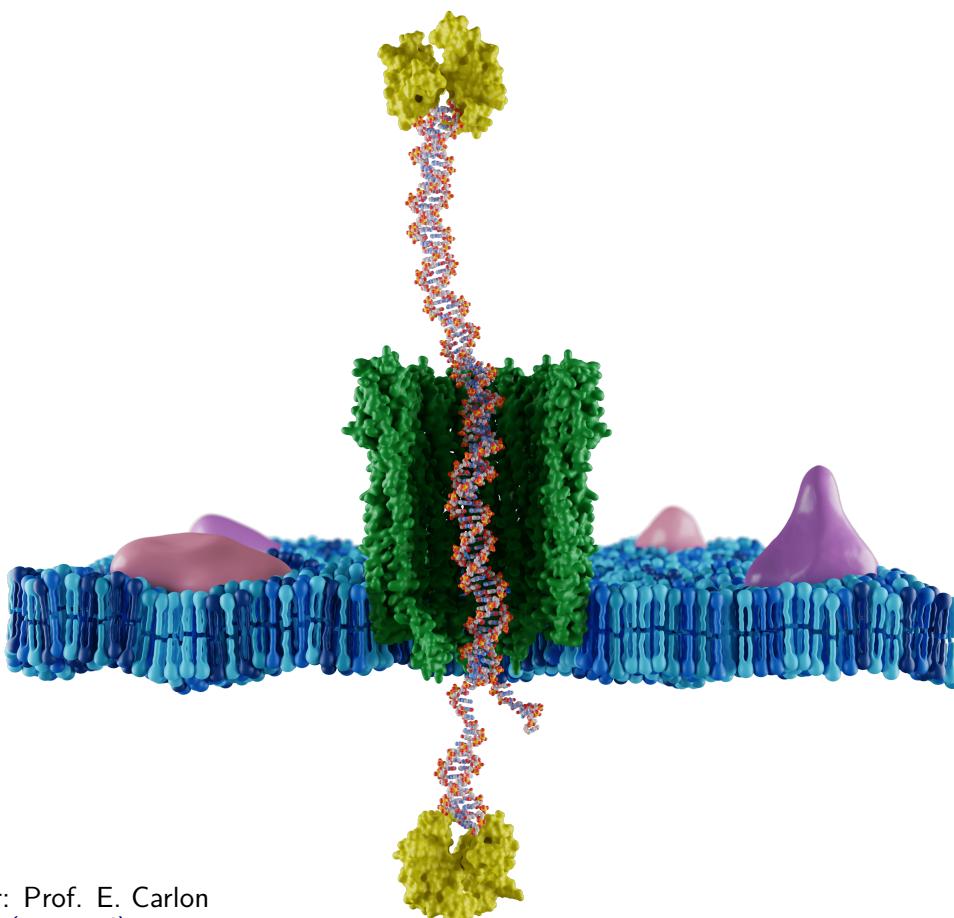


Coarse-grained simulations of the DNA nanopiston

**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	write caption	2
1.3	Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis)	5
1.4	zelf nog namaken met blender!!!	8

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	1
1.3 Polymer Physics	3
1.4 Computer Simulations	5
2 The DNA Nanopiston	9
3 Adapting the Model	11
A 1D Confined Diffusion	13
Bibliography	15
Acknowledgements	17

CHAPTER

1

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

1.2 Biological Nanopores

Biological nanopores are small perforations in the bi lipid membrane, created by a pore forming protein. These proteins are toxins produced by bacteria, as means of killing targeted cells. They work by perforating the cell membrane of a cell, of which the contents is spilled into the environment though osmoses, thereby killing the cell.

The small scale of these protein structures, a few nanometers across,

due to their small scale ideal for spectroscopy.

short introduction to their structural.

should be noted that also solid state nano pores exist poking hole in semi conductor. Prescision is much less so currently bio pores are better. Interesting future.

1. INTRODUCTION

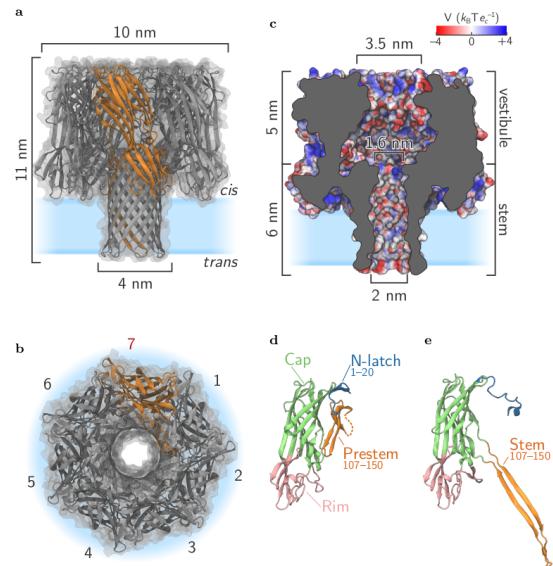


Figure 1.1: write caption

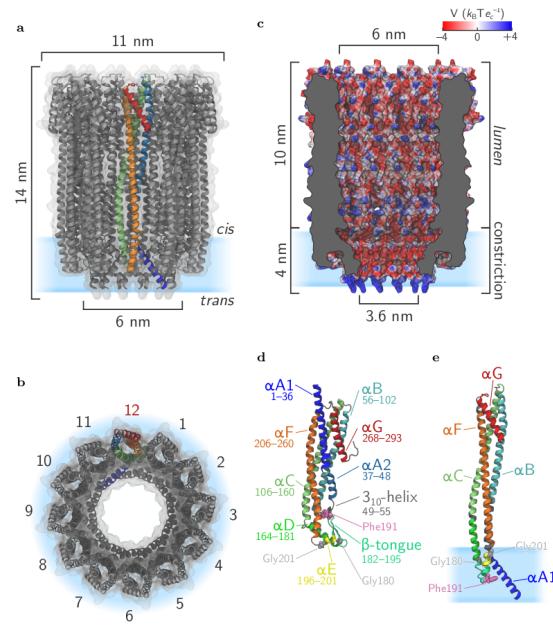


Figure 1.2: write caption

1.2.1 α -Hemolysin (α HL)

1.2.2 Cytolysin A (ClyA)

1.2.3 Ionic current spectroscopy

It should be noted that besides these biological nanopores, there are also solid state nanopore under development. nog niet zo goed als bio pores, maar hebben het voordeel dat ze heel goed afgesteld kunnen worden -> dit is de toekomst!

1.3 Polymer Physics

What is a polymer?

$$E_{WLC} = -K \sum_{i=1}^{N-1} \hat{u}_i \cdot \hat{u}_{i+1} = -K \sum_{i=1}^{N-1} \cos \theta_i, \quad (1.1)$$

$$\begin{aligned} Z_{WLC}(N, T) &= \int d\theta_1 d\theta_2 \dots d\theta_N \sin \theta_1 \sin \theta_2 \dots \sin \theta_N \exp \left[\beta K \sum_{i=1}^N \cos \theta_i \right] \\ &= \left[\int_0^\pi d\theta \sin \theta e^{\beta K \cos \theta} \right]^N \end{aligned} \quad (1.2)$$

$$Z_{WLC}(1, T) = \int_0^\pi d\theta e^{\beta K \theta} = \frac{2 \sinh(\beta K)}{\beta K} \quad (1.3)$$

Using the relation:

$$\langle \cos \theta_{i+1} \rangle = \frac{\partial \log Z_{WLC}(1, T)}{\beta \partial K}, \quad (1.4)$$

one can calculate the mean dot product between two consecutive bonds:

$$\langle \vec{u}_i \cdot \vec{u}_{i+1} \rangle = \langle \cos \theta_{i+1} \rangle = \frac{1}{\tanh(\beta K)} - \frac{1}{\beta K}. \quad (1.5)$$

Which in the limit of $\beta K \gg 1$ leads to:

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

This condition is true for low temperatures or large stiffness between neighbouring bonds (large K). In the following equations the persistence length l_b is mathematically introduced:

$$\langle \vec{t}_1 \cdot \vec{t}_{n+1} \rangle = \langle \vec{t}_1 \cdot \vec{t}_n \rangle \langle \cos \theta_n \rangle = a^2 \langle \cos \theta \rangle^n = a^2 \exp \left[-\frac{na}{l_b} \right], \quad (1.7)$$

1. INTRODUCTION

where $l_b \equiv \frac{aK}{k_b T}$ and a is the constant bond length for the discrete WLC model.

The bending persistence length l_b can be related to the end-to-end vector \vec{R} in the continuum limit. For example can be rewritten using the arc-length parameter s where $0 \leq s \leq L$ and $L = Na$ as:

$$\langle \hat{u}(q) \cdot \hat{u}(q + s) \rangle = e^{-s/l_b}. \quad (1.8)$$

The continuum version of the end-to-end vector is:

$$\vec{R} = \int_0^L \hat{u}(s) ds. \quad (1.9)$$

Taking the square of this expression and averaging over it, we obtain the following:

$$\begin{aligned} \langle \vec{R}^2 \rangle &= \int_0^L ds ds' \langle \hat{u}(s) \cdot \hat{t}(s') \rangle \\ &= 2l_b L \left\{ 1 - \frac{l_b}{L} \left(1 - e^{-L/l_b} \right) \right\}. \end{aligned} \quad (1.10)$$

Explain here two limiting cases of above expression. This leads to a interpretation of what the persistence length means intuitively. -> length over which the correlations between bond vectors are lost.

1.4 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 formalism that are only analytically solvable in few idealised scenarios. Applying these formulas to a problem consisting of just more than two particles already leads 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. 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, this knowledge played an important part in the design of the famous Delta Works.

Another example more topical 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 corresponding to their construct. Bernal in his 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



dit weg laten???
misschien te veel
bla bla

Figure 1.3: Example of an expanded model of a simple liquid (J L Finney, Ph.D thesis)

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. However, . . .

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

1.4.1 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 particle 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 simulations. 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, this choice depends interely on the system at hand. When working with an isolated system – i.e. microcanonical ensemble –, where the total energy is conserved, logically a 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 concerved energy are bounded.

On the otherhand, when a system is in contact with a heatbath –i.e. canonical ensemble–, not the total energy is concerved but rather the temperature of the simulation needs to 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 timestep. Poplular examples of thermostats are the Nosé-Hoover

thermostat or the Langevin thermostat. The latter regulates the temperature by introducing a 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 ← 0 to n − 1 do
4   if isOddNumber(ai) then
5     newList.append(ai + 1)      // Some thought-provoking comment.
6   else
7     // Another comment
8     newList.append(ai + 2)
9   end if
10 end for
11 return newList

```

1.4.2 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 few particles over short time-scales.

During these simulations the most costly calculations involve the non-bonded interaction in the system. These interatomic interactions make the computational complexity for rudimentary MD simulations scale as $O(N^2t)$, here 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 interaction 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 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 a systems thermodynamical properties, rather than the precise small scale dynamics. Here larger pseudo atoms are designed, based upon some structure in the system after which the pseudo interactions are tweaked to accurately reproduces the system's dynamics.

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

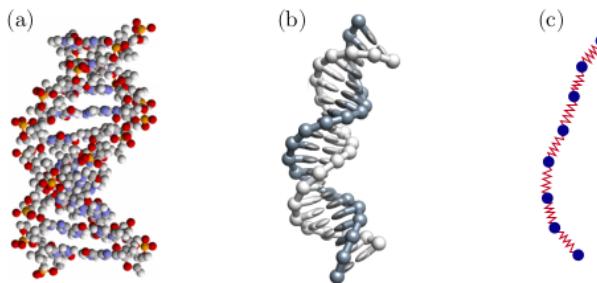


Figure 1.4: zelf nog namaken met blender!!!

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

CHAPTER 3

Adapting the Model

All models are wrong, but some are useful.

— George Box

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

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

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

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,

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

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

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

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

Celestijnlaan 200d bus 2412

3000 LEUVEN, BELGIË

tel. + 32 16 32 71 24

fys.kuleuven.be

