ABSTRACT

Title of dissertation: POLYMER CONCEPTS IN BIOPHYSICS

Gregory Morrison, Doctor of Philosophy, 2008

Dissertation directed by: Devarajan Thirumalai
Department of Chemistry and Biochemsitry

The recent advent of experimental techniques that study biological systems on the level of a single molecule have lead to a number of exciting new results. These experiments have a variety of applications in understanding both the kinetics and equilibrium properties of biomolecules. By applying the concepts of polymer physics to these single molecule experiments, we are able to more fully understand the physical picture underlying a number of experimental observations. In this thesis, we use a variety of polymer models to develop a better understanding of many single molecule experiments. We show that the kinetics of loop formation in biopolymers can be generally understood as a combination of an equilibrium and dynamic part for a number of different polymer models. We study the extension of a homopolymer as a function of applied tension, and develop a simple theoretical framework for determining the effect of interactions on the stretching of the chain. We show that the measured hopping rates in a laser optical tweezer experiment are necessarily altered by the experimental setup, and suggest a method to accurately infer the correct hopping rates using accurately measured free energy profiles. We

show that the effect of the experimental setup can be understood using a novel polymer model. Finally, we propose a Hamiltonian-based method to study the properties of spherically confined wormlike chains, which accurately determines the equilibrium properties of the system for strongly confined chains. In these studies, we are able to better understand the behavior of many disparate systems using relatively simple arguments from polymer theory.

POLYMER CONCEPTS IN BIOPHYSICS

by

Gregory Morrison

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Devarajan Thirumalai, Chair/Advisor
Professor Robert Dorfman
Professor John Weeks
Professor George Lorimer
Professor Arthur La Porta

Acknowledgments

There are many people I need to thank for all of their help in making my work here possible, and my life here enjoyable.

First and foremost, I would like to thank my advisor, Dave Thirumalai, for sparking my interest in all of these problems. Throughout my time at Maryland, Dave has presented me with a number of interesting problems, and given me the time for my mind to wander through the theory while ensuring that I stay on track. He has also given me a number of opportunities for collaboration with talented researchers, which have been both productive and quite enjoyable. Dave, thank you for teaching me how to be a scientist.

I must also thank all of my colleagues in our group, with whom I have had a number of conversations, both on personal and professional levels. Their advice and experience has been of enormous value to me. I must particularly thank Changbong Hyeon and Ngo Toan, with whom I have had the great pleasure of co-authoring papers. Our time working together was and continues to be a source of great inspiration for me, and has greatly increased my understanding of these fascinating biological problems. Thank you all for your help.

My time here in Maryland would have been much more difficult without the friendship of my housemates. Kellie Clancy has been my friend for my entire adult life, and I would not be here without her support. She has seen me through many tough times, and helped me grow as a person as well as hone my intellect. As promised, the right hand side of eq. 3.21 is dedicated to her. John Minaert has

been my friend since the beginning of graduate school, and has consistently offered me refuge from the stress in my life. I must also thank him for acting as a sounding board for my work over the years, and the fact that he (very occasionally) inspires me to want to build something. Finally, Sarah Weiss kept me sane through the process of writing this dissertation by helping me relax and making me laugh, no matter how stressed I was. Thank you all for being there.

I would never had made it to this point without the support of my parents. I am grateful for the extensive educational opportunities they provided me, their love and support throughout my life, and the lessons they have taught me. I take from my father a love of nature and a desire to get things done well as opposed to fast. My mother gave me a love of conversations late into the night, and a desire to learn more simply for the sake of knowing it. You have both given me so much, and all of what is good in me comes from you. Thank you, and I love you.

Table of Contents

List of Figures

List of Abbreviations

FRET	Föster Resonance Energy Transfer
FJC	Freely Jointed Chain
WLC	Wormlike Chain
MC	Monte Carlo
CBMC	Configuration Bias Monte Carlo
GRM	Generalized Rouse Model
NBA	Native Basin of Attraction
UBA	Unfolded Basin of Attraction
LOT	Laser Optical Tweezer
H-RNA-H	Handle + RNA + Handle
ES	Edwards and Singh
FEC	Force-Extension Curve
SCE	Self-Consistent Equation
AFM	Atomic Force Microscopy
SAW	Self-Avoiding Walk
WF	Wilemski and Fixman
SSS	Szabo, Schulten, and Schulten
TC	Thick Chain

Chapter 1

Introduction

1.1 Single Molecule Experiments in Biophysics

The field of biophysics is broad, encompassing a number of biologically relevant studies, from protein sequence analysis to cellular motility to protein folding and beyond. Recent advances in single molecule experiments have allowed experimentalists to directly study the behavior of important biomolecules under a number of different conditions [1, 2]. With these new experimental techniques, many properties of biologically relevant systems can be probed in new ways, providing new information about the structural [3] and kinetic [4] properties of biomolecules. These new experimental advances require a similar increase in our theoretical understanding of single molecule systems, in order to correctly interpret the experimental results, and suggest further experiments to be attempted. Many single molecule studies can be approached theoretically within the framework of polymer physics, providing powerful physical insights into the complex world of biology.

The question of loop closure in biomolecules is of great interest, both experimentally [5, 6, 7, 8, 9] and theoretically [10, 11, 12, 13]. The loop closure problem is relevant to determining the kinetics of RNA or protein folding, and can be used to determine the intrinsic stiffness of DNA. Loop closure times can be experimentally determined on the single molecule level using the Förster Resonance Energy Trans-

fer (FRET) experimental technique [2, 14, 15]. In a FRET experiment, donor and acceptor dyes are attached to the ends of a biomolecule, and light of a particular wavelength is shone on the system to excite the donor dye. If the donor and acceptor are sufficiently close together, the excited energy of the donor can be transferred to the acceptor via a dipole-dipole interaction. The efficiency of this energy transfer (i.e. the probability that the donor's energy is transferred to the acceptor) is given by $E = [1 + (R_{DA}/R_0)^6]^{-1}$, with R_{DA} the distance between the donor and acceptor, and R_0 a constant, depending on the details of the dyes. By monitoring the emissions of photons from the donor and acceptors, the time-dependent distance between the dyes can be monitored, allowing a direct experimental measurement of endpoint fluctuations in the biomolecule. As the timescales of loop closure are experimentally accessible, it is of importance to theoretically understand this process [10, 11]. In particular, can we predict the timescales of loop closure using simple polymer theory? How does the loop closure process depend on the particular polymer model? And finally, can we determine any general principles for understanding the process of loop closure? These questions will be addressed in chapter 2, where we show that the process of loop formation can be simply understood in terms of an equilibrium and dynamic contribution for multiple polymer models.

With the advent of single molecule pulling experiments, it is possible to directly measure the effect of an external tension on a biopolymer [16, 17, 18]. This allows a direct measurement of the folding landscapes of RNA [19, 20] and proteins [21, 22], and can estimate the bending rigidity of stiff chains. One technique for applying an external tension to a biomolecule is a Laser Optical Tweezer (LOT) experiment [23]

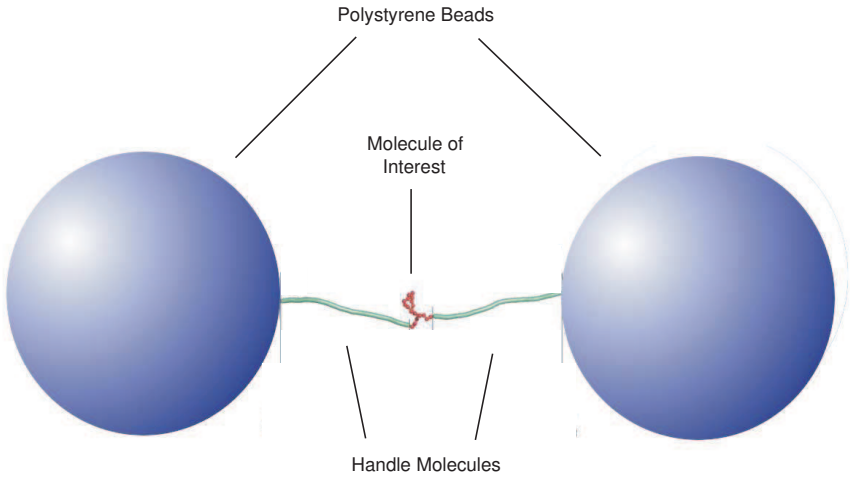

Figure 1.1: Schematic of a Laser Optical Tweezer Experiment. Labeled are the optically trapped beads, the handle molecules, and the molecule of interest.

(shown schematically in Fig. 1.1). By shining a tightly focused laser on a polystyrene bead, the bead becomes trapped, with a harmonic restoring force towards the center of the beam. In a LOT experiment, handle molecules are bonded to the endpoints of the biomolecule of interest, which are in turn attached to the polystyrene beads. One bead is pinned at the origin by a fixed laser while the position of the other optical trap is moved, causing an external tension to be applied to the system. One important question one can ask about LOT experiments is whether we can predict the behavior of the force-extension curves (FEC's) under a constant eternal tension. Can we develop scaling laws to understand the FEC's? And how do the interactions of the biomolecule change the behavior of the FEC's? One may also wonder whether the experimental setup changes the intrinsic behavior of the molecule of interest (i.e. the behavior in the absence of the handles and beads). Do the handles and beads alter the behavior of the molecule of interest? If the handles do effect the molecule's behavior, what choice of the properties of the handles will allow the

closest measurement of the intrinsic behavior of the biomolecule? These questions will be addressed in chapters 3 and 4. In chapter 3, we show that the known scaling laws for the FEC's under constant force do not accurately reproduce the behavior of short chains under tension, and develop a theory that does capture the essential features of a homopolymer under tension. In chapter 4, we show that the handles that are attached to the molecule of interest do alter the kinetic behavior of the system, regardless of their stiffness, but stiff handles can accurately measure the equilibrium properties of the system. We also show that the intrinsic kinetics of the molecule of interest can be inferred from the measured free energy profiles of the system.

The effect of confinement on biomolecules is an area of active experimental [24, 25] and theoretical [26, 27, 28] interest. Histone wrapping of DNA and strong adsorbtion to curved surfaces are essential in many biological processes, showing the importance of surface confinement. The viral encapsulation of its genome [27, 28], and the resulting structured configurations of the confined DNA [29, 30], are also of great interest. Finally, the ability to fabricate nanochannels allows experiments of confined biomolecules in many geometries [31]. Single molecule pulling experiments have allowed the direct measurement of the behavior of histone wrapped DNA [24], as well as the loading forces and encapsulated pressures in viruses [25]. However, many aspects of confined biomolecules remain undetermined experimentally, and poorly understood theoretically. Due to the importance of confined biomolecules, we wish to develop a simple theory to understand the properties of these systems. Can we determine the structures of confined biopolymers, and predict the pressure due to

confinement? In chapter 5, we develop a simple, Hamiltonian-based theory to study the effects of spherical confinement of a stiff chain, and show that the development of structure can be well understood using a locally defined order parameter.

1.2 Simulation Techniques

Computer simulations have become an essential aspect of biophysics. Simulations allow us to study the equilibrium and kinetic behavior of complex biomolecules with a level of detail that is inaccessible using experimental methods. In order to apply the principles of polymer physics to biological systems, we must treat biomolecules on a coarse grained level. In coarse graining, we replace the complex interactions found in the system of interest with minimal, unified interactions that are simpler to work with, while still capturing the essential features of the system. In our studies, we will treat biomolecules as chain of monomers, whose interactions can be generally divided into a backbone bonding potential, which keeps the distance between neighboring monomers fixed, a bending potential if the chain is stiff, and a non-local, inter-monomer potential, which accounts for non-nearest neighbor interactions. Non-local potentials must be chosen such that it closely mimics the interactions observed for the biological system of interest. The details of the models we will use are discussed in the next section.

Monte Carlo (MC) techniques are an important method in the simulation of polymer physics and coarse grained models of biomolecules [32, 33]. By iteratively making small changes to the state of a polymer, the configurational space available

to the chain can be efficiently explored. The pivot algorithm, a common implementation of the MC technique, samples the conformational space iteratively as follows: given an initial polymer, a trial polymer is generated by selecting a random monomer i and rotating the bond between monomers i and $i+1$ or $i-1$ by a random angle θ. The energy of each conformation (E_0 and E_{test}) is computed, and the new configuration is accepted using the Metropolis criterion. If $\Delta E = E_0 - E_{test} < 0$, the trial chain is accepted, whereas if $\Delta E > 0$, the new configuration is accepted with probability $P_{acc} = e^{-\beta \Delta E}$, with $\beta = 1/k_B T$. This acceptance rule ensures that the trial chains are Boltzmann distributed, such that the average of an observable A in the simulation is given by

$$\langle A \rangle = \int \prod_n d^d \mathbf{r}_n A(\{\mathbf{r}_n\}) e^{-\beta V[\{\mathbf{r}_n\}]} / \int \prod_n d^d \mathbf{r}_n e^{-\beta V[\{\mathbf{r}_n\}]} \approx \frac{1}{M} \sum_{i=1}^{M} A_i, \tag{1.1}$$

after M samples have been taken. The pivot algorithm implicitly takes the backbone potential into account, by keeping the distance between neighboring monomers fixed.

Another powerful Monte Carlo technique is Configurational Bias Monte Carlo (CBMC), which enables the generation of equilibrated and uncorrelated chains of N bonds [32, 34]. This is accomplished by iteratively growing the chain out of a number of generated trial bonds. The procedure is as follows: suppose we have created $n-1$ bonds of a chain, and wish to add the n^{th}. We generate l trial bonds, $\{\mathbf{u}_j\}_{j=1}^{l}$, and compute the weight for each trial bond $\mathbf{u}_j$ as $w_n^{(j)} = \exp(-\beta E_j)$. A particular trial bond m is added to the end of the chain with probability $w_n^{(m)}/W_n$, with $W_n = \sum_{j=1}^{l} w_n^{(j)}$ (effectively an estimate of the partition function for the n^{th} bond). Once the full chain is grown using this method, we compute the total

weight, $W_{tot} = \prod_{n=1}^{N} W_n$, and compare this to the weight of another chain, grown with the CMBC algorithm with weight W_{tot}^{old}. The new chain is accepted using the Metropolis criterion, $P_{acc} = \min(1, W_{tot}/W_{tot}^{old})$. Repeated application of this scheme allows the creation of many chains which satisfy Boltzmann statistics, with the advantage of each chain being completely uncorrelated from the previous chain. Free energies are also easily computed, with $\beta F = -\log(\langle W_{tot} \rangle)$. This ability to calculate F explicitly is another advantage of the CBMC algorithm. While the number of trial bonds l is arbitrary, it is essential that l be chosen large enough such that energetically favorable bonds are often tested (i.e. the estimate of the partition function must be accurate).

Monte Carlo simulations are useful in the study of the equilibrium properties of coarse grained biomolecules, but the methods above can not give us any information about the kinetics of these systems. Instead, Langevin dynamics simulations can be used to determine the kinetic properties of biological systems [32]. The equations of motion of a coarse grained homopolymer in an implicit solvent can be written as

$$m\frac{\partial^2 \mathbf{r}_n(t)}{\partial t^2} = -\gamma\frac{\partial \mathbf{r}_n(t)}{\partial t} - \frac{\delta H[\mathbf{r}_n(t)]}{\delta \mathbf{r}_n(t)} + \vec{\eta}_n(t), \tag{1.2}$$

where $\mathbf{r}_n(t)$ is the position of the n^{th} monomer of mass m_n, and γ is the friction coefficient. The Hamiltonian $H[\mathbf{r}_n]$ describes the intra-chain interactions, minimally containing the backbone bonding of the chain, and also includes the backbone stiffness and inter-monomer interactions when needed. In our work, the solvent is addressed implicitly using the random white noise force $\vec{\eta}_n(t)$, with $\langle \eta_n^{(i)}(t) \rangle = 0$ and $\langle \eta_n^{(i)}(t) \times \eta_m^{(j)}(t') \rangle = 2k_B T\gamma \times \delta_{nm}\delta_{ij}\delta(t-t')$ (with i and j denoting the components

of the force). In our Langevin simulations, the equations of motion are integrated using the velocity Verlet algorithm [32]. In the case of high friction, the inertial term is negligible compared to the dissipative term in eq. 1.2, and the equations of motion are approximately Brownian,

$$\gamma \frac{\partial \mathbf{r}_n(t)}{\partial t} = -\frac{\delta V[\mathbf{r}_n(t)]}{\delta \mathbf{r}_n(t)} + \vec{\eta}_n(t). \tag{1.3}$$

We will primarily be interested in the Brownian limit of the Langevin equation when studying the dynamics of a system, as most biological systems satisfy the high friction requirement, and the kinetics are more accurately determined in this regime.

1.3 Polymer Models

Polymer physics involves the study of a chain of interacting monomers. One of the simplest polymer models [35, 36] is the Freely Jointed Chain (FJC), with a fixed bonding distance and no inter-monomer interactions, shown schematically in Fig. 1.2a. The distribution in phase space for a FJC with N bonds and with bond spacing a is

$$\Psi_{FJC} = \prod_{n=1}^{N} \delta(|\mathbf{r}_{n+1} - \mathbf{r}_n| - a), \tag{1.4}$$

where $\langle \cdots \rangle_{FJC} = \int \prod_n d^3\mathbf{r}_n \ (\cdots)\Psi_{FJC} / \int \prod_n d^3\mathbf{r}_n \ \Psi_{FJC}$. One experimental observable of interest, which roughly measures the 'size' of the polymer, is the end-to-end distance of the polymer $\mathbf{R} = \mathbf{r}_{N+1} - \mathbf{r}_1 = \sum_{n=1}^{N} \mathbf{r}_{n+1} - \mathbf{r}_n \equiv a \sum_{n=1}^{N} \mathbf{u}_n$, with the bond vector $\mathbf{u}_n = (\mathbf{r}_{n+1} - \mathbf{r}_n)/a$. The FJC is rotationally invariant, so $\langle \mathbf{R} \rangle = 0$, but

$$\langle \mathbf{R}^2 \rangle = a^2 \sum_{nm} \langle \mathbf{u}_n \cdot \mathbf{u}_m \rangle = a^2 \sum_{n} \langle \mathbf{u}_n^2 \rangle + a^2 \sum_{n \neq m} \langle \mathbf{u}_n \cdot \mathbf{u}_m \rangle = Na^2, \tag{1.5}$$

where the last equality follows from the fact that $\langle \mathbf{u}_n^2 \rangle \equiv 1$ for all n, and the bond angles are uncorrelated, with $\langle \mathbf{u}_n \cdot \mathbf{u}_m \rangle = 0$ for $n \neq m$.

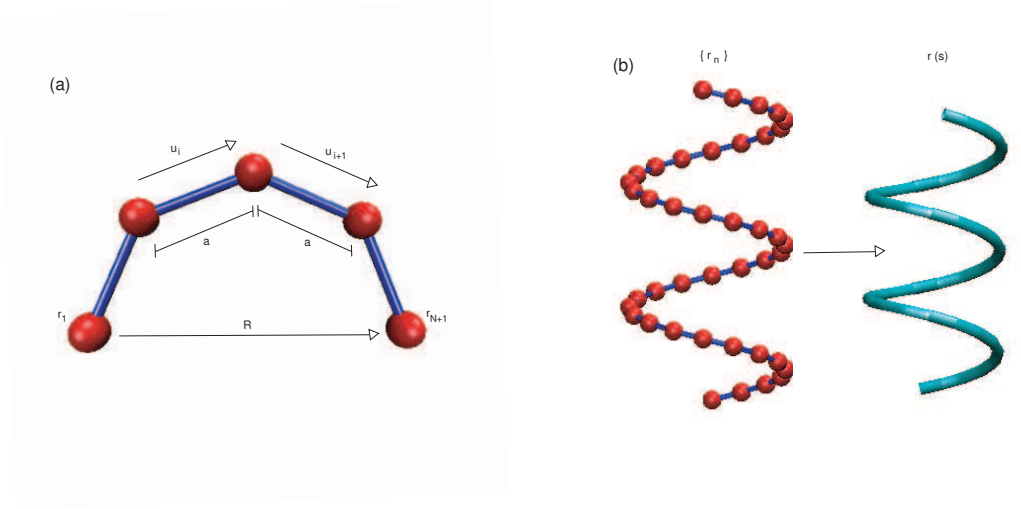

Figure 1.2: (a) Schematic of a Freely Jointed Chain. Shown are the bond vectors and end-to-end distance. (b) Schematic of the continuum representation of a polymer.

The FJC is closely mimicked by the Gaussian chain, which treats the bonds between monomers as harmonic springs, rather than rigid constraints. This approximate representation for the FJC is much simpler to work with than the rigid constraints for the FJC, and can be expressed in terms of the discrete Hamiltonian

$$\beta H_G = \frac{3}{2a^2} \sum_{n=1}^{N} (\mathbf{r}_{n+1} - \mathbf{r}_n)^2. \tag{1.6}$$

The Gaussian chain is simpler than the FJC to work with, and gives the expected $\langle \mathbf{R}^2 \rangle = Na^2$. The end-to-end distribution function (or Green's function) is also simple to work out exactly,

$$P_G(\mathbf{R}; N) = \left(\frac{3}{2\pi N a^2} \right)^{3/2} \exp\left(- \frac{3\mathbf{R}^2}{2Na^2} \right). \tag{1.7}$$

In the thermodynamic limit (with $N \to \infty$, $a \to 0$, while holding the length and $Na = L$ fixed), the discrete Hamiltonian H_G can be written in the continuum limit

(with $\mathbf{r}_n \to \mathbf{r}(s)$ and $\lim_{a\to 0}(\mathbf{r}_{n+1} - \mathbf{r}_n)/a \to \partial\mathbf{r}(s)/\partial s$) as

$$\beta H_G[\mathbf{r}(s)] = \frac{3}{2a}\int_0^L ds\Big(\frac{\partial \mathbf{r}(s)}{\partial s}\Big)^2. \tag{1.8}$$

This continuum representation is often useful in obtaining analytic results for the Gaussian chain, shown schematically in Fig. 1.2b. The continuum Gaussian chain is equivalent to a random walk, as the Green's function for the end-to-end distance satisfies $(\partial/\partial N - a^2/6\ \nabla^2_{\mathbf{r}_L})P(\mathbf{r}_L - \mathbf{r}_0; N) = 0$. The continuum representation, and the resulting connection to the diffusion equation, is quite useful for determining the behavior of a Gaussian chain in a variety of situations.

Many biological molecules are stiff, with an intrinsic resistance to bending [35, 36]. This resistance to bending can be introduced through an energy penalty in the distribution in phase space,

$$\begin{aligned}\Psi_{WLC} &= \prod_{n=1}^{N}\delta(|\mathbf{u}_n| - 1)\exp\Big(\frac{l_p}{a}\sum_{n=1}^{N-1}\frac{\mathbf{u}_{n+1}\cdot\mathbf{u}_n}{a^2}\Big) \\ &\propto \prod_{n=1}^{N}\delta(|\mathbf{u}_n| - 1)\exp\Big(-\frac{l_p}{2a}\sum_{n=1}^{N-1}(\mathbf{u}_{n+1} - \mathbf{u}_n)^2\Big).\end{aligned} \tag{1.9}$$

This provides an energy penalty for inducing a bend between bonds n and $n+1$, while keeping the lengths of the bond vectors fixed. In the continuum limit (with $\mathbf{u}_n \to \mathbf{u}(s)$ and $(\mathbf{u}_{n+1} - \mathbf{u}_n)/a \to \partial\mathbf{u}(s)/\partial s = \partial^2\mathbf{r}(s)/\partial s^2$), the distribution in eq. 1.9 becomes

$$\Psi_{WLC}[\mathbf{u}(s)] = \prod_s \delta(|\mathbf{u}(s)| - a)\exp\Big[-\frac{l_p}{2}\int_0^L ds\Big(\frac{\partial\mathbf{u}(s)}{\partial s}\Big)^2\Big] \tag{1.10}$$

As in the case of the Gaussian Hamiltonian (eq. 1.8), eq. 1.10 is equivalent to the diffusion equation, $(\partial/\partial L - l_p/4\ \nabla^2_{\mathbf{u}})G(\mathbf{u}_L, \mathbf{u}_0; L) = 0$ on the unit sphere, with the

resulting Green's function [37]

$$G(\mathbf{u}_L, \mathbf{u}_0; L) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m(\theta_0, \varphi_0) Y_l^m(\theta_L, \varphi_L) e^{-l(l+1)L/2l_p}. \tag{1.11}$$

This Green's function allows the correlations in the bond vectors to be determined as

$$\langle \mathbf{u}(0) \cdot \mathbf{u}(L) \rangle = e^{-L/l_p}, \tag{1.12}$$

showing that the persistence length is the range over which the bonds are correlated. While many averages for the WLC are known exactly, the behavior of a WLC remains poorly understood in a number of biologically relevant situations, making the wormlike chain a fascinating system of study. Further discussion of the behavior of a WLC can be found in chapters 4 and 5.

Single molecule techniques allow us to pull on the ends of a polymer, with a force $\mathbf{f}$ applied to one end of the chain and $-\mathbf{f}$ to the other. This gives rise to the additional energetic term $H[\mathbf{r}(s), \mathbf{f}] = H[\mathbf{r}(s)] - \mathbf{f} \cdot [\mathbf{r}(L) - \mathbf{r}(0)] = H[\mathbf{r}(s)] - \mathbf{f} \cdot \int_0^L ds\, \mathbf{u}(s)$. The effect of tension on the linear extension of the flexible polymer models (both the FJC and Gaussian chains) can be computed directly. In particular, the linear extension, $\langle \mathbf{R} \cdot \hat{\mathbf{f}} \rangle \equiv \langle Z \rangle$, which is generally used as the experimental observable for real biomolecules, can be determined for any polymer model as

$$\langle Z \rangle = \frac{\partial}{\partial(\beta \mathbf{f})} \Big[\int \mathcal{D}[\mathbf{r}(s)] e^{-\beta H + \beta \mathbf{f} \cdot (\mathbf{r}_L - \mathbf{r}_0)} \Big] \tag{1.13}$$

In the case of the FJC, we find that $\langle Z \rangle_{FJC} = Na[\coth(a\beta f) - 1/(a\beta f)]$, with $f = |\mathbf{f}|$. For a Gaussian chain, the extension is proportional to the tension, with

$\langle Z \rangle_G = Na(a\beta f)/3$. This shows that the Gaussian chain is extensible, with the average extension proportional to $\mathbf{f}$ for all external tension. Thus, while the Gaussian model is extremely useful for describing a flexible chains under certain situations, it can not fully describe an inextensible system such as the FJC. The linear extension for a WLC is significantly more difficult to determine analytically, but is well approximated by [1] $l_p\beta f = x - 1/4 + 1/4(1-x^2)$, with $x = \langle Z \rangle / L$. Further discussion of the effect of tension on polymers can be found in chapters 3 and 4.

Additional interactions between non-nearest neighbors are essential in relating the concepts of polymer physics to systems of real biological interest. One of the simplest and most essential of these is the excluded volume interaction, which accounts for the fact that multiple monomers can not occupy the same space. These self-avoiding chains are difficult to deal with analytically, but a simple scaling arguments due to Flory [6] allow us to determine the behavior of a self-avoiding chain for large N. If the 'size' of the self avoiding polymer is R, the free energy for a d-dimensional self-avoiding chain can be written as

$$F \sim \frac{R^2}{N} + v_0 N \times \frac{N}{V} \sim \frac{R^2}{N} + v_0 \frac{N^2}{R^d} \tag{1.14}$$

where we have neglected all numerical coefficients. The first term is the entropy of a simple Gaussian chain (the exponent in eq. 1.7), and the second term accounts for the interactions of monomers with excluded volume v_0, which are localized to a volume $V \sim R^d$. Minimization of the free energy with respect to R gives

$$R \sim R_F \sim N^{\frac{3}{2+d}} \equiv N^{\nu} \tag{1.15}$$

defining the Flory exponent $\nu = 3/(2+d)$, and R_F the Flory radius. The Flory

exponent has been shown to be exact in 1 and 2 dimensions, and remarkably accurate in 3 dimensions, showing the power of such simple scaling arguments. The Flory radius is the only scaling variable in the problem of self-avoiding chains, so the distribution of end-to-end distances can be written in terms of $x = |\mathbf{R}|/R_F$, and the properties of self-avoiding chains will be invariant to changes in v_0 or N for large enough N when written in terms of the scaled variable x. Self avoiding (purely repuslive) homopolymers are also referred to polymers in a good solvent, in contrast to self-attracting homopolymers, referred to as polymers in a poor solvent. In a poor solvent, each monomer on the chain feels a short ranged attraction to every other monomer. The scaling of a poor solvent can be determined using a Flory scaling argument as well, with $R \sim N^{1/3}$ for poor solvents. However, the relative attraction between monomers decreases with increasing temperature, with a transition from poor to good occurring at the Θ-temperature. At $T = \Theta$, the homopolymer has the scaling $R \sim N^{1/2}$ (identical to the random walk), and is expected to behave similar to a FJC. The effect of solvent quality on homopolymers is further discussed in chapters 2 and 3.

The kinetic behavior of polymers is often quite difficult to determine analytically. One simple analytic approach to determining the timescales of polymer fluctuations is the Rouse model, which describes the time-dependent behavior of a Gaussian chain in the overdamped limit of the Langevin equation (eq. 1.3). In the continuum limit [35], the time-dependent Gaussian Hamiltonian is

$$\beta H_G[\mathbf{r}(s,t)] = \frac{3}{2a^2}\int_0^N ds\, \left(\frac{\partial \mathbf{r}(s,t)}{\partial s}\right)^2. \tag{1.16}$$

If we take the center of mass of the chain to be at the origin, this can be broken into normal modes, with $\mathbf{r}(s,t) = \sum_{n=1}^{N} \tilde{\mathbf{r}}_n(t)\cos(n\pi s/N)$, and the equations of motion become

$$\frac{\partial \tilde{\mathbf{r}}_n}{\partial t} = -\frac{3n^2\pi^2 D}{N^2 a^2}\tilde{\mathbf{r}}_n + \vec{\eta}_n(t), \tag{1.17}$$

with the diffusion coefficient $D = k_B T/\gamma$. This is exactly solvable, with $\tilde{\mathbf{r}}_n(t)$ expressible in terms of $e^{-n^2 t/\tau_R}$, giving the Rouse time $\tau_R = N^2 a^2/3D\pi^2$. τ_R is the slowest mode of the Gaussian chain, and gives the timescale of global fluctuations of the polymer. Many time-dependent averages can be computed in terms of τ_R with little difficulty. The Langevin equation can also be solved exactly in the discrete case (eq. 1.6) using normal modes as well. Further discussion of the kinetics of simple polymer models can be found in chapters 2 and 4.

1.4 Summary

In this thesis, we will apply the theoretical approaches of polymer physics to systems of biological interest. In chapter 2, we address an inconsistency between two analytic theories to determine the loop closure time of a noninteracting Gaussian chain. The theory of Wilemskiy and Fixman [38] (WF), which determines the loop closure time in terms of a sink-sink correlation function, finds $\tau_c \sim N^2$, which has been shown to agree well with simulations and experiments. However, the theory of Szabo, Schulten, and Schulten [11] (SSS), which determines the closure time as diffusion in terms of an effective potential for the endpoints depending only on the equilibrium properties of the chain, shows that $\tau_c \sim N^{3/2}$. We show that, if

the dynamics of the endpoints are correctly taken into account, using an effective diffusion coefficient that addresses the kinetics of the endpoints, the SSS theory can be made to agree with the WF theory, with the kinetics of loop formation broken into equilibrium and dynamical parts. We also show that this result is quite general, and can be used to address the closure for an interacting chain as well. The ability to break the problem of loop closure into two experimentally accessible parts allows for a simple and intuitive understanding of closure in real biological systems.

In chapter 3, we address the equilibrium behavior of an interacting homopolymer under the application of an external tension, f. Pincus determined that, for large N, the extension of a self-avoiding chain in the direction of the force should scale as [39] $\langle Z \rangle \sim f^{2/3}$. We analytically show, using the methods of Edwards and Singh (ES) [40], that this scaling regime does indeed exist for large N. However, we find that 'large N' requires a surprisingly long chain, with the expected scaling emerging only for $N > 9 \times 10^4$. We show using simulations that the ES theory agrees well with simulations, and that the Pincus scaling is not observed for shorter chains ($N \leq 1600$). We show that this is due to a failing of the scaling arguments used to determine the $f^{2/3}$ scaling, which is not applicable to short chains. We also study the behavior of a homopolymer in a poor solvent, and show that the extension of such a chain shows rich behavior as a function of external tension.

In LOT experiments, handles are attached to the molecules of interest in order to study the equilibrium and kinetic properties of the molecule of interest. In chapter 4, we study the effect of attaching these linkers to the P5GA RNA hairpin. We show that the intrinsic equilibrium distribution functions are accurately reproduced only

using very stiff handles. We also show that all handles, either flexible or stiff, will change the kinetic properties of the system (the folding and unfolding times), with stiff handles having the largest effect on the hopping times. We develop a simple Generalized Rouse Model (GRM) to explain these effects, which treats the complicated RNA interactions as a unified harmonic bond with a cutoff, and the handles as Gaussian chains. We find that the GRM is able to almost quantitatively reproduce the effect of attaching both flexible and stiff linkers. Finally, we show directly that the kinetic properties of the chain under any applied force can be best determined by accurately measuring the equilibrium distribution at the midpoint force f_m, and using Kramer's theory to determine the folding or unfolding rates.

In chapter 5, we study the equilibrium behavior of a spherically confined WLC. By applying the mean field method [41, 42] to a WLC confined to the surface of a sphere, we are able to recover all known scaling laws, as well as reproduce the exact average end-to-end distance for the confined system. We also determine the effect of external tension on a surface confined WLC, of interest in studying the unravelling of DNA wrapped around a histone. Finally, we apply the mean field method to a WLC confined to the interior of a sphere, which is of great interest in studying the behavior of DNA encapsulated in a virus. We show that long, volume confined WLC's behave as a surface confined chains, and that the structure of the encapsulated chain can be understood by examining the correlations in the local winding axis. We also determine the pressure on the sphere due to the confined WLC, and show that the entropy of confinement is insufficient to determine the huge pressures observed in real viruses, with intra-chain interactions playing a huge

role in the pressure.

Chapter 2

The Kinetics of Loop Closure in Polymer Chains

2.1 Introduction

Contact formation between the ends of a long polymer (cyclization) has been intensely studied both experimentally [5, 6] and theoretically [38, 11, 12, 13, 43, 44, 45]. Recently, the kinetics of loop formation has become increasingly important largely because of its relevance to DNA looping [7, 46] as well as protein [8, 47, 48, 49, 50, 51, 52] and RNA folding [9]. The ease of cyclization in DNA, which is a measure of its intrinsic flexibility [46, 53], is important in gene expression and interactions with proteins and RNA. In addition, the formation of contacts between residues (nucleotides) near the loop [44] may be the key nucleating event in protein (RNA) folding. Experiments studying loop formation of proteins [8, 47, 54] and RNA [3, 55] have prompted a number of theoretical studies [43, 56, 57] that build on the pioneering work of Wilemski and Fixman [38] (WF) and Szabo, Schulten, and Schulten [11] (SSS). The WF formalism determines the loop closure time τ_c by solving the diffusion equation in the presence of a sink term. The sink function accounts for the possibility that contact between the ends of a polymer chain occurs whenever they are in proximity. The time for forming a loop is related to a suitable time integral of the sink-sink correlation function. SSS developed a much simpler theory to describe the dependence of the rate of end-to-end contact formation in

an ideal chain on the polymer length N. The SSS approximation [11] describes the kinetics of contact formation between the ends of the chain as a diffusive process in an effective potential, derived from the end-to-end probability distribution $P(\mathbf{R}_{ee})$

The validity of using the dynamics in a potential of mean force, $F(\mathbf{R}_{ee}) \sim -k_B T \log[P(\mathbf{R}_{ee})]$, to obtain τ_c hinges on local equilibrium being satisfied, i.e. all processes except the one of interest must occur rapidly. In the case of cyclization kinetics in simple systems (the Rouse model or self-avoiding polymer chains), the local equilibrium approximation depends minimally on the cyclization time τ_c, and the internal chain relaxation time τ_R. In the limit $\tau_c/\tau_R \gg 1$, one can envision the motions of the ends as occurring in the effective free energy $F(\mathbf{R}_{ee})$, because the polymer effectively explores the available volume before the ends meet. By solving the diffusion equation for an ideal chain for which $F(\mathbf{R}_{ee}) \sim 3k_B T \mathbf{R}_{ee}^2/2\bar{R}_{ee}^2$, with $\bar{R}_{ee} \sim b\sqrt{N}$, where b is the monomer size, subject to absorbing boundary conditions, SSS showed that the mean first passage time for contact formation ($\sim \tau_c$) is $\tau_{SSS} \sim \tau_0 N^{\frac{3}{2}}$, where τ_0 is a microscopic time constant The simplicity of the SSS result, which reduces contact-formation kinetics to merely computing $P(\mathbf{R}_{ee})$, has resulted in its widespread use to fit experimental data on polypeptide chains [8, 47, 54]. The dependence of τ_c on N using the SSS theory differs from the WF predictions. In addition, simulations also show that τ_c deviates from the SSS prediction [58, 59, 60, 61], which predicts $\tau_c \sim N^2$. The slower dependence of τ_{SSS} on N can be traced to the failure of the assumption that all internal chain motions occur faster than the process of interest.

The theory based on the WF formalism and simulations show the closure time

$\tau_{WF} \sim N^{1+2\nu}$ ($\nu \approx 3/5$ for self-avoiding walk and $\nu = 1/2$ for the Rouse chain), where D_c is a diffusion constant. In this chapter, we show that the WF result for Rouse chains, τ_{WF}, can be obtained within the SSS framework, provided an effective diffusion constant that accounts for the relaxation dynamics of the ends of the chains is used, instead of simply using the monomer diffusion coefficient D_0. Thus, the simplicity of the SSS approach can be preserved while recovering the expected scaling result [38, 12] for the dependence of τ_c on N. (ii) The use of the Rouse model may be appropriate for polymers or polypeptide chains near Θ-conditions; however, in both good and poor solvents interactions between monomers determine the statics and dynamics of the polymer chains. The chain will swell in good solvents ($\nu \approx 3/5$) whereas in poor solvents, polymers and polypeptide chains adopt compact globular conformations. In these situations, interactions between the monomers or the amino acid residues affect τ_c. The monomer-monomer interaction energy scale, ϵ_{LJ}, leading to the chain adopting a swollen or globular conformation influences both ν and the chain relaxation dynamics, and hence affects τ_c. Because analytic theory in this situation is difficult, we provide simulation results for τ_c as a function of ϵ_{LJ} and for $10 < N \leq 100$.

2.2 Derivation of τ_{WF} for the Rouse Model using the SSS Approximation

The Rouse chain consists of N beads, with successive beads connected by a harmonic potential that keeps them at an average separation b, the Kuhn length.

Contact formation between the chain ends can occur only if thermal fluctuations result in monomers 1 and N being within a capture radius a. There are three relevant time scales that affect loop closure dynamics: $\tau_0 \approx b^2/D_0$, the fluctuation time scale of a single monomer, τ_{ee}, the relaxation time associated with the fluctuations of the end-to-end distance, and τ_R, the relaxation time of the entire chain. Because loop formation can occur only if the ends can approach each other, processes that occur on time scale τ_{ee} must be coupled to looping dynamics, which must be accounted for in the kinetics of the system.

The Langevin equation for a Gaussian chain in the continuum limit is [36]

$$\gamma \frac{\partial \mathbf{r}(s,t)}{\partial t} = -\frac{\delta H_0[\mathbf{r}(s,t)]}{\delta \mathbf{r}(s,t)} + \vec{\eta}(s,t), \tag{2.1}$$

where $\vec{\eta}(s,t)$ a white noise force with $\langle \vec{\eta}(s,t) \rangle = 0$, $\langle \vec{\eta}(s,t) \cdot \vec{\eta}(s',t') \rangle = 6\gamma k_B T \delta(t - t')\delta(s - s')$. γ is the friction coefficient, and $D_0 = k_B T/\gamma$ is the microscopic diffusion coefficient. By writing $\mathbf{r}(s,t) = \mathbf{r}_0 + 2\sum_{n=1}^{N-1} \mathbf{r}_n(t)\cos(n\pi s/N)$, the Gaussian Hamiltonian H_0 becomes

$$H_0 = \frac{3}{2b^2}\int_0^N ds \Big(\frac{\partial \mathbf{r}(s,t)}{\partial s}\Big)^2 = \frac{3}{2Nb^2}\sum_n n^2\pi^2 \mathbf{r}_n^2(t). \tag{2.2}$$

The equation of motion for each mode

$$\dot{\mathbf{r}}_n(t) = -\frac{3n^2\pi^2 D_0}{N^2 b^2}\mathbf{r}_n(t) + \vec{\eta}_n(t). \tag{2.3}$$

can be solved independently, and the solutions naturally reveal the time scale for global motions of the chain, $\tau_R = N^2 b^2/3D_0\pi^2 \sim N^2 b^2/D_0$. We note that τ_R is much larger than the relevant time scale for internal motions of the monomers, $\tau_1 \approx b^2/D_0$ for large N. Solving eq. 2.3 directly, we find the fluctuations in the

end-to-end distance $\mathbf{R}_{ee}$ are given by

$$\langle \delta \mathbf{R}_{ee}^2(t) \rangle = 16 N b^2 \sum_{n \text{ odd}} \frac{N^2}{n^4 \pi^4} \sin^2 \left(\frac{n\pi}{N} \right) \left(1 - e^{-n^2 t/\tau_R} \right). \tag{2.4}$$

with $\langle \delta \mathbf{R}_{ee}^2(t) \rangle \equiv \langle [\mathbf{R}_{ee}(t) - \mathbf{R}_{ee}(0)]^2 \rangle$. The details of the calculation leading to eq. (2.4) are given in Appendix A.1. If we define an effective diffusion constant using

$$D(t) = \frac{\langle \delta \mathbf{R}_{ee}^2(t) \rangle}{6t}, \tag{2.5}$$

then $D(0) = 2D_0$, as is expected for the short time limit [11, 60]. On time scales on the order of $\tau_R \sim N^2$, we find $D(\tau_R) \sim D_0/N$. This is identical to the diffusion constant for the center of mass of the chain [36], and is expected for the diffusion constant for global chain motion.

The theory of Szabo, Schulten, and Schulten [11] (SSS) determines the loop closure time by replacing the difficult polymer problem, having many degrees of freedom, with a single particle diffusing in a potential of mean force. With this approximation, τ_c, which can be related to the probability that the contact is not formed, becomes

$$\tau_c = \frac{1}{\mathcal{N}} \int_a^{Nb} dr \frac{1}{D(r)P(r)} \Big(\int_r^{Nb} dr' P(r') \Big)^2 + \frac{1}{\kappa \, \mathcal{N} P(a)}, \tag{2.6}$$

where κ is the rate of loop closure given $|\mathbf{R}_{ee}| = a$, $P(r)$ is the equilibrium end-to-end distribution of the chain, and $\mathcal{N} = \int_a^{Nb} dr \; P(r)$. Here, we will consider only a chemically irreversible process, with the binding rate constant $\kappa \to \infty$. In the case of the non-interacting Gaussian chain, $P(r) \sim r^2 \exp(-3r^2/2Nb^2)$, and we find (for $N \gg 1$ and with the assumption $D(r) = D_0 = \text{const}$, the loop closure time is [11]

$$\tau_{SSS} \approx \frac{1}{3} \sqrt{\frac{\pi}{6}} \frac{N^{\frac{3}{2}} b^3}{D_0 a}. \tag{2.7}$$

The scaling of $\tau_{SSS} \sim N^{\frac{3}{2}}$ given in eq. (2.7) disagrees with other theories [38, 43] and numerous simulations [58, 59, 60, 61] that predict $\tau_c \sim N^2$ for $Nb^2 \gg a^2$ and $a \geq b$. It has been noted [56, 62] that the SSS theory may be a lower bound on the loop closure time for a freely draining Gaussian chain, and that an effective diffusion coefficient that is smaller than D_0 is required to fit the simulated [56] and experimental [2] data using τ_{SSS}.

As noted by Doi [12], the relevant time scale for loop closure is not simply the global relaxation time. Fluctuations in $\mathbf{R}_{ee}$ are given not only by the longest relaxation time (i.e. the ground state approximation), but also from important contributions that arise from higher modes. This gives rise to the differences between the Harmonic Spring and Rouse models [12, 59]. In the Harmonic Spring model, the chain is replaced with only one spring, connecting the two ends of the chain (the spring constant is chosen to reproduce the end-to-end distribution function). The higher order modes give rise to excess fluctuations on a scale $\sim 0.4\sqrt{N}b = R'$, and their inclusion is necessary to fully capture the physics of loop closure. In the approximation of a particle diffusing in an effective potential (as in the SSS theory), this time scale is simple to determine. If we consider only the x component of $\mathbf{R}_{ee}$, we can treat it as a particle diffusing in a potential $U_{eff}(R_x) = 3R_x^2/2Nb^2 - O(1)$, with diffusion constant $D = 2D_0$ (the factor of 2 arising from $D(0) = 2D_0$, see eq. 2.5). In this case, we find

$$\langle \delta R_x^2(t) \rangle = \frac{2}{3} Nb^2 \Big(1 - e^{-t/\tau_{ee}} \Big), \tag{2.8}$$

and $\langle \mathbf{R}_{ee}^2(t) \rangle = 3\langle \delta R_x^2(t) \rangle$, with $\tau_{ee} = Nb^2/6D_0$ the natural relaxation time of the

end-to-end distance. Because we have evaluated τ_{ee} using diffusion in an effective potential, the dependence of τ_{ee} on N should be viewed as a mean field approximation.

We can determine the effective diffusion constant on the time scale τ_{ee}, which includes the relaxation of $\mathbf{R}_{ee}(t)$ at the mean field level. We define the effective diffusion constant as

$$D_{ee} = \lim_{t \sim \tau_{ee}} \frac{\langle \delta \mathbf{R}_{ee}^2(t) \rangle}{6t}. \tag{2.9}$$

with $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ given in eq. (2.4), which includes all of the modes of the chain, and not simply the lowest one (see eq. 2.5). Noting that $\tau_{ee}/\tau_R \sim N^{-1} \ll 1$ for large N, we can convert the sum in eq. (2.4) into an integral:

$$\begin{aligned} \langle \delta \mathbf{R}_{ee}^2 \rangle &\approx \frac{2\sqrt{2}}{\pi} N^{\frac{3}{2}} b^2 \int_0^\infty dx \, \frac{\sin^2(bx/\sqrt{3D_0 t})}{x^4} \left(1 - e^{-x^2}\right) && (2.10) \\ &\approx 8b\sqrt{\frac{3D_0 t}{\pi}}. && (2.11) \end{aligned}$$

In particular, for $t \approx \tau_{ee}/2 = Nb^2/12D_0$,

$$D_{ee} \approx \frac{8\, D_0}{\sqrt{N\pi}} - \frac{16 D_0}{3N} + O(N^{-\frac{3}{2}}). \tag{2.12}$$

We expect these coefficients to be accurate to a constant on the order of unity. The effective diffusion constant D_{ee} takes the higher order modes of the chain into account, and should capture the essential physics of the loop closure. In other words, on the time scale τ_{ee}, resulting in $D_{ee} \sim N^{-\frac{1}{2}}$, the monomers at the chain ends find themselves within a volume $\sim a^3$, so that contact formation is possible. Substituting D_{ee} into eq. (2.7) gives

$$\tau_c \approx \frac{N^2 b^3 \pi}{24\sqrt{6}\, D_0 a} \sim \tau_{WF}, \tag{2.13}$$

in the limit of large N. Thus, within the SSS approximation, the N^2 dependence of τ_c may be obtained, provided the effective diffusion constant D_{ee} is used.

The importance of using a diffusion constant that takes relaxation dynamics of $\mathbf{R}_{ee}$ into account has also been stressed by Portman [56]. However, eq. (2.13) does not account for the possibility of $\tau_c \sim N^{\alpha_\tau}$, with $1.5 < \alpha_\tau < 2$, as observed with simulations by Pastor et. al. [58] when the capture radius $a < b$. Additionally, the closure time in eq. (2.13) depends on the capture radius as a^{-1}, which disagrees with the a-independent prediction of Doi [12]. Both of these discrepancies are discussed below, using insights garnered from simulations.

In order to measure $\mathbf{R}_{ee}(t)$ and τ_c for a non-interacting Freely Jointed chain, we have performed extensive Brownian dynamics simulations, the details of which are given in Appendix A.2. For comparison with the analytic theory, we calculated the modified SSS first passage time, with $P(r)$ given in eq. (A.4), and D_{ee} given in eq. (2.12). The results are shown in Fig. 2.1. We find that the behavior of τ_c depends strongly on the ratio a/b.

For $N \leq 100$ and $a \geq b$, we find that the modified SSS theory using the effective diffusion constant D_{ee} in eq. (2.12) gives an excellent fit to the data, as a function of both N and a (Fig. 2.1(A)). Thus, modeling the loop closure process as a one-dimensional diffusive process in a potential of mean force is appropriate, so long as a diffusion coefficient that takes the dynamics of the chain ends into account is used.

For $N \geq 100$ and $a \geq b$, we notice significant deviations in the data from the theoretical curves. The data points appear to converge as a is varied for large

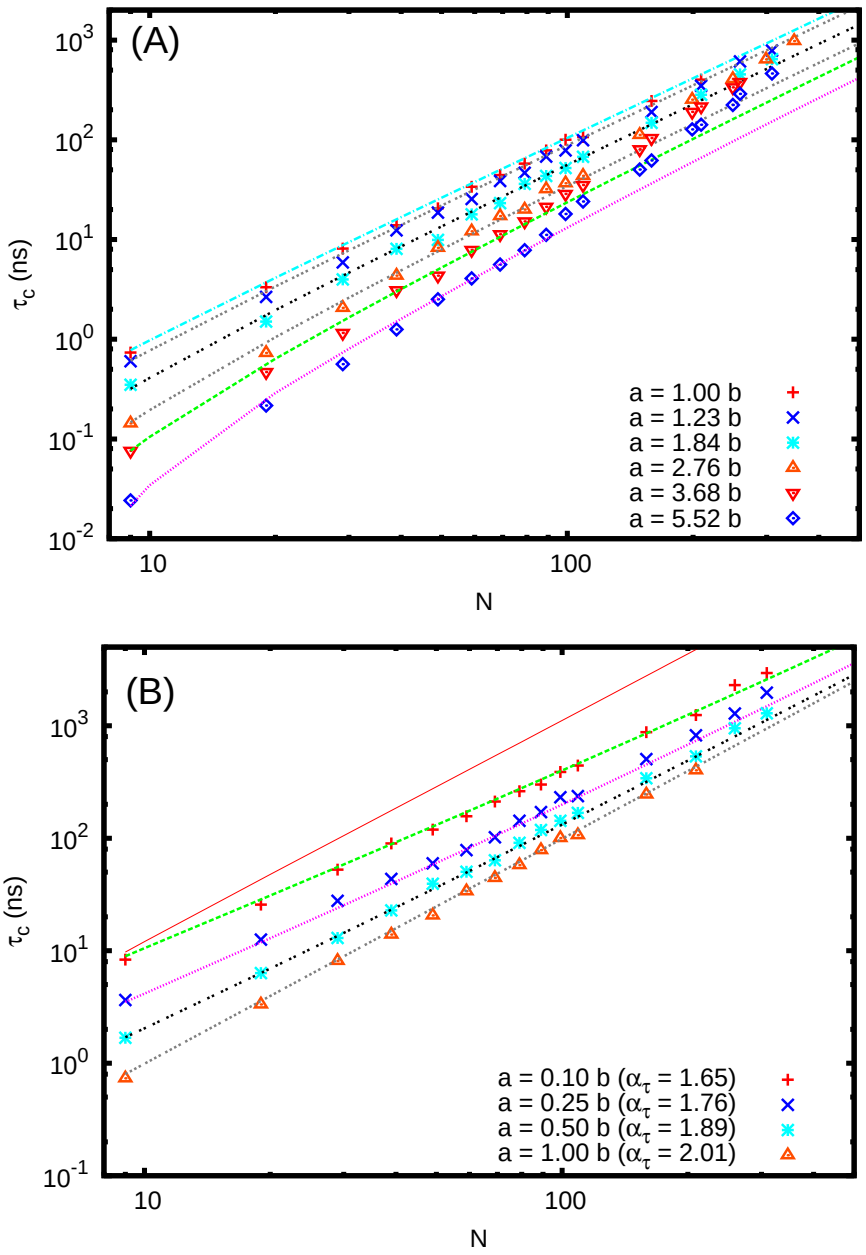

Figure 2.1: Dependence of τ_c on N for various values of a. The symbols correspond to different values of the capture radius. (A): The values of a/b are 1.00 (+), 1.23 ($\times$), 1.84 ($*$), 2.76 (Δ), 3.68 (∇), and 5.52 ($\diamond$). The lines are obtained using eq. (2.6) with $\kappa \to \infty$. The diffusion constant in eq. (2.6) is obtained using $D = \langle \delta \mathbf{R}_{ee}^2(\tau_{ee}/2)/3\tau_{ee} \rangle$, with $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ given in eq. (2.10). (B): The values of a/b are 0.10 (+), 0.25 ($\times$), 0.50 ($*$), and 1.00 (Δ). The lines are the theoretical predictions using eq. (2.15). The poor fit using eq. (2.13) with $a = 0.1b$ (solid line) shows that the two-stage mechanism has to be included to obtain accurate values of τ_c. The effective exponent α_τ, obtained by fitting $\tau_c \sim N^{\alpha_\tau}$, is shown in parentheses.

N, suggesting the emergence of Doi's [12] predicted scaling of $\tau_c \sim N^2 a^0$. This departure from the predictions of eq. (2.13) suggests that the one-dimensional mean field approximation, which gives rise to the a dependence of τ_c, breaks down. Even our modified theory, which attempts to include fluctuations in $\mathbf{R}_{ee}$ on a mean field level leading to D_{ee}, cannot accurately represent the polymer as a diffusive process with a single degree of freedom for large N. In this regime, the many degrees of freedom of the polymer must be explicitly taken into account, making the WF theory [38] more appropriate.

The condition $a < b$ is non-physical for a Freely Jointed Chain with excluded volume, and certainly not relevant for realistic flexible chains in which excluded volume interaction between monomers would prevent the approach of the chain ends to distances less than b. However, that for Wormlike Chains, with the statistical segment $l_p > b$, the equivalent closure condition $a < l_p$ is physically realistic, although we do not address chain stiffness directly here. In this case, we find $\tau_c \sim N^{\alpha_\tau}$, with $1.5 < \alpha_\tau < 2$, in agreement with the simulation results of Pastor et. al. [58] (see Fig. 2.1(B)). In deriving D_{ee}, we assumed [12] that the relaxation of the end-to-end vector is rate limiting. Once $|\mathbf{R}_{ee}| \sim R' \approx 0.4\sqrt{N}b$, the faster internal motions of the chain will search the conformational space rapidly, so that τ_c is dominated by the slower, global motions of the chain (i.e. it is diffusion limited). This assumption breaks down if $a \ll b$, because the endpoints must continue their search using the rapid internal motions on a time scale $\sim b^2/D_0$ (as $D(0) = 2D_0$; see eq. 2.5). In the limit of small a, the memory of the relaxation of the ends of the chain is completely lost. In this case, our derivation of D_{ee}, using a mean field approach, can not accu-

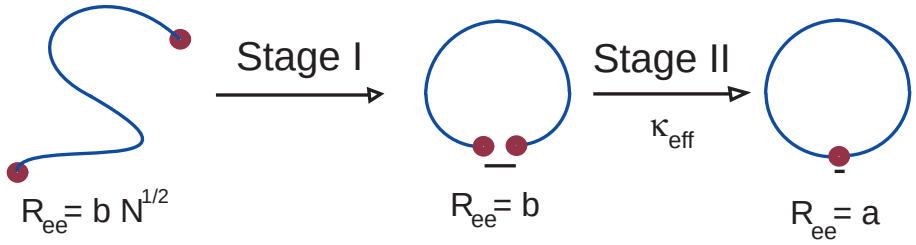

Figure 2.2: Sketch of the two-stage mechanism for loop closure for Rouse chains when $a < b$. Although unphysical, this case is of theoretical interest. In the first stage, fluctuations in $\mathbf{R}_{ee}$ result in the ends approaching $|\mathbf{R}_{ee}| = b$. The search of the monomers within a volume b^3 ($> a^3$), which is rate limiting, leads to a contact in the second stage.

rately describe the finer details when the endpoints search for each other over very small length scales, and hence the theory must be modified in this regime.

We view the loop closure for small a as a two step process (Fig. 2.2), with the first being a reduction to $|\mathbf{R}_{ee}| \sim b$. The first stage is well modeled by our modified SSS theory (see Fig. 2.1(A)) using the effective diffusion coefficient in eq. (2.12). The second stage involves a search for the two ends within a radius b, so that contact can occur whenever $|\mathbf{R}_{ee}| = a < b$. The large scale relaxations of the chain are not relevant in this regime. We therefore introduce a scale-dependent diffusion coefficient

$$D_{ee}(r) \approx \begin{cases} 8D_0/\sqrt{N\pi} & r > b \\ 2D_0 & r \leq b \end{cases}. \tag{2.14}$$

Substitution of eq. (2.14) into eq. (2.6) (with $P(r)$ given by eq. (A.4)) yields

$$\tau_c(a) \approx \frac{N^2 b^2 \pi}{24\sqrt{6}\, D_0} + \frac{N^{3/2} b^2 (b-a)\sqrt{\pi}}{6\sqrt{6} D_0 a}. \tag{2.15}$$

for $a \leq b$. In Fig. 2.1(B), we compare the predictions of eq. (2.15) for the closure time to the simulated data for $a \leq b$. The fit is excellent, showing that the simple

scale-dependent diffusion coefficient (eq. (2.14)), that captures the two stage mechanism of cyclization when $a < b$, accurately describes the physics of loop closure for small a. By equating the two terms in eq. (2.15), we predict that the $N^{3/2}$ scaling will begin to emerge when $N \leq 16b^2(a/b-1)^2/a^2\pi$, consistent with the predictions of Chen et. al. [60].

An alternate (but equivalent) description of the process of loop formation for small a can also be given. After the endpoints are within a sphere of radius b, chain fluctuations will drive them in and out of the sphere many times before contact is established. This allows us to describe the search process using an effective rate constant κ_{eff}, schematically shown in Fig. 2.2. For small a, the loop closure (a search within radius b) becomes effectively rate limited as opposed to diffusion limited [2] contact formation. The search will be successful, in the SSS formalism, on a time scale

$$\tau_{b\to a} \approx \frac{1}{2D_0\mathcal{N}'}\int_a^b \frac{dr}{P(r)}\Big(\int_r^b dr'\ P(r')\Big)^2, \tag{2.16}$$

with $\mathcal{N}' = \int_a^b dr\, P(r)$. Again, we have taken $D = 2D_0$ in this regime, because loop formation is dominated by the fast fluctuations of the monomers, which occur on the time scale of b^2/D_0 (see eq. 2.5). For $a \approx b$, $\tau_{b\to a} \approx (a-b)^2/6D_0$, whereas $\tau_{b\to a} \approx b^3/6aD_0$ as $a \to 0$. $\tau_{b\to a}$ can be used to define the effective rate constant $\kappa_{eff} \propto (b-a)/\tau_{b\to a}$. This can be substituted into eq. (2.6), and gives the approximate loop closure time as $a \to 0$

$$\tau_c(a) - \tau_c(b) \approx \frac{1}{\kappa_{eff}\mathcal{N}'P(b)} \propto \frac{N^{3/2}b^3}{D_0 a}, \tag{2.17}$$

reproducing the same scaling for small a as in eq. (2.15). The parameter κ_{eff}

is analogous to the reaction limited rate [2], reminiscent of the two-state kinetic mechanism used to analyze experimental data. If the search rate within the capture region (given by κ_{eff}) is small, then we expect the exponent $\alpha_\tau < 2$. Indeed, the experiments of Buscaglia et. al. suggest that α_τ changes from 2 (diffusion-limited) to 1.65 (reaction-limited). Our simulation results show the same behavior $\alpha_\tau = 2$ for $a/b \geq 1$, which corresponds to a diffusion limited process, and $\alpha_\tau \approx 1.65$ for $a/b = 0.1$, in which the search within $a/b < 1$ becomes rate limiting.

2.3 Loop Closure for Polymers in Good and Poor Solvents

The kinetics of loop closure can change dramatically when interactions between monomers are taken into account. In good solvents, in which excluded volume interactions between the monomers dominate, it is suspected that only the scaling exponent in the dependence of τ_c on N changes compared to Rouse chains. However, relatively little is known about the kinetics of loop closure in poor solvents in which enthalpic effects, that drive collapse of the chain, dominate over chain entropy. Because analytic work is difficult when monomer-monomer interactions become relevant, we resort to simulations to provide insights into the loop closure dynamics. We use a Lennard-Jones interaction to represent the inter-monomer interactions (see Appendix A.2). The second virial coefficient, defining the solvent quality, is given approximately by

$$v_2(\epsilon_{LJ}) = \int d^3\mathbf{r}\Big[1 - \exp\Big(- \beta H_{LJ}(\mathbf{r})\Big)\Big], \tag{2.18}$$

with $\beta = 1/k_B T$. In a good solvent $v_2 > 0$, while in a poor solvent $v_2 < 0$. A plot of v_2 as a function of ϵ_{LJ} given in Fig. 2.3(A) shows that $v_2 > 0$ when $\beta\epsilon_{LJ} < 0.3$ and $v_2 < 0$ if $\beta\epsilon_{LJ} > 0.3$. Throughout this chapter, we will refer to $\beta\epsilon_{LJ} = 0.4$ as weakly hydrophobic and $\beta\epsilon_{LJ} = 1.0$ as strongly hydrophobic. The classification of the solvent quality based on eq. (2.18) is approximate. The precise determination of the Θ-point ($v_2 \approx 0$) requiring the computation of v_2 for the entire chain. For our purposes, this approximate demarcation between good, Θ, and poor solvents based on eq. (2.18) suffices.

To fully understand the effect of solvent quality on the cyclization time, our simulations run over $\beta\epsilon_{LJ} = i/10$, with $1 \leq i \leq 10$. In our simulations, N was varied from 7 to 300 for each value of ϵ_{LJ}, with a fixed capture radius of $a = 2b = 0.76$nm, with the loop closure time identified as the mean first passage time. In Fig. 2.3(B), we show the scaling of the radius of gyration $\langle \mathbf{R}_g^2 \rangle$ as a function of N. We find $\langle \mathbf{R}_{ee}^2 \rangle \sim N^{\frac{6}{5}}$ for the good solvent and $\langle \mathbf{R}_{ee}^2 \rangle \sim N$ for the Θ solvent ($\beta\epsilon_{LJ} = 0.3$). In poor solvents ($\beta\epsilon_{LJ} > 0.3$), the large N scaling of $\langle \mathbf{R}_{ee}^2 \rangle \sim N^{\frac{2}{3}}$ is not observed for the values of N used in our simulations. Similar deviation from the expected scaling of $\langle \mathbf{R}_{ee}^2 \rangle$ with N have been observed by Rissanou et. al. [63] for short chains in a poor solvent. Simulations using much longer chains ($N \geq 5000$) may be required to observe the expected scaling exponent of 2/3.

The loop closure time for the chains in varying solvent conditions is shown in Fig. 2.4(A) and (B), and it is clear that the solvent quality drastically changes the loop closure time. The values of τ_c for the good solvent ($\beta\epsilon_{LJ} = 0.1$) are nearly three orders of magnitude larger than in the case of the strong hydrophobe

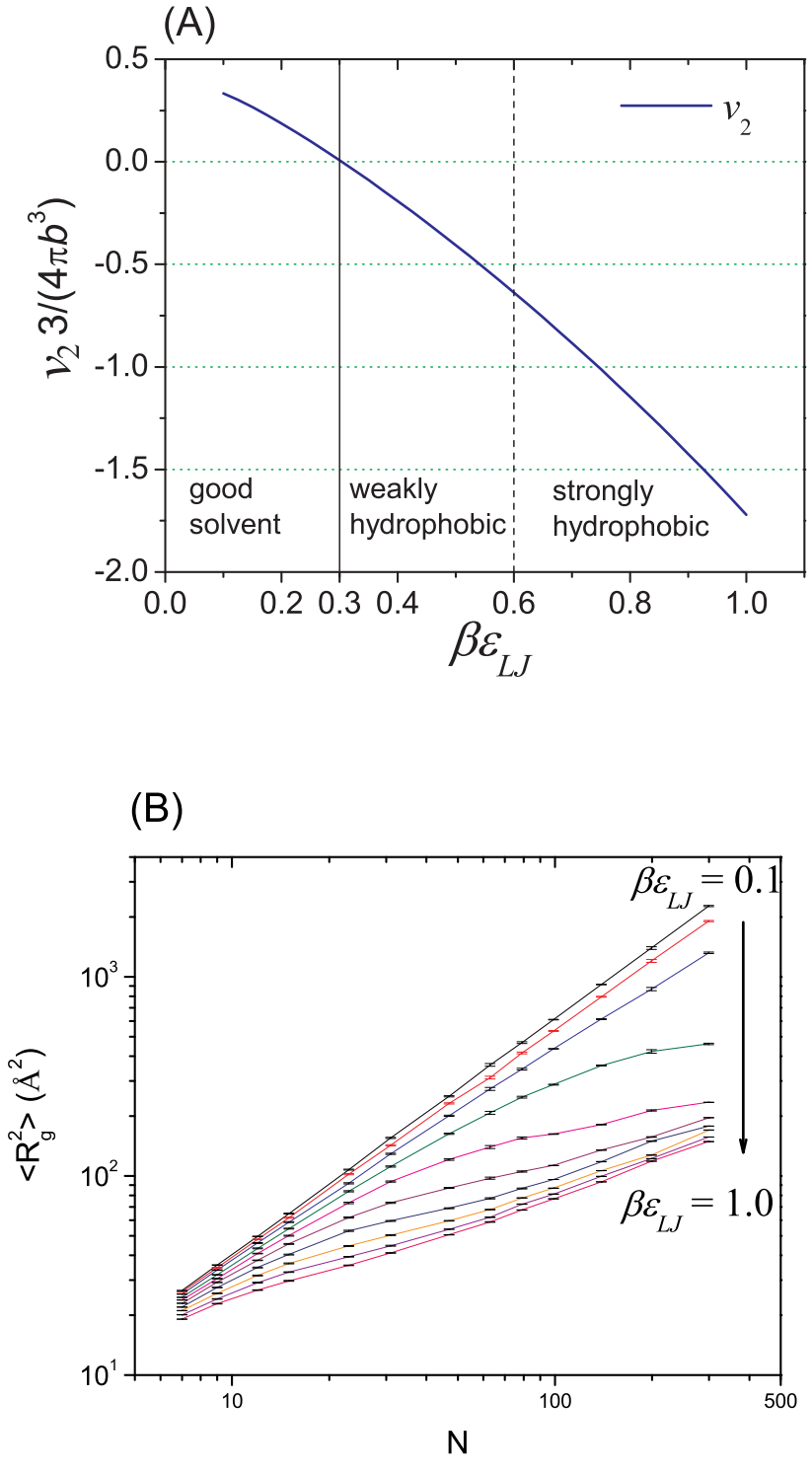

Figure 2.3: (A): Second virial coefficient as a function of ϵ_{LJ}, from eq. (2.18). The classification of solvent quality based on the values of v_2 are shown. (B): The variation of $\langle \mathbf{R}_g^2 \rangle$ with N for different values of ϵ_{LJ}. The value of $\beta\epsilon_{LJ}$ increases from 0.1 to 1.0 (in the direction of the arrow).

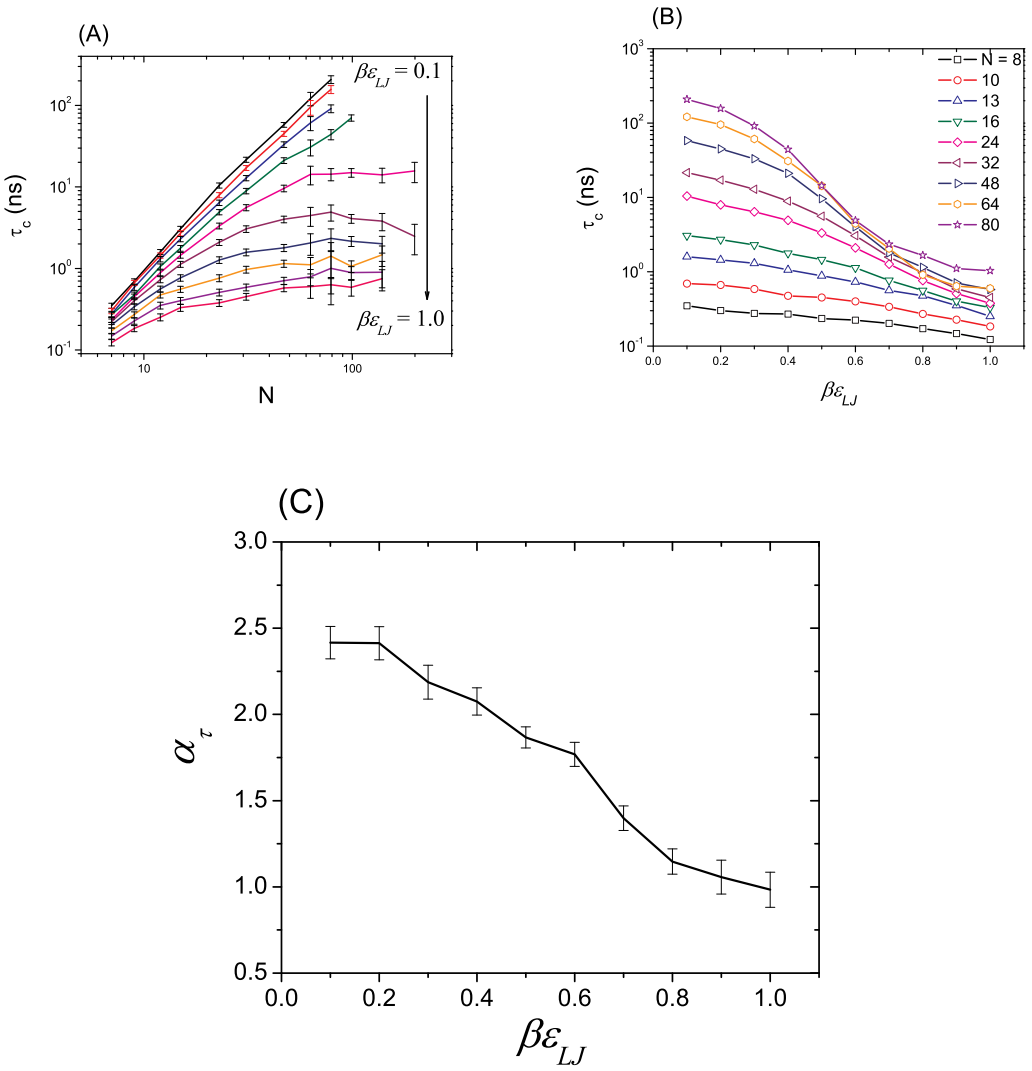

Figure 2.4: (A): Loop closure time as a function of N for varying solvent quality. The values of $\beta\epsilon_{LJ}$ increase from 0.1 to 1.0 from top to bottom, as in Fig. 3(A). (B): τ_c as a function of ϵ_{LJ}, which is a measure of the solvent quality. The values of N are shown in various symbols. (C): Variation of the scaling exponent of $\tau_c \sim N^{\alpha_\tau}$ as a function of ϵ_{LJ}.

($\beta\epsilon_{LJ} = 1.0$) for $N = 80$ (Fig. 2.4(A)). For N in the range of 20 to 30 (typically used in experiments on tertiary contact formation in polypeptide chains) the value of τ_c is about 20ns in good solvents, whereas in poor solvents τ_c is only about 0.3ns. The results are vividly illustrated in Fig. 2.4(B), which shows τ_c as a function of ϵ_{LJ} for various N values. The differences in τ_c are less pronounced as N decreases (Fig. 2.4(B)). The absolute value of τ_c for $N \approx 20$ is an order of magnitude less than obtained for τ_c in polypeptides [2]. There could be two inter-related reasons for this discrepancy. The value of D_0, an effective diffusion constant in the SSS theory, extracted from experimental data and simulated $P(\mathbf{R}_{ee})$ is about an order of magnitude less than the D_0 in our paper. Secondly, Buscaglia et al. [2] used the WLC model with excluded volume interactions whereas our model does not take into account the effect of bending rigidity. Indeed, we had shown in an earlier study [62] that chain stiffness increases τ_c. Despite these reservations, our values of τ_c can be made to agree better with experiments using $\eta \approx 5$cP [45] and a slightly larger value of b. Because it is not our purpose to quantitatively analyze cyclization kinetics in polypeptide chains we did not perform such comparison.

We also find that the solvent quality significantly changes the scaling of $\tau_c \sim N^{\alpha_\tau}$, as shown in Fig. 2.4(C). For the range of N considered in our simulations, τ_c does not appear to vary as a simple power law in N (much like $\langle \mathbf{R}_g^2 \rangle$; see Fig. 2.3(B)) for $\beta\epsilon_{LJ} > 0.3$. The values of τ_c in poor solvents shows increasing curvature as N increases. However, if we insist that a simple power law describes the data then for the smaller range of N from 7 to 32 (consistent with the methods of other authors [54, 2, 50]), we can fit the initial slopes of the curves to determine an effective

exponent α_τ (2.4(C)), i.e. $\tau_c \approx \tau_0 N^{\alpha_\tau}$. In the absence of sound analytical theory, the extracted values of α_τ should be viewed as an effective exponent. We anticipate that, much like the scaling laws for $\langle \mathbf{R}_g^2 \rangle$, the final large N scaling exponent for τ_c will only emerge for [63] $N \geq 5000$, which is too large for accurate simulations. However, with the assumption of a simple power law behavior for small N, we find that the scaling exponent precipitously drops from $\alpha_\tau \approx 2.4$ in the good solvent to $\alpha_\tau \approx 1.0$ in the poor solvent. Our estimate of α_τ in good solvents is in agreement with the prediction of Debnath and Cherayil [43] ($\alpha_\tau \approx 2.3 - 2.4$) or Thirumalai [64] ($\alpha_\tau \approx 2.4$), and is fairly close to the value obtained in previous simulations[61] ($\alpha_\tau \approx 2.2$). The differences between our results and ref. [61] may be related to the choice of the Hamiltonian. Podtelezhnikov and Vologodskii [61] used a harmonic repulsion between monomers to represent the impenetrability of the chain, and took $a/b < 1$ in their simulations.

In contrast to the good solvent case, our estimate of α_τ in poor solvents is significantly lower than the predictions of Debnath and Cherayil [43], who suggested $\alpha_\tau \approx 1.6 - 1.7$, based on a modification of the WF formalism [38]. However, fluorescence experiments on multiple repeats of the possibly weakly hydrophobic glycine and serine residues in D_2O have found $\tau_c \sim N^{1.36}$ for short chains [54] and $\tau_c \sim N^{1.05}$ for longer chains [50], in qualitative agreement with our simulation results. The qualitative agreement between simulations and experiments on polypeptide chains suggest that interactions between monomers are more important than hydrodynamic interactions, which are not considered here.

The dramatically smaller loop closure times in poor solvents than in good

solvents (especially for $N > 20$; see Fig 2.4(B)) suggest further study. In poor solvents, the chain adopts a compact globular conformation with the monomer density $\rho b^3 \sim \mathcal{O}(1)$, where $\rho \approx N/R_g^3$. For large N, when entanglement effects may dominate, it could be argued that in order for the initially spatially separated chain ends ($|\mathbf{R}_{ee}|/a > 1$) to meet, contacts between the monomer ends with their neighbors must be broken. Such unfavorable events might require overcoming enthalpic barriers ($\approx \bar{Q} \times \epsilon_{LJ}$, where $\bar{Q}$ is the average number of contacts for a bead in the interior of the globule), which would increase τ_c. Alternatively, if the ends search for each other using a diffusive, reptation-like mechanism without having to dramatically alter the global shape of the collapsed globule, τ_c might decrease as ϵ_{LJ} increases (i.e. as the globule becomes more compact). It is then of interest to ask whether looping events are preceded by global conformational changes, with a large scale expansion of the polymer that allows the endpoints to search the volume more freely, or if the endpoints search for each other in a highly compact, but more restrictive, ensemble of conformations.

In order to understand the mechanism of looping in poor solvents, we analyze in detail the end-to-end distance $|\mathbf{R}_{ee}(t)|$ and the radius of gyration $|\mathbf{R}_g(t)|$ for two trajectories (with $\beta\epsilon_{LJ} = 1$ and $N = 100$). One trajectory has a fast looping time ($\tau_{cF} \approx 0.003$ns), while the looping time in the other is considerably slower ($\tau_{cS} \approx 4.75$ns). Additionally, we compute the time-dependent variations of the coordination number, $Q(t)$ for each endpoint. We define two monomers i and j to be in 'contact' if $|\mathbf{r}_i - \mathbf{r}_j| \geq 1.23b$ (chosen since $V_{LJ}(1.23b) = -\epsilon_{LJ}/2$, see Appendix A.2), and define $Q_1(t)$ and $Q_N(t)$ to be the total number of monomers in contact

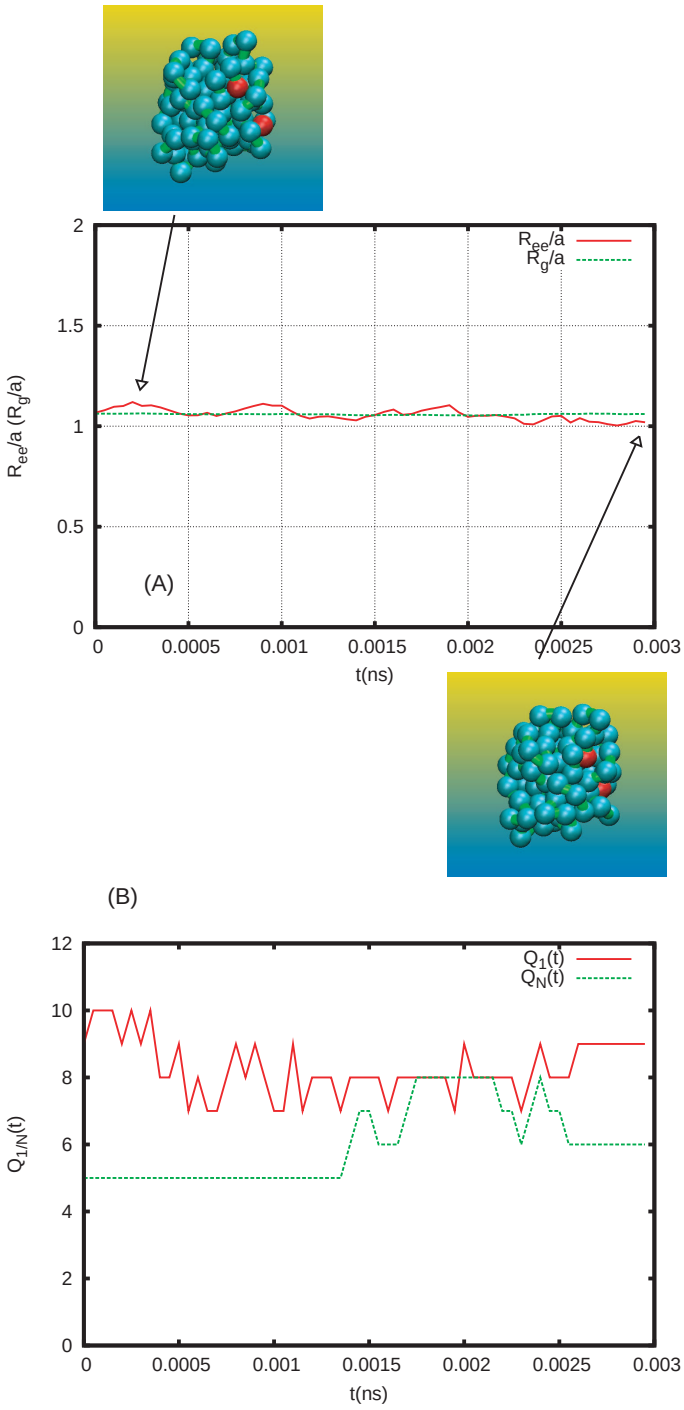

Figure 2.5: Mechanism of loop closure for a trajectory with a short ($\sim$ 0.003ns) first passage time. The values of N and $\beta\epsilon_{LJ}$ are 100 and 1.0 respectively. (A): Plots of $|\mathbf{R}_{ee}|$ and $|\mathbf{R}_g|$ (scaled by the capture radius (a) as a function of time. The structures of the globules near the initial stage and upon contact formation between the ends are shown. The end to end distance is in red. (B): The time-dependent changes in the coordination numbers for the first ($Q_1(t)$) and last ($Q_N(t)$) monomers during the contact formation.

with monomers 1 and N respectively. We do not include nearest neighbors on the backbone when computing the coordination number, so the geometrical constraints gives $0 \leq Q(t) \leq 11$. With this definition, an endpoint on the surface of the globule will have $Q = 5$. These quantities are shown in Figs 2.5 and 2.6.

The trajectory with the fast looping time ${\tau_c}^F$ (Fig 2.5) shows little variation in either $|\mathbf{R}_g|$ or $|\mathbf{R}_{ee}|$. We find $|\mathbf{R}_{ee}| \approx |\mathbf{R}_g|$, suggesting that the endpoints remain confined within the dense globular structure throughout the looping process. This is also reflected in the coordination numbers for both of the endpoints, with both $Q_1(t)$ and $Q_N(t)$ are in the range $5 \leq Q(t) \leq 10$ throughout the simulation. The endpoints in this trajectory, with the small loop closure time τ_c^F, always have a significant number of contacts, and traverse the interior of the globule when searching for each other. Similarly, we also found that the trajectory with a slow looping time τ_c^S (Fig 2.6) shows little variation in $\mathbf{R}_g$ throughout the run. The end-to-end distance, however, shows large fluctuations over time, and $\langle \mathbf{R}_{ee}^2 \rangle \geq 2\langle \mathbf{R}_g^2 \rangle$ until closure. This suggests that, while the chain is in an overall globular conformation (small, constant $\mathbf{R}_g^2$), the endpoints are mainly found on the exterior of the globule. This conclusion is again supported by the coordination number, with $Q(t) \leq 5$ for significant portions of the simulation. While the endpoints are less restricted by nearby contacts and able to fluctuate more (as they are outside of the globule), they spend much longer searching for each other. Thus, it appears that the process of loop formation in poor solvents, where enthalpic effects might be expected to dominate for $N = 100$ occurs by a diffusive process, with no significant entanglement effects observed in our simulations.

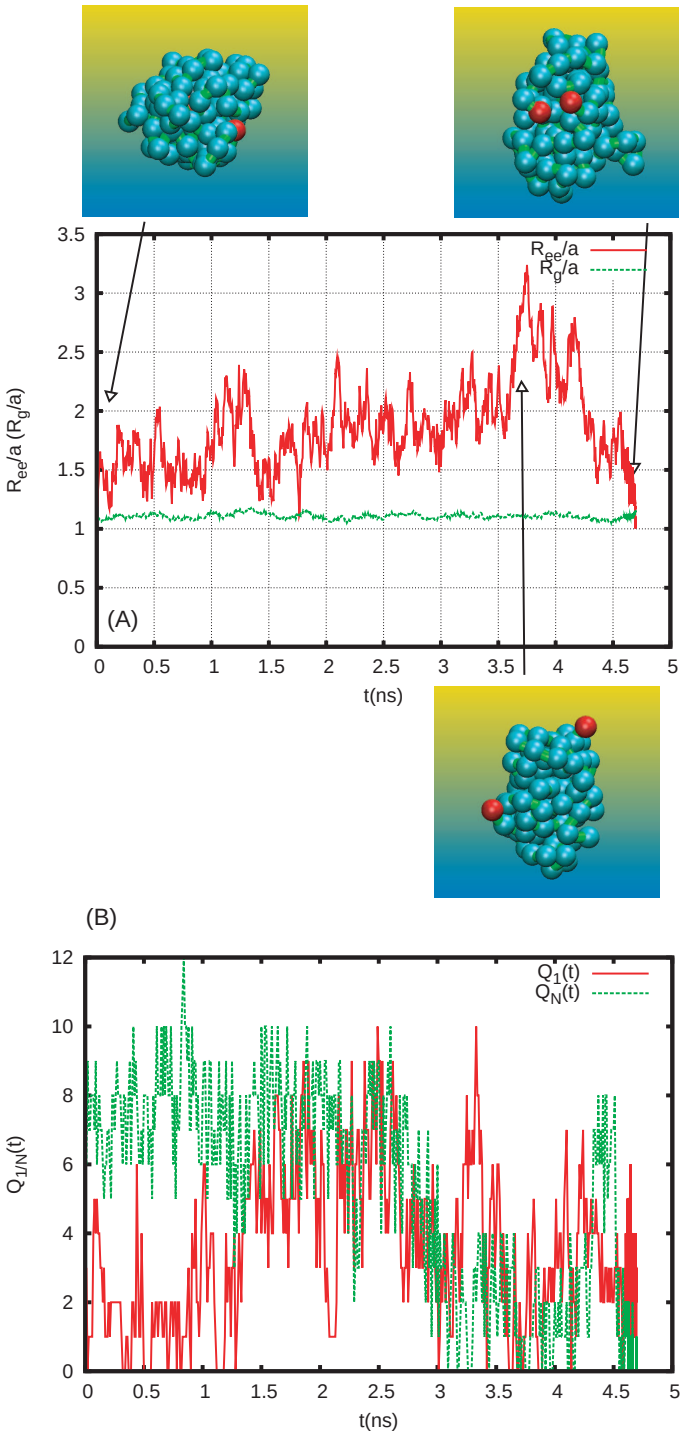

Figure 2.6: Same as Fig. 2.5, except the data are for a trajectory with a first passage time for contact formation that is about 4.7ns. (A): Although the values of $|\mathbf{R}_g|$ are approximately constant, $|\mathbf{R}_{ee}|$ fluctuations greatly. (B): Substantial variations in $Q_1(t)$ and $Q_N(t)$ are observed during the looping dynamics, in which both ends spend a great deal of time on the surface of the globule.

We note that trajectories in which the first passage time for looping is rapid (with ${\tau_c}^i < \tau_c$ for trajectory i) have at least one endpoint with a high coordination number ($Q > 5$) throughout the simulation. In contrast, for most slow-looping runs (with ${\tau_c}^i > \tau_c$), we observe long stretches of time where both endpoints have a low coordination number ($Q < 5$). These results suggest that motions within the globule are far less restricted than one might have thought, and loop formation will occur faster when the endpoints are within the globule than if the endpoints are on the surface (i.e. entanglement effects are irrelevant for loop closure with $N \leq 100$). The longer values of τ_c are found if the initial separation of the end points is large, which is more likely if they are on the surface than buried in the interior. The absence of any change in $|\mathbf{R}_g(t)|$ in both the trajectories, which represent the extreme limits in the first passage time for looping, clearly shows that contact formation in the globular phase is not an activated process.

The above results suggest a very general mechanism of loop closure for interacting polymers. The process of contact formation for a given trajectory depends on the initial separation $\mathbf{R}_{ee}$, and the dynamics of the approach of the ends. Thus, τ_c should be determined by the distribution of $P(\mathbf{R}_{ee})$ (an equilibrium property), and an effective diffusion coefficient $D(t)$ (a dynamic property). We have shown for the Rouse model that such a deconvolution into equilibrium and dynamic parts, which is in the spirit of the SSS approximation, is accurate in obtaining τ_c for a wide range of N and a/b, and will see below that a similar framework is applicable to more general chains.

The decomposition of looping mechanisms into a convolution of equilibrium

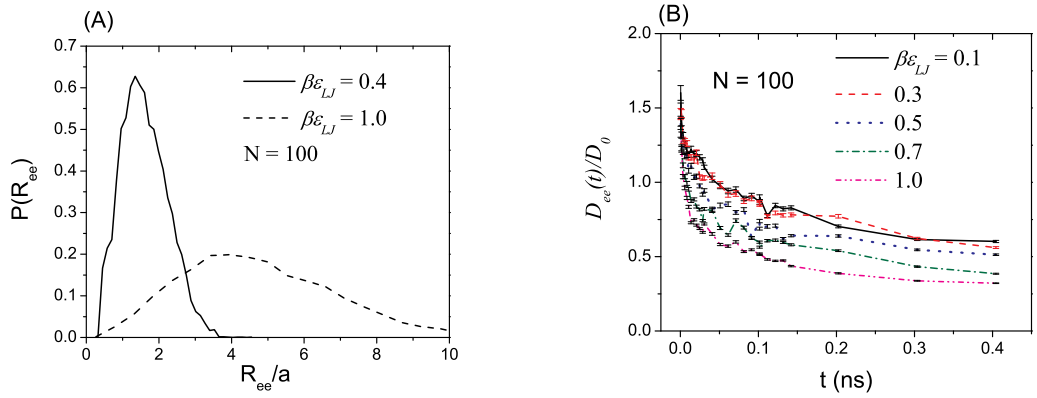

Figure 2.7: (A): Distribution of end-to-end distances for a weakly ($\beta\epsilon_{LJ} = 0.4$) and strongly ($\beta\epsilon_{LJ} = 1.0$) hydrophobic chain. (B): Diffusion constant $D_{ee}(t)$ in units of D_0 for varying solvent quality. The diffusion constant is defined using $D_{ee}(t) = \langle \delta \mathbf{R}_{ee}^2(t) \rangle / 6t$. The values of ϵ_{LJ} are shown in the inset.

and dynamical parts explains the large differences in τ_c as the solvent quality changes. We find, in fact, that the equilibrium behavior of the endpoints dominates the process of loop formation, with the kinetic processes being only weakly dependent on the solvent quality for short chains. In Fig. 2.7(A), we plot the end-to-end distribution function for weakly ($\beta\epsilon_{LJ} = 0.4$) and strongly ($\beta\epsilon_{LJ} = 1$) hydrophobic polymer chains. The strongly hydrophobic chain is highly compact, with a sharply peaked distribution. The average end-to-end distance is significantly lower than is the weakly hydrophobic case. While the distribution function is clearly strongly dependent on the interactions, the diffusion coefficient $D(t)$ is only weakly dependent on the solvent quality (Fig. 2.7(B)). The values of $D(t) = \langle \delta \mathbf{R}_{ee}^2 \rangle / 6t$ are only reduced by a factor of about 2 between the $\beta\epsilon_{LJ} = 0.1$ (good solvent, with a globally swollen configuration) and the $\beta\epsilon_{LJ} = 1.0$ (poor solvent, with a globally globular configuration) on intermediate time scales. We note, in fact, that the good solvent and Θ solvent cases have virtually identical diffusion coefficients throughout

the simulations (Fig 2.7(B)). This suggests that the increase in τ_c (Fig 2.4) between the Rouse and the good solvent chains is primarily due to the broadening of the distribution $P(\mathbf{R}_{ee})$, i.e. the significant increase in the average end-to-end distance in the good solvent case, $\langle \mathbf{R}_{ee}^2 \rangle \sim N^{2\nu}$, with $\nu = 3/5$.

In fact, if we take $D \approx 2D_0$ in eq. (2.6) and numerically integrate the distribution function found in the simulations for $N = 100$, $\tau_c(\beta\epsilon_{LJ} = 1.0)$ and $\tau_c(\beta\epsilon_{LJ} = 0.4)$ differ by two orders of magnitude, almost completely accounting for the large differences seen in Fig. 2.4(B) between the two cases. Because $P(\mathbf{R}_{ee})$ can, in principle, be inferred from FRET experiments [65, 66] the theory outlined here can be used to quantitatively predict loop formation times. In addition, FRET experiments can also be used to assess the utility of polymer models in describing fluctuations in single stranded nucleic acids and polypeptide chains.

It is also of interest to determine the mechanisms of interior loop formation. We performed simulations for $N = 80$ by first computing the time for cyclization τ_c using the methods above. In another set of simulations, two interacting linkers each containing 20 beads were attached to the ends of the $N = 80$ chain. For the resulting longer chain we calculated τ_l for $l = 80$ as a function of $\beta\epsilon_{LJ}$. Such a calculation is relevant in the context of single molecule experiments in which the properties of a biomolecule (RNA) is inferred by attaching linkers with varying polymer characteristics. It is important to choose the linker characteristics that minimally affects the dynamic properties of the molecule of interest. The ratio $\tau_{l=80}/\tau_c$ depends on $\beta\epsilon_{LJ}$ and changes from 2.6 (good solvents) to 2.0 under Θ conditions, becoming unity in poor solvents (Fig. 2.8). Analysis on the dependence

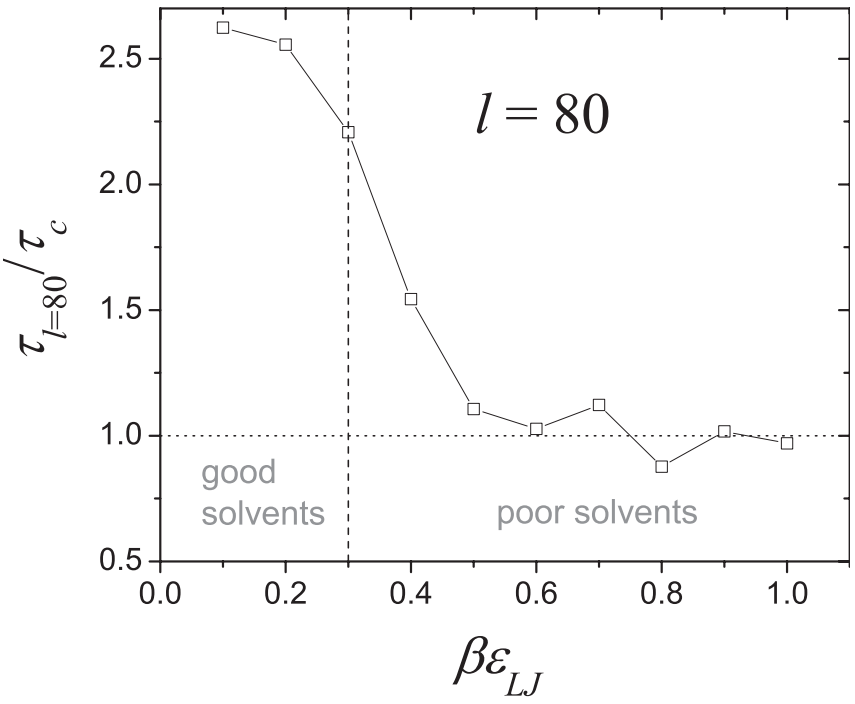

Figure 2.8: The ratio $\tau_{l=80}/\tau_c$ as a function of $\beta\epsilon_{LJ}$ for a chain with two linkers (each of 20 beads) that are attached to beads 20 and 100. In good solvents, the interior loop closure kinetics is about 2.5 times slower than the end-to-end one with the same loop length. In poor solvents, however, there is virtually no difference between the two.

of the diffusion coefficients of interior-to-interior vector D_{ij} ($i = 20$ and $j = 100$) and end-to-end vector (of original chain without linkers) D_{ee} on solvent conditions indicates that on the time scales relevant to loop closure time (analogous to τ_{ee} for the Rouse chain), D_{ij} reduces to about one half of D_{ee} in good and Θ solvents, whereas the two are very similar in poor solvents. The changes in the diffusion coefficient together with the equilibrium distance distribution explains the behavior in Fig. 2.8.

2.4 Conclusions

A theoretical description of contact formation between the chain endpoints is difficult because of the many body nature of the dynamics of a polymer. Even for the simple case of cyclization kinetics in Rouse chains, accurate results for τ_c are difficult to obtain for all values of N, a, and b. The present work confirms that, for large N and $a/b > 1$, the looping time must scale as N^2, a result obtained using the WF formalism [38, 12]. Here, we have derived $\tau_c \sim N^2$ (for $N \gg 1$ and $a \geq b$) by including the full internal chain dynamics within the simple and elegant SSS theory [11]. We have shown that, for $N < 100$ and especially in the (unphysical) limit $a/b < 1$, the loop closure time $\tau_c \sim \tau_0 N^{\alpha_\tau}$ with $1.5 < \alpha_\tau \leq 2$. In this limit, our simulations show that loop closure occurs in two stages with vastly differing time scales. By incorporating these processes into a scale-dependent diffusion coefficient, we obtain an expression for τ_c that accurately fits the simulation data. The resulting expression for τ_c for $a < b$ (eq. (2.15)) contains both the $N^{\frac{3}{2}}$ and N^2 scaling behavior,

as suggested by Pastor et. al. [58]

The values of τ_c for all N change dramatically when interactions between monomers are taken into account. In good solvents, $\tau_c \sim \tau_0 N^{\alpha_\tau}$ ($\alpha_\tau \approx 2.4$) in the range of N used in the simulations, in reasonable agreement with earlier theoretical estimates [64, 43]. The simulation results for τ_c in poor solvents show a rich behavior, with τ_c decreasing over two orders of magnitude as the solvent quality changes. In the range of N considered here, a power law fit of τ_c with N ($\tau_c \sim N^{\alpha_\tau}$) also shows that the exponent α_τ depends on the solvent quality. Analysis of the trajectories that monitor loop closure shows that contact between each end of the chains is established by diffusive, reptation-like motion within the dense, compact globular phase.

The large variations of τ_c as the solvent quality changes suggests that there ought to be significant dependence of loop formation rates on the sequence in polypeptide chains. In particular, our results suggest that as the number of hydrophobic residues increase, τ_c should decrease. Similarly, as the number of charged or polar residues increase, the interactions between amino acids will become more repulsive, increasing τ_c. Our work, however, neglects sequence effects, which are also likely to be important. It may be possible to directly test the predictions in Fig. 2.4(B) by varying the solvent quality for polypeptides. A combination of denaturants (makes the solvent quality good) and PEG (makes it poor) can be used to measured τ_c in polypeptide chains. We expect the measured τ_c should be qualitatively similar to the findings in Fig. 2.4(B).

The physics of loop closure for small and intermediate chain lengths ($N \leq 300$)

is rather complicated, due to contributions from various time and length scales (global relaxation and internal motions of the chains). The contributions from these sources are often comparable, making the process of looping dynamics difficult to describe theoretically. To better understand the interplay of these scales, we have explored wide ranges of conceivable parameters, namely the chain length N, capture radius a, and conditions of the solvents expressed in terms of ϵ_{LJ}. By combining analytic theory and simulations, we have shown that, for a given N, the looping dynamics in all solvent conditions is primarily determined by the initial separation of the end points. The many body nature of the diffusive process is embodied in $D(t)$, which does not vary significantly as the solvent quality changes for a fixed N. We find that the process of loop formation can be divided quite generally into equilibrium and kinetic contributions.

Chapter 3

The Extension of a Homopolymer Under an External Tension

3.1 Introduction

Single molecule nanomanipulation methods have been used to measure the response of biological macromolecules to mechanical force. Such measurements give direct estimates of the elasticity of DNA [67], RNA [68], proteins [69, 70], and polysaccharides [71]. Single molecule experiments allow the study of the tension-induced stretching of RNA [17, 18, 72] and proteins [18, 73]. A theoretical understanding of the resulting force-extension curves (FECs) is desired. Sequence effects [74] make this a very difficult problem, but stretching of homopolymers by force provides a potentially simpler case for which the FECs can be calculated.

In a pioneering paper, Pincus [39] considered the strong stretching of homopolymers in a good solvent. The strong stretching limit corresponds to a large enough force, f, such that $N^{\nu}a < \langle Z \rangle \ll Na$, where $\nu = 3/5$ is the Flory exponent, N is the number of monomers, a is the size of a monomer, and $\langle Z \rangle = \langle z_N - z_0 \rangle$ is the mean tension-induced end-to-end distance (with $\mathbf{f} = f\hat{\mathbf{z}}$). Pincus showed that the size of the stretched polymer should be determined by an interplay between the Flory radius $R_F = N^{\nu}a$ and the tensile screening length (or the blob size) [39], $\xi_P = k_B T/f$. When f is small, then $x = R_F/\xi_P \ll 1$, while in the opposite limit, $x \gg 1$. The scaling assumption is that for arbitrary f, the average end-to-end

distance can be written as

$$\langle Z \rangle = R_F \Phi(R_F/\xi_P) \tag{3.1}$$

Assuming such a scaling argument holds, the FEC can be divided into three regimes.

(*i*) For small f, we expect a linear increase in the extension of the chain. At low forces, $\Phi(x) \approx x$, and hence $\langle Z \rangle \propto R_F^2 \times (a\beta f)$. (*ii*) In the strongly stretched limit, which arises for intermediate forces, the value of $\langle Z \rangle$ can be obtained by dividing the chain into a sequence of aligned tensile blobs (along the force axis) whose size is [75] $\xi_P \sim (\beta f)^{-1}$. The monomers contained within each blob behave as an unperturbed self-avoiding walk. In this case $\xi_P = (\beta f)^{-1} \sim N_b^{\nu}$, with N_b the number of monomers in a blob. The linear extension of the chain is then given by $\langle Z \rangle \sim \xi_P \times N/N_b \sim N(\beta f)^{\frac{1}{\nu}-1} \sim N(\beta f)^{\frac{2}{3}}$. We will refer to this intermediate scaling regime as the Pincus regime. It should be stressed that this argument is valid only if $\xi_P/a \gg 1$, which may not be satisfied for a stiff polymer, or a flexible polymer with small N (see below). (*iii*) For extremely large forces (beyond the strong stretching regime), we expect the excluded volume to become irrelevant, as the bonds between monomers become fully aligned with the z-axis, and inter-monomer interactions become rare. The FEC in this regime will be model-dependent, with $\langle Z \rangle \approx Na^2\beta f/3$ for an extensible chain, and $\langle Z \rangle \approx Na$ for a inextensible chain. We will refer to this behavior as the non-universal regime.

The Pincus scaling description of the stretching of homopolymers is well known. However, as far as we are aware, a microscopic derivation of the FEC anticipated by Pincus has not been directly shown. More importantly, it is unclear how the

FEC of polymers with finite N compares with the predictions of the scaling theory. In other words, for finite values of N (on the order of 1000), how pronounced is the Pincus regime? In this paper, we develop a self-consistent, variational theory based on the Edwards-Singh method [40] to determine the average extension of a homopolymer in a good solvent. The theory gives excellent agreement with simulations. Surprisingly, neither the theoretical predictions nor simulations display the Pincus regime for $N = 100$ or $N = 1600$. We show that this is due to a finite-size effect, and show that the Pincus regime emerges only for $N \geq 10^5$. Only when N is sufficiently large is the concept of the tensile blob (with $\xi_P \sim aN_b^{\nu}$) satisfied, where $N \gg N_b \gg 1$. Additionally, we show using both theory and simulations that the tension induced stretching of homopolymers in a poor solvent exhibits a first order transition between an ensemble of collapsed states and rod-like conformations. The nature of the transition is dependent on the solvent quality, which is measured in terms of the relative attraction between the monomers. The theoretical predictions for the poor solvent case are only in qualitative agreement with the simulations, with simulations showing that tension-induced transitions occur via a hierarchy of structures, depending on the solvent quality. Force-dependent structure factors show that, for a weakly hydrophobic polymer, the transition to the fully extended state occurs through a variety of structures, depending on the length scales of interest.

3.2 Polymers Under Tension in a Good Solvent

Extensible Polymer: The continuum Hamiltonian for a self-avoiding Gaussian

chain under tension is taken to be

$$\beta H_0 \;=\; \frac{3}{2a^2}\int_0^N ds\ \dot{\mathbf{r}}^2(s) - \beta f \int_0^N ds\ \dot{z}(s) + \Delta_2, \tag{3.2}$$

where f is aligned with the z-axis, $\beta = 1/k_B T$. The excluded volume is represented by

$$\Delta_2 \;=\; \frac{v_0}{2}\int_0^N ds \int_0^N ds'\ \delta[\mathbf{r}(s) - \mathbf{r}(s')], \tag{3.3}$$

with v_0 the strength of the self-avoiding interaction, with $v_0 > 0$ in a good solvent. To compute the force-extension curves (FECs) and compare them to simulations, we use a self-consistent variational method, originally proposed by Edwards and Singh [40]. Following the convention in single molecule experiments, we use the term FEC for the extension changes upon application of force. However, throughout the paper, we will derive and plot the extension $\langle Z \rangle$ as a function of f. A reference Hamiltonian

$$\beta H_1 \;=\; \frac{3}{2a^2\lambda^2}\int_0^N ds\ \dot{\mathbf{r}}^2(s) - \beta f \int_0^N ds\ \dot{z}(s), \tag{3.4}$$

is chosen, and the parameter λ is determined self-consistently. Because we are interested in calculating the FEC of a self-avoiding chain, the relevant quantity is the dependence of $\langle Z(f)\rangle_0 = \langle z_N - z_0\rangle = \int_0^N ds\, \langle \dot{z}(s)\rangle_0$ on f, where $\langle \cdots \rangle_0$ indicates the average with respect to the Hamiltonian βH_0 (eq. 3.2). Because it is not possible to compute the exact average $\langle Z\rangle_0$, we calculate the difference between $\langle Z\rangle_0$ and $\langle Z\rangle_1$ (where $\langle \cdots \rangle_1$ is the average with respect to βH_1), assuming that $\Delta_1 + \Delta_2$ is small, with

$$\Delta_1 \;=\; \frac{3}{2a^2}\Big(1 - \frac{1}{\lambda^2}\Big)\int_0^N ds\ \dot{\mathbf{r}}^2(s). \tag{3.5}$$

To first order in $\Delta_1 + \Delta_2$, we obtain [40]

$$\langle Z\rangle_0 - \langle Z\rangle_1 = \langle Z(\Delta_1 + \Delta_2)\rangle_1 - \langle Z\rangle_1\langle \Delta_1 + \Delta_2\rangle_1. \tag{3.6}$$

A self-consistent equation for λ is obtained by insisting that $\langle Z\rangle_0 \approx \langle Z\rangle_1$, which leads to the condition

$$\langle Z(\Delta_1 + \Delta_2)\rangle_1 = \langle Z\rangle_1\langle \Delta_1 + \Delta_2\rangle_1. \tag{3.7}$$

Throughout this work, we compute averages with respect to βH_1 only, so the subscripts on $\langle\cdots\rangle$ will be dropped. The terms involving Δ_1, Δ_2 are easily calculated using

$$\langle Z\Delta_1\rangle - \langle Z\rangle\langle\Delta_1\rangle = \frac{1}{2}\lambda(\lambda^2-1)\,\frac{\partial\langle Z\rangle}{\partial\lambda} = \frac{\lambda^2(\lambda^2-1)Na^2}{6}\beta f \tag{3.8}$$

and

$$\langle Z\Delta_2\rangle - \langle Z\rangle\langle\Delta_2\rangle = \frac{v_0}{2}\int_0^N ds\int_0^N ds'\frac{\partial}{\partial(\beta f)}\langle\delta[\mathbf{r}(s)-\mathbf{r}(s')]\rangle \tag{3.9}$$

with $\langle\delta[\mathbf{r}(s)-\mathbf{r}(s')]\rangle = (3/2\pi a^2\lambda^2|s-s'|)^{\frac{3}{2}}\exp\left(-|s-s'|\lambda^2a^2\beta^2f^2/6\right)$ (the details of the calculations are given in Appendix B.2). Using Eqs. (3.7), (3.8), and (3.9), the self-consistent equation for λ becomes

$$\begin{aligned}
\lambda^2-1 &= \frac{v\sqrt{N}}{\lambda^3}\int_\delta^1 du\frac{1-u}{\sqrt{u}}\;e^{-Nu\lambda^2\varphi^2/6} \\
&= \frac{6v}{\lambda^5\varphi^2\sqrt{N}}\Big\{e^{-N\lambda^2\varphi^2/6} - \sqrt{\delta}\,e^{-\delta N\lambda^2\varphi^2/6} \\
&\quad + \frac{\varphi}{\lambda^3}\sqrt{\frac{N\pi}{6}}\Big(1-\frac{3}{N\lambda^2\varphi^2}\Big)\Big[\mathrm{erf}\Big(\lambda\varphi\sqrt{\frac{N}{6}}\Big) - \mathrm{erf}\Big(\lambda\varphi\sqrt{\frac{\delta N}{6}}\Big)\Big]\Big\},
\end{aligned} \tag{3.10}$$

where we have defined the dimensionless excluded volume parameter $v = (3/2\pi)^{\frac{3}{2}}\,v_0/a^3$, the dimensionless force $\varphi = a\beta f$, and where $\mathrm{erf}(x)$ is the error function. We have also included a cutoff, δ, in the integral over u (with $u = |s-s'|/N$), to account for

the finite separation between the monomers, which is neglected in the continuum representation of the Hamiltonian in Eq. 3.4. As the system is truly discrete in both simulations and experiments, such a cutoff may be necessary in the continuum limit. We expect $\delta \sim \lambda/N$, since the discrete monomers are separated by a distance $|\mathbf{r}_{i+1} - \mathbf{r}_i| \approx \lambda a$ on an average in the reference Hamiltonian, βH_1. The cutoff is usually only imposed in theories that have a self-energy divergence [40, 76, 77, 78], and is generally not required if there is no divergence inherent in the theory, as is the case here. However, we will see that this cutoff is essential in order to reproduce the FECs obtained in simulations. Given a solution λ to the SCE, the linear end-to-end distance is given by $\langle Z \rangle = Na^2\lambda^2\beta f/3$ (see Appendix B.2).

It is not difficult to show that, as $f \to 0$, a solution to Eq. 3.7 is $\lambda \approx \lambda_0 \propto (v^2 N)^{\frac{1}{10}}$, giving the expected linear regime, $\langle Z \rangle \sim N^{\frac{6}{5}} v^{\frac{2}{5}} a \times (a\beta f)$. We immediately see that this gives the correct scaling with N and v for low forces, with $\langle Z \rangle \approx \langle \mathbf{R}^2 \rangle_{f=0} \times (\beta f)/3$. We also note that, if we set $\delta = 0$, we exactly recover (in our notation) the original, tension-free ($f = 0$) self-consistent equation for a self-avoiding chain, $\lambda^2 - 1 = \sqrt{6N/\pi^3}\, v_0 a^3/\lambda^3$ developed by Edwards and Singh [40].

For intermediate f, we can obtain the correct Pincus scaling for large N. If we assume $\lambda \approx \lambda_0$, we find $N\lambda_0^2\varphi^2 \gg 1$ when $\varphi \approx a\beta f_T \sim N^{-\frac{3}{5}} v^{-\frac{1}{5}}$, defining the transition force f_T into the strongly stretched Pincus regime. For $f \geq f_T$, we can neglect terms of order N^{-1} and $\exp(-N\lambda^2\varphi^2/6)$ for large N, and set $\mathrm{erf}(\lambda\varphi\sqrt{N/6}) \approx 1$. This gives the approximate SCE

$$\lambda^2 - 1 \approx \frac{v\sqrt{6\pi}}{\lambda^4\varphi}\left[1 - \mathrm{erf}\left(\lambda\varphi\sqrt{\frac{N\delta}{6}}\right)\right] + O(N^{-1}). \tag{3.11}$$

With $\delta \sim \lambda/N$, we see that we can neglect the error function in this regime as well if $a\beta f_T \sim \lambda_0^{-\frac{3}{2}} v^{\frac{1}{10}} N^{-\frac{9}{20}} \ll 1$. If N is sufficiently large to satisfy this requirement, the SCE becomes $\lambda^2 - 1 \approx v\sqrt{6\pi}/\lambda^4\varphi + O(vN^{-\frac{1}{4}})$. We thus find the approximate solution in the Pincus regime $\lambda \approx \lambda_P \propto (v/\varphi)^{\frac{1}{6}}$. For large N and intermediate forces, we find $\langle Z \rangle \propto N v^{\frac{1}{3}} f^{\frac{2}{3}}$, as is expected of the Pincus scaling [39]. Note that neglecting terms of order $vN^{-\frac{1}{4}}$ may be valid only for extremely large N (on the order of $N \sim 10^5$). Thus, the onset of the non-linear scaling regime depends on both v and N, as was anticipated by Pincus.

For sufficiently large φ, we can neglect terms of order φ^{-1} in Eq. 3.10, to find an extended or rod-like solution $\lambda \approx \lambda_E = 1$. This root gives $\langle Z \rangle \approx Na^2\beta f/3$, identical to the non-interacting average for an extensible chain. This is not surprising; as the tension becomes large, the excluded volume interaction is not relevant. We also note that, in this regime, the chain will become greatly overextended. As was shown by Pincus, the extension beyond the non-linear regime is non-universal and depends on the precise model used for the homopolymer [39].

Inextensible Polymer: Because the extensible polymer can overstretch for large forces, which does not occur for real polymers that are linked by covalent bonds, we wish to develop a theory for an approximately inextensible model. We were also motivated to consider the inextensible model because Monte Carlo simulations (which are less computationally expensive, allowing for longer chains to be studied) hold the distance between successive beads fixed at precisely a, independent of f.

From the discrete, non-interacting, spring-like Hamiltonian

$$H[\{\mathbf{r}_n\}] = \frac{k}{2a^2}\sum_n (|\Delta\mathbf{r}_n| - a)^2 - \beta f \sum_n \Delta z_n, \tag{3.12}$$

the average end-to-end distance, as well as the fluctuations in the transverse and longitudinal directions can be computed exactly. Defining $X = x_N - x_0$ (with x transverse to the force axis), we find

$$\begin{aligned}
\frac{\langle Z\rangle}{Na} &= \frac{1}{\mathcal{N}}\int_0^\infty dx\ x^2 e^{-k(x-1)^2/2}\cosh(\varphi x) - \frac{1}{\varphi} && (3.13)\\
\frac{\langle Z^2\rangle - \langle Z\rangle^2}{Na^2} &= \frac{1}{\mathcal{N}}\int_0^\infty dx\ x^3 e^{-k(x-1)^2/2}\sinh(\varphi x) - \Big(\frac{\langle Z\rangle}{Na}\Big)^2 - \frac{2}{\varphi}\Big(\frac{\langle Z\rangle}{Na}\Big) && (3.14)\\
\frac{\langle X^2\rangle}{Na^2} &= \frac{1}{\varphi}\Big(\frac{\langle Z\rangle}{Na}\Big), && (3.15)
\end{aligned}$$

where $\mathcal{N} = \int_0^\infty dx\ xe^{-k(x-1)^2/2}\sinh(\varphi x)$, and we have used $\langle Z^2\rangle = N\langle z_n^2\rangle + N(N-1)\langle z_n\rangle^2$ in Eq. 3.14. We approximate the Hamiltonian in Eq. 3.12 with a continuous chain using an Inextensible Gaussian Hamiltonian (IGH) [79, 80]

$$H_I[\mathbf{r}(s)] = \frac{3}{2a^2}\int_0^N ds\Big(\frac{\dot{x}^2(s) + y^2(s)}{\alpha_1^2(k,\varphi)} + \frac{\dot{z}^2(s)}{\alpha_3^2(k,\varphi)}\Big) - \beta g(k,\varphi)\int_0^N ds\dot{z}(s) \tag{3.16}$$

where α_1 and α_3 are the effective spring constants in the longitudinal and transverse directions, respectively, and g is an effective tension. The spring constants α_1 and α_3, and the effective tension g, are functions of both k and φ. Using the IGH in eq. 3.16, we find

$$\begin{aligned}
\langle Z\rangle &= Na\alpha_3^2\beta g/3,\\
\langle Z^2\rangle - \langle Z\rangle^2 &= Na^2\alpha_3^2/3,\\
\langle X^2\rangle &= Na^2\alpha_1^2/3. && (3.17)
\end{aligned}$$

Equating the averages in Eq. 3.17 with those in Eqs. (3.13)-(3.15) explicitly gives the desired IGH in terms of k and f. The full expression for the α_i's and g are

quite lengthy for general k and f, and we omit them here. However, in the limit as $k \to \infty$, we recover the Freely Jointed Chain (FJC) averages. In the FJC limit, the expressions for the α_i's and g are quite simple, and we find

$$\begin{aligned} \alpha_1^2 &= \frac{3}{\varphi^2}\Big(\varphi \coth(\varphi) - 1\Big), \\ \alpha_3 &= \frac{3}{\varphi^2}\Big(1 - \varphi^2 \mathrm{csch}^2(\varphi)\Big), \\ \alpha\beta g &= \varphi \frac{\alpha_1^2}{\alpha_3^2}. \end{aligned} \tag{3.18}$$

These spring constants, $\alpha_i(k \to \infty, f)$, were derived by Hatfield and Quake using a different method [80], but the effective tension g has been thus far unreported.

We note that this approximate FJC Hamiltonian gives the simple Gaussian behavior for $\varphi \to 0$, whereas in the limit of $\varphi \to \infty$, we can easily show that the distributions give the expected form of $P(X) = \delta(X)$, $P(Y) = \delta(Y)$, and $P(Z) = \delta(Z - Na)$, with $X = x_N - x_0$, and $Y = y_N - y_0$. We therefore expect that the IGH to be an good approximation for an inextensible chain in the limits of small and large f, with possible deviations from the correct distribution for intermediate f. Because of the more complicated form of the Hamiltonian in Eq. 3.16, exact analytic work is difficult in the inextensible case. We can, however, generate a self-consistent equation using Eq. 3.7 to determine the FEC of a self-avoiding inextensible chain in a manner similar to the extensible case. Using the reference Hamiltonian

$$H_r = \frac{3}{2a^2\lambda^2} \int_0^N ds \Big(\frac{\dot{x}^2(s) + y^2(s)}{\alpha_1^2} + \frac{\dot{z}^2(s)}{\alpha_3^2} \Big) - \beta g \int_0^N ds\, \dot{z}(s), \tag{3.19}$$

and defining

$$\Delta_1^{(IGH)} = \frac{3}{2a^2}\Big(1 - \frac{1}{\lambda^2}\Big) \int_0^N ds \Big(\frac{\dot{x}^2(s) + y^2(s)}{\alpha_1^2} + \frac{\dot{z}^2(s)}{\alpha_3^2} \Big), \tag{3.20}$$

we can, to first order in $\Delta_1^{(IGH)} + \Delta_2$ (with Δ_2 the excluded volume interaction, eq. 3.3), develop the self-consistent equation $\langle Z(\Delta_1^{(IGH)} + \Delta_2)\rangle = \langle Z\rangle \langle \Delta_1^{(IGH)} + \Delta_2\rangle$, similar to Eq. 3.7. The final form of the inextensible SCE is similar to that of eq. 3.10, with

$$\lambda^2 - 1 = \frac{v\sqrt{N}}{\lambda^3 \alpha_1^2 \alpha_3} \int_\delta^1 du \; \frac{1-u}{\sqrt{u}} e^{-N\lambda^2 \alpha_3^2 \gamma^2 u/6}, \tag{3.21}$$

with $\gamma = a\beta g$ the dimensionless effective tension. It is possible, albeit complicated, to show that the solution to Eq. 3.21, with $k \gg 1$, will be divided into approximately the same scaling regions as we found in the extensible case (which we do not explicitly show here). The roots to the inextensible SCE, determined using Eq. 3.21, are similar to the extensible roots from Eq. 3.7, with significant differences in the two models occurring only for $a\beta f \geq 1$. As in the case of the extensible chain, the expected Pincus scaling of $\langle Z\rangle \sim f^{\frac{2}{3}}$ emerges only for very large N. Thus, both for the extensible chain and the IGH with excluded volume interactions, the linear behavior and the Pincus regime are obtained (for large N). The behavior of the FEC in the limit of very large force is clearly model dependent, as predicted by Pincus [39].

In order to determine if the extensible theory accurately predicts the effect of excluded volume on a self-avoiding polymer under tension, we have performed Langevin simulations with $N = 100$ at various stretching forces (the details are given in Appendix B.1. To compare our theory (with the self consistent equation given in eq. 3.10) to simulations, we need two fitting parameters, v_0 and δ. We determine v_0 by fitting the simulated FEC in the linear, low force regime, and

obtain δ by a global fit of the theoretical predictions of the FEC to the results of the simulation. The scaling laws for the extension as a function of force can not be accurately determined by simply fitting a linear [33, 81, 82] or log-log [83] plot of the FEC. Such fits implicitly assume that there exists a well-defined scaling regime, where $\langle Z \rangle \sim f^y$ with y constant. In order to determine the various scaling regimes of the FEC without imposing such an assumption, we will define the force dependent effective scaling exponent $y(\varphi)$ such that

$$y(\varphi) = \partial \log(\langle Z \rangle)/\partial \log(\varphi). \tag{3.22}$$

For a true scaling regime to exist, we expect $y(\varphi)$=const for a non-vanishing range of ϕ. In Figure 3.1, we show the best fit of the extensible theory (eq. 3.10) compared to the simulations for the polymer in a good solvent ($v_0 > 0$). We find $v \approx 58.6$ and $\delta = 1.6\lambda/N$ agrees well with the simulation data. We note that this gives $v_0 \approx 178a^3$, significantly larger than the hard-core second virial coefficient of $v_0 = 4\pi a^3/3$. It is known from the Edwards-Singh calculation [40] (with $f = 0$) that if higher order terms are included in deriving the self-consistent equation (Eq. 3.7), they merely renormalize v without altering the scaling behavior. A similar behavior is expected when $f \neq 0$. As a result of the renormalization of v, we find that the extracted value of v from simulations is larger than the value calculated directly from the second virial coefficient.

We see that the theoretical predictions depend very strongly on the choice of cutoff, with the $\delta = 0$ theoretical FEC showing very poor agreement with the simulated data for $a\beta f \geq 0.1$. This is somewhat surprising, as a cutoff in the continuum

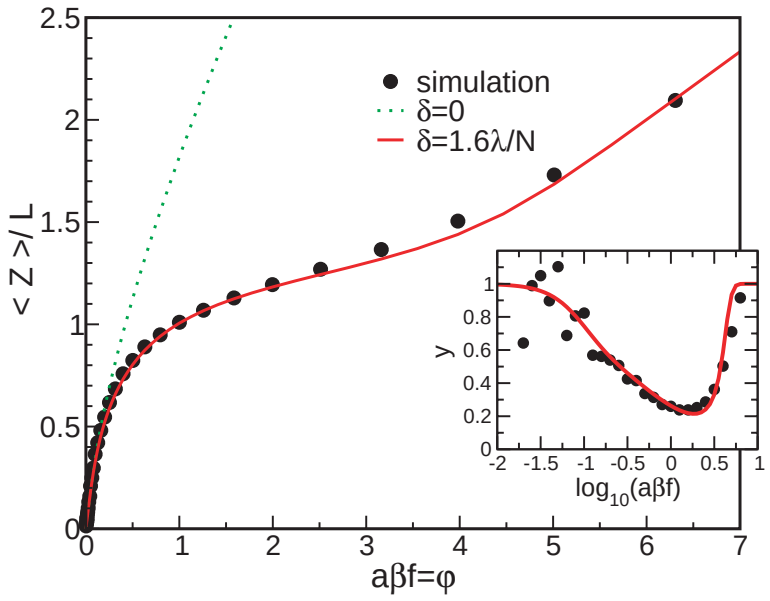

Figure 3.1: $\langle Z \rangle$ as a function of φ for varying φ. The dots are the simulation results with $N = 100$. The linear regime corresponds to the dimensionless excluded volume parameter $v \approx 58.6$. The best visual fit (solid line) is obtained with $\delta = 1.6\lambda/N$. Also shown are the fits with $\delta = 0$ (dotted line). The inset compares the theoretical predictions (solid line) and the simulations results (dots) for the effective scaling exponent, $y(\varphi)$.

limit approximation generally is used only to avoid self-energy divergences in the theory [40, 76, 77, 78], which are not present here. We also note that neither the theory nor the simulation predicts a Pincus-like scaling of $y \approx 2/3$. We will find that this is due to the fact that the notion of the unperturbed tensile blob is not applicable for $N = 100$ (see below).

In order to asses the conditions under which the Pincus regime can be observed, we plot the theoretical effective scaling exponent $y(\varphi)$ for increasing N in Fig. 3.2(a). While there is no clear Pincus regime for N=100, the expected 2/3 scaling emerges for larger N. Variation in v (i.e. changing repulsion between monomers) only effects the depth of the trough in the non-universal regime (see Fig. 3.2(a)), so adjusting v can not yield the expected Pincus scaling for smaller N. Figure 3.2 also shows that a very large $N \sim 10^6$ is required in order to see the 2/3 scaling over a large force range. For small values of N, the inequality $N \gg (\xi_P/a)^{\frac{1}{\nu}} \gg 1$ required to observe the Pincus scaling is not satisfied, and the tensile blobs are thus effectively stretched. The width Δf over which the strong stretching is observed can be computed using the self-consistent theory. If we define the Pincus regime such that $\partial y(\varphi)/\partial\varphi \leq \epsilon$ (with y defined in Eq. 3.22) for some tolerance ϵ, we can numerically determine the dependence of the width of the Pincus regime with respect to N. The width of the Pincus regime, Δf, is shown in Fig. 3.2(b) for $\epsilon = 0.05$, along with a fit $\Delta f \approx 0.018 - 1600N^{-1}$. In the inset, we show the transition force into the Pincus regime, f_T, along with the expected scaling of $N^{-\frac{3}{5}}$. We can extrapolate that the minimum number of monomers, N_{min}, for a self-avoiding polymer to show that the Pincus regime emerges only when $N_{min} \approx 9 \times 10^4$ for $\epsilon = 0.05$. Larger values of N

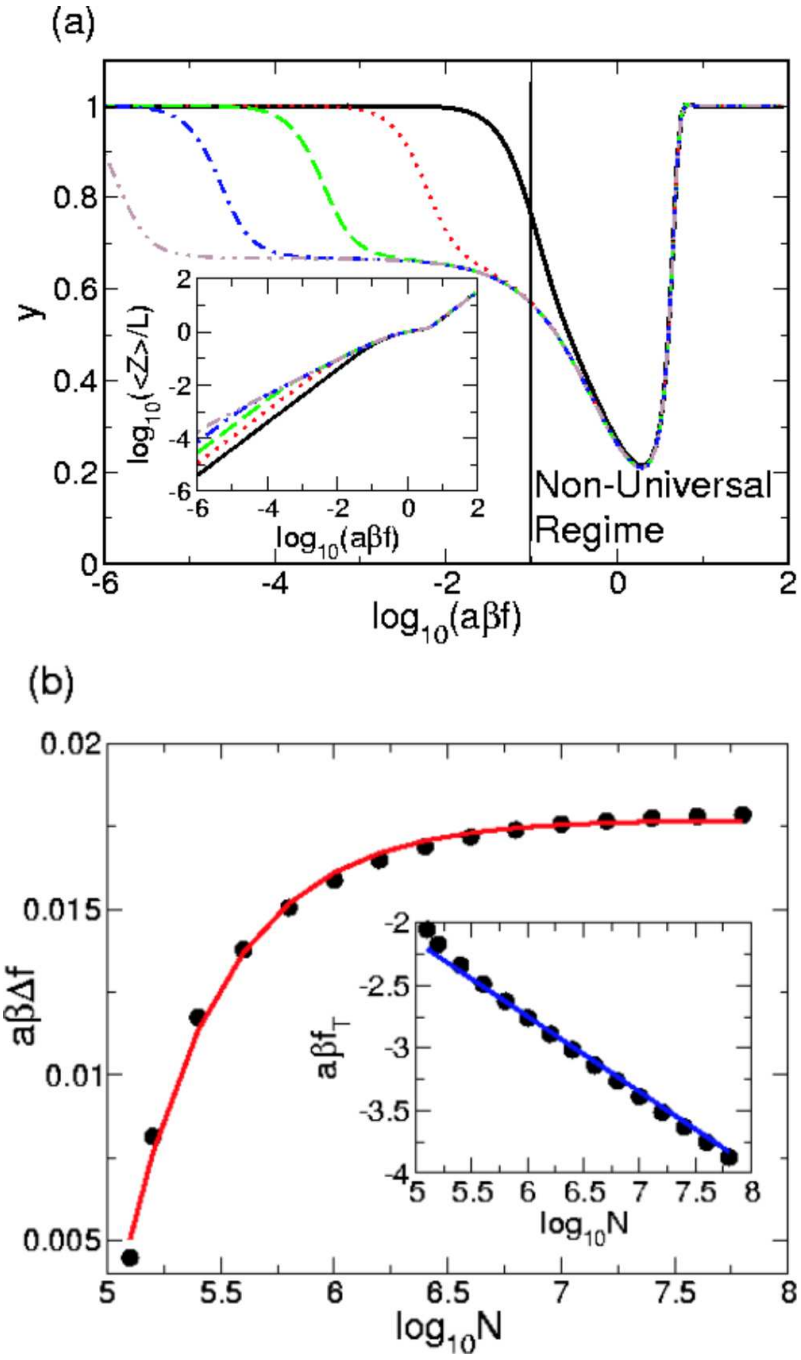

Figure 3.2: (a) The effective scaling exponent $y(\varphi)$ for $N = 10^2$ (—), $N = 10^4$ $(\cdots)$, $N = 10^6$ $(--)$, $N = 10^8$ $(- \cdot -)$, and $N = 10^{10}$ $(- \cdot \cdot -)$, all with $v = 58.6$ and $\delta = 1.6\lambda/N$ obtained theoretically. The inset shows the log-log plot of the extension vs. force, for the same parameters. (b) The width of the Pincus regime Δf as a function of N for $\epsilon = 0.05$. The inset shows the initial Pincus transition force f_T as a function of N. Also shown is the predicted $N^{-\frac{3}{5}}$ scaling.

are required for the Pincus scaling to continue over an observable interval of f. This finite size effect is remarkable, because when $f = 0$ the Flory exponent ($\nu \approx 0.6$) can be accurately obtained with $N < 100$ [84]. Because N_{min} is too large for accurate simulations, it is not possible to explicitly demonstrate the Pincus scaling of $f^{2/3} in$ *silico*. In principle, single molecule AFM or optical tweezer experiments can be used to confirm the predictions for such large N.

In order to test our inextensible theory, we determine the best fit to a Monte Carlo simulation of a thick chain [33, 81] with $N = 1600$. The thick chain is an inextensible, hard-core excluded volume model, with a configuration rejected if a triplet of monomers lie within a circle of radius a (see Appendix B.3 for details). Our variational Hamiltonian in Eq. 3.19 is generated using the spring constant $k = 10^4$ in Eq. 3.16. In Fig. 3.3, we compare the FEC and effective scaling exponent (Eq. 3.22) for the simulations and the inextensible theory, in Eq. 3.21. The theoretical predictions are in very good agreement with the FEC obtained using Monte Carlo simulations (Fig. 3.3(a)). We find $v \approx 15.7$ gives a good fit for the simulation data for low forces, and again $\delta = 1.6\lambda/N$ gives a good global fit to the simulated data.

In Fig 3.3(b), we see that there is a deflection in $y(\varphi) \approx 2/3$ at $\varphi \approx 0.1$ (denoted Pincus-like), corresponding to a Pincus-like regime observed in the simulations, and accurately predicted by the theory based on Eq. 3.21. Such a deflection near $y = 2/3$ is predicted by the theory for both the extensible and inextensible Hamiltonians, and can be clearly seen in Fig. 3.2(a) for $N = 10^4$. This deflection shows that the Pincus regime is beginning to emerge, but the width of the regime

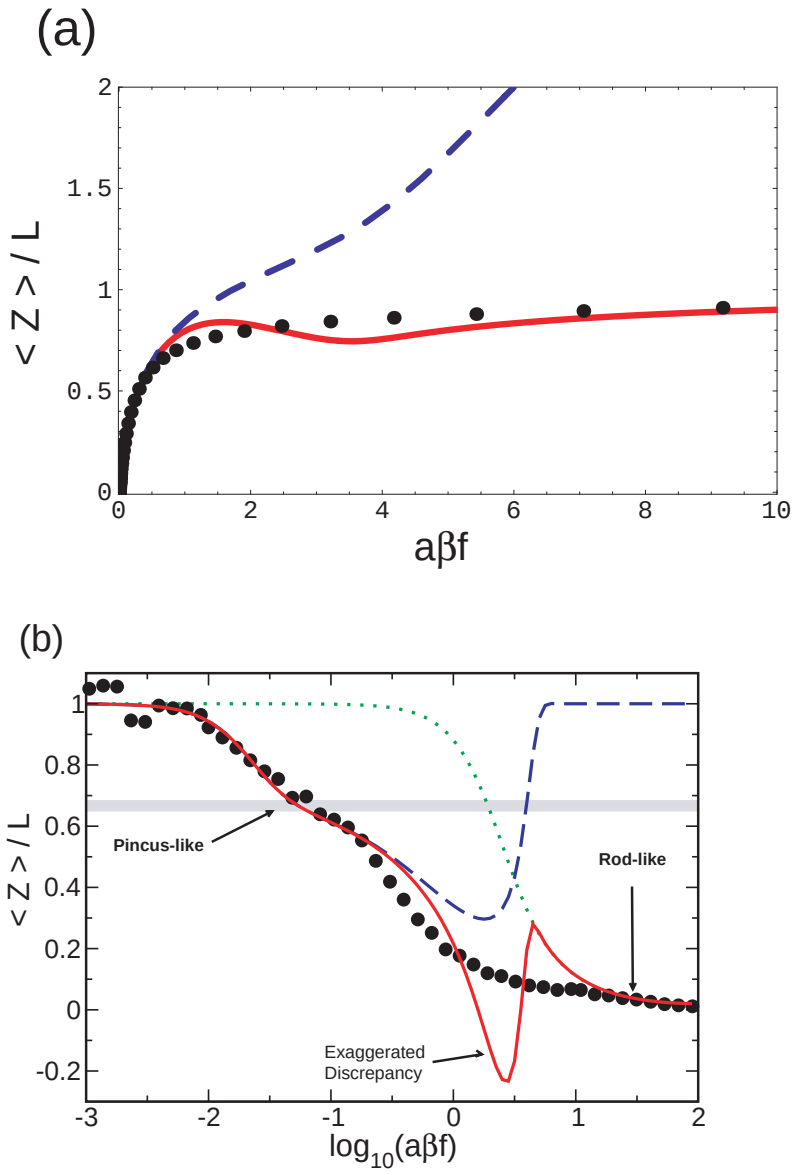

Figure 3.3: (a) Force extension curve for an inextensible chain with $N = 1600$. Shown are the simulation data (dots), along with the best fit for the IGH (solid line) and extensible Hamiltonian (dashed line), with $v = 15.7$ and $\delta = 1.6\lambda/N$. (b): Effective scaling exponent $y(\varphi)$ for the inextensible FJC. Shown are the theoretical exponents for the IGH (solid) and extensible Hamiltonian (dashed). Also shown is the non-interacting FJC exponent (dotted line).

Δf is vanishingly small. We also see the expected return to the non-interacting FJC behavior for large f. The fit is, however, quite poor for $a\beta f \approx 1-4$, where the effective scaling exponent differs greatly from the simulation data. Figure 3.3(a) shows that the poor fit for intermediate f originates with a slight overestimation in $\langle Z \rangle$ vs. f near $a\beta f \approx 1$, followed by an underestimation in $\langle Z \rangle$ near $a\beta f \approx 3$. This over- and underestimation produces a FEC that is not monotonically increasing with f, a completely non-physical result. The small differences between the theoretical and simulated FECs are greatly exaggerated by the effective scaling exponent in the intermediate force range.

The reason for the discrepancy between theory and simulation for intermediate forces is that, in the approximate representation for the (nearly) inextensible chain, extensions from $|\Delta \mathbf{r}_n| = a$ are allowed (see Appendix B.3). As a result, the chain can stretch somewhat, with an average monomer spacing exceeding a. For this reason, less force is required to extend the chain at intermediate forces, producing an overestimate of the FEC. The consequences of the minor disagreement between theory and simulations are exaggerated when the effective exponent $y(\varphi) = \partial \log(\langle Z \rangle)/\partial \log(\varphi)$ is computed (Fig 3.3(b)). We see, however, that both the extensible and inextensible polymer models in a good solvent accurately predict the Pincus-like regime observed for $a\beta f \sim 10^{-2} - 10^{-1}$. At high forces, the response to the force depends on the precise model used to account for chain connectivity. As a result, the predictions for the extensible and inextensible polymer models are vastly different when $a\beta f > 1$.

3.3 Reexamination of the Blob Concept for Finite N

In order to better understand the unexpected scaling behavior of the FEC's for finite N, a more detailed study of the physical processes of extension are required. There are three mechanisms by which the average extension of an extensible chain can increase as a function of force. The first is orientation of the polymer along the force axis. We expect that, for small f, the force will cause alignment with the z axis, with little perturbation of the chain conformation. In the second mechanism, the extension of the polymer is determined by an interplay between ξ_P (a length scale below which f is not relevant), and N (which effectively determines the number of aligned blobs along the force direction). We expect this mechanism will occur for intermediate forces, and for sufficiently large N, cause the emergence of the Pincus regime. As these blobs are stretched, $\langle Z \rangle$ will increase without significantly affecting the alignment along the f axis. For large forces we expect overextension to dominate, when the chain is fully aligned and the monomers on a length scale ξ_P are stretched. In order to see these physical mechanisms of the extension in the simulations of finite, extensible polymers with $N = 100$, we compute the effective force-induced alignment exponent ω, given by $\langle Z/|\mathbf{R}| \rangle \sim f^{\omega}$, and the effective overextension exponent μ, given by $\langle Z/L \rangle \sim f^{\mu}$. If the polymer is perfectly aligned along the z-axis, we expect that the exponent $\omega \to 0$ (i.e. $\langle Z/|\mathbf{R}| \rangle$=const). Likewise, if the extension is dominated by the overextension of the chain, we expect $\mu \to 1$ ($i.e. \langle Z \rangle \sim L(f) \propto f$)The variations of the effective exponents ω and μ for $N = 100$ as f changes are shown in Fig. 3.4(a). We see that the polymer aligns with the z

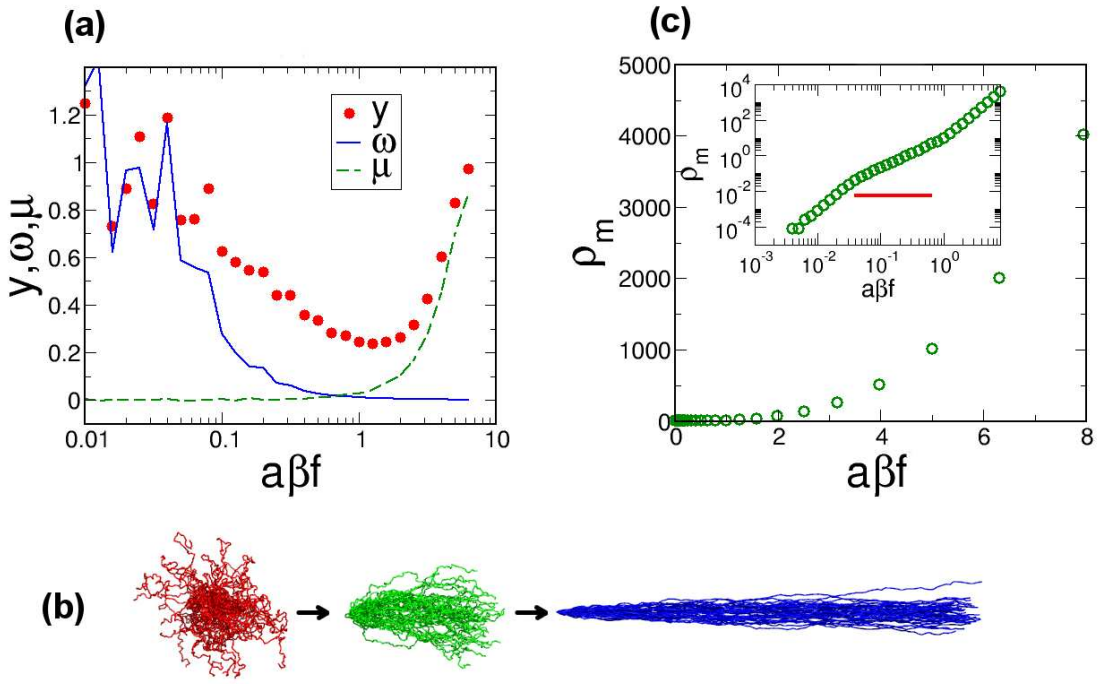

Figure 3.4: (a) Contributions to the effective scaling exponent $y(\varphi)$. Shown are $y(\varphi)$ (dots), as well as the alignment exponent ω (solid line) and overstretching exponent μ (dashed line). (b) The ensembles of structures at $a\beta f = 0$ (red), 0.5 (green), and 8.0 (blue) are given to demonstrate the three step mechanism of the extensible chain stretching, i.e., (i) alignment, (ii) disruption of tensile blob, and (iii) overstretching. (c) The blob density as a function of force. The inset shows a log-log plot of the monomer density inside the blob, showing three distinct scaling regimes. Scaling relation ρ_m vs f is obtained by fitting the data above red line in the inset.

axis at relatively small forces, with full alignment ($\omega \to 0$) occurring for $\varphi \approx 0.1$. Overextension does not begin until $\varphi \approx 3$ (in the non-universal regime, see Fig 3.4(a)), giving a wide range of forces in which stretching of the monomers inside of the blobs contributes to the behavior of $\langle Z \rangle$. Representative snapshots of the chain configuration in the three regimes are shown in Fig. 4(b).

The absence of a clear signature of the Pincus regime, even for $N = 1600$, is intimately related to the breakdown of the inequality $N \gg (\xi_P/a)^{\frac{1}{\nu}} \gg 1$. For large

enough N, when the Pincus regime in the FEC is observed (Fig 3.2(a)), the size of the blob $\xi_P \approx k_B T/f$ is expected to scale as $\xi_P \approx a N_b^{\nu}$, where N_b (presumed to be much greater than unity) is the number of monomers inside of the blob. The monomer density, ρ_m, inside the blob will scale as

$$\rho_m \sim N_b/\xi_p^3 \sim \xi_p^{1/\nu-3} \sim (f/k_B T)^{3-1/\nu}. \tag{3.23}$$

In good solvents, $\nu = 3/5$, and hence ρ_m is given by $\rho_m \sim f^{4/3} \equiv f^w$. If the effective value for w with finite N exceeds $w = 4/3$, as could be the case when the force locally stretches the chain segments inside ξ_P, we will find $\langle Z \rangle \sim f^y$ with $y \neq 2/3$ in the intermediate force regime.

In order to provide insights into the effective blob response to f for the self-avoiding extensible polymer of the finite size (N=100), we have calculated the dependence of the monomer density inside the blob on f. To obtain the scaling behavior between monomer density and the force from the simulations, we perform the following steps:

1. Make a sphere of radius $b = \xi_P/2$, with $\xi_P (= k_B T/f)$, centered on the i^{th} monomer and count the number of monomers (N_b) within the sphere whose volume is b^3. The density of monomers within the sphere center at i^{th} monomer is $\rho_m(i) \sim N_b/b^3$.

2. Move to the $(i+1)^{th}$ monomer, and compute the density again.

3. When $i = N$, the average density is computed using $\langle \rho_m(f) \rangle = 1/N \sum_{i=1}^{N} \rho_m(i)$.

4. Repeat this procedure for the ensemble of structures obtained at each force.

Although this method of computing the monomer density from the polymer structures is very crude, the scaling exponent between ρ_m and f should not be affected by the details of the calculation. The results are shown in Fig. 3.4(c). We find that $\rho_m \sim f^{1.6}$ in the intermediate force regime (data above the red base line in the inset of Fig. 3.4(c)) . From Eq. 3.23, a density scaling of $f^{1.6}$ implies $\xi_p \sim N_b^{0.71} \neq N_b^{0.60}$, which indicates that there is no force range in which ideal blobs can be observed for small N. The observed scaling exponent for N_b is greater than that for a simple self-avoiding walk, which suggests that the monomers inside of the blob do not behave as unperturbed SAW's. Thus, the fundamental premise used in the blob argument used to derive the Pincus regime breaks down for small N. The tensile force is felt by the monomers within the blobs, which swell due to the stretching of monomers inside ξ_P. The density of monomers inside the blob scales differently than is expected for large values of N, and provides the microscopic reason why $\langle Z \rangle \sim f^x$ with $x < 2/3$ for a finite-sized self-avoiding chain. As N increases, the intermediate force regime can be large enough so that $\rho_m \sim f^{\frac{4}{3}}$, which is needed to see the Pincus scaling $\langle Z \rangle \sim f^{\frac{2}{3}}$.

3.4 Homopolymer in a Poor Solvent

In a poor solvent, the second virial coefficient (v_0) becomes negative, with the strength of the attractive interactions between monomers exceeding those between the monomers and the solvent. As a result, the polymer adopts collapsed, globular

conformations at temperatures below the Flory Θ temperature. In poor solvents, the Edwards model is modified to include an effective three-body interaction, to ensure that averages of physical observables are convergent. The extensible Hamiltonian in a poor solvent is $\beta H_P = \beta H_0 + \Delta_3$, where H_0 is defined in Eq. 3.2 and

$$\Delta_3 = \frac{w_3}{6}\int_0^N ds \int_0^N ds' \int_0^N ds''\ \delta[\mathbf{r}(s) - \mathbf{r}(s')]\ \delta[\mathbf{r}(s') - \mathbf{r}(s'')]. \tag{3.24}$$

The self-consistent equation for the extension in this case becomes $\langle Z(\Delta_1 + \Delta_2 + \Delta_3)\rangle = \langle Z\rangle\langle \Delta_1 + \Delta_2 + \Delta_3\rangle$, similar to Eq. 3.7. We have already determined the Δ_1 and Δ_2 terms, and need only compute $\langle Z\Delta_3\rangle - \langle Z\rangle\langle\Delta_3\rangle = a\partial/\partial\varphi\langle\Delta_3\rangle$. The SCE for an extensible polymer in a poor solvent can then be written

$$\begin{aligned}\lambda^2 - 1 \;=\; & \frac{v\sqrt{N}}{\lambda^3}\int_\delta^1 du\ \frac{1-u}{\sqrt{u}} e^{-N\lambda^2\varphi^2 u/6} \\ & + \frac{w}{\lambda^6}\int_\delta^1 du_1 \int_\delta^{1-u_1} du_2 \frac{(1-u_1-u_2)(u_1+u_2)}{u_1^{3/2}u_2^{3/2}} e^{-N\lambda^2\varphi^2(u_1+u_2)/6},\end{aligned} \tag{3.25}$$

where we have defined the dimensionless $w = (3/2\pi)^3\ w_3/a^6$. The inextensible self-consistent equation is similar, with a similar root structure, and we will therefore omit such a calculation here. Again, we have included a cutoff δ in the integrals, as was done for the two-body case. However, the three-body integral in Eq. 3.25 is clearly divergent for $\delta = 0$ (scaling as $\delta^{-1/2}$), unlike the two-body term. This divergence must be removed for the self-consistent equation to converge in the limit of $N \to \infty$, by renormalizing w. For $f = 0$, we can evaluate the three-body integral exactly, and find that, with $\delta \sim \lambda/N$ and as $N \to \infty$, it diverges as $\sim 16/3\sqrt{\delta}$. The $N \to \infty$ divergence is therefore removed if we renormalize $w = \bar{w}/\sqrt{N}$. It can be shown that the self-consistent equation has a globular solution

$\lambda \approx \lambda_g = (4\bar{w}/|v|)^{\frac{1}{3}} N^{-\frac{1}{6}}$ for $f = 0$ and large N, giving the expected $\langle \mathbf{R}^2 \rangle \sim N^{\frac{2}{3}}$ for a homopolymer in a poor solvent. However, the final term of Eq. 3.25 can not be evaluated exactly for non-zero φ, so we must resort to numerical work in order to determine the roots for larger forces.

We find that Eq. 3.25 has three unique roots beyond a critical force f_c, which correspond to collapsed (λ_c), extended (λ_E), and intermediate (λ_b) structures. Numerically, we find $0 < \lambda_c \leq \lambda_g$, and $\lambda_E \approx 1$ for $f > f_c$. Our interpretation of λ_c as corresponding to a collapsed state is only qualitative, because an extensible homopolymer (used as the reference Hamiltonian in the calculations) in a poor solvent does not have a unique 'collapsed' state. With the interpretation of λ_g and λ_E as the roots signifying the two local minima of the free energy for the collapsed and extended states, we can interpret the intermediate root λ_b as a local maximum of the free energy, i.e. the barrier (or saddle point) between the two states. Again, this interpretation is qualitative only, because there is no well defined 'barrier' between the collapsed and extended states. In Fig. 3.5(a) we show the extension $\langle Z \rangle / L = \varphi \lambda^2 / 3$ for the three solutions to the self-consistent equation (3.25) for $v = -5$ and $\bar{w} = 1$ (arbitrarily chosen), and with $\delta = 1.6\lambda/N$. We see $a\beta f_c \approx 3.5$ is the critical force at which the extended and intermediate roots emerge. We expect f_c to depend on the particular values of v and $\bar{w}$, and expect it will be an increasing function of $|v|/\bar{w}$. In this triple-root regime, the polymer will be in bistable equilibrium between the collapsed ensemble and extended state, suggesting the development of a pearl-necklace structure for intermediate f. The collapsed and intermediate solutions coalesce for a finite $f = f_E$ (Fig 3.5(a) inset). For $f > f_E$, λ_c and λ_b vanish,

leaving the extended root λ_E the only solution to Eq. 3.25. This shows, as expected, that the inter-monomer interactions become irrelevant for sufficiently high force, and $\langle Z \rangle \sim Na \times (a\beta f)/3$ as $f \to \infty$. A schematic picture of the free energy as a function of the extension $\langle Z \rangle$ is shown in Fig. 3.5(b-e), for varying force, showing our qualitative interpretation of the solutions to the self-consistent equation (3.25).

A similar multi-root structure has been previously predicted for a polymer in a poor solvent with electrostatic interactions [78, 85, 86]. These references note the emergence of multiple roots beyond a critical value of the backbone charge density (in this respect, equivalent to the tension), and qualitatively identify the meaning of the multiple roots as we have. However, because the Edwards Singh method can not predict the barrier height or the depth of the minima, we can not quantitatively predict $\langle Z \rangle$ for a polymer in a poor solvent. This qualitative picture, namely the tension-induced globule to rod transition which should occur when $f > f_c$, is confirmed using explicit simulations of force-induced stretching of a homopolymer in a poor solvent. The simulations (see below) also provide a microscopic picture of the structural transitions that occur as w_3, appearing in Eq. 3.25, increases.

In Figure 3.6, we show the average linear extension for weakly hydrophobic (a) and strongly hydrophobic (b) polymers as a function of force. The weakly hydrophobic polymer does show a transition between two linear scaling regimes, with the low force behavior of $\langle Z \rangle \approx \langle \mathbf{R}^2 \rangle_{f=0} \times (\beta f)/3$, and the high force behavior returning to the non-interacting $\langle Z \rangle = Na^2 \beta f/3$. The transition is very smooth, and does not show the expected first order transition due to the weak nature of the

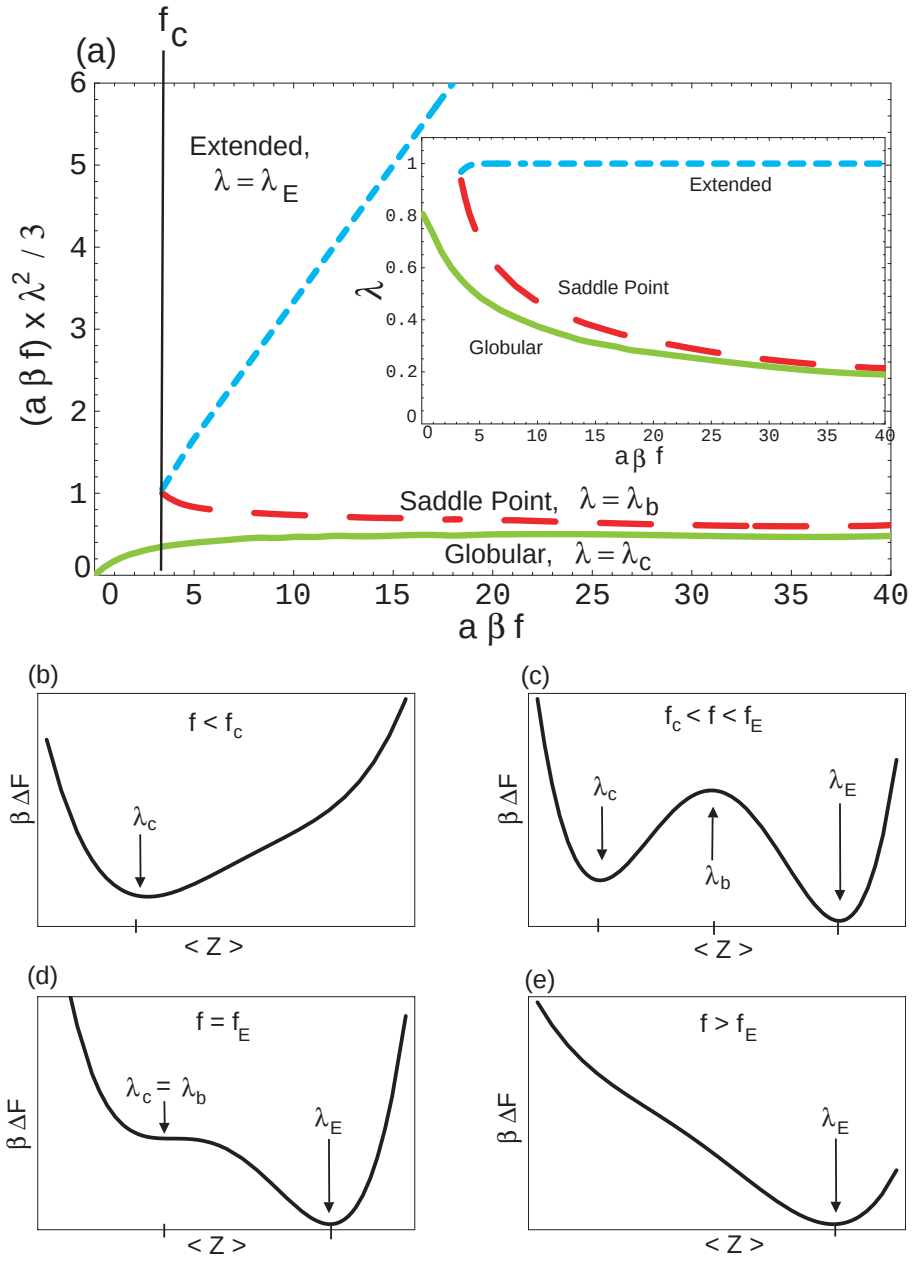

Figure 3.5: (a): The extension $\langle Z\rangle/L = (a\beta f)\lambda^2/3$ for the three roots of the self-consistent equation in a poor solvent for $v = -5$ and $\bar{w} = 1$, with $\delta = 1.6\lambda/N$: λ_g (solid line), λ_b (dashed line), and λ_E (dotted line). The three values for λ are shown in the inset. (b) For $f < f_c$, the polymer is globular. (c) In the force range $f_c < f < f_E$, the chain conformations are a combination of globular and extended states. (d) at $f = f_E$, the globular configuration is marginally stable. (e) For $f > f_E$, the chain is in the fully extended state.

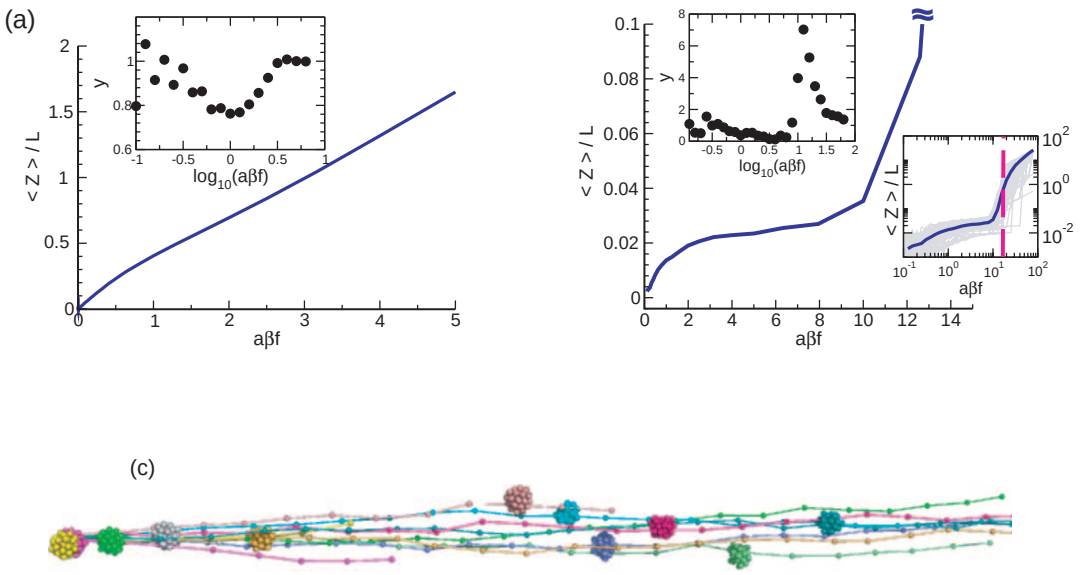

Figure 3.6: (a) Extension as a function of force for a weakly hydrophobic polymer (ε in Eq. $B.2$ is 0.5). (b) Same as (a), except the chain is strongly hydrophobic ($\varepsilon = 1.5$). The insets show the effective scaling exponent $y(\varphi)$ (Eq. 3.22). The transition to the extended state in (a) appears continuous. For the strongly hydrophobic polymer, the globule $\rightarrow$ rod like transition is sharp. The transition force depends on the energetic details of the globule. The heterogeneity of the transition is manifested as the broad variations of transition force. The ensemble of structures found at the globule-to-rod transition force ($f_c = 1.8k_BT/a$) are shown in (c).

interactions, as shown in the inset. The strongly hydrophobic chain does show a first order transition around $a\beta f_c \approx 1.8$, but with broad dispersion. Variations in the critical unbinding force are substantial from molecule to molecule, due to the microscopic heterogeneity of the globular structures. The observed plateau in Fig 3.6(b) is due to full alignment of the globule along the z-axis, as was the case for the self-avoiding polymer (Fig 3.4(a)), and seen in the theoretical predictions (Fig 3.5). There is a large range of forces over which the FEC does not resemble either the globular or fully extended states, showing the bistable equilibrium between the two.

3.5 The Scattering Function Under Force

The analysis using scattering experiments is extremely useful for investigating the overall polymer configurations, because the scattering intensity as a function of transfer momentum ($I(q) = \langle 1/N^2 \sum_{i<j} \exp(i\mathbf{q} \cdot \mathbf{r}_{ij}) \rangle$) provides structural information on all length scales, while the FEC only provides information about the extension of the chain. By comparing with the well-known scaling relations of $I(q)$ with respect to q, one can envision the configurations of the polymer over all length scales. For example, for the various structures we expect $I(q) \sim q^{-x}$, with $x = 2$ (Gaussian chain), $x = 1$ (rod), $x = 4$ (globules), and $x = 5/3$ (polymer in a good solvent) [87]. We calculated the scattering intensity by integrating the distance

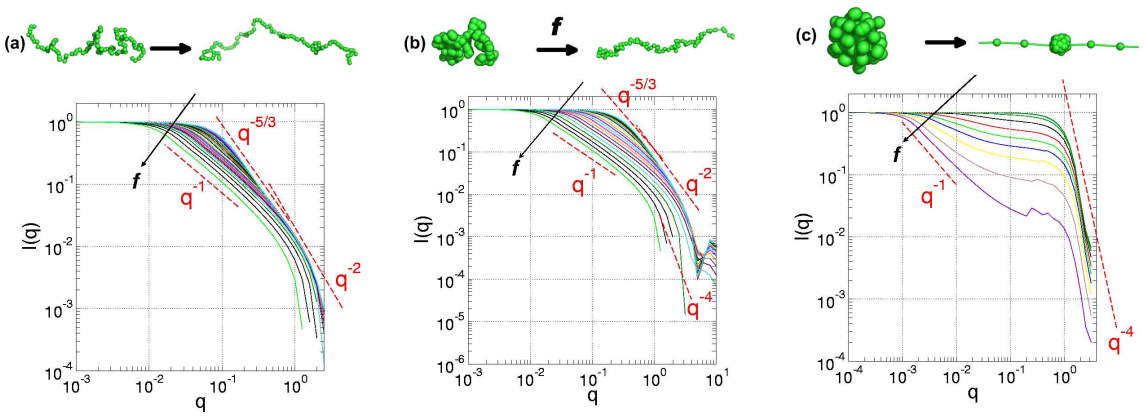

Figure 3.7: (a) $I(q)$ for a homopolymer in good solvent under varying tension, (b) $I(q)$ for weakly hydrophobic homopolymer under tension, (c) $I(q)$ for strongly hydrophobic homopolymer under tension. The arrows in (a), (b), and (c) indicate increasing f values. The tension-induced structural changes of a homopolymer are illustrated in three solvent conditions (good, near theta, and poor solvent conditions)

distribution function obtained from the ensemble of structures,

$$I(q,f) = \int d^3\mathbf{r} P(\mathbf{r},f) e^{i\mathbf{q}\cdot\mathbf{r}} = 4\pi \int_0^\infty dr\ r^2 P(r,f)\ \frac{\sin qr}{qr}, \tag{3.26}$$

with $q = |\mathbf{q}|$. In our simulations performed under varying tension values, we obtained $4\pi r^2 P(r,f)$ directly from the ensemble of structures by collecting the histograms between the interval of $(r, r+dr)$ with $dr = 0.2a$.

An inspection of the scattering intensity $I(q)$ of a homopolymer in different solvent conditions, shown in Fig. 3.7, along with snapshots of representative structures, succinctly summarizes the series of dynamic processes during the tension-induced structural transitions. (i) In good solvents (Fig 3.7(a)), the entire chain of N=100 is characterized by the tensile blob under a small force ($I(q) \sim q^{-5/3}$ for $q \sim (0.1-1)$). As f increases, the tensile blobs continuously change to the rod state, which is indicated by $I(q) \sim q^{-1}$. (ii) For the weakly hydrophobic condition (with $\epsilon_{LJ} = 0.5$,

see Fig. 3.7(b)), the chain has hierarchy of structures in terms of distinct length scales. When f is small, both signatures of Gaussian ($I(q) \sim q^{-2}$) and globule structure ($I(q) \sim q^{-4}$) are found on small length scales $q^{-1} \leq 1$, while the chain is characterized by the polymer in a good solvent for $q^{-1} \geq 1$. As f increases, the globule to rod transition of the self-avoiding chain takes place continuously. (iii) For the strongly hydrophobic condition (Fig (c)), the whole chain is collapsed to compact globule ($I(q) \sim q^{-4}$. The globular structure is maintained so that all $I(q)$'s are practically identical for $q^{-1} \geq 1$ as long as $f < f_c$. When f becomes greater than f_c, a sharp transition occurs, reflecting the globule ($I(q) \sim q^{-4}$) to the rod ($I(q) \sim q^{-1}$) transition. The first order nature of force-induced stretching has been previously described using scaling arguments [88].

3.6 Conclusions

We have developed a general theory for describing the response of homopolymers to an external force for arbitrary values of N, the number of monomers. By using both an extensible and inextensible model for the polymer in a good solvent, we show that the theoretical results are in accord with the predictions of the Pincus scaling laws for large N. The mean chain extension depends linearly on the force at small f, and scales as $\langle Z \rangle \sim f^{\frac{2}{3}}$ for intermediate f and sufficiently large N. Simulations of an extensible chain with $N = 100$ and the thick chain model with $N = 1600$ were performed to validate the theory. The theoretical predictions for

the force-extension curves are in excellent agreement with the simulation results. Surprisingly, the expected Pincus scaling is not observed in simulations, even for $N = 1600$. The theory predicts that the width Δf for observing the Pincus regime for $N \sim O(10^3)$ is vanishingly small. Only when N exceeds $\sim 10^5$ can the strong stretching limit ($\langle Z \rangle \sim f^{\frac{2}{3}}$) be unambiguously observed. The failure to observe the Pincus scaling is linked to the breakdown of the notion that the monomers inside the well-defined tensile blobs are unperturbed. For $N \sim O(10^3)$, the monomers inside each blob feel the effect of force, which essentially violates the required inequality $N \gg (\xi_P/a)^{\frac{1}{\nu}} \gg 1$. Other scaling theories that depend on the unperturbed tensile blob concept, such as the statistics and dynamics of a cylindrically confined self-avoiding chain [77, 89, 90], may also show a similar finite size effect.

Applying tension to a polymer in a poor solvent produces a much richer set of structures, because of the attractive monomer-monomer interactions that, with $f = 0$, leads to globule formation provided $T < T_\Theta$. For this case, the theoretical analysis predicts that the globule to stretched (i.e. rod-like conformation) transition should occur abruptly via a first-order transition when f exceeds a critical force. While the simulation results are in accord with the theoretical predictions, they show several structural transformations, depending on the quality of the solvent. The hierarchy of structures are reflected in the force-dependent structure factor. For weakly hydrophobic polymers ($T \approx T_\Theta^+$) at small forces, the scattering function $I(q)$ shows signatures of Gaussian and globular structures at large length scales (small q), whereas at small length scales the polymer behaves as a self-avoiding chain. At large forces, the transition to a rod-like conformation occurs. These

structural transitions occur continuously as f increases for a weakly hydrophobic chain. Strongly hydrophobic chains ($T < T_\Theta$) adopt globular structures at small forces. The conformation remains globular as long as $f < f_c \approx k_B T_\Theta / R_g$. The globular nature of the conformation is reflected in the $I(q) \sim q^{-4}$ scaling. If $f > f_c$, there is an abrupt transition to the rod-like state, which is reflected in the $I(q) \sim q^{-1}$ scaling.

The predictions made here can be, in principle, validated with single molecule AFM or optical tweezers experiments. Our simulations show that the forces required to stretch the homopolymer ($N \approx 100$) is on the order of about 30 pN, which are easily accessible in current experiments.

Chapter 4

The Effect of Linkers on the Equilibrium and Kinetic Properties of RNA Hairpins

4.1 Introduction

A greater understanding of the process of protein or RNA folding, needed to understand many biologically relevant systems [91, 92] and diseases [93], can be attained by the study of the underlying energy landscape of the system [94, 95, 19]. It is possible to construct the shape of the energy landscape, including the energy scales of ruggedness [96, 97], using dynamical trajectories that are generated by applying a constant force (f) to the ends of proteins and RNA [98, 20, 99, 23]. For an experiment spanning a time such that the system samples the conformational space ergodically, the distribution $P(X)$ should converge to the equilibrium distribution function $P_{eq}(X)$. Using this strategy, laser optical tweezer (LOT) experiments have been used to obtain the sequence-dependent folding landscape of a number of RNA and DNA hairpins [100, 20, 101, 23], using $X = R_m$, the end-to-end distance of the hairpin that is conjugate to f, as a natural reaction coordinate.

In LOT experiments, the hairpin is held between two long handles (DNA [23] or DNA/RNA hybrids [100]), whose ends are attached to polystyrene beads (Fig. 4.1a). The dynamics of the RNA extension ($z_m = z_{3'} - z_{5'} \approx R_m$, provided transverse

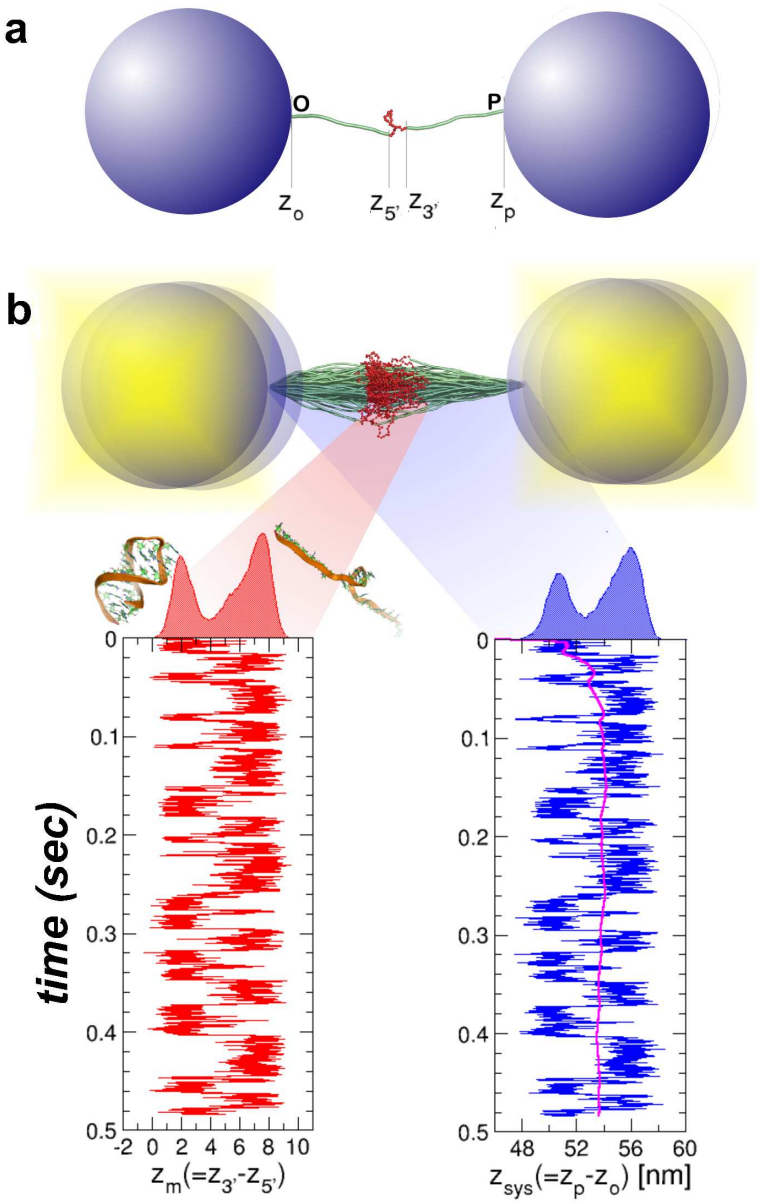

Figure 4.1: **a**. A schematic diagram of the optical tweezers setup used to measure the hairpin's folding landscape. Two RNA/DNA hybrid linkers are attached to the 5' and 3' ends of the RNA hairpin, and a constant force is applied to one end through the bead. **b**. The ensemble of sampled conformations of the H-RNA-H system during the hopping transitions obtained using L=25 nm and l_p = 70 nm. The illustration is created using the simulated structures collected every 0.5 ms. An example of the time trace of each component of the system, at f = 15.4 pN, is given L for both linkers is 25 nm. $z_m(= z_{5'} - z_{3'})$ and $z_{sys}(= z_p - z_o)$ measure the extension dynamics of the RNA hairpin and of the entire system respectively. The time averaged value $\overline{z}_T(t) = \frac{1}{t}\int_0^t d\tau z_{sys}(\tau)$ for the time trace of z_{sys} is shown as the bold line, displaying the ergodicity of the system. The histograms of the extension are shown on top of each column.

fluctuations are small) in the presence of f can not be observed directly in an experiment, and is indirectly obtained in a LOT experiment by monitoring the distance between the attached polystyrene beads ($z_{sys} = z_p - z_o$), one of which is optically trapped at the center of the laser focus (Fig 4.1a). The goal of these experiments is to extract the intrinsic folding landscape ($\beta F^o_{eq}(z_m) \equiv -\log[P_{eq}(z_m)] \approx -\log[P_{eq}(R_m)]$) and the intrinsic dynamics of the hairpin (intrinsic meaning in the absence of handles), using the f-dependent trajectories $z_{sys}(t)$. To achieve these goals in a LOT experiment, the fluctuations in the handles should minimally perturb the dynamics of the hairpin in order to probe the true dynamics of a molecule of interest. However, depending on the handle length, L, and the handle stiffness, l_p, the intrinsic fluctuations of the handles can not only disort the signal from the hairpin, but also directly affect its dynamics. The first is a problem that pertains to the measurement process, while the second is a problem of the coupling between the instruments and the dynamics of RNA.

Here, we use coarse-grained molecular simulations of a RNA hairpin to show that in order to obtain accurate $\beta F_{eq}(R_m)$ under a constant external tension, the linkers used in the LOT must be stiff, i.e., the value of L/l_p has to be small. Surprisingly, the force-dependent folding and unfolding rates that are directly measured using the time traces, $z_m(t)$, are closer to the intrinsic values (the rates as measured without handles attached) for larger L/l_p, i.e. more flexible handles. Most importantly, accurate estimates of the f-dependent hopping rates over a wide range of f-values, in the absence of handles, can be made using $\beta F_{eq}(R)$, in the presence of handles obtained at $f = f_m$, the transition midpoint at which the native basin

of attraction (NBA) and the unfolded basin of attraction (UBA) of the RNA are equally populated. The essential physics of the system is captured using a generalized Rouse model (GRM), in which there is a favorable interaction between the two non-covalently linked ends. The GRM gives quantitative agreement with the simulation results. We provide a framework for using the measured folding landscape of nucleic acid hairpins at $f \approx f_m$ to obtain f-dependent folding and unfolding times and the transition state movements as f is varied [102, 103, 104, 105, 106, 107].

4.2 Modeling the LOT experiments

In order to extract the folding landscape from LOT experiments, the time scales associated with the dynamics of the beads, handles, and the hairpin have to be well-separated [108, 109, 110]. The bead fluctuations are described by the overdamped Langevin equation $\gamma dx_p/dt = -kx_p + F(t)$ where k is the spring constant associated with the restoring force, and the random white-noise force $F(t)$ satisfies $\langle F(t)\rangle = 0$ and $\langle F(t)F(t')\rangle = 2\gamma k_B T\delta(t-t')$. The bead relaxes to its equilibrium position on a time scale $\tau_r = \gamma/k$. In terms of the trap stiffness, k_p, and the stiffness k_m associated with the Handle-RNA-Handle (H-RNA-H; see Fig. 1), $k = k_p + k_m$. With $\gamma = 6\pi\eta a$, $a = 1\mu$m, $\eta \approx 1$cP, $k_p \approx 0.01$ pN/nm [111], and $k_m \approx 0.1$ pN/nm, we find $\tau_r \leq 1$ ms. In LOT experiments [108, 110], separation in time scales is satisfied such that $\tau_U^o \approx \tau_F^o \gg \tau_r$ at $f \approx f_m$, where τ_U^o and τ_F^o are the intrinsic values of the RNA (un)folding times in the absence of handles.

Since z_m is a natural reaction coordinate in force experiments, the dispersion of the bead position may affect the measurement of $F_{eq}(z_m)$. At equilibrium, the fluctuations in the bead positions satisfy $\delta x_{eq}^2 \sim k_BT/(k_p + k_m) \sim k_BT/k_m$, and hence k_m should be large enough to minimize the dispersion of the bead position. The force fluctuation, $\delta f_{eq}^2 \sim k_BTk_p^2/(k_p + k_m)$, is negligible in the LOT because $k_p \ll k_m$, and as a result $\delta f_{eq}/f_m \sim 0$, since $\delta f_{eq} \approx 0.1$ pN while $f_m \sim 15$ pN. Thus, we model the LOT setup by assuming that the force and position fluctuations due to the bead are small, and exclusively focus on the effect of handle dynamics on the folding landscape and hopping kinetics of RNA (Figs. 4.1a-b).

To study the effects of handles on hairpin formation, we used the self-organized polymer (SOP) model of the P5GA hairpin [72], and applied a force $f = f_m \approx 15.4$ pN (see Appendix C.1 for details of the simulation). The force is exerted on the end of the handle attached to the 3' end of the RNA (point P in Fig.4.1a), while fixing the other end (point O in Fig.4.1a). Simulations of P5GA with L=25 nm and $l_p = 70$ nm show that the extension of the entire system ($z_{sys} = z_p - z_0$) fluctuates between two limits centered around $z_{sys} \approx 50$ nm and $z_{sys} \approx 56$ nm (Fig. 4.1b). The time-dependent transitions in z_{sys} between 50 nm and 56 nm correspond to the hopping of the RNA between the NBA and the UBA. Decomposing the system's extension as $z_{sys} = z_H^{5'} + z_m + z_H^{3'}$, where $z_H^{5'}(= z_{5'} - z_o)$ and $z_H^{3'}(= z_p - z_{3'})$ are the extensions of the handles parallel to the force direction (Fig. 4.1a), shows that $z_{sys}(t)$ accurately reflects the transitions in $z_m(t)$ (Fig. 4.1b). Because the simulation time is long enough for the harpin to ergodically explore the conformations between the NBA

and UBA, the histograms collected from the time traces amount to the equilibrium distributions $P_{eq}(X)$ where $X = z_{sys}$ or z_m (Fig 4.1B. To establish that the time traces are ergodic, we show that $\overline{z}_T(t) = \frac{1}{t}\int_0^t d\tau z_{sys}(\tau)$ reaches the thermodynamic average ($\approx \int_{-\infty}^{\infty} z_{sys} P_{eq}(z_{sys}) dz_{sys}$=53.7 nm) after $t \geq 0.1$ sec (the magenta line on $z_{sys}(t)$ in Fig. 4.1b).

Fig. 4.1b shows that the handles fluctuate along the force axis, even in the presence of tension, which results in slight differences between $P_{eq}(z_{sys})$ and $P_{eq}(z_m)$. Comparison between the free energy profiles obtained from $z_{sys}(t)$ and $z_m(t)$ can be used to investigate the effect of the characteristics of the handles on the free energy landscape. To this end, we repeated the force-clamp simulations by varying the contour length ($L = 5 - 100$ nm) and persistence length ($l_p = 0.6$ and 70 nm) of the handles. Fig. 4.2 shows that the discrepancy between the measured free energy $F_{eq}(z_{sys})$ (dashed lines in green) and the molecular free energy $F_{eq}(z_m)$ (solid lines in red) increases for the more flexible and longer handles The solid black lines show the intrinsic free energy (with no handles), and shows that attaching handles stabilizes the folded state. For large L and small l_p, the basins of attraction in $F_{eq}(z_m)$ are not well resolved. The largest deviation between $F_{eq}(z_{sys})$ and $F_{eq}(z_m)$ is found when $l_p = 0.6$ nm and $L = 25$ nm ($L/l_p \approx 40$) (the graph enclosed by the orange box in Fig. 4.2a). In contrast, the best agreement between $F_{eq}(z_{sys})$ and $F_{eq}(z_m)$ is found for $l_p = 70$ nm and $L = 5$ nm (the graph inside the magenta box in Fig. 4.2), which corresponds to $L/l_p \approx 0.07$. Thus, measurement of the free energy landscape requires the use of stiff handles. In the LOT experiments, $L/l_p \approx 6 - 7$ [100, 20, 23].

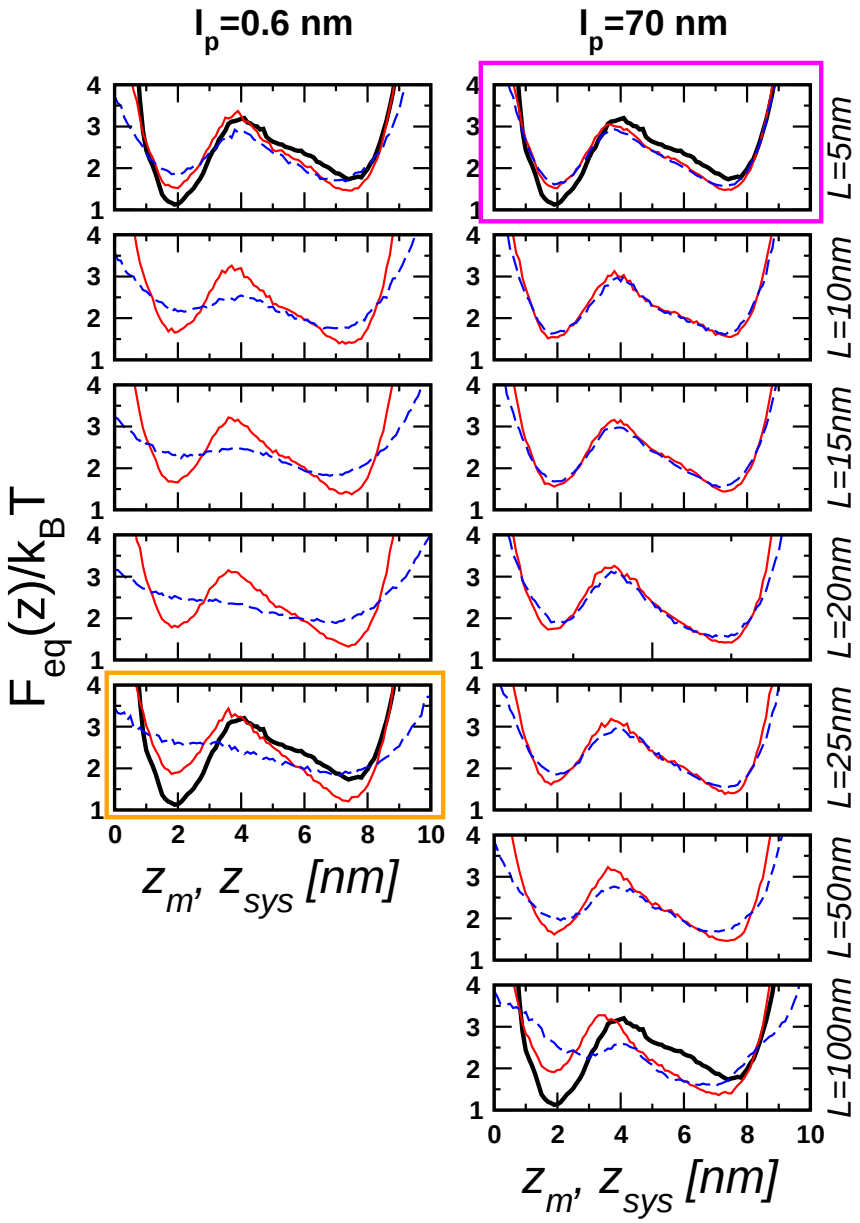

Figure 4.2: The free energy profiles, $F_{eq}(z_{sys})$ (dashed line in green) and $F_{eq}(z_m)$ (solid line in red), calculated using the histograms obtained from the time traces $z_{sys}(t)$ and $z_m(t)$ for varying L and l_p. $F_{eq}(z_{sys})$ and $F_{eq}(z_m)$ for a given l_p and different L are plotted in the same graph to highlight the differences. The intrinsic free energy $F^o_{eq}(z_m)$, the free energy profile in the *absence* of handle, is shown in black. The condition that produces the least deviation ($l_p = 70$ nm, $L = 5$ nm) and the condition of maximal difference ($l_p = 0.6$ nm, $L = 25$ nm) between $F_{eq}(z_m)$ and $F_{eq}(z_{sys})$ are enclosed in the magenta and orange boxes, respectively.

4.3 The Generalized Rouse Model

In order to establish the generality of the relationship between the free energy profiles as measured by z_m and those measured by z_{sys}, we introduce an exactly solvable Generalized Rouse Model (GRM) that minimally represents the system of RNA and handles (Fig. 4.3a). We mimic the hairpin using a Gaussian chain with N_0 bonds and Kuhn length a. The endpoints of the RNA mimic are harmonically trapped in a potential with stiffness k, as long as they are within a cutoff distance $c = 4$nm. Two handles, each with N_h bonds and Kuhn length b, are attached to the ends of the RNA. We fix one endpoint of the entire chain at the origin, and apply a force $f_m \approx 15.4$ pN to the other end. This gives the discrete Hamiltonian

$$\begin{aligned}\beta H &= \frac{3}{2b^2}\sum_{i=1}^{N_h}(\mathbf{r}_{i+1}-\mathbf{r}_i)^2 + \frac{3}{2b^2}\sum_{i=N_h+N_0+1}^{2N_h+N_0}(\mathbf{r}_{i+1}-\mathbf{r}_i)^2 - \beta\mathbf{f}\cdot(\mathbf{r}_N-\mathbf{r}_1) + \beta k_0\mathbf{r}_1^2 \\ &\quad + \frac{3}{2a^2}\sum_{i=N_h+1}^{N_h+N_0}(\mathbf{r}_{i+1}-\mathbf{r}_i)^2 + \beta V[\mathbf{r}_{N-N_h+1}-\mathbf{r}_{N_h+1}], \qquad (4.1)\end{aligned}$$

where

$$V[\mathbf{r}] = \begin{cases} k\mathbf{r}^2 & |\mathbf{r}| \le c \\ kc^2 & |\mathbf{r}| > c \end{cases}. \qquad (4.2)$$

The first two terms in equation (4.1) are the discrete connectivity potentials for the two handles, each with N_h bonds ($N_h + 1$ monomers), and with Kuhn length b. The mechanical force $\mathbf{f}$ in the third term is applied along the z direction, with $|\mathbf{f}| = f_m = 15.4$ pN. We also fix the 5' end of the system with a harmonic bond of strength $k_0 = 2.5\times10^4$ pN·nm^{-1} in the fourth term of eq. 4.1. The fifth term mimics the RNA hairpin with N_0 bonds and spacing $a = 0.5$ nm. Interactions between

the two ends of the RNA hairpin are modeled as harmonic bond with strength $k \approx 0.54$ pN·nm^{-1} that is cut off at $c = 4$ nm (eq. 4.2). When $|\mathbf{R}_m|$ exceeds 4nm, the bond is broken, mimicking the unfolded state. This minimal model treats the complicated interactions of the P5GA hairpin as a unified interaction at the endpoints, appropriate because of the all-or-none transition from the extended to the folded state seen in the simulations (Fig. 4.1).

The distribution functions of both R_m ($\approx z_m$ at high f) and R_{sys} ($\approx z_{sys}$ at high f) are exactly solvable in the continuum representation,

$$\beta H[\mathbf{r}(s)] = \frac{3}{2b^2}\int_0^{N_h} ds\ \mathbf{r}^2(s) + \frac{3}{2a^2}\int_{N_h}^{N-N_h} ds\ \mathbf{r}^2(s) + \frac{3}{2b^2}\int_{N-N_h}^{N} ds\ \mathbf{r}^2(s) \qquad (4.3)$$
$$+\beta k_0 \mathbf{r}_0^2/2 - \beta f(z_L - z_0) + \beta V[\mathbf{r}(N-N_h) - \mathbf{r}(N_h)]$$

with $V[\mathbf{r}]$ given in eq. 4.2. The condition that the 5' end of the chain is fixed (i.e. $k_0 \to \infty$) can be easily satisfied in the continuum limit with a delta function restriction ($\delta[\mathbf{r}_0]$) in the averages. As each segement of the chain is Gaussian, we can then compute

$$\langle \cdots \rangle_{GRM} = \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 d^3\mathbf{r}_3 (\cdots)\ Z_G(\mathbf{r}_1; N_h, b) Z_G(\mathbf{r}_2 - \mathbf{r}_1; N_0, a) \qquad (4.4)$$
$$\times Z_G(\mathbf{r}_3 - \mathbf{r}_2; N_h, b) e^{\beta \mathbf{f}\cdot\mathbf{r}_3 - \beta V[\mathbf{r}_2 - \mathbf{r}_1]}$$

where $\mathbf{r}(N_h) = \mathbf{r}_1$, $\mathbf{r}(N - 2N_h) = \mathbf{r}_2$, and $\mathbf{r}(N) = \mathbf{r}_3$, and where each segment of the chain has the Gaussian propagator

$$Z_G(\mathbf{r}; N, a) = \left(\frac{3}{2\pi N a^2}\right)^{\frac{3}{2}} \exp\left(-\frac{3\mathbf{r}^2}{2Na^2}\right) \qquad (4.5)$$

We are interested in the distribution functions $P_{eq}(\mathbf{R}_{sys}) = \langle \delta[\mathbf{r}(N) - \mathbf{R}_{sys}] \rangle_{GRM}$ and $P_{eq}(\mathbf{R}_m) = \langle \delta[\mathbf{r}(N - N_h) - \mathbf{r}(N_h) - \mathbf{R}_m] \rangle_{GRM}$. Because of the relatively high

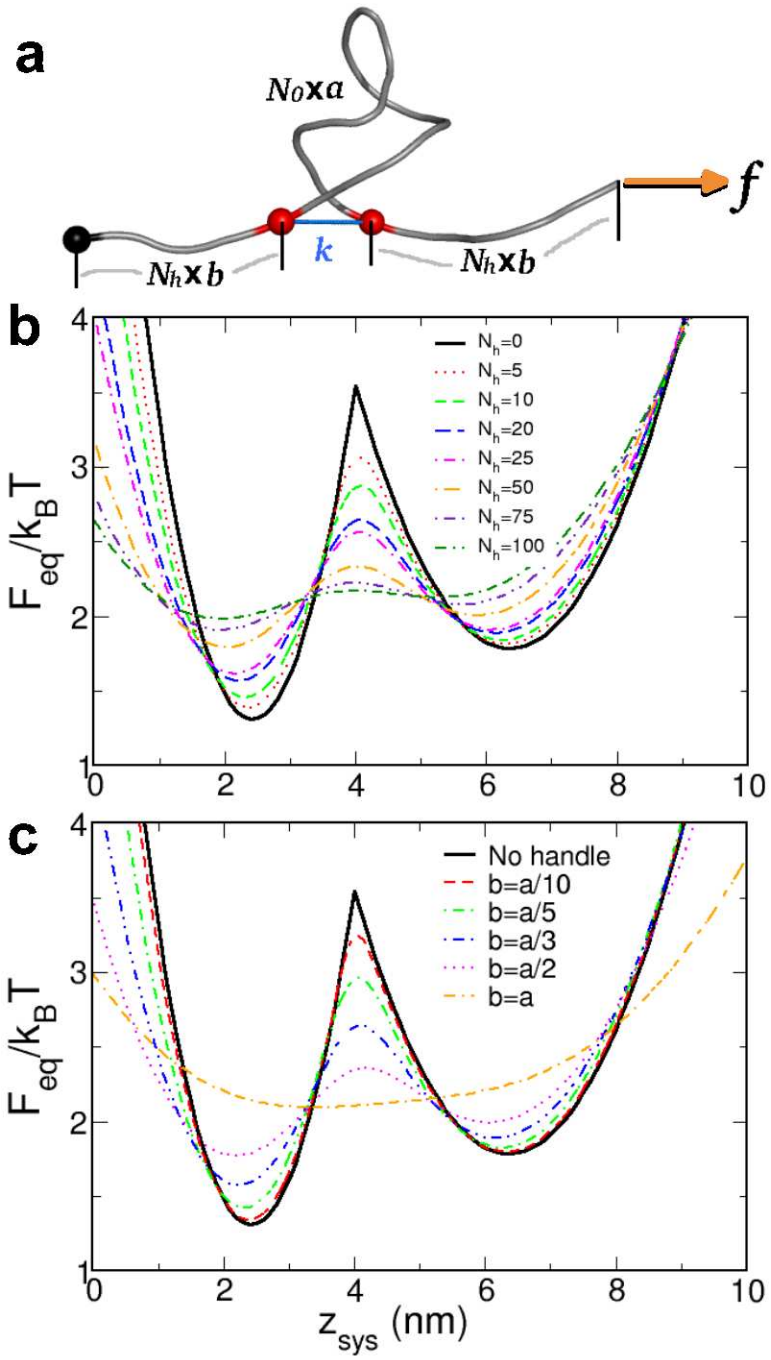

Figure 4.3: Free energy profiles for the GRM. **a**. A schematic diagram of the GRM, showing the number of monomers (N_0 and N_h) and Kuhn lengths (a and b) in each region of the chain, the harmonic interaction between the ends of the RNA mimic, and the external tension. **b**. The free energy profile for a fixed $b(= a/3)$ and increasing N_h as a function of $R_{sys} \approx z_{sys}$. The barrier heights decrease and the well depths increase for increasing N_h. **c**: The free energy profile for fixed $N_h = 20$ and varying b. The barrier heights decrease and the well depths increase for increasing b. In both b and c, the profiles are shifted so that the positions of the local maxima and minima coincide with those of the intrinsic free energy (with $N_h = 0$).

external tension (15.4pN) applied to the system, we note that $P_{eq}(z_m) \approx P_{eq}(R_m)$ and $P_{eq}(z_{sys}) \approx P_{eq}(R_{sys})$. Both of these averages can be computed directly, with

$$P_{eq}(z_m) = \frac{R_m}{\mathcal{N}} e^{-3R_m^2/2a^2N_0 - \beta V[R_m]} \sinh(\beta f R_m) \tag{4.6}$$

$$P_{eq}(z_{sys}) = \sqrt{\frac{3}{\pi b^2 N_h}} \frac{\sinh(\beta f R_{sys})}{\mathcal{N}} \int_0^\infty dy\ y \sinh\left(\frac{3yR_{sys}}{N_h b^2}\right) \times e^{3y^2/2N_0a^2 - 3(y-R_{sys})^2/4N_h b^2 - \beta V[y]} \tag{4.7}$$

$$\mathcal{N} = \int_0^\infty dy\, y e^{-3y^2/2a^2N_0 - \beta V[y]} \sinh(\beta f y) \tag{4.8}$$

To model the P5GA system using the GRM, we choose our parameters (k, N_0, and a) such that the handle-free distributions are closely matched by $P_{eq}(z_m)$. Note that $P_{eq}(R_m)$ in the GRM, eq. 4.6, is independent of the handle characteristics (N_h and b), which agrees with the relative invariance of $P_{eq}(R_m)$ seen in the simulations (solid red lines in Fig. 4.1). We find that choosing $N_0 = 20$, $a = 0.5$nm, and $k = 0.54$pN/nm gives good agreement with the P5GA free energies (solid black line in Fig. 4.3b-c).

While the stiffness in the handles of the simulated system (Fig. 4.1) cannot be accurately modeled using a Gaussian chain, the primary effect of attaching the handles is to alter the fluctuations of the endpoints of the RNA. By equating the longitudinal fluctuations for the WLC, $\langle \delta \mathbf{R}_{||}^2 \rangle_{WLC} \sim L l_p^{-1/2} (\beta f)^{-3/2}$, with the fluctuations for the Gaussian handles, $\langle \delta \mathbf{R}_{||}^2 \rangle_G \sim Lb$, we estimate that the effective persistence length of the handles scales as $l_p^{eff} \sim b^{-2} f^{-3}$. Thus, smaller spacing in the Gaussian handles in the GRM will mimic stiffer handles in the H-RNA-H system. The free energies computed for the GRM, shown in Fig. 4.3b-c, are consistent with the results of the simulations. The free energy profiles deviate significantly

from $F^o_{eq}(z_m)$ as N_h increases or 'stiffness' decreases. The relevant variable that determines the accuracy of $F_{eq}(z_{sys})$ is $N_h b^2 \sim L/l_p^{eff}$, with the free energies remaining unchanged if $N_h b^2$ is kept constant. The GRM confirms that accurate measurement of the folding landscape using z_{sys} requires stiff handles.

4.4 The Effect of Handles on the Kinetics of Hairpin Formation

Because LOT experiments can also be used to measure the force-dependent rates of hopping between the NBA and the UBA, it is important to assess the influence of the dynamics of the handles on the intrinsic hopping kinetics of the RNA hairpin. In other words, how should the structural characteristics of the linkers be chosen so that the measured hopping rates using the time traces $z(t)$ and the intrinsic rates are as close as possible?

We first performed force clamp simulations of P5GA in the absence of handles to obtain the intrinsic (or ideal) folding (τ_F^o) or unfolding (τ_U^o) times, that serve as a reference for the H-RNA-H system. To obtain the boundary conditions for calculating the mean refolding and unfolding times, we collected the histograms of the time traces and determined the positions of the minima of the NBA and UBA, $z_F = 1.9$ nm and $z_U = 7.4$ nm (Fig. 4.4a). The analysis of the time traces provides the transition times in which z_m reaches $z_m = z_U$ starting from $z_m = z_F$. The mean unfolding time τ_U is obtained using $\tau_U = 1/N \sum_i \tau_U(i)$, with the mean folding time similarly calculated. The values of τ_U^o and τ_F^o computed from the time trace of $z_m(t)$

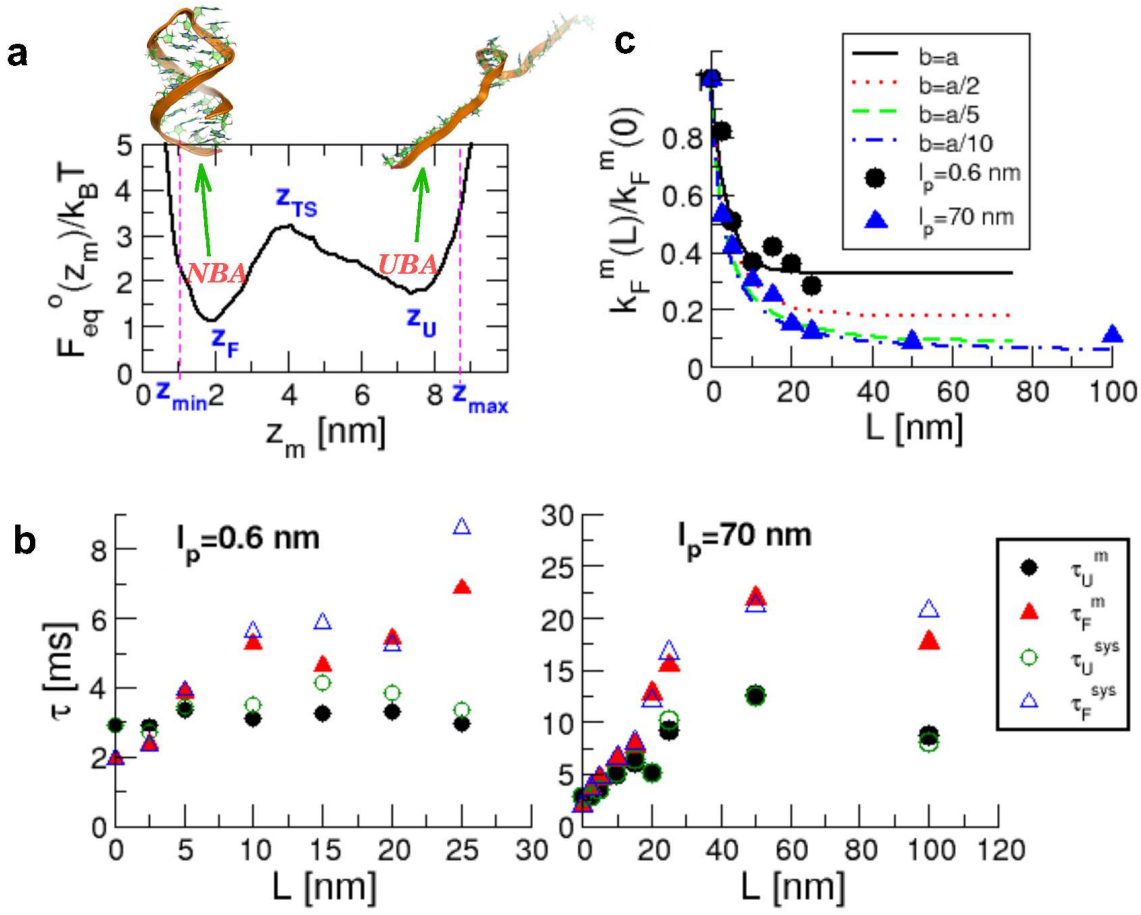

Figure 4.4: **a**. The free energy profile for P5GA with $L = 0$ nm. **b**. The transition times at $f = f_m$, obtained using $z_m(t)$ (filled symbols) and $z_{sys}(t)$ (empty symbols). The ratio $\tau^m_{U(F)}/\tau^{sys}_{U(F)} \approx 1$, which shows that $z_{sys}(t)$ mirrors the hopping of P5GA. **c**. Folding rate $k^m_F(L)/k^m_F(0)$ as a function of L for varying b, using the GRM. The plots show b/a =1, 1/2, 1/5, and 1/10. The simulation results for P5GA are also shown as symbols, to emphasize that the GRM accounts for the hopping kinetics in the H-RNA-H system accurately.

are 2.9 ms and 1.9 ms, respectively. At $f_m = 15.4$ pN and $L = 0$ nm, the equilibrium constant $K_{eq} = \tau_F^o/\tau_U^o = 0.67$, which shows that the bare molecular free energy is slightly tilted towards NBA at $f = 15.4$ pN (see also Fig. 4.2).

The values of the folding (τ_F^m) and unfolding(τ_U^m) times were also calculated for the P5GA hairpin with attached handles (Fig. 4.1). As the length of the handles increases both τ_U^m and τ_F^m increase gradually, and the equilibrium distribution shifts towards the UBA, i.e. $K_{eq} = \tau_F^m/\tau_U^m$ increases (Fig. 4.4b). Strikingly, the use of flexible handles results in minimal deviations of τ_U^m and τ_F^m from their intrinsic values (Fig.4.4b). The attachment of handles (stiff or flexible) to the 5' and 3' ends restricts their movement, which results in a decrease in the number of paths to the NBA and UBA. Thus, both τ_U^m and τ_F^m increase (Fig.4.4b). As the stiffness of the handles increases, the extent of pinning increses. These arguments show that flexible and short handles, that have the least restriction on the fluctuations of the 5' and 3' ends of the hairpin, cause minimal perturbation to the intrinsic RNA dynamics, and hence the hopping rates.

Because the experimentally accessible quantity is the extension of the H-RNA-H system (z_{sys}), it is important to show that the transition times can be reliably obtained using $z_{sys}(t)$. Although $z_{sys}(t)$ differs from $z_m(t)$ in amplitude, the "phase" between the two quantities track each other reliably throughout the simulation, even when the handles are long and flexible. We calculated τ_U^{sys} and τ_F^{sys} by analyzing the trajectories $z_{sys}(t)$ using the same procedure used to compute their intrinsic values. Comparison of τ_U^{sys} (τ_F^{sys}) and τ_U^m (τ_F^m) for both stiff and flexible handles shows excellent agreement at all L values (Fig.4.4b). Thus, it is possible to infer

the RNA dynamics $z_m(t)$ by measuring $z_{sys}(t)$. However, attaching handles will still alter the hopping times, increasing both the folding and unfolding rate noticeably (see Fig. 4.4), so that $\tau^{sys}_{U(F)} \approx \tau^{m}_{U(F)} < \tau^{0}_{U(F)}$.

The simulation results can be fully understood using the GRM (Fig 4.3a), for which we can exactly solve the overdamped Langevin equation using the discrete representation of the Gaussian chain. We can use the theory of Wilemski and Fixman (WF) [112] to determine an approximate loop closure time for the H-RNA-H system. The WF theory writes the loop closure time in terms of a sink-sink correlation function

$$\tau_F^m \approx \int_0^\infty dt\; C(\mathbf{R}_m) P_{eq}(\mathbf{R}_m, t|\mathbf{R}_m^0) C(\mathbf{R}_m^0) P_{eq}(\mathbf{R}_m^0) \tag{4.9}$$

with $P_{eq}(\mathbf{R}_m, t|\mathbf{R}_m^0)$ the time-dependent propagator for $\mathbf{R}_m$, giving the probability of finding $\mathbf{r}_{N-N_h+1}(t) - \mathbf{r}_{N_h+1}(t) = \mathbf{R}_m$ given the initial separation $\mathbf{r}_{N-N_h+1}(0) - \mathbf{r}_{N-N_h+1}(0) = \mathbf{R}_m^0$. This propagator is calculated exactly in terms of the GRM interaction matrix $\mathbf{M}$ (given in Appendix C.2, eq. C.5 with $k = 0$), with

$$P_{eq}(\mathbf{R}_m, t|\mathbf{R}_m^0) = \Big(\frac{3}{2\pi\Delta x_{eq}^2\sqrt{1-h^2(t)}}\Big)^3 \exp\Big(-\frac{3}{2\Delta x_{eq}^2[1-h^2(t)]} \times\Big[(\bar{\mathbf{R}}_m^0)^2 + \bar{\mathbf{R}}_m^2 - 2h(t)\bar{\mathbf{R}}_m^0\cdot\bar{\mathbf{R}}_m\Big]\Big) \tag{4.10}$$

with $\bar{\mathbf{R}_m} = \mathbf{R}_m - \langle\mathbf{R}_m\rangle$ (see Appendix C.3 for details). Δx_{eq} and $h(t)$ can be written in terms of the interaction matrix $\mathbf{M}$, given in Appendix C.2. We find

$$\begin{aligned}
\Delta x_{eq}^2 &= a^2\Big[\Big(\mathbf{M}^{-1}\Big)_{N_h+1,N_h+1} + \Big(\mathbf{M}^{-1}\Big)_{N-N_h+1,N-N_h+1} - 2\Big(\mathbf{M}^{-1}\Big)_{N_h+1,N-N_h+1}\Big] \\
h(t) &= \frac{a^2}{\Delta x_{eq}^2}\Big[\Big(\mathbf{M}^{-1}\mathbf{G}(t)\Big)_{N_h+1,N_h+1} + \Big(\mathbf{M}^{-1}\mathbf{G}(t)\Big)_{N-N_h+1,N-N_h+1} \\
&\qquad -2\Big(\mathbf{M}^{-1}\mathbf{G}(t)\Big)_{N_h+1,N-N_h+1}\Big]
\end{aligned} \tag{4.11}$$

and where $\mathbf{G}(t) = \exp(-3D\mathbf{M}t/a^2)$ (see Appendix C.3). Δx_{eq} and $h(t)$ are most easily determined by diagonalizing the matrix $\mathbf{M}$, which must be done numerically for $b \neq a$. We take $C(\mathbf{R}_m) = \delta[|\mathbf{R}_m| - c]$, i.e. the GRM is considered 'folded' when the interactions in the RNA mimic are turned on (see eq. 4.2). Because of the high external tension (15.4pN), we approximate $P_{eq}(\mathbf{R}_m, t|\mathbf{R}_m^0) \approx P_{eq}(z_m, t|z_m^0)$. We then finally find

$$\tau_F^m \propto \int_0^\infty dt\ P_{eq}(c, t|c) \tag{4.12}$$

The refolding rate of the RNA hairpin under tension is analogous to $k_F^m = \tau_F^m$. A plot of $k_F^m(L)/k_F^m(0)$ versus L (Fig. 4.4c) illustrates that smaller deviations from the handle-free values occur when l_p is small. We use the WF theory to determine an approximate time of contact formation ($\tau_F^m = (k_F^m)^{-1}$) as a function of b (i.e. increasing handle 'stiffness') and N_h. Moreover, Fig. 4.4c shows that the refolding rate decreases for increasing N_h regardless of the stiffness of the chain. The saturating value of k_F^m as $N_h \to \infty$ depends on b, with 'stiffer' handles having a much larger effect on the folding rate, as seen in the simulations.

We also find the dependence of k_F on L agrees well with the behavior observed in the simulation of P5GA. The ratio $k_F^m(L)/k_F^m(0)$ for $b = a$ agrees well with the trends of the flexible linker ($l_p = 0.6$ nm) for all of the simulated lengths, with both saturating at $k_F(L) \approx 0.35k_F(0)$ for large L. The trends for 'stiffer' chains (smaller b) in the GRM qualitatively agree with the P5GA simulation with stiff handles ($l_p = 70$ nm), with remarkably good agreement for $0.1 \leq b/a \leq 0.2$ over the entire range of L. The GRM, which captures the physics of both the equilibrium

and kinetic properties of the more complicated H-RNA-H, provides a theoretical basis for extracting kinetic information from experimentally (or computationally) determined folding landscapes.

4.5 Free energy landscapes and hopping rates

Stiff handles are needed to obtain $F_{eq}(z_{sys})$ [23] that resembles $F^o_{eq}(z_m)$, whereas the flexible handles produce hopping rates that are close to their handle-free values. These two findings appear to demand two mutually exclusive requirements in the choice of the handles in LOT experiments. However, if z_m is a good reaction coordinate, then it should be possible to extract the hopping rates using accurately measured $F_{eq}(z_{sys})(\approx F_{eq}(z_m) \approx F^o_{eq}(z_m))$ at $f \approx f_m$, using handles with small L/l_p. The (un)folding times can be calculated using the mean first passage time (Kramers' rate expression) with appropriate boundary conditions [113],

$$\begin{aligned} \tau_U^{KR} &= \int_z^{z_U} dy e^{\beta F_{eq}(y)} \frac{1}{D_U} \int_{z_{min}}^{y} dx e^{-\beta F_{eq}(x)}, \\ \tau_F^{KR} &= \int_{z_F}^{z} dy e^{\beta F_{eq}(y)} \frac{1}{D_F} \int_{y}^{z_{max}} dx e^{-\beta F_{eq}(x)}, \end{aligned} \tag{4.13}$$

where z_{min}, z_{max}, z_U and z_F are defined in Fig. 4.4a. The effective diffusion coefficient $D_F(D_U)$ is obtained by equating τ_F^{KR} (τ_U^{KR}) in equation (4.13), with $F_{eq}(z_m) = F^o_{eq}(z_m)$, to the simulated τ_F^o (τ_U^o). We calculated the f-dependent $\tau_U^m(f)$ and $\tau_F^m(f)$ by evaluating equation (4.13) using $F^o_{eq}(z_m|f) = F^o_{eq}(z_m|f_m) - (f - f_m)\cdot z_m$. The calculated and simulated results for P5GA are in good agreement (Fig 4.5a-b). At the higher force ($f = 16.8$ pN), the statistics of hopping transition within our simulation time is not sufficient to establish ergodicity. As a result, the simulation

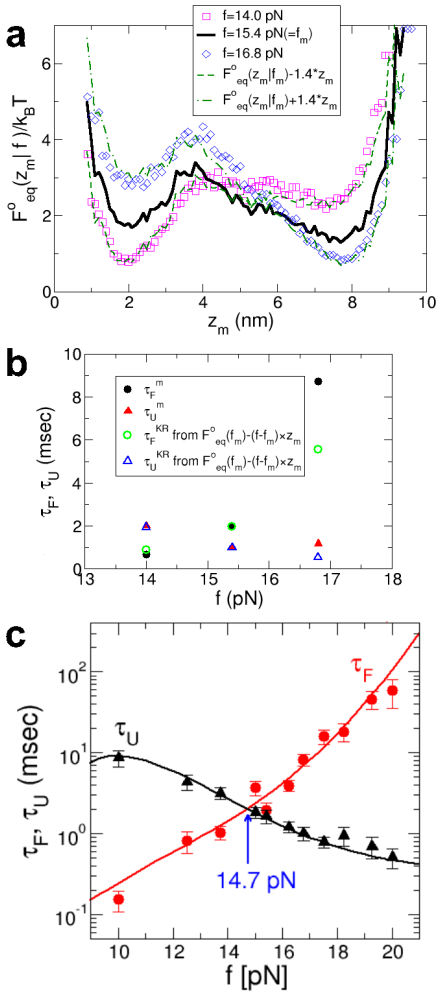

Figure 4.5: **a**. Comparison of the measured free energy profiles (symbols) with the shifted free energy profiles $\beta F^o(z_m|f_m) - \beta(f - f_m) \cdot z_m$. **b**. Folding and unfolding times as a function of force $f = 14$ pN $< f_m$, $f = 15.4$ pN $\approx f_m$, and $f = 16.8$ pN $> f_m$. $\tau^m_{F(U)}$ is obtained from the time trace in Figure 2B in Ref [72] at each force, while $\tau^m_{F(U)}$ is computed using the tilted profile $\beta F^o_{eq}(z_m|f) = \beta F^o_{eq}(z_m|f_m) - \beta(f - f_m) \cdot z_m$ in equation (4.13). **c**. Folding and unfolding times using the GRM. Symbols are a direct simulation of the GRM (error bars are standard deviation of the mean). The solid lines are obtained using the Kramers theory (equation (4.13)). We choose $D_U \approx 3D_0$, so that that the simulated and Kramers times agree at $f = f_m$. The position of each basin of attraction as a function of force for the GRM is given by $z_U \approx N_0 a^2 \beta f/3$ and $z_F \approx N_0 a^2 \beta f/(3 + 2N_0 a^2 \beta k)$.

results are not as accurate at high forces (data not shown). To further show that the use of $F^o_{eq}(z_m|f)$ in equation (4.13) gives accurate hopping rates, we calculated $\tau^o_U(f)$ for the GRM and compared the results with direct simulations of the handle-free GRM, which allows the study of a wider range of forces (using the Hamiltonian in eq. 4.1). The results in Fig. 4.5c show that $F^o_{eq}(z_m|f)$ indeed gives very accurate values for the transition times from the UBA and NBA over a wide force range.

4.6 Conclusions

The self-assembly of RNA and proteins may be viewed as a diffusive process in a multi-dimensional folding landscape. To translate this physical picture into a predictive tool, it is important to discern a suitable low-dimensional representation of the complex energy landscape, from which the folding kinetics can be extracted. Our results show that, in the context of nucleic acid hairpins, precise measurement of the sequence-dependent folding landscape of RNA is sufficient to obtain good estimates of the f-dependent hopping rates in the absence of handles. It suffices to measure $F_{eq}(z_{sys}) \approx F_{eq}(z_m) \approx F^o_{eq}(z_m)$ at $f = f_m$ using stiff handles, while $F_{eq}(z_m|f)$ at other values for f can be obtained by tilting $F_{eq}(z_m|f_m)$. The accurate computation of the hopping rates using $F_{eq}(z_m)$ show that z_m is an excellent reaction coordinate nucleic acid hairpins under tension. Further theoretical and experimental work is needed to test if the proposed framework can be used to predict the force dependent hopping rates for other RNA molecules that fold and unfold through populated intermediates.

Chapter 5

Spherical Confinement of Wormlike Chains

5.1 Introduction

The Wormlike Chain (WLC) model [114] has been shown to accurately model the intrinsic stiffness (or resistance to bending) of DNA [115, 1] under a variety of conditions. The WLC model incorporates this stiffness into a persistence length l_p ($\approx$ 50nm$\approx$ 150bp for DNA), the length scale over which the polymer is difficult to bend. When confined upon or within a curved geometry, the interplay between the intrinsic stiffness of the polymer and the curvature of the confinement may drastically change the behavior of the polymer.

Confinement of a polymer to a curved surface has been an area of active interest. In eukaryotes, the first level of chromosomal compaction of DNA is histone wrapping, with the DNA wrapped around the cylindrical histone (with radius 4.2nm and height 2.4nm) about 1.7 times [116, 24, 26]. The stability of this tightly bent structure is essential in understanding the development of the chromosome. Many authors have studied theoretically the behavior of polymers confined to cylinders [117, 118, 119] and spheres [117, 120, 121, 122]. The symmetry of spherical confinement An exact solution for the end-to-end distance $\mathbf{R}$ of a WLC confined to the surface of a sphere [120] from a microscopic Hamiltonian has been shown to be correct using simulations [121].

Cylindrical confinement is of particular interest due to the change in the effective dimensionality of the system [77] (with tight cylindrical confinement becoming effectively one dimensional as the radius vanishes), inducing changes in the length scales of the system [123, 124, 125, 34, 126]. The ability to fabricate nanochannels allows direct study of such a confined system, and many theoretical results have been confirmed [127, 31].

Experiments on the dsDNA-containing bacteriophages $\phi 29$ [128, 129] and $\epsilon 15$ [29], as well as the T [130, 131, 30, 132, 133], P [134], and λ [135, 136, 137, 138] phage classes, have all determined a number of details of the structures, pressures, and ejection timescales of many viruses. Regardless of the shape of the viral capsid, it is generally seen that the DNA orders itself in concentric rings [130, 135, 29, 131, 30, 136, 139], with the spacing between rings $\sim$0.3nm. The 5' endpoint of the DNA has been shown to be found near the center of the encapsulated mass of DNA [132, 133], with the 3' end found near the capsid wall [138]. Single molecule experiments [67] have shown the pressure on the capsid walls to be on the order of 60atm, inducing a significant resistance to the DNA encapsulation. These observations have generated a number of theoretical studies, primarily interested in the packaged structure [28, 140, 141, 142, 143], inter-strand spacing [144, 27, 145], energy or pressure [146, 144, 28, 142, 147, 148, 149, 150], and the loading or ejection process [27, 151, 28, 140, 152, 142]. These studies have predicted a number of properties of an encapsulated chain. While the specific geometry of the confining viral capsid varies from phage to phage, the properties of the encapsulated DNA can be studied using spherical [141, 152, 140] or cylindrical [144, 145] to a very good approximation.

The study of confined WLC's in these simple geometries is thus of great relevance to improving our understanding of the properties of viruses.

In order to study the effects of both surface and volume confinement on the behavior of a WLC, we will use the mean field method [41, 153, 42] of Ha and Thriumalai. This method has been successful in computing a number of different properties of a WLC in a number of different situations [153, 62], and can be used to approximate the effect of intra-chain interactions [154]. Additionally, the mean field method has been shown to produce a tractable theory for a closed, smooth chain on the surface of a sphere [122]. This chapter is organized as follows. In Sec. 5.2, we determine the mean field theory for an open WLC, and show that it reproduces all known averages and scaling laws. We also show that the mean field theory accurately reproduces the correct scaling coefficient of the free energy of confinement. In Sec. 5.3, we determine the behavior of a surface-confined WLC under the application of an external force. In Sec. 5.4, we adapt the mean field theory to the case of volume confinement. We show that the volume confined chain behaves similar to a surface confined chain, and explicitly determine the pressure due to confinement. The mean field theory is able to accurately reproduce a number of properties observed in the simulations. We also show that the structural order of the confined WLC can be understood by defining a local winding axis $\mathbf{a}_i$. Finally, in Sec. 5.5, we consider a self-avoiding WLC confined to the interior of a sphere. We show that the the development of ordered shells can be again understood via the local winding axis, and find that ordered structures depend sensitively on the relative stiffness (l_p/R).

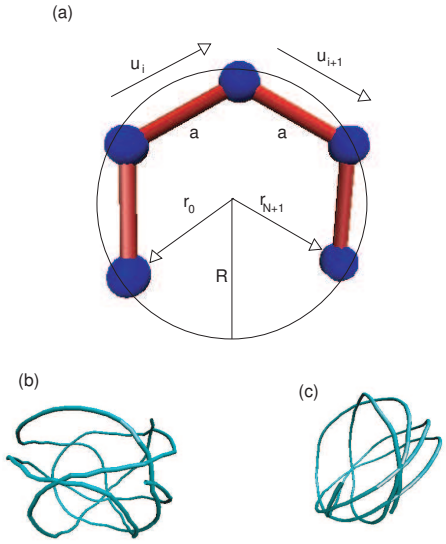

Figure 5.1: (a) Schematic of the problem of surface confinement, showing the position and bond vectors $\mathbf{r}_n$ and $\mathbf{u}_n$. (b) One structure of a surface confined WLC, with $R = 8a$ and $l_p = 20a$. (c) Same as (b), with $R = 8a$ and $l_p = 100a$.

5.2 Confinement to the Surface of a Sphere

We will begin by developing the Mean Field formalism for a WLC with persistence length l_p, fixed inter-monomer spacing a, and length $L = Na$, confined to the surface of the sphere of radius R. We define $\mathbf{r}_n = (x_n, y_n, z_n)$ the position of the n^{th} monomer, and the bond spacing $\mathbf{u}_n = \Delta\mathbf{r}_n = \mathbf{r}_{n+1} - \mathbf{r}_n$ with $|\mathbf{u}_n| \equiv a$. (shown in Fig. 5.1). The distribution of the chain in phase space is

$$\begin{aligned}\Psi_S(\{\mathbf{r}_n\}) &\propto \prod_n \delta(\mathbf{r}_n^2 - R^2)\ \delta(\Delta\mathbf{r}_n^2 - a^2)\ e^{-l_p/2a^3\,(\Delta\mathbf{r}_{n+1} - \Delta\mathbf{r}_n)^2} \\ &\propto \int_{-i\infty}^{i\infty} \prod_{n=1}^{N+1} dk_n d\lambda_n \exp\Big[-\frac{1}{2} a l_p \frac{(\Delta\mathbf{r}_{n+1} - \Delta\mathbf{r}_n)^2}{a^4} - a\lambda_n\Big(\frac{\Delta\mathbf{r}_n^2}{a^2} - 1\Big) \\ &\qquad -ak_n\Big(\frac{\mathbf{r}_n^2}{R^2} - 1\Big)\Big], \qquad (5.1)\end{aligned}$$

where the second line follows from the first after a Fourier transform of the delta functions (with the Fourier variables $\{\lambda_n\}$ and $\{k_n\}$. Following Ha and Thirumalai [41, 42], we can write the partition function as $Z = \int \prod_n d^3\mathbf{r}_n \Psi_S(\{\mathbf{r}_n\}) \equiv \int \prod_n d\lambda_n dk_n \exp(-\mathcal{F}_S[\{\lambda_n, k_n\}])$, which defines the free energy functional for surface confinement, $\mathcal{F}_S$. We can write $\mathcal{F} = \mathcal{F}_x + \mathcal{F}_y + \mathcal{F}_z - a\sum_n(\lambda_n + k_n)$, with the free energy functional $\mathcal{F}_x$ given by

$$e^{-\mathcal{F}_x} = \int \prod_n dx_n \; e^{-\mathcal{H}_x} \qquad \mathcal{H}_x = a\sum_n \left(\frac{l_p}{2}\,\frac{(\Delta x_{n+1} - \Delta x_n)^2}{a^4} + \lambda_n \frac{\Delta x_n^2}{a^2} + k_n \frac{x_n^2}{R^2}\right) \quad (5.2)$$

In the Mean Field approximation, we assume $\mathcal{F}$ is sharply peaked around $\{\lambda_n, k_n\} = \{\lambda_n^*, k_n^*\}$, so that $Z \sim \exp(-\mathcal{F}^*)$. The optimal values of λ and k will occur when $\mathcal{F}$ is minimized, so we must solve $\partial\mathcal{F}/\partial\lambda_n = \partial\mathcal{F}/\partial\kappa_n = 0$ (i.e. a saddle point approximation). In this approximation, we have dropped the integrals over k_n and λ_n, so that these Fourier variables play the role of spring constants restricting the position and bending of the chain, respectively.

Since the discrete Hamiltonian is quadratic in the x_n's, we can write $\mathcal{F} = 3/2\;\log[\mathrm{Det}(\mathbf{Q})] - a\sum_n(\lambda_n + k_n) + \mathrm{const}$, with $\mathbf{Q}$ a symmetric $(N+1)\times(N+1)$ tridiagonal matrix, given in Appendix D (eqs D.2-D.4). Solving these coupled equations becomes intractable for large N, and we must make additional approximations. The symmetry of the matrix is respected by the substitution $\lambda_n \to \lambda$ and $k_n \to k$, except for exactly three elements near the endpoints (see Appendix D for more details). This is similar to the excess endpoint fluctuation terms found in the unconfined theory[41, 42], where λ was shown to be constant except at the endpoints, requiring an additional mean field variable to suppress endpoint fluctuations, δ. A

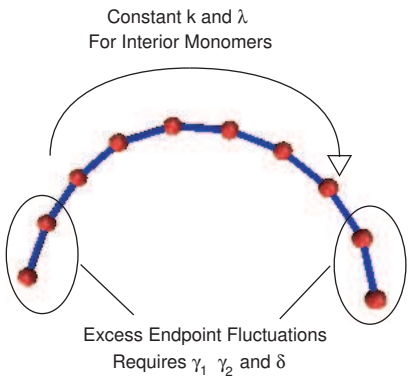

Figure 5.2: Schematic of mean field variables. k and λ are constant for all interior points, with γ_1, γ_2, and δ suppressing excess fluctuations at the endpoints

similar and suggests choosing $k_1 = k_N = k + \gamma_1/a - \gamma_2/R$, $k_2 = k_{N-1} = k + \gamma_2/R$, $\lambda_1 = \lambda_{N-1} = \lambda + \delta/a - a\gamma_2^2/R^2$, with $k_n = k$ and $\lambda_n = \lambda$ otherwise (see Fig. 5.2). The specific forms of λ_1, k_1, and k_2 are chosen to ensure convergence of the continuum limit. With this assumption, we need only minimize $\mathcal{F}$ with respect to these new mean field variables in order to determine $\mathcal{F}^*$. Substitution of these values into the Hamiltonian in eq. 5.2 and taking the continuum limit (with $a \to 0$, $N \to \infty$, and $Na \to L$) we can separate the Hamiltonian into interior and endpoint terms, $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_e$, with

$$\mathcal{H}_0 = \int_0^L ds \left(\frac{l_p}{2} \ddot{x}^2(s) + \lambda \dot{x}^2(s) + \kappa x^2(s) \right) \tag{5.3}$$

$$\mathcal{H}_e = \delta(u_0^2 + u_L^2) + \gamma_1 \left(\frac{x_0^2}{R^2} + \frac{x_L^2}{R^2} \right) + 2\gamma_2 \left(u_0 \frac{x_0}{R} - u_L \frac{x_L}{R} \right), \tag{5.4}$$

where we have defined $u_0 = \dot{x}(0)$ and $u_L = \dot{x}(L)$, with $\dot{x} = \partial x(s)/\partial s$. The free

energy functional $\mathcal{F}_x$ in the continuum limit then becomes

$$\mathcal{F}_x = -\log\Big[\int d^4\mathbf{x} \int \mathcal{D}[x(s)] \exp(-\mathcal{H}_0 - \mathcal{H}_e)\Big], \tag{5.5}$$

with $\mathbf{x} = (x_0, x_L, u_0, u_L)$, and the total free energy becomes $\mathcal{F}_S = \mathcal{F}_x + \mathcal{F}_y + \mathcal{F}_z - \lambda L - \kappa L - 2\delta - 2\gamma_1$. The path integral in eq. 5.5 can be evaluated exactly [155], and we find

$$Z_0(\mathbf{x}) \equiv \int \mathcal{D}[x(s)]\ \exp(-\mathcal{H}_0[x(s)]) = K \exp\Big(\mathbf{x}\cdot\mathbf{M}\mathbf{x}\Big), \tag{5.6}$$

where $\mathbf{M}$ is a 4×4 matrix. $\mathbf{M}$ and K are evaluated in Appendix D, and given explicitly in eqs. D.9 and D.16. This relatively simple propagator allows us to calculate many averages with little difficulty. In computing $\mathbf{M}$ and K, we find that the confined wormlike chain depends on the two frequencies

$$\omega_i = \Big(\frac{\lambda}{l_p} \pm \sqrt{1 - \frac{2kl_p}{\lambda^2}}\ \Big)^{\frac{1}{2}}, \tag{5.7}$$

with averages on the mean field level expressed in terms of $\cosh(L\omega_i)$ and $\sinh(L\omega_i)$. These quantities can greatly simplified in the limit of large $L\omega_i$, which we will term strong confinement (see below).

The total free energy functional now becomes

$$\mathcal{F}_S \;=\; -3\log\Big(\int d^4\mathbf{x}\ Z(\mathbf{x})\Big) - \lambda L - \kappa L - 2\delta - 2\gamma_1 \quad Z(\mathbf{x}) = Z_0(\mathbf{x}) e^{-\mathcal{H}_e} \tag{5.8}$$

with Z_0 given in eq. 5.6. The optimal parameters λ, k, δ, γ_1, and γ_2 are obtained by solving the five coupled Mean Field equations,

$$\frac{\partial \mathcal{F}_S}{\partial \lambda} = \frac{\partial \mathcal{F}_S}{\partial k} = \frac{\partial \mathcal{F}_S}{\partial \delta} = \frac{\partial \mathcal{F}_S}{\partial \gamma_1} = \frac{\partial \mathcal{F}_S}{\partial \gamma_2} = 0. \tag{5.9}$$

Note that the λ, κ, δ and γ_1 derivatives immediately imply, respectively,

$$\frac{1}{L}\int_0^L ds\ \langle \mathbf{u}^2(s)\rangle = 1 \qquad \frac{1}{L}\int_0^L ds\ \langle \mathbf{r}^2(s)\rangle = R^2 \qquad \langle \mathbf{u}_0^2 + \mathbf{u}_L^2\rangle = 2 \qquad \langle \mathbf{r}_0^2 + \mathbf{r}_L^2\rangle = 2R^2$$

This suggests that the Mean Field approximation is equivalent to replacing the local requirements $\mathbf{u}^2(s) = 1$ and $\mathbf{r}^2(s) = R^2$ to the global conditions $\langle \mathbf{u}^2(s)\rangle = 1$ and $\langle \mathbf{r}^2(s)\rangle = R^2$. λ plays the role of a spring constant that keeps the bond spacing fixed on average, while k is a spring constant that keeps $\langle \mathbf{r}^2\rangle = R$ on average. The γ_2 derivative in eq. 5.9 implies $\langle \mathbf{r}_0 \cdot \mathbf{u}_0 - \mathbf{r}_L \cdot \mathbf{u}_L\rangle = 0$, which is expected for confinement to the surface of a sphere, as $\mathbf{u}(s)$ is always tangential to the surface of the sphere (i.e. $\mathbf{u}(s) \perp \hat{\mathbf{r}}$ for all s).

The solutions to the mean field equations (eq. 5.9) can be determined exactly for all L, l_p, and R, giving

$$\lambda = \frac{9}{8l_p} - \frac{l_p}{R^2} \qquad k = \frac{l_p}{2R^2} \qquad \delta = \frac{3}{4} \qquad \gamma_1 = \frac{3}{4} \qquad \gamma_2 = -\frac{l_p}{2R} \tag{5.10}$$

We note that λ, the effective spring constant between monomers which ensures inextensibility globally, becomes negative if $R < 2\sqrt{2}\,l_p/3$. For a free WLC, λ is found to be strictly positive, and creates an attractive interaction between neighboring monomers. Without this attractive interaction, the chain tends to 'explode' on average for the free chain, with $\langle \mathbf{u}^2(s)\rangle_{\lambda=0} > 1$. If the radius is sufficiently small, though, the bond vectors will be compressed, and we find $\langle \mathbf{u}^2(s)\rangle_{\lambda=0} < 1$. A repulsive interaction between monomers, with a negative value for λ, is required to prevent the chain from 'imploding' for $R < 2\sqrt{2}\,l_p/3$. With the solutions to the

mean field equations in eq. 5.10, eq. 5.7 becomes

$$\omega_i = \frac{3}{4l_p}\left(1 \pm \sqrt{1 - \frac{16l_p^2}{9R^2}}\right). \tag{5.11}$$

We note that, in the limit of large R, $L\omega_1 \sim L/l_p$ and $L\omega_2 \sim Ll_p/R^2$. Our demarcation of strong confinement, $L\omega_i \gg 1$, then requires stiff chains ($L \gg l_p$) and sufficiently small radii ($R \ll \sqrt{Ll_p}$). The ω_i's become complex if $R < 4l_p/3$, which leads to oscillatory behavior in both the partition function (see Appendix D), and the resulting average properties of the system.

The correlation functions in both position and bending can be computed, using the solutions to the mean field equations (eq. 5.10), giving

$$\begin{aligned}
\langle \mathbf{r}(s) \cdot \mathbf{r}(s') \rangle &= R^2 e^{-|\Delta s|/\zeta_S}\left[\cosh\left(\frac{|\Delta s|}{\zeta_S}\,\Omega_S\right) + \frac{1}{\Omega_S}\sinh\left(\frac{|\Delta s|}{\zeta_S}\,\Omega_S\right)\right] \\
\langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle &= e^{-|\Delta s|/\zeta_S}\left[\cosh\left(\frac{|\Delta s|}{\zeta_S}\,\Omega_S\right) - \frac{1}{\Omega_S}\sinh\left(\frac{|\Delta s|}{\zeta_S}\,\Omega_S\right)\right],
\end{aligned} \tag{5.12}$$

with $\zeta_S = 4l_p/3$, $\Omega_S = \sqrt{1 - 16l_p^2/9R^2}$, and $\Delta s = s - s'$. We can also explicitly confirm that the Mean Field roots satisfy the inextensibility and confinement constraints on average, finding $\langle \mathbf{u}^2(s) \rangle \equiv 1$ and $\langle \mathbf{r}^2(s) \rangle \equiv R^2$ for all s. Thus, the Mean Field theory replaces the rigid local constraints on the system by average local constraints. Note that, while our theory gives the exact second moment in both position and bending, we find $\langle \mathbf{u}^4 \rangle = \langle \mathbf{r}^4 \rangle / R^4 = 5/3 \neq 1$ as the rigid constraints would require.

In the limit as $R \to \infty$, we find $\langle \mathbf{u}(0) \cdot \mathbf{u}(L) \rangle \to e^{-3L/2l_p}$ in eq. 5.12, which is not the expected limit of the bending correlation function. We would instead expect the bending correlations for surface confinement to converge on the two dimensional

solution ($\langle \mathbf{u}(0) \cdot \mathbf{u}(L) \rangle = e^{-L/2l_p}$ in two dimensions) for large radius. This suggests a mean field persistence length, l_0, much like in the unconfined theory of Thirumalai and Ha [41, 42]. Substitution of $l_p = 3l_0$ into eq. 5.12 results in the expected limits of the correlation function as $R \to \infty$. Thirumalai and Ha found a similar result for a three dimensional unconfined WLC, $l_p = 3l_0/2$. Their argument was that, since additional forbidden configurations have been allowed by replacing the δ functions in eq. 5.1 with Gaussians, the mean field persistence length is smaller than the true persistence length. In the confined theory, we allow three dimensional configurations by replacing the confining δ functions with Gaussians, which would be forbidden by the surface confinement in addition to relaxing the rigid inter-monomer constraints. For this reason, we would expect more permitted configurations, so l_p/l_0 in the confined case should be greater than that of the unconfined case. In practice, l_p is often determined by fitting experimental or simulation data to a suitable polymer model. Here, the renormalization of l_p within the mean field theory is not a serious concern. With these definitions for the mean field persistence length, we find $\Omega_S = \sqrt{1 - 16l_0^2/R^2}$ and the decay length of the correlations becomes $\zeta_S = 1/4l_p$. After substitution into eq. 5.12, we find that the *exact* calculation of Spakowitz and Wang [120] is reproduced, valid for all values of L, l_0, and R. The ability to calculate these averages exactly shows the power of the mean field method.

We can determine the Free Energy and pressure for a WLC confined to the surface of a sphere (which does not require the substitution of $l_p = 3l_0$),

$$\beta F \sim \mathcal{F} = \frac{9L}{8l_p} + \frac{Ll_p}{2R^2} + \text{const} \qquad \beta PV \sim \frac{Ll_p}{3R^2} \tag{5.13}$$

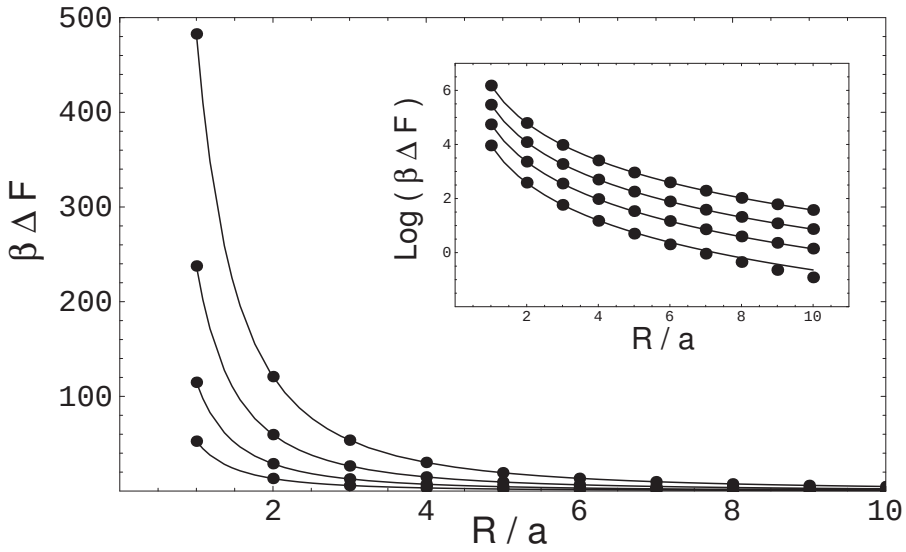

Figure 5.3: Free energy $\beta\Delta F = \beta F(R) - \beta F(\infty)$ as a function of R for a surface-confined wormlike chain. Shown are l_p/a=20, 10, 5 and 2. The inset shows $\log(\beta\Delta F)$ as a function of R, displaying the good agreement even for large R.

where the pressure is computed from $\beta P = -\partial\mathcal{F}/\partial V$. This is identical to the scaling predicted by Odijk for a tightly bent WLC [117] as is expected. Additionally, the coefficient of the scaling law agrees with that predicted by Mondescu and Muthukumar [119] for the Freely Jointed Chain. The accuracy of the correlation functions in eq. 5.12 have already been explicitly shown using Monte Carlo simulations [121]. However, it is of interest to show that the coefficients of the free energy determined by the mean field method agree well with simulations. We use the Configurational Bias Monte Carlo (CBMC) method [156] to determine the free energy of a surface confined wormlike chain with $L = 50a$, for various values of l_p and R. The theoretical curves for $\beta F \sim Ll_p/2R^2 + const$ are accurate to within $\sim$5%.

5.3 Surface Confined Stiff Chains under Tension

The true utility of the mean field method is in its ability to study the effect of additional potentials in problems involving confined WLC's with relatvie ease. In this section, we will study the effect of an external tension, $\mathbf{f}$, on a surface confined WLC. For the free chain, the Mean Field method has been shown [153] to give excellent agreement previous theoretical and experimental results [1]. Such a calculation for the surface-confined WLC will also verify that the mean field method satisfies the confinement on average, even under the extreme situation of a strong pulling force.

The distribution in phase space of a confined WLC under tension can be written as $\Psi_s(\mathbf{f}) = \Psi_s \exp[-\beta \mathbf{f} \cdot (\mathbf{r}_1 - \mathbf{r}_{N+1})]$, with $\beta = 1/k_B T$ and Ψ_S given in eq. 5.1. Because the external tension does not induce an energetic term quadratic in $\{\mathbf{r}_n\}$, the mean field theory in the previous section can be used with little change. The discrete free energy functional can be written as $\mathcal{F} = \mathcal{F}_x + \mathcal{F}_y + \mathcal{F}_z - \beta f(z_{N+1} - z_1) - a\sum_n(\lambda_n + k_n)$, with $\mathcal{F}_x$ given in eq. 5.2 ($\mathcal{F}_y$ and $\mathcal{F}_z$ are defined similarly), and we have taken $\mathbf{f} = f\hat{\mathbf{z}}$. None of the terms involving k_n or λ_n are altered with the addition of force, and we can again rewrite the quadratic terms of the Hamiltonian in terms of a symmetric, tridiagonal matrix $\mathbf{Q}$ (explicitly given in Appendix D, eqs D.2-D.4). We again make the substitution $k_1 = k_N = k + \gamma_1/a - \gamma_2/R$, $k_2 = k_{N-1} = k + \gamma_2/R$, $\lambda_1 = \lambda_{N-1} = \lambda + \delta/a - a\gamma_2^2/R^2$, with $k_n = k$ and $\lambda_n = \lambda$ otherwise. In the continuum limit ($N \to \infty$, $a \to 0$, and $Na \to L$), we find

$$e^{-\mathcal{F}} \quad = \quad e^{-\mathcal{F}_x - \mathcal{F}_y - \mathcal{F}_z - \beta f(z_L - z_0) + \lambda L + kL + 2\delta + 2\gamma_1}$$

$$= \left(\int d^4\mathbf{x}\, Z(\mathbf{x}) \right)^2 \left(\int d^4\mathbf{z}\, Z(\mathbf{z})\, e^{-\beta f(z_L - z_0)} \right) e^{\lambda L + \kappa L + 2\delta + 2\gamma_1}, \quad (5.14)$$

with $Z(\mathbf{x})$ is given in eq 5.8. The mean field free energy can be evaluated, yielding

$$\mathcal{F}(\mathbf{f}) = \mathcal{F}(0) + \frac{(R\beta f)^2}{2}\left[\left(\mathbf{M}^{-1} \right)_{11} - \left(\mathbf{M}^{-1} \right)_{12} \right] \quad (5.15)$$

where $\mathcal{F}(0)$ is the free energy at $f = 0$ (eq. 5.8), and we have used

$$\int d^N\mathbf{x}\ \exp\Big(-\mathbf{x}\cdot\mathbf{M}\mathbf{x} + \mathbf{v}\cdot\mathbf{x} \Big) = \left(\frac{\pi^N}{\det(\mathbf{M})} \right)^{\frac{1}{2}} \exp\left(\frac{1}{4}\mathbf{v}\cdot\mathbf{M}^{-1}\mathbf{v} \right)$$

Under the assumption that $L\omega_i \gg 1$ (a strongly confined chain, see eq. 5.7), it is not difficult to show that the solutions to the mean field equations (eq. 5.9) become

$$\lambda = \frac{9}{8l_p} - \frac{l_p}{R^2} \qquad k = \frac{l_p}{2R^2} \qquad \delta = \frac{3}{4} \qquad \gamma_1 = \frac{3}{4}\sqrt{1 + \frac{4(\beta f R)^2}{9}} \qquad \gamma_2 = -\frac{l_p}{2R} \quad (5.16)$$

Under the application of a force, only the behavior of the endpoints of a strongly confined chain is effected, reflected in the fact that only γ_1 depends on f. The interior monomer behavior should be insensitive to the force, so it is not surprising that λ and k are independent of the force. The extension as a function of the external tension can be computed via $\langle \mathbf{R} \rangle = -\partial\mathcal{F}/\partial(\beta\mathbf{f})$. For a strongly confined chain, we find

$$\frac{\langle z_L - z_0 \rangle}{R} = \frac{1}{3} R\beta f\ \frac{\langle \mathbf{R}^2 \rangle_0}{R^2}\left(1 - \frac{\langle \mathbf{R}_0^2 \rangle}{4R^2}\left[1 - \sqrt{1 + 4(R\beta f)^2/9} \right] \right)^{-1}, \quad (5.17)$$

with $\langle \mathbf{R}^2 \rangle_0$ the average end-to-end distance with $f = 0$ (we note that $\langle \mathbf{R}^2 \rangle_0 = 2R^2(1 - \langle \mathbf{r}_0 \cdot \mathbf{r}_L \rangle) \neq 0$ for all non-vanishing L, see eq. 5.12). While the force-extension curves for a confined WLC increase monotonically as a function of f, the

system has rather complicated behavior as a function of R. In Fig. 5.4, we see the extension of the chain as a function of R is oscillatory for small R, due to the non-monotonic values of $\langle \mathbf{R}^2 \rangle_0$.

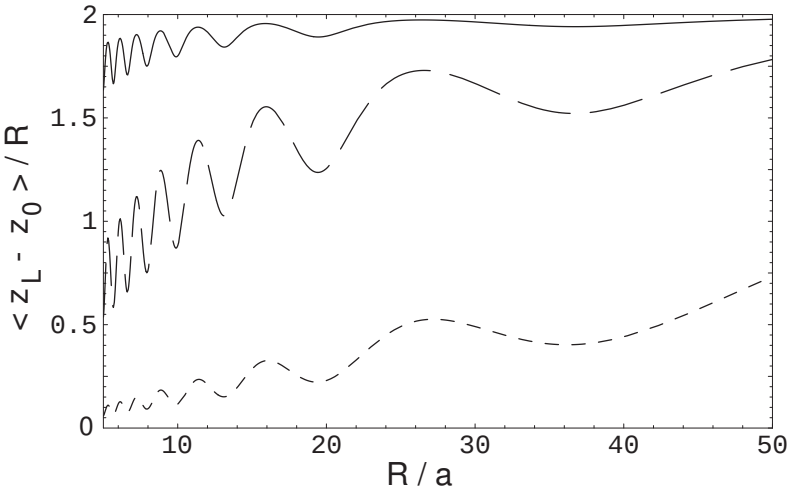

Figure 5.4: Linear extension under a constant external tension as a function of the radius. For all curves, $L = 250a$ and $l_p = 50a$. The applied tensions are $a\beta f = 10$ (solid), 1 (dashed) and 0.1 (dotted).

Eq. 5.17 has the asymptotic limits

$$\langle z_L - z_0 \rangle \sim \begin{cases} \beta f \langle \mathbf{R}^2 \rangle_0 / 3 & R\beta f \ll 1 \\ 2R - \frac{3}{\beta f}(4\frac{R^2}{\langle \mathbf{R}^2 \rangle_0} - 1) + O(f^{-2}) & R\beta f \gg 1 \end{cases} \tag{5.18}$$

The extension in the low force limit is identical to the expression found for a Gaussian chain under tension, with $\langle \mathbf{R} \cdot \hat{\mathbf{z}} \rangle^{(G)} = Na^2\beta f/3 = \beta f \langle \mathbf{R}^2 \rangle_0^{(G)}/3$. The large force limit has the expected limiting extension of $z_L - z_0 = 2R$ for a chain that is confined to the surface of a sphere. Surprisingly, though, the scaling of $\langle z_L - z_0 \rangle - 2R \sim f^{-1}$ differs from the scaling of the unconfined chain in the high force limit, $\langle z_L - z_0 \rangle - L \sim f^{-1/2}$. The change in the large-force scaling laws is linked to the fact that only the endpoints are affected by the force for a surface confined chain. In interior, the extension of

the chain comes about by alignment of all bonds with the force axis. However, when confined to the surface of the sphere, the extension occurs primarily by to the translation of the endpoints to the poles of the sphere, rather than a global realignment in the bond vectors. This is reflected in the fact that γ_1, which controls the position of the endpoints, is the only mean field parameter dependent on f. We note as well that the $f^{-1/2}$ scaling is seen in the unconfined mean field theory, and comes about due to the fact that λ (which determines the behavior of all of the bonds) becomes a function of f.

It is also possible to numerically solve the mean field equations for small L/R, and at very high external tensions we find

$$\langle z_L - z_0 \rangle \approx \begin{cases} 2R\sin(L/2R) & L \leq \pi R \\ 2R & L > \pi R \end{cases} \tag{5.19}$$

This is the exact end-to-end distance of a fully stretched chain confined to the surface of a sphere. The mean field method thus satisfies the confining constraints on average, even under extreme pulling, and again predicts the lower moments exactly.

We can also determine the Free Energy of a confined WLC under tension:

$$\beta F = \frac{9L}{8l_p} + \frac{Ll_p}{2R^2} - 3\sqrt{1 + \frac{4(\beta f R)^2}{9}} + 3\log\left[1 + \sqrt{1 + \frac{4(\beta f R)^2}{9}}\ \right] + \text{const}, \tag{5.20}$$

where we have neglected terms of order $O(e^{-3L/4l_p})$. The force-dependent terms in the free energy are not extensive because the tension only strongly effects the endpoints of the chain. βF becomes tension-dominated when f exceeds a critical force $R\beta f_c \sim Ll_p/4R^2$. Because the only force-scale in the problem is $\beta PA \sim$

Ll_p/R^2, with A the surface area of the sphere, the scaling of this critical force is expected. We expect the leading coefficient to be correct, due to the accuracy of our expression for the free energy of a WLC without the external tension (see Fig. 5.3). It It is amusing to estimate f_c for a strand of DNA wrapped around a histone [116, 24, 26], with $l_p \approx 50$nm, $L \approx 43$nm, and $R \approx 4$nm. We find the tension dominates the free energy when $f > f_c \approx 34$pN, which is significantly larger than the force required at each unwrapping event seen in single molecule experiments on histones [24]. However, as it has been observed that the tilting of the histone with respect to the force axis is of great importance when determining the behavior of the system [116], which the mean field theory does not take into account, we can view our result only as an upper bound on the unravelling force.

5.4 Wormlike chains confined to the interior of a sphere

The mean field theory for computing the average properties of a surface confined chain can be extended to studying the effects of volume confinement. The distribution in phase space of a WLC confined to the interior of a sphere is

$$\Psi_V(\{\mathbf{r}_n\}) \propto \prod_n \Theta(R^2 - \mathbf{r}_n^2)\ \delta(\Delta\mathbf{r}_n^2 - a^2)\ e^{-l_p(\Delta\mathbf{r}_{n+1} - \Delta\mathbf{r}_n)^2/a^3} \tag{5.21}$$

where $\Theta(x)$ is the Heaviside step function, ensuring that each monomer is contained within the sphere. The other two terms in eq. 5.21 are identical to the ones in Ψ_S (eq. 5.1). The similarities between the two distributions suggest that volume confinement can be treated on the mean field level as well. Two configurations for a volume confined WLC are shown in Fig. 5.5 for $R = 8a$ for two different values

of the persistence length ($l_p/a = 20$ and 100), and are qualitatively similar to the structures seen for the surface confined chains (Fig. 5.1). Unfortunately, the Θ function in eq. 5.21, ensuring the chain is within the interior of the sphere of radius R, can not be dealt with as simply as the δ functions in eq. 5.1 at the mean field level. It is not difficult to show that, for a single particle confined within a sphere, simply minimizing the Fourier Transform of the Θ function does not give the correct value of $\langle \mathbf{r}^2 \rangle$. However, the similarities between eq. 5.21 and eq. 5.1 suggest that we can formally write

$$\Psi_V \propto \int \prod_n dk_n d\lambda_n \exp\Big[-\frac{1}{2} a l_p \frac{(\Delta \mathbf{r}_{n+1} - \Delta \mathbf{r}_n)^2}{a^4} - a\lambda_n \Big(\frac{\Delta \mathbf{r}_n^2}{a^2} - 1 \Big) - a k_n \frac{\mathbf{r}_n^2}{R^2} - g(ak_n) \Big], \qquad (5.22)$$

where g is an undetermined function, chosen such that free energy minimization will satisfy the constraints ($\mathbf{u}^2(s) = a^2$ and $\mathbf{r}^2(s) \leq R^2$) on average. This form is identical to the surface confinement case, except for g. Because of this, we can immediately see that the same substitution of interior (i.e. $k_n \to k$ and $\lambda_n \to \lambda$) and endpoint terms will satisfy the symmetry of the $\mathbf{Q}$ (see Appendix D, and eqs D.2-D.4). This allows the problem of volume confinement in the continuum limit to be written in terms of the mean field variables λ, k, δ, γ_1, and γ_2, with the Free Energy expressible as $\mathcal{F} = \mathcal{F}_x + \mathcal{F}_y + \mathcal{F}_z - G[\lambda, k, \delta, \gamma_1, \gamma_2]$. F_x is defined in eq. 5.5, and G constrains the minimization of $\mathcal{F}$ (i.e. contains the as yet undetermined Lagrange multipliers).

The treatment of volume confinement at the mean field level is more difficult than the case of surface confinement for a number of reasons. In the case of surface

Figure 5.5: Two configurations for a stiff chain confined to the interior of a sphere. Shown are $L = 200a$, $R = 8a$, and $l_p/a = 20$ (a) and 100 (b).

confinement, we replaced the strict constraint of $\mathbf{r}^2(s) \equiv R$ with the global constraint $\frac{1}{L}\int_0^L ds\langle \mathbf{r}^2(s)\rangle = R^2$. It was possible to show, however, that the mean field solution implied $\langle \mathbf{r}^2(s)\rangle = R^2$ for all s under surface confinement. A similar condition will hold for the volume confined chain, with

$$\frac{1}{L}\int_0^L ds \frac{\langle \mathbf{r}^2(s)\rangle}{R^2} \equiv \rho, \tag{5.23}$$

where ρR^2 is the average monomer position (with the average taken over both configurational space as well as along the chain). ρ can not be determined from the mean field theory itself, and must be supplied using some other method (see below). Eq. 5.23 is equivalent to the requirement at the mean field level

$$\frac{\partial \mathcal{F}}{\partial k} = L\rho \tag{5.24}$$

(see eq. 5.3). The average position of the endpoints within the sphere need not be identical to the average position for interior points of the chain, i.e. $\langle \mathbf{r}^2(s)\rangle \neq$ const

for volume confinement. At the mean field level, this can be treated approximately by the restriction

$$\frac{\langle \mathbf{r}_0^2 \rangle}{R^2} = \frac{\langle \mathbf{r}_L^2 \rangle}{R^2} \equiv \rho_0, \tag{5.25}$$

with $\rho_0 \neq \rho$. This is implemented at the mean field level using

$$\frac{\partial \mathcal{F}}{\partial \gamma_1} = 2\rho_0 \tag{5.26}$$

(see eq. 5.4). Much like ρ, we can not determine ρ_0 theoretically, and additional information is required to determine it. Because $\mathbf{u}(s)$ need not be perpendicular to $\hat{\mathbf{r}}$, $\langle \mathbf{u}(s) \cdot \mathbf{r}(s) \rangle \neq 0$, which must be accounted for at the mean field level as well. There is an inherent asymmetry in the quantity $\mathbf{r}(s) \cdot \mathbf{u}(s)$, that can be seen as follows: If we simply relabel the monomers, such that $s \to L - s$, we find $\mathbf{r}_0 \to \mathbf{r}_L$ and $\mathbf{u}_0 \to -\mathbf{u}_L$ (since $\mathbf{u}(s) = \partial \mathbf{r}(s)/\partial s$). As the average behavior of the endpoints is independent of the labeling, this immediately implies that $\langle \mathbf{u}_0 \cdot \mathbf{r}_0 \rangle = -\langle \mathbf{u}_L \cdot \mathbf{r}_L \rangle$. We then restrict

$$\frac{\langle \mathbf{u}_L \cdot \mathbf{r}_L \rangle}{R} = -\frac{\langle \mathbf{u}_0 \cdot \mathbf{r}_0 \rangle}{R} \equiv \rho_c. \tag{5.27}$$

ρ_c represents the average correlation between the position and the bending at the endpoints of the chain. If the endpoints of a confined chain are found near the wall of the sphere, the direction of the bond vectors at the endpoint will be restricted, pointing away from the wall of the sphere and giving $\langle \mathbf{r}_L \cdot \mathbf{u}_L \rangle > 0$. For the free chain, we expect $\rho_c(R \to \infty) = 0$, since there is no restriction on the direction of the bending. Eq. 5.27 is expressed at the mean field level as

$$\frac{\partial \mathcal{F}}{\partial \gamma_2} = -4\rho_c \tag{5.28}$$

(see eq. 5.4). The lagrange multipliers for both λ and δ remain unchanged at the mean field level, with $\partial\mathcal{F}/\partial\lambda = L$ and $\partial\mathcal{F}/\partial\delta = 2$ (see eq. 5.8).

Up to the three undetermined mean field parameters (ρ, ρ_0, and ρ_c), we can write the mean field free energy for volume confinement as

$$\mathcal{F}_V = \mathcal{F}_x + \mathcal{F}_y + \mathcal{F}_z - \lambda L - 2\delta - \rho k L - 2\rho_0\gamma_1 + 4\rho_c\gamma_2 \tag{5.29}$$

where $\mathcal{F}_x$ is identical to the one dimensional free energy functional for the surface case (eq 5.5). The three mean field parameters ρ, ρ_0, and ρ_c can not be determined using the mean field theory, and are found using simulations. We use Langevin dynamics simulations to determine the equilibrium behavior of a WLC confined to the interior of a sphere. The details of our simulations are given in Appendix D.3.

The mean field equations for volume confinement, given in eq. 5.9, are quite similar to the surface confinement equations, and can be solved in the limit of strong confinement (i.e. large $L\omega_i$, see eq. 5.7). We find

$$\begin{gathered}\lambda = \frac{9}{8l_p} - \frac{l_p}{\rho R^2} \qquad k = \frac{l_p}{2\rho^2 R^2} \qquad \delta = \frac{3(\rho_0 + \rho_c^2)}{4(\rho_0 - \rho_c^2)} \\ \gamma_1 = -\frac{3}{4\rho} + \frac{3}{2(\rho_0 - \rho_c^2)} \qquad \gamma_2 = -\frac{l_p}{2\rho R} - \frac{3\rho_c}{2(\rho_0 - \rho_c^2)}\end{gathered} \tag{5.30}$$

which gives $\omega_i = 3/4l_p(1 \pm \sqrt{1 - 16l_p^2/9\rho R^2})$. The solutions for the ω_i's, which define the average behavior over the entire length of the chain, are identical to those found for a wormlike chain confined to the surface of a sphere of radius $\sqrt{\rho}\,R$ (see eq. 5.11). However, the endpoint terms differ from the surface confined system, allowing for fluctuations in the average monomer positions near the ends of the chain.

In Figure 5.6, we show the average monomer positions as a function of s for varying R and l_p. Fig. 5.6a shows that "interior" monomer behavior (where

$\langle \mathbf{r}^2(s) \rangle \approx$ const) begins to emerge within the range $2R \geq s \geq L - 2R$, with endpoint effects clearly dominating the behavior over the range $2R$ on either end. In Fig. 5.6b, we see that increasing the persistence length of the chain while keeping R fixed changes the values of ρ and ρ_0 (reflected in the increase of $\langle \mathbf{r}^2(s) \rangle_{int}/R^2$, $\langle \mathbf{r}_0^2 \rangle/R^2$, and $\langle \mathbf{r}_L^2 \rangle/R^2$) but does not significantly alter the qualitative behavior of $\langle \mathbf{r}^2(s) \rangle$ as a function of s. We note that the average position of the interior monomers is approximately constant (i.e. $\langle \mathbf{r}^2(s) \rangle_{int} \approx \rho R^2$, consistent with the fact that the ω_i's suggest confinement to a sphere of radius $\sqrt{\rho}\ R$.

The range of the endpoint effect makes physical sense: if $\mathbf{r}_0$ is near the boundary of the sphere, the bending energy near the endpoint will be lower if $\mathbf{u}_0$ is directed towards the center of the sphere, as opposed to being directed towards the wall. This implies that monomers near the endpoints will be found closer to the origin of the sphere on average, as is seen in Fig. 5.6. This endpoint effect will dominate the behavior of the chain until another monomer comes into contact with the wall, a distance of at most $2R$.

We perform a number of simulations in order to determine the mean field parameters ρ, ρ_0 and ρ_c, varying L, l_p, and R. We find that, for large L, there is virtually no variation in any of the mean field parameters. Since ρ determines the effective surface confinement (see Fig. 5.6b), increasing L does not change ρ, as the chain simply wraps further around the effective surface at $\sqrt{\rho}\ R$. ρ_0 and ρ_c are likewise independent of the length of the chain, as they dominate the behavior of the endpoints, and have only a strong effect over a length $2R$ along the chain (see Fig. 5.6a). Since the only remaining length scales are l_p and R, we expect that all

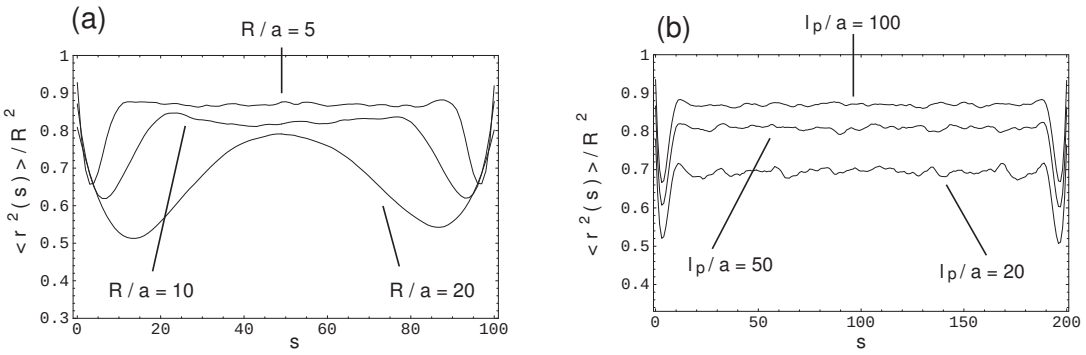

Figure 5.6: $\langle \mathbf{r}^2(s) \rangle$ vs. s. (a): $L = 100a$ and $l_p = 100a$, with (from highest to lowest) R/a=5, 10, and 20. The average monomer position is dominated by endpoint effects for $s \leq 2R$ and $s \geq L - 2R$. (b): $L = 200a$ and $R = 5a$, with (from highest to lowest) l_p/a=100, 50, and 20. With R fixed, variations in l_p change only the value of ρ, but does not alter the behavior of the monomers.

of the mean field parameters depend only on the ratio l_p/R. In Fig. 5.7, we show the behavior of the three mean field parameters (ρ, ρ_0, and ρ_c) as a function of l_p/R, determined from a simulation with $L = 80a$ and $R = 5a$. Changing the values of L or R do not change the measured values of ρ, ρ_0 and ρ_c, as long as l_p/R is held fixed, as expected (data not shown). We find for long, stiff chains (with $L/R \gg 1$ and $l_p/R \gg 1$) that $\rho \leq 0.9$, $\rho_0 \leq 0.8$, and $\rho_c \leq 0.5$. Most physical systems of interest (the viral packing of DNA, for example) will be in this strongly confined regime.

Another difference between surface and volume confinement is the relationship between the true and mean field persistence lengths. In the limit as $R \to \infty$, we expect to recover the unconfined three dimensional correlation function $\langle \mathbf{u}_0 \cdot \mathbf{u}_L \rangle = e^{-L/l_p}$. For the interior monomers (where the system is confined approximately to a sphere of radius $\sqrt{\rho}\, R$, see Fig 5.6), we can determine the bending correlation function as we did for the surface confined chain, which gives $\langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle \sim e^{-3|\Delta s|/2l_p}$ in the limit of $R \to \infty$. This only converges to the correct correlation function if

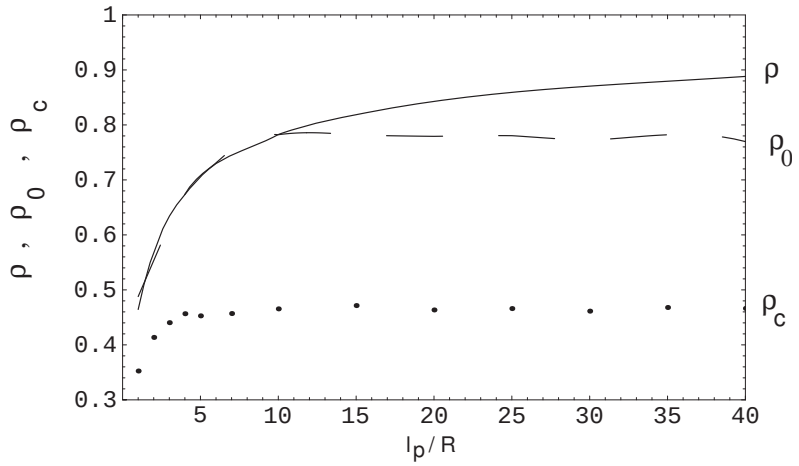

Figure 5.7: The mean field parameters ρ, ρ_0, and ρ_c as a function of R. ρ (solid line) defines the radius of the effective surface confinement for the interior monomers (eq. 5.23). ρ_0 (dashed line) is the average scaled endpoint distance (eq. 5.25). ρ_c (dotted line) gives the correlation between the position and bond vector at the endpoints (eq. 5.27).

the volume-confined mean field persistence length is $l_0^V = 3l_p/2$ for large R, the same relation between the l_p and l_0 in the unconfined theory [41, 42]. However, because the system is approximately confined to the surface of a sphere of radius $\sqrt{\rho}R$, and the surface-confined mean field persistence length is $l_0^S = 3l_p \neq l_0^V$, the relationship between l_0 and l_p is R-dependent. The ratio l_0/l_p to be independent of L for large L, since we have seen that the development of near-surface confinement depends only on the raito l_p/R (Fig 5.7). We then define $l_0 = \alpha(l_p/R)l_p$, and expect $3/2 \leq \alpha(l_p/R) \leq 3$ for all l_p/R, i.e. a wormlike chain confined to the interior of a sphere will behave somewhere in between a free wormlike chain in 3 dimensions, and a surface confined wormlike chain. This gives the correlation functions for the interior monomers (i.e. $s, s' > 2R$ and $s, s' < L - 2R$),

$$\langle \mathbf{r}(s) \cdot \mathbf{r}(s') \rangle \;=\; \rho R^2 e^{-|\Delta s|/\zeta_V} \Big[\cosh\left(\frac{|\Delta s|}{\zeta_V}\Omega_V\right) + \frac{1}{\Omega_V} \sinh\left(\frac{|\Delta s|}{\zeta_V}\Omega_V\right) \Big]$$

$$\langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle \;=\; e^{-|\Delta s|/\zeta_V} \Big[\cosh\Big(\frac{|\Delta s|}{\zeta_V} \Omega_V \Big) - \frac{1}{\Omega_V} \sinh\Big(\frac{|\Delta s|}{\zeta_V} \Omega_V \Big) \Big], \qquad (5.31)$$

with $\zeta_V = 4\alpha l_p/3$ and $\Omega_V = \sqrt{1 - 16\alpha^2 l_p^2/9\rho R^2}$. Near the endpoints, the correlation functions become more complicated, due to the dependence of the behavior of the endpoints on ρ_0 and ρ_c.

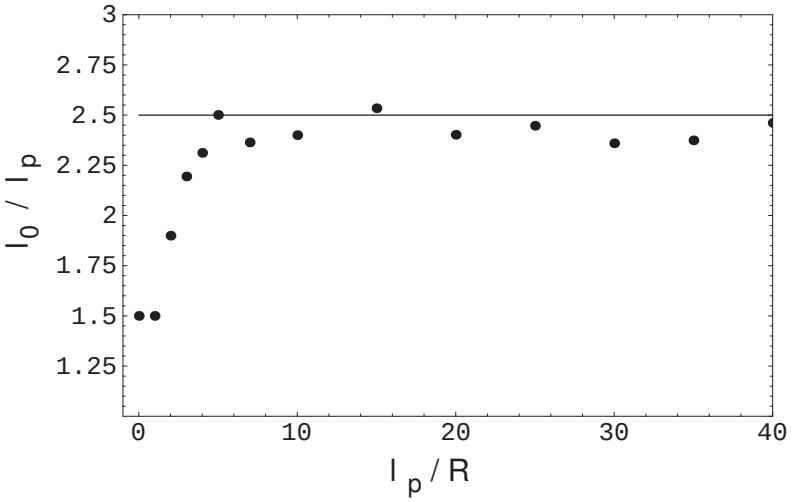

Figure 5.8: $\alpha(l_p/R) = l_0/l_p$ as a function of l_p/R. Points are the results of simulations, the line is the approximate limiting value for strong confinement ($l_p/R \gg 1$) of $\alpha \approx 2.48 \approx 5/2$.

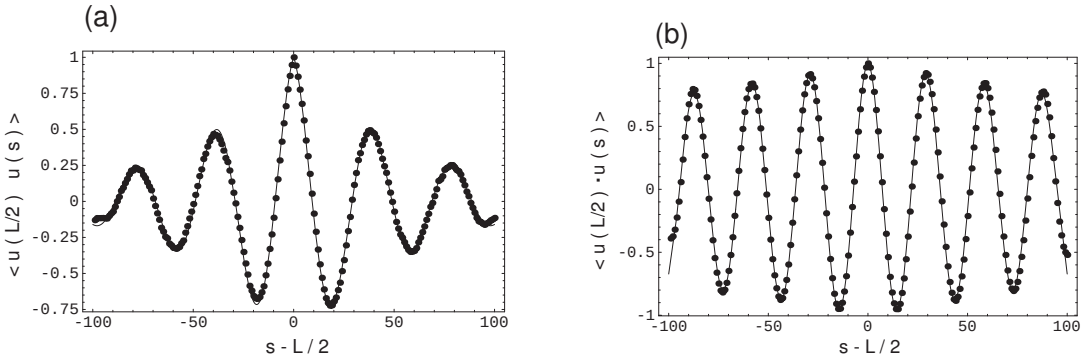

Figure 5.9: Bending correlation function $\langle \mathbf{u}(L/2) \cdot \mathbf{u}(s) \rangle$ as a function of s for a chain with $L = 200a$. The points are simulation data, the solid lines are the theoretical results in eq. 5.31. (a) has $l_p = 20a$ and $R = 8a$, and (b) has $l_p = 100a$ and $R = 5$.

In order to determine $l_0/l_p = \alpha(l_p/R)$, we must again turn to our simulation

results. We perform a number of low friction Langevin Dynamics simulations (see Appendix D.3 for details) with $L = 80a$ and $R = 5a$, with varying l_p. The bending correlation function is fit using eq. 5.31, with only α as a fitting parameter. The resulting values are shown in Fig. 5.8. We find that α does indeed vary with only l_p/R (changing R does not change α as long as l_p/R is held constant), and the saturating value becomes $\alpha(\infty) \approx 5/2$. It is not surprising that α never reaches the surface-confined value of $\alpha = 3$: since $\rho < 1$ for all values of l_p/R, the conformational space available for the chain is larger than a truly surface-confined chain. The decay length $\zeta_V = 4\alpha l_p/3 \approx 3.3 l_p$, which is the length scale over which the bending and positions become decorrelated, is strictly less than than the decay length for surface confinement, $\zeta_S = 4l_p$, due to the larger number of configurations that are available to volume confined chains. Using the values of α (Fig 5.8) and ρ (Fig. 5.7) determined from the $N = 80$ simulations, we compare the bending correlation functions for a simulation with $L = 200a$ to the results of eq. 5.31, shown in Fig. 5.9. The agreement between simulation and theory is excellent with no additional fitting parameters, and the agreement the position correlation function is equally as good (data not shown).

Further information about the structure of a stiff chain confined to a sphere can be determined using the local winding axis of the chain. The unit local winding axis of bonds i and $i+1$ (the axis about which $\mathbf{u}_i$ and $\mathbf{u}_{i+1}$ wind) is given by [121] $\hat{\mathbf{a}}_i = \mathbf{a}_i/|\mathbf{a}_i|$, with

$$\mathbf{a}_i = \mathbf{u}_i \times \mathbf{u}_{i-1} \ , \tag{5.32}$$

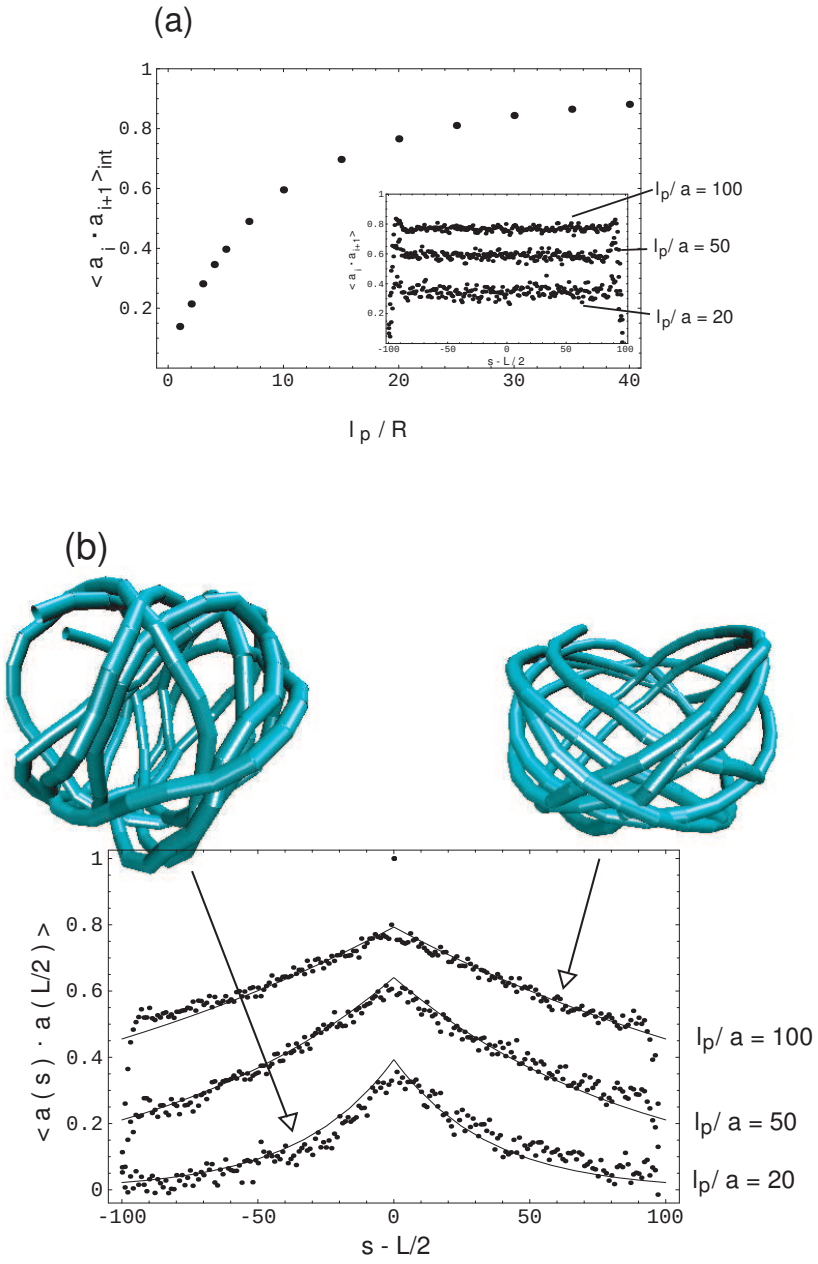

Figure 5.10: (a) Average correlations in the nearest neighbor winding axis as a function of l_p/R. The average is taken over interior points only (i.e. $2R \leq s \leq L - 2R$). The inset shows the average nearest neighbor correlation for $L = 200a$, $R = 5a$, and $l_p/a = 100$, 50, and 20. The correlations drop sharply near the endpoints. (b) $\langle \hat{\mathbf{a}}(L/2) \cdot \hat{\mathbf{a}}(s) \rangle$ as a function of s for $L = 200a$ and $R = 5$. From highest to lowest, l_p/a=100,50, and 20. The points are simulation data, the lines the theoretical curves, $\langle \hat{\mathbf{a}}(s) \cdot \hat{\mathbf{a}}(s') \rangle \propto e^{-4|\Delta s|/3l_0}$. A typical structure for the cases $l_p = 100a$ and $l_p = 20a$ are shown on top; the clear correlation of the winding axis is evident for the stiffer chain.

Analytical work with the local winding axis is difficult, because

$$\hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} = \frac{\cos(\theta_{i-1,i+1})}{\sin(\theta_{i-1,i})\sin(\theta_{i,i+1})} - \cot(\theta_{i+1,i})\cot(\theta_{i-1,i}) \tag{5.33}$$

where we have defined $\cos(\theta_{i,j}) = \mathbf{u}_{i+1} \cdot \mathbf{u}_j$. The details of this result are shown in Appendix D.4. While directly computing the average of eq. 5.33 is intractable, the symmetry of the problem quickly shows that, for a free WLC, $\langle \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} \rangle = 0$ (since $\mathbf{u}_{i+1}$ may be freely rotated about the $\mathbf{u}_i$ axis without changing $\theta_{i,i+1}$). However, such a symmetry does not exist for confined chains, and we therefore restrict our study to the simulation results. We find that, for interior monomers, the local winding axes are highly correlated, with $\langle \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} \rangle \leq 0.9$ for strongly confined chains (Fig. 5.10a). Once again, variations in L and R do not alter $\langle \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} \rangle$, as long as l_p/R is held fixed. The endpoints of the chain are not strongly correlated to the interior behavior (Fig. 5.10a, inset), with a precipitous drop to $\langle \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} \rangle \leq 0.1$ at the endpoints. This sharp drop suggests that the endpoints of the chain behave more like an unconfined chain than do the interior monomers, which is consistent with our physical picture of the origin of the endpoint effects (see the discussion of ρ_c above, eq. 5.27). Correlations between the winding axes for interior monomers as a function of their separation Δs is exponentially distributed (Fig. 5.10b), with

$$\langle \hat{\mathbf{a}}(s) \cdot \hat{\mathbf{a}}(s') \rangle \approx \langle \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} \rangle e^{-4|\Delta s|/3l_0}. \tag{5.34}$$

In eq. 5.34, the decay length $3l_0/4 = 3\alpha l_p/4 \approx 1.9 l_p$ is determined by fitting the simulations with $L = 80a$, $R = 5a$, and varying l_p. The theoretical curves in Fig. 5.10b use this correlation length, with no additional fitting parameters. This shows that the length scale $l_0 = \alpha(l_p/R)l_p$ defines the length scale over which correlations

in the confined system decay. The sharp drop in the axis correlation function near the endpoints shows that the wrapping near the ends of the chain is uncorrelated to the interior wrapping. Forrey and Muthukumar [142] use $\langle \mathbf{a}_i \cdot \hat{\mathbf{z}} \rangle$as their order parameter in order to study the wrapping of DNA within the $\phi 29$ phage (with the DNA loaded into the capsid along the z-axis). The find that the local winding axes become decorrelated with the z-axis, as is expected due to the lack of correlations between the interior and the endpoints (Fig. 5.10).

The free energy and pressure of the volume confined WLC can be computed using our mean field roots (eq. 5.30). The expressions are somewhat more lengthy than in the surface case, due to the endpoint terms involving ρ_0 and ρ_c, but in the limit of small R (relevant for most physical systems) we find

$$\beta F \sim \frac{L l_p}{2\rho R^2} - \frac{2\rho_c l_p}{\rho R} - 3\log(R) + \text{const} \qquad \beta PV \sim \frac{L l_p}{3\rho R^2} - \frac{2\rho_c l_p}{3\rho R} + 1, \quad (5.35)$$

with $\beta P = -\partial \mathcal{F}/\partial V$. The R^{-1} terms in the free energy and pressure are not present in the surface confined case, and are due entirely to the nonuniformity in $\langle \mathbf{r}^2(s) \rangle$ as a function of s (see Fig. 5.6), reflected in the fact that this term is proportional to ρ_c. It is interesting to note that the L-independent, R^{-1} term is negative, which is due to the fact that portions of the chain near the endpoints will be found on average closer to the center of the sphere than the interior monomers (as seen in Fig 5.6), thus reducing their contribution to the pressure. In order to determine the pressure from the simulations, we compute

$$PA = \sum_i \mathbf{f}_{i \to wall} \cdot \hat{\mathbf{r}} \quad (5.36)$$

with $A = 4\pi R^2$ the surface area of the sphere, and $\mathbf{f}_{i \to wall}$ the force of the i^{th}

monomer on the wall. In Fig. 5.11, the simulated results along with the mean field expression for the pressure (of which eq. 5.35 is the limit of small R). We find the agreement is excellent for a large range of L, l_p and R, particularly for small R where endpoint effects are less important.

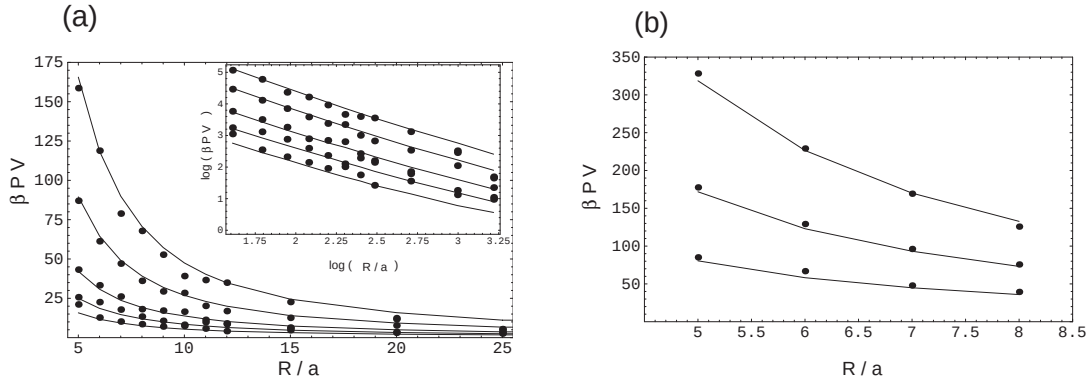

Figure 5.11: βPV as a function of R for varying l_p and L. In both, the dots are simulation results, and the solid line is the theoretical result with no fitting parameters. ρ, ρ_0 and ρ_c are taken directly from the simulation results for each value of L, R, and l_p. (a): $L = 100a$ and (from highest to lowest) l_p/a=100, 50, 20, 10, and 5. The inset is a log-log plot for this data. (b): $L = 200$ and (from highest to lowest), l_p/a=100, 50, and 20.

We can compare our results to the experimental pressures determined by Smith et. al [67], using the ϕ29 virus. The ϕ29 viral capsid is not spherical, with an icosohedral shell of radius ≈21nm and height ≈54nm, but has a volume equivalent to a sphere of radius ≈ 26nm. The fully packed virus contains a strand of DNA of length 6.6μm, with persistence length 50nm. If we neglect the excluded volume, electrostatic, and solvent-induced interactions of the DNA (a rather severe approximation), and take $\rho \approx 0.9$ (the saturating value of ρ, see Fig. 5.7), we find $P \approx 1$kPa ($=10^{-3}$pN/nm^2), almost 4 orders of magnitude lower than the 6MPa measured in the experiments. It is clear that the behavior of a strongly confined wormlike chain

is critically dependent on the intra-chain interactions, in agreement with a number of other studies [144, 27, 145, 28, 142, 148], with the contribution due to entropy making a negligible contribution to the pressure.

5.5 Confined Excluded Volume Chains

The effect of confinement on an excluded volume self-avoiding chain is difficult to predict analytically. For a non-interacting chain, we have shown that the relevant variable for virtually all of the properties of the chain is l_p/R. For an excluded volume chain, we expect a new variable to arise: the packing fraction $\phi = V_{chain}/V_{sphere}$. In order to study the problem of a confined, self-avoiding WLC, we resort only to simulations. The details of our simulation can be found in Appendix D.3. In our simulations, we take the chain to be a sequence of soft spheres of radius $a/2$, so that the volume fraction is $\phi = Na^3/8R^3$. A wide range of volume fractions are found in nature in the context of viral packing, with ϕ ranging from 0.07 to almost 0.5 [27], with genomes in the range $N \sim 10^4 - 10^6$. It is computationally intractable to simulate such long chains, and we will restrict ourselves here to $L = 200a$ and $R = 8a$ ($\phi \approx 0.05$). Higher packing fractions require significantly longer equilibration times, and (as seen below) this packing fraction allows us to see a clear development of structure. Such short chains do not allow for a direct comparison to any viral packaging problems, but is of interest in studying the structures of confined, self-avoiding WLC's.

In Fig. 5.12, show typical configurations for a self-avoiding WLC confined

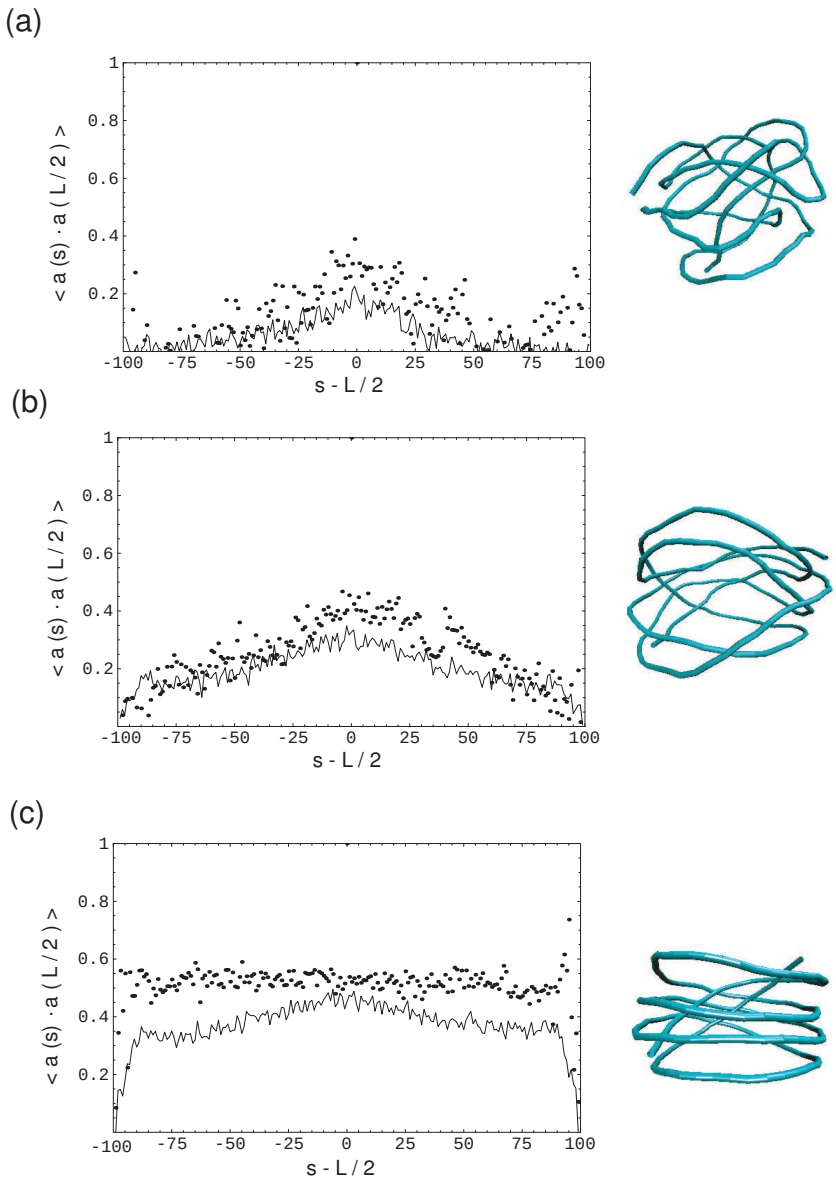

Figure 5.12: The corelations in the winding axis for a self avoiding WLC. In all figures, the dots are the correlations in the winding axes for the outer shell only. The lines are the correlations in the winding axes for the non-interacting WLC. Shown are $R = 8a$, $L = 200a$, and $l_p/a = 20$ (a), 50 (b), and 100 (c). In all cases, the intra-chain interaction increases the correlations of the winding axes, which becomes more clear for stiff chains. As the outer layer becomes more organized, the correlations in the winding axes become almost constant. Typical structures for each simulation are shown next to the figures. The development of the outer shell, and resulting correlations in the winding axis, is evident for increasingly stiff chains.

to the interior of a sphere. For reasonably flexible chains (Fig. 5.12a), there does is no clear development of structure in the confiugrations of the chain (compare to Fig. 5.5). As the stiffness increases, the excluded volume chain develops a clear, spool-like structure, as noted by many authors [29, 131, 136, 27, 28, 142]. For very stiff chains (Fig. 5.12c), the outer shell is highly correlated, and differs greatly from its non-interacting counterpart (compare to Fig. 5.5). The highly correlated nature of the outer shell is confirmed by considering the correlations in the winding axes for both the self-avoiding and non-interacting cases. In Fig. 5.12, we also see that intra-chain interactions tend to increase both the magnitude and the range of the winding axis correlations in the outer shell. As the chain becomes very stiff, the outer shell has an almost constant correlation in the winding axes, with $\langle \hat{\mathbf{a}}(s) \cdot \hat{\mathbf{a}}(s') \rangle_{outer} \approx \langle \hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} \rangle_{outer}$ for all s and s'. This is consistent with the simulations of Cerdà et. al. [121], which showed that flexible excluded volume chains confined to the surface of a sphere are disordered, but helicoidal structures develop for sufficiently stiff chains. Further simulations are required to fully understand the specifics of the onset of structure for self-avoiding WLC's under confinement.

5.6 Conclusions

We have shown that spherical confinement of WLC's can be well understood by applying the mean field theory of Ha and Thirumalai [41, 153, 42]. For a surface confined chain we can determine many average properties of the WLC, by replacing the rigid constraints of inextensibility ($\Delta \mathbf{r}_n^2 \equiv a^2$) and confinement ($\mathbf{r}_n^2 \equiv R^2$)

with average constraints. We have shown that the mean field approach reproduces the exact results of Spakowitz and Wang [120], and reproduces the correct scaling coefficient of the free energy of confinement. The mean field approach is also able to study the effects of additional potentials on confined WLC's. We determine the scaling and free energy of a surface confined WLC under tension, which may be of use in better understanding the wrapping of DNA around histones [116, 24]. The force-extension curve (FEC) for a strongly confined WLC is shown to differ greatly from the unconfined FEC, with a strikingly different method of extension.

We show that the mean field method can approximately determine the behavior of a WLC confined to the interior of a sphere. Interior monomers (far from the endpoints) to be approximately surface confined, with $\langle \mathbf{r}^2(s) \rangle \approx 0.9R^2$ for strongly confined chains, but endpoint effects dominate the behavior of the chain for $s < 2R$ or $s > L - 2R$. Structural information about the confined chain can be determined by examining the correlations in the local winding axis, and we find that strongly confined stiff chains are highly structured even without intra-chain interactions. The mean field estimates of the pressure due to confinement show that the extreme pressures inside of a viral capsid are not strongly dependent on simple confinement entropy, but depend critically on the intra-chain interactions. Simulations of short, self-avoiding WLC's show that the development of structure depends on the stiffness of the chain, and can again be understood on a qualitative level by examining the local winding axis. Interestingly, the development of structure for a volume confined WLC appears similar to the surface confined case, suggested by the effective surface confinement felt by the non-interacting chain.

Appendix A

Appendix for Chapter 2

A.1 The Center of Mass Average

In formulating the fluctuations of the end-to-end distance vector, $\langle \delta \mathbf{R}_{ee}^2 \rangle$, it is important to take into account the failings of the continuum model of the Freely Jointed Chain. A simple calculation of $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ with $\mathbf{R}_{ee}(t) = \mathbf{r}(N,t) - \mathbf{r}(0,t)$ as determined from eq (2.1) gives

$$\langle \delta \mathbf{R}_{ee}^2(t) \rangle = 16Nb^2 \sum_{n \text{ odd}} \frac{1}{n^2\pi^2}\left(1 - e^{-n^2 t/\tau_R}\right) \tag{A.1}$$

We will refer to this result as the standard analytic average. However, the non-physical boundary conditions imposed on the continuum representation, with $\partial \mathbf{r}/\partial s \equiv 0$ at the endpoints, will strongly affect the accuracy of this result.

To minimize the effect of the boundary conditions on averages involving the end-to-end distance, we compute averages with respect to the differences between the centers of mass of the first and last bonds, using

$$\mathbf{R}_{ee}(t) \approx \int_{N-1}^{N} ds\ \mathbf{r}(s,t) - \int_{0}^{1} ds\ \mathbf{r}(s,t). \tag{A.2}$$

We will refer to this as the center of mass average. Using this representation, $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ is given in eq (2.4).

In Fig. A.1, we compare the values of $D(t)$ obtained from $\langle \delta \mathbf{R}_{ee}^2(t) \rangle$ (in eq. (2.5)) for $N = 19$ and $b = 0.39$. In both cases, b is taken as a fitting parameter.

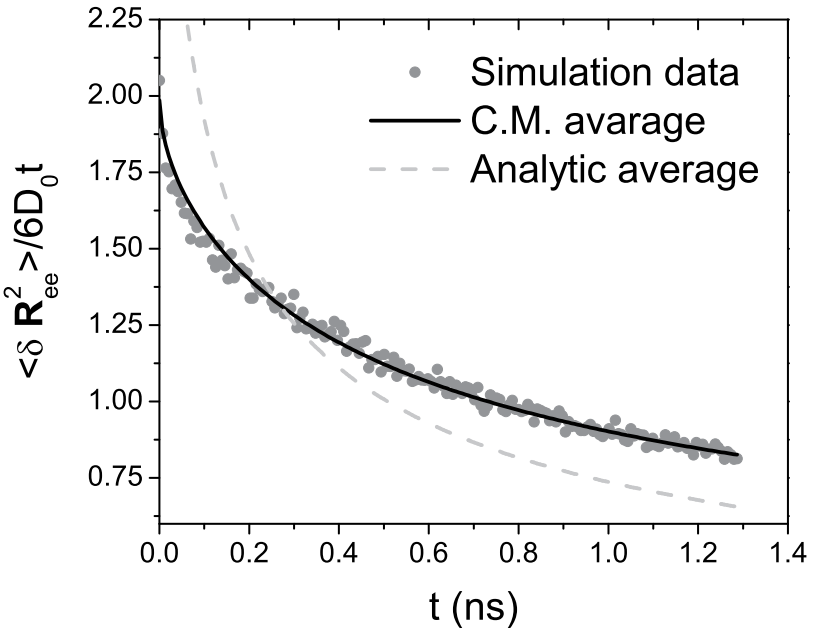

Figure A.1: Measured Diffusion Coefficient as a function of time for the Rouse chain with $N = 19$ and $b = 0.39$nm. Symbols are the simulation data, the dashed line (standard average) is obtained using eqs. (A.1) and (2.5) (with best fit $b \approx 0.26$nm), and the solid line is the center-of-mass average derived using eqs (2.4) and (2.5) (with best fit $b \approx 0.41$nm).

The center of mass average, which fits the data quite well, has a best fit of $b = 0.41$ (a difference of 5%), whereas the standard average does not give accurate results. For this reason, all averages involving $\mathbf{R}_{ee}$ are computed using the center of mass theory.

A.2 Simulation Details

In the case of the non-interacting chain, we model the connectivity of the chain using the Hamiltonian

$$\beta H = \frac{k_s}{2} \sum_{i=1}^{N} \left(1 - \frac{|\mathbf{r}_{i+1} - \mathbf{r}_i|}{b_0}\right)^2, \tag{A.3}$$

with $b_0 = 0.38$nm, and a spring constant $k_s = 100$. We note that $\langle(\mathbf{r}_{i+1} - \mathbf{r}_i)^2\rangle^{\frac{1}{2}} \approx 0.39$nm for this Hamiltonian, which we take as the Kuhn length b when fitting the data. For large N, the differences between the FJC and Rouse models are not relevant, and hence the scaling of τ_c with N for these two models should be identical. The microscopic diffusion coefficient was taken as $D_0 = 0.77\text{nm}^2/\text{ns}$. The equations of motion in the overdamped limit were integrated using the Brownian dynamics algorithm [18], with a time step of $\Delta t = 10^{-4}$ns. The end-to-end distribution $P(r)$ is easily computed for the model in eq. (A.3), giving the expression for large k_s

$$P(r) = 2r \int_0^\infty dq\, q \sin(qr) \left(\frac{b_0 q \cos(b_0 q) + k_s \sin(b_0 q)}{b_0 q(1 + k_s)} e^{-b_0^2 q^2/k_s^2}\right)^{N-1}, \tag{A.4}$$

which must be numerically integrated.

The Hamiltonian used in our simulations of a polymer in a good or poor solvent

is $H = H_{FENE} + H_{LJ}$, where

$$H_{FENE} = -\frac{kb^2}{2}\sum_{i=1}^{N}\log\left[1-\left(\frac{|\mathbf{r}_{i+1}-\mathbf{r}_i|-b}{R_0}\right)^2\right] \tag{A.5}$$

models the chain connectivity, with $k = 22.2k_BT$, and $b = 0.38$nm. The choice $R_0 = 2b/3$ (diverging at $|\mathbf{r}_{i+1}-\mathbf{r}_i| = b/3$ or $5b/3$) allowed for a larger timestep than using [18] $R_0 = b/2$, and increased the efficiency of conformational sampling. The interactions between monomers are modeled using the Lennard-Jones potential,

$$H_{LJ} = \epsilon_{LJ}\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\left[\left(\frac{b}{\mathbf{r}_{ij}}\right)^{12} - 2\left(\frac{b}{\mathbf{r}_{ij}}\right)^{6}\right], \tag{A.6}$$

with $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. The Lennard-Jones interaction between the covalently bonded beads $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ are neglected to avoid excessive repulsive forces.

In our simulations, we computed the mean first passage time directly. For the noninteracting chain, we generated the initial conditions by Monte Carlo equilibration. The chains in good or poor solvent were initially equilibrated using parallel tempering (replica exchange) Monte Carlo [32] to ensure proper equilibration, with each replica pertaining to one value of ϵ_{LJ}. Starting from each equilibrated initial configuration, the equations of motion were integrated until $|\mathbf{R}_{ee}| \leq a$ for the first time, with the first passage time computed for multiple values of N and a. The loop closure time τ_c was identified with the mean first passage time, obtained by averaging over 400 independent trajectories.

Appendix B

Appendix For Chapter 3

B.1 Simulation Details

To calculate the equilibrium FEC of a self-avoiding polymer, we performed low friction Langevin dynamics simulations using the Hamiltonian

$$\beta H = \frac{3}{2a^2}\sum_{i=1}^{N-1}(|\mathbf{r}_i - \mathbf{r}_j|^2 - a^2) + \sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\varepsilon\left(\frac{a}{|\mathbf{r}_{i+1} - \mathbf{r}_i|}\right)^{12} - \beta f(z_N - z_1), \quad \text{(B.1)}$$

with $a = 1$, $\varepsilon = 100$, and $N = 100$. We set $k_BT = 1/\beta = 1$ in the simulations. The first term in Eq. B.1 describes the chain connectivity in the extensible form that, in the continuum limit, becomes $3/2a^2 \int_0^N ds\, \dot{\mathbf{r}}^2(s)$. We model the excluded volume interactions between the monomers using a r^{-12} repulsion term (the second term in eq. B.1). Because of the large ε value, the summation does not include neighboring monomers (i and $i+1$) to avoid excessive repulsive forces. The last term in eq. B.1 denotes the potential due to tension acting on the ends of the polymer. Thus, this model can be viewed as the discrete representation of the Hamiltonian in Eq. 3.2. The natural time unit is $\tau_L = (ma^2/\varepsilon_h)^{1/2}$. We chose $\zeta = 0.05\tau_L^{-1}$ and $h = 0.002\tau_L$. To begin the simulations, we generate 200 initial random polymer conformations, and thermally equilibrate those structures for 5×10^6 h with $f = 0$. Subsequently, a constant force is applied in the z-direction to one end of each polymer, with the other end held in a fixed position. The force exerted is increased as $f_j = 10^{-3+0.1j}$

k_BT/a with $j = 1, 2, \ldots, 39$. The integer j is increased every 5×10^6 h. For each force step, we neglect the first 2×10^6 steps to ensure that the chain has equilibrated at f_j, and collect the statistics of polymer conformations every 10^4 integration time steps for the remaining time steps.

The simulation procedure used to study the stretching of a homopolymer in a poor solvent is identical to the one described in good solvent case, except for the Hamiltonian used. The Hamiltonian in a poor solvent is

$$\beta H = \frac{3}{2a^2}\sum_{i=1}^{N-1}(|\mathbf{r}_{i+1}-\mathbf{r}_i|^2-a^2)+\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\varepsilon\left[\left(\frac{a}{|\mathbf{r}_i-\mathbf{r}_j|}\right)^{12}-2\left(\frac{a}{|\mathbf{r}_i-\mathbf{r}_j|}\right)^{6}\right]-\beta f(z_N-z_1), \tag{B.2}$$

where $\varepsilon = 0.5$ and 1.5 are used for different solvent conditions, and where the other parameters are the same as in the good solvent case. The nature of the polymer is characterized by the second virial coefficient $v_2 = \int d\mathbf{r}\{1 - e^{-\beta V_{int}(r)}\}$, where $V_{int}(r)$ is the second term of Eq. (B.2). When $\varepsilon \approx 0.3$, v_2 approaches zero, and corresponds to the theta condition ($T = T_\Theta$).

B.2 Self-Consistent equation for λ

In this appendix, we provide the details for the calculations of $\langle Z\rangle_1$ and $\langle\delta[\mathbf{r}(s)-\mathbf{r}(s')]\rangle_1$ that appear in Eqs. (3.8) and (3.9).

$$\begin{aligned}\langle Z\rangle_1 &= \frac{\int\mathcal{D}\mathbf{r}(s)Ze^{-\beta H_1}}{\int\mathcal{D}\mathbf{r}(s)e^{-\beta H_1}} = \frac{\partial}{\partial(\beta f)}\log\left[\int\mathcal{D}\mathbf{r}(s)e^{-\beta H_1}\right]\\ &= \frac{\partial}{\partial(\beta f)}\log\left[\int\mathcal{D}\mathbf{r}_\perp(s)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\mathbf{r}_\perp^2(s)}\int\mathcal{D}z(s)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\left(\dot{z}(s)-\frac{a^2\lambda^2\beta f}{3}\right)^2+\frac{Na^2\lambda^2}{6}\beta^2f^2}\right]\\ &= \tfrac{1}{3}Na^2\lambda^2\beta f\end{aligned} \tag{B.3}$$

$$\begin{aligned}\langle\delta[\mathbf{r}(s)-\mathbf{r}(s')]\rangle_1 &= \int\int\int\int d\mathbf{r}(N)d\mathbf{r}(s')d\mathbf{r}(s)d\mathbf{r}(0)\; G(\mathbf{r}(N)-\mathbf{r}(s')|N-s')\delta[\mathbf{r}(s)-\mathbf{r}(s')] \\ &\quad \times G(\mathbf{r}(s')-\mathbf{r}(s)|s'-s)G(\mathbf{r}(s)-\mathbf{r}(0)|0) \\ &= G(0|s'-s), \end{aligned} \tag{B.4}$$

where the propagator $G(\cdots)$ is decomposed into transverse and longitudinal components, $G(\mathbf{R}|N) = G_\perp(\mathbf{R}_\perp|N)G_\parallel(Z|N;f)$, each of which can be exactly solved. We find

$$\begin{aligned} G_\perp(\mathbf{R}_\perp|N) &= \frac{\int \mathcal{D}\mathbf{r}_\perp(s)\delta\left(\int_0^N ds\dot{\mathbf{r}}_\perp(s)-\mathbf{R}_\perp\right)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\dot{\mathbf{r}}^2(s)}}{\int \mathcal{D}\mathbf{r}_\perp(s)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\dot{\mathbf{r}}^2(s)}} \\ &= \frac{\int \mathcal{D}\mathbf{r}_\perp(s)\int\frac{d^2\mathbf{k}}{(2\pi)^2}e^{i\mathbf{k}\cdot\left(\int_0^N ds\dot{\mathbf{r}}_\perp(s)-\mathbf{R}_\perp\right)}e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\dot{\mathbf{r}}^2(s)}}{\int \mathcal{D}\mathbf{r}_\perp(s)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\dot{\mathbf{r}}^2(s)}} \\ &= \int\frac{d^2\mathbf{k}}{(2\pi)^2}e^{-\frac{Na^2\lambda^2}{6}\left(\mathbf{k}-\frac{3\mathbf{R}_\perp}{Na^2\lambda^2}\right)^2-\frac{3\mathbf{R}_\perp^2}{2Na^2\lambda^2}} \\ &= \left(\frac{3}{2\pi Na^2\lambda^2}\right)\exp\left(-\frac{3\mathbf{R}_\perp^2}{2Na^2\lambda^2}\right), \end{aligned} \tag{B.5}$$

where the δ-function is first converted to the Fourier representation, and then the path integral is performed. The propagator $G_\perp(\mathbf{R}_\perp|N)$ is obtained after performing the integral in Fourier space. The longitudinal component is also similarly calculated under the Hamiltonian with linear force term,

$$\begin{aligned} G_\parallel(Z|N;f) &= \frac{\int \mathcal{D}z(s)\delta\left(\int_0^N ds\dot{z}(s)-Z\right)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\dot{z}^2(s)+\beta f\int_0^N ds\dot{z}(s)}}{\int \mathcal{D}z(s)e^{-\frac{3}{2a^2\lambda^2}\int_0^N ds\dot{z}^2(s)+\beta f\int_0^N ds\dot{z}(s)}} \\ &= \left(\frac{3}{2\pi Na^2\lambda^2}\right)^{1/2}\exp\left[-\frac{3}{2Na^2\lambda^2}\left(Z-\frac{Na^2\lambda^2\beta f}{3}\right)^2\right]. \end{aligned} \tag{B.6}$$

Thus, we find

$$\langle\delta[\mathbf{r}(s)-\mathbf{r}(s')]\rangle_1 = G(0|s'-s) = \left(\frac{3}{2\pi|s'-s|a^2\lambda^2}\right)^{3/2}\exp\left[-\frac{|s'-s|a^2\lambda^2\beta^2f^2}{6}\right]. \tag{B.7}$$

In obtaining Eq. (3.7), we used the extension Z as the observable to determine the optimal value of λ. Alternatively, one can also obtain the SCE for λ using the transverse fluctuation of polymer $\mathbf{R}_\perp^2 = X^2 + Y^2$ where X and Y are the projections of the end-to-end distance vector $\mathbf{R}$, i.e., $\langle \mathbf{R}_\perp^2(\Delta_1 + \Delta_2)\rangle - \langle \mathbf{R}_\perp^2\rangle\langle \Delta_1 + \Delta_2\rangle = 0$. Computations involving $\mathbf{R}_\perp^2$ are significantly simpler than those involving the end-to-end distance vector, $\mathbf{R}^2$, because the propagators in the x and y directions are decoupled from the force-dependent propagator in the z direction. Using the same methods as before with our original variational Hamiltonian in Eq. (3.4), we find

$$\langle \mathbf{R}_\perp^2 \Delta_1\rangle - \langle \mathbf{R}_\perp^2\rangle\langle \Delta_1\rangle = \frac{2Na^2\lambda^2(\lambda^2-1)}{3}, \tag{B.8}$$

$$\langle \mathbf{R}_\perp^2 \Delta_2\rangle - \langle \mathbf{R}_\perp^2\rangle\langle \Delta_2\rangle = -\frac{a^2\lambda^2 v_0}{3}\int_0^N ds \int_0^N ds'\, G(0|s'-s)(s'-s) \tag{B.9}$$

SCE using Eqs. (B.8) and (B.9) results in

$$\lambda^2 - 1 = \frac{v\sqrt{N}}{\lambda^3}\int_\delta^1 du \frac{1-u}{\sqrt{u}}\, e^{-Nu\lambda^2\varphi^2/6}\ , \tag{B.10}$$

which is identical to the SCE obtained using the linear end-to-end distance (Z) as the generating observable in Eq. (3.10). Thus, the computation of the FEC is not dependent on whether the mean extension or the transverse fluctuations are used in determining the self-consistent equation.

B.3 The Thick Chain Model

In order to verify the theoretical predictions for the polymer described by the (nearly) Inextensible Gaussian Hamiltonian (IGH) with excluded volume interactions, we have simulated the FEC using the thick chain (TC) model for a self-avoiding polymer. In the TC model, the polymer is viewed as a chain with a finite uniform thickness D, and is represented as a succession of beads with position vectors $\mathbf{r}_0, ..., \mathbf{r}_N$. All of the bond vectors $\Delta\mathbf{r}_n = \mathbf{r}_{n+1} - \mathbf{r}_n$ $(n = 0, .., N)$ have the same modulus a. Therefore, unlike the IGH where $\langle|\Delta\mathbf{r}_n|\rangle \approx a$ in Eq. (3.16), the bond length restriction $|\Delta\mathbf{r}_n| = a$ is strictly enforced in the TC model. The Hamiltonian of the chain under tension is given by

$$\mathcal{H}_{TC} = \sum_{i,j,k} V(R_{i,j,k}) - \mathbf{f} \cdot (\mathbf{r}_N - \mathbf{r}_0), \tag{B.11}$$

where the first term enforces the self-avoidance, and the second term represents the external force. In particular,

$$V(R_{ijk}) = \begin{cases} 0, & R_{ijk} > D \\ \infty, & R_{ijk} \leq D, \end{cases} \tag{B.12}$$

where R_{ijk} is radius of the circle going through the triplet of beads (i, j, k). Physically, the first term in the Hamiltonian (Eq. (B.11)) ensures the self-avoidance of the chain by rejecting both local self-intersection (the local radius of curvature must be no smaller than D) and interpenetration of any two portions of the chain at some finite arc-length. Intuitively, it allows only configurations satisfying the thickness constraints, that the radii of circles going through all the triplets of beads (i, j, k) are greater than D.

In order to characterize the stretching response of a thick chain with $D/a = 1$ and $N = 1600$ we performed Monte Carlo simulations using the following scheme. Starting from an arbitrary initial chain conformation that satisfies the thickness constraints, the exploration of the available configuration space was performed by distorting conformations by means of pivot and crankshaft moves. The new structures were accepted or rejected according to the standard Metropolis criterion (the infinite strength of the three-body penalties of Eq. (B.11) was enforced by always rejecting configurations violating the circumradius constraints). The relative elongation of the chain was calculated for increasing values of the applied stretching force. For each run, after equilibration, we measure the autocorrelation time and sampled a sufficient number of independent conformations to achieve a relative error of at most 10^{-3} in the average chain elongation. For moderate to high forces, this typically entailed the collection of 10^4 independent structures, whereas a 10-fold increase of sampling was required at small forces due to the broad distribution of the end-to-end separation along the force direction. At small forces, conformational fluctuations can be even larger than the mean extension, which makes achieving converged results for $\langle Z \rangle$ more difficult.

Appendix C

Appendix for Chapter 4

C.1 Simulation Hamiltonian

The Hamiltonian for the RNA hairpin with N nucleotides, which is modeled using the self-organized polymer (SOP) model [72], is

$$H_{SOP} = -\frac{kR_0^2}{2}\sum_{i=1}^{N-1}\log\left(1-\frac{(r_{i,i+1}-r_{i,i+1}^o)^2}{R_0^2}\right)+\sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\epsilon_h\left[\left(\frac{r_{i,j}^o}{r_{i,j}}\right)^{12}-2\left(\frac{r_{i,j}^o}{r_{i,j}}\right)^{6}\right]\Delta_{i,j} + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\epsilon_l\left(\frac{\sigma}{r_{i,j}}\right)^{12}(1-\Delta_{i,j})+\sum_{i=1}^{N-2}\epsilon_l\left(\frac{\sigma^*}{r_{i,i+2}}\right)^{6}, \quad \text{(C.1)}$$

where $r_{i,j} = |\mathbf{r}_i - \mathbf{r}_j|$ and $r_{i,j}^o$ is the distance between monomers i and j in the native structure. The first term enforces backbone chain connectivity using the finite extensible nonlinear elastic (FENE) potential, with $k \approx 1.4 \times 10^4$ pN·nm^{-1} and $R_0 = 0.2$ nm. The Lennard-Jones interaction (second term in equation (C.1)) describes interactions only between native contacts (defined as $r_{i,j}^o \leq 1.4$ nm for $|i-j| > 2$), with $\Delta_{i,j} = 1$ if monomers i and j are within 1.4 nm in the native state, and $\Delta_{i,j} = 0$ otherwise. Non-native interactions are treated as purely repulsive (the third term in equation (C.1)) with $\sigma = 0.7$ nm. We take $\epsilon_h = 4.9$ pN·nm and $\epsilon_l = 7.0$ pN·nm for the strength of interactions. In the fourth term, the repulsion between the i^{th} and $(i+2)^{th}$ interaction sites along the backbone has $\sigma^* = 0.35$ nm to prevent disruption of the native helical structure.

The handles are modeled using the Hamiltonian

$$H_{handles} = \frac{k_S}{2} \sum_{i=1}^{N-1} (r_{i,i+1} - r_0)^2 - k_A \sum_{i=1}^{N-2} \hat{\mathbf{r}}_{i,i+1} \cdot \hat{\mathbf{r}}_{i+1,i+2}. \tag{C.2}$$

The neighboring interaction sites, with an equilibrium distance $r_0 = 0.5$ nm, are harmonically constrained with a spring constant $k_S \approx 1.4 \times 10^4$ pN·nm^{-1}. In the second term of eq. (C.2), the strength of the bending potential, k_A, determines the handle flexibility. We choose two values, k_A =7.0 pN·nm and k_A =561 pN·nm to model flexible and semiflexible handles respectively, and assign $k_A = 35$ pN·nm to the junction connecting two ends of the RNA and the handles. We determine the corresponding persistence length for the two k_A values as $l_p = 0.6$ and 70 nm (see SI text). The contour length of each handle is varied from $N = 5 - 200$.

C.2 The GRM Hamiltonian Matrix

Defining $\mathbf{R}_N(t) = \{x_1(t), \ldots, x_{N+1}(t)\}$, we can write the Hamiltonian in eq. 4.1 as

$$\beta H = \frac{3}{2a^2}\, \mathbf{R}_N^T \mathbf{M} \mathbf{R}_N - \beta f(z_{N+1} - z_1) \tag{C.3}$$

where $\mathbf{M}$ can be written in the block matrix form

$$\mathbf{M} = \begin{pmatrix} a^2\mathbf{M}_1/b^2 & 0 & \\ 0 & \mathbf{M}_2 & 0 \\ 0 & 0 & a^2\mathbf{M}_3/b^2 \end{pmatrix}. \tag{C.4}$$

$\mathbf{M}_1$ and $\mathbf{M}_3$ both symmetric $N_h \times N_h$ matrices, and $\mathbf{M}_2$ is a symmetric $(N-2N_h) \times (N-2N_h)$ matrix. The symmetric block matrices have non-vanishing elements given

by

$$\mathbf{M}^{(1)}_{i,i} = 2 - \delta_{i,1} + b^2 k_0/3a^2 \qquad \mathbf{M}^{(3)}_{i,i} = 2 - \delta_{i,N+1} \tag{C.5}$$

$$\mathbf{M}^{(2)}_{i,i} = 2 + (a^2/b^2 + k - 1)[\delta_{i,N_h+1} - \delta_{i,N-Nh+1}] \tag{C.6}$$

$$\mathbf{M}^{(1)}_{i,i+1} = \mathbf{M}^{(2)}_{i,i+1} = \mathbf{M}^{(3)}_{i,i+1} = -1 \tag{C.7}$$

$$\mathbf{M}^{(2)}_{1,N_h+1} = k \tag{C.8}$$

In the case when the interaction is cut off (i.e. $|\mathbf{r}_{N-N_h+1} - \mathbf{r}_{N_h+1}| \leq c$), the matrices are identical with the replacement $k \to 0$. In the case $k = 0$, $k_0 = 0$, and $b = a$ (i.e. no interaction, free endpoints, and identical RNA and handle spacing), we recover the standard Rouse matrix, $\mathbf{M}_{i,i} = 2 - \delta_{i,1} - \delta_{i,N+1}$ and $\mathbf{M}_{i,i+1} = -1$, as expected.

C.3 Derivation of the GRM Propagator

For the Generalized Rouse Model, with a Hamiltonian given in eq. C.3, each component of a monomer's position evolves under the Langevin equation

$$\begin{aligned} \zeta \frac{\partial \mathbf{X}}{\partial t} &= -\frac{3k_B T}{2a^2}\mathbf{M}\cdot\mathbf{X} + \mathbf{N}(t') + \mathbf{F}(t') \\ \mathbf{X}(t) &= \mathbf{G}(t)\cdot\mathbf{X}(0) + \frac{1}{\zeta}\int_0^t dt'\ \mathbf{G}(t-t')\Big(\mathbf{N}(t') + \mathbf{F}(t')\Big) \end{aligned} \tag{C.9}$$

where $\mathbf{X} = \{x_1(t), \ldots, x_{N+1}(t)\}$ contains the positions $\mathbf{N} = \{\eta_1^{(x)}, \ldots, \eta_{N+1}^{(x)}\}^T$ contains the random, white noise forces, and $\mathbf{F} = \{f_1^{(x)}(t), \ldots, f_{N+1}^{(x)}(t)\}^T$ contains the time-dependent force, and where we have defined

$$\mathbf{G}(t) = \exp\Big(-\mathbf{M}\frac{3Dt}{a^2}\Big) \tag{C.10}$$

The y and z components evolve under an identical Langevin equation, and since the GRM is purely Gaussian, they do not interact.

Because the GRM is Gaussian, we can write the propagator for the distance between monomers n and m, $\Delta\mathbf{r} = \mathbf{r}_n(t) - \mathbf{r}_m(t)$ as $P(\Delta\mathbf{r}, t|\Delta\mathbf{r}_0)$=$P(\Delta x, t|\Delta x_0)$ $P(\Delta y, t|\Delta y_0)$ $P(\Delta z, t|\Delta z_0)$. For the x component, we have

$$P(\Delta x, t|\Delta x_0) = \Big\langle \delta[x_n(t) - x_m(t) - \Delta x]\ \delta[x_n(0) - x_m(0) - \Delta x_0] \Big\rangle \tag{C.11}$$

$$= \int \frac{dk\, dq}{(2\pi)^2} e^{-ik\Delta x - iq\Delta x_0} \Big\langle \exp\Big(ik[x_n - x_m] + iq[x_n(0) - x_m(0)] \Big) \Big\rangle \tag{C.12}$$

The term $ik[x_n - x_m] + iq[x_n(0) - x_m(0)]$ in eq. C.12 can be written as $ik[x_n - x_m] + iq[x_n(0) - x_m(0)] = \mathbf{T}_n - \mathbf{T}_m$, where $\mathbf{T} = \mathbf{T}^{(1)} + \mathbf{T}^{(2)} + \mathbf{T}^{(3)}$ and

$$\mathbf{T}^{(1)} = \Big(ik\mathbf{G}(t) + iq\mathbf{I} \Big)\mathbf{X}(0) \tag{C.13}$$

$$\mathbf{T}^{(2)} = \frac{ik}{\zeta} \int_0^t dt' \mathbf{G}(t - t')\mathbf{N}(t') \tag{C.14}$$

$$\mathbf{T}^{(3)} = \frac{ik}{\zeta} \int_0^t dt' \mathbf{G}(t - t')\mathbf{F}(t') \tag{C.15}$$

We must therefore average $\exp(\mathbf{T}_n^{(1)} - \mathbf{T}_m^{(1)})$ over the initial conditions and $\exp(\mathbf{T}_n^{(2)} - \mathbf{T}_m^{(2)})$ over the noise, in order to determine the desired propagator.

In order to compute the noise averages, we write

$$\langle \cdots \rangle_{noise} = \frac{1}{\mathcal{N}} \int \prod_l \mathcal{D}[N_l(t)]\ \Big(\cdots \Big) \exp\Big(-\frac{1}{4k_B T\zeta} \int_{-\infty}^{\infty} dt\ N_l^2(t) \Big) \tag{C.16}$$

$\exp(\mathbf{T}_n^{(2)} - \mathbf{T}_m^{(2)})$ can then easily be evaluated by completing the square, with

$$-\frac{\beta}{4\zeta} \sum_l \int_0^t dt' N_l^2(t') + \frac{ik}{\zeta} \sum_l \int_0^t dt' \Big(G_{nl}(t - t') - G_{ml}(t - t') \Big) N_l(t') \tag{C.17}$$

$$= -\frac{\beta}{4\zeta} \sum_l \int_0^t dt' \Big[N_l(t') - \frac{2ik}{\beta} \Big(G_{nl}(t - t') - G_{ml}(t - t') \Big) \Big]^2 \tag{C.18}$$

$$-\frac{k^2}{\beta\zeta} \sum_l \int_0^t \Big(G_{nl}(t - t') - G_{ml}(t - t') \Big)^2$$

$$= -\frac{\beta}{4\zeta}\int_0^t dt'\ \bar{\mathbf{N}}(t')\bar{\mathbf{N}}^T(t') \tag{C.19}$$
$$-\frac{k^2}{\beta\zeta}\int_0^t dt'\Big([\mathbf{G}(t')\mathbf{G}^T(t')]_{nn}+[\mathbf{G}(t')\mathbf{G}^T(t')]_{mm}-2[\mathbf{G}(t')\mathbf{G}^T(t')]_{nm}\Big)$$

where we have defined $\bar{\mathbf{N}}_l = \mathbf{N}_l(t') - 2ik\,k_BT[G_{nl}(t-t') - G_{ml}(t-t')]$. The path integral over $\mathbf{N}$ can be computed after a change of variables ($\mathbf{N} \to \bar{\mathbf{N}}$). From eq. C.10, we find

$$\int_0^t dt'\mathbf{G}(t')\mathbf{G}^T(t') = \frac{a^2}{6D}\mathbf{M}^{-1}\Big(\mathbf{I}-\mathbf{G}(2t)\Big) \tag{C.20}$$

so that

$$\Big\langle e^{\mathbf{T}_n^{(2)}-\mathbf{T}_m^{(2)}}\Big\rangle = \exp\Big[-\frac{k^2a^2}{6}\sum_{lj}\Big(\delta_{nl}-\delta_{ml}\Big)\Big(\mathbf{M}^{-1}\Big)_{lj}\Big(\delta_{jn}-\delta_{jm}\Big)$$
$$+\frac{k^2a^2}{6}\sum_{lj}\Big(\mathbf{G}_{nl}-\mathbf{G}_{ml}\Big)\Big(\mathbf{M}^{-1}\Big)_{lj}\Big(\mathbf{G}_{jn}-\mathbf{G}_{jm}\Big)\Big] \tag{C.21}$$

Note that this noise average is completely independent of the details of the force, and depends *only* on the GRM Hamiltonian.

The average over initial coordinates is with respect to the Hamiltonian

$$H = -\frac{3}{2a^2}\mathbf{X}^T(0)\mathbf{M}\mathbf{X}(0) + \beta\mathbf{R}(0)\mathbf{F}^0 \tag{C.22}$$

where we have defined $\mathbf{F}^0 = \mathbf{F}(0)$. We can rewrite for simplicity $\mathbf{T}_n^{(1)} - \mathbf{T}_m^{(1)} = \mathbf{V}\cdot\mathbf{X}(0)$, with

$$[\mathbf{V}]_l \equiv ik([\mathbf{G}_{nl}(t) - \mathbf{G}_{ml}(t)]) + iq(\delta_{nl}-\delta_{ml}) \tag{C.23}$$

The average over initial coordinates is now trivial, since $\int d\mathbf{X}\exp(-\mathbf{X}\mathbf{M}\mathbf{X} + [\mathbf{V} + \mathbf{F}^0]\mathbf{X}) \propto \exp([\mathbf{V}+\mathbf{F}^0]\mathbf{M}^{-1}[\mathbf{V}+\mathbf{F}^0]/4)$. Exploiting the symmetries of the matrices,

we can rewrite this as

$$\frac{a^2}{6}(\mathbf{F}^0)^T\mathbf{M}^{-1}\mathbf{F}^0 - \frac{a^2k^2}{6}\sum_{lj}\Big(\mathbf{G}_{nl}(t)-\mathbf{G}_{ml}(t)\Big)(\mathbf{M}^{-1})_{lj}\Big(\mathbf{G}_{jn}-\mathbf{G}_{jm}(t)\Big) \quad \text{(C.24)}$$
$$-\frac{a^2kq}{3}\sum_{lj}\Big(\mathbf{G}_{nl}(t)-\mathbf{G}_{ml}(t)\Big)(\mathbf{M}^{-1})_{lj}\Big(\delta_{jn}-\delta_{jm}\Big)$$
$$-\frac{a^2q^2}{6}\sum_{lj}\Big(\delta_{nl}-\delta_{ml}\Big)(\mathbf{M}^{-1})_{lj}\Big(\delta_{jn}-\delta_{jm}\Big)$$
$$+\frac{ika^2}{3}\sum_{lj}\Big(\mathbf{G}_{nl}(t)-\mathbf{G}_{ml}(t)\Big)(\mathbf{M}^{-1})_{lj}\mathbf{F}^0_j + \frac{iqa^2}{3}\sum_{lj}\Big(\delta_{nl}-\delta_{ml}\Big)(\mathbf{M}^{-1})_{lj}\mathbf{F}^0_j$$

Combining the force-independent terms of all of the averages gives

$$\langle\exp(ik[x_n(t)-x_m(t)]+iq[x_n(0)-x_m(0)])\rangle = \exp\Big(-\frac{(k^2+q^2)}{6}\Delta x^2_{eq} - \frac{kq}{3}\Delta x^2_{eq}h(t) + B(t,\mathbf{F})\Big) \quad \text{(C.25)}$$

with

$$\Delta x^2_{eq} = a^2\Big[\Big(\mathbf{M}^{-1}\Big)_{nn} + \Big(\mathbf{M}^{-1}\Big)_{mm} - 2\Big(\mathbf{M}^{-1}\Big)_{nm}\Big] \quad \text{(C.26)}$$

$$\Delta x^2_{eq}h(t) = a^2\Big[\Big(\mathbf{M}^{-1}\mathbf{G}(t)\Big)_{nn} + \Big(\mathbf{M}^{-1}\mathbf{G}(t)\Big)_{mm} - 2\Big(\mathbf{M}^{-1}\mathbf{G}(t)\Big)_{nm}\Big] \quad \text{(C.27)}$$

where $\Delta x^2_{eq} = \lim_{t\to\infty}\langle[x_n(t)-x_m(t)]^2\rangle|_{\mathbf{f}=0}$, and where B includes the contribution from the external tension (see below).

The force-dependent terms are most easily considered by computing the contributions from k and q separately. There is only one q-dependent term, giving the contribution

$$\frac{iqa^2}{3}\sum_{lj}\Big(\delta_{nl}-\delta_{ml}\Big)(\mathbf{M}^{-1})_{lj}\mathbf{F}^0_j = iq\langle x\rangle_0 \quad \text{(C.28)}$$

with $\langle x\rangle_0$ the average initial extension of the chain under the force $\mathbf{F}^0$. This term is independent of the time-dependent details of the force.

The k dependent terms are more complicated, but we can rewrite

$$\frac{ika^2}{3}\sum_{lj}\Big(\mathbf{G}_{nl}(t)-\mathbf{G}_{ml}(t)\Big)(\mathbf{M}^{-1})_{lj}\mathbf{F}^0_j = ik\langle x\rangle_0$$
$$-\frac{ik}{\zeta}\int_0^t dt'\Big[\Big(\mathbf{G}(t-t')\mathbf{F}^0\Big)_n-\Big(\mathbf{G}(t-t')\mathbf{F}^0\Big)_m\Big] \qquad \text{(C.29)}$$

Combining this term with the non-averaged portion of the propagator, we find the force-dependent contribution becomes

$$B(t,\mathbf{F}) = i(k+q)\langle z\rangle_0 + ika^2\Delta f_{eff}(t) \qquad \text{(C.30)}$$
$$a^2\Delta f_{eff}(t) \equiv \frac{1}{\zeta}\Big(\int_0^t dt'\ \mathbf{G}(t-t')[\mathbf{F}(t')-\mathbf{F}^0]\Big)_n-\frac{1}{\zeta}\Big(\int_0^t dt'\ \mathbf{G}(t-t')[\mathbf{F}(t')-\mathbf{F}^0]\Big)_m$$

We now have the final form of the propagator:

$$P(\Delta x,t|\Delta x_0)=\int\frac{dk\,dq}{(2\pi)^2}\exp\Big(-\frac{\langle\Delta x^2_{eq}\rangle}{6}(k^2+q^2)-\frac{\langle\Delta x^2\rangle_{eq}h(t)}{3}kq$$
$$+k[\bar{\Delta}x-a^2\Delta f_{eff}(t)]+iq\bar{\Delta}x_0\Big) \qquad \text{(C.31)}$$

with $\bar{\Delta}x=\Delta x-\langle\Delta x\rangle_0$. Integration over k and q gives

$$P(\Delta x,t|\Delta x_0)=\frac{3}{2\pi\Delta x^2_{eq}\sqrt{1-h^2(t)}}\exp\Big(-\frac{3}{2\Delta x^2_{eq}[1-h^2(t)]}\Big[\bar{\Delta}x_0^2+$$
$$(\bar{\Delta}x-a^2\Delta f_{eff}(t))^2-2h(t)\bar{\Delta}x_0(\bar{\Delta}x-a^2\Delta f_{eff}(t))\Big]\Big) \qquad \text{(C.32)}$$

In the case of a constant tension, $\Delta f_{eff}(t)\equiv 0$, and we can write the full, three dimensional propagator as

$$P(\Delta\mathbf{r},t|\Delta\mathbf{r}_0)=\Big(\frac{3}{2\pi\Delta x^2_{eq}\sqrt{1-h^2(t)}}\Big)^3\exp\Big(-\frac{3}{2\Delta x^2_{eq}[1-h^2(t)]}$$
$$\times\Big[\bar{\Delta}\mathbf{r}_0^2+\bar{\Delta}\mathbf{r}^2-2h(t)\bar{\Delta}\mathbf{r}\cdot\bar{\Delta}\mathbf{r}_0\Big]\Big) \qquad \text{(C.33)}$$

with $\bar{\Delta}\mathbf{r}=\Delta\mathbf{r}-\langle\Delta\mathbf{r}\rangle_0$.

Appendix D

Appendix For Chapter 5

D.1 The $\mathbf{Q}$ Matrix

Defining $\mathbf{x}^N = (x_1, \ldots, x_{N+1})$, we can rewrite the Hamiltonian in eq. 5.2 as

$$\mathcal{H}_x = a \sum_n \left(\frac{l_p}{2} \frac{(\Delta x_{n+1} - \Delta x_n)^2}{a^4} + \lambda_n \frac{\Delta x_n^2}{a^2} + k_n \frac{x_n^2}{R^2} \right) = \left(\mathbf{x}^N\right)^T \mathbf{Q} \mathbf{x}^N \tag{D.1}$$

with $\mathbf{Q}$ a tridiagonal matrix with elements:

$$\mathbf{Q}_{i,i+2} = \mathbf{Q}_{i+2,i} = \frac{l_p}{2a^4} \tag{D.2}$$

$$\mathbf{Q}_{i,i+1} = \mathbf{Q}_{i+1,i} = \begin{cases} -\frac{l_p}{a^4} - \frac{2\lambda_i}{a^2} & i = 1, N \\ -\frac{2l_p}{a^4} - \frac{2\lambda_i}{a^2} & \text{else} \end{cases} \tag{D.3}$$

$$\mathbf{Q}_{i,i} = \begin{cases} \frac{l_p}{2a^4} + \frac{\lambda_1}{a^2} + \frac{ak_1}{R^2} & i = 1 \\ \frac{l_p}{2a^4} + \frac{\lambda_N}{a^2} + \frac{ak_{N+1}}{R^2} & i = N+1 \\ \frac{5l_p}{2a^4} + \frac{\lambda_{i-1}+\lambda_i}{a^2} + \frac{ak_i}{R^2} & i = 2, N \\ \frac{3l_p}{a^4} + \frac{\lambda_{i-1}+\lambda_i}{a^2} + \frac{ak_i}{R^2} & \text{else} \end{cases} \tag{D.4}$$

The structure of $\mathbf{Q}$ is unchanged under the transformation $k_1 = k_{N+1}$, $k_2 = k_N$, $\lambda_1 = \lambda_N$, $k_i = k$ for $2 < i < N$ and $\lambda_i = \lambda$ for $1 < i < N$.

D.2 Evaluation of the 1-D Confined Propagator

We are interested in evaluating the path integral in eq. 5.5,

$$Z_0(\mathbf{x}) = \int \Delta[x(s)] \exp\left(-\frac{l_p}{2} \int_0^L d\sigma\, \ddot{x}^2(s) - \lambda \int_0^L ds\, \dot{x}^2(s) - \frac{\kappa}{R^2} \int_0^L ds\, x^2(s) \right), \tag{D.5}$$

subject to the boundary conditions $x(0) = x_0$, $u(0) = u_0$, $x(L) = x_L$, and $u(L) = u_L$. We can write $x(s) = f(s) + g(s)$, where $g(0) = g(L) = \dot{g}(0) = \dot{g}(L) = 0$ and

$$\frac{l_p}{2} f^{(4)}(s) - \lambda \ddot{f}(s) + \kappa f(s) = 0$$

with $f(0) = x_0,\ f(L) = x_L,\ \dot{f}(0) = u_0,$ and $f(L) = u_L$. If f satisfies the above differential equation, a simple integration by parts gives

$$\begin{aligned} Z_0(\mathbf{x}) &= K(L) \exp\Big(-\frac{l_p}{2}\Big[\ddot{f}\dot{f} - f^{(3)} f\Big]_0^L - \lambda[u_L x_L - u_0 x_0]\Big) \quad \text{(D.6)} \\ K(L) &= \int \Delta[g(\sigma)] \exp\Big(-\frac{l_p}{2}\int_0^L d\sigma\ \ddot{g}^2(\sigma) - \lambda \int_0^L d\sigma\ \dot{g}^2(\sigma) \\ -\frac{\kappa}{R^2}\int_0^L d\sigma\ g^2(\sigma)\Big) & \quad \text{(D.7)} \end{aligned}$$

where g and g' vanish at the boundaries. The exponential term in eq. D.6 can be evaluated by solving the differential equation for f, giving

$$Z_0(\mathbf{x}) = K(L) e^{-\mathbf{x}\cdot\mathbf{M}\mathbf{x}}, \quad \text{(D.8)}$$

where

$$\mathbf{M} = \frac{l_p R}{2d(L)} \begin{pmatrix} R\,m_{11} & R\,m_{12} & R\,m_{13} & m_{14} \\ R\,m_{12} & R\,m_{11} & -m_{14} & -R\,m_{13} \\ R\,m_{13} & -m_{14} & m_{33}/R & m_{34}/R \\ m_{14} & -R\,m_{13} & m_{34}/R & m_{33}/R \end{pmatrix} + \frac{\lambda}{2} \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad \text{(D.9)}$$

Here, we have defined

$$\begin{aligned} d(L) &= \frac{2\omega_1\omega_2(1 - \cosh(L\omega_1)\cosh(L\omega_2)) + (\omega_1^2 + \omega_2^2)\sinh(L\omega_1)\sinh(L\omega_2)}{\omega_1\omega_2(\omega_1^2 - \omega_2^2)} \\ m_{11} &= \omega_1 \sinh(L\omega_1)\cosh(L\omega_2) - \omega_2 \cosh(L\omega_1)\sinh(L\omega_2) \\ m_{12} &= \omega_2 \sinh(L\omega_2) - \omega_1 \sinh(L\omega_1) \end{aligned}$$

$$
\begin{aligned}
m_{13} &= \frac{\omega_1^2 - \omega_2^2}{2\omega_1\omega_2}\sinh(L\omega_1)\sinh(L\omega_2) \\
m_{14} &= \cosh(L\omega_1) - \cosh(L\omega_2) \\
m_{33} &= \frac{1}{\omega_2}\cosh(L\omega_1)\sinh(L\omega_2) - \frac{1}{\omega_1}\sinh(L\omega_1)\cosh(L\omega_2) \\
m_{34} &= \frac{1}{\omega_1}\sinh(L\omega_1) - \frac{1}{\omega_2}\sinh(L\omega_2)
\end{aligned}
$$

with

$$
\omega_i = \left[\frac{\lambda}{l_p}\Big(1 \pm \sqrt{1 - \frac{2\kappa l_p}{\lambda^2}}\Big)\right]^{\frac{1}{2}}. \tag{D.10}
$$

Note that the full propagator $Z(\mathbf{x})$in eq. 5.8 can be written as $Z(\mathbf{x}) = Z_0(\mathbf{x})\,\exp(-\mathbf{x}\cdot\mathbf{G}\mathbf{x})$, with the matrix $\mathbf{G}$ containing terms suppressing excess end-point fluctuations, given by

$$
\mathbf{G} = \begin{pmatrix}
\gamma_1/R^2 & 0 & \gamma_2/R & 0 \\
0 & \gamma_1/R^2 & 0 & -\gamma_2/R \\
\gamma_2/R & 0 & \delta & 0 \\
0 & -\gamma_2/R & 0 & \delta
\end{pmatrix} \tag{D.11}
$$

In general, average values will involve calculating the determinant of some matrix involving $\mathbf{M}$. Simplification of the determinant is a tediuous process, but it is useful to note that any matrix with the above symmetries has a determinant

$$
\begin{aligned}
\mathrm{Det}(\mathbf{M}) &= A_1^2 - A_2^2 \qquad (D.12) \\
A_1 &= m_{13}^2 + m_{14}^2 + m_{12}m_{34} - m_{11}m_{13} \\
A_2 &= 2m_{13}m_{14} + m_{12}m_{33} - m_{11}m_{34}
\end{aligned}
$$

We can calculate $K(L)$ by the evaluation of a simple integral. Following the standard method of Feynman [155], we can write the propagator from (x_0, u_0) to

(x_L, u_L) as a sum over all intermediate points, (x_s, u_s),

$$\begin{aligned} Z_0(x_0, x_L, u_0, u_L; L) &= \int_{-\infty}^{\infty} dx_s du_s \, Z_0(x_0, u_0, x_s, u_s; s) \times Z_0(x_s, u_s, x_L, u_L; L-s) \qquad \text{(D.13)} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_s du_s \; K(s) \exp\Big(- \mathbf{x}_1^T \cdot \mathbf{M}(s) \mathbf{x}_1 \Big) \qquad \text{(D.14)} \\ &\qquad \times K(L-s) \exp\Big(- \mathbf{x}_2^T \cdot \mathbf{M}(L-s)\, \mathbf{x}_2 \Big) \end{aligned}$$

where $\mathbf{x}_1 = (x_0, x_s, u_0, u_s)$ and $\mathbf{x}_2 = (x_s, x_L, u_s, u_L)$. $\mathbf{M}$ has already been determined (eq. D.9), and $K(L)$, given in eq. D.7, is independent of all x_i's and u_i's. The integral in eq. D.14 is tedious to evaluate, but yields

$$Z_0(\mathbf{x}; L) = K(L) e^{-\mathbf{x}^T \cdot \mathbf{M}(L)\, \mathbf{x}} = K(s) K(L-s) \; \frac{2\pi}{l_p} \Big(\frac{d(s) d(L-s)}{(\omega_1^2 - \omega_2^2) d(L)} \Big)^{\frac{1}{2}} e^{-\mathbf{x}^T \cdot \mathbf{M}(L)\, \mathbf{x}} \qquad \text{(D.15)}$$

We then find

$$K(L) = \frac{l_p}{2\pi} \sqrt{\frac{\omega_1^2 - \omega_2^2}{d(L)}} \; e^{-L\eta}, \qquad \text{(D.16)}$$

where η is an arbitrary constant, independant of L, l_p, and R, adding an irrelevant constant to the free energy.

$\mathbf{M}$ and K can be greatly simplified in the limit of strong confinement ($L\omega_i \gg 1$, see the main text), with

$$\begin{aligned} \mathbf{M} &= \frac{l_p R \omega_1 \omega_2}{2} \begin{pmatrix} R(\omega_1 + \omega_2) & 0 & 1 & 0 \\ 0 & R(\omega_1 + \omega_2) & 0 & -1 \\ 1 & 0 & (\omega_1^{-1} + \omega_2^{-1})/R & 0 \\ 0 & -1 & 0 & (\omega_1^{-1} + \omega_2^{-1})/R \end{pmatrix} \\ K &\propto \sqrt{\omega_1 \omega_2}\, (\omega_1 + \omega_2) e^{-L(\omega_1 + \omega_2)/2} \qquad \text{(D.17)} \end{aligned}$$

This strongly confined representation is significantly easier to work with when computing the mean field solutions.

To ensure that our calculation of Z_0 has the correct limiting behavior, we find

$$\lim_{R\to\infty}\int_{-\infty}^{\infty} dx_L Z_0(\mathbf{x}) = e^{-kLx_0^2}\left(\frac{l_p\Omega}{2\pi\sinh(\Omega L)}\right)^{\frac{1}{2}}\exp\Big(-\frac{l_p\Omega}{2\sinh(\Omega L)} \times\Big[(u_0^2+u_L^2)\cosh(\Omega L)-2u_0u_L\Big]+O(1/R)\Big) \tag{D.18}$$

which is identical to the unconfined propagator found in the work of Thirumalai and Ha [41, 42], except for the the term $e^{-kLx_0^2}$. Since, as $R\to\infty$, the system becomes translationaly invariant, we would expect x_0 to decouple from the bending terms, and that integration over x_0 would lead to a divergent integral $\propto V\to\infty$. The integral over the initial position then simply adds an irrelevant constant to the free energy. We therefore have the relationship

$$Z_U(u_0,u_L) = \lim_{R\to\infty}\frac{1}{V}\int_{-\infty}^{\infty} dx_0 dx_L\ Z(\mathbf{x}) \tag{D.19}$$

identical to the propagator used by Ha and Thirumalai [41, 42], up to a multiplicative constant.

D.3 Details of the simulations for volume confinement

Because the Configurational Bias Monte Carlo (CBMC) method [156] is only applicable to Hamiltonians involving nearest neighbor interactions, we can not use this method to accurately determine the equilibrium behavior of a spherically confined wormlike chain. We have therefore performed a number of Langevin Dynamics simulations with varying L, l_p, and R when considering volume confinement. The Hamiltonian used is

$$\beta H = \frac{k}{2a^2}\sum_{i=0}^{N}(|\mathbf{r}_{i+1}-\mathbf{r}_i|-a)^2 - \frac{l_p}{a}\sum_{i=0}^{N-1}\mathbf{u}_i\cdot\mathbf{u}_{i+1} + \epsilon_S\sum_{i=0}^{N+1}\left(\frac{a}{|\mathbf{r}_i|-(R+a)}\right)^{12} \tag{D.20}$$

with $L = Na$. The first term ensures the connectivity of the chain, and we take $k = 10^4$ throughout, ensuring very stiff bonds. The second term accounts for the bending stiffness of the chain, with persistence length l_p. We have confirmed directly that this Hamiltonian in the unconfined case (i.e. only the first two terms of eq. D.20 are used) gives $\langle \mathbf{u}(s) \cdot \mathbf{u}(s') \rangle = e^{-|\Delta s|/l_p}$ to within $\sim 5\%$. The third term of the Hamiltonian approximately confines the chain to the interior of a sphere of radius R, using a Lennerd-Jones repulsion. The confinement energy is on the order of $\epsilon_S \, k_B T$ when $|\mathbf{r}_i| = R$, and increases sharply for larger $|\mathbf{r}_i|$. We choose $\epsilon_S = 1$ throughout the simulations, which restricts $|\mathbf{r}_i|/R \leq 1.01$ for all of the parameters we considered. To determine the equilibrium properties of the system, we use the low friction limit, with $\eta = 0.1$, and a timestep of $h = 0.001$ (in dimensionless units, or equivalently with the mass $m = 1$, spacing $a = 1$, and $k_B T = 1$).

In order to study the behavior of an excluded volume chain, we instead use the hamiltonian

$$\beta H_1 = \beta H_0 + \epsilon_{EV} \sum_{i=0}^{N-1} \sum_{j=i+2}^{N+1} \left(\frac{a}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{12} \tag{D.21}$$

We choose $\epsilon_{EV} = 1$ throughout the excluded volume simulations.

D.4 Calculation of the winding axis

Because of the spherical symmetry of the problem, we are free to choose our coordinate system such that it simplifies the calculation. We define $\hat{\mathbf{u}}_{i-1} = \hat{\mathbf{z}}$, determining the z-axis, and $\hat{\mathbf{u}}_i = (\sin(\Theta_i), 0, \cos(\Theta_i))$, defining the x-axis. We take our third bond to be $\hat{\mathbf{u}}_{i+1} = (\sin(\Theta_{i+1})\cos(\varphi), \sin(\Theta_{i+1})\sin(\varphi), \cos(\Theta_{i+1}))$. With

$\theta_{i,j}$ the angle between bonds i and j, we see $\theta_{i-1,i} = \Theta_i$ and $\theta_{i-1,i+1} = \Theta_{i+1}$. It is convenient to eliminate the azimuthal angle φ when computing $\cos(\theta_{i,i+1}) = \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_{i+1}$, giving

$$\cos(\varphi) = \cot(\theta_{i,i+1}) \cot(\Theta_i) - \cos(\Theta_{i+1}) \csc(\theta_{i,i+1}) \csc(\Theta_i) \tag{D.22}$$

In this coordinate system, $\hat{\mathbf{a}}_i = \hat{\mathbf{y}}$, and $|\mathbf{a}_{i+1}| = \sin(\theta_{i,i+1})$. To compute the dot product between the two winding axes, we need only $\mathbf{a}_{i+1} \cdot \hat{\mathbf{y}} = \cos(\Theta_i) \sin(\Theta_{i+1}) \cos(\varphi) - \cos(\Theta_{i+1}) \sin(\Theta_i)$. Eq. 5.33 is recovered upon substitution of $\cos(\varphi)$ in $\hat{\mathbf{a}}_i \cdot \hat{\mathbf{a}}_{i+1} = \mathbf{a}_{i+1} \cdot \hat{\mathbf{y}} / |\mathbf{a}_{i+1}|$.

Bibliography

[1] C. Bustamante, Z. Bryant, and S. B. Smith. 10 years of tension: Single-molecule DNA mechanics. *Nature*, 421:423, 2003.

[2] M. Buscaglia, L. Lapidus, W. Eaton, and J. Hofrichter. Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophys. J.*, 91:276–288, 2006.

[3] M. Woodside, W. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci.*, 103:6190, 2006.

[4] I. J. Chang, J. C. Lee, J. R. Winkler, and H. B. Gray. The protein-folding speed limit: Intrachain diffusion times set by electron-transfer rates in denatured $Ru(NH_3)_5$ (His-33)-Zn-cytochrome c. *Proc. Natl. Acad. Sci.*, 100:3838–3840, 2003.

[5] M. A. Winnik. *Cyclic Polymers*. Elsvier, New York, 1986. edited by J. A. Semylen.

[6] P. J. Flory. *Statistical Mechanics of Chain Molecules*. New York. Interscience Publishers, 1969.

[7] T. E. Cloutier and J. Widom. DNA twisting flexibility and the formation of sharply looped proteinDNA complexes. *Proc. Natl. Acad. Sci.*, 102(10):3645–3650, 2005.

[8] L. Lapidus, W. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci.*, 97(13):7220–7225, 2000.

[9] D. Thirumalai and C. Hyeon. Rna and protein folding: Common themes and variations. *Biochemistry*, 44:4957–4970, 2005.

[10] G. Wilemski and M. Fixman. Diffusion-controlled intrachain reactions of polymer. I Theory. *J. Chem. Phys.*, 60(3):866–877, 1974.

[11] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *J. Chem. Phys.*, 72(8):4350–4357, 1980.

[12] M. Doi. Diffusion-controlled reaction of polymers. *Chem. Phys.*, 9:455–466, 1975.

[13] I. Sokolov. Cyclization of a polymer: First-passage problem for a non-markovian process. *Phys. Rev. Lett.*, 90(8):080601, 2003.

[14] A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz, and S. Weiss. Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci.*, 97(10):5179–5184, 2000.

[15] A. Navon, V. Ittah, P. Landsman, H. A. Scheraga, and E. Haas. Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease a. folding initiation structures in the c-terminal portions of the reduced protein. *Biochem.*, 40:105, 2001.

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

[17] C. Hyeon and D. Thirumalai. Chemical theory and computation special feature: Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci.*, 102:6789–6794, 2005.

[18] C. Hyeon, R. I. Dima, and D. Thirumalai. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure*, 14:1633, 2006.

[19] D. Thirumalai and C. Hyeon. RNA and Protein folding: Common Themes and Variations. *Biochemistry*, 44(13):4957–4970, 2005.

[20] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci.*, 103:6190–6195, 2006.

[21] L. J. Lapidus, P. J. Steinbach, W. A. Eaton, A. Szabo, and J. Hofrichter. Effects of chain stiffness on the dynamics of loop formation in polypeptides. appendix:testing a 1-dimensional diffusion model for peptide dynamics. *J. Phys. Chem. B*, 106:11628–11640, 2002.

[22] L.J. Lapidus, W.A. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA*, 97:7220–7225, 2000.

[23] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314:1001–1004, 2006.

[24] S. Mihardja, A. J. Spakowitz, Y. Zhang, and C. Bustamante. Effect of force on mononucleosomal dynamics. *Proc. Natl. Acad. Sci.*, 103:15871, 2006.

[25] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes andD. L. Anderson, and C. Bustamante. The bacteriophage ϕ29 portal motor can package DNA against a large internal force. *Nature*, 413:748, 2001.

[26] N. Marky and G. S. Manning. A theory of DNA dissociation from the nucleosome. *J. Mol. Biol.*, 254:50–61, 1995.

[27] P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev, and R. Phillips. Forces during bacteriophage DNA packaging and ejection. *Biophys. J.*, 88:851–866, 2005.

[28] S. Tzlil, J. T. Kindt, W. M. Gelbart, and A. Ben-Shaul. Forces and pressures in DNA packaging and release from viral capsids. *Biophys. J.*, 84:1616–1627, 2003.

[29] W. Jiang, J. Chang, J. Jakana, P. Weigele, and W. Chiu. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature*, 439:612, 2006.

[30] P. Serwer, S. Khan, S. Hayes, R. Watson, and G. A. Griess. The conformation of packaged bacteriophage t7 dna: Informative images of negatively stained t7. *J. Struct. Biol.*, 120:32, 1997.

[31] S. Köster, D. Steinhauser, and T. Pfohl. Brownian motion of actin filaments in confining microchannels. *J. Phys.: Cond. Matt.*, 17:S4091, 2005.

[32] D. Frenkel and B. Smit. *Understanding Molecular Simulation - From Algorithms to Applications*. Academic Press, San Diego, CA, 2002.

[33] N. Toan, D. Marenduzzo, and C. Micheletti. Inferring the diameter of a biopolymer from its stretching response. *Biophys. J.*, 89:80, 2005.

[34] M. Dijkstra, D. Frenkel, and H. N. W. Lekkerkerker. Confinement free energy of semiflexible polymers. *Physica A*, 193:374, 1993.

[35] M. Doi. *Introduction to Polymer Physics*. Clarendon Press, Oxford, 1996.

[36] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Clarendon Press, Oxford, 1988.

[37] G. B. Arfken and H. J. Weber. *Mathematical Methods for Physicists*. Harcourt, San Diego, 5 edition, 2001.

[38] G. Wilemski and M. Fixman. Diffusion-controlled intrachain reactions of polymers. I: theory. *J. Chem. Phys.*, 60(3):866–877, 1974.

[39] P. Pincus. Excluded volume effects and stretched polymer chains. *Macromolecules*, 9:386–388, 1976.

[40] S. Edwards and P. Singh. Size of a polymer molecule in solution. *J. Chem. Soc. Faraday Trans. II*, 75:1001–1019, 1979.

[41] B. Y. Ha and D. Thirumalai. A mean-field model for semiflexible chains. *J. Chem. Phys.*, 103:9408, 1995.

[42] D. Thirumalai and B. Y. Ha. *Theoretical and Mathematical Models in Polymer Research*. Acadamia, New York, 1988.

[43] P. Debnath and B. Cherayil. Dynamics of chain closure: Approximate treatment of nonlocal interactions. *J. Chem. Phys.*, 120(5):2482–2489, 2004.

[44] Z. Guo and D. Thirumalai. Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers*, 36:83–102, 1995.

[45] N. M. Toan, D. Marenduzzo, P. R. Cook, and C. Micheletti. Depletion effects and loop formation in self-avoiding polymers. *Phys. Rev. Lett.*, 97:178302, 2006.

[46] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci.*, 102(15):5397–5402, 2005.

[47] L. Lapidus, P. Steinbach, W. Eaton, A. Szabo, and J. Hofrichter. Effects of chain stiffness on the dynamics of loop formation in polypeptides. *J. Phys. Chem. B*, 106:11628–11640, 2002.

[48] I-J. Chang, J. C. Lee, J. R. Winkler, and H. B. Gray. Bioinorganic chemistry special feature: The protein-folding speed limit: Intrachain diffusion times set by electron-transfer rates in denatured $Ru(NH_3)_5$(His-33)-Zn-cytochrome *c*. *Proc. Natl. Acad. Sci.*, 100:3838–3840, 2003.

[49] J. C. Lee, B. T. Lai, J. J. Kozak, H. B. Gray, and J. R. Winkler. α-synulein tertiary contact formation. *J. Phys. Chem. B.*, 111:2107–2112, 2007.

[50] R. Hudgins, F. Huang, G. Gramlich, and W. Nau. A flourescence-based method for direct measurement of submicroseond intramolecular contact formation in biopolymers. *J. Amer. Chem. Soc.*, 124(4):556–564, 2002.

[51] S. J. Hagen, C. W. Carswell, and E. M. Sjolander. Rate of intrachain contact formation in an unfolded protein: temperature and denaturant effects. *J. Mol. Biol.*, 305:1161–1171, 2001.

[52] H. Neuweiler, A. Schulz, M. Bohmer, J. Enderlein, and M. Sauer. Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level. *J. Am. Chem. Soc.*, 125:5324–5330, 2003.

[53] D. Shore and R. Baldwin. Energetics of DNA twisting. *J. Mol. Biol.*, 170:957–981, 1983.

[54] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber. The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci.*, 96:9597–9601, 1999.

[55] M. Woodside, P. Anthony, W. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314:1001, 2006.

[56] J. Portman. Non-gaussian dynamics from a simulation of a short peptide: Loop closure rates and effective diffusion coefficients. *J. Chem. Phys.*, 118(5):2381–2391, 2003.

[57] C. Camacho and D. Thirumalai. Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci.*, 92:1277–1281, 1995.

[58] R. Pastor, R. Zwanzig, and A. Szabo. Diffusion limited first contact of the ends of a polymer: Comparison of theory with simulation. *J. Chem. Phys.*, 105(9):3878–3882, 1996.

[59] M. Sakata and M. Doi. Computer simulation of the intramolecular reaction of polymers. *Poly. J.*, 8(5):409–413, 1976.

[60] J. Chen, H.-K. Tsao, and Y.-J. Sheng. Diffusion-controlled first contact of the ends of a polymer: Crossover between two scaling regimes. *Phys. Rev. E*, 72:031804, 2005.

[61] A. Podtelezhnikov and A. Vologodskii. Simulations of polymer cyclization by brownian dynamics. *Macromol.*, 30:6668–6673, 1997.

[62] C. Hyeon and D. Thirumalai. Kinetics of interior loop formation in semiflexible chains. *J. Chem. Phys.*, 124:104905, 2006.

[63] A. Rissanou, S. Anastasiadis, and I. Bitsanis. Monte carlo study of the coil-to-globule transition of a model polymeric system. *J. Poly. Sci. B*, 44(24):3651–3666, 2006.

[64] D. Thirumalai. Time scales for the formation of the most probable tertiary contacts in proteins with applications to cytochrme c. *J. Phys. Chem. B*, 103:608–610, 1999.

[65] Ted A. Laurence, Xiangxu Kong, Marcus Jäger, and Shimon Weiss. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *PNAS*, 102:17348–17353, 2005.

[66] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *PNAS*, 104:105–110, 2007.

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

[68] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, 2001.

[69] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H. E. Gaub. How strong is a covalent bond? *Science*, 283:1727–1730, 1999.

[70] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. The study of protein mechanics with the atomic force microscope. *Trends in Biochem. Sci.*, 24:379–384, 1999.

[71] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276:1109–1112, 1997.

[72] C. Hyeon and D. Thirumalai. Mechanical unfolding of RNA : from hairpins to structures with internal multiloops. *Biophys. J.*, 92:731–743, 2007.

[73] M. Vendruscolo, E. Paci, C. Dobson, and M. Karplus. Three key residues form a critical contact network in a protein folding transition state. *Nature*, 409:641–645, 2001.

[74] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314:1001–1004, 2006.

[75] P. G. de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, Ithaca and London, 1979.

[76] C. Cordeiro, M. Mosilana, and D. Thirumalai. Shape of confined polymer chains. *J. Phys. II (France)*, 7:433–447, 1997.

[77] G. Morrison and D. Thirumalai. The shape of a flexible polymer in a cylindrical pore. *J. Chem. Phys*, 122:194907, 2005.

[78] B. Y. Ha and D. Thirumalai. Conformations of a polyelectrolyte chain. *Phys. Rev. A*, 46:R3012, 1992.

[79] H. Benoit, B. Duplessix, R. Ober, M. Daoud, J. P. Cotton, B. Farnoux, and G. Jannink. Scattered intensity by a polymer chain in a sample with external constraints. *Macromolecules*, 8:451–453, 1975.

[80] J. Hatfield and S. Quake. Dynamic properties of an extended polymer in solution. *Phys. Rev. Lett.*, 82:3548–3551, 1999.

[81] N. Toan and D. Marenduzzo. Inferring the effective thickness of polyelectrolytes from stretching measurements at various ionic strengths: applications to DNA and RNA. *J. Phys.: Cond. Mat.*, 18:S269, 2006.

[82] R. Maurice and C. Matthai. Force-extension curves for a single polymer chain under varying solvent conditions. *Phys. Rev. E*, 60:3165, 1999.

[83] I. Webman and J. Lebowitz. Elastic properties of a polymer chain. *Phys. Rev. A*, 23:316–320, 1981.

[84] D. Ceperley, M. Kalos, and J. Lebowitz. Computer simulation of the dynamics of a single polymer chain. *Phys. Rev. Lett.*, 41:313, 1978.

[85] A. Dua and T. Viglis. Self-consistent variational theory for globules. *Europhys. Lett.*, 71:49–55, 2005.

[86] H. Schiessel and P. Pincus. Counterion-condensation-induced collapse of highly charged polyelectrolytes. *Macromolecules*, 31:7953–7959, 1998.

[87] J. S. Higgins and H. C. Benoit. *Polymers and neutron scattering.* Oxford University Press,, New York, 1994.

[88] A. Halperin and E. Zhulina. Stretching polymer brushes in poor solvents. *Macromolecules*, 24:5393, 1991.

[89] S. Jun and B. Mulder. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proc. Natl. Acad. Sci.*, 103:12388–12393, 2006.

[90] F. Brochard and P. G. de Gennes. Dynamics of confined polymer chains. *J. Chem. Phys.*, 67:52, 1977.

[91] A. Fersht. *Structure and Mechanism in Protein Science : A Guide to Enzyme Catalysis and Protein Folding.* W. H. Freeman Company, 1998.

[92] J.A. Doudna and T.R. Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(11):222–228, 2002.

[93] C. M. Dobson. Protein misfolding, evolution and disease. *Trends Biochem. Sci.*, 24:329–332, 1999.

[94] K. A. Dill and H. S. Chan. From Levinthal to Pathways to Funnels. *Nature Struct. Biol.*, 4:10–19, 1997.

[95] J. N. Onuchic and P. G. Wolynes. Theory of protein folding. *Curr. Opin. Struct. Biol.*, 14:70–75, 2004.

[96] C. Hyeon and D. Thirumalai. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc. Natl .Acad. Sci.*, 100:10249–10253, 2003.

[97] R. Nevo, V. Brumfeld, R. Kapon, P. Hinterdorfer, and Z. Reich. Direct measurement of protein energy landscape roughness. *EMBO reports*, 6(5):482, 2005.

[98] M. Schlierf, H. Li, and J. M. Fernandez. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl. Acad. Sci.*, 101:7299–7304, 2004.

[99] J. Brujić, R. I. Hermans Z, K. A. Walther, and J. M. Fernandez. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nature Physics*, 2:282–286, 2006.

[100] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, Jr., and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, 2001.

[101] B. Onoa, S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco, Jr., and C. Bustamante. Identifying Kinetic Barriers to Mechanical Unfolding of the *T. thermophila* Ribozyme. *Science*, 299:1892–1895, 2003.

[102] D. K. Klimov and D. Thirumalai. Stretching single-domain proteins: Phase diagram and kinetics of force-induced unfolding. *Proc. Natl. Acad. Sci.*, 96(11):6166–6170, 1999.

[103] C. Hyeon and D. Thirumalai. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci.*, 102:6789–6794, 2005.

[104] M. Manosas, D. Collin, and F. Ritort. Force-dependent fragility in RNA hairpins. *Phys. Rev. Lett.*, 96:218301, 2006.

[105] D. K. West, D. J. Brockwell, P. D. Olmsted, S. E. Radford, and E. Paci. Mechanical Resistance of Proteins Explained Using Simple Molecular Models. *Biophys. J.*, 90:287–297, 2006.

[106] O. K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96:108101, 2006.

[107] C. Hyeon and D. Thirumalai. Measuring the energy landscape roughness and the transition state location of biomolecules using single molecule mechanical unfolding experiments. *J. Phys.: Condens. Matter*, 19:113101, 2007.

[108] M. Manosas and F. Ritort. Thermodynamic and Kinetic Aspects of RNA pulling experiments. *Biophys. J.*, 88:3224–3242, 2005.

[109] C. Hyeon and D. Thirumalai. Forced-unfolding and force-quench refolding of RNA hairpins. *Biophys. J.*, 90:3410–3427, 2006.

[110] M. Manosas, J. D. Wen, P. T. X. Li, S. B. Smith, C. Bustamante, I. Tinoco, Jr., and F. Ritort. Force Unfolding Kinetics of RNA using Optical Tweezers. II. Modeling Experiments. *Biophys. J.*, 92:3010–3021, 2007.

[111] K. Svoboda and S. M. Block. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.*, 23:247–285, 1994.

[112] G. Wilemski and M. Fixman. Diffusion-controlled intrchain reactions of polymers. II Results for a pair of terminal reactive groups. *J. Chem. Phys.*, 60(3):878–890, 1974.

[113] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University press, New York, 2001.

[114] O. Kratky and G. Porod. *Rec. Trav. Chim.*, 68:1106, 1949.

[115] H. Yamakawa. *Helical Wormlike Chains in Polymer Solution*. Springer, Berlin, 1997.

[116] I. M. Kulić and H. Schiessel. DNA spools under tension. *Phys. Rev. Lett.*, 92:228101, 2004.

[117] T. Odijk. Physics of tightly curved semiflexible polymer chains. *Macromolecules*, 26:6897, 1993.

[118] C. Lin, Y. Tsai, and C. Hu. Wrapping conformations of a polymer on a curved surface. *Phys. Rev. E*, 75:031903, 2007.

[119] R. Mondescu and M. Muthkuhmar. Brownian motion and polymer statistics on certain curved manifolds. *Phys. Rev. E*, 57:4411, 1998.

[120] A. Spakowitz and Z. Wang. Semiflexible polymer confined to a spherical surface. *Phys. Rev. Lett.*, 91:166102, 2003.

[121] J. Cerdá, T. Sintes, and A. Chakrabarti. Excluded volume effects on polymer chains confined to spherical surfaces. *Macromolecules*, 38:1469–1477, 2005.

[122] E. Katzav, M. Adda-Bedia, and A. Boudaoud. A statistical approach to close packing of elastic rods and to DNA packaging in viral capsids. *Proc. Natl. Acad. Sci*, 103:18900–18904, 2006.

[123] T. Odijk. On the statistics and dynamics of confined or entangled stiff polymers. *Macromol.*, 16:1340, 1983.

[124] T. W. Burkhardt. Free energy of a semiflexible polymer in a tube and statistics of a randomly accelerated particle. *J. Phys. A: Math. Gen.*, 30:L167–L172, 1997.

[125] T. W. Burkhardt. Free energy of a semiflexible polymer confined along an axis. *J. Phys. A: Math. Gen.*, 28:L629–L635, 1995.

[126] A. Lamura, T. W. Burkhardt, and G. Gompper. Helical polymer in cylindrical confining geometries. *Phys. Rev. E*, 70:051804, 2004.

[127] F. Wagner, G. Lattanzi, and E. Frey. Conformations of confined biopolymers. *Phys. Rev. E*, 75:050902(R), 2007.

[128] Y. Tao, N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell*, 95:431–437, 1998.

[129] W. Ernshaw and S. Casjens. DNA packaging by the double-stranded DNA bacteriophages. *Cell*, 21:319–331, 1980.

[130] J. Lepault, J. Dubochet, W. Baschong, and E. Kellenberger. Organization of double-stranded DNA in bacteriophages: A study by cryo-electron microscopy of vitrified samples. *EMBO J.*, 6:1507–1512, 1987.

[131] M. Cerritelli, N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, and A. C. Steven. Encapsidated conformation of bacteriophage t7 dna. *Cell*, 91:271, 1997.

[132] L. W. Black, W. W. Newcomb, J. W. Boring, and J. C. Brown. Ion etching of bacteriophage t4: Support for a spiral-fold model of packaged DNA. *Proc. Natl. Acad. Sci.*, 82:7960, 1985.

[133] L. W. Black and D. J. Silverman. Model for DNA packaging into bacteriophage t4 heads. *J. Virol.*, 643:643, 1978.

[134] Z. Zhang, B. Greene, P. A. Thuman-Commike, J. Janka, P. Prevelige, J. King, and W. Chiu. Visualization of the maturation transition in bacteriophage p22 by electron microscopy. *J. Mol. Biol.*, 297, 2000.

[135] J. Widom and R. L. Baldwin. Tests of spool models for DNA packaging in phage λ. *J. Mol. Biol*, 171:419, 1983.

[136] E. C. Mendelson, W. W. Newcomb, and J. C. Brown. Ar^+ plasma-induced damage to DNA in bacteriophage λ: Implications for the arrangement of DNA in the phage head. *J. Virol.*, 66:2226, 1992.

[137] S. L. Novick and J. D. Baldeshwieler. Flouresence measurement of the kinetics of DNA injection by bacteriophage λ into liposomes. *Biochem.*, 27:7919, 1988.

[138] J. C. Brown and W. W. Newcomb. Ion etching of bacteriophage λ: Evidence that the right end of the DNA is located at the outside of the phage DNA mass. *J. Virol.*, 60:564, 1986.

[139] N. V. Hud. Double-stranded dna organization in bacteriophage heads: An alternative toroid-based model. *Biophys. J.*, 69:1355, 1995.

[140] J. Kindt, S. Tzlil, A. Ben-Shaul, and W. M. Gelbart. DNA packaging and ejection forces in bacteriophage. *Proc. Natl. Acad. Sci*, 98:13671–13674, 2001.

[141] J. C. LaMarque, T. L. Le, and S. C. Harvey. Packaging double-helical DNA into viral capsids. *Biopolymers*, 73:348–355, 2003.

[142] C. Forrey and M. Muthukumar. Langevin dynamics simulations of genome packaging in bacteriophage. *Biophys. J.*, 91:25–41, 2006.

[143] W. S. Klug and M. Ortiz. A director-field model of DNA packaging in viral capsids. *J. Mec. Phys. Solids*, 51:1815–1847, 2003.

[144] P. K. Purohit, J. Kondev, and R. Phillips. Mechanics of DNA packaging in viruses. *Proc. Natl. Acad. Sci*, 100:3173–3178, 2003.

[145] P. K. Purohit, J. Kondev, and R. Phillips. Force steps during viral DNA packaging? *J. Mech. Phys. Solids*, 51:2239–2257, 2003.

[146] T. Odijk. Hexagonally packed DNA within bacteriophage t7 stabilized by curvature stress. *Biophys. J.*, 75:1223–1227, 1998.

[147] D. Marenduzzo and C. Micheletti. Thermodynamics of DNA inside a viral capsid: The role of DNA intrinsic thickness. *J. Mo. Biol.*, 330:485–492, 2003.

[148] S. C. Riemer and V. A. Bloomfield. Packaging of DNA in bacteriophage heads: Some considerations on energetics. *Biopolymers*, 17:785, 1978.

[149] R. Metzler and P. G. Dommersnes. Helical packaging of semiflexible polymers in bacteriophages. *Eur. Biophys. J.*, 33:497, 2004.

[150] T. Sakaue. Semiflexible polymer confined in closed spaces. *Macromol.*, 40:5206, 2007.

[151] M. M. Inamdar, W. M. Gelbart, and R. Phillips. Dynamics of ejection from bacteriophage. *Biophys. J.*, 91:411–420, 2006.

[152] A. J. Spakowizt and Z. G. Wang. DNA packing in bacteriophage: Is twist important? *Biophys. J.*, 88:3192, 2005.

[153] B. Y. Ha and D. Thirumalai. Semifexible chains under tension. *J. Chem. Phys.*, 106:4243, 1997.

[154] P. Hansen and R. Podgornik. Wormlike chains in the large d-limit. *J. Chem. Phys*, 114:8637, 2001.

[155] R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals.* McGraw-Hill, New York, 1965.

[156] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, San Diego, 2 edition, 2002.