A Generalized Theory of DNA Loop Formation: Considering the effects of Binding Topology and Sequence-Dependent Curvature

by

David P. Wilson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Jens-Christian Meiners, Chair
Professor Dante Amidei
Professor Noel C. Perkins
Assistant Professor Jennifer P. Ogilvie
Scientist Alexei V. Tkachenko, Brookhaven National Laboratory

Study without reflection is a waste of time,
Reflection without study is dangerous.
— Confucius

Acknowledgments

I am grateful for the opportunity to work in the Meiners Laboratory at the University of Michigan under the the guidance of my thesis advisor Jens-Christian Meiners, who allowed flexibility and independence in my research. I greatly appreciated his keen physical insight and drive to lunge headstrong into new research areas. I also want to thank Alexei Tkachenko, who tirelessly assisted my research and encouraged me throughout. I want to thank Dante Amidei for strenuously insisting that I demonstrate the relevance and interest of these biophysical systems to physicists. I would also like to thank Jennifer P. Ogilvie for her steady encouragement and assistance when I changed research focus to biophysics.

Working in the Meiners laboratory allowed the delightful collaboration. I would like to thank Gerhard Blab, a postdoc in the Meiners lab for his guidance. I would like to thank Yih-Fan Chen, a former student of the Meiners lab for his collaboration and encouragement. I would like to extend a special thank you to Krishnan Raghunathan, who tirelessly read so many drafts of my work, your insight and attention to detail was greatly appreciated.

I was blessed to have the opportunity to work with Noel C. Perkins whose quiet patience provided a delightful learning environment, while challenging me to always provide a physical interpretation for my work. I would like to thank Sachin Goyal, a student of Noel C. Perkins, for his delightful enthusiasm and selfless mentoring throughout the beginning of my work, without his efforts this project would not have begun. I would also like to thank Todd D. Lillian, a student of Noel C. Perkins, for his patience and assistance throughout this endeavor, as well as his collaboration through providing data for chapter 3, 4 and 5. Finally I am grateful to Andy Hirsh, a student of Noel C. Perkins, for providing me much of the data for chapter 4 and chapter 6.for his long hours in a room which was way too hot, helping me get data for my thesis.

I would like to thank Rob Phillips of Cal Tech, for allowing me to come and participate in his Biophysics Bootcamp, and for his Put your head on the chopping block

challenges a the Biophysical Society meetings.

Finally I would like to thank my wife Margaret Wilson, who was a wonderful source of support, and to whom I have dedicated this thesis. She was able to foresee the successful completion of this work years before I did, her faith, support and friendship have sustained me throughout the years.

I am grateful to the Physics Department at the University of Michigan for allowing me to teach as a means of financial support and to the research funding provided to me by Jens-Christian Meiners, which was partially supported by the NIH GMS65934.

Table of Contents

List of Figures

Figure

List of Appendices

Appendix

Chapter 1

Introduction

1.1 Biological Background

The cell is a very crowded environment, containing many large structures such as organelles and actin networks and protein bound DNA. To illustrate the crowding, consider that every human cell contains 46 chromosomes within the cells nucleus, and that each of those chromosome has an average length of $2-3\ cm$, for a total length of $\sim 1\ m$ of DNA per cell. The nucleus of the cell is on average a few microns in diameter, which requires an enormous amount of compactification of DNA. The process of compactifying DNA is further complicated by the fact that the DNA must remain readily accessible on the local genetic level for gene expression to occur. Therefor some form of mechanical control must exist in tandem with any biochemical controls if DNA is to be packaged and unpackaged for genetic transactions without destroying the genetic information contained within. The enormity of this mechanical system compels us to study the effects of DNA curvature and twist on a smaller scale in order to characterize their effects.

1.1.1 DNA Structure

All DNA consists only of four different nucleotide subunits, adenine (A), cytosine (C), guanine (G) and thymine (T), which combined with the sugar-phosphate backbone comprise the DNA bases. These bases bind together to form long polymer chains. Two of these chains wrap around one another due to hydrophobic interactions of the nucleotides, and produce the double helix structure, see fig. 2.1. The double helix is held together by hydrogen bonding between the base pairs, which alway pair up in

the same manner, (A-T) and (G-C). DNA is a fairly slender molecule with an average diameter of 2 nm, . DNA is also very stiff on the scale of 50 nm $\sim$ 150 base pairs, which is referred to as the bending persistence length. The wrapping of the DNA base-pairs produces an uneven groove structure, with the major and minor groove having a width of $2.2/nm$ and 1.2 nm respectively, see fig. 2.1. Another important physical feature of DNA is the inhererent directionality along its backbone referred to as $5'$ to $3'$, of which complementary strands of DNA must pair up in opposing order, $3'$ to $5'$. The $5'$ end has a terminal phosphate group, and the $3'$ end has a terminal hydroxyl group. See [6] for more information on DNA structure.

1.1.2 Lac Operon of *E. coli*

The breadth of DNA-protein structure formation is enormous, and we will focus on gene regulation of prokaryotic systems, which are a group of organisms which lack a cell nucleus or other membrane bound organelles. Of these organisms, we will use the lac operon of *Escherichia coli*, or *E. coli* for short, as our illustrative example for gene regulation through DNA loop formation, as proteins found here are homologous *i.e* they are derived from a common ancestor, to those of Archaebacteria and Eukaryotic systems. *E. coli* is perhaps the best studied prokaryotic system, and serves as the hydrogen atom for biophysics.

E. coli is capable of digesting glucose and lactose sugars. Digestion of lactose requires a specific protein, known as β-galactosidase, which *E. coli* is capable of producing through gene expression. The group of three genes which control this ability, combined with the lac repressor (LacI) and the DNA binding operators are known as the Lac Operon, see fig. 1.1. The DNA operator sites within the lac operon facilitate binding of a protein known as the lactose repressor, which is independently encoded on a gene known as LacI. We will now turn our attention to the lac repressor protein, see [11] for more information.

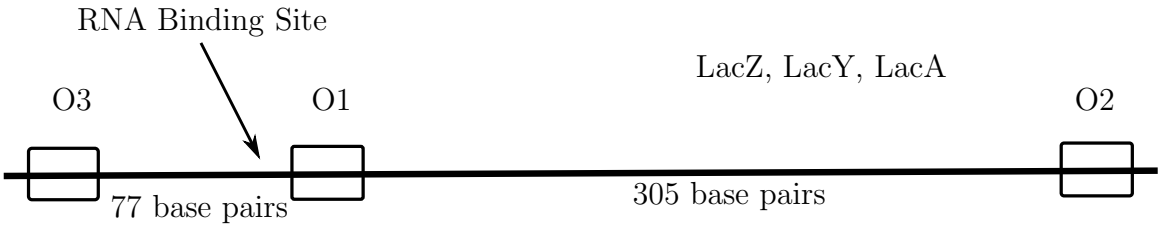

Figure 1.1 The lac operon of *E. coli* contains 3 DNA binding operators, O_1, O_2 and O_3 and 3 genes LacZ, LacY and LacA, referred to as LacZYA. There are three possible combinations of binding any two of the DNA operators, O_1/O_2, O_1/O_3 and O_2/O_3. In addition to what is shown here, the lac operon contains an additional gene for the lac repressor with its own promoter site.

DNA Operators

DNA operators are specific ordered sequence of base pairs which have a high binding affinity to a regulatory protein. Often these regulatory proteins are capable of some binding to any sequence of DNA, although a random sequence of DNA base pairs will typically have such a low binding affinity, that the protein will simply diffuse away as quickly as it arrived. The wild-type lac operon contains three DNA operators, see fig. 1.1.

1.2 Gene Expression

Gene expression occurs when a protein named RNA polymerase binds to DNA at a specific site on the DNA backbone known as the promoter site. The RNA polymerase then begins to move along the DNA from the $5'$ to $3'$ direction, unzipping and exposing the DNA bases inside the double helix. The RNA polymerase creates a complementary strand of mRNA to the DNA gene sequence, which is known as transcription. This mRNA is complementary to the DNA bases, with the exception of uracil (U) in place of thymine (T). The mRNA is then transported to the ribosome, where the Ribosome then translates the m-RNA sequence into amino acids. The subsequent amino sequence forms the associated protein, which then diffuses or is otherwise transported away from the ribosomes to conduct its purpose within the cell.

Wild-type *E. coli* expresses three genes, LacZ, LacY and LacA (LacZYA), which

lay in that order after the promoter site. LacZYA are expressed as β-galactosidase, permease, β-galactosidase transacetylase, respectively. Of these three genes, only β-galactosidase is known to participate in a metabolic pathway, and mutations of of LacYA do not appear to affect the cells ability to digest lactose.

1.2.1 Gene Regulation

Gene regulation refers to an organisms ability to turn on and off its genes. Producing proteins costs the cell valuable energy, which it must be able to control in order to survive. Regulation in lac repressor-like systems results from disrupting the ability of RNA polymerase to begin transcription, since the locating the promoter site, *e.g.* such as LacZYA, is a critical first step to expressing these genes.

Over time *beta*-galactoside will break down naturally in the cell and constant production is required for digesting of lactose. If no lactose is present, or if a better fuel source is available, *E.coli* must turn off LacZYA. In *E. coli* this regulation is accomplish by the lactose repressor, lacI. When there is no lactose present, lacI binds to the DNA operators and inhibits the start of transcription. When lactose is added to the cell, β-galactoside is produced.

Lactose binds to the lac repressor near the C-terminus, and causes a conformational change within the protein that reduces its ability to bind DNA, and inhibit repression, see fig. 1.2 for more information. The binding of the lac repressor to DNA and subsequent removal by lactose sugar is known as a genetic switch.

LacI Crystal Structure

The crystal structure of LacI was determined by Lewis et al [26]; where lac repressor protein and DNA operator segments formed co-crystals and were then mapped by x-ray diffraction experiments. Successful crystalization of LacI required the formation of DNA-protein co-crystals, demonstrating that LacI can simultaneously bind to two distant DNA operator sites, forming a loop. The wild-type *E. coli* DNA contains three distinct operator regions, separated by 77 and 310 base pair, respectively. The central operator O_1, see fig. 1.1, has the strongest affinity of the three operators, however it was shown by Oehler et al. [42] that all three operators were needed for full repression of the Lac Operon.

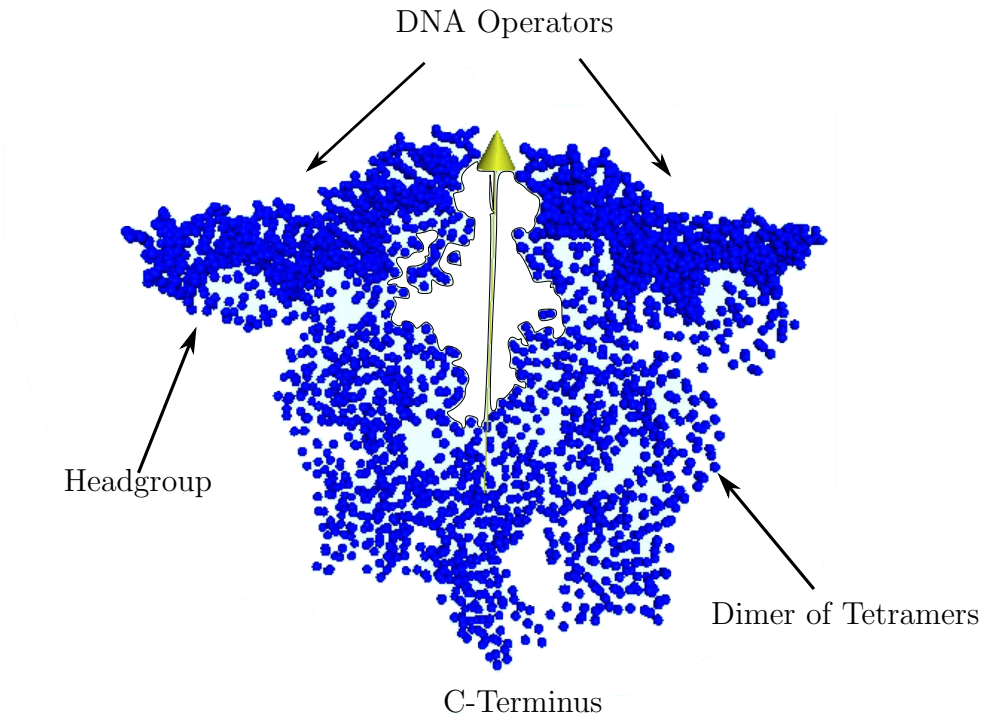

Figure 1.2 The lac repressor protein (LacI) is a tetramer which is formed into two dimeric arms, joined at the C-terminus by Van der Waals interactions. The LacI head groups are shown here bound to DNA operators. The protein structure was determined by Lewis et al. [26].

1.3 Gene Regulation via DNA Loop Formation

There are several types of DNA-protein loop systems [1] that can be split into two main groups, those which use an architectural protein to help form the DNA into a loop, and those which form a loop solely through thermal fluctuations. We focus on the later, as it is the simplest system we can examine, allowing us to gain a physical foothold from which we can build an intuitive understanding of this multifaceted, complex biophysical world. We now exclusively focus on the lac repressor system on *E. coli*

1.3.1 Phase Dependence of Lac Repressor Binding

Careful experiments by NAME et al. [] have shown that incremental additions of DNA base pairs between two operator sites will dramatically affect the ability of *E.*

coli to regulate LacZYA. The repression level of *E. coli* was measured and shown to be periodic with a period of ~ 10.5 base pairs, suggesting that the helical repeat of DNA plays a role. The sharp cyclic structure of the repression curve is known as DNA phasing, which we explore in detail in chapter 4 and chapter 5. The phasing dependence suggests a purely physical mechanism, as nothing to do with the affinity of DNA-protein binding is directly affected.

1.3.2 Cooperativity of DNA operators

Experiments by Oehler et al. [42] showed that all three DNA operators are required for full repression of LacZYA. They measured the repression level after carefully removing each operator in turn, and found that all three possible loops were utilized in repression. This result was unexpected given the shortest loop 77 base pairs is one half of a persistence length, and looping should be improbable given the stiffness of DNA.

1.3.3 Reduction of Repression through Lac Dimers

Additional experiments by Dong et al [12] showed that if you mutate away the C-terminus of the lac repressor, the ability of the cell to repress LacZYA is greatly reduced, demonstrating the need for simultaneous binding to the lac repressor. This work also demonstrates the stabilizing presence of the loop, as dimers separately occupy binding sites were unable to effectively regulate lacZYA.

1.4 Statistical Mechanics

While some systems use Architectural proteins to prebend the DNA prior to binding by a regulatory protein, the lac operon of *E. coli* does not [1]. It is a thermally driven mechanical looping process on a scale where DNA is fairly rigid, making it a great statistical mechanics problem. We approach the description by computing partition functions of the open and looped state of DNA, in order to compute the Stockmayer Jacobson J factor.

Calculating the probability that contact will occur between two distant ends of a polymer under prescribed orientations is a long-standing question of considerable significance in polymer physics. This problem was rigorously defined in the context of

polyelectrolyte condensation as the ratio of equilibrium constants for cyclization and bimolecular association by introduction of the Jacobson Stockmayer (J) factor [21]. Yamakawa and Stockmayer expanded on this work using the Kratky-Porod wormlike chain model (WLC) to compute the J factor of angle-independent DNA ring-closure probabilities [62]. Shimada and Yamakawa then included twist alignment of the end points [49], known as phasing, to explain the measured oscillatory cyclization rates by Shore and Baldwin on DNA shorter than 500 base pairs [50]. Shimada and Yamakawa calculated the J factor for the ring and and unconstrained loop, by treating DNA as a homo-polymer with coincident end points and parallel tangent vectors, as well as with coincident end points with unconstrained tangent vectors, respectively.

1.5 DNA and Protein modeling

We study the mechanics of this protein-mediated loop formation and characterize the variables which control loop formation and stability, by treating the DNA as an elastic polymer. Specifically we use the elastic rod model developed by Goyal et al. [15], which coarsely grains the DNA into a continuous approximation. This model extends the worm-chain model (WLC) to include twist and sequence dependent elasticity and curvature. The DNA operators occur in three forms in nature, and a fourth form known as the palindromic operator is also used in laboratory experiments due to its enhanced binding affinity to the lac repressor. We will only consider palindromic operators in this work, since they allow the largest number of binding configurations and modeling simplicity. Since DNA is known to have a dual groove structure along its double helix, we construct a model which is capable of treating DNA as a heteropolymer with two distinct bending elasticities, and a single torsional elasticity.

Our semi-analytic computation of the J factor generalizes this closure probability to include arbitrary end-point locations, binding orientations, sequence-dependent curvature and elasticity while reproducing the earlier results of Shimada and Yamakawa for the ring and unconstrained loop. As many biologically relevant cases do not fit neatly into one of these special cases, we generalize the computation of J factors to cover any nearly planar shape, as well as to build our intuition for protein mediated DNA loop formation.

We numerically calculate J factors based on a semi-analytic continuous elastic rod formulation that includes as inputs specified end point locations and orientations of the DNA. This formulation goes beyond the homogeneous straight elastic rod of

Balaeff et al[3, 2, 4] by allowing the inclusion of intrinsic curvature and stiffness based upon sequence-dependent effects. We also compute thermal fluctuations, which contribute non-trivially to the free energy cost of loop formation. This aspect has some similarities to the recent work of Zhang and Crothers [63, 65] who used a discrete model to compute thermal fluctuations, although their J factors disagreed with the previous results of Shimada and Yamakawa [49].

Previous works by Olson et al. have modeled DNA as an elastic rod with sequence specific properties using individual base pairs as their elements to examine normal modes [35, 43, 44, 46, 36, 9, 45, 53]. While Monte Carlo methods have been successfully used to compute J factors [55, 47], they are in general computationally taxing, making it potentially difficult to separate out the individual effects of curvature and stiffness, or make the distinction between enthalpic and entropic contributions; by contrast, our computation of the J factor based on a desired equilibrium shape takes only minutes on a desktop computer.

Many DNA-binding proteins impose very specific boundary conditions on DNA loop formation. Previous results by Swigon et al., Segall et al. and Purohit et al have shown that boundary condition constraints on the DNA end points play a significant role in the facilitation of loop formation [53, 41, 48]. Boundary conditions have also been suggested by Tkachenko [54] as an explanation for the striking disagreements between the cyclization rates measured by Du et al. [14, 13] and Cloutier et al. [61, 8]. Therefore, any useful model for these interactions must accommodate such arbitrary boundary conditions. Thus, the J factor framework gives quantitative insights into the mechanics of protein-mediated DNA loop formation and is important for multiscale models of larger DNA-protein assemblies such as chromatin and nucleosomes. The nucleosome is DNA-protein complex which is made up of a persistence length of DNA wrapped 1.6 times around a core histone protein.

1.6 Guide to contents

Within this thesis you will find

Chapters

- **Chapter 2: Theory** - In this chapter we provide a detailed calculation of the J factor for arbitrary DNA configurations. We rigorously define the elastic rod

model as well as provide a firm mathematical basis for the remaining chapters.

- **Chapter 3: Planar Loops** - In this chapter we study DNA loop formation without a mediating protein. This work provides an initial physical foundation for which can relate the results of DNA-lac binding. Here we apply the generalized J factor methods to planar DNA conformations. This chapter the computation of the effective torsional persistence length ℓ_τ^* for nearly planar loops. We verify the J factors for the ring and unconstrained loop by Shimada et al. [49].
- **Chapter 4: Lac Induced Loops** - *E. coli* uses thermal fluctuations to regulate its DNA through loop formation which can be described with our J factor formulation. This section deals with intrinsically straight DNA. We discuss how to connect our formalism to experimental predictions using FRET and TPM. Our results for straight DNA bound to Lac repressor are then compared to the Monte Carlo results of Towles et al [55].
- **Chapter 5: Intrinsically Curved DNA** - Here we study the effects of intrinsic curvature on DNA loop formation and see that our results are consistent with recent Gel electrophoresis results of Mehta et al. [37] and FRET measurements of Morgan et al. [38].
- **Chapter 6: Extensions** - Here we discuss our ongoing study of heterogeneous stiffness of DNA, as well as twist-bend coupling of DNA. We also discuss extensions to include Molecular Dynamics computations of the lac repressor, as well as inclusion of architectural proteins. Finally we recommend several experimental systems which can be used to validate our predictions.
- **Chapter 7: Summary of Contributions** - A concise summary of the specific findings and contributions of this work.

Appendices

- **Appendix A: Notation** - A list of all variables, and their dimensionless products. We also explain how to convert the J factors to units of Molarity.
- **Appendix B: Rotations** - A clear and concise discussion about the choice of finite rotations, given they do not commute.
- **Appendix C: Linear Operators** - An explicit list of the Linear Operators for the displacement angle formulation of the solution to the normal modes of the Hamiltonian using the Galerkin Method.
- **Appendix D: Hamiltonian Matrix** - An explicit list of Hamiltonian matrix

operators, and detailed instructions of how to form the Hamiltonian Matrix for the Ritz Method.

- **Appendix E: Delta Function Constraints** - The full details of the computation of the constraint matrix V for the looped state of DNA.

Chapter 2

Theory

2.1 Introduction to J factor

Given many cellular regulatory functions are governed by thermally driven DNA loop formation, we set out to calculate the Stockmayer Jacobson J factor [21], which is the probability density of finding two properly aligned DNA end points within a given capture volume. Loop formation is governed by two effects, enthalpy and entropy. Short lengths of DNA are enthalpically prohibited from forming loops by the large energetic cost of bending DNA, and long lengths of DNA are entropically prohibited by the large conformational space open to the molecule and the unlikelihood of the two end-points coming into contact due to thermal fluctuations.

The J factor is defined as [21, 63]

$$J = 8\pi^2 \frac{Z_\ell}{Z_o} \tag{2.1}$$

where Z_ℓ is the partition function for the subset of DNA molecules which are subject to the looped state constraints and Z is the partition function of the unlooped or open DNA molecules which are unconstrained. The $8\pi^2$ coefficient arises due to the constraints of the loop formation. We want to compute the total probability of DNA loop formation, which requires that we multiply the ratio of partition functions $\frac{Z_\ell}{Z}$ by 4π to account for all orientations of the loop over solid angle. The additional factor of 2π arrises from the torsional alignment of end-points of the helical axis. Understanding the role of binding conditions on DNA loop formation is one of the primary goals of this work.

2.1.1 Partition Functions

In order to compute the partition functions of the J factor we construct a Hamiltonian $\mathcal{H}$ about the open $\mathcal{H}^o$ and looped $\mathcal{H}^\ell$ states of DNA. The looped state Hamiltonian $\mathcal{H}^\ell$ includes the internal energy of the DNA, E_ℓ, the enthalpic cost of loop formation with respect to the ground state, as well as the thermal fluctuations about the equilibrium state, $\beta H^\ell = \beta E_\ell + \frac{1}{2}\sum_i \lambda_i \xi_i^2$, where $\beta = 1/k_B T$, k_B is the Boltzmann constant and T is the temperature. Small deformation due to thermal fluctuations about these states can be expressed in terms of the normal mode basis. We describe deformations to the equilibrium state by three displacements $\vec{u}$ and three rotation angles θ_1, θ_2, ψ, see fig. 2.3 and fig. 2.1. We determine the entropic cost of DNA loop formation by computing the change in the conformational space available to the open and looped states. This change is directly related to the ratio of the eigenmodes which describe thermal fluctuations about the two equilibrium states.

The looped state partition function is

$$Z_\ell = \int [\mathrm{d}\xi_i]\, e^{-\beta H^\ell} \delta^3(|\vec{u}(L)|)\delta(\theta_{1L})\delta(\theta_{2L})\delta(\psi_L) \tag{2.2}$$

where ξ_i is the amplitude of the i^{th} normal mode of the DNA and $[\mathrm{d}\xi_i] = \prod_i^\infty \mathrm{d}\xi_i$ is a shorthand for the integration variables over the space of eigenmode amplitudes. The six δ-functions enforce the boundary constraints on the DNA loop, and are expanded upon later in section 2. The δ-functions contribute dimensions of inverse volume to the J factor.

We take the open state to be the ground state of our system, and consider its energy to be only thermal excitations about the ground state, given by the open state Hamiltonian $H^o = \frac{1}{2}\sum_i \lambda_i \xi_i^2$, with λ_i is the eigenvalue of the i^{th} normal mode. The open state partition function is

$$Z = \int [\mathrm{d}\xi_i]\, e^{-\beta H^o} \tag{2.3}$$

We use the overall sequence length to make quantities dimensionless. The curvature vectors $\vec{\kappa}$ carry dimension of L^{-1}, the stiffness Matrix $B(s)$ carries the dimensions of L and then the integration over $\mathrm{d}s$ carries the dimensions of L . The arc-length is parameterized by $S = sL$ where $0 \leq s \leq 1$. Dimensional scaling is key to understanding the J factor, for this reason we write the explicit units of length when they contribute to this scaling, *e.g.* ℓ_p/L and ℓ_τ/L. For simplicity all other quanti-

ties in this chapter are assumed to have been made dimensionless, unless otherwise indicated, see Appendix A for more information.

2.1.2 Persistence Length

The length of DNA being considered affects which model is appropriate to describe it. On longer lengths, thousands of base pairs, the worm-like chain (WLC) is employed, which is a continuum limit of the freely jointed chain [34]. The WLC does not include torsion of the molecule and for shorter lengths we need to include this degree of freedom. On macroscopic scales when we perturb a rod, we see the perturbation affect the entire rod, or more precisely the motions are all correlated. This concept of length of correlation of perturbations is understood through the concept of persistence length. We can compute the bending persistence length by calculating the correlation function of two tangent vectors $\hat{t}(s_1)$ and $\hat{t}(s_2)$

$$\langle \hat{t}(s_1) \cdot \hat{t}(s_2) \rangle \propto \exp(-|s_1 - s_2|/\ell_p), \tag{2.4}$$

where ℓ_p is the bending persistence length, the characteristic length scale over which thermal fluctuations of DNA remain correlated, see fig. 2.3. We also employ the torsional persistence length ℓ_τ, which sets the characteristic length scale over which torsional deformations remain correlated. We use commonly accepted averaged values for the bending and torsional persistence length of 50 nm and 75 nm, respectively, [17, 52, 5, 4]. On occasion we relate the torsional persistence length to the bending persistence length as $\ell_\tau = \sigma \ell_p$.

We treat the statistical properties of DNA, such as bending ℓ_p and torsional persistence lengths ℓ_τ, as intrinsic elasticity parameters for the elastic rod model of DNA. We use a stiffness tensor, $B(s)$ to represent the elasticity of DNA. We then use an elastic rod model following Goyal et al. [15] to compute the equilibrium states of DNA under various boundary conditions, which can be inferred from LacI-DNA co-crystals in the case of gene regulation. Once these equilibria are known, we can compute their respective normal modes and subsequently construct the J factor.

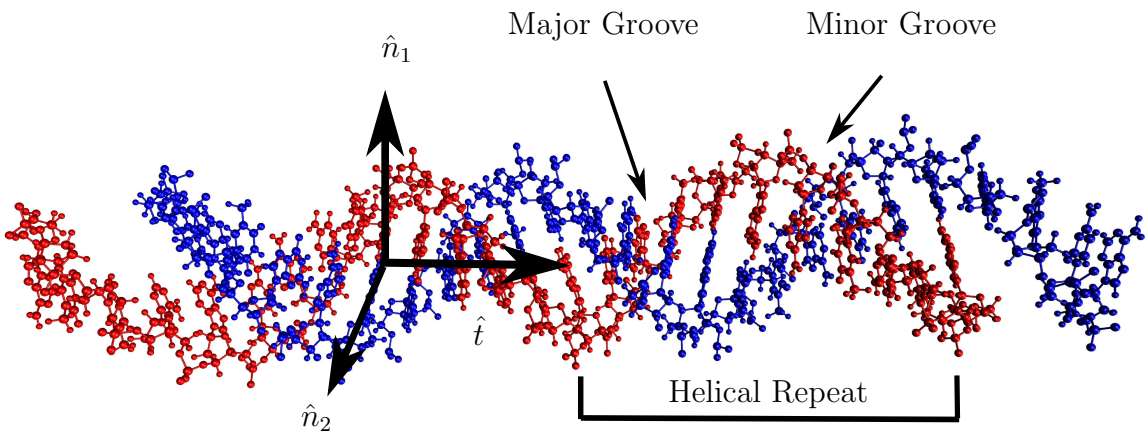

Figure 2.1 The double helix structure of DNA gives rise to a major and minor groove, which have a spacing of 2.2 nm and 1.2 nm respectively. DNA also repeats its axial orientations, known as the helical repeat, on average every 10.5 base pairs. The unequal groove spacing motives us to use a body fixed coordinate system of $\hat{t}, \hat{n}_1, \hat{n}_2$, where we align the tangent vector with the centerline of the DNA molecule, and the normal vectors with the Major and Minor grooves. The unequal groove spacing leads to heterogeneous bending stiffness ℓ_1 and ℓ_2 along the minor and major grooves, respectively.

2.2 Elastic Rod Model

In this section we explain how we use an elastic rod model to describe conformational changes to DNA based on thermal fluctuations about its various equilibrium conformations. We compare differences of DNA curvature provided by the elastic rod model following Goyal et al [15]. We are interested in the mechanical equilibrium state without thermal fluctuations $T = 0$. The open state of DNA is taken to be a random sequence where the non-uniform base pair stacking faults typically average out over a helical repeat and are therefore considered straight. Other sequences, such as A-tract bends have high specific intrinsic curvature and are considered in chapter 5. We use a body fixed coordinate system as opposed to the Frenet-Serret formulation. DNA requires us to make this distinction since it is not a homopolymer with isotropic stiffness, instead it has two distinct bending elasticities corresponding to the major and minor grooves, as explained in chapter 1. We investigate heterogeneous stiffness in chapter 6.

2.2.1 Body Fixed Coordinate System

A local coordinate system is then assigned to each cross section of the DNA model. These cross sections can represent individual DNA basepairs, and are assumed to rotate with the major and minor grooves of the DNA, the period or helical repeat is often taken to be ~ 10.5 base pairs, see Goyal et al. [15, 27]. Then as the DNA is bent into a looped state which obeys the specified boundary conditions, each individual cross section is tracked, allowing us to use the same local body fixed coordinate system to describe local deformations in the open and looped states of DNA, avoiding the need for a Jacobian to compare the two different coordinate systems. The body fixed basis vectors we employ are the tangent $\hat{t}$ and two normal vectors $\hat{n}_1$ and $\hat{n}_2$. The tangent is chosen to align with the sugar phosphate backbone of the DNA double helix, the normal directions are chosen to align with the minor and major grooves of the DNA, respectively, see fig. 2.1 and fig. 2.2. We consider two sets of basis vectors for each state

$$\hat{t}, \hat{n}_1, \hat{n}_2 \qquad \text{(Equilibrium State Basis)}$$
$$\tilde{t}, \tilde{n}_1, \tilde{n}_2 \qquad \text{(Deformed State Basis)}$$

In general the body fixed coordinate vectors do not align with the Frenet-Serret [23] space curved defined coordinate system, *e.g.* the normal vectors need not be directed inwards to the curve see fig. 2.2. The resulting skew symmetric relationship of vector derivatives is

$$\begin{bmatrix} \hat{t}' \\ \hat{n}_1' \\ \hat{n}_2' \end{bmatrix} = \begin{bmatrix} 0 & \kappa_2 & -\kappa_1 \\ -\kappa_2 & 0 & \tau \\ \kappa_1 & -\tau & 0 \end{bmatrix} \begin{bmatrix} \hat{t} \\ \hat{n}_1 \\ \hat{n}_2 \end{bmatrix} \tag{2.5}$$

where the prime on $\hat{t}'$ indicates a derivative with respect to the arc length parameter s. The standard Frenet-Serret space-curve description would set $\kappa_1 = 0$.

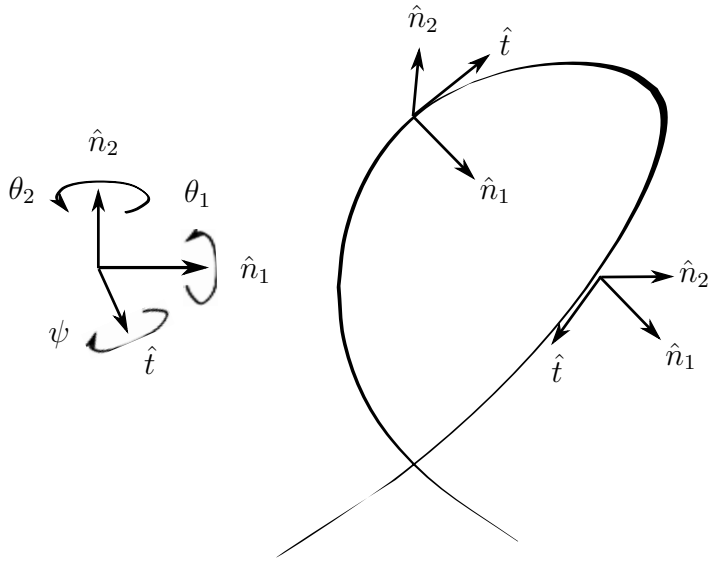

Figure 2.2 The deformation variables ψ, θ_1, θ_2 describe rotations about the body fixed basis vectors $\hat{t}, \hat{n}_1, \hat{n}_2$, respectively. The body fixed coordinate system does not match the Serret-Frenet coordinates in general, as demonstrated by $\hat{n}_1$ point inwards and outwards to the curve. The body fixed coordinates are chosen to align with the major and minor grooves of the DNA molecule, see fig. 2.1

2.2.2 Rotation Matrices and Deformations

We are interested in modeling deformations of DNA due to thermal fluctuations about its equilibrium state. We use three local deformation angles $(\psi, \theta_1, \theta_2)$ to describe deformations about the equilibrium state, we refer to their collective deformation as $\xi = (\psi, \theta_1, \theta_2)$, which should not be confused with an infinitesimal rotation. Deformations about the tangent vector are given by ψ, whereas deformations about the two normal vectors $\hat{n}_1$ and $\hat{n}_2$ are given by θ_1 and θ_2, respectively. These deformations also lead to position dependent displacements of the rod centroid from the equilibrium location, referred to as $\vec{u}$, see fig. 2.3.

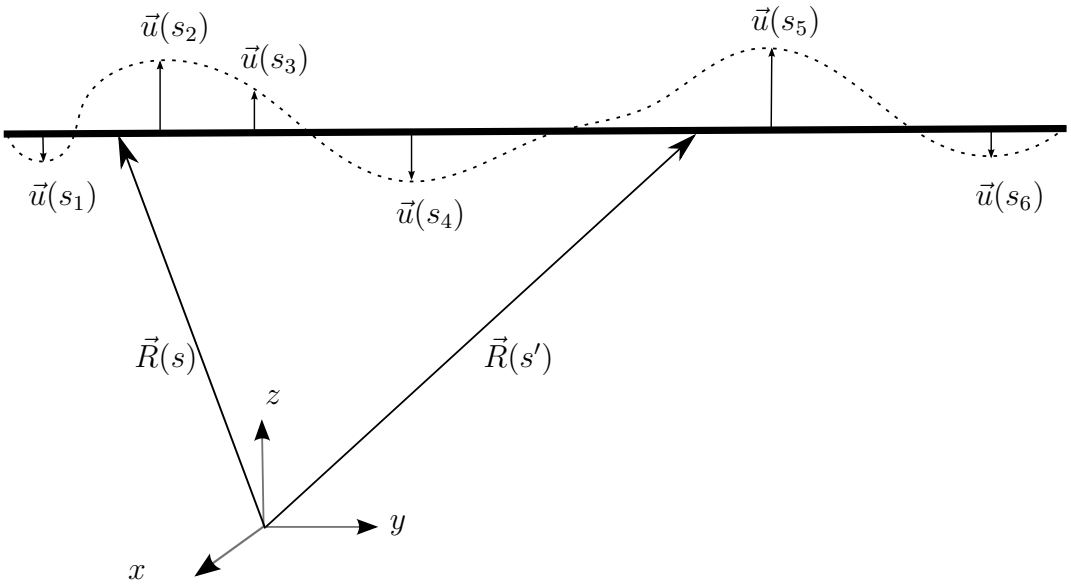

Figure 2.3 Thermal fluctuations cause arbitrary deformations of the DNA molecule from the equilibrium states described by $\vec{R}$ to $\tilde{R} = \vec{R} + \vec{u}$. The looped state boundary conditions require that the end-points displacements vanish $\vec{u}(s = 1) = 0$ The displacement deformations $\vec{u}$ are about the body fixed coordinate system, see fig. 2.1 and fig. 2.2. The bending ℓ_p and torsional persistence length ℓ_τ are the characteristic length over which the tangent vectors at $\vec{R}(s)$ and $\vec{R}(s')$ remain correlated due to thermal fluctuations.

While we do not use these displacements as deformation variables, we do compute them when considering the constraints placed on our DNA molecule. We use this choice of coordinates as opposed to those of displacements because they naturally arise in the strain energy of the rod as dependent variables when you neglect the contributions of kinetic energy. By contrast, we would employ displacement variables if we were simultaneously interested in formulating the kinetic energy of the rod.

The axial stiffness of DNA is extremely high when compared to the more modest bending and torsional stiffness. We understand this fact by recognizing the energy to untwist the DNA double-helix is much more than that needed to bend or twist it due to hydrophobic interactions of the base pairs [6]. Our choice of deformation variables mean that our system is inextensible, which simplifies our theory. Our choice

of deformation angles about the equilibrium states are given by

$$R(\psi)R(\theta_1)R(\theta_2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\theta_1 & 0 & -\sin\theta_1 \\ 0 & 1 & 0 \\ \sin\theta_1 & 0 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & \sin\theta_2 & 0 \\ -\sin\theta_2 & \cos\theta_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

using the small angle approximation, keeping to 2^{nd} order terms,

$$\sim \begin{bmatrix} 1 - \frac{1}{2}(\theta_1^2 + \theta_2^2) & \theta_2 & -\theta_1 \\ -\theta_2 + \theta_1\psi & 1 - \frac{1}{2}(\theta_2^2 + \psi^2) & \psi \\ \theta_1 + \theta_2\psi & -\psi + \theta_1\theta_2 & 1 - \frac{1}{2}(\theta_1^2 + \psi^2) \end{bmatrix} \tag{2.6}$$

where since we are only considering small variations from the equilibrium, we use a small angle approximation, keeping to Gaussian Order, or 2^{nd} order.

The order of rotations needs to be carefully considered, as they do not commute when working to 2^{nd} order in deformations. The choice of order must remain consistent with our definitions of the deformation angles, see Appendix B for more information. The infinitesimal rotations that connect the deformed state to the equilibrium state vectors are

$$\begin{bmatrix} \tilde{t} \\ \tilde{n}_1 \\ \tilde{n_2} \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{2}(\theta_1^2 + \theta_2^2) & \theta_2 & -\theta_1 \\ -\theta_2 + \theta_1\psi & 1 - \frac{1}{2}(\theta_2^2 + \psi^2) & \psi \\ \theta_1 + \theta_2\psi & -\psi + \theta_1\theta_2 & 1 - \frac{1}{2}(\theta_1^2 + \psi^2) \end{bmatrix} \begin{bmatrix} \hat{t} \\ \hat{n}_1 \\ \hat{n}_2 \end{bmatrix}. \tag{2.7}$$

This choice uniquely defines ψ as the angle between the perturbed state $\tilde{n}_1$ and equilibrium state $\tilde{n}_2$ as well as θ_1, between $\hat{n}_2$ and $\tilde{t}$, and θ_2, between the tangent $\tilde{t}$ and $\hat{n}_1$. More information is available on this choice of rotation matrices in Appendix B. Henceforth the perturbed or deformed states are represented by tildes such as $\tilde{t}$ or $\tilde{\kappa}$ and the equilibrium terms are indicated by a hat $\hat{t}$ or by a vector $\vec{\kappa}$. Equilibrium components have no designations *e.g.* κ_2. The deformed state basis vectors $\{\tilde{t}, \tilde{n}_1, \tilde{n}_2\}$

remain orthonormal to 2^{nd} order in deformation variables

$$\begin{aligned}
\tilde{t}\cdot\tilde{t} &= (1-\theta_1^2-\theta_2^2)(\hat{t}\cdot\hat{t}) + (\theta_2^2)(\hat{n}_1\cdot\hat{n}_1) + (\theta_1^2)(\hat{n}_2\cdot\hat{n}_2) = 1\\
\tilde{n}_1\cdot\tilde{n}_1 &= (\theta_2^2)\hat{t}\cdot\hat{t} + (1-\theta_2^2-\psi^2)\hat{n}_1\cdot\hat{n}_1 + (\psi^2)\hat{n}_2\cdot\hat{n}_2 = 1\\
\tilde{n}_2\cdot\tilde{n}_2 &= (\theta_1^2)\hat{t}\cdot\hat{t} + (\psi^2)\hat{n}_1\cdot\hat{n}_1 + (1-\theta_1^2-\psi^2)\hat{n}_2\cdot\hat{n}_2 = 1\\
\tilde{t}\cdot\tilde{n}_1 &= (-\theta_2+\theta_1\psi)\hat{t}\cdot\hat{t} + (\theta_2)\hat{n}_1\cdot\hat{n}_1 + (-\theta_1\psi)\hat{n}_2\cdot\hat{n}_2 = 0\\
\tilde{t}\cdot\tilde{n}_2 &= (\theta_1+\theta_2\psi)\hat{t}\cdot\hat{t} + (-\theta_2\psi)\hat{n}_1\cdot\hat{n}_1 + (-\theta_1)\hat{n}_2\cdot\hat{n}_2 = 0\\
\tilde{n}_1\cdot\tilde{n}_2 &= (-\theta_1\theta_2)\hat{t}\cdot\hat{t} + (-\psi+\theta_1\theta_2)\hat{n}_1\cdot\hat{n}_1 + (\psi)\hat{n}_2\cdot\hat{n}_2 = 0
\end{aligned} \tag{2.8}$$

2.2.3 Basis Vectors and Curvature Components of the Deformed State

We now use our orthonormal deformation basis to express changes to local curvature due to thermal fluctuations about the equilibria of the open and looped state of DNA, as in fig. 2.3. In total there are four different curvature vectors,

$$\vec{\kappa}^o = (\kappa_1^o(s), \kappa_2^o(s), \tau^o(s)) \xrightarrow{\xi(\theta_1,\theta_2,\psi)} \tilde{\kappa}^o = (\tilde{\kappa}_1^o(s), \tilde{\kappa}_2^o(s), \tilde{\tau}^o(s)) \qquad \text{(Open State)}$$

$$\vec{\kappa}^\ell = \left(\kappa_1^\ell(s), \kappa_2^\ell(s), \tau^\ell(s)\right) \xrightarrow{\xi(\theta_1,\theta_2,\psi)} \tilde{\kappa}^\ell = \left(\tilde{\kappa}_1^\ell(s), \tilde{\kappa}_2^\ell(s), \tilde{\tau}^\ell(s)\right) \qquad \text{(Looped State)}$$

First we need to compute the derivatives of the basis vectors so that we may compute the components of the deformation curvature vector in a manner similar to eq. (2.5). Then differentiating the new basis vectors eq. (2.7) with respect to the arc length s, indicated by the prime, leads to,

$$\begin{bmatrix}\tilde{t}'\\ \tilde{n}_1'\\ \tilde{n}_2'\end{bmatrix} = \begin{bmatrix}1-\frac{1}{2}(\theta_1^2+\theta_2^2) & \theta_2 & -\theta_1\\ -\theta_2+\theta_1\psi & 1-\frac{1}{2}(\theta_2^2+\psi^2) & \psi\\ \theta_1+\theta_2\psi & -\psi+\theta_1\theta_2 & 1-\frac{1}{2}(\theta_1^2+\psi^2)\end{bmatrix}' \begin{bmatrix}\hat{t}\\ \hat{n}_1\\ \hat{n}_2\end{bmatrix}$$
$$+ \begin{bmatrix}1-\frac{1}{2}(\theta_1^2+\theta_2^2) & \theta_2 & -\theta_1\\ -\theta_2+\theta_1\psi & 1-\frac{1}{2}(\theta_2^2+\psi^2) & \psi\\ \theta_1+\theta_2\psi & -\psi+\theta_1\theta_2 & 1-\frac{1}{2}(\theta_1^2+\psi^2)\end{bmatrix} \begin{bmatrix}\hat{t}\\ \hat{n}_1\\ \hat{n}_2\end{bmatrix}'$$

then using eq. (2.5)

$$
\begin{aligned}
&= \Bigg\{ \begin{bmatrix} 1-\frac{1}{2}\left(\theta_1^2+\theta_2^2\right) & \theta_2 & -\theta_1 \\ -\theta_2+\theta_1\psi & 1-\frac{1}{2}\left(\theta_2^2+\psi^2\right) & \psi \\ \theta_1+\theta_2\psi & -\psi+\theta_1\theta_2 & 1-\frac{1}{2}\left(\theta_1^2+\psi^2\right) \end{bmatrix}' \\
&\quad + \begin{bmatrix} 1-\frac{1}{2}\left(\theta_1^2+\theta_2^2\right) & \theta_2 & -\theta_1 \\ -\theta_2+\theta_1\psi & 1-\frac{1}{2}\left(\theta_2^2+\psi^2\right) & \psi \\ \theta_1+\theta_2\psi & -\psi+\theta_1\theta_2 & 1-\frac{1}{2}\left(\theta_1^2+\psi^2\right) \end{bmatrix} \begin{bmatrix} 0 & \kappa_2 & -\kappa_1 \\ -\kappa_2 & 0 & \tau \\ \kappa_1 & -\tau & 0 \end{bmatrix} \Bigg\} \begin{bmatrix} \hat{t} \\ \hat{n}_1 \\ \hat{n}_2 \end{bmatrix}
\end{aligned} \tag{2.9}
$$

where we use eq. (2.7) to compute the deformed basis derivatives. The final form of the deformed state basis vectors and their derivatives of eq. (2.9) keeping to 2^{nd} order are,

$$
\begin{aligned}
\tilde{t}' = &\left(-\kappa_1\theta_1-\kappa_2\theta_2-\theta_1\theta_1'-\theta_2\theta_2'\right)\hat{t} \\
&+\left(\kappa_2+\theta_2'+\tau\theta_1-\frac{1}{2}\kappa_2(\theta_1^2+\theta_2^2)\right)\hat{n}_1 \\
&+\left(-\kappa_1-\theta_1'+\tau\theta_2+\frac{1}{2}\kappa_1(\theta_1^2+\theta_2^2)\right)\hat{n}_2
\end{aligned} \tag{2.10}
$$

$$
\begin{aligned}
\tilde{n}_1' = &\left(-\kappa_2-\theta_2'+\kappa_1\psi+(\theta_1\psi)'+\frac{1}{2}\kappa_2(\theta_2^2+\psi^2)\right)\hat{t} \\
&+\left(-\kappa_2\theta_2-\tau\psi+\kappa_2\theta_1\psi-\theta_2\theta_2'-\psi\psi'\right)\hat{n}_1 \\
&+\left(\tau+\psi'+\kappa_1\theta_2-\kappa_1\theta_1\psi-\frac{1}{2}\tau(\theta_2^2+\psi^2)\right)\hat{n}_2
\end{aligned} \tag{2.11}
$$

$$
\begin{aligned}
\tilde{n}_2' = &+\left(\kappa_1+\theta_1'+\kappa_2\psi+(\theta_2\psi)'-\kappa_2\theta_1\theta_2-\frac{1}{2}\kappa_1(\theta_1^2+\psi^2)\right)\hat{t} \\
&+\left(-\tau-\psi'+\kappa_2\theta_1+(\theta_1\theta_2)'+\kappa_2\theta_2\psi+\frac{1}{2}\tau(\theta_1^2+\psi^2)\right)\hat{n}_1 \\
&+\left(-\tau\psi-\kappa_1\theta_1-\theta_1\theta_1'-\psi\psi'+\tau\theta_1\theta_2-\kappa_1\theta_2\psi\right)\hat{n}_2
\end{aligned} \tag{2.12}
$$

We now solve for the deformation curvature components as a function of $(\psi, \theta_1, \theta_2)$.

$$\begin{bmatrix} \tilde{t}' \\ \tilde{n}_1' \\ \tilde{n}_2' \end{bmatrix} = \begin{bmatrix} 0 & \tilde{\kappa}_2 & -\tilde{\kappa}_1 \\ -\tilde{\kappa}_2 & 0 & \tilde{\tau} \\ \tilde{\kappa}_1 & -\tilde{\tau} & 0 \end{bmatrix} \begin{bmatrix} \tilde{t} \\ \tilde{n}_1 \\ \tilde{n}_2 \end{bmatrix} \tag{2.13}$$

We solve for $\tilde{\kappa}_2$ by computing $\tilde{t}' \cdot \tilde{n}_1$. In a same manner we solve for all of the deformation curvature components. This procedure is the same for open $\tilde{\kappa}_1^o, \tilde{\kappa}_2^o, \tilde{\tau}^o$ and looped $\tilde{\kappa}_1^\ell, \tilde{\kappa}_2^\ell, \tilde{\tau}^\ell$ state deformation curvature components, therefor when sufficient clarity exists, we drop the superscripts o, ℓ.

The three deformation curvature components are computed using eq. (2.13), it can also be seen that the definitions are consistent, *e.g.* $\tilde{n}_1' \cdot \tilde{n}_2 = -\tilde{n}_2' \cdot \tilde{n}_1$. If rotations are chosen incorrectly, then the deformed curvature components are inconsistent, *e.g.* $\tilde{n}_1' \cdot \tilde{n}_2 \neq -\tilde{n}_2' \cdot \tilde{n}_1$.

$$\begin{aligned}
\tilde{\tau} &= \tilde{n}_1' \cdot \tilde{n}_2 = -\tilde{n}_2' \cdot \tilde{n}_1 \\
&= \tau + (\psi' - \kappa_2\theta_1 + \kappa_1\theta_2) + \left(-\theta_1\theta_2' - \frac{1}{2}\tau(\theta_1^2 + \theta_2^2)\right) + \mathcal{O}(3) \\
\tilde{\kappa}_1 &= \tilde{n}_2' \cdot \tilde{t} \ = -\tilde{t}' \cdot \tilde{n}_2 \\
&= \kappa_1 + (\theta_1' - \tau\theta_2 + \kappa_2\psi) + \left(\psi\theta_2' + \tau\theta_1\psi - \frac{1}{2}\kappa_1(\theta_2^2 + \psi^2)\right) + \mathcal{O}(3) \\
\tilde{\kappa}_2 &= \tilde{t}' \cdot \tilde{n}_1 \ = -\tilde{n}_1' \cdot \tilde{t} \\
&= \kappa_2 + (\theta_2' + \tau\theta_1 - \kappa_1\psi) + \left(-\psi\theta_1' + \kappa_1\theta_1\theta_2 + \tau\theta_2\psi - \frac{1}{2}\kappa_2(\theta_1^2 + \psi^2)\right) + \mathcal{O}(3)
\end{aligned} \tag{2.14}$$

These deformation curvature components, which are expanded about the two equilibrium states $\vec{\kappa}^{o,\ell}$ are directly used to compute the Hamiltonian eq. (2.21). The intrinsic curvature components of the open state $\vec{\kappa}^o$ are not explicitly present in this form, as they have already been factored into the determination of the equilibrium curvature $\vec{\kappa}$.

2.2.4 Equilibrium Boundary Conditions

We use the elastic rod model developed by Goyal et al. [15] to model DNA and to determine the equilibrium curvature and torsion of the DNA loop state by specifying boundary conditions in terms of end-point location and orientations. The looped

state is treated as a clamped-clamped system, limiting the allowable deformations, see section 2 for more details. The open state configuration requires that there be no force or moment on the DNA end-points, which leads to straight DNA for wild-type sequences and a super-helical structure for the A-tract bends. These A-tract sequences induce intrinsic curvature $\vec{\kappa}^o \neq 0$ to the open state, see chapter 5 for more information.

2.2.5 Multiple Equilibrium States: Topology and Twist

For a given set of boundary conditions, there exist multiple equilibrium states, although not all are equally probable. There exist multiple binding topologies of DNA to LacI protein in the *E. coli* system which we explore in chapter 4. The binding topologies for LacI are inferred from the protein crystal structures by Lewis et al [26]. For boundary conditions, we specify locations and angles for the DNA; this includes an angle about the tangent vector, referred to as the phase angle. For any given topology, there two minimal energy equilibrium states, referred to as over- and under-twisted. The classification is as such,

$$\Delta T_w = \frac{1}{2\pi}\int_0^1 \mathrm{d}s\left(\tau^\ell(s) - \tau^o(s)\right), \text{where if } \begin{cases} \Delta T_w < 0, & \text{Under-twisted,} \\ \Delta T_w > 0, & \text{Over-twisted.} \end{cases} \tag{2.15}$$

Often only one solution contributes to the J factor as it is considerably more energetically favorable than the other state, for more informations, see fig. 4.1 in chapter 4.

2.3 Hamiltonian

We now use the deformation curvature vector $\tilde{\kappa}$ to construct a Hamiltonian which describes thermal fluctuations about equilibrium. We then expand the Hamiltonian in terms of the deformation angles $(\theta_1, \theta_2, \psi)$, keeping to 2^{nd} order, such that $H^{o,\ell} = H^{o,\ell}_{eq} + \Delta E$. The *in vivo* and *in vitro* environments are extremely viscous for DNA, which leads to over-damped thermal excitations of the DNA, for this reason we neglect the kinetic energy contributions to the Hamiltonian. The potential energy describes the work needed to deform the open state DNA into the looped state, as well as the change to the associated thermal fluctuations about each state. In addition electrostatic interactions arise due to the fact that DNA has a negatively

charged backbone, although under physiological conditions, screening occurs which limits these interactions to a distance of only a few nanometers [20], and we neglect them within this work; We discuss electrostatics again in chapter 6.

Now that we are able to compute the changes to curvature due to thermal fluctuations, we want to understand the energetic cost of these deformations caused by thermal fluctuations about the equilibrium state. Enthalpic contributions to the free energy are computed by determining the work needed to bend and twist the DNA from its reference zero strain energy open-state to the equilibrium looped-state. Entropic contributions to the free energy are computed by determining the change of the normal modes available to the open and looped states. Together these contributions to the free energy describe the relative stability and probability of loop formation of the DNA looped state. We employ two numeric techniques to solve for the eigenmode spectrum of the open and looped states.

All of our eigenvalues scale linearly with $\frac{\ell_p}{L}$, and are a function of their respective boundary conditions. We relate the average value of ℓ_τ to average value of ℓ_p, therefore we explicitly factor out the dimensional scaling as $\lambda \to \lambda\frac{\ell_p}{L}$ for clarity. The looped state Hamiltonian can be separated into two components as

$$\beta H^\ell = \overbrace{\frac{1}{2}\beta E_\ell}^{\text{Enthalpy}} + \overbrace{\frac{1}{2}\sum_{n=1}^{\infty}\lambda_n^\ell \frac{\ell_p}{L}\xi_n^2}^{\text{Entropy}\Rightarrow\Lambda} \tag{2.16}$$

where E_ℓ is the loop formation energy and λ_n is the n^{th} eigenvalue and ξ_n is the n^{th} eigenmode amplitude. The change of eigenmodes from the open to looped state contribute to the entropic coefficient Λ of the J factor, explained below. We divide the J factor by Avogadro's number $\mathcal{N}_A$ to make the units in terms of Molarity ($\mathcal{M}$), see Appendix A for more information. The J factor is then written as

$$J = \frac{8\pi^2}{\mathcal{N}_A}\frac{\int\prod_{i=1}^{N}\mathrm{d}\xi_i e^{-\beta H^\ell}\delta^3(|\vec{u}_L|)\delta(\theta_{1L})\delta(\theta_{2L})\delta(\psi_L)}{\int\prod_{i=1}^{N}\mathrm{d}\xi_i e^{-\beta H^o}} \tag{2.17}$$

where

$$u_{(1,2,3)L} \triangleq L\sum_{i=1}^{3N} u^i_{1,2,3}(1), \text{ and } u^i_{1,2,3}(1) = \int_0^1 \tilde{t}^i \mathrm{d}s \tag{2.18}$$

The J factor is the concentration of properly aligned DNA end points within a system.

The enthalpic terms contribute to the exponential in the J factor, while the entropic terms contribute to the pre-factor of the exponential of the J factor. We then write the J factor as

$$J = \underbrace{\Lambda}_{\substack{\text{Entropic} \\ \text{Contribution}}} \times \underbrace{e^{-\beta E_\ell - c_f}}_{\substack{\text{Enthalpic} \\ \text{Contribution}}} \tag{2.19}$$

where Λ includes all of the entropic contributions *i.e.* those from the change of thermal fluctuations between the looped and open state, we refer to Λ as the entropic coefficient. The loop formation energy, or enthalpic contributions to the J factor are given by βE_L. The remaining term $c_f \sim L/(4\ell_p)$ and is necessary force term introduced by Shimada and Yamakawa [49].

In principle the spectrum of eigenmodes is infinite, although we only need to consider a subset of this space up to the M^{th} mode, as after this mode, the open state and looped state modes yield the same contribution to the J factor, which we discuss in depth in section 2. We want the Hamiltonian to have this form so that we can explicitly compute eq. (2.1) in terms of Gaussian integrals as

$$\int \prod_{i=1}^{M} \mathrm{d}\xi_i e^{-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\xi_i^2} = \prod_{i=1}^{M} \sqrt{\frac{2\pi L}{\lambda_i \ell_p}} \tag{2.20}$$

This result is helpful as we be able to integrate over our eigenmode amplitudes ξ_i once we have solved for the Hamiltonian.

2.3.1 Constructing the Hamiltonian

We explicitly construct the Hamiltonian using the deformation curvature components that we have derived above. In practice we do this construction twice, once about the open state H^o and once about the closed state H^ℓ. The Hamiltonian is

$$\beta H = \frac{1}{2}\int_0^1 \mathrm{d}s \left(\tilde{\kappa}^{o,\ell}(s) - \vec{\kappa}^o(s)\right)^{\mathrm{T}} B(s) \left(\tilde{\kappa}^{o,\ell}(s) - \vec{\kappa}^o(s)\right) \tag{2.21}$$

where,

$$B(s) = \frac{1}{L}\begin{bmatrix} \ell_1(s) & \ell_{12}(s) & \ell_{1\tau}(s) \\ \ell_{12}(s) & \ell_2(s) & \ell_{2\tau}(s) \\ \ell_{1\tau}(s) & \ell_{2\tau}(s) & \ell_\tau(s) \end{bmatrix} \tag{2.22}$$

We have absorbed β and L into $B(s)$ so that its components are now dimensionless, see Appendix A for more information. In writing the stiffness tensor in this manner, we explicitly allow inhomogeneous bending and torsional stiffness as a function of DNA sequence, as well as twist-bend and bend-bend coupling. However we assume the off diagonals [44, 55] of $B(s)$ are small when compared to the diagonal elements, and we therefor treat them as zero throughout this work. We discuss non-homogeneous bending stiffness and twist-bend coupling $\ell_{2\tau}$ in chapter chapter 6. In essence we assume that the body-fixed frame coincides with the principal (eigen) directions of the stiffness tensor, see fig. 2.1. In general for wild-type and random sequences of DNA, we assume that there is no intrinsic curvature, $\vec{\kappa}^o = 0$, while for A-tracts we include sequence specific intrinsic curvature $\vec{\kappa}^0 = \vec{\kappa}^0(s) \neq 0$, see chapter 5 for more details.

Expanding in terms of our deformation variables θ_1, θ_2, ψ about the equilibrium, and keeping to 2^{nd} order we find

$$\begin{aligned}
\ell_1(\tilde{\kappa}_1 - \kappa_1^o)^2 =& \ell_1 \left(\Delta\kappa_1^2 + 2\Delta\kappa_1(\theta_1' - \tau\theta_2 + \kappa_2\psi) + \theta_1'^2 - 2\tau\theta_2\theta_1' + 2\kappa_2\psi\theta_1' + 2\Delta\kappa_1\psi\theta_2' \right. \\
&\left. + \theta_2^2(\tau^2 - \kappa_1\Delta\kappa_1) + \psi^2(\kappa_2^2 - \kappa_1\Delta\kappa_1) + 2\Delta\kappa_1\tau\theta_1\psi - 2\kappa_2\tau\theta_2\psi\right) \\
\ell_2(\tilde{\kappa}_2 - \kappa_2^o)^2 =& \ell_2 \left(\Delta\kappa_2^2 + 2\Delta\kappa_2(\theta_2' + \tau\theta_1 - \kappa_1\psi) + \theta_2'^2 + 2\tau\theta_1\theta_2' - 2\kappa_1\psi\theta_2' - 2\Delta\kappa_2\psi\theta_1' \right. \\
&+ \theta_1^2(\tau^2 - \kappa_2\Delta\kappa_2) + \psi^2(\kappa_1^2 - \kappa_2\Delta\kappa_2) + 2\kappa_1\Delta\kappa_2\theta_1\theta_2 - 2\kappa_1\tau\theta_1\ \psi \\
&\left. + 2\tau\Delta\kappa_2\theta_2\psi\right) \\
\ell_\tau(\tilde{\tau} - \tau^o)^2 =& \ell_\tau \left(\Delta\tau^2 + 2\Delta\tau(\psi' - \kappa_2\theta_1 + \kappa_1\theta_2) + \psi'^2 - 2\kappa_2\theta_1\psi' + 2\kappa_1\theta_2\psi' - 2\Delta\tau\theta_1\theta_2' \right. \\
&\left. + \theta_1^2(\kappa_2^2 - \tau\Delta\tau) + \theta_2^2(\kappa_1^2 - \tau\Delta\tau) - 2\kappa_1\kappa_2\theta_1\theta_2\right)
\end{aligned} \tag{2.23}$$

where $\Delta\kappa_{1,2} = \kappa_{1,2} - \kappa_{1,2}^o$ and $\Delta\tau = \tau - \tau^o$ is the difference between the spontaneous curvature of the intrinsic and equilibrium shapes. The equilibrium loop energy is then given by

$$\beta E_\ell = \frac{1}{2}\int_0^1 \mathrm{d}s \left(\frac{\ell_1}{L}\Delta\kappa_1^2 + \frac{\ell_2}{L}\Delta\kappa_2^2 + \frac{\ell_\tau}{L}\Delta\tau^2\right) \tag{2.24}$$

2.4 Solving for Normal Modes

In order to solve for the normal modes of the Hamiltonian, we employ two similar methods using known functions to approximate the solutions to the normal modes. The first method involves a straightforward variation of the Hamiltonian with respect to our deformation variables $(\theta_1, \theta_2, \psi)$, and leads to a system of coupled second-order differential-equations. These equations are not analytically solvable, so we must employ a numerical method to approximate their solution. The method we use is known as the Galerkin method and employs comparison functions that satisfy the boundary conditions, although they do not individually satisfy the differential equations. The second method uses the same comparison functions to directly construct the Hamiltonian and sidestep the second order problem, solving what is essentially a system of first order differential equations. The second method can only be used when our boundary conditions allow, as explained below. All of the normal mode contributions factor into the entropic coefficient Λ, which we break into components, indicated as Λ_x^y where x indicates which mode and constraints are being considered and $y = o, \ell$ indicates the open or looped states, respectively.

Galerkin Method and Comparison Functions

The Galerkin method involves using known functions, referred to as comparison functions, which satisfy the boundary conditions specified by the problem, although they do not satisfy the differential equations. The unknown solution to the system of differential equations is written in terms of the deformation variables, which are then expanded in terms of these comparison functions with arbitrary coefficients. When enough comparison functions are used, assuming the comparison function basis is complete, an approximate solution of coefficients can be found, which then combined with the comparison functions, satisfy the system of differential equations.

We collect our deformation variables within the Hamiltonian to construct a matrix representation in terms of deformation variables $(\theta_1, \theta_2, \psi)$. We then expand these variables in terms of known comparison functions, $\phi_n^{(a)}$ and unknown coefficients $\alpha_n^{(a)}$

as

$$\begin{bmatrix}\theta_1(s)\\ \theta_2(s)\\ \psi(s)\end{bmatrix} = \sum_{n=1}^{N} \begin{bmatrix}\alpha_n^{(1)}\phi_n^{(1)}(s)\\ \alpha_n^{(2)}\phi_n^{(2)}(s)\\ \alpha_n^{(3)}\phi_n^{(3)}(s)\end{bmatrix} = \boldsymbol{\phi}\boldsymbol{\alpha}$$

$$= \overbrace{\begin{bmatrix} \phi_1^{(1)} & \mathbf{0} & \mathbf{0} & \phi_2^{(1)} & \mathbf{0} & \mathbf{0} & \cdots & \phi_N^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \phi_1^{(2)} & \mathbf{0} & \mathbf{0} & \phi_2^{(2)} & \mathbf{0} & \cdots & \mathbf{0} & \phi_N^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \phi_1^{(3)} & \mathbf{0} & \mathbf{0} & \phi_2^{(3)} & \cdots & \mathbf{0} & \mathbf{0} & \phi_N^{(3)} \end{bmatrix}}^{\boldsymbol{\phi}} \overbrace{\begin{bmatrix}\alpha_1^{(1)}\\ \alpha_1^{(2)}\\ \alpha_1^{(3)}\\ \vdots\\ \alpha_N^{(1)}\\ \alpha_N^{(2)}\\ \alpha_N^{(3)}\end{bmatrix}}^{\boldsymbol{\alpha}} \tag{2.25}$$

where $a = 1, 2, 3$ and indexes $(\theta_1, \theta_2, \psi)$, and n indicates which basis function. This expansion allows different comparison functions for each deformations variable, allowing for different boundary conditions on the end points. We use the same expansion for the Linear Operators and Hamiltonian Construction.

Boundary Conditions

We treat the DNA looped state as a rod clamped at both ends, which we refer to as clamped-clamped. These boundary conditions mean that no displacement or angular deformations are allowed at the ends of the DNA. The boundary conditions are

$$\begin{aligned} \theta_1(0) = \theta_1(1) = 0, \quad u_1(0) = u_1(1) = 0 \\ \theta_2(0) = \theta_2(1) = 0, \quad u_2(0) = u_2(1) = 0 \\ \psi(0) = \psi(1) = 0, \quad u_3(0) = u_3(1) = 0. \end{aligned} \tag{2.26}$$

Modifications of these boundary conditions to allow for open-open geometries are explained below.

2.4.1 Galerkin Method: Linear Operators

We construct our Linear Operators by taking a variation $\boldsymbol{\delta}$ of our expanded Hamiltonian eq. (2.21) with respect to $(\theta_1, \theta_2, \psi)$ similar to those used by Lu et al [32]. We

start with

$$\boldsymbol{\delta}\left(\beta H^{o,\ell}\right) = \frac{1}{2}\int_0^1 \mathrm{d}s \left\{ \boldsymbol{\delta}(\ell_1(s)\left(\tilde{\kappa}_1^{o,l} - \kappa_1^o\right)^2) + \boldsymbol{\delta}(\ell_2(s)\left(\tilde{\kappa}_2^{o,l} - \kappa_2^o\right)^2) + \boldsymbol{\delta}(\ell_\tau(s)\left(\tilde{\tau}^{o,l} - \tau^o\right)^2)\right\} \tag{2.27}$$

The first order terms of $\left(\tilde{\kappa}^{o,\ell} - \vec{\kappa}^o\right)^2$ define the equilibrium curvature $\vec{\kappa}^{o,\ell}$ solution of the boundary-value problem

$$\left(\ell_1\Delta\kappa_1\right)' - \ell_2\tau\Delta\kappa_2 + \ell_\tau\kappa_2\Delta\tau = 0 \tag{2.28}$$

$$\left(\ell_2\Delta\kappa_2\right)' + \ell_1\tau\Delta\kappa_1 - \ell_\tau\kappa_1\Delta\tau = 0 \tag{2.29}$$

$$\left(\ell_\tau\Delta\tau\right)' + (\ell_2 - \ell_1)\kappa_1\kappa_2 + \ell_1\kappa_1^o\kappa_2 = 0 \tag{2.30}$$

where we have assumed there is no domain loading force along the DNA molecule, *e.g* the only forces and moments which determine the equilibrium loop state are applied at the DNA end-points.

If our stiffness tensor $B(s)$ is constant with respect to arc length s and we assume isotropic bending stiffness $\ell_1 = \ell_2$ in the absence of intrinsic curvature $\vec{\kappa}^i = 0$, then eq. (2.30) leads to $\tau' = 0$ or more clearly, the twist must be constant along the DNA molecule for all equilibrium solutions. We are interested in the 2^{nd} order terms as they contribute to the normal modes. Variation of the first deformation curvature component yields,

$$\begin{aligned}\boldsymbol{\delta}(\ell_1(\tilde{\kappa}_1 - \kappa_1^o)^2) =& 2\left(\ell_1\theta_1' - \ell_1\tau\theta_2 + \ell_1\kappa_2\psi\right)\boldsymbol{\delta}\theta_1' + 2\left(\ell_1\Delta\kappa_1\psi\right)\boldsymbol{\delta}\theta_2' \\ &+ \left[2\ell_1\tau\Delta\kappa_1\psi\right]\boldsymbol{\delta}\theta_1 \\ &+ \left[-2\ell_1\tau\theta_1' + 2\ell_1(\tau^2 - \kappa_1\Delta\kappa_1)\theta_2 + 2\ell_1\kappa_2\tau\psi\right]\boldsymbol{\delta}\theta_2 \\ &+ \left[2\ell_1\kappa_2\theta_1' + 2\ell_1\Delta\kappa_1\theta_2' + 2\ell_1\Delta\kappa_1\tau\theta_1 - 2\ell_1\kappa_2\tau\theta_2\right. \\ &\left.+2\ell_1(\kappa_2^2 - \kappa_1\Delta\kappa_1)\psi\right]\boldsymbol{\delta}\psi. \end{aligned} \tag{2.31}$$

Our deformation variables are θ_1, θ_2, ψ, so we need to remove the derivatives from $\boldsymbol{\delta}\theta_1', \boldsymbol{\delta}\theta_2'$. We integrate by parts these derivatives onto the the other terms in the

Hamiltonian using $\int u\mathrm{d}v = uv - \int v\mathrm{d}u$, which yields

$$\begin{aligned}
&\int_0^1 \mathrm{d}s \left\{2\left(\ell_1\theta_1' - \ell_1\tau\theta_2 + \ell_1\kappa_2\psi\right)\boldsymbol{\delta}\theta_1' + 2\left(\ell_1\Delta\kappa_1\psi\right)\boldsymbol{\delta}\theta_2'\right\} \\
&\quad = 2\left(\ell_1\theta_1' - \ell_1\tau\theta_2 + \ell_1\kappa_2\psi\right)\boldsymbol{\delta}\theta_1\big|_o^1 + 2\left(\ell_1\Delta\kappa_1\psi\right)\boldsymbol{\delta}\theta_2\big|_o^1 \\
&\qquad - \int_0^1 \mathrm{d}s \left\{2\left(\ell_1\theta_1' - \ell_1\tau\theta_2 + \ell_1\kappa_2\psi\right)'\boldsymbol{\delta}\theta_1 + 2\left(\ell_1\Delta\kappa_1\psi\right)'\boldsymbol{\delta}\theta_2\right\}
\end{aligned} \tag{2.32}$$

whereas we hold $\boldsymbol{\delta}\theta_1(0) = \boldsymbol{\delta}\theta_1(1) = \boldsymbol{\delta}\theta_2(0) = \boldsymbol{\delta}\theta_2(1) = \boldsymbol{\delta}\psi(0) = \boldsymbol{\delta}\psi(1) = 0$ at the end points as is normal with variational problems, so our boundary terms *e.g.* $2\left(\ell_1\Delta\kappa_1\psi\right)\boldsymbol{\delta}\theta_2\big|_o^1 = 0$. The full list of boundary terms appears in Appendix C. Then plugging eq. (2.32) back into eq. (2.31) we get

$$\begin{aligned}
&\boldsymbol{\delta}(\ell_1(\tilde{\kappa}_1 - \kappa_1^o)^2) \\
&\quad = \left[-2(\ell_1\theta_1')' + 2(\ell_1\tau\theta_2)' - 2(\ell_1\kappa_2\psi)' + 2\ell_1\tau\Delta\kappa_1\psi\right]\boldsymbol{\delta}\theta_1 \\
&\quad + \left[-2\ell_1\tau\theta_1' - 2(\ell_1\Delta\kappa_1\psi)' + 2\ell_1(\tau^2 - \kappa_1\Delta\kappa_1)\theta_2 + 2\ell_1\kappa_2\tau\psi\right]\boldsymbol{\delta}\theta_2 \\
&\quad + \left[2\ell_1\kappa_2\theta_1' + 2\ell_1\Delta\kappa_1\theta_2' + 2\ell_1\Delta\kappa_1\tau\theta_1 - 2\ell_1\kappa_2\tau\theta_2 + 2\ell_1(\kappa_2^2 - \kappa_1\Delta\kappa_1)\psi\right]\boldsymbol{\delta}\psi.
\end{aligned} \tag{2.33}$$

We then take variations of the other curvature components, and collecting variables in terms of $\boldsymbol{\delta}\theta_1, \boldsymbol{\delta}\theta_2, \boldsymbol{\delta}\psi$. The results is a system of three coupled differential equations,

$$\begin{aligned}
\left(\mathcal{L}_{11}\theta_1 + \mathcal{L}_{12}\theta_2 + \mathcal{L}_{13}\psi\right)\boldsymbol{\delta}\theta_1 &= 0 \\
\left(\mathcal{L}_{21}\theta_1 + \mathcal{L}_{22}\theta_2 + \mathcal{L}_{23}\psi\right)\boldsymbol{\delta}\theta_2 &= 0 \\
\left(\mathcal{L}_{31}\theta_1 + \mathcal{L}_{32}\theta_2 + \mathcal{L}_{33}\psi\right)\boldsymbol{\delta}\psi &= 0.
\end{aligned} \tag{2.34}$$

To illustrate the form of the linear operators $\mathcal{L}$, we explicitly write $\mathcal{L}$ here

$$\mathcal{L}_{11} = 2(-\ell_1)\frac{d^2}{\mathrm{d}s^2} + 2(-\ell_1')\frac{d}{\mathrm{d}s} + 2\ell_2(\tau^2 - \kappa_2\Delta\kappa_2) + \ell_\tau(\kappa_2^2 - \tau\Delta\tau))\mathbf{1}. \tag{2.35}$$

The remaining linear operators and all boundary terms are located in Appendix C. The above eq. (2.34) would generally be considered the equations of motion if we also considering kinetic and hydrodynamic friction terms. We are only interested in the effect of thermal fluctuations on the DNA, as we know due to over-dampening of the hydrodynamic forces, any kinetic contribution will quickly dissipate. We then consider the action of a thermal fluctuation on the elastic rod. Any fluctuation can be decomposed into the normal modes spectrum of the elastic rod, so we then set

eq. (2.34) equal to a generalized force λ times an angular displacement, in order to solve for the normal modes of the system.

$$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \mathcal{L}_{13} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \mathcal{L}_{23} \\ \mathcal{L}_{31} & \mathcal{L}_{32} & \mathcal{L}_{33} \end{bmatrix} \begin{bmatrix} \theta_1(s) \\ \theta_2(s) \\ \psi(s) \end{bmatrix} = \lambda \begin{bmatrix} \theta_1(s) \\ \theta_2(s) \\ \psi(s) \end{bmatrix} \tag{2.36}$$

where we now expand as in eq. (2.25), and left multiply by $\boldsymbol{\phi}^T$, and integrate over the arc length s, which results in the following eigenvalue equation,

$$\begin{aligned} \boldsymbol{\phi}^T \boldsymbol{\mathcal{L}}\, \boldsymbol{\phi}\, \boldsymbol{\alpha} &= \lambda\, \boldsymbol{\phi}^T \boldsymbol{\phi} \boldsymbol{\alpha} \\ \left[\int_o^1 \mathrm{d}s\ \boldsymbol{\phi}^T \boldsymbol{\mathcal{L}} \boldsymbol{\phi}\right] \boldsymbol{\alpha} &= \lambda \left[\int_o^1 \mathrm{d}s\ \boldsymbol{\phi}^T \boldsymbol{\phi}\right] \boldsymbol{\alpha} \\ \tilde{\boldsymbol{L}} \boldsymbol{\alpha} &= \lambda \tilde{\boldsymbol{M}} \boldsymbol{\alpha}, \end{aligned} \tag{2.37}$$

where $\tilde{\boldsymbol{L}} \triangleq \left[\int_o^1 \mathrm{d}s\ \boldsymbol{\phi}^T \boldsymbol{\mathcal{L}} \boldsymbol{\phi}\right]$ and $\tilde{\boldsymbol{M}} \triangleq \left[\int_o^1 \mathrm{d}s\ \boldsymbol{\phi}^T \boldsymbol{\phi}\right]$. If the basis functions are not orthonormal, then a generalized eigenvalue problem results as $\tilde{\boldsymbol{M}} \neq \mathbf{1}$. The form of $\tilde{\boldsymbol{M}}$ appears in Appendix C. Our choice of basis functions is orthonormal, therefore $\tilde{\boldsymbol{M}} = \mathbf{1}$, and therefore the eigenvalues and vectors of $\tilde{\boldsymbol{L}}$ are the same as those of our Hamiltonian. The structure of $\tilde{\boldsymbol{L}}$ is a block matrix by construction

$$\tilde{\boldsymbol{L}} = \begin{bmatrix} \boldsymbol{L}(1,1) & \boldsymbol{L}(1,2) & \cdots & \boldsymbol{L}(1,3N) \\ \boldsymbol{L}(2,1) & \boldsymbol{L}(2,2) & \cdots & \boldsymbol{L}(2,3N) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{L}(3N,1) & \boldsymbol{L}(3N,2) & \cdots & \boldsymbol{L}(3N,3N) \end{bmatrix} \tag{2.38}$$

where,

$$\boldsymbol{L}(m,n) = \int_0^1 \mathrm{d}s \begin{bmatrix} \phi_m^1 \mathcal{L}_{11} \phi_n^1 & \phi_m^1 \mathcal{L}_{12} \phi_n^2 & \phi_m^1 \mathcal{L}_{13} \phi_n^3 \\ \phi_m^2 \mathcal{L}_{21} \phi_n^1 & \phi_m^2 \mathcal{L}_{22} \phi_n^2 & \phi_m^2 \mathcal{L}_{23} \phi_n^3 \\ \phi_m^3 \mathcal{L}_{31} \phi_n^1 & \phi_m^3 \mathcal{L}_{32} \phi_n^2 & \phi_m^3 \mathcal{L}_{33} \phi_n^3 \end{bmatrix} \tag{2.39}$$

We then integrate each element over the arc length s and diagonalize the matrix to find the eigenvalues and eigenvectors. As we increase our expansion parameter N, we find that the eigenvalues converge from above. Our choice of uniform boundary conditions allows use of the same set of comparison functions for each of the clamped-clamped DNA deformations and we therefore drop the m index in eq. (2.25)

$$\phi_n(s) = \sqrt{2}\sin(n\pi s) \tag{2.40}$$

where n refers to one of the comparison functions within this basis. The comparison functions are normalized such that $\int_0^1 \mathrm{d}s\phi_m^2 = 1$. We are able to chose different functions as needed and a further refinement of this work would be to eliminate the boundary conditions and replace them with boundary potentials, allowing for more accurate DNA-protein interactions, see chapter 6.

Arbitrary Boundary Conditions

In determining the space of conformations available to the open state of DNA, we need to also consider arbitrary displacements and angular orientations of the DNA end points. For a given end-point force or moment due to a thermal fluctuation, the result is a displacement or angular orientation of the end point due to this boundary loading. The response to this boundary loading is a new equilibrium shape which has a vanishing net force and moment of the end points. We then use the Galerkin method to solve for the response function due to the boundary value loading, and integrate over the these positions and orientations in order to quantify the difference from an open to clamped end. It is always possible to reorient your coordinate system so that the displacement and angular orientation of one end-points is zero, therefore we only consider deformations to one end of the DNA. We consider a fluctuation ξ as

$$\xi = \xi^o + \xi^a \tag{2.41}$$

where ξ^o is any component of the fluctuation which leaves the end point angles unchanged, and ξ^a is the remaining component which leads to an arbitrary orientation and ultimately arbitrary location; we explore the connection of angle to location later in subsection 2. We then expand the fluctuation ξ in terms of appropriate comparison functions

$$\xi = \boldsymbol{\phi}\boldsymbol{\alpha}^{(o)} + \boldsymbol{\gamma}\boldsymbol{\alpha}^{(a)} \tag{2.42}$$

where ϕ is the same matrix of comparison functions introduced above and γ are new functions appropriate with the new desired boundary values; the superscripts (o) and (a) refer to no boundary angle deformation and arbitrary boundary angle

deformation. Then using the Galerkin method

$$\begin{aligned}\int \mathrm{d}s \phi^T \mathcal{L}\phi\alpha^{(o)} + \int \mathrm{d}s \phi^T \mathcal{L}\gamma\alpha^{(a)} &= 0 \\ \tilde{L}\alpha^{(o)} + \Gamma\alpha^{(a)} &= 0 \end{aligned} \tag{2.43}$$

where $\tilde{L} \triangleq \int \mathrm{d}s \phi^T \mathcal{L}\phi$ and $\Gamma \triangleq \int \mathrm{d}s \phi^T \mathcal{L}\gamma$. Rearrangement leads to

$$\alpha^{(a)} = -\Gamma^{-1}\tilde{L}\alpha^{(o)} \tag{2.44}$$

where Γ^{-1} is the inverse of matrix Γ. In many cases we chose $\gamma = \zeta s$, where ζ is the scaling term which we then integrate over, *e.g* in the case of torsional relaxation $\zeta = \psi_L^\ell$ and in the case of open state fluctuations $\zeta = \theta_{1L}^o, \theta_{2L}^o$ or ψ_L^o, where the subscript L indicates the displacement of the DNA endpoint located at $s = 1$.

2.4.2 Ritz Method: Hamiltonian Matrix Construction

Our second approach is numerically simpler; instead of constructing a set of linearized differential equations, we instead build the Hamiltonian Matrix H in terms of our comparison function basis. We then expand the Hamiltonian in terms of our basis functions and then reorganize the terms into a matrix form. This matrix is then diagonalized and its eigenmodes are used to compute the change of thermal fluctuations from the open to looped state, this technique is known as the Ritz method and is equivalent to the Galerkin Method, [10]. As we increase the number of basis function N, our eigenmodes converge to those which satisfy the boundary conditions and the differential equations. Neglecting the cross terms of $B(s)$, the Hamiltonian can be grouped by order of deformation variables as

$$\begin{aligned}\beta H =& \frac{1}{2}\int_0^1 \mathrm{d}s \left(\tilde{\kappa}^{o,\ell}(s) - \vec{\kappa}^o(s)\right)^{\mathrm{T}} B(s) \left(\tilde{\kappa}^{o,\ell}(s) - \vec{\kappa}^o(s)\right) \\ =& \beta H(\mathcal{O}(0)) + \beta H(\mathcal{O}(1)) + \beta H(\mathcal{O}(2)) \end{aligned} \tag{2.45}$$

where,

$$\begin{aligned}
\beta H(\mathcal{O}(0)) =& \frac{1}{2}\int_0^1 \mathrm{d}s\left(\ell_1\Delta\kappa_1^2+\ell_2\Delta\kappa_2^2+\ell_\tau\Delta\tau^2\right) && \text{(Loop Formation Energy)}\\
\beta H(\mathcal{O}(1)) =& \frac{1}{2}\int_0^1 \mathrm{d}s\left(2\ell_1\Delta\kappa_1(\theta_1'-\tau\theta_2+\kappa_2\psi)+2\ell_2\Delta\kappa_2(\theta_2'+\tau\theta_1-\kappa_1\psi)\right.\\
&\left.+2\ell_\tau\Delta\tau(\psi'-\kappa_2\theta_1+\kappa_1\theta_2)\right) && \text{(Equilibrium Terms)}\\
\beta H(\mathcal{O}(2)) =& \frac{1}{2}\int_0^1 \mathrm{d}s\left[\ell_1(\theta_1'^2-2\tau\theta_2\theta_1'+2\kappa_2\psi\theta_1'+2\Delta\kappa_1\psi\theta_2'+\theta_2^2(\tau^2-\kappa_1\Delta\kappa_1)\right.\\
&+\psi^2(\kappa_2^2-\kappa_1\Delta\kappa_1)+2\Delta\kappa_1\tau\theta_1\psi-2\kappa_2\tau\theta_2\psi)\big)\\
&+\ell_2(\theta_2'^2+2\tau\theta_1\theta_2'-2\kappa_1\psi\theta_2'-2\Delta\kappa_2\psi\theta_1'+\theta_1^2(\tau^2-\kappa_2\Delta\kappa_2)\\
&+\psi^2(\kappa_1^2-\kappa_2\Delta\kappa_2)+2\kappa_1\Delta\kappa_2\theta_1\theta_2-2\kappa_1\tau\theta_1\psi+2\tau\Delta\kappa_2\theta_2\psi)\big)\\
&+\ell_\tau(\psi'^2-2\kappa_2\theta_1\psi'+2\kappa_1\theta_2\psi'-2\Delta\tau\theta_1\theta_2'+\theta_1^2(\kappa_2^2-\tau\Delta\tau)\\
&\left.+\theta_2^2(\kappa_1^2-\tau\Delta\tau)-2\kappa_1\kappa_2\theta_1\theta_2)\right] && \text{(Normal Mode Terms)}
\end{aligned}$$

Where we refer to the first order terms as the equilibrium terms because their variation with respect to the deformation variables lead to our equilibrium boundary-value problem as in eq. (2.30). We then construct the Hamiltonian Matrix out of the normal mode terms from above.

$$\beta H = \frac{1}{2}\int_0^1 \mathrm{d}s\begin{bmatrix}\theta_1 & \theta_2 & \psi\end{bmatrix}\begin{bmatrix}\mathcal{H}_{11} & \mathcal{H}_{12} & \mathcal{H}_{13}\\ \mathcal{H}_{21} & \mathcal{H}_{22} & \mathcal{H}_{23}\\ \mathcal{H}_{31} & \mathcal{H}_{32} & \mathcal{H}_{33}\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \psi\end{bmatrix} \tag{2.46}$$

where the complete form of $\mathcal{H}$ is given in Appendix D. Then using the basis function expansion from section $\tilde{\mathcal{H}}$ is a $3N\times 3N$ matrix exactly in the manner as eq. (2.37).

$$\begin{aligned}
\beta H =& \frac{1}{2}\int_0^1 \mathrm{d}s\left(\frac{\ell_1}{L}\Delta\kappa_1^2+\frac{\ell_2}{L}\Delta\kappa_2^2+\frac{\ell_\tau}{L}\Delta\tau^2\right)+\frac{1}{2}\alpha^{\mathrm{T}}\left(\int_0^1 \phi^{\mathrm{T}}(s)\mathcal{H}(s)\phi(s)\mathrm{d}s\right)\alpha\\
=& E_\ell+\frac{1}{2}\sum_{i=1}^{N}\lambda_i\frac{\ell_p}{L}\xi_i^2
\end{aligned} \tag{2.47}$$

where E_ℓ is the cost of forming the loop as defined in eq. (2.24).

Hamiltonian Matrix Construction

The construction of the Hamiltonian Matrix H is straightforward for terms that do not contain derivatives on the deformation variables. In the basis of θ_1, θ_2, ψ we have

$$\begin{aligned}
&\theta_1^2\left(\ell_2(\tau^2-\kappa_2\Delta\kappa_2)+\ell_\tau(\kappa_2^2-\tau\Delta\tau)\right)\\
&+\theta_2^2\left(\ell_1(\tau^2-\kappa_1\Delta\kappa_1)+\ell_\tau(\kappa_1^2-\tau\Delta\tau)\right)\\
&+\psi^2\left(\ell_1\psi^2(\kappa_2^2-\kappa_1\Delta\kappa_1)+\ell_2(\kappa_1^2-\kappa_2\Delta\kappa_2)\right)\\
&\Leftrightarrow\\
&\begin{bmatrix}\theta_1 & \theta_2 & \psi\end{bmatrix}
\begin{bmatrix}
\ell_2(\tau^2-\kappa_2\Delta\kappa_2) & 0 & 0\\
+\ell_\tau(\kappa_2^2-\tau\Delta\tau) & & \\
0 & \ell_1(\tau^2-\kappa_1\Delta\kappa_1) & 0\\
 & +\ell_\tau(\kappa_1^2-\tau\Delta\tau) & \\
0 & 0 & \ell_2(\kappa_1^2-\kappa_2\Delta\kappa_2)\\
 & & +\ell_1(\kappa_2^2-\kappa_1\Delta\kappa_1)
\end{bmatrix}
\begin{bmatrix}\theta_1\\ \theta_2\\ \psi\end{bmatrix}
\end{aligned} \tag{2.48}$$

Then off-diagonal terms are constructed to be symmetric as such,

$$\begin{aligned}
&-2\ell_\tau\kappa_1\kappa_2\theta_1\theta_2\\
&-2\ell_2\kappa_1\tau\theta_1\psi \Leftrightarrow \begin{bmatrix}\theta_1 & \theta_2 & \psi\end{bmatrix}
\begin{bmatrix}
0 & -\ell_\tau\kappa_1\kappa_2 & -\ell_2\kappa_1\tau\\
-\ell_\tau\kappa_1\kappa_2 & 0 & -\ell_1\kappa_2\tau\\
-\ell_2\kappa_1\tau & -\ell_1\kappa_2\tau & 0
\end{bmatrix}
\begin{bmatrix}\theta_1\\ \theta_2\\ \psi\end{bmatrix}.\\
&-2\ell_1\kappa_2\tau\theta_2\psi
\end{aligned} \tag{2.49}$$

The terms containing derivatives, *e.g.* $-2\tau\theta_2\theta_1'$ need to be carefully constructed to be symmetric. Consider one of the contributions to H_{21}

$$-2\tau\theta_2\theta_1' \Leftrightarrow \begin{bmatrix}\theta_1 & \theta_2 & \psi\end{bmatrix}
\begin{bmatrix}
0 & 0 & 0\\
-2\tau\frac{d}{ds} & 0 & 0\\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}\theta_1\\ \theta_2\\ \psi\end{bmatrix} \tag{2.50}$$

The matrix could be made symmetric by integration by parts to move the derivative, although that construction would not necessarily be symmetric either. The solution is to instead create a symmetric $\mathcal{H}'_{sym}$ in a straightforward manner,

$$\mathcal{H}'_{sym} \triangleq \frac{1}{2}(\mathcal{H}' + \mathcal{H}'^T) \tag{2.51}$$

which results in a symmetric matrix containing all the 2^{nd} order curvature terms that contain a single derivative on a deformation variable. We illustrate this construction

by writing

$$-2\tau\theta_2\theta_1' \Leftrightarrow \begin{bmatrix}\theta_1 & \theta_2 & \psi\end{bmatrix}\begin{bmatrix}0 & -\tau\frac{\overleftarrow{d}}{ds} & 0\\ -\tau\frac{\overrightarrow{d}}{ds} & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \psi\end{bmatrix} \tag{2.52}$$

where the direction of the arrow on $\frac{d}{ds}$ indicates the direction that the derivative acts. The remaining terms in the Hamiltonian are $(\theta_1')^2, (\theta_2')^2, (\psi')^2$) terms which are constructed the same was as in eq. (2.48), although the basis used is $\theta_1', \theta_2', \psi'$). The final form of the Hamiltonian is very similar to that of the linear operators in Equation 2.38

$$\tilde{\boldsymbol{H}} = \begin{bmatrix}\boldsymbol{H}(1,1) & \boldsymbol{H}(1,2) & \cdots & \boldsymbol{H}(1,3N)\\ \boldsymbol{H}(2,1) & \boldsymbol{H}(2,2) & \cdots & \boldsymbol{H}(2,3N)\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{H}(3N,1) & \boldsymbol{H}(3N,2) & \cdots & \boldsymbol{H}(3N,3N)\end{bmatrix} \tag{2.53}$$

where,

$$\boldsymbol{H}(m,n) = \int_0^1 \mathrm{d}s \begin{bmatrix}\phi_m H_{11}\phi_n & \phi_m H_{12}\phi_n & \phi_m H_{13}\phi_n\\ \phi_m H_{21}\phi_n & \phi_m H_{22}\phi_n & \phi_m H_{23}\phi_n\\ \phi_m H_{31}\phi_n & \phi_m H_{32}\phi_n & \phi_m H_{33}\phi_n\end{bmatrix} \tag{2.54}$$

we have again used the fact that our boundary conditions are the same for each deformation variable, when they differ we use the linear operator methods. Additional details of the Hamiltonian Matrix construction are shown in Appendix D.

2.4.3 Enforcing Constraints on the Looped State Normal Modes with Delta Functions.

Our choice of comparison functions obey the angular constraints $\delta(\theta_1), \delta(\theta_2), \delta(\psi)$, however they do not obey the end-point displacement constraints $\delta^3(|\vec{u}|)$. We therefore need to explicitly compute the contributions of the δ-functions. The i^{th} eigenmode

is given by

$$\vec{\xi}_i(s) = \xi_i \begin{bmatrix} \psi_i(s) \\ \theta_{1i}(s) \\ \theta_{2i}(s) \end{bmatrix} = \xi_i \sum_{n=1}^{N} \begin{bmatrix} \alpha_{0n}^{(i)} \\ \alpha_{1n}^{(i)} \\ a_{2n}^{(i)} \end{bmatrix} \phi_n(s) \tag{2.55}$$

where the coefficients α are the components of the numerically computed eigenmodes.

The endpoint displacement $\vec{u}_i$ scales with the eigenmode amplitude ξ_i, we then determine the end-point displacement due to this eigenmode by integrating over the entire length of DNA. The components of the displacement vector at the end point, given by $\vec{u}(L)$ for the i^{th} eigenmode are

$$\vec{u}_i(L) = L \int_o^1 \mathrm{d}s \; \tilde{t}^i(s) = \xi_i L \int_o^1 \mathrm{d}s \; \left(\theta_2^i(s)\hat{n}_1(s) - \theta_1^i(s)\hat{n}_2(s)\right) \tag{2.56}$$

using the relationship to deformed curvature components and basis vectors eq. (2.7). We evaluate the end point constraint in terms of $\hat{x}, \hat{y}, \hat{z}$ in the coordinate system of the stress-free DNA. We therefore express our basis vectors in terms of x, y, z components, the same coordinate system as the DNA position vector $\vec{R}$,

$$\vec{u}_i(L) = u_{1i}(L)\hat{x} + u_{2i}(L)\hat{y} + u_{3i}(L)\hat{z} = \xi_i L \left(v_{1i}(1)\hat{x} + v_{2i}(1)\hat{y} + v_{3i}(1)\hat{z}\right) \tag{2.57}$$

where,

$$\begin{aligned} v_{1i} &= \int_o^1 \mathrm{d}s \left(\theta_2^i(s)\hat{n}_{1x}(s) - \theta_1^i(s)\hat{n}_{2x}(s)\right)\hat{x} \\ v_{2i} &= \int_o^1 \mathrm{d}s \left(\theta_2^i(s)\hat{n}_{1y}(s) - \theta_1^i(s)\hat{n}_{2y}(s)\right)\hat{y} \\ v_{3i} &= \int_o^1 \mathrm{d}s \left(\theta_2^i(s)\hat{n}_{1z}(s) - \theta_1^i(s)\hat{n}_{2z}(s)\right)\hat{z} \end{aligned} \tag{2.58}$$

Conjugate Variable Method

Now the displacements components u_{ji} within the δ-functions can be written in terms of the eigenmode amplitudes ξ_i using a Fourier Transformation

$$\delta(u_1) = \delta\left(\sum_{i=1}^{N} u_{1i}(L)\right) = \frac{1}{2\pi} \int \mathrm{d}\mu_1 \prod_{i=1}^{N} \exp(-iL\mu_1 v_{1i}\xi_i) \tag{2.59}$$

then expanding all of displacement constraints we get

$$\delta(u_1(L))\delta(u_2(L))\delta(u_3(L)) = \int \frac{\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\mu_3}{(2\pi)^3}\prod_{i=1}^{N}\exp\left(-iL\xi_i\sum_{j=1}^{3}\mu_j v_{ji}\right) \tag{2.60}$$

We then use the expansion of the δ-functions in terms of the eigenmode amplitudes ξ_i to explicitly compute the looped state partition function Z_ℓ. We will write $\tilde{Z}_\ell$ to be the looped state partition function, modulo the enthalpic contributions $\exp\left(-\beta E_\ell - c_f\right)$

$$\begin{aligned}\tilde{Z}_\ell &= \int\prod_{i=1}^{\infty}\mathrm{d}\xi_i\exp\left(-\frac{1}{2}\lambda_i\left(\frac{\ell_p}{L}\right)\xi_i^2\right)\delta(u_1(L))\delta(u_2(L))\delta(u_3(L)) \\ &= \int\frac{\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\mu_3}{(2\pi)^3}\prod_{i=1}^{\infty}d\xi_i\exp\left(-\frac{1}{2}\lambda_i\left(\frac{\ell_p}{L}\right)\xi_i^2 - iL\xi_i\sum_{j=1}^{3}\mu_j v_{ij}\right). \end{aligned}\tag{2.61}$$

We now complete the square and shift the integration over ξ_i

$$\begin{aligned}\tilde{Z}_\ell &= \int\frac{\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\mu_3}{(2\pi)^3}\prod_{i=1}^{\infty}d\xi_i\exp\left(-\frac{1}{2}\lambda_i\left(\frac{\ell_p}{L}\right)\xi_i^2\right)\exp\left(-\frac{1}{2}\frac{L^3}{\ell_p\lambda_i}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2\right) \\ &= \left(\frac{2\pi L}{\ell_p}\right)^{N/2}\prod_{i=1}^{N}\frac{1}{\sqrt{\lambda_i}}\int\frac{\mathrm{d}\mu_1 d\mu_2 d\mu_3}{(2\pi)^3}\exp\left(-\frac{1}{2}\frac{L^3}{\ell_p\lambda_i}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2\right). \end{aligned}\tag{2.62}$$

We write the product over the eigenvalues in terms of the determinant of the Hamiltonian, $\det H = \prod_{i=1}^{N}\frac{1}{\sqrt{\lambda_i}}$. The remaining integration over $\mathrm{d}\mu_i$ is now computed by collecting the conjugate variables into a vector as $\mu \triangleq [\mu_1, \mu_2, \mu_3]$,

$$\frac{1}{\lambda_i}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2 = \frac{1}{\lambda_i}\begin{bmatrix}\mu_1\mu_2\mu_3\end{bmatrix}\begin{bmatrix} v_{1i}^2 & v_{1i}v_{2i} & v_{1i}v_{3i} \\ v_{1i}v_{2i} & v_{2i}^2 & v_{2i}v_{3i} \\ v_{1i}v_{3i} & v_{2i}v_{3i} & v_{3i}^2\end{bmatrix}\begin{bmatrix}\mu_1 \\ \mu_2 \\ \mu_3\end{bmatrix} \tag{2.63}$$

where we compute a constraint matrix V_i for every eigenmode ξ_i. We then define the constraint matrix V as the sum over individual eigenmode constraint matrices

$$V \triangleq \sum_{i}^{N}\frac{1}{\lambda_i}\begin{bmatrix} v_{1i}^2 & v_{1i}v_{2i} & v_{1i}v_{3i} \\ v_{1i}v_{2i} & v_{2i}^2 & v_{2i}v_{3i} \\ v_{1i}v_{3i} & v_{2i}v_{3i} & v_{3i}^2\end{bmatrix} \tag{2.64}$$

then computing the integration over the conjugate variables of eq. (2.62), we compute the looped state partition function

$$\begin{aligned}\tilde{Z}_\ell &= \int \prod_{i=1}^{\infty} \mathrm{d}\xi_i \exp\left(-\frac{1}{2}\lambda_i \left(\frac{\ell_p}{L}\right)\xi_i^2\right) \delta(u_1(L))\delta(u_2(L))\delta(u_3(L)) \\ &= \left(\frac{2\pi L}{\ell_p}\right)^{N/2} \left(\prod_{i=1}^{N} \frac{1}{\sqrt{\lambda_i}}\right) \int \frac{\mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3}{(2\pi)^3} \exp\left(-\frac{1}{2}\frac{L^3}{\ell_p}\vec{\mu}^T V \vec{\mu}\right) \\ &= \left(\frac{2\pi L}{\ell_p}\right)^{N/2} \frac{1}{\sqrt{\det H^\ell}} \frac{1}{(2\pi)^3} \left(\frac{\ell_p}{L^3}\right)^{3/2} \sqrt{\frac{(2\pi)^3}{\det V}}. \end{aligned} \tag{2.65}$$

The full details of the partition function calculation appears in Appendix E.

2.4.4 Contributions of the Looped State Normal Modes

Now that we have corrected for the δ functions, we examine other eigenmodes which present additional challenges when computing the J factor, as well as quantify the entropic contributions of the looped state eigenmodes Λ^ℓ

Zero Modes

The first special case we discuss is the zero-stiffness eigenmode of the ring. When instrinsically straight-DNA ($\vec{\kappa}^o$) is bent into a ring; a state with aligned end-point tangent vectors, an eigenmode with zero stiffness arises, known as the zero mode. As this eigenvalue approaches zero $\lambda_o \to 0$, the integral $\int \mathrm{d}\xi_o \exp(-\lambda_o \xi^2) \to \infty$ appears to diverge. However the J factor remains finite as the eigenmode causes the ring to rotate out of the plane and back onto itself, making the deformation periodic and hence the integral remains finite. The lowest eigenmode for planar shapes happens to be linear $\lambda_1 \propto \Theta$ near $\Theta = 0$; for this reason, we separate out the lowest looped eigenmode in the calculation in order to make the treatment of this mode clearer. The details of this computation appear in chapter 3.

The lowest mode contribution can be written as

$$\Lambda_1^\ell = \int \mathrm{d}\xi_1 \exp(-\frac{1}{2}\frac{\ell_p}{L}\lambda_1^\ell \xi_1^2) = \begin{cases} \sqrt{\frac{2\pi}{\lambda_1^\ell}\frac{L}{\ell_p}} & , \text{ In general} \\ 2\pi & , \text{ Zero mode} \end{cases} \tag{2.66}$$

where the details of the zero mode calculation appear in chapter 3. It is also important to note that the contributions of λ_1 to the delta function computations of V

are included with the higher order modes. These modes are very straightforwardly handled as Gaussian integrals

$$\begin{aligned}\Lambda^{\ell}_{2:N} =& \int \prod_{i=2}^{N} \mathrm{d}\xi_i \exp(-\frac{1}{2}\lambda_i^{\ell}\frac{\ell_p}{L}\xi_i^2)\delta^3(|\vec{u}(L)|)\delta(\theta_{1L})\delta(\theta_{2L}) \\ =& \sqrt{\left(\frac{\ell_p}{2\pi L^3}\right)^3 \frac{1}{\det V}} \prod_{i=2}^{N} \sqrt{\frac{2\pi}{\lambda_i^{\ell}}\frac{L}{\ell_p}}\end{aligned} \tag{2.67}$$

where we can also write $\det \mathcal{H}^{\ell} = \prod_{i=1}^{N} \lambda_i^{\ell}\frac{L}{\ell_p}$. Note that the determinant formally diverges, although as we see below, the ratio of open to looped determinants is finite.

Torsional Constraint Relaxation

Understanding the role of binding constraints on DNA loop formation is one of the primary goals of this work. To this end we relax the torsional binding constraint for nearly planar loops. This physical system allows the computation of an effective torsional persistence length ℓ_τ^* in a straightforward manner. We relax the torsional constraint by considering the effect of an arbitrary moment on one of the DNA endpoints, and then relate the angular displacement to the input moment using the Linear Operators introduced in the Galerkin Method.

The eigenmode spectrum can be split into two parts, those that have vanishing endpoint angle $\psi(1)$ and those that have an arbitrary angle and position ψ_L

$$\prod_{i=1}^{N} \lambda_i = \begin{cases} \psi(1) = 0, & \text{Clamped Modes} \\ \psi(1) = \psi_L, & \text{Unconstrained Mode} \end{cases} \tag{2.68}$$

The eigenmodes which obey the torsional constraint $\psi(1) = 0$ are included in eq. (2.67). In total, the angular alignment, referred to as phasing, can affect the J factor by several orders of magnitude. Relaxing the torsional constraint as described in subsection 2 yields the relationship between moment and angular displacement, which then allows us to quantify the contribution of the $\psi(1)$ constraint on the looped state. The effect of this constraint is twofold, it affects the Gaussian integral over ψ_L as well as contributes to the constraint matrix V. We report the Gaussian integral results

separately as

$$\Lambda^{\ell}_{\psi_L} = \int \mathrm{d}\psi_L \exp(-\frac{1}{2}\lambda^{\ell}_{\psi}\frac{\ell_\tau}{L}\psi_L^2) = \sqrt{\frac{2\pi}{\lambda^{\ell}_{\psi}}\frac{L}{\ell_\tau}} \tag{2.69}$$

as indicated in eq. (2.68). When the torsional constraint is relaxed, the J factor no longer scales as a function of ℓ_τ, which we discuss in section 2. See chapter 3 for more details on the effective torsional persistence length ℓ^*_τ.

2.4.5 Open State Normal Modes

The open state boundary conditions require that there are no forces or moments on the end-points. These higher order boundary conditions then allow there to be arbitrary positions $\vec{u}$ and angular orientations $(\psi, \theta_1, \theta_2)$ of the end-points. However, our choice of comparison functions limit the angular displacements of the end-points, which must be corrected for properly. We complete the mode spectrum by adding to it those modes which cause an angular displacement at the end points. The open state normal modes can be broken down into two groups,

Eigenmodes with End Points $\psi(1) = \theta_1(1) = \theta_2(1) = 0$

These modes are the solutions to the open boundary conditions where the modes have no endpoint angular deformations, and correspond to the modes we have found using our basis functions above in eq (2.40). They are expressed in a straightforward way as

$$\beta H^o = \frac{1}{2}\sum_{n=1}^{\infty}\lambda^o_n\frac{\ell_p}{L}\xi_n^2 \tag{2.70}$$

where there is no enthalpic term as the open state is treated as the ground state. The open state entropic contributions of the clamped modes are split up to facilitate comparison to the looped state as

$$\Lambda^o_1 = \int \mathrm{d}\xi_1 \exp(-\frac{1}{2}\lambda^o_1\frac{\ell_p}{L}\xi_1^2) = \sqrt{\frac{2\pi}{\lambda^o_1}\frac{L}{\ell_p}} \tag{2.71}$$

$$\Lambda^o_{2:N} = \int \prod_{i=2}^{N}\mathrm{d}\xi_i \exp(-\frac{1}{2}\lambda^o_i\frac{\ell_p}{L}\xi_i^2) = \prod_{i=2}^{N}\sqrt{\frac{2\pi}{\lambda^o_i}\frac{L}{\ell_p}} \tag{2.72}$$

where we can also write $\det \mathcal{H}^o = \prod_{i=1}^{N} \lambda_i^o \frac{L}{\ell_p}$. The determinant formally diverges, although as we see below, the ratio of open to looped determinants is finite.

Eigenmodes with End Point $\theta_1(1), \theta_2(1), \psi(1) =$ arbitrary

In the open state, there are no constraints on the DNA endpoints, so we must also consider arbitrary moments and forces on the endpoints. Using the Galerkin methods in subsection 2 we compute the effect of fluctuations which leave $\theta_1(1), \theta_2(1), \psi(1) \neq 0$. The entropic contributions simplify for an intrinsically straight rod as

$$\Lambda_{\psi}^{o} = \int \mathrm{d}\psi_L \exp(-\frac{1}{2}\frac{\ell_\tau}{L}\psi_L^2) = \sqrt{\frac{2\pi L}{\ell_\tau}} \tag{2.73}$$

$$\Lambda_{\theta_1}^{o} = \int \mathrm{d}\theta_{1L} \exp(-\frac{1}{2}\frac{\ell_p}{L}\theta_{1L}^2) = \sqrt{\frac{2\pi L}{\ell_p}} \tag{2.74}$$

$$\Lambda_{\theta_2}^{o} = \int \mathrm{d}\theta_{2L} \exp(-\frac{1}{2}\frac{\ell_p}{L}\theta_{2L}^2) = \sqrt{\frac{2\pi L}{\ell_p}} \tag{2.75}$$

although in the case of intrinsic curvature, as we see in chapter 5, the general form is a matrix determinant as we can not easily separate out the effects of forces and moments at the boundary.

2.5 Assembling the J factor Calculation

We now collect all of the enthalpic and entropic contributions to the J factor

$$J = \frac{8\pi^2}{\mathcal{N}_A} \frac{\Lambda_1^\ell}{\Lambda_1^o} \frac{\Lambda_{2:N}^\ell}{\Lambda_{2:N}^o} \frac{\Lambda_{\psi_L}^\ell}{\Lambda_{\psi_L}^o} \frac{1}{\Lambda_{\theta_{1L}}^o \Lambda_{\theta_{2L}}^o} e^{-\beta E_\ell - c_f}. \tag{2.76}$$

The enthalpic contributions to loop formation enter in the exponential as

$$\beta E_\ell = \frac{1}{2} \int_0^1 \mathrm{d}s \ \left(\vec{\kappa}(s) - \vec{\kappa}^o(s)\right)^T B(s) \left(\vec{\kappa}(s) - \vec{\kappa}^o(s)\right). \tag{Enthalpy}$$

The entropic contributions from the lowest eigenmodes from eq. (2.66) and eq. (2.71) enter as

$$\frac{\Lambda_1^\ell}{\Lambda_1^o} = \begin{cases} \sqrt{\frac{\lambda_1^o}{\lambda_1^\ell}}, & \text{In General} \\ \sqrt{\frac{\lambda_1^o}{2\pi}\frac{\ell_p}{L}}, & \text{Zero Mode.} \end{cases} \tag{2.77}$$

The entropic contributions of eq. (2.67) and eq. (2.72) , which are the modes which cause no angular displacement at the endpoints are given by

$$\frac{\Lambda_{2:N}^\ell}{\Lambda_{2:N}^o} = \frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^{9/2}\sqrt{\frac{1}{(2\pi)^3 \det V}}\prod_{i=2}^{N}\sqrt{\frac{\lambda_i^o}{\lambda_i^\ell}} \tag{2.78}$$

where we have purposely altered the length scaling terms to be in terms ℓ_p^3. Then the torsional constraint term of the looped mode in eq. (2.69) as well as the torsional constraint correction from the open state in eq. (2.73) yield

$$\frac{\Lambda_{\psi_L}^\ell}{\Lambda_{\psi_L}^o} = \begin{cases} \sqrt{\frac{\ell_\tau}{2\pi L}}, & \text{In general} \\ \sqrt{\frac{1}{\lambda_\psi^\ell}}, & \text{No } \psi_L \text{ constraint} \end{cases} \tag{2.79}$$

we use the form the J factor without a torsional constrain ψ_L to derive an effective torsional stiffness ℓ_τ^* in chapter 3. The final contributions to the J factor come from correcting for arbitrary displacement and orientations of $\theta_1(L)$ and $\theta_2(L)$ in eq. (2.74) and eq. (2.75), for the open state equilibrium due to thermal fluctuations

$$\frac{1}{\Lambda_{\theta_{1L}}^o \Lambda_{\theta_{2L}}^o} = \frac{1}{2\pi}\frac{\ell_p}{L}. \tag{2.80}$$

Carefully combining all of the contributions to the J factor, we arrive at the following form for the **general case** J factor in terms of eigenvalues

$$J = \frac{1}{\mathcal{N}_A \ell_p^3}\sqrt{\frac{2}{\pi}}\left(\frac{\ell_p}{L}\right)^{11/2}\sqrt{\frac{\ell_\tau}{2\pi L}}\sqrt{\frac{1}{\det V}}\prod_{i=1}^{N}\sqrt{\frac{\lambda_i^o}{\lambda_i^\ell}}e^{-\beta E_\ell - c_f}. \tag{2.81}$$

Then writing the **general case** J factor in terms of determinants

$$J = \frac{1}{\mathcal{N}_A \ell_p^3}\sqrt{\frac{2}{\pi}}\left(\frac{\ell_p}{L}\right)^{11/2}\sqrt{\frac{\ell_\tau}{2\pi L}}\sqrt{\frac{1}{\det V}\frac{\det \mathcal{H}^o}{\det \mathcal{H}^\ell}}e^{-\beta E_\ell - c_f}. \tag{2.82}$$

Note that when bending stiffness is non-isotropic, the overall length scaling $(\ell_p/L)^{11/2}$ remains the same. The units of the J factor are inverse volume, it is customary to convert this to molarity $\mathcal{M}$. Then for the an equilibrium configuration with a zero mode *e.g.* the ring

$$J = \frac{1}{\mathcal{N}_A \pi} \frac{1}{\ell_p^3} \left(\frac{\ell_p}{L}\right)^6 \sqrt{\frac{\ell_\tau}{2\pi L}} \sqrt{\frac{\lambda_1^o}{\det V}} \prod_{i=2}^{N} \sqrt{\frac{\lambda_i^o}{\lambda_i^\ell}} e^{-\beta E_\ell - c_f} \qquad \text{(Torsionally Fixed Ring)}$$

when we relax the torsional constraint subsection 2, we then recover the the J factor for the torsionally free ring

$$\begin{aligned} J =& \frac{1}{\mathcal{N}_A \pi} \frac{1}{\ell_p^3} \left(\frac{\ell_p}{L}\right)^6 \sqrt{\frac{\lambda_1^o}{\lambda_\psi^\ell \det V}} \prod_{i=2}^{N} \sqrt{\frac{\lambda_i^o}{\lambda_i^\ell}} e^{-\beta E_\ell - c_f} \qquad \text{(Torsionally Free Ring)} \\ =& 32\pi^3 \frac{1}{\mathcal{N}_A \ell_p^3} \left(\frac{\ell_p}{L}\right)^6 e^{-2\pi^2 \frac{\ell_p}{L} - \frac{L}{4\ell_p}} \end{aligned} \tag{2.83}$$

as found by Shimada and Yamakawa [49]. The J factor also allows us to determine the free energy cost of DNA loop formation, see chapter 4, as well as the ratio of looping lifetimes, see chapter 5 for more information.

Truncation of Determinants

While the product of eigenvalues in eq. (2.67) and eq. (2.72) do not converge, the ratio of determinants of the Hamiltonians for the open and looped state are finite. In the limit of large N, the differences between the open and looped eigenmodes vanish, as the spatial resolution of the eigenmodes is below the curvature of the states, and only sees a straight segment. This fact allows us to truncate the ratio of modes after the m^{th} eigenmode as $\lambda_{m+1}^o/\lambda_{m+1}^\ell \to 1$. In practice, $m \sim 20$, which for DNA of the order of ℓ_p corresponds to spatial node distances of 7.5 base pairs, which can safely be regarded as continuous. The ratio of eigenvalues is

$$\frac{\det \mathcal{H}^o}{\det \mathcal{H}^\ell} = \frac{\lambda_1^o \cdots \lambda_m^o}{\lambda_1^o \cdots \lambda_m^o} \frac{\lambda_{m+1}^o \cdots \lambda_N^o}{\lambda_{m+1}^\ell \cdots \lambda_N^\ell} = \frac{\lambda_1^o \cdots \lambda_m^o}{\lambda_1^\ell \cdots \lambda_m^\ell}. \tag{2.84}$$

2.6 Scaling

Understanding scaling allows us to fully characterize the systems at hand. The eigenvalues scale as ℓ_p/L which accounts for the different scalings based on DNA conformation.

$$J = \Lambda(\Theta, \Phi, \Psi) \frac{1}{\mathcal{N}_A \ell_p^3} \left(\frac{\ell_p}{L}\right)^{11/2} \exp\left(-\frac{1}{2}\frac{\ell_p}{L}\int \mathrm{d}s\,(\vec{\kappa} - \vec{\kappa}^o)^2 - \frac{L}{4\ell_p}\right) \tag{2.85}$$

where angles Θ, Φ, Ψ represent arbitrary orientation angles of the DNA end-points, *e.g.* see fig. 3.2 of chapter 3. Shimada and Yamakawa [49] provided two fundamental examples, the ring and the averaged loop, which we discuss in depth in chapter 3. The torsionally unconstrained ψ_L = arb. ring consists of DNA without a binding protein. The DNA has aligned endpoint tangents, and the J factor is

$$J = 32\pi^3 \frac{1}{\mathcal{N}_A \ell_p^3} \left(\frac{\ell_p}{L}\right)^6 e^{-2\pi^2 \frac{\ell_p}{L} - \frac{L}{4\ell_p}} \tag{Ring}$$

$$\tag{2.86}$$

here we see that the scaling is $(\ell_p/L)^6$, as opposed to $(\ell_p/L)^{11/2}$ as presented in eq. (2.85) above. The reason for this deviation is the presence of the zero mode, and its associated entropic contribution presented in eq. (2.66). The next case is that of the unconstrained loop, in which all orientations of Θ such that the end-points are coincident, averaged over solid angle, see chapter 3 for more information. The J factor is

$$J_{\text{unc.}} = 110 \frac{1}{\mathcal{N}_A \ell_p^3} \left(\frac{\ell_p}{L}\right)^5 e^{-14.054 \frac{\ell_p}{L} - \frac{L}{4\ell_p}}. \tag{Unconstrained}$$

$$\tag{2.87}$$

again the scaling dimension differ from $(\ell_p/L)^6$, as the enthalpy is expressed as $E = E(\Theta) \propto \Theta^2$ and when this angle is integrated over, an additional factor of $\sqrt{L/\ell_p}$ is introduced. These results are shown in more detail in chapter 3.

Chapter 3

Planar Loops

3.1 Introduction

Calculating the probability that contact will occur between two distant ends of a polymer under prescribed orientations is a long-standing question of considerable significance in polymer physics. Yamakawa and Stockmayer expanded the Jacobson Stockmayer (J) factor [21] using the Kratky-Porod wormlike chain (WLC) to compute the angle-independent DNA ring-closure probabilities, otherwise known as cyclization of DNA [62]. Cyclization is the process of forming loops of DNA without using a DNA binding protein. Shimada and Yamakawa then included twist alignment of the end points [49], which we refer to as phase angle Ψ, to explain the oscillatory with respect to length, cyclization rates, measured by Shore and Baldwin on DNA shorter than 500 base pairs [50]. This phase dependence Ψ is based upon the DNA sequence length and average helical repeat of 10.5 base pairs. If the DNA end-points are not properly aligned when they come into contact, a large amount of twist will be required for binding, reducing the probability of cyclization.

Shimada and Yamakawa calculated the J factor for the ring and unconstrained loop averaged over solid angle [49], by treating DNA as a homo-polymer with coincident end points and parallel tangent vectors, as well as with coincident end points with unconstrained tangent vectors, respectively, see fig. 3.2. Our semi-analytic computation of the J factor generalizes this closure probability to include arbitrary end-point locations, binding orientations, sequence-dependent curvature and elasticity while reproducing the earlier results of Shimada and Yamakawa for the ring and unconstrained loop. As many biologically relevant cases do not fit neatly into one of these special cases, we generalize the computation of J factors to cover any nearly planar

shape, as well as to build our intuition for protein mediated DNA loop formation.

3.1.1 DNA Cyclization

Our work can be used to predict DNA cyclization rates based upon sequence dependent curvature and stiffness, as well as a function of DNA loop shapes. DNA cyclization begins by purposely placing unpaired bases at the ends of a segment of double stranded DNA, see fig. 3.1. These unpaired bases are referred to as overhangs and are chosen to be complimentary to each other *e.g.* one end may be $A-T-G-C$ and the other would be $G-C-A-T$. Then when thermal fluctuations bring the overhangs into contact with one another, the bases pair up and a loop of double stranded DNA is formed. This loop is however transient as they bases are only held together by Van der Waal interactions; the same fluctuations that form the loop are able to break it. To stabilize the loop, a protein called ligase is used to fuse together the sugar-phosphate backbone by forming a covalent bond. Although the details of this enzyme are not relevant to our work, it is important to note that the concentration of ligase must be carefully regulated [14, 13], or the measured J factor will be incorrect. Ligase can non-specifically bind to DNA, inducing local curvature, and if the concentration of ligase is too high, the reaction kinetics will also be disrupted, causing the measured J factor to be too large. If not enough ligase is present, then loops of DNA will form and break down before the ligase can stabilize them, causing the measured cyclization rate to be too low.

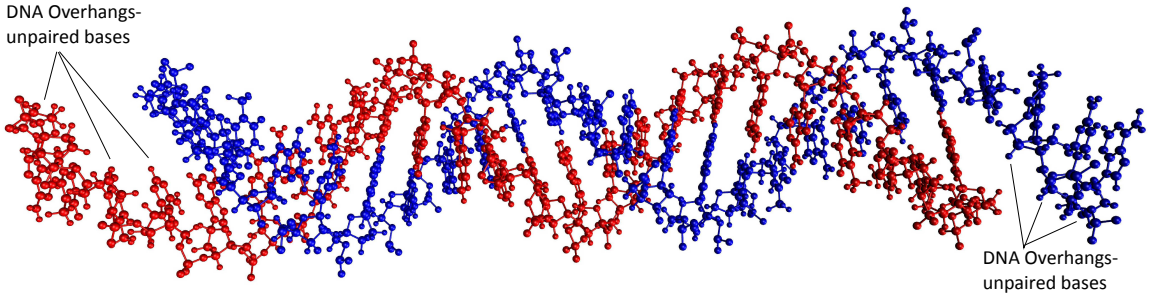

Figure 3.1 An example of unpaired DNA overhangs. The overhangs are chosen to be complementary to each other, such that $3'$ to $5'$ of one strand pairs up with $5'$ to $3'$ of the complementary overhang. This image was made using [19].

In experiments, the DNA overhangs must also be chosen carefully. Ideally they would consist of equal number of AT and GC pairs, as all GC will give an overestimate of the cyclization rate, and all AT will give an underestimate [14, 13, 8]. The

reason too many AT yield a lower J factor is that the binding of these base pairs is weak when compared to GC and it is likely that loops will break down before the ligase can repair the backbone. The reason too many GC yield a larger J factor is due to this stronger binding, as the experiment is intended to measure equilibrium rates, and depends on the loop break down being fast, which is not the case with all GC. The length of these overhangs is another critical component, as if they are too short, they will have a harder time staying stable until the ligase can bind the DNA. If the overhangs are too long, then the cyclization rate will be too high as the single stranded DNA will be able to explore a larger than intended volume by the cyclized sequence length. The J factor is measured through using Gel electrophoresis experiments to separate DNA based on its radius of gyration, and the relative percentages of cyclized to open state DNA is measured. The study of DNA cyclization is useful way to characterize the physical properties of DNA, such as phase angle dependence, sequence dependent stiffness and curvature. In this chapter we treat DNA as a homopolymer for the purposed of benchmarking the results of Shimada and Yamakawa [49]. In chapter 6 we explore planar results with $\ell_1 \neq \ell_2$.

3.1.2 Equilibrium States

In this chapter we study planar DNA loops in mechanical equilibrium with coincident end-points. The open state is intrinsically straight $\vec{\kappa}^0$ DNA with homogeneous bending stiffness $\ell_1 = \ell_2 = \ell_p$. The loops can by specifying three angle of relative orientation Θ, Φ, Ψ, which are defined as the angles between the end-point tangent vectors $\hat{t}(0)$ and $\hat{t}(1)$, as seen in fig. 3.2. We focus on planar loops ($\Phi = 0$) with clamped-clamped boundary conditions, whose curvatures can be analytically computed using elliptical integrals [24]. In practice, however since we consider all values of Θ, we instead use the rod model [15] to determine the equilibrium curvatures. In addition, we study these loops without the torsional constraint $\delta(\psi_L)$, free rotation of the overhangs is allowed even after the bases pair.

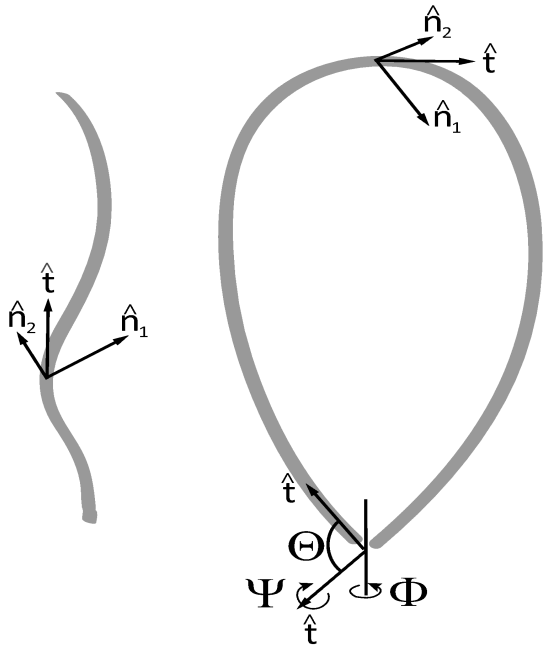

Figure 3.2 We specify the coincident end-point equilibrium loops with by three angles Θ, Φ, Ψ, which are functionally equivalent to Euler angles. Planar loops are a subset of these, given by $\Phi = \Psi = 0$. The left figure is an arbitrary open state with with boundary conditions requiring no force or moment on the end-points. A set of body-fixed coordinate vectors are displayed for reference. The right figure is the looped state with the same body-fixed coordinate vectors from the open state shown for reference. We consider more general loop shapes in chapter 4.

Loop Examples

The ring is defined by $\Theta = \Phi = 0$, where $\Psi = 0, \ldots, 2\pi$. We refer to Ψ as the phase angle and it should not be confused with ψ the phase angle deformation induced by a thermal fluctuations. The teardrop $\Theta \sim 0.54\pi \sim 97^{\text{o}}$ and $\Phi = 0$. The teardrop in interesting because it has vanishing end-point curvatures, so no moments are required to hold it in place. The teardrop is the minimum energy planar loop with $E_\ell/k_BT = 14.054\ \ell_p/L$, this results from vanishing end-point curvatures, which greatly reduce the looping energy. The final specific shape we note is the hairpin $\Theta = \pi,\ \phi = 0$. These three shapes are plotted below in fig. 3.3.

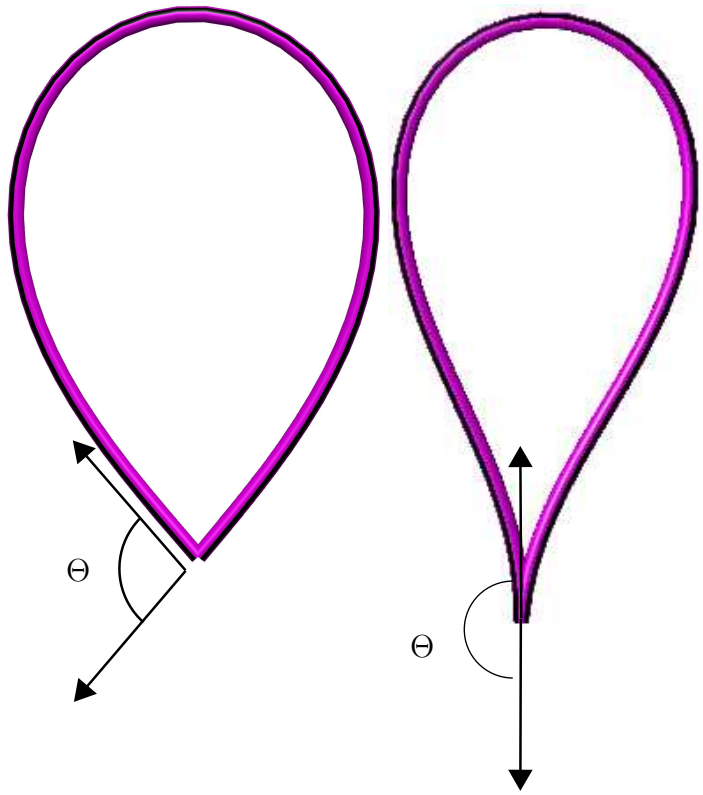

Figure 3.3 Two examples of planar loops, which are defined by $\Psi = \Phi = 0$ and arbitrary Θ. The loop on the left is the teardrop, which is defined by the vanishing curvature of the end-points. The loop on the right is the hairpin, defined by anti-parallel tangent vectors $\Theta = \pi$, the large curvature of the end-points make this loop difficult to form, see fig. 3.6.

3.2 Planar Hamiltonian

The equilibrium curvature $\vec{\kappa}^{\ell}(s)$ state of the ring is given by

$$\vec{\kappa}^{\ell}(s) = (2\pi \sin(\omega s/2), 2\pi \cos(\omega s/2), \omega) \tag{3.1}$$

where ω is the native twist of DNA defined by the helical repeat, which on average is 10.5 base pairs or $\sim 34.3°$ per base pair. However in this chapter we only consider DNA with homogeneous bending stiffness, which means that we can set $\omega = 0$ and instead of using the body-fixed coordinates, we can use the Frenet-Serret coordinates instead. In essence due to the lack of twist, the space-curve of the equilibrium shape is sufficient to compute the needed equilibrium curvatures $\vec{\kappa}^{\ell} = (0, \kappa_2(s), 0)$. We will examine many loop shapes from the ring, to the teardrop to the hairpin, see fig. 3.3. We can derive the planar equilibria and curvatures by considering a straight rod with

an axial force placed up on, as in [24], although in practice it is more convenient to use the rod model [15] to compute the range of loop formation angle $\Theta = 0$ the ring, to $\Theta = \pi$, the hairpin. A few examples of curvatures $\kappa_2^\ell(s)$ found using the rod model are present in fig. 3.4.

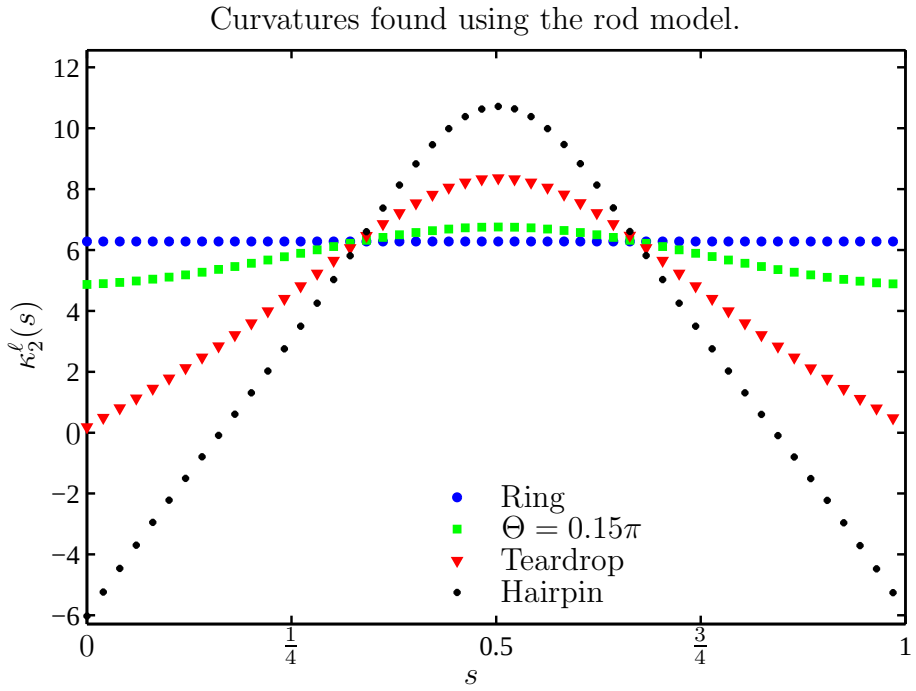

Figure 3.4 The planar loops without twist can be described with a single curvature component κ_2, which is computed using the rod model [15]. The curvature of the ring is constant $\kappa_2 = 2\pi \ \forall s$, while the teardrop is no constant with vanishing end-point curvature.

3.2.1 Deformation Curvatures

We allow deformations due to thermal fluctuations to the planar loops $\kappa_1 = \tau = 0$, which are formed from intrinsically straight DNA $\kappa_1^o = \kappa_2^o = \tau^o = 0$. The deformation Hamiltonian (2.21) simplifies to

$$\begin{aligned}\ell_1\tilde{\kappa}_1^2 + \ell_2\tilde{\kappa}_2^2 =& \ell_2\kappa_2^2 + 2\ell_2\kappa_2\theta_2' + \ell_1\theta_1'^2 + \ell_2\theta_2'^2 + 2(\ell_1 - \ell_2)\kappa_2\psi\theta_1' \\ &+ (\ell_1 - \ell_2)\kappa_2^2\psi^2 - \ell_2\kappa_2^2\theta_1^2. \qquad (3.2)\\ \ell_\tau\tilde{\tau}^2 =& \ell_\tau(\psi' - \kappa_2\theta_1)^2. \qquad (3.3)\end{aligned}$$

The assumption of homogeneous stiffness $\ell_1 = \ell_2 = \ell_p$, further reduces $\ell_1 \tilde{\kappa_1}^2 + \ell_2 \tilde{\kappa_2}^2$ to

$$\ell_p(\tilde{\kappa}_1^2 + \tilde{\kappa}_2^2) = \ell_p \kappa_2^2 + 2\ell_p \kappa_2 \theta_2' + \ell_p(\theta_1'^2 - \kappa_2^2 \theta_1^2) + \ell_p \theta_2'^2 \tag{3.4}$$

Then the planar Hamiltonian decouples into planar and out of plane deformations as

$$\begin{aligned}
\beta H(\mathcal{O}(0)) &= \frac{1}{2}\frac{\ell_p}{L}\int_0^1 \mathrm{d}s\ \kappa_2^2 && \text{(Loop Formation Energy)} \\
\beta H(\mathcal{O}(1)) &= \frac{\ell_p}{L}\int_0^1 \mathrm{d}s\ \kappa_2 \theta_2' && \text{(Equilibrium Term)} \\
\beta H(\mathcal{O}(2)) &= \underbrace{\frac{1}{2}\frac{\ell_p}{L}\int_0^1 \mathrm{d}s\ \theta_2'^2}_{\text{In Plane}} + \underbrace{\frac{1}{2}\frac{\ell_p}{L}\int_0^1 \mathrm{d}s\ \left(\theta_1'^2 - \kappa_2^2\theta_1^2\right) + \frac{1}{2}\frac{\ell_\tau}{L}\int_0^1 \mathrm{d}s\ \left(\psi' - \kappa_2\theta_1\right)^2}_{\text{Out of Plane}} && \text{(3.5)} \\
& && \text{(Normal Modes Terms)}
\end{aligned}$$

The planar Hamiltonian contributes the same eigenmodes as does the open state θ_2^o modes which do not cause angular displacement at the end points. The contribution of these modes do not strictly cancel out, as there are displacement constraints on the looped state modes *e.g.* the looped state modes contribute to the constraint matrix.

3.3 Normal Modes

The open and looped state eigenmodes which cause no angular displacement at the end-points are computed using the Hamiltonian Matrix method from subsection 2. The contribution of the open state modes which cause angular displacements of the end-point are found using the Linear Operator methods of subsection 2. The numerics of the ring reveal a zero eigenmode, which we will verify the analytically. The linear operators for the planar loop are

$$(\delta\theta_2) : \theta_2''(s) = 0 \tag{3.6}$$

$$(\delta\theta_1) : \theta_1''(s) + \kappa_2(s)\theta_1(s) + \frac{\ell_\tau}{\ell_p}\kappa_2(s)(\psi'(s) - \kappa_2(s)\theta_1(s)) = 0 \tag{3.7}$$

$$(\delta\psi) : (\psi'(s) - \kappa_2(s)\theta_1(s))' = 0. \tag{3.8}$$

We see from eq. (3.8) that $\psi'(s) - \kappa_2(s)\theta_1(s) = \phi = \text{constant}$. The torsional component of the out of plane Hamiltonian is then simplified as

$$\frac{1}{2}\frac{\ell_\tau}{L}\int_0^1 \mathrm{d}s\ (\psi' - \kappa_2\theta_1)^2 = \frac{1}{2}\frac{\ell_\tau}{L}\phi^2 \tag{3.9}$$

Then substituting $\psi'(s) - \kappa_2(s)\theta_1(s) = \phi$ into (3.7) we arrive at

$$\theta_1''(s) + \kappa_2^2(s)\theta_1(s) + \frac{\ell_\tau}{\ell_p}\kappa_2(s)\phi = 0$$

then for the ring $\kappa_2(s) = 2\pi$,

$$\theta_1''(s) + 4\pi^2\theta_1(s) + 2\pi\frac{\ell_\tau}{\ell_p}\phi = 0. \tag{3.10}$$

3.3.1 The zero eigenvalue normal mode of the ring.

The ring offers a distinct challenge when computing the normal modes of the loop, as it contains a zero eigenvalue. This eigenmode causes the entire ring to rotate about the clamped end-points and is consistent with the boundary constraints imposed by $\delta^3(\vec{u}), \delta(\theta_1), \delta(\theta_2)$. We remove the zero mode from the numeric computation and compute its contribution to the J factor by hand.

If we consider the minimal torsion solution $\phi = 0$ of eq. (3.10), subject the the boundary conditions $\theta_1(0) = \theta_1(1) = 0$ and $\psi_0 = 0$ and $\psi(1) = \psi_L$, we obtain the zero mode

$$\begin{aligned}\theta_1^{(1)} &= \xi_1 \sin(2\pi s)\\ \psi^{(1)} &= \xi_1 \left(1 - \cos(2\pi s)\right).\end{aligned} \tag{3.11}$$

By inspection we see that eq. (3.11) satisfies $\delta(\theta_1), \delta(\psi)$ as well as trivially satisfies $\delta(\theta_2)$. The end-point displacement $\vec{u}_1(1)$ of this mode

$$\vec{u}(1) = -L\int_0^1 \mathrm{d}s\theta_1(s)\hat{n}_2(s) = -\xi_1 L\int_0^1 \mathrm{d}s\sin(2\pi s)\ \hat{n}_2 = 0 \tag{3.12}$$

satisfies $\delta^3(\vec{u})$, and is therefore allowed by our boundary conditions and must be considered. We then substitute the zero mode into the deformation Hamiltonian eq. (3.5),

and find

$$\beta H = \frac{1}{2}(4\pi^2)\frac{\ell_p}{L}\int_0^1 \mathrm{d}s(\sin^2 2\pi s - \cos^2 2\pi s) + \frac{1}{2}\frac{\ell_\tau}{L}\phi = 0, \tag{3.13}$$

that this mode does not deform the equilibrium shape, or increase the energy of the ring, hence it is properly referred to as the zero (eigen) value mode.

Zero Mode Contribution to the J factor

When we examine the contribution of the zero mode to the entropic coefficient Λ_1^ℓ

$$\Lambda_1^\ell = \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \exp(-\frac{1}{2}\tilde{\lambda}_1\xi_1^2) = \int_{-\infty}^{\infty} \mathrm{d}\xi_1(1), \quad \text{diverges} \tag{3.14}$$

it appears that the J factor diverges, although a more carefully consideration of this mode reveals that it yields a finite contribution to the J factor.

Resolution of the zero mode

Consider now a clamped ring which lies within the xy plane as in fig. 3.5. The end-points which are clamped, are located at the origin. The displacement of any cross-section of the rod is then given by

$$\begin{aligned}\vec{u}(s) &= -\xi_1 L \int \mathrm{d}s \sin(2\pi s)\hat{n}_2 = \frac{\xi_1 L}{2\pi}\left(\cos 2\pi s - 1\right)\hat{n}_2 = -\frac{\xi_1 L}{2\pi}(\cos 2\theta + 1)\hat{n}_2 \\ &= -\frac{2\xi_1 L}{2\pi}\cos^2\theta\ \hat{z}\end{aligned} \tag{3.15}$$

where $\hat{n}_2 = \hat{z},\ \forall s$, since only $u_3 \neq 0$, we refer to the displacement as u_3. The distance to any cross section element from the y-axis is given by

$$x = \frac{2L}{2\pi}\cos^2\theta. \tag{3.16}$$

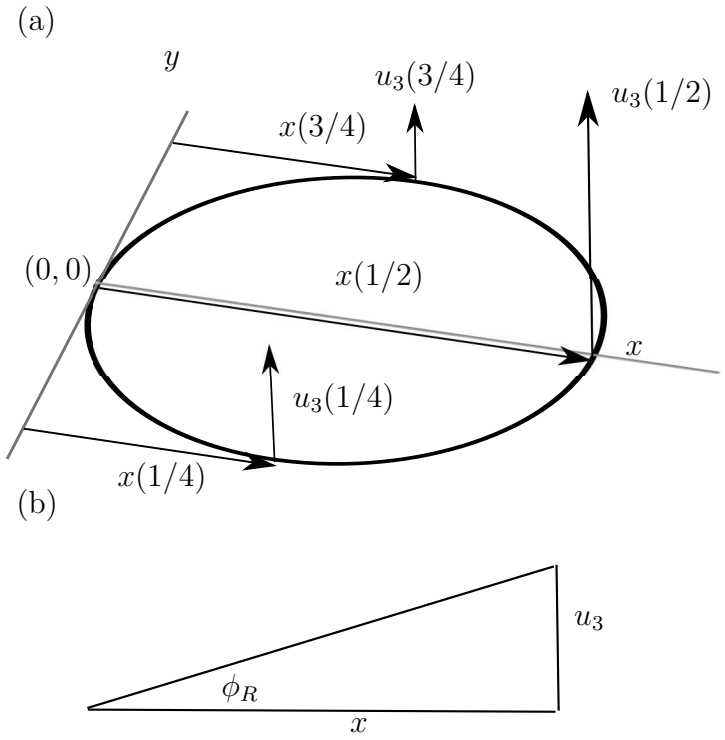

Figure 3.5 Zero Mode of ring - (a) The deformation is different at each point along the ring, depending on the distance from the y-axis. (b) The deformation cause the ring to rigidly rotate about the clamped end-points located at $(0,0)$.

Our choice of deformation variables occludes the fact that the zero mode is actually a rigid body rotation since the rotation angle ϕ_R is the same for all points off the y axis

$$\phi_R \triangleq \frac{u_3}{x} = \xi_1 \frac{-\frac{2L}{2\pi}\cos^2\theta}{\frac{2L}{2\pi}\cos^2\theta} = -\xi_1 = \text{constant } \forall\ s, \tag{3.17}$$

where we treat $u_3 \approx \phi_R x$, see fig. 3.5. We now see that ξ_1 can be related to the rotation angle ϕ_R, and is therefore periodic and bounded. Then returning to eq. (3.14), we see that Λ_1^ℓ is finite and yields

$$\Lambda_1^\ell = \int_{-\infty}^{\infty} \mathrm{d}\xi_1 \exp(-\frac{1}{2}\lambda_1 \frac{\ell_p}{L}\xi_1^2) = \int_{-\infty}^{2\pi} \mathrm{d}\phi_R = 2\pi \tag{3.18}$$

The zero mode influences the scaling dimensions of the J factor by not canceling out the corresponding $\sqrt{L/\ell_p}$ from the open state, resulting in a scaling factor of $\left(\frac{\ell_p}{L}\right)^6$ instead of $\left(\frac{\ell_p}{L}\right)^{11/2}$.

3.3.2 Relaxing the torsional constraint for the ring.

We will now use eq. (3.10) to relax the torsional constraint $\delta(\psi_L)$ for the planar loop and compute the torsionally unconstrained J factor. We must first however consider the case in which $\phi = 0$, with the torsional constraint $\psi_L = 0$, which we will see is the zero eigenvalue normal mode.

We relax the torsional constraint by removing $\delta(\psi_L)$ from the J factor calculation in order to compute an effective torsional persistence length ℓ_τ^*, the computation of ℓ_τ^* is shown in section 3. The removal of the torsional constraint requires that we integrate over any arbitrary orientations of ψ_L, as we did for the open state modes in subsection 2.

We now solve for the torsional displacement allowed by removing of $\delta(\psi_L)$. The solution of the differential equations of eq. (3.10) subject to the boundary conditions $\theta_1(0) = \theta_1(1) = 0$ and $\psi_0 = 0$ and $\psi(1) = \psi_L$ is

$$\begin{aligned}\theta_1 =& \frac{1}{2\pi}\frac{\ell_\tau}{\ell_p - \ell_\tau}\psi_L(\cos(2\pi s) - 1)\\ \psi =& \frac{1}{2\pi}\frac{\ell_\tau}{\ell_p - \ell_\tau}\psi_L \sin(2\pi s) + \psi_L s.\end{aligned} \tag{3.19}$$

This contribution of this solution to the Hamiltonian is

$$\int \mathrm{d}\psi_L \exp(-\beta H) = \int \mathrm{d}\psi_L \exp(-\frac{1}{2}\frac{\ell_p}{L}\frac{\ell_\tau}{\ell_p - \ell_\tau}\psi_L^2) = \sqrt{\frac{2\pi L}{\ell_p}\frac{\ell_p - \ell_\tau}{\ell_\tau}}. \tag{3.20}$$

At first glance this solution appears unusual, as in general $\ell_\tau > \ell_p$, and the result of (3.20) is therefore negative. This calculation is however incomplete, as this torsionally unconstrained solution also contributes to the $\delta^3(|\vec{u}|)$ constraints, which will result in an overall positive contribution. Then computing the contribution of the torsional mode to the constraint matrix we see that

$$\vec{u}_\tau = -L\int_0^1 \mathrm{d}s \left[\frac{1}{2\pi}\frac{\ell_\tau}{\ell_p - \ell_\tau}\psi_L(\cos(2\pi s) - 1)\right]\hat{z} = -\psi_L L\frac{1}{2\pi}\left(\frac{\ell_p}{\ell_p - \ell_\tau}\right) \tag{3.21}$$

where we compute the contribution to the total constraint matrix from this mode as

$$V_\tau = \frac{1}{\lambda_\tau}\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & v_3^2\end{bmatrix} = \frac{1}{(2\pi)^2}\frac{\ell_p}{(\ell_p - \ell_\tau)}\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{bmatrix} \tag{3.22}$$

The constraint matrix without the contribution of the this torsional mode is

$$V_{sub} = \begin{bmatrix} 0.0127 & 0 & 0 \\ 0 & 0.0127 & 0 \\ 0 & 0 & 0.0507 \end{bmatrix} \tag{3.23}$$

using the isotropic stiffness values of $\ell_\tau = 1.5\ell_p = 75$. Then the total constraint matrix is

$$V = \begin{bmatrix} 0.0127 & 0 & 0 \\ 0 & 0.0127 & 0 \\ 0 & 0 & -0.0253 \end{bmatrix} \tag{3.24}$$

which has a negative determinant. Then combining this result with eq. (3.20) to compute Λ, we obtain a real result for the J factor. The total J factor without the torsional constraint is independent of ℓ_τ for $\ell_\tau > \ell_p$. An instability exists for $\ell_\tau \leq \ell_p$, *e.g.* nicked DNA, which requires an alternative treatment of the problem.

3.4 Contributions to the J factor

The J factors for nearly planar loops are shown for a range of loop formation angles Θ. We compute the J factors with the torsional constraint $J_{(\psi_L=0)}$ and without the torsional constraint $J_{(\psi_L\neq 0)}$ in order to compute an effective torsional persistence length ℓ_τ^* as outlined in section 3. We say nearly planar, because as we will see, bending and twist become coupled due to the loop formation. This coupling affords easier out of plane thermal deformations, so our equilibrium shapes should be thought of as nearly planar loops.

3.4.1 Enthalpic Contributions

The loop formation energy as a function of angle Θ and its associated enthalpic contribution to the J factor is seen in fig. 3.6. The Teardrop shape has the lowest strain (enthalpic) energy, $14.4\frac{\ell_p}{L}\ k_BT$, of any of the in-plane shapes, and is where the endpoint curvatures of the loop vanish. The loop formation energy $E(\Theta)$ can be well fit to quadratic, yielding a Gaussian contribution to the J factor. The loop formation

energy is

$$\beta E(\Theta) = \frac{1}{2}\int \kappa_p^2 \mathrm{d}s = 2.02(\Theta - 0.54\pi)^2 + 14.054, \tag{3.25}$$

In order to recover the unconstrained J factor averaged over solid angle of Shimada and Yamakawa [49], we use eq. (3.25) in the limit of $\ell_p \to 0$. This limit is necessary because the J factor is integrated over loop formation angle Θ, which does not scale as a function of length.

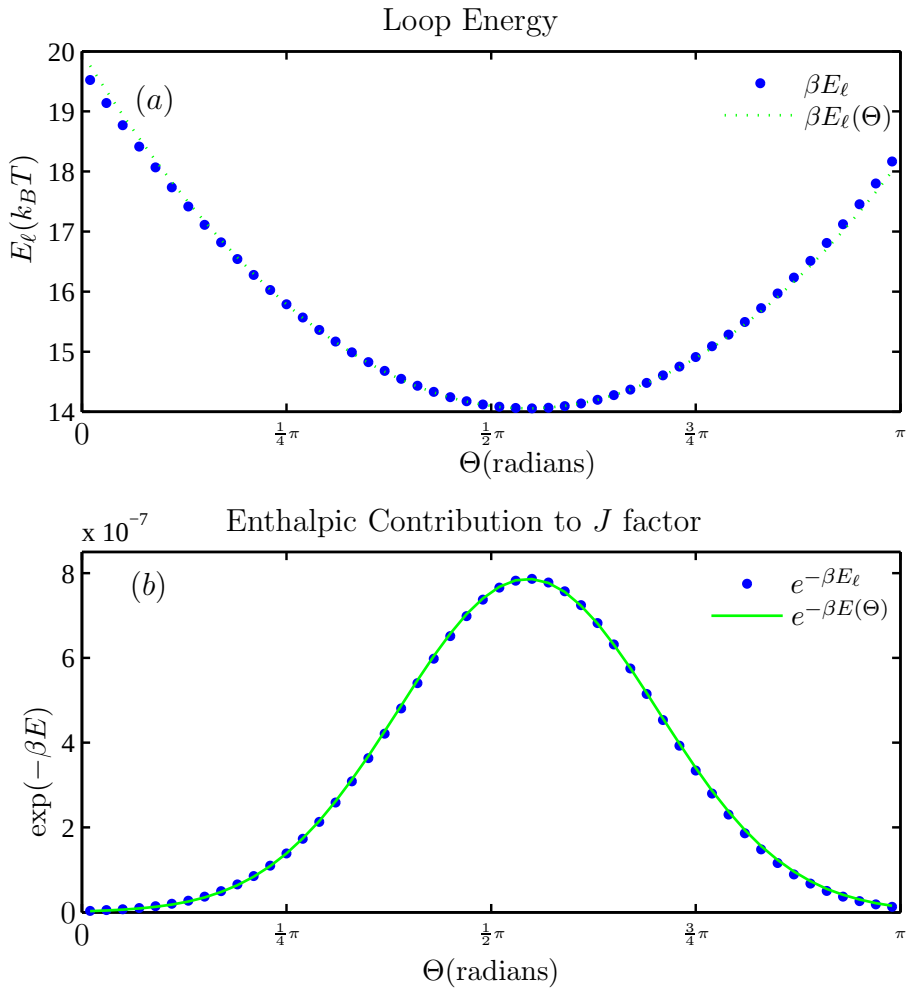

Figure 3.6 (a) The loop formation energy E_ℓ is strongly dependent on loop formation angle Θ. The rod model [15] computations are given by the solid circles, the numeric fit $\beta E_\ell(\Theta)$ is given by the dashed lines. (b) A demonstration that the quadratic fit of βE_ℓ is a good approximation. The enthalpic contributions vary strongly with loop formation energy. The length in (a) and (b) is one persistence length ℓ_p.

3.4.2 Entropic Coefficient

The entropic coefficient $\Lambda(\Theta)$ is computed for torsionally constrained and unconstrained DNA loops of sequence length $L = \ell_p$. The loop formation angle is swept out from the Ring $\Theta = 0$, to a Teardrop $\Theta \sim 0.54\pi$ and finally to the Hairpin $\Theta = \pi$, as seen in fig. 3.8. The lowest eigenvalue of the in-plane loops can be well approximated as $\lambda_1^{\ell} = 2\pi\Theta$ for relatively small angles Θ, see fig. 3.7.

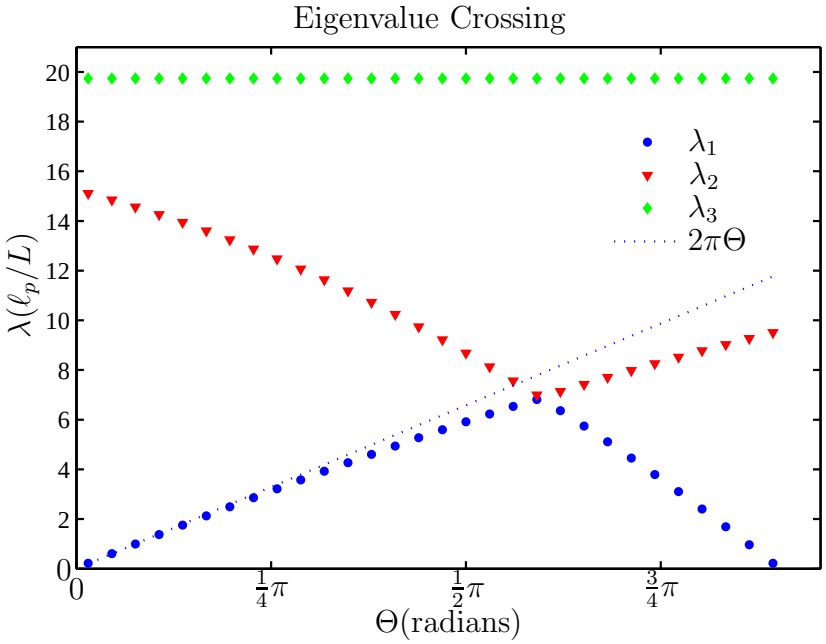

Figure 3.7 The three lowest looped eigenvalues as a function of Θ. The lowest mode λ_1 can be approximated using $2\pi\Theta$ for relatively small angles. The lowest mode at $\Theta = 0$ is symmetric, and the second lowest mode is antisymmetric, they swap positions where the eigenvalues meet, at which time the antisymmetric mode begins to dominate the J factor.

Factoring out the contribution of λ_1 from $\Lambda(\Theta)$ reveals a slowly varying component with respect to Θ, which we call $\gamma(\Theta)$, see fig. 3.8. The slowly varying component is

$$\gamma(\Theta) = 21.66\Theta^2 - 21.10\Theta + 147.82. \tag{3.26}$$

We will alter $\gamma(\Theta)$ when we numerically fit the J factor with analytic functions in section 3.

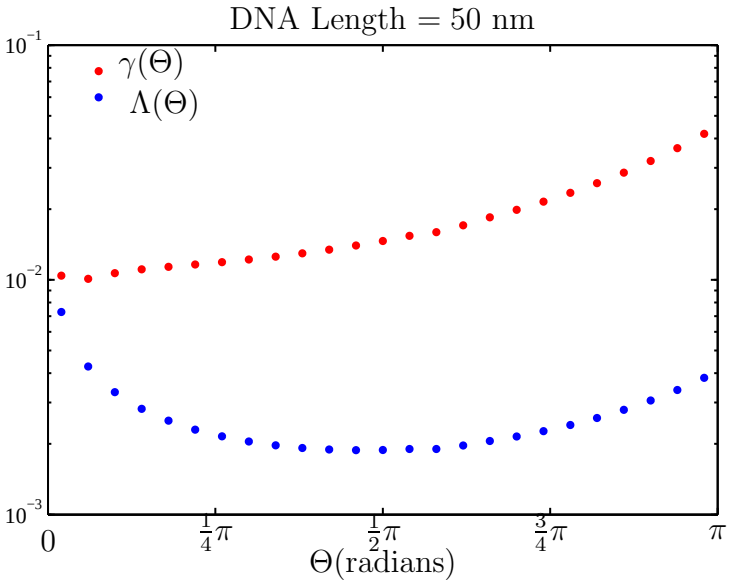

Figure 3.8 The entropic coefficient $\Lambda(\Theta)$ is largely dominated by contributions from the lowest eigenmode of the loop, $\lambda_1^\ell(\Theta)$. To illustrate this dependence, we write $\Lambda(\Theta) = f(\lambda_1^\ell(\Theta))\gamma(\Theta)$, where $f(\lambda_1^\ell)$ contains only the contributions of the lowest eigenmode λ_1^ℓ. The function $f(\lambda_1^\ell)$ is given by $\Lambda_{2:N}^\ell/\Lambda_{2:N}^o$, and $\gamma(\Theta)$ is given in eq. (3.26). It is then clear that $\gamma(\Theta)$ is a slowly varying function on the interval $(\Theta = 0)$ to $(\Theta = 0.54\pi)$, and then steadily increases on the interval from $\Theta = 0.54\pi$ to Hairpin $(\Theta = \pi)$. The shift in behavior of $\gamma(\Theta)$ occurs after the lowest eigenmode changes from symmetric to antisymmetric. Even for relatively short DNA $\Lambda(\Theta)$ is shown to effect the J factor by an order of magnitude in fig. 3.11. The units are in terms of Molarity, although this plot is not intended to represent a physical quantity, rather it shows the dependence of Λ and subsequently J to the lowest eigenmode.

3.4.3 The J factor dependence on Θ and Length

The J factor is strongly dependent on Θ as seen in fig. 3.9. The sensitivity near the ring is due largely to entropic effects Λ, while the teardrop and the hairpin have nearly the same entropic contribution, the enthapy dominates and results in a large difference in their J factors.

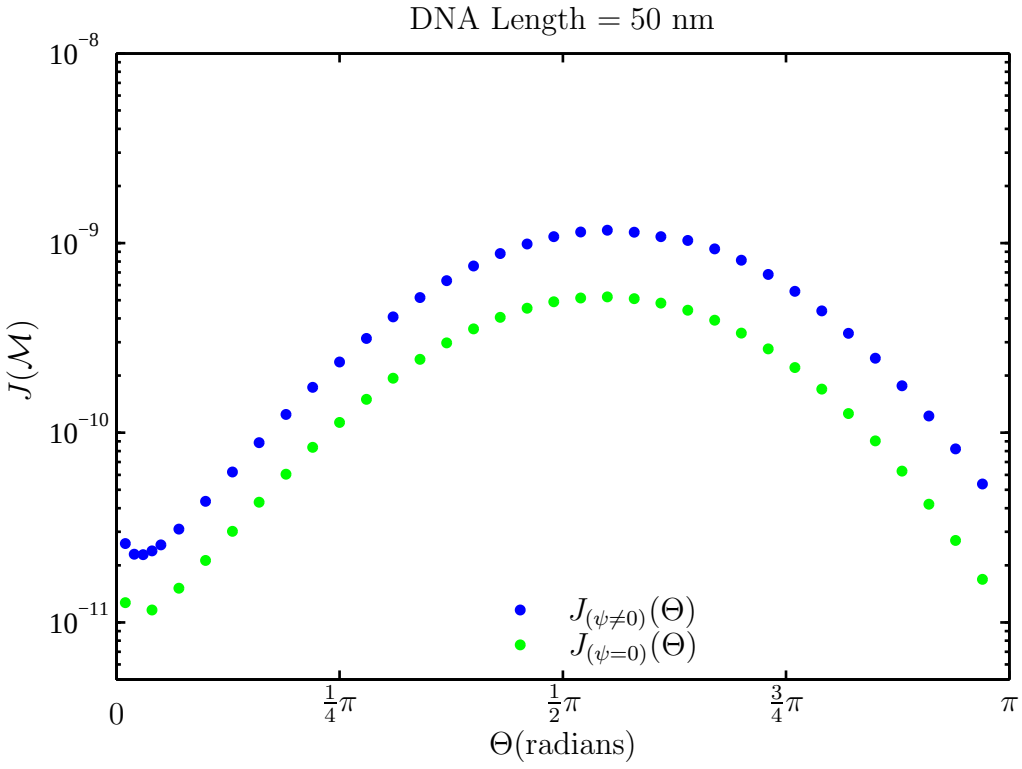

Figure 3.9 The J factor computed with the torsional constraint $J_{\psi_L=0}(\Theta)$ is shown by the darker circles, and without the torsional constraint $J_{(\psi\neq 0)}(\Theta)$ is shown by the lighter circles. The DNA length was chosen to be $L = \ell_p$ to sidestep scaling concerns and to clearly show the dependence of loop formation angle Θ.

The J factor for the ring, teardrop and hairpin as a functions of length are

$$J_{\text{ring}} = 32\pi^3 \frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^6 \exp\left(-2\pi^2 \frac{l_p}{L} - \frac{L}{4l_p}\right)$$

$$J_{\text{teardrop}} = 151 \frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^{11/2} \exp\left(-14.054 \frac{\ell_p}{L} - \frac{L}{4\ell_p}\right) \tag{3.27}$$

$$J_{\text{hairpin}} = 140 \frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^{11/2} \exp\left(-18.165 \frac{\ell_p}{L} - \frac{L}{4\ell_p}\right). \tag{3.28}$$

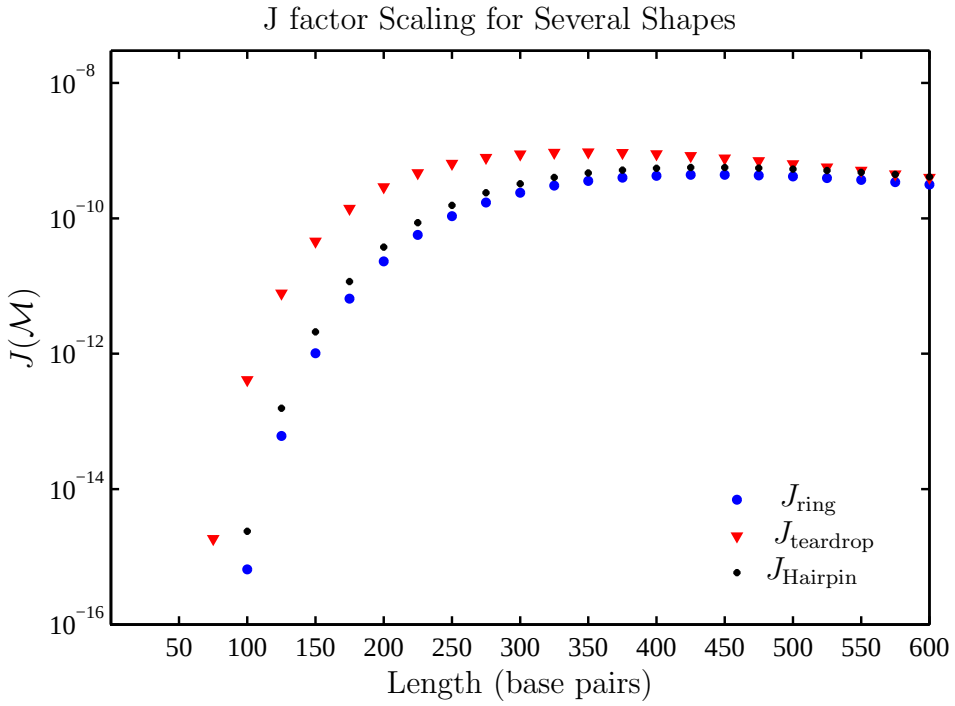

Figure 3.10 The J factor for three planar loop shapes as a function of length.

3.5 Demonstration of the Importance of Entropic Effects on Short Length Scales

It is clear that for short lengths of DNA, enthalpy is the major contribution to the J factor, so naturally one is inclined to consider whether or not variations of the entropic coefficient are even needed for these short lengths. We demonstrate that entropy is essential to computing the J factor for all lengths by plotting our numeric calculation of the J factor against two J factors created by only considering enthalpic variations due to Θ coupled with constant entropic coefficients . We use the entropic coefficient for the ring $\Lambda(\Theta_0)$, and that of the unconstrained loop averaged over solid angle $\langle\Lambda(\Theta)\rangle$. We then treat $\Lambda \neq \Lambda(\Theta)$, or more precisely, ignore any variations due to changing entropic contributions. The results are quite striking for DNA of $L = \ell_p$, as this DNA is quite stiff and enthalpy dominates the J factor, although it is unable to set the absolute scale of the looping probability, as seen in fig. 3.11. There are regions of $J = J(\Theta)$ where entropic variations due to Θ have a small effect, as shown by the teardrop region of $\Theta = 0.54\pi$, although near the ring, $\Theta = 0$, the variations of Λ are quite pronounced. In summary, the enthalpic contributions can tell us which

of two equilibrium loops is worth considering, although it can not provide us with a sense of scale in terms of end point concentrations of properly aligned end-points.

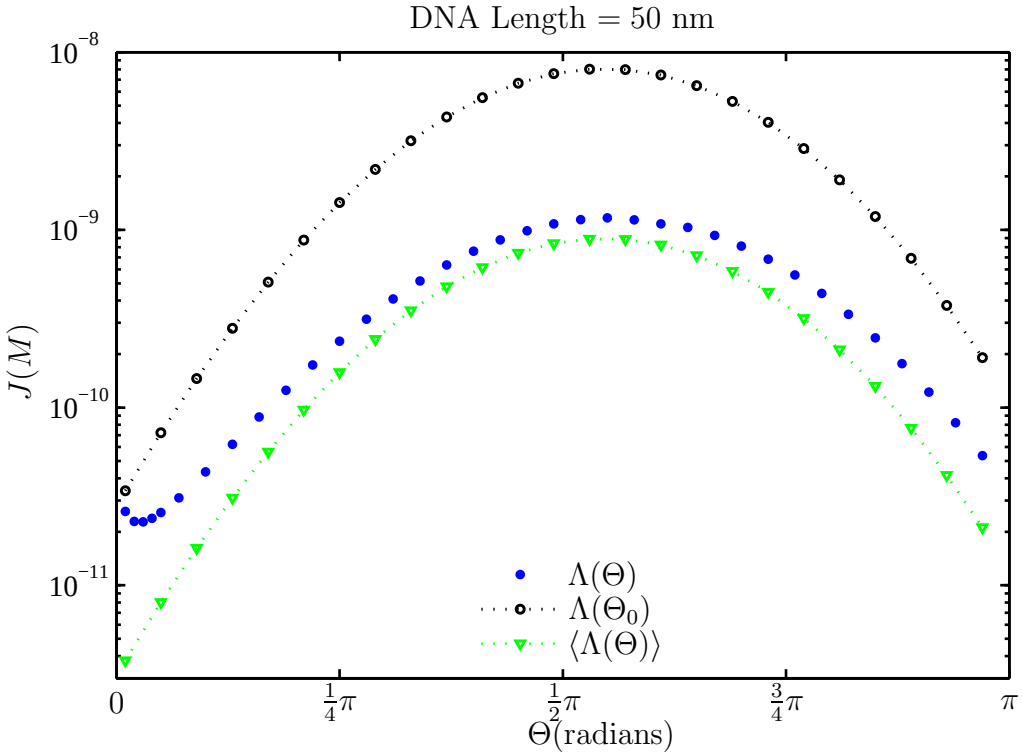

Figure 3.11 - A comparison of three J factors with the identical enthalpic contributions $\exp(-\beta E_\ell(\Theta) - c_f)$, and different entropic contributions Λ. The entropic coefficient computed for each value of Θ considered is represented by solid circles. We compare this result to that of an approximation using the ring Λ_0 and unconstrained $\langle\Lambda(\Theta)\rangle$ loop by assuming these contributions are constant for all values of Θ, given by the dotted circles and dotted triangles, respectively. In this way we demonstrate via several orders of magnitude difference that the entropic changes are vital to the calculation of the J factor. Small angles and Hairpin structures are poorly described by the enthalpic only unconstrained loop extrapolations of the J factor. These results are for DNA of length 50 nm and increase in difference as length is increased.

3.6 Effective Torsional Persistence Length

For planar loops we computed the J factor with the torsional constraint $J_{(\psi_L=0)}$ and without the torsional constraint $J_{(\psi_L\neq 0)}$. The main difference between these two results is the torsional stiffness to length scaling. The torsionally constrained rod has an additional scaling term $\sqrt{\ell_\tau/L}$, while $J_{(\psi_L\neq 0)}$ is independent of the torsional persistence length ℓ_τ, which we have confirmed numerically. We relaxed the torsional

constraint by considering arbitrary rotations of ψ_L in addition to the other normal modes. The change to the J factor is similar to that of the correction to the open state modes which cause angular displacement of the end-points. The J factors can then related by an effective torsional persistence length ℓ_τ^* as

$$J_{(\psi_L \neq 0)} = \int_{-\infty}^{\infty} \mathrm{d}\psi_L \exp(-\frac{1}{2}\lambda_\psi \frac{\ell_\tau}{L}\psi_L^2) J_{(\psi_L=0)} = \sqrt{\frac{2\pi}{\ell_\tau^*}} J_{\psi_L=0} \tag{3.29}$$

where the deformation of the rod cause by the angular displacement ψ_L is still constrained by $\delta(\theta_1), \delta(\theta_2), \delta^3(|\vec{u}|)$. It is also important to note that λ_ψ is not strictly an eigenvalue, as the differential equation constructed using the linear operators is a boundary value problem; it is however equivalent to the torsional stiffness of any arbitrary moment about the end-point tangent which results in an angular displacement ψ_L subject to the boundary constraints of the looped rod. We then solve for ℓ_τ as

$$\ell_\tau^* = 2\pi \left(\frac{J(\Theta)_{\psi=0}}{J(\Theta)_{\psi \neq 0}} \right)^2. \tag{3.30}$$

The effective torsional persistence length is in essence the torsional softening which occurs due to equilibrium loop shape leading to coupling of the bending and torsional degrees of freedom within the out of plane deformation Hamiltonian. The effective torsional persistence length can be written as a torsional and bending spring in series

$$\frac{1}{\ell_\tau^*} = \frac{1}{\ell_\tau} + \alpha(\Theta)\frac{1}{\ell_p}, \tag{3.31}$$

$$\ell_\tau^* = \ell_\tau \frac{1}{1 + \alpha(\Theta)\sigma} \tag{3.32}$$

where $\sigma \triangleq \ell_\tau/\ell_p$. The function $\alpha(\Theta)$ contains all of the angular dependence of ℓ_τ^*, which we plot ℓ_τ^* and $\alpha(\Theta)$ in fig. 3.12. As the geometry of the loop changes with angle Θ, the torsional stiffness ℓ_τ of the rod couples more strongly with the bending stiffness ℓ_p, which results in decrease of the effective torsional stiffness of the system. To illustrate this point, one may consider a straight rod that is bent into a ring of constant curvature, if one of the end points is twisted, the ring will initially remain planar. By contrast, if one considers twisting one end of a teardrop shape, the loop will immediately leave the plane, demonstrating the coupling of torsion and bending. The effective torsional persistence length represents the conversion between twist and writhe.

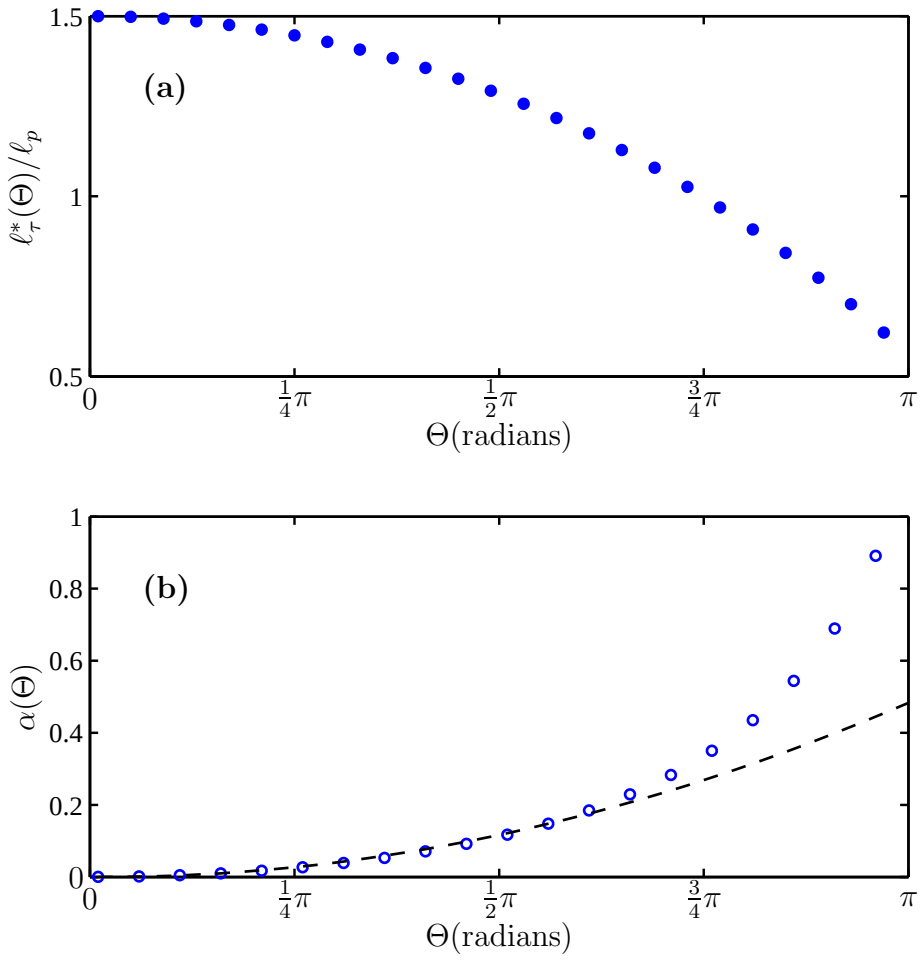

Figure 3.12 **(a)** The effective torsional persistence length, ℓ_τ^* in units of ℓ_p as a function of loop formation angle, Θ. The Ring has pure torsional modes with stiffness $\lambda_i \frac{\ell_\tau}{L}$ and as Θ increase the bending and torsional modes become coupled, reducing the effective torsional persistence length. **(b)** The torsion-bending coupling $\alpha(\Theta)$ shown as open circles, is quadratic from the Ring to the Teardrop as seen by the dashed line. From the Teardrop to the Hairpin, $\alpha(\Theta)$ is cubic in Θ.

In fig. 3.12, it is clear that $\alpha(\Theta)$ has as simple quadratic dependence up until the Teardrop shape and afterwards becomes cubic in Θ

$$\alpha(\Theta) = \frac{1}{2\pi^2}\Theta^2 - \frac{1}{6\pi^3}\Theta, \quad 0 \leq \Theta \leq 0.55\pi, \tag{3.33}$$

$$\alpha(\Theta) = 0.42\Theta^3 - 2.55\Theta^2 + 5.46\Theta - 3.87, 0.55\pi \leq \Theta \leq \pi \tag{3.34}$$

3.7 Numerical Fit for Planar J factors

In this section we will fit entropic coefficient $\Lambda(\Theta)$ to analytic functions for ease of use. The J factor for the planar loops can be written as

$$J(\Theta) = \Lambda(\Theta)\exp\left(-\frac{\ell_p}{L}\beta E(\Theta) - \frac{L}{4\ell_p}\right) \tag{3.35}$$

$$\tag{3.36}$$

where $\Lambda(\Theta)$ can be split into two pieces, one which dominates the variations of $\Lambda(\theta)$ and another slowly varying component $\gamma(\Theta)$

$$\Lambda(\Theta) = \underbrace{\left[\frac{\Lambda_1^\ell}{\Lambda_1^o}\right]}_{\substack{\text{Dominant variation} \\ \text{with } \Theta}} \times \underbrace{\left[\frac{\sqrt{2}}{\sqrt{\pi\lambda_\psi \det V}}\left(\prod_{i=2}^{N}\sqrt{\frac{\lambda_i^o}{\lambda_i^l}}\right)\left(\frac{\ell_p}{L}\right)^{11/2}\right]}_{\gamma(\Theta),\ \text{Slowly varying with } \Theta} \tag{3.37}$$

The dominant variations are due to the lowest looped eigenmode λ_1^ℓ, which suggests that we base our numeric fit on the influence of $\Lambda_1^\ell/\Lambda_1^o$. We begin by redefining ξ_1 to extract the scaling dimensions, $\xi_1 = \sqrt{\frac{L}{\ell_p}}\tilde{\xi}_1$, which leads to

$$\frac{\Lambda_1^\ell}{\Lambda_1^o} = \sqrt{\frac{\lambda_1^o\ell_p}{2\pi L}}\int_{-\infty}^{+\infty} d\xi_1 \exp(-\frac{1}{2}\frac{\ell_p}{L}\lambda_1^l\xi_1^2) = \sqrt{\frac{\lambda_1^o}{2\pi}}\int_{-\infty}^{+\infty} d\tilde{\xi}_1 \exp(-\frac{1}{2}\lambda_1^l\tilde{\xi}_1^2). \tag{3.38}$$

Now we consider $\lambda_1(1-\cos\xi_1)$ because when the normal mode amplitude is small, when compared to 1, $\lambda_1(1-\cos\xi_1) \sim \lambda_1\xi_1^2$. This substitution for the lowest eigenmode allows us to introduce a Bessel function into the numeric fit for J. The Hyperbolic Bessel Function, $I_0(z)$ is defined as

$$I_0(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}x \exp(\pm z\cos x). \tag{3.39}$$

Then recalling the zero mode change of limits of integration, we shift to $-\pi$ to π and substitute the Hyperbolic Bessel Function into eq. (3.38). The ratio of entropic coefficients is now

$$\begin{aligned}\frac{\Lambda_1^\ell}{\Lambda_1^o} &= \sqrt{\frac{\lambda_1^o}{2\pi}}\int_{-\infty}^{+\infty} \mathrm{d}\tilde{\xi}_1 \exp(-\frac{1}{2}\lambda_1^l\tilde{\xi}_1^2) \approx \sqrt{\frac{\lambda_1^o}{2\pi}}\int_{-\pi}^{\pi} \mathrm{d}\xi_1 \exp(\lambda_1^l(1-\cos\xi_1) \\ &= \sqrt{2\pi\lambda_1^o}\exp(-\lambda_1^l)I_0(\lambda_1^l) = \sqrt{2\pi\lambda_1^o}\exp(-2\pi\Theta)I_0(2\pi\Theta)\end{aligned} \tag{3.40}$$

It is important to keep in mind that there is no inherent meaning to using the Hyperbolic Bessel Function, it simply serves as a good fitting function, and these details are provided to motivate why it was used. We showed in fig. 3.7 that the lowest looped eigenmode could be approximated as $\lambda_1^\ell = 2\pi\Theta$ for small Θ. We will use this fit for then entire range of Θ in this numeric fit. The total entropic coefficient is then give by

$$\Lambda(\Theta) = I_0(2\pi\Theta)e^{-2\pi\Theta}\gamma(\Theta)\frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^{11/2} \tag{3.41}$$

where the slowly varying component $\gamma(\Theta)$ is now given by

$$\gamma(\Theta) = 365\Theta^2 - 525\Theta + 32\pi^3, \tag{3.42}$$

where the ordinate intersection has been chosen to align with the torsionally unconstrained ring. The slowly varying component is different from eq. (3.26) as it now contains the corrections of using the Hyperbolic Bessel Function, the ordinate intersection, as well as the lowest looped eigenvalue approximation. The total planar J factor is then given by

$$J(\Theta) = I_0(2\pi\Theta)e^{-2\pi\Theta}\gamma(\Theta)\frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^{11/2}\exp\left(-\frac{\ell_p}{L}\beta E(\Theta) - \frac{L}{4\ell_p}\right) \tag{3.43}$$

$$\beta E(\Theta) = \frac{1}{2}\int \kappa_p^2 \mathrm{d}s = 2.02(\Theta - 0.54\pi)^2 + 14.054, \tag{3.44}$$

where Θ is the loop tangent angle in radians and units of Molarity. This numeric fit is excellent, with a maximum error of less than 1%, for all Θ. The length scaling of the fit must be slight modified from $(\ell_p/L)^{11/2}$ to $(\ell_p/L)^6$ for the ring ($\Theta = 0$) due to the presence of the zero mode.

3.8 Comparison to Known Results

Shimada and Yamakawa provided two special cases for the in-plane J factors [49]: the Ring defined by aligned tangents, $\Theta = 0$, and the unconstrained loop. Zhang and Crothers [63] arrived at different results for these J factors using their base pair harmonic mode computations, although the reason for the difference remains unclear. We have reproduced the ring and unconstrained loop results to within 0.01% using our continuum rod model in agreement with Shimada and Yamakawa. The two models should not yield significantly differ results since the radius of curvature is much greater than the DNA bond length. We also compute nearly the same ring and loop energy as Zhang and Crothers, and they computation of normal modes should be comparable, although they do not mention the zero eigenvalue for the ring, which may indicate that they did not treat it properly. Finally Zhang and Crothers constructed their delta functions constraint matrix differently than our method, which could yield a different determinant. These factors may be the lead to the discrepancy in their results.

3.9 Conclusions

The results presented here are for near planar DNA loops with coincident end points and arbitrary loop tangent angle Θ. In this paper we present results for DNA of the length $50nm$ or approximately 14 helical repeats, so we will assume $\ell_1 = \ell_2$. We treat the DNA as a homogeneous polymer with bending and torsional persistence lengths of 50 nm and 75 nm, respectively.[52, 17, 5]. While DNA is a heteropolymer with

anisotropic bending persistence lengths $\ell_1(s) \neq \ell_2(s)$, this anisotropy largely averages out after a few helical repeats (~ 10.5 base pairs) as demonstrated by Kehrbaum and Maddocks [22].

We have developed a generalized approach for computing J factors of arbitrary loop shapes, which may include sequence-dependent stiffness and curvature and reproduces the results of Shimada and Yamakawa for the ring and unconstrained loop. We have shown that the J factor varies strongly for near planar loop shapes as a function of loop tangent angle Θ for intrinsically straight DNA with isotropic bending stiffness. The in-plane J factors can be well fit with analytic functions for all Θ. We have defined an effective torsional persistence length ℓ_τ^* and subsequence torsion-bending coupling $\alpha(\Theta)$ which are shown to vary significantly as a function of loop formation angle, Θ. Finally, our calculation is computationally very quick, taking only a few minutes per J factor for any set of input boundary conditions.

Chapter 4

LacI Induced Loops

4.1 Protein Mediated Loop Formation

DNA loop formation can occur in many ways, with multiple binding and architectural proteins, as well as several binding sites on the DNA molecule itself. We begin with the simplest protein mediated loop formation, which involves binding of two DNA operators to a single protein. The canonical example of this type of looping is the Lac Operon in *E. coli*, as there is only one binding protein known as the Lac Repressor, and three DNA binding sites, known as operators. Binding of the lac Repressor to any two of these sites will constitute the loops which we focus on within this chapter. In many ways, *E. coli* serves as the Hydrogen Atom for biophysics, as it is a relatively simple and very well characterized system. Experiments by Müeller-Hill, et al. [39] have demonstrated that the ability of DNA to form loops and repressor the LacZYA genes were affected strongly by small changes in the DNA length. This effect was semi-periodic which correlated with the 10.5 base pair helical repeat of DNA, which was similar to the phasing cyclization data of Shore et al [50]. Subsequent crystalization of the lac repressor and DNA operator segments into co-crystal by Lewis et al. [26], confirmed that the observed phasing was due to specific end-point orientation requirements for binding to the Lac Repressor. This phasing is due to the fact that every base base pair rotates the end-point orientation of DNA by about 34°.

4.1.1 Lac Repressor

The Lac Repressor is a homotetramer bound into two dimeric arms, adjoined at the C-terminus, which two DNA binding head groups, see fig. 4.1. The orientation of the

DNA at the operator regions is also highly specific, as the head groups of the LacI protein will only bind to properly oriented DNA, requiring us to consider the helical repeat of DNA, and determine the phase angle Ψ at the DNA end point relative to the protein. Several distinct loops are possible, which we refer to as binding or loop topologies since they can not be converted into another of these loop topology without first breaking the LacI-DNA binding. This general DNA-protein complex has spatially separated end-points as well as prescribed angles (Θ, Φ, Ψ) which are obtained from DNA-protein co-crystals with LacI protein serving as the canonical example [26].

4.2 Binding Topologies

We classify the DNA loop topologies by the dot product of the two end point tangent vectors. When the dot product is positive, we say it is a parallel (P) topology, and when the dot product is negative, we say it is anti-parallel (A) topology. The inter-operator DNA also has a directionality defined by the $5'$ to $3'$ direction, so the direction which the DNA binds to the LacI protein also serves to define two topological distinctions, forward and reverse. The inter-operator length does not include the 21 base pair DNA operators bound to the LacI head groups. If the DNA enters the LacI head group with the $5'$ to $3'$ direction facing into the protein, we refer to that as the forward direction, and the opposite as the reverse direction. In total these variation account for 8 topologies. The orientation of DNA to LacI requires a specific angular alignment which requires the DNA major groove to have a specific orientation relative to the LacI head group, therefore we need to consider the DNA phase angle Ψ, which is determined by the helical repeat. We consider the two lowest energy bound state solutions, referred to as the under- and over-twisted states, as mentioned in chapter 2. There are many different twist states corresponding to Ψ *e.g* $\Psi = 3\pi, 4\pi, 5\pi$, although the energetic cost is so large as to prohibit their consideration. The addition of phase angle Ψ brings our total number of possible binding topologies to 16. We find however that after making some simplifications that in general we only need to consider the two topologies, $P1$ and $A1$, see fig. 4.1.

The wild type *E. coli* system has three operators separated by 77 and 305 base pairs. The central operator O_1 has the strongest binding of the three. The strength of this binding is partially do to the nearly palindromic nature of the DNA sequence. The lack of palindromic DNA operators means that all binding topologies of DNA do

not bind with equal strength, and some may not be physically possible, *e.g* the $P1$ loop may have a high J factor, although the operator may not be able to bind under this orientation, therefore the loop will not form. In this work however, we treat all operators as palindromic, or O_{sym} and therefore consider the most probable binding topologies. There are several reasons loop topologies are important. In this work we have assumed ideal operators, although in wild type systems this is not typically the case. It is possible for the binding strength of the DNA to an operator to be strong in one direction and then weak in the other direction for non-palindromic operators. The implication of this fact is that while one loop topology may be very probable, the likelihood that the DNA will actually bind to the protein or even remain stably bound is very low. It is not possible to crystalize DNA loops which are bound to the lac repressor, and therefore unable to determine their structure using X-ray diffraction. Therefore, another method to determine the structure, and our model predicts which topologies should form and be stable.

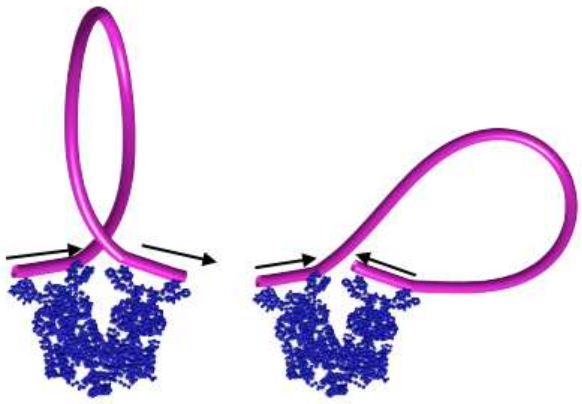

Figure 4.1 The two main topologies we consider are characterized by the manner with which their end-point tangent vectors, shown above, bind to the lac repressor, a roughly v-like structure, shown in blue. The plot on the left is the parallel topology $P1$ and the plot on the right is the anti-parallel topology $A1$. Each of these two binding topologies have two twist states. In general there are 8 distinct topologies with two twist solutions each, bringing the total number of topologies to be considered to 16. Twisting and bending the DNA will not change the binding topology.

4.2.1 Twist

After considering the bending required to get the DNA end point near the lacI head group, the end point must also be twisted to align with the phase angle Ψ required by the helical repeat. The helical repeat determines the orientation of the second DNA

operator relative to the first DNA operator when forming a loop. For example, it may be easy to form a loop when bound to one operator, but the second operator could be 180° rotated from the required boundary conditions. Meeting this requirement leads to an overall change to the DNA twist, ΔT_w.

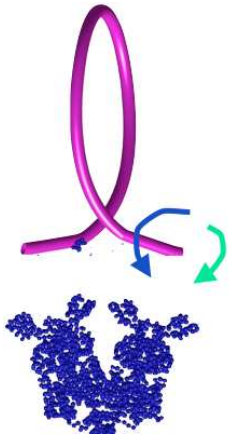

Figure 4.2 There are two rotations of the DNA end points which align the DNA with the LacI head groups and allow binding to the Lac protein. One rotation increases the twist of the DNA and the other decreases the twist of the DNA. The state shown here without additional twisting is referred to as the minimal twist state as it generally has the lowest twist energy.

As mentioned in chapter 2 these two solutions are classified by how they change the twist of the DNA

$$\Delta T_w = \frac{1}{2\pi}\int_0^1 \mathrm{d}s\left(\tau^{\ell}(s) - \tau^{o}(s)\right), \text{where } \begin{cases} \Delta T_w < 0, & \text{Under-twisted,} \\ \Delta T_w > 0, & \text{Over-twisted.} \end{cases} \tag{4.1}$$

In practice we find two solutions for each phase angle as we sweep through all possibilities of Ψ, although we do not necessarily determine if the loop is over- or under-twisted as we are generally interested in if a loop forms of a given topology, rather than its twist state.

4.2.2 Topological Simplifications

While the exact *in-vivo* structure of LacI is not known, LacI was crystalized by Lewis et al. [26] and subsequent molecular dynamics computations have shown the crystal structure to be a good approximation of the equilibrium shape [58, 57]. The LacI protein is not perfectly symmetric about the central axis, and given the limitations

of the structural knowledge of LacI, we neglect this difference and treat LacI as if it were symmetric. We are able to capture this lack of asymmetry by altering our boundary conditions, although without knowing for certain the *in vivo* structure, and considering we treat the protein as perfectly rigid, we neglect the difference for now. We also consider the DNA operators to be palindromic O_{sym}, therefore the forward and reverse direction of the DNA can also be neglected. These simplifications leave us with two parallel $P1$ and $P2$ and one antiparallel $A1$ solution along with the over- and under-twisted solutions, reducing the total number of possible binding topologies with two twist states that we need to consider to 6, although we also find that $P2$ is much higher in energy and therefore we do not consider it here.

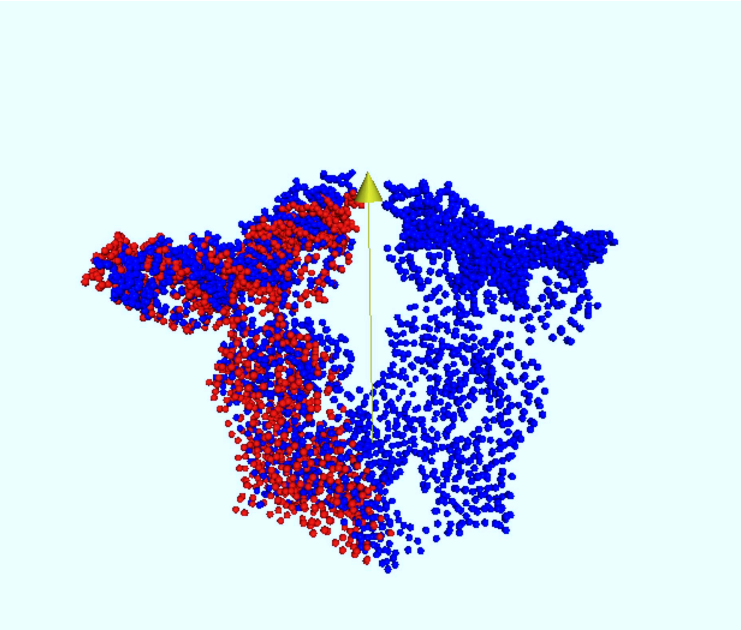

Figure 4.3 Since the Lac Repressor is a homotetramer, we expect the protein to be symmetric about the central axis. We demonstrate the small asymmetry of Lac Repressor by rotating the right half of the Lac Repressor protein by 180° about the central axis shown above. We treat the protein as if it were perfectly symmetric [30], which results in a deviation of the crystal structure of less than 10 Angstroms and a rotation of less than 10 degrees from those found using X-ray crystallography by Lewis et al. [26]

4.3 Loop formation energy as a function of length

In this chapter we consider intrinsically straight DNA $\vec{\kappa}^o = 0$ with isotropic bending stiffness $\ell_p = \ell_1 = \ell_2$. We consider DNA equilibrium states with inter-operator DNA lengths from 50 to 650 base pairs. These equilibrium states are found using the rod

model of Goyal et al. [15]. The enthalpic cost of loop formation with minimal twist energy as a function of length is given in fig. 4.8. In practice, the rod model is used to compute all angles of twist by slowly rotating one of the end points through 4π, while finding an equilibrium at each angle, as demonstrated by Lillian et al. [27].

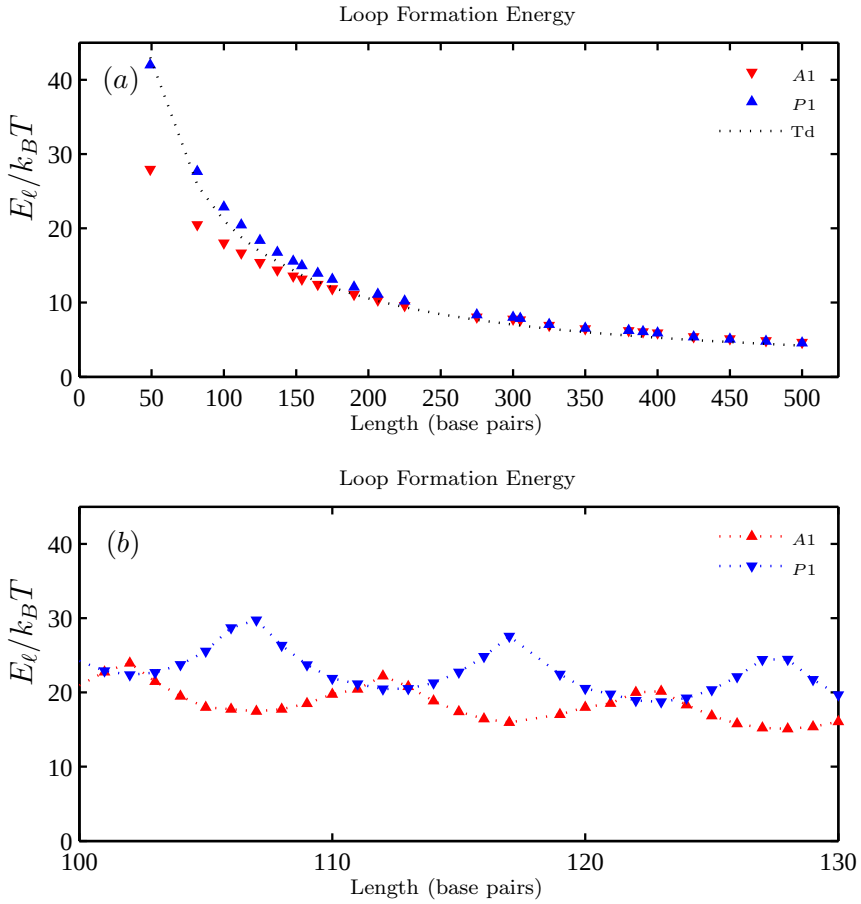

Figure 4.4 (a) The loop formation energy as a function of length for the minimum twist state, which does not take into account the helical repeat when angularly aligning Ψ for the loop binding. The teardrop has also been plotted for comparison of loop formation energy, and it is clear that the separation of end-points due to the lac repressor has a large effect on the required loop formation energy. (b) The loop formation energy has a strong phase dependence Ψ, which caries over to the J factor computation. The phasing allows $P1$ to be comparable to $A1$, as opposed to the minimal twist state. We assume a helical repeat of 10.5 base pairs.

We see that the Anti-parallel loops $A1$ have the lowest loop formation energy for short lengths of DNA, and then as length is increased, the $P1$ and $A1$ loops become comparable. The energetic cost of $P2$ loops are so much higher that we do not

consider them until longer lengths of DNA.

4.4 Normal Modes

We proceed with computing the normal modes as outlined in chapter 2. For a given length and topology, we compute the normal modes for each equilibrium state value of ψ. In chapter 3 we found that the eigenmodes were decoupled into planar and out-of-plane modes. The lac mediated loops are strongly coupled, involving all three deformation variables θ_1, θ_2, ψ, over the space of modes, until the modes become insensitive to the loop curvature, as described in section 2. In general a few more modes are needed for lac mediated loops than for the planar self-contact loops. An example of a normal mode for $A1$ and $P1$ are given in fig. 4.5 and fig. 4.6, respectively.

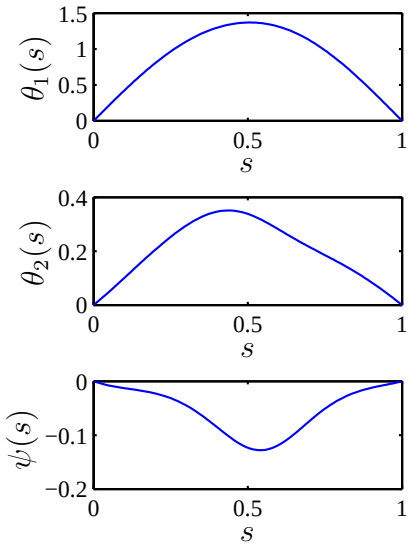

Figure 4.5 The lowest positive eigenmode for DNA of length $L = 150$ basepairs for the $A1$ topology. In contrast to the planar loops where θ_2 was decoupled from θ_1 and ψ, the normal modes of the DNA-lacI system typically include all three deformation variables θ_1, θ_2, ψ.

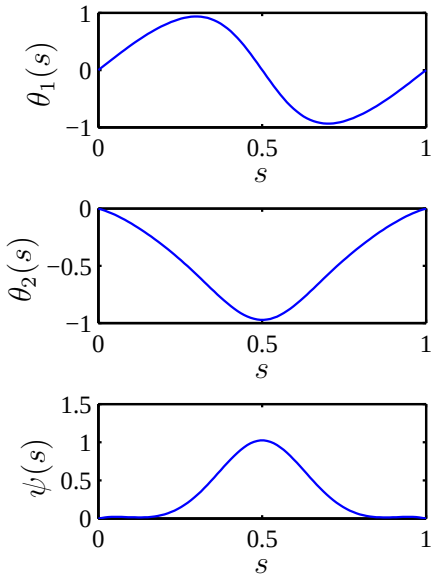

Figure 4.6 The lowest positive eigenmode for DNA of length $L = 150$ basepairs for the $P1$ topology. The mode is highly coupled in θ_1, θ_2, ψ. This mode is nearly the same stiffness as the $A1$ loop above, although its behavior in terms of θ_1 and θ_2 is quite different.

The normal modes are very sensitive to phase angle. We see that the lowest mode remains the most sensitive compared to the other modes. The negative eigenmode does not cause concern because the constraint matrix corrects for this value. The interpretation is that the loop can relax to a lower energy state if we allow displacements of the end-points, which is to be true if it were not for the presence of $\delta^3(|\vec{u}|)$, see fig. 4.7.

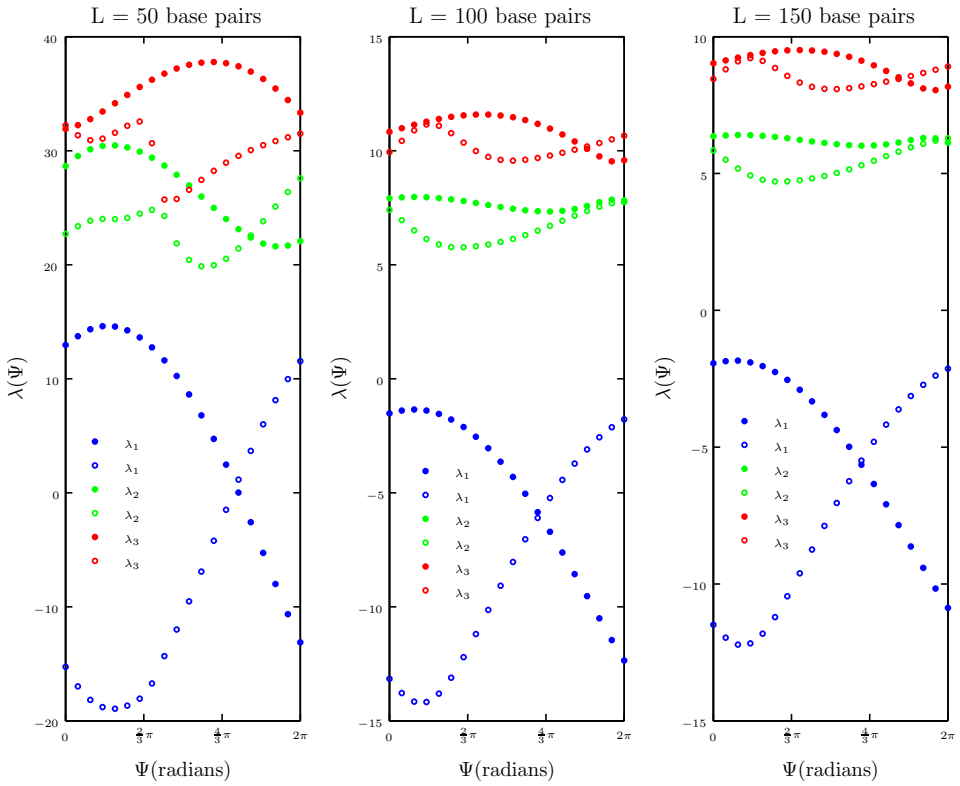

Figure 4.7 The three lowest eigenvalues for both twist states of the $A1$ binding topology as a function of Ψ for three lengths, $L = 50, 100, 150$ base pairs. The eigenvalues have a strong phase angle dependence, although we will see below that when combined with the δ functions to form Λ, does not lead to a large phase dependence in the entropic coefficient.

4.5 J factor Computations

We compute the J factor as a function of length L and Phase Angle Ψ and topology $A1$ and $P1$. We find that in general, the $A1$ loop topology is the most probable topology for DNA loops of less than $\sim$ 200 base pairs. For loops longer than 300 base pairs, the effect of phasing on preferred topology is greatly diminished. For a given length we compute the J factor for for $\Psi = 0, \ldots, 2\pi$, for both over- and under-twisted DNA loops for the two topologies we consider. From these computations, we initially consider the most probable loop formation state, previously referred to as the minimum twist state. We begin by considering DNA loops which satisfy the boundary conditions required for DNA-lac co-crystals [26], although we do enforce the phase angle requirement specified by the helical repeat. We instead consider the most

probable loop formed, irrespective of the phase angle. Typically this loop requires little twist, so we refer to it as the minimum twist state. We then find that in terms of length alone, the $A1$ topology is preferred for DNA less than 200 base pairs. We also find that after 300 base pairs, the $P1$ loop begins to become more probable than the $A1$, though the difference in their probabilities is small. These loops in general are highly non-planar, and have comparably probable binding topologies. The role of enthalpy is to prohibit short loops from forming due to the amount of curvature per base pair required, and then at longer lengths, enthalpy becomes less dominant as entropic effects take over. The role of entropy starts out small, and then becomes dominant at longer lengths. The entropic contributions set the physical scale of the J factor in terms of inverse volume, which we then convert to Molarity.

4.5.1 Preferred Topologies as a function of Length for Minimum Twist State

We begin by considering the two equilibrium states with minimum twist energy. The enthalpic contributions strongly affect the J factor at lengths under a persistence length, and then after nearly two persistence lengths, the loop formation energy becomes a small effect on the J factor, see fig. 4.8.

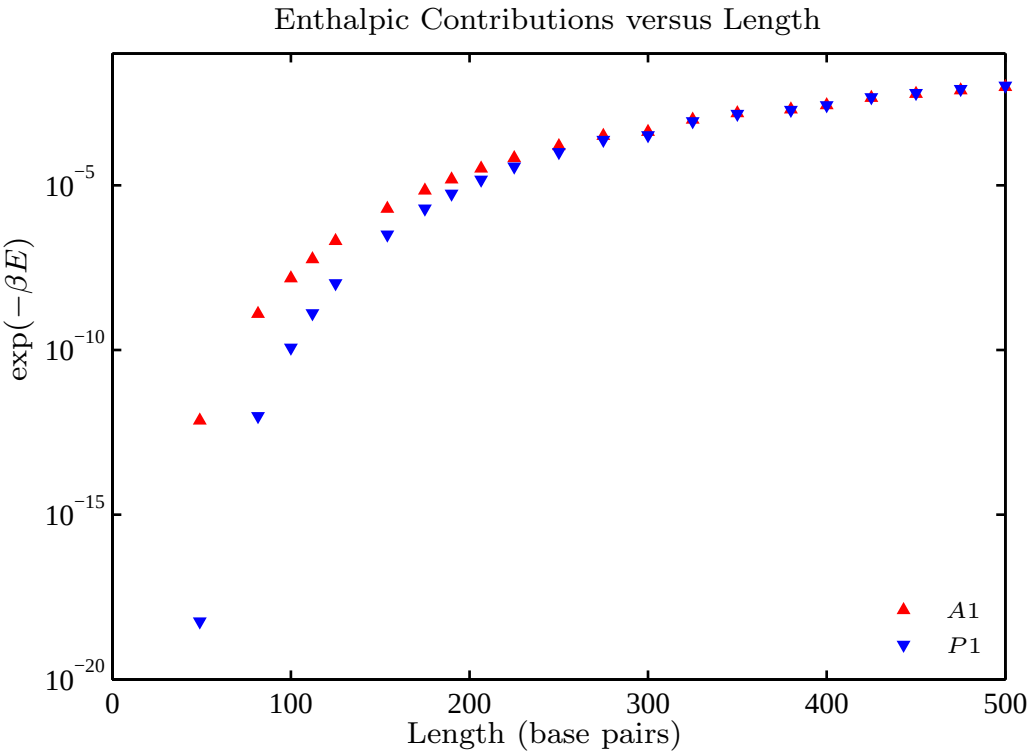

Figure 4.8 The enthalpic contributions to the J factor decrease with length, as the loop formation energy decreases. In general it is easier to form the $A1$ loop for DNA less than 200 base pairs in length. When we consider phasing, this plot will show $A1$ and $P1$ getting closer at shorter lengths, although in general $A1$ requires less bending per base pair to form a loop and therefore remains easier to form.

The entropic contribution to the J factor for $P1$ and $A1$ loops with minimal twist are given in fig. 4.9, when loop formation energy decreases, entropic effects begin to dominate the J factor. The effect of the entropic coefficient increases with length, the pure dimensional scaling , given by ρ is

$$\rho = \frac{1}{\ell_p^3}\left(\frac{\ell_p}{L}\right)^{11/2}\sqrt{\frac{\ell_\tau}{2\pi L}} \tag{4.2}$$

where the contribution of $\sqrt{\frac{\ell_\tau}{2\pi L}}$ did not appear for the planar loops since we considered them to be torsionally free at the end-points.

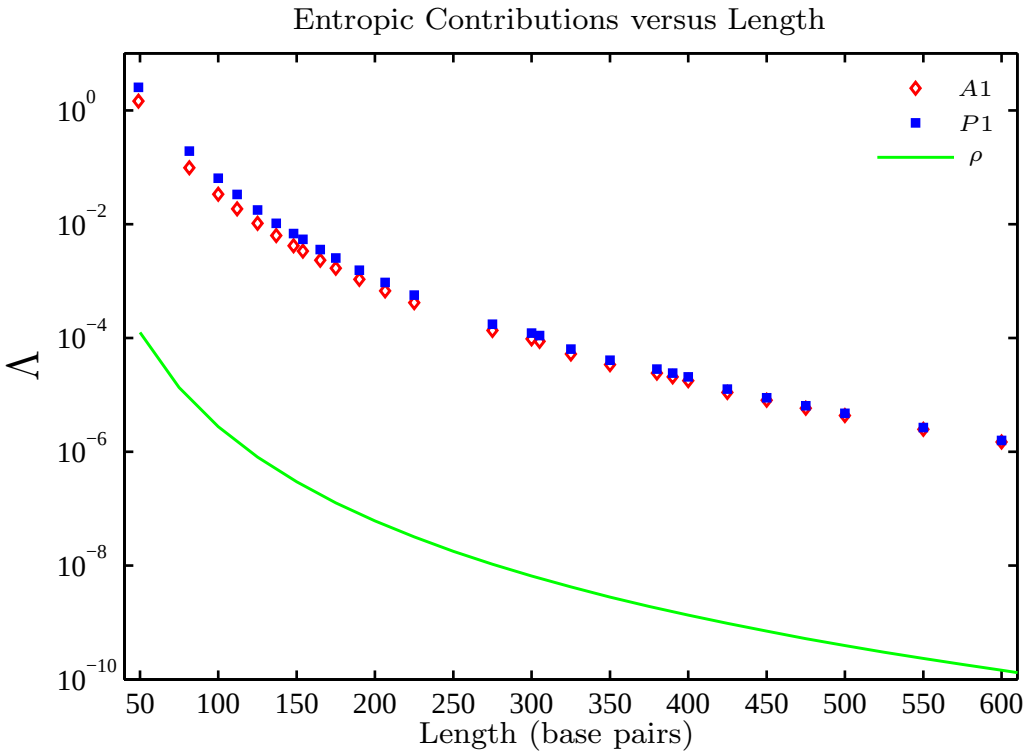

Figure 4.9 The entropic contributions span 5 orders of magnitude over this length range. The scaling term ρ is shown for clarity. While the units are technically that of Molarity, this plot is intended to show only the minimum twist entropic contribution and should not be considered separately.

The entropic contributions to the J factor span 5 orders of magnitude over the length range of $50 - 650$ base pair, or roughly $16 - 220$ nm. The length scaling is as expected in chapter 2. The entropic coefficient slightly favors the $A1$ binding topology, although the numeric difference between the entropic coefficient for these two topologies is very small overall for intrinsically straight DNA bound in the minimum twist state.

The J factor for DNA under 200 base pairs is larger for the $A1$ binding topology. On longer length scales, the difference in the required curvature per base pair is low enough that the $A1$ and $P1$ loops become comparable without considering the effects of phasing. The reason for the preferred topologies is the end-point curvatures for $P1$ remain much larger than those of $A1$ resulting in a suppression of $P1$ binding topologies. We now have a measure of large and small probabilities for DNA loop formation from $50 - 650$ base pairs. These results for the J factor will serve as our basis of comparison for intrinsically curved states, as seen in chapter 5.

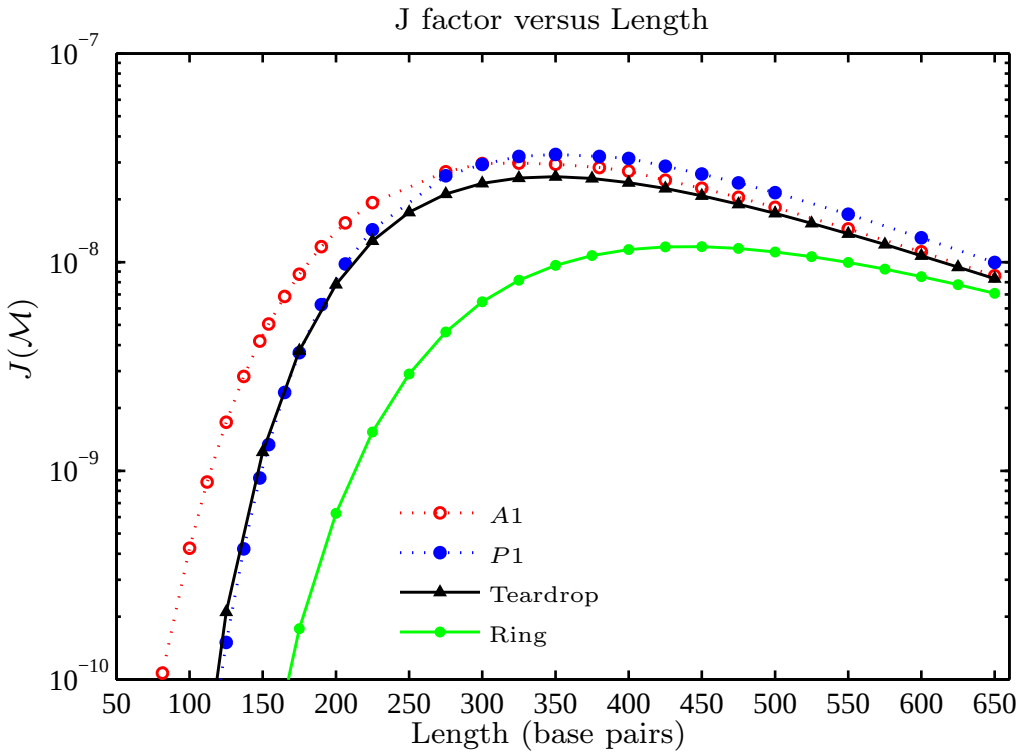

Figure 4.10 The J factor for the minimal twist states as a function of length show a clear preference for $A1$ over $P1$ for lengths under 200 base pairs. The separation of DNA end-points required for looping due to the distance between the LacR head groups makes the looping more probable, as seen by comparison with the teardrop, which is the lowest energy loop for planar shapes. Also shown here is the ring, which remains the least probable loop throughout due to the difficulty of aligning the end-point tangents. The bending and torsional persistence lengths are 50 and 75 nm, respectively.

4.5.2 The Preferred Topology is a function of Phase Angle Ψ

The computation of the J factor for the minimal torsion loop revealed that for lengths under 200 base pairs, the $A1$ topology was preferred. We now examine how phase angle affects this preferred topology. For a given length of DNA we compute all potential phase angles and then use the proper helical repeat to determine which J factor is needed. On average DNA has a helical repeat of 10.5 base pairs although since each base pair difference changes the phase angle by roughly 34 degrees, sequence effects can alter this repeat, even if only locally while leaving the overall sequence with a helical repeat of 10.5 base pairs on average. While the J factor for each of the two twist states for each DNA binding topology are different, the question we are seeking

to answer is what is the probability that a loop of a given topology will form, not what is the probability of the over-twisted state, so we sum the two J factors for each length. In general only one of the two twist states contribute meaningfully, as is seen in fig. 4.11. We see that after a certain angle Ψ it if easier to loop if instead the DNA twisted in the other direction. In general, only one solution meaningfully contributes to the J factor, except in regions where the two states are comparable.

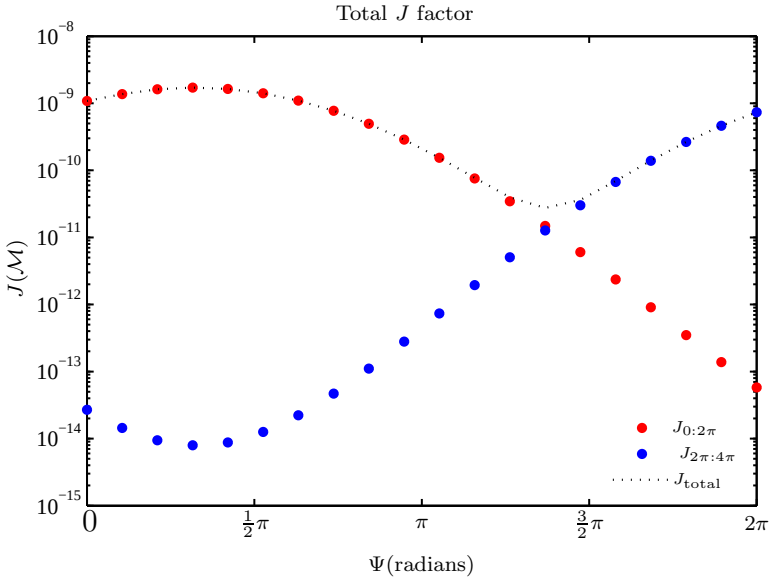

Figure 4.11 We present the two J factor solutions for every possible phase angle. In general only one of the solutions contributes meaningfully to the J factor, as seen by the dotted line. In principle one is over-twisted, and the other is under-twisted, although both would be equally capable of gene regulation where they meet near $\Psi \sim 3\pi/2$, so their combined contribution is used to compute the total J factor. The actual twist of each state would need to be determined by integrating the torsion τ^{ℓ}. Note that there normally appears to be a cusp in the J factor, which is due to this summation and not some inherent numerical instability.

Now we turn our attention to the effect of twist as a function of length. We see in fig. 4.12 that $P1$ retains a large phase sensitivity from $50 - 200$ base pairs, whereas $A1$ becomes less sensitive to phase. We see that as DNA approaches 200 base pairs that the difference in binding topologies becomes less important as phasing can bring the two loops close enough to be roughly equally probable. At least two methods can be used to test these predictions. The first method is Tethered Particle Microscopy (TPM) by computing ratios of looped lifetimes for different lengths. The ratio of

open τ^o to looped τ^ℓ lifetimes is given by

$$\frac{\tau^\ell}{\tau^o} = \frac{J}{c_o} \exp\left(\frac{\epsilon}{k_B T}\right) \tag{4.3}$$

where ϵ is the binding strength of the LacI protein and DNA operators, and c_o is a reference concentration, normally take to be $1\mathcal{M}$. While we do not know the precise value of ϵ, it is the same for each topology.

The second method is Fluroescense Resonance Energy Transfer (FRET) by placing markers ahead of and behind the DNA operator segments. Then by measuring the FRET efficiency, a determination of which loop topology was present would be possible, and for regions where both topologies are equally probable, an equal mixture of the two states should be seen, see chapter 6 for more information.

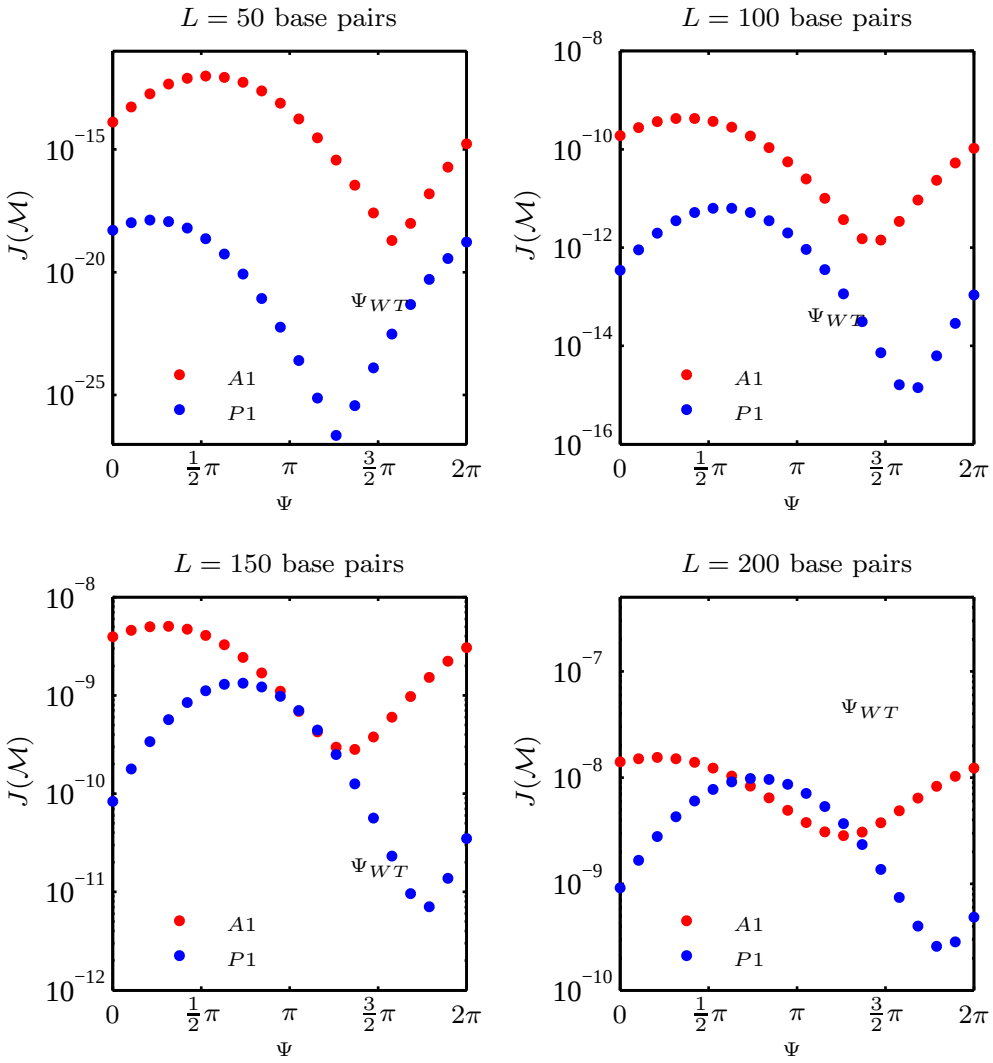

Figure 4.12 The J factor for $A1$ and $P1$ become comparable as length is increased due to phasing effects of the DNA operators. The wild type helical repeat phase angle determined by 10.5 base pairs, has been indicated on each plot by Ψ_{WT} These computations could be used to determine FRET efficiency.

4.5.3 The Dependence of Length on Phase Angle

We now examine the effect of the phase angle Ψ on the J factor as a function of length. We assume that DNA has a helical repeat of 10.5 base pairs, although the exact number does not change the qualitative results, only the relative location of the peaks and valleys within the J factor. The value of the phase angle is determined by

dividing the length by the helical repeat and then using the remainder to determine the proper orientation to determine the proper angle for the $A1$ and $P1$ loops. In fig. 4.13 we separately plot the enthalpic and entropic contributions, as well as the total J factor over a range of 30 base pairs in order to demonstrate the effect of phasing. The amount of phase dependence within the enthalpic contributions dominate the phasing of the J factor. The entropic contributions do not in general show a strong phasing effect, although this fact is not surprising when you consider the exponential nature of the enthalpic contributions versus the multiplicative nature of the entropic terms within the J factor. The entropic contributions of the normal modes also enter in through a square root, further suppressing any variations for a short length. We will see however that when intrinsic curvature is added to the DNA, that this is not always the case in chapter 5. The introduction of phasing shows that while the minimal torsion loop shows a clear preference for the $A1$ topology for DNA under 200 base pairs, that phasing can change this picture and allow $P1$ to be competitive when $A1$ is difficult to form.

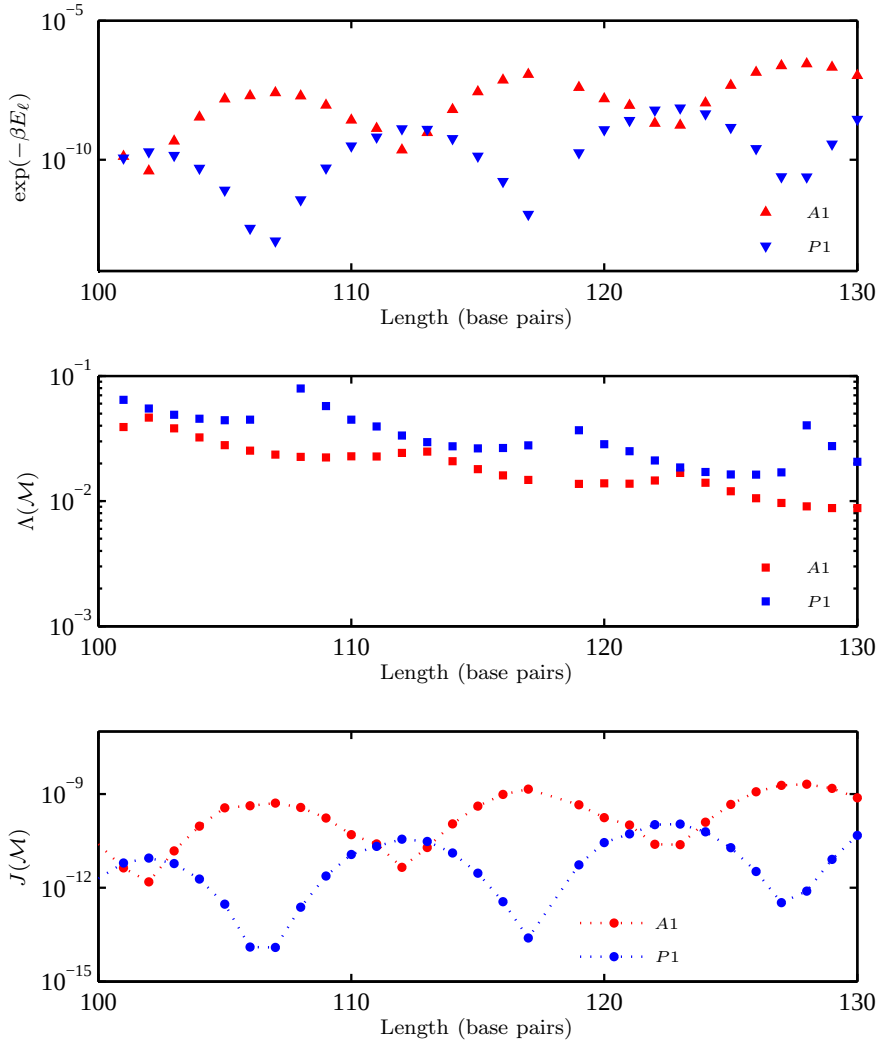

Figure 4.13 The top plot shows the enthalpic contributions to the J factor for a helical repeat of 10.5 base pairs. Both loop topologies show a strong phase dependence, although $P1$ has a greater variation. The middle plot shows the entropic contributions for the J factor to be relatively phase independent, and while not shown here, continue to behave this way ever at longer lengths of DNA. The bottom plot shows the J factor for $A1$ and $P1$, and demonstrates that variations in the enthalpic contributions as a function of length tend to dominate the J factor.

The J factor exhibits a strong phase angle dependence over a great range of loop lengths. In fig. 4.14 we assume a helical repeat of 10.5 base pairs then examine its effect for a large range of lengths. We find, as expected, that for shorter lengths, the phase dependence has a large effect on the J factor, up to several orders of magni-

tude. Then for longer lengths of DNA we see that the effect of phasing is diminished, having less than an order of magnitude effect on the J factor. We also see that the $P1$ loop has a stronger dependence on the phase angle than the $A1$ loop. This strong dependence is due to the difficulty of forming the DNA loop in a smaller volume and requiring more curvature per base pair for poorly phased DNA loops.

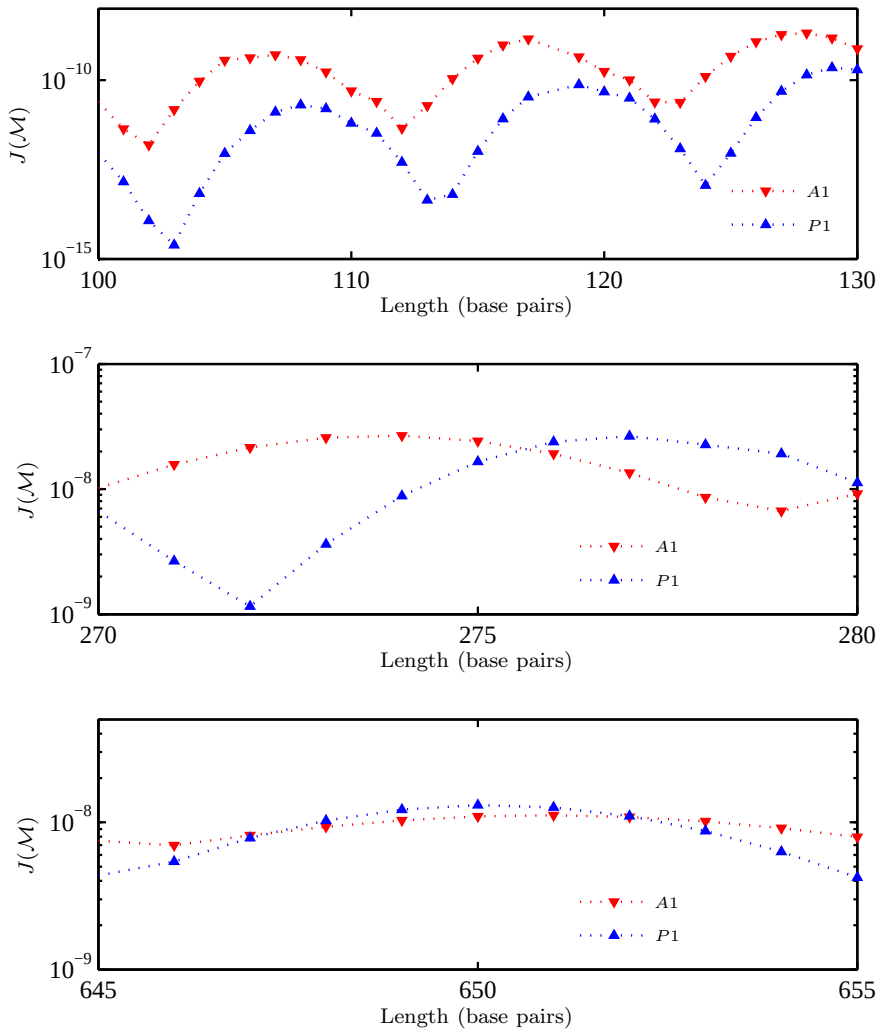

Figure 4.14 The J factor as a function of length and phase angle Ψ for a helical repeat of 10.5 base pairs. The top plot shows that for short lengths of DNA ~ 115 base pairs, the operator phasing can have a 5 orders of magnitude effect on the J factor. We see in the middle plot that near the peak of the J factor, ~ 275 base pairs, the operator phasing has about an order of magnitude effect on the J factor. Finally in the bottom plot, we see that for very long lengths of DNA ~ 650 base pairs that phasing has a considerably diminished effect of less that an order of magnitude. Throughout however $P1$ continues to have the strongest phasing dependence for intrinsically straight DNA.

4.6 J factor Components

Our independent computation of enthalpic and entropic contributions to the J factor allows a greater qualitative understanding of the overall behavior with respect to length. In fig. 4.15 we present the minimal torsion J factor for the $A1$ binding topology. On this plot, we overlay the enthalpic and entropic contributions in order to illustrate their effect on the J factor. From the perspective of increasing loop length, the J factor is initially constrained by the enormous amount of energy needed for such short loop formation, as seen by the solid line. then as entropic contributions become larger, they begin to suppress the J factor as the enthalpic contributions yield. The net result is peak in the J factor near 300 base pairs, and then entropic effects begin to dominate, bringing the J factor down again. This entropic dominance can be understood by recognizing that there is a larger conformational space available to the longer DNA molecule, and therefore it requires a larger change of entropy to form the loop.

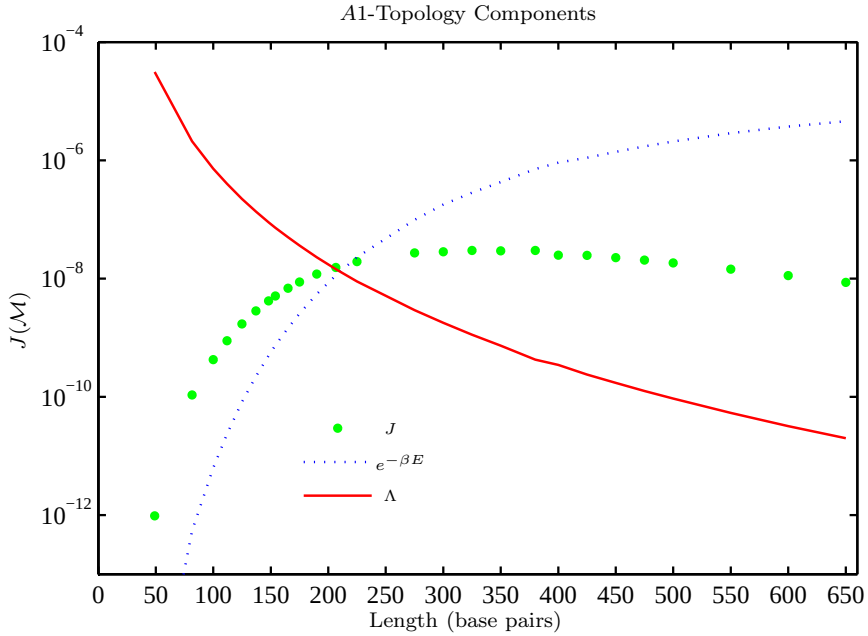

Figure 4.15 Initially the J factor is dominated by the loop formation energy E_ℓ, although E_ℓ scales as ℓ_p/L, so as length increases, the enthalpic contribution $\exp(-\beta E_\ell)$ tends towards 1. The entropic terms scale as $(\ell_p/L)^{11/2}$, so as length increases, Λ dominates as it tends towards 0, and pulls the J factor peak down with length. There is a region where the enthalpic resistance to looping is lessening as quickly as the conformational space is increasing for the entropic terms and pulls down the looping probability. Within this region the J factor plateaus for a period, and then as enthalpy can lessen no more, the J factor then begins decreasing again.

4.7 Free Energy versus Length

The J factor can also be related to the change of free energy ΔG cost of loop formation

$$J = c_o e^{-\beta \Delta G} \tag{4.4}$$

and c_o is a reference concentration, normally take to be 1 $\mathcal{M}$. The free energy cost of loop formation was recently computed using Monte Carlo methods by Towles et al. [55] for DNA mediate loop formation. These results are in good agreement with our minimal twist solutions; although we disagree as to the magnitude of phase dependence on the J factor. We predict a larger torsional variation in the free energy, although we would expect this difference as Towles et al. [55] used a different form of the stiffness tensor $B(s)$, where they included twist-bend coupling, which should lessen the effect of phasing, especially for short lengths of DNA where any additional flexibility will greatly lower enthalpic contributions, see fig. 4.8. Towles et al. [55] also computed the protein binding differently, considering the DNA operators as each binding to two sub-regions of the Lac repressor head groups as well as treated the full asymmetry of LacI. The fact that we are in agreement for the minimal twist state, demonstrates a validation of earlier simplifications. There are also some clear disagreements with Towles et al., as they find the $P2$ loop to remain comparable to the $P1$ loop for very short lengths of DNA L 100 base pairs, which seems contradictory to the prevailing data, as the curvature required to form $P2$ necessitates a very large loop formation cost, and greatly suppresses the J factor. In addition, they show a relative phase insensitivity for $A2$ and not for $A1$ over all lengths considered. It is unclear to us why this would be the case at this time.

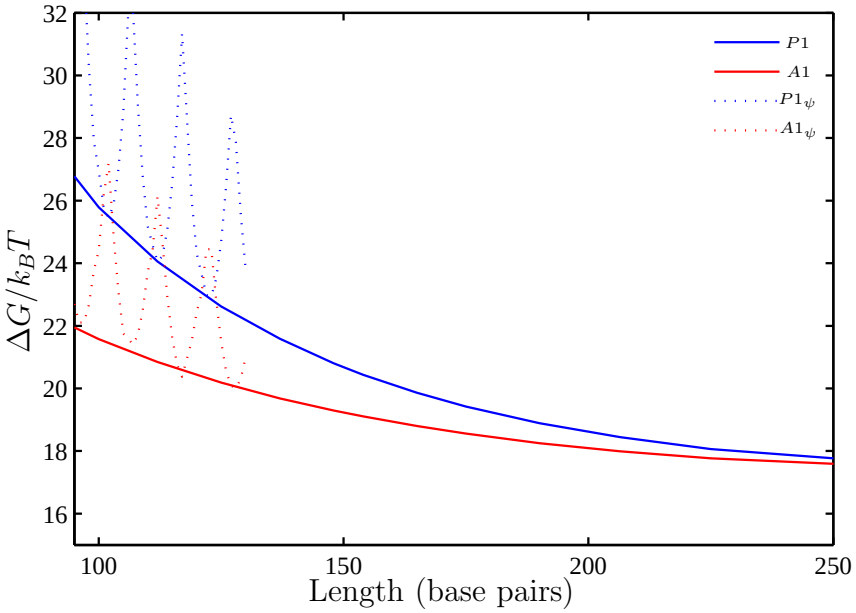

Figure 4.16 The free energy for the minimum twist states for $A1$ and $P1$ as a function of length. We find larger torsional variation than Towles et al. [55].

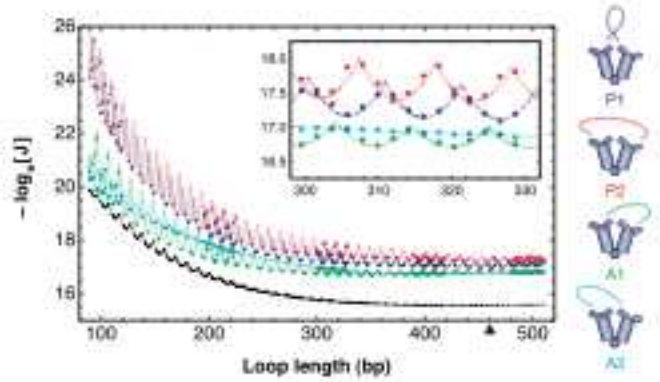

Figure 4.17 Monte Carlo results by Towles and Nelson, the light blue/green are $A1$ loops, and the red/dark blue are $P1$ and $P2$ loops.

The loop formation energy for $P2$ at short length scales is much higher than those of $A1$ and $P1$, so it is unlikely these Monte Carlo results are correctly interpreted at those lengths.

4.8 Conclusions

We have computed J factors for Protein mediated DNA loop formation which focus on the Lac Repressor system. We have shown that DNA looping depends strongly on length, phase angle and allowable binding topology using our J factor computation. We independently compute the enthalpic and entropic contributions to the J factor for arbitrary boundary conditions. We have also computed the J factor for the $A1$ and $P1$ binding topologies and discussed the implications of a preferred binding topologies for wild type systems, as well as for modeling more general systems for structural biologists. We have shown that the operator phase angle Ψ is critical for determining the most probable binding topology for intrinsically straight DNA bound to the lac repressor for DNA lengths under ~ 225 base pairs. Over the length scales of $50 - 650$ base pairs, the $P1$ binding topology continues to have the strongest phasing dependence for intrinsically straight DNA.

We are also able to make some predictions about the wild type system. All things being equal, that is without considering proteins such as HU which induce non-specific curvature into DNA, we would expect the wild type system to have $A1$ as its 77 base pair loop topology, and $A1$ and $P1$ in roughly equal proportions for the 305 and 382 base pair loops. We are also able to indicate a regime of inter-operator DNA lengths could be used to test the predictions of our model using Tethered Particle Microscopy and bulk FRET measurements. We quantitatively agree with Towles et al. [55] when looking at the Gibbs free energy for the minimal torsion state of the $A1$ and $P1$ topologies. Our computation of any topology for a given length and full range of Ψ from 0 to 4π takes under 5 minutes, making this model extremely computationally efficient.

Chapter 5

Intrinsic Curvatures

5.1 Introduction

In this chapter we study the effect of intrinsic curvature $\vec{\kappa}^o \neq 0$ on DNA loop formation. This intrinsic curvature typically results from sequence-dependent base pair stacking. Each set of base pairs tilt the centerline of the DNA helix away from a straight line, *e.g.* imagine building a tower with wedges rather than square blocks. On average however this tilting and twisting of a random DNA sequence will average out, resulting in a linear molecule, see fig. 5.1. Intrinsic curvature can also be induced by architectural proteins which we will discuss further in chapter 6. In the case of architectural proteins, bending and torsional stiffness of the DNA-protein complex would have to be carefully considered. In this chapter we study DNA sequences with intrinsic curvature resulting from Adenine tracts (A-tracts), which have be shown to dramatically increase loop stability [37] in lac repressor mediated DNA loop formation. Previous works by Goyal, Lillian et al. have shown that in principle, intrinsic curvature due to A-tracts can greatly reduce loop formation energy E_ℓ and preferred looping topology [16, 29].

A major hinderance to DNA loop formation of relatively short DNA is the large energetic barrier required to surpass in order to form the loop. However, if the DNA has an intrinsic curvature which is aligned with the desired loop curvature, then the overall loop formation energy E_ℓ will be greatly lowered as

$$\beta E_\ell = \frac{1}{2}\int_0^1 \mathrm{d}s \;\; (\vec{\kappa}(s) - \vec{\kappa}^o(s))^T \, B(s) \, (\vec{\kappa}(s) - \vec{\kappa}^o(s)) \, . \tag{5.1}$$

In this chapter we continue to treat the stiffness tensor $B(s)$ as diagonal with homo-

geneous bending stiffness, $\ell_1 = \ell_2 = \ell_p$.

The intrinsic curvature of an A-tract is due to the repeated placement of AT base pairs on the length scale of a helical repeat. The individual twist and tilt effects due to this ordering are cumulative as a result. Coordinated placement of each A-tract leads to the A-tract bends. This cumulative affect of this coordinated tilting and twisting is to induce super-helical curvature and torsion of the DNA molecule in addition to the normal double stranded helix, see fig. 5.1. We will consider A-tracts that are approximately 80 base pairs in length [29]. According to the SHS model [29], the principle curvature and torsion of the helical domain is 0.0098 rad/A$^\circ$ $\sim$ 2.81 deg/nm and 0.0074 rad/A$^\circ$ $\sim$ 2.12 deg/nm, respectively. This SHS structure is a close approximation, varying by less than $1nm$ along the DNA sequence, of the structure provided by the model.it tool which uses the tri-nucleotide model to predict DNA structure [40].

5.2 Experimental Studies of A-tract Bends

A series of DNA constructs containing A-tracts and two DNA operators were designed by Mehta et al [37] to test the hypothesis that intrinsic curvature should make loop formation easier. In order their constructs contained a DNA operator, then a region of straight DNA, referred to as a linker region, then their A-tract sequence, followed by another linker region, then finally terminated by a DNA operator. They varied the length of the two linker regions in order to rotate the DNA binding operator orientation relative to the A-tract bend, and modify the phase angles ψ_1 and ψ_2, see fig. 5.1. In this manner they could sweep out both phase angles with only a modest change to overall sequence length. They found that their constructs formed loops with the lac repressor and remained looped for a significant amount of time when compared to intrinsically straight DNA. In general, wild-type or intrinsically straight DNA loops are found to last for nearly 30 minutes, while these designed A-tracts sequences lasted for over 24 hours. The primary difference between the two sequences is the intrinsic curvature, which suggests that eq. (5.1) leads to the proper intuition when considering intrinsic curvature. In addition to the Gel electrophoresis experiments, Morgan et al. [38] were also able to measure the loop binding topology for these constructs using FRET methods.

These designed sequences are interesting because they allow us to carefully consider the effects of intrinsic curvature in a controlled manner. Since DNA is highly

compactified within cells, we know that it must exist in a highly bend or curved state, understanding the effects of curvature on DNA loop formation is essential to understanding DNA within a cellular environment.

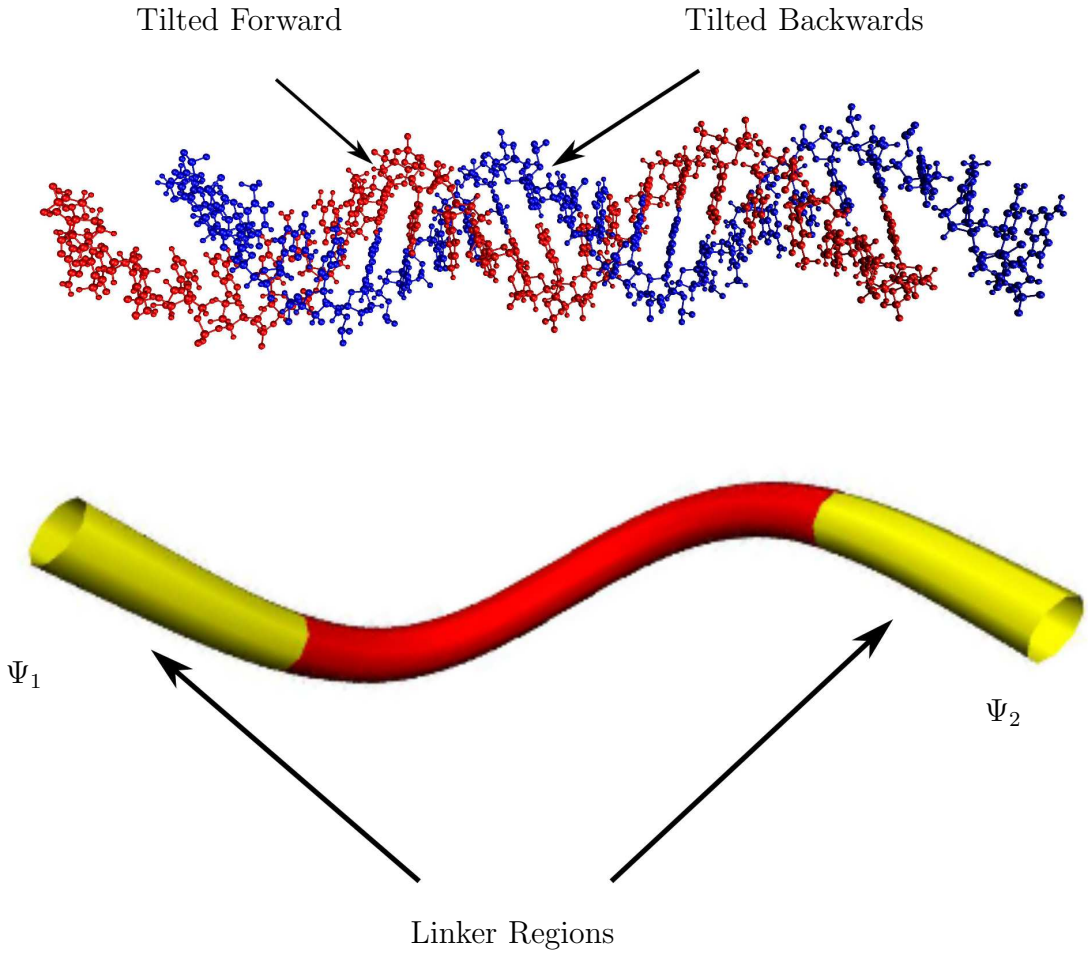

Figure 5.1 The top figure shows that a random sequence of base pairs is on average straight, or lacks intrinsic curvature $\vec{\kappa}^o = 0$. We see that some base pairs tilt forward, while other base pairs tilt backwards. The bottom figure shows the Straight Helical Straight approximation (SHS) of Lillian et al. [29], for the zero temperature A-tract design sequences of Mehta et al. [37]. There are two straight linker regions (light gray/yellow) are flanked by DNA operators (not shown). These linker regions border the central curved A-tract region (dark gray/red). The intrinsic curvature of the A-tract is due to carefully repeating AT base pairs along the DNA length such that their collective tilt is complementary and the result is super-helical intrinsic curvature. Imagine pulling out blocks from a Jenga(tm) tower from the same side and then gently twisting the structure. The length of the linker regions control the two phase angles Ψ_1 and Ψ_2, two phase angles are now required due to the intrinsic curvature. By adding and subtracting a base pair from the left and right hand side, respectively, these phase angles can be changed. According to the SHS model [29], the principle curvature and torsion of the helical domain is $0.0098 rad/\mathring{A}$ and $0.0074 rad/\mathring{A}$, respectively.

5.3 Modeling the A-tract Structure

These constructs were numerically studied by Goyal et al. [16] and shown to be much more energetically favorable than straight DNA. Continuing this work, Lillian et al [30] theoretically investigated the A-tracts bends design space spanned by two phase angles Ψ_1 and Ψ_2, see fig. 5.1.

The A-tract sequences we examine are flanked by two idealized LacI operators. They have an inter-operator DNA length of 142 base pairs, which is of the order of ℓ_p, so we do not encounter any length scaling difficulties when comparing different sequences. The typical linker region lengths are on the order of 30 base pairs each, and the A-tract length is on the order of 80 base pairs. The collective structure of the linker-A-tract-linker DNA sequence is referred to as the Straight Helical Straight (SHS) approximation [30]. While the initial binding angle was not required for straight symmetry due to rotation symmetry about the end-point tangents, the A-tracts require two phase angles Ψ_1 and Ψ_2 to describe the binding properly. It is important to remember that while the topologies are referred to as parallel and antiparallel, the binding domains on the lac repressor are not parallel.

The total DNA length of 142 base pairs is kept fixed by adding and subtracting one base pair from the two linker regions. The exploration of this design space revealed that parallel binding topologies were now possible, while the same length of straight DNA strongly favors antiparallel binding for all orientations of the phase angle Ψ.

5.4 Results

We compute J factor over the design space of the A-tract sequences to determine how intrinsic curvature affects DNA loop formation as well as the preferred binding topology. The lac repressor mediated loop formation is now dependent on two phase angles Ψ_1 and Ψ_2 due to the addition of intrinsic curvature. We also examine how much of a change relative to straight DNA exists. We also compare how our predictions of loop lifetimes compare to the results of Mehta et al. [37] Finally we compare our J factor topology predictions to those computed by Morgan et al. [38].

5.4.1 The effect of curvature on loop formation

The intrinsic curvature of the $A-tracts$ greatly reduce the cost of DNA loop formation, some loops are as low as $4.8k_BT$, in contrast on this length scale, the lowest minimal torsion $A1$ loop for straight DNA is approximately 12 k_BT. Overall the loop formation energy is in general much less for at least one A-tracts bend binding topology, even in regimes where the twist energy is very high, as we see in fig. 5.2.

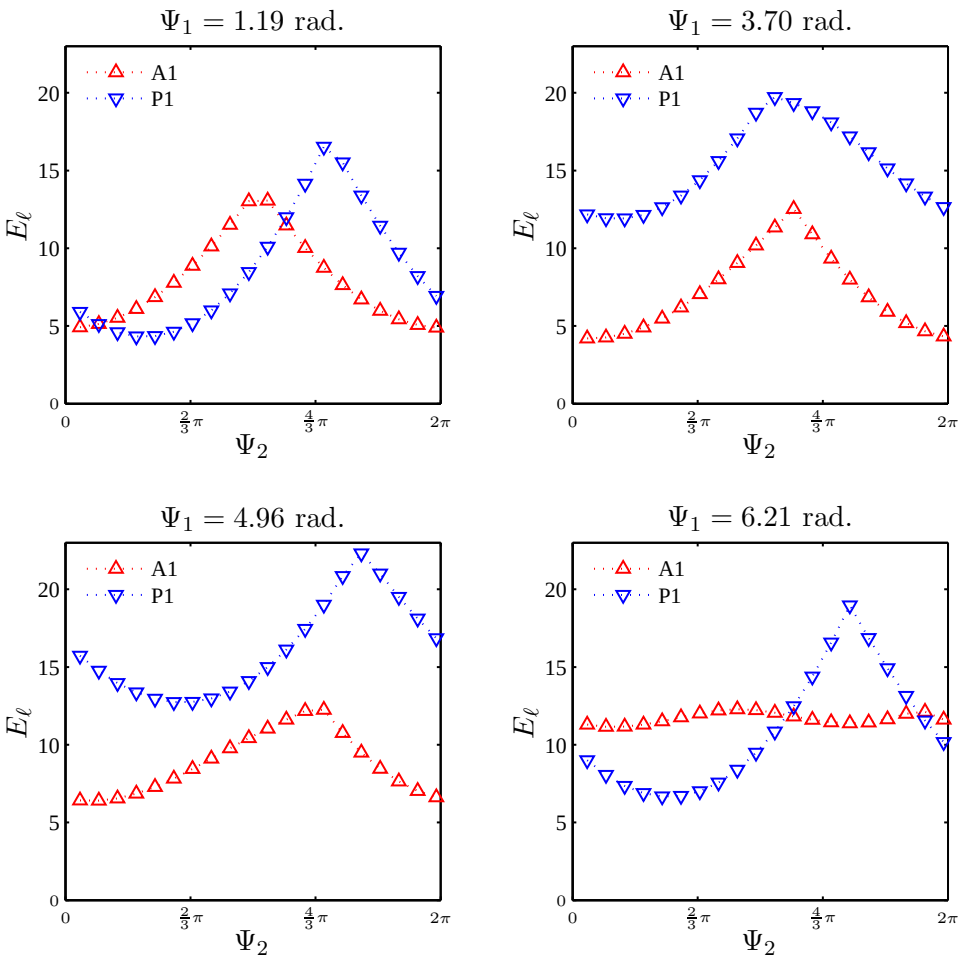

Figure 5.2 The energy required to form a loop with pre-bent DNA is generally less than straight DNA for at least one of the DNA binding topologies. Exceptions occur in regions where phasing is in extreme opposition to the loop forming. In contrast to straight DNA, we see that for this 142 base pair DNA that both $A1$ and $P1$ binding are allowed and comparable over a large range of Ψ_1 and Ψ_2. As Ψ_1 is rotated, the most probable loop changes quickly from $A1$ to $P1$ and back, based on Ψ_2.

The exponential suppression to the J factor loop formation energy exponentially

suppresses DNA loop formation as $\exp(-\beta E_\ell$, although as we say in chapter 4, increasing DNA length lowered this barrier and greatly enhanced the J factor, analogously the intrinsic curvature generally lowers the loop formation barrier.

5.4.2 Binding Topologies

As with the intrinsically straight DNA, the A-tracts have two preferred topologies $A1$ and $P1$, with $P2$ being so much higher energetically that we do not consider it here. In contrast to the intrinsically straight DNA, the A-tracts sequences do not show a clear preference to parallel or antiparallel binding topologies. We begin by examining how the energetics of the topologies change with respect to the new phase angle Ψ_1 and the previous Ψ_2. These angles are determined by the straight DNA linker regions. Adding a base pari before or after the A-tract shifts the phase angle by $\sim 34°$. We see a much larger contrast in terms of enthalpic contribution to the J factor for the A-tract sequences, we even see that for some cases, the straight DNA is preferred, although for the vast majority of cases, the A-tracts have a much lower energy and therefore higher J factor. We understand the regions where the straight DNA is preferred by considering the prebent sequence being oriented such that the bending required for loop formation is such that it needs to bend against its intrinsic curvatures, thus raising the loop formation energy dramatically.

Antiparallel Binding Topologies

For straight DNA in chapter 4 we observed that the $A1$ binding topology was relatively insensitive to this phase angle in terms of order of magnitude difference of the J factor, this fact is no longer generally true for these A-tract sequences. We do see that for $\Psi = 6.21\ rad$ a region of relative phase insensitivity, see fig. 5.3.

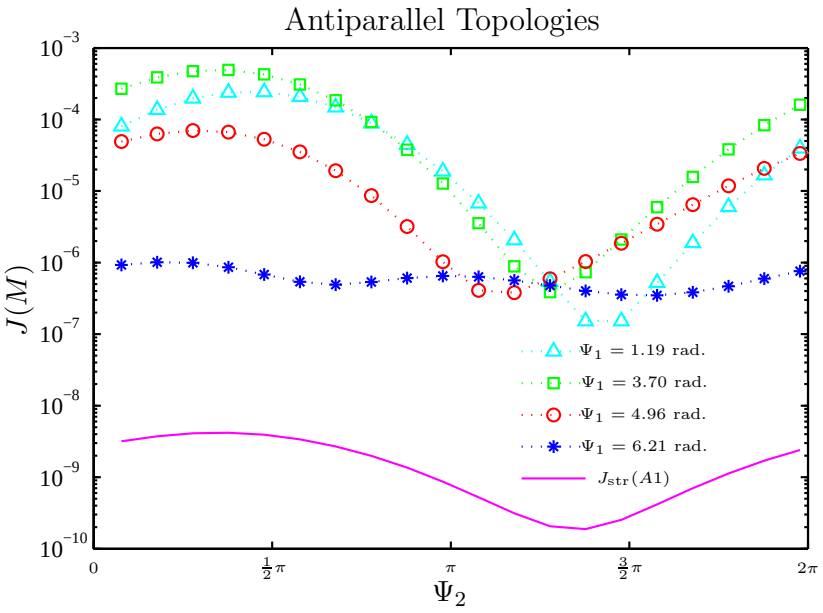

Figure 5.3 The J factor for the Antiparallel A-tract loops for various phase angles Ψ_1 are always higher than the best Antiparallel straight DNA of the same length (solid line). Here we see a strong dependence on Ψ_2, compared to the same length of straight DNA shown by the solid line. We also see that the J factor is 10^5 times larger for some configurations of Ψ_1 and Ψ_2. As we rotate through Ψ_1 we see a hill and valley structure, beginning with $\Psi_1 = 1.19$ rad. (triangles), we rise with $\Psi = 3.70$ rad. (squares) and then fall with $\Psi = 4.96$ rad. (circles) and then reach the valley near $\Psi_1 = 6.21$ rad. (asterisks).

Parallel Binding Topologies

The Parallel loops show a very similar phase dependence as the straight rod, although the magnitude of the difference in J factor is much larger for the pre-sequences, although in general for all values of Ψ_1 and Ψ_2, the pre-bending increases the likelihood of DNA loop formation, see fig. 5.4. This consistency of looping probability for the parallel loops make these topologies competitive with antiparallel topologies, as opposed to the straight rod where the antiparallel was preferred on this length scale for nearly all phase angles.

5.5 Preferred Looping Topologies

We now compare the two binding topologies for several combinations of Ψ_1 and Ψ_2 We find that we agree with the FRET determination of the loop binding topologies measured by Morgan et al. [38], see fig. 5.5

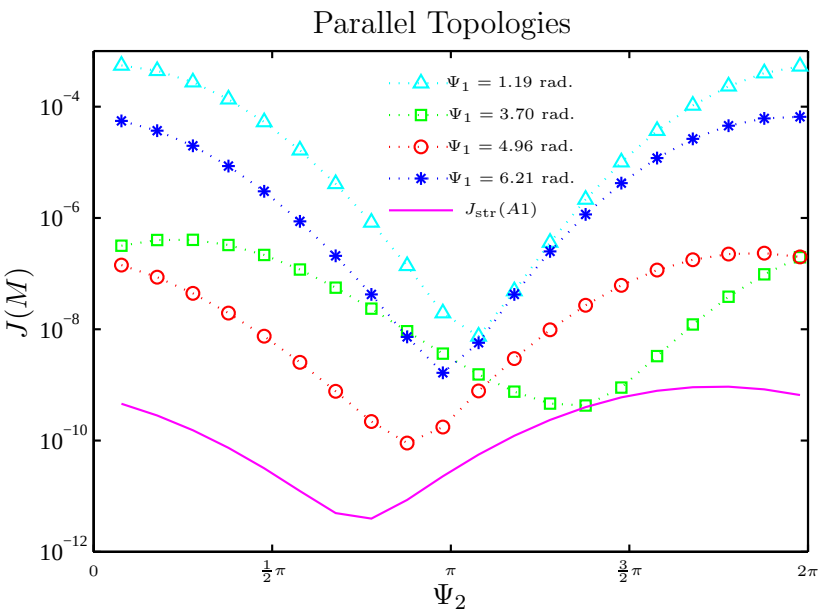

Figure 5.4 The J factor for the Parallel A-tract loops for various phase angles Ψ_1 are generally higher than the best Antiparallel straight DNA of the same length (solid line). Here we continue to see a strong dependence on Ψ_2. This dependence is larger in magnitude than for straight DNA P1 loops. We also see that the J factor is 10^6 times larger for some configurations of Ψ_1 and Ψ_2. As we rotate through Ψ_1 we see a hill and valley structure, beginning with $\Psi_1 = 1.19$ rad. (triangles), we fall with $\Psi = 3.70$ rad. (squares) and then fall further with $\Psi = 4.96$ rad. (circles) and then climb back up near $\Psi_1 = 6.21$ rad. (asterisks). We have not yet found any regions of phase insensitivity for $P1$ loops.

The FRET data of Morgan et al. [38] and Loop Formation Energy consideration of Lillian [29] indicate that 9C14 is a $P1$ loop, although the J factor calculation reveals that there may also be a very probable $A1$ loop as well. The reason for this dual binding topology comes from considering the over- and under-twisted DNA binding topologies for $A1$. Neither of the states are as probable as $P1$, although their combined probability is comparable, see fig. 5.6

5.5.1 Loop Lifetimes

The Gel electrophoresis experiments of Mehta et al. [37] shows that the loop stability of these A-tract sequences is much higher than that of straight DNA. We can use the J factor to compute the ratio of open τ^o to looped τ^ℓ lifetimes by

$$\frac{\tau^\ell}{\tau^o} = \frac{J}{c_o} \exp\left(\frac{\epsilon}{k_B T}\right) \tag{5.2}$$

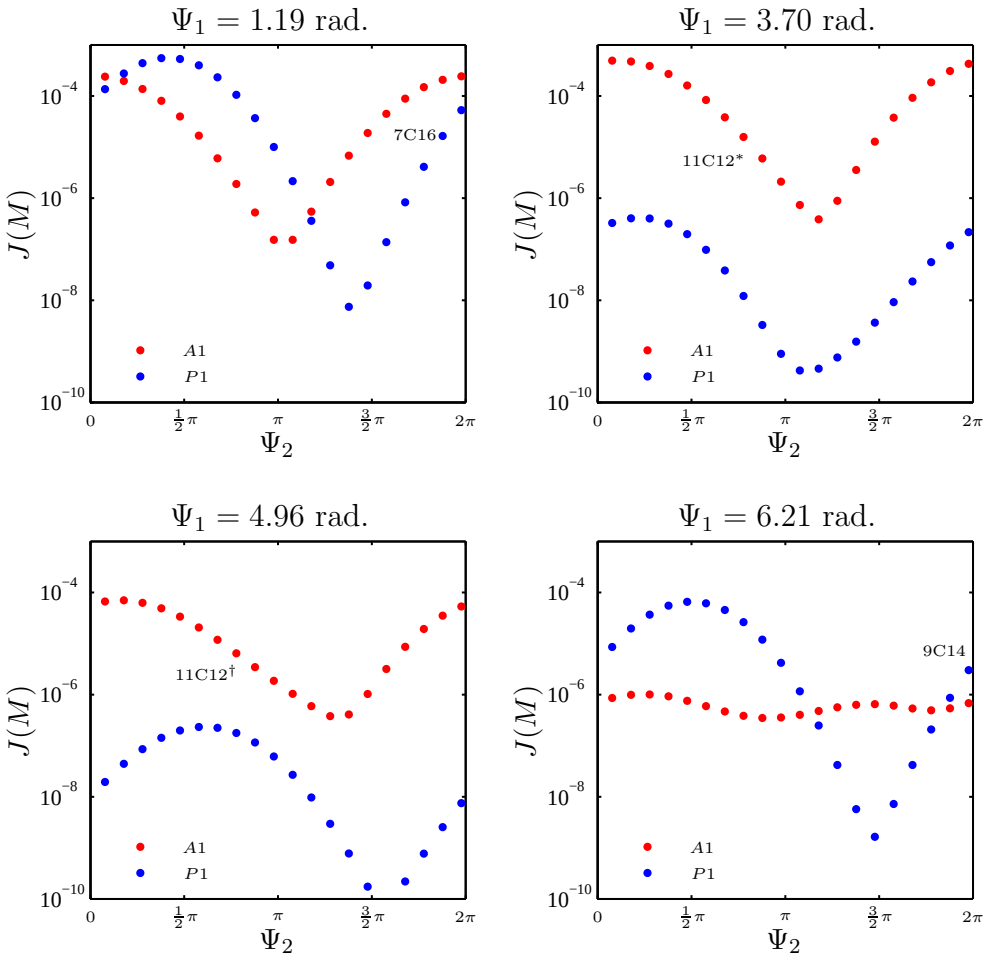

Figure 5.5 Comparing the J factors for several values of Ψ_1 and Ψ_2 we are able to indicate the expected topologies for the A-tract constructs of Mehta et al. [37]. We verify the FRET results of Morgan et al. [38]. We find taht 7C16, on the top left plot is $A1$ binding, and that 11C12 is most probably $A1$ binding as well. The 11C12 sequence lies between $\Psi = 3.70$ rad. and the $\Psi = 4.96$ rad. plots, at $\Psi_1 = 4.5$ rad. and $\Psi_2 = 1.7$ rad. The third sequence, 9C14 is seen on the bottom right plot as $P1$ binding, although the difference between $A1$ and $P1$ is small. We will look closer at 9C14 in fig. 5.6.

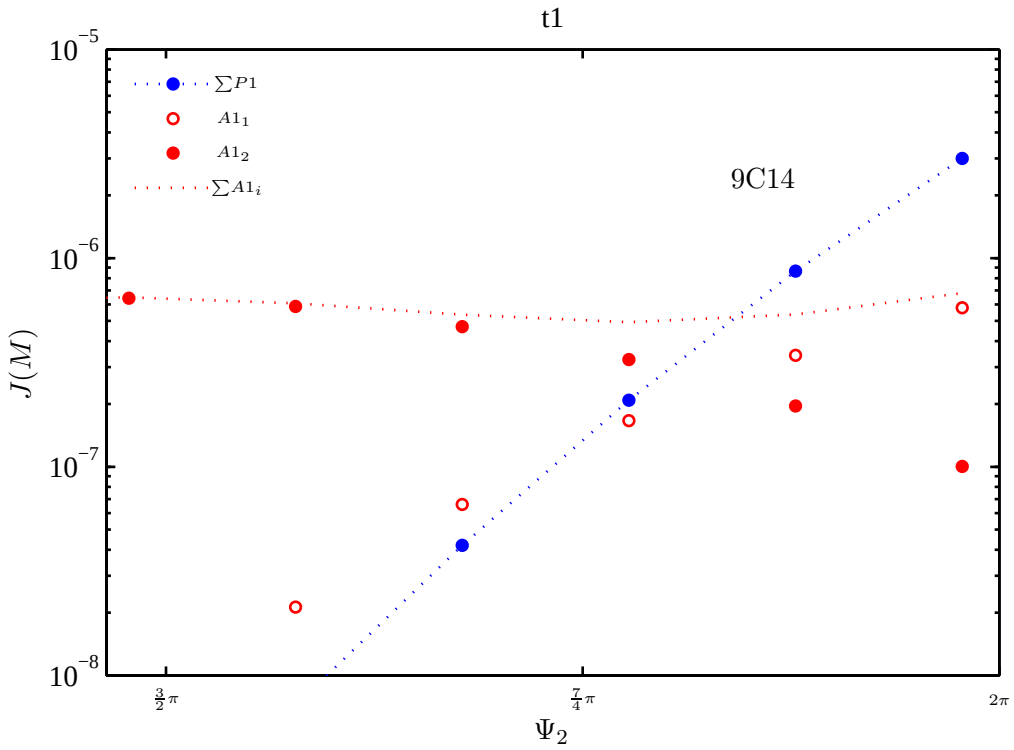

Figure 5.6 The J factor for the 9C14 sequence indicated the most probable loop topology as being $P1$, however careful consideration of the two possible $A1$ states, for over- and under-twisted DNA binding reveals that the $\sum A1_i$ states roughly $\frac{1}{2}$ as probable as $P1$ making it more competitive than either of the two individual states as previously considered. One of the strengths of our model is locating the regions where two topologies can coexist.

where ϵ is the binding strength of the LacI protein and DNA operators, and c_o is a reference concentration, normally take to be 1 $\mathcal{M}$. While we do not know the precise value of ϵ, it is only a function of operator strength and not binding topology. Therefore we can write a ratio of looped lifetimes for the A-tracts to Intrinsically Straight DNA using their respective J factors

$$\frac{\tau_A^\ell}{\tau_A^o} = \frac{J_A}{J_S}\frac{\tau_S^\ell}{\tau_S^o} \sim 10^5 \frac{\tau_S^\ell}{\tau_S^o} \tag{5.3}$$

where τ_A^ℓ/τ_A^ℓ is the ratio of looped to open lifetimes for the A-tract sequences, and τ_S^ℓ/τ_S^ℓ is the ratio of looped to open lifetimes for the straight sequences. The measurements by Mehta et al. indicate that the A-tracts bound to lac repressor are stable for at least 24 hours, while the straight DNA was stable for about 30 minutes. These results could be further tested through TPM measurements of different A-tract sequences and then comparing the ratio of their lifetimes.

5.6 Conclusions

We present the the first semi-analyitc computations of J factors for sequences containing intrinsic curvatures. It is important to remind the reader that we have not taken into account the spatial location of the lac repressor protein, therefore some of these topologies may be physically unavailable due to steric blocking by the lac repressor. However, throughout the design space, the protein mediated A-tract loops show a greatly enhanced J factor, on the order of 10^5 higher than equivalent straight sequences of the best binding conditions. In general, the A-tract induced curvature facilitates DNA loop formation, even in cases where phasing effects are quite large. It appears that the worst phased A-tract has at least one binding topology which is comparable in J factor as the best equivalent straight DNA for DNA of the order of 1 persistence length. The methods introduced here could also be used to compute looping for systems which contain architectural proteins *e.g.* Hu or IHF, by asking how probably is loop formation after the architectural protein binds to the DNA, assuming the bending and torsional stiffness is the bound region were reasonably well known.

Chapter 6

Extensions

6.1 Future Extensions

We propose a few ways that the our model can be extended. These topics are roughly listed in order of increasing challenge. We begin by discussing Heterogeneous elasticity, which we are able to capture in the rod model as well as the normal mode analysis, although our current results are very preliminary.

6.1.1 Heterogeneous Elasticity

We have modeled DNA as a homogeneous molecule with uniform bending stiffness, while in reality, DNA has a major and minor groove that have different bending stiffness values. Experimentally it is very challenging to measure the two stiffness values separately, so bulk or averaged effective persistence lengths ℓ_p are measured instead. Kehrbaum et al. [22] have shown through computational results that the difference from heterogeneous stiffness quickly averages out quickly past a few helical repeats. Others have challenged this assumption in the presence of off-diagonal coupling terms in the stiffness tensor [53, 55]. A clear experimental system which shows the difference in these two assumptions, or the necessity of a more complicated structure for the stiffness tensor $B(s)$ has not yet emerged.

We should be able to provide clear predictions for experimental systems, based on small variations in stiffness values, or helical phasing in a computationally efficient manner. To this end we have begun computing equilibrium configurations for planar loops as well as lac repressor loops. Our computations thus far find that for loops as short as 42 base pairs, there is no significant loop formation energy difference from

those with homogeneous stiffness for a diagonal stiffness tensor $B(s)$. Computations of normal modes have shown similar phase dependence, however a full computation of these effects is still in progress.

Approach

The typical effective bending stiffness is 50 nm. Monte Carlo results by Towles et al [55] have incorporated heterogeneous stiffness values based on the di-nucleotide model of Coleman et al [9], given our good agreement with the Monte Carlo torsionally minimum solutions, we have chosen $\ell_1 = 67\ nm$, $\ell_2 = 37\ nm$, which yield an effective sequence stiffness of 50 nm. We compute the effective bending stiffness by

$$\frac{1}{\ell_p^*} = \frac{1}{2}\left(\frac{1}{\ell_1} + \frac{1}{\ell_2}\right) \tag{6.1}$$

where the form resembled a parallel circuit, although for $\ell_1 = \ell_2$, the stiffness should be ℓ_p, therefore the factor of $1/2$ is added. These computations are currently ongoing and results are expected shortly. In the future, we will also add off-diagonal terms to $B(s)$, such as twist-bend coupling $\ell_{1\tau}$, and characterize their influence to DNA looping.

6.1.2 Sequence Dependent Elasticity

Sequence dependence elasticity can be spatially localized as regions with many AT base pairs tend to be softer than those with many GC base pairs. The reason for this difference is AT pairs only have two hydrogen bonds, while GC have three hydrogen bonds. As mentioned in chapter 1 DNA in eukaryotic systems is compactified using histone binding proteins, which form nucleosomes. It has been suggested [31, 60] that these sequences of DNA which bind to the histones are more flexible than wild-type DNA. The rod model can be used to compute the equilibrium shapes under prescribed boundary conditions of these systems with local or overall softening of the persistence lengths, and our J factor computation is readily capable of this computation. In chapter 5 we saw that sequence dependent curvature was able to dramatically increase the J factor and its variations based on phase angles. The parameters for these local stiffness values have been proposed by several models [33, 25, 9, 59].

6.1.3 Protein Binding Potentials

Currently the J factor computation assumes the protein-DNA binding is very rigid, and should be extended to include a full treatment of the binding protein. This change could be accomplished by replacing the delta functions which hold the DNA rigidly to the protein with more realistic binding potentials of the head groups which bind the DNA operator segments. This addition should be relatively straightforward, provided the binding potentials are known and easily expressed into angular coordinates. The extension of this work to include proteins would necessitate a more complete description of the DNA-protein binding regions which would involve modifying our $\delta(\vec{u})$ functions above in eq. (2.2) of chapter 2 as was done in previous Monte Carlo works [65, 53, 48].

6.1.4 Architectural Proteins

We learned in chapter 5 that intrinsic curvature can greatly lower the energetic barrier required for looping and dramatically increase the J factor. There are many proteins which bind specifically and non-specifically to DNA and induce curvature to the molecule. *E. coli*has several of these proteins, *e.g.* HU, IHF and CAP-cAMP to name a few. It is known that CAP-cAMP directly influences production of β-galactosidase, as it participates in the selective digestion of Glucose over Lactose [11]. It should be possible to model the influence of these proteins as intrinsic curvature and stiffness of the DNA, if self-contact is also taken into account with the protein. These types of systems are a natural extension of the J factor computation.

6.1.5 Multi-scale model

Once the protein binding potentials are incorporated, the rod model could be interfaced with a molecular dynamics simulation of the LacI protein. Then updates of the protein location and orientation could affect the rod model equilibria and provide a more accurate measure of the available thermal modes of the system, leading to a more complete computation of the entropic cost of DNA-protein loop formation. Multi-scale computations of this form have done by Baeleff et al. [3, 2, 58, 57, 4], however, they only treated the DNA as a simple elastic rod, without sequence dependent elasticity or stiffness, and similarly neglected entropic effects, and were therefor unable to compute J factors.

6.1.6 Self-Contact and Electrostatics

The rod model [16] has been extended by Lillian et al. [28] to include self-contact of DNA as well as electrostatic interactions of the DNA molecule. We have not however included these effects in the equilibria that are presented here, as they increase the computational cost. When sweeping out phase angles, one DNA end-point is rotated through 4π, turning off self-contact allows the rod to pass through itself, without getting tangled and twisted. If an equilibrium required self contact, the normal modes as computed here would be insufficient to handle this self-contact, as they change behavior dramatically. The contribution from electrostatics would affect the various equilibrium topologies differently, as $P1$ tend to have the DNA strands in close proximity, and they would therefor experience more repulsion and higher loop formation energies. Electrostatic coupling for thermal excitations is likely very small, and the normal modes would not be greatly affected.

6.2 Experimental Systems

We propose a few experimental methods which can test our predictions.

6.2.1 Cyclization

There is a lot of interest in measuring and understanding J factors for DNA on short length scales $\sim$ 100 base pairs [61, 14, 8, 13]. It has been suggested that end-point constraints play a role in these differing results [54]. It is also possible that sequence-dependent elasticity plays a role as well. Our computation can, with the appropriate inputs, quickly map out the J factor landscape for DNA on this length as a function of sequence-dependent elasticity and be used to predict cyclization rates, which can be measured in a straightforward manner.

6.2.2 TPM measurements of Ratios of Loop Lifetimes.

Tethered Particle Microscopy (TPM) can be used to measure the looped and unlooped lifetimes of DNA tethered to a microsphere in the presence of lac repressor [51, 56, 41, 18, 7]. We can predict the ratio of looped lifetimes for different sequences and predict their relative ratios. While we do not know the exact binding affinity of

DNA operators and lac repressor, we can form a ratio of the looped lifetimes as

$$\frac{\tau_A^\ell}{\tau_A^o} = \frac{J_A}{J_S}\frac{\tau_S^\ell}{\tau_S^o} \sim 10^5 \frac{\tau_S^\ell}{\tau_S^o} \tag{6.2}$$

see chapter 5 for more information.

6.2.3 FRET Measurements of Topology

We are able to predict the preferred binding topologies based on our J factor computations. These predictions can be tested using Fluorescent Energy Resonance Transfer (FRET) by attaching Fluorescent markers Cy3-Cy5, which are appropriate for typical separation distances on this length scale, before and after the DNA operators. The FRET efficiency measured in bulk systems would then be compared to our computation of the preferred topology. We are also able to identify lengths of DNA where $A1$ and $P1$ should occur in relatively equal proportions, which can also be demonstrated using FRET.

6.2.4 Inverse Rod Modeling and Extraction of physical DNA parameters

The largest barrier between model results and experimental comparison is the lack of precise biophysical parameters for DNA. It is unknown precisely how much stiffer a region of GC is to AT, or how much twist and tilt of the DNA backbone is affected by sequence, although there have been several models which attempt to extract these parameters, such as the di-nucleotide [9]and tri-nucleotide models [40]. It should be possible to alter the inputs for stiffness and curvature using the rod model to sweep out an entire range of J factors based on different initial conditions, and then to in tandem run Monte Carlo simulations to numerically connect the J factors and test the reasonability of these input conditions. In this way, we can extract the effect of variable sequence to the looping probability directly. With enough computations, it should be possible to extract up to a few base pair resolution effects of sequence dependence.

Chapter 7

Summary of Contributions

In this dissertation we have made significant contributions towards the understanding of the role of enthalpy and entropy in determining the Stockmayer Jacobson J-factor. In chapter 3, we studied DNA looping in the absence of a DNA binding protein. In chapter 4, we studied intrinsically straight DNA bound to LacI protein and in chapter 5 we studied the LacI binding protein with the addition of sequence specific DNA curvature. In chapter 6 we explored extensions and experimental tests of our generalized J factor computation.

7.1 Major Contributions

We present a semi-analytic Generalized J factor computation that

1. computes the normal modes of the open and looped states of DNA, allowing us to quantify the entropic cost of DNA loop formation and subsequently compute of the J factor, see chapter 2.
2. allows arbitrary DNA end-point locations and orientations, such as protein mediated loop formation.
3. fully includes entropic effects and allows a prediction of physical quantities without the use of tuning parameters, see chapter 2 and chapter 4.
4. independently computes the enthalpic and entropic cost of DNA loop formation.
5. verifies the results of Shimada and Yamakawa [49] for the ring and unconstrained loop, settling the debate raised by Zhang et al. [64] as to whether or not the original paper was in error.
6. is capable of incorporating sequence-dependent curvature $\vec{\kappa}^o$, *e.g.* A-tract sequences of Mehta et al. [37].

7. is able to predict ratios of loop lifetimes for straight DNA as well as A-tract sequences.
8. is capable of incorporating sequence dependent and heterogeneous stiffness $\ell_1(s) \neq \ell_2(s)$.
9. is very computationally efficient. For a given set of boundary conditions, elasticity and/or curvature our computation requires only a few minutes.

7.2 Specific Findings

1. We have predicted J factors for $A1$ and $P1$ binding topologies for intrinsically straight DNA for lengths 50 to 650 base pairs in length based on experimental measurements of DNA persistence length, which does not rely on any free parameters, see chapter 4.
2. We have shown our results for intrinsically straight DNA are consistent with previous Monte Carlo work by Towles et al [55], see chapter 4
3. We confirmed the preference for $A1$ binding topology for DNA loops under 200 base pairs for intrinsically straight DNA, which contradicts the popular image of the $P1$ loop as shown on the Science cover in 1996 by Lewis et al. [26], see chapter 4.
4. Predicted J factors for A-tracts which are up to 10^5 times greater than equivalent straight DNA, which supports the Gel electrophoresis data of Mehta et al. [37], see chapter 5.
5. We have made specific predictions about the relative looping probabilities between $A1$ and $P1$ loops which may be measured using FRET, see chapter 4 and chapter 5.
6. We have predicted that the 9C14 sequence of Morgan et al. [38] has 2 comparable topologies, when an argument based on loop formation energy alone would typically indicate only one, see chapter 5.
7. We have shown that $P1$ and $A1$ are allowable for DNA containing A-tract bends near one persistence length, which is typically dominated by $A1$ for straight DNA, our results agree with the measurements of Mehta et al. [37], see chapter 5.
8. We have reproduced with our semi-analytic approach the ring and unconstrained loop results of Shimada and Yamakawa [49], in contrast to the recent Monte Carlo work of Zhang and Crothers [63], see chapter 3.

9. We generalized the J factor for planar loops, covering the ring to the teardrop and continuing on to the hairpin. We have presented a convenient numeric approximation which can be easily used to compute planar J factors, see chapter 3.
10. We have computed the effective torsional persistence length ℓ_τ for planar shapes and clearly demonstrated how bending and torsional stiffness couple, see chapter 3.
11. We have found that the enthalpic energy and trends in the J factor for DNA with $\ell_1 \neq \ell_2$ are consistent with $\ell_1 = \ell_2$ for DNA as short as 49 base pairs, validating our simplification to homogeneous bending stiffness when considering DNA beyond a few helical repeats, see chapter 6.

Appendices

Appendix A

Notation

"Consistency of notation is the Hobgoblin of a small mind."
— Lowell Brown, University of Washington 2001.

All inputs to the model are made dimensionless by multiplying by the overall sequence length L and $\beta = \frac{1}{k_B T}$. The dimensionless quantities we use are listed in the variable descriptions. We use o to designate open, or unlooped state quantities. We use ℓ to designate looped or closed state quantities.

A.1 Latin Alphabet Notation

Variable	Descriptions
$B(s)$	Stiffness Tensor, $[\beta B(S)L^{-1}]$
c_f	Higher order force term of Shimada and Yamakawa, $c_f = \frac{L}{4\ell_p}$
E_ℓ	Loop formatin energy, $[\beta E_\ell]$
ΔE	The equilibrium thermal energy of DNA.
$\mathcal{H}_{mn}$	The (m, n) Hamiltonian matrix operator.
$\tilde{\mathcal{H}}$	The integrated product of the open or looped Hamiltonian matrix operators and comparison functions as outline in the Ritz Method, see Appendix D.
$H^{o,\ell}$	Scalar Hamiltonian for the the open and looped state, $[\beta H]$.
J	The Stockmayer Jacobson J factor.
k_B	Boltzmann constant.
L	The total inter-operator DNA length.

Variable	Dimensionless Form Descriptions
$\mathcal{L}_{mn}$	The (m, n) Linear operator.
$\tilde{\mathcal{L}}$	The integrated product of Linear Operators and comparison functions as outline in the Galerkin Method, see Appendix C.
ℓ_p	The average bending persistence length of DNA, $[\ell_p L^{-1}]$.
ℓ_τ	The average torsional persistence length of DNA, $[\ell_\tau L^{-1}]$.
ℓ_τ^*	The effective torsional persistence length of DNA due to equilibrium state twist-bend coupling, $[\ell_\tau^* L^{-1}]$.
ℓ_1	Persistence length along the minor groove of DNA, $[\ell_1 L^{-1}]$.
ℓ_2	Persistence length along the major groove of DNA, $[\ell_2 L^{-1}]$.
$\mathcal{M}$	Molarity = 1 mol/Liter.
N	The number of basis functions considered in our approximation.
$\mathcal{N}_A$	Avogadro's Number.
$\hat{n}_1(s)$	Equilibrium unit normal vector, directed along the minor groove.
$\tilde{n}_1(s)$	Deformed state unit normal vector, directed along the minor groove.
$\hat{n}_2(s)$	Equilibrium unit normal vector, directed along the major groove.
$\tilde{n}_2(s)$	Deformed state unit normal vector, directed along the major groove.
$\mathcal{O}(0, 1, 2)$	Order parameter for deformation variables *e.g* $\kappa\psi$ is order 1.
$\vec{R}(s)$	Equilibrium State position vector of the DNA sugar-phosphate backbone.
$\tilde{R}(s)$	Deformed State position vector of the DNA sugar-phosphate backbone.
$R(\theta_1)$	Deformation Rotation matrix about the $\hat{n}_1$ direction.
$R(\theta_2)$	Deformation Rotation matrix about the $\hat{n}_2$ direction.
$R(\psi)$	Deformation Rotation matrix about the $\hat{t}$ direction.
s, S	The arc length parameter for the DNA, $s = (0, 1)$, $S = (0, L)$.
T	Temperature.
T_w	Twist of the DNA molecule.
ΔT_w	Difference in twist from the looped to open state, $\int_0^1 \mathrm{d}s \left(\tau^\ell(s) - \tau^o(s)\right)$
$\hat{t}$	Equilibrium unit tangent vector, directed along sugar-phosphate backbone.
$\tilde{t}$	Deformed state unit tangent vector, directed along sugar-phosphate backbone.
$\vec{u}$	Deformed state displacement vector from the equilibrium state, $\tilde{R} = \vec{R} + \vec{u}$.

Variable	Description
u_1	Deformed state displacement along the tangent $\hat{t}$.
u_2	Deformed state displacement along the normal $\hat{n}_1$.
u_3	Deformed state displacement along the normal $\hat{n}_2$.
V	The total constraint matrix formed from our δ functions.
V_i	The constraint matrix for the ξ_i eigenmode.
v_{1i}	The displacement along the equilibrium tangent vector $\hat{t}$ modulo the eigenmode amplitude ξ_i and length L.
v_{2i}	The displacement along the equilibrium normal vector $\hat{n}_1$ modulo the eigenmode amplitude ξ_i and length L.
v_{3i}	The displacement along the equilibrium normal vector $\hat{n}_2$ modulo the eigenmode amplitude ξ_i and length L.
Z	The open state partition function.
Z_ℓ	The looped or closed state partition function.
$\tilde{Z}_\ell$	The looped or closed state partition function modulo $\exp(-\beta E_\ell - c_f)$.

A.1.1 Greek Alphabet Notation

Variable	Description
$\boldsymbol{\alpha}, \alpha_i, \alpha^o, \alpha^a$	Unknown coefficients of the comparison functions ϕ.
β	inverse Boltzmann Constant and Temperature, $\beta = (k_B T)^{-1}$.
Γ	Arbitrary boundary condition term, $\Gamma \triangleq \int \mathrm{d}s\ \phi^T \mathcal{L} \gamma$, see chapter 2.
$\boldsymbol{\gamma}$	Matrix of comparison functions for an arbitrary angular displacement.
$\gamma(\Theta)$	The slowly varying function of the in plane J factor as a function of Loop
$\delta(\theta_1)$	Constraint on θ_1 at $s = 0$ and $s = 1$
$\delta(\theta_2)$	Constraint on θ_2 at $s = 0$ and $s = 1$
$\delta(\psi)$	Constraint on ψ at $s = 0$ and $s = 1$
$\Delta\kappa_1$	Difference of looped and open state curvature $\Delta\kappa_1 = \kappa_1^\ell - \kappa_1^o$
$\Delta\kappa_2$	Difference of looped and open state curvature $\Delta\kappa_2 = \kappa_2^\ell - \kappa_2^o$

Variable	Description
$\boldsymbol{\delta\theta_1}$	Variation of deformation variable θ_1
$\boldsymbol{\delta\theta_2}$	Variation of deformation variable θ_2
$\boldsymbol{\delta\psi}$	Variation of deformation variable ψ
ζ	Scaling or Amplitude of the arbitrary displacement mode.
Θ	End-point planar component of the tangent vector alignment angle.
θ_1	Deformation angle about $\hat{n}_1$
θ_2	Deformation angle about $\hat{n}_2$
θ_{1L}	End-point angular displacement about $\hat{n}_1$
θ_{2L}	End-point angular displacement about $\hat{n}_2$
κ_1	First component of the equilibrium curvature vector $[\vec{\kappa}L]$
κ_1^o	First component of the open state equilibrium curvature vector $[\kappa_1^o L]$
κ_1^ℓ	First component of the looped state equilibrium curvature vector $[\kappa_1^\ell L]$
κ_2	Second component of the equilibrium curvature vector $[\kappa_2 L]$
κ_2^o	Second component of the open state equilibrium curvature vector $[\kappa_2^o L]$
κ_2^ℓ	Second component of the looped state equilibrium curvature vector $[\kappa_2^\ell L]$
$\vec{\kappa}^o$	Equilibrium open state curvature vector, $[\vec{\kappa}^o L]$
$\tilde{\kappa}^o$	Deformed state open state curvature vector, $[\tilde{\kappa}^o L]$
$\vec{\kappa}^\ell$	Equilibrium state looped state curvature vector, $[\vec{\kappa}^\ell L]$
$\tilde{\kappa}^\ell$	Deformed state looped state curvature vector, $[\tilde{\kappa}^\ell L]$
Λ	The Entropic Coefficient of the J factor.
Λ^o	The open state component of the Entropic Coefficient of the J factor.
Λ^ℓ	The looped state component of the Entropic Coefficient of the J factor.
$\lambda_i^{o,\ell}$	The eigenvalue of the eigenvector $\vec{\xi}_i$.
λ_0^ℓ	The zero eigenvalue of the ring.
λ_1	The lowest open or looped state eigenvalue.

Variable	Description
μ_1	Conjugate variable of u_1.
μ_2	Conjugate variable of u_2.
μ_3	Conjugate variable of u_3.
σ	Ratio of torsional to bending persistence length $\frac{\ell_\tau}{\ell_p}$.
τ	Third component of the equilibrium curvature vector $[\tau L]$
τ^o	Third component of the open state equilibrium curvature vector $[\tau^o L]$
τ^ℓ	Third component of the looped state equilibrium curvature vector $[\tau^\ell L]$
$\tilde{\tau}$	Third component of the deformed state curvature vector $[\tilde{\tau} L]$
$\Delta\tau$	Difference of looped and open state torsion $\Delta\tau = \tau^\ell - \tau^o$
Φ	Out of plane Loop formation angle.
ϕ_n	The n^{th} comparison functions.
$\boldsymbol{\phi}$	Matrix of known comparison functions.
ξ	Arbitrary deformation to the equilibrium state.
$\vec{\xi}_i$	An eigenvector or Normal mode of the open or looped state.
ξ_i	Amplitude of eigenvector $\vec{\xi}_i$.
ξ^o	Component of an arbitrary end-point displacement Amplitude which leaves the angular displacement zero at the end-point.
ξ^a	Component of an arbitrary end-point displacement Amplitude which leaves the angular displacement non-zero at the end-point.
Ψ	Torsional Loop formation angle, referred to as a phase angle.
$\Psi_{1,2}$	The two phase angles required for A-tract loop formation.
ψ	Deformation angle about $\hat{t}$
ψ_L	End-point angular displacement about $\hat{t}$
ω	Helical frequency of DNA.
ρ	Dimensional scaling of the J factor.

A.2 J factor units

The J factor is often expressed in terms of Molarity. The units of Molarity are moles per liter, $1\ L = 1\ dm^3 = 10^3\ cm^3 = 10^{24}\ nm^3$. The conversion is then $1\text{mole} = 6.0221 \times 10^{23}\text{particles}$. We calculate the J factor in terms of molecules/nm^3. Then so our conversion is $10^{24}\ nm^3/6.0221 \times 10^{23}\text{molecules} \approx 1.661\ nm^3$.

$$J\left(\frac{1\text{ molecule}}{nm^3}\right) \times \frac{10^{24} nm^3}{1L} \times \frac{1\text{Mole}}{\mathcal{N}_A} = 1.661\ J\left(\frac{\text{Moles}}{\text{Liter}}\right) \tag{A.1}$$

where $\mathcal{N}_A$ is Avogodro's Number.

Appendix B

Rotations

We keep up to second order in terms of deformation variables θ_1, θ_2, ψ which are induced by our rotation matrices $R(\theta_1), R(\theta_2), R(\psi)$, and therefore the rotation matrices do not commute. There are 6 possible applications for the rotation matrices to induce our deformations of the elastic rod. The rotation matrices we use are

$$R(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-\frac{1}{2}\psi^2 & \psi \\ 0 & -\psi & 1-\frac{1}{2}\psi^2 \end{bmatrix}$$

$$R(\theta_1) = \begin{bmatrix} \cos\theta_1 & 0 & -\sin\theta_1 \\ 0 & 1 & 0 \\ \sin\theta_1 & 0 & \cos\theta_1 \end{bmatrix} \sim \begin{bmatrix} 1-\frac{1}{2}\theta_1^2 & 0 & -\theta_1 \\ 0 & 1 & 0 \\ \theta_1 & 0 & 1-\theta_1^2 \end{bmatrix}$$

$$R(\theta_2) = \begin{bmatrix} \cos\theta_2 & \sin\theta_2 & 0 \\ -\sin\theta_2 & \cos\theta_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1-\frac{1}{2}\theta_2^2 & \theta_2 & 0 \\ -\theta_2 & 1-\frac{1}{2}\theta_2^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

keeping to second order enforces the in-extensibility constraint on the DNA. The rotation order we have chosen is $R(\psi), R(\theta_1), R(\theta_2)$

$$\begin{bmatrix} \tilde{t} \\ \tilde{n}_1 \\ \tilde{n_2} \end{bmatrix} = \begin{bmatrix} 1-\frac{1}{2}\left(\theta_1^2+\theta_2^2\right) & \theta_2 & -\theta_1 \\ -\theta_2+\theta_1\psi & 1-\frac{1}{2}\left(\theta_2^2+\psi^2\right) & \psi \\ \theta_1+\theta_2\psi & -\psi+\theta_1\theta_2 & 1-\frac{1}{2}\left(\theta_1^2+\psi^2\right) \end{bmatrix} \begin{bmatrix} \hat{t} \\ \hat{n}_1 \\ \hat{n}_2 \end{bmatrix}. \tag{B.1}$$

Which leads to the correct definition of the deformation angles, such that ψ is the angle between $\tilde{n}_1$ and $\hat{n}_2$. In addition θ_2 is the angle about which $\hat{n}_1$ rotates to produce $\tilde{n}_1$ and careful consideration shows for a right handed coordinate system, θ_1 to be negative. An example of a rotation which does not produce the correct definitions

of the phase angles is

$$R(\psi)R(\theta_2)R(\theta_1) \sim \begin{bmatrix} 1-\frac{1}{2}\left(\theta_1^2+\theta_2^2\right) & \theta_2 & -\theta_1 \\ -\theta_2+\theta_1\psi & 1-\frac{1}{2}\left(\theta_2^2+\psi^2\right) & \psi+\theta_1\theta_2 \\ \theta_1+\theta_2\psi & -\psi & 1-\frac{1}{2}\left(\theta_1^2+\psi^2\right) \end{bmatrix} \tag{B.2}$$

where we see that ψ is incorrect.

Appendix C

Linear Operators

C.1 Linear Operators

The variation of the Hamiltonian and subsequent collection of variables in terms of variations of the deformation variables $\boldsymbol{\delta}\theta_1, \boldsymbol{\delta}\theta_2, \boldsymbol{\delta}\psi$ leads to the following linear operators and boundary terms.

C.1.1 Explicit form of the Linear Operators

The linear operators are

$$
\begin{aligned}
L_{11}\theta_1 =& 2\left[-\ell_1\right]\theta_1'' + 2[-\ell_1']\theta_1' + 2[\ell_2(\tau^2 - \kappa_2\Delta\kappa_2) + \ell_\tau(\kappa_2^2 - \tau\Delta\tau)]\theta_1 \\
L_{12}\theta_2 =& 2[(\ell_1 + \ell_2)\tau - \ell_\tau\Delta\tau]\theta_2' + 2[(\ell_1\tau)' + \ell_2\kappa_1\Delta\kappa_2 - \ell_\tau\kappa_1\kappa_2]\theta_2 \\
L_{13}\psi =& 2[\ell_2\Delta\kappa_2 - \ell_1\kappa_2 - \ell_\tau\kappa_2]\psi' + 2[-(\ell_1\kappa_2)' + \ell_1\tau\Delta\kappa_1 + (\ell_2\Delta\kappa_2)' - \ell_2\kappa_1\tau]\psi \\
L_{21}\theta_1 =& 2[-(\ell_1 + \ell_2)\tau + \ell_\tau\Delta\tau]\theta_1' + 2[-(\ell_2\tau)' + \ell_2\kappa_1\Delta\kappa_2 + (\ell_\tau\Delta\tau)' - \ell_\tau\kappa_1\kappa_2]\theta_1 \\
L_{22}\theta_1 =& 2[-\ell_2]\theta_2'' + 2[-\ell_2']\theta_2' + 2[\ell_1(\tau^2 - \kappa_1\Delta\kappa_1) + \ell_\tau(\kappa_1^2 - \tau\Delta\tau)]\theta_2 \\
L_{23}\psi =& 2[-\ell_1\Delta\kappa_1 + \ell_2\kappa_1 + \ell_\tau\kappa_1]\psi' + 2[-(\ell_1\Delta\kappa_1)' - \ell_1\kappa_2\tau + (\ell_2\kappa)' + \ell_2\tau\Delta\kappa_2]\psi \\
L_{31}\theta_1 =& 2[\ell_1\kappa_2 - \ell_2\Delta\kappa_2 + \ell_\tau\kappa_2]\theta_1' + 2[\ell_1\Delta\kappa_1\tau - \ell_2\kappa_1\tau + (\ell_\tau\kappa_2)']\theta_1 \\
L_{32}\theta_2 =& 2[\ell_1\Delta\kappa_1 - \ell_2\kappa_1 - \ell_\tau\kappa_1]\theta_2' + 2[-\ell_1\kappa_2\tau + \ell_2\tau\Delta\kappa_2 - (\ell_\tau\kappa_1)']\theta_2 \\
L_{33}\psi =& 2[-\ell_\tau]\psi'' + 2[-\ell_\tau']\psi' + 2[\ell_1(\kappa_2^2 - \kappa_1\Delta\kappa_1) + \ell_2(\kappa_1^2 - \kappa_2\Delta\kappa_2)]\psi
\end{aligned}
\tag{C.1}
$$

The linear operators are self-adjoint,

$$\int_0^1 \mathrm{d}s\phi_m L_{mn}\phi_n = \int_0^1 \mathrm{d}s\phi_n L_{nm}\phi_m \tag{C.2}$$

which leads to a symmetric Hamiltonian, which guarantees us real eigenvalues, although no positive eigenvalues, which are corrected by our delta functions.

C.1.2 Boundary Terms

In the theory chapter, we added bold text to the variations $\boldsymbol{\delta}\theta$ to avoid confustion with the $\delta^3(|\vec{u}|)$ functions, since there is no risk of confusion here, we drop the additional bold text. We recall that for variation problems we take our end points to be fixed, $\delta\theta_1(0) = \delta\theta_1(1) = 0$, $\delta\theta_2(0) = \delta\theta_2(1) = 0$ and $\delta\psi(0) = \delta\psi(1) = 0$ The boundary terms are arising from $\delta\left(\kappa_1 - \kappa_1^o\right)^2$ are

$$\begin{aligned}
&\ell_1\Delta\kappa_1\delta\theta_1|_0^1 = 0\\
&\ell_1\theta_1'\delta\theta_1|_0^1 = 0\\
&\ell_1\tau\theta_2\delta\theta_1|_0^1 = 0\\
&\ell_1\kappa_2\psi\delta\theta_1|_0^1 = 0\\
&\ell_1\Delta\kappa_1\psi\delta\theta_2|_0^1 = 0
\end{aligned}$$

which all individually vanish. The boundary terms are arising from $\delta\left(\kappa_2 - \kappa_2^o\right)^2$ are

$$\begin{aligned}
&\ell_2\Delta\kappa_2\delta\theta_2|_0^1 = 0\\
&\ell_2\theta_2'\delta\theta_2|_0^1 = 0\\
&\ell_2\tau\theta_1\delta\theta_2|_0^1 = 0\\
&\ell_2\kappa_1\psi\delta\theta_2|_0^1 = 0\\
&\ell_2\Delta\kappa_2\psi\delta\theta_1|_0^1 = 0
\end{aligned}$$

which all individually vanish. The boundary terms are arising from $\delta\left(\tau-\tau^{o}\right)^{2}$ are

$$\begin{aligned}
&\ell_{\tau}\Delta\tau\delta\psi|_{0}^{1}=0\\
&\ell_{\tau}\psi'\delta\psi|_{0}^{1}=0\\
&\ell_{\tau}\kappa_{2}\theta_{1}\delta\psi|_{0}^{1}=0\\
&\ell_{\tau}\kappa_{1}\theta_{2}\delta\psi|_{0}^{1}=0\\
&\ell_{\tau}\Delta\tau\theta_{1}\delta\theta_{2}|_{0}^{1}=0
\end{aligned}$$

which all individually vanish. We therefore do not need to consider them further given our boundary conditions.

Appendix D

Hamiltonian Matrix Construction

We use the Ritz Method to discretize the Hamiltonian functional, which reduces it to a function over the coefficient space of the comparison functions. The construction of the Hamiltonian matrix is straightforward as outlined in chapter 2.

D.1 Ritz Method: Hamiltonian Matrix

The Hamiltonian Matrix is composed of the second order deformation variables, given by

$$\begin{aligned}\beta H(\mathcal{O}(2)) =&\frac{1}{2}\int_0^1 \mathrm{d}s\,\big[\ell_1(\theta_1'^2 - 2\tau\theta_2\theta_1' + 2\kappa_2\psi\theta_1' + 2\Delta\kappa_1\psi\theta_2' + \theta_2^2(\tau^2 - \kappa_1\Delta\kappa_1)\\ &+\psi^2(\kappa_2^2 - \kappa_1\Delta\kappa_1) + 2\Delta\kappa_1\tau\theta_1\psi - 2\kappa_2\tau\theta_2\psi)\big)\\ &+\ell_2(\theta_2'^2 + 2\tau\theta_1\theta_2' - 2\kappa_1\psi\theta_2' - 2\Delta\kappa_2\psi\theta_1' + \theta_1^2(\tau^2 - \kappa_2\Delta\kappa_2)\\ &+\psi^2(\kappa_1^2 - \kappa_2\Delta\kappa_2) + 2\kappa_1\Delta\kappa_2\theta_1\theta_2 - 2\kappa_1\tau\theta_1\psi + 2\tau\Delta\kappa_2\theta_2\psi)\big)\\ &+\ell_\tau(\psi'^2 - 2\kappa_2\theta_1\psi' + 2\kappa_1\theta_2\psi' - 2\Delta\tau\theta_1\theta_2' + \theta_1^2(\kappa_2^2 - \tau\Delta\tau)\\ &+\theta_2^2(\kappa_1^2 - \tau\Delta\tau) - 2\kappa_1\kappa_2\theta_1\theta_2)\big]\\ &=\frac{1}{2}\int_0^1 \mathrm{d}s\,\mathcal{H}\end{aligned} \tag{D.1}$$

Then we construct the Hamiltonian matrix directly through

$$\mathcal{H} = \begin{bmatrix}\theta_1 & \theta_2 & \psi\end{bmatrix}\begin{bmatrix}\mathcal{H}_{11} & \mathcal{H}_{12} & \mathcal{H}_{13}\\ \mathcal{H}_{21} & \mathcal{H}_{22} & \mathcal{H}_{23}\\ \mathcal{H}_{31} & \mathcal{H}_{32} & \mathcal{H}_{33}\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \psi\end{bmatrix} \tag{D.2}$$

D.1.1 Terms with no derivatives

The components of $\mathcal{H}$ which do not involve derivatives of the deformation variables, $\mathcal{H}^0_{ab}$ are symmetric by construction,

$$\begin{aligned}
\mathcal{H}^0_{11} &= (\ell_\tau - \ell_2)(\kappa_2^2 - \tau^2) + \ell_2\kappa_2\kappa_2^o + \ell_\tau\tau\tau^o \\
\mathcal{H}^0_{12} &= \mathcal{H}^0_{21} = (\ell_2 - \ell_\tau)\kappa_1\kappa_2 - \ell_2\kappa_1\kappa_2^o \\
\mathcal{H}^0_{13} &= \mathcal{H}^0_{13} = (\ell_1 - \ell_2)\kappa_1\tau - \ell_1\tau\kappa_1^o \\
\mathcal{H}^0_{22} &= (\ell_\tau - \ell_1)(\kappa_1^2 - \tau^2) + \ell_1\kappa_1\kappa_1^o + \ell_\tau\tau\tau^o \\
\mathcal{H}^0_{23} &= (\ell_2 - \ell_1)\kappa_2\tau - \ell_2\tau\kappa_2^o \\
\mathcal{H}^0_{33} &= (\ell_1 - \ell_2)(\kappa_2^2 - \kappa_1^2) + \ell_1\kappa_1\kappa_1^o + \ell_2\kappa_2\kappa_2^o
\end{aligned} \tag{D.3}$$

D.1.2 Terms with one derivatives

then we also have terms with a single derivative of the deformation variables, $\mathcal{H}'_{ab}$ and are not symmetric in this form,

$$\begin{aligned}
\mathcal{H}'_{12} &= (2(\ell_2 - \ell_\tau)\tau + 2\ell_\tau\tau^o)\frac{\mathrm{d}}{\mathrm{ds}} \\
\mathcal{H}'_{13} &= (-2\ell_\tau\kappa_2)\frac{\mathrm{d}}{\mathrm{ds}} \\
\mathcal{H}'_{21} &= (-2\ell_1\tau)\frac{\mathrm{d}}{\mathrm{ds}} \\
\mathcal{H}'_{23} &= (2\ell_\tau\kappa_1)\frac{\mathrm{d}}{\mathrm{ds}} \\
\mathcal{H}'_{31} &= (2(\ell_1 - \ell_2)\kappa_2 + 2\ell_2\kappa_2^o)\frac{\mathrm{d}}{\mathrm{ds}} \\
\mathcal{H}'_{32} &= (2(\ell_1 - \ell_2)\kappa_1 - 2\ell_1\kappa_1^o)\frac{\mathrm{d}}{\mathrm{ds}}.
\end{aligned} \tag{D.4}$$

where we make the contribution symmetric by

$$\mathcal{H}'_{sym} \triangleq \frac{1}{2}\left(\mathcal{H}' + \mathcal{H}'^T\right) \tag{D.5}$$

D.1.3 Terms with two derivatives

The final terms are very straightforward as we write,

$$\mathcal{H}'' = \begin{bmatrix} \theta_1' & \theta_2' & \psi' \end{bmatrix} \begin{bmatrix} \ell_1 & 0 & 0 \\ 0 & \ell_2 & 0 \\ 0 & 0 & \ell_\tau \end{bmatrix} \begin{bmatrix} \theta_1' \\ \theta_2' \\ \psi' \end{bmatrix} \tag{D.6}$$

D.1.4 Summary

Then putting eq. (D.3), eq. (D.4), eq. (D.6) we have

$$\mathcal{H} = \mathcal{H}^0 + \mathcal{H}'_{sym} + \mathcal{H}'' \tag{D.7}$$

which agrees with the results of the linear operators, although only involves first order derivatives, which makes our numerics cleaner.

Appendix E

Delta Function Constraints

Conjugate Variable Method

The δ-functions can be written in terms of the amplitudes of the i^{th} eigenmode by a Fourier Transformation

$$\begin{aligned}\delta(u_1) =&\delta\left(\sum_{i=1}^{N} u_{1i}(L)\right) = \delta\left(\sum_{i=1}^{N} \xi_i L v_{1i}\right) = \frac{1}{2\pi}\int \mathrm{d}\mu_1 \exp\left(-i\mu_1 \sum_{i=1}^{N} L v_{1i}\xi_i\right)\\ =&\frac{1}{2\pi}\int \mathrm{d}\mu_1 \prod_{i=1}^{N} \exp\left(-i\mu_1 L v_{1i}\xi_i\right). \end{aligned} \tag{E.1}$$

The product of the δ-functions is

$$\delta(u_1(L))\delta(u_2(L))\delta(u_3(L)) = \int \frac{\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\mu_3}{(2\pi)^3}\prod_{i=1}^{N}\exp(-iL\xi_i\sum_{j=1}^{3}\mu_j v_{ji}). \tag{E.2}$$

We write $\tilde{Z}_\ell$ as the looped state partition function, modulo the enthalpic contributions

$$\tilde{Z}_\ell = \frac{Z_\ell}{\exp\left(-\beta E_\ell - c_f\right)}. \tag{E.3}$$

We then use the Fourier Transformation of the δ-functions to explicitly compute the

looped state partition function Z_ℓ.

$$\tilde{Z}_\ell = \int \prod_{i=1}^{\infty} \mathrm{d}\xi_i \exp\left(-\frac{1}{2}\lambda_i \left(\frac{\ell_p}{L}\right)\xi_i^2\right) \delta(u_1(L))\delta(u_2(L))\delta(u_3(L))$$
$$= \int \frac{\mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3}{(2\pi)^3} \prod_{i=1}^{\infty} d\xi_i \exp\left(-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\xi_i^2 - iL\xi_i \sum_{j=1}^{3} \mu_j v_{ij}\right). \tag{E.4}$$

Now that we have an expansion for the δ-functions in terms of the amplitudes of the i^{th} mode ξ_i, we can explicitly compute the integration over the normal modes of our Hamiltonian,

$$\tilde{Z}_\ell = \int \prod_{i=1}^{\infty} \mathrm{d}\xi_i \exp\left(-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\xi_i^2\right) \delta(u_1(L))\delta(u_2(L))\delta(u_3(L))$$
$$= \int \frac{\mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3}{(2\pi)^3} \prod_{i=1}^{\infty} d\xi_i \exp\left(-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\xi_i^2 - iL\xi_i \sum_{j=1}^{3} \mu_j v_{ij}\right)$$
$$= \int \frac{\mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3}{(2\pi)^3} \prod_{i=1}^{\infty} \mathrm{d}\xi_i \exp\left(-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\left(\xi_i^2 + \frac{2iL^2}{\lambda_i L}\xi_i \sum_{j=1}^{3} \mu_j v_{ij}\right)\right). \tag{E.5}$$

We then complete the square in terms of ξ_i

$$\tilde{Z}_\ell = \int \frac{\mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3}{(2\pi)^3} \prod_{i=1}^{\infty} \mathrm{d}\xi_i \exp\left(-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\left(\xi_i + \frac{iL^2}{\lambda_i \ell_p}\sum_{j=1}^{3}\mu_j v_{ij}\right)^2 - \frac{1}{2}\frac{L^3}{\lambda_i \ell_p}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2\right)$$

and shift the integration over $\xi_i + \dfrac{iL^2}{\lambda_i \ell_p}\displaystyle\sum_{j=1}^{3}\mu_j v_{ij} \to \xi_i$,

$$\tilde{Z}_\ell = \int \frac{\mathrm{d}\mu_1 \mathrm{d}\mu_2 \mathrm{d}\mu_3}{(2\pi)^3} \prod_{i=1}^{\infty} d\xi_i \exp\left(-\frac{1}{2}\lambda_i \frac{\ell_p}{L}\xi_i^2\right) \exp\left(-\frac{1}{2}\frac{L^3}{\ell_p \lambda_i}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2\right) \tag{E.6}$$

We only keep up to the N^{th} eigenmode ξ_N as modes beyond this mode cancel out with corresponding open state modes. Modes past the N^{th} are still constrained by the delta functions, so we formally keep them for now

$$\tilde{Z}_\ell = \left(\frac{2\pi L}{\ell_p}\right)^{N/2} \prod_{i=1}^{\infty} \frac{1}{\sqrt{\lambda_i}} \int \frac{\mathrm{d}\mu_1 d\mu_2 d\mu_3}{(2\pi)^3} \exp\left(-\frac{1}{2}\frac{L^3}{\ell_p \lambda_i}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2\right) \tag{E.7}$$

We are left with the integral over conjugate variables μ_i. We collect conjugate vari-

ables in terms of a vector $\mu \triangleq [\mu_1, \mu_2, \mu_3]$. Now expanding,

$$\begin{aligned}\frac{1}{\lambda_i}\left(\sum_{j=1}^{3}\mu_j v_{ij}\right)^2 =& \frac{1}{\lambda_i}\left(\mu_1^2 v_{1i}^2 + \mu_2^2 v_{2i}^2 + \mu_3^2 v_{3i}^2\right. \\ &\left. + 2\mu_1\mu_2 v_{1i}v_{2i} + 2\mu_1\mu_3 v_{1i}v_{3i} + 2\mu_2\mu_3 v_{2i}v_{3i}\right) \\ =& \frac{1}{\lambda_i}\begin{bmatrix}\mu_1\mu_2\mu_3\end{bmatrix}\begin{bmatrix} v_{1i}^2 & v_{1i}v_{2i} & v_{1i}v_{3i} \\ v_{1i}v_{2i} & v_{2i}^2 & v_{2i}v_{3i} \\ v_{1i}v_{3i} & v_{2i}v_{3i} & v_{3i}^2\end{bmatrix}\begin{bmatrix}\mu_1 \\ \mu_2 \\ \mu_3\end{bmatrix}\end{aligned} \tag{E.8}$$

where we get a constraint matrix V_i for every eigenmode ξ_i. We see that the contribution of each mode is suppressed by the eigenvlaue, therefor modes beyond N, will not contribute in a meaningful way, we then drop them and keep up to mode N for consistency, although numerically the first few modes have a significant V_i. We then define the constraint matrix as

$$V \triangleq \sum_{i=1}^{N}\frac{1}{\lambda_i}\begin{bmatrix} v_{1i}^2 & v_{1i}v_{2i} & v_{1i}v_{3i} \\ v_{1i}v_{2i} & v_{2i}^2 & v_{2i}v_{3i} \\ v_{1i}v_{3i} & v_{2i}v_{3i} & v_{3i}^2\end{bmatrix} \tag{E.9}$$

Then plugging V back into eq. (E.7)

$$\begin{aligned}\tilde{Z}_\ell &= \left(\frac{2\pi L}{\ell_p}\right)^{N/2}\left(\prod_{i=1}^{N}\frac{1}{\sqrt{\lambda_i}}\right)\int\frac{\mathrm{d}\mu_1\mathrm{d}\mu_2\mathrm{d}\mu_3}{(2\pi)^3}\exp\left(-\frac{1}{2}\frac{L^3}{\ell_p}\vec{\mu}^T V\vec{\mu}\right) \\ &= \left(\frac{2\pi L}{\ell_p}\right)^{N/2}\frac{1}{\sqrt{\det H^\ell}}\frac{1}{(2\pi)^3}\left(\frac{\ell_p}{L^3}\right)^{3/2}\sqrt{\frac{(2\pi)^3}{\det V}}\end{aligned} \tag{E.10}$$

Where we have used the identity between the determinant and product of eigenvalues

$$\det H = \prod_{i=1}^{N}\frac{1}{\sqrt{\lambda_i}}. \tag{E.11}$$

Bibliography

[1] S. Adhya, S. Semsey, and K. Virnik. A gamut of loops: meandering DNA. *Trends in biochemical sciences*, 30(6):334–41, 2005.

[2] A. Balaeff, C. R. Koudella, L. Mahadevan, and K. Schulten. Modelling DNA loops using continuum and statistical mechanics. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 362(1820):1355–71, July 2004.

[3] A. Balaeff, L. Mahadevan, and K. Schulten. Elastic rod model of a DNA loop in the Lac operon. *Phys. Rev. Lett*, 83:4900–4903, 1999.

[4] A. Balaeff, L. Mahadevan, and K. Schulten. Modeling DNA loops using the theory of elasticity. *Physical Review E*, 73(3):1–23, 2006.

[5] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante. Ionic effects on the elasticity of single DNA molecules. *PNAS*, 94:6185–6190, 1997.

[6] C. R. Calladine, H. R. Drew, B. Luisi, and A. A. Travers. *Understanding DNA, Third Edition: The Molecule and How it Works*. Academic Press, 2004.

[7] Y. F. Chen, J. N. Milstein, and J. C. Meiners. Femtonewton Entropic Forces Can Control the Formation of Protein-Mediated DNA Loops. *Physical Review Letters*, 104(4):1–4, January 2010.

[8] T. E. Cloutier and J. 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):3645–50, March 2005.

[9] B. D. Coleman, W. K. Olson, and D. Swigon. Theory of sequence-dependent DNA elasticity. *The Journal of Chemical Physics*, 118(15):7127, 2003.

[10] R. Courant and D. Hilbert. *Methods of Mathematical Physics, Vol. 1*. Wiley-VCH, 1989.

[11] T. M. Devlin. *Textbook of Biochemistry with Clinical Correlation*. Wiley-Liss, 2010.

[12] F Dong, S Spott, O Zimmermann, B Kisters-Woike, B Müller-Hill, and a Barker. Dimerisation mutants of Lac repressor. I. A monomeric mutant, L251A, that binds Lac operator DNA as a dimer. *Journal of molecular biology*, 290(3):653–66, July 1999.

[13] Q. Du, M. Vologodskaia, H. Kuhn, M. Frank-Kamenetskii, and A. Vologodskii. Gapped DNA and cyclization of short DNA fragments. *Biophys. J*, 88(6):4137–45, June 2005.

[14] Q. Du, M. Vologodskaia, A. Vologodskii, C. Smith, and N. Shiffeldrim. Cyclization of short DNA fragments and bending fluctuations of the double helix. *PNAS*, 102:5397–402, 2005.

[15] S. Goyal, L. Lee, and N. C. Perkins. Nonlinear dynamics and loop formation in Kirchhoff rods with implications to the mechanics of DNA and cables. *Journal of Computational Physics*, 209:371–389, 2005.

[16] S. Goyal, T. D. Lillian, S. Blumberg, J. C. Meiners, and E. Meyhofer. Intrinsic Curvature of DNA In uences LacR-Mediated Looping . *Biophys. J*, 93(December):4342–4359, 2007.

[17] P. J. Hagerman. Flexibility Of DNA. *Annual review of biophysics and biophysical chemistry*, 17(1):265–286, January 1988.

[18] L. Han, H. G. Garcia, S. Blumberg, K. B. Towles, J. F. Beausang, P. C. Nelson, and R. Phillips. Concentration and Length Dependence of DNA Looping in Transcriptional Regulation. *Cancer Research*, 4(5):e5621, 2009.

[19] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[20] J. N. Israelachvili. *Intermolecular and Surface Forces with Applications to Colloidal and Biological Systems*. Academic Press, 1985.

[21] H. Jacobson and W. H. Stockmayer. Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. *Journal of Chemical Physics*, 18(12):1600, 1950.

[22] S. Kehrbaum and J. H. Maddocks. Effective properties of elastic rods with high intrinsic twist. *IMACS World Congress*, 16th, 2000.

[23] E. Kreyszig. *Differential Geometry*. Dover Publications, 1991.

[24] L. D. Landau, L. P. Pitaevskii, E. M. Lifshitz, and A. M. Kosevich. *Theory of Elasticity, Third Edition: Volume 7*. Butterworth-Heinemann, 1986.

[25] F. Lankas, J. Sponer, P. Hobza, and J. Langowski. Sequence-dependent elastic properties of DNA. *Journal of molecular biology*, 299(3):695–709, 2000.

[26] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, 271(5253):1247–54, March 1996.

[27] T. D. Lillian. Computational elastic rod model applied to dna looping. *Engineering Conference*, 2007.

[28] T. D. Lillian. *Mechanics and Function of DNA Looping and Supercoiling*. PhD thesis, University of Michigan, 2010.

[29] T. D. Lillian, S. Goyal, J. D. Kahn, E. Meyhofer, and N. C. Perkins. Computational Analysis of Looping of a Large Family of Highly Bent DNA by LacI. *Biophys. J*, 95(12):pp5832–5842, 2008.

[30] T. D. Lillian, S. Goyal, E. Meyhofer, J. D. Kahn, and N. C. Perkins. Computational elastic rod theory captures DNA and protein flexibility in the Lac-repressor complex. *Biophys. J*, 94((1 Meeting Abstracts)):294, 2008.

[31] P. T. Lowary and J. 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, February 1998.

[32] C. L. Lu and N. C. Perkins. Nonlinear Spatial Equilibria and Stability of Cables Under Uni-axial Torque and Thrust. *Journal of Applied Mechanics*, 61(4):879, 1994.

[33] G. S. Manning, J. H. Maddocks, and J. D. Kahn. A continuum rod model of sequence-dependent DNA structure. *AIP*, 105(13):5626–5646, 1996.

[34] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, December 1995.

[35] N. L. Marky, R. L. Jernigan, W. K. Olson, and V. 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.

[36] A. Matsumoto and W. K. Olson. Sequence-Dependent Motions of DNA: A Normal Mode Analysis at the Base-Pair Level. *Biophys. J*, 83(July):22–41, 2002.

[37] R. A. Mehta and J. D. Kahn. Designed hyperstable Lac repressor.DNA loop topologies suggest alternative loop geometries. *Journal of molecular biology*, 294(1):67–77, November 1999.

[38] M. A. Morgan, K. Okamoto, J. D. Kahn, and D. S. English. Single-molecule spectroscopic determination of lac repressor-DNA loop conformation. *Biophys. J*, 89(4):2588–96, October 2005.

[39] J. Müller, S. Oehler, and B. Müller-Hill. Repression of lac Promoter as a function of distance, Phase and Quality of an Auxiliary lac Operator. *J. Mol. Biol.*, 257:21–29, 1996.

[40] M. G. Munteanu, K. L. Vlahovicek, S. Parthasaraty, I. Simon, and S. Pongor. Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.*, 23(9):341–346, 1998.

[41] P. C. Nelson, D. E. Segall, and R. Phillips. Volume-exclusion effects in tethered-particle experiments: bead size matters. *Phys. Rev. Lett*, 96:088306, 2006.

[42] S. Oehler, E. R. Eismann, H. Kramer, and B. Müller-Hill. The three operators of the lac operon cooperate in repression. *EMBO Journal*, 9(4):973 – 979, 1990.

[43] W. K. Olson. Simulating DNA at low resolution. *Current Opinion in Structural Biology*, 6(6):242–256, 1996.

[44] W. K. Olson, A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(19):11163–8, September 1998.

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

[46] W. K. Olson and V. B. Zhurkin. Modeling DNA Deformations. *Curr. Opin. Struct. Biol*, 10:286–297, 2000.

[47] S. Panyukov and Y. Rabin. Fluctuating elastic rings: statics and dynamics. *Phys. Rev. E*, 64:011909, July 2001.

[48] P. K. Purohit and P. C. Nelson. Effect of supercoiling on formation of protein-mediated DNA loops. *Phys. Rev. E*, 74(06):1907, 2006.

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

[50] D. Shore, J. Langowski, and R. L. Baldwin. DNA flexibility Studied by Covalent Closure of Short Fragments into circles. *Biochemistry*, 78(8):4833–4837, 1981.

[51] S. B. Smith, L. Finzi, and C. Bustamante. Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads. *Science*, 258(5085):1122–1126, 1992.

[52] T. R. Strick, J. F. Allemand, D. Bensimon, and V. Croquette. The Elasticity of a Single Supercoiled DNA Molecule. *Science*, 271(5257):1835–1837, 1996.

[53] D. Swigon, B. D. Coleman, and W. K. Olson. Modeling the Lac repressor-operator assembly: The influence of DNA looping on Lac repressor conformation. *PNAS*, 103(26):9879–9884, 2006.

[54] A. V. Tkachenko. Role of boundary constraints in DNA cyclization. *q-bio/0703026*, 2007.

[55] K. B. Towles, J. F. Beausang, H. G. Garcia, R. Phillips, and P. C. Nelson. First-principles calculation of DNA looping in tethered particle experiments. *Physical biology*, 6(2):25001, January 2009.

[56] B. van Den Broek, F. Vanzi, D. Normanno, G. J. L. Wuite, and F. S. Pavone. Real-time observation of DNA looping dynamics of type IIE restriction enzymes NaeI and NarI. *Nucleic acids research*, 34(1):167–174, 2006.

[57] E. Villa, A. Balaeff, and K. Schulten. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6783–8, May 2005.

[58] E. Villa, L. Mahadevan, K. Schulten, and A. Balaeff. Multiscale method for simulating protein-DNA complexes. *Multiscale Model. Simul.*, 107:3967, 2004.

[59] K. L. Vlahovicek, L. Kajan, and S. Pongor. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucl. Acids Res.*, 31:3686–3687, 2003.

[60] J. Widom. Role of DNA sequence in nucleosome stability and dynamics. *Quarterly reviews of biophysics*, 34(3):269–324, August 2001.

[61] J. Widom and T. E. Cloutier. Spontaneous sharp bending of double-stranded DNA. *Molecular cell*, 14(3):355–62, May 2004.

[62] H. Yamakawa and W. H. Stockmayer. Statistical Mechanics of Wormlike Chains. II. Excluded Volume Effects. *Journal of Chemical Physics*, 57(7):2843, 1972.

[63] Y. Zhang and D. M. Crothers. High-throughput approach for detection of DNA bending and flexibility based on cyclization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6):3161–6, March 2003.

[64] Y. Zhang and D. M. Crothers. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J*, 84:136–153, 2003.

[65] Y. Zhang, A. E. McEwen, D. M. Crothers, and S. D. Levene. Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. *PloS one*, 1(1):e136, January 2006.