Biopolymer Simulations: From Next-Generation Genomics to Consumer Products

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Xiaolan Li

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Kevin D. Dorfman

April, 2018

Acknowledgements

I would like to first and foremost thank my advisor Prof. Kevin Dorfman for taking me on and allowing me to carry out scientific research projects that both challenging and impactful. We had a lot of inspiring discussions that help me gain fundamental and comprehensive knowledge of various subjects. I greatly appreciate his deep understanding and enthusiasm on research topics that guided students through their toughest moments in the research. He had supported me when I pursued a summer internship to further develop my professional goal. He also provided me with a number of opportunities to collaborate with both researchers and entrepreneurs. Kevin has been a great mentor and advisor to me. Thank you.

I would like to thank my other committee members, Prof. Dave Morse, Prof. Xiang Cheng and Prof. Ilja Siepmann for reading and commenting on my work. I also would like to thank Akash Arora, Hui-Min Chuang, Guo Kang Cheong and Vaidyanathan Sethuraman for proofreading parts of the thesis.

I appreciate the great support from the people in Dorfman group over the past five years. I feel so fortunate to work with you: Zhicheng Long, Sarit Dutta, Doug Tree, Damini Gupta, Julian Struksheats, Akash Arora, Pranav Agrawal, Seunghwan Shin, Guo Kang Cheong and Vaidyanathan Sethuraman. Especially, I would like to thank Abhiram Muralidhar, who mentored me when I joined the group, taught me a lot of PERM techniques, and provided me with valuable career advice. Aashish Jain has been a great coworker and we have a lot of stimulating discussions on various research topics. Michael McGovern helped me with learning LAMMPS when I started the methylcellulose project. Hui-Min has been a great friend who added much laughter to my time in the group. I also appreciate the collaboration with the team at BioNano Genomics. Especially I want to thank Jeff Reifenberger for his suggestions and inputs

on the direction of part of this doctoral work.

I would like to thank my parents, Yi Li and Jinmei Zhao, for all their support and resource that have led me to my Ph.D. study at Minnesota. You have inspired me to pursue a Ph.D., taught me to always strive for the best in the work and always enjoy the beauty of life. I am grateful to my friends for their support and accompany, which made me feel I was not alone in the U.S..

Finally, I acknowledge the funding and resources that support this doctoral work. The work presented here was supported by Bill and Triana Silliman Fellowship and First-Year Graduate Student Fellowship in the Department of Chemical Engineering and Material Science; the National Science Foundation (NSF) Grant No. CBET-1262286 to Kevin Dorfman and NSF Grant No. CBET-1254340 Camille & Henry Dreyfus Foundation to Charles Schroeder; NSF through the University of Minnesota Materials Science Research and Engineering Center under Award No. DMR-1420013. During my last year of doctoral study, I was also supported by Doctoral Dissertation Fellowship in the Graduate School of the University of Minnesota. The computational resources were provided in part by the Minnesota Supercomputing Institute at the University of Minnesota.

Abstract

Biopolymers have many unique properties which play an essential and pervasive role in everyday life, thus making them attractive for engineering applications. Understanding how the particular properties of biopolymers give rise to important applications in technology remains a long-standing challenge. Although biopolymers can have different chemistries, they share some common physical properties: high molecular weights, stiff backbones, and complex internal structures. Computer simulation, therefore, plays quite an important role since it provides a way to study a generic model that, by changing the parameters appearing in the model, permits studying a wide variety of biopolymers. Specifically, we focus on two such biopolymers: DNA and methylcellulose. This thesis focuses on studying the universal properties of the two aforementioned biopolymers using novel molecular simulation techniques.

DNA attracts particularly strong interest not only because of its fascinating double-helix structure but also because DNA carries biological information. Genomic mapping is emerging as a new technology to provide information about large-scale genomic structural variations. In this context, the conformation and properties of the linearized DNA are only beginning to be understood. With a Monte Carlo chain growth method known as pruned-enriched Rosenbluth method, we explore the force-extension relationship of stretched DNA. In this scenario, external forces and confinement are two fundamental and complementary aspects. We begin by stretching a single DNA in free solution. This allows separation of restrictions imposed by forces from that by walls. This work shows that the thickness of DNA plays an important role in the force-extension behavior. The key outcome is a new expression that approximates the force-extension behavior with about 5% relative error for all range of forces. We then analyze slit-confined DNA stretched by an external force. This work predicted a new regime in the force-extension behavior that features a mixed effect of both sensible DNA volume and sensible wall effects. We anticipate such a complete description of the force-extension of DNA will prove useful for the design of new genomic mapping technologies.

The dissertation also involves another biopolymer, methylcellulose, which has an extremely wide range of commercial uses. Methylcellulose is thermoresponsive polymer that undergoes a morphological transition at elevated temperature, forming uniform

diameter fibrils. However, mechanisms behind the solution-gel transition are poorly understood. Following the computational studies by Huang *et al.* [1], we apply Langevin dynamics simulations to a coarse-grained model that produces collapsed ring-like structures in dilute solution with a radius close to the fibrils observed in experiments. We show that the competition between the dihedral potential and self-attraction causes these collapsed states to undergo a rapid conformational change, which helps the chain to avoid kinetic traps by permitting a transition between collapsed states.

We expect our findings from computational studies of biopolymers will not only provide a deep understanding of semiflexible polymer physics but also inspire novel engineering applications relying on the properties of biopolymers.

Contents

List of Tables

List of Figures

Statement of the Author's Contributions

Much of the research contained in this dissertation was originated from collaborative efforts. Some chapters have also appeared as articles in various journals. Because many of the articles list multiple authors, I will point out my contributions to the research.

I wrote most of Chapters 1, 2, and 3. Multiple figures in Chapters 1, 2, and 3 are adopted from publications without my contribution, and the related publications are cited in the corresponding figure captions.

Chapter 4 is based on X. Li, C. M. Schroeder, and K. D. Dorfman, "Modeling the stretching of wormlike chains in the presence of excluded volume," *Soft Matter*, vol. 11, pp. 5947-5954, 2015. In this paper, we use pruned-enriched Rosenbluth method (PERM) to explore the force-extension behavior of real semiflexible polymer chains stretched in free solution. The original PERM code was written by two former members of Dorfman group, D. R. Tree, and A. Mulralidhar. I developed an updated version of the code to implement external forces into the existing algorithm, and the data in this publication was generated with the updated version of the code. In this publication, I performed simulations and data analysis. All the authors shared interpretation of the results. K. D. Dorfman and I wrote most of the paper. C. M. Schroeder contributed to the discussion and the writing of introduction.

Chapter 5 is based on X. Li and K. D. Dorfman, "Effect of excluded volume on the force-extension of wormlike chains in slit confinement," *J. Chem. Phys.*, vol. 144, p. 104902, 2016. This publication contains results from the aforementioned PERM code to study the configuration of real semiflexible chains that are stretched and confined simultaneously. In this publication, I modified the PERM code to have both force potential and confinement at the same time, performed simulations, and analyzed the data. All the authors shared interpretation of the results and the writing of the paper.

Chapter 6 is based on X. Li, F. S. Bates, and K. D. Dorfman, "Rapid conformational fluctuations in a model of methylcellulose," *Phys. Rev. Mat.*, vol. 1, no. 2, p. 025604, 2017. This paper used Langevin Dynamics simulations to study the gelation mechanism of methylcellulose. I performed the simulation with LAMMPS package [3] (ver. Jun 2014) and analyzed the data. K. D. Dorfman and I interpreted the data and wrote most of the paper. F. S. Bates contributed to the discussion and the writing of the

introduction.

Chapter 1

Introduction and Motivation

1.1 Genome Mapping

Deoxyribonucleic acid (DNA) carries the genetic instructions that are crucial in the growth, development, functioning, and reproduction of living organisms [4, 5], and thus attracts extraordinary interests in sequencing and deciphering the code embedded in DNA. Dating back to the early 1970s, the first DNA sequences were obtained by academic researchers using two-dimensional chromatography [6–9]. From then on, knowledge of DNA sequences has become indispensable in all fields of biological research. The first generation of sequencing technology is Sanger chain termination method [10] developed in 1977. With this achievement, researchers were able to sequence DNA in a reliable and reproducible manner, although it was expensive and inefficient at that time.

In 2005, an innovative approach, which is called "next-generation sequencing" (NGS) technique, has revolutionized the genomic science. This method reads an enormous amount of short segments (~500 base pairs, or bp) through automation and massive parallelization, which lead to the ability to sequence the entire genome of an organism in a much quicker and cheaper way [12–15]. As shown in Fig. 1.1, the NGS steps consist of library preparation, amplification, sequencing, and alignment [11]. The final result is a consensus sequence from the ensemble of short reads [16]. The initial sequencing and mapping of the human genome were estimated to have cost about 3 billion US dollars [12, 17], but the platforms are improving at the rate of Moore's law since 2008.

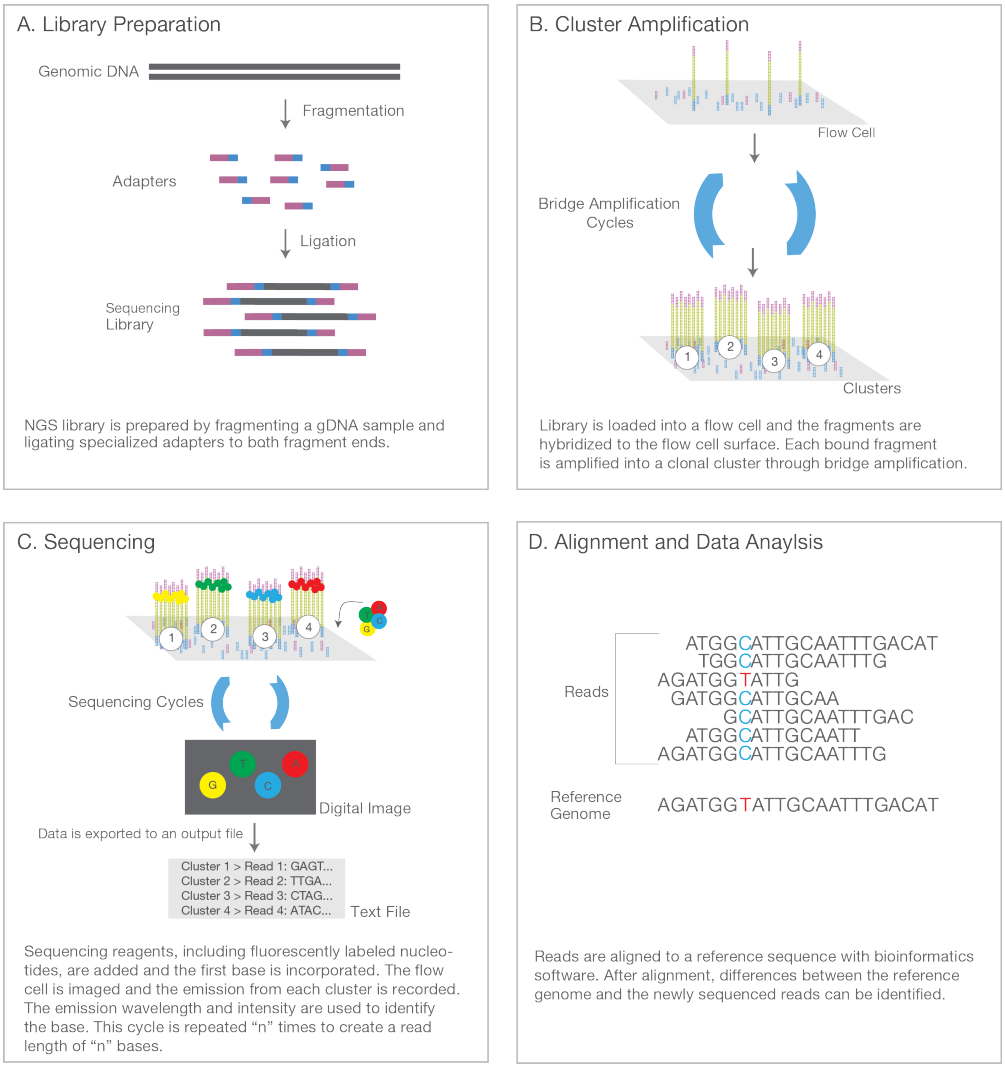

Figure 1.1: Schematic of a next-generation sequencing technique. This method includes four steps: (A) library preparation, (B) cluster amplification, (C) sequencing, and (D) alignment and data analysis. Reproduced from [11].

However, since NGS technologies were originally designed to analyze genes instead of genomes, there are many inherent limitations. First of all, current sequencing technologies rely on short read lengths, which prevent them from extracting the full spectrum of information associated with the genome [19–21]. Such short read lengths are orders of magnitude smaller compared to the size of a human genome, which contains about 3

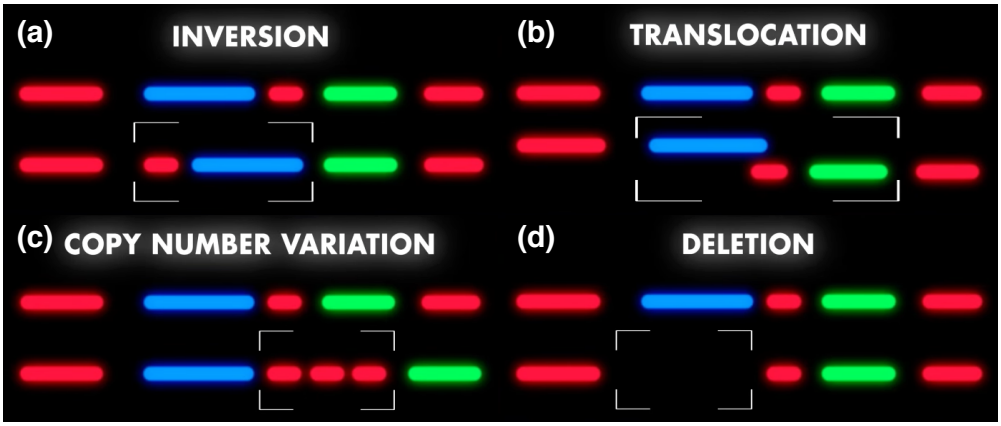

Figure 1.2: Various structural variations can happen on the genome. Different colors represent different segments of genome. Reproduced from [18].

billion base pairs. Nevertheless, there are a lot of studies suggesting that the structural variations in a large genomic distance are crucial to the genomic functions of living organisms [19, 22, 23]. Such structural variations include but are not limited to inversion, translocation, copy number variation and deletion of sequences as shown in Fig. 1.2. Moreover, these short segments come from ensembles of DNA molecules originating from many different cells, thus are not capable of detecting the structural variation in a large genomic distance of individuals within a species, or even the structural variation of cells within the same organism (e.g. cancer cells). Amplification adds another layer of the limitation of NGS in that it duplicates different segments with a different amount, thus put a bias on the "statistical consensus" of the final merged sequence. Beyond the issues described above, it is not at all obvious that we need to read every base pair to make a useful diagnosis on the genomic level [24]. As a result, we seek for a *de novo* method that is able to detect structural variation in a large genomic distance without reading every single base pair.

Meanwhile, there is a growing interest in platforms that can readout genomic information directly from single long DNA molecules, with the ultimate goal of single-cell, single-genome analysis. The rewards are tempting: we will be able to improve read length, detect structural variation directly, and circumvent amplification bias [20,21,25]. A promising class of approach is the genome mapping technique, which measure the genomic distances of large pieces of DNA molecules ranging from several kilobase pairs

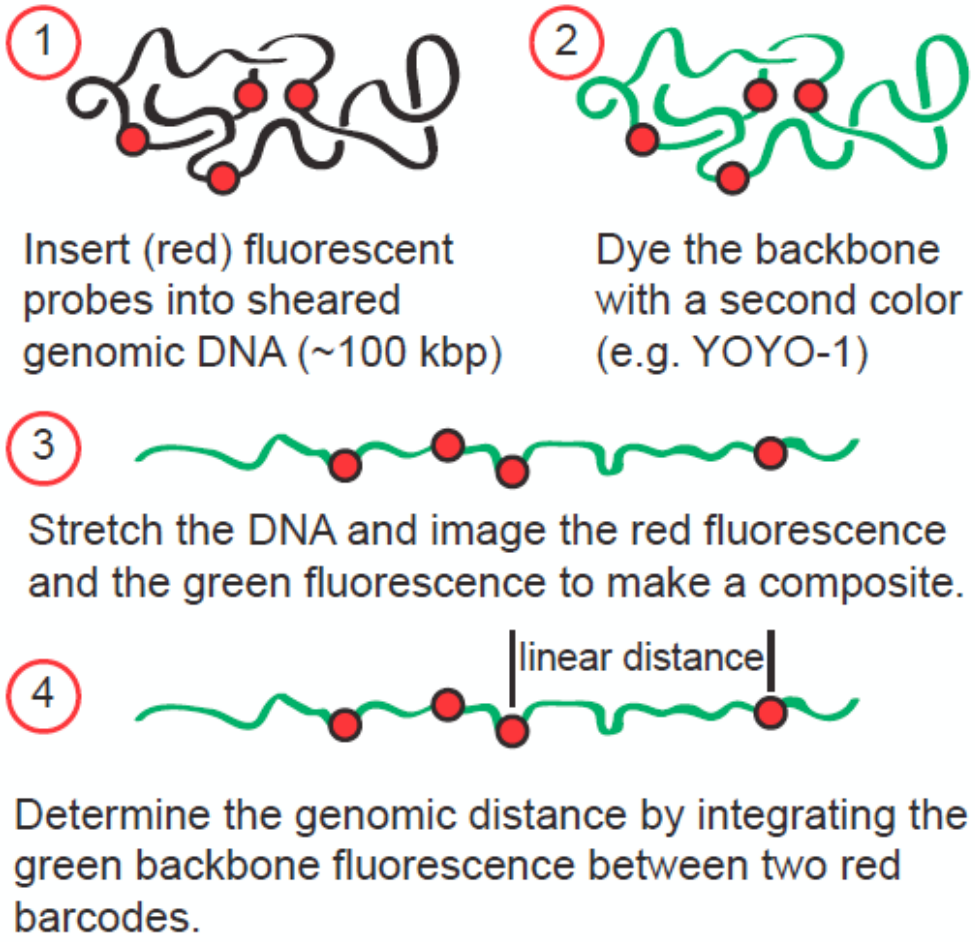

Figure 1.3: Schematic of a typical genomic mapping process. Image from Kevin D. Dorfman.

up to one megabase pairs without sequence details.

Figure 1.3 shows a schematic of the genomic mapping process. The first step is to insert fluorescent probes on the sequence-specific spots on DNA. Then we dye the whole backbone evenly with a second fluorescent color, stretch the DNA molecule into a linear configuration, and image the elongated molecule with a fluorescent microscope [21, 26–29]. The genomic distances between the red sequence-specific markers are determined by the green fluorescent intensity. Repeating the measurement with lots of overlapping

long DNA fragments and assembling them together eventually gives a genome map. This genome map can act as a template for assessing structural variation. For example, consider a species that has 1% of individuals that have two copies of gene X of a typical length of 5 kbp, while the remainder has only one copy of gene X. By performing genomic mapping measurement many times, one can obtain a genome map suggesting 99% of the genomes measured contain one copy of gene X and 1% of the genomes contain two copies. On the contrary, NGS is not able to detect easily the copy number variation because of the short read length.

A key step in the genomic mapping is to linearize a long piece of DNA. Through linearization, DNA reveals ordered information and reduced fluctuation, which then permits high-resolution imaging and fluorescent intensity analysis. In fact, through linearization, the resolution of mapping can be reduced from ~10 Mbp for DNA in metaphase chromosomes to ~1 kbp, leaving the domain of individual genes within reach [30].

Linearization technically challenging in terms of DNA manipulation at the single molecule level. However, thanks to the development of nanofabrication technology, nanomaterials, and super-resolution optical imaging, there are several ways to accomplish this task. The first class of methods is to stretch DNA with external forces. Fluid flow, for example, can be applied to stretch DNA molecule either on a specially treated surface [31, 32] or through extensional flow [33]. The former, known as "molecular combing", exploits surface forces to simultaneously stretch and assemble millions of DNA molecules on the surface of a vinyl silane treated glass carrier. In the original version of this process, a glass slide is dipped into the purified DNA solution and the DNA molecules attach to the glass surface by one end. As the slide withdrawn, the DNA molecules are stretched over the surface. This technology has been commercialized in company Genomic Vision and company OpGen. Another approach stretches DNA molecules through elongational flow [33, 34]. The DNA molecules were first labeled and stretched in a microfluidic device, and then driven through fluorescence detectors. This kind of method is known as "Direct Linear Analysis" (DLA).

An alternative way to linearize DNA is to impose confinement. The nanoconfinement is defined as a hollow geometry with the confining dimension typically in dozens to hundreds of nanometers, which is much smaller than the size of genomic DNA molecules

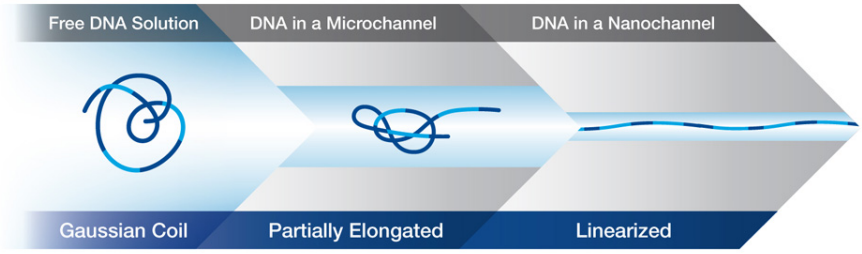

Figure 1.4: A long genomic DNA in free solution can be linearized in confinement. Reproduced from [35].

in free solution. The geometry could be, but not restricted to nanoslit, rectangle or square nanochannel, nanotube, triangular nanochannel, or even spatial gradient structures. Several studies have been devoted to investigating those confinement effects on DNA [36–39]. The various techniques and evolution are comprehensively reviewed in [24]. Cao *et al.* are pioneers in experimenting confined DNA [40]. They put single DNA molecules in a continuous spatial gradient nanoconfinement which smoothly narrows the cross section from the micron to nanometer length scale, thus traps and elongates DNA. This technology has later been commercialized by the company Bio-Nano Genomics, although a different rectangle channel is used. The difficulty of genomic mapping with nanoconfinement lies in (*i*) manufacturing a small geometry with long length and (*ii*) the compromise of DNA loading efficiency into very small nanofeatures. Both of the methods, i.e., the external force and the nanoconfinement, alter the DNA configuration profoundly in nanoscale, thus requires deep understanding in the physics of DNA and how it interacts with the environment.

1.2 Methylcellulose Gelation

My thesis also involves another biopolymer, methylcellulose. Methylcellulose is particularly important with an extremely wide range of commercial uses, such as a thickener in pharmaceuticals, cosmetics and food products [41, 42]. Methylcellulose is a cellulose-based polymer with hydroxyl groups (-OH) partially replaced by methoxy moieties (-CH_3) at the C-2, C-3 and/or C-6 positions. The cellulose backbone, together with the intramolecular hydrogen bonds, gives methylcellulose a linear and rigid structure.

Since there is an irregular distribution of substituents along the cellulosic backbone, methylcellulose always appears as a heterogeneous polymer and the degree of substitution (DS) defines the average number of methyl groups per anhydroglucose unit. The DS ranges from zero for unsubstituted cellulose to 3 for the fully substituted polymer and characterizes aqueous solubility of the polymer. For example, a DS of 1.64 to 1.92 yields maximum water solubility, while a lower DS leads to methylcelluloses that are only soluble in caustic solution due to intramolecular hydrogen bonding, and a higher DS produces methylcelluloses that are soluble only in organic solvents because of the hydrophobic methyl groups [43].

Methylcellulose is also of considerable scientific interest that has been studied both experimentally and theoretically, especially for its gelation mechanism. As shown in Fig. 1.5, unlike other water-soluble materials such as inorganic salts, most methylcellulose is readily soluble in water at low temperature, while thermoreversible gelation occurs at elevated temperature accompanied by an increase in optical turbidity [45–47].

Owing to its gelation properties, methylcellulose is investigated extensively to develop materials with innovative physico-chemical properties. Over the past decades, this lower critical solution temperature (LCST) phase behavior in aqueous systems [48–56] has been studied extensively [44–46,48–55,57–61]. However, there is no consensus on the relationship between the LCST phase behavior and the gelation process. Many have attributed this behavior to viscoelastic liquid-liquid phase separation [45,46,48–55,57–60], where the polymer-rich and polymer-deficient phases are kinetically trapped. In an effort to address this phase-separation hypothesis, some studies have attempted to build a temperature-composition phase diagram [44, 48, 50]. However, unfortunately, there is no conclusive evidence to clearly show the coexistence and stability curves for the phase diagram, which are crucial for the liquid-liquid phase separation behavior.

Recently, several experimental studies on the structure and physical properties of aqueous methylcellulose solutions provided compelling evidence for a heterogeneous fibrillar morphology [45–48,52,62,63], a structural motif common to semiflexible polymers found in living organisms [64–66]. This fibrillar morphology and the associated viscoelastic response of these solutions are fundamentally different from what would be expected from a morphology based solely on liquid-liquid phase separation. Understanding the structure of the fibrils as well as their mechanical properties could guide

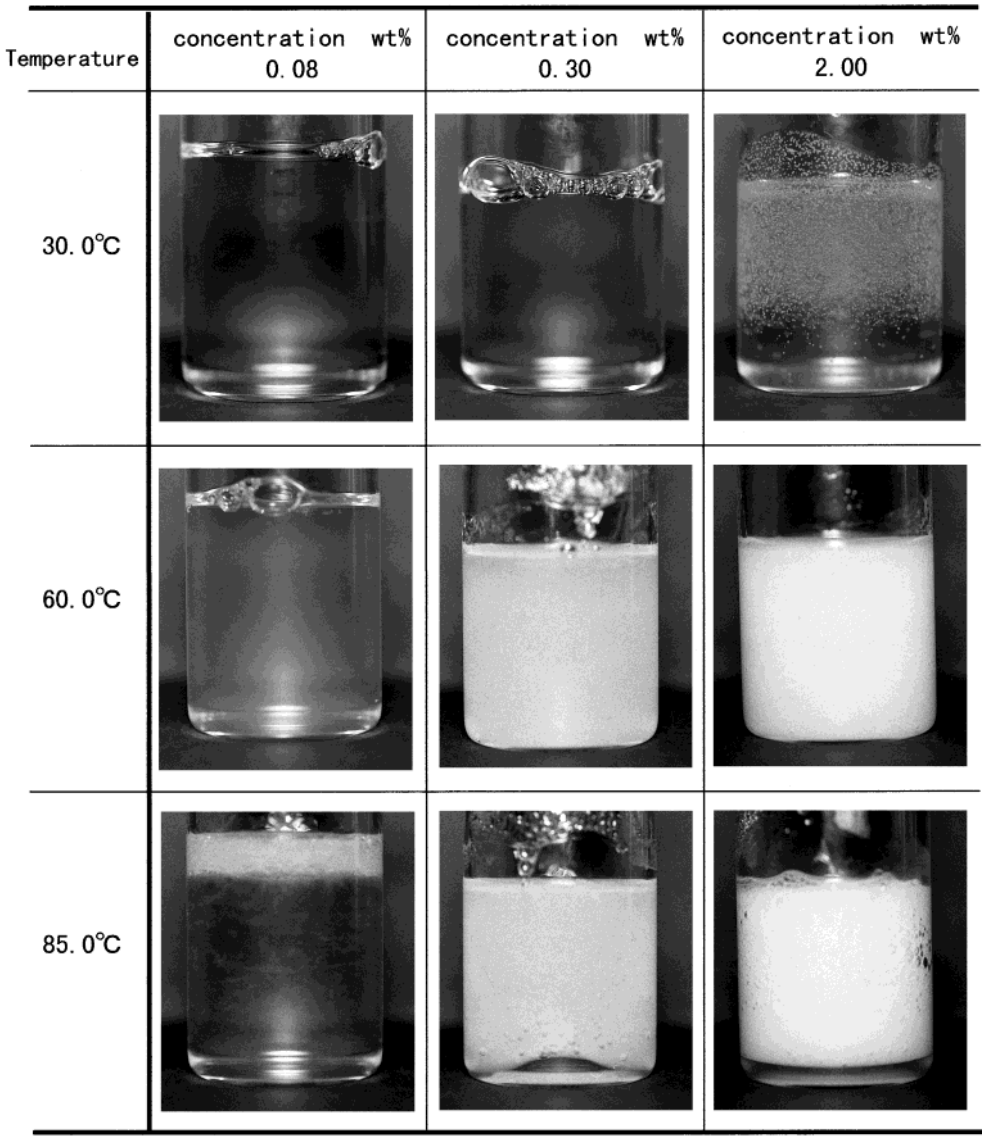

Figure 1.5: Methylcellulose dissolved in water at various concentration 0.08, 0.30, and 2.00 wt% kept at 30 °C, 60 °C, and 85 °C for about 30 min. Reproduced from [44].

efforts in the design of new and useful materials. Lott *et al.* [45, 46] quantified the fibrillar structure of aqueous methylcellulose gels with a combination of (real space) cryogenic transmission electron microscopy (cryo-TEM, Fig. 1.6) and (reciprocal space) small-angle neutron scattering (SANS) techniques. Fibrillar dimensions of 14 ± 1 nm

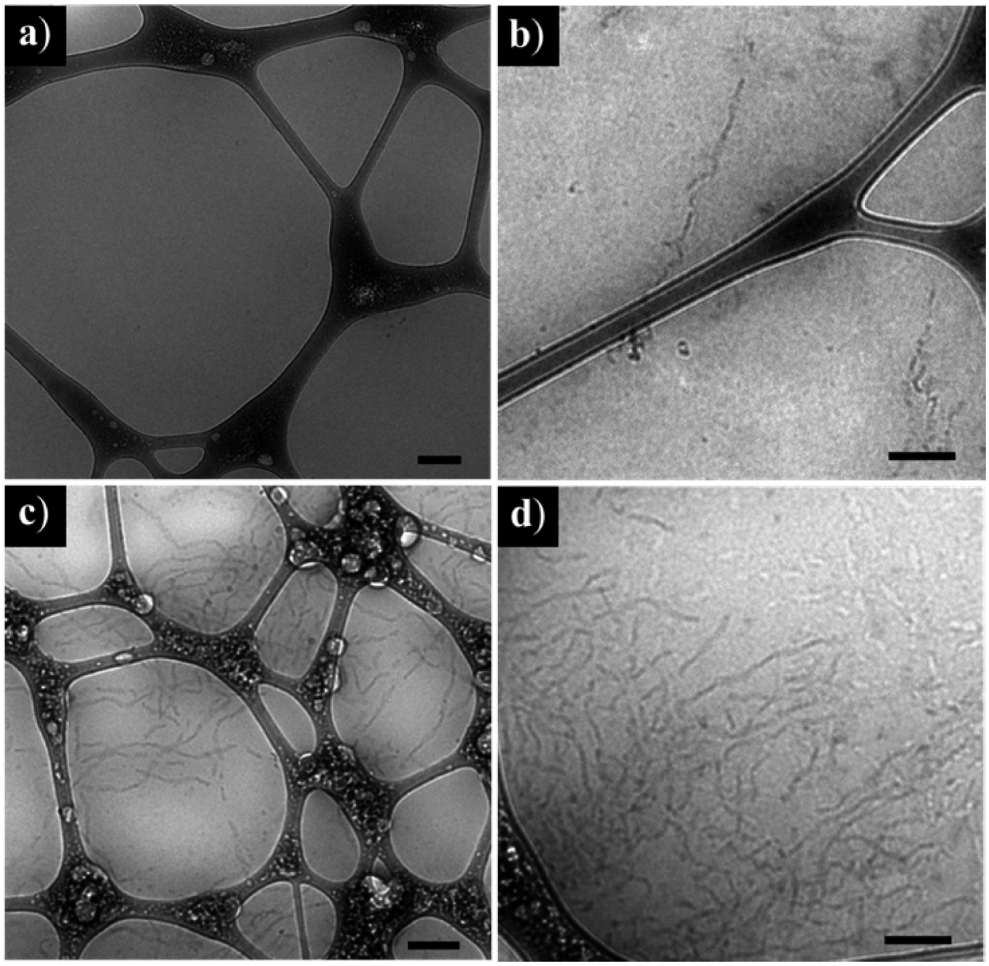

Figure 1.6: Cryo-TEM images of the fibrillar morphology of methylcellulose 0.20 wt % aqueous solutions of 300,000 g/mol at (a) 50 °C, (b) 55 °C, (c) 60 °C, and (d) 65 °C. All scale bars are 200 nm. Reproduced from [45].

were obtained by fitting the SANS data with a model based on the form factor for flexible cylinders with a polydisperse radius. Surprisingly, this diameter is independent of mass fraction and molecular weight M_w within the range of 0.01% to 3.79% and 49 to 530 kg/mol, respectively. McAllister *et al.* [62] further investigated the thermodynamics of aqueous methylcellulose solutions. They found that the growth of the fibrils cannot be explained solely using classical Flory-Huggins solution theory [67,68]. They proposed

that methylcellulose fibrils reflect a kinetically trapped state that balances thermodynamic equilibrium and local orientational order. While the gelation mechanism is still an open question, these experimental studies provided an unambiguous picture of the fibrillar morphology, although the detailed molecular arrangement within the fibrils has not been established experimentally.

1.3 Research Outline

Biopolymers have many unique properties and play an essential and pervasive role in everyday life, thus making them attractive for both scientific interest and engineering applications. In the genomic mapping technology, it is important to understand how to make a reliable prediction of the DNA extension and achieve the reduction of fluctuation when applying external forces and nanoconfinement. Despite the advancement of genomic mapping experiments, there is a lack of theoretical understanding and numerical evidence in the physics of DNA and its interaction with the environment at the nanoscale. In fact, the DNA configuration is affected by a lot of factors, such as electrostatics, hydrodynamics, the confinement geometry and external forces. In addition, the double-helix structure of DNA further complicates the physics. DNA is a stiff polymer with persistence length (characteristic length of making a 180° turn) of 50 nm, which is of the same length scale of most nanofluidic devices. Due to the complexity of the parameter space, it is impossible for theorists to propose a thorough analytic solution, nor for experimentalists to fully explore the possible designs by trial-and-error.

The same situation applies towards understanding the gelation mechanism of methylcellulose. Although recent experimental studies [46] on the gel structure revealed the unambiguous fibrillar morphology, the detailed gelation mechanism remains poorly understood, in particular at higher polymer concentrations. There are several computational efforts [1, 69] that provide plausible models for fibril formation in dilute solution. However, they leave questions remained to be answered, especially how to remove the mis-collapsed states and what are the important physical properties that determine the configuration of the methylcellulose.

Although DNA and methylcellulose, and other biopolymers not mentioned can have different chemistries, they share some common physical properties: high molecular

weights, stiff backbones, and complex internal structure. Computer simulation, therefore, plays quite an important role since it provides a way to study a generic model that, by changing the parameters appearing in the model, permits studying a huge variety of biopolymers. In addition, simulation also provides opportunities to explore conditions which can be challenging for experiments but crucial for understanding the basic properties of biopolymers.

As such, our research focuses on *applying computational tools to biopolymer problems* with direct applications. In particular, we are interested in stretched DNA for genomic mapping, and the gelation mechanism of methylcellulose for consumer products. These two projects involve the application of novel simulation techniques to biopolymer models and allow building connections between theoretical predictions and experimental studies. In an effort to address the genomic mapping, we are interested in understanding how an external force and confinement alter the already complex DNA properties. This allows us to simultaneously investigate the basic physical properties of biopolymers while advancing this promising genomic technology. Such a general understanding of biopolymer and the development of computational tools can also be applied to study the mechanism of methylcellulose gelation, especially the spontaneous formation of high-aspect-ratio fibrils, and the escaping from local energy traps due to the intramolecular interactions. We anticipate that the knowledge gained from this study will provide insights into the discrepancy between the high polymer concentrations, where fibrils form in experiments, and the low polymer concentration required in simulations in order to get ring structures. Moreover, the development of a generic biopolymer model, along with advanced simulation techniques, would provide a powerful tool to explore the physical properties of biopolymers in general.

In line with the above goal, **Chapter 2** summarized the theories, experiments, and simulations surrounding the physics of DNA and methylcellulose. We start with describing the discrete wormlike chain model (DWLC), which is a good representation for most of the biopolymers including DNA. We then review the physical properties of DNA in various conditions, such as in confinement or in free solution, with or without external stretching forces. We also discuss the simulation progresses surrounding the gelation mechanism of methylcellulose.

Chapter 3 reviews the simulation methodology we used to simulate DNA as a semiflexible chain in various conditions. Specifically, we motivate the introduction of pruned-enriched Rosenbluth method (PERM), which is an advanced chain growth Monte Carlo method of choice, by first introducing simple sampling and Rosenbluth sampling and highlighting their inefficiency in simulating a long piece of DNA with excluded volume effects. We then proceed to introduce an off-lattice version of PERM and how could it be applied to study the DWLC of interest. The parameterization and simulation methodology of methylcellulose is introduced in Chapter 6.

Stretching a polymer chain in the presence of an external force is a classic problem in polymer physics. In general, entropic and enthalpic intramolecular interactions must be considered for an accurate description of polymer elasticity. In **Chapter 4**, we propose an interpolation formula (the EV-WLC relation) for the force-extension behavior of wormlike chains in the presence of hard-core excluded volume interactions, analogous to the classic interpolation formula from Marko and Siggia for ideal wormlike chains [70]. Using PERM simulations of asymptotically long, discrete wormlike chains in an external force, we show that the error in the EV-WLC interpolation formula to describe discrete wormlike chains is systematically smaller than the error in the Marko-Siggia interpolation formula, except for the saturation region in which both formulas have the same limiting behavior. We anticipate that the EV-WLC interpolation formula will prove useful in the coarse-graining of wormlike chain models for dynamic simulations. Related results for the excess free energy due to excluded volume provide strong support for the physical basis of the Pincus regime.

With the knowledge gained from stretching a real semiflexible chain in free solution, **Chapter 5** focused on stretching a real semiflexible chain in nanoslit confinement. We again use PERM simulations to develop a quantitative phase diagram for the stretching of a real wormlike chain confined in a slit. Our simulations confirm the existence of a "confined Pincus" regime in slit confinement, analogous to the Pincus regime in free solution, where excluded volume effects are sensible. The lower bound for the confined Pincus regime in the force-molecular weight plane, as well as the scaling of the extension with force and slit size, agree with an existing scaling theory for this regime. The upper bound of the confined Pincus regime depends on the strength of the confinement. For strong confinement, the confined Pincus regime ends when the contour length in the

Pincus blob is too short to have intrablob excluded volume. As a result, the chain statistics become ideal and the confined Pincus regime at low forces is connected directly to ideal chain stretching at large forces. In contrast, for weak confinement, the confined Pincus regime ends when the Pincus blobs no longer fit inside the slit, even though there is sufficient contour length to have excluded volume inside the Pincus blob. As a result, weak confinement leads to a free-solution Pincus regime intervening between the confined Pincus regime for weak forces and ideal chain stretching at strong forces. Our results highlight shortcomings in existing models for the stretching of wormlike chains in slits. Excitingly, after publishing this result, our theoretical prediction was confirmed by experimental evidence [71]. We anticipate that such a complete description of the force-extension of real wormlike chains will prove useful for both the interpretation of experimental data and the design of new technologies.

In light of an understanding of the generic semiflexible model, in **Chapter 6**, we switch to studying methylcellulose gelation mechanism. We use Langevin dynamics simulations to investigate a coarse-grained model for methylcellulose that produces collapsed ring-like structures in the dilute solution with a radius close to the fibrils observed in experiments. We show that the competition between the dihedral potential and self-attraction causes these collapsed states to undergo a rapid conformational change, which helps the chain to avoid kinetic traps by permitting a transition between collapsed states. If the dihedral potential is removed, the chains do not escape from their collapsed configuration, whereas at high dihedral potentials, the chains cannot stabilize the collapsed state. We provide systematic data on the effect of the dihedral potential in a model of methylcellulose and discuss the implication of these previously overlooked, rapid conformational fluctuations on the spontaneous formation of high-aspect-ratio fibrils.

Finally, **Chapter 7** summarizes the important finding of this dissertation and proposed possible future research directions.

Chapter 2

Background

This chapter provides a brief background and reviews recent literature concerning the physics of DNA and methylcellulose as semiflexible chains, which are necessary to understand the material presented in the following chapters. Section 2.1 begin with describing the definition of wormlike chain (WLC), which is a coarse-grained representation of DNA and other semiflexible polymers. In this section, we also introduce the discrete wormlike chain (DWLC) model which facilitates the simulation process of DNA along with the force field implementation. Section 2.2 and Section 2.3 review recent studies concerning the physics of DNA and other semiflexible polymers in both free solutions and nanoslit confinement. In particular, we will discuss the static properties of DNA in the presence of an external force, for the entropic and enthalpic intramolecular interactions must be considered for an accurate description of polymer elasticity. For the remainder of this chapter, Section 2.4 reviews some important theoretical and simulation studies for methylcellulose, especially the different configurations at low temperature and high temperature, and its gelation mechanism. We begin by reviewing an atomistic methylcellulose model and highlight some of its important properties in elevated temperature. In particular, we are interested in obtaining the coarse-grained model parameters from an all-atom simulation. We then review coarse-grained simulation studies with a focus on the gelation mechanism that drives the once dissolved chains into fibril structure under high temperature.

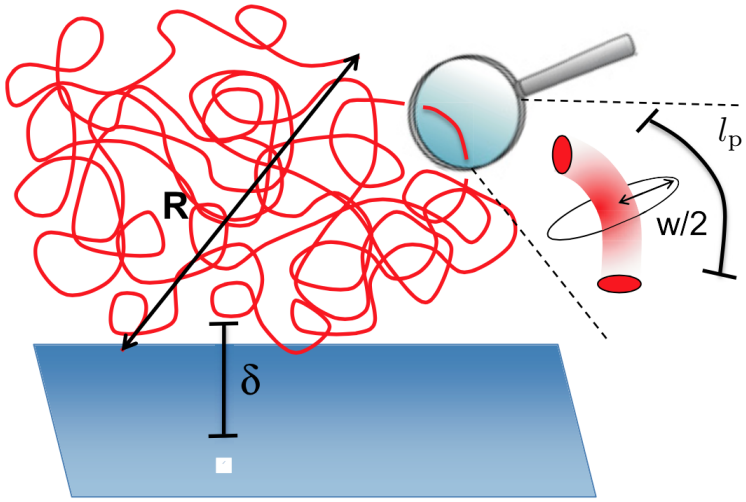

Figure 2.1: Schematic of DNA as a wormlike chain with the end-to-end length (R), persistence length (l_p), effective width (w) and wall-DNA depletion width (δ). Reproduced from [74].

2.1 Discrete Wormlike Chain Model for DNA Simulation

It has been shown that DNA behaves like a semi-flexible polymer [70, 72, 73] with a local bending stiffness. This implies that shorter DNA molecules are rod-like, yet longer segments behave as flexible coils and thus can be described with wormlike chain models.

Figure 2.1 shows a typical DNA conformation near a flat surface. In genomic mapping, the surface can be the confinement imposed by nanochannel or nanoslit. There are several key physical parameters characterizing a single DNA chain. The contour length L denotes the length of the chain when fully extended. The end-to-end distance R is the length between the two ends of the chain. The persistence length l_p characterizes the length of the flexibility of the chain beyond which DNA begins to lose directional correlation and forms random coil-like structure. In addition, DNA exhibits self-avoiding features between distant segments of the chain characterized by an effective hard-core width w. The depletion zone δ characterizes the exclusion region from the wall. δ reflects the entropic contribution of flexibility and repulsive electrostatic interactions between DNA and the surface.

In this section, our goal is to describe a statistical mechanical model that is capable

of describing a large piece of DNA with the length matching genomic mapping scale, meanwhile capturing the sub-persistence length behavior, in other words, the "worm-like" chain behavior. On the one hand, an all-atom molecular model is able to capture the chemical details and physical properties of DNA. However, from a computational efficiency perspective, an all-atom model is only capable of simulating short DNA with a few hundreds of base pairs because of the complicated chemical structure. This short chain length is way below the typical 10^5 base pairs commonly appearing in genomic mapping applications. On the other hand, a model of DNA with a stiff rod representing a Kuhn length $b \approx 2l_{\mathrm{p}}$ and rotating freely without resistance expedites calculation and is able to simulate long DNA chains. Nonetheless, this over-coarse-grained model sacrifices the local-bending stiffness which characterizes the "wormlike" behavior.

The popular wormlike chain (WLC) [75–77] model is a good balance for coarse-grained DNA with atoms aggregated in some way, while still capturing the sub-persistence length behavior. Commonly known as the Kratky-Porod model, this mode envisions an isotropic rod that is continuously flexible. For a polymer of contour length L, we define s to be the path along the polymer as $s \in (0, L)$, $\vec{r}(s)$ to be the position vector along the chain, $\hat{t}(s)$ to be the unit tangent vector at s. We then have

$$\hat{t}(s) \equiv \frac{\partial \vec{r}(s)}{\partial s} \tag{2.1}$$

and the end-to-end distance vector $\vec{R}$ is

$$\vec{R} = \int_0^L \hat{t}(s) ds. \tag{2.2}$$

Orientation change of the unit tangent vector $\hat{t}(s)$ characterizes the local stiffness of the wormlike chain, as the more reluctant the chain is to change the orientation, the stiffer the chain. The orientation correlation function decays exponentially along s:

$$\langle \hat{t}(s) \cdot \hat{t}(0) \rangle = \langle \cos\theta(s) \rangle = e^{-s/l_{\mathrm{p}}}. \tag{2.3}$$

After we have defined the local curvature factor as $\partial \hat{t}(s)/\partial s$, the bending energy of

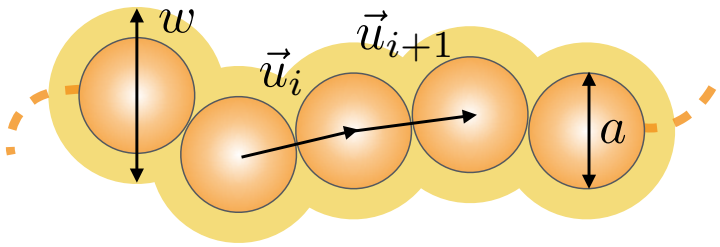

Figure 2.2: Schematic of the discrete wormlike chain (DWLC) model. This model consists of a series touching beads with diameter a, which is also the bond length. In addition, there is a bending potential between two nearest beads characterized by the relative orientation between two neighboring unit bond vectors $\vec{u}_i$ and $\vec{u}_{i+1}$. The effective width w defines the region for the excluded volume repulsion, as shown in the yellow-shaded area. The dashed line represents the rest of the chain.

a chain is

$$U_{\text{bend}} = \frac{1}{2}\epsilon \int_0^L \left(\frac{\partial \hat{t}(s)}{\partial s} \right)^2 ds \tag{2.4}$$

with ϵ being a bending constant that characterizes the local stiffness of the chain.

Although this WLC model captures the bending stiffness of the polymer chain in a coarse-grained level, there are two more things that prevent this model to be directly applied to DNA simulation, and thus requires modification. First of all, Eq. 2.4 calculates bending energy in a continuum approach, while in mesoscopic simulations, discretization is required to keep track of mass positions. Moreover, this model only captures the ideal behavior of DNA and does not consider the thickness of the polymer chain. However, DNA and other polymers are often long enough to exhibit excluded volume behavior in good solvent through steric and electrostatic interactions. In such a way, a discretized model that includes both bending stiffness and excluded volume is necessary.

The discrete WLC model (DWLC) [78–80] we use here is a coarse-grained polymer model that is able to zoom in a length scale under l_{p}, and also exhibits repulsive interactions between monomer units. As shown in Fig. 2.2, the DWLC model is constructed with a series of $N+1$ touching beads and inextensible bonds of length a. The bending

energy in Eq. 2.4 can be discretized as

$$U_{\text{bend}} = \frac{2\epsilon}{a} \sum_{i=1}^{N} (1 - \vec{u}_i \cdot \vec{u}_{i+1}). \tag{2.5}$$

$\vec{u}_i$ and $\vec{u}_{i+1}$ are two consecutive unit bond vectors. We can further define θ_i as the angle between $\vec{u}_i$ and $\vec{u}_{i+1}$, and define a dimensionless bending constant $\kappa = 2\beta\epsilon/a$ with β being the inverse thermal energy $(k_{\text{B}}T)^{-1}$. Then Eq. 2.5 becomes

$$\beta U_{\text{bend}} = \kappa \sum_{i=1}^{N} (1 - \cos\theta_i). \tag{2.6}$$

It is important to relate the persistence length l_{p}, which is the length over which correlations in the direction of the tangent are lost, to the dimensionless bending constant κ. With the equilibrium probability density function for $\cos\theta_i$ [81], one can obtain a relationship between κ and l_{p} as

$$\frac{l_{\text{p}}}{a} = \frac{\kappa}{\kappa + 1 - \kappa \coth\kappa}. \tag{2.7}$$

To add excluded volume, $N + 1$ spherical beads are placed at the bond joints, and the hard-core repulsive potential for DNA is introduced between non-contiguous beads. The range of the hard-core repulsion is defined by the effective width w. U_{EV} is the hard-core excluded volume energy defined as,

$$\beta U_{\text{EV}}(\mathbf{r}_{ij}) = \begin{cases} \infty, & |r_{ij}| \le w \\ 0, & |r_{ij}| > w, \end{cases} \tag{2.8}$$

where $|r_{ij}|$ is the positive distance between bead centers at i and j. Note that we commonly have $w \ge a$, since unphysical chain crossing can occur when $w < a$. The total potential energy U_{total} of a configuration can be expressed as

$$U_{\text{total}} = U_{\text{bend}} + U_{\text{EV}}. \tag{2.9}$$

Note that other potentials could be added to U_{total} to represent DNA under various condition, such as an external force field which we will discuss later.

Finally, although we are able to study long DNA by developing a coarse-grained

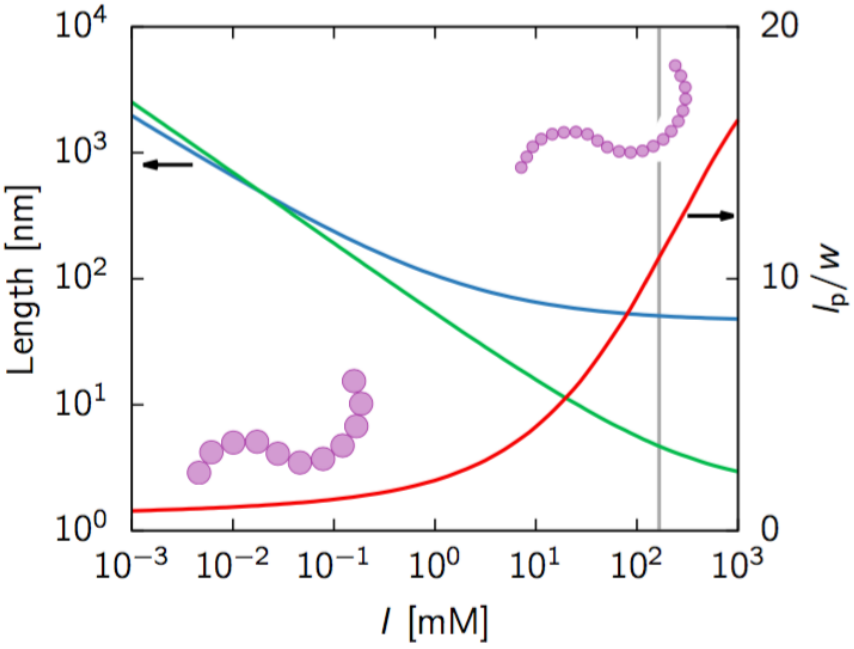

Figure 2.3: Ionic strength dependece of the persistence length (blue line), effective width (green line), and monomer anistropy l_p/w (red line) for DNA. The vertical gray line represents the ionic strength of 165 mM (5 × TBE) [82]. The schematic shows two DWLC models under different ionic strength. Modified from [83].

DWLC model with only two parameters l_p and w, we omit electrostatic details which could be rather important for shorter DNA chains less than 100 basepairs. It is noteworthy that many of the DNA properties have an ionic strength dependence, thus a neutral wormlike chain model for DNA oversimplifies the electrostatics and might not be a good approximation to softer electrostatic potentials, especially at low ionic strength. DNA is a polyelectrolyte with nominally 2.0 and effectively as low as 0.5 negative charges per basepair [84]. The electrostatic interactions between negatively charged phosphate groups on the backbone are screened by ions in solution. By decreasing the ionic strength, or the salt concentration in the buffer, the screening effect between negatively charged phosphate groups reduces. As a result, the value of DNA properties such as w and l_p increase. Furthermore, the repulsive interaction between the polyelectrolyte and the chagred bounding surfaces increases. Materials such as poly-dimethylsilozane

(PDMS) and untreated SiO_2 that are commonly used for creating nanostructured devices, acquire a negative surface charge when immersed in an aqueous solution with physiological pH [85, 86]. This repulsive electrostatic interaction renders the effective confinement size a function of ionic strength. Figure 2.3 shows the dependence of l_p and w on the ionic strength plotted using the empirical relation from Dobrynin [82, 87] and Stigter's theory [88]. Note that both relations fail at low ionic strength because of the non-linearity of the Poisson-Boltman equation due to the reduced screening effects of electrostatic interactions. However, to simplify the model and accelarate the simulation process, we limit our scope to high ionic strength buffer where the electrostatic interactions are screened by the excessive ions in the buffer. This assumption allows us to use a neutral model with effective l_p and w that lumps electrostatic details. For the purpose of predicting configurations and thermodynamic properties in the length scales comparable to genomic length DNA, a neutral model with electrostatics implemented implicitly is sufficient, and would further facilitate the calculation process. Throughout this thesis and for the case of DNA, unless otherwise specificied, we use an ionic strength of 165 mM, which corresponds to 5 × TBE. This buffer condition gives $l_p = 53$ nm and $w = 4.6$ nm, as marked with a gray line in Fig. 2.3.

2.2 DNA as a Semiflexible Polymer in Free Solution

2.2.1 Static Properties

We first consider a wormlike chain in free solution without external stretching forces. One equilibrium property of interest is the end-to-end distance that is commonly used as a measure of the polymer size. Since the conformation of the unconfined molecule is rotationally invariant, $\langle R \rangle$ is zero. Thus, the mean square end-to-end distance $\langle R^2 \rangle$ is often used instead. For an ideal wormlike chain following Kratky-Porod model, $\langle R^2 \rangle$ is given by [89, 90],

$$\langle R^2 \rangle = 2l_p L - 2l_p^2 \left[1 - \exp\left(-\frac{L}{l_p} \right) \right]. \tag{2.10}$$

Eq. 2.10 is valid for ideal chain with all range of the chain length L. There are two limiting cases: for short chains with $L \ll l_p$, statistically it is difficult for the chain to change orientation significantly, so the wormlike chain is essentially rod-like with

$R \approx L$; for long chains with $L \gg l_\mathrm{p}$, the second term in Eq. 2.10 becomes less important, leading to $\langle R^2 \rangle \approx 2Ll_\mathrm{p}$. Note that in the second case we have $R \sim L^{0.5}$, which is the metric of a random walk in free solution. This is straightforward: in the long chain limit, when measuring the size of the entire chain instead of zooming into the details of local stiffness, each Kuhn length ($\approx 2l_\mathrm{p}$) can be viewed as an anisotropic rigid rod rotating invariantly.

Another static property characterizing the size of the chain is the radius of gyration R_g, which can be directly measured through experimental techniques such as X-ray and neutron scattering experiments. By definition, R_g is the root mean square distance of chain segments from their center of mass. After combining with ideal wormlike chain model, the R_g^2 can be expressed as [89, 91]

$$R_\mathrm{g}^2 = Ll_\mathrm{p}\left\{\frac{1}{3} - \frac{l_\mathrm{p}}{L} + 2\frac{l_\mathrm{p}^2}{L^2} - 2\frac{l_\mathrm{p}^3}{L^3}\left[1 - \exp\left(-\frac{L}{l_\mathrm{p}}\right)\right]\right\}. \tag{2.11}$$

It is noteworthy that the derivations of Eq. 2.10 and Eq. 2.11 are based on Kratky-Porod model, and thus are only valid for an ideal polymer chain. The ideal chain model assumes that polymer segments can overlap with each other as if the chain were a phantom chain. However, two segments cannot occupy the same space at the same time. Flory proposed a theory, commonly known as Flory theory, to account for the excluded volume repulsion by proposing the free energy of a chain as,

$$\beta F \sim \frac{R^2}{Ll_\mathrm{p}} + \frac{\mathcal{N}^2 v_\mathrm{ex}}{R^3} \tag{2.12}$$

where $\mathcal{N} \equiv L/l_\mathrm{p}$ is the number of persistence length of a chain and $v_\mathrm{ex} \equiv l_\mathrm{p}^2/w$ is the excluded volume of each persistence length. Equation 2.12 shows that the free energy of a wormlike chain in a good solvent consists of two competing terms: the first term is the entropy contribution that resists the stretching of the chain, while the second term favors separated segments. Minimizing the total free energy with respect to R at equilibrium gives

$$R \sim l_\mathrm{p}\left(\frac{w}{l_\mathrm{p}}\right)^{2\upsilon-1}\left(\frac{L}{l_\mathrm{p}}\right)^{\upsilon}. \tag{2.13}$$

The exponent $\upsilon = 3/5$ is famously known as Flory exponent. Although it has been

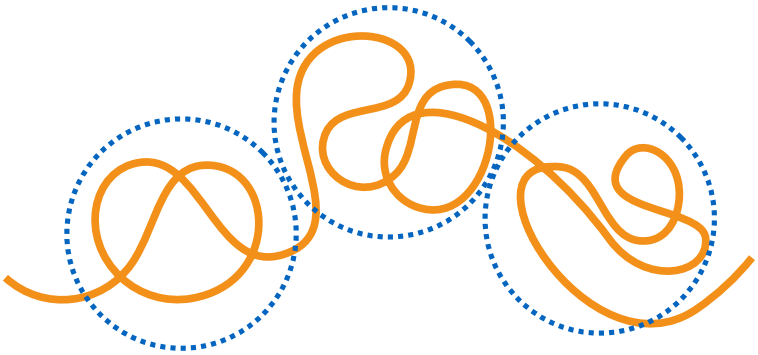

Figure 2.4: Schematic of thermal blobs in free solution. The orange curve represents a single wormlike chain, and the blue circles are the conceived thermal blobs. The polymer chain can be thought of as being composed of blobs within which the chain executes random walks.

proven [77, 92–94] that Eq. 2.12 is incorrect, Eq. 2.13 is still valid because of the cancellation of errors. For theories and numerical calculations of a more accurate Flory exponent of υ = 0.5876, interested readers should refer to more sophisticated methods [95, 96]. Note that in the so-called θ solvent where polymer chain behaves as if it were an ideal chain, υ is reduced to 1/2. Therefore, polymer with excluded volume interactions has a larger size and behaves like a swelled object.

The influence of excluded volume interaction increases as the length of the chain L increases, since the more segments the chain has, the more likely the segments would interact with each other. We can thus propose a blob that is associated with the balance between thermal energy and the excluded volume contribution. As shown in Fig. 2.4, a blob contains certain contour length l_{th} at which excluded volume free energy becomes order of $k_{\mathrm{B}}T$. Equating the second term in Eq. 2.12 to $k_{\mathrm{B}}T$ gives

$$l_{\mathrm{th}} \sim \frac{l_{\mathrm{p}}^3}{w^2}. \tag{2.14}$$

On the length scale below l_{th}, the polymer behaves like ideal chain whereas above l_{th}, excluded volume interaction prevails. Similarly, we can define blobs caused by other forces or effects, such as external forces of tension (Section 2.2.2) and confinement (Section 2.3.1) by equating the free energy contribution to $k_{\mathrm{B}}T$.

2.2.2 Stretching Semiflexible Polymers in Free Solution

At the length scale relevant to genomic mapping applications, we are interested in the overall mechanical properties of DNA and how it alters the conformations in various conditions, rather than understanding specific chemical bonds and detailed interactions. In this scope, the elastic properties of DNA play an important role in physical properties related to stretching, bending and twisting, and thus are essential in determining the overall conformation. Moreover, understanding the elastic picture opens up opportunities in other applications such as DNA binding to drugs, genetic regulation, and compact packaging of DNA in living cells.

A single wormlike chain adopts a random coil conformation which maximizes its entropy. There is enormous number of ways to arrange a chain in a coiled structure, but few possible ways for the chain to be near its full length. In other words, the stretched state is entropically unfavorable. The associated entropic force results from the reduction of the number of possible configurations and increases as a random coil pulled from the ends. The entropic force is rather weak, typically of the order of pN.

In order to work against entropy, an external force has to be applied. In the situation of force stretching, the DNA chain is subjected to a tensional uncoiling along the force direction and reaches to an equilibrium extension at the balance between the entropic force and the externally applied force. For simplicity, here we will discuss a uniform force field with a constant external force vector $\vec{f}$. In practice, this uniform field can be approximated by hydrodynamic flows [97] or by the electric field acting on a DNA molecule in an ionic solution [98,99]. The net external force relies on a small imbalance between the field acting on the DNA and the hydrodynamic resistance acting on the shear layer of the flow. In this section, we examine the consequences where the first bead is anchored at origin, and each bead experiences the same amount of constant external force. The total force potential of a chain with $N+1$ beads can be expressed as

$$U_{\text{force}} = \sum_{i=1}^{N+1} \vec{f} \cdot \vec{r}_i. \tag{2.15}$$

Figure 2.5 illustrates various regimes for a real wormlike chain under different stretching forces. In this context, we define another type of blob, the tensile blob, also known as Pincus blob [101], of the size $\xi = k_{\text{B}}T/f$. Similar to the thermal blob

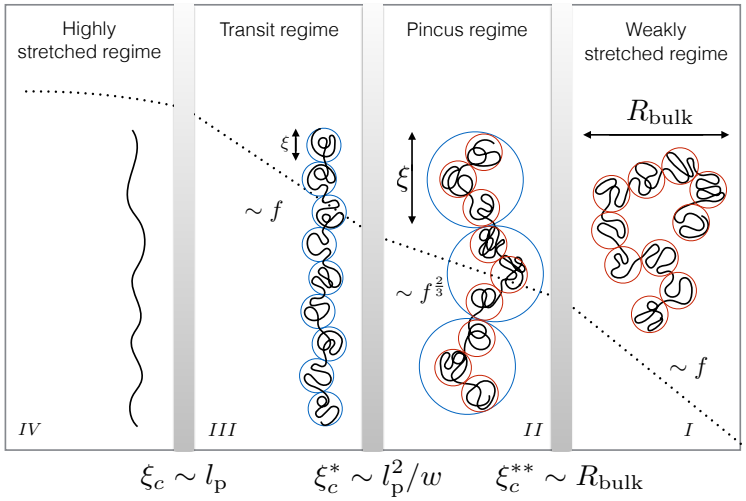

Figure 2.5: Schematic illustration and log-log scaling curve of wormlike chain under stretching force. $\xi = k_{\text{B}}T/f$ is the characteristic length of tensile blobs. Gray bars indicate transition bounds. The red circles denotes the thermal blobs, where the excluded volume interaction equals $k_{\text{B}}T$, and the blue circles denote the tensile blobs. The four regimes are separated by three characteristic lengths, ξ_c, ξ_c^*, and ξ_c^{**} corresponding to the persistence length l_{p}, the thermal blob size l_{p}^2/w, and the unperturbed size R_{bulk}. Generated following in the format in [100].

that features excluded volume, backfolding or overlapping of a pair of tensile blobs is at an energy cost of $k_{\text{B}}T$. In regime I as shown in Fig. 2.5, the force is extremely small so that the size of the tensile blob ξ is larger than the size of the polymer R_{bulk}. As a result, the force would not induce larger conformational change, but rather acts as a weak perturbation. In this regime, the ensemble average of extension $\langle X \rangle$ is along the force direction and is dependent on contour length. $\langle X \rangle$ slightly deviates from R_{bulk} by f in a Hookean way. The polymer chain escapes the weakly stretched regime by either increase the strength of the force or increase the length of the chain until

$$\xi_c^{**} = k_{\text{B}}T/f = R_{\text{bulk}}. \tag{2.16}$$

Outside the weakly stretched regime, we may assume blobs are more or less lined-up along the direction of deformation, thus $\langle X \rangle$ is linear with L thus eliminates any contour-length dependence properties. Pincus theory [101] shows the extension of the

stretched polymer should be determined by the competition between the tensile blob size with the Flory radius $R_{\mathrm{F}} \sim L^{3/5} l_{\mathrm{p}}^{1/5} w^{-1/3}$. By equating R_{F} to $k_{\mathrm{B}}T/f$ and having $\langle X \rangle = \mathcal{N}_f \xi$ with $\mathcal{N}_f$ being the number of tensile blobs, one has the force for real polymer scales non-linearly with extension as

$$\langle X \rangle / L \approx l_{\mathrm{p}}^{1/3} w^{1/3} (k_{\mathrm{B}}T/f)^{-2/3} \tag{2.17}$$

and the fluctuation in extension σ as

$$\frac{\sigma^2}{L l_{\mathrm{p}}} \approx \left(\frac{k_{\mathrm{B}}T}{f} \frac{w}{l_{\mathrm{p}}} \right)^{1/3} \tag{2.18}$$

More details on the derivation of Pincus scaling theory and the validating simulation results are available in Chapter 4. The Pincus regime ends where the tensile blobs shrink to be smaller than the thermal blob so that inside each of the tensile blob, there is not enough segment of polymer to interact with each other to experience excluded volume effects. The transition boundary between regime *II* and regime *III* can be determined by equating tensile blob size to thermal blob size

$$\xi_{\mathrm{c}}{}^{*} = l_{\mathrm{p}}^2 / w. \tag{2.19}$$

For regime *III* and *IV* in Fig.2.5, the theories are based on Kratky-Porod model [102] without any excluded-volume effects taken into account. In this case, the blobs are separated purely by tensile forces and the size of a tensile blob is smaller than a thermal blob. In the continuum wormlike chain model, by adding the force potential term to Eq. 2.4, we have

$$U_{\text{bend-force}} = \frac{1}{2}\epsilon \int_0^L \left(\frac{\partial \hat{t}(s)}{\partial s} \right)^2 ds - f \int_0^L \frac{\partial x(s)}{\partial s} ds \tag{2.20}$$

where x is the projection of $\hat{t}$ in force direction along the polymer backbone. Although we could rely on numerical methods to solve Eq. 2.20, we shall not dwell on the exact

numerical methods to derive force-extension relationship, but simply quote the approximate interpolation formulas [70]:

$$\frac{fl_{\mathrm{p}}}{k_{\mathrm{B}}T} = \frac{\langle X\rangle}{L} + \frac{1}{4(1-\langle X\rangle/L)^2} - \frac{1}{4}, \tag{2.21}$$

In regime *III* where the force is relatively weak but still strong enough to escape Pincus regime, this interpolation reduces to:

$$\frac{\langle X\rangle}{L} \approx \frac{2}{3}\frac{fl_{\mathrm{p}}}{k_{\mathrm{B}}T}, \tag{2.22}$$

and in regime *IV* with strong force, we have

$$\langle X\rangle/L \approx 1 - \sqrt{4fl_{\mathrm{p}}/k_{\mathrm{B}}T}. \tag{2.23}$$

The Marko-Siggia interpolation is widely used in predicting wormlike chain extension and fitting the experimental data to obtain the persistence length. It has a rather undemanding formula while generating satisfying results that deviate only a few percent from the exact numerical solution of Eq. 2.20 for ideal wormlike chain. However, it does not come as a surprise that Marko-Siggia interpolation cannot be applied to Pincus regime where the excluded volume region dominates because the derivation of the scaling theory behind Eq. 2.21 is based on ideal wormlike chain model.

The huge advances in single-molecule manipulation techniques over the past three decades make testing the elasticity of DNA feasible. The experimental studies reviewed here measure force-extension relations, thus provide a quantitative test of wormlike chain elasticity and its mechanical flexibility. To date, there are four classes of single-molecule stretching experiments with different types of external applied force. The first class of force is a magnetic field. Smith *et al.* [72] developed a microdevice in which the free end of DNA was attached to a 3-μm-diameter magnetic bead. The DNA experienced a uniform tension along the chain in a known magnetic field, thus allowing measurement of force-extension relation. Because the magnetic bead can be imaged, there was not fluorescent dye needed to stain the entire DNA molecule. The second class of force is the electric field [98, 99]. With one end anchored, DNA as a polyelectrolyte was stretched by the electrophoretic force acting directly on the DNA.

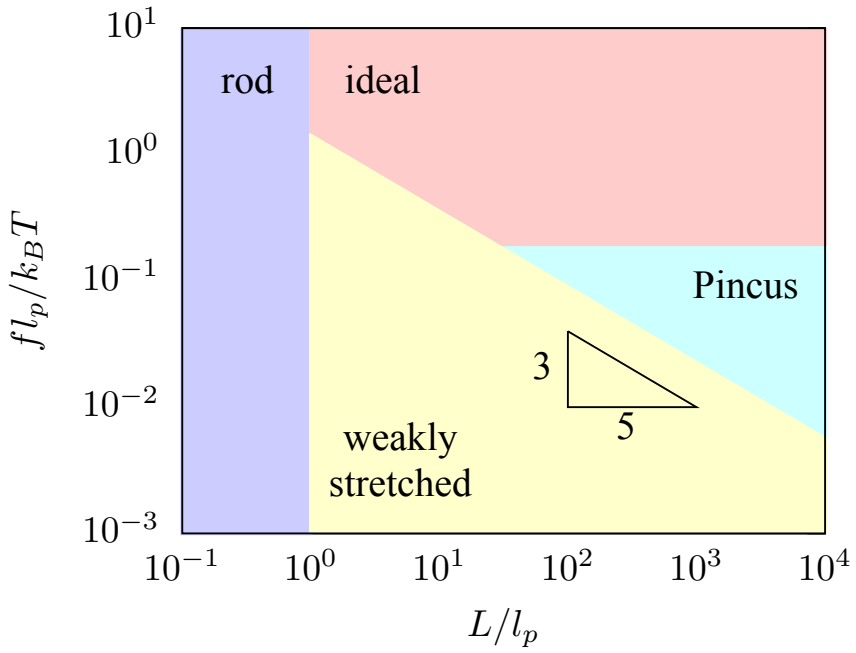

Figure 2.6: Phase diagram of wormlike chain stretched in a free solution. The dark-blue shaded area shows short chains has rod-like behavior. The yellow shaded area indicated long chains under weak stretching forces corresponding to regime *I* in Fig. 2.5. The light-blue shaded area corresponds to regime *II*. The pink shaded area shows idea chain behavior corresponding to regime *III* and *IV*. The triangle shows the scaling law of boundaries between weakly stretched regime and Pincus/ideal regime.

The third class of force is hydrodynamic flow drag [97], where the hydrodynamic force past the coil. These three fields are closer to the kind of stretching that might occur naturally during gel electrophoresis or in shear flows. Moreover, the devices stretch DNA in a continuous process, and therefore are capable of high throughput analysis. Last but not least, techniques of optical tweezers [103,104] with feedback control provide reliable testing of DNA elasticity. In the experiments, one end of a DNA was bonded to a coverglass surface or a bead held with a micropipette, with the other end tethered to a microscopic bead. The coverglass or the bead moved with respect to the microscopic bead. However, the other end, the microscopic bead was not entirely fixed, but rather free to move within a certain range or the "trap". The change of the light intensity, along with the resulting electronic feedback circuit, prevented the bead from escaping

the trap. This system is capable of measuring force-extension relationship with both low (~0.1 pN) and high (~50 pN) forces, therefore provides a reliable experimental system for studies of polymer mechanical flexibility. It is noteworthy that the experiments reviewed above generally permit nonuniform tension instead of the homogenous force field in the theoretical studies, and are closer to the DNA stretched in the real world such as gel electrophoresis and nanopore stretching.

2.3 DNA as a Semiflexible Polymer in Nanoslit Confinement

2.3.1 de Gennes' Blob Theory in Nanoslit Confinement

As mentioned in the last section, DNA often acts as a model polymer for researchers to explore polymer physics. A nanoslit offers one of the simplest confinement that constrains the orientation of polymer in only one dimension, which leads to the "free" span of polymer in a pseudo-2D way. This compression would greatly alter the behavior of confined polymer, thus allows us to study fundamental polymer physics ranging from thermo dynamic properties to dynamic properties. In the scope of the current contribution, we mainly focus on the former class of problems, in particular the equilibrium extension of confined DNA with or without the external forces. The latter typically concerns diffusive behavior and relaxation times, and interested readers should refer to [2, 37, 38, 105]. With the advances in nanofabrication technologies together with fluorescence microscopy imaging, it is possible to observe and measure the static and dynamic behavior of DNA directly. Such devices can be used to not only test existing understanding of polymer physics but also guide the design and optimization of the genomic mapping technologies.

The slit confinement is often referred to as two infinite parallel plates immersed in buffer solution. Note that in the real world, there is no truly unbounded slit that completely eliminates the confinement in the other two dimensions, but rather the slit would be approximated by a nanochannel with high aspect ratio, i.e., the size of the plane is much wider than the height of the slit, and also much greater than the in-plane radius of gyration $R_{g,||}$ of the polymer chain so that the polymer would not feel the wall

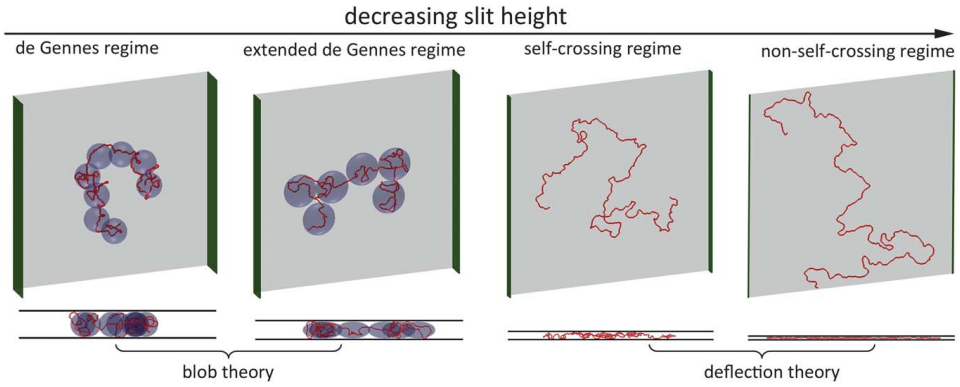

Figure 2.7: Top and side views of regimes of slit-confined wormlike chain proposed by Dai *et al.* [2]. Decreasing slit height causes the transition from the weakly confined de Gennes regime all the way to strong confined Odijk regime. Reproduced from [2].

effects.

The equilibrium conformation of a confined polymer is determined by the competition of three lengths: the 3D radius of gyration R_{g}, the distance between two plates H, and the persistence length l_{p}. As a results, multiple regimes of confinement versus extension can be distinguished. In the following discussion of this chapter, we refer the R_{g} in 3D as $R_{\mathrm{g,bulk}}$ to differentiate it from the in-plane radius of gyration $R_{\mathrm{g},\parallel}$. In the weak confinement where $H \gg R_{\mathrm{g,bulk}}$, similar to the weakly stretched regime in free solution in Section 2.2.2, the wormlike chain stays in a coiled structure with one dimension slightly compressed. By decreasing slit height to $H < R_{\mathrm{g,bulk}}$, the de Gennes regime rises as shown in Fig. 2.7. The conformation of the wormlike chain consists of a series of so-called de Gennes blobs of size H, with real chain statistics inside the blobs. These blobs are isometric, and because the size of the de Gennes blob is greater that the size of the thermal blob, the excluded volume interactions cause the polymer to swell in all dimensions, and the free energy caused by the excluded volume interaction inside a de Gennes blob is larger than $k_{\mathrm{B}}T$. For this regime, the blob size R_{b} scales as,

$$R_{\mathrm{b}} \sim H \sim L_{\mathrm{b}}^{3/5} w^{1/5} l_{\mathrm{p}}^{1/5} \tag{2.24}$$

where L_b is the contour length inside a blob. The Flory free energy is [2, 106]

$$\frac{F}{k_B T} \sim \frac{X^2}{(L/L_{blob})H^2} + \frac{(L/L_{blob})^2 H^2}{X^2} \tag{2.25}$$

Differentiating Eq. 2.25 with respect to X for minimum free energy and substituting Eq. 2.24 gives

$$\langle X \rangle \sim L^{3/4} H^{-1/4} w^{1/4} l_p^{1/4} \tag{2.26}$$

Note that we have $\langle X \rangle \sim L^{3/4}$ instead of $\langle X \rangle \sim L$ as in the case of stretching DNA with external force.

Further decreasing H to be below the size of thermal blob l_p^2/w leads to insufficient contour length inside a blob to experience excluded volume effect. In order to maintain the swelled behavior as a real wormlike chain, the blob in the so-called extended de Gennes regime ($2l_p < H < l_p^2/w$) can be conceived in an ellipsoid shape with a shorter axis of H and a longer axis defined as B. By assuming the ideal chain behavior inside a blob and equating the excluded volume energy of a blob to $k_B T$, we have

$$B \sim L_b^{1/2} l_p^{1/2} \sim L_b w^{1/2} H^{1/2} \tag{2.27}$$

which gives

$$L_b \sim l_p H/w \tag{2.28}$$

$$B \sim l_p H^{1/2}/w^{1/2} \tag{2.29}$$

By replacing the spherical blob with ellipsoid in Eq. 2.25, Dai *et al.* [2] obtained the Flory free energy of the extended de Gennes regime as

$$\frac{F}{k_B T} \sim \frac{X^2}{(L l_p)} + \frac{L^2 w}{X^2 H} \tag{2.30}$$

and the extension $\langle X \rangle$ is the same as Eq. 2.26. A recently simulation study [107] provided evidence of extended de Gennes regime in nanoslit. Highly anisotropic chains ($l_p \gg w$) were studied so that a broad window exists between the lower bound and the higher bound for the extended de Gennes regime $l_p < H < l_p^2 w$.

As the H further decreases to $H < l_{\mathrm{p}}$, the blobs no longer exist because of the strong confinement in the slit dimension. Odijk [108] proposed a deflection model that consists of a series of segments that deflect between the slit planes. The absence of confinement in the other two dimensions causes the segments to be isotropically aligned parallel to the plane surface. Because the confinement is so strong, the slit-confined chain can be effectively viewed as a 2D chain with a weak perturbation in the confining dimension [109]. This theory has been developed further by Tree *et al.* [110] to incorporate effects of real chains and tested through numerical calculation. In the "self-crossing" Odijk regime, the H is large enough for the deflection segment to experience excluded volume interactions, while in the "non-self-crossing" Odijk regime, as the name suggested, the chain can no longer cross itself.

Much experimental progress has been made in understanding slit-confined polymer conformation. In moderate confinement, the experimental evidence has been provided [111–113] to confirm the postulated blob theory by measuring the polymer equilibrium size as a function of H and L. The experimental studies in the Odijk regimes are rather controversial, especially the transition from moderate confinement to strong confinement. Bonthuis *et al.* [112] measured the in-plane radius of gyration $R_{\mathrm{g},\parallel}$ and found a sudden change at $H \approx 100$ nm $\approx 2l_{\mathrm{p}}$. For a smaller H, the $R_{\mathrm{g},\parallel}$ remains invariant to H. By contrast, Tang *et al.* [113] observed a more gradual transition and suggested that there is no Odijk regime. The problem of determining the size of polymer in the slit has attracted a lot of research effor and is still an ongoing topic.

Finally, it is noteworthy that there are a lot of similarities between DNA in slit confinement and DNA under external forces. In the case of confinement, the DNA elongates through the geometry restriction imposed by the slit, which is similar to external force f. The size of the de Gennes blob and tensile blob are H and $k_{\mathrm{B}}T/f$ respectively and are all in competition with the size of the thermal blob l_{p}^2/w. Applying the two effects simultaneously by stretching a slit-confined real wormlike chain in an external force would enhance the degree of extension. Due to the complicated physical picture on the topic, we take a closer look at the stretched regime in slits in Chapter 5.

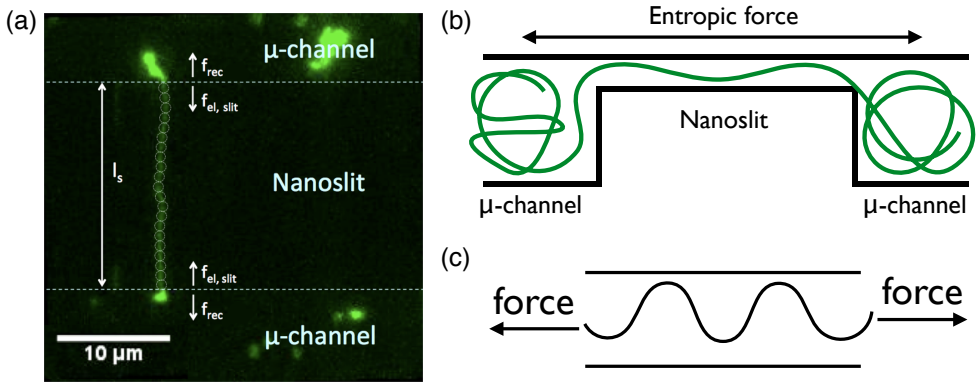

Figure 2.8: (a) The configuration of DNA in the tug-of-war experiments with both ends coiledin microchannels and the middle segment trapped in nanoslit. Reproduced from [114]. (b) The cross-sectional schematic of DNA stretched by two symmetric entropic forces. (c) The system can be modeled as a long DNA trapped in an infinite large slit with two symmetric external forces stretching both ends.

2.3.2 Stretching Semiflexible Polymers in Nanoslit Confinement

Many studies have been focused on understanding the force-extension behaviors of semiflexible polymers in nanoslit [105, 114–126]. One motivation comes from exploring the force-induced chain deformation in confinement and the backbone entropic elasticity. In the experimental setups, a single DNA molecule spans in a nanoslit connecting either two larger reservoirs [117, 119, 120] or a number of discrete nanopits [105, 123–126]. Yeh *et al.* [114,117] studied the dependence of the entropy-driven polymer statics and dynamics under the entropic force in the nanoslit. As shown in Fig. 2.8(a), a single molecule is in the "tug-of-war" configuration with both ends coiling inside the microchannels with the size of hundreds of microns, and the middle part of DNA is trapped in the nanoslit with the size of dozens or hundreds of nanometers. The difference in configurational entropy arising from the difference in mass of relaxed DNA to the slit-confined DNA, which exerts a stretching force on the middle portion of the chain in the slit. Fig. 2.8(b) shows the cross-sectional schematic of the trapped DNA. The confinement-induced entropic forces exerted at micro-nano interfaces are symmetric, thus we can simplify the system of interest as a long DNA trapped in an infinite large slit with two symmetric external forces stretching both ends, and the external forces are balanced by the entropic elastic

forces. These kind of experiments help the understanding of both static and dynamic behavior of DNA in a confined system and shed lights on the single molecule device design. Another motivation is that the external force and the slit confinement would stretch the DNA simultaneously to a strong extensional deformation, thus enhances imaging focus and resolution. Balducci *et al.* [127] induced hydrodynamic extensional flows to stretch DNA in a cross-slot with a 300-nm nanoslit. They showed clear evidence that both external forces and the confinement lead to an improved stretching and a longer relaxation time of the chain when compared to the cases with just one effect present.

The configuration of the wormlike chain is determined by balancing the thermal blob size, de Gennes blob size, and tensile blob size. Taloni *et al.* [114] studied the transition from weak confinement to strong confinement under various of forces. They proposed the scaling arguments that merged the scaling laws in Section 2.3.1 for wormlike chain in slits in the absence of force with the scaling laws in Section 2.2.2 for the stretching of a real wormlike chain in the absence of confinement. When the stretching force is strong, the excluded volume effects between different segments of the polymer chain inside the slit are minimal, so that it is appropriate to model the problem as stretching a confined ideal chain in an external force, with the blobs or deflection segments aligned in the force direction. However, as the force decreases, excluded volume interactions become crucial so that real chain statistics with Flory exponent must be taken into account. Successful scaling laws should decide when theories developed for the stretching of ideal chains can be safely applied, and when to modify the model to account for excluded volume. In particular, the controversy lies in determining the force required to escape the Pincus regime in strong confinement. Because of the confusion on the topic, we briefly reviewed the scaling arguments originally set forth by Taloni *et al.* [114] and proposed a modification to their theory for strong confinement in Chapter 6.

2.4 Recent Computational Studies on Methylcellulose Gelation Mechanism

In order to study the gelation mechanism, several computational studies have been conducted to explore how methylcellulose chains [1,69,129–133] collapse in various solution

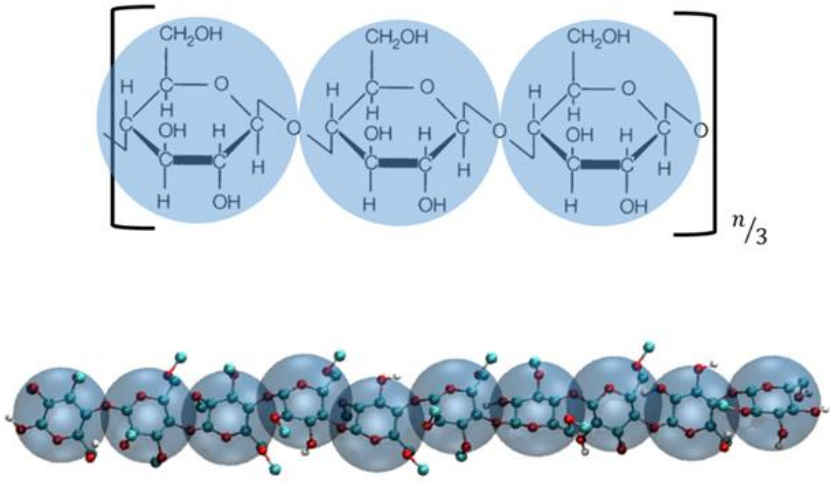

Figure 2.9: Schematic of the coarse-grained model of methylcellulose. Each bead represents one monomer. The beads are connected via harmonic or FENE springs. The bonded interactions include bond potential, bending potential and dihedral potential that apply to any two, three and four consecutive beads respectively. The intermolecular non-bonded interaction is modeled with Lennard-Jones potential. Reproduced from [128].

conditions. In a particularly relevant work, Kong *et al.* [129] systematically studied the collapse transition of a generic semiflexible polymer model with variable backbone stiffness and self-attraction strength. They constructed a phase diagram of the resulting collapsed states including toruses, globules, and bundles. In a subsequent study, Huang *et al.* [1, 130] developed a multi-scale simulation specifically targeting methylcellulose chemistry. With respect to understanding the connection between fibril formation and gelation, it is important to have a chemically realistic model since the sol-gel transition is closely related to the strength of methylcellulose self-attractive interactions.

To this end, Huang *et al.* first conducted atomistic simulations for 10-mer methylcellulose oligomers [130]. In this work, they constructed methylcellulose monomers with all possible combinations of methyl substituents. Homo-oligomers and random oligomers were constructed then to evaluate different atomistic force fields through the conformational preference model dimer cellobiose. The results from the atomistic simulation were then used to parameterize the force fields appearing in their coarse-grained methylcellulose model. As shown in Fig. 2.9, the coarse-grained methylcellulose was

represented with a bead-spring model, with each bead located in the center-of-mass of each monomer. With isolated chain simulations, Huang *et al.* [1] found that at 50 °C, methylcellulose chains with 600 monomers or more can adopt a ring-like structure in a dilute aqueous system. Importantly, while this ring structure is not a high-aspect ratio fibril, it is plausible that they are precursors to forming methylcellulose fibrils. Based on this coarse-grained model [1], Ginzburg *et al.* [69] developed a statistical mechanical model for the behavior of multiple methylcellulose chains and proposed that the fibril is a one-dimensional stack formed by methylcellulose rings.

Although these computational studies [1, 69, 129] provide a plausible model for fibril formation in dilute solution, they leave open questions remained to be answered, especially how to remove the mis-collapsed states and what are the important physical properties that determine the configuration of the methylcellulose model. We address these questions in Chapter 6.

Chapter 3

Simulation Method

Although biopolymers can have different chemistries, they share some common physical properties: high molecular weights, stiff backbones, and complex internal structure. Computer simulation, therefore, plays quite an important role since it provides a way to study a generic model (Fig. 3.1). By changing the parameters appearing, this model permits studying a huge variety of biopolymers including both DNA and methylcellulose.

In addition, simulation provides opportunities to explore conditions that are crucial for understanding the basic properties of biopolymers. Such conditions can be challenging for experiments due to the cost and the limited capabilities of fabricating devices and also for analytical treatment due to the complexity of the parameter space and the limitation of applying over-simplified theoretical models. Simulation has its own advantages in that it can generate information relatively quickly to evaluate a large part of the parameter space numerically, thus providing a thorough understanding of biopolymers in various conditions. In this work, we have focused largely on studying physical properties of biopolymers through simulation, which allows building connections between theoretical predictions and experimental studies.

In this dissertation, we are mainly interested in the force-extension relationship of stretched DNA as a wormlike chain, and the gelation mechanism of methylcellulose. Towards the DNA simulation, our focus will be on describing the pruned-enriched Rosenbluth method (PERM). In particular, we will explain our implementation of off-lattice PERM with external force potentials. We will also explain Langevin dynamics with a focus on exploring methylcellulose gelation mechanisms in Chapter 6.

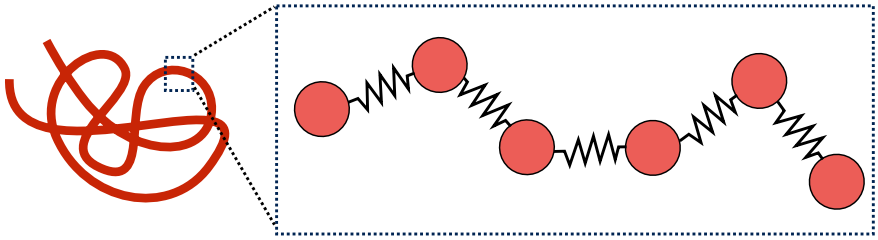

Figure 3.1: A generic model (right) for a biopolymer chain (left). Each bead represents a segment of the biopolymer chain, and the beads interact with each other by bonded forces and non-bonded forces.

3.1 Monte Carlo Chain Growth Simulation

Monte Carlo methods have been used extensively to calculate thermodynamic properties of polymer [134]. The essential idea is to use importance sampling to sample configurations or states with a probability proportional to Boltzmann factor, which allows sampling of the energy favorable parts of the phase space that contribute most to the thermodynamic average. The conventional Monte Carlo method, known as the Metropolis Monte Carlo method [135] first proposes an initial state, then permits moves in the phase space and accepts or rejects the moves with a certain probability associated with the free energy. Although this method can accelerate the process of predicting configurations compared to brute-force search, it is insufficient to reach the experimentally relevant molecular weights of the DNA. The maximum number of beads that can be afforded by Metropolis Monte Carlo simulation is of $\mathcal{O}(10^3)$, while as least one magnitude more is required to reach the long chain limit where the theories of polymer physics are valid and still capture the sub-persistence length behavior [83]. Moreover, in the conventional Monte Carlo simulation, one has to perform independent simulations for different molecular weights. As a result, multiple conventional Monte Carlo simulations at various lengths are required to draw any relationship for scaling laws as a function of molecular weights.

Monte Carlo chain growth simulation, however, is capable of simulating longer polymer chains and estimating thermodynamic properties as a function of the molecular weight. In Monte Carlo chain growth simulation, polymer chains are built as self-avoiding walks (SAW) by adding one bead at each step. Self-avoidance means that

no overlapping configuration is allowed in which the same space is occupied by two or more beads. To illustrate the concept of Monte Carlo chain growth method, especially PERM, I will use a self-avoiding walk (SAW) with 201 beads (200 steps) linear chain on a 2D lattice as an example. To start the process, the first bead is placed at a random site on the 2D lattice, and each of the successive bead is placed at a neighboring free site of the previous bead (as a linear polymer chain), until there is no more free site to grow, or the chain reaches it maximum length (200 steps). Here, we intend to calculate thermodynamic averages of the chain configuration, such as the mean radius of gyration $\langle R_{\mathrm{g}} \rangle$ and the mean end-to-end distance $\langle X \rangle$ of the 2D SAW chain. In order to get an accurate estimation of the thermodynamic average, the goal here is to generate enough samples that reach the maximum contour length.

In this section, our focus will be to describe the methodology of PERM simulation and in particular, our implementation of off-lattice PERM with discrete wormlike chain model under external stretching forces. We will motivate the use of PERM by introducing two rather elementary chain growth methods, which are Simple sampling (Section 3.1.1) and Rosenbluth-Rosenbluth method (Section 3.1.2) and showing their limitations in simulating long polymer chains of interest. We then introduce the on-lattice PERM (Section 3.1.3) to simulate long chains that can reach molecular weights suitable for comparison with experiments and polymer physics theories. We will compare the performance of the three on-lattice method in the context of a SAW of a linear chain on a 2D lattice with 201 beads. Finally, we move beyond 2D lattice simulation to 3D off-lattice SAW simulation and explain how to implement the external forces and confinement simultaneously into the discrete wormlike chain model (DWLC) along with PERM (Section 3.1.4). Our discussion in this section is mainly inspired by notes on PERM by Prellberg [136] and Muralidhar [137].

3.1.1 Simple Sampling

A straightforward way of sampling SAW is by Simple Sampling, which directly implements the probabilistic growth. In this process, each walk has an equal weight. We will show in this section that while this sampling method is easy to implement and statistically correct, a more sophisticated sampling method is required to sample SAW with more steps.

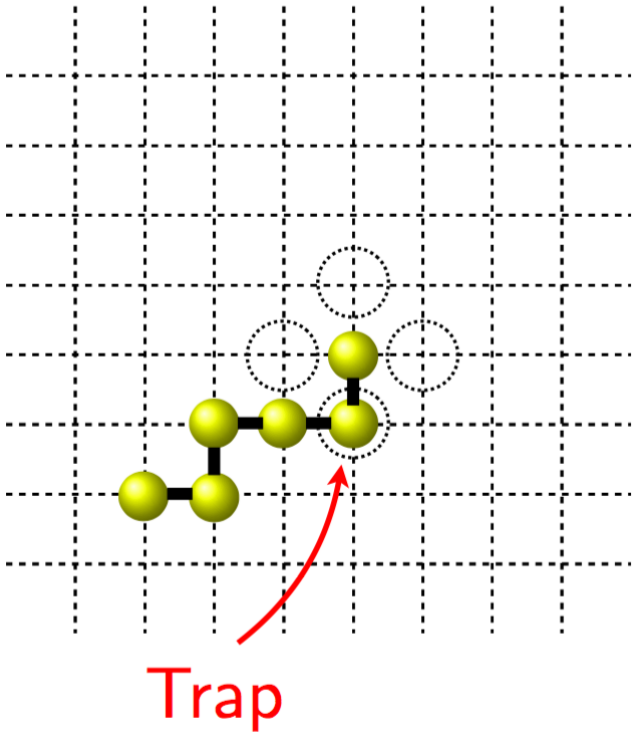

Figure 3.2: Simple sampling of a self-avoiding walk (SAW) on a 2D lattice. A trap means the next step from the growing end overlaps with the existing beads. Because simple sampling samples all the neighboring sites with equal probabilities, the chain often goes into traps. Reproduced from [137].

We start the process by placing the first bead at an arbitrary lattice node at the 0^{th} step and fix it as the origin of the coordinate system. The first bead is now the "growing" end of the polymer chain. For the i^{th} step, we scan the neighborhood of the growing end and pick one neighboring node with equal probability, which is 1/4 on the 2D lattice because each node has four neighbors. We then add the $(i+1)^{\text{th}}$ bead at the chosen lattice node up until the chain reaches the prescribed maximum number of beads, which is 201 in this context, or the chosen lattice node at current step is occupied by the existing beads that have already been grown in the current linear chain, shown as a "trap" in Fig. 3.2. If the process is not terminated by either of the two conditions, the growing process proceeds by making the $(i+1)^{\text{th}}$ bead the new growing end and picking among new neighbors with equal probability. One such realization from placing

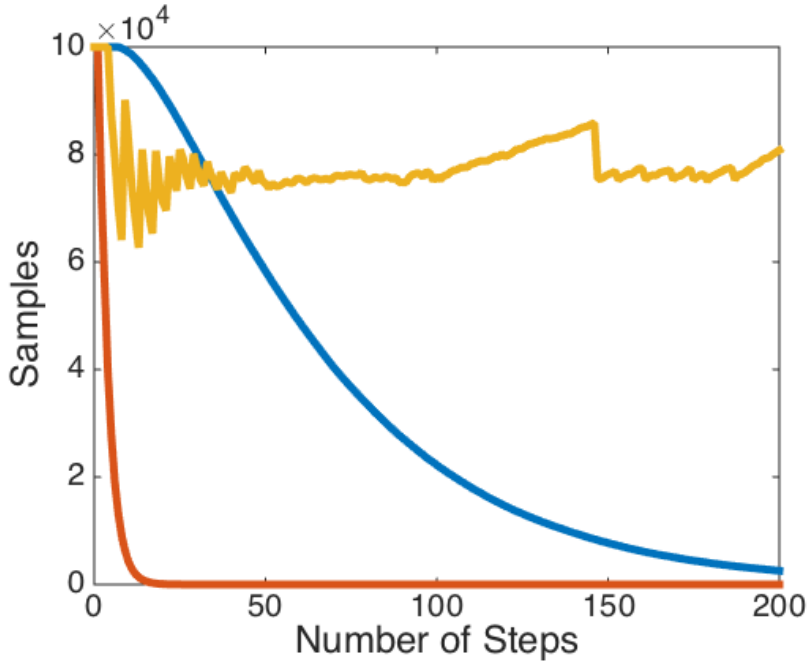

Figure 3.3: Comparison of the performance of different Monte Carlo chain growth method for 2D SAW with 200 steps (201 beads). The red curve is from simple sampling, the blue from Rosenbluth-Rosenbluth Method, and the yellow curve is from PERM. The number of tours performed for each method is 10^5.

the first bead towards terminating the growth is called a tour. We then start over again to develop another tour until we have enough tours to estimate the ensemble average of thermodynamic properties. For example, the ensemble average of the radius of gyration of a chain with n steps $\langle R_{g,n} \rangle$, can be obtained by averaging the radius of gyration obtained from different tours,

$$\langle R_{g,n} \rangle = \frac{\sum_{t=1}^{t=N_t} R_{g,n}^t}{N_t}, \tag{3.1}$$

where $R_{g,n}^t$ is the radius of gyration of a chain with n steps in the t^{th} tour and N_t is the total number of tours in the simulation. Note that this is an arithmetic average over N_t because every configuration is developed unbiased with an equal probability.

Although Simple Sampling is straightforward, it is inefficient for sampling self-avoiding chains with $N \geq 10$. This is easy to explain: for the best scenario of a straight chain, there is still a probability of $1/4$ that the current growing end will choose the lattice site occupied by the proceeding bead, thus terminating the process. In general, as the number of steps N increase, the probability of sampling a SAW with simple sampling decreases exponentially with N, because there are 2.638^N SAWs [136] for 4^N random walks on a 2D lattice. For a chain with 10 steps (11 beads), the probability of sampling a SAW without any termination before reaching to the end is $(2.638/4)^{10} = 0.0157$, and for the chain of $N = 201$, the probability decreases to 6.96×10^{-37}. Figure 3.3 shows the number of samples as a function of the length of the chain. We observe that simple sampling is extremely inefficient to sample SAW for large N, which is often required in the Monte Carlo chain growth simulation.

3.1.2 Rosenbluth-Rosenbluth Method

An improved sampling algorithm was proposed in 1956 by Rosenbluth and Rosenbluth [138]. The basic idea to grow self-avoiding configurations by only sampling the next step from empty lattice sites instead of sampling from all the neighboring site with equal probability. In this way, the process terminates only if the polymer chain reaches the maximum length, or the chain was "trapped" in a dead end with no space to grow, as in Fig. 3.4. We will show in this section that comparing to simple sampling, this method is able to sample SAW with more steps before termination, However, the occurrence of traps, or the attrition problem, still increases exponentially with the contour length of the polymer chain.

As shown in Fig. 3.5, in the original Rosenbluth-Rosenbluth method [138], at the 0^{th} step, the first bead is place at the origin, and the weight of the current step, ω_0 is trivially equal to 1. We use ω_i to denote the weight or the number of free sites of the i^{th} step. For a new tour, there are four possibilities with no intersections, giving $\omega_1 = 4$. When a subsequent bead is added, one has to scan the neighborhood to find the free sites to grow instead of selecting from all the neighboring nodes without bias. If there is more than one free site, one needs to choose among them uniformly. If there is no free site, the chain is killed. The weight of a configuration after N steps, W_N, is the

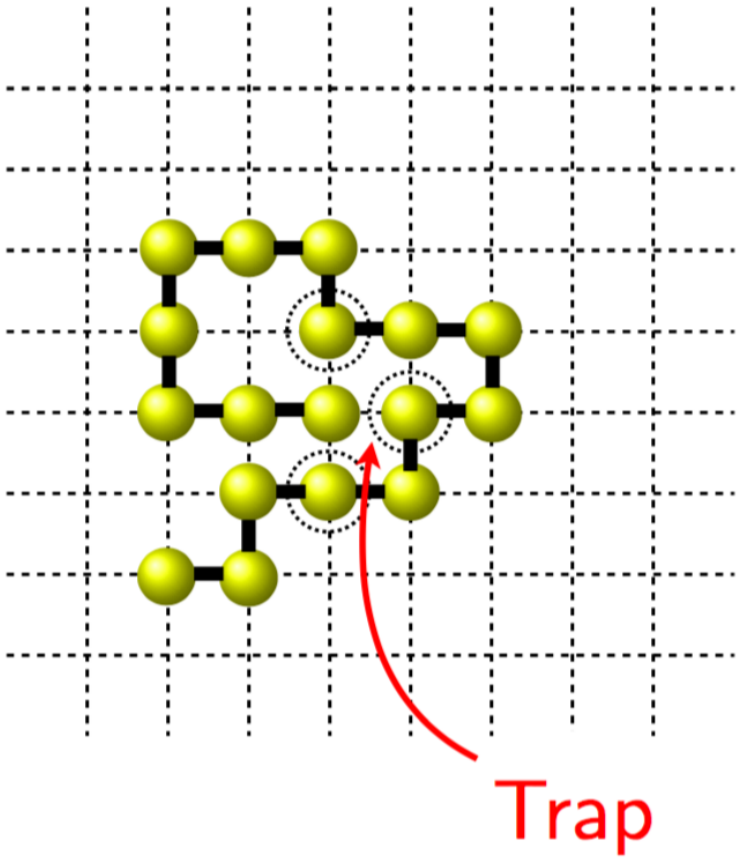

Figure 3.4: Rosenbluth-Rosenbluth method of SAW on a 2D lattice. Traps often occur even if Rosenbluth-Rosenbluth method only sample from neighboring free sites. Reproduced from [137].

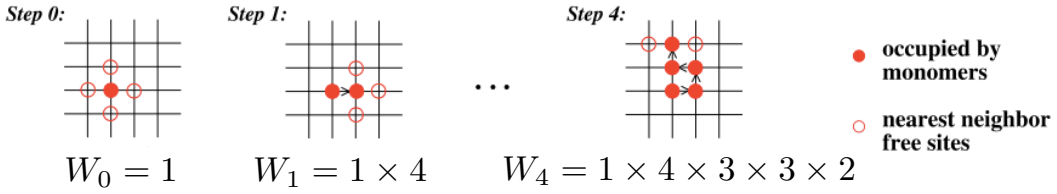

Figure 3.5: Schematic of SAW from the 0^{th} step to the 4^{th} step. Weights from each step are shown at the bottom. The Rosenbluth bias is used such that free sites are picked uniformly. Reproduced from [139].

product of ω_i of the chain,

$$W_N = \prod_{i=0}^{N} \omega_i \tag{3.2}$$

W_N simply represents the number of possible self-avoilding configurations of a polymer chain with N steps.

Similar to the simple sampling, the ensemble average of the thermodynamic properties can be estimated by the statistics from each tour. For example, the ensemble average of the radius of the gyration $\langle R_{\mathrm{g},N} \rangle$ can be obtained by averaging the $R_{\mathrm{g},N}$ together with the weight W_N,

$$\langle R_{\mathrm{g},N} \rangle = \frac{\sum_{t=1}^{t=N_{\mathrm{t}}} W_N^t R_{\mathrm{g},N}^t}{\sum_{t=1}^{t=N_{\mathrm{t}}} W_N^t}, \tag{3.3}$$

where W_N^t is the weight of the configuration with N steps in the t^{th} tour. Note that this is a weighted average over the number of tours N_{t} because each configuration is biased towards self-avoiding configurations by only choosing among the free sites on the lattice. As a result, each configuration is associated with an appropriate weight to account for the bias.

Nevertheless, this scheme is not without deficiency in that when the chain length N becomes very large, only an exponentially small fraction of chains survive. Figure 3.3 shows the improvement gained by Rosenbluth-Rosenbluth method over simple sampling and its limitation. The number of chains that reach over 100 steps without termination decreases exponentially, leading to a poor sampling for large chain length. Moreover, as stated in [140], the ensemble average can be dominated by several "powerful" configurations with enormous weight, given the large variation of the weights and the limited number of the tours that can be afforded by the computational capability.

3.1.3 Pruned-enriched Rosenbluth Method (PERM)

PERM was invented as an improvement of the Rosenbluth-Rosenbluth method that is excellent for long polymer chains [141]. This method corrects "bad" configurations that easily go into traps, thus circumvents the attrition problems. At some point along the growth process, it becomes clear that the chain is in a "healthy" condition or has

no free space to grow. PERM is able to grow more configurations and prevent chains from going into traps through population control: when the weight of the chain is large, in other words, the chain has more free space to grow, PERM introduces branches or enrichment to the chain growth; when the weight is relatively low, PERM prunes the chain to prevent the chain from being trapped in a dead end. In contrast to growing a linear structure, chains with tree-like structures are grown in order to search in a depth-first fashion [142].

There are two main types of PERM in terms of deciding when to perform the prune and enrichment [140, 141]: non-blind PERM and blind PERM. A non-blind PERM means that a target weight distribution W_{target} is already known a *priori* and can be used to decide whether the calculated weight at the current i^{th} step W_i is too large or too small [141]. If W_i is too large, we tend to make more copies and divide W_i by the number of copies. If W_i is too small, we tend to terminate the growth of this branch. A blind PERM is when the sampling distribution is not known a *priori* [140]. Instead of a given target weight distribution, one estimates and updates the target weight from the weighted average of the previous tours. This is the variant of PERM that we used in our work because in general, we do not know the sample distribution, especially when the system is complicated. In this section, we will show how non-blind PERM works through a 1D random walk with the target weight known a *priori*, then we will discuss how to perform a blind PERM with both 1D random walk and 2D SAW as examples.

Let us first consider a 1D random walk, as shown in Fig. 3.6(a). We introduce the 1D random walk as an example because the target weight, or the probability distribution, is known as a *priori*: it is simply a binomial distribution with $p_{n,k_n} = \frac{1}{2^n}\binom{n}{k_n}$. p_{n,k_n} denotes the probability of the n^{th} step having k step(s) choosing to go to the right direction. At the 0^{th} step, a bead starts from the origin. Steps to the right are assigned with positive coordinates. A simple sampling would suggest that a new step should be either left or right with equal probability so that each step would be independent of all the previous steps. However, it is not a surprise that this walk can "never" reach the positions that are far away from the origin, in other words very large k_n or very small k_n, because of the nature of the binomial distribution.

PERM can overcome this sampling limitation by generating configurations with different k_n with a uniform probability, and assigning the appropriate weight accordingly.

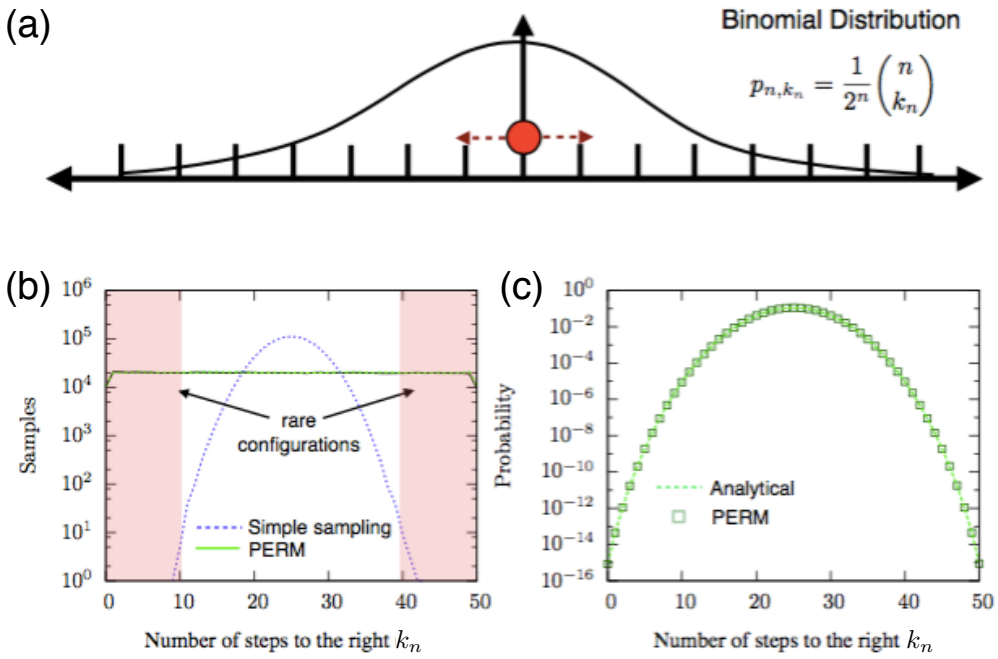

Figure 3.6: PERM of 1D random walk for $n = 50$ steps and 1,000,000 tours. (a) A 1D random walk with the probability distribution equivalent to a binomial distribution. (b) The total number of samples for each value of k are shown for simple sampling (blue-dash line), PERM with predetermined weight (red-dash line) and blind PERM (green line). Red-shaded area is denotes as rare configurations with $k \leq 10$ and $k \geq 40$. (c) Comparison of *priori* binomial probability (green-dash line) and estimated probability from blind PERM (black squares) on a logarithmic scale.

For example, in the 1D random walk, an analytical target weight

$$W_{\text{target},n,k_n} = \frac{n+1}{2^n}\binom{n}{k_n} \tag{3.4}$$

is assigned so that the rule of importance sampling is satisfied with the uniform sampling distribution $p_{n,k_n,\text{uni}} = 1/(n+1)$,

$$p_{n,k_n,\text{uni}} W_{\text{target},n,k_n} = p_{n,k_n} = \frac{1}{2^n}\binom{n}{k_n}. \tag{3.5}$$

This target weight W_{target,n,k_n} is used to decide if the weight of the current walk is

too large or too small. Now suppose that there is a walk with weight W_{n,k_n}. In the ideal scenario, we want W_{n,k_n} to be close to W_{target,n,k_n} with minimal fluctuations. If the ratio $R_{n,k_n} = W_{n,k_n}/W_{\text{target},n,k_n} < 1$, PERM decides the current weight is too small and this walk will stop growing (get pruned) with probability $p_{n,k_n} = 1 - R_{n,k_n}$ or continue to grow with probability $p_{n,k_n} = R_{n,k_n}$. If $R_{n,k_n} > 1$, $\lfloor R_{n,k_n} \rfloor + 1$ copies of the walk will grow (enrichment) with probability $p_{n,k_n} = R_{n,k_n} - \lfloor R_{n,k_n} \rfloor$, or $\lfloor R_{n,k_n} \rfloor$ copies with probability $p_{n,k_n} = 1 - R_{n,k_n} + \lfloor R_{n,k_n} \rfloor$. After determining how many copies to grow, the W_{n,k_n} is divided by the number of copies of the current step. As a result, PERM generates a collection of configurations with weights W_{n,k_n} around the target weight W_{target,n,k_n}, thus each of the configurations is of an equal chance of being generated. In each tour or the fulfillment of the walk, PERM leads to a large tree-like structure with correlated walks grown from one original point at length zero. For other more complicated distributions without W_{target,n,k_n} known a *priori*, W_{target,n,k_n} can be estimated by a weighed average of all the previously sampled configuration,

$$W_{\text{target},n,k_n} \leftarrow (n+1) W_{n,k_n} / \sum_{l=0}^{n} W_{n,l} \tag{3.6}$$

Figure 3.6(b) shows the simulation results of a 50-step 1D random walk with Simple Sampling, blind PERM, and non-blind PERM. As we can see, PERM is more or less a uniform sampling method which is able to generate "rare" configurations. This attribute makes PERM excellent for exploring various configurations of a polymer chain. Figure 3.6(c) shows the probability sampled from a "blind" PERM, and the comparison with the target probability distribution $p_{n,k_n} = \frac{1}{2^n}\binom{n}{k_n}$.

Similar concepts can be extended to 2D SAW with pruning and enrichment, as shown in Figure 3.7. The weight of the i^{th} steps is still the Rosenbluth weight ω_i. The target weight is the average weight of all the previous SAWs, and the prune/enrichment decision is made based on the ratio between the weight of the current tour and the target weight, as discussed in the 1D random walk. The statistics can be calculated similarly to Eq. 3.3. Figure 3.8 shows the square end-to-end distance R^2 calculated from 2D SAW PERM. As we can see, it successfully reproduces the Flory scaling $\nu \simeq 0.75$ in 2D [143]. Grassberger reported that PERM is capable of simulating chains up to $N = 1,000,000$ with high statistics in Θ condition [141].

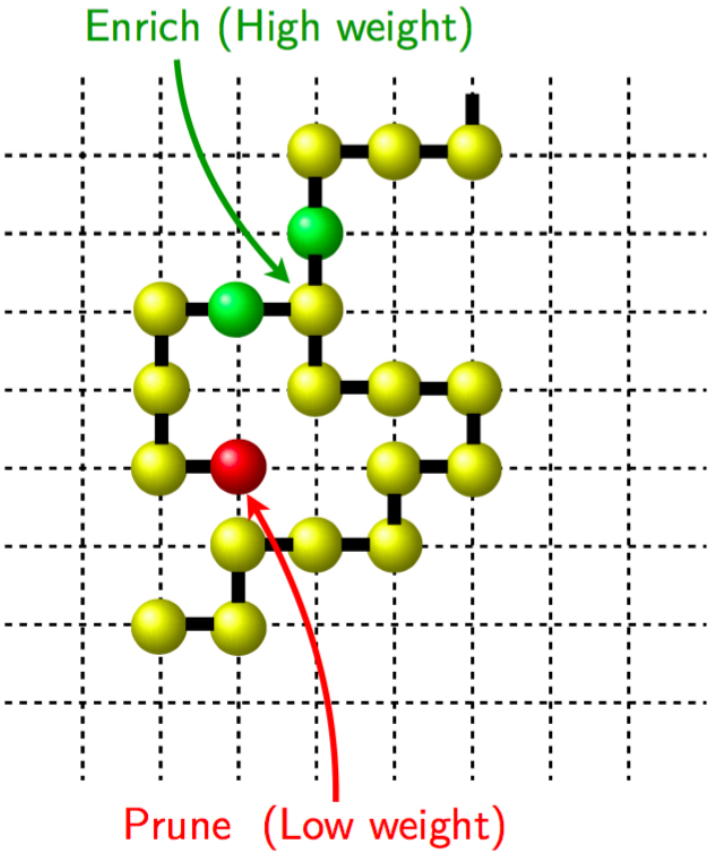

Figure 3.7: Illustration of PERM of a SAW on a 2D lattice. The low weight branches tend to be pruned, and the high weight branches tend to make more copies. Reproduced from [137].

Figure 3.3 shows that compared to simple sampling and Rosenbluth-Rosenbluth method, more samples can be generated with PERM simulation for the full range of the molecular weights, thus overcoming the attrition problem. However, comparing with the 1D case, it is not quite a uniform sampling, with the number of samples fluctuating around the total number of tours. This is due to the fact that in the blind PERM, the target weight is calculated and updated on the fly, reflecting the information gained in previous tours. As a result, the tours performed earlier with a "poor" estimation of the target weight as a reference will result in fluctuations. Nevertheless, these problems can be addressed by increasing the number of tours and getting a good guess of the target

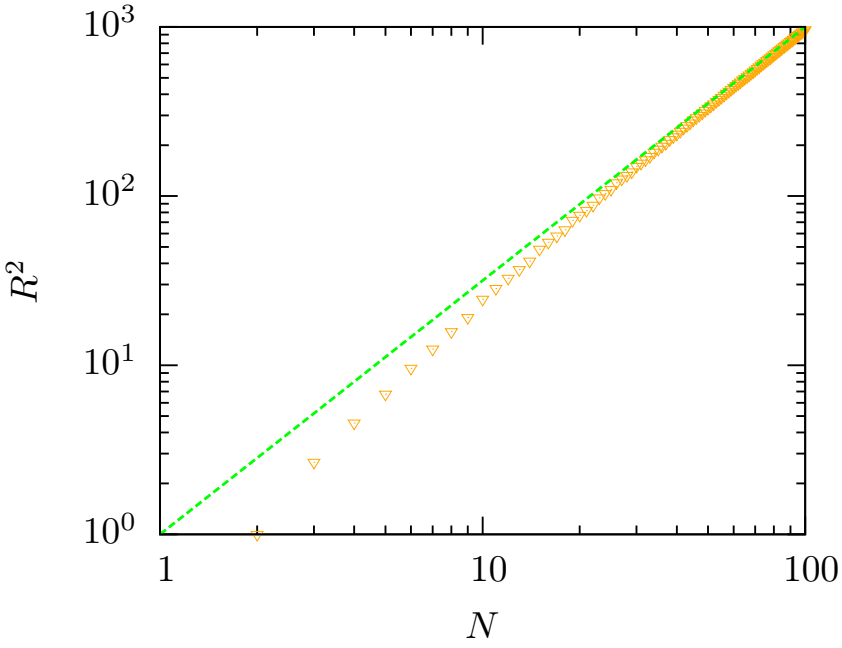

Figure 3.8: Squared end-to-end distance R^2 calculated from 2D SAW PERM simulation. Orange triangles show simulation results. Green-dashed line shows the Flory scaling.

weight before entering the productive simulation, as we will discuss in the next section.

3.1.4 PERM with Discrete Wormlike Chain Model

Thus far, our focus has been on lattice simulations using a variety of sampling methods. The 2D and 3D lattice PERM have been successfully applied to study polymers in various conditions [118, 144, 145]. However, the lattice model also introduces artifacts because only 90° bending are permitted, thus making it (*i*) inadequate to reproduce the sub-persistence length features which are crucial to study wormlike chains and (*ii*) less versatile to include a force field, such as external forces. In order to get a more accurate description of the discrete wormlike chain model under external forces, we sacrifice the computational efficiency of an on-lattice model and move on to discuss how we can perform off-lattice simulations using PERM. In what follows, we explain a 3D off-lattice PERM with SAW that we use to simulate semiflexible polymers with external forces, along with some computational tricks to accelerate the simulation process.

In statistical mechanics, the probability of observing a configuration is proportional to its Boltzmann factor, which is a function of the state energy. Along with the basic idea of Monte Carlo simulation, the i^{th} growth step should be accepted with a probability associated with its state energy U_i. In the lattice PERM and Rosenbluth-Rosenbluth method with SAW, the weight ω_i, or the number of neighboring free sites, can be reinterpreted as

$$\omega_i = \sum_{j=1}^{k} a_i^{(j)}. \tag{3.7}$$

$a_i^{(j)}$ is the atmosphere of the j^{th} site on the i^{th} step

$$a_i^{(j)} = \exp\left(-\beta U_i^{(j)}\right),\ j = 1 \text{ to } k. \tag{3.8}$$

k is the total number of neighboring sites, with $k = 4$ on a 2D lattice and $k = 8$ on a 3D lattice. $U_i^{(j)}$ is the potential energy for the j^{th} lattice site. In the SAW, the $U_i^{(j)}$ is a hard-core excluded-volume free energy (Eq. 2.8), so that $a_i^{(j)}$ is either 0 or 1 indicating the availability of the trial site, and ω_i is the total number of free sites on the lattice.

We made two major changes in the off-lattice PERM: (i) picking the next step with a bending angle ranging from 0° to 180° without binning, and (ii) expanding the state energy $U_i^{(j)}$ beyond just having hard-core excluded volume free energy. To illustrate the off-lattice PERM simulation, we show here how to grow one more step of a 3D off-lattice chain with consideration of the following free energy of a discrete wormlike chain model under external forces,

$$U_{\text{total}} = U_{\text{bend}} + U_{\text{EV}} + U_{\text{force}}. \tag{3.9}$$

U_{bend}, U_{EV} and U_{force} are the same as Eq. 2.6, Eq. 2.8 and Eq. 2.15, respectively.

The chain grows in a similar fashion as discussed in Section 3.1.3. In order to propose the next step with an appropriate weight, we need to take all the of three energy terms into consideration. At some point of the simulation, the chain is in a configuration as shown in Fig. 3.9(a) with i beads. The center of the $(i+1)^{\text{th}}$ bead can be anywhere on the surface of a sphere with the origin being the center of the i^{th} bead and the radius being the bond length a. The atmosphere of the j^{th} trial position is

$$a_i^{(j)} = a_{\text{bend},i}^{(j)} \times a_{\text{EV},i}^{(j)} \times a_{\text{force},i}^{(j)},\ j = 1 \text{ to } k \tag{3.10}$$

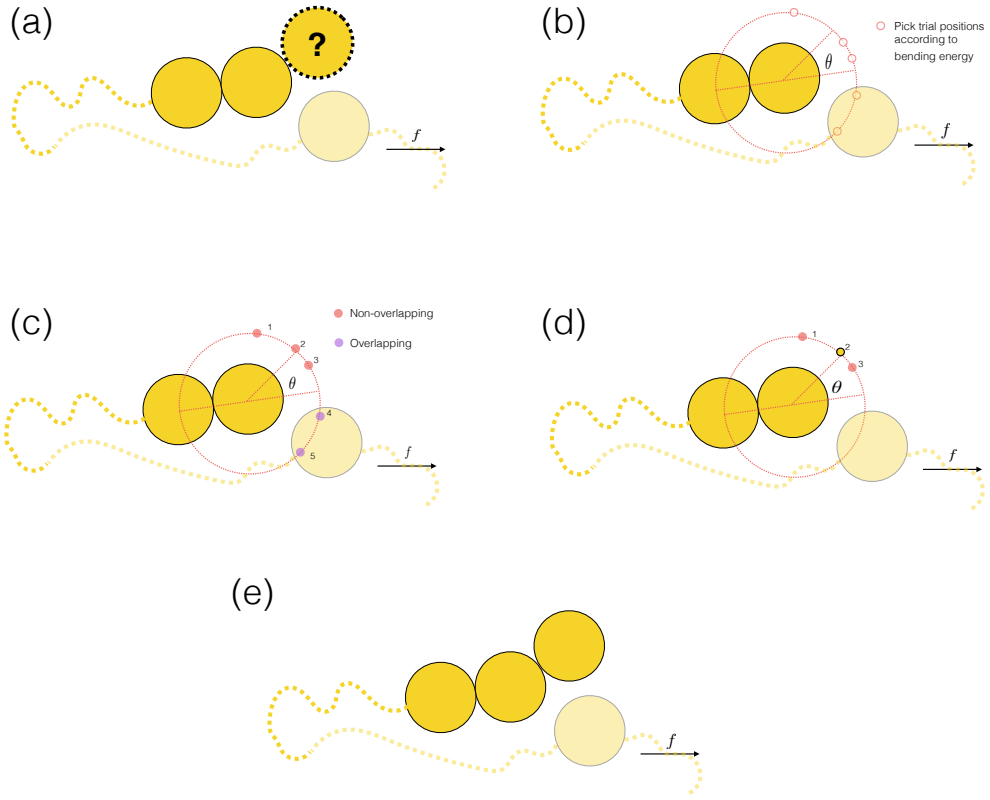

Figure 3.9: Schematics of 3D off-lattice PERM with discrete wormlike chain model under external forces.

and the weight of the i^{th} step is Eq. 3.7.

As shown in Fig. 3.9(b), in the current version of PERM, we generate $k = 5$ trial sites on the surface of the sphere with the probability proportional to $\exp(-\beta U_{\text{bend}})$. Choosing an appropriate k is important to this algorithm. On the one hand, a small k can often lead to the risk that all the trial sites overlap with the existing beads, just like the "traps" in the Rosenbluth-Rosenbluth method, thus $\omega_i = 0$ and the whole branch gets pruned. On the other hand, a high value of k is at the expense of computational time, which is linear in k. The choice $k = 5$ is a satisfactory value that was developed in our group empirically. To incorporate U_{bend}, there are three ways to propose a trial position: picking uniformly and balance with weight; picking with a numerical probability proportional to Boltzmann weight and picking with an analytical probability

proportional to Boltzmann weight. Although all of the methods are statistically correct and follow the rule of importance sampling ($p_1 w_1 = p_2 w_2 \propto \exp(-\beta U)$ with 1 and 2 being two different sampling methods and p being the probability distribution function) the computational efficiency differs a lot due to the pointed shape of the bending energy distribution.

The first method is to sample a trial position uniformly on the surface of the sphere, and put Boltzmann factor associated with bending energy, $\exp(-\beta U_{\text{bend}})$, on the weight side. We could propose the j^{th} ($j \in [1, k]$) trial position randomly and uniformly on the surface of the sphere, and assign the weight of the bending step to be $a^{(j)}_{\text{bend},i} = \exp(-\beta U^{(j)}_{\text{bend},i})$ to ensure consistency with importance sampling. The potential problem of this method lies in the large variation of the weight. For a typical bending constant $\kappa = 10$, the distribution of $a^{(j)}_{\text{bend},i}$ can range over 5 orders of magnitude from $\theta = 0°$ to $\theta = 180°$. This large variation will lead to a wide fluctuation of the final weights, thus fails to provide a balanced ratio between prune and enrichment.

The second approach is to put Boltzmann factor associated with bending energy on the probability side and sample the trial positions numerically. To propose the j^{th} ($j \in [1, k]$) trial position, we first propose m positions randomly and uniformly on the surface of the sphere, and pick one trial position with a probability

$$p^{(j)}_{\text{bend},i,q} = \exp\left(-U^{(j)}_{\text{bend},i,q}/k_{\text{B}}T\right) / \sum_{q=1}^{m} \exp\left(-U^{(j)}_{\text{bend},i,q}/k_{\text{B}}T\right). \tag{3.11}$$

The $a^{(j)}_{\text{bend},i}$ of the j^{th} ($j \in [1, k]$) trial position is $\sum_{q=1}^{m} \exp(-U^{(j)}_{\text{bend},i,q}/k_{\text{B}}T)$. This process repeats k times to propose k trial positions. Although this method reduces the variation in weights, the challenging part is the numerical sampling. As the value of κ increases, the sharpness of the Boltzmann factor increases, so that a very large m is required to get a reasonable collection of samples to describe the distribution, and a total number of $m \times k$ calculations are required, thus making the simulation inefficient.

The third approach is still to put Boltzmann factor associated with bending energy on the probability side, but now the samples are generated inherently satisfying the probability distribution $p^{(j)}_{\text{bend},i} \propto \exp(-U^{(j)}_{\text{bend},i}/k_{\text{B}}T)$. In other words, the k trial samples are now generated according to an analytical probability distribution function (PDF) directly. The atmosphere of the bending energy of the j^{th} trial position is simply

$a^{(j)}_{\text{bend},i} = 1$. This method requires the existence of an analytical invertible PDF to locate the trial positions on the surface of the sphere with a series of the proposed probability. Fortunately, this can be achieved because the bending potential we are using, $\exp[-\kappa(1-\cos\theta)]$, is analytically invertible. This method circumvents the fluctuation of the weight as in the first method, and the numerical inefficiency in the second method due to the limited number of samples available to describe the distribution, thereby improves the computational efficiency and allows growing very long chains with $\mathcal{O}(10^5)$ beads. Therefore, we use this method to propose k trial positions on the surface of the sphere with the probability proportional to bending energy Boltzmann factor. As shown in Fig. 3.9(b), the $k = 5$ trial moves are generated directly from the bending energy inverted PDF. The trial positions favor the orientation along the direction of the previous bond vector.

To account for the self-avoiding nature of the polymer chain, we need to check if each of the trial positions overlaps with the existing beads. According to Eq. 2.8, the Boltzmann factor associated with U_{EV}, or the atmosphere of the excluded volume energy of the j^{th} trial position $a^{(j)}_{\text{EV},i}$, is either 0 or 1, indicating the overlapping and non-overlapping configurations, respectively. As shown in Fig. 3.9(c), among the $k = 5$ trial positions, the 4^{th} and 5^{th} trail positions overlap with the existing beads that have been previously grown in this tour, thus $a^{(4)}_{\text{EV},i}$ and $a^{(5)}_{\text{EV},i}$ are 0, so that there is no possibility to continue to grow the next step in these two trial positions.

Checking self-avoidance can be time-consuming in an off-lattice simulation. For on-lattice simulation, one only needs to check the neighboring sites, which is 4 in the 2D case and 8 in the 3D case. However, in the off-lattice simulation, there is no limitation on the orientation, so that the step of checking the collision need to loop over all the previously grown beads. This leads to $\mathcal{O}(N)$ in each step and a total processing time of $\mathcal{O}(N^2)$. To address this challenge, we introduce neighboring list based on the Verlet neighbor list [146]. The details are available in [137]. The idea is to maintain a list of all beads within a given cutoff distance of the current growing end. Whenever calculating the self-avoidance, one only check with the beads within the neighbor list instead of checking with all the existing beads, and thus accelerate the simulation process. The neighbor list is constructed and updated along with the growth of the chain and reconstructed when a newly added bead is out of the cutoff zone. In our simulation, we fix the cutoff distance

to be $22w$. The neighboring list makes our algorithm much more efficient compared to the PERM code without the neighbor list.

The final step is to pick among the non-overlapping trial position with the probability

$$p^{(j)}_{\text{force},i} = \exp\left(-U^{(j)}_{\text{force},i}/k_{\text{B}}T\right)/\sum_{j=1}^{k}\exp\left(-U^{(j)}_{\text{force},i}/k_{\text{B}}T\right),\ j \text{ in 1 to } k,\ U^{(j)}_{\text{EV},i} \text{ not 0.} \quad (3.12)$$

The atmosphere of the external force of the j^{th} trial position is

$$a^{(j)}_{\text{force},i} = \exp\left[-\beta\left(-\vec{f}\cdot\vec{r}^{(j)}_i\right)\right]. \quad (3.13)$$

Thus, according to Eq. 3.10, the total atmosphere $a^{(j)}_i$ can also be expressed as

$$a^{(j)}_i = \exp\left[-\beta\left(U_{\text{EV}} - \vec{f}\cdot\vec{r}^{(j)}_i\right)\right] \quad (3.14)$$

We then select one of the k trial moves with probability

$$p_i{}^{(j)} = \frac{a^{(j)}_i}{\omega_i} \quad (3.15)$$

where ω_i is calculated from Eq. 3.7. During a tour, we track the cumulative weight for a configuration W_n with Eq. 3.2. The prune and enrichment methodology is the same as in Section 3.1.3.

Finally, Algorithm 1 summarizes our PERM implementation with hard-core repulsion and external forces. Part of the pseudocode was adopted from [136] and [137].

We performed PERM simulations with a Message Passing Interface (MPI)-based parallel Fortran code using a master-slave algorithm (Fig. 3.10). The feasibility of parallelization lies in that each tour is relatively independent, however requires some information from all the completed tours as a reference (e.g., the target weight). Once initialized, the master processor assigns jobs (tours) to slave processors along with the necessary statistics. Each slave processor then runs jobs, gets back to the master processor by sending weight of the current tour and other thermodynamic statistics, and asks for new jobs until the maximum number of tours is reached.

Algorithm 1 Implementation of off-lattice SAW PERM with external forces

$s. \leftarrow 0$; w. $\leftarrow 0$; $Tours \leftarrow 0$; $n \leftarrow 1$; $w_1 \leftarrow 1$; a $\leftarrow 0$; $Copy_1 \leftarrow 1$; $s_1 \leftarrow s_1 + 1$
$r_1 \leftarrow$ Set first bead position; Start neighbor list; Sum. $Z_1 \leftarrow$ Sum. $Z_1 + w_1$
while $Tours < MaxTours$ **do**
    if $n = MaxBeads$ or $a = 0$ **then** $Copy_n \leftarrow 0$
    else $Ratio \leftarrow w_n/W_n$; $p \leftarrow Ratio$ mod 1; Generate random number $\rho \in [0; 1]$
        if $\rho < p$ **then** $Copy_n \leftarrow \lfloor Ratio \rfloor + 1$
        else $Copy_n \leftarrow \lfloor Ratio \rfloor$
        end if
        if (Blind) **then** $W_n =$ Sum. $Z_n/(Tours + 1)$; $w_n = W_n$
        else $w_n = W_n$
        end if
    end if
    if $Copy_n = 0$ **then**
        while $n > 1$ and $Copy_n = 0$ **do** $n \leftarrow n - 1$
        end while
        if $n > 1$ **then** Construct new neighbor list
        end if
    end if
    if $n = 1$ and $Copy_n = 0$ **then**
        $Tours \leftarrow Tours + 1$; $a \leftarrow 0$; $Copy_1 \leftarrow 1$; $s_1 \leftarrow s_1 + 1$; Sum. $Z_1 \leftarrow$ Sum. $Z_1 + w_1$
    else
        Update/Construct new neighbor list
        if $a > 0$ **then**
            $Copy_n \leftarrow Copy_n - 1$
            Draw 5 trial positions on the surface of a sphere from the DWLC bending energy distribution
            Check self-avoidance of the 5 trial positions
            Calculate atmosphere of the force energy $a_n^{(j)}$ with j in non-overlapping trial site(s).
            Calculate weight of the force energy ω_n with Eq. 3.7 with j in non-overlapping trial site(s)
            Pick one of the non-overlapping sites randomly with $p_n = a_n^{(j)}/\omega_n$
            $n \leftarrow n + 1$; $w_n \leftarrow w_n \times a$; $r_n \leftarrow$ Picked site; $s_n \leftarrow s_n + 1$; Sum. $Z_n + w_n$
            if (Production run) **then** #Statistics for thermodynamic averaging
                Sum. $R_n \leftarrow$ Sum. $R_n + w_n \times R_n$
            end if
        end if
    end if
end while
$Z_n \leftarrow$ Sum. $Z_n/Tours$
if (Production run) **then** $R_n \leftarrow$ Sum.R_n/Sum. Z_n
end if

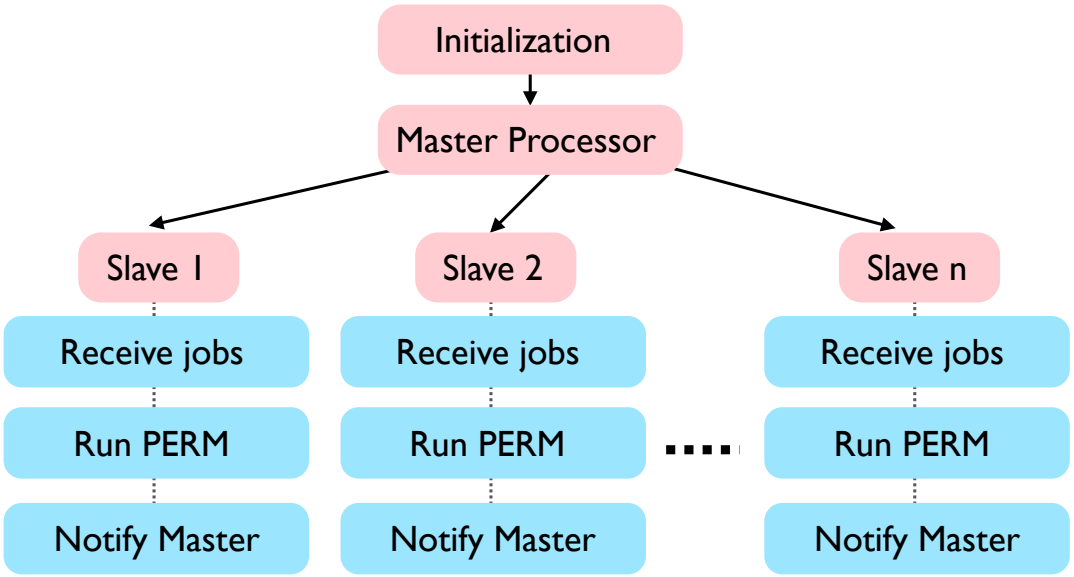

Figure 3.10: Schematic of master-slave parallel algorithm implementation in PERM. Master processor assigns jobs to slave processors and keeps track of all the statistics. Slave processors run PERM simulations tour by tour and send statistics back to master processor.

Chapter 4

Modeling the Stretching of Wormlike Chains in the Presence of Excluded Volume

This chapter is based on the publication
X. Li, C. M. Schroeder, and K. D. Dorfman, "Modeling the stretching of wormlike chains in the presence of excluded volume"
Soft Matter, vol. 11, pp. 5947-5954, 2015 [147]

4.1 Introduction

Stretching a polymer chain in the presence of an external force is a classic problem in polymer physics [92]. In general, entropic and enthalpic intramolecular interactions must be considered for an accurate description of polymer elasticity [89]. An ideal flexible polymer chain with no enthalpic interactions (which, for our purposes, means no excluded volume) can be modeled as a simple random walk with a Gaussian distribution function for the end-to-end extension. In the limit of low forces, equilibrium thermodynamics suggests that the force f required to stretch a chain is linear in the extension. For an ideal freely-jointed chain, the (dimensionless) entropic force is given

by the Hookean expression (similar to Eq. 2.22)

$$F^{\mathrm{H}} = \frac{3}{2}z \tag{4.1}$$

where $F = fl_{\mathrm{p}}/k_{\mathrm{B}}T$ is the dimensionless force for a chain of persistence length l_{p} with $k_{\mathrm{B}}T$ being the Boltzmann factor, and $z = X/L$ is the fractional extension for a chain of contour length L with X being the extension along the force direction. For most practical situations, the elasticity needs to be modified to consider the effects of high forces (i.e. finite extensibility) and excluded volume interactions. In the limit of high forces, a partition function approach can be used to describe the elasticity of an ideal freely-jointed chain,

$$F^{\mathrm{FJC}} = \frac{1}{2}\mathcal{L}^{-1}(z) \tag{4.2}$$

where $\mathcal{L}(x) = \coth(x) - x^{-1}$ is the Langevin function, which has no analytical inverse.

Wormlike chains represent a different class of macromolecules with a uniform distribution of bending stiffness along the polymer backbone. Over 20 years ago, Marko and Siggia [70] proposed the interpolation formula (similar to Eq. 2.21)

$$F^{\mathrm{WLC}} = z + \frac{1}{4(1-z)^2} - \frac{1}{4} \tag{4.3}$$

to describe the extension of a wormlike chain under tension. The Marko-Siggia formula correctly limits to Eq. 4.1 in the low-force limit and the saturation value $F^{\mathrm{WLC}} \cong [2(1-z)]^{-2}$ in the high-force limit. Equation 4.3 only deviates by a few percent from the force-extension result computed numerically from the Hamiltonian of an ideal wormlike chain [70]. As a result, the Marko-Siggia force relation has found widespread use, most notably in the description of the force-extension behavior of DNA [148] below the B-to-S transition at 70 pN [103, 149]. It is used ubiquitously as the spring-force in bead-spring models of wormlike chains such as DNA [150, 151].

For real polymer chains, incorporation of excluded volume interactions is challenging and has been considered using renormalization approaches [92]. In a classic paper, Pincus used scaling theory to show that the restoring force for real polymers scales

non-linearly with extension in the limit of low forces [101],

$$F^{\text{real}} \propto z^{3/2}, \qquad z \ll 1 \tag{4.4}$$

Clearly, the elastic behavior for real polymer chains is strikingly different than the Hookean response for ideal chains.

A key limitation of the Marko-Siggia and Langevin force relations is that these relations were obtained for ideal chains, and thus cannot account for the excluded volume interactions and the concomitant nonlinear-low force elasticity of a real chain. For very stiff chains, where the persistence length l_{p} is much larger than the backbone width w, the excluded volume is weak [83, 152]. However, for single-stranded DNA and many synthetic polymers, the monomer anisotropy ratio l_{p}/w is modest and excluded volume effects can be important [83]. It would be highly desirable to have an interpolation formula similar to Eq. 4.3 to interpret force-extension experiments with such molecules [153, 154]. Such a formula is even more important for modeling the behavior of these polymers in flow using coarse-grained, bead-spring models. For efficient modeling, each spring must represent a large number of persistence lengths. When bead-spring models are used to study polymer dynamics at the relatively low flow strengths encountered in many experimental systems, excluded volume *within* a spring can become important.

In the present chapter, we propose an interpolation formula for wormlike chains that connects the Pincus regime [101] in the presence of strong excluded volume interactions to the Marko-Siggia result [70] for ideal wormlike chains. Using simulations of a discrete wormlike chain model, we show that this interpolation formula provides a good description of the force-extension behavior for all values of the monomer anisotropy ratio l_{p}/w we studied over experimentally relevant values of the fractional extension. A key challenge in our work is simulating chains with high resolution of the chain backbone up to a sufficiently high molecular weight to observe the Pincus regime [155]. While it is possible to reach such high molecular weights by reducing the number of degrees of freedom with a lattice model [121] or by reducing the resolution of the chain backbone with a bead-rod model [100], the off-lattice pruned-enriched Rosenbluth method (PERM) used previously to study discrete wormlike chains in free solution [83] and in confinement [81, 110, 156–158] is readily adapted to the force-extension proble [121].

Using this approach, we are able to simulate asymptotically long chains down to small values of the fractional extension ($z \approx 0.1$) over a wide range of l_p/w values, thereby accessing all of the relevant regimes. The results of these simulations not only allow us to assess the accuracy of our interpolation formula relative to the Marko-Siggia force relation, but also provide strong support for the existence of the Pincus regime.

4.2 Interpolation Formula for the Stretching of Real Wormlike Chains

We propose that Eq. 4.3 should be replaced by an excluded volume-wormlike chain (EV-WLC) interpolation formula consisting of two parts,

$$F = F_{\text{low}} + F_{\text{high}} \tag{4.5}$$

The quantity

$$F_{\text{low}} = \frac{z^{1.5}}{0.21(w/l_p)^{1/2} + (2/3)z^{1/2}} \tag{4.6}$$

is the dominant contribution for small z, with the constant 0.21 determined from a fit to our simulation data. Conversely, the term

$$F_{\text{high}} = \frac{1}{4(1-z)^2} - \left(\frac{1}{4} + \frac{z}{2}\right) \tag{4.7}$$

is the dominant contribution at high z. Note that the leading-order term in F_{high} is $O(z^2)$ for small z. Since the constant 0.21 in Eq. 4.6 was determined by fitting to simulation data for a discrete wormlike chain model, this parameter may differ for an interpolation formula describing a continuous wormlike chain model. However, it is worth keeping in mind that the overall form of the EV-WLC interpolation formula (i.e., the limiting behavior and the crossovers between different regimes) does not assume a discrete wormlike chain model.

The rationale for this formula is threefold:

First, when the chain is strongly stretched ($z \approx 1$), excluded volume should not be important and the Marko-Siggia result for ideal chains applies. It is readily confirmed that Eq. 4.5 reduces to Eq. 4.3 in this limit. Note that this saturation value is correct

for a continuous wormlike chain. For a discrete wormlike chain, which we will use for our simulations here, the saturation value shifts from the wormlike chain behavior $F \sim (1-z)^{-2}$ to the freely-jointed chain result $F \sim (1-z)^{-1}$ for sufficiently high forces [159].

Second, for small values of the extension, the leading-order behavior of Eq. 4.5 should reduce to Pincus's scaling result for weak stretching in the presence of excluded volume [101]. Pincus's theory is based on the existence of a tensile screening length $\xi_t = k_B T/f$ that competes with the Flory radius $R_F = L^{3/5} l_p^{1/5} w^{1/5}$ for a wormlike chain of contour length L [89]. The force-extension behavior can be obtained by a scaling argument where tensile blobs of size ξ_t contain a contour length

$$L_{blob} = \xi_t^{5/3} l_p^{-1/3} w^{-1/3} \tag{4.8}$$

The fractional extension is then given by $z = \xi_t / L_{blob}$, leading to [101]

$$F \sim z^{3/2} (l_p/w)^{1/2} \tag{4.9}$$

This is indeed the leading-order behavior of F_{low}. Since $F_{high} \sim O(z^2)$ for small z, and it is also the leading-order behavior of Eq. 4.5 for small z.

Third, F_{low} should exhibit a crossover from Pincus behavior to ideal wormlike chain behavior. The Pincus regime crosses over to the ideal scaling regime when the tensile blob size is commensurate with the thermal blob size, $\xi_t \cong l_p^2/w$ [100]. The crossover point is

$$F^* \cong w/l_p \tag{4.10}$$

with a corresponding fractional extension

$$z^* \cong w/l_p \tag{4.11}$$

This is indeed the fractional extension where the two quantities in the denominator of F_{low} are balanced.

Before moving on, we should note that Eq. 4.5 is not intended to be a model for freely-jointed chains. The force-extension interpolation behavior of freely-jointed chains has been addressed previously in a similar interpolation approach by Rahadkrishnan

and Underhill [160]. Equation 4.5 should not reduce to a freely-jointed chain model in the limit $l_p = w$ because the saturation behavior of the wormlike chain model is qualitatively different than a freely-jointed chain model [159]. For this reason, we only consider cases $l_p > w$ to test the EV-WLC formula.

4.3 Simulation Method

We obtained force-extension data for a discrete wormlike chain model [79, 83] using pruned-enriched Rosenbluth method (PERM) simulations [141]. Our simulations are the off-lattice analog of previous lattice simulations by Hsu and Binder [121]. The discrete wormlike chain model consists of a series of N inextensible bonds of length a. We use touching beads such that $a = w$, where w is the width of the chain. The contour length of the chain is thus $L = Nw = (N_b - 1)w$, where N_b is the number of beads. A bending energy (Eq. 2.6) is imposed between contiguous trios of beads. The bending energy κ is related to the persistence length by Eq. 2.7. Excluded volume interactions are treated by a hard core potential between non-contiguous beads (Eq. 2.8).

In each tour of the PERM simulations, the first bead is placed at the origin. Due to the translational invariance of the imposed force, this initial condition leads to no loss of generality. For the n^{th} chain growth step, we make $k = 5$ trial moves by selecting points on the unit sphere from the discrete wormlike chain distribution in the absence of excluded volume or the external force [81]. As a result, the reference state for our simulations is an ideal discrete wormlike chain at zero force. The j^{th} trial move to place the n^{th} bead at position $\mathbf{r}_n^{(j)}$ is assigned an atmosphere similar to Eq. 3.14,

$$a_n^{(j)} = \exp\left[-\beta\left(U_{\text{EV}} - \mathbf{f}\cdot\mathbf{r}_n^{(j)}\right)\right] \tag{4.12}$$

where U_{EV} is the excluded volume caused by placing this bead and $\mathbf{f}$ is the force. We then select one of the k trial moves with probability

$$p_n^{(j)} = \frac{a_n^{(j)}}{\omega_n} \tag{4.13}$$

where

$$\omega_n = \sum_{j=1}^{k} a_n^{(j)} \tag{4.14}$$

is the Rosenbluth weight for step n.

During a given tour, we track the cumulative weight of a configuration,

$$W_n = \prod_{i=0}^{n} \omega_n. \tag{4.15}$$

and enforce pruning and enriching steps via Grassberger's algorithm [141]. If at some step n a chain's cumulative weight is too high relative to the target weight, we "enrich" by generating a copy of the configuration and splitting the weight W_n between the two copies. Conversely, if at some step n the chain's cumulative weight is too low relative to the target weight, it is "pruned" and growth terminates at that step. In Grassberger's algorithm, the target weight is adjusted on-the-fly based on the current status of the simulation to improve sampling efficiency [141]. For each value of l_p/w, we conducted at least 10^5 tours so that the standard error of the mean, assumed to be the sampling error, is small compared to the symbol size in the plots. The details of the algorithm are elaborated in Chapter 3.

As a chain growth method, PERM naturally produces equilibrium data as a function of molecular weight. For a chain consisting of n steps, the average fractional extension is

$$z_n = \frac{\sum_t W_n^{(t)} z_n^{(t)}}{\sum_t W_n^{(t)}} \tag{4.16}$$

where $W_n^{(t)}$ is the cumulative weight of configuration t in the ensemble and $z_n^{(t)}$ is the corresponding extension of the configuration in that tour

$$z_n^{(t)} = \frac{\mathbf{r}_n^{(t)} \cdot \mathbf{f}}{(n-1)w} \tag{4.17}$$

with $\mathbf{r}_n^{(t)}$ the vector position of the n^{th} bead of configuration t and $(n-1)w$ is the contour length at step n. We run our simulations to sufficiently high molecular weights such that z becomes independent of n. The number of beads used for the data in this

chapter appear in Appendix A. As a result, we simply report the asymptotic value of z in what follows. Evidence in support of this claim is provided (see Appendix A).

In the course of our discussion, it will also prove useful to compute the excess free energy ΔF^{EV} caused by excluded volume. For this calculation, we repeat our simulations at a force f setting $U_{\mathrm{EV}} = 0$. In PERM, the free energy for growth out to step n relative to the reference state is

$$\beta F_n = -\ln\langle W_n\rangle \tag{4.18}$$

where the angle brackets indicate an average value. The excess free energy for chains grown out to step n is then given by [157]

$$\beta\Delta F_n^{\mathrm{EV}} = -\ln\frac{\langle W_n\rangle}{\langle W_n^{\mathrm{ideal}}\rangle} \tag{4.19}$$

where W_n^{ideal} is the cumulative weight from PERM simulations in the absence of excluded volume.

4.4 Results

We begin by comparing the simulation data we obtained for real discrete wormlike chains to the response of ideal continuous wormlike chains given by the Marko-Siggia interpolation formula in Eq. 4.3. As shown in Fig. 4.1, the force-extension behavior at high stretch is insensitive to the monomer anisotropy ratio l_{p}/w. However, the elastic behavior depends on the monomer anisotropy ratio at low forces. For a stiff chain, the deviation between the Marko-Siggia interpolation formula and the simulation data is small, even at rather small values of the fractional extension. For the more flexible chain, the deviation from the Marko-Siggia interpolation formulation is substantial and persists over a wide range of fractional extensions. In both cases, the EV-WLC interpolation formula reasonably captures both the saturation behavior at high forces and the deviation from the Marko-Siggia interpolation formula at low forces.

We included the Pincus scaling in Eq. 4.4 beside the low-force data in Fig. 4.1. The data do appear to follow Pincus scaling for sufficiently low forces, and we will address this issue in a quantitative manner shortly. For the moment, it suffices to note that the

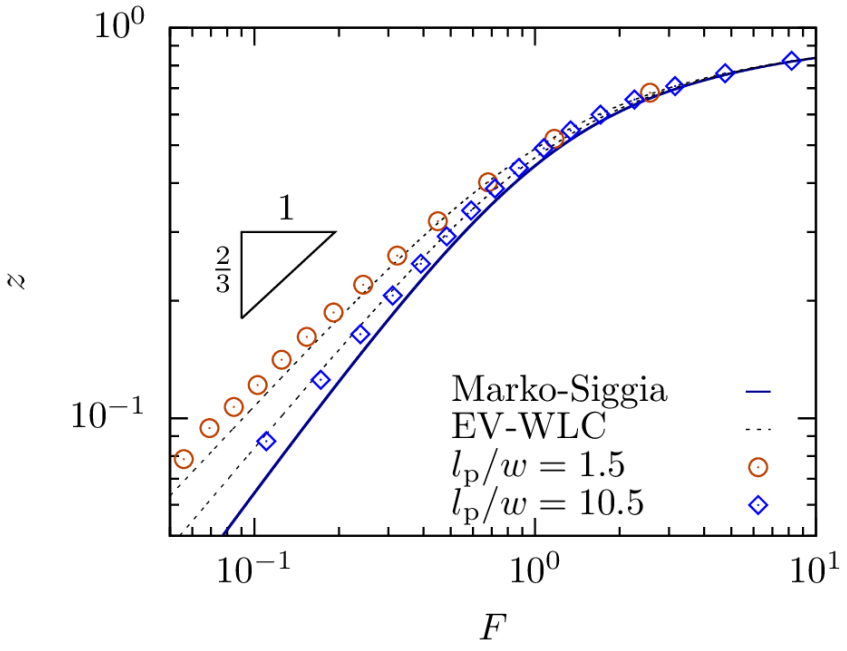

Figure 4.1: Comparison between discrete wormlike chain simulation data and the Marko-Siggia interpolation formula (solid line) in Eq. 4.3 and the EV-WLC interpolation formula (dashed lines) in Eq. 4.5 for a relatively flexible chain ($l_p/w = 1.5$) and a stiff chain ($l_p/w = 10.5$). The triangle indicates the Pincus scaling in Eq. 4.4. Similar plots for other values of l_p/w are provided (see Appendix A).

Pincus scaling is not a good description of the chain for all forces, which follows directly from its derivation [101]. As a result, we defer the error in the Pincus scaling to a later point, and focus for the moment exclusively on the Marko-Siggia interpolation formula and the EV-WLC interpolation formula.

The most important question to resolve, from a practical standpoint, is when the stretching of semiflexible chains should be modeled by the Marko-Siggia interpolation formula in Eq. 4.3 and when the EV-WLC interpolation formula in Eq. 4.5 provides a better description. To answer this question in a quantitative manner, we evaluated the error in these formulas for discrete wormlike chains as

$$\epsilon = \frac{|\tilde{z} - z|}{z} \tag{4.20}$$

with z being the value obtained from the simulation and $\tilde{z}$ being the value from the interpolation formulas in Eqs. 4.3 or 4.5. Naturally, the error is a function of the force. Figure 4.2 shows the error for the data in Fig. 4.1. As expected, the error in the Marko-Siggia formula increases as the force decreases due to excluded volume effects. Moreover, for the stiff chain with $l_{\mathrm{p}}/w = 10.5$, we see that the Marko-Siggia interpolation formula indeed only exhibits errors of a few percent once the excluded volume effects are suppressed at high forces. The error also increases for the EV-WLC formula as the force decreases, since the interpolation formula only approximately captures the crossover between Pincus scaling and the Hookean response.

The data in Fig. 4.2 also provide insight into modeling the stretching of double-stranded DNA, which is a very common model polymer whose monomer anisotropy in a high ionic strength buffer is similar to $l_{\mathrm{p}}/w = 10.5$ [83]. Our data for discrete wormlike chains support the use of the Marko-Siggia interpolation formula in models of double-stranded DNA in flow [150, 151]. However, it is worth noting that l_{p}/w decreases as the ionic strength decreases because the electrostatic interactions affect the persistence length and the width differently [83]. For very low ionic strengths, the EV-WLC interpolation formula may prove to be more accurate for double-stranded DNA than the Marko-Siggia formula.

We have obtained data at many different monomer anisotropies, and plots similar to Fig. 4.1 and Fig. 4.2 for these other values of l_{p}/w are provided (see Appendix A). Figure 4.3 summarizes the overall result, reporting the average error over all forces where we have obtained data. The average error for the EV-WLC interpolation is always smaller than the average error in the Marko-Siggia interpolation formula, independent of l_{p}/w. A closer inspection of the error as a function of force in Fig. 4.2 and the additional data provided (see Appendix A) shows that this trend persists for all values of the force except the saturation regime, where the two interpolation formulas are essentially the same. Thus, we expect that the EV-WLC interpolation formula in Eq. 4.5 will prove quite useful for modeling relatively flexible wormlike chains.

The error in both the Marko-Siggia and EV-WLC formulas both increase as the chain becomes more flexible. For the Marko-Siggia interpolation formula, we suspect that much of this error is due to a failure capture the low-force behavior, as Fig. 4.3 reports the average value of the error over all forces. For the EV-WLC formula, we previously

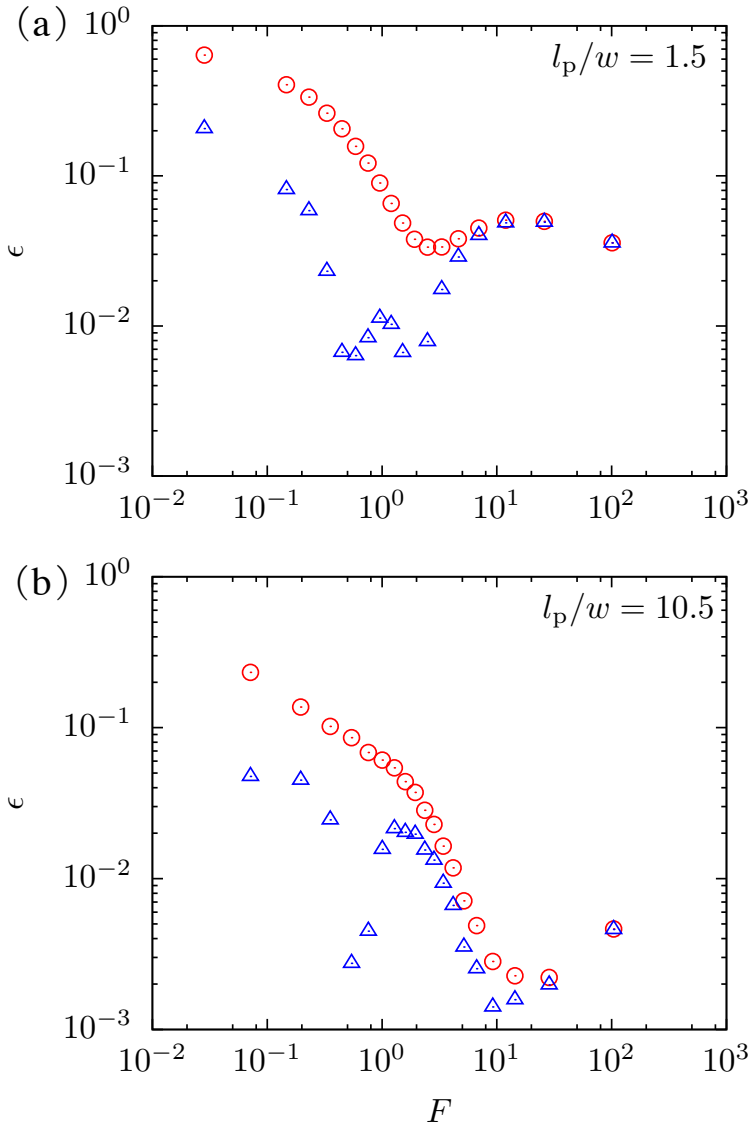

Figure 4.2: Plot of the error ϵ (Eq. 4.20) between simulation data and the Marko-Siggia interpolation formula (red circles) and the EV-WLC interpolation (blue triangles) as a function of dimensionless force F for (a) a relatively flexible chain ($l_p/w = 1.5$) and (b) a stiff chain ($l_p/w = 10.5$). Similar plots for other values of l_p/w are provided (see Appendix A).

proposed that the error arises primarily due to the approximate way that Eq. 4.5 treats the cross-over between the Pincus scaling and Hookean behavior. However, for both the

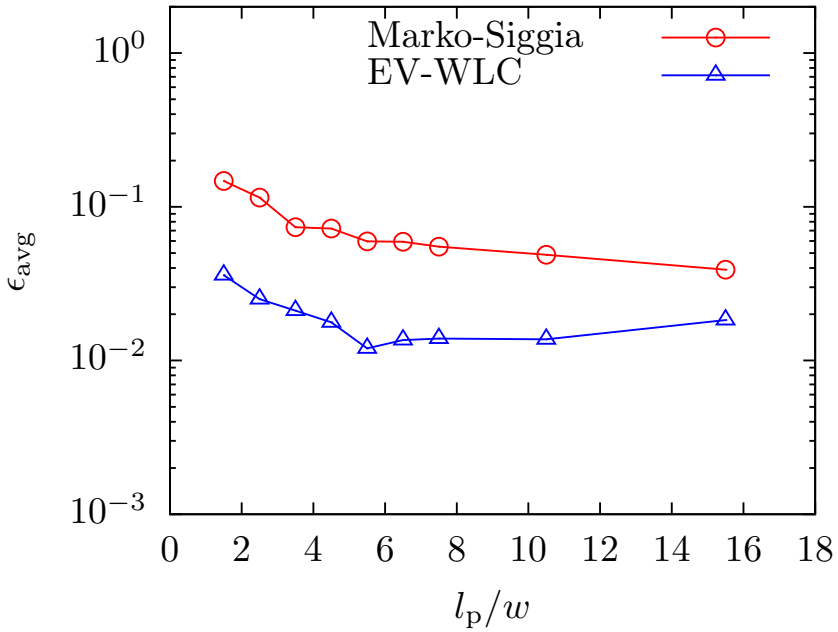

Figure 4.3: Average error, ϵ_{avg}, for the Marko-Siggia and EV-WLC interpolation formulas as a function of l_{p}/w.

Marko-Siggia and EV-WLC formulas, some of the error may also arise from the use of theories for continuous chains to describe data obtained from simulations of a discrete wormlike chains. Indeed, as l_p/w decreases, the discreteness of the model becomes increasingly important. For both interpolation formulas, the error in the interpolation formula increases as the discreteness of the model increases.

In the course of obtaining the force-extension data required to produce Fig. 4.3, we obtained a large amount of data that should correspond to the Pincus regime. Thus, it is worthwhile to take a moment to see whether our data are consistent with Eqs. 4.9 and 4.11 and to assess quantitatively the error between the Pincus force law and the simulation data. For this purpose, we also included data for a freely-jointed chain ($l_{\mathrm{p}} = w$) in Fig. 4.4. While the freely-jointed chain does not give the same limiting behavior as a wormlike chain at high extensions, it produces a Pincus regime. Figure 4.4a provides a rescaled force-extension plot demonstrating the collapse in the Pincus regime with a crossover corresponding to Eq. 4.11. To test the scaling in Eq. 4.9,

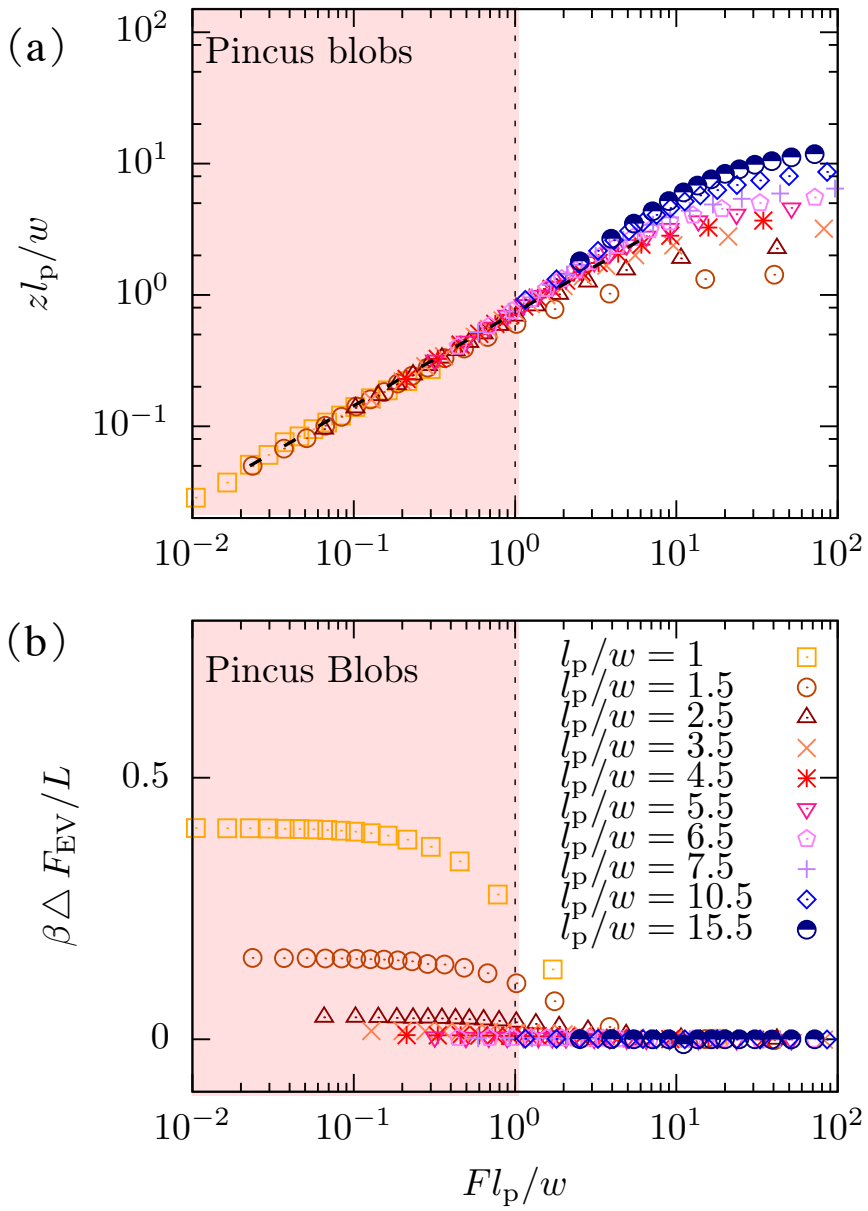

Figure 4.4: Plot of (a) the rescaled extension zl_p/w versus the rescaled force Fl_p/w and (b) excess free energy per unit length, $\beta\Delta F_{EV}/L$, for different values of l_p/w. The vertical dot-dashed line denotes the boundary of the (shaded) Pincus regime. The dashed line in panel (a) is the regression result to the Pincus regime. The symbols for different values of l_p/w are the same in panel (a) and (b).

we extracted the data corresponding to the Pincus regime and used linear regression to determine the prefactor and exponent for the scaling law. This analysis led to an exponent $F^{0.71}$. This exponent is consistent with Pincus's analysis using the Flory radius [89] $R_F = L^{\nu}(l_{\mathrm{p}}w)^{(1-\nu)/2}$, which leads to [100]

$$F \sim z^{\nu/(1-\nu)}(l_{\mathrm{p}}/w)^{(2\nu-1)/(1-\nu)} \tag{4.21}$$

The change in the exponent from $z \sim F^{2/3}$ in Eq. 4.9 to $z \sim F^{0.71}$ in Eq. 4.21 using $\nu = 0.587597$ as the Flory exponent [96] is identical to the case of the scaling law for the extension of semiflexible polymers confined in channels in the de Gennes regime [80]. For the data in Fig. 4.4 corresponding to the Pincus regime, the force-extension

$$F = 1.5z^{1.42}\left(\frac{l_p}{w}\right)^{0.42} \tag{4.22}$$

where the prefactor is obtained from the linear regression in Fig. 4.4, leads to an average error of 0.036. Figure 4.4b provides the corresponding values of the excess free energy due to excluded volume, demonstrating that the onset of excluded volume interactions is coincident with the Pincus scaling for the chain extension.

4.5 Discussion

The key result of this chapter is the development and evaluation of the EV-WLC interpolation formula in Eq. 4.5. Based on the error analysis in Fig. 4.3 it appears that the EV-WLC interpolation formula provides a good description of the force-extension behavior of wormlike chains. However, we need to be careful about extending the EV-WLC interpolation formula in Eq. 4.5 to polymers with isotropic monomers where l_{p} becomes very close to the chain width w. As pointed out by Dobrynin *et al.* [159], at very high forces, the saturation behavior switches from the wormlike chain result to the flexible chain result (i.e. the Langevin function) as the bending energy decreases. In principle, it should be possible incorporate this crossover for the saturation behavior for arbitrary stiffness κ into our EV-WLC interpolation formula, since Dobrynin *et al.* [159] have already determined how to interpolate between the flexible and wormlike stretching for ideal chains.

The EV-WLC interpolation formula is most useful for modest ratios of l_p/w, and these values characterize a number of important polymer systems. Single-stranded DNA is a polymer with enormous biological relevance that exhibits nearly isotropic monomers. There is a growing experimental interest in using single-stranded DNA as a model polymer [152, 161]. Using biochemical synthesis methods, single-stranded DNA molecules with $\approx 10^4$ bases containing designer sequences with minimal base paring can readily be synthesized and uniformly labeled with fluorescent dyes, thereby enabling the direct visualization of single chain dynamics using fluorescence microscopy [161]. Due to the very small persistence length of single-stranded DNA ($l_p \approx$ 1-2 nm under modest salt concentrations) [153], single-stranded DNA chains with contour lengths $L \approx$ 15-20 μm correspond to $N_K \approx$ 7,500-10,000 Kuhn segments compared to only $N_K \approx$ 150-190 for double-stranded DNA of similar contour length. The ability to study single chain dynamics of long chain, highly flexible polymers opens a new window into observing non-linear phenomena and chain dynamics in flow, which are heavily influenced by dominant EV and intramolecular hydrodynamic interactions [162]. From this perspective, the non-equilibrium flow dynamics of highly flexible polymers such as single-stranded DNA is expected to differ *qualitatively* compared to linear λ-DNA of similar contour length L. Our enthusiasm towards using the EV-WLC model for single-stranded DNA is tempered by the possibility that the results could be affected by torsional constraints, which are included in some coarse-grained models, such as the 3-SPN model [163], but not in others, such as OX-DNA [164]. Ultimately, the importance (or lack thereof) of torsion on the force-extension behavior of ssDNA is a question that needs to be resolved experimentally. Our EV-WLC formula provides a framework for addressing this question, since it assumes no torsional potential.

On the chemistry side, a broad class of synthetic polymers would also be described by the EV-WLC interpolation formula. In particular, we anticipate that the EV-WLC formula will describe the elastic behavior of synthetic polymers that have bulky side groups but do not form helical structures, thereby maintaining modest values of l_p/w. In many ways, we do not yet know which polymers will be described by the EV-WLC formula because the low-force elasticity has not yet been rigorously investigated for most synthetic polymers using single molecule force spectroscopy. Whereas AFM can faithfully measure the high-force elasticity of single polymers, magnetic tweezers are

one method capable of interrogating the low-force regime; however, this approach has only been applied to a handful of polymers such as single-stranded DNA [153] and poly(ethylene glycol) [154].

Poly(ethylene glycol) or PEG presents an interesting case in the context of developing interpolation formulas for chain elasticity. Gaub and coworkers [165] performed AFM measurements on PEG and observed that the polymer forms water-mediated superstructures in aqueous solutions. As a result, and perhaps unexpectedly, PEG is well described by the Marko-Siggia force relation (Eq. 4.3) in the limit of high forces in water. On the other hand, stretching PEG in an aprotic solvent (hexadecane) resulted in force-extension curves that were well fit by the inverse Langevin function (Eq. 4.2), which is characteristic of stretching a flexible polymer in a theta solvent in the absence of EV interactions. Recently, magnetic tweezers were used to probe the force-extension behavior of PEG in the low force regime [154], which revealed that PEG exhibits both the Pincus and Hookean regimes in aqueous solutions. However, the Pincus regime only survives up to very small extensions $z \approx 0.06$, perhaps due to local rigidification of the polymer backbone due to the formation of superstructures in aqueous solution. From this view, it is clear that the existence of solvent-polymer interactions for PEG results in an increase in monomer rigidity and somewhat unexpected behavior. Overall, the lessons from these results clearly illustrate that the details of the chemistry, solvent interactions, and local molecular structure are key to determining the emergent force-extension behavior for any macromolecule.

The availability of the EV-WLC interpolation formula opens up a new avenue for coarse-grained modeling of such polymers in flow. In typical bead-spring models, a polymer chain is described by a series of beads (friction points) connected by massless springs. The entropic penalty for stretching the chain is captured by a spring force, while enthalpic effects arising from intramolecular excluded volume interactions are imposed by a pairwise potential between beads. An alternate approach is to use Eq. 4.5 to simultaneously capture the effects of stretching and the internal excluded volume interactions due to the subchain represented by the spring. We envision that such a model could prove very useful for modeling the dynamics of such polymers in flow.

While our primary emphasis in this chapter is the development and testing of the

EV-WLC interpolation formula, the results we have obtained for stretching in the Pincus regime should also be viewed in light of the existing simulation and experimental literature. From the simulation side, our data are part of a growing body of literature [100,121,155] demonstrating the existence of the Pincus regime, that is, a low-force non-linear elasticity for polymers in a good solvent. Our key contributions in this respect are methodological, showing that off-lattice PERM simulations of a discrete wormlike chain model can reach sufficiently high molecular weights to observe Pincus scaling even for rather stiff chains, and in the thermodynamics of the Pincus regime, with Fig. 4.4b clearly demonstrating that the Pincus regime exists due to excluded volume interactions. From the experimental side, moving forward, it will be worthwhile to see how our data relate to the force-extension properties of single-stranded DNA [153] and new classes of synthetic polymers that can be studied using magnetic tweezers [154].

4.6 Concluding Remarks

In this chapter, we have shown that an interpolation formula that incorporates excluded volume interactions leads to more accurate predictions of the force-extension behavior of discrete wormlike chains than the classic Marko-Siggia interpolation formula, which was developed for ideal continuous wormlike chains. The EV-WLC interpolation formula will prove particularly important for polymers with relatively small range of monomer anisotropies l_{p}/w, as these values characterize many important experimental systems such as single-stranded DNA and synthetic polymers that contain bulky side groups but do not form helices. We anticipate that the EV-WLC interpolation formula will prove useful as a model for the force-extension behavior of such polymers as such experimental data become available [153, 154]. Even more importantly, we expect that the EV-WLC interpolation formula will provide a quantitatively accurate force law for coarse-grained simulations of these polymers in flow.

Chapter 5

Effect of Excluded Volume on the Force-extension of Wormlike Chains in Slit Confinement

This chapter is based on the publication
X. Li and K. D. Dorfman, "Effect of excluded volume on the force-extension of wormlike chains in slit confinement"
J. Chem. Phys., vol. 144, p. 104902, 2016 [166]

5.1 Introduction

Recent interest in the theory [114,121] and simulation [115,121,122] of the stretching of wormlike chains in slit-like confinement has been motivated in part by experiments where a DNA molecule spans a nanoslit connecting either two large reservoirs [117,119,120] or a number of discrete nanopits [105, 123–126]. The difference in configurational entropy arising from the difference in mass of relaxed DNA in each of the reservoirs (or nanopits) exerts a stretching force on the intervening portion of the chain in the slit. If this stretching force is strong, then excluded volume interactions between different segments of the DNA chain inside the slit are minimal. In this case, it is appropriate to treat the problem as the stretching of a confined, ideal chain in an external force [115, 122].

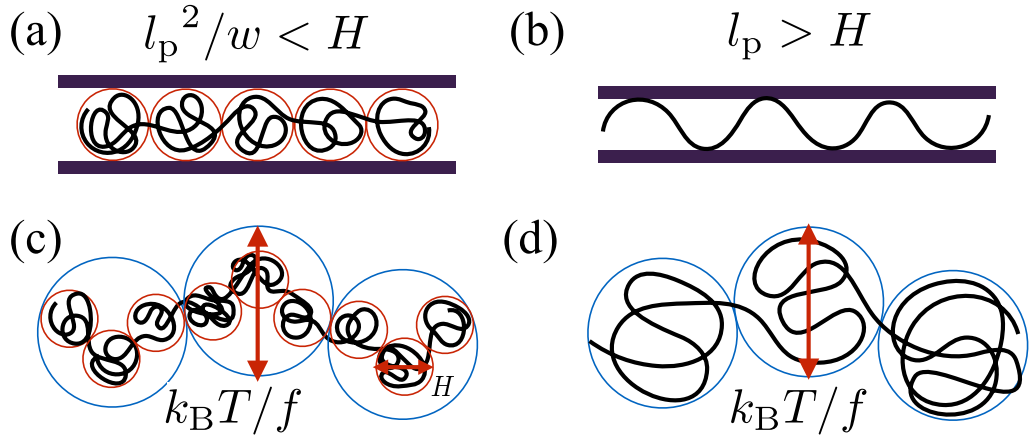

Figure 5.1: Schematic illustration of stretching in slit confinement. In the absence of a force, in weak confinement (the de Gennes regime) the chain is organized into a series of blobs (a) while for strong confinement (the Odijk regime), the chain is organized into a series of deflection segments (b). In the presence of a weak force and confinement, the blobs (c) or the deflection segments (d) are organized in the plane into a series of Pincus blobs.

However, as the stretching force decreases, excluded volume interactions between parts of the chain in the slit become important. Efficient modeling of these experiments requires knowing when theories developed for the stretching of ideal chains [115, 122] can be safely applied and, if necessary, how to modify the model to account for excluded volume [114, 121].

In the present chapter, we use pruned-enriched Rosenbluth Method (PERM) simulations [141] to test the scaling arguments originally set forth by Taloni *et al.* [114] for the role of excluded volume during the force-extension of real wormlike chains. Explicitly, we consider chains of contour length L, persistence length l_p, and width w confined in a slit of height H that are stretched by an external force f. Figure 5.1 summarizes the key idea behind their scaling theory [114], which merges the ideas underlying Odijk's theory [167] for confinement in slits in the absence of a force with Pincus's theory [101] for the stretching of real wormlike chains in the absence of confinement. In the absence of a force, weak confinement leads to the formation of a "chain of blobs" (the de Gennes regime, Fig. 5.1(a)) while strong confinement leads to the formation of a "chain

of deflection" segments (the Odijk regime, Fig. 5.1(b)). In the presence of a weak force, the chain of blobs (Fig. 5.1(c)) or chain of deflection segments (Fig. 5.1(d)) is arranged in the plane into a series of Pincus (tensile) blobs of size $\xi_T = k_B T/f$, where $k_B T$ is the Boltzmann factor. The latter idea is analogous to Pincus's theory [101] for the force-extension of real wormlike chains in free solution — the key difference is that the appropriate coarse-graining of the chain in confinement is blobs or deflection segments, depending on the strength of confinement, while the equivalent coarse-graining in free solution is in terms of Kuhn lengths.

We refer to the chain configurations in Fig. 5.1(c) and Fig. 5.1(d) as the "confined Pincus regime." Our main goal here is to test the proposed scaling laws [114] for the force-extension in the confined Pincus regime, as well as the proposed scaling laws for upper and lower boundaries for this regime in the force-molecular weight phase plane. In the course of this analysis, we further show that the force-extension behavior of the chain for forces exceeding the upper bound of the confined Pincus regime is well described by models developed for unconfined wormlike chains [70, 101, 168]. Taken together, our analysis provides a quantitatively accurate model for the force-extension behavior of long ($L \gg l_p$) wormlike chains in strong and weak confinement and phase diagrams for understanding the different regimes of stretching in slit confinement. We also consider briefly the effects of finite molecular weight and moderate confinement, as these are experimentally relevant scenarios.

5.2 Scaling Theory

We begin by recalling briefly the key results from the scaling theory by Taloni *et al.* [114] and a modification to their theory for strong confinement. The details of the derivation are provided (see Appendix B). The key idea is balancing the tension blob size $\xi_T = k_B T/f$ with the size $R = L^{3/4}(l_p w/H)^{1/4}$ of a slit-confined chain in the absence of a force. This analysis (see Appendix B) yields

$$z \cong \left(\frac{f l_{p,\parallel}}{k_B T}\right)^{1/3} \left(\frac{w}{H}\right)^{1/3}, \tag{5.1}$$

where $z \equiv Z/L$ is the fractional extension of the chain in the direction of the force. In the latter, $l_{p,\parallel}$ is the effective persistence length, which is obtained from the characteristic decay of the bond vector autocorrelation function in confinement. While there are a number of quantitative reports [2, 110, 122, 169] of the effective persistence length as a function of slit height and chain stiffness, we focus here primarily on weak confinement, $H > {l_p}^2/w$, and strong confinement, $H < l_p$. As a result, we simply use the three-dimensional result $l_{p,\parallel} = l_p$ in weak confinement and the two-dimensional result $l_{p,\parallel} = 2l_p$ in strong confinement [169]. We also provide data on the role of $l_{p,\parallel}$ as part of the Appendix B.

The lower bound for the confined Pincus regime corresponds to short chains that cannot form many Pincus blobs [114]. At the scaling level, the minimum contour length is the molecular weight of a Pincus blob (see Appendix B),

$$L_{\min} \sim l_{p,\parallel} \left(\frac{k_B T}{f l_{p,\parallel}} \right)^{4/3} \left(\frac{H}{w} \right)^{1/3}. \tag{5.2}$$

The upper bound of the confined Pincus regime is a more subtle point, depending on the degree of confinement but not the contour length. Taloni *et al.* [114] proposed that the confined Pincus regime ends when the Pincus blob size is equal to the slit height, leading to (see Appendix B)

$$f_{\max}^{\text{weak}} \cong \frac{k_B T}{H} \tag{5.3}$$

and, with Eq. 5.1, a maximum fractional extension (see Appendix B)

$$z_{\max}^{\text{weak}} \cong \left(\frac{l_{p,\parallel} w}{H^2} \right)^{1/3}. \tag{5.4}$$

We posit that this result only applies in weak forces. For strong forces, the appropriate upper bound is when the contour length in a Pincus blob is shorter than the minimum contour length for real chain behavior in Odijk's theory for confined chains (see Appendix B),

$$f_{\max}^{\text{strong}} \cong \frac{k_B T}{l_{p,\parallel}} \left(\frac{w}{H} \right)^{1/2}. \tag{5.5}$$

The corresponding fractional extension is (see Appendix B)

$$z_{\max}^{\text{strong}} \cong \left(\frac{w}{H}\right)^{1/2}. \tag{5.6}$$

The detailed rationale for our conjecture is provided in Appendix B.

Inasmuch as Eq. 5.1 arises from balancing the size of a confined subchain with the tension blob size (see Appendix B), it is valid for all regimes of slit-like confinement provided that the appropriate value of $l_{\text{p},\parallel}$ is used. For example, in the extended de Gennes regime [2, 80] corresponding to $2l_{\text{p}} < H < l_{\text{p}}^2/w$, we would expect that Eq. 5.1 will apply with $l_{\text{p},\parallel} \approx l_{\text{p}}$ [2]. The key challenge in addressing this regime of intermediate confinement is the location of the crossover. The characteristic channel in the extended de Gennes regime of confinement is $O(l_{\text{p}}^2/w)$, whereupon Eqs. 5.3 and 5.5 are identical at the accuracy available for a scaling law. However, since this channel size represents the upper bound for the extended de Gennes regime, we expect that the crossover for strong confinement in Eq. 5.5 will be the more accurate model.

5.3 Simulation Method

To test the predictions of this scaling theory, we model the confined chain with the discrete, off-lattice wormlike chain model used in our simulations that has been elaborated in Chapter 2 and Chapter 3. The chain consists of a series of N inextensible bonds of length equal to the backbone width w of the chain, corresponding to touching beads. The contour length of the chain is thus $L = Nw = (N_{\text{b}} - 1)w$, where N_{b} is the number of beads and N is the number of bonds. Such a fine-grained model is required to capture the sub-persistence length behavior, which is crucial for strong confinement.

A bending energy is introduced between consecutive bonds as Eq. 2.6. The bending parameter κ is related to the persistence length in free solution by Eq. 2.7. Excluded volume interactions are treated by a hard core potential for real chains (Eq. 2.8). Similar to the excluded volume, chains can also interact with slit via a hard core potential

$$\beta U_{\text{EV}} = \begin{cases} \infty, & |r_{i,wall}| > H/2, \\ 0, & |r_{i,wall}| \le H/2. \end{cases} \tag{5.7}$$

Note that the parameter H is the effective slit height $H_{\text{wall}} - w$, using w as the excluded volume distance for chain-wall interactions for two walls separated by a distance H_{wall}. Similar to our prior work [168], the force is implemented as an external potential in the model. In the simulation, we fix one end of the chain while applying a force on the growing end. We have described how to implement the force in our simulation elsewhere [168].

We obtained the force-extension data in a slit using pruned-enriched Rosenbluth method (PERM) simulations [141]. We run our simulations to sufficiently high molecular weights such that the extension becomes independent of N. The number of beads for the long chain data in this chapter appear in the Appendix B.

We focus primarily on the two qualitatively different regimes of confinement illustrated in Fig. 5.1: weak confinement ($l_{\text{p}}{}^2/w < H$), corresponding to the de Gennes regime, and strong confinement ($l_{\text{p}} > H$), corresponding to the Odijk regime, with a brief diversion to demonstrate the confined Pincus regime for the extended de Gennes regime ($2l_{\text{p}} < H < l_{\text{p}}^2/w$). The latter analysis focuses on asymptotically long chains, where the scaling theory is expected to be valid. At the conclusion of our analysis, we consider typical experimental molecular weights for DNA (λ-phage and T4 DNA) in a 25 nm slit, where finite-size effects may be more pronounced.

5.4 Results

5.4.1 Confined Pincus Regime

We begin our analysis by testing the prediction for the scaling of the extension with force in the confined Pincus regime (Eq. 5.1). As seen in Fig. 5.2(a) and 5.2(d), which correspond to strong confinement (a) and weak confinement (d) with their expected crossover points, the scaling law in Eq. 5.1 collapses the data for both weak and strong confinement inside the confined Pincus regime (shaded area). Moreover, using an appropriate value of $l_{\text{p},\parallel}$, the data for weak and strong confinement also collapse onto a single master curve in the confined Pincus regime. The data at weak forces is noisy, due to strong thermal fluctuations, and thus difficult to converge with our computational resources.

To be more quantitative in this conclusion, we extracted the data corresponding

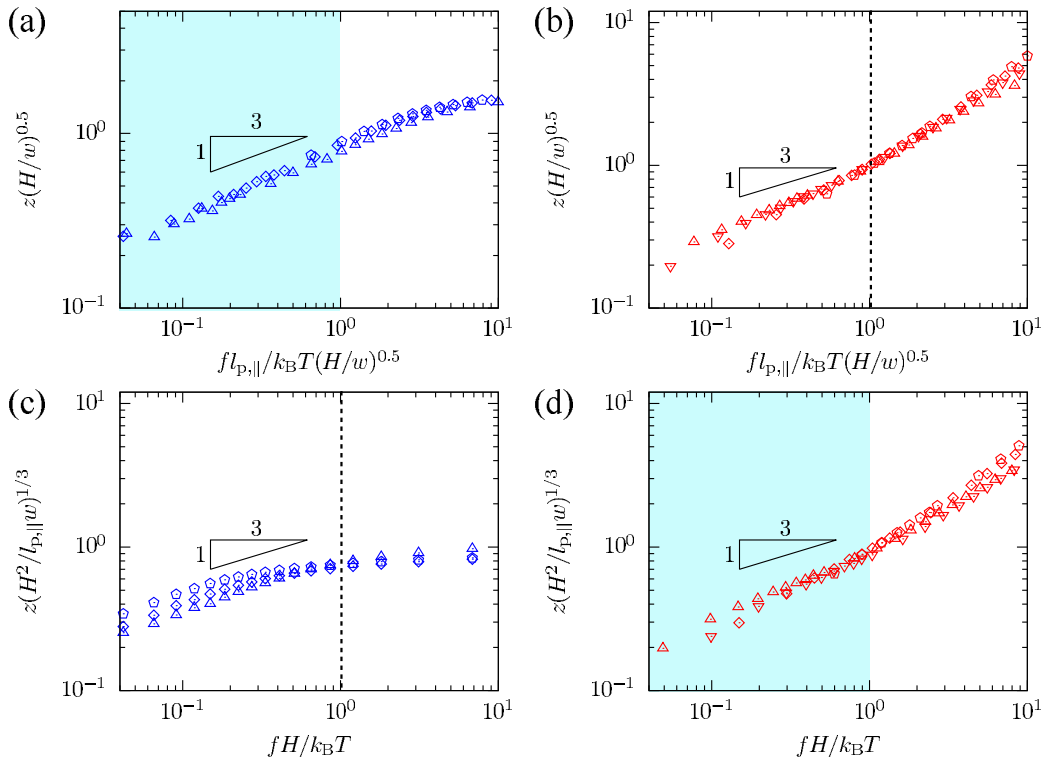

Figure 5.2: Plot of the rescaled extension versus the rescaled force for different values of l_p/w and H. The rescaled axes are chosen to test the maximum force and extension for strong confinement (Eqs. 5.5 and 5.6) and weak confinement (Eqs. 5.3 and 5.4). The panels correspond to (a) data for strong confinement using the crossovers for strong confinement, (b) data for weak confinement using the crossovers for strong confinement, (c) data for strong confinement using crossovers for weak confinement, and (d) data for weak confinement using crossovers for weak confinement. In each case, the shaded area or dashed line shows the confined Pincus regime predicted by these scaling arguments. Blue symbols are simulation results for strong confinement with triangle: $l_p/w = 5.5$, $H/w = 4$; diamond: $l_p/w = 10.5$, $H/w = 4$; pentagon: $l_p/w = 15.5$, $H/w = 4$. Red symbols are simulation results for weak confinement with triangle: $l_p/w = 5.5$, $H/w = 49$; inverted triangle: $l_p/w = 5.5$, $H/w = 99$; diamond: $l_p/w = 10.5$, $H/w = 149$; pentagon: $l_p/w = 15.5$, $H/w = 299$. The black triangle shows the scaling law in Eq. 5.1.

to the scaling $z \sim f^{1/3}$ in the confined Pincus regime and used linear regression to determine the prefactor and exponent for the scaling law. Not only are the exponents similar (0.3916 ± 0.0167 for the Odijk regime and 0.3861 ± 0.0998 for the de Gennes regime), but the prefactors are almost identical (0.8246 for the Odijk regime and 0.8385

for the de Gennes regime) provided we use the effective persistence length in strong confinement. As a result, Eq. 5.1 appears to be an excellent description of the data in the confined Pincus regime independent of the degree of confinement. Repeating the derivation of the confined Pincus regime scaling laws using $\nu = 0.588$ as the Flory exponent in 3D and $\nu = 0.75$ as Flory exponent in 2D [95] furnishes

$$z \sim \left(\frac{f l_{\mathrm{p},\parallel}}{k_{\mathrm{B}}T}\right)^{1/3} H^{-0.367} w^{0.299} l_{\mathrm{p},\parallel}^{0.068}. \tag{5.8}$$

Repeating the linear regression of the data in Figs. 5.2(a) and 5.2(d) using Eq. 5.8 yields exponents of 0.3430 ± 0.0592 for strong confinement and 0.3557 ± 0.0134 for weak confinement, which are closer to the expected value of 1/3.

The particular rescaled axes in Fig. 5.2 were selected to test the upper bound for the confined Pincus regime, with the shading in Fig. 5.2 assuming a prefactor of unity for the cross-over points. Inasmuch as the scaling laws appear to extend beyond the shading region, this suggests a different, but still $O(1)$, prefactor. Moreover, the prefactors for the crossover appear to be different for weak and strong confinement. We also see that the proposed criteria for the crossovers in strong confinement (Eqs. 5.5 and 5.6) in Fig. 5.2(a) indeed capture the crossover in the data strong confinement, where as the crossovers in weak confinement (Eqs. 5.3 and 5.4, Fig. 5.2(c)) are not a good description. In contrast, the data in Figs. 5.2(b) and (d) do not allow us to make a similar conclusion about the data for weak confinement, as both the cutoffs for weak confinement (Eqs. 5.3 and 5.4, Fig. 5.2(d)) and strong confinement (Eqs. 5.5 and 5.6, Fig. 5.2(c)) appear to provide similar results. The reason for this ambiguity is that our particular slit sizes satisfy the inequality $H > l_{\mathrm{p}}^2/w$ but do not satisfy the strong inequality $H \gg l_{\mathrm{p}}^2/w$. As a result, the maximum forces given by Eqs. 5.3 and 5.5 are similar.

It is quite challenging to simulate the stretching of long chains in the de Gennes regime using sub-persistence length resolution, since the number of beads required to reach the long-chain limit becomes prohibitive. However, such long chains in wide slits are required to test the crossover points for $f_{\max}^{\mathrm{weak}}$ and $z_{\max}^{\mathrm{weak}}$ in Eqs. 5.3 and 5.4. Fortunately, the de Gennes regime is valid for both semiflexible chains and freely-jointed chains, as the only restriction on the theory is that the subchain within a blob be a

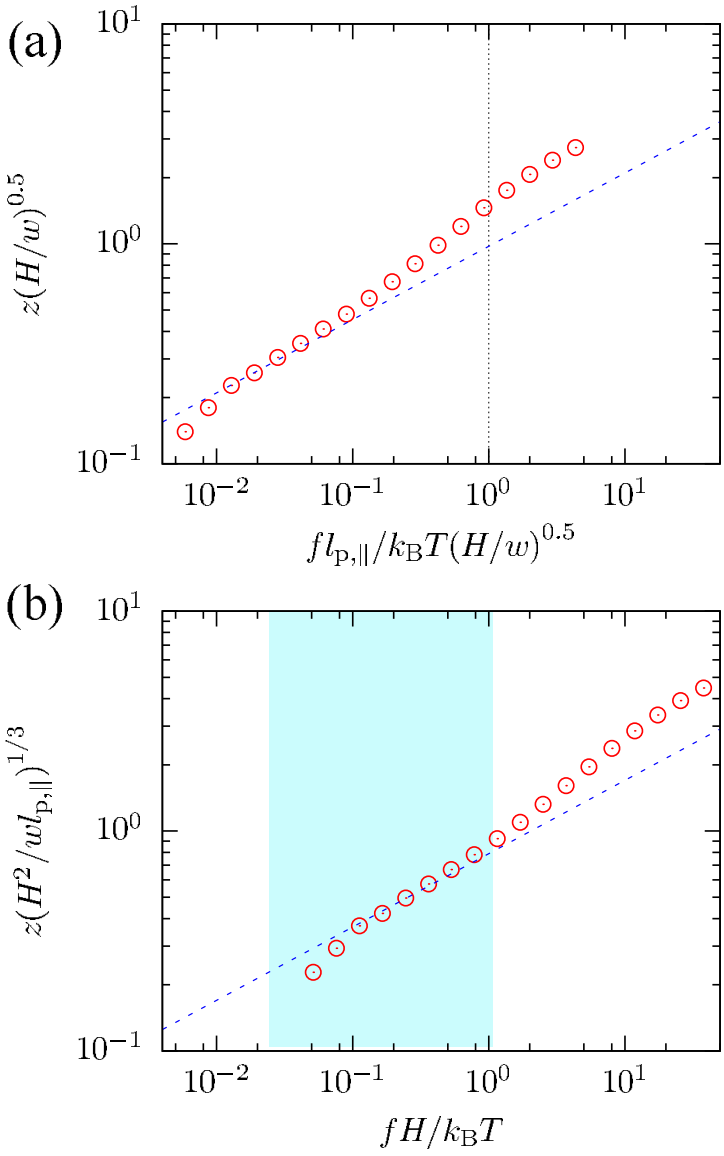

Figure 5.3: Plot of the rescaled extension versus the rescaled force for a freely-jointed chain (l_{p}/w = 0.5) in weak confinement (H/w = 19). The rescaled axes are chosen to test the maximum force and extension given by (a) Eqs. 5.5 and 5.6 and (b) Eqs. 5.3 and 5.4. The triangle indicates the scaling for the confined Pincus regime. The dashed line in (a) shows the crossover given by Eq. 5.5. The shaded area in (b) shows the confined Pincus regime given by Eq. 5.3.

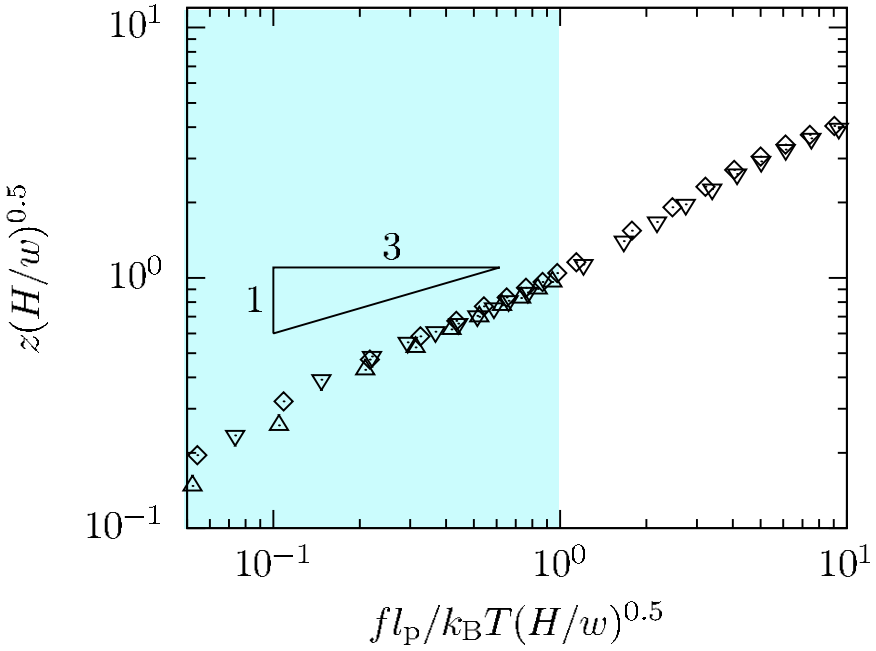

Figure 5.4: Plot of the rescaled extension versus the rescaled force for different values of l_p/w and H. The rescaled axes are chosen to test the strong confinement crossovers given by Eqs. 5.5 and 5.6. The symbols are simulation results for intermediate confinement with triangle: l_p/w = 10.5, H/w = 99; inverted triangle: l_p/w = 10.5, H/w = 49 and diamond: l_p/w = 14.5, H/w = 49. The black triangle shows the scaling law in Eq. 5.1. The shaded area shows the confined Pincus regime.

self-avoiding random walk [109]. We thus simulated force-extension behavior for a freely-jointed chain (i.e., $U_{\rm bend}$ = 0 in Eq. 2.6) in a wide slit H/w = 19, corresponding to weak confinement, and we were able to reach the long-chain limit for most forces. This channel has a wider separation between the two possible upper boundaries for the force, with $f_{\rm max}^{\rm weak}$ almost an order of magnitude smaller than $f_{\rm max}^{\rm strong}$. Figure 5.3(b) suggests that the scaling $z \sim f^{1/3}$ ends at the predicted crossover given by Eqs. 5.3 and 5.4 for very weak confinement — there is an uptick in the exponent for $f > k_BT/H$. In contrast, the scaling of the extension with the force in Fig. 5.3(a) changes well before the cutoff given by Eq. 5.5. In both panels of Fig. 5.3, note that at very small forces the force-extension behavior deviates from the confined Pincus regime because the chains are too short.

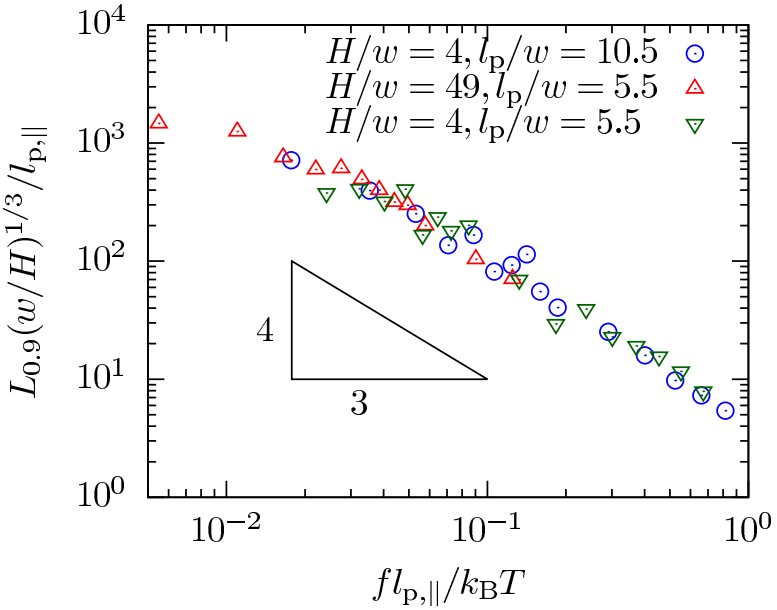

Figure 5.5: Rescaled contour length for the lower bound of the confined Pincus regime as a function of the rescaled force for three chains with different stiffness in different confinement. The contour length $L_{0.9}$ is the contour length at which the fractional extension reaches 90% of its saturation value for a given value of the force. The triangle shows the scaling in Eq. 5.2.

We have also considered the case of intermediate confinement for slit heights $2l_p < H < l_p^2/w$, which correspond to the extended de Gennes regime [2, 80]. Figure 5.4 confirms the existence of a confined Pincus regime in the extended de Gennes regime, with the crossover being captured by the strong confinement criteria (Eqs. 5.5 and 5.6). The corresponding plot for the weak confinement criteria (Eqs. 5.3 and 5.4), which also work reasonable well, is provided in Appendix B. The ambiguity in the selection of the crossover point is consistent with our previous discussion; for these channel sizes, both the strong confinement and weak confinement results are essentially the same since the chain only has modest monomer anisotropy.

We next turn our attention to the lower bound for the Pincus regime given by Eq. 5.2, i.e. the chain length sufficient to begin to form Pincus blobs. The onset of the confined Pincus scaling is a gradual one, so the transition point is not well defined. However, we know that the fractional extension along the direction of the force vanishes at zero

force because the chain configuration is isotropic in the plane in the absence of a force and eventually saturates to some value of z for large forces. We arbitrarily selected the contour length $L_{0.9}$ at which the fractional extension reached 90% of its saturation value at a given force as the onset of the confined Pincus regime. A detailed explanation of the data analysis protocol is provided in Appendix B.

Figure 5.5 shows that Eq. 5.2 is an accurate description for the scaling law for the lower bound of the confined Pincus regime, collapsing data in both strong confinement and weak confinement. While it is possible to make a quantitative analysis similar to what we did for Fig. 5.2, the arbitrariness of using 90% of the fractional extension given by Eq. 5.1 to define the onset of the Pincus regime makes any further quantitative analysis less insightful.

5.4.2 Stronger Forces

While our primary goal is to test the scaling theory for the confined Pincus regime [114], establishing the upper boundaries for this regime requires simulating chains at higher forces. As such, our simulations also provide insights into chain stretching beyond the confined Pincus regime. In particular, we already see in Fig. 5.2 that the data for strong and weak confinement differ qualitatively after exiting the confined Pincus regime. As a result, we consider these two cases separately.

Figure 5.6 shows the strong confinement behavior for an asymptotically long chain with properties similar to DNA. After exiting the confined Pincus regime, the force-extension data for the real chain is almost identical to simulation data obtained from companion simulations of a confined *ideal* chain, where the intra-chain excluded volume was turned off. After the end of the confined Pincus regime, the fractional extension of the real chain and ideal chain are indistinguishable. Moreover, we found that the two-dimensional Marko-Siggia interpolation formula,

$$\frac{f l_{\mathrm{p},\parallel}}{k_{\mathrm{B}}T} = \frac{3}{4}z + \frac{1}{8(1-z)^2} - \frac{1}{8}, \tag{5.9}$$

is a reasonably good approximation for the extension at high forces.

The behavior at strong forces in weak confinement is qualitatively different than that in strong confinement. Figure 5.7 illustrates this point using a relatively flexible

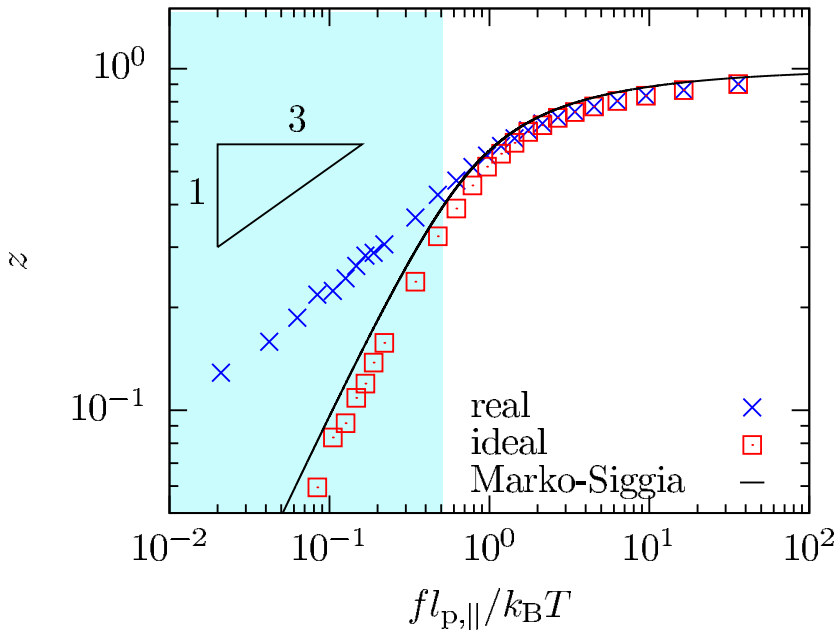

Figure 5.6: Plot of the fractional extension versus the rescaled force for strong confinement for $l_p/w = 10.5$ and $H/w = 4$. The solid black line is Eq. 5.9. The shaded area shows the confined Pincus regime.

chain in a very wide slit. For weak forces, we see the confined Pincus scaling law in Eq. 5.1. At very small forces there is a deviation from the confined Pincus scaling towards an apparently larger exponent. This is not a deviation from confined Pincus behavior, but rather a finite length effect. For such small forces and large slits, extremely long chains of $O(10^5)$ beads are needed in order to enter the confined Pincus regime, exceeding our computational resources. At the highest forces, we see that (i) the real chain and ideal chain data are again indistinguishable but (ii) now the extension data in weak confinement agree well with the Marko-Siggia interpolation formula in three dimensions,

$$\frac{fl_{p,\|}}{k_BT} = z + \frac{1}{4(1-z)^2} - \frac{1}{4}. \tag{5.10}$$

Comparing the high-force data in Fig. 5.6 to those in Fig. 5.7, it appears that the three-dimensional Marko-Siggia formula in Eq. 5.10 does a better job modeling the data

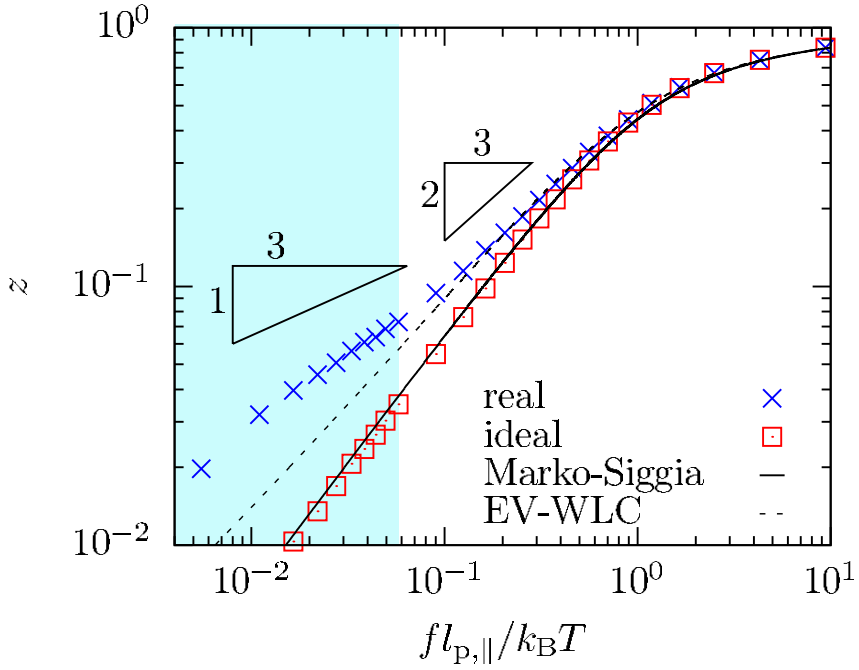

Figure 5.7: Plot of the fractional extension versus rescaled force for weak confinement for l_{p}/w = 5.5 and H/w = 99. The solid black line is Eq. 5.10 and the dashed black line is Eq. 5.11. The shaded area shows the confined Pincus regime. A similar plot for l_{p}/w = 5.5 and H/w = 49 is included (see Appendix B) to further support the confined Pincus regime scaling.

for weak confinement than the two-dimensional Marko-Siggia formula in Eq. 5.9 does for strong confinement. The reason lies in the nature of the confinement when the chain is almost completely stretched in the direction of the force. In weak confinement, the highly stretched chain is effectively unconfined. While those configurations that are near the wall may experience some residual effects of confinement, most configurations are not near the wall and thus deflect about the axis of the force in exactly the same way as in the absence of confinement. In contrast, the chain in strong confinement is not literally confined to a plane. Rather, long-wavelength deflections are cut-off by the walls [122]. As a result, the Marko-Siggia formula for 2D confinement overestimates the stretching since it assumes no deflections out of the plane. If we relax our approximation $l_{\mathrm{p},\parallel} = 2l_{\mathrm{p}}$ for strong confinement and use a more realistic value, the agreement between

the Marko-Siggia formula for 2D and the simulation data is improved. A plot showing how the result is affected by using a $l_{p,\parallel}$ as a function of H is provided (see Appendix B).

In contrast to the strong confinement case, there appears to be a second power law with scaling $z \sim f^{2/3}$, which is consistent with the Pincus regime in free solution [101]. However, the range of forces over which $z \sim f^{2/3}$ is rather narrow. To demonstrate that the region $z \sim f^{2/3}$ in Fig. 5.7 is indeed the same as the free-solution Pincus behavior, we developed two additional pieces of evidence. The first supporting data are already in Fig. 5.7. In Chapter 4 [168], we developed an interpolation formula

$$\frac{f l_{p,\parallel}}{k_B T} = \frac{z^{1.5}}{0.21(w/l_{p,\parallel})^{1/2} + (2/3)z^{1/2}} + \frac{1}{4(1-z)^2} - \frac{1}{4} - \frac{z}{2} \tag{5.11}$$

for the force-extension of wormlike chains in free solution. This "EV-WLC" interpolation formula captures both the free-solution Pincus regime, which arises from excluded volume, as well as the ideal wormlike chain behavior described by the Marko-Siggia interpolation formula in Eq. 5.10. As we can see in Fig. 5.7, the EV-WLC interpolation formula provides a very good description of the force-extension behavior once the Pincus blobs are smaller than the slit height.

Figure 5.8 provides the second piece of evidence. Recall that the Pincus theory in free solution [101, 168] has the force-extension behavior

$$z \cong \left(\frac{f l_p}{k_B T}\right)^{2/3} \left(\frac{w}{l_p}\right)^{1/3} \tag{5.12}$$

and ends at a maximum force

$$f_{max}^{Pincus} \cong k_B T \left(\frac{w}{{l_p}^2}\right), \tag{5.13}$$

corresponding to the point where the chain inside a free-solution Pincus blob no longer makes a self-avoiding random walk. The rescaled axes in Fig. 5.8(a) show an approximate collapse of the data using the scaling in Eq. 5.12. We would not expect a perfect collapse of the data to Eq. 5.12 since there is still a residual effect due to the walls that gradually decreases as the force increases. Nevertheless, the overall collapse of the data supports our claim that the dominant contribution to the force-extension is the physics

of the Pincus regime in the absence of confinement. For clarity, we have also replotted data from Fig. 5.2(d) in Fig. 5.8(b) using the same rescaling of the force as Fig. 5.8(a). We see that the collapse of the data to the confined Pincus regime in Fig. 5.8(b) starts to fail at the same forces that the collapse of the data to the free solution Pincus theory in Fig. 5.8(a) begins to succeed.

The narrowness of the region where $z \sim f^{2/3}$ reflects the challenge in simulating this regime, rather than a failure of the scaling theory. For an unconfined chain, we were able simulate a decade of forces where $z \sim f^{2/3}$ [168]. For a confined chain, the range of forces where $z \sim f^{2/3}$ is bounded by

$$\frac{k_{\mathrm{B}}T}{H} < f < k_{\mathrm{B}}T\left(\frac{w}{{l_{\mathrm{p}}}^2}\right). \tag{5.14}$$

In principle, we could extend the range of forces where $z \sim f^{2/3}$ to lower forces by increasing the slit height H. However, reaching the long chain limit becomes increasingly more difficult as H increases.

Overall, Fig. 5.7 and Fig. 5.8 support a model of the force-extension in weak confinement with a transition between a confined Pincus regime to a free-solution Pincus regime. In other words, the chain remains in a Pincus regime all the way until it reaches ideal chain behavior, but when H is smaller than the tension blob size the two-dimensional Pincus blobs no longer exist. However, inside each of the blobs, there is still enough contour length to form a self-avoiding random walk [89]. As a result, once the force exceeds $f_{\mathrm{max}}^{\mathrm{weak}}$, the effect of confinement is minimal and the chain now forms three-dimensional Pincus blobs.

5.4.3 Phase Diagrams for Stretching in Slit Confinement

The overall results of our simulations and analysis of the scaling laws are summarized by the phase diagrams in Fig. 5.9. For short chains, $L < l_{\mathrm{p}}$, there should be a rather uninteresting rod-like regime that we have not explored here. The physics of strong confinement of these short, rod-like chains is already relatively complicated [110] without considering the additional complexity of the applied force. In weak confinement, the rod-like regime is trivial since it is essentially unconfined. In either case, at a qualitative level, we would expect the main effect of the applied force is to rotate the rod into the

direction of the force, with the stretching becoming a smaller effect than is the case for long chains.

Outside of the rod-like regime, both strong and weak confinement have a regime of weak stretching for short chains and weak forces [114]. This is effectively a linear-response (Hookean) regime in confinement, analogous with the same regime for stretching of wormlike chains in free solution [92]. We have not provided any data to support the existence of this regime because it is very challenging to obtain accurate measurements of z due to its small value and large fluctuations. Figure 5.5 shows that, independent of the type of confinement, there is a transition into a confined Pincus regime with increasing molecular weight that is given by the scaling result in Eq. 5.2. For weak confinement, the confined Pincus regime is connected to a free-solution Pincus regime once the force exceeds $f_{\max}^{\text{weak}}$. Further increasing the force beyond $f_{\max}^{\text{Pincus}}$ produces ideal chain behavior. Since the free-solution Pincus regime is rather narrow for our simulations, we simply denote this region of the phase diagram as a "free-solution" regime to emphasize the ability to model the stretching of the confined chain with the EV-WLC interpolation formula [168]. We do not observe a free-solution Pincus regime in strong confinement because the Pincus blobs in the plane are removed by the force before they no longer fit inside the slit. At the largest forces, independent of the regime of confinement, there is an ideal chain regime where the force-extension behavior is very similar to that in free solution. The value of the slit height determines the boundary between a free-solution or ideal regime and the confined Pincus regime. When the chain is weakly confined, i.e., when ${l_{\text{p}}}^2/w < H$, the boundary is determined by Eq. 5.3. When the chain is strongly confined, i.e., when ${l_{\text{p}}}^2/w > H$, the boundary is determined by Eq. 5.5.

While we have not provided any supporting data, it is reasonable to assume that the weak stretching regime is directly connected to ideal chain behavior for short (but not yet rod-like) chains. Such chains are effectively ideal over the entire range of force and their force-extension is presumably well described by existing models for the stretching of ideal chains in slits [115, 122].

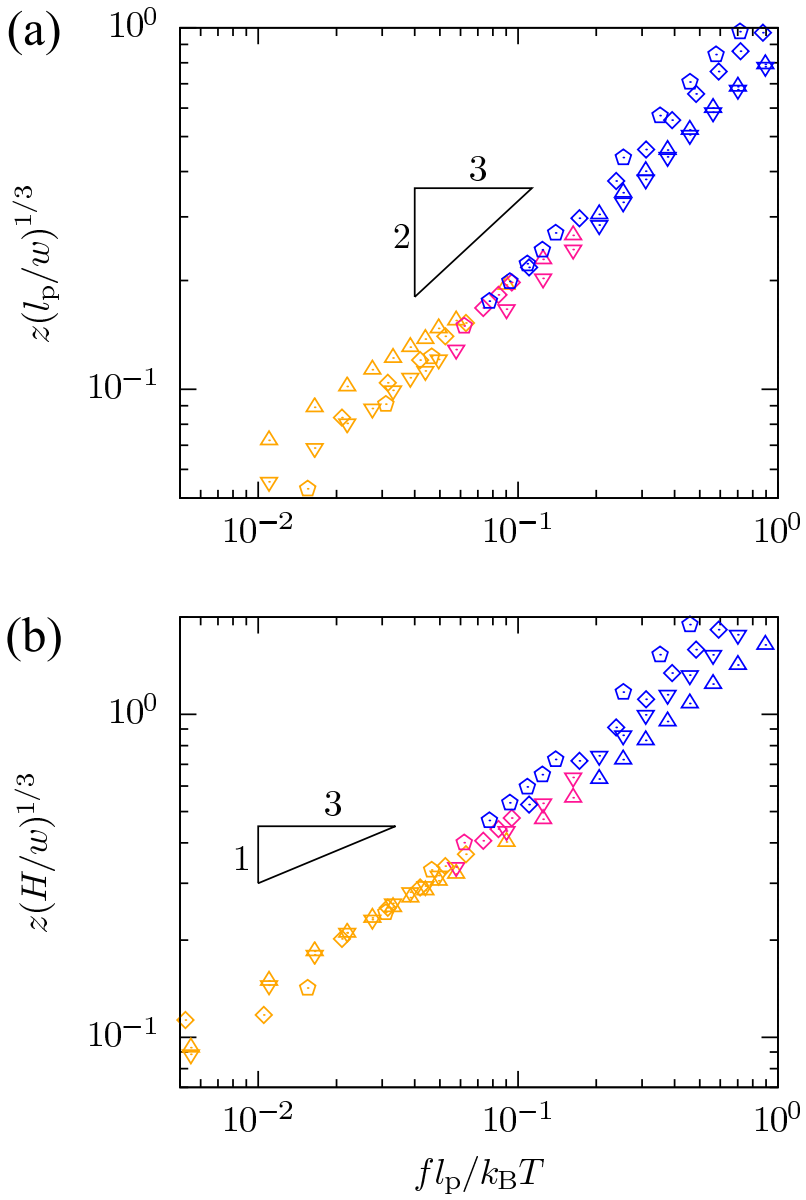

Figure 5.8: Rescaled force-extension data for weak confinement to demonstrate collapse of the data to the theory for the Pincus regime in (a) free solution and (b) confinement. Yellow symbols are simulation data in confined-Pincus regime with $k_BT/f > H$; pink symbols are simulation data in free-Pincus regime with $k_BT/f < H$ and $k_BT/f > {l_p}^2/w$; green symbols are in ideal chain regime with $k_BT/f < {l_p}^2/w$. Triangle: l_p/w = 5.5, H/w = 49; inverted triangle: l_p/w = 5.5, H/w = 99; diamond: l_p/w = 10.5, H/w = 149; pentagon: l_p/w = 15.5, H/w = 299.

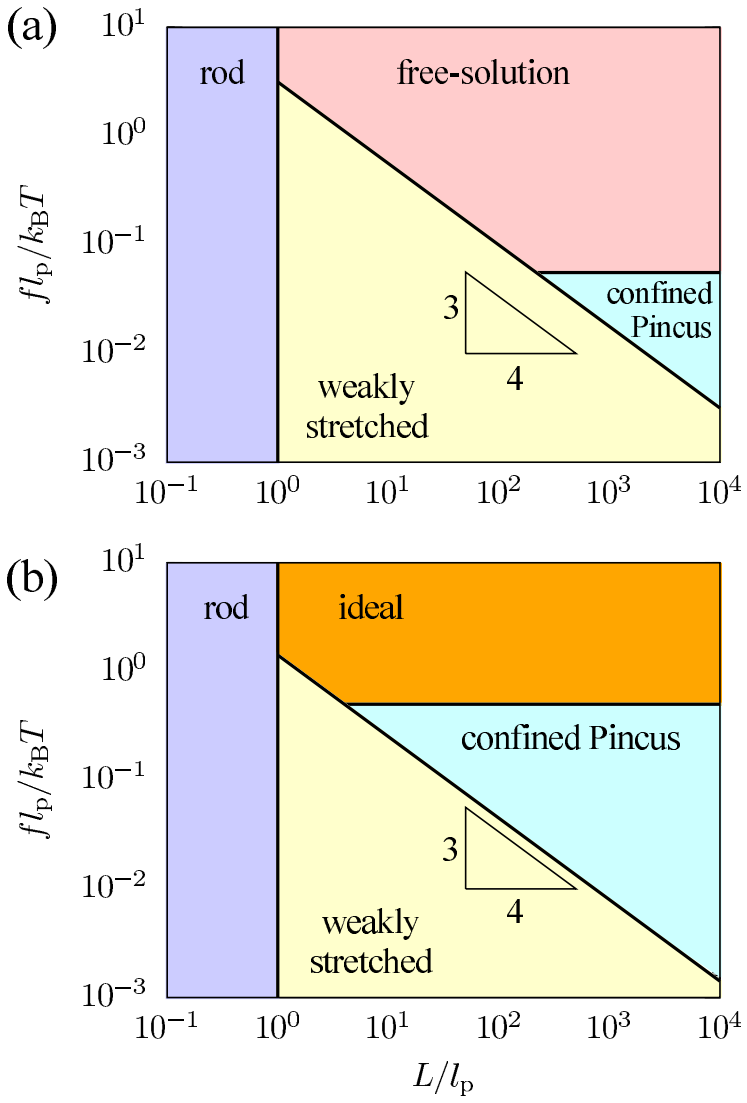

Figure 5.9: Phase diagrams for force-extension of wormlike chains in slit confinement. (a) Weak confinement for a slit height $H/w = 99$ and $l_p/w = 5.5$. The boundary between the weakly stretched regime and other regimes is given by Eq. 5.2. The boundary between the confined Pincus regime and the free-solution regime is given by Eq. 5.3. The stretching in the free solution regime is approximated by Eq. 5.11, which includes both free-solution Pincus behavior and ideal-chain behavior. (b) Strong confinement for a slit height $H/w = 4$ and $l_p/w = 10.5$. The boundary between the weakly stretched regime and other regimes is given by Eq. 5.2. The boundary between the confined Pincus regime and the ideal regime is given by Eq. 5.5. The stretching in the ideal regime is approximated by Eq. 5.9.

5.4.4 Applications to DNA

Thus far we have focused on asymptotically long chains, with the aim of testing the existence and scaling theory for the confined Pincus regime. While these situations are of most interest from the standpoint of polymer theory, many experimental scenarios using DNA involve shorter chains. We thus simulated λ DNA (48.5 kilobase pairs) and T4 DNA (169 kilobase pairs) as model systems for DNA experiments. Note that these particular sequences of DNA are commonly used in the literature due to their ready commercial availability in purified forms, but there is no intrinsic reason why longer DNA cannot be used for experiments. Indeed, the pioneering experiments on nanochannel confinement [170] used λ-DNA concatemers up to 12-λ (582 kilobase pairs), and entire chromosomes were used in subsequent channel confinement experiments [171, 172]. For our DNA model, we use an effective width $w = 5$ nm [80] based on Stigter's theory [173] for a high ionic strength buffer, with λ-DNA corresponding to a chain of $N = 3361$ beads (16.8 μm length) and T4 DNA corresponding to a chain of $N = 11971$ beads (59.85 μm length). We set the persistence length to $l_{\mathrm{p}} = 52.5$ nm, which is a reasonable value at high ionic strength [82, 148]. The actual channel width is $H_{\mathrm{wall}} = 25$ nm, slightly larger than the lower bound using standard fabrication methods [174], which leads to an accessible width of $H = 20$ nm.

Figure 5.10 shows the simulated fractional extension for both DNA molecules as a function of the force. As anticipated from our previous results, a confined Pincus regime exists in both cases for weak forces. However, the span of this regime is narrow for λ DNA (Fig. 5.10(a)), with much of the low force data corresponding to the weak stretching regime arising from its finite molecular weight. As the molecular weight increases (Fig. 5.10(b)), the range of forces corresponding to the confined Pincus regime increases.

5.5 Discussion

The key objective of this chapter was to test the scaling theory by Taloni *et al.* [114] for the confined Pincus regime. Our data support their theory for the scaling of the force-extension in Eq. 5.1, the lower bound in molecular weight in Eq. 5.2, and the upper bound for weak confinement in Eq. 5.3. The key difference between our results

and their scaling theory [114] emerges for strong confinement, where the data support the crossover given by Eq. 5.5 rather than the result in Eq. 5.3 proposed by Taloni *et al.* [114]. The reason for this discrepancy is that the Pincus regime can end when the Pincus blobs no longer fit inside the slit (which is the case for weak confinement) or when the subchain within the Pincus blob is no longer a self-avoiding walk (which is the case for strong confinement).

Taloni *et al.* [114] also proposed a scaling theory to describe the force-extension behavior for stronger forces in slit confinement. The key idea in their theory for stronger forces is an analogy between stretching imposed by the external force and stretching imposed by compression in a nanochannel. Once the confined Pincus regime ends, their theory proposes that the extension of the chain in a force becomes the same as the extension of the chain in an effective channel set by the size of the tension blobs. Our simulation data do not support this theory for higher forces, consistent with prior work identifying the differences between stretching in free solution and compression in a nanochannel when the force exceeds the Pincus regime in free solution [100]. Rather, the data in Fig. 5.6 and Fig. 5.7 support a model where the force-extension is essentially the same as that in the absence of confinement. Importantly, we identified a free-solution Pincus regime over small range of forces (Fig. 5.7) that cannot emerge from an analogy between stretching and nanochannel confinement.

Our force-extension simulation results for stretching in strong confinement extends the previous work by Hsu and Binder [121], who used PERM simulations to study self-avoiding, semiflexible chains on a square lattice. Similar to our results for the confined Pincus regime, they found sensible excluded volume interactions under a weak stretching force when the chain is restricted to the plane. Indeed, they found that the resulting two-dimensional Pincus blobs have a force-extension behavior

$$z \sim \left(\frac{f l_{\mathrm{p},\parallel}}{k_{\mathrm{B}} T} \right)^{1/3}, \tag{5.15}$$

which has the same scaling with the applied force as Eq. 5.2. The difference between their two-dimensional study and our results for strong confinement is that we allow the fluctuations in the direction perpendicular to the slit walls, which requires accounting for the slit height H relative to the size of the excluded volume interactions, w.

Our results have important implications for previous work on modeling the stretching of wormlike chains in slit confinement. In a recent publication, de Haan and Schendruk [122] proposed several interpolation formulas for the stretching of ideal wormlike chains in confinement. Our results indicate that their model is not applicable to real polymers under weak forces, as their interpolation in this regime is based on Kraky-Porod model [102], which describes the polymer as an entropic spring with no interactions between non-neighboring monomers. As illustrated in Fig. 5.6 and Fig. 5.7, the correct description at weak forces for real chains is the confined Pincus theory [114]. Moreover, as the force increases, the correct description for stretching of real chains in weak confinement is a free-solution Pincus behavior that also cannot arise from an ideal chain theory. A qualitative difference between our model for real chains and their model for ideal chains is that the stretching of ideal chains can be described by a single interpolation formula that spans all slit heights, from strong confinement to weak confinement [122]. In contrast, our phase diagrams for real chains in Fig. 5.9 show that an additional free-solution Pincus regime arises in weak confinement, rendering it qualitatively different than strong confinement.

Our model also differs qualitatively from that proposed by Chen *et al.* [115], who addressed this problem using Brownian dynamics simulations of a bead-spring polymer model with repulsive interactions between the beads. The latter confined polymer model is most appropriate for weak confinement, where the presence of the walls does not substantially alter the conformational entropy of the subchain that gives rise to the the spring force. Their model for the force-extension behavior in weak confinement is a modification of the three-dimensional Marko-Siggia interpolation formula in Eq. 5.10 to account for a finite extension z_0 at zero force caused by the excluded volume interactions,

$$\frac{f l_{\mathrm{p},\parallel}}{k_{\mathrm{B}}T} = \frac{1}{4}\left[\frac{1}{(1-z)^2} - \frac{1}{(1-z_0)^2}\right] + (z - z_0). \tag{5.16}$$

Although they include the excluded volume interactions qualitatively in the model, they found that this interpolation works better for larger force ($f l_{\mathrm{p},\parallel}/k_{\mathrm{B}}T > 1$). Our results explain the discrepancy between Eq. 5.16 and their simulation data, as well as the good agreement between their simulation data and Eq. 5.16 at higher forces [115]. For weak forces, Eq. 5.16 does not reduce to the confined Pincus scaling in Eq. 5.1, which we

now know is the correct description for real wormlike chains under a weak stretching force. Thus, it is unsurprising that Eq. 5.16 is not accurate for weak forces. For strong forces, we found that the original three-dimensional Marko-Siggia interpolation formula given by Eq. 5.10 is a good model for the stretching. Since z_0 is small, it is readily apparent that Eq. 5.16 approaches the three-dimensional Marko-Siggia formula in the high force limit. Thus, it is reasonable that Eq. 5.16 is a good approximation to the force-extension behavior in weak confinement under strong forces.

While we have achieved good agreement between simulation and scaling theory, it is challenging to verify the scaling theory for weak extension experimentally. To date, most experiments have focused on the strong stretching regime. For example, Yeh *et al.* [117] have probed the force-extension relation of a single molecule confined in a nanoslit. In their experimental system, two symmetric micro-nano interfaces were connected by a nanoslit. The conformational entropy difference between the portion of polymer chain confined in the nanoslit and the portion in the reserviors induces an entropic force, which stretches the DNA segment inside the nanoslit. However, their experimental setup is only capable of studying strong forces with strong confinement, since the chain is sensitive to elastic retraction and thus recoils from the nanoslit into the reserviors. Another experimental possibility is the system of Lin *et al.* [175], where one end of λ-DNA is tethered and the other is attached to a magnetic bead. They measured the elastic response of the portion of stretched DNA that is confined through a slit between two reservoirs.

While these experiments dealt with strong stretching ($z > 0.5$), after we published our result, Yeh *et al.* [71] further studied the force-extension relation for a wide range of the extension by applying external electric fields to stretch DNA tethered to microspheres anchored at a nanoslit entrance. Excitingly, this experiment validates our theoretical prediction, particularly the "confined Pincus" regime in slit confinement.

5.6 Concluding Remarks

Through a combination of scaling theory [114] and simulations, we have constructed a phase diagram for the stretching of semiflexible polymers confined in nanoslits with a slit height H. Most notably, our results demonstrate strong evidence in support of a

confined Pincus regime in all ranges of confinement, as well as a free-solution Pincus regime in weak confinement. The latter additional regime points towards a qualitative difference between stretching in the Odijk and de Gennes regimes of confinement. To a large extent, the results confirm the scaling theory by Taloni *et al.* [114] for the confined Pincus regime, with the exception of the maximum force for the confined Pincus regime under strong confinement. Our results have also clarified shortcomings in existing models for the stretching of wormlike chains in slits [114, 115, 122], including (i) showing that the inapplicability of the compression-stretching analogy under strong forces in free solution [100] is also valid in confinement [114]; (ii) demonstrating the inability of ideal chain models to capture the force-extension behavior of real polymers under weak forces [122], similar to the case in free solution; and (iii) showing that existing interpolation formulas [115, 122], which do not account for any of the Pincus regimes in Fig. 5.9, cannot capture the force-extension behavior in slit confinement. While we focused on the confined Pincus regime for long chains (corresponding to DNA with contour lengths from 45 μm to 360 μm), we still see the evidence of confined Pincus regime at relatively low forces for λ DNA and T4 DNA. However, when the force decreases further, the force-extension behavior quickly goes to the weakly stretched regime, due to the finite chain length. Although it may be challenging to experimentally test the phase diagrams in Fig. 5.9, we anticipate that such a complete description of the force-extension of real wormlike chains will prove useful for both the interpretation of experimental data and the design of new technologies.

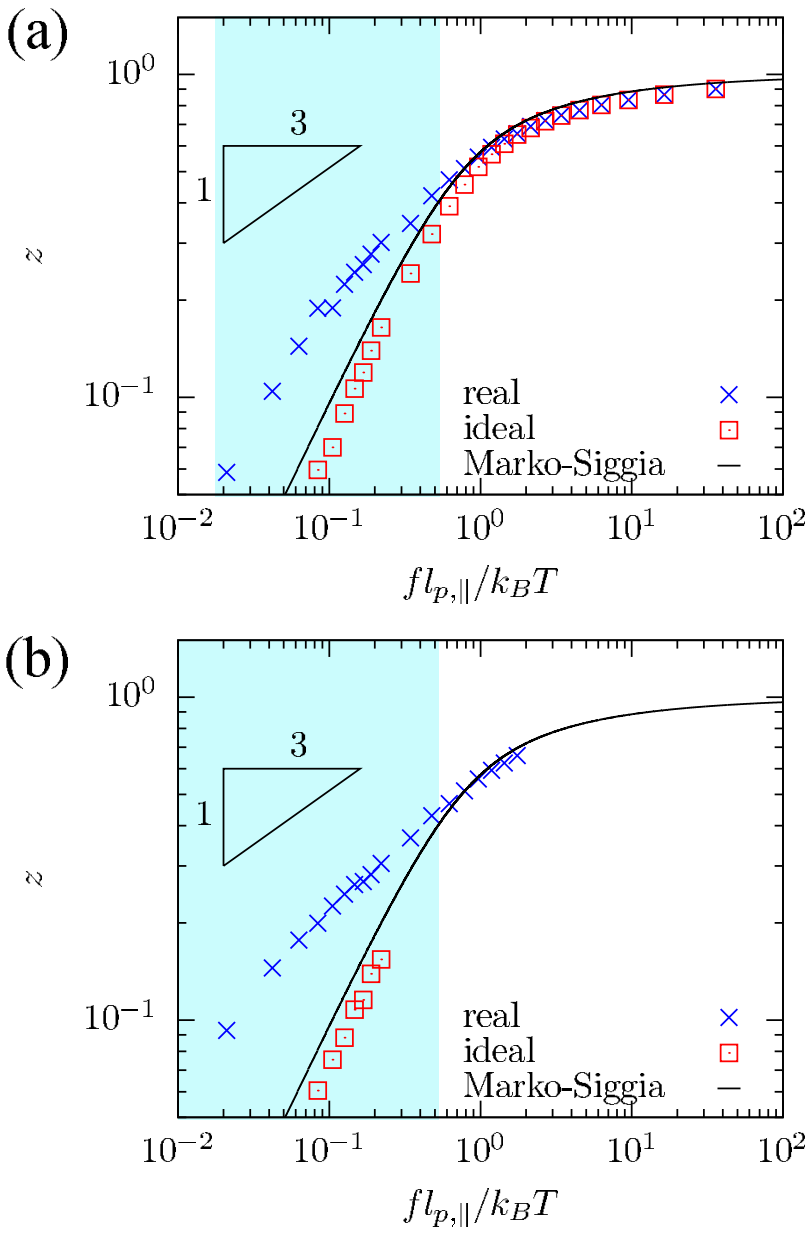

Figure 5.10: Plot of the fractional extension versus rescaled force for l_p/w = 10.5 and H/w = 4 where (a) corresponds to λ DNA (N = 3361 beads) and (b) corresponds to T4 DNA (N = 11971 beads). The solid black line is Eq. 5.9. The shaded area shows the confined Pincus regime.

Chapter 6

Rapid Conformational Fluctuations in a Model of Methylcellulose

This chapter is based on the publication
X. Li, F. S. Bates, and K. D. Dorfman, "Rapid conformational fluctuations in a model of methylcellulose"
Phys. Rev. Mat., vol. 1, no. 2, p. 025604, 2017 [176]

6.1 Introduction

As mentioned in Section 1.2, methylcellulose (MC) is an important class of biopolymers investigated extensively to develop materials with innovative physico-chemical properties. Depending on the degree of substitution, MC in water can form a hydrogel upon heating. Recent experimental studies [45,46] on the gel structure reveal a fibrillar morphology with a uniform diameter. Fibril formation cannot be explained solely by established equilibrium concepts such as Flory-Huggins solution theory [67, 68], thus requiring a new model. From a computational standpoint, this is not an easy problem since the fibrils have a relatively large axial extent ($\sim \mu$m), but the chemical details of the polysaccharide units must be included in order to capture the intramolecular

interactions leading to fibril formation.

Several computational studies have been conducted to study the gelation mechanism and have been reviewed in Section 2.4. Although these computational studies [1, 69] provide a plausible model for fibril formation in dilute solution, they leave open two important questions that we address here. First, other collapsed states, such as hairpins and folded bundles, could appear as long-lived intermediate states, or even final (metastable) states [1]. How these "misfolded" collapsed states are removed and how to prevent these misfolded states from propagating through fibril formation are unanswered questions. Second, although atomistic simulations suggest [1, 130] that the hydrophobic interaction, i.e., the non-bonded self-attraction, is the major driving force for MC gelation, the coarse-grained MC model was parameterized from atomistic simulations of 10-mer oligomers. The dihedral potential obtained from this mapping involves four successive monomers, and it is not obvious that the potential obtained from mapping 10-mers will translate to longer chains. Thus, it is important to investigate the sensitivity of the dihedral potential before drawing conclusions regarding MC gelation, as it competes with the non-bonded interactions that drive the formation of collapsed structure.

We show here that significant insights into the gelation mechanism can be obtained through long-time simulations of single MC chains and a model problem of ring association, taking advantage of the model proposed by Huang *et al.* [1]. In particular, we identify a previously overlooked, rapid conformational fluctuation produced by the dihedral potential that we posit is important for both escaping misfolded states and guiding fibril formation for multiple chain simulation. We also study the effects on the dihedral potential on the frequency of this rapid conformational change and shape of the collapsed states, demonstrating that the fluctuations in the collapsed state are governed by the balance between the dihedral potential and non-bonded interactions. While this coarse-grained model appears to capture many of the features of the collapse of a single methylcellulose chain, we show that several important features for methylcellulose gelation remain unanswered and suggest directions for further model improvements.

6.2 Model and Simulation Method

6.2.1 Coarse-grained Model of Methylcellulose

The model and the simulation method are described in detail by Huang *et al.* [1]. For completeness, we outline the key details here and recapitulate the various model parameters in Appendix C. In the simulations described throughout this work, MC is modeled as a heterogenous polymer with all eight types of monomers, i.e., cellulose monomer, 2-MC, 3-MC, 6-MC, 2,3-MC, 2,6-MC, 3,6-MC and 2,3,6-MC. The order of the monomers was generated randomly and the composition profile was chosen to match the mole fraction of the commercial polymer METHOCEL A [43], as tabulated in Appendix C. Each bead in the coarse-grained (CG) model represents one monomer. The bead-bead interaction potential,

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{nonbonded}} \tag{6.1}$$

contains four parts: harmonic bond, angle, dihedral interactions, and nonbonded potentials. The first three potentials are of bonded types and expressed as

$$U_{\text{bond}} = \frac{1}{2} K_{\text{b}} (l - l_0)^2 \tag{6.2}$$

$$U_{\text{angle}} = \frac{1}{2} K_{\theta} (\theta - \theta_0)^2 \tag{6.3}$$

and

$$U_{\text{dihedral}} = K_{\text{d}} [1 + d \cos(n\phi)] \tag{6.4}$$

Here K_{b}, K_{θ} and K_{d} are the bond, angle and dihedral force constants, respectively, l_0 is the equilibrium bond length, and θ_0 is the equilibrium bending angle. In Eq. 6.4, d and n are the phase constants. Huang *et al.* [1] performed atomistic simulations of 10-mer homogenous MC oligomers and determined the bonded parameters for the CG model by mapping the radial distribution functions (RDF) from the atomistic model to their CG counterparts. Although there are 8 different homopolymers, the intramolecular atomistic RDFs are similar among all monomer types. We thus use the same set of the bonded parameters for all monomers, and the parameters are summarized in

Appendix C. In what follows, we will be particularly interested in K_{d}, which controls the dihedral strength for four consecutive beads on a chain.

The nonbonded interactions have the form of a truncated and shifted Lennard-Jones (LJ) 9-6 potential,

$$U_{\text{nonbonded}} = \begin{cases} \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^9 - \left(\frac{\sigma_{ij}}{r}\right)^6 - \left(\frac{\sigma_{ij}}{r_{\mathrm{c}_{ij}}}\right)^9 + \left(\frac{\sigma_{ij}}{r_{\mathrm{c}_{ij}}}\right)^6\right] & r < r_{\mathrm{c}_{ij}} \\ 0 & r \geq r_{\mathrm{c}_{ij}} \end{cases} \tag{6.5}$$

where i and j denote two types of MC monomers. The intermolecular nonbonded interaction parameters among the same type of monomers, i.e., σ_{ii}, $r_{c_{ii}}$ and ε_{ii}, were obtained again by Huang *et al.* [1] by matching the atomistic simulation of short homogenous oligomers to the CG models. The bead size and cut-off radius, σ and r_{c}, depend on the type of monomers and the ambient temperature, while ε is also an explicit function of contour length. Note that the "effective temperature" is captured by the nonbonded parameters rather than being an explicit function of temperature in the CG simulation. We have tabulated the nonbonded parameters in Appendix C for a representitive low temperature (25 °C) and a representitive high temperature (50 °C). At 25 °C, MC is soluble in water, while gelation can occur at 50 °C [177]. For the nonbonded interactions between different types of monomers, we used a geometric mixing rule to calculate the σ_{ij}, ε_{ij} and $r_{\mathrm{c}_{ij}}$ in the heterogenous MC chains. The mixing rules are also included for completeness in Appendix C.

6.2.2 Simulation Method

Following prior work [1], we simulated this coarse-grained model using Langevin dynamics with the velocity-Verlet intergrator in the LAMMPS package [3] (ver. Jun 2014) in the NVE ensemble. Simulations were set up using LJ units with three fundamental scales: m for unit mass, σ for unit distance, and ε for unit energy. We chose the unit mass to be the average molecular weight of all 8 different types of MC monomers, 188 Da. Although different types of monomers have different mass depending on how many hydroxy groups are substituted, the additional complications introduced by accounting for small changes in mass are not consequential within a coarse-grained, Langevin

dynamics simulation. We set the unit length to be 0.515 nm, which is the simulated averaged center-of-mass separation of monomers from previous atomistic simulations [1]. The unit energy, ε is set to be $k_{\mathrm{B}}T$ with k_{B} being the Boltzmann factor and T = 298 K. We ran our simulations using a Langevin thermostat at the unit temperature, and the "effective temperature" is captured by the self-attraction force field between MC monomers as described above. A summary of all the dimensionless parameters and their corresponding dimensional parameters are tabulated in Appendix C.

In addition to these three fundamental quantities, the damping parameter in the Langevin dynamics algorithm plays an important role in determining the relationship between simulation time scale and real time scale. Although the unit time is not a fundamental unit and can be expressed as $\tau = (m\sigma^2/\varepsilon)^{1/2}$, its dimensional counterpart can not be calculated directly from the dimensional value of m, σ and ε because they contain no information about the solvent. Thus, the damping parameter is specified in time units and is regarded as inversely related to solvent viscosity. We set the damping parameter to be 10τ, leading to the estimate τ = 0.028 ns. This value was reported by Huang *et al.* [1] and we confirmed it by measuring diffusion coefficients in our simulations at T = 25 °C.

Unless otherwise specified, we initiated our simulations with random sequences for heterogeneous MC. We simulated 1000-mer MC chains with periodic boundary conditions and a box size length of 600σ. The simulation results reported for each value of K_{d} include at least three independent trajectories. To study the collapse of the chain above the gelation temperature, the MC chain is first equilibrated at low temperature (25 °C) for 10^7 steps at a time step of $5 \times 10^{-4}\tau$, and an instantaneous temperature jump is introduced by switching the nonbonded parameters from 25 °C to 50 °C. Then the simulation proceeds for at least 5×10^7 steps with time step of $5 \times 10^{-4}\tau$.

6.2.3 Data Analysis

We study the rapid conformational fluctuations by monitoring the shape of the MC chain, and we will call this rapid change a "flipping event," which we discuss later in more detail. Specifically, three quantities are used for structural characterization: the eigenvalues of the gyration tensor λ, the radius of gyration R_{g}, and the relative shape anisotropy κ^2.

The gyration tensor quantifies the second moments of monomer positions on a polymer chain,

$$S_{mn} \equiv \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)} \tag{6.6}$$

where $r_m^{(i)}$ is the m^{th} Cartesian coordinate of the position of the i^{th} monomer. The coordinate system has been chosen so that the center of mass lies at the origin. Since the gyration tensor is a symmetric 3×3 matrix, diagonalization produces the principal moments of the gyration tensor, i.e., the squared eigenvalues. The eigenvalues are ordered such that ${\lambda_x}^2 \leq {\lambda_y}^2 \leq {\lambda_z}^2$. The eigenvalues of the gyration tensor measure the extensions in the principle axis system, thus give the dimensions of an object.

The principal moments can be combined to give two scalar quantities that describe the shape of the polymer conformation, thereby enabling us to locate the flipping events without the need to consider the detailed configurations of the chain. The squared radius of gyration is the sum of the principle moments of the gyration tensor,

$${R_{\text{g}}}^2 = {\lambda_x}^2 + {\lambda_y}^2 + {\lambda_z}^2. \tag{6.7}$$

We found that $\Delta R_{\text{g}}/R_{\text{g}}$, i.e., the relative difference of radius of gyration between two time frames, is a useful proxy for capturing the conformational change. We chose the time difference for computing ΔR_{g} to be 10^5 steps, or 50τ. A peak of $\Delta R_{\text{g}}/R_{\text{g}}$ signals a sudden conformational change in the polymer size, and thus a potential flipping event. We set the threshold of a flipping event to be $\Delta R_{\text{g}}/R_{\text{g}} \geq 0.1$ to screen out typical changes in R_{g} due to thermal fluctuations, and we also cluster peaks within 50τ to be a single flipping event in order to avoid overcounting large conformational changes immediately following a flipping event that result from the instability of the relatively open chain configuration.

Finally, the relative shape anisotropy [178, 179] is defined as

$$\kappa^2 = 1 - 3\frac{\lambda_x^2\lambda_y^2 + \lambda_x^2\lambda_z^2 + \lambda_y^2\lambda_z^2}{\left({\lambda_x}^2 + {\lambda_y}^2 + {\lambda_z}^2\right)^2}. \tag{6.8}$$

This shape descriptor reflects both the symmetry and dimensionality of a polymer conformation; the minimal value $\kappa^2 = 0$ indicates a highly symmetric conformation while

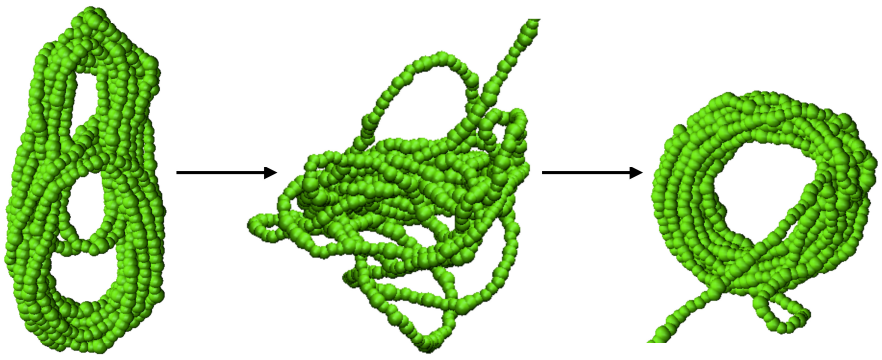

Figure 6.1: Snapshots of a flipping event of a 1000-mer MC chain at 50 °C. The time lag between each snapshot is 7 ns, corresponding to 250 τ.

the maximal value $\kappa^2 = 1$ indicates all beads lying on a line. For a planar symmetric structure, for example a ring conformation, κ^2 is around 1/4 [180].

6.3 Single Chain Results

6.3.1 Flipping Events in an Isolated MC Chain

We first simulated an isolated MC chain with N = 1000 at 50 °C, where the self-interactions between MC monomers are strong enough to form collapsed structures in the model of Huang *et al.* [1] It has been shown by simulation [1, 69] that, at elevated temperatures, a single MC chain with 600 monomers or more can form a ring structure. We confirmed this result in our simulations.

Interestingly, we observed that after the chain forms a collapsed structure, large conformation changes can happen within a short period of time that do not correspond to the expected breathing modes of a ring. Figure 6.1 shows a typical flipping event in a single 1000-mer MC chain. The total simulation time is $3.4 \times 10^4 \tau$, which corresponds to 952 ns. We started the high temperature simulation with an initial configuration generated from a room temperature simulation, as described in the Methods. Because at low temperature the self-attraction is insufficient to collapse the chain, the MC chain is initially a random coil. At t = 0, we elevated the temperature and found that the

conformation of the MC chain first becomes a collapsed structure at t = 280 ns, consistent with previous work [1,69], and undergoes the small shape fluctuations one would expect for a collapsed ring. However, at t = 592 ns, as shown in the first snapshot of Fig. 6.1, the chain adopts a bifocal structure. Within the next 7 ns, the chain goes through a huge conformational change with the whole collapsed structured disturbed. Owing to the self-attraction forces, this loose structure is not stable. After another 7 ns, the chain collapsed back into another relatively tight structure, thus converting from one collapsed conformation to another. We have included the rotating views of the three representative snapshots in Appendix C. The dynamics in this particular example are not a special case; flipping happens quite often throughout the course of simulation. We posit that flipping events are of great importance because they impart an ability to alter rapidly the collapsed structures, which are regarded as local energy minima and represent metastable states.

If we assume that a sudden change of the conformation indicates a flipping event, we can identify the flipping events by monitoring the changes of radius of gyration. Figure 6.2 shows how R_g and $\Delta R_g/R_g$ evolve with respect to simulation time for three independent trajectories with N = 1000 at 50 °C. Although different trajectories differ in their details, as expected from a stochastic simulation, the overall trends indicate that the chain size and flipping frequencies are robust to the initial conditions and the thermal noise. At the start of the simulation, R_g is relatively high because the MC chain begins in a coiled state. As simulation proceeds, R_g decreases due to the collapse of the MC chain, and reaches a plateau at $t \sim 1 \times 10^4 \tau$. The average value of the plateau for R_g is around 6 nm. After this time, R_g stays in a small range with typical thermal fluctuations. At a few time points, R_g goes through a sudden change, indicating the flipping events. These fluctuations are challenging to identify from R_g alone, so the lower plot of Fig. 6.2 shows $\Delta R_g/R_g$ with respect to time. This plot clearly shows the location of sudden change of the chain conformation, readily distinguishing the flipping events from more subtle thermal fluctuations. To automate the identification of flipping events, we set the threshold of a flipping event to be above $\Delta R_g/R_g$ = 0.1.

Of particular interest are the kinetics of the flipping events. We calculated the time interval between two flipping events by subtracting the time of a flipping event from its

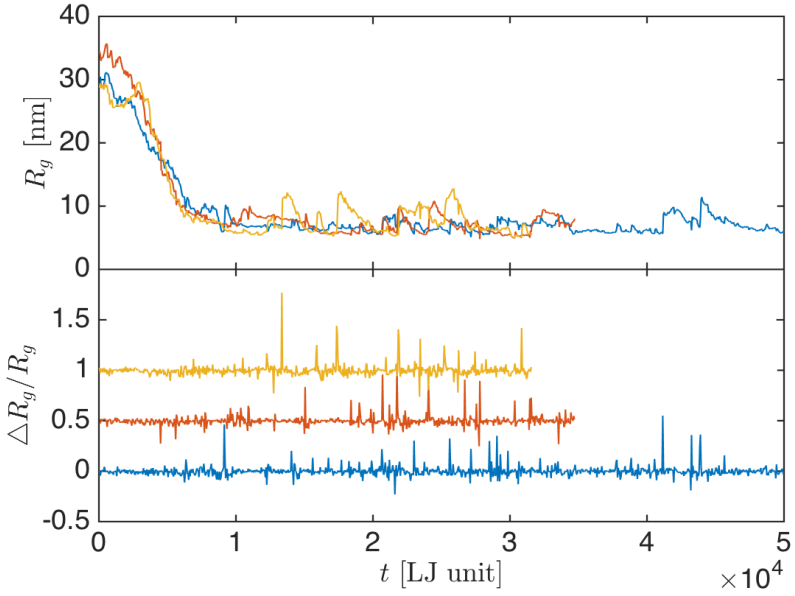

Figure 6.2: Radius of gyration and the relative difference of R_{g} for a 1000-mer MC chain as a function of simulation time. Upper: R_{g} versus time for three independent trajectories with 1000 monomers. The three trajectories are of total duration $3.16\times10^4\tau$, $3.475\times10^4\tau$ and $5\times10^4\tau$. Before $t=0$ the chain was equilibrated at room temperature. At $t=0$ the temperature was elevated to 50 °C. Bottom: $\Delta R_{\mathrm{g}}/R_{\mathrm{g}}$ versus time for the same three trajectories. An offset of 0.5 on y-axis for each trajectory has been made for clarity. A total of 14, 18 and 25 flipping events are identified for top to bottom trajectories, respectively.

successor,

$$\Delta t_n = t_{n+1} - t_n. \tag{6.9}$$

Figure 6.3 shows a histogram of the resulting time interval distribution. Along with the three trajectories in Fig. 6.2, we added two more trajectories to improve the sampling. The median of the time interval distribution is 950τ. Compared to a typical flipping time duration of 250τ, as shown in Fig. 6.1, the flipping event happens quite often and thus is able to alter the chain conformation within a short period of time.

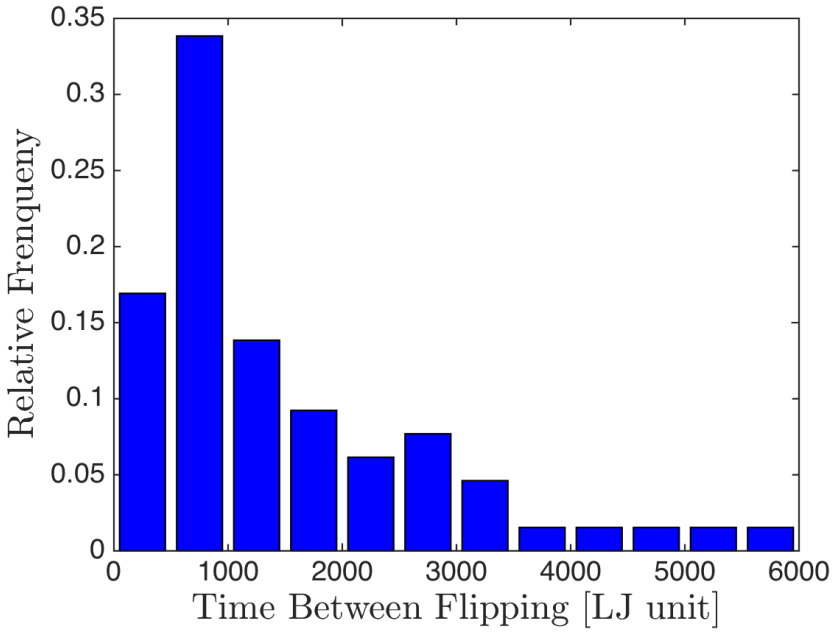

Figure 6.3: Histogram of the time between two flipping events for 1000-mer MC chain at 50 °C. The data correspond to a total 65 flipping events obtained from 5 independent trajectories.

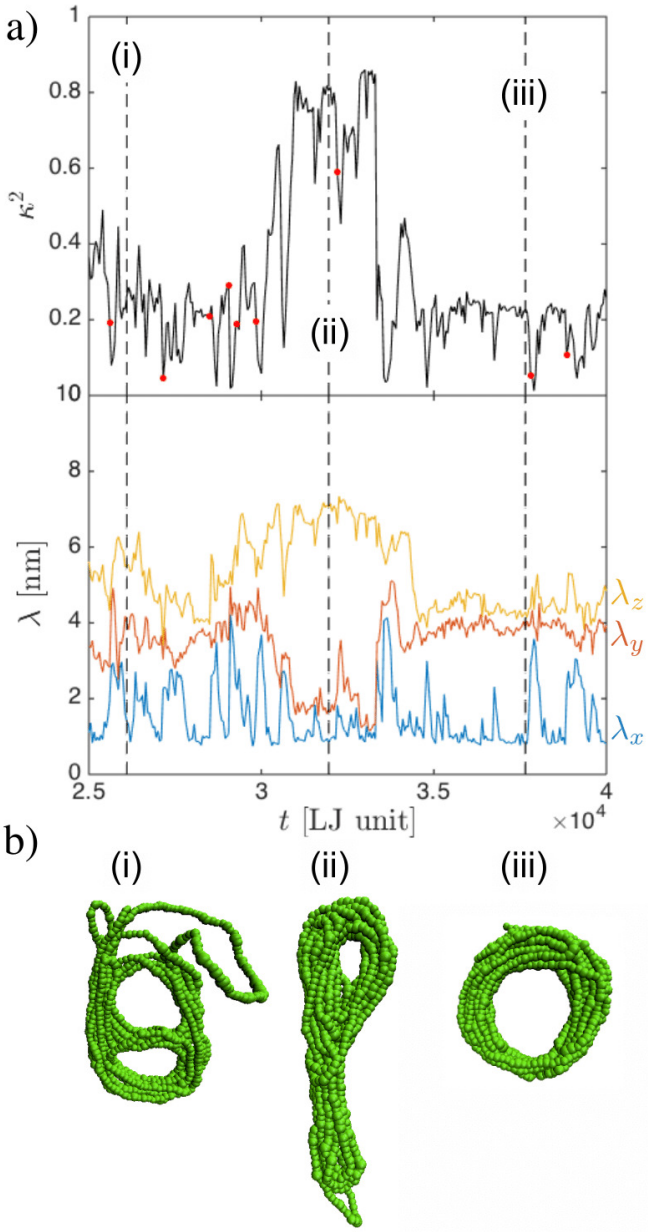

Figure 6.4: (a) Two-panel plot of the shape anisotropy κ^2 and the individual eigenvalues of the gyration tensor λ_i versus time for a 1000-mer MC chain. The temperature jumps from 25 °C to 50 °C at the time t = 0 (not indicated in the figure). The (red) dots indicate flipping events. (b) Snapshots at (i) 731 ns, (ii) 889 ns and (iii) 1054 ns, which correspond to $2.61 \times 10^4\tau$, $3.18 \times 10^4\tau$ and $3.77 \times 10^4\tau$, respectively. The corresponding time points in panel (a) are indicated by the dashed vertical lines.

We found that the eigenvalues of the gyration tensor ($\lambda_x,\ \lambda_y,\ \lambda_z$) and the relative shape anisotropy (κ^2) are particularly insightful for characterizing the chain configurations before and after flipping events. In Fig. 6.4(a), we plot part of one trajectory with N = 1000 for an isolated MC chain at 50 °C. Similar plots for other trajectories are included in Appendix C. We also show three representative conformation snapshots in Fig. 6.4(b). During much of the simulation, the shape anisotropy is around 0.25, which indicates a planar symmetric structure. However, we see fluctuations throughout much of the range of κ^2, ranging from 0.1 and 0.85, indicating other shapes. In the three representative conformations in Fig. 6.4(b), the first is of a misfolded bifocal structure with some loose "arms." Because this structure is not stable, and in the process of re-orientating itself, the corresponding value of κ^2 is part of a large fluctuation at the time indicated by (i) in Fig. 6.4(a). After several flipping events, the chain conformation adopts a highly anisotropic bundled structure with a relatively high value of $\kappa^2 \approx 0.8$. The last structure shows a return to a ring shape with κ^2 close to 0.25, indicating a planar symmetric structure. Thus κ^2 is very helpful in discriminating shapes between bundled and planar configurations, the two most common collapsed structures observed in our simulations.

However, κ^2 does not discriminate between cases (i) and (iii) in Fig. 6.4, i.e., between planar symmetric structures with different shapes. Thus, it is useful to consider also the eigenvalues of the associated gyration tensor, which measure the extension in the principle axis system. Some small fluctuations, as well as large jumps, are also observed Fig. 6.4(a) for the individual eigenvalues. We observed that for the third structure, corresponding to the desired ring structure, λ_y and λ_z are almost identical. These two eigenvalues indicate the outer diameter of the axisymmetric ring conformation, while the smaller eigenvalue λ_x indicates the thickness of the ring. The eigenvalues around (i) are fluctuating substantially, consistent with our discussion of the shape anisotropy. The eigenvalues for (i) are also very different from those for (iii), and this difference allows us to identify the two structures. For (ii), we just have one large eigenvalue (λ_z) and two smaller eigenvalues (λ_x and λ_y), indicating a bundled structure. Thus, κ^2 and eigenvalues of the gyration tensor are complementary indicators that together describe the highly structured conformations of the MC chain.

To further explore the effects of the flipping event on the shape of the collapsed

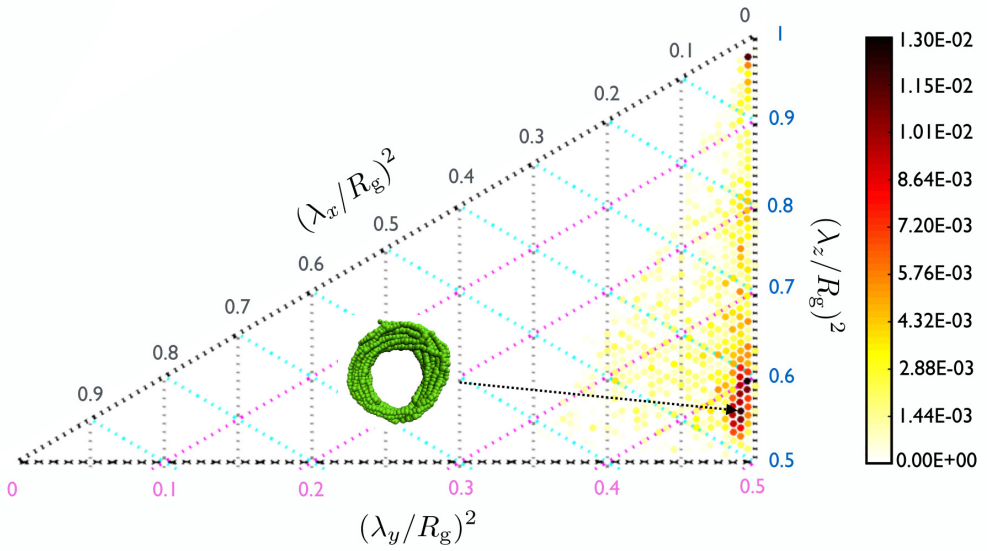

Figure 6.5: Ternary plot of the rescaled eigenvalues $(\lambda/R_g)^2$ for 1000-mer MC chains at 50 °C. The eigenvalues are ordered as $\lambda_x \le \lambda_y \le \lambda_z$ and satisfy $\lambda_x^2+\lambda_y^2+\lambda_z^2 = R_g^2$. Data points are sampled every 50τ. The data correspond to the second half of 5 independent trajectories to remove the effects of the initial configuration and make sure the MC chain has already reached or visited a collapsed state. The 2006 data points are binned with bin size of $\Delta(\lambda_x/R_g)^2 = 0.01$ and $\Delta(\lambda_y/R_g)^2 = 0.01$. The colormap shows the probability of observing the combination of eigenvalues in a given bin. A representative MC conformation within the highest probability bin is included.

states for an isolated MC chain, Fig. 6.5 presents a ternary plot for the eigenvalues of the gyration tensor using the five trajectories from Fig. 6.3. The idea is to learn how the dimensions of the conformation are distributed throughout the course of a simulation, as well as to evaluate the likelihood of different collapsed shapes. We plot three rescaled eigenvalues $(\lambda/R_g)^2$ in the order of ${\lambda_x}^2 \le {\lambda_y}^2 \le {\lambda_z}^2$. A striking feature is that the most probable conformation corresponds to $\lambda_x < \lambda_y \approx \lambda_z$; this indicates that most of the simulation time is spent in a ring configuration. Also, the eigenvalues are distributed due to the combined effects of regular thermal fluctuations, flipping events, and the unstable conformation following the flipping events. Overall, Fig. 6.5 provides a detailed picture of how different shapes are distributed in the course of simulation for isolated 1000-mer MC chains at elevated temperature.

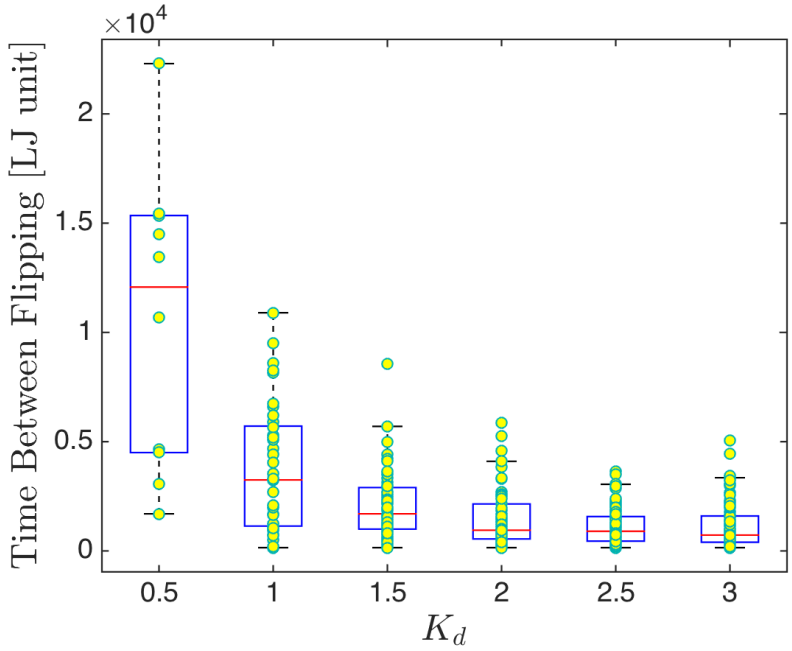

Figure 6.6: Box plot of the time between two flipping events with respect to the strength of the dihedral potential, K_{d}, at 50 °C. The box represents the interquartile range, which contains 50% of the values. The whiskers extend to cover 99.3% of the values. The line across the box is the median value. Each circle shows one data point, with the ones above the whiskers denoting outliers.

6.3.2 Effect of Dihedral Potential

The results discussed thus far were obtained using the coarse-grained MC model parameters proposed by Huang *et al.* [1] We also investigated how these results change as a function of the parameters, and identified the dihedral potential as the most important contributor to the flipping events. Thus, it is illuminating to understand, in a systematic way, the role of the dihedral potential. While arbitrary choices of the dihedral potential will no longer correspond to a coarse-grained MC chain, they allow us to understand the sensitivity of the results to the particular value of K_{d} appearing in the model by Huang *et al.* [1]

We first quantify the frequency of the flipping events as a function of the strength of the dihedral potential, K_{d}. We chose K_{d} to be in the range of 0 to 3.0, with a spacing

of 0.5. We also attempted to increase K_d further, but found that the calculations were infeasible due to the very small simulation time step required to prevent the simulation from crashing. For each K_d, we obtained at least three independent trajectories for at least $2.5 \times 10^4 \tau$. Figure 6.6 shows a box plot of the distribution of the time interval between flipping events as a function of K_d. Separate histograms (similar to Fig. 6.3) for the time interval distribution for each K_d are included in Appendix C. Note that $K_\mathrm{d} = 0$ is not included here because no flipping events were detected. Figure 6.6 shows that the median of the time between flipping decreases as K_d increases. Furthermore, beyond $K_\mathrm{d} = 2.0$, the time between flipping events reaches a plateau of around $1.0 \times 10^3 \tau$. Overall, Fig. 6.6 clearly shows the likelihood of flipping events strongly depends on K_d. Since flipping helps the MC chain explore different free energy minima and different collapsed states, we can conclude that K_d is a key factor in any MC model.

To more deeply understand the effects of K_d and the flipping on the shape of the collapsed states, we created box plots for the eigenvalues for different K_d in Fig. 6.7. Separate histograms for the eigenvalues for each K_d are available in Appendix C. The first thing to notice is that, similar to Fig. 6.5, the median values of the eigenvalues follow the pattern $\lambda_x < \lambda_y \approx \lambda_z$ for all K_d. Thus, the polymers are dominantly in a ring conformation, independent of K_d. There also seems to be two distinct regimes for λ_x. For $K_\mathrm{d} \leq 1.0$, the median of λ_x decreases to slightly below 1 nm. We suspect this is because i) the polymer stays mainly in a ring structure in this regime with few flipping events, and ii) the outer diameter of the ring increases with K_d in this regime, which can be observed from the increase in λ_y and λ_z. Thus the thickness of the ring, embodied in λ_x, decreases. Furthermore, in this regime, the number of outliers increases as K_d increases due to the increase of the frequency of flipping events. In contrast, for $K_\mathrm{d} > 1.0$, the aforementioned increase in the frequency of flipping events leads to both the median and the box size of λ_x increasing, corresponding to an increase in the probability of observing unstable loose conformations. Because the flipping events happen more often, λ_x values that indicate loose conformations and appear as outliers in box plots for $K_\mathrm{d} \leq 1.0$, start to blend into the middle quartiles of the distribution for larger values of K_d. As a result, we observe a concomitant decrease of the number of outliers for λ_x in the range of $K_\mathrm{d} > 1.0$ in Fig. 6.7. For λ_y and λ_z, both the median and the number of outliers increase as K_d goes up, which is also the result of the increase

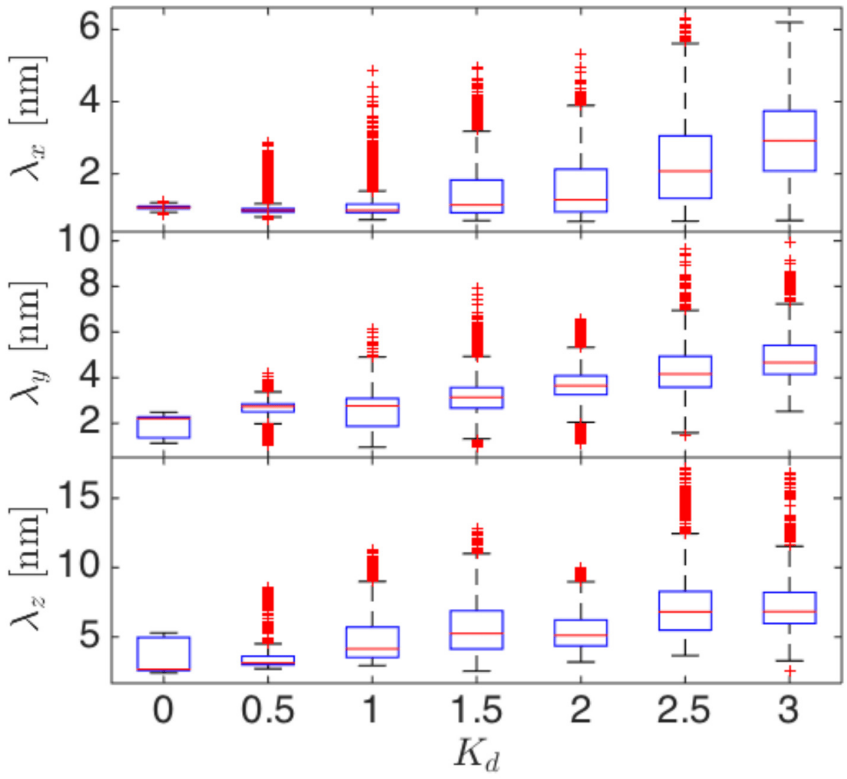

Figure 6.7: Box plots showing the eigenvalues of the gyration tensor ($\lambda_x \leq \lambda_y \leq \lambda_z$) as a function of the dihedral potential K_{d}. The description of the box plots are similar to Fig. 6.6, except here the outliers are plotted with the (red) cross symbols.

in the flipping frequency.

6.4 Implications for Fibril Assembly

Thus far, we have looked into isolated MC chains and shown how flipping events alter the conformation of a single MC chain. However, MC forms high aspect ratio fibrils in experiments, which involves multiple polymers. Although it is challenging to simulate such a large system, we can launch smaller simulations and still gain insights into the fibril formation mechanism.

We thus want to examine, in a simple system, how a flipping event can aid the

assembly of a series of smaller rings into a larger tubular structure, even if the initial condition is unlikely to be observed in practice. We return again to the model of Huang *et al.* where $K_{\mathrm{d}} = 2.0$. We first generate the initial ring conformation by simulating a single 1000-mer MC chain at $K_{\mathrm{d}} = 2.0$. Afterwards, we make four replicas and place the resulting five rings face-to-face in series with the center-to-center distance along the axis being 5.4 nm, as seen in the first snapshot of Fig. 6.8. We then launch the simulation for 3×10^7 steps with a time step of $1 \times 10^{-4}\tau$ under the same elevated temperature, 50 °C.

As the simulation proceeds, the three rings on the top form one tube, while the two rings on the bottom form a second tube. The blue chain, which is on the top of the first tube, then undergoes a flipping event, which allows this chain to stick out several "arms". Then one of the arms extended out to pass through the hollow space inside the first tube and reached the top face of the second tube. Because the loose conformation of the blue chain is not stable due to the self-attraction, it collapses back into the ring structure. In the meantime it drags the second tube to attach to its tube. After this, the unified proto-fibril relaxes into a more symmetric structure, and then the blue ring goes through another flipping event. The final structure is a single fibril-like structure that has grown in the longitude direction.

The dynamics in Fig. 6.8 imply that the flipping events induced by the dihedral potential can facilitate the assembly of distinct rings, which then aid in the formation of the longer fibril structure, in particular to facilitate alignment of the proto-fibrils. We also noticed in Fig. 6.8 that there are two flipping events in the course of the simulation, and the time interval between these two flipping events is 1430τ. This lies within the time interval distribution in Fig. 6.3 for $K_{\mathrm{d}} = 2.0$, albeit larger than the median. Furthermore, the two flipping events in Fig. 6.8 all occur for the blue chain, which has one face not attached to other chains. We thus suspect that the energy barrier imposed by neighboring chains plays an important role in the frequency of flipping. If a chain is trapped in the middle of the proto-fibril, it has a lower mobility that prevents escape from its neighbors and flipping into a loose structure. As such, the interior of the tube is stabilized by excluded volume and attractive interactions, while the faces of the tube can flip. The net result is a preference towards axial growth of the fibrils.

While Fig. 6.8 makes a promising connection between flipping of an individual chain

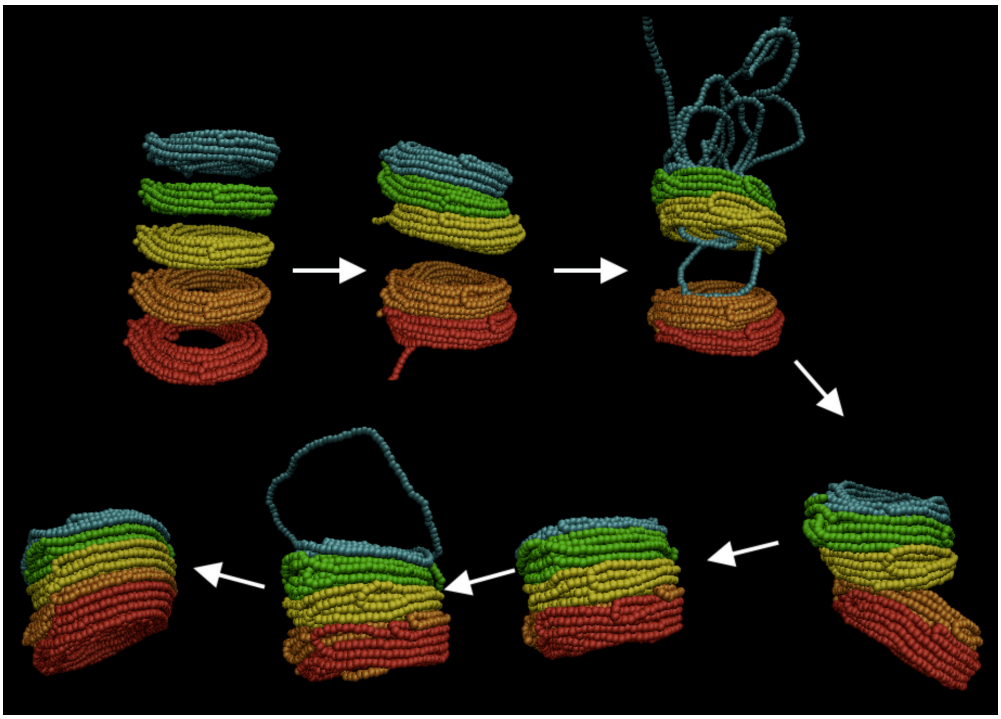

Figure 6.8: Snapshots of initial, intermediate, and final structures in the five-chain simulation at 50 °C. The ring is formed from a single chain simulation with N = 1000. The initial configuration was constructed by placing five replicas of a ring with the spacing of 5.4 nm. Each color represents one MC chain.

and the dynamics of fibril assembly, there are some limitations of the model and simulation method that attenuate our enthusiasm. First, in the course of simulation, we initiate the simulation by equilibrating the MC chain at 25 °C, and then make an instantaneous temperature jump to 50 °C to form the ring structure. This method ignores the time required for heat transfer, which is a potential but not critical problem. Heat transfer limitations certainly make the estimation of the time for chain to collapse inaccurate, but the presence of flipping events occurs at the higher temperature. Inasmuch as fibril formation takes longer than the time to heat the sample, flipping events could still play a role in the sol-gel transition.

Another important point to consider is the effect of MC concentration. Our single-chain simulations are always in the dilute limit based on the simulation box size. We

observed that as we increase the polymer concentration in our simulation and start from a relaxed solution at 25 °C, making a sudden temperature jump to 50 °C leads to the MC chains becoming entangled and trapped in a network, instead of forming rings. Moreover, the branches of these networks do not necessarily have a uniform diameter, contrary to what has been observed in experiments. The issue with polymer density is consistent with what we see in Fig. 6.8: the frequency of the flipping is reduced due to the more "crowded" environment, thus making it less likely to cross an energy barrier into another collapsed state. It is definitely of great interest to study a larger system with multiple MC chains at a higher concentration. Unfortunately, the system size and simulation time required to directly simulate gelation are well beyond the computational limits, and suggest the need for a simpler (or softer) model to study gelation.

A way to circumvent this situation and still get an estimate of fibril formation within the present model is to use biased initial configurations. Instead of a completely relaxed-coil structure, the initial configuration can be set as a partially relaxed and a partially ring-like structure. In reality, this initial configuration could correspond to a fast nucleation where part of an MC chain collapses while the rest remains as relaxed intermediate segments or dangling ends. In this way, it might be possible to study the interaction of MC chains at a higher concentration while also preventing the entangled structure. This biased initial configuration could also help to explain the "bridge" or "ghost" structure that connects two successive fibrils along the axial direction, as observed in experiments [45–47]. In these experiments, MC fibrils consist of collapsed rings that are interlinked by either stacking or wrapping onto the end of an existing fibril or another ring. It is further proposed that the MC gelation is a two-step process [47] beginning with a single or proto-tube nucleus as a precursor, and a secondary nucleation and growth process. This biased partial-coil/partial-relaxed initial configuration could act as an intermediate state in this two-step process to promote our understanding of the secondary nucleation while saving computational effort in the primary nucleation process. Alongside the flipping events, as we described above, this biased initial configuration could also contribute MC free ends that connect between different groups of proto-tubes. These free ends could also rationalize the "bridge" or "ghost" structure, as observed in experiments.

Finally, there is a potential problem using an implicit solvent. Although the gelation

in the coarse-grained model is driven by the self-attractive interactions of the MC chain at elevated temperature, in reality, the hydrogen bonding between hydroxy groups on MC and solvent molecules may play an important role in the gelation mechanism. The nature of the interactions between water and MC in the fibril state remains an open question. Moreover, the detailed structure of water in the vicinity of the polymer chains also may be important. While small angle neutron scattering data [46] indicate that the fibrils contain approximately 60% water, neither the latter experiments nor cryo-TEM measurements [45–47] provide definitive evidence that the water is located within the core of a fibril. It is possible that the water and the hydrophilic groups of MC are coordinated, rather than being locally segregated as predicted by the present model [1]. Addressing the details of water-MC coordination in the gel state by simulation is exceedingly difficult, requiring sufficient contour length to form a fibril while maintaining atomistic resolution. The requisite simulation is at least an order of magnitude larger than the atomistic simulations [130] used to determine the coarse-grained parameters used here [1], and likely infeasible. As a result, it is possible that while the coarse-grained, non-bonded interaction potentials could provide an accurate representation of that between a pair of very short MC chains, they may not capture the nature of the interactions in the relatively polymer-dense environment of the fibrils. Elucidating the detailed interactions between water and MC within a fibril may unlock a key step towards understanding fibril formation.

6.5 Concluding Remarks

Recent experimental studies [45, 46] on the structure of methylcellulose gels revealed a fibrillar morphology, overturning the prevailing model of physical gelation for these materials. While the experimental evidence for fibril formation is clear, the detailed gelation mechanism remains poorly understood. In this chapter, building on the simulation method in Huang *et al.* [1], we showed that the previously overlooked flipping events, characterized by a sudden chain conformational change, can help a methycellulose chain to re-orient itself from one collapsed state to another, as well as facilitating assembly of multiple rings into stacks. We further studied the methylcellulose chain with a range of dihedral potential strengths, and found that relatively small changes in

the strength of dihedral potential could have strong effects on the flipping statistics.

Chapter 7

Conclusion and Discussion

As stated, we aimed at applying computational tools to study biopolymers numerically with engineering applications. A large part of this dissertation focuses on stretching DNA in free solution and confinement, which is important for advancing genomic mapping technology. In this scenario, external forces and confinement are two fundamental and complementary aspects. We modified the pruned-enriched Rosenbluth method (PERM) that has been previously developed in our group to explore the force-extension behavior of DNA molecules. The most important finding of this work is that we showed the thickness of a DNA chain plays an important role in the force-extension behavior in various conditions, which has been overlooked in previous studies. Practically, our work provides handy tools for interpreting experiment results and developing efficient simulation models. We anticipate such a complete description of the force-extension of DNA will prove useful for the design of new genomic mapping technologies.

Methylcellulose is another biopolymer of interest. We applied Langevin Dynamics simulations to the computational study on the methylcellulose gelation mechanism. We found a "flipping" event, introduced purely by physical properties, that facilitates the assembly of the separated methylcellulose chains into a long fibril. This finding might be a key step in solving the mysterious gelation mechanism of methylcellulose.

Although the chemical details of these two biopolymers are different, they share common physical properties. Therefore, computer simulation is able to study various biopolymers with a generic model by changing the parameters and specifications. In

addition, our comprehensive simulation also provides opportunities to explore conditions which are rather challenging for experiments but crucial for understanding the basic properties of biopolymers, which allows building connections between theoretical predictions and experimental studies.

In Chapter 4, a single DNA chain is stretched in free solution. This allows separation of restriction imposed by forces from that by nanoconfinement. The key outcome of this part is a new formula that approximates the force-extension behavior with about 5% relative error for any value of the force. This is a great improvement comparing to the classic widely-used (more than 2,000 citations) Marko-Siggia interpolation formula [70]. The latter formula was developed for "phantom" polymers without thickness and usually has around 10% relative error for high forces, but is not even qualitatively correct for weak forces. The new formula developed in my dissertation is by far the best interpolation valid over the whole range of external forces.

Going beyond the case of an unconfined biopolymer, in Chapter 5 we applied this DNA model and simulation technique to explore the effect of slit confinement together with an external force. Most notably, this work predicted a new regime (phase) in the force-extension behavior under confinement with numerical evidence. This new regime features a mixed effect of both sensible DNA volume and sensible wall effects, thereby demonstrating conclusively that several previous works on modeling the stretching of confined DNA miss the key physics. Excitingly, after publishing this result, my theoretical prediction was confirmed by experimental evidence [71]. Moreover, we have constructed a phase diagram for the stretched DNA in nanoslits, which is straightforward to use in experiments.

In Chapter 6, we expect that in methycellulose simulations, the flipping events will prove particularly important for semiflexible self-attractive polymer models, as they help the chain to escape from local energy traps due to the intramolecular interactions. Furthermore, our study suggests that the flipping event facilitates the assembly of separated methylcellulose chains into forming a long fibril. We anticipate that this rapid conformational change will provide insights into the discrepancy between the high polymer concentrations, where fibrils form in experiments, and the low polymer concentration required in simulations in order to get ring-like structures.

Finally, we conclude by discussing the limitations of our work and future opportunities. In the discrete wormlike chain (DWLC) model, we consider DNA as a neutral polymer and use the hard-core potential to represent the real chain behavior. However, DNA is a polyelectrolyte that has negtive charges on the backbone. This leads to electrostatic interactions with ions in the buffer, charged surface, and the external electric field. Thus, lumping electrostatic effects into effective neutral parameters such as effective width w and persistence length l_{p} could potentially introduce errors, especially in the cases of low ionic strength and tight confinement where the overlapping of double layers becomes important. The challenge of including electrostatic interactions explicitly lies in solving the highly nonlinear Poisson-Boltzmann equation, as finding the numerical solution of a second-order partial differential equation for each step repeatedly could be extremely demanding in computational resources. Although substantial challenges exist, the future opportunity lies in moving forward to introducing computational practical electrostatic interactions, such as DNA-wall repulsion described by the linear Debye-Huckel potential. Other generic potentials, interpolations, or even a pre-computed table of solutions could also be introduced to give a better description of electrostatic interactions. In the simulation of methylcellulose, we are unable to directly simulate gelation at a higher concentration due to computational limits. Questions also remain regarding the dynamics of gelation as a function of heating rate, since we had an unrealistic sudden temperature jump throughout the simulation. Elucidating the connections between heat transfer, concentration and gelation is an interesting avenue for future studies and will provide new insights into the methylcellulose gelation mechanism.

Bibliography

[1] W. Huang, R. Ramesh, P. K. Jha, and R. G. Larson, "A systematic coarse-grained model for methylcellulose polymers: Spontaneous ring formation at elevated temperature," *Macromolecules*, vol. 49, no. 4, pp. 1490–1503, 2016.

[2] L. Dai, J. J. Jones, J. R. C. van der Maarel, and P. S. Doyle, "A systematic study of DNA conformation in slitlike confinement," *Soft Matter*, vol. 8, pp. 2972–2982, 2012.

[3] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.*, vol. 117, no. 1, pp. 1–19, 1995.

[4] J. D. Watson, F. H. Crick, *et al.*, "Molecular structure of nucleic acids," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[5] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, no. 5258, p. 561, 1970.

[6] R. Padmanabhan, E. Jay, and R. Wu, "Chemical synthesis of a primer and its use in the sequence analysis of the lysozyme gene of bacteriophage T4," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 71, no. 6, pp. 2510–2514, 1974.

[7] R. Wu, "Nucleotide sequence analysis of DNA," *Nature*, vol. 236, no. 68, pp. 198–200, 1972.

[8] E. Jay, R. Bambara, R. Padmanabhan, and R. Wu, "DNA sequence analysis: a general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping," *Nucleic Acids Res.*, vol. 1, no. 3, pp. 331–354, 1974.

[9] W. Ray, C. D. Tu, and R. Padmanabhan, "Nucleotide sequence analysis of DNA xii. the chemical synthesis and sequence analysis of a dodecadeoxynucleotide which binds to the endolysin gene of bacteriophage lambda," *Biochem. Biophys. Res. Commun.*, vol. 55, no. 4, pp. 1092–1099, 1973.

[10] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 74, no. 12, pp. 5463–5467, 1977.

[11] Illumina, *An Introduction to Next-Generation Sequencing Technology*. Illumina, Inc. (San Diego, CA).

[12] I. H. G. S. Consortium *et al.*, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, p. 860, 2001.

[13] I. H. G. S. Consortium *et al.*, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, no. 7011, p. 931, 2004.

[14] M. L. Metzker, "Sequencing technologiesthe next generation," *Nat. Rev. Genet.*, vol. 11, no. 1, p. 31, 2010.

[15] T. P. Niedringhaus, D. Milanova, M. B. Kerby, M. P. Snyder, and A. E. Barron, "Landscape of next-generation sequencing technologies," *Anal. Chem.*, vol. 83, no. 12, pp. 4327–4341, 2011.

[16] J. R. Miller, S. Koren, and G. Sutton, "Assembly algorithms for next-generation sequencing data," *Genomics*, vol. 95, no. 6, pp. 315–327, 2010.

[17] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, and R. A. Holt, "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[18] "Genomic Vision." http://www.genomicvision.com (accessed Feb 25, 2018).

[19] L. Feuk, A. R. Carson, and S. W. Scherer, "Structural variation in the human genome," *Nat. Rev. Genet.*, vol. 7, no. 2, pp. 85–97, 2006.

[20] F. Persson and J. O. Tegenfeldt, "DNA in nanochannelsdirectly visualizing genomic information," *Chem. Soc. Rev.*, vol. 39, no. 3, pp. 985–999, 2010.

[21] R. K. Neely, P. Dedecker, J.-i. Hotta, G. Urbanavičiūtė, S. Klimašauskas, and J. Hofkens, "DNA fluorocode: a single molecule, optical map of DNA with nanometre resolution," *Chem. Sci.*, vol. 1, no. 4, pp. 453–460, 2010.

[22] R. Redon, S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, *et al.*, "Global variation in copy number in the human genome," *Nature*, vol. 444, no. 7118, p. 444, 2006.

[23] A. J. Iafrate, L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer, and C. Lee, "Detection of large-scale variation in the human genome," *Nat. Genet.*, vol. 36, no. 9, p. 949, 2004.

[24] K. D. Dorfman, S. B. King, D. W. Olson, J. D. Thomas, and D. R. Tree, "Beyond gel electrophoresis: Microfluidic separations, fluorescence burst analysis, and DNA stretching," *Chem. Rev.*, vol. 113, no. 4, pp. 2584–2667, 2012.

[25] S. L. Levy and H. G. Craighead, "DNA manipulation, sorting, and mapping in nanofluidic systems," *Chem. Soc. Rev.*, vol. 39, no. 3, pp. 1133–1152, 2010.

[26] K. Jo, D. M. Dhingra, T. Odijk, J. J. de Pablo, M. D. Graham, R. Runnheim, D. Forrest, and D. C. Schwartz, "A single-molecule barcoding system using nanoslits for DNA analysis," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, no. 8, pp. 2673–2678, 2007.

[27] S. K. Das, M. D. Austin, M. C. Akana, P. Deshpande, H. Cao, and M. Xiao, "Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes," *Nucleic Acids Res.*, vol. 38, no. 18, pp. e177–e177, 2010.

[28] Y. Ebenstein, N. Gassman, S. Kim, J. Antelman, Y. Kim, S. Ho, R. Samuel, X. Michalet, and S. Weiss, "Lighting up individual DNA binding proteins with quantum dots," *Nano Lett.*, vol. 9, no. 4, pp. 1598–1603, 2009.

[29] W. Reisner, N. B. Larsen, A. Silahtaroglu, A. Kristensen, N. Tommerup, J. O. Tegenfeldt, and H. Flyvbjerg, "Single-molecule denaturation mapping of DNA in nanofluidic channels," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 30, pp. 13294–13299, 2010.

[30] M. Levy-Sakin and Y. Ebenstein, "Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy," *Curr. Opin. Biotech.*, vol. 24, no. 4, pp. 690–698, 2013.

[31] A. Bensimon, A. Simon, A. Chiffaudel, V. Croquette, F. Heslot, and D. Bensimon, "Alignment and sensitive detection of DNA by a moving interface," *Science*, vol. 265, no. 5181, pp. 2096–2098, 1994.

[32] J. Jing, J. Reed, J. Huang, X. Hu, V. Clarke, J. Edington, D. Housman, T. S. Anantharaman, E. J. Huff, B. Mishra, *et al.*, "Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, no. 14, pp. 8046–8051, 1998.

[33] E. Y. Chan, N. M. Goncalves, R. A. Haeusler, A. J. Hatch, J. W. Larson, A. M. Maletta, G. R. Yantz, E. D. Carstea, M. Fuchs, G. G. Wong, *et al.*, "DNA mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags," *Genome Res.*, vol. 14, no. 6, pp. 1137–1146, 2004.

[34] T. T. Perkins, D. E. Smith, and S. Chu, "Single polymer dynamics in an elongational flow," *Science*, vol. 276, no. 5321, pp. 2016–2021, 1997.

[35] "Bionano genomics." http://www.bionanogenomics.com (accessed Mar 1, 2018).

[36] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin, "Statics and dynamics of single DNA molecules confined in nanochannels," *Phys. Rev. Lett.*, vol. 94, no. 19, p. 196101, 2005.

[37] A. Balducci, P. Mao, J. Han, and P. S. Doyle, "Double-stranded DNA diffusion in slitlike nanochannels," *Macromolecules*, vol. 39, no. 18, pp. 6273–6281, 2006.

[38] C.-C. Hsieh, A. Balducci, and P. S. Doyle, "Ionic effects on the equilibrium dynamics of DNA confined in nanoslits," *Nano Lett.*, vol. 8, no. 6, pp. 1683–1688, 2008.

[39] J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, *et al.*, "The dynamics of genomic-length DNA molecules in 100-nm channels," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 30, pp. 10979–10983, 2004.

[40] H. Cao, J. O. Tegenfeldt, R. H. Austin, and S. Y. Chou, "Gradient nanostructures for interfacing microfluidics and nanofluidics," *Appl. Phys. Lett.*, vol. 81, no. 16, pp. 3058–3060, 2002.

[41] D. Klemm, B. Heublein, H.-P. Fink, and A. Bohn, "Cellulose: Fascinating biopolymer and sustainable raw material," *Angew. Chem. Int. Ed.*, vol. 44, no. 22, pp. 3358–3393, 2005.

[42] S. Kamel, N. Ali, K. Jahangir, S. Shah, and A. El-Gendy, "Pharmaceutical significance of cellulose: A review," *Express Polym. Lett.*, vol. 2, no. 11, pp. 758–778, 2008.

[43] "Dow answer center." https://dowac.custhelp.com (accessed Feb 24, 2017).

[44] M. Takahashi, M. Shimazaki, and J. Yamamoto, "Thermoreversible gelation and phase separation in aqueous methyl cellulose solutions," *J. Polym. Sci. Part B Polym. Phys.*, vol. 39, no. 1, pp. 91–100, 2001.

[45] J. R. Lott, J. W. McAllister, S. A. Arvidson, F. S. Bates, and T. P. Lodge, "Fibrillar structure of methylcellulose hydrogels," *Biomacromolecules*, vol. 14, no. 8, pp. 2484–2488, 2013.

[46] J. R. Lott, J. W. McAllister, M. Wasbrough, R. L. Sammler, F. S. Bates, and T. P. Lodge, "Fibrillar structure in aqueous methylcellulose solutions and gels," *Macromolecules*, vol. 46, no. 24, pp. 9760–9771, 2013.

[47] J. W. McAllister, P. W. Schmidt, K. D. Dorfman, T. P. Lodge, and F. S. Bates, "Thermodynamics of aqueous methylcellulose solutions," *Macromolecules*, vol. 48, no. 19, pp. 7205–7215, 2015.

[48] S. Arvidson, J. Lott, J. McAllister, J. Zhang, F. Bates, T. Lodge, R. Sammler, Y. Li, and M. Brackhagen, "Interplay of phase separation and thermoreversible gelation in aqueous methylcellulose solutions," *Macromolecules*, vol. 46, no. 1, pp. 300–309, 2012.

[49] N. Sarkar, "Kinetics of thermal gelation of methylcellulose and hydroxypropylmethylcellulose in aqueous solutions," *Carbohydr. Polym.*, vol. 26, no. 3, pp. 195–203, 1995.

[50] C. Chevillard and M. A. Axelos, "Phase separation of aqueous solution of methylcellulose," *Colloid Polym. Sci.*, vol. 275, no. 6, pp. 537–545, 1997.

[51] M. Hirrien, C. Chevillard, J. Desbrieres, M. Axelos, and M. Rinaudo, "Thermogelation of methylcelluloses: New evidence for understanding the gelation mechanism," *Polymer*, vol. 39, no. 25, pp. 6251–6259, 1998.

[52] K. Kobayashi, C.-I. Huang, and T. P. Lodge, "Thermoreversible gelation of aqueous methylcellulose solutions," *Macromolecules*, vol. 32, no. 21, pp. 7070–7077, 1999.

[53] L. Li, P. Thangamathesvaran, C. Yue, K. Tam, X. Hu, and Y. Lam, "Gel network structure of methylcellulose in water," *Langmuir*, vol. 17, no. 26, pp. 8062–8068, 2001.

[54] J. P. A. Fairclough, H. Yu, O. Kelly, A. J. Ryan, R. L. Sammler, and M. Radler, "Interplay between gelation and phase separation in aqueous solutions of methylcellulose and hydroxypropylmethylcellulose," *Langmuir*, vol. 28, no. 28, pp. 10551–10557, 2012.

[55] T. Chatterjee, A. I. Nakatani, R. Adden, M. Brackhagen, D. Redwine, H. Shen, Y. Li, T. Wilson, and R. L. Sammler, "Structure and properties of aqueous methylcellulose gels by small-angle neutron scattering," *Biomacromolecules*, vol. 13, no. 10, pp. 3355–3369, 2012.

[56] L. Li, "Thermal gelation of methylcellulose in water: Scaling and thermoreversibility," *Macromolecules*, vol. 35, no. 15, pp. 5990–5998, 2002.

[57] L. Li, H. Shan, C. Yue, Y. Lam, K. Tam, and X. Hu, "Thermally induced association and dissociation of methylcellulose in aqueous solutions," *Langmuir*, vol. 18, no. 20, pp. 7291–7298, 2002.

[58] J. Desbrieres, M. Hirrien, and S. Ross-Murphy, "Thermogelation of methylcellulose: Rheological considerations," *Polymer*, vol. 41, no. 7, pp. 2451–2461, 2000.

[59] M. A. Villetti, V. Soldi, C. Rochas, and R. Borsali, "Phase-separation kinetics and mechanism in a methylcellulose/salt aqueous solution studied by time-resolved small-angle light scattering (SALS)," *Macromol. Chem. Phys.*, vol. 212, no. 10, pp. 1063–1071, 2011.

[60] S. Guillot, D. Lairez, and M. Axelos, "Non-self-similar aggregation of methylcellulose," *J. Appl. Crystallogr.*, vol. 33, no. 3, pp. 669–672, 2000.

[61] H. Tanaka, "Viscoelastic phase separation," *J. Phys. Condens. Matter*, vol. 12, no. 15, p. R207, 2000.

[62] J. W. McAllister, J. R. Lott, P. W. Schmidt, R. L. Sammler, F. S. Bates, and T. P. Lodge, "Linear and nonlinear rheological behavior of fibrillar methylcellulose hydrogels," *ACS Macro Lett.*, vol. 4, no. 5, pp. 538–542, 2015.

[63] R. Bodvik, A. Dedinaite, L. Karlson, M. Bergström, P. Bäverbäck, J. S. Pedersen, K. Edwards, G. Karlsson, I. Varga, and P. M. Claesson, "Aggregation and network formation of aqueous methylcellulose and hydroxypropylmethylcellulose solutions," *Colloids Surf. A*, vol. 354, no. 1, pp. 162–171, 2010.

[64] J. V. Shah and P. A. Janmey, "Strain hardening of fibrin gels and plasma clots," *Rheol. Acta*, vol. 36, no. 3, pp. 262–268, 1997.

[65] V. K. Praveen, S. J. George, R. Varghese, C. Vijayakumar, and A. Ajayaghosh, "Self-assembled π-nanotapes as donor scaffolds for selective and thermally gated fluorescence resonance energy transfer (FRET)," *J. Am. Chem. Soc.*, vol. 128, no. 23, pp. 7542–7550, 2006.

[66] F. MacKintosh, J. Käs, and P. Janmey, "Elasticity of semiflexible biopolymer networks," *Phys. Rev. Lett.*, vol. 75, no. 24, p. 4425, 1995.

[67] P. J. Flory, "Phase equilibria in solutions of rod-like particles," *Proc. Roy. Soc. A*, vol. 234, no. 1196, pp. 73–89, 1956.

[68] P. J. Flory, "Statistical thermodynamics of semi-flexible chain molecules," *Proc. Roy. Soc. A*, vol. 234, no. 1196, pp. 60–73, 1956.

[69] V. V. Ginzburg, R. L. Sammler, W. Huang, and R. G. Larson, "Anisotropic self-assembly and gelation in aqueous methylcellulose – theory and modeling," *J. Polym. Sci. Part B Polym. Phys.*, vol. 54, no. 16, pp. 1624–1636, 2016.

[70] J. F. Marko and E. D. Siggia, "Stretching DNA," *Macromolecules*, vol. 28, no. 26, pp. 8759–8770, 1995.

[71] J.-W. Yeh and K. Szeto, "Stretching of tethered DNA in nanoslits," *ACS Macro Lett.*, vol. 5, no. 10, pp. 1114–1118, 2016.

[72] S. B. Smith, L. Finzi, and C. Bustamante, "Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads," *Science*, vol. 258, no. 5085, pp. 1122–1126, 1992.

[73] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, "Single-molecule studies of DNA mechanics," *Curr. Opin. Struc. Biol.*, vol. 10, no. 3, pp. 279–285, 2000.

[74] W. Reisner, J. N. Pedersen, and R. H. Austin, "DNA confinement in nanochannels: physics and biological applications," *Rep. Prog. Phys.*, vol. 75, no. 10, p. 106601, 2012.

[75] J. S. Pedersen, "Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting," *Adv. Colloid Interface Sci.*, vol. 70, pp. 171–210, 1997.

[76] M. Doi, *The theory of polymer dynamics*. No. 73, Oxford University Press, 1988.

[77] M. Rubinstein and R. Colby, *Polymers Physics*. Oxford, 2003.

[78] C. Storm and P. Nelson, "Theory of high-force DNA stretching and overstretching," *Phys. Rev. E*, vol. 67, no. 5, p. 051906, 2003.

[79] J. Wang and H. Gao, "A generalized bead-rod model for Brownian dynamics simulations of wormlike chains under strong confinement," *J. Chem. Phys.*, vol. 123, no. 8, p. 084906, 2005.

[80] Y. Wang, D. R. Tree, and K. D. Dorfman, "Simulation of DNA extension in nanochannels," *Macromolecules*, vol. 44, no. 16, pp. 6594–6604, 2011.

[81] D. R. Tree, Y. Wang, and K. D. Dorfman, "Extension of DNA in a nanochannel as a rod-to-coil transition," *Phys. Rev. Lett.*, vol. 110, no. 20, p. 208103, 2013.

[82] C. C. Hsieh, A. G. Balducci, and P. S. Doyle, "Ionic effects on the equilibrium dynamics of DNA confined in nanoslits," *Nano Lett.*, vol. 8, pp. 1683–1688, 2008.

[83] D. R. Tree, A. Muralidhar, P. S. Doyle, and K. D. Dorfman, "Is DNA a good model polymer?," *Macromolecules*, vol. 46, no. 20, pp. 8369–8382, 2013.

[84] D. Stigter, "A comparison of Manning's polyelectrolyte theory with the cylindrical gouy model," *J. Chem. Phys.*, vol. 82, no. 14, pp. 1603–1606, 1978.

[85] I. Wong and C.-M. Ho, "Surface molecular property modifications for poly (dimethylsiloxane)(PDMS) based microfluidic devices," *Microfluid Nanafluid*, vol. 7, no. 3, pp. 291–306, 2009.

[86] S. H. Behrens and D. G. Grier, "The charge of glass and silica surfaces," *J. Chem. Phys.*, vol. 115, no. 14, pp. 6716–6721, 2001.

[87] A. V. Dobrynin, "Effect of counterion condensation on rigidity of semiflexible polyelectrolytes," *Macromolecules*, vol. 39, no. 26, pp. 9519–9527, 2006.

[88] D. Stigter, "Interactions of highly charged colloidal cylinders with applications to double-stranded DNA," *Biopolymers*, vol. 16, no. 7, pp. 1435–1448, 1977.

[89] M. Rubinstein and R. H. Colby, *Polymer physics*, vol. 23. Oxford University Press New York, 2003.

[90] N. Saitô, K. Takahashi, and Y. Yunoki, "The statistical mechanical theory of stiff chains," *J. Phys. Soc. JPN*, vol. 22, no. 1, pp. 219–226, 1967.

[91] H. Benoit and P. Doty, "Light scattering from non-gaussian chains," *J. Phys. Chem. U.S.*, vol. 57, no. 9, pp. 958–963, 1953.

[92] P.-G. De Gennes, *Scaling concepts in polymer physics*. Cornell university press, 1979.

[93] L. Schäfer, *Excluded volume effects in polymer solutions: as explained by the renormalization group*. Springer Science & Business Media, 2012.

[94] J. Des Cloizeaux, "On the absence of flory terms in the energy and in the entropy of a polymer chain," *J. Phys. Paris*, vol. 37, no. 5, pp. 431–434, 1976.

[95] B. Li, N. Madras, and A. D. Sokal, "Critical exponents, hyperscaling, and universal amplitude ratios for two-and three-dimensional self-avoiding walks," *J. Stat. Phys.*, vol. 80, no. 3-4, pp. 661–754, 1995.

[96] N. Clisby, "Accurate estimate of the critical exponent ν for self-avoiding walks via a fast implementation of the pivot algorithm," *Phys. Rev. Lett.*, vol. 104, no. 5, p. 055702, 2010.

[97] T. T. Perkins, D. E. Smith, R. G. Larson, and S. Chu, "Stretching of a single tethered polymer in a uniform flow," *Science*, vol. 268, no. 5207, pp. 83–87, 1995.

[98] J. M. Schurr and S. B. Smith, "Theory for the extension of a linear polyelectrolyte attached at one end in an electric field," *Biopolymers*, vol. 29, no. 8-9, pp. 1161–1165, 1990.

[99] S. B. Smith and A. J. Bendich, "Electrophoretic charge density and persistence length of DNA as measured by fluorescence microscopy," *Biopolymers*, vol. 29, no. 8-9, pp. 1167–1173, 1990.

[100] L. Dai and P. S. Doyle, "Comparisons of a polymer in confinement versus applied force," *Macromolecules*, vol. 46, no. 15, pp. 6336–6344, 2013.

[101] P. Pincus, "Excluded volume effects and stretched polymer chains," *Macromolecules*, vol. 9, no. 3, pp. 386–388, 1976.

[102] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle," *Recl. Trav. Chim. Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 1949.

[103] S. B. Smith, Y. Cui, and C. Bustamante, "Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules," *Science*, pp. 795–799, 1996.

[104] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching DNA with optical tweezers," *Biophys. J.*, vol. 72, no. 3, pp. 1335–1346, 1997.

[105] A. R. Klotz, M. Mamaev, L. Duong, H. W. de Haan, and W. W. Reisner, "Correlated fluctuations of DNA between nanofluidic entropic traps," *Macromolecules*, vol. 48, pp. 4742–4747, 2015.

[106] L. Dai, D. R. Tree, J. R. van der Maarel, K. D. Dorfman, and P. S. Doyle, "Revisiting blob theory for DNA diffusivity in slitlike confinement," *Phys. Rev. Lett.*, vol. 110, no. 16, p. 168105, 2013.

[107] G. K. Cheong, X. Li, and K. D. Dorfman, "Evidence for the extended de gennes regime of a semiflexible polymer in slit confinement," *Phys. Rev. E*, vol. 97, no. 2, p. 022502, 2018.

[108] T. Odijk, "Scaling theory of DNA confined in nanochannels and nanoslits," *Phys. Rev. E*, vol. 77, no. 6, p. 060901, 2008.

[109] M. Daoud and P. G. de Gennes, "Statistics of macromolecular solutions trapped in small pores," *J. Phys. Paris*, vol. 38, pp. 85–93, 1977.

[110] D. R. Tree, W. F. Reinhart, and K. D. Dorfman, "The Odijk regime in slits," *Macromolecules*, vol. 47, no. 11, pp. 3672–3684, 2014.

[111] Y.-L. Chen, M. Graham, J. de Pablo, G. Randall, M. Gupta, and P. Doyle, "Conformation and dynamics of single DNA molecules in parallel-plate slit microchannels," *Phys. Rev. E*, vol. 70, no. 6, p. 060901, 2004.

[112] D. J. Bonthuis, C. Meyer, D. Stein, and C. Dekker, "Conformation and dynamics of DNA confined in slitlike nanofluidic channels,"

[113] J. Tang, S. L. Levy, D. W. Trahan, J. J. Jones, H. G. Craighead, and P. S. Doyle, "Revisiting the conformation and dynamics of DNA in slitlike confinement," *Macromolecules*, vol. 43, no. 17, pp. 7368–7377, 2010.

[114] A. Taloni, J.-W. Yeh, and C.-F. Chou, "Scaling theory of stretched polymers in nanoslits," *Macromolecules*, vol. 46, pp. 7989–8002, 2013.

[115] Y.-L. Chen, P.-K. Lin, and C.-F. Chou, "Generalized force-extension relation for wormlike chains in slit confinement," *Macromolecules*, vol. 43, pp. 10204–10207, 2010.

[116] P.-K. Lin, J.-F. Chang, C.-H. Wei, P. Tsao, W. Fann, Y.-L. Chen, *et al.*, "Partial hydrodynamic screening of confined linear and circular double-stranded DNA dynamics," *Phys. Rev. E*, vol. 84, no. 3, p. 031917, 2011.

[117] J.-W. Yeh, A. Taloni, Y.-L. Chen, and C.-F. Chou, "Entropy-driven single moelcule tug-of-war of DNA at micro-nanofluidic interfaces," *Nano Lett.*, vol. 12, pp. 1597–1602, 2012.

[118] H.-P. Hsu and K. Binder, "Semiflexible macromolecules with discrete bond angles confined in nanoslits: A Monte Carlo test of scaling concepts," *Macromolecules*, vol. 46, no. 19, pp. 8017–8025, 2013.

[119] K. L. Kounovsky-Shafer, J. P. Hernandez-Ortiz, K. Jo, T. Odijk, J. J. de Pablo, and D. C. Schwartz, "Presentation of large DNA molecules for analysis as nanoconfined dumbbells," *Macromolecules*, vol. 46, pp. 8356–8368, 2013.

[120] R. Marie, J. N. Pedersen, D. L. V. Bauer, K. H. Rasmussen, M. Yusuf, E. Volpi, H. Flyvbjerg, A. Kristensen, and K. U. Mir, "Integrated view of genome structure and sequence of a single DNA molecule in a nanofluidic device," *Proc. Natl. Acad. Sci. USA*, vol. 110, pp. 4893–4898, 2013.

[121] H.-P. Hsu and K. Binder, "Stretching semiflexible polymer chains: Evidence for the importance of excluded volume effects from Monte Carlo simulation," *J. Chem. Phys.*, vol. 136, no. 2, p. 024901, 2012.

[122] H. W. de Haan and T. N. Shendruk, "Force-extension of DNA in a nanoslit: Mapping between the 3D and 2D limits," *ACS Macro Lett.*, vol. 4, pp. 632–635, 2015.

[123] W. Reisner, N. B. Larsen, H. Flyvbjerg, J. O. Tegenfeldt, and A. Kristensen, "Directed self-organization of single DNA molecules in a nanoslit via embedded nanopit arrays," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, pp. 79–84, 2009.

[124] J. T. Del Bonis-O'Donnell, W. Reisner, and D. Stein, "Pressure-driven DNA transport across an artificial nanotopography," *New J. Phys.*, vol. 11, p. 075032, 2009.

[125] A. R. Klotz, H. B. Brandao, and W. W. Reisner, "Diffusion resonance of nanoconfined polymers," *Macromolecules*, vol. 45, pp. 2122–2127, 2012.

[126] A. R. Klotz, M. Mamaev, H. W. de Haan, J. Z. Y. Chen, and W. W. Reisner, "Measuring the confinement free energy and effective width of single polymer chains via single-molecule tetris," *Macromolecules*, vol. 48, pp. 5028–5033, 2015.

[127] A. G. Balducci, J. Tang, and P. S. Doyle, "Electrophoretic stretching of DNA molecules in cross-slot nanoslit channels," *Macromolecules*, vol. 41, no. 24, pp. 9914–9918, 2008.

[128] W. Huang, "Multi-scale modeling of cellulosic polymers for optimal drug delivery properties in solid dispersion formulation," 2017.

[129] M. Kong, I. Saha Dalal, G. Li, and R. G. Larson, "Systematic coarse-graining of the dynamics of self-attractive semiflexible polymers," *Macromolecules*, vol. 47, no. 4, pp. 1494–1502, 2014.

[130] W. Huang, I. Saha Dalal, and R. G. Larson, "Analysis of solvation and gelation behavior of methylcellulose using atomistic molecular dynamics simulations," *J. Phys. Chem. B*, vol. 118, no. 48, pp. 13992–14008, 2014.

[131] H. Noguchi, S. Saito, S. Kidoaki, and K. Yoshikawa, "Self-organized nanostructures constructed with a single polymer chain," *Chem. Phys. Lett.*, vol. 261, no. 4-5, pp. 527–533, 1996.

[132] D. T. Seaton, S. Schnabel, D. P. Landau, and M. Bachmann, "From flexible to stiff: Systematic analysis of structural phases for single semiflexible polymers," *Phys. Rev. Lett.*, vol. 110, no. 2, p. 028103, 2013.

[133] H. Noguchi and K. Yoshikawa, "Folding path in a semiflexible homopolymer chain: A Brownian dynamics simulation," *J. Chem. Phys.*, vol. 113, no. 2, pp. 854–862, 2000.

[134] K. Kremer and K. Binder, "Dynamics of polymer chains confined into tubes: Scaling theory and Monte Carlo simulations," *J. Chem. Phys.*, vol. 81, no. 12, pp. 6381–6394, 1984.

[135] N. Metropolis and S. Ulam, "The Monte Carlo method," *J. Am. Stat. Assoc.*, vol. 44, no. 247, pp. 335–341, 1949.

[136] T. Prellberg, "From Rosenbluth sampling to PERM-rare event sampling with stochastic growth algorithms," *arXiv preprint arXiv:1205.3734*, 2012.

[137] A. Muralidhar, *Equilibrium properties of DNA and other semiflexible polymers confined in nanochannels*. PhD thesis, University of Minnesota, 2016.

[138] M. N. Rosenbluth and A. W. Rosenbluth, "Monte Carlo calculation of the average extension of molecular chains," *J. Chem. Phys.*, vol. 23, no. 2, pp. 356–359, 1955.

[139] H.-P. Hsu and P. Grassberger, "A review of Monte Carlo simulations of polymers with PERM," *J. Stat. Phys*, vol. 144, no. 3, pp. 597–637, 2011.

[140] T. Prellberg and J. Krawczyk, "Flat histogram version of the pruned and enriched Rosenbluth method," *Phys. Rev. Lett.*, vol. 92, no. 12, p. 120602, 2004.

[141] P. Grassberger, "Pruned-enriched Rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000," *Phys. Rev. E*, vol. 56, no. 3, p. 3682, 1997.

[142] F. Wall and J. J. Erpenbeck, "New method for the statistical computation of polymer dimensions," *J. Chem. Phys.*, vol. 30, no. 3, pp. 634–637, 1959.

[143] T. Vilgis, P. Haronska, and M. Benhamou, "Branched polymers in restricted geometry: Flory theory, scaling and blobs," *J. Phys. II*, vol. 4, no. 12, pp. 2187–2196, 1994.

[144] H.-P. Hsu and P. Grassberger, "2-dimensional polymers confined in a strip," *Eur. Phys. J. B*, vol. 36, no. 2, pp. 209–214, 2003.

[145] H.-P. Hsu and P. Grassberger, "Polymers confined between two parallel plane walls," *J. Chem. Phys.*, vol. 120, no. 4, pp. 2034–2041, 2004.

[146] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford University Press, 2017.

[147] X. Li, C. M. Schroeder, and K. D. Dorfman, "Modeling the stretching of wormlike chains in the presence of excluded volume," *Soft Matter*, vol. 11, no. 29, pp. 5947–5954, 2015.

[148] C. Bustamante, J. F. Marko, E. D. Siggia, S. Smith, *et al.*, "Entropic elasticity of lambda-phage DNA," *Science*, pp. 1599–1599, 1994.

[149] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron, "DNA: an extensible molecule," *Science*, pp. 792–794, 1996.

[150] E. S. Shaqfeh, "The dynamics of single-molecule DNA in flow," *J. Non-Newton Fluid*, vol. 130, no. 1, pp. 1–28, 2005.

[151] J. R. Graham and M. T. Leary, "A review of empirical capital structure research and directions for the future," *Annu. Rev. Financ. Econ.*, vol. 3, no. 1, pp. 309–345, 2011.

[152] F. Latinwo and C. M. Schroeder, "Model systems for single molecule polymer dynamics," *Soft Matter*, vol. 7, no. 18, pp. 7907–7913, 2011.

[153] O. A. Saleh, D. McIntosh, P. Pincus, and N. Ribeck, "Nonlinear low-force elasticity of single-stranded DNA molecules," *Phys. Rev. Lett.*, vol. 102, no. 6, p. 068301, 2009.

[154] A. Dittmore, D. B. McIntosh, S. Halliday, and O. A. Saleh, "Single-molecule elasticity measurements of the onset of excluded volume in poly (ethylene glycol)," *Phys. Rev. Lett.*, vol. 107, no. 14, p. 148301, 2011.

[155] M. J. Stevens, D. B. McIntosh, and O. A. Saleh, "Simulations of stretching a strong, flexible polyelectrolyte: using long chains to access the Pincus scaling regime," *Macromolecules*, vol. 46, no. 15, pp. 6369–6373, 2013.

[156] A. Muralidhar, D. R. Tree, and K. D. Dorfman, "Backfolding of wormlike chains confined in nanochannels," *Macromolecules*, vol. 47, no. 23, pp. 8446–8458, 2014.

[157] A. Muralidhar, D. R. Tree, Y. Wang, and K. D. Dorfman, "Interplay between chain stiffness and excluded volume of semiflexible polymers confined in nanochannels," *J. Chem. Phys.*, vol. 140, no. 8, p. 084905, 2014.

[158] A. Muralidhar and K. D. Dorfman, "Kirkwood diffusivity of long semiflexible chains in nanochannel confinement," *Macromolecules*, vol. 48, no. 8, pp. 2829–2839, 2015.

[159] A. V. Dobrynin, J.-M. Y. Carrillo, and M. Rubinstein, "Chains are more flexible under tension," *Macromolecules*, vol. 43, no. 21, pp. 9181–9190, 2010.

[160] R. Radhakrishnan and P. T. Underhill, "Models of flexible polymers in good solvents: relaxation and coil–stretch transition," *Soft Matter*, vol. 8, no. 26, pp. 6991–7003, 2012.

[161] C. Brockman, S. J. Kim, and C. M. Schroeder, "Direct observation of single flexible polymers using single stranded DNA," *Soft Matter*, vol. 7, no. 18, pp. 8005–8012, 2011.

[162] R. G. Larson, *The structure and rheology of complex fluids*, vol. 150. Oxford University Press New York, 1999.

[163] T. A. Knotts IV, N. Rathore, D. C. Schwartz, and J. J. de Pablo, "A coarse grain model for DNA," *J. Chem. Phys.*, vol. 126, no. 8, p. 02B611, 2007.

[164] T. E. Ouldridge, A. A. Louis, and J. P. Doye, "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model," *J. Chem. Phys.*, vol. 134, no. 8, p. 02B627, 2011.

[165] F. Oesterhelt, M. Rief, and H. Gaub, "Single molecule force spectroscopy by AFM indicates helical structure of poly (ethylene-glycol) in water," *New J. Phys.*, vol. 1, no. 1, p. 6, 1999.

[166] X. Li and K. D. Dorfman, "Effect of excluded volume on the force-extension of wormlike chains in slit confinement," *J. Chem. Phys.*, vol. 144, no. 10, p. 104902, 2016.

[167] T. Odijk, "Scaling theory of stretched polymers in nanoslits," *Phys. Rev. E*, vol. 77, p. 060901(R), 2008.

[168] X. Li, C. M. Schroeder, and K. D. Dorfman, "Modeling the stretching of wormlike chains in the presence of excluded volume," *Soft Matter*, vol. 11, pp. 5947–5954, 2015.

[169] H.-P. Hsu, W. Paul, and K. Binder, "Polymer chain stiffness vs. excluded volume: A Monte Carlo study of the crossover towards the worm-like chain model," *Europhys. Lett.*, vol. 92, p. 28003, 2010.

[170] J. O. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, and R. H. Austin, "The dynamics of genomic-length DNA molecules in 100-nm channels," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 30, pp. 10979–10983, 2004.

[171] K. H. Rasmussen, R. Marie, J. M. Lange, W. E. Svendsen, A. Kristensen, and K. U. Mir, "A device for extraction, manipulation and stretching of DNA from single human chromosomes," *Lab Chip*, vol. 11, pp. 1431–1433, 2011.

[172] C. Freitag, C. Noble, J. Fritzsche, F. Persson, M. Reiter-Schad, A. N. Nilsson, A. Graneli, T. Ambjornsson, K. U. Mir, and J. O. Tegenfeldt, "Visualizing the entire DNA from a chromosome in a single frame," *Biomicrofluidics*, vol. 9, p. 044114, 2015.

[173] D. Stigter, "Interactions of highly charged colloidal cylinders with applications to double-stranded DNA," *Biopolymers*, vol. 16, pp. 1435–1448, 1977.

[174] P. Mao and J. Han, "Fabrication and characterization of 20 nm planar nanofluidic channels by glass-glass and glass-silicon bonding," *Lab Chip*, vol. 5, pp. 837–844, 2005.

[175] J. Lin, F. Persson, J. Fritzsche, J. O. Tegenfeldt, and O. A. Saleh, "Bandpass filtering of DNA elastic modes using confinement and tension," *Biophys. J.*, vol. 102, no. 1, pp. 96–100, 2012.

[176] X. Li, F. S. Bates, and K. D. Dorfman, "Rapid conformational fluctuations in a model of methylcellulose," *Phys. Rev. Mat.*, vol. 1, no. 2, p. 025604, 2017.

[177] N. Sarkar, "Thermal gelation properties of methyl and hydroxypropyl methylcellulose," *J. Appl. Polym. Sci.*, vol. 24, no. 4, pp. 1073–1087, 1979.

[178] W. L. Mattice and U. W. Suter, *Conformational Theory of Large Molecules: the Rotational Isomeric State Model in Macromolecular Systems*. Wiley-Interscience, 1994.

[179] D. N. Theodorou and U. W. Suter, "Shape of unperturbed linear polymers: Polypropylene," *Macromolecules*, vol. 18, no. 6, pp. 1206–1214, 1985.

[180] K. Šolc, "Shape of a random-flight chain," *J. Chem. Phys.*, vol. 55, no. 1, pp. 335–344, 1971.

Appendix A

Supporting Infomation to Chapter 4

A.1 Fractional Extension vs. Contour Length

Figure A.1 provides representative data showing that the fractional extension z becomes independent of contour length L for sufficiently long chains.

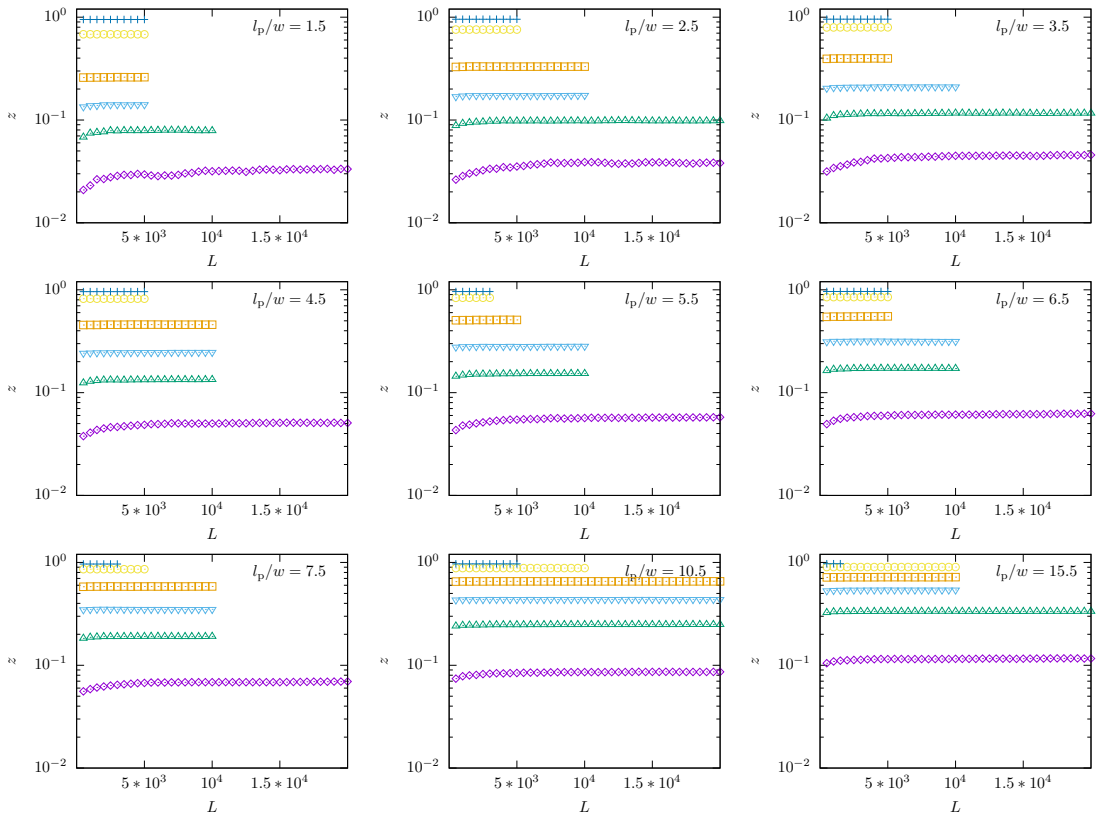

Figure A.1: Fractional extension z as a function of contour length $L = (N_b - 1)w$ for N_b beads of size w for different values of the force Fw/l_p: 0.0105 (purple $\diamond$), 0.0373 (green $\triangle$), 0.0833 (cyan $\triangledown$), 0.2152 (orange $\square$), 1.7100 (yellow $\circ$), 18.0000 (blue $+$). Each panel corresponds to a different value of l_p/w.

Tables A.1 and A.2 provide the number of beads N_b used for each data point in Chapter 4.

Table A.1: Number of beads used for the data in Figs. 4.1, 4.3, 4.4. The values in the first row are different l_p/w. The simulation data for the freely-jointed chain (l_p/w = 1.0) are only used for Fig. 4.4.

Fw/l_p	1.0	1.5	2.5	3.5	4.5	5.5	6.5	7.5
0.0000	40001	40001	40001	40001	40001	40001	40001	40001
0.0105	40001	20001	20001	20001	40001	40001	20001	20001
0.0164	40001	20001	20001	20001	40001	80001	20001	20001
0.0227	40001	10001	20001	20001	20001	10001	20001	20001
0.0296	40001	10001	20001	20001	20001	10001	20001	20001
0.0373	20001	10001	20001	20001	10001	10001	10001	10001
0.0461	20001	10001	10001	20001	10001	10001	20001	10001
0.0563	20001	10001	10001	10001	10001	10001	10001	10001
0.0684	20001	10001	10001	10001	10001	5001	10001	10001
0.0833	20001	5001	10001	10001	10001	10001	10001	10001
0.1023	20001	5001	10001	5001	10001	10001	5001	10001
0.1275	10001	5001	5001	20001	10001	5001	5001	10001
0.1627	10001	10001	5001	5001	10001	5001	5001	10001
0.2152	10001	5001	10001	5001	10001	5001	5001	10001
0.3000	10001	5001	5001	5001	5001	3201	5001	5001
0.453	10001	5001	5001	5001	5001	5001	5001	5001
0.7807	10001	5001	5001	5001	5001	5001	5001	5001
1.7100	5001	5001	5001	5001	5001	3201	5001	5001
6.7133	3201	5001	5001	5001	5001	3201	5001	3201
18.000	5001	5001	5001	5001	5001	3201	5001	3201

Fw/l_{p}	10.5	15.5
0.0000	40001	40001
0.0105	40001	80001
0.0164	40001	40001
0.0227	40001	40001
0.0296	20001	40001
0.0373	20001	40001
0.0461	10001	40001
0.0563	20001	40001
0.0684	20001	40001
0.0833	20001	10001
0.1023	10001	20001
0.1275	10001	20001
0.1627	20001	20001
0.2152	20001	10001
0.3000	20001	5001
0.453	10001	5001
0.7807	10001	10001
1.7100	10001	10001
6.7133	10001	10001
18.000	5001	5001

Table A.2: Number of beads used for Fig. 4.2

l_p/w = 1.5		l_p/w = 2.5		l_p/w = 3.5		l_p/w = 4.5	
Fw/l_p	L	Fw/l_p	L	Fw/l_p	L	Fw/l_p	L
0.0189	40001	0.0144	20001	0.0121	40001	0.0106	40001
0.0529	20001	0.0400	10001	0.0610	20001	0.0291	40001
0.0980	20001	0.0735	10001	0.0794	20001	0.0532	20001
0.1536	20001	0.1143	10001	0.0944	10001	0.0821	20001
0.2202	10001	0.1625	10001	0.1337	20001	0.1158	5001
0.2989	10001	0.2187	10001	0.1791	5001	0.1547	10001
0.3920	10001	0.2842	5001	0.2315	5001	0.1992	5001
0.5027	5001	0.3609	5001	0.2922	5001	0.2504	5001
0.6361	5001	0.4518	5001	0.3634	5001	0.3101	5001
0.8002	5001	0.5616	5001	0.4485	5001	0.3806	5001
1.0078	5001	0.6980	5001	0.5528	3201	0.4664	5001
1.2803	5001	0.8739	5001	0.6856	3201	0.5745	5001
1.6557	5001	1.1118	5001	0.8630	3201	0.7174	5001
2.2059	5001	1.4552	5001	1.1158	3201	0.9193	5001
3.0819	5001	1.9944	5001	1.5089	3201	1.2303	5001
4.6453	5001	2.9464	5001	2.1971	5001	1.7710	5001
7.9517	5001	4.9446	5001	3.6327	5001	2.8932	3201
17.2787	5001	10.5555	5001	7.6490	5001	6.0228	3201
67.3484	5001	40.6124	5001	29.1270	5001	22.7337	3201

l_p/w = 5.5		l_p/w = 6.5		l_p/w = 7.5		l_p/w = 10.5	
Fw/l_p	L	Fw/l_p	L	Fw/l_p	L	Fw/l_p	L
0.0095	40001	0.0087	40001	0.0081	40001	0.0068	40001
0.0262	40001	0.0240	20001	0.0222	20001	0.0186	40001
0.0477	10001	0.0436	20001	0.0404	20001	0.0337	20001
0.0735	10001	0.0670	20001	0.0620	10001	0.0517	40001
0.1034	10001	0.0942	20001	0.0870	10001	0.0723	20001
0.1378	10001	0.1253	10001	0.1135	5001	0.0957	10001
0.1770	5001	0.1606	10001	0.1478	5001	0.1220	10001
0.2218	5001	0.2008	10001	0.1845	5001	0.1517	10001
0.2737	5001	0.2471	10001	0.2265	5001	0.1853	10001
0.3347	5001	0.3012	10001	0.2754	5001	0.2241	5001
0.4083	5001	0.3660	5001	0.3338	3201	0.2697	5001
0.5002	5001	0.4466	5001	0.4057	3201	0.3252	5001
0.6208	5001	0.5514	5001	0.4989	3201	0.3960	5001
0.7897	5001	0.6973	5001	0.6276	3201	0.4924	5001
1.0481	5001	0.9189	3201	0.8220	3201	0.6359	3201
1.4945	5001	1.2997	3201	1.1545	3201	0.8780	3201
2.4167	5001	2.0832	3201	1.8361	3201	1.3697	3201
4.9815	5001	4.2560	3201	3.7224	3201	2.7220	3201
18.6584	5001	15.8327	3201	13.7577	3201	9.8951	3201

l_p/w = 15.5	
Fw/l_p	L
0.0056	10001
0.0152	10001
0.0275	10001
0.0420	5001
0.0586	5001
0.0772	10001
0.0981	5001
0.1214	5001
0.1477	5001
0.1775	5001
0.2121	5001
0.2536	5001
0.3056	5001
0.3751	5001
0.4765	3201
0.6450	3201
0.9826	3201
1.9033	3201
6.7673	5001

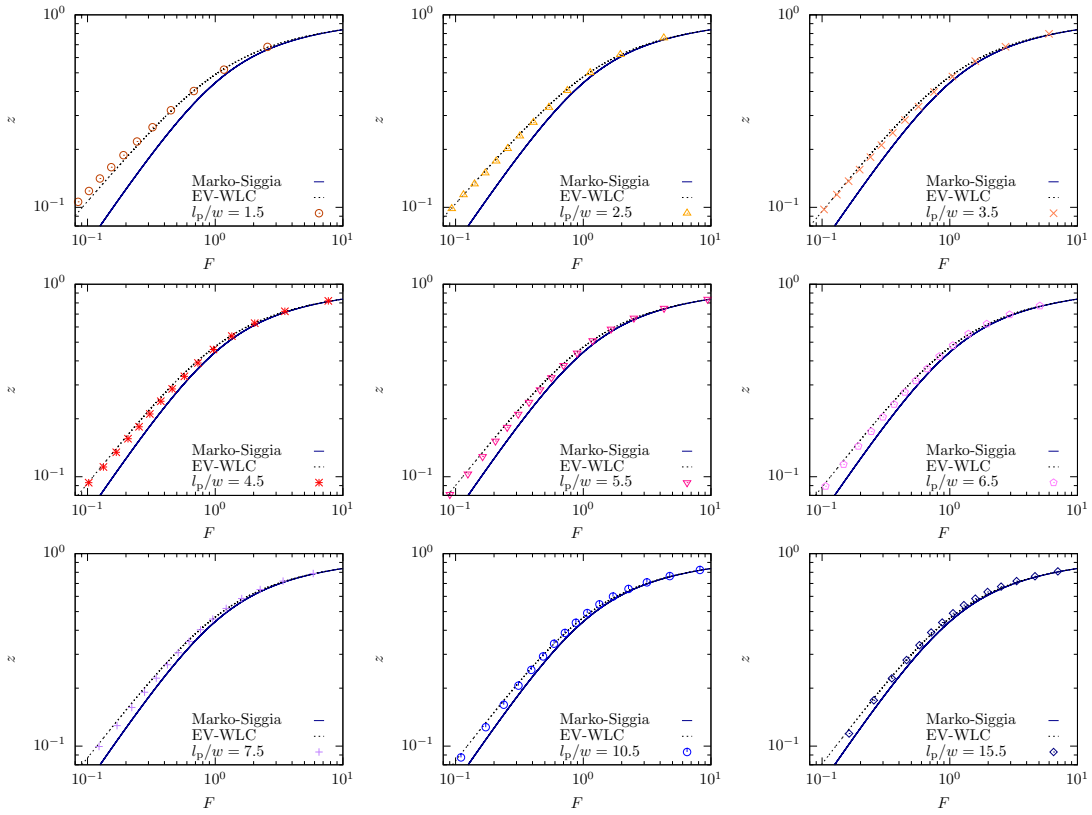

Figure A.2: Figure 4.1 of the main text for all values of l_p/w.

A.2 Comparison between Simulation Data and Interpolations with Various Monomer Anisotropies

Figure A.2 is the equivalent of Fig. 4.1 in the main text for all values of l_p/w and Fig. A.3 is the equivalent of Fig. 4.2 in the main text for all values of l_p/w.

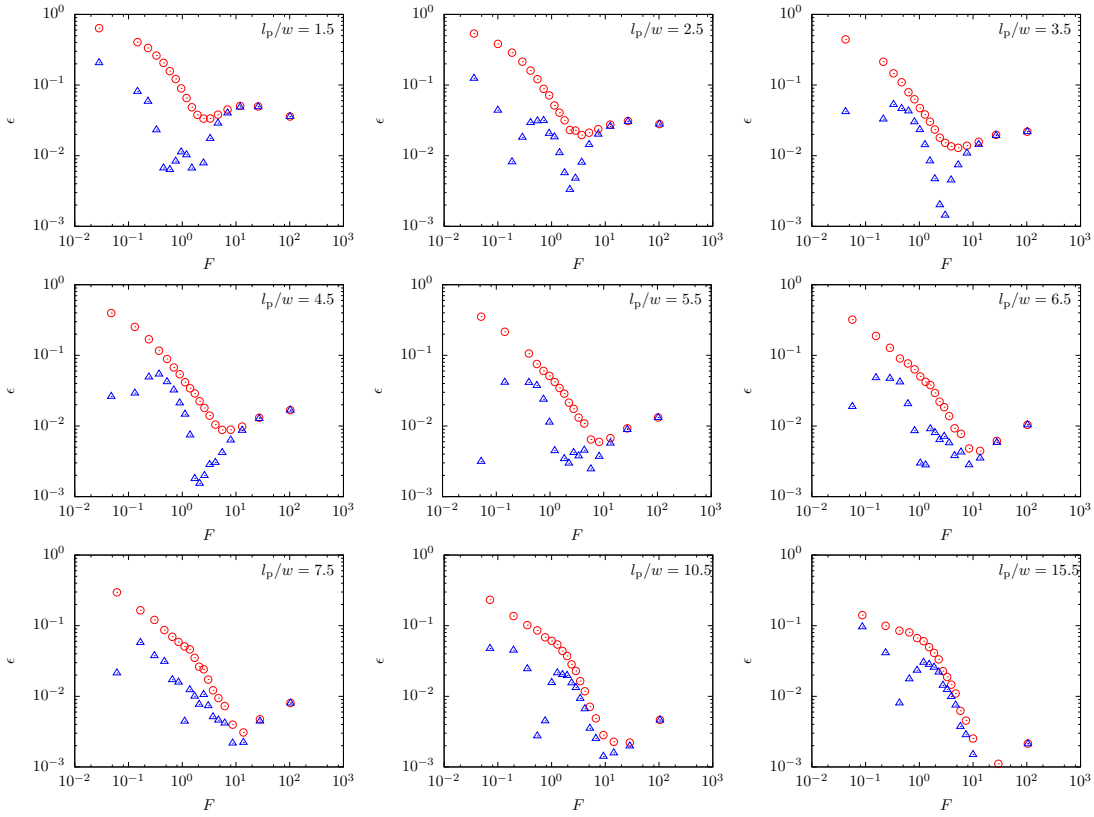

Figure A.3: Figure 4.2 of the main text for all values of l_p/w for the Marko-Siggia interpolation formula (red $\circ$) and the EV-WLC interpolation formula (blue $\triangle$). Each panel corresponds to a different value of l_p/w.

A.3 Interpolation Error with Various l_p/w

Appendix B

Supporting Infomation to Chapter 5

B.1 Scaling Theory

B.1.1 Details of the Derivation in Chapter 5.2

For sufficiently long chains, Odijk [167] showed that excluded volume is important for all slit heights in the absence of a force. His scaling theory predicts that the root-mean-squared end-to-end distance of a chain confined to a slit is

$$R \cong L^{3/4} \left(\frac{w l_{\mathrm{p}}}{H} \right)^{1/4}. \tag{B.1}$$

Odijk's theory is valid for any slit height H, but the prefactor depends on whether the chain consists of blobs ($H \gtrsim l_{\mathrm{p}}$) or deflection segments ($H \lesssim l_{\mathrm{p}}$) [167]. This discrepancy is removed by replacing the persistence length of the chain, l_{p}, with the effective persistence length, $l_{\mathrm{p},\parallel}$,

$$R \cong L^{3/4} \left(\frac{w l_{\mathrm{p},\parallel}}{H} \right)^{1/4}. \tag{B.2}$$

The effective persistence length is obtained from the characteristic decay of the bond vector autocorrelation function in confinement, as described in the main text. Note that Eq. B.2 also has an additional correction for strong confinement due to the orientation

of the deflection segments [110]. This additional correction is small, so we ignore it at the level of approximation here.

In the presence of a weak force, Taloni *et al.* [114] suggest that the confined chain can be considered as a series of Pincus (tensile) blobs of size [101]

$$\xi_{\mathrm{T}} = \frac{k_{\mathrm{B}}T}{f}. \tag{B.3}$$

The idea in Pincus's theory [101] applied to confined chains [114] is that the contour length L_{b} inside a Pincus blob of size ξ_{T} is the same as the contour length inside a subchain of size R. Setting Eq. B.2 equal to Eq. B.3 and solving for the contour length, we find that the section of the chain inside each Pincus blob is of length

$$L_{\mathrm{b}} \cong l_{\mathrm{p},\parallel} \left(\frac{k_{\mathrm{B}}T}{f l_{\mathrm{p},\parallel}}\right)^{4/3} \left(\frac{H}{w}\right)^{1/3}. \tag{B.4}$$

The number of such blobs in a chain of total contour length L is $N = L/L_{\mathrm{b}}$ and each blob is extended a distance ξ_{T} in the direction of the force. In this confined Pincus regime, the total extension of the chain in the direction of the force is $Z = N\xi_{\mathrm{T}}$, leading to [114]

$$z \cong \left(\frac{f l_{\mathrm{p},\parallel}}{k_{\mathrm{B}}T}\right)^{1/3} \left(\frac{w}{H}\right)^{1/3}, \tag{B.5}$$

where $z \equiv Z/L$ is the fractional extension of the chain in the direction of the force. As Eq. B.5 combines the results for confined chains and Pincus theory, it seems reasonable to refer to it as the confined Pincus regime.

The lower bound for the confined Pincus regime corresponds to short chains that cannot form many Pincus blobs [114]. In other words, the contour length of the chain needs to exceed the contour length inside the Pincus blob. At the scaling level, the minimum contour length is simply the molecular weight of a Pincus blob given by Eq. B.4,

$$L_{\min} \sim l_{\mathrm{p},\parallel} \left(\frac{k_{\mathrm{B}}T}{f l_{\mathrm{p},\parallel}}\right)^{4/3} \left(\frac{H}{w}\right)^{1/3}. \tag{B.6}$$

The upper bound of the confined Pincus regime is a more subtle point, depending on the degree of confinement but not the contour length. Taloni *et al.* [114] proposed that

the confined Pincus regime ends when the Pincus blob size is equal to the slit height, leading to

$$f_{\max}^{\text{weak}} \cong \frac{k_{\text{B}}T}{H} \tag{B.7}$$

and, with Eq. B.5, a maximum fractional extension

$$z_{\max}^{\text{weak}} \cong \left(\frac{l_{\text{p},\parallel} w}{H^2}\right)^{1/3}. \tag{B.8}$$

However, if we recall the theory for the Pincus regime in free solution [101], we note that a Pincus regime can also end if the contour length inside the Pincus blob is too short to make a self-avoiding random walk. For the case of confinement, this alternate upper bound corresponds to the case where the contour length L_{b} given by Eq. B.4 is shorter than the minimum contour length for real chain behavior in Odijk's theory for confined chains. The latter restriction on the contour length is $H l_{\text{p},\parallel}/w$, which corresponds to the crossover between the excluded volume scaling in Eq. B.2 and the ideal chain scaling $R \sim (L l_{\text{p},\parallel})^{1/2}$. Setting $L_{\text{b}} = H l_{\text{p},\parallel}/w$ leads to

$$f_{\max}^{\text{strong}} \cong \frac{k_{\text{B}}T}{l_{\text{p},\parallel}} \left(\frac{w}{H}\right)^{1/2} \tag{B.9}$$

and corresponding fractional extension

$$z_{\max}^{\text{strong}} \cong \left(\frac{w}{H}\right)^{1/2}. \tag{B.10}$$

Our choice of superscript labels in Eqs. B.7-B.10 suggests when these different upper bounds are applicable. Explicitly, the ratio of the maximum forces is

$$\frac{f_{\max}^{\text{weak}}}{f_{\max}^{\text{strong}}} = \left(\frac{l_{\text{p},\parallel}^2}{Hw}\right)^{1/2}. \tag{B.11}$$

Weak confinement corresponds to the case $H \gg l_{\text{p}}^2/w$, in which case we expect $f_{\max}^{\text{weak}}$ in Eq. B.7 to be the correct maximum force. In strong confinement, it proves useful to rewrite Eq. B.11 as

$$\frac{f_{\max}^{\text{weak}}}{f_{\max}^{\text{strong}}} = \left(\frac{l_{\text{p},\parallel}/w}{H/l_{\text{p},\parallel}}\right)^{1/2}. \tag{B.12}$$

By definition, wormlike chains correspond to the case $l_\mathrm{p}/w > 1$. Since strong confinement further requires that $H \ll l_\mathrm{p}$, we expect that Eq. B.9 is the correct description for the maximum force in small slits.

B.1.2 Derivation of the Confined Pincus Regime Using $\nu \neq 3/5$

Here we derive a somewhat more sophisticated version of the confined Pincus regime where we use a 3D Flory exponent $\nu = 0.588$ and the 2D Flory exponent $\nu = 3/4$. In this case, the 3D blob has the Flory radius

$$R_F = L^\nu w^{2\nu-1} l_\mathrm{p}^{2-3\nu} \tag{B.13}$$

The contour length L_* inside a blob of size $R_F = H$ is then

$$L_* \sim \left(H w^{1-2\nu} l_\mathrm{p}^{3\nu-2}\right)^{1/\nu} \tag{B.14}$$

The size of the confined chain is then

$$R \sim \left(\frac{L}{L_*}\right)^{3/4} H \tag{B.15}$$

whereupon the Odijk result for the size of a confined chain in Eq. B.1 becomes

$$R \sim \left[\frac{L}{(H w^{1-2\nu} l_\mathrm{p}^{3\nu-2})^{1/\nu}}\right]^{3/4} H \tag{B.16}$$

Equation B.16 reduces to Eq. B.1 for $\nu = 3/5$.

If we set R in Eq. B.16 equal to the tensile blob size $k_\mathrm{B}T/f$, we find that each Pincus blob now contains a contour length

$$L_b \sim \left(\frac{k_\mathrm{B}T}{fH}\right)^{4/3} \left(H w^{1-2\nu} l_\mathrm{p}^{3\nu-2}\right)^{1/\nu} \tag{B.17}$$

The latter result reduces to Eq. B.4 for $\nu = 3/5$.

The extension of the chain follows from the derivation for $\nu = 3/5$, giving

$$z \sim \frac{k_\mathrm{B}T}{f L_b} \tag{B.18}$$

which gives

$$z \sim \left(\frac{f}{k_{\mathrm{B}}T}\right)^{1/3} H^{(4/3)-1/\nu} w^{2-(1/\nu)} l_{\mathrm{p}}^{(2/\nu)-3} \tag{B.19}$$

The latter equation reduces to Eq. B.5 for $\nu = 3/5$. If we use $\nu = 0.588$, we have

$$z \sim \left(\frac{f}{k_{\mathrm{B}}T}\right)^{1/3} H^{-0.367} w^{0.299} l_{\mathrm{p}}^{0.401} \tag{B.20}$$

B.2 List of Simulation Parameters

Tables B.1, B.2, B.3 and B.4 provide the number of beads N_{b} used for each data point in Chapter 5.

Table B.1: Number of beads used for Figs. 5.2, 5.5, 5.6, 5.7, 5.8.

	l_p/w = 5.5	l_p/w = 5.5	l_p/w = 5.5	l_p/w = 10.5	l_p/w = 10.5
	H/w = 4	H/w = 49	H/w = 99	H/w = 4	H/w = 149
fw/k_BT	N_b	N_b	N_b	N_b	N_b
0.0005	-	40001	40001	-	40001
0.001	40001	40001	40001	40001	40001
0.002	40001	40001	40001	40001	40001
0.003	40001	40001	40001	40001	40001
0.004	40001	40001	40001	40001	40001
0.005	40001	40001	40001	40001	40001
0.006	40001	40001	40001	40001	40001
0.007	40001	40001	40000	40001	40001
0.008	40001	40001	40001	40001	40001
0.009	40001	40001	40001	40001	40001
0.0105	20001	20001	20001	20001	20001
0.0164	20001	20001	20001	20001	20001
0.0227	20001	20001	20001	20001	20001
0.0296	20001	20001	20001	20001	20001
0.0373	20001	20001	20001	20001	20001
0.0461	20001	20001	20001	20001	20001
0.0563	20001	20001	20001	20001	20001
0.0684	20001	20001	20001	20001	20001
0.0833	20001	5001	10001	20001	20001
0.1023	10001	5001	10001	5001	20001
0.1275	10001	5001	10001	5001	-
0.1627	10001	5001	10001	5001	-
0.2152	10001	5001	10001	5001	-
0.3	10001	5001	10001	5001	-
0.453	10001	5001	10001	5001	-
0.7807	10001	5001	10001	5001	-
1.71	10001	5001	10001	5001	-
6.7133	10001	5001	10001	5001	-
18	3201	5001	10001	5001	-

	l_p/w = 15.5	l_p/w = 15.5
	H/w = 4	H/w = 299
fw/k_BT	N_b	N_b
0.0005	-	-
0.001	-	40001
0.002	-	40001
0.003	-	40001
0.004	-	40001
0.005	-	40001
0.006	-	40001
0.007	-	40001
0.008	-	40001
0.009	-	40001
0.0105	10001	-
0.0164	10001	20001
0.0227	5001	20001
0.0296	20001	20001
0.0373	5001	20001
0.0461	5001	20001
0.0563	5001	20001
0.0684	5001	20001
0.0833	5001	20001
0.1023	5001	20001
0.1275	5001	-
0.1627	5001	-
0.2152	5001	-
0.3	5001	-
0.453	5001	-
0.7807	5001	-
1.71	5001	-
6.7133	5001	-
18	5001	-

Table B.2: Number of beads used for the data in Fig. 5.3. The simulations are for the freely-jointed chain ($l_\mathrm{p}/w = 1.0$) with $H/w = 19$.

$fw/k_\mathrm{B}T$	N_b
0.0027	20001
0.0040	20001
0.0059	20001
0.0087	10001
0.0129	10001
0.0190	10001
0.0279	10001
0.0412	10001
0.0607	10001
0.0895	5001
0.1318	5001
0.1943	5001
0.2863	5001
0.4220	5001
0.6219	5001
0.9165	5001
1.3507	5001
1.9905	5001

Table B.3: Number of beads used for the data of ideal chains in Figs. 5.6, 5.7.

	$l_\mathrm{p}/w = 5.5, H/w = 99$	$l_\mathrm{p}/w = 10.5, H/w = 4$
$fw/k_\mathrm{B}T$	N_b	N_b
0	20001	20001
0.0005	20001	20001
0.001	20001	20001
0.002	20001	20001
0.003	20001	20001
0.004	20001	20001
0.005	20001	20001
0.006	20001	20001
0.007	20001	20001
0.008	20001	20001
0.009	20001	20001
0.0105	20001	20001
0.0164	10001	10001
0.0227	10001	10001
0.0296	10001	10001
0.0373	10001	10001
0.0461	10001	10001
0.0563	10001	10001
0.0684	10001	10001
0.0833	10001	10001
0.1023	5001	5001
0.1275	5001	5001
0.1627	5001	5001
0.2152	5001	5001
0.3	5001	5001
0.453	5001	5001
0.7807	5001	5001
1.71	5001	5001
6.7133	5001	5001
18	5001	5001

Table B.4: Number of beads used for the data of ideal chains in Fig. 5.10.

	$l_\mathrm{p}/w = 10.5, H/w = 49$	$l_\mathrm{p}/w = 10.5, H/w = 99$	$l_\mathrm{p}/w = 15.5, H/w = 49$
$fw/k_\mathrm{B}T$	N_b	N_b	N_b
0.0005	40001	40001	40001
0.001	40001	40001	40001
0.002	40001	40001	40001
0.003	40001	40001	40001
0.004	40001	40001	40001
0.005	40001	40001	40001
0.006	40001	40001	40001
0.007	40001	40001	40001
0.008	40001	40001	40001
0.009	40001	-	40001
0.0105	20001	-	20001
0.0164	20001	-	20001
0.0227	20001	-	20001
0.0296	20001	-	20001
0.0373	20001	-	20001
0.0461	20001	-	20001
0.0563	20001	-	20001
0.0684	20001	-	20001
0.0833	20001	-	20001
0.1023	5001	-	10001
0.1275	5001	-	10001
0.1627	5001	-	10001
0.2152	5001	-	10001
0.3	5001	-	10001
0.453	5001	-	10001
0.7807	5001	-	10001
1.71	5001	-	10001
6.7133	5001	-	10001
18	5001	-	10001

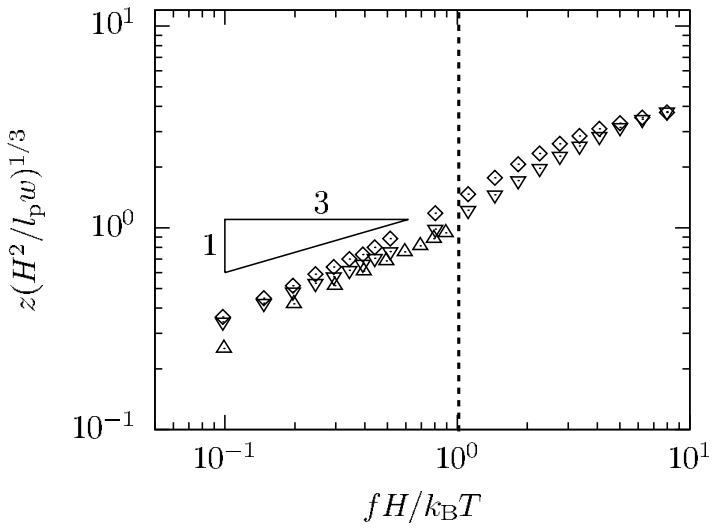

Figure B.1: Plot of the rescaled extension versus the rescaled force for different values of l_p/w and H. The rescaled axes are chosen to test the strong confinement crossovers given by Eqs. B.7 and B.8. The symbols are simulation results for intermediate confinement with triangle: l_p/w = 10.5, H/w = 99; inverted triangle: l_p/w = 10.5, H/w = 49 and diamond: l_p/w = 14.5, H/w = 49. The black triangle shows the scaling law in Eq. B.5. The dashed line shows the crossover given by Eq. B.7.

B.3 Additional Figures

Figure B.1 replots the data in Fig. 5.4 of the main text using the crossovers for weak confinement.

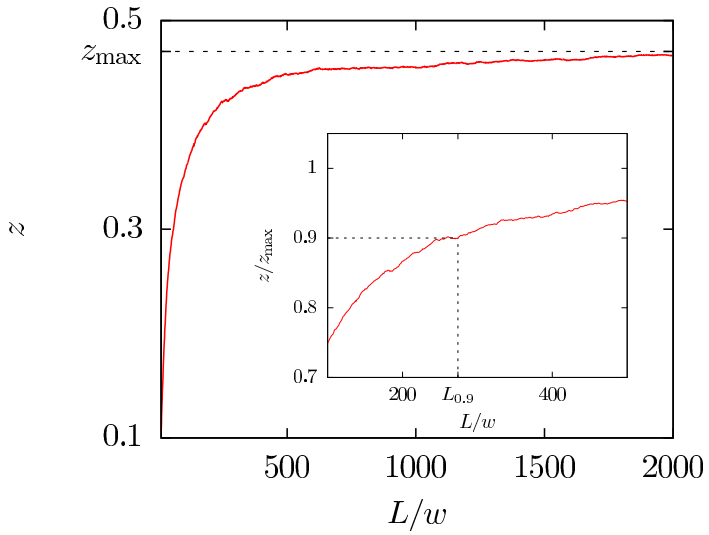

Figure B.2: Plot of the fractional extension versus contour length for l_p/w = 10.5 and H/w = 4 under force f = 0.0296 in confined Pincus regime. The dashed line shows the saturated extension $z_{\max}$, i.e., where z reaches its maximum and is independent of contour length. Inset: Plot of the fractional extension rescaled by $z_{\max}$ versus contour length. Dashed lines show the contour length $L_{0.9}$ where the relative extension reaches 90% of its maximum.

Figure B.2 shows how the result in Fig. 5.5 of the main text is generated. The saturation value of the extension is only gradually reached as a function of contour length. For this value of the force, short lengths correspond to the weak stretching (Hookean) regime while the long lengths correspond to the confined Pincus regime. To identify the onset of the confined Pincus regime, we need to determine when the fractional extension $z(L)$ is close enough to its saturation value $z_{\max} = z(L \to \infty)$.

We chose to use the point at which the fractional extension reaches 90% of the saturation value. The main panel of Fig. B.2 shows the fractional extension as a function of molecular weight. We then divide by the value of z at the largest molecular weight, taking this value to be the saturation value $z_{\max}$. The inset then shows $z/z_{\max}$. The point at which this ratio is 0.9 is defined as the length $L_{0.9}$.

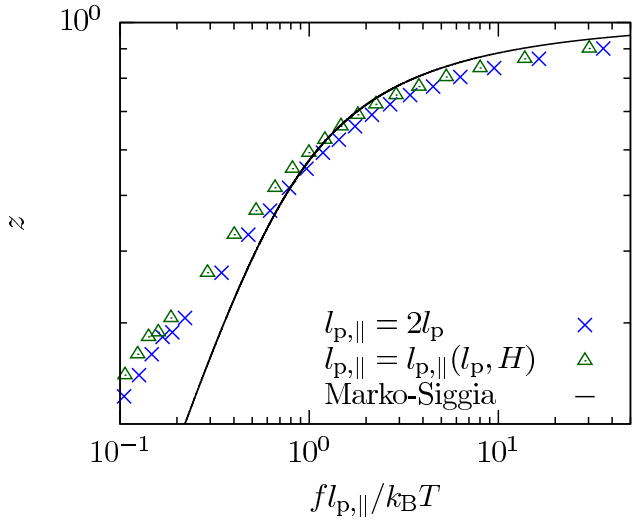

Figure B.3: Plot of the fractional extension versus the rescaled force for strong confinement for $l_\mathrm{p}/w = 10.5$ and $H/w = 4$ for real chains. Green triangles show the simulation data with $l_{\mathrm{p},\parallel}$ as a function of both H and l_p. For the latter data, we assign $l_{\mathrm{p},\parallel} = 17.696w$ or, equivalently, $l_{\mathrm{p},\parallel} = 1.6853l_\mathrm{p}$, following [2]

Figure B.3 shows how the result in Fig. 5.6 of the main text is affected by using a more accurate value of $l_{\mathrm{p},\parallel}$ for the particular value of H/w.

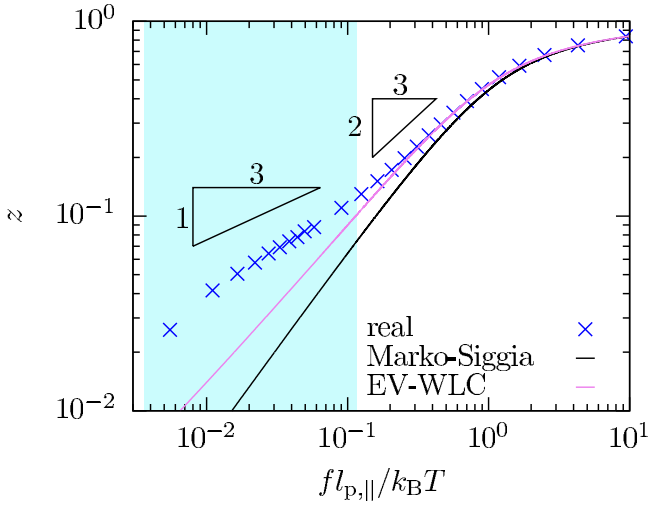

Figure B.4: Plot of the fractional extension versus rescaled force for weak confinement for $l_{\mathrm{p}}/w = 5.5$ and $H/w = 49$. The solid black line is the 3D Marko-Siggia formula and the solid purple line is the EV-WLC interpolation formula. The shaded area shows the confined Pincus regime.

Figure B.4 shows a relatively flexible WLC confined in a wide slit. This figure reinforces the conclusions reached in Fig. 5.7 of the main text.

Appendix C

Supporting Infomation to Chapter 6

C.1 List of Simulation Parameters

We summarize the mole fractions of methylcellulose monomer substitution type in Table C.1, the bonded coarse-grained model parameters in Table C.2 and the nonbonded parameters in Table C.3.

Table C.1: Average Mole Fraction of Methylcellulose Monomers in METHOCEL A Chemistry [1] [a]

monomer type	mol %
cellulose	0.05
2-MC	0.13
3-MC	0.02
6-MC	0.10
2,3-MC	0.10
2,6-MC	0.26
3,6-MC	0.05
2,3,6-MC	0.29

[a] The degree of substitution (DS) is 1.83.

Table C.2: Summary of the bonded parameters [1]

	Dimensional Units		Dimensionless Units	
l_0	0.515	nm	1	σ
K_b	2478.28	kJ/(mol nm^2)	1000	ε/σ^2
θ_0	2.88	rad	165	deg
K_θ	74.35	kJ/(mol rad^2)	30	ε/rad^2
n, d			1	
K_d	4.96	kJ/mol	2	ε

Table C.3: Summary of the intermolecular parameters of the LJ 9-6 potential [a]

Low Temperatures (25 °C)					
	A	B	$\varepsilon_{ii,1000}$	σ_{ii}	$r_{c,ii}$
cellulose (C)	4.1324	0.2913	0.5525	1.2383	2.0250
2-MC	3.6765	0.1801	1.0596	1.0687	1.7476
3-MC	3.4880	0.2310	0.7073	1.2722	2.0804
6-MC	5.2975	0.2969	0.6814	1.0517	1.7198
2,3-MC	3.9416	0.2456	0.7226	1.3062	2.1359
2,6-MC	3.8093	0.2072	0.9104	1.1196	1.8308
3,6-MC	2.9452	0.2055	0.7122	1.3401	2.1914
2,3,6-MC	1.9479	0.0400	1.4776	1.4079	2.3024
High Temperatures (50 °C)					
	A	B	$\varepsilon_{ii,1000}$	σ_{ii}	$r_{c,ii}$
cellulose (C)	2.8070	0.1139	1.2780	1.1988	1.9695
2-MC	2.4410	0.0222	2.0940	1.0687	1.7476
3-MC	2.5033	0.0840	1.4012	1.2892	2.1082
6-MC	2.4760	0.0140	2.2478	1.1144	1.8308
2,3-MC	2.4819	0.0362	1.9328	1.3001	2.1359
2,6-MC	2.9161	0.0483	2.0888	1.0975	1.8031
3,6-MC	2.2424	0.0489	1.5996	1.3170	2.1637
2,3,6-MC	1.9172	0.0292	1.5670	1.4521	2.3856

[a] $\varepsilon_{ii,N} = AN^{-B}$ with N being the number of beads.

In the heterogenous methylcellulose chains, we use a geometric mixing rule to calculate the σ_{ij}, ε_{ij} and r_{cij} values between different types of the monomers:

$$\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}} \tag{C.1}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{C.2}$$

$$r_{\mathrm{c}ij} = \sqrt{r_{\mathrm{c}ii}r_{\mathrm{c}jj}} \tag{C.3}$$

C.2 Supplementary Plots for Various K_{d}

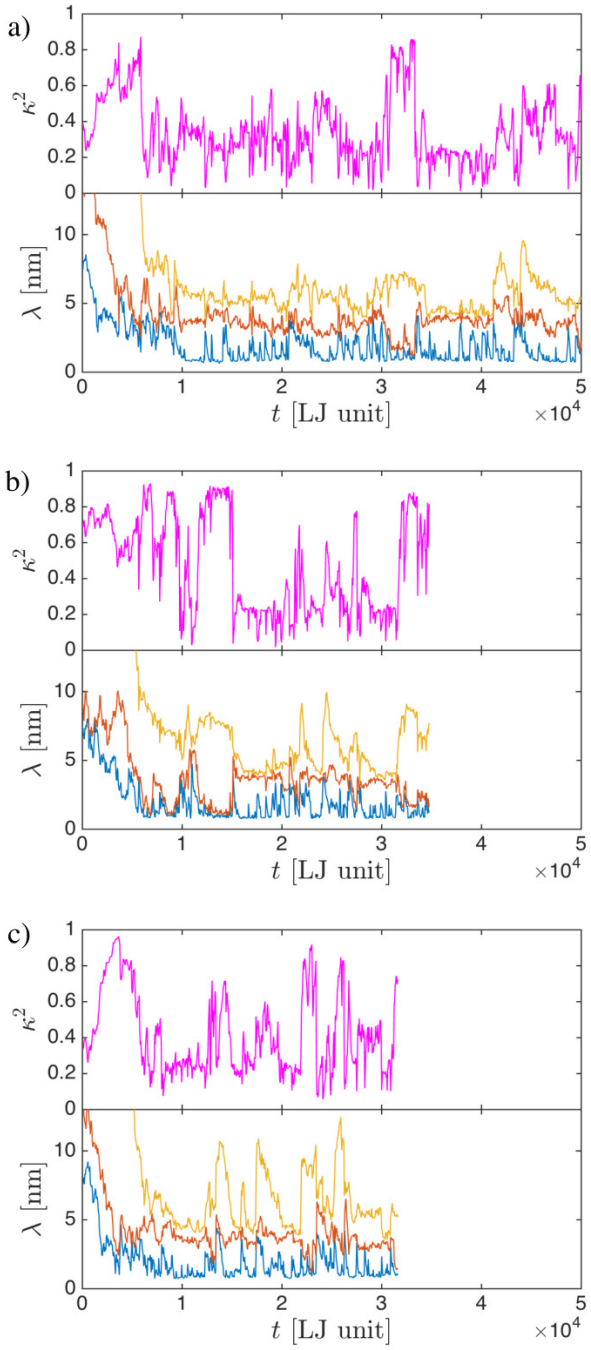

Figure C.1: Plots of κ^2 and eigenvalues versus time for three independent 1000-mer methylcellulose simulations with $K_\mathrm{d} = 2.0$ at 50 °C. λ_x, λ_y and λ_z are denoted in blue, red and yellow, respectively.

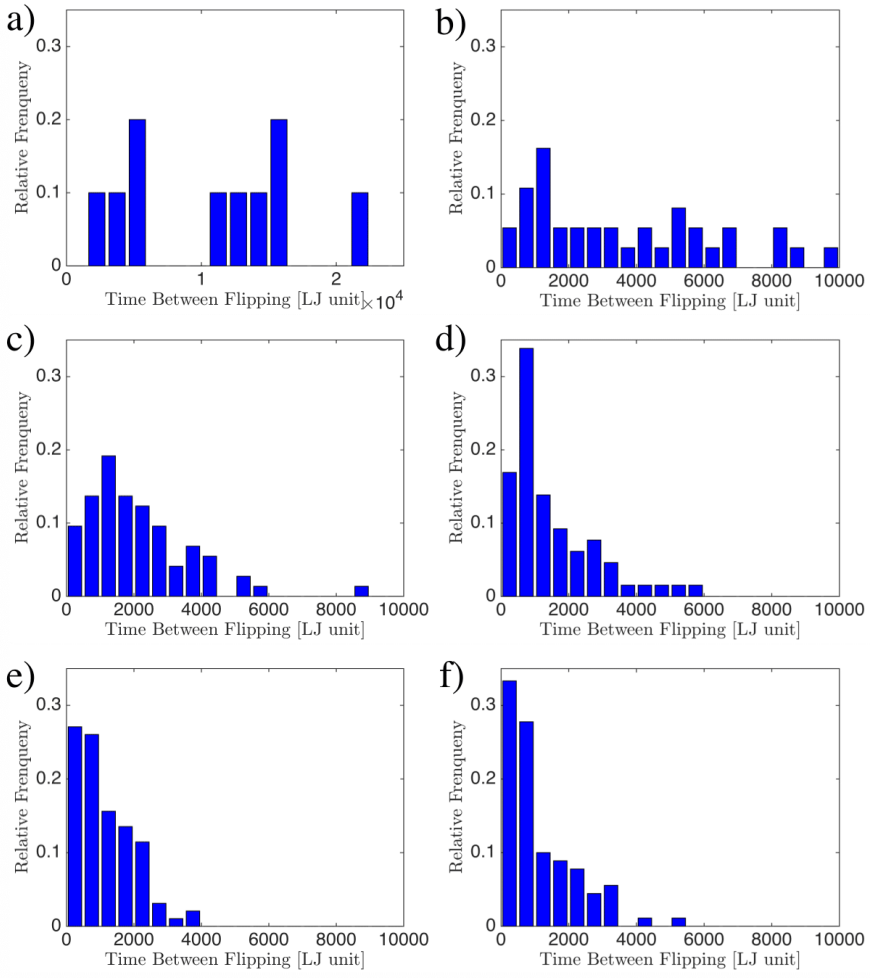

Figure C.2: Histograms showing the time interval distribution between two flipping events with respect to K_{d}. Each histogram contains at least 3 independent trajectories of 1000-mer methylcellulose simulations at 50 °C

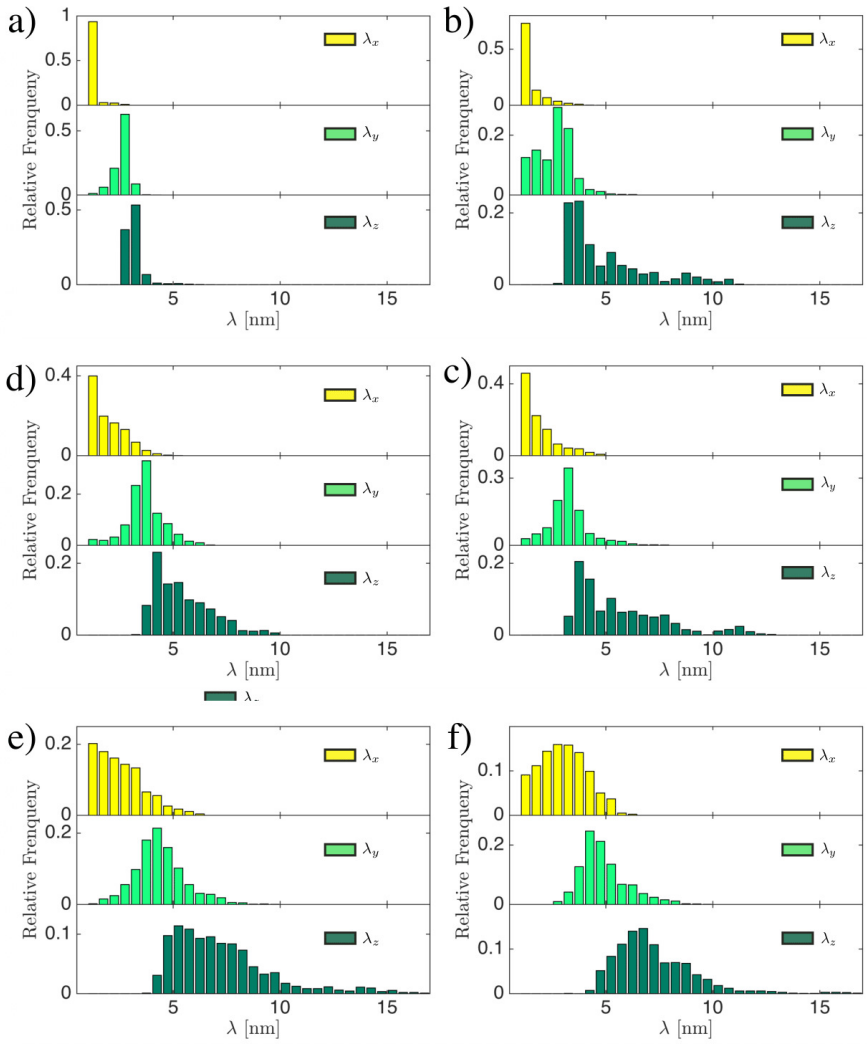

Figure C.3: Histogram of eigenvalues ($\lambda_x < \lambda_y < \lambda_z$) for 1000-mer methylcellulose chains at 50 °C. Data points are sampled every 50τ from the second half of at least 3 independent trajectories.