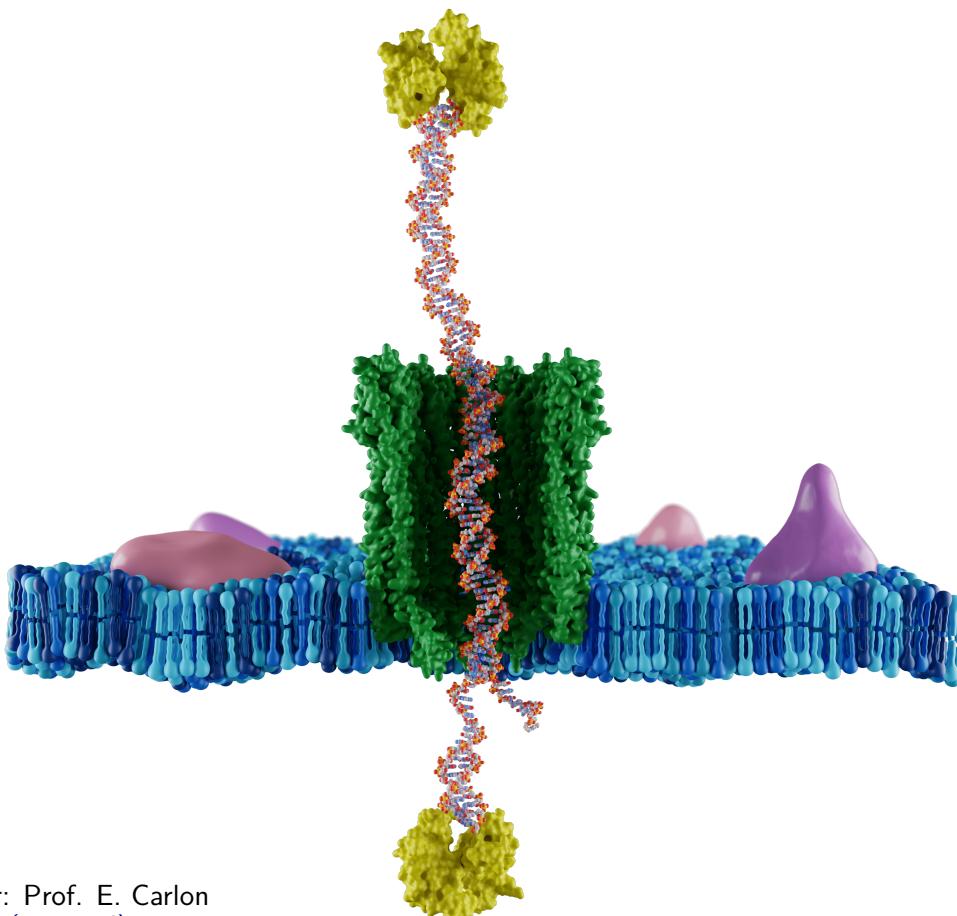


Coarse-grained simulations of the DNA nanopiston

**Jan Stevens**

Supervisor: Prof. E. Carlon
Affiliation (optional)

Tutor: *(optional)*
Affiliation (optional)

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

Academic year 2020-2021

© Copyright by KU Leuven

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

Abstract

abstract

Vulgariserende Samenvatting

Summary in dutch.

asdf

Summary in Layman's Terms

Summary in english.

List of Figures

1.1	write caption	2
1.2	zelf nog maken	3
1.3	zelf nog maken	5
1.4	caption nog maken	5
1.5	Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis)	9
1.6	zelf nog maken	12
3.1	Structure of the OxDNA model with the different interactions. Figure was taken form [].	15

List of Tables

Contents

Abstract	i
Vulgariserende Samenvatting	iii
Summary in Layman's Terms	v
List of Figures	ix
List of Tables	xi
Contents	xiii
1 Introduction	1
1.1 Thesis outline	1
1.2 Biological Nanopores	2
1.3 Deoxyribonucleic acid (DNA)	5
1.4 Polymer Physics	6
1.5 Computer Simulations	9
2 The DNA Nanopiston	13
2.1 Rotaxane Formation	13
2.2 Operating principles	13
2.3 Coarse-grained simulations	13
3 Adapting the Model	15
3.1 OxDNA	15
4 Toehold displacement reaction	17
4.1 Forward Flux sampling	17
4.2 Simulation technique	17
5 Simulations of the Rotaxane	19
5.1 Mixed Rotaxane	19
5.2 Conformational Fluctuations of the Rotaxane	19
5.3 Toehold Displacement Reaction	19
6 Conclusions and Perspectives	21
6.1 Results & Conclusions	21
6.2 Future Perspectives	21
A 1D Confined Diffusion	23
Bibliography	25
Acknowledgements	27

Introduction

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

— Richard P. Feynman, *The Feynman Lectures on Physics*²

1.1 Thesis outline

betere openingszin

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

Most commonly known is the bacterial flagella motor, providing an efficient way for bacteria to roll and tumble through their environment. The flagella consist of a stator connected to the cell membrane and a rotor providing the rotary motion. The work is produced by the flow of cations through the stator, inducing changes in the electrostatic interactions between the two parts of the flagella, generating unidirectional motion.

Whilst being so abundantly present in nature, fabricating synthetic molecular machines turns out to be a difficult task. One of the biggest hurdles in this process arises from the length-scale of these machines. Often times not these structures are not larger than a few nanometres, making the dominant forces result from random thermal fluctuations. These stochastic forces result in the Brownian motion of these molecular machines. Using these random motion no useful work can be extracted, to overcome this hurdle often time synthetic molecular machines are embedded in surfaces like bilipid layers.

dissipating heat, soft large structures. polymers, DNA.

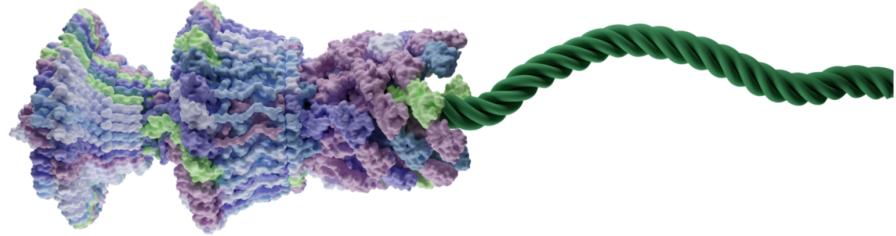


Figure 1.1: write caption

DNA piston can be characterised as a autonomous molecular machine, which turns over chemical fuel to perform work continuously. bayoumi et al.

aim of the molecular machine is to perform selective transport of dna through the membrane. This has been done before following an external bias, but special is that the machine operates also opposing a external bias. The physics that makes this special property possible is entropy and will be discussed futher in detail in later chapters of this thesis.

thesis outline, first chapter is a short introduction to important concepts. chapter two is a discussion of the dna nanopiston, largely based on the paper recently published by bayoumi et al. Next adaptation of the model is discussed in chapter 3. The results of these simulations are discussed in the last chapter.

1.2 Biological Nanopores

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

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

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

1.2.1 α -Hemolysin (α -HL)

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

The α -HL pore(PDBID:...) is an oligomeric complex with multiple naturally occurring variations. The most typical configuration is a heptameric structure, meaning that there are seven protomers found in the complex. The secondary structure elements consist principally of β -sheets, making it 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.

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

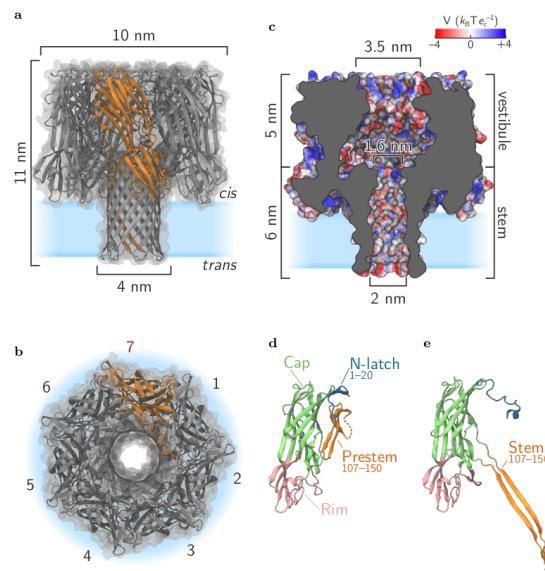


Figure 1.2: zelf nog maken

1.2.2 Cytolysin A (ClyA)

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

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

Structurally the shape of ClyA resembles that of two hollow cylinders stacked on top of each other. This cylinder approximation will be important later on in this thesis, where it will be used to create a simplified model of the nanopore. The total height of the complex is 14nm and the maximum width is measured to be 11nm. The lumen's size of this nanopore differentiates it from the previously discussed α -HL. The cis entrance of the lumen measures 6nm, while the constricted side of the pore is still 3.6nm in diameter. In contrast to the α -HL, the inside surface of ClyA has a net negative charge, making it cation sensitive. This excess charge will induce an important coulomb interaction between the pore and negatively charged analytes.

1.2.3 Ionic current spectroscopy

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

This ion current through the pore can accurately be measured. If the pore is empty we refer to the measured current as the open pore current. However the applied electric field also induces forces upon analytes dissolved in the liquid. The net result of these interactions is a flux of analytes towards and in some cases through the nanopore. Analytes located inside of the nanopore partially block the ion current through the pore, reducing the measured current. Using machine learning algorithms, the time series of these current fluctuations can be measured and identified with particular analytes in the solution. These methods are so precise, that they allow for single cell spectroscopy.

It should be noted that besides these biological nanopores, there are also inorganic

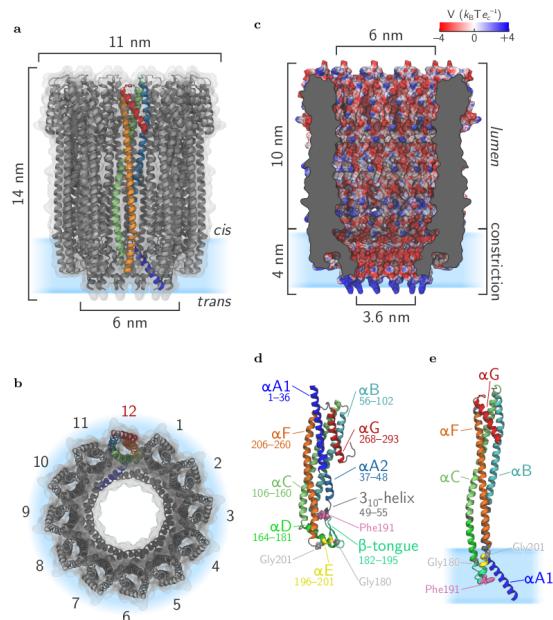


Figure 1.3: zelf nog maken

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

1.3 Deoxyribonucleic acid (DNA)

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



theory of heredity.

The structure of DNA was first determined by Rosalind Franklin using X-ray crystallography. Their research concluded that DNA consists of two individual strands coiled around each other in a double helical structure. Each strand is a chain of monomers, which we call nucleotides. A nucleotide is made up of a deoxyribose sugar, phosphate group and one of four nitrogenous bases: cytosine(C), guanine(G), adenine(A) or thymine(T). The covalent bonds that give both strands structure are formed between consecutive phosphate groups, together they make up the backbone of the strand. 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, forming only A-T and C-G pairs.

Since the binding of the two strands is mediated by hydrogen bonding, the 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, it is characterised by a right-handed helix and the coplanarity between the complementary bases as shown in Fig. A helical twist of B-DNA consists of around 10 bp's having a net helical pitch of 0.34nm. During this thesis when analysing DNA we refer to the B-DNA form.

When Studying DNA the statistical theory polymer physics is a useful tool. A atomistic resolution is not needed to accurately describe processes involving longer length scales. Reducing the complexity of the DNA to the monomers level is often justified, allowing us to use more general results of polymer physics.

1.4 Polymer Physics

A polymer is a biomolecule made up of building blocks called monomers, linked together to form a chain. The configuration of this chain is determined by the position vector of each monomer, denoted as $\{r_0, r_1, \dots, r_N\}$. 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 $|\mathbf{u}_i| = a$.

Various different model can be used to model a polymer, the most simple one is called the Freely Jointed Chain (FJC). This model is an example of an ideal flexible polymer, in which no excluded volume interactions or polymer bending is 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 using the uniform distribution

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

where a is the fixed bond length.

The above described model provides a relatively accurate description of long polymers, however the assumption that consecutive monomers are uncorrelated becomes problematic at small length scales. The Kratky-Porod, or discrete wormlike chain, model solves this problem by taking the energetic cost of bending the polymer into account. Mathematically this is done introducing a bending rigidity between neighbouring bonds in the form of a coupling constant, $\kappa > 0$. Each polymer configuration is assigned an energy using the equation,

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

where $\hat{\mathbf{u}} = \mathbf{u}/a$ is the unit bond-vector and θ_i the angle between the bond-vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_{i+1}$. The lowest energy state of this discrete wormlike chain is a straight rodlike configuration, where the bond angles 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, the partition function factorises into a product of single bond-vector partition as,

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

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

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

From the found partition function we can now determine the bond-vector correlation function. Using the definition of the partition function, we determine the average cosine of

the angle between consecutive bonds to be,

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

Studying the conformation of polymers is often times done by working in the at low temperatures or with a large bending rigidity, κ , 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 an orthonormal basis defined by the normal and tangential directions of the preceding vector $\hat{\mathbf{u}}_{n-1}$ 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 bonds. Performing the factorisation yields

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

here 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 decreasing. The defined quantity represents the characteristic length scale of the polymer over which the correlations between bond-vectors is lost.

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

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

1.5 Computer Simulations

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

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

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

An example of one of these mechanical simulations is the Waterloopkundig Laboratorium or currently the waterloopbos, a scale model of important Dutch waterways, where the influence of waves on harbours and docks was studied. This simulation provided revolutionary insights into the behaviour of water and played an important part in the design of the famous Delta Works.

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

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

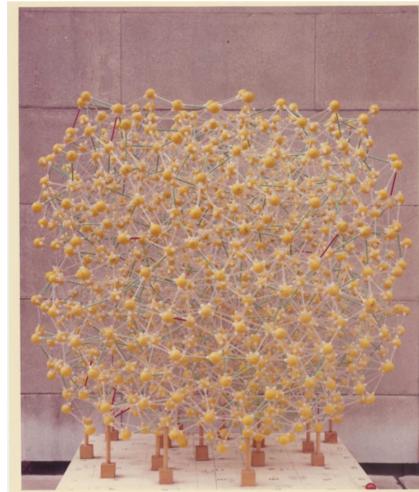


Figure 1.5: Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis)

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

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

1.5.1 Molecular Dynamics Simulations

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

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

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

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

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

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

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

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

Algorithm 1: The Velocity Verlet algorithm

```

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

1.5.2 Coarse Grained modelling

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

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

1. INTRODUCTION

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

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

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

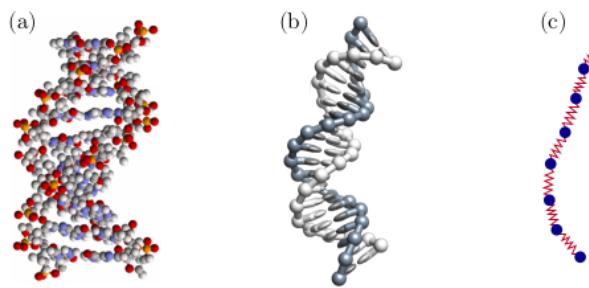


Figure 1.6: zelf nog maken

CHAPTER 2

The DNA Nanopiston

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

— Pierre-Simon Laplace

Duidelijk duiden dat dit hoofdstuk gebaseerd is op de paper van stefanos en bajoumi.

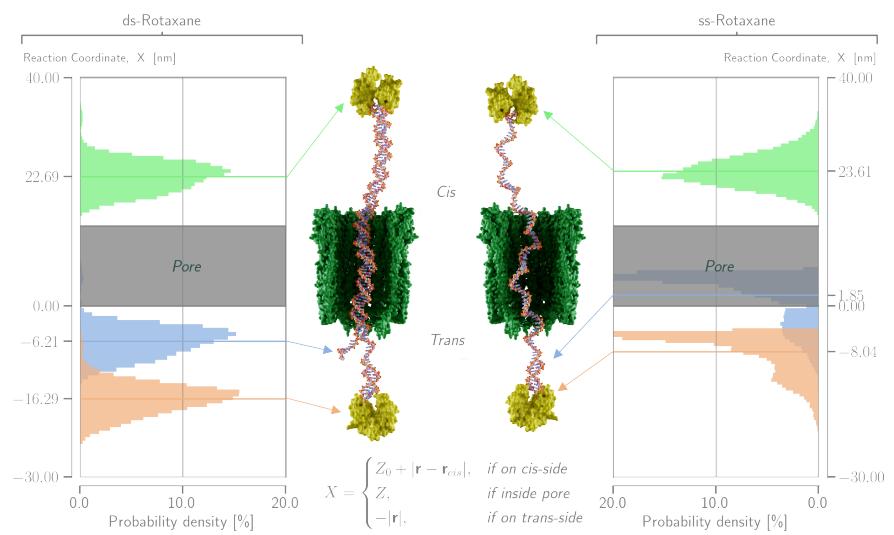
2.1 Rotaxane Formation

2.2 Operating principles

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

2.3 Coarse-grained simulations

2. THE DNA NANOPISTON



CHAPTER 3

Adapting the Model

All models are wrong, but some are useful.

— George Box

3.1 OxDNA

OxDNA is a coarse-grained model of DNA developed by Thomas E. Ouldridge et al. at Oxford university. 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 next compared to experimental data to tweak the interactions, characterising the approach as "top down" coarse-graining. The interactions defined in the OxDNA model can be summarized as,

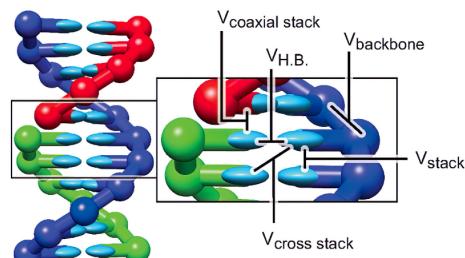


Figure 3.1: Structure of the OxDNA model with the different interactions. Figure was taken from [1].

3. ADAPTING THE MODEL

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

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 interaction site is an excluded volume interaction located at the backbone. These interactions simulate the covalent bonding between consecutive phosphate groups using the FENE (finitely extensible nonlinear elastic) bond type.

The last interaction site is again located at base where it provides a base stacking interaction between consecutive nucleotides. The nucleotides stacking in DNA is crucial for the formation of the characteristic helical structure. Using these stacking interactions this structure is implicitly imposed in the model. Contrasting the common approach of explicitly imposing the Double helix structure in other coarse grained-models like 3SPN and Martini. 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 single stranded DNA strands and the hybridisation reactions play an important role. Since both of these 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 simulate computationally intensive simulations like DNA hybridisation.

CHAPTER 4

Toehold displacement reaction

All models are wrong, but some are useful.

— George Box

4.1 Forward Flux sampling

Use franklin paper.

Local minima, order paramters, measure flux over first order param. Factorise probabilties.
Choose a FFS variant explain the process.

math.

enumerate stappen plan.

show image of trajectories.

4.2 Simulation technique

CHAPTER 5

Simulations of the Rotaxane

All models are wrong, but some are useful.

— George Box

5.1 Mixed Rotaxane

5.2 Conformational Fluctuations of the Rotaxane

5.3 Toehold Displacement Reaction

6

CHAPTER

Conclusions and Perspectives

All models are wrong, but some are useful.

— George Box

6.1 Results & Conclusions

6.2 Future Perspectives

APPENDIX A

1D Confined Diffusion

Studying the dynamics of the mixed rotaxane highlighted the importance of entropic interactions between the nano pore and the DNA strand. Here we observed that a fully double stranded DNA polymer represented a special case. The uniformity of the \mathcal{X} histogram corresponding to this 0 nt mixed rotaxane suggests a free diffusive motion of the rotaxane in a bounded one-dimensional domain. This isotropic behaviour was previously also observed in the bead-spring simulations by Bayoumi et al.¹

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

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

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

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

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

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

The general solution is given by the linear combination,

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

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

As expected, the mean squared distances saturates to $\langle \Delta x^2 \rangle = L^2/6$ in the long-time limit $t \gg L^2/D$.

$$\langle \Delta x^2 \rangle = \frac{L^2}{6} \left[1 - \frac{96}{\pi^4} e^{-\frac{D\pi^2}{L^2} t} \right]$$

Autonomous and active transport operated by an entropic dna piston. Nano Letters, 21(1):762–768. PMID: 33342212.

- [2] Feynman, Richard P. (Richard Phillips), .-. (c1963). The Feynman lectures on physics. Reading, Mass. : Addison-Wesley Pub. Co., c1963-1965. Vol. 2 has subtitle: The electromagnetic field; 3 has subtitle: Quantum mechanics.;Includes bibliographical references and indexes.

- [1] Bayoumi, M., Nomidis, S. K., Willem, K., Carlon, E., and Maglia, G. (2021).

Bibliography

Acknowledgements

...

DEPARTEMENT NATUURKUNDE EN STERRENKUNDE

Celestijnenlaan 200d bus 2412

3000 LEUVEN, BELGIË

tel. + 32 16 32 71 24

fys.kuleuven.be

