

**Dynamics of DNA knots and links**

by

**Abby Pecoske Fulton**

BA, Carthage College, 2013

MS, Oregon State University, 2015

Submitted to the Graduate Faculty of  
the Dietrich School of Arts and Sciences

in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH  
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Abby Pecoske Fulton

It was defended on

July 13, 2021

and approved by

David Swigon, Department of Mathematics University of Pittsburgh

Javier Arsuaga, U.C. Davis Department of Molecular and Cellular Biology and Department  
of Mathematics

Armin Schikorra, University of Pittsburgh Department of Mathematics

Anna Vainchtein, University of Pittsburgh Department of Mathematics

Dissertation Director: David Swigon, Department of Mathematics University of Pittsburgh

Copyright © by Abby Pecoske Fulton  
2021

# Dynamics of DNA knots and links

Abby Pecoske Fulton, PhD

University of Pittsburgh, 2021

The goal of this work is to describe the dynamics of DNA knots and links in an ionized fluid. To do so, we employ three models: 1. The Generalized Immersed Boundary (GIB) method, which is a deterministic method that accounts for the fluid, structure interaction of an immersed DNA molecule in an ionized fluid; 2. The Stochastic Generalized Boundary (SGIB) Method, which is an extension of the GIB method that also takes into account the random thermal fluctuations within the fluid; 3. The Sequence Dependent SGIB method, which is a new extension of the SGIB method that accounts for the elastic properties of a specified DNA sequence. Using the GIB and SGIB methods, we explore the energy landscape of a closed DNA segment in a trefoil knot configuration. We first analyze the symmetry of stable knotted equilibrium configurations, approximate saddle configurations, and examine elastic energy throughout the deterministic process. We then use the SGIB method to model DNA knot dynamics as a continuous time Markov chain. We classify and find boundaries within the energy landscape using Procrustes distance. Finally, we obtain a steady state distribution for the Markov process given a fixed linking number and compare this to the Gibb's distribution from energy estimates obtained from the GIB method. Lastly, using the SD-SGIB method, we also explore the effects of sequence dependence in the formation of kinetoplast DNA (kDNA), which has a chainmail-like linked DNA structure. We do so by finding the distribution of centroid distances of two kDNA minicircles.

---

This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided.

## Table of contents

<b>Preface</b> . . . . .	xi
<b>1.0 Introduction</b> . . . . .	1
1.1 Background . . . . .	2
1.2 DNA modeling assumptions . . . . .	5
1.3 DNA topology . . . . .	6
<b>2.0 Equilibrium configurations of DNA knots</b> . . . . .	11
2.1 Introduction . . . . .	11
2.2 Generalized Immersed Boundary Method . . . . .	12
2.3 Methods . . . . .	18
2.4 Results . . . . .	21
2.4.1 Effects of $Lk_{rel}$ and $Cs$ on final configurations . . . . .	23
2.4.2 Dynamics of transitions . . . . .	25
2.4.3 Phase two equilibria configurations . . . . .	28
2.5 Conclusion . . . . .	32
<b>3.0 Stochastic dynamics of DNA knots</b> . . . . .	34
3.1 Introduction . . . . .	34
3.2 Stochastic Generalized Immersed Boundary Method . . . . .	35
3.3 Methods . . . . .	38
3.3.1 Kendall shape space and Procrustes analysis . . . . .	38
3.3.2 Classification of stochastic configurations . . . . .	39
3.3.3 Stochastic knots as a Markov process . . . . .	41
3.4 Results . . . . .	42
3.4.1 Procrustes analysis . . . . .	42
3.4.2 Markov process of stochastic knots . . . . .	49
3.5 Conclusion . . . . .	58
<b>4.0 Dynamics of DNA links</b> . . . . .	60

4.1 Introduction . . . . .	60
4.2 Sequence Dependent Stochastic Generalized Immersed Boundary Method . .	61
4.2.1 Dynamical sampling of minicircle centroid distances . . . . .	65
4.3 Kinetoplast DNA sequence dependent effects . . . . .	67
4.3.1 Sequence Dependent SGIB Method . . . . .	69
4.4 Conclusion . . . . .	76
<b>5.0 Conclusion . . . . .</b>	<b>77</b>
<b>Appendix A. Mean holding times and transition rates . . . . .</b>	<b>79</b>
<b>Appendix B. SD-SGIB parameters . . . . .</b>	<b>88</b>
<b>Appendix C. Kinetoplast DNA sequences . . . . .</b>	<b>92</b>
<b>Bibliography . . . . .</b>	<b>95</b>

## List of tables

1	GIB computational parameters . . . . .	22
2	SGIB computational parameters . . . . .	40
3	Knot configuration equivalence class representatives for positive and negative $Lk_{rel}$ . . . . .	49
4	kDNA computational parameters . . . . .	68
5	Sequence dependent elasticity parameters . . . . .	68
6	Elastic coefficients for $f_{ij}$ representing twisting and bending moduli in units $kT/deg^2$	90
7	Elastic coefficients for $g_{ij}$ representing the coupling between bending, shearing, and slide in units $kT/deg/\text{\AA}$ . . . . .	90
8	Elastic coefficients for $h_{ij}$ representing shift, slide and rise energies in units $kT/\text{\AA}^2$	91

## List of figures

1	Left: DNA [80] Center: Cosserat rod with central axis, $C_1$ , exterior axis, $C_2$ , and frame $\{\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3\}$ Right: Sequence-dependent discrete base pair model. . . . .	7
2	Electrostatic charges along the phosphate backbone are moved to the central axis of the DNA rod (Center) and discrete base pair sequence (Right) [80]. . . . .	7
3	Linking number $Lk(C_1, C_2) = 6$ , $Wr(C_1) = 0$ , $Tw(C_1, C_2) = 6$ for two curves, the central axis, $C_1$ and the winding curve $C_2$ in red. . . . .	9
4	Writhe of 1 and 2 for an oriented curve, with signed crossing convention. . . . .	10
5	Twist of 0, $\frac{1}{2}$ , and 1 for two oriented curves, $C_1$ and $C_2$ . . . . .	10
6	Rod mechanics . . . . .	14
7	DNA starting configuration . . . . .	22
8	Phase one stable equilibrium configurations parametrized by $Lk_{rel} = \pm 1, \pm 2, \dots \pm 10$ for $Cs = 0.1M$ and $Cs = 0.01M$ . . . . .	24
9	Dynamic progression of knot configuration starting from the standard trefoil parametrization for $Lk_{rel} = -9$ , $Cs = 0.1M$ . Timestamps: from left to right and top to bottom frames: $0.0003\mu s$ , $0.4203\mu s$ , $0.6903\mu s$ , $1.2003\mu s$ , $2.7303\mu s$ , $5.0403\mu s$ , $6.9603\mu s$ , $8.5803\mu s$ , $20.4303\mu s$ . . . . .	26
10	Elastic energies throughout dynamic simulation for $Lk_{rel} = -7$ and $Cs = 0.1M$ Energy values are in units of kT, and velocities in $\mu m/s$ . From the top left to bottom right panels: 1. bending energy, 2. twist energy, 3. shear energy, 4. stretch energy, 5. electrostatic energy, 6. excess twist, 7. fluid kinetic energy, 8. maximum velocity of the DNA and fluid, 9. minimum distance between points along the DNA. . . . .	27
11	Saddle configuration for various linking numbers, with $Cs = 0.1M$ . From left to right and top to bottom, displayed are saddle node configurations for $Lk_{rel} = -4, -7, -8, -9, -10, 5, 8, 9, 10$ . . . . .	29

12	Phase two stable equilibrium configurations for $Lk_{rel} = -1, -2, \dots, -10$ with corresponding energy values in kT. Configurations inside the black boxes have $D_3$ symmetry, those inside green boxes have $C_2$ symmetry, while the rest are asymmetric.	30
13	Phase two stable equilibrium configurations for $Lk_{rel} = 1, 2, \dots, 10$ with corresponding energy values in kT. Configurations inside the black boxes have $D_3$ symmetry, those inside green boxes have $C_2$ symmetry, while the rest are asymmetric.	31
14	3-Dimensional projection of knot configurations in Kendall shape space for negative $Lk_{rel}$ .	44
15	Dendrogram for negative $Lk_{rel}$ knot configurations representing the minimum Procrustes distance between equilibria.	45
16	3-Dimensional projection of knot configurations in Kendall shape space for positive $Lk_{rel}$ .	46
17	Dendrogram for positive $Lk_{rel}$ knot configurations representing the minimum Procrustes distance between equilibria.	47
18	Top: $Lk_{rel} = -8$ stochastic knot configurations with corresponding $d_{min}$ equilibrium configuration classifications. Bottom: traces of $10 \times d_{min}$ to each knot class representative with state classification denoted by minimizing state.	50
19	Transition rates for $Lk_{rel} = -3$ .	55
20	Top panels: Traces of $10 \times d_{min}$ with minimizing state starting at the state $C_{-3,-3}$ on the left and $C_{-3,-9}$ on the right. Bottom left panel: mean holding times in seconds for states $C_{-3,-3}$ and $C_{-3,-9}$ . Bottom right panel: Comparison of stationary distributions $\pi$ (blue) and $\hat{\pi}$ (orange) for $Lk_{rel} = -3$ .	56
21	Stationary distribution $\pi$ (blue) and Gibb's distribution $\hat{\pi}$ (orange) for $Lk_{rel} = -3, -4, -5, -6, -7, 8, -9, -10$ from top the top left panel to the bottom right panel.	57
22	<i>T. brucei</i> centroid distance (microns) distribution and centroid distance time traces for 20 Hopf-linked minicircles.	70
23	<i>C. fasciculata</i> centroid distance (microns) distribution and centroid distance time traces for 20 Hopf-linked minicircles.	71
24	<i>C. fasciculata</i> dynamic progression illustrating sticking with corresponding distance plot.	73

25	Diffusion coefficient for <i>C. fasciculata</i> minicircles. Initial centroid distance of 0 microns (blue) and 0.1616 microns (orange) . . . . .	75
26	Mean holding times and transition rates for $Lk_{rel} = -4$ . . . . .	80
27	Mean holding times and transition rates for $Lk_{rel} = -5$ . . . . .	81
28	Mean holding times and transition rates for $Lk_{rel} = -6$ . . . . .	82
29	Mean holding times and transition rates for $Lk_{rel} = -7$ . . . . .	83
30	Mean holding times and transition rates for $Lk_{rel} = -8$ . . . . .	84
31	Mean holding times and transition rates for $Lk_{rel} = -9$ . . . . .	85
32	Mean holding times and transition rates for $Lk_{rel} = -10$ . . . . .	86
33	Mean holding times and transition rates for $Lk_{rel} = -4$ . . . . .	87

## Preface

Finishing my Ph.D. to become a mathematician was a lifelong dream that at many times, I doubted could actually happen. This work would not have been possible without the support of so many amazing people in my life who helped me every step of the way. I am eternally grateful, and you all hold a place in my heart.

Thank you to my advisor, David Swigon, for your support and encouragement since even before taking me on as a student. You personally guided me throughout the preliminary and comprehensive exam process and stayed kind and positive through many tears. I have greatly enjoyed our meetings throughout the years, and I will miss sharing stories of my hiking and biking adventures with you. Through your guidance and example, I have grown to be a better researcher and teacher, so thank you.

I was very fortunate to have a supportive committee: Javier Arsuaga, Armin Schikorra, and Anna Vainchtein throughout the overview and thesis writing process, and I have enjoyed working with all of you. My meetings with our kDNA collaborators, Javier Arsuaga and Pengyu Liu, were a great driving force and source of sanity throughout the pandemic.

Within the Pitt math department, I have had numerous supportive professors and colleagues. The entire math bio seminar group has taught me to confidently present my work, despite being terrified to give my first seminar talk. To my cohort at Pitt and friends passing through Pittsburgh: Nicole Vorderobermeier, Youngmin Park, Dan and Sam Ehrmann, Ben Letson, Michael McLaughlin, Pam and Joe Fiordilino, Pam Delgado, Ivan Ramirez, Derek Orr, and Astrid Berge, thank you for your unending encouragement and all of the fun times we shared in and out of the office. I need to specifically thank Marcello Codianni and Michael Lindsey. You two became my family away from home, kept me sane through the pandemic, and helped me solve and talk through so many prelim, comps, research, and teaching related problems.

Before graduate school at Pitt, I was fortunate enough to have many supportive mentors and programs to introduce me to the foundations of mathematics. I learned to passionately study and teach from my mentors in the Carthage College mathematics department: Mark

Snavely, Charlotte Chell, Aaron Trautwein, Erik Tou, and Erlan Wheeler. The Carleton College Summer Mathematics Program for Women introduced me to the rigor that would be expected of me as a graduate student and provided me a network of women that I am happy to call friends and colleagues. My master's advisor, David Koslicki, and my professors and colleagues at Oregon State helped me to develop a strong work ethic. From the countless late nights studying, to foosball games and jam sessions, Dane Skinner, Joe Umhoeffer, Will Mayfield, and Debora Piemonti were always around to support me.

To my family: Mom, Shawna and Jeff, Katie and Trey, Dad and Kelly, and all of the Fultons, Kislias and Kopps, you have always supported my dreams and encouraged me to further pursue education. Through all of my struggles in grad school, you consistently believed in me and never let me quit. I would not have made it this far without you all.

Lastly, I have to thank my husband, Nathan Fulton. Thank you for your technical support: for introducing me to cluster computing, for the late night debugging, and for being a sounding board to work through difficult concepts. More importantly, throughout all of the changes in college and grad school, your love and support have remained constant. Sharing similar life goals and career aspirations with you always made me feel that we were in this together; I've had a partner by my side who personally understands the struggles I've faced. You challenged me to push myself past every hurdle, you held me through many tears, and you believed in me more than I believed in myself. Nathan, I am so blessed to have you in my life. Thank you, my love.

## 1.0 Introduction

Since the discovery of DNA, mathematical models have played a key role in understanding its molecular structure, genomic function, and the cellular processes involving this macromolecule. This discovery has led to advances in the medical field and biology as a whole, and it sparked development of the fields of genetic engineering, genomics, and molecular biology. While a lot of progress has been made in understanding DNA as a whole, there is still much to learn about its function in complex cellular processes.

The modeling process itself gave rise to mathematical frameworks created to better understand the role of DNA at the molecular and cellular level. Our work focuses on connecting several of these modeling frameworks to the broader discipline of mathematics. We study DNA topology by specifically considering DNA knots and links. We consider two types of models: a continuum model, in which DNA is thought of as an electrostatically charged Cosserat rod in fluid, and a discrete model, in which DNA is considered to be a sequence of base pairs interacting with a fluid. We then use these models to study the energy landscape of DNA elastic knots with excess twist and the dynamics of a network of linked kinetoplast DNA minicircles.

Chapter 2 studies the space of DNA knot equilibria with excess twist, their symmetries, transition configurations, and the elastic energies of the DNA knot configurations. Here we use the Generalized Immersed Boundary (GIB) Method [41] to model the dynamics of DNA immersed in a fluid. In the GIB method, we idealize DNA as a charged Cosserat rod and use Navier-Stokes equations to describe the motion of the fluid. This chapter provides an idea of the state space of knot equilibria, but gives an incomplete characterization of the dynamics of DNA. Using a deterministic model does not account for thermal fluctuations within the fluid and only gives equilibrium states within the complete energy landscape.

Chapter 3 characterizes the dynamics of DNA knots as a continuous time Markov chain. Here, we use the Stochastic Generalized Immersed Boundary (SGIB) Method [73] to model the interaction of DNA, idealized as a charged Cosserat rod, within a fluid using Stokes equations to model fluid flow. This method gives a more realistic account of the dynamics of

DNA in a fluid by accounting for thermal fluctuations within the fluid. Using Kendall shape space analysis and a Procrustes distance [40] classifier, we find equivalence classes of knot equilibria and the barriers between the deterministic states obtained in chapter 2. We then use the SGIB knot traces as a continuous time Markov chain within this state space, and obtain transition rates, and mean holding times for each of these random walks. Finally, we use the transition rates to obtain a stationary Gibb's distribution for knot configurations for a fixed linking number and compare this to the distribution estimated from only the elastic energy of the system.

Chapter 4 considers the dynamics of DNA links. Here, we use a novel method, the Sequence Dependent Stochastic Generalized Immersed Boundary (SD-SGIB) Method, to dynamically model the interaction of two kinetoplast DNA minicircles. In the SD-SGIB method, we idealize DNA as a sequence of base pairs immersed in a fluid. This method accounts for the differing elastic properties of the exact base pair sequence of a DNA molecule, as opposed to the GIB and SGIB methods, which idealize DNA as a homogeneous, isotropic elastic rod. We then analyze the method, and use the SD-SGIB method to find a distribution of the center of masses of two Hopf-linked kDNA minicircles for two kinetoplastida: *T. brucei* and *C. fasciculata*.

## 1.1 Background

A DNA duplex is a double-helical polymer with two polynucleotide strands. Each of the two strands has a sugar-phosphate backbone and nucleotide base pairs. The two strands are joined in a Watson-Crick double-helix [77], with complementary base pairs, A to T or G to C, joined by hydrogen bonds. There is a single charge along the backbone at each base pair. This structure of DNA lends the polymer to have bending and torsional stiffness, and be nearly inextensible.

One of the forms of DNA we are interested in studying is a DNA plasmid, or closed ring. Due to processes of replication, transcription, recombination and DNA compaction the topological structure of DNA changes, and the DNA plasmid becomes supercoiled and

knotted. In vivo, supercoiling is the natural state of a plasmid in the cell. DNA is either naturally plectonemic for bacterial DNA plasmids or solenoidal, meaning spooled around histones in eukaryotic cell nuclei.

The linking number of DNA is a topological invariant and is defined by the Gauss linking number of two closed curves. Topoisomerases and Recombinases are enzymes that are responsible for changing this DNA topology. Topoisomerases specifically maintain the plectonemic or solenoidal structure of DNA in a cell. They do so by either cutting and ligating a single strand or two strands of DNA. Topoisomerase I cuts a single strand of DNA and changes the linking number of DNA by 1. Similarly, Topoisomerase II cuts both strands of DNA and changes the linking number by 2. [76]. The occurrence of DNA knots has been verified experimentally by Rybenkov et al., who found a distribution of DNA knots in the presence of Topoisomerase II [67].

DNA knots arise in several cellular processes, and the presence of DNA knots has also been found experimentally [74, 67]. In the case of bacteriophage packaging, a linear strand of DNA is packed into the bacteriophage capsid, and becomes knotted. This knotting process and the probability of knotting has been modeled by [1, 5, 4, 3]. Similarly, the role of Topoisomerase II in the process of unknotting has been modeled in [37], but not as a dynamical process.

Linked DNA plasmids also are formed through the process of cellular replication. The linking structure of entangled DNA loops is studied in [30], and others have studied the unlinking process of DNA [71, 69]. Linked segments are particularly pronounced in the structure of mitochondrial DNA of Kinetoplastid protozoa. Kinetoplastids are parasitic organisms with a unique mitochondrial DNA structure. Two examples of harmful Kinetoplasts that are *Trypanosoma brucei* which causes African sleeping sickness, and *Trypanosoma cruzi*, which causes Chagas disease.

This kinetoplast DNA, or kDNA, consists of a collection of around 5000 DNA minicircles and 20 to 30 maxicircles linked in a sparse chainmail-like structure [68]. The formation of this structure is not completely understood. Arsuaga et al. have proposed that the formation of the kDNA network is a product of confinement [2, 25, 24, 26], in which they prove that the probability of network formation increases to one with the density of rigid minicircles

[24]. Another model by Ogbadoyi et al. proposes that with no linking, minicircles would be lost due to segregation [58]. The role of histone-like proteins or structure preserving enzymes has been investigated in [39, 82]. Similarly, topological models have been used to better understand the network topology of kDNA [2, 25, 15, 26, 24, 51, 48, 47, 28]. We are interested in modeling a small network of kDNA minicircles dynamically.

Jun O’Hara defined the energy of a knot based upon the electrostatic energy of a thin rod [60]. This O’Hara energy is a functional on the space of knot conformations that preserves prime knot types under gradient descent. Freedman, He, and Wang later proved the existence of energy minimizers for knots of a particular class [31]. The knot energy minimizers give a notion of an ideal knot shape. Since proving the existence of these knot energy minimizers, many researchers have sought to characterize properties of energy minimizers for various knot energies including Möbius energies [10, 11, 9, 12].

The energy minimizing knot configurations are typically found by minimizing a knot energy with a gradient decent algorithm. Ideal extensible knot configurations have been studied by minimizing bending energies both numerically [8] and analytically [34]; however, we are particularly interested in the dynamical process of finding energy minimizing configurations. Simplifying knot cascades ending an unlinking process similar to the action of Recombinase has been studied analytically in [49, 50]. Similarly, the energy spectrum of links was studied in [55, 66].

Because of the bending and torsional stiffness properties of DNA, it is natural to model the double helical structure as a charged elastic rod. We are interested in studying the dynamics of DNA knots, by modeling DNA as an electrostatically charged elastic rod. Since O’Hara knot energies are similar to electrostatic energies, this should give an idea of the energy landscape of extensible knots with twist, and the dynamics of these configurations.

Using the theory of Kirchoff rods, supercoiling of elastic rings and knots as well as applications to DNA have been studied, in [18, 21, 20, 17, 19]. These works model DNA as an inextensible Kirchoff rod with bending and twisting energy, but electrostatic energy is not accounted for. Lim et al. studied the effects of a closed rod with twist and bend in fluid numerically [46] using an extension to Peskin’s Immersed Boundary method [70]. This model was extended to include electrostatics and a hardcore potential to model the dynamics

of DNA plasmids immersed in fluid in the Generalized Immersed Boundary method [41]. Stochastic DNA plasmid models [73] based upon the Stochastic Immersed boundary method [6]. As a result, we seek to model the dynamics of DNA trefoil knots with excess twist using both the GIB and SGIB methods.

## 1.2 DNA modeling assumptions

Throughout this work, we make three main sets of modeling assumptions about the structure of DNA and its surrounding fluid. In chapters 2 and 3, we make the same simplifying assumptions about the structure of a DNA segment, which coincide with the assumptions made in the GIB [46, 41] and SGIB [73] methods. We think of the double-helical structure of DNA as an elastic rod immersed in fluid. Figure 1 illustrates the correspondence between a DNA segment and Cosserat rod. Instead of emphasizing the mechanical properties between base pairs, we assume that the DNA segment has uniform twist, bend and shear constants along the entire segment of DNA. This translates into assuming the Cosserat rod is homogeneous and isotropic. We also assume that the charges positioned along the sugar-phosphate backbone of a DNA molecule are moved to the central axis of the rod. In figure 2 we see that the charges are moved from the backbone to the central axis of the Cosserat rod in the GIB and SGIB methods.

To drive dynamics of a knotted DNA segment, we will impose an initial excess twist in the rod. However, this is not to be confused with the initial pitch of the helical structure. A double-helical segment of DNA at equilibrium has no excess twist in the rod segment. Thus imposing excess twist induces strain in the helical structure and as we have observed, drives the dynamics of a knotted DNA segment.

Throughout chapters 2 and 3, as in the GIB method, [41, 46], and SGIB [73] method, we assume the DNA is immersed in an incompressible fluid, with a low Reynolds number. We also make the assumption that the fluid contains a dissolved monovalent salt. Thus we would expect electrostatic energy to affect the overall dynamics of the system. However, in the GIB method [41, 46], an throughout chapter 2, the rod is immersed in an incompressible

fluid, with no thermal fluctuations, and fluid dynamics are given by Navier-Stokes flow. Throughout chapter 3, we assume the fluid has random thermal fluctuations, and we use a corresponding stochastic model. We also model fluid dynamics using Stokes flow throughout this chapter.

Throughout chapter 4, our assumptions about the idealized structure of DNA and its surrounding fluid differ from chapters 2 and 3. First, we assume that the DNA segment is immersed in a ionized fluid, but we no longer model the DNA by an elastic rod. Instead, we follow the theory of sequence-dependent DNA elasticity [16], and consider a DNA segment as a series of stacked base pairs, each with their own orientation and kinematic properties between adjacent base pairs. As a result, the specific DNA sequence affects the mechanical properties along the segment of DNA, and the segment no longer has homogeneous and transversely isotropic properties. The right panel of figure 1 illustrates the discrete, sequence-dependent base pair model. We do however still assume that the electrostatic charges along the DNA segment are located along the central axis of the rod. In figure 2, the charges from the sugar-phosphate backbone are moved to the center of each base-pair slab.

The fluid is still assumed to be incompressible with low Reynolds number, and the dynamics of the fluid is given by Stokes flow. We assume that the fluid contains a dissolved, monovalent salt, as done previously, but for computational simplicity, instead of treating the electrostatic interaction using a Debye-Hückel screening term [78], we treat the electrostatic interaction as in [67], in which the effect of electrostatic repulsion gives DNA a larger effective diameter.

### 1.3 DNA topology

We can begin to define concepts in DNA topology by thinking of a strand of DNA as two oriented curves. The first curve,  $C_1$ , lies along the central axis, through the hydrogen bonds of the two helical strands, while the second curve,  $C_2$ , lies along the DNA backbone of one of the strands. We choose the axial curve as the first curve, as opposed to using both backbones as our two curves, because in the continuous representation of DNA, we think of

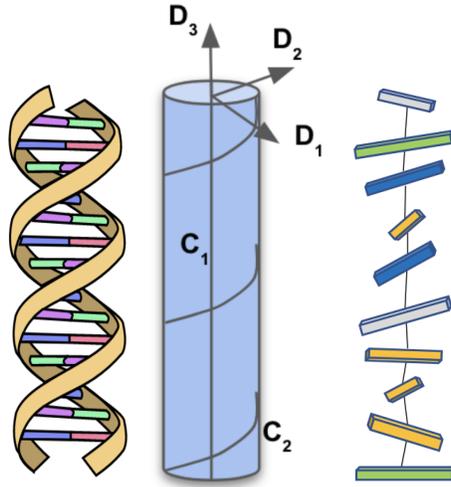


Figure 1: Left: DNA [80] Center: Cosserat rod with central axis,  $C_1$ , exterior axis,  $C_2$ , and frame  $\{\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3\}$  Right: Sequence-dependent discrete base pair model.

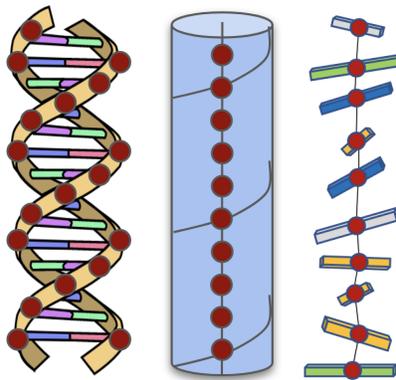


Figure 2: Electrostatic charges along the phosphate backbone are moved to the central axis of the DNA rod (Center) and discrete base pair sequence (Right) [80].

the central axis as the axis of our Cosserat rod. Since a DNA plasmid is a closed knotted or unknotted ring, let  $C_1$  and  $C_2$  be two oriented, closed curves in  $\mathbb{R}^3$ . Given  $C_1$  and  $C_2$ , the linking number describes how woven the two curves are. We can think of this as half the number of signed crossings of two closed curves in any planar projection. Formally, for two differentiable curves, we define the linking number of  $C_1$  and  $C_2$  by taking the Gauss integral [22], [72]:

$$Lk(C_1, C_2) = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{\mathbf{t}_1(s_1) \times \mathbf{t}_2(s_2) \cdot (\mathbf{x}_1(s_1) - \mathbf{x}_2(s_2))}{|\mathbf{x}_1 - \mathbf{x}_2|^3} ds_1 ds_2, \quad (1.1)$$

where  $C_i$  is parametrized by  $\mathbf{x}_i(s)$  as a function of arc length,  $s$ , and  $\mathbf{t}_i = \frac{d\mathbf{x}_i(s)}{ds}$ . An example of two oriented curves with linking number of 6 can be found in figure 3. The two curves  $C_1$  and  $C_2$  are given by the central axis, and the red winding curve respectively.

We can similarly define the writhe and twist for a strand of DNA. Given a single oriented, closed curve,  $C_1$ , the writhe represents the amount of chiral deformation in a curve. This is then either the average number of signed crossings over all planar projections, or formally for a closed, differentiable curve  $C_1$ , the writhe is defined as [22], [72]:

$$Wr(C_1) = \frac{1}{4\pi} \oint_{C_1} \oint_{C_1} \frac{\mathbf{t}_1(s_1) \times \mathbf{t}_2(s_2) \cdot (\mathbf{x}_1(s_1) - \mathbf{x}_2(s_2))}{|\mathbf{x}_1 - \mathbf{x}_2|^3} ds_1 ds_2. \quad (1.2)$$

An example of a planar curve with writhe of 1 and -2 is shown in figure 4. The signed crossing convention follows the standard signed crossing convention [22].

We can similarly define the twist of a DNA segment. Given  $C_1$  and  $C_2$ , the twist measures the amount the amount one curves winds about another. For two differentiable curves,  $C_1$  and  $C_2$ , in which  $\mathbf{t}_1(s)$  is the unit tangent to  $C_1$ , and  $\mathbf{d}(s) = \mathbf{x}_2(\sigma(s)) - \mathbf{x}_2(s)$  is perpendicular to  $\mathbf{t}_1(s)$ , the twist is defined as [72], [22]:

$$Tw(C_2, C_1) = \frac{1}{2\pi} \oint_{C_1} (\mathbf{t}_1(s) \times \mathbf{d}(s)) \cdot \mathbf{d}'(s) ds. \quad (1.3)$$

An example of twist of 0,  $\frac{1}{2}$ , and 1 is shown in figure 5.

Two curves with a nonzero linking number are said to be topologically linked. Two oriented curves with linking number of 1 is known as a Hopf link. For a given knot type, the linking number is topologically invariant. While neither the twist nor writhe is invariant,

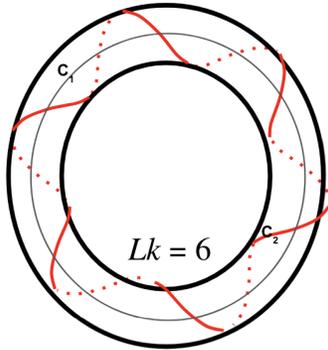


Figure 3: Linking number  $Lk(C_1, C_2) = 6$ ,  $Wr(C_1) = 0$ ,  $Tw(C_1, C_2) = 6$  for two curves, the central axis,  $C_1$  and the winding curve  $C_2$  in red.

a result by Calugareanu [14] and White [79] is that the linking number is the sum of the writhe and twist:

$$Lk(C_1, C_2) = Wr(C_1) + Tw(C_2, C_1) \quad (1.4)$$

With regards to DNA, various enzymes change the topology of DNA during processes of cellular replication. Topoisomerase I and Topoisomerase II change the linking number of DNA by  $\pm 1$  or  $\pm 2$  respectively by strand passage. These enzymes typically simplify the topology of knotted DNA during cellular replication. Recombinase also similarly simplifies DNA topology by reducing crossings, but at a slower rate than the Topoisomerases.

Typically, a DNA plasmid with high writhe exhibits supercoiling. These supercoiled structures are known as plectonemes. We are interested in exploring the dynamics of DNA knots by increasing the initial twist in the DNA segment to see how this affects the end structure or the stable configuration of the knotted structure. Because the linking number remains invariant for a given knot type, we would expect that an initial increase in twist could lead to an increase in writhe, when the twist relaxes.

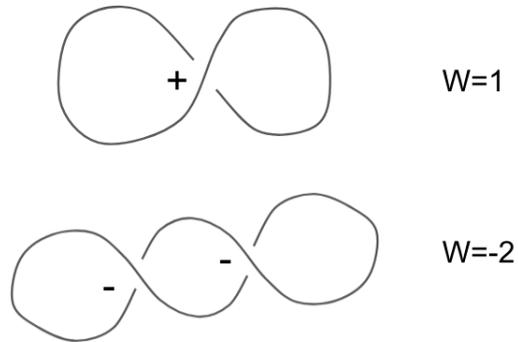


Figure 4: Writhe of 1 and 2 for an oriented curve, with signed crossing convention.

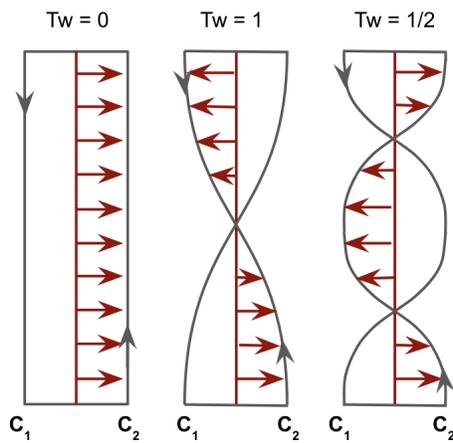


Figure 5: Twist of 0,  $\frac{1}{2}$ , and 1 for two oriented curves,  $C_1$  and  $C_2$ .

## 2.0 Equilibrium configurations of DNA knots

### 2.1 Introduction

DNA is a double-helical polymer consisting of two polynucleotide strands connected by nucleotide base pairs bound by hydrogen bonds. The charged sugar-phosphate backbone has residues located every 0.34 nm, each with a single electronic charge. These charges and the sugar-phosphate backbone itself cause DNA to have bending and torsional stiffness. The residues along the backbone also cause a self-repulsive force that reduces the chance of self-contact of a strand of DNA.

The mechanical properties of DNA are particularly important during processes of cellular transcription and replication, when DNA becomes knotted and forms supercoiled structures, or plectonemes. Both Topoisomerase enzymes and Recombinase enzymes change the topology of DNA throughout these cellular processes. We are interested in modeling the dynamics of these supercoiled DNA knot configurations.

Both knot and plectoneme structures have been observed *in vivo* and *in vitro*. Vinograd and Lebowitz observed supercoiled closed DNA rings, or plasmids [75]. Using gel electrophoresis, Rybenkov et al. found a distribution of knot types, for low-crossing number torus knots [67], and among the closed DNA segments, the most common knot type observed, apart from the unknot, was the trefoil knot. As a result, we focus our studies on modeling the dynamics of the trefoil knot type.

Due to its mechanical structure, we model DNA as in the Generalized Immersed Boundary (GIB) method [41, 46]. DNA's charged sugar-phosphate backbone causes bending and torsional stiffness. It also greatly reduces the chance of a strand coming into self-contact, meaning that until the DNA is in the presence of enzymes, no strand passage naturally occurs. We thus model DNA by an intrinsically straight, homogeneous, isotropic, elastic rod. For simplicity, we assume that the charges from the phosphate backbone are distributed uniformly along the central axis. Self-contact forces are included as (i) a long-range electrostatic repulsive force, and (ii) a hard-core potential that prevents strand passage of two segments

of the DNA.

In vivo, DNA is immersed in a solution with dissolved counterions. The effect of electrostatic repulsion is dampened by the concentration of monovalent and divalent salt ions in the solvent. The GIB method models the fluid-structure interaction of DNA idealized as a rod immersed in a solvent, and the electrostatic dampening effect is modeled as in the theory of counterion condensation, due to Manning [53].

In 2004, Coleman and Swigon found equilibrium structures of closed, knotted and unknotted rods [19]. However, electrostatic energy was not accounted for in their model. Lim et al. [46] generalized the Immersed Boundary method to closed circular, elastic rods, and numerically obtained equilibrium structures of a closed, twisted rod in a fluid. Using an extension of the GIB method, Lim et al. modeled the dynamics of closed, unknotted electrostatically charged rods immersed in a fluid with various dissolved molar salt concentrations in [41]. They found equilibrium structures of these unknotted, closed rods using an extension of the GIB method, but no other knot types were considered.

Our goal is to find equilibrium and saddle configurations of the DNA elastic rod model, which includes electrostatic and contact energies. We then describe the main features of such configurations and divide them into classes according to their symmetries. We also describe main features of the dynamics of a knot configuration during the process of equilibration.

## 2.2 Generalized Immersed Boundary Method

The Generalized Immersed Boundary (GIB) Method [46], and its extension to DNA specifically [41], describes the dynamics of a closed, electrostatically charged, elastic rod immersed in fluid. The GIB method builds upon Peskin's Immersed Boundary Method [70], and is a generalization of the method in [46], which applied the immersed boundary method to closed, rods with intrinsic twist immersed in a fluid.

The Immersed Boundary Method was first used to study the dynamics of heart valves and their interaction with surrounding fluid [61, 62, 63]. The classical IB method is applied to two or three-dimensional structures immersed in a three-dimensional fluid. The GIB method

[46] extended this to include a thin Cosserat rod, or one-dimensional structure in fluid. This method incorporates the fluid-structure interaction using a smooth approximation of a delta function to distribute forces and torque from the rod onto the fluid. This technique both gives the Cosserat rod an effective diameter and avoids the issue of a singularity with a standard delta function as a kernel.

In the extension of the GIB method [41], DNA is modeled by a closed, intrinsically straight, electrostatically charged, elastic rod. For simplicity, the elastic rod is also assumed to be homogeneous and isotropic. Thus, we assume that the kinematic properties are uniform throughout the segment of DNA, as opposed to depending on the specific sequence of the DNA plasmid, and that the DNA has no propensity to twist or bend in any particular direction.

The closed rod is assumed to be immersed in a fluid whose dynamics is described by an incompressible, Navier-Stokes flow with no-slip conditions. This follows the convention of the Immersed Boundary Method [70]. The motion of the fluid is described using the Eulerian coordinates, and the motion of the rod is described using the Lagrangian coordinates. The GIB accounts for the fluid-structure interaction of a one-dimensional rod in a fluid. This accounts for the forces and moments from the rod acting on the fluid and as well as the forces from the fluid onto the rod. The velocity of rod is the same as the fluid, meaning we impose the no-slip conditions. The interaction between the Eulerian and Lagrangian variables is accounted for using a smooth approximation of the three-dimensional Dirac delta function, which we refer to as a generalized delta function, as in the GIB method [41, 46].

Because of its kinematic properties, we model the double-helical structure of DNA as a closed, elastic rod whose mechanics are governed by the Cosserat rod theory. The rod is described by the position of the central axis,  $\mathbf{X}(s, t)$  at time  $t$ , and the frame at each point along the rod,  $\mathbf{d} = (\mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3)$ , at time  $t$ . This frame tracks the stretch and torque within the rod itself. As opposed to an inextensible, Kirchoff rod, the GIB method [46] allows for no single vector of the frame to lie tangent to the rod, meaning the material parameter  $s$  is not necessarily the arc length. These modifications allow the rod to be slightly extensible.

The Cosserat rod theory accounts for bending, twisting, stretching and shearing of the rod. It naturally takes into account the internal moments and forces of the rod. The force,

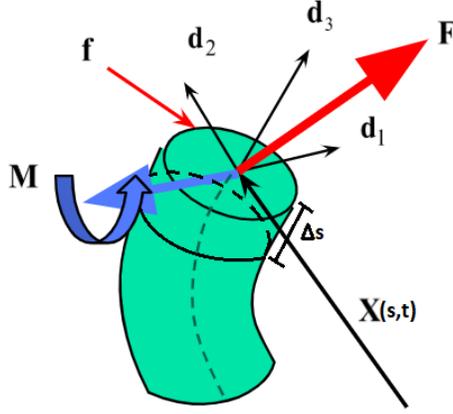


Figure 6: Rod mechanics

$\mathbf{F}(s, t)$ , and the moment,  $\mathbf{M}(s, t)$ , transmitted across the rod at position  $s$  are defined as in the GIB method [41].

In what follows,  $a_1$ , and  $a_2$  represent the bending moduli of the rod about  $\mathbf{D}^1$  and  $\mathbf{D}^2$  respectively;  $a_3$  is the twisting modulus of the rod;  $b_1$ , and  $b_2$  are the shear force constants;  $b_3$  is the stretch-force constant. The elastic energy intrinsic to the Cosserat rod is described by shear, bending, twist, and stretch energies:

$$\begin{aligned}
 E = \frac{1}{2} \int & \left[ a_1 \left( \frac{\partial \mathbf{D}^2}{\partial s} \cdot \mathbf{D}^3 \right)^2 + a_2 \left( \frac{\partial \mathbf{D}^3}{\partial s} \cdot \mathbf{D}^1 \right)^2 + a_3 \left( \frac{\partial \mathbf{D}^1}{\partial s} \cdot \mathbf{D}^2 \right)^2 \right. \\
 & \left. + b_1 \left( \mathbf{D}^1 \cdot \frac{\partial \mathbf{X}}{\partial s} \right)^2 + b_2 \left( \mathbf{D}^2 \cdot \frac{\partial \mathbf{X}}{\partial s} \right)^2 + b_3 \left( \mathbf{D}^3 \cdot \frac{\partial \mathbf{X}}{\partial s} - 1 \right)^2 \right] ds.
 \end{aligned} \tag{2.1}$$

Note that the total energy, however, also includes electrostatic energies, which is discussed in detail later.

In our case, for simplicity, the rod has transversely isotropic material properties, meaning  $a_1 = a_2 = a$ , and  $b_1 = b_2 = b$ . This is a minor assumption meaning the DNA does not have a propensity to twist or bend in either direction.

The dynamics of the fluid is governed by incompressible Navier-Stokes equations, where  $\mathbf{u}(\mathbf{x}, \mathbf{t})$  is the fluid velocity,  $p(\mathbf{x}, t)$  the fluid pressure,  $\rho$  the fluid density, and  $\mu$  the viscosity

of the fluid:

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \mu \delta \mathbf{u} + \mathbf{f}^b, \quad (2.2)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.3)$$

The force density  $\mathbf{f}^b(\mathbf{x}, t)$  represents the external body force from the rod to the fluid resulting from the mechanical imbalance of the immersed DNA. This takes into account the force and torque from the rod applied to the fluid, as well as the total self-contact forces and electrostatic repulsion from the rod applied to the fluid.

The full system of equations describing the rod and fluid dynamics is as in [41] and is as follows:

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \mu \delta \mathbf{u} + \mathbf{f}^b, \quad (2.4)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.5)$$

$$0 = \mathbf{f} + \frac{\partial \mathbf{F}}{\partial s}, \quad (2.6)$$

$$0 = \mathbf{m} + \frac{\partial \mathbf{M}}{\partial s} + \frac{\partial \mathbf{X}}{\partial s} \times \mathbf{F}, \quad (2.7)$$

$$\mathbf{F} = F_1 \mathbf{D}^1 + F_2 \mathbf{D}^2 + F_3 \mathbf{D}^3, \quad (2.8)$$

$$\mathbf{M} = M_1 \mathbf{D}^1 + M_2 \mathbf{D}^2 + M_3 \mathbf{D}^3, \quad (2.9)$$

$$F_1 = b_1 \mathbf{D}^1 \cdot \frac{\partial \mathbf{X}}{\partial s} \quad F_2 = b_2 \mathbf{D}^2 \cdot \frac{\partial \mathbf{X}}{\partial s} \quad F_3 = b_3 \left( \mathbf{D}^3 \cdot \frac{\partial \mathbf{X}}{\partial s} - 1 \right) \quad (2.10)$$

$$M_1 = a_1 \frac{\partial \mathbf{D}^2}{\partial s} \cdot \mathbf{D}^3 \quad M_2 = a_2 \frac{\partial \mathbf{D}^3}{\partial s} \cdot \mathbf{D}^1 \quad M_3 = a_3 \frac{\partial \mathbf{D}^1}{\partial s} \cdot \mathbf{D}^2 \quad (2.11)$$

$$\begin{aligned} \mathbf{f}^b(\mathbf{x}, t) = & \int (-\mathbf{f}(s, t)) \delta_c(\mathbf{x} - \mathbf{X}(s, t)) ds + \frac{1}{2} \nabla \times \int (-\mathbf{m}(s, t)) \delta(\mathbf{x} - \mathbf{X}(s, t)) ds \\ & + \int (-\mathbf{f}^c(s, t)) \delta_c(\mathbf{x} - \mathbf{X}(s, t)) ds + \sum_{i=1}^n (-\mathbf{f}_i^e(t)) \delta_c(\mathbf{X} - \mathbf{X}_i(t)), \end{aligned} \quad (2.12)$$

$$\frac{\partial \mathbf{X}(s, t)}{\partial t} = \mathbf{U}(s, t) = \int \mathbf{u}(s, t) \delta_c(\mathbf{x} - \mathbf{X}(s, t)) d\mathbf{x}, \quad (2.13)$$

$$\mathbf{W}(s, t) = \frac{1}{2} \int (\nabla \times \mathbf{u}) \delta_c(\mathbf{x} - \mathbf{X}(s, t)) d\mathbf{x}, \quad (2.14)$$

$$\frac{\partial \mathbf{D}_i(s, t)}{\partial t} = \mathbf{W}(s, t) \times \mathbf{D}_i(s, t), \quad i = 1, 2, 3. \quad (2.15)$$

Here  $-\mathbf{f}$  and  $-\mathbf{m}$  are the force and torque densities from the rod applied to the fluid. The terms  $\mathbf{f}^c(s, t)$  and  $\mathbf{f}_i^e(t)$  represent the steric repulsion and electrostatic forces defined below. The locally averaged fluid velocity and angular fluid velocity are given by  $\mathbf{U}(s, t)$  and  $\mathbf{W}(s, t)$ , and equation (2.15) updates the frame in the fluid.

The GIB method accounts for the fluid-structure interaction with a smooth approximation of a three-dimensional delta function, effectively giving the rod thickness within the fluid, and applying the force and torque generated by the rod onto the fluid. This is accounted for by a convolution of the function,  $\delta_c(\mathbf{x}) = \delta_c(x_1)\delta_c(x_2)\delta_c(x_3)$ , with each of the forces and moments from the rod applied to the fluid, including the total self-contact force and electrostatic forces. Here,  $\delta_c(\mathbf{x})$  is defined as follows [63]:

$$\delta_c(\mathbf{x}) = \frac{1}{c^3} \phi\left(\frac{x_1}{c}\right) \phi\left(\frac{x_2}{c}\right) \phi\left(\frac{x_3}{c}\right), \quad (2.16)$$

where the function  $\phi$  is defined by:

$$\phi(r) = \begin{cases} \frac{3-2|r|+\sqrt{1+4|r|-4r^2}}{8}, & \text{if } |r| \leq 1, \\ \frac{5-2|r|-\sqrt{-7+12|r|-4r^2}}{8}, & \text{if } 1 \leq |r| \leq 2, \\ 0, & \text{if } |r| \geq 2. \end{cases} \quad (2.17)$$

To ensure the force and torque of the rod are applied to the fluid only at neighboring grid points to the position of the rod, the following, with  $c$  being an integer multiple of the fluid mesh spacing  $h$ , must hold for all  $\mathbf{X}$ :

$$\sum_{\mathbf{j}} \delta_c(\mathbf{j} - \mathbf{X}) h^3 = 1, \quad (2.18)$$

$$\sum_{\mathbf{j}} (\mathbf{j}h - \mathbf{X}) \delta_c(\mathbf{j}h - \mathbf{X}) h^3 = 0. \quad (2.19)$$

The index  $\mathbf{j}$  is any vector with integer components.

In the extended GIB method [41], the self-contact force  $\mathbf{f}^c(s, t)$  at position  $s$  along the rod and time  $t$  follows a Hooke's law. Each local self-contact force is proportional to the amount of compression within the rod. The total self-contact force then is the sum over all local points of self-contact. The amount of compression is relative to the diameter of DNA,  $D = 20\text{\AA}$ , and the proportionality constant,  $g$ , is chosen sufficiently large so that there is no strand crossing during a self-contact.

That is, the self-contact force at position  $s$  given by the sum of contact points,  $s'$ , with  $s$ :

$$\mathbf{f}^c(s, t) = \sum_{s'} \hat{\mathbf{f}}^c(s, s', t), \quad (2.20)$$

where

$$\hat{\mathbf{f}}^c(s, s', t) = g(D - |\mathbf{X}(s, t) - \mathbf{X}(s', t)|) \frac{\mathbf{X}(s, t) - \mathbf{X}(s', t)}{|\mathbf{X}(s, t) - \mathbf{X}(s', t)|} \quad (2.21)$$

for any two material points  $s$  and  $s'$  satisfying  $|s - s'| > 2D$ ,  $|\mathbf{X}(s, t) - \mathbf{X}(s', t)| \leq D$ , and  $(\mathbf{X}(s, t) - \mathbf{X}(s', t)) \cdot \mathbf{T}(s, t) = (\mathbf{X}(s, t) - \mathbf{X}(s', t)) \cdot \mathbf{T}(s', t) = 0$  [41].

To model the electrostatic interaction of DNA, along the central axis of the rod, a single charge is located at each base pair, or every  $0.34nm$ . The total electrostatic force is similarly a sum of all electrostatic repulsive forces at each charge along the rod. Each local repulsive charge force is a screened Coulombic force following the counterion condensation theory. That is, each local repulsive charge is a repulsive force with a Debye-Hückel screening [78].

More precisely, the Coulombic force  $\mathbf{f}_i^e$  is defined by

$$\mathbf{f}_i^e(t) = \sum_{j \neq i}^n \mathbf{F}_{ij}^e(t), \quad (2.22)$$

where

$$\mathbf{F}_{ij}^e(t) = \frac{q_i q_j \exp^{\kappa d_{ij}(t)}}{4\pi\epsilon_0\epsilon_W d_{ij}(t)} \left( \frac{1}{d_{ij}(t)} + \kappa \right) \frac{(\mathbf{X}_i(t) - \mathbf{X}_j(t))}{d_{ij}(t)}, \quad (2.23)$$

and  $n$  is the number of base pairs; the position and net charge of the  $i$ th base pair is given by  $\mathbf{X}_i$  and  $q_i$  respectively; and  $d_{ij} = |\mathbf{X}_i(t) - \mathbf{X}_j(t)|$  is the distance between charges  $q_i$  and  $q_j$ . Also, the constants  $\epsilon_0$ , and  $\epsilon_W$  represent the permittivity of free space and the dielectric constant of water at 300K, and  $q_i$  is  $0.48e$ , where  $e$  is the elementary charge of an electron [52, 81]. For a molar salt concentration of  $C_s$  moles per litre, the Debye screening parameter is  $\kappa = 0.329\sqrt{C_s}\text{\AA}^{-1}$ .

The electrostatic potential is given by the sum:

$$E_{electro} = \sum_{i < j} \frac{q_i q_j e^{-\kappa d_{ij}}}{4\pi\epsilon_0\epsilon_W d_{ij}}. \quad (2.24)$$

### 2.3 Methods

The numerical method we used is as in the GIB method and implemented in MATLAB on a cluster. The rod and fluid each have their own mesh, with the fluid expressed in the Eulerian coordinates, and the DNA rod expressed in the Lagrangian coordinates. For simulations, we discretize the rod at each base pair, and consider closed DNA segments with 600 base pairs. Thus, the DNA is a closed segment, with periodic boundary conditions for the position of the rod at time  $t$ ,  $\mathbf{X}(s, t)$  and each of the vectors in the frame,  $\{\mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3\}$  at time  $t$ . The initial position of the rod is described by a standard trefoil knot parametrization:

$$\mathbf{X}(s, 0) = (X_1(s, 0), X_2(s, 0), X_3(s, 0)), \quad (2.25)$$

$$X_1(s, 0) = \frac{1}{p} \cos\left(\frac{ps}{r_0}\right) \left(1 + \frac{D}{r_0} \cos\left(\frac{qs}{r_0}\right)\right), \quad (2.26)$$

$$X_2(s, 0) = \frac{1}{p} \sin\left(\frac{ps}{r_0}\right) \left(1 + \frac{D}{r_0} * \cos\left(\frac{qs}{r_0}\right)\right), \quad (2.27)$$

$$X_3(s, 0) = \frac{D}{r_0} \sin(qs/r_0), \quad (2.28)$$

where  $D$  is as defined above, the diameter of the rod,  $r_0$  is the radius of a circle with circumference equal to a 600 base pair plasmid. Figure 7 illustrates the starting configuration of the DNA trefoil knot. The charges are located at every base pair, with a spacing of 0.34 nm. The parameters  $p$  and  $q$  represent the number of times the central axis wraps around the rotational axis of symmetry of a torus and the central circle of the torus respectively. This holds for general  $(p, q)$ -torus knots, and in the case of a right-handed trefoil,  $p = 2$  and  $q = 3$ .

The rod is given an initial excess twist,  $Lk_{rel}$ , with uniform twist density,  $Lk_{rel}/(2\pi r_0)$ . This is achieved by arranging the initial orientation of the material frame,  $\{\mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3\}$ . Note that for a circular configuration, the excess twist parameter is the same as the change in linking number,  $\Delta Lk$ . For the right-handed trefoil knot, the linking number is  $-3$ , so  $Lk_{rel} = \Delta Lk + 3$ . The frame is parametrized by material parameter  $s$ , and  $\mathbf{z}$  is the unit tangent to the  $z$ -axis:

$$\mathbf{D}^3(s) = \frac{r_0}{\left\| r_0 \frac{d\mathbf{X}(s,0)}{ds} \right\|} \frac{d\mathbf{X}(s,0)}{ds}, \quad (2.29)$$

$$\mathbf{R}(s) = \frac{\mathbf{D}^3 \times \mathbf{z}}{\|\mathbf{D}^3 \times \mathbf{z}\|}, \quad (2.30)$$

$$\mathbf{E}(s) = \frac{\mathbf{R} \times \mathbf{D}^3}{\|\mathbf{R} \times \mathbf{D}^3\|}, \quad (2.31)$$

and

$$\mathbf{D}^1(s) = \cos\left(Lk_{rel} \frac{s}{r_0}\right) \mathbf{E}(s) + \sin\left(Lk_{rel} \frac{s}{r_0}\right) \mathbf{R}(s), \quad (2.32)$$

$$\mathbf{D}^2(s) = -\sin\left(Lk_{rel} \frac{s}{r_0}\right) \mathbf{E}(s) + \cos\left(Lk_{rel} \frac{s}{r_0}\right) \mathbf{R}(s), \quad (2.33)$$

where

$$0 \leq s \leq 2\pi r_0. \quad (2.34)$$

Initially,  $\mathbf{D}^3$  lies tangent to the initial axial curve  $\mathbf{X}(s, 0)$ . The unit vectors  $\mathbf{R}(s)$  and  $\mathbf{E}(s)$  are useful vectors perpendicular to  $\mathbf{D}^3$ . The frame vectors  $\mathbf{D}^1$  and  $\mathbf{D}^2$  are expanded in this normal frame and describe the number of initial twists the axial curve is given,  $Lk_{rel}$ . Note that as opposed to a starting circular configuration, because the starting trefoil configuration

is not an equilibrium solution for the system, a uniform twist density is sufficient to drive the dynamics of the immersed DNA and surrounding fluid.

The fluid starts at rest and is given periodic boundary conditions with a space grid of  $10\text{\AA}$ . This means the system is driven by the initial excess twist in the rod. We make the assumption, as in [41], that the fluid contains monovalent salt ions with a concentration of either  $Cs = 0.1\text{M}$  or  $Cs = 0.01\text{M}$  to examine the effects of electrostatics on the dynamics of the system.

Given an initial knot configuration and twist density, the algorithm is as follows [41]: 1. Compute the external force,  $\mathbf{f}$ , and moment,  $\mathbf{m}$ , densities from the rod onto the fluid. These are computed using the position  $\mathbf{X}$  of the rod at time  $t$  and the frame at every position along the rod,  $\{\mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3\}$  using the discretized balance equations. 2. Spread  $\mathbf{f}$ , and  $\mathbf{m}$  from the rod to the fluid using the discrete delta function. 3. Update fluid velocity  $\mathbf{u}$  by solving the Navier-Stokes equations using an FFT. 4. Interpolate  $\mathbf{u}$  at the rod discretization points. 5. Interpolate the rod through the fluid. 6. Check the CFL condition to ensure convergence [46].

To find knotted equilibrium configurations, we had two simulation phases. In the first phase, for each fixed  $Cs = 0.1\text{M}$  with  $Cs = 0.01\text{M}$  and  $Lk_{rel}$  ranging from -10 to 10, we found a single stable configuration by starting with the standard parametrization of the (3,2)-torus knot, and imposing a uniform twist density by an integer number of twists, of  $Lk_{rel}$ . This means that both the rod and the frames have a periodic boundary condition, i.e.  $\mathbf{X}^1 = \mathbf{X}^{601}$ , and  $\mathbf{D}^{i,1} = \mathbf{D}^{i,601}$ , for  $i = 1, 2, 3$ .

In the second phase of simulations, we wanted to further explore the space of equilibrium configurations. To do so, from each of the initial stable configurations from phase 1, for a fixed  $Cs = 0.1\text{M}$ , we perturb the initial equilibrium configuration by imposing a new excess link with uniform twist density. For each excess linking number,  $Lk_{rel}$ , this gave an entire family of stable equilibria.

The uniform twist density is imposed through the initial position of the  $\{\mathbf{D}^1, \mathbf{D}^2, \mathbf{D}^3\}$  frame. For each linking number the system is driven by twisting energy and runs until reaching a steady state. Each equilibrium configuration is parametrized by a molar ionic strength of the surrounding fluid, and excess linking number,  $Lk_{rel}$ .

For a knot configuration, we are interested in the dynamics of the coiling process, so we need to compute energy values, linking number, writhe, and twist. To do so, the total energy is a sum of the elastic energy, defined in the GIB section, with the steric and electrostatic energies. Thus, the total elastic energy includes bending, stretching, shearing, twisting, steric, and electrostatic energies. We also find the kinetic energy of the fluid and the rod and keep track of the fluid and rod velocities throughout a dynamic simulation. The energy of the rod is given by:

$$E_{tot} = E + E_{electro}, \quad (2.35)$$

and the kinetic energy of the fluid is

$$\frac{1}{2}\rho \int |\mathbf{u}|^2 dt. \quad (2.36)$$

The linking number for a given knot configuration remains invariant throughout dynamic simulations, assuming there is no strand passage. This follows from the theorem by Calugareanu and White [14], [79]. Since we fix the linking number in the beginning, we can use it to compute either the writhe or twist. Linking number, writhe, and twist are defined in equations 1.1, 1.2, and 1.3 of chapter 1 respectively, where  $C_1 = \mathbf{X}(s, \cdot)$  for a fixed time, and  $C_2$  is the curve on the exterior of the rod traced by the  $\mathbf{D}^1(s, \cdot)$  vector for a fixed time.

## 2.4 Results

We used two simulation phases to find equilibrium configurations and the dynamics between stable equilibria. In the first phase of simulations, we start with the standard parametrization of a trefoil, as in figure 7. In the second phase, we start with one of the established stable equilibria from phase one. Starting with the standard trefoil parametrization, for a fixed  $Cs = 0.1M$  or  $Cs = 0.01M$ , we impose a uniform twist density, by an integer number of excess link in the rod. The stable solutions can be found in figure 8. Similarly, the stable solutions starting from these phase one equilibria can be found in figures 12 and 13.

parameters	symbol	value
grid size	$N \times N \times N$	$64^3$
domain size	$L \times L \times L$	$L = 487.014 \text{ \AA}$
time step	$\Delta t$	$3 \times 10^{-11}$
fluid density	$\rho$	$1 \text{ g cm}^3$
fluid viscosity	$\mu$	$0.01 \text{ g (cm s)}^{-1}$
permittivity of free space	$\epsilon_0$	$3.45 \times 10^{-7} e^2 / (g \text{ \AA}^3 s^{-2})$
dielectric constant of water	$\epsilon_W$	$77.7$
electric charge	$q$	$0.24 \times 2e$
molar salt concentration	$Cs$	$0.01, 0.1M$
Debye screening parameter	$\kappa$	$0.329 \sqrt{Cs} \text{ \AA}^{-1}$
bending modulus	$a_1 = a_2 = a$	$1.3 \times 10^{-19} \text{ g cm}^3 \text{ s}^{-2}$
twist modulus	$a_3$	$1.82 \times 10^{-19} \text{ g cm}^3 \text{ s}^{-2}$
shear modulus	$b_1 = b_2 = b$	$5 \times 10^{-5} \text{ g cm}^3 \text{ s}^{-2}$
stretch modulus	$b_3 = b$	$5 \times 10^{-5} \text{ g cm}^3 \text{ s}^{-2}$
number of base pairs	$n$	$600$
radius	$r$	$324.676 \text{ \AA}$
diameter	$D$	$20 \text{ \AA}$

Table 1: GIB computational parameters



Figure 7: DNA starting configuration

Starting with an initial configuration of a standard trefoil knot with initial excess link,  $Lk_{rel}$ , regardless of  $Lk_{rel}$  and  $Cs$ , the system passes through typical stages of progression. A sample trajectory is seen in figure 9 for a 600 Bp DNA segment, with  $Lk_{rel} = -9$  and  $Cs = 0.1M$ .

The system is driven by excess twist energy resulting in a slight collapse of the axial curve. Following this is a rapid equilibration of the twist density until the knotted DNA reaches one or more points of self contact. This equilibration of the twist density is what results in the formation of plectonemes, or regions of high writhe, with terminal or central loops. The equilibration of initial excess twist is followed by a slower folding process, driven by the electrostatic repulsion and bending energy. Finally, we observe a slithering motion of the nearly equilibrated configuration, where two segments of the rod slide past one-another in a position of near contact.

We have also observed other equilibrium structures without any plectonemes present, where there is still a rapid equilibration of the initial excess twist density until one or more points of self-contact occurs. Often, if no plectonemes are present, this initial pretzel-like configuration is quite compact. This process is still typically followed by a slower bending process and possibly slight slithering. A canonical example of this compact configuration is the  $\Delta Lk = -5$ ,  $Cs = 0.1M$  knot shown in figure 8.

#### 2.4.1 Effects of $Lk_{rel}$ and $Cs$ on final configurations

Figure 8 illustrates the effects of linking number and ionic strength on the limiting knot configurations. An increase in the ionic strength of the fluid neutralizes the electrostatic repulsion effects of the counterions within the fluid. Thus, as  $Cs$  increases, the axial curve comes closer to self contact. This appears as a ballooning effect for lower values of  $Cs$ . This property is consistent across all values of  $Lk_{rel}$ . Similarly, as the absolute value of  $Lk_{rel}$  increases, we see longer plectonemes, or regions of supercoiling, and more compact knotted structures, also leading to more regions of near self-contact of the axial curve. This phenomenon is seen for values of  $|Lk_{rel}|$  greater than or equal to 4 at  $Cs = 0.1M$ , but is only noticeable for  $Lk_{rel} = -9, -10, 10$ , for  $Cs = 0.01M$ . Note however that because of both

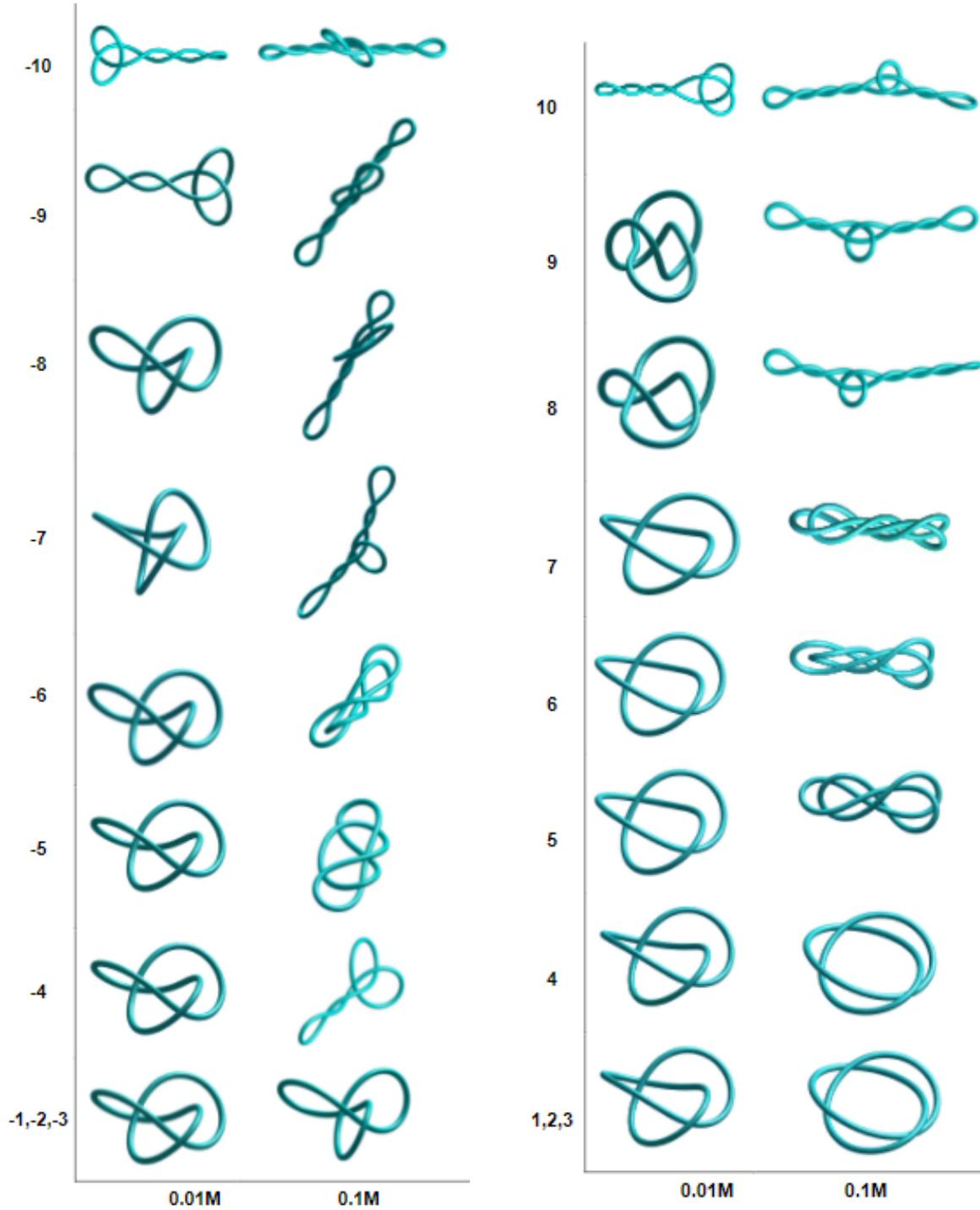


Figure 8: Phase one stable equilibrium configurations parametrized by  $Lk_{rel} = \pm 1, \pm 2, \dots \pm 10$  for  $Cs = 0.1M$  and  $Cs = 0.01M$ .

electrostatics and the hardcore potential, the DNA rod does not actually make contact with itself in the equilibrium configurations. If self-contact occurs, the hardcore potential acts as a spring force to repel the DNA rod at each point of contact, while the electrostatic force is a screened repulsive force at each base pair.

#### 2.4.2 Dynamics of transitions

Throughout the process of equilibration, the system follows gradient flow-like dynamics. This process occurs on the time scale of  $10 - 200\mu s$ . For a DNA knot configuration, the total energy in,  $kT$ , is a sum of twist, bending, shear, stretch, and electrostatic energies. The system is largely driven by the initial excess twist energy. As the twist energy equilibrates, the rod reaches one or more points of self-contact, resulting in a large increase in electrostatic energy. This effect is demonstrated by the simultaneous increase in excess bend and electrostatic energy as twist energy rapidly decreases until near equilibration. The elastic energies, fluid kinetic energy, and fluid and DNA velocities are displayed in figure 10.

The main contributors to the total energy of the elastic rod are the twist, and electrostatic energies followed, by the bending energy. The initial increase in the bending energy is a result of the equilibration of excess twist density. Then, as the slower bending process unfolds the knotted DNA, bending energy values decrease until reaching equilibration. Shear and stretch contribute little to the total elastic energy of the rod, and shearing requires relatively the same amount of time to equilibrate as bending, twist, and electrostatic energies. However, stretch equilibration is roughly one hundred times faster than all other mechanical processes and equilibration of the twist energy is reached on the order of  $0.1\mu s$ , meaning the DNA is very nearly inextensible.

The fluid flow also follows gradient flow-like dynamics. Both the fluid kinetic energy and fluid velocity follow a near exponential decay with possible local minima until reaching equilibration. The fluid velocity tracks with the velocity of the DNA immersed in the fluid. In this case, the dynamics of the fluid is solely driven by the dynamics of the twisted DNA configuration.

While the final configuration obtained is a stable steady state, we can also approximate

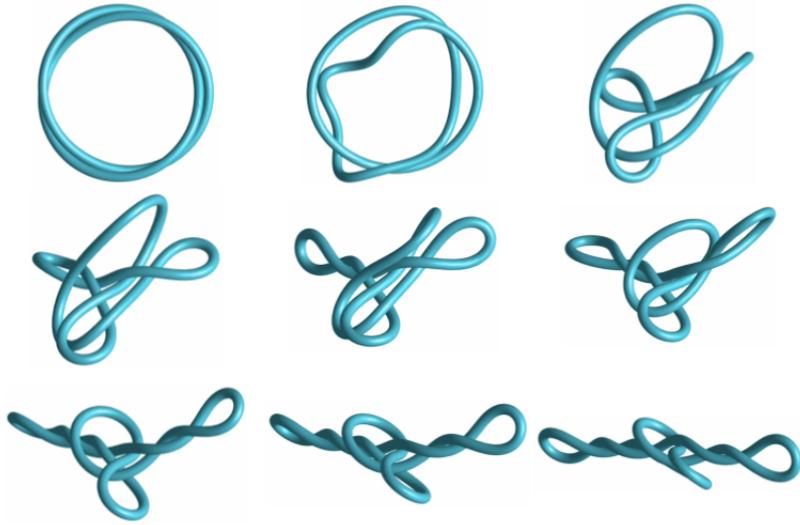


Figure 9: Dynamic progression of knot configuration starting from the standard trefoil parametrization for  $Lk_{rel} = -9$ ,  $Cs = 0.1M$ . Timestamps: from left to right and top to bottom frames:  $0.0003\mu s$ ,  $0.4203\mu s$ ,  $0.6903\mu s$ ,  $1.2003\mu s$ ,  $2.7303\mu s$ ,  $5.0403\mu s$ ,  $6.9603\mu s$ ,  $8.5803\mu s$ ,  $20.4303\mu s$ .

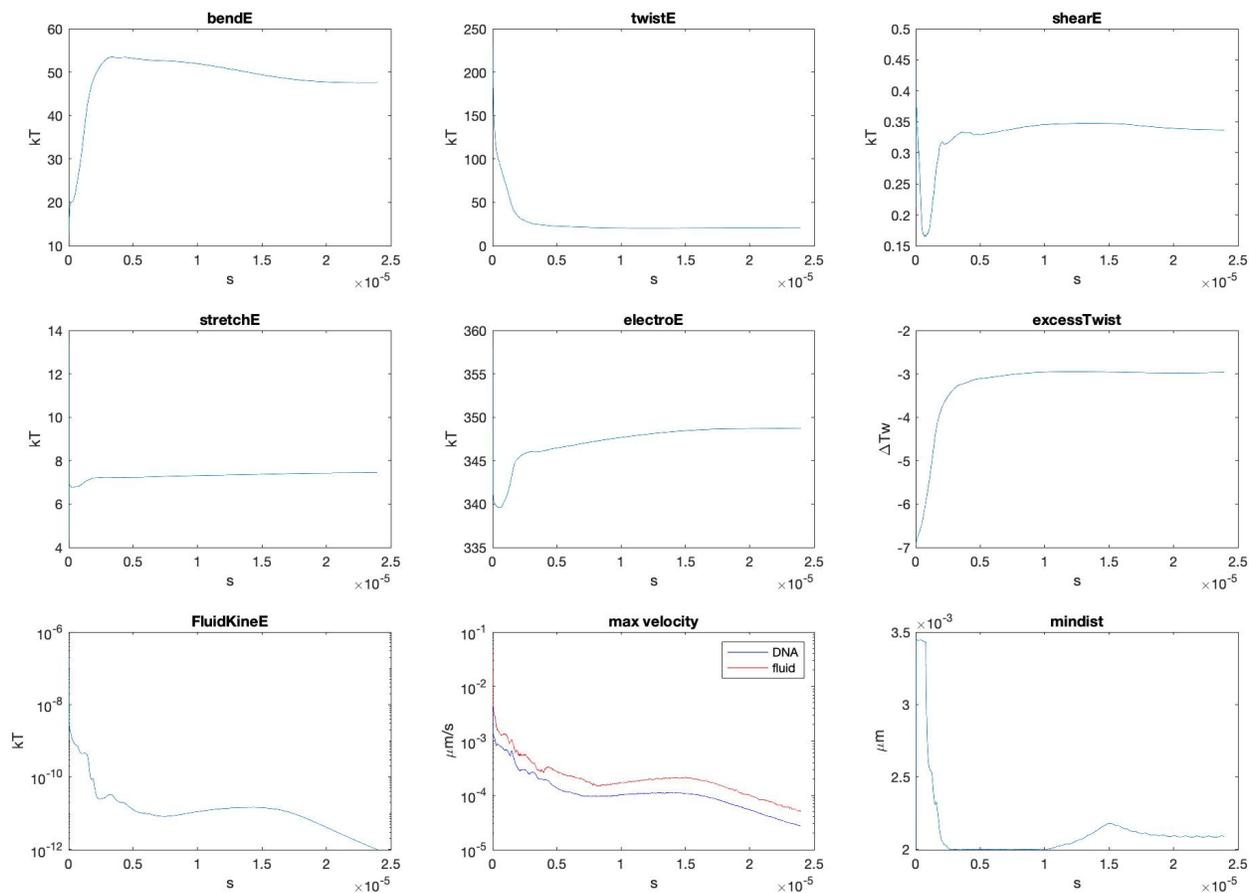


Figure 10: Elastic energies throughout dynamic simulation for  $Lk_{rel} = -7$  and  $Cs = 0.1M$ . Energy values are in units of kT, and velocities in  $\mu m/s$ . From the top left to bottom right panels: 1. bending energy, 2. twist energy, 3. shear energy, 4. stretch energy, 5. electrostatic energy, 6. excess twist, 7. fluid kinetic energy, 8. maximum velocity of the DNA and fluid, 9. minimum distance between points along the DNA.

saddle-like equilibrium structures. When the fluid and rod velocities decelerate, the system passes near a saddle-node. This frequently occurs after the rapid equilibration of the twist density, and before the slower bending process occurs. Example of saddle nodes are shown in figure 11. These were obtained by finding the knot configurations where the DNA velocity slows before accelerating again.

### 2.4.3 Phase two equilibria configurations

The stable equilibrium knot configurations in figures 12 and 13 were obtained through the second phase of exploring the  $Lk_{rel}, Cs$  parameter space. The configurations on the diagonal are those obtained in the first phase of simulations, and are identical to those shown in figure 8 for positive and negative  $Lk_{rel}$ . From these initial stable configurations, we perturbed the linking number by a uniform twist density, through changing  $\Delta Lk$  by integer values. This corresponds to moving vertically throughout the table starting from the diagonal.

Each of these configurations contains a 600 Bp DNA sequence for  $Cs = 0.1M$  and integer  $Lk_{rel}$  ranging from -10 to 10. Note that  $Lk_{rel} = 0$  is not displayed, but this is the linking number for the unperturbed standard trefoil. We considered only the  $Cs = 0.1M$  parameter space, because at  $Cs = 0.01M$ , the initial limiting configurations were much less diverse across all linking numbers than for  $Cs = 0.1M$ .

Throughout simulations, because there is no strand passage of the DNA segment, the linking number remains invariant. Moving across a row of figures 12 and 13 illustrates the variety of equilibria found while preserving linking number.

Notice that we have discovered entire new families which persist through various linking numbers. We observe some common features as before: plectonemes or regions of supercoiling occur with larger values of  $|Lk_{rel}|$ . The families of configurations with plectonemes and either terminal loops or central loops are more likely to bifurcate and change symmetry as  $Lk_{rel}$  changes, compared to the more compact stable structures. Consider for example the  $Lk_{rel} = -3$  column. As we move vertically throughout the column, by perturbing  $Lk_{rel}$ , the configuration passes through the standard trefoil knot,  $Lk_{rel} = -1, -2$ , an opened version of that in  $Lk_{rel} = -3$ , a configuration with two loops and a plectoneme at  $Lk_{rel} = -4, -5$ ,

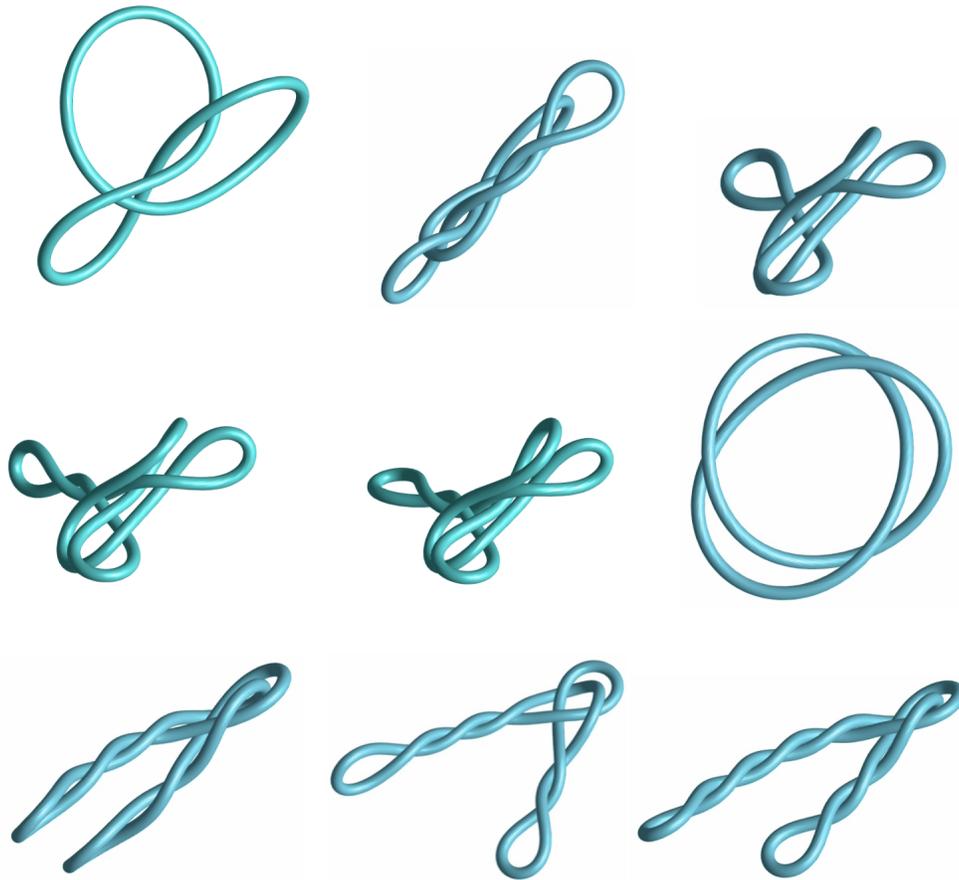


Figure 11: Saddle configuration for various linking numbers, with  $Cs = 0.1M$ . From left to right and top to bottom, displayed are saddle node configurations for  $Lk_{rel} = -4, -7, -8, -9, -10, 5, 8, 9, 10$ .

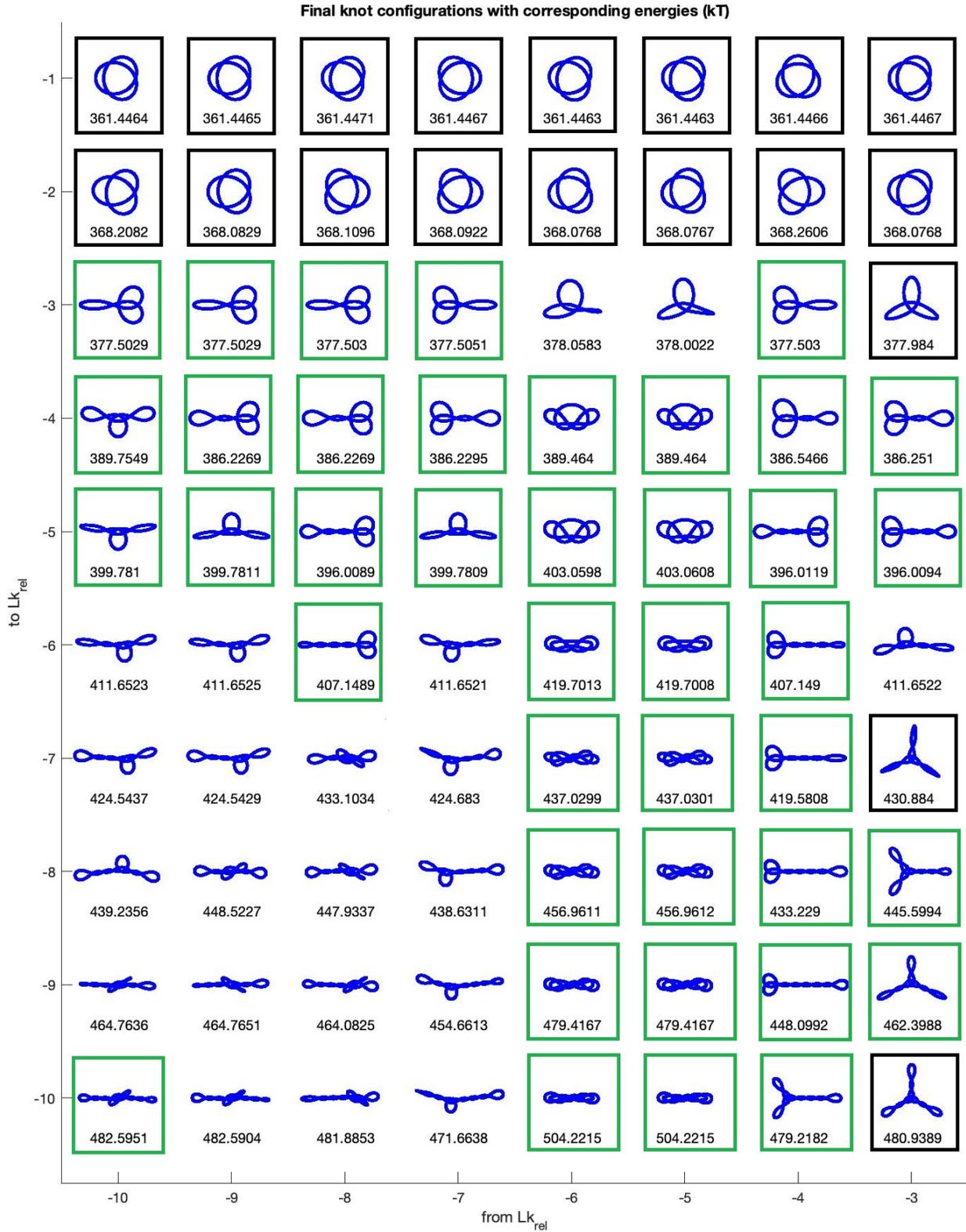


Figure 12: Phase two stable equilibrium configurations for  $Lk_{rel} = -1, -2, \dots, -10$  with corresponding energy values in kT. Configurations inside the black boxes have  $D_3$  symmetry, those inside green boxes have  $C_2$  symmetry, while the rest are asymmetric.

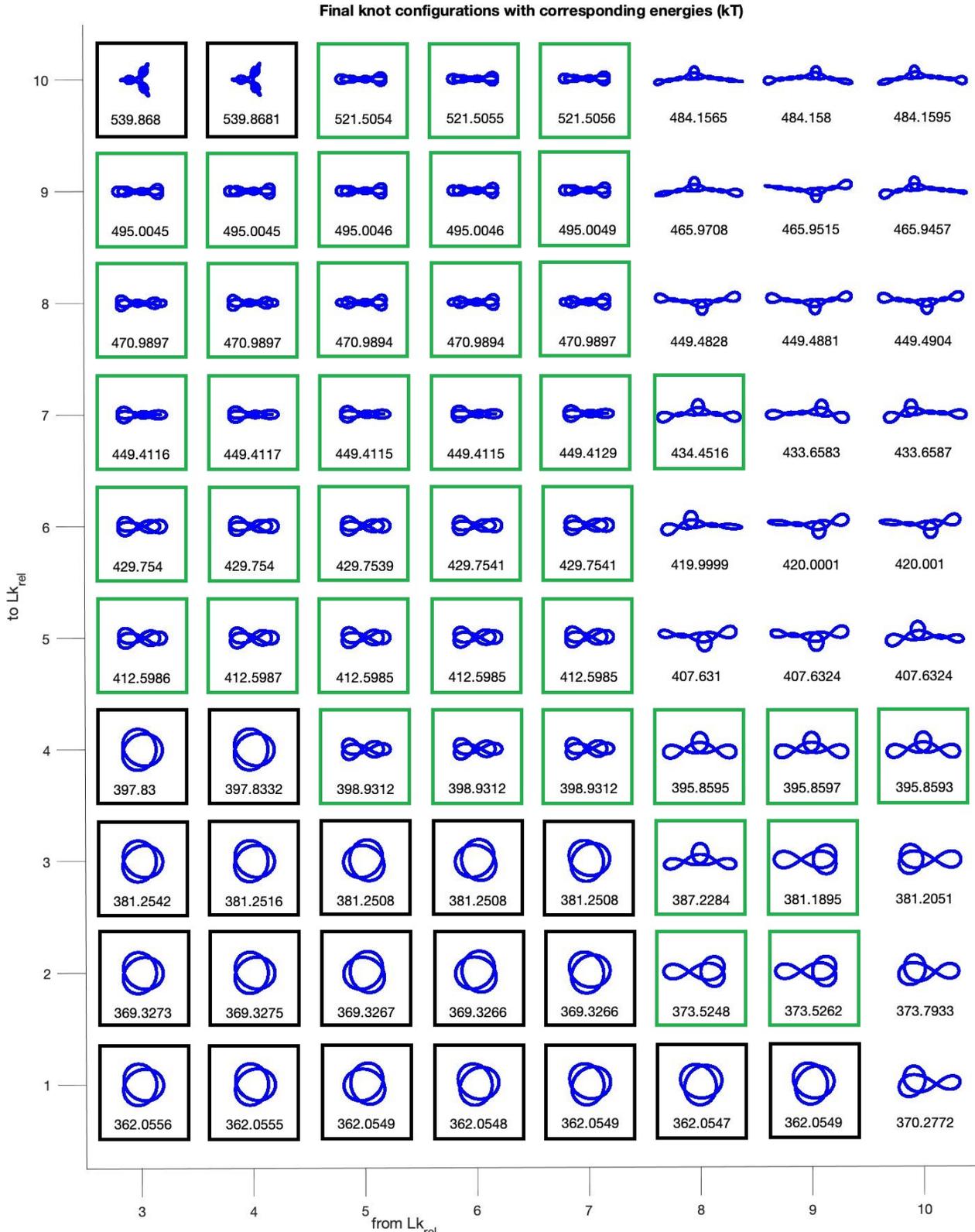


Figure 13: Phase two stable equilibrium configurations for  $Lk_{rel} = 1, 2, \dots, 10$  with corresponding energy values in kT. Configurations inside the black boxes have  $D_3$  symmetry, those inside green boxes have  $C_2$  symmetry, while the rest are asymmetric.

a central loop with two plectonemes at  $Lk_{rel} = -6$ , and three plectonemes with 3 terminal loops at  $Lk_{rel} = -7, -8, -9, -10$ .

If we compare this the  $Lk_{rel} = -5$  column, we see that the more compact, pretzel-like configurations are more robust when the linking number is perturbed. They do not develop long plectonemes by perturbing the linking number, and the configuration remains compact. This is consistent for positive and negative  $Lk_{rel}$  values.

Note that we also discovered new asymmetric knot configurations. We have found two types of symmetries for the trefoil knot with excess twist. Those that have a three fold rotational and reflection symmetry, or  $D_3$  symmetry, and those that have one axis of rotational symmetry, or belong to the  $C_2$  symmetry group. Figures 12 and 13 partition the space of stable equilibria by their respective symmetry groups. It is interesting that as we perturb the value of  $Lk_{rel}$ , we see a process of both symmetry breaking and formation. Starting from  $Lk_{rel} = -3$ , the knot configuration jumps branches between structures with  $D_3$  symmetry to asymmetry, to  $C_2$ , and back to  $D_3$  symmetries.

The dynamics of the phase two steady states are similar to the phase one steady states. For  $Lk_{rel}$  in the range of -1 to -10, perturbing the phase one equilibrium solution by a negative value of  $Lk_{rel}$  further drives the system by an excess in twist energy. The dynamics appear to be a continuation of the process observed in phase one. Perturbing these phase one equilibria by a positive  $Lk_{rel}$  has an uncoiling effect on the equilibrium configuration. The system is still driven by excess twist, just in the opposite direction.

## 2.5 Conclusion

Using the GIB method, we were able to determine stable equilibria of an elastic trefoil knot for various linking numbers, approximate saddle configurations, and observe the elastic energy values throughout the dynamic process. The equilibrium configurations were consistent with the results found by Coleman and Swigon [19] for torus knots. We observed similar stable knot configurations for the trefoil knot, with the exception of those that have significant self-contact. Given that the GIB method additionally incorporates electrostatic

energy, we would not expect to find configurations which come into close contact for large portions of the DNA segment to be stable.

Given this additional constraint, we found stable equilibria to share similar features of plectonemic structures with high  $\Delta Lk$  and similar symmetries. Additionally, we found many new families of stable equilibria at lower excess linking numbers, and we explored a wider range of values for excess linking number.

The benefit of finding knot energy minimizers through a dynamical process as opposed to a gradient descent algorithm is that we can find transient states and observe energy values throughout the process. From this deterministic dynamical system, we gain a broader view of the knot energy landscape as a whole.

## 3.0 Stochastic dynamics of DNA knots

### 3.1 Introduction

While exploring the dynamics of DNA knots using the GIB method gave many insights into the way in which knotted DNA behaves immersed in a fluid, we still only have a glimpse of biologically realistic dynamics using a deterministic model. At the length scale of a DNA plasmid, thermal fluctuations play a large part in the overall dynamics. To further study these dynamics, we will incorporate thermal forcing using the Stochastic Generalized Immersed Boundary (SGIB) Method [73].

The SGIB method takes into account the random thermal forcing within the fluid, and also accounts for interaction of an elastic rod immersed in fluid as it builds upon the GIB method [46, 41]. So, we will model the structure of a DNA plasmid as in the GIB and SGIB methods. As in the GIB method, we assume that a closed DNA plasmid is immersed in an ionized fluid. The double-helical structure of DNA is assumed to be a closed, homogeneous, isotropic, intrinsically straight elastic rod. Thus, the DNA has uniform elastic properties throughout the entire segment. We also assume that the charges, which lie along the sugar-phosphate backbone of DNA are moved to the central axis of the Cosserat rod.

While the immersed DNA molecule is modeled as in Chapter 1, the numerical method describing the dynamics of the fluid change for the SGIB method. Since the ratio of inertial force to viscous force, or the Reynolds number, of the fluid is low, the nonlinear advection term in the Navier-Stokes equations can be ignored. Thus the SGIB method models the fluid dynamics using incompressible Stokes flow. Additionally, there is a term added to the force density acting on the fluid which accounts for a uniformly random thermal forcing.

Building upon our results from Chapter 2, we seek to gain insights into the dynamics and energy landscape of DNA trefoil knots with excess twist. To do so, we classify the equilibrium structures found in Chapter 2 using an equivalence relation from a Kendall shape space. We then use the SGIB method to simulate knotted configurations with a given excess linking number. Viewing this as a continuous time Markov process, our goal is to

determine the mean holding time that the stochastic knot configurations spend within each basin of attraction of a stable equilibrium, find the transition rates for the CTMC, obtain the transition probabilities between these equilibria, and compare these to the energy estimates obtained in chapter 2.

### 3.2 Stochastic Generalized Immersed Boundary Method

The Stochastic Generalized Immersed Boundary (SGIB) Method, developed by Swigon et al. [73] is a stochastic extension of the GIB method [46] and the GIB method extended to DNA plasmids [41]. This stochastic extension takes into account the random thermal fluctuations within a fluid at the scale of macromolecules and cellular structures, and it accounts for torsional drag of an immersed structure in fluid. The stochastic extension used is in the style of the Stochastic IB method by Atzberger et al. [6], in which random thermal forcing is added to the fluid equations. This method is different from a Brownian dynamics approach [13, 29, 64, 65], in which the fluid has a significantly lower density than the immersed structure, known as Dissipative Particle dynamics, [7, 27, 29, 38, 35, 43, 54, 56], which works well at larger length scales.

Because the SGIB method builds upon the GIB method, the Cosserat rod equations, balancing the forces and moments of the rod remain the same. We define electrostatic and steric forces,  $\mathbf{f}^e$  and  $\mathbf{f}^c$ , in equations 2.22, and 2.21 in chapter 1 and as in the GIB method. Similarly, the force density  $\mathbf{f}^b(\mathbf{x}, t)$  is defined in equation 2.12 in chapter 2 and represents the external body force from the rod to the fluid resulting from the mechanical imbalance of the immersed DNA [46, 41]. However, for the equations of fluid motion, because the system has a low Reynolds number, the non-linear advection term is neglected, resulting in a time-dependent, incompressible, Stokes flow. Thus, the fluid flow for the SGIB method is as follows, where  $\mathbf{u}(\mathbf{x}, t)$  is the fluid velocity and  $p(\mathbf{x}, t)$  is the fluid pressure [73]:

$$\rho \mathbf{u}_t = -\Delta p + \mu \nabla \mathbf{u} + \mathbf{f}^{tot}, \quad (3.1)$$

$$\Delta \cdot \mathbf{u} = 0. \quad (3.2)$$

Now, the force density  $\mathbf{f}^{tot}(\mathbf{x}, t)$  acting on the fluid is a sum of the the external body force from the rod onto the fluid, and the thermal force:  $\mathbf{f}^{tot} = \mathbf{f}^b(\mathbf{x}, t) + \mathbf{f}^{thm}$ . The thermal force is assumed to be uniform across the domain of the fluid, and is represented in Fourier space as Gaussian white noise, where  $B_k(t)$  is standard 3-dimensional Brownian motion, and  $D_k$  is a mode-dependent forcing term [6]:

$$\hat{\mathbf{f}}^{thm} dt = \rho \sqrt{2D_k} dB_k(t). \quad (3.3)$$

Following the method from [6], to derive the numerical method, we use a finite difference scheme for the fluid equations:

$$\rho \frac{d\mathbf{u}_{\mathbf{m}}}{dt} = \frac{\mu}{\Delta x^2} (\mathbf{u}_{\mathbf{m}^-} - 2\mathbf{u}_{\mathbf{m}} + \mathbf{u}_{\mathbf{m}^+}) - \frac{1}{2\Delta x} (p_{\mathbf{m}^+} - p_{\mathbf{m}^-}) + \mathbf{f}^{total}(\mathbf{x}_{\mathbf{m}}, t), \quad (3.4)$$

where the fluid variables are on a periodic grid with length  $L$  in each direction, with  $N$  discretization points and spacing  $\Delta x = L/N$ . Similarly,  $\mathbf{m} = (m_1, m_2, m_3)$  has integer values and lies along the integer lattice, and  $\mathbf{x}_{\mathbf{m}}$  is the position of the grid point with index  $\mathbf{m}$ . Similarly,  $\mathbf{u}_{\mathbf{m}}$ , and  $p_{\mathbf{m}}$ , represent the discretized velocity and pressure at position  $\mathbf{x}_{\mathbf{m}}$ , as in [6].

If we apply a discrete Fourier transform, the system decouples into a system of ordinary differential equations:

$$\frac{d\hat{\mathbf{u}}_{\mathbf{k}}}{dt} = -\alpha_{\mathbf{k}} \hat{\mathbf{u}}_{\mathbf{k}} - i\rho^{-1} \hat{p}_{\mathbf{k}} \hat{\mathbf{g}}_{\mathbf{k}} + \rho^{-1} \hat{\mathbf{f}}_{\mathbf{k}}^{total}, \quad \hat{\mathbf{u}}_{\mathbf{k}} = \frac{1}{N^3} \sum_{\mathbf{m}} \mathbf{u}_{\mathbf{m}} \exp(-i2\pi\mathbf{k} \cdot \mathbf{m})/N, \quad (3.5)$$

$$\hat{\mathbf{u}}_{\mathbf{k}} \cdot \hat{\mathbf{g}}_{\mathbf{k}} = 0, \quad \mathbf{u}_{\mathbf{m}} = \sum_{\mathbf{k}} \hat{\mathbf{u}}_{\mathbf{k}} \exp(i2\pi\mathbf{k} \cdot \mathbf{m})/N, \quad (3.6)$$

where  $\hat{\mathbf{u}}_{\mathbf{k}}$  is the discrete Fourier transform of  $\mathbf{u}_{\mathbf{m}}$ . Here,

$$\alpha_{\mathbf{k}} = \frac{2\mu}{\rho\Delta x^2} \sum_{j=1}^3 (1 - \cos(2\pi\mathbf{k}^{(j)}/N)) \quad (3.7)$$

represents a mode-dependent decay coefficient, and the discrete transform of the second order finite difference operator is given by

$$\hat{\mathbf{g}}_{\mathbf{k}}^{(j)} = \frac{\sin(2\pi\mathbf{k}^{(j)}/N)}{\Delta x}. \quad (3.8)$$

We impose an additional constraint,  $\overline{\hat{\mathbf{u}}_{\mathbf{N}-\mathbf{k}}} = \hat{\mathbf{u}}_{\mathbf{k}}$ , to ensure we have a system of real differential equations.

Due to the incompressibility, since  $\hat{\mathbf{g}}_{\mathbf{k}}$  is orthogonal to both  $\hat{\mathbf{u}}_{\mathbf{k}}$  and  $\frac{d\hat{\mathbf{u}}_{\mathbf{k}}}{dt}$ , taking the dot product of  $\hat{\mathbf{g}}_{\mathbf{k}}$  with equation (3.5) yields:

$$\hat{p}_{\mathbf{k}} = \frac{-i\hat{\mathbf{g}}_{\mathbf{k}} \cdot \hat{\mathbf{f}}_{\mathbf{k}}^{total}}{|\hat{\mathbf{g}}_{\mathbf{k}}|^2}. \quad (3.9)$$

Thus we can use a projector to control the incompressibility:

$$\mathcal{P}_{\mathbf{k}}^{\perp} = \left( I - \frac{\hat{\mathbf{g}}_{\mathbf{k}}\hat{\mathbf{g}}_{\mathbf{k}}^T}{\|\hat{\mathbf{g}}_{\mathbf{k}}\|^2} \right). \quad (3.10)$$

Finally, we get the resulting stochastic differential equation for fluid velocity and pressure in Fourier space:

$$d\hat{\mathbf{u}}_{\mathbf{k}} = -\alpha_{\mathbf{k}}\hat{\mathbf{u}}_{\mathbf{k}}dt + \rho^{-1}\mathcal{P}_{\mathbf{k}}^{\perp}\hat{\mathbf{f}}_{\mathbf{k}}^b dt + \sqrt{2D_{\mathbf{k}}}\mathcal{P}_{\mathbf{k}}^{\perp}dB_{\mathbf{k}}(t). \quad (3.11)$$

At this step, we see that using a Stokes flow as opposed to a Navier-Stokes flow is advantageous, because if we assume the force field  $\hat{\mathbf{f}}_{\mathbf{k}}$  remains constant within one time step, the equation 3.11 is explicitly integrable over one time step using standard techniques from Itô calculus. Thus, we get a discrete update formula for  $\hat{\mathbf{u}}_{\mathbf{k}}$ :

$$\hat{\mathbf{u}}_{\mathbf{k}}((n+1)\Delta t) = e^{-\alpha_{\mathbf{k}}\Delta t}\hat{\mathbf{u}}_{\mathbf{k}}(n\Delta t) + \frac{1}{\rho\alpha_{\mathbf{k}}}(1 - e^{-\alpha_{\mathbf{k}}\Delta t})\mathcal{P}_{\mathbf{k}}^{\perp}\hat{\mathbf{f}}_{\mathbf{k}}^b(n\Delta t) + \mathcal{P}_{\mathbf{k}}^{\perp}\sigma_{\mathbf{k}}\eta_{\mathbf{k}}^n, \quad (3.12)$$

where  $\eta_{\mathbf{k}}^n$  is a 3D standard normal Gaussian random variable [6].

We can similarly derive an update formula for the position of the DNA within one time step:

$$\mathbf{X}_j((n+1)\Delta t) - \mathbf{X}_j(n\Delta t) = \sum_{\mathbf{m}} \delta_c(\mathbf{x}_{\mathbf{m}} - \mathbf{X}_j(n\Delta t)) \int_{n\Delta t}^{(n+1)\Delta t} \mathbf{u}_{\mathbf{m}}(s)ds\Delta x^3, \quad (3.13)$$

where  $\mathbf{X}_j(n\Delta t)$  is the  $j$ th component of the position vector  $\mathbf{X}$  at time  $n\Delta t$ . That is, the updated position of DNA is given by averaging integrated fluid velocities over a spatial neighborhood centered at the old position of DNA [6].

The full numerical algorithm is as follows: 1. Evaluate the balance equations of the DNA and compute the resultant forces and moments at each discretization point along the DNA; 2. Convert forces and moments external to the DNA to forces and moments acting on the fluid using the discrete delta function; 3. Update the fluid velocity using the Stokes flow in Fourier space; 4. Update the velocity of the DNA 5. Propagate the DNA through the fluid.

### 3.3 Methods

For the stochastic process, our goal was to obtain mean holding times spent near each equilibrium configuration and determine transition rates between these stable equilibria. We simulated stochastic knot dynamics using the SGIB method, classified these configurations using a Kendall shape space and Procrustes distance [40], and then estimated the mean holding times and transition probabilities for the invariant distribution.

#### 3.3.1 Kendall shape space and Procrustes analysis

From the family of equilibrium configurations obtained in chapter 2, we first partitioned the set of stable equilibria by the linking number and an equivalence relation obtained using similarity in a Kendall shape space. Two shapes,  $S^1$  and  $S^2$ , defined by a discrete set of points  $S^i = \{\mathbf{s}_1^i, \mathbf{s}_2^i, \dots, \mathbf{s}_k^i\} \in \mathbb{R}^n$ , are equivalent if they are the same up to dilation and rigid isometry. Thus, the Kendall shape space is the quotient space  $(\mathbb{S}^{n(k-1)-1}/\mathbf{SO}(\mathbf{n}), \|\cdot\|_p)$  with the Frobenius norm, and the metric  $d_p(\hat{S}^1, \hat{S}^2) = \|\hat{S}^1 - \hat{S}^2\|_p$  for equivalence classes  $\hat{S}^1, \hat{S}^2$  known as the Procrustes distance [40].

Since each of the configurations has 600 Bp, in  $\mathbb{R}^3$ , the knot equivalence classes live in the space  $(\mathbb{S}^{3(599)-1}/\mathbf{SO}(\mathbf{3}), \|\cdot\|_p)$ . We neglect the extra base pair in the Procrustes analysis, because repeated points make the distance  $d_p$  not well-defined.

We also wanted our distance in the quotient space to be invariant under cyclic permutation relabelings of the choice of DNA discretization points. Thus, we defined a new distance of configurations  $C_{\alpha,a} = \{x_1^{\alpha,a}, x_2^{\alpha,a}, \dots, x_k^{\alpha,a}\}$  and  $C_{\beta,b} = \{x_1^{\beta,b}, x_2^{\beta,b}, \dots, x_k^{\beta,b}\}$  to be

$$d_{min}(C_{\alpha,a}, C_{\beta,b}) = \min_{\sigma} d_p(C_{\alpha,a}^{\sigma}, C_{\beta,b}), \quad (3.14)$$

where  $\sigma$  is a cyclic permutation of indices  $\{1, 2, \dots, k\}$ , and  $C_{\alpha,a}^{\sigma} = \{x_{\sigma(1)}^{\alpha,a}, x_{\sigma(2)}^{\alpha,a}, \dots, x_{\sigma(k)}^{\alpha,a}\}$ .

Using this quotient space with the new distance  $d_{min}$ , we classified our family of equilibrium configurations by the change in the linking number,  $Lk_{rel}$ , and chose the equivalence class representative to be the configuration with the lowest energy value for a fixed linking number. This is because our stable configurations are obtained dynamically and thus constitute approximations to the true energy minimizers. For a dynamical process, the natural

choice of equivalence class representative is the knot configuration with the lowest elastic energy.

For a fixed linking number  $Lk_{rel}$  corresponding to a single row of figures 12 and 13, to obtain the equivalence classes and equivalence class representatives, we classified the knot configurations using a tolerance of  $\epsilon_p = 0.071$ . Thus, for a fixed relative linking number,  $Lk_{rel} = \alpha$ , two configurations  $C_{\alpha,a}$  and  $C_{\alpha,b}$  belong to the same equivalence class if  $d_{min}(C_{\alpha,a}, C_{\alpha,b}) \leq \epsilon_p$ .

### 3.3.2 Classification of stochastic configurations

Next, we ran stochastic simulations of these knot types using the SGIB method. Simulations were run using MATLAB on a cluster. For both the deterministic and stochastic configurations, we consider a closed, knotted segment of DNA, with 600 base pairs but discretize by every fourth base pair. The initial position is given by the equivalence class representatives from the classification in Kendall shape space, with a specified  $Lk_{rel}$  value. The fluid, as before, is given periodic boundary conditions, with a mesh discretization of  $64^3$ . A complete table of parameter values can be found in table 2 for the SGIB method. Note that we only considered simulations with molar salt concentration of  $Cs = 0.1M$ , because the family of configurations for  $Cs = 0.01M$  was too limited.

Our goal was to classify these stochastic configurations using Procrustes analysis and find the equilibrium configuration that the stochastic knot most closely resembled as a function of time. To do so, we classified the stochastic knot configurations,  $\mathbf{X}_\alpha(t)$ , by taking the minimum Procrustes distance from each of the stable equilibria equivalence classes for a given  $Lk_{rel} = \alpha$ . For the Procrustes analysis,  $k = 150$ , as the segment of DNA has 151 discretization points. We compared the stochastic configurations to the equivalence class representatives for a fixed linking number using  $d_{min}$ . Then, as a function of time, we classified  $\mathbf{X}_\alpha(t)$  by finding the stable equilibrium equivalence class representative,  $C_{\alpha,a}$ , that minimized  $d_{min}(\mathbf{X}_\alpha(t), C_{\alpha,a})$ . We denote this minimizing path by  $\chi_\alpha(t)$ . This allows us to think of the stochastic simulation as a one-dimensional random walk throughout the state space of stable knot configurations.

parameters	symbol	value
grid size	$N \times N \times N$	$64^3$
domain size	$L \times L \times L$	$L = 1136.366 \text{ \AA}$
time step	$\Delta t$	$7 \times 10^{-11}$
fluid density	$\rho$	$1 \text{ g cm}^3$
fluid viscosity	$\mu$	$0.01 \text{ g (cm s)}^{-1}$
permittivity of free space	$\epsilon_0$	$3.45 \times 10^{-7} e^2 / (g \text{ \AA}^3 s^{-2})$
dielectric constant of water	$\epsilon_W$	$77.7$
electric charge	$q$	$0.24 \times 2e$
molar salt concentration	$Cs$	$0.01, 0.1M$
Debye screening parameter	$\kappa$	$0.329 \sqrt{Cs} \text{ \AA}^{-1}$
bending modulus	$a_1 = a_2 = a$	$1.3 \times 10^{-19} \text{ g cm}^3 \text{ s}^{-2}$
twist modulus	$a_3$	$1.82 \times 10^{-19} \text{ g cm}^3 \text{ s}^{-2}$
shear modulus	$b_1 = b_2 = b$	$5 \times 10^{-5} \text{ g cm}^3 \text{ s}^{-2}$
stretch modulus	$b_3 = b$	$5 \times 10^{-5} \text{ g cm}^3 \text{ s}^{-2}$
number of base pairs	$n$	$150$
radius	$r$	$324.676 \text{ \AA}$
diameter	$D$	$20 \text{ \AA}$

Table 2: SGIB computational parameters

### 3.3.3 Stochastic knots as a Markov process

Recall that throughout a deterministic or stochastic simulation, the linking number of the transient knot configuration remains invariant, without strand passage of the DNA. For a fixed linking number  $Lk_{rel}$ , we think of the path through the stochastic transient states as a one-dimensional random walk throughout the stable knot configuration state space using the minimum Procrustes distance classifier. We model the stochastic process as a homogeneous, continuous time Markov process. For each fixed  $Lk_{rel}$ , our goal is to find the transition rates and mean holding times for the CTMC. We then use these to find the stationary distribution for the stochastic process, and compare this to an estimate of the Gibb's distribution determined from the deterministic elastic energy values.

For a fixed  $\alpha$ , and for a continuous time Markov process  $\chi_\alpha(t)$  over the space of equilibrium configurations,  $\{C_{\alpha,a}\}$ , the probability of transitioning from state  $b$  at time 0 to state  $a$  at time  $t$  is given by the matrix  $P(t)$  and satisfies  $P'(t) = P(t)Q$  [32]. Here,  $P(t)$  has entries  $p_{a,b} = Prob(\{\chi(t) = a | \chi(0) = b\})$ . The matrix  $Q$  is the transition rate matrix, and gives the transition rates of the underlying discrete time Markov chain. We ultimately want to find the steady state distribution  $\pi$  that satisfies  $P\pi = \pi$  and compare this to the Gibb's distribution,  $\hat{\pi}$  obtained solely from the elastic energy.

To find these, we ran the stochastic simulations for negative  $Lk_{rel}$ , and  $Cs = 0.1M$ , tracking every 20th time step until reaching an equilibrium distribution. For a fixed  $Lk_{rel} = \alpha$ , we estimate the occupancy times near each equilibrium configuration,  $C_{\alpha,a}$ , as the total time spent near  $C_{\alpha,a}$ , according to the  $d_{min}$  classification. We also find the transition probabilities of the underlying discrete time Markov process.

For a fixed linking number  $Lk_{rel} = \alpha$ , we seek the mean time spent at state  $a$ ,  $T_a$ , the transition rates,  $Q$ , and the stationary distribution  $\pi$ . We found the mean time  $T_a$  as the sum of the occupancy times across every initial starting configuration. For a continuous time Markov process, the transition time from one state to another follows a Poisson process. The transition rates given by matrix  $Q$ , and the mean time spent in state  $a$  satisfy [32]:

$$T_a = -1/q_{aa}. \tag{3.15}$$

We then estimated the total transition probabilities  $P_{a,b}$  for the underlying discrete time Markov process across every initial starting configuration, and we used the occupancy time to obtain the remaining transition rates using the relation [32]:

$$P_{ab} = -\frac{q_{ab}}{q_{bb}}. \quad (3.16)$$

Note that the stationary distribution  $\pi$  also satisfies  $Q\pi = 0$ . After estimating the transition rate matrix  $Q$ , we solved for the stationary distribution  $\pi$ . Finally, we were able to estimate the Gibb's distribution from the deterministic system,  $\hat{\pi}$ , solely from the elastic energy of a given state,  $\phi_i$  [32]:

$$\hat{\pi}_i = \frac{\exp(-\phi_i)}{\sum_j \exp(-\phi_j)}. \quad (3.17)$$

## 3.4 Results

### 3.4.1 Procrustes analysis

Using the equivalence relation from the Kendall shape space, we classified the stable knot equilibria for  $Lk_{rel} = \pm 1, \pm 2, \dots, \pm 10$ , and  $Cs = 0.1M$ . Figures 14 and 16 show the collection of equilibrium configurations for the negative and positive  $Lk_{rel}$  parameter range respectively. These graphs show the 3-dimensional projection of knots in the Kendall shape space, with axes representing a closest fit projection of the Procrustes distance between the configurations.

We see that knot types which appear to be similar under rigid isometry and rotations are clustered together under the classical multidimensional scaling 3-dimensional projection. In particular, the projection captures key characteristics that are unique to each knot type. The families of knotted configurations which resemble the standard trefoil parametrization are clustered together. The family of compact, pretzel-like configurations are clustered as expected, and those with long plectonemes with either midpoint or terminal loops are similarly grouped. This is consistent for positive and negative values of  $Lk_{rel}$ .

One disadvantage of looking at the 3-dimensional projection is that some key information from the minimum Procrustes distance clustering is lost. Some knot types appear farther

apart under the projection. See for example the knots labeled 44, 54, and 64. Visually, we would like to classify these as belonging to the same family, but they appear far apart after the projection.

To compensate for this, we also used a dendrogram to illustrate the clustering by Procrustes distance. Within the energy landscape of knot configurations, the minimum Procrustes distance classification provides barriers between energy wells. Figures 15 and 17 show the dendrogram for negative and positive values of  $Lk_{rel}$  respectively. These dendrograms are trees constructed from the minimum Procrustes distance between configurations. Each is constructed by finding the distance between clusters of objects. So the distance between two configurations is given by the lowest shared node in the graph. The x-axis of the dendrogram represents the configurations labeled both as ordered pairs, and as an integer from 1 to 80 as a reference for the 3-dimensional projected knots. In the ordered pair labeling,  $C_{\alpha,a} = (\alpha, a)$ ,  $\alpha$  represents the final  $Lk_{rel}$  and  $a$  represents the starting  $Lk_{rel} = a$  configuration. From the dendrogram, we see that the minimum Procrustes distance closely corresponds to similarity of configurations as defined by the Kendall shape space. For example, in figure 15, the entire collection of knots with final  $Lk_{rel} = -1, -2$  all share a common node with  $d_{min} < 0.02$ , which aligns with our intuition. Because all of these very closely resemble the starting trefoil parametrization (and one another), we would expect them to be similar in the Kendall shape space.

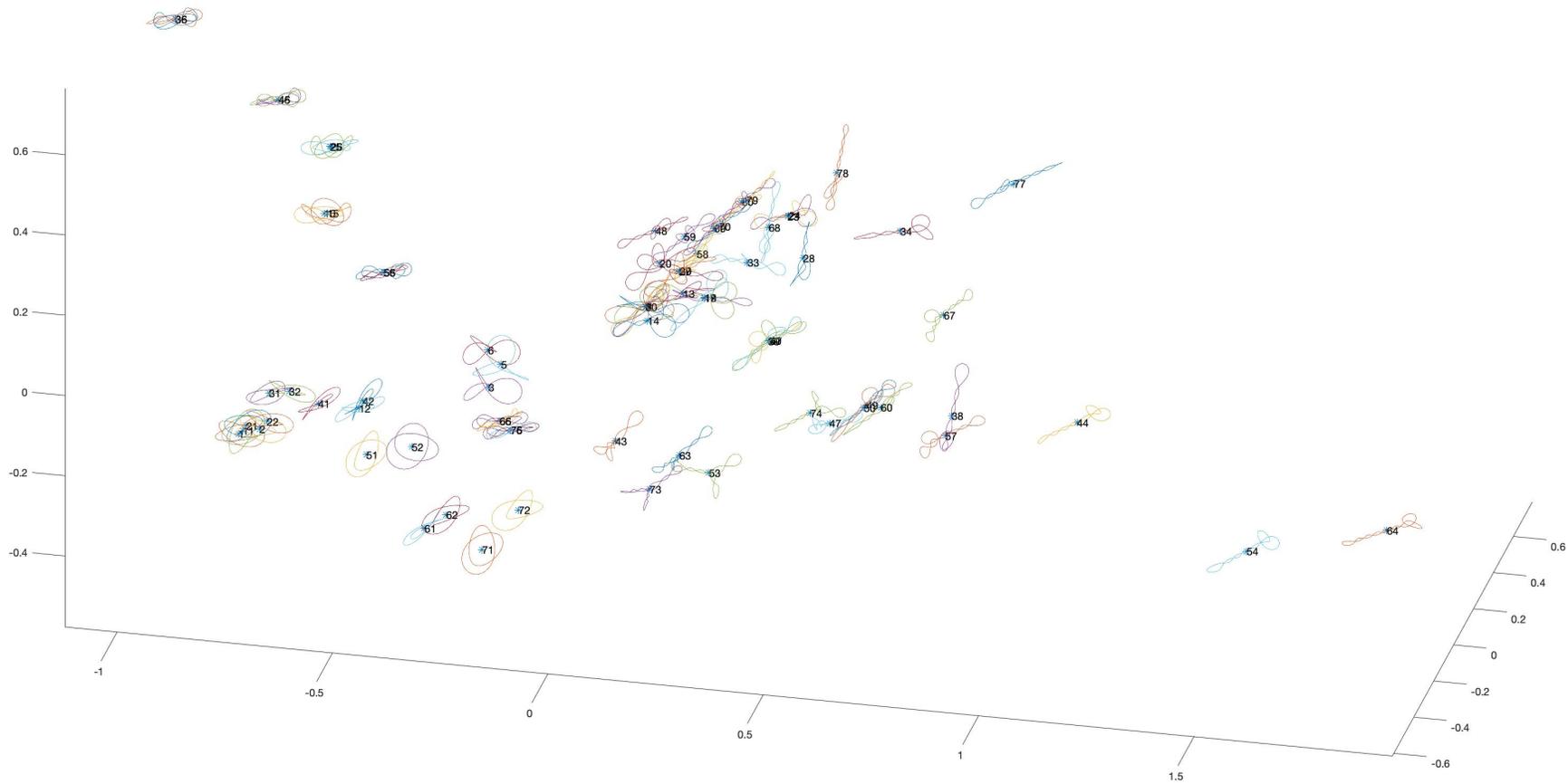


Figure 14: 3-Dimensional projection of knot configurations in Kendall shape space for negative  $Lk_{rel}$ .

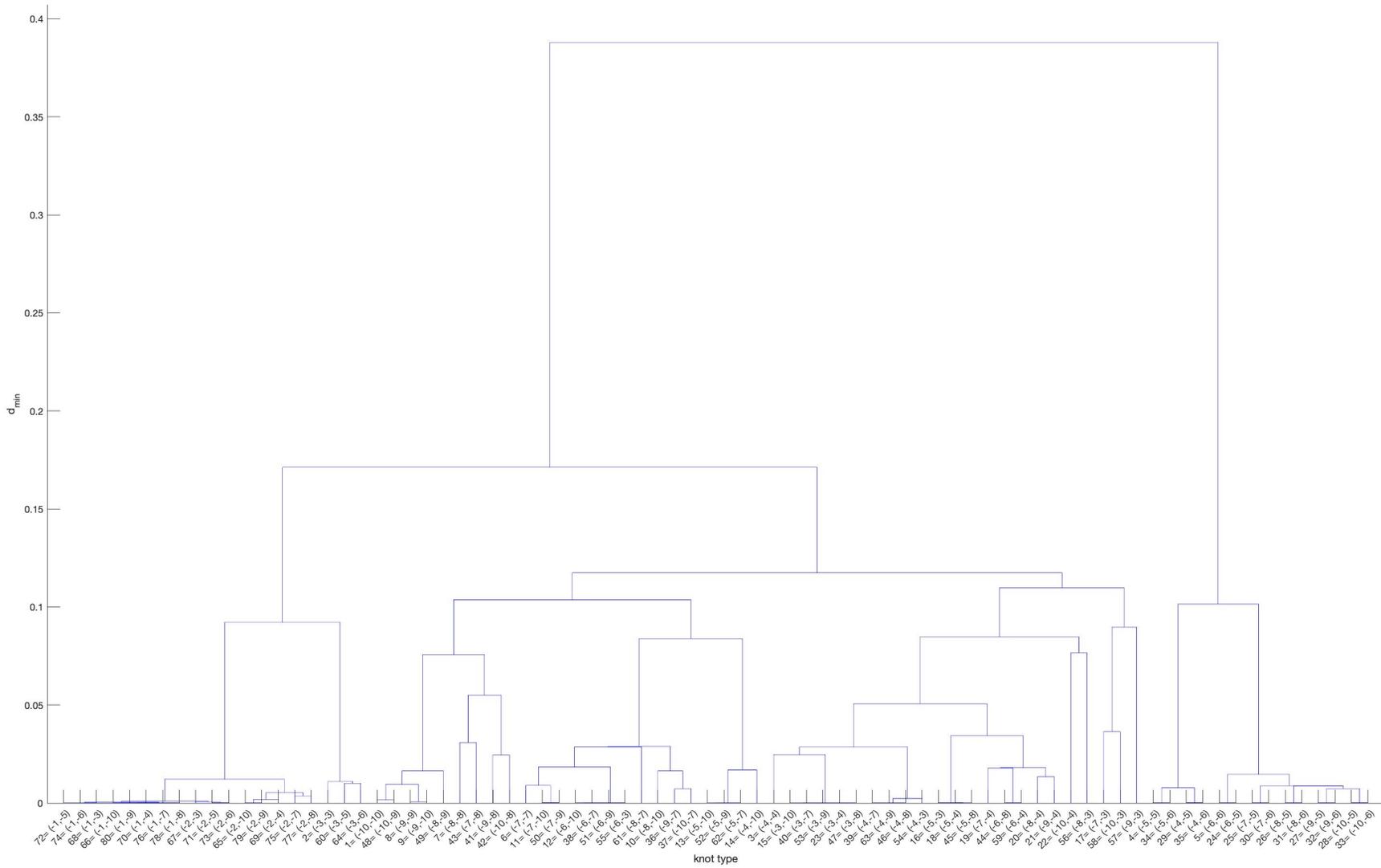


Figure 15: Dendrogram for negative  $Lk_{rel}$  knot configurations representing the minimum Procrustes distance between equilibria.

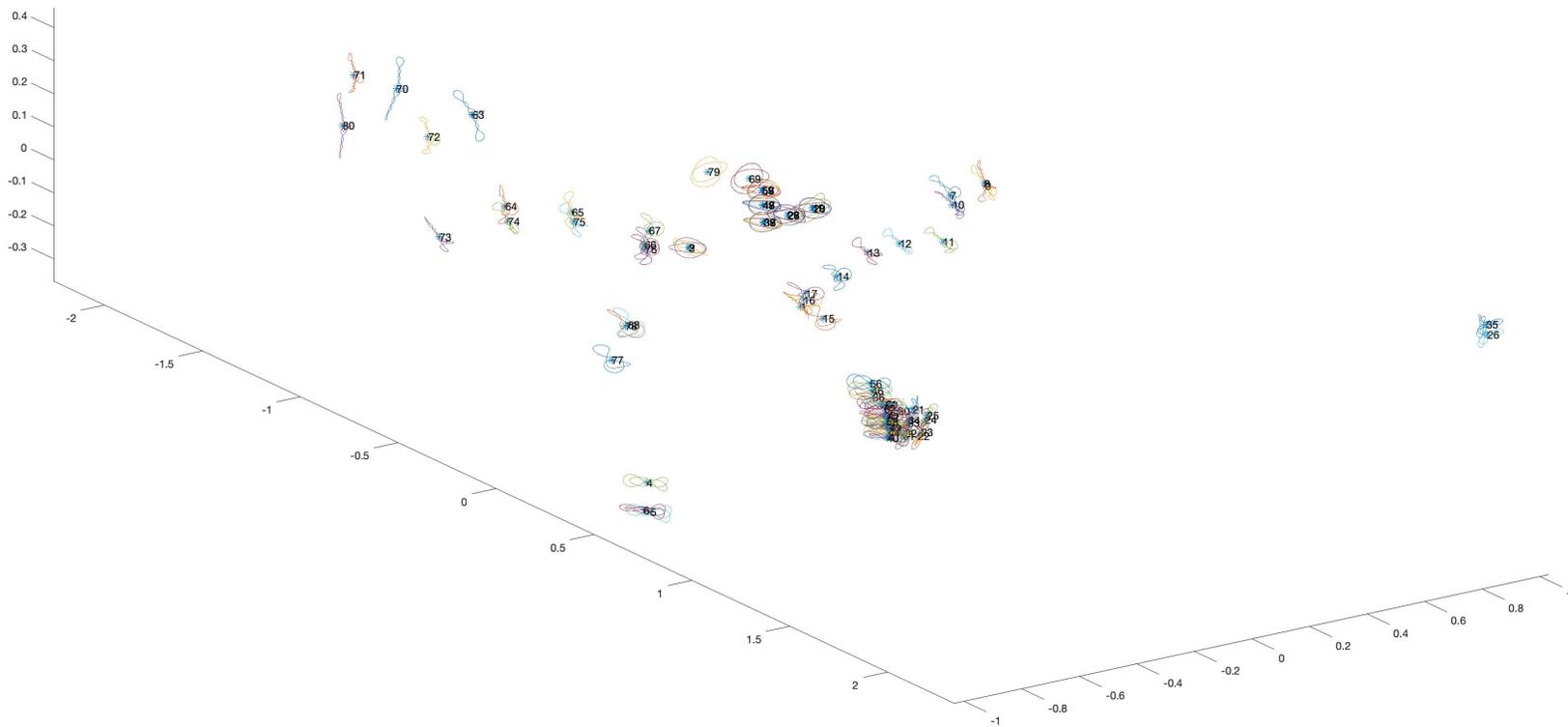


Figure 16: 3-Dimensional projection of knot configurations in Kendall shape space for positive  $Lk_{rel}$ .

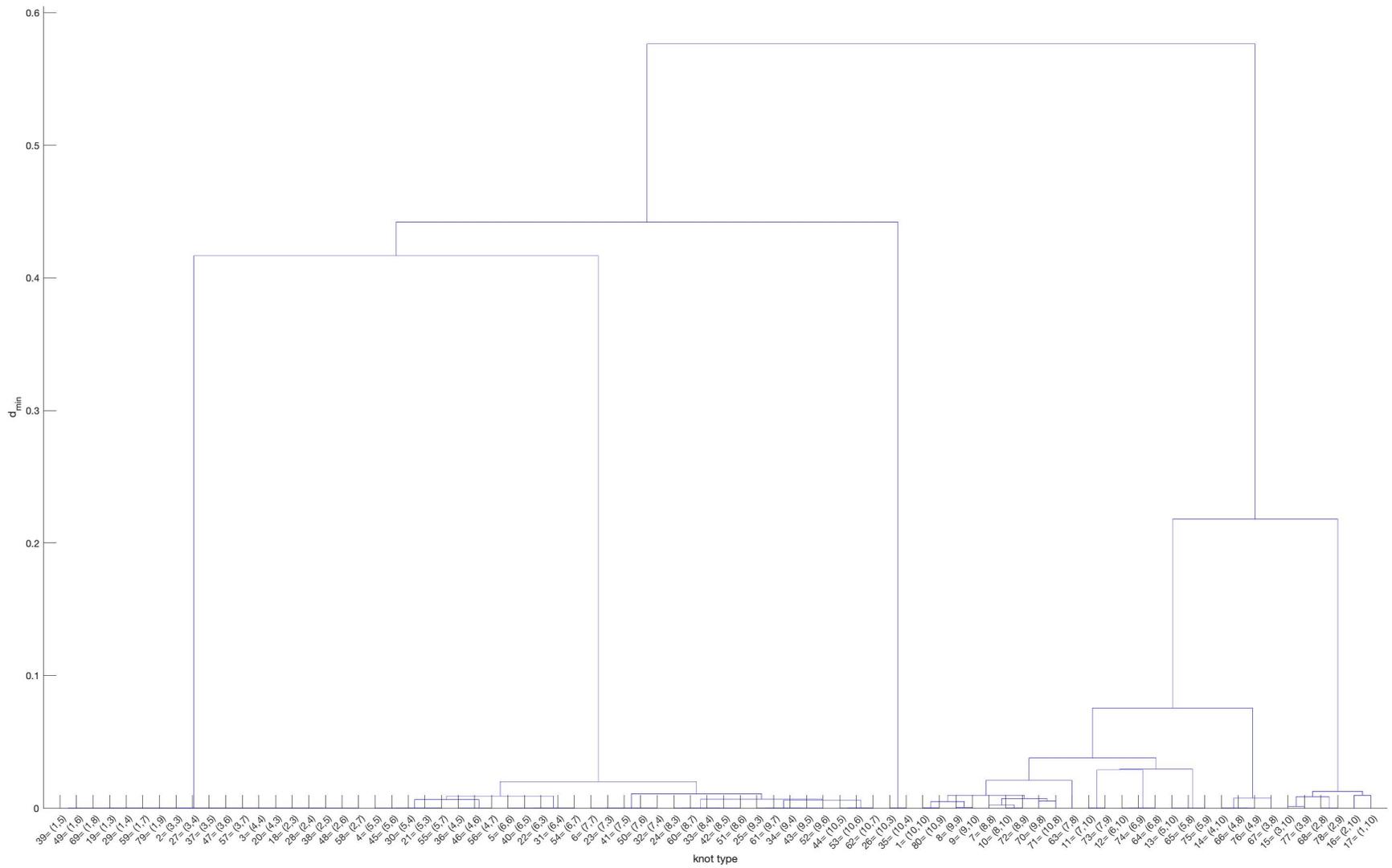


Figure 17: Dendrogram for positive  $Lk_{rel}$  knot configurations representing the minimum Procrustes distance between equilibria.

Using the minimum Procrustes distance, we classified the stable equilibrium configurations by equivalence classes for each fixed  $Lk_{rel}$ . While the Procrustes analysis can classify configurations across linking number, we chose to first partition by linking number, because the linking number remains invariant in a dynamic simulation without DNA strand passage. Using the dendrograms and based on the minimum Procrustes distance, we classified configurations with a tolerance of  $\epsilon_p = 0.071$ . This gives equivalence classes for positive and negative  $Lk_{rel}$ . The equivalence class representatives are listed in table 3 for each fixed  $Lk_{rel}$ .

From the minimum Procrustes distance classification and the knot tables in figures 12 and 13, we see that the representatives of each equivalence class have the lowest energy values. This corresponds to these configurations being the closest approximation to the stable equilibria when compared to other knots. For negative  $Lk_{rel}$ , generally we see more diversity of knot types for higher values of  $|Lk_{rel}|$ , with  $Lk_{rel} = -1, -2$  only having one equivalence class, and  $Lk = -8$  having 7 distinct classes. To illustrate the similarity of equivalence classes, for  $Lk_{rel} = -3$ , the equivalence classes contain the following knot types:  $[C_{-3,-3}] = \{C_{-3,-3}, C_{-3,-5}, C_{-3,-6}\}$  and  $[C_{-3,-9}] = \{C_{-3,-4}, C_{-3,-7}, C_{-3,-8}, C_{-3,-9}, C_{-3,-10}\}$ . Notice that for the class  $[C_{-3,-9}]$  all of the equilibria are nearly identical in the Kendall shape space, while the other two equilibria in  $[C_{-3,-3}]$  are asymmetric, compared to  $C_{-3,-3}$  which has  $D_3$  symmetry. We chose to classify these as the same knot type class, because they have a small pairwise  $d_{min}$  distance, and  $C_{-3,-5}$  and  $C_{-3,-6}$  likely only differ from  $C_{-3,-3}$ , because they are farther from convergence to the true equilibrium configuration.

After classifying the deterministic knot configurations, we similarly classified the stochastic knot configurations using the minimum Procrustes distance and the equivalence classes for each fixed  $Lk_{rel}$ . Figure 18 gives an example of stochastic knot configurations for  $Lk_{rel} = -8$ , and their corresponding classification. The time traces show the  $d_{min}$  value for each of the stable equilibrium configurations when compared to the stochastic simulation. We also display the classified state underneath the traces, represented as the minimizing state in the legend.

$Lk_{rel}$	Knot Types	$Lk_{rel}$	Knot Types
-1	$C_{-1,-5}$	1	$C_{1,8}, C_{1,10}$
-2	$C_{-2,-5}$	2	$C_{2,6}, C_{2,8}$
-3	$C_{-3,-3}, C_{-3,-9}$	3	$C_{3,6}, C_{3,8}, C_{3,9}$
-4	$C_{-4,-6}, C_{-4,-7}, C_{-4,-10}$	4	$C_{4,3}, C_{4,5}, C_{4,10}$
-5	$C_{-5,-6}, C_{-5,-7}, C_{-5,-8}$	5	$C_{5,6}, C_{5,8}$
-6	$C_{-6,-5}, C_{-6,-7}, C_{-6,-8}$	6	$C_{6,5}, C_{6,8}$
-7	$C_{-7,-3}, C_{-7,-4}, C_{-7,-6}, C_{-7,-8}, C_{-7,-9}$	7	$C_{7,5}, C_{7,8}, C_{7,9}$
-8	$C_{-8,-3}, C_{-8,-4}, C_{-8,-6}, C_{-8,-7}, C_{-8,-8}, C_{-8,-9}, C_{-8,-10}$	8	$C_{8,5}, C_{8,8}$
-9	$C_{-9,-3}, C_{-9,-4}, C_{-9,-5}, C_{-9,-7}, C_{-9,-8}, C_{-9,-10}$	9	$C_{9,3}, C_{9,10}$
-10	$C_{-10,-3}, C_{-10,-4}, C_{-10,-6}, C_{-10,-7}, C_{-10,-8}, C_{-10,-9}$	10	$C_{10,3}, C_{10,5}, C_{10,8}$

Table 3: Knot configuration equivalence class representatives for positive and negative  $Lk_{rel}$ .

### 3.4.2 Markov process of stochastic knots

After classifying the stochastic knot configurations, we are able to frame the overall stochastic process as a continuous time Markov chain through the space of knot configurations. The minimum Procrustes distance classification gives an indication of energy barriers between the stable equilibria for a fixed  $Lk_{rel}$ .

For fixed  $Lk_{rel} = -3, -4, \dots, -10$ , we first calculated the mean holding times,  $T_i$ , and used this and the transition probabilities to find the transition rate matrix,  $Q$  for the CTMC. Lastly, from  $Q$ , we were able to obtain the stationary distribution  $\pi$  and compare this to the Gibb's distribution,  $\hat{\pi}$  generated from the elastic energies found in chapter 2.

For each  $Lk_{rel}$ , our final results are as follows.

$Lk_{rel} = -3$ , with states  $[C_{-3,-3}, C_{-3,-9}]$ :

$$Q = 1 \cdot 10^8 \cdot \begin{pmatrix} -1.0037 & 0.0505 \\ 1.0037 & -0.0505 \end{pmatrix}, \quad \pi = \begin{pmatrix} 0.0479 \\ 0.9521 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.3820 \\ 0.6180 \end{pmatrix}. \quad (3.18)$$

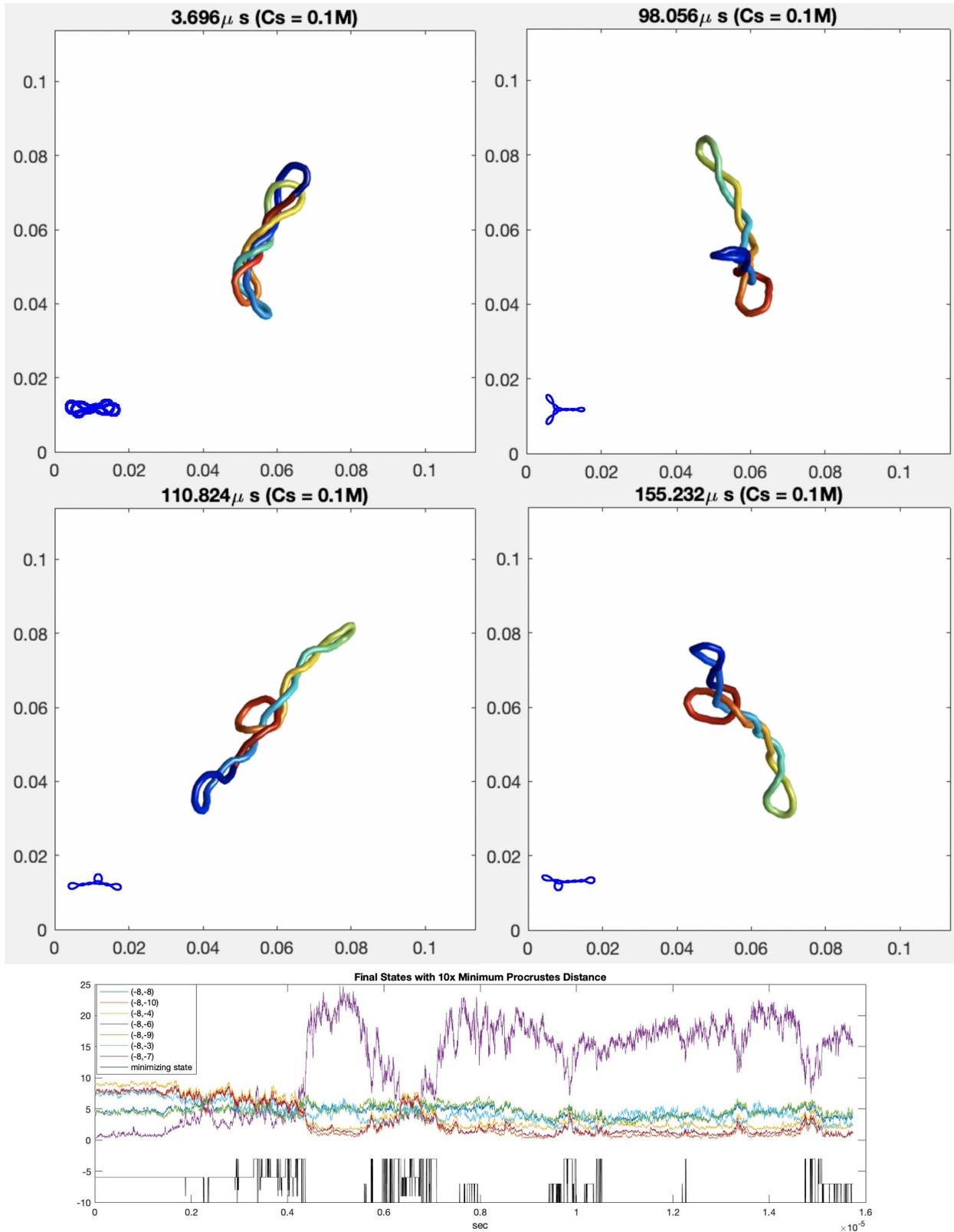


Figure 18: Top:  $Lk_{rel} = -8$  stochastic knot configurations with corresponding  $d_{min}$  equilibrium configuration classifications. Bottom: traces of  $10 \times d_{min}$  to each knot class representative with state classification denoted by minimizing state.

$Lk_{rel} = -4$ , with states  $[C_{-4,-10}, C_{-4,-6}, C_{-4,-8}]$ :

$$Q = 1 \cdot 10^7 \cdot \begin{pmatrix} -5.4058 & 1.5106 & 0.1601 \\ 0.5045 & -3.6685 & 0.0215 \\ 4.9013 & 2.1580 & -0.1816 \end{pmatrix}, \quad \pi = \begin{pmatrix} 0.0312 \\ 0.0099 \\ 0.9589 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0275 \\ 0.0368 \\ 0.9357 \end{pmatrix}. \quad (3.19)$$

$Lk_{rel} = -5$ , with states  $[C_{-5,-6}, C_{-5,-8}, C_{-5,-7}]$ :

$$Q = 1 \cdot 10^7 \cdot \begin{pmatrix} -1.8279 & 0.0425 & 0.3055 \\ 0.5156 & -0.7357 & 2.6396 \\ 1.3123 & 0.6932 & -2.9451 \end{pmatrix}, \quad \pi = \begin{pmatrix} 0.0507 \\ 0.7501 \\ 0.1992 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0008 \\ 0.9767 \\ 0.0225 \end{pmatrix}. \quad (3.20)$$

$Lk_{rel} = -6$ , with states  $[C_{-6,-5}, C_{-6,-7}, C_{-6,-8}]$ :

$$Q = 1 \cdot 10^7 \cdot \begin{pmatrix} -2.6333 & 0.9537 & 0.1001 \\ 1.7555 & -3.2813 & 0.3838 \\ 0.8778 & 2.3276 & -0.4839 \end{pmatrix}, \quad \pi = \begin{pmatrix} 0.0785 \\ 0.1341 \\ 0.7874 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0000 \\ 0.0225 \\ 0.9775 \end{pmatrix}. \quad (3.21)$$

$Lk_{rel} = -7$ , with states  $[C_{-7,-3}, C_{-7,-4}, C_{-7,-6}, C_{-7,-8}, C_{-7,-9}]$ :

$$Q = 1 \cdot 10^7 \cdot \begin{pmatrix} -4.7778 & 0.0533 & 0.2397 & 1.5110 & 0.6492 \\ 1.4333 & -0.6177 & 0 & 2.8159 & 4.5898 \\ 0.1433 & 0 & -1.9175 & 0.7555 & 0.0151 \\ 1.0989 & 0.0749 & 1.4382 & -8.5165 & 0.6492 \\ 2.1022 & 0.4895 & 0.2397 & 3.4341 & -5.9033 \end{pmatrix},$$

$$\pi = \begin{pmatrix} 0.0289 \\ 0.8475 \\ 0.0108 \\ 0.0201 \\ 0.0927 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0000 \\ 0.9930 \\ 0.0000 \\ 0.0000 \\ 0.0069 \end{pmatrix}. \quad (3.22)$$

$Lk_{rel} = -8$ , with states  $[C_{-8,-8}, C_{-8,-10}, C_{-8,-4}, C_{-8,-6}, C_{-8,-9}, C_{-8,-3}, C_{-8,-7}]$ :

$$Q = 1 \cdot 10^8 \cdot \begin{pmatrix} -1.1103 & 0.0009 & 0.0025 & 0.0650 & 0.1163 & 0.0280 & 0.0397 \\ 0.0123 & -0.3132 & 0.0098 & 0.0150 & 0.5584 & 0.4629 & 0.4777 \\ 0.1727 & 0.0281 & -0.0449 & 0 & 0.0349 & 0.1305 & 0.1913 \\ 0.1850 & 0.0017 & 0 & -0.2849 & 0.2210 & 0.0559 & 0 \\ 0.1604 & 0.0199 & 0.0003 & 0.0950 & -1.1284 & 0.0497 & 0 \\ 0.0987 & 0.0645 & 0.0056 & 0.1100 & 0.1978 & -0.9849 & 0.0857 \\ 0.4811 & 0.1981 & 0.0266 & 0 & 0 & 0.2579 & -0.7943 \end{pmatrix},$$

$$\pi = \begin{pmatrix} 0.0074 \\ 0.2144 \\ 0.6338 \\ 0.0180 \\ 0.0078 \\ 0.0297 \\ 0.0888 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0000 \\ 0.0024 \\ 0.9931 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0045 \end{pmatrix}. \quad (3.23)$$

$Lk_{rel} = -9$ , with states  $[C_{-9,-10}, C_{-9,-4}, C_{-9,-5}, C_{-9,-7}, C_{-9,-8}, C_{-9,-3}]$ :

$$Q = 1 \cdot 10^7 \cdot \begin{pmatrix} -5.9362 & 0.0073 & 0.3695 & 0.3415 & 2.1408 & 1.3660 \\ 0.0487 & -0.5476 & 0 & 0.7358 & 0.4773 & 0.1102 \\ 0.1622 & 0 & -4.0025 & 0 & 0.1880 & 0.4517 \\ 1.3624 & 0.4405 & 0.0616 & -2.2521 & 0.6943 & 2.6549 \\ 2.3680 & 0.0827 & 0.9852 & 0.1992 & -5.4533 & 1.4541 \\ 1.9949 & 0.0170 & 2.5862 & 0.9756 & 1.9528 & -6.0368 \end{pmatrix},$$

$$\pi = \begin{pmatrix} 0.0683 \\ 0.4621 \\ 0.0177 \\ 0.2745 \\ 0.0767 \\ 0.1007 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0000 \\ 0.9986 \\ 0.0000 \\ 0.0014 \\ 0.0000 \\ 0.0000 \end{pmatrix}. \quad (3.24)$$

$Lk_{rel} = -10$ , with states  $[C_{-10,-4}, C_{-10,-6}, C_{-10,-7}, C_{-10,-8}, C_{-10,-9}, C_{-10,-3}]$ :

$$Q = 1 \cdot 10^7 \cdot \begin{pmatrix} -8.3453 & 1.8765 & 0.7743 & 5.0731 & 1.8614 & 2.1494 \\ 0.4576 & -2.5553 & 0.0018 & 0.0290 & 0 & 0.1019 \\ 3.8855 & 0.1996 & -1.0254 & 1.9423 & 1.8930 & 0.1834 \\ 1.5345 & 0.0399 & 0.1184 & -8.4937 & 0.9465 & 0.2343 \\ 0.5474 & 0.0399 & 0.0955 & 0.7827 & -6.2784 & 0.5603 \\ 1.9203 & 0.3993 & 0.0354 & 0.6667 & 1.5775 & -3.2292 \end{pmatrix},$$

$$\pi = \begin{pmatrix} 0.1284 \\ 0.0284 \\ 0.6543 \\ 0.0396 \\ 0.0363 \\ 0.1129 \end{pmatrix}, \quad \hat{\pi} = \begin{pmatrix} 0.0005 \\ 0.0000 \\ 0.9993 \\ 0.0000 \\ 0.0000 \\ 0.0001 \end{pmatrix}. \quad (3.25)$$

The transition rate diagram for  $Lk_{rel} = -3$  is shown in figure 19. The time traces, mean holding times, stationary distribution from the CTMC,  $\pi$ , and Gibb's distribution from the elastic energies are displayed in figure 20. The time traces at the top display the minimum Procrustes distance from the two stable equilibria  $C_{-3,-3}$ , and  $C_{-3,-9}$ , with a graph of the corresponding minimizing state as determined by the Procrustes analysis. The time trace plot on the left was started from the state  $C_{-3,-3}$ , and the right was started from the state  $C_{-3,-9}$ . One can see that regardless of the starting state, the system quickly transitions to the  $C_{-3,-9}$  state, and spends the majority of the time in this state.

The bottom panel displays the mean holding times and stationary distributions. We see that the mean holding times closely correspond with both the time traces, and the stationary distribution  $\pi$ . The jump process from one state into another for a CTMC is a Poisson process. This gives information about the transition rates, and specifically the values of  $q_{i,i}$  for states  $i$ . According to the transition rates, we expect the system to transition more frequently from state  $C_{-3,-3}$  than from  $C_{-3,-9}$ . Thus the barrier from  $C_{-3,-9}$  to  $C_{-3,-3}$  is much lower than the barrier from  $C_{-3,-3}$  to  $C_{-3,-9}$ .

The two distributions give us information about the system as a whole compared to the deterministic system. The Gibb's distribution generated from the energies of the two stable equilibria,  $\hat{\pi}$  only accounts for the local minimum value of an energy well. According to  $\hat{\pi}$ , we would expect the CTMC to spend roughly 61% of the time near state  $C_{-3,-9}$  and 39% of the time near state  $C_{-3,-3}$ . However, from the traces, and the mean holding times, we see that this is not the case. One possible explanation would be that the traces observed are transient, and not representative of the steady state; however these very closely align with the stationary distribution  $\pi$ . Thus, the distribution  $\pi$  is capturing properties about the dynamics of the system that the energy values of stable knot configurations alone do not account for.

By generating  $\pi$  from the CTMC,  $\pi$  more accurately captures properties of the energy landscape. While this distribution accounts for the energy values, it also accounts for the free energy of the system. Comparing  $\pi$  to  $\hat{\pi}$ , we see that the state  $C_{-3,-9}$  has a much higher probability of occurring, and  $C_{-3,-3}$  has a much lower probability of occurring in the stochastic process than the deterministic system. Thus  $C_{-3,-9}$  has a relatively high entropy, and a lower free energy. Entropy gives a measure of disorder of the equilibrium state, meaning  $C_{-3,-3}$  has a narrow energy well. One possible explanation for this is  $C_{-3,-3}$  has  $D_3$  symmetry. To transition to  $C_{-3,-9}$  with  $C_2$  symmetry means breaking the symmetry with respect to any axis of symmetry. Symmetry breaking is an example of higher disorder, so this state appears with higher probability, and there are more ways for the system to break symmetry than to add symmetry to a knot configuration.

We have discussed the stochastic process for  $Lk_{rel} = -3$  in full detail. The transition rates and mean holding times for each  $Lk_{rel}$  can be found in appendix A. Figure 21 details the stationary distribution for all eight linking numbers,  $Lk_{rel} = -3, -4, \dots -10$ . For  $Lk = -4$ , the two distributions agree fairly well. Both  $\pi$  and  $\hat{\pi}$  predict a high probability of the configuration with two loops and one plectoneme as the most likely configuration,  $C_{-4,-8}$ , and the two other stable configurations have a low probability of occurring. This indicates that the depth of the energy wells alone is a good indication of the overall behavior of the stochastic process.

Consistently, we observe that the knot configuration family with two loops and one

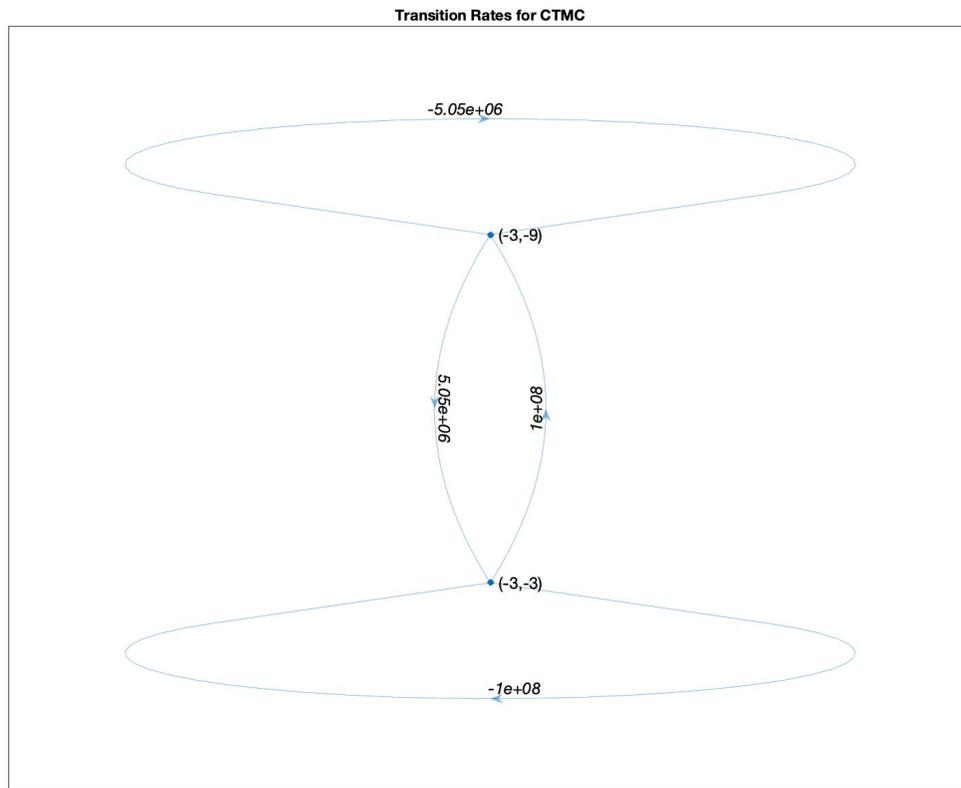


Figure 19: Transition rates for  $Lk_{rel} = -3$ .

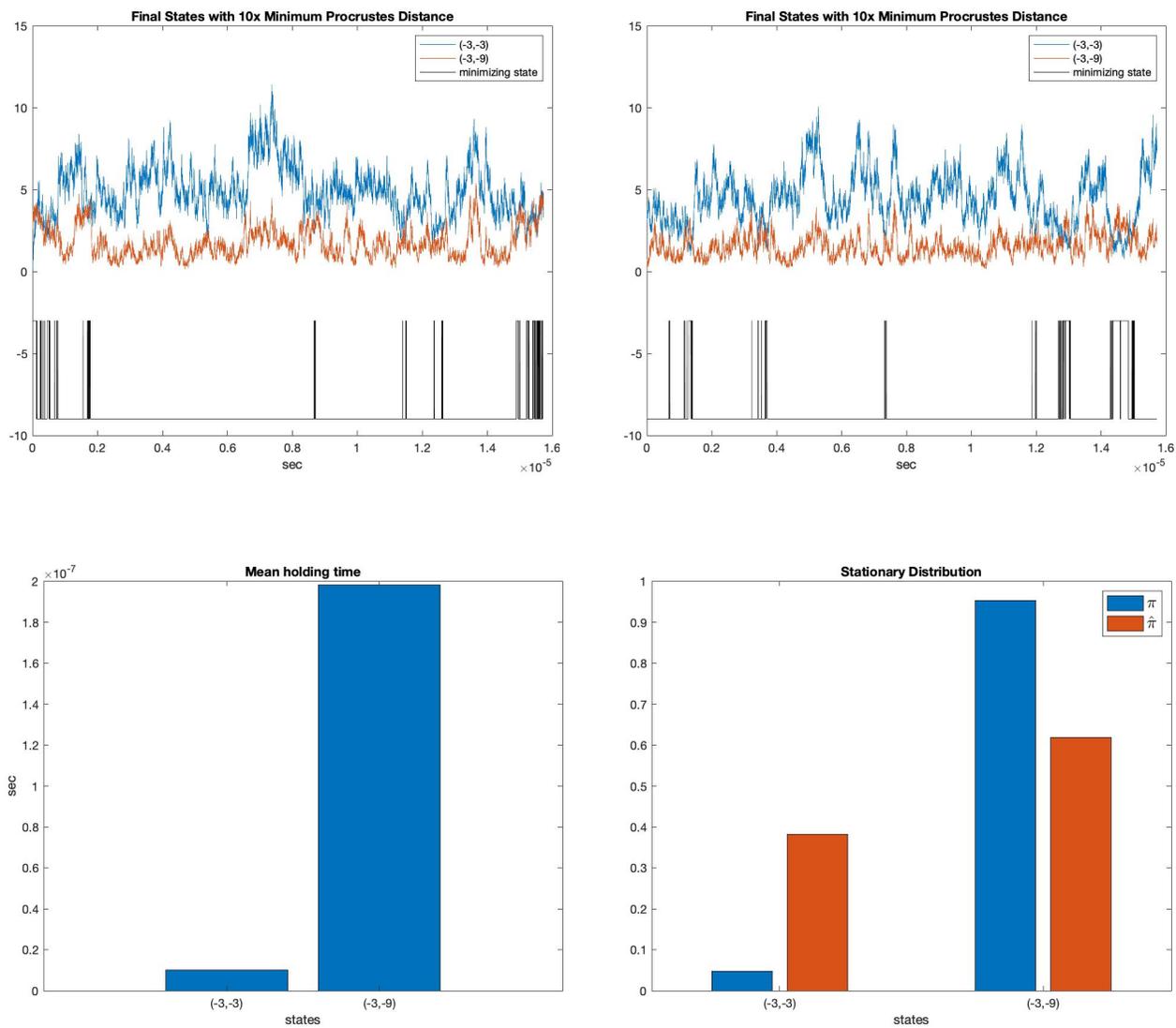


Figure 20: Top panels: Traces of  $10 \times d_{min}$  with minimizing state starting at the state  $C_{-3,-3}$  on the left and  $C_{-3,-9}$  on the right. Bottom left panel: mean holding times in seconds for states  $C_{-3,-3}$  and  $C_{-3,-9}$ . Bottom right panel: Comparison of stationary distributions  $\pi$  (blue) and  $\hat{\pi}$  (orange) for  $Lk_{rel} = -3$ .

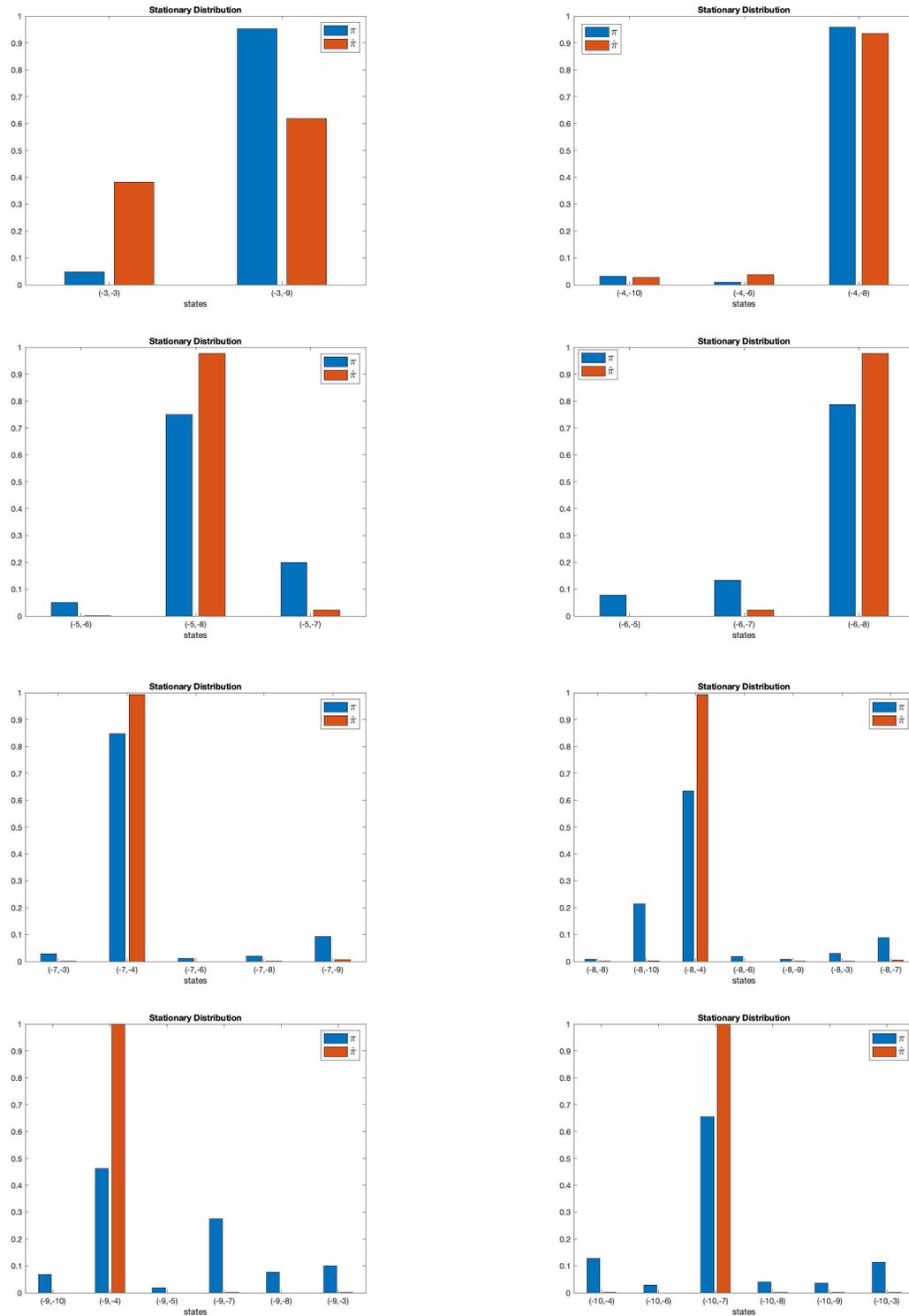


Figure 21: Stationary distribution  $\pi$  (blue) and Gibb's distribution  $\hat{\pi}$  (orange) for  $Lk_{rel} = -3, -4, -5, -6, -7, 8, -9, -10$  from top the top left panel to the bottom right panel.

plectoneme occur with high probability across all linking numbers. This family is the  $C_{\alpha,-4}$  for  $\alpha \geq -9$  branch, and  $C_{\alpha,-8}$  for  $\alpha \in \{-3, -4, -5, -6\}$  configurations. Along this branch, as the excess twist increases, so does the elastic energy. As the elastic energy increases, the Gibb's distribution  $\hat{\pi}$  overestimates the probability of encountering this configuration in the CTMC.

The pretzel family of configurations has highest energy values across linking numbers. These are the  $C_{\alpha,-5}$  and  $C_{\alpha,-6}$  for  $\alpha \leq -4$  configurations. This has a very low probability for both  $\pi$  and  $\hat{\pi}$ , however the distribution  $\hat{\pi}$  consistently underestimates the entropic contribution of the system, as  $\pi$  is slightly, but consistently higher for these configurations.

The next likely configuration across all  $Lk_{rel}$  is the family with a central loop and two plectonemes. These are the configurations  $C_{\alpha,-10}$  for  $\alpha \in \{-4, -5, -6, -7, -8\}$  and  $C_{\alpha,-9}$  for  $\alpha \in \{-5, -6, -7\}$ . Notice that the deterministic system vastly underestimates the probability of observing this state in the CTMC.

The other asymmetric states with either a central pretzel and two plectonemes, or states with 3 plectonemes having either  $C_2$  or  $D_3$  symmetry occur with low probability in the CTMC across all  $Lk_{rel}$ . The probability of observing these states is also consistently underestimated by the distribution  $\pi$ .

### 3.5 Conclusion

Using the stochastic generalized immersed boundary method, we modeled the dynamics of DNA knots with excess twist. Using a stochastic model gives more biologically accurate framework. This also allows us to model the dynamics of DNA knots using a continuous time Markov chain.

Using the stable equilibrium configurations obtained in chapter 2, we partitioned the space of configuration using Kendall shape space and a minimum Procrustes distance equivalence relation. These equilibria are the states for the CTMC. We then classified the stochastic knots using the minimum Procrustes distance. This classification provides boundaries within the energy landscape of elastic knots with excess twist. After finding the  $d_{min}$  classification,

we found transition rates, mean holding times, and stationary distributions for stochastic knots with a fixed linking number. Finally, we compared the stationary distribution  $\pi$  to the distribution  $\hat{\pi}$ .

Overall, we found the stochastic process to more accurately represent the dynamics of the DNA knots, as it accounts for the entropy of the system. Thus, using the framework of a continuous time Markov process gives us a better understanding of the energy landscape of DNA knots than the deterministic system alone.

## 4.0 Dynamics of DNA links

### 4.1 Introduction

Kinetoplastids are parasitic organisms with a unique mitochondrial DNA structure, known as a kinetoplast. This kinetoplast DNA, or kDNA, consists of a collection of approximately 5000 DNA minicircles and 20 to 30 maxicircles linked in a sparse chainmail-like structure [68]. The formation of this structure is not completely understood. The role of histone link proteins or structure preserving enzymes has been investigated in [39, 82]. Similarly, topological models have been used to better understand the network topology of kDNA [2, 25, 15, 26, 24, 51, 48, 47, 28]. We seek to model a small network of kDNA minicircles dynamically in order to obtain a distribution of the distance between centroids of two such minicircles.

Studying the dynamics of molecular structures often requires making simplifying assumptions in order to gain insights into the behavior of the system as a whole, while keeping the model computationally feasible. While this approach often captures key features of the molecular system, these simplifying assumptions may omit critical details about the system as a whole. In the case of dynamics of DNA segments, plasmids, knots, and links, the particular DNA sequence changes the kinematic properties of the molecule. The Immersed Boundary method [70] and Generalized Immersed Boundary Method [46] provide a model and numerical method that accounts for the fluid-structure interaction of a molecule immersed in fluid. In [41], Lim et al. applied this to DNA plasmids, while accounting for the electrostatic repulsion and hardcore potential of a DNA molecule.

In chapter 2, we studied the dynamics of DNA trefoil knots using this extension of the GIB method. To do so, we had to assume that the double-helical structure of DNA behaved like a homogeneous, isotropic, intrinsically straight elastic rod. While Cosserat rod theory provides a nice framework for the dynamics of DNA, these assumptions neglect the kinematics of a particular DNA sequence and also do not account for thermal forcing of the fluid.

The Stochastic Immersed Boundary Method [6] provides a way of incorporating uniform thermal forcing throughout the fluid to study the fluid-structure interaction at a microscopic scale. Swigon et al. created the Stochastic Generalized Immersed Boundary Method [73] to incorporate the effects of uniform thermal forcing while studying the dynamics of DNA plasmids. In chapter 3, we applied the SGIB method to model the dynamics of DNA knots immersed in a fluid under the assumption that DNA behaves like a homogeneous, isotropic elastic rod.

The exact sequence of DNA has been shown to affect the elastic properties of the DNA molecule [59, 44, 33]. Coleman et al. created a sequence dependent model of DNA elasticity [16] and found equilibrium configurations of 150 Bp DNA o-rings. While this theory incorporates sequence dependence of DNA, it does so by minimizing the total elastic energy of the DNA.

Our goal is to incorporate the kinematics of sequence dependence into the SGIB method to dynamically study the effects of a specified DNA sequence. We introduce the sequence dependent SGIB method as a way of incorporating the effects of a specific DNA sequence on the dynamics of DNA immersed in a fluid. We then apply this method to kDNA minicircles, and find the centroid distance distribution for two Hopf-linked minicircles for two Kinetoplastids: *Trypanosoma brucei* and *Crithidia fasciculata*.

## 4.2 Sequence Dependent Stochastic Generalized Immersed Boundary Method

The Sequence Dependent SGIB method is an adaptation of the SGIB method [73] in which DNA is modeled as a sequence of base pairs immersed in a fluid. This method, along with other immersed boundary methods [70, 46, 41, 73], dynamically accounts for the fluid-structure interaction of an immersed elastic structure in fluid. In contrast to these continuum, Cosserat rod models, each base pair is represented discretely as in the sequence-dependent DNA elasticity model [16].

In the Sequence Dependent SGIB method, we represent DNA as a sequence of  $N + 1$  base pairs with  $N$  spaces between base pairs. Each base pair is represented as a rectangular slab

with position  $\mathbf{X}^n(t)$  through the axial curve of DNA and orientation given by the orthonormal triad  $\{\mathbf{D}_1^n, \mathbf{D}_2^n, \mathbf{D}_3^n\}$  centered at the center of mass of each slab, for each  $n$ . This is analogous to discretizing the axial curve  $\mathbf{X}(s, t)$  and material frame  $\{\mathbf{D}_1(s, t), \mathbf{D}_2(s, t), \mathbf{D}_3(s, t)\}$  from the GIB method and SGIB methods [46, 41, 73].

We introduce relative position vectors  $\mathbf{r}_n = \mathbf{X}^{n+1} - \mathbf{X}^n$ , which defines a polygonal curve along the central axis of the DNA. The vector  $\mathbf{D}_2^n$  lies in the plane of the slab, runs along the longer axis of the slab, and is directed towards the DNA backbone. The vector  $\mathbf{D}_1^n$  lies in the plane of the rectangular slab, and runs along the shorter axis of the slab. Similarly, the  $\mathbf{D}_3^n$  vector is perpendicular to both  $\mathbf{D}_1^n$  and  $\mathbf{D}_2^n$ , and is normal to the plane of the slab.

The conformation of two neighboring base-pairs, represented by two slabs, is parametrized using the Cambridge University Engineering Department Helix Computation Scheme as follows [36]: The relative orientation of the slabs is determined by three angles: tilt, roll and twist, denoted by  $\theta_1^n, \theta_2^n$  and  $\theta_3^n$  and the relative positions are given by shift, slide and rise and denoted by  $\rho_1^n, \rho_2^n$  and  $\rho_3^n$ . The elastic energy of two consecutive base pairs  $\{n, n+1\}$  is given by the quadratic function

$$\psi^n(\theta_1^n, \theta_2^n, \theta_3^n, \rho_1^n, \rho_2^n, \rho_3^n) = \sum_{i=1}^3 f_{ij}(\theta_i^n - \bar{\theta}_i^n)(\theta_j^n - \bar{\theta}_j^n) \quad (4.1)$$

$$+ \sum_{i=1}^3 g_{ij}(\theta_i^n - \bar{\theta}_i^n)(\rho_j^n - \bar{\rho}_j^n) \quad (4.2)$$

$$+ \sum_{i=1}^3 h_{ij}(\rho_i^n - \bar{\rho}_i^n)(\rho_j^n - \bar{\rho}_j^n) \quad (4.3)$$

where  $\bar{\theta}_i^n$  and  $\bar{\rho}_i$  ( $i = 1, 2, 3$ ) are relaxed DNA parameters, and the coefficients  $f_{ij}$  and  $g_{ij}$  are the stiffness constants [16].

For the homogeneous, isotropic elastic rod model used in chapters 2 and 3, DNA is an idealized B-DNA structure, meaning the relaxed parameter values and stiffness constants are independent of the DNA sequence. They are defined by taking  $\theta_1 = \theta_2 = \rho_1 = \rho_2 = 0, \theta_3 = 34.3^\circ, \rho_3 = 0.34nm$ , the bending is assumed to be isotropic with the persistence length of  $50nm$ , which corresponds to  $f_{11} = f_{22} = \frac{kT}{4.84^2 deg^2}$ , twisting modulus is assumed to be  $\frac{f_{33}}{f_{22}} = 1.4$ , and translational deformations are assumed to be negligible by making  $f_{44} = f_{55} = f_{66} = \frac{5000kT}{(nm)^2}$ .

For the sequence dependent SGIB method, we study the effects of the DNA sequence; thus here we utilize sequence dependent values of the parameters and coefficients that are

consistent with physical measurements of the deformations of DNA molecule as observed in X-ray crystal structures of DNA fragments [59]. Alternative values can be estimated from molecular dynamics simulations [44]. A complete table of these parameters is given in table 5.

The total energy of each minicircle is given by the sum of  $\psi^n$  for all consecutive base pairs, where  $\psi^n$  is the sum of the interaction of between the  $n$ th and  $(n + 1)$ st base pair,  $\psi = \sum_{n=1}^N \psi^n$ .

For each  $n$ , the balance equations, describing the balance of the DNA's internal and external forces and moments is given by

$$\mathbf{f}^n - \mathbf{f}^{n-1} = \boldsymbol{\phi}^n, \quad \mathbf{m}^n - \mathbf{m}^{n-1} = \mathbf{f}^n \times \mathbf{r}^n + \boldsymbol{\mu}^n \quad (2 \leq n \leq N) \quad (4.4)$$

For each  $n$ , the force and moment that the  $(n + 1)$ st exerts on  $n$ th base pair is given by  $\mathbf{f}^n$  and  $\mathbf{m}^n$  respectively, and the external force and moment of  $n$ th base pair acting on the fluid is given by  $\boldsymbol{\phi}^n$  and  $\boldsymbol{\mu}^n$  [16].

As in the SGIB method [6, 73], the motion of the fluid is modeled using a time-dependent, incompressible, Stokes flow with velocity,  $\mathbf{u}(\mathbf{x}, t)$ , pressure  $p(\mathbf{x}, t)$ , density  $\rho$ , and viscosity  $\mu$ :

$$\rho \mathbf{u}_t = -\nabla p + \mu \Delta \mathbf{u} + \mathbf{f}^{tot}, \quad \nabla \cdot \mathbf{u} = 0. \quad (4.5)$$

The fluid is assumed to have a no slip condition, meaning the DNA moves with the velocity of the fluid. The force density  $\mathbf{f}^{tot}(\mathbf{x}, t)$  acting on the fluid accounts for two forces: the force resulting from the DNA acting on the fluid, and the thermal forcing. The SGIB method assumes a uniform random thermal forcing throughout the fluid, as in [6].

The full system of equations describing the rod and fluid dynamics is as follows:

$$\rho \frac{\partial \mathbf{u}}{\partial t} = -\nabla p + \mu \Delta \mathbf{u} + \mathbf{f}^{tot}, \quad (4.6)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (4.7)$$

$$\mathbf{f}^n - \mathbf{f}^{n-1} = \boldsymbol{\phi}^n, \quad (2 \leq n \leq N) \quad (4.8)$$

$$\mathbf{m}^n - \mathbf{m}^{n-1} = \mathbf{f}^n \times \mathbf{r}^n + \boldsymbol{\mu}^n \quad (2 \leq n \leq N) \quad (4.9)$$

$$\mathbf{f}^n = f_1^n \mathbf{D}_1^n + f_2^n \mathbf{D}_2^n + f_3^n \mathbf{D}_3^n, \quad (4.10)$$

$$\mathbf{m}^n = m_1^n \mathbf{D}_1^n + m_2^n \mathbf{D}_2^n + m_3^n \mathbf{D}_3^n, \quad (4.11)$$

$$f_i^n = \frac{\partial \psi^n}{\partial \rho_j^n} \left( Z_{jk}(-\gamma^n) Y_{kl} \left( -\frac{1}{2} \kappa^n \right) Z_{li}(-\zeta^n) \right), \quad m_i^n = \Gamma_{ij}^n \left( \frac{\partial \psi^n}{\partial \theta_j^n} + \frac{\partial \psi^n}{\rho_k^n} {}_j \Lambda_{kl}^n \rho_l^n \right), \quad (4.12)$$

$$\begin{aligned} \mathbf{f}(\mathbf{x}, t) &= \sum_{n=1}^N (-\phi^n(t)) \delta_c(\mathbf{x} - \mathbf{X}^n(t)) + \frac{1}{2} \nabla \times \sum_{n=1}^N (-\boldsymbol{\mu}^n(t)) \delta(\mathbf{x} - \mathbf{X}^n(t)) \\ &+ \sum_{n=1}^N (-\mathbf{f}^{c,n}(t)) \delta_c(\mathbf{x} - \mathbf{X}^n(t)) + \sum_{i=1}^n (-\mathbf{f}^{e,n}(t)) \delta_c(\mathbf{X} - \mathbf{X}_n(t)), \end{aligned} \quad (4.13)$$

$$\frac{\partial \mathbf{X}(s, t)}{\partial t} = \mathbf{U}(s, t) = \int \mathbf{u}(s, t) \delta_c(\mathbf{x} - \mathbf{X}(s, t)) d\mathbf{x}, \quad (4.14)$$

$$\mathbf{W}(s, t) = \frac{1}{2} \int (\nabla \times \mathbf{u}) \delta_c(\mathbf{x} - \mathbf{X}(s, t)) d\mathbf{x}, \quad (4.15)$$

$$\frac{\partial \mathbf{D}_i(s, t)}{\partial t} = \mathbf{W}(s, t) \times \mathbf{D}_i(s, t), \quad i = 1, 2, 3. \quad (4.16)$$

Here  $D_{ij}^n = \mathbf{D}_1^n \cdot \mathbf{D}_j^{n+1} = Z_{ij}(\zeta^n) Y_{kl}(\kappa^n) Z_{lj}(\eta^n)$ . We define  $\rho_i^n$  in terms of  $r_i^n$  and  $\theta_i^n$  by  $\rho_i^n = Z_{ij}(-\gamma^n) Y_{jk}(-\frac{1}{2}\kappa^n) Z_{kl}(-\zeta^n) r_l^n$ , and the orthonormal mid-basis  $\bar{\mathbf{D}}_i^n$  by  $\mathbf{D}_i^n \cdot \bar{\mathbf{D}}_j^n = Z_{jk}(\zeta^n) Y_{kl}(\frac{1}{2}\kappa^n) Z_{lj}(\gamma^n)$ . The angles  $\zeta^n$ ,  $\kappa^n$ ,  $\eta^n$  follow the Euler-angle system, and the matrices  $Y_{ij}(\alpha)$ ,  $Z_{ij}(\alpha)$ , the quadratic form  $\Gamma_{ij}^n$ , and the skew-symmetric matrices  ${}_j \Lambda_{kl}^n$  are defined in Appendix B.

We define  $\delta_c(\mathbf{x})$ , the smooth approximation of the 3-dimensional delta function, in equation 2.16 in chapter 2 [46, 41]. The steric and electrostatic forces,  $\mathbf{f}^{c,n}(t)$  and  $\mathbf{f}^{e,n}$  are defined in equations 2.21 and 2.22 respectively in chapter 1 [41]. Here the total self-contact force,  $\mathbf{f}^c(t) = \sum_{n=1}^N \mathbf{f}^{c,n}(t)$ , is the sum of every local contact force at each point of contact at position  $\mathbf{X}^n(t)$ .

The force density  $\mathbf{f}^{tot}(\mathbf{x}, t)$  acting on the fluid is a sum of the the external body force from the rod onto the fluid, and the thermal force:  $\mathbf{f} = \mathbf{f}(\mathbf{x}, t) + \mathbf{f}^{thm}$ . The thermal force is assumed to be uniform across the domain of the fluid, and is represented in Fourier space as

Gaussian white noise. This is defined in equation 3.3 in chapter 2 [6, 73], and the procedure of solving for the fluid velocity and position of DNA in one time step follows the method from [6]. Using discretized balance equations of the Stokes flow decouples the system into a system of ordinary differential equations in Fourier space. Representing the thermal force as Gaussian white noise in Fourier space with Stokes flow means that the resultant stochastic differential equation is solvable in one time step using Itô integration, and the position of the DNA is updated accordingly.

The full numerical algorithm of the Sequence Dependent SGIB method is as follows: 1. Evaluate the balance equations of the DNA, and compute the resultant forces and moments at each discretization point along the DNA using the sequence dependent kinematic variables,  $\theta_1^n, \theta_2^n, \theta_3^n, \rho_1^n, \rho_2^n, \rho_3^n$ ; 2. Convert forces and moments external to the DNA to forces and moments acting on the fluid using the discrete delta function; 3. Update the fluid velocity using Stokes flow in Fourier space; 4. Update the velocity of the DNA 5. Propagate the DNA through the fluid.

#### 4.2.1 Dynamical sampling of minicircle centroid distances

To determine the effects of sequence dependence on the dynamics of kNDA, we wanted to find the distribution of centroid distances of two Hopf-linked minicircles for two particular kinetoplast DNA sequences: *T. brucei* and *C. fasciculata*. To do so, we used the SD-SGIB method to dynamically simulate the interaction of two minicircles. Simulations were run using MATLAB on a cluster.

For each minicircle, the segment of DNA was given periodic boundary conditions, meaning  $X^1(0) = X^{N+1}(0)$ , and  $D_i^1(0) = D_i^{N+1}(0)$  for  $i = 1, 2, 3$ . The kinematic parameters  $\theta_1^n, \theta_2^n, \theta_3^n, \rho_1^n, \rho_2^n, \rho_3^n$  were defined for the DNA sequences of *T. brucei* and *C. fasciculata* according to [59, 44] and are listed in table 5. The minicircle sequences for *T. brucei* and *C. fasciculata* with  $N = 1014$  and  $N = 2515$  base pairs respectively are listed in appendix C and were taken from [57]. The fluid is also assumed to have periodic boundary conditions with  $\hat{N} = 64^3$ .

The starting configuration for each minicircle was generated using the Gaussian sampling

method from [23] with subsegment pairing and closure detection. The procedure takes advantage of the quadratic nature of the elastic energy of the DNA, which implies that the probability of any particular deformation (represented by parameters  $\theta_i, \rho_i$ ) is given by a multivariate normal density that can easily be sampled to obtain a collection of deformed configurations of the segment. To enforce a closure of the segment, we sampled independently two halves of the sequence and then identified those halves for which the ends came into proximity. We chose the most likely starting configuration with the linking number of 95 for *T. brucei* and linking number of 239 for *C. fasciculata*. Each minicircle was discretized by every base pair, and our time step of  $\Delta t = 2 \cdot 10^{-10}$  for *T. brucei* and  $\Delta t = 1 \cdot 10^{-9}$  for *C. fasciculata*.

For modeling kDNA minicircles, the Debye-Hückel electrostatic screening effects of the SGIB method were neglected. Instead, we chose an effective diameter of DNA of 65 Angstroms, corresponding to solvent ionic strength of 0.1M. This was given to be the effective diameter of a segment of DNA according to [67], as a function of ionic strength of the solvent. We treat the electrostatic interaction as a hardcore potential, rather than using the electrostatic Debye-screening method [52, 81] as in [73, 41]. A full set of parameters is listed in table 4. Note that here, we list the constants for electrostatics for the general SD-SGIB method. We had to adjust the stiffness parameter  $g$ , from the GIB methods and SGIB methods for stability of the method.

The probability density was obtained by simulating 20 configurations of *T. brucei* and *C. fasciculata* Hopf-linked minicircles, and running simulations until we observed sufficient mixing of the centroid distance trajectory paths. We consider two minicircles topologically linked if the linking number as defined in equation 1.1 in chapter 1 is nonzero. The linking number for these minicircles was computed from the algorithm as in [42, 45]. Here, we start with two DNA minicircle segments (generated by the Gaussian sampling method) joined by a Hopf-link. The 20 starting configurations were initialized with centroid distances spaced uniformly from a distance of 0 Angstroms to the diameter of the starting configurations of 1131 Å for *T. brucei* and 3232 Å for *C. fasciculata*. Each pair of minicircles was also given a random rotation to start, under the condition that the minicircles had to remain linked. We then obtained the probability density of  $R$ , the distance between two centroids, finding the

total time spent at distance  $R$ , and dividing by the surface area of a sphere with radius  $R$ , so that we obtain a rotationally invariant probability density.

### 4.3 Kinetoplast DNA sequence dependent effects

From the 20 simulations for both *T. brucei* and *C. fasciculata*, we computed the distance between the centroids of the two minicircles. The traces of these distances along with a histogram of the centroid distances is displayed in figures 22 and 23 respectively. These were generated by adding an initial burn-in period to exclude the dependence on the starting distribution.

For *T. brucei*, we see that there is sufficient mixing for the time traces, so we would expect the distribution to estimate the steady state distribution for the stochastic process. From the centroid distribution, we see a peak in the radial distance around  $0.035\mu\text{m}$ , and a sharp decay in the tail of the distribution.

The result for *T. brucei* is consistent with results from [48]. Using a Gaussian sampling method [23] to generate *T. brucei* and *C. fasciculata* minicircle configurations, Liu et al. found the linking probabilities for two minicircles for sequences with effective diameter of  $D = 65\text{\AA}$ . They determined the linking probability for a fixed centroid distance, but this was not obtained dynamically. Because kDNA minicircles naturally fluctuate in fluid, we sought to obtain the distribution dynamically and were interested in the distribution of centroid distances given that two minicircles must stay linked.

It is worth noting that the probability of a small centroid distance decays near zero as well. Diao et al. find the linking probability of two Hopf-linked minicircles modeled as rigid circles [26]. With rigid circles and no volume exclusion effects, the linking probability is linear, and increases to a probability of one as the centroid distance decreases to zero. Our results are consistent with [25, 26] that show the linking probability for two minicircles near zero decreases for freely-jointed chains and rigid circles with volume exclusion.

If we compare this with the traces and centroid distance distribution, we see that the distribution for *C. fasciculata* is very different, and there is little mixing of the traces. This

parameters	symbol	value
grid size	$\hat{N} \times \hat{N} \times \hat{N}$	$64^3$
domain size	$L \times L \times L$	T.b: $L = 274.351 \text{ \AA}$ , C.f: $L = 680.467 \text{ \AA}$
time step	$\Delta t$	T.b: $2 \times 10^{-10}$ , C.f: $1 \times 10^{-9}$
fluid density	$\rho$	$1 \text{ g cm}^3$
fluid viscosity	$\mu$	$0.01 \text{ g (cm s)}^{-1}$
permittivity of free space	$\epsilon_0$	$3.45 \times 10^{-7} e^2 / (g \text{ \AA}^3 s^{-2})$
dielectric constant of water	$\epsilon_W$	77.7
electric charge	$q$	$0.24 \times 2e$
molar salt concentration	$C_s$	0.01, 0.05, 0.1M
Debye screening parameter	$\kappa$	$0.329 \sqrt{C_s} \text{ \AA}^{-1}$
number of base pairs	$N$	T.b:1014, C.f:2515
radius	$r$	T.b: $54.870 \text{ \AA}$ , C.f: $136.093 \text{ \AA}$
diameter	$D$	$65 \text{ \AA}$

Table 4: kDNA computational parameters

parameters	CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
tilt ( $\theta_1$ )	$0.00^\circ$	$-0.02^\circ$	$0.00^\circ$	$-1.31^\circ$	$-0.04^\circ$	$-1.30^\circ$	$-1.51^\circ$	$0.00^\circ$	$0.46^\circ$	$0.00^\circ$
roll ( $\theta_2$ )	$4.32^\circ$	$4.98^\circ$	$2.93^\circ$	$3.79^\circ$	$5.04^\circ$	$0.43^\circ$	$1.83^\circ$	$1.01^\circ$	$1.79^\circ$	$0.70^\circ$
twist ( $\theta_3$ )	$34.73^\circ$	$34.98^\circ$	$37.08^\circ$	$32.65^\circ$	$33.06^\circ$	$35.18^\circ$	$35.49^\circ$	$29.86^\circ$	$31.36^\circ$	$33.58^\circ$
shift ( $\rho_1$ )	$0.00 \text{ \AA}$	$-0.05 \text{ \AA}$	$0.00 \text{ \AA}$	$0.09 \text{ \AA}$	$-0.04 \text{ \AA}$	$0.02 \text{ \AA}$	$-0.29 \text{ \AA}$	$0.00 \text{ \AA}$	$0.26 \text{ \AA}$	$0.00 \text{ \AA}$
slide ( $\rho_2$ )	$0.3 \text{ \AA}$	$0.3 \text{ \AA}$	$0.1 \text{ \AA}$	$-0.3 \text{ \AA}$	$-0.4 \text{ \AA}$	$-0.2 \text{ \AA}$	$-0.1 \text{ \AA}$	$-0.7 \text{ \AA}$	$-0.6 \text{ \AA}$	$0.2 \text{ \AA}$
rise ( $\rho_3$ )	$3.4 \text{ \AA}$	$3.4 \text{ \AA}$	$3.3 \text{ \AA}$	$3.3 \text{ \AA}$	$3.4 \text{ \AA}$	$3.3 \text{ \AA}$	$3.3 \text{ \AA}$	$3.2 \text{ \AA}$	$3.3 \text{ \AA}$	$3.4 \text{ \AA}$

Table 5: Sequence dependent elasticity parameters

is because for many of these simulations, we observed sticking for the SD-SGIB method. Thus, we sought to quantify the sticking behavior, and to find solutions to prevent this from happening in the numerical method.

### 4.3.1 Sequence Dependent SGIB Method

The SD-SGIB method is a novel method for modeling dynamics of sequence dependent DNA immersed in a fluid. Using our new method, we were able to model sequence dependent effects of kDNA minicircles. In the case of *T. brucei*, Sequence Dependent SGIB method worked effectively; however this method has computational limitations. We will describe the advantages of the new method, and discuss its limitations, including defining the main issue of sticking.

There are two main limitations of the SD-SGIB method. The first is that the numerical method scales with the length of the segment of the DNA. This is a general property of IB methods, [41, 73], and is one reason we chose to model two Hopf-linked minicircles, as opposed to a large network of minicircles.

The second is the issue of sticking. We say that a simulation is sticking, if two strands of DNA move together as one for an extended period of time. This occurs as an artifact of the immersed boundary method. The forces and moments from the DNA acting on the fluid are spread to neighboring fluid mesh points through the smooth approximation of the 3-dimensional delta function  $\delta_c$ . If points along the DNA occupy the same mesh space, without a large force from electrostatic repulsion, these will move together throughout the fluid, as opposed to having natural fluctuations due to the thermal forcing in the fluid.

We are able to identify sticking by finding the pairwise distance between DNA base pairs. If the distance for a significant portion of the DNA remains below a minimum threshold, we identify this as sticking. We considered finding the covariance matrix of the position of base pairs as a function of time. As the DNA strands move together, their positions should be positively correlated, but because of no-slip conditions, all of the DNA positions are positively correlated throughout a dynamic simulation, and it is too difficult to determine the effects from sticking alone.

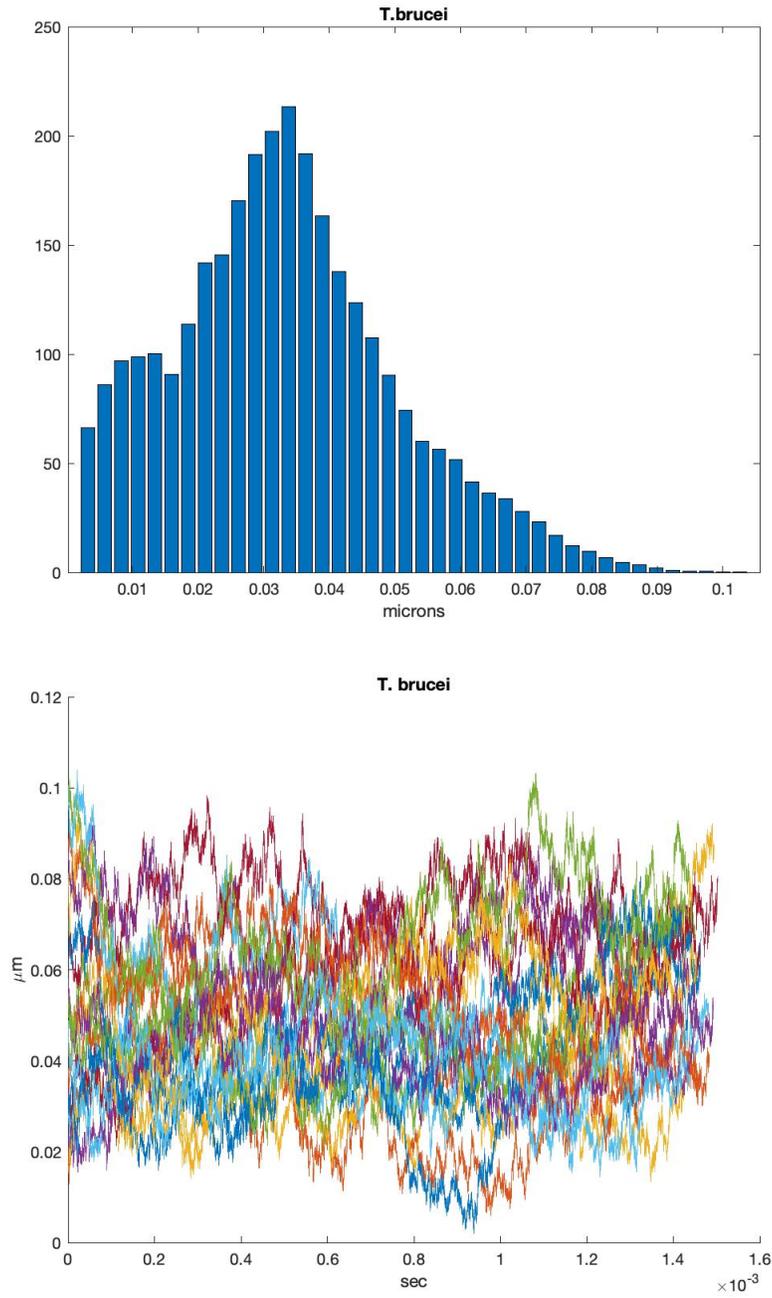


Figure 22: *T. brucei* centroid distance (microns) distribution and centroid distance time traces for 20 Hopf-linked minicircles.

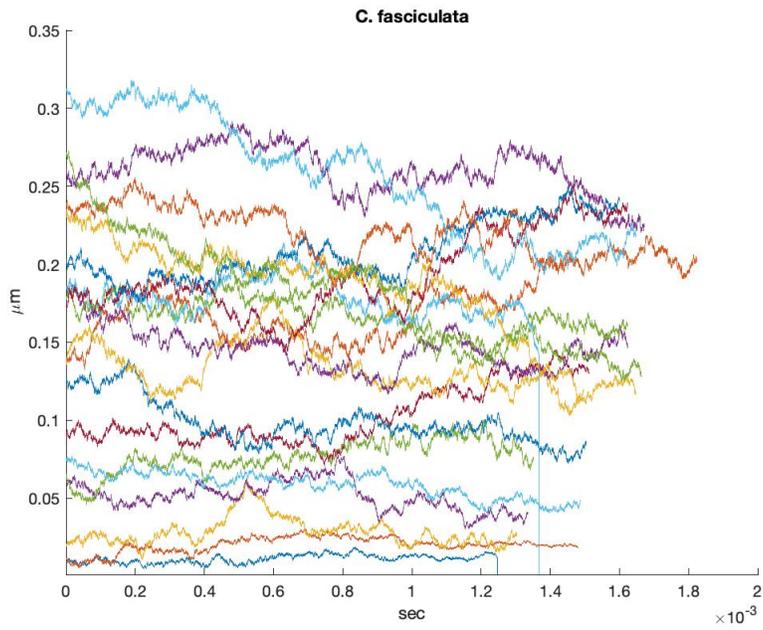
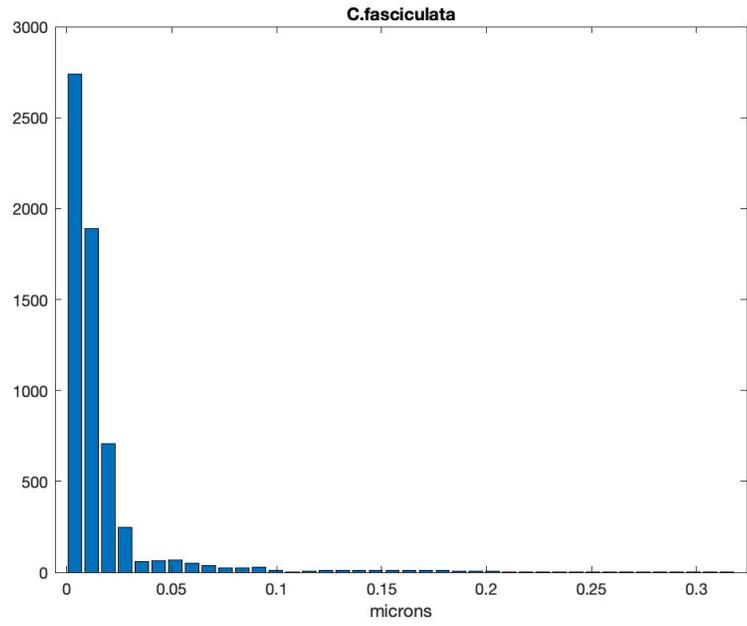


Figure 23: *C. fasciculata* centroid distance (microns) distribution and centroid distance time traces for 20 Hopf-linked minicircles.

We generated distance plots throughout the sequence dependent simulations for both *T.brucei* and *C.fasciculata*. We used a threshold of  $1.3 \cdot D$  and  $1.2 \cdot D$  to determine if two strands of DNA were sticking throughout a simulation. As an example of sticking, figure 24 displays a dynamic progression of two *C. fasciculata* minicircles, and the corresponding distance plots. The distance plot shows a color map of the pairwise distance between each base pair. Between any two base pairs, a distance less than a tolerance of  $1.2 \cdot D$  is displayed in yellow, and distances above this threshold are displayed in blue. The first minicircle has labels 1 through 2515, and the second is labeled as 2516 through 5030. The yellow band along the diagonal represents the distance from one base pair to itself, so we look for persistent yellow bands off of this diagonal.

With an effective diameter of  $D = 65\text{\AA}$ , we see that the simulation of *C. fasciculata* shows signs of the SD-SGIB method sticking. In the left panel, at two different locations along both minicircles, there is a growing segment of DNA that begins to move as one strand, until roughly one third of the minicircle strands move as a single strand of DNA. In the panel on the right, we see that this corresponds to the appearance of yellow bands in the distance plot that persists and grow in size throughout the dynamic progression. In contrast, the simulations for *T. brucei* with effective diameter of  $D = 65\text{\AA}$  never exhibited sticking. Sometimes, it is possible to see a yellow band appear and even fluctuate momentarily, but for *T. brucei*, these bands disappear eventually. The close proximity of the DNA strands is due to natural drift and diffusion of the stochastic process.

Another method of identifying sticking, is to compute the diffusion coefficient for the centroid distances as a function of time. For a one-dimensional random walk  $X(t)$ , the diffusion coefficient is approximated by  $D_{Diff} = \frac{(\Delta x)^2}{2\Delta t}$ . Thus we calculated the diffusion coefficient for the traces for centroid distances. A simulation exhibits sticking if the diffusion coefficient is small. This is because there are fewer degrees of freedom for the center of mass of two minicircles to fluctuate when one or several portions of the DNA segments are moving as a single strand. Note that a large diffusion coefficient does not necessarily mean that there is no sticking present. Take for example two minicircles that start far apart and stick near the location of the Hopf link. These DNA segments are allowed to fluctuate like flapping wings, despite sticking.

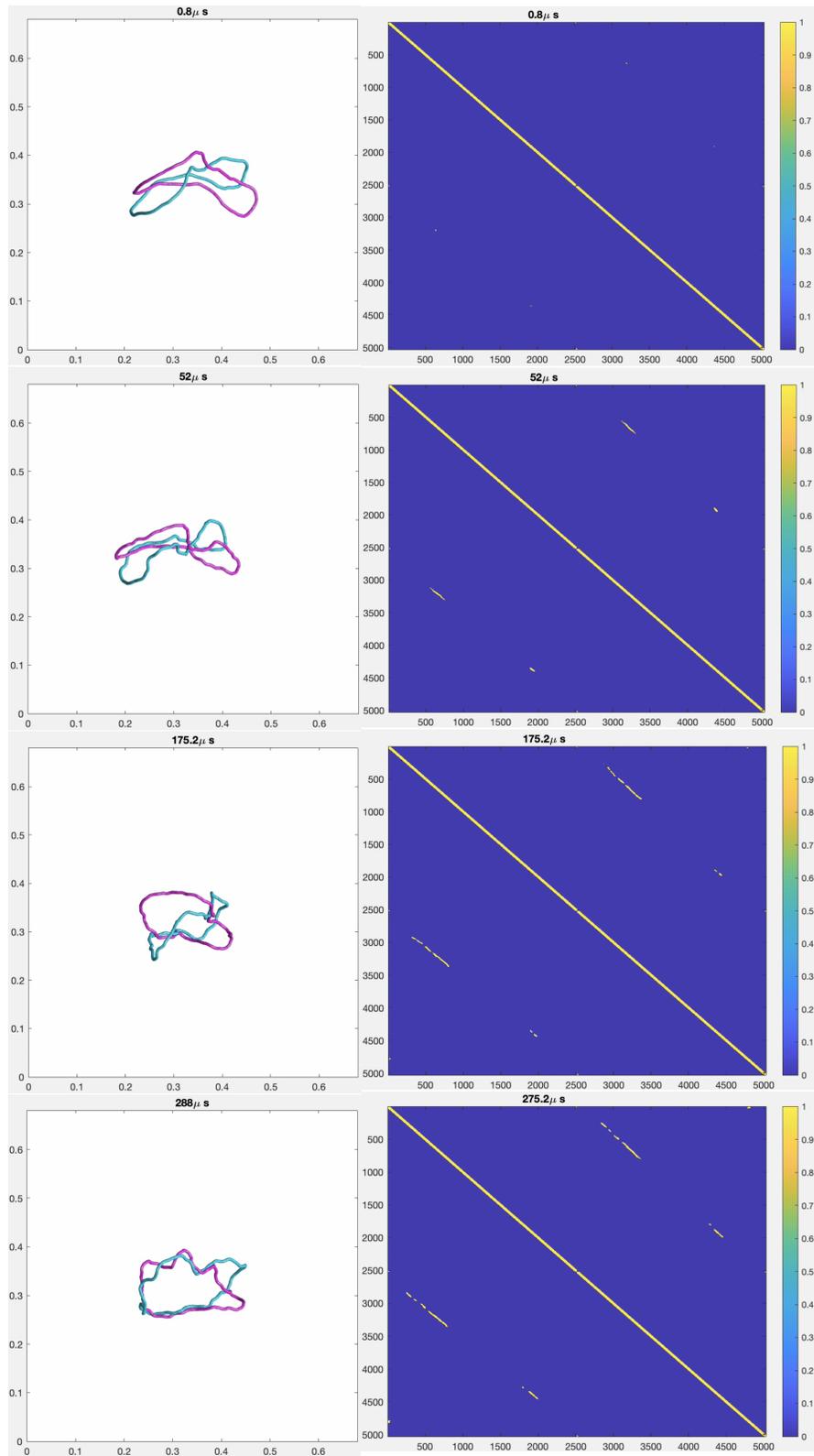


Figure 24: *C. fasciculata* dynamic progression illustrating sticking with corresponding distance plot.

We estimated the diffusion coefficient for two *C. fasciculata* minicircle centroid distances. In figure 25, we see that for *C. fasciculata* the diffusion coefficient for the sticking simulation is much lower than the simulation with no sticking. This gives a procedure for checking for sticking in real time as well, although a large diffusion coefficient does not necessarily mean that there is not sticking for the SD-SGIB method.

We tried several ways to prevent the sticking phenomenon for the sequence dependent minicircles. Depending on the chosen parameter values for the SD-SGIB method, we observed sticking for both kDNA sequences. The first method was to increase the 20Å diameter of DNA to an effective diameter of 65Å. This was effective for preventing sticking for *T. brucei*, but not for *C. fasciculata*. For  $\hat{N} = 64$ , the fluid domain length to grid sizes gives a mesh width of 4.287Å and 10.632Å for *T. brucei* and *C. fasciculata* respectively. The second method we tried for *C. fasciculata* was to change the fluid mesh. When we increased the mesh to  $128^3$ , computations were no longer feasible. For  $\hat{N} = 64$  and  $dt = 1 \cdot 10^{-9}$ , it took 51927.66 seconds to run 12000 iterations of a *C. fasciculata minicircle* simulation. Comparatively, for  $\hat{N} = 128$ , and  $dt = 3 \cdot 10^{-10}$ , this would require a runtime of roughly 26257546.6 seconds to run obtain a simulation of the same length of time. That is roughly 500 times longer because of the refined fluid mesh and significantly smaller value of  $dt$ . When we discretize by every base pair, even with only a mesh of  $64^3$ , the dynamical simulations already take three weeks running continuously on a cluster, thus for  $\hat{N} = 128$  the current method is not computationally feasible. We thought to reduce the number of base pairs, but that would neglect the sequence dependent effects.

The only method we found to be effective for *C. fasciculata*, was to include electrostatic repulsion. We observed sticking for  $Cs = 0.1, 0.05M$  and neither of these solved the sticking phenomenon. The problem is that with  $Cs = 0.01M$ , there was no sticking throughout the simulation. With such high electrostatic repulsion, corresponding to an effective diameter of 150Å [67], this would no longer be biologically accurate.

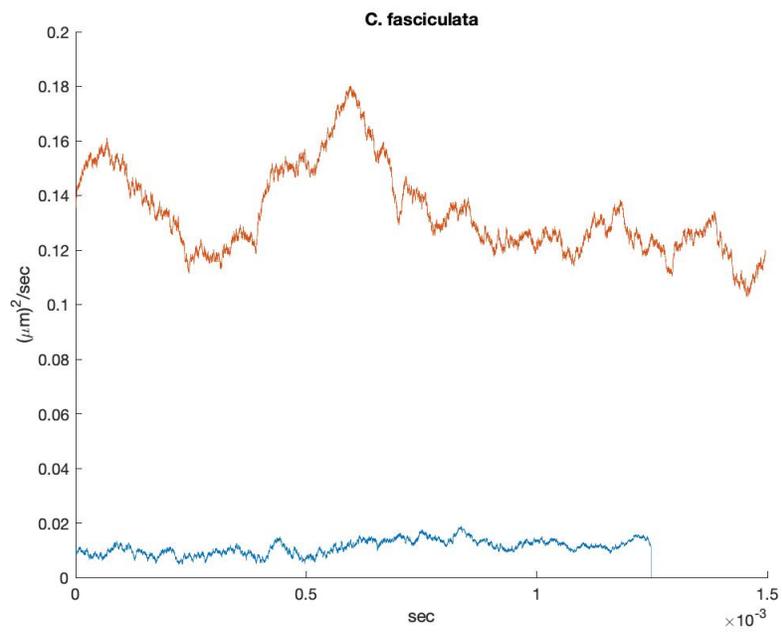


Figure 25: Diffusion coefficient for *C. fasciculata* minicircles. Initial centroid distance of 0 microns (blue) and 0.1616 microns (orange)

## 4.4 Conclusion

We created a novel method for modeling the effect of sequence dependent in DNA dynamics. This SD-SGIB method incorporates base pair specific elastic properties of DNA while still dynamically modeling the DNA interaction with fluid.

Using the SD-SGIB method, we modeled the dynamics of two kDNA Hopf-linked minicircles. For *T. brucei*, we found a distribution for the center of masses of two linked minicircles dynamically. In the case of modeling *C. fasciculata*, because of the length of the DNA sequence, the SD-SGIB method encounters the phenomenon of sticking. We examined possible solutions to this issue, and found including electrostatic repulsion to be effective.

The average minicircle centroid distance could indicate when the network is in a stress-free state. This is because we would expect that in a stress free configuration, the pairwise minicircle centroid distance to be within a small error of the mean. So at large scales this gives a measure of the stress-free or strained network configurations based on the diameter of the network.

By incorporating sequence dependence with the fluid-structure interaction method, the SD-SGIB method more accurately models the dynamics of DNA. This was effective in modeling the dynamic of DNA links for small networks of DNA sequences with length constraints.

## 5.0 Conclusion

Mathematical modeling has played a key role in understanding the structure and function of DNA since its very discovery. This has sparked advances across several fields of biology and medicine, and it inspired new mathematical frameworks aimed at explaining the function of DNA. We studied the topology of DNA by modeling dynamics of DNA knots and links.

Using a continuum rod model, we found knotted equilibrium structures for various excess linking numbers and a distribution for the various knot types. Because DNA knots and knot topology change throughout the process of cellular replication, this may give insights into the enzymes and mechanisms responsible for maintaining the topology of DNA. Using a discrete base pair model, we dynamically modeled a pair of kDNA minicircles. This could give an indication of the topology of a stress free kinetoplast DNA network and in turn shed light on the cellular processes involved in maintaining this network structure.

To model the topology of DNA dynamically, we considered three numerical methods. The first was the Generalized Immersed Boundary Method, which considers DNA to be a thin elastic rod immersed in fluid. Using this method we found the symmetries of knot equilibrium structures for various excess linking numbers, and deterministically described the dynamics of transitions between these states. The second method we used was the Stochastic Generalized Immersed Boundary Method. Because this method accounts for thermal fluctuation of the surrounding fluid, the SGIB method gives a more biologically realistic model of DNA dynamics. Using the SGIB method, we estimated the probability of DNA knot configurations with excess twist. The SGIB method assumes DNA to be homogeneous and transversely isotropic, and neglects the elastic effects of specific base pairs. Thus, we created a novel method, the Sequence Dependent Stochastic Generalized Immersed Boundary Method to account for sequence dependence. Using the SD-SGIB method, we studied the effects of sequence dependence on kinetoplast DNA by dynamically simulating two Hopf linked minicircles. We also found the method to be effective at modeling sequence dependent DNA for small networks of DNA minicircles with length constraints.

By modeling DNA knots and links, we dynamically explored the energy landscape of

elastic knots with excess twist. Using the deterministic model, we found stable equilibrium knot configurations and saddle configurations. We also determined how the elastic energy of these knot configurations changed throughout the equilibration process. This deterministic method provides an idea of the overall dynamics but gives an incomplete picture based on elastic energy values alone.

Thus, we consider a stochastic model over the space of knot configurations. Thinking of this process as a continuous time Markov chain gives more realistic dynamics by accounting for the entropy of the system. The deterministic system provided a state space of knots given a fixed linking number for the CTMC. Using the Kendall shape space and a minimum Procrustes distance equivalence relation, we found energy barriers between these states. We then obtained transition rates between knot states and the mean holding times for each of the equilibrium configurations. Finally, we used the mean holding times and transition rates to find a steady state distribution of the stochastic process. By comparing this to a distribution estimated from the elastic energies alone, we found that the stochastic system gives a more accurate representation of the knot energy landscape.

DNA topology is important in understanding the function of DNA and the cellular mechanisms involved in maintaining its structure. Modeling DNA knots and links, provides a rich mathematical framework to dynamically explore this topology, and through this modeling process, we gain insights into the behavior of DNA as a whole.

## Appendix A Mean holding times and transition rates

The following are the mean holding times and transition rates for  $Lk = -4, -5, -6, -7, -8, -9, 10$ .

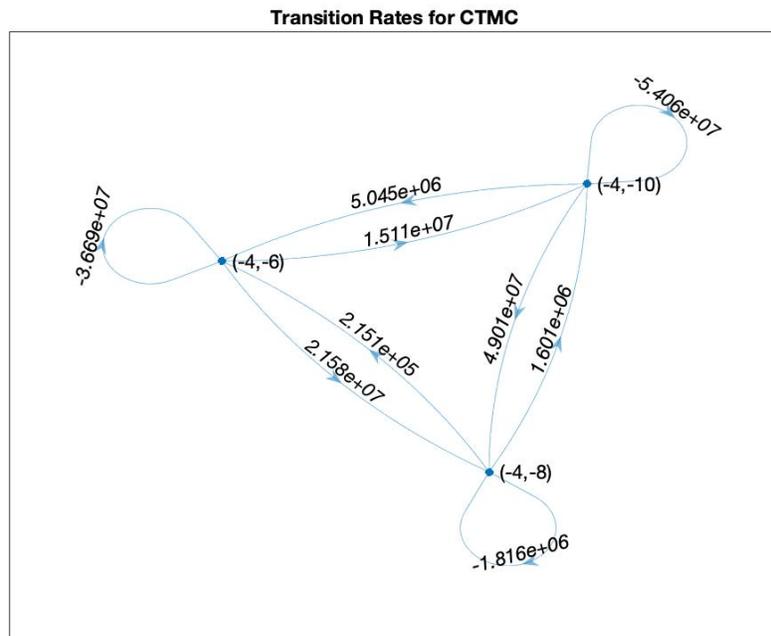
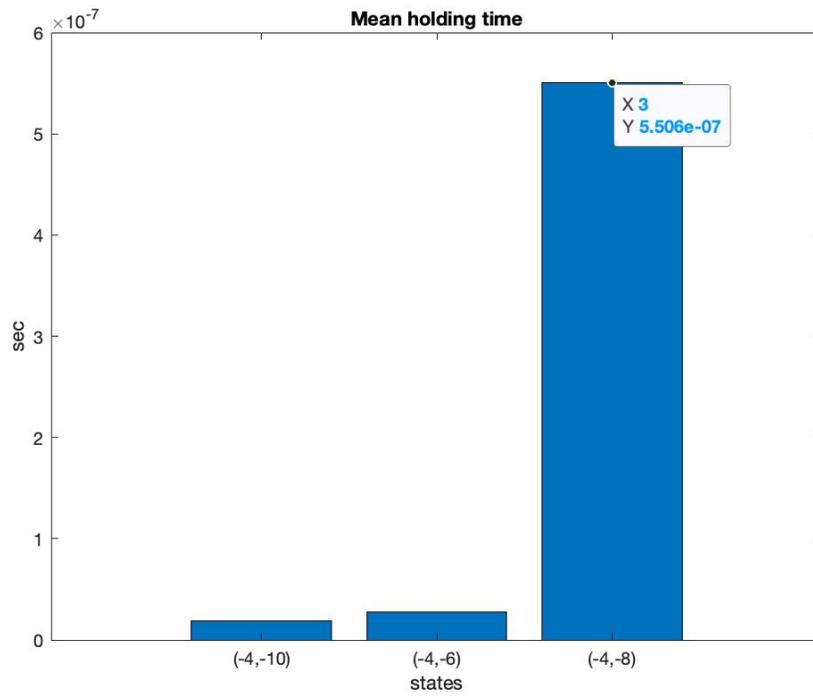


Figure 26: Mean holding times and transition rates for  $Lk_{rel} = -4$ .

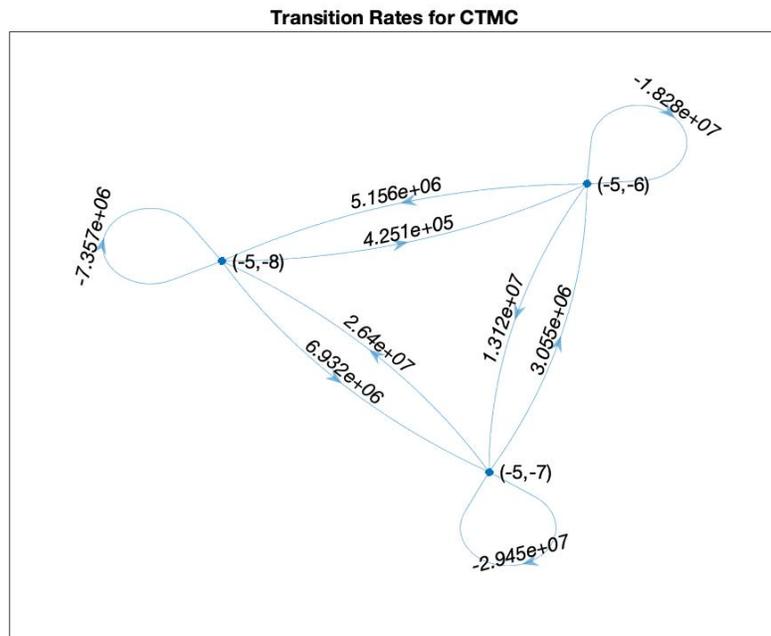
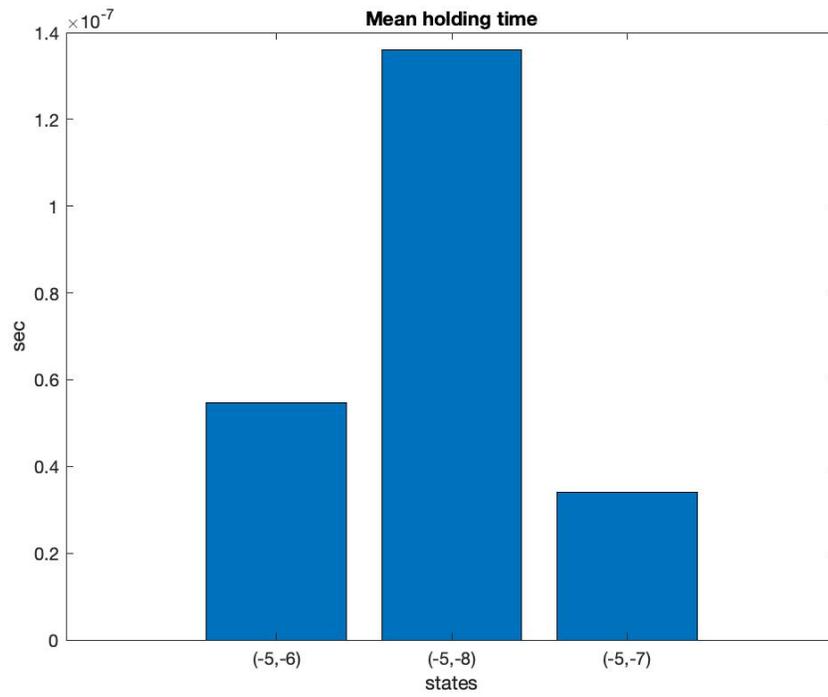


Figure 27: Mean holding times and transition rates for  $Lk_{rel} = -5$ .

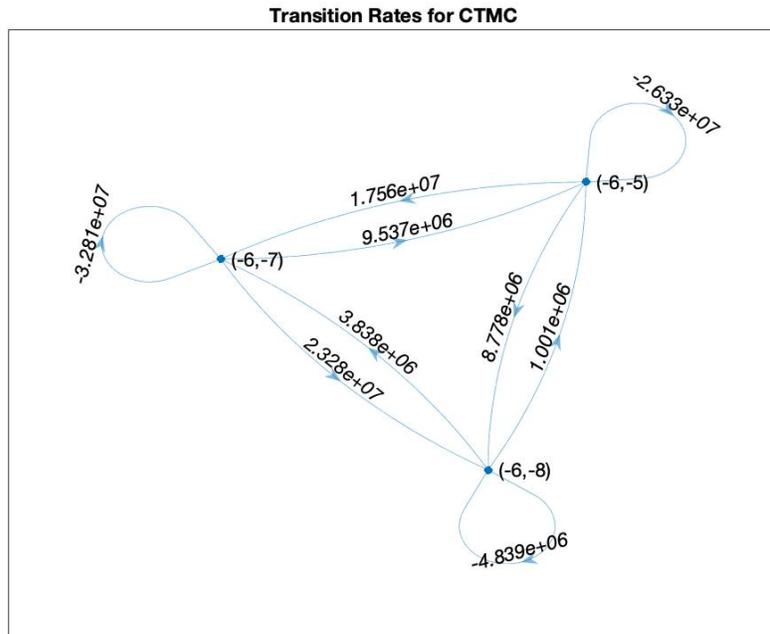
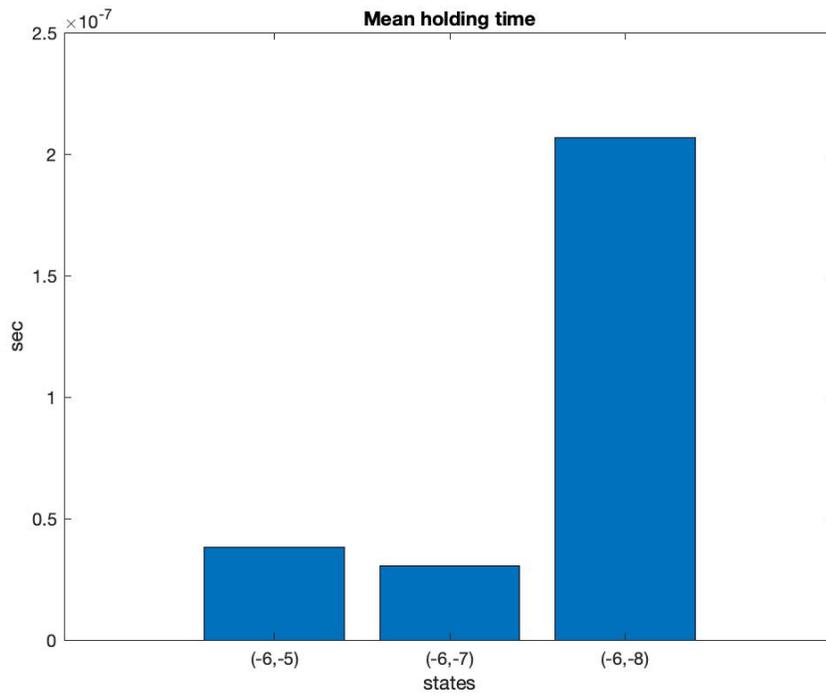


Figure 28: Mean holding times and transition rates for  $Lk_{rel} = -6$ .



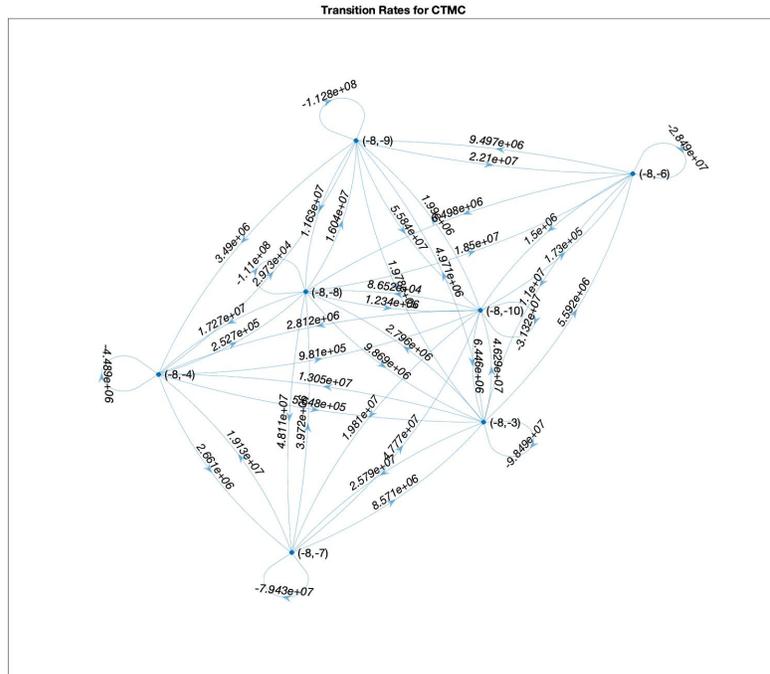
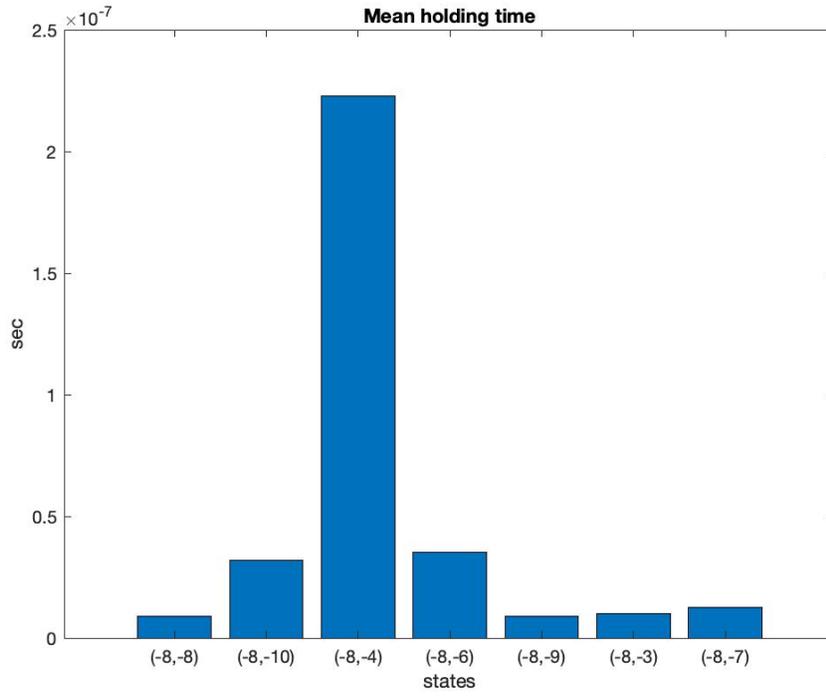


Figure 30: Mean holding times and transition rates for  $Lk_{rel} = -8$ .

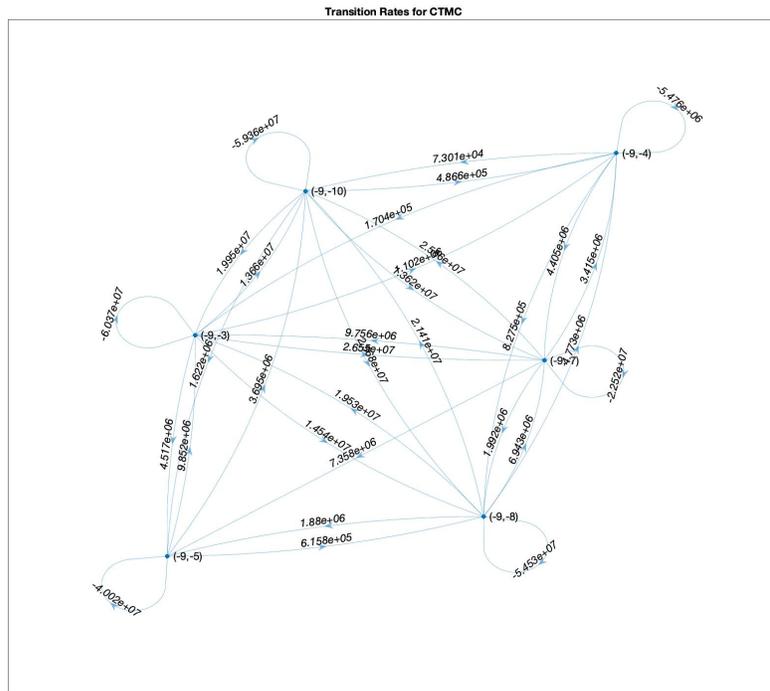
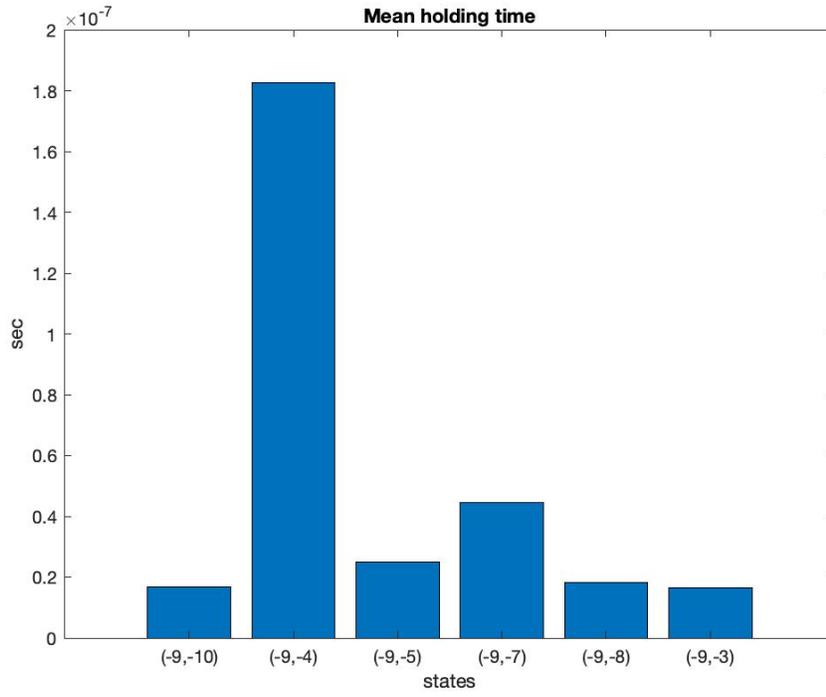


Figure 31: Mean holding times and transition rates for  $Lk_{rel} = -9$ .

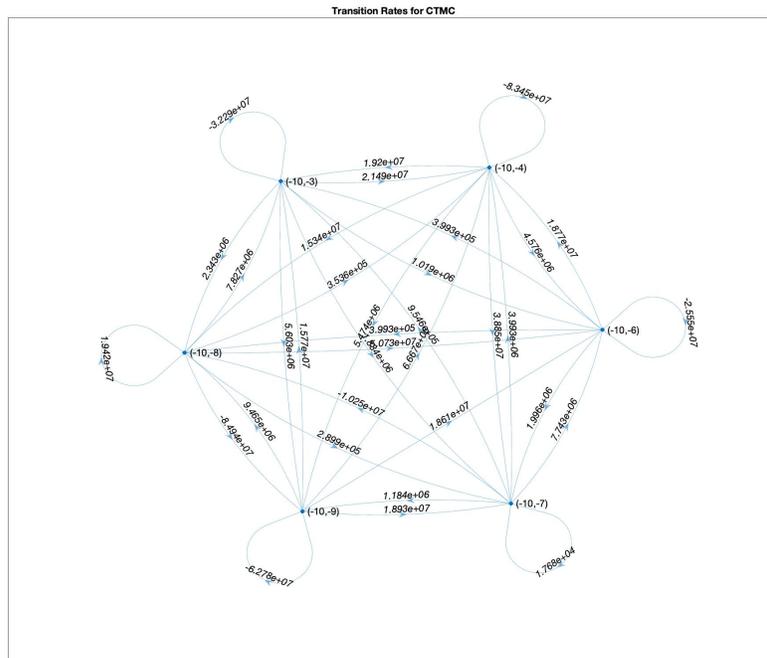
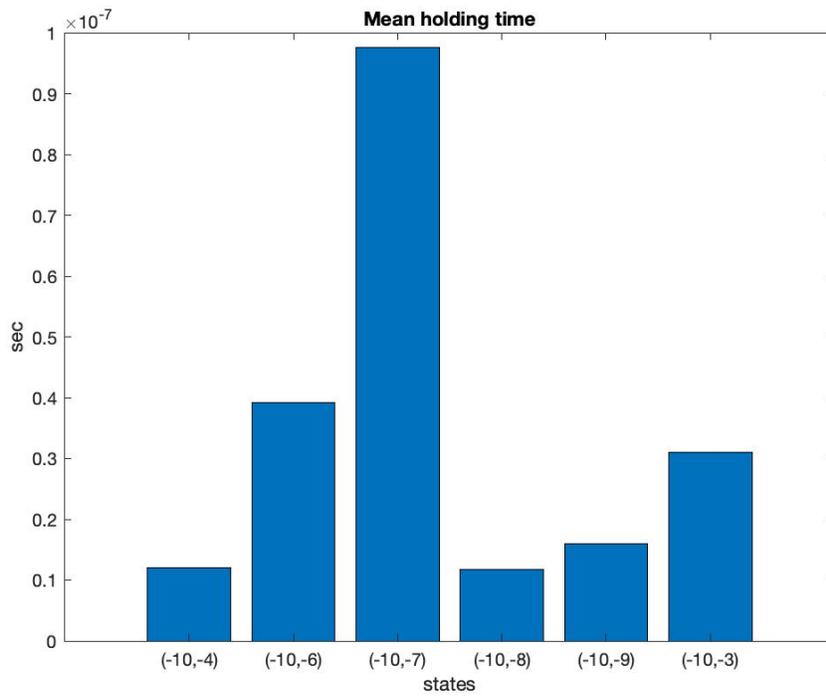


Figure 32: Mean holding times and transition rates for  $Lk_{rel} = -10$ .

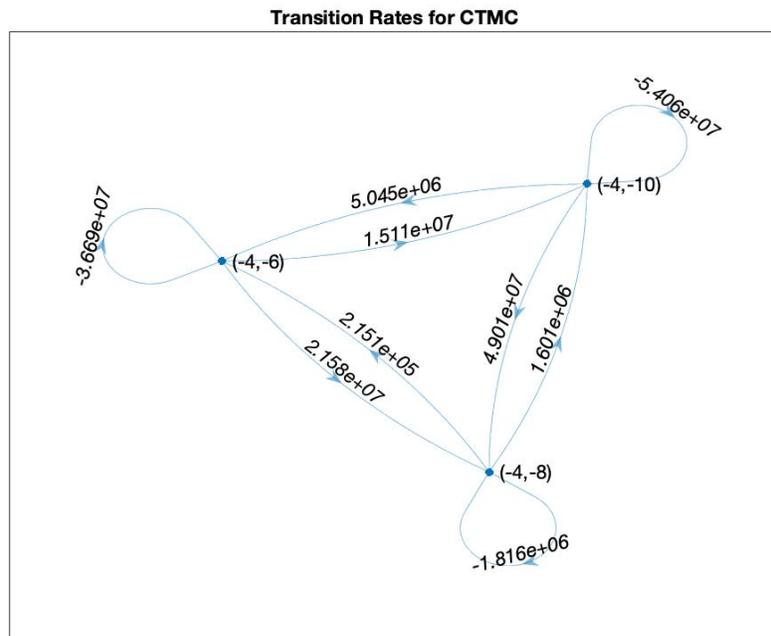
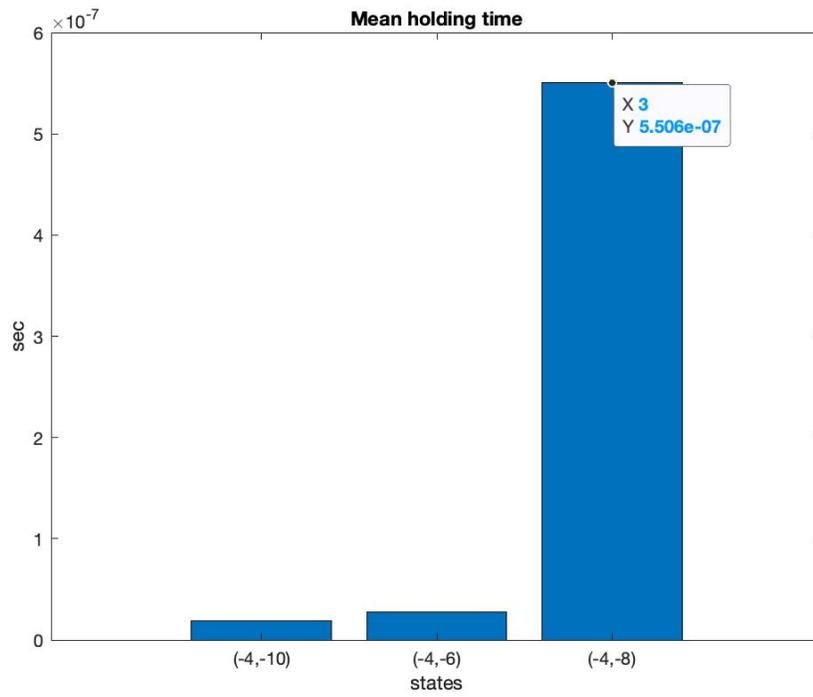


Figure 33: Mean holding times and transition rates for  $Lk_{rel} = -4$ .

## Appendix B SD-SGIB parameters

From the SD-SGIB method, the Matrix  $\Gamma_{ij}^n$  is defined as

$$[\Gamma_{ij}^n] = \begin{pmatrix} -\frac{\theta_1^n}{\kappa^n} \sin \zeta^n + \frac{\theta_2^n \cos \zeta^n}{2 \tan \frac{1}{2} \kappa^n} & -\frac{\theta_2^n}{\kappa^n} \sin \zeta^n - \frac{\theta_1^n \cos \zeta^n}{2 \tan \frac{1}{2} \kappa^n} & \tan \frac{1}{2} \kappa^n \cos \zeta^n \\ \frac{\theta_1^n}{\kappa^n} \cos \zeta^n + \frac{\theta_2^n \sin \zeta^n}{2 \tan \frac{1}{2} \kappa^n} & \frac{\theta_2^n}{\kappa^n} \cos \zeta^n - \frac{\theta_1^n \sin \zeta^n}{2 \tan \frac{1}{2} \kappa^n} & \tan \frac{1}{2} \kappa^n \sin \zeta^n \\ & -\frac{\theta_2^n}{2} & \frac{\theta_1^n}{2} & 1 \end{pmatrix} \quad (\text{B.1})$$

and

$$[Z_{ij}(\alpha)] = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (\text{B.2})$$

$$[Y_{ij}(\alpha)] = \begin{pmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix}, \quad (\text{B.3})$$

which defines  $\theta_1^n, \theta_2^n, \theta_3^n$  by the relations:  $\zeta^n = \frac{1}{2}\theta_3^n - \gamma^n$ ,  $\kappa^n = \sqrt{(\theta_1^n)^2 + (\theta_2^n)^n}$ ,  $\eta^n = \frac{1}{2}\theta_3^n + \gamma^n$ , and  $\tan \gamma^n = \frac{\theta_1^n}{\theta_2^n}$  from the Euler-angle system [18].

For each  $j$ , we define the components of the skew matrix  $[_j\Lambda_{kl}^n]$  as follows [16]: For  $j = 1$ :

$${}_1\Lambda_{12}^n = \frac{\theta_2^n \left( 1 - \cos \left( \frac{1}{2} \kappa^n \right) \right)}{(\kappa^n)^2}, \quad (\text{B.4})$$

$${}_1\Lambda_{13}^n = \frac{\theta_1^n \theta_2^n \left( 2 \sin \left( \frac{1}{2} \kappa^n \right) - \kappa^n \right)}{2(\kappa^n)^3}, \quad (\text{B.5})$$

$${}_1\Lambda_{23}^n = \frac{1}{2} + \left( \frac{\theta_2^n}{\kappa^n} \right)^2 \frac{2 \sin \left( \frac{1}{2} \kappa^n \right) - \kappa^n}{2\kappa^n}. \quad (\text{B.6})$$

For  $j = 2$ :

$${}_2\Lambda_{12}^n = \frac{\theta_1^n \left( \cos \left( \frac{1}{2} \kappa^n \right) - 1 \right)}{(\kappa^n)^2}, \quad (\text{B.7})$$

$${}_2\Lambda_{13}^n = \left(\frac{\theta_1^n}{\kappa^n}\right)^2 \frac{\kappa^n - 2 \sin\left(\frac{1}{2}\kappa^n\right)}{2\kappa^n} - \frac{1}{2}, \quad (\text{B.8})$$

$${}_2\Lambda_{23}^n = \frac{\theta_1^n \theta_2^n \left(\kappa^n - 2 \sin\left(\frac{1}{2}\kappa^n\right)\right)}{2(\kappa^n)^3}. \quad (\text{B.9})$$

For  $j = 3$ :

$${}_2\Lambda_{12}^n = \frac{1}{2} \cos\left(\frac{1}{2}\kappa^n\right), \quad (\text{B.10})$$

$${}_3\Lambda_{13}^n = -\frac{\theta_1^n}{2\kappa^n} \sin\left(\frac{1}{2}\kappa^n\right), \quad (\text{B.11})$$

$${}_3\Lambda_{23}^n = -\frac{\theta_2^n}{2\kappa^n} \sin\left(\frac{1}{2}\kappa^n\right). \quad (\text{B.12})$$

The coefficients of matrices  $f_{ij}$ ,  $g_{ij}$ , and  $h_{ij}$  as in the method [16] are from [59, 44]. For each successive pair of bases, the coefficients of these matrices in the index order  $(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)$  listed vertically are listed in tables 6, 7 and 8.

CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
0.107	0.121	0.135	0.197	0.157	0.149	0.133	0.190	0.130	0.120
0.000	-0.004	0.000	0.014	0.010	0.007	-0.003	0.000	0.011	0.000
0.000	-0.003	0.000	0.034	0.003	0.002	0.019	0.000	0.011	0.000
0.000	-0.004	0.000	0.014	0.010	0.007	-0.003	0.000	0.011	0.000
0.038	0.061	0.053	0.067	0.068	0.064	0.044	0.068	0.077	0.095
0.018	0.020	0.031	0.024	0.014	0.034	0.024	0.024	0.021	0.012
0.000	-0.003	0.000	0.034	0.003	0.002	0.019	0.000	0.011	0.000
0.018	0.020	0.031	0.024	0.014	0.034	0.024	0.024	0.021	0.012
0.103	0.066	0.056	0.080	0.086	0.101	0.086	0.091	0.099	0.069

Table 6: Elastic coefficients for  $f_{ij}$  representing twisting and bending moduli in units  $kT/deg^2$

CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
-0.352	-0.316	-0.164	-0.282	-0.330	-0.263	-0.369	-0.162	-0.113	-0.248
0.000	0.028	0.000	0.038	0.063	0.017	0.028	0.000	0.065	0.000
0.000	-0.029	0.000	0.137	0.125	0.168	-0.038	0.000	0.018	0.000
0.000	-0.038	0.000	-0.080	0.009	0.086	0.021	0.000	0.175	0.000
0.038	0.029	-0.014	-0.088	0.116	-0.211	-0.071	0.047	-0.051	0.355
-0.117	-0.127	-0.088	-0.068	-0.104	-0.254	-0.237	-0.122	-0.118	-0.227
0.000	0.019	0.000	-1.411	-0.867	-0.673	-0.679	0.000	0.057	0.000
0.003	-0.008	-0.163	-0.403	-0.026	-0.120	0.141	0.229	0.240	0.724
-0.357	-0.310	-0.514	-0.803	-0.587	-0.254	-0.383	-0.350	-0.566	-0.635

Table 7: Elastic coefficients for  $g_{ij}$  representing the coupling between bending, shearing, and slide in units  $kT/deg/\text{\AA}$

CG	CA	TA	AG	GG	AA	GA	AT	AC	GC
2.510	3.167	3.459	2.963	3.106	7.161	5.158	4.063	4.077	2.449
0.000	0.725	0.000	0.128	0.081	0.783	1.840	0.000	1.476	0.000
0.000	0.074	0.000	0.656	1.177	2.772	3.285	0.000	-0.047	0.000
0.000	0.725	0.000	0.128	0.081	0.783	1.840	0.000	1.476	0.000
3.521	2.262	2.007	4.498	3.841	8.107	4.162	8.886	11.452	5.582
3.214	2.362	2.330	3.732	4.365	1.777	2.124	5.428	6.243	6.349
0.000	0.074	0.000	0.656	1.177	2.772	3.285	0.000	-0.047	0.000
3.214	2.362	2.330	3.732	4.365	1.777	2.124	5.428	6.243	6.349
22.628	20.875	37.210	34.532	33.325	37.297	24.466	38.484	32.711	31.530

Table 8: Elastic coefficients for  $h_{ij}$  representing shift, slide and rise energies in units  $kT/\text{\AA}^2$

## Appendix C Kinetoplast DNA sequences

The exact DNA sequences for *T. brucei* and *C. fasciculata* from [57] are as follows:

*T. brucei*:

GAAAAAACC CAAAATCTTATGGGCGTGCAAAAATACACATACACAAATCC  
CGTGCTATTTTTGGGCTGTTTTTTAGGTCCGAGGTACTTCGAAAGGGGTTG  
GTGTAATACACACATGGTTTTTCTCGAGATTCAGGGTTTTGGGGTGAT  
ATCTAGTGTAATTAATATTGTGTTTTTATAGTCTACTTAAGGAATAAAAT  
ATAGTAATAGATAAATATATAAGTTAGATATATAGCAATTATAATTAAAC  
TGAAGAGGTTATAATACCTCGAAACTCGCGGTGATGATTTTTATATTTATT  
TCTTATATTACTATTTATTAATTTATTCTCATTCTCGGGATTACCTAGTG  
GGAAAGAAATGAGATAATAGATATGTATTGTAGTATTATAATGATATATA  
TAGATATAAGATCAACAAAACCTGCCATTTTCTATAGTGAGTGATGATATA  
TTATAATTAAATGATATTATTATTAATAATCTATTTATTATTTTATTTATT  
TATGGAGGATGAAATTAATGGGATTATTCGGTGGTAGAGTGGGATTAATG  
TGATAAATACTGCTTCATATCTGCGTCTAGAAAGGTAAAATAATTTAATA  
GATAAGTAGTAGTAATATATTAAGTTAAATATGATATATATAATGACTAA  
CTAAACTGATAAAGCAGTAGAAGAGACGATGTAATATATATAATTATTAA  
TTACTGTTAATATATCTATTATTTTATTTTATTACTAAGGGAAAGAGAAG  
ATATTAATAGATAGAACAATGAATAGATAATAAAGTGAAGAATTATATAA  
ATCCAATAAAAGTGAAATAGAATCTGAGAGACTGTGATTTATACCTGTAA  
AATGAGATTTATATTTATACTGTATCTATTATTTTATTTATGATTGGAGT  
GGTGAGATAATGGAGGGAATCAGAAGTGAAGCACAGAGTTATTAATTGAA  
AGAGATTAGTAGTGGATGTAAGAAAATTATGGAAAATCGGGTAAAAATCG  
AAGAAAATGGCTTG

*C. fasciculata*:

CCTCCTGGCAGGGGGTTTGGCGGGGTTCTAGCCCGATTCGGGGCGTTCT  
GCGGGGGTTTTTTTCTGGTCTGGGCGCGGGTTTGGGCTGGTTTGGGCTGG

GTTTGGACTGTTTGTGCTAGTTGGGCGCTACGGACTGCTTTGCGATGGTG  
CGCGGGGGGGTGGTTTCACCACTATTCTGATTGTTGTTTTCGCTCCTTGG  
TGGGGTTTATATGCGCTCCGTTCCGGTCGTATTCTGGAATTTTGGGGTTTG  
CCAAAAGTGAACCTCCGACATTTCTCGCGGGGTTAATATATAGACTAGAC  
GCGTCGTTGTTAATTTTGCCATGGGTGTGTTTGTGTTGTTCTGGTGCCCG  
GAGGCTGATTTCCGGGGTCCCGCGAAAATCAGAAACGGTCTCGGGTAGG  
GGCGTTCTGCGAAAATCGACTTTTGATACAGGAAATCCCGTTCAAAAATC  
GTGATTTTTTCAATTTTGGAGGCAAACCTGGGGATTTCCGGGGTTGGTGTA  
GTATTTCTGGGTCCGGGGGTCTGAGGGGTCCAATACCTTCTGATAGAT  
TCGCCTTTTATAGGCGTTCTGCTCGTTACTTTTATAACTTTAGTTGCTCT  
TATGTTTGCTATAAATATATAGCTTTGATTTCTAGACTTGCTTGCGTTTA  
AAGTTGTTTGCGCGGGCTTCCTGTGGGTTTTGTTTTGGTGTGGCTTGTTA  
TTTGTGATTTTGCTAGTTTCTTTGCGGTTTTGTCTATTTTTAGTTGTTTT  
GTGTATTTGTACTTTACGTTTTTTGGTTGTTGTGGCTTTGCGTTTTTATA  
CTGCTTTGCTGGCTTGGTTGGTTATGTTGGCTTGTGGTTTTGTTTTTATT  
TTGTGTGTTTCGTGGGTGTTGATGTTTTTGTGTTTTTTGGTTGCTTTTGTA  
GCTTTAGGGTGGTTACTATTAGTTTTCTTTTTGTTTTCGCTTTTGTTCTG  
GGGTTTGTGATTAGCTTTGGGGGTTTCGTGGTTGTTGTGCCTGTGTTATT  
TAGTTGTGTCCCACGGTGGGTTCCGGCTGCTGGTTGGGTGTGCTTACTGTT  
TCTTGTTATGTTGGTATGTATGCTATGTTGCTGCTAGTTGTTTTTATGGT  
TTTGCCTTGTCTGTTGCGTGTGTATGTGTTTATTTATTTGATTGTTTAG  
ATTGTTTTAATAACTTTGTGTTGCATTTGTTTTAGATTTAAAAGGCTTGT  
TGTTGTGTTGTTGTGTTGTTGCTATTGTTTTGATTTGTCTTTGCTGCTCA  
CTGCGTGGTACACATTGATTGCTCGAGTGCATGTTGTGTTGATAGCTTC  
TTGTAGTTTTTCGTTGTTTGTAAATGTTGGTGTGGTGTGGTGTCTCG  
GTTGCCACCTGTGGTTTCTTTAAGTGTGTTGTTGCTGTTTATTTGTTGTT  
TGTTGGTTATTGGTTTATTGTTTGCATTAGCCTTCTGTGGGTTTGAAACT  
GTTGTATTCTTGTTTACTTGGGTGGTTTATCTTGATTTGGCTTTATTGTT  
GGGTACTTGTGTTGTTTGTGTTTATGCTCTTTCTTTGTTGCTGGT  
GCTTGCTGAACTGTTTGTGGTTGTTGGGGCGTGTGGGTTTGAGGGTGT

TTTGGGGTGGTTTGGGGTGCCCGCGAAATATCAGAAATGGTCTCGGGTAG  
GGGCGTTCTGCGAAAATCGACTTTTGATACAGGAAATCCCGTTCAAAAAT  
GGCAGTTTTCTCGATTTTGGAGGCTCGGCTGGGATTTCCGGGGTTGGTGT  
AGTCATTCTGGGTCCGGGCGGGTCTGGCGGGGGTTCTGTTAAACGCGGG  
GGTTGCTTCAGTGCTGTTATTCATCCGCTTCGAAGTTAATTTTCGTTGTT  
TAGCTTGTAGTTTGCTCTGTGGGGTTCTGAAATTGCCATTTTGGCGCTT  
TTTATCGTTGGGTGTGTACGATTGCGCGGCGTCGCTTTCGACGACGGGGC  
CGAGTGTTCTTGCACGAGGTCGGGAGCGCTAGCCCGTCGTTGAATGCAAG  
TGCAACATACGTGAGGCCGCGGACGAGCCCCGTCCCTGAAAGGGGAGGAG  
GCTAGTTGACGCTAGGCCGGAGCCACGAATGGCGAGCAAAGCTAGCCCGA  
GCCATGAACGCGAACGGCCGGGGAGACTTGCCGGGGAAAGGGGAGGGTCA  
AGTACCAGGCTCGAACAGTATAACAACGACAAGACGCCGCTGCATCGCCAT  
ACTTTTATCTTTCGCACATTCATGTGTGAACTAGTTTGCTTTAACACGGT  
GCCTCGTTTAACTCTTGCGGGTGGTAGACAGACTCTAAAGCAGATGCG  
TAGACGTTTCAGATTTTGATTTTGGAGTGCGTTTTTGGCCATTTTTTGCCC  
ATTTTTCCCTTAAAATTCAATAAAATTGCGGGATTTTTTACCATTTTTGT  
CGATTTTTGGGGTATTTTCGCTGTTTTTTGGCATTTTTTTGGCCATTTTTC  
CTTGATTTTTGGGCACTTTTCGGGCTCCAAAAAAGTAACCTCGCGATTTTC  
GCCTGGAATTTTAGGC

## Bibliography

- [1] Javier Arsuaga and Yuanan Diao. DNA knotting in spooling like conformations in bacteriophages. *Computational and Mathematical Methods in Medicine*, 9(3-4):303–316, 2008.
- [2] Javier Arsuaga, Yuanan Diao, Michele Klingbeil, and Victor Rodriguez. Properties of topological networks of flexible polygonal chains. *Computational and Mathematical Biophysics*, 2(1):98–106, 2014.
- [3] Javier Arsuaga, Joaquim Roca, and De Witt Sumners. *Topology of viral DNA*. World Scientific Publishing, 2010.
- [4] Javier Arsuaga, Mariel Vázquez, Paul McGuirk, Sonia Trigueros, De Witt Sumners, and Joaquim Roca. DNA knots reveal a chiral organization of DNA in phage capsids. *Proceedings of the National Academy of Sciences*, 102(26):9165–9169, 2005.
- [5] Javier Arsuaga, Mariel Vázquez, Sonia Trigueros, De Witt Sumners, and Joaquim Roca. Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proceedings of the National Academy of Sciences*, 99(8):5373–5377, 2002.
- [6] Paul J. Atzberger, Peter R. Kramer, and Charles S. Peskin. A stochastic immersed boundary method for fluid-structure dynamics at microscopic length scales. *Journal of Computational Physics*, 224:1255 – 1292, 2006.
- [7] J. Bonet Avalos and A.D. Mackie. Dissipative particle dynamics with energy conservation. *EPL (Europhysics Letters)*, 40(2):141, 1997.
- [8] Sören Bartels and Philipp Reiter. Stability of a simple scheme for the approximation of elastic knots and self-avoiding inextensible curves. *arXiv preprint arXiv:1804.02206*, 2018.
- [9] Simon Blatt, Philipp Reiter, and Armin Schikorra. Harmonic analysis meets critical knots. Critical points of the möbius energy are smooth. *Trans Amer Math Soc*, 368(9):6391–6438, 2016.

- [10] Simon Blatt, Philipp Reiter, and Armin Schikorra. On O’hara knot energies i: Regularity for critical knots. *arXiv preprint arXiv:1905.06064*, 2019.
- [11] Simon Blatt, Philipp Reiter, Armin Schikorra, and Nicole Vorderobermeier. Scale-invariant tangent-point energies for knots. *arXiv preprint arXiv:2104.10238*, 2021.
- [12] Simon Blatt and Nicole Vorderobermeier. On the analyticity of critical points of the möbius energy. *Calculus of Variations and Partial Differential Equations*, 58(1):1–28, 2019.
- [13] John F. Brady and Georges Bossis. Stokesian dynamics. *Annual Review of Fluid Mechanics*, 20(1):111–157, 1988.
- [14] G. Calugareanu. Sur les classes d’isotopie des noeuds tridimensionnels et leurs invariants. *Czechoslovak Math J.*, 11(86):588–625, 1961.
- [15] Junghuei Chen, Carol A. Rauch, James H. White, Paul T. Englund, and Nicholas R. Cozzarelli. The topology of the kinetoplast DNA network. *Cell*, 80(1):61–69, 1995.
- [16] Bernard D. Coleman, Wilma K. Olson, and David Swigon. Theory of sequence-dependent DNA elasticity. *The Journal of Chemical Physics*, 118(15):7127–7140, 2003.
- [17] Bernard D. Coleman and David Swigon. Theory of supercoiled elastic rings with self-contact and its application to DNA plasmids. *Journal of Elasticity and the Physical Science of Solids*, 60(3):173–221, 2000.
- [18] Bernard D. Coleman and David Swigon. Theory of self-contact in DNA molecules modeled as elastic rods. *ATTI DEI CONVEGNI LINCEI-ACCADEMIA NAZIONALE DEI LINCEI*, 177:281–296, 2001.
- [19] Bernard D. Coleman and David Swigon. Theory of self-contact in Kirchoff rods with applications to supercoiling of knotted and unknotted DNA plasmids. *Phil. Trans. R. Soc. Lond.*, 362:1281 – 1299, 2004.
- [20] Bernard D. Coleman, David Swigon, and Irwin Tobias. Elastic stability of DNA configurations. ii. Supercoiled plasmids with self-contact. *Physical Review E*, 61(1):759, 2000.

- [21] Bernard D Coleman, Irwin Tobias, and David Swigon. Theory of the influence of end conditions on self-contact in DNA loops. *The Journal of Chemical Physics*, 103(20):9101–9109, 1995.
- [22] Peter Cromwell. *Knots and Links*. Cambridge University Press, 2004.
- [23] Luke Czapla, David Swigon, and Wilma K. Olson. Sequence-dependent effects in the cyclization of short DNA. *Journal of Chemical Theory and Computation*, 2(3):685–695, 2006.
- [24] Y. Diao, K. Hinson, R. Kaplan, M. Vazquez, and J. Arsuaga. The effects of density on the topological structure of the mitochondrial DNA from trypanosomes. *Journal of Mathematical Biology*, 64(6):1087–1108, 2012.
- [25] Y. Diao, K. Hinson, Y. Sun, and J. Arsuaga. The effect of volume exclusion on the formation of DNA minicircle networks: implications to kinetoplast DNA. *Journal of Physics A: Mathematical and Theoretical*, 48(43):435202, oct 2015.
- [26] Yuanan Diao, Victor Rodriguez, Michele Klingbeil, and Javier Arsuaga. Orientation of DNA minicircles balances density and topological complexity in kinetoplast DNA. *Plos One*, 10(6):e0130998, 2015.
- [27] Albert Einstein. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.
- [28] Flapan Erica and Helen Wong. *Topology and Geometry of Biopolymers*, volume 746. American Mathematical Society, 2020.
- [29] Donald L. Ermak and J. Andrew McCammon. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics*, 69(4):1352–1360, 1978.
- [30] Lang Feng, Ruojie Sha, Nadrian C. Seeman, and Paul M. Chaikin. Topological interaction by entangled DNA loops. *Physical Review Letters*, 109(18):188301, 2012.
- [31] Michael H. Freedman, Zheng-Xu He, and Zhenghan Wang. Möbius energy of knots and unknots. *Annals of Mathematics*, 139(1):1–50, 1994.
- [32] Crispin W. Gardiner et al. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.

- [33] Stephanie Geggier and Alexander Vologodskii. Sequence dependence of DNA bending rigidity. *Proceedings of the National Academy of Sciences*, 107(35):15421–15426, 2010.
- [34] Alexandra Gilsbach, Philipp Reiter, and Heiko von der Mosel. Symmetric elastic knots. *arXiv preprint arXiv:2105.08558*, 2021.
- [35] H.P. Grimm, A.B. Verkhovskiy, A. Mogilner, and J-J. Meister. Analysis of actin dynamics at the leading edge of crawling cells: implications for the shape of keratocyte lamellipodia. *European Biophysics Journal*, 32(6):563–577, 2003.
- [36] M.A. El Hassan and C.R. Calladine. The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *Journal of Molecular Biology*, 251(5):648–664, 1995.
- [37] Xia Hua, Diana Nguyen, Barath Raghavan, Javier Arsuaga, and Mariel Vázquez. Random state transitions of knots: a first step towards modeling unknotting by type ii topoisomerases. *Topology and its Applications*, 154(7):1381–1397, 2007. Special Issue: The Third Joint Meeting Japan-Mexico in Topology and its Applications.
- [38] B. Garc Ía-archilla, J.M. Sanz-Serna, and Robert D. Skeel. Long-time-step methods for oscillatory differential equations. In *SIAM J. Sci. Comput*, 1999.
- [39] Robert E. Jensen and Paul T. Englund. Network news: the replication of kinetoplast DNA. *Annual Review of Microbiology*, 66:473–491, 2012.
- [40] David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.
- [41] Sookyung Lim Yongsam Kim and David Swigon. Dynamics of an electrostatically charged elastic rod in fluid. *Proc. R. Soc.*, A(467):569 – 590, 2010.
- [42] Konstantin Klenin and Jörg Langowski. Computation of writhe in modeling of supercoiled DNA. *Biopolymers: Original Research on Biomolecules*, 54(5):307–317, 2000.
- [43] Anthony J.C. Ladd. Numerical simulations of particulate suspensions via a discretized Boltzmann equation. Part 1. Theoretical foundation. *Journal of Fluid Mechanics*, 271:285–309, 1994.

- [44] Richard Lavery, Krystyna Zakrzewska, David Beveridge, Thomas C. Bishop, David A. Case, Thomas Cheatham III, Surjit Dixit, B. Jayaram, Filip Lankas, Charles Laughton, et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Research*, 38(1):299–313, 2010.
- [45] Michael Levitt. Protein folding by restrained energy minimization and molecular dynamics. *Journal of Molecular Biology*, 170(3):723–764, 1983.
- [46] Sookyung Lim, Anca Ferent, X. Sheldon Wang, and Charles S. Peskin. Dynamics of a closed rod with twist and bend in fluid. *SIAM J. Sci. Comput.*, 31(1):273 – 302, 2008.
- [47] P. Liu, R. Polischuk, Y. Diao, and J. Arsuaga. Characterizing the topology of kinetoplast DNA using random knotting. *Topology and Geometry of Biopolymers*, 746:17, 2020.
- [48] Penyu Liu, Abby Pecoske Fulton, Javier Arsuaga, and David Swigon. DNA sequence effects on the topology of kinetoplast networks. *preprint*, 2021.
- [49] Xin Liu and Renzo L. Ricca. Knots cascade detected by a monotonically decreasing sequence of values. *Scientific Reports*, 6(1):1–7, 2016.
- [50] Xin Liu, Renzo L. Ricca, and Xin-Fei Li. Minimal unlinking pathways as geodesics in knot polynomial space. *Communications Physics*, 3(1):1–9, 2020.
- [51] Julius Lukeš, D. Lys Guilbride, Jan Votýpka, Alena Zíková, Rob Benne, and Paul T. Englund. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryotic Cell*, 1(4):495–502, 2002.
- [52] Gerald S. Manning. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly Reviews of Biophysics*, 11(2):179–246, 1978.
- [53] GS Manning. Counterion condensation theory of attraction between like charges in the absence of multivalent counterions. *The European Physical Journal E*, 34(12):1–18, 2011.
- [54] C.A. Marsh and J.M. Yeomans. Dissipative particle dynamics: The equilibrium for finite time steps. *EPL (Europhysics Letters)*, 37(8):511, 1997.

- [55] Henry Keith Moffatt. The energy spectrum of knots and links. *Nature*, 347(6291):367–369, 1990.
- [56] Petri Nikunen, Mikko Karttunen, and Ilpo Vattulainen. How would you integrate the equations of motion in dissipative particle dynamics simulations? *Computer Physics Communications*, 153(3):407–423, 2003.
- [57] National Center of Biotechnology Information. Nucleotide [Internet] Bethesda (md): National Library of Medicine (US), 2004.
- [58] Emmanuel O. Ogbadoyi, Derrick R. Robinson, and Keith Gull. A high-order transmembrane structural linkage is responsible for mitochondrial genome positioning and segregation by flagellar basal bodies in trypanosomes. *Molecular biology of the cell*, 14(5):1769–1779, 2003.
- [59] Wilma K. Olson, Andrey A. Gorin, Xiang-Jun Lu, Lynette M. Hock, and Victor B. Zhurkin. DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proceedings of the National Academy of Sciences*, 95(19):11163–11168, 1998.
- [60] Jun O’Hara. Energy of a knot. *Topology*, 30(2):241–247, 1991.
- [61] Charles S. Peskin. Flow patterns around heart valves: a numerical method. *Journal of Computational Physics*, 10(2):252–271, 1972.
- [62] Charles S. Peskin. Numerical analysis of blood flow in the heart. *Journal of Computational Physics*, 25(3):220–252, 1977.
- [63] Charles S. Peskin. Fluid dynamics of the heart and its valves. In *APS Division of Fluid Dynamics Meeting Abstracts*, pages 3INV–02, 1997.
- [64] Arjun Raj and Charles S. Peskin. The influence of chromosome flexibility on chromosome transport during anaphase A. *Proceedings of the National Academy of Sciences*, 103(14):5349–5354, 2006.
- [65] Peter Reimann and Peter Hänggi. Introduction to the physics of Brownian motors. *Applied Physics A*, 75(2):169–178, 2002.

- [66] Renzo L. Ricca and Francesca Maggioni. On the groundstate energy spectrum of magnetic knots and links. *Journal of Physics A: Mathematical and Theoretical*, 47(20):205501, 2014.
- [67] Valentin V. Rybenkov, Nicholas R. Cozzarelli, and Alexander V. Vologodskii. Probability of DNA knotting and the effective diameter of the DNA double helix. *Proceedings of the National Academy of Sciences*, 90(11):5307–5311, 1993.
- [68] Theresa A. Shapiro and Paul T. Englund. The structure and replication of kinetoplast DNA. *Annual Review of Microbiology*, 49(1):117–143, 1995.
- [69] Koya Shimokawa, Kai Ishihara, Ian Grainge, David J. Sherratt, and Mariel Vázquez. FtsK-dependent XerCD-dif recombination unlinks replication catenanes in a stepwise manner. *Proceedings of the National Academy of Sciences*, 110(52):20906–20911, 2013.
- [70] Charles S. Peskin. The immersed boundary method. *Acta Numerica*, 11:479–517, 2002.
- [71] Robert Stolz, Masaaki Yoshida, Reuben Brasher, Michelle Flanner, Kai Ishihara, David J. Sherratt, Koya Shimokawa, and Mariel Vázquez. Pathways of DNA unlinking: a story of stepwise simplification. *Scientific Reports*, 7(1):1–11, 2017.
- [72] David Swigon. The mathematics of DNA structure mechanics, and dynamics. *The IMA Volumes in Mathematics and its Applications*, 150:293 – 319, 2009.
- [73] David Swigon, Sookyung Lim, and Yongsam Kim. Dynamical simulations of DNA supercoiling and compression. *Biochem. Soc. Trans.*, 41(2):554 – 558, 2013.
- [74] Sonia Trigueros, Javier Arsuaga, Maria E. Vazquez, De Witt Summers, and Joaquim Roca. Novel display of knotted DNA molecules by two-dimensional gel electrophoresis. *Nucleic Acids Research*, 29(13):e67–e67, 2001.
- [75] Jerome Vinograd and Jacob Lebowitz'. Physical and topological properties of circular DNA. *The Journal of General Physiology*, 49(6):103–125, 1966.
- [76] James C Wang. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Reviews Molecular Cell Biology*, 3(6):430–440, 2002.

- [77] James D Watson and Francis H.C. Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.
- [78] T.P. Westcott, I. Tobias, and W.K. Olson. Modeling self-contact forces in the elastic theory of DNA supercoiling. *J. Chem. Phys.*, 107:3967–3980, 1997.
- [79] James H. White. Self-linking and the Gauss integral in higher dimensions. *American Journal of Mathematics*, 91(3):693–728, 1969.
- [80] Wikipedia contributors. Structure — Wikipedia, the free encyclopedia, 2021. [Online; accessed 28-July-2021].
- [81] Loren Dean Williams and L. James Maher III. Electrostatic mechanisms of DNA deformation. *Annual Review of Biophysics and Biomolecular Structure*, 29(1):497–521, 2000.
- [82] C. Wilson Xu, Jane C. Hines, Michele L. Engel, David G. Russell, and Dan S. Ray. Nucleus-encoded histone H1-like proteins are associated with kinetoplast DNA in the trypanosomatid *Crithidia fasciculata*. *Molecular and Cellular Biology*, 16(2):564–576, 1996.