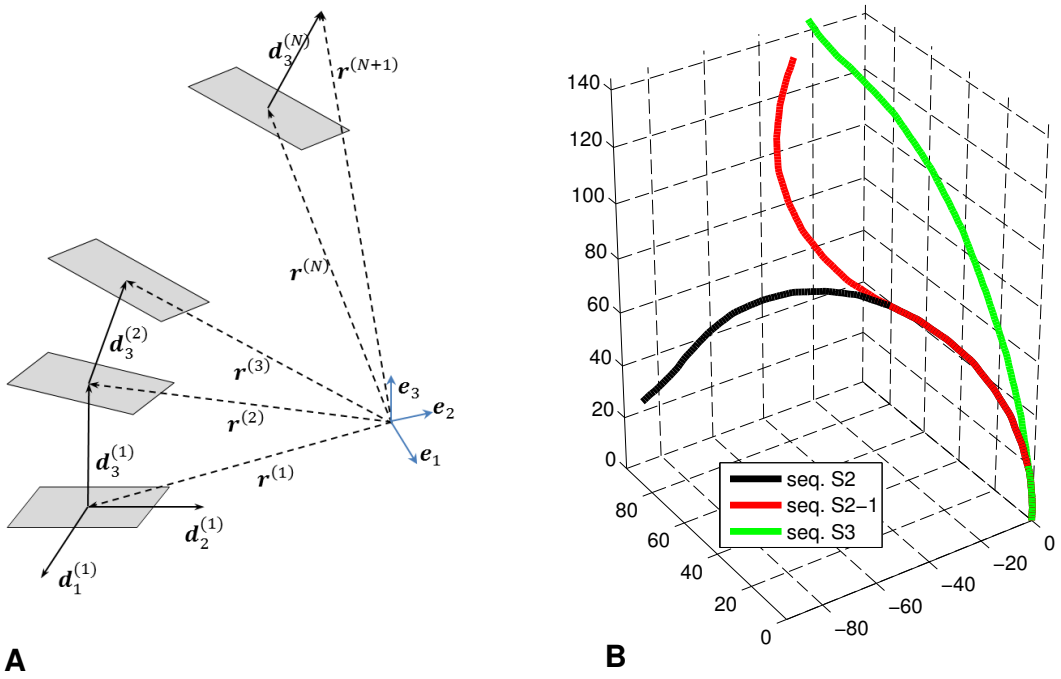


Figure 3.11: (A) A schematic that defines base-pair coordinates via a set of vectors $\mathbf{r}^{(i)}$ which identify the center of each basepair's plane and the local basis ($\mathbf{d}_j^{(i)}$, $j = 1, 2, 3$) which determine their spatial orientation. (B) Example of the helix trajectories of some DNA sequences in this study, calculated using the parameters from the Anselmi-Scipioni model[84]. The unit of length in three axes is base pair.

where N is the number of base pair. The location of the center of base pair i is given by:

$$\mathbf{r}^{(i)} = \mathbf{e}_3 + \mathbf{d}_3^{(2)} + \cdots + \mathbf{d}_3^{(i-1)}, \quad i = 3, \cdots, N \tag{3.9}$$

Note that the final virtual vector $\mathbf{r}^{(N+1)}$ is useful when describing a closed DNA conformation (DNA minicircles) since one just needs to set a distance constraint on the two ends $\mathbf{r}^{(N+1)} = \mathbf{r}^{(1)} = 0$.

As an example, Figure 3.11B represents some helical trajectories of several 186-bp DNA sequences which include sequence S2, sequence S3, and sequence S2-1, calculated using the parameters taken from the Anselmi-Scipioni model[84].

3.4.2 Adding the thermal fluctuations

In Section 3.4.1, we introduced the wedge-angle DNA model which describes the relative oritentation of two base pairs by three angles which are roll, tilt, and twist. There are two

contributions for each deformation angle: one from the intrinsic bendedness (represented by ρ_0, τ_0, Ω_0) and the other from thermal fluctuations (represented by the difference between the actual angle and the ground-state angle, for example, $\rho - \rho_0$). The parameters (ρ_0, τ_0, Ω_0) describe a minimum energy configuration or a 'frozen' state of the dsDNA molecule. Thermal fluctuations, however, cause deviations of these parameters from their ground-state values. To describe the thermal fluctuations of the DNA, we use an elastic potential for the bending energy of each base-pair step. The potential depends quadratically on the deviation of the bending angle from its ground-state value:

$$E(\rho, \tau, \Omega) = \frac{1}{2}k_\rho(\rho - \rho_0)^2 + \frac{1}{2}k_\tau(\tau - \tau_0)^2 + \frac{1}{2}k_\Omega(\Omega - \Omega_0)^2. \tag{3.10}$$

Coupling between different types of step deformations (for example, roll-tilt coupling) is not considered here.

The probability that a dinucleotide step adopts a configuration (ρ, τ, Ω) is proportional to the Boltzman factor of the step deformation energy

$$p(\rho, \tau, \Omega) \propto \exp\left(-\frac{E(\rho, \tau, \Omega)}{k_B T}\right) = \exp\left(-\frac{(\rho - \rho_0)^2}{2\sigma_\rho^2}\right) \exp\left(-\frac{(\tau - \tau_0)^2}{2\sigma_\tau^2}\right) \exp\left(-\frac{(\Omega - \Omega_0)^2}{2\sigma_\Omega^2}\right) \tag{3.11}$$

where $k_B T$ is the thermal energy which is about 0.6 kcal/mol at room temperature. In this equation, $\sigma_\rho = \sqrt{\frac{k_B T}{k_\rho}}$, $\sigma_\tau = \sqrt{\frac{k_B T}{k_\tau}}$, and $\sigma_\Omega = \sqrt{\frac{k_B T}{k_\Omega}}$ are the variances of the Gaussian distributions for the three helical angles.

Interestingly, the bendability of dsDNA in different directions is not uniform. It was estimated that $(\sigma_\rho, \sigma_\tau, \sigma_\Omega) \approx (6^\circ, 3^\circ, 4^\circ)$ for a generic dsDNA molecule[89]. This means that dsDNA deforms more easily in the direction of roll than in the direction of tilt by almost two times. These quantities were also found to be sequence-dependent. For example, data from the crystallized structures of protein-DNA complexes[90] discovered that the TA step is the most flexible one since it possesses large roll fluctuation compared to other steps (see Table 3.3).

3.4.3 Numerical prediction of the J factor

We randomly sampled numbers from a Gaussian distribution, multiplied them with the roll, tilt, and twist variances, and added them to the ground-state roll, tilt, and twist angles to

Table 3.3: The variances for the angle fluctuations of some dinucleotide steps from the Olson-Zhurkin model[90].

	Roll σ_ρ (°)	Tilt σ_τ (°)	Twist σ_Ω (°)
AA	6.1360	3.7552	3.3208
AC	4.4340	3.5250	3.5775
AG	3.8691	3.7552	3.8362
AT	5.5782	2.8440	3.8362
TA	7.4963	3.0718	4.6902

obtain roll, tilt, and twist angles for each dinucleotide in thermal equilibrium. Using these angles, we generated coordinates of dsDNA configurations[91]. The histogram of end-to-end distances (r) was normalized and divided by $4\pi r^2$ to obtain $J(r)$, the probability density of the end-to-end distance being equal to r. j_M was obtained by estimating $J(r)$ at 5.5 nm and converting it to molar concentration. Typically, 10^5 configurations were enough to yield reliable extrapolation of j_M for >150-bp dsDNA, and 10^6 configurations for <150-bp dsDNA. The number of chains whose end-to-end distance is within the annealing range (5.5

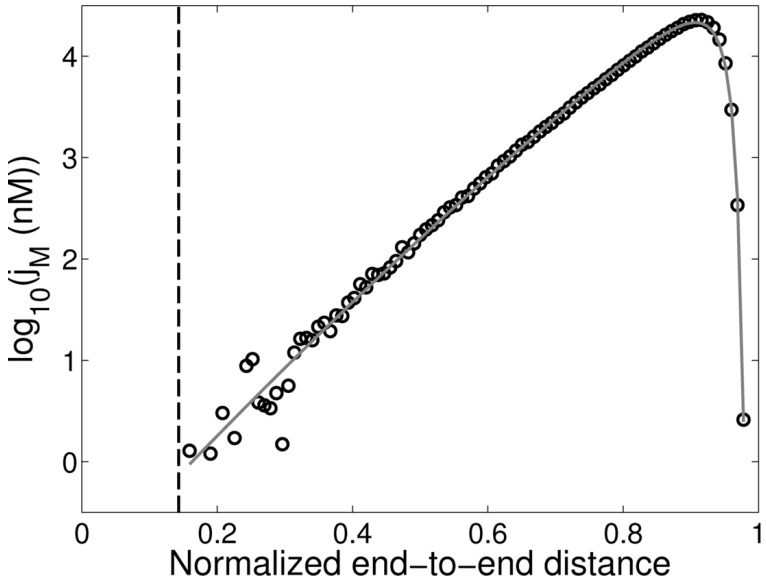


Figure 3.12: The logarithm of j_M as a function of the end-to-end distance. Black circles represent the mean probability density in each bin expressed in molar units. The x-axis represents the end-to-end distance of a dsDNA normalized by its contour length. The default bin width is equal to the helical rise of DNA base pair. For small end-to-end distances, the bin width is adjusted so that the integrated probability over a bin is not zero. The gray curve is the fit to the black circles as explained in the main text. The vertical dashed line indicates the end-to-end distance of 5.5 nm where j_M is obtained.

nm) is very small, especially for straighter dsDNAs. For example, sequence S4 typically yields none or a few loops out of 2 million trials. Therefore, estimation of j_M at a small distance is error-prone. To remedy this problem, we extrapolated the dependence of the logarithm of j_M on the end-to-end distance to obtain j_M at small distances[92]. We used a rational function to fit this curve: $y = ax + b + c/(x-d)^2$, where b corresponds to j_M near zero end-to-end distance. An example of this extrapolation is shown in Figure 3.12.

3.4.4 Optimizing helical parameters from gel mobility

A curved dsDNA costs less bending energy than a straight one, and thus has a higher j_M. To compare j_M between different dsDNAs, one must know their intrinsic shape. Koo and Crothers obtained a relationship between the deflection angle of a dsDNA and its apparent size on a gel[75]. However, the absolute deflection angle of a reference dsDNA was determined from its j_M according to the WLC model. Since we wanted to test the WLC model, we used the relationship between the apparent size and the small-angle fluctuation of the DNA helical axis known as the central dispersion (σ^2)[84]. According to Scipioni and coworkers, the apparent size of a dsDNA is related to the Boltzmann probability that its curvature is straightened out in thermal equilibrium, and is thus proportional to its central dispersion. The central dispersion has two contributions, $\sigma^2 = \sigma_s^2 + \sigma_d^2$ where σ_s^2 is the static contribution from the intrinsic bendedness of the sequence and σ_d^2 is the dynamic contribution from the thermal fluctuations of the molecule. Since the static term is overwhelming compared to the dynamic term for bent DNAs, the second contribution is ignored in this treatment[84]. Here, we followed the calculation by De Santis and Scipioni [93] to estimate the values for the central dispersion of several DNA sequences and correlated them with their measured electrophoretic mobilities to deduce the compatible helical parameters for the wedge-angle model.

The average curvature vector $\mathbf{C}(n)$, which measures the orientational deviation of the curved helical axis at position n compared to the first nucleotide's local helical axes, is calculated as:

$$\mathbf{C}(n) = \sum_{s=n_1}^{n_2} \mathbf{d}_s \exp\left(\frac{2\pi\imath s}{\nu}\right), \tag{3.12}$$

where $n = (n_1 + n_2)/2$, $\imath$ is the imaginary unit, ν is the helical periodicity which is set to be 10.4, and $\mathbf{d}_s = \rho_s - \imath\tau_s$ is the local deviation of the sth base pair plane from the canonical B-DNA which is represented as a complex vector in terms of the roll (ρ) and tilt (τ) angles.

The central angular dispersion σ^2 can be calculated from $\mathbf{C}(n)$, which includes several averaging of the central dispersion and its complex conjugate $\mathbf{C}(n)^*$:

$$\sigma^2 = \langle \mathbf{C}(n)\mathbf{C}(n)^* \rangle - \langle \mathbf{C}(n) \rangle \langle \mathbf{C}(n) \rangle^*, \tag{3.13}$$

or more explicitly,

$$\sigma^2 = \frac{1}{N-1} \sum_{n=1}^{N-1} \left[\left(\sum_{s=1}^{n} \mathbf{d}_s e^{\frac{2\pi\imath s}{\nu}} \right) \left(\sum_{s=1}^{n} \mathbf{d}_s e^{\frac{2\pi\imath s}{\nu}} \right)^* \right] - \left[\left(\frac{1}{N-1} \sum_{n=1}^{N-1} \sum_{s=1}^{n} \mathbf{d}_s e^{\frac{2\pi\imath s}{\nu}} \right) \left(\frac{1}{N-1} \sum_{n=1}^{N-1} \sum_{s=1}^{n} \mathbf{d}_s e^{\frac{2\pi\imath s}{\nu}} \right)^* \right]. \tag{3.14}$$

The apparent size is proportional to the central dispersion and can be expressed as: $R = a\sigma^2 + b$. To deduce the most compatible set of helical parameters that could predicts the gel mobility of the tested sequences, we attempted to minimize the total difference between the predicted apparent size and the measured apparent size of seven synthetic DNA sequences $\sum_{allsequence} (R_{simul.} - R_{exp.})^2$. We used the Levenberg-Marquardt nonlinear least squares algorithm in MATLAB to find helical parameters that best predict the mobility of the seven artificial dsDNAs. We optimized twenty two variables in total: twenty angular parameters corresponding to roll and tilt angles for all ten unique dinucleotides and two additional parameters (a, b) for the linear relationship between the apparent size (R) and the central dispersion (σ^2). We used 0.000156 and 0.97 as the initial values of a and b as published[93] and initialized the angular parameters based on seven different helical parameter sets, termed BT[88], OZ[90], CD[94], AS[84], LB[95], CHARMM27[96], and PARMBSC0[96]. These helical parameter sets were determined from gel electrophoresis, crystal structures, or MD simulations. The twist angle was relatively insensitive to this optimization and was thus kept fixed[93]. The three dimensional dsDNA structure of a sequence was determined by the thirty angular variables, from which the central dispersion was calculated as published[93].

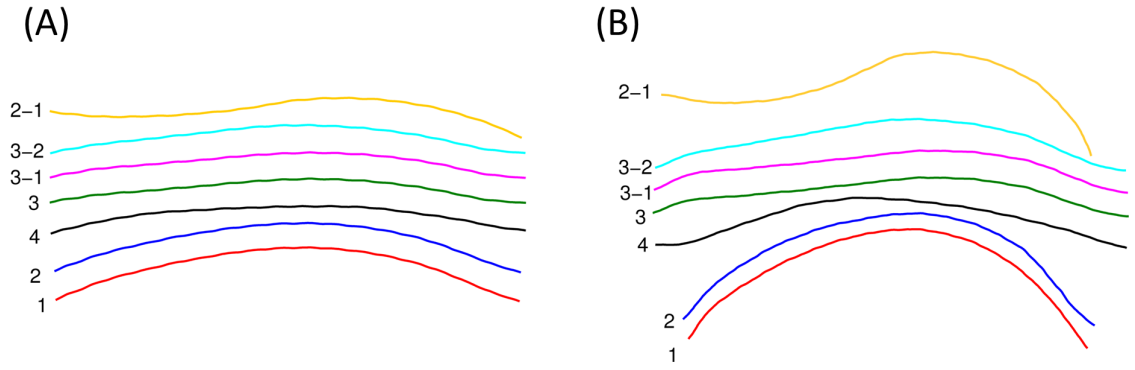


Figure 3.13: (A) Two dimensional projection of the intrinsic shapes of the synthetic dsDNAs. The three dimensional configuration of dinucleotide coordinates was generated using the helical parameter set derived from Olson and Zhurkin (OZ)[90], and we performed principle component analysis to find the projection plane most parallel to this configuration. (B) The intrinsic shapes of the synthetic dsDNAs predicted by Bolshoy and Trifonov (BT)[88] parameter set.

This method enables us to obtain the most compatible adjustments of these seven dinucleotide parameter sets from the literature to match the observed gel mobility pattern without a priori assumption on the shape of any DNA molecule. The shapes of dsDNAs predicted by two of the optimized parameter sets are presented in Figure 3.13.

3.5 Comparison of experiment with WLC simulation

3.5.1 The effect of intrinsic curvature on the looping probability of DNA sequence

To compute j_M, we used the optimized sets of angular parameters with equal bending rigidity parameters in both roll and tilt directions ($\sigma_\rho = \sigma_\tau = 4.8°, \sigma_\Omega = 4.2°$) that correspond to the bending and twisting persistence lengths of 49.7 nm and 65 nm, respectively[47]. Calculations using anisotropic bending parameters ($\sigma_\rho = 6°, \sigma_\tau = 3°, \sigma_\Omega = 4.2°$) did not show significantly different results. Because the bending energy of a dinucleotide is proportional to the square of its angular displacement, the angular displacement is normally distributed in thermal equilibrium. Hence, we randomly sampled angular displacement values from a Gaussian distribution and built $10^5 - 10^6$ chain configurations for each dsDNA. The value of j_M was obtained by extrapolating the logarithm of j_M at small end-to-end distance of 5.5 nm

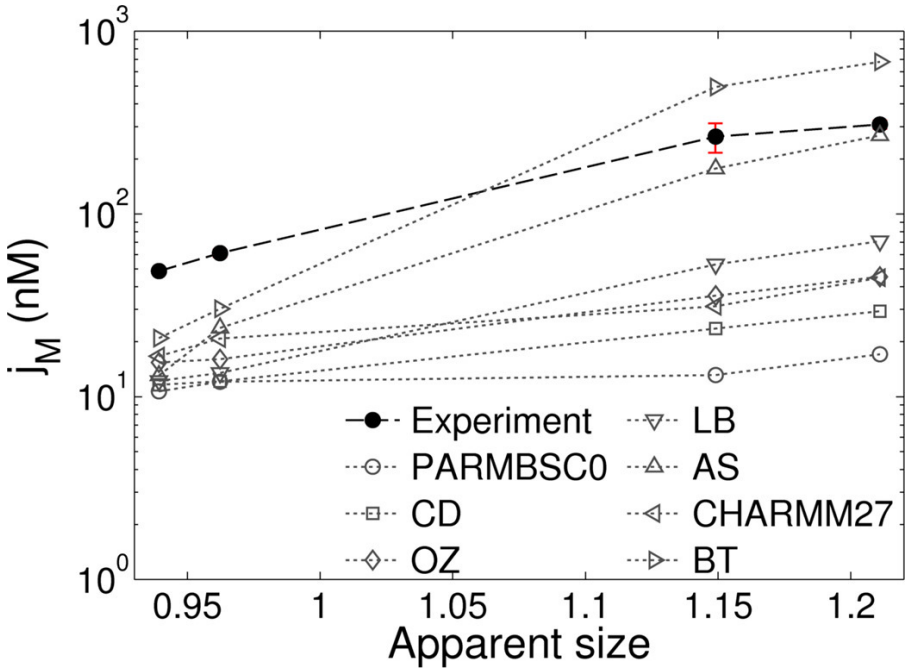


Figure 3.14: Dependence of j_M on curvature. j_M of dsDNAs uniformly curved by design (sequence S1, S2, S3 and S4) were experimentally determined according to Equation 3.6. The mean values of three independent measurements are shown as solid spheres, and the error bars (representing SEM) are shown in red behind them. For comparison, j_M values calculated from simulation using seven different parameter sets are shown on the same plot. Except for the two values obtained using the BT parameter set (right-pointing triangles), all j_M values are predicted to be lower than measured according to the dinucleotide chain model.

As shown in Figure 3.14, in most combinations of DNA sequence and helical parameter set, the predicted j_M was lower than its experimental counterpart. Although the parameter set derived from the study of Bolshoy and Trifonov (BT) yields a higher j_M value than the experimental value for curved dsDNAs, it significantly overestimates the fold change in j_M. We note that this parameter set tends to predict larger deflection angles compared to others (Figure 3.13). All parameters sets were able to predict the increasing order of j_M. The ones that predict larger deflection angles tend to predict a larger change in j_M between the most curved and the straightest dsDNAs than measured. The ones that predict a similar change in j_M, however, underestimate j_M overall. This analysis suggests that it is difficult to reconcile the looping probability with the observed gel mobility within the sequence-dependent worm-like model.

3.5.2 Looping probability of a genomic sequence DNA as a function of length

Using the same parameter sets, we also computed j_M of the genomic sequence dsDNA as a function of length. As shown in Figure 3.15, the BT parameter set produces j_M that closely matches the experimental data. CHARMM27 also works well except at 112 bp. Other parameter sets can produce the overall rate of increase in j_M, but underestimates the absolute value of j_M. We note that BT and CHARMM27 parameter sets can well predict the measured j_M values because they also predict relatively large curvatures for the genomic sequence dsDNAs used (Figure 3.16).

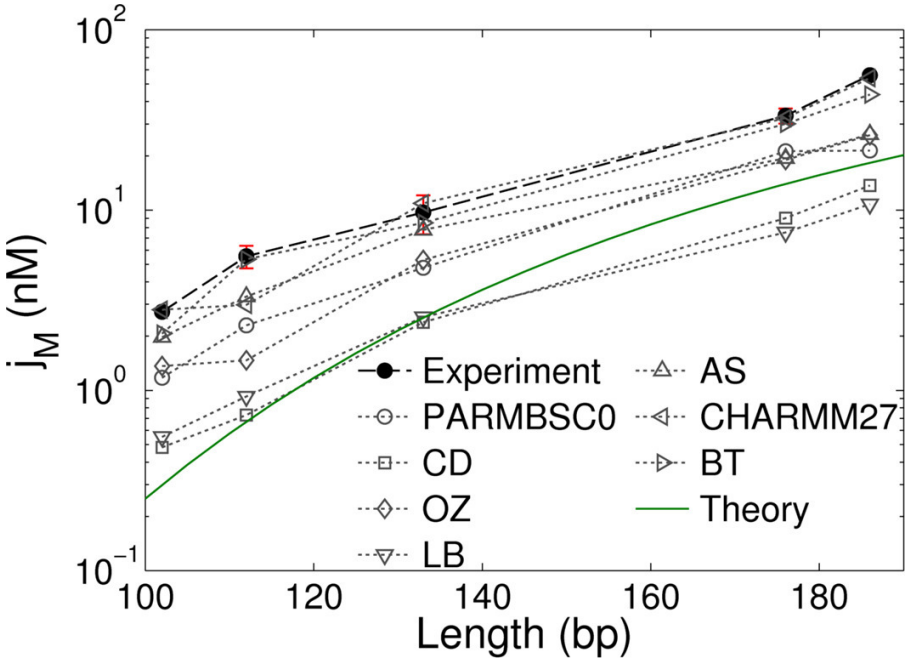


Figure 3.15: Dependence of j_M on length. j_M of the genomic sequence dsDNAs with varying length (Sequence 102, 112, 133, 176, and 186) were experimentally determined according to Equation 3.6. The mean values of three independent measurements are shown as solid spheres, and the error bars (representing SEM) are shown in red behind them. For comparison, j_M values calculated from simulation using seven different parameter sets are shown on the same plot. Most j_M values are predicted to be lower than measured according to the dinucleotide chain model. The green curve represents j_M of a uniform, isotropic worm-like chain calculated from an analytical estimation by Douarche and Cocco[24] with the constraint of 5.5 nm end-to-end distance.

Moreover, even for the CD or LB parameter set whose prediction deviates most from experimental values, the discrepancy is not more than 10-fold between 100 and 200 bp. In comparison, j_M of an intrinsically straight, isotropic chain calculated from an analytical

estimation by Douarche and Cocco[24] is substantially lower than the measured j_M (green curve, Figure 3.15), and the gap between them widens as length decreases. This result shows that the measured length dependence of j_M alone can be explained by the worm like chain model with a proper choice of dinucleotide parameters.

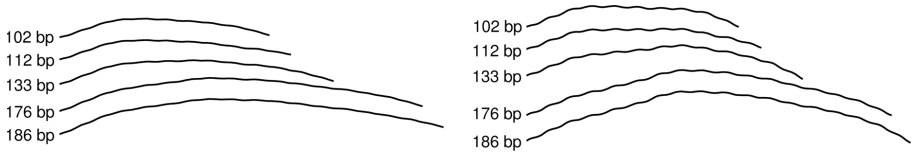


Figure 3.16: The intrinsic shapes of the genomic sequence dsDNAs with varying length predicted by OZ parameter set (left) and CHARMM27 parameter set (right).

3.6 Discussion

In summary, we studied looping kinetics of dsDNA shorter than 200 bp using single-molecule FRET. The measured first looping times were much longer than the estimated first diffusive encounter time between the ends of the dsDNA, which allowed us to extract the equilibrium looping probability density j_M from the looping rate. j_M increased with curvature and decreased with length as expected. We computed j_M using helical parameter sets consistent with the observed gel mobility and found that it was lower than the measured j_M in most cases. We also demonstrate that it is difficult to test the WLC model of dsDNA with j_M vs. length measurement alone as some parameter sets could reproduce the measured j_M between 100 and 200 bp. Hence, we propose the experimental comparison between gel mobility and looping probability of dsDNA at a fixed length as a powerful, alternative approach to test consistency of the WLC model.

In a ligase-based DNA cyclization assay, j_M is defined as the concentration of monomers that result in the same fraction of dimers and circles in equilibrium, and can be estimated from the transient first-order rates of circle and dimer formation[17]. However, this measured j_M is sensitive to the ligase concentration used and can be overestimated as pointed out by Du and Vologodskii[38]. Furthermore, nonspecific binding DNA ligase to DNA can also affect j_M[72]. In contrast, our FRET-based looping assay similar to Vafabakhsh and Ha[42] is free from these concerns.

In this study, we used the mean first looping time (τ_{loop}) for theoretical comparison for two reasons. First, Equation 3.3 formulates looping kinetics in terms of the mean first passage time by averaging the first passage time over the unsynchronized initial end-to-end distribution of a polymer[78]. Second, using the mean first looping time (τ_{loop}) instead of the unlooped state lifetime reduces the risk of overestimating the frequency of rare events. For example, a routine dwell time analysis includes events that are observed during the finite acquisition time only, but excludes slow events that could have been observed had the acquisition time been longer. Also, a long exposure required to capture rare events can elevate the temperature, thus accelerating the apparent looping rate.

The computation of j_M by simulation requires geometric constraints for the looped state. In a ligation-based DNA cyclization assay, j_M exhibits length-dependent oscillation, indicative of torsional stress involved in covalent loop closure. However, we did not observe such oscillation in the looping rate over a 10-bp window (Figure 3.6B), which suggests that annealing between the sticky ends in our assay does not require as precise an alignment between them as in covalent loop closure. In comparison, a larger helical phase dependence of the looping rate was observed by Vafabakhsh and Ha[42], and in another protein-mediated looping study, a single nick was not enough to eliminate this helical phase dependence[97]. We suspect that the torsional stress is alleviated in our assay because even upon annealing, there remains a few-base gap between opposite ends of one strand. Moreover, the dye molecules present in the sticky ends might prevent proper stacking of bases around them. Hence, for j_M calculation, we applied end proximity as the only constraint.

In this study, dsDNA was attached to the surface through a terminal base. This raises the question of how the surface attachment would affect j_M. According to our simulation, the j_M of a dsDNA terminally anchored to surface is lower than that of a dsDNA free in solution[98]. Therefore, the j_M values we report here should be taken as low estimates, which imply that the discrepancy in j_M between a real dsDNA and a worm-like chain might be even larger.

Our results suggest that it is difficult to reconcile the energetics of small-angle dsDNA

bending with that of large-angle or sharp bending and underscore the importance of sequence intrinsic curvature for DNA flexibility at short length scales. Gel mobility can be theoretically interpreted as the probability of a curved dsDNA to be straightened[84]. Thus, the angular parameters derived based on gel mobility of mildly curved dsDNAs are optimized to describe small-angle bending, and we show that they underestimate j_M which describes sharp bending. The most discussed mechanism for subelastic dsDNA is the formation of defects such as kinks and bubbles in the dsDNA structure[32]. However, the existence of stable kinks or bubbles in dsDNAs as long as 190 bp is very unlikely at the high salt concentrations we used[38]. Another mechanism that does not require bubble formation comes from a recent molecular dynamics study[99]. In that study, the authors proposed that counterions preferentially congregate on the concave side of the sharply bent dsDNA, and therefore, the free energy cost of bending dsDNA increases linearly with bending angle at large bending angles. It is also possible that the dinucleotide chain model fails at large bending angles because nearest-neighbor effects become significant[100]. Further studies on the effect of salt, temperature, and sequence on dsDNA looping will help elucidate the mechanism of extreme dsDNA flexibility.

CHAPTER IV

UNLOOPING ASSAY TO TO INVESTIGATE THE BENDING MECHANICS OF DNA UNDER STRONG BENDING CONDITIONS

Sharp bending of dsDNA plays an essential role in genome structure and function. However, the elastic limit of dsDNA bending remains controversial. Here, we measured the opening rates of small dsDNA loops with contour lengths ranging between 40 and 200 bp using single-molecule FRET. The relationship of loop lifetime to loop size revealed a critical transition in bending stress. Above the critical loop size, the loop lifetime changed with loop size in a manner consistent with elastic bending stress, but below it, became less sensitive to loop size, indicative of softened dsDNA. We show that our result is in quantitative agreement with the KWLC model and previously reported looping probabilities of dsDNA between 50 and 200 bp. Our findings shed new light on the energetics of sharply bent dsDNA.

4.1 Introduction

Strong bending of dsDNA, which refers to deflection of larger than $\sim 2.4^\circ$ between adjacent base pairs (or equivalently, one turn per persistence length). Since the WLC is valid only within the elastic limit of dsDNA, the actual bending energy of dsDNA in such processes may deviate from the WLC prediction. From the ligase-dependent cyclization experiment, Widom *et al.* found that the J factors of dsDNAs shorter than 150 bp were measured to be several orders of magnitude higher than the WLC model predictions[28, 29]. This was considered to be the critical size for the validity of the WLC model. Later, Du *et al.* carefully considered a technical point in the bulk assay and re-estimated the critical size to be lower than 100 bp[38]. Other experiments were conducted to explore the critical size of the elastic bending of dsDNA. CryoEM imaging[40] detected no sharp bending points like kinks or melting bubbles in the 94bp DNA minicircle from the Widom's study[28] while a study based on DNA minicircle digestion[39] found that DNA as short as 60-70bp are likely to possess such local irregularity, but not DNA minicircles larger than 90bp.

The controversy emerged again when in a recent study, Vafabakhsh and Ha used single-molecule FRET to measure the J factors of dsDNAs much shorter than 100 bp[42]. In this method, DNA loop formation can be detected without external agents such as protein or a bead that can bias the equilibrium looping probability of dsDNA[72, 73, 71]. They found that J factors in the range between 65 and 110 bp determined from looping kinetics were a few orders of magnitude higher than the WLC model prediction. The results from this study suggest a significant departure of dsDNA from either the WLC model or 45-53-nm persistence length. However, other experimental factors could have led to an overestimation of the J factor: (1) using synthetic oligos may introduce mismatched base pairs[36], (2) high salt conditions (1 M sodium or 10 mM magnesium) can increase DNA curvature and flexibility[101, 102, 103, 104] and/or (3) long sticky ends used in the experiment can increase the apparent looping probability[105, 44]. (1) can be addressed by using PCR-based DNA assembly[36, 45], but (2) and (3) cannot be easily addressed because lowering salt concentration or shortening the sticky ends severely reduce the frequency of looping events observable by FRET for short DNA molecules.

In this chapter, I will describe a different FRET-based approach to test the WLC model at short length scales. The key idea is that stability of end-to-end annealed DNA loops is highly sensitive to loop size due to internal bending stress as depicted in Figure 4.1B. In our FRET assay, the looped state of a dsDNA is stabilized by formation of a transient linker duplex of ∼10 bp between its sticky ends. The lifetime of this linker duplex depends on the shear force exerted along its helical axis by the looped DNA. Since different DNA models make different predictions about how this shear force depends on the loop length, we can experimentally test these models by measuring linker lifetime vs. loop size.

Our unlooping-based approach has unique capabilities that complement the ligation-based or FRET-based J factor measurements: (1) unlooping rates can be measured with good statistics in moderate salt conditions where looping of short dsDNA rarely occurs; (2) only the molecules that were able to loop are followed in the loop breakage assay, which automatically filters out dysfunctional molecules; and (3) the unlooping rate is related to the shear force, which is easier to compute than the J factor.

Using this unlooping assay, we measured the lifetime of small DNA loops as a function of loop size in the strong bending regime. We found that the loop lifetime decreases with decreasing loop size, indicative of increasing bending stress. The bending stress, however, ceased to increase elastically below a critical loop size, reminiscent of a structural transition in dsDNA. The critical loop size increased in the presence of magnesium, indicating the role of divalent ions in softening dsDNA. We show that our data cannot be explained by a continuous polymer model with a single flexibility parameter, but instead calls for a model with an additional internal state such as a KWLC model. Using such a model, we can estimate the free energy of kink formation to be larger than $\sim 18 k_B T$, and $\sim 12 k_B T$ with 5 mM magnesium. We also show how the apparent discrepancy between previous J factor measurements and discrepancy among several experimental approaches such as cyclization and DNA minicircle digestion can be quantitatively resolved by our KWLC model.

4.2 Shear force and loop breakage dynamics

Since it costs energy to bend dsDNA, a looped DNA will exert mechanical force on its two ends to restore its unlooped conformation. We consider a very simple rod model in which the DNA with contour length L is uniformly bent into an arc with radius R and end-to-end distance r (Figure 4.1A). We can estimate the shear force of the bent molecule in case $r \approx 0$. For a uniformly bent elastic rod with a radius of curvature R and an end-to-end

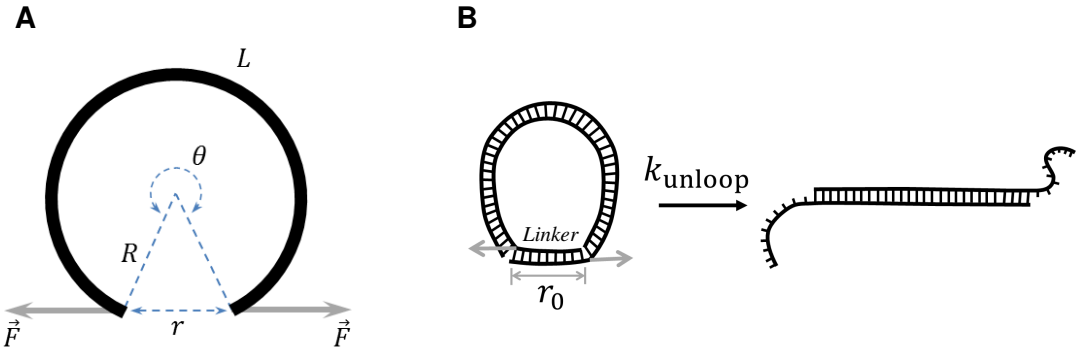


Figure 4.1: (A) A simple model to estimate the shear force of a uniformly bent elastic rod. (B) The shear force exerted on the linker duplex by the loop. The force exerted in the shear direction (gray arrows) accelerates dissociation of the linker duplex according to the Bell relationship (Equation 4.5).

distance r, the bending energy (Equation 1.7) becomes $E^{WLC}(r) = k_B T L_P L/2R^2$, where L_P is the persistence length and L is the contour length of the rod. The shear force at small end-to-end distance $r \approx 0$ can be calculated as:

$$f^{WLC}(r \approx 0) = -\left.\frac{\mathrm{d}E}{\mathrm{d}r}\right|_{r\to 0} = \frac{k_B T L L_P}{R^3}\left.\frac{\mathrm{d}R}{\mathrm{d}r}\right|_{r\to 0}. \tag{4.1}$$

Using a geometric condition for the dependence of the end-to-end distance r on the central angle $\theta = L/R$ (Figure 4.1A), we can relate r to R by: $r = 2R\sin(\theta/2) = 2R\sin(L/2R)$. From this formula, one can evaluate the derivative $\left.\frac{\mathrm{d}R}{\mathrm{d}r}\right|_{r\to 0}$ as

$$\left.\frac{\mathrm{d}R}{\mathrm{d}r}\right|_{r\to 0} = \left[\left.\frac{\mathrm{d}r}{\mathrm{d}R}\right|_{R\to L/2\pi}\right]^{-1} = 2\pi, \tag{4.2}$$

which gives the value of the shear force

$$f^{WLC}(r \approx 0) = \frac{4\pi^2 k_B T L_P}{L^2} \propto \frac{1}{L^2}. \tag{4.3}$$

If the bending energy depends differently on the bending angle, for example, a linear rather than a quadratic dependence as in the LSEC model[34, 35], the scaling law for the length dependence of the force might be different. Assuming $E(\theta) = E^{LSEC}(\theta) = k_B T\alpha|\theta|$, one can estimate that

$$f^{LSEC}(r \approx 0) = \frac{2\pi k_B T\alpha}{L} \propto \frac{1}{L}. \tag{4.4}$$

This estimation suggests that measuring the length dependence of the shear force in tiny DNA loops may help to differentiate among different models of DNA bending at short length scales where the bending energy is the dominant factor in the total free energy. A more elaborate calculation considers the shape of DNA loop in its minimum energy configuration, which is a teardrop shape. Thermal fluctuations, which manifest in an additional entropic term in the total free energy of the bending DNA loop, also affect the shear force. These points will be covered in the following sections where we compare the force estimations from experiments with the values calculated from Monte Carlo simulations to explore the elastic limit of dsDNA bending.

In our smFRET DNA looping assay, the measurable kinetics of loop breakage has a direct link to the shear force. Consider a linear dsDNA being bent and stabilized by long, sticky annealing overhangs, which we term the linker duplex (Figure 4.1B). Under this condition, the presence of a shear force will lower the energy barrier of the linker duplex in the annealing state, which facilitates the duplex dissociation via shearing. If there is a single energy barrier in the dissociation pathway of the annealing duplex, the lifetime (τ) of the linker duplex of length r_0 subjected to a shear force (f) can be modeled by the Bell relationship[106]

$$\tau(f) = \tau(0) \exp\left(-\frac{f\Delta r_0}{k_B T}\right), \tag{4.5a}$$

$$\log \tau(f) = \log \tau(0) - \frac{f\Delta r_0}{k_B T}, \tag{4.5b}$$

where Δr_0 is the elongation of the linker duplex at the transition state. Meanwhile, the dependence of shear force on loop size can be calculated from the thermodynamic relation

$$f(r_0) = -k_B T \left.\frac{\partial \log P(r)}{\partial r}\right|_{r_0} \tag{4.6}$$

where $P(r)$ is the equilibrium radial distribution function of end-to-end distance r of a polymer. The central idea of this study is that we measure the loop lifetime τ as a function of loop size L and compute the shear force using Monte Carlo simulation and Equation 4.6. We then correlate them using Equation 4.5b to extract two fitting parameters, $\tau(0)$ and Δr_0. These predicting parameters are compared to additional experimental measurements to check for their consistency, which help us to identify the DNA model that best explains our experiment and simulation data.

4.3 Experimental design and methods

4.3.1 Materials

The DNA molecules used in this unlooping assay have a double-stranded part with variable length from 37 bp to 189 bp and 13-nucleotide (nt) long single-stranded complementary overhangs (sticky ends). One overhang contains Cy3, and the other contains Cy5 and biotin (Figure 4.2A). The sequences of these overhangs are ATAG/iCy5/GAATTTACC, where /iCy5/ represents the internally labeled Cy5, and GGTAAATTCACTAT with the underlined 'A' inserted as a spacer opposite to iCy5 to increase the likelihood of base pairing around iCy5 that interrupts the backbone. All DNA molecules are derived from a master sequence that is ∼50% in GC content and does not have curvature-inducing patterns such as GGGCCC or A-tracts. The master sequence was constructed by annealing the ends of two 113-nt long single-stranded DNAs over a 16-nt region and extending their 3′-ends using DNA polymerase. The 210-bp master DNA was purified by gel electrophoresis, and PCR-amplified with dangling-end primers to generate DNAs with common terminating sequences. The annealing location of one of the primers was varied to generate DNAs with different lengths. These PCR products were used as templates in another round of PCR to incorporate fluorescent labels and a biotin as previously described[45]. Strands were exchanged between these PCR products to obtain the final DNA constructs for our experiment. Detailed sequences are listed below (from $5' \rightarrow 3'$, bold regions represent the universal adapters):

Master 210 bp DNA: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTA

CCTACAAGCCCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCC
TTCTTAACGACACGTGAAGGGACGAACCGCATACTTACGATCAGGCATAGAT
CTTACACCGTAGCAGGTAGTGCCAGGCATCGTGTTCGTAACCTTACTTCAAC
CATTCGAGCTCGTTGTTG

189 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCCTTCTTAACG
ACACGTGAAGGGACGAACCGCATACTTACGATCAGGCATAGATCTTACACCGT
AGCAGGTAGTGCCAGGCATCG**CATTCGAGCTCGTTGTTG**

168 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCCTTCTTAACG
ACACGTGAAGGGACGAACCGCATACTTACGATCAGGCATAGATCTTACACCGT
CATTCGAGCTCGTTGTTG

147 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCCTTCTTAACG
ACACGTGAAGGGACGAACCGCATACTTACGAT**CATTCGAGCTCGTTGTTG**

136 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCCTTCTTAACG
ACACGTGAAGGGACGAACCGC**CATTCGAGCTCGTTGTTG**

126 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCCTTCTTAACG
ACACGTGAAGG**CATTCGAGCTCGTTGTTG**

115 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTACGTCTGCCACCCTTCTTAACG
CATTCGAGCTCGTTGTTG

94 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCCCCGCCAGTTA**CATTCGAGCTCGTTGTTG**

84 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC
CCAATAGCGTACGGGATCATCC**CATTCGAGCTCGTTGTTG**

74 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC

CCAATAGCGTAC**CATTCGAGCTCGTTGTTG**

63 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTACCTACAAGC C**CATTCGAGCTCGTTGTTG**

53 bp: **GTGCCAGCAACAGATAGC**CTATCCATAGACTATTA **CATTCGAGCTCGTTGTTG**

42 bp: **GTGCCAGCAACAGATAGC**CTATCC**CATTCGAGCTCGTTGTTG**

37 bp: **GTGCCAGCAACAGATAGC**C**CATTCGAGCTCGTTGTTG**

4.3.2 Single-molecule unlooping assay

The DNA molecules were immobilized on a PEG-coated glass surface through NeutrAvidin-biotin interaction (Figure 4.2A), which is similar to the smFRET DNA looping experiment described in Chapter 3. Cy3 and Cy5, the donor-acceptor pair for FRET are incorporated near the sticky ends of the DNA so that loop stabilization by the sticky ends results in high FRET efficiency. A biotin linker extends from one end for surface immobilization (Figure 4.2A). The power of the 532 nm laser was $\sim 5\mu$W when measured after the microscope objective before reaching the critical angle of incidence.

In the loop breakage assay, immobilized DNA molecules were first incubated in 2 M NaCl buffer for up to an hour to generate looped molecules. The high salt concentration and long time incubation was employed to generate as many looped molecules as possible[42]. Once equilibrium is reached, we introduced the imaging buffer (5 mM PCA, 100 mM PCD, 1 mM Trolox) that contains 2 M NaCl to start image acquisition. After 20 seconds, new imaging buffer with 50-200 mM NaCl was perfused into the imaging channel at a flow rate of 75 μL/min, which corresponds to $\sim$1 cm/s in flow velocity through the channel. The typical dimension of the channel cross-section is 0.075 mm $\times$ 2.0 mm. The Cy3 and Cy5 fluorescence intensities are continuously monitored. We recorded the times it takes for molecules to unloop from single-molecule time traces (Figure 4.2B), built the survival time histogram of the number of remaining high-FRET molecules, and fitted it with a single exponential function to extract the linker lifetime (Figure 4.2C).

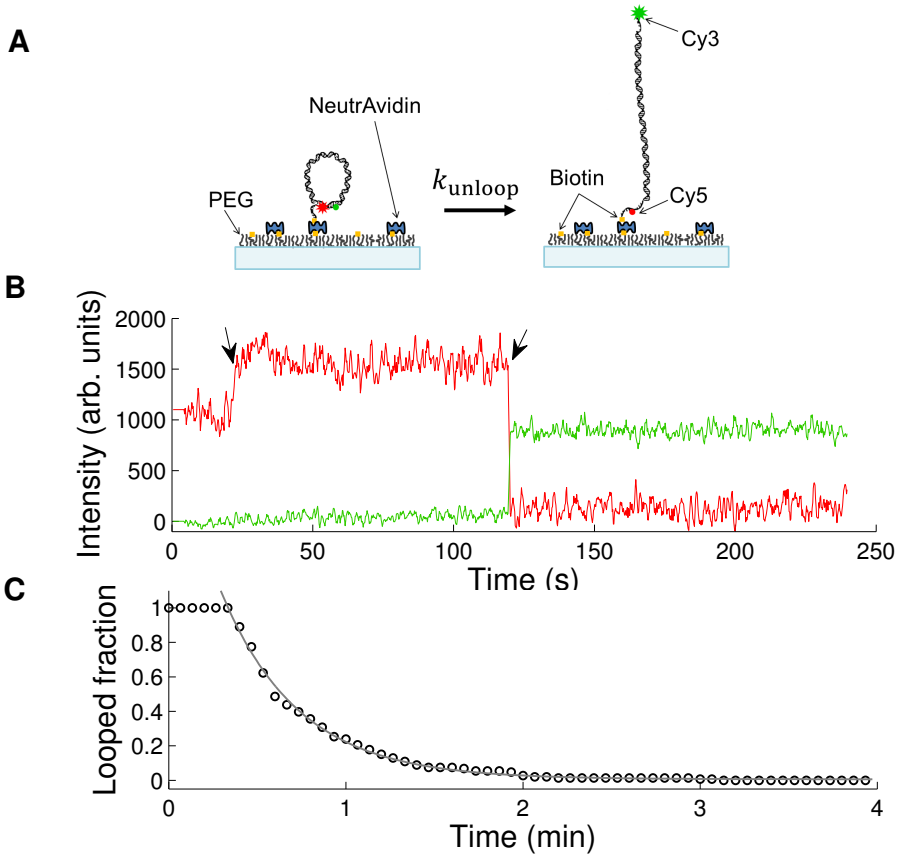


Figure 4.2: Loop breakage assay. (A) The shear force exerted on the linker duplex by the loop. The force exerted in the shear direction (gray arrows) accelerates dissociation of the linker duplex according to the Bell relationship (Equation 4.5). (A) DNA design. A short DNA with sticky ends can be captured in the looped state when the two sticky ends are annealed. The looped state (left) and the unlooped state (right) correspond to high FRET and low FRET, respectively. For single-molecule experiments, DNA molecules are immobilized through a biotinylated end to a PEG-coated glass surface. In $[Na^+] = 2$ M, a significant fraction of molecules exist in the looped state. Decreasing [NaCl] from 2 M to 50 mM by flow induces breakage of DNA loops. (B) A representative time trace of Cy3 (green) and Cy5 (red) intensities from a single molecule. The change in salt concentration causes an increase in the Cy5 intensity due to an unknown reason (marked by a black arrow). Upon loop breakage, Cy5 intensity drops, and Cy3 intensity jumps (marked by a black arrow). (C) The time decay of the number of dsDNA loops upon salt concentration drop. The molecules begin to unloop shortly after perfusion of 50 mM $[Na^+]$ buffer. The decay curve is fitted with a single exponential function to extract the lifetime of the DNA loop.

4.4 Smaller DNA loops are less stable

We conducted this salt drop experiment in 50 mM NaCl for different lengths of DNA molecules ranging from 40 to 200 bp. In this length range, the bending energy dominates the free energy of looping. Since the total bending energy of the loop increases as the loop size decreases (Figure 4.7A), we expect smaller loops to become less stable. In support of this notion, the linker lifetime decreased as the DNA length was reduced (Figure 4.3).

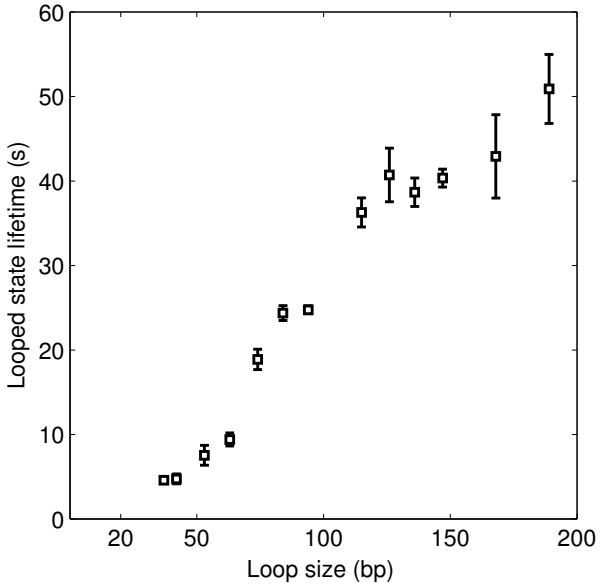


Figure 4.3: Looped state lifetime vs. loop size. The loop lifetime at various size was measured in 50 mM NaCl. The error bar is the standard error of the mean from at least 4 measurements.

4.5 Shear force calculations

4.5.1 DNA models and parameters

Since the DNAs used in this study have random sequence with negligible intrinsic curvature, we used a discrete, isotropic DNA model, which is the random-ϕ model[107]. In this model, dsDNA was treated as a chain of rigid monomers or links. Because of thermal fluctuations, link $i+1$ is deviated from link i by a local bending angle of $\theta_{i,i+1}$. Since we assume that there is no preferred local axis for the bending, the torsion angle $\phi_{i,i+1}$ are equally populated. The random-ϕ model is a simplified version of the wedge-angle model which was described in Chapter 3. In the random-ϕ model, the roll angle ρ is θ, the tilt angle τ is set to be zero,

and the twist angle Ω (or ϕ) is randomly distributed over 2π.

We considered two continuous polymer models: the WLC model and the LSEC model. The WLC model is the canonical elastic DNA model with a quadratic dependence of deformation energy on bending angle. In comparison, the LSEC model assumes a linear relationship between them, and has been proposed as a phenomenological DNA model in the strong bending regime[34, 35]. The bending energy was assigned to each angle between adjacent monomers. Thus, the Hamiltonian was the sum of the total bending energy of the polymer from all monomer steps $\sum_{i=1}^{N-1} \frac{k}{2}\theta_{i,i+1}^2$ for WLC and $\sum_{i=1}^{N-1} B|\theta_{i,i+1}|$ for LSEC where $\theta_{i,i+1}$ is the angle between the i-th monomer and the i+1-th monomer. For the WLC model, each base pair was treated as a monomer, similar to the dinucleotide model. On the other hand, the LSEC model assumes that DNA has a linear bending energy function with a monomer length of $\sim$7bp[35, 34].

The length of the monomer and the value of the rigidity constant are chosen so that the known statistical mechanical properties of the polymer can be reproduced by simulation. In particular, the elastic parameters of both models are strongly constrained by the persistence length of $\sim$50 nm in the long limit. For the WLC model, we chose the bending rigidity constant k to be $73.53k_BT$ for each 1-bp long monomer. For the LSEC model, we chose $B = 7.84k_BT$ for each 7-bp long monomer (2.37 nm). To confirm that the chosen parameters in both models correctly reflect the elasticity of dsDNA in the long limit, we simulated a set of DNA conformations in thermal equilibrium using Gaussian sampling simulation. We then collected the end-to-end distance in each DNA conformation and calculated the mean square end-to-end distance $\langle R^2 \rangle$ of the DNA as a function of the contour length L up to 300 nm. A linear dsDNA molecule longer than the persistence length can be well described as a worm-like chain, and the analytical formulae for the mean-square end-to-end distance is available. The simulated mean square end-to-end distance is then compared to a theoretical calculation assuming an elastic chain

$$\langle R^2 \rangle = 2L_pL\left(1 - \frac{L_p}{L}\left(1 - e^{-L/L_p}\right)\right), \tag{4.7}$$

where L_p is the persistence length and L is the contour length. As shown in Figure 4.6A,

the two models fit well with the analytical formulae with a persistence length of 50nm.

We also considered the KWLC model[36] that allows for kink formation at large bending angles. The bending energy for the (i,i+1)-th dinucleotide step is $\min\left(\frac{1}{2}k\theta_{i,i+1}^2, h + (\theta_{i,i+1} - b)^6\right)$. In this formula, k is the bending rigidity which is the same as in the WLC model, and h is the energy barrier for kinking[36]. b was fixed to 0.3 radians (if not mentioned otherwise) to allow the kinks to adopt bending angles up to 90 °.

4.5.2 Monte Carlo simulation for dsDNA bending with Metropolis criterion

To sample the chain conformations in thermal equilibrium, the Metropolis-Monte Carlo procedure includes a consecutive number of displacements of the chain parts. A new trial conformation can be accepted or rejected depending on whether its total energy follows a Boltzmann distribution. If the trial conformation is rejected, the current conformation is stored and used as the template to generate new conformations until the new one is accepted. The starting conformation is chosen arbitrarily. The rule for accepting or rejecting a new conformation is the followed:

1. If the total energy of the new conformation, E_{new} , is lower than the energy of the previous conformation, E_{old}, the new conformation is accepted.

2. If the energy of the new conformation is greater than the energy of the previous conformation, the new conformation is accepted with the probability $p = \exp(-\Delta E/k_B T)$, where $\Delta E = E_{new} - E_{old}$.

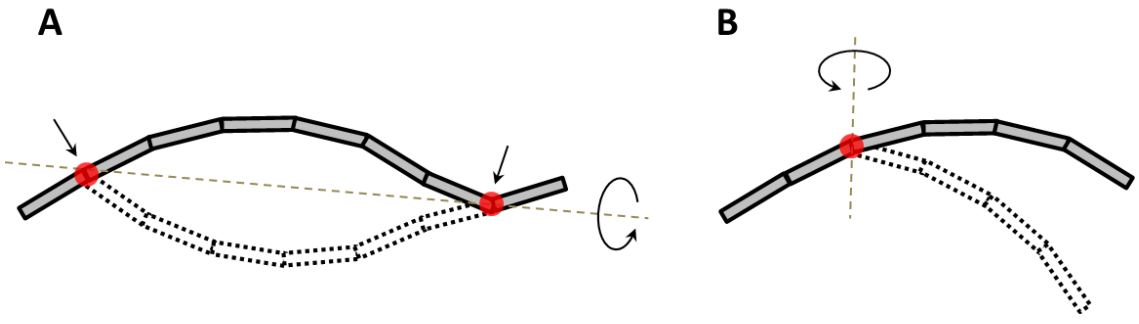


Figure 4.4: Crankshaft (A) and pivot (B) rotations. The vertices for the rotation axis are marked by red circles.

To generate new conformations, we employed two types of the displacements[108]. The first type it the crankshaft rotation. In this rotation method, a subchain is rotated by a randomly chosen angle, α, around an axis that connects two randomly chosen vertices of the chain (Figure 4.4A). This type of displacement preserves the end-to-end distance of the chain. In the second type of displacement, which is the pivot rotation, a randomly chosen subchain, which includes one end of the chain, is rotated by a random angle, α, around a randomly oriented axis passing through the internal end of the subchain (Figure 4.4B). The values of α are usually uniformly distributed over an interval $[-\alpha_0, \alpha_0]$. The value α_0 can be adjusted to increase the acceptance rate of the trial conformations.

4.5.3 Biased Monte Carlo simulation for simulating small DNA loops

Simulating small loop conformations using Gaussian sampling (importance sampling) is time-consuming. For example, if we estimate the fraction of a 190bp DNA in the looped conformation with a 5nm end-to-end distance using a non-biased MC simulation (see Section 4.5.2), we can see that this fraction is extremely small ($< 10^{-6}$), see Figure 4.5. With a typical data set of $\sim 5 \times 10^5$, such small number cannot be calculated accurately. Therefore, we used a biased sampling method to calculate the conformational distribution of rare events, which is also called the "umbrella sampling" method. The method is based on the idea of introducing an artificial potential, $U(x)$, near the reaction coordinate of interest. Here, x refers to the reaction coordinate. With a proper choice of $U(x)$, the probabilities of very rare conformations can be greatly increased. We applied this method here with a biased potential $U(x)$ on the end-to-end distance ($x \equiv r$) to increase the fraction of chains in the looped states with the end-to-end distances around $r = r_0$.

Here, we used a harmonic potential $U(r) = K(r - r_0)^2/2$ with stiffness K that restrains the end-to-end distance near r_0. When the biased potential is the dominant term in the Hamiltonian, one should expect that the chain distribution will resemble a Gaussian distribution as the result of a harmonic potential $U(r)$. Except for the bias potential for umbrella sampling, we did not apply constraints on relative bending or torsional angles between the

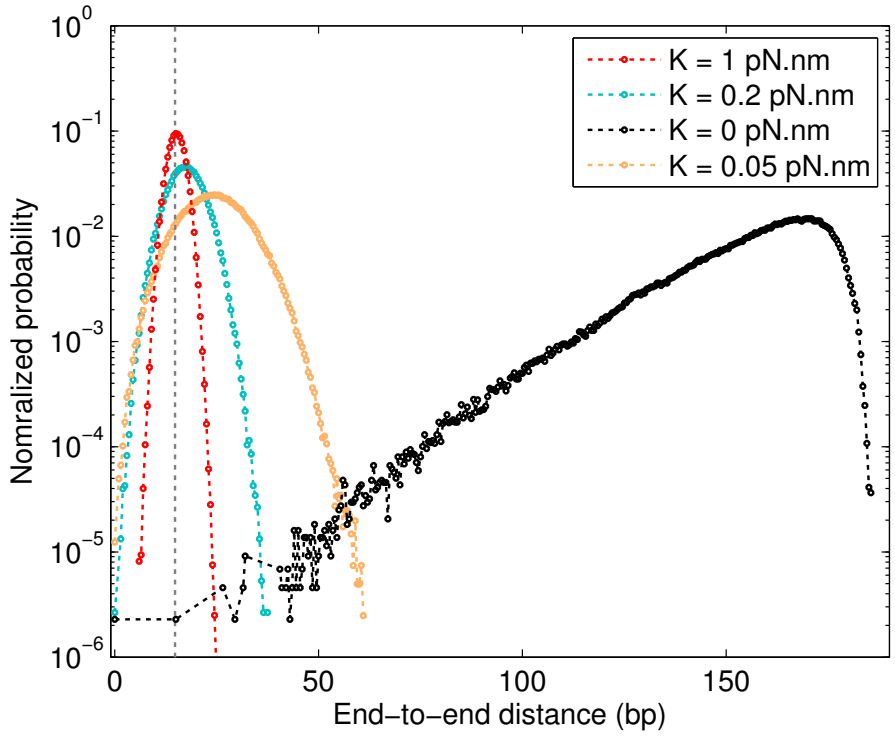


Figure 4.5: The normalized radial probability distribution $P(r)$ of a 190-bp DNA with 50nm persistence length, sampled at different K values for the biased potential $U(r) = K(r - r_0)^2/2$, where r_0 is the linker length (vertical dashed line). The curve with $K = 0$ corresponds to a distribution in a non-biased sampling without the restrained potential $U(r)$. Sample size: 5×10^5 chains, bin size: 0.5 bp.

two ends because flexible gaps at the ends of the linker effectively relax bending and torsional stress. The lack of angular constraints in the loop geometry of our DNA construct is supported by the observation that the J-factor of DNA with gaps does not oscillate with the helical phase of DNA[109], in contrast to intact DNA circles[105].

By increasing the strength K of the biased potential, the fraction of molecules in the looped state can be greatly enhanced, as apparent from Figure 4.5. For example, a biased potential with K as small as 1 pN.nm increases the fraction of conformations with $r \approx r_0$ to $\sim 10\%$. The spring constant K for the biasing potential in the case of the WLC model and the LSEC model was set to 8 pN.nm/(1 bp)^2 and 400 pN.nm/(7 bp)^2, respectively.

In principle, the force can be obtained from the derivative of the unbiased radial probability distribution at r_0 according to Equation 4.6, but it is impractical in general because it is difficult to estimate $P(r \approx r_0)$ directly. With the biased potential restraining the two

ends of the chain, the biased force (f^b) is then given by[110]

$$f^b(r) = f^u(r) + K(r - r_0). \tag{4.8}$$

Thus, the unbiased force (f^u) is equal to f^b if evaluated at r_0, which enables us to use Equation 4.8 to calculate f^u directly from a biased radial probability distribution. Since derivatives are sensitive to statistical noise, we instead used an approximation that contains averaging[111]

$$f(r_0) = -\frac{k_B T}{\mathrm{var}(\delta r)} \langle \delta r \rangle, \tag{4.9}$$

where δr is the deviation of the end-to-end distance from r_0. $\langle \delta r \rangle$ and $\mathrm{var}(\delta r)$ are the mean and the variance of these deviations, respectively. Note that this estimation is valid only when the biased distribution is very close to a Gaussian distribution, which can be obtained by a large enough K value (Figure 4.5). Pivot moves were used to sample the conformational space of the chain, and Metropolis criterion was applied to accept conformations consistent with the Boltzmann distribution. The chain was equilibrated for 10^5 MC steps starting from the minimum energy conformation, and approximately 5×10^6 conformations after equilibration were used to obtain $P(r)$. The calculated force for a specific loop size did not depend on the value of K. For the WLC and the LSEC models with monotonically increasing bending energy, the calculated force varied little between simulations. For the KWLC model with a discontinuous slope, the calculated force for small loop sizes was more variable and, therefore, we increased the number of simulations until the SEM was smaller than 8% of the mean.

4.5.4 Shear force for DNA loops with varying size

We performed biased MC simulations to calculate the shear force as a function of loop size in two DNA models: the WLC model and the LSEC model. While both models can describe the elastic property of dsDNA at large length scales (Figure 4.6A), the LSEC model generally predicts high-curvature conformations more frequently than the WLC model[34, 35]. In agreement with this conclusion, our MC simulations using the WLC and the LSEC models yield similar shear force for DNA around 200bp (Figure 4.6B). Furthermore, for

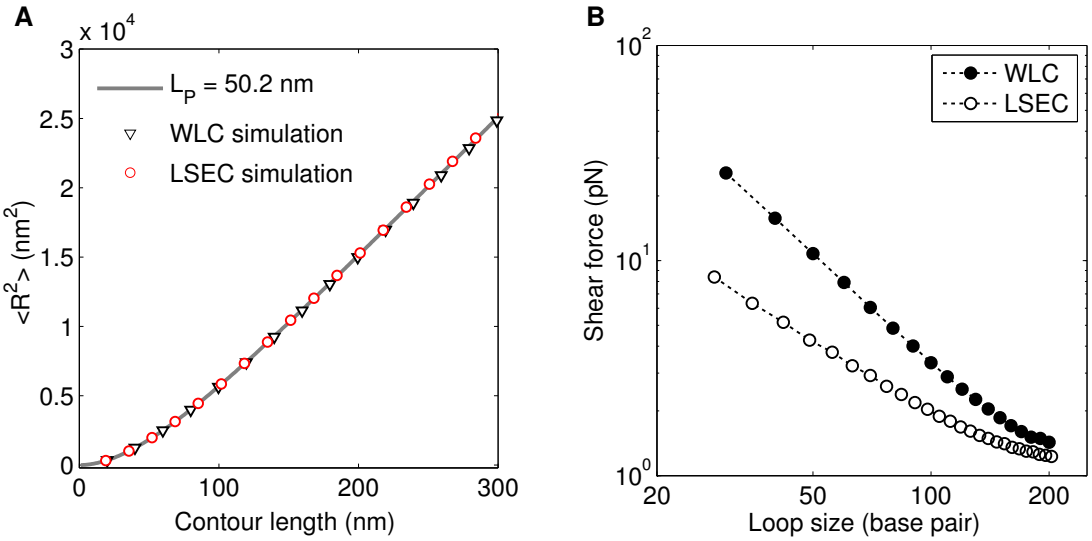


Figure 4.6: (A) The mean square end-to-end distance of dsDNA chain with varying contour length up to 300 nm. DNA chains were sampled in a non-biased manner ($K = 0$) using the WLC model (red circles) and the LSEC model (black triangles) and the mean square end-to-end distance was calculated from the chain ensemble. In comparison, we also showed a theoretical calculation of the mean square end-to-end distance assuming a worm-like chain (Equation 4.7). (B) Line-scatter log-log plot of calculated shear force vs. loop size. The shear force is calculated from the biased MC simulation using the WLC model with a persistence length of 50 nm (filled circles) or the LSEC model (open circles). For each loop size, we performed three simulations, each with $\sim 5 \times 10^6$ accepted conformations. The errorbar size is typically smaller than the size of the symbol.

shorter DNAs, we found that the LSEC model produces a significantly weaker shear force and a more moderate length-dependence than the WLC model (Figure 4.6B), which can be qualitatively explained by the simple rod model presented in Section 4.2.

For DNA much shorter than the persistence length, we expect that the contribution from thermal fluctuations to the shear force should become relatively small compared to the bending energy. To understand at which length limit that the entropic force is unavoidable, we implemented the elastic rod approximation to estimate the shape and the bending energy of DNA in the ground state[24, 25]. This is a semi-analytical approach in which the bending energy of a DNA polymer in its minimum energy configuration with an end-to-end distance r can be calculated. We extracted the bending energy $E(r)$ as a function of the end-to-end distance for these teardrop-shaped DNA loops and calculated the shear force by taking the derivative, $\frac{\mathrm{d}E}{\mathrm{d}r}\big|_{r=r_0}$. We also used constrained nonlinear optimization toolbox in MATLAB

to directly minimize the total bending energy for a comparison. As expected, our calculated WLC shear force from MC simulations is in excellent agreement with the result from the elastic rod approximation below 100 bp (Figure 4.7A). We also note that the calculated shear force depends only weakly on r_0 near the value chosen for our analysis (Figure 4.7B).

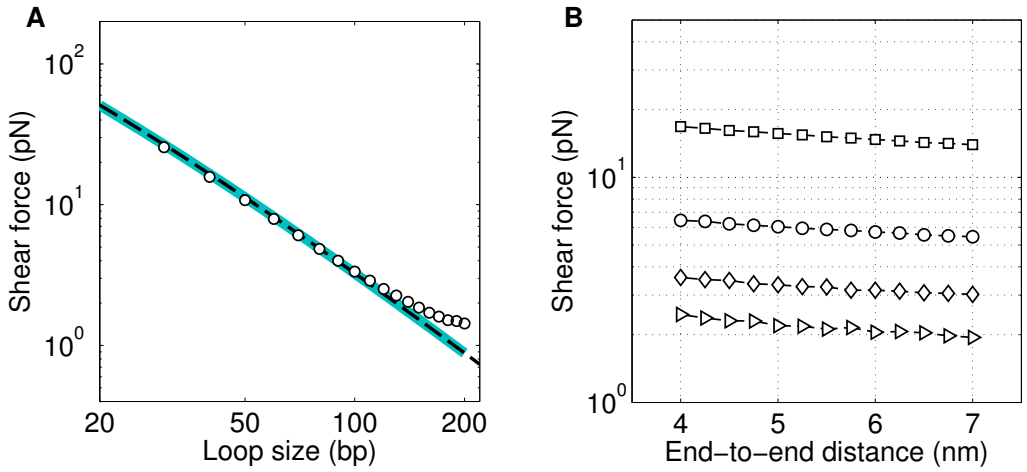


Figure 4.7: (A) The shear force vs. loop size. Open circles represent shear force obtained from MC simulations for the WLC. The relationship is plotted on log-log axes to highlight the scaling. The elastic rod approximation was also used to calculate the shear force by using either elliptic integrals (black dashed curve)[24, 25] or the constrained nonlinear optimization in MATLAB (thick cyan curve). Two calculation methods yielded identical results. All forces were evaluated at 5-nm end-to-end distance. (B) Shear force vs. end-to-end distance. The mean shear force was calculated from the WLC model for different loop sizes and different end-to-end distances (r_0). The shear forces at different loop sizes (square: 40 bp, circle: 70 bp, diamond: 100 bp, triangle: 130 bp) decrease only slightly as a function of the end-to-end distance. Since the linker duplex is extended by ~ 1 nm before dissociation, our estimated force can be variable by $\sim 5\%$ for all loop sizes tested.

4.6 Analysis of the linker lifetime vs. shear force reveals softening transition of a worm-like chain

We plotted the logarithm of the measured lifetime vs. the calculated forces, which is expected to be a straight line according to Equation 4.5. As shown in Figure 4.8, the overall relationship follows a straight line between 60 and 200 bp, but deviates from it at smaller loop sizes.

To analyze the linker lifetime vs. shear force, we performed linear regression with the

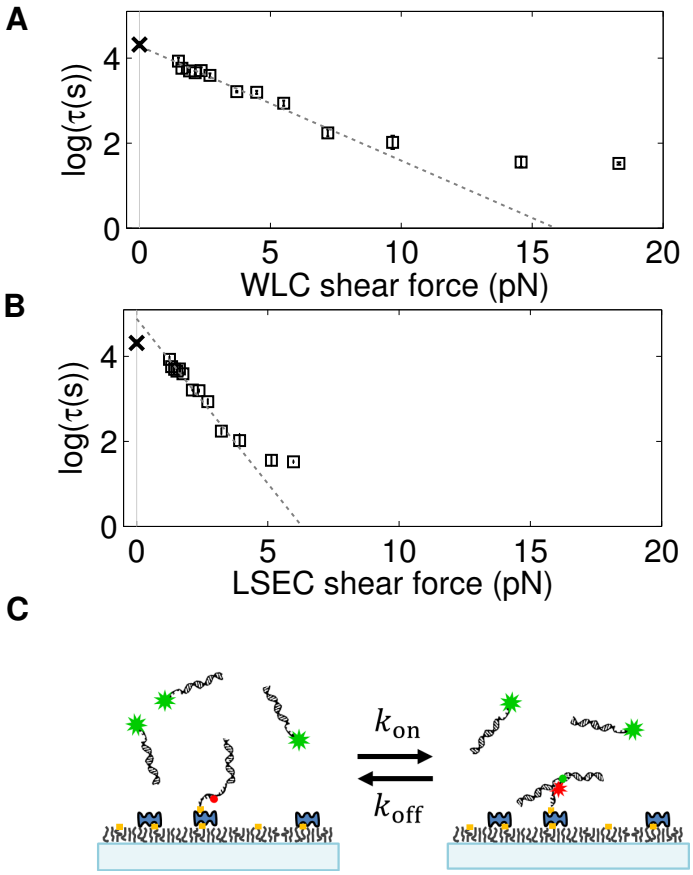


Figure 4.8: The relationship between linker lifetime and shear force. The natural logarithm of the lifetime measured in 50 mM NaCl is plotted as a function of shear force calculated from the WLC (A) and the LSEC model (B). Linear regression yields the zero-force lifetime ($\tau(0)$) and the separation distance (Δr_0) for duplex dissociation. Only the lifetimes for DNA loops larger than 60 bp are included in the regression. Data for loop size less than 60 bp are excluded from the linear regression based on the RMSE analysis (see Figure 4.9). (C) The zero-force lifetime ($\tau(0)$) measurement from dissociation kinetics of a linear dimer. In this experiment, the linker formed between the sticky ends of the DNA molecules does not experience a shear force, and therefore, the dissociation lifetime corresponds to $\tau(0)$. DNA molecules are composed of an 18-bp duplex and a 13-mer single-stranded overhang, and are identical to the end-segments of the DNA molecule as depicted in Figure 4.2A. The zero-force lifetimes (marked '$\times$') averaged from four measurements are plotted in (A) and (B) for comparison with the two models. The error bar for this data point is smaller than the symbol.

'robustfit' function (MATLAB). We also examined how the goodness of fit changes with the range of fitting using the standard regression error or RMSE (root mean squared error) as an indicator. As shown in Figure 4.9, the RMSEs for both WLC and LSEC models increase significantly when points below ∼60 bp were included. This analysis indicates that Equation 4.5 does not hold below this length because the calculated forces are overestimated compared to the actual forces exerted on the linker. This indicates a softening transition of the loop where the actual force becomes weaker than the force predicted by each model. Therefore, we did not include these points when extracting the fitting parameters, $\tau(0)$ and Δr_0 for WLC and LSEC models.

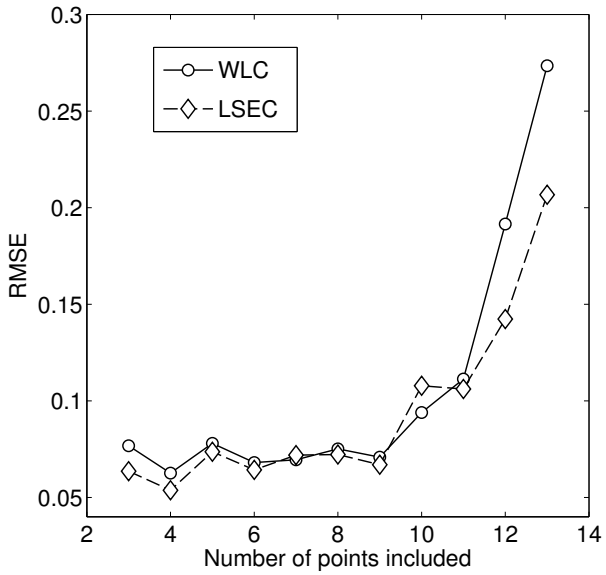


Figure 4.9: RMSE analysis. The linear regression was performed with the 'robustfit' function (MATLAB) on the logarithm of loop lifetime vs. shear force. To identify outliers, we compared the RMSE (root mean squared error) values resulting from different ranges of fitting. For example, the last point is obtained when the entire range of 13 loop sizes from 189 bp down to the smallest 37 bp were included in the fitting. Including the last few points significantly increases the regression error, which indicates that the linear relationship predicted by Equation 4.5 no longer holds for loop sizes smaller than 60 bp. Thus, we did not include three points corresponding to 37, 42, and 53 bp in the regression when extracting the fitting parameters, $\tau(0)$ and Δr_0.

The relationship in the linear regime can be fitted with Equation 4.5b to obtain the negative slope (Δr_0) and the y-intercept ($\tau(0)$), both of which are related to the dissociation kinetics of the linker duplex. Since the WLC and LSEC models predict markedly different

Δr_0 (1.10 ± 0.14 nm vs. 3.18 ± 0.48 nm) and $\tau(0)$ (72.24 ± 10.28 sec vs. 132.83 ± 6.20 sec), we can compare these fitting parameters with experimental values to identify the correct model before the softening transition.

The linker lifetime with zero shear force, $\tau(0)$, can be measured using the same linker without the loop. For this experiment, we prepared two separate DNA molecules identical to the end-segments of the DNA used in the unlooping assay so that they can form the same linker without the shear force. The partially hybridized DNA molecules for $\tau(0)$ measurement (italic: double-stranded region) are:

Cy3-DNA:

5′ - Cy3-GGTAAATTCACTAT*CAACAACGAGCTCGAATG* - 3′

3′ - *GTTGTTGCTCGAGCTTAC* - 5′ (blocking oligo)

and Cy5-DNA:

5′ BiotinTEG - GAAACATAG/ iCy5 /GAATTTACC*GTGCCAGCAACAGATAGC* - 3′

3′ - *CACGGTCGTTGTCTATCG* - 5′ (blocking oligo)

We mixed equal amounts of the two partially hybridized DNA molecules in annealing buffer (100mM NaCl, 10 mM Tris·HCl pH 7.0, 1 mM EDTA) to obtain a final concentration of 5 μM. The mixture was heated at 95°C for 5 minutes, slowly cooled down to room temperature, and loaded on a polyacrylamide gel (19:1 Acryl:Bis, 15% (w/v) in TBE 1X pH 8.0). Linear dimers were extracted from the gel using an electroelution kit (G-CAPSULE, 786-001, G-Biosciences) after running the gel at 10 V/cm for $\sim$1 hour (Figure 4.10A). We immobilized the Cy5 DNA on the surface and introduced the Cy3 DNA at $\sim$20 nM concentration (Figure 4.8C). Linker formation and separation resulted in two-state fluctuation in Cy5 intensity due to FRET (see a typical time-lapsed trace for the Cy5 intensity in Figure 4.10B). Linker separation could be well-described by first-order kinetics, from which the lifetime was extracted. We find that the measured $\tau(0)$ (marked '$\times$' in Figure 4.8) agrees well with the WLC model prediction, but not with LSEC.

On the other hand, Δr_0 was previously measured to be 1 Å per base pair by pulling short DNA duplexes at opposite 5′-ends[112]. In our stretched linker duplex, the total

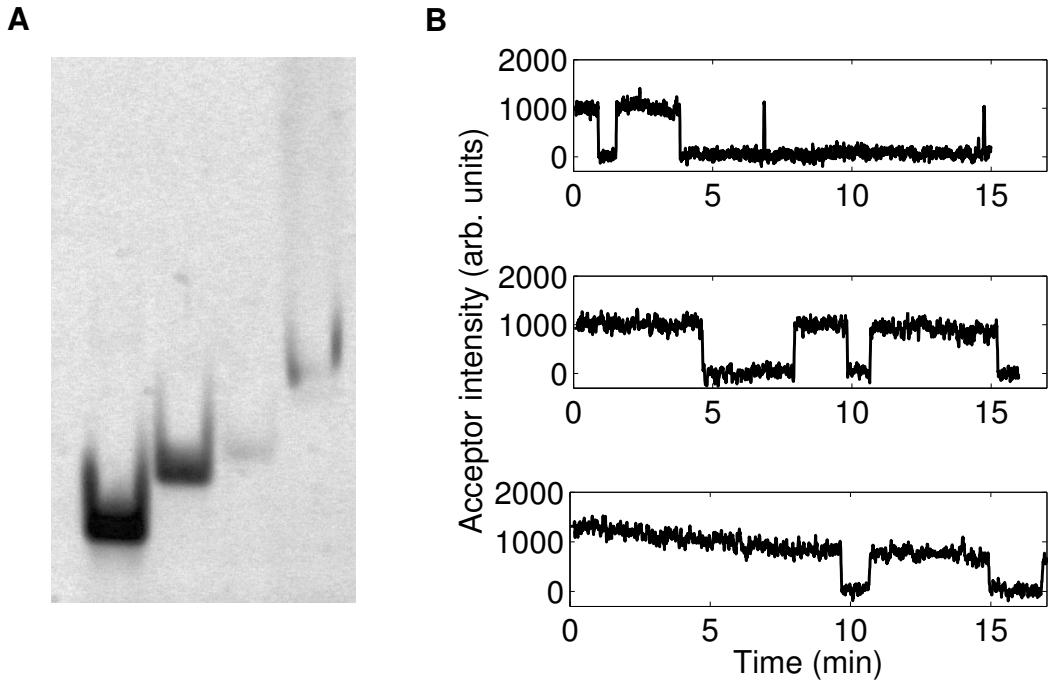


Figure 4.10: $\tau(0)$ measurement. (A) Polyacrylamide gel image of the hybridized oligos. From left to right, primer 1 only, 1:1 mixture of primer 1 and its blocking oligo, primer 2 only, Lane 4: 1:1 mixture of primer 2 and its blocking oligo. (B) Typical time traces of reversible linker formation and separation in 50, 100 and 200 mM [Na^+] (from top to bottom). Linker formation results in a burst in Cy5 intensity due to FRET. The survival probability of the dimer since $t = 0$ is fitted with a single exponential function to extract the linker lifetime at zero force $\tau(0)$. The concentration of the free monomer was adjusted to obtain similar binding rates at different [Na^+].

number of complementary base pairs is 13, but the largest number of consecutive base pairs is 9 due to Cy5 in the backbone. Therefore, Δr_0 can be estimated to be in the range of 0.9 nm to 1.3 nm, which includes the prediction of the WLC model but not the LSEC model. Since both parameters Δr_0 and $\tau(0)$ are compatible with the WLC model, but not with the LSEC model, we conclude that the free energy of dsDNA loop as small as 60 bp is better described by the WLC model.

4.7 Softening transition is sensitive to magnesium

To confirm that our conclusion is not affected by duplex dissociation kinetics, we conducted the unlooping assay at different [Na^+] concentrations. As expected, the linker lifetime $\tau(0)$ was prolonged at higher salt concentrations (Figure 4.11A and Figure 4.10B). Despite

changes in loop lifetimes as a function of [Na^+], all curves exhibit a softening transition near 60 bp, and all $\tau(0)$ values (marked '$\times$' in Figure 4.11A) overlap nicely with the values extrapolated by the WLC model. This result further supports our conclusion that the WLC model correctly describes the free energy of dsDNA bending prior to the softening transition.

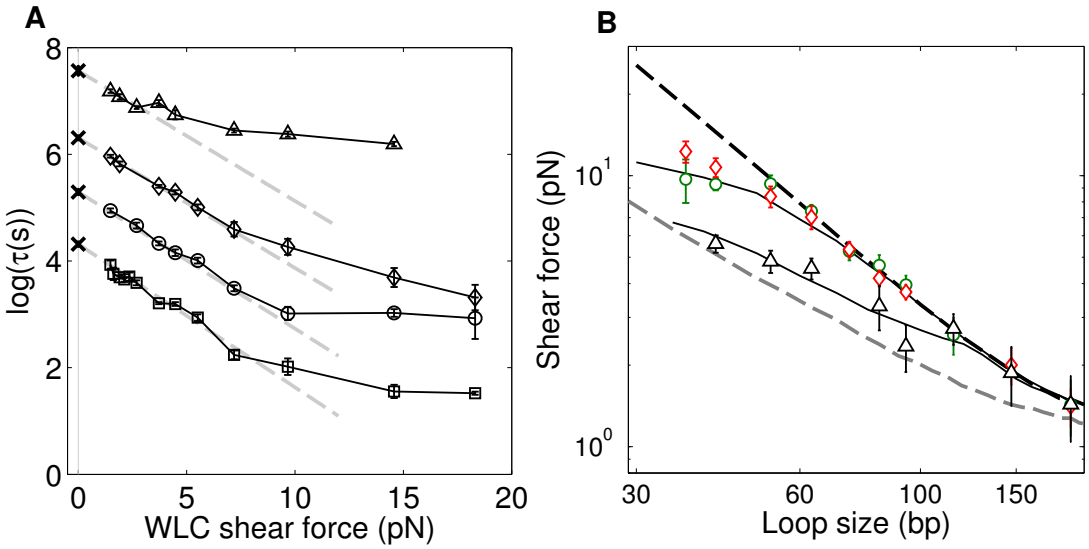


Figure 4.11: The effect of sodium and magnesium salt on strong dsDNA bending. (a) The logarithm of the measured loop lifetime is plotted as a function of the predicted WLC shear force in different [Na^+]s (squares: 50 mM, circles: 100 mM, diamonds: 200 mM) and in 5 mM [Mg^{2+}] (triangles). In sodium buffers, the softening transition appears at ~8 pN, which corresponds to a loop size of 60 bp whereas in 5 mM [Mg^{2+}], it is noticeable at ~3 pN, which corresponds to a loop size of 100 bp. For the WLC model, we assumed a constant persistence length of 50 nm[77, 113]. The lifetimes at zero force ('$\times$' symbols) were measured from the dimer dissociation experiments. For each salt condition, 3 separate measurements were performed. Linear fitting of data points in the elastic regime (gray dashed lines) yields almost identical negative slopes independent of [Na^+] concentration. (b) Shear force extracted from the unlooping experiment in 100 mM [Na^+] (green circles), 200 mM [Na^+] (red diamonds) and 5 mM [Mg^{2+}] (black triangles) are compared with predictions from three DNA models (the axis is in log-log scale). We optimized h and b of the KWLC model to fit the softening transition. The black curves represent the shear forces calculated from the KWLC model. For 50-200 mM [Na^+], best fit yields $h = 22k_BT$ and $b = 0.3$, and $h = 17k_BT$ and $b = 0.7$ for 5 mM [Mg^{2+}]. Also shown are the forces calculated from the WLC model (dashed black curve) and the LSEC model (dashed gray curve). The KWLC model with the lower h is similar to the LSEC model while the KWLC model with the higher h is similar to WLC. Below a certain loop size, the KWLC model predicts a smaller shear force than the WLC model because a kink relieves some of the bending stress.

We also investigated how magnesium affects strong bending of dsDNA. Magnesium is

essential for the activity of the ligase in the cyclization assay, and the restriction enzyme in DNA minicircle digestion. Therefore, almost all enzyme-based experiments on strong dsDNA bending have been performed in the presence of magnesium at relatively high concentrations (5-10 mM). Interestingly, we found that in the presence of 5 mM [Mg^{2+}], the softening transition of dsDNA occurs near 100 bp (Figure 4.11A). This result indicates that magnesium can dramatically increase the apparent flexibility of dsDNA in the strong bending regime.

4.8 Discussion

Using a FRET-based unlooping assay, we probed the energetics of dsDNA bending in the strong bending regime. We measured the loop lifetime as a function of loop size. In standard Na^{+} concentrations between 50 and 200 mM, the observed relationship in the range between 60 and 200 bp was consistent with the WLC model. Below 60 bp, we observed that dsDNA loses elastic rigidity, which leads to a weaker dependence of the shear force on the loop size. The critical loop size where softening occurs corresponds to a maximum bending angle of 7°/bp in a teardrop shape. In the presence of 5 mM [Mg^{2+}], the critical loop size increased to 100 bp, corresponding to 4°/bp. This result suggests that in cyclization experiments that typically use 10 mM [Mg^{2+}], subelastic bending can enhance the looping probability of dsDNA shorter than 100 bp.

The interpretation of our results relies on the Bell relationship between duplex lifetime and stretching force[106]. In general, a bond can dissociate through several different pathways[114], which may give rise to a nontrivial relationship between bond lifetime and the applied force. However, our assumption of the Bell model is justified by previous experimental studies[112, 115]. Notably, a DNA duplex pulled at the opposite 5′-ends by AFM, in the same shear geometry as in our DNA loop, exhibited strand separation kinetics consistent with a single energy barrier along the mechanical separation path. Also, the Chemla group recently demonstrated that DNA duplex dissociation under a constant tensile force follows the Bell relationship by combining fluorescence with optical tweezers[115]. In that study, the relationship between Δr_0 and duplex length (L) was extracted to be $\Delta r_0 = 0.096 \times L$

(nm), and has been more precisely determined as $\Delta r_0 = 0.256 \times (L - 6)$ (nm) (personal communication with Dr. Chemla). Either estimation puts Δr_0 to be in the range consistent with the WLC model but not with the LSEC model.

We showed that continuous polymer models with a single flexibility parameter are not sufficient to explain the observed loop lifetime vs. loop size relationship. It was pointed out that finite end-to-end distance of a loop can give rise to a non-monotonic dependence of looping probability on loop size[116]. But such geometric effect emerges only when the loop size is comparable to the end-to-end distance. Even with this geometric consideration, the shear force always changes monotonically with loop size (Figure 4.6B). Therefore, the observed deviation in loop lifetime vs. loop size relationship cannot be due to this effect. The breakdown of continuous models below the critical loop size is likely due to structural transition in the dsDNA helix such as kink formation that was originally proposed by Crick and Klug[30]. For free DNA, kinks are rare, transient deformations only occurring at a rate of $10^{-4}-10^{-5}$[117, 118], but they appear more frequently in sharply bent DNA[37, 119, 120]. Kinks are expected to have a significantly reduced bending rigidity, but our data indicates that the kink rigidity is not zero because, otherwise, the loop lifetime just below the critical loop size would be similar to $\tau(0)$. Therefore, we rule out the KWLC model with completely flexible kinks[31]. Instead we assume kinks to possess non-zero rigidity according to the KWLC model proposed by Vologodskii and Frank-Kamenetskii[36]. In this model, the dinucleotide bending energy (E) is given by

$$E = \min\left(\frac{1}{2}k\theta^2, h + (\theta - b)^6\right) \tag{4.10}$$

where k is the bending rigidity identical to that of the WLC model, h is the energy barrier of kinking, and b specifies the range of bending angles at the kink. This model is conceptually similar to the meltable worm-like chain model[32, 33].

To set the lower limit on the energy barrier for kink formation, we varied h while fixing b in our simulation to find h that is most compatible with the observed critical length of 60 bp. The parameter b was chosen to be 0.3 which allows kink angles up to 90°[36] based on other calculations and molecular dynamics simulations[37, 41]. As shown in Figure 4.11B,

$h = 22k_BT$ and $b = 0.3$ can produce a transition in the shear force below 60 bp, which is consistent with our observation. Using this h value, we can also calculate the free energy of kink formation ΔG_k. To calculate the free energy of kink formation, we adopted the computational method in [121], which is also conceptually similar to a more theoretical approach[33]. We considered the dinucleotide bending energy $E(\theta)$ with both the elastic bending term and kinking term using the functional form in Equation 4.10. The critical kink angle (β) was defined as the intercept of the two terms. The equilibrium probability density ($p(\theta)$) or the partition function of the bending angle θ is proportional to the multiplicity of $\sin(\theta)$ and the Boltzmann factor

$$p(\theta) \sim \sin(\theta) \exp(-E(\theta)/k_BT). \tag{4.11}$$

The kinking probability (P_k) is the probability for θ to exceed the critical kink angle β, which is

$$P_k = \frac{\int_\beta^\pi \sin(\theta) \exp(-E(\theta)/k_BT) d\theta}{\int_0^\pi \sin(\theta) \exp(-E(\theta)/k_BT) d\theta}. \tag{4.12}$$

The free energy of kink formation ΔG_k can be directly calculated from P_k as $\Delta G_k = -k_BT \log(P_k)$. In the condition with sodium only, the free energy of kink formation (ΔG_k) was calculated to be $\Delta G_k \approx 18k_BT$, which is similar to the upper limits of previous estimations[121, 120]. In comparison to $h = 22k_BT$ and $b = 0.3$ in the KWLC model, the lifetime vs. loop size relationship taken at 5 mM [Mg^{2+}] yields $h = 17k_BT$ and $b = 0.7$. These parameters correspond to a lower free energy of kink formation of $\Delta G_k = 12k_BT$, similar to a previous estimation[33], and larger kink angles up to 110°.

Using the parameters, h and b, constrained by our data, we can also determine the probability of kink formation in a DNA minicircle as a function of loop size. We performed a restrained MC simulation of DNA minicircles of various sizes (see Appendix A) to measure the frequency of large angle deflections in thermal equilibrium. In our simulation, we only consider the effect of bending stress on kink formation. As shown in Figure 4.12, in the absence of magnesium, kink formation is negligible even in 60-bp loops due to a high energy barrier. In the presence of 5 mM [Mg^{2+}], however, the kinking probability increases sharply with decreasing loop size, approaching unity at 70 bp while remains insignificant for DNA

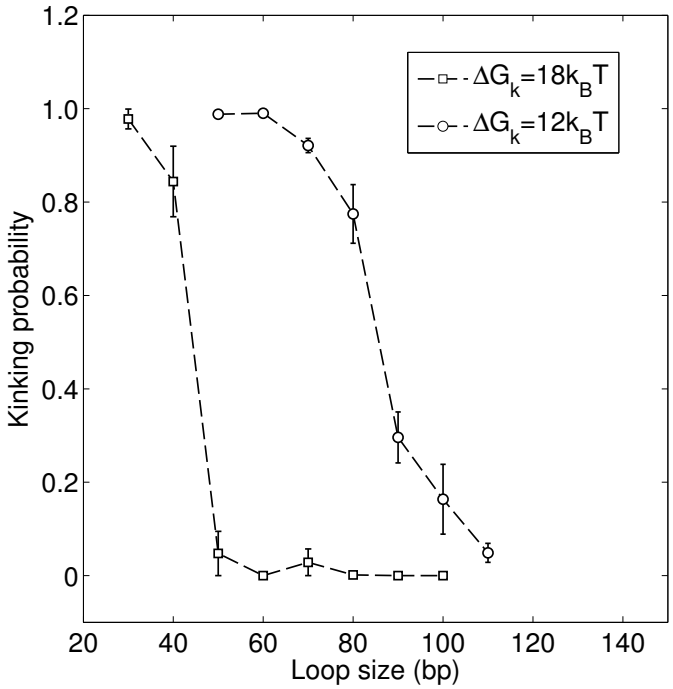


Figure 4.12: Kinking probability in DNA minicircles. The kinking probabilities of DNA minicircles were calculated as a function of loop size using the KWLC model with different free energies of kink formation (squares: $\Delta G_k = 18k_BT$ ($h = 22k_BT$, $b = 0.3$), circles: $\Delta G_k = 12k_BT$ ($h = 17k_BT$, $b = 0.7$). The SEM error bar for each loop size was calculated from 5 simulations.

over 100 bp. This simulation result agrees well with a previous minicircle digestion study that detected kinks in 60-bp minicircles due to bending stress alone[39]. Our simulations suggest that in such small DNA minicircles, since the kink is very stable, the enzyme can effectively detect and digest the local helical disruptions.

Our results suggest that magnesium can promote subelastic bending above a critical bending angle of 4°/bp by stabilizing large-angle deformations. This interpretation is similar to the conclusion of a recent study with DNA vises[120]. Therefore, we considered whether magnesium-facilitated softening could explain high J factors reported previously[42]. We thus calculated the J factor as a function of length using the KWLC model with $h = 17k_BT$ and $b = 0.7$ constrained by the data taken at 5 mM [Mg^{2+}]. We calculated the J factor by using the Weighted Histogram Analysis Method (WHAM) with multiple chain distributions obtained from several biased MC simulations (see Appendix A). As shown in Figure 4.13, while the KWLC model produces J factors similar to the WLC model prediction above 100

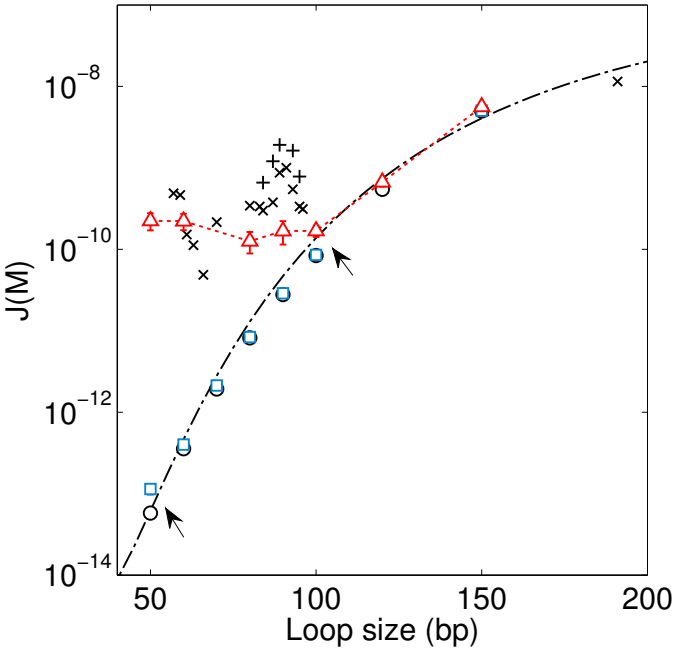


Figure 4.13: J factor comparison. The J factor was computed using a fixed end-to-end distance of 5 nm without end-to-end angular or torsional constraints. We used the weighted histogram analysis method to calculate J factors predicted by the WLC model (black circles), the KWLC model with $h = 17k_BT$ and $b = 0.7$ (red triangles), and with $h = 22k_BT$ and $b = 0.3$ (blue squares). The dash-dotted line is the WLC-prediction according to the theoretical approximation[24]. For comparison, the J factors from Vafabakhsh and Ha[42] are also shown. Symbols marked '×' indicate J factors measured from surface-immobilized DNA in 1 M $[Na^+]$ and symbols marked '+' from vesicle-encapsulated DNA at 10 mM $[Mg^{2+}]$. For consistency with our calculation, we shifted the original data[42] by 10 bp to the left to account for the linker length. The arrows indicate where the KWLC model deviates from the WLC model. We performed 4 simulations to generate the SEM error bars for each J factor.

bp as previously demonstrated[38], it also predicts substantially higher J factors for DNA below 100 bp, matching the J factors determined from the single-molecule FRET cyclization study[42] within a factor of 10. The agreement between our KWLC model and the result of the FRET study may be closer if the difference in the buffer condition (5 mM vs. 10 mM $[Mg^{2+}]$). In the absence of magnesium, however, our KWLC model predicts that the WLC model will be valid at least down to 55 bp (blue squares, Figure 4.13). This may explain why some studies lacking magnesium did not observe enhanced dsDNA flexibility at short length scales[122, 123].

Our unlooping assay enables investigation of strong dsDNA bending in buffer conditions not compatible with the ligation-based cyclization, the FRET-based cyclization, and the AFM assay. In the ligation assay, magnesium must be present at high concentrations for ligase activity. For AFM, magnesium is necessary to bind DNA to the surface[124]. In the FRET-based cyclization assay, high magnesium or sodium concentration is necessary to produce a statistically significant number of looping events. In this study, we demonstrated that effects of small amounts of monovalent and divalent ions on the elastic limit of dsDNA can be studied separately. Moreover, the unlooping assay is more well-suited to the study of kink formation than the cyclization assay because the probability of kink formation increases with bending stress. Our unlooping assay is similar in some ways to previous methods employing small DNA loops[125, 126, 120]. In these studies, electrophoretic mobility or intramolecular FRET of these loops was measured to investigate kinking. Our approach differs from theirs in two ways. First, we measure kinetic decay of the looped state instead of equilibrium distribution between alternative conformations in the looped state. Second, we do not need to include stretching or twisting energy in the Hamiltonian for single-stranded parts or twisted dsDNA. Therefore, our method allows a more direct link between the measurable quantities and dsDNA bending rigidity and holds great promise for studying the effect of sequence, salt, and temperature on strong dsDNA bending.

CHAPTER V

FUTURE PROSPECTS - SEQUENCE DEPENDENT KINKABILITY

5.1 Introduction

DNA kinks, which are the local disruptions in the helical structure, were proposed nearly 40 years ago[30]. They can play an important role in packaging DNA since they allow dsDNA to adopt more compact structures. The exact nature of this disruption is not known. It can correspond to unstacking between two adjacent base pairs or opening of a single base pair. It was suggested from an MD simulation[127] that while the GC-rich oligomers adopted a classical kink I, which is unstacking of a single base-pair step, distortions in A tracts tended to span three consecutive base pairs, which were termed kink II. In terms of energy, opening of a single base pair is much more costly than unstacking of a single dinucleotide because single base pair opening requires not only melting of the base pair, but also unstacking of two dinucleotides. However, a melting spot is generally more flexible than an unstacked step. Several experiments have established that the base pair opening requires a free energy of disruption of 7 kcal/mol[128, 129, 130].

Kinks can be realized in protein-DNA complexes such as in the nucleosomes[131]. For example, the TA step, which has the lowest stacking energy in conjunction with weak base pairing interaction, was found to be located on the inside, facing the surface of the octamer in some of the strongest positioning sequences[132, 133, 134]. Perhaps, the positioning of TA steps at the contact points between the DNA and the histone octamer can minimize distortion energy because of the kinkability of TA step[135, 136].

The importance of kinks was also suggested in the context of DNA looping. End-to-end ligation of a linear DNA into a circle has been traditionally employed to measure the persistence length of dsDNA[21, 17]. Using this cyclization assay, Widom and coworker

found that $\sim$ 100-bp dsDNA loops much more frequently than predicted by the continuous worm-like chain (WLC) model[28]. While this result was later disputed because of the saturating concentration of ligase used in that study[38], it popularized the idea that kinks in the double helix can enhance flexibility at small length scales[32, 31]. Our recent study using smFRET has established that the WLC model is valid for random dsDNA as short as 60-bp in the condition with low to medium sodium concentration. In the condition that is comparable to cyclization experiment with $5 - 10$ mM $MgCl_2$, the critical size increases to 100 bp[48].

In this study, we used smFRET looping assay to investigate the sequence dependence of the looping propensity of dsDNA in the short length scales below 100bp. While the sequence dependence of the persistence length of dsDNA has been discovered[27], it might only be applicable to weak bending regime where the WLC approxmation is valid. Indeed, the extracted bending rigidity of dinucleotides from this study has no correlation with the frequency of dinucleotide bending found in crystallography[27, 135]. This poses a very interesting question on how sequence content affects the looping probability of short DNAs in the regime where kinks can occur[39, 48].

5.2 Methods

Single-stranded DNA oligos with 84 and 63 nucleotides were ordered from Eurofins Operon. I performed PCR with a pair of DNA primers to generate the corresponding DNA duplexes. The 84-bp and 63-bp dsDNAs were used as the DNA templates for the second round of PCR to generate DNAs with fluorophore and biotin labels. Strand exchange was carried out as described in Chapters 3 and 4 to construct the DNA sample with all modifications. Details about the sequences can be retrieved from Appendix B. The four primers used in this study are:

Primer 1: GGGCATCGTACACC,

Primer 2: /5Cy3/ ACGGATTCTG<u>A</u>GGGCATCGTACACC

Primer 3: /5BioTEG/AAAACAAAA/iCy5/<u>A</u>CAGAATCCGT<u>A</u>CCCACTCGTCCTTG,

Primer 4: CCCACTCGTCCTTG.

First, the universal adapter regions GGGCATCGTACACC and CCCACTCGTCCTTG have rich GC-contents near the 5'-end to minimize possible end fraying (melting of base pairs). Second, we added two additional nucleotide spacers (A) around the annealing region of the linker duplex to prevent any stacking between the ends of the linker duplex and the DNA loop. Last, the annealing region of the sticky ends is chosen as ACGGATTCTG for its strong stability that can stabilize the DNA sequence as short as 60-bp into the looped state[42].

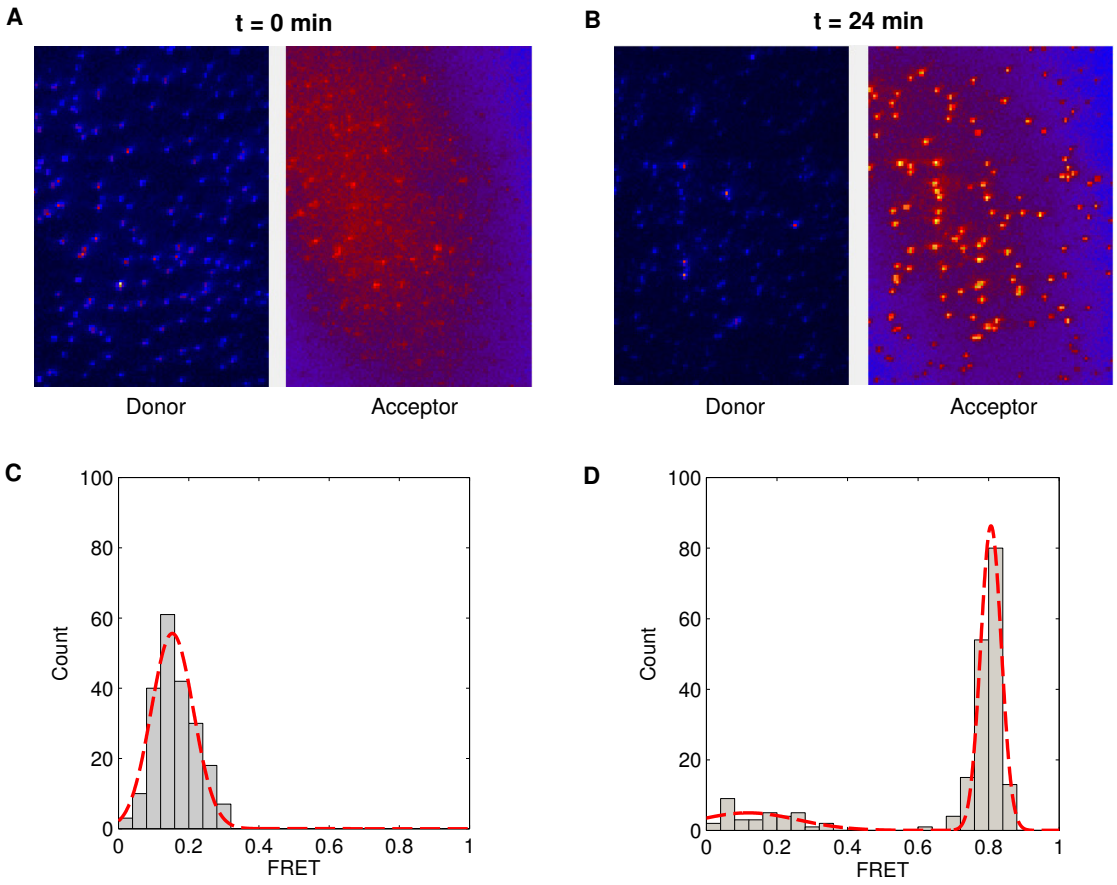


Figure 5.1: smFRET looping experiment for the 84bp TA601 DNA (see text). (A) and (B) Captures of the donor and acceptor channels taken at two different moments after the introduction of an imaging buffer with 1.5 M NaCl. The looping of DNA is indicated by an increase in the number of molecules with high acceptor intensity and a decrease in the number of molecules with high donor intensity. (C) and (D) Typical FRET histograms from $\sim$ 300 DNA molecules after being incubated in high-salt buffer for 0 minutes (A) and 24 minutes (B). A double Gaussian function was used to fit the histogram. Fraction of looped molecules is calculated by dividing the area under the Gaussian curve with high FRET by the summation of the two areas.

The experiment was implemented as previously described[42]. The first step is to introduce an imaging buffer with only a few millimolar sodium concentration into the sample chamber. Almost all molecules are in the unlooped state which are marked by the high donor intensity (Figure 5.1A). A new imaging buffer with 1.5 M NaCl was quickly flushed through the channel to initiate DNA looping (Figure 5.1B). We recorded the FRET values of all molecules from different fields of view at defined moments and built the FRET histogram. The FRET histogram shows two peaks, which correspond to the unlooped (low FRET) and looped (high FRET) fraction. Figures 5.1C and 5.1D show the two FRET histograms taken at two different moments after the introduction of high-salt buffer (1.5 M) which indicate the transition of the population from the unlooped to the looped conformation. We introduced fresh buffer every 2 hours to minimize the acidification of the buffer although it was reported that this effect is insignificant for the PCD/PCA compared to the GODCAT oxygen scavenging system[60].

We used a double Gaussian function to fit the histogram(Figures 5.1C and 5.1D) and extracted the fraction of looped molecule over time. This fraction was calculated as the ratio of the area under the curve with high FRET over the summation of the two areas for low and high FRET. The fraction of the looped molecule was reported as a function of time which can be fitted using a single exponential function to extract the apparent looping rate (Figure 5.2A). We used an exponential function with a base line to fit the data

$$N(t) = C(1 - e^{-Rt}), \tag{5.1}$$

where R is the apparent looping rate and C is the saturation value of the high FRET fraction. C never reaches 1 because there is a small fraction of molecules with inactive Cy5 that does not show high FRET even in the looped state[42]. We found that the fraction of molecule that carry active Cy3 and Cy5 is ~ 0.8 for most samples in our experiment. As a result, the looped fraction of several DNAs progressed to the same saturation level of 0.8 (Figure 5.2A). For DNA that loops very slowly, the saturation level becomes lower (Figure 5.2B) since the unlooping rate is not significantly larger than the looping rate.

5.3 *Preliminary Results*

5.3.1 Looping dynamics of some DNA sequences with varying nucleosome affinity

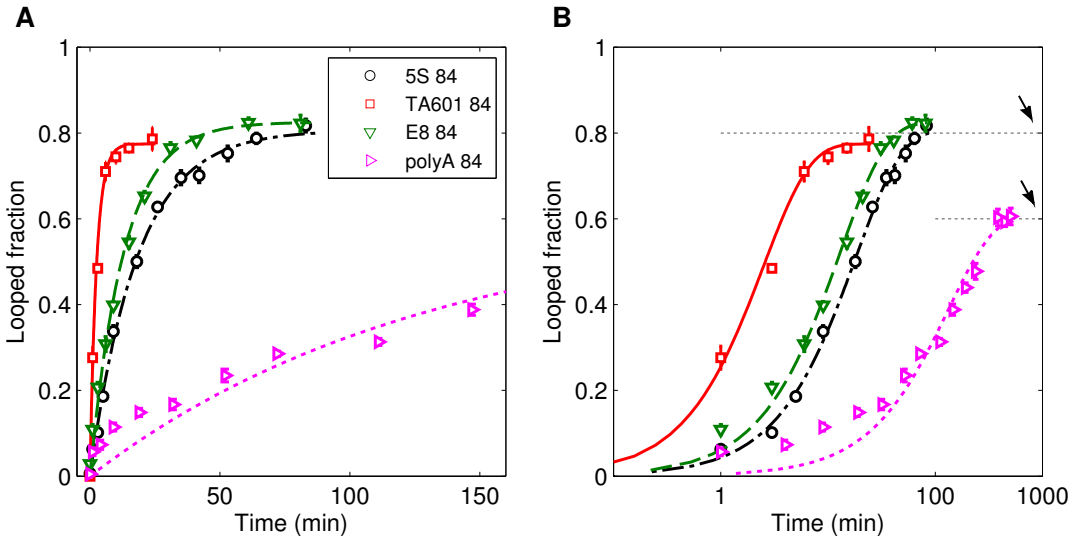


Figure 5.2: (A) The fraction of looped molecules for several 84bp DNAs which include TA601, E8, 5S, and a sequence with several poly(dA:dT) tracts. We used a single exponential function with a baseline (Equation 5.1) to fit the data and extract the apparent looping rate R. (B) Same as (A) but plotted in the log scale to illustrate large variation in the looping kinetics of the tested sequences. The arrows indicate the saturation levels of the curves. In both figures, the errorbars (SEM) represent the variation of the fraction recorded from five different fields of view.

DNA flexibility has been considered as an important factor to determine nucleosome propensity since it will cost less energy to compact a more flexible DNA into the DNA-protein complex[137, 28]. It was observed that the free energy of nucleosome formation has a strong correlation with the free energy of DNA cyclization for DNA at ∼100 bp[28]. We want to test this hypothesis for DNAs in the sub-100-bp long regime.

I changed the middle 56bp region of the 84-bp long DNA to some sequences with varying nucleosome affinity. The sequences that I tried were the Widom 601 DNA, which was known for its strong nucleosome preference[132, 28], the 5S rDNA sequence from sea urchin[28], the random E8 sequence[28], and a sequence with several poly(dA:dT) tracts[138]. If the E8 sequence is taken as the reference sequence, the free energy of nucleosome formation

of the Widom 601 DNA is $\Delta G_{nuc} = -3k_BT$, which indicates a strong preference to histone octamers[132]. For the 5S rDNA sequence, $\Delta G_{nuc} = -1.6k_BT$[28]. In contrast, the poly(dA:dT) sequence is a nucleosome-depleted sequence since the poly(dA:dT) tract is known for its disfavor of nucleosome formation[138]. The free energy to form nucleosome for this sequence is positive and relatively high, $\Delta G_{nuc} = +2k_BT$ relative to 5S[139].

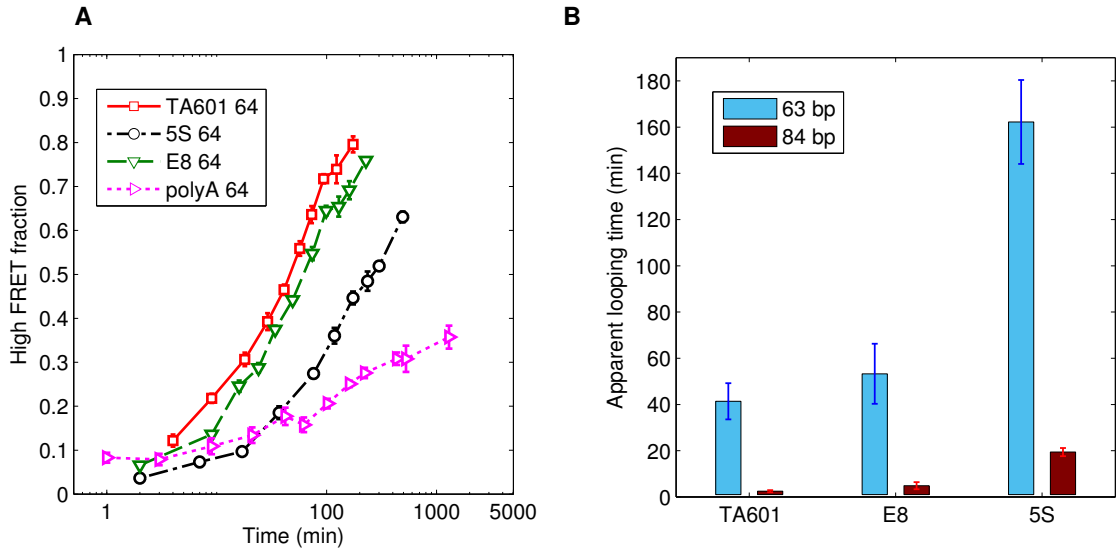


Figure 5.3: (A) The fraction of looped molecules for several 63bp DNA. (B) The apparent looping time ($1/R$) for several 63bp and 84bp DNAs. The errorbars represent the fitting error of the R value in Equation 5.1.

As shown in Figure 5.2B, these 84-bp sequences have significantly different looping dynamics. Among the four sequences, the Widom 601 DNA loops the fastest with a typical looping time of $2-3$ minutes, while the poly(dA:dT) loops significantly slower, taking $\sim$ 130 minutes to reach equilibrium. This result is somewhat consistent with previous cyclization data at larger length scales[28], except that we observed the 5S sequence loops slower than the generic E8 sequence. Our result on the looping of the poly(dA:dT) sequence also suggests a stiff poly(dA:dT) tract which is in agreement with previous studies on the flexibility of this oligonucleotide[140, 141, 142]. This observation is also in line with a recent smFRET looping study[42] but in contrast to the other looping experiment using the tethered particle assay[138]. It is possible that the difference in the boundary condition for looping in [138] can account for this discrepancy. We shortened the DNA fragments to

63bp and measured their looping dynamics. The shorter DNAs loop significantly slower (Figures 5.3 A and 5.3B) but they preserve the trend well (Figure 5.3B) despite the fact that the center pattern has been severely shortened from 56bp to 35bp.

5.3.2 Looping dynamics as a function of GC percentage

We are also interested in studying the dependence of the looping kinetics on the GC content of the sequence. We varied the 35-bp regions in a random manner to generate a set of 63-bp DNAs with the GC content of the 35-bp region from 40% to 80% (see Appendix B). We avoided sequence patterns which could produce significant curvature such as the A-tracts or the GGGCCC pattern. Gel electrophoresis was used to confirm that these sequences do not have detectable intrinsic bendedness (data not shown). The smFRET looping assay reveals that among the sequences we tested, there is no apparent dependence of the looping rates on GC content (Figure 5.4A).

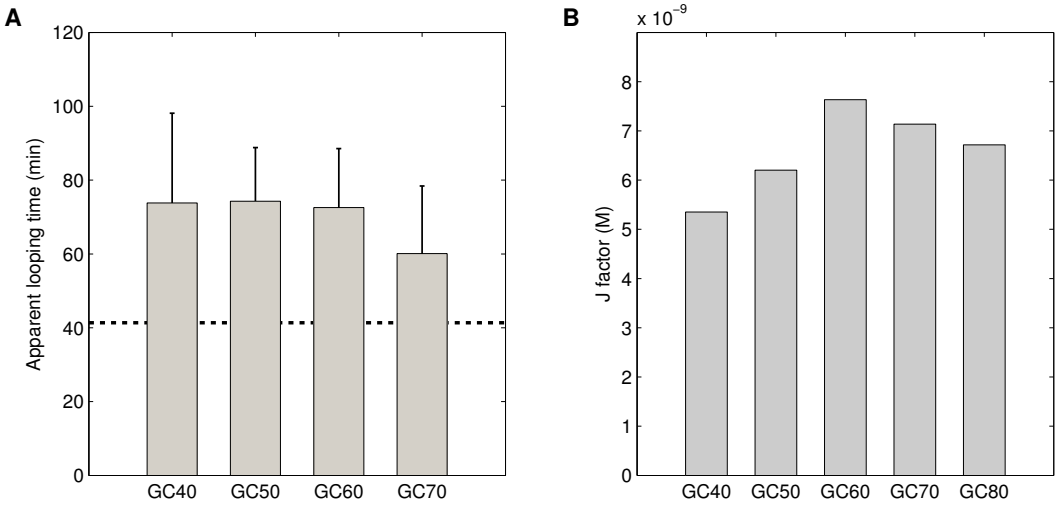


Figure 5.4: (A) The apparent looping time of several synthetic DNAs with varying GC contents. The GC percentage value indicates the percentage of G/C base in the centered 35-bp region of the 63-bp DNA. Sequence details are presented in Appendix B. (B) Predicted J factors of the sequences with varying GC content. The persistence length of these DNAs were calculated based on a dinucletotide assumption[27].

Since the sequence effect on the bending rigidity of DNA in the weak bending regime has been characterized[27], we tested this prediction on our sequence set. From the estimated average persistence length of these sequences, we calculated the J factor of the looped

conformation with an end-to-end distance of 4 nm using an isotropic model for the DNA with no intrinsic bendedness[24]. Note that the looping time is inversely proportional to the J factor (Equation 3.6). Although the applicability of this sequence-dependent dinucleotide model can be questionable in the strong bending condition, the numerical calculations also predict a minimal variation in the J factors of these sequences with varying GC content (Figure 5.4B).

5.3.3 Some DNAs with repeating patterns show enhanced flexibility

I tried to explore which sequence feature can dramatically enhance flexibility. We have already observed that the phasing of local sequence curvature in accordance with the helical repeat, which is 10.5 bp, can result in an increase in the sequence curvature and the looping rates[45]. The same idea may be applied to sequence bendability in the context of DNA looping. If the flexible dinucleotides are arranged in-phase with the helical repeat, one should expect an increase in the sequence bendability and therefore, an increase in the J factor. It is interesting to note that the Widom 601 sequence, a synthetic DNA fragment with the highest affinity to form nucleosomes *in vitro*, has a very distinctive distribution of the TA step with a periodicity of 10.1 bp, which is very close to the helical repeat[132]. We showed here that in the context of looping, a 35-bp fraction of the Widom 601 DNA, which essentially has only 3 units of the Widom 601 pattern (TTAAACGCACGTACGCGCTGTCTACCGCGTTTTAA), can also enhance the looping rate by a factor of 2 compared to DNA with random base distribution (Figure 5.5, also see Figure 5.3). Indeed, it can be expected that more flexible sequences might possess higher affinity to proteins in general.

Interestingly, we found that other repeating patterns can similarly enhance the looping rate. We tried the pattern TTAGGG, which constitutes the repeating unit of human telomeres[143]. With $3-5$ repeating blocks of this pattern, centered within the 35-bp regime in sequences G3, G4, and G5 (see Appendix B for sequence content), we also observed significantly enhanced flexibility, up to 3 fold increase, compared to random DNA (Figure 5.5). It is apparent from the sequence content that the TA steps in this pattern possess a periodicity of 6 bp which is about half of the helical repeat. It is therefore possible

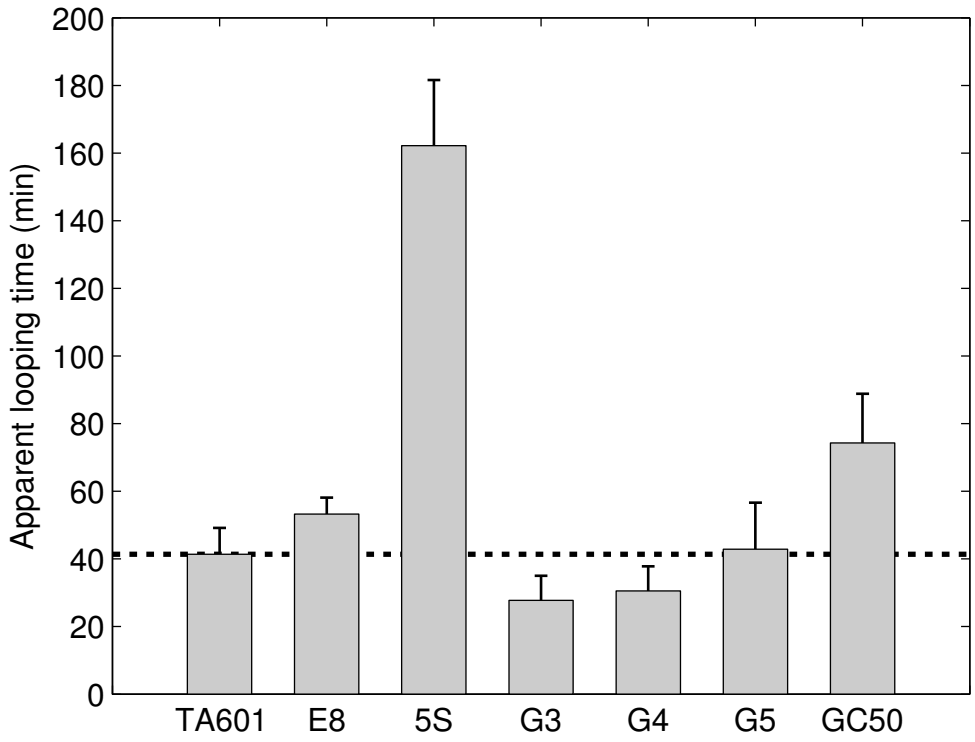


Figure 5.5: The apparent looping time of several DNA sequences with some particular patterns. G3, G4, and G5 are sequences with 3, 4, and 5 consecutive repeating blocks 'TTAGGG'. A ssDNA with 4 consecutive TTAGGG blocks can self-fold into an unique secondary structure which is called the G-quadruplex[143]. The dotted line indicates the apparent looping time of the sequence with the Widom 601 pattern.

that the step phasing can be optimized to get an even higher level of sequence bendability. Finally, we note that the 5S DNA, which is another nucleosome DNA, seems to have higher bending rigidity in the context of looping compared to random DNA (Figure 5.5). This observation highlights the need in taking into account specific interactions between DNA and histone octamers, along with the mechanical property of the DNA sequence underlying the nucleosome, to adequately reflect the sequence dependence of nucleosome affinity.

5.4 Final remarks

In this project, we used the smFRET looping assay to study the sequence dependence of the looping probability of DNA as short as 63 bp. At this length scales, the effect of local disruptions such as kinks on the helical structure cannot be ignored. Therefore, the experiment of DNA looping in short length scales can deepen our understanding on the

mechanical property of the dsDNA in the extreme bending condition. As we have already seen, a variation in the sequence content of only 35 bp, which is about $3-4$ helical repeats, could have strong effect on the bendability of the sequence. Understanding the energetics of DNA at sharp bending angles can give more insights into the structures and dynamics of DNA-protein complexes which involve small DNA loops. In the case of nucleosomes, a strong link between DNA flexibility and nucleosome stability has long been acknowledged from several experimental evidences[28, 144].

Some of the future directions are to test a large set of sequences and to perform rigorous statistical analysis[132] to identify sequence features that correlate with high looping probability. It is tempting to combine the looping assay with the unlooping assay to study the energetics of DNA bending for short DNA with varying sequence content. A full set of DNA flexibility parameters at the dinucleotide step level that takes into account both the sequence dependence of DNA kinkability in the strong bending regime and the sequence dependence of DNA flexibility in the weak bending regime[27] might fully explain DNA bending at all length scales. It is also important to come up with a new strategy for a large-scale study to categorize DNA based on their looping propensity, especially for small DNA minicircles under 100 bp. A very recent study[145] has revealed a set of rules for DNA looping that bear some similarities to nucleosome positioning sequences[132].

APPENDIX A

SUPPLEMENTARY METHODS FOR CHAPTER IV

A.1 J factor calculation

The J factor is calculated by Weighted Histogram Analysis Method (WHAM)[146]. A number of umbrella sampling simulations were carried out, each having its own restraint energy $U_j(r_k)$ where j is the simulation index, and k is the bin index. In the j-th simulation, one obtains the number of counts $n_{j,k}$ in the k-th bin with the total counts equal to $N_j = \sum_k n_{j,k}$. Using the bias factor in each bin $c_{j,k} = \exp(-U_j(r_k))$, we can obtain the radial probability density of the unrestrained chain (p_k^0)

$$p_k^0 = \frac{\sum_j n_{j,k}}{\sum_j f_{k,j} N_j}, \tag{A.1a}$$

$$f_{k,j} = \frac{c_{j,k}}{\sum_k c_{j,k} p_k^0}. \tag{A.1b}$$

These equations were solved iteratively by updating the equations until p_k^0 converges. We adjusted the spring constant and restraint coordinates so that there is significant overlap between adjacent histograms. Typically, each individual histogram was built from 10^6 chains. The J factor was obtained by normalizing p_k^0, dividing it by $4\pi r^2$, and converting it to molar units. An example of umbrella samplings and the derivation of the radial probability distribution and J factor for a 70-bp worm-like chain DNA are illustrated in Figure A.1.

A.2 Minicircle simulations

The MC simulation for a DNA minicircle was implemented as previously described[121]. We applied the KWLC bending energy to each link and calculated the total bending energy of the minicircle. Random conformations generated by crankshaft rotations were selected based on the Metropolis criteria. In one course of simulation, 15×10^6 conformations were typically collected. To enhance the sampling efficiency, we randomly picked angles for the

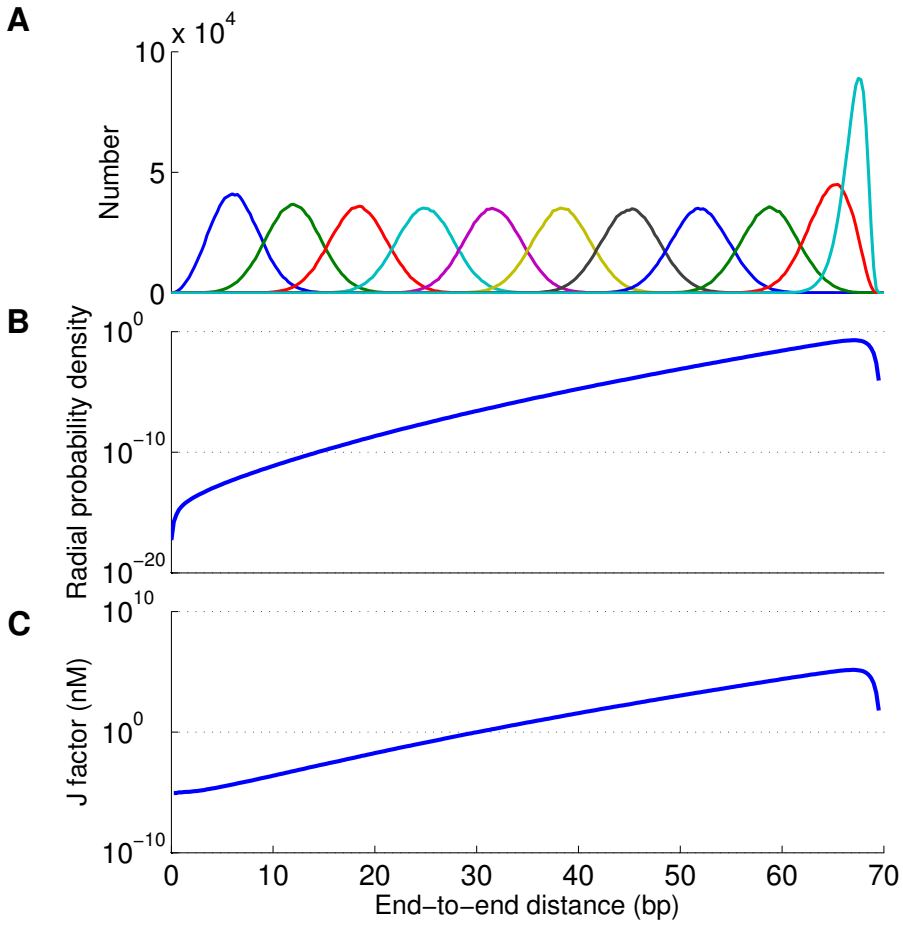


Figure A.1: J factor calculation by the weighted histogram analysis method. (a) Umbrella sampling was performed at every 10-bp step. The spring constant was chosen so that neighboring histograms overlap significantly. Each histogram was obtained from 10^6 MC conformations after 100,000 thermalization steps. (b) The radial probability distribution was obtained by iterating through Equation A.1. (c) The J factor in nanomolar units can be obtained by dividing the amplitude of the radial probability distribution by $4\pi r^2 \Delta r$ and multiplying by 4.24×10^{10}.

crankshaft rotation from two uniform distributions across two intervals, $[-90°, 90°]$ and $[-10°, 10°]$. For each accepted conformation, all the dinucleotide angles were recorded to determine if the minicircle has kinks. A kink was assigned if the bending angle exceeds the critical kink angle defined as the intercept of the two energy terms in Equation 4.10. For each loop size, we calculated the kinking probability, which is the fraction of accepted conformations with at least one kink.

APPENDIX B

SUPPLEMENTARY METHODS FOR CHAPTER V

The universal adapter regions are marked in bold.

B.1 63-bp DNA sequences with varying GC contents

40% GC (GC40): **CCCACTCGTCCTTG**AACGTTTGAACTTCTAACAACACGCAG ATAGTACC**GGTGTACGATGCCC**

50% GC (GC50): **CCCACTCGTCCTTG**AGCAACCAGAACCGCTTGGGACTC TATCAGAGAGT**GGTGTACGATGCCC**

60% GC (GC60): **CCCACTCGTCCTTG**CGTGTGGAGTGATAGCACCCAGTG GCCTCCCTGAG**GGTGTACGATGCCC**

70% GC (GC70): **CCCACTCGTCCTTG**GGCGCTCCACACCGGAGCACGCTC TACCGACCTGG**GGTGTACGATGCCC**

80% GC (GC80): **CCCACTCGTCCTTG**CGCGGTGGCGGTCCGCCGGAACTC CTCCGCGCGAG**GGTGTACGATGCCC**

B.2 63-bp DNA sequences with varying nucleosome affinity

E8 (reference [28]): **CCCACTCGTCCTTG**TCCACGGTGCTGATCCCCTGTGCTGT TGGCCGTGT**GGTGTACGATGCCC**

5S (reference [138]): **CCCACTCGTCCTTG**AACTTTCATCAAGCAAGAGCCTA CGACCATACCAT**GGTGTACGATGCCC**

TA601 (reference [132]): **CCCACTCGTCCTTG**TTAAACGCACGTACGCGCTG TCTACCGCGTTTTAA**GGTGTACGATGCCC**

polyA (from reference [138]): **CCCACTCGTCCTTG**AACTGAAAAAGAGAAAA ATAAGAAAATCTTCTAGA**GGTGTACGATGCCC**

B.3 84-bp DNA sequences with varying nucleosome affinity

E8 (reference [28]): **CCCACTCGTCCTTG**TTTATTTATCGCCTCCACGGTGCTGA TCCCCTGTGCTGTTGGCCGTGTTATCTCGA**GGTGTACGATGCCC**

5S (reference [138]): **CCCACTCGTCCTTG**TTAAATAGCTTAACTTTCATCAA GCAAGAGCCTACGACCATACCATGCTGAATATA**GGTGTACGATGCCC**

TA601 (reference [132]): **CCCACTCGTCCTTG**CTAGCACCGCTTAAACGCAC GTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGG**GGTGTACGATGCCC**

polyA (from reference [138]): **CCCACTCGTCCTTG**GATGAAAAAAAAACTG AAAAAGAGAAAAATAAGAAAATCTTCTAGAACGTTCCGAA**GGTGTACGAT-GCCC**

B.4 Sequences with TTAGGG repeats

G4: **CCCACTCGTCCTTG**ATAGATTAGGGTTAGGGTTAGGGTTAGGGTCCTC A**GGTGTACGATGCCC**

G5: **CCCACTCGTCCTTG**ATTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG TCA**GGTGTACGATGCCC**

G3: **CCCACTCGTCCTTG**ATAGACTCCTTAGGGTTAGGGTTAGGGTATCC TCA**GGTGTACGATGCCC**

APPENDIX C

PERMISSIONS TO REUSE PREVIOUSLY PUBLISHED MATERIALS

1. Licensed content title: Measuring shape-dependent looping probability of DNA

   Citation: Tung T Le and Harold D Kim. Measuring shape-dependent looping probability of DNA. *Biophysical journal*, 104(9):2068-2076, 2013.

   Licensee: Tung Le

   License Date: May 04, 2015

   License Numbers: 3621901323540

   Licensed content publisher: Elsevier

   Type Of Use: Thesis/Dissertation

2. Licensed content title: Probing the elastic limit of DNA bending

   Citation: Tung T Le and Harold D Kim. Probing the elastic limit of DNA bending. *Nucleic acids research*, 42(16):10786-10794, 2014.

   Licensee: Tung Le

   License Date: April 30, 2015

   License Numbers: 3618761072772, 3618761077107

   Licensed content publisher: Oxford University Press

   Type Of Use: Thesis/Dissertation

REFERENCES

[1] Nadrian C Seeman. Nanomaterials based on dna. *Annual review of biochemistry*, 79: 65, 2010.

[2] JD Watson and FHC Crick. A structure for deoxyribose nucleic acid. *Nature*, 421 (6921):397–3988, 1953.

[3] TS Baker, NH Olson, and SD Fuller. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiology and Molecular Biology Reviews*, 63(4):862–922, 1999.

[4] Laura M Bond, Justin P Peters, Nicole A Becker, Jason D Kahn, and L James Maher. Gene repression by minimal lac loops in vivo. *Nucleic acids research*, 38(22):8072–8082, 2010.

[5] Andrew J Andrews and Karolin Luger. Nucleosome structure (s) and stability: variations on a theme. *Annual review of biophysics*, 40:99–117, 2011.

[6] Karolin Luger, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature*, 389(6648):251–260, 1997.

[7] Hernan G Garcia, Paul Grayson, Lin Han, Mandar Inamdar, Jané Kondev, Philip C Nelson, Rob Phillips, Jonathan Widom, and Paul A Wiggins. Biological consequences of tightly bent dna: the other life of a macromolecular celebrity. *Biopolymers*, 85(2): 115–130, 2007.

[8] Wilma K Olson and Victor B Zhurkin. Working the kinks out of nucleosomal dna. *Current opinion in structural biology*, 21(3):348–357, 2011.

[9] O Kratky and G Porod. Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12):1106–1122, 1949.

[10] Michael Rubinstein and R Colby. *Polymers Physics*. Oxford, 2003.

[11] Masao Doi. *The theory of polymer dynamics*. Number 73. oxford university press, 1988.

[12] John G Elias and Don Eden. Transient electric birefringence study of the persistence length and electrical polarizability of restriction fragments of dna. *Macromolecules*, 14(2):410–419, 1981.

[13] Jan Bednar, Patrick Furrer, Vsevolod Katritch, Alicja Stasiak, Jacques Dubochet, and Andrzej Stasiak. Determination of dna persistence length by cryo-electron microscopy. separation of the static and dynamic contributions to the apparent persistence length of dna. *Journal of molecular biology*, 254(4):579–594, 1995.

[14] Steven B Smith, Yujia Cui, and Carlos Bustamante. Overstretching b-dna: the elastic response of individual double-stranded and single-stranded dna molecules. *Science*, 271(5250):795–799, 1996.

[15] C Bouchiat, MD Wang, J-F Allemand, T Strick, SM Block, and V Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical journal*, 76(1):409–413, 1999.

[16] David Shore, Jorg Langowski, and Robert L Baldwin. Dna flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences*, 78(8):4833–4837, 1981.

[17] William H Taylor and Paul J Hagerman. Application of the method of phage t4 dna ligase-catalyzed ring-closure to the study of dna structure: Ii. nacl-dependence of dna flexibility and helical repeat. *Journal of molecular biology*, 212(2):363–376, 1990.

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

[19] Carlos Bustamante, John F Marko, Eric D Siggia, and Steven B Smith. Entropic elasticity of lambda-phage dna. *Science*, 265(5178):1599–1600, 1994.

[20] Michelle D Wang, Hong Yin, Robert Landick, Jeff Gelles, and Steven M Block. Stretching dna with optical tweezers. *Biophysical journal*, 72(3):1335, 1997.

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

[22] Maria Vologodskaia and Alexander Vologodskii. Contribution of the intrinsic curvature to measured DNA persistence length. *Journal of Molecular Biology*, 317(2): 205–213, March 2002. ISSN 00222836. URL `http://linkinghub.elsevier.com/retrieve/pii/S0022283601953665`.

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

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

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

[26] David S Horowitz and James C Wang. Torsional rigidity of dna and length dependence of the free energy of dna supercoiling. *Journal of molecular biology*, 173(1):75–91, 1984.

[27] Stephanie Geggier and Alexander Vologodskii. Sequence dependence of dna bending rigidity. *Proceedings of the National Academy of Sciences*, 107(35):15421–15426, 2010.

[28] Timothy E Cloutier and Jonathan Widom. Spontaneous sharp bending of double-stranded dna. *Molecular cell*, 14(3):355–362, 2004.

[29] TE Cloutier and Jonathan Widom. Dna twisting flexibility and the formation of sharply looped protein–dna complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10):3645–3650, 2005.

[30] FHC Crick and A Klug. Kinky helix. *Nature*, 255(5509):530–533, 1975.

[31] Paul A Wiggins, Rob Phillips, and Philip C Nelson. Exact theory of kinkable elastic polymers. *Physical Review E*, 71(2):021909, 2005.

[32] Jie Yan and John F Marko. Localized single-stranded bubble mechanism for cyclization of short double helix dna. *Physical review letters*, 93(10):108108, 2004.

[33] David A Sivak and Phillip L Geissler. Consequences of local inter-strand dehybridization for large-amplitude bending fluctuations of double-stranded dna. *The Journal of chemical physics*, 136(4):045102, 2012.

[34] Paul A Wiggins, Thijn Van Der Heijden, Fernando Moreno-Herrero, Andrew Spakowitz, Rob Phillips, Jonathan Widom, Cees Dekker, and Philip C Nelson. High flexibility of dna on short length scales probed by atomic force microscopy. *Nature nanotechnology*, 1(2):137–141, 2006.

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

[36] Alexander Vologodskii and Maxim D Frank-Kamenetskii. Strong bending of the dna double helix. *Nucleic acids research*, 41(14):6785–6792, 2013.

[37] Filip Lankaš, Richard Lavery, and John H Maddocks. Kinking occurs during molecular dynamics simulations of small dna minicircles. *Structure*, 14(10):1527–1534, 2006.

[38] Quan Du, Chaim Smith, Nahum Shiffeldrim, Maria Vologodskaia, and Alexander Vologodskii. Cyclization of short dna fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15):5397–5402, 2005.

[39] Quan Du, Alexander Kotlyar, and Alexander Vologodskii. Kinking the double helix by bending deformation. *Nucleic acids research*, 36(4):1120–1128, 2008.

[40] Davide Demurtas, Arnaud Amzallag, Eric J Rawdon, John H Maddocks, Jacques Dubochet, and Andrzej Stasiak. Bending modes of dna directly addressed by cryo-electron microscopy of dna minicircles. *Nucleic acids research*, page gkp137, 2009.

[41] JS Mitchell, CA Laughton, and Sarah A Harris. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled dna. *Nucleic acids research*, 39(9):3928–3938, 2011.

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

[43] Philip C Nelson. Spare the (elastic) rod. *Science*, 337(6098):1045–1046, 2012.

[44] Alexander Vologodskii, Quan Du, Maxim D Frank-Kamenetskii, et al. Bending of short dna helices. *Artif. DNA: PNA XNA*, 4:1–3, 2013.

[45] Tung T Le and Harold D Kim. Measuring shape-dependent looping probability of dna. *Biophysical journal*, 104(9):2068–2076, 2013.

[46] Tung T Le and Harold D Kim. Studying dna looping by single-molecule fret. *JoVE (Journal of Visualized Experiments)*, (88):e51667, 2014.

[47] Luke Czapla, David Swigon, and Wilma K Olson. Sequence-dependent effects in the cyclization of short dna. *Journal of Chemical Theory and Computation*, 2(3):685–695, 2006.

[48] Tung T Le and Harold D Kim. Probing the elastic limit of dna bending. *Nucleic acids research*, 42(16):10786–10794, 2014.

[49] Alexey K. Mazur and Mounir Maaloum. Dna flexibility on short length scales probed by atomic force microscopy. *Physical Review Letters*, 112(6):068104, 2014.

[50] Alexey K Mazur and Mounir Maaloum. Atomic force microscopy study of dna flexibility on short length scales: smooth bending versus kinking. *Nucleic acids research*, page gku1192, 2014.

[51] Theodor Förster. *Delocalized excitation and excitation transfer*. Florida State University, 1965.

[52] Robert M Clegg. Förster resonance energy transferfret what is it, why do it, and how it's done. *Laboratory techniques in biochemistry and molecular biology*, 33:1–57, 2009.

[53] David L Andrews. A unified theory of radiative and radiationless molecular energy transfer. *Chemical Physics*, 135(2):195–201, 1989.

[54] Paul R Selvin. Fluorescence resonance energy transfer. *Methods in enzymology*, 246: 300–334, 1995.

[55] Asif Iqbal, Sinan Arslan, Burak Okumus, Timothy J Wilson, Gerard Giraud, David G Norman, Taekjip Ha, and David MJ Lilley. Orientation dependence in fluorescent energy transfer between cy3 and cy5 terminally attached to double-stranded nucleic acids. *Proceedings of the National Academy of Sciences*, 105(32):11176–11181, 2008.

[56] Taekjip Ha, Th Enderle, DF Ogletree, DS Chemla, PR Selvin, and S Weiss. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences*, 93(13):6264–6268, 1996.

[57] Taekjip Ha. Single-molecule fluorescence resonance energy transfer. *Methods*, 25(1): 78–86, 2001.

[58] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule fret. *Nature methods*, 5(6):507–516, 2008.

[59] Colin Echeverría Aitken, R Andrew Marshall, and Joseph D Puglisi. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophysical journal*, 94(5):1826–1835, 2008.

[60] Xinghua Shi, John Lim, and Taekjip Ha. Acidification of the oxygen scavenging system in single-molecule fluorescence studies: in situ sensing with a ratiometric dual-emission probe. *Analytical chemistry*, 82(14):6132–6138, 2010.

[61] Ivan Rasnik, Sean A McKinney, and Taekjip Ha. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature methods*, 3(11):891–893, 2006.

[62] Thorben Cordes, Jan Vogelsang, and Philip Tinnefeld. On the mechanism of trolox as antiblinking and antibleaching reagent. *Journal of the American Chemical Society*, 131(14):5018–5019, 2009.

[63] Larry J Friedman, Johnson Chung, and Jeff Gelles. Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. *Biophysical journal*, 91(3):1023–1031, 2006.

[64] Taekjip Ha, Ivan Rasnik, Wei Cheng, Hazen P Babcock, George H Gauss, Timothy M Lohman, and Steven Chu. Initiation and re-initiation of dna unwinding by the escherichia coli rep helicase. *Nature*, 419(6907):638–641, 2002.

[65] Stanley D Chandradoss, Anna C Haagsma, Young Kwang Lee, Jae-Ho Hwang, Jwa-Min Nam, and Chirlmin Joo. Surface passivation for single-molecule protein studies. *JoVE (Journal of Visualized Experiments)*, (86):e50549, 2014.

[66] Alexander Gust, Adrian Zander, Andreas Gietl, Phil Holzmeister, Sarah Schulz, Birka Lalkens, Philip Tinnefeld, and Dina Grohmann. A starting point for fluorescence-based single-molecule measurements in biomolecular research. *Molecules*, 19(10): 15824–15865, 2014.

[67] Boyang Hua, Kyu Young Han, Ruobo Zhou, Hajin Kim, Xinghua Shi, Sanjaya C Abeysirigunawardena, Ankur Jain, Digvijay Singh, Vasudha Aggarwal, Sarah A Woodson, et al. An improved surface passivation method for single-molecule studies. *Nature methods*, 11(12):1233–1236, 2014.

[68] Harold D Kim. *Single molecule studies of dynamic biological processes*. PhD thesis, Stanford University, 2004.

[69] Chongli Yuan, Elizabeth Rhoades, Xiong Wen Lou, and Lynden A Archer. Spontaneous sharp bending of dna: role of melting bubbles. *Nucleic acids research*, 34(16): 4554–4560, 2006.

[70] Chongli Yuan, Huimin Chen, Xiong Wen Lou, and Lynden A Archer. Dna bending stiffness on small length scales. *Physical review letters*, 100(1):018102, 2008.

[71] Lin Han, Hernan G Garcia, Seth Blumberg, Kevin B Towles, John F Beausang, Philip C Nelson, and Rob Phillips. Concentration and length dependence of dna looping in transcriptional regulation. *PloS one*, 4(5):e5621, 2009.

[72] Chongli Yuan, Xiong Wen Lou, Elizabeth Rhoades, Huimin Chen, and Lynden A Archer. T4 dna ligase is more than an effective trap of cyclized dsdna. *Nucleic acids research*, 35(16):5294–5302, 2007.

[73] Carlo Manzo, Chiara Zurla, David D Dunlap, and Laura Finzi. The effect of nonspecific binding of lambda repressor on dna looping dynamics. *Biophysical journal*, 103 (8):1753–1761, 2012.

[74] Noam Kaplan, Irene K Moore, Yvonne Fondufe-Mittendorf, Andrea J Gossett, Desiree Tillo, Yair Field, Emily M LeProust, Timothy R Hughes, Jason D Lieb, Jonathan Widom, et al. The dna-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236):362–366, 2008.

[75] Hyeon-Sook Koo and Donald M Crothers. Calibration of dna curvature and a unified description of sequence-directed bending. *Proceedings of the National Academy of Sciences*, 85(6):1763–1767, 1988.

[76] Gianni Prosseda, Alessia Mazzola, Maria Letizia Di Martino, Denis Tielker, Gioacchino Micheli, and Bianca Colonna. A temperature-induced narrow dna curvature range sustains the maximum activity of a bacterial promoter in vitro. *Biochemistry*, 49(13):2778–2785, 2010.

[77] Christoph G Baumann, Steven B Smith, Victor A Bloomfield, and Carlos Bustamante. Ionic effects on the elasticity of single dna molecules. *Proceedings of the National Academy of Sciences*, 94(12):6185–6190, 1997.

[78] Attila Szabo, Klaus Schulten, and Zan Schulten. First passage time approach to diffusion controlled reactions. *The Journal of Chemical Physics*, 72(8):4350–4357, 1980.

[79] Jan Wilhelm and Erwin Frey. Radial distribution function of semiflexible polymers. *Physical Review Letters*, 77(12):2581, 1996.

[80] Axel E Nkodo, Jean M Garnier, Bernard Tinland, Hongji Ren, Claude Desruisseaux, Laurette C McCormick, Guy Drouin, and Gary W Slater. Diffusion coefficient of dna molecules during free solution electrophoresis. *Electrophoresis*, 22(12):2424–2432, 2001.

[81] Alexei A Podtelezhnikov and Alexander V Vologodskii. Dynamics of small loops in dna molecules. *Macromolecules*, 33(7):2767–2771, 2000.

[82] Suckjoon Jun, John Bechhoefer, and B-Y Ha. Diffusion-limited loop formation of semiflexible polymers: Kramers theory and the intertwined time scales of chain relaxation and closing. *EPL (Europhysics Letters)*, 64(3):420, 2003.

[83] Changbong Hyeon and D Thirumalai. Kinetics of interior loop formation in semiflexible chains. *The Journal of chemical physics*, 124(10):104905, 2006.

[84] C Anselmi, P De Santis, R Paparcone, M Savino, and A Scipioni. From the sequence to the superstructural properties of dnas. *Biophysical chemistry*, 95(1):23–47, 2002.

[85] Robert S Manning, John H Maddocks, and Jason D Kahn. A continuum rod model of sequence-dependent dna structure. *The Journal of chemical physics*, 105(13):5626–5646, 1996.

[86] RE Dickerson. Definitions and nomenclature of nucleic acid structure components. *Nucleic acids research*, 17(5):1797–1803, 1989.

[87] David S Goodsell and Richard E Dickerson. Bending and curvature calculations in b-dna. *Nucleic acids research*, 22(24):5497, 1994.

[88] A Bolshoy, P McNamara, RE Harrington, and EN Trifonov. Curved dna without aa: experimental estimation of all 16 dna wedge angles. *Proceedings of the National Academy of Sciences*, 88(6):2312–2316, 1991.

[89] Wilma K Olson, Nancy L Marky, Robert L Jernigan, and Victor B Zhurkin. Influence of fluctuations on dna curvature: a comparison of flexible and static wedge models of intrinsically bent dna. *Journal of molecular biology*, 232(2):530–554, 1993.

[90] Wilma K Olson, Andrey A Gorin, Xiang-Jun Lu, Lynette M Hock, and Victor B Zhurkin. Dna sequence-dependent deformability deduced from protein–dna crystal complexes. *Proceedings of the National Academy of Sciences*, 95(19):11163–11168, 1998.

[91] Yongli Zhang and Donald M Crothers. Statistical mechanics of sequence-dependent circular dna and its application for dna cyclization. *Biophysical journal*, 84(1):136–153, 2003.

[92] Paul J Hagerman. Analysis of the ring-closure probabilities of isotropic wormlike chains: application to duplex dna. *Biopolymers*, 24(10):1881–1897, 1985.

[93] Pasquale De Santis, Antonio Palleschi, Maria Savino, and Anita Scipioni. Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of dna curvature. *Biochemistry*, 29(39):9269–9273, 1990.

[94] CR Calladine, Horace R Drew, and Maxine J McCall. The intrinsic curvature of dna in solution. *Journal of molecular biology*, 201(1):127–137, 1988.

[95] Yongxing Liu and David L Beveridge. A refined prediction method for gel retardation of dna oligonucleotides from dinucleotide step parameters: reconciliation of dna bending models with crystal structure data. *Journal of Biomolecular Structure and Dynamics*, 18(4):505–526, 2001.

[96] Alberto Perez, Filip Lankas, F Javier Luque, and Modesto Orozco. Towards a molecular dynamics consensus view of b-dna flexibility. *Nucleic acids research*, 36(7):2379–2394, 2008.

[97] Niels Laurens, David A Rusling, Christian Pernstich, Ineke Brouwer, Stephen E Halford, and Gijs JL Wuite. Dna looping by foki: the impact of twisting and bending rigidity on protein-induced looping dynamics. *Nucleic acids research*, 40(11):4988–4997, 2012.

[98] James T Waters and Harold D Kim. Equilibrium statistics of a surface-pinned semiflexible polymer. *Macromolecules*, 46(16):6659–6666, 2013.

[99] Justin Spiriti, Hiqmet Kamberaj, Adam MR de Graff, MF Thorpe, and Arjan van der Vaart. Dna bending through large angles is aided by ionic screening. *Journal of Chemical Theory and Computation*, 8(6):2145–2156, 2012.

[100] Richard Lavery, Krystyna Zakrzewska, David Beveridge, Thomas C Bishop, David A Case, Thomas Cheatham, Surjit Dixit, B Jayaram, Filip Lankas, Charles Laughton, et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in b-dna. *Nucleic acids research*, 38(1):299–313, 2010.

[101] Ivan Brukner, Slavoljub Susic, Mensur Dlakic, Ana Savic, and Sándor Pongor. Physiological concentration of magnesium ions induces a strong macroscopic curvature in gggccc-containing dna. *Journal of molecular biology*, 236(1):26–32, 1994.

[102] Ioulia Rouzina and Victor A. Bloomfield. DNA bending by small, mobile multivalent cations. *Biophysical Journal*, 74(6):3152–3164, 1998.

[103] Loren Dean Williams and L James Maher III. Electrostatic mechanisms of dna deformation. *Annual review of biophysics and biomolecular structure*, 29(1):497–521, 2000.

[104] Nancy C Stellwagen and Yongjun Lu. Effect of magnesium ions and temperature on the sequence-dependent curvature of dna restriction fragments. *Journal of Physics: Condensed Matter*, 22(49):494110, 2010.

[105] Justin P Peters and L James Maher. Dna curvature and flexibility in vitro and in vivo. *Quarterly reviews of biophysics*, 43(1):23–63, 2010.

[106] George I Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342): 618–627, 1978.

[107] John A Schellman. Flexibility of dna. *Biopolymers*, 13(1):217–226, 1974.

[108] Alexander Vologodskii. Simulation of equilibrium and dynamic properties of large dna molecules. In *Computational Studies of RNA and DNA*, pages 579–604. Springer, 2006.

[109] Quan Du, Maria Vologodskaia, Heiko Kuhn, Maxim Frank-Kamenetskii, and Alexander Vologodskii. Gapped dna and cyclization of short dna fragments. *Biophysical Journal*, 88(6):4137–4145, 2005.

[110] Johannes Kästner and Walter Thiel. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method:umbrella integration. *The Journal of chemical physics*, 123(14):144104, 2005.

[111] Wonmuk Hwang. Calculation of conformation-dependent biomolecular forces. *The Journal of chemical physics*, 127(17):175104, 2007.

[112] Torsten Strunz, Krisztina Oroszlan, Rolf Schäfer, and Hans-Joachim Güntherodt. Dynamic force spectroscopy of single dna molecules. *Proceedings of the National Academy of Sciences*, 96(20):11277–11282, 1999.

[113] Jay R Wenner, Mark C Williams, Ioulia Rouzina, and Victor A Bloomfield. Salt dependence of the elasticity and overstretching transition of single dna molecules. *Biophysical journal*, 82(6):3160–3169, 2002.

[114] Michael Hinczewski, J Christof M Gebhardt, Matthias Rief, and D Thirumalai. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proceedings of the National Academy of Sciences*, 110(12):4500–4505, 2013.

[115] Matthew J Comstock, Taekjip Ha, and Yann R Chemla. Ultrahigh-resolution optical trap with single-fluorophore sensitivity. *Nature methods*, 8(4):335–340, 2011.

[116] Neeraj J Agrawal, Ravi Radhakrishnan, and Prashant K Purohit. Geometry of mediating protein affects the probability of loop formation in dna. *Biophysical journal*, 94(8):3150–3158, 2008.

[117] Ari Zeida, Matías Rodrigo Machado, Pablo Daniel Dans, and Sergio Pantano. Breathing, bubbling, and bending: Dna flexibility from multimicrosecond simulations. *Physical Review E*, 86(2):021903, 2012.

[118] Maxim D Frank-Kamenetskii and Shikha Prakash. Fluctuations in the dna double helix: A critical review. *Physics of Life Reviews*, 2014.

[119] O-chul Lee and Wokyung Sung. Enhanced bubble formation in looped short double-stranded dna. *Physical Review E*, 85(2):021902, 2012.

[120] Alexander P Fields, Elisabeth A Meyer, and Adam E Cohen. Euler buckling and nonlinear kinking of double-stranded dna. *Nucleic acids research*, 41(21):9881–9890, 2013.

[121] Xiaozhong Zheng and Alexander Vologodskii. Theoretical analysis of disruptions in dna minicircles. *Biophysical journal*, 96(4):1341–1349, 2009.

[122] Robert Schöpflin, Hergen Brutzer, Oliver Müller, Ralf Seidel, and Gero Wedemann. Probing the elasticity of dna on short length scales by modeling supercoiling under tension. *Biophysical journal*, 103(2):323–330, 2012.

[123] Xuesong Shi, Daniel Herschlag, and Pehr AB Harbury. Structural ensemble and microscopic elasticity of freely diffusing dna by direct measurement of fluctuations. *Proceedings of the National Academy of Sciences*, 110(16):E1444–E1451, 2013.

[124] S Mantelli, P Muller, S Harlepp, and M Maaloum. Conformational analysis and estimation of the persistence length of dna using atomic force microscopy in solution. *Soft Matter*, 7(7):3412–3416, 2011.

[125] Hari Shroff, David Sivak, Jake J Siegel, AL McEvoy, Merek Siu, Andrew Spakowitz, Phillip L Geissler, and Jan Liphardt. Optical measurement of mechanical forces inside short dna loops. *Biophysical journal*, 94(6):2179–2186, 2008.

[126] Juan Wang, Hao Qu, and Giovanni Zocchi. Critical bending torque of dna is a materials parameter independent of local base sequence. *Physical Review E*, 88(3): 032712, 2013.

[127] Jeremy Curuksu, Jiri Sponer, and Martin Zacharias. Elbow flexibility of the kt38 rna kink-turn motif investigated by free-energy molecular dynamics simulations. *Biophysical journal*, 97(7):2004–2013, 2009.

[128] Alexander V Lukashin, Alexander V Vologodskii, Maxim D Frank-Kamenetskii, and Yuril L Lyubchenko. Fluctuational opening of the double helix as revealed by theoretical and experimental study of dna interaction with formaldehyde. *Journal of molecular biology*, 108(4):665–682, 1976.

[129] Maurice Guéron, Michel Kochoyan, and Jean-Louis Leroy. A single mode of dna base-pair opening drives imino proton exchange. *Nature*, 328(6125):89–92, 1987.

[130] Daniel Coman and Irina M Russu. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in dna. *Biophysical journal*, 89(5):3285–3292, 2005.

[131] Timothy J. Richmond and Curt A. Davey. The structure of DNA in the nucleosome core. *Nature*, 423(6936):145–150, May 2003. ISSN 0028-0836. URL `http://www.nature.com/nature/journal/v423/n6936/abs/nature01595.html`.

[132] PT Lowary and Jonathan Widom. New dna sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology*, 276(1):19–42, 1998.

[133] Hans R Widlund, Prasad N Kuduvalli, Martin Bengtsson, Hui Cao, Thomas D Tullius, and Mikael Kubista. Nucleosome structural features and intrinsic properties of the tataaacgcc repeat sequence. *Journal of Biological Chemistry*, 274(45):31847–31852, 1999.

[134] M Roychoudhury, A Sitlani, J Lapham, and DM Crothers. Global structure and mechanical properties of a 10-bp nucleosome positioning motif. *Proceedings of the National Academy of Sciences*, 97(25):13608–13613, 2000.

[135] MA El Hassan and CR Calladine. Conformational characteristics of dna: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 355(1722):43–100, 1997.

[136] Eugene YD Chua, Dileep Vasudevan, Gabriela E Davey, Bin Wu, and Curt A Davey. The mechanics behind dna sequence-dependent properties of the nucleosome. *Nucleic acids research*, 40(13):6338–6352, 2012.

[137] Jonathan Widom. Role of dna sequence in nucleosome stability and dynamics. *Quarterly reviews of biophysics*, 34(03):269–324, 2001.

[138] Stephanie Johnson, Yi-Ju Chen, and Rob Phillips. Poly (da: dt)-rich dnas are highly flexible in the context of dna looping. *PloS one*, 8(10):e75799, 2013.

[139] Yair Field, Noam Kaplan, Yvonne Fondufe-Mittendorf, Irene K Moore, Eilon Sharon, Yaniv Lubling, Jonathan Widom, and Eran Segal. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology*, 4(11):e1000216, 2008.

[140] Tali E Haran and Udayan Mohanty. The unique structure of a-tracts and intrinsic dna bending. *Quarterly reviews of biophysics*, 42(01):41–81, 2009.

[141] Edward C Sherer, Sarah A Harris, Robert Soliva, Modesto Orozco, and Charles A Laughton. Molecular dynamics studies of dna a-tract structure and flexibility. *Journal of the American Chemical Society*, 121(25):5981–5991, 1999.

[142] Jean Louis Leroy, Eric Charretier, Michel Kochoyan, and Maurice Gueron. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to dna curvature. *Biochemistry*, 27(25):8894–8898, 1988.

[143] Helen Hwang, Noah Buncher, Patricia L Opresko, and Sua Myong. Pot1-tpp1 regulates telomeric overhang structural dynamics. *Structure*, 20(11):1872–1880, 2012.

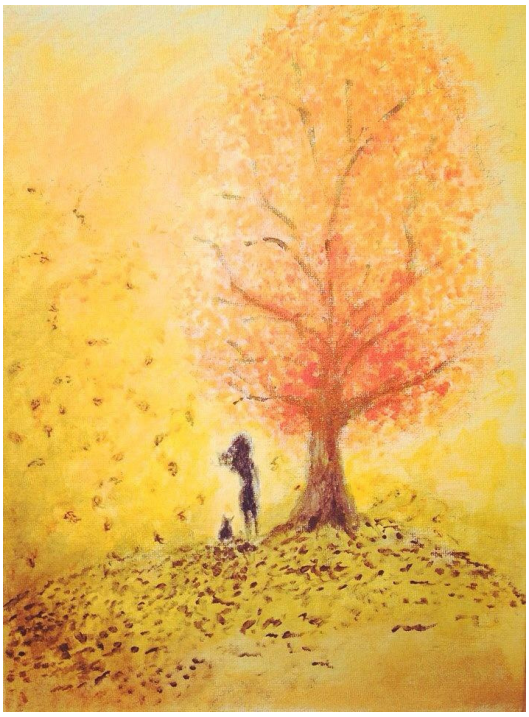
[144] Thuy TM Ngo, Qiucen Zhang, Ruobo Zhou, Jaya G Yodh, and Taekjip Ha. Asymmetric unwrapping of nucleosomes under tension directed by dna local flexibility. *Cell*, 160(6):1135–1144, 2015.

[145] Gabriel Rosanio, Jonathan Widom, and Olke C Uhlenbeck. In vitro selection of dnas with an increased propensity to form small circles. *Biopolymers*, 2015.

[146] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.

VITA

Tung Le received his bachelor in physics from Hanoi National University of Education in 2008. He then came to Georgia Institute of Technology to study biophysics as a graduate student. After the first year working with Dr. Toan Nguyen to investigate magnesium-induced DNA condensation by computer simulations, he moved into the field of experimental single-molecule biohphysics with the support from Dr. Harold Kim. Using single-molecule Fluorescence Resonance Energy Transfer, he studied bending mechanics of double-stranded DNA molecules in the strong bending condition. For future research, Tung wants to perform other single-molecule manipulation techniques and apply them to study various biological structures in the cells. Outside the lab, Tung enjoys painting and drawing illustrations.



Alone in the Woods