A Coarse-Grained Model for DNA Loop Extrusion

Thesis submitted to
Indian Institute of Science Education and Research Pune
in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by
Adrian John Pinto

Supervisors: Dr. Apratim Chatterji and Dr. Peter Virnau
Department of Theoretical Physics, IISER Pune
Institute of Physics, JGU

Certificate

This is to certify that this dissertation entitled A Coarse-Grained Model for DNA Loop Extrusion towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study carried out by Adrian John Pinto at Indian Institute of Science Education and Research and Johannes Gutenberg University Mainz under the supervision of Dr. Apratim Chatterji, Associate Professor, Department of Physics, IISER Pune, and Prof. Dr. Peter Virnau, Professor, Institute of Physics, JGU Mainz, during the academic year 2023-2024.

Adrian John Pinto

Dr. Apratim Chatterji

Dr. Peter Virnau

Commitee:

Dr. Apratim Chatterji

Dr. Peter Virnau

Dr. Vijayakumar Chikkadi

To Jolene and Jason

Declaration

I hereby declare that the matter embodied in the report entitled A Coarse-Grained Model for DNA Loop Extrusion are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, and Institute of Physics, Johannes Gutenberg University, Mainz, under the supervision of Dr. Apratim Chatterji and Dr. Peter Virnau and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Adrian John Pinto
20191015

Acknowledgements

I would like to thank my supervisors, Dr. Apratim Chatterji and Dr. Peter Virnau for their guidance and advice. Their continued support and invaluable feedback has kept me motivated throughout the project and has enabled me to finish the thesis. I would like to thank Maurice Pierre Schmitt for being an amazing mentor, always open for discussions should I get stuck on any problem. Danke, Mauriq! I would also like to thank Dr. Biswajit Pradhan and his group for the experimental data and insightful meetings which helped us steer the simulations in the right direction. I extend my gratitude to every member of Dr. Chatterji's as well as Dr. Virnau's group. Without their continued support, I wouldn't have been able to achieve this feat.

The thesis wouldn't have been possible if it wasn't for the super-computing facility at both the institutes. Parambrahma and Iona helped tremendously in my endeavors, so I would like to acknowledge them. I would like to thank KVPY for helping my studies with scholarship as well as SFB 1551 for the financial support during my thesis. And lastly, I would like to thank my family and friends, whose unconditional support always help me do my best.

Abstract

We craft a coarse-grained model which captures the statics and dynamics of loop extrusion experiments performed at the Max Planck Institute of Biophysics quantitatively. After getting a working model for extrusion, we parameterize our model to the experimental results of Biswajit et al. and attempt to uncover the physical mechanisms behind the observations. In the experiments, the DNA is grafted at two ends and exposed to ATP along with the loop extruding protein SMC 5/6, where the extrusion stalls at a particular loop size. Using our parameterized values, we have managed to map the loop sizes of our simulation with those of the experiment, and explain the origin for stalling of the extrusion. We also conduct detailed analyses of the model dynamics employing principles of statistical physics, elucidating the underlying mechanisms driving the observed simulation outcomes. Additionally, we propose a model for generating torsional strain and producing coiled structures within the polymer. This is done to generate physically relevant structures called plectonemes, which are formed in DNA.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Overview

Deoxyribonucleic acid, or DNA, has been the focal point of molecular biology research since its structure was elucidated by Watson and Crick in the year 1953 [1]. There always has been growing curiosity regarding the packaging of a molecule made up of large number of base pairs into a micron-sized cell [2]. In humans, for instance, the DNA is approximately 2 meters long and is packed into a cell of a few microns, which seems like a monumental feat at first glance. But there are a surprisingly large number of processes that are responsible for the compaction as well as maintaining the necessary structural form that is crucial for the functioning of the molecule that has evolved to become the genetic molecule [3].

The canonical DNA is a right-handed double helix made up of two chains of polynucleotides that run anti-parallel to each other. These chain are held together by hydrogen bonds between the bases, commonly called base pairs. Each base pair is 0.34nm [3] in size, and the number of base pairs effectively decides the length of the DNA or vice versa. In the experiments performed by Pradhan et al. [4], λ-DNA was used, which has a total of 48502 base pairs, hence translates to a total length $16.2\mu m$. In the coarse-grained simulation scheme, these base pairs precisely form the monomers, albeit a few tens will correspond to one monomer.

The existence of the DNA in a highly folded state is not a random process. It occurs not only for DNA packaging, rather it is an active and regulated process which is responsible for maintaining the structure as well as the regulatory functions of the cell [5]. The folding of the DNA occurs hierarchically, where the double helix is wrapped around histone octamers to form nucleosome chromatin fibres of around few nm length scale, which are further folded to form topologically associating domains (TADs) [6]. The folding of the chromatin fibres is an active process regulated by the structural maintenance of chromosomes (SMC) protein complexes. During this process the DNA extrudes loops, which will be the focus of the thesis.

The helical state of the DNA is also responsible for torsional strain along the contour. Normally, the DNA is said to be in a torsionally relaxed state. Supercoiling occurs when the tor-

sional strain deviates from the relaxed state and either results in the formation of overwound or underwound regions [7]. These configurations are basically consequences of the conservation of the linking number, but they have profound implications on the most crucial process in the cell such as transcription. In a section of the thesis, an attempt at modeling supercoiling will also be presented, along with the drawbacks stating why the model did not work as intended.

1.1 Experimental Overview and Motivation

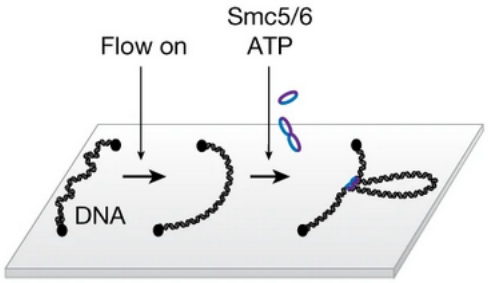
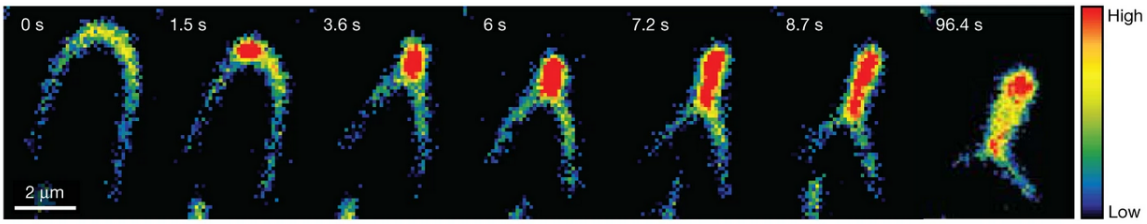


Figure 1.1: **Schematic of experimental setup with a grafted DNA in the presence of flow.** The figure has been reproduced with permission from the author [4].

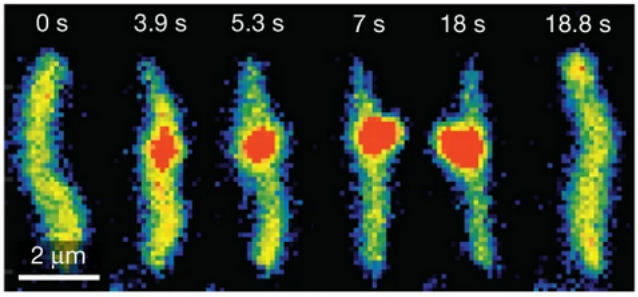
The SMC complexes are primarily responsible for chromosome organization, regulation of genomic processes, segregation of chromosomes as well as DNA repair [5]. In the experiments performed by Pradhan et al. [4], the SMC 5/6 complex isolated from *Saccharomyces cerevisiae* was used to perform the extrusion experiment.

The ends of the λ-DNA molecule were grafted to a passive glass surface in the setup. A buffer flow was applied perpendicular to the axis of the DNA and the molecule was allowed to stretch, followed by imaging through total internal reflection microscopy. Later, the Smc5/6 complex was added along with ATP and the observation was that the DNA was initially concentrated into one spot, and gradually elongated into a loop, which is an obvious indication of the extrusion. The visualization was also done in the absence of buffer flow, in which case the loosely compact DNA punctum increased in size over the course of the experiment [4]. The experimental images are displayed in figure 1.2. These clearly show the onset of extrusion in the presence of Smc5/6 complex and ATP.

The goal of this project is to craft a coarse-grained model that can explain the dynamics as well as the statics of the experiments performed by Biswajit et al. at the Max Planck Institute of



(a) Images of DNA extrusion under buffer flow



(b) Images of DNA extrusion without flow

Figure 1.2: **Series of Images of DNA extrusion mediated by Smc5/6 complex.** The snapshots are that of the experiment imaged by total internal microscopy. The figures have been reproduced with permission from the author [4].

Biophysics. The motivation behind this work is to get a predictive model of DNA loop extrusion which is able to reproduce experiments at the experimental length and time scale. Further, we aim to understand the mechanism behind the stalling of extrusion. In the long term, we aim to understand how the loop extrusion interacts with the DNA topology to efficiently get packed inside the nuclear volume or form Topologically Associated Domains (TADs) and how the loop extrusion would proceed should it encounter entangled regions, and how it might change the conformation of the DNA. We want to achieve this via polymer simulation and using a coarse-grained description that can be parameterized effectively to match the experimental time and length scales.

1.2 Thesis Outline

In the fist chapter, a brief overview of the experimental setup as well as the methods used to perform the experiments will be presented without going over the biological details behind the findings.

The second chapter entails the theoretical description of the simulation methods as well the modeling details used to perform extrusion simulation. In section 2.1, we present the molecular dynamics schema and briefly explain Langevin and Brownian Dynamics, which are the methods used to simulate the polymer. Moving on, in section 2.2, we go over the loop extrusion model and present every component of the model under separate subsections. In section

2.2.4, we also explain the basis for choosing the simulation parameters for both DNA displacement experiment as well the extrusion experiment. To conclude the section, we present our attempt at modeling torsion in the polymer chain.

The third chapter encompasses the results accompanied by discussion, where we analyze all our findings as well as the plots in depth. Firstly, in section 3.1, we present an easy method to calibrate the length scale of the simulation with that of the experiment, each subsection explaining how we did it for two different experimental conditions. Following the length scale calibration, we calibrate the time scale 3.2, where we see a mismatch between the DNA displacement time scale as well as extrusion time scale. We briefly address this issue, and present ways to improve the speed of the simulation. In section 3.3, we see how the active force influences the relative extension and analyze the three different regimes of the active force. This section is immediately followed by tension analysis in 3.4, where we provide evidence for some of the claims made in the section 3.3 by analyzing the tension in the polymer chain.

In section 3.5, we see how the extrusion speed changes during the extrusion simulation, and compare it with the experiment to see both qualitative as well as quantitative matches. We analyze the effect of discrete active force in section 3.6, where the extrusion force was applied every nth iteration. In section 3.7, we plot the probability distribution of the loop sizes, and thereby free energy, in the presence and absence of active forces. We move on to present an indirect way to measure the stalling forces in section 3.8 and provide evidence for why extrusion stalls in section 3.9 using force-extension curves, followed by effects of grafting distance on the loop size in 3.10. In the following section, we simulation the extrusion in the presence of flow in order to replicate the experimental system in the presence of buffer flow. In section 3.11, we investigate entropy and loops to show the effects of entropy in deciding the conformation of a ring polymer.

Lastly, in section 3.12 we analyze the results of the supercoiling model, and present the prominent drawbacks of the model, and explain why it did not represent the experimental model. We briefly explain the ways to form plectoneme like structures in sections 3.11.3 and 3.11.4 before moving on to the conclusion in chapter 4.

Chapter 2

Methodology and Computational Schemes

In this chapter, we present the salient aspects of the model as well the integration schemes used to replicate the experiments. Using polymer models to simulate the DNA has been around for decades now, and we make an attempt to build upon the existing DNA models and implement necessary modifications to simulate loop extrusion. We also come up with a model to implement four-body potentials to simulate torsion in the contour.

2.1 Molecular Dynamics Simulation Schemes

Molecular dynamics has been extensively used to study systems which are hard to solve analytically. Prominent among the subclass are many body problems, where the dynamics of the system can be obtained by integrating the positions and velocity of particles over time numerically. Over the past few years, molecular dynamics methods have given valuable insights on polymeric systems [8], bio-molecules [9], and other many body systems owing to the extensive progress in simulation methods and hardware. At the core of the method, the molecular dynamics entails integrating equations of motions using integration schemes like Velocity Verlet [10] to calculate positions and velocity in small time steps of Δt.

The position and velocity of the particles in the simulation can be updated using the Velocity Verlet integration scheme. The derivation is described in the appendix, but the result can be summarized as [11],

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2m}\vec{F}(t)\Delta t^2 \tag{2.1}$$

$$\vec{v}(t+\Delta t) = \vec{v}(t) + \frac{1}{2m}\left[\vec{F}(t+\Delta t) + \vec{F}(t)\right]\Delta t. \tag{2.2}$$

2.1.1 Langevin Dynamics

Langevin equations describe the trajectories of the particles subjected to damping and noise. These contributions are added to the standard Newton's equations of motion and, in the numer-

ical schema, integrated using the Velocity Verlet or any other standard integration techniques. The Langevin equation is given as [12],

$$\vec{F}_{\text{noise}} + \vec{F}_{\text{damping}} = m\vec{a} \tag{2.3}$$

The damping term can be taken to be proportional to the particles velocity (negative!), hence by introducing the damping constant γ, one can write the equation as,

$$\vec{F}_{\text{noise}} - \gamma\vec{v} = m\vec{a} \tag{2.4}$$

In the simulations, the particles are integrated forward in time using these equations. The equations essentially have a built in thermostat to maintain the temperature, called Langevin Thermostat, which translates to a canonical ensemble (NVT). In the Langevin equation, we can add a total force term to incorporate the potentials and constraints in the system. In HOOMD-blue [13], the equation for integration is as follows,

$$m\frac{d\vec{v}}{dt} = \vec{F}_{\text{T}} - \gamma\vec{v} + \vec{F}_{\text{R}} \tag{2.5}$$

where F_T is the total force due to all the potentials and constraints, and F_R is the random force, $\langle |F_R|^2 \rangle = 6k_BT\gamma/\Delta t$. The random force proportional to k_BT is chosen in accordance with the fluctuation-dissipation theorem, which has been briefed over in the appendix. The thermal noise gives the particles velocities from a thermal bath at the desired temperature and the damping term balances it, thereby maintaining a temperature that fluctuates around k_BT.

2.1.2 Brownian Dynamics

In the over-damped limit, the damping terms dominate the inertial terms and the inertial terms can be ignored. The corresponding equation in the absence of inertial terms is the Brownian equation. If we ignore the inertial term, then the particles will accelerate quickly and reach a velocity dictated by the damping constant alone. This is the diffusive regime, and the particles will thermalize with the bath. The equation can be written as [12],

$$\gamma v(t) = F_{\text{external}} + F_{\text{noise}} \tag{2.6}$$

where F_{external} is the culmination of all the external forces acting on the particles.

2.2 The Model: Loop Extrusion

In a real biological system, the DNA is extruded via a family of proteins called SMC or structural maintenance of chromosomes. The elucidation of the complete structure is beyond the scope of this model, but we use a coarse-grained approach to effectively reproduce dynamics as well as statics of the loop extruding proteins. In our model, a bead-spring polymer is used to model the DNA which is a common approach in many simulations of DNA [14] [15]. In computer simulations, the SMC protein has the shape of a handcuff, with two rings connected to each other. This is modeled using rigid body dynamics as elaborated in the following section. We use HOOMD-blue [13] to model our polymer. The package can run simulations of particles on both CPU and GPU.

2.2.1 Semi-Flexible Polymer: The DNA

The DNA is simulated as a semi-flexible polymer using a coarse-grained description of the base pairs. The method of mapping the base pairs into monomers is outlined in the section on parameters. The beads are modeled as hard spheres with an interaction potential between them given by a WCA potential of the form [8],

$$V_{\mathrm{WCA}} = \begin{cases} 4\varepsilon\left[\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^{6}}{r^{6}}\right] + \varepsilon, & \forall r < 2^{1/6}\sigma \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

The potential ensures that the monomers of the chain do not overlap each other. The inter-particle forces can be found just by taking the negative derivative of the above potential written as,

$$F_{\mathrm{WCA}} = 4\varepsilon\left[\frac{12\sigma^{12}}{r^{13}} - \frac{6\sigma^{6}}{r^{7}}\right], \quad \forall r < 2^{1/6}\sigma. \tag{2.8}$$

The σ of the beads was set to 1 indicating a radius of 0.5 for each bead. The ε for simulation was also set to 1, which is the constant dictating the strength of the potential. The mass of the beads was set to 1 in simulation units.

We introduce a FENE potential between adjacent monomers to keep monomers connected to each other forming a linear chain polymer. The FENE potential has the form [16],

$$V_{\mathrm{FENE}}(r) = -\frac{1}{2}kr_0^2 \ln\left[1 - \left(\frac{r}{r_0}\right)^2\right]. \tag{2.9}$$

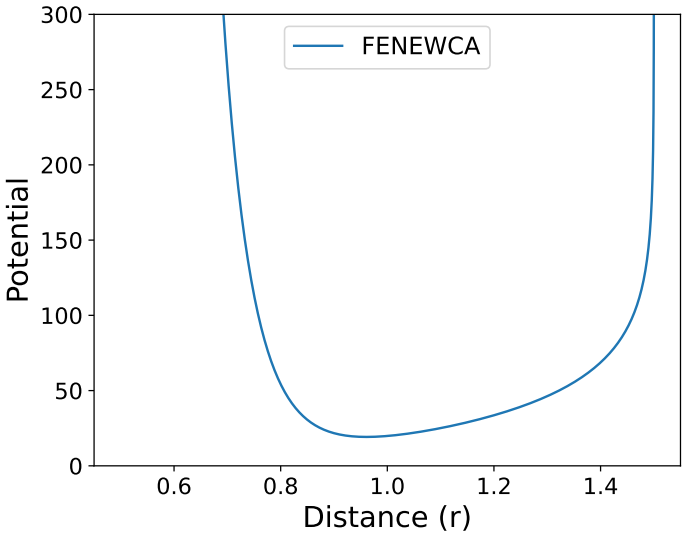


Figure 2.1: **FENEWCA potential curve for parameter values k = 30, r_0 = 1.5, and ε = 30.** The FENE + WCA potential is used to simulate the bead-spring polymer.

The value of k is set to 30, the r_0 is 1.5 and the potential combined with WCA is given in the figure 2.1. The FENE force is the negative derivative of the potential and can be written as,

$$F_{\text{FENE}}(r) = -\frac{kr_0^2}{r_0^2 - r^2} r. \tag{2.10}$$

The polymer simulated with the above potentials is just a flexible polymer. In order to make it semi-flexible as required to model DNA, a bending cost is introduced via a potential of the form

$$V_{\text{Bending}} = \varepsilon_b(1 - \cos(\theta - \pi)). \tag{2.11}$$

The implementation of the potential in HOOMD-blue [13] is slightly different because of the difference in the angle convention, but the underlying physics remains the same. Assuming the angle between the three adjacent monomers to be θ, the potential is minimal when all the monomers are in a straight line, i.e. $\theta = 180°$. There is an energy cost for any angles that deviate from this minima, hence the polymer has semiflexibility incorporated into it by the potential. The choice of the strength of the potential has been elaborated in the later section, as this determines the persistence length of the polymer and we had to match ours with that of experimental DNA.

2.2.2 The Handcuff: SMC Protein Complexes

The handcuff is composed of two rigid rings of ten particle each, connected to each other using FENE potential at two anchors. The anchors are connection between two monomers of different rings colored red in the figure 2.2, and they are connected using FENE potential. There is a

bending cost associated with the anchors, which ensures that the preferred configuration of the handcuff is the one in which both the rings lie in the same plane. To implement the bending cost, the three body potential used for the semiflexibility of the polymer is used here. The mass of each of the rings is set to 2.75, which makes them almost 3 times as heavy as the monomers themselves.

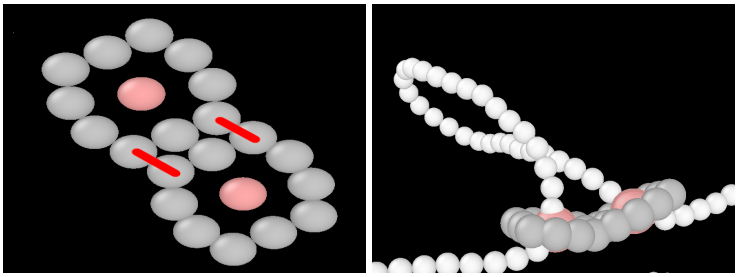


Figure 2.2: **The handcuff snippet from the simulation captured using ovito [17].** The ring particles are colored grey, and the central particle used in the integration scheme is colored peach. The anchors holding the two rings together is colored red.

The rigid body (rings in our case) is an incompressible body made up of a central particle and a set of constituent particles. All the constituent particles interact with every other particle in the system, but the integration scheme in HOOMD-blue [13] computes the net force and torque experienced by all the constituent particles and integrates the position and velocity of the central particle accordingly. The semi-flexible polymer is inserted into this handcuff, starting from the first ring and into the second, so that the polymer forms a loop.

2.2.3 The Wall: Grafting the DNA

The simulation is done in a large box but the first and the last monomers of the polymer are grafted to a wall. The wall-monomer interaction is turned off for the monomers in contact with the wall. In line with the experiments, the wall is located in the xz plane, and both the monomers are attached to the same wall. The interactions of the monomers with the wall are mediated by 9/3 form of the usual Lennard-Jones potential [18]. The name is derived from the exponents used in the standard Lennard-Jonnes potential. The attractive part is to the power of 9, while the repulsive part is to the power of 3. In the most general form [19], the potentials of such kind are called Mie potentials, and the implementation of the potential found in HOOMD-blue [13] has the form,

$$U(r) = \left(\frac{n}{n-m}\right)\left(\frac{n}{m}\right)^{\frac{m}{n-m}} \varepsilon\left[\left(\frac{\sigma}{r}\right)^{n} - \left(\frac{\sigma}{r}\right)^{m}\right]. \tag{2.12}$$

The values of n and m were set to 9 and 3 respectively, while the usual values of $\varepsilon = 1$ and $\sigma = 1$ were used.

2.2.4 Deciding The Parameters

The general model used in this work was developed by Wettermann et al. [15]. In this work, a Kratky-Porod-like chain was parameterized and tested by reproducing experimental knotting probabilities. The solvent conditions of the experiment had to be taken into account in our model since the solvent present around the DNA screens the electrostatic repulsion of the charged DNA. In order to account for this, we used the implicit solvent approach of Stinger [20] and Rybenkov [21], also used by Wettermann et al. [15], where the excluded volume, the diameter of the DNA d, is used to represent these screened interactions. The diameter of the DNA is a function of the salt concentration as shown in figure 2.3 and this dependence has been explicitly mapped in the paper [20].

To match the persistence length of the simulation with that of the DNA, we need to know the

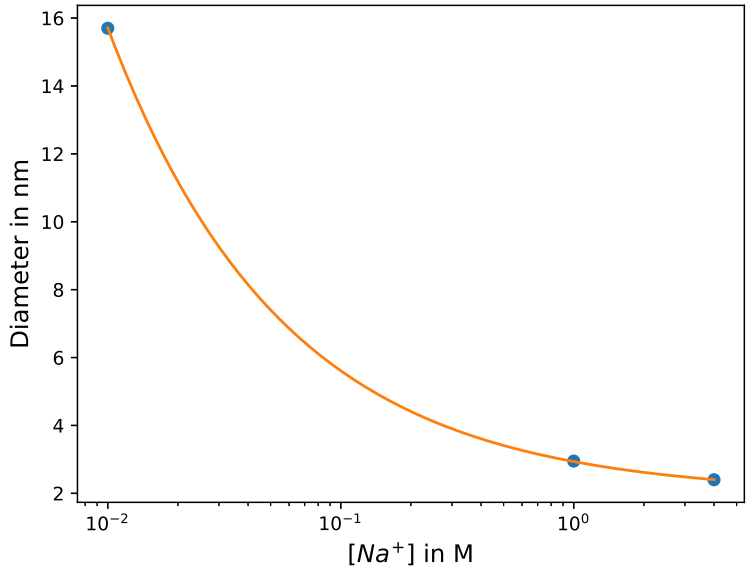


Figure 2.3: **Effective diameter of the DNA as a function of salt concentration** [20]. At higher salt concentrations, the repulsive interactions between the base pairs are screened, consequently decreasing the effective diameter of the DNA. The plot has been reproduced using values from Wettermann's paper [15].

bending potential used to model semi-flexibility, which in our case, is given by a $\cos\theta$ potential. The energy needed for the polymer to go from a linear configuration to a bent configuration is given by the thermal energy. Hence by simple algebra and using cosine expansion [22], we can write,

$$\varepsilon_b(\cos\theta - \cos\pi) \approx k_b T.$$

If we assume $\beta = \pi - \theta$, we can write,

$$\begin{aligned} \varepsilon_b(-\cos\beta + 1) &\approx k_b T \\ \varepsilon_b(-1 + \beta^2/2 + 1) &\approx k_b T \\ \beta^2 &\approx 2k_b T/\varepsilon_b. \end{aligned}$$

From the results of worm-like chain [23], the persistence length for a small angle of bending is given by,

$$l_p = \frac{2d}{\beta^2}. \tag{2.13}$$

Hence, plugging this in the above algebra, we get the relation between the persistence length and bending epsilon

$$l_p \approx \frac{\varepsilon_b d}{k_b T}. \tag{2.14}$$

There are two experiments performed by our collaborators: the DNA displacement experiment and the extrusion experiment [4]. In the fluctuation experiment, the DNA was grafted at the two ends, and the maximum extent of fluctuation was calculated. The exact method is elaborated in a later section. In the extrusion experiment, loop extrusion was studied in the presence of SMC 5/6 and ATP. Both the experiments were performed under different conditions, so we calculate the parameters for each of them here.

In the fluctuation experiment performed on grafted DNA, the salt concentrations are given in the table. The total Na^+ concentration used in the experiment is 55 mM, and from the figure 2.3, the diameter of the DNA for the given salt concentration is 7.13 nm. Each base pair of the DNA accounts for $1/3$ nm, hence the total number of monomer in the simulation can be written as,

$$N \approx \#base pairs \frac{1\text{nm}}{3d}. \tag{2.15}$$

Number of base pairs	48502
Salt Concentration	50 mM NaCl, 2.5 mM $MgCl_2$
Intercalating Dyes	100 nM Sytox Orange

Table 2.1: **The experimental conditions for DNA displacements.**

Hence, the total number of monomers in the simulation are approximately 2269, and every monomer in the simulation corresponds to ≈ 23 base pairs.

The presence of intercalating dye brings down the persistence length of the DNA from 50 nm to 35 nm. Plugging this value in the equation for calculating bending epsilon, we get,

$$\varepsilon_b \approx 35 \text{ nm}/7.13 \text{ nm}$$
$$\varepsilon_b \approx 4.911.$$

Number of monomers	2269
Graft distance	645
FENEWCA constant between monomers	30
FENEWCA r_0	1.5
κ bending	4.911
WCA σ	1
WCA ε	1

Table 2.2: **The simulation parameters for DNA displacements.**

In the extrusion experiments [4], the experimental parameters used were different from the ones used for the fluctuation experiment. The salt concentration was higher, owing to which we had to change the coarse graining and equilibrate the polymer before performing the extrusion. Our collaborators used 100 mM of NaCl and 7.5 mM $MgCl_2$, along with some intercalating dye, which reduced the persistence length.

Number of base pairs	48502
Salt concentration	100 mM *NaCl*, 7.5 mM $MgCl_2$
Intercalating dyes	100 nM Sytox Orange

Table 2.3: **The experimental conditions for DNA extrusion experiment.**

Accordingly, the diameter of the DNA from figure 2.3 accounts to 5.3436 nm, which in turn would correspond to 3026 base pairs from equation 2.15. Substituting this value of d in the equation to get the strength of the bending potential, we get,

$$\varepsilon_b \approx 35 \text{ nm}/5.344 \text{ nm}$$
$$\varepsilon_b \approx 6.549.$$

In summary, the simulation parameters used to simulate the extrusion experiment are given in table 2.4.

2.2.5 Initialization and Equilibration

The polymer has to be initialized before the simulation can start. This can be achieved in various ways, but the system has to be equilibrated to ensure that it reaches a stable and repre-

Number of monomers	3026
Graft distance	910
FENEWCA constant between monomers	30
FENEWCA r_0	1.5
κ bending	6.549
WCA σ	1
WCA ε	1

Table 2.4: **The simulation parameters for DNA extrusion experiment.**

sentative state before any of the extracted data can be analysed. Once an equilibrium state has been achieved from any initial state, the thermodynamic properties stabilize and do not change drastically. This should reflect the true behavior of polymers, or in our case, of DNA in real situations.

The entire chain was initialized as a rectangle, with the first and the last monomer attached to the wall. We repeated the initialization for both the set of parameters, but the figure 2.4 is plotted for parameters of the extrusion experiment, where there are 3026 monomers and the grafting distance is 910. Using Langevin integration scheme of HOOMD-blue [13] for the polymer with dt = 0.01, we ran the simulation for 1×10^9 iterations. To ascertain whether we had run the simulation long enough, Rg^2 of the polymer was plotted. The radius of gyration is a measure of the spatial compactness of the polymer chain. It is the average distance of the monomers from the center of mass of the polymer. The formula for calculating the radius of gyration can be written as follows [23],

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_{cm})^2}. \tag{2.16}$$

Once the radius of gyration of the system had stabilized as in figure 2.4, the positions of the monomers were extracted to be used as an initial configuration in the rest of the simulation. The high values of Rg^2 are caused by the grafting of the DNA at the two ends. The maximum distance between the center of mass and the last monomer close to the wall can go up to half of the grafting distance (450 a.u.). The initialization was done without the handcuff because the chain by itself had to reach equilibrium before the insertion of the handcuff. The handcuff was inserted once the final configuration of the equilibrated chain was attained. This configuration is shown in the figure 2.5.

2.2.6 The Active Forces

The SMC protein complex extrudes the DNA via the use of ATP. In our simulation, the two rings of the handcuff act as the openings through which the polymer is supposed to be pulled.

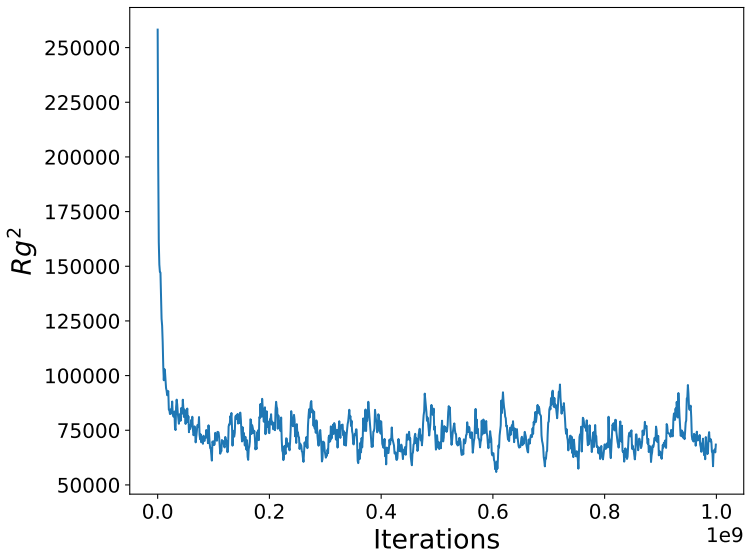


Figure 2.4: **The square of the radius of gyration plotted against the iterations.** The data was extracted every 10^5 time steps and we ran the simulation for 1×10^9 time steps. The fluctuations clearly signify that the system has attained equilibrium. The high values on the y-axis are caused by the grafting distance since the maximum distance between the center of mass and the monomer close to the wall can be around half the grafting distance, which is set to 910.

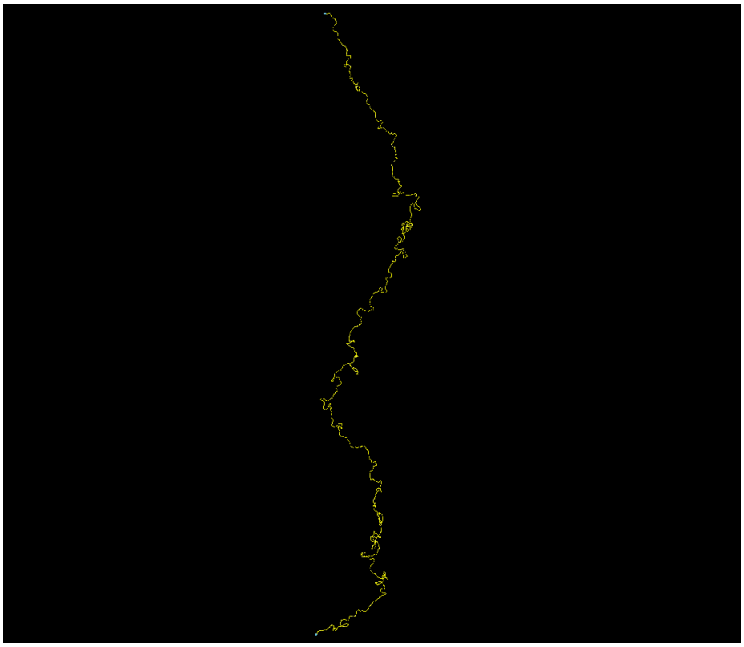


Figure 2.5: **A simulation snapshot of the final configuration captured using ovito [17].** The grafting distance is set to 910 and there are 3026 monomers.

This is done by computing the normal component of the two rings and then applying a force to the monomer in proximity to the ring along the normal direction. Equal and opposite force is also applied to the ring for equivalence. Both the rings had separate amplitudes for this force, so it is possible to simulate proteins that pulled the DNA on one side as well as the proteins

pulling from both sides using our approach. In order to account for the periodic binding and unbinding of the ATP to the proteins, it is possible to apply this active force periodically and study the effects.

The form of the force is just a constant force, and there is no definite potential we are using. The magnitude of this force is a parameter to be tuned, and the choice plays an important role in deciding the dynamics of the chain. This has been explored in the later sections, and many results from the simulation are matched back and forth with the experiments. The form of the potential used here can be explored in depth, but in our coarse-grained model, we used just the most basic form possible.

2.2.7 The Flow Force

The extrusion experiments were also performed under buffer flow [4] in a particular direction. This allowed the DNA to open up reliably, and the extrusion mediated by the SMC complex was more evident in the images shown in the experimental overview section. In order to simulate the buffer flow, we apply a constant force in the preferred direction and allow the system to relax. The magnitude of the force was set to a small value of about 0.0001 so that there would not be any unphysical effect due to the flow. The constant force is applied to every monomer as well as the handcuff, since in the experiments, the buffer flow (almost) equally affects the constituents (we have ignored any effects due to the walls).

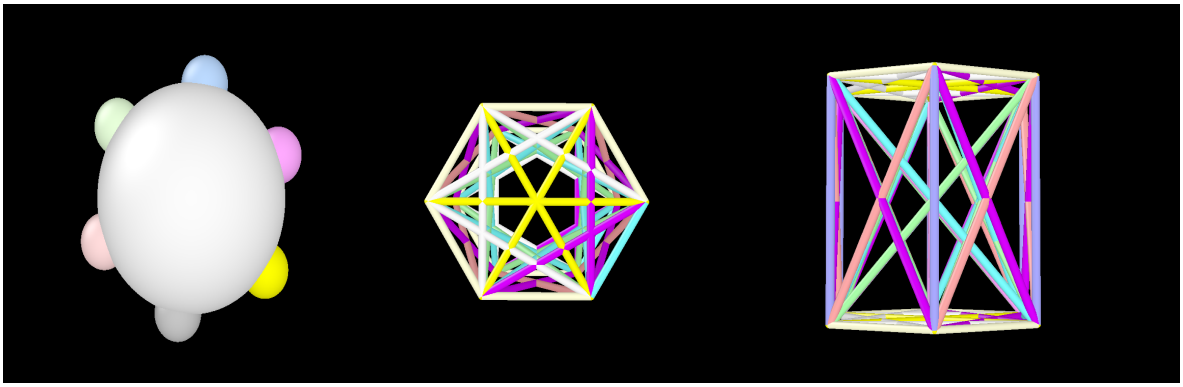
2.3 The Model: Supercoiling

Besides working on the extrusion model, we also investigated supercoiling in this work which is presented as an independent section. As we saw in the previous section on loop extrusion, the DNA is usually modeled using a bead-spring polymer, where monomers are interconnected by some potentials like FENE or harmonic. Here, the basis for using the polymer picture remains the same, but our ultimate goal is to model the coiled structures called plectonemes that are formed in the DNA because of the torsional strain. We use ideas from the model by Mitra et al. for modeling the helix formation [22] and Sinha et al. for modeling hydra dynamics [24].

In our usual bead spring picture, the spring is attached from the center of one monomer to the center of another. This assures that there is rotational symmetry, owing to which there would be no way to introduce torsion without the inclusion of some other four body potentials. So, even if the adjacent monomers were rotated in opposite directions, there would be no change in the distance, as such, there would be no change in the potential and thereby no force. We tried to break this symmetry by connecting the springs from the periphery of the monomers rather than only the center, and not just by one, but instead by more than one spring.

This would reasonably cause some effects to arise due to the rotation of the monomers.

The virtual particles placed on the periphery of the central monomer acted as the adhering



(a) Particles (b) Top View of Bonds (c) Side view of Bonds

Figure 2.6: **A schematic of the supercoiling model.** The central big particle is the monomer and the smaller beads are the virtual particles that act as adhering points for the springs. The figures (b) and (c) show the presence of harmonic bonds between the virtual particles.

point for the spring potential to act. In all the springs, we used a harmonic potential, which is of the form,

$$V(r) = \frac{1}{2}k(r - r_0)^2 \tag{2.17}$$

where all parameters were tunable, and the value of r_0 was set to 0.8 for linear springs. By linear springs, we mean the springs between two overlapping virtual particles when the monomers are exactly stacked with all the virtual particles overlapping in the default configuration. The choice of r_0 for the remaining springs was decided once we set the value for linear springs.

Considering only linear connections, if the adjacent monomers were rotated right now, all the connections would be stretched by equal amounts. The way to restore the potentials back to the initial configurations would be either to rotate back to the initial configuration before rotation or just close the distance between the monomers. Both scenarios are indistinguishable for the potential as it only keeps track of the distance without actually knowing about the inherent dynamics. But our intention is to generate the torsion, so that there is corresponding restoring torque acting on the monomers.

One way to do this would be to have diagonal connections. This would mean having harmonic potential between non-adjacent virtual particles, or in simple words springs. Now, let us analyze again what would happen when we rotate the adjacent monomers. The linear springs would be extended, and the diagonal springs would contract. The only way to relieve the system of stress would be to restore the initial configuration, as moving the monomers closer wouldn't be favorable to the diagonals, and moving the monomers further apart would be unfavorable to

the linear ones. Hence, there is restoring torque developed in the system to counteract our own external torque.

We decided on the number of virtual particles arbitrarily. The higher the number of particles, more efficient the restoring torque generating mechanism would be, but overall it would mean higher computational cost. Of course, we also risk making the system not resemble the DNA due to the rigidity of the chain, which is one of the foremost drawbacks of this model, as will be explained in a later section. After some trials, we settled on six as the optimal number of virtual particles. The virtual particles are attached to the center of the monomer with a spring of $\kappa = 10000$ value so that they would be restricted to the surface of the monomers. Further, the virtual particles of the same monomer are also interconnected to each other so that they stay in the plane.

The system is composed of many such pairs connected to each other to form a linear chain. The first and the last monomers are always held stationary over the course of the simulation. This is necessary because the monomers could relieve any torsional stress if one of the ends is free. For simulating the system, we use Brownian Dynamics with an integration time step of 10^{-5}. The κ of all interconnecting springs was chosen to be 300.

Chapter 3

Results and Discussion

In this chapter, we will highlight the results and describe in detail the analysis performed to obtain the results. Initially, we will discuss the calibration methods used to match the length and time scale of the experiment to that of the simulation. This is very crucial to make sense of any simulation data presented throughout the rest of the chapter. Most of the immediate findings of the plots and results will be discussed in the underlying sections themselves, giving some backing to the claims made in each sections.

3.1 Length Scale Calibration

For matching the length scale in the simulation with that of the experiment, we need to calibrate the length scale. This can be done in many ways using the parameters defined in the deciding the parameters section. One way is to use the total length of the polymer in the simulation and match it with the total length of the DNA of the experiment, and map 1 real unit in nm to simulation units.

3.1.1 DNA Displacement Simulation

As we already estimated, the total number of base pairs in our simulation is 2269. Between the monomers, we use FENE potential with parameters k = 30, $r_0 = 1.5$ and $\varepsilon = 1$. The expectation value of the bond length with the FENE potential with these parameters comes out to be 0.9677. This is our average bond distance hence the total length of the polymer in simulation units (arbitrary units a.u.) can be written as,

$$L = 2269 * 0.9677 = 2195.7$$

But we already know that the total length of the DNA in real units is $16.17 \mu m$. Hence,

$$\frac{2195.7}{16167} \text{a.u.} = 1\text{nm}$$

$$0.136\text{a.u.} = 1\text{nm} \tag{3.1}$$

3.1.2 Extrusion Simulation

We perform a similar analysis for the extrusion parameters. The number of base pairs resulting from the experimental parameters is 3026. As such, due to similar parameters of the FENE potential, the total length of the polymer in simulation units can be written as,

$$L = 3026 * 0.9677 = 2928.26$$

Since the DNA used in the extrusion experiment is the same, the total length of the DNA is still $16.17\mu m$. Hence,

$$\frac{2928.26}{16167}\text{a.u.} = 1\text{nm}$$
$$0.181\text{a.u.} = 1\text{nm} \tag{3.2}$$

3.2 Time Scale Calibration

The force in simulation units is arbitrary, so it has to be converted to real units to make any reasonable analysis and compare them with the experiments. Since we already know the length scale, getting a time scale would give physical significance to the force amplitudes of our simulation and let us compare our values with the extrusion forces of the experiment.

3.2.1 DNA Displacement Time Scale

We analyzed displacements of the DNA to compare the experimental timescale to our simulations. To do this, we use an experiment performed by our collaborators where the DNA on a plate was grafted at the two ends and the data was extracted for calculating the mean square displacement. To replicate the experiment, we grafted our polymer on two ends and after the system had attained equilibrium, extracted the position coordinates of all the monomers. To calculate the mean square displacement, we took the x-coordinates of the polymer along particular ”zlines” as shown in the figure 3.1. This just means that we took a particular line in the xy plane, say z = 1, and extracted x-coordinates of the particles about this line. The MSD calculation was done only for these x-coordinates in line with the experiments, where only the x-coordinate can be extracted from the 2D microscopy images.

The mean square displacement is a standard measure of a particle’s deviation from a reference

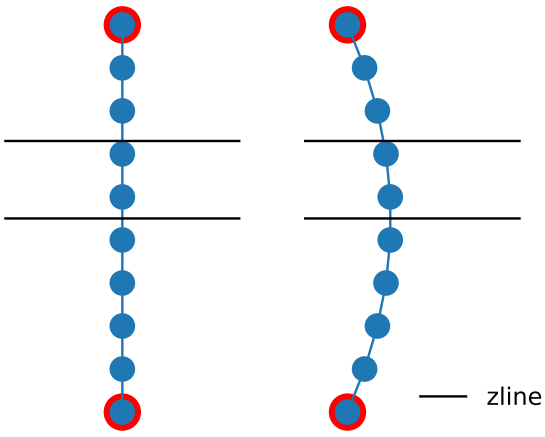


Figure 3.1: **A 2D representation of the method used to calibrate the fluctuation time scale**. As shown in the figure, the monomer positions are limited due to the grafting points illustrated by the red outlines, owing to which the MSD reaches saturation. We evaluate the MSD about these "zlines" using only the x-coordinates.

position over a given time t. If the MSD saturates, it gives an estimate of the maximum spatial extent. The mean square displacement at a time t is calculated as follows,

$$MSD = \left\langle |\vec{x}(t) - \vec{x_0}|^2 \right\rangle = \frac{1}{N} \sum_{i=1}^{N} |\vec{x}^{(i)}(t) - \vec{x}^{(i)}(0)|^2, \tag{3.3}$$

where the vector $\vec{x}$ can be a multidimensional position. We modify the general equation for our case. Since we only take x-coordinates into account, the formula for lag time τ can be written as

$$MSD(\tau) = \left\langle |x(t_2) - x(t_1)|^2 \right\rangle, \tau = t_2 - t_1. \tag{3.4}$$

In our simulation, as mentioned before, the DNA was grafted at two ends and the system was allowed to equilibrate after initializing it in a rectangular conformation. The grafting distance was fixed at 625 and we used 2269 monomers after calibrating with the experimental parameters. The drag coefficient of the simulation was set to 1, and the simulation was run for about 26e9 iterations. This run took almost 19 days, and data was extracted every 5e4 iterations to plot the MSD values as shown in figure 3.2.

In the experiment perfomed by Pradhan et al., the displacements of the DNA were analyzed particular pixels away from the grafted ends (figure 3.3). The base pairs fluctuating around the central line had the maximum extent of MSD which kept progressively decreasing on either side. We observe the same in the simulations as the values of MSD are maximum around the central portion of the polymer and values decreased on either side. This is expected because

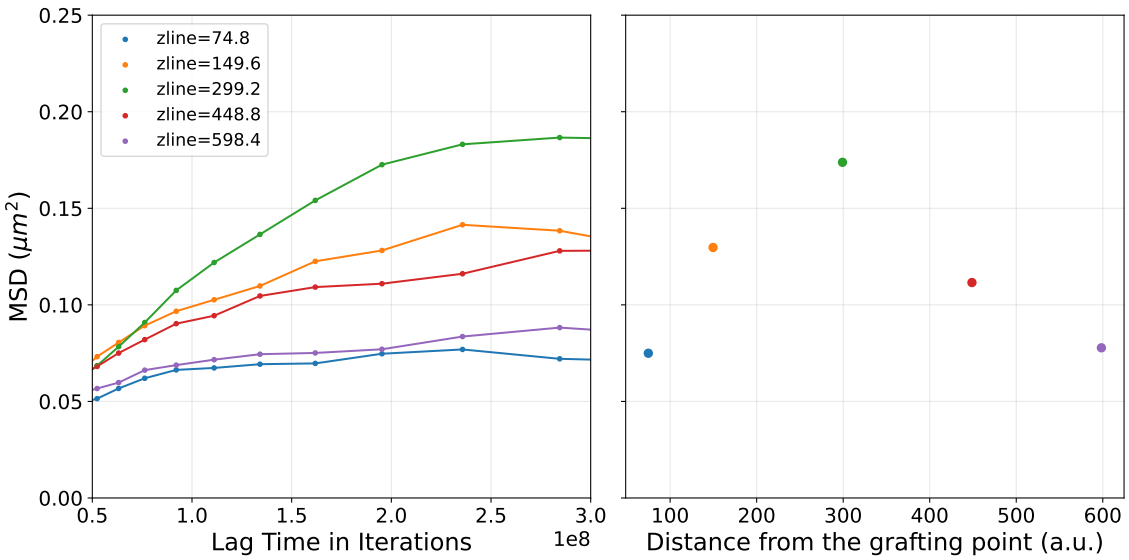


Figure 3.2: **MSD values of DNA displacements plotted over iterations for gamma 1**. The colors represent corresponding distance along the z direction; for instance the green line represents MSD at a distance of 299.2 from the grafting point at z=0 as shown by the plot on the right. The values of MSD in the plot on the right are extracted at lag time 2×10^8.

the base pairs in the center can move the maximum before being pulled by the base pairs on either side which can move less. This repeats over the entire length of the DNA, and we get the corresponding decrease in the MSD. The plateau occurs because the DNA is grafted. We can see from figure 3.3 that the values of the MSD obtained from the experiment match well with those of our simulation.

To calibrate the time scale between the experiment and the simulation, we tried to compute after how many iterations the MSD values started to saturate, and compared it to the corresponding experimental time. Comparing figures 3.2 and 3.4, we can see that approximately, 2×10^8 iterations correspond to 1s for gamma value of 1. This means there is a significant computational cost to reasonably simulate the time scale of the experiment. Again, if it was just finding the fluctuations of the DNA, this would still be achievable, but if we were to extrapolate the time scale to the process of extrusion, then it becomes a monumental task just to simulate one experiment.

The workaround to this problem is to make the diffusion more prominent by increasing the diffusion constant. This can be rationalized by the Stokes-Einstein [25] equation which provides the diffusion constant of a particle under Brownian motion. The equation is given by,

$$D = \frac{k_B T}{6\pi\eta r v}. \tag{3.5}$$

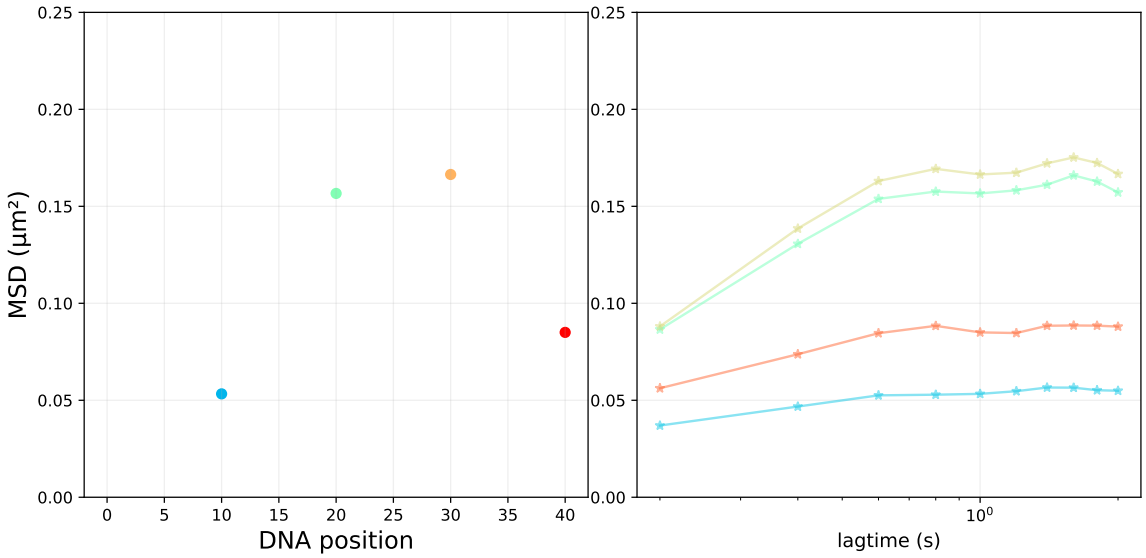


Figure 3.3: **MSD values of DNA displacements in the experiment**. The left side of the plot shows the MSD values extracted at time $t = 1$s from the plot on the right. The DNA position is measured in pixels in the experiment, which have been mapped to our simulation units in all other simulation plots. The figures have been reproduced with permission from the author.

In this case, the diffusion constant is inversely proportional to the drag coefficient. So, even for our polymer system, we should be able to speed up the diffusion of the chain by decreasing the value of the drag coefficient. This should make the MSD saturate faster, given that the chain dynamics have hastened due to the increased diffusion constant. To test this solution, we repeated the same simulation for a for $\gamma = 0.1$. The results were as expected, and this is evident from the figure plotted below 3.4. In this case, the order of time scales decreased by a factor of ten, and the MSD plot saturated much faster. Due to the higher mobility as a result of a lowered drag coefficient, the simulation time step dt at which the system updates at $\gamma = 0.1$ are stable remain to be analyzed.

This run was much faster than the previous one. We ran the simulation for about 8×10^9 iterations, which took about 5 days on gpu. The averaging took much less time than the previous run, so the computational cost was reduced drastically by changing the value of drag coefficient. The simulation system was different from the previous one. The grafting distance was still fixed at 625 and we used 2269 monomers. Again, by comparing figure 3.4 with the figure from the experiment 3.2, it takes approximately 3×10^7 iterations for the MSD values to saturate. This means that 1s in the experimental time corresponds to 3×10^7 iterations.

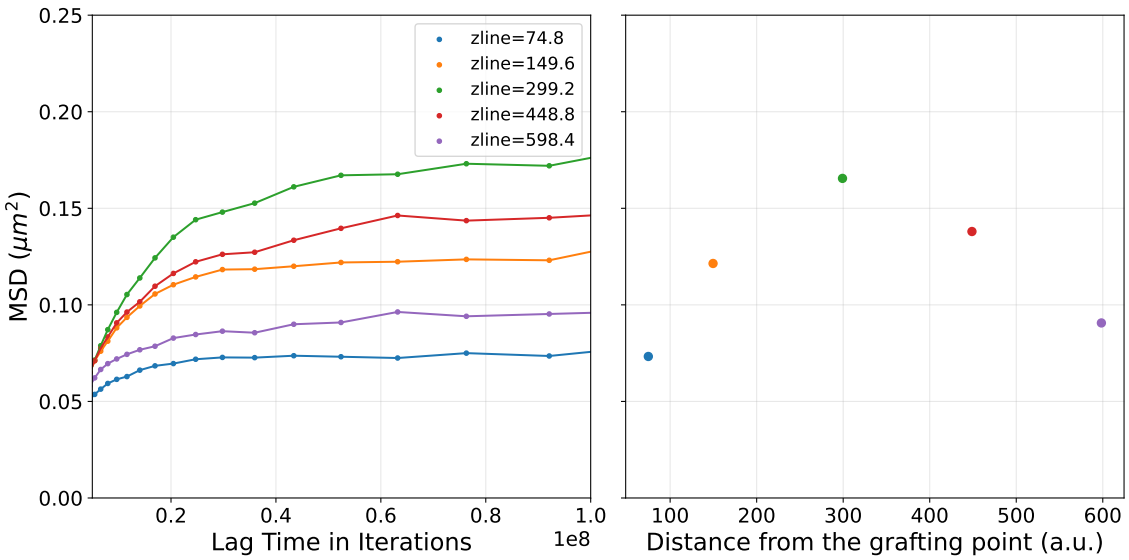


Figure 3.4: **MSD values of DNA displacements plotted over iterations for gamma 0.1**. The colors represent corresponding distance along the z direction; for instance the green line represents MSD at a distance of 299.2 from the grafting point at z=0 as shown by the plot on the right. The values of MSD in the plot on the right are extracted at lag time 5e7.

3.2.2 Extrusion Time Scale

The time scale used for calibrating the force of extrusion can be obtained from the time taken for the relative extension to saturate. The elaboration of the relative extension has been done in the results section, but to give a succinct definition, it gives an estimate of the amount of DNA that been extruded through the SMC protein, or in our case through the handcuff. This has to saturate, given that there is a limit to how much DNA can be extruded. We compare the time taken for saturation from similar values of relative extension and extract a time scale.

The analysis for extracting the time scale using extrusion was done using for $\gamma = 0.1$. As discussed in the previous section, the fluctuations are enhanced for smaller values of gamma, owing to which the dynamics of the chain is faster. This enables us to reduce the computational cost without losing any physics of the mechanism. The plot of relative extension over time of the experiment performed by Pradhan et al. has been attached in 3.9. It can be seen from the plot that the time taken for the relative extension to go from 0.5 to saturation is close to 5 seconds.

The comparison of the experimental plot in figure 3.9 with that of our simulation yields the corresponding extrusion time scale. We can see that the iterations taken for the relative extension to saturate from 0.5 in simulation is approximately 0.6×10^7 iterations. This would mean that 5s in real experimental time would correspond to these many iterations. Thus, 1 second in real time corresponds to 1.1×10^6 iterations. Note that in previous section, we showed that 1s

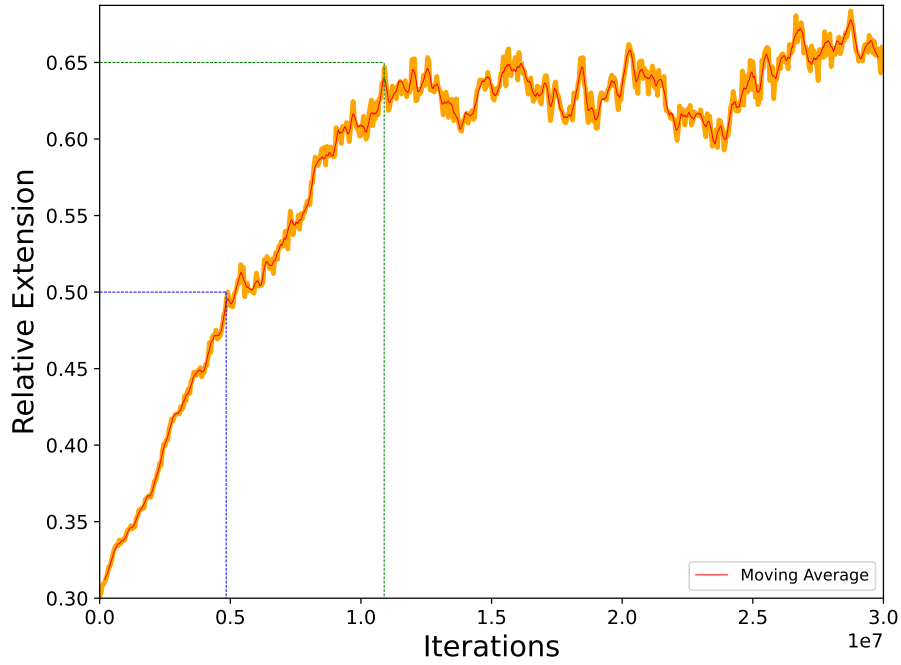


Figure 3.5: **Relative extension as a function of iterations for calibration of time scale.** The moving average has been plotted over a window of 100 convolutions. The extrusion has reached saturation and the corresponding iterations required for the relative extension to saturate from 0.5 has been calculated by drawing the dotted lines.

corresponds to 3×10^7 iterations which is one order of magnitude greater than this.

The time scales for DNA displacement and extrusion are very different. We can think of a few possible reasons why this might be the case. Firstly, the experiments were performed under two different conditions, and even though we have accounted for this by changing our parameters, we don't really have any idea of other possible experimental variables between the two experiments. Secondly, the addition of Smc5/6 and ATP might hasten the chain dynamics and this might cause the difference in the time scale. We don't have concrete analysis for any of these assumptions, but the solution is to perform DNA displacement experiment first followed by extrusion on the same setup and then do the comparison again. This might at least tell us why there is a variability. It's also possible that the difference in the time scale is because of the biological mechanism behind the process of extrusion. The extrusion happens in steps, inter-spaced by periods where there is no extrusion [26]. So, the extrusion in simulation which occurs in a continuous manner seems to be much faster.

3.3 The Active Force and Relative Extension

As mentioned in the model section, the active force is a tunable parameter in the simulation, and depending on the magnitude, it produces different dynamics in our system. To recall briefly,

the active force is the force applied on the monomers in contact with either of the rings of the handcuff for the extrusion to start or propagate. Consequently, we apply equal and opposite reaction force on the handcuff itself. Both these forces will be collectively referred to as active forces or extrusion forces interchangeably throughout the following sections. We are explicitly modeling the protein which pulls from both the rings in line with the experiments, as such our active force will be applied on both the rings.

The magnitude of the active was initially set to arbitrary in units of k_bT/σ. There are three regimes here:

1. The active force is comparable to the thermal force ($\approx 1k_bT/\sigma$)
2. The active force is much higher than the thermal force
3. The active force is much lower than the thermal force

It is evident that extrusion doesn't occur in the third case, because the fluctuations are predominant, and the active force is not enough to pull the monomer through the loop. The interesting cases are the first and the second one, where the active force is either comparable or much higher than the thermal energy scale. Intuitively, we can conclude when the active force is dominant, the monomers will be completely extruded, and the polymer between the handcuff and either grafting points will be completely taut. This can be analyzed effectively by calculating a parameter called relative extension.

In the experiments conducted by Pradhan et al., in order to characterize the amount of extrusion, a parameter called relative extension was used. It was calculated as follows,

$$R_{ext} = \frac{48502d}{(I_{up} + I_{down})L_c} \tag{3.6}$$

where 48502 is the total number of base pairs in the lambda DNA, d is the end to end distance of the double tethered DNA, and Lc is the total contour of the DNA which is equal to $16\mu m$. I_{up} and I_{down} are the sizes of the DNA above and below the extruded region, calculated by measuring the intensity during imaging [4].

We converted the parameters of the experiment to simulation to characterize the extrusion in our model. The 48502 base pairs of the λ DNA were nothing but the total number of monomers, N, in our simulation. This was dependent on the salt concentration they used, and hence differed according to their experimental conditions. The parameter d was equivalent to the grafting distance d_{graft} in our case, and the total length of the DNA, Lc was equal to $0.9677N$, where is the total monomers. This is because the equilibrium bond length of our FENE springs was equal 0.9677, hence the factor. The rest of the parameters were obvious, as I_{up} was just the

total monomer above the handcuff and I_{down} was the total monomers below the handcuff. So in simulations, the extension can be written as,

$$R_{ext} = \frac{Nd_{graft}}{(N_{up} + N_{down})N * 0.9677}. \tag{3.7}$$

The simulation of the extrusion was performed in the three different regimes mentioned above. The last one was as expected, and the conclusion was that the active force should be above the threshold value for extrusion to occur. This was similar to a first order transition, where extrusion did not occur below certain threshold, and started once we crossed the threshold. This was hard to characterize exactly because of the inherent randomness in the system and the floating point precision, but the qualitative behavior was as expected.

Quantitatively, the relative extension gives an estimate of the amount of extrusion in the system. If the value is close to 1, this means the region of the grafted polymer between the handcuff and the graft points is taut. Conversely, values around 0.4 (in our parameter regime) imply that extrusion has hardly happened. The idea relative saturation 'saturating' before the complete extrusion wasn't something we had pondered upon. The hypothesis was that the monomers between the handcuff and the wall would be completely pulled through the rings, and the extrusion would stop once the polymer had become taut. Of course, this was the case, but only in the regime when the activity was much higher than the thermal scales.

In the plot 3.6, the activity value was set to a value much higher than the thermal scale. This would mean that the fluctuations of the polymer would be inconsequential to the extrusion event, and hence any effects due to the randomness in the system wouldn't matter in the schema of extrusion. As expected, complete extrusion occurred in this case and the relative extension saturated at values close to 0.92. We observed the same behavior even when the activity was increased to higher values (for values above 100 or somewhere close, the simulation crashed), and the regions of the polymer between the handcuff and the grafting points remained taut. The plots have been omitted to avoid redundancy, but they looked no different from 3.6.

We performed multiple simulations each with decreasing value of the active force until we reached the minimum possible threshold where extrusion was no longer possible. In the regime where the order of the active forces was comparable to the thermal forces, that is the first regime, we started to see the interesting dynamics. The saturation started to occur before the complete extrusion, and the value of saturation periodically decreased as we decreased the value of the extrusion force. This meant there were other forces counter balancing our own extrusion forces. In this regime the fluctuations were as important as the extrusion forces themselves.

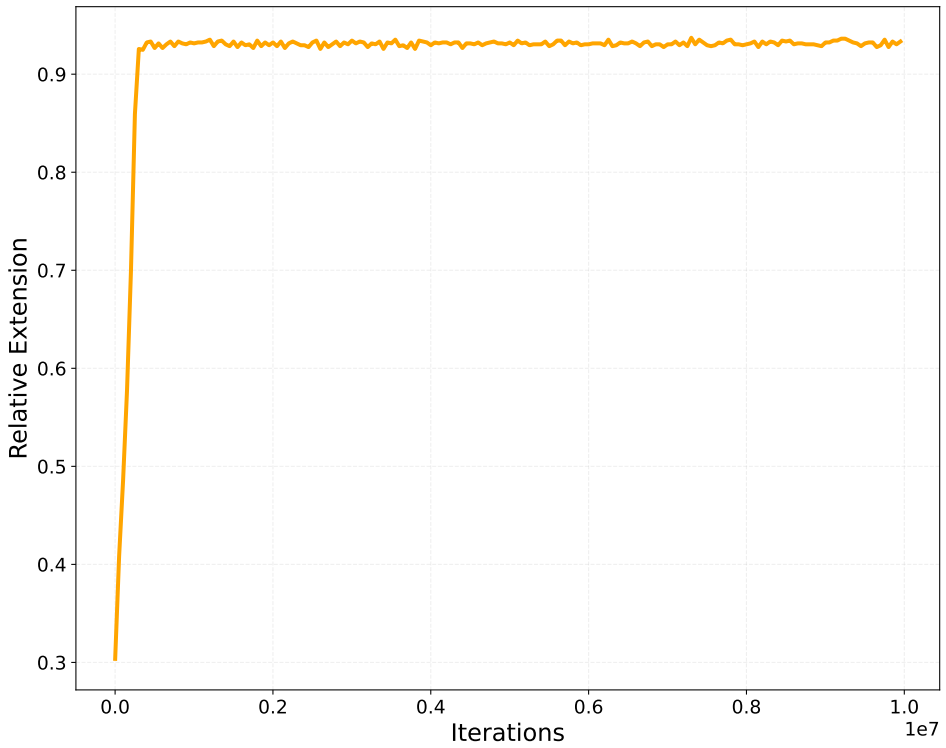


Figure 3.6: **Relative extension of the polymer when active force >> thermal force**. The activity value in the simulation was set to 10 k_bT/σ, which was an order of magnitude greater than the thermal force.

There are a couple of contributing players in the stalling force. Firstly, the polymer does not prefer loops because it decreases the configurational entropy of the molecule. In simple words, the presence of loops decreases the total ways in which the polymer can rearrange itself. As such, there is always an opposing force against the extrusion forces, which increases as the size of the loop increases. Secondly, the fluctuations of the non-extruded part of the polymer might also be the contributors to this stalling. Both these cases might be intimately linked, and from the later analyses it seems that the fluctuations play a dominant role in deciding the optimal loop size. These factors contribute to stalling only when we are in the first regime, where the forces are comparable to the thermal energy scales.

The plot 3.7 and figure 3.8 shows the importance of fluctuations in the extrusion event. The activity is maintained at 0.2, which is comparable to the thermal energy scales and we can see that the relative extension saturates at a much lower value than the case where the activity values are high. In a later section, we will show that the fluctuations of the rest of the chain will tend to minimize any loops present in the polymer.

Comparing our result to the experiment performed by Pradhan et al. 3.9, the DNA wasn't extruded completely in the experiment either and the relative extension saturated around the

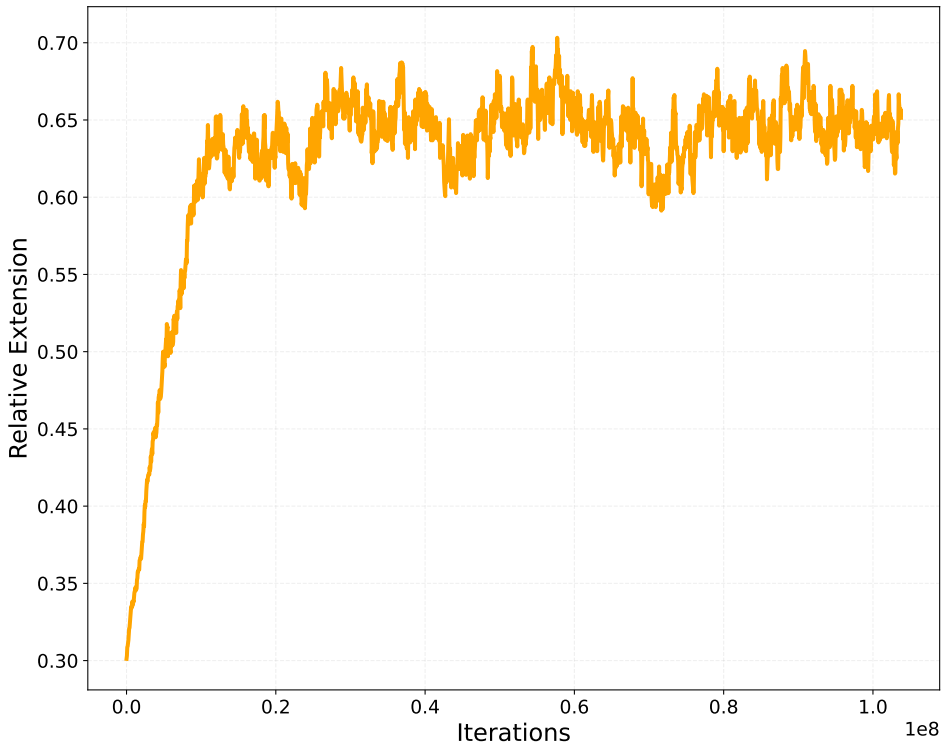
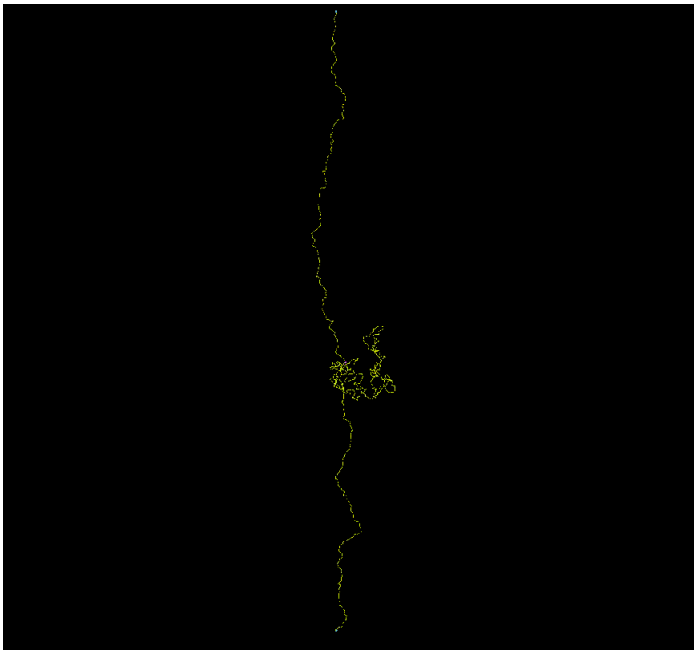


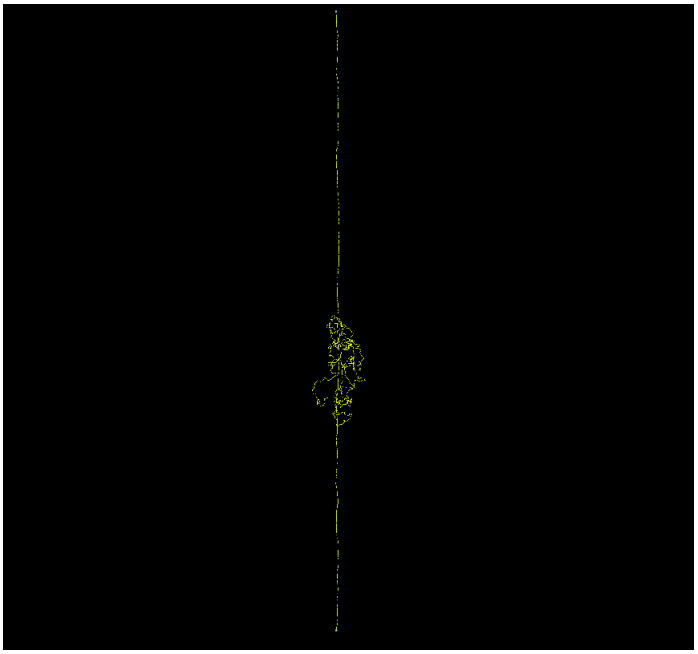
Figure 3.7: **Relative extension of the polymer when active force ≈ thermal force**. The activity value in the simulation was set to 0.2 k_bT/σ, which was comparable to the thermal force. The relative extension has saturated before the polymer between the grafting points and handcuff has become taut.

similar values as ours when we parameterized our force values to 0.2 k_BT/σ. This is an obvious implication of the fact that the extrusion forces of the SMC protein used in the experiment were small enough for the fluctuations of the DNA to have an effect in the process of extrusion. Hence, the extrusion process stopped because of these stalling forces even in the experiment.

The collection of the points in figure 3.10 shows how the relative extension saturates to different values of activity. The values of activity are small enough so that they are comparable to the thermal energy scales, and we can see that the size of loop progressively increases with increasing values of extrusion forces. The plot by itself may not be so surprising, but when put up against the fact that there is no extrusion despite the presence of more monomers to be pulled through is quite riveting. Initially there is a small linear regime followed by a plateau, which signified the transfer into the second regime where the extrusion stops when the entire non-extruded polymer becomes taut. This is where tension comes into play to stall the extrusion, and the relative extension saturates almost at the same values for all higher extrusion forces.



(a) Acitivity $0.2k_BT/\sigma$



(b) Activity $20k_BT/\sigma$

Figure 3.8: **Simulation snapshot of extrusion captured using ovito [17].** The yellow splotches are the monomers, and extruded monomers are concentrated at a region in both the snapshots. The distinction between the two regimes is very evident from the snapshots, as in the higher activity case, the non-extruded polymer is completely taut.

3.4 Tension Analysis

In order to better characterize the points made in the last section and gauge the effects of stalling forces, we analyzed the tension in the polymer chain. This would help us tell apart the role of tension and forces due to fluctuations on the process of extrusion, and thus concretely establish

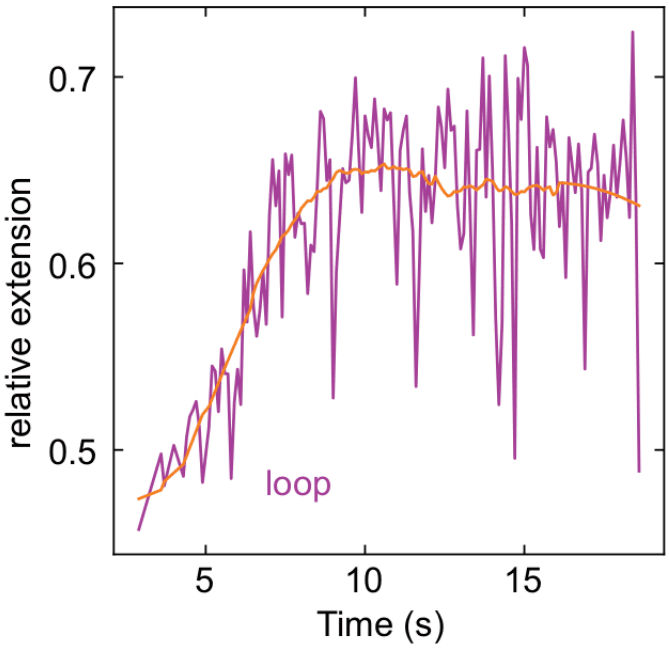


Figure 3.9: **SMC5/6-meidated loop extrusion kinetics extracted from a single looping event**. The above plot is from the experiment, and it shows how the relative extension of the DNA changes as a function of time. The figure has been reproduced with permission from the author [4].

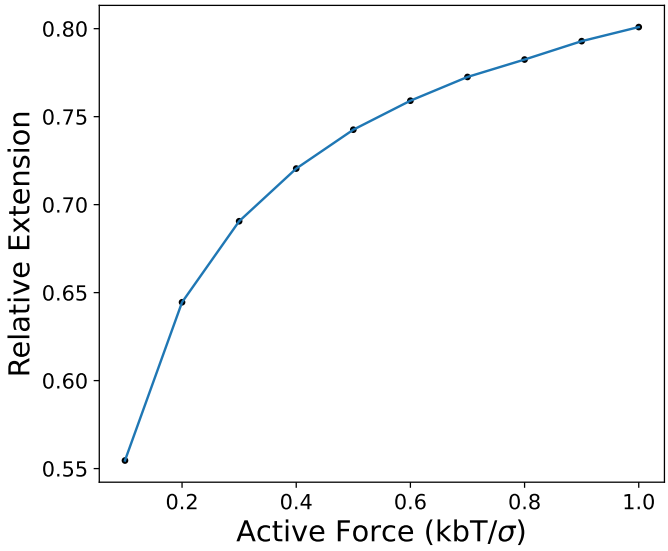


Figure 3.10: **Relative extension as a function of extrusion forces**. The averaging for relative extension has been done after the equilibrium has been attained. The transition away from the thermal energy scales is evident in the plot.

the fact that fluctuations of the chain are indeed dominant factors in saturation of the relative extension in the regime where the energy scales are comparable to the thermal ones. Initially, we perform some analysis on simple systems where the polymer is grafted on either one end and both ends without any extrusion process. The qualitative results are obvious for these con-

figurations, so this would concretely establish the correctness of our method.

Let us first characterize the two possible scenarios for extrusion to stop completely. First, the polymer between the grafting points and the handcuff becomes completely taut and there would be no way to get more monomers past the rings of the handcuff. Second, the stalling forces would be just enough to stop the active forces from continuing the process of extrusion. In the first case, the tension along the chain should be different between the extruded part and the non extruded part, because the chain in taut, and the springs are stretched. In the second case, the tension should be similar across the chain as no stretching has occured yet.

The method we adapt to measure the tension along the chain is to find the average bond length across each of the monomer and use it in the equation of force due to the FENEWCA potential to get an estimate of the extension force along the polymer chain. This, we effectively treat it as tension, and any increase from the default value would indicate that there is extension in the chain. The equation for the FENEWCA potential is given by,

$$V(r) = -\frac{1}{2}kr_0^2 \ln\left[1-\left(\frac{r}{r_0}\right)^2\right] + 4\varepsilon\left[12\left(\frac{\sigma^{12}}{r^{12}}\right) - 6\left(\frac{\sigma^6}{r^6}\right)\right]. \tag{3.8}$$

The negative derivative of the potential is the force, hence the tension in the polymer can be characterized by using the equation,

$$F(\langle r\rangle) = -k\left[\frac{r_0^2}{r_0^2 - \langle r\rangle^2}\right]\langle r\rangle + 4\varepsilon\left[12\left(\frac{\sigma^{12}}{r^{13}}\right) - 6\left(\frac{\sigma^6}{r^7}\right)\right]. \tag{3.9}$$

To analyze whether the equation gave us the right results, we grafted a polymer chain at one end and gave flow in the direction normal to the wall. This would cause the polymer to stretch in the direction of the flow, and consequently, the magnitude of the tension would be maximum at the portion of the polymer close to the wall and would progressively decrease along the chain. This is because the last monomer would be just pulled the flow force, the adjacent one will be pulled by the flow force as well the last monomer, and the numbers keep on adding until we reach the monomer right next to the wall. This will be pulled by all the monomers away from the wall, hence would have maximum magnitude of tension. The plot 3.11 is in line with our assumptions, hence, qualitatively it produces the expected results.

The simulation for tension analysis was also performed by grafting both ends of the chain. The wall was positioned in the xz plane, and the flow was given normal to the wall, along the y direction. The magnitude was high enough for the polymer to stretch yet stay intact. Consequently, the chain started stretching along the direction of the flow. The monomers adja-

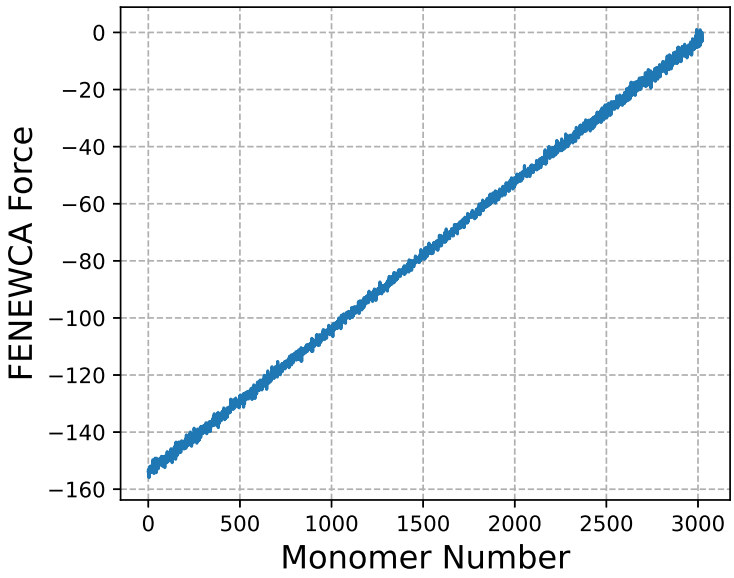
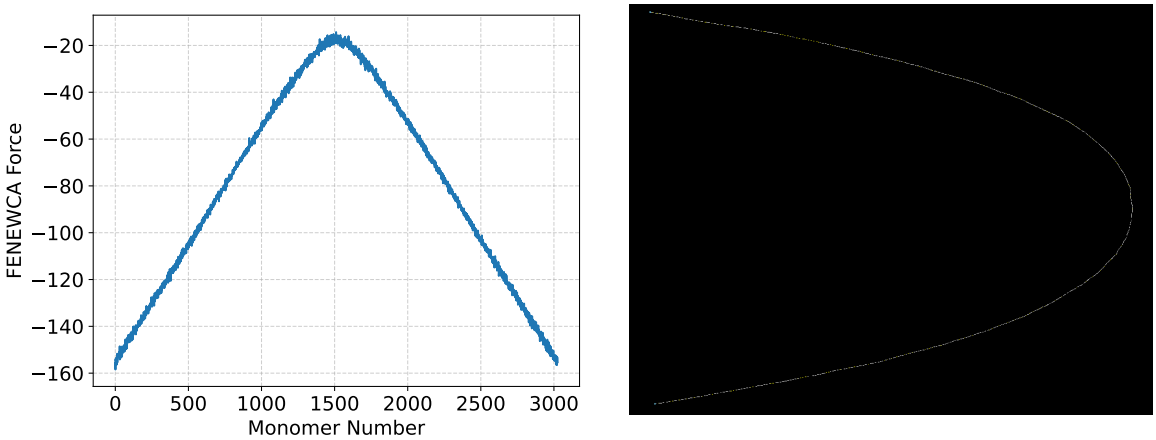


Figure 3.11: **The tension along the polymer for a chain grafted at one end in the presence of flow**. The magnitude of the tension is maximum at the grafted end, as it is pulled by all the monomers on the away from the grafting point, and the magnitude of the force progressively decreases along the chain.

cent to the walls were once again pulled by all the monomers away from the walls, hence had maximum magnitude of tension, which kept on progressively decreasing as we moved to the monomers toward the middle of the polymer. As shown in 3.12, the simulation results confer to our intuition regarding the system.



(a) Tension along the contour

(b) Simulation snapshot

Figure 3.12: **The tension along the polymer for a chain grafted at two ends in the presence of flow**. The grafting distance is set to 845. The magnitude of tension is maximum at the grafted ends, and it progressively decreases as we move to the center of the chain. The FENE force in the plot is averaged over last 500 iterations.

From the above results, we have verified qualitatively that our method to analyze the tension

works. Now, we perform similar analysis for polymer without any handcuff and extrusion, in the absence of flow. The polymer was once again grafted at two ends and allowed to fluctuate freely without stretching. We can expect the FENEWCA force to fluctuate around a mean value because there is nothing overextending the springs anywhere, except the ones right next to the walls. They are of different nature, so we have omitted them from the plot. The mean value of the force would be expected value for a grafted chain with or without the handcuff, given that there is no stretching in the chain. Setting this as our reference value, we would expect the chain without any stretching to fluctuate about this value.

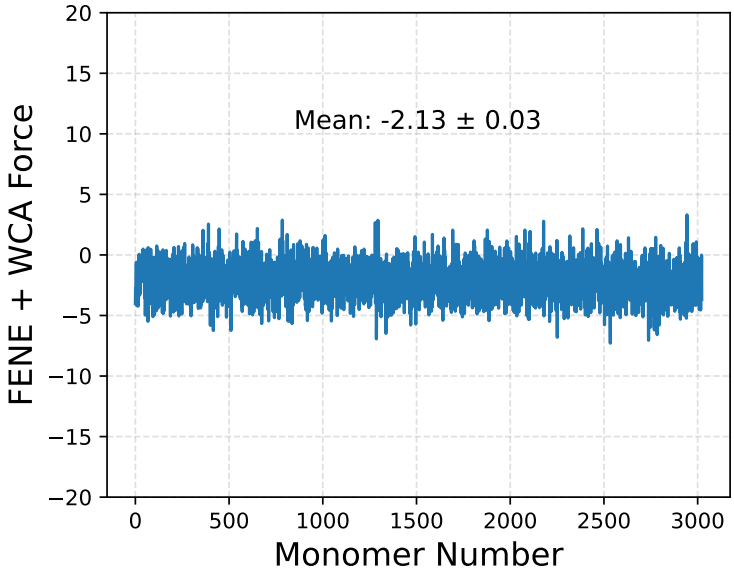


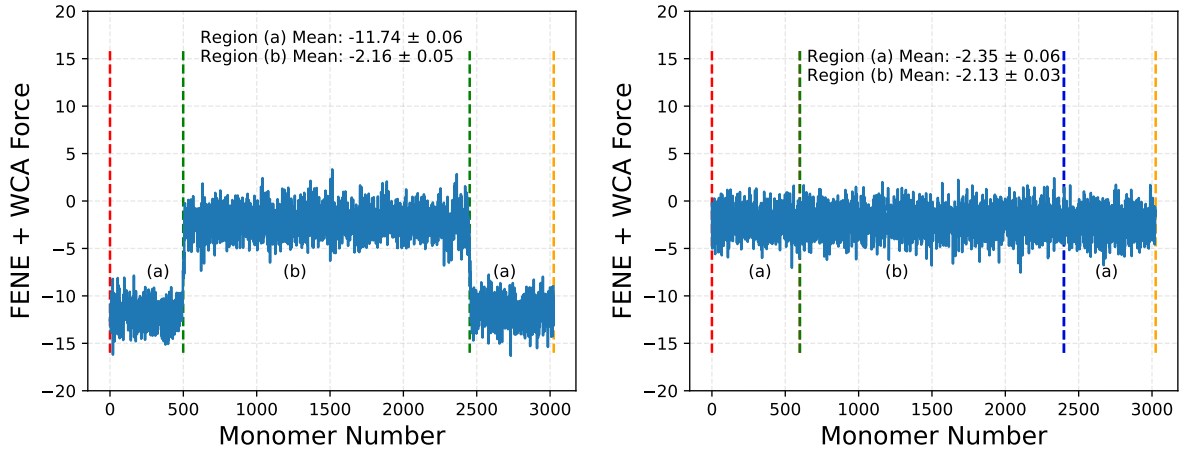
Figure 3.13: **The tension along the polymer for a chain grafted at two ends without flow**. The grafting distance is set to 845. The FENE + WCA potential between the monomers adjacent to the wall have been ignored in the plot and mean value calculation because of their different nature. The FENE + WCA force in the plot is averaged over the last the 500 iterations.

We now analyze the tension in the three different regimes. Obviously, when the active force value is less than the threshold, no extrusion occurs and the plot would be no different from the one for shown in figure 3.13 because this case is no different from having no extrusion forces. The simulation gave us the same result, so we have omitted the plot to avoid redundancy. In the regime where active forces are much greater than the thermal forces, the polymer chain becomes taut, so we expect the tension to be maximum in the taut part of the chain, and to be minimum in the extruded part. The minimum tension should fluctuate around our reference value, because there is no stretching in the extruded part of the chain.

It is evident from the plot 3.14a that the polymer is taut as expected when the extrusion forces are high. The extruded region is still fluctuating around the reference value, which means there

is no stretching observed in the extruded part, whereas the taut chain shows some stretching. The claim here is that the extrusion stops purely because of the tension in this case, and the activity or extrusion force is not enough to cause more stretching and pull more monomers through the rings of the handcuff. Entropic effects aren't evident here because we are beyond the thermal energy scales. This is the qualitative behavior expected when the relative extension reaches to values close to 1, that is when most of the free polymer has been extruded through the rings of the handcuff.

The interesting regime is when the magnitude of the extrusion forces are comparable to that of



(a) Extrusion force 10 k_BT/σ

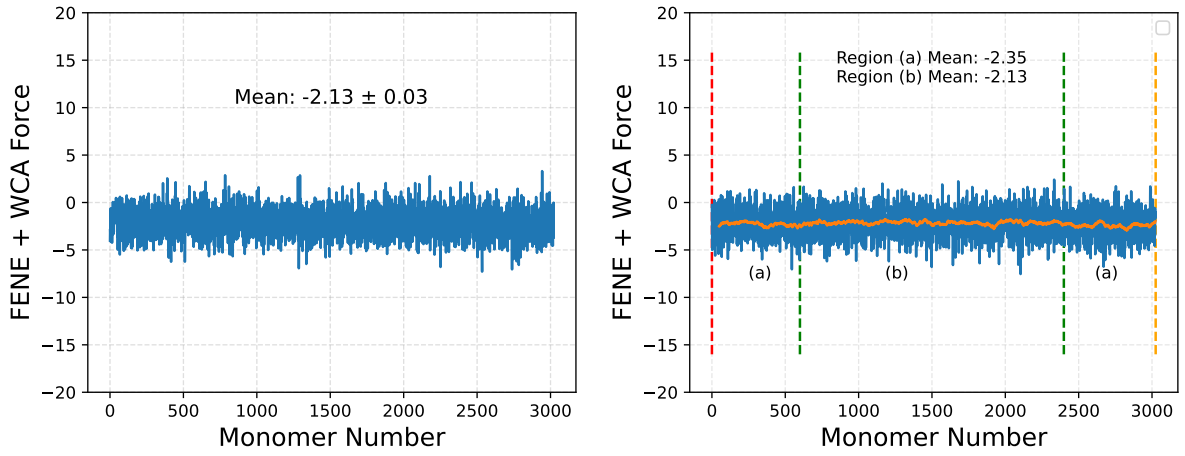
(b) Extrusion force 0.2 k_BT/σ

Figure 3.14: **Comparison of the tension along the polymer for a grafted chain after the extrusion has occurred for extrusion force in two regimes**. The grafting distance is set to 845. The FENE + WCA potential between the monomers adjacent to the wall have been ignored in the plot. The FENE + WCA force in the plot is averaged over the last 500 iterations. From the plots, it is quite clear the difference in the means between the extruded and non-extruded regions is approximately equal to the corresponding activity of the handcuff.

the thermal energy scales. The exact tension analysis was performed for this regime too by giving an activity of 0.2 to the handcuff, and letting the system reach a steady state. This state was characterized by looking at the fluctuations of the relative extension around a mean value. In the figure 3.14b, the difference is not very evident because the fluctuations are dominant here. There is some force which is opposing the extrusion forces and causing the system to reach a steady state. This is indeed the force due to the fluctuation of the chain, and it is opposing further extrusion from happening.

3.5 The Extrusion Speed

The extrusion speed is the rate of extrusion. In the experiments, it is the rate at which the DNA base pairs are extruded through the SMC protein complex. Extrapolating to our simulation, it



(a) No Extrusion

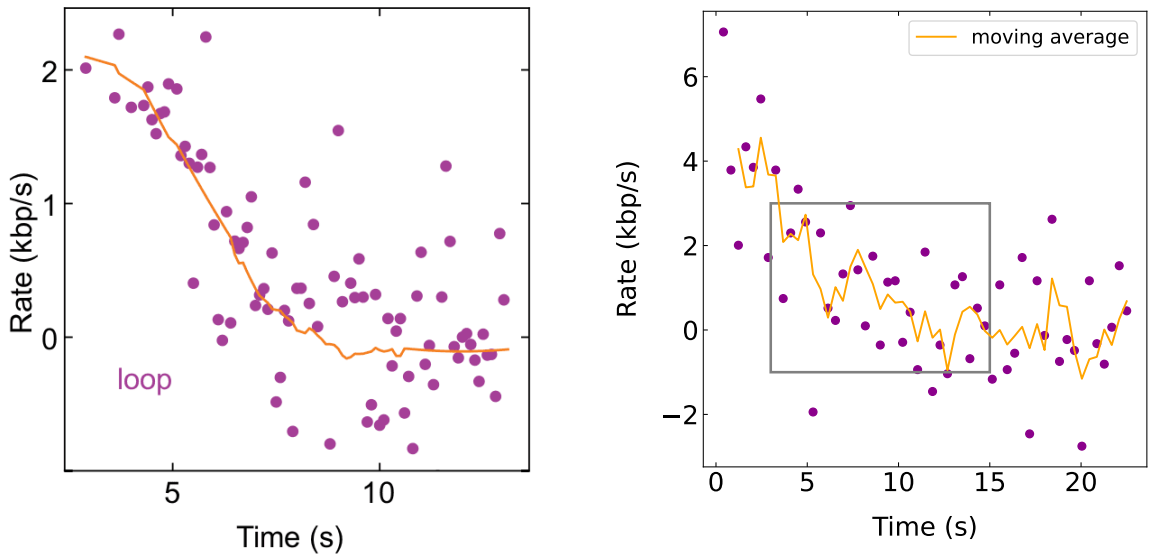
(b) Extrusion force 0.2 k_BT/σ

Figure 3.15: **Comparison of the tension along the polymer chain in the presence and absence of extrusion plotted with moving average**. The FENE + WCA potential between the monomers adjacent to the wall have been ignored in the plot. The FENE + WCA force in the plot is averaged over last the 500 iterations. The regions indicate the portion of the polymer which has been extruded (region b) and the non-extruded portion. As we can see, the difference in the mean values between region (a) and (b) uyis close to 0.2 in the plot on the right, which is exactly the activity given to the handcuff.

is simply the number of monomers passing through the rings of the handcuff every iteration during the process extrusion. To compare the experimental extrusion speed to the simulation, we need to use the time scale we got from extrusion, as well as the course-graining analysis elaborated in the previous section, since the units of experimental extrusion speed are kilo base pairs per second.

The rate of extrusion in the simulation depends on the extrusion force. The higher the extrusion force, the faster is the extrusion, hence the larger is the rate. This makes sense, given that the relative extension saturates at increasing values as we ramp up the activity while still being in the thermal regime. In the experiments, the extrusion was done with only one kind of SMC protein, so we were not able to verify how the relative extrusion scales with the activity. Both qualitatively and quantitatively, the rates between experiments and simulation we found matched quite well 3.16, and the rate decreased over increasing iterations.

The decrease in the extrusion speed might be attributed to the increase in the forces needed to stretch the linear segment as the size of the loop increases, and the fluctuations of the rest of the chain become predominant. But it also possible that the handcuff simply has less and less of free monomers to pull through the rings as more of the monomers are extruded. It is hard to ascertain which of these two factors is the dominant contributor, but in the thermal regime, it seems that the force-extension affects do play a role in decreasing the extrusion speed, given



(a) Experiment (b) Simulation

Figure 3.16: **The extrusion speed as a function of time for both experiments and simulation**. The extrusion speed decreases over time, and from figure 3.17 we can ascertain that the difference in rate at the start is because of the difference in the relative extension at the beginning. For similar scales, the rates between experiments and simulation are very similar as shown by the box on the right. The figure on the left has been reproduced with permission from the author [4].

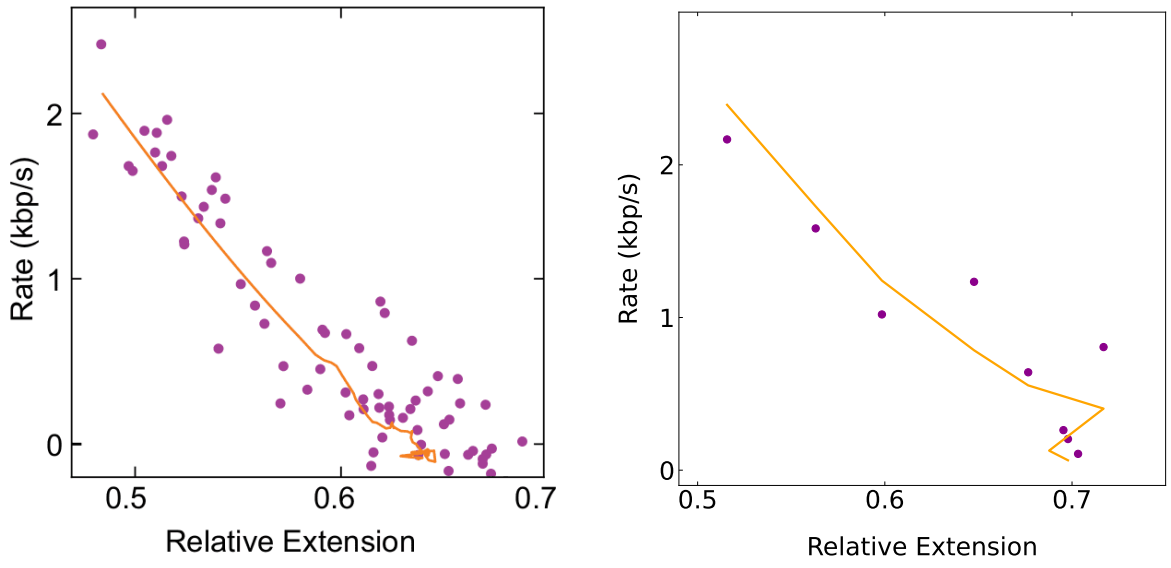
that the extrusion stops despite having free monomers to be pulled through the handcuff.

As we can see from figure 3.17, we observe the same qualitative behavior in the simulation as that of the experiment. The units in the simulation were converted to the real units in the plot just to get the quantitative estimates and see if the values between the simulation and experiments match. This conversion was trivial, but to calculate the extrusion speed in the simulation, we accounted for the monomer in contact with the handcuff inter-spaced by $\approx 10^6$ iterations just to get the necessary statistics. The reason is self explanatory as the real unit time corresponds to 1.1×10^6 iterations. The statistics attained due to this averaging was instead not very good, but we were able to draw out some inferences from the plot.

The equation used for the converting the simulation units to real units is as follows,

$$\text{rate (in ru)} = \frac{\text{rate (in su)} \times \text{(total base pairs)}}{\text{(total monomers)} \times 1.1 \times 10^6} \tag{3.10}$$

where ru is real units, su is simulation units, total base pairs are the number of base pairs in λ DNA which is 48502, total monomers is number of monomers encompassing the coarse-grained model which is equal to 3026. The factor in the denominator is to convert the iteration in simulation units to seconds.



(a) Experiment

(b) Simulation

Figure 3.17: **The extrusion speed as a function of relative extension for both experiments and simulation**. The orange line is the moving average in both the simulation and the experiment. The reduced sampling points in the simulation are because of greater averaging for each bin in the absence of which the data was too noisy for any inference.The relative extension saturates around 0.65, and it is evident from the fact that the rate reaches close to zero in both the simulation and experiments. The figure on the left has been reproduced with permission from the author [4].

3.6 Activity Frequency

As we have seen in the previous sections, the SMC protein complex responsible for extrusion was modelled using a handcuff with two rings. The activity was simply the extrusion force applied to the monomer in contact with the ring normal to the ring. In all the previous simulation results, the extrusion force was applied every iteration, that is the monomers were pulled through the rings of the handcuff every time step in the simulation. But it is possible to apply active force every other iteration, and from a simulation point of view, it is quite simple.

Interestingly, when the frequency of iteration was comparable to the DNA displacement time scale, as long the rate of extrusion was kept the constant, all the simulations saturated to a similar value of relative extension. In other words, 0.2 activity per iteration, 0.4 activity every 2 iterations, 0.6 activity every 3 iterations, 0.8 activity every 4 iterations, all saturated to almost 0.65 relative extension. For us, this meant we could speed up the simulation quite a bit. The plots in figure 3.18 attest to the fact that it is only extrusion force per iteration that makes a difference in the relative extension. We could, in principle, exploit this result to get a speed up in the simulation.

The rate limiting step in our simulation is the application of active force. This is an expen-

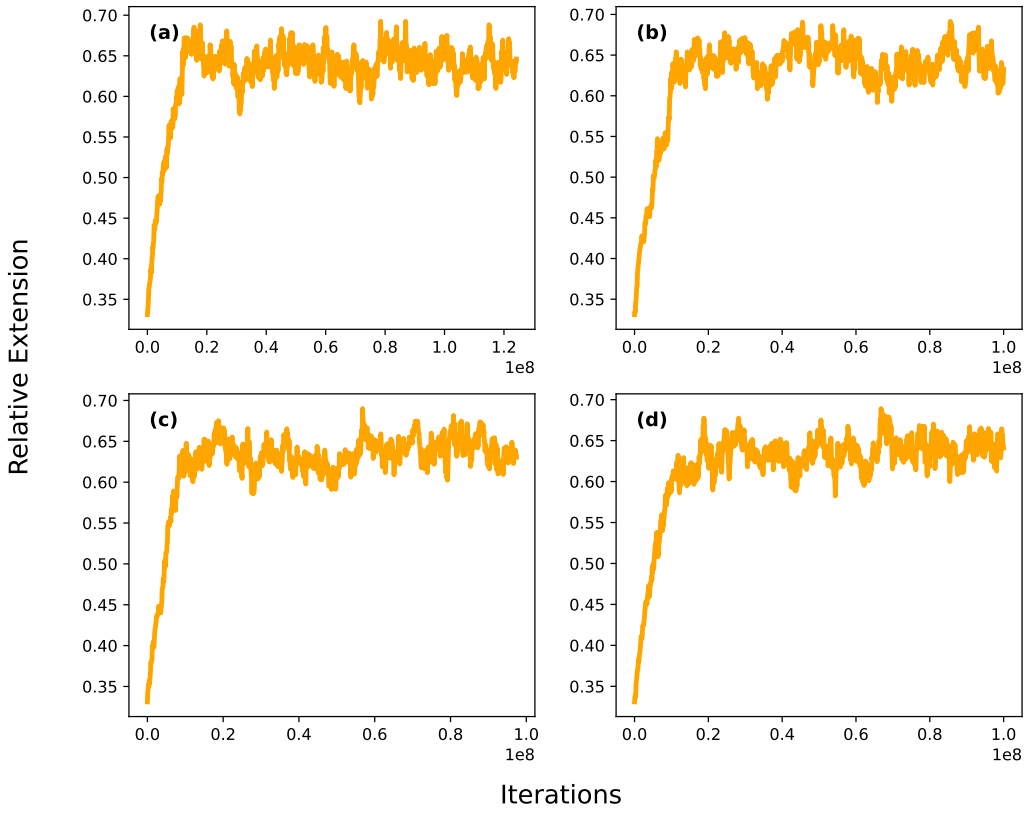


Figure 3.18: **The saturation of relative extension for different frequencies of extrusion forces, all cases giving a rate of 0.2 activity per iteration**. (a) Extrusion force is 0.4 with frequency 2, (b) Extrusion force is 0.6 with frequency 3, (c) Extrusion force is 0.8 with frequency 4, (d) Extrusion force is 1.6 with frequency 8. The relative extension in all the figures saturates at similar values, emphasizing that it is only the rate that matters when the frequency of the active forces is within the fluctuation time scale of the polymer.

sive step for multiple reasons. Firstly, the monomers in contact with the handcuff have to be found, and for this we compute distance with every other monomer and take the minimum. Second, the normal has to be evaluated, and despite not doing this every iteration, it's still quite expensive because custom functions in HOOMD-blue [13] run on CPU, unless the code itself is explicitly written in CUDA. As such, we can get almost twice speed up decreasing the frequency of extrusion force. So far, there seems to be no problem trying to increase the speed of the simulation this way, but, unless stated, all plots in the simulation have been produced by giving activity every iteration.

The result seems to be viable only to speed up simulation for now, but supposedly, the real systems might show similar behavior as that of our simulations. Given that only rate of extrusion matters, it seems inconsequential to see whether the SMC protein extrudes continuously over time, or in discontinuous fashion. In either cases, the same relative extension can be obtained. It is indeed possible that the saturation we see in the experiment is because of higher activity value over inter spaced intervals, but there is no problem is reproducing this behav-

ior by giving low activity values over every iteration. Both these situations give rise to same physics of the polymer, and so far, we haven't seen any difference between the two scenarios.

3.7 Distribution of Loop Sizes and Free Energy

In this section, we analyze the distribution of the loop sizes both in the presence and in the absence of extrusion forces and try to get a free energy landscape of the system. The calculation of the free energy from the probability distribution is done using the standard way, where the natural logarithm multiplied by $-k_BT$ would give us the required landscape. This will be analyzed against the size of the loops to give an estimate of what sizes of the loops would be preferred for different activity values.

In the absence of activity, the size of the loop formed by the two rings of the handcuff would be

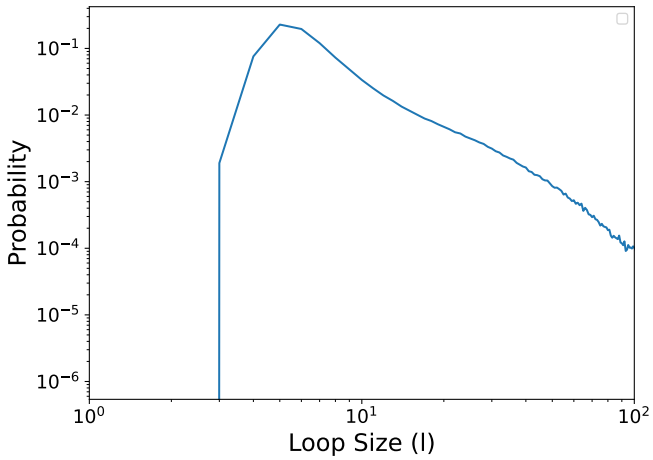


Figure 3.19: **Probability distribution of the loop size in the absence of activity.** The maximum is peaked around 5 because of the finite size of the handcuff and the steric repulsion between the loop and the rings of the handcuff. The plot is in log-log scale.

minimized. This is because the forces due to the fluctuations of the rest of the chain wouldn't allow the loop to increase in size. As such, the probability distribution of the loop size peaks around some small finite value as shown in the figure 3.19. The steric repulsion between the handcuff and the loop would not allow the size to decrease any further. The general characteristics of the curve is conserved across different system with loops, given that we are using the same handcuff for loop formation.

The equations used to evaluate the free energy are mentioned in the appendix. To recapitulate,

$$A = -k_BT \ln P(l) \tag{3.11}$$

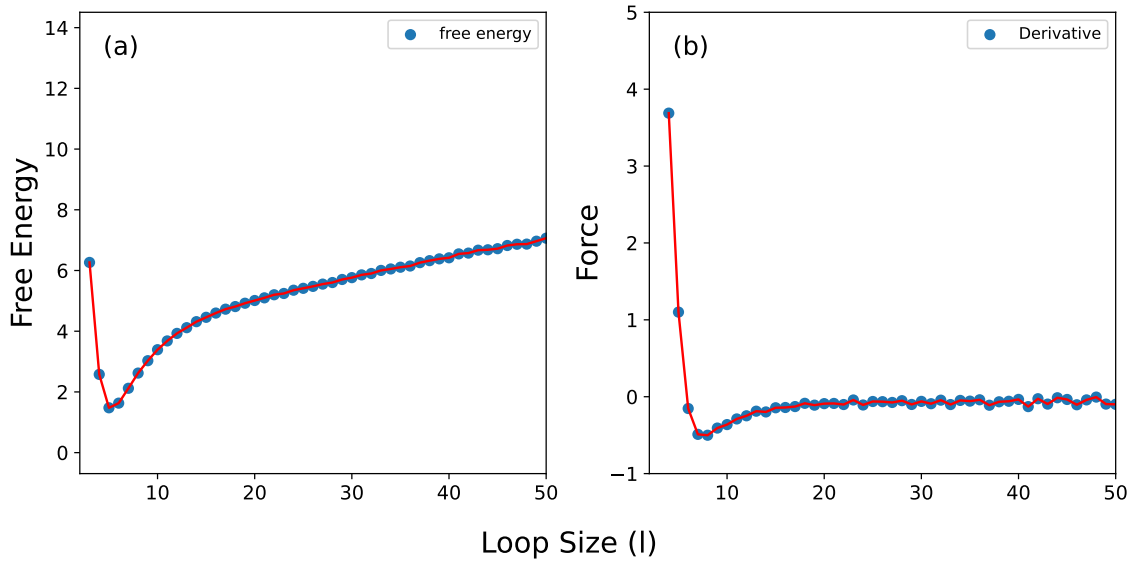


Figure 3.20: **Free energy and the derivative as a function of loop size in the absence of activity.** The free energy in (a) has been plotted by taking logarithm of the probability distribution and multiplying by $-k_BT$. The force in (b) is just the derivative of the free energy.

where l is the loop size and A is the free energy. To compute force,

$$F = -\frac{\partial A}{\partial l} \tag{3.12}$$

where F is the force.

We can take the natural logarithm of the probability distribution to get the free energy landscape. The figure 3.20 suggests that the minima of the free energy occurs around a loop size of 5. The free energy values are higher for anything lower because of the steric repulsion, and this is evident from the plot. Further, the fluctuations wouldn't let the loop size increase, which is also clear from the increasing free energy values beyond the minima. The information contained is nothing new from the previous plot, but the reasons for such a probability distribution become more concrete from the free energy landscape.

We also plot the negative of the first derivative of the free energy. To do this, we chose two adjacent points on the free energy plot, and take the slope between them. The value obtained is set as the derivative of the second point in the derivative plot. As such, the derivative plot starts from +1 on the x axis compared to the free energy plot, and this is not easily visible because of the x axis scale. It is very clear from the figure 3.20 that initially, for very small loop sizes, there is some positive force representative of the steric repulsion forces in the system. The force goes through zero and then becomes negative, which represents the forces due to the fluctuations of the rest of the chain that prevent the loop from increasing in size. The subtle balance between two regimes decides the size of the loop here, and even in the presence of

activity where extrusion forces replace the steric repulsion.

Interestingly, we can also see that magnitude of the force in figure 3.20 is of the order of the extrusion forces we need to give in order to start the process of extrusion. The force curve tells us why the extrusion starts around the values of 10^{-2}. That's the force needed to overcome the free energy barrier here for the onset of the extrusion. Again, owing to the discreteness of the plot, it is very difficult to get the exact estimates, but the order of magnitudes are almost equal.

The free energy analysis can be extended to the system with activity. The extrusion forces are kept low enough so that they are comparable to the thermal energy scales before starting the extrusion. We wait until the system has reached equilibrium to perform analysis. This is done by assuring that the size of the loop fluctuates around a mean, which is the ideal loop size for the given value of the extrusion force. We sample the data following the saturation and find the probability distribution of the loop size and plot it on a log-log scale 3.21. It is clear that the probability distribution peaks around a preferred value of loop size which is unique for the given value of activity.

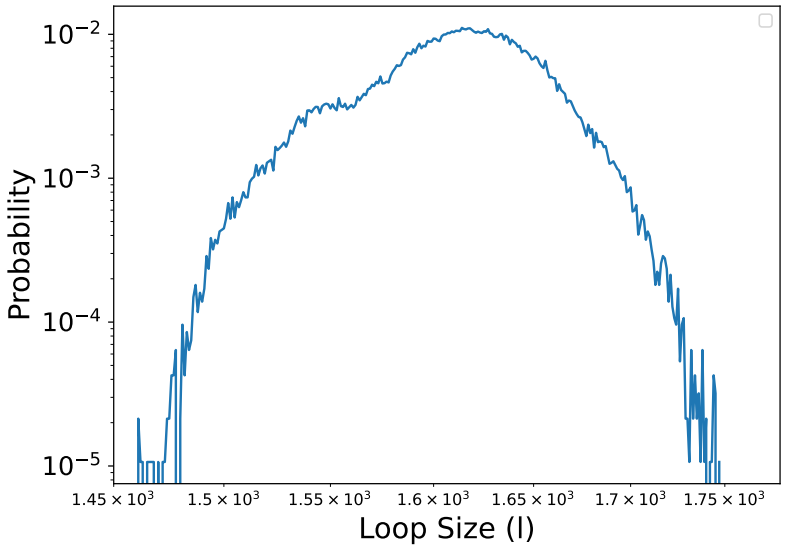


Figure 3.21: **Probability distribution of the loop size in the presence of activity.**. The sampling of data for plotting the probability distribution is done after the equilibrium configuration has been attained, which is characterized by fluctuations of the loop size about a mean value.

We calculate the free energy landscape of the extruded polymer by using the same approach as before, but we will restrict the scale to the regime of our interest. As we can see from the plot 3.22, there is a free energy minima around 1614, which translates to saying that the loop size fluctuates around this value once the equilibrium has been achieved. Also, from the derivative of the free energy, we can see that there are forces counterbalancing the extrusion forces that were applied manually. Whenever the size of the loop becomes less than the mean, the positive

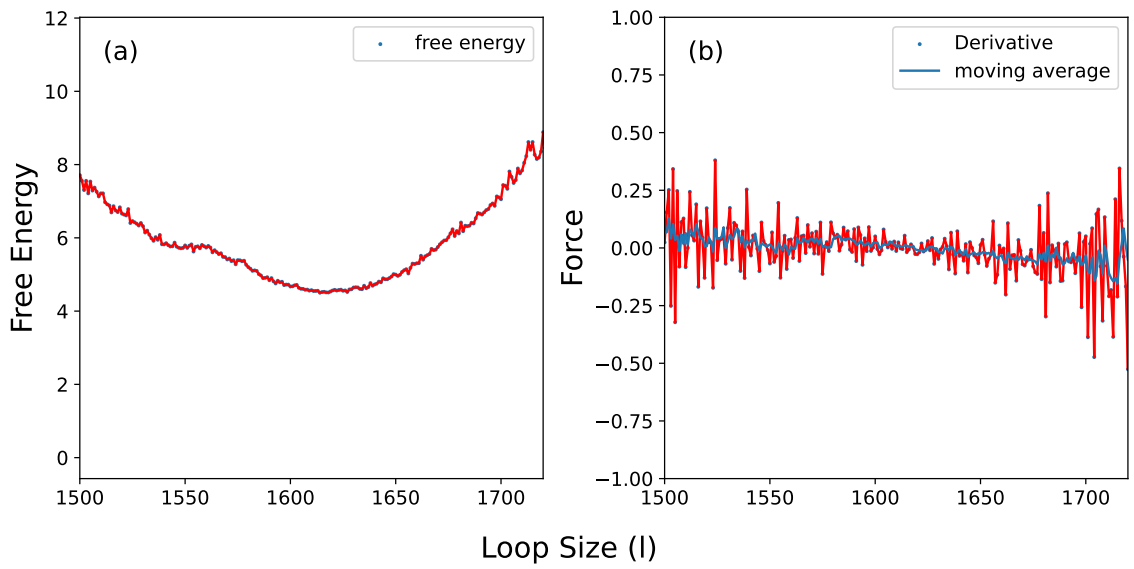


Figure 3.22: **Free energy and the derivative as a function of loop size in the absence of activity.** The (b) is just the derivative of the free energy and the moving average is plotted by taking a window size of 5.

extrusion forces act to bring it back to the mean, and if the size becomes more than the mean, the forces due to fluctuations of the rest of the chain bring it back. These counterbalancing forces are comparable in magnitude only in this regime.

3.8 From Extrusion Forces to Stalling Forces

Evaluating the stalling forces for different values of extrusion forces is a non-trivial task. This is because characterizing the force due to fluctuations of the chain is extremely hard. Further, in addition to the fluctuations of the chain not part of the loop, there are also fluctuations of the loop itself which might be influencing the possible equilibrium configuration of the system. Taking into account all of these, it seems naive to call the force evaluated by taking the derivative of the free energy as the stalling force of the system. Adding to this, there is also a problem of resolution in the derivative plot owing to the discrete nature of the size of the loop.

We coarsely measure the stalling forces using the extrusion forces and try to get a qualitative relation between how the stalling forces change as the size of the loop. The impetus for doing so is as follows: The system attains some mean value for the loop size after it achieves equilibrium. This means that the extrusion force which is causing an increase in the size of the loop is counter balanced by the stalling force almost exactly, as such for the particular loop size, we have an estimate of the stalling force which is equal to the extrusion force. This claim is by no means exact, but it gives us as estimate of the stalling force as a function of the loop size.

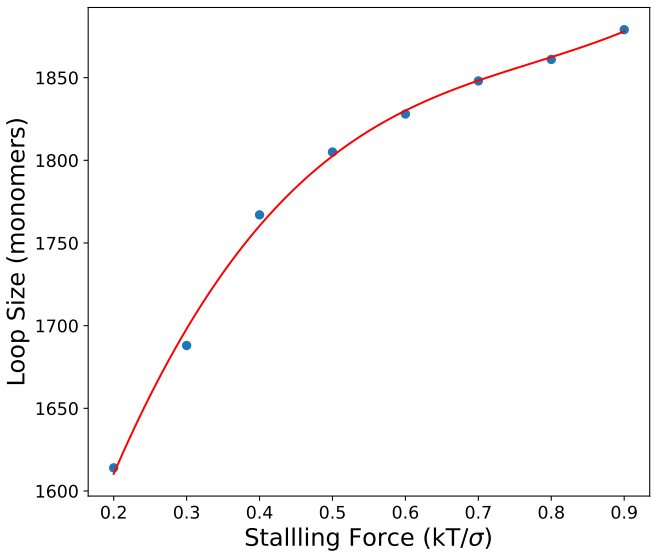


Figure 3.23: **Stalling force as a function of the size of the loop.** The regime under observation is still restricted to the thermal energy scale. The data points have been obtained by taking the minima of the free energy for every value of the active force.

A similar plot has presented in figure 3.10 but here we use the loop size instead of relative extension. Also, instead of just averaging, the free energy landscape has been used to get the most optimal size of the loop. For activity values starting from 0.1, every 0.1 step, we ascertain the probability distribution and find the minima for the free energy before taking the corresponding value for the loop size as the optimal size. We get a overall similar kind of curve as before with more inference on how the size of the loops changes with the stalling force.

3.9 The Force Extension Curve: A Reason for Stalling

To analyze the stretching of polymer in the presence of external forces, enthalpic as well as entropic contributions have to be considered [23]. Both these factors are essential when considering the polymer elasticity. The force applied to stretch the polymer chain goes into reduction of the conformational entropy of the chain [27]. In a simple example of an ideal freely-jointed chain, the there are no enthalpic contributions because of the absence of excluded volume interactions [28], and the entropic contributions result in a behavior no different from Hookean spring. So, the equation for force-extension can be written as [23],

$$f = \frac{k_B T}{l_p} \frac{3}{2} z \tag{3.13}$$

where z is the extension, given by r_{ee}/L where r_{ee} is the end to end distance of the polymer chain and L is the contour length, l_p is the persistence length.

The DNA is a semi-flexible polymer, and even in our coarse-grained description, the DNA is modeled using a semi-flexible polymer. So the simple expression of Hookean spring is no longer valid here. But worm-like chain (WLC) model can be used to describe polymers that are semi-flexible [27]. So, effectively, if we have a force-extension relation for a WLC model, then it can effectively give us a force-extension relation for the DNA. Till date, there is no analytic expression for this, but there is a interpolated expression pioneered by Marko and Siggia [27] that matches with the force-extension curve of the DNA quite well. The interpolation formula is given by,

$$f = \frac{k_B T}{4 l_p}\left[4z + \frac{1}{1-z^2} - 1\right]. \tag{3.14}$$

This expression fits the force-extension curve of a semi-flexible polymer as shown for DNA by Marko and Siggia [27]. To test it, we graft our semi-flexible polymer of 3026 monomers to a wall at one end, and apply force to the monomer at the other end. There are excluded volume interactions in our polymer chain, so it is not quite a WLC chain, but we would like to check the validity of the equation before we move on to our grafted chain with the handcuff.

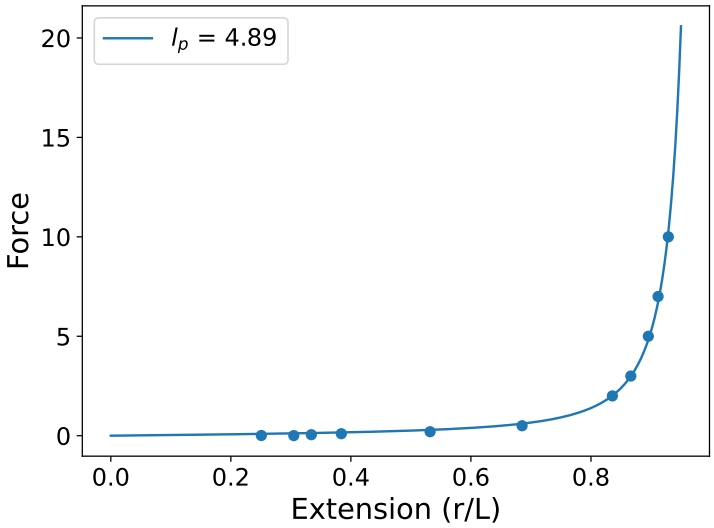


Figure 3.24: **Force extension curve of a linear chain grafted at one end.** A constant force is applied to the last monomer and the corresponding extension in the polymer chain is calculated by finding the distance between the first and the last monomer and dividing it by the total contour length.

To compute the extension in the polymer chain, we first calculate the distance between the first and the last monomer, which is just the end to end distance, r_{ee}. Diving it by the total contour length would give us the corresponding extension of the chain. We perform multiple simulation, each time using a different force and letting the system equilibrate before extracting data

for analysis. As we can see from the plot, the data points fall on the curve, and the persistence length from the fit comes out to be 4.89. We made sure our semi-flexible chain resembled the DNA with persistence length 35nm.So, the persistence length in simulation units is,

$$l_{\mathrm{p}} = \frac{3026 \times 35\mathrm{nm}}{16167\mathrm{nm}} = 6.55\mathrm{a.u.} \tag{3.15}$$

From the plot 3.24, we get a persistence length that is quite close considering the fact that equation 3.15 does not take into account excluded volume interactions. This clearly tell us that we can analyze the force-extension curves of semi-flexible polymer to a reasonable accuracy using the above equation.

Now, we try to use the equation in our extrusion model. We start with the initial configuration and let the system extrude until an equilibrium state has been reached. The distance between one of the rings of the handcuff and a grafting point is the end to end distance of one half of the chain that is between them and hasn't been extruded. If we divide this end to end distance by the total length of the polymer between the handcuff and the grafting point, we get a value for extension. Of course, this value changes because of the fluctuations of the polymer as well the handcuff diffusion, but it doesn't affect the analysis. The stretching force here is just the activity of one of the ring, because this is precisely the force with which the linear chain is being pulled during extrusion.

The persistence length is quite close to the actual value even though the fit is not quite perfect 3.25. This tells us that the there is some resistance to stretching during the extrusion, and it is this force that stalls the extrusion despite the presence of more monomers that can be pulled through the handcuff. As the extension increases, that is, more and more monomers are extruded, the force needed to extrude also increases and there is a point where the activity is balanced by this force and extrusion stalls. This also explains the decrease in the rate of extrusion over iterations.

Motivated by the above result, we try to extrapolate this analysis to the entire chain. This simply means that we are using both segments of the linear chain, and using twice the force because both rings of the handcuff are involved in the process of extrusion. Here, instead of using one parameter fit, we now use two parameters to get the optimum fit. The fit parameters are just end to end distance and the persistence length. This is because the total length of the linear segments changes because of the extrusion as more monomers are added to the loop with the increase in extrusion force. Hence, the constant parameter is the end to end distance which is just the sum of the distance between the handcuff and the wall on either side.

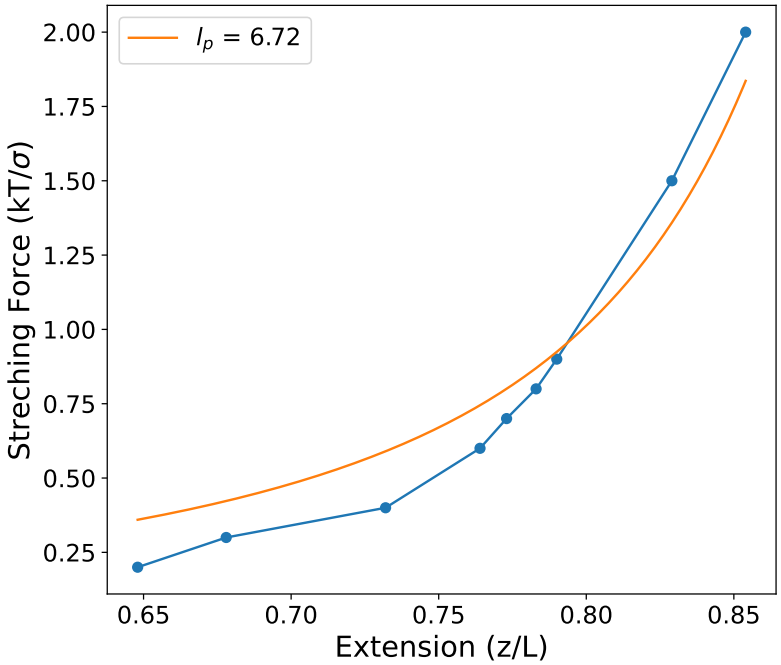


Figure 3.25: **Force extension curve for the one half of the linear chain between the handcuff and grafting point**. The extension is calculated by taking the ratio of distance between the handcuff and one of the grafting points and the total contour length of the linear segment between them. The stretching force is just the activity.

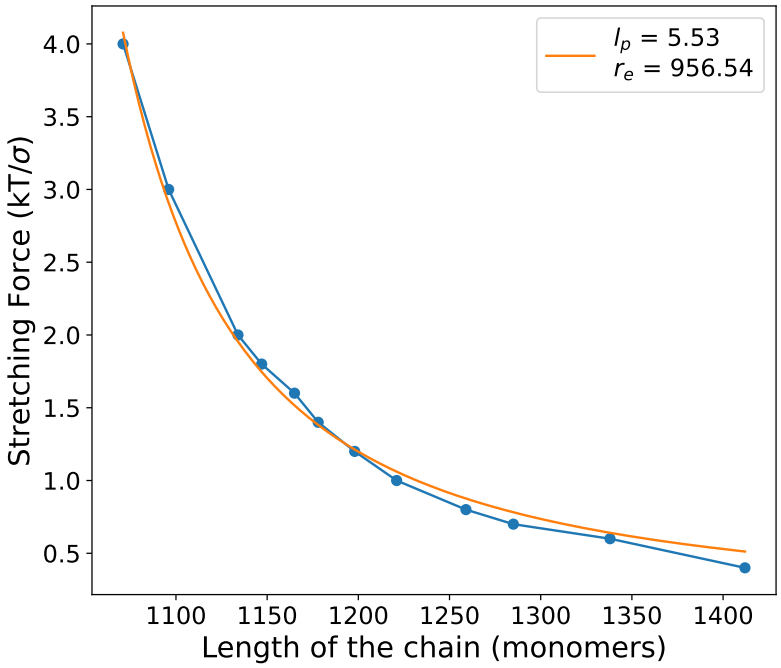


Figure 3.26: **Force extension curve for both linear segments between the handcuff and the grafting points.** The total length of the chain is just the sum of both the linear segments. The two parameters in the fit are the persistence length and the end to end distance.

As we can see from the plot 3.26, the persistence length is quite close to the persistence length of our semi-flexible polymer. The grafting distance is 910, as we already know from the parameter section. The higher value of end to end distance in the fit parameter can be explained from the fact that the handcuff lies in a different plane from that of the grafting points, as such the end to end distance is greater than the grafting distance. Further, the end to end distance fluctuates even after the system has attained equilibrium. The exact value of the end to end distance fluctuates around 910 to 940, so the fit is on the higher side.

From the analysis done in this section, we can claim that the extrusion stalls because of entropic stretching. This actually explains many results we saw in the previous sections like complete extrusion at higher values of activity, decrease in extrusion speed over time, and, prominently, it provides a reason for why the entrusion stalls in both the experiments as well as the simulation.

3.10 Effect of Graft Distance on Loop Size

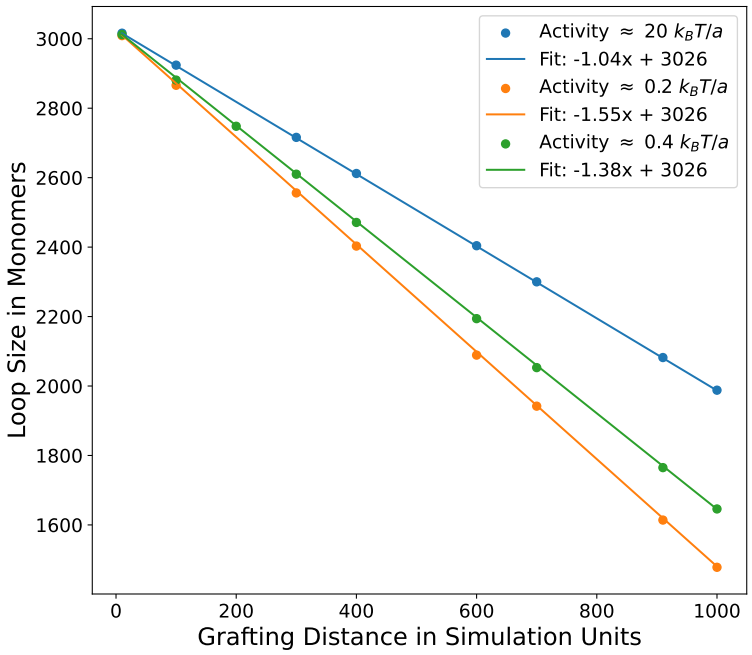


Figure 3.27: **The effect of grafting distance on the loop size for difference activity values**. All the lines are linear, which is to be expected since as the grafting distance increases, there is a linear decrease in the number of monomers available for extrusion. The intercept is 3026, which is just the total number of monomers when the grafting distance is zero, hence it is not a fit parameter.

The distance between the grafting points was held constant throughout the analysis performed in the last few sections. Now, we try to ascertain the effects of changing the grafting distance on the size of the loop formed by the extruded portion of the polymer. The simulations

are performed in both regimes, i.e. when the active forces are comparable to the thermal energy scales as well as when they are much higher, and we give the same value of the extrusion forces for all the grafting distances. All independent parameters apart from the grafting distance are held constant.

The relation between loop size and the grafting distance is linear. This doesn't come out as a surprise, since as the distance between the grafting points decreases, the number of monomers available for extrusion increases linearly. But the interesting part lies in the slope of the curve. For the regime where entropic effects are prevalent (activity value is 0.2), we can clearly that the slope of the fit is -1.54 instead of -1. This means that there is something else in play other than just the linear increase in the number of monomers. The possible explanation for this factor lies in the prevalent fluctuations of the chain in this regime. When the grafting distance is high, due to the increased fluctuations of the chain between the grafting points and the handcuff, extrusion stalls earlier. Now, the affects of these fluctuating forces keep on reducing as we decrease the grafting distance owing to which the loop size increases faster than the rate at which the monomers are being freed up. This explains the higher magnitude of the slope in the figure 3.27.

The control system is also plotted in the figure 3.27, where for performing extrusion the activity values are set to 20 k_BT/σ which are much higher than the permissible thermal energy scales. From the slope, it is evident that in this regime there is one to one correspondence between the increase in the grafting distance and the decrease in the size of the loop. The slope of the fit is one, which indeed supports the claim that the entire increase in the grafting distance is found in the size of the loop. This points to the presence of entropic effects in the other regime, where activity values are low.

The simulation snapshots have been captured using ovito [17] in figure 3.8 after the system has attained equilibrium. We can clearly see that when the activity value is much higher than the thermal energy scales, the polymer between the handcuff and the grafted ends is completely taut as in figure 3.8b, whereas when the activity values are low, the polymer hasn't been completely extruded as seen in 3.8a even though the extrusion has stalled. The handcuff along with the rings is hidden beneath the assemblage of yellow splotches, and interestingly, it diffuses along the length of the contour in both the cases.

3.11 Extrusion in Flow

We also perform extrusion in the presence of flow to replicate the experiments performed in buffer flow. A constant force normal to the wall was applied to every monomer followed by extrusion. The flow force was kept small so as to let the extrusion forces dominate and allow

maximum extrusion. The initialization and equilibration of the system was done as before: we initialized the polymer in a rectangular configuration and let it relax without flow before inserting the handcuff. The flow was applied to this final configuration, and as is evident from the simulation snapshot, the polymer opened up quite nicely before the extrusion started, and the yellow splotches from before formed the extended loop like that in the experiment [4]. We haven't exactly mapped the flow simulation to the experiments, so there is some work to be done in this direction.

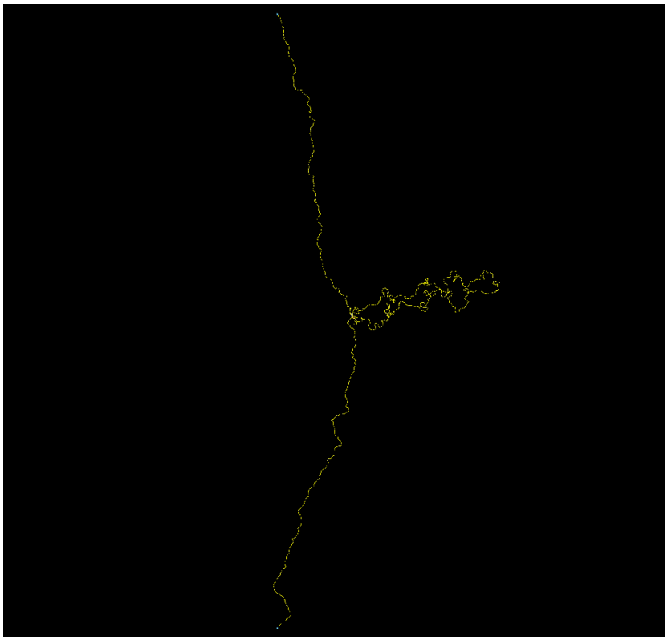


Figure 3.28: **Simulation snapshot of extrusion in the presence of flow captured using ovito [17].** The extrusion force in the snapshot is 0.2 k_BT/σ and flow force is set to 0.0001 k_BT/σ. All the monomers are colored yellow, except the handcuff, which is not visible in the snapshot.

3.12 Entropy, Loops, and Conformations

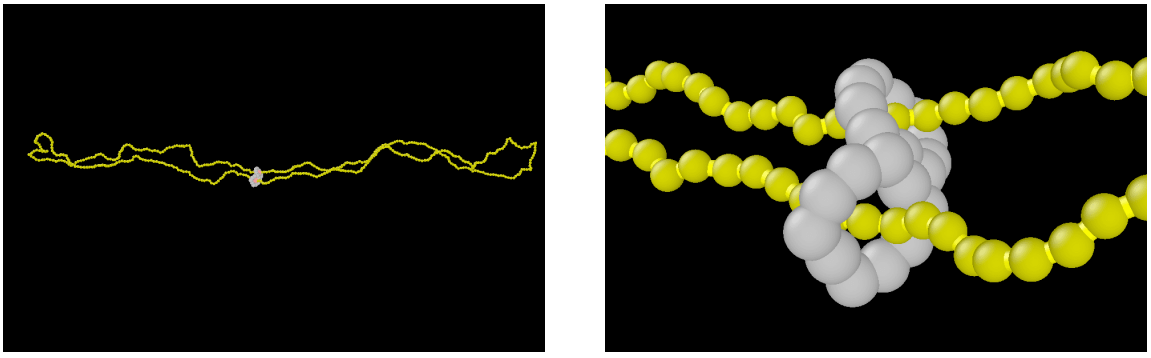
So far, through the previous sections, we were able to get a working model of extrusion that matched the length and time scale of the experiment. Further, we were also able to explain the stalling of extrusion before the chain becomes entirely taut. But this explains the experiment where the DNA is grafted at the two ends, but in the real systems this is not usually the case. The role of entropy might be important in vivo than the force-extension relation, so we plan to lay a preliminary foundation for further investigation through this section.

The task of quantifying entropic force is non-trivial, and we have yet to find a suitable schema that lets us measure the entropic forces existing in our system. This section stands to emphasize the fact that it's not just a weak force that doesn't have any physical significance, rather, more often than not, it plays a vital role in deciding the conformation of a polymer. In most general

scenarios, all system would tend to move toward configurations of increasing disorder, and the polymer systems are no different.

The ring polymers are a class of topologically constrained molecules. One simple act of join two ends of a linear polymer changes the dynamics of a single molecule drastically. Loops of a fixed size are nothing but ring polymers because they have a common point of contact. Of course, in our case, we have a loop that changes in size throughout the course of the simulation, but the system reaches a stage where the relative extension is fluctuating about a mean value, which means the loop size does too. The presence of loops in a linear chain, or in a ring polymer for that matter, decreases the number of accessible configurations. Hence the entropic forces always try to maximize the total conformations possible by decreasing the loop sizes.

The model used here is slightly different from the one used to model the SMC mediated loop extrusion. The linear polymer used throughout has been closed at one end to get a circular polymer, and it is no longer grafted. The rest of the parameters and modeling details are no different from the extrusion model. The polymer is still made semi-flexible using the three body potential, and the magnitude of the potential has been kept the same. The WCA interactions give some definite volume to the monomers and prevent their overlap and FENE potential ensures that the bonds between the neighbors are not essentially rigid. To summarize, we have a ring polymer consisting of 500 monomers, and a rigid handcuff with two rings, ten particles each. The number of monomers are arbitrary and the study could have done with any number of monomers to reproduce similar behavior.



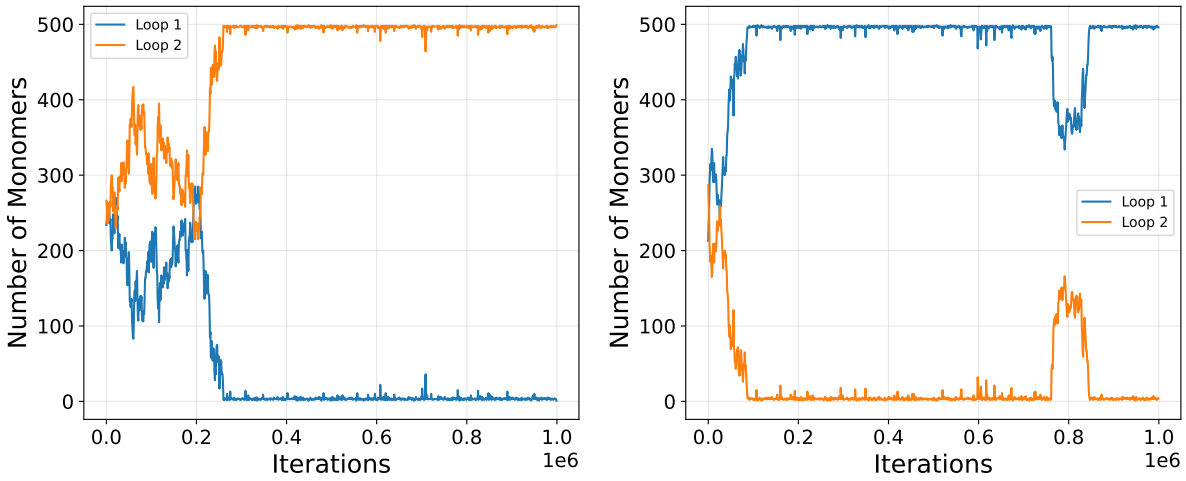
(a) Snapshot of the entire system (b) Snapshot of the handcuff

Figure 3.29: **A simulation snapshot of a ring polymer with a handcuff placed at the middle** [17]. The polymer is passed through both the rings of the handcuff giving rise to two loops of almost equal sizes.

To initialize the system, the linear polymer is passed through both the rings of the handcuff before the first and the last monomer are joined using FENE bond potential. The handcuff divides the ring polymer into two loops, and we made sure to keep the loops almost similar in

size with similar number of monomers. The snapshots in figure 3.29 are after the first thousand iterations, and it evolved from a initial configuration which was just a rectangle to a distorted rectangle. The scheme used for integration is again Langevin with the same dt (0.01) used for extrusion simulation, and we the ran the system for about 100 million iterations.

The system evolves until equilibrium before performing an analysis, and we would expect one of the loops to become very small. This is because the fluctuations of the either of the loops would tend to minimize the other, and move toward a configuration which is entropically favorable. In this case, since having one loop is unavoidable, either of the one would get minimized. The forces when the ring is in the middle might be similar at the beginning, but owing to the randomness of the system due to the fluctuations, the handcuff wanders about the central position, until the force from one of the loop dominates and completely pushes the handcuff toward one side.



(a) Loop 2 diminishes Loop 1

(b) Loop 1 diminishes Loop 2

Figure 3.30: **The size of loops of the ring polymer over the course of the simulation**. The handcuff is placed at the middle of the ring polymer, so either the left or the right loop can diminishe the other. The two plots are from different simulations, indicating that it is completely random which loop gets to dominate.

The figure 3.30 clearly drives home the point that one of the loop will always diminish the other. Ten independent ensemble runs were given, each with the handcuff in the middle as the initial configuration. About almost half the time, the loop 1 eliminated the other, and vice versa in the rest. This is a clear implication that the entropic forces are indeed involved in deciding the configurations possible when the energy scales are in of the order of the thermal energy. Plots for other ensembles were not included in the analysis to avoid redundancy. In some cases, the handcuff overcame the energy barrier for loop formation, and did give rise to loops as show in 3.30b, but they were transient and the configurations were not energetically favorable.

3.13 Supercoiling: Analysis and Drawbacks

In the next few sections, we would look at the results of the supercoiling model and present some of the key drawbacks to state why the model did not work as intended. Our primary goal of coming up with a model was to see if the springs would give rise to four body potentials and help us in generating torsional strain in the system without the need for invoking any four body potentials. Even though the model worked, we failed to keep the underlying polymer model intact. As such, we got a mechanically correct model, but it was far from a normal working model of a polymer.

3.13.1 Diagonal Connections and Semiflexibility

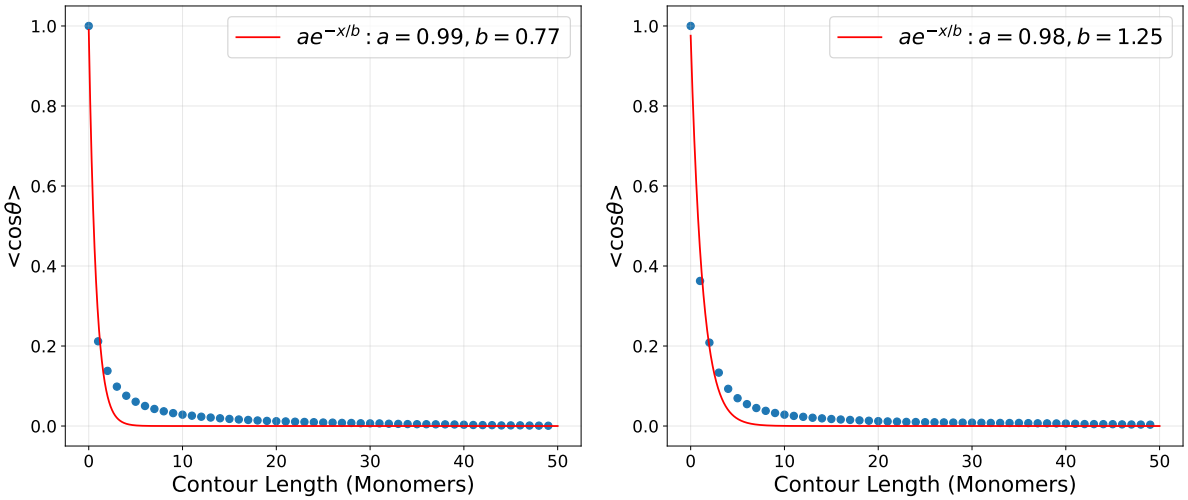
The virtual particles are the point of attachments of the springs on top of the monomers, where the harmonic potential acts. The number of virtual particles on the monomers are arbitrary, and they can vary from two all the way to very large numbers. The more the number of virtual particles, the higher is the computation cost, hence we need to decide on the optimum value which effectively translates the torsion to writhes. We analyze the effects of the number of virtual particles in this section.

To check for the semiflexibility, we measure the tangent-tangent correlation function. This is simply the angle between two tangent formed by joining any two monomers, averaged over the entire chain. The correlation decays as the monomers become separated in space. Mathematically, the expression can be written as,

$$\langle cos\theta \rangle = \exp^{-\frac{l}{P}} \tag{3.16}$$

where P is the persistence length and l is the contour length.

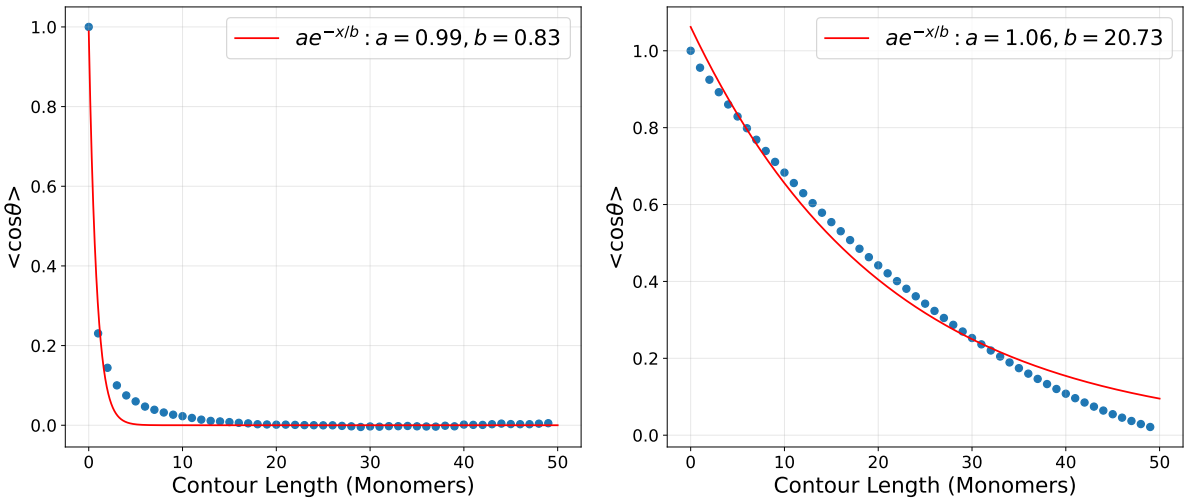
We first perform the simulation with two virtual particles placed in the diametric opposite direction with only linear connections. In this case, the polymer acts no different from a normal bead spring polymer with a single spring attached from the center. There is no bending cost either in this case, and the polymer essentially behaves like a flexible polymer with a higher springs constant. The κ of all the springs is maintained at 100. Now, we introduce diagonal connections in the springs, and even though the persistent length increases marginally, it's nothing significant. This stands to state the diagonal springs are responsible for introducing some kind effective bending cost in the system, though it is not very evident in case of two virtual particles. We plot persistence length as a function of the contour length in the figure 3.31 for both these scenarios, and we can see that persistence length increases when there are diagonal connections.



(a) No diagonal connections

(b) Diagonal and linear connections

Figure 3.31: **The tangent-tangent correlation function plotted against the contour length in monomers for two virtual particles**. The fit parameter b gives the persistence length in the plot. In the absence of diagonal connections, the system effectively acts like a flexible polymer. Even though diagonals connections do increase the persistence length, it's quite insignificant in the case of two virtual particles.



(a) No diagonal connections

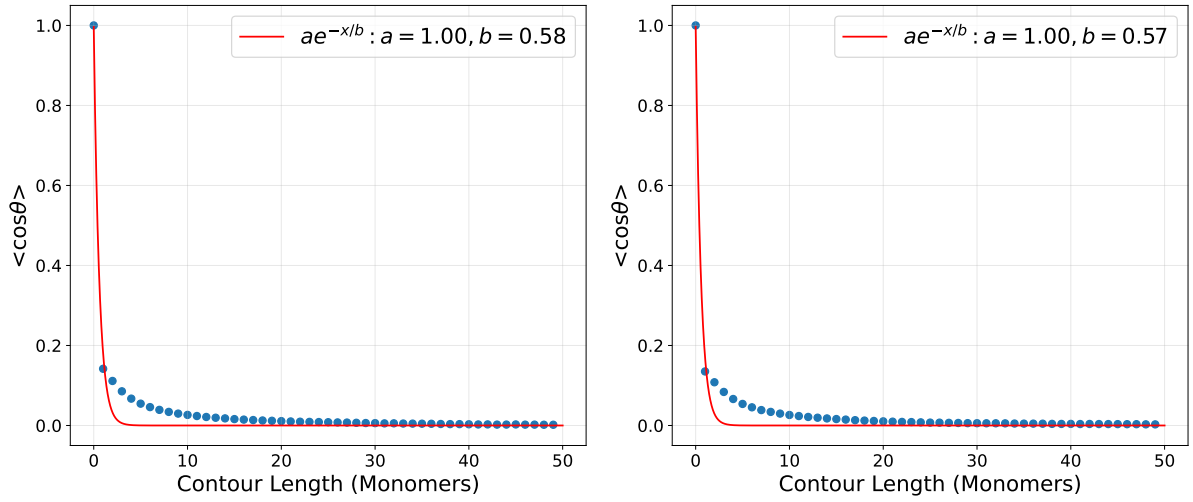
(b) Diagonal and linear connections

Figure 3.32: **The tangent-tangent correlation function plotted against the contour length for three virtual particles**. The persistence length increases significantly in the presence of diagonal connections, whereas the polymer remains flexible in the presence of linear connections only.

The analysis is extended to three virtual particles, and we measured the persistence length in the two cases. When only linear springs were present, the persistence remained below one, an obvious indication of the fact that the linear springs do not contribute to the stiffness of the polymer chain. In the presence of non-linear springs, i.e. when the non-adjacent virtual particles were also connected, the persistence length increases quite significantly, and this increase

depended on the strength of the springs. For the 100 spring constant, the persistence length increased closed to 21 as is evident from the plot.

A similar behavior can be observed for higher number of virtual particles. The polymer with linear connections only always show similar decay of the persistence length to account for the flexibility. The reason is that the polymer, in the presence of only linear connections, can bend via sheer and compression without actually changing the length of the linear springs. This effectively allows the polymer to attain configurations which would not be possible in the presence of diagonal springs as any sheer or compression would change the length of the diagonal and linear connections in the opposite manner. The figure 3.33 drives home the point that semiflexibility arises when there are both diagonal and linear connections.



(a) Four virtual particles

(b) Five Virtual Particles

Figure 3.33: **The tangent-tangent correlation function plotted against the contour length for linear connections**. The polymer remains flexible in the presence of only linear connections irrespective of the number of virtual particles.

3.13.2 Spring Energies

In equilibrium, when there are no external forces present, all the harmonic spring energies would fluctuate around $0.5k_BT$, if they are independent degrees of freedom, which is the case if there is one spring between the two monomers. When there are multiple springs connecting the two monomers, not all the springs are independent, as such individual springs might fluctuate at values lesser than $0.5k_BT$. In the presence of compression or torsion, the spring energies would increase, and they would fluctuate at higher values.

The simulation for computing the spring energies was done with 120 monomers, each of which had six virtual particles. The system was allowed to equilibrate first, following which, the

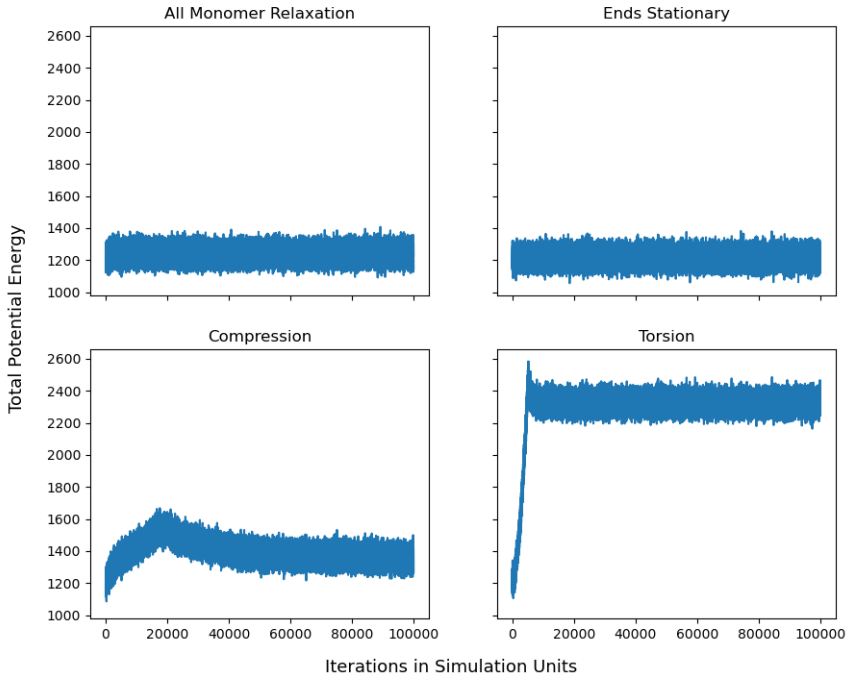


Figure 3.34: **Potential Energy of the spring over iterations.** The total harmonic spring energies have been plotted for the springs interconnecting two monomers. The springs within the same monomers have been ignored for all calculations.

energies were extracted over time. In the compression case, the monomers at the end were compression by 40 simulation units and the corresponding energy of the springs was calculated. As we can see from figure 3.34, holding the ends stationary does not affect the springs in any manner, and compression increases the energies, which is as expected. In the torsion case, the first monomer was rotated eight times, and the torsional strain was translated across the whole chain over the course of iterations. Almost most of the springs changed in length, and the total energy of the springs almost doubled in the torsion case.

3.13.3 Formation of Plectoneme like structures

The general idea for the formation of plectonemes is the conversion of twists in the system to writhe. For this, we start the simulation with the usual polymer chain where the first and the last monomer are held fixed. The κ of all the interconnecting springs was maintained at 300, which was high enough to effectively relax the torsional chain in the polymer by formation of writhes. Again, it was possible to generate the writhes wtih lesser κ, but we needed to increase the torsional strain by giving additional rotations.

The number of monomers in the system were fixed to 60, each containing six virtual particles containing all possible springs. The system was first compressed slowly by 20 simulation units

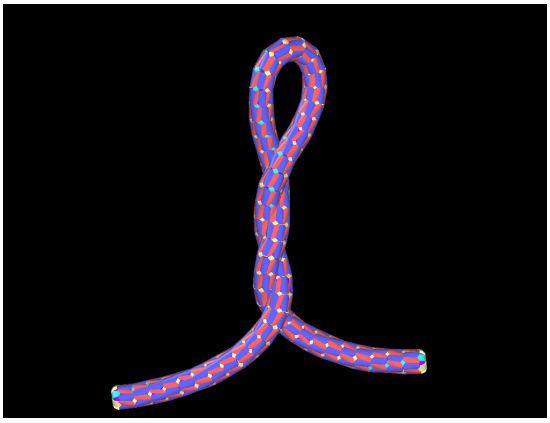


Figure 3.35: **Simulation snapshot of a single plectoneme like structure.** The bonds are increased in size for ease of visualization.

over the course of 10 million iterations. In the entire process, the first and the last monomer were held fixed, that is, they were not integrated with the other monomers, rather compressed manually. Later, the first monomer was rotated for six *pi* rotations over 4 million iterations and the torsion was allowed to propagate through through the length of the polymer. The consequent linking number in the simulation was 6, and the maximum writhes that could form in the system was 6. Not all the torsional strain was converted to writhes as the system relaxed, and some twist still persisted in the chain as is evident from the figure 3.35.

3.13.4 Multiple Plectoneme like Structures

Getting more than one plectoneme isn't possible just by twirling one end and holding the other in place. It always results in a single writhe irrespective of the torsional strain in the system. So, generating more than one of these writhes is a multi-step process. The entire chain is divided into two parts, and writhes are generated individually in both the segments before letting the chain relax.

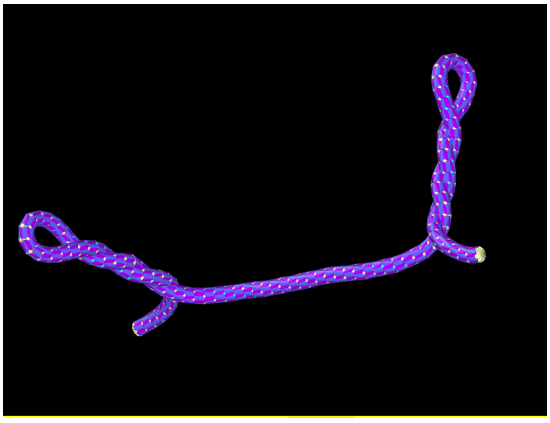


Figure 3.36: **Simulation snapshot of a two plectoneme like structure.** The bonds are increased in size for ease of visualization.

The total number of monomers here was fixed to 120 which is just twice the number of monomers we used in generating the single plectoneme. Initially, the first and the central monomer were held fixed and the system was compressed by 20 simulation units as before. Later, the first monomer was rotated to generated torsion, and the system was allowed to relax until it formed the first plectoneme. This process was repeated for the other half of the chain, where the compression was done as usual by moving the last monomer toward the middle one, and the last monomer was rotated in the opposite direction to generate necessary torisional strain in the system. Once the system relaxed, two plectoneme like structures were generating in the countour as shown in figure 3.36.

The problems of the model become very evident here, as the writhe formed as such remain stationary without change even after the system has relaxed for a long time. In real systems though, plectonemes diffuse [29] along the the chain contour, or they undergo hopping [29] during which a plectoneme suddenly disappears and reappears at another site on the contour of the DNA simultaneously in accordance with the conservation of the linking number. For any of these processes the occur in our model, there is a very large energy barrier which the system has to overcome and the chances of this happening are almost negligible. As such, even though mechanically correct, the model hardly resembles the supercoiled DNA.

Chapter 4

Conclusion and Future Work

In our work, we have been able to come up with a model for loop extrusion that effectively captures the statics and dynamics of the loop extrusion observed in the experiment. Our model was parameterized to the experiments, and we were able to map both the length and time scale of our simulation to those of the experiments and compare our results. The model has given interesting insight into the physics behind the stalling of extrusion, and how the force-extension and fluctuations of the chain plays an important role in stalling the extrusion when the extrusion forces are of the order of thermal forces. This also explains why the extrusion has stopped much before the the DNA has become taut even in the experiments.

Apart from the simulation results, the interesting aspect is how nature is energy efficient. In our simulation, the minimum force required for the extrusion is close to 0.05 k_BT/σ, whereas to simulate the real system we would need a force close to 0.2 k_BT/σ. This is just above the threshold, which is an indication of the fact that not much energy is spent in getting to the required conformation, even though there is potential to do so. There are also many other SMC complexes, which might have different extrusion forces, but it's possible that most of them will be in the regime where entropic effects are important.

Generating torsion in the polymer by inclusion of multiple springs did not work as intended. The model worked well mechanically, but showed none of the characteristics of plectonemes in real systems. The alternative way to model the system is to use the four body potential [30], which includes use of virtual particles in addition to the monomers, albeit only while calculating the dihedral angle.

The future work on extrusion entails making the theoretical description for the stalling of extrusion more concrete. There are many unanswered questions, importantly the quantification of the entropic forces in play during the extrusion. Once, we have a solid framework in place, we will make quantitative predictions on how the graft distance is related to the relative extrusion, influence of ionic conditions on the loop size, the effects of DNA size on the extrusion, and try

to verify our claims against the experiments.

We also aim to investigate the mismatch between the time scale of DNA displacement and the extrusion. Firstly, it would be necessary to perform both the experiments under the same conditions so as to remove all the experimental variables. We would then see if performing simulations in the overdamped limit, and whether switching to Brownian dynamics would solve the problem of time scale mismatch. We would also like to investigate if modeling the extrusion like in the real system [26] would help us solve this problem. To perform these simulation, we have to improve the efficiency of the code first, and maybe writing our own GPU script might be the way out should the available libraries not work out.

The SMC mediated extrusion is not limited to linear DNA, but rather, it also occurs in plasmids, or other circular forms of DNA [31]. We also to aim to extend our model to perform extrusion in a ring polymer in line with the experiments.

Our model explains the stalling of extrusion in the experiments and simulations, but the actual mechanism in vivo might be different. There is no grafting in vivo, so the role of extension mediated stalling might be much lesser and entropy might play more important role if extrusion stalls at all. We would like to continue our investigation in ring polymers and see whether the extrusion stalls because of the fluctuations of one half of the loop. Further research would also involve investigating the topological effects of DNA like occurrence of knots, and the incorporation of a model which allows for supercoiling and the study of supercoiling in the context of extrusion.

Bibliography

[1] Audrey Borus. *James Watson, Francis Crick, Rosalind Franklin, and Maurice Wilkins: The Scientists who Revealed the Structure of DNA*. The Rosen Publishing Group, Inc, 2020.

[2] Petra Anne Levin and Esther R Angert. "Small but mighty: cell size and bacteria". In: *Cold Spring Harbor perspectives in biology* 7.7 (2015), a019216.

[3] Nigel Chaffey. *Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular biology of the cell. 4th edn.* 2003.

[4] Biswajit Pradhan et al. "The Smc5/6 complex is a DNA loop-extruding motor". In: *Nature* (2023), pp. 1–6.

[5] Iain F Davidson and Jan-Michael Peters. "Genome folding through loop extrusion by SMC complexes". In: *Nature reviews Molecular cell biology* 22.7 (2021), pp. 445–464.

[6] Chris R Calladine and Horace Drew. *Understanding DNA: the molecule and how it works*. Academic press, 1997.

[7] Ivan Junier et al. "DNA supercoiling in bacteria: state of play and challenges from a modeling viewpoint". In: *arXiv preprint arXiv:2303.10819* (2023).

[8] Thomas E Gartner III and Arthi Jayaraman. "Modeling and simulations of polymers: a roadmap". In: *Macromolecules* 52.3 (2019), pp. 755–786.

[9] Martin Karplus and J Andrew McCammon. "Molecular dynamics simulations of biomolecules". In: *Nature structural biology* 9.9 (2002), pp. 646–652.

[10] Loup Verlet. "Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules". In: *Physical review* 159.1 (1967), p. 98.

[11] Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press, 1996.

[12] Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford university press, 2001.

[13] Joshua A Anderson, Jens Glaser, and Sharon C Glotzer. "HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations". In: *Computational Materials Science* 173 (2020), p. 109363.

[14] Gerard T Barkema, Debabrata Panja, and JMJ Van Leeuwen. “Semiflexible polymer dynamics with a bead-spring model”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2014.11 (2014), P11008.

[15] Sarah Wettermann, Ranajay Datta, and Peter Virnau. “Influence of ionic conditions on knotting in a coarse-grained model for DNA”. In: *Frontiers in Chemistry* 10 (2023), p. 1627.

[16] N. W. Lima et al. “Molecular dynamics simulation of polymerlike thin films irradiated by fast ions: A comparison between FENE and Lennard-Jones potentials”. In: *Phys. Rev. B* 94 (19 Nov. 2016), p. 195417. DOI: 10.1103/PhysRevB.94.195417. URL: https://link.aps.org/doi/10.1103/PhysRevB.94.195417.

[17] Alexander Stukowski. “Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool”. In: *MODELLING AND SIMULATION IN MATERIALS SCIENCE AND ENGINEERING* 18.1 (JAN 2010). ISSN: 0965-0393. DOI: {10.1088/0965-0393/18/1/015012}.

[18] Farid F Abraham and Y Singh. “The structure of a hard-sphere fluid in contact with a soft repulsive wall”. In: *The Journal of Chemical Physics* 67.5 (1977), pp. 2384–2385.

[19] Johann Fischer and Martin Wendland. “On the history of key empirical intermolecular potentials”. In: *Fluid Phase Equilibria* 573 (2023), p. 113876.

[20] Dirk Stigter. “Interactions of highly charged colloidal cylinders with applications to double-stranded DNA”. In: *Biopolymers: Original Research on Biomolecules* 16.7 (1977), pp. 1435–1448.

[21] Valentin V Rybenkov, Nicholas R Cozzarelli, and Alexander V Vologodskii. “Probability of DNA knotting and the effective diameter of the DNA double helix.” In: *Proceedings of the National Academy of Sciences* 90.11 (1993), pp. 5307–5311.

[22] Debarshi Mitra and Apratim Chatterji. “Transient helix formation in charged semiflexible polymers without confinement effects”. In: *Journal of Physics: Condensed Matter* 33.4 (2020), p. 044001.

[23] M Rubinstein and RH Colby. *Polymer Physics Oxford Univ. Press*. 2003.

[24] Suyash Naik et al. “Differential tissue stiffness of body column facilitates locomotion of Hydra on solid substrates”. In: *Journal of Experimental Biology* 223.20 (2020), jeb232702.

[25] Jürgen Renn. “Einstein’s invention of Brownian motion”. In: *Annalen der Physik* 517 (2005), pp. 23–37.

[26] Cees Dekker et al. “How do molecular motors fold the genome?” In: *Science* 382.6671 (2023), pp. 646–648.

[27] John F Marko and Eric D Siggia. “Stretching dna”. In: *Macromolecules* 28.26 (1995), pp. 8759–8770.

[28] Xiaolan Li, Charles M Schroeder, and Kevin D Dorfman. "Modeling the stretching of wormlike chains in the presence of excluded volume". In: *Soft Matter* 11.29 (2015), pp. 5947–5954.

[29] Marijn TJ van Loenhout, MV De Grunt, and Cees Dekker. "Dynamics of DNA supercoils". In: *Science* 338.6103 (2012), pp. 94–97.

[30] Fabrizio Benedetti et al. "Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes". In: *Nucleic acids research* 42.5 (2014), pp. 2848–2855.

[31] Biswajit Pradhan et al. "Loop extrusion-mediated plasmid DNA cleavage by the bacterial SMC Wadjet complex". In: *bioRxiv* (2024). DOI: 10.1101/2024.02.17.580791. eprint: https://www.biorxiv.org/content/early/2024/02/17/2024.02.17.580791.full.pdf. URL: https://www.biorxiv.org/content/early/2024/02/17/2024.02.17.580791.

[32] Kerson Huang. *Statistical mechanics*. John Wiley & Sons, 2008.

[33] Daniel V Schroeder. *An introduction to thermal physics*. 1999.

[34] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.

Appendix A

The statistical physics of polymer is the descriptive study of the possible conformers as well the the dynamics of these conformations over space and time. The polymers are macromolecules which are of great importance, both from a biological as well as a technological point of view. The space and time scales under study throughout the course of the thesis are much larger than atomic scales, so classical statistics would suffice to describe most of the dynamics seen in the model. Prominently, concepts of energy, temperature, free energy, entropy, would be recurrent, as such, they will be addressed briefly in this section.

A.1 The Ensembles and Energies

In a mechanically and adiabatically isolated system, there is no exchange of heat or matter between the system and the surrounding. So, the energy of the system, the number of molecules as well the total volume is always a constant. The set of microstates of this kind of system form the microcanonical ensemble [32]. Unless perturbed externally, there is no way for the system to change its energy over time. The conserved internal energy of this system has a simple form, which can be derived from first principles [32] or just written based on microscopic observations in the guise of the first law of thermodynamics.

$$dU = TdS - PdV \tag{A.1}$$

where the first term is just the heat, and the second term is the work done.

The isolated system could be modified and kept in touch with a heat bath. This breaks away the adiabatic isolation and heat can enter into the system. As such, the energy is no longer a constant, but the thermal bath ensures that the temperature of the system becomes the sames as the bath. The zeroth law of thermodynamics gives the macroscopic validity of this claim, but this can be derived even from the microstate picture [32]. The collection of microstates of such a system in which the temperature is kept constant is called a canonical ensemble and the conserved quantity here is the Helmholtz free energy, given by the equation,

$$A = U - TS \tag{A.2}$$

where U is the internal energy, T is the temperature, S is the entropy. In the differential form, this just becomes,

$$\begin{aligned} dA &= dU - SdT - TdS \\ dA &= TdS - PdV - SdT - TdS \\ dA &= -PdV - SdT \end{aligned}$$

It is the helmholtz free energy that will be minimized in a canonical ensemble while choosing the most probable microstate. This can be proved easily with the help of Clasius inequality [32], whose mathematical form simply states that $dS > \frac{\delta q}{T}$. By insertion of Clausius inequality into equation A.1, we can see that for a mechanically isolated system ($\delta W = 0$) at constant temperature,

$$dU - TdS \leq 0 \tag{A.3}$$

Using this inequality in the differential from of free energy, we get

$$dA = dU - TdS \leq 0 \tag{A.4}$$

This essentially states that the Helmholtz free energy, A, is simply a decreasing function in a canonical ensemble, that is when the temperature and the volume are held constant. Under equilibrium, the energy no longer changes, and the helmholtz function attains a constant value, which is simply the minimum possible value. Hence, to get the most preferred microstate, finding the minimum of the free energy would suffice.

All the simulation in the thesis are done with a thermostat, which means the temperature remains the same (or fluctuates about a mean). So, essentially the conformations of the polymers in our system are in canonical ensemble, owing to the free energy landscape of the simulations would essentially tell us the preferred conformations. The has been exploited effectively to get a general idea of the kinds of forces in play during the extrusion events.

A.2 Multiplicity and Entropy

A microstate is a specific configuration of the system that takes into account the position as well the momentum of all particles involved in describing the system, and a macrostate is the macroscopic property resulting from the culmination of all these coordinates, for instance pressure, temperature, volume. Now, multiplicity refers to all possible microstates of a particular

macrostate [33]. The internal energy and entropy can be defined using the states picture.

For a particular macrostate, the internal energy is the mean of the energy of that microstate multiplied by the probability of the occurrence of that microstate in a statistical ensemble. In mathematical terms, it is simply the mean over all microstates which can be written as follows,

$$U = \sum_{\mu} p_{\mu} H(\mu) \tag{A.5}$$

The absolute entropy is simply the logarithm of the multiplicity in a microcanonical ensemble. This can we written as,

$$S = k_B ln\Omega \tag{A.6}$$

where omega is the multiplicity and k_B is the boltzmann constant. In case of canonical ensemble, the microstates depends on the probability of the microstates.

$$S = -k_B \sum_{\mu}^{\Omega} p_{\mu} \ln p_{\mu} \tag{A.7}$$

A.3 Probability Distribution and Free Energy

The partition function can be defined as the sum over all possible microstates states of a system weighted by the Boltzmann factor. This is the normalization factor that appears in the probability distribution function for a given microstate in a canonical ensemble [34]. Mathematically, the partition function can be expressed as,

$$Z = \sum_{\mu} e^{-\frac{H(\mu)}{k_B T}} \tag{A.8}$$

where μ stands for a particular microstate, and the sum ranges to all possible microstates of the system.

To derive the relation between Helmholtz free energy and the probability distribution, we consider a mechanically isolated system in contact with a heatbath. This assures that the temperature of the system is always a constant and we are in a canonical ensemble. The probabilities for a given microstate μ at temperature T having generalized coordinates **x**, can be written in terms of the partition function as follows [34],

$$p(\mu) = \frac{e^{-\beta H(\mu)}}{Z} \tag{A.9}$$

where $\beta = \frac{1}{k_B T}$. This derivation is non-trivial, and has been methodically derived in the reference, so the details are left out. But it stands to state that the energy no longer has a definite value, rather it is a random variable. So, we can define a probability distribution associated with the energy random variable (E) by switching the random variable using the dirac-delta function. Mathematically, this is equivalent to writing [34],

$$p(E) = \sum_m up(u)\delta(H(\mu) - E) = \frac{e^{-\beta E}}{Z}\sum_\mu \delta(H(\mu) - E) \tag{A.10}$$

The sum is just the total number of microstates with energy E. This is simply the Ω we have come across in the previous section and can be written in terms of entropy as $\Omega = e^{\left(\frac{S}{k_B}\right)}$. Hence, the above equation can be simplified to get,

$$p(E) = \frac{\Omega(E)e^{-\beta E}}{Z} = \frac{1}{Z}exp\left[\frac{S(E)T}{k_B T} - \frac{E}{k_B T}\right] \tag{A.11}$$

The term in the square brackets is the negative of the Helmholtz free energy, as such, the equation can written as,

$$p(E) = \frac{1}{Z}e^{-\frac{A(E)}{k_B T}} \tag{A.12}$$

Taking natural logarithm on both sides and some rearrangement, we get,

$$A(E) = -k_B T \ln p(E) \tag{A.13}$$

which states that we can find the free energy of the system by calculating the probability distribution function. The minimum of this free energy landscape would give us the most preferred configuration, and in case of the simulation, this information is very crucial.

Appendix B

B.1 The Velocity Verlet

The equations concerning the governing the integration can be derived as follows,

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \frac{d\vec{x}}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{x}}{dt^2}\Delta t^2$$
$$\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{v}\Delta t + \frac{1}{2m}\vec{F}\Delta t^2$$

where we have done the Taylor expansion in the first line, and used the fact that $\vec{F} = m\frac{d^2\vec{x}}{dt^2}$ to get the final expression. For the velocity, we do the Taylor expansion again, and evaluate the final expression as follows,

$$\begin{aligned}
\vec{v}(t+\Delta t) &= \vec{v}(t) + \frac{d\vec{v}}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{v}}{dt^2}\Delta t^2 \\
&= \vec{v}(t) + \frac{\vec{F}}{m}\Delta t + \frac{1}{2m}\frac{d\vec{F}}{dt}\Delta t^2 \\
&= \vec{v}(t) + \frac{\vec{F}}{m}\Delta t + \frac{1}{2m}\left[\vec{F}(t+\Delta t) - \vec{F}(t)\right]\Delta t \\
\vec{v}(t+\Delta t) &= \vec{v}(t) + \frac{1}{2m}\left[\vec{F}(t+\Delta t) + \vec{F}(t)\right]\Delta t
\end{aligned}$$

These discretized equations govern the evolution of particle position and velocity over time.

B.2 The Noise Term

The random component of the force, or the noise term, in the Langevin equation has unique characteristics of its own. The changes in this force are very erratic, and don't retain in any memory. In mathematical construct, the time average average of the force equates to zero, $\langle F \rangle = 0$, and the auto-correlation function of the force goes as $\left\langle \vec{F}(t)\vec{F}(t') \right\rangle = q\delta(t-t')$ [12]. If

we assume a particle moving with initial velocity v_0, then the solution to the Langevin equation can be written as [12],

$$v(t) = v_0 e^{-\eta t} + \int_0^t ds e^{-\gamma(t-s)} \vec{F}_{noise}(s) \tag{B.1}$$

If we take the velocity squared and write the average, we get [12],

$$\left\langle v^2(t) \right\rangle = v_0^2 e^{-2\gamma t} + \frac{q}{2\gamma}\left[1 - e^{-2\gamma t}\right] \tag{B.2}$$

From the equipartition theorem, we know that if a particle is moving in a thermal bath at temperature T, then

$$\left\langle \frac{mv^2(t)}{2} \right\rangle = \frac{k_B T}{2} \tag{B.3}$$

From the above equation, we get the magnitude of the noise term, and the result is a special case of fluctuation dissipation theorem.

$$q = 2\gamma \frac{k_B T}{m} \tag{B.4}$$