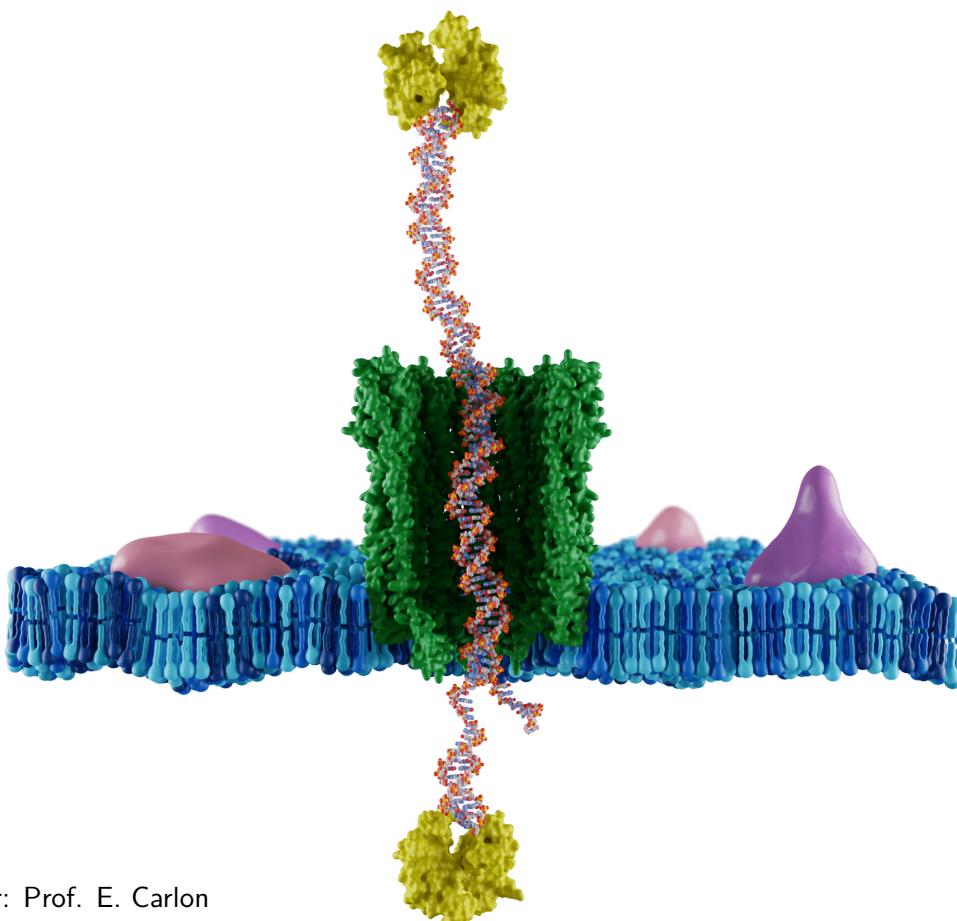


Coarse-grained simulations of the DNA nanopiston



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

Autonomous molecular machines are ubiquitous in the machinery of life, collectively driving the molecular processes in our cells. Inspired by these biological machines, scientists develop synthetic devices performing specialised operations at the nanoscale. In this thesis we study a specific molecular machine designed by Bayoumi et al.[1], which is composed of a DNA-neutravidin piston trapped inside a ClyA nanopore.

Using the free energy of DNA hybridisation this molecular machine is able to perform autonomous and active transport of DNA cargo both following or opposing an external bias force. During each operating cycle of the nanopiston a DNA cargo is transported from the *cis*- to the *trans*-side of the membrane, in which the piston is embedded.

Due to the length scale associated with molecular machines performing in depth experimental studies has been proven to be challenging. During this thesis we aim to shed light on the operating principles of the nanopiston by using molecular dynamics simulations. Motivated by the computational cost of classical all-atom simulations a coarse-grained model of the DNA nanopiston is designed.

Entropic interactions between the DNA piston and the nanopore are thought to play an important role in facilitating the DNA transport. Studying these effects reveal two distinct origins of entropic forces. Large double stranded DNA is kept predominantly outside of the pore's constriction, by the entropic penalty of confinement. Whereas, the flexible single stranded DNA also endeavours to maximize its available configurational space by opposing confinement.

Consecutively the conformational fluctuations of the nanopiston are studied. These simulations clearly show the importance of the entropic interactions promoting the operation cycle. The entropic penalty of confining the flexible single stranded DNA components of the piston in the nanopore enable the continuation of the hybridisation reactions.

In an attempt to study a full piston cycle, the hybridisation reactions driving the operation are simulated using our coarse-grained model. Due to the inherent difficulty of simulating these reactions an advanced sampling method called forward flux sampling is needed. While performing these simulations the main limitation of our coarse-grained model was encountered. The compliance of the biological nanopore is found to be essential in facilitating the hybridisation pathways, but is not yet incorporated in our current model.

Vulgariserende Samenvatting

Als we door een biologische bril naar de natuur kijken zijn cellen de kleinste bouwstenen van een organisme. Deze basiseenheid van de natuur is opgebouwd uit vele moleculaire machines, die gezamenlijk het leven mogelijk maken. Denk hierbij aan de flagellen waarmee bacteriën zich voortbewegen of de enzymen die ons erfelijk materiaal kopiëren bij de celdeling. Wat deze biologische machines zo speciaal maakt, is hun kleine omvang. Vaak zijn ze namelijk niet veel groter dan enkele nanometers. Geïnspireerd door de vele taken die deze machines kunnen volbrengen, onderzoeken wetenschappers hoe zij deze zelf kunnen namaken en optimaliseren.

In dit onderzoek bestuderen wij een specifiek voorbeeld van een synthetische moleculaire machine, die instaat is om DNA moleculen te transporteren. Deze machine kan vergeleken worden met een zuiger uit een verbrandingsmotor, waarbij tijdens elke cyclus een molecule wordt getransporteerd. Door de microscopische omvang van dit apparaat is het moeilijk gebleken om de exacte mechanismen te bestuderen. Om toch een inzicht te verkrijgen in deze mechanismen kunnen computersimulaties worden gebruikt.

Door een computermodel van deze moleculaire machine te ontwerpen, kunnen we de microscopische bewegingen van de individuele componenten volgen. Vanwege de complexiteit van dit biologische apparaat is het niet mogelijk om elk atoom explicet te modelleren. Dit probleem wordt opgelost door gebruik te maken van een coarse-grained model, waarbij sommige moleculaire details worden genegeerd. In deze thesis dient onze computer als een computationele microscoop om de kleinste processen te bestuderen.

Het uiteindelijke doel van deze studie is om beter te begrijpen hoe deze moleculaire machine functioneert. Op deze manier kunnen onze simulaties bijdragen tot het ontwikkelen van nieuwere en betere versies. De vooruitgang in dit onderzoeksgebied gaat zeer snel, waardoor wetenschappers voorzichtig durven uit te kijken naar het ontwerpen en fabriceren van een volledig synthetische cel.

Summary in Layman's Terms

If we look at nature from a biological perspective, cells are the smallest building blocks of an organism. This basic unit of nature is made up of many molecular machines that together make life possible. An example are the engines with which bacteria move around or the machines that copy our genetic material during cell division. What makes these devices so special is their small size, often they are not much larger than a few nanometres. Inspired by the many tasks these small machines can accomplish, scientists are busy investigating how they can design them in their lab.

In this work, we study a specific example of a synthetic molecular machine, capable of transporting DNA strands. This machine can be compared to a piston in an internal combustion engine, where a DNA strand is transported during each cycle. Due to the microscopic size of this device, it is difficult to determine the exact working mechanisms. To gain an insight in these mechanisms, computer simulations are used.

By designing a computer model of this molecular machine, we can analyse the microscopic movements of its individual components. Because of the complexity of this biological device it is not possible to model every atom explicitly. This problem is solved by using a 'coarse-grained' model, where some molecular details are ignored. In this thesis, our computer serves the role of a computational microscope by which we study these tiny processes.

The ultimate goal of this study is to better understand how this molecular machine functions and guide the development of improved versions. The progress in this research field is moving fast, compelling scientists to cautiously start thinking of designing and fabricating a fully synthetic cell.

List of Figures

1.1	Render of a flagellar motor-hook complex, based on the cryo-EM structure of the <i>Salmonella</i> flagellar motor.	2
1.2	Detailed illustration of ionic current spectroscopy.	4
1.3	Structural overview of the α -Hemolysin (α -HL) pore forming protein.	5
1.4	Structural overview of the Cytolysin A (ClyA) pore forming protein.	6
1.5	Figure showing the double helical structure of DNA.	8
1.6	Illustration of the Kratky-porod model.	9
1.7	Example of an expanded model of a simple liquid.	11
1.8	Comparing different degrees of coarse graining of DNA	14
2.1	short lof/lot caption	17
2.2	short lof/lot caption	19
2.3	This is a figure	21
3.1	Structure of the OxDNA model	24
3.2	This is a figure [35]	25
3.3	write caption [36]	26
3.4	write caption[.]	29
4.1	short lof/lot caption	33
4.2	short lof/lot caption	35
4.3	write caption	37

Contents

Abstract	i
Vulgariserende Samenvatting	iii
Summary in Laymans's Terms	v
List of Figures	ix
Contents	xi
1 Introduction	1
1.1 Thesis outline	1
1.2 Biological Nanopores	3
1.3 Deoxyribonucleic acid (DNA)	7
1.4 Polymer Physics	8
1.5 Computer Simulations	11
2 The DNA Nanopiston	15
2.1 Rotaxane Formation	16
2.2 Operating principles	18
2.3 Coarse-grained simulations	20
3 Improving the Model	23
3.1 OxDNA	23
3.2 DNA Thermodynamics	24
3.3 Forward Flux sampling	27
3.4 Computational setup	29
4 Simulations of the Rotaxane	31
4.1 Mixed Rotaxane	32
4.2 Conformational Fluctuations of the Rotaxane	36
4.3 Hybridisation Reactions	37
5 Conclusions and Perspectives	39
5.1 Results & Conclusions	39
5.2 Future Perspectives	40
A One Dimensional Confined Diffusion	43
Bibliography	47
Acknowledgements	51

CHAPTER

1

Introduction

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

— Sidney Coleman

1.1 Thesis outline

All organisms in nature tirelessly perform work to keep their cellular functions intact, opposing the ever increasing entropy in their environment. 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 their corresponding length-scale. Often times these machines 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 of these thermal fluctuations in their environment molecular machines naturally perform a stochastic motion that complicates their functioning. Extracting useful work from these freely tumbling structures is almost impossible. To overcome this limitation most synthetic molecular machines are embedded in a larger complex providing necessary stability.[2]

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 flagellar motor, shown in Figure 1.1. The function of this machine is providing an efficient way for bacteria to roll and tumble throughout their environment. Just like in electrical motors, the flagellar motor consists of a stator and a rotor. The stator is anchored into the

1. INTRODUCTION

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 flagellar motor generates a unidirectional motion.[3]

Similarly to macroscopic engines, heat is produced during the operation of molecular machines. When the structure is not capable of dissipating this heat efficiently, an excessive build-up compromises its durability. To mitigate this problem, large and soft molecules are often 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. This makes DNA a popular material in nanotechnology.

The central topic of this thesis is studying a synthetic molecular machine consisting of a DNA nanopiston embedded into a phospholipid membrane, as depicted on the cover page. The structure was designed by Bayoumi et al. with the aim of performing selective transport of DNA through a membrane.[1] This nanopiston can be characterised as an autonomous molecular machine, which turns over chemical fuel to continuously perform work.

In the first chapter a comprehensive introduction of important concepts regarding the DNA nanopiston is given. Having laid this theoretical foundation, the structure and operation cycle of the DNA nanopiston is discussed in chapter two. Next the computational model used in this thesis is presented in chapter three. In chapter four, the results of these simulations is discussed. Finally chapter five will offer a discussion of these results and recommendations for further research.

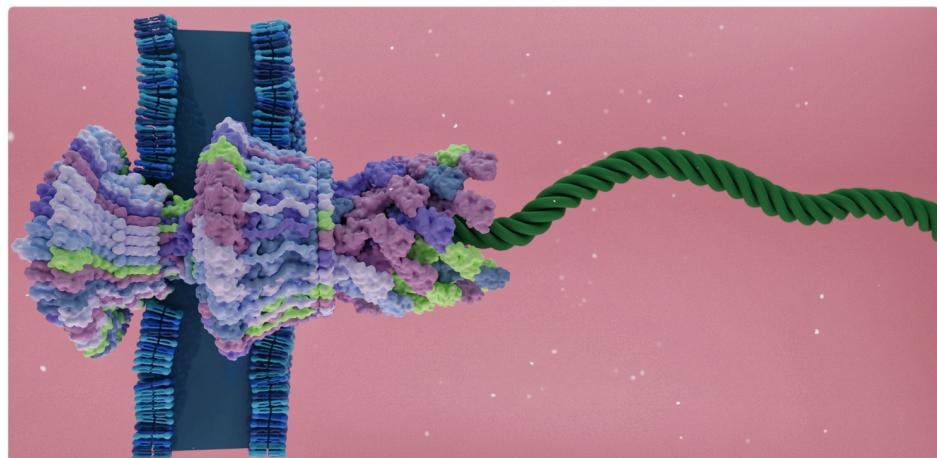


Figure 1.1: Render of a flagellar motor-hook complex embedded in a cell membrane. The merged density map of the flagellar motor was obtained from the cryo-EM structure of the Salmonella flagellar motor and prepared using ChimeraX. Both the cell membrane and the flagella are interpretations. Image was rendered using Blender.[4]–[6]

1.2 Biological Nanopores

Biological nanopores are small perforations in a lipid bilayer membrane, created by the central cylindrical cavity of a folded protein assembly. The majority of these proteins are toxins produced by pathogenic bacteria. Their function in nature is to perforate the membrane of a cell, causing the cell to depolarize, disrupting vital cell functions. The perforation also induces an osmotic potential, causing cell nutrients to spill into the environment. Both effects eventually result in the killing of the cell.[7]

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. Retrieving information from nano-scale processes has proven to be a challenging task. Developing sensors to probe this small length scale is thereby very relevant. This is the exact task nanopores provide a possible solution to, i.e. spectroscopy at the smallest scale.

1.2.1 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 is depicted in figure 1.2. A reservoirs, filled with a saline solution, is partitioned into two compartments by a lipid bilayer. The membrane is perforated using a pore forming protein, for example α -HL. 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 of the nanopore partially block the flow through the pore, reducing the measured ion current. 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.[8]

The construction of the molecular machine discussed in this thesis is based on this spectroscopy technique. Instead of studying analytes in the reservoirs, here the configuration of DNA thread trapped inside of the pore can be analysed using the measured ion current.

It should be noted that besides these biological nanopores there are also inorganic nanopores in development.[9] An example 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 costs, 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, solid state nanopores will prove to be an important asset in the future of nanotechnology.

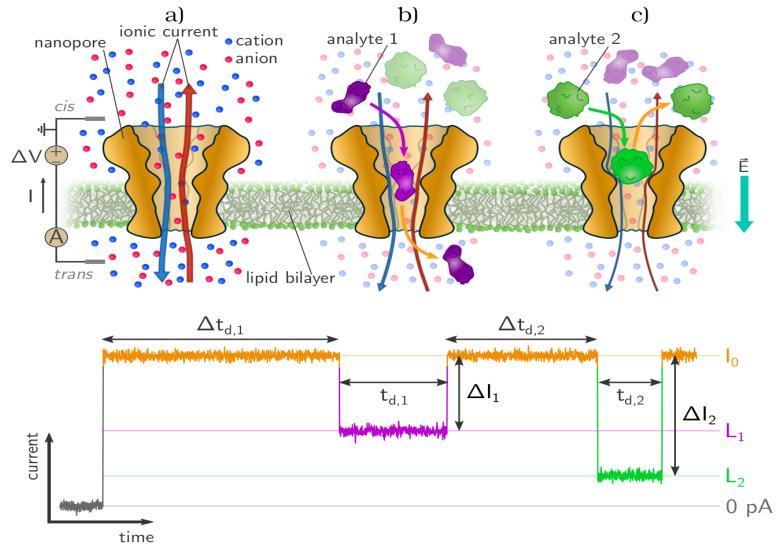


Figure 1.2: Illustration of the three primary scenarios encountered during ionic current spectroscopy are depicted. Applying a potential difference over the membrane induces a flow of cations (blue) and anions (red) through the pore. If the pore is unobstructed, the mediated ion current is referred to the open-pore current (I_0), shown in a). After a certain time $\Delta t_{d,1}$, analyte 1 approaches the pore entrance and partially blocks the pore current to the current level L_1 , as depicted in b). Due to the size of this analyte it is able to translocate through the pore with a characteristic dwell time $t_{d,1}$. Both the duration of translocation and the reduction in ion current dependents on the specific analyte properties . In c) a larger analyte is shown to block the pore current to a lower level L_2 . The larger size of this analyte forces it to exit back into the cis-compartment with a characteristic dwell time $t_{d,2}$. Figure taken from reference [10].

1.2.2 α -Hemolysin (α -HL)

The α -Hemolysin (α -HL) protein is the most commonly used pore forming protein to create artificial nanopores. It is produced by the *Staphylococcus aureus*, a bacterium commonly found in the human skin microbiome.[11]

The α -HL pore (PDBID¹:7AHL[12]) is an oligomeric complex with multiple naturally occurring variations. The most typical configuration is a heptameric structure, meaning that there are seven subunits, called protomers, making up the pore complex. The secondary structure elements consist principally of β -sheets ², making them a member of the β -barrel pore-forming toxins. Through both electrostatic and hydrophobic interactions, the α -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 β -barrel into the membrane.[13]

¹An identifier for entries in the Protein Data Bank (PDB), a database for structural data of large biological molecules

²A secondary protein structure created by hydrogen bonding between parallel or antiparallel polypeptide strands forming extended -pleated sheets.

Structurally the shape of α -HL resembles that of a hollow mushroom, see Figure 1.3. The total height of the complex is 11 nm and the maximum width is measured to be 10 nm. The internal chamber of the pore located at the cis-side of the membrane is called the lumen. The lumen of α -HL is quite constricted having a diameter of merely 3 nm. At the membrane the lumen chamber transitions into a protein stem, referred to as the constriction of the pore. Here the diameter of the chamber is reduced to a minimum of 1.5 nm. On the wall of α -HL's inside chambers, the charges are relatively uniformly distributed.

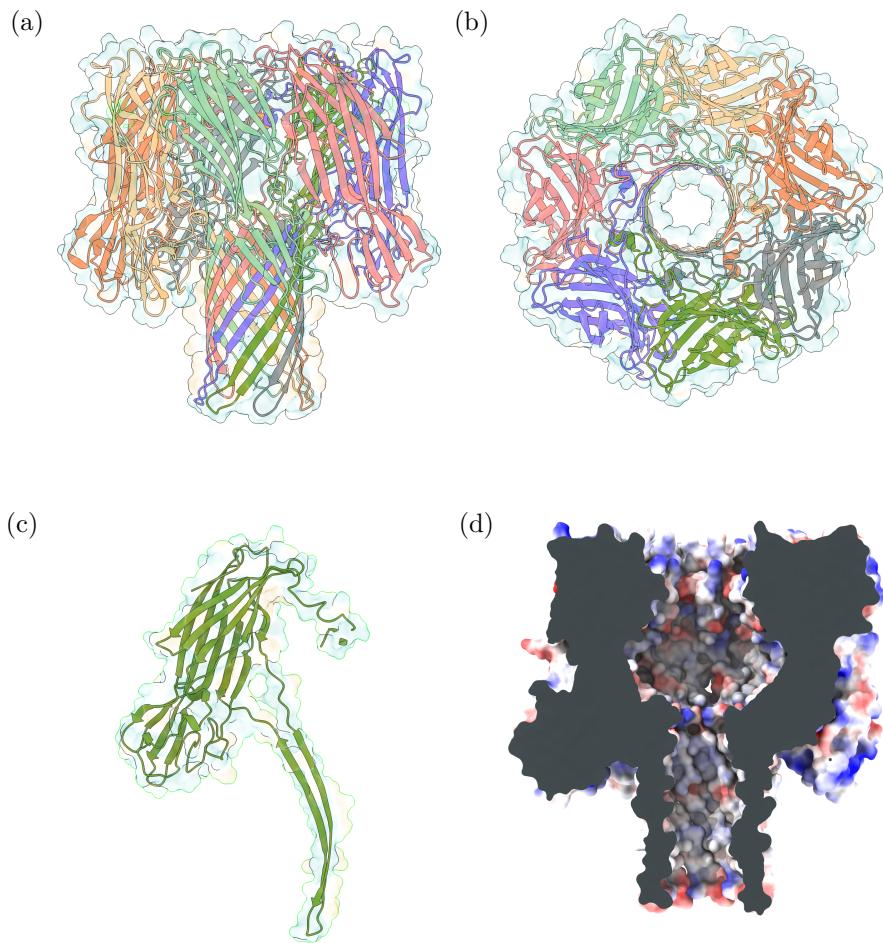


Figure 1.3: (a) Side view and (b) top view of the heptameric α -hemolysin (α -HL) protein structure. The mushroom shape arising from the protomer assembly is clearly shown. (c) A protomer subunit from the pore complex, where the two antiparallel β strands composing the β barrel are visible. (d) Estimation of the charge distribution on the cross-section of the merged density surface of α -HL. All images were rendered using ChimeraX. [6], [12]

1. INTRODUCTION

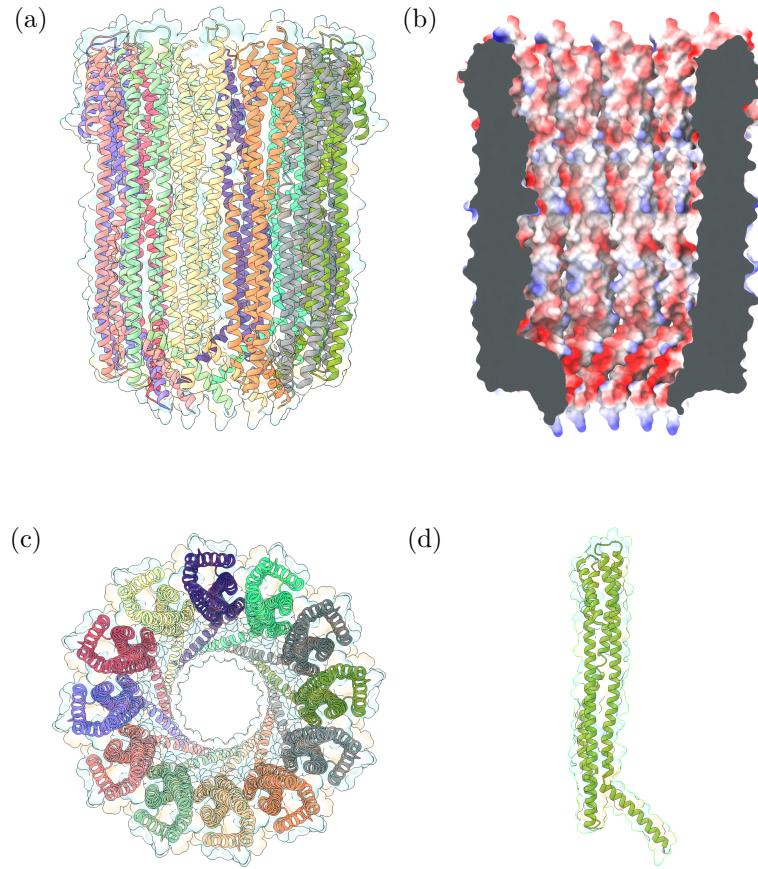


Figure 1.4: (a) Side view and (b) top view of the dodecameric Cytolysin A (ClyA) protein structure. The cylindrical shape arising from the protomer assembly is clearly shown. (c) A protomer subunit from the pore complex, where the α -tongue is visible. (d) Estimation of the charge distribution on the cross-section of the merged density surface of ClyA. All images were rendered using ChimeraX. [6], [14]

1.2.3 Cytolysin A (ClyA)

The Cytolysin A (ClyA) is a larger type of pore forming protein first found to be secreted by *E. coli* strains. [15] The larger size of its lumen allows for different types of applications compared to smaller complexes like α -HL. The larger diameter of the pore's stem allows for translocation of double stranded DNA, which is most relevant for this thesis.

The ClyA pore (PDBID:6MRT[14]) is an oligomeric complex most typically found in a dodecameric configuration, meaning that there are twelve subunits, called protomers, making up the pore complex. In nature small variations on this configuration are found. The secondary structure elements consist principally of α -helices³, making it a member of the

³A secondary protein structure created by stabilising a coiled peptide chain with hydrogen bonds, forming right handed-helix configuration.

α -pore-forming toxins. The protein formation is induced by the hydrophobic interactions between the monomers β -hairpin and the solvent. The main structural rearrangement in this process consists of swinging out this hairpin and inserting the resulting α -tongue into the membrane. After this transition, the membrane-bounded monomers oligomerize to form the final pore structure.[16]

Structurally the shape of ClyA resembles that of two hollow cylinders stacked on top of each other, see Figure 1.4. 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 α -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 α -HL, the inside surface of ClyA has a net negative charge, promoting the capturing of positive ions. On the contrary, this excess charge will induce significant coulomb repulsion between the pore and negatively charged analytes.

1.3 Deoxyribonucleic acid (DNA)

Deoxyribonucleic acid (DNA) is a long biopolymer composed of two strands, commonly found in its characteristic double helix structure, shown in Figure 1.5. 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, when Friedrich Meicher discovered a new microscopic substance found on discarded surgical bandages.[17] 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", plays an important role as a blueprint for the perpetuation of living matter.[18]

The structure of DNA was first determined by Rosalind Franklin using X-ray crystallography. Later this research was published by Watson and Crick, who concluded that DNA consists of two individual strands forming a double helical structure.[19] 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 rise to a selection rule, allowing to only form 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 Figure 1.5. A helical twist of B-DNA consists of around 10 base pairs, having a net helical pitch of 0.34 nm . During this thesis, when analysing DNA we refer to the B-DNA form.

Thermal fluctuations in the environment dominate the DNA strands dynamics, giving rise to an inherently stochastic system. Consequently, a statistical theory of polymer physics is an essential tool in describing these dynamics. 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.



Figure 1.5: Figure of the double helical structure of DNA, here in the B-DNA form as it is the most commonly observed in nature. The phosphate backbones are represented as ribbons, enclosing the bases forming coplanar hydrogen bonds. Image was made in Blender, using generated PDB data.

1.4 Polymer Physics

A polymer is a biomolecule made up of building blocks called monomers, linked together to form a chain. In our simple mathematical model 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 simplest 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.

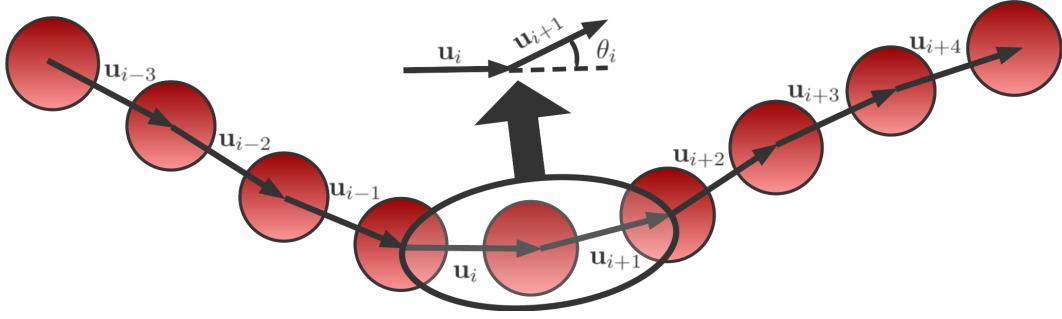


Figure 1.6: Figure illustrating the Kratky-Porod model. The angle between consecutive bonds, as highlighted in the figure, is used to impose bending rigidity in the polymer.

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 [20], or discrete wormlike chain, solves this problem by taking the energetic cost of bending the polymer into account. Mathematically this is done by 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 θ_i is the angle between neighbouring bond-vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_{i+1}$, see Figure 1.6. The lowest energy state of this discrete wormlike chain is a straight rodlike configuration, where the bond angles θ_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

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

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,

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

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,

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

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,

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

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

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

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

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

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

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

introducing a new polymer quantity, the bending persistence length

$$l_b \equiv \frac{a\kappa}{k_b T}.\tag{1.10}$$

This general result in polymer physics states that the correlations between bond-vectors is exponentially decays. 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 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 computer 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 'Waterloopkundig Laboratorium', a hydrological laboratory located in Delft [22]. 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 topical example is the use of mechanical simulations to study the structure of water. In the early 20th century Prof. J.D. Bernal and his fellow researchers built various ball and stick models of water to analyse the possible 3D configurations of water molecules in a liquid [21]. Their research eventually explained the peculiar physical properties of water from an atomistic perspective. Despite how 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 Prof. 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 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... [23]



Figure 1.7: Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis). [21]

After the first computer simulations were performed in the Los Alamos laboratory [24], the popularity of simulations rapidly increased. The remarkable explanatory power of simulations, combined with the relative easy construction of computer models, led 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,

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

The motion of the particles are governed by the forces f_i acting upon them, which are usually derived from the interaction potential \mathcal{U} . Solving these differential equations is achieved by employing a discretized time integration scheme. The discretization resolution is conventionally called the time step of the simulations denoted by Δt .

There are a large number of different integrations schemes that one can choose from, in which 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. This means that the errors on the total energy due to discretization are bounded.

On the other hand, when a system is in contact with a thermal reservoir –i.e. canonical ensemble– the total energy is not conserved, but rather the temperature of the simulation should be 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 time step. 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,

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

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

1.5.2 Coarse Grained modelling

As most things 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 interactions⁴ 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 arises from the fact that for each individual particle all the other particles are contributing to its energy 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 favourable 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 $O(Nt)$. [25]

From the scaling complexity we conclude that decreasing the total number of particles in the system drastically improves the computational cost of an MD simulation. To reduce the number of particles in a MD simulation a method called coarse-graining can be used. In all atom simulations each atom is explicitly represented as a particle in the simulation. Contrarily, coarse-grained simulation group together multiple atoms to form generalised pseudo-atoms, with their respective pseudo-interaction. During this simplification some atom details are neglected, resulting in a model that only approximates our initial system. However, useful results can be obtained from these simulations if we remain aware of the made approximations during the simplification.

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 top down approach, which is the alternative method, focuses more on the precise reproduction of experimental results. Here larger pseudo-atoms are designed based upon characteristic patterns in the structure. Next the pseudo-interactions are tweaked to accurately reproduce the system's dynamics.

In the field of DNA simulations coarse graining turned out to be a very important method. Previously all atom simulations of DNA polymers were restricted to simulations of less than hundred basepairs over only a few microseconds. The development of coarse-grained models allowed for the simulation of large scale systems, often encountered in DNA technology. A few examples of commonly used coarse grained models of DNA are Martini [26], 3SPN [27]

⁴referring to the

1. INTRODUCTION

and oxDNA [28].

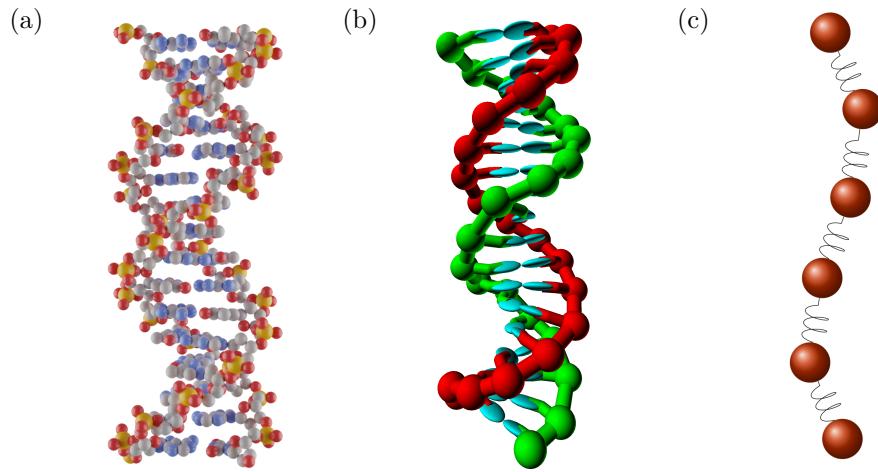


Figure 1.8: Comparing different coarse-grained models of DNA. (a) The most evident model is of course the All-atom model, providing an atomic description of the DNA strand. (b) A second model called OxDNA utilises the individual nucleotides as pseudo-atoms, thereby reducing the complexity while preserving much of the DNA structure. (c) Further increasing the level of coarse-graining leads to simplified polymer models, as the bead-and-spring model shown here. The double helix structure of DNA is not incorporated in this type of model, allowing only for the large scale behaviour to be studied.

CHAPTER

2

The DNA Nanopiston

Chapter Reference

The contents of the chapter is based on:

- Bayoumi, M., Nomidis, S. K., Willems, K., Carlon, E., and Maglia, G. (2021). Autonomous and active transport operated by an entropic dna piston. *Nano Letters*, 21(1):762768. PMID: 33342212.

Recently, a DNA nanopiston based molecular machine has been developed by the Maglia research group. Its main function constitutes the turning over of chemical fuel, in the form of ssDNA, into autonomous and active transport. The design is based upon the group's earlier work, where they developed a protein rotaxane [29], consisting of a polypeptide thread trapped in a ClyA nanopore by two stopper proteins. This rotaxane could be moved between two stable states inside the nanopore using an electric potential, acting as a molecular switch.

Motivated by the results from this research, the DNA nanopiston was developed by Bayoumi et al. in the Maglia group. In this new molecular machine the rotaxane constitutes of a DNA strand instead of a polypeptide thread. Utilising the thermodynamic transitions of DNA, this complex is capable of actively transporting DNA cargo-strands through the nanopore.

In this chapter the work of Bayoumi et al. is discussed, giving an overview of the construction and operating cycle of this molecular machine. At the end of this chapter the molecular dynamics simulations from the paper of Bayoumi et al. are discussed, as they were the main inspiration behind our project.

2.1 Rotaxane Formation

Synthetic molecular machines are often times embedded in larger complexes, providing the necessary structure for their operation. For this reason biological nanopores are a suitable starting point in the development of molecular machines. These transmembrane proteins spontaneously self-assemble into well-defined structures, embedded into a lipid bilayer. Extensive research has been performed towards designing methods and tools to tailor their structural and electrostatic properties for specific use cases, originally focused on ionic current spectroscopy. This large back catalogue of research can now be employed into building out their utility as an ideal building block for membrane bound molecular machines.

The design of the DNA nanopiston is centred around the biological nanopore Cytolysin A (ClyA). A modified variant ClyA-AS is used, which has been specially engineered for use in ionic current spectroscopy.[\[30\]](#) By virtue of the large inner lumen of ClyA, translocation of dsDNA through the pore is possible. The molecular machine utilises this property by capturing a DNA rotaxane structure inside of the pore, anchoring it to the lipid membrane. The rotaxane is composed of a DNA complex connected to two neutravidin protein stoppers via biotin, which serves to keep the structure captured in the pore. The DNA complex consists of three ssDNA strands: ssDNA 1, ssDNA 2, and ssDNA 3, shown in Figure

Formation of the DNA nanopiston takes place in a saline filled reservoir, where a ClyA nanopore is embedded into a membrane partitioning the reservoir into a cis- and trans-side. To start the formation process Neutravidin ($0.5 \mu M$), ssDNA 1 (5–biotinylated, 100 nt, $1.2 \mu M$) and ssDNA 2 (80 nt, $1 \mu M$) are added to the cis-compartment. Since the first 70 nucleotides of ssDNA 1 are complementary to the last 70 nucleotides of ssDNA 2, the two strands will partially hybridise. This results in a DNA duplex structure, on one side connected to a neutravidin stopper and on the other two ssDNA overhangs.

Applying a voltage of $+100 mV$ over the reservoir results in a net force guiding the DNA complex from the cis towards the trans-side. The applied potential drives the capturing of the DNA inside the nanopore, observed as a drop in the current through the pore (Fig ..). The complex remains indefinitely captured inside until the applied potential is reversed, restoring the open-pore current.

Finally, neutravidin ($1 \mu M$) and ssDNA 3 (3-biotinylated, 20 nt, $2 \mu M$) are brought into the solution at the trans-side, while keeping the potential at $+100 mV$. The longest overhang of the captured DNA thread is formed by the 30 free nucleotides of ssDNA 1. The added strand, ssDNA 3, is fully complementary with the last 20 nucleotides of this overhang, resulting in the hybridisation of both strands. To verify if the hybridisation has successfully taken place, the voltage over the reservoir is reversed. If no difference is measured in the pore-current, we conclude that the stable structure is formed, capturing the rotaxane inside the pore by the two protein stoppers.

After the hybridisation we find the final structure, which is referred to as the rotaxane-ds. Here the suffix 'ds' alludes to the predominantly dsDNA composition of this rotaxane configuration.

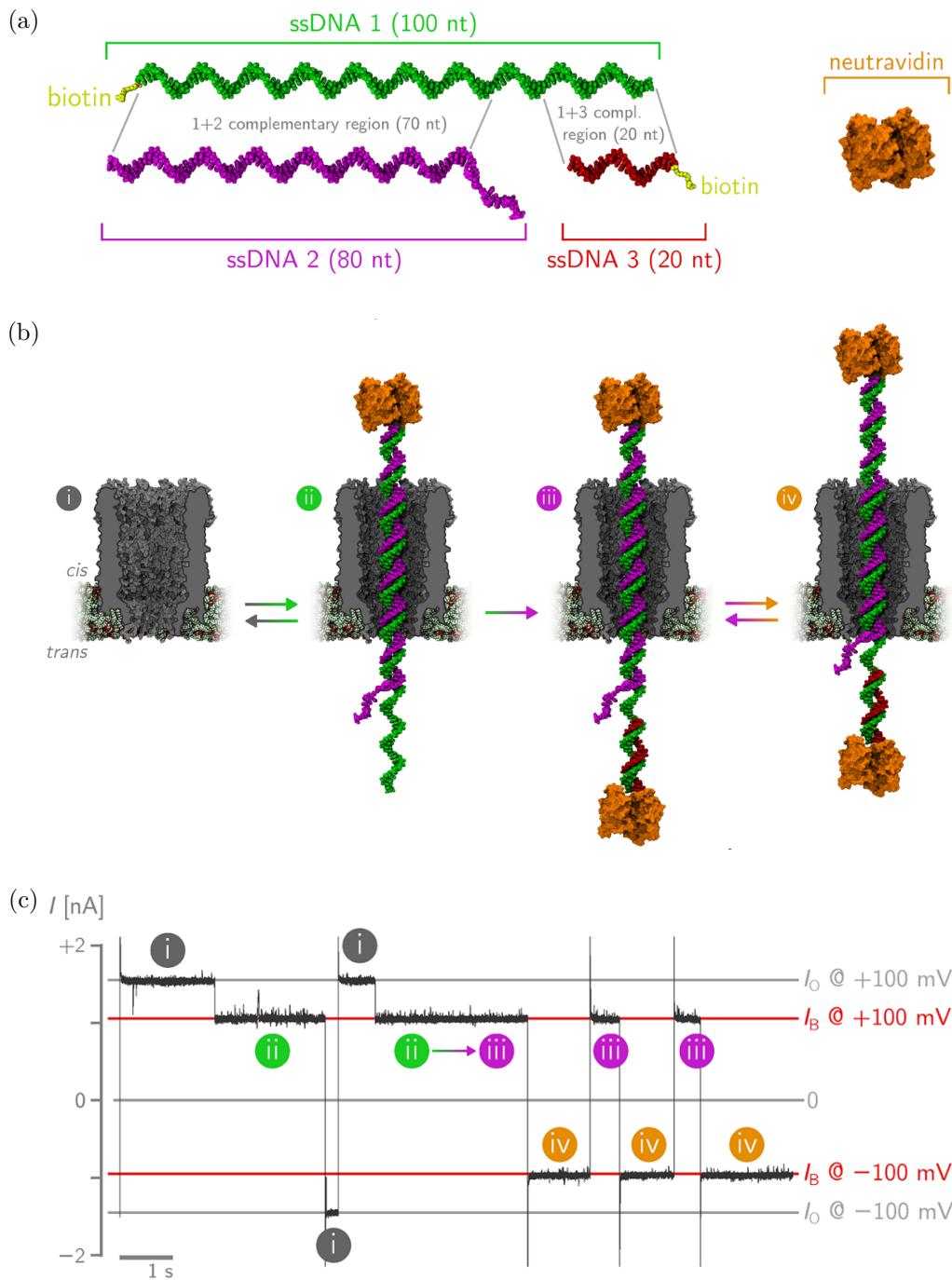


Figure 2.1: This is a figure [.]

2.2 Operating principles

Having successfully constructed the DNA nanopiston, the operation cycle (Figure ...) can now be discussed. For convenience, we take the rotaxane-ds configuration as the starting point of this cycle. The power stroke of the molecular machine is initiated by bringing the appropriate chemical fuel, ssDNA 4 ($0.5 \mu M$), into solution at the trans-side. The DNA strand is fully complementary to ssDNA 2, thereby inducing a toehold-mediated strand displacement. Here the flexible overhang located at the end of ssDNA 2, referred to as the toehold, is used to mediate the hybridisation of ssDNA 2 and 4.

During the strand displacement reaction, it is hypothesised that different transient states can possibly occur. One of the possible scenario's describes the hybridisation happening inside of the nanopore. This scenario is deemed to be unlikely, since this process would require three strands of ssDNA to be simultaneously present inside the constriction of the nanopore. Alternatively, the hybridisation can take place outside of the nanopore, in the trans-side of the reservoir. This process implies that the neutravidin protein would enter the lumen of the pore, which has been showed to be possible by previous studies. [31] This variation of the transient state is thereby thought to be the most probable.

The resulting configuration is called the rotaxane-ss in view of the fact that it is predominately composed of ssDNA. During this process a DNA duplex, composed of the ssDNA 2 and 4, is released into the trans-side of the reservoir.

Subsequently the cargo strand, $0.5 \mu M$ of ssDNA 2, is brought into solution at the cis-side, inducing the piston's recovery stroke. In this process the cargo hybridises with the rotaxane-ss, re-establishing a rotaxane-ds structure and completing the cycle. Each piston iteration transports one cargo strand, from the cis- to the trans-side of the nanopore, turning over one fuel strand in the process.

Important to note is that no external potential is specified for operating the DNA nanopiston. In contrast to earlier DNA transporters, the piston is able to function in a range of applied transmembrane biases. Experimentally, it is verified that the cycle operates at positive, $+20 mV$, $+50 mV$ and $+100 mV$ and negative, $-20 mV$. The limited range observed for negative voltages is most likely resulting from the inability of the fuel strand to hybridise with the toehold of rotaxane-ds. This hybridisation reaction is a rate limiting process, resulting in faster cycles at positive than at negative applied bias, shown in Figure Other factors, like the accessibility of cargo strands during the hybridisation of rotaxane-ss also influence the cycle rate.

The ability of the nanopiston to transport cargo both with and against an external bias is an important property of this molecular machine. It suggests that the externally applied bias might not be necessary for its functioning. Suspected is that the entropic interactions between the DNA strands and the nanopore are expected to play an important role in this behaviour. To accurately understand these effects, further analysis is needed.

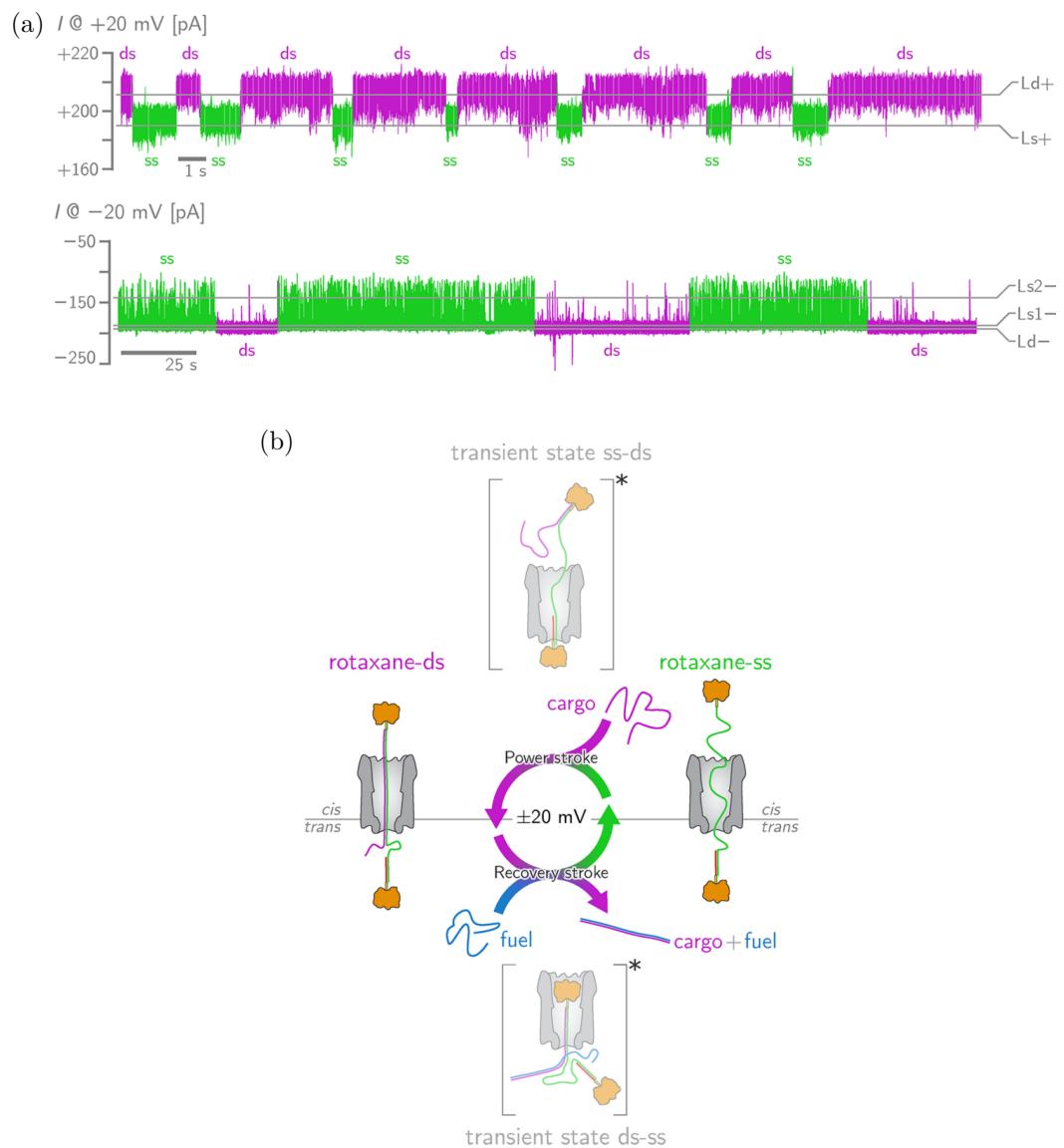


Figure 2.2: This is a figure

2.3 Coarse-grained simulations

During the discussion of the nanopiston's operation cycle it was hypothesised that entropic interactions play an important role in its functioning. Studying these phenomena is experimentally rather difficult, since all results are deduced from the measured current traces and known interactions between the different elements of the system. To obtain a more in-depth understanding of these interactions, computer simulations are needed to provide the necessary insights.

In view of these challenges, a coarse-grained model of the DNA nanopiston was devised by Bayoumi et al. The model is used to perform molecular dynamics simulations of the conformational fluctuations of the rotaxane at zero bias. Performing the simulations was done by using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) and its implementation of a Langevin intergrator. [32]

The coarse-grained model is composed of four types of interacting pseudo-atoms, all taking into account excluded volume interactions via repulsive Lennard-Jones potentials. First of all, the ClyA nanopore is represented by three vertically stacked open cylinders, with diameters 6 nm, 5.5 nm and 2.9 nm from the cis- to the trans-side. To take into account the electrostatic DNA-nanopore interactions, the cylinder radii are appropriately adjusted from the pore's geometry reported in. [33] These electrostatic interactions predominantly arise from an excess negative charge in the constriction of ClyA, resulting in a reduced effective size of the trans-cylinder. Note that the pore is excluded from the Langevin integration, resulting in a static pore model.

A semiflexible bead-and-spring model is used to simulate the rotaxane. Each spherical bead represents one ssDNA nucleotide or five dsDNA base pairs, respectively having a diameter of 1 nm and 2.2 nm. The bond connecting each consecutive pair of beads is represented by a harmonic spring. Determining the bond stiffness is done by means of the equipartition theorem, from which we find

$$k_{bond} = \frac{3k_b T}{\langle a \rangle}, \quad (2.1)$$

here $\langle a \rangle$ is the equilibrium bond length taken to be 0.68 nm or 1.7 nm for ssDNA and dsDNA respectively. In this model the bending rigidity of the DNA polymer is taken into account. This effect is modelled as previously seen in the discrete worm like chain, where the angle between consecutive bond vectors is assigned a harmonic potential. Using equation ..., the bending rigidity is determined by

$$\kappa_{bend} = \frac{l_p \kappa_b T}{\langle a \rangle}, \quad (2.2)$$

where l_p is the persistence length of the ssDNA (2.2 nm) or dsDNA (45 nm) strands. The difference in persistence length results in the relatively large flexibility of ssDNA.

The last components of the coarse-grained model are the neutravidin protein stoppers, which capture the rotaxane inside of the nanopore. During experiments it was observed that the neutravidin proteins can enter the lumen of ClyA. Using this information, the size of the

spherical stoppers was fitted to reproduce this behaviour in simulations. The fitted value is a diameter of 7 nm. The lipid bilayer in which the pore is embedded also interacts with these proteins, which is implemented by a reflective boundary at the lower entrance of the nanopore.

Having established a coarse-grained model of the DNA nanopiston, a computational analysis of the conformational fluctuations can be performed. A first conclusion drawn from these simulations is the importance of the toehold of rotaxane-ds. It was observed that the toehold was kept outside of the constriction of ClyA, resulting from the high entropic cost of confining the flexible strand of ssDNA. This affinity of the toehold to be outside of the nanopore exposes it for initiating the strand displacement, even if an opposing bias is applied.

The hybridisation of rotaxane-ss is also supported by entropic interactions. The interface between the ssDNA and dsDNA parts is kept outside of the nanopore, resulting from the entropic cost of confining the dsDNA in the constriction of ClyA. Keeping the interface outside of the pore prevents the sequestering of the toehold during hybridisation.

These simulations clearly highlight the importance of entropic interactions between the DNA rotaxane and the nanopore during the piston's operation cycle. However, the model used by Bayoumi et al.[1] imposes limitations on the possible research that can be performed. More advanced methods are needed, which are discussed in the following chapters of the thesis.

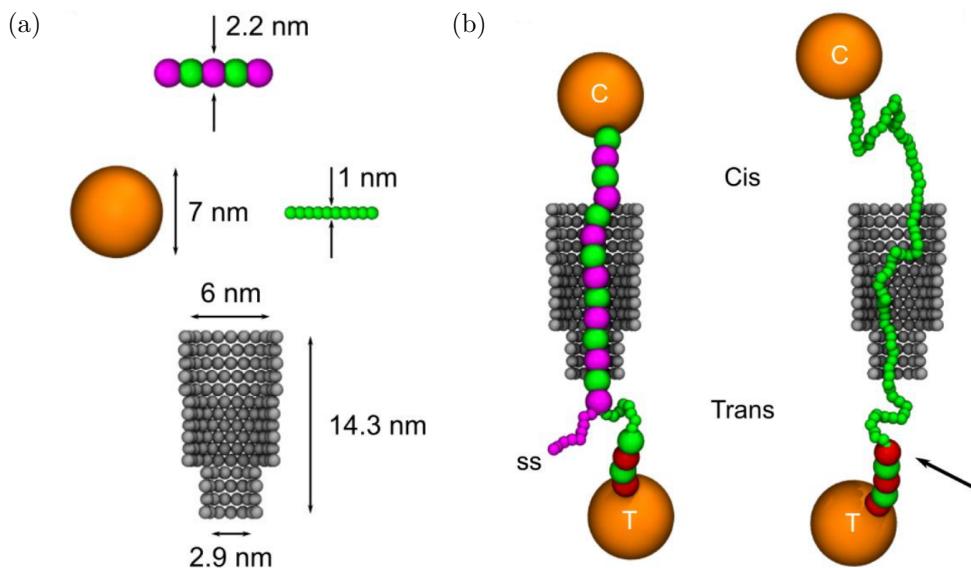


Figure 2.3: This is a figure

CHAPTER 3

Improving the Model

All models are wrong, but some are useful.

— George Box

Although the simulations discussed in the previous chapter provided an answer to important questions regarding the piston's operation, some remarks need to be made. Due to the nature of this bead-and-spring model it does not allow for hybridisation reactions to be simulated. Improving upon this model by using a more advanced coarse-grained representation of DNA is therefore needed, providing also a more realistic description of the conformational fluctuations. In this thesis we improved upon this model by using the OxDNA coarse-grained model for DNA.

3.1 OxDNA

OxDNA is a coarse-grained model of DNA developed by Thomas E. Ouldridge et al. at the University of Oxford.[34][28] 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,

3. IMPROVING 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 [26] en 3SPN [27], 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.

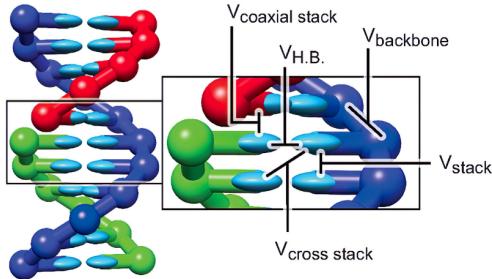


Figure 3.1: Structure of the OxDNA model with the different interactions. Figure was taken from. [28]

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 [26] en 3SPN [27], 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.

3.2 DNA Thermodynamics

The field of DNA thermodynamics focuses on understanding how the structure of DNA varies with temperature. Due to the nature of 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, shown in Figure 3.2, driven by a reduction in free energy due to the bonding of complementary base-pairs. The latter is called DNA melting, a process

observed at high temperatures. This dissociation is energetically driven, since the reduction in free energy due to base-pair hybridisation is no longer a favourable trade-off with the reduction in configurational entropy in the duplex structure.

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

Initiating a hybridisation reaction between two strands of ssDNA incurs a thermodynamic penalty. This penalty originates in the decrease of configurational entropy, when the strands start to form a duplex. This has a consequence that initial contacts in these reactions often dissociate, due to the initial energy barrier that needs to be crossed before the full hybridisation becomes energetically favourable. Even when an initial contact results in the stabilisation of a dsDNA duplex for select base-pairs, the configuration often times is not conducive to full duplex formation. Especially in repetitive sequences, the chance of a mismatched initial hybridisation is significant.

Another limiting factor to the rate constant of hybridisation reactions is that these transitions are not characterised by a single state, but rather by an ensemble of possible transition pathways. The number of pathways increases dramatically when the strand sequences are repetitive, giving rise to hybridisation pathways facilitated by Inchworm and pseudo-knot displacements[35].

The combination of the unstable initialisation of hybridisation reactions together with its many transition pathways, complicates the analysis of the full reaction kinetics observed in DNA hybridisation.

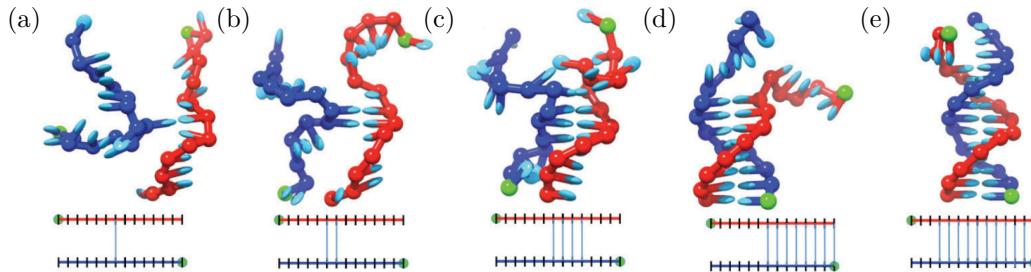


Figure 3.2: This is a figure [35]

The other important thermodynamic transition in the operating cycle of the nanopiston is a toehold mediated strand displacement. Initially this reaction consists out of two components. The first is an imperfect duplex structure, formed by a substrate strand and an incumbent strand. The two strands are partially complementary by having either a mismatch in their base-pair sequence or a surplus of base-pairs on the substrate strand. The non-hybridised part of the substrate constitutes a flexible strand of ssDNA that is referred to as the toehold.

The second component is called the invasive strand, and is fully complementary with the

3. IMPROVING THE MODEL

substrate. It is energetically favourable for the invading strand to disrupt the imperfect duplex structure and form a fully Watson-Crick complementary dsDNA with the substrate strand. This displacement reaction results in an overall reduction in the free energy of the system, since the strand displacement increases the total number of hybridised base-pairs.

The process of strand displacement starts with the hybridisation of the toehold and the invading strand. Once this initial hybridisation has occurred the invading strand can start to contest hybridised base-pairs of the imperfect duplex. Disrupting the base-pairing of the duplex is referred to as fraying, while the reverse process where new base-pairs are formed is called zippering. During this process the invading strand competes with the incumbent strand to form base-pairs with the substrate.

The reaction can be modelled using an one-dimensional energy landscape, called the intuitive energy landscape (IEL) model[36], shown in Figure 3.3. In the shape of the energy landscape we recognise two distinct features. The first features is the initial energy barrier, also seen in DNA hybridisation. This energy barrier again arises from a reduction in configurational entropy, when the initial binding happens. The second feature is the plateau, representing the change in free energy when the strand displacement takes place. This plateau gives rise to a relatively slow reaction, which can be explained using a simple toy model.

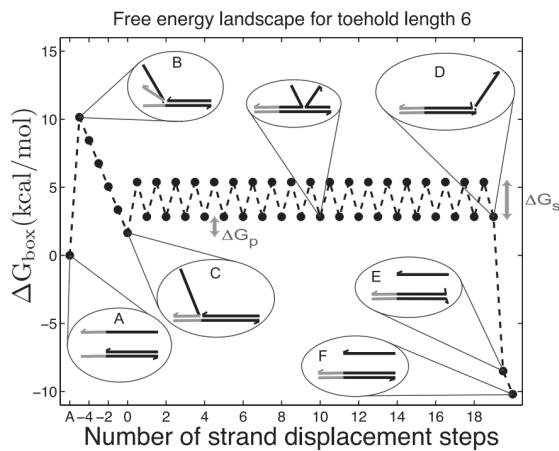


Figure 3.3: write caption [36]

Considering a scenario, where the substrate consists of $N + 1$ nucleotide, with only one of these nucleotides constituting the toehold. If we now assume that both the incumbent and invasive strand contest each others hybridised base-pairs at the same rate, the displacement reaction can be modelled as a random walk. The model is reduced to a famous problem in probability theory, called the gamblers ruin. It can be shown that the reaction rate scales as $1/3N$. Combining this result with fact that there exist various possible branch migration pathways, it can be concluded that studying the full reaction kinetics of toehold displacement reaction is computationally expensive for large strands.[36]

These two types of thermodynamic transitions, central in the operation cycle of the DNA nanopiston, are relatively difficult to study due to their complex reaction kinetics and initial energy barrier. Accurately analysing the reaction constants of these rare events, requires the use of advanced sampling techniques.

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 difference in the time-scales corresponding to 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, the hybridisation and toehold displacement reactions studied in this thesis can be classified as a rare event. Due to the large temporal spacing between these rare events, simulating them with a brute-force approach is inefficient. In this case, molecular dynamics simulations spend a lot of computational resources on simulating the waiting 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 encompasses 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 configurational space are explored. Notable examples of these methods are the adaptive biasing force algorithm [37], basis function sampling[38] and umbrella sampling [39].

The second class of methods, known as path sampling methods, do not influence the systems Hamiltonian, but rather interface directly with the simulated 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 [40] and forward flux sampling [41] [42]. 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, Ω , using a set of nonintersecting hypersurfaces. By design, we choose this order parameter to be a function, $\lambda(\cdot) : \Omega \rightarrow \mathcal{R}$, monotonically increasing from the initial state A to 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, λ_A and λ_B , construct the interfaces separating the two local energy basins from the rest of the phase space. Finally, this procedure

3. IMPROVING THE MODEL

can be done for a N -number of interfaces partitioning the space between A and B , for which we require

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

Note that this method 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. Analogous to the ambiguous choice of a collective variable in free energy methods, constructing these hypersurfaces is often more an art than 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 , 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 λ_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 λ_0 and the subsequent probability of reaching the final state from this interface. Using the previously defined interfaces, we can now factorise 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, since 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, differing in the approach by which they calculate the probability $P(\lambda_n | \lambda_0)$. During this thesis I chose to use the Rosenbluth-like (RB) method [41]. 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 λ_0 interface by running simulations in the A basin. Keeping track of the fraction of successful runs, $\langle \Phi_{A,0} \rangle / \langle h_A \rangle$ is evaluated.
- (ii) Fire k_0 trial runs from generated configurations on λ_0 until they cross λ_1 or cross back to λ_0 . Store the final configurations of the successful simulations.

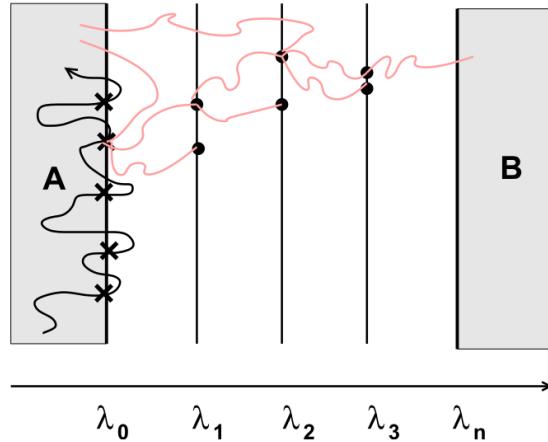


Figure 3.4: write caption[.]

- (iii) Sample one of the saved configuration on the λ_1 interface and use it to shoot k_1 runs to the next interface λ_2 .
- (iv) Iterate the previous steps until the trajectories reach λ_n or no more configurations are available.
- (v) If not successful, sample a stored configuration on λ_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 path b starting at the initial basin and reaching interface λ_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 crossing interface j 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 Computational setup

The model used to study the DNA nanopiston, is largely based on the model previously devised by Bayoumi et al. The main variation between the two models lays in the different coarse-grained models, used to simulate the DNA strands. As discussed in Chapter 2, the

3. IMPROVING THE MODEL

Bayoumi model uses a bead-spring approach to simulate DNA strands, where we use the more sophisticated model oxDNA. This DNA model gives a better representation of the dynamics of DNA strands, at the same time allowing for accurate simulations of the thermodynamic transitions in the DNA nanopiston.

The simulations are performed using the popular molecular dynamics simulator, LAMMPS [32]. Employing the Lammps implementation of oxDNA developed by Henri et al [43], it becomes possible to study the interactions between oxDNA strands and externally defined particles. The initial configurations of the simulations are generated using the Moltemplate package [44], a general purpose molecule builder for LAMMPS.

The molecular dynamics simulations performed in this thesis utilises a Langevin thermostat, more precisely the Dot-C Langevin integrator also implemented by Henri et al. This is a LAMMPS implementation of the Langevin C integrator developed by Davidchack et al. [45], falling in the class of rigid-body Langevin-type integrators. This type of thermostat separates the stochastic and deterministic parts of a Langevin thermostat to efficiently take into account the extra degrees of freedom in the system, arising from the non-spherical shape of the oxDNA beads. As is common practice in MD simulations, the diffusion coefficient of the oxDNA strand is chosen larger than the value of physical DNA. This is done to speed up the simulations, while ensuring its physical accuracy.

The model is used to more accurately study the conformational fluctuations of the DNA Rotaxane and develop understanding of the entropic interactions between the DNA and the nanopore. Next the thermodynamic transitions in the operation cycle of the piston are simulated using a forward flux sampling algorithm. The FFS algorithm is implemented as a Python script, performing the simulations by interfacing with the Python API of LAMMPS.

CHAPTER 4

Simulations of the Rotaxane

...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*

In this chapter the results of various simulations utilising the OxDNA based rotaxane model are presented. The aim of this computational analysis is to shed light on the entropic effects which play a central role in the nanopiston's operating cycle. By extension, the thermodynamic transitions providing the free energy governing its operation are studied.

To better understand the entropic interactions between the DNA rotaxane and the nanopore a specially engineered rotaxane variant is studied. This other class of rotaxane is called the mixed rotaxane, composing of different ds- and ssDNA fractions in an attempt to isolate the influence of entropic interactions. Having explored the effects of the entropic contributions, next the conformational fluctuations of the nanopiston are simulated and analysed. Lastly, an attempt is made to simulate the thermodynamic transitions driving the piston cycle. These particular simulations are made possible by the use of a forward flux sampling algorithm.

4.1 Mixed Rotaxane

Having identified the entropic interactions as key component of the nanopiston cycle, studying them is an important but challenging task. The problem arises when we aim to specify the entropic contributions to the conformal fluctuations of the rotaxane. The main factor complicating this analysis is the multiplicity of the interactions acting upon the nanopiston.

The first category of forces is composed of the external forces arising from the potential difference applied over the membrane. This potential difference induces an electric field both outside and inside of the nanopore, influencing the movement of molecules in these regions. The most significant contributions can be identified as the electrophoretic and electro-osmotic forces acting upon the rotaxane. The second class of forces is composed of the intrinsic forces. These forces do not arise from the electric field, but rather originate from the direct interactions between the rotaxane and the nanopore itself. Under this category fall the electrostatic, steric and entropic forces, limiting the conformational freedom of the piston.

In order to study the role of entropy in the conformational fluctuations of the nanopiston, these effects need to be isolated from the other contributions present in the system. To achieve this Bayoumi et al. devised a new variation of the rotaxane called the mixed rotaxane. In this rotaxane variation the toehold is removed, resulting in a thread composed of solely ds- and ssDNA. By varying the DNA composition of this rotaxane the changes in entropic interactions can be analysed. More specifically, the total length of the rotaxane was constant, i.e. the total number of nucleotides and basepairs are fixed at 100, while the ssDNA part was varied from 0nt to 30nt in steps of 10nt. The effects of these changes in composition were first studied using I-V measurements in the range of $[-100mV, +100mV]$. During an I-V measurement a range of voltages is applied over the lipid membrane and at each step in this voltage sweep the current through the pore is measured. A detailed explanation of this method is described in. [46] The results from these experiments are presented in Figure....

The mixed rotaxane composed entirely of dsDNA yields an almost linear I-V trace in the measured voltage range. This result indicates a constant obstruction of the nanopore over the entire voltage sweep. On the other hand, the mixed rotaxane composed of a 10nt ssDNA strand shows a drastically different I-V trace. Decreasing the voltage below $-20mV$ results in current fluctuations between two distinct levels. A second transition can be observed when the voltage is further decreased below $-70mV$. Here the measured current becomes independent from the applied voltage, suggesting a partial blockage of the current through the nanopore. These same characteristics can be identified in the I-V trace of the mixed rotaxane with a 20nt ssDNA strand. Remarkably, the voltage range in which the current fluctuations are observed is moved to a more negative voltage range from $-30mV$ to $-85mV$. When the fraction of ssDNA in the mixed rotaxane is further increased to 30nt, the characteristic blockage is no longer observed and the linear I-V trace is recovered.

To interpret the observed characteristics in the I-V traces, molecular dynamics simulation where performed of the mixed rotaxanes. The rotaxanes were simulated with no bias, i.e. $0mV$, to solely highlight the changes in the entropic effects. Quantifying the conformational

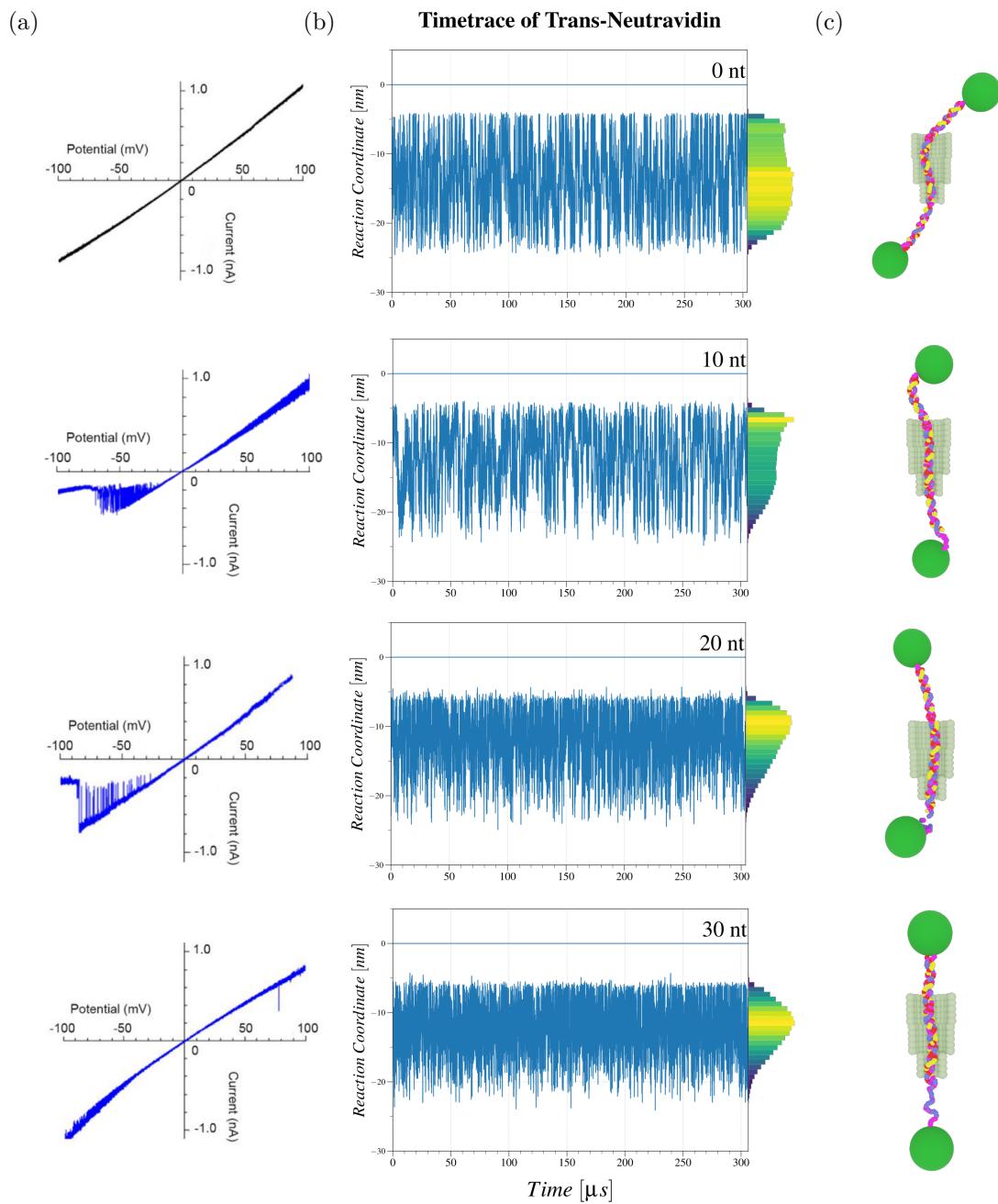


Figure 4.1: This is a figure.

4. SIMULATIONS OF THE ROTAXANE

fluctuations of the rotaxane is done by tracing the trans-stopper protein during the simulation using a reaction coordinate defined in our system. This reaction coordinate, X , is determined as,

$$X = \begin{cases} z_0 + |\mathbf{r} - \mathbf{r}_{cis}|, & \text{if on cis-side} \\ z, & \text{if inside pore} \\ -|\mathbf{r}|, & \text{if on trans-side} \end{cases} \quad (4.1)$$

Here the z-axis is aligned with the symmetry axis of the nanopore, placing the origin of the coordinate system at center of the pore's trans-entrance. With respect to this coordinate system, the center of pore's cis-entrance is located at $r_{cis} = (0, 0, z_0)$.

In Figure ... the measured time trace of the X -coordinate is presented for each mixed rotaxane type. The horizontal line indicates the origin of the reaction coordinate, representing the trans-entrance of the pore. On the right side of these graphs the X histograms of the time traces are presented, indicating the positional distribution of the trans-stopper during the simulations. It should be noted that the time traces are not obtained from one continuous simulation, but rather aggregated from various independent simulations performed in parallel, aiming to reduce the simulation time.

For the rotaxane fully composed of dsDNA, i.e. $0nt$, a uniform X histogram is found. This result indicates free diffusion of the rotaxane within a bounded one dimensional domain, namely the nanopore. In the case where $10nt$ and $20nt$ are substituted into the composition of the rotaxane a peak is observed in the X histogram. This peak indicates a tendency of the trans-stopper to move towards the entrance of the nanopore. This observation is in agreement with the previously discussed I-V curves, since the presence of the stopper at the pore entrance partially blocks the current flow. This tendency is driven by an entropic force arising from the smaller ssDNA strand, being more easily captured in the constriction of the pore compared to the large dsDNA strand.

Next, we observe that increasing the length of the ssDNA part of the mixed rotaxane gradually shifts the histogram's peak further away from the pore entrance. Confining a portion of the freely fluctuating ssDNA strand inside of the pore, drastically reduces the number of available configurational microstates and thereby also its entropy. This change in entropy induces an opposing entropic force. An equilibrium configuration is found, when both the large dsDNA is kept outside of the constriction, while allowing a maximal length of ssDNA strand to freely fluctuate outside of the pore. These combined effects shift the peak of the distribution further away from the pore entrance as the number of nucleotides is increased. An ever larger electrophoretic force is required to sustain full pore blockage, which is why no blockage is observed in the experiment's voltage range. The results found in these simulations are in accordance with the corresponding simulations performed by Bayoumi et al.[], using the bead-and-spring model.

As the final part in our analysis of the mixed rotaxanes, the confined diffusion of the $0nt$ mixed rotaxane is studied. The uniform distribution of the X histogram suggests that the rotaxane vertically fluctuates inside of the pore like a rigid rod. To study this diffusive behaviour, the mean square displacements(MSD) of the trans-stopper is determined within a

range of time intervals.¹ Calculating the MSD is done by utilising a rolling-window analysis on the measured time trace. The MSD is an important quantity used to characterise diffusive processes, since it allows us to verify whether an external force influences the motion. From analytical calculations, presented in appendix A, we find that one dimensional confined diffusion can be described up to the second order as,

$$\langle \Delta x \rangle = \frac{L^2}{6} - \frac{96}{\pi^4} e^{-\frac{D\pi^2 t}{L^2}},$$

here L is the length of the confined region, i.e. the hight of the pore, and D the diffusion constant of the particle. This model was fitted to our simulations, for which the result is presented in Figure... From this we can conclude the *Ont* mixed rotaxane can be well described using the simple model of one dimensional confined diffusion, confirming our postulation. This model predicts the diffusion coefficient of the mixed rotaxane as, $69.08 \pm 0.02 \mu s^{-1}$, giving us an idea of the speed of the rotaxane dynamics.

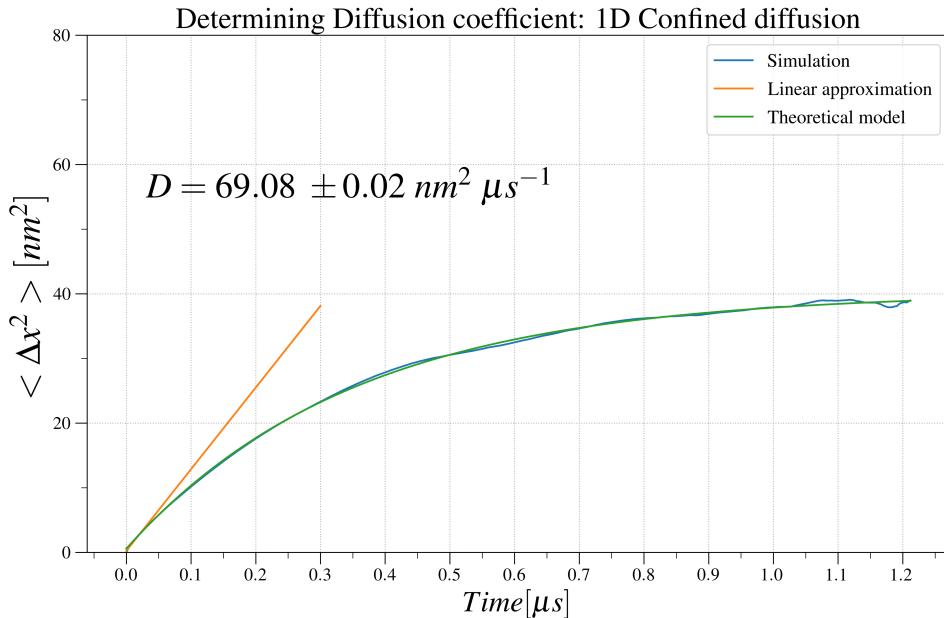


Figure 4.2: write caption

¹reference tidynamics by Pierre de Buyl

4.2 Conformational Fluctuations of the Rotaxane

Having gained insight into the entropic interactions between the DNA and the nanopore, now the stable intermediate states of the nanopiston's operation cycle are studied. Both the rotaxane-ss and -ds are composed of stiff dsDNA and flexible ssDNA parts, which characterise their conformational fluctuations by the entropic interactions. The two rotaxanes types are simulated in a nanopore, using the predefined coarse-grained model, in absence of an external bias, i.e. 0 mV. The results are presented in Figure ...

Analysing the conformational fluctuations of the rotaxane-ds, we observe that the ssDNA toehold remains predominantly outside of the pore. This effect can be explained by taking into account the entropic cost of capturing the flexible strand of ssDNA into the constriction of the pore. Since the geometry of the rotaxane prohibits the toehold from reaching the cis-side of the pore, the toehold can only freely fluctuate on the trans-side of the pore. Throughout the simulation this entropic force keeps the toehold in the trans-reservoir and thereby placing the cis-protein stopper close to the entrance of the pore. This entropic interaction plays an important role in the operation of the nanopiston. Even when an external voltage difference induces an upward electrophoretic force on the rotaxane, the competing entropic force keeps the toehold outside of the constriction and thereby exposing it for hybridisation with a fuel strand. This also explains the halting of the piston cycle at high voltages. In this case the entropic force is overcome by the induced electrophoretic force, sequestering the toehold inside of the pore inhibiting the binding of fuel strands.

In the same figure the results for the rotaxane-ss are presented. An upwards entropic force is observed in the positional histograms, arising from the high flexibility of the long ssDNA strand. This force originates from the increase in configurational microstates available to the rotaxane-ss, when the ssDNA strand is allowed to freely fluctuate in the cis-reservoir. The flexibility of the ssDNA strand overcomes the entropic penalty of confining the dsDNA into the constriction of the pore. Capturing the interface between the ssDNA and dsDNA parts of the rotaxane-ss inside of the pore promotes the operation cycle. In this case a longer ssDNA strand is exposed to the cis-reservoir, better facilitating the hybridisation with cargo strands. This can be seen in the positional histogram of the cis-protein stopper, where the large fluctuations of the ssDNA strand allow it to venture far away from the nanopore.

These results explain the functional importance of the entropic interactions in the nanopiston's operating cycle. Comparing our findings with the results by Bayoumi et al., we see that both models are in reasonable accordance. However, in our simulations the interface of the rotaxane-ss is observed entering inside the pore's constriction, while in the bead-and-spring model this behaviour is not observed. This difference can be attributed to the more accurate simulating of ssDNA by OxDNA, mainly arising from the more precise parametrisation of the model and the ability for consecutive bases to unstack.

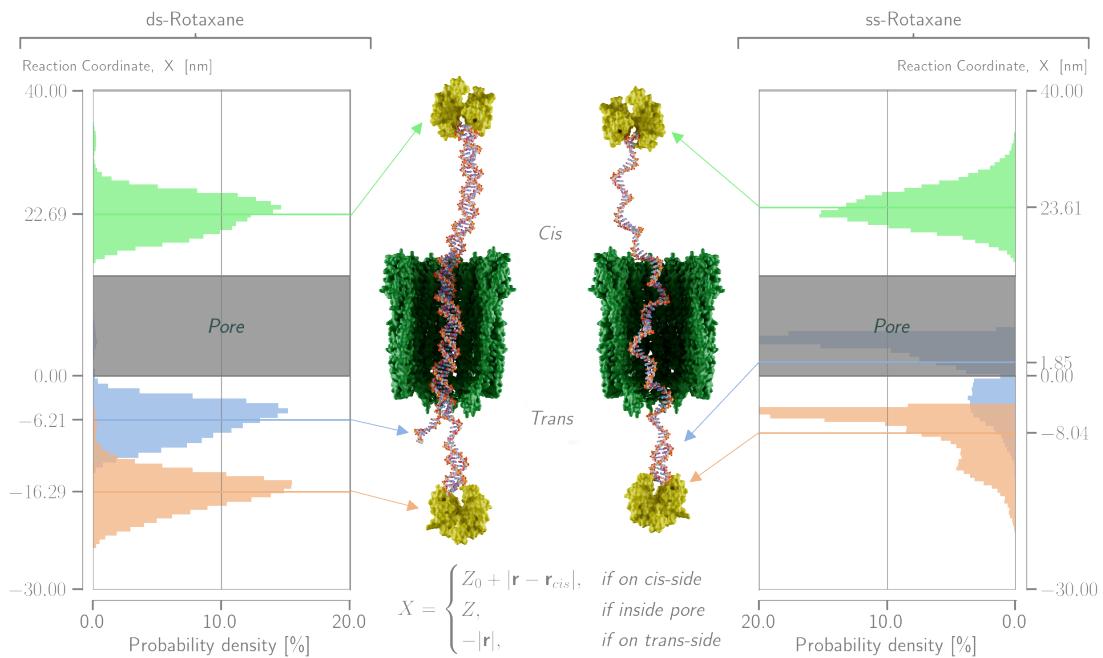


Figure 4.3: write caption

4.3 Hybridisation Reactions

Hybridisation reactions are central to the operation cycle of the DNA nanopiston. By providing the required free energy, they facilitate the transitions between the two stable states of the cycle, rotaxane-ss and rotaxane-ds. Combining these transitions with the previously discussed entropic interactions a ratcheting mechanism enables the extraction of useful work from the inherently stochastic system.

The associated length and time scales of these reactions make their in-depth analysis experimentally challenging. For this reason scientists resort to computational simulations for high resolution analyses of these reactions. To study the DNA hybridisation occurring during the piston operation cycle, we utilised our OxDNA based model of the piston which is simulated using molecular dynamics. As previously discussed in chapter three, the intrinsic energy landscape associated with these reactions complicates brute force simulations of the transitions. To overcome these limitations a forward flux sampling algorithm was employed. Illustrating the viability of this technique, first both the hybridisation and the toehold displacement reactions of the rotaxanes were simulated outside of the nanopore. From these simulations it was confirmed, that using the forward flux sampling algorithm the full ensemble of transition path ways present in the hybridisation reactions can be studied.

4. SIMULATIONS OF THE ROTAXANE

However, performing these same simulations with the rotaxanes placed inside of the nanopore inhibited the reaction from occurring. During the final stages of both the toehold displacement and hybridisation reaction, the geometry of the rotaxane forces three ssDNA strands inside of the pore's constriction. The diameter of the pore was modelled to carefully capture both the electrostatic and excluded volume interactions between the ClyA pore and the DNA, resulting in a diameter of 2.9 nm , compared to the 1 nm width of the ssDNA strand. Due to the static nature of our coarse-grained pore model, the diameter of the pore constriction does not facilitate the three ssDNA strand to enter the constriction simultaneously.

In literature we find an indication that the α -helices constituting the pore's constriction allow for a reasonable amount of structural fluctuations of the constriction.[\[10\]](#) Our simulations indicate that the compliancy of the pore entrance is essential in the hybridisation reaction of the rotaxanes, but were not taken into account in our model. This limitation of our coarse-grained model impedes the full analysis of the DNA nanopiston's operating cycle. Improving upon this model by including the compliance of the pore's constriction was attempted, but did not succeed within the time constraints of this thesis.

Conclusions and Perspectives

Blurring the line between life and the artificial.

— *Man-made molecular machines*, Watson & Cockcroft

In this work the novel molecular machine devised by Bayoumi et al. has been studied using molecular dynamics simulations. The primary objective of this research was to shed light on the operating principles, facilitating the autonomous and active transport, which characterise this DNA nanopiston. In this chapter the obtained results are summarized, after which recommendations for future research are made.

5.1 Results & Conclusions

Understanding the underlying processes driving the operation of molecular machines is an important corner stone in furthering their development. As a result of the scale and complexity of these structures, performing experimental studies aiming to shed light on these interactions has been proven to be challenging. In this work we aimed to provide an insight into the operation of the DNA nanopiston using molecular dynamics simulations.

This thesis builds further upon simulations presented in the original paper, where a coarse-grained model of the DNA nanopiston was developed based on both theoretical and experimental considerations. The central component of this model is the DNA rotaxane which was simulated using a bead-and-spring model. To increase the scope of the research we designed a new coarse-grained model of the DNA nanopiston, utilising a more accurate representation of DNA, namely OxDNA. This new model yields a more realistic description of DNA, while also enabling us to simulate DNA hybridisation reactions.

5. CONCLUSIONS AND PERSPECTIVES

The entropic penalty of confining single stranded DNA inside of the nanopore was presumed to play an important role in the mechanisms driving the DNA piston. We studied these entropic effects by simulating a specifically engineered class of rotaxanes, called the mixed rotaxanes. These simulations indicated that there are two competing entropic interactions present in the rotaxane-pore complex. A first entropic force was found to arise from confining the large dsDNA strands of the rotaxane inside of the pore. Competing with this force were the smaller but more flexible ssDNA strands, which endeavour to maximize their available configurational space by opposing the confinement of the pore. As the lengths of these ssDNA strands were increased, this latter interaction started to dominate.

After establishing this essential understanding of the entropic interactions, the two stable states of the piston's operating cycle were simulated (i.e. rotaxane-ss and rotaxane-ds). These simulations indicated that the sequestering of the rotaxane-ds overhang was prevented by the entropic penalty of confining it in the pore. The established entropic force enabled the hybridisation with fuel strands even opposing an upward electrophoretic force, differentiating the DNA nanopiston as an active transporter. Due to the rotaxane geometry this hybridisation reaction resulted in the formation of rotaxane-ss in a low-entropy state. As a consequence the rotaxane-ss spontaneously migrated in the trans-cis direction exposing the flexible ssDNA strand to the cis-reservoir, enclosing the dsDNA fraction inside of the pore. This upward motion played an important role in exposing the ssDNA fraction of rotaxane-ss to the cis-reservoir, where it can hybridise with a cargo strand. Here we concluded that the interplay of the hybridisation reactions and the entropic interactions collectively drive the molecular machine.

In the last part of this thesis an attempt was made to study the ensemble of hybridisation pathways providing the free energy to drive the nanopiston. The study of these thermodynamic transitions was complicated by the complex reaction kinetics and an initial energy barrier. To overcome this difficulty an forward flux sampling algorithm was used to analyse these transitions. From the performed simulations we concluded that the static representation of the ClyA pore in our coarse-grained model inhibits these hybridisation reactions. The compliance of the pore constriction is found to be essential in facilitating the hybridisation reactions. However, this was not yet incorporated in our model.

The performed computational analysis of the DNA nanopiston confirmed previous research on the operating principles of the piston cycle. The performed exploratory research motivates further research towards the fundamental aspects of this nanopiston. Eventually, this knowledge will aid the development of novel molecular devices, further blurring the line between life and the artificial.

5.2 Future Perspectives

During this thesis we were able to deepen our understanding of the DNA nanopiston. However, some important questions remained unanswered. While studying the hybridisation reactions, which drive the molecular machine, we encountered the limitations of our coarse-grained model. Simulations indicated that the compliancy of the nanopore is an integral component

in facilitating these reactions. In our current model the ClyA nanopore is represented as a static complex, not allowing the diameter fluctuations that are needed for the hybridisation to take place. Future research should attempt to incorporate a more accurate representation of the ClyA nanopore in the coarse-grained model of the DNA nanopiston.

Various approaches can be taken to accomplish this improved accuracy. A first proposition would be to expand further upon the existing model, by incorporating the constriction of the pore in the Langevin integrator. Using experimental data, we can parametrise the interactions between the constituent beads of the pore and reproduce the diameter fluctuations of the constriction of the ClyA pore. Another possible solution could be to use an already parametrised coarse-grained model. An example is the Martini force-field, which allows for accurate simulations of transmembrane proteins such as the ClyA pore. In this new model both OxDNA or the Martini forcefield could be used to simulate the DNA strand. These possible improvements would increase the accuracy of our model, probably enabling the simulation of a full piston cycle.

Molecular dynamics simulations will prove to be a useful tool in the development of new molecular machines. Future research is focused on designing molecular devices that are able to not only transport DNA, but also other molecules through a membrane. Eventually the aim would be to implement these new molecular machines into a sophisticated droplet-based device with emergent properties, i.e. a fully synthetic cell.

APPENDIX A

One Dimensional Confined Diffusion

This appendix provides a detailed description of the motion of a Brownian particle confined to a one dimensional domain. Due to the statistical nature of this motion we will discuss the evolution of the probability density function $\phi(x, t)$, which represents the probability of finding the particle on the position x at time t . Once the evolution of this probability function is known, the statistical properties can be evaluated. Here we will discuss the mean square displacement (MSD) as it is used to analyse the motion of the fully double stranded mixed rotaxane.

A central result in statistical mechanism is that the evolution of ϕ is described by the diffusion equation, given by

$$\frac{\partial \psi}{\partial t} = D \frac{\partial^2 \psi}{\partial x^2}, \quad (\text{A.1})$$

here D is the diffusion coefficient of our particle. The confinement is imposed through reflecting boundary conditions, equivalent to a vanishing particle current at the boundaries of our domain, $j = -D \frac{\partial \psi}{\partial x} = 0$. Solving the heat the equation can be done using the method of separation of variables, here we assume that the solution can be expressed as $\psi(x, t) = f(x)g(t)$. Upon substitution [A.1](#) becomes,

$$\frac{\dot{g}}{g} = \frac{\ddot{f}}{f} = -\alpha, \quad (\text{A.2})$$

here the two expressions are implied to be constant since the variables are independent. The partial differential equation is now treated as two separate ordinary differential equations, for which we find,

$$t : \quad \dot{g} = -\alpha g(t) \Rightarrow g(t) = e^{-\alpha t}, \quad (\text{A.3})$$

$$x : D\ddot{f} = -\alpha f(x) \Rightarrow f(x) = A \sin(Kx) + B \cos(Kx) \quad (\text{A.4})$$

$$= B \cos\left(\frac{\pi n x}{L}\right). \quad (\text{A.5})$$

In the latter expression the boundary conditions impose $A = 0$ and the following relation,

$$\frac{\alpha}{D} = \frac{\pi^2 n^2}{L^2}. \quad (\text{A.6})$$

By substituting the found results into the assumed form of the solution we find the general solution by the linear combination,

$$\psi(x, t) = \sum_{n=0}^{+\infty} C_n \cos\left(\frac{\pi n x}{L}\right) e^{-\frac{D\pi^2 n^2}{L^2} t} \quad (\text{A.7})$$

$$\psi(x, 0) = \delta(x - x_0) = \sum_{n=0}^{+\infty} C_n \cos\left(\frac{\pi n x}{L}\right) \quad (\text{A.8})$$

$$\psi(x, t) = \frac{1}{L} \left[1 + \sum_{n=1}^{+\infty} \cos\left(\frac{\pi n x_0}{L}\right) \cos\left(\frac{\pi n x}{L}\right) e^{-\frac{D\pi^2 n^2}{L^2} t} \right]. \quad (\text{A.9})$$

This expression describes the behaviour of a Brownian particle in a one dimensional confined domain. Using the found expression the MSD is calculated to be,

$$\langle \Delta x^2 \rangle = \langle (x - x_0)^2 \rangle \quad (\text{A.10})$$

$$= \frac{L^2}{6} \left[1 - \frac{96}{\pi^4} \sum_{k=0}^{+\infty} \frac{1}{(2k+1)^4} e^{-\frac{D(2k+1)^2 \pi^2}{L^2} t} \right]. \quad (\text{A.11})$$

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 of the exponential and find,

$$\langle \Delta x^2 \rangle = \frac{L^2}{6} - \frac{16L^2}{\pi^4} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^4} + \frac{16Dt}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} + \mathcal{O}\left(\frac{D^2 t^2}{L^4}\right). \quad (\text{A.12})$$

Using the two convergent series,

$$\sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \frac{\pi^2}{8} \quad \text{and} \quad \sum_{k=0}^{\infty} \frac{1}{(2k+1)^4} = \frac{\pi^4}{96} \quad (\text{A.13})$$

the free diffusion is recovered at short time scales, [47]

$$\langle \Delta x^2 \rangle = 2Dt \quad \text{for } t \ll L^2/D. \quad (\text{A.14})$$

Bibliography

- [1] M. Bayoumi, S. K. Nomidis, K. Willems, E. Carlon, and G. Maglia, „Autonomous and Active Transport Operated by an Entropic DNA Piston“, *Nano Letters*, vol. 21, no. 1, pp. 762–768, 2021, PMID: 33342212.
- [2] M. A. Watson and S. L. Cockcroft, „Man-made molecular machines: membrane bound“, *Chem. Soc. Rev.*, vol. 45, pp. 6118–6129, 22 2016.
- [3] Y. Sowa and R. M. Berry, „Bacterial flagellar motor“, *Quarterly Reviews of Biophysics*, vol. 41, no. 2, pp. 103–132, 2008.
- [4] J. Tan, X. Zhang, X. Wang, *et al.*, „Structural basis of assembly and torque transmission of the bacterial flagellar motor“, *Cell*, vol. 184, no. 10, 2665–2679.e19, May 2021.
- [5] B. O. Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.
- [6] E. F. Pettersen, T. D. Goddard, C. C. Huang, *et al.*, „UCSF ChimeraX: Structure visualization for researchers, educators, and developers“, *Protein Science*, vol. 30, no. 1, pp. 70–82, 2021.
- [7] M. D. Peraro and F. G. van der Goot, „Pore-forming toxins: ancient, but never really out of fashion“, *Nature Reviews Microbiology*, vol. 14, no. 2, pp. 77–92, Feb. 2016.
- [8] S. Howorka and Z. Siwy, „Nanopore analytics: sensing of single molecules“, *Chem. Soc. Rev.*, vol. 38, pp. 2360–2384, 8 2009.
- [9] C. Dekker, „Solid-state nanopores“, *Nature Nanotechnology*, vol. 2, no. 4, pp. 209–215, Apr. 2007.
- [10] K. Willems, *Computational modeling of biological nanopores*, 2021.
- [11] S. Bhakdi and J. Tranum-Jensen, „Alpha-toxin of *Staphylococcus aureus*“, eng, *Microbiological reviews*, vol. 55, no. 4, pp. 733–751, Dec. 1991, PMC372845[pmcid].
- [12] L. Song, M. R. Hobaugh, C. Shustak, *et al.*, „Structure of Staphylococcal -Hemolysin, a Heptameric Transmembrane Pore“, *Science*, vol. 274, no. 5294, pp. 1859–1865, 1996.
- [13] T. Sugawara, D. Yamashita, K. Kato, *et al.*, „Structural basis for pore-forming mechanism of staphylococcal -hemolysin“, *Toxicon*, vol. 108, pp. 226–231, 2015.
- [14] W. Peng, M. de Souza Santos, Y. Li, D. R. Tomchick, and K. Orth, „High-resolution cryo-EM structures of the *E. coli* hemolysin ClyA oligomers“, *PLoS ONE*, vol. 14, no. 5, pp. 1–17, May 2019.
- [15] M. Mueller, U. Grauschoff, T. Maier, R. Glockshuber, and N. Ban, „The structure of a cytolytic α -helical toxin pore reveals its assembly mechanism“, *Nature*, vol. 459, no. 7247, pp. 726–730, Jun. 2009.

BIBLIOGRAPHY

- [16] S. Benke, D. Roderer, B. Wunderlich, *et al.*, „The assembly dynamics of the cytolytic pore toxin ClyA“, *Nature Communications*, vol. 6, no. 1, p. 6198, Feb. 2015.
- [17] f. Miescher, „Ueber die chemische Zusammensetzung der Eiter-zellen“, *Med.-Chem. Unters*, no. 4, pp. 441–460, 1871.
- [18] O. T. Avery, C. M. MacLeod, and M. McCarty, „Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III“, *Journal of Experimental Medicine*, vol. 79, no. 2, pp. 137–158, Feb. 1944.
- [19] J. D. WATSON and F. H. C. CRICK, „Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid“, *Nature*, vol. 171, no. 4356, pp. 737–738, Apr. 1953.
- [20] O. Kratky and G. Porod, „Röntgenuntersuchung gelöster Fadenmoleküle“, *Recueil des Travaux Chimiques des Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 1949.
- [21] J. L. Finney, „Bernal and the structure of water“, *Journal of Physics: Conference Series*, vol. 57, pp. 40–52, Feb. 2007.
- [22] P. v. G. Kees Vreugdenhil Gerard Alberts, „Waterloopkunde“, *Nieuw Archief voor Wiskunde*, vol. 02, pp. 262–276, Sep. 2001.
- [23] J. D. Bernal, „The Bakerian Lecture, 1962. The Structure of Liquids“, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 280, no. 1382, pp. 299–322, 1964.
- [24] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, „Equation of State Calculations by Fast Computing Machines“, *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [25] C. Sagui and T. Darden, „Multigrid methods for classical molecular dynamics simulations of biomolecules“, *The Journal of Chemical Physics*, vol. 114, no. 15, pp. 6578–6591, 2001.
- [26] P. C. T. Souza, R. Alessandri, J. Barnoud, *et al.*, „Martini 3: a general purpose force field for coarse-grained molecular dynamics“, *Nature Methods*, vol. 18, no. 4, pp. 382–388, Apr. 2021.
- [27] G. S. Freeman, D. M. Hinckley, and J. J. de Pablo, „A coarse-grain three-site-per-nucleotide model for DNA with explicit ions“, *The Journal of Chemical Physics*, vol. 135, no. 16, p. 165 104, 2011.
- [28] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, „DNA Nanotweezers Studied with a Coarse-Grained Model of DNA“, *Phys. Rev. Lett.*, vol. 104, p. 178 101, 17 Apr. 2010.
- [29] A. Biesemans, M. Soskine, and G. Maglia, „A Protein Rotaxane Controls the Translocation of Proteins Across a ClyA Nanopore“, *Nano Letters*, vol. 15, no. 9, pp. 6076–6081, Sep. 2015.
- [30] M. Soskine, A. Biesemans, B. Moeyaert, *et al.*, „An engineered ClyA nanopore detects folded target proteins by selective external association and pore entry“, *Nano letters*, vol. 12, no. 9, pp. 4895–4900, Sep. 2012.
- [31] B. Lu, C. Stokes, M. Fahie, *et al.*, „Protein Motion and Configurations in a Form-Fitting Nanopore: Avidin in ClyA“, *Biophysical Journal*, vol. 115, no. 5, pp. 801–808, Sep. 2018.
- [32] S. Plimpton, „Fast Parallel Algorithms for Short-Range Molecular Dynamics“, *Journal of Computational Physics*, vol. 117, no. 1, pp. 1–19, 1995.

- [33] L. Franceschini, T. Brouns, K. Willems, E. Carlon, and G. Maglia, „DNA Translocation through Nanopores at Physiological Ionic Strengths Requires Precise Nanoscale Engineering“, *ACS Nano*, vol. 10, no. 9, pp. 8394–8402, Sep. 2016.
- [34] T. Ouldridge, „Coarse-grained modelling of DNA and DNA self-assembly“, Ph.D. dissertation, Oxford University, UK, 2011.
- [35] T. E. Ouldridge, P. Sulc, F. Romano, J. P. K. Doye, and A. A. Louis, „DNA hybridization kinetics: zippering, internal displacement and sequence dependence“, *Nucleic Acids Research*, vol. 41, no. 19, pp. 8886–8895, Aug. 2013.
- [36] N. Srinivas, T. E. Ouldridge, P. Sulc, *et al.*, „On the biophysics and kinetics of toehold-mediated DNA strand displacement“, eng, *Nucleic acids research*, vol. 41, no. 22, pp. 10 641–10 658, Dec. 2013, gkt801[PII].
- [37] J. Comer, J. C. Gumbart, J. Hénin, *et al.*, „The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask“, *The Journal of Physical Chemistry B*, vol. 119, no. 3, pp. 1129–1151, Jan. 2015.
- [38] J. K. Whitmer, C.-c. Chiu, A. A. Joshi, and J. J. de Pablo, „Basis Function Sampling: A New Paradigm for Material Property Computation“, *Phys. Rev. Lett.*, vol. 113, p. 190 602, 19 Nov. 2014.
- [39] J. Kästner, „Umbrella sampling“, *WIREs Computational Molecular Science*, vol. 1, no. 6, pp. 932–942, 2011.
- [40] C. Dellago, P. G. Bolhuis, and P. L. Geissler, „Transition Path Sampling“, in *Advances in Chemical Physics*. John Wiley and Sons, Ltd, 2002, ch. 1, pp. 1–78.
- [41] R. J. Allen, D. Frenkel, and P. R. ten Wolde, „Simulating rare events in equilibrium or nonequilibrium stochastic systems“, *The Journal of Chemical Physics*, vol. 124, no. 2, p. 024 102, 2006.
- [42] R. J. Allen, C. Valeriani, and P. R. ten Wolde, „Forward flux sampling for rare event simulations“, *Journal of Physics: Condensed Matter*, vol. 21, no. 46, p. 463 102, Oct. 2009.
- [43] O. Henrich, Y. A. Gutiérrez Fosado, T. Curk, and T. E. Ouldridge, „Coarse-grained simulation of DNA using LAMMPS“, *The European Physical Journal E*, vol. 41, no. 5, p. 57, May 2018.
- [44] A. I. Jewett, D. Stelter, J. Lambert, *et al.*, „Moltemplate: A Tool for Coarse-Grained Modeling of Complex Biological Matter and Soft Condensed Matter Physics“, *Journal of Molecular Biology*, vol. 433, no. 11, p. 166 841, 2021, Computation Resources for Molecular Biology.
- [45] R. L. Davidchack, T. E. Ouldridge, and M. V. Tretyakov, „New Langevin and gradient thermostats for rigid body dynamics“, *The Journal of Chemical Physics*, vol. 142, no. 14, p. 144 114, 2015.
- [46] G. Maglia, A. J. Heron, D. Stoddart, D. Japrung, and H. Bayley, „Chapter 22 - Analysis of Single Nucleic Acid Molecules with Protein Nanopores“, in *Single Molecule Tools, Part B:Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*, ser. Methods in Enzymology, N. G. Walter, Ed., vol. 475, Academic Press, 2010, pp. 591–623.
- [47] T. Bickel, „A note on confined diffusion“, *Physica A: Statistical Mechanics and its Applications*, vol. 377, no. 1, pp. 24–32, 2007.

Acknowledgements

Statistical biophysics group

Carlon

Ouders + Corien

Vrienden.

DEPARTEMENT NATUURKUNDE EN STERRENKUNDE

Celestijnenlaan 200d bus 2412

3000 LEUVEN, BELGIË

tel. + 32 16 32 71 24

fys.kuleuven.be

