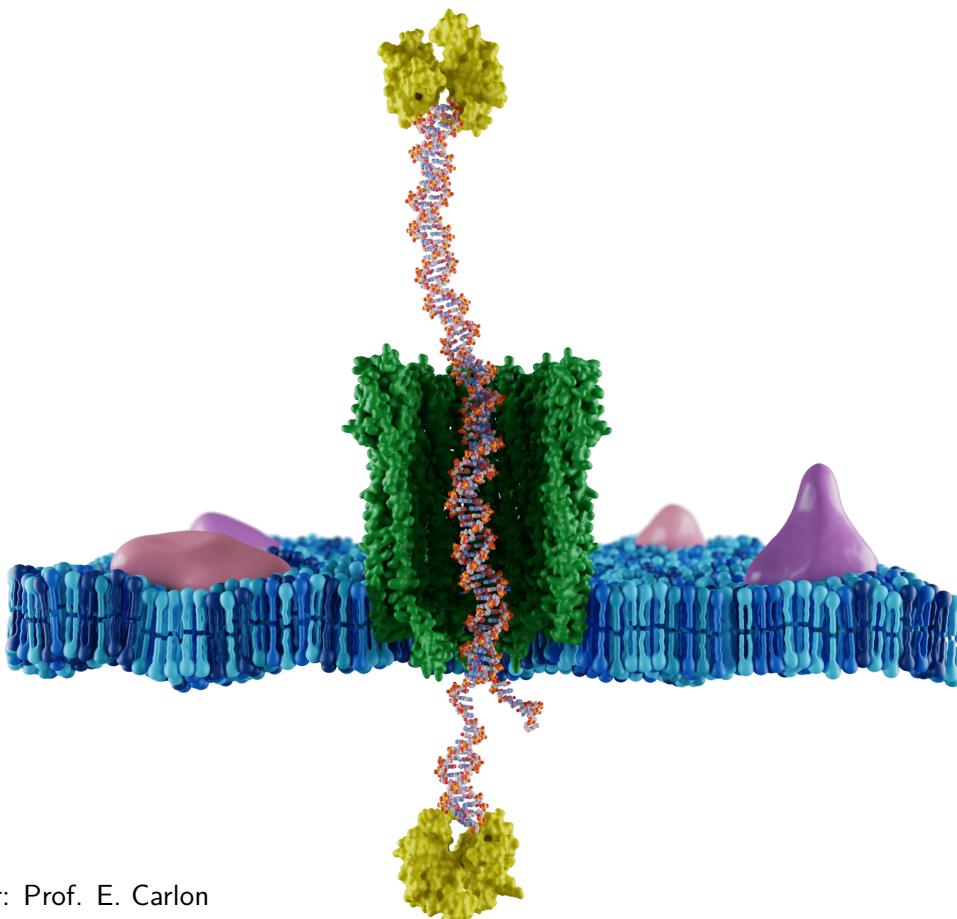


# Coarse-grained simulations of the DNA nanopiston



**Jan Stevens**

Supervisor: Prof. E. Carlon

Thesis presented in  
fulfillment of the requirements  
for the degree of Master of Science  
in Physics

Academic year 2020-2021



© Copyright by KU Leuven

Without written permission of the promoters and the authors it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to KU Leuven, Faculteit Wetenschappen, Geel Huis, Kasteelpark Arenberg 11 bus 2100, 3001 Leuven (Heverlee), Telephone +32 16 32 14 01. A written permission of the promoter is also required to use the methods, products, schematics and programs described in this work for industrial or commercial use, and for submitting this publication in scientific contests.



# **Abstract**

abstract



# Vulgariserende Samenvatting

Summary in dutch.

asdf



# **Summary in Layman's Terms**

Summary in english.







# List of Figures

1.1	write caption . . . . .	2
1.2	This is a figure . . . . .	4
1.3	This is a figure . . . . .	5
1.4	caption nog maken . . . . .	6
1.5	Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis)[.]	10
1.6	This is a figure . . . . .	13
2.1	This is a figure [.] . . . . .	16
2.2	This is a figure . . . . .	17
2.3	This is a figure . . . . .	18
3.1	Structure of the OxDNA model with the different interactions. Figure was taken form []. (H.B. is hydrogen bond) . . . . .	19
3.2	This is a figure [.] . . . . .	20
3.3	write caption[.] . . . . .	22
3.4	write caption[.] . . . . .	24
4.1	asdf asdf asdf asdf asd fasd fasdf asf asdfasdfa asdfasdf asfasf. asdf asdf asdf asdf asd fasd fasdf asf asdfasdfa asdfasdf asfasf. asdf asdf asdf asd fasd fasdf asf asdfasdfa asdfasdf asfasf. asdf asdf asdf asd fasd fasdf asf asdfasdfa asdfasdf asfasf. . . . .	28
4.2	write caption . . . . .	29
4.3	write caption . . . . .	29



# **List of Tables**



# Contents

<b>Abstract</b>	<b>i</b>
<b>Vulgariserende Samenvatting</b>	<b>iii</b>
<b>Summary in Laymans's Terms</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>Contents</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Thesis outline . . . . .	1
1.2 Biological Nanopores . . . . .	3
1.3 Deoxyribonucleic acid (DNA) . . . . .	6
1.4 Polymer Physics . . . . .	7
1.5 Computer Simulations . . . . .	10
<b>2 The DNA Nanopiston</b>	<b>15</b>
2.1 Rotaxane Formation . . . . .	15
2.2 Operating principles . . . . .	15
2.3 Coarse-grained simulations . . . . .	16
<b>3 Adapting the Model</b>	<b>19</b>
3.1 OxDNA . . . . .	19
3.2 DNA Thermodynamics . . . . .	20
3.3 Forward Flux sampling . . . . .	22
3.4 Simulation technique . . . . .	25
<b>4 Simulations of the Rotaxane</b>	<b>27</b>
4.1 Mixed Rotaxane . . . . .	27
4.2 Conformational Fluctuations of the Rotaxane . . . . .	27
4.3 Toehold Displacement Reaction . . . . .	27
<b>5 Conclusions and Perspectives</b>	<b>31</b>
5.1 Results & Conclusions . . . . .	31
5.2 Future Perspectives . . . . .	31
<b>A 1D Confined Diffusion</b>	<b>33</b>
<b>Acknowledgements</b>	<b>35</b>



# Introduction

*...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.*

---

— Richard P. Feynman, *The Feynman Lectures on Physics*<sup>7</sup>

## 1.1 Thesis outline

All organisms in nature tirelessly perform work, struggling against an ever increasing entropy. This work is collectively performed by countless molecular machines, all contributing to their specific tasks.

Despite being so abundantly present in nature, fabricating synthetic molecular machines turns out to be a difficult task. One of the biggest hurdles in this process arises from the length-scale of these machines. Often times these structures are not larger than a few nanometres, making the typical energy associated with the bonds and distortions of their structure comparable to thermal energy fluctuations.

As a result, the random thermal fluctuations in the environment of these molecular machines induce the stochastic motion that complicates their functioning. Extracting useful work from this Brownian motion, performed by the freely tumbling structures, is almost impossible. To overcome this limitation, most synthetic molecular machines are embedded in a larger complex providing necessary structure.

This phenomenon is also observed in nature, for instance in the interfacing of protein complexes with the phospholipid bilayer of cells. A widely known example is the bacterial flagella motor, which provides an efficient way for bacteria to roll and tumble through their environment. Just like in electrical motors, the flagella consists of a stator and a rotor. The

## 1. INTRODUCTION

---

stator is anchored into the cell membrane, while the rotor is allowed to freely rotate. The work is produced by the flow of cations through the stator, inducing changes in the electrostatic interactions between the two parts of the flagella, generating unidirectional motion.

Analogues to macroscopic engines, during the operation of molecular machines, heat is produced. When the structure is not capable of dissipating this heat efficiently, an excessive build up limits the life cycle of the complex. To mitigate this problem, often times large and soft molecules are used in the design of nanomachines. A logical choice is the use of polymers, which can effectively dissipate heat as a result of their flexibility. Due to the programmability of DNA, using the Watson-Crick interactions, the DNA polymer provides additional aptitude, making it a popular material in nanotechnology.

The central topic of this thesis is studying the DNA nanopiston, designed by Bayoumi et al., a DNA based molecular machine embedded into a phospholipid membrane. This nanopiston can be characterised as an autonomous molecular machine, which turns over chemical fuel to continuously perform work. The aim of this complex is to perform selective transport of DNA through a membrane. The operation cycle of previously designed DNA transporters require a supporting external bias, where the nanopiston operates also against an external bias. The physics driving this machine is entropy and will be discussed in detail in this thesis.

In the first chapter, an introduction is given into important concepts regarding the DNA nanopiston. Having laid this theoretical foundation, the structure and operation cycle of the DNA nanopiston are discussed in chapter two. Next the simulation model used in this thesis is presented in chapter three. The results of these simulations are finally discussed in the chapter four. The final chapter five of this thesis offers a discussion of the results and recommendations for further research.

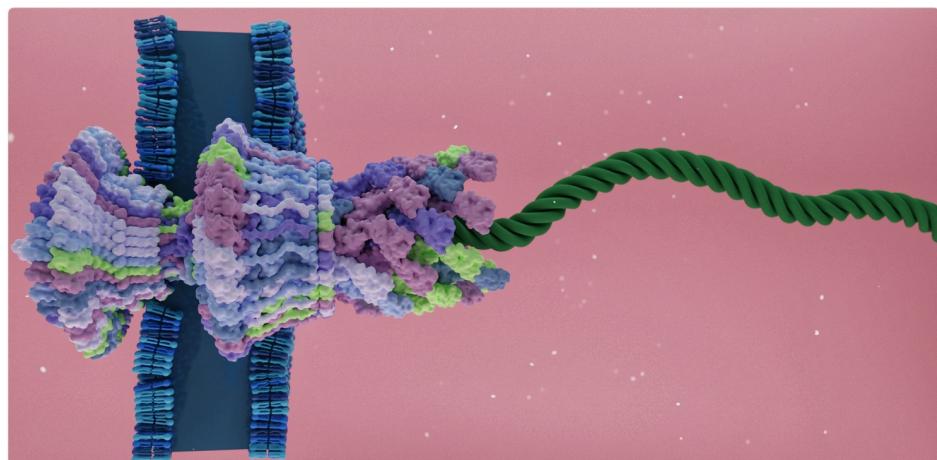


Figure 1.1: write caption

## 1.2 Biological Nanopores

Biological nanopores are small perforations in a lipid bilayer, created by a pore forming protein. The majority of these proteins are toxins produced by pathogenic bacteria. Their function in nature is to perforate the membrane of a cell, causing cell depolarization and inducing an osmotic potential. These effects disrupt vital cell functions or spill its nutrients into the environment, often times resulting in the killing of the cell.

The reason scientists are interested in studying nanopores is related to their size. These protein structures are generally only a few nanometres in diameter, making them comparable in size to the tiny transistors found in modern computers. Working at these small scales has the unavoidable complication, that it becomes difficult to retrieve information from nano scale processes. Developing sensors to probe this exotic length scale is thereby very relevant. This is the exact problem nanopores provide a solution to, i.e. spectroscopy at the smallest scale.

Before delving into ionic current spectroscopy, the primary application of these nanopores, first a brief overview will be given of the structural properties of two popular biological nanopores.

### 1.2.1 $\alpha$ -Hemolysin ( $\alpha$ -HL)

The  $\alpha$ -Hemolysin ( $\alpha$ -HL) protein is the most commonly used pore forming proteins to create biological nanopores. It is produced by the *Staphylococcus aureus*, a bacterium commonly found in human microbiota.

The  $\alpha$ -HL pore(PDBID:...) is an oligomeric complex with multiple naturally occurring variations. The most typical configuration is a heptameric structure, meaning that there are seven protomers found in the complex. The secondary structure elements consist principally of  $\beta$ -sheets, making it a member of the  $\beta$ -barrel pore-forming toxins. Through both electrostatic and hydrophobic interactions, the  $\alpha$ -HL is bound to the membrane of a target cell. Here the monomers assemble to a 'prepore' complex that transitions to the stable pore complex by inserting the  $\beta$ -barrel into the membrane.

Structurally the shape of  $\alpha$ -HL resembles that of a hollow mushroom. The total height of the complex is 11nm and the maximum width is measured to be 10nm. The internal chamber of the pore located at the cis side of the membrane is called the lumen. The lumen of  $\alpha$ -HL is quite constricted measuring a diameter of 3nm. At the membrane, the lumen chamber transitions into a protein stem, referred to as the constriction of the pore further reducing the diameter of the chamber to a minimum of 1.5nm. Over the inside surface of  $\alpha$ -HL the charges are relatively uniformly distributed, which will play an important role in

further applications.

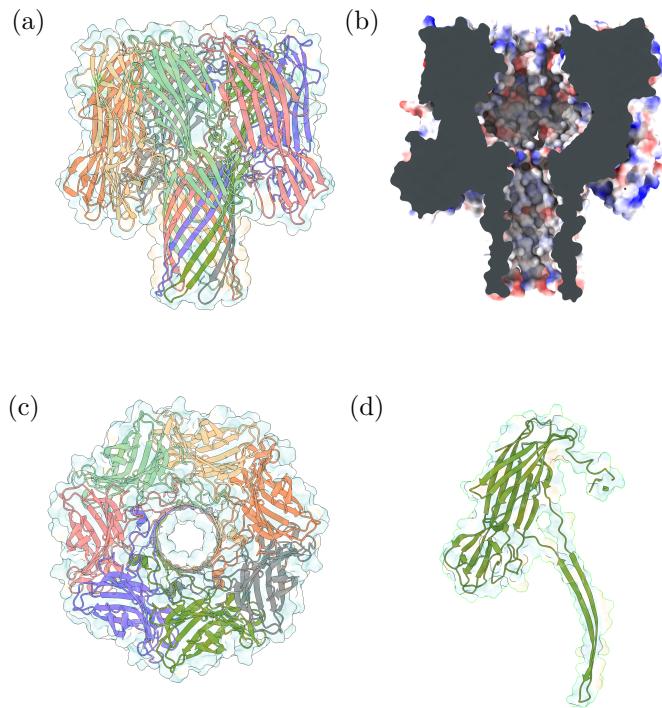


Figure 1.2: This is a figure

### 1.2.2 Cytolysin A (ClyA)

The Cytolysin A (ClyA) is a larger type of pore forming protein, first found to be secreted by *E. coli* strains. The larger size of its lumen allows for different types of applications, compared to smaller complexes like  $\alpha$ -HL. Most relevant for this thesis, is the fact that the larger diameter of the pore's stem allows for translocation of double stranded DNA.

The ClyA pore (PDBID:6MRT[cite]) is an oligomeric complex most typically found in a dodecameric configuration, meaning that there are twelve protomers found in the complex. In nature there are found small variations on this configuration. The secondary structure elements consist principally of  $\alpha$ -helices, making it a member of the  $\alpha$ -pore-forming toxins. The protein formation is induced by the hydrophobic interactions between the  $\beta$ -hairpin and the solvent. The main structural rearrangement in this process consists of swinging out this  $\beta$ -tongue and inserting it into the membrane. After this transition, the membrane-bounded monomers oligomerize to form the final pore structure.

Structurally the shape of ClyA resembles that of two hollow cylinders stacked on top of each other. This cylinder approximation will be important later on in this thesis, where it

will be used to create a simplified model of the nanopore. The total height of the complex is 14nm and the maximum width is measured to be 11nm. The lumen's size of this nanopore differentiates it from the previously discussed  $\alpha$ -HL. The cis entrance of the lumen measures 6nm, while the constricted side of the pore is still 3.6nm in diameter. In contrary to the  $\alpha$ -HL, the inside surface of ClyA has a net negative charge, making it cation sensitive. This excess charge will induce a significant coulomb interaction between the pore and negatively charged analytes.

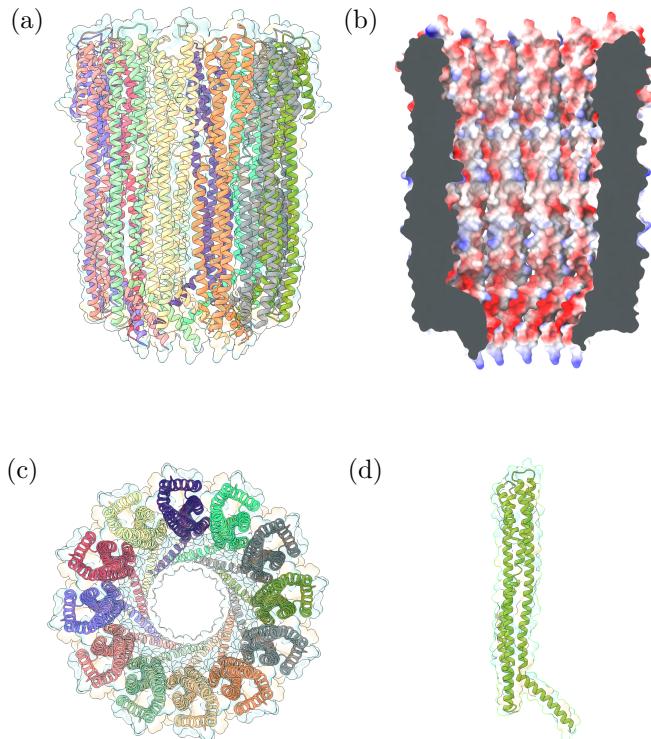


Figure 1.3: This is a figure

### 1.2.3 Ionic current spectroscopy

In recent years the study of nanopores became a popular research domain, mainly due to the development of the nanopore-based ionic current spectroscopy. For the case of biological nanopores, this method depicted in figure ... A lipid bilayer is perforated using a pore forming protein, for example  $\alpha$ -HL. The membrane separates two compartments filled with a saline solution. When a potential difference is created over the membrane, the nanopore mediates an ion current between the two liquid-filled compartments.

This ion current through the pore can accurately be measured. If the pore is empty we refer to the measured current as the open pore current. However the applied electric field also induces forces upon analytes dissolved in the liquid. The net result of these interactions is a flux of analytes towards and in some cases through the nanopore. Analytes located inside

## 1. INTRODUCTION

---

of the nanopore partially block the ion current through the pore, reducing the measured current. Using machine learning algorithms, the time series of these current fluctuation can be measured and identified with particular analytes in the solution. These methods are so precise, that they allow for single cell spectroscopy.

It should be noted that besides these biological nanopores, there are also inorganic nanopores under development. An example of inorganic nanopores are solid state nanopores, created by making perforations in a semi-conductor wafer. While currently not as accessible as biological nanopores, mainly due to their high production cost, this method has some major advantages. First of all the material properties provide a chemical robustness not present in biological nanopores. The production process also allows for easy scalability and customisability. While currently not as widely used as biological nanopores, due to their customisability and robustness, solid state nanopores will prove to be an important asset in future nanotechnology.

### 1.3 Deoxyribonucleic acid (DNA)

Deoxyribonucleic acid (DNA) is a long biopolymer composed of two strands, commonly found in its characteristic double helix structure. DNA is most famously known for storing the genetic code of organisms in the nucleus of their cells. The existence of this genetic code was already postulated by the Greek philosopher Aristotle. He developed a heredity theory, based upon "blueprints", in which he tried to explain why physical traits were passed on from generation to generation. This theory would go unnoticed until in 1869 Friedrich Meicher discovered a new microscopic substance found on discarded surgical bandages. He would call this substance "Nuclein" since it originated from the nucleus of the cell. Later it was found that this new substance, currently known as "Deoxyribonucleic acid" or DNA, plays an important role as a blueprint for the perpetuation of living matter.

The structure of DNA was first determined by Rosalind Franklin using X-ray crystallography. Later this research was published by Watson and Crick [1], who concluded that DNA consists of two individual strands, coiled around each other in a double helical structure. Each strand is a chain of monomers, which we call nucleotides. A nucleotide is made up of a deoxyribose sugar, phosphate group and one of four nitrogenous bases: cytosine(C), guanine(G), adenine(A) or thymine(T). The covalent bonds that give both strands structure are formed between consecutive phosphate groups, which together make up the DNA backbone. To form the double helix, two backbones are held together by selective hydrogen bonds occurring between corresponding bases of opposing strands. These dipole interactions give

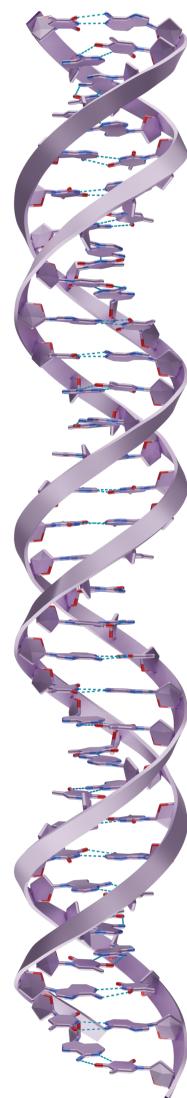


Figure 1.4: caption nog maken

rise to a selection rule, forming only A-T and C-G base pairs.

Since the binding of the two strands is mediated by hydrogen bonding, association and dissociation is possible. The study of these processes is called DNA thermodynamics. The dissociation process of double stranded DNA (dsDNA) is called DNA melting, resulting in two individual strands of single stranded DNA (ssDNA). The reverse process is called DNA hybridisation, which is the selective binding of complementary nucleotides to form dsDNA.

The double helix structure of DNA comes in three different types, B-DNA, A-DNA and Z-DNA, all having a slightly different geometric arrangement. In nature the B-form is most commonly observed, which is characterised by a right-handed helix and the coplanarity between the complementary bases as shown in Fig. ... . A helical twist of B-DNA consists of around 10 base pairs, having a net helical pitch of  $0.34\text{nm}$ . During this thesis, when analysing DNA we refer to the B-DNA form.

When studying DNA the statistical theory of polymer physics is a useful tool. An atomistic resolution is not always needed to accurately describe processes involving relatively long DNA strands. Reducing the complexity of DNA to the monomer level is often justified, allowing us to use more general results in polymer physics.

## 1.4 Polymer Physics

A polymer is a biomolecule made up of building blocks called monomers, linked together to form a chain. The configuration of this chain is determined by the position vector of each monomer, denoted as  $\{\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_N\}$ . The link between each consecutive pair of monomers is called the bond-vector, defined as  $\mathbf{u}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$ . During this discussion we will assume these bonds to be inextensible, i.e. having a fixed bond length of  $|\mathbf{u}_i| = a$ .

Various different models can be used to describe a polymer. The most simple version is called the Freely Jointed Chain (FJC). This model is an example of an ideal flexible polymer, in which excluded volume interactions or polymer bending rigidity are not taken into account. In this model, it is assumed that each bond-vector is completely uncorrelated with its adjacent bonds. Mathematically this is represented by assigning the bond-vector orientation an uniform probability distribution

$$g(\mathbf{u}) = \frac{1}{4\pi a} \delta(|\mathbf{u}| - a), \quad (1.1)$$

where  $a$  is the fixed bond length.

The above described model provides a relatively accurate description of long polymers. However, the assumption that consecutive monomers are uncorrelated becomes problematic at small length scales. The Kratky-Porod model, or discrete wormlike chain, solves this problem by taking the energetic cost of bending the polymer into account. Mathematically this is

done introducing a bending rigidity between neighbouring bonds in the form of a coupling constant,  $\kappa > 0$ . Each polymer configuration is assigned an energy using the equation,

$$E_{WLC} = -\kappa \sum_{i=1}^N \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_{i+1} = -\kappa \sum_{i=1}^N \cos \theta_i, \quad (1.2)$$

where  $\hat{\mathbf{u}} = \mathbf{u}/a$  is the unit bond-vector and  $\theta_i$  is the angle between neighbouring bond-vectors  $\hat{\mathbf{u}}_i$  and  $\hat{\mathbf{u}}_{i+1}$ . The lowest energy state of this discrete wormlike chain is a straight rodlike configuration, where the bond angles  $\theta_i$  are minimized.

To calculate the bond-vector correlation function, we first determine the partition function,  $Z_{WLC}$ , of the system. Identifying the single monomer contributions, this quantity factorises into a product of single bond-vector partition functions as

$$\begin{aligned} Z_{WLC}(N, T) &= \int_0^\pi \cdots \int_0^\pi d\theta_1 \dots d\theta_N \sin \theta_1 \dots \sin \theta_N e^{\beta \kappa \sum_{i=1}^{N-1} \cos \theta_i} \\ &= \left[ \int_0^\pi d\theta \sin \theta e^{\beta \kappa \cos \theta} \right]^N \\ &= [Z_{WLC}(1, T)]^N, \end{aligned} \quad (1.3)$$

where  $\beta = 1/\kappa_b T$  is the inverse temperature. It rests us to determine the single bond-vector partition function. Carrying out the integration yields the result,

$$Z_{WLC}(1, T) = \int_0^\pi d\theta e^{\beta \kappa \theta} = \frac{2 \sinh(\beta \kappa)}{\beta \kappa}. \quad (1.4)$$

From the found partition function we can now determine the bond-vector correlation function. Using the definition of the partition function, we determine the average cosine of the angle between consecutive bonds to be,

$$\begin{aligned} \langle \cos \theta_{i+1} \rangle &= \frac{\partial \log Z_{WLC}(1, T)}{\partial (\beta \kappa)} \\ &= \frac{1}{\tanh(\beta \kappa)} - \frac{1}{\beta \kappa}. \end{aligned} \quad (1.5)$$

Studying the conformation of polymers is often times done assuming a low temperature or large bending rigidity, where we find that the above expression simplifies. In the limit,  $\beta \kappa \gg 1$ , the lowest order approximation yields,

$$\langle \cos \theta \rangle \approx 1 - \frac{1}{\beta \kappa}. \quad (1.6)$$

Decomposing the bond-vector  $\hat{\mathbf{u}}_{n+1}$  in terms of an orthonormal basis, defined by the normal and tangential directions of the preceding vector  $\hat{\mathbf{u}}_n$ , gives

$$\hat{\mathbf{u}}_{n+1} = \hat{\mathbf{u}}_n \cos \theta_n + \hat{\mathbf{u}}_n^\perp \sin \theta_n. \quad (1.7)$$

This decomposition allows us to express the correlation between distant bond-vectors in terms of the correlation between neighbouring bonds-vectors. The factorisation yields

$$\langle \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_{i+m} \rangle = \langle \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_{i+m-1} \rangle \langle \cos \theta \rangle = \dots = \langle \cos \theta \rangle^m, \quad (1.8)$$

where we used the fact that the sinusoidal terms vanish due to symmetry. Exploring this result in the limit,  $\beta\kappa \gg 1$ , we find the expression

$$\langle \hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_{i+m} \rangle = e^{m \log(1 - \frac{1}{\beta\kappa})} \approx e^{-na/l_p}, \quad (1.9)$$

introducing a new polymer quantity, the bending persistence length

$$l_b \equiv \frac{a\kappa}{k_b T}. \quad (1.10)$$

This general result in polymer physics states that the correlations between bond-vectors is exponentially decreasing. The defined quantity represents the characteristic length scale of the polymer, over which the correlations between bond-vectors is lost.

Two limiting cases can be explored. Firstly, in the case where the persistence length is much larger than the polymer's length,  $l_p \gg na$ , all bond-vectors are correlated, i.e. the polymer approximates a straight rod. For the reverse case, where  $l_p \ll na$ , it can easily be shown that the polymer behaves as a stochastic walk.

The persistence length is a central result in the theory of polymer physics, providing a measurable quantity related to the bending rigidity of a polymer. During the simulations performed in this thesis, the notion of bending persistence length is used to discuss the flexibility of the DNA polymer.

## 1.5 Computer Simulations

The theory of classical mechanics is often regarded as the first major breakthrough in the field of physics. For every aspiring physicist this is still the starting point of their studies. Unfortunately, getting to know these relatively simple laws of nature, leads to the inescapable realisation that these theories are expressed in mathematical formalisms that are only analytically solvable in few idealised scenarios. Applying these formulas to a problem consisting of just more than two particles already leads to practically unsolvable equations.

Although it is often times not possible to find an exact solution to equations related to complex physical systems, finding reasonable approximations to their solution is achievable. One popular method to analyse the dynamics of complex systems is the use of simulations.

Simulations have a rich history within physics and engineering, starting even before the invention of the computer.

An example of one of these mechanical simulations is the Waterloop-kundig Laboratorium, a hydrological laboratory, located in Delft[.]. This laboratory houses a scale model of important Dutch waterways, where the influence of waves on harbours and docks was studied. The simulation provided revolutionary insights into the behaviour of water and played an important part in the design of the famous Delta Works.

Another more relevant example is the use of mechanical simulations to study the structure of water. In the early 20th century physicist J.D. Bernal and his fellow researchers build various ball and stick models of water to analyse the possible 3D configurations of water molecules in a liquid[.]. Their research eventually explained the peculiar physical properties of water from a atomistic perspective. However useful these mechanical simulations turned out to be, the biggest drawback of the method was the extreme cost of labour involved with their construct. As Bernal alluded to in his famous 1962 lecture,

... I took a number of rubber balls and stuck them together with rods of a selection of different lengths ranging from 2.75 to 4 inch. I tried to do this in

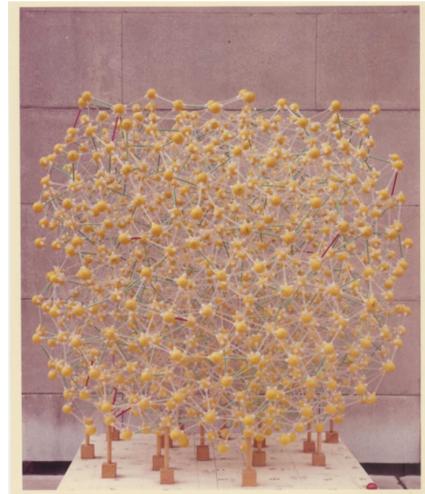


Figure 1.5: Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis)[.]

---

the first place as casually as possible, working in my own office being interrupted every five minutes or so and not remembering what I had done before the interruption...[.]

After the first computer simulations where performed in the Los Alamos labs[.], the popularity of simulations rapidly increased. The remarkable explanatory power of simulations, combined with the relative easy construction of computer models, lead to a fast adoption of computer simulations in the scientific community. Within the context of this thesis, computer simulations are used to study the mechanics of the DNA polymer. Due to the high number of atoms in a typical system, it is generally not possible to find an analytical solution to their equations of motion. In this context, simulations are often used to gain an insight into the complex dynamics of the system and guide the developments of more simple approximate theories. The simulations act as a bridge between the microscopic constituents of the systems and the macroscopic properties we want to understand.

### 1.5.1 Molecular Dynamics Simulations

Molecular Dynamics (MD) is a computer simulation technique, used to analyse the dynamics of a classical many-body system. The central idea of this method is to generate all the trajectories in a system of  $N$  particles by numerically integrating the classical equations of motion,

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{f}_i, \quad \mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} \mathcal{U}_i, \quad \text{for } i \in N.$$

The motion of the particles are governed by the forces  $f_i$  acting upon them, which are usually derived from the interaction potentials  $\mathcal{U}_i$ . Solving these differential equations is achieved by employing a discretized time integration scheme. Algorithm 1 shows the typical structure of a molecular dynamics simulation. The discretization resolution is conventionally called the time step of the simulations denoted by  $\Delta t$ .

There are a large number of different integrations schemes that one can choose from, where the choice depends entirely on the system at hand. When working with an isolated system – i.e. microcanonical ensemble –, logically an energy conserving integrator is needed. The canonical choice for this type of integration scheme is the Velocity-Verlet algorithm. This algorithm is an example of leapfrog integration, where the updating of the positions and velocities are interleaved at different points in time. The major strength of this type of algorithm is that it turns out to be a symplectic integrator, which means the errors on the conserved energy are bounded.

On the other hand, when a system is in contact with a thermal reservoir –i.e. canonical ensemble– not the total energy is conserved, but rather the temperature of the simulation is fixed. To achieve this, a thermostat is implemented in the MD simulation. A typical thermostat attempts to negate any drift in temperature by appropriately importing or exporting energy to the system after each timestep. Popular examples of thermostats are the Nosé-Hoover thermostat or the Langevin thermostat. The latter regulates the temperature by

introducing an implicit solvent to the simulation that gives rise to random thermal kicks. The resulting equations of motion are the Langevin equations given by,

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla\mathcal{U}_i - \gamma_i \frac{d\mathbf{r}_i}{dt} + \xi_i(t), \quad (1.11)$$

where  $\gamma_i$  is known as the friction coefficient and  $\xi_i(t)$  a random force acting upon the particles. The combination of the last two terms fully capture the statistical consequences of the solvent interacting with the system.

---

**Algorithm 1:** The Velocity Verlet algorithm

---

```

Input : Configuration of the system at  $t = 0$ 
1 newList = []
2 /* For odd elements in the list, we add 1, and for even
   elements, we add 2. */ 
3 for  $i \leftarrow 0$  to  $n - 1$  do
4   if isOddNumber( $a_i$ ) then
5     newList.append( $a_i + 1$ ) // Some thought-provoking comment.
6   else
7     // Another comment
8     newList.append( $a_i + 2$ )
9   end if
10 end for
11 return newList
```

---

### 1.5.2 Coarse Grained modelling

As most thing do, molecular dynamics simulations have their pitfalls. A commonly encountered problem is the rapidly increasing computational cost, when the number of particles in the system increase. If not addressed, this would limit the scope of MD simulations to systems of a few particles over short time-scales.

During these simulations the most costly calculations involve the non-bonded interactions in the system. These interatomic interactions make the computational complexity for rudimentary MD simulations scale as  $O(N^2t)$ , where  $N$  is the number of particles in the system and  $t$  the simulation time. This bad scaling behaviour comes from the fact, that for each individual particle all the other particles are contributing to its interaction potential. To improve this scaling behaviour, the non-bonded interactions in a MD simulation are almost always truncated. This localization of the interatomic interactions has the nice effect that not all atoms are involved in every calculation. Efficient algorithms, like the multigrid method[.], have been derived to improve the scaling complexity of MD simulations up to  $O(Nt)$ .

Coarse graining is a method to further optimize molecular dynamics simulations. In contrary to all atom simulations, where each atom is explicitly represented in the simulation, in coarse grained simulations multiple atoms are grouped together to form generalised pseudo-atoms with their respective pseudo-interaction.

There are two distinctly different ways to construct a coarse grained model. The first method starts from the all atom model of the system and generalises nearby atoms into larger pseudo-atoms, this is called the bottom up approach. The second method focuses more on the precise reproduction of experimental results, rather than the precise small scale dynamics. Here larger pseudo-atoms are designed, based upon characteristic patterns in the structure, after which the pseudo-interactions are tweaked to accurately reproduce the system's dynamics.

In the case of DNA simulations, coarse graining turned out to be a very important method. Previous all atom simulations of DNA polymers were restricted to simulations of less than hundred basepairs over only a few microseconds. Studying large scale systems, often encountered in DNA technology, was only possible after the development of coarse grained models. A few examples of commonly used coarse grained models of DNA are Martini[.], 3SPN[.] and oxDNA[.]

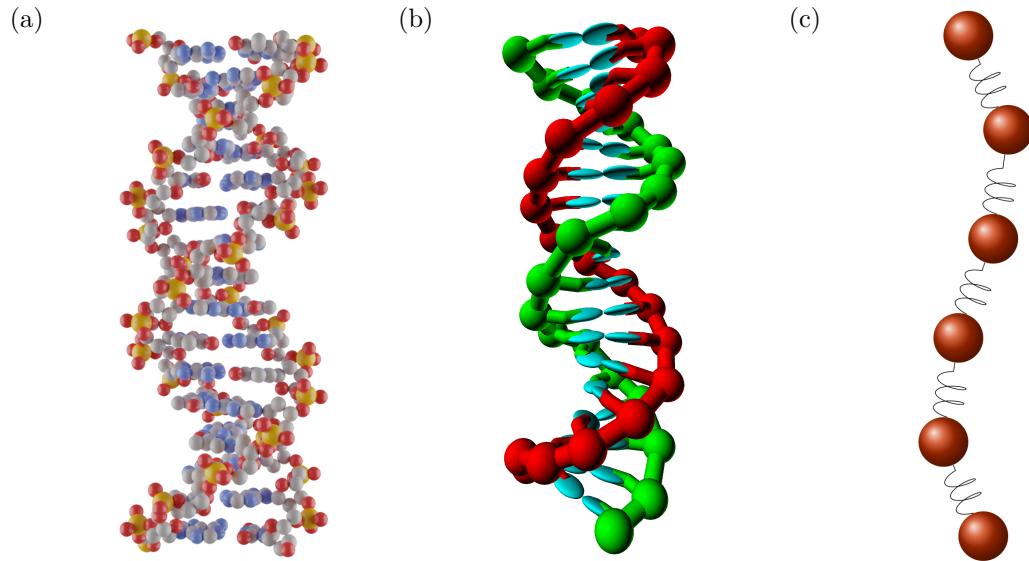


Figure 1.6: This is a figure



# CHAPTER 2

## The DNA Nanopiston

*The career of a young theoretical physicist consists of treating the harmonic oscillator in ever-increasing levels of abstraction.*

---

— Sidney Coleman

Duidelijk duiden dat dit hoofdstuk gebaseerd is op de paper van Stefanos en Bajoumi. Biological transporter. This chapter will discuss the function + operation + computational analysis with beadspring model and LAMMPS

### 2.1 Rotaxane Formation

talk about the constituents. Neutravidin, nanopore, DNA -> toehook.

Formation due to change in voltage. Measure current and infer the state of the complex.

### 2.2 Operating principles

Due to brownian motion both autonomous and non-autonomous systems use ratcheting to eventually deliver work. Using this method minimal synthetic machines can deliver elegance and performance shedding the 'baggage' of biological evolution.

Walk through the operation cycle.

Discuss the importance of entropy. power stroke is toehold disp reaction. Results in two entropy states. This entropic energy provides the recovery stroke. Next hybridisation with the fuel strand makes us come back to the initial configuration.

## 2. THE DNA NANOPISTON

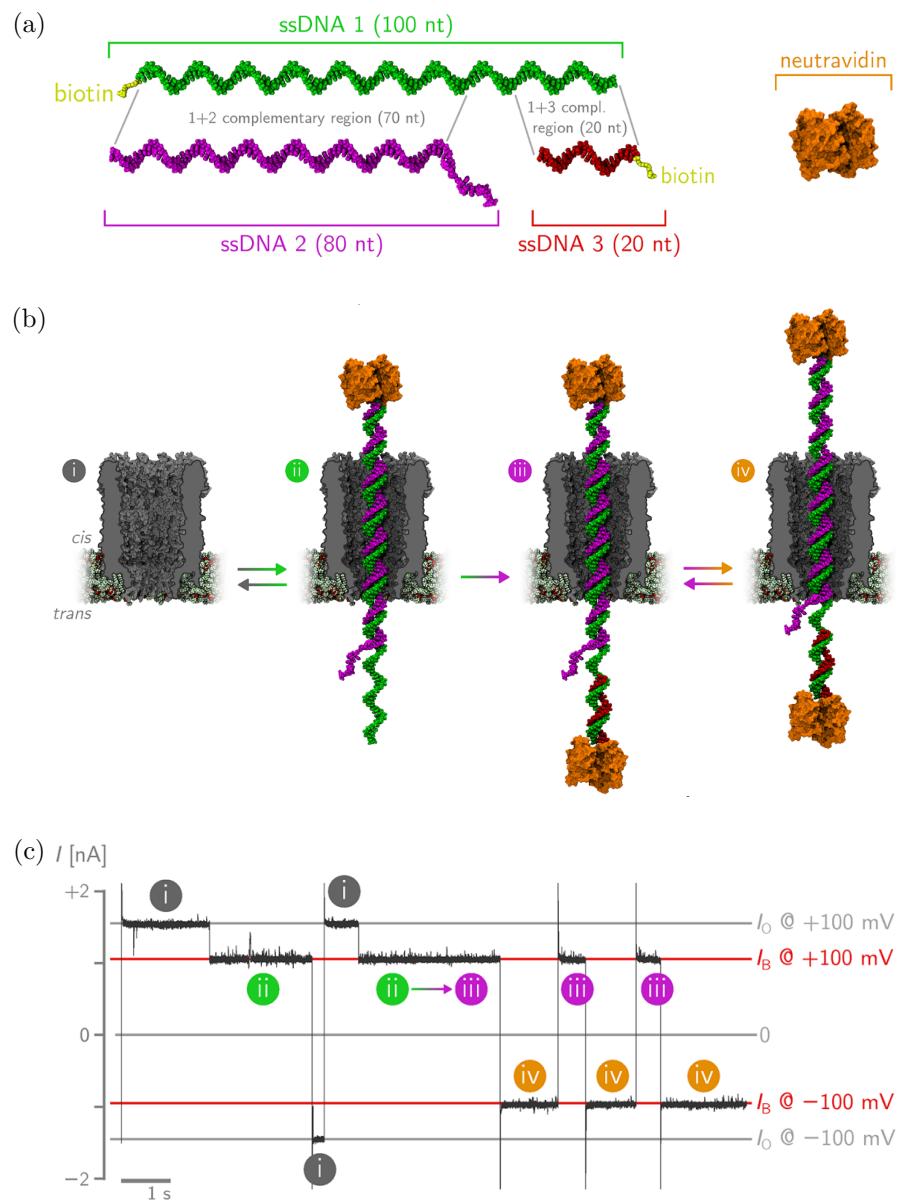


Figure 2.1: This is a figure [.]

### 2.3 Coarse-grained simulations

Understanding the Entropic interactions + providing a way to look into detail what happens. Chemist only measure current, we can provide a new window into the world of the nanopore.

LAMMPS.

Discuss bead spring model DNA + repulsive LJ interactions to simulate excluded volume

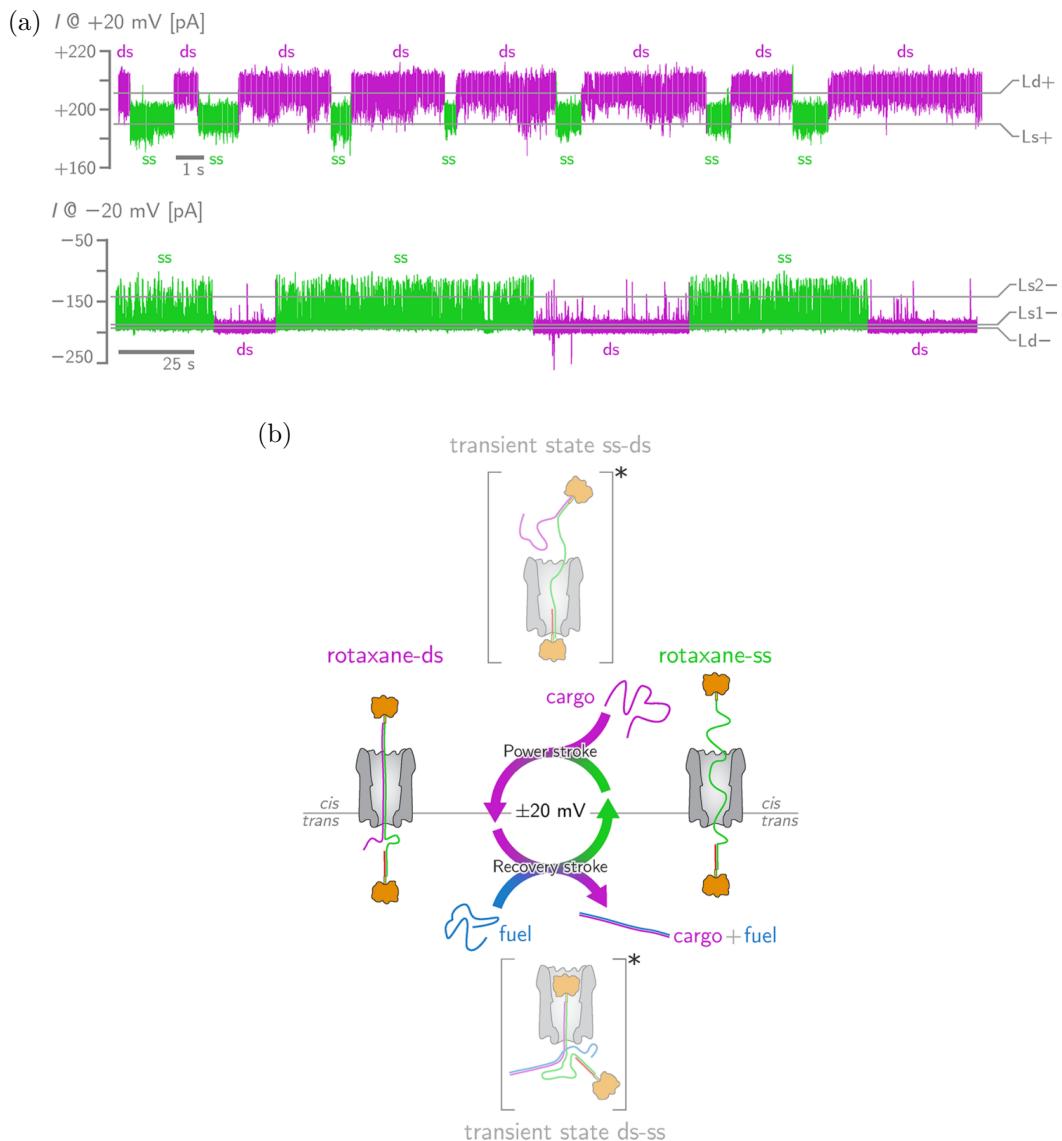


Figure 2.2: This is a figure

interactions.

The bond strength is chosen to be, by transforming the equation for persistence length  $k_{\text{SsDNA}} = \text{persistenceSsDNA} * \text{boltzmann} * \text{temperature} / \text{lengthNt}$   $k_{\text{DsDNA}} = \text{persistenceDsDNA} * \text{boltzmann} * \text{temperature} / \text{lengthBp}$   $k_{\text{bend}} = l_p * k_{bt} * T / \langle a \rangle$  Kuhn segments for ssDNA en dsDNA are different!

Gaussian probably distribution of the beads, i.e. the beads show ideal chain behaviour. Equipartition theorem.

## 2. THE DNA NANOPISTON

---

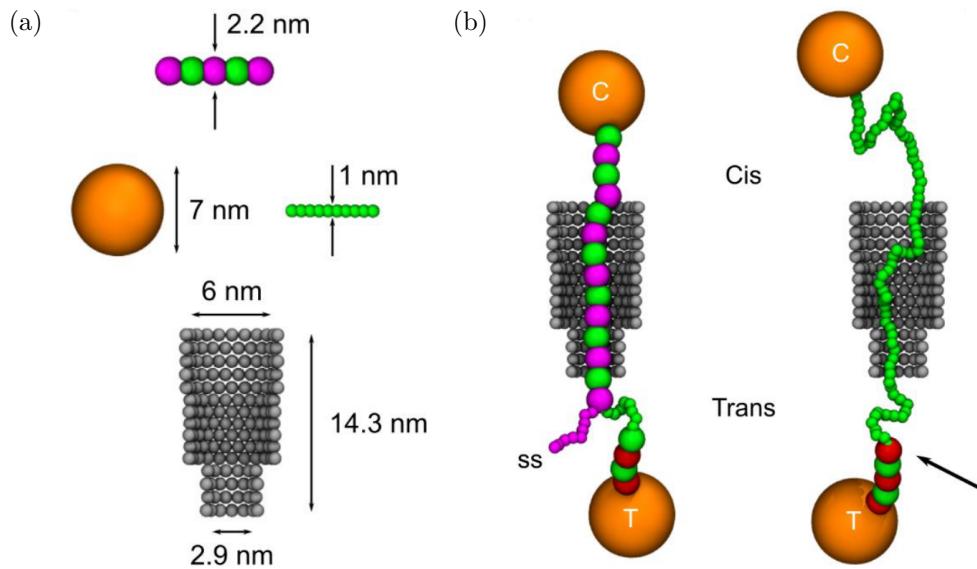


Figure 2.3: This is a figure

$$k_{bond} = 3k_b T / \langle a \rangle$$

Design nanopore and neutravidin pore. Based on electrostatic calculations + size stopper is fitted using simulations, requiring blockage to occur.

Langevin integrator is used. As is common in simulations of coarse-grained models, we use a higher diffusion coefficient than for physical DNA.

discuss limits of the model. Limited accuracy of the CG model since it does not capture the full structure of DNA accurately. For example the double helix structure is not captured. Another consequence is that the DNA hybridisation can not be simulated using this model, since both ssDNA nucleotides and dsDNA basepairs are represented by simple beads. To further analyse and understand the operation cycle of the nanopiston a more accurate CG model is needed.

# 3

## CHAPTER

# Adapting the Model

*All models are wrong, but some are useful.*

---

— George Box

## 3.1 OxDNA

OxDNA is a coarse-grained model of DNA developed by Thomas E. Ouldridge et al. at the University of Oxford. The central aim of the project was to develop a coarse-grained model of DNA, that could be used in the design of DNA technology. For the development of these technologies a model was needed that accurately captured the structural, mechanical and thermodynamical properties of DNA, while keeping the computational cost low.

The OxDNA model represents each nucleotide in the DNA strand as a rigid unit. Each rigid nucleotide has three independent interaction sites, each capturing a different aspect of the model. The interactions between these pseudo-atoms are compared to experimental data to calibrate the interactions, characterising their approach as "top down" coarse-graining. The interactions defined in the OxDNA model can be summarized as,

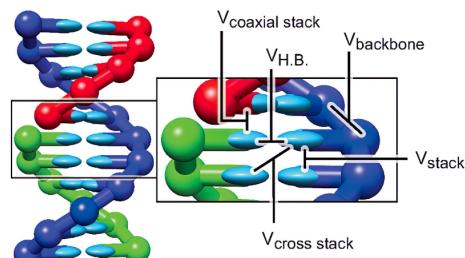


Figure 3.1: Structure of the OxDNA model with the different interactions. Figure was taken from [1]. (H.B. is hydrogen bond)

### 3. ADAPTING THE MODEL

---

$$V = \sum_{\text{nearest neighbours}} \left[ V_{\text{backbone}} + V_{\text{stack}} + V'_{\text{exc}} \right] \\ + \sum_{\text{other pairs}} \left[ V_{\text{H.B.}} + V_{\text{cross stacking}} + V_{\text{exc}} + V_{\text{coax stack}} \right]. \quad (3.1)$$

The first interaction site is the hydrogen-bonding/base excluded volume site, incorporating the hybridisation of complementary nucleotides into the model. The hydrogen-bonding interactions are not fixed, allowing for OxDNA to simulate dsDNA, ssDNA and their thermodynamic transitions.

The second is an excluded volume interaction site located at the backbone. This site's main role is to simulate the covalent bonding between consecutive phosphate groups. These permanent bonds provide structure to the ssDNA strands by forming the backbone.

The last interaction site is again located at the base, where it provides a base stacking interaction between consecutive nucleotides. The nucleotide stacking in DNA is crucial for the formation of the characteristic helix structure. Using these stacking interactions, this structure is implicitly imposed in the OxDNA model. This is in contrast with the traditional approach, used in coarse grained-models like Martini[.] en 3SPN[.], where the double helix structure is explicitly constructed. This implicit structure allows for the unstacking of nucleotides, which especially in ssDNA is an important contribution to the flexibility of the strand.

During the simulations of the DNA nanopiston, both the flexibility of the ssDNA strands and the DNA thermodynamics play an important role. Since both aspect of DNA are accurately captured by the OxDNA model, it provides a logical choice for our simulations. The low number of degrees of freedom in the model allows us to study computationally intensive simulations like DNA thermodynamics.

## 3.2 DNA Thermodynamics

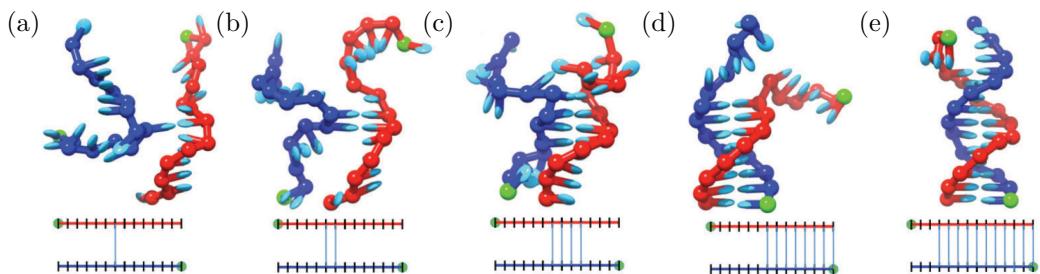


Figure 3.2: This is a figure [.]

The field of DNA thermodynamics focuses on understand how the structure of DNA varies with temperature. Due to nature of the hydrogen bond interactions that give rise to the structure of dsDNA, the association and dissociation of the DNA duplex is possible. The former is called DNA hybridisation, which is driven by a reduction in free energy due to the bonding of complementary base-pairs. While the latter is called DNA melting, a process seen at high temperatures, where the trade-off in configurational entropy during duplex formation is no longer favourable compared to the reduction in free energy.

During the discussion of the DNA nanopiston, we stated that thermodynamic transitions are the driving force behind its operating cycle. The power stroke in the cycle is induced by a toehold mediated strand displacement reaction, while a hybridisation reaction facilitates the recovery stroke.

Initiating strand hybridisation reaction incurs a thermodynamic penalty, this penalty originates in the decrease of configurational entropy when the strand start to form a duplex. This has a consequence that initial contacts in these reactions often dissociate. Even when an initial contact results in the stabilisation of the duplex for select base-pairs, the configuration often times is not conducive to full duplex formation.

The main reasons is that these transitions are not characterised by a single state but rather by a ensemble of possible transition pathways. The number of pathways increase dramatically when the strand sequences are repetitive, giving rise to hybridisation pathways called the Inchworm and the pseudo knot.

The combination of unstable initialisation of the hybridisation reaction together with the many transition pathways complicates the study of hybridisation with MD simulations

Deze nog ergens gebruiken:

fraying to describe the disruption of base pairs at the end of a duplex; if all base pairs fray, the duplex melts or dissociates.

zippering refers to when a new base pair forms at the end of an existing duplex, shown in figure 3.2

---

The other important thermodynamic transition is the Toehold mediated strand displacement. This reaction involved three DNA strand, of which two partially complementary strands from an imperfect duplex structure. The two strands can be partially complementary by having either a mismatch in basepair or a number of basepairs. The non hybridised part of the strand constitutes a flexible ssDNA strand that is refered to as the toehold.

The third strand in this reaction is fully complementary with one of the strands in the duplex, making it energetically favourable for the .

strand displacement increases number of hybridised base pairs, fully WatsonCrick complementary -> descreasing the overall free energy of the system. Overall, displacement is thermodynamically driven forward by the net gain in base pairs due to the toehold

### 3. ADAPTING THE MODEL

---

Once the toehold has bound, there are two possibilities: (i) the toehold base pair could dissociate, leading to the dissociation of the invader or (ii) the nearest base pair of the substrate-incumbent complex could fray, allowing the invader to compete to replace that base pair and complete the first step of branch migration.

as may seem reasonable, that the rate at which either base pair frays is similar, process (ii) should be approximately half as fast as process (i). strand displacement kinetics depends on toehold length. Initiating strand displacement incurs a thermodynamic penalty (making is difficult to simulate).

This is because, once the substrate- incumbent base pair frays, there is a 50the invader replacing the frayed base pair, and a 50step.

Therefore, the probability of successfully completing the remaining steps ofbranch migration before going back to the toehold-only- bound state is 1/20, from the gamblers ruin analysis

model branch migration at a more detailed level. we analyze a 1D (single-pathway) model of toehold-mediated strand displacement called the intuitive energy landscape (IEL) model -> refer to graph on image.

---

Conclusion: Reactions do not happen that easy, difficult to simulate, especially the toehold displacement reaction. An advanced sampling technique is needed to sample these rare events.

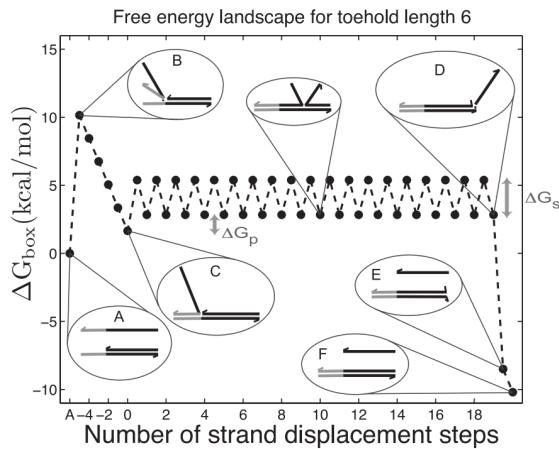


Figure 3.3: write caption[.]

### 3.3 Forward Flux sampling

Computational methods are used to study a wide variety of phenomena, ranging from large meteorological events to chemical reactions at the atomic scale. One class of phenomena

that is omnipresent in all these fields are the rare events. A rare event is an event caused by stochastic fluctuations in the system, characterised by a large gap in the time-scales of the duration of the events and their temporal spacing. The infrequency of their occurrence in combination with their short duration, makes them hard to study with both experimental and computational approaches.

Using this definition many natural processes can be classified as rare events, among which the hybridisation and toehold displacement reactions studied in this thesis. Due to the large temporal spacing of these rare events, brute-force molecular dynamics simulations would simulate a lot of wait time between events. To effectively probe the kinetics of these rare events, advanced sampling methods are needed.

A large ensemble of advanced sampling methods have been developed and can be largely divided into two classes. The first class are the free energy methods, based upon applying a biasing potential onto chosen collective variables. These potentials bias the Hamiltonian of the system in such a way that rare parts of its configuration space are explored. Notable examples of these methods are adaptive biasing force algorithm[.], basis function sampling[.] and umbrella sampling[.].

The second class of methods, known as path sampling methods, do not influence the systems Hamiltonian, but rather interface directly with the simulations trajectories. The transition path ensemble is usually sampled by perturbing an initial transition path or partitioning the phase space in subregions. Examples of these methods are transition path sampling [.] and forward flux sampling[.][.]. The latter will be used in our hybridisation simulations, motivated by its relative simplicity.

Forward Flux Sampling (FFS) starts with identifying two local minima,  $A$  and  $B$ , in the energy landscape of our system, for which we want to sample the transition path ensemble. Next an order parameter,  $\lambda(x)$ , is defined with the aim of partitioning the phase space,  $\Omega$ , using a set of nonintersecting hypersurfaces. By design, we choose this order parameter to be a function,  $\lambda(.) : \Omega \rightarrow \mathcal{R}$ , monotonically increasing from the initial state  $A$  and too the final state  $B$ .

Using this function the two local minima can now be specified as  $A := \{x : \lambda(x) < \lambda_A\}$  and  $B := \{x : \lambda(x) \geq \lambda_B\}$ . The chosen levels of order,  $\lambda_A$  and  $\lambda_B$ , construct the interfaces separating the two local energy basins from the rest of the phase space. Finally this procedure can be done for a  $N$ -number of interfaces partitioning the space between  $A$  and  $B$ , we find

$$\lambda_A = \lambda_0 < \lambda_1 < \dots < \lambda_{N-1} < \lambda_N = \lambda_B \quad (3.2)$$

Note that this methods does not require an in depth knowledge of the systems energy landscape, however the choice of order parameter will heavily influence the efficiency of the simulation. Analogues to the ambiguous choice of a collective variable in free energy methods, making these decisions is often more an art then a science.

The ultimate aim of these methods is to get a grasp of the kinetics of rare events. In quantitative terms this means determining the rate constant of the transition from  $A$  to  $B$ ,

### 3. ADAPTING THE MODEL

---

denoted as  $k_{AB}$ . The expression used to calculate  $k_{AB}$  is:

$$k_{AB} = \frac{\langle \Phi_{A,n} \rangle}{\langle h_A \rangle} = \frac{\langle \Phi_{A,0} \rangle}{\langle h_A \rangle} P(\lambda_n | \lambda_0) \quad (3.3)$$

where  $\langle \Phi_{A,n} \rangle$  is the steady-state flux of trajectories starting in  $A$  and reaching the final interface  $\lambda_n$  (i.e. reaching  $B$ ) and  $\langle h_A \rangle$  is the average fraction of time that a trajectory spends in the basin of local minima  $A$ . In the above equation this steady state flux is factorised into the flux of trajectories starting in  $A$  and crossing  $\lambda_0$  and the subsequent probability of reaching the final state from this interface. Using the previously defined interfaces, we can now factorize the events' probability into transition probabilities between the individual interfaces as

$$P(\lambda_n | \lambda_0) = \prod_{i=0}^{n-1} P(\lambda_{i+1} | \lambda_i). \quad (3.4)$$

Estimating these transition probabilities can be done by shooting trajectories starting from one interface to the next, while keeping track of the fraction of attempts successfully crossing the next interface. Since not the entire energy landscape between the minima has to be crossed, measuring these small transitions can be more easily done.

Note that this set-up allows for simulations of both equilibrium and out-of-equilibrium systems, it does not require detailed balance like other sampling techniques. Non equilibrium systems are ubiquitous in soft matter physics, illustrating another strength of the method. Different variants on the FFS method have been devised, they differ in the approach by which

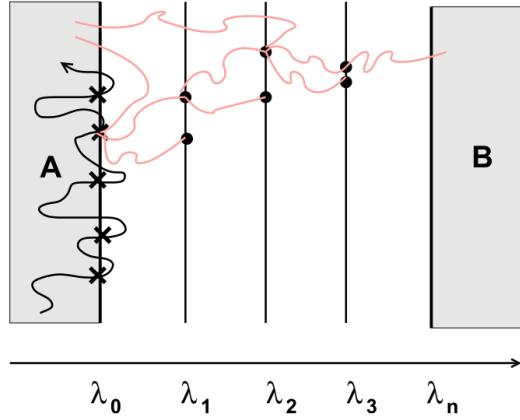


Figure 3.4: write caption[.]

they calculate the probability  $P(\lambda_n | \lambda_0)$ . During this thesis I choose to use the Rosenbluth-like (RB) method [zie citation allen review]. The choice is motivated by its resemblance with well known Monte Carlo Simulations and recursive nature, making it easy to implement. This method generates unbranched transition paths from state  $A$  to state  $B$ , making them easy to analyse. The algorithm is described in six steps:

- (i) Generate configurations on the  $\lambda_0$  interface by running simulations in the  $A$  basin. Keeping track of the fractions  $\langle \Phi_{A,0} \rangle / \langle h_A \rangle$  is evaluated.
- (ii) Fire  $k_0$  trial runs from generated configurations on  $\lambda_0$  until they cross  $\lambda_1$  or cross back to  $\lambda_0$ . Store the final configurations of the successful simulations.
- (iii) Sample one of the saved configuration on the  $\lambda_1$  interface and use it to shoot  $k_1$  runs to the next interface  $\lambda_2$ .
- (iv) Iterate the previous steps until the trajectories reach  $\lambda_n$  or no more configurations are available.
- (v) If not successful, sample a stored configuration on  $\lambda_0$  and repeat the steps (i) to (iv).
- (vi) Finally compute  $P(\lambda_n | \lambda_0)$  using a weighted average of individual transition probabilities as described below.

Calculating the transition probabilities is done by taking a weighted average of the attempted trial runs. The trial runs of path  $b$  starting at interface  $\lambda_i$  is assigned a weight  $w_{i,b}$  as

$$w_{i,b} = \prod_{j=0}^{i-1} S_{j,b} / k_j. \quad (3.5)$$

where  $S_{j,b}$  is the number of successful trajectories during the generation of path  $b$ . Using these weights the transition probability is computed using

$$P(\lambda_{i+1} | \lambda_i) = \frac{\sum_b w_{i,b} S_{i,b} / k_i}{\sum_b w_{i,b}}. \quad (3.6)$$

## 3.4 Simulation technique



# CHAPTER 4

## Simulations of the Rotaxane

*Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the beings which compose it, if moreover this intelligence were vast enough to submit these data to analysis, it would embrace in the same formula both the movements of the largest bodies in the universe and those of the lightest atom; to it nothing would be uncertain, and the future as the past would be present to its eyes.*

---

— Pierre-Simon Laplace

### 4.1 Mixed Rotaxane

### 4.2 Conformational Fluctuations of the Rotaxane

### 4.3 Toehold Displacement Reaction

#### 4. SIMULATIONS OF THE ROTAXANE

---

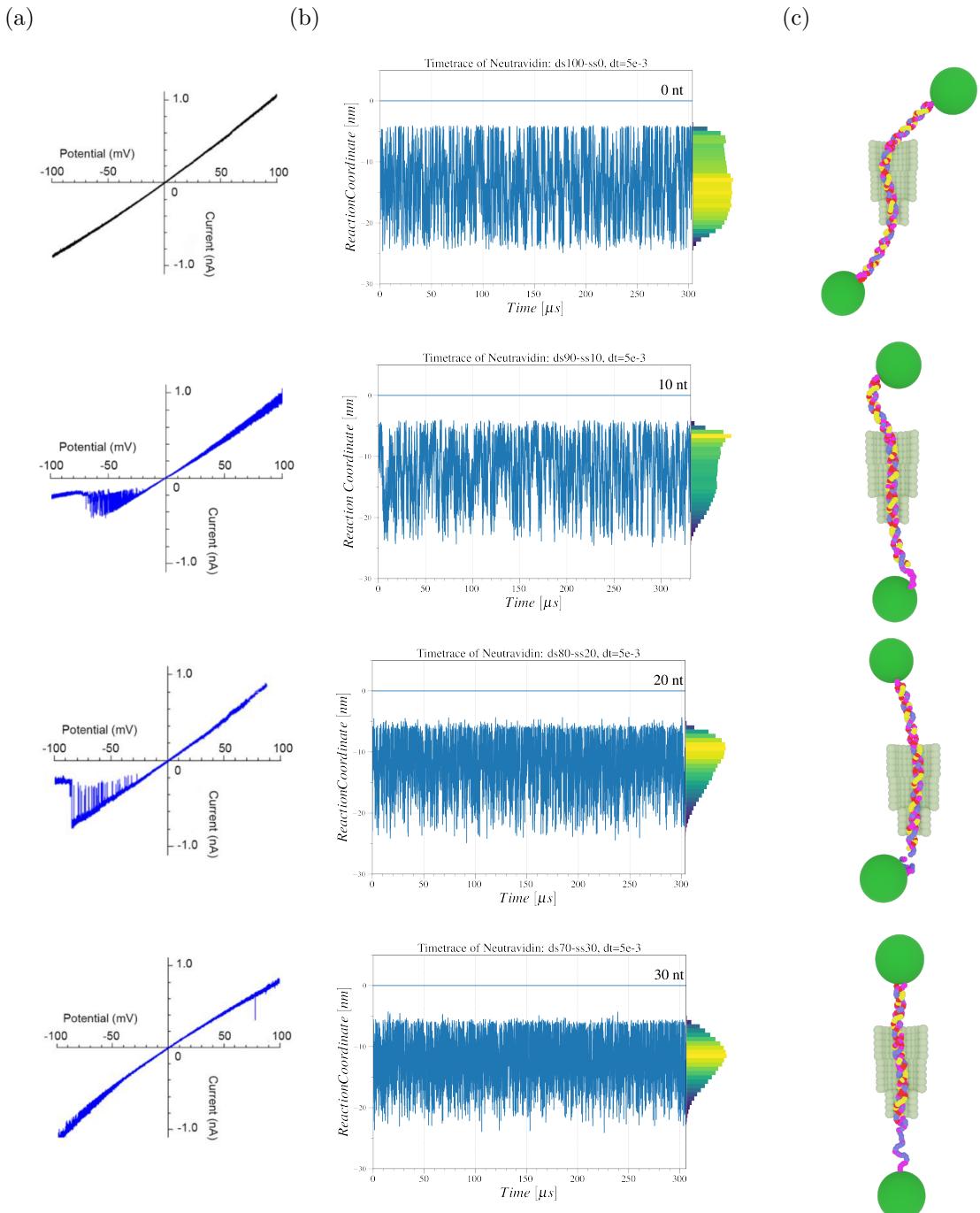


Figure 4.1: asdf asdf asdf asdf asd fasdf fasdf asf asdfasdfa asdfasdf asfasf. asdf asdf asdf asd fasdf fasdf asf asdfasdfa asdfasdf asfasf. asdf asdf asdf asdf asd fasdf fasdf asf asdfasdfa asdfasdf asfasf. asdf asdf asdf asdf asd fasdf fasdf asf asdfasdfa asdfasdf asfasf.

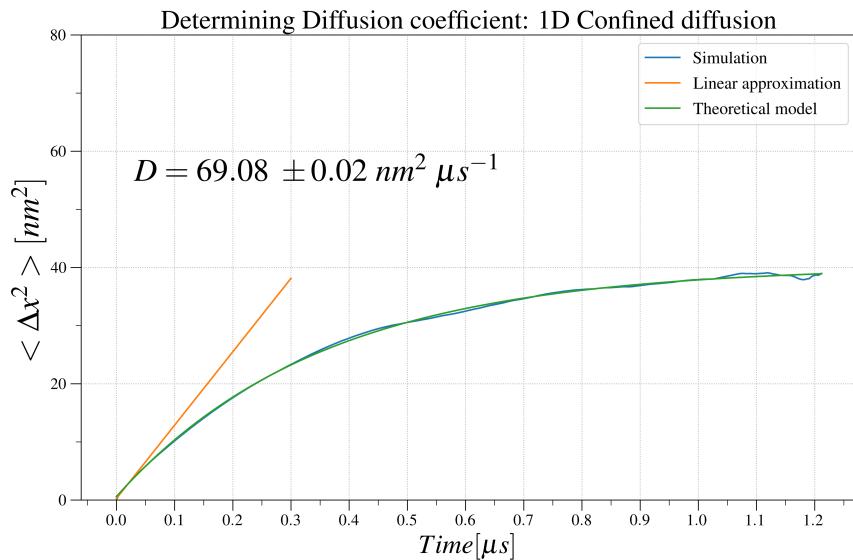


Figure 4.2: write caption

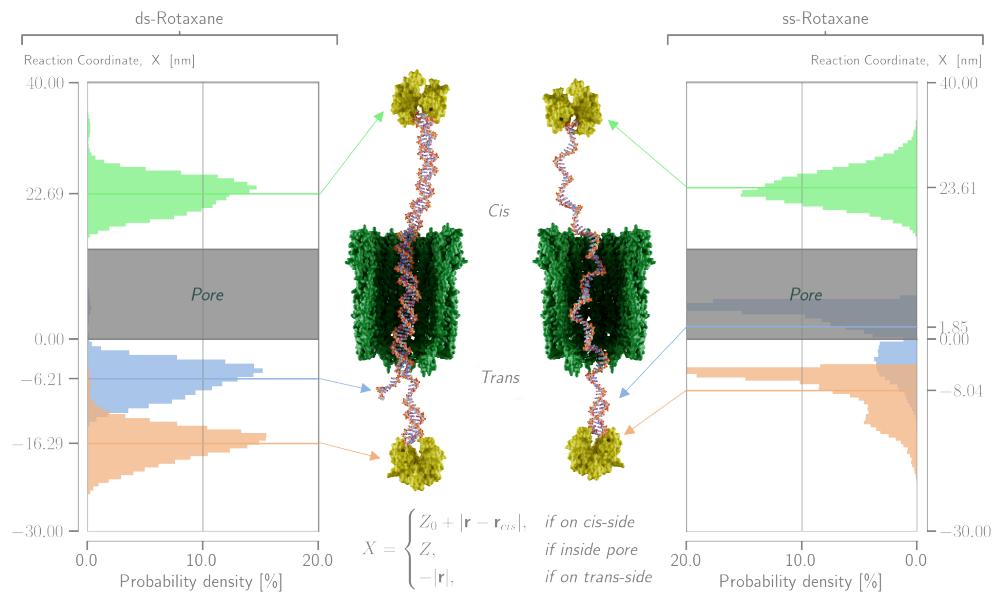


Figure 4.3: write caption



# CHAPTER 5

## Conclusions and Perspectives

*All models are wrong, but some are useful.*

---

— George Box

### 5.1 Results & Conclusions

### 5.2 Future Perspectives



# APPENDIX A

## 1D Confined Diffusion

Studying the dynamics of the mixed rotaxane highlighted the importance of entropic interactions between the nano pore and the DNA strand. Here we observed that a fully double stranded DNA polymer represented a special case. The uniformity of the  $\mathcal{X}$  histogram corresponding to this 0 nt mixed rotaxane suggests a free diffusive motion of the rotaxane in a bounded one-dimensional domain. This isotropic behaviour was previously also observed in the bead-spring simulations by Bayoumi et al.<sup>?</sup>

$$\langle \Delta x^2 \rangle \simeq 2nDt.$$

$$\frac{\partial \psi}{\partial t} = D \frac{\partial^2 \psi}{\partial x^2}, \psi(x, t) = f(x)g(t)$$

Reflecting boundary conditions  $j = -D \frac{\partial \psi}{\partial x} = 0$ . Current vanishes at the boundaries

$$t : \quad \dot{g} = -\alpha g(t) \Rightarrow g(t) = e^{-\alpha t}$$

$$\begin{aligned} x : \quad D \ddot{f} &= -\alpha f(x) \Rightarrow f(x) = A \sin(Kx) + B \cos(Kx) \\ &= B \cos\left(\frac{\pi n x}{L}\right) \end{aligned}$$

$$\frac{\alpha}{D} = \frac{\pi^2 n^2}{L^2}$$

The general solution is given by the linear combination,

$$\begin{aligned}\psi(x, t) &= \sum_{n=0}^{+\infty} C_n \cos\left(\frac{\pi n x}{L}\right) e^{-\frac{D\pi^2 n^2}{L^2} t} \\ &= \frac{1}{L} \left[ 1 + \sum_{n=1}^{+\infty} \cos\left(\frac{\pi n x_0}{L}\right) \cos\left(\frac{\pi n x}{L}\right) e^{-\frac{D\pi^2 n^2}{L^2} t} \right] \\ \langle \Delta x^2 \rangle &= \langle (x - x_0)^2 \rangle \\ &= \frac{L^2}{6} \left[ 1 - \frac{96}{\pi^4} \sum_{k=0}^{+\infty} \frac{1}{(2k+1)^4} e^{-\frac{D(2k+1)^2 \pi^2}{L^2} t} \right]\end{aligned}$$

As expected, the mean squared distances saturates to  $\langle \Delta x^2 \rangle = L^2/6$  in the long-time limit  $t \gg L^2/D$ . To explore the other limiting case  $t \ll L^2/D$  we perform a Taylor expansion and find

$$\langle \Delta x^2 \rangle = \frac{L^2}{6} - \frac{16L^2}{\pi^4} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^4} + \frac{16Dt}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} + \mathcal{O}\left(\frac{D^2 t^2}{L^4}\right).$$

[cite in bickel]

$$\sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{\pi^2}{8} \quad \text{and} \quad \sum_{k=0}^{\infty} \frac{1}{(2k+1)^4} = \frac{\pi^4}{96}$$

$$\langle \Delta x^2 \rangle = 2Dt \quad t \ll L^2/D$$

# Acknowledgements

...





**DEPARTEMENT NATUURKUNDE EN STERRENKUNDE**

Celestijnlaan 200d bus 2412

3000 LEUVEN, BELGIË

tel. + 32 16 32 71 24

fys.kuleuven.be

