TUDelft

Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Translocation of heterogeneous polymers through a nanopore

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

M.C.A. de Kever

Delft, the Netherlands
May 2011

TUDelft

MSc THESIS APPLIED MATHEMATICS

"Translocation of heterogeneous polymers through a nanopore"

M.C.A. de Kever

Delft University of Technology

Abstract: Translocating a chain of different beads through a very small pore can be used as a first step of modelling a DNA chain that passes through a nanopore. This translocation process offers a variety of possibilities in chemical and biological processes, for instance rapid DNA sequencing. In this thesis the chain is modelled as a polymer with different types of monomers as beads. The translocation dynamics of heterogeneous polymers through nanopores can be modelled using the LJ and FENE potentials and different interaction strengths between the monomers of the polymer and the pore. The translocation time gives important information of the chain sequence, depending on the length of the polymer. The waiting time is defined as the time a specific monomer stays inside the pore. This waiting time in particular gives useful results considering the chain sequence. Simulations reveal that the waiting time of the last monomer can define the type of monomer under consideration. Monomers with a high interaction with the pore will stay inside considerably longer. We found that from the average waiting time it is possible to retrieve the original sequence of the beads constituting the chain.

Keywords: Translocation, Heterogeneous Polymers, Nanopores, Translocation time, Waiting time, DNA sequencing

Daily supervisor

Dr. J.L.A. Dubbeldam

Responsible professor

Prof.dr.ir. A.W. Heemink

Other thesis committee members

Dr. N.H. Dekker

Dr. C. Kraaikamp

May 2011

Delft, the Netherlands

Preface

This report is a written survey of the Master Thesis part of the Master Applied Mathematics with a specialisation in Computer Science and Engineering at Delft University of Technology in the field of Mathematical Physics.

For a short summary of the content, I'd like to refer to the abstract. For the Matlab code, I'd like to refer to the appendix.

My special thanks go out to Dr. Johan Dubbeldam for supervising this project and his assistance during my research. He always had time for answering all my questions and revising my work.

I would also like to thank Wilco Ketelaar, operational executive at TFOZ, Guus Gussenhoven, forensic advisor FSO Zuid, and Dr. Jos Herbergs from TMFI for answering all my questions regarding forensic research and their enthousiasm. And Dr. N.H. Dekker for informing me about the research done at the faculty of Applied Physics concerning translocation.

In conclusion I would also like to thank Prof.dr.ir A.W. Heemink, Dr. N.H. Dekker and Dr. C. Kraaikamp for being member of my MSc Thesis Committee.

Thank you!

Delft, May 15, 2011

Margit de Kever

Contact: M.C.A.deKever@student.tudelft.nl

CONTENTS

List of Figures

CHAPTER 1

INTRODUCTION

When a forensic investigator enters a crime scene, he starts searching for evidence that could help reveal the truth of what has happened. The type of crime needs to be identified, this can vary from burglary to murder. He will search for fingerprints, footprints, gunpowderresidue and other types of evidence. Bloodspatter can be used for reconstructing a crime. One of the most powerful types of evidence are traces of DNA. The classic traces are blood, saliva and sperm. But there are also other DNA traces that can't be seen by the naked eye. These are called biological contact traces. For instance during a struggle cells from the skin of the offender can be left behind on the victim. The investigator collects all traces he can find. During this investigation he needs to determine what the significance of every trace is. Traces of blood can belong to the offender yet they can also belong to the victim. Finally he can end up with 100 to 200 different traces. According to the type of crime he selects a number of traces he thinks has the most value for revealing the truth. Usually this number is between 5 or 10 different traces. Of course traces of blood are easier to investigate than contact traces. The selection of traces will be sent to a forensic institute, for instance the NFI or TMFI. It depends on the type of crime how long it takes to investigate the traces. The fastest they can get results is in a day. Usual it takes about a week. In a big murder investigation it can take up to a month before everything is processed.

When the traces arrive at the forensic institute they will be thoroughly examined. This takes up to four days. During the first day the DNA is isolated from the trace and quantified. The second day the DNA fragments are multiplied using polymerase chain reactions. Using short tandem repeat analysis loci on the DNA from two or more samples are compared, where a locus is the specific location of a gene or DNA sequence on a chromosome. This takes about an hour for 16 samples. On day three the process of day two is repeated in order to be extra sure of the results. On the fourth day they report all the results of the examination.

The forensic investigator would like to have the results of the DNA examination as fast as possible in order to find the truth of what has happened as quickly as possible. The developments in this area are going fast. New methods lead towards solving cases faster and better. Could there be a faster way of examining the DNA?

1.1 DNA

DNA (Deoxyribonucleic acid) consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases. It is the sequence of these four bases along the backbone that encodes information.

1.2 TRANSPORTATION OF MOLECULES THROUGH MEMBRANES

Transportation of polymer molecules through a nanopore is the subject of a lot of research nowadays both experimental and theoretical. It is important in biological systems and technological applications.
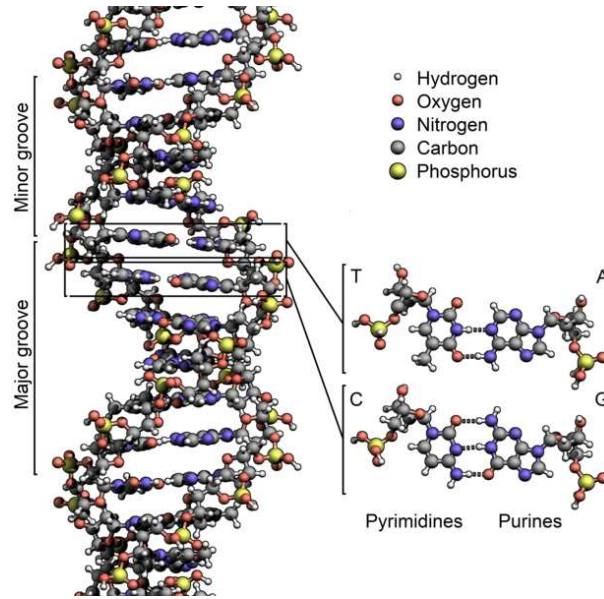
Figure 1.1: DNA, source: www.wikipedia.org

The translocation of biomolecules through channels in membranes is a very important process within or across biological cells for both delivery and signaling purposes. In (bio)chemistry, forced translocation is used in selection and purification of larger molecules, in medicine, it plays an important role in drug delivery. Also DNA and RNA translocation across nuclear pores, protein transport through membrane channels and virus injection into cells are important examples of phenomena in which long molecules traverse through a nanopore (Luo et al., 2008b), (Dekker, 2007).

All these processes of translocation have their potential technological applications, such as for example rapid DNA sequencing, gene therapy and controlled drug delivery. Actually intensive research may lead to a fast and efficient sequence detection. As mentioned in the introduction, now the sequencing of DNA happens through chemical processes, which are costly and time consuming. Rapid DNA sequencing using nanopores might help to speed up the process of examining the DNA traces the forensic investigator had found!

DNA is double stranded but for research purposes in the early stages of research processes, single stranded DNA sequences are often used. Our research is based on translocation of homogeneous polymers. It's important to note that inhomogeneities in the structure and interactions between polymer and other molecules might have a significant effect on translocation dynamics (Luo et al., 2008c).

1.3   Translocation through a nanopore

During translocation a chain moves through a very small pore. The chain consists of several elements, the most simple chain consists of monomers of a single species. This chain is called a homogeneous polymer. A monomer is a unit that can consist of severval base pares. For example you could say: 5kbp DNA can be modelled as 20-30 monomers. The next step is a chain that consists of two different types of monomers, say A and C. Such a polymer is called a heterogeneous polymer. The combination of the two sorts of monomers can attain different series. For example series like A_nC_n with A and C the two different monomers and n the number of monomers. But also $(A_nC_n)_m$, where m is the amount of times the series repeat itself is an interesting combination. And also a random combination could be possible. Both homogeneous and heterogeneous polymers will be investigated in this thesis, but we will restrict ourselves to two-species chains in this project. Research is still mainly based on two-species chains as even this simple model is still hard to understand. Though there are more possibilities for the chain. In the end the idea is to translocate a DNA or RNA chain containing four different monomers that are also interconnected to each other, they don't align in a straight chain.
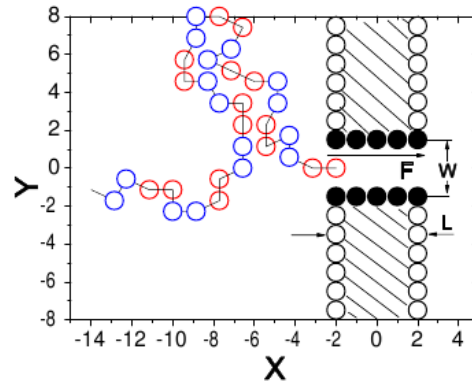
Figure 1.2: Schematic representation of a heterogeneous polymer ready for translocation

The small <u>pore</u> is actually called a <u>nanopore</u>. One can compare the size of a nanopore with the size of the diameter of a single hair divided by 80 000, that is a few nm. This kind of size is way too small to see with our eyes, but can be observed with AFM (atomic force microscope) images. The pore will interact with the chain. For instance, the A monomers of the chain can feel an attractive force when they are inside the pore. In order for a chain to traverse a pore, a force is applied to the part of the chain inside the pore. When the force is very small, the translocation process takes a long time, and translocation could actually fail. Therefore in experiments typically voltage differences of a few 100mV are applied across a pore, which implies a sufficiently strong driving force that usually guarantees succesful translocation.

Most theoretical research focuses on the dependence of the translocation time on relevant parameters. This way you can study how the translocation dynamics depend on the details of the DNA sequences. The parameters under consideration are for instance the polymer chain length N, pore length L and pore width W, driving force F, sequence and secondary structure and polymer-pore interactions. Also the dynamics of a single segment passing through the pore during translocation in an important issue. An interesting open key question is if DNA translocation through a nanopore can be used to determine the detailed sequence structure of the molecule (Luo et al., 2008b). In this thesis heterogeneous polymers with two types of monomers will be investigated. It turns out that we will be able to discriminate between A and C type monomers when the interaction strength with the pore for the different species differs a factor 3. This is a promising result and a step in the direction of actually sequencing a full DNA profile, although many challenges are still to be overcome.
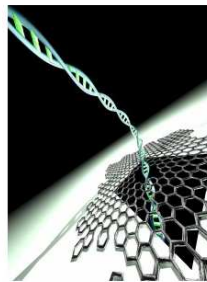
Figure 1.3: A sketch of a polymer traversing a graphene nanopore, source: ceesdekkerlab.tudelft.nl
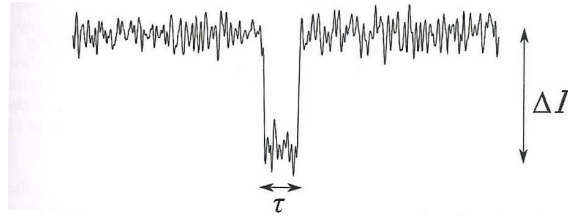
Figure 1.4: Schematic representation the blockade signal

1.4 An experiment

Let's start with a well-known typical translocation experiment. In physics the following routine gives interesting results. An electric field drives single stranded DNA and RNA molecules through a solid-state nanopore immersed in a salt-solution and the passage of each molecule is signalled by a blockade in the channel current. The length and size of the blockade can be used to characterize the kind of polymer that was under investigation. Comparing the translocation time distributions of polydeoxyadenylic acid $[(\text{poly}(dA)_{100})]$ and polydeoxycytidylic acid $[(\text{poly}(dC)_{100})]$ DNA molecules, it is found that the translocation time distribution of $[(\text{poly}(dA)_{100})]$ is much longer. This is due to the polymer-pore interactions on the dynamics of the biopolymer. There is a stronger attractive interaction of $[(\text{poly}(dA)_{100})]$ with the pore (Luo et al., 2008a).

1.5 Physics and Mathematics

The previous experiment shows how one is able to discriminate between two different polymers. In figure 1.5 there is an example of the blockade signal. There are two features we can see in this figure. How long the blockade is and how big it is. This blockade corresponds to a polymer that has filled the pore. This way the current through the pore is blocked. When the pore is empty the signal returns back to its original state. A lot of research in physics focusses on the actual translocation of the polymer through the nanopore, see for instance (Dekker, 2007), (Dekker and Kowalczyk, 2011), (Dekker et al.), (Krudde, 2009). Unfortunately in physics it is still impossible to sequence DNA with single base-pair resolution. Though in this thesis we will focus on the mathematics of the translocation process and we are able to discriminate each monomer. This is done by monitoring the waiting time for each monomer. That is the time it takes monomer n to reach the pore for the first time given that monomer $n-1$ is in the pore at time $t = 0$. This quantity is not accessible in an experiment which explains the difficulties underlying DNA-sequency using solid-state nanopores.

1.6 Assignment

In this thesis the translocation time distributions of polymer molecules through a nanopore in a membrane in two-dimensions will be investigated. In particular, the dependence of the translocation time on the sequence occuring in heterogeneous chains will be studied.

We will study polymers with different sequences for instance polymers with repeating blocks A_nC_n for various values of n. We will also investigate the distribution for the residence time of each monomer of the polymers inside the pore.

The experiment described shows a real case of translocation of a chain through a nanopore. But all investigations in this thesis, only deal with simulations of the translocation processes on a computer.

1.6.1 Expectations

We have certain expectations of the results. The translocation time should depend on the length of the polymer, the length of the seperate blocks and the orientation of which base enters the pore first. The final emptying of the pore will probably take the longest, as the final emptying of the pore corresponds to a rare

crossing of the barrier (Luo et al., 2008a). Translocation dynamics was observed in many experiments and computations to be anomalous. The contribution of the different processes to the translocation time will depend in different ways on the attraction strength of the pore (Luo et al., 2008c). The pattern exihibited by the waiting times of the individual bases and their periodicity can unambiguously determine the values of m, n and N respectively.

1.6.2 OVERVIEW OF THIS THESIS

So in this thesis a polymer chain will be modelled that translocates through a nanopore. This thesis is divided in two parts. In the first part we examine the mathematical concepts necessary in translocation and investigate the literature. In chapter 2 the basic topics of our statistical model will be explained. These topics are stochastic differential equations, the Smoluchowski equation, the Fokker-Planck equation and the Langevin equation. In chapter 3 we start with a basic approach of our problem using the Rouse model. In chapter 4 we use a few articles to explain the different types of models that explain the basics of translocation. In the second part we build a numerical two-dimensional model and use this to implement a program in Matlab that we use to produce some interesting result. The numerical model for the translocation will be built in three steps. The first step is modelling the movement of a homogeneous polymer chain through space. Using several forces both conservative and random the polymer chain attains an equilibrium state. This is done in chapter 5. In the second step a nanopore is inserted in the model. When the nanopore is inserted additional forces enter the model, these forces arise from the interaction between the monomers of the polymer with those of the nanopore. Also external forces can be added, for instance a voltage difference. This is done in chapter 6. In chapter 7 the third stage is explained in which a homogeneous polymer chain is replaced by a heterogeneous polymer. This last stage is the first step towards sequencing DNA. During the simulations of the steps in the previous three chapters we use an external force. In order to compare our results with those of the theory investigated in the first part of this thesis, we also investigate in chapter 8 a few cases where the external force is smaller.

After building the model, it will be implemented in Matlab. The program will simulate different polymer chains passing through a nanopore. The chains will differ in length and structure. Also the pore can attain different sizes. In the literature data is averaged over 2000 to 10^4 independent runs (Luo et al., 2007). In this thesis most results are averaged over 2000 independent runs. The explanation of the code can be found in chapter 9.

PART I

LITERATURE STUDY

Chapter 2

Basic Knowledge

The translocation process that we like to study is inherently stochastic. Therefore we will briefly review a number of concepts that will be used in modelling the translocation process.

2.1 Stochastic Differential Equations

The theory behind the model of this thesis is based on stochastic differential equations. In this chapter the basics will be explained. Let's start with a basic ordinary differential equation.

$$\frac{dx}{dt} = f(t,x) \tag{2.1}$$

This equation can also be written as

$$dx = f(t,x)dt \tag{2.2}$$

or

$$x(t) = x_0 + \int_{t_0}^{t} f(s,x(s))ds \tag{2.3}$$

where $x(t) = x(t|x_0,t_0)$ is a solution for the initial condition $x(t_0) = x_0$. Define l.i.m. as the limit in mean square sense for the limit of a stochastic series as:

$$\text{l.i.m.}_{t\to\infty} X_t = X. \tag{2.4}$$

This means

$$\lim_{t\to\infty} E(X_t - X)^2 = 0. \tag{2.5}$$

Now a stochastic differential equation can be defined with Brownian motion.

$$dX_t = f(t,X_t)dt + g(t,X_t)d\beta_t \tag{2.6}$$
$$X_{t_0} = X_0, \tag{2.7}$$

where $d\beta_t$ comes from the random Brownian motion. Define a random proces W as 'white noise' if it's mean is zero and the covariance of times s and t is the Dirac delta function

$$E[W(t)] = 0 \tag{2.8}$$
$$E[W(s)W(t)] = \delta(t-s). \tag{2.9}$$

White noise does not have a fysical meaning, only a mathematical ideal representation of the fysical world. Brownian motion cannot be differentiated, because white noise is discontinuous everywhere. The Fouriertransform of the covariance is given by

$$F(\omega) = \int_{\infty}^{\infty} e^{i\omega t}\delta(t)dt = 1. \tag{2.10}$$

White noise can be seen as a Gaussian process. Brownian motion is related to white noise by

$$\beta(t) = \int_{0}^{t} W(s)ds, \text{ for } 0 \leq t \leq \infty, \tag{2.11}$$

and can be written as

$$d\beta(t) = \beta(t + dt) - \beta(t) = W(t)dt. \tag{2.12}$$

Taking $\beta_{t_0} = \beta_0$ gives for the mean and covariance

$$E[\beta(t)] = \beta_0 \tag{2.13}$$
$$E[(\beta(t) - \beta_0)^2] = t - t_0. \tag{2.14}$$

For the stochastic differential equation (2.6)this gives

$$E[d\beta(t)] = 0 \tag{2.15}$$
$$E[(d\beta(t))^2] = dt. \tag{2.16}$$

All $d\beta(t_i)$ are independent of each other and the initial condition $\beta(t)$ and the distribution of $d\beta(t)$ is Gaussian. Now the stochastic differential equation can be interpreted as a stochastic integral and

$$X_t = X_0 + \int_{t_0}^{t} f(s, X_s)ds + \int_{t_0}^{t} g(s, X_s)d\beta_s, \text{ for } 0 \leq t_0 \leq t \leq \infty. \tag{2.17}$$

The second integral is a stochastic integral. This integral cannot be solved using regular calculus. For this we use Ito calculus. Assume $E[|g(t, x)|^2] < \infty$ for all $t \in [0, T]$ and $g(t, x)$ is continuous on the interval $[0, T]$ and independent of $d\beta_t$. Take $\mathcal{P}$ as a partition of $[0, T]$:

$$0 = b_0 < t_1 < ... < t_n = T \tag{2.18}$$

The integral can now be defined as

$$\int_{\mathcal{P}} g(t, x)d\beta_t = \text{l.i.m.}_{\Delta t \to \infty} \sum_{i=0}^{n-1} g(t_i', x)(\beta_{t_{i+1}} - \beta_{t_i}), \tag{2.19}$$

with $\Delta t = \max_i(t_{i+1} - t_i)$ and $t_i' \in [t_i, t_{i+1}]$. For Ito calculus* take $t_i' = t_i$.

2.2 SMOLUCHOWSKI EQUATION

In the previous chapter we encountered Brownian motion. The effect of Brownian motion appears most clearly in diffusion. Small particles initially placed at a certain point will spread out in time. Let $C(x, t)$ be the concentration at (x, t). Fick's law now states that if the concentration is not uniform, there exists a flux $q(x, t)$ which is proportional to the spatial gradient of the concentration

$$q(x, t) = -D\frac{\partial C}{\partial x}, \tag{2.20}$$

with D the diffusion constant. During a one-dimensional transport process, the mass is transported in the x-direction. A graphic explanation of this process is displayed in the figure.

Two parallel surfaces are placed orthogonal on the x-axis a distance Δx apart. Take the concentration $C(x, t)$ at (x, t). In between the surfaces there exists a mass of $C(x, t)\Delta x$. In time this mass changes. The difference in mass is given by

---

*For the method of Stratonovich take $t_i' = \frac{1}{2}(t_i + t_{i+1})$.
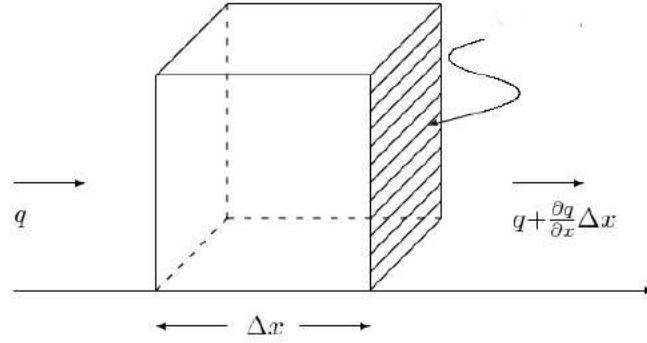
Figure 2.1: Visualisation transport process

$$\frac{\partial C}{\partial t}\Delta x. \tag{2.21}$$

This should be equal to the difference in flux through the two surfaces. This difference is given by

$$\frac{\partial q}{\partial x}\Delta x. \tag{2.22}$$

We found the so-called continuity equation

$$\frac{\partial C}{\partial t} = -\frac{\partial q}{\partial x}. \tag{2.23}$$

And using Fick's law, the diffusion equation is found

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2}. \tag{2.24}$$

If there is an external potential $U(x)$, Fick's law must be modified. The potential $U(x)$ exerts a force $F = -\frac{\partial U}{\partial x}$ on a particle, and gives a velocity v, which is linear in F so

$$v = -\frac{1}{\xi}\frac{\partial U}{\partial x}, \tag{2.25}$$

where ξ is the friction constant. This velocity of a particle gives an additional flux Cv, so that the total flux now turns into

$$q = -D\frac{\partial C}{\partial x} - \frac{C}{\xi}\frac{\partial U}{\partial x}. \tag{2.26}$$

In the equilibrium state the concentration is given by the Boltzmann distribution

$$C_{eq}(x) \propto e^{\left(-\frac{U(x)}{k_B T}\right)}. \tag{2.27}$$

Using this we can derive the Einstein relation

$$D = \frac{k_B T}{\xi}. \tag{2.28}$$

The Einstein relation is a special case of a more general theorem, the fluctuation dissipation theorem, which we will encounter later on.

Combining this new flux equation with the continuity equation gives the Smoluchowski equation

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\frac{1}{\xi}\left(k_B T\frac{\partial C}{\partial x} + C\frac{\partial U}{\partial x}\right). \tag{2.29}$$

2.3 The Fokker-Planck Equation

The name of the FPE comes from the work of Fokker(1914) and Planck(1917) which where both investigators of Brownian motion. The FPE is also called Kolmogorov's Equation or Smoluchowski Equation. The Fokker-Planck equation in one dimension is

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x,t)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x,t)f(x,t)]. \tag{2.30}$$

What makes the FPE interesting is that it is valid for conditional probability

$$f(x,t) = p(x,t|x_0 t_0), \tag{2.31}$$

for any initial x_0, t_0, with the initial condition $p(x,t_0|x_0,t_0) = \delta(x-x_0)$.

And the stochastic process described by a conditional probability satisfying the FPE is equivalent to the SDE (stochastic differential equation)

$$dx(t) = A[x(t),t]dt + \sqrt{B[x(t),t]}dW(t). \tag{2.32}$$

As the FPE is a second-order parabolic partial differential equation, an initial condition and two boundary conditions are needed. The intial condition is already mentioned a few lines above. For the boundary conditions several options are possible.

Using this Fokker-Planck equation we can connect the theory of the stochastic differential equations with the theory of the Smoluchowski equation.

2.4 Langevin equation

Unfortunately solving the Fokker-Planck equation is in general difficult as it is a general second order partial differential equation with unknown coefficients. An alternative description of Brownian motion is to study the equation of motion of the Brownian particle writing the random force $f(t)$ explicitly in the Langevin form:

$$\xi\frac{\partial x}{\partial t} = -\frac{\partial U}{\partial x} + f(t) \tag{2.33}$$

In this equation the random force is the sum of the forces working on a particle. To connect the Langevin equation to the Smoluchowski equation we assume that the distribution of $f(t)$ is Gaussian with:

$$\langle f(t)\rangle = 0 \tag{2.34}$$
$$\langle f(t)f(t')\rangle = 2\xi k_B T\delta(t-t') \tag{2.35}$$

This gives a one on one equivalence relation between the Langevin equation and the Fokker-Planck equation. In the appendix this equivalence will be proven for the case where $U = 0$.

CHAPTER 3

ROUSE MODEL

The Rouse model is the simplest model describing the dynamics of a simple chain in a solution. We will follow the line of thought as presented in (Doi and Edwards, 1986). In this model, the dynamics of the polymer are modeled using Brownian motion. For this model, two assumptions are made. There is no hydrodynamic interaction and no excluded volume interaction. Because there is no excluded volume interaction, the chains of this model are called ghost chains and the system can be viewed as beads connected with springs. The equation of motion of the beads can be discribed by the Langevin equation

$$\frac{\partial}{\partial t} r_n(t) = \sum_m H_{nm} \left[\frac{\partial U}{\partial r_m} + F_m^R \right], \tag{3.1}$$

where $H_{nm} = \frac{\delta_{nm}}{\xi}$, the mobility tensor and $U = \frac{k}{2} \sum_{n=2}^N (r_n - r_{n-1})^2$, the interaction potential, with $k = \frac{3k_B T}{b^2}$. Now the Langevin equation can be written as a linear equation for $n = 2, 3, ..., N - 1$

$$\xi \frac{dr_n}{dt} = -k(2r_n - r_{n+1} - r_{n-1}) + F_n^R \tag{3.2}$$

and for $n = 1, N$

$$\xi \frac{dr_1}{dt} = -k(r_1 - r_2) + F_1^R, \qquad \xi \frac{dr_N}{dt} = -k(r_N - r_{N-1}) + F_N^R. \tag{3.3}$$

The distribution of the random force F_n^R is Gaussian, with:

$$\langle F_n^R \rangle = 0 \tag{3.4}$$
$$\langle F_{n\alpha}^R F_{m\beta}^R \rangle = 2\xi k_B T \delta_{nm} \delta_{\alpha\beta} \delta(t - t') \tag{3.5}$$

By regarding n as a continous variable, and taking the continous limit, the previous equation can be rewritten as

$$\xi \frac{\partial r_n}{\partial t} = k \frac{\partial^2 r_n}{\partial n^2} + F_n^R. \tag{3.6}$$

The limits turn into

$$\frac{\partial r_n}{\partial n}\Big|_{n=0} = 0, \qquad \frac{\partial r_n}{\partial n}\Big|_{n=N} = 0. \tag{3.7}$$

with $r_0 = r_1$ and $r_{N+1} = r_N$. The moments of the random forces slightly change according to

$$\langle F_n^R \rangle = 0 \tag{3.8}$$
$$\langle F_{n\alpha}^R F_{m\beta}^R \rangle = 2\xi k_B T \delta(n - m) \delta_{\alpha\beta} \delta(t - t') \tag{3.9}$$

The Rouse model has a few characteristics. For instance, the mean square displacement of the position of the centre of mass is given by

$$\left\langle (R_{CM}(t) - R_{CM}(0))^2 \right\rangle = 6\frac{k_B T}{N\xi}t. \tag{3.10}$$

with the centre of mass given by

$$R_{CM} = \frac{1}{N} \int_0^N r_n dn. \tag{3.11}$$

The self diffusion constant of the centre of mass is defined by[*]

$$D_g = \lim_{t\to\infty} \frac{1}{2t} \left\langle (R_{CM}(t) - R_{CM}(0))^2 \right\rangle = \frac{k_B T}{N\xi}. \tag{3.12}$$

The characteristic relaxation time τ_r of a polymer can be defined as the longest relaxation time

$$\left\langle \vec{R_N}(t)\vec{R_N}(0) \right\rangle = e^{\left(\frac{-t}{\tau_r}\right)}, \tag{3.13}$$

where $R_N(t) = r_N(t) - r_0(t)$ is the end-to-end vector and

$$\tau_r = \frac{\xi N^2 b^2}{3\pi^2 k_B T}. \tag{3.14}$$

So for the Rouse model $\tau_r \propto N^2$. Another way to obtain N dependence of quantities like τ and the self diffusion coefficient D_g is by the theory of scaling arguments. According to (Doi and Edwards, 1986) scaling can be applied to the model. Take

$$N \to \frac{N}{\lambda} \quad b \to b\lambda^\nu \quad \xi \to \xi\lambda \tag{3.15}$$

and it can be shown that

$$D_g = \frac{k_B T\lambda}{N\xi\lambda} = \frac{k_B T}{N\xi} \propto N^{-1} \quad \tau_r = \frac{\xi\lambda N^x b^2 \lambda^{2\nu}}{3\pi^2 k_B T\lambda^x} = \frac{\xi N^x b^2 \lambda^{2\nu+1}}{3\pi^2 k_B T\lambda^x} \propto N^{1+2\nu}, \tag{3.16}$$

because we need to take $x = 1+2\nu$. The exponent ν is the Flory exponent given for $d = 3$ as $\nu = 0.588$ and at $d = 2$ as $v = 0.75$.

---

[*]In 3D instead of dividing by two, one needs to divide by six.

CHAPTER 4

TRANSLOCATION

We now have the tools to build a stochastic model of translocation. Though there are several phenomeno-logical theories that explain the basics of translocation. In this chapter the most relevant models will be briefly reviewed.

We start with the basic model of a chain of length N translocating from the cis-side to the trans-side of a membrane. We can calculate the free energy. This is also called the Helmholtz free energy, a thermodynamical state function, which measures the energy in a closed system at constant temperature and volume.

$$F = E - TS, \tag{4.1}$$

where F is the free energy, E is the internal energy of the system, T is the temperature and S is the entropy. This free energy is minimal when the system is in equilibrium. Now we make three assumptions. First, we can divide the chain in two parts which are both in equilibrium with their environment. Second, the cis- and transside environments are equal, there exists no driving force, which would bias the translocation. Though we will also examine the case with force. The third assumption is the most important. Translocation of a polymer chain , which is in reality a process in 3D, can be modelled as effectively taking place in one dimension, by only considering the length of the translocated part of the chain. This assumption greatly simplifies the analysis.

The entropy of both parts of the chain on the cis- and trans-sides can be calculated:

$$
\begin{aligned}
S_I &= k_B \log[(N-n)^{-\gamma} \mu^{N-n}] \\
S_{II} &= k_B \log[n^{-\gamma} \mu^n] \\
S_{tot} &= S_I + S_{II} \\
&= N k_B \log \mu - \gamma k_B \log[n(N-n)]
\end{aligned}
\tag{4.2}
$$

where μ is a connectivity constant which depends for example on dimension, n is the number of monomers on the trans-side, N is the length of the polymer and γ a constant for the steric hinderance by the wall.

Together these equations give (where we assume $E = 0$):
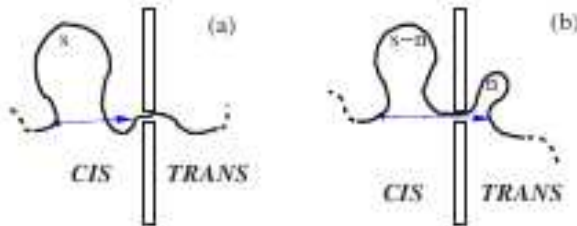
Figure 4.1: Schematic representation the translocation process from the cis-side to the trans-side

$$F = E - TS = -Nk_BT\log\mu + k_BT\gamma\log[n(N-n)].$$ (4.3)

There are two equivalent ways to go from this to a stochastic dynamical system, either by a stochastic differential equation or by the use of the Fokker-Planck equation. Assume that there is a single coordinate n which completely describes the stochastic process. The force acting on n is the entropic force given by F

$$F = -\frac{\partial}{\partial n}k_BT\gamma\ln[n(N-n)] = \frac{k_BT\gamma(N-2n)}{n(N-n)}$$ (4.4)

Inserting this in the Langevin equation 2.4 describes the process:

$$\xi\frac{dn}{dt} = -\frac{k_BT\gamma(N-2n)}{n(N-n)} + B(t)$$ (4.5)

The Langevin equation includes a Brownian force $B(t)$. For this force we have:
$$\langle B(t)\rangle = 0$$
$$\langle B(t_1)B(t_2)\rangle = \Gamma\delta(t_1 - t_2),$$
with $\Gamma = 2k_BT\xi$, see also the next chapter. From the Langevin equation the Ito SDE can be derived

$$\xi dn(t) = -\frac{k_BT\gamma(N-2n)}{n(N-n)}dt + \sqrt{\Gamma}dW(t)$$ (4.6)

Integrating equation (4) gives

$$n(t) = n(t_0) + \frac{1}{\xi}\int_{t_0}^{t}-\frac{k_BT\gamma(N-2n)}{n(N-n)}dt' + \frac{1}{\xi}\int_{t_0}^{t}\sqrt{\Gamma}dW(t)$$ (4.7)

taking $n(t_0) = 0$ it is given that

$$\langle n(\Delta t) - n_0\rangle = -\frac{k_BT\gamma(N-2n)}{\xi n(N-n)}\Delta t$$ (4.8)

$$\langle(n(\Delta t) - n_0)^2\rangle = \frac{\Gamma\Delta t}{\xi^2}$$ (4.9)

The Fokker-Planck equation corresponding to the Langevin equation equals:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}[\frac{k_BT\gamma(N-2n)}{n(N-n)\xi}P] + \frac{k_BT}{\xi}\frac{\partial^2}{\partial x^2}P$$ (4.10)

4.1   Comparing Numerical results with Theoretical results

The numerical results in this thesis can be compared to some theoretical results. For this we use the theory of the Fokker-Planck Equation. This was the equation that gives the time evolution of the probability density function for a system. For instance, we can calculate the first passage time using the FPE. In this section a short review is given of the corresponding Backward FPE, which we will further use to derive a differential equation for the mean first passage time, following (Gardiner, 2004).

The Backward Fokker-Planck Equation

$$\frac{\partial G}{\partial \omega} = -A\frac{\partial G}{\partial \phi} + B\frac{\partial^2 G}{\partial \phi^2}$$ (4.11)

In fact this is the Hermitian conjugate of the Fokker-Planck operator. Since the the Fokker-Planck equation is given by:

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial}{\partial x}[A(x,t)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x,t)f(x,t)] = L_{FP}G,$$ (4.12)

then

$$L_{FP}^\dagger = -A\frac{\partial G}{\partial x} + B\frac{\partial^2 G}{\partial x^2}. \tag{4.13}$$

4.1.1 MEAN FIRST PASSAGE TIME

The Fokker-Planck equation is very useful for calculating first passage times. The position of a monomer can be described by the FPE. The time it takes the monomer to remain in a certain region around the initial position can be calculated using the backward Fokker-Planck equation. Two different cases will be discussed.

Let's start will two absorbing barriers, so $a \leq x \leq b$. This means that when the monomer reaches a or b it is removed from the interval. At time $t = 0$ the monomer of interest is at position x_0. This x_0 is somewhere in the interval (a, b). The probability that at time t the particle is still inside this interval is:

$$\int_a^b p(x', t|x, 0)dx' = G(x, t) \tag{4.14}$$

The time that the particle leaves (a, b) is called T and rewritting the previous equation gives:

$$\mathrm{Prob}(T \geq t) = \int_a^b p(x', t|x, 0)dx' = G(x, t). \tag{4.15}$$

The system is time homogeneous therefore, $p(x', t|x, 0) = p(x', 0|x, -t)$ and now the backward FPE can be used:

$$\partial_t p(x', t|x, 0) = A(x)\partial_x p(x', t|x, 0) + \frac{1}{2}B(x)\partial_x^2 p(x', t|x, 0) \tag{4.16}$$

and hence $G(x, t)$ obeys the equation

$$\partial_t G(x, t) = A(x)\partial_x G(x, t) + \frac{1}{2}B(x)\partial_x^2 G(x, t). \tag{4.17}$$

The boundary conditions give the following information, as: $p(x', 0|x, 0) = \delta(x - x')$ it is known that

$$\begin{aligned} G(x, 0) &= 1 & a \leq x \leq b \\ &= 0 & \text{elsewhere} \end{aligned} \tag{4.18}$$

and if $x = a$ or $x = b$, the monomer is absorbed, so then

$$\mathrm{Prob}(T \geq t) = G(a, t) = G(b, t) = 0. \tag{4.19}$$

As $G(x, t)$ is the probability that $T \geq t$, the mean of any function of T can be defined as:

$$\langle f(T) \rangle = -\int_0^\infty f(t)dG(x, t), \tag{4.20}$$

and the mean first passage time is $\langle T \rangle = T(x) = -\int_0^\infty t\partial_t G(x, t)dt = \int_0^\infty G(x, t)dt$. For this last equality integration by parts is used and $\mathrm{Prob}(T \geq \infty) = 0$, $G(x, t)$ goes to zero fast enough.[*]
Using

$$\int_0^\infty \partial_t G(x, t)dt = G(x, \infty) - G(x, 0) = -1,$$

and integrating 4.17 over $(0, \infty)$ an ODE can be derived

$$A(x)\partial_t T(x) + \frac{1}{2}B(x)\partial_x^2 T(x) = -1, \tag{4.21}$$

with as boundary conditions $T(a) = T(b) = 0$. The solution for this ODE is

---

[*]As such $\langle T^n \rangle = T_n(x) = \int_0^\infty t^{n-1}G(x, t)dt$

$$T(x) = \frac{2\left[\left(\int_a^x \frac{dy}{\psi(y)}\right)\int_x^b \frac{dy'}{\psi(y')}\int_a^{y'} \frac{\psi(z)dz}{B(z)} - \left(\int_x^b \frac{dy}{\psi(y)}\right)\int_a^{y'} \frac{dy'}{\psi(y')}\int_a^{y'} \frac{\psi(z)dz}{B(z)}\right]}{\int_a^b \frac{dy}{\psi(y)}},$$ (4.22)

where

$$\psi(x) = \exp\left[\int_a^x \frac{2A(x')}{B(x')}dx'\right]$$ (4.23)

However, the problem under consideration in this thesis does not have two absorbing boundaries, but one absorbing and one reflecting boundary. Therefore slight adjustment needs to be made. One of the absorbing boundaries (a) is replaced by a reflecting one. The new boundary conditions are:

$$\partial_x G(a,t) = 0$$
$$G(b,t) = 0$$

Solving 4.21 again with the new boundary equations gives:

$$T(x) = 2\int_x^b \frac{dy}{\psi(y)}\int_a^y \frac{\psi(z)}{B(z)dz},$$ (4.24)

with $\psi(x)$ given by equation (4.23) and we assume $a < b$. In the numerical part of this thesis most simulations of translocation include an additional force exerted on the monomers in the nanopore. For instance in fysical experiments mentioned in the introduction a potential difference between the two sides of the pore exists. This model of translocation is called driven translocation. In this thesis one has $A = -U'(s)$, with $U(s) = \ln(s(1-s))$, $B = 2D$, $a = x_0$ and $b = x_0 + 1$. And the $\psi(x)$ results in

$$\begin{aligned}\psi(x) &= \exp\left[\int_a^x \frac{2A(x')}{B(x')}dx'\right] \\ &= \exp\left[\int_{x_0}^x \frac{-2U'(x')}{2D}dx'\right] \\ &= \exp\left[\frac{-1}{D}\ln(x(1-x)) - \ln(x_0(1-x_0))\right] \\ &= \left[\frac{x(1-x)}{x_0(1-x_0)}\right]^{\frac{-1}{D}}\end{aligned}$$

take $(x_0(1-x_0))^{\frac{1}{D}} = C$
and:

$$T(x) = 2\int_x^{x_0+1} \frac{(y(1-y))^{1/D}}{C_1}\int_{x_0}^y \frac{C_2}{2D(z(1-z))^{1/D}}dzdy$$

where C_1 and C_2 are defined like C.

When $U(s) = \ln(s(1-s))$ the external force is not included. To include the driving force used in this thesis take $U(s) = \ln(s(1-s)) + ks$ and find:

$$\begin{aligned}\psi(x) &= \exp\left[\int_a^x \frac{2A(x')}{B(x')}dx'\right] \\ &= \exp\left[\int_{x_0}^x \frac{-2U'(x')}{2D}dx'\right] \\ &= \exp\left[\frac{-1}{D}\left(\ln(x(1-x)) - \ln(x_0(1-x_0)) + k(x-x_0)\right)\right] \\ &= C_{exp}\left[\frac{x(1-x)}{x_0(1-x_0)}\right]^{\frac{-1}{D}}e^{\frac{-kx}{D}}\end{aligned}$$

take $(x_0(1-x_0))^{\frac{1}{D}} = C$
and:

$$T(x) = 2\int_x^{x_0+1} \frac{(y(1-y))^{1/D}e^{\frac{kx}{Dx_0}}}{C_1}\int_{x_0}^y \frac{C_2}{2D(z(1-z))^{1/D}e^{\frac{kz}{Dz_0}}}dzdy.$$

For an example of this mean first passage time see figure 4.2[†].

[†]The mean first passage time for the driven translocation does not differ significantly from the undriven translocation
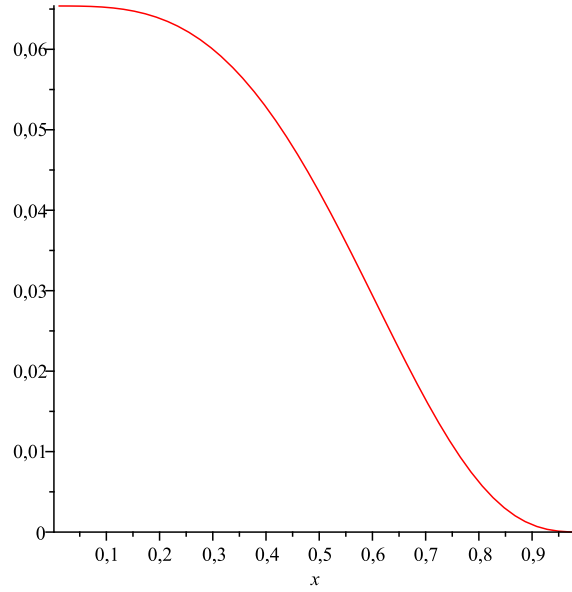
Figure 4.2: Example of mean first passage time starting from position x

4.2   SUNG AND PARK, AN EQUILIBRIUM MODEL

In (Sung and Park, 1996) the authors introduce a three-dimensional translocation model where the interaction of the polymer segments with the membrane is "considered to be only of steric origin". Here the modelling is also done in one-dimension. The authors obtain the free energy F and from that derive the equation of motion. As mentioned before, during translocation several configurations of the polymer are prohibited by the membrane. This leads to a reduction of the entropy of the polymer and an increase of its free energy. They take an ideal chain, where the first monomer is placed on a wall in the yz plane. They take as a boundary condition that all the other monomers cannot cross this wall. Now they obtain the probability of finding the final monomer at some place r, given the first one at r_0 on the surface $G(r, r_0; n)$ by using the method of images. This probability is proportional to the number of configurations in free space, which is the Gaussian distribution, minus the number of configurations for which the chain does cross the surface,

$$
\begin{aligned}
G(r, r_0; n) &= G_0(r, r_0; n) - G_0(r, -r_0; n) \\
&= \left[\frac{2\pi n b^2}{3}\right]^{-\frac{3}{2}} e^{\frac{-3(r-r_0)^2}{2nb^2}} - \left[\frac{2\pi n b^2}{3}\right]^{-\frac{3}{2}} e^{\frac{-3(r+r_0)^2}{2nb^2}} \\
&= \left[\frac{2\pi n b^2}{3}\right]^{-\frac{3}{2}} \frac{6x\epsilon}{nb^2} e^{\frac{-3(r)^2}{2nb^2}}
\end{aligned}
$$

where $r_0 = (\epsilon, 0, 0)$ and ϵ is an arbitrarily small distance of the anchored segment from the surface. The steric constraint factor of a chain is given as

$$
Z_S(n) = \int_{x>0} G(r, r_0; n) dr < 1,
\tag{4.25}
$$

and scales as $n^{-\frac{1}{2}}$. The partition function as the boundary condition is absent is given by

$$
Z_B(n) \sim e^{-\beta n \mu}
\tag{4.26}
$$

where $\beta = \frac{1}{k_B T}$ and μ is the chemical potential per segment defined by $\mu = \left[\frac{\partial F(n)}{\partial n}\right]_T$ in the limit $n \to \infty$. Now the free energy $F(n)$ given from the full partition function is given as:

$$
F(n) = -k_B T \times \ln[Z_S(n) Z(B(n))] = \frac{1}{2} k_B T \ln n + \mu n + C
\tag{4.27}
$$

where C is a constant term independent of n which arises from the integral in equation 4.2. At this stage the boundary condition is absent, since this is a static model. To reintroduce the boundary condition in the model, they decompose the chain into two independent chains each in the opposite half spaces "cis" and "trans". Now the total free energy of the system can be calculated as:

$$
\begin{aligned}
F(n) &= F(n) + F(N-n) \\
&= \tfrac{1}{2} k_B T \ln[n(N-n)] + n\Delta\mu + C
\end{aligned}
$$

The $\Delta\mu$ part of this last equation is the excess chemical potential per segment of the trans side relative to that of the cis side. In homogeneous media this contribution does not appear and exactly this leads to our entropic barrier. Namely, the free energy with $\Delta\mu = 0$ has a symmetric barrier which is of entropic origin, which, for a long chain, is nearly flat except near $n = 1$ or $n = N - 1$.

Furthermore Sung and Park treat the translocation process as a diffusive random process which is described by a Fokker-Planck equation for the probability distribution of n, $\frac{\partial}{\partial t}P(n,t) = L_{FP}(n)P(n,t)$, where the operator is

$$
L_{FP}(n) = \frac{1}{b^2}\left[\frac{\partial}{\partial n}D(n)e^{-\beta\mathcal{F}(n)}\left(\frac{\partial}{\partial n}e^{\beta\mathcal{F}(n)}\right)\right] \tag{4.28}
$$

$D(n)$ is the chain diffusivity constant and it is given by $D = \frac{k_B T}{\Gamma} \sim N^{-\nu}$, Γ is the chain friction proportional to $N^{-\nu}$. The exponent ν is 1 if the hydrodynamic interaction between the segments is neglected. This is the case in the Rouse model. And $\nu = \frac{1}{2}$ if it is included. This is called the Zimm model. Now the mean first passage time can be defined as:

$$
\tau = b^2 \int_1^{N-1}\left[\frac{1}{D(n)}e^{\beta\mathcal{F}(n)}\int_1^n\left(e^{-\beta\mathcal{F}(n')}\right)dn'\right]dn \tag{4.29}
$$

And assuming that D does not change, they conclude that the model predicts that the translocation time scales with N as $\tau = \frac{(Nb)^2}{2D} \sim (Nb)^{2+\nu}$, where b is the length of a Kuhn segment and these results can be checked experimentally. This is also the case for the one dimensional diffusion of a single Brownian particle. For instance take $D = \frac{1}{N}$ and $\nu = 1$ as is the case in the Rouse model, and find $\tau \sim N^3$. If Γ is independent of N, $\tau \sim N^2$. Typically the exponent ν is between 0 and 1.

4.3   Anomalous Diffusion model

The results in the previous section, where the translocation time is found to be $\tau \propto N^2$ seem quite right. Though while comparing this with the characteristic Rouse time $\tau_{\text{Rouse}} \propto N^{2\nu+1}$, with $\nu = 0.588$ for $d = 3$ and $\nu = 0.75$ for $d = 2$ something odd appears as $\tau_{\text{Rouse}} \gg \tau$. This means that it would take a free polymer longer to diffuse a distance of the order of its gyration radius than it takes to translocate.

To resolve this problem, in (Dubbeldam et al., 2007b) they have taken a different approach, to overcome the equilibrium hypothesis. First of all they consider the translocation process to be anomalous diffusion and they use a fold model (Metzler and Klafter, 2003), (Panja et al., 2007), (Kantor and Kardar, 2004). Second, they derive analytical results to compare them with Monte Carlo simulations.

Again they turn the three dimensional problem into a one-dimensional problem, by focussing on the translocation coordinate called folds, s. As usual, the chain of length n is assumed to be in equilibrium with a free energy of entropic nature. Then the translocation coordinate follows Brownian motion and the Smoluchowski equation is used with the free energy as external potential.

Again we write down the free energy of both the cis and trans side and the total free energy:

$$
\begin{aligned}
\frac{F^t(n)}{T} &= -n\ln\kappa - (\gamma_1 - 1)\ln n \\
\frac{F^c(n)}{T} &= -(s-n)\ln\kappa - (\gamma_1 - 1)\ln(s-n) \\
\frac{F(n)}{T} &= -s\ln\kappa - (\gamma_1 - 1)\ln[n(s-n)]
\end{aligned} \tag{4.30}
$$

where κ is the connectivity constant and γ_1 is the surface entropic exponent which equals $\gamma_1 \approx 0.945$. Now the entropic activation barrier can be calculated as

$$
\Delta E(s) = (F)(s/2) - (F)(1) = (1 - \gamma_1)T\ln s \tag{4.31}
$$

Now take the pure Rouse time $\tau \propto S^{2v+1}$ and add the activation energy, imposed by the membrane with the nanopore. Then the characteristic time scales as

$$t = \tau_R(s)e^{\Delta E(s)} \propto s^{2v+2-\gamma_1} \tag{4.32}$$

Estimating the mean-squared displacement of the s coordinate gives:

$$\langle s^2 \rangle \propto t^{\frac{2}{2v+2-\gamma_1}} \tag{4.33}$$

Now the anomalous diffusion law can be stated as: $\langle s^2 \rangle \propto t^a$ with $a = \frac{2}{2v+2-\gamma_1}$ and the translocation time equals $\tau \propto N^{2v+2-\gamma_1}$

As it is known from the literature that fractional diffusion gives rise to a subdiffusive dependence of $\langle s^2 \rangle$ on t, that is, $\langle s^2(t) \rangle \sim t^\alpha$ with $\alpha < 1$, this is the starting point for further analysis.

$$\frac{\partial}{\partial t}W(s,t) = {}_0D_t^{1-a}K_\alpha\frac{\partial^2}{\partial s^2}W(s,t), \tag{4.34}$$

where $W(s,t)$ is the probability diffusion function for having a segment s at time t in the pore and the <u>fractional Riemann-Liouville operator</u> ${}_0D_t^{1-\alpha}W(s,t) = \frac{1}{\Gamma(\alpha)}\frac{\partial}{\partial t}\int_0^t \frac{W(s,t')}{(t-t')^{1-\alpha}}dt'$.

The analysis is based on the more general analysis of the fractional Fokker-Planck equation. Solving the equation above is solving a so-called boundary value problem. Take as interval $0 \leq s \leq N$. There are two boundary conditions corresponding to the reflecting and absorbing case $\frac{\partial}{\partial s}W(s,t)_{s=0} = 0|$ and $W(s = N) = 0$. The initial distribution is $W(s, t = 0) = \delta(s - s_0)$. The solution is a sum over all eigenfunctions, called $\phi_n(s)$. And the solution can be represented as:

$$W(s,t) = \sum_0^\infty T_n(t)\phi_n(s) \tag{4.35}$$

Solving equation 4.3, one finds:

$$W(s,t) = \frac{2}{N}\sum_{n=0}^\infty \cos\left(\frac{(2n+1)\pi s_0}{2N}\right)\cos\left(\frac{(2n+1)\pi s}{2N}\right) \times E_\alpha\left[\left(\frac{-(2n+1)^2\pi^2}{4N^2}\right)K_\alpha t^\alpha\right] \tag{4.36}$$

where $\Gamma(a)$ is the Gamma function, and K_α is the generalized diffusion constant and $E_\alpha(x)$ is the Mittag-Leffler function defined as:

$$E_\alpha(x) = \sum_{k=0}^\infty \frac{x^k}{\Gamma(1+\alpha k)}. \tag{4.37}$$

The actual distribution of translocation times can be calculated using the first-passage time distribution (FPTD),

$$Q(s_0,t) = -\frac{d}{dt}\int_0^N W(s,t)ds \tag{4.38}$$

$$= \frac{\pi K_\alpha t^{\alpha-1}}{N^2}\sum_{n=0}^\infty (-1)^n(2n+1)\cos\left(\frac{(2n+1)\pi s_0}{2N}\right) \tag{4.39}$$

$$\times E_{\alpha,\alpha}\left[\left(-\frac{(2n+1)^2\pi^2}{4N^2}\right)K_\alpha t^\alpha\right] \tag{4.40}$$

where $E_{\alpha,\alpha}$ is the generalized Mittag-Leffler function given by

$$E_{\alpha,\alpha}(x) = \sum_{k=0}^\infty \frac{x^k}{\Gamma(\alpha+\alpha k)}. \tag{4.41}$$

If $t \to \infty$, $Q(t) \sim \frac{\alpha N^2}{2\Gamma(1-\alpha)K_\alpha t^{1+\alpha}}$. They compare the scaling of $Q(t)$ with t for large t with numerical simulations.

More information is contained in the moments of the probability distribution of the translocation coordinate $s(t)$. In particular the first and second monoments[‡].

$$
\langle s \rangle (t) = \frac{\int_0^N sW(s,t)ds}{\int_0^N W(s,t)ds} \tag{4.42}
$$

$$
= N - \frac{2N \sum_{n=0}^\infty \frac{1}{(2n+1)^2} E_\alpha \left(-\frac{(2n+1)^2\pi^2}{4N^2} K_\alpha t^\alpha \right)}{\pi \sum_{n=0}^\infty \frac{(-1)^n}{(2n+1)^2} E_\alpha \left(-\frac{(2n+1)^2\pi^2}{4N^2} K_\alpha t^\alpha \right)} \tag{4.43}
$$

$$
\langle s^2 \rangle (t) = \frac{\int_0^N s^2 W(s,t)ds}{\int_0^N W(s,t)ds} \tag{4.44}
$$

$$
= N^2 - \frac{8N^2 \sum_{n=0}^\infty \frac{(-1)^n}{(2n+1)^3} E_\alpha \left(-\frac{(2n+1)^2\pi^2}{4N^2} K_\alpha t^\alpha \right)}{\pi^2 \sum_{n=0}^\infty \frac{(-1)^n}{(2n+1)^2} E_\alpha \left(-\frac{(2n+1)^2\pi^2}{4N^2} K_\alpha t^\alpha \right)} \tag{4.45}
$$

$$
\langle s \rangle (t=0) = 0 \tag{4.46}
$$

$$
\langle s \rangle (t \to \infty) = \frac{N}{3} \tag{4.47}
$$

$$
\langle s^2 \rangle (0) - \langle s \rangle^2 (0) = 0 \tag{4.48}
$$

$$
\langle s^2 \rangle (t \to \infty) - \langle s \rangle^2 (t \to \infty) = \frac{N^2}{18} \tag{4.49}
$$

$$
\tag{4.50}
$$

4.4 Driven polymer translocation

In addition to the anomalous diffusion model, the same authors wrote an article, where the translocation of the polymers is driven (Dubbeldam et al., 2007a). In order to take this additional force into account, the equations of the previous chapter slightly differ. The Fractional Fokker-Planck equation 4.3 changes into:

$$
\frac{\partial}{\partial t} W(s,t) = {}_0D_t^{1-a} \left[\frac{\partial}{\partial s} \frac{U'(s)}{\xi_\alpha} + K_\alpha \frac{\partial^2}{\partial s^2} \right] W(s,t) \tag{4.51}
$$

where $\xi_\alpha = \frac{T}{K_\alpha}$ and $U(s)$ is the external field. This external field is a simple linear function given by $U(s) \geq ks$. The k can for instance be seen as the chemical potential difference $k = \Delta\mu$. The solution of the FFPE changes in:

$$
W(s,t) = \frac{2}{N} e^{\frac{f(n-n_0)}{2}} \sum_{n=0}^\infty \sin\left(\frac{n\pi s_0}{N}\right) \sin\left(\frac{n\pi s}{N}\right) \times E_\alpha \left(-\left(\frac{f^2}{4} + \frac{n^2\pi^2}{N^2}\right) K_\alpha t^\alpha \right) \tag{4.52}
$$

Using the calculations from (Lubensky and Nelson, 1999) the first passage time distribution is given by:

$$
Q(t) = \frac{\alpha}{4\pi^{\frac{1}{2}} ft} \left[\frac{\Gamma(1+\alpha)}{K_\alpha t^\alpha} \right]^{\frac{1}{2}} \left[\frac{N^2 \Gamma(1+\alpha)}{K_\alpha t^\alpha} - 2 \right] \times \exp\left(-\frac{\left[N - f\frac{K_\alpha t^\alpha}{\Gamma(1+\alpha)} \right]^2}{4\frac{K_\alpha t^\alpha}{\Gamma(1+\alpha)}} \right) \tag{4.53}
$$

And the translocation time scales as:

$$
\tau \propto \frac{1}{f} N^{2\nu+1-\gamma_1} \tag{4.54}
$$

[‡]The solution of the last equation is a correction of the original paper!

4.5 Heterogeneous Translocation

All the previous theory was based on homogeneous translocation. This means that all the beads of the translocating chains are the same. Though in DNA there are four different bases. There don't exist many theoretical papers on the translocation of heterogeneous chains. J. Dubbeldam and F. Redig explain the basics in (Dubbeldam and Redig, 2010).

They take two different beads A, B and these are randomly placed in a chain of N beads as $\eta(i)$ for $i = 1, ..., n$. The pore has different affinity for the molecules that belong to A and B. The environment on both sides of the membrane is the same. The position of the translocated part is measured in unit $\frac{1}{n}$. So X_t is always between 0 and 1, where $X_t = 0$ corresponds to the situation where the chain has fully translocated. Again the free energy $F(x)$ of the chain configuration can be calculated using the simple potential

$$F(x) = \ln(x(1-x)) \tag{4.55}$$

Now the continuous time random walk on $T_n := (0, \frac{1}{n}, .., \frac{n-1}{n}, 1)$ is defined for functions $f : T_n \to \mathbb{R}$ by

$$L_\eta f(x) = \left(I_{\eta(\lfloor nx \rfloor)=A} \tau_A + I_{\eta(\lfloor nx \rfloor)=B} \tau_B \right) L_n f(x) \tag{4.56}$$

with

$$
\begin{aligned}
L_n f(x) &= e^{-\frac{\beta}{2}\left(F\left(x+\frac{1}{n}\right)-F(x)\right)} \left(f\left(x + \frac{1}{n}\right) - f(x) \right) \\
&+ e^{-\frac{\beta}{2}\left(F\left(x-\frac{1}{n}\right)-F(x)\right)} \left(f\left(x - \frac{1}{n}\right) - f(x) \right)
\end{aligned}
\tag{4.57}
$$

for $x \notin 0, 1$ and $L_n f(0) = 0$, $L_n f(1) = f(1 - \frac{1}{n}) - f(1)$ and $\beta = \frac{1}{k_B t}$. The last term accounts for the entropy. This random walk is absorbing at $x = 0$ and becomes reflecting at $x = 1$. In the limit of $n \to \infty$ the random walk can by replaced by a diffusion process. The autors take

$$\psi(x, \eta) = \mathbb{E}_{\eta, x}^n(T_0) \tag{4.58}$$

with $\psi(x, \eta)$ the mean first passage time for a polymer starting at x and having a bead sequence given by η. And they take $L_\eta \psi = -\frac{1}{\chi}$, with $\chi = I_{\eta(\lfloor nx \rfloor)=A} \tau_A + I_{\eta(\lfloor nx \rfloor)=B} \tau_B$. Taking the average and using the ergodic theorem results in

$$\lim_{n \to \infty} L_n \psi = - \left(p_A \frac{1}{\tau_A} + p_B \frac{1}{\tau_B} \right). \tag{4.59}$$

Using Taylor expansion they take

$$\mathcal{L} f(x) := n^2 L_n f(x) = -\beta F'(x) f'(x) + f''(x) \tag{4.60}$$

This means that we have returned back to our familiar diffusion process. Again a the translocation time distribution comes down to

$$T(x) = c \int_0^x \int_t^1 \frac{(t(1-t))^\beta}{(u(1-u))^\beta} du dt$$

where

$$c = \left(p_A \frac{1}{\tau_A} + p_B \frac{1}{\tau_B} \right) \tag{4.61}$$

PART II

NUMERICAL RESULTS

CHAPTER 5

THE BASIC MODEL

5.1 THEORY

First the behaviour of a polymer moving in free space is investigated. The polymer chains themselves are modeled as bead-spring chains of Lennard-Jones (LJ) particles with finite extension nonlinear elastic (FENE) potential. The short-range (repulsive and possible attractive) LJ potential, models the excluded volume effects and Van der Waals interactions between all pairs of beads (Huopaniemi et al. (2007)). So it ensures that particles cannot be at the same place at the same time and they cannot overlap. The potential energy associated with the Lennard-Jones interaction (forces) is given by:

$$U_{\mathrm{LJ}}(r) = 4\varepsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6] + \varepsilon, \tag{5.1}$$

for $r \leq 2^{\frac{1}{6}}\sigma$ and 0 for $r > 2^{\frac{1}{6}}\sigma$. Here σ is the diameter of a monomer and ε is the depth of the potential.

The fact that the monomers are connected to each other in a chain is modeled by the so-called FENE potential, which reads

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2}kR_0^2 \ln(1 - \frac{r^2}{R_0^2}). \tag{5.2}$$

Here r is the distance between two consecutive monomers, k is the spring constant and R_0 is the maximum allowed separation between connected monomers.

Each monomer is subjected to conservative, frictional and random forces. For modelling the movement of the polymer, Langevin dynamics are used. This means that the behaviour of each monomer is governed by Newton's second law, which results in the following differential equation:

$$m\ddot{r}_i = -\nabla(U(r_i)) - \xi v_i + F_{\mathrm{ext}} + F_i^R, \tag{5.3}$$

for each monomer i. Here m is the mass of a monomer, r_i is the location of the monomer, $U(r_i)$ is the total potential that works on a monomer such that $-\nabla(U(r_i))$ is the force on the particles that is calculated from the LJ and FENE potentials, v_i is the velocity of each monomer, ξ is the frictional constant, F_{ext} an external force and F_i^R is a random force working on each monomer arising from collisions with the solvent molecules. The frictional force $F_i^F = -\xi v_i$ for each individual monomer comes from hydrodynamic drag.

The random force includes the Brownian motion of the monomers resulting from the random bombardment of solvent molecules and is represented by F_i^R and satisfies the fluctuation-dissipation theorem (Huopaniemi et al. (2007)). As the stochastic force $F_i^R(t)$ is assumed to be Gaussian distributed, we need only to consider the first two monoments of the distribution of $F_i^R(t)$, which are given as $\langle F_i^R(t) \rangle = 0$ and $\langle F_i^R(t_1)F_i^R(t_2) \rangle = \Gamma\delta(t_1 - t_2)$, with $\Gamma = 2k_BT\xi m$.
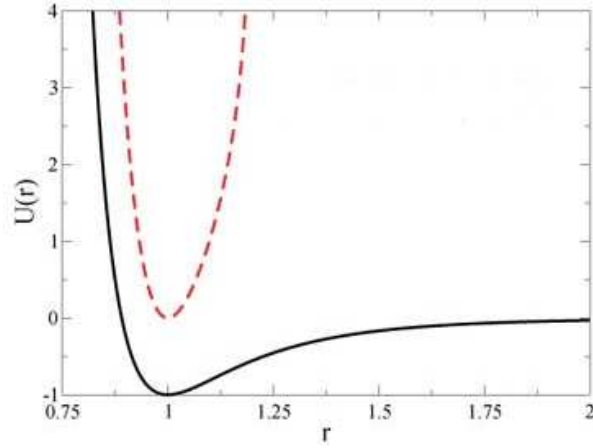
Figure 5.1: The Lennard-Jones potential (solid line) and the FENE potential (dotted line)

Derivation of the fluctuation dissipation theorem

Start with the Langevin equation:

$$\frac{dp}{dt} = -\xi p + f(t), \tag{5.4}$$

it is known that,

$$\langle f(t_1) f(t_2) \rangle = \Gamma \delta(t_1 - t_2),$$

and take,

$$p = mv.$$

Solving this equation gives:

$$p(t) = p_0 e^{-\xi t} + e^{-\xi t} \int_0^t e^{\xi t'} f(t') dt'. \tag{5.5}$$

When calculating $\langle p^2 \rangle$ one gets:

$$
\begin{aligned}
\langle p^2 \rangle &= p_0^2 e^{-2\xi t} + e^{-2\xi t} \int_0^t \int_0^{t''} e^{\xi t'} \langle f(t') f(t'') \rangle e^{\xi t''} dt' dt'' + 2p_0 e^{-2\xi t} \int_0^t e^{\xi t'} \langle f(t') \rangle dt' \quad (5.6) \\
&= p_0^2 e^{-2\xi t} + e^{-2\xi t} \int_0^t \int_0^{t'} e^{\xi t'} \Gamma \delta(t' - t'') e^{\xi t''} dt' dt'' + 2p_0 e^{-2\xi t} \int_0^t e^{\xi t'} \langle f(t') \rangle dt' \quad (5.7) \\
&= p_0^2 e^{-2\xi t} + e^{-2\xi t} \left[\Gamma \left(\int_0^t e^{2\xi t'} dt' \right) \right] + 2p_0 e^{-2\xi t} \int_0^t e^{\xi t'} \langle f(t') \rangle dt' \quad (5.8) \\
&= p_0^2 e^{-2\xi t} + e^{-2\xi t} [\Gamma [\frac{1}{2\xi} [e^{2\xi t'}]_0^t] + 2p_0 e^{-2\xi t} \int_0^t e^{\xi t'} \langle f(t') \rangle dt' \quad (5.9) \\
&= p_0^2 e^{-2\xi t} + \frac{\Gamma}{2\xi} [1 - e^{-2\xi t}] + 2p_0 e^{-2\xi t} \int_0^t e^{\xi t'} \langle f(t') \rangle dt' \quad (5.10) \\
&= \frac{\Gamma}{2\xi} + p_0 e^{-2\xi t} [p_0 - \frac{\Gamma}{2\xi p_0} + 2 \int_0^t e^{\xi t'} f(t') dt']. \quad (5.11)
\end{aligned}
$$

The second term in the last equation goes to zero for $t \to \infty$. And all that remains is

$$\langle p^2 \rangle = \frac{\Gamma}{2\xi}. \tag{5.12}$$

It was already defined that $p = mv$ so,

$$\langle p^2 \rangle = m^2 \langle v^2 \rangle = \frac{\Gamma}{2\xi}, \tag{5.13}$$

moreover by the equipartition theorem the velocity fluctuations are related by the temperature by

$$\frac{1}{2}m \langle v^2 \rangle = \frac{1}{2}k_B T \tag{5.14}$$

$$m \langle v^2 \rangle = k_B T. \tag{5.15}$$

Substituting gives:

$$mk_b T = \frac{\Gamma}{2\xi} \tag{5.16}$$

$$\Rightarrow \Gamma = 2\xi k_B T m. \tag{5.17}$$

5.1.1 ROUSE TIME

As there is no membrane present in this fase, the characteristic time for the cis-trans translocation can be calculated. This would be the pure Rouse time $t_R \propto s^{2\nu+1}$. Therefore this model is also called the Rouse model. The Rouse time is defined as the time that it takes for a free polymer to diffuse a distance of the order of its gyration radius (Dubbeldam et al., 2007b).

5.1.2 INTEGRATION IN TIME

For the integration in time of the model, the Verlet algorithm is used. This algorithm is based on a Taylor expansion of the coordinate of a monomer at time $t + \Delta t$ and $t - \Delta t$ about time t

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\partial^3 r}{\partial t^3}\frac{\Delta t^3}{3!} + \mathcal{O}(\Delta t^4), \tag{5.18}$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\partial^3 r}{\partial t^3}\frac{\Delta t^3}{3!} + \mathcal{O}(\Delta t^4). \tag{5.19}$$

Adding these two equations and subtracting $r(t - \Delta t)$ on both sides gives

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 + \mathcal{O}(\Delta t^4). \tag{5.20}$$

This new position is accurate to order Δt^4. The velocity is calculated from the trajectory and accurate to order Δt^2

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2). \tag{5.21}$$

The velocity will be used for calculating for instance the kinetic energy or temperature. This algorithm is in particular useful as it conserves the total linear momentum of the system and the total energy of the system.

5.2 SIMULATION

The model in this chapter is implemented in Matlab. In the following figures, some of the characteristics of the model are depicted for chains of length $n = 11, 21, 51, 101$. These can be checked using theoretical values. The slope of the centre of mass should be equal to

$$D = \frac{k_B T}{N\xi}. \tag{5.22}$$

And the radius of gyration, with $\nu = 0.75$ should go to

$$R_g^2 = \frac{1}{6} N^{2\nu} b^2. \tag{5.23}$$

The temperature T should stay constant at the prescribed value. The total linear momentum should be conserved, following

$$\frac{d}{dt} \sum_{i=1}^{N} m_i v_i(t) = 0. \tag{5.24}$$

Also the total energy (kinetic and potential energy) should be conserved. The kinetic energy should follow

$$E_{\text{kin}} = \sum_{i=1}^{N} \frac{m v_i^2}{2} = N k_b T. \tag{5.25}$$

All figures are based on 2000 simulations, except for those corresponding to the polymer of length $N = 101$ For this polymer only 380 simulations were used. From the figure we can see that the gyration radii go to the right equilibrium except for $N = 11$. In fact a polymer of length $N = 11$ is almost too short to give reliable results. The momentum for all polymers stays around zero with small fluctuations. The energy is approximately constant. In the figures 5.4, 5.5 we can see the log-log plot of the center of mass. We clearly observe a cross over behaviour which is consistent with the literature.

Because the gyration radius and the kinetic energy are related to the length of the polymer, its interesting to see how these quantities depend on N. In figures 5.6, 5.7 we can indeed see that the dependence can be fit using a power law. For the radius of gyration we find $\sim N^{1.54}$ and for the kinetic energy we find $\sim N^{1.01}$, which agrees with theoretical values.
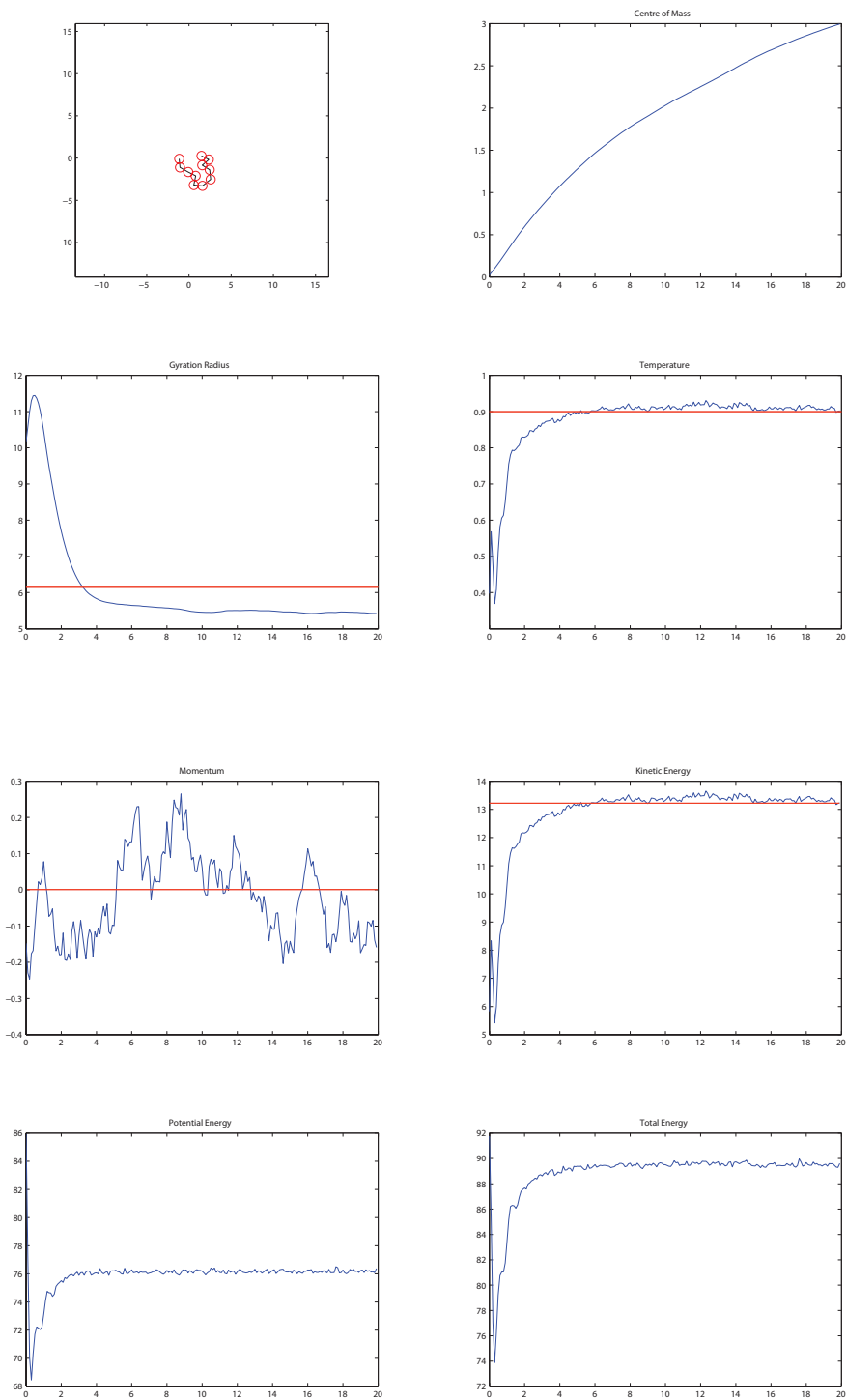
Figure 5.2: No Pore N=11.
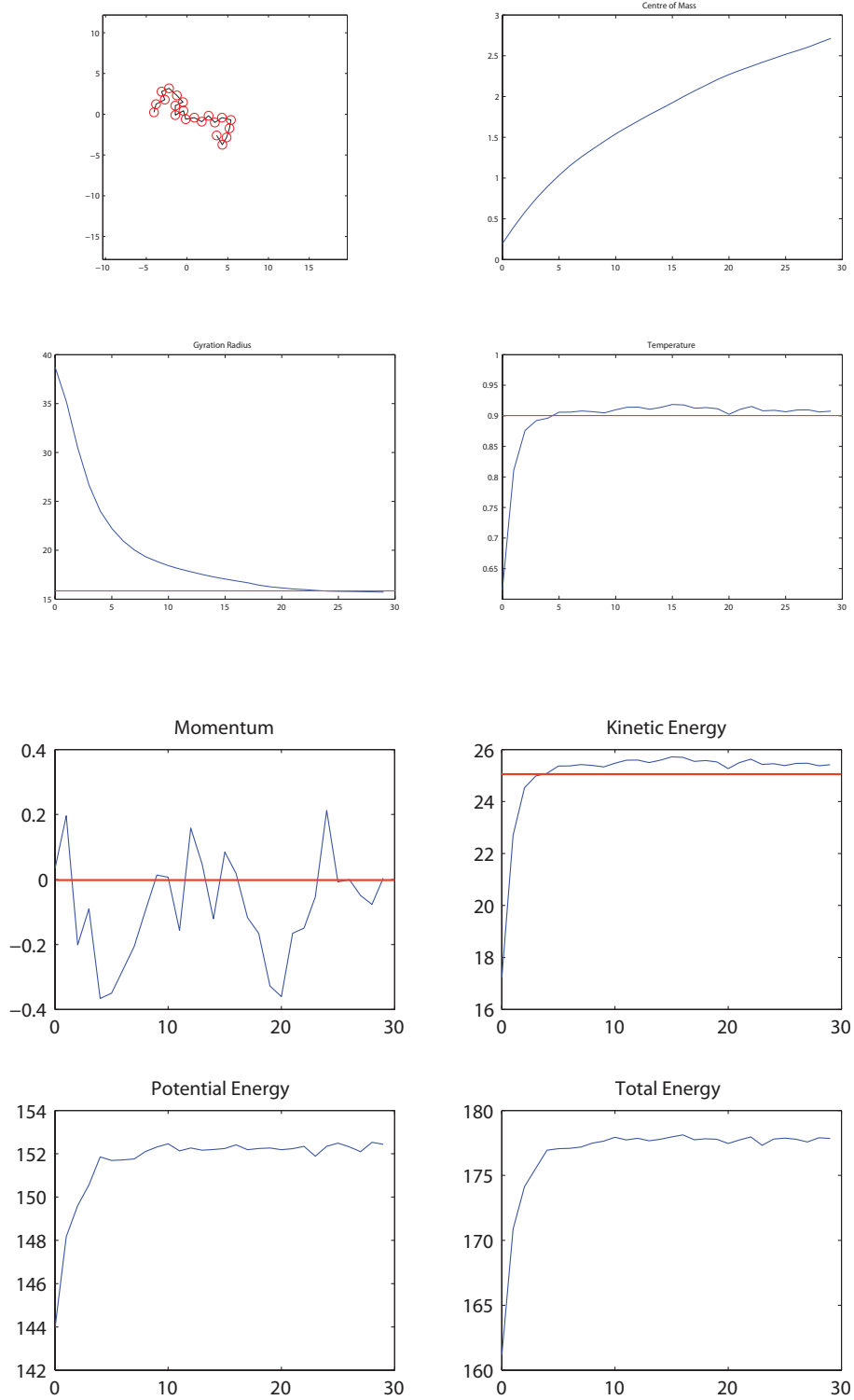
Figure 5.3: No Pore N=21.
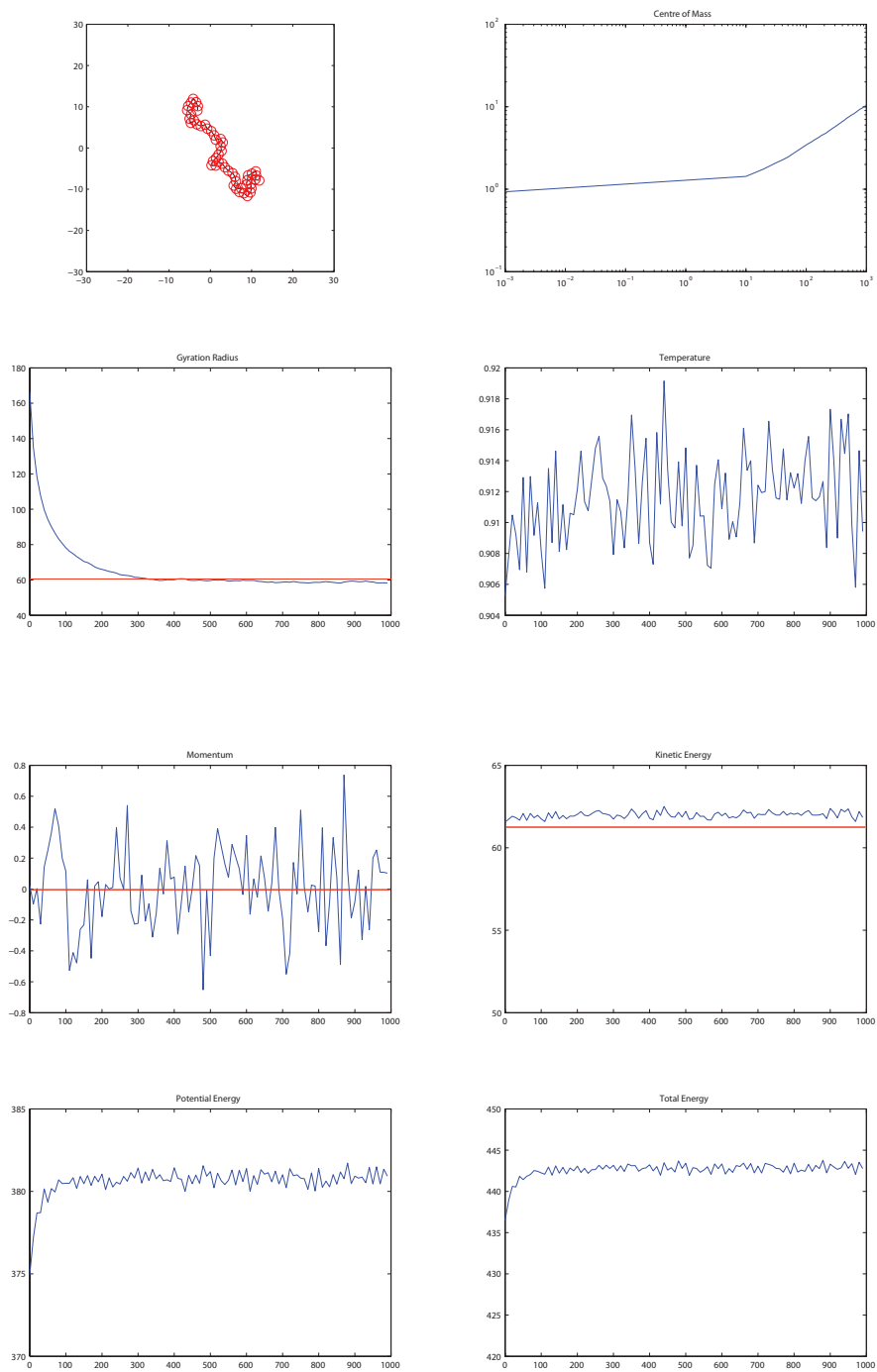
Figure 5.4: No Pore N=51.
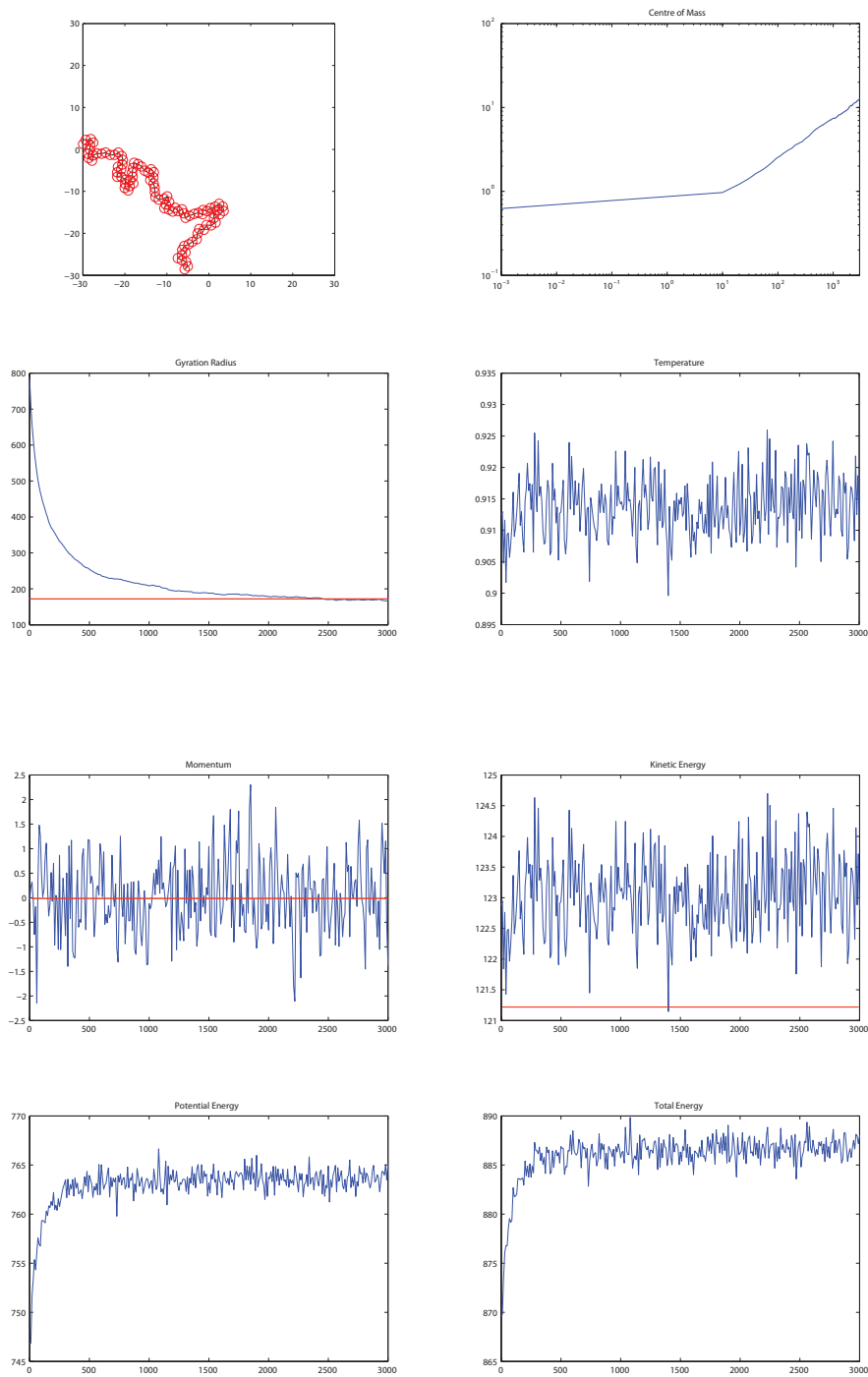
Figure 5.5: No Pore N=101.
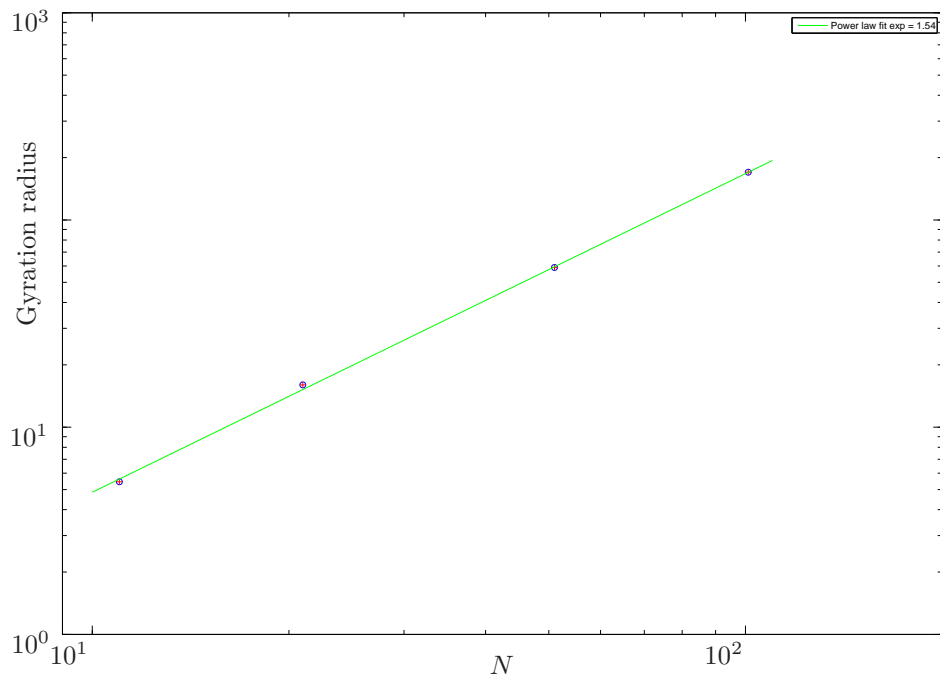
Figure 5.6: Radius of gyration vs N, Power law fit $\sim N^{1.54}$
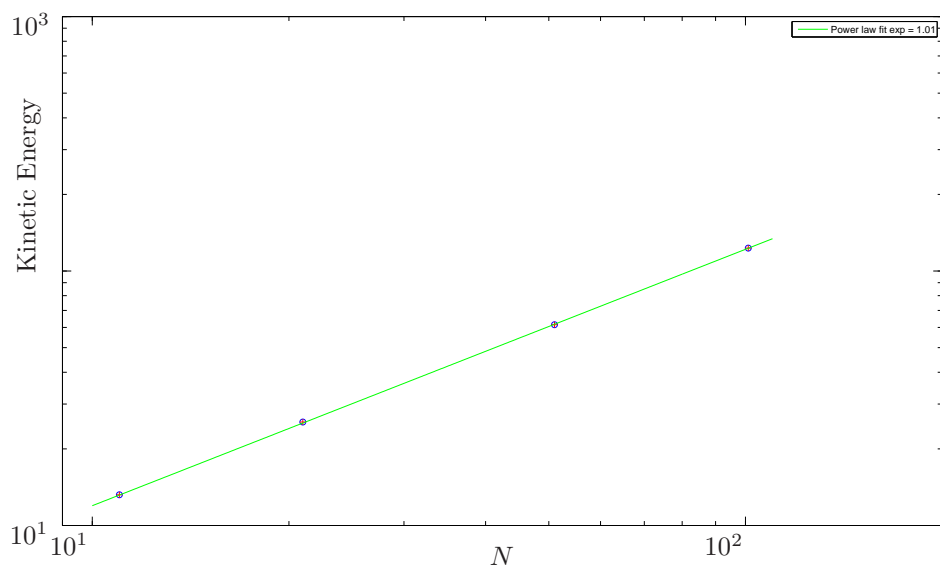
Figure 5.7: Kinetic energy vs N, Power law fit $\sim N^{1.01}$

CHAPTER 6

HOMOGENEOUS TRANSLOCATION

The next step in the model is the actual translocation of a homogeneous polymer through a nanopore. During translocation a few fases will be encountered. In the first fase a polymer chain enters the pore. Than the translocation process begins. This process is based on certain assumptions, the entropic barrier that needs to be crossed is the main one. But also the interaction between the monomers and the pore is very important. After translocation we want to be able to interpret the results of our simulations, so we need to introduce some typical translocation terms, like successful translocation and waiting time.

6.1 THE PORE

Let's start with expanding our model. A membrane is inserted in the space where the polymer chains from the previous chapter were moving freely around. Now besides the interaction between the monomers themselves, which was captured in the LJ potential and the FENE potential, the interaction with the membrane also needs to be modelled.

The model for the membrane is the following. A rigid wall of thickness L will be defined, with in its center a pore of width W, which is assumed to be small enough to allow only a single segment passage (Melchionna et al., 2009). The wall is formed by stationary particles at a distance of σ from each other. The segments cannot cross the membrane elsewhere.

The interaction of the wall-monomers with the monomers of the polymer is modelled by a Lennard-Jones potential as well. This potential is basically the same as the one that is used for the interaction between the monomers of the polymer. Except for the cut-off and the interaction strength. Here a cut-off of 2.5σ is used and an interaction strength of ε_{pm}. This means that the interaction can either be attractive or repulsive depending on the position of the monomers of the polymer from the monomers of the membrane (Luo et al., 2007)[*]. Let's recall the LJ potential

$$U_{LJ}(r) = 4\varepsilon_{pm}[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] + \varepsilon_{pm}, \tag{6.1}$$

The only thing that is changed, is the ε. This constant is changed into a constant dependent of the interaction strength of the particular pore-monomer it refers to. In this chapter about homogeneous translocation this constant is called ε_{pm}. Note that so far the model includes two different interaction strengths: ε_{mm}, the interaction strength between monomers from the polymer and ε_{pm}, the interaction strength between monomers from the polymer and monomers from the pore.

6.2 EXTERNAL FORCE

Another aspect that can be added to the polymer-pore model is an external force. This external force is defined as $F_{\text{ext}} = F\vec{x}_i$ where F is the external force strength exerted on the monomers in the pore, and $\vec{x}$ is a unit vector in the direction along the pore axis. The external force can, for instance be a driving force like a difference in the electric potentials at the different sides of the pore. This way the polymer

---

[*]This interaction is due to a chemical potential gradient (Dubbeldam et al., 2007b).
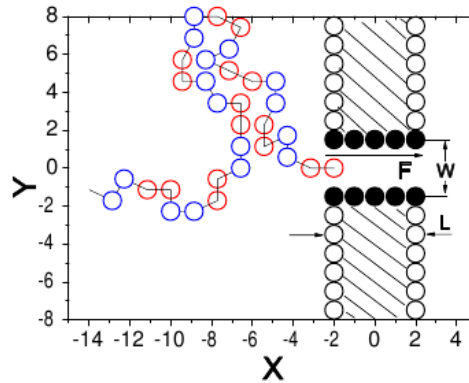
Figure 6.1: Schematic representation of a heterogeneous polymer ready for translocation

can be driven through the nanopore. In this thesis an external force is assumed. Adding this external force to the model accelerates the translocation. This way also simulations run faster.

Adding this external force involves changing the Langevin equation:

$$m\ddot{r}_i = -\nabla(U(r_i)) - \xi v_i + F_{\text{ext}} + F_i^R, \tag{6.2}$$

Besides the driving force used in this thesis other external forces are used in translocation research. An extra example, a pulling force, which is important for ATM (Atomic Force Microscope) experiments, is briefly explained in the addenda.

6.3   INITIAL CONFIGURATION

Initially the membrane with the pore is inserted in the model. A possibility for the next step is inserting a polymer chain in the free space and to wait for it to move towards the pore and for it to enter the pore and hope for it to translocate. This would be too time consuming. Though the success rate of a chain being captured by the pore is an interesting subject, it will not be discussed here, but some analysis on the process has been performed by Muthukumar and one can find a detailed discussion in (Muthukumar, 2010). But instead of inserting the monomer in the free space, it can also be placed with its first monomer at the opening of the pore. In this thesis results are based on a slightly different approach. In the initial configuration the centre monomer of the polymer is placed inside the pore. This prevents the polymer for leaving the chain without being fully translocated. For instance the polymer could have returned to the side of the polymer where it had entered. Let's call the side of the membrane where the polymer chain 'comes' from the cis-side and the side of the membrane the polymer translocates to the trans-side. In the matlab code it is assumed that the cis-side is to the left and the trans-side to the right of the membrane.

So the centre of the polymer is placed inside the pore. Still the translocation cannot be started yet. First the straight polymer that's inserted needs to relax and reach equilibrium. This can be done by holding the centre monomer and let the rest start to move. So the middle monomer is not allowed to move untill the rest of the polymer has attained an equilibrium position, by undergoing thermal collisions described by the Langevin thermostat (Luo et al., 2007). In figure 6.3 a polymer is displayed in a straight configuration and a equilibrium configuration. **Note:** The ends of the polymer in the straight configuration are not visible.

During simulations it is not necessary for doing this over and over again. When an equilibrium position is attained, it can be used for all following simulations. This is done by taking the equilibrium position, let it run for a little while longer and start the translocation process. This saves time and still allows for a different random position at the start of the translocation every time. In figures 6.4, 6.5, 6.6 and 6.7 some numerical results for free polymer chains are presented. For the polymer chains of lengths $11, 21, 51, 101$
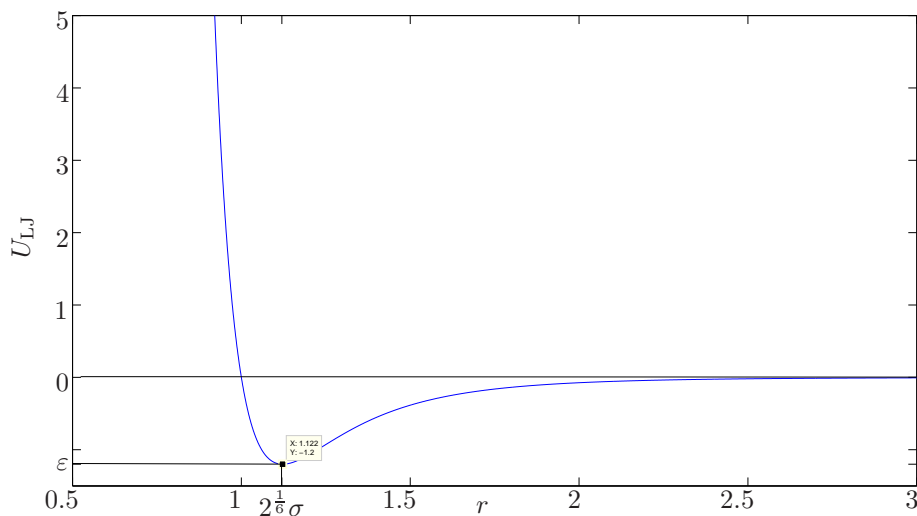
Figure 6.2: The LJ potential minus ε

a simulation is done in order to find equilibrium positions for them. This position is reached when the gyration radii at both sides of the pore have converged. The gyration radius is defined by

$$\text{Rg}^2 = \frac{1}{N} \sum_{k=1}^{N} \left\langle (\mathbf{x}_k - \mathbf{x}_{mean})^2 \right\rangle \tag{6.3}$$

with $\mathbf{x}_{mean} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k$. And recall that in equilibrium the radius of gyration should go to $\frac{1}{6} * N^{2\nu}$. In the figures below is displayed resp. the radius of gyration from simulations and theory and a final configuration.

6.4   SUCCESSFUL TRANSLOCATION

After attaining its equilibrium starting position, the centre monomer of the polymer chain is free to move as well. Now the translocation process starts. When the chain has moved out of the pore, the translocation is stopped.

Most theoretical research focusses on the translocation time of the polymer passing through the nanopore. Using this information, one would like to be able to link the translocation time with several aspects of the translocation of the polymer, for instance the length of the polymer or the external force exerted on the monomers in the pore. This way the simulation results can be compared with the results from actual experiments.

There are a multiple equivalent definitions of the translocation time. Some are based on theory others on practical results. For instance the translocation time is defined in (Luo et al., 2008a) as the time interval between the entrance of the first segment into the pore and the exit of the last segment and in (Sung and Park, 1996) as the mean first passage time for the barrier crossing, obtained from the Fokker-Planck equation. The first definition is more practical, where the second is based on theory. The last definition is based on the so-called entropic barrier. This barrier will be fully explained in the following section.

In this thesis, the polymer starts halfway the pore. So the translocation time is defined as the time interval that is elapsed between the moment the middle monomer is released and the exit of the last bead. This way successful translocations are very likely to take place. And adding the external force to the model translocation will always be succesful. After translocation of many polymers the data will be averaged to find the translocation time. The average is usually taken over 2000 runs with different initial

equilibrium positions. When the data is averaged over a different number of runs, this will be stated specifically.

In real experiments translocation is not always successful. To get some insight in this phenomenon, the translocation probability P_{trans} can be calculated, as the fraction of runs leading to succesful translocation for given initial conditions. Also in experiments the measured blockage time consists of both the time that a polymer succesfully translocated, as the time a polymer took to enter the pore and return back to the cis-side of the pore. For this purpose the residence time τ_r is defined as the weighted sum of these two (Luo et al. (2007)). Note that in case of forcing this amount to the same thing.

The total translocation time can be split up in three different process. The initial filling of the pore, the transfer of the polymer from the cis to the trans-side and the final emptying of the pore (Luo et al., 2008a), Luo et al. (2007). All these processes have their own features considering translocation time. As mentioned above, the initial filling of the pore will not be investigated here. But in the next chapter the final emptying of the pore will reveal interesting properties of the polymer chains under consideration. A global impression of the three stages of translocation is shown in 6.8.

6.4.1   Waiting time

The ultimate goal of the translocation processes is sometimes considered to be to sequence DNA. This implies that individual monomers are to be distinguished while traversing the pore. To be able to test whether it would in principle be possible to discriminate different monomers by the time it takes each monomer to travel through the pore we introduce the so-called waiting time. So besides investigating the translocation time it is also interesting to investingate this. The waiting time (also called residence time of the monomers) of monomer s is defined as the average time between the events that the monomer s and the monomer $s+1$ exit the pore for the first time. This gives some insight in how long each monomer stays inside the pore. In particular the difference between the three different processes mentioned above could be perceived, in particular, the emptying of the pore is visible for monomers that have a high interaction strength with the pore. Or, as will be done in the next chapter, a difference between monomers with different interaction strengths with the pore. To be able to differentiate between different monomers would take a huge step in future research of sequencing DNA. And as will be shown in the next chapter, this seems to actually be possible! Unfortunately this is not possible yet in real experiments, as was mentioned in the introduction.

6.5   Entropic barrier

What happens during translocation? Because the polymer chain is forced to go through the nanopore and unable to pass the membrane, certain configurations are impossible. This leads to an entropic energy barrier (Luo et al., 2007). For a succesful translocation it is necessary that the polymer crosses this entropic energy barrier. This entropic activation barrier is defined as $\Delta E(s)$. Because the barier is present, is slows down the transition rate. This was also encountered in chapter 4. Remember in (Sung and Park, 1996) the translocation seen as a stochastic diffusion process crossing the free energy barrier was calculated from the chain configuration partition function. The total free energy is:

$$\mathcal{F}(n) = \frac{1}{2} k_B T \ln[n(N-n)] + C. \tag{6.4}$$

So $\Delta E(N) = \mathcal{F}(\frac{N}{2}) - \mathcal{F}(1)^\dagger$. The characteristic time then scales as $t(s) = t_R \exp[\Delta E(s)] \propto s^{2\nu+2-\gamma_1}$ (Dubbeldam et al., 2007b). In figure 6.5 the free energy is visualised as a function of the translocation coordinate and the entropic energy barrier is depicted.

---

†In (Luo et al., 2007) the energy difference is calculated differently as they say refering to the final emptying of the pore: 'This is due to the activated nature of the translocation process with a free energy difference of $\Delta \tilde{F} = L[\varepsilon_{pm} - \frac{F}{2} - f(N)]$ between the final and the initial state'.

6.6 RESULTS

On the following pages the results of the translocation of polymers of length $N = 11, 21, 5, 101$ are displayed. For the lengths $N = 11, 21, 51$ the translocation time and waiting time distribution is displayed. The translocation time distributions look like Gaussians with heavy right tails. The waiting time distributions reveal properties of the translocation process. First there is a start up time. The polymer needs to adjust to its new dynamics. Than the process will speed up. At the end of the translocation process the emptying of the pore takes a bit longer. The pore is attracted to the polymer and won't let go of it. Unfortunately in the waiting time distributions in this chapter and the following chapters some unusual peaks exist. It is unknown where they come from, but they might be explained by the configuration of the polymer at the cis-side of the membrane. When the polymer is pulled through the pore, the part of the polymer of the cis-side might get strangled up. It will take some time to unwind before the next monomer can translocate. As the peaks happen mostly at the same place, it could also be a numerical error. In figure 6.6 the translocation times are plotted against the length of the polymers and can be fitted using a power fit law. In figure 6.12 the translocation and waiting time distribution of a polymer of length $N = 101$ is shown in detail. We can see that indeed a Gaussian fits to the data. The heavy right tail can be fitted using Exponential, Double exponential and a Power law fit. **Note:** the mode is the value that occurs most frequently in a data set or a probability distribution.
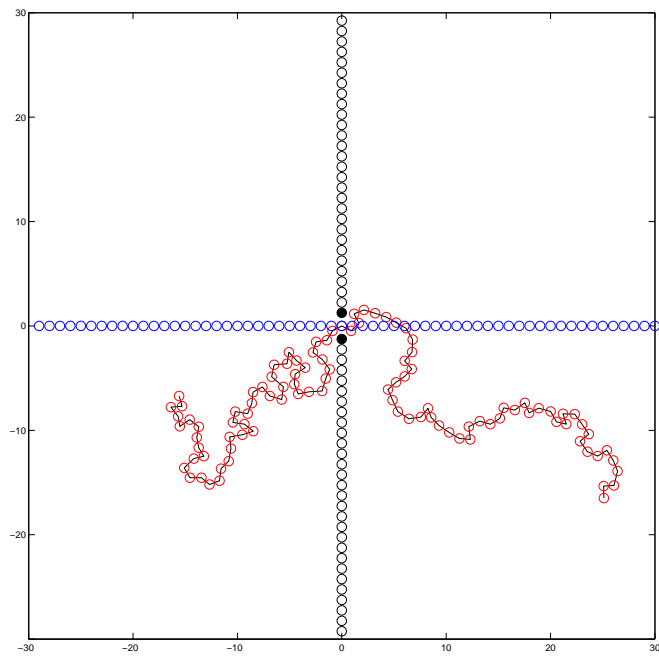
Figure 6.3: Straight configuration vs equilibrium configuration
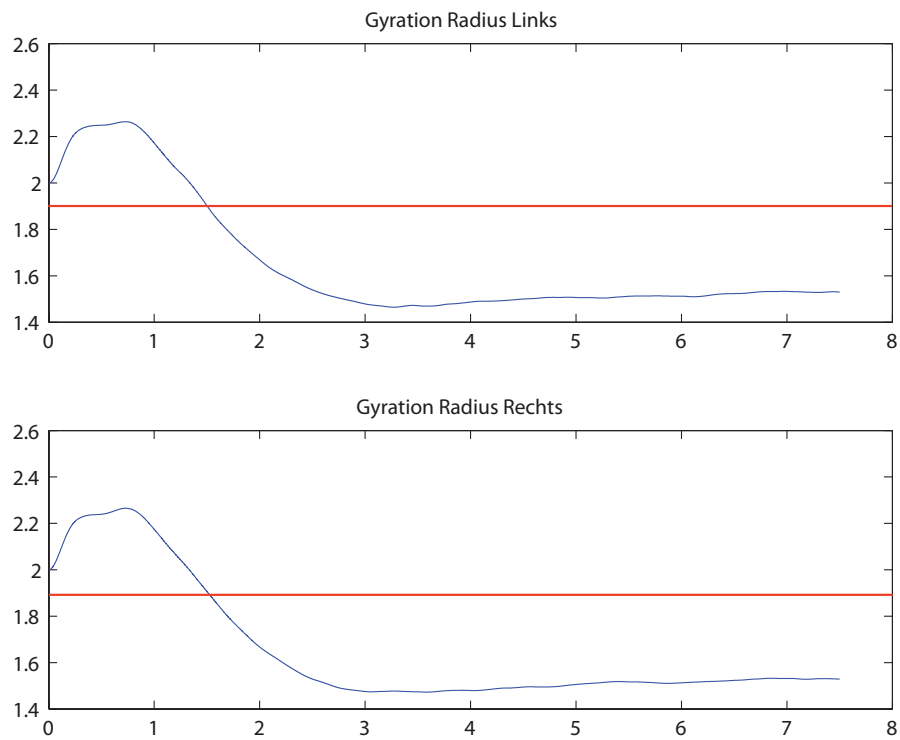
(a) Gyration radii
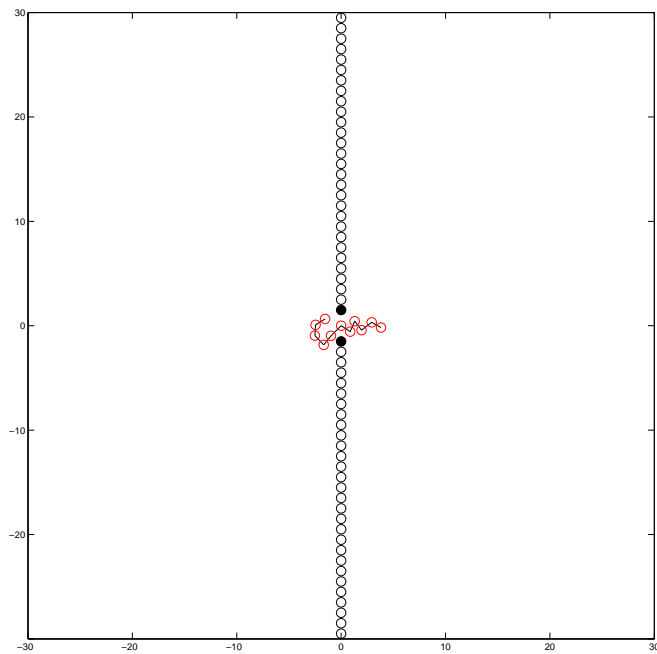
(b) Final configuration

Figure 6.4: Equilibrium position $N = 11$.
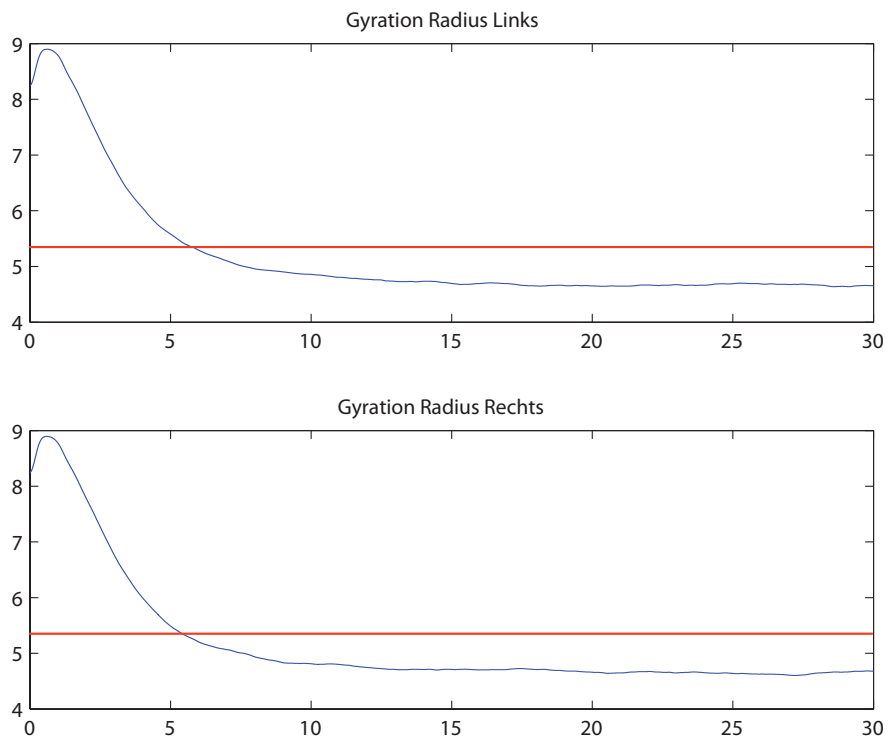
(a) Gyration radii
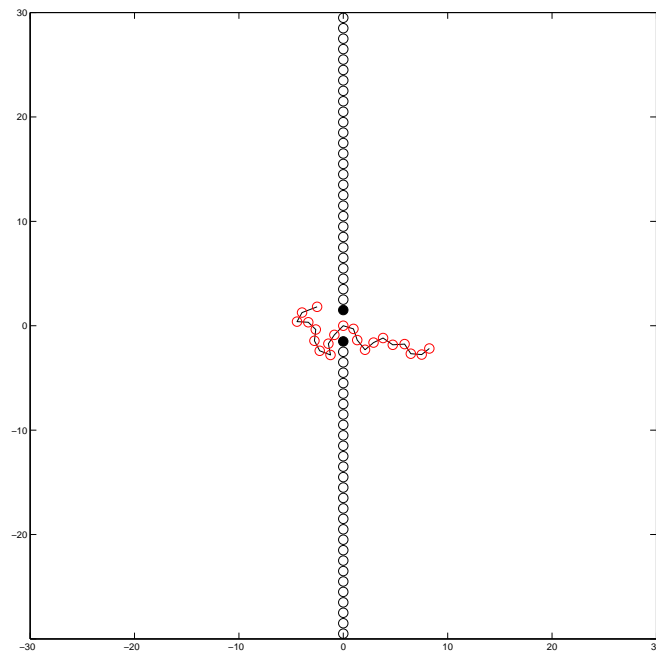
(b) Final configuration

Figure 6.5: Equilibrium position $N = 21$.
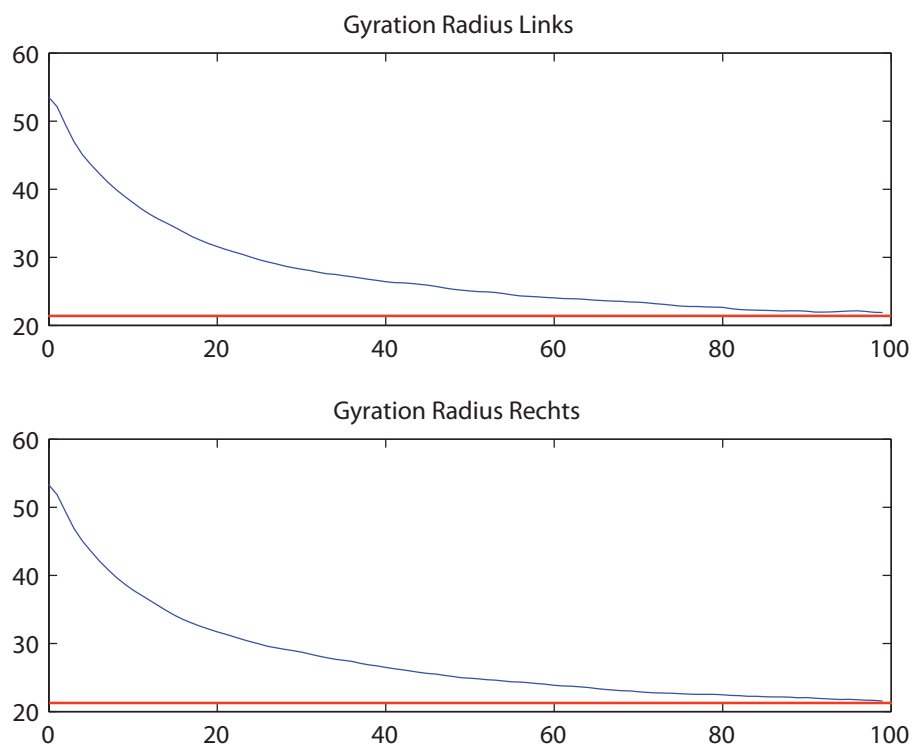
(a) Gyration radii
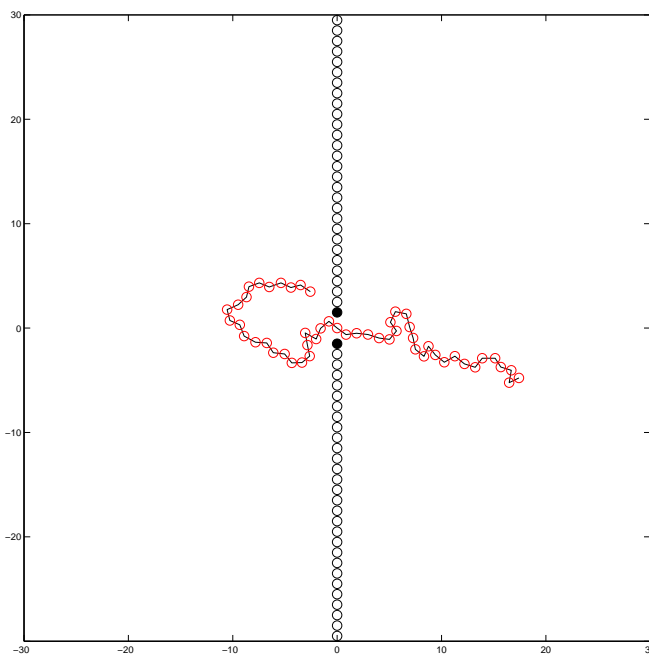
(b) Final configuration

Figure 6.6: Equilibrium position $N = 51$.
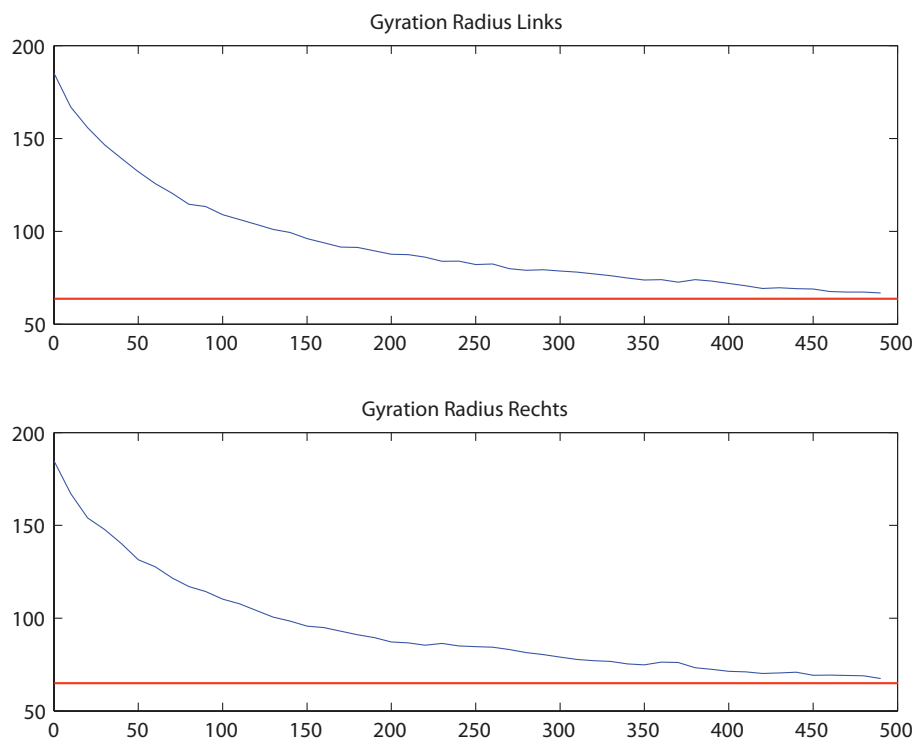
(a) Gyration radii
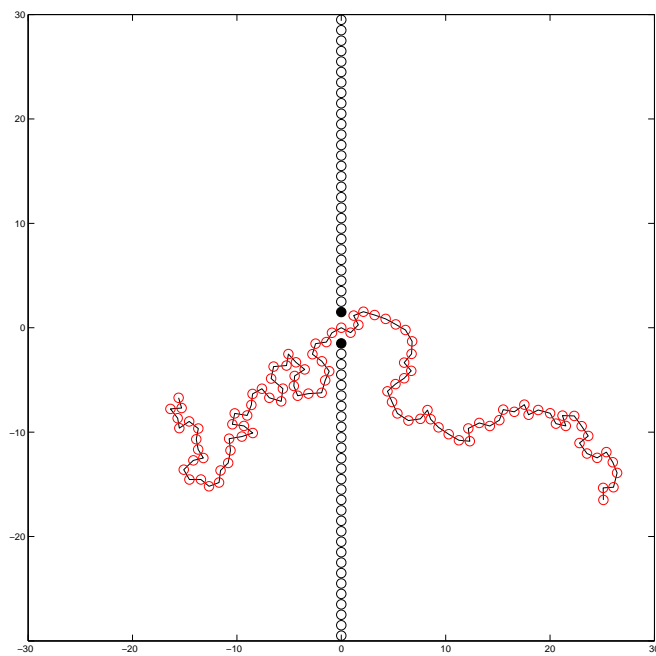
(b) Final configuration

Figure 6.7: Equilibrium position $N = 101$.
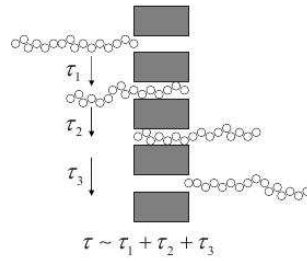
Figure 6.8: Three stages of translocation, τ_1 the initial filling of the pore, τ_2 the translocation, τ_3 the emptying of the pore
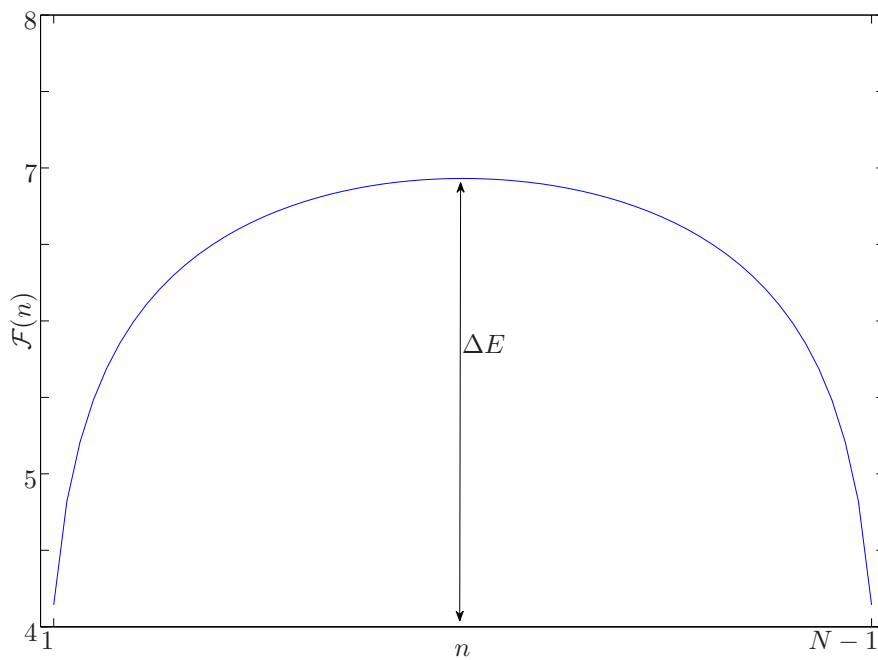
Figure 6.9: Free energy $\mathcal{F}(n)$ as a function of the translocation coordinate n

Figure 6.10: τ vs N, Power law fit $\sim N^{1.26}$

(a) Translocation time distribution $N = 11$

(b) Translocation time distribution $N = 21$

(c) Translocation time distribution $N = 51$

(d) Waiting time distribution $N = 11$

(e) Waiting time distribution $N = 21$
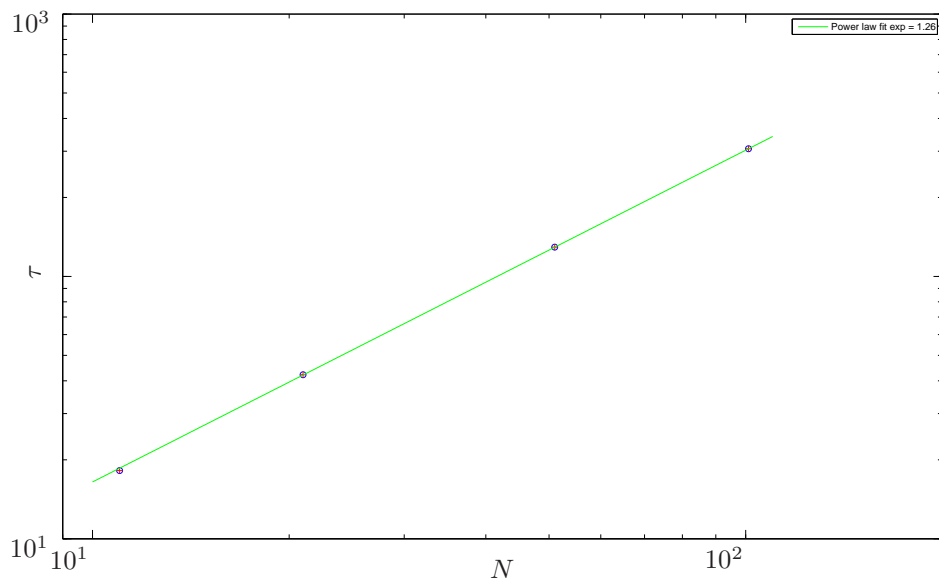
(f) Waiting time distribution $N = 51$
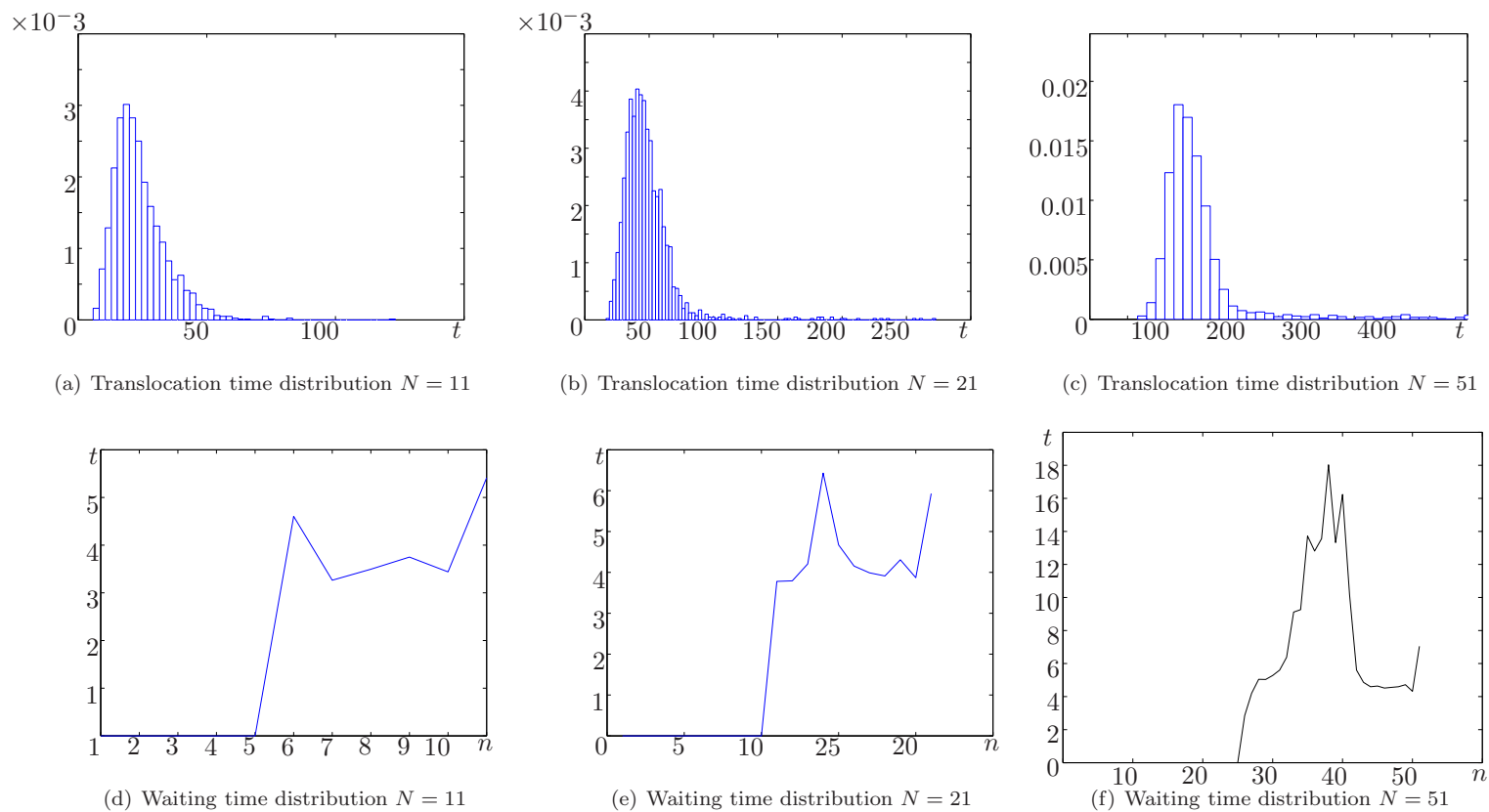
Figure 6.11: Translocation time and Waiting time distributions for homogeneous polymers with length $N = 11, 21, 51$

(a) Translocation

(b) Waiting time distribution

Mode FPT	306.54
Mean FPT	310.25
Gaussian fit	$\mu = 301.4$
	$\sigma = 51.24$
Exp fit	exp = -0.037
Double exp fit	exp1 = -0.24
	exp2 = -0.024
Power law fit	exp = -12.87
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.031

(c) Data

(d) Translocation time log-log distribution

(e) Translocation time distribution

(f) Translocation time cumulative distribution
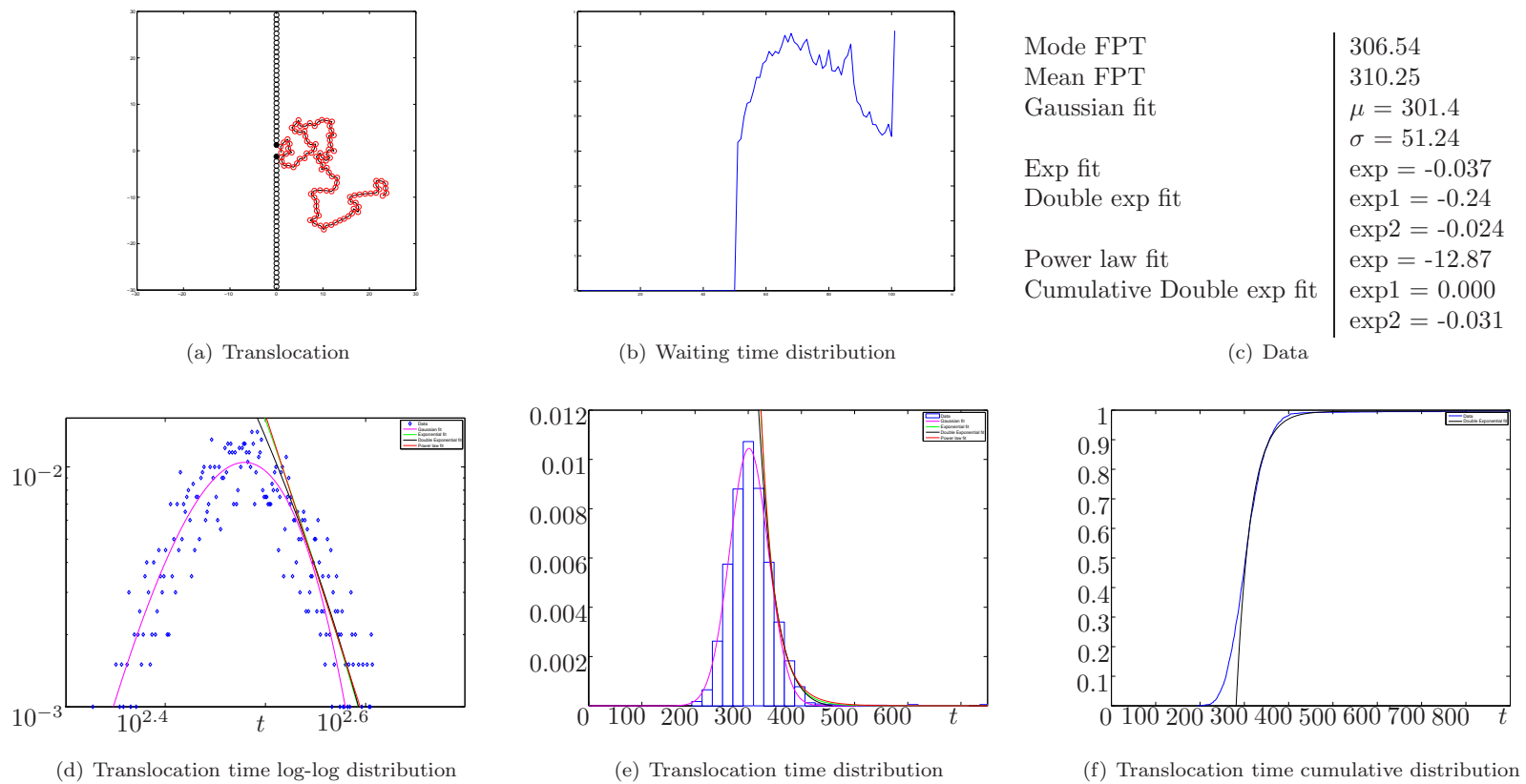
Figure 6.12: Homogeneous polymer $N = 101$.

CHAPTER 7

HETEROGENEOUS TRANSLOCATION

Instead of homogeneous polymers, this chapter focusses on heterogeneous polymers. This means that the polymer will consist of multiple different monomers. These monomers differ in the way they interact with the nanopore. In this chapter polymers with two different monomers will be investigated. The model itself doesn't change much. Only the base-pore interaction modelled bij the LJ potential is modified. The interaction strengths are specific for each base. For example ε_{pA} and ε_{pC}.

Note: In this chapter the A type monomers are labeled red and these are the ones with a high interaction strength and the C type ones, labeled blue, have a low interaction strength with the pore.

Having a high interaction strength means that this monomer is highly attracted to the pore and prefers to stay inside more than the monomers with a low attraction strength. This will influence the total translocation time. By the use of simulations we will try to infer some features of heterogeneous polymers translocation through a nanopore. For a number of different sequences we investigate the translocation time distribution and the waiting time. In particular we investigate a number of polymers with random sequences. We find surprisingly that for the interactions chosen, it is possible to discriminate between the A and C type monomers in the chain.

RED VS BLUE

Let's start with a very simple example. Almost all papers involving the translocation of heterogeneous polymers mention the dependence of the translocation time on the orientation of which type of monomer of the polymer enters the pore first. For a length of $N = 11$ we take two different sequences. They both have five successive blue monomers and six successive red monomers but their orientation is turned. Simulations now give for the blue orientated polymer a translocation time of 22.33 time unit and for the red oriented polymer a translocation time of 14.45 time unit. This is exactly what is expected. As the blue oriented polymer ends with 6 red monomers, it is expected that these prefer to stay inside the pore with their high interaction strength and it will take a lot longer before this polymer leaves the pore.

7.1 HALF RED, HALF BLUE

The next sequence under investigation is a polymer with the first $\frac{1}{2}N$ monomers blue, the middle monomer red and the last $\frac{1}{2}N$ monomers red as one of the polymers of the previous example. As the simulation starts with the polymer half way the pore, these simulations should not give results that differ much with the original simulations with homogeneous polymers. Though the translocation process could go slightly faster. For the polymers of length $N = 11, 21, 51$ the translocation time is plotted versus N in figure 7.1. Fitting the data with a power law gives $\tau \propto N^{1.23}$. **Remark:** The data of $N = 11$ seems to be a bit off. It should be smaller, as it should be smaller than homogeneous translocation. As said before, chains of length $N = 11$ could be too short to produce reasonable data.
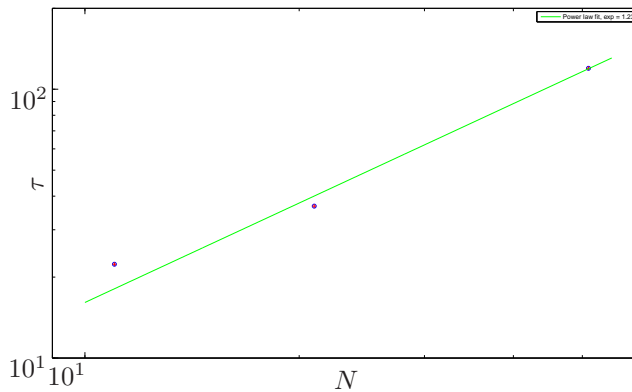
Figure 7.1: τ vs N Sequence half red, half blue, Power law fit $\sim N^{1.23}$

7.2   A FOURTH PART RED, A FOURTH PART BLUE, A FOURTH PART RED AND A FOURTH PART BLUE

Because our translocation process starts with the polymer already halfway the pore the previous section could not give much new information. In this section the two half sides of the polymer are divided in two parts again. Again the translocation time versus the length of the polymer for $N = 21, 51, 101$ is plotted, but now for these new sequences, in figure 7.2. Fitting the data with a power law gives $\tau \propto N^{1.17}$. The translocation process goes faster than in the case of homogeneous translocation. This is indeed what one would expect, there are more blue monomers translocating that go faster through the pore than the red monomers.

Let's have a closer look at the translocation process of the polymer with length $N = 101$ (figure 7.4). The distribution of the translocation time can be fitted with the Gaussian, Exponential, Double Exponential and Power law fit. And the cumulative distribution with the Double Exponential fit. In figure 7.4(b) the waiting time is plotted. Something unusual happens there and this is rather interesting. First the last red monomer of the first half passes through the pore. As this one won't be replaced by another red attractive monomer, but by a blue one, this last red monomer stays inside the pore for a long time. The first blue monomers from the second half shoot through the pore very quickly. Then the translocation process slows down a little. As the red ones are approaching, it speeds up again. The first new red monomers stay again a long while in the pore and then the translocation process speeds up. Again the final red monomer takes a long while inside the pore before leaving it, as was already mentioned for the homogeneous case. **Remark:** Again the high peak around monomer $n = 85$ could be
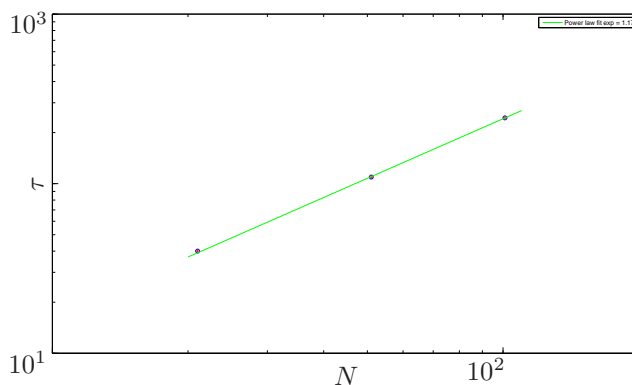
Figure 7.2: τ vs N Sequence a 4th red, a 4th blue, a 4th red, a 4th blue, Power law fit $\sim N^{1.17}$

explained by configuration issues.

For $N = 101$ the simulation was also done for the other orientation (red starts translocating first), see figure 7.5. As the translocation process ends with blue monomers, the translocation time is faster than in the reversed orientation case. Again take a look at the waiting time of the monomers. It is possible to recognise where the new blue sequence starts!. First about 25 red monomers pass through the nanopore. As the end of the red sequence is approaching, the monomers stay inside the pore for a longer amount of time. Especially when the transfer from red to blue happens this seems to hold up the translocation process. The red monomers want to stay inside and as they won't be replaced by other red monomers they don't like to let the blue ones enter the pore. After the blue ones have entered the pore the translocation process quickly speeds up.

7.3 ALTERNATING RED AND BLUE

The next sequence consists of alternating red and blue monomers. Because there are no clusters of the red subsequent monomers that slow down the translocation process, the translocation process goes faster. For the lengths of $N = 11, 21, 51, 101$ the translocation time τ vs the number of monomers N is plotted in figure 7.3. This gives a power law fit with: $\tau \propto N^{1.24}$.

Features of the translocation process of the alternating sequence of length $N = 101$ is shown in figure 7.6. Again the waiting time of the polymer shows a lot of the sequence of the polymer. There is definitely an almost fully consistent alternating pattern. And the last monomer shows to be a red one. Changing the red ones in blue ones and the other way around gives figure 7.7.

7.4 RANDOM SEQUENCES

In this section the behaviour of random sequences will be explored. For five different chains of four different lengths the translocation time and waiting time are investigated, see figures 7.8, 7.9, 7.11 and 7.12. The translocation times seem to be distributed the same way the previous translocation times of the previous chains were distribution. The distribution can be fitted by a Gaussian and the tail is slightly heavy, which can be fitted by a exponential or power law. The double exponential fit seems at sight again to be the best fit for the tail. This can also be concluded for the cumulative distribution. The translocation process goes faster than in the homogeneous case, as there are blue monomers present.

The most interesting are the waiting times. Almost fully consistent can the sequence of the randomly placed blue and red monomers be read out of the waiting time figure. The red dots in the average waiting time (see the (b) part of the figures) correspond to an A-type monomer in the pore, a blue dot to a C-type monomer, counting from the middle monomer from right to left. After a minor decrease at the start, the waiting time increases for subsequent red monomers and decreases for blue monomers. The upper peaks correspond to red monomers and the lower peaks correspond to the blue monomers. This way a unknown sequence can be identified by investigating the waiting times. It is expected that under perfect conditions, for instance starting at the beginning of the nanopore and 10^4 simulations will give perfect results.

Figure 7.3: τ vs N Alternating sequence, Power law fit $\sim N^{1.24}$

(a) Translocation

(b) Waiting time distribution

Mode FPT	244.38
Mean FPT	290.94
Gaussian fit	$\mu = 254.3$
	$\sigma = 46.42$
Exp fit	exp = -0.041
Double exp fit	exp1 = -0.042
	exp2 = -0.003
Power law fit	exp = -15.78
Cumulative Double exp fit	exp1 = -0.000
	exp2 = -0.044

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.4: Sequence a 4th red, a 4th blue, a 4th red, a 4th blue, $N = 101$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	234.50
Mean FPT	263.45
Gaussian fit	$\mu = 256.5$
	$\sigma = 47.7$
Exp fit	exp = -0.032
Double exp fit	exp1 = -0.032
	exp2 = 0.023
Power law fit	exp = -9.7
Cumulative Double exp fit	exp1 = -0.000
	exp2 = -0.028

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.5: Sequence a 4th blue, a 4th red, a 4th blue, a 4th red, $N = 101$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	247.96
Mean FPT	283.94
Gaussian fit	$\mu = 253.3$
	$\sigma = 45.63$
Exp fit	exp = -0.033
Double exp fit	exp1 = -0.033
	exp2 = 0.022
Power law fit	exp = -19.5
Cumulative Double exp fit	exp1 = -0.000
	exp2 = -0.042

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.6: Alternating sequence starting with a red monomer, $N = 101$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	256.11
Mean FPT	276.21
Gaussian fit	$\mu = 268.6$
	$\sigma = 44.16$
Exp fit	exp $= -0.032$
Double exp fit	exp1 $= -0.050$
	exp2 $= -0.050$
Power law fit	exp $= 26.53$
Cumulative Double exp fit	exp1 $= -0.004$
	exp2 $= -0.013$

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.7: Alternating sequence starting with a red monomer, $N = 101$.

(a) Translocation

(b) Waiting time

Mode FPT	303.81
Mean FPT	316.08
Gaussian fit	$\mu = 292.8$
	$\sigma = 57.47$
Exp fit	exp = -0.026
Double exp fit	exp1 = -0.015
	exp2 = -0.020
Power law fit	exp = -8.65
Cumulative Double exp fit	exp1 = -0.000
	exp2 = -0.026

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution
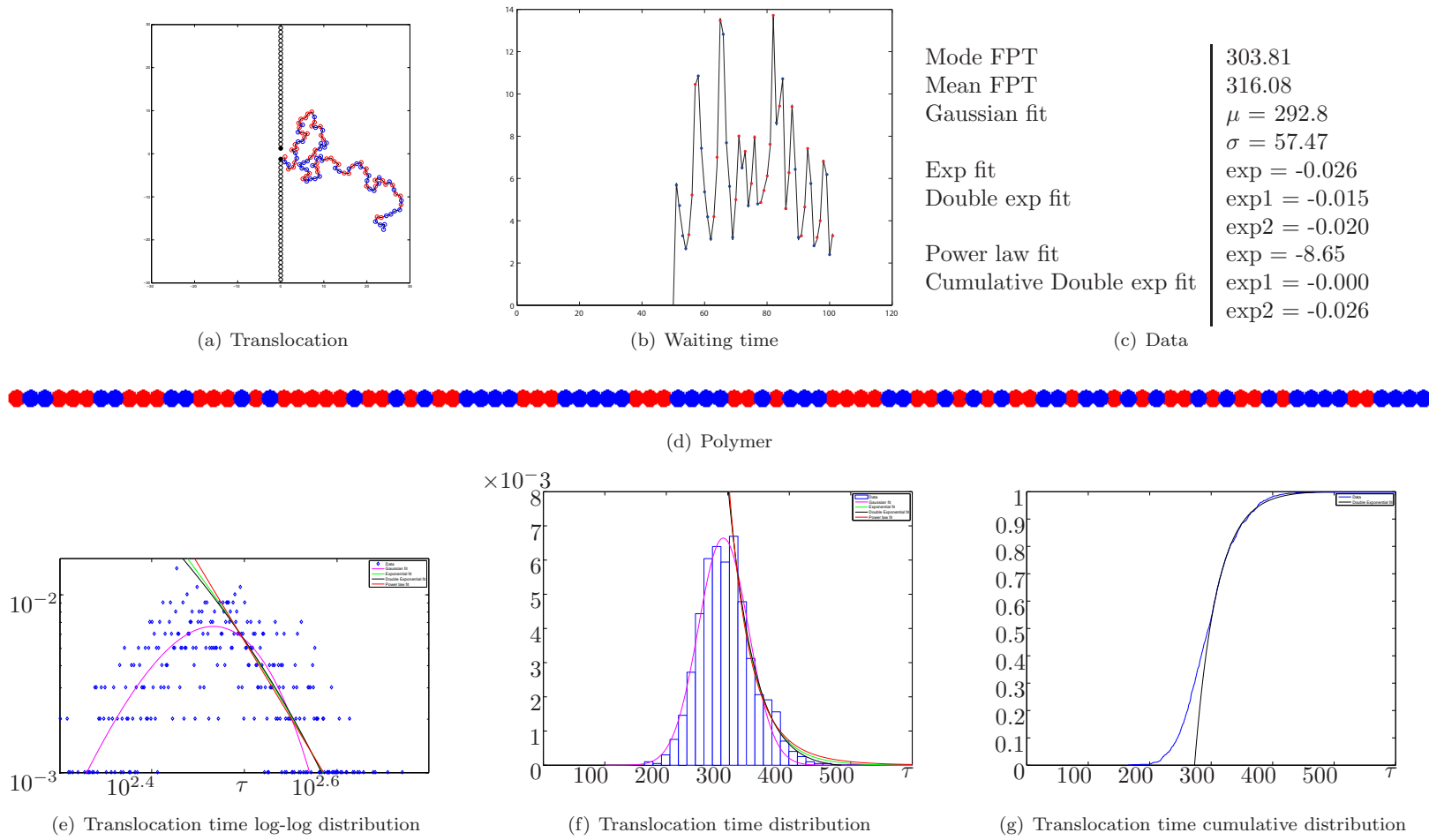
Figure 7.8: Random sequence 1 $N = 101$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	283.21
Mean FPT	291.22
Gaussian fit	$\mu = 284.5$
	$\sigma = 54.2$
Exp fit	exp = -0.031
Double exp fit	exp1 = -0.033
	exp2 = -0.040
Power law fit	exp = -10.01
Cumulative Double exp fit	exp1 = -0.000
	exp2 = -0.035

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.9: Random sequence 2 $N = 101$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	107.59
Mean FPT	262.08
Gaussian fit	$\mu = 106.7$
	$\sigma = 27.07$
Exp fit	exp = -0.054
Double exp fit	exp1 = -0.55
	exp2 = -0.006
Power law fit	exp = -7.31
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.033

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.10: Random sequence $N = 51$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	42.8
Mean FPT	40.71
Gaussian fit	$\mu = 37.87$
	$\sigma = 16.27$
Exp fit	exp = -0.075
Double exp fit	exp1 = -0.84
	exp2 = -0.086
Power law fit	exp = -4.0
Cumulative Double exp fit	exp1 = -0.000
	exp2 = -0.009

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.11: Random sequence $N = 21$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	15.97
Mean FPT	23.92
Gaussian fit	$\mu = 20.94$
	$\sigma = 12.65$
Exp fit	exp = -0.085
Double exp fit	exp1 = -0.083
	exp2 = -0.056
Power law fit	exp = -2.60
Cumulative Double exp fit	exp1 = -0.0002
	exp2 = -0.101

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 7.12: Random sequence $N = 11$.

Chapter 8

Changing forces

In addition to the previous chapters, this chapter presents the results of a few cases where the external force is changed. The results are presented in figures (8.1, 8.2, 8.3, 8.4, 8.5, 8.6). In the previous chapters the external force was always taken equal to 2. This makes the translocation process to go faster and this way, simulations will take less time. Though this makes it hard to compare results with results from the theoretical investigation of the literature study part. To make simulation times resonable, the following simulations were run: 2 polymers of length $N = 51$ with an external force equal to 1, simulated a 1000 times; 2 polymers of length $N = 21$ with an external force equal to 0.5, simulated 500 times and finally 2 polymers of length $N = 11$ with an external force equal to 0.1, simulated 100 times. The two different sequences were a homogeneous sequence and an alternating sequence. It is important to note that the fewer the simulations, the worse the results. Especially the cases of $N = 11$ are not reliable. Though three important conclusions are worth noting. First of all, translocation takes considerably longer than in the cases of a big external force, which was to expected. Second, the behaviour determined in the previous chapters can again be seen in the lower external force cases. For instance the emptying of the pore when the last monomer(s) are red and the alternating waiting time for alternating sequences. The third conclusion is interesting compared to the theoretical results. As the external force goes to zero, the exponent of the power law fit goes to 1.8 as was the case in (Dubbeldam et al., 2007b).

(a) Translocation

(b) Waiting time distribution

Mode FPT	226.80
Mean FPT	422.64
Gaussian fit	$\mu = 266.4$
	$\sigma = 90.84$
Exp fit	exp = -0.010
Double exp fit	exp1 = -0.014
	exp2 = -0.025
Power law fit	exp = -3.6
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.007

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 8.1: External force $F = 1$, $N = 51$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	227.60
Mean FPT	375.38
Gaussian fit	$\mu = 227.8$
	$\sigma = 80.91$
Exp fit	exp = -0.011
Double exp fit	exp1 = -0.058
	exp2 = -0.032
Power law fit	exp = -4.2
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.003

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution
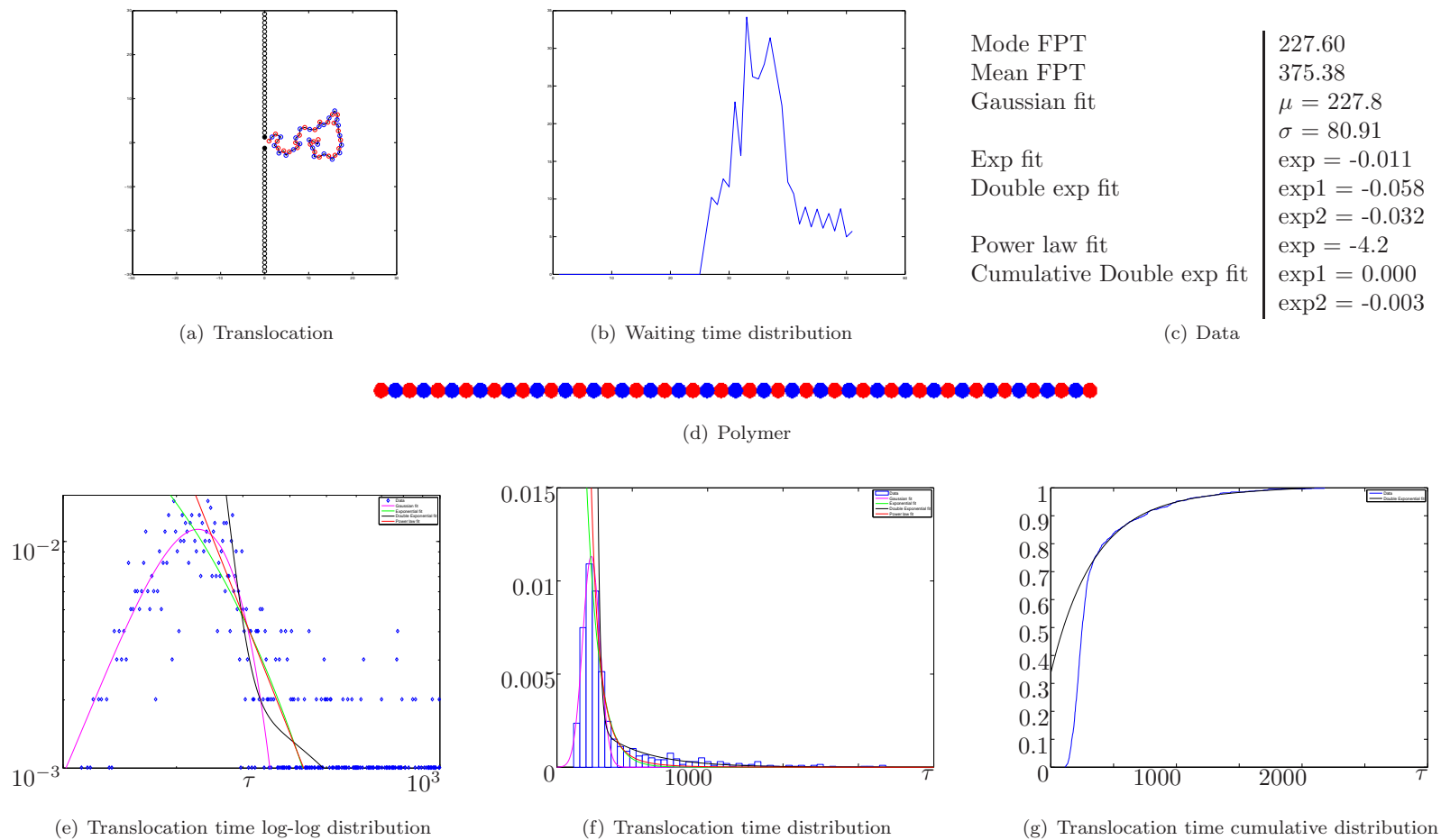
Figure 8.2: External force $F = 1$, alternating sequence, $N = 51$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	128.23
Mean FPT	227.15
Gaussian fit	$\mu = 173.3$
	$\sigma = 133$
Exp fit	exp = -0.008
Double exp fit	exp1 = -0.005
	exp2 = -0.005
Power law fit	exp = -3.1
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.011

(c) Data

(d) Polymer

(e) Translocation time log-log

(f) Translocation time

(g) Translocation time cumulative
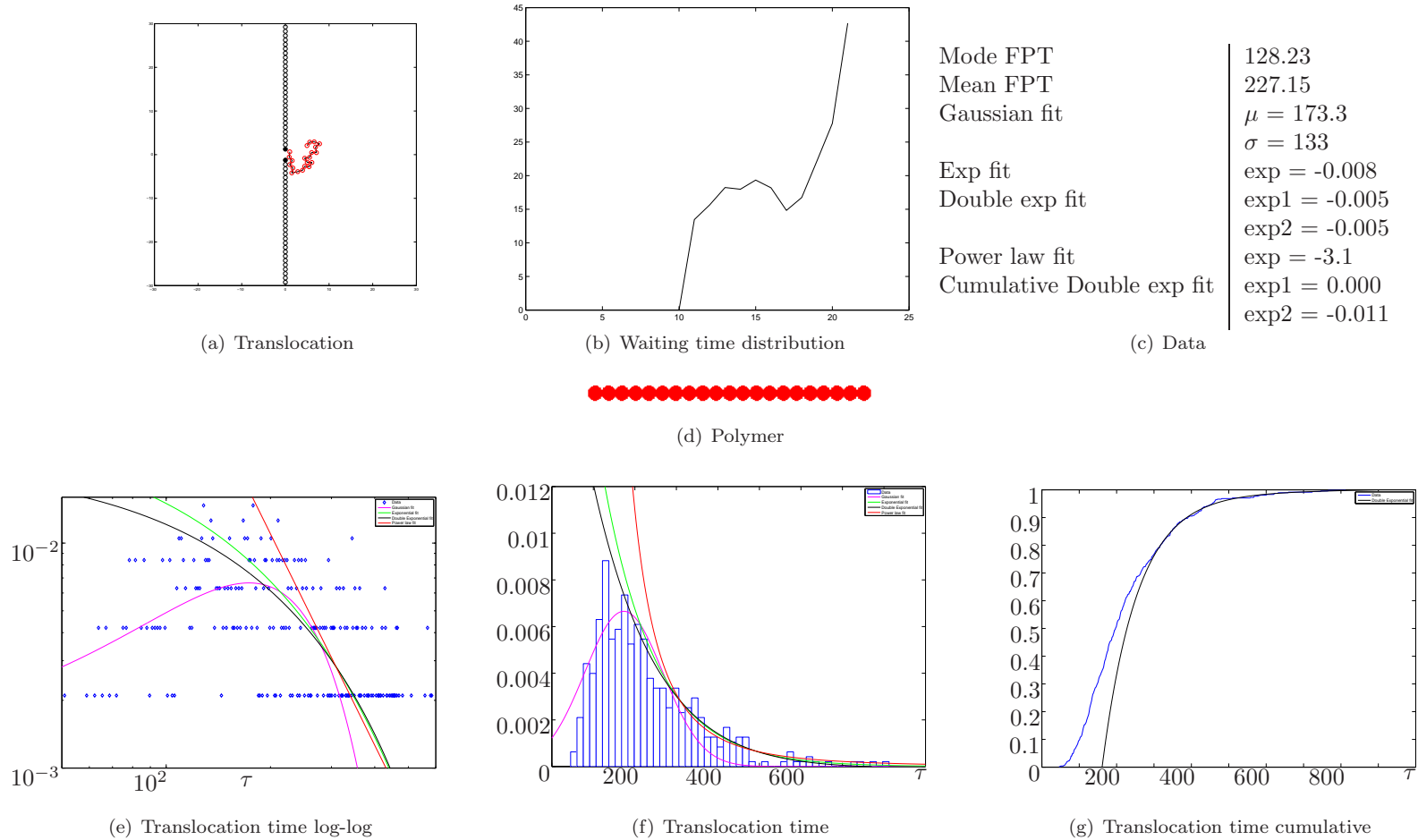
Figure 8.3: External force $F = 0.5$, $N = 21$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	130.02
Mean FPT	166.65
Gaussian fit	$\mu = 140.6$
	$\sigma = 87.54$
Exp fit	exp = -0.011
Double exp fit	exp1 = -0.007
	exp2 = -0.010
Power law fit	exp = -2.8
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.013

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution
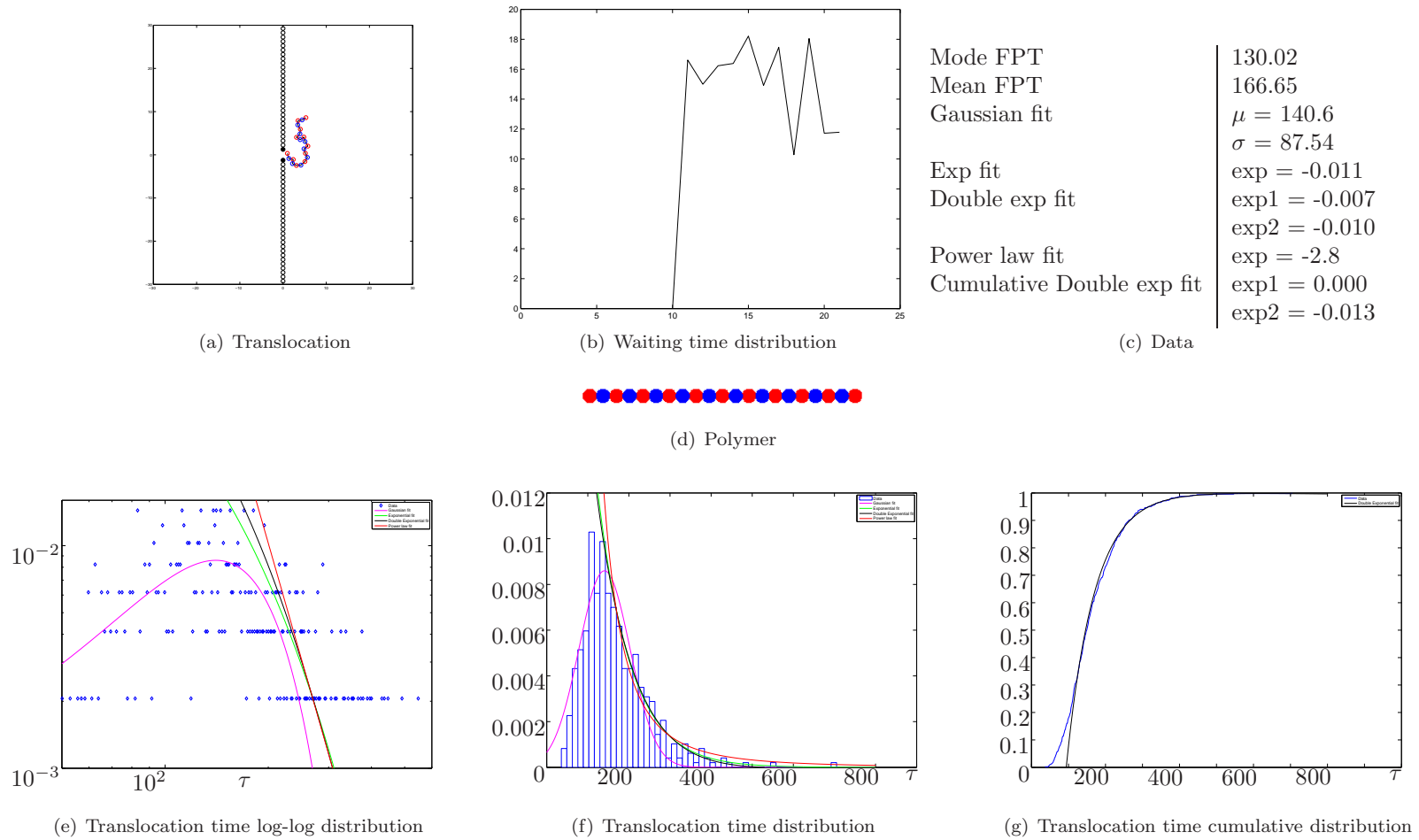
Figure 8.4: External force $F = 0.5$, alternating sequence $N = 21$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	47.57
Mean FPT	342.40
Gaussian fit	$\mu = 19.81$
	$\sigma = 65.66$
Exp fit	exp = -0.003
Double exp fit	exp1 = -0.0014
	exp2 = -0.0014
Power law fit	exp = -0.933
Cumulative Double exp fit	exp1 = 0.000
	exp2 = -0.004

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution
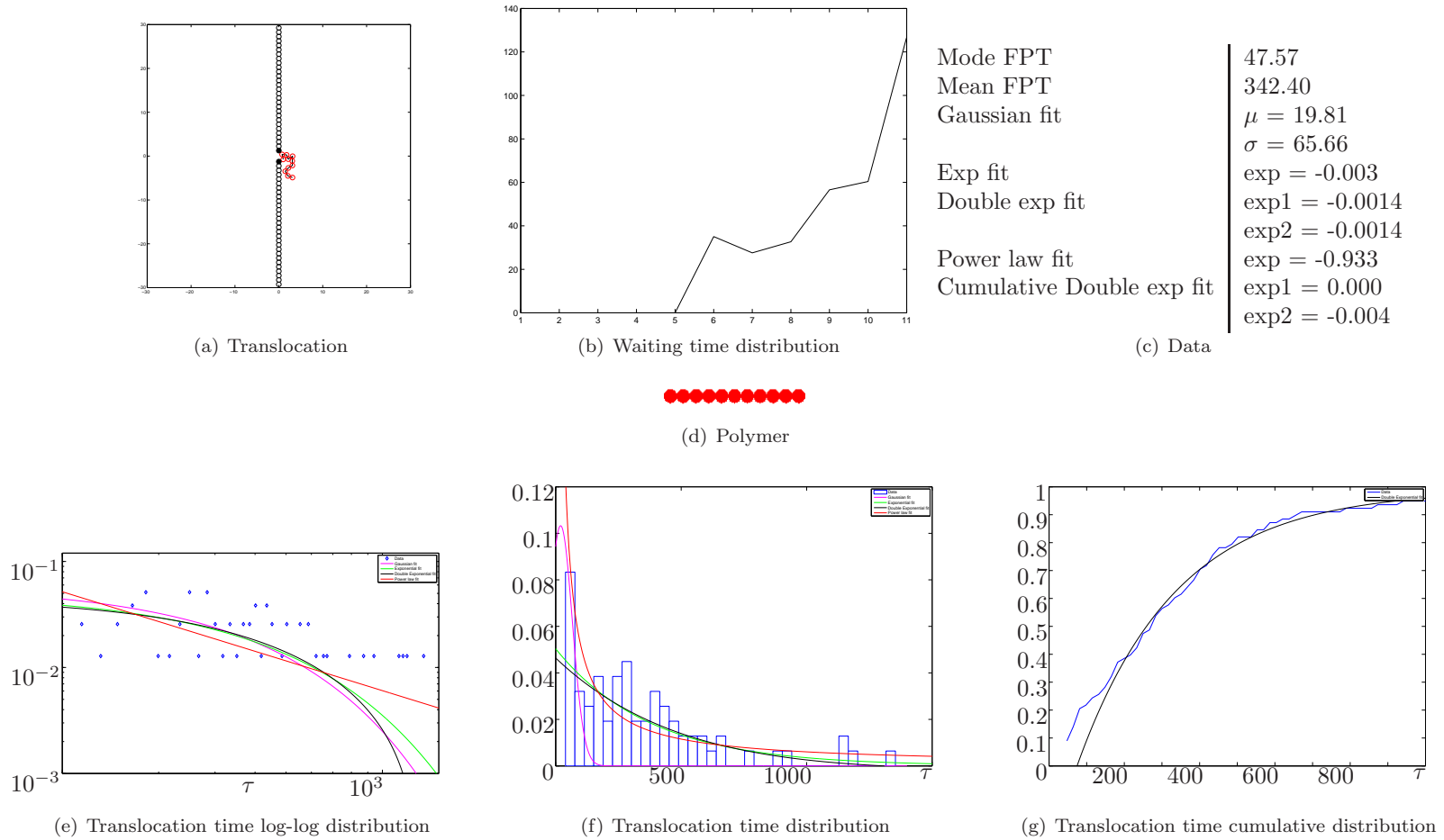
Figure 8.5: External force $F = 0.1$, $N = 11$.

(a) Translocation

(b) Waiting time distribution

Mode FPT	110.4
Mean FPT	195.01
Gaussian fit	$\mu = 89.03$
	$\sigma = 85.65$
Exp fit	exp = -0.008
Double exp fit	exp1 = -0.258
	exp2 = -0.005
Power law fit	exp = -1.76
Cumulative Double exp fit	exp1 = -0.898
	exp2 = -0.007

(c) Data

(d) Polymer

(e) Translocation time log-log distribution

(f) Translocation time distribution

(g) Translocation time cumulative distribution

Figure 8.6: External force $F = 0.1$, alternating sequence $N = 11$.

Chapter 9

Program

To get the figures in this thesis a self-written program is used. This program is written in Matlab and based on the theory of this thesis and (Vlugt et al., 2008). In this chapter the most important routines are mentioned and the numerous abilities of the program will be highlighted. The Matlab code itself can be found in the addendum. The different fases below can also be found explained in short in the code.

9.1 Manual of the code

Init

Obviously the program starts with defining a number of things. The init fase is divided in four subfases. In the first fase used-defined constants need to be given that will define the purpose of the program. In the second fase simulation specific variables need to be defined. These are for instance the length and width of the pore or the length of the polymer. More on these used-defined constants can be found in the next section. In the third fase the constants of the main method need to be defined. These are defined in table 9.1.

In fase four some preparing activities take place. These are: preallocating space for big matrices, taking the squares of distance constants and defining the measures of the figures and the area in which the translocation will take place.

Pore

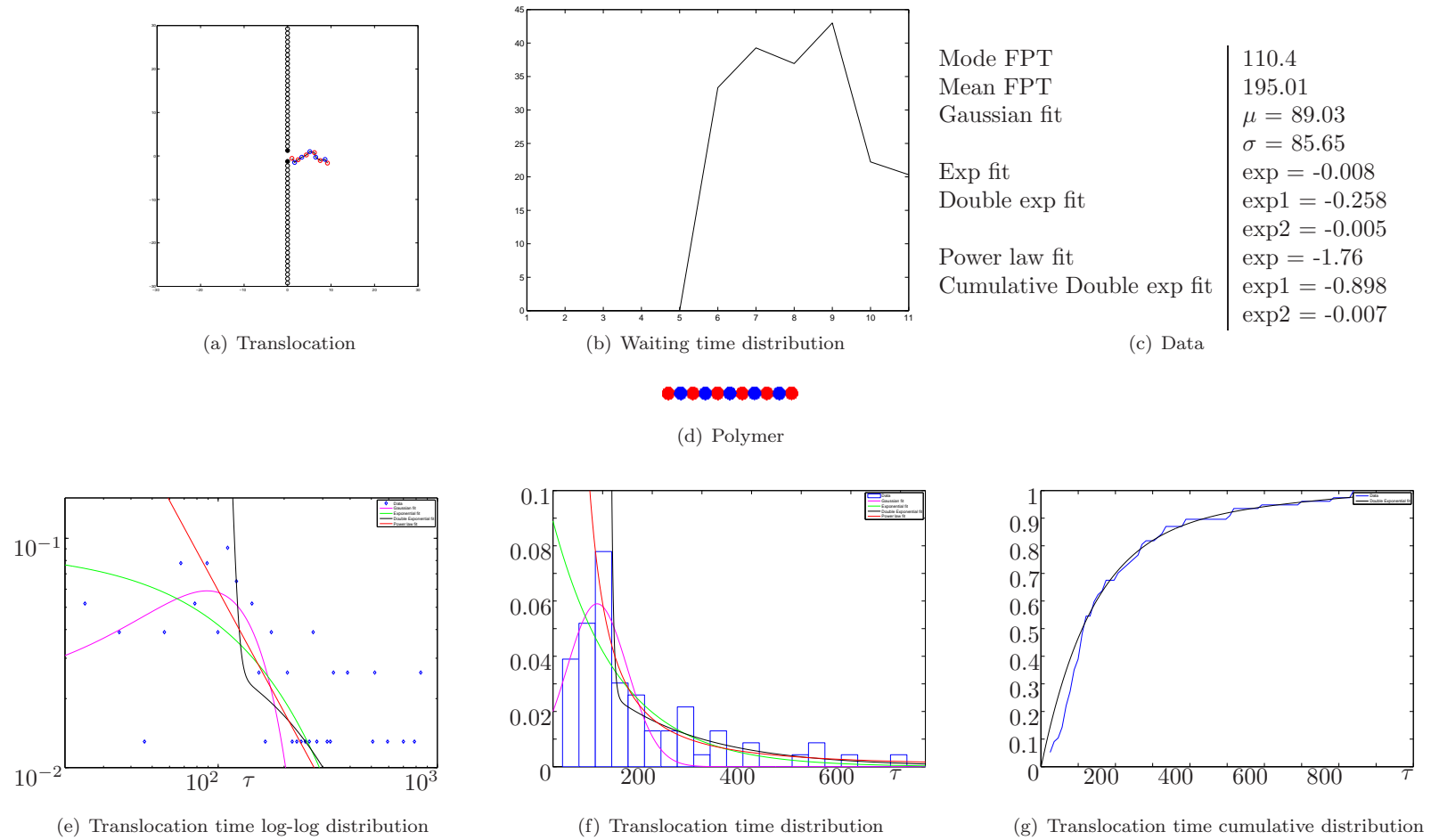Most simulations are used for translocating a polymer. For translocation obviously a pore is necessary. This pore will be built during initialisation. First a membrane is built. The size of the membrane is dependent of the size of the polymer. The size of the pore itself is user-defined. In principle the height of the pore can be defined in n or in $0.5n$ with $n \in \mathbb{Z}$ and the width of the pore in n with $n \in \mathbb{Z}$ as the membrane consists of at least two outer rows of monomers, except obviously when the width is equal to one. For visualization and differences in attration strengths the monomers of the pore and the monomers

Table 9.1: constants

σ	$=$	1	diameter of monomer
m	$=$	1	mass of monomer
T	$=$	0.9	temperature
R	$=$	2	maximum allowed seperation between connected monomers
k	$=$	7	spring constant
k_B	$=$	$\frac{1.2}{T}$	Boltzmann constant
ξ	$=$	0.7	friction constant
ε	$=$	1.2	energy scale
R_C	$=$	$2^{\frac{1}{6}} * \sigma$	cut-off distance
R_{CP}	$=$	$2.5 * \sigma$	cut-off distance attracting pore

(a) Pore, width=5, length=3   (b) Pore, width=1, length=1
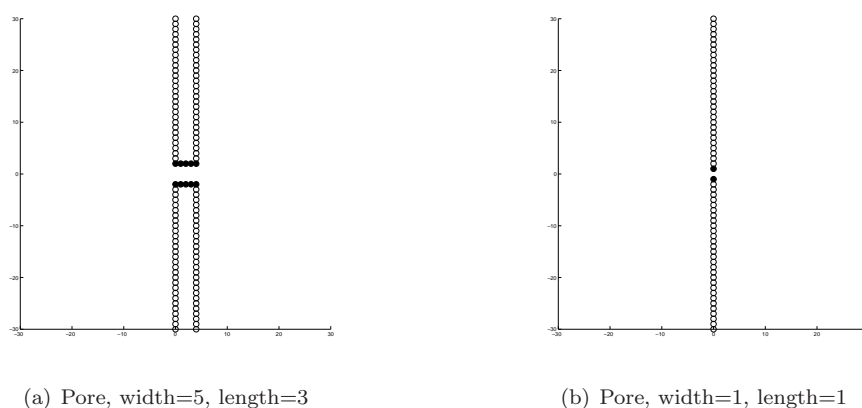
Figure 9.1: Different sizes of pores

of the membrane are distinguished. After building the pore the number of monomers that interact with the polymer is made smaller for calculation conveniences. Now the simulation can start.

INITIAL POSITION POLYMER

For the intial position of the polymer (with or without pore) three different options are possible. First of all the simulation can start with the polymer in a straight line. The second option is to use the configuration of a saved configuration. And the third option is a random configuration. The second option is the most useful. While using the first you need a lot of time to let the polymer go to an equilibrium position, but this will likely always work out. The last option might go to the equilibrium faster but the possibility that an error occurs becomes bigger. Though, for safety, additional restraints are placed on the monomers close to the pore. After the polymer is placed you can get a visualisation of the polymer (and the pore). Besides the initial configuration, the first distance matrices are defined and also the initial velocity. The random initial configuration is found by trying random configurations and calculating whether the monomers aren't either to close to each other or too far away.

START TRANSLOCATION

Before starting the translocation the program checks three things. First it calculates the monomers that are closest to each other. This is done for speeding up the program. This way it won't be necessary to calculate every time the interactions between all monomers. For this we need to assume that these distances won't change too many too quickly. The second check involves defining which monomer is in the pore before the translocation process starts, this way we can track when the monomer leaves the pore. The final check involves the conditions whether the program should still run. These conditions are in case of a pore whether the polymer is still in the pore and no error has occured and in case of running an equilibrium configuration or a time defined movement of the polymer whether it is finished. In the case of an equilibrium configuration the external force needs to be redefined as zero. Now the program starts translocating the chain. The amount of translocations being performed is user-defined.

INTERACTION PORE-POLYMER

In the occurence of a pore, the interaction will be calculated of the monomers of the pore with the monomers of the polymer. This interaction is three-fold, the three interactions are: interaction with membrane monomers, interaction with the pore with red monomers or with blue monomers. In the first case there exists only a repulsive interaction strength and in the second and third cases their also exists a attractive strength which is different in case of different monomers of the polymer (red or blue). A new distance matrix needs to be calculated. And the interaction will only be calculated for the monomers that are close enough, these are within the cut-off distance. So the interaction from the LJ potential is only calculated for a selected number of monomers. This is also defined in the LJ potential. When the
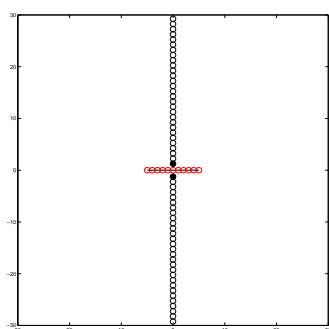
(a) Polymer straight starts halfway the pore
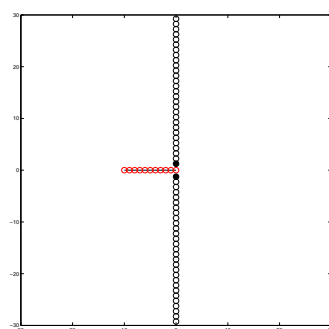
(b) Polymer straight starts at the left side of the pore
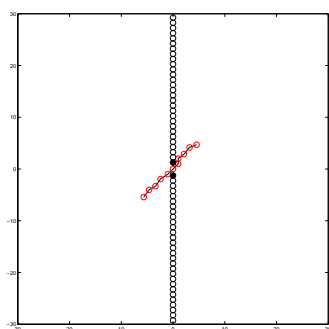
(c) Random polymer starts halfway the pore
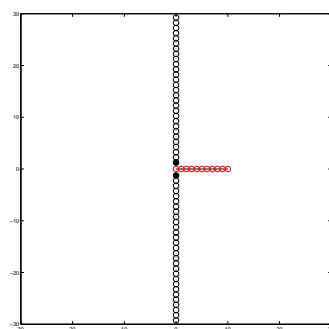
(d) Polymer straight starts at the right side of the pore

Figure 9.2: Different initial positions

distance is too big, the potential is defined to be zero. For information purposes besides the potential also the energy associated with the potential will also be calculated in this step. As the monomers inside the pore also undergo a external force exerted on them, these monomers are also found in this step.

INTERACTION MONOMERS

After the interaction of the polymer with the pore, the interaction between the monomers of the polymer itself is calculated. This interaction is based on the LJ potential and the FENE potential. Again also the energy associated with the potentials are calculated. These calculations are not done for each momoner with every other monomer, but again for a selected number of monomers, that are within the range of the cut-off distance. Though these are still too many calculations. So the number of monomers that will be checked is lowered by using information from previous calculations. Only those monomers which are in a circle of used-defined radius of a specific monomer will be used for interaction with that monomer a few steps long. This selection is compiled in another step. **N.B.** the initial selection is made in the START TRANSLOCATION subsection.

FORCES

In just two lines of code the random force is calculated and all forces working on the different monomers of the polymer are added to create the total force working on these monomers. These forces are the ones calculated from the LJ and FENE potential, the force calculated from the interaction with the pore, the external force, the random force and the hydrodynamic drag.

Random force: $F_r = 1/\sqrt{\Delta t} * \text{rand} * \sqrt{(2 * k_b * T * \xi)}$

All forces: $f = f + fp + F_{ext} - \xi * v + F_r$

UPDATE POSITION POLYMER

Now the position of the polymer is updated according to the verlet algorithm explained in chapter (5.1.2).

$x_{n+1} = 2 * x_n - x_{n-1} + \delta^2 * f/m$

Also the new velocity is defined according to the verlet algorithm and the old value x_{n-1} is updated. In case there is no x_{n-1}, $(n = 1)$, the value is set to a userdefined value. This is either x_n or based on a Taylor expansion.

$v = (x_{n+1} - x_{n-1})/2 * \Delta t$

Note: At certain places in the code there seems to be an extra adjustment to variables. In these cases periodic boundary conditions were taking into account, but these turned out to be redundant.

CHECK EQUILIBRIUM OPTION

At this point the program checks again whether the option equilibrium is on. When this is the case, the specific monomer of the polymer that needs to stay in the same place all the time, needs to be placed back. Also the external force value is updated. In the case of an equilibrium configuration the external force needs to be redefined as zero.

UPDATE MONOMERS OF INTEREST

Now the selection of monomers that are within a certain radius from eachother needs to be adjusted. This action is two-fold. Most of the time for the monomers in the selection their new distance will be calculated. But at certain times (again used-defined) the distances between all monomers is calculated and a new selection is made. At this point the program also checks whether no error has occured concerning the maximum allowed distance between two subsequent monomers. During simulations this never seems to be the case.

WHICH MONOMER HAS LEFT THE PORE?

For the distribution of the waiting times, it is necessary to know when a monomer has left the pore. So each time the next monomer has left the pore for the first time it is recorded and the time between the successive events is also recorded. Besides being necessary for the waiting time, this is also used as a check whether the polymer is translocating through the pore in the right way.

UPDATE VARIABLES OF INTEREST

At this point a number of values need to be saved in order to compare results at the end of the simulation. These are radii of gyration, centre of mass, potential energy, kinetic energy, temperature and momentum. In case the visualisation is on, the latest configuration will appear on screen. In order to save memory, these values do not need to be updated every time.

CHECK CONDITIONS

In the initial fase the program checked the conditions. Now it checkes them again. These conditions were, in case of a pore whether the polymer is still in the pore and no error has occured and in case of running an equilibrium configuration or a time defined movement of the polymer whether it is finished. If the conditions aren't met, the simulation of the single polymer stops.

CHECK THE RESULTS

The simulation of one single polymer has ended. The translocation time is recorded and the program checks whether it translocated succesfully. For this to happen it needs to have translocated the full polymer and no error has occurred. This error can happen in two ways. First the distance between two monomers became too big. This almost never happens. Or unfortunately the polymer has passed through the membrane, from the trans-side to the cis-side. This error seems to happen occasionally, but cannot be explained yet. If a mistake occurs the program shows a visualization of this. The program also shows how many simulations were performed. After this, the a new simulation is started, until the initial set number of simulations is reached.

RESULTS

For the different settings of the program different results can be displayed.

- In all cases:
  - The final configuration
  - The simulation time

- In case of no pore:
  - Centre of mass
  - Gyration radius
  - Temperature
  - Momentum
  - Kinetic energy
  - Potential energy
  - Total energy

- In case of equilibrium configuration
  - Gyration radius of cis- and trans-side

- In case of translocation
  - Translocation time
  - Histogram translocation time
  - Waiting time

9.2 User-defined constants

What needs to be done?

The program has several options for running different kind of simulations. First of all, of course, you can define whether you want to insert a nanopore in your model for simulation. For the initial configuration there are three options, a straight line, a random configuration or a saved configuration from a previous simulation. And the polymer can start at the left side of the pore, in the middle and in principe at the right side of the pore. Though for this last option you have to manually change the rest of the code at few places. But translocation from left to right is the same as from right to left, so this option seems redundant. You can also decide to hold one monomer at one place and let the rest of the polymer move till it has found an equilibrium position, for this you also need to list the time when it should be at a equilibrium configuration. After it has reached this position, you can choose to stop the simulation or let it run further and translocate. For visualisation during simulation you can also adjust how fast you would like to see the polymer translocate. But be aware, this can takes a lot of time and memory. The init option is used for the first run of your simulation. If you don't want to use this, you can also let the first step be based on your present value of the position of the polymer. The options will be defined by giving them a value of $0, 1, 2$. See the code for specific values.

What does my polymer looks like?

For defining the polymer there are several values that need to be filled in. You have to define the length of the polymer, and how the 'a' and 'b' monomers are defined. And the interaction strengths of these monomers with the a pore. If a saved configuration is used, you need to load this.

What does my pore look like?

For the pore all you need to define are the length and width of the pore, and the external force exerted on the monomers inside the pore.

How many information needs to be saved?

In order to keep the amount of time and memory needed for simulations reasonable, there are several values you can change. For the selection of the monomers that interact with each other, you can define how many times this selection stays the same and the radius of the circle for the monomers to decide which ones are closest to each other. You can also define how many monomers of the membrane are taken for calculating the interaction with the monomers of the polymer.

How many simulations?

Of course you can adjust the number of simulations you want to run. Also you can define the number of timesteps you want to take per run. This is only possible when you don't want to translocate. You always have to define the value of the timestep.

CHAPTER 10

CONCLUSION

Translocating a chain of different beads through a very small pore can be used as a first step of modelling a DNA chain that passes through a nanopore. This translocation process offers a variety of possibilities in chemical and biological processes, for instance rapid DNA sequencing. In this thesis the chain is modelled as a polymer with different types of monomers as beads. The translocation dynamics of heterogeneous polymers through nanopores can be modelled using the LJ and FENE potentials and different interaction strengths between the monomers of the polymer and the pore. The translocation time gives important information of the chain sequence, depending on the length of the polymer. The waiting time is defined as the time a specific monomer stays inside the pore. This waiting time in particular gives useful results considering the chain sequence. Simulations reveal that the waiting time of the last monomer can define the type of monomer under consideration. Monomers with a high interaction with the pore will stay inside considerably longer. We found that from the average waiting time it is possible to retrieve the original sequence of the beads constituting the chain.

Besides the numerical model, a theoretical stochastic model is presented, based on several different papers. This way an all-round model is covered, containing an equilibrium model, an anomalous diffusion model, a driven translocation model and a heterogeneous translocation model.

The translocation process during simulations is driven, with a force of 2 units. Extra simulations show that reducing this force will lead to translocation times that agree with theoretical results for undriven translocation.

The polymers under consideration vary in length between $N = 11, 21, 51, 101$ and in sequence. In particular several random sequences are simulated. The waiting time distributions of these random chains show the possibilities of sequencing using translocation.

The program used for the simulations is a self written program in Matlab.

10.1 RECOMMENDATIONS

Instead of simulating heterogeneous polymers of two different types of monomers, one could add one or two more types of monomers to the model. The one-on-one relationship between the sequence and its corresponding waiting time could then be really tested and investigated

This thesis is based on simulations of translocations through nanopores. At this point it is not possible yet to define the waiting times of single monomers in real experiments. Though, the ionization potentials for different monomers are different. Therefore in the future it might be that laser-induced attractive interactions along with fluorescence spectroscopy could be used to make a device that will detect either the waiting time or the number of monomers inside the pore as a function of time (Luo et al., 2008c).

Bibliography

Dekker, C. and Kowalczyk, S . Nanogaatjes voor dna-analyse. Nederlands Tijdschrift voor Natuurkunde, (mei):78–82, 2011.

Dekker, N.H. , Krudde, V. , Janssen, X.J.A. , and Hout, M. van den. Distinguishable populations report on interactions of single dna molecules with solid-state nanopores.

Dekker, C. . Solid-state nanopores. Nature Nanotechnology, 2(APRIL):209–215, 2007.

Doi, M. and Edwards, S.F. . The Theory of Polymer Dynamics. Clarendon Press, 1986.

Dubbeldam, J. and Redig, F. . Translocating heterogeneous polymer chains. 2010.

Dubbeldam, J.L.A. , Milchev, A. , Rostiashvili, V.G. , and Vilgis, T.A. . Driven polymer tranlocation through a nanopore: A manifestation of anomalous diffusion. EPL, 79(18002):1–6, 2007a.

Dubbeldam, J.L.A. , Milchev, A. , Rostiashvili, V.G. , and Vilgis, T.A. . Polymer translocation through a nanopore: A showcase of anomalous diffusion. Physical Review E, 76(010801):1–4, 2007b.

Gardiner, C.W. . Handbook of Stochastic Methods. Springer-Verlag, 2004.

Huopaniemi, I. , Luo, K. , Ala-Nissila, T. , and Ying, S.-C. . Polymer translocation through a nanopore under a pulling force. Physical Review E, 75(061912):1–6, 2007.

Kantor, Y. and Kardar, M. . Anomalous dynamics of forced translocation. Physical Review, 69(021806):1–12, 2004.

Krudde, V. . Translocations of poly(a) and dna through small solid-state nanopores, 2009.

Lubensky, D.K. and Nelson, D.R. . Driven polymer translocation through a narrow pore. Biophysical Journal, 77(October):1824–1838, 1999.

Luo, K. , Ala-Nissila, T. , Ying, S.-C. , and Bhattacharya, A. . Influence of polymer-pore interactions on translocation. Physical Review Letters, 99(148102):1–4, 2007.

Luo, K. , Ala-Nissila, T. , Ying, S.-C. , and Bhattacharya, A. . Translocation dynamics with attractive nano-pore-polymer interactions. Physical Review E, 78(061918):1–8, 2008a.

Luo, K. , Ala-Nissila, T. , Ying, S.-C. , and Bhattacharya, A. . Sequence dependence of dna translocation through a nanopore. Physical Review Letters, 100(058101):1–4, 2008b.

Luo, K. , Ala-Nissila, T. , Ying, S.-C. , and Bhattacharya, A. . Dynamics of dna translocation through an attractive nanopore. Physical Review E, 78(061911):1–7, 2008c.

Melchionna, S. , Bernaschi, M. , Fyta, M. , Kaxiras, E. , and Succi, S. . Quantized biopolymer translocation through nanopores: Departure from simple scaling. Physical Review E, 79(030901):1–4, 2009.

Metzler, R. and Klafter, J. . Letter to the editor: When translocation dynamics becomes anomalous. Biophysical Journal, 85:2776–2779, 2003.

Muthukumar, M. . Theory of capture rate in polymer translocation. The Journal of Chemical Physics, 132(195101):1–10, 2010.

Panja, D. , Barkema, G.T. , and Ball, R.C. . Origin and effects of anomalous dynamics on unbiased polymer translocation. (25-07), 2007.

Sung, W. and Park, P.J. . Polymer translocation through a pore in a membrane. Physical Review Letters, 77(4):783–786, 1996.

Vlugt, T.J.H. , Eerden, J.P.J.M. van der, Dijkstra, M. , Smit, B. , and Frenkel, D. . Introduction to Molecular Simulation and Statistical Thermodynamics. Delft, The Netherlands, 2008.

Addendum

Investigating translocation nowadays has a lot of different approaches. Another approach than the one followed in this thesis, is the one where a pulling force is exerted on the first leading monomer of the polymer. A reallife development is the possibility to manipulate single molecules, where the motion of a polymer can be controlled by optical tweezers. There also exists a new sequencing technique that is based on a combination of magnetic and optical tweezers for controlling the DNA motion. Therefore research also focusses on theoretically investigating translocation under a pulling force.

The pulling force is represented as $F_{pulling} = F\hat{x}$, where F is the pulling force strength on the first monomer and $\hat{x}$ is a unit vector in the direction perpendicular to the wall. This force can then be included in the main model of Eq.(6.2) (Huopaniemi et al., 2007).

The Morse potential

Instead of the Lennard-Jones potential in some papers such as (Dubbeldam et al., 2007a) the Morse potential is used. The Morse potential is named after physicist Philip M. Morse, and the potential models the potential energy. The Morse potential is given by

$$U(r) = \varepsilon \left(1 - e^{-a(r-r_e)}\right)^2,\tag{10.1}$$

where r_e is the equilibrium bond distance and a controls the 'width' of the potential.

An example of the one-on-one relationship between the Langevin equation and the Fokker-Planck equation

Following (Doi and Edwards, 1986) we consider the Brownian motion of a free particle, so we know $U = 0$. The Langevin equation then becomes

$$\xi\frac{dx}{dt} = f(t).\tag{10.2}$$

Integrating this gives

$$x(t) = x(t_0) + \frac{1}{\xi}\int_0^t f(t')dt'.\tag{10.3}$$

It is well known in literature that $x(t)$ must be a Gaussian. The probability distribution of $x(t)$ must be

$$\Psi(x,t) = (2\pi B)^{\frac{-1}{2}} \exp\left(-\frac{(x-A)^2}{2B}\right),\tag{10.4}$$

with

$$A = \langle x(t)\rangle \qquad B = \left\langle (x(t) - A)^2\right\rangle.\tag{10.5}$$

Calculating these moments give for A

$$A = x(t_0) + \frac{1}{\xi} \int_0^t \langle f(t') \rangle \, dt' = x_0 \tag{10.6}$$

and for B

$$
\begin{aligned}
B &= \left\langle \left(\frac{1}{\xi} \int_0^t f(t') dt' \right) \left(\frac{1}{\xi} \int_0^t f(t'') dt'' \right) \right\rangle & (10.7) \\
&= \frac{1}{\xi^2} \int_0^t \int_0^{t'} \langle f(t') f(t'') \rangle \, dt'' dt' & (10.8) \\
&= \frac{2k_B T}{\xi} \int_0^t \int_0^{t'} \delta(t' - t'') dt'' dt' & (10.9) \\
&= \frac{2k_B T}{\xi} t & (10.10) \\
&= 2DT. & (10.11)
\end{aligned}
$$

So the probability distribution of $x(t)$ is given by

$$\Psi(x,t) = (4\pi Dt)^{\frac{-1}{2}} \exp\left(-\frac{(x-x_0)^2}{4Dt} \right), \tag{10.12}$$

which is exactly the solution of the Fokker-Plank equation

$$\frac{\partial \Psi}{\partial t} = D \frac{\partial^2}{\partial x^2} \Psi \tag{10.13}$$

CALCULATION OF THE NUMBER OF WALKS THAT YOU CAN TRAVEL, STARTING FROM 0, AND COUNTING N STEPS NEVER CROSSING THE ORIGIN

$$P(x, t+\Delta t) = P(x+\Delta x, t)p_L + P(x-\Delta x, t)p_R + P(x,t)[1-(p_L+p_R)] \tag{10.14}$$

where p_L is the posibility that you step to the left and p_R is the possibility that you step to the right. Taking the Taylor expansion will give

$$
\begin{aligned}
P(x,t) + \frac{\partial P}{\partial t}\Delta t &= P(x,t)p_L + \frac{\partial P}{\partial x}p_L \Delta x + \frac{\partial^2 P}{\partial x^2}p_L\frac{(\Delta x)^2}{2} & (10.15) \\
&+ P(x,t)p_L + \frac{\partial P}{\partial x}p_R \Delta x + \frac{\partial^2 P}{\partial x^2}p_R\frac{(\Delta x)^2}{2} & (10.16) \\
&+ P(x,t) - P(x,t)(p_L + p_R) & (10.17) \\
\frac{\partial P}{\partial t}\Delta t &= \frac{\partial P}{\partial x}\Delta x(p_R - p_L) + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}(\Delta x)^2[p_L + p_R] & (10.18)
\end{aligned}
$$

We take both p_L and p_R equal to $\frac{1}{2}$ and we take $\Delta x = 1$ and $\Delta t = 1$ and arrive at

$$\frac{\partial P}{\partial t} = \frac{1}{2}\frac{\partial^2 P}{\partial x^2} \tag{10.19}$$

This is a second order differential equation. We need two boundary conditions and one initial condition.

$$
\begin{aligned}
P(x,0) &= \delta(x - x_0) & (10.20) \\
P(0,t) &= 0 & (10.21) \\
P(N,t) &= 0, \text{ for } N \to \infty & (10.22)
\end{aligned}
$$

Solving this results in

$$P(x,t) = \frac{1}{\sqrt{2\pi t}}\left(\exp\left(\frac{-(x-x_0)^2}{2t} \right) - \exp\left(\frac{-(x+x_0)^2}{2t} \right) \right). \tag{10.23}$$

Remark

Citation (Luo et al., 2008c): "The present model has some limitations: Due to the coarse-grained nature of the model, it is clear that the quantative details of some results presented here depend on the microscopics of the real-world experimental setup. However, dimensionality plays an unimportant role. Regarding the issue of hydrodynamics, recent molecular dynamics and lattice Boltzmann simulation results show that hydrodynamics is screened out in a narrow pore. Finally for the present case where we model single stranded DNA chains, the bending stiffness of the chain is also not expected to play a role. "

Code of the Matlab Program

```
% MASTER THESIS Program 1.0 Chain with nanopore 2D and two different monomers_____

clear all
clc
close all

tic
figure

%INPUT_____
init=1;                         % init=1 if verlet algorithm is used for initial step, else initial step is based on random first placing
beeld=0;                        % beeld=1 when visualization of polymer is required
wacht=0.001;                    % pause during plotting
config=2;                       % config=1 if first configuration is straight line, config=2 if equilibriumconfig is used else random first configuration
nanopore=1;                     % nanopore=1 if there is a nanopore
start=0.5;                      % start=0 when polymer starts at the beginning of the pore, 0.5 when it starts halfway, and 1 when it starts at the end of the pore
equilibrium=0;                  % first let polymer attain equilibrium position
only=0;                         % if only equilibrium needs to be attained, else after equilibriumend, starts translocating
equilibriumend=5000;

nrresults=1;
exponent=0.5;                   % exponent=1 when size polymer equals half size pore, exponent=0.5 when sqrt size polymer equals half size pore
timeloop=32;
straal=3;

c=1;                    %number of simulations
nn=5000;                %number of timesteps
rem=10000;
n=51;                   %number of particles
xload=load('xeq51');
%vload=load('veq51');
simulationend=5*n^3;

%Random sequence
aapje=randperm(n);
aapjea=aapje(1:round(n/2));
aapjeb=aapje((round(n/2)+1):n);
mona=sort(aapjea);
mont=zeros(n,1);
mont(mona')=1;
monb=sort(aapjeb);          %distribution monomer b

%Or predefined sequence

% mona=[(1:1:n)];           %distribution monomer a
% mont=zeros(n,1);
% mont(mona')=1;
% monb=[(0:0:n)];           %distribution monomer b

varpa=3;                    %attraction monomer a
varpb=1;                    %attraction monomer b

length=1.5;                 %diameter porie
width=1;                    %breedte porie


tijd=zeros(n,1);

%CONSTANTEN_____

sigma=1;                    %diameter of monomer
m=1;                        %mass of monomer
T=0.9;                      %temperature
R=2;                        %maximum allowed seperation between connected monomers
R2=R^2;
k=7;                        %spring constant
kb=1.2/T;
xi=0.7;
var=1.2;                    %energy scale

xb=(n-1)*6;                     %square box!!!!
yb=xb;
height=xb/2-(length/2+0.5);
heightnonv=floor((n-1)^exponent)-(length/2+0.5);
xbnonv=floor(((n-1)^exponent)*2);

rc=((2^(1/6))*sigma);       %cut-off distance
rcp=(2.5*sigma);            %cut-off distance attracting pore
rc2=rc^2;
rcp2=rcp^2;

delta=0.001;                %time interval

Fextn=2;
Fexte=0;
if nanopore==1
    Fextc=Fextn;                %external force
else
    Fextc=Fexte;
end

Fext=zeros(n,2);
Fexta=zeros(n,2);
Fextv=zeros(n,2);

diagonal=diag(ones(n-1,1),1);
diagonal=diagonal(1:end-1,:);
diagonal=find(diagonal);

%PORIE_____
if nanopore==1
    xx=[zeros(height,1),(length/2+1.5:1:xb/2)'];
    xx=[xx; xx(:,1) -1*xx(end:-1:1,2)];
    if width~=1
        xx=[xx; xx(:,1)+width-1 xx(:,2)];
    end
```

```
    sizexx=size(xx,1);
    xx=[xx; (0:1:width-1)' (length/2+0.5)*ones(width,1); (0:1:width-1)' -1*(length/2+0.5)*ones(width,1)];
    wand=(1:1:sizexx(1));
    sizeporie=size(xx,1);
    porie=(sizexx(1)+1:1:sizeporie(1));
    xxwand=xx(wand,:);
    xxporie=xx(porie,:);

    %nanopore visualization is previous code, now let's reduce the pore
    xx=[zeros(heightnonv,1),(length/2+1.5:1:xbnonv/2)'];
    xx=[xx; xx(:,1) xx(:,2)-(xbnonv/2+ceil(length/2)+1)];
    if width~=1
        xx=[xx; xx(:,1)+width-1 xx(:,2)];
    end
    sizexx=size(xx,1);
    xx=[xx; (0:1:width-1)' (length/2+0.5)*ones(width,1); (0:1:width-1)' -1*(length/2+0.5)*ones(width,1)];
    sizeporie=size(xx,1);


    %PREPARE RESULTS_____
    F=zeros(c,nrresults);
    tau=zeros(4,1);
    if equilibrium==1
        GRA=zeros(equilibriumend/rem,c);
        GRB=zeros(equilibriumend/rem,c);
    end
else

    POT=zeros(nn/rem,c);
    CM=POT;
    GR=POT;
    E=POT;
    TC=POT;
    M=POT;
end

cc=1;

for b=1:c

    %INIT_____
    if config==1;
        x=[(-n:1:-1)'+start*(n+width)+1-2*start,zeros(n,1)];
        rd=zeros(n-1,n);
        for i=1:(n-1)
            for j=(i+1):n
                xd=abs(x(i,1)-x(j,1));
                yd=abs(x(i,2)-x(j,2));
                rd(i,j)=((xd)^2+(yd)^2)^0.5;
            end
        end
        v=2*(rand(n,2)-0.5);
    elseif config==2
        x=xload.x;
        v=2*(rand(n,2)-0.5);
        rd=zeros(n-1,n);
        for i=1:(n-1)
            for j=(i+1):n
                xd=abs(x(i,1)-x(j,1));
                yd=abs(x(i,2)-x(j,2));
                rd(i,j)=((xd)^2+(yd)^2)^0.5;
            end
        end
    else
        initdummy=1;
        rd=zeros(n-1,n);
        while initdummy>0;
            x=-1*rand(n,2)-0.5*sigma;
            dummy=tril(ones(n));
            x=dummy*x;
            x=x+start*n;

            tel=start*(n+width)+1-2*start;
            x(tel,:)=[0 0];

            if start==0.5
                x(tel+1,:)=[-1 -1];
                x(tel-1,:)=[1 1];
            elseif start==0
                x(tel+1,:)=[-1 -1];
            else
                x(tel-1,:)=[1 1];
            end
            initdummy=0;
            for i=1:n-1
                for j=(i+1):n
                    xd=abs(x(i,1)-x(j,1));
                    yd=abs(x(i,2)-x(j,2));
                    rd(i,j)=((xd)^2+(yd)^2)^0.5;
                    if i==j-1
                        if rd(i,j)>R
                            initdummy=1;
                        end
                    end
                end
            end

        end
        x=mod((x+(xb/2)),xb)-xb/2;
        x=x(n:-1:1,:);
        v=2*(rand(n,2)-0.5);
    end

    r2=rd.^2;

    update=find(r2<straal*rc2 & r2>0);
    sup=size(update);


    %Plot initial distribution
```

```
if beeld==1;
    xmona=x(mona,:);
    xmonb=x(monb,:);
    plot(x(:,1),x(:,2),'k-');
    hold on
    plot(xmona(:,1),xmona(:,2),'o','MarkerSize',10,'MarkerEdgeColor','red','MarkerFaceColor','none');
    plot(xmonb(:,1),xmonb(:,2),'o','MarkerSize',10,'MarkerEdgeColor','blue','MarkerFaceColor','none');
    if nanopore==1
        plot(xxporie(:,1),xxporie(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','black');
        plot(xxwand(:,1),xxwand(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','none');
    end
    axis equal
    axis([-xb/(2*(n-1)/10),xb/(2*(n-1)/10),-yb/(2*(n-1)/10),yb/(2*(n-1)/10)]);
end

fout=0;
count=0;

%FORCE_____
a=1;
stapjes=0;
tijdn=zeros(n,1);
monomer=(1-start)*(n+width)+1-2*(1-start);

if nanopore==1
    %conditie=(x(1,1)<=width-0.5 & x(end,1)>=0-0.5 & fout==0);
    conditie=(monomer~=0 & fout==0);
    Fextc=Fextn;
    if equilibrium==1
        Fextc=Fexte;
        if only==1
            conditie=(a<=equilibriumend);
        end
    end
else
    conditie=(a<=nn);
end

while conditie

    energie=0;

    f=zeros(n,2);
    f2=f;
    fp=f;
    fp2=f;

    GULJ=zeros(n-1,n);
    GUFENE=GULJ;
    %          GUFENE2=GULJ;

    if nanopore==1
        GULJP=zeros(n,sizeporie(1));
        GULJP2=zeros(n,sizeporie(1));
        monat=mona;

        xaap1=repmat(x(:,1),1,sizeporie);
        xxaap1=repmat(xx(:,1)',n,1);
        xaap2=repmat(x(:,2),1,sizeporie);
        xxaap2=repmat(xx(:,2)',n,1);
        xdpm=abs(xaap1-xxaap1);
        ydpm=abs(xaap2-xxaap2);

        rdp=((xdpm).^2+(ydpm).^2).^0.5;
        rdp2=rdp.^2;

        closem=find(rdp2(:,1:sizexx)<=rc2);
        %ULJ(i,j)=4*var*[(sigma/rd(i,j))^12-(sigma/rd(i,j))^6]+var;
        energie=energie+sum(4.*var.*(sigma./rdp(closem)).^12-(sigma./rdp(closem)).^6+var);
        GULJP(closem)=48.*var.*(1./rdp(closem)).*((sigma./rdp(closem)).^6).*(((sigma./rdp(closem)).^6)-0.5);

        XY=[n,sizexx];
        [X,Y] = ind2sub(XY,closem);
        sizeclosem=size(closem,1);

        for ii=1:sizeclosem
            fp(X(ii),1)=fp(X(ii),1)+GULJP(closem(ii))*(x(X(ii),1)-xx(Y(ii),1))/abs(rdp(closem(ii)));
            fp(X(ii),2)=fp(X(ii),2)+GULJP(closem(ii))*(x(X(ii),2)-xx(Y(ii),2))/abs(rdp(closem(ii)));
        end


        poreclosem=find(rdp2(:,sizexx+1:sizeporie)<=rcp2)+sizexx*n;
        XY2=[n,sizeporie-sizexx];
        [X2,Y2] = ind2sub(XY2,poreclosem);
        sizeporeclosem=size(poreclosem,1);

        poreclosema=poreclosem(ismember(X2, mona));
        energie=energie+sum(4.*varpa.*(sigma./rdp(poreclosema)).^12-(sigma./rdp(poreclosema)).^6+varpa);
        GULJP(poreclosema)=48.*varpa.*(1./rdp(poreclosema)).*((sigma./rdp(poreclosema)).^6).*(((sigma./rdp(poreclosema)).^6)-0.5);

        poreclosemb=poreclosem(ismember(X2, monb));
        energie=energie+sum(4.*varpb.*(sigma./rdp(poreclosemb)).^12-(sigma./rdp(poreclosemb)).^6+varpb);
        GULJP(poreclosemb)=48.*varpb.*(1./rdp(poreclosemb)).*((sigma./rdp(poreclosemb)).^6).*(((sigma./rdp(poreclosemb)).^6)-0.5);


        for ii=1:sizeporeclosem
            fp(X2(ii),1)=fp(X2(ii),1)+GULJP(poreclosem(ii))*(x(X2(ii),1)-xx(Y2(ii),1))/abs(rdp(poreclosem(ii)));
            fp(X2(ii),2)=fp(X2(ii),2)+GULJP(poreclosem(ii))*(x(X2(ii),2)-xx(Y2(ii),2))/abs(rdp(poreclosem(ii)));
        end


        externek=find(x(:,1)<=(0+width) & x(:,1)>=-1);
        Fext=Fextv;
        Fext(externek,1)=Fextc;

    end
```

```
updatespec=find(r2<=rc2 & r2>0);
energie=energie+sum(4.*var.*(sigma./rd(updatespec)).^12-(sigma./rd(updatespec)).^6+var);
GULJ(updatespec)=48.*var.*(1./rd(updatespec)).*((sigma./rd(updatespec)).^6).*(((sigma./rd(updatespec)).^6)-0.5);

IJ = [n-1,n];

updatespec=unique([updatespec; diagonal]);
sups=size(updatespec,1);

[I,J] = ind2sub(IJ,updatespec);

energie=energie+sum(-1.*0.5.*k.*R2.*log(1-(rd(diagonal).^2./R2)));
GUFENE(diagonal)=0.5.*k.*R^2.*(1./(1-(rd(diagonal).^2./(R^2)))).*-2.*rd(diagonal)./(R^2);

xvare=(x(I,1)-x(J,1))./abs(rd(updatespec));
xvart=(x(I,2)-x(J,2))./abs(rd(updatespec));

for ii=1:sups
    f(I(ii),1)=f(I(ii),1)+GULJ(updatespec(ii)).*xvare(ii)+GUFENE(updatespec(ii)).*xvare(ii);
    f(J(ii),1)=f(J(ii),1)-1.*GULJ(updatespec(ii)).*xvare(ii)-1.*GUFENE(updatespec(ii)).*xvare(ii);
    f(I(ii),2)=f(I(ii),2)+GULJ(updatespec(ii)).*xvart(ii)+GUFENE(updatespec(ii)).*xvart(ii);
    f(J(ii),2)=f(J(ii),2)-1.*GULJ(updatespec(ii)).*xvart(ii)-1.*GUFENE(updatespec(ii)).*xvart(ii);
end


Fr=(1/(sqrt(delta)))*randn(n,2)*(2*kb*T*xi)^0.5;

f=f+fp+Fext-xi*v+Fr;

if a==1;
    if init==1
        xo=x+v*delta+(delta^2/(2*m))*f;
        xo=mod((xo+(xb/2)),xb)-xb/2;
    else
        xo=x;
    end
    xcmo=sum(m.*x)/sum(n*m);
end

xn=2*x-xo+delta^2*f/m;

if equilibrium==1 && a<=equilibriumend
    xn(n*(1-start)+start,:)=[0 0];
    Fextc=Fexte;
else
    Fextc=Fextn;
end

if mod(a,timeloop)==0

    for i=1:(n-1)
        for j=(i+1):n

            xd=abs(xn(i,1)-xn(j,1));
            if xd>(xb/2)
                xd=xb-xd;
            end

            yd=abs(xn(i,2)-xn(j,2));
            if yd>(yb/2)
                yd=yb-yd;
            end

            rd(i,j)=((xd)^2+(yd)^2)^0.5;
            if i==j-1
                if rd(i,j)>R
                    fout=1;
                    disp(' fout ')
                    f;

                end

            end
        end
    end
    r2=rd.^2;
    update=find(r2<straal*rc2 & r2>0);
    sup=size(update,1);
else
    for ii=1:sups

        xd=abs(xn(I(ii),1)-xn(J(ii),1));
        if xd>(xb/2)
            xd=xb-xd;
        end

        yd=abs(xn(I(ii),2)-xn(J(ii),2));
        if yd>(yb/2)
            yd=yb-yd;
        end

        rd(updatespec(ii))=((xd)^2+(yd)^2)^0.5;
        if i==j-1
            if rd(updatespec(ii))>R
                fout=1;
                disp(' fout ')
                f;
                f2;
            end

        end
    end
    r2=rd.^2;
end

%periodic boundary conditions
%rd=rd-xb*round(rd/xb);

xn=mod((xn+(xb/2)),xb)-xb/2;

%CHECK_____




energie=energie+sum(-1.*0.5.*k.*R2.*log(1-(rd(diagonal).^2./R2)));
```

```
v=(xn-xo)/(2*delta);

%UPDATE_____

xo=x;
x=xn;
stapjes=stapjes+1;
if x(monomer,1)>width
    tijdn(monomer)=stapjes*delta;
    monomer=monomer-1;
    stapjes=0;
end

if nanopore~=1
    if mod(a,rem)==0

        POT(a/rem,cc)=energie;

        M(a/rem,cc)=sum(sum(m*v));

        TC(a/rem,cc)=(sum(sum(m*(v.^2))))/(2*n*kb);

        E(a/rem,cc)=sum((m/2)*sum(v.^2));

        xcm=sum(m.*x)/sum(n*m);
        CM(a/rem,cc)=(sum((xcm-xcmo).^2))^0.5;

        xm=repmat(((1/n)*sum(x)),n,1);
        GR(a/rem,cc)=(1/n)*sum(sum((x-xm).^2));
    end
end

if equilibrium==1
    if mod(a,rem)==0
        nreq=(n-1)/2;
        xma=repmat(((1/nreq)*sum(x(1:nreq,:))),nreq,1);
        GRA(a/rem,cc)=(1/nreq)*sum(sum((x(1:nreq,:)-xma).^2));
        xmb=repmat(((1/nreq)*sum(x((nreq+2):end,:))),nreq,1);
        GRB(a/rem,cc)=(1/nreq)*sum(sum((x((nreq+2):end,:)-xmb).^2));
    end
end

if beeld==1;

    xmona=x(mona,:);
    xmonb=x(monb,:);
    plot(x(:,1),x(:,2),'k-');
    hold on
    plot(xmona(:,1),xmona(:,2),'o','MarkerSize',10,'MarkerEdgeColor','red','MarkerFaceColor','none');
    plot(xmonb(:,1),xmonb(:,2),'o','MarkerSize',10,'MarkerEdgeColor','blue','MarkerFaceColor','none');

    if nanopore==1
        plot(xxwand(:,1),xxwand(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','none');
        plot(xxporie(:,1),xxporie(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','black');
    end

    hold off
    axis equal
    axis([-xb/(2*(n-1)/10),xb/(2*(n-1)/10),-yb/(2*(n-1)/10),yb/(2*(n-1)/10)]);
    pause(wacht)
end


a=a+1;

if nanopore==1
    %conditie=(x(1,1)<=width-0.5 & x(end,1)>=0-0.5 & fout==0);
    conditie=(monomer~=0 & fout==0);
    if only==1
        conditie=(a<=equilibriumend);
    end
else
    conditie=(a<=nn);
end
end

countdummy=0;
if sum(x(:,1)<0)~=0
    count=count+1;
    countdummy=1;
end

F(cc)=a*delta;
if fout~=1 && monomer==0 && countdummy~=1
    cc=cc+1;
    tijd=tijd+tijdn;
end

if monomer~=0
    figure
    xmona=x(mona,:);
    xmonb=x(monb,:);
    plot(x(:,1),x(:,2),'k-');
    hold on
    plot(xmona(:,1),xmona(:,2),'o','MarkerSize',10,'MarkerEdgeColor','red','MarkerFaceColor','none');
    plot(xmonb(:,1),xmonb(:,2),'o','MarkerSize',10,'MarkerEdgeColor','blue','MarkerFaceColor','none');

    if nanopore==1
        plot(xxwand(:,1),xxwand(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','none');
        plot(xxporie(:,1),xxporie(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','black');
    end

    hold off
    axis equal
    axis([-xb/(2*(n-1)/10),xb/(2*(n-1)/10),-yb/(2*(n-1)/10),yb/(2*(n-1)/10)]);
end

b;
disp(b)
disp(monomer)
```

```
    disp(count)

end

cc=cc-1;
disp(' Aantal simulaties ')
disp(cc)

if equilibrium==1
    GRAT=sum(GRA,2)/cc;
    figure
    subplot(2,1,1)
    plot((1:rem:nn)*delta, GRAT)
    title('Gyration Radius Links');
    GRBT=sum(GRB,2)/cc;
    subplot(2,1,2)
    plot((1:rem:nn)*delta, GRBT)
    title('Gyration Radius Rechts');
    figure
    xmona=x(mona,:);
    xmonb=x(monb,:);
    %subplot(2,2,1);
    plot(x(:,1),x(:,2),'k-');
    hold on
    plot(xmona(:,1),xmona(:,2),'o','MarkerSize',10,'MarkerEdgeColor','red','MarkerFaceColor','none');
    plot(xmonb(:,1),xmonb(:,2),'o','MarkerSize',10,'MarkerEdgeColor','blue','MarkerFaceColor','none');

    if nanopore==1
        plot(xxwand(:,1),xxwand(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','none');
        plot(xxporie(:,1),xxporie(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','black');
    end

    hold off
    axis equal
    axis([-xb/(2*(n-1)/10),xb/(2*(n-1)/10),-yb/(2*(n-1)/10),yb/(2*(n-1)/10)]);

elseif nanopore==1
    figure
    hist(F(:,nrresults),round(b/5))
    save('F51rand1.mat','F')
    figure
    tijd=tijd/cc;
    tijd=tijd(n:-1:1);
    save('tijd51rand1.mat','tijd')
    plot(tijd)
    figure
    xmona=x(mona,:);
    save('mona51rand1.mat','mona')
    xmonb=x(monb,:);
    plot(x(:,1),x(:,2),'k-');
    hold on
    plot(xmona(:,1),xmona(:,2),'o','MarkerSize',10,'MarkerEdgeColor','red','MarkerFaceColor','none');
    plot(xmonb(:,1),xmonb(:,2),'o','MarkerSize',10,'MarkerEdgeColor','blue','MarkerFaceColor','none');

    if nanopore==1
        plot(xxwand(:,1),xxwand(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','none');
        plot(xxporie(:,1),xxporie(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','black');
    end

    hold off
    axis equal
    axis([-xb/(2*(n-1)/10),xb/(2*(n-1)/10),-yb/(2*(n-1)/10),yb/(2*(n-1)/10)]);
else
    CMT=sum(CM,2)/cc;
    figure
    subplot(2,2,1)
    xmona=x(mona,:);
    xmonb=x(monb,:);
    plot(x(:,1),x(:,2),'k-');
    hold on
    plot(xmona(:,1),xmona(:,2),'o','MarkerSize',10,'MarkerEdgeColor','red','MarkerFaceColor','none');
    plot(xmonb(:,1),xmonb(:,2),'o','MarkerSize',10,'MarkerEdgeColor','blue','MarkerFaceColor','none');

    if nanopore==1
        plot(xxwand(:,1),xxwand(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','none');
        plot(xxporie(:,1),xxporie(:,2),'o','MarkerSize',10,'MarkerEdgeColor','black','MarkerFaceColor','black');
    end

    hold off
    axis equal
    axis([-xb/(2*(n-1)/10),xb/(2*(n-1)/10),-yb/(2*(n-1)/10),yb/(2*(n-1)/10)]);

    subplot(2,2,2)
    plot((1:rem:nn)*delta, CMT)
    title('Centre of Mass');
    GRT=sum(GR,2)/cc;

    subplot(2,2,3)
    plot((1:rem:nn)*delta, GRT)
    title('Gyration Radius');
    TCT=sum(TC,2)/cc;

    subplot(2,2,4)
    plot((1:rem:nn)*delta, TCT)
    hold on

    title('Temperature');
    MT=sum(M,2)/cc;
    figure

    subplot(2,2,1)
    plot((1:rem:nn)*delta, MT)
    title('Momentum');
    ET=sum(E,2)/cc;

    subplot(2,2,2)
    plot((1:rem:nn)*delta, ET)
    title('Kinetic Energy');
    POTT=sum(POT,2)/cc;

    subplot(2,2,3)
```

```
    plot((1:rem:nn)*delta, POTT)
    title('Potential Energy');
    TOTEN=POTT+ET;

    subplot(2,2,4)
    plot((1:rem:nn)*delta, TOTEN)
    title('Total Energy');
end

t2=toc;
tijdinmin=t2/60;
disp(' tijd ')
disp(tijdinmin)
```