Mechanics and Function of DNA Looping and Supercoiling

by

Todd D. Lillian

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

Doctoral Committee:

Professor Edgar Meyhöfer, Co-Chair
Professor Noel C. Perkins, Co-Chair
Professor Ronald G. Larson
Associate Professor Jens-Christian Meiners
Associate Professor Jason D. Kahn, University of Maryland

Acknowledgments

I am thankful for the opportunity I have had to work with my advisors Prof. Noel Perkins and Prof. Edgar Meyhöfer. Throughout my graduate research, Prof. Perkins (Mechanical Engineering, University of Michigan) has provided sound direction to my research and has offered me the flexibility to pursue many of my own research interests. Prof. Meyhöfer (Mechanical Engineering, University of Michigan) has helped me better understand molecular biology and the value of stepping away from the details of my calculations to see how they fit into the big picture.

I would also like to thank the remaining members of my committee: Prof. R. Larson, Prof. J-C. Meiners, and Prof. J. Kahn. I enjoyed discussions with Prof. R. Larson (Chemical Engineering, University of Michigan), especially those relating to dimensional analysis and DNA dynamics. Prof. J-C. Meiners (Physics and Biophysics, University of Michigan) and his students have helped me to better understand the biology of the Lac repressor system. My research has been strengthened by Prof. J. Kahn (Chemistry and Biochemistry, University of Maryland) and his students who have shared their experimental observations and expertise; and I especially enjoyed visiting their lab and seeing some experiments.

I am especially grateful for Dr. Sachin Goyal (currently a postdoc at Emory University) for patiently mentoring me through the steep learning curve as I began my research, and as he concluded his PhD. He also provided me with the computational model for DNA which is the foundation for my research.

I have enjoyed working with Troy Lionberger throughout my PhD. I am especially grateful for the opportunity to be included in his collaboration with Andrzej Stasiak and Davide Demurtas which resulted in Chapter 5

I would like to thank Prof. I. Andricioaei at the University of California - Irvine, and his students, Maryna Taranova and Jeff Wereszczynski, for our collaborations. Their MD simulations facilitated the work described in Chapter 7.

I am also thankful for many other faculty and students with whom I have had the privilege to collaborate and discuss common research interests, including: Seth Blumberg, Aaron Haeusler, Andy Hirsh, Troy Lionberger, Alexei Tkachenko, Tom Weglarz, and David

Wilson.

Finally, I am grateful for the financial support I have benefited from throughout the duration of my PhD studies. The University of Michigan Department of Mechanical Engineering supported me with a one year Department Fellowship and for one semester with a Graduate Student Instructorship. Through my advisors, I have also received continued support from NSF grants CMS-0439574, CMS-0510266 and CMMI-0825488.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Appendices

Appendix

Chapter 1

Introduction

DNA is an essential molecule that enables the storage and retrieval of genetic information. Following the discoveries of Watson and Crick [114] and Franklin and Gosling [30] over 50 years ago, the double helical structure of DNA has been widely publicized. Specifically, DNA is composed of four nucleotide subunits (Adenine (A), Thymine (T), Cytosine (C), and Guanine (G)) that bind together forming long polymer chains. Two chains of nucleotides wrap around each other forming the famous double helix. The double helix is held together by hydrogen bonds between the nucleotides on one strand with complementary nucleotides of the adjoining strand. The resulting basepairs obey the following pairing rules: A binds with T, and C binds with G. The specific sequence of nucleotides carries genetic information.

In their classic paper, Watson and Crick state, “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material [114].” Thus, the structure of the molecule is significant to its function. Subsequent studies have confirmed this observation and have also highlighted the important relationships between the mechanical properties of DNA and its biological functions. As an example, consider the length scales describing DNA as suggested by [13]. Its diameter is approximately 2 *nm*, and for human DNA molecules its length is about 3 *cm*. Several such molecules are packaged into a cell nucleus which is approximately 5 μm in diameter. To be contained within the cell nucleus, the extremely long molecule must be compacted by four orders of magnitude. Interestingly, the cell controls this compaction while also ensuring accessibility of genetic information. Clearly the mechanical properties of DNA including its bending and twisting enable this considerable compaction.

1.1 DNA Looping[1]

A premier example that illustrates the biological significance of DNA mechanics is that of protein-mediated DNA looping (see, for example [68, 88]). Looping proteins bind simultaneously to separated sites on a DNA molecule to form a loop from the intervening DNA. The loop can both respond to changes in bending and torsion or induce them to affect downstream functions, such as regulating transcription or facilitating protein-DNA assembly during replication, recombination, and condensation [28, 90]. For example, the activator protein NtrC facilitates transcription by interacting with the RNA polymerase holoenzyme via looping, and the loop can respond to DNA bending. Repressor proteins such as the Lac and Gal repressors (LacI and GalR) of the bacterium *E. coli* repress transcription upon introducing a DNA loop, and GalR may act by establishing a promoter conformation inconsistent with initiation. For these and other systems, the intrinsic curvature of DNA may facilitate looping [16, 70] by essentially ‘pre-bending’ the helical axis of otherwise unstressed DNA. Similarly, proteins such as CAP and IHF induce large DNA bends (about 90° for CAP, 180° for IHF) to facilitate looping; see, for example [3, 60, 99]. Although our focus is on looping of DNA by LacI of the bacterium *E. coli*, relevant similarities exist with other DNA-protein complexes forming similarly short DNA loops.

1.1.1 Experiments on the Lac Repressor DNA Loop Complex

X-ray crystallography revealed the structure of the Lac repressor core tetramer and complete protein including short DNA operator fragments [31, 60] as illustrated in Fig. 1.1. Analysis of these structures suggest the possible role of protein flexibility concentrated in the apex of the protein ‘V’ (yellow circle of Fig. 1.1) and/or in the head domains (green circles of Fig. 1.1) [31, 60]. Lower resolution microscopy images of the entire protein DNA complex [58, 83, 108] further support the possibility of protein conformational changes. In addition, crystallography identifies symmetries in the arrangement of four equivalent subunits, arranged as a dimer of dimers [31, 60]. Furthermore, palindromic DNA operator fragments formed the best crystals and were therefore used in the crystal structure of [60]. The symmetry of palindromic operators, however, leads to ambiguity in the precise manner in which DNA binds to the protein (binding topology) as mentioned in [37, 98].

One of the major challenges limiting our understanding of the mechanics and dynamics of these DNA loops is their short length scale. DNA, with a 2 *nm* diameter is much smaller than the wavelength of light in the visual spectrum and therefore cannot be visualized in a

---

[1] In this section we quote significant passages from our previous publication [61].

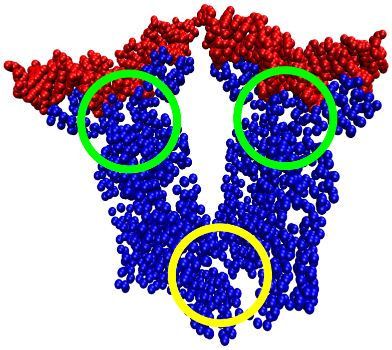


Figure 1.1 Crystal structure of the Lac repressor protein in complex with short DNA operator fragments as determined by [60]. The protein C-alpha atoms are illustrated with blue spheres while the complete set of atoms for the operator fragments are given by red spheres. The two green circles, near the head domains, highlight locations of suspected flexibility. The yellow circle highlights a region of suspected flexibility in the apex of the protein ‘V’. This illustration was created using the software of [46].

light microscope. Conventional electron microscopy (EM) and atomic force microscopy (AFM), however, have been successful at resolving DNA to the scale of its diameter. However, in these methods the DNA is removed from biologically relevant conditions, flattened onto a planar substrate and often stained before it can be visualized. The mechanical state of DNA in these images could well reflect the process of attaching it to a surface and provide only a questionable representation of its conformation in physiological solution. Additionally, these static images provide little insight into the dynamics of the molecule.

Cryo electron microscopy (cryo-EM) is a promising technique for imaging single DNA molecules. It overcomes many of the challenges faced by conventional EM and AFM by eliminating the need to attach the molecule to a substrate. Instead, in cryo-EM, a DNA construct remains in a thin film (~ 40 *nm* thick) of its physiological solution while it undergoes flash freezing. The freezing occurs on a short time scale [27] and thereby limits the molecule’s opportunity to respond to the abrupt change in temperature. Therefore the state of the molecule is captured nearly at an instant in time. When two Cryo-EM images of a single DNA molecule are considered simultaneously (a stereo pair), its three-dimensional helical axis can be resolved to within about 1 *nm*; but its individual nucleotides cannot be resolved [47]. Therefore, it’s not possible to relate or ‘register’ a molecule’s known nucleotide sequence with its cryo-EM image without any identifying feature like an end or a bound protein. Additionally, the torsional state of strain around the molecule is unknown.

This reconstruction procedure is complicated, relies upon user inputs, assumes the axis is smooth, and is limited by the resolution of the individual cryo-EM images [47]. The differences among reconstructions of molecules from a single population reflect the thermal fluctuations of the molecule about its minimum energy configuration.

Although cryo-EM has not been employed to visualize Lac repressor loops, it has been used to study the geometry of DNA minicircles of a similar length. In [1] two populations of 158 *bp* DNA minicircles were studied. In this study, the two populations were identical in sequence except for an 18 *bp* site containing either a CAP binding site or TATA box. The objective of this study was to detect differences in the intrinsic curvature of the two constructs. Even with a sample size of 31 and 64 for CAP and TATA minicircles respectively, the thermal fluctuations of the minicircles were significant enough to “blur” any differences. A further study considered shorter 94 *bp* minicircles and focused on the detection of kinks in the minicircles [24]. As a control, nicks (broken phosphodiester bond) and gaps (missing basepairs) were introduced into the backbone of some of the minicircles. This study showed no conclusive evidence for kinks or sharp bends in the covalently closed or nicked constructs. In contrast, the gapped minicircles showed sharp bends indicative of high localized flexibility.

One of the best experimentally characterized behaviors of Lac repressor looping is the dependence of loop stability (as measured by, for example, repression level) on the length of the inter-operator DNA. Similar to results from cyclization experiments (see, for example [92]), results from looping experiments exhibit an oscillatory component in the free energy of looping with a period of about a helical turn (see, for example, [6, 8, 58, 74]). This oscillatory component is a direct result of the required torsional alignment of the operator sites when bound to the Lac repressor. That is, upon the addition of basepairs to the inter-operator DNA, the relative angular orientation between operator sites changes, due to the helical construction of DNA, and therefore the conformation of the loop adjusts to maintain proper alignment of the DNA operators with respect to the protein binding sites. (The relative angular orientation between operator sites is referred to as ‘operator phasing’.) However, as cautioned in [37, 119], the distinct closure requirements for looping versus cyclization yield significant quantitative differences in the thermodynamics, particularly for the sub-persistence length of the inter-operator DNA of the DNA-LacI complex. Many of these looping studies are complicated by a myriad of factors that influence looping in-vivo. For example, the work of Becker et al. [6] highlights the strong influence of the architectural protein HU in stabilizing DNA loops. In addition, in-vivo loop stability is often measured by repression level, which may also be influenced by other cellular factors.

By contrast, in-vitro experiments, as in [57, 58], are more controlled and allow for multiple methods to interrogate DNA loops. For instance, Krämer et al. examine the looping behavior of inter-operator DNA sequences spanning the range of 153 – 168 *bp* with experiments including gel shift assays, DNA footprinting, and permanganate sensitivity. They further probe looping of supercoiled DNA minicircles. The collection of experiments provides insight into phasing effects on both the energetics and topology of DNA loops. More specifically these experiments probed the following loop characteristics: stability, topology (i.e., linking number), size, sensitivity to DNA digestion enzymes, and effects of supercoiling substrate DNA. Despite this extensive experimental effort, much remains unknown regarding the energetics or topology of these DNA loops.

Some of the most revealing experiments on DNA looping in-vitro are those of the Kahn lab at the University of Maryland [29, 70, 73] which probe the combined effects of phasing and intrinsic curvature on protein-mediated looping. In these studies, significant intrinsic curvature is introduced into three designed sequences (denoted 11C12, 9C14 and 7C16) by incorporating A-tract bends into the inter-operator DNA [70]. Abundant experimental data have been published for these sequences including the following loop characteristics: stability, size, topology, and binding topology. Although, individually these experiments each provide important insight into DNA looping, the collection of data makes these experiments particularly insightful. That said, the FRET experiments [29, 73] are especially interesting in that they directly address the role of protein flexibility and DNA binding topology. These FRET experiments suggest that the 9C14 sequence forms predominantly with the 'parallel' binding topology in which the protein 'V' is not extended [73]. Despite this compelling data, many questions remain. For example, the role of protein flexibility and binding topology for the 11C12 sequence remains uncertain.

1.1.2 Models for the Lac Repressor-DNA Loop Complex

The mechanics of DNA looping and cyclization, although notably different [119], also share similarities which lead to several common modeling strategies. The most detailed model for DNA mechanics is the atomistic representation afforded by molecular dynamics (MD). MD calculations track the position of the individual atoms composing DNA and the surrounding physiological buffer as they interact and as functions of time. At the atomistic scale, direct simulation of DNA looping through MD remains an elusive goal due to the sheer size and disparate time and length scales needed to resolve looping dynamics [89]. Even so, short time simulations (about 80 *ns*) on DNA minicircles of about 100 *bp* are possible with significant computational effort [59].

The large computational resources required for fine grain, all-atom MD, motivates the need for coarse grain models. One method of coarse graining DNA is to consider each basepair as a rigid body. In the resulting 'discrete' models [20, 50, 78, 98, 111, 119, 121], the basepairs are approximated as rigid bodies connected to their nearest neighbors by translational and rotational springs. The discrete models, which have the potential to resolve details at the basepair scale, necessarily require the basepair scale resolution of model parameters. The elastic rod approximation (see, for example, [4, 38, 65, 80]), which has been employed to study DNA supercoiling and looping, is another coarse-grained model describing the long-length scale mechanics of the molecule [77, 89]. This continuum model necessarily averages the elasticity properties of DNA over at least a helical turn. Nevertheless the rod model is capable of accounting for sequence-dependent properties including sequence-dependent intrinsic curvature and sequence-dependent stiffness; see, for example, the formulations in [38, 65]. Both discrete and continuum models may approximate effects including intrinsic curvature, electrostatics, self contact, and hydrodynamics. However, a major challenge is to properly characterize all of these physical interactions when the experimental data is nonexistent or incomplete. Coarse grain models can also be extended to account for thermal effects by incorporating Brownian dynamics, statistical mechanics, or Monte Carlo methods [2, 50, 53, 71, 91, 98, 116, 119, 121]. In this context, these coarse grain methods are known by alternate names. For instance, the 'worm-like chain' model is simply an elastic rod model that accounts for Brownian dynamics. Often the 'worm-like chain' model is discretized to facilitate computations ('bead-rod' model) and this formulation resembles the discrete methods we have discussed.

Models of the DNA-Lac repressor complex start with the known crystal structure of the Lac repressor including the bound DNA operators [60]. For instance, the rod model does so by employing rod position and orientation boundary conditions deduced from the Lac repressor crystal structure. However, a challenge remains in defining boundary conditions with imprecise knowledge of the protein structure for the looped complex. Recent studies have explored the role of protein flexibility. For example, [50, 98, 119, 120] employ simplified protein models in which the protein is allowed to deform from the 'V' configuration of the crystal structure into an extended conformation suggested by experiments. Alternatively, the multi-scale model of [106, 107] explicitly accounts for protein flexibility over short time scales. In this multi-scale model the DNA loop is treated as an equilibrium elastic rod coupled to an MD model for the protein. The findings highlight the role of flexibility concentrated in the head domains and suggest that the protein is stiffer than expected against protein 'V' deformations. However, these results were obtained by exaggerating the reaction forces of the DNA on the protein in order to accelerate the relaxation in the (necessarily) limited integration time.

Comparing computational models with experiments on looping helps establish model parameters, check model accuracy, and shed light on the physical processes in looping. A common comparison with experiments is the dependence of repression level (a measure of loop stability) on inter-operator length. Computational models clearly show that the oscillations in loop stability with length are a result of operator phasing (see for example [37]). Perhaps the most intensive comparison to this phasing data are those of Zhang et al. that fit four model parameters to in-vivo experimental data of repression level [120]. Interestingly, the best parameter fit corresponds to unusually high DNA and protein flexibility, suggesting a mechanism for increased DNA flexibility in-vivo.

The more extensive in-vitro experiments on the looping of straight DNA [57, 58] and of intrinsically curved DNA [29, 70, 73] facilitate in depth comparisons with models. In contrast to in-vivo repression level experiments that primarily measure looping energetics, the in-vitro experiments also explore loop topology. Theoretical studies have compared predictions of loop topology with experimental measurements and analyzed sensitivity to DNA digestion enzymes [98], linking number [50], and binding topology [50, 98].

1.2 DNA Supercoiling

Another behavior that illustrates the strong connection between DNA mechanics and biological function is that of DNA supercoiling. Many significant cellular processes, including transcription and replication, induce DNA supercoils. Additionally, the action of many proteins that interact with DNA, may be enhanced or repressed by the degree of supercoiling. (In fact, the affinity of the Lac repressor to DNA operator sites is known to depend upon the degree of supercoiling [57].) Consequently, supercoiling must be regulated by enzymes (topoisomerases) in the cell; see, for example, [26]. Interestingly, topoisomerases in cancerous cells have been a target of chemotherapy drugs. For example, topotecan inhibits human type I topoisomerase and thereby results in cell death. Topoisomerases are categorized as either type I or II based on the mechanism used to alter supercoiling: type I topoisomerases transiently cut a single DNA backbone while type II topoisomerases transiently cut both backbones. Both type I and II enzymes are further classified by their mechanism of operation. Here we focus on type IB topoisomerases (topo IB). The crystal structure of topo IB has been determined [81] and its operation can be summarized in the following steps: 1) bind and clamp around a double stranded DNA with weak sequence specificity [7], 2) sever a single backbone and bind the free 3’ end, 3) allow a “controlled rotation” [14, 54, 95, 115] about the intact backbone to relieve positive or negative supercoils, and 4) religate the broken backbone. Examples of type IB topoisomerases include vaccinia topoisomerase and human topoisomerase I. (The roman numeral “I” in human topoisomerase I denotes its order of discovery and not its type [34].) Although we shall focus on the topo IB, similarities exist with other systems involving supercoiled DNA.

1.2.1 Single Molecule Experiments on Supercoiled DNA and Topo IB

Recent advances in single molecule experimental techniques provide substantial insight into the mechanics and dynamics of supercoiled DNA. Especially relevant to topoisomerases are the techniques that probe the torsional mechanics of DNA, including magnetic tweezers, which manipulate magnetic beads bound to DNA. These experiments often consist of a single DNA molecule tethering a magnetic bead (about $1\mu m$ in diameter) to a glass cover slip. A permanent magnet is then used to exert a force on the magnetic bead. While measuring the quasi-static force extension curve of DNA under tension, Smith et al. [93] identified the potential use of magnetic beads to apply a torque in addition to a force. A torque is generated due to the misalignment of the bead’s dipole with an external magnetic field. However, in their initial experiment they were unable to apply torques because the DNA was bound at a single point at the cover slip and bead, allowing the DNA to swivel freely. Subsequent experiments rectified this problem. For example, Strick et al. [97] were able to twist the DNA by simply rotating a permanent magnet over the magnetic bead tethered by DNA. This allowed them to generate a force-extension relation for supercoiled molecules.

These initial experiments have been extended in several ways. For example, Harada et al. [42] decorated a magnetic bead with smaller fluorescent beads. This enabled them to measure the rotation of the bead and thereby measure the twist of DNA induced by RNA polymerase during transcription. Another example is that of Bryant et al. [10] who introduced an additional and smaller fluorescent bead, referred to as a “rotor bead,” near a single strand nick within a long length of DNA. Here oscillations of this bead were tracked as highly supercoiled DNA relaxed. Crut et al. performed yet another variation of these magnetic bead experiments by combining a laser trap with a magnetic trap [22]. They were able to measure the relaxation of supercoiled DNA under tension. Most relevant to our interest in topo IB are the experiments of Koster et al. [54–56]. In these experiments topo IB is introduced to a highly supercoiled DNA constrained by a magnetic bead. As topo IB relieves the supercoils, a tensile force applied to the bead extends the DNA. Using a calibration curve, this extension is related to the reduction in supercoiling. These experiments measured the effects of tension and and drug topotecan on supercoil relaxation. Interestingly, these experiments show that topo IB does not allow the DNA to freely rotate, but its grip on the

DNA provides “friction” that slows the relaxation.

Despite the significant experimental efforts to characterize torsional dynamics, the experimentally measured dynamics are overwhelmingly dominated by the drag of the large tethered bead (bead diameter is micron-scale compared to the nano-scale diameter of DNA). For instance, the equilibrium model for DNA deformations of Crut et al. [22] confirms the time scale of supercoiled DNA is much faster than that of the bead. Computational modeling is a promising tool to help build our understanding of the dynamics of supercoiled DNA.

1.2.2 Modeling Supercoiled DNA and Topo IB

Developing a model for the relaxation of supercoils by topo IB is a significant challenge. Fortunately, an enormous MD effort ($>$ 100 cpu years) by the Andricioaei lab, at the University of California - Irvine, characterized the energetics and structural changes of topo IB in complex with a short fragment of DNA ($\sim 20\ bp$) [115]. A major result of this study is a free-energy landscape as a function of one reaction coordinate, the relative rotation between the segments of DNA on either side of the single strand nick. Representing a longer biologically relevant length of DNA (much longer than a persistence length) is computationally prohibitive for MD. Therefore a separate model for the long length scale dynamics of supercoiled DNA is required in addition to this MD model.

The DNA modeling strategies introduced above in the context of Lac repressor looping could also apply to the relaxation of supercoils by topo IB. However, effects that are often neglected in the context of Lac repressor loops may become significant when considering the dynamics of plectonemic supercoils. For example, on the short length scale associated with LacI-DNA looping (usually less than about 2 persistence lengths) the elastic energy dominates thermal energy in the free energy cost of looping. For these short DNA loops, the tight bending and electrostatics associated with self contact interactions are energetically prohibitive. In addition, large scale dynamic deformations are usually not modeled in DNA looping. In contrast, modeling the relaxation of a plectonemic supercoil formed from a long length of DNA requires full consideration of these effects. In a plectoneme, the “end loop” experiences significant bending, the “ply” experiences significant electrostatic interactions, and the whole length is subject to thermal excitation and hydrodynamic drag. Furthermore, during relaxation, the entire molecule undergoes a large conformational changes from an initial plectonemic state to a final extended state.

Despite the theoretical challenges listed above, analytical, Monte Carlo, and statistical mechanics methods have all been employed to account for many of the complicated physical effects required to describe the rotation-extension relationship for a DNA molecule bound to a coverslip on one end and a bead on the other; see for example [19, 67, 79, 112, 118]. Specifically, these methods have successfully described the experimentally observed shortening of a DNA tether upon the rotation of an attached magnetic bead. These models have even explained asymmetry of the rotation-extension relationship due to positive vs. negative supercoils, the relative plectonemic vs. extended linear DNA content, and the buckling transition as a plectoneme forms on an extended DNA molecule. Pertinent to our interest, Marko combines his rotation-extension relationship with a simplified description of the topo IB to describe the rate of supercoil relaxation by topo IB [67]. However, like most of the other models, here the long length scale transient dynamics of the supercoiled DNA are not represented.

Relatively few studies have modeled the transient dynamics of supercoiled DNA. Instead most modeling efforts that consider dynamics focus on the steady state dynamics of supercoiled circular DNA induced by thermal fluctuations; see for example [17, 45, 48]. However, the Brownian dynamics studies of Mielke et al. [72] and Wada and Netz [113] are two interesting exceptions. Mielke et al. [72] consider supercoiling induced by RNA polymerase. In their study they follow transient dynamics of plectoneme formation and dissipation of twist along a DNA molecule. Wada and Netz [113] introduce a system in order to study the dissipation of twist into a linear DNA molecule. Specifically, they consider a length of linear DNA with one end under constant torsional rotations and the other free. For low angular velocity they show that the molecule dissipates twist by rotating about the local helical axis along the entire length of the molecule. In contrast, for large angular velocity, the twist dissipation involves the formation of plectonemes near the rotating end.

1.3 Research Objective

The previous studies of LacI looping reviewed above largely rely on either theoretical models or experimental methods with surprisingly little interplay between these approaches. As a result there remain many unanswered questions regarding the energetics and topology of these DNA loops. Clearly, detailed theoretical models can provide tremendous insight toward explaining and motivating experimental studies. Thus, a major objective of this research is to carefully re-examine results from existing experimental studies on DNA looping, with particular focus on studies by the Kahn group at the University of Maryland, to obtain quantitative confirmation of the theoretical models developed herein. In addition, we seek to employ these models to motivate future experimental studies.

In contrast to LacI looping, the relaxation of supercoils by topo IB has received little attention by theoretical analysis. As with LacI looping we are also interested in establishing a strong connection between theoretical models and recent experimental observations. However, before such a connection can be made, we must advance theory. Thus, another major objective of this research is to extend our elastic rod theory to describe the relaxation of supercoils in the topo IB system. In addition, we plan to leverage recent MD simulations by the Andricioaei lab, at the University of California - Irvine, to describe the free energy landscape for DNA rotating within the topo IB enzyme [115].

1.4 Scope of Research

We have selected two example systems to understand the role of DNA mechanics in cellular functions, namely: 1) looping mechanics of DNA by Lac repressor, and 2) relaxation of DNA supercoils by topo IB. These examples are chosen due to the opportunity to collaborate with other computational and experimental research groups and make comparisons with extensive experimental observations. The scope and organization of our study into these two systems is presented next.

1.4.1 Looping Mechanics of DNA by Lac Repressor

Our early research of DNA mechanics focused on Lac repressor mediated DNA looping [37]. In this dissertation research, we extend this work through four new studies. These studies include: 1) analysis of looping for intrinsically straight DNA, 2) analysis of looping for intrinsically curved DNA, 3) effects of Lac repressor flexibility on DNA looping, and 4) analysis of the geometry of DNA minicircles. Each study corresponds to a chapter and is summarized below.

1) Analysis of looping for intrinsically straight DNA. Inspired by the experimental observations of [57, 58], in Chapter 2 we study the influence of operator phasing on loop energetics and topology for nominally straight DNA.

2) Analysis of looping for intrinsically curved DNA. In Chapter 3 we continue prior experimental and computational efforts [29, 37, 70, 73] to understand the role of sequence dependent curvature in DNA looping. Here we significantly extend previous modeling efforts [37] by exploring the possible looping behaviors of a very large family of curved sequences that had not yet been synthesized. Specifically, we draw detailed comparisons with the available experimental data including: loop stability, linking number, binding topology, and gel migration velocity. Computed results identify sequences with distinct looping behaviors including hyper versus hypo stable loops, loops with distinct binding topology, etc. As a result of this work we began a collaboration with the Kahn lab to experimentally characterize these new sequences, to evaluate the predictive capability of the computational model, and to identify directions in which the theoretical work should be further extended.

3) Effects of Lac Repressor flexibility on DNA looping. Building evidence suggests that the structure of Lac repressor protein is not rigid, but is flexible. Therefore, neglecting protein conformational changes in studies 1 and 2 above limit their impact. In Chapter 4, we extend our computational model to explicitly include protein flexibility using an extended rod theory to create a multi-body approximation for the protein.

4) Analysis of the geometry of DNA minicircles. In Chapter 5 we follow up on a recent opportunity to collaborate with researchers imaging DNA minicircles using cryo-electron microscopy. Specifically, we develop metrics to describe the geometry of these highly stressed and supercoiled minicircles. Considering the experimentally observed and theoretically computed thermal distribution of these metrics, we differentiate minicircles with various levels of supercoiling. Doing so sheds light on the nature of DNA bending, including the possibility of kinks, experienced by these minicircles.

1.4.2 Relaxation of DNA Supercoils by Topo IB

After establishing the foundation for modeling the mechanics of static DNA loops with the above studies, we make the transition to our second model system, relaxation of supercoils by topo IB. The dynamics of supercoiled DNA brings several exciting new challenges in accounting for hydrodynamics, thermal excitation, sharp bending deformations (and possibly kinking), self contact and electrostatics in the elastic rod model. Therefore we must first extend the rod model to include some of these physical phenomena before exploring the relaxation of supercoiled DNA by topo IB. Specifically, we develop a method to account for electrostatics and self contact in the elastic rod model. We then exercise this new model to simulate the relaxation of DNA supercoils by topo IB. Each study corresponds to a chapter and is summarized below.

1) Electrostatics and self contact. In order to simulate plectonemes and other supercoiled geometries, in Chapter 6 we develop an efficient algorithm to account for electrostatics and self contact in an elastic rod formulation. (Such self contact is not generally expected to occur in LacI DNA loops because of the associated high electrostatic energetic penalty; see, for example, [4].) Here, we employ this algorithm to simulate the quasi-static loading leading to supercoiled DNA and draw comparisons with existing experimental and theoretical studies.

2) Modeling the relaxation of DNA supercoils. Finally, in Chapter 7 we perform our first multi-scale simulation of the relaxation of DNA supercoils by topo IB. For this study we make an approximation for the hydrodynamic drag on DNA. To simulate the interaction between DNA and topo IB, we pair our elastic rod model with a reduced order model of topo IB. The reduced order model is based on extensive MD conducted by the Andricioaei lab at the University of California - Irvine.

Chapter 2

Computational Analysis of Looping for Intrinsically Straight DNA

In this chapter we implement a computational elastic rod model to explore looping by the Lac repressor (LacI) protein for intrinsically straight DNA. This chapter sets the stage for the following chapters, especially Chapter 3 in which we consider the effects of intrinsically curved DNA, and Chapter 4 in which we consider the effects of protein flexibility. Here, we begin by reviewing the computational model used for our studies of DNA looping. Then, we characterize the effects of operator phasing for straight DNA.

2.1 Review of Elasto-Dynamic Rod Model [1]

We begin with a brief summary of the elasto-dynamic rod model as already reported in [37] so that the reader can understand the extension we contribute herein. While we must necessarily sacrifice some details for the sake of brevity, we encourage the interested reader to consult [36–38] for a complete development of these ideas.

We begin by approximating the long-length scale structures of DNA as a flexible rod having elastic properties as determined from experiments [10–12, 96], molecular dynamics simulations [9] and other biophysical techniques. The long-length scale structures of interest include loops and plectonemes. The spatial resolution limit of our model is approximately one-helical turn of the molecule (i.e., about 3.5 *nm*).

Figure 2.1 illustrates a segment of DNA and an element of a rod with equivalent averaged elastic properties. The shape of the rod (shape of the helical axis of DNA) is parameterized by the three-dimensional centerline curve $\vec{R}(s,t)$ and the cross-section fixed frame $\mathbf{a}(s,t)$

---

[1] In this section we quote significant passages from our previous publications [37, 62, 63].

where s denotes the contour-length coordinate measured from one end and t denotes time. Here we take the third basis vector of $\mathbf{a}(s,t)$ to be aligned with the tangent of the rod. (A separate basepair fixed frame (not shown) may be defined relative to this cross-section fixed frame.) The shape of the rod is also described by the curvature and twist vector $\vec{\kappa}(s,t)$ (defined as the spatial rate of rotation of $\mathbf{a}(s,t)$ [38]). Under stress-free conditions, the helical axis is in general not straight but conforms to a curved/twisted space curve. This 'intrinsic curvature' of DNA is captured by $\vec{\kappa}_0(s)$ and it depends on the basepair sequence. The change in curvature/twist, $\vec{\kappa}(s,t)-\vec{\kappa}_0(s)$, produced by any subsequent deformation of the helical axis (e.g., by protein binding), generates an internal moment $\vec{q}(s,t)$ and internal force $\vec{f}(s,t)$. The myriad inter-atomic interactions superimpose to yield a long-length scale material law, which is often assumed to be linearly elastic (see, for example, [4, 65, 75, 89, 98]). Here, that law is

$$\vec{q}(s,t) = \mathbf{B}(\vec{\kappa}(s,t)-\vec{\kappa}_0(s)) \tag{2.1}$$

where the diagonal stiffness tensor $\mathbf{B}$ includes both bending and torsional stiffness. (In this expression, $\mathbf{B}$ is diagonal when $\vec{q}(s,t)$, $\vec{\kappa}(s,t)$, and $\vec{\kappa}_0(s)$ are expressed in terms of the local frame $\mathbf{a}(s,t)$.) Consequently, the elastic energy functional for the rod is

$$E(t) = \int_0^L \frac{1}{2}[\{\vec{\kappa}(s,t)-\vec{\kappa}_o(s)\}^T \mathbf{B}\{\vec{\kappa}(s,t)-\vec{\kappa}_o(s)\}]ds. \tag{2.2}$$

Here, $E(t)$ is the elastic deformation energy and L is the total length of the molecule. As in [4, 65], we use commonly used values of the bending and torsional stiffness (50 *nm* and 75 *nm* respectively) which can be found from experimental measurements of the persistence lengths for DNA bending/torsion [5, 40, 97]. The above law renders the rod model non-homogeneous, that is, sequence-dependent by explicitly accounting for the effects of intrinsic curvature/twist $\vec{\kappa}_0(s)$. By allowing off-diagonal elements and including explicit dependence on s in $\mathbf{B}$, the rod formulation has been extended in [38] to account for sequence-dependent stiffness, anisotropy, and tension-torsion coupling [43, 66].

The dynamics of the rod is governed by the differential equations of motion below that are integrated using boundary conditions that correspond to physical loads or kinematic constraints [38]. We describe the kinematics of the molecule by the linear velocity $\vec{v}(s,t)$ and the angular velocity $\vec{\omega}(s,t)$ of the rod cross-section. The following four vector equations of rod theory [38] are numerically integrated to solve for the four vector unknowns $(\vec{f},\vec{q},\vec{v},\vec{w})$ when combined with (2.1):

$$\frac{\partial \vec{f}}{\partial s} + \vec{\kappa}\times\vec{f} = m\left(\frac{\partial \vec{v}}{\partial t} + \vec{\omega}\times\vec{v}\right) - \vec{F}_{body}, \tag{2.3}$$

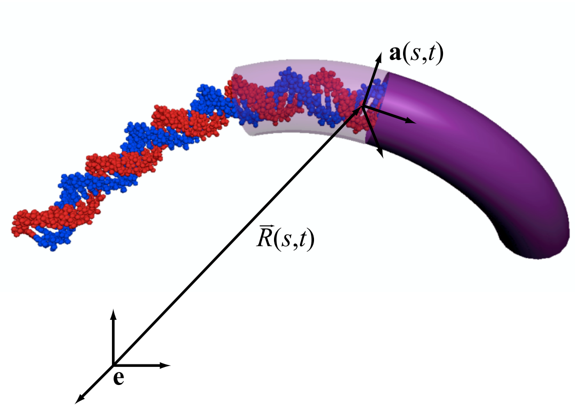


Figure 2.1 The all-atom structure of DNA as represented by a flexible rod with equivalent averaged elastic properties. The position vector $\vec{R}(s,t)$ locates the helical axis of DNA as a function of the contour-length coordinate s and time t with respect to the inertial frame $\mathbf{e}$. $\mathbf{a}(s,t)$ represents a body-fixed frame of the rod as a function of s and t.

$$\frac{\partial \vec{q}}{\partial s} + \vec{\kappa} \times \vec{q} = \mathbf{I}\frac{\partial \vec{\omega}}{\partial t} + \vec{\omega} \times \mathbf{I}\vec{\omega} + \vec{f} \times \hat{t} - \vec{Q}_{body}, \tag{2.4}$$

$$\frac{\partial \vec{v}}{\partial s} + \vec{\kappa} \times \vec{v} = \vec{\omega} \times \hat{t}, \tag{2.5}$$

$$\frac{\partial \vec{\omega}}{\partial s} + \vec{\kappa} \times \vec{\omega} = \frac{\partial \vec{\kappa}}{\partial t}. \tag{2.6}$$

Equations (2.3) and (2.4) represent the balance laws for linear and angular momentum of an element of DNA, respectively. Equations (2.5) and (2.6) are kinematical constraints that describe the (assumed) inextensible helical axis and the required compatibility between curvature and angular velocity, respectively. In this dynamic formulation, $m(s)$ denotes the DNA mass per unit contour length, $\mathbf{I}(s)$ denotes the tensor of principal mass moments of inertia per unit contour length, $\vec{F}_{body}(s,t)$ denotes any distributed body forces per unit contour length, $\vec{Q}_{body}(s,t)$ denotes any distributed body moments per unit contour length, and $\hat{t}(s,t)$ denotes the helical axis unit tangent vector. In Chapter 6 we will develop an expression for electrostatic forces and in Chapter 7 we will develop and expression for hydrodynamic drag; both expressions will be incorporated into the model through $\vec{F}_{body}(s,t)$ and $\vec{Q}_{body}(s,t)$.

The field equations, (2.3)-(2.6), are solved numerically using finite differencing and employing the generalized-alpha method in both space and time [18, 36, 38]. Discretization produces a system of nonlinear equations at every timestep that we solve for the field variables using Newton-Raphson iteration. This iterative method requires the computation of the Jacobian for the system of equations that can be particularly involved if $\vec{F}_{body}(s,t)$ and/or $\vec{Q}_{body}(s,t)$ are complicated functions of the field variables. A summary of the numerical algorithm is summarized in Chapter A.

2.2 Model Simplifications

A comprehensive study of looping for intrinsically curved DNA sequences requires significant computational effort because numerous combinations of parameters are required, including those defining the following features: 1) the geometry of the intrinsic curvature, 2) the location (phasing) of both the operator sites and the intrinsic bends, and 3) the manner in which LacI binds to the operators. We discuss next a number of simplifications which yield an efficient analysis of looping of intrinsically straight DNA. In Chapter 3 we will describe further simplifications for a study of intrinsically curved DNA.

In this chapter we simplify our study of LacI mediated looping by employing a homogeneous (straight) elastic rod representation for the DNA. Specifically, here we assume that any intrinsic curvature of the inter-operator DNA sequence is negligible compared to the bending and twisting that it sustains upon binding with LacI. (In Chapter 3 we will relax this assumption, and account for the intrinsic curvature of a large family of DNA constructs.) Therefore, an initially stress-free and straight DNA conformation provides the initial conditions for the differential equations describing the dynamics of the rod. The subsequent computation of the conformation of the looped DNA-LacI complex follows the methods described in [37].

For straight DNA, the phasing of the operator sites is determined by the helical repeat of the DNA. The term phasing refers to the torsional alignment, about the helical axis, of the two DNA operator sites. With the addition of a single inter-operator basepair, the torsional alignment changes by about 34°, and repeats every helical turn (about 10.5 bp). In addition to the phasing change associated with additional basepairs, the length of the inter-operator sequence also increases. Here we take one parameter, the length of the DNA in helical turns, to completely define the length of the elastic rod and the phasing (or relative torsional alignment) of the operators. We further assume a constant 34.6 $\AA$ length per helical turn.

Assuming a rigid LacI, the protein crystal structure dictates the position and orientation of the operators and thus the boundary conditions for the rod model. In general, there exist 8 possible binding topologies, where topology here refers to the orientation of each operator relative to the protein and the direction of the connecting DNA [33, 37]. Figure 2.2 illustrates four topologies and reviews our notation for their classification [37]. In addition, for each binding topology, one must further consider at least two twist isomers (one over- and one under-twisted). Hence, in general, each sequence requires simultaneous consideration of 8 possible binding topologies with both over- and under-twisted topoisomers. However, by assuming symmetry of the protein or DNA about their respective dyadic axes, the 8 possible binding topologies reduce to 4 unique topologies. (For example, a symmetric protein results in four pairs of indistinguishable binding topologies: **P1F** and **P1R**, **P2F** and **P2R**, **A1F** and **A1R**, and **A2F** and **A2R**.) The 4 binding topologies further reduce to 3 upon assuming both symmetric protein and DNA. (Specifically, we have three sets of indistinguishable binding topologies: **P1F** and **P1R**; **P2F** and **P2R**; and **A1F**, **A1R**, **A2F**, and **A2R**. Therefore we will refer to these topologies as **P1**, **P2**, and **A** respectively.) The intrinsically straight DNA introduced above and the symmetric protein assumption introduced next possesses the required symmetries to reduce the number of binding topologies to 3. Hence, under these mild assumptions, our analysis here requires only 6 computations at every length of inter-operator DNA (3 binding topologies $\times$ 2 topoisomers) relative to the 16 computations (8 binding topologies $\times$ 2 topoisomers) required otherwise.

Because the Lac repressor is a homotetramer, we expect its structure to be symmetric. Figure 2.3 illustrates the near-symmetry of the protein crystal structure: a 180° rotation of the right half (blue), about the dyadic axis (black arrow), superimposes nearly perfectly on the left half (red). Quantitatively, the boundary conditions for the rod model resulting from this assumption of protein symmetry deviate by less than $10\,\AA$ (in position) and 10° (in orientation) from those computed ‘exactly’ from the crystal structure [60]. Looping energetics were found to be relatively insensitive to a similar simplifying assumption used in [120], in which the three-dimensional protein structure considered herein was approximated by an even simpler two-dimensional structure.

2.3 Results and Discussion

Figure 2.4 presents the loop elastic energy (kT) as a function of inter-operator length (helical turns) where the energies for **A**, **P1** and **P2** binding topologies are denoted by the blue, red and black dotted curves respectively. Figure 2.4(a) considers phasing over a wide range of lengths (5 to 45 helical turns), while Fig. 2.4(b) focuses on a smaller range but includes

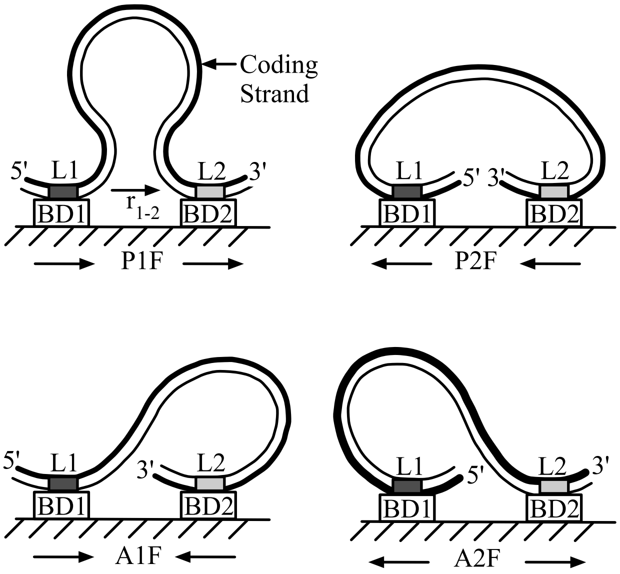


Figure 2.2 Classification of binding topologies for the LacI-DNA complex [33, 37]. There are eight possible ways for the two protein binding domains (BD1 and BD2) to bind to the two DNA operator locations (L1 and L2), four of which are illustrated here. These eight possibilities are distinguished using a 3-character code [37] which slightly extends the 2-character code of [33]. The first character (**P** or **A**) indicates if the operators are 'parallel' (**P**) or 'anti-parallel' (**A**) with respect to each other along the 5′ to 3′ direction. The second character (**1** or **2**) indicates the orientation (5′ to 3′) of the operator at L1; if it points towards the inside of the protein, it is assigned the number **1**; if it points towards the outside, it is assigned the number **2**. The third character (**F** or **R**) indicates if L1 binds to BD1 (**F** for 'forward') or to BD2 (**R** for 'reverse'). All four **F** topologies are illustrated here. With the assumption of a symmetric LacI the four **R** topologies are indistinguishable from the **F** topologies.

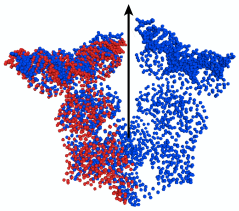
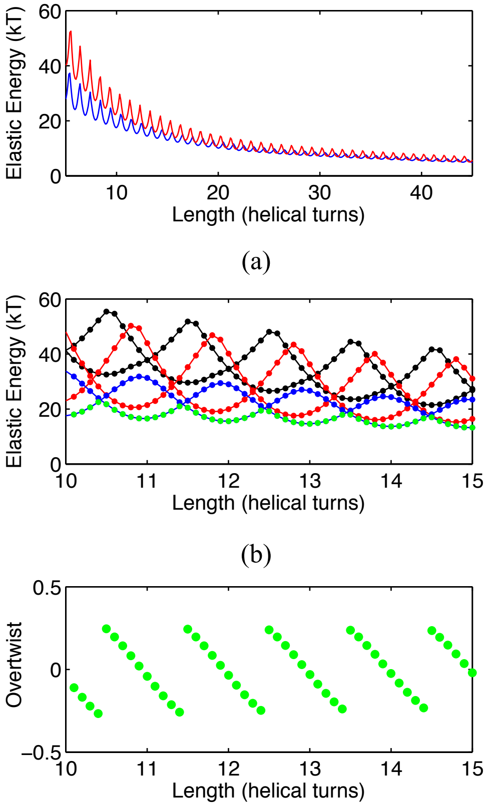


Figure 2.3 The X-ray co-crystal structure of the LacI-DNA complex [60]. In this figure, the right half (blue) of the crystal structure is rotated about the black axis by 180° and then superimposed on the left half (red). The differences between the blue and red C-alpha atoms on the left side illustrate the very slight asymmetry in the crystal structure.



(a)

(b)

(c)

Figure 2.4 (a) Total length (including phasing and as measured in helical turns) effects on the minimum elastic deformation energy (kT) for straight DNA. This plot considers a wide range of lengths and thus for clarity shows the minimum energy **A** (blue) and **P1** (red) binding topologies. (b) Phasing effects on the elastic energy (kT) of six possible loop equilibria for straight inter-operator DNA. Blue, red and black dotted curves denote results for **A**, **P1** and **P2** binding topologies, respectively. The dotted green curve highlights loops with the minimum elastic energy for any length of inter-operator DNA. (c) Phasing effects on the over-twist of the minimum energy loops in (b).

all three binding topologies and two topoisomers. The dotted green curve of Fig. 2.4(b) identifies the loop within the ensemble of loops possessing the minimum elastic energy. Figure 2.4(c) plots the twist in the minimum energy loops of Fig. 2.4(b).

Three major observations are immediately apparent upon considering Fig. 2.4. First, the average elastic energy decreases with increasing length for any binding topology. Second, the elastic energy oscillates with a period of about a helical turn. Third, the **A** binding topology is the energetically most favorable over a large range of inter-operator lengths.

The first observation is directly explainable with elementary rod theory. Rod theory predicts that the average elastic energy scales as $1/L$ for increasing length L [101]. Thus, for loops greater than 100 bp in length, the addition or subtraction of a single basepair alters the bending strain energy by about 1% or less. At first glance this decrease in energy suggests an increase in stability with increasing length. However, at longer lengths, entropic effects dominate the free energy of looping and our neglect of thermal effects is then no longer appropriate. Many studies suggest a length scale that separates 'short loops' that are dominated by elastic energy effects from 'long loops' dominated by entropic effects. The transition is believed to be on the order of a few persistence lengths [2, 77, 119]. Although our computations in Fig. 2.4(a) extend into the range in which entropy may become significant, we focus our attention on short elastic energy dominated loops (about 1 persistence length, 150 *bp*, or 14 helical turns) and comparisons among sequences of similar length, for which we agree with [65] that entropic effects are both relatively small and constant. Furthermore, short inter-operator lengths have been shown to maximize repression in-vivo [74]. In fact, Müller et al. show a maximal repression for an inter-operator length of 70.5 *bp* (measured between operator centers, or 56 *bp* as measured here and in [37]).

As one would expect, the oscillations observed in the energy curves of Fig. 2.4 are a direct result of operator phasing; one cycle corresponding closely to one complete 360° change in the torsional alignment of the operators. In contrast to the modest length effects discussed above, Fig. 2.4 shows an appreciable oscillation of strain energy upon the addition (or subtraction) of basepairs with an amplitude of about 30% of the average elastic energy for 10 helical turns of DNA; see also [37]. Consequently, we avoid describing changes in phasing as a function of basepairs (which is sensitive to the assumed number of basepairs per helical turn) and instead describe changes in phasing in the units of helical turns.

We further observe that for a given binding topology, oscillations decay with increasing length. This aligns with experimental measurements of looping rates that exhibit diminishing oscillations with increasing inter-operator length. In fact, Müller et al. could not detect changes in repression due to phasing for lengths greater than 400 *bp* [74]. Our results, in Fig. 2.4(a), show that phasing effects have significantly died out (oscillations on the order of thermal energy 1 kT) by about 30-40 helical turns or about two persistence lengths. At these lengths and longer, the flexibility of the DNA is quite significant. Also in agreement are the results of Zhang et al. that predict phasing effects to be significantly reduced beyond lengths of about 300 *bp* [119].

Interestingly, we expect any protein flexibility would reduce the stiffness of the overall system and thereby reduce the importance of phasing as suggested by [8, 119]. However, the agreement between the length at which phasing drops off considering a rigid protein and that determined experimentally provides at least some evidence that the protein is relatively stiff compared to the torsional stiffness of the DNA loop. We will address protein flexibility in more detail in Chapter 4.

Results in Fig. 2.4(a) for homogeneous (straight) DNA demonstrate that the **A** binding topology yields the energetically favorable loop (minimum elastic energy) over the range of 5 to about 20 helical turns. The dominance of the **A** topology is consistent with other computational models of the Lac repressor including [98]. Interestingly, experimental and computational work suggest that a similar protein, the Gal-repressor, also prefers antiparallel loops [33, 108].

Upon including other effects (e.g., the addition of intrinsic curvature, sequence-dependent elasticity, entropy, etc.), however, it is possible that loops with the **P1** binding topology may become energetically favorable at this length scale. Such transitions are observable at longer inter-operator lengths (greater than about 200 *bp*). The transitions appear as overlapping energy curves for the **A** and **P1** binding topologies in Fig. 2.4(a) and thus an additional (second) energetic 'cusp' is formed in each helical turn by creating a minimum energy curve from the curves for the **A** and **P1** binding topologies. In fact, Saiz et al. [85] identify such cusps (referred to as a "second component") in their Fourier analysis of experimental data [6, 74] for repression levels. They further attribute the cusps to transitions in binding topologies [87] by considering the different ways the DNA operators can bind to the protein as we do here. In a separate study involving HU [6], a similar second frequency component was recognized and attributed to IPTG bound loops. Interestingly, this cross over between binding topologies also reduces the amplitude of oscillations otherwise expected due to phasing alone.

A somewhat more subtle observation is that the computed minimum energy loops in Fig. 2.4(b)-(c) transition from under-twisted to over-twisted topoisomers near the observable energetic 'crests and troughs,' as also suggested in [91, 102, 119]. As a result, the phasing at each (local) energetic minimum produces a loop nearly free of torsional stress. An intriguing question remains regarding the possible biological significance of these near 'torsion-free' loop states which coincide with the optimum phasing for minimum elastic energy. In contrast, at the energetic crests the twist is discontinuous suggesting discontinuous writhe and hence a dramatic change in the energetically preferred loop geometry. This is a direct result of the crossing of energy curves corresponding to the over-twisted and under-twisted topoisomers of the **A** binding topology. This suggests that DNA sequences corresponding to these energetic crests could form 2 distinct loops of similar energy. Furthermore, one could consider a dynamic inter-conversion of significantly different loops as the Lac repressor binds and unbinds the DNA. Experimentally, it would be interested to design such a sequence and observe two populations of loops (perhaps using gel migration experiments) and transitions between them (perhaps with single molecule FRET).

2.4 Conclusions

In this study of Lac repressor looping we contribute a quantitative understanding of the relationship between DNA mechanics and looping for intrinsically straight DNA. This serves as the foundation for Chapter 3 and Chapter 4 which add intrinsically curved DNA and protein flexibility respectively. We also make three major observations that agree with experiments and theory. First, the average elastic energy decreases with increasing length for any binding topology. Second, the elastic energy oscillates with a period of about a helical turn. Third, the **A** binding topology is the energetically most favorable over a large range of inter-operator lengths.

Chapter 3

Computational Analysis of Looping of a Large Family of Highly Bent DNA by LacI[1]

We now extend the analysis of the previous chapter, Chapter 2, to consider intrinsically curved inter-operator DNA. Experimentally, the Kahn lab at the University of Maryland [29, 70, 73] has probed the combined effects of operator phasing and intrinsic curvature on protein-mediated looping using designed sequences that contain A-tract bends in the inter-operator DNA [70]. Three such bent sequences, referred to as 11C12, 9C14 and 7C16, form hyperstable loops proposed to have markedly different conformations, although they are not necessarily optimal representatives of any particular conformation [70]. In a recent study [37], we demonstrate that upon explicitly incorporating sequence-dependent intrinsic curvature in the DNA constitutive law, the computational rod model correctly predicts a broad range of the experimental results for the Kahn constructs. In particular, the theory successfully predicts the operator orientations (loop topologies) known from FRET measurements, the linking number distribution known from cyclization assays of the LacI-DNA complex, the relative loop stabilities known from competition assays, and the relative loop size inferred from gel mobility assays.

We consider the bent sequences above to be but three examples from a large family of molecules having the embedded A-tract. Molecules within this family are distinguished by the phasing of the A-tract relative to the terminal operators which bind to LacI. The energetics and topology of the looped DNA-LacI complexes depend critically upon this phasing, and it is uncertain if any of the three sequences synthesized to date have yielded optimal characteristics such as stability or conformational uniformity. In this study, we materially extend [37] by establishing a two-parameter representation for a complete description of the

---

[1] In this chapter we quote significant passages from our previous publication [61].

operator phasing of an entire family of related molecules. We exercise the computational rod model over the two-dimensional ‘design space’ of sequences to evaluate the landscape of possible looping characteristics including loop energy/stability. Doing so illustrates that combining the operator phasing experiment with the unique variety of experiments possible on hyperstable loops may allow us to design loops with intriguing conformational switching properties as well as to test the model’s assumptions about protein and DNA flexibility.

3.1 Methods

In Section 2.2 we indicated that a comprehensive study of looping can be simplified by making appropriate approximations for the following features: 1) the geometry of the intrinsic curvature, 2) the location (phasing) of both the operator sites and the intrinsic bends, and 3) the manner in which LacI binds to the operators. In this chapter we make use of many of the approximations introduced in Chapter 2. However, instead of representing DNA as a homogeneous straight rod, we now introduce a new simplification to aid in the modeling of a large family of intrinsically curved DNA molecules.

Figure 3.1 introduces a simplified ‘Straight-Helical-Straight’ (SHS) representation for describing intrinsic curvature with the rod model [37]. Bent sequences related to the three synthesized sequences introduced in [70] are modeled as a helically supercoiled bent A-tract domain (dark gray) flanked by two straight linker domains (light gray). Figure 3.1 also illustrates the SHS rod representation superimposed upon an atomistic representation of the stress-free and zero temperature conformation of one such sequence, 11C12 from [70]. The close agreement between the two representations follows from the fact that repeating A-tracts largely bend the helical axis of DNA into a helical supercoil; see, for example, [13, 25]. The pitch and the radius of the superhelix depend on the details of the dependence of curvature on sequence. In Appendix B we show that the SHS representation approximates the helical axis of all three bent sequences of [70] to within an RMS error of less than the radius of DNA (less than 10 $\AA$).

The SHS representation is symmetric in that the right and left halves are identical to within a 180° rotation about the out-of-plane axis at the midpoint. This symmetry is not a fundamental limitation of the rod model, which is capable of accounting for arbitrary intrinsic curvature [37], but here we use the symmetry to reduce the computational cost of evaluating an entire family of bent sequences. The mathematical definition of the SHS representation, as provided in Appendix B, ultimately defines the intrinsic curvature/twist $\vec{\kappa}_o(s)$ employed in the above elastic energy functional for the computational rod model. The

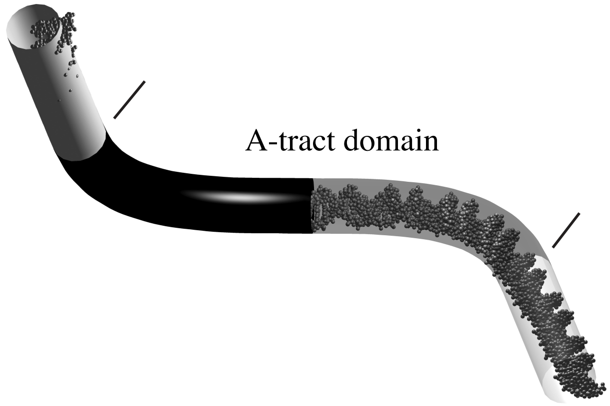


Figure 3.1 The Straight-Helical-Straight (SHS) approximation for highly bent DNA sequences. The light and dark gray tube illustrates the SHS approximation to the stress-free, zero temperature conformation of the bent sequence 11C12 introduced in [70]. The light gray portions of the tube are the straight "linker" domains and the dark gray portion is the superhelically curved "A-tract" domain. The transparent half of the tube reveals an atomistic representation of the 11C12 sequence provided by the web-based tool "model.it" [109]. (Model.it is just one model for the intrinsic curvature of DNA. Each model is somewhat different, and therefore results in conflicting intrinsic curvature. We consider sensitivity to our use of this model in Appendix C.) Other models also exist to predict the intrinsic curvature of DNA.) The SHS representation is symmetric: the left and right halves are related by a 180° rotation about the out-of-plane axis at the midpoint. Appendix B provides the mathematical definition of the SHS representation. The SHS representation approximates the helical axes of the bent sequences [70] to within an RMS error of less than the radius of DNA (less than $10\,\mathring{A}$).

symmetry of the SHS representation, similar to the straight DNA of Chapter 2, reduces the number of independent binding topologies that must be considered. The SHS representation of the initial DNA conformation provides the initial conditions for the differential equations describing the dynamics of the rod.

Individual molecules within the family of bent sequences are distinguished by the phasing of their operators relative to the A-tract bend, which determines the orientation of the bound protein relative to the bend center. (For straight DNA, only a single phasing parameter is required, describing simply the torsional alignment of the terminal operator sites.) The orientation of the helix is defined by a vector triad, the three principal directions of the space curve corresponding to the tangential, normal and bi-normal vectors. Figure 3.2 defines the two phasing parameters, Θ_1 and Θ_2, by illustrating the torsional alignment of two triads. The gray triad represents the orientation of the helix at the end of the A-tract domain, and the black triad the orientation of the first basepair of the operator within the linker domain. The angle Θ_1 between the triads defines the torsional alignment or 'phasing' of the A-tract

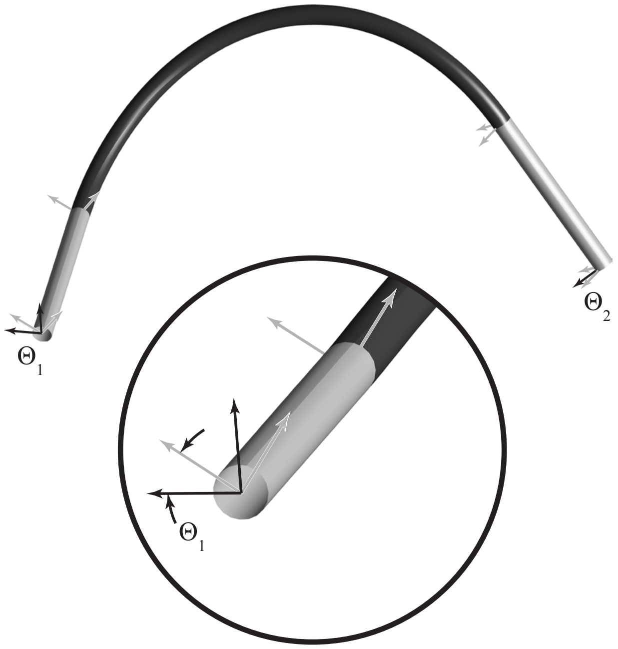


Figure 3.2 SHS representation showing the definition of the phasing parameters Θ_1 and Θ_2 in terms of triads aligned with the principal directions (tangential, normal and bi-normal unit vectors) of the DNA. The gray triad is aligned with the end of the helical (A-tract) segment, and the black one is at the first basepair of the operator at the end of the straight (linker) segment. The enlarged view shows that Θ_1 is the angle formed between corresponding vectors for the gray and black triads. The Θ_1 and Θ_2 phasing parameters vary between 0 and 1 helical turns upon adding or subtracting basepairs to the linker domain, as in the three molecules in [70]. Θ_1 and Θ_2 define the torsional alignment of the operators relative to each other and to the central superhelical domain. The first and second numbers in the designations of the three bent sequences (11C12, 9C14 and 7C16) define their linker lengths and correspond to the phasing parameters Θ_1 and Θ_2 respectively.

to this operator. An analogous definition holds for Θ_2 at the opposite end.

For intrinsically straight DNA, the torsional phasing of the operators has a much larger effect on the elastic energy of looping than the modest effect of small changes (< 3 *nm*) in contour length; see, for example Chapter 2 and [37]. Accordingly, we simplify our calculations by exploring up to complete 360° changes of the phasing parameters (Θ_1 and Θ_2) while keeping the overall contour length of the SHS representation constant. Experimentally, the phasing parameters are changed by adding or subtracting basepairs in the linker domains [70], where adding a single basepair changes the associated phasing parameter by approximately 34°. The first and second numbers in the designations of the three bent sequences (11C12, 9C14 and 7C16) define their linker lengths and correspond to the phasing parameters Θ_1 and Θ_2 respectively. We report Θ_1 and Θ_2 in units of 'helical turns' where one 'helical turn' represents a 360° change in the torsional alignment of the associated operator relative to the A-tract domain. Thus, the family of bent sequences with a common A-tract domain reduces to a two-parameter family of curves distinguished by Θ_1 and Θ_2. Allowing these phasing parameters to vary through their entire range (0 to 1) yields a large family of sequences, the majority of which have not been synthesized, but which includes the three synthesized sequences of [70].

We estimate the positions of the designed sequences in the Θ_1-Θ_2 design space by mapping the rod representations of the stress-free conformations [37] onto the SHS representation introduced in this chapter. To this end, we first superimpose the helical axes of the SHS representation with that obtained in [37] based on a consensus tri-nucleotide model for DNA [109]. Next, we project the triads defining the operator phasing from the rod representations onto the SHS representation as in Fig. 3.2. This projection determines the values of the phasing parameters (Θ_1 and Θ_2) for the SHS representation that best approximate the corresponding rod representation. Note that since the rod representation for the helical axis [37] originates from an assumed DNA model, the consensus tri-nucleotide model [109], the approximate values of Θ_1-Θ_2 are estimates that will vary if an alternative DNA model is used or if the DNA helical repeat changes significantly.

3.2 Results

The looping of three highly bent DNA sequences by the Lac repressor (LacI) protein has been characterized using a wide array of experimental techniques [29, 70, 73]. In a recent theoretical study [37], we demonstrate that the computational rod model correctly predicts the major experimental findings for these three specific sequences. We now significantly extend these predictions to the entire two-parameter family of related bent sequences, as described by the SHS approximation defined above and in Appendix B. While many of these sequences have not yet been synthesized, the theoretical results below reveal intriguing possibilities for future experiments.

Computational (e.g., [37, 98, 119, 120]) and experimental (e.g., [6, 8, 74]) studies of phasing effects for 'straight' DNA often report the free energy (or repression level) as a function of a single independent phasing parameter (often the contour length in basepairs). Such results must now be extended to demonstrate the simultaneous dependence on two independent phasing parameters for the highly bent sequences. The two phasing parameters Θ_1 and Θ_2 (see Fig. 3.2), given in units of helical turns, distinguish bent sequences within this family.

Figure 3.3 illustrates the computed elastic energy cost of looping for all bent sequences over the possible sequence 'design space': $0 \leq \Theta_1 \leq 1$ and $0 \leq \Theta_2 \leq 1$. The result is an energy 'contour map' where the contours map the loci of designs having equivalent elastic energy (kT) cost for loop formation. For each specified value of Θ_1 and Θ_2, we report the elastic energy of the minimum energy looped conformation, from among all three possible binding topologies and considering both over- and under-twisted topoisomers. (In Fig. 3.3 the computed elastic energy is symmetric about the diagonal $\Theta_1 = \Theta_2$. This symmetry is a direct result of the prescribed symmetries of the protein and bent DNA.) Also illustrated in Fig. 3.3 are three specific computed loops: the minimum energy loop that forms with the **A** (anti-parallel) binding topology sketched in Fig. 2.2 (4.0 kT), the minimum energy loop that forms with the **P1** (parallel 1) binding topology (4.2 kT), and the overall maximum energy loop, which also forms with the **A** binding topology (12.3 kT). These energetic differences are experimentally testable as in [70].

The rod model is versatile and efficient, especially considering the approximations utilized herein. Consequently, we can readily explore the sensitivity to many assumptions used to generate Fig. 3.3. For example, in Appendix C we present preliminary results illustrating the effects of changes in the SHS representation on the landscape of Fig. 3.3. Specifically, we test the sensitivity to the magnitude of the curvature in the bent domain and the position of the bend along the length of the rod. In addition, we present in Chapter 4 the sensitivity of Fig. 3.3 to flexibility of the Lac repressor protein.

The looped complexes formed by the original three bent sequences exhibited significant differences in gel migration behavior [70]. Although we cannot predict quantitative electrophoretic mobilities for these complexes, we expect the radius of gyration (R_g) to correlate to the migration speed, as suggested previously for the three synthesized sequences [37]. In Fig. 3.4, we report the radius of gyration of the minimum-energy DNA-LacI complex for each value of Θ_1 and Θ_2. (The R_g is estimated as the root mean square distance of each atom in the DNA-LacI complex from the average position of all the atoms. We lump the average number of atoms per basepair into the associated spatial grid point in our discretized rod model, and we lump the average number of atoms per amino acid into a single point located at the C-alpha atoms in the crystal structure, because only the C-alpha atoms are resolved in the crystal structure [60].) In Fig. 3.4 we again demarcate the regions of the energetically preferred binding topologies as identified in Fig. 3.3. These borders between energetically preferred binding topologies often result in discontinuities in R_g. In addition to discontinuities arising from a change in preferred binding topology, discontinuities may also arise from a change in preferred topoisomer (i.e., over- or under-twisted). Figure 3.4

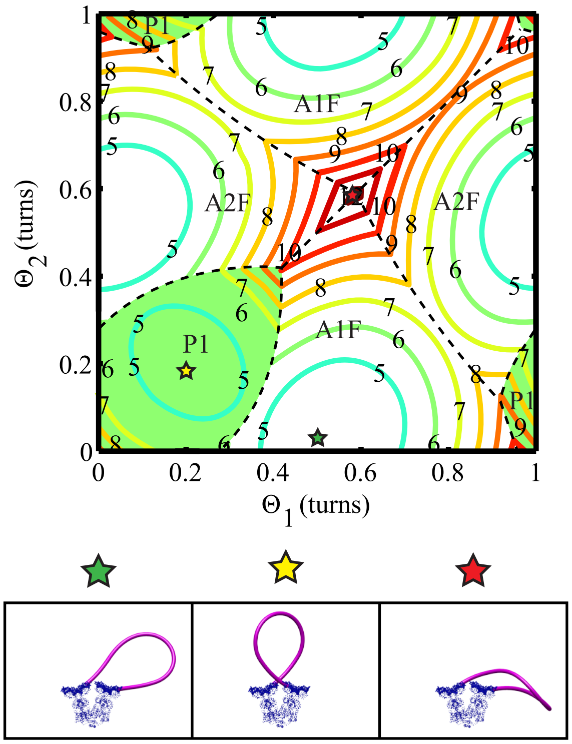


Figure 3.3 The energetic cost of looping over the entire Θ_1 and Θ_2 sequence ‘design space’ of highly bent sequences with a common bent A-tract domain. The elastic energy (kT) of looping for the minimum energy conformations is reported, considering all binding topologies for both over- and under-twisted topoisomers. Distinct (minimum-energy) binding topologies are delineated with dashed black lines. The green regions correspond to bent sequences that preferentially bind with parallel (**P1**) binding topologies. All other sequences, the white region, preferentially bind with anti-parallel (**A**) binding topologies. Three computed loops are shown: the minimum energy loop with the **A** binding topology (4.0 kT), the minimum energy loop with the **P1** binding topology (4.2 kT), and the overall maximum energy loop, which adopts the **A** binding topology (12.3 kT). (The assumed DNA and protein symmetries make the **A1** and **A2** binding topologies indistinguishable.)

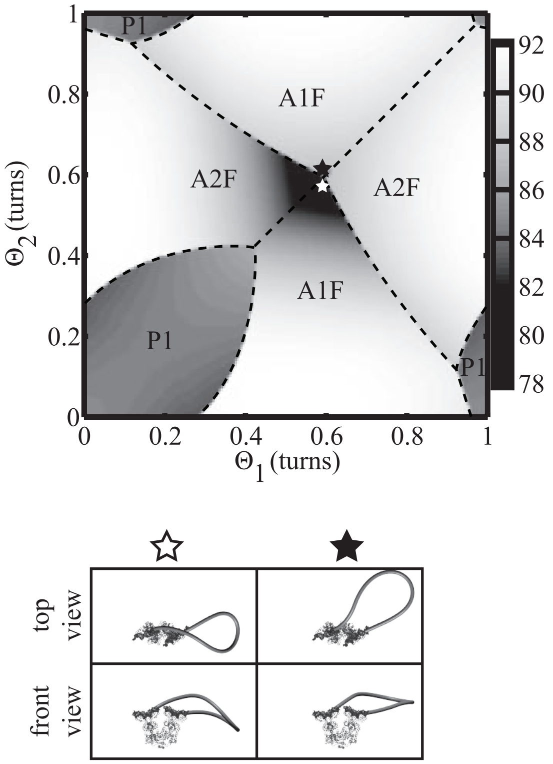
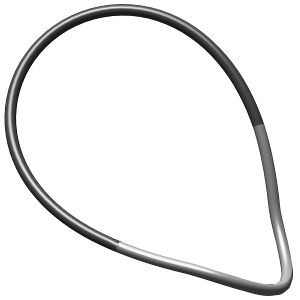


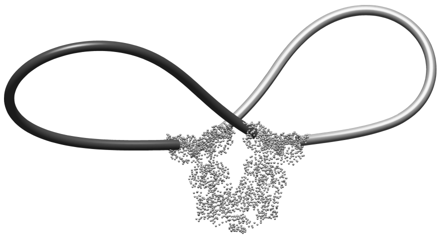
Figure 3.4 Computed radius of gyration (*Å*) of the minimum energy looped complexes reported in Fig. 3.3. Preferred binding topologies are demarcated with dashed lines. Some of these borders between binding topologies result in discontinuities in the R_g. Discontinuities may also arise from the change in preferred topoisomer (i.e., over- versus under-twisted) as in the two example loops illustrated below ($\Delta Tw = -0.13$ and $\Delta Tw = 0.07$ for the loops designated by the white and black stars respectively).

illustrates two sample looped complexes, one on each side of a discontinuity. ($\Delta Tw = -0.13$ and $\Delta Tw = 0.07$ for the loops designated by the white and black stars respectively.)

Mehta and Kahn [70] carried out ligation reactions that yielded minicircles formed by cyclizing the free ends of ~ 350 *bp* DNA molecules following formation of embedded ~ 150 *bp* DNA-LacI loops. They measured ΔLk, the difference between the *Lk* of DNA cyclized while bound to LacI and the *Lk* of DNA cyclized in the absence of LacI. They interpreted the results in terms of changes in the LacI conformation (open and closed forms), but our earlier work showed that the results can also be explained by considering over- or under-twisting in the loop as well as various binding topologies of the loop. Extending this analysis and the experiments to the entire family of bent sequences should resolve this issue of protein bending flexibility vs. DNA twisting and binding topology changes.



(a)



(b)

Figure 3.5 Computed minicircles mimicking the experimental procedure of [70]. Light gray segments represent the highly bent sequence (SHS representation) and the dark gray segments represent the intrinsically straight sequences forming the 'tails.' Since the operator DNA is considered fixed by the protein crystal, is a visible gap between the light and dark gray rods. (a) Computed minicircle cyclized in the absence of LacI. For this sample minicircle $\Delta Tw^a = 0.38$ and $Wr^a = 0.16$. (The superscript (*a*) is used to denote that Tw and Wr refer to the state depicted in (a).) (b) Minicircle cyclized subsequent to the formation of the looped DNA-LacI complex. ΔLk is defined as the Lk of the minicircle cyclized in the absence of LacI minus the Lk of the minicircle cyclized subsequent to the formation of the LacI-DNA loop. This sample minicircle is formed from the minimum energy loop illustrated in Fig. 3.3 with $\Delta Tw^b = 0.14$, $Wr^b = -0.56$. The resultant change in Lk is: $\Delta Lk = \Delta Tw^b + Wr^b - (\Delta Tw^a + Wr^a) = -1$. (Because of computational error, here ΔLk is rounded.)

Figure 3.5(a) illustrates DNA cyclized in the absence of LacI, which establishes the baseline Lk for computing ΔLk. Figure 3.5(b) illustrates the minicircle formed by cyclizing the free ends of the DNA after first forming the DNA-LacI complex for the minimum energy loop that forms with the **A** (anti-parallel) binding topology (and also illustrated in Fig. 3.3).

The cyclization experiment was modeled numerically using the computational rod model (Fig. 3.5) by dividing a minicircle into an SHS domain and two straight domains. The SHS domain, with its prescribed intrinsic curvature (light gray segments in Fig. 3.5), ultimately forms the primary DNA loop, and the straight domain (dark gray segments in Fig. 3.5) represents the DNA 'tails' outside the inter-operator region (with a length of 211 *bp*). For a given set of phasing parameters Θ_1 and Θ_2, we first simulate the formation of a minicircle in the absence of LacI, as in Fig. 3.5(a). Because the two ends of the SHS representation are connected by a straight segment of DNA and there is no bound protein, only one phasing parameter is required, the sum of Θ_1 and Θ_2. Next, we select the minimum energy looped DNA-LacI complex as in Fig. 3.3. The operator DNA is considered fixed by the protein crystal structure, and based on the kinetic stability of these complexes we assume that the geometry and topology of this loop remain unchanged during ligation. Starting from this state, we simulate the formation of a secondary loop formed by the tails, which now ligate using the only available binding topology. After choosing the energetic minimum from the under- and over-twisted topoisomers, we compute the *Lk* of the final minicircle with LacI; refer to the example computed minicircle in Fig. 3.5(b). Because the length and phasing of the straight DNA tails that form the secondary loop are held constant, only one minimum energy loop of each binding topology must be computed. As a result, all **A1** primary loops (formed with arbitrary phasing of a bent sequence) are closed with identical **A2** secondary loops, with a constant calculated elastic energy of 15.5 kT. All **P1** primary loops are closed with identical **P2** secondary loops, with energy 19.6 kT. We report in Fig. 3.6 the computed ΔLk between the loops formed with and without LacI and over the entire design space. The preferred binding topology of the primary loop is denoted in the figure (thin dashed lines). In addition, we identify in Fig. 3.6 the discrete changes in *Lk* of the energetically preferred minicircle formed in the absence of LacI (thick dashed lines), which affects the baseline for calculating ΔLk.

3.3 Discussion

In a previous paper [37], we established that the computational rod model predicts the major experimental findings for the looping of three bent sequences synthesized to date [29, 70, 73]. In this chapter, we introduce the SHS representation for an entire family of these bent sequences and systematically explore possible looping behaviors. Doing so reveals new insights on how intrinsic curvature influences the looping of DNA, which in turn motivates future experimental studies.

The differences in the looping behaviors of the three highly bent sequences (7C16, 9C14, and 11C12) [70] originate from the differences in phasing of the A-tract domain. In particular, the three sequences shift the A-tract domain by 2 *bp* relative to the ends of the linker domains, which leads to a substantial (approximately 70°) torsional phase difference of the A-tract from one sequence to another. The energy contour plot of Fig. 3.7 illustrates

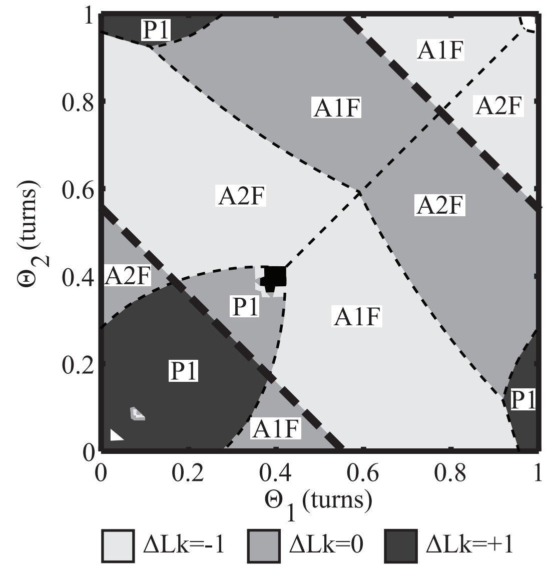


Figure 3.6 Map of relative linking number (ΔLk) for minicircles formed from the DNA-LacI complex, with ΔLk defined as in Fig. 3.5. ΔLk is computed over the entire design space. Preferred binding topologies are demarcated with dashed lines as in Fig. 3.4. The thick dashed lines denote discrete changes in ΔLk resulting from discrete changes in Lk of the DNA minicircle formed in the absence of LacI. Isolated discontinuities in ΔLk within the **P1** region along the diagonal ($\Theta_1 = \Theta_2$) arise from changes in writhe (± 2) due to changes in the handedness (sign) of the crossings of the primary and secondary loops; see, for example, [13]. (The ΔLk jumps at a few isolated areas within the **P1** region. Further investigation is necessary to determine if these are real or artifacts of the numerical approximations.)

the approximate positions of the three synthesized sequences within the Θ_1-Θ_2 design space, estimated as described in Section 3.1. Note that the three synthesized sequences span only a modest fraction of the overall design space, which has been largely unexplored to date. The assignment of Θ_1 and Θ_2 depends on an assumed helical repeat and DNA curvature model, but the relative properties of DNA constructs differing by the indicated number of turns would be unaffected by errors in the absolute positions of the existing molecules.

The larger view of the design space illustrated in Fig. 3.7 is obtained by ‘tiling’ the energy contour plot of Fig. 3.3 assuming that the loop elastic energy remains periodic in the two phasing parameters Θ_1 and Θ_2. (We again emphasize that any variation in the loop elastic energy due to the small changes in the sequence contour length are inconsequential relative to those induced by changes in phasing [37]). To highlight this periodicity, Fig. 3.8 presents one cross-section of the energy landscape of Fig. 3.7. The periodic variation in elastic energy for this family of bent sequences is strikingly similar to the periodic variation in repression level (or elastic energy and free energy) for nominally straight DNA due to

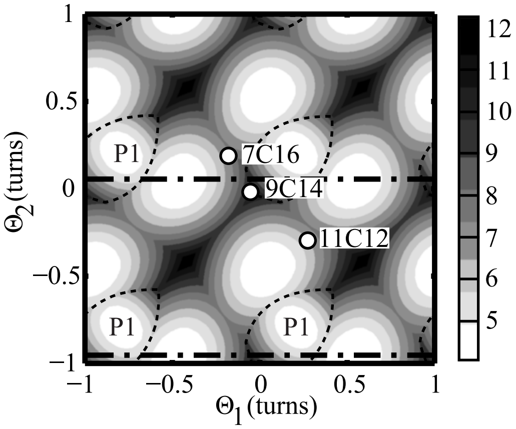


Figure 3.7 Elastic energy contour map (Fig. 3.3) extended using tiling. The thin dashed lines outline the regions corresponding to the **P1** binding topology; the remaining region corresponds to the **A** binding topology. The indicated locations of the synthesized sequences are a result of mapping the representation of intrinsically curved DNA used in [37] onto the SHS representation. The thick dashed lines indicates the plane for the cross-section of Fig. 3.8. (These two lines are separated by one turn and therefore result in identical cross-sections.)

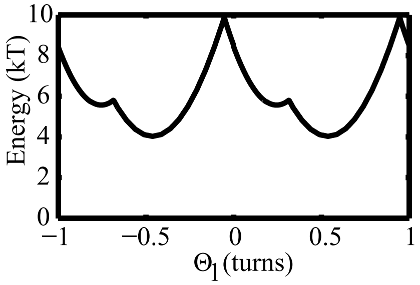


Figure 3.8 Cross-section of the energy landscape along the thick dashed lines of Fig. 3.7.

torsional phasing as discussed, for example, in [6, 8, 58, 74]. Figure 3.7 illustrates that there exist distinct regions of the design space where the preferred binding topology is **P1** (outlined with thin dashed lines) versus **A**. Thus, by appropriate experimental design, one may bias the preferred binding topology. Also, the elastic energy cost of looping varies significantly - by more than a factor of two - over the design space. Again, this suggests that by appropriate experimental design, one may significantly influence the loop stability. We explore these possibilities further below.

Although it is evident from Fig. 3.7 that the '**A** region' (outlined with thin dashed lines) is larger than the '**P1** region', both occupy significant fractions of the design space (i.e., both regions encompass many possible bent sequences). By contrast, only the **A** binding topology is preferred for intrinsically straight DNA of similar contour length, assuming no protein flexibility [98, 120]. Thus, for the '**P1** region' of the design space, properly phased intrinsic curvature is necessary and sufficient to decrease the elastic energy of **P1** enough to overcome the bias towards the **A** binding topology. As with intrinsically straight DNA on these length scales (approximately 140 bp), the **P2** binding topology is never preferred [120].

The experimental and computational data for the three synthesized sequences already provide evidence for two distinct looped states: qualitatively different topoisomer distributions and FRET behavior have been observed for the 9C14 and 11C12 molecules. The existence of the two states was proposed to arise primarily from flexibility of LacI at the base of the 'V' [29, 50, 70]. The results presented here and in [37] suggest an alternative explanation, that distinct looped states could arise from distinct preferred binding topologies, as also proposed in [87] for straight DNA. The role of LacI flexibility has been examined experimentally and computationally but a clear consensus has not yet emerged. For example, an MD model of the LacI protein suggests a stiff protein 'V' [107] with flexible 'head' domains. By contrast, electron microscopy and X-ray scattering demonstrate a flexible 'V' region in solution [83, 100]. DNA looping models that include approximate treatments of protein flexibility align reasonably well with experimental data [50, 98, 120] and thereby support a flexible 'V'.

Whether protein flexibility or binding topology is the root cause of the observed multiple looped states is a persistent question which could be resolved by further experiment. Some of the most compelling experimental data derive from FRET assays on DNA molecules where donor and acceptor fluorophores are attached at opposite ends of the inter-operator DNA sequence. Single molecule measurements on the looped LacI-9C14 complex give a high FRET efficiency, corresponding to a distance of about 35 $\AA$, indicating that the 9C14 sequence forms primarily **P1** loops with no observable protein deformation [73]. This experiment is well suited to detect the **P1** topology, but is unable to distinguish other binding topologies or protein flexibility because of their expected low FRET efficiency. Repositioning one fluorophore to the DNA tail just outside the operator domain could allow FRET detection of the **A** binding topology, as also suggested in [98]. A high FRET efficiency would confirm the **A** binding topology, whereas a low FRET efficiency would confirm protein flexibility. Another experiment to probe protein flexibility would start with a loop suspected of having large protein deformation and exhibiting low FRET efficiency. The loop could then be cleaved using a restriction enzyme or alternatively nicked at selected sites, significantly reducing its stiffness. Negligible protein deformation would be implicated if the FRET efficiency did not change. Conversely, a significantly increased FRET efficiency would confirm protein deformation in the initial loop. However, if the system equilibrates rapidly after cleaving the DNA, the sandwich complex could also form at random and thereby confound the experiment.

A second major result apparent in Fig. 3.3 is that the energetic cost of looping varies by more than a factor of two (from approximately 4 kT to 12 kT) across the family of bent sequences. Consequently, one would expect large differences in loop formation and breakdown rates and loop stability for representative bent sequences which span the design space. (If we assume that the entropic contribution to the free energy of looping is nearly the same for all sequences in the family of bent sequences, we calculate the relative equilibrium constants between the minimum and maximum elastic energy loops to be: $k_{loop,min}/k_{loop,max}$ =exp$[12-4] \sim 3000$.) The energy contour map in Fig. 3.3 could be used as a guide to design new sequences to minimize (or maximize) the elastic energy cost of looping, starting from the three previously synthesized sequences. Sequences could be synthesized specifically to probe the energetic extremes leading to comparatively hyper- and hypo-stable loops. Interestingly, the elastic energy cost of looping for the local minima within the **P1** and **A** regions are nearly equivalent, suggesting that hyper-stable **P1** and **A** loops could each be produced. They would prove useful in future experiments due to their extreme stability.

Another possibility arises from considering the borders between the **P1** and **A** regions. The bent sequences near these borders may form loops with distinct binding topologies but with near-equivalent energetic costs, so that one would expect to observe a near-equal distribution of loops having two distinct binding topologies. In single molecule experiments (e.g., SM-FRET), one might further hope to observe interconversions among the different looped states and the unlooped state. Though Edelman et al. [29] suggest that such interconversions occur on a long time scale for hyperstable loops because of their energetic stability, the inter-conversions could be accelerated with low concentrations of the inducer

IPTG. Any such transitions were not observable in the SM-FRET study of [73] because the looped complexes freely diffused through the microscope's field of view.

The radius of gyration R_g for the computed loops reported in Fig. 3.4 is a measure of loop size and thus one factor influencing the speed of a looped complex through a gel. However, the gel matrix or ionic conditions in the gel may alter the preferred loop topology. Among the three sequences synthesized to date, the experimental mobilities for loops that form the **P1** binding topology are larger (suggesting a more compact loop) than the mobilities of loops believed to have the **A** binding topology or an open form LacI [37]. This result holds true for much of the bent sequence design space, as seen by the dashed lines of Fig. 3.4 demarcating the preferred binding topology. (An exception exists in a very small region in the center of Fig. 3.4 where compact loops with the **A** binding topology are predicted.) We further observe in Fig. 3.4 that discontinuities in R_g are a result of changes in preferred binding topologies and topoisomers. The discontinuities in loop size suggest the design of gel assays to detect distinct binding topologies and/or topoisomers as a function of the two phasing parameters Θ_1 and Θ_2.

The computed change in linking number ΔLk is also discontinuous across regions of preferred binding topologies as well as preferred topoisomers. In addition, we observe that ΔLk=+1 occurs only within the '**P1** region.' This fact suggests yet another experiment to detect loops that preferentially form with the **P1** binding topology as the discrete changes in ΔLk are relatively easy to detect.

Calculations of ΔLk depend on the topologies of three important components: the minicircle formed in the absence of LacI, the primary loop, and the secondary loop. The influence of the minicircle formed in the absence of LacI becomes pronounced at the discontinuities marked by the thick dashed lines in Fig. 3.6. These discontinuities result from a change in the energetically preferred topoisomer as the total length of the minicircle changes. As previously discussed, this length depends upon the sum of Θ_1 and Θ_2. Thus, these discontinuities follow the *diagonal* lines ($\Theta_1 = -\Theta_2$) in Fig. 3.6. The primary loop has the most pronounced effect on ΔLk. The primary loop is formed by the most energetically favored binding topology and topoisomer available for a pair of phasing parameters Θ_1 and Θ_2. Changes in preferred binding topology or topoisomer with changes in these phasing parameters result in the majority of the discontinuities observable in Fig. 3.6. Finally, the topology of the secondary loop has relatively little impact on ΔLk because it remains constant, regardless of phasing, among all complexes in which the primary loops have the same binding topology. The role of the secondary loop is, however, quite pronounced at very isolated points within the **P1** region along the diagonal ($\Theta_1 = \Theta_2$) of Fig. 3.6. These points arise from the changes in writhe (± 2) due to changes in the handedness (sign) of the crossings of the primary and secondary loops; see, for example, [13]. In these regions the effect of electrostatic repulsion, which we do not consider, may be important.

In addition to measuring the ΔLk of the minicircles, Mehta and Kahn [70] also measured the corresponding cyclization rates. These rates are strongly influenced by the energetic cost of forming the secondary loop from the DNA tails outside the operator region. Experiments on the three synthesized sequences show that loops having ΔLk=+1 form more slowly than all others. Figure 3.6 supports this observation. In particular, note that ΔLk=+1 arises only for loops with the **P1** binding topology. During cyclization, the tails must therefore form a secondary loop with the (remaining) **P2** binding topology. However, as noted in the discussion above of the energetic cost of looping, the **P2** binding topology leads to the least energetically favorable loops, a finding also in agreement with [120]. In fact, we calculate the energetic cost of forming a loop from the DNA tails with the **A** and the **P2** binding topologies to be 12.5 kT and 19.6 kT respectively.

The landscapes presented here suggest that it would be useful to synthesize larger sets of molecules in which Θ_1 and Θ_2 are varied systematically. Significant differences in gel migration speed between sequences differing by a single basepair could indicate a sudden change in preferred binding topology or preferred topoisomer. Gel assays offer a simple means to distinguish loops that generate different topoisomers. FRET or tethered particle microscopy techniques can address loop topology and the conformation of the protein.

These computational results also support suggestions based on experiment that LacI-DNA loops are dynamic entities, and that intrinsic DNA curvature in the loops can control their shapes. Cellular DNA bending proteins could certainly have the same effect. Dynamic transitions arising from compositional, chemical or thermal signals in the cell might play an important role in the control of gene expression.

3.4 Conclusions

This study materially extends the theoretical results of [37] by analyzing the properties of a family of LacI-DNA loops, using a computational rod model for highly bent inter-operator DNA. The extension follows from considering bent sequences composed of two straight linker domains which flank a common helically-supercoiled (A-tract) domain. A unifying straight-helical-straight (SHS) representation for the unstressed state leads to a family of molecules that are distinguished by two phasing parameters, Θ_1 and Θ_2, specifying the torsional phase of the A-tract relative to each of the operators. By exploiting the near-symmetry of LacI and the inter-operator DNA, we significantly reduce the computational effort in analyzing looping for the entire space of molecules (i.e., over all possible torsional phases $\Theta_1 \in [0,1]$ and $\Theta_2 \in [0,1]$ turns). The resulting computations of loop energy and topology reveal new observations of looping mechanics and further suggest compelling experimental studies.

First, the entire two-parameter 'design space' is composed of two subspaces of bent sequences that prefer to bind to LacI with parallel (**P1**) versus anti-parallel (**A**) binding topologies. Thus, the distinct looped states observed for three previously synthesized bent sequences [29, 70] could arise from differences in preferred binding topology instead of solely from protein flexibility. Second, sequences located near the parallel versus anti-parallel 'borders' in the design space form loops with distinct binding topologies but with near-equivalent energetic cost. Thus, such loops might readily interconvert between **P1** and **A** binding topologies (and unlooped states) due to thermal energy. Third, the energetic cost of looping varies by more than a factor of two (from approximately 4 kT to 12 kT) over the design space, confirming the significant influence of DNA intrinsic curvature on looping. As a result, it should be possible to design bent sequences with controlled looping kinetics and stability.

The ability to synthesize any sequence in the family of bent sequences discussed above creates several possibilities for future experimental studies. For instance, the energy contour map (Fig. 3.7) may guide the design of new sequences that minimize (or maximize) the elastic energy cost of looping leading to hyper-stable (or hypo-stable) looped complexes. Second, to probe whether multiple looped states [29, 70] develop from changes in binding topology or protein flexibility, one may extend the previous FRET experiments [29, 73] by repositioning one fluorophore just outside the operator domain to detect the **A** binding topology. Gel assays may provide a ready means to observe the distinct (discontinuous) changes in the size of the looped complexes (R_g) and/or the distinct changes in linking number (ΔLk) within known regions of the design space. Finally, possible biological roles of putative loop switching could be addressed in-vivo with repression experiments similar to those originally used to identify the presence of LacI-anchored DNA loops.

Chapter 4

The Effects of Lac Repressor Flexibility on DNA Looping

In previous chapters we approximate the Lac repressor as a rigid protein defined by its crystal structure [31, 60]. However, there is accumulating evidence suggesting the Lac repressor is flexible. For example, observations from X-ray crystallography [31, 60], electron microscopy [57, 58, 83, 94], atomic force microscopy [108], neutron and X-ray scattering [15, 69, 100], and single molecule tethered particle motion [41, 76, 84, 105, 117] all provide experimental evidence supporting a flexible protein.

Additional evidence for Lac repressor flexibility comes from bulk and single molecule FRET studies by the Kahn lab at the University of Maryland [29, 73]. Currently, the Kahn lab is performing an aggressive series of bulk FRET measurements on 25 sequences forming a 5×5 grid over the design space introduced in Chapter 3; and the lab is planning a related series of single molecule FRET assays. In their previous FRET studies, both donor and acceptor fluorophores were placed within the inter-operator DNA near the operator sites. This configuration is most sensitive to loops formed with the **P1** binding topology. The current study considers an additional three configurations of the FRET markers: (1) the donor repositioned just outside the operator site while the acceptor remains within the inter-operator DNA, (2) the acceptor is repositioned outside the operator site while the donor remains within the inter-operator DNA, and (3) both donor and acceptor are repositioned outside the operator sites. The four FRET configurations are denoted D0A0, D1A0, D0A1, and D1A1 respectively. (The ‘D0’ or ‘D1’ refers to the position of donor fluorophore inside and outside the inter-operator DNA respectively. Similarly, the ‘A0’ or ‘A1’ refers to the position of the acceptor fluorophore inside and outside the inter-operator DNA respectively.) Results from these FRET studies are expected to shed light on the role of protein flexibility and binding topology in DNA looping.

Many theoretical and computational studies of Lac repressor looping build additional evidence in support of protein flexibility; see for example [50, 86, 98, 103, 107, 120]. Despite this abundance of data, there remain many unknowns about the degree of flexibility and the associated modes of deformation. However, the experimental and computational efforts to date suggest the possibility of concentrated flexibility at the apex of the protein 'V' and/or at the head domains. Here we consider varying degrees of flexibility at these sites and thereby account for the commonly proposed mechanisms of protein deformation.

4.1 Representing Protein Flexibility

Our protein modeling approach is to consider the protein as an extension of the elastic rod representation for the DNA. Specifically, we construct the protein from several 'stiff' elastic rod domains that are coupled by localized flexible domains. Consequently, the stiffness varies along the contour length of the elastic rod, but remains constant within individual domains. Each domain is assigned intrinsic curvature and stiffness to properly account for the geometry and stiffness of the protein and DNA. Figure 4.1 depicts how we represent both DNA and protein with a single elastic rod with larger diameter rod domains being stiffer than smaller diameter domains. Although the ratio between bending to torsional stiffness is likely different from domain to domain, we maintain a constant ratio of 2:3 through all domains due to uncertainty of the protein stiffness. The stiffnesses matrix ($\mathbf{B}_{protein,i}$) of domain i is scaled relative to the stiffnesses matrix of DNA ($\mathbf{B}_{DNA}$). We adjust the scaling factor (β_i) to explore the effects of a variable degree of flexibility; that is

$$\mathbf{B}_{protein,i} = \beta_i \mathbf{B}_{DNA}. \tag{4.1}$$

To make meaningful comparisons between concentrated flexibility at the apex of the 'V' and at the head domains, the contour length of these domains are both fixed at 10 $\AA$ (about a 3 bp length). In our model we assign the stiff (rigid) protein domains a stiffness 10 times greater than the DNA domain, $\beta = 10$. For the flexible domains, we explore order-of-magnitude changes in stiffness, $\beta = \{10, 1.0, 0.1, 0.01\}$. We do this by conducting simulations at each stiffness level and combination of flexibility in the 'V' or head domains.

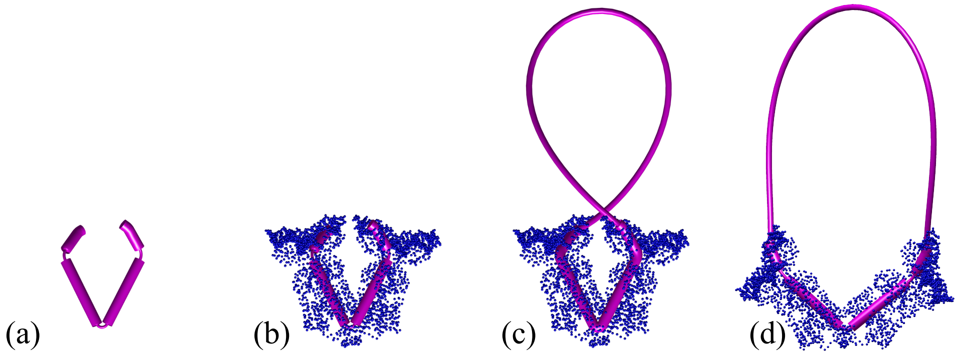


Figure 4.1 (a) Elastic rod representation of the protein for concentrated flexibility in the apex of the 'V' and in the head domains. (b) The same rod representation aligned with the protein structure. (c) Another rod representation for the protein (this time with high stiffness throughout) together with the rod representation for the DNA. (d) A flexible rod representation for the protein with rod representation for DNA. Here the all atom structure of the deformed protein has been approximated from the elastic rod representation.

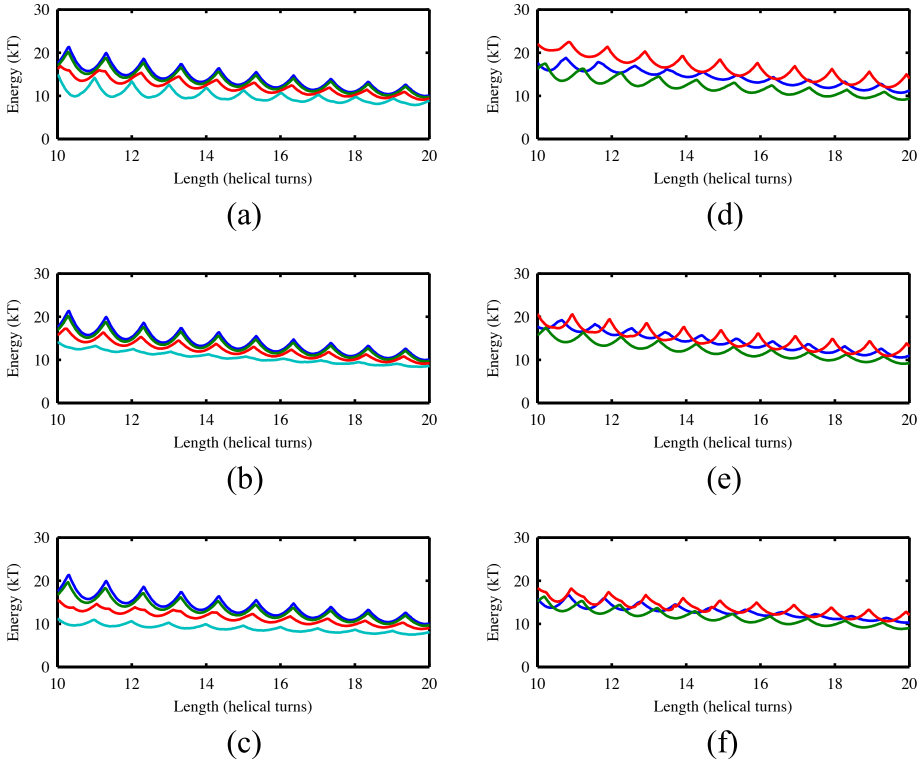
4.2 Results and Discussion

Here we present and discuss results for our calculations considering protein flexibility. As extensions to our results in Chapter 2 and Chapter 3, we consider the effects of protein flexibility on the looping of both intrinsically straight and intrinsically curved DNA.

4.2.1 Intrinsically Straight DNA

In Fig. 4.2 we present results for the elastic energy of looping as a function of operator separation for our flexible protein model. In Fig. 4.2(a)-(c) effects of four levels of stiffness concentrated at the protein 'V' (a), head domains (b), and both 'V' and head domains (c) are considered. The stiffness of these flexible domains relative to that of DNA is reduced sequentially: $\beta = 10$ (blue), $\beta = 1.0$ (green), $\beta = 0.1$ (red), and $\beta = 0.01$ (teal). The respective curves are associated with the minimum energy loop considering all binding topologies and topoisomers.

For flexibility concentrated in the head domains (Fig. 4.2(b)) or the combination of head and 'V' domains (Fig. 4.2(c)), the amplitude of the oscillations in energy (phasing effects) die out with decreasing stiffness. For a rigid protein, these oscillations arise from changes in the torsional alignment of the two operator sites with the addition of basepairs. In order to bind to a rigid Lac repressor, the operators must twist to align with the protein binding sites. Depending on the binding topology and the phasing of the operator sites, the required



(a) (d)

(b) (e)

(c) (f)

Figure 4.2 Here we plot the elastic energy (kT) as a function of length for a flexible protein. (a)-(c) Here we present the effects of four levels of stiffness concentrated at the protein 'V' (a), head domains (b) and both 'V' and head domains (c). Here the stiffness is measured relative to the stiffness used for the DNA and taking on the following values: $\beta = 10$ (blue), $\beta = 1.0$ (green), $\beta = 0.1$ (red), and $\beta = 0.01$ (teal). The minimum energy among all topoisomers and and binding topologies is plotted. (d)-(f) Here we present the effects of protein flexibility for the individual binding topologies. For these results the stiffness of flexible domains remains a constant 0.1 and is concentrated at the protein 'V' (d), head domains (e) and both 'V' and head domains (f). The energy is plotted considering the minimum of both topoisomers for all binding topologies: **P1** (blue), **A** (green), and **P2** (red).

twist changes. The addition of flexibility concentrated in the head domains allows the protein binding sites to adjust, in three dimensions, and thereby reduce the torsional stress sustained by the loop (and the boundary conditions of the DNA loop). Interestingly, however, phasing effects remain significant for concentrated flexibility at the 'V' (Fig. 4.2(a)), even for the most extreme flexibility considered ($\beta = 0.01$ teal curve). Experimental observations [41, 74] show phasing effects even for long lengths (about 300 bp) of inter-operator DNA and therefore suggests relatively high stiffness located in the head domains. Therefore, the protein is either relatively stiff throughout, or the flexibility is primarily localized in the 'V' domain.

Figure 4.2(a)-(c) also suggests the scale at which protein flexibility effects are significant. For example, there is little change from a rigid protein approximation to the addition of flexibility when $\beta = 10$ (blue) or $\beta = 1.0$ (green). Therefore in this regime a rigid approximation for the protein is appropriate. In contrast however, the energetics are more sensitive to changes in stiffness in the range of $\beta = 0.1$ (red). Within this regime, it is important to accurately represent the flexibility of the protein. Interestingly, the computational analysis presented in [120] arrives at a stiffness for the protein on the order of our $\beta = 0.1$. For protein stiffness much lower than $\beta = 0.1$, the flexible domains can be approximated as spherical joints and the energetics are relatively insensitive to changes in stiffness.

Another interesting observation from the results in Fig. 4.2(a)-(c) is that the local energetic minima and maxima shift with decreasing stiffness. For example, if we consider the energetic maxima for the case with flexibility concentrated in the 'V' domain (Fig. 4.2(a)), they become nearly a half helical turn out of phase when reducing the stiffness from the stiffest (blue) to the most flexible (teal). Although phasing effects significantly diminish with decreasing stiffness within the head domains, the energetic minima and maxima also shift; see Fig. 4.2(b).

Figure 4.2(d)-(f) illustrates the effect of protein flexibility on the energy of loops with different binding topologies. Here the stiffness of the flexible domains is held constant at $\beta = 0.1$ with concentrated flexibility at the protein 'V' (Fig. 4.2(d)), head domains (Fig. 4.2(e)) and both 'V' and head domains (Fig. 4.2(f)). As when the protein is considered rigid, the **A** topology (green) dominates as the energetically preferred binding topology. By not considering other binding topologies or modes of flexibility, most previous computational studies of Lac repressor looping suggest that concentrated flexibility in the 'V' domain leads to the **P1** binding topology being energetically dominant [50, 98, 107, 120]. In addition, Fig. 4.2(d)-(f) shows that the energetic differences between binding topologies are more pronounced when flexibility is introduced into the 'V' domain (Fig. 4.2(d)), and become less significant when introduced into the head domains (Fig. 4.2(e)-(f)). Therefore, our results suggest that all binding topologies should be considered when accounting for protein flexibility.

4.2.2 Intrinsically Curved DNA

Now we consider the effects of protein flexibility on the large family of intrinsically curved DNA molecules introduced in Chapter 3. Figure 4.3 presents the minimum energy for looping considering all binding topologies and topoisomers and the effects of protein flexibility. Here the stiffness of the flexible domains is reduced from $\beta = 10$ (Fig. 4.3(a)) by introducing sequentially more flexibility in the 'V' (Fig. 4.3(b)-(d)), head domains (Fig. 4.3(e)-(g)), and combined 'V' and head domains (Fig. 4.3(h)-(j)). Interestingly, the energy landscape remains qualitatively similar for all but the cases with the most flexibility (Fig. 4.3(d), (g), and (j)) where the landscape is relatively flat. In addition, the energetic minima (about 4 kT) remains nearly constant as β is reduced from 10 (Fig. 4.3(a)) to 0.1 (Fig. 4.3(c), (f), and (i)). This suggests that there is a baseline energetic cost of forming a loop with this family of sequences that is relatively insensitive to the stiffness of the protein. We could calculate this baseline energetic cost by calculating the energy in the limit of a protein with no stiffness in the flexible domains. The flatness of the energy landscape of Fig. 4.3(j) suggests that it is a good approximation of this limit.

In Fig. 4.3 the energetically preferred binding topology is indicated with shading. Similar to the energetic landscape, the preferred binding topologies remain qualitatively similar, for β greater than about 0.1. This suggests that many of the results compiled in Chapter 3 for intrinsically curved DNA (including binding topology, for example) remain qualitatively unchanged for β greater than about 0.1. Surprisingly, for low stiffness, the **P2** binding topology becomes preferred in some regions; see Fig. 4.3(d), (g), and (j). (We do not account for possible steric interferences between DNA and protein; see, for example, [4].) However, as already indicated for intrinsically straight DNA, the energetic differences between binding topologies is small at low protein stiffness.

Motivated by the ongoing experimental efforts of the Kahn lab at the University of Maryland, we predict the distances between FRET markers for the large family of intrinsically curved DNA. In addition to the D0A0 fluorophore configuration used in [29, 73], we estimate distances for the other three configurations: D1A0, D0A1, and D1A1. Figure 4.4 presents FRET distances ($\AA$) for all configurations (D0A0, D1A0, D0A1, and D1A1) with $\beta = 0.1$ concentrated in the 'V' domain, head domains, or both 'V' and head domains. The distances are presented in gray scale such that white and black correspond to short (high efficiency) and long (low efficiency) distances; and because of the expected errors in the experimentally measured distances, we discretize the distances into 4 levels.

The predictions of Fig. 4.4 will help us interpret the ongoing FRET experiments in the Kahn lab. We expect that there will be some differences between our predictions and the experimental observations due to model assumptions. In addition, the experimental measurements could have significant error. However, we have begun identifying features in the predictions that will likely be observable despite experimental error and model assumptions. One such feature is that the FRET distance landscapes for D1A0 and D0A1 are reflections of each other about the diagonal line $\Theta_1 = \Theta_2$. Although, this is a result of the symmetry of the DNA and protein assumed in our model, we expect that any asymmetry in the actual system is minor. In addition, because of these symmetries, the D0A0 and D1A1 FRET configurations are symmetric about the same line. Another feature present in Fig. 4.4 is the

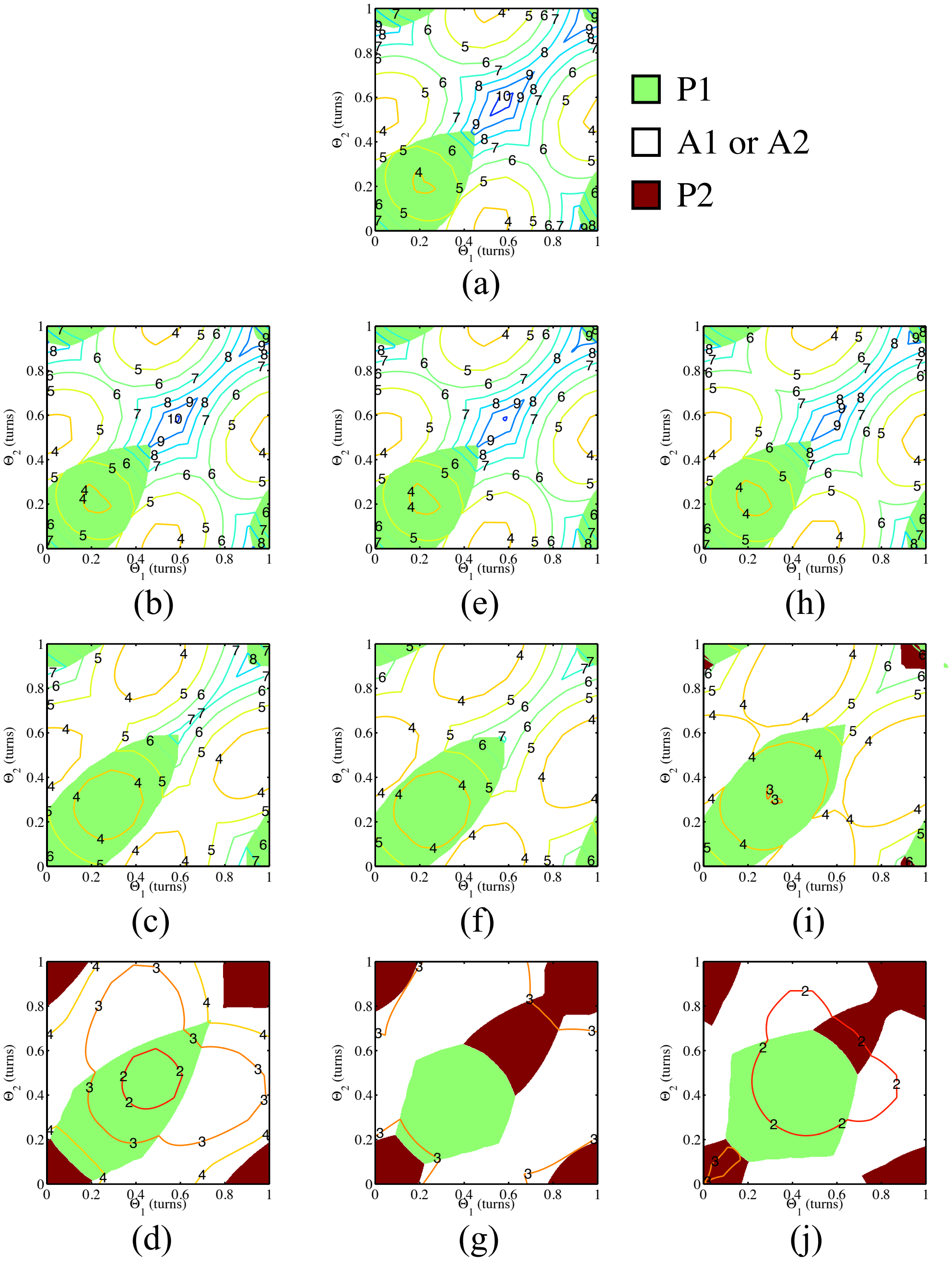


Figure 4.3 Elastic energy (kT) of the energetic minimum loop as a function of phasing parameters Θ_1 and Θ_2 considering protein flexibility. The energetically preferred binding topology is shaded as indicated. (a) Here the rod representation for the protein is 10 times stiffer than the DNA. (b)-(d) Here the apex of the protein 'V' is reduced in stiffness taking on values of 1.0, 0.1, 0.01 times the stiffness of the DNA rod. (e)-(g) Here the head domains of the protein are reduced in stiffness taking on values of 1.0, 0.1, 0.01 times the stiffness of the DNA rod. (h)-(j) Here both the apex of the protein 'V' and the head domains are reduced in stiffness taking on values of 1.0, 0.1, 0.01 times the stiffness of the DNA rod.

sharp differences in FRET distances across much of the diagonal line $\Theta_1 = \Theta_2$ for D1A0 and D0A1. Similar differences occur along a curve nearly perpendicular to this line.

4.3 Conclusions

In this chapter we have extended the elastic rod model for DNA looping used in Chapter 2 and Chapter 3 to account for flexibility of the Lac repressor protein. By exercising this model and considering intrinsically straight DNA we have shown that phasing effects remain significant when flexibility is introduced to the ‘V’ domain; however they diminish with increasing flexibility introduced in the head domains. Because of the significant phasing effects observed in experiments, the head domains are likely relatively stiff. Our results also suggest that loop energetics are most sensitive to changes in protein stiffness when $\beta = 0.1$. For much greater stiffness, the protein may be considered rigid; while for much lower stiffness, the protein allows free rotation at the sites of localized flexibility. For straight DNA our results also suggest that the optimal phasing adjusts with increasing flexibility. For example, the optimal minimum energy loop shifts almost a half helical turn when the stiffness is reduced from $\beta = 10$ to $\beta = 0.01$. Finally, our results show that the **A** binding topology remains energetically dominant suggesting that more than just the **P1** topology must be considered when accounting for protein flexibility.

We further exercised this model for protein flexibility to consider looping of the large family of intrinsically curved DNA constructs introduced in Chapter 3. By doing so we have shown that both the energy and preferred binding topology landscapes remain qualitatively similar for $\beta \geq 0.1$. We also find that for $\beta \geq 0.1$ the energetic minima over the entire family remains about 4 kT. Finally, we predict the distances between FRET markers corresponding to experiments underway in the Kahn lab at the University of Maryland. These predictions suggest the experimental observations for the D0A0 and D1A1 FRET configurations should be symmetric, while the D1A0 and D0A1 configurations should be reflections of one another. Additionally, the predictions identify borders of in the design space that separate sharp differences in FRET distances that should be readily observable in experiments.

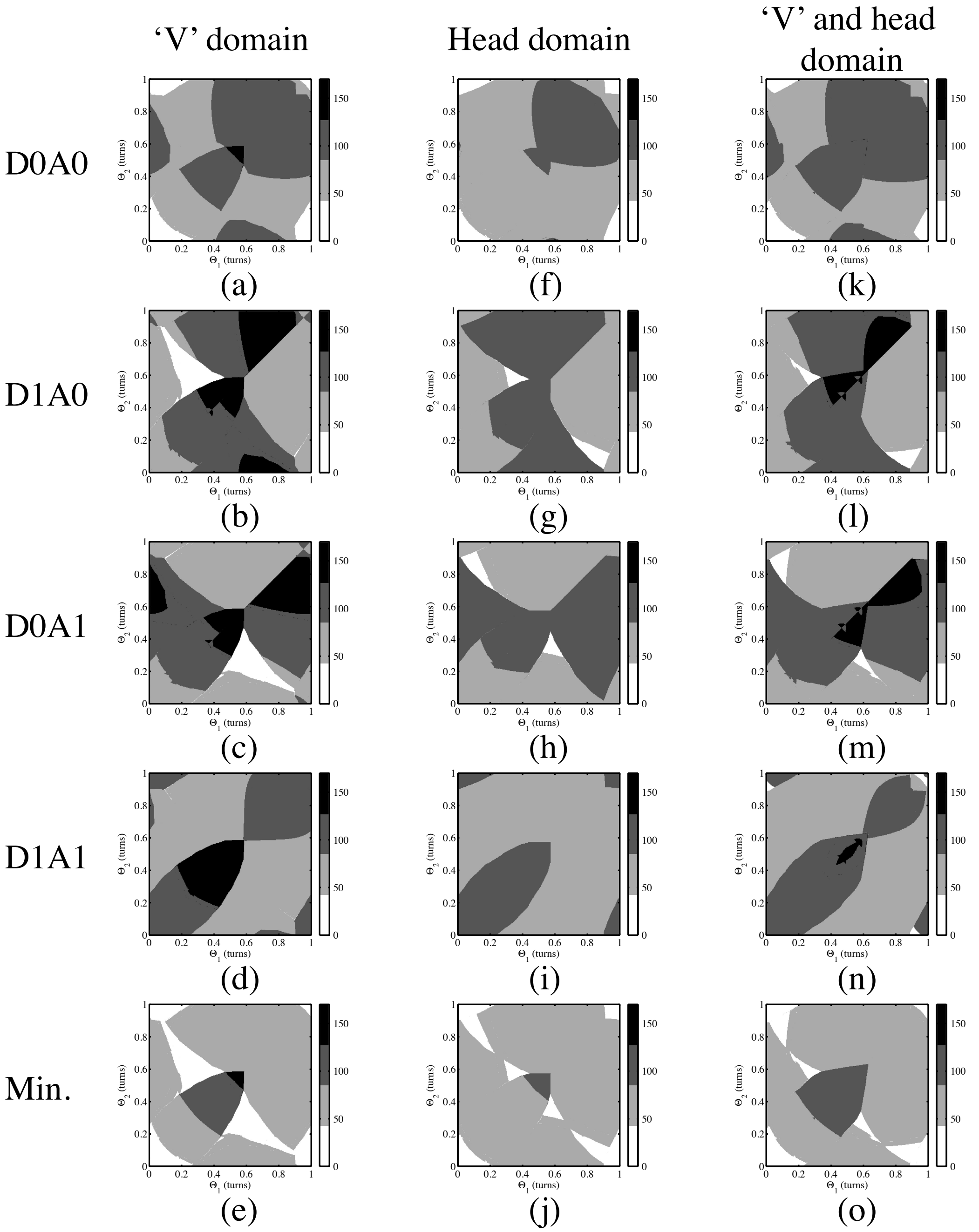


Figure 4.4 Estimated FRET marker distance ($\mathring{A}$) of the energetic minimum loop as a function of phasing parameters Θ_1 and Θ_2 considering protein flexibility. The distance is presented in grayscale with black and white corresponding to low and high FRET efficiency respectively. Because the corresponding FRET experiments are expected to have significant error we discretize the distances into 4 bins. Here the stiffness of the flexible domains is held constant at 0.1 times that of DNA and is localized in the 'V' domain (a)-(e), head domains (f)-(j), and 'V' and head domains (k)-(o). The rows correspond to FRET marker configurations, with the last row representing the minimum distance (maximum efficiency) from the above rows.

Chapter 5

Analysis Metrics for the Geometry of DNA Minicircles Observed by Cryo-EM

Recently, the Stasiak group at the Université de Lausanne has implemented cryo-EM to study minicircles of 100, 106, and 108 *bp*. Because of their different lengths, each of the minicircles experiences different bending and torsional stresses. These three minicircles were previously the subject of a transcription assay aimed at measuring the effect of superhelical stress on transcription [64]. Bal31 digestion assays revealed relative levels of twist within the three minicircles (100 $bp < 108$ $bp < 106$ *bp*). Energetically, we expect the minicircles to form with linking number (ΔLk) in the range of -0.5 to 0.5. Using the twist angles of [49], the predicted ΔLk in the individual molecules are -0.29, 0.15, and 0.04 for the 100, 106, and 108 *bp* minicircles, respectively. The objective of the follow-up cryo-EM study is to further characterize the differences in superhelical stress with a particular emphasis on the detection of possible kinks. Additionally, the study attempts to quantify the partitioning of ΔLk between twist (Tw) and writhe (Wr) for the minicircles. To date, cryo-EM was used to observe these three minicircles each under two buffer conditions. The 100 and 106 *bp* minicircles were observed with and without Mg^{+2}; while the 108 *bp* minicircle was observed with and without both Mg^{+2} and ethidium bromide. Table 5.1 introduces a nomenclature for the populations of minicircles and their conditions.

Here we leverage this cryo-EM data to study the mechanics of DNA on a sub-persistence length scale. In particular, our objective is to establish a theoretical framework to analyze these minicircles. To this end, we consider several metrics to describe the geometry of the minicircles. For the cryo-EM data, we compare the distributions of these metrics among the populations of minicircles. We employ a Brownian dynamics model to predict their expected distributions and the number of samples required to detect differences between distributions. We further leverage the Brownian dynamics model to gain insight into the mechanics of these minicircles by comparing the predicted distributions to their experimentally observed

Minicircle	Length (bp)	Conditions	n
MC100	100		14
MC100Mg	100	Mg^{+2}	10
MC106	106		10
MC106Mg	106	Mg^{+2}	10
MC108	108		11
MC108EthMg	108	EthBr+Mg^{+2}	11

Table 5.1 Six sample populations of minicircles were imaged using cryo-EM. Each of the three minicircles were imaged under two conditions. Both the 100 and 106 *bp* minicircles minicircles were imaged in both the absence and presence of Mg^{+2}. The 108 *bp* minicircle was imaged in the absence and presence a solution containing ethidium bromide and Mg^{+2}. The sample size (n) is also given for each population.

counterparts. Finally, we identify the metrics with the most promising ability to distinguish the minicircles.

5.1 Methods

Geometric Measures of DNA Minicircles

The Stasiak group at the Université de Lausanne has provided us with the cryo-EM reconstructions from their recent studies on several DNA minicircles. The reconstruction of each minicircle consists of a series of 150 points distributed around its helical axis. From the reconstructions we calculate the following metrics to analyze the geometry of the minicircles:

- **Bend Angles:** To gain insight into the bending stiffness of DNA we consider the the degree of bending in the minicircles. We introduce the 'bend angle' as a metric to quantify the bending of DNA and its distribution over a population of minicircles. To calculate bend angles, we represent the reconstruction of a minicircle with 21 segments of equal length (about 1 segment for every 5 *bp* or 1.65 *nm*). The number of segments (or equivalently, bend angles) should not be too few or too many. Specifically, dividing the minicircle into too few segments would tend to filter bend angle information from the cryo-EM reconstructions. Meanwhile using too many segments provides no additional information because of the limited image resolution and smoothing during the reconstruction process. (One factor influencing our choice of segment length, 1.65 *nm*, was that it should be larger than the approximate image resolution, ~ 1 *nm*.) Furthermore, as the discretization length is reduced, neighboring bend angles become increasingly correlated. (The bend angles are already correlated within these sub-persistence length minicircles.) Therefore, we believe our use of 21 bend angles per minicircle is a reasonable compromise. (An additional factor motivating our choice of 21 bend angles is the computational ease of the associated Brownian dynamics simulations discussed below.) The bend angles are simply the angles between neighboring segments; and by definition, they are positive. These bending angles are similar to the curvature distributions calculated in [1]. For each minicircle, we consider 5 possible 'registrations' of the 21 vertices distributed around the minicircle. From one registration to the next, the location of the vertices are shifted by $\frac{1}{5}^{th}$ of a segment length. By doing so we can capture sharp bends that might otherwise be filtered out. As a result, we compute $21 \times 5 = 105$ bend angles per minicircle.

- **Maximum Bend Angles:** To assist in the identification of possible kinks, we consider the maximum bend angle observed in an individual minicircle. The maximum bend angle is simply the maximum of the collection of bend angles, described above, for the given reconstruction. We expect a kink, if present, to manifest itself as an exceptionally large bend angle.
- **In-plane Bend Angles:** As another measure of the bending stiffness, we calculate in-plane bend angles. While the bend angles discussed above account for the two axis bending of the molecule, the in-plane bend angles measure bending about a single (out-of-plane) axis. In contrast to the bend angles described above, the in-plane bend angles can take on both positive and negative values. Consequently, we expect the in-plane bend angles to take on a more Gaussian-like distribution than the strictly positive bend angles described above.

  As a first step in this analysis, we must define the 'plane' of a reconstructed minicircle. To this end, we interpolate evenly spaced points around the contour length of the reconstruction. We use the same number of points as basepairs in this interpolation. Next, we assign a unit mass to each point and calculate the inertia matrix for the minicircle. Then we reorient the minicircle in space to align the principal axes of inertia with the X-Y-Z directions. The maximum principal moment of inertia is aligned with the Z-axis, and the minimum and intermediate principal moments of inertia are aligned with the X and Y axes. The X-Y plane is considered the plane of the reconstructed minicircle. Now, we interpolate 21 segments of equal length around the minicircle. These segments are projected onto the X-Y plane from which the in-plane bend angles are measured between adjacent segments; see Fig. 5.1. They are positive when the principal normal of the curve points toward the interior of the minicircle and negative when it points away. Again, we consider 5 different registrations of the 21 segments.

- **Ellipticity:** As a measure of the overall geometry of a minicircle with a single scalar quantity, we calculate ellipticity. Ellipticity is the ratio of the major to minor diameter of the best fit ellipse for a minicircle. Ellipticity was introduced in [24] to quantify the geometry of minicircles with single strand gaps on opposite sides. To calculate ellipticity, we first compute the plane of the reconstructed minicircles as described above. The coordinates of the minicircle are projected onto the X-Y plane and the ellipticity is calculated from the best fit ellipse to this projection.
- **In-plane Modal Decomposition:** As a more detailed measure of the overall geometry of a minicircle, we further decompose the minicircles into in-plane normal modes and compute the corresponding modal energies. One of the major differences in geometry of minicircles with a given sequence arise from thermal fluctuations. Random thermal fluctuations should deform each minicircle from an equilibrium configuration into a unique deformed configuration. We can describe such an arbitrary perturbation from equilibrium with a basis of thermal (normal) modes. A modal stiffness is associated with each mode; and because of equipartition, we expect the modal energies for a population of constructs to follow a Boltzmann distribution. (That is each mode is treated as a single degree of freedom with an associated stiffness. Therefore, from equipartition, each mode has a defined distribution (Boltzmann) of energies.) We predict that minicircles with a small magnitude ΔLk form energetically more favorably. Therefore we expect $-0.5 \leq \Delta Lk \leq 0.5$; although we are uncertain about the exact value of ΔLk for the minicircles. For simplicity, we use the in-plane normal modes associated with a $\Delta Lk = 0.0$ minicircle for the decomposition of all the cryo-EM reconstructions regardless of ΔLk. (The $\Delta Lk = 0.0$ minicircle has the unique property that in-plane and out-of-plane deformations are decoupled in the normal modes.) To calculate modal energies, we decompose the in-plane bend angles of an individual minicircle into in-plane normal modes. (We use 21 bend angles per minicircle and consider only one registration, as the modal energies are relatively insensitive to registration.) The modal amplitudes are then converted into modal energies using the modal stiffnesses. For a derivation of normal modes and the method of decomposing bend angles into normal modes, see Appendix D. By using the in-plane normal modes of the $\Delta Lk = 0.0$ minicircle to decompose minicircles with non-zero ΔLk, there is no guarantee that the distribution of energy within a mode will follow the Boltzmann distribution. (However, we will show using Brownian dynamics that the

for $-0.5 \leq \Delta Lk \leq 0.5$ the modal energies are well described by the Boltzmann distribution.) The calculated modal energies are sensitive to the assumptions we made to calculate the normal modes and stiffnesses (including the bending persistence length). However, because the assumptions are consistent from construct to construct they provide a useful measure of geometry for comparisons.

- **Writhe:** Although the cryo-EM images provide an enlightening description of the molecule's helical axis, they do not provide direct information about how the molecule is torsionally deformed. Specifically, the torsional stress as a function of position along the contour cannot be determined from the cryo-EM images. However, we can compute writhe (Wr) for the minicircles, which provides some insight into the overall twist (Tw) of the molecules. For a closed minicircle, linking number (Lk) is fixed and the sum $Tw + Wr = Lk$. To calculate Wr, we exercise an algorithm described in [52] on the coordinate data interpolated at the basepair level.

Statistical Analysis

The existing cryo-EM data is limited and consists of only about 10 samples of each minicircle construct. Because of this small sample size, we must be careful and draw conclusions from statistically significant observations. The metrics discussed above, when applied to a sample population of minicircles, will result in a distribution of values. At the most basic level we can describe the sample distributions using the sample mean and standard deviation. These quantities are especially useful in describing Gaussian distributions. Unfortunately, we don't expect the distributions associated with many of these metrics to be Gaussian. (For example, bend angles are defined as positive quantities. Therefore their distribution will be skewed.) Therefore, we employ the two-sample Kolmogorov-Smirnov (K-S) test to measure statistical differences between sample distributions for the different constructs. The two-sample K-S test is based on the cumulative distribution functions of two sample populations. In this test, the null hypothesis asserts that the two sample distributions are taken from one population distribution. The null hypothesis is rejected when the maximum vertical distance between cumulative sample distribution functions ($d_{1,2}$) is larger than a critical value; that is when

$$d_{1,2} > K_\alpha \sqrt{\frac{n_1 + n_2}{n_1 n_2}}. \tag{5.1}$$

Here, K_α is a constant that reflects the desired confidence level ($K_\alpha = 1.36$ at 95% confidence); and the integers n_1 and n_2 are the number of samples comprising the two sample distributions. Therefore, the two-sample K-S test provides a means to compare the sample

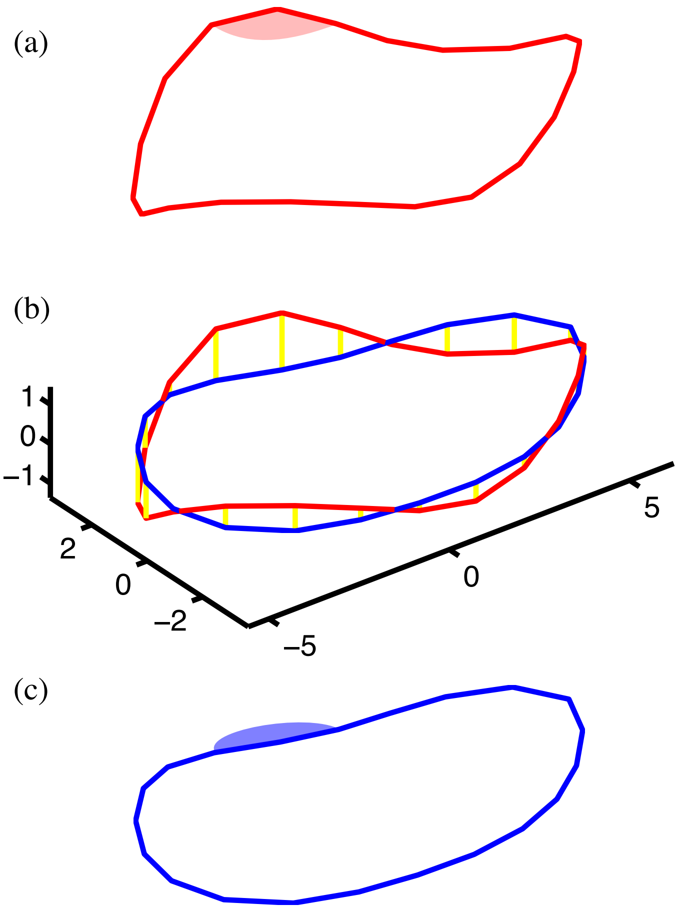


Figure 5.1 (a) Three dimensional reconstruction of a DNA minicircle represented by 21 segments. A single bend angle is highlighted. (b) Projection of the minicircle in (a) onto a plane. (c) Two dimensional projection of the minicircle given in (a). A single in-plane bend angle is highlighted.

distributions for distinct constructs and identify significant differences to within a prescribed confidence level.

Interestingly, if we estimate the distance d expected between actual population distributions we can estimate the number of samples necessary to detect differences between distributions. By inverting (5.1) and taking $n_1 = n_2 = n$, we can estimate the number of samples required from both populations as,

$$n = 2\left(\frac{K_\alpha}{d}\right)^2. \tag{5.2}$$

We will employ a Brownian dynamics model to estimate d and thereby predict the number of experimental samples required to distinguish two distributions with a given confidence.

Ideally, we could attain a sufficiently large sample size such that the sample distribution would accurately represent the actual distribution. Realistically however, the time-intensive experiments limit the number of cryo-EM reconstructions to a modest number ($n \lesssim 20$). Therefore, here we predict if the number of samples required to represent a population of minicircles is experimentally attainable. If we prescribe acceptable confidence limits on the experimentally observed cumulative distribution functions ($\pm d$), we can invoke the one-sided K-S test to calculate the number of samples required to describe the actual populations. The one-sided K-S test is similar to the two-sided test; but it tests the hypothesis that a sample distribution comes from a known distribution. The number of samples required to describe the actual distribution within a given tolerance is

$$n = \left(\frac{K_\alpha}{d}\right)^2. \tag{5.3}$$

Here, n is the number of samples forming the sample distribution. If, for example, we are interested in describing the complete distribution we might take $d = 0.05$ with 95% confidence ($K_\alpha = 1.36$) and find $n = 740$. Even if we relax this tolerance to $d = 0.10$, the required sample size is only reduced to $n = 185$. Clearly, an accurate description of the distribution of an entire population of minicircles remains outside a feasable experimental effort. However, as experimental methods continue to improve, this sample size may become attainable.

Brownian Dynamics Simulations

We exercised Brownian Dynamics (BD) simulations to theoretically predict thermal equilibrium distributions of geometrical metrics. These simulations were conducted using a formulation based on the algorithm presented in [53]. Using BD we represent minicircles of 100, 106, and 108 *bp* using 21 rigid segments (each of 1.65 *nm* in length). We expect the effects of differences in the ΔLk to have a pronounced effect on the distributions. In contrast, small differences in length between minicircles should have only a minor effect on the distributions. If, for example, we take the helical repeat of DNA to be 10.5 *bp* and consider the differences between the 106 and 108 *bp* minicircles the difference in ΔLk is significant, $\Delta Lk = (108-106)/10.5 = 20\%$, while the difference in length is negligible, $(108-106)/106 = 2\%$. Our simulations use 21 segments approximating the length of about 105 *bp*, which we use to represent the 100, 106 and 108 *bp* lengths. We simulate the range of $0.0 \leq \Delta Lk \leq 0.5$, and recognize the symmetries between the positive and negative states to predict thermal equilibrium distributions for $-0.5 \leq \Delta Lk \leq 0.0$.

Harmonic potentials are used to describe the extension of individual segments and bending/torsion between adjacent segments. Following the variables defined in [53], we choose the following parameter values describing extension, bending and torsional stiffnesses respectively: $\delta = 0.08$, $\alpha_b = 15.07$ (approximating a 50 nm bending persistence length), and $C = 3.1 \times 10^{-28}\ Nm^2$ (approximating a 75 nm torsional persistence length). (The product $2\alpha_b kT$ can also be thought of as bending stiffness at individual vertices.) Additionally, because the minicircles remain nearly circular for $|\Delta Lk| \lesssim 1.0$, there is no self contact, and we assume electrostatic repulsion forces are already considered in the apparent bending stiffness; see, for example [110].

Because we are interested in describing the thermal equilibrium distributions and not the dynamics, we need only represent the elasticity of the molecules with accuracy; see, for example [53]. Therefore we chose convenient approximations for parameters governing the hydrodynamics and neglect hydrodynamic interactions. In addition, we use a bead radius of 1.28 nm for translational hydrodynamic drag and an effective cylinder radius of 1.2 nm for the rotational drag about the helical axis.

The BD timestep size (5.0 ps) was chosen such that a BD simulation of a linear DNA chain accurately produce the expected mean squared fluctuations for stretch, torsion and cosine of the bend angles for BD segments (see, [53]). Table 5.2 shows that the timestep size results in accurate distributions of stretch, torsion and cosine of the bend angles.

Quantity	Expected	Computed	%Error
Stretch $\frac{\langle s - \langle s \rangle \rangle}{l_o^2}$	0.0064	0.0063	1.56%
Torsion $\langle \tau^2 \rangle$	0.0215	0.0215	0.00%
Bending $\langle cos(\beta) \rangle$	0.968	0.967	0.10%

Table 5.2 Expected [53] and computed BD stretch, torsion, and bending averages.

To verify that the BD simulations ran long enough to represent the thermal ensemble, we calculate the autocorrelation function for the parameters of interest. In Fig. 5.2 we plot the natural log of the auto correlation of Wr, maximum bend angle and ellipticity for $\Delta Lk = 0.0$. The inverse of the slope of these curves is a measure of the decay rate for a given metric. Using our choice of model parameters we measure the decay rate as measured from Wr, maximum bend angle, and ellipticity to be about 2.25 ns, 0.1 ns and 5.0 ns respectively. Thus, for our purposes, it takes about 10 ns (2×5 ns) for the minicircle to forget a perturbation. Or, every 10 ns the minicircle takes on an uncorrelated new state. Therefore, our BD simulations, running 2 million time-steps or 10 μs, build distributions from about 1000 independent measures.

Our objective in running BD simulations is to predict thermal distributions of geometri-

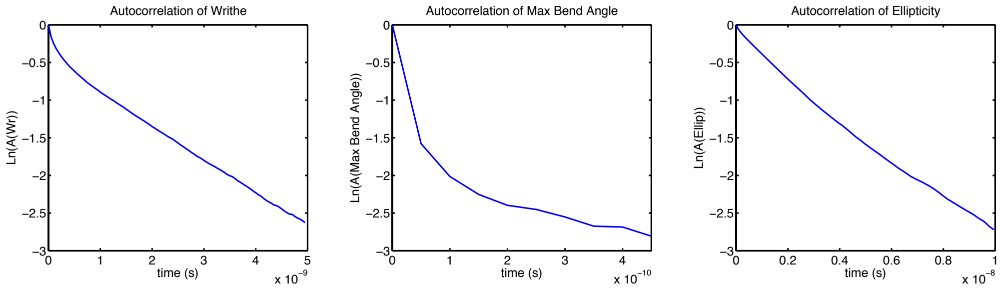


Figure 5.2 Natural log of the autocorrelation functions for Wr, max bend angle, and ellipticity for $\Delta Lk = 0.0$. These autocorrelations were calculated from simulations run using a Brownian dynamics algorithm based on [53].

cal metrics that we can compare with corresponding quantities measured from the cryo-EM images. Therefore we consider the same metrics listed above for describing the cryo-EM reconstructions. Whereas the cryo-EM reconstruction results in 150 points describing the helical axis, the BD simulations result in only 21 vertices. Therefore, the procedures described above for calculating the geometrical metrics need only minor adjustments for use on the BD simulations. For example, the inertia matrix as calculated from basepair resolution of the helical axis is calculated from only 21 vertices for the BD simulations.

5.2 Results and Discussion

Here we present and discuss results for metrics describing the geometry of DNA minicircles as observed by cryo-EM and as predicted by theory.

Bend Angles

In Table 5.3 we report the averages and standard deviations of the bend angles for the cryo-EM reconstructions. Because the bend angles are by definition positive, we do not expect them to follow a Gaussian distribution. For the BD simulations with ΔLk ranging from -0.5 to 0.5, the average and standard deviation of the bend angle is nearly constant with values of 0.35 and 0.15 (radians) respectively; see Table 5.3. Interestingly, we can estimate this average bend angle from the vector sum of estimates for the average in-plane and out-of-plane bend angles. Because the minicircle must form a circle, the in-plane average bend angle ($\bar{\theta}_{in}$) must be

$$\bar{\theta}_{in} = \frac{2\pi}{21} = 0.299. \tag{5.4}$$

Using equipartition, we can easily estimate the average out-of-plane bend angle ($\bar{\theta}_{out}$) provided we assume that each bend is independent (defining the out of plane bend angle, θ_{out}, as a positive quantity). First we equate the average bending energy at vertex to the the average energy expected by equipartition,

$$\alpha_B kT \bar{\theta}_{out}^2 = \frac{1}{2}kT. \tag{5.5}$$

Subsequently, if we isolate the average out of plane bend angle we have,

$$\bar{\theta}_{out} = \left(\frac{1}{2\alpha_B}\right)^{1/2} = 0.182. \tag{5.6}$$

The vector sum of these two components yields an estimated average bend angle of 0.35 in line with the average computed from BD simulations; see Table 5.3.

Minicircle	**n**	**Avg.**	**Std.**
MC100	14	0.373	0.159
MC100Mg	10	0.376	0.181
MC106	10	0.354	0.150
MC106Mg	10	0.379	0.177
MC108	11	0.333	0.110
MC108EthMg	11	0.355	0.145
BD Sim. ($\Delta Lk = 0.0$)		0.35	0.15
BD Sim. ($\Delta Lk = 0.5$)		0.35	0.15

Table 5.3 Mean and standard deviation of bend angles (radians).

With a limited sample size for the cryo-EM reconstructions, it is important to be careful with the two-sample K-S test. In particular, the bend angle distributions that we calculate are a compilation of all the distributions of bend angles for individual minicircles. The distribution of bend angles of an individual minicircle must conspire to form a circle. Therefore, the closure of the minicircle constrains the bend angles. (The exact constraint is complicated to formulate because of the three-dimensional geometry of the minicircles.) Additionally, with a sub-persistence length of DNA ($\sim 2/3\ L_p$) we expect the bend angles to be highly correlated within a minicircle. To be conservative, we consider the collection of 105 bend angles (21 bend angles times 5 registrations) for a reconstruction of an individual minicircle to be one independent measure. Specifically, if we have 10 reconstructions of a DNA construct and calculate 105 bend angles, then we effectively only have 10 independent measures of bend angle rather than 1050. Therefore, n_1 and n_2 in (5.1) are taken as the total number of minicircles and not the total number of bend angles. Following this assumption, the two-sample K-S test is unable to distinguish between distributions. Specifically, the null hypothesis, that all pairs of constructs have the same bend angle distributions, is not rejected. Similarly, the BD simulations show the maximum vertical distance between cumulative distribution functions is about $d = 0.01$. This suggests that an enormous number of samples (nearly 37000) would be required to experimentally resolve differences between bend angle distributions (using the two-sample K-S test) due to differences in ΔLk. However with a less conservative approximation of the effective number of independent measures of bend angles per minicircle, the number of required samples may be reduced. This observation may also be sensitive to the number of segments into which a minicircle is divided. This dependence on discretization length further suggests that bend angle distributions are a poor metric to distinguish minicircles.

Following both BD data and existing cryo-EM data, it appears that comparing bend angle distributions is not an effective metric to distinguish minicircles. Interestingly, we also expect a similar approach employing curvature distributions [1] to suffer the same challenges, because bend angles and curvatures are so closely related.

Maximum Bend Angles

Table 5.4 presents the mean and standard deviation for the maximum bend angles. It's important to note that we don't expect this distribution to follow a Gaussian distribution. Interestingly, the statistics for MC108 are noticeably different compared to the other constructs. Like the bend angles above, the BD simulations show that the average and standard deviation of the maximum bend angles are nearly constant considering the range of ΔLk between -0.5 and +0.5 with values of 0.65 and 0.08 respectively; see Table 5.4. Except for the MC108 minicircles, this average and standard deviation is lower than the mean and standard deviations for the cryo-EM reconstructions.

Minicircle	**n**	**Avg.**	**Std.**
MC100	14	0.763	0.148
MC100Mg	10	0.814	0.170
MC106	10	0.718	0.176
MC106Mg	10	0.799	0.264
MC108	11	0.556	0.103
MC108EthMg	11	0.718	0.110
BD Sim. ($\Delta Lk = 0.0$)		0.65	0.08
BD Sim. ($\Delta Lk = 0.5$)		0.66	0.08

Table 5.4 Mean and standard deviation of maximum bend angles (radians).

The smaller average and standard deviation of MC108 suggests that it is more circular and stiffer than the others. (For a population of perfectly circular constructs, the average maximum bend angle and average bend angle would both be 0.299; see (5.4). A larger average maximum bend angle is therefore a measure of the departure from circularity.) This suggests that either the MC108 minicircle experiences an increased relative stiffness (to make it more circular), or the other minicircles experience a reduction in relative stiffness. If we assume that the MC108 minicircle is the most torsionally relaxed ($\Delta Lk \sim 0.0$), the superhelical stress in the others could be the cause of their reduced stiffness. The smaller average for MC108 compared to the BD simulations further suggests that the bending persistence length is larger than the BD model assumes. However, this could likely be an artifact of a small sample size or the reconstruction procedure which gives preference to a smooth helical axis.

One reason for considering maximum bend angles is to identify possible kinks. If kinking eliminates the bending stiffness at a point, we would expect the observed average maximum bend angles to be much larger than observed in Table 5.4. Therefore, if kinks occur in these minicircles, they likely have some non-zero bending stiffness. It must be noted, however, that the reconstruction procedure would tend to smooth out any potential kinks. Although it is uncertain to what degree it does so. In addition, identifying kinks by using the maximum bend angle could also be sensitive to the discretization length scale. Longer discretization lengths could filter out sharp kinks.

The BD simulations predict the maximum vertical distance between cumulative distribution functions to be at about $d = 0.05$ considering the range of $-0.5 \leq \Delta Lk \leq 0.5$. Therefore, the BD simulations imply that many samples (about 750) would be required to experimentally resolve these differences using the two-sample K-S test. Interestingly, however, the two-sample K-S test detects a significant difference between the distributions of maximum bend angles for the MC108 minicircles and all but the MC108EthMg minicircles. Specifically, the test rejects the null hypothesis with 95% confidence, that the MC108 minicircles have the same distribution of maximum bend angles as any one of the other minicircles (excluding MC108EthMg). (At a lower confidence level, 90%, the null hypothesis is rejected between the MC108 and the MC108EthMg minicircles.) Figure 5.3 further illustrates the uniqueness of the MC108 minicircles by presenting the cumulative distribution functions for their maximum bend angles. As a result, the maximum bend angle has potential as a metric to distinguish the geometry of these minicircles.

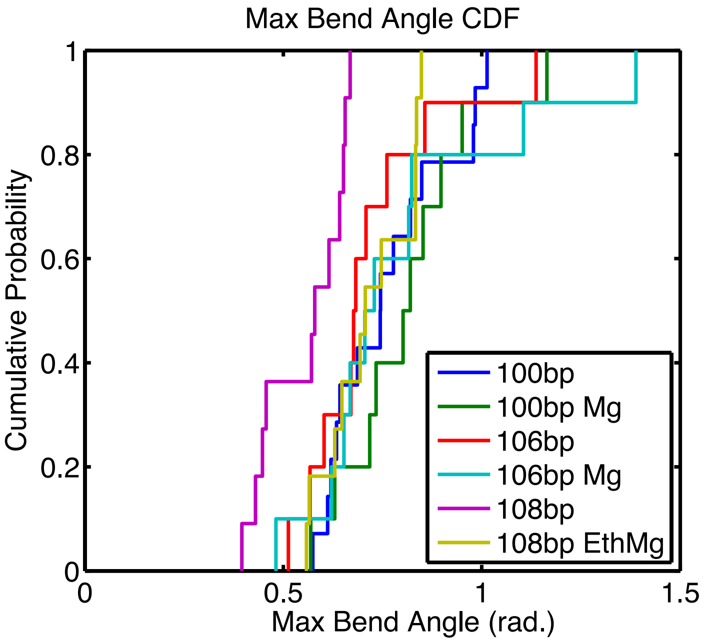


Figure 5.3 Cumulative distribution functions for the maximum bend angles (radians) observed in the cryo-EM reconstructions.

In-plane Bend Angles

Table 5.5 presents the mean and standard deviation of the in-plane bend angles and corresponding values obtained by BD. Note the average in-plane bend angle is identical for every construct. This is a requirement of the constraint that the in-plane bend angles form a closed planar polygon; see Appendix D for the constraint equations. By employing the following approximation for this constraint we can estimate the bending stiffness of the constructs. Specifically, our approximation is to take the actual average in-plane bend angle, 0.299 (see (5.4)), and approximate thermal fluctuations about this average to be uncorrelated between vertices. Following this approximation, the in-plane bend angles form a Gaussian distribution. The standard deviation of the in-plane bend angle is therefore a measure of the molecule's bending stiffness. To estimate the bending stiffness, we begin by writing the change in the average bending energy (from its equilibrium value, $\bar{\theta}_{in} = 0.299$) of each vertex and using equipartition

$$\left\langle \alpha_B kT \left(\theta_{in} - \bar{\theta}_{in}\right)^2 \right\rangle = \frac{1}{2} kT. \tag{5.7}$$

Solving this expression for $2\alpha_B$, the bending stiffness at each vertex yields

$$2\alpha_B = \frac{1}{\left\langle \left(\theta_{in} - \bar{\theta}_{in}\right)^2 \right\rangle}, \tag{5.8}$$

where $\left\langle \left(\theta_{in} - \bar{\theta}_{in}\right)^2 \right\rangle$ is the variance of the in-plane bend angles. The persistence length is related to bending stiffness ($2\alpha_B$) following Equation (3) of [53]; and Table 5.5 also presents the estimated persistence lengths. To gain some understanding of the accuracy of this method we apply it to estimate the bending persistence length for the BD simulations. The $\Delta Lk = 0.0$ and $\Delta Lk = 0.5$ minicircles correspond to a persistence length of 57 *nm* and 54 *nm* respectively. The in-plane bend angles support the assertion above that the MC108 minicircles appear to be significantly stiffer than the other minicircles. However, its calculated persistence length, 139 *nm* (with 95% confidence limits of about 45 *nm* and 285 *nm*), seems unrealistically high. This large persistence length is likely an artifact of the reconstruction procedure that tends to smooth the helical axis.

Minicircle	**n**	**Avg.**	**Std.**	**Est.** L_p
MC100	14	0.299	0.162	63.1
MC100Mg	10	0.299	0.186	47.9
MC106	10	0.299	0.160	64.8
MC106Mg	10	0.299	0.155	68.9
MC108	11	0.299	0.109	139.0
MC108EthMg	11	0.299	0.142	81.1
BD Sim. ($\Delta Lk = 0.0$)		0.30	0.17	57
BD Sim. ($\Delta Lk = 0.5$)		0.30	0.18	51

Table 5.5 Mean and standard deviation of in-plane bend angles (radians). Also, an estimate for persistence length (L_p) in *nm* is calculated from the standard deviation of the in-plane bend angles.

For a more rigorous approximation of in-plane bend angle distributions we employ BD simulations. These simulations calculate a nearly constant average and standard deviation of the in-plane bend angles, considering ΔLk between -0.5 and +0.5, with values of 0.30 and 0.17 respectively; see Table 5.5. Additionally, the BD simulations predict a value of about $d = 0.01$ for the maximum difference between cumulative distribution functions. This suggests a large number (nearly 37000) of samples would be required to distinguish minicircles. Accordingly, when applying the two-sample K-S test to the experimentally observed distributions of in-plane bend angles the null hypothesis is not rejected (at the 95% confidence level). That is there is insufficient data to suggest that the different constructs have different distributions of in-plane bend angles despite the apparent differences in standard deviations observed in Table 5.5. Therefore the in-plane bend angle distributions are considered a poor metric to distinguish minicircles and suffers from the same challenges as the (two axis) bend angle distributions.

Ellipticity

Table 5.6 presents the average and standard deviation of the ellipticity of the cryo-EM reconstructions. Because ellipticity is defined to be greater than or equal to one, the distribution of ellipticity is not expected to be Gaussian. As with the other metrics, the MC108 construct stands out with its smaller averages and standard deviations suggesting a more circular and perhaps stiffer configuration. The BD simulations result in a nearly constant average and standard deviation of ellipticity of about 1.08 and 0.05 respectively for the range of $-0.5 \leq \Delta Lk \leq 0.5$; see Table 5.6.

Minicircle	**n**	**Avg.**	**Std.**
MC100	14	1.194	0.090
MC100Mg	10	1.185	0.120
MC106	10	1.152	0.117
MC106Mg	10	1.174	0.103
MC108	11	1.104	0.064
MC108EthMg	11	1.134	0.075
BD Sim. ($\Delta Lk = 0.0$)		1.08	0.05
BD Sim. ($\Delta Lk = 0.5$)		1.10	0.05

Table 5.6 Mean and standard deviation of ellipticity.

The BD simulations also show differences in cumulative distribution functions of ellipticity are about $d = 0.1$, suggesting a large number of samples required (about 185) to distinguish distributions. When comparing the sample distributions between constructs, at the 95% confidence level the two-sample K-S test cannot reject the null hypothesis (that the ellipticity distributions are identical between all pairs of constructs). However, the K-S test does reject the null hypothesis at the 90% level when comparing the sample distributions for the MC100 and MC106 and for the MC100 and MC108 constructs. Therefore, ellipticity appears to have some potential for distinguishing minicircles. Furthermore, ellipticity benefits from reducing the geometry of a minicircle into single scalar value and its relative insensitivity to discretization length.

In-plane Modal Decomposition

Table 5.7 presents the calculated mean and standard deviation of in-plane modal energies for the cryo-EM reconstructions. We expect the distribution of modal energies to be well approximated by the Boltzmann distribution, for which the mean and standard deviations are equivalent. Furthermore, if the persistence length of these minicircles is 50 *nm* then the average and standard deviation of the modal energy should both be about 1 kT. If their persistence length were smaller (larger) the average and standard deviation of the modal energies would still be equal and would be larger (smaller) than 1 kT. The MC108 minicircles stand out with lower average and standard deviation of modal energies. Because the modal energy is a measure of the amplitude of the mode, it is also a measure of the departure from circularity. Therefore, the lower average modal energies suggests that the MC108 minicircles take on a more circular configuration.

Minicircle	**n**	**Mode 1**		**Mode 2**		**Mode 3**		**Mode 4**	
		Avg.	**Std.**	**Avg.**	**Std.**	**Avg.**	**Std.**	**Avg.**	**Std.**
MC100	14	4.04	2.94	2.06	2.30	1.15	1.47	0.56	0.47
MC100Mg	10	4.03	4.15	3.06	3.50	1.78	1.66	1.33	1.40
MC106	10	3.21	4.27	1.81	1.19	1.88	1.28	0.67	0.54
MC106Mg	10	3.32	2.65	1.01	1.07	2.11	2.28	0.58	0.22
MC108	11	1.46	1.44	1.23	1.83	0.69	1.00	0.18	0.16
MC108EthMg	11	2.27	1.85	1.92	1.84	1.37	1.84	0.41	0.34
BD Sim. ($\Delta Lk = 0.0$)		1.01	1.01	1.02	1.03	1.03	1.04	1.03	1.05
BD Sim. ($\Delta Lk = 0.5$)		1.26	1.23	1.12	1.16	1.08	1.07	1.05	1.06

Table 5.7 Calculated mean and standard deviation of in-plane modal energy (kT) for modes 1-4.

The BD simulations show that modal energy distributions are relatively insensitive to ΔLk of the minicircle. Figure 5.4 illustrates this insensitivity. Specifically, it presents four plots (one for each of the first four modes) and plots the average computed modal energy (kT) as a function of ΔLk considering four cases: (blue) nominal bending stiffness (50 *nm* persistence length) at all vertices, (red) bending stiffness reduced to 50% at all vertices (25 *nm* persistence length), (magenta) bending stiffness at one vertex is reduced to 25% (representing a kink), and (green) bending stiffness is reduced at all vertices to 50% and at one vertex to 25%. This figure shows that for the nominal bending stiffness the, modal energy is nearly constant, and therefore insensitive. Additionally, for the first and second mode d is approximately 0.09 and 0.04 respectively considering $-0.5 \leq \Delta Lk \leq 0.5$. This difference decreases with increasing mode number. This therefore suggests that many samples (more than about 200) would be required to distinguish minicircles. However, when comparing the modal energy distributions for pairs of constructs for the Cryo-EM data, we

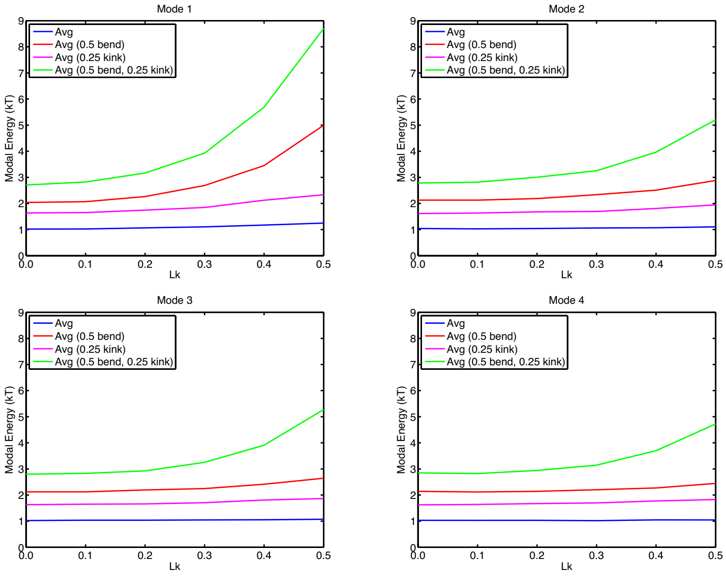


Figure 5.4 Predicted average modal energies for the first 4 modes considering a range of ΔLk. Additionally, we explore the effects of reducing the bending stiffness to 50% and introducing a possible kink by reducing the bending stiffness at one vertex to 25% of the bending stiffness of the other vertices. Specifically, the curves are as follows: (blue) nominal bending stiffness (50 *nm* persistence length) at all vertices, (red) bending stiffness reduced to 50% at all vertices (25 *nm* persistence length), (magenta) bending stiffness at one vertex is reduced to 25% (representing a kink), and (green) bending stiffness is reduced at all vertices to 50% and at one vertex to 25%.

reject the null hypothesis (that the distributions are identical) at 90% and 95% confidence for:

- Mode 1: MC108 and MC100 (95%), MC106 and MC100 (90%)
- Mode 2: MC108 and MC106 (95%)
- Mode 3: MC106 and MC100 (95%), MC108 and MC106 (95%)
- Mode 4: MC108 and MC100 (95%), MC108 and MC106 (95%), MC108 and MC106Mg (95%), MC108EthMg and MC108 (90%)

These significant differences observed in the cryo-EM reconstructions suggest that in-plane modal decomposition may be an effective metric to distinguish minicircles.

Consistent with equipartition, the BD simulations show that we should expect an average of 1 kT and standard deviation of 1 kT for every mode with $\Delta Lk = 0.0$; see Table 5.7. (The average modal energy is 1 kT rather than $1/2$ kT because each mode has two associated degrees of freedom, amplitude and phase.) For $\Delta Lk = 0.5$, the average and standard deviation of the modal energy both remain close to 1 kT. In contrast, the modal energy distributions calculated for the cryo-EM reconstructions do not match the expected distributions for several possible reasons. A primary source of this discrepancy is believed to be a consequence of the reconstruction procedure. During reconstruction, an optimization algorithm is used to approximate the helical axis of the minicircle. Because of the low resolution of the cryo-EM images, image processing techniques are supplemented with a bending energy penalty term in the optimization algorithm. The algorithm may therefore filter (short wavelength) sharp bends in the helical axis. Consequently, we expect this filter to reduce the apparent modal energies with increasing mode number (decreasing wavelength). Beyond some critical mode this filter will reduce the apparent modal energy to such an extent that the computed modal energy is meaningless. To approximate this critical mode we first consider the resolution of cryo-EM. Lets assume that cryo-EM is unable to resolve length-scales smaller than the diameter of DNA (2 *nm*). This suggests that there are about 16 independent measures of position around the circumference of a 100 *bp* minicircle ($100\ bp \times 0.33\ nm/bp/2\ nm \sim 16$). Further assuming that about 4 such points are needed to resolve a single modal wavelength suggests that modes with up to 4 full wavelengths can be resolved. The third in-plane mode here consists of 4 full wavelengths; see Fig. D.2 in Appendix D. Although, we present results through mode 4, we believe that modes higher than about mode 3 cannot be considered reliable. As repeated several times the MC108 minicircle appears to be the most circular, and therefore it is reassuring that the average energy decreases with increasing mode number as expected. In contrast, the average modal energy of the third mode of the MC106Mg minicircle is about twice that of its second mode.

Another possible reason that the modal energies don't appear to match the expected distributions is the possible existence of kinks. To understand the influence of kinks on the distribution of energy among modes, we simulated minicircles with a localized reduction of bending stiffness. Specifically, to represent a kink, we reduced the bending stiffness to 25% its original value at one vertex in the BD minicircles. The BD simulations show that if there is only one kink then the modal energy for each mode is almost uniformly increased over all modes; compare the blue and magenta curves of Fig. 5.4. In contrast, we anticipate that if more than one kink were present, the energy would likely not be evenly distributed among the modes. For instance, if there were two kinks on opposite sides of a minicircle, we anticipate that the first mode would be excited the most.

Yet another possible reason that the modal energy distributions are unexpected could be that the in-plane modes of the $\Delta Lk = 0.0$ case are a poor approximation for the modes of minicircles with non-zero ΔLk. To measure the changes in the modal energy distributions when applying these modal decompositions to minicircles with nonzero ΔLk, we employed BD simulations; see Fig. 5.4. As already mentioned, this figure shows that for the nominal bending stiffness the modal energy is nearly constant, and therefore insensitive, for the range of $0.0 \leq \Delta Lk \leq 0.5$. We also conducted BD simulations with an extreme reduction in bending stiffnes of is 50% of the nominal value (red curves of Fig. 5.4). Changing the bending stiffness also changes the ratio between bending and torsional stiffness. This ratio is important in determining the degree of coupling between in-plane and out-of-plane modes for non-zero ΔLk minicircles. As this ratio is reduced, the minicircle builds preference to bend out of plane rather than twist in response to thermal fluctuations. Even with this extreme reduction in stiffness, the modal energy is fairly insensitive to $\Delta Lk \approx 0$. However, for the first mode, the energy becomes sensitive beyond about $\Delta Lk = 0.25$; and this sensitivity decreases with increasing mode number. In fact, this sensitivity is almost eliminated in the second mode. Because of the low sensitivity to ΔLk, we believe that decomposing the minicircles into in-plane normal modes for $\Delta Lk = 0.0$ is a reasonable approximation.

We can also employ this in-plane modal decomposition to explore the persistence length of the minicircles. Because the MC108 appears to be the stiffest and most circular construct we focus on it. Specifically, we believe it is least likely to experience kinking or any other increased flexibility that would inhibit estimation of the persistence length. We further assume that it is relaxed, $\Delta Lk = 0.0$ and that the reconstruction procedure does not filter the lowest mode. Following these assumptions, the average modal energy (1.46 kT) for the 108 *bp* minicircle suggests that the persistence length is 1/1.46 times the persistence length associated with the modal stiffness (50 *nm*); and therefore the estimated persistence length is 34 *nm*. (The computed modal energy is proportional to the bending stiffness of the DNA relative to the bending stiffness used to compute the modal stiffnesses.) At the 95% confidence level the lower and upper bounds for the persistence length are 17 *nm* and 57 *nm* respectively. This is a sizable range and is a consequence of a small sample size.

As indicated, the in-plane modal decompositions of the experimental data appears to be an effective metric to distinguish minicircles. This suggests a full three-dimensional modal decomposition may also be an effective metric. Full three-dimensional normal modes would account for the coupling between in-plane and out-of-plane deformations experienced by minicircles with $\Delta Lk \neq 0.0$ and thereby better account for the differences between the constructs. However, one complication of this approach is that ΔLk for the individual minicircles is unknown.

Writhe

Table 5.8 presents the average and standard deviation of Wr for the constructs. Once again, the MC108 construct stands out with its smaller average and standard deviation, suggesting that it is more planar/circular and stiffer. The BD simulations predict that the average Wr varies from 0.00 for $\Delta Lk = 0.0$ to about 0.017 at $\Delta Lk = 0.5$; meanwhile, the standard deviation of Wr remains nearly constant at about 0.02; see Table 5.8. The averages in Table 5.8 appear consistent with this prediction, but the standard deviations are several times larger. Overall, the minicircles are all nearly planar, despite their differences in ΔLk.

Minicircle	n	Avg.	Std.
MC100	14	0.017	0.072
MC100Mg	10	0.031	0.074
MC106	10	0.018	0.063
MC106Mg	10	-0.003	0.092
MC108	11	0.009	0.028
MC108EthMg	11	-0.022	0.042
BD Sim. ($\Delta Lk = 0.0$)		0.00	0.019
BD Sim. ($\Delta Lk = 0.5$)		0.017	0.023

Table 5.8 Mean and standard deviation of writhe.

Considering $0.0 \leq \Delta Lk \leq 0.5$, the maximum distances between cumulative distribution functions is about $d = 0.3$ corresponding to a more feasible number of samples $n \sim 40$. For the experimental data, at the 95% confidence level the two-sample K-S test does not reject the null hypothesis (that all pairs of Wr distributions are identical). For a lower confidence level, 90%, the two-sample K-S test shows differences between sample distributions for MC108 and MC108EthMg and between MC100Mg and MC108EthMg. Overall, the cryo-EM and BD simulations suggest that Wr could be a good metric to distinguish minicircles.

It is surprising that the average Wr of the MC108 minicircles is larger than that of the MC108EthMg minicircles; and that the two-sample K-S test showed that the distributions are significantly different. We expect ethidium bromide to intercalate between basepairs and thereby reduce the Tw of the minicircle. (Also, ethidium bromide is expected to stiffen the molecule.) Because the Lk of the minicircle remains fixed and $Lk = Tw + Wr$, we would expect the Wr of the MC108EthMg to be higher than the MC108 minicircles.

5.3 Conclusions

In this study we have presented a theoretical framework to analyze cryo-EM reconstructions of DNA minicircles. In particular, we have considered several possible metrics that could be used to quantify the geometry of individual DNA minicircles. We then used these metrics to study the sample populations of minicircles observed in cryo-EM. We further employed a Brownian dynamics algorithm to predict the distributions for these metrics and to estimate the sample size necessary to detect differences between them. From this exercise we have identified those metrics having the greatest potential for distinguishing the geometrical properties of the minicircles.

Experimentally, we observe the statistically significant differences at the 95% confidence level among distributions of maximum bend angles and modal energies. At a lower confidence level, 90%, ellipticity and Wr also arise as statistically significant discriminators of minicircle geometry. In contrast, the bend angle and in-plane bend angles appear to be poor discriminators. BD simulations lead to the conclusion that Wr may discriminate the species of minicircles with a reasonable number (about 40) of samples whereas all other metrics would require far greater sample sizes.

All the metrics proposed herein suggest that the MC108 minicircle is the most similar to the $\Delta Lk = 0.0$ minicircle of our BD model. Based on in-plane modal decomposition of this minicircle, we estimate the bending persistence length of the tightly bent DNA minicircles to be 34 *nm* (with 95% confidence limits of about 17 *nm* and 57 *nm*).

Chapter 6

Electrostatics and Self Contact in an Elastic Rod Approximation for DNA[1]

Throughout the previous chapters we have emphasized, using the example of Lac repressor looping, that the mechanics and biological function of DNA are strongly coupled. Here we consider DNA supercoiling as another example. Many essential cellular processes, including transcription and replication, require or impose twisting and/or bending strains upon the molecule (see Fig. 6.1). These strained states, including deformations that result from the DNA twisting or wrapping upon itself like an overly twisted telephone cord, are referred to as supercoils. A supercoiled topology is quantified by its linking number (*Lk*), and for a closed loop the *Lk* is a topological invariant; no matter how the DNA of the closed loop is deformed (but not broken), the *Lk* remains constant (see, for example, [32]). The *Lk* is the sum of two components, twist (Tw) and writhe (Wr), which depend on the stress state of the molecule. Tw is proportional to the torsional stress in the molecule while Wr is a measure of the three-dimensional bending that often results in intertwining of the helical axis about itself.

In the crucial process of transcription, the DNA code is read and a single stranded (mRNA) copy is produced. During transcription an enzyme (RNA polymerase) locally untwists a short segment of DNA and breaks the hydrogen bonds joining the two chains of the double helix; see Fig. 6.1. The enzyme produces considerable twist in the downstream DNA which may then induce Wr, hence supercoils. The dynamics of DNA in many cellular processes, including the interplay between Tw and Wr during transcription, remain largely unresolved and are the focus of ongoing research.

In the previous chapters we have applied the elastic rod model to study looping by Lac repressor. Our model provides considerable insight into the energetics and topology of

---

[1] In this chapter we quote significant passages from our previous publication [62].

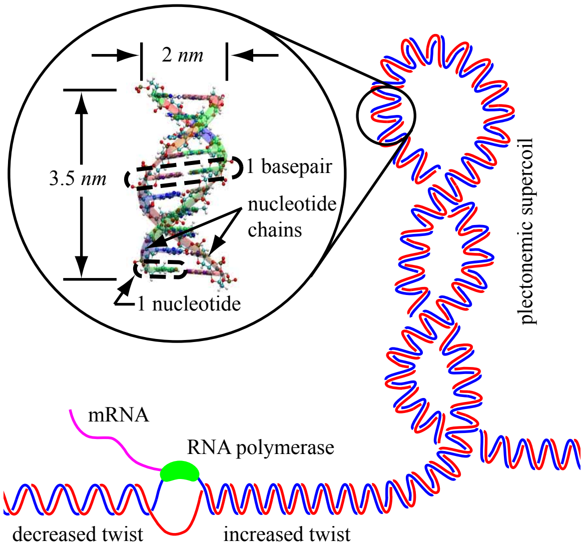


Figure 6.1 At a short length scale (inset), DNA is composed of two nucleotide chains that bond together (basepairing) and twist around each other to form the familiar double helix. During transcription, an enzyme (RNA polymerase) induces DNA supercoiling which may result in long length scale bending and twisting of the DNA helical axis (supercoiling).

DNA looping [37, 61]. Although adequate for looping, our model neglects electrostatic interactions which are essential when considering processes that result in DNA self contact and plectoneme formation.

The objective of this chapter is to contribute an electrostatic extension for the elastodynamic rod model developed in [38] and reviewed in Section 2.1 to enable the analysis of DNA supercoil dynamics. Here we present an electrostatic formulation for this rod model and outline the computational strategy for its implementation. Then we present sample results and draw comparisons with results from existing (equilibrium) models.

6.1 Incorporating Electrostatics

DNA is an electrostatically charged molecule with negative charges distributed along the backbone of the double helix. Consequently, as distant segments along the molecule's contour length approach one another they generate repulsive electrostatic forces. Several approximations have been proposed to account for electrostatic effects in DNA mechanics models. The simplest of these neglect the details of electrostatics in favor of imitating the overall mechanical behavior. For example, models of mechanical contact (see, for example, [110]) or simple functions describing self contact forces (see, for example [39, 51]) have

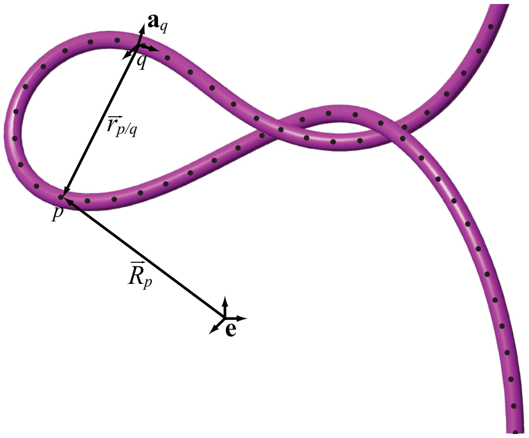


Figure 6.2 The electrostatic force is dependent upon all pairwise vectors $\vec{r}_{p/q}$ (the position ($\vec{R}_p$) of an electric charge at gridpoint p on the DNA axis relative to another at q). Also depicted are the fixed inertial reference frame $\mathbf{e}$ and a body fixed frame $\mathbf{a}_q$ for gridpoint q.

been proposed. Parameters for these models include an equivalent radius for mechanical contact of supercoiled DNA (larger than the physical radius of the molecule) which is meant to account for electrostatic repulsion.

In contrast, herein we approximate the true electrostatic interactions with an electrostatic potential function. To this end, we consider a uniform distribution of point charges along the axis of the elastic rod model. The magnitude and density of these charges are prescribed to create an equivalent charge distribution for DNA in an ionic solution. Positive ions surrounding DNA tend to screen a portion of the molecule's charge and thereby reduce the effective charge for self interactions; see, for example, [110]. The following Debye-Hückel potential describes these electrostatic interactions between charges along DNA

$$E_{elec} = \sum_{p=1}^{N} \sum_{q=p+1}^{N} \frac{\nu^2 l_o^2 e^{-\kappa_D r_{p,q}}}{D r_{p,q}}; \tag{6.1}$$

see, for example, [110]. The total electrostatic potential energy is the sum of all pairwise interactions of N point charges located at points p and q. In (6.1), $r_{p,q}$ is the distance between point charges (see Fig. 6.2), ν is the charge density, l_o is the length of DNA between adjacent point charges, $D = 4\pi\varepsilon$ with ε being the permittivity of water, and $1/\kappa_D$ is the Debye length. For simplicity in our derivation point charges are assigned to the spatial gridpoints used to discretize the elastic rod such that l_o is the spatial discretization size. In general, l_o could be larger or smaller than the spatial discretization size. As a general rule, Vologodskii and Cozzarelli [110] suggest that $l_o \leq 2/\kappa_D$ for the effects of electrostatics to be independent of l_o.

The electrostatic force ($\vec{F}_{p,q}$) acting on a point charge at p due to the interaction with another point charge at q follows from the electrostatic potential (6.1)

$$\vec{F}_{p,q} = -\nabla E_{elec} = -\frac{\partial E_{elec}}{\partial \vec{r}_{p,q}}. \tag{6.2}$$

Substituting (6.1) into (6.2) yields

$$\vec{F}_{p,q} = \frac{\nu^2 l_o^2}{D} \frac{e^{-\kappa_D r_{p,q}}}{r_{p,q}^2} \left(\kappa_D + \frac{1}{r_{p,q}} \right) \vec{r}_{p,q}. \tag{6.3}$$

To fully account for electrostatics in the elastic rod formulation one should account for *all* interactions between all point charges. Thus we define $\vec{F}_p$ as the net electrostatic force on point p,

$$\vec{F}_p = \sum_{q \neq p} \frac{\nu^2 l_o^2}{D} \frac{e^{-\kappa_D r_{p,q}}}{r_{p,q}^2} \left(\kappa_D + \frac{1}{r_{p,q}} \right) \vec{r}_{p,q}. \tag{6.4}$$

The interaction forces $\vec{F}_p$ can then be inserted into the discretized form of the governing equations for the elastic rod as components of $\vec{F}_{body}$ in (2.3). Bear in mind that $\vec{F}_{body}$ is a 'distributed' force, and so we must divide $\vec{F}_p$ by the discretization length (l_o) when forming $\vec{F}_{body}$.

6.2 Numerical Implementation

Upon discretizing (2.3)–(2.6) we arrive at a system of nonlinear equations at each timestep following [38]. We solve this system of equations using Newton-Raphson iteration for the field variables ($\vec{v}$, $\vec{\omega}$, $\vec{\kappa}$, $\vec{f}$) at the next timestep. To this end, the system of equations is linearized and a Jacobian matrix is assembled from the coefficients of the field variables ($\vec{v}$, $\vec{\omega}$, $\vec{\kappa}$, $\vec{f}$). To calculate these coefficients we compute partial derivatives of all the terms in the equations with respect to all of the field variables. Calculation of the Jacobian is complicated by the fact that $\vec{r}_{p/q}$ in (6.4) depends on an integrated form of the field variables ($\vec{v}$, $\vec{\omega}$). To proceed, we must therefore develop the dependence of $\vec{r}_{p,q}$ on these field variables following the conceptual steps below.

We begin with an expression for the relative position vectors in terms of the position of gridpoints along the DNA contour length

$$\vec{r}_{p/q}^{\,t+1} = \vec{R}_p^{t+1} - \vec{R}_q^{t+1}. \tag{6.5}$$

Here the superscript $t+1$ denotes the next timestep. In particulare, we develop the dependence of the electrostatic force at the $t+1$ timestep upon the field variables at the $t+1$ timestep. The field variables at all previous timesteps are considered known when updating the Jacobian (or constant during the linearization process). Next, the update formula for R_p^{t+1} from [38] reveals

$$\vec{R}_p^{t+1} = \vec{R}_p^t + \Delta t \left\{(1-\gamma)[\mathbf{L}_p^t]^T \vec{v}_p^t + \gamma[\mathbf{L}_p^{t+1}]^T \vec{v}_p^{t+1}\right\} \tag{6.6}$$

where γ is a weighting parameter chosen for the generalized alpha method [18] and here we set $\gamma = 1$. Δt is the timestep size, and $\mathbf{L}_p^t$ and $\mathbf{L}_p^{t+1}$ are direction cosine matrices for $\mathbf{a}_p^t$ and $\mathbf{a}_p^{t+1}$ respectively. That is

$$\mathbf{L}_p^{t+1} = \begin{bmatrix} \mathbf{a}_{p1}^{t+1} \cdot \mathbf{e}_1 & \mathbf{a}_{p1}^{t+1} \cdot \mathbf{e}_2 & \mathbf{a}_{p1}^{t+1} \cdot \mathbf{e}_3 \\ \mathbf{a}_{p2}^{t+1} \cdot \mathbf{e}_1 & \mathbf{a}_{p2}^{t+1} \cdot \mathbf{e}_2 & \mathbf{a}_{p2}^{t+1} \cdot \mathbf{e}_3 \\ \mathbf{a}_{p3}^{t+1} \cdot \mathbf{e}_1 & \mathbf{a}_{p3}^{t+1} \cdot \mathbf{e}_2 & \mathbf{a}_{p3}^{t+1} \cdot \mathbf{e}_3 \end{bmatrix}. \tag{6.7}$$

(The numerical subscripts 1, 2, and 3 on the reference frames $\mathbf{a}_p^{t+1}$ and $\mathbf{e}$ are used to denote their first, second and third unit basis vectors respectively.) Here the direction cosine matrices are used to express the velocities ($\vec{v}_p^t$ and $\vec{v}_p^{t+1}$), which in our algorithm we express in terms of their respective local frames ($\mathbf{a}_p^t$ and $\mathbf{a}_p^{t+1}$), in terms of the inertial frame ($\mathbf{e}$). (We express the position vectors $\vec{R}_p^t$ and $\vec{R}_p^{t+1}$ in terms of the inertial frame $\mathbf{e}$.) In the numerical formulation the direction cosine matrix is updated each timestep by the following relation

$$\mathbf{L}_p^{t+1} = \exp\left\{-\tilde{\theta}_p^{t+1}\right\} \mathbf{L}_p^t, \tag{6.8}$$

where $\tilde{\theta}_p^{t+1}$ is a 3×3 skew symmetric matrix formed from the elements of the vector

$$\vec{\theta}_p^{t+1} = \Delta t \left\{(1-\gamma)\vec{\omega}_p^t + \gamma \vec{\omega}_p^{t+1}\right\} \tag{6.9}$$

such that its cross product with any vector $\vec{z}$ is $\vec{\theta}_p^{t+1} \times \vec{z} = \tilde{\theta}_p^{t+1} \vec{z}$ [38]. (We also express $\vec{\theta}_p^t$ and $\vec{\theta}_p^{t+1}$ in terms of their respective body fixed frames $\mathbf{a}_p^t$ and $\mathbf{a}_p^{t+1}$.) The $\exp\{\}$ operator is the matrix exponential operator. For the skew symmetric argument, $-\tilde{\theta}_p^{t+1}$, this operation can be expressed as

$$\exp\left\{-\tilde{\theta}_p^{t+1}\right\} = \mathbf{I} + \frac{\tilde{\theta}_p^{t+1}}{\theta_p^{t+1}} \sin\left\{\theta_p^{t+1}\right\} + \left(\frac{\tilde{\theta}_p^{t+1}}{\theta_p^{t+1}}\right)^2 \left(1 - \cos\left\{\theta_p^{t+1}\right\}\right). \tag{6.10}$$

Equations (6.5)–(6.10) represent the key steps defining how the electrostatic forces depend on $\vec{\omega}$ and $\vec{v}$.

We recognize that there are many possible strategies to reduce the computational cost of calculating both the electrostatic forces and the associated Jacobian. Our strategies are as follows.

1. *Using Infrequent Jacobian Updates.* The Jacobian, in the Newton-Raphson iteration, points the next iteration in the direction of steepest descent toward the solution of the nonlinear system of equations. If the Jacobian is simply close to pointing in the direction of steepest descent, the method often converges to the solution. Consequently, we have found it unnecessary to recompute the electrostatic contribution to the Jacobian every Newton-Raphson iteration or even every timestep. Instead, the electrostatic contribution is held fixed for many timesteps (each with several Newton-Raphson iterations) while the non-electrostatic portion of the Jacobian is updated every iteration of every timestep. Doing so may require slightly more Newton-Raphson iterations to achieve convergence, but this is far less than the cost of extensive updates for the electrostatic contribution.
2. *Using Incomplete Jacobian.* Again, we take advantage of the idea that the Jacobian need only point in the general direction of steepest descent. The terms in the electrostatic part of the Jacobian that are most significant are those that correspond to pairs of point charges in close proximity. Therefore, we can choose a cutoff distance d_c such that only interactions with $r_{p/q} \leq d_c$ contribute to the Jacobian, while all others are neglected. We find $d_c \sim 15/\kappa_D$ to be a good starting point for the cutoff distance.
3. *Ignoring nearest neighbors.* Equation (6.1) suggests that all interactions between points p and q should be considered when calculating the electrostatic interactions. Fortunately, however, when points p and q are neighbors along the axis of the rod their interaction force remains nearly constant for small changes in the local bending or twisting. Therefore these nearest neighbor interactions contribute little to the Jacobian and can reasonably be ignored. In addition, these nearest neighbor interactions tend to have a stiffening effect on the bending of the molecule and this stiffening is sensitive to discretization length [23]. Therefore, nearest neighbor interactions are sometimes lumped with the elastic bending stiffness of the molecule to provide an effective bending stiffness that also incorporates nearest neighbor electrostatics.

6.3 Results and Discussion

In this section we discuss sample results from our elasto-dynamic rod model and make comparisons with predictions from the existing equilibrium models.

6.3.1 Stability of an elastic rod bent into a circular ring

As a first example, consider an intrinsically straight segment of DNA that is bent into a circular ring such that the two ends stack together (forming a DNA plasmid). Depending on the number of basepairs in the segment, the DNA must distribute torsional stress along its length for the ends to properly stack, while remaining in a circular configuration. The stress free stacking of DNA is such that adjacent basepairs are oriented with about a 34.6° rotation about the helical axis (or equivalently, about 10.4 *bp* per helical turn of DNA). If there are an integer number of helical turns in the segment, the ends can be joined together forming a perfect circle without torsional stress. The elastic rod model allows us to continuously vary the number of basepairs to investigate the effects of forming a circle from different lengths of DNA. We exercise our model to study an elastic rod that is initially deformed into a ring in a torsionally relaxed configuration and then slowly rotate the ends of the rod relative to one another to vary the torsional stress. The torsional rotation of one end relative to the other changes the linking number (*Lk*) of the closed loop such that each full turn changes the linking number by one. We explore the effects of this *Lk* change on the topology of the closed loop and compare our results to the stability calculations of [21] obtained using an equilibrium rod model with idealized mechanical self contact.

An important result is the critical *Lk* (bifurcation value) at which the loop buckles from a circular configuration into a figure-8 configuration as studied in [21]. Given the parameters describing the rod geometry and elasticity in [21], one arrives at distinct *Lk* for the bifurcation from circular to figure-8 configuration (*Lk*=1.16) and from figure-8 to circular configuration (*Lk*=1.03). We approximate these critical values of *Lk* and predict the corresponding non-equilibrium transitions with our dynamic rod model. To this end, we adjust the values for the constants describing electrostatic forces to arrive at an effective radius for mechanical contact approximately equal to the radius of the elastic rod model [21]. For our simulation, the initial condition is a slightly imperfect (non-planar) circle. This facilitates the initiation of the buckling transition in the simulation.

In Fig. 6.3 we present the results of our simulation starting with a plot of *Tw* and *Lk* as functions of time in Fig. 6.3(a). In this simulation *Lk* is prescribed by the rotation (torsional boundary conditions) of the two ends of the rod as a function of time. In Fig. 6.3(b) we

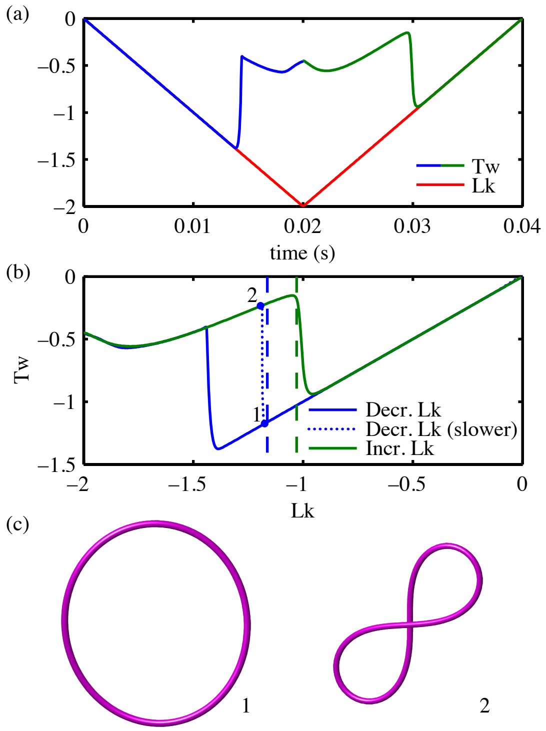


Figure 6.3 Electrostatic self contact in circular twisted rod. (a) The time evolution of *Tw* and *Lk*. *Tw* is plotted using blue (green) to distinguish times when *Lk* is decreasing (increasing). (b) *Tw* vs. *Lk*. The critical linking numbers for transitions between the circular and figure-8, as reported in [21], are shown with dashed vertical lines: blue for decreasing *Lk* and green for increasing *Lk*. The solid blue and green lines represent *Tw* as a function of *Lk* for decreasing and increasing *Lk*, respectively. The additional dotted blue line is attained by lowering the rate of end rotation to 0.2% of its original value in order to show convergence to the critical value of *Lk*. (c) Two deformed states calculated with the slower loading rate and corresponding to the points 1 and 2 in (b).

plot *Tw* vs. *Lk* for both decreasing (blue) and increasing (green) *Lk*. Additionally we plot a curve of *Tw* vs. *Lk* obtained using a lower rate of end rotation. The two vertical lines mark the critical values of *Lk* for transitions from the circular to the figure-8 configuration (blue dashed line) and from the figure-8 to the circular configuration (green dashed line).

Our rod formulation is dynamic and therefore does not, in general, reproduce the equilibrium predictions of [21]. However, as the rate of end rotation is reduced, the dynamic formulation converges to the purely static model [21] (see the dotted blue curve of Fig. 6.3(b)). Interestingly, the transition from figure-8 to circle configuration is much less sensitive to the end rotation rate than that from circle to figure-8. Our dynamic formulation augments the purely static predictions of [21] by uncovering the non-equilibrium transitions between the circular and figure-8 configurations.

Our model is further distinguished by employing an electrostatic potential in lieu of the idealized mechanical contact of [21]. As mentioned above, we chose electrostatic constants consistent with a contact radius used in [21]. The close agreement between our calculations and the predicted critical *Lk* of [21] suggests that the differences in these approaches to account for self contact are minor for this example. In fact, a model for contact is not even necessary to calculate the critical *Lk* for transition from the circular to figure-8 configuration, because the transition begins without self contact. By contrast, the transition from the figure-8 to the circular configuration begins in a self contact configuration and therefore could be more sensitive to the model for contact; but, this does not appear to be the case in this example.

6.3.2 Plectoneme Formation

A second example is motivated by plectoneme formation (DNA supercoiling). Here we parallel the work of Heijden et al. [104] who study the stability of an elastic rod that is twisted while otherwise clamped at both ends. For one example, the clamps are separated by a distance of half the rod's contour length. One end of the rod is rotated about its axis and the resultant equilibrium conformations are computed. We evaluate this system using a dynamic formulation and again adjust the values for the constants describing electrostatic forces to arrive at an effective radius for mechanical contact close to the radius of their elastic rod model. In addition we rotate both ends (in opposite directions) in an attempt to balance any asymmetries in loading that might induce dynamics not experienced in the static model [104]. We explore the stability of plectonemes arising through a relative end rotation of 3.5 turns (22 rad.). As with the previous example, we begin the simulation with initial conditions that facilitate buckling transitions.

Figure 6.4(a) presents sequential snapshots of the dynamic evolution of the elastic rod while the ends are rotated. Interestingly, the rod appears to depart from a symmetric configuration as the ends are rotated and a plectoneme forms. This asymmetry does not appear in the static equilibrium calculations of [104] which suggests that this is sensitive to the *dynamic* end rotation. As indicated, in our dynamic simulation we rotate both ends to reduce the possibility of departing from symmetry. Nevertheless small numerical noise (e.g., round-off error and/or the initial conditions) present in our computational method, must bias the solution toward an asymmetric structure.

Figure 6.4(b) is a bifurcation diagram for the (non-dimensional) torque experienced at the end of the rod for a range of end rotations. The solid black curve represents the predictions of [104] and the dashed blue and red curves are produced from our dynamic

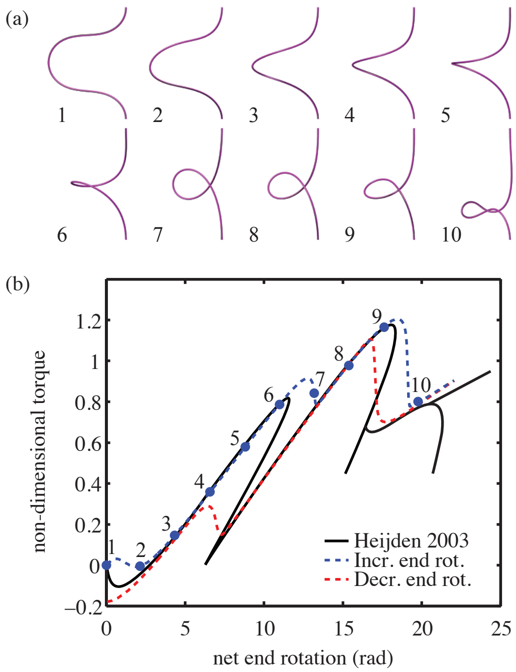


Figure 6.4 Plectoneme formation in a twisted, but otherwise clamped, elastic rod. (a) A series of snapshots of the rod configuration with increasing end rotation. (b) A plot of the non-dimensional torque at one end of the rod as a function of the net end rotation. The black curve represents the the calculated equilibrium curve of Fig. 10(a) in [104]. The other two dashed curves are calculated from our dynamic model and present results for increasing (blue) and decreasing (red) end rotation. The numbered points correspond to the numbered series of snapshots in (a).

model with increasing and decreasing end rotation respectively.

Aside from the explainable dynamic departures from equilibrium, the rod model follows very closely the static predictions of [104]. As with the previous example, the dynamic model experiences hysteresis when subjected to increasing then decreasing loads (i.e., increasing vs. decreasing end rotation). This hysteresis results from non-equilibrium transitions that occur when the dynamic rod continues along a stable equilibrium path into, or near, an instability. For increasing end rotation, the dynamic rod model tends to overshoot the stable pathway before making a transition to another stable pathway. This suggests that viscous damping delays the transition. (Note that relative to future applications concerning DNA, DNA inertial effects are usually insignificant compared to the large viscous forces experienced by the molecule.)

In contrast, for decreasing end rotation, the the dynamic rod model tends to transition before reaching the end of a stable pathway. These transitions, however, may be more sensitive to the model for self contact. Note, the bifurcation diagram of [104] indicates where changes occur in the number of points of self contact. In comparing Fig. 6.4(b) to the equivalent figure in [104], it appears that the dynamic transitions are associated with a change in the number of points of self contact. (In contrast to a mechanical contact model, points of self contact are not clearly defined in an electrostatic formulation.) Specifically, when tracing the curve representing increasing (or decreasing) end rotation, each dynamic transition appears to correspond to an increase (or decrease) in the number of points of self contact. The snapshots numbered 9 and 10 in Fig. 6.4(a) show a change from 1 to 2 points of self contact consistent with this explanation.

For the two dynamic curves of Fig. 6.4(b), the departures from the static equilibrium predictions occurring near zero end rotation are distinguished from the dynamic transitions discussed above. During these ‘departures,’ there are no points of self contact, nor any changes in the number of contacts. For the case of decreasing end rotation approaching zero, the dynamic curve overshoots the equilibrium as a result of viscous damping (much like the transitions discussed above for *increasing* end rotation). Because the simulation is initialized to a relaxed configuration at zero end rotation, the curve for increasing end rotation aligns with the static prediction; and it immediately departs upon increasing the end rotation. Again, viscous damping prevents the dynamic simulation from following the static equilibrium curve. To further understand the behavior near zero end rotation it is useful to consider extending the bifurcation diagram to include negative end rotations. The resulting diagram becomes symmetric about the axis of zero end rotation. For this extended figure we observe how the dynamic pathway would now overshoot the static equilibrium pathway because of its sharp ‘S’ curve.

6.4 Conclusions

Here we outline the steps needed to incorporate a model for electrostatic interactions into our elasto-dynamic rod approximation for DNA. By employing this approximation, we demonstrate, via specific examples, that it is consistent with predictions from existing equilibrium approximations that consider mechanical self contact. With this new capability, the elasto-dynamic rod model is now well positioned to study DNA supercoiling with the added influence of electrostatics.

Chapter 7

Modeling the Dynamic Relaxation of DNA Supercoils by Topoisomerase IB

The addition of electrostatics and self contact to our elastic rod model in Chapter 6 now positions us to perform the first (to our knowledge) multi-scale simulation of the relaxation of DNA supercoils by topo IB. To do so, we introduce an approximation for hydrodynamic drag experienced by the DNA. To simulate the interaction between DNA and topo IB, we pair our elastic rod model with a reduced order model of topo IB obtained by the Andricioaei lab at the University of California - Irvine using MD.

7.1 Approximation for Hydrodynamic Drag

The transient dynamics of DNA are significantly damped by the hydrodynamics. Here we offer a first approximation for the effects of hydrodynamic drag using estimated drag coefficients for translation and axial (torsional) rotation. We arrive at the drag coefficients per unit length by neglecting hydrodynamic interactions and considering the drag on a straight rigid rod with equivalent contour length. We further assume that the drag forces on this straight rod are uniformly distributed along its length and are proportional to the rod's velocity. Following these assumptions, the drag coefficients (per unit length) for axial ($C_{\|}$) and lateral ($C_{\perp}$) motions, and rotations about the helical axis (C_{axial}) are:

$$C_{\|} = \frac{2\pi\mu}{\ln\left(\frac{L}{d}\right) - 0.2}, \tag{7.1}$$

$$C_{\perp} = \frac{4\pi\mu}{\ln\left(\frac{L}{d}\right) + 0.84}, \tag{7.2}$$

$$C_{axial} = \pi \mu d^2; \tag{7.3}$$

see [44], for example. Here, $\mu = 1.3 \times 10^{-3}\ kg/ms$ is the viscosity of the surrounding buffer, $L = 68\ nm$ is the DNA contour length used in the example system below, and $d = 2\ nm$ is the diameter of the molecule. The translational drag coefficients are only weakly dependent upon L for $d \ll L$ and therefore provide a reasonable approximation for drag. A similar approximation is described in [122]. As with the electrostatic forces described in Chapter 6, these drag relations present themselves through the variables $\vec{F}_{body}$ and $\vec{Q}_{body}$ in the governing equations ((2.3)-(2.6)). Specifically, the drag force

$$\vec{F}_{drag} = -\begin{bmatrix} C_\perp & 0 & 0 \\ 0 & C_\perp & 0 \\ 0 & 0 & C_\parallel \end{bmatrix} \vec{v} \tag{7.4}$$

and torque

$$\vec{Q}_{drag} = -\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_{axial} \end{bmatrix} \vec{\omega} \tag{7.5}$$

contribute to $\vec{F}_{body}$ and $\vec{Q}_{body}$ respectively. (In these two expressions, $\vec{F}_{drag}$, $\vec{Q}_{drag}$, $\vec{v}$, and $\vec{\omega}$ are expressed in terms of the local frame **a**.)

7.2 Modeling the topo IB - DNA system

Motivated by the recent tethered particle experiments [54, 55] described in Chapter 1, we apply the elasto-dynamic rod model to understand the transient dynamics of supercoil relaxation induced by topo IB; see Fig. 7.1(a). For this initial study, we focus on the relaxation of a short 200 *bp* plectoneme following the nicking of a single backbone by topo IB.

For our study here we exercise the electrostatic approximation introduced in Chapter 6 using the following parameter values: $N = 200$, $\nu = 0.608\ e^-/\AA$ (assuming a 0.1 M salt concentration, see [110]), $l_o = 0.34\ nm$, $\kappa_D = 1.04\ 1/nm$, and $\varepsilon = 80 \times 8.854 \times 10^{-12}\ F/m$. Additionally, we exclude 40 nearest neighbor interactions (corresponding to a contour length of 13.6 *nm*) and thereby focus on electrostatic interactions at distant sites as measured along the contour length. Our choice of 13.6 *nm* is consistent with the work of [110] which suggests that the cutoff should be between 10 and 20 *nm*.

To arrive at the initial conditions for our simulation of relaxation by topo IB, we first

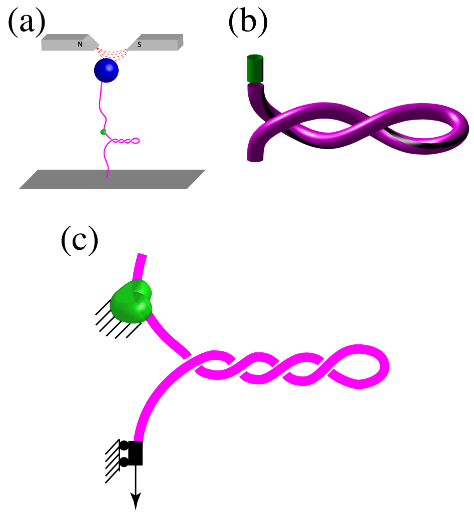


Figure 7.1 (a) Schematic of a tethered particle experiment using a magnetic trap to detect the relaxation of supercoils due to topo IB. (b) Illustration of the initial plectoneme (initial conditions) for our simulation. The purple (plectoneme) domain represents the DNA that is explicitly modeled using the rod formulation. The green domain, represents an (unmodeled) continuation of the DNA on the opposite side of the nicking site. The interface between the two domains is where topo IB acts. (c) Enlarged schematic and free-body diagram of (a) showing the boundary conditions. The upper boundary conditions describe the fixed position and orientation of topo IB with an approximation to the topo IB-induced torque on DNA determined from the free energy landscape of [115]. The lower end is fixed against all rotations and lateral translations. However, this end is allowed to translate along the tangential direction under the action of an applied (constant) tension.

simulate the formation of a plectoneme, in equilibrium, from an otherwise straight segment of DNA [39, 62]. To do this with the model, we initiate the computation in a buckled column configuration with both ends clamped, their tangents aligned and separated by a distance of about $L/3$. Then we apply a rotation about the tangent of one end to reach $\Delta Lk = 4$. The linking number ($\Delta Lk = 4$) is obtained from the sum of the Wr (upon constructing a closure following [82]) and ΔTw of the DNA. Finally, we release the translational constraint along the tangent of one end and apply a constant tension (0.2 pN). This process yields a highly stressed plectoneme having internal (strain) energy of 133 kT.

For the purpose of studying the transient relaxation following nicking, we now treat this fully formed plectoneme at rest as the initial conditions; refer to Fig. 7.1(b). The long purple domain represents the plectoneme that is modeled whereas the short green domain represents an un-modeled continuation of the DNA on the opposite side of the enzyme. The interface between the purple and green domains represents the location where topo IB makes a single strand nick thereby permitting the highly energetic plectoneme to relieve energy by rotating about the intact strand. The black stripe serves to visualize the twist state

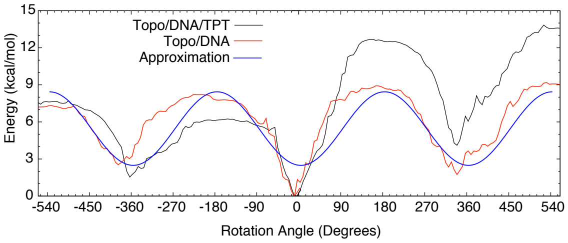


Figure 7.2 Free energy landscape calculated by umbrella sampling as reported in [115]. The blue curve represents an approximate sinusoidal approximation for use in the rod model.

of the supercoiled DNA.

For the duration of the simulation, the bottom end is held under a constant tension (0.2 *pN*) while otherwise fixed against rotations and lateral translations. In the (initial) equilibrium state, the top end of the rod is held fixed, prohibiting both translation and rotation. However, upon nicking by topo IB, this end is allowed to rotate about a fixed axis parallel to the helical axis and located on the circumference of the rod (along the black stripe illustrated in Fig. 7.1(b)). Rotation about this eccentric axis represents rotation about the intact (un-nicked) backbone of the DNA. As supercoils are relieved, topo IB imparts a reaction torque on the DNA. Recently, MD simulations revealed an energy landscape for topo IB in complex with a short fragment of DNA as a function of the rotation angle about the eccentric axis [115]. Figure 7.2 illustrates a sinusoidal approximation to the potential calculated using MD. We assume that this energy landscape represents the free energy of topo IB as a function of the rotation angle, that the potential is periodic in the rotation angle, and that it does not depend on any other quantities (such as the speed of rotation). Following these approximations, the torque ($T_{topo\ IB}$) imparted to the DNA by topo IB is given by (the negative gradient of the potential)

$$T_{topo\ IB} = -5\ kT \sin(\theta)\,, \tag{7.6}$$

where kT is the product of the Boltzmann constant and temperature; and θ is the angular rotation of the top end measured from the equilibrium state.

To translate the relation for torque on the end ($T_{topo\ IB}$) to rotation of the end (θ), in (7.6), into boundary conditions is not straight forward for our model. Our formulation solves for four vector unknowns ($\{\vec{v}, \vec{\omega}, \vec{\kappa}, \vec{f}\}$); therefore we must relate these unknowns to $T_{topo\ IB}$ and θ. We relate $T_{topo\ IB}$ to the $\vec{\kappa}(L,t)$ at the the end ($s = L$) of the DNA by considering the torsional component of $\vec{q}(L,t)$ using (2.1),

$$T_{topo\ IB} = \hat{t}^T(L,t)\mathbf{B}\vec{\kappa}(L,t). \tag{7.7}$$

Additionally, we relate $\theta(t)$ to an integral of $\vec{\omega}(L,t)$ about the local tangent $\hat{t}(L,t)$,

$$\theta(t) = \int_0^t \hat{t}^T(L,t)\vec{\omega}(L,\tau)d\tau, \tag{7.8}$$

where we utilize the generalized-alpha method [18] for the discretization and evaluation of the integral.

7.3 Results and Discussion

The following figures illustrate the simulated relaxation of the plectoneme upon nicking by topo IB. Figure 7.3 plots the topological variables (ΔLk, ΔTw, and Wr) of the plectoneme as functions of time. In addition, the conformations at three specific times (corresponding to $\Delta Lk = \{3,\ 2,\ 1\}$) are illustrated. Relaxation of supercoils slows as the simulation progresses through time, possibly reflecting the overwhelming influence of drag as the energy is reduced. Interestingly, the DNA ultimately relaxes to a configuration with $\Delta Lk \sim 1$, and remains at a local minima of the free energy. The global free energy minimum corresponds to the fully extended and untwisted configuration. Figure 7.4 plots the contributions to the free energy (elastic strain energy and topo IB free energy) and sheds light on the energy barrier between this local minimum and the global minimum. As mentioned, the simulation relaxes to an equilibrium with $\Delta Lk \sim 1$ which corresponds to a minimum in potential for topo IB. Therefore to escape the local minimum, the system must overcome a barrier with height approximately the same as the energy difference from minimum to maximum of the topo IB potential (about 10 kT); which is significant compared to thermal energy (1 kT).

In reference to Fig. 7.3, the simulations suggest that supercoil relaxation occurs on a $0.1 - 1.0\ \mu s$ timescale. Because of increased hydrodynamic drag, significantly larger plectonemes could relax at significantly lower rates rendering this process observable in future single molecule experiments.

Figure 7.5 illustrates the reaction torque at each end of the DNA as functions of time. Initially, the reaction torques are identical as the DNA is at equilibrium. Upon snipping by topo IB, the reaction torques follow distinct paths prior to achieving the same (lower) value at the final equilibrium. Notice that the reaction torque at the end bound to topo IB experiences large dynamic variations in magnitude and direction. As prescribed by the

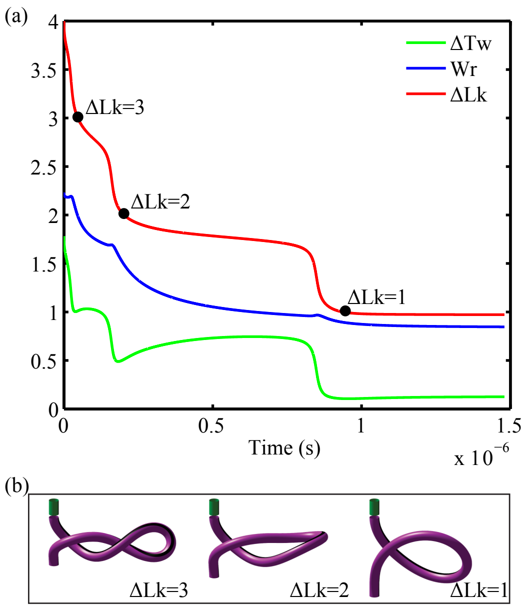


Figure 7.3 (a) Topological variables (ΔLk, ΔTw, and Wr) of the plectoneme as functions of time throughout the relaxation. (b) The conformations at three specific times corresponding to $\Delta Lk = \{3,\ 2,\ 1\}$.

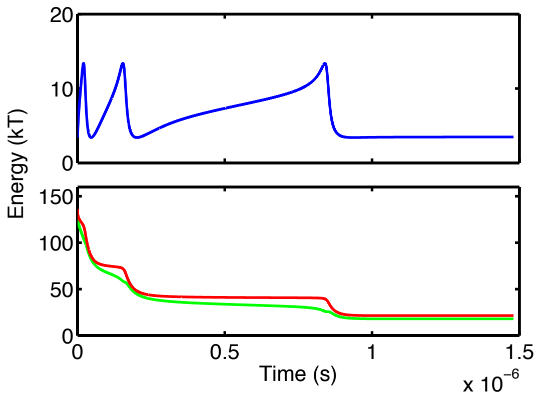


Figure 7.4 (above) Potential of the topo IB alone as a function of time. (below) Strain energy of the elastic rod (green) and sum of the topo IB potential and elastic rod strain energy (red) as functions of time.

torque boundary conditions ((7.6)), the reaction torque at this end is based on the free energy landscape computed by MD which depends only on the rotation of this end. The periodic nature of the potential results in short time intervals in which the torque drives relaxation (negative torque) and other intervals during which the torque resists relaxation

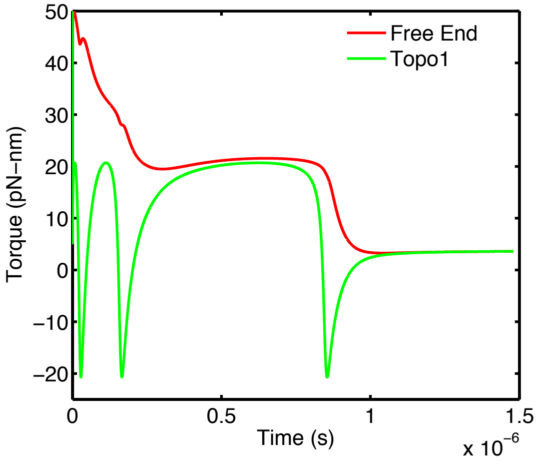


Figure 7.5 Reaction torques on the DNA. The red curve corresponds to the reaction torque at the free (tension loaded) end while the green curve corresponds to the reaction torque at the opposite end and bound to topo IB.

(positive torque). During intervals with negative torque, the plectoneme experiences fast conformational changes and associated rapid changes in ΔLk as seen previously in Fig. 7.3. By contrast, the torque on the (tension loaded) free end experiences a gradual and stepwise reduction in torque suggesting a quasi-static process. Again, this gradual relaxation results from the significant hydrodynamic drag. Specifically, the hydrodynamic drag along the length of the molecule has the effect of filtering out the rapid kinetics of the torque at the topo IB interface from the torque experienced by the free end. This suggests that experiments that seek to observe rapid kinetics, with a magnetic trap and large magnetic bead for example, may be challenged by the effects of hydrodynamic drag.

We recognize that the omission of thermal excitation in our model could significantly alter the results presented here. However, we expect the effects of thermal excitation to be minor in the regime in which the strain energy of the DNA is much larger than thermal energy (1 kT). Therefore the highly stressed initial condition, suggests that the initial stages of relaxation may provide a reasonable representation to the true dynamics.

7.4 Conclusion

Here we perform the first multi-scale model of the relaxation of DNA supercoils by topo IB. Our model employs a novel method based on MD simulations to represent the interaction between DNA and topo IB. For the simulation presented here we observe a supercoil relaxation timescale of about $0.1 - 1.0\ \mu s$.

Chapter 8

Summary of Specific Findings and Major Contributions

In this dissertation we have made significant contributions toward understanding the mechanics and function of DNA looping and supercoiling. Here we outline specific findings and major contributions for DNA looping as detailed in Chapters 2-5 and for DNA supercoiling as detailed in Chapters 6-7. We first summarize our specific findings and and follow with a summary of our major contributions.

8.1 Specific Findings

Through our computational studies of DNA looping (Chapters 2-4) we make several specific observations regarding loop geometry and energetics; and here we summarize three of the most significant observations. First, we find that the anti-parallel (**A**) binding topologies generally dominate as energetic minima for both intrinsically straight and intrinsically curved DNA. However, for intrinsically curved DNA there are regions in which the **P1** binding topology dominates suggesting the existence of two distinct loop states for intrinsically curved DNA. Second, the elastic energy of looping the large family of intrinsically curved sequences spans the range of $5-12\ kT$. Interestingly, both **A** and **P1** binding topologies yield comparable energetic minima ($\sim 5\ kT$). Finally, loop energetics and topology is sensitive to changes in protein stiffness when $\beta \sim 0.1$. (Recall that β is a measure of protein stiffness relative to the stiffness of DNA.) Specifically, for stiffnesses of $\beta > 0.1$, the energetic and preferred binding topology landscapes remain qualitatively insensitive to changes in stiffness.

In our analysis of the cryo-EM reconstructions of DNA minicircles (Chapter 5) we identified metrics most likely to distinguish the geometry of populations of individual molecules.

Specifically, we find that the max bend angles, modal energies, and writhe are likely the most powerful discriminators among populations. Interestingly, all the metrics we consider suggest that the MC108 minicircles are the most circular. Furthermore, the modal energy distribution for MC108 suggests that the DNA's persistence length is 34 *nm*.

Dynamic simulations following our efforts to account for self contact and electrostatics in Chapter 6 show good agreement with previous equilibrium theory of self contacting elastic rods. By extending this in Chapter 7 we compute a first estimate for the relaxation timescale of supercoils by topo IB ($0.1 - 1.0\ \mu s$). Significantly, we calculate dynamics on a biologically relevant time scale for a DNA-topo IB complex much larger than is possible to simulate using MD.

8.2 Major Contributions

In Chapters 2-5 we explore the mechanics and function of DNA looping. Throughout these chapters we exercise our elastic rod model in new directions and draw comparisons with existing experimental data. In doing so, we make many specific observations and conclusions relevant to understanding DNA looping as already detailed in Chapters 2-5. Below we highlight four major contributions resulting from the research in these chapters.

1. *A method of representing a large family of intrinsically curved DNA.* Three intrinsically curved sequences have been studied by the Kahn lab at the University of Maryland [29, 70, 73] and feature an A-tract bent domain flanked by two straight domains. Loops formed from these three sequences have shown interesting properties including hyperstability and distinct loop geometries. To better understand looping of an entire family of intrinsically curved sequences (including these three), we develop the SHS representation in Chapter 3. This representation is composed of a helically bent rod domain flanked by two straight rod domains. We account for differences between sequences in this family by introducing two phasing parameters. This SHS representation allows us to efficiently compute looping behaviors for a large family of possible designed sequences. It further helps us isolate and understand the effects of changes in phasing from sequence to sequence.
2. *A method to account for LacI flexibility.* As indicated in Chapter 1, experimental and computational evidence suggest that the LacI protein is not rigid but flexible. These studies provide evidence of concentrated flexibility in the head domains and/or in the 'V' domain; however, the precise modes of flexibility and their associated stiffnesses have yet to be determined. In Chapter 4 we develop a method to account for flexibility

of the LacI protein. Specifically, we extend the elastic rod model to represent possible rigid and flexible domains within the protein. By employing our representation for flexibility over a wide range of stiffnesses for concentrated flexibility in the head and/or 'V' domains, we build understanding of its effects on looping. We also draw comparisons with past experiments (and prepare for comparing with current and future experiments) to help characterize the nature of flexibility of the LacI protein.

3. *Methods to compare cryo-EM images of DNA minicircles.* Recent cryo-EM work has resulted in several three-dimensional reconstructions each of three different DNA minicircles (of 100, 106, and 108 *bp*). However, because of the limited sample size, the influence of thermal fluctuations, and the limited resolution associated with cryo-EM and the reconstruction process, the geometric and energetic differences between the minicircles are not easily attained. To this end, we develop and identify a few metrics that quantify their geometry and hold promise to discriminate between the three constructs. Our analysis of the cryo-EM reconstructions and Brownian dynamics simulations of minicircles suggest that maximum bend angles, modal energies, ellipticity and Wr may be significant discriminators.
4. *Computational predictions that motivate new experiments.* Throughout Chapters 2-5 we find our rod model agrees with many existing experimental observations of DNA loops. By exercising our model in directions previously unexplored, we make several interesting new predictions. For example, our predictions suggest sequences that form two distinct loops (e.g. different binding topologies or protein conformation) with similar energetics and therefore may interconvert. Our predictions in Chapter 3 have already motivated an extensive experimental effort in the Kahn lab at the University of Maryland.

After establishing the foundation for modeling the mechanics of static DNA loops in Chapters 2-5, we make the transition to our second model system, relaxation of supercoils by topo IB. Consequently, in Chapters 6-7 we necessarily make model improvements as well as complete a first simulation of the relaxation of supercoils by topo IB. Here we enumerate two major contributions from these chapters.

1. *An efficient computational algorithm for self contact and electrostatics.* While the looping studies of Chapters 2-5 do not consider DNA self contact, our interest in the relaxation of supercoils by topo IB requires an accurate and efficient model representation of plectonemic supercoils. Therefore, in Chapter 6 we extend our elastic rod model to account for DNA self contact and electrostatic interactions. In addition, we implement strategies to reduce the significant computational cost of accounting for these interactions.

2. *A first multi-scale model of the relaxation of supercoiled DNA by topo IB*. The range of time and length scales associated with the relaxation of supercoiled DNA make for significant computational modeling challenges. For example, the short length scale (~ 1 *nm*) conformational changes of topo IB induce long length scale (greater than the persistence length) conformational changes in supercoiled DNA. We overcome many of these challenges in Chapter 7 with our first multi-scale simulation of the relaxation of supercoiled DNA by topo IB. The multi-scale resulted from combining recent MD modeling of topo IB by the Andricioaei lab at the University of California - Irvine with our elastic rod model for DNA.

Appendices

Appendix A

Numerical Algorithm

Here we outline the finite difference algorithm used for our elastic rod model for DNA. Our algorithm is written in MATLAB® (Natick, Massachusetts, USA) and is freely availably upon request (tlillian@umich.edu or ncp@umich.edu). We begin by rewriting the system of nonlinear partial differential equations ((2.3)-(2.6)) describing the elastic rod formulation as

$$M\dot{Y} + KY_s + F = 0. \tag{A.1}$$

Here Y is a 12×1 vector of field variables $\left\{\vec{v}, \vec{\omega}, \vec{\kappa}, \vec{f}\right\}$ with $\dot{Y} = \frac{\partial Y}{\partial t}$ and $Y_s = \frac{\partial Y}{\partial s}$ as partial derivatives with respect to space and time respectively. The matrices M and K are 12×12 matrices defined as

$$M = \begin{bmatrix} \Theta & \Theta & \Theta & \Theta \\ \Theta & \Theta & I & \Theta \\ \Theta & \mathbf{I} & \Theta & \Theta \\ mI & \Theta & \Theta & \Theta \end{bmatrix},$$

and

$$K = -\begin{bmatrix} I & \Theta & \Theta & \Theta \\ \Theta & I & \Theta & \Theta \\ \Theta & \Theta & \mathbf{B} & \Theta \\ \Theta & \Theta & \Theta & I \end{bmatrix}.$$

Here, Θ is a 3×3 zero matrix, I is a 3×3 identity matrix, m is the mass per unit length, $\mathbf{I}$ is a 3×3 the inertia matrix per unit length, and $\mathbf{B}$ is the 3×3 stiffness tensor. The 12×1

vector F is defined as

$$F = \left[\begin{array}{c} \vec{\omega} \times \hat{t} - \vec{\kappa} \times \vec{v} \\ -\vec{\kappa} \times \vec{\omega} \\ -\left(\frac{\partial \mathbf{B}}{\partial s} \vec{\kappa} - \frac{\partial (\mathbf{B} \vec{\kappa}_o)}{\partial s} \right) + \vec{\omega} \times \mathbf{I} \vec{\omega} + \vec{f} \times \hat{t} - \vec{\kappa} \times \mathbf{B} \left(\vec{\kappa} - \vec{\kappa}_o \right) - \vec{Q}_{body} \\ m \left(\vec{\omega} \times \vec{v} \right) - \vec{\kappa} \times \vec{f} - \vec{F}_{body} \end{array} \right]. \qquad \text{(A.2)}$$

Here, $\hat{t}$ is the tangent of the rod, $\vec{F}_{body}$ is a body force per unit length on the rod, and $\vec{Q}_{body}$ is a body moment per unit length on the rod. The matrices M, K, and F in general are functions of the arclength coordinate (s) and time (t). In addition, F may also be a function of the configuration of the rod (the position and orientation of the rod axis as a function of s). For example, in Chapter 6 a dependence on configuration arises from the inclusion of electrostatic forces.

We write boundary conditions (that in general are functions of time) in the following form

$$\left[\begin{array}{cc} C_0 & C_L \end{array} \right] \left[\begin{array}{c} Y_{s=0} \\ Y_{s=L} \end{array} \right] = c. \qquad \text{(A.3)}$$

Here C_0 and C_L are 12×12 matrices and c is a 12×1 vector. The entries of C_0, C_L and c are chosen to describe the boundary conditions on the rod. To represent Dirichlet boundary conditions on $\vec{v}$ and $\vec{\omega}$ at both ends, for example, the first 6 rows and columns of C_0 and the second 6 rows and first 6 columns of C_L would both form 6×6 identity matrices while the remaining entries would all be zeros. In this special case, the entries of c would be the prescribed values of $\vec{v}$ and $\vec{\omega}$ at either end.

To solve this system of equations we discretize it into N_s spatial gridpoints. We use j to index grid-points from 1 to N_s. We now write (A.1) for a point $(j - \frac{1}{2})$ halfway between grid-points $j-1$ and j

$$(M\dot{Y})_{j-\frac{1}{2}} + (KY_s)_{j-\frac{1}{2}} + (F)_{j-\frac{1}{2}} = 0. \qquad \text{(A.4)}$$

We approximate $(M\dot{Y})_{j-\frac{1}{2}}$ with

$$(M\dot{Y})_{j-\frac{1}{2}} = \frac{1}{2}(M_{j-1}\dot{Y}_{j-1} + M_j\dot{Y}_j), \qquad \text{(A.5)}$$

$(KY_s)_{j-\frac{1}{2}}$ with

$$(KY_s)_{j-\frac{1}{2}} = \frac{1}{2}(K_{j-1} + K_j)\frac{1}{\Delta s}(-Y_{j-1} + Y_j), \qquad \text{(A.6)}$$

and $(F)_{j-\frac{1}{2}}$ with

$$(F)_{j-\frac{1}{2}} = \frac{1}{2}(F_{j-1} + F_j). \tag{A.7}$$

Substituting (A.5)-(A.7) into (A.4), simplifying, and arranging into matrix form we have

$$\begin{bmatrix} M_{j-1} & M_j \end{bmatrix} \begin{bmatrix} \dot{Y}_{j-1} \\ \dot{Y}_j \end{bmatrix} + \begin{bmatrix} -(K_j + K_{j-1})\frac{1}{\Delta s} & (K_j + K_{j-1})\frac{1}{\Delta s} \end{bmatrix} \begin{bmatrix} Y_{j-1} \\ Y_j \end{bmatrix} + F_{j-1} + F_j = 0 \tag{A.8}$$

Considering $1 \leq j \leq N_s$, we have $(N_s - 1)$ equations like (A.8). For more compact notation, we now define the following $12N_s \times 1$ vectors:

$$\bar{\dot{Y}} = \begin{bmatrix} \dot{Y}_1 \\ \dot{Y}_2 \\ \dot{Y}_3 \\ \vdots \\ \dot{Y}_{N_s} \end{bmatrix}, \tag{A.9}$$

$$\bar{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_{N_s} \end{bmatrix}, \tag{A.10}$$

and

$$\bar{F} = \begin{bmatrix} F_1 + F_2 \\ F_2 + F_3 \\ F_3 + F_4 \\ \cdots \\ F_{N_s - 1} + F_{N_s} \\ c \end{bmatrix}. \tag{A.11}$$

We further define the following $12N_s \times 12N_s$ matrices:

$$\bar{M} = \begin{bmatrix} M_1 & M_2 & 0 & \cdots & 0 \\ 0 & M_2 & M_3 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & M_{N_s-1} & M_{N_s} \\ 0 & \cdots & & & 0 \end{bmatrix} \tag{A.12}$$

and

$$\bar{K} = \frac{1}{\Delta s} \begin{bmatrix} -K_2 - K_1 & K_2 + K_1 & 0 & \cdots & 0 \\ 0 & -K_3 - K_2 & K_3 + K_2 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -K_{N_s} - K_{N_s-1} & K_{N_s} + K_{N_s-1} \\ \Delta s C_0 & 0 & \cdots & 0 & \Delta s C_L \end{bmatrix}. \tag{A.13}$$

Now assembling equations (A.8) (for $1 \leq j \leq N_s$) and (A.3) we have

$$\bar{M}\dot{\bar{Y}} + \bar{K}\bar{Y} + \bar{F} = 0. \tag{A.14}$$

Note that (A.14) is continuous in time. We now integrate it in time using the generalized alpha method [18]; see also [35, 38].

$$(1-\alpha_m)\bar{M}\dot{\bar{Y}}^i + \alpha_m\bar{M}\dot{\bar{Y}}^{i+1} + (1-\alpha_k)\bar{K}\bar{Y}^i + \alpha_k\bar{K}\bar{Y}^{i-1} + (1-\alpha_k)\bar{F}^i + \alpha_k\bar{F}^{i-1} = 0 \tag{A.15}$$

Here the superscript i is an index that denotes the discrete temporal grid-point. The variables α_m and α_k are numerical parameters for the generalized alpha method. Because $\bar{M}$ and $\bar{K}$ may themselves be functions of time we perform a weighted average following the generalized alpha method

$$\bar{M} = (1-\alpha_m)\bar{M}^i + \alpha_m\bar{M}^{i-1}, \tag{A.16}$$

$$\bar{K} = (1-\alpha_k)\bar{K}^i + \alpha_k\bar{K}^{i-1}. \tag{A.17}$$

Now substituting (A.16) and (A.17) into (A.15) we have

$$\begin{aligned}(1-\alpha_m)[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\dot{\bar{Y}}^i+\alpha_m[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\dot{\bar{Y}}^{i-1}\\ +(1-\alpha_k)[(1-\alpha_k)\bar{K}^i+\alpha_k\bar{K}^{i-1}]\bar{Y}^i+\alpha_k[(1-\alpha_k)\bar{K}^i+\alpha_k\bar{K}^{i-1}]\bar{Y}^{i-1}\\ +(1-\alpha_k)\bar{F}^i+\alpha_k\bar{F}^{i-1}=0.\end{aligned} \tag{A.18}$$

Here

$$\bar{Y}^i=\bar{Y}^{i-1}+\Delta t[(1-\gamma)\dot{\bar{Y}}^{i-1}+\gamma\dot{\bar{Y}}^i] \tag{A.19}$$

and therefore

$$\dot{\bar{Y}}^i=\frac{1}{\gamma}[\frac{1}{\Delta t}(\bar{Y}^i-\bar{Y}^{i-1})-(1-\gamma)\dot{\bar{Y}}^{i-1}]. \tag{A.20}$$

Here, γ is a numerical parameter of the generalized alpha method. The parameters α_m, α_k, and γ may all be chosen to adjust numerical dissipation and accuracy of the generalized alpha method. Now substituting (A.20) into (A.18) we have

$$\begin{aligned}(1-\alpha_m)[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\frac{1}{\gamma}[\frac{1}{\Delta t}(\bar{Y}^i-\bar{Y}^{i-1})-(1-\gamma)\dot{\bar{Y}}^{i-1}]\\ +\alpha_m[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\dot{\bar{Y}}^{i-1}+(1-\alpha_k)[(1-\alpha_k)\bar{K}^i+\alpha_k\bar{K}^{i-1}]\bar{Y}^i\\ +\alpha_k[(1-\alpha_k)\bar{K}^i+\alpha_k\bar{K}^{i-1}]\bar{Y}^{i-1}+(1-\alpha_k)\bar{F}^i+\alpha_k\bar{F}^{i-1}=0.\end{aligned} \tag{A.21}$$

For convenience, we group terms for $\bar{Y}^i$, $\bar{Y}^{i-1}$, and $\dot{\bar{Y}}^{i-1}$ and define their coefficients as A_1, A_2, and A_3 respectively. Specifically, we define

$$A_1=(1-\alpha_m)[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\frac{1}{\gamma\Delta t}+(1-\alpha_k)[(1-\alpha_k)\bar{K}^i+\alpha_k\bar{K}^{i-1}], \tag{A.22}$$

$$A_2=-(1-\alpha_m)[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\frac{1}{\gamma\Delta t}+\alpha_k[(1-\alpha_k)\bar{K}^i+\alpha_k\bar{K}^{i-1}], \tag{A.23}$$

and

$$A_3=(1-\alpha_m)[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}]\frac{1}{\gamma}[-(1-\gamma)]+\alpha_m[(1-\alpha_m)\bar{M}^i+\alpha_m\bar{M}^{i-1}] \tag{A.24}$$

such that (A.21) becomes

$$A_1\bar{Y}^i+A_2\bar{Y}^{i-1}+A_3\dot{\bar{Y}}^{i-1}+(1-\alpha_k)\bar{F}^i+\alpha_k\bar{F}^{i-1}=0. \tag{A.25}$$

Note that $\bar{F}^i$ may in general be a function of $\bar{Y}^i$.

To find the solution ($\bar{Y}^i$) of this system of nonlinear equations we use Newton-Raphson iterations. We index our iterations with the left superscript l. To initiate the iterations we define

$$ {}^{0}\bar{Y}^i = \bar{Y}^{i-1} + \Delta t \dot{\bar{Y}}^{i-1}. \tag{A.26} $$

Newton-Raphson iterations are based upon linearizing (A.25) about ${}^{l}\bar{Y}^i$, the l^{th} estimate of $\bar{Y}^i$. To do so we define the Jacobian

$$ {}^{l}\bar{J}^i = (1-\alpha_k) \left. \frac{\partial \bar{F}^i}{\partial \bar{Y}^i} \right|_{{}^{l}\bar{Y}^i} + A_1. \tag{A.27} $$

Now we iteratively solve the the following equation for ${}^{l}\bar{Y}^i$:

$$ {}^{l}\bar{Y}^i = {}^{l-1}\bar{Y}^i - \left({}^{l-1}\bar{J}^i\right)^{-1} \left(A_1 {}^{l-1}\bar{Y}^i + A_2 \bar{Y}^{i-1} + A_3 \dot{\bar{Y}}^{i-1} + (1-\alpha_k) {}^{l-1}\bar{F}^i + \alpha_k \bar{F}^{i-1}\right) \tag{A.28} $$

until some error tolerance is satisfied; for example, when $||{}^{l}\bar{Y}^i - {}^{l-1}\bar{Y}^i||$ is sufficiently small.

To aid in the calculation of the Jacobian, we define G^i_j as

$$ G^i = \frac{\partial F^i_j}{\partial Y^i_j}. \tag{A.29} $$

Therefore, in the special case when $\vec{F}_{body}$ and $\vec{Q}_{body}$ are zero vectors and $\mathbf{B}$ is constant we have

$$ G^i_j = \begin{bmatrix} -\tilde{\kappa}^i_j & -\tilde{\hat{t}} & \tilde{v}^i_j & \Theta \\ \Theta & -\tilde{\kappa}^i_j & \tilde{\omega}^i_j & \Theta \\ \Theta & -\widetilde{\left(\mathbf{I}\omega^i_j\right)} + \tilde{\omega}^i_j \mathbf{I} & \widetilde{\left(\mathbf{B}(\kappa^i_j - \kappa_o)\right)} - \tilde{\kappa}^i_j \mathbf{B} & -\tilde{\hat{t}} \\ m\tilde{\omega}^i_j & -m\tilde{v}^i_j & \tilde{f}^i_j & -\tilde{\kappa}^i_j \end{bmatrix}. \tag{A.30} $$

Here, the $\tilde{()}$ operator generates a skew symmetric matrix, such that for any pair of 3-vectors $\vec{a}$ and $\vec{b}$, $\vec{a} \times \vec{b} = \tilde{a}\vec{b}$. Now we can define $\bar{G}^i$ (similar to (A.16) and (A.17)) as

$$ \bar{G}^i = \frac{\partial \bar{F}^i}{\partial \bar{Y}^i} = \begin{bmatrix} G^i_1 & G^i_2 & 0 & \cdots & 0 \\ 0 & G^i_2 & G^i_3 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & G^i_{N_s-1} & G^i_{N_s} \\ \frac{\partial c^i}{\partial Y^i_1} & 0 & \cdots & 0 & \frac{\partial c^i}{\partial Y^i_{N_s}} \end{bmatrix}. \tag{A.31} $$

Finally, this expression is evaluated at ${}^{l}\bar{Y}^i$ and substituted into (A.27).

During the algorithm we also compute the position $\vec{R}$ and orientation of the rod as a function of the arclength coordinate (s). (Our algorithm here is based on integrating curvature, however we could instead base it on integrating velocity and angular velocity; see Section 6.2.) To do so we first define

$$\vec{\theta}^i_j = \frac{\Delta s}{2}\left\{\vec{\kappa}^i_{j-1} + \vec{\kappa}^i_j\right\}. \tag{A.32}$$

Next we calculate

$$\mathbf{L}^i_j = \exp\left\{-\tilde{\theta}^i_j\right\}\mathbf{L}^i_{j-1}, \tag{A.33}$$

where $\exp\{\}$ is the matrix exponential operator and $\mathbf{L}^i_j$ is the direction cosine matrix

$$\mathbf{L}^i_j = \begin{bmatrix} \mathbf{a}^i_{j1}\cdot\mathbf{e}_1 & \mathbf{a}^i_{j1}\cdot\mathbf{e}_2 & \mathbf{a}^i_{j1}\cdot\mathbf{e}_3 \\ \mathbf{a}^i_{j2}\cdot\mathbf{e}_1 & \mathbf{a}^i_{p2}\cdot\mathbf{e}_2 & \mathbf{a}^i_{j2}\cdot\mathbf{e}_3 \\ \mathbf{a}^i_{j3}\cdot\mathbf{e}_1 & \mathbf{a}^i_{j3}\cdot\mathbf{e}_2 & \mathbf{a}^i_{j3}\cdot\mathbf{e}_3 \end{bmatrix}. \tag{A.34}$$

Here, $\mathbf{a}^i_j$ is a body fixed reference frame at spatial gridpoint j along the arclength. Its basis vectors are $\mathbf{a}^i_{j1}$, $\mathbf{a}^i_{j2}$, and $\mathbf{a}^i_{j3}$, with $\mathbf{a}^i_{j3} = \hat{t}^i_j$. Finally, we compute $\vec{R}^i_j$

$$\vec{R}^i_j = \vec{R}^i_{j-1} + \frac{\Delta s}{2}\left\{\hat{t}^i_{j-1} + \hat{t}^i_j\right\}. \tag{A.35}$$

During many of the simulations here, the end ($s = 0$) is held fixed such that $\vec{R}^i_1$ and $\mathbf{L}^i_1$ are prescribed and held constant. In general they could vary with time; in this case we integrate velocity and angular velocity following the development in Section 6.2 to determine $\vec{R}^i_1$ and $\mathbf{L}^i_1$.

To summarize our algorithm, we present an outline below.

- Set $i = 0$
- Define initial conditions $\bar{Y}^i$ and $\dot{\bar{Y}}^i$
- Loop in time with index i
    - Update i using $i = i + 1$
    - Reset Newton-Raphson index, $l = 0$
    - Calculate ${}^0\bar{Y}^i$, the first estimate of $\bar{Y}^i$, by using (A.26)
    - Loop for Newton-Raphson iterations
        - Update l using $l = l + 1$
        - Calculate ${}^{l-1}\vec{R}^i_j$ and ${}^{l-1}\mathbf{L}^i_j$ for all j using (A.32)-(A.35)
        - Update A_1, A_2, and A_3 when necessary (for example if the boundary conditions change)

        - Calculate ${}^{l-1}\bar{F}^i$
        - Calculate ${}^{l-1}\bar{J}^i$ using (A.27)
        - Solve (A.28) for ${}^{l}\bar{Y}^i$
        - Exit loop if ${}^{l}\bar{Y}^i$ is sufficiently converged, otherwise continue Newton-Raphson loop
    - Set $\bar{Y}^i = {}^{l}\bar{Y}^i$
    - Calculate $\dot{\bar{Y}}^i$ from (A.20)
    - Calculate $\vec{R}^i_j$ and $\mathbf{L}^i_j$ for all j using (A.32)-(A.35)
    - Exit loop when integrated through time sufficiently long
- Algorithm complete, post-process output

Appendix B

SHS Representation

In the Section 3.1, we introduce the Straight-Helical-Straight (SHS) representation for the intrinsic curvature of highly bent sequences containing an A-tract domain (helical supercoil) flanked by two straight linker domains. The definition of the SHS representation is provided in Fig. B.1. This definition employs four geometric parameters which describe the helical axis of ds-DNA: 1) L_s, the length of the straight domains, 2) L_t, the total length of the sequence, 3) κ, the principal curvature of the helical domain, and 4) τ, the geometric torsion of the helical domain. Recall that molecules within this family are distinguished by the two phasing parameters Θ_1 and Θ_2 which define the torsional phasing of the operators in the linker domains to the start of the helical (A-tract) domain.

The values of the four parameters (Table B.1), described above, provide an excellent approximation to the intrinsic curvature of the three bent Kahn sequences 11C12, 9C14, and 7C16 synthesized to date [70]. The least squares fit of the SHS representation to the helical axis of the unstressed, zero temperature conformation of the sequence 11C12 predicted using [109]. Figure 3.1 shows that the RMS position error between the helical axes of these two representations (Table B.2) for each of the three Kahn sequences is less than the radius of DNA (approximately 10 $\AA$) confirming that the intrinsic curvature of the bent sequences are well approximated with a single SHS curve using the parameters from Table B.1.

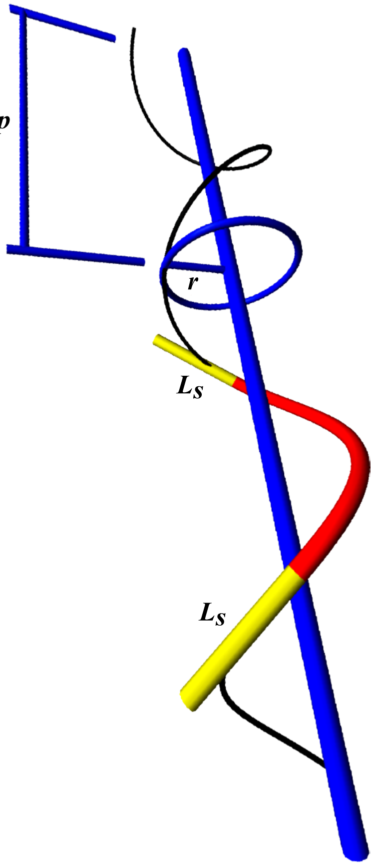


Figure B.1 The definition of the SHS representation. The helical domain (red curve) is a portion of a helix (black curve) with helical radius (r) and helical pitch $(2\pi p)$. The helical radius and pitch define the principal curvature and geometric torsion of the helical axis of the A-tract through the relations $r = \frac{\kappa}{\kappa^2+\tau^2}$ and $p = \frac{\tau}{\kappa^2+\tau^2}$. The two straight linker domains (yellow), with length L_s, are tangent to the helical domain and include the operators at their terminal ends.

Sequence	L_s ($\AA$)	L_t ($\AA$)	κ (rad/$\AA$)	τ (rad/$\AA$)
all Kahn sequences	94	468	0.0098	0.0074

Table B.1 Geometrical parameters defining the SHS representation. The parameters L_s, L_t, κ, and τ denote the length of each straight domain, the total contour length, the principal curvature of the helical domain, and the geometric torsion of the helical domain, respectively.

Sequence	RMS ($\AA$)
11C12	3.2
9C14	9.7
7C16	9.7

Table B.2 The RMS position errors between the helical axis of each bent sequence (stress-free, zero temperature conformation predicted by [109]) and the unified SHS representation introduced herein. The RMS error for each of the three sequences remains less than 10 $\AA$.

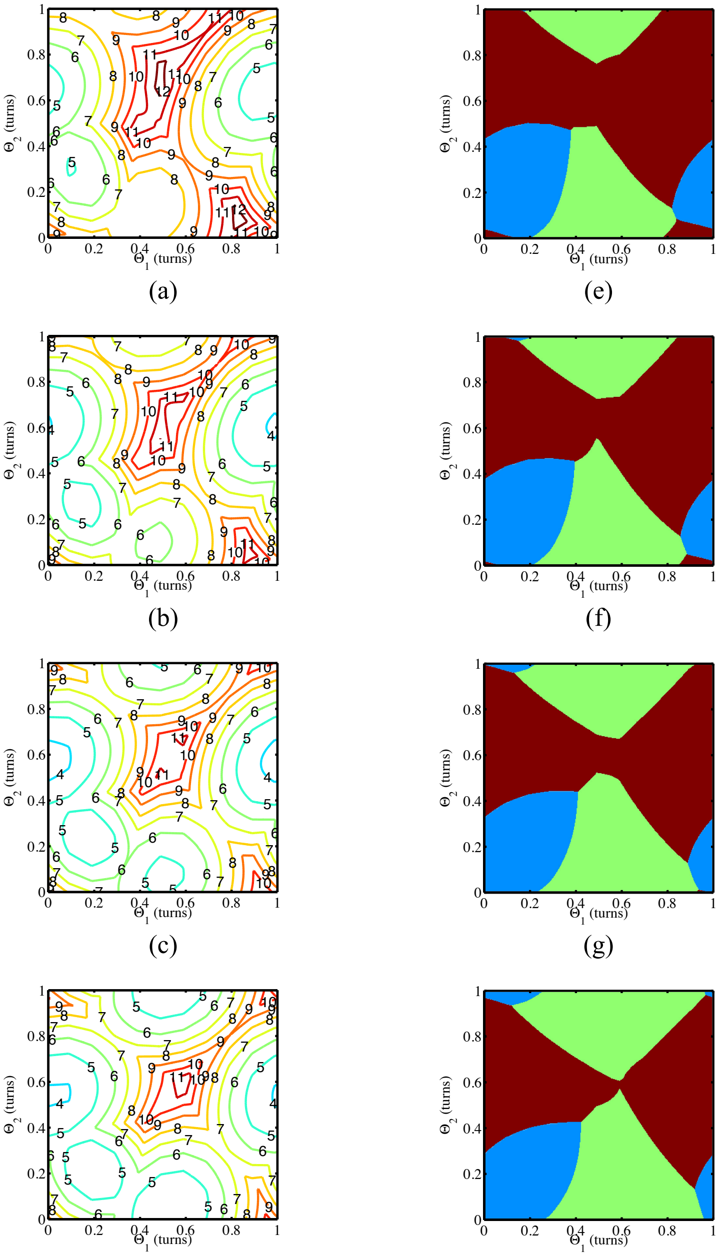
Appendix C

Sensitivity to SHS Representation

In this appendix we present preliminary results that illustrate the versatility of the rod model. Specifically, the rod model can efficiently compute new energy landscapes like that of Figure 3.3 resulting from changes in the values of the parameters describing the SHS representation.

In Figure C.1 we consider the effects on looping energy (kT) and binding topology of offsetting the helical bend of the SHS representation from its original centered location. Here the overall length of the rod and the length of the bend remain constant, but the lengths of the two straight linker domains are adjusted. The linker domains are adjusted by increments of 25%, such that (a) and (e), (b) and (f), (c) and (g), and (d) and (h) corresponding to linker lengths of 25% and 175%, 50% and 150%, 75% and 125%, and 100% and 100% of their original lengths respectively. Their original lengths (L_s) were provided in Table B.1.

In Figure C.2 we consider the effects on looping energy (kT) of the magnitude of the curvature of the intrinsic bend. Specifically, here the geometric torsion of the SHS rod is set to zero generating a planar SHS representation. The principal curvature is then adjusted from 50% in (a), to 100% in (b), and finally to 150% in (c) of its original value (κ of Table B.1).



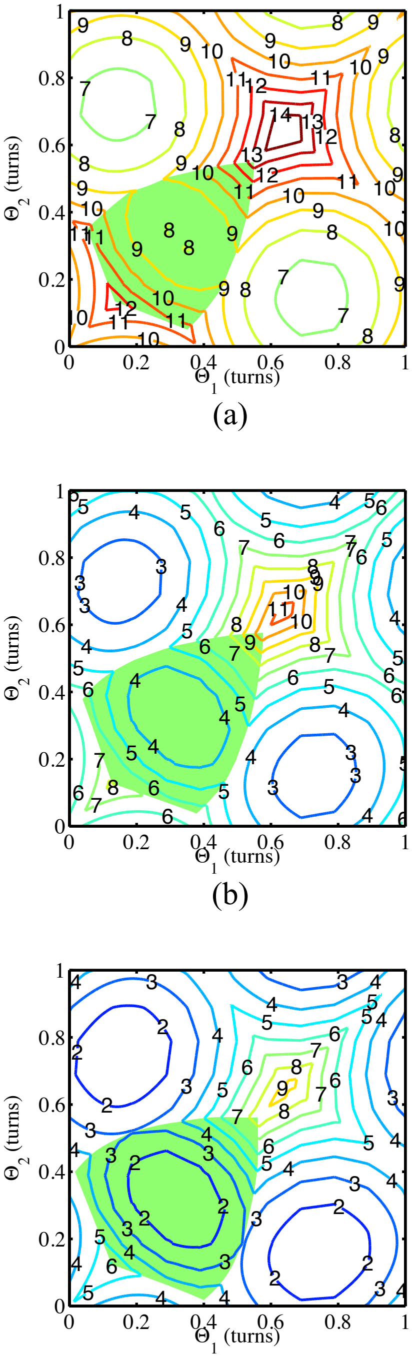
(a) (e)

(b) (f)

(c) (g)

(d) (h)

Figure C.1 Here we consider the effects on looping energy (kT) (a)-(d) and binding topology (e)-(h) of offsetting the helical bend of the SHS representation from its original centered location. For (e)-(h), the energetically preferred binding topology is highlighted, blue for **P1**, green for **A1** and brown for **A2**. Here the overall length of the rod and the length of the bend remain constant, but the lengths of the two straight linker domains are adjusted. The linker domains are adjusted by increments of 25%, such that (a) and (e), (b) and (f), (c) and (g), and (d) and (h) corresponding to linker lengths of 25% and 175%, 50% and 150%, 75% and 125%, and 100% and 100% of their original lengths (L_s of Table B.1) respectively.



(a)

(b)

(c)

Figure C.2 Here we consider the sensitivity of the looping energy (kT) to changes of the magnitude of the intrinsic curvature of the bent domain. Specifically, here the geometric torsion of the SHS rod is set to zero generating a planar SHS representation. The principal curvature is then adjusted from 50% in (a), to 100% in (b), and finally to 150% in (c) of its original value (κ of Table B.1). In addition, the preferred binding topology is **P1** for the green shaded regions and **A1** or **A2** for the unshaded regions.

Appendix D

In-plane Modes

Here we derive the in-plane normal modes for a minicircle with $\Delta Lk = 0$. Here, the three dimensional geometry of the minicircle is represented by N rigid segments of length l; see Fig. D.1. Our formulation is written in terms of angles (θ_i) measuring the orientation of each segment ($i = \{1, 2, ..., N\}$) and bend angles (β_i) measuring the angle between neighboring segments, such that for $i = \{2, 3..., N\}$

$$\beta_i = \theta_i - \theta_{i-1}, \tag{D.1}$$

and for $i = 1$

$$\beta_1 = \theta_1 - (\theta_N + 2\pi). \tag{D.2}$$

The following two constants on the angles θ_i enforce the closed geometry of a minicircle:

$$0 = \sum_{i=1}^{N} l \cos \theta_i, \tag{D.3}$$

$$0 = \sum_{i=1}^{N} l \sin \theta_i. \tag{D.4}$$

The bending strain energy is a function of the bending stiffness $2\alpha_B kT$ and the bend angle β_i at each vertex $i = \{1, 2, ..., N\}$,

$$V_{elastic} = \alpha_B \sum_{i=1}^{N} \beta_i^2. \tag{D.5}$$

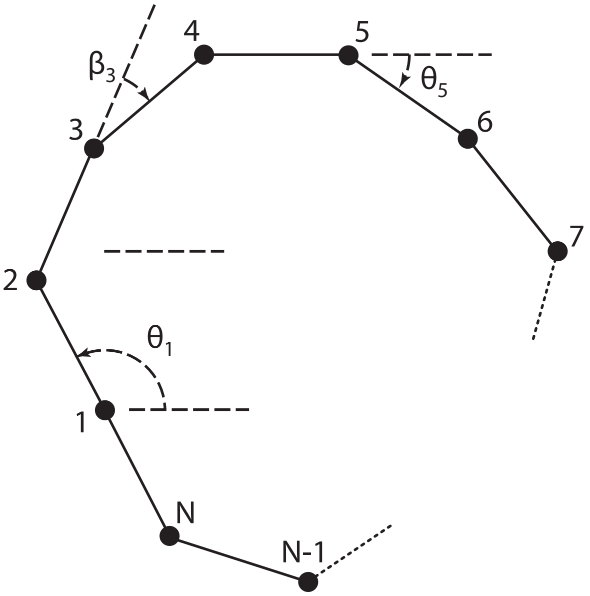


Figure D.1 Diagram of a minicircle represented by N rigid segments forming a closed polygon. Here, each vertex is numbered 1 to N. The angular orientation of each segment relative to the horizontal is measured by an angle θ_i. One example bend angle (β_3) and two example angles (θ_1 and θ_5) are shown.

Rewriting the elastic energy ($V_{elastic}$) in terms of θ_i results in:

$$V_{elastic} = \alpha_B \left((\theta_1 - 2\pi - \theta_N)^2 + \sum_{i=2}^{N} (\theta_i - \theta_{i-1})^2 \right). \tag{D.6}$$

We define V as the sum of the in-plane elastic bending energy and Lagrange multiplier terms (with multipliers μ_1 andμ_2) enforcing the constraints given by (D.3) and (D.4).

$$V = V_{elastic} - \mu_1 \sum_{i=1}^{N} l \cos\theta_i - \mu_2 \sum_{i=1}^{N} l \sin\theta_i \tag{D.7}$$

Considering the effect of applied moments to each segment M_i, the work done on the system is

$$W = \sum_{i=1}^{N} M_i \theta_i. \tag{D.8}$$

At equilibrium

$$\delta V + \delta W = \sum_{i=1}^{N} \left(\frac{\partial V}{\partial \theta_i} + \frac{\partial W}{\partial \theta_i} \right) \delta \theta_i + \frac{\partial V}{\partial \mu_1} \delta \mu_1 + \frac{\partial V}{\partial \mu_2} \delta \mu_2 = 0. \tag{D.9}$$

Consequently, all the coefficients to the variation terms ($\delta\theta_i$, $\delta\mu_1$ and $\delta\mu_2$) must be zero.

$$\frac{\partial V}{\partial \theta_i} + \frac{\partial W}{\partial \theta_i} = 0 \tag{D.10}$$

$$\frac{\partial V}{\partial \mu_1} = 0 \tag{D.11}$$

$$\frac{\partial V}{\partial \mu_2} = 0 \tag{D.12}$$

For convenience, we define:

$$Y = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_N & \mu_1 & \mu_2 \end{bmatrix}^T, \tag{D.13}$$

$$\tilde{M} = \frac{\partial W}{\partial Y} = \begin{bmatrix} M_1 & M_2 & M_3 & \cdots & M_N & 0 & 0 \end{bmatrix}^T, \tag{D.14}$$

$$\tilde{V} = \frac{\partial V}{\partial Y} = \begin{bmatrix} \frac{\partial V}{\partial \theta_1} & \cdots & \frac{\partial V}{\partial \theta_N} & \frac{\partial V}{\partial \mu_1} & \frac{\partial V}{\partial \mu_2} \end{bmatrix}^T, \tag{D.15}$$

$$\tilde{V} = \begin{bmatrix} 2\alpha_B kT\left(-2\pi - \theta_N + 2\theta_1 - \theta_2\right) + \mu_1 l \sin\theta_1 - \mu_2 l \cos\theta_1 \\ \vdots \\ 2\alpha_B kT\left(-\theta_{i-1} + 2\theta_i - \theta_{i+1}\right) + \mu_1 l \sin\theta_i - \mu_2 l \cos\theta_i \\ \vdots \\ 2\alpha_B kT\left(2\pi - \theta_{N-1} + 2\theta_N - \theta_1\right) + \mu_1 l \sin\theta_N - \mu_2 l \cos\theta_N \\ -\sum_{i=1}^{N} l \cos\theta_i \\ -\sum_{i=1}^{N} l \sin\theta_i \end{bmatrix}. \tag{D.16}$$

Therefore we can write

$$\tilde{V} + \tilde{M} = 0. \tag{D.17}$$

If we define Y_0 such that $\tilde{V}(Y_0) = 0$ ($\tilde{M} = 0$) and Taylor expand $\tilde{V}$ about Y_0, we have

$$\tilde{V}(Y_0 + \Delta Y) = \tilde{V}\big|_{Y_0} + \left.\frac{\partial \tilde{V}}{\partial Y}\right|_{Y_0} \Delta Y + O^2. \tag{D.18}$$

Therefore

$$0 \approx \tilde{V}\big|_{Y_0} + \left.\frac{\partial \tilde{V}}{\partial Y}\right|_{Y_0} \Delta Y + \Delta\tilde{M}. \tag{D.19}$$

Rearranging this expression we have

$$\left.\frac{\partial \tilde{V}}{\partial Y}\right|_{Y_0} \Delta Y = -\Delta\tilde{M}. \tag{D.20}$$

The eigenvalue problem that we consider requires the applied moments (the first N components of $\Delta\tilde{M}$) to be proportional to the angular deformations (the first N components of ΔY) for each mode. The constant of proportionality is the modal stiffness or eigenvalue λ^k of the k^{th} mode. We express the eigenvectors as $\begin{bmatrix} v^k \\ f^k \end{bmatrix}$ composed of the vector v^k that describes changes in θ_i and the vector f^k that describes changes in μ_1 and μ_2 respectively. This results in the following generalized eigenvalue problem

$$\left.\frac{\partial \tilde{V}}{\partial Y}\right|_{Y_0} \begin{bmatrix} v^k \\ f^k \end{bmatrix} = \lambda^k \begin{bmatrix} I_{N,N} & \Theta_{N,1} & \Theta_{N,1} \\ \Theta_{1,N} & 0 & 0 \\ \Theta_{1,N} & 0 & 0 \end{bmatrix} \begin{bmatrix} v^k \\ f^k \end{bmatrix}. \tag{D.21}$$

Here $I_{N,N}$ is the $N \times N$ identity matrix, $\Theta_{N,1}$ is the $N \times 1$ zero matrix, and $\Theta_{1,N}$ is the $1 \times N$ zero matrix.

Because the matrices forming the eigenvalue problem are singular, some eigenvalues are either 0 or infinite. For our particular eigenvalue problem there is only one 0 eigenvalue mode, which describes a rigid body rotation of the entire minicircle. The modes associated with infinite eigenvalues describe deformations that are excluded because of the constraints and therefore we neglect these modes. The remaining eigenvalues are positive and real. Normal modes associated with the lowest three non-zero eigenvalues are depicted in Fig. D.2. As a result of the closed circular geometry, each eigenvalue is two-fold degenerate; the two eigenvectors ($v^{k,1}$ and $v^{k,2}$) differ only in their 'phase' around the circle. We choose the pairs of eigenvectors such that $v^{k,1}$ and $v^{k,2}$ are orthonormal. From these normal modes ($v^{k,1}$ and $v^{k,2}$) in terms of changes in angles θ_i, we define $w^{k,1}$ and $w^{k,2}$ as the corresponding normal modes in terms of changes in bend angles by exercising the following equations for

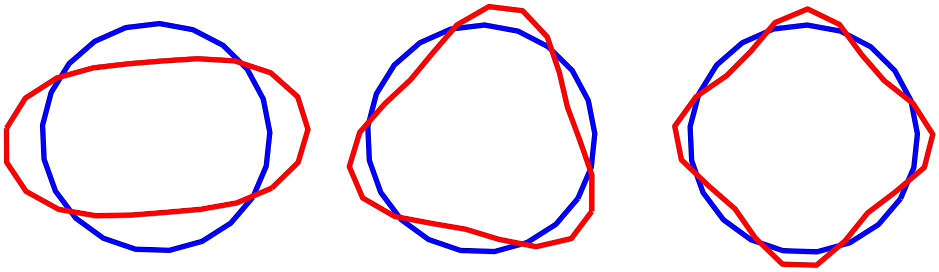


Figure D.2 Three normal modes (red) associated with the lowest three non-zero eigenvalues (modal stiffnesses). The relaxed equilibria are also illustrated (blue).

$i = \{2, 3, ..., N\}$

$$w_i^k = v_i^k - v_{i-1}^k, \tag{D.22}$$

and for $i = 1$

$$w_1^k = v_1^k - v_N^k. \tag{D.23}$$

To calculate the modal energy we first calculate the modal amplitude (α^k) using

$$\alpha^k = v^k \cdot (\beta - \bar{\beta}), \tag{D.24}$$

where $\beta - \bar{\beta}$ is an N-vector of changes in bend angle from equilibrium to the current deformed state, and v^k is an N-vector for the k^{th} normal mode. The modal energy (E^k) for the k^{th} mode is quadratic in α^k,

$$E^k = \frac{1}{2}\lambda^k \left(\alpha^k\right)^2. \tag{D.25}$$

Bibliography

[1] Amzallag, A., C. Vaillant, M. Jacob, M. Unser, J. Bednar, J. D. Kahn, J. Dubochet, A. Stasiak, and J. H. Maddocks. 2006. 3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy. *Nucl. Acids Res.* 34:e125.

[2] Balaeff, A., C. R. Koudella, L. Mahadevan, and K. Schulten. 2004. Modelling DNA loops using continuum and statistical mechanics. *Philos. T. Roy. Soc. A.* 362:1355–1371.

[3] Balaeff, A., L. Mahadevan, and K. Schulten. 2004. Structural basis for cooperative DNA binding by CAP and Lac repressor. *Structure* 12:123–132.

[4] Balaeff, A., L. Mahadevan, and K. Schulten. 2006. Modeling DNA loops using the theory of elasticity. *Phys. Rev. E* 73:031919.

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

[6] Becker, N. A., J. D. Kahn, and L. J. Maher. 2005. Bacterial repression loops require enhanced DNA flexibility. *J. Mol. Biol.* 349:716–730.

[7] Been, M. D., R. R. Burgess, and J. J. Champoux. 1984. Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. *Nucl. Acids Res.* 12:3097–3114.

[8] Bellomy, G. R., M. C. Mossing, and M. T. Record. 1988. Physical properties of DNA in vivo as probed by the length dependence of the lac operator looping process. *Biochemistry* 27:3900–3906.

[9] Beveridge, D. L., G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(C(p)G) steps. *Biophys. J.* 87:3799–3813.

[10] Bryant, Z., M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante. 2003. Structural transitions and elasticity from torque measurements on DNA. *Nature* 424:338–341.

[11] Bustamante, C., Z. Bryant, and S. B. Smith. 2003. Ten years of tension: single-molecule DNA mechanics. *Nature* 421:423–427.

[12] Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. *Science* 265:1599–1600.

[13] Calladine, C. R., H. R. Drew, B. F. Luisi, and A. A. Travers. 2004. Understanding DNA: the molecule and how it works. 3rd edition. Elsevier Academic Press.

[14] Champoux, J. J. 2001. DNA topoisomerases: Structure, function, and mechanism. *Annu. Rev. Biochem.* 70:369–413.

[15] Charlier, M., J. C. Maurizot, and G. Zaccai. 1980. Neutron scattering studies of lac repressor. *Nature* 286:423–425.

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

[17] Chirico, G., and J. Langowski. 1994. Kinetics of DNA supercoiling studied by Brownian dynamics simulation. *Biopolymers* 34:415–433.

[18] Chung, J., and G. M. Hulbert. 1993. A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized-alpha method. *J. Appl. Mech.* 60:371–375.

[19] Clauvelin, N., B. Audoly, and S. Neukirch. 2009/5/6/. Elasticity and Electrostatics of Plectonemic DNA. *Biophys. J.* 96:3716–3723.

[20] Coleman, B. D., W. K. Olson, and D. Swigon. 2003. Theory of sequence-dependent DNA elasticity. *J. Chem. Phys.* 118:7127–7140.

[21] Coleman, B. D., D. Swigon, and I. Tobias. 2000. Elastic stability of DNA configurations. II. Supercoiled plasmids with self-contact. *Phys. Rev. E* 61:759–770.

[22] Crut, A., D. A. Koster, R. Seidel, C. H. Wiggins, and N. H. Dekker. 2007. Fast dynamics of supercoiled DNA revealed by single-molecule experiments. *Proc. Natl. Acad. Sci.* 104:11957–11962.

[23] Delrow, J. J., J. A. Gebe, and J. M. Schurr. 1997. Comparison of hard-cylinder and screened Coulomb interactions in the modeling of supercoiled DNAs. *Biopolymers* 42:455–470.

[24] Demurtas, D., A. Amzallag, E. J. Rawdon, J. H. Maddocks, J. Dubochet, and A. Stasiak. 2009. Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles. *Nucl. Acids Res.* 37:2882–2893.

[25] Drak, J., and D. M. Crothers. 1991. Helical repeat and chirality effects on DNA gel electrophoretic mobility. *Proc. Natl. Acad. Sci.* 88:3074–3078.

[26] Drlica, K. 1984. Biology of bacterial deoxyribonucleic acid topoisomerases. *Microbiol. Mol. Biol. R.* 48:273–289.

[27] Dubochet, J., M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz. 1988. Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21:129–228.

[28] Echols, H. 1990. Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination. *J. Biol. Chem.* 265:14697–14700.

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

[30] Franklin, R. E., and R. G. Gosling. 1953. Molecular configuration in sodium thymonucleate. *Nature* 171:740–741.

[31] Friedman, A. M., T. O. Fischmann, and T. A. Steitz. 1995. Crystal-structure of lac repressor core tetramer and its implications for DNA looping. *Science* 268:1721–1727.

[32] Fuller, F. B. 1971. The writhing number of a space curve. *Proc. Natl. Acad. Sci.* 68:815–819.

[33] Geanacopoulos, M., G. Vasmatzis, V. B. Zhurkin, and S. Adhya. 2001. Gal repressosome contains an antiparallel DNA loop. *Nature Struct. Biol.* 8:432–436.

[34] Gellert, M. 1981. DNA topoisomerases. *Annu. Rev. Biochem.* 50:879–910.

[35] Gobat, J. I., and M. A. Grosenbaugh. 2001. Application of the generalized-alpha method to the time integration of the cable dynamics equations. *Comput. Method. Appl. M.* 190:4817–4829.

[36] Goyal, S. 2006. A dynamic rod model to simulate mechanics of cables and DNA. Ph.D. thesis, University of Michigan.

[37] Goyal, S., T. Lillian, S. Blumberg, J.-C. Meiners, E. Meyhöfer, and N. C. Perkins. 2007. Intrinsic curvature of DNA influences LacR-mediated looping. *Biophys. J.* 93:4342–4359.

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

[39] Goyal, S., N. C. Perkins, and C. L. Lee. 2008. Non-linear dynamic intertwining of rods with self-contact. *Int. J. Non Linear Mech.* 43:65–73.

[40] Hagerman, P. J. 1988. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* 17:265–286.

[41] Han, L., H. G. Garcia, S. Blumberg, K. B. Towles, J. F. Beausang, P. C. Nelson, and R. Phillips. 2009. Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS ONE* 4:e5621.

[42] Harada, Y., O. Ohara, A. Takatsuki, H. Itoh, N. Shimamoto, and K. Kinosita. 2001. Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase. *Nature* 409:113–115.

[43] Healey, T. J. 2002. Material symmetry and chirality in nonlinearly elastic rods. *Math. Mech. Solids.* 7:405–420.

[44] Howard, J. 2001. Mechanics of motor proteins and the cytoskeleton. Sinauer Associates, Sunderland, MA.

[45] Huang, J., T. Schlick, and A. Vologodskii. 2001. Dynamics of site juxtaposition in supercoiled DNA. *Proc. Natl. Acad. Sci.* 98:968–973.

[46] Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *J. Mol. Graphics* 14:33–38.

[47] Jacob, M., T. Blu, C. Vaillant, J. H. Maddocks, and M. Unser. 2006. 3-D shape estimation of DNA molecules from stereo cryo-electron micro-graphs using a projection-steerable snake. *IEEE T. Image Process.* 15:214–227.

[48] Jian, H., T. Schlick, and A. Vologodskii. 1998. Internal motion of supercoiled DNA: Brownian dynamics simulations of site juxtaposition. *J. Mol. Biol.* 284:287–296.

[49] Kabsch, W., C. Sander, and E. N. Trifonov. 1982. The ten helical twist angles of B-DNA. *Nucl. Acids Res.* 10:1097–1104.

[50] Kahn, J. D., R. Cheong, L. M. Edelman, R. A. Mehta, and M. A. Morgan. 2006. Flexibility and control of protein-DNA loops. *Biophysical Reviews and Letters* 1:327 – 341.

[51] Klapper, I. 1996. Biological applications of the dynamics of twisted elastic rods. *J. Comput. Phys.* 125:325–337.

[52] Klenin, K., and J. Langowski. 2000. Computation of writhe in modeling of supercoiled DNA. *Biopolymers* 54:307–317.

[53] Klenin, K., H. Merlitz, and J. Langowski. 1998. A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.* 74:780–788.

[54] Koster, D. A., V. Croquette, C. Dekker, S. Shuman, and N. H. Dekker. 2005. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature* 434:671–674.

[55] Koster, D. A., F. Czerwinski, L. Halby, A. Crut, P. Vekhoff, K. Palle, P. B. Arimondo, and N. H. Dekker. 2008. Single-molecule observations of topotecan-mediated TopIB activity at a unique DNA sequence. *Nucl. Acids Res.* 36:2301–2310.

[56] Koster, D. A., K. Palle, E. S. M. Bot, M.-A. Bjornsti, and N. H. Dekker. 2007. Antitumour drugs impede DNA uncoiling by topoisomerase I. *Nature* 448:213–217.

[57] Krämer, H., M. Amouyal, A. Nordheim, and B. Müller-Hill. 1988. DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor. *EMBO J.* 7:547–556.

[58] Krämer, H., M. Niemöller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Müller-Hill. 1987. Lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.* 6:1481–1491.

[59] Lankas, F., R. Lavery, and J. H. Maddocks. 2006. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* 14:1527–1534.

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

[61] Lillian, T. D., S. Goyal, J. D. Kahn, E. Meyhöfer, and N. C. Perkins. 2008. Computational analysis of looping of a large family of highly bent DNA by LacI. *Biophys. J.* 95:5832–5842.

[62] Lillian, T. D., and N. C. Perkins. 2009. Electrostatics and self contact in an elastic rod approximation for DNA. *Proc. ASME IDETC/CIE 2009* .

[63] Lillian, T. D., N. C. Perkins, and S. Goyal. 2008. Computational elastic rod model applied to DNA looping. *Proc. ASME IDETC/CIE 2007* 5:1449–1456.

[64] Lionberger, T. A. 2010. Regulating gene expression through DNA mechanics: Tightly looped DNA represses transcription. Ph.D. thesis, University of Michigan.

[65] Manning, R. S., J. H. Maddocks, and J. D. Kahn. 1996. A continuum rod model of sequence-dependent DNA structure. *J. Chem. Phys.* 105:5626–5646.

[66] Marko, J. F. 1997. Stretching must twist DNA. *Europhys. Lett.* 38:183–188.

[67] Marko, J. F. 2007. Torque and dynamics of linking number relaxation in stretched supercoiled DNA. *Phys. Rev. E* 76:021926.

[68] Matthews, K. S. 1992. DNA looping. *Microbiol. Mol. Biol. Rev.* 56:123–136.

[69] Mckay, D. B., C. A. Pickover, and T. A. Steitz. 1982. Escherichia coli lac repressor is elongated with its operator DNA binding domains located at both ends. *J. Mol. Biol.* 156:175–183.

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

[71] Merlitz, H., K. Rippe, K. V. Klenin, and J. Langowski. 1998. Looping dynamics of linear DNA molecules and the effect of DNA curvature: A study by Brownian dynamics simulation. *Biophys. J.* 74:773–779.

[72] Mielke, S. P., W. H. Fink, V. V. Krishnan, N. Gronbech-Jensen, and C. J. Benham. 2004. Transcription-driven twin supercoiling of a DNA loop: A Brownian dynamics study. *J. Chem. Phys.* 121:8104–8112.

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

[74] Müller, J., S. Oehler, and B. Müller-Hill. 1996. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.* 257:21–29.

[75] Munteanu, M. G., K. Vlahovicek, S. Parthasarathy, I. Simon, and S. Pongor. 1998. Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.* 23:341–347.

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

[77] Olson, W. K. 1996. Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* 6:242–256.

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

[79] Purohit, P. K. 2008. Plectoneme formation in twisted fluctuating rods. *J. Mech. Phys. Solids.* 56:1715–1729.

[80] Purohit, P. K., and P. C. Nelson. 2006. Effect of supercoiling on formation of protein-mediated DNA loops. *Phys. Rev. E* 74:061907.

[81] Redinbo, M. R., L. Stewart, P. Kuhn, J. J. Champoux, and W. G. J. Hol. 1998. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* 279:1504–1513.

[82] Rossetto, V., and A. C. Maggs. 2003. Writhing geometry of open DNA. *J. Chem. Phys.* 118:9864–9874.

[83] Ruben, G. C., and T. B. Roos. 1997. Conformation of Lac repressor tetramer in solution, bound and unbound to operator DNA. *Microsc. Res. Tech.* 36:400–416.

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

[85] Saiz, L., J. M. Rubi, and J. M. G. Vilar. 2005. Inferring the in vivo looping properties of DNA. *Proc. Natl. Acad. Sci.* 102:17642–17645.

[86] Saiz, L., and J. M. Vilar. 2007. Multilevel deconstruction of the in vivo behavior of looped DNA-protein complexes. *PLoS ONE* 2:e355.

[87] Saiz, L., and J. M. G. Vilar. 2006. In vivo evidence of alternative loop geometries in DNA-protein complexes. *arXiv:q-bio.BM/0602012* .

[88] Schleif, R. 1992. DNA looping. *Annu. Rev. Biochem.* 61:199–223.

[89] Schlick, T. 1995. Modeling superhelical DNA: recent analytical and dynamic approaches. *Curr. Opin. Struct. Biol.* 5:245–262.

[90] Semsey, S., K. Virnik, and S. Adhya. 2005. A gamut of loops: meandering DNA. *Trends Biochem. Sci.* 30:334–341.

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

[92] Shore, D., and R. L. Baldwin. 1983. Energetics of DNA twisting: I. Relation between twist and cyclization probability. *J. Mol. Biol.* 170:957–981.

[93] Smith, S. B., L. Finzi, and C. Bustamante. 1992. Direct mechanical measurements of the elasticity of single DNA-molecules by using magnetic beads. *Science* 258:1122–1126.

[94] Steitz, T. A., T. J. Richmond, D. Wise, and D. Engelman. 1974. The lac repressor protein: molecular shape, subunit structure, and proposed model for operator interaction based on structural studies of microcrystals. *Proc. Natl. Acad. Sci.* 71:593–597.

[95] Stewart, L., M. R. Redinbo, X. Y. Qiu, W. G. J. Hol, and J. J. Champoux. 1998. A model for the mechanism of human topoisomerase I. *Science* 279:1534–1541.

[96] Strick, T., J. F. Allemand, V. Croquette, and D. Bensimon. 2000. Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol.* 74:115–140.

[97] Strick, T. R., J. F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette. 1996. The elasticity of a single supercoiled DNA molecule. *Science* 271:1835–1837.

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

[99] Swigon, D., and W. K. Olson. 2008. Mesoscale modeling of multi-protein-DNA assemblies: The role of the catabolic activator protein in Lac-repressor-mediated looping. *Int. J. Non Linear Mech.* 43:1082–1093.

[100] Taraban, M., H. Zhan, A. E. Whitten, D. B. Langley, K. S. Matthews, L. Swint-Kruse, and J. Trewhella. 2008. Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. *J. Mol. Biol.* 376:466–481.

[101] Timoshenko, S. P., and J. M. Gere. 1961. Theory of elastic stability. McGraw-Hill.

[102] Tobias, I., D. Swigon, and B. D. Coleman. 2000. Elastic stability of DNA configurations. I. General theory. *Phys. Rev. E* 61:747–758.

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

[104] van der Heijden, G. H. M., S. Neukirch, V. G. A. Goss, and J. M. T. Thompson. 2003. Instability and self-contact phenomena in the writhing of clamped rods. *Int. J. Mech. Sci.* 45:161–196.

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

[106] Villa, E., A. Balaeff, L. Mahadevan, and K. Schulten. 2004. Multiscale method for simulating protein-DNA complexes. *Multiscale Model Simul.* 2:527–553.

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

[108] Virnik, K., Y. L. Lyubchenko, M. A. Karymov, P. Dahlgren, M. Y. Tolstorukov, S. Semsey, V. B. Zhurkin, and S. Adhya. 2003. "Antiparallel" DNA loop in gal repressosome visualized by atomic force microscopy. *J. Mol. Biol.* 334:53–63.

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

[110] Vologodskii, A., and N. Cozzarelli. 1995. Modeling of long-range electrostatic interactions in DNA. *Biopolymers* 35:289–296.

[111] Vologodskii, A., and N. R. Cozzarelli. 1996. Effect of supercoiling on the juxtaposition and relative orientation of DNA sites. *Biophys. J.* 70:2548–2556.

[112] Vologodskii, A. V., and J. F. Marko. 1997. Extension of torsionally stressed DNA by external force. *Biophys. J.* 73:123–132.

[113] Wada, H., and R. R. Netz. 2009. Plectoneme creation reduces the rotational friction of a polymer. *Europhys. Lett.* 87:38001 (6pp).

[114] Watson, J. D., and F. H. C. Crick. 1953. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature* 171:737–738.

[115] Wereszczynski, J. M. 2008. Simulations of nucleic acids under stress, in solution, and complexed to proteins. Ph.D. thesis, University of Michigan.

[116] Wilson, D. P., T. Lillian, S. Goyal, A. V. Tkachenko, N. C. Perkins, and J.-C. Meiners. 2007. Understanding the role of thermal fluctuations in DNA looping. *In* Proc. SPIE, S. M. Bezrukov, editor, volume 6602. SPIE, 660208.

[117] Wong, O. K., M. Guthold, D. A. Erie, and J. Gelles. 2008. Interconvertible Lac repressor-DNA loops revealed by single-molecule experiments. *PLoS Biology* 6.

[118] Yang, Z., Z. Haijun, and O.-Y. Zhong-can. 2000. Monte Carlo implementation of supercoiled double-stranded DNA. *Biophys. J.* 78:1979–1987.

[119] Zhang, Y., A. E. McEwen, D. M. Crothers, and S. D. Levene. 2006. Statistical-mechanical theory of DNA looping. *Biophys. J.* 90:1903–1912.

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

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

[122] Zhu, Q., J. Zeng, M. S. Triantafyllou, and D. K. P. Yue. 2006. Direct numerical simulation of single-molecule DNA by cable dynamics. *J. Microelectromech. Syst.* 15:1078–1087.