AMERICAN UNIVERSITY OF BEIRUT

Kinetics of the Ejection of a Polymer from a Capsid: a Molecular Dynamics Approach

by
Rodrique Georges Badr

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Physics
of the Faculty of Arts and Sciences
at the American University of Beirut

Beirut, Lebanon
May 2018

AMERICAN UNIVERSITY OF BEIRUT

Kinetics of the Ejection of a Polymer from a Capsid: a Molecular Dynamics Approach

by

Rodrique Georges Badr

Approved by:

Dr. Leonid Klushin, Professor Physics	Advisor
Dr. Jihad Touma, Professor Physics	Member of Committee
Dr. Michel Kazan, Professor Physics	Member of Committee

Date of thesis defense: May 9, 2018

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name: Badr Rodrigue Georges
Last First Middle

☒ Master's Thesis ○ Master's Project ○ Doctoral Dissertation

☒ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes after: **One ___ year from the date of submission of my thesis, dissertation or project.**
Two ___ years from the date of submission of my thesis , dissertation or project.
Three ___ years from the date of submission of my thesis , dissertation or project.

R Badr May 16 2018
Signature Date

This form is signed when submitting the thesis, dissertation, or project to the University Libraries

Acknowledgements

My first acknowledgment goes to Dr. Leonid Klushin my thesis supervisor. His mentoring was invaluable for my growth and progress while working on this project. I would also like to thank him for his patience during the past year and a half.

In addition, I would like to thank my committee members Dr.'s Jihad Touma and Michel Kazan for reviewing my work and for their notes during my defense.

One big thank you is due to Joumana Abi Fallah. Her endless effort to make all the bureaucratic processes smooth makes everyone's lives easier. We also appreciate all the times she listened to our nagging. Thank you Joumana for being the mother in our department.

I would like to thank Dr. Ghassan Antar for the opportunity he presented me as an undergraduate to do research. This opportunity is one of the main reasons I have the skills I needed to tackle this project and properly think about the problem.

Special thanks go to Lama Tannoury for her priceless support, I can honestly say that finishing this would have been ten times harder without her. In our five years of friendship she was the reason I was able to survive all the pressure and mishaps of university and daily life. So thank you Lama for being an amazing friend.

Finally I would like to thank my friends for their support and precious discussion. I want to thank Antranik Sefilian for always pushing me forward and encouraging me to do my best and challenge myself. Wassim Sleiman, Mariam Khaldieh, Hana Baroudi, Kafa Alameh, Ali Khalife, and Jaad Tannous for all the fun times that delayed the completion of this work, and the comments and criticisms that made the work and myself better.

An Abstract of the Thesis of

Rodrique Georges Badr for Master of Science
Major: Physics

Title: Kinetics of the Ejection of a Polymer from a Capsid: a Molecular Dynamics Approach

In this thesis we investigate the problem of a polymer chain ejecting from a cylindrically symmetric capsid. We approach the problem through molecular dynamics simulations using a bead spring model for the chain, and a strongly repulsive potential for the walls of the capsid. We formulate two theories to get predictions for the ejection times, one deterministic and one stochastic. We also introduce a method to calculate local pressure distribution in cylindrical coordinates. The simulation results show some deviations from the theory which is an indication that we need to improve the theoretical model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We have witnessed in the past few decades a rapid advancement in applications of physical concepts in medical sciences and biology, driven by the better understanding of soft matter. Soft matter is found in all living organisms, and takes different forms: membranes, proteins, DNA strands etc... This makes the concepts developed in soft matter physics readily applicable to biological systems by making simple, yet effective, analogies between the system of interest and one of the models thoroughly studied in soft matter.

The dynamics of confined polymers has gained a lot of importance in several fields of science, especially since confined polymers appear in numerous places (the nucleus of a cell, DNA in a viral capsid...). Recently a lot of effort has been put in experimental, theoretical, and numerical methods to better understand the mechanisms and scales of the storage and ejection of DNA from the capsids of bacteriophages. What sparked the interest in the ejection of polymers from capsids was the experiment done by Hershey and Chase in 1952 [3]. The experiment is mostly known for the confirmation of DNA as the main genetic material.

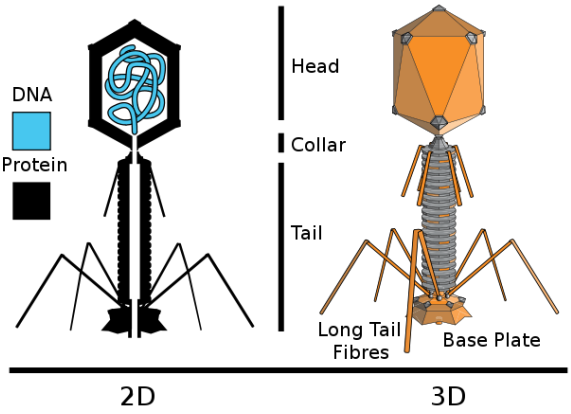

Figure 1.1: Structure of a Bacteriophage [1].

However, the experiment also showed that bacteriophages passively inject the host cell with their DNA while leaving the capsid outside of the host, which left scientists with various questions concerning the process of the ejection, the time scales involved, along with the effects of different geometries on those time scales.

The structure of bacteriophages can be simply described by a head in which the genetic material (DNA) is stored, a long tail, and a neck which connects the tail to the head, as shown in Figure 1.1. Inside the head, which will henceforth be referred to as the chamber, the DNA is so densely packed that pressures can reach values up to 6 MPa [4]. It is widely believed that this high pressure is the main driving force for the ejection process.

Experimental work on the ejection of DNA from phages generally shows that the process is divided into multiple stages which are separated by dynamic pauses [4, 5]. In addition, a study concerning the effect of temperature on the ejection by de Frutos et al. [5] shows that at low temperatures the ejection is slowed down dramatically. However, the curves describing the length of DNA ejected versus

time look parallel in a log-lin plot, which suggests a law

$$L \propto ln(t \times f(T)) \tag{1.1}$$

where L is the length ejected, t is time, and $f(T)$ is a function of temperature.

On the theoretical level, Inamdar et al. [6] were able to make analytical progress and get an expression of the time needed for the ejection of a polymer from a capsid by adopting the approach of diffusion in an external field. The prediction is in the form of a formula for the mean first-passage time that follows from the Fokker-Planck equation. We use a similar method described in section 2.6. Milchev et al. [7] studied the ejection of a polymer from a nanopore. The ejection from a nanopore is similar to the final stage of the ejection of DNA from a bacteriophage (when no part of the DNA is left in the chamber, part of it is in the tail, and the rest is outside). They were able to get a scaling law for the ejection time that can be evaluated directly to get prediction.

This topic has also been extensively studied numerically. Milchev et al. [7] accompanied their analytical analysis with Monte Carlo (MC) simulations of a polymer ejecting from a pore. Their results showed good agreement with the predicted behavior and scaling laws. Other numerical experiments were conducted by Ali et al. [8, 9]. These experiments were Molecular Dynamics (MD) simulations of a polymer ejecting from a spherical capsid with a hole on its top. Results have shown that if the data is not averaged over different simulations, pauses similar to those observed experimentally [4, 5] appear, which were explained as pauses needed for the polymer to rearrange itself inside the capsid in

order to be able to continue the ejection. Also, Brownian Dynamics simulations of again a polymer being ejected from a spherical capsid were done by Kindt et al. [10]. This work focused on the variation of the force resisting the loading of an extra monomer into the capsid for polymers that interact through an attractive-repulsive potential and through a purely repulsive potential.

Inspired by the work done on the subject, in this thesis we investigate the problem of a polymer in a cylindrical capsid through molecular dynamics simulations, studying the equilibrium properties and the ejection kinetics.

Chapter 2

Theoretical Background

In nature, atoms bind together to form molecules, from simple molecules like hydrogen H_2 and oxygen O_2, to more complicated molecules like the proteins in our bodies. When molecules of the same nature start binding together they make polymers. A polymer is a defined as a sequence of elementary repeating units called **monomers**. The monomers themselves are individual molecules, and they are linked together in a chain like fashion, so that we sometimes use the term **chain** to refer to polymers. The different models and tools used to study polymer physics are outlined below, starting from the basic definitions, up to more involved dynamics.

This chapter relies mostly on the works authored by M. Rubinstein and R. Colby [11] for the sections on the different types of chains and scaling laws (see also the book by P. G. de Gennes [12]), an article by T. Nakamura et al. [13] for microscopic pressure, an article by Milchev et al. [7] for the section on the ejection from a nanopore, and C. W. Gardiner [14] for the section on stochastic dynamics.

2.1 Ideal Chains

If we label the positions of the monomers of a polymer by $\boldsymbol{r_i}$, where i goes from $1 \rightarrow N$ and label the monomers from tip to end, an ideal chain is defined as a chain where the interaction between monomers vanishes for $|i-j| >> 1$. In other words, monomers that are far apart along the chain do no interact with each other, which leads to the overlap of monomers in certain places. Interactions in an ideal chain have to do with the flexibility of the chain, and several models exist that we will not be going over here. For reference see [11].

Due to the nature of interactions in ideal chains, on large enough length scales the chain conformations can be mapped onto random walks of N steps, step size a, and end-to-end vector $\boldsymbol{R}$. Using random walk statistics, we find that the average end-to-end length of an ideal chain is $\langle R_0 \rangle = aN^{1/2}$. We can also find an expression for the Helmholtz free energy $F = E - TS$ of an ideal chain, where E is the internal energy of the chain, T is the temperature, and S is the entropy. It is useful to note here that due to the range of interactions in an ideal chain, the internal energy of the chain does not depend on the end-to-end vector. To begin, the probability distribution of end-to-end vectors for a 3 dimensional random walk can be written as

$$P_{3d}(N, \boldsymbol{R}) = (\frac{3}{2\pi N a^2})^{3/2} exp(-\frac{3\boldsymbol{R}^2}{2Na^2})$$

and in terms of the end-to-end distance $R \equiv |\boldsymbol{R}|$

$$P_{3d}(N, R) 4\pi R^2 dR = 4\pi (\frac{3}{2\pi N a^2})^{3/2} exp(-\frac{3R^2}{2Na^2}) R^2 dR$$

Using the above distribution we can evaluate the entropy $S(N, \boldsymbol{R}) = kln(\Omega(N, \boldsymbol{R}))$ where Ω is the number of conformations of chain of N monomers and end-to-end vector $\boldsymbol{R}$. The distribution of end-to-end vectors is related to Ω through

$$P_{3d}(N, \boldsymbol{R}) = \frac{\Omega(N, \boldsymbol{R})}{\int \Omega(N, \boldsymbol{R}) d\boldsymbol{R}}$$

From this relation the entropy can be found to be

$$S(N, \boldsymbol{R}) = -\frac{3}{2} k \frac{\boldsymbol{R}^2}{Na^2} + S(N, 0)$$

where the second term does not depend on $\boldsymbol{R}$. Finally, the free energy is

$$F(N, \boldsymbol{R}) = \frac{3}{2} kT \frac{\boldsymbol{R}^2}{Na^2} + F(N, 0)$$

where again the second term does not depend on $\boldsymbol{R}$.

2.2 Real Chains

To make the model more realistic, we need to account for the possibility of interaction between any monomers. This is achieved through the introduction of the excluded volume v, which is the volume that a monomer occupies such that two monomers cannot overlap. Polymers however are typically dissolved in solvents, and depending on the type of solvent the interaction between monomers will be replaced by an effective interaction, resulting in an effective excluded volume. The effective interactions can be modeled as a potential that is repulsive at small distances and attractive at larger distance, the most common example being the Lennard-Jones potential shown in figure 2.1. When the repulsive (attractive)

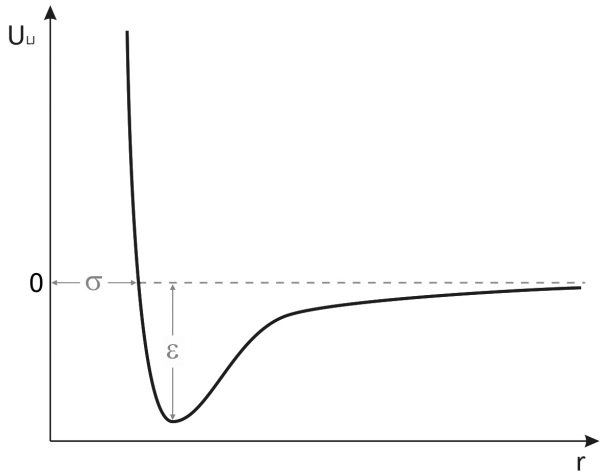

Figure 2.1: Lennard-Jones Potential.

part dominates, the excluded volume will be positive (negative).

2.2.1 Types of Solvents

In a polymer solution, we can differentiate between two species, the polymer and the solvent. Three types of interactions can be distinguished:

- Monomer-Monomer interactions characterized by the mean interaction energy u_{MM}
- Solvent-Solvent interactions characterized by the mean interaction energy u_{SS}
- Monomer-Solvent interactions characterized by the mean interaction energy u_{MS}

A dimensionless quantity that characterizes the difference in interaction energies in a solution is the **Flory interaction parameter**

$$\chi = \frac{z}{2}\frac{(2u_{MS} - u_{MM} - u_{SS})}{kT} \tag{2.1}$$

And using this the effective excluded volume can be defined

$$v = (1 - 2\chi)v_0$$

where v_0 is the excluded volume from hard-core repulsion. From this definition we can distinguish between the following types of solvents:

- Athermal solvent: for high temperatures where $\chi \to 0$ the value of the excluded volume approaches v_0. At this limit the excluded volume is a very slowly varying function of temperature, hence the name athermal solvent. The effective interaction here can be modeled as purely repulsive.

- Good Solvent: for intermediate temperatures $0 < \chi < 1/2$ and $0 < v < v_0$. In this regime the the monomers feel slight attraction, which reduces the excluded volume to values below the hard-core volume. The effective interaction here can be modeled as purely repulsive, but with less stiffness than that of an athermal solvent.

- Theta solvent: for $\chi = 1/2$ the excluded volume is exactly equal to zero, which is the case for ideal chains, and the polymer will obey ideal chain statistics. In this case there is no interaction between monomers far away from each other along the chain.

- Poor solvent: this is the regime for low temperatures where $\chi > 1/2$, resulting in a negative excluded volume and the collapse of the polymer into a globule. Here the interaction needs to have both attraction and repulsion.

2.2.2 Real Chain in a Good Solvent

In a good solvent a real chain tends to swell to a size $R > R_{ideal}$ due to the repulsive interactions between the monomers. The equilibrium size of a real chain depends on the interplay between entropy and interaction. A model that captures this is the **Flory theory** for a polymer in a good solvent. Flory assumed that the entropic contribution is the same as for an ideal chain

$$F_{ent} \approx kT\frac{R^2}{Na^2}$$

up to numerical factors. For the interaction part, we imagine the monomers to be uniformly distributed in a box of volume R^3, which is equal to the **pervaded volume** of the chain . The volume per monomer is R^3/N, and if the excluded volume is v, the probability of two monomers overlapping is $\frac{v}{R^3/N}$ and the number of overlapping monomers is $\frac{vN^2}{R^3}$. The energy of excluded volume interactions is on the order kT up to numerical factors of order unity, which results in an estimate for the interaction free energy

$$F_{int} = kTv\frac{N^2}{R^3}$$

and finally the total free energy of a real chain in a good solvent is

$$F = kT\left(\frac{R^2}{Na^2} + v\frac{N^2}{R^3}\right) \tag{2.2}$$

Minimizing this free energy with respect to R gives the equilibrium size of a free chain

$$R_F = v^{1/5}a^{2/5}N^{3/5} \tag{2.3}$$

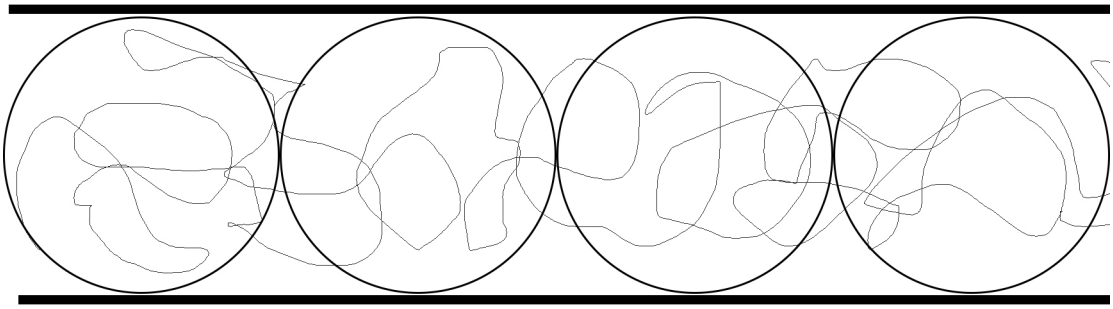

Figure 2.2: A chain in a narrow tube. The blob size is set by the diameter of the tube. On scales smaller than the blob size the chain behaves like a free chain.

2.3 Scaling Approaches

When dealing with confined chains it is possible to get an idea about free energies using simple scaling arguments. For scaling, it is important to introduce the concept of a **blob**. A blob defines a length scale below which the chain conformation is unaffected by external effects. For a real chain, on scales below the blob size the size of the chain obeys eq. 2.3, while on larger scales the size depends on the type of confinement as described below. As a chain grows from sizes smaller than the blob sizes to larger sizes and starts feeling the effect of confinement, the difference in entropy starts to be significant. A significant change in entropy is on the order $\Delta S/k \approx 1$, resulting in a change in free energy $\Delta F \approx kT$; this makes the free energy stored in each blob on the order of kT.

2.3.1 Real Chain in a Narrow Tube

We consider a real chain confined in an infinitely long rigid cylindrical tube of diameter $a << d << R_F$. The effect of confinement is felt on length scales comparable to the diameter of the tube, which sets the blob size to d (see figure 2.2). We assume our chain has N monomers and monomer length a, and in a

good solvent $v \approx a^3$ so that using using eq. 2.3 $R_F = aN^{3/5}$. Let g be the number of monomers in an individual blob and $N_b = N/g$ be the number of blobs, on scales smaller than d the chain behaves like a free chain and we have

$$d = ag^{3/5} \Rightarrow g = \left(\frac{d}{a}\right)^{5/3}$$

and the number of blobs is

$$N_b = N\left(\frac{a}{d}\right)^{5/3}$$

Since each blob carries energy on the order kT, the free energy and equilibrium size of a real chain in a tube are respectively

$$F = A_F N_b kT = A_F kTN\left(\frac{a}{d}\right)^{5/3} \tag{2.4}$$

$$R_{||} = N_b d = A_e aN\left(\frac{a}{d}\right)^{2/3} \tag{2.5}$$

where A_F and A_e are dimensionless model dependent prefactors. We can also write the free energy as

$$F = BkT\frac{R_{||}}{d} \tag{2.6}$$

where B here is a dimensionless model independent prefactor. B depends however on the geometry of the confinement, and was shown to have values close to 5 for cylindrical geometries [15].

2.3.2 Real Chain in a Large Chamber: Semi-Dilute Solution

The free energy of a polymer solution depends on the concentration/volume fraction ϕ. The volume fraction of a single real chain in a good solvent is the ratio of the volume of the monomers Na^3 to the pervaded volume R_F^3

$$\phi^* = \frac{Na^3}{R_F^3} \tag{2.7}$$

$$\phi^* = N^{-4/5} \tag{2.8}$$

where the last equally is obtained after replacing R_F with eq. 2.3 and $v = a^3$. ϕ^* is called the overlap concentration. When the concentration of a polymer solution is less than ϕ^* the solution is said to be dilute and the polymers do not overlap. For a dilute solution of polymers, the osmotic pressure follows the van't Hoff law

$$\Pi = \frac{kT}{b^3}\frac{\phi}{N}$$

As the concentration approaches ϕ^* the polymers start to overlap and the osmotic pressure will become a stronger function of the concentration

$$\Pi = \frac{kT}{b^3}\frac{\phi}{N} f\left(\frac{\phi}{\phi^*}\right)$$

where $f\left(\frac{\phi}{\phi^*}\right) \approx 1$ for $\phi < \phi^*$. For semi-dilute solutions we have $\phi > \phi^*$ and we assume that f is a power law in the concentration

$$f\left(\frac{\phi}{\phi^*}\right) \approx \left(\frac{\phi}{\phi^*}\right)^z$$

which results in an expression for the osmotic pressure

$$\Pi \approx \frac{kT}{a^3}\phi^{1+z}N^{4z/5-1} \tag{2.9}$$

The osmotic pressure in semi-dilute solutions must not depend on the number of monomers in each chain, so we can find the value of z by setting the power of N in eq. 2.9 to zero, we find

$$z = 5/4$$

and replacing in eq. 2.9

$$\Pi \approx \frac{kT}{a^3}\phi^{9/4} \tag{2.10}$$

Using this, we get the free energy per volume of a semi-dilute solution

$$\frac{F}{V} \approx \frac{kT}{a^3}\phi^{9/4} \tag{2.11}$$

For our problem, we have a chain with N monomers confined in a cylindrical chamber of volume V so that the concentration is $\phi = \frac{Na^3}{V}$ and the free energy per volume is

$$\frac{F}{V} = A_\mu kT(a^3)^{5/4}\left(\frac{N}{V}\right)^{9/4} \tag{2.12}$$

where A_μ is again a dimensionless numerical prefactor that is model dependent.

2.4 Microscopic Pressure

In order to characterize the equilibrated initial conditions of the polymer in the capsid, we looked at pressure distribution on the microscopic level. For a system of N particles with coordinates $\boldsymbol{r}_i$ and momenta $\boldsymbol{p}_i$ the microscopic definition of pressure is

$$\frac{\partial}{\partial t}\left[\sum_{i=1}^{N} \boldsymbol{p}_i \delta(\boldsymbol{x}-\boldsymbol{r}_i)\right] = -\nabla.\boldsymbol{P}(\boldsymbol{x}) \tag{2.13}$$

If the particles interact through a potential that has translational invariance, such that $\sum_{i=1}^{N} \nabla_i U(\boldsymbol{r}_i) = 0$, we have

$$\boldsymbol{P}(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i \otimes \boldsymbol{p}_i}{m_i}\delta(\boldsymbol{x}-\boldsymbol{r}_i) - \sum_{i=1}^{N} \nabla_i U(\boldsymbol{r}_i) \otimes \int_{C_{0i}} d\boldsymbol{l}\,\delta(\boldsymbol{x}-\boldsymbol{l})$$

where C_{0i} is a contour that connects $\boldsymbol{r}_i$ to an arbitrary point $\boldsymbol{r}_0$, $\boldsymbol{l}$ is a length element along C_{0i}, and $\otimes$ denotes a dyadic product. The first term includes the kinetic (ideal gas) contribution, while the second term includes the contribution from interactions. In our work we only deal with two body interactions so that the pressure can be written as

$$\boldsymbol{P}(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i \otimes \boldsymbol{p}_i}{m_i}\delta(\boldsymbol{x}-\boldsymbol{r}_i) + \sum_{i=1}^{N-1}\sum_{j>i} \boldsymbol{F}(\boldsymbol{r_{ij}}) \otimes \int_{C_{ji}} d\boldsymbol{l}\,\delta(\boldsymbol{x}-\boldsymbol{l}) \tag{2.14}$$

where now C_{ji} connects $\boldsymbol{r}_i$ to $\boldsymbol{r}_j$, $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$, and the intermolecular force $\boldsymbol{F}(\boldsymbol{r_{ij}})$ is radial and directed along the line joining the particles; it is assumed to have the form $\boldsymbol{F}(\boldsymbol{r_{ij}}) = f(\boldsymbol{r_{ij}})\boldsymbol{r_{ij}}$, and is the force acting on particle i due to its interaction with particle j. Henceforth, the individual particle contribution to the interaction term in the pressure will be referred to as $\boldsymbol{P}_{ij}(\boldsymbol{x})$, and its individual

components are

$$P_{ij}^{\alpha\beta}(\boldsymbol{x}) = \int_{C_{ji}} (d\boldsymbol{l}.\boldsymbol{e}_\alpha)(\boldsymbol{F}.\boldsymbol{e}_{\boldsymbol{\beta}})\delta(\boldsymbol{x}-\boldsymbol{l})$$

where $\alpha/\beta = x, \rho, \phi$ in cylindrical coordinates, and the force term was absorbed into the integral since it does not depend on $\boldsymbol{l}$. The unit vectors $\boldsymbol{e_\alpha}$ are

$$\begin{aligned}
\boldsymbol{e}_x &= \hat{x} \\
\boldsymbol{e}_\rho &= \frac{\boldsymbol{\rho}}{|\boldsymbol{\rho}|} \\
\boldsymbol{e}_\phi &= \boldsymbol{e}_x \times \boldsymbol{e}_\rho
\end{aligned}$$

In order to compute the pressure tensor in our simulations, we average equation 2.14 over finite volume elements

$$P_{ij}^{\alpha\beta}(\mathcal{V}) = \frac{1}{|\mathcal{V}|}\int_{\mathcal{V}} d\boldsymbol{x} P_{ij}^{\alpha\beta}(\boldsymbol{x})$$

where $\mathcal{V}$ refers to a specific slice and $|\mathcal{V}|$ is the volume of the slice. The full expressions in cylindrical coordinates that we used in the calculations are listed in appendix B. For the expressions in other coordinate systems and their relevant applications see [13, 16, 17].

2.5 Deterministic Ejection Kinetics

Using some simple arguments, we can write equations of motion for the number of particles remaining in a section of the capsid. The ejection from the full capsid can be divided into two stages. The first stage consist of the emptying of the chamber. The second stage is when the chamber is empty and we have ejection

from the tube.

2.5.1 The Initial Conditions

Before the ejection starts, the chain is in an equilibrium state defined by the equality of chemical potentials between the chamber and the tube. From eqs. 2.4 and 2.12 we can find the chemical potentials in the tube and the chamber respectively

$$\mu_{tube} = A_F kT \left(\frac{a}{d}\right)^{5/3} \tag{2.15}$$

$$\mu_{ch} = A_\mu kT (a^3)^{5/4} \left(\frac{N}{V}\right)^{5/4} \tag{2.16}$$

We can see that the chemical potential in the tube is fully determined by the diameter of the tube. This means that the equilibrium chemical potential is determined by the diameter of the tube, and this gives us the equilibrium number of monomers in the chamber for fixed chamber volume V, or in other words, the diameter of the tube determines the equilibrium density in the chamber. We distinguish two cases: if the equilibrium number of monomers in the chamber is larger than the total number of monomers, the chain will fully reside in the chamber and we will have an empty tube. If the equilibrium number is less than the total number of monomers, part of the chain will reside in the chamber and part of it will be in the tube. The latter case is what we are interested in and is illustrated in figure 2.3.

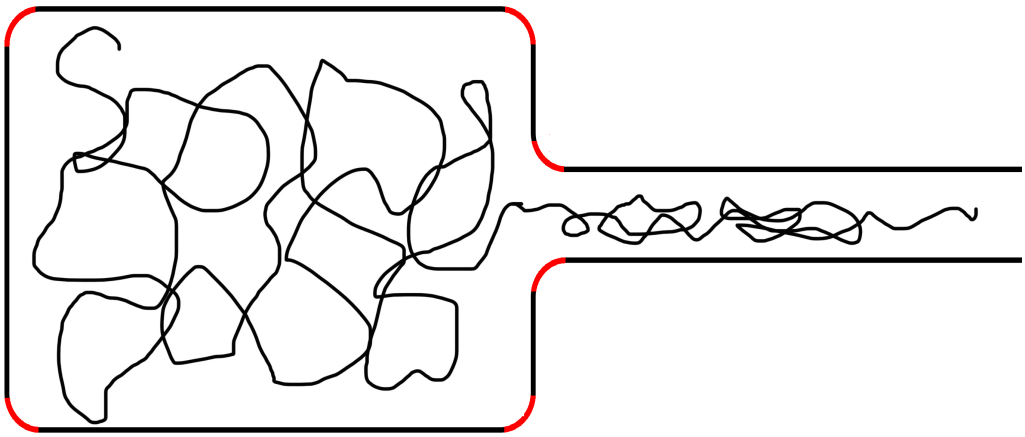

Figure 2.3: Equilibrium configuration of a long chain in a capsid.

2.5.2 Emptying the Chamber

During the first stage we assume that the number of monomers in the tube is constant, and the ejection is driven by the difference in chemical potential between the chamber and the outside. In order to write an equation of motion, we equate the work done by drag forces to the change in free energy, specifically the rate of change of the free energy is equal to the power dissipated by friction

$$\frac{\partial F}{\partial t} = P_{diss}$$

The total free energy is a sum of three contributions from the chamber, the tube, and the outside, and the same applies for the power dissipated

$$F = F_{ch} + F_{tube} + F_{out}$$

$$P_{diss} = -\zeta(n_{ch}v_{ch}^2 + n_{tube}v_{tube}^2 + n_{out}v_{out}^2)$$

where n_i and v_i are the number of monomers and drift velocity in the correspond-

ing region respectively, and ζ is the drag coefficient. Since the outside region is infinitely large the free energy of the portion of the chain outside the chamber is that of a free chain, and is taken as the reference for our free energies. Since the number of monomers in the tube is constant we take $F_{tube} = cst$. In addition, we assume that the drift velocities in the chamber and the outside region are very small compared to the drift velocity in the tube; this means that the tube's contribution to the dissipated power dominates the contributions from the chamber and the outside region, and the latter are neglected. Having all this we can write

$$\frac{\partial F_{ch}}{\partial t} = -n_{tube}\zeta v_{tube}^2$$

the drift velocity in the tube can be written $v_{tube} = \lambda \dot{n}_{ch}$, where λ is the length per monomer in the tube and the dot denotes a time derivative. We can also use the chain rule to write $\frac{\partial F}{\partial t} = \dot{n}_{ch}\frac{\partial F}{\partial n_{ch}}$, where

$$\frac{\partial F}{\partial n_{ch}} = \mu_{ch}(n_{ch}) = A_\mu kT\left(\frac{n_{ch}a^3}{V}\right)^{5/4}$$

is the chemical potential in the chamber and is obtained from eq. 2.12. Finally, combining all of the above, we have an evolution equation for the number of monomers in the chamber

$$\dot{n}_{ch} = -A_\mu \frac{kT}{n_{tube}\zeta\lambda^2}\left(\frac{n_{ch}a^3}{V}\right)^{5/4} = -\frac{\mu_{ch}}{n_{tube}\zeta\lambda^2} \tag{2.17}$$

This equation can be integrated to give the time t_1 needed to empty the chamber (with $kT = 1$ and $a^3 = 1$)

$$t_1 = 4\frac{\zeta\lambda^2 V^{5/4} n_{tube}}{A_\mu}\left[(n_{ch}^f)^{-1/4} - (n_{ch}^i)^{-1/4}\right] \tag{2.18}$$

which results in infinite time for ejection if we choose the final number of monomers in the chamber n_{ch}^{f} to be zero. This is an indicator that the current approach is not sufficient, and the role of diffusion needs to be considered, which is described in section 2.6. Eq. 2.18 can be used however to get an overestimate of the time needed to empty the chamber if we choose $n_{ch}^{f} \neq 0$.

Looking at eq. 2.5, we can define λ as

$$\lambda = R_{||}/N = A_e a \left(\frac{a}{d}\right)^{2/3}$$

so that λ is fully determined by the diameter of the tube. This also allows us to write $n_{tube} = l/\lambda$ where l is the length of the tube. And replacing these expressions in eq. 2.18 we find $t_1 \propto lV^{5/4}d^{-2/3}$

2.5.3 Ejection from the Tube

If we start with a chain long enough so that it has a tail outside the tube, it will feel a force that is entropic in origin and independent of the size of the tail [18].

We have seen in section 2.3 that the free energy of a polymer confined in a tube can be written as

$$F = B\frac{kT}{D}R_{||}$$

For a chain in an infinite tube, the dimensionless prefactor $B \approx 5$. It was show however in [15] that for an open tube, the chain experiences some stretching which results in the value $\tilde{B} \approx 3$.

During tube ejection, the chain occupies a length x inside the tube, and we have

$$F = \tilde{B}\frac{kT}{D}x \tag{2.19}$$

and the average (entropic) force acting on the chain is

$$f = \tilde{B}\frac{kT}{D}. \tag{2.20}$$

Since the chain is in the overdamped regime, we can balance the entropic force with the drag force

$$n(t)\zeta\frac{dx}{dt} = -f \tag{2.21}$$

where $v_d = \frac{dx}{dt}$ is the drift velocity in the tube, ζ is the drag coefficient, and $n(t)$ is the number of monomers in the tube. Taking λ as the length per monomer in the tube, we have $x = \lambda n(t)$. Assuming λ is independent of time, we can change the force balance equation to an equation for the number of monomers in the tube

$$n(t)\frac{dn}{dt} = -\frac{\tilde{B}}{\zeta\lambda}\frac{kT}{D} \tag{2.22}$$

Integrating this equation gives the time $t2$ needed for the polymer to leave the tube completely (with $kT = 1$)

$$t_2 = \frac{\zeta\lambda D}{2\tilde{B}}\left[n_i^2 - n_f^2\right] \tag{2.23}$$

2.6 Stochastic Ejection

As discussed above, the deterministic equations that were derived result in an infinite time of ejection from the chamber. To overcome this the effects of diffusion

need to be included in the analysis, which is done through the Fokker-Planck Equation (FPE). Then using the FPE we can calculate a mean **First Passage Time** which should give the mean time for ejection from the capsid.

2.6.1 The Fokker-Planck Equation

The FPE is an equation for the evolution of conditional probabilities $P(x,t|x_0,t_0)$ that a variable $\boldsymbol{X}$ has value x at time t, given that it had value x_0 at time t_0. We can distinguish between the forward (eq. 2.24) and backward (eq. 2.25) FPE which in one dimension are

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t} = -\frac{\partial}{\partial x}[\mathcal{A}(x,t)P(x,t|x_0,t_0)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\mathcal{D}(x,t)P(x,t|x_0,t_0)] \quad (2.24)$$

$$\frac{\partial P(x,t|x_0,t_0)}{\partial t_0} = -\mathcal{A}(x_0,t_0)\frac{\partial}{\partial x_0}[P(x,t|x_0,t_0)] - \frac{1}{2}\mathcal{D}(x_0,t_0)\frac{\partial^2}{\partial x_0^2}[P(x,t|x_0,t_0)] \quad (2.25)$$

Where $\mathcal{A}$ is the drift coefficient, and $\mathcal{D}$ is the diffusion coefficient. If the drift and diffusion coefficients are independent of time the process is said to be **homogeneous**. The forward equation gives the probability of $\boldsymbol{X}$ having value x at time t given that we started at x_0 at time t_0. The backward equation gives the probability of the value being x_0 at time t_0 given that the final value is at x at time t. In order to solve the forward (backward) equation we need initial (final) conditions in addition to boundary conditions. We can distinguish several type of boundary conditions; however, for reasons that will be apparent later we will only mention two types of boundary conditions for the backward equation (for a comprehensive list of boundary conditions for both equations see [14]):

- Reflecting boundary: if $x = a$ is a reflecting boundary, so that it cannot be crossed, it means that the flow of probability through the boundary must vanish $\left.\frac{\partial P(x,t|x_0,t_0)}{\partial x_0}\right|_{x_0=a} = 0$

- Absorbing boundary: if $x = a$ is an absorbing boundary, the probability of being there must vanish $P(x, t|a, t_0) = 0$

Commonly x is the position of a certain particle; in our work however, we look at the probability distribution for the number of monomers in a certain region, so that for example the conditional probability $P(n_{ch}, t|n_0, 0)$ gives the probability of having n_{ch} monomers in the chamber at a time t if we initially had n_0 monomers in it.

2.6.2 First Passage Time for a Homogeneous Process

We are interested in knowing the mean time needed for a particle that starts at a position $a < x < b$ to leave the region bounded by a and b. This time is called the first passage time $T(x)$. Using the definition of conditional probabilities and the backward FPE, we can arrive at a differential equation for $T(x)$ (see [14])

$$\mathcal{A}(x)\frac{\partial T}{\partial x} + \frac{1}{2}\mathcal{D}(x)\frac{\partial^2 T}{\partial x^2} = -1$$

Subject to the boundary conditions at a and b. If we choose one boundary to be absorbing and the other reflecting, the solution of the equation is

$$T(x) = 2\int_x^b \frac{dy}{\psi(y)} \int_a^y \frac{\psi(z)}{\mathcal{D}(z)} dz \qquad \begin{array}{r} a \text{ reflecting} \\ b \text{ absorbing} \\ a < b \end{array} \tag{2.26}$$

$$T(x) = 2\int_a^x \frac{dy}{\psi(y)} \int_y^b \frac{\psi(z)}{\mathcal{D}(z)} dz \qquad \begin{array}{r} a \text{ absorbing} \\ b \text{ reflecting} \\ a < b \end{array} \tag{2.27}$$

$$\psi(x) = exp\left(\int_a^x \frac{2\mathcal{A}(x')}{\mathcal{D}(x')} dx' \right) \tag{2.28}$$

2.6.3 First Passage Time for our Problem

To get predictions for our problem make the analogy between the position x of a particle diffusing in one dimension and the number of monomers in the capsid N. Our stochastic variable N can have values between the initial number of monomers inside the capsid N_0 and 0 monomers inside the capsid. We choose an absorbing at 0 and a reflecting boundary at N_0.

We need to find the equivalents of drift and diffusion for our evolving variable N. We again separate the ejection into two stages. When we still have monomers in the chamber we see from looking at eq. 2.17 that the drift can be written as

$$\mathcal{A}_{ch} = -\frac{\mu_{ch}}{n_{tube}\lambda^2\zeta}$$

where n_{tube} and λ are assumed to be constant. The chemical potential can be thought of as the equivalent of a force that drives the change in the number of monomers. Now since drift is the force multiplied by mobility, the equivalent of mobility is $M = \frac{1}{n_{tube}\lambda^2\zeta}$, and then the diffusion coefficient is

$$\mathcal{D}_{ch} = \frac{kT}{n_{tube}\lambda^2\zeta}$$

When the chamber is empty and we only have monomers in the tube, eq. 2.17 is still valid but with μ_{ch} replaced with μ_{tube}, or by looking at eq. 2.22 the drift during this time is

$$\mathcal{A}_{tube} = -\frac{\mu_{tube}}{N\lambda^2\zeta} = -\frac{\tilde{B}kT}{N\lambda\zeta d}$$

where the last equality can be verified by replacing the expression for μ_{tube} and after a little manipulation. Finally the diffusion coefficient is again

$$\mathcal{D}_{tube} = \frac{kT}{N\lambda^2\zeta}$$

Using those expressions we can define the diffusion and drift coefficients for the whole process

$$\mathcal{D}(N) = \begin{cases} \frac{kT}{\lambda^2\zeta}\frac{1}{N} & \text{for } 0 \leq N \leq n_{tube} \\ \\ \frac{kT}{\lambda^2\zeta}\frac{1}{n_{tube}} & \text{for } n_{tube} \leq N \leq N_0 \end{cases} \tag{2.29}$$

$$\mathcal{A}(N) = \begin{cases} -\frac{\tilde{B}kT}{\lambda\zeta D}\frac{1}{N} & \text{for } 0 \leq N \leq n_{tube} \\ \\ -\frac{A_\mu kT}{\lambda^2\zeta n_{tube}}\left(\frac{N-n_{tube}}{V}\right)^{5/4} & \text{for } n_{tube} < N \leq N_0 \end{cases} \tag{2.30}$$

Now $\psi(N)$ can be calculated

$$\psi(N) = \begin{cases} exp\left(-\frac{2\tilde{B}\lambda}{D}N\right) & \text{for } 0 \leq N \leq n_{tube} \\ \\ exp\left(-\frac{2\tilde{B}\lambda}{D}n_{tube} - \frac{8}{9}\frac{A_\mu}{V^{5/4}}(N-n_{tube})^{9/4}\right) & \text{for } n_{tube} \leq N \leq N_0 \end{cases} \tag{2.31}$$

and finally we get the mean first passage time using eq. 2.28

$$T(N_0) = 2\int_0^{n_{tube}} \frac{G(y)}{\psi(y)}dy + 2\int_{n_{tube}}^{N_0} \frac{F(y)}{\psi(y)}dy \tag{2.32}$$

$$G(y) = \int_y^{N_{max}} \frac{\psi(z)}{\mathcal{D}(z)}dz \qquad \text{for } 0 \leq y \leq n_{tube} \tag{2.33}$$

$$F(y) = \int_y^{N_{max}} \frac{\psi(z)}{\mathcal{D}(z)}dz \qquad \text{for } n_{tube} \leq y \leq N_0 \tag{2.34}$$

After replacing the expressions for $\psi(z)$ and $\mathcal{D}(z)$ in the equation for G the integral splits into two integrals ,G_1 for $z < n_{tube}$ and G_2 for $z > n_{tube}$. G_2 is

similar to the integral for F; after making a change of variable $t = \frac{8}{9}\frac{A_\mu}{V^{5/4}}(z - n_{tube}^{9/4})$, and some manipulations, both can be evaluated in terms of the Gamma function and the incomplete Gamma functions defined as

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$$

$$\Gamma_l(a,x) = \int_0^x t^{a-1}e^{-t}dt \qquad \text{is the lower incomplete Gamma function}$$

$$\Gamma_u(a,x) = \int_x^\infty t^{a-1}e^{-t}dt \qquad \text{is the upper incomplete Gamma function}$$

The result is

$$G(y) = \frac{\zeta D^2}{4\tilde{B}^2 kT}\left[\left(2\frac{B\lambda}{D}y + 1\right)e^{-2\frac{\tilde{B}\lambda}{D}y} - \left(2\frac{\tilde{B}\lambda}{D}n_{tube} + 1\right)e^{-2\frac{\tilde{B}\lambda}{D}n_{tube}}\right] \tag{2.35}$$
$$+ \frac{4}{9}n_{tube}\frac{\lambda^2\zeta}{kT}e^{-2\frac{\tilde{B}\lambda}{D}n_{tube}}\left(\frac{9V^{5/4}}{8A_\mu}\right)^{4/9}\left(\Gamma\left(\frac{4}{9}\right) - \Gamma_u\left(\frac{4}{9}, t_{max}\right)\right)$$

$$F(y) = \frac{4}{9}n_{tube}\frac{\lambda^2\zeta}{kT}e^{-2\frac{\tilde{B}\lambda}{D}n_{tube}}\left(\frac{9V^{5/4}}{8A_\mu}\right)^{4/9}\left(\Gamma\left(\frac{4}{9}\right) - \Gamma_l\left(\frac{4}{9}, t_{min}\right) - \Gamma_u\left(\frac{4}{9}, t_{max}\right)\right) \tag{2.36}$$

$$t_{min} = \frac{8}{9}\frac{A_\mu}{V^{5/4}}(y - n_{tube})^{9/4} \;;\; t_{max} = \frac{8}{9}\frac{A_\mu}{V^{5/4}}(N_0 - n_{tube})^{9/4} \tag{2.37}$$

And finally the mean first passage time is evaluated through numerical integration of eq. 2.32.

Chapter 3

Computational Approach

Our main approach to the problem was through molecular dynamics (MD) simulations. In order to run the simulations a model for the polymer as well as an algorithm to solve the equations of motion.

3.1 Building a Model

A computational model for a polymer needs to account for several things, from the scales involved to the types of interaction.

3.1.1 Coarse Graining

The first step is choosing the level of coarse graining. The level of coarse graining, or the scale at which the polymer is being studied, depends on the specific topic of study. When simulating a system related to DNA, several choices are available starting from the smallest scales where the full structure of DNA is apparent, to larger scales where it is modeled as a simple semi-flexible polymer. For a review on the different types of coarse graining for DNA and chromatin see [19].

3.1.2 The Solvent

In reality, polymers are studied in solutions where they are dissolved in solvents. The quality of the solvent affects the equilibrium size of the polymer, where a good solvent makes the polymer swell, while a poor solvent leads to the collapse of the polymer. The different types of solvents are summarized is section 2.2.1. Computationally, the solvent can be accounted for either explicitly or implicitly. In explicit solvent models, along with the polymer, there are solvent particles that interact with each other and the monomers, and the quality of the solvent is specified by the affinity between the different particles, also as described in section 2.2.1. The second way is that of implicit solvent, where the solvent is simply accounted for through the type of non-bonded interactions. In the good solvent regime the interaction is purely repulsive, while it has an attractive part in the poor solvent regime.

3.1.3 The Langevin Thermostat

The sizes of most monomers in polymer solutions are small enough to be affected by collisions with solvent particles. This results in Brownian diffusion and friction. The fluctuation-dissipation theorem relates those two effects through the Einstein-Smoluchowski relation

$$\mathcal{D} = \frac{kT}{\zeta}$$

where ζ is the drag coefficient and $\mathcal{D}$ is the diffusion coefficient. This relation can be included in the equations of motion by accounting for a drag force $F_d = -\zeta\dot{x}$ and a random force $F_r = \sqrt{2\zeta kT}R(t)$ where $R(t)$ is a stationary Gaussian process satisfying

$$\langle R(t) \rangle = 0$$

$$\langle R(t)R(t')\rangle = \delta(t-t')$$

3.1.4 Interactions

In most models, interactions between particles in a coarse grained polymer are divided into two-body, three-body, or four-body interactions. Two body interactions are in turn divided into two types: bonded, and non-bonded interactions. Bonded interactions arise between nearest neighbors, and the interaction potential is chosen in such a way that the particles stay bonded to each other. Non-bonded interactions arise between polymers that are far away from each other along the chain, and this interaction reflects the quality of the solvent in implicit solvent models. Three-body interactions are related to the bending of the polymer chain. If the chain is not fully flexible, bending will be penalized by a potential which depends on the relative positions of three successive particles along the chain (cf. Figure 3.1). As for four-body interactions, they are specified by a potential that depends on the torsion angle, which is specified by the relative angles of the first and last particles in a sequence of four consecutive monomers along the chain.

3.1.5 Confinement

The polymer is confined by rigid repulsive walls. The geometry and the interaction of the polymer with the walls need to be specified and are detailed in section 3.2.1.

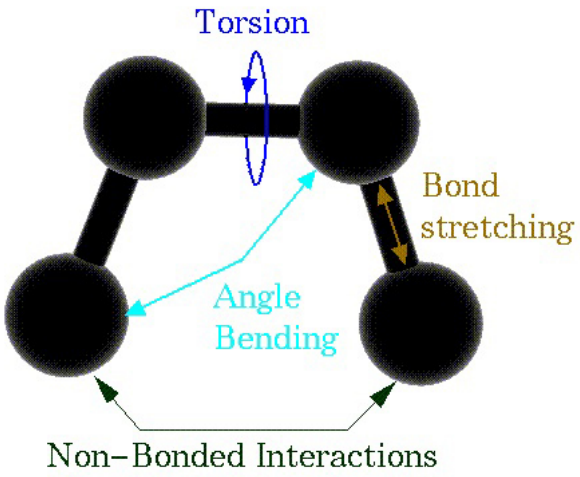

Figure 3.1: Different Types of Interactions.

3.2 Our Model

In our simulations we used a bead-spring model for the polymer. The polymer is modeled as a chain of beads connected by nonlinear springs, and we use the implicit solvent approach in the good solvent regime. The polymer is fully flexible such that we ignore three-body and four-body interactions (see section 3.1.4). The chosen types of interaction are the following:

- Bonded interactions: a combination of the Finitely Extensible Nonlinear Elastic (FENE) potential and the repulsive part of the Lennard-Jones (LJ) potential, also known as the Weeks-Chandler-Andersen (WCA) potential.

$$U(r_{ij}) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] & \text{for } r_{ij} < r_{WCA} = 2^{1/6}\sigma \\ -\frac{1}{2}kR_0^2 ln\left[1 - \left(\frac{r_{ij}}{R_0}\right)^{2}\right] & \text{for } r_{ij} \leq R_0 \\ \infty & \text{for } r_{ij} > R_0 \end{cases} \tag{3.1}$$

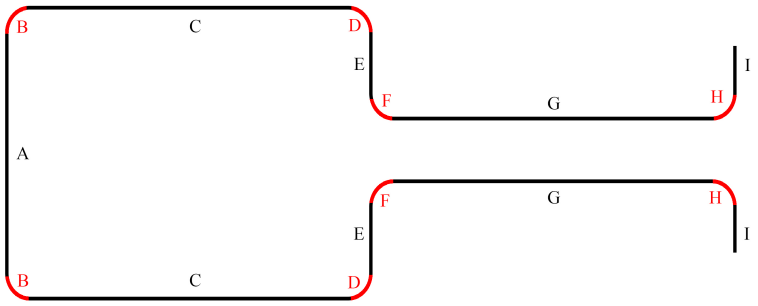

Figure 3.2: Capsid Geometry.

- Non-bonded interactions: To run the simulations in the good solvent regime we chose the WCA potential for non-bonded interactions

$$U(r_{ij}) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \epsilon & \text{for } r_{ij} < r_{WCA} = 2^{1/6}\sigma \\ 0 & \text{for } r_{ij} > r_{WCA} \end{cases} \tag{3.2}$$

For our calculations we chose $\sigma = 1$, $\epsilon = k_BT = 1$, $k = 30k_BT/\sigma^2 = 30$, $R_0 = 1.5\sigma = 1.5$, and for the capsid walls we chose $\sigma_b = 0.1$, and the mass of a monomer is taken to be $m = 1$. To make the analogy with biological systems, taking the monomer size to be on the order of the persistence length of single stranded DNA, our length unit is about $30\mathring{A}$, which encapsulates about 10 bases, making the mass of a monomer about 5×10^{-23} Kg. Taking $T = 300K$, the time unit in our simulation is equivalent to about 1ns.

3.2.1 Details of the Confining Capsid

The capsid geometry is illustrated in Figure 3.2. The corners are modeled as quarter circles with radius equal to the range of the repulsive interaction $r_{WCA} = 2^{1/6}\sigma$. The walls interact with the particles via the same repulsive potential used for non-bonded interactions (eq. 3.2) but with $\sigma = 0.1$. The force $F_{wall}(\boldsymbol{r}_m - \boldsymbol{r_c})$

is thought of as originating from a point on the wall with position $\boldsymbol{r}_c$ and acts along the line connecting that point to the particle at position $\boldsymbol{r}_m$. The position of the origin of the force for the different boundaries A-I is given in table 3.1 where x_m, y_m, and z_m are the monomer coordinates, x_p, y_p, and z_p are the positions of the centers of curvature for the corners, which are in turn summarized in table 3.2, $r_{pm} = |\boldsymbol{r}_p - \boldsymbol{r}_m|$ is the distance between the relevant center of curvature and the monomer, D and L are the diameter and length of the chamber, d and l are the diameter and length of the tube respectively.

Surface Label	x_c	y_c	z_c
A	0	y_m	z_m
Corners B, D, F, H	$\frac{r_{WCA}(x_m - x_p)}{r_{pm}} + x_p$	$\frac{r_{WCA}(y_m - y_p)}{r_{pm}} + y_p$	$\frac{r_{WCA}(z_m - z_p)}{r_{pm}} + z_p$
C	x_m	$\frac{y_m}{z_m} z_c$	$sgn(z_m) \frac{D/2}{\sqrt{1+\frac{y_m^2}{z_m^2}}}$
E	L	y_m	y_m
G	x_m	$\frac{y_m}{z_m} z_c$	$sgn(z_m) \frac{d/2}{\sqrt{1+\frac{y_m^2}{z_m^2}}}$
I	$L + l$	y_m	y_m

Table 3.1: Coordinates of the force centers from the different surfaces as functions of monomer coordinates x_m, y_m, and z_m. x_p, y_p, and z_p are the positions of the centers of curvature for the corners. $r_{pm} = |\boldsymbol{r}_p - \boldsymbol{r}_m|$. D and L are the diameter and length of the chamber, d and l are the diameter and length of the tube respectively. r_{WCA} is the range for the WCA interaction for the boundary. $sgn(x)$ is the sign of x.

Surface Label	x_p	y_p	z_p
B	r_{WCA}	$\frac{y_m\left(D/2-r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$	$\frac{z_m\left(D/2-r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$
D	$L-r_{WCA}$	$\frac{y_m\left(D/2-r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$	$\frac{z_m\left(D/2-r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$
F	$L+r_{WCA}$	$\frac{y_m\left(d/2+r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$	$\frac{z_m\left(d/2+r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$
H	$L+l-r_{WCA}$	$\frac{y_m\left(d/2+r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$	$\frac{z_m\left(d/2+r_{WCA}\right)}{\sqrt{y_m^2+z_m^2}}$

Table 3.2: Coordinates of the centers of curvature for the corners. x_m, y_m, and z_m are the monomer coordinates. x_p, y_p, and z_p are the positions of the centers of curvature for the corners. D and L are the diameter and length of the chamber, d and l are the diameter and length of the tube respectively. r_{WCA} is the range for the WCA interaction for the boundary.

3.2.2 The Effective Confinement Size

When dealing with theoretical calculations, the walls of confinement are always assumed to be rigid, while the confinement in the simulations is achieved through a smooth potential (see figure 3.3)A. In order to be able to compare the simulation results with the theory we introduce an effective diameter of confinement in the following way.

For a rigid box of length L, where the potential is zero inside but infinite outside, evaluating the partition function yields $Z = \int_{-\infty}^{+\infty} e^{-\frac{U(x)}{kT}} dx = L$. Now introducing a smooth repulsive potential to emulate the walls such that

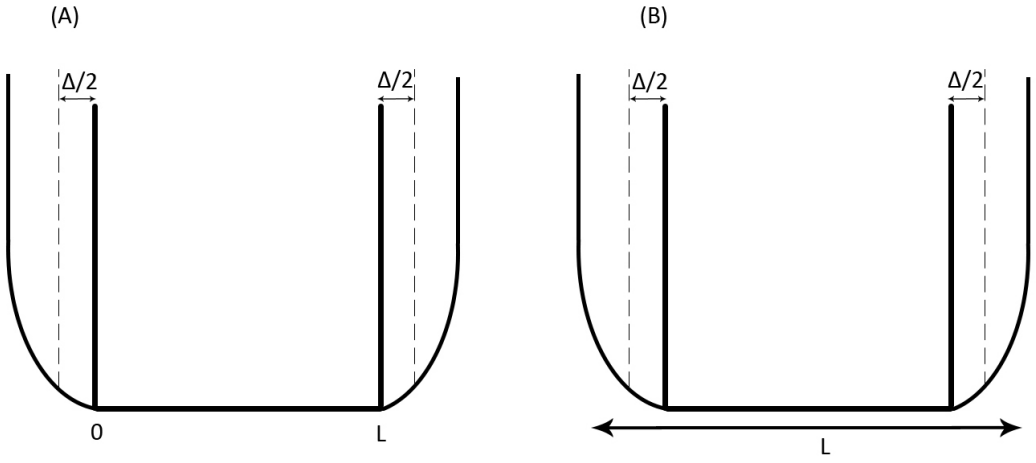

Figure 3.3

$$U(x) = \begin{cases} U_W(x) & \text{for } x < 0 \\ 0 & \text{for } 0 \leq x \leq L \\ U_W(L-x) & \text{for } x > L \end{cases} \tag{3.3}$$

with the condition that $U(x)$ is continuous at $x = 0$ and $x - L$. The partition function becomes

$$\begin{aligned} Z &= \int_{-\infty}^{+\infty} e^{-\frac{U(x)}{kT}} \\ &= \int_{-\infty}^{0} e^{-\frac{U_W(x)}{kT}} dx + \int_{0}^{L} dx + \int_{0}^{+\infty} e^{-\frac{U_W(L-x)}{kT}} dx \\ &= L + \Delta \end{aligned} \tag{3.4}$$

where $\Delta = \int_{-\infty}^{0} e^{-\frac{U_W(x)}{kT}} dx + \int_{0}^{+\infty} e^{-\frac{U_W(L-x)}{kT}} dx$ is the increase in the box size due to the smoothening of the walls. Specifying the potential and evaluating Δ allows us to get the effective dimensions after specifying all the parameters (see figure 3.3). When it comes to the effective diameter of a cylindrical box, we need

to find the effective area of confinement instead of the effective length, and then from that find the effective diameter. Doing the calculation however showed that the numerical values obtained this way are very close to what we get from the previous way, and so we adopt the same correction for both the diameters and the lengths.

In our simulations we define the box sizes by the position of the singularity of the WCA potential as shown in figure 3.3B. This leads to confinement sizes that are slightly less than the chosen diameters D and d and chamber length L as defined in the simulations. Evaluating the partition function for the WCA repulsive walls (eq. 3.2) leads to a correction (for $kT = \epsilon = 1$)

$$\Delta = 0.2139\sigma \tag{3.5}$$

And finally the effective parameters are

$$L_{eff} = L - 2.031\sigma \tag{3.6}$$

$$D_{eff} = D - 2.031\sigma \tag{3.7}$$

3.3 The Algorithm

A very important thing to consider while setting up molecular dynamics simulations is the choice of algorithm to integrate the equations of motion. The considerations that go into the choice are those of precision and computational efficiency.

3.3.1 The Verlet Algorithm

The aim is to solve differential equations of the form:

$$\dot{x} = v(t)$$

$$\ddot{x} = \frac{f(x, v, t)}{2m}$$

This is done by updating the current positions $x(t)$ to their new values $x(t + \Delta t)$. Doing a Taylor expansion on the new and the old positions we get

$$x(t + \Delta t) = x(t) + \dot{x}(t)\Delta t + \ddot{x}(t)\Delta t^2 + \dddot{x}\Delta t^3 + \mathcal{O}(\Delta t^4)$$

$$x(t - \Delta t) = x(t) - \dot{x}(t)\Delta t + \ddot{x}(t)\Delta t^2 - \dddot{x}\Delta t^3 + \mathcal{O}(\Delta t^4)$$

adding the two above equations we get

$$x(t + \Delta t) \approx 2x(t) - x(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 + \mathcal{O}(\Delta t^4)$$

It can be seen that given an initial condition and a subsequent position, and if the forces depend only on position, all positions can be calculated without having to calculate velocities. Calculating velocities is important in our simulations however because of the presence of viscosity. Several methods can be derived from the Verlet method to compute velocities. Most of the computational time is spent calculating forces, which makes the choice of method for the calculation of

velocities irrelevant for speed; however, the methods differ in accuracy. We use the Velocity Verlet method outlined below

1. Define initial positions $x_i(t)$ and velocities $v_i(t)$
2. Compute forces $f_i(t) = -\frac{dU(x_i)}{dx_i} + F_d(v_i) + F_r$ where $U(x_i)$ is the particle potential, F_d and F_r are defined in section 3.1.3
3. Update velocities by half a step using $v_i(t + \frac{1}{2}\Delta t) = v_i(t) + \frac{f_i(t)}{m}\frac{\Delta t}{2}$
4. Update positions by a full time step using $x_i(t + \Delta t) = x_i(t) + v_i(t + \frac{1}{2}\Delta t)$
5. Compute new forces $f_i(t + \Delta t)$
6. Update velocities to full time step $v_i(t + \Delta t) = v_i(t + \frac{1}{2}\Delta t) + \frac{f_i(t+\Delta t)}{m}\frac{\Delta t}{2}$
7. Repeat steps 3-5 until satisfied

3.3.2 The Verlet List

In most MD simulations, the only type of interaction that is accounted for is pair-wise interaction. This means that for each iteration we need to calculate the forces for $N(N-1)/2$ pair interactions. In our simulations interactions are truncated, saving some time on force calculations, we however still need to calculate the distances between $N(N-1)/2$ pairs to check which particles interact with each other, so that the simulation time scales like N^2. A way to reduce the dependence of the computational time on the number of particles is to use a Verlet List [20]. The first step in preparing the list consists of introducing a cutoff radius r_v slightly larger than the cutoff radius for the forces r_{WCA}. For a particle i, the list will include all particles within a distance r_v from i. During each iteration, only the distances of particles in the list will be calculated, which is a calculation of order

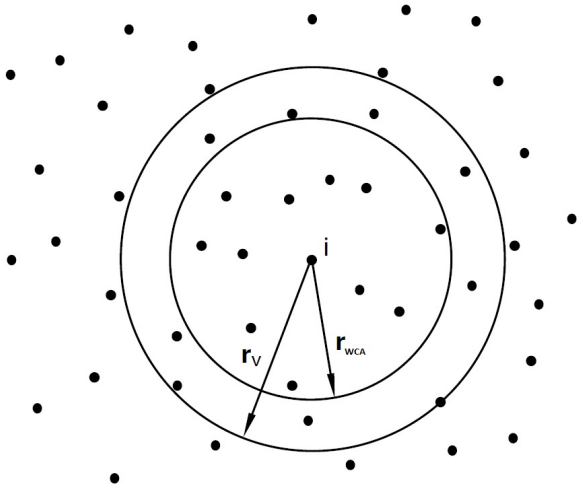

Figure 3.4: The two radii for the Verlet list. Particles within r_v are included in the list for particle i, while particles within r_{WCA} interact with i [2].

N, and the particle i will interact with particles within the interaction range r_{WCA}. From time to time the list needs to be updated; the update, an operation of order N^2, will be performed whenever the displacement of any particle in the simulation exceeds $r_v - r_{WCA}$. The use of a Verlet list in the algorithm reduces the dependence of the computational time from N^2 to about $N^{3/2}$ [2].

3.4 Calculating the Pressure Distribution

As mentioned in section 2.4 we wish to calculate the local pressure tensor elements averaged over volume slices. We choose the Irving-Kirkwood contour (eq. B.2) for the interaction part of the pressure (see appendix B), and average over cylindrical slices (see figure C.1). We wish to calculate the interaction contribution for two monomers i and j with positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ respectively. The contour going from j to i is defined as

$$\boldsymbol{l} = \boldsymbol{r}_j + \lambda \boldsymbol{r}_{ji} = \boldsymbol{r}_j + \lambda(\boldsymbol{r}_i - \boldsymbol{r}_j)$$

with $0 < \lambda < 1$. In cylindrical coordinates has components

$$\boldsymbol{l}_x = (x_j + \lambda x_{ji})\hat{\boldsymbol{x}}$$

$$\boldsymbol{l}_\rho = (y_j + \lambda y_{ji})\hat{\boldsymbol{y}} + (z_j + \lambda z_{ji})\hat{\boldsymbol{z}}$$

$$\boldsymbol{l}_\phi = \boldsymbol{l}_x \times \boldsymbol{l}_\rho$$

We choose to calculate the pressure distribution for radii between 0 and $D/2$, and axial positions between 0 and x_{max}. To use the formulas derived in appendix B we need to find λ_1 and λ_2 which define the portion of the contour inside a specific slice. The algorithm is detailed in appendix C.

Chapter 4

Simulation Results

4.1 Infinite Tube

After setting up the simulations we wished to test their validity by comparing with predictions from the scaling laws discussed in section 2.3. To do this we ran simulations for a polymer in an infinite tube and studied the dependence of the end-to-end distance on the diameter and the number of monomers. We varied the number of monomers between 50 and 200, and the diameter between 1 and 50. The results are shown in figure 4.1.

As explained in section 2.3, the scaling regime is achieved for $a << d << R_F$ where $a \approx 1$ is the size of an individual monomer and R_F is the end-to-end distance for a free real chain. From figure 4.1a we can see that the end-to-end distance changes following three regimes. The first regime is when $d_{eff} \approx a$ or $d_{eff} < 3$; here the chain is forced to have almost straight conformations due to the very strong confinement. The second regime is for intermediate diameters; here we are in the scaling regime, where we expect the end-to-end distance to change proportionally to $d^{-2/3}$, to check this we plot the end-to-end distance versus the

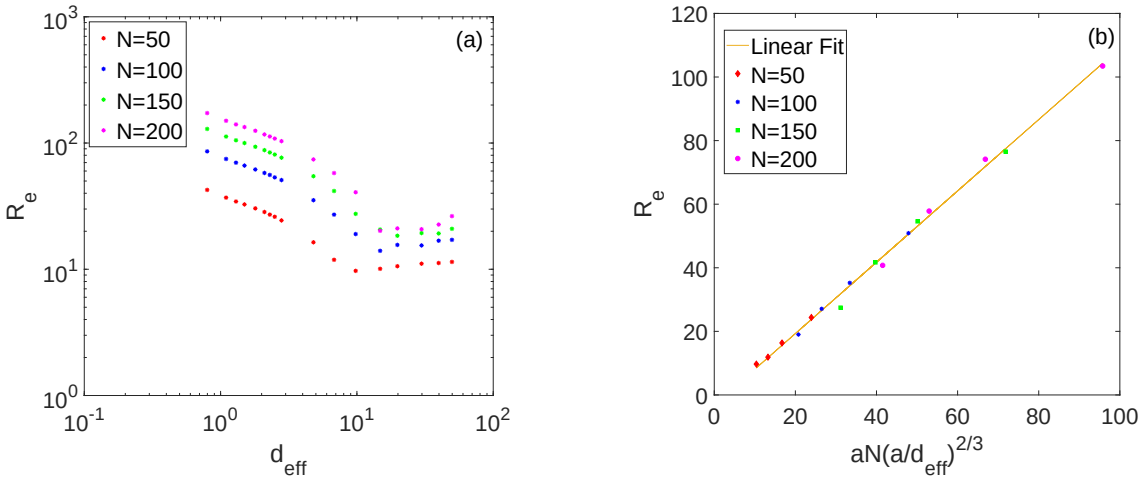

Figure 4.1: *(a)* Variation of the end-to-end distance for a polymer in a tube for different diameters, we can distinguish 3 regimes, for $d_{eff} < 3$, $3 < d_{eff} \lesssim 10$, and $d_{eff} \gtrsim 10$. *(b)* End-to-end distance for polymers of different lengths versus the scaling variable $aN(a/d)^{2/3}$ for diameters $3 < d_{eff} < 10$ shows very good agreement with the scaling theory result.

scaling variable $aN(a/d)^{2/3}$ (figure 4.1b), which shows very good agreement with the expected scaling law; the slope of the line is the numerical coefficient A_e that appears in eq. 2.5

$$A_e = 1.06$$

Finally the third regime is achieved large diameters; here the chain is not under strong confinement anymore and the size of the chain increases with the diameter.

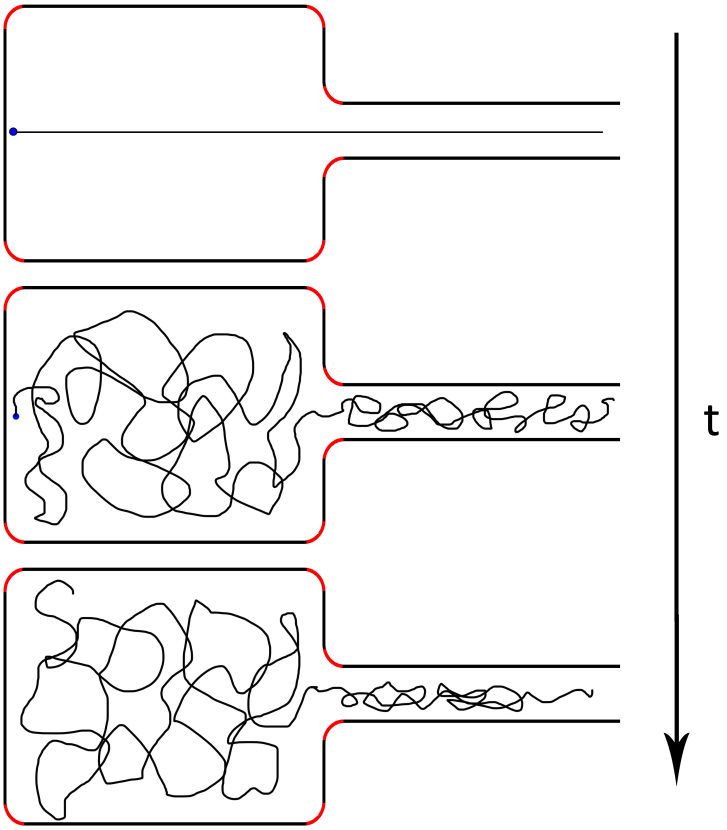

Figure 4.2

4.2 Ejection

4.2.1 Preparing the Initial Conditions

We wish to start the ejection process with an equilibrated conformation of the polymer in the capsid. To generate the initial condition we start with a fully extended chain and fix the first monomer at the bottom of the capsid. We run the simulation for 2×10^7 (10^4 time units) steps, during which time the chamber starts to fill up. Then the fixed monomer is released and the polymer is left to equilibrate for an additional 6×10^7 (3×10^4 time units). If the polymer is long enough we will end up with a configuration where we have some of the monomers reside in the chamber and the rest are in the tube (see figure 4.2).

Theoretical considerations support the idea that this conformation minimizes the confinements free energy if the tube is long enough. Figure 4.3 shows the time evolution of the number densities in the chamber and the tube.

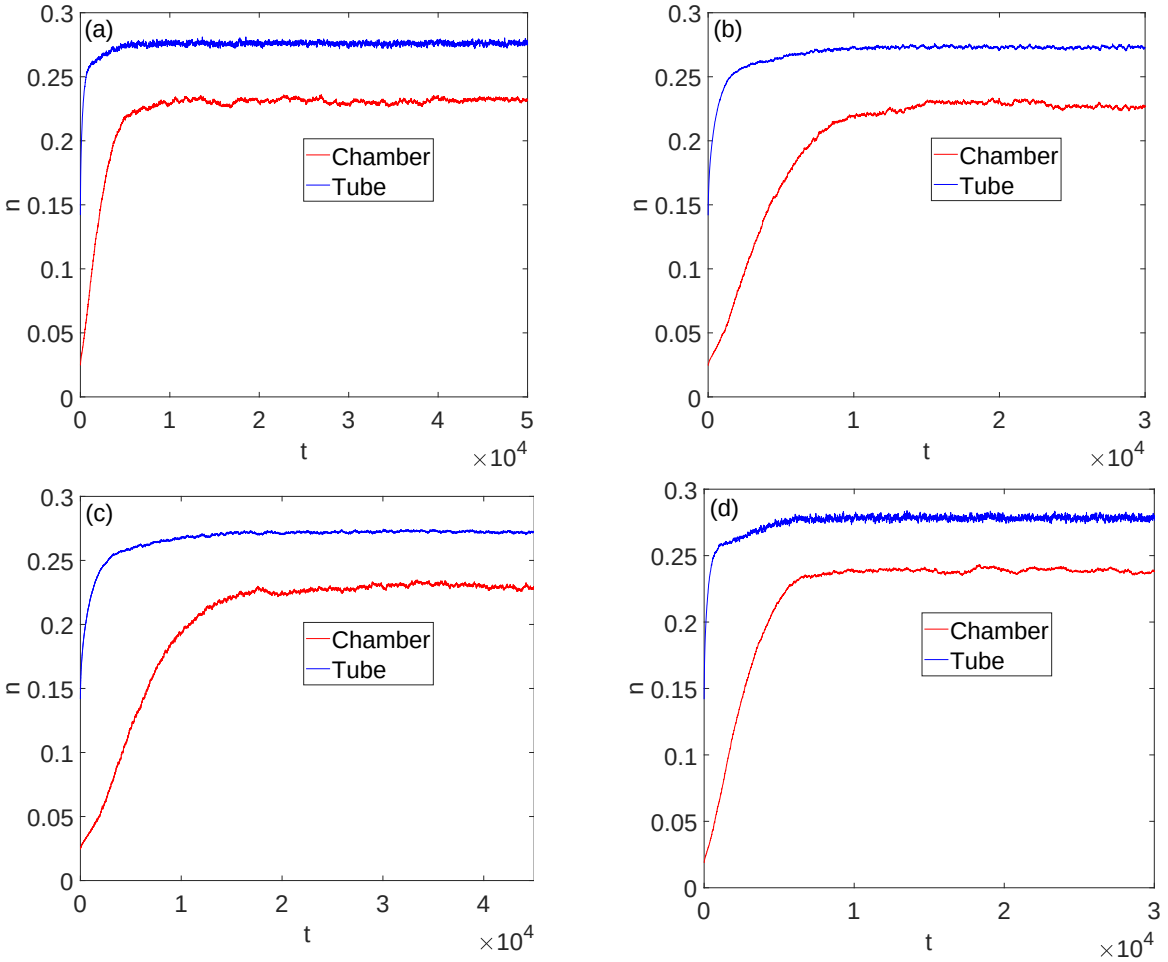

Figure 4.3: Time evolution of the number density n in the chamber and the tube for (a) $N = 200,\ D = 7,\ L = 10,\ d = 3$, (b) $N = 300,\ D = 7,\ L = 10,\ d = 3$, (c) $N = 400,\ D = 7,\ L = 10,\ d = 3$, (d) $N = 200,\ D = 8,\ L = 10,\ d = 3$.

Something to note here is the lack of dependence of the density in the chamber on the length of the chain (comparing figures 4.3a, b, and c) and the diameter of the chamber (comparing figures 4.3a and d). This confirms the what we mentioned in section 2.5.1 about the equilibrium density in the chamber being determined entirely by the diameter of the tube.

Typical CPU times for the preparation of one initial configuration for $N = 200$ is about 1.5 hours, with a total CPU time of about 13 days for the 200 configurations prepared for each set of parameters. Running 10 simulations in parallel reduced this to about 30 hours for 200 initial conditions. The CPU time for $N = 400$ is about 7 hours for one initial configuration.

4.2.2 Density and Pressure Profiles of Equilibrated Chains

Using the method described in section 2.4 and appendices B and C, we calculated the pressure profile for an equilibrated chain. Results for a chain of length $N = 200$ are shown in figure 4.4 which also shows the local density distribution, and figures 4.5 and 4.6 show the variation in the axial direction and the radial direction respectively. The density distribution shows that there is a larger density in the tube, which fits the results shown in figure 4.3. Looking at the pressure profiles raised certain questions concerning the consistency between the acquired profiles and the established mechanical equilibrium. Looking at figure 4.5, we notice that there are small nonzero pressure gradients as we move in the axial direction at fixed radius, the large gradients near $x = 15$ being numerical artifacts. Normally, gradients of pressure indicate the presence of forces in the direction of negative gradients, and in the presence of external forces the gradients in pressure are canceled by external forces to achieve mechanical equilibrium. However, this is the case when dealing with simple liquids and gases. What we have in our case is a chain that exhibits elastic properties, and is under the effect of various stresses. The statement for mechanical equilibrium in this case is inferred from equation 2.13 as $\nabla \cdot \boldsymbol{P}(\boldsymbol{x}) = 0$. This equation involves off diagonal terms of the pressure tensor and their derivatives. Checking this condition and obtaining the

pressures in the non-equilibrium situation is left for future investigation.

A simulation with 200 monomers and 6×10^7 steps (30000 time units) takes about 2.3 hours of CPU time to finish. The size of the relevant data is about 13MB for a spacial resolution of 0.01 in the radial direction and 0.1 in the axial direction for the capsid size we studied.

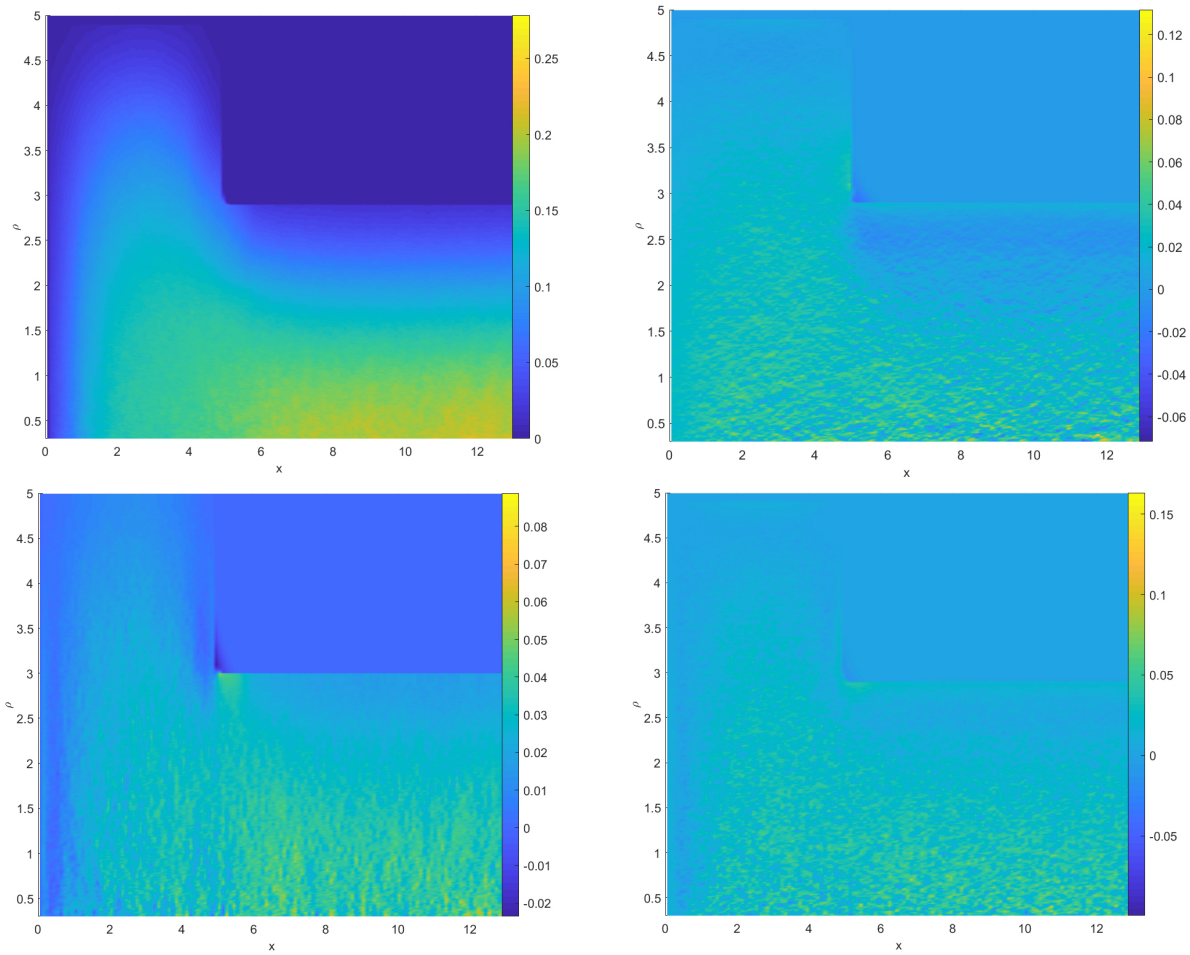

Figure 4.4: Top left: Density distribution. Top right: $P^{xx}(x, \rho)$. Bottom left: $P^{\rho\rho}(x, \rho)$. Bottom right: $P^{\phi\phi}(x, \rho)$. The parameters are $N = 200,\ D = 10,\ L = 5,\ d = 6$.

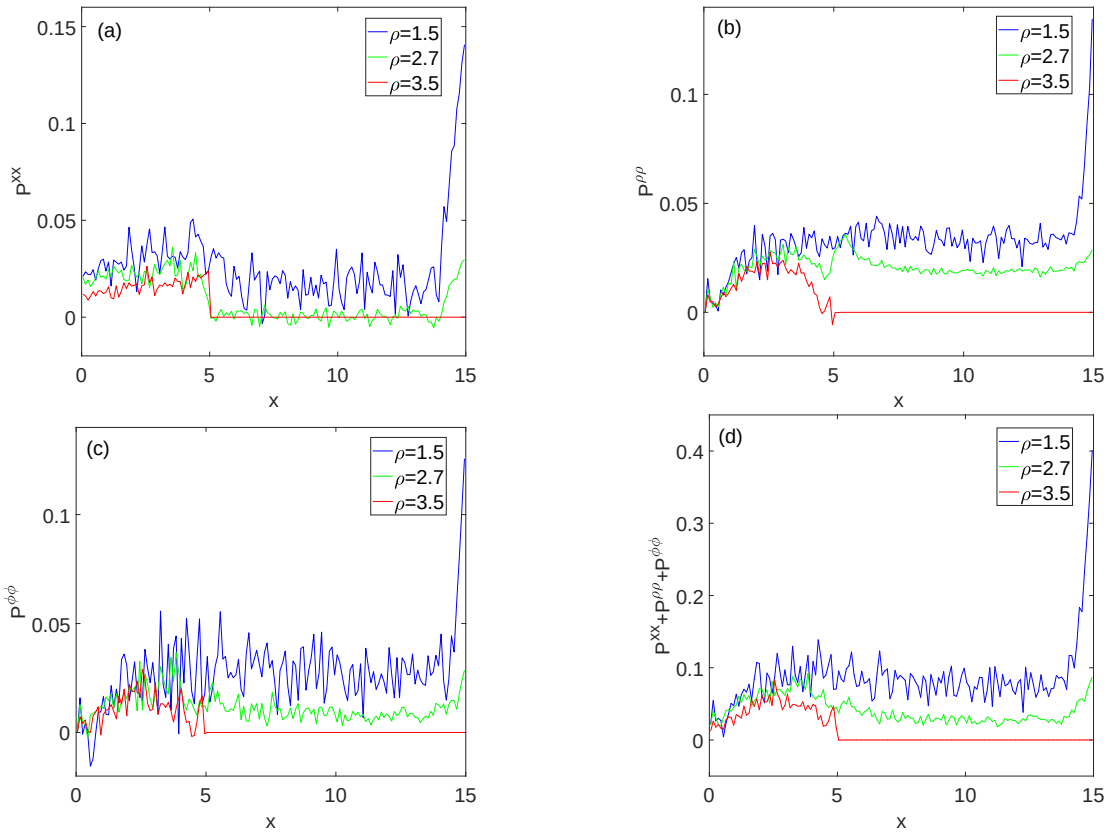

Figure 4.5: Pressure tensor components versus axial position for different radii. (a) P^{xx}, (b) $P^{\rho\rho}$, (c) $P^{\phi\phi}$, (d) $tr(\boldsymbol{P})$.

4.2.3 Determination of the Prefactor A_μ

In order to get the theoretical predictions described in sections 2.5.2 and 2.6, we need to determine the coefficient A_μ. As stated above, at equilibrium the chemical potential in the chamber and the tube are equal. The expressions for the chemical potentials can be obtained from the free energies found in section 2.3. With our choices of $kT = 1$ and $a = 1$ we get

$$A_\mu = BA_e d^{-5/3} \left(\frac{N}{V}\right)^{-5/4} \tag{4.1}$$

where d is the diameter of the tube, $V = \frac{\pi}{4} D^2 L$ is the volume of the chamber, and N is the number of monomers in the chamber. Now we can use the number

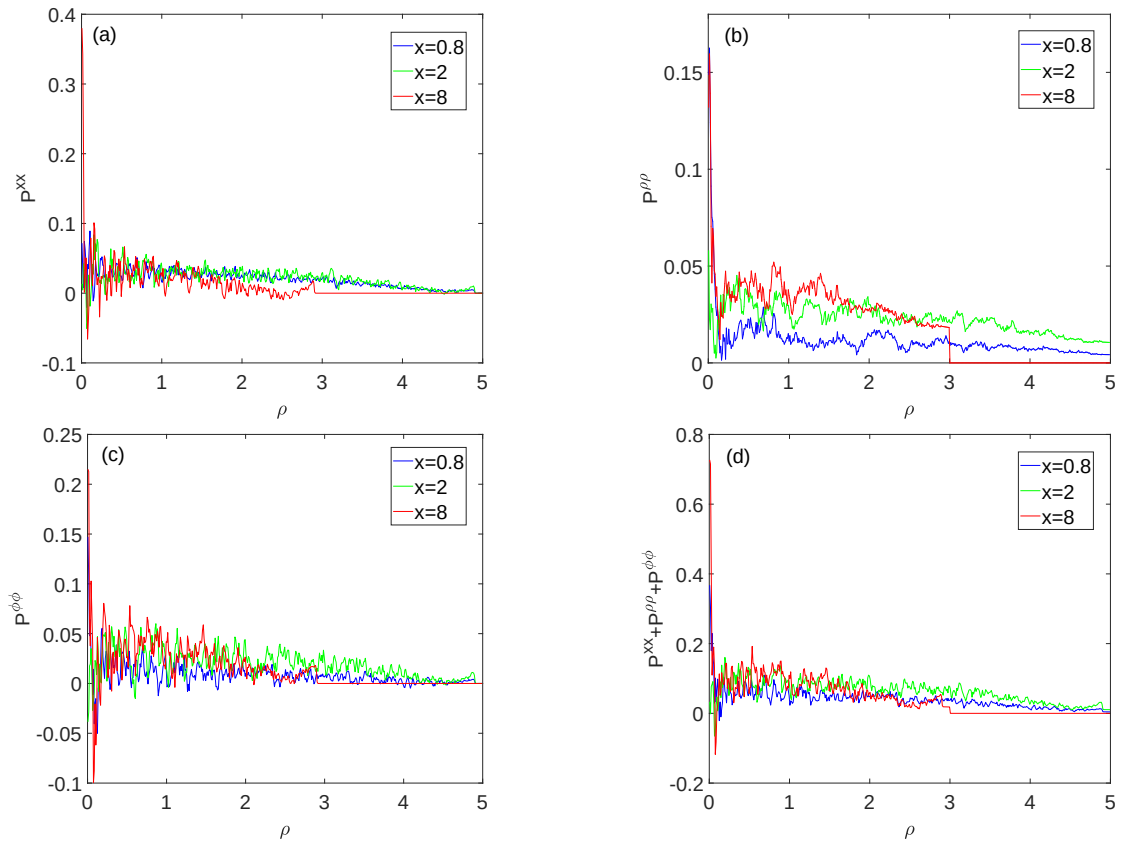

Figure 4.6: Pressure tensor components versus radial position for different axial positions. (a) P^{xx}, (b) $P^{\rho\rho}$, (c) $P^{\phi\phi}$, (d) $tr(\boldsymbol{P})$.

densities obtained from the initial condition preparations to find A_μ. Using $B = 5$, $A_e = 1.06$, $d = d_{eff}$, and using the equilibrium values for the density in the chamber we find

$$A_\mu \approx 5.8$$

4.2.4 Forced Pulling Simulations

In the derivation carried out in section 2.5.2 we made the assumption that the drift velocity in the chamber is very small, so that the chamber contribution to the dissipated power can be ignored. To validate this assumption we ran simula-

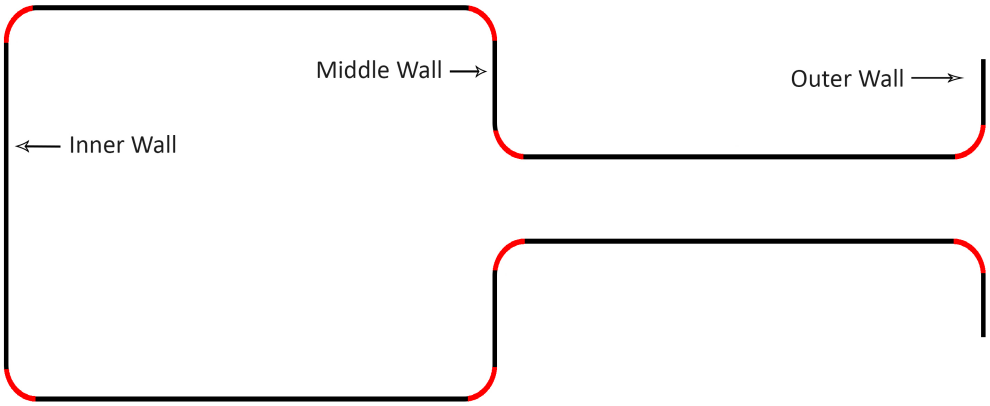

Figure 4.7: Capsid cartoon showing our definitions of the inner, middle, and outer wall.

tions with an infinitely long tube, and we applied a pulling force on the outermost monomer in the tube that pulls in the direction going from the chamber to the tube. This was done for polymers of different lengths ($N = 200, 300, 400$) and for different pulling forces ($f_{pull} = 1, 2, 3$). Figure 4.8 shows the values of the speed in the axial direction in the chamber and the tube v_{drift} vs time. For both strong and weak pulling forces, there is a trend in the drift velocity in the tube to decrease then increase as a function of time. The reason for this is that as the density in the chamber decreases, the axial positive force originating from the inner wall goes down. However, the retarding force from the middle wall remains significant, and slows down the ejection process. Once the chamber is almost empty though, the only axial force remaining is the pulling force, and this leads to an increase of the drift velocity again. This is confirmed by looking at the wall forces and drift velocity in the tube as a function of the number of monomers in the chamber shown in figure 4.9. In both 4.9a and 4.9b it can be seen that for large N_{ch} the contributions from both walls are comparable, then start deviating as N_{ch} decreases and the net force from the middle wall becomes larger than from the inner wall. The drift velocity follows a trend that is compatible with what is

found for the wall forces.

Regardless of the trend in the drift velocity, it can be seen that at for a strong pulling force v_{drift} in the tube is much larger than in the chamber, which fits the assumptions made in section 2.5.2. For small pulling force however, there are times when the velocities in the chamber and the tube are comparable, which is a sign that the equations may need to be modified.

CPU time needed for the last monomer leaves the chamber is about 0.12 hours for $f = 3$ and $N = 200$, and 2 hours for $f = 1$ and $N = 400$.

4.2.5 Spontaneous Ejection Simulations

The main aim of this work is to investigate the spontaneous ejection of a chain from a capsid driven by entropic forces. The initial configuration is the chain equilibrated with an infinite tube. We take this configuration and cut the tube short. If the tube opening is far away from the end of the chain, the chain will be in a long lived metastable state. To make the system unstable, one needs to form a tail extending outside the capsid. If the system is in the metastable state, the ejection time will be dominated by the time needed to form a tail outside. To start with an unstable configuration in the simulations, we cut the tube at a length such that a part of the chain is already outside and a tail is formed (see figure 4.10). We ran simulations for different parameters but fixed tube diameter and the results are summarized is table 4.1.

As a first test, we checked whether the ejection times fit the expectations from the deterministic equation. As mentioned in section 2.5.2, we expect the chamber evacuation times to go as $t_1 \propto lV^{5/4}d^{-2/3}$. Since the tube diameter

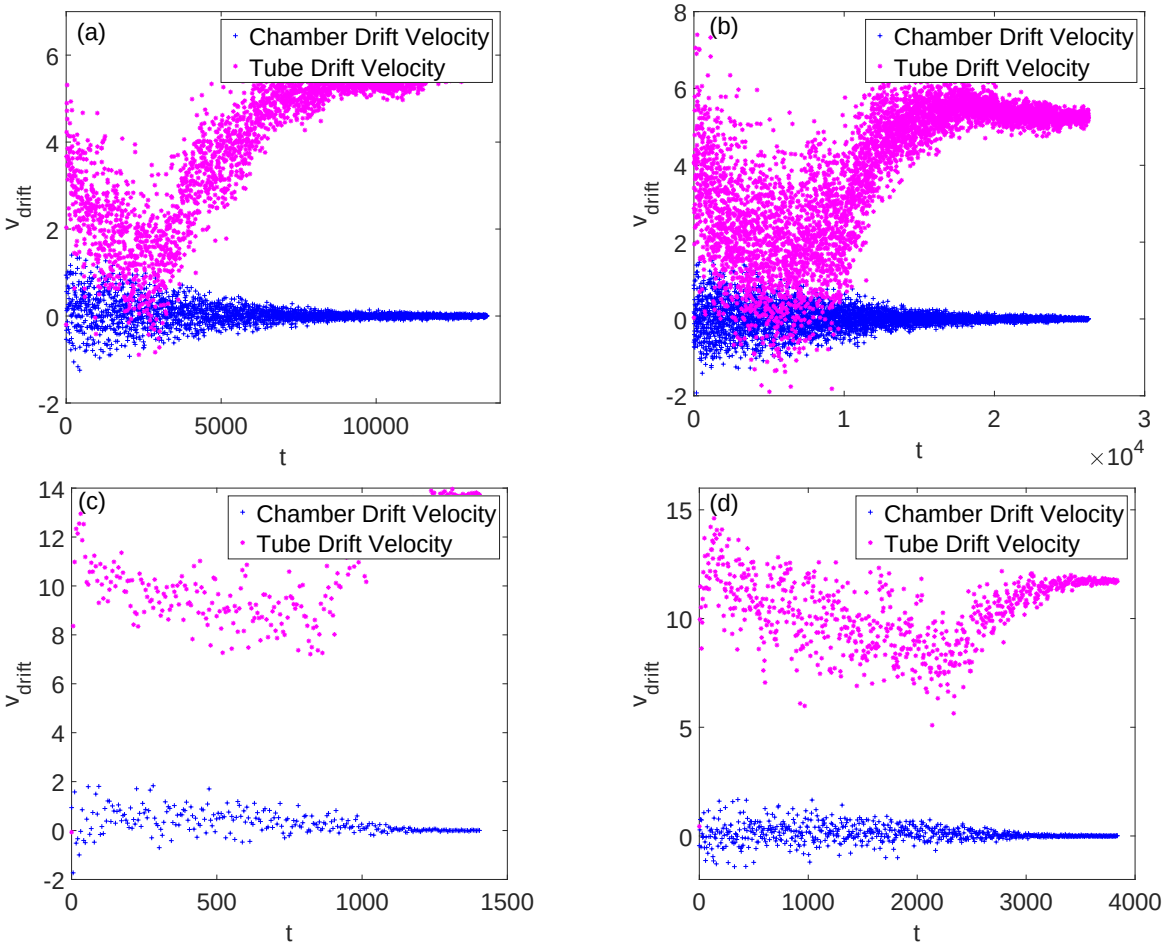

Figure 4.8: Drift velocity summed over all monomers in the chamber and in the tube for (a) $N = 200,\ f_{pull} = 1$, (b) $N = 400,\ f_{pull} = 1$, (c) $N = 200,\ f_{pull} = 3$, (d) $N = 400,\ f_{pull} = 3$.

was not varied, we plotted the chamber evacuation times t_{ch} versus the variable $l(D^2L)^{5/4}$ in figure 4.11 which shows good agreement with the prediction.

Next we looked at the distribution of ejection times. Figure 4.12 shows histograms of total ejection times for capsids of the same size, but different chain lengths. The distributions exhibit a certain width, which is an indications of the stochastic nature of the ejection process, and gives better justification for the Fokker-Planck Equation approach.

Now looking at the table and comparing different values, we can see the

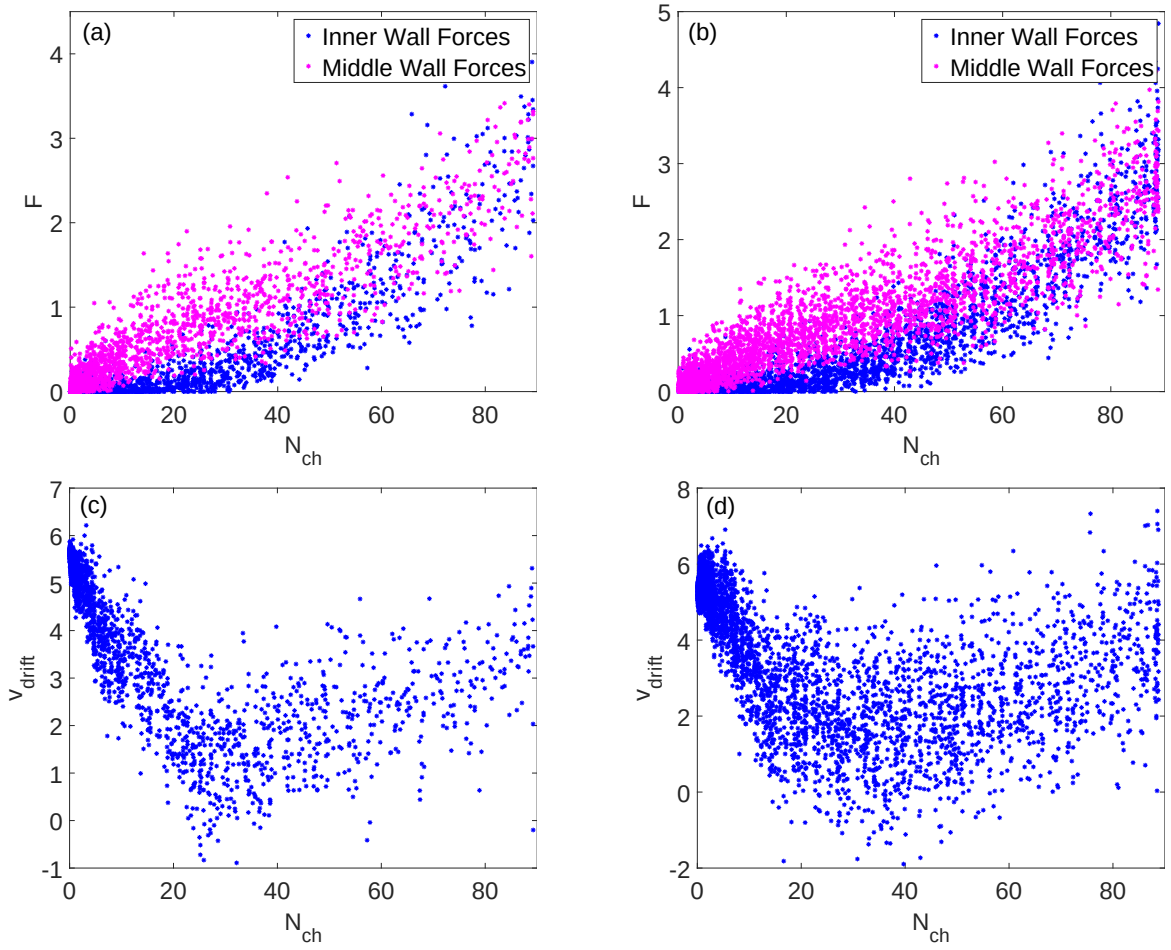

Figure 4.9: Above, wall forces against the number of monomers in the chamber N_{ch} for (a) $N = 200,\ f_{pull} = 1$, (b) $N = 400,\ f_{pull} = 1$. Below, tube drift velocity against the number of monomers in the chamber N_{ch} for (c) $N = 200,\ f_{pull} = 1$, (d) $N = 400,\ f_{pull} = 1$.

following:

1. The deterministic expectation t_{det} always overestimates the ejection times, which is something expected.

2. The theoretical first passage times overestimate the ejection times for tubes longer than 20, but underestimate them for $l = 20$.

3. Comparing the simulation times for rows 1, 4, and 6, we see that the ejection

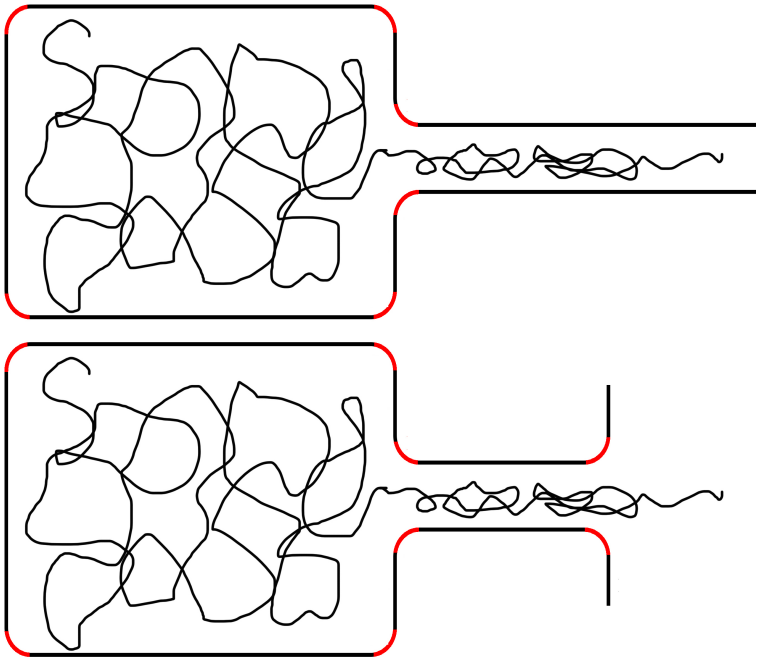

Figure 4.10: Top: Geometry during equilibration with infinite tube. Bottom: Geometry during the ejection simulation.

time for $N = 200$ is larger than for $N = 300, 400$, but is almost the same for the latter.

4. Comparing rows 1, 2, and 3 we see that the ejection time increases with both D and l.

5. Comparing rows 2, 5, and 7, we see that the ejection time changes significantly with the length of the tube (in theory the ejection time must not depend on the total length of chain).

Point 2 can be explained by looking at the graphs in figure 4.13. In 4.13a and b we can see that for a short tube ($l = 20$) the theoretical net force overestimates the simulation forces most of the time and by significant amounts. While looking at 4.13c we see that for a long tube ($l = 100$) the theoretical curves still overestimate the simulation points during the first part of chamber evacuation, but it

Row	N,D,L,l	t_{ch}	t_{total}	t_{det}	t_{FPT}
1	200,7,10,20	1617.2	1804.6	2190.7	1396.4
2	200,7,10,45	2703.5	3381.3	5130.1	3673.8
3	200,8,10,20	2445.3	2641.0	3158.5	1853.9
4	300,7,10,20	1355.1	1506.8	2190.7	1396.4
5	300,7,10,100	4875.0	7496.2	12509	10838
6	400,7,10,20	1333.9	1502.3	2190.7	1396.4
7	400,7,10,150	7345.5	12884	20479	19837

Table 4.1: Ejection times for different parameters we chose. The tube diameter was fixed at $d = 3$ ($d_{eff} = d - 0.2031$). N, D, L, and l are respectively the total number of monomers, the diameter of the chamber, the length of the chamber, and the length of the tube (a correction of 0.2031 is to be understood for all size parameters except l). t_{ch} is the chamber evacuation time from the simulations, t_{total} is the total ejection time from simulations, t_{det} is the theoretical total ejection time as derived in sections 2.5.2 and 2.5.3 (n_{ch}^f was chosen to be 1 in eq. 2.18), and t_{FPT} is the first passage time from eq. 2.32.

is largely underestimated for $n_{ch} < 20$, and this may be the cause that t_{FPT} is larger than the simulation time for long tubes. We can also see from figure 4.13d during the final stages of the ejection (N¡50), the theoretical force overestimates the simulation results. In addition we can see that the assumption that the force is constant during tube ejection does not seem to be accurate. The discrepancy may also be due to the fact that for shorter tubes the friction in the chamber is comparable to the friction in the tube as can be seen from figure 4.14.

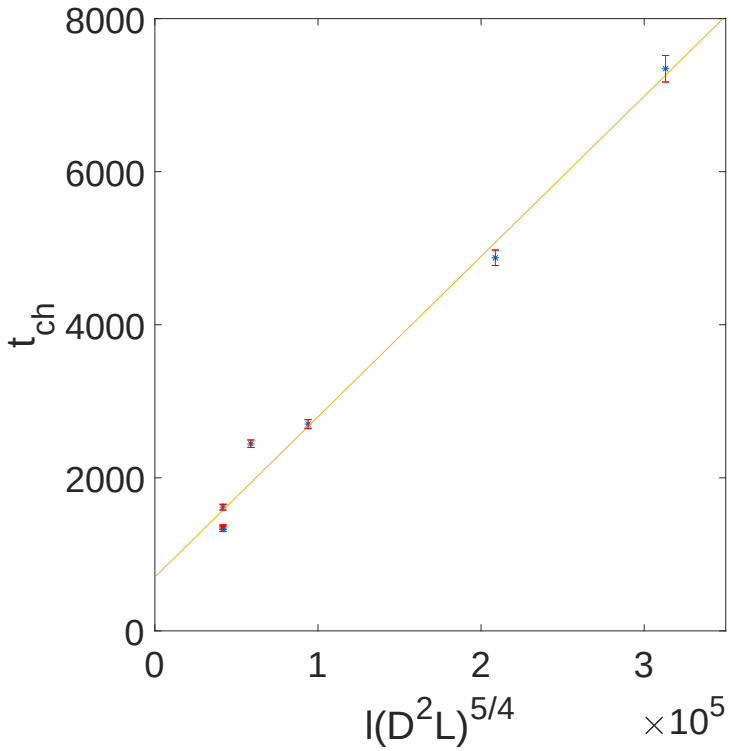

Figure 4.11: t_{ch} versus the variable $l(D^2L)^{5/4}$ as derived from eq. 2.18 shows good agreement (the line is a linear fit).

There is no obvious reason why the ejection time for row 1 is longer than for rows 4 and 6. Looking at the forces for all three we see that they are almost the same. Also the initial numbers of monomers in the chamber and the tube are the same for all three.

The CPU time for the passive ejection simulations was between 0.1 hours for short tubes and chains, and 2 hours for long chains and long tubes.

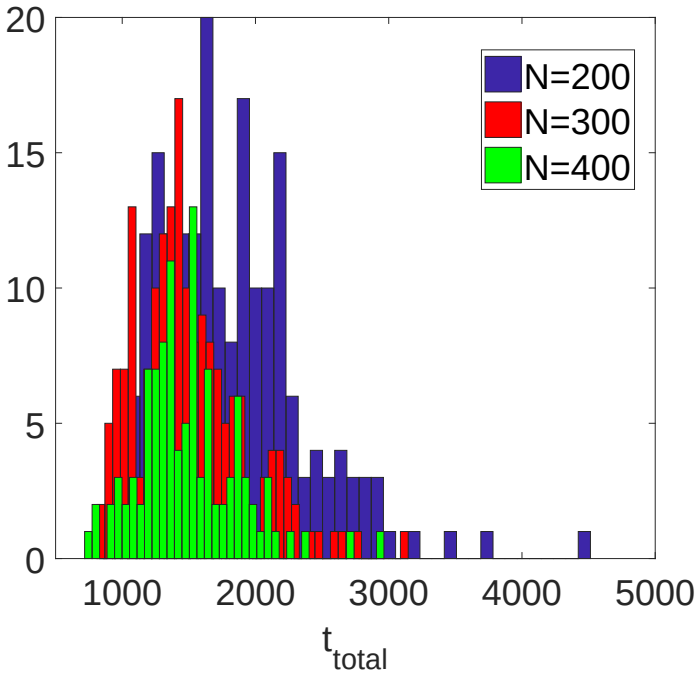

Figure 4.12: Histograms of the total ejection times for $D = 7,\ L = 10,\ d = 3,$ and $l = 20$. The histograms for $N = 300$ and $N = 400$ are similar but they are different from that of $N = 200$.

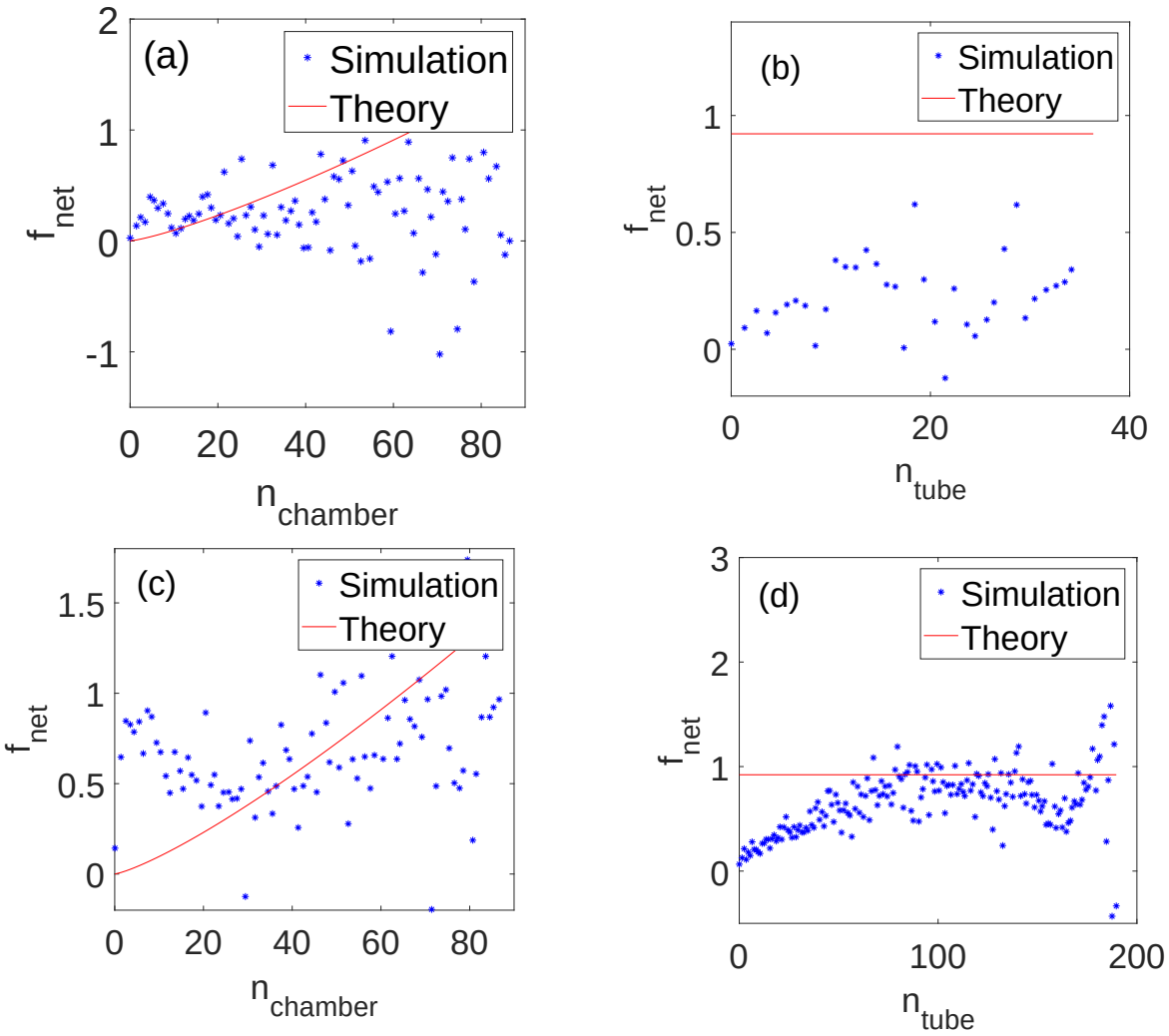

Figure 4.13: Net force in the axial direction versus (a) the number of monomers in the chamber for $N = 300,\ l = 20$; (b) the number of monomers in the tube for $N = 300,\ l = 20$; (c) the number of monomers in the chamber for $N = 300,\ l = 100$; (d) the number of monomers in the tube for $N = 300,\ l = 100$.

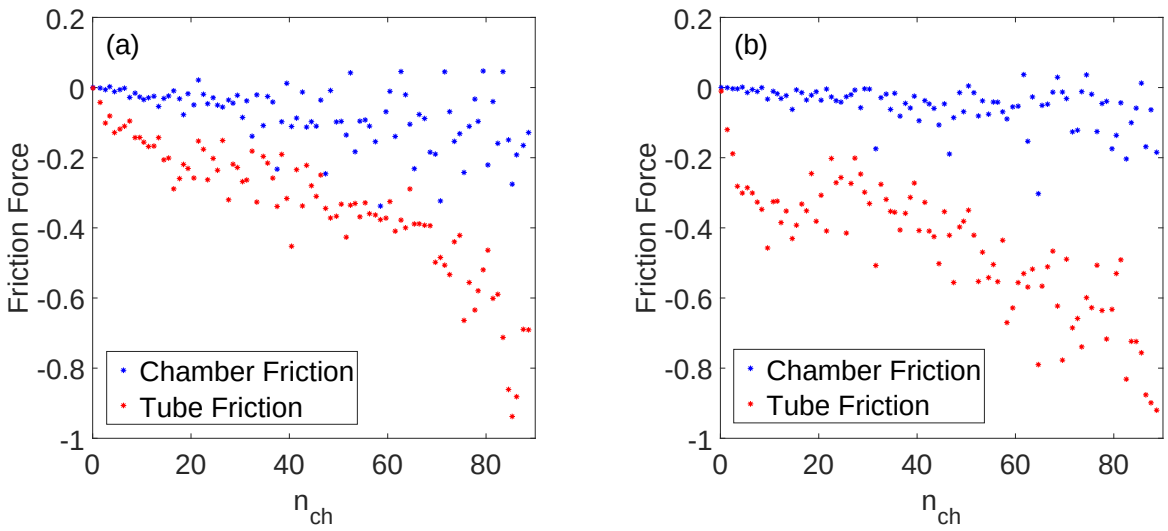

Figure 4.14: Net friction in the tube and in the chamber for $N = 300$, (a) $l = 20$, (b) $l = 100$

Chapter 5

Conclusion and Future Work

In this work, we investigated different aspects of a polymer chain in a cylindrically symmetric capsid. In order to study the equilibrium properties of the chain, we derived expressions for volume averaged local pressure in cylindrical coordinates(section 2.4, appendix B), and developed an algorithm for the proper calculation of the pressure (section 3.4, appendix C). The acquired profiles show some gradients which at first glance seems inconsistent with the established mechanical equilibrium, but upon a closer look at the condition for mechanical equilibrium this may not be the case. Future work will include testing whether the condition for mechanical equilibrium $\nabla \cdot \boldsymbol{P}(\boldsymbol{x}) = 0$ is satisfied by the equilibrated chain, in addition to the calculation of the pressure distribution of the out of equilibrium chain as an additional way to characterize the ejection.

When it comes to the ejection from the capsid, we were able to derive a deterministic formula for the time needed to evacuate the chamber using simple energetic arguments and scaling formulas (section 2.5.2). Combined with the analysis in [7] from which eq. 2.23 was derived, an overestimation of the total

ejection time from the capsid can be obtained. To get better estimates for the ejection time, it turns out we need to include the effects of diffusion into our analysis. This was done through the use of the Fokker-Planck equation, and we derived an equation for the ejection time (section 2.6). The final result is in the form of an integral that is evaluated numerically. In the derivation of the theoretical expressions we made certain assumptions that were tested in section 4.2.4. The results show that the theory needs some modification, and this was also confirmed in the results from the passive ejection simulations, where it was found that the expressions for the chemical potentials used in the theoretical derivations both under and overestimate of the net forces (section 4.2.5). Finding a better description of the ejection stages is necessary for a complete understanding of the problem. Future work will involve separating the process into more accurate stages described by accurate expression for the relevant chemical potentials.

Appendix A

Abbreviations

MD	Molecular Dynamics Simulations
MC	Monte Carlo Simulations
FPE	Fokker-Planck Equation
LJ	Lennard-Jones
WCA	Weeks-Chandler-Andersen (repulsive part of the Lennard-Jones potential)

Appendix B

Averaged Pressure Tensor Elements

Definitions:

- The intermolecular forces are radial and directed along the line joining the particles. They are assumed to have the form

$$\boldsymbol{F}(\boldsymbol{r}_{ij}) = f(\boldsymbol{r}_{ij})\boldsymbol{r}_{ji}$$

  where

$$\boldsymbol{r}_{ji} = \boldsymbol{r}_i - \boldsymbol{r}_j$$

  and $\boldsymbol{r}_i$ is expressed in cylindrical coordinates $\boldsymbol{r}_i = (x_i, \rho_i, \phi_i)$

- x, ρ, and ϕ subscripts refer to the different components in cylindrical coordinates where the axial direction is along the x-axis, for example:

$$\begin{aligned}
\boldsymbol{F}_x(\boldsymbol{r}_{ij}) &= f(\boldsymbol{r}_{ij})(xi - xj)\hat{x} \\
\boldsymbol{F}_\rho(\boldsymbol{r}_{ij}) &= f(\boldsymbol{r}_{ij})[(yi - yj)\hat{y} + (zi - zj)\hat{z}] \\
\boldsymbol{F}_\phi(\boldsymbol{r}_{ij}) &= \boldsymbol{F}_x \times \boldsymbol{F}_\rho
\end{aligned}$$

and in the same way

$$\begin{aligned}\boldsymbol{x}_i &= x_i\hat{x}\\ \boldsymbol{\rho}_i &= y_i\hat{y} + z_i\hat{z}\\ \boldsymbol{\phi}_i &= \boldsymbol{x}_i \times \boldsymbol{\rho}_i\end{aligned}$$

- The basis vectors are

$$\begin{aligned}\hat{x}\\ \hat{\rho} &= \frac{y\hat{y} + z\hat{z}}{\sqrt{y^2+z^2}}\\ \hat{\phi} &= \hat{x} \times \hat{\rho}\end{aligned}$$

Here we write down the expressions for the interaction part of the pressure tensor. All expression are averaged over volume elements (see section 2.4)

$$P_{ij}^{\alpha\beta}(\mathcal{V}) = \frac{1}{|\mathcal{V}|}\int_{\mathcal{V}} d\boldsymbol{x} \int_{C_{ji}} (d\boldsymbol{l}\cdot\boldsymbol{e}_\alpha)(\boldsymbol{F}\cdot\boldsymbol{e}_{\boldsymbol{\beta}})\delta(\boldsymbol{x}-\boldsymbol{l}) \tag{B.1}$$

For the integration contour we choose the Irving-Kirkwood contour [21] defined by

$$\boldsymbol{l} = \boldsymbol{r}_j + \lambda\boldsymbol{r}_{ji} \tag{B.2}$$
$$d\boldsymbol{l} = \boldsymbol{r}_{ji}d\lambda \tag{B.3}$$

where $0 \leq \lambda \leq 1$. This way the integral becomes

$$P_{ij}^{\alpha\beta}(\mathcal{V}) = \frac{1}{|\mathcal{V}|}\int_{\mathcal{V}} d\boldsymbol{x} \int d\lambda(\boldsymbol{r}_{ji}\cdot\boldsymbol{e}_\alpha)(\boldsymbol{F}\cdot\boldsymbol{e}_{\boldsymbol{\beta}})\delta(\boldsymbol{x}-\boldsymbol{l}) \tag{B.4}$$

If we evaluate the integral over the volume first, we can eliminate the $\delta(\boldsymbol{x}-\boldsymbol{l})$ by setting $\boldsymbol{x} = \boldsymbol{l}$, or replacing any relevant $\boldsymbol{x}$ component by the corresponding $\boldsymbol{l}$ component, for example for $P_{ij}^{\rho\rho}$

$$P_{ij}^{\rho\rho} = \frac{1}{|\mathcal{V}|}\int_{\mathcal{V}} d\boldsymbol{x} \int d\lambda(\boldsymbol{r}_{ji}\cdot\hat{\boldsymbol{\rho}})(\boldsymbol{F}\cdot\hat{\boldsymbol{\rho}})\delta(\boldsymbol{x}-\boldsymbol{l})$$

Using $\hat{\rho} = \frac{y\hat{y}+z\hat{z}}{\sqrt{y^2+z^2}}$ and $\boldsymbol{F}(\boldsymbol{r}_{ij}) = f(\boldsymbol{r}_{ij})\boldsymbol{r}_{ji}$ we get

$$\begin{aligned} P_{ij}^{\rho\rho} &= \frac{1}{|\mathcal{V}|}\int_{\mathcal{V}} d\boldsymbol{x} \int d\lambda \left(\frac{y_{ji}y + z_{ji}z}{\sqrt{y^2+z^2}}\right)\left(\frac{f y_{ji}y + f z_{ji}z}{\sqrt{y^2+z^2}}\right)\delta(\boldsymbol{x}-\boldsymbol{l}) \\ &= \frac{f}{|\mathcal{V}|}\int_{\mathcal{V}} d\boldsymbol{x} \int_{\lambda_1}^{\lambda_2} d\lambda \left(\frac{y_{ji}y + z_{ji}z}{y^2+z^2}\right)^2 \delta(\boldsymbol{x}-\boldsymbol{l}) \\ &= \frac{f}{|\mathcal{V}|}\int_{\lambda_1}^{\lambda_2} d\lambda \left(\frac{y_{ji}l_y + z_{ji}l_z}{l_y^2+l_z^2}\right)^2 \end{aligned} \tag{B.5}$$

where $l_y = y_j + \lambda y_{ji}$, and similarly for z. The integration bounds λ_1 and λ_2 define the segment of the contour that is inside the integration volume. The integrals were evaluated using computer algebra and yielded the following expressions for the different components:

$$P_{ij}^{xx} = \frac{f}{|\mathcal{V}|}(x_i - x_j)^2 \Big[\lambda\Big]_{\lambda_1}^{\lambda_2} \tag{B.6}$$

$$P_{ij}^{x\rho} = P_{ij}^{\rho x} = \frac{f}{|\mathcal{V}|}(x_i - x_j)\Big[\big|\boldsymbol{l}_\rho(\lambda)\big|\Big]_{\lambda_1}^{\lambda_2} \tag{B.7}$$

$$P_{ij}^{x\phi} = P_{ij}^{\phi x} = \frac{1}{|\mathcal{V}|}\frac{\boldsymbol{F_x}\cdot(\boldsymbol{\rho_j}\times\boldsymbol{\rho_i})}{|\rho_{ji}|}\Big[ln(|\rho_{ji}|\lambda + \big|\boldsymbol{l}_\rho(\lambda)\big| + \boldsymbol{\rho_j}\cdot\frac{\boldsymbol{\rho}_{ji}}{|\rho_{ji}|})\Big]_{\lambda_1}^{\lambda_2} \tag{B.8}$$

$$P_{ij}^{\rho\rho} = \frac{f}{|\mathcal{V}|}\left[\lambda|\rho_{ji}|^2 - |\boldsymbol{\rho_j}\times\boldsymbol{\rho_i}|arctan\left(\frac{\boldsymbol{\rho_{ji}}\cdot\boldsymbol{l_\rho}(\lambda)}{|\boldsymbol{\rho_j}\times\boldsymbol{\rho_i}|}\right)\right]_{\lambda_1}^{\lambda_2} \tag{B.9}$$

$$P_{ij}^{\rho\phi} = P_{ij}^{\phi\rho} = \frac{f}{|\mathcal{V}|}|\boldsymbol{\rho_j}\times\boldsymbol{\rho_i}|\Big[ln(|\boldsymbol{l_\rho}(\lambda)|^2)\Big]_{\lambda_1}^{\lambda_2} \tag{B.10}$$

$$P_{ij}^{\phi\phi} = \frac{f}{|\mathcal{V}|}|\boldsymbol{\rho_j}\times\boldsymbol{\rho_i}|\left[arctan\left(\frac{\boldsymbol{\rho_{ji}}\cdot\boldsymbol{l_\rho}(\lambda)}{|\boldsymbol{\rho_j}\times\boldsymbol{\rho_i}|}\right)\right]_{\lambda_1}^{\lambda_2} \tag{B.11}$$

Appendix C

Details of Pressure Calculation Algorithm

Here we detail the steps followed to calculate the interaction contribution to the pressure. The aim is to find the bounds λ_1 and λ_2 that define the part of the contour inside a particular slice.

1. Divide the capsid into slices of thickness dX axially and dR radially. The number of radial slices is $N_R = \frac{D/2}{dR}$, the number of axial slices is $N_x = \frac{x_{max}}{dX}$, and the total number of slices is $N_{total} = N_R \times N_x$. The slices are labeled from 1 to N_{total} moving radially first then in the axial direction. See for example figure C.1a.

2. From figure C.1b we can see that it is possible for the contour to pass from large radius to smaller and then back. To account for this we find the position along the contour with the minimal distance ρ from the axis. This is achieved by minimizing $|\boldsymbol{l}_\rho(\lambda)|^2$ with respect to λ.

$$|\boldsymbol{l}_\rho|^2 = y_j^2 + z_j^2 + 2\lambda(y_i y_j + z_i z_j - y_j^2 - z_j^2) + \lambda^2[(y_j - y_i)^2 + (z_j - z_i)^2]$$

$$\frac{d|\boldsymbol{l}_\rho|^2}{d\lambda} = 2\lambda[(y_j - y_i)^2 + (z_j - z_i)^2] + 2(y_i y_j + z_i z_j - y_j^2 - z_j^2) = 0$$

$$\lambda_{min} = \frac{(y_i y_j + z_i z_j - y_j^2 - z_j^2)}{(y_j - y_i)^2 + (z_j - z_i)^2}$$

   if $0 < \lambda_{min} < 1$ then the shortest distance is between the two monomers and we have double intersections. and we choose $\rho_{min} = |\boldsymbol{l}_\rho|(\lambda_{min})$. Otherwise we have only individual intersections and we take $\rho_{min} = min(\rho_i, \rho_j)$. We also need $\rho_{max} = max(\rho_i, \rho_j)$.

3. For a position (ρ, x), the indices of the radial and axial slice where the point belongs are:

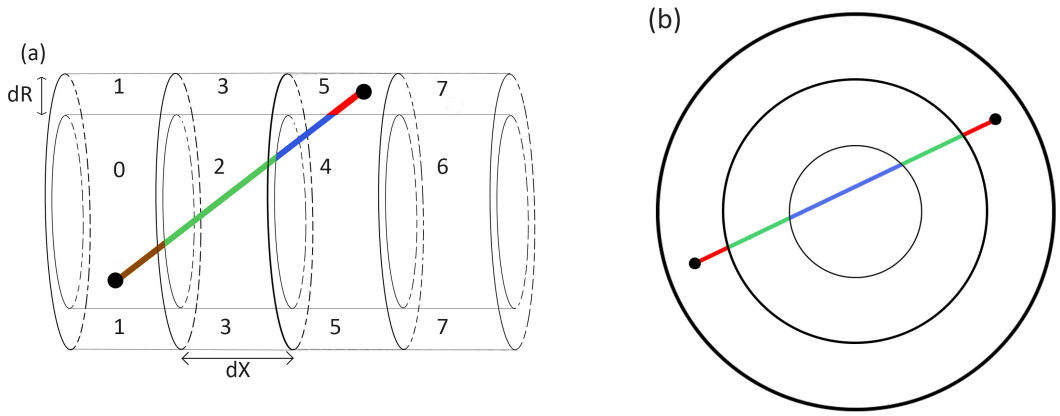

Figure C.1: Irving-Kirkwood contours used for the determination of the pressure. Colors indicate the slice to which the segments contribute. (a) The red segment contributes to slice 5, blue to slice 4, green to slice 2, and brown to slice 0. (b) It can be seen that as we move along the contour, it is possible to move to different slices then come back to the original one radially.

$$ind_\rho = floor\left(\frac{\rho}{D/2} \times N_r\right)$$

$$ind_x = floor\left(\frac{x}{x_{max}} \times N_x\right)$$

To get the total number of intersections with the boundaries of the slices we calculate $ind_{\rho_{min}}$, $ind_{\rho_{max}}$, ind_{ρ_i}, ind_{ρ_j}, ind_{x_i}, and ind_{x_j}. The number of intersections in the axial direction is

$$N_{intX} = |ind_{x_i} - ind_{x_j}|$$

and in the radial direction it is

$$N_{intR} = |ind_{\rho_i} - ind_{\rho_j}| + 2[min(ind_{\rho_i}, ind_{\rho_j}) - ind_{\rho_{min}}]$$

4. To find the λ's that correspond to intersections along the axial direction, loop over the number of intersections N_{intX} to find the intersection points

$$x_{int} = (ind_{x_j} - i + 1)dX \quad \text{if } x_i < x_j$$

$$x_{int} = (ind_{x_j} + i)dX \quad \text{if } x_i > x_j$$

where $= 1 \rightarrow N_{intX}$. Then find the corresponding λ using

$$\lambda_i = \frac{x_{int} - x_j}{x_i - x_j}$$

5. To find the λ's that correspond to intersections along the radial direction, loop from ind_{ρ_min} ind_{ρ_min} to find the intersection points

$$\rho_{int} = (j+1)dR$$

where $j = ind_{\rho_{min}} \rightarrow ind_{\rho_{max}} - 1$. Then to find the corresponding λ we need to solve the quadratic equation

$$\rho_{int}^2 = y_j^2 + z_j^2 + 2\lambda(y_i y_j + z_i z_j - y_j^2 - z_j^2) + \lambda^2[(y_j - y_i)^2 + (z_j - z_i)^2]$$

which is done easily using $\lambda_i = \frac{-b \pm \sqrt{\Delta}}{2a}$. If the solution is between 0 and 1 it is stored, otherwise it is not.

6. Sort the λ's found in steps 4 and 5 from smallest to largest.

7. Use the formulas in appendix B to calculate the value for the pressure, and add the contribution to the slice where $(\rho_{mid}, x_{mid}) = \left(\rho(\frac{\lambda_i + \lambda_{i+1}}{2}), x(\frac{\lambda_i + \lambda_{i+1}}{2})\right)$ is found whose index is

$$ind = ind_{\rho_{mid}} + ind_{x_{mid}} N_R$$

Bibliography

[1] "Wikimedia commons, english: Svg redraw of image:tevenphage.png, retrieved from https://commons.wikimedia.org/wiki/file:tevenphage.svg," 2008.

[2] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, vol. 1. Elsevier, 2001.

[3] A. D. Hershey and M. Chase, "Independent functions of viral protein and nucleic acid in growth of bacteriophage," *The Journal of general physiology*, vol. 36, no. 1, pp. 39–56, 1952.

[4] S. Mangenot, M. Hochrein, J. Rädler, and L. Letellier, "Real-time imaging of dna ejection from single phage particles," *Current Biology*, vol. 15, no. 5, pp. 430–435, 2005.

[5] M. De Frutos, L. Letellier, and E. Raspaud, "Dna ejection from bacteriophage t5: analysis of the kinetics and energetics," *Biophysical journal*, vol. 88, no. 2, pp. 1364–1370, 2005.

[6] M. M. Inamdar, W. M. Gelbart, and R. Phillips, "Dynamics of dna ejection from bacteriophage," *Biophysical journal*, vol. 91, no. 2, pp. 411–420, 2006.

[7] A. Milchev, L. Klushin, A. Skvortsov, and K. Binder, "Ejection of a polymer chain from a nanopore: Theory and computer experiment," *Macromolecules*, vol. 43, no. 16, pp. 6877–6885, 2010.

[8] I. Ali, D. Marenduzzo, and J. Yeomans, "Polymer packaging and ejection in viral capsids: shape matters," *Physical review letters*, vol. 96, no. 20, p. 208102, 2006.

[9] I. Ali, D. Marenduzzo, and J. Yeomans, "Ejection dynamics of polymeric chains from viral capsids: effect of solvent quality," *Biophysical journal*, vol. 94, no. 11, pp. 4159–4164, 2008.

[10] J. Kindt, S. Tzlil, A. Ben-Shaul, and W. M. Gelbart, "Dna packaging and ejection forces in bacteriophage," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13671–13674, 2001.

[11] M. Rubinstein and R. H. Colby, *Polymer physics*, vol. 23. Oxford university press New York, 2003.

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

[13] T. Nakamura, W. Shinoda, and T. Ikeshoji, "Novel numerical method for calculating the pressure tensor in spherical coordinates for molecular systems," *The Journal of chemical physics*, vol. 135, no. 9, p. 094106, 2011.

[14] C. Gardiner, *Stochastic methods*, vol. 4. springer Berlin, 2009.

[15] L. H. Tannoury, *Dynamic origins of entropic force: thermodynamic theory vs. molecular dynamics simulations*. PhD thesis, 2017.

[16] T. Nakamura, S. Kawamoto, and W. Shinoda, "Precise calculation of the local pressure tensor in cartesian and spherical coordinates in lammps," *Computer Physics Communications*, vol. 190, pp. 120–128, 2015.

[17] A. Milchev, "Effects of polymer stiffness on surface tension and pressure in confinement," *The Journal of Chemical Physics*, vol. 143, no. 6, p. 064701, 2015.

[18] L. I. Klushin, A. M. Skvortsov, H.-P. Hsu, and K. Binder, "Dragging a polymer chain into a nanotube and subsequent release," *Macromolecules*, vol. 41, no. 15, pp. 5890–5898, 2008.

[19] N. Korolev, L. Nordenskioeld, and A. P. Lyubartsev, "Multiscale coarse-grained modelling of chromatin components: Dna and the nucleosome," *Advances in colloid and interface science*, vol. 232, pp. 36–48, 2016.

[20] L. Verlet, "Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules," *Physical review*, vol. 159, no. 1, p. 98, 1967.

[21] J. Irving and J. G. Kirkwood, "The statistical mechanical theory of transport processes. iv. the equations of hydrodynamics," *The Journal of chemical physics*, vol. 18, no. 6, pp. 817–829, 1950.