

# 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[,], 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.

**CENTERED**

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.[.] 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 ....

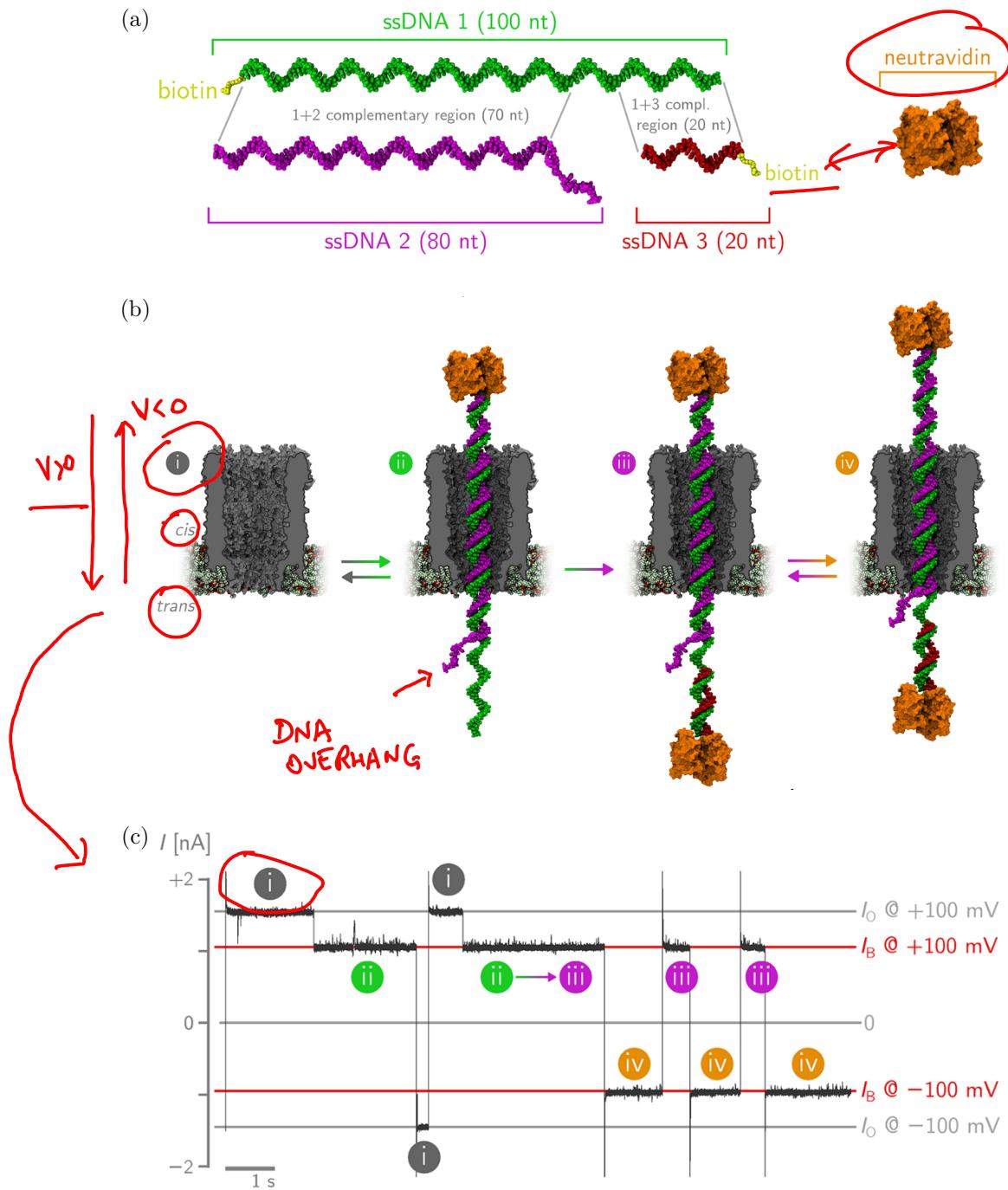
**SINGLE STRANDED DNAs**

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  $+100 \text{ mV}$  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  $+100 \text{ mV}$ . 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.

*EXPLAIN  
COLORS  
OF THE  
STRANDS*

*ARE THESE  
DETAILS  
RELEVANT?*

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.[.] 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.

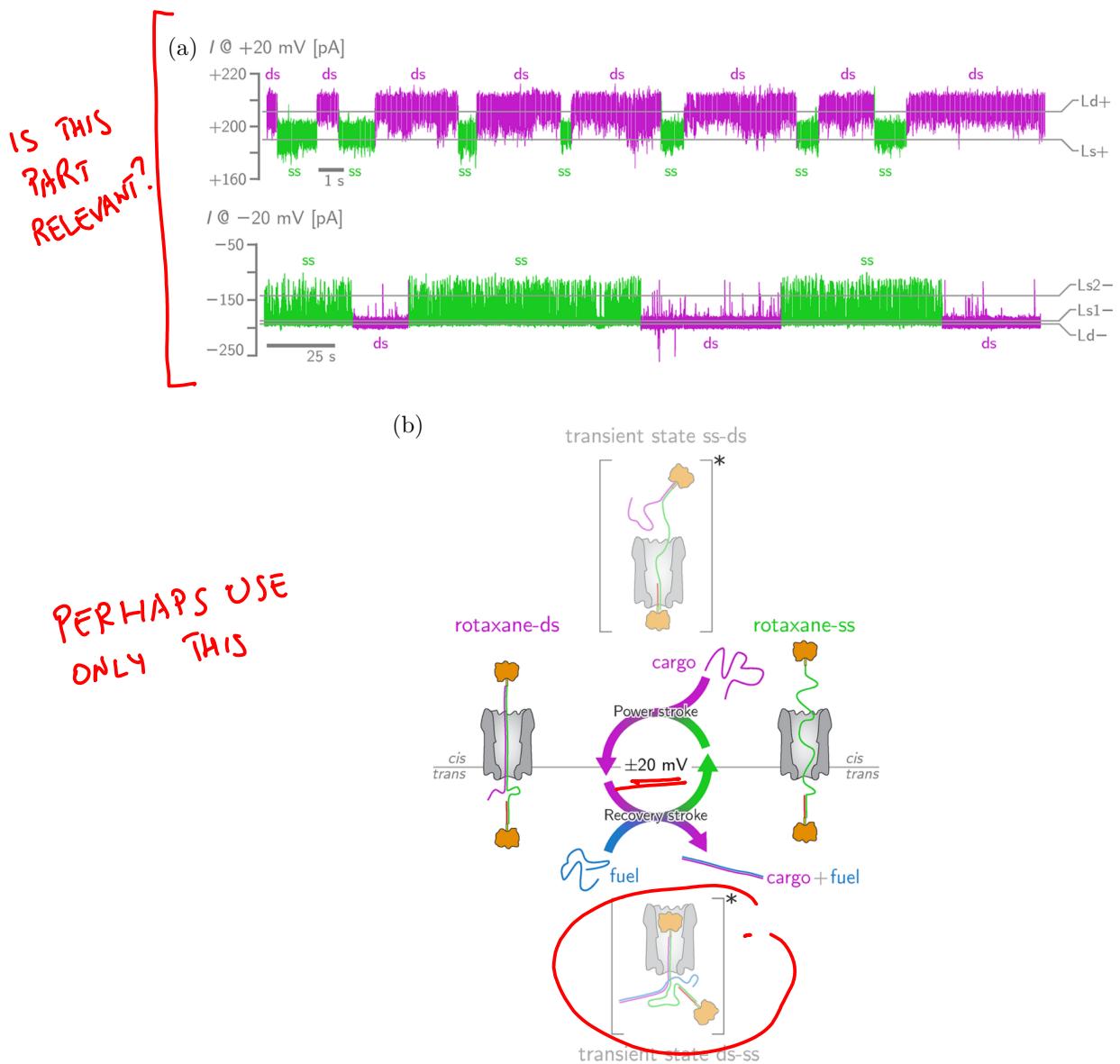
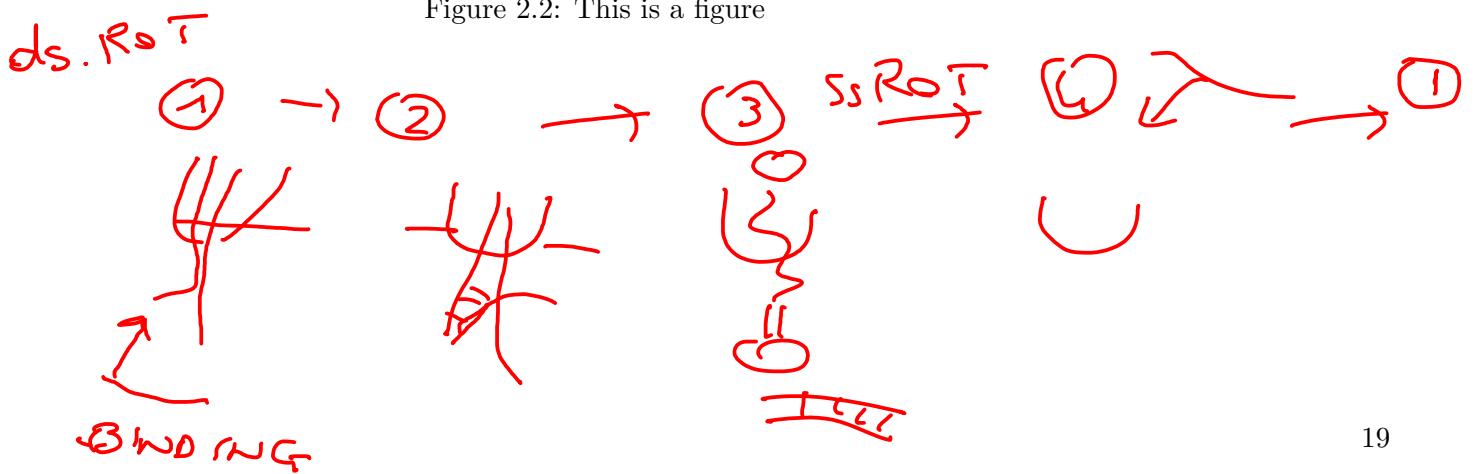


Figure 2.2: This is a figure



## 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 integrator.[.]

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[.] 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

↑  
CITATION ?

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.[.] 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.

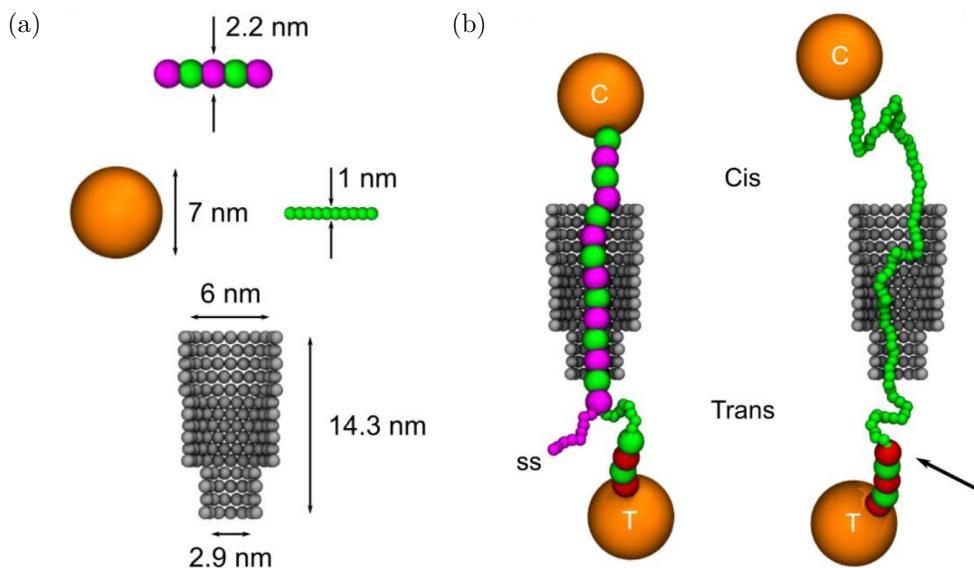


Figure 2.3: This is a figure



# 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.[.] 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[.] en 3SPN[.], where the double helix structure is explicitly constructed. This implicit structure allows for the unstacking of nucleotides, which especially in ssDNA is an important contribution to the flexibility of the strand.

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

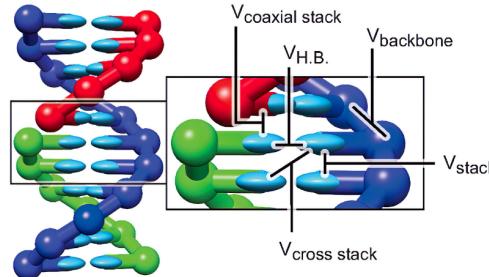


Figure 3.1: Structure of the OxDNA model with the different interactions. Figure was taken from [.]

## 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 fig. ...., 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 increase dramatically when the strand sequences are repetitive, giving rise to hybridisation pathways facilitated by Inchworm and pseudo-knot displacements[.]

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.

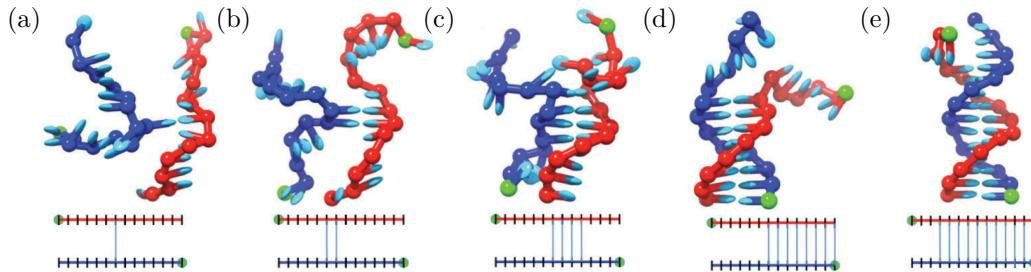


Figure 3.2: This is a figure [.]

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 substrate. It is energetically favourable for the invading strand to disrupt the imperfect duplex

### 3. IMPROVING THE MODEL

---

structure and form a fully WatsonCrick 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[.], shown in figure ... . 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.

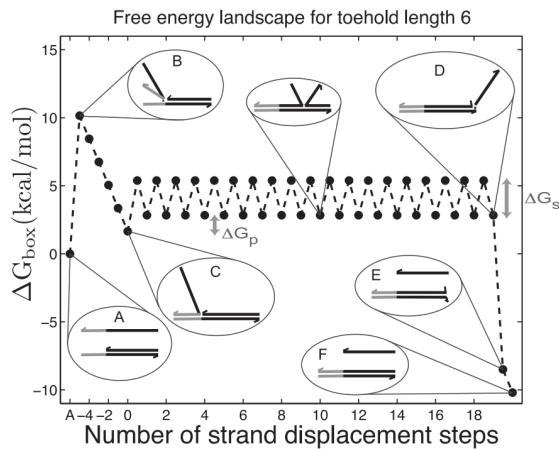


Figure 3.3: write caption[.]

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.[.]

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 of 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[.], basis function sampling[.] and umbrella sampling[.]

The second class of methods, known as path sampling methods, do not influence the systems Hamiltonian, but rather interface directly with the 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[.] and forward flux sampling[.][.]. The latter will be used in our hybridisation simulations, motivated by its relative simplicity.

Forward Flux Sampling (FFS) starts with identifying two local minima,  $A$  and  $B$ , in the energy landscape of our system, for which we want to sample the transition path ensemble. Next an order parameter,  $\lambda(x)$ , is defined with the aim of partitioning the phase space,  $\Omega$ , using a set of nonintersecting hypersurfaces. By design, we choose this order parameter to be a function,  $\lambda(\cdot) : \Omega \rightarrow \mathcal{R}$ , 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,  $\lambda_A$  and  $\lambda_B$ , 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  $\lambda_n$  (i.e. reaching  $B$ ) and  $\langle h_A \rangle$  is the average fraction of time that a trajectory spends in the basin of local minima  $A$ . In the above equation this steady state flux is factorised into the flux of trajectories starting in  $A$  and crossing  $\lambda_0$  and the subsequent probability of reaching the final state from this interface. Using the previously defined interfaces, we can now 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 [zie citation allen review]. The choice is motivated by its resemblance with well known Monte Carlo Simulations and recursive nature, making it easy to implement. This method generates unbranched transition paths from state  $A$  to state  $B$ , making them easy to analyse. The algorithm is described in six steps:

- (i) Generate configurations on the  $\lambda_0$  interface by running simulations in the  $A$  basin. Keeping track of the 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  $\lambda_0$  until they cross  $\lambda_1$  or cross back to  $\lambda_0$ . Store the final configurations of the successful simulations.

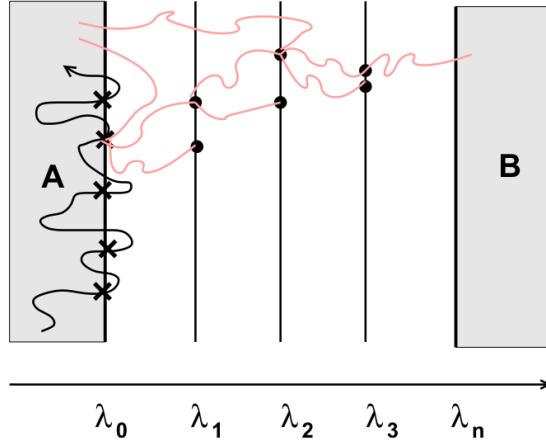


Figure 3.4: write caption[.]

- (iii) Sample one of the saved configuration on the  $\lambda_1$  interface and use it to shoot  $k_1$  runs to the next interface  $\lambda_2$ .
- (iv) Iterate the previous steps until the trajectories reach  $\lambda_n$  or no more configurations are available.
- (v) If not successful, sample a stored configuration on  $\lambda_0$  and repeat the steps (i) to (iv).
- (vi) Finally compute  $P(\lambda_n|\lambda_0)$  using a weighted average of individual transition probabilities as described below.

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

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

where  $S_{j,b}$  is the number of successful trajectories 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[.]. Employing the Lammps implementation of oxDNA developed by Henri et al[.], 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[.], 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. [.], 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*[1]

DISCUSSED IN  
CHAPTER 3

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, composed 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.

AS THEY ARE COMPOSED  
WILL BE  
SHOW A FIGURE  
Out 2out  
---

## 4.1 Mixed RotaxaneS

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  $[-100\text{mV}, +100\text{mV}]$ . 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[. bay] The results from these experiments are presented in Figure....

PERHAPS  
HERE  
THE  
FIGURE

ARE THESE  
ACRONYMS CLEAR  
TO THE READER?

EXPLAIN THAT  
(a) ARE  
EXPERIMENTS  
(b) SIMULATIONS  
(c)

SHIFTED ?

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 where simulated with no bias, i.e.

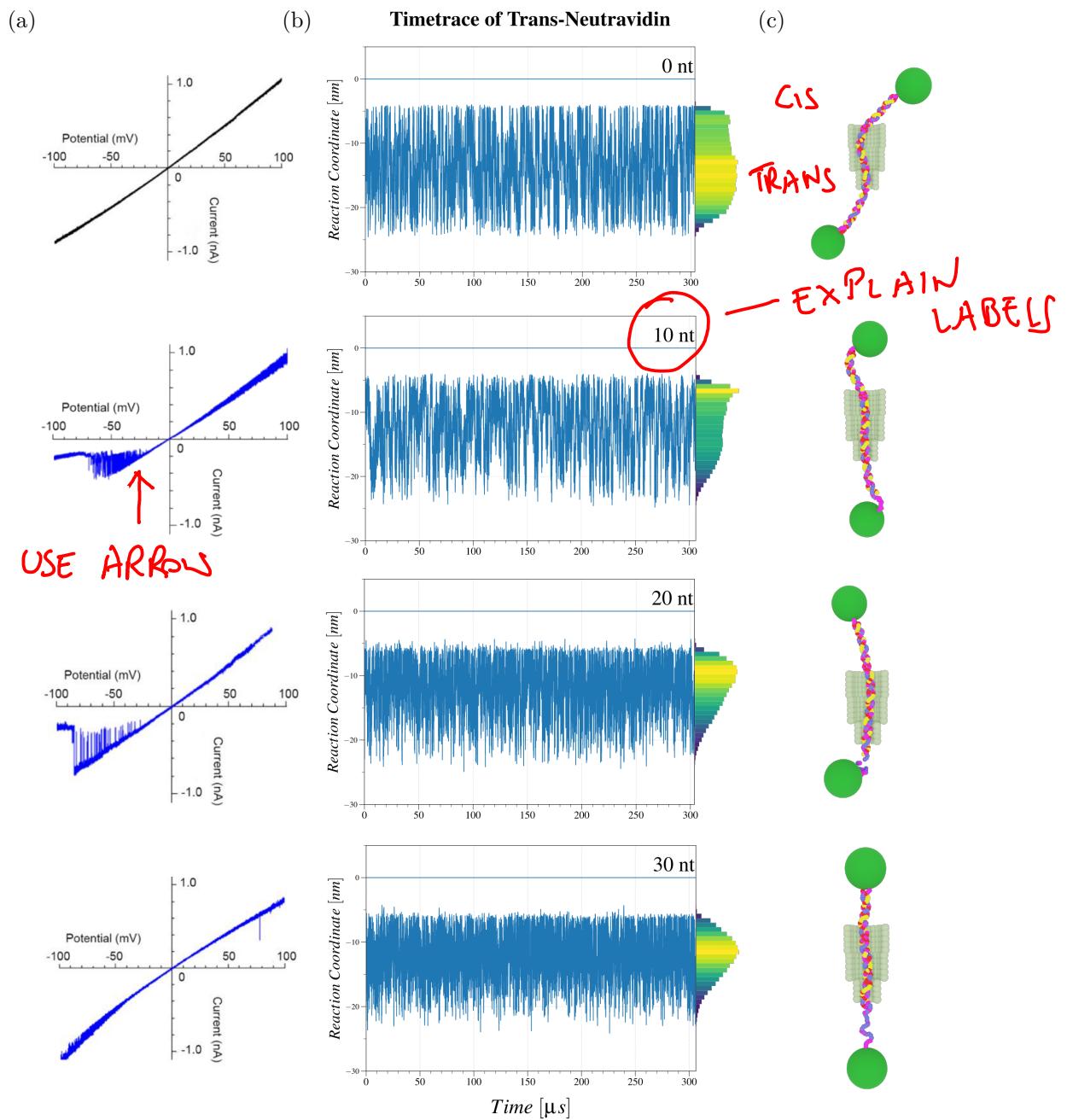


Figure 4.1: This is a figure.

#### 4. SIMULATIONS OF THE ROTAXANE

YOU SHOULD  
INDICATE IT  
WITH AN  
ARROW IN  
FIG.

0mV, to solely highlight the changes in the entropic effects. Quantifying the conformational 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,

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

DEFINED FOR  
ANY NUCLEOTIDE  
OR NEUT.  
IN THE  
SYSTEM!

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  $\mathbf{r}_{cis} = (0, 0, z_0)$ .

OF THE TRANS-STOPPER..

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.

| EXPLAIN  
THE DNA  
ACTUALLY  
BENDS  
(SEE  
FIG 4.1)

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

HERE YOU COULD START A NEW SUBSECTION:  
L.1.2 0nt ROTAXANE AND 1o DIFFUSION

behaviour, the mean square displacements(MSD) of the trans-stopper is determined within a range of time intervals.<sup>1</sup> 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,

$$\langle \Delta x^2 \rangle = \frac{L^2}{6} - \frac{96}{\pi^4} e^{-\frac{D\pi^2 t}{L^2}}, \quad (\text{Eq. Number})$$

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  $O_{nt}$  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,  $69.08 \pm 0.02 \mu\text{s}^{-1}$ , giving us an idea of the speed of the rotaxane dynamics.

FOR DIFFUS  
IN AN  
INFINITE  
DOMAIN

$$\langle \Delta x^2 \rangle \approx 2Dt$$

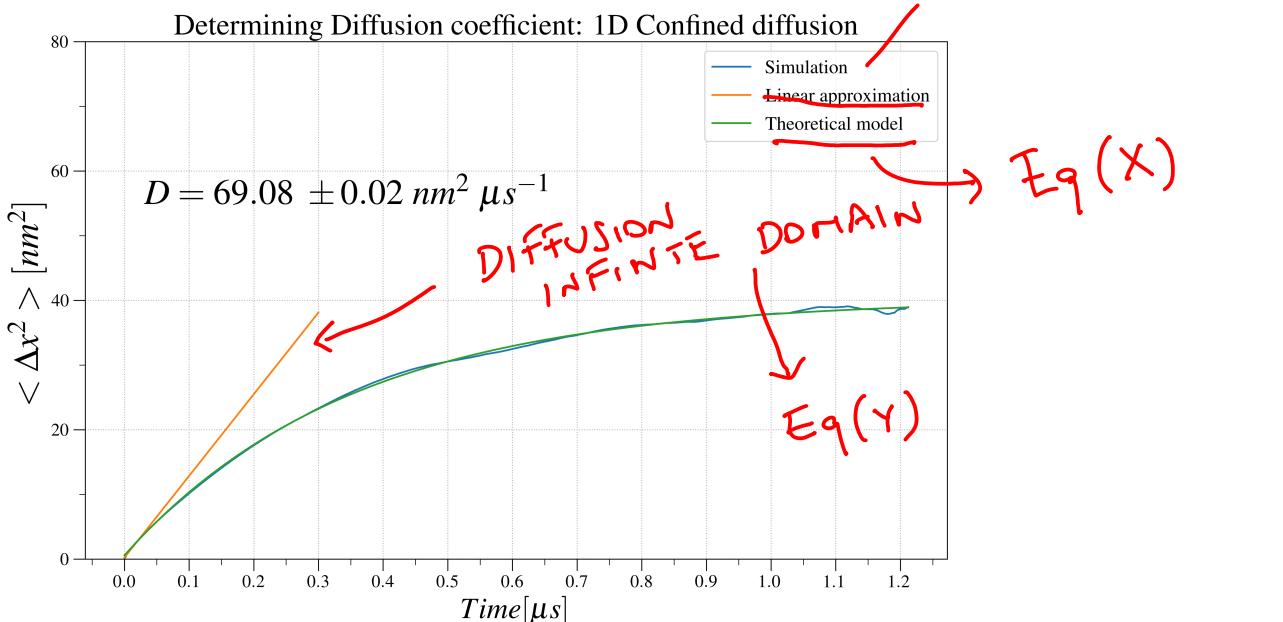


Figure 4.2: write caption

<sup>1</sup>reference tidynamics by Pierre de Buyl

ds AND ss

## 4.2 Conformational Fluctuations of the Rotaxane<sup>s</sup>

IN THE  
MIXED ROT.  
↑

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 ...

YOU SHOULD  
EXPLAIN THAT  
YOU ARE  
ANALYZING  
HISTOGRAMS  
OF X REACT  
COORDINATE  
GIVEN BY  
EQ....



(BLUE HISTOGRAM !)

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.

SAY EXPLICITLY  
THAT HISTOGRAMS  
SHIFTS UPWARD

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.

SIMUL.

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 spring-and-bead 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.

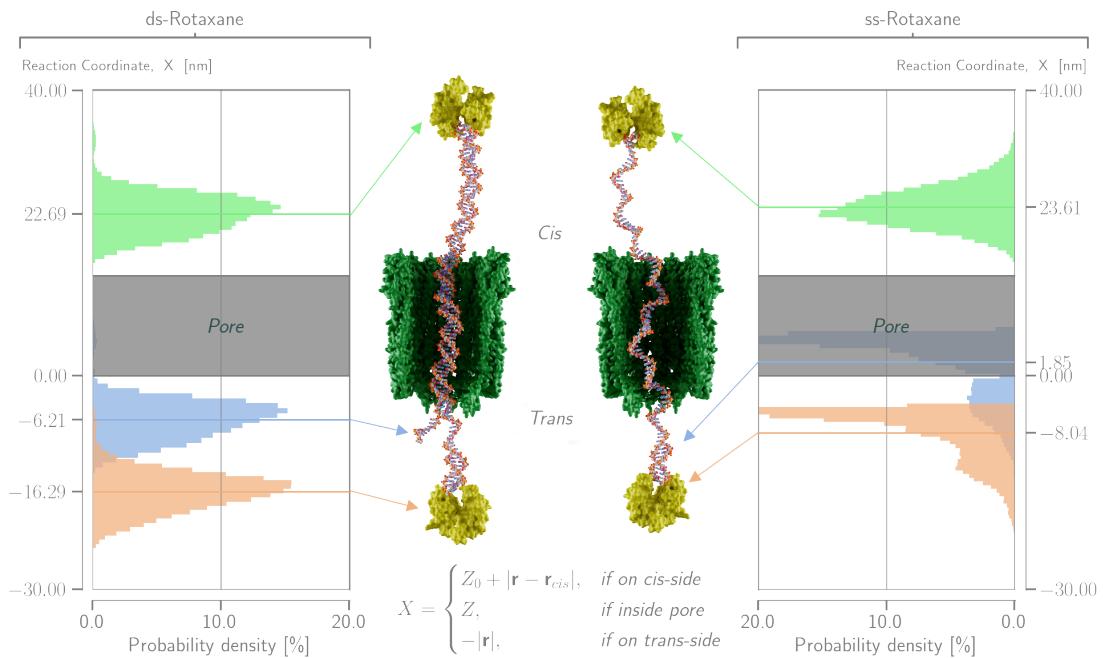


Figure 4.3: write caption

## 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.

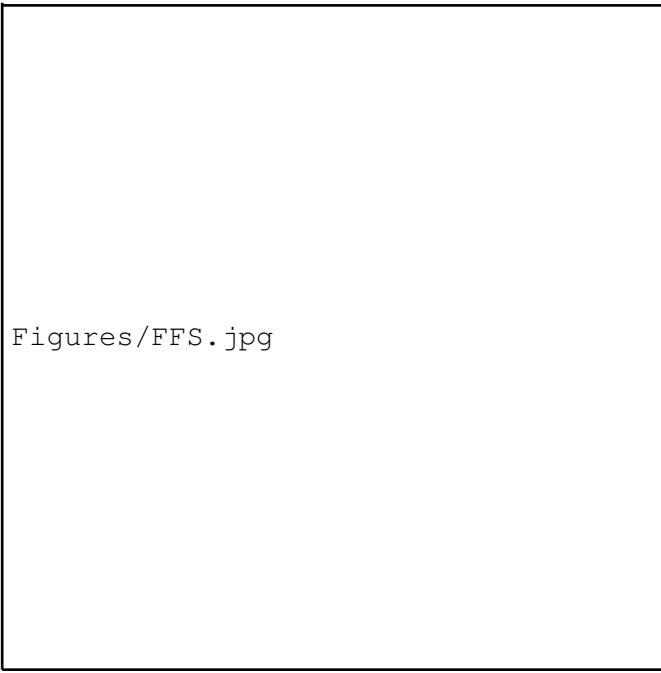
The associated length and time scales of these reactions make their in-depth analysis experimentally challenging. For this reason scientists resort to computational simulations for high resolution analyses of these reactions. To study the DNA hybridisation occurring during the piston operation cycle, 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  $2.9\text{ nm}$ , compared to the  $1\text{ nm}$  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  $\alpha$ -helices constituting the pore's constriction allow for a reasonable amount of structural fluctuations of the constriction.[.] 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.



Figures/FFS.jpg

Figure 4.4: Nog maken